VMware Implementation with IBM System Storage DS5000

- Introduction to VMware design and architecture
- VMware and storage planning and configuration
- VMware Site Recovery Manager implementation

Sangam Racherla
Diego Silva Cogo
Mario David Ganem
Hartmut Lonzer
Minja Obradovic
Gabriel Requejo

ibm.com/redbooks

Redbooks

International Technical Support Organization

VMware Implementation with IBM System Storage DS5000

June 2013

SG24-8052-00

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

First Edition (June 2013)

This edition applies to VMware vSphere ESXi 5, IBM Midrange Storage DS5000 running V7.8x firmware, and IBM System Storage DS Storage Manager V10.8x.

Contents

Notices

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

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

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

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

Trademarks

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

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

AIX®
BladeCenter®
DB2®
DS4000®
DS8000®
FlashCopy®
IBM®
Power Systems™
Redbooks®
Redbooks (logo) ®
System p®
System Storage DS®
System Storage®
System x®
Tivoli®
XIV®
xSeries®

The following terms are trademarks of other companies:

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

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

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

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

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

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

Preface

In this IBM® Redbooks® publication, we compiled best practices for planning, designing, implementing, and maintaining IBM Midrange Storage Solutions. In this publication, we use IBM System Storage® DS5000 storage subsystem for the implementation procedures, and the same procedures can be used for implementations with DCS3700 or DS3500 storage subsystems. We also compiled configurations for a VMware ESX and VMware ESXi Server host environment.

Setting up an IBM Midrange Storage Subsystem is a challenging task. Our principal objective in this book is to provide you with a sufficient overview to effectively enable storage area network (SAN) storage and VMware. There is no single configuration that is satisfactory for every application or situation. However, the effectiveness of the VMware implementation is enabled by careful planning and consideration. Although the compilation of this publication is derived from an actual setup and verification, we did not stress test or test for all possible use cases that are used in a limited configuration assessment.

Because of the highly customizable nature of a VMware ESXi host environment, you must consider your specific environment and equipment to achieve optimal performance from an IBM Midrange Storage Subsystem. When you are weighing the recommendations in this publication, you must start with the first principles of input/output (I/O) performance tuning. Each environment is unique and the correct settings that are used depend on the specific goals, configurations, and demands for the specific environment.

This publication is intended for technical professionals who want to deploy VMware ESXi and VMware ESX Servers with IBM Midrange Storage Subsystems.

Authors

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

Sangam Racherla is an IT Specialist and Project Leader working at the ITSO in San Jose, CA. He has 12 years of experience in the IT field and has been with the ITSO for the past eight years. Sangam has extensive experience in installing and supporting the ITSO lab equipment for various IBM Redbooks publications projects. He has expertise in working with Microsoft Windows, Linux, IBM AIX®, IBM System x®, and IBM System p® servers, and various SAN and storage products. Sangam holds a degree in Electronics and Communication Engineering.

Diego Silva Cogo is an IT professional, specialized in cloud and virtualization solutions. He has more than 11 years of experience in the IT industry. Diego resides in Porto Alegre, Brazil, where he works as a VMware IT Specialist for the Gerdau Long Steel North America Strategic Outsourcing account. He holds a Network Technology Bachelor's Degree and more than 20 Microsoft certifications.

Mario David Ganem is an IT professional, specialized in cloud computing and storage solutions. He has 15 years of experience in the IT industry. Mario resides in Buenos Aires, Argentina, where he works as an Infrastructure IT Architect in the IBM Global Delivery Center. Prior to starting his career in IBM in 2006, Mario worked in many companies, such as Hewlett-Packard, Compaq, and Unisys. He has developed the internal virtualization products curricula training, which he is teaching to DCA professionals. He holds, among others, many industry certifications from Microsoft, RedHat, VMware, Novell, Cisco, CompTIA, HP, and Compaq.

Hartmut Lonzer is a Technical Consultant in the Partnership Solution Center Southwest/Germany. As a former Storage FTSS member, his main focus is on Storage and System x. He is responsible for educating and supporting the IBM Business Partners and clients in his area in technical matters. His experience regarding DS Storage goes back to the beginning of this product. He has been with IBM in various technical roles for 35 years.

Minja Obradovic is an IT Specialist in Technical Support Services in IBM Serbia since 2007. He is a member of CEEMEA Virtual Front End Midrange Storage Support team providing Storage systems support in Central Eastern Europe, Middle East, and Africa. His primary focus is post-sales Storage Virtualization, SAN, and Midrange Storage support. His areas of expertise include SAN switches (Brocade and Cisco MDS), SAN Storage (DS3000, IBM DS4000®, V7000, and IBM XIV®), IBM xSeries®, and IBM Power Systems™.

Gabriel Requejo is an IT professional with more than 15 years of experience in the IT field. For the last four years, he has worked on several Virtualization projects for different companies around the world using VMware products with different IBM Storage solutions. Before this, Gabriel worked for several domestic and international clients providing Citrix and Windows solutions using IBM System x servers. Gabriel also has experience on AIX, Linux, Novell, and Microsoft products.

Thanks to the following people for their contributions to this project:

Jon Tate
David Bennin
David Watts
Tamikia Barrow
Shari Deiana

IBM International Technical Support Organization

Dahn Le
Richard Mancini
Gene Cullum
Alexander Watson
John Sexton
Mousa Hammad
Kerry Anders
Jason M Schwarz
Patrick Caporale
Jim Gallagher

IBM

Doug Merrill
John Bish
David Worley
Stephen Carl

NetApp

Brian Steffler
Steven Tong

Brocade Communications Systems, Inc.

Simon Pfau

VMware (Germany)

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Part 1

Planning

In Part 1, we provide the conceptual framework for understanding IBM Midrange Storage Systems in a storage area network (SAN) and vSphere environment. We include recommendations, hints, and tips for the physical installation, cabling, and zoning. Even though performance figures are not included, we discuss the performance and tuning of various components and features to guide you when you are working with IBM Midrange Storage.

Before you start any configuration of the IBM Midrange Storage Subsystem in a VMware vSphere environment, you must understand the following concepts to guide you in your planning:

- Recognizing the IBM Midrange Storage Subsystem feature set
- Balancing drive-side performance
- Understanding the segment size of logical drives
- Knowing about storage system cache improvements
- Comprehending file system alignment
- Knowing how to allocate logical drives for vSphere ESXi hosts
- Recognizing server hardware architecture
- Identifying specific vSphere ESXi settings

Assistance in planning for the optimal design of your implementation is provided in the next chapters.

1

Introduction of IBM VMware Midrange Storage Solutions

In this chapter, we introduce you to the IBM VMware Midrange Storage Solutions and provide an overview of the components that are involved.

1.1 Overview of IBM VMware Midrange Storage Solutions

Many enterprises have implemented VMware or plan to implement it. VMware provides more efficient use of assets and lower costs by consolidating servers and storage. Applications that previously ran in under-utilized dedicated physical servers are migrated to their own virtual machine (or virtual server) that is part of a VMware ESX cluster or a virtual infrastructure.

As part of this consolidation, asset utilization typically can be increased from under 10% to over 85%. Applications that previously had dedicated internal storage now can use a shared networked storage system that pools storage to all of the virtual machines and their applications. Backup, restore, and disaster recovery become more effective and easier to manage. Because of the consolidated applications and their mixed workloads, the storage system must deliver balanced performance and high performance to support existing IT service-level agreements (SLAs). The IBM Midrange Storage Systems provide an effective means to that end.

IBM Midrange Storage Systems are designed to deliver reliable performance for mixed applications including transaction and sequential workloads. These workloads include applications that are typical of a virtual infrastructure, including email, database, web server, file server, data warehouse, and backup profiles. IBM offers a complete line of storage systems from entry-level systems to midrange systems to enterprise-level systems that are certified to work with VMware vSphere ESX Server.

The IBM Midrange Storage systems that are discussed in this publication include the DS5100, DS5300, and DS5020 models. These systems are included in the references throughout the manuals as DS-Series. We discuss these storage subsystems in greater detail later in Chapter 3, “Planning the VMware vSphere storage system design” on page 29.

All of these systems offer shared storage that enables all of VMware’s advanced functionality:

- vSphere Distributed Resource Scheduler (DRS)
- vCenter Site Recovery Manager (SRM)
- vSphere High Availability (HA)
- vSphere Fault Tolerance (FT)
- vSphere Virtual Machine File System (VMFS)
- vSphere vMotion
- VMware vSphere Storage vMotion

The IBM DS5000 storage systems include the following features:

- IBM DS5000 storage systems offer the highest performance and the most scalability, expendability, and investment protection that is available in the IBM Midrange portfolio.
- IBM DS5000 storage subsystem offers enterprise-class features and availability. This storage system can handle the largest and most demanding virtual infrastructure workloads.
- IBM DS5000 storage systems support up to 448 drives (1.34 PB) using the EXP5000 enclosure and up to 480 drives (1.44 PB) of high-density storage with the EXP5060 enclosure.

VMware vCenter Site Recovery Manager 5 (SRM) is supported on IBM DS3500, IBM DCS3700, and DS5000.

1.2 IBM VMware storage solutions

Many companies consider and employ VMware virtualization solutions to reduce IT costs while increasing the efficiency, utilization, and flexibility of their hardware. In fact, 100,000 clients have deployed VMware, including 90% of Fortune 1000 businesses. Yet maximizing the operational benefits from virtualization requires network storage that helps optimize the VMware infrastructure.

The IBM Midrange Storage Solutions for VMware offer clients:

- Flexibility: Support for iSCSI and Fibre Channel shared storage, plus host bus adapter (HBA) and storage port multi-pathing and boot from SAN.
- Performance: Outstanding high-performance block-level storage that scales with the VMware virtual machine file system (VMFS), independently verified high performance by the Storage Performance Council SPC-1 and SPC-2 benchmarks, and balanced performance delivered by the IBM Midrange Storage Systems for mixed applications running in a virtual infrastructure.
- Horizontal scalability: From entry-level through midrange to enterprise-class network storage with commonality of platform and storage management.
- Hot Backup and Quick recovery: Non-disruptive backup solutions using IBM Tivoli® and NetBackup with and without VMware vStorage APIs for Data Protection, providing quick recovery at the file or virtual machine level
- Disaster recovery: DS5000 Enhanced Remote Mirror that offers affordable disaster recovery with automatic failover with VMware vCenter Site Recovery Manager 5 (SRM).
- Affordability: Low total cost of ownership (TCO) shared storage with included IBM Storage Manager Software and no separate software maintenance fees, cost-effective tiered storage within the same storage system, leveraging Fibre Channel drives for high performance and Serial Advanced Technology Attachment (SATA) drives for economical capacity.
- Efficiency: Data Services features, such as IBM FlashCopy® and VolumeCopy, enable VMware Centralized Backup to disk to eliminate backup windows, and provide required network storage for VMware ESX Server features, such as VMware vSphere vMotion, VMware vSphere Storage vMotion, VMware vSphere Distributed Resource Scheduler (DRS), and VMware vSphere High Availability (HA).

VMware vSphere includes components and features that are essential for managing virtual machines. The following components and features form part of the VMware vSphere suite:

- vSphere ESXi
- vSphere vCenter Server
- vSphere VMFS
- vSphere Fault Tolerance (FT)
- vSphere vMotion
- vSphere High Availability (HA)
- vSphere Distributed Resource Scheduler (DRS)
- vSphere Storage vMotion (sVMotion)
- vSphere Storage DRS
- Profile-Driven Storage
- vSphere Distributed Power Management (DPM)
- vSphere Storage I/O control (SIOC)
- vSphere Network I/O control

1.2.1 VMware vSphere ESXi architecture

VMware vSphere ESXi is virtual infrastructure partitioning software that is designed for server consolidation, rapid deployment of new servers, increased availability, and simplified management. It helps to improve hardware utilization and save space, IT staffing, and hardware costs.

VMware vSphere virtualizes the entire IT infrastructure, including servers, storage, and networks. It groups these heterogeneous resources and transforms the rigid, inflexible infrastructure into a simple and unified manageable set of elements in the virtualized environment. With vSphere, IT resources can be managed like a shared utility and quickly provisioned to different business units and projects without worrying about the underlying hardware differences and limitations.

Many people might have had earlier experience with VMware's virtualization products in the form of VMware Workstation or VMware Server. VMware vSphere ESXi is quite different from other VMware products. It runs directly on the hardware. It is considered a bare-metal solution, offering a mainframe-class virtualization software platform that enables the deployment of multiple, secure, independent virtual machines on a single physical server.

VMware vSphere ESXi allows several instances of operating systems, such as Microsoft Windows Server, Red Hat, SuSE Linux, Mac OS, and more, to run in partitions that are independent of one another. Therefore, this technology is a key software enabler for server consolidation. You can move existing, unmodified applications and operating system environments from many older systems onto a few new high-performance System x platforms.

Real cost savings can be achieved by reducing the number of physical systems to manage. Reducing the number of systems saves floor space and rack space and reduces power consumption. It eliminates the headaches that are associated with consolidating dissimilar operating systems and applications that require their own OS instance.

The architecture of VMware vSphere ESXi is shown in Figure 1-1 on page 7.

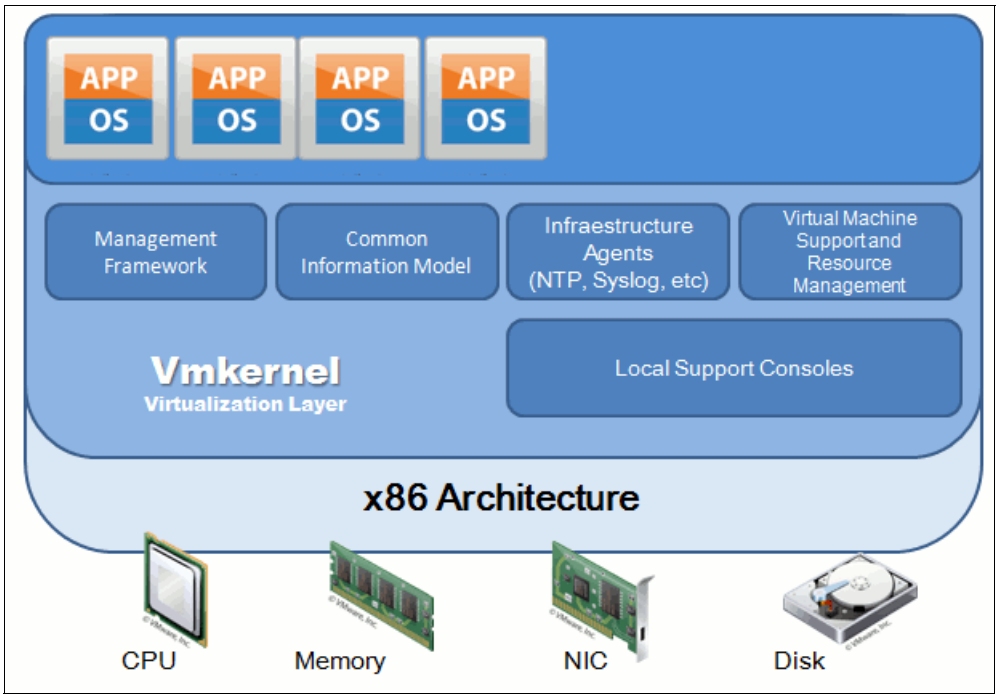

Figure 1-1 VMware vSphere ESXi architecture

VMware vSphere ESXi and vSphere vCenter Server help you to build cost-effective, high-availability solutions by using failover clustering between virtual machines. Until now, system partitioning (the ability of one server to run multiple operating systems instances simultaneously) has been the domain of mainframes and other large midrange servers. But with VMware vSphere products, dynamic logical partitioning can be enabled on IBM System x systems.

Instead of deploying multiple servers that are scattered around a company and running a single application on each server, the servers can be consolidated physically and simultaneously enhance system availability. VMware Hypervisor (vSphere ESXi) allows to each server to run multiple operating systems and applications in virtual machines, providing centralized IT management. Because these virtual machines are isolated from the other machines, if one machine fails, it does not affect the others. Not only is VMware ESXi software great for optimizing hardware usage, it can also offer the added benefits of higher availability and scalability.

1.2.2 Overview of using VMware vSphere with SAN

A storage area network (SAN) is a highly effective solution for supporting and provisioning VMware products. SAN configuration requires careful consideration of components, including host bus adapters (HBAs) on the host servers, SAN switches, storage processors, disks, and storage disk arrays. A SAN topology has at least one switch present to form a SAN fabric. SAN can provide high-performance characteristics and feature functions, such as FlashCopy, Volumecopy and mirroring.

1.2.3 Benefits of using VMware vSphere with SAN

By using SAN with VMware vSphere, you can improve data accessibility and system recovery.

Using SAN with VMware vSphere also offers these benefits:

- SAN with VMware vSphere effectively stores data redundantly and eliminates single points of failure.
- Data centers can quickly negotiate system failures.
- The VMware ESXi hypervisor provides multipathing by default and automatically supports virtual machines.
- SAN with VMware ESXi systems extends failure resistance to servers.
- VMware vSphere with SAN makes high availability and automatic load balancing affordable for more applications than if dedicated hardware is used to provide standby services.
- Because shared main storage is available, building virtual machine clusters that use Microsoft Cluster Server (MSCS) becomes possible.
- If virtual machines are used as standby systems for existing physical servers, shared storage is essential and a viable solution.
- vSphere vMotion capabilities migrate virtual machines seamlessly from one host to another.
- Solutions that use vSphere High Availability (HA), with a SAN for a cold standby, help to ensure an immediate and automatic failure response.
- vSphere Distributed Resource Scheduler (DRS) helps ensure more efficient resource utilization.
- Inside of a DRS Cluster, a VMware ESXi host can be moved into maintenance mode and the system migrates all the running virtual machines to another VMware ESXi host automatically.
- vSphere Storage vMotion can be used as a storage tiering tool to move data to another datastore and type of storage platform. Virtual machine storage disks can be relocated to different locations with no downtime. They are transparent to the virtual machine or the user.
- SAN high-performance characteristics and feature functions empower your Disaster Recovery solution when you use vCenter Site Recovery Manager (SRM).

The transportability and encapsulation of VMware virtual machines complement the shared nature of SAN storage. When virtual machines are on SAN-based storage, you can shut down a virtual machine on one server and power it up on another server. Or, you can suspend it on one server and resume operation on another server on the same network in a matter of minutes. With this ability, you can migrate computing resources and maintain consistent shared access.

1.2.4 VMware vSphere and SAN use cases

Using VMware vSphere with SAN is effective for the following tasks:

- Maintenance with zero downtime: When performing maintenance, use vSphere DRS or VMware vMotion to migrate virtual machines to other servers.
- Load balancing: Use vSphere vMotion or vSphere DRS to migrate virtual machines to other hosts for load balancing.
- Storage consolidation and simplification of storage layout: Host storage is not the most effective method to use storage available. Shared storage is more manageable for allocation and recovery.
- Disaster recovery: Storing all data on SAN can greatly facilitate remote storage of data backups.

1.3 Overview of VMware vStorage APIs for Data Protection

vStorage APIs for Data Protection are the next generation of VMware's data protection framework that enables backup products to perform centralized, efficient, off-host LAN-free backup of vSphere virtual machines. This feature was introduced with vSphere 4.0 to replace the old backup integrated solution, which was known as VMware Consolidated Backup (VCB).

The schema for VMware vStorage APIs for Data Protection is shown in Figure 1-2.

.

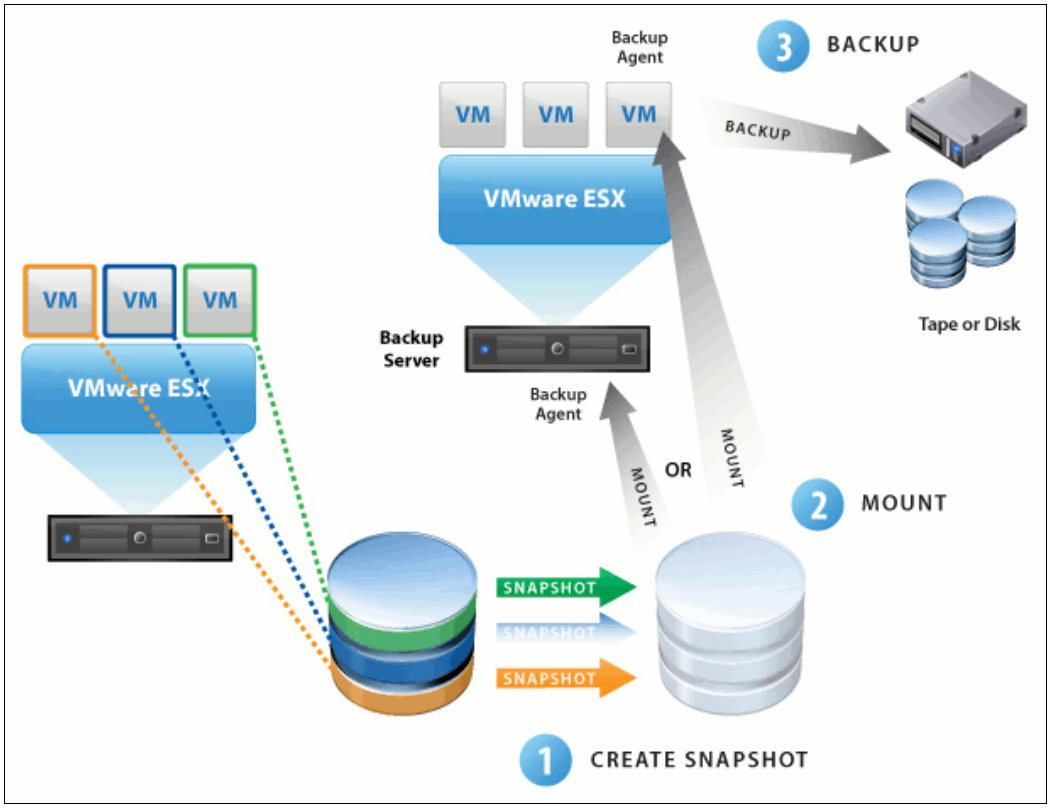

Figure 1-2 VMware vStorage APIs for Data Protection

With vStorage APIs for Data Protection you can perform these tasks:

- Integrate with existing backup tools and technologies that are already in place.
- Perform full and incremental file backups of virtual machines.
- Perform full image backup of virtual machines.
- Centrally manage backups to simplify management of IT resources.

Improve performance with Centralized Virtual Machine Backup

Centralized Virtual Machine Backup helps you eliminate backup traffic from your network to improve the performance of production virtual machines:

- Eliminating backup traffic with LAN-free virtual machine backup utilizing tape devices
- Reducing the load on the VMware vSphere ESXi and allowing it to run more virtual machines

vStorage APIs for Data Protection leverage the snapshot capabilities of VMware vStorage VMFS to enable backup across SAN without requiring downtime for virtual machines. As a result, backups can be performed non-disruptively at any time of day without requiring extended backup windows and the application and user downtime that is associated with backup windows.

vStorage APIs for Data Protection are designed for all editions of vSphere. It is supported by many backup products, such as IBM Tivoli Storage managerSM, Symantec NetBackup, CA ArcServe, and VizionCore vRanger, among others.

For additional information, see the following web location:

http://www.vmware.com/products/vstorage-apis-for-data-protection/overview.html

1.4 Overview of VMware vCenter Site Recovery Manager (SRM)

VMware vCenter Site Recovery Manager (SRM) provides business continuity and disaster recovery protection for virtual environments. Protection can extend from individual replicated datastores to an entire virtual site. VMware's virtualization of the data center offers advantages that can be applied to business continuity and disaster recovery.

The entire state of a virtual machine (memory, disk images, I/O, and device state) is encapsulated. *Encapsulation* enables the state of a virtual machine to be saved to a file. Saving the state of a virtual machine to a file allows the transfer of an entire virtual machine to another host.

Hardware independence eliminates the need for a complete replication of hardware at the recovery site. Hardware that is running VMware vSphere ESXi server at one site can provide business continuity and disaster recovery protection for hardware that is running VMware vSphere ESXi server at another site. This capability eliminates the cost of purchasing and maintaining a system that sits idle until disaster strikes.

Hardware independence allows an image of the system at the protected site to boot from disk at the recovery site in minutes or hours instead of days.

vCenter Site Recovery Manager leverages array-based replication between a primary site and recovery site, such as the IBM DS Enhanced Remote Mirroring (ERM) functionality. The workflow that is built into SRM automatically discovers which datastores are set up for replication between the protected and recovery sites. SRM can be configured to support bidirectional protection between two sites.

vCenter Site Recovery Manager provides protection for the operating systems and applications that are encapsulated by the virtual machines that are running on VMware vSphere ESX Servers.

A vCenter Site Recovery Manager server must be installed at the protected site and at the recovery site. The protected and recovery sites must be managed by their own vCenter Server. The SRM server uses the extensibility of the vCenter Server to provide:

- Access control
- Authorization
- Custom events
- Event-triggered alarms

The architecture of a vCenter Site Recovery Manager environment is shown in Figure 1-3.

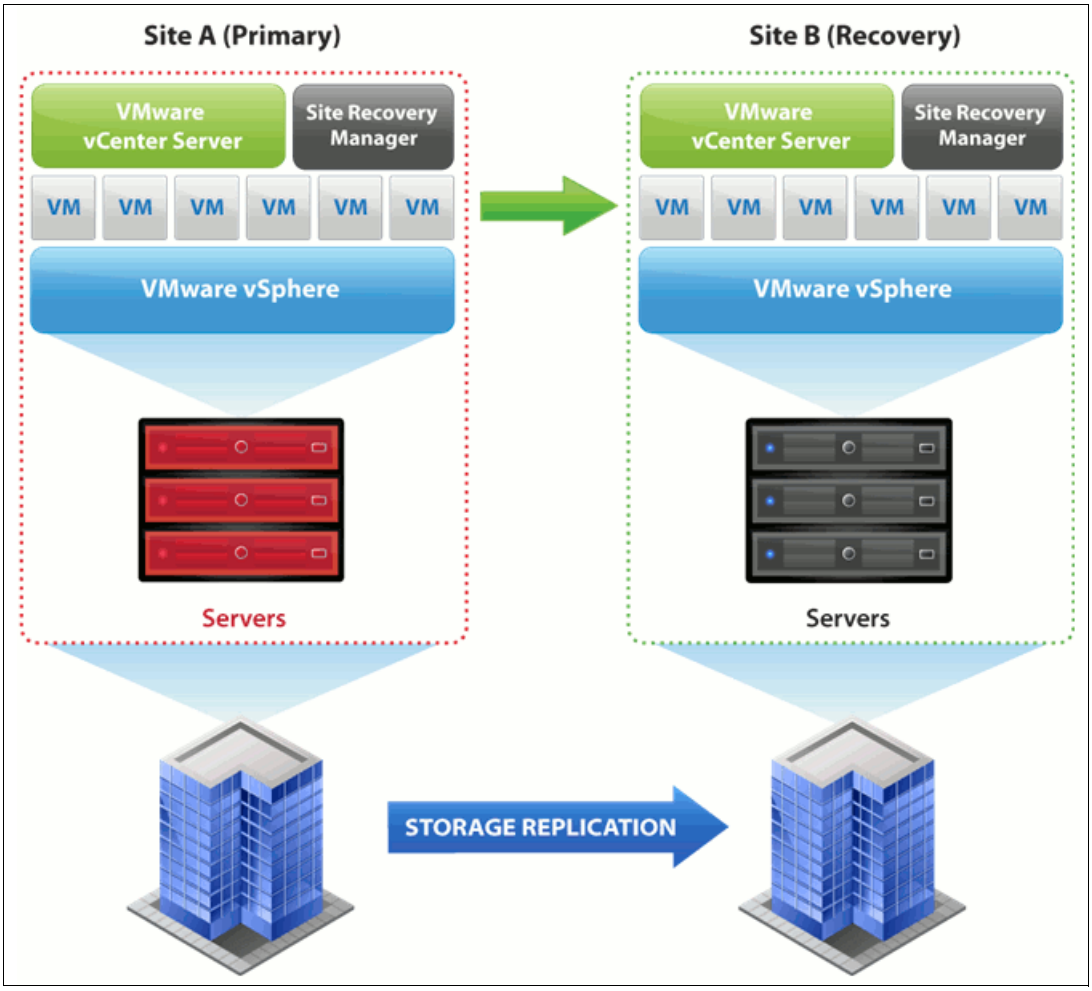

Figure 1-3 VMware vCenter Site Recovery Manager

vCenter Site Recovery Manager has the following prerequisites:

- Each site must have at least one data center. The SRM server operates as an extension to the vCenter server at a site. Because the SRM server depends on vCenter for some services, you must install and configure vCenter Server at the protected site and at the recovery site.
- Pre-configured array-based replication: If you are using array-based replication, identical replication technologies must be available at both sites.

- A supported database engine that uses Open Database Connectivity (ODBC) for connectivity in the protected site and in the recovery site.
- An SRM license that is installed on the vCenter license server at the protected site and the recovery site. Additionally, vSphere needs to be licensed sufficiently for SRM to protect and recover virtual machines.
- The recovery site must have hardware, network, and storage resources that can support the same virtual machines and workloads as the protected site.
- The sites need to be connected by a reliable IP network. If you are using array-based replication, ensure that your network connectivity meets the arrays' network requirements.
- The recovery site needs to have access to public and private networks that are comparable to the protected site's networks.

For additional information, see the following website:

http://www.vmware.com/products/site-recovery-manager/overview.html

For more detailed information, visit the IBM DS Series Portal, which contains updated product materials and guides:

http://www.ibmdsseries.com

For more information about the VMware vCenter Site Recovery Manager, see Chapter 8, "VMware Site Recovery Manager (SRM5) implementation" on page 179.

2

Security design of the VMware vSphere Infrastructure architecture

In this chapter, we discuss the security design and associated items of the VMware vSphere Infrastructure architecture.

2.1 Introduction to VMware vSphere Infrastructure

VMware vSphere Infrastructure is the most widely deployed software suite for optimizing and managing IT environments through virtualization from the desktop to the data center. The only production-ready virtualization suite, vSphere Infrastructure is proven at more than 20,000 clients of all sizes, and it is used in a wide variety of environments and applications. vSphere Infrastructure delivers transformative cost savings, as well as increased operational efficiency, flexibility, and IT service levels.

vSphere Infrastructure incorporates a number of features that directly address the security concerns of the most demanding data center environments:

- A virtualization layer designed from the ground up to run virtual machines in a secure manner and still provide high performance.
- Compatibility with SAN security practices. vSphere Infrastructure enforces security policies with logical unit number (LUN) zoning and LUN masking.
- Implementation of secure networking features. VLAN tagging enhances network security by tagging and filtering network traffic on VLANs. And, Layer 2 network security policies enforce security for virtual machines at the Ethernet layer in a way that is not available with physical servers.
- Integration with Microsoft Active Directory. vSphere Infrastructure allows you to base access controls on existing Microsoft Active Directory authentication mechanisms.

vSphere Infrastructure, the latest generation of VMware vSphere data center products, includes several key enhancements that further address the security needs and challenges of modern IT organizations:

- Custom roles and permissions: vSphere Infrastructure enhances security and flexibility with user-defined roles. You can restrict access to the entire inventory of virtual machines, resource pools, and servers by assigning users to these custom roles.
- Resource pool access control and delegation: vSphere Infrastructure secures resource allocation at other levels in the company. For example, when a top-level administrator makes a resource pool available to a department-level user, all virtual machine creation and management can be performed by the department administrator within the boundaries assigned to the resource pool.
- Audit trails: vSphere Infrastructure maintains a record of significant configuration changes and the administrator who initiated each one. You can export reports for event tracking.
- Session management: vSphere Infrastructure enables you to discover and, if necessary, terminate vCenter user sessions.

VMware has implemented internal processes to ensure VMware products meet the highest standards for security. The VMware Security Response Policy documents VMware's commitments for resolving possible vulnerabilities in VMware products so that clients can be assured that any such issues will be corrected quickly. The VMware Technology Network (VMTN) Security Center is a one-stop shop for security-related issues involving VMware products. It helps you stay up-to-date on all current security issues and to understand considerations related to securing your virtual infrastructure.

For more information, see these websites:

- VMware Security Response Policy

  http://www.vmware.com/support/policies/security_response.html

- VMware Technology Network (VMTN) Security Center

  http://www.vmware.com/technical-resources/security/index.html

The success of this architecture in providing a secure virtualization infrastructure is evidenced by the fact that many large, security-conscious clients from areas, such as banking and defense, have chosen to trust their mission-critical services to VMware virtualization.

VMware vSphere Infrastructure architecture and security features

From a security perspective, VMware vSphere Infrastructure consists of three major components:

- Virtualization layer, which consists of these components:
  - VMkernel
  - Virtual machine monitor (VMM)
  - Management framework
  - Common information model
  - Infrastructure agents
  - Virtual machine support
  - Resource management
  - Local support consoles
- Virtual machines
- Virtual networking layer

2.2 Virtualization layer

VMware vSphere ESXi presents a generic x86 platform by virtualizing four key hardware components: processor, memory, disk, and network. An operating system is then installed into this virtualized platform. The virtualization layer or VMkernel, running into Hypervisor, is a kernel designed by VMware specifically to run virtual machines. It controls the hardware utilized by VMware ESXi Server hosts and schedules the allocation of hardware resources among the virtual machines. Because the VMkernel is fully dedicated to supporting virtual machines and is not used for other purposes, the interface to the VMkernel is strictly limited to the API required to manage virtual machines. There are no public interfaces to VMkernel, and it cannot execute arbitrary code.

VMware vSphere ESXi provides additional VMkernel protection with the following features:

- Memory Hardening: The ESXi kernel, user-mode applications, and executable components, such as drivers and libraries, are located at random, non-predictable memory addresses. Combined with the non-executable memory protections made available by microprocessors, this provides protection that makes it difficult for malicious code to use memory exploits to take advantage of vulnerabilities.
- Kernel Module Integrity: Digital signing ensures the integrity and authenticity of modules, drivers, and applications as they are loaded by the VMkernel. Module signing allows ESXi to identify the providers of modules, drivers, or applications and whether they are VMware certified.

- Trusted Platform Module (TPM): Each time ESXi boots, it measures the VMkernel and a subset of the vSphere Installation Bundle (VIB) loaded modules and stores the measurements into Platform Configuration Register (PCR) 20 of the TPM. This behavior is enabled by default and cannot be disabled. Hardware support for this feature is fully tested and supported by VMware and its other equipment manufacturers (OEM) partners.

The VMkernel alternates among all the virtual machines on the host in running the virtual machine instructions on the processor. Every time a virtual machine's execution is stopped, a context switch occurs. During the context switch, the processor register values are saved and the new context is loaded. When a certain virtual machine's turn comes around again, the corresponding register state is restored.

Each virtual machine has an associated virtual machine monitor (VMM). The VMM uses binary translation to modify the guest operating system kernel code so that it can run in a less privileged processor ring. This is analogous to what a Java virtual machine (JVM) does using just-in-time translation. Additionally, the VMM virtualizes a chip set for the guest operating system to run on. The device drivers in the guest cooperate with the VMM to access the devices in the virtual chip set. The VMM passes the request to the VMkernel to complete the device virtualization and support the requested operation.

Note: The VMM utilized by VMware ESXi is the same as the one used by other VMware products that run on host operating systems, such as VMware Workstation or VMware Server. Therefore, all comments related to the VMM apply to all VMware virtualization products.

2.2.1 Local support consoles

In VMware vSphere ESXi 5, the Console OS (which was provided in all known prior versions of ESX) has been removed. Now, all VMware agents have been ported to run directly on the VMkernel. The Infrastructure services are provided natively through modules included with the VMkernel. Other authorized third-party modules, such as hardware drivers and hardware monitoring components, can run in VMkernel as well. Only modules that have been digitally signed by VMware are allowed on the system, creating a tightly locked-down architecture. Preventing arbitrary code from running on the ESXi host greatly improves the security of the system.

For more information about the Support Console improvements, see the following URL:

http://www.vmware.com/products/vsphere/esxi-and-esx/overview.html

Securing local support consoles

To protect the host against unauthorized intrusion and misuse, VMware imposes constraints on several parameters, settings, and activities. You can loosen the constraints to meet your configuration needs. If you do, make sure that you are working in a trusted environment and have taken enough other security measures to protect the network as a whole and the devices connected to the host.

Consider the following recommendations when evaluating host security and administration:

- Limit user access: To improve security, restrict user access to the management interface and enforce access security policies, such as setting up password restrictions. The ESXi Shell has privileged access to certain parts of the host. Therefore, provide only trusted users with ESXi Shell login access. Also, strive to run only the essential processes, services, and agents, such as virus checkers and virtual machine backups.

- Use the vSphere Client to administer your ESXi hosts: Whenever possible, use the vSphere Client or a third-party network management tool to administer your ESXi hosts instead of working though the command-line interface (CLI) as the root user. Using vSphere Client lets you limit the accounts with access to the ESXi Shell, safely delegate responsibilities, and set up roles that prevent administrators and users from using capabilities they do not need.
- Use only VMware sources to upgrade ESXi components: The host runs various third-party packages to support management interfaces or tasks that you must perform. VMware does not support upgrading these packages from anything other than a VMware source. If you use a download or patch from another source, you might compromise management interface security or functions. Regularly check third-party vendor sites and the VMware knowledge base for security alerts.

2.3 CPU virtualization

Binary translation is a powerful technique that can provide CPU virtualization with high performance. The VMM uses a translator with the following properties:

Binary Input is binary x86 code, not source code.

Dynamic Translation happens at run time, interleaved with execution of the generated code.

On demand Code is translated only when it is about to execute. This eliminates the need to differentiate code and data.

System level The translator makes no assumptions about the code running in the virtual machine. Rules are set by the x86 architecture, not by a higher-level application binary interface.

Subsetting The translator's input is the full x86 instruction set, including all privileged instructions; output is a safe subset (mostly user-mode instructions).

Adaptive Translated code is adjusted in response to virtual machine behavior changes to improve overall efficiency.

During normal operation, the translator reads the virtual machine's memory at the address indicated by the virtual machine program counter, classifying the bytes as prefixes, opcodes, or operands to produce intermediate representation objects. Each intermediate representation object represents one guest instruction. The translator accumulates intermediate representation objects into a translation unit, stopping at 12 instructions or a terminating instruction (usually flow control). Buffer overflow attacks usually exploit code that operates on unconstrained input without doing a length check. The classical example is a string that represents the name of something.

Similar design principles are applied throughout the VMM code. There are few places where the VMM operates on data specified by the guest operating system, so the scope for buffer overflows is much smaller than in a general-purpose operating system.

In addition, VMware programmers develop the software with awareness of the importance of programming in a secure manner. This approach to software development greatly reduces the chance that vulnerabilities will be overlooked. To provide an extra layer of security, the VMM supports the buffer overflow prevention capabilities built in to most Intel and AMD CPUs, known as the NX or XD bit. The Intel hyperthreading technology allows two process threads to execute on the same CPU package. These threads can share the memory cache

on the processor. Malicious software can exploit this feature by having one thread monitor the execution of another thread, possibly allowing theft of cryptographic keys.

VMware vSphere ESXi virtual machines do not provide hyperthreading technology to the guest operating system. VMware vSphere ESXi, however, can utilize hyperthreading to run two different virtual machines simultaneously on the same physical processor. However, because virtual machines do not necessarily run on the same processor continuously, it is more challenging to exploit the vulnerability. However, if you want a virtual machine to be protected against the small chance of the type of attack we previously discussed, VMware vSphere ESXi provides an option to isolate a virtual machine from hyperthreading. For more information, see the Knowledge Base article at this website:

http://kb.vmware.com/kb/1728

Hardware manufacturers have begun to incorporate CPU virtualization capabilities into processors. Although the first generation of these processors does not perform as well as the VMware software-based binary translator, VMware will continue to work with the manufacturers and make appropriate use of their technology as it evolves.

2.4 Memory virtualization

The RAM allocated to a virtual machine by the VMM is defined by the virtual machine's BIOS settings. The memory is allocated by the VMkernel when it defines the resources to be used by the virtual machine. A guest operating system uses physical memory allocated to it by the VMkernel and defined in the virtual machine's configuration file.

The operating system that executes within a virtual machine expects a zero-based physical address space, as provided by real hardware. The VMM gives each virtual machine the illusion that it is using such an address space, virtualizing physical memory by adding an extra level of address translation. A *machine address* refers to actual hardware memory, and a *physical address* is a software abstraction used to provide the illusion of hardware memory to a virtual machine. This paper uses "physical" in quotation marks to highlight this deviation from the usual meaning of the term.

The VMM maintains a pmap data structure for each virtual machine to translate "physical" page numbers (PPNs) to machine page numbers (MPNs). Virtual machine instructions that manipulate guest operating system page tables or translation lookaside buffer contents are intercepted, preventing updates to the hardware memory management unit. Separate shadow page tables, which contain virtual-to-machine page mappings, are maintained for use by the processor and are kept consistent with the physical-to-machine mappings in the pmap. This approach permits ordinary memory references to execute without additional overhead because the hardware translation lookaside buffer caches direct virtual-to-machine address translations read from the shadow page table. As memory management capabilities are enabled in hardware, VMware takes full advantage of the new capabilities and maintains the same strict adherence to isolation.

The extra level of indirection in the memory system is extremely powerful. The server can remap a "physical" page by changing its PPN-to-MPN mapping in a manner that is completely transparent to the virtual machine. It also allows the VMM to interpose on guest memory accesses. Any attempt by the operating system or any application running inside a virtual machine to address memory outside of what has been allocated by the VMM causes a fault to be delivered to the guest operating system, typically resulting in an immediate system crash, panic, or halt in the virtual machine, depending on the operating system. This is often termed *hyperspacing*, when a malicious guest operating system attempts I/O to an address space that is outside normal boundaries.

When a virtual machine needs memory, each memory page is zeroed out by the VMkernel before being handed to the virtual machine. Normally, the virtual machine then has exclusive use of the memory page, and no other virtual machine can touch it or even see it. The exception is when transparent page sharing is in effect.

Transparent page sharing (TPS) is a technique for using memory resources more efficiently. Memory pages that are identical in two or more virtual machines are stored one time on the host system's RAM, and each of the virtual machines has read-only access. Such shared pages are common, for example, if many virtual machines on the same host run the same operating system. As soon as any one virtual machine tries to modify a shared page, it gets its own private copy. Because shared memory pages are marked copy-on-write, it is impossible for one virtual machine to leak private information to another through this mechanism. Transparent page sharing is controlled by the VMkernel and VMM and cannot be compromised by virtual machines. It can also be disabled on a per-host or per-virtual machine basis.

The *ballooning* (balloon driver) technique or a guest balloon is a driver that is part of the VMware Tools. It is loaded into the guest operating system as a pseudo-device driver. The balloon driver process (vmmemctl) recognizes when a VM is idle and exerts artificial pressure on the guest OS causing it to swap out its memory to disk. If the hypervisor needs to reclaim virtual machine memory, it sets a proper target balloon size for the balloon driver, making it "inflate" by allocating guest physical pages within the virtual machine.

Ballooning is a completely different memory reclamation technique compared to page sharing, but working together, Transparent Page Sharing and the balloon driver let ESX Server comfortably support memory over-commitment.

ESXi has a third memory reclaim technology, which is known as *hypervisor swapping*. It is used in the cases where ballooning and page sharing are not sufficient to reclaim memory. To support this, when starting a virtual machine, the hypervisor creates a separate swap file for the virtual machine. Then, if necessary, the hypervisor can directly swap out guest physical memory to the swap file, which frees host physical memory for other virtual machines.

For more information about memory reclamation technologies, see the "Understanding Memory Resource Management in VMware ESX Server" document:

http://www.vmware.com/files/pdf/perf-vsphere-memory_management.pdf

2.5 Virtual machines

Virtual machines are the containers in which guest operating systems and their applications run. By design, all VMware virtual machines are isolated from one another. Virtual machine isolation is imperceptible to the guest operating system. Even a user with system administrator privileges or kernel system level access on a virtual machine's guest operating system cannot breach this layer of isolation to access another virtual machine without privileges explicitly granted by the VMware vSphere ESXi system administrator.

This isolation enables multiple virtual machines to run securely while sharing hardware and ensures both their ability to access hardware and their uninterrupted performance. For example, if a guest operating system running in a virtual machine crashes, other virtual machines on the same VMware vSphere ESXi host continue to run. The guest operating system crash has no effect on these areas:

- The ability of users to access the other virtual machines
- The ability of the running virtual machines to access the resources they need
- The performance of the other virtual machines

Each virtual machine is isolated from other virtual machines running on the same hardware as shown in Figure 2-1. While virtual machines share physical resources, such as CPU, memory, and I/O devices, a guest operating system in an individual virtual machine cannot detect any device other than the virtual devices made available to it.

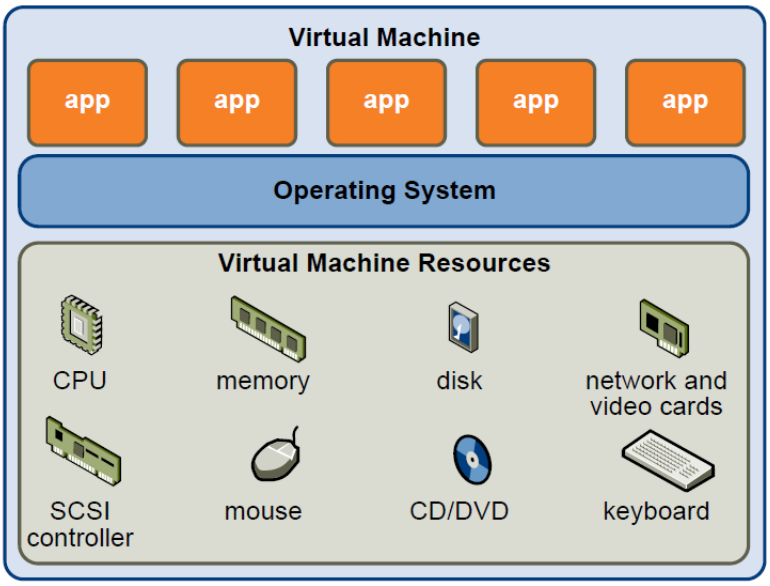

Figure 2-1 Virtual machine isolation

Because the VMkernel and VMM mediate access to the physical resources and all physical hardware access takes place through the VMkernel, virtual machines cannot circumvent this level of isolation. Just as a physical machine can communicate with other machines in a network only through a network adapter, a virtual machine can communicate with other virtual machines running on the same VMware vSphere ESXi host only through a virtual switch. Furthermore, a virtual machine communicates with the physical network, including virtual machines on other VMware vSphere ESXi hosts, only through a physical network adapter.

Considering the virtual machine isolation in a network context, you can apply these rules:

- If a virtual machine does not share a virtual switch with any other virtual machine, it is completely isolated from other virtual networks within the host.
- If no physical network adapter is configured for a virtual machine, the virtual machine is completely isolated from any physical networks.
- If you use the same safeguards (firewalls, antivirus software, and so on) to protect a virtual machine from the network as you do for a physical machine, the virtual machine is as secure as the physical machine.

You can further protect virtual machines by setting up resource reservations and limits on the ESXi host. For example, through the fine-grained resource controls available in ESXi host, you can configure a virtual machine so that it always gets at least 10% of the host's CPU resources, but never more than 20%.

Resource reservations and limits protect virtual machines from performance degradation if another virtual machine tries to consume too many resources on shared hardware. For example, if one of the virtual machines on an ESXi host is incapacitated by a denial-of-service or distributed denial-of-service attack, a resource limit on that machine prevents the attack from taking up so many hardware resources that the other virtual machines are also affected. Similarly, a resource reservation on each of the virtual machines ensures that, in high

resource demands by the virtual machine targeted by the denial-of-service attack, all the other virtual machines still have enough resources to operate.

By default, VMware vSphere ESXi imposes a form of resource reservation by applying a distribution algorithm that divides the available host resources equally among the virtual machines while keeping a certain percentage of resources for use by system components, such as the service console. This default behavior provides a degree of natural protection from denial-of-service and distributed denial-of-service attacks. You set specific resource reservations and limits on an individual basis if you want to customize the default behavior so the distribution is not equal across all virtual machines on the host.

2.6 Virtual networking layer

The virtual networking layer consists of the virtual network devices through which virtual machines and the service console interface with the rest of the network. VMware vSphere ESXi Server relies on the virtual networking layer to support communications between virtual machines and their users. In addition, VMware vSphere ESXi Server hosts use the virtual networking layer to communicate with iSCSI SANs, network-attached storage (NAS), and so on. The virtual networking layer includes virtual network adapters and the virtual switches.

2.6.1 Virtual standard switches (VSS)

The networking stack was completely rewritten for VMware vSphere ESXi Server using a modular design for maximum flexibility. A virtual switch is “built to order” at run time from a collection of small functional units:

- The core layer 2 forwarding engine
- VLAN tagging, stripping, and filtering units
- Virtual port capabilities specific to a particular adapter or a specific port on a virtual switch
- Level security, checksum, and segmentation offload units

When the virtual switch is built at run time, VMware ESXi Server loads only those components it needs. It installs and runs only what is actually needed to support the specific physical and virtual Ethernet adapter types used in the configuration. Therefore, the system pays the lowest possible cost in complexity and makes the assurance of a secure architecture all the more possible. Virtual standard switch architecture is shown in Figure 2-2 on page 22.

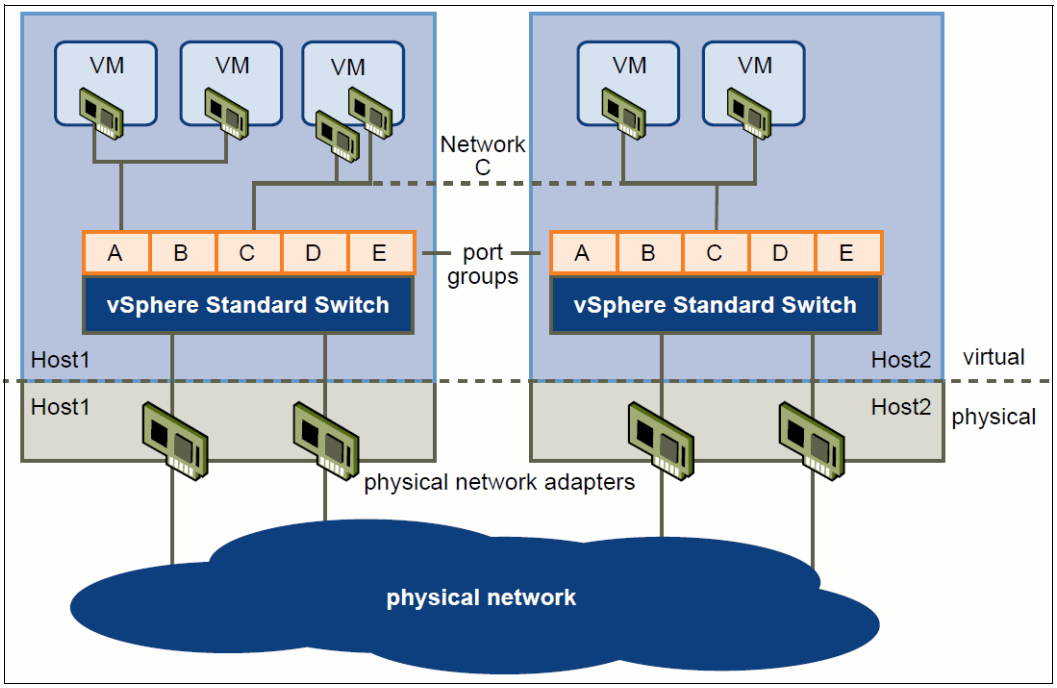

Figure 2-2 Virtual standard switch architecture

2.6.2 Virtual Distributed Switches (VDS)

A vSphere Distributed Switch (VDS) functions as a single virtual switch across all associated hosts. This ability allows virtual machines to maintain consistent network configuration as they migrate across multiple hosts. Each VDS is a network hub that virtual machines can use. A VDS can route traffic internally between virtual machines or link to an external network by connecting to physical Ethernet adapters. Each VDS can also have one or more distributed port groups assigned to it. Distributed port groups aggregate multiple ports under a common configuration and provide a stable anchor point for virtual machines connecting to labeled networks. Figure 2-3 on page 23 shows the Virtual Distributed Switch architecture.

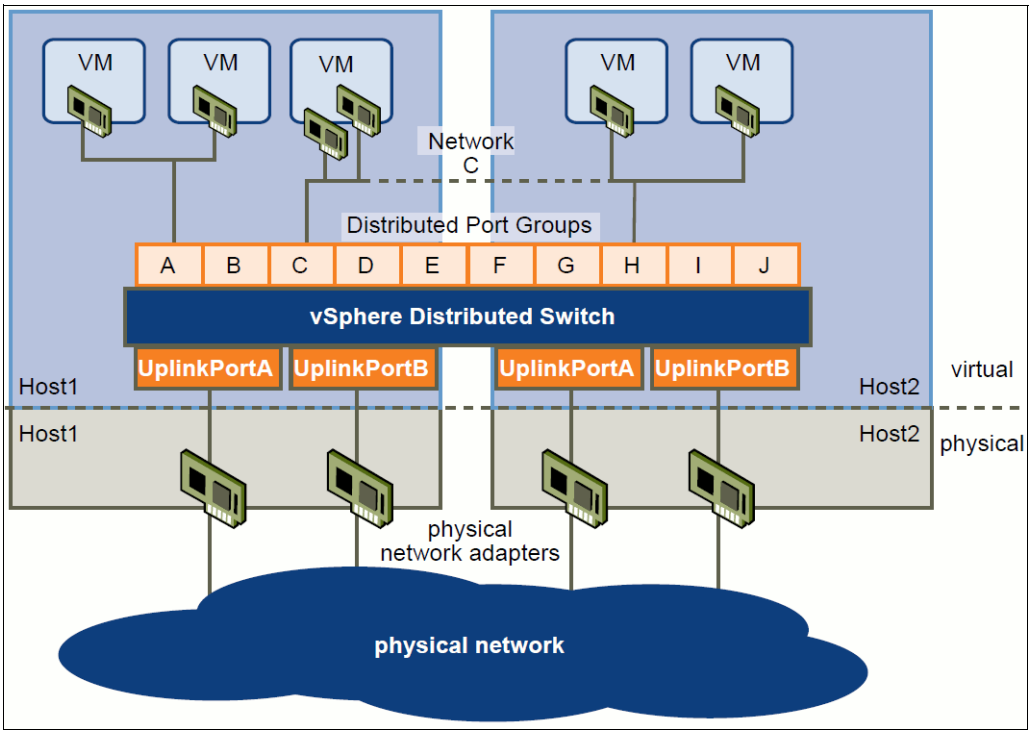

Figure 2-3 Virtual Distributed Switch architecture

2.6.3 Virtual switch VLANs

VMware ESX Server supports IEEE 802.1q VLANs, which you can use to further protect the virtual machine network, service console, or storage configuration. This driver is written by VMware software engineers according to the IEEE specification. VLANs let you segment a physical network so that two machines on the same physical network cannot send packets to or receive packets from each other unless they are on the same VLAN. There are three configuration modes to tag (and untag) the packets for virtual machine frames:

- Virtual machine guest tagging (VGT mode): You can install an 802.1Q VLAN trunking driver inside the virtual machine. Tags will be preserved between the virtual machine networking stack and external switch when frames are passed from or to virtual switches.
- External switch tagging (EST mode): You can use external switches for VLAN tagging. This is similar to a physical network, and VLAN configuration is normally transparent to each individual physical server.
- Virtual switch tagging (VST mode): In this mode, you provision one port group on a virtual switch for each VLAN, then attach the virtual machine's virtual adapter to the port group instead of the virtual switch directly. The virtual switch port group tags all outbound frames and removes tags for all inbound frames. It also ensures that frames on one VLAN do not leak into another VLAN.

2.6.4 Virtual ports

The virtual ports in vSphere ESXi Server provide a rich control channel for communication with the virtual Ethernet adapters attached to them. ESXi Server virtual ports know authoritatively what the configured receive filters are for virtual Ethernet adapters that are attached to them. This means no learning is required to populate forwarding tables.

Virtual ports also know authoritatively the “hard” configuration of the virtual Ethernet adapters attached to them. This capability makes it possible to set policies, such as forbidding MAC address changes by the guest and rejecting forged MAC address transmission, because the virtual switch port can essentially know with certainty what is “burned into ROM” (actually, stored in the configuration file, outside control of the guest operating system).

The policies available in virtual ports are much harder to implement, if possible at all, with physical switches. Either someone must manually program the ACLs into the switch port, or you must rely on weak assumptions, such as “first MAC seen is assumed to be correct.”

The port groups used in ESXi Servers do not have a counterpart in physical networks. You can think of them as templates for creating virtual ports with particular sets of specifications. Because virtual machines move around from host to host, ESXi Server needs a good way to specify, through a layer of indirection, that a given virtual machine must have a particular type of connectivity on every host on which it might run. Port groups provide this layer of indirection, enabling vSphere Infrastructure to provide consistent network access to a virtual machine, wherever it runs.

Port groups are user-named objects that contain enough configuration information to provide persistent and consistent network access for virtual Ethernet adapters:

- Virtual switch name
- VLAN IDs and policies for tagging and filtering
- Teaming policy
- Layer security options
- Traffic shaping parameters

Therefore, port groups provide a powerful way to define and enforce security policies for virtual networking as shown in Figure 2-4.

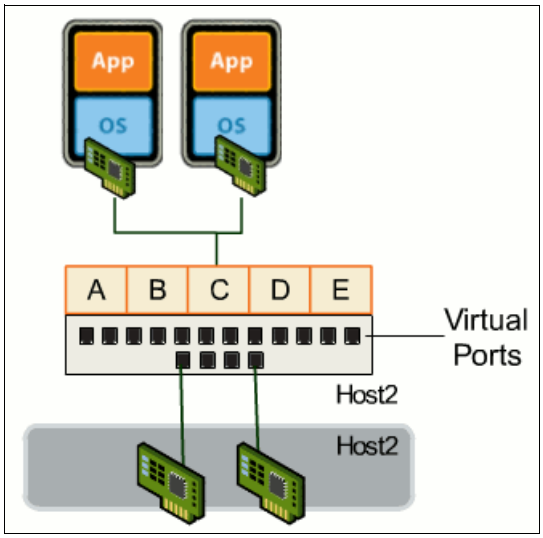

Figure 2-4 Virtual Ports

2.6.5 Virtual network adapters

vSphere Infrastructure provides several types of virtual network adapters that guest operating systems can use. The choice of adapter depends upon factors, such as support by the guest operating system and performance, but all of them share these characteristics:

- They have their own MAC addresses and unicast/multicast/ broadcast filters.

- They are strictly layered Ethernet adapter devices.
- They interact with the low-level VMkernel layer stack using a common API.

Virtual Ethernet adapters connect to virtual ports when you power on the virtual machine on which the adapters are configured, when you take an explicit action to connect the device, or when you migrate a virtual machine using vSphere VMotion. A virtual Ethernet adapter updates the virtual switch port with MAC filtering information when it is initialized and whenever it changes. A virtual port can ignore any requests from the virtual Ethernet adapter that violate the level 2 security policy in effect for the port.

2.6.6 Virtual switch isolation

A common cause of traffic leaks in the world of physical switches is cascading, which is often needed because physical switches have a limited number of ports. Because virtual switches provide all the ports you need in one switch, there is no code to connect virtual switches. vSphere ESXi Server provides no path for network data to go between virtual switches at all. Therefore, it is relatively easy for ESXi Server to avoid accidental violations of network isolation or violations that result from malicious software running in a virtual machine or a malicious user. The ESXi Server system does not have complicated and potentially failure-prone logic to make sure that only the correct traffic travels from one virtual switch to another. Instead, it simply does not implement any path that any traffic can use to travel between virtual switches. Furthermore, virtual switches cannot share physical Ethernet adapters, so there is no way to fool the Ethernet adapter into performing a loopback or something similar that causes a leak between virtual switches.

In addition, each virtual switch has its own forwarding table, and there is no mechanism in the code to allow an entry in one table to point to a port on another virtual switch. Every destination the switch looks up must match ports on the same virtual switch as the port where the frame originated, even if other virtual switches' lookup tables contain entries for that address.

A would-be attacker likely finds a remote code execution bug in the VMkernel to circumvent virtual switch isolation. Because ESXi Server parses so little of the frame data, primarily just the Ethernet header, this is difficult.

There are natural limits to this isolation. If you connect the uplinks of two virtual switches, or if you bridge two virtual switches with software running in a virtual machine, you are exposed to the same kinds of problems you might see in physical switches.

2.6.7 Virtual switch correctness

It is important to ensure that virtual machines or other nodes in the network cannot affect the behavior of the virtual switch.

VMware vSphere ESXi Server guards against such influences in the following ways:

- Virtual switches do not learn from the network in order to populate their forwarding tables. This eliminates a likely vector for denial-of-service (DoS) or leakage attacks, either as a direct DoS attempt or, more likely, as a side effect of some other attack, such as a worm or virus, as it scans for vulnerable hosts to infect.
- Virtual switches make private copies of any frame data used to make forwarding or filtering decisions, which is a critical feature and is unique to virtual switches.

It is important to ensure that frames are contained within the appropriate VLAN on a virtual switch.

ESXi Server does so in the following ways:

- VLAN data is carried outside the frame as it passes through the virtual switch. Filtering is a simple integer comparison. This is really just a special case of the general principle that the system must not trust user accessible data.
- Virtual switches have no dynamic trunking support.
- Virtual switches have no support for what is referred to as *native VLAN*.

Dynamic trunking and native VLAN are features in which an attacker might find vulnerabilities that can open isolation leaks. This is not to say that these features are inherently insecure, but even if they are implemented securely, their complexity can lead to mis-configuration and open an attack vector.

2.7 Virtualized storage

VMware vSphere ESXi Server implements a streamlined path to provide high-speed and isolated I/O for performance-critical network and disk devices. An I/O request that is issued by a guest operating system first goes to the appropriate driver in the virtual machine.

VMware vSphere ESXi Server provides two emulations of storage controllers:

- LSI Logic or BusLogic SCSI devices: The corresponding driver is loaded into the guest operating system as either an LSI Logic or a BusLogic driver. The driver typically turns the I/O requests into accesses to I/O ports to communicate to the virtual devices using privileged IA-32 IN and OUT instructions. These instructions are trapped by the virtual machine monitor and then handled by device emulation code in the virtual machine monitor that is based on the specific I/O port being accessed. The virtual machine monitor then calls device-independent network or disk code to process the I/O. For disk I/O, VMware ESX Server maintains a queue of pending requests per virtual machine for each target SCSI device. The disk I/O requests for a single target are processed in round-robin fashion across virtual machines by default. The I/O requests are then sent down to the device driver that is loaded into ESXi Server for the specific device on the physical machine.
- Paravirtual SCSI (PVSCSI) adapter: Following basically the same I/O redirection concept, VMware provides the paravirtualized SCSI adapters, where this high-performance storage adapter provides better throughput and lower CPU utilization for virtual machines. It is best suited for environments where guest applications are very I/O intensive, using applications such as Microsoft SQL Server, Oracle MySQL, and IBM DB2® among others.

2.8 SAN security

A host that runs VMware vSphere ESXi Server is attached to a Fibre Channel SAN in the same way that any other host is. It uses Fibre Channel host bust adapters (HBAs). The drivers for those HBAs are installed in the software layer that interacts directly with the hardware. In environments that do not include virtualization software, the drivers are installed on the operating system, but for vSphere ESXi Server, the drivers are installed in the VMkernel (Virtualization Layer). vSphere ESXi Server includes the native vSphere Virtual Machine File System (vSphere VMFS), which is a high performance cluster file system and volume manager that creates and manages virtual volumes on top of the LUNs that are presented to the ESXi Server host. Those virtual volumes, usually referred to as *datastores* and *virtual disks,* that are allocated to specific virtual machines.

Virtual machines have no knowledge or understanding of Fibre Channel. The only storage that is available to virtual machines is on SCSI devices. A virtual machine does not have virtual Fibre Channel HBAs; instead, it only has virtual SCSI adapters. Each virtual machine can see only the virtual disks that are presented to it on its virtual SCSI adapters. This isolation is complete, with regard to both security and performance. A VMware virtual machine has no visibility into the worldwide name (WWN), the physical Fibre Channel HBAs, or even the target ID or other information about the LUNs upon which its virtual disks reside. The virtual machine is isolated to such a degree that software that executes in the virtual machine cannot even detect that it is running on a SAN fabric. Even multipathing is handled in a way that is transparent to a virtual machine. Furthermore, virtual machines can be configured to limit the bandwidth that they use to communicate with storage devices, which prevents the possibility of a denial-of-service attack against other virtual machines on the same host by one virtual machine taking over the Fibre Channel HBA.

Consider the example of running a Microsoft Windows operating system inside a vSphere ESXi virtual machine. The virtual machine sees only the virtual disks that the ESXi Server administrator chooses at the time that the virtual machine is configured. This operation of configuring a virtual machine to see only certain virtual disks is effectively LUN masking in the virtualized environment. It has the same security benefits as LUN masking in the physical world, and it is just done with another set of tools.

Software executing in the virtual machine, including the Windows operating system, is aware of only the virtual disks that are attached to the virtual machine. Even if the Windows operating system attempts to issue a SCSI command, Report LUNs, for example, to discover other targets, vSphere ESXi Server prevents it from discovering any SCSI information that is not appropriate to its isolated and virtualized view of its storage environment. Additional complexities in the storage environment arise when a cluster of vSphere ESXi Server hosts is accessing common targets or LUNs. The vSphere VMFS ensures that all of the hosts in the cluster cooperate to ensure correct permissions and safe access to the VMFS volumes. File locks are stored on disk as part of the volume metadata, and all ESXi Server hosts that use the volumes are aware of the ownership. Ownership of files and various distributed file system activities are rendered exclusive and atomic by the use of standard SCSI reservation primitives. Each virtual disk (sometimes referred to as a *.vmdk file*) is exclusively owned by a single powered-on virtual machine. No other virtual machine on the same or another ESXi Server host is allowed to access that virtual disk. This situation does not change fundamentally when there is a cluster of vSphere ESXi Server hosts, with multiple virtual machines powered on and accessing virtual disks on a single VMFS volume. Because of this fact, vSphere VMotion, which enables live migration of a virtual machine from one ESXi Server host to another, is a protected operation.

2.9 VMware vSphere vCenter Server

VMware vSphere vCenter Server provides a central place where almost all management functions of VMware Infrastructure can be performed. vCenter relies on Windows security controls and therefore must reside on a properly managed server with network access limited to those ports that are necessary for it to interoperate with all of the other VMware vSphere components and features. It is role-based and tied to Active Directory or heritage NT domains, making it unnecessary to create custom user accounts for it. vCenter also keeps records of nearly every event in the vSphere ESXi Server system, allowing the generation of audit trails for compliance.

vSphere vCenter manages the creation and enforcement of resource pools, which are used to partition available CPU and memory resources. A resource pool can contain child resource pools and virtual machines, allowing the creation of a hierarchy of shared resources. Using

resource pools, you can delegate control over resources of a host or cluster. When a top-level administrator makes a resource pool available to a department-level administrator, that administrator can then perform all virtual machine creation and management within the boundaries of the resources to which the resource pool is entitled. More important, vSphere vCenter enforces isolation between resource pools, so that resource usage in one pool does not affect the availability of resources in another pool. This action provides a coarser level of granularity for containment of resource abuse in addition to the granularity that is provided on the vSphere ESXi Server host level.

vSphere vCenter has a sophisticated system of roles and permissions to allow fine-grained determination of authorization for administrative and user tasks, based on user or group and inventory item, such as clusters, resource pools, and hosts. Using this system, you can ensure that only the minimum necessary privileges are assigned to people to prevent unauthorized access or modification.

VMware vCenter Server lets administrators rapidly provision VMs and hosts using standardized templates, and ensures compliance with vSphere host configurations and host and VM patch levels with automated remediation. VMware vCenter Server also gives administrators control over key capabilities, such as vSphere vMotion, Distributed Resource Scheduler, High Availability, and Fault Tolerance.

For more information about vSphere vCenter architecture and features, see this website:

http://www.vmware.com/files/pdf/techpaper/Whats-New-VMware-vCenter-Server-50-Technical-Whitepaper.pdf

3

Planning the VMware vSphere storage system design

Planning carefully is essential to any new storage installation. And to choose the proper equipment and software with their best settings for your installation can be a challenge. Well-thought-out design and planning prior to the implementation can help you get the most of your investment in the present and protect it for the future, which can include throughput capability, size, and resources that are necessary to handle the volume of traffic allied with the required capacity.

This chapter provides guidelines needed to assist in the planning of storage systems for a VMware vSphere environment.

3.1 VMware vSphere ESXi Server storage structure: Disk virtualization

In addition to the disk virtualization that is offered by a SAN, VMware further abstracts the disk subsystem from the guest operating system (OS). It is important to understand this structure to make sense of the options for best practices when connecting VMware vSphere ESXi hosts to a SAN-attached subsystem.

3.1.1 Local storage

The disks that vSphere ESXi host uses for its boot partition are usually local disks that have a partition/file structure akin to the Linux file hierarchy. The disks are internal storage devices inside your ESXi host and external storage devices outside and connected to the host directly through different protocols. vSphere ESXi supports various internal and external local storage devices (disks), including SCSI, IDE, SATA, USB, and SAS storage systems. Because local storage devices do not support sharing across multiple hosts, the recommendation is to use them only for storing some template or ISO files.

3.1.2 Networked storage

Networked storage consists of external storage systems that your ESXi host uses to store virtual machine files remotely. Typically, the host accesses these systems over a high-speed storage network. Networked storage devices are shared. Datastores on networked storage devices can be accessed by multiple hosts concurrently. IBM DS Storage Systems attached to vSphere ESXi hosts support the following networked storage technologies.

Fibre Channel (FC) storage

FC storage stores virtual machine files remotely on an FC storage area network (SAN). An FC SAN is a high-speed network that connects your hosts to high-performance storage devices. The network uses FC protocol to transport SCSI traffic from virtual machines to the FC SAN devices. To connect to the FC SAN, your host needs to be equipped with FC host bus adapters (HBAs) and FC (fabric) switches to route storage traffic.

Figure 3-1 on page 31 shows a host with an FC HBA connected to a fibre array (storage) through a SAN fabric switch. The LUN from a storage array becomes available to the host. The virtual machine access to the LUNs is accomplished through a Virtual Machine File System (VMFS) datastore.

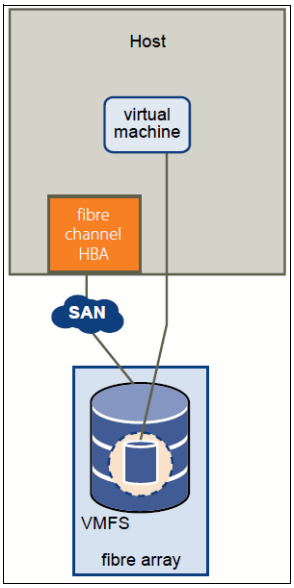

Figure 3-1 vSphere ESXi basic FC storage configuration

Internet Small Computer System Interface (iSCSI)

iSCSI is an industry standard development to enable transmission of SCSI block commands over the existing IP network by using TCP/IP. The virtual machine files are remotely stored on storage with iSCSI capabilities. iSCSI SANs use Ethernet connections between host servers and high-performance storage subsystems.

iSCSI SAN uses a client/server architecture, in which the client (vSphere ESXi host), called iSCSI initiator, operates on your host. It initiates iSCSI sessions by issuing SCSI commands and transmitting them, encapsulated into iSCSI protocol, to a server (storage system). The server is known as an iSCSI *target*. The iSCSI target represents a physical storage system on the network. The iSCSI target responds to the initiator's commands by transmitting required iSCSI data.

VMware supports different types of initiators.

Hardware iSCSI adapter

A hardware iSCSI adapter is a third-party adapter that offloads iSCSI and network processing from your host. Hardware iSCSI adapters are divided into categories:

- Dependent hardware iSCSI adapter: Depends on VMware networking and iSCSI configuration and management interfaces that are provided by VMware.
- Independent hardware iSCSI adapter: Implements its own networking and iSCSI configuration and management interfaces.

Software iSCSI adapter

A software iSCSI adapter is VMware code built into the VMkernel. It allows your host to connect to the iSCSI storage device through standard network adapters. The software iSCSI adapter handles iSCSI processing while communicating with the network adapter. With the software iSCSI adapter, you can use iSCSI technology without purchasing specialized hardware.

Figure 3-2 on page 32 shows supported vSphere ESXi iSCSI initiators and a basic configuration.

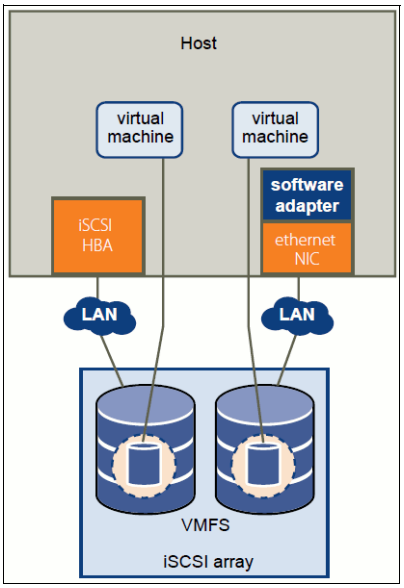

Figure 3-2 vSphere ESXi iSCSI supported initiators and basic configuration

For more information about iSCSI and FC storage basics, see the *IBM System Storage DS5000 Series Hardware Guide*, SG24-8023.

3.1.3 SAN disk usage

VMware vSphere continues to emphasize support for SAN-based disks. SAN disk is used on vSphere Server in the following manner:

- After the IBM Midrange Storage Subsystem is configured with arrays, logical drives, and storage partitions, these logical drives are presented to the vSphere Server.
- Two options exist for using these logical drives within vSphere Server:

  Option 1 Formatting these disks with the VMFS: This option is most common because several features require that the virtual disks are stored on VMFS volumes.

  Option 2 Passing the disk through to the guest OS as a raw disk: No further virtualization occurs; instead, the OS writes its own file system onto that disk directly as though it is in a stand-alone environment without an underlying VMFS structure.
- VMFS volumes house the virtual disks that the guest OS sees as its real disks. These virtual disks are in the form of a virtual disk file with the extension *.vmdk*.
- The guest OS either reads from or writes to the virtual disk file (`.vmdk`) or writes through the vSphere ESXi abstraction layer to a raw disk. In both cases, the guest OS treats the disks as though they are real.

Figure 3-3 on page 33 shows logical drives to vSphere VMFS volumes.

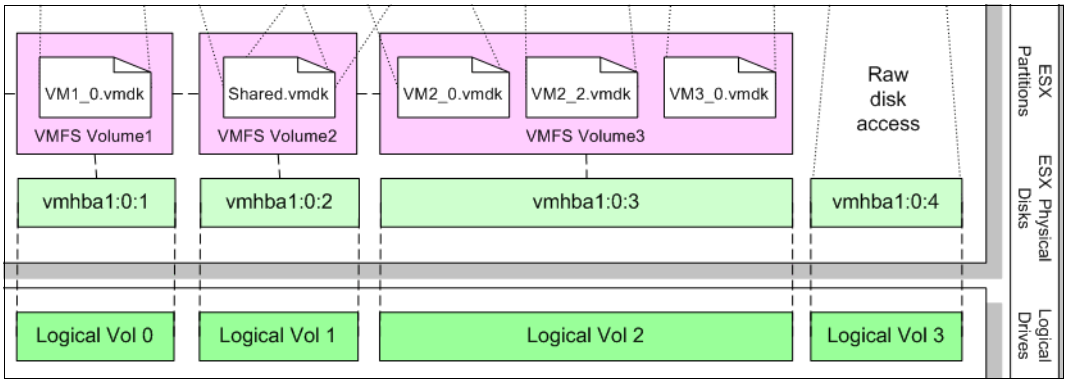

Figure 3-3 Logical drives to vSphere VMFS volumes

3.1.4 Disk virtualization with VMFS volumes and .vmdk files

The VMware vSphere Virtual Machine File System (VMFS) is the file system created by VMware specifically for the vSphere Server environment. It is designed to handle very large disks (LUNs) and store the virtual machine (*.vmdk*) files. VMFS volumes store these types of information:

- Virtual machine disk files (`.vmdk`)
- Virtual machine configuration files (`.vmx`)
- Memory images from suspended virtual machines
- Snapshot files for the `.vmdk` files manually created or set to a disk mode of non-persistent, undoable, or append.

The virtual machine `.vmdk` files represent what is seen as a physical disk by the guest OS. These files have many distinct benefits over physical disks (although several of these functions are available through the advanced functions of an IBM Midrange Storage System):

- They are portable, so they can be copied from one vSphere ESXi host to another, either for moving a virtual machine to a new ESXi host or to create backup or test environments. When copied, they retain all of the structure of the original disk and if it is the virtual machine's boot disk, it includes all of the hardware drivers that are necessary to make it run on another vSphere ESXi host (although the `.vmx` configuration file also needs to be replicated to complete the virtual machine).
- They are easily resized (using **`vmfsktools`** or *vSphere Client*) if the virtual machine needs more disk space. This option presents a larger disk to the guest OS that requires a volume expansion tool for accessing the additional space.
- They can be mapped and remapped on a single vSphere ESXi host to keep multiple copies of a virtual machine's data. Many more `.vmdk` files can be stored for access by a vSphere host than are represented by the number of virtual machines that are configured.

3.1.5 VMFS access mode: Public mode

Public mode is the default mode for VMware ESXi Server and the *only* option for VMware ESX 3.x and above.

VMFS-3 partitions also allow multiple vSphere Servers to access the VMFS volume concurrently and use file locking to prevent contention on the `.vmdk` files.

Introduced with vSphere5, VMFS-5 continues providing the same file locking mechanism as VMFS-3 to prevent the contention on the `.vmdk` files.

> **Note:** Starting with VMFS-3, there is no longer a shared mode. Virtual machine clustering now occurs with raw device mapping (RDM) in physical or virtual compatibility mode.

3.1.6 vSphere Server .vmdk modes

vSphere Server has two modes of operation for `.vmdk` file disks that can be set from within the vSphere ESXi.

The server management user interface is seen during the creation of the `.vmdk` files or afterward by editing an individual virtual machine's settings. The modes are listed:

Persistent Similar to normal physical disks in a server. vSphere Server writes immediately to a persistent disk.

Non-persistent Changes that were made since the last time a virtual machine was powered on are lost when that VM is powered off (soft reboots do not count as being powered off).

3.1.7 Specifics of using SAN arrays with vSphere ESXi Server

Using a SAN with an vSphere ESXi Server host differs from traditional SAN usage in various ways, which we discuss in this section.

Sharing a VMFS across vSphere ESXi Servers

The vSphere Virtual Machine File System (VMFS), shown in Figure 3-4 on page 35, is designed for concurrent access from multiple physical machines and enforces the appropriate access controls on virtual machine files.

vSphere VMFS can perform these functions:

- Coordinate access to virtual disk files: ESXi Server uses file-level locks, which the *VMFS Distributed Lock Manager* manages. This feature prevents a virtual machine from being powered on by multiple servers at the same time.
- Coordinate access to VMFS internal file system information (metadata): vSphere ESXi Server coordinates accurate shared data.

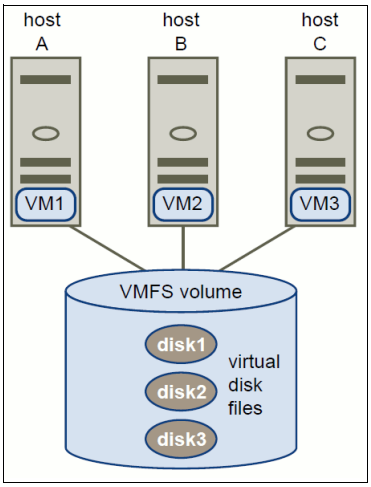

Figure 3-4 VMFS across ESXi hosts

Metadata updates

A VMFS holds files, directories, symbolic links, RDMs, and so on, and the 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 the following operations:

- Creating, growing, or locking a file
- Changing a file's attributes
- Powering a virtual machine on or off
- Creating or deleting a VMFS datastore
- Expanding a VMFS datastore

LUN display and rescan

A SAN is dynamic. The LUNs that are available to a certain host can change based on several factors:

- 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.

3.1.8 Host types

Every LUN has a slightly different behavior depending on the type of host that is accessing it. The host type determines how the storage subsystem controllers work with each operating system on the hosts to which they are connected. For VMware hosts, a special host type is available: VMware. If you are using the default host group, ensure that the default host type is also VMware.

Note: If you change the host type while the storage subsystem and host are running, you need to follow these guidelines:

- The controllers do not need to be rebooted after the change of host type.
- The host must be rebooted.
- Change the host type under low I/O conditions.

3.1.9 Levels of indirection

If you are used to working with traditional SANs, the levels of indirection can initially be confusing for these reasons:

- You cannot directly access the virtual machine operating system that uses the storage. With traditional tools, you can monitor only the VMware ESXi Server operating system, but not the virtual machine operating system. Use the vSphere Client to monitor virtual machines.
- Each virtual machine is configured by default with one virtual hard disk and one virtual SCSI controller during installation. You can modify the SCSI controller type and SCSI bus sharing characteristics by using the vSphere Client to edit the virtual machine settings. You can also add hard disks and virtual SCSI controllers to your virtual machine.
- The HBA that is visible to the SAN administration tools is part of the VMware vSphere ESXi Server, not the virtual machine.
- The VMware vSphere ESXi Server system performs multipathing for you, the VMkernel multipathing plug-in that ESXi provides by default is the VMware Native Multipathing Plug-in (NMP). The NMP is an extensible module. The following tasks are in scope:
  - Manages physical path claiming and unclaiming
  - Registers and unregisters logical devices
  - Associates physical paths with logical devices
  - Processes I/O requests to logical devices
  - Supports management tasks, such as abort or reset of logical devices

3.2 Choosing the IBM Midrange Storage Subsystem for a VMware implementation

Unfortunately, there is not a straightforward answer to this question. All of the IBM Midrange Storage Systems can provide excellent functionality for attaching to VMware vSphere Servers. The answer depends on the specific requirements that a vSphere Server is intended to address and the expectations that need to be met in terms of performance, availability, capacity, growth, and so on.

One thing is certain, the sizing requirements for capacity and performance do not change when a vSphere Server is being considered as opposed to a group of individual physical servers. Some consolidation of SAN requirements can be achieved. Other requirements remain, for example, because of under-utilization. Consolidation is often possible regarding the number of physical HBAs that can be required and, therefore, also the number of SAN switch ports that are also required to connect those HBAs. Because both of these items come at a considerable cost, any reduction can represent significant savings. It is also common to find low-bandwidth utilization of HBAs and SAN switch ports in a non-consolidated environment, thus also adding to the potential for consolidation of these items.

Alternatively, it is common that individual physical disk utilization is high and, therefore, reducing the number of physical disks is often inappropriate. Like in all SAN implementations, consider both the immediate requirements of the project and the possibilities for reasonable future growth.

3.3 Overview of IBM Midrange Storage Systems

In this section, we provide a brief overview of the IBM Midrange Storage Systems to help you decide which storage subsystem is best suited for your VMware environment. For detailed descriptions of IBM Midrange Storage Systems, see these books:

- *IBM System Storage DS5000 Series Hardware Guide*, SG24-8023
- *IBM System Storage DS5000 Series Implementation and Best Practices Guide*, SG24-8024
- *IBM System Storage DS3500 Introduction and Implementation Guide*, SG24-7914
- *IBM System Storage DCS3700 Introduction and Implementation Guide*, SG24-8037

3.3.1 Positioning the IBM Midrange Storage Systems

The IBM DS storage family is suitable for a broad range of business needs. From entry-level IBM System Storage DS3000 series, to the midrange IBM System Storage DS5000 series, to high-performance IBM System Storage DS8000® series, the IBM DS storage family covers the needs of small businesses all the way up to dealing with large enterprise requirements.

The IBM Midrange Storage Systems, also referred to as the IBM System Storage DS3000 and DS5000 series, are designed to meet the demanding open-systems requirements of today and tomorrow, while establishing a new standard for lifecycle longevity with field-replaceable host interface cards. Seventh-generation architecture delivers relentless performance, real reliability, multidimensional scalability, and unprecedented investment protection.

The IBM System Storage DS3000 and DS5000 series consists of the following storage systems:

- DS3500 Express:
  - DS3512 Express Single Controller Storage System (1746A2S)
  - DS3512 Express Dual Controller Storage System (1746A2D)
  - DS3524 Express Single Controller Storage System (1746A4S)
  - DS3524 Express Dual Controller Storage System (1746A4D)
  - DS3524 Express DC Dual Controller Storage System (1746T4D)
- DCS3700 (1818-80C)
- DS5020 Express Disk System (1814-20A)

  Targeted at growing midrange sites requiring reliability, efficiency, and performance value
- DS5100 Disk System (1818-51A)

  Targeted at cost-conscious midrange sites requiring high-end functionality and pay-as-you-grow scalability
- DS5300 Disk System (1818-53A)

  Targeted at environments with compute-intensive applications and large-scale virtualization/consolidation implementations

Figure 3-5 on page 38 shows the positioning of the products within the DS3000 and DS5000 Midrange series.

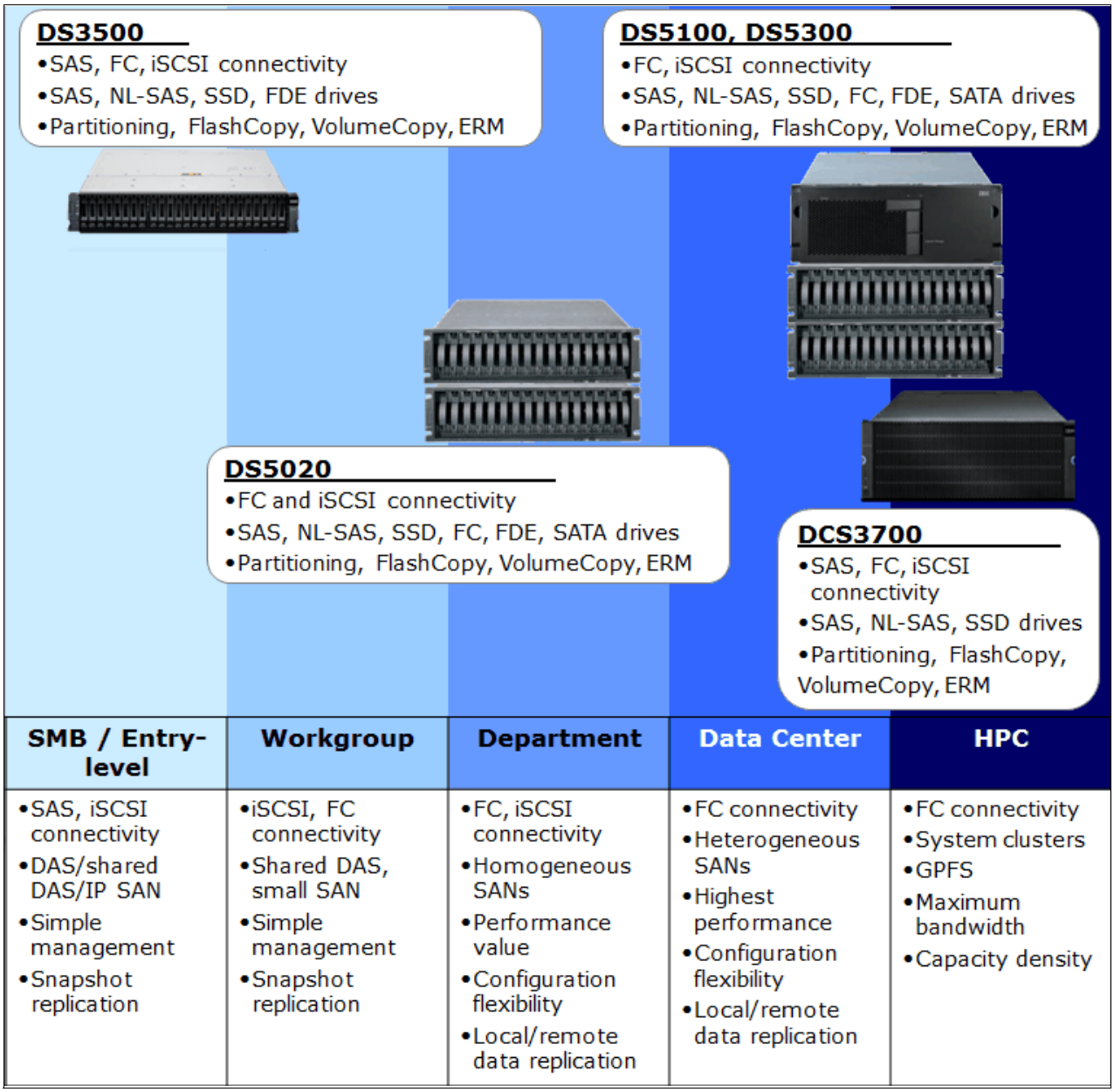

Figure 3-5 Product positioning within the Midrange DS3000 and DS5000 series

For more information about the positioning and the characteristics of each of the family members of the IBM Midrange System Storage, see the *IBM System Storage DS5000 Series Hardware Guide*, SG24-8023, and *IBM System Storage DS5000 Series Implementation and Best Practices Guide*, SG24-8024.

3.4 Storage subsystem considerations

This section presents several important application-specific considerations.

3.4.1 Segment size

The segment size that we discuss in the following section refers to the data partitions of your VMware installation. It is recommended to separate your OS partitions from your data partitions. Base the segment size on the type and expected I/O size of the data. Store sequentially read data on logical drives with small segment sizes and with dynamic prefetch enabled to dynamically read-ahead blocks. For the procedure to choose the appropriate disk segment size, see “Calculating optimal segment size” on page 39.

Oracle

Little I/O from Oracle is truly sequential in nature except for processing redo logs and archive logs. Oracle can read a full-table scan all over the disk drive. Oracle calls this type of read a *scattered read*. The Oracle’s sequential data read is for accessing a single index entry or a single piece of data. Use small segment sizes for an Online Transaction Processing (OLTP) environment with little or no need for read-ahead data. Use larger segment sizes for a Decision Support System (DSS) environment where you perform full table scans through a data warehouse.

Remember three important things when considering block size:

- Set the database block size lower than or equal to the disk drive segment size. If the segment size is set at 2 KB and the database block size is set at 4 KB, this procedure takes two I/O operations to fill the block, which results in performance degradation.
- Make sure that the segment size is an even multiple of the database block size. This practice prevents partial I/O operations from filling the block.
- Set the parameter *db_file_multiblock_read_count* appropriately. Normally, you want to set the *db_file_multiblock_read_count* as shown:

  `segment size = db_file_multiblock_read_count * DB_BLOCK_SIZE`

You also can set the *db_file_multiblock_read_count* so that the result of the previous calculation is smaller but in even multiples of the segment size. For example, if you have a segment size of 64 KB and a block size of 8 KB, you can set the *db_file_multiblock_read_count* to 4, which equals a value of 32 KB, which is an even multiple of the 64-KB segment size.

SQL Server

For SQL Server, the page size is fixed at 8 KB. SQL Server uses an extent size of 64 KB (eight 8-KB contiguous pages). For this reason, set the segment size to 64 KB. Read “Calculating optimal segment size” on page 39.

Exchange server

Set the segment size to 64 KB or multiples of 64. See “Calculating optimal segment size”.

Calculating optimal segment size

The IBM term *segment size* refers to the amount of data that is written to one disk drive in an array before writing to the next disk drive in the array. For example, in a RAID 5, 4+1 array with a segment size of 128 KB, the first 128 KB of the LUN storage capacity is written to the first disk drive and the next 128 KB to the second disk drive. For a RAID 1, 2+2 array, 128 KB of an I/O is written to each of the two data disk drives and to the mirrors. If the I/O size is larger than the number of disk drives times 128 KB, this pattern repeats until the entire I/O is completed.

For very large I/O requests, the optimal segment size for a RAID array is one that distributes a single host I/O across all data disk drives.

The formula for optimal segment size is:

```
LUN segment size = LUN stripe width ÷ number of data disk drives
```

For RAID 5, the number of data disk drives is equal to the number of disk drives in the array minus 1, for example:

```
RAID5, 4+1 with a 64 KB segment size => (5-1) * 64KB = 256 KB stripe width
```

For RAID 1, the number of data disk drives is equal to the number of disk drives divided by 2, for example:

```
RAID 10, 2+2 with a 64 KB segment size => (2) * 64 KB = 128 KB stripe width
```

For small I/O requests, the segment size must be large enough to minimize the number of segments (disk drives in the LUN) that must be accessed to satisfy the I/O request, that is, to minimize segment boundary crossings. For IOPS environments, set the segment size to 256 KB or larger so that the stripe width is at least as large as the median I/O size.

When using a logical drive manager to collect multiple storage system LUNs into a Logical Volume Manager (LVM) array or volume group(VG), the I/O stripe width is allocated across all of the segments of all of the data disk drives in all of the LUNs. The adjusted formula is shown:

```
LUN segment size = LVM I/O stripe width / (# of data disk drives/LUN * # of
LUNs/VG)
```

To learn the terminology so that you can understand how data in each I/O is allocated to each LUN in a logical array, see the vendor documentation for the specific LVM.

Best practice: For most implementations, set the segment size of VMware data partitions to 256 KB.

3.4.2 Midrange Storage Systems cache features

On the DS3500, DCS3700, and DS5000 series, there are two cache features that are worth describing. These features are the permanent cache backup and the cache mirroring.

The *permanent cache backup* feature provides a cache hold-up and de-staging mechanism to save cache and processor memory to a permanent device. This feature replaces the reliance on batteries in older models to keep the cache alive for a period of time when power is interrupted.

Disk drive cache has permanent data retention in a power outage. This function is accomplished by using flash drives. The batteries only power the controllers until data in the cache is written to the flash drives. When the storage subsystem is powered back up, the contents are reloaded to cache and flushed to the logical drives.

When you turn off the storage subsystem, it does not shut down immediately. The storage subsystem writes the contents of cache to the flash drives before powering off. Depending on the amount of cache, the storage subsystem might take up to several minutes to actually power off.

Note: When upgrading cache, dual inline memory modules (DIMMs) need to be upgraded together with flash drives.

The other feature used for cache protection is the *dedicated write cache mirroring*. When this feature is enabled, all cache is mirrored between controllers. In the case of a controller failure, write cache is not lost because the other controller mirrored the cache. When write cache mirroring is enabled, there is no impact to performance.

3.4.3 Enabling cache settings

Always enable read cache. Enabling read cache allows the controllers to process data from the cache if it was read before and thus the read is significantly faster. Data remains in the read cache until it is flushed.

Enable write cache to let the controllers acknowledge writes as soon as the data reaches the cache instead of waiting for the data to be written to the physical media. For other storage systems, a trade-off exists between data integrity and speed. The IBM DS3500, DCS3700, and DS5000 storage subsystems are designed to store data on both controller caches before being acknowledged. To protect data integrity, cache mirroring must be enabled to permit dual controller cache writes.

Enable write-cache mirroring to prevent the cache being lost in a controller failure.

Whether you need to prefetch cache depends on the type of data that is stored on the logical drives and how that data is accessed. If the data is accessed randomly (by way of table spaces and indexes), disable prefetch. Disabling prefetch prevents the controllers from reading ahead segments of data that most likely will not be used, unless your logical drive segment size is smaller than the data read size requested. If you are using sequential data, cache prefetch might increase performance because the data can be pre-stored in cache before reading.

3.4.4 Aligning file system partitions

Align partitions to stripe width. Calculate stripe width by the following formula:

```
segment_size / block_size * num_drives
```

For example, using this formula, 4+1 RAID 5 with a 512-KB segment equals 512 KB / 512 Byte * 4 drives= 4096 Bytes.

3.4.5 Premium features

Premium features, such as FlashCopy and VolumeCopy, are available for both the virtual drive and for the raw device mapping (RDM) device. For virtual drives, VMware has tools to provide these functions. For RDM devices, the IBM Midrange Storage Subsystem provides the following premium features:

- VolumeCopy
- FlashCopy
- Enhanced Remote Mirroring
- Storage Partitioning

3.4.6 Considering individual virtual machines

Before you can effectively design your array and logical drives, you must determine the primary goals of the configuration: performance, reliability, growth, manageability, or cost. Each goal has positives, negatives, and trade-offs. With the goals determined for your environment, follow the guidelines discussed in this chapter to implement them. To get the best performance from the IBM storage subsystem, you must know the I/O characteristics of the files to be placed on the storage system. After you know the I/O characteristics of the files, you can set up a correct array and logical drive to support these files.

Web servers

Web server storage workloads typically contain random small writes. RAID 5 provides good performance. It has the advantage of protecting the system from one drive loss and has a lower cost by using fewer disk drives.

Backup and file read applications

The IBM Midrange Storage Systems perform very well for a mixed workload. There are ample resources, such as IOPS and throughput, to support backups of virtual machines and not affect the other applications in a virtual environment. Addressing performance concerns for individual applications takes precedence over backup performance.

However, there are applications that read large files sequentially. If performance is important, consider using RAID 10. If cost is also a concern, RAID 5 protects from disk drive loss with the fewest disk drives.

Databases

Databases are classified as one of the following categories:

- Frequently updated databases: If your database is frequently updated and if performance is a major concern, your best choice is RAID 10, even though RAID 10 is the most expensive because of the number of disk drives and expansion drawers. RAID 10 provides the least disk drive overhead and provides the highest performance from the IBM storage systems.
- Low-to-medium updated databases: If your database is updated infrequently or if you must maximize your storage investment, choose RAID 5 for the database files. RAID 5 lets you create large storage logical drives with minimal redundancy of disk drives.
- Remotely replicated environments: If you plan to remotely replicate your environment, carefully segment the database. Segment the data on smaller logical drives and selectively replicate these logical drives. Segmenting limits WAN traffic to only what is absolutely needed for database replication. However, if you use large logical drives in replication, initial establish times are larger and the amount of traffic through the WAN might increase, leading to a poor database performance. The IBM premium features, Enhanced Remote Mirroring (ERM), VolumeCopy, and FlashCopy, are useful for replicating remote environments.

3.4.7 Determining the best RAID level for logical drives and arrays

In general, RAID 5 works best for sequential large I/Os (> 256 KB), and RAID 5 or RAID 1 works best for small I/Os (< 32 KB). For I/O sizes in between, the RAID level can be dictated by other application characteristics. Table 3-1 on page 43 shows the I/O size and optimal RAID level.

Table 3-1 I/O size and optimal RAID level

I/O size	RAID level
Sequential, large (>256 KB)	RAID 5
Small (<32 KB)	RAID 5 or RAID 1
Between 32K B and 256 KB	RAID level does not depend on I/O size.

RAID 5 and RAID 1 have similar characteristics for read environments. For sequential writes, RAID 5 typically has an advantage over RAID 1 because of the RAID1 requirement to duplicate the host write request for parity. This duplication of data typically puts a strain on the drive-side channels of the RAID hardware. RAID 5 is challenged most by random writes, which can generate multiple disk drive I/Os for each host write. Different RAID levels can be tested by using the DS Storage Manager Dynamic RAID Migration feature, which allows the RAID level of an array to be changed and maintain continuous access to data.

Table 3-2 shows the RAID levels that are most appropriate for specific file types.

Table 3-2 Best RAID level for file type

File type	RAID level	Comments
Oracle Redo logs	RAID 10	Multiplex with Oracle
Oracle Control files	RAID 10	Multiplex with Oracle
Oracle Temp datafiles	RAID 10, RAID 5	Performance first/drop re-create on disk drive failure
Oracle Archive logs	RAID 10, RAID 5	Determined by performance and cost requirements
Oracle Undo/Rollback	RAID 10, RAID 5	Determined by performance and cost requirements
Oracle datafiles	RAID 10, RAID 5	Determined by performance and cost requirements
Oracle executables	RAID 5	
Oracle Export files	RAID 10, RAID 5	Determined by performance and cost requirements
Oracle Backup staging	RAID 10, RAID 5	Determined by performance and cost requirements
Exchange database	RAID 10, RAID 5	Determined by performance and cost requirements
Exchange log	RAID 10, RAID 5	Determined by performance and cost requirements
SQL Server log file	RAID 10, RAID 5	Determined by performance and cost requirements
SQL Server data files	RAID 10, RAID 5	Determined by performance and cost requirements
SQL Server Tempdb file	RAID 10, RAID 5	Determined by performance and cost requirements

Use RAID 0 arrays only for high-traffic data that does not need any redundancy protection for device failures. RAID 0 is the least used RAID format but provides for high-speed I/O without the additional redundant disk drives for protection.

Use RAID 1 for the best performance and to provide data protection by mirroring each physical disk drive. Create RAID 1 arrays with the most disk drives possible (30 maximum) to achieve the highest performance.

Use RAID 5 to create arrays with either 4+1 disk drives or 8+1 disk drives to provide the best performance and to reduce RAID overhead. RAID 5 offers good read performance at a reduced cost of physical disk drives compared to a RAID 1 array.

Note: If protection for a two drive failure is needed, use RAID 6. It has the same performance as RAID 5 but uses an extra drive for additional protection.

Use RAID 10 (RAID 1+0) to combine the best features of data mirroring of RAID 1, plus the data striping of RAID 0. RAID 10 provides fault tolerance and better performance compared to other RAID options. A RAID 10 array can sustain multiple disk drive failures and losses as long as no two disk drives form a single pair of one mirror.

3.4.8 Server consolidation considerations

There is a misconception that simply adding up the amount of storage required for the number of servers that will be attached to a SAN is good enough to size the SAN. The importance of understanding performance and capacity requirements is very high but is even more relevant to the VMware environment because the concept of server consolidation is also part of the equation. Figure 3-6 demonstrates a consolidation of four physical servers into a single VMware ESXi Server to explain the considerations.

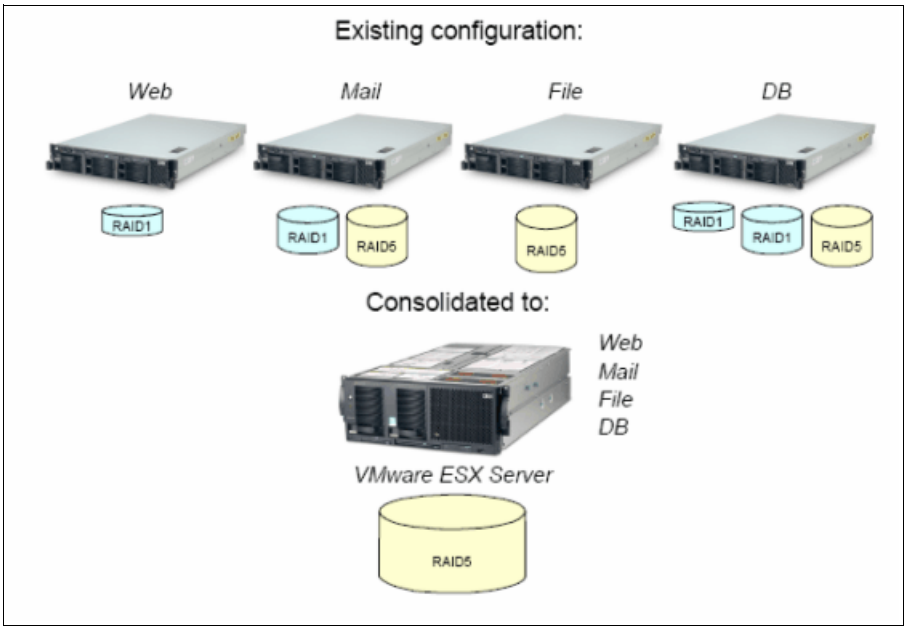

Figure 3-6 Unrealistic storage consolidation

In Figure 3-6, an attempt is made to take the capacity requirement that is calculated from the four existing servers and use that as a guide to size a single RAID 5 array to host all four virtual environments.

It is unlikely that assigning a single RAID 5 LUN to the vSphere Server host in this way supplies enough disk performance to service the virtual machines adequately.

> **Note:** While the following guidelines help to increase the performance of a VMware ESXi Server environment, it is important to realize that the overhead of the VMware ESXi Server virtualization layer still exists. In cases where 100% of the native or non-virtualized performance is required, an evaluation of the practicality of a VMware environment must occur.

An assessment of the performance of the individual environments can show that there is room for consolidation with smaller applications. The larger applications (mail or DB) require that similar disk configurations are given to them in a SAN environment as they had in the previous physical environment.

Figure 3-7 illustrates that a certain amount of storage consolidation might indeed be possible without ignoring the normal disk planning and configuration rules that apply for performance reasons. Servers with a small disk I/O requirement can be candidates for consolidation onto fewer LUNs; however, servers that have I/O-intensive applications require disk configurations that are similar to those of their physical counterparts. It might not be possible to make precise decisions how to best configure the RAID array types and which virtual machine disks must be hosted on them until after the implementation. In an IBM Midrange Storage Systems environment, it is safe to configure several of these options later through the advanced dynamic functions that are available on the storage subsystems.

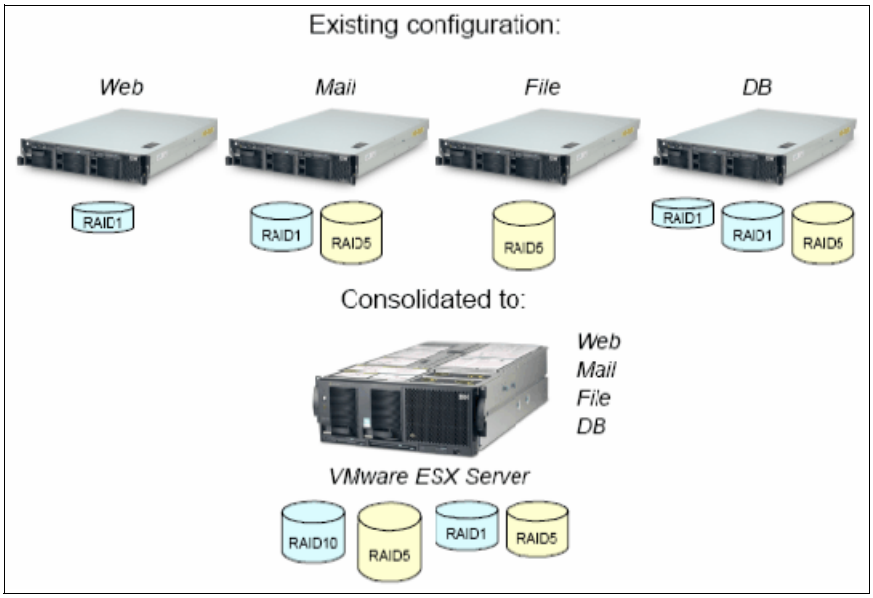

Figure 3-7 Potential realistic storage consolidation

These changes might include the following actions:

- Adding more disks (capacity) to an array using the Dynamic Capacity Expansion function (before creating VMFS datastores on the LUN)
- And, joining two VMFS volumes together in a volume set
- Changing the array type from RAID 5 to RAID 10 using the Dynamic RAID-Level Migration function
- Or, changing the segment sizing to better match our application using the Dynamic Segment Sizing function

Note: Dynamic Volume Expansion is not supported for VMFS-formatted LUNs.

3.4.9 VMware ESX Server storage configurations

There are many ways to implement VMware ESXi Servers that are attached to IBM Midrange Storage Systems. Variants range from the number of HBAs, switches, and paths that are available for a VMware ESXi Server, to multiple VMware ESXi Servers sharing access to logical drives on the IBM Midrange Storage Systems.

Configuring according to a common base of settings allows for growth from one configuration to another with minimal impact. It is therefore recommended to review all of the configurations with your growth plan in mind (as much as possible) so that best practices can be applied from the initial installation and last through a final configuration as it develops over time.

This principle correlates with the installation and configuration details that we give throughout this document. Compile the settings that need to be made into a common set for all configurations with additional minimal changes listed for specific configurations as required.

At the time of writing, DS Storage Manager software is not available for VMware ESX Server operating systems. Therefore, to manage DS5000 Storage Subsystems with your VMware ESXi Server host, you must install the Storage Manager client software (SMclient) on a Microsoft Windows or Linux management workstation, which can be the same workstation that you use for the browser-based VMware ESXi Server Management interface.

VMware ESXi Server restrictions

Certain VMware ESXi server restrictions exist for storage.

SAN and connectivity restrictions

In this section, we discuss SAN and connectivity restrictions for storage:

- VMware ESXi Server hosts support host-agent (out-of-band) managed DS5000 configurations only. Direct-attach (in-band) managed configurations are not supported.
- VMware ESXi Server hosts can support multiple host bus adapters (HBAs) and DS5000 devices. However, there is a restriction on the number of HBAs that can be connected to a single DS5000 Storage Subsystem. You can configure up to two HBAs per partition and up to two partitions per DS5000 Storage Subsystem. Additional HBAs can be added for additional DS5000 Storage Subsystems and other SAN devices, up to the limits of your specific subsystem platform.
- When you use two HBAs in one VMware ESXi Server, LUN numbers must be the same for each HBA that is attached to the DS5000 Storage Subsystem.
- Single HBA configurations are allowed, but each single HBA configuration requires that both controllers in the DS5000 are connected to the HBA through a switch. If they are connected through a switch, both controllers must be within the same SAN zone as the HBA.

  Important: Having a single HBA configuration can lead to the loss of access data in a path failure.

- Single-switch configurations are allowed, but each HBA and DS5000 controller combination must be in a separate SAN zone.

- Other storage devices, such as tape devices or other disk storage, must be connected through separate HBAs and SAN zones.

Partitioning restrictions

In this section, we discuss partitioning restrictions for storage:

- The maximum number of partitions per VMware ESXi Server host, per DS5000 Storage Subsystem, is two.
- All logical drives that are configured for VMware ESXi Server must be mapped to a VMware ESXi Server host group.

> **Note:** Set the host type of all your VMware ESXi Servers to VMware. If you are using the default host group, ensure that the default host type is VMware.

- Assign LUNs to the VMware ESXi Server starting with LUN number 0.
- Do not map an access (UTM) LUN (LUN ID 31) to any of the VMware ESXi Server hosts or host groups. Access (UTM) LUNs are used only with in-band managed DS5000 configurations, which VMware ESXi Server does not support at this time.

Failover restrictions

In this section, we discuss failover restrictions for storage:

- You must use the VMware ESXi Server failover driver for multipath configurations. Other failover drivers, such as RDAC, are not supported in VMware ESXi Server configurations.
- The default failover policy for all DS5000 Storage Subsystems is now most recently used (MRU).
- Use the VMware host type in VMware ESXi Server configurations (2.0 and higher).
- The VMware host type automatically disables AVT/ADT.

Other restrictions

In this section, we discuss other restrictions for storage:

- Dynamic Volume Expansion is not supported for VMFS-formatted LUNs.

> **Recommendation:** Do not boot your system from a SATA device.

Cross-connect configuration for VMware vSphere ESXi

A cross-connect storage area network (SAN) configuration is required when VMware vSphere ESXi hosts are connected to IBM Midrange Storage Systems. Each host bus adapter (HBA) in a vSphere ESXi host must have a path to each of the controllers in the DS storage subsystem. Figure 3-8 on page 48 shows the cross connections for VMware server configurations.

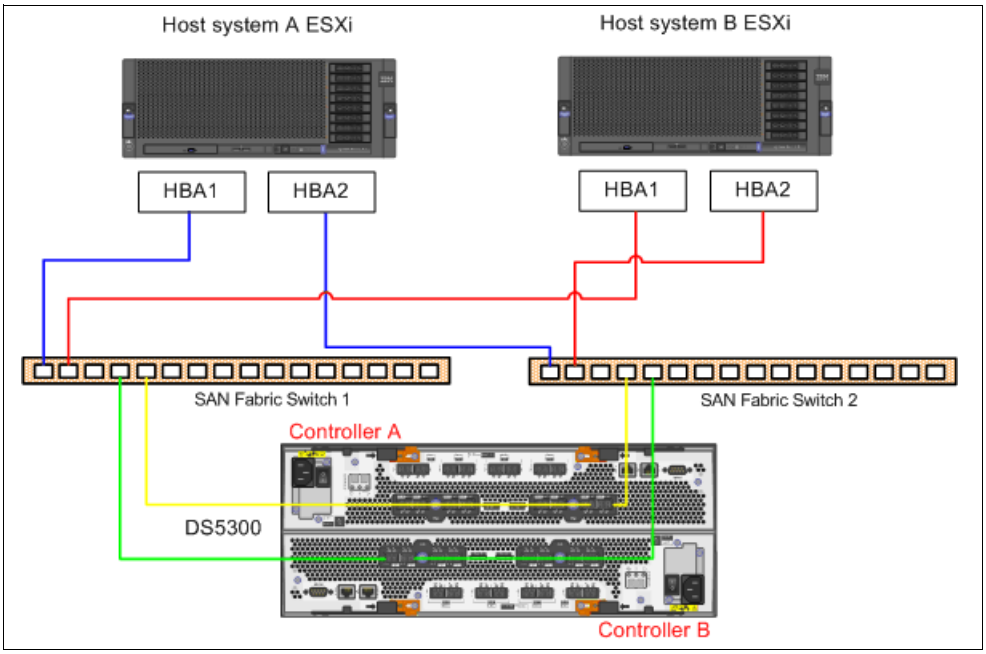

Figure 3-8 Cross-connect configuration for vSphere ESXi connections

A single path to both controllers can lead to either unbalanced logical drive ownership or thrashing under certain conditions. The ownership of all logical drives can be forced to one of the controllers. Depending on which path the VMware ESXi Server finds first, the single active controller on that path can be forced to assume ownership of all LUNs, even those for which that controller is not the preferred owner. This process limits the storage performance for the VMware ESXi Server.

In configurations that involve multiple VMware ESXi Servers that are attached to the IBM DS Midrange Storage Systems, the behavior is exacerbated. When one VMware ESXi Server performs LUN discovery, it can lead to thrashing or bouncing logical drive ownership between the controllers.

To avoid these problems, VMware advises that you set up four paths between the server and the storage system. At least two vSphere ESXi host HBA ports must be used and both HBA ports must see both controllers.

A loss of one of the paths can lead to less than optimal performance because logical drives owned by the controller on the lost path are transferred to the other controller with the surviving path.

If performance is also a concern, consider adding additional connections from one of the storage system's available host ports to the switch.

To preserve logical drive ownership, each controller is cross-connected to the other switch. The disadvantage of this type of switching is that the additional storage system host ports are consumed for the zone and cannot be used to address other performance concerns. If you are seeking to prevent logical drive ownership transfer, consider using the additional controller to switch connections in multiple zones.

The previous recommendations prevent thrashing but do not sufficiently address performance concerns. Only one of the paths can be active, because the first HBA port that the vSphere ESXi host configured is used to communicate with both controllers. To maximize performance, you must spread the load between more paths.

3.4.10 Configurations by function

This section discusses different configurations that are available when using multiple vSphere hosts.

A vSphere VMFS volume can be set as one of these modes:

- A VMFS volume that is visible by only one vSphere ESXi host. We call this mode the *independent* VMFS module. When you have multiple vSphere ESXi hosts, independent VMFS modules can be set through LUN masking (partitioning). This type of configuration is rarely needed and not recommended. It might be implemented when there is a requirement to separate the vSphere hosts' virtual machines. For example, two companies or departments share a SAN infrastructure but need to retain their own servers/applications.
- A VMFS volume that is visible by multiple vSphere ESXi hosts. This mode is the default. This VMFS mode is called *public* VMFS.
- A VMFS volume that is visible by multiple vSphere ESXi hosts and stores virtual disks (`.vmdk`) for split virtual clustering. This VMFS mode is called *shared* VMFS.

Public VMFS might be implemented for the following reasons:

- vSphere high availability (HA) using two (or more) vSphere ESXi hosts with shared LUNs, allowing one vSphere ESXi host to restart the workload of the other vSphere ESXi host if needed. With public VMFS, virtual machines can be run on any host, ensuring a level of application availability in a hardware failure on the vSphere hosts.

  This is possible, as multiple vSphere Servers have access to the same VMFS volumes, and a virtual machine can be started from potentially any vSphere Server host (although not simultaneously). It is important to understand that this approach does not protect against `.vmdk` file corruption or failures in the storage subsystem unless the `.vmdk` file is somehow replicated elsewhere.
- vSphere vMotion allows a running virtual machine to be migrated from one vSphere host to another without being taken offline. In scenarios where a vSphere Server needs to be taken down for maintenance, the virtual machines can be moved without being shut down and as they receive workload requests.
- vSphere Storage vMotion allows you to relocate virtual machine disk files between and across shared storage locations, maintaining continuous service availability.
- Clustering is another method to increase the availability of the environment and is only supported by VMware vSphere using Microsoft Clustering Services (MSCS) on Windows guests. Clustering cannot only transfer the workload with minimal interruption during maintenance, but near continuous *application availability* can be achieved in an OS crash or hardware failure, depending on which of the following configurations is implemented:
  - Local virtual machine cluster increases availability of the OS and application. Many server failures relate to software failure; therefore, implementing this configuration can help reduce software downtime. This configuration does not however increase hardware availability, and this might need to be considered when designing the solution.
  - Split virtual machine cluster increases availability of the OS, application, and vSphere ESXi host hardware by splitting the cluster nodes across two vSphere ESXi hosts. In an OS or vSphere ESXi host hardware failure, the application can fail over to the surviving vSphere host/virtual machine cluster node.

– Physical/virtual machine (hybrid) cluster increases availability of the OS, application, and server hardware where one node is a dedicated physical server (non-ESX), and the other node is a virtual machine. Implementations of this kind are likely to occur where the active node of the cluster requires the power of a dedicated physical server (that is, four or more processors, or more than 3.6-GB memory) but where the failover node can be less powerful, yet remain for availability purposes.

  The physical/virtual machine (hybrid) cluster might also be implemented where a number of dedicated physical servers are used as active nodes of multiple clusters failing over to their passive cluster nodes that all exist as virtual machines on a single vSphere Server. Because it is unlikely that all active nodes fail simultaneously, the vSphere ESXi host might only need to take up the workload of one cluster node at a time, therefore, reducing the expense of replicating multiple cluster nodes on dedicated physical servers. However, the physical server (that is, not the vSphere Server) can only have a non-redundant SAN connection (a single HBA and a single storage controller); therefore, we do not actively advocate the use of this solution.

Configuration examples

The examples in this section show the configuration options that are available when multiple vSphere hosts attach to shared storage partitions.

High availability

The configuration in Figure 3-9 shows multiple vSphere Servers connected to the same IBM Midrange Storage Subsystem with a logical drive (LUN) shared between the servers. (This configuration can have more than just two vSphere ESXi hosts.)

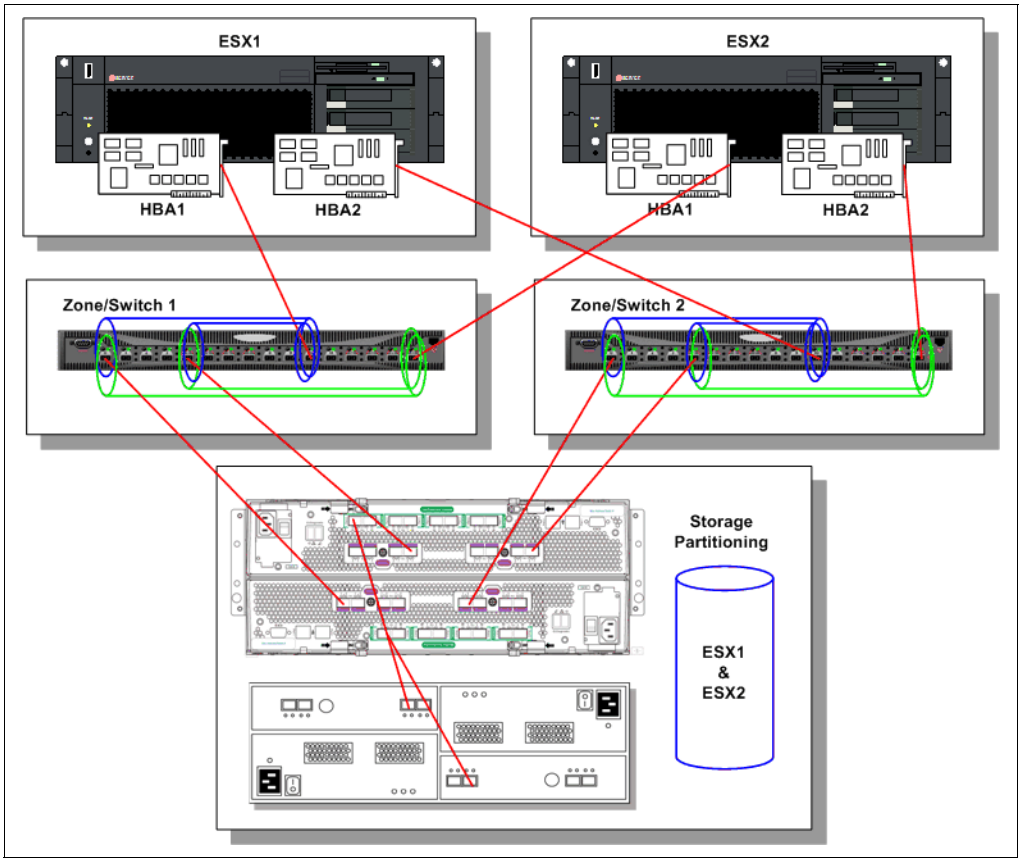

Figure 3-9 Multiple servers sharing a storage partition configuration sample

vSphere vMotion

The configuration for vSphere vMotion functions the same as the configuration in the preceding high availability (HA) section.

Clustering (guest OS level)

> **Note:** Guest Clustering is only supported by VMware using Microsoft Clustering Services (MSCS) on Windows guests, and only in a two node per cluster configuration.

There are many ways to implement MSCS with VMware vSphere ESXi, depending upon the level of requirements for high availability and whether physical servers are included in the mix.

In the following sections, we review the ways that MSCS might be implemented.

A local virtual machine cluster configuration is shown in Figure 3-10. VMFS volumes are used with the access mode set to public for all of the virtual machine disks. This design requires that both virtual machines must run on top of the same ESXi physical host. Therefore, you lose the hardware redundancy provided by VMware HA functions described previously.

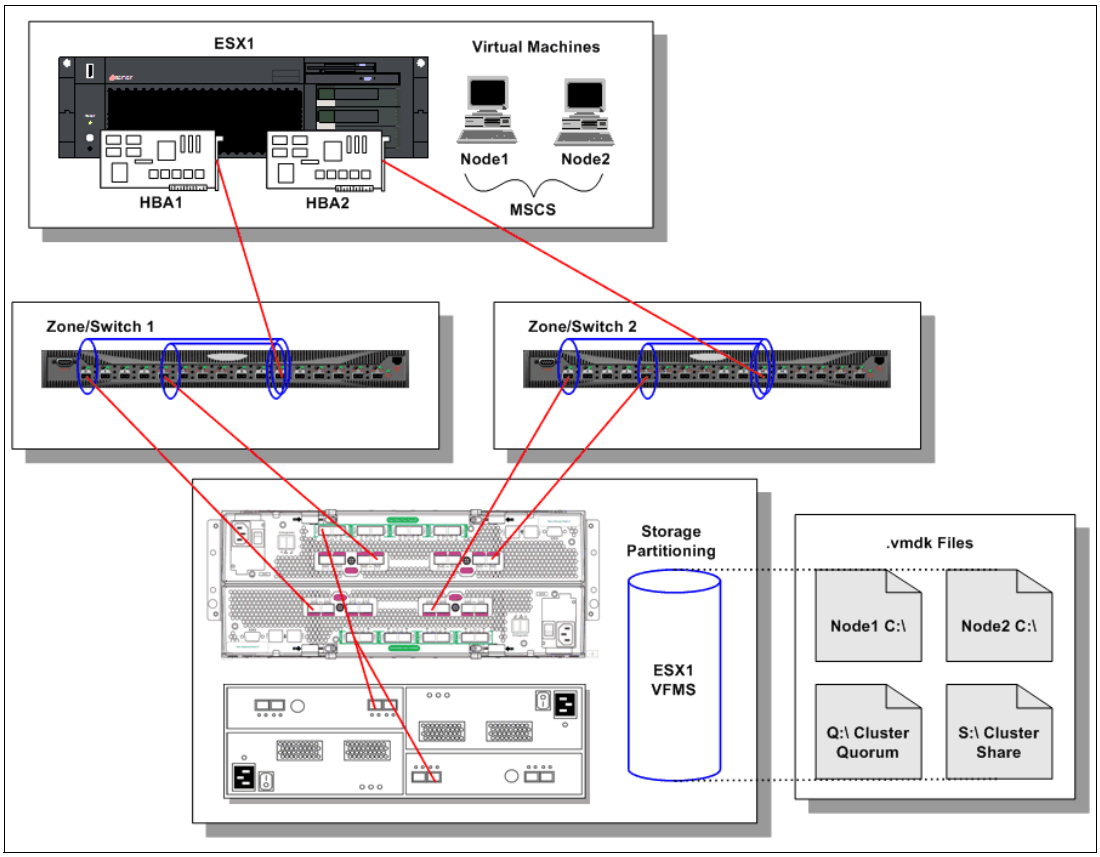

Figure 3-10 Local virtual machine cluster

A split virtual machine cluster configuration is shown in Figure 3-11 on page 52. VMFS volumes are used with the access mode set to public for all virtual machine `.vmdk` files (OS boot disks). Raw volumes are used for the cluster shares. The cluster shares can be `.vmdk` files on shared VMFS volumes, but limitations make using raw volumes easier to implement. There are other caveats about VMware function availability when using MSCS clustering functions with virtual machines.

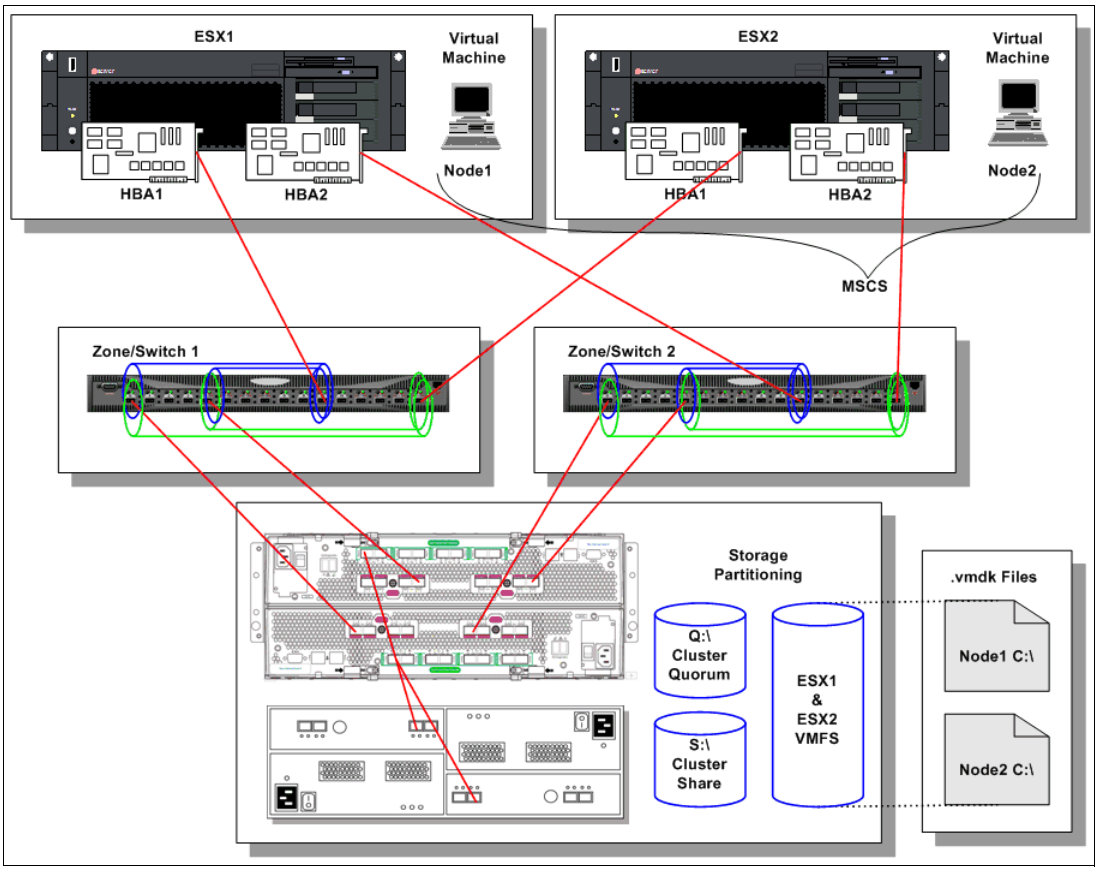

Figure 3-11 Split virtual machine cluster

For more information about vSphere ESXi and Microsoft Cluster Services implementation and support, see this website:

http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-mscs-guide.pdf

For Microsoft Cluster Service (MSCS) support on ESX/ESXi, see VMware KB Article 1004617:

http://kb.vmware.com/kb/1004617

3.4.11 Zoning

Zoning for a VMware vSphere ESXi Server environment is essentially the same as a non-ESXi environment. It is considered a good practice to separate the traffic for stability and management reasons. Zoning follows your standard practice where, in reality, it is likely that multiple servers with different architectures (and potentially different cable configurations) are attached to the same IBM Midrange Storage Subsystem. In this case, additional hosts are added to the appropriate existing zones, or separate zones are created for each host.

A cross-connect SAN configuration is required when vSphere ESXi hosts are connected to IBM Midrange Storage Systems. Each HBA in a vSphere ESXi host must have a path to each of the controllers in the DS Storage Subsystem.

Figure 3-12 on page 53 shows a sample configuration with multiple switches and multiple zones.

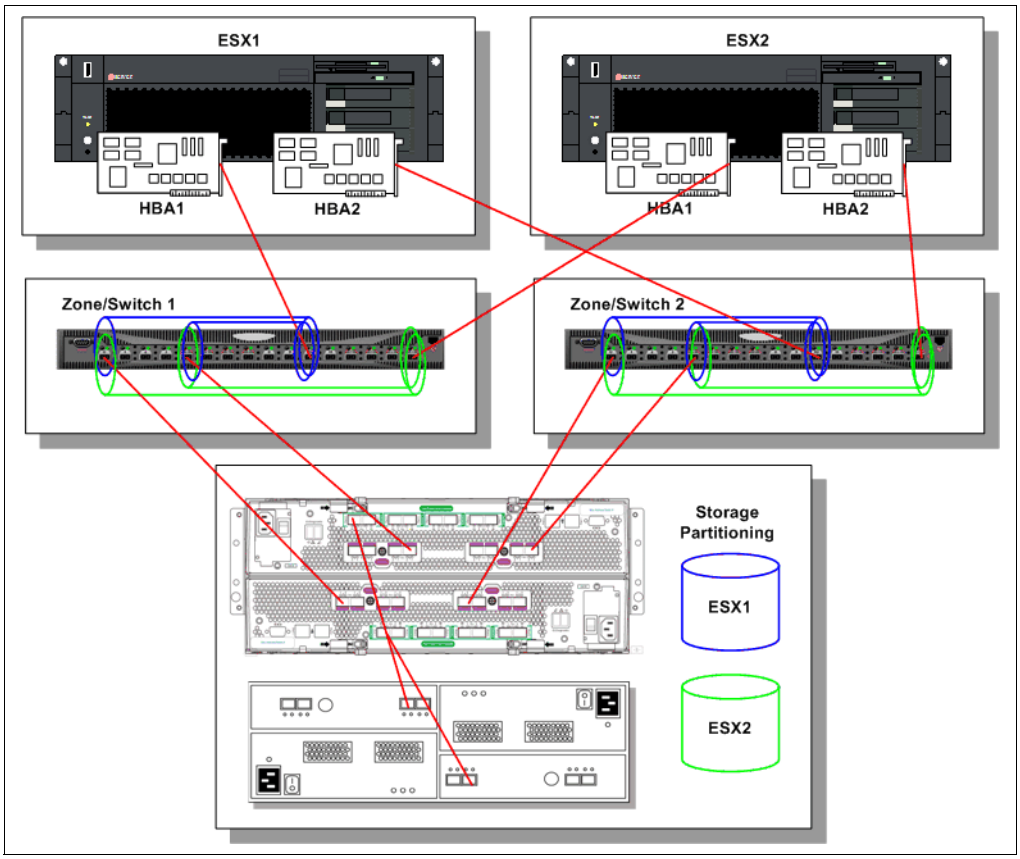

Figure 3-12 Multiple switches with multiple zones

For more information about zoning the SAN switches, see *Implementing an IBM b-type SAN with 8 Gbps Directors and Switches*, SG24-6116, or *Implementing an IBM/Cisco SAN*, SG24-7545.

4

Planning the VMware vSphere Server design

Careful planning is essential to any new VMware vSphere installation. In this chapter, we provide guidelines to help you to plan your VMware vSphere environment.

4.1 Considering the VMware vSphere Server platform

The server platform contains the server hardware and the system software. When considering the hardware and operating system on which you want to run an application, such as Oracle database, there are many issues to consider:

- High availability: Is Oracle Real Application Clusters (Oracle RAC) needed at the guest OS level to provide HA capabilities? Are other clustering solutions, such as Microsoft Clustering Services, required at the guest OS level (virtual machines)? Is vSphere DRS or vSphere vMotion needed to support high availability?
- Scalability: If the database is expected to grow and more hardware resources are required to provide the future performance that the client needs, Oracle can provide a scalable approach to accommodate growth potential. Oracle databases, vSphere HA cluster, vSphere DRS, and vSphere Motion can accommodate scalability for virtual machines.
- Number of concurrent sessions: Determine the number of concurrent sessions and the complexity of these transactions before deciding what virtual hardware and operating system to use for the database.
- Amount of disk I/Os per second (IOPS): If the database performs many IOPS, consider vSphere ESXi Server hardware that supports multiple host bus adapters (HBAs). Also, consider the number of disk drive spindles needed to provide the IOPS that are forecasted by the application.
- Size: If you have a small database or a few users, a small-to-medium size hardware platform is justified.
- Cost: If cost is a factor for purchasing hardware, the x86 platform is a more cost-effective platform. The x86 provides outstanding performance for the money.

4.1.1 Minimum server requirements

For a complete up-to-date list of the prerequisites for installing vSphere ESXi, see the following VMware website:

http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-installation-setup-guide.pdf

4.1.2 Maximum physical machine specifications

For more information about the maximum hardware capabilities of the vSphere ESXi, see the following VMware website:

http://www.vmware.com/pdf/vsphere5/r50/vsphere-50-configuration-maximums.pdf

4.1.3 Recommendations for enhanced performance

The following list outlines a basic configuration. In practice, you can use multiple physical disks, which can be SCSI disks, Fibre Channel (FC) LUNs, or RAID LUNs.

The following items are recommended for enhanced performance:

- A second disk controller with one or more drives, dedicated to the virtual machines. The use of PVSCSI is another alternative for hardware or applications that drive a high amount of I/O throughput. Refer to this article for further information:

  http://kb.vmware.com/kb/1010398

- Sufficient RAM for each virtual machine and the support
- Dedicated Ethernet cards for network-sensitive virtual machines.

For best performance, all of the data that is used by the virtual machines must be on the physical disks allocated to virtual machines. Therefore, these physical disks must be large enough to hold disk images that will be used by all the virtual machines.

Similarly, you must provide enough RAM for all of the virtual machines plus the local support console-related services.

> **Note:** To ensure the best possible I/O performance and workload management, VMware vSphere ESXi provides its own drivers for supported devices. Be sure that the devices you plan to use in your server are supported.

For additional detail on I/O device compatibility, download the VMware ESX Server I/O Adapter Compatibility Guide from the VMware website:

http://partnerweb.vmware.com/comp_guide/pdf/vi_io_guide.pdf

You must ensure that there is enough free disk space available to install the guest operating system and applications for each VM on the disk that they will use.

For general performance recommendations, check the updated “Performance Best Practices document”:

http://www.vmware.com/pdf/Perf_Best_Practices_vSphere5.0.pdf

4.1.4 Considering the server hardware architecture

Available bandwidth depends on the server hardware. The number of buses adds to the aggregate bandwidth, but the number of HBAs sharing a single bus can throttle the bandwidth.

Calculating aggregate bandwidth

An important limiting factor in I/O performance is the I/O capability of the server that hosts the application. The aggregate bandwidth of the server to the storage system is measured in MBps and contains the total capability of the buses to which the storage system is connected. For example, a 64-bit PCI bus clocked at 133 MHz has a maximum bandwidth calculated by the following formula:

```
PCI Bus Throughput (MB/s) = PCI Bus Width / 8 * Bus Speed

64-bit /8 * 133 MHz = 1062 MB/s ~ = 1GB/s
```

Table 4-1 shows PCI-X bus throughput.

Table 4-1 PCI-X bus throughput

MHz	PCI bus width	Throughput (MB/s)
66	64	528
100	64	800
133	64	1064
266	64	2128

MHz	PCI bus width	Throughput (MB/s)
533	64	4264

Sharing bandwidth with multiple HBAs

Multiple HBAs on a bus share this single source of I/O bandwidth. Each HBA might have multiple FC ports, which typically operate at 1 Gbps, 2 Gbps, 4 Gbps, or 8 Gbps. As a result, the ability to drive a storage system can be throttled by either the server bus or by the HBAs. Therefore, whenever you configure a server or whenever you analyze I/O performance, you must know how much server bandwidth is available and which devices are sharing that bandwidth.

VMware vSphere ESXi path failover and load distribution

vSphere ESXi has a built-in failover driver to manage multiple paths called *Native Multipath Plug-In (NMP)*. At startup, or during a rescan that might be issued from the vCenter Console, all LUNs or logical drives are detected. When multiple paths to a logical drive are found, the failover driver (NMP) is configured and uses the default Most Recently Used (MRU) policy. The IBM Midrange Storage Subsystem is an active/passive storage system where logical drive ownership is distributed between two controllers. The individual logical drives are presented to the vSphere ESXi host by both controllers. The vSphere ESXi host configures both controllers as possible owners of a LUN, even though only one controller owns the LUN. The ESXi host is able to distinguish between the active controller, the controller that owns a logical drive, and the passive controller. The active controller is the preferred controller.

Note: Additional multi-path drivers, such as the Redundant Disk Array Controller (RDAC), are not supported by vSphere ESXi.

The Native Multipath plug-in (NMP) failover driver provides three policies or Path Selection Plug-Ins (PSPs):

- Fixed: The fixed policy is intended for active/active devices and is not recommended for the IBM Midrange Storage Systems. If the fixed policy is selected for logical drives that are presented by the IBM Midrange Storage Subsystem, thrashing can result.
- Most recently used (MRU): The MRU policy is intended for active/passive devices and is a requirement for configurations with IBM Midrange Storage Systems.

  Note: Using active/passive arrays with a fixed path policy can potentially lead to path thrashing. For more information about active/active and active/passive disk arrays and path thrashing, see the *SAN System Design and Deployment Guide*:

  http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-storage-guide.pdf

- Round robin (RR): The host uses an automatic path selection algorithm rotating through all active paths when connecting to active/passive arrays, or through all available paths when connecting to active/active arrays. RR is the default for a number of arrays and can be used with both active/active and active/passive arrays to implement load balancing across paths for different LUNs.

Concerns and recommendations

A single path to both controllers can lead to either unbalanced logical drive ownership or thrashing under certain conditions. The ownership of all logical drives can be forced to one of the controllers. Depending on the path that the vSphere ESXi host finds first, the single active

controller on that path can be forced to assume ownership of all LUNs, even those for which that controller is not the preferred owner. This process limits the storage performance for the vSphere ESXi host.

In configurations involving multiple vSphere ESXi hosts attached to the IBM Midrange Storage Systems, the behavior is exacerbated. When one ESXi host performs LUN discovery, logical drive ownership can lead to thrashing or bouncing ownership between the controllers.

To avoid these problems, VMware advises that you set up four paths between the server and the storage system as shown in Figure 4-1. At least two vSphere ESXi host HBA ports must be used and both HBA ports must reach both controllers.

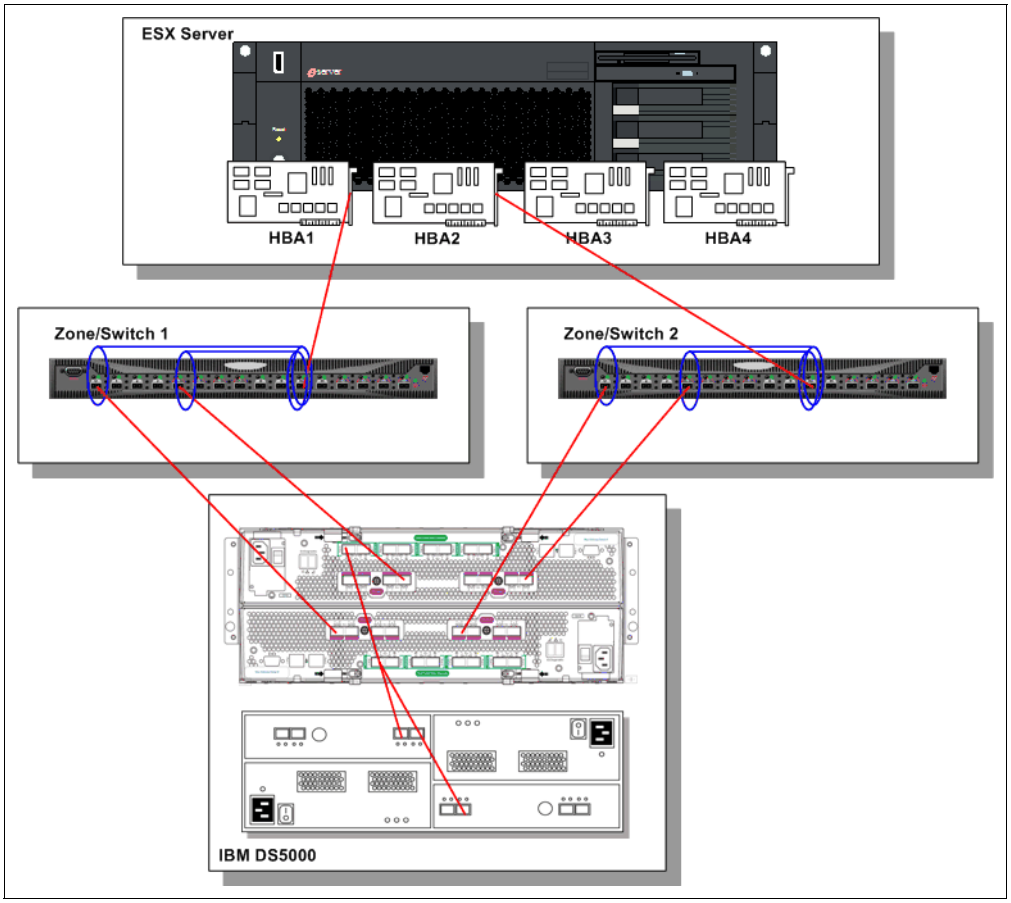

Figure 4-1 Paths between the vSphere ESXi host and the DS5000 Storage System

To preserve logical drive ownership, each controller is cross-connected to the other switch. The disadvantage of this type of switching is that the additional storage system host ports are consumed for the zone and cannot be used to address other performance concerns. If you are seeking to prevent logical drive ownership transfer, consider using the additional controller to switch connections in multiple zones.

The previous recommendations prevent thrashing but do not sufficiently address performance concerns. Only one of the paths can be active, because the first HBA port that is configured by the vSphere ESXi host is used to communicate with both controllers. To maximize performance, you must spread the load between more paths.

Example of server path failover and load distribution

A vSphere ESXi host has eight paths that consist of eight server FC HBA ports (four dual-port FC HBAs), eight storage system host ports, and a pair of switches. In a simple configuration that depends only on the ESXi host, the MRU failover policy implements all individual paths. However, the additional ESXi host's HBA ports do not add benefit because only two of the eight paths are used.

To increase the I/O performance, spread the load across more ESXi host's HBA ports and more storage system host ports. You can implement this process by creating multiple groups of four-path configurations.

There are several elements that are necessary to perform this task:

1. Combine pairs of vSphere ESXi host HBA ports with pairs of IBM DS5000 storage subsystem host ports by using zoning on the SAN switches.
2. Logically divide the vSphere ESXi host's pairs of HBA ports into separate storage partitions on the storage system.
3. Assign specific logical drives, which are balanced between controllers, to the storage partition.

Zoning the switches defines a specific path to the storage system. This path is refined with the storage partitioning and the creation of the logical host definition. After specific LUNs are presented to the logical host, the path definition is complete.

You can benefit from this strategy by the number of supported LUNs. The vSphere ESXi host supports a maximum of 256 LUNs or paths to LUNs. Relying on just the failover driver's MRU policy severely limits the actual number of LUNs found. In practice, only sixteen actual LUNs are supported in an eight-server port configuration.

In a configuration with 44 physical LUNs, a certain path shows 88 LUNs, including active LUNs and standby LUNs. If there are eight FC HBA ports, 88 LUNs are available on each port. The resulting 704 LUNs greatly exceed vSphere ESXi host capabilities. By following the recommended practice, you can increase the quantity of supported LUNs to 128.

The multiple zone and storage partitioning configuration better distributes the load by using four of eight available paths to the storage system. You can scale this strategy by adding additional pairs of vSphere ESXi host HBA ports, zones, storage system host ports, and storage partitions.

Figure 4-2 on page 61 shows the recommended best practice for configuring multiple zones and storage partitioning. If implemented in a clustered vSphere ESXi host environment, all of the vSphere ESXi hosts must share a common configuration.

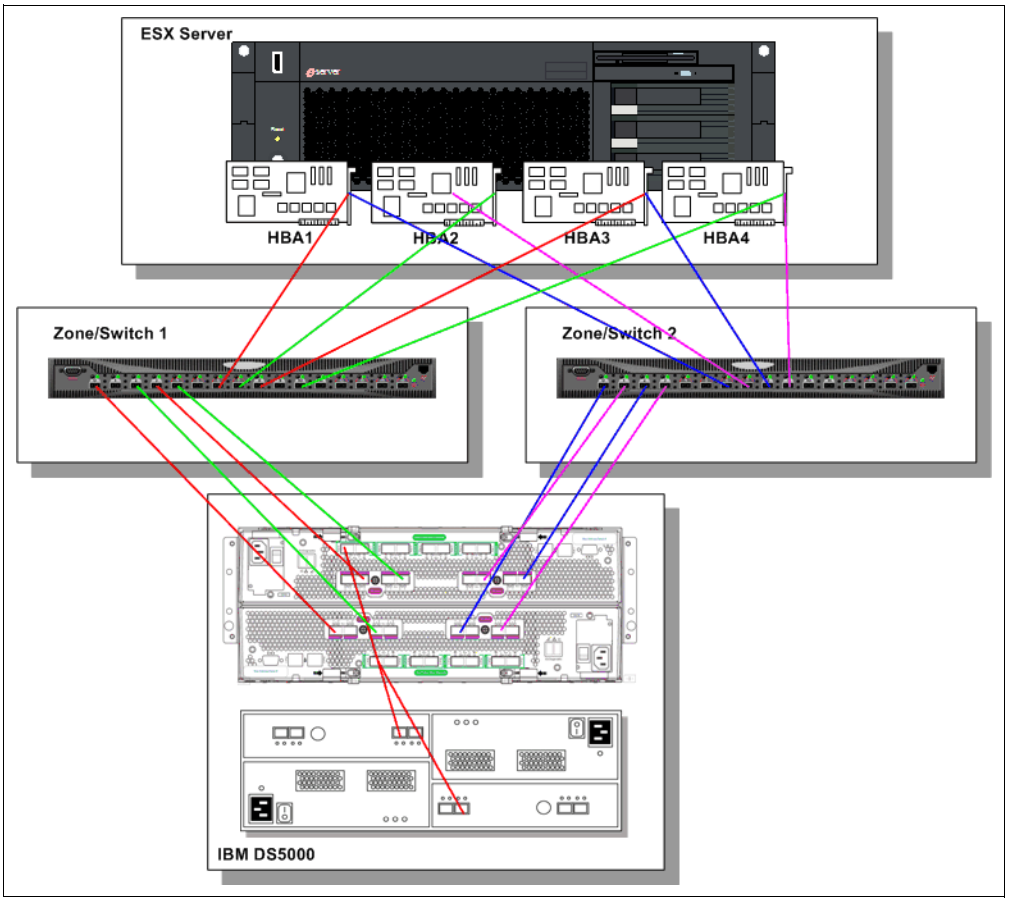

Figure 4-2 Best practice for configuring multiple zones and storage partitioning

4.1.5 General performance and sizing considerations

The goal of this section is not to describe an approach for sizing and performance, but rather to point out specific characteristics of a vSphere ESXi host implementation.

When it comes to performance, it is important to remember that you must not automatically expect a virtual machine to exhibit the same performance characteristics of the physical server it emulates. This is not to say that a virtual machine cannot cope with performance-intense workloads. However, if achieving the highest performance is a major goal or requirement, VMware might not be the correct choice. The same goes for workloads requiring large symmetrical multiprocessor systems (SMPs) (typically of more than two CPUs). In any case, it is important to agree on the minimum acceptable performance figures, and then document them to perform a Proof of Concept (POC), if performance is the main concern.

CPU overhead

The virtualization process introduces a CPU overhead that needs to be considered when sizing VMware solutions. The percentage of overhead depends on the nature of the workload. As a guideline (and from numbers observed with actual implementations), you can use the following approach:

- Computation-intense workload: overhead negligible (1 - 3%)
- Disk I/O-intense workload (less than 10%)
- Network I/O-intense workload (5% or even greater)

In reality, you might see a mixed workload that results in an average overhead of 10%. Software iSCSI overhead also was reduced compared to previous versions of VMware ESX Server. ESXi 5.0 includes the following performance enhancements:

- 160 logical CPUs and 2048 virtual CPUs per host: ESXi 5.0 provides headroom for more virtual machines per host and the ability to achieve even higher consolidation ratios on larger machines that are considered “monster” virtual machines.
- 64-bit VMkernel: The VMkernel, a core component of the ESXi hypervisor, is since version 4.0 a 64-bit. This provides greater host physical memory capacity and more seamless hardware support than earlier releases.

The vSphere ESXi scheduler includes several features and enhancements that help improve the throughput of all workloads, with notable gains in I/O intensive workloads:

- Relaxed co-scheduling of vCPUs, introduced in earlier versions of VMware ESX Server, is further fine-tuned, especially for SMP virtual machines.
- vSphere ESXi (4.0 and higher) scheduler utilizes new finer-grained locking that reduces scheduling overheads in cases where frequent scheduling decisions are needed.
- The new scheduler is aware of processor cache topology and takes into account the processor cache architecture to optimize CPU usage.
- For I/O-intensive workloads, interrupt delivery and the associated processing costs make up a large component of the virtualization overhead. The scheduler enhancements greatly improve the efficiency of interrupt delivery and associated processing.

4.2 Operating system considerations

This section describes items to consider when using a particular operating system and how that operating system affects partition alignments.

4.2.1 Buffering the I/O

The type of I/O (buffered or unbuffered) provided by the operating system to the application is an important factor in analyzing storage performance issues. Unbuffered I/O (also known as *raw I/O* or *direct I/O*) moves data directly between the application and the disk drive devices. Buffered I/O is a service provided by the operating system or by the file system. Buffering improves application performance by caching write data in a file system buffer, which the operating system or the file system periodically moves to permanent storage. Buffered I/O is generally preferred for shorter and more frequent transfers. File system buffering might change the I/O patterns generated by the application. Writes might coalesce so that the pattern seen by the storage system is more sequential and more write-intensive than the application I/O itself. Direct I/O is preferred for larger, less frequent transfers and for applications that provide their own extensive buffering (for example, Oracle). Regardless of I/O type, I/O performance generally improves when the storage system is kept busy with a steady supply of I/O requests from the host application. Become familiar with the parameters that the operating system provides for controlling I/O (for example, maximum transfer size).

4.2.2 Aligning host I/O with RAID striping

For all file systems and operating system types, you must avoid performance degrading segment crossings. You must not let I/O span a segment boundary. Matching I/O size (commonly, by a power of two) to array layout helps maintain aligned I/O across the entire disk drive. However, this statement is true only if the starting sector is correctly aligned to a

segment boundary. Segment crossing is often seen in the Microsoft Windows operating system. The way that partition alignment works depends on the version of Windows used and the version in which the partition alignment was created. In Windows Server 2008, partition alignment is usually performed by default. The default for disks larger than 4 GB is 1 MB; the setting is configurable and is found in the registry:

```
HKLM\SYSTEM\CurrentControlSet\Services\VDS\Alignment
```

For partitions that are created by Windows 2000 Server or Windows Server 2003, start at the 64th sector. Starting at the 64th sector causes misalignment with the underlying RAID striping and allows the possibility for a single I/O operation to span multiple segments.

Because the alignment of file system partitions can affect performance, since vSphere ESXi 5.0, every new Virtual Machine File System (VMFS-3) or VMFS-5 partition is automatically aligned along the 1-MB boundary. For a VMFS-3 partition that is created by using an earlier version of ESX/ESXi that aligned along the 64-KB boundary, and if that file system is then upgraded to VMFS-5, it retains its 64-KB alignment and needs to be aligned manually.

4.2.3 Recommendations for host bus adapter settings

The following HBA guidelines are recommended:

- Use the default HBA settings of the HBA vendor.
- Use the same model of HBA in the vSphere ESXi host. Mixing HBAs from various vendors in the same vSphere ESXi host is not supported.
- Ensure that the Fibre Channel HBAs are installed in the correct slots in the host, based on slot and bus speed. Balance PCI bus load among the available buses in the server.
- Make sure that each server has enough HBAs to allow maximum throughput for all the applications hosted on the server for the peak period. I/O spread across multiple HBAs provides higher throughput and less latency for each application.
- To provide redundancy in an HBA failure, make sure that the server is connected to a dual redundant fabric.

4.2.4 Recommendations for Fibre Channel switch settings

The following Fibre Channel switch settings are recommended:

- Enable in-order delivery: Recommended settings are available from the supplier of the storage system. For example, on Brocade switches, verify that the In-Order Delivery parameter is enabled.
- Inter-switch links: In a multi-switch SAN fabric, where I/O traverses inter-switch links, make sure to configure sufficient inter-switch link bandwidth.
- Disable trunking on the Fibre Channel switch: When using a Cisco Fibre Channel switch, the IBM Midrange Storage Subsystem host ports and the Fibre Channel HBA ports on the server cannot be configured on the switch with the trunking enabled. The use of the trunking feature can cause thrashing of logical drive ownership on the storage system. Trunking is set to automatic by default. You can change trunking to non-trunk under the Trunk Config tab.

4.2.5 Using Command Tag Queuing

Command Tag Queuing (CTQ) refers to the controller's ability to line up multiple SCSI commands for a single LUN and run the commands in an optimized order that minimizes rotational and seek latencies. Although CTQ might not help in certain cases, such as single-threaded I/O, CTQ never hurts performance and therefore is generally recommended. The IBM models vary in CTQ capability, generally up to 2048 per controller. Adjust the CTQ size to service multiple hosts. CTQ is enabled by default on IBM storage systems, but you also must enable CTQ on the host operating system and on the HBA. Refer to the documentation from the HBA vendor.

The capability of a single host varies by the type of operating system, but you can generally calculate CTQ as follows:

```
OS CTQ Depth Setting = Maximum OS queue depth (< 255) /Total # of LUNs
```

Note: If the HBA has a lower CTQ capacity than the result of the previously mentioned calculation, the HBA's CTQ capacity limits the actual setting.

4.2.6 Analyzing I/O characteristics

Analyze the application to determine the best RAID level and the appropriate number of disk drives to put in each array:

- Is the I/O primarily sequential or random?
- Is the size of a typical I/O large (> 256 KB), small (< 64 KB), or in-between?
- If this number is unknown, calculate a rough approximation of I/O size from the statistics reported by the IBM DS Storage Manager Performance Monitor using the following formula:

  ```
  Current KB/second ÷ Current I/O/second = KB/I/O
  ```

- What is the I/O mix, that is, the proportion of reads to writes? Most environments are primarily Read.
- What read percent statistic does IBM DS Storage Manager Performance Monitor report?
- What type of I/O does the application use: buffered or unbuffered?
- Are concurrent I/Os or multiple I/O threads used?

In general, creating more sustained I/O produces the best overall results, up to the point of controller saturation. Write-intensive workloads are an exception to this rule.

4.2.7 Using VMFS for spanning across multiple LUNs

The following VMFS best practices are recommended:

- One VMFS volume per LUN and carve up the VMFS volume into many VMDKs.

  Although vSphere ESXi supports using several smaller LUNs for a single VMFS, spanning LUNs is not recommended. You can improve performance by using a single, correctly sized LUN for the VMFS. Fewer larger LUNs are easier to manage.
- Separate heavy workloads onto separate LUNs as needed. You can create several VMFS volumes to isolate I/O-intensive virtual machines or use raw device mappings (RDMs) as an alternate way of isolating I/O-intensive virtual machines to reduce contention.
- Mix virtual machines with different peak access times.

For more information, see the following document:

http://www.vmware.com/pdf/vmfs-best-practices-wp.pdf

4.2.8 Using the VAAI plug-in

The VMware vStorage API for Array Integration (VAAI) enables disk arrays to provide the native functions already performed today. Performance is enhanced for standard vSphere operations as shown in Figure 4-3.

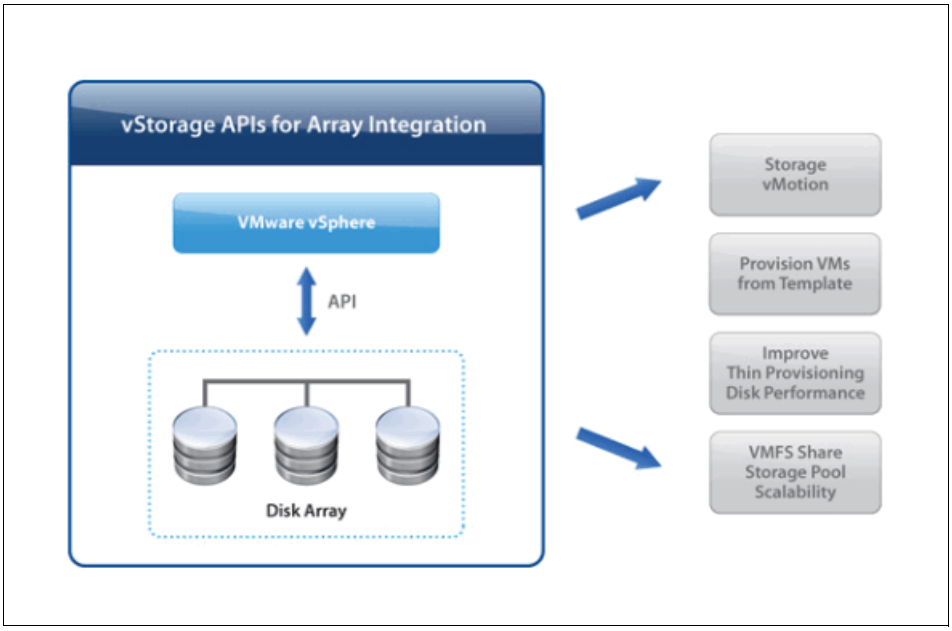

Figure 4-3 VMware vStorage API for Array Integration functionality

The VAAI plug-in also provides a more granular means of handling data, eliminating the need for VMFS to use storage from the storage pool. Instead, simple operations, such as moving a VM, starting it, creating a VM from a template, taking snapshots, or even stopping a VM, can be performed more efficiently. This is especially true when many vSphere hosts share a single datastore. The Storage APIs for Array Integration are included with the Enterprise and Enterprise Plus editions of VMware vSphere.

Note: The VAAI plug-in is supported by IBM System Storage DCS3700 and DS3500 only.

Part 2

Configuration

In Part 2, we provide detailed steps for the VMware ESXi Server installation and storage-related setup and configuration.

5

VMware ESXi Server and storage configuration

In this chapter, we outline the process for installing the VMware ESXi 5 Server and the configuration settings that are necessary to connect to IBM System Storage Midrange Disk. In this case, we use DS5000 storage subsystem but you can follow the same procedure with DCS3700 or DS3500 storage subsystems.

5.1 Storage configuration

First, you need to configure your storage on the IBM Midrange Storage Subsystem. Complete the following steps as part of the storage configuration:

1. Connections:
   - Fibre Channel: Zone your VMware ESXi Server to your IBM Midrange Storage Subsystem. Ensure that your VMware environment has sufficient paths and connections for redundancy and high availability. See Figure 4-1 on page 59.
   - iSCSI: Check the storage subsystem iSCSI configuration.
2. Create a logical unit number (LUN): Create a LUN of a size that fits your VMware partition requirements.
3. Storage partition: From the IBM DS Storage Manager mapping window, create a VMware host and define the host ports for the following components:
   - Fibre Channel HBAs, as shown in Figure 5-1
   - iSCSI host interface card (HICs), as shown in Figure 5-2 on page 71

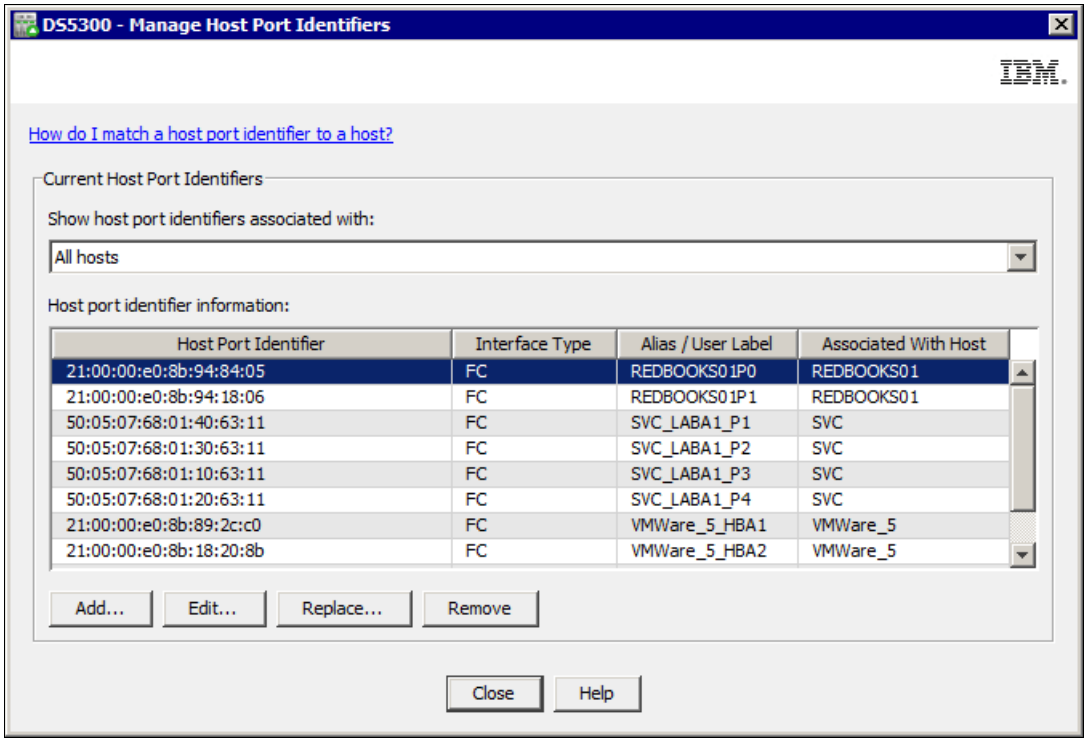

Figure 5-1 Storage partitioning for an FC connection

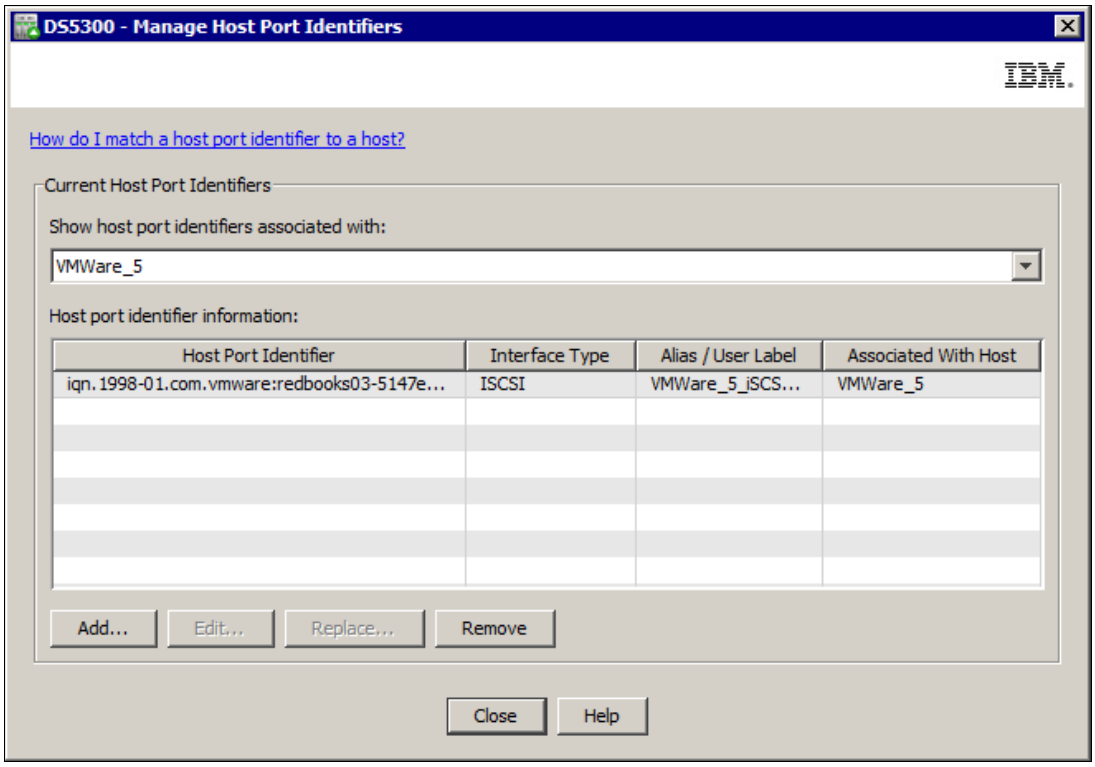

Figure 5-2 Storage partitioning for an iSCSI connection

Note: Use VMware for the host type for all VMware hosts. If you are using the default host group, ensure that the default host type is VMware.

4. LUN mapping: Map the LUN that we created in step 2 on page 70 to the host partition that you created in the preceding step.

   Figure 5-3 shows an example of a valid LUN mapping for installation purposes.

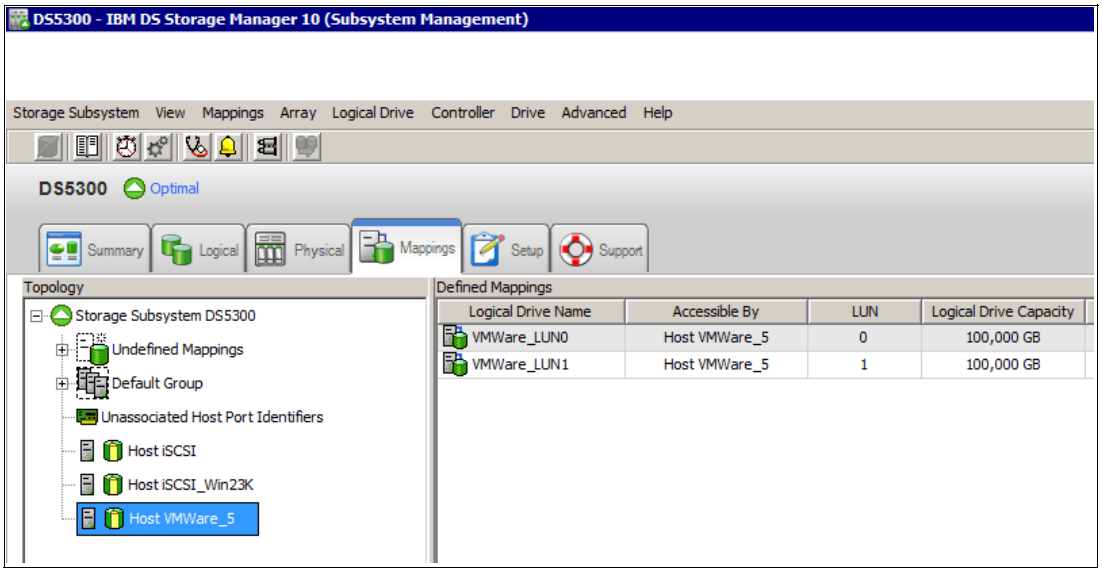

Figure 5-3 LUN mapping for VMware installation

For step-by-step instructions to configure IBM Midrange Storage Systems, see these resources:

- *IBM System Storage DS5000 Series Hardware Guide*, SG24-8023
- *IBM System Storage DS5000 Series Implementation and Best Practices Guide*, SG24-8024
- *IBM System Storage DS3500 Introduction and Implementation Guide*, SG24-7914
- *IBM System Storage DCS3700 Introduction and Implementation Guide*, SG24-8037

Note: The IBM System Storage DS® Storage Manager cannot be installed on the VMware ESXi 5 Server, but instead on a Linux or Windows management workstation. However, it can be installed on the same management workstation that you used for the browser-based VMware Management Interface.

Important: Note the restrictions listed in "VMware ESXi Server restrictions" on page 46 for VMware ESXi Server storage configurations.

5.1.1 Notes about mapping LUNs to a storage partition

We provide recommendations about LUN mapping that are specific to VMware ESXi Servers (ESXi 5 and older).

Use this information when you map your LUNs on the VMware ESXi Server:

- Map the LUNs using consecutive numbers, starting with LUN 0. Map LUNs to numbers 0, 1, 2, 3, 4, 5, and so on, without skipping any numbers.
- On each partition, you must map a LUN 0.
- If your configuration does not require LUN sharing (single or multiple independent VMware ESXi Servers and a local virtual cluster), each logical drive must be mapped either directly to a host, or to a host group with a single host as a member.
- Default LUN ID 31 (Access Logical Drive) is not supported and must be removed from the mappings list for each VMware ESXi host and host group.
- LUN sharing across multiple VMware ESXi Servers is only supported when you are configuring VMware vMotion enabled hosts or Microsoft Cluster nodes. On LUNs that are mapped to multiple VMware ESXi Servers, you must change the access mode to Shared.

  You can map the LUNs to a host group for the VMware ESXi Servers, so they are available to all members of the host group. For additional information about Windows Clustering with VMware ESX Server, see the *VMware ESXi Server Installation Guide*:

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

5.1.2 Steps for verifying the storage configuration for VMware

Complete the following steps to verify that your storage setup is fundamentally correct and that you can see the IBM Midrange Storage Subsystem on your VMware ESXi Server:

1. Boot the server.
2. On the initialization of the QLogic BIOS, press Ctrl+Q to enter the Fast!UTIL setup program.

3. Select the first host bus adapter that is displayed in the Fast!UTIL panel as shown in Figure 5-4.

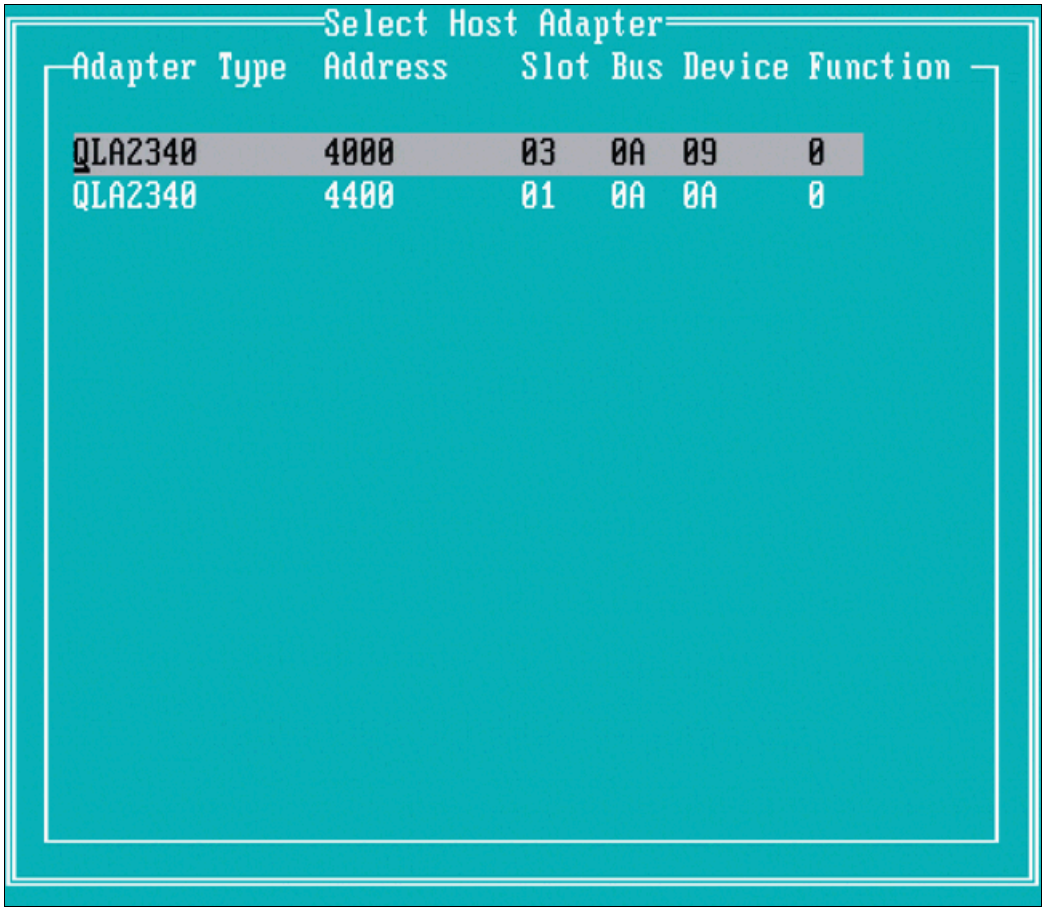

Figure 5-4 Fast!UTIL Select Host Adapter panel

4. Select **Host Adapter Settings**, and press Enter.
5. Select **Scan Fibre Devices**, and press Enter. Figure 5-5 on page 74 shows the resulting output.

   If you do not see a DS5000 controller, verify the cabling, switch zoning, and LUN mapping.

```
Scan Fibre Channel Loop
ID      Vendor   Product          Rev     Port Name        Port ID

128     No device present
129     IBM      1818      FAStT  0730  201700A0B86E32A0    010000
130     IBM      1818      FAStT  0730  202600A0B86E32A0    010700
131     No device present
132     No device present
133     No device present
134     No device present
135     No device present
136     No device present
137     No device present
138     No device present
139     No device present
140     No device present
141     No device present
142     No device present
143     No device present
```

Figure 5-5 Scanning for Fibre Channel devices

Note: Depending on how the configuration is cabled, you might see multiple instances.

If using iSCSI, connection can be verified after the VMware ESXi Server is installed and set up, by pinging the IP address of the storage subsystem iSCSI HIC ports.

5.2 Installing the VMware ESX Server

We provide the procedure and details to install the VMware ESX server.

5.2.1 Prerequisites

Refer to the minimum server hardware configuration requirements that we described in 4.1.1, “Minimum server requirements” on page 56.

You need a VMware ESXi 5 Installer CD/DVD or a USB flash drive. In addition, complete the following information in Table 5-1 before you begin.

Table 5-1 VMware ESXi Server information

Field	Value
Server name (FQDN)	_____.<*domain*>.com
Management IP address	___.___.___.___
Subnet mask	___.___.___.___
Default gateway	___.___.___.___

Field	Value
Primary DNS server IP address	___.___.___.___
Secondary DNS server IP address	___.___.___.___
iSCSI primary interface	___.___.___.___
iSCSI secondary interface	___.___.___.___

If you are using iSCSI to connect to your storage, you need to assign two IP addresses for iSCSI connectivity.

5.2.2 Configuring the hardware

Power off the server hardware if it is powered on and continue with the following steps:

1. If needed, install additional network adapters.
2. If needed, install Fibre Channel host bus adapter (HBA) cards.
3. If needed, install iSCSI HICs or network adapters to use for iSCSI.
4. After the chassis is closed and the machine is reracked, plug in all associated cables except the SAN Fibre Channel cables.
5. Configure the BIOS and RAID.

Configuring the server BIOS

Follow these steps:

1. Check all firmware and update as necessary (BIOS, HBA, and internal RAID).
2. Ensure that your server BIOS is set up to accommodate virtualization technology. Refer to your server vendor-specific documentation for guidelines.

Configuring the server HBA

Note: In this example, we used a pair of QLogic QLA2340 cards.

Follow these steps to enable the HBA BIOS:

1. Press Ctrl-Q when prompted during the boot process to configure the QLogic BIOS.
2. Select the first QLogic card entry. Press Enter, as shown in Figure 5-6 on page 76.

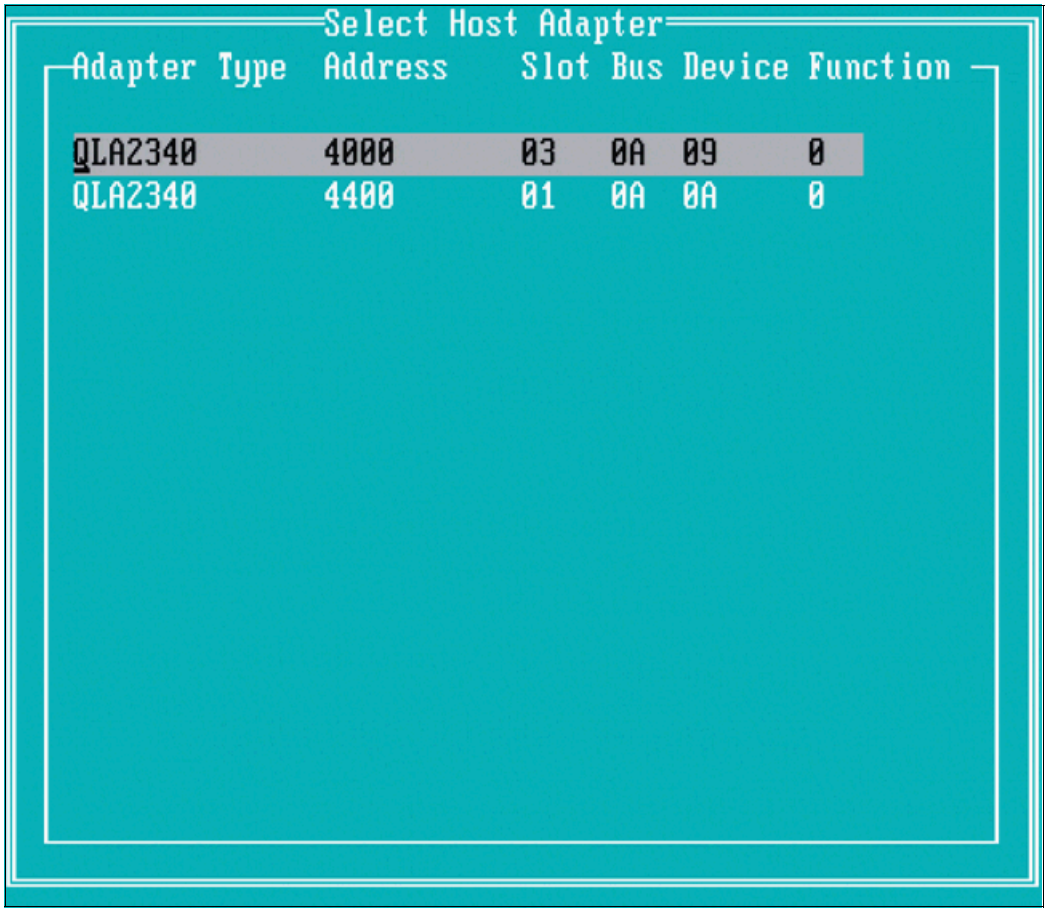

Figure 5-6 Selecting the HBA adapter

3. In Figure 5-7, select **Configuration Settings** for the selected HBA card. Press Enter.

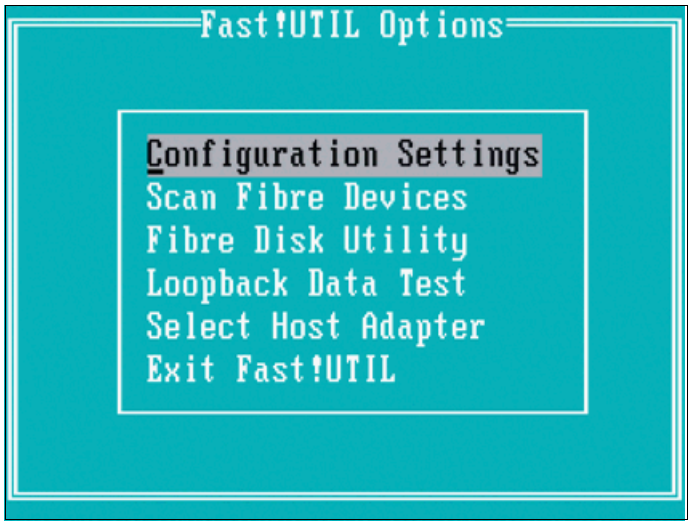

Figure 5-7 HBA configuration settings

4. If you are booting VMware ESXi from SAN, set Host Adapter BIOS to **Enabled**. Otherwise, leave it Disabled as shown in Figure 5-8 on page 77.

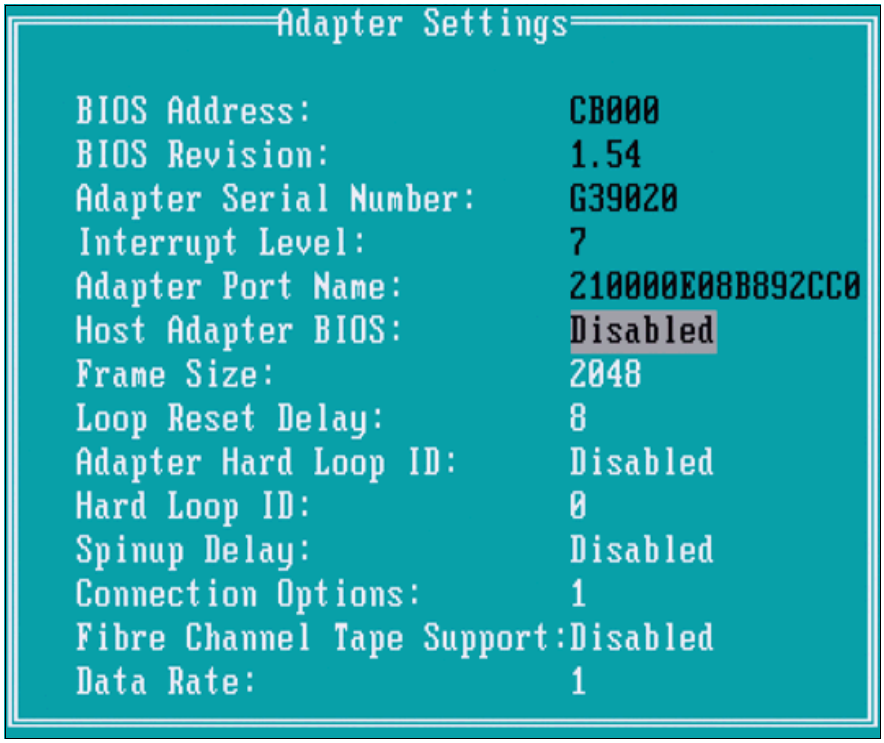

Figure 5-8 HBA settings

5. Press Esc to exit and select **Advanced Adapter Settings**. Check that these settings are correct as shown in Figure 5-9:
   - Enable LIP Reset: No
   - Enable LIP Full Login: Yes
   - Enable Target Reset: Yes

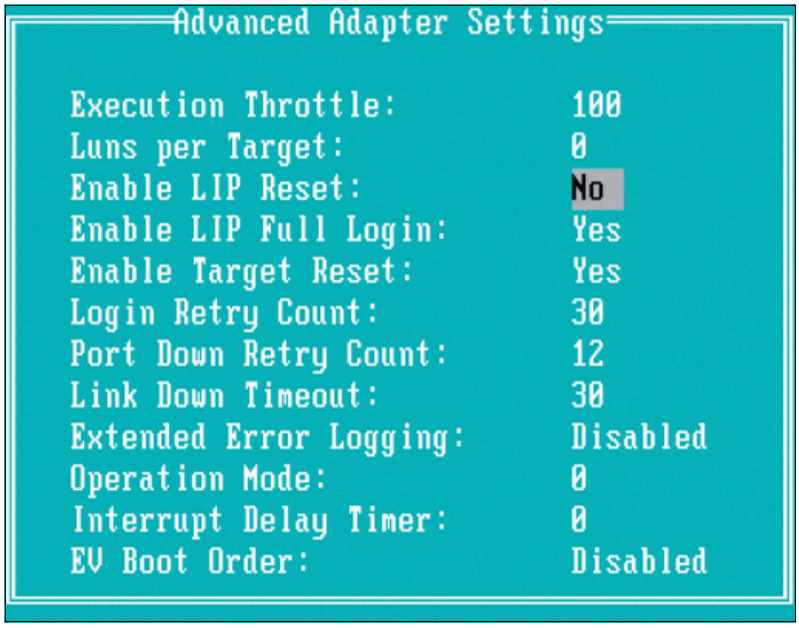

Figure 5-9 Advanced Adapter Settings

6. Press the Esc key to exit. Select **Save Changes** when prompted as shown in Figure 5-10. Press Enter.

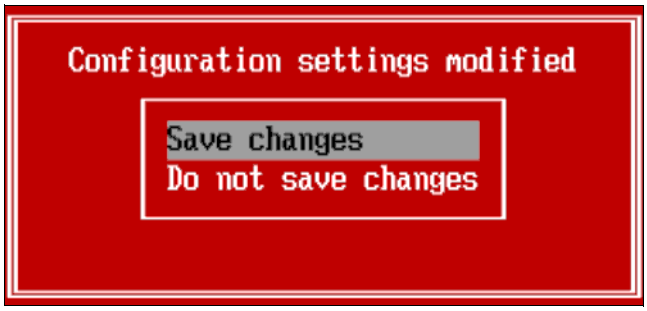

Figure 5-10 Saving changes

7. If additional HBAs are present, highlight **Select Host Adapter** on Figure 5-6 on page 76 and press Enter. Repeat step 2 on page 75 through step 6 for each additional HBA.
8. When the entire configuration is complete, select **Exit Fast!UTIL** as shown in Figure 5-11.

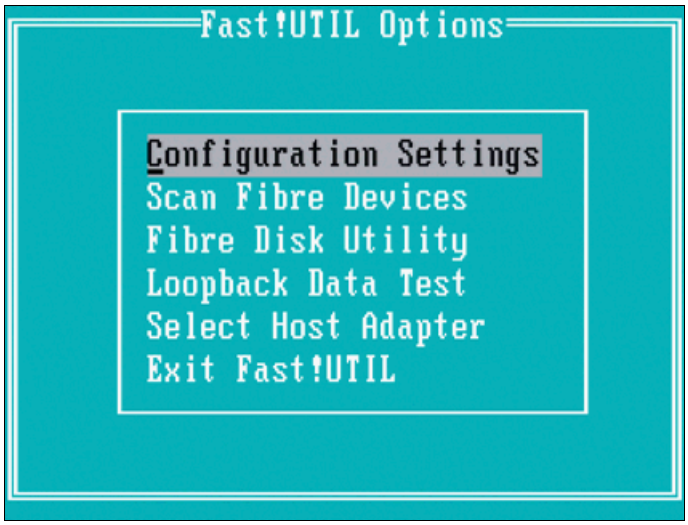

Figure 5-11 Fast!UTIL options

9. You are prompted with the panel that is shown in Figure 5-12. Select **Reboot System**.

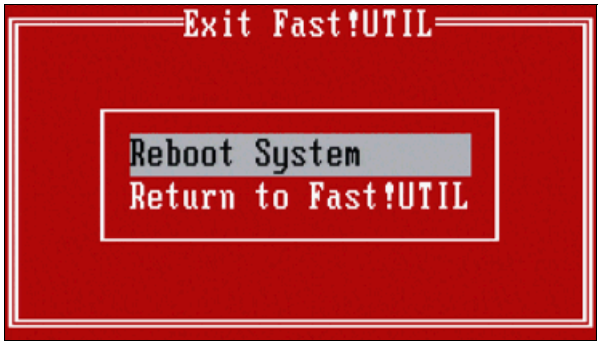

Figure 5-12 Fast!UTIL exit panel

Configuring the server RAID

Follow these steps to configure the server RAID:

1. At the prompt during reboot, press Ctrl-A to enter the Controller Configuration menu.
2. Configure the internal server RAID controller for a RAID1 or RAID 10 configuration, which allows you to preserve a working OS set in a drive failure. Performance on the local drives is not as critical as the performance of the actual virtual machines' datastores.

> **Note:** It is recommended to have in place at least one disk that is configured as "Hot Space" (HSP) to use for future disk failures and to avoid unnecessary delays on the RAID rebuild process.

5.2.3 Configuring the software on the VMware ESX Server host

We show the procedure to install the VMware ESXi Server software on the VMware ESXi Server host.

> **Note:** In this procedure, we use VMware ESXi 5.0.0-469512. Some of the steps might reflect that version number. Substitute the version number that you are installing.

Follow these steps:

1. On the initial panel, select **ESXi-5.0.0-469512-standard Installer**, as shown in Figure 5-13.

Figure 5-13 VMware ESXi Server Boot Menu panel

2. After the installer starts to boot, you see panels that list all the modules that are loading. When you get to the panel that is shown on Figure 5-14 on page 80, press Enter to continue.

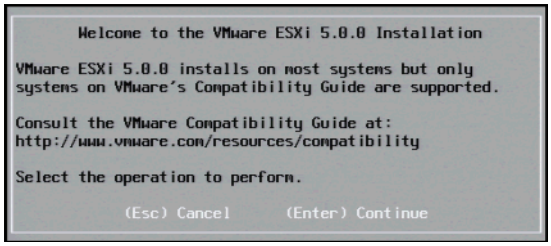

Figure 5-14 VMware ESXi Server Welcome panel

3. Review the End User License Agreement panel, which is shown on Figure 5-15, and accept the license terms by pressing F11.

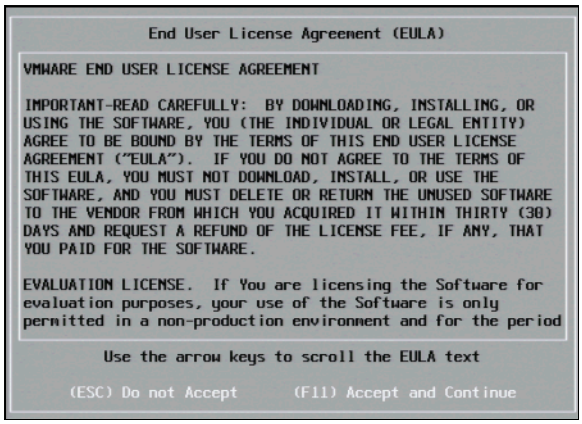

Figure 5-15 VMware ESXi Server End User License Agreement (EULA)

4. The installer continues to scan for available devices. On the panel that is shown on Figure 5-16, select the drive on which you want to install VMware ESXi Server. Press Enter to continue.

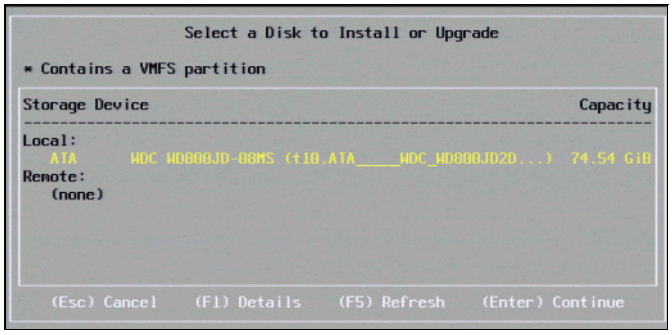

Figure 5-16 VMware ESXi Server Select a Disk to Install or Upgrade panel

5. On the keyboard layout selection panel that is shown on Figure 5-17 on page 81, select your keyboard layout and press Enter to continue.

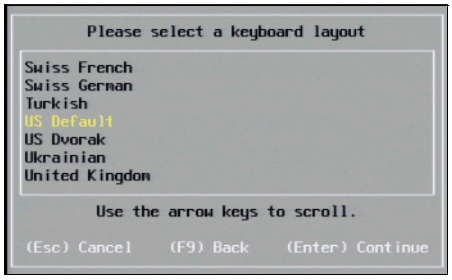

Figure 5-17 VMware ESXi Server keyboard layout selection

6. On the next panel, you need to set up the password for the root user. You can leave this blank but it is not recommended. Type the password twice to confirm it as shown in Figure 5-18 and press Enter to continue.

Figure 5-18 VMware ESXi Server root user password setup

Note: A blank password can be used at the installation and configured later by using the system customization interface.

7. You need to confirm the installation options as shown in Figure 5-19. If the settings are correct, proceed with installation by pressing F11.

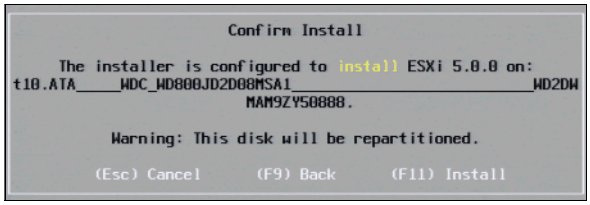

Figure 5-19 VMware ESXi Server Confirm Install panel

Important: In a server rebuild, be careful with the disk selection. It is recommended that you disable HBA cards (from the BIOS) before you start the installation.

8. Wait for the installation to complete. The panel that is shown in Figure 5-20 on page 82 appears. Press Enter to reboot the system and start the VMware ESXi Server.

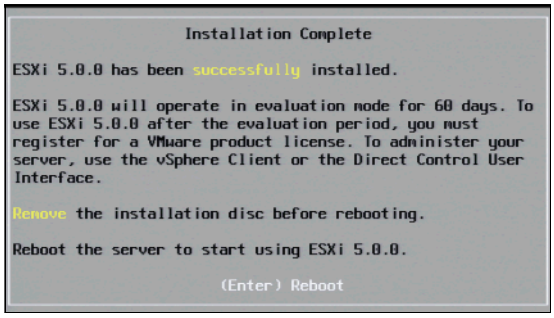

Figure 5-20 VMware ESXi Server Installation Complete panel

9. After the server starts, the VMware ESXi Server is shown (Figure 5-21). You must set up a few parameters to manage the VMware ESXi Server. Press F2 to set up the necessary parameters.

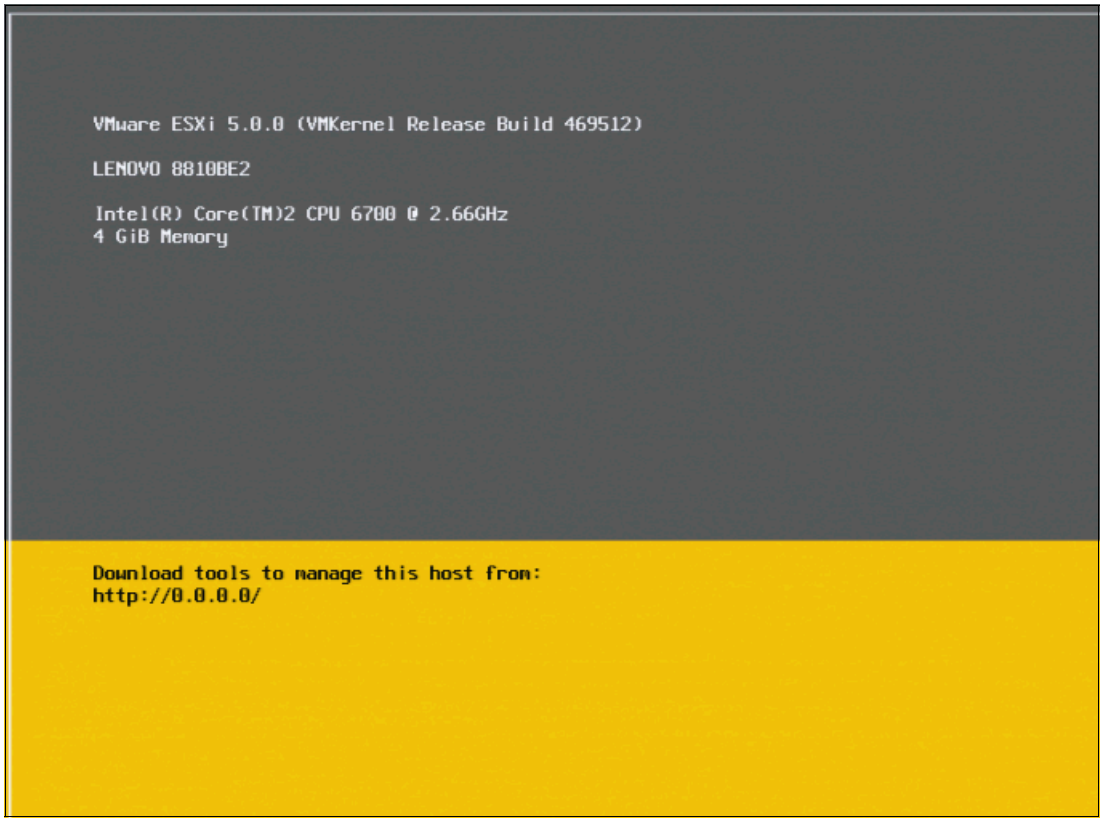

Figure 5-21 VMware ESXi Server initial panel

10. Select **Configure Management Network** as shown in Figure 5-22 on page 83 and press Enter.

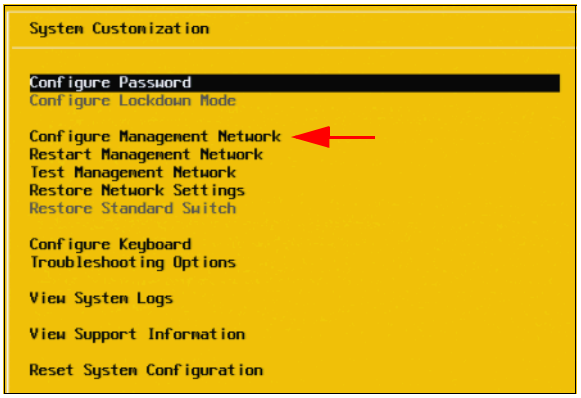

Figure 5-22 VMware ESXi Server System Customization panel

11. Select **IP Configuration** as shown in Figure 5-23 and press Enter.

> **Note:** You can also set additional parameters as shown in Figure 5-23 on page 83, such as Domain Name System (DNS) settings, IPv6, and VLANs, depending on your network configuration.

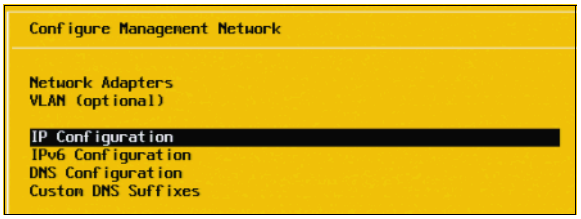

Figure 5-23 VMware ESXi Server Configure Management Network panel

12. Type your network configuration details as shown in Figure 5-24 or select **Use dynamic IP address and network configuration**, depending on your network setup. Press Enter to continue.

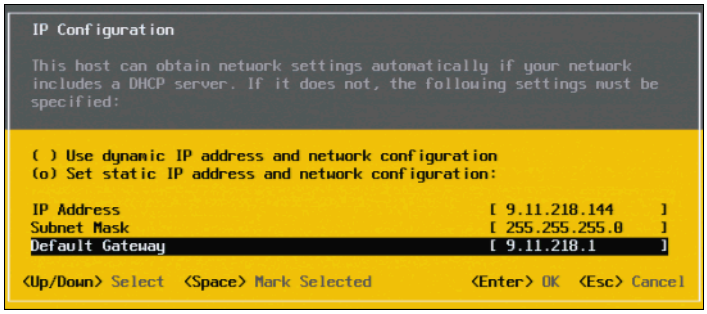

Figure 5-24 VMware ESXi Server IP Configuration panel

13. After setting up all the network parameters, you return to the panel shown on Figure 5-23. Press Esc to exit. The panel in Figure 5-25 on page 84 is shown. Press Y to confirm the network changes that you made. Restart the management network to apply the changes.

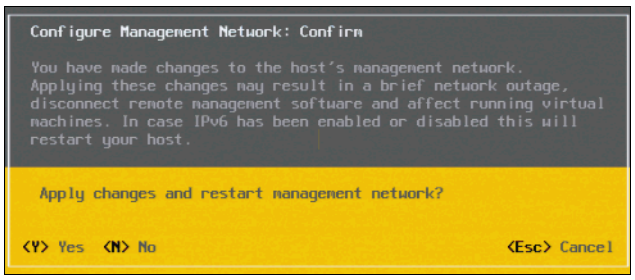

Figure 5-25 VMware ESXi Server management network change confirmation panel

5.2.4 Connecting to the VMware ESXi Server

You are ready to connect to the VMware ESXi Server. Perform the following procedure on the management workstation that you will use to administer the VMware ESXi Server:

1. Using your web browser, connect to the host name or IP address of the newly created VMware ESXi Server. We used Firefox as the Internet browser application. The error message in Figure 5-26 is normal.

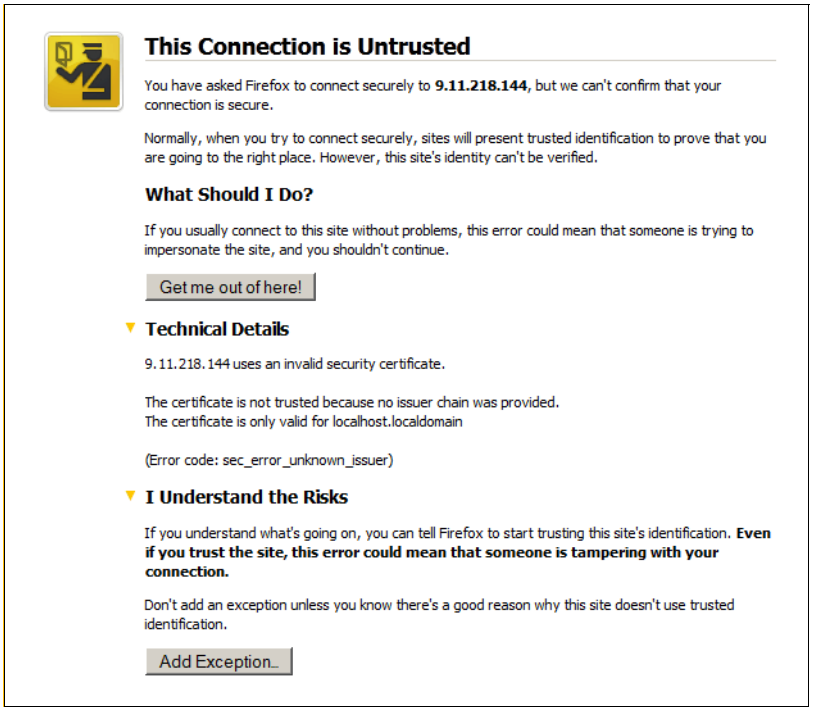

Figure 5-26 Connecting to VMware ESXi Server

2. Click **Add Exception**. Figure 5-27 on page 85 is presented to acquire the Secure Sockets Layer (SSL) Certificate. Click **Get Certificate** and then click **Confirm Security Exception**.

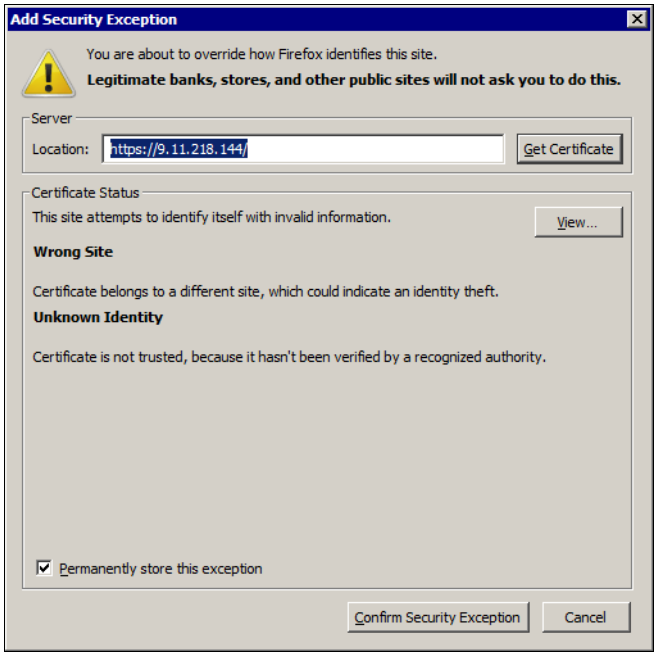

Figure 5-27 Add Security Exception

3. You are presented with the panel that is shown in Figure 5-28 on page 86. Click **Download vSphere Client** to download the setup package on the system that you will use as the initial administrative workstation.

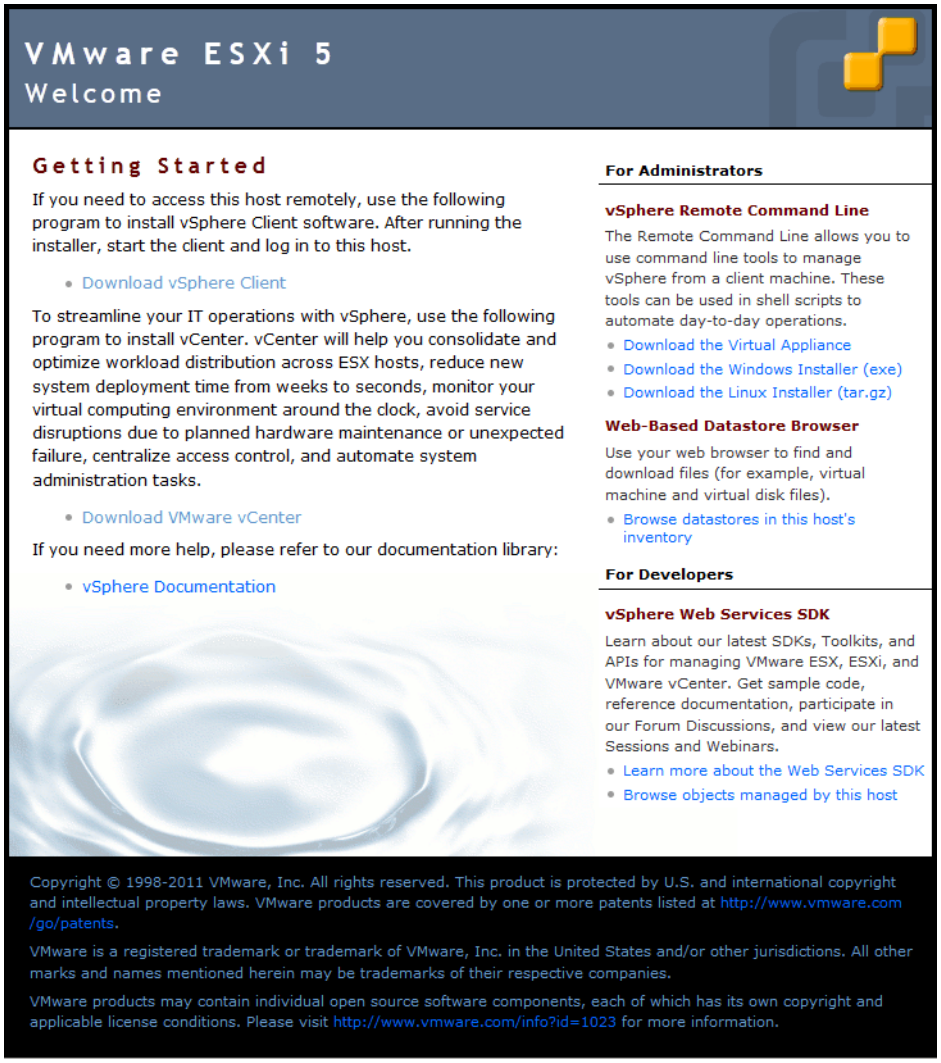

Figure 5-28 VMware ESXi Server Welcome web page

4. Run the downloaded setup package on your administrative workstation.

> **Note:** The setup package has several hardware and software requirements and recommendations as listed at this website:
>
> `http://pubs.vmware.com/vsphere-50/index.jsp?topic=/com.vmware.vsphere.install.doc_50/GUID-7C9A1E23-7FCD-4295-9CB1-C932F2423C63.html`

5. Choose the setup language as shown in Figure 5-29 and click **OK**.

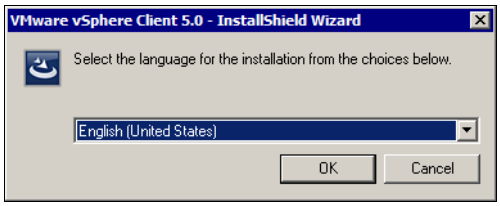

Figure 5-29 VMware vSphere Client setup language selection panel

6. On the welcome panel that is shown in Figure 5-30, click **Next**.

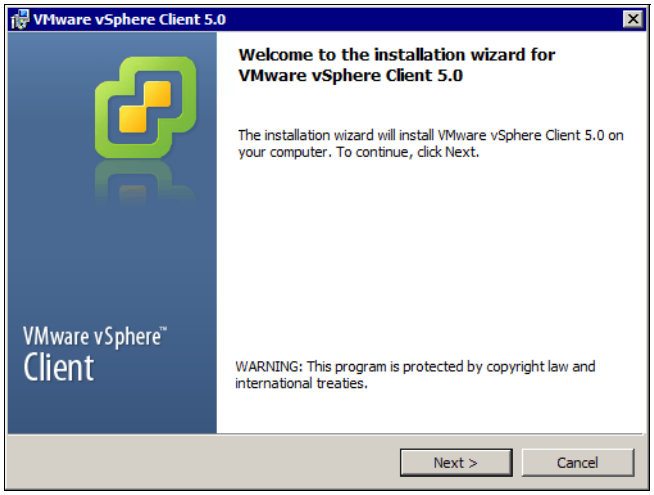

Figure 5-30 VMware vSphere Client setup welcome panel

7. Review the patent agreement that is shown in Figure 5-31 and click **Next**.

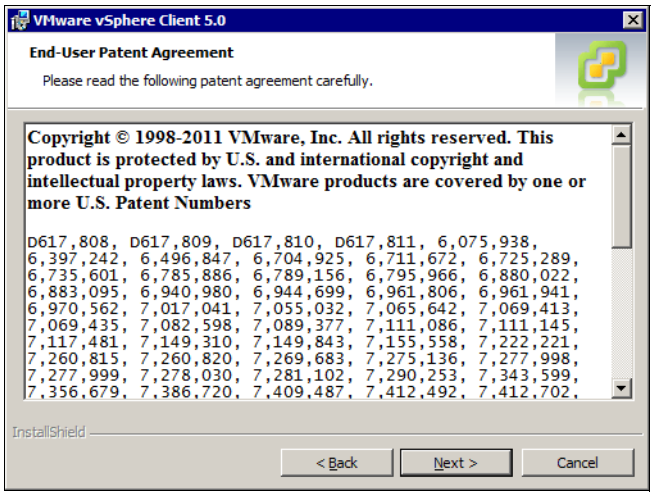

Figure 5-31 VMware vSphere Client setup patent agreement panel

8. Review and agree to the license agreement that is shown in Figure 5-32 on page 88 and click **Next**.

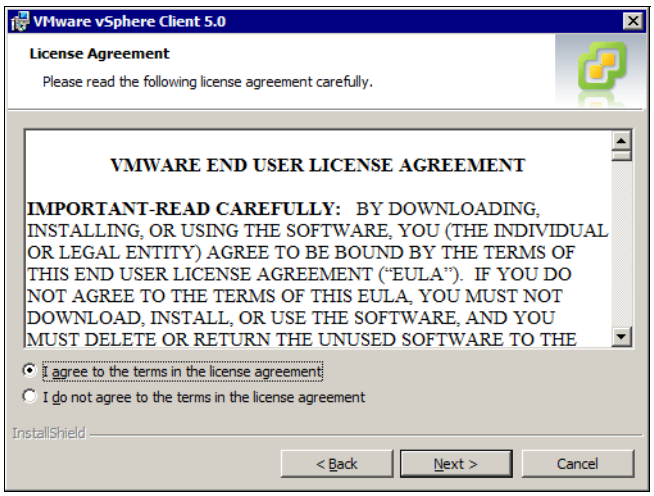

Figure 5-32 VMware vSphere Client setup license agreement panel

9. Type your name and organization in the panel that is shown in Figure 5-33 and click **Next**.

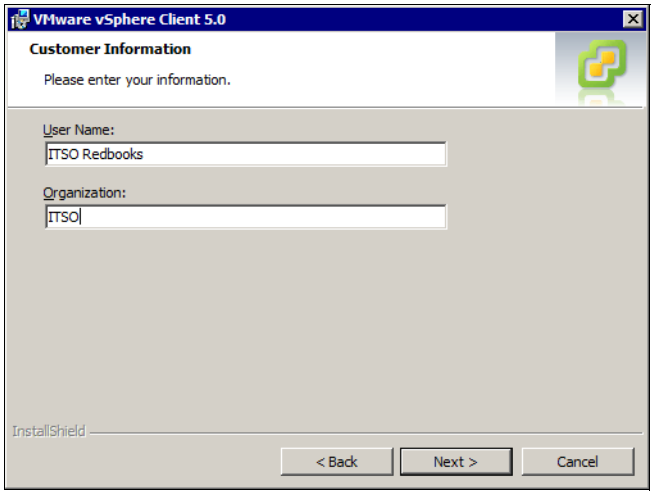

Figure 5-33 VMware vSphere Client setup customer information panel

10. Select the folder in which to install the VMware vSphere Client as shown in Figure 5-34 on page 89 and click **Next**.

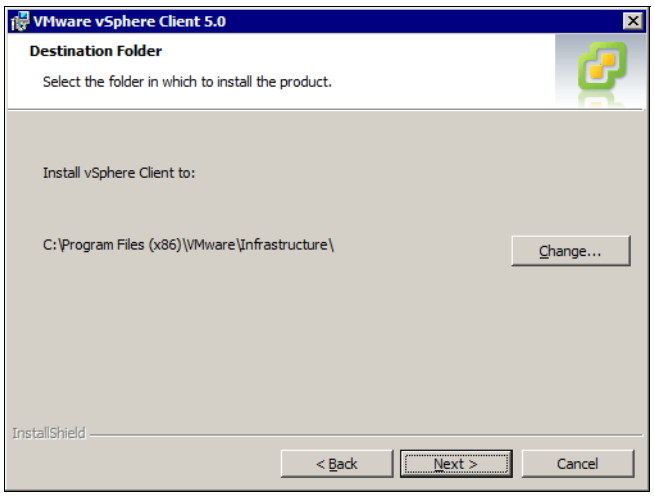

Figure 5-34 VMware vSphere Client setup destination folder panel

11. Install VMware vSphere Client 5.0 by clicking **Install** as shown in Figure 5-35.

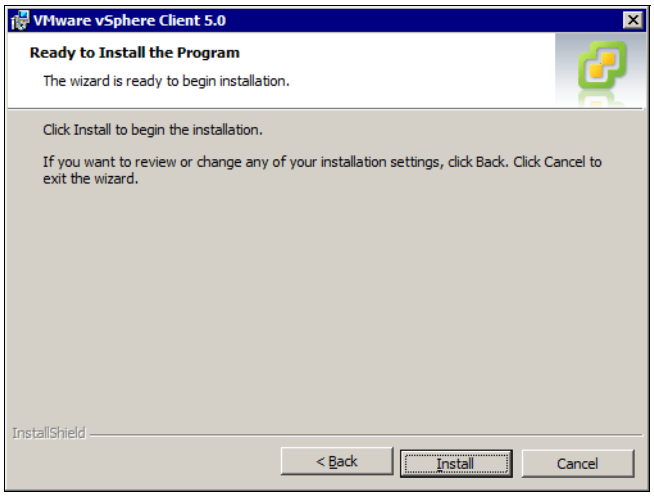

Figure 5-35 VMware vSphere Client setup installation panel

12. After the installation is complete, on the final panel, which is shown in Figure 5-36 on page 90, click **Finish**.

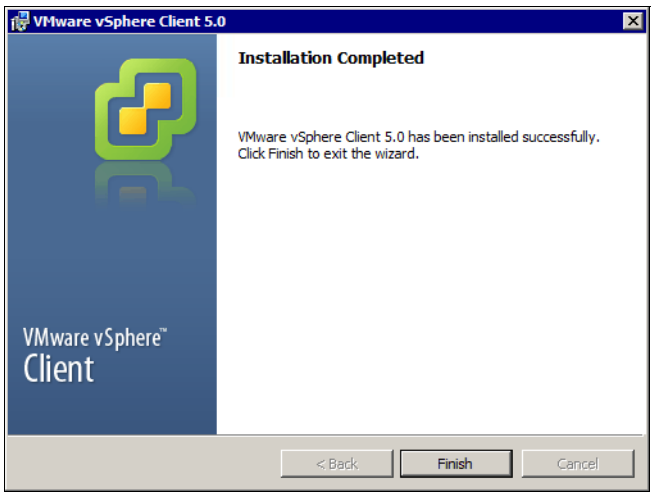

Figure 5-36 VMware vSphere Client 5.0 setup final panel

13. Now, you are ready to connect to the VMware ESXi Server by using the VMware vSphere Client. Run the newly installed client, type the VMware ESXi Server IP address or host name, and log in as `root`, as shown in Figure 5-37.

Figure 5-37 VMware vSphere Client logon

5.2.5 Creating additional virtual switches for guests' connectivity

Follow these steps:

1. Using the VMware vSphere Client, connect to the VMware ESXi Server (log in as `root`).
2. Click the **Configuration** tab and select **Networking** as shown in Figure 5-38 on page 91.

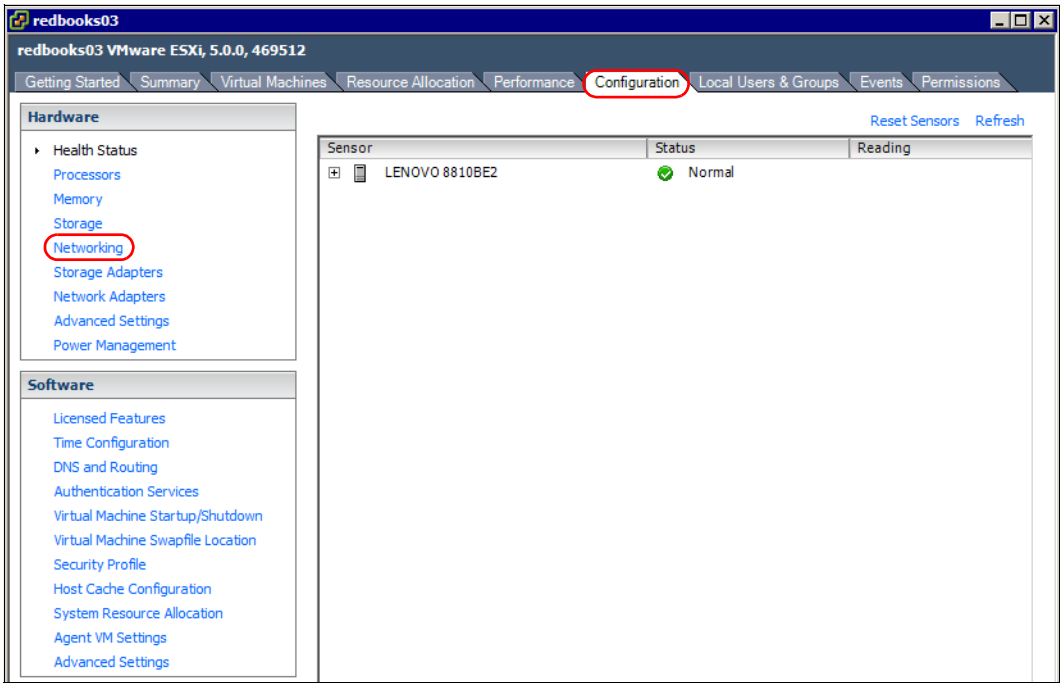

Figure 5-38 vSphere Configuration panel

3. In the upper-right corner, click **Add Networking**, as shown in Figure 5-39.

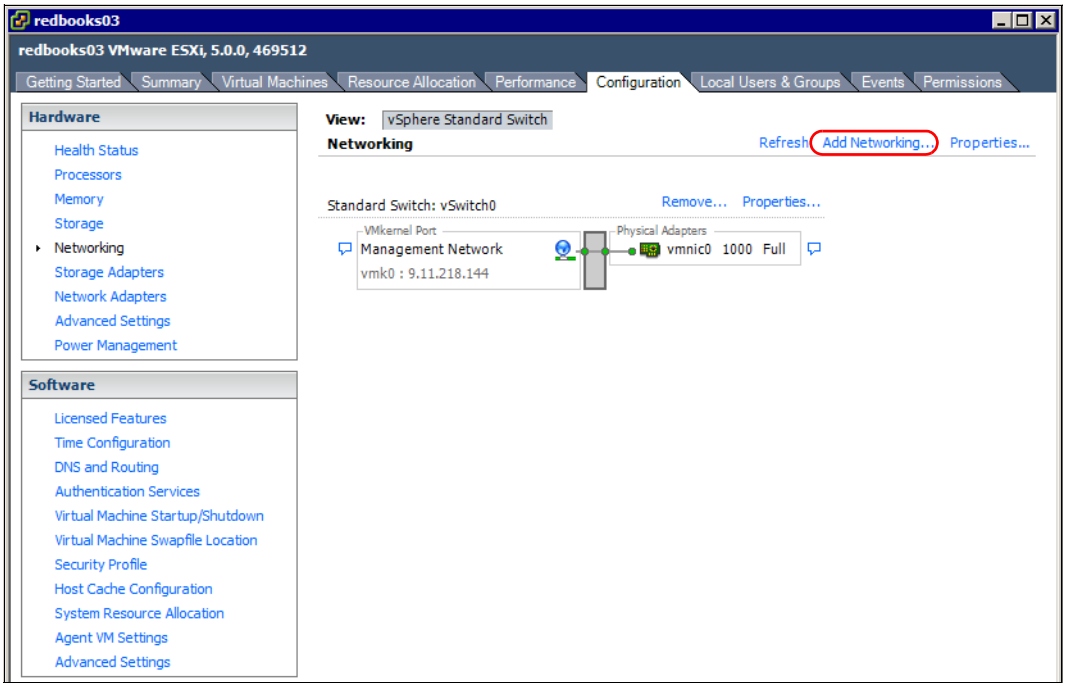

Figure 5-39 Add Networking

4. Select **Virtual Machine** as the Connection Type, and click **Next**, as shown in Figure 5-40 on page 92.

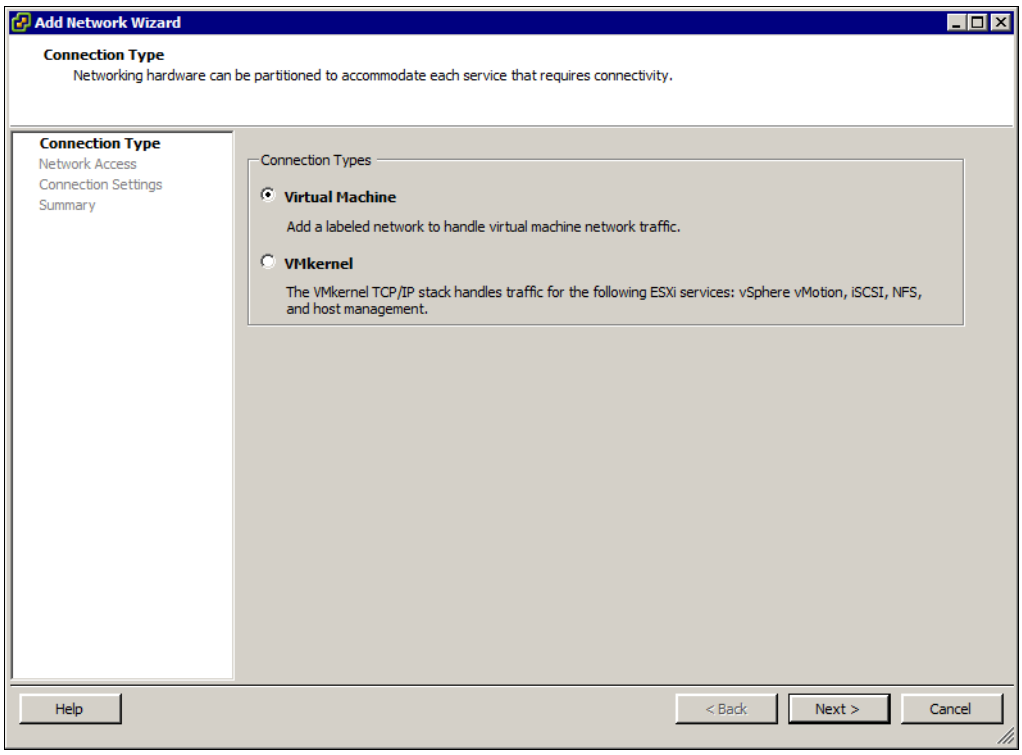

Figure 5-40 Connection Type

5. Select one of the remaining LAN adapters (in this case, vmnic1), as shown in Figure 5-41. Click **Next**.

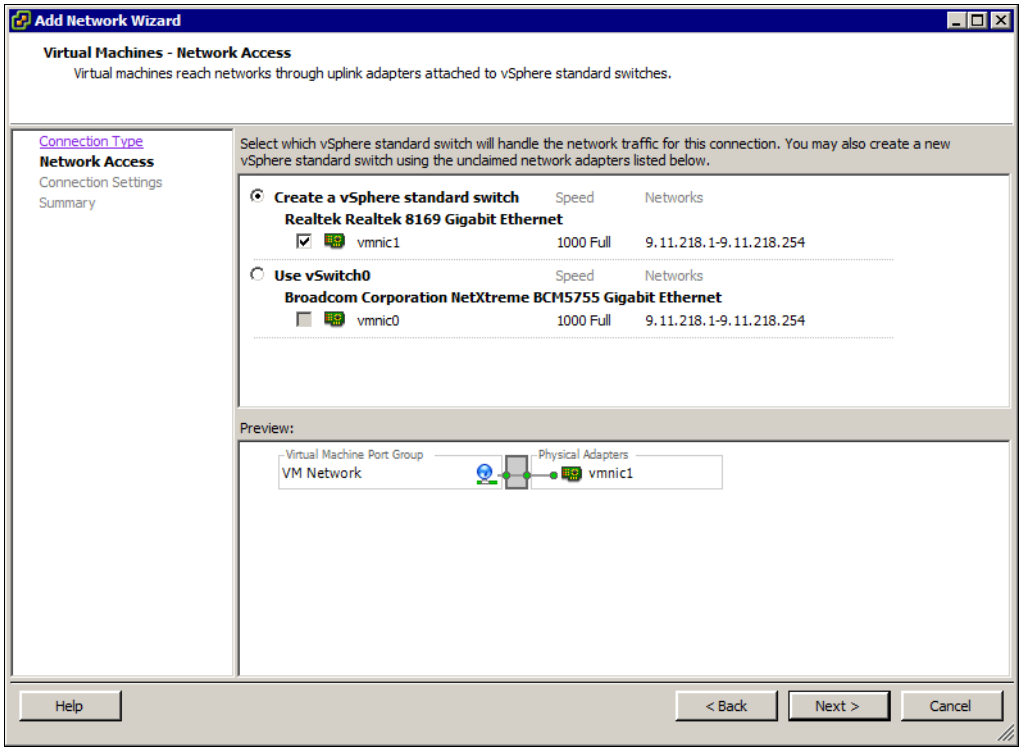

Figure 5-41 Selecting Network Access

6. On the Connection Settings panel, type your network name (Network Label), as shown in Figure 5-42, and click **Next**.

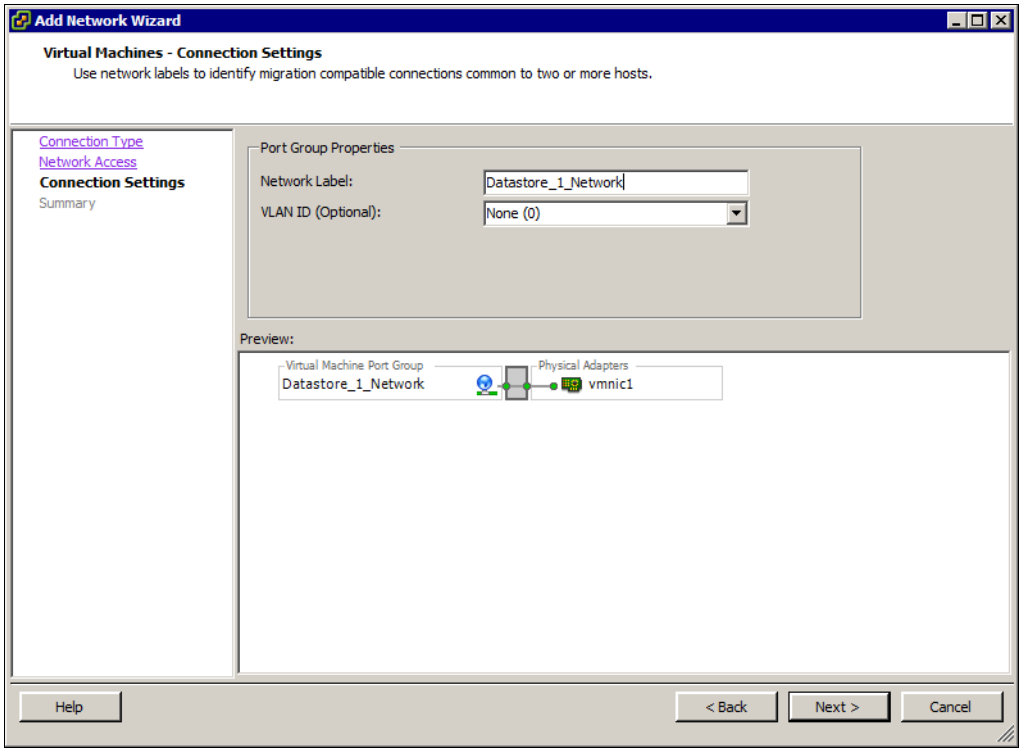

Figure 5-42 Virtual Network Label

7. Check your settings on the summary panel shown in Figure 5-43 on page 94 and click **Finish**.

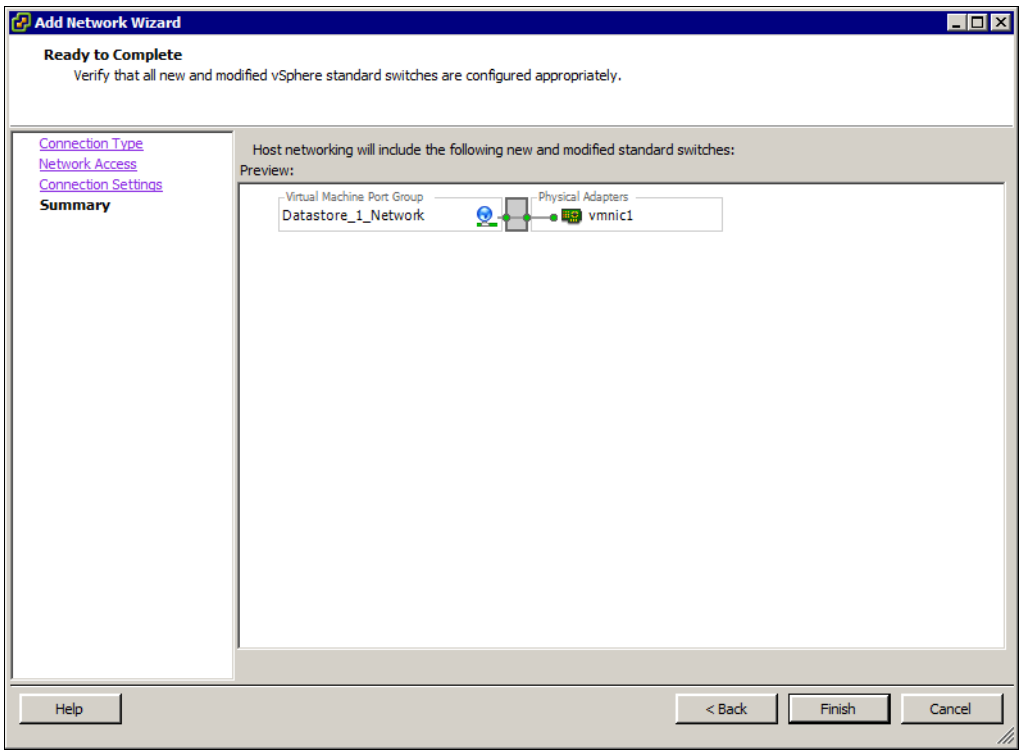

Figure 5-43 Network summary panel

5.2.6 Connecting to SAN storage by using iSCSI

The following procedure explains how to connect to your storage by using iSCSI. We are using an iSCSI software initiator to connect to the storage. A software iSCSI adapter is a part of VMware code. It uses standard Ethernet adapters to connect to your iSCSI storage.

Another option is to purchase hardware iSCSI initiators. You can obtain more information about types of hardware iSCSI initiators in the VMware documentation that is listed in “Related publications” on page 255.

There are several steps in configuring an iSCSI software initiator:

1. Activate the software iSCSI adapter.
2. Configure networking for iSCSI.
3. Configure iSCSI discovery addresses.

Note: VMware ESXi has several restrictions that apply to connecting iSCSI SAN devices:

- ESXi does not support iSCSI-connected tape devices.
- You cannot use virtual machine multipathing software to perform I/O load balancing to a single physical LUN.
- ESXi does not support multipathing when you combine independent hardware adapters with either software or dependent hardware adapters.

Activating the software iSCSI adapter

Follow this procedure to activate the software iSCSI adapter, which is also known as an *initiator*:

1. Use the VMware vSphere Client to connect to the VMware ESXi Server (log in as `root`).
2. Click the **Configuration** tab and select **Storage Adapters** as shown in Figure 5-44.

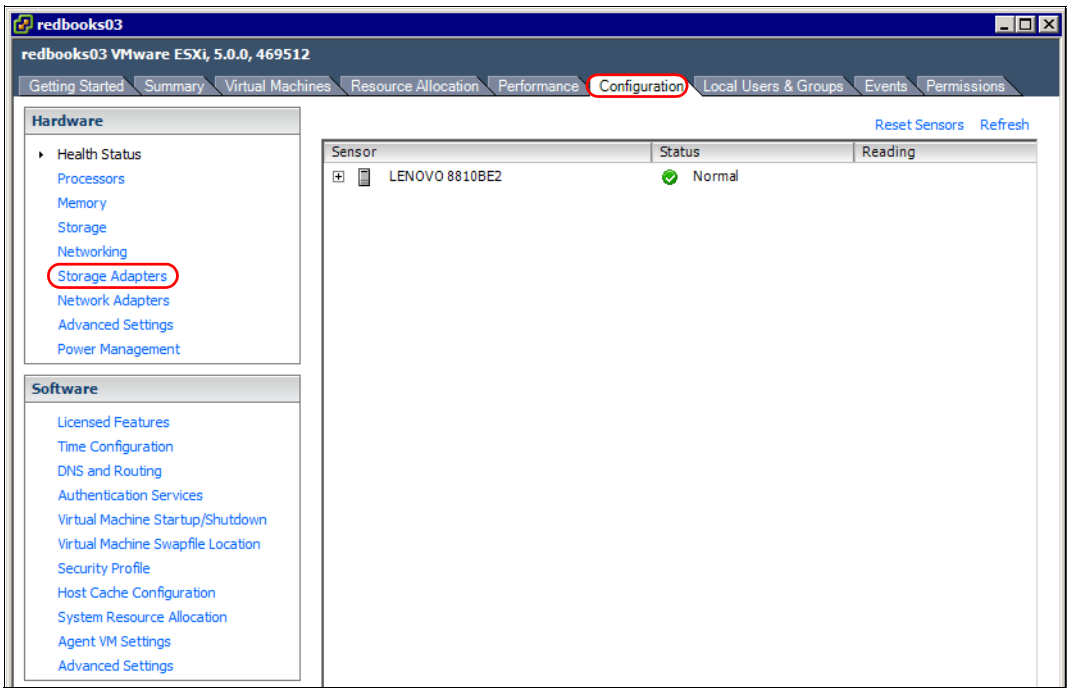

Figure 5-44 vSphere Configuration panel

3. Click **Add** as shown in Figure 5-45 on page 96.

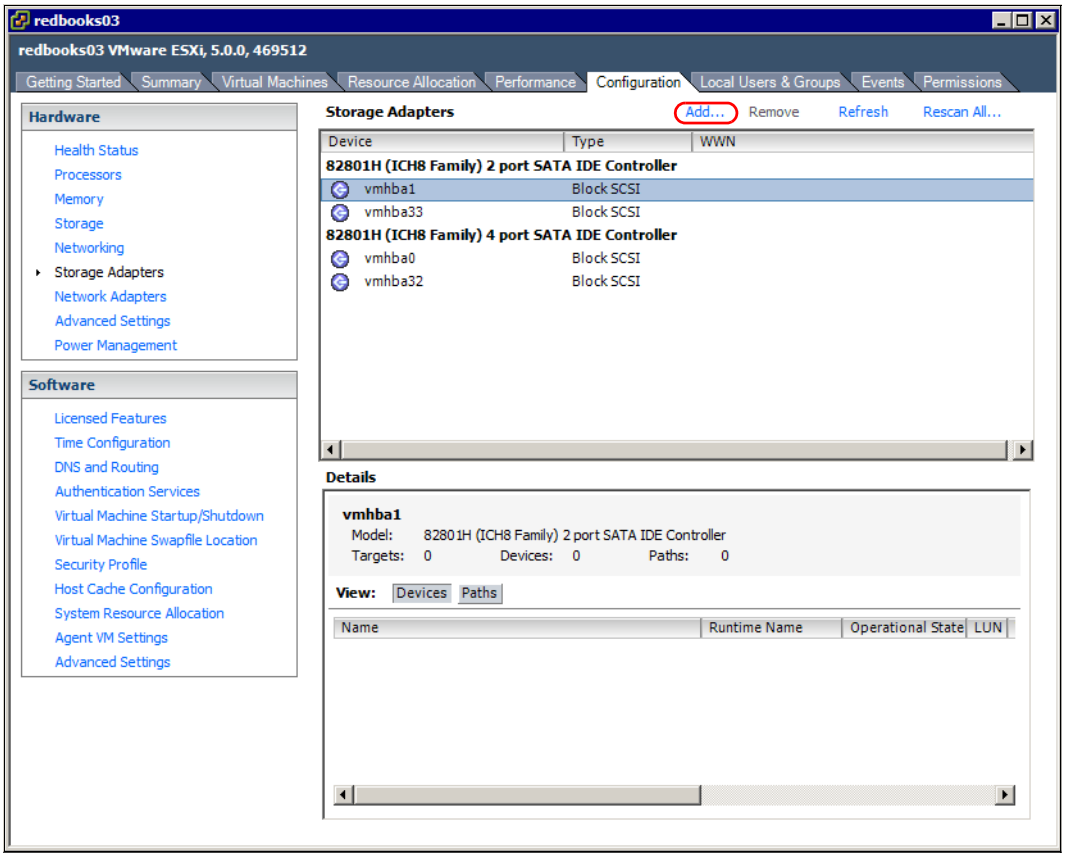

Figure 5-45 vSphere Storage Adapters panel

4. Select **Add Software iSCSI Adapter** (if not selected) and click **OK** as shown in Figure 5-46.

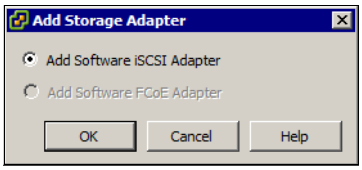

Figure 5-46 Add Software Adapter

5. Click **OK** to confirm adding a software iSCSI adapter on the panel that is shown in Figure 5-47.

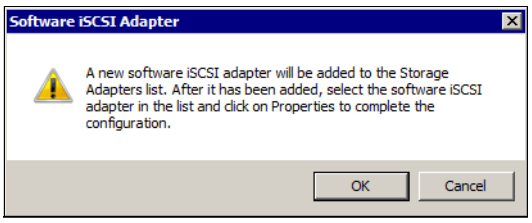

Figure 5-47 Software iSCSI adapter confirmation panel

6. Now, the software iSCSI adapter is activated as shown in Figure 5-48 on page 97.

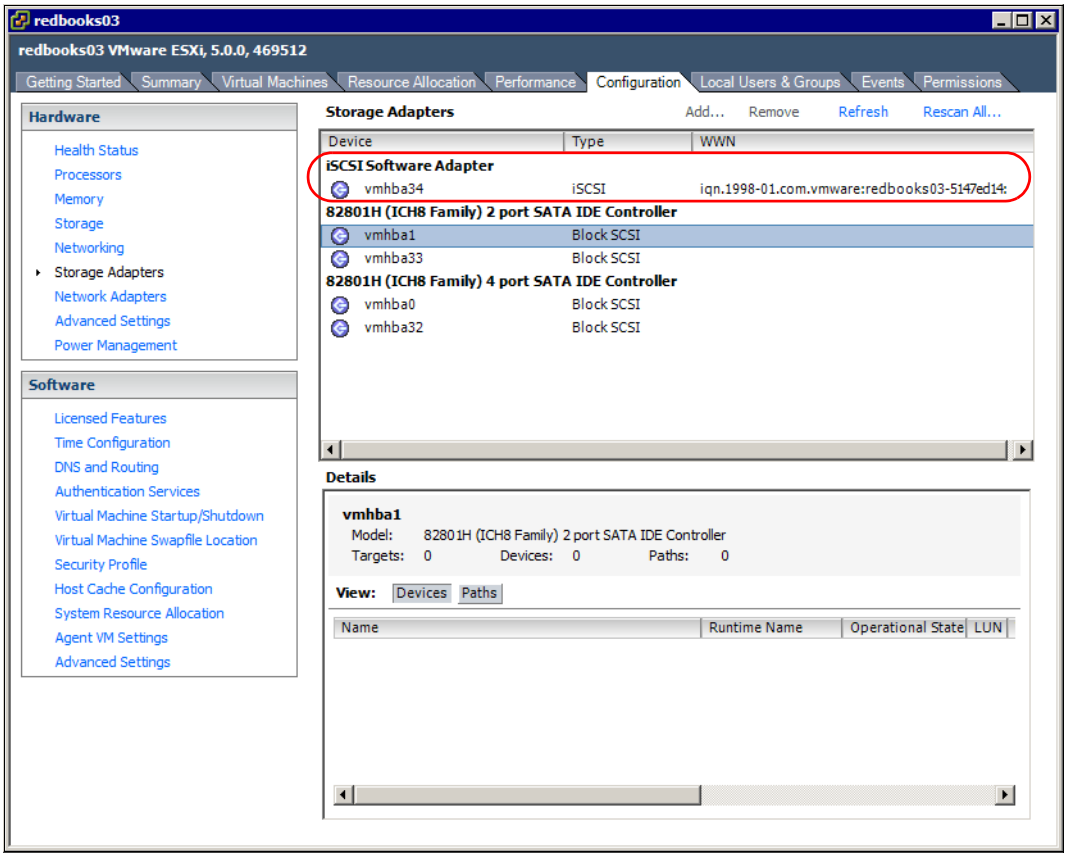

Figure 5-48 iSCSI Software Adapter installed

Configuring networking for iSCSI

We use two network adapters for the iSCSI connection to the storage subsystem. They need to be added to a separate vSphere switch, and they both need to be assigned a separate IP address. This procedure shows the necessary steps:

1. On the Configuration tab, click **Networking** in the Hardware Section and then select **Add Networking** as shown in Figure 5-49 on page 98.

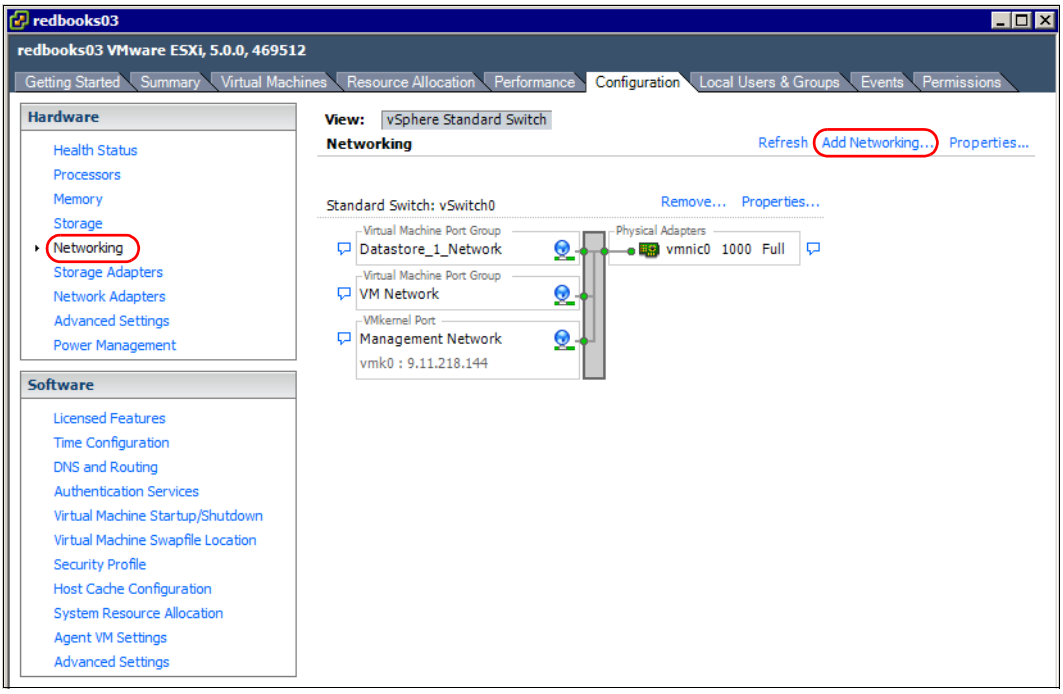

Figure 5-49 vSphere Networking panel

2. In the initial panel of the Add Network Wizard, select **VMkernel** for the connection type as shown in Figure 5-50 and click **Next**.

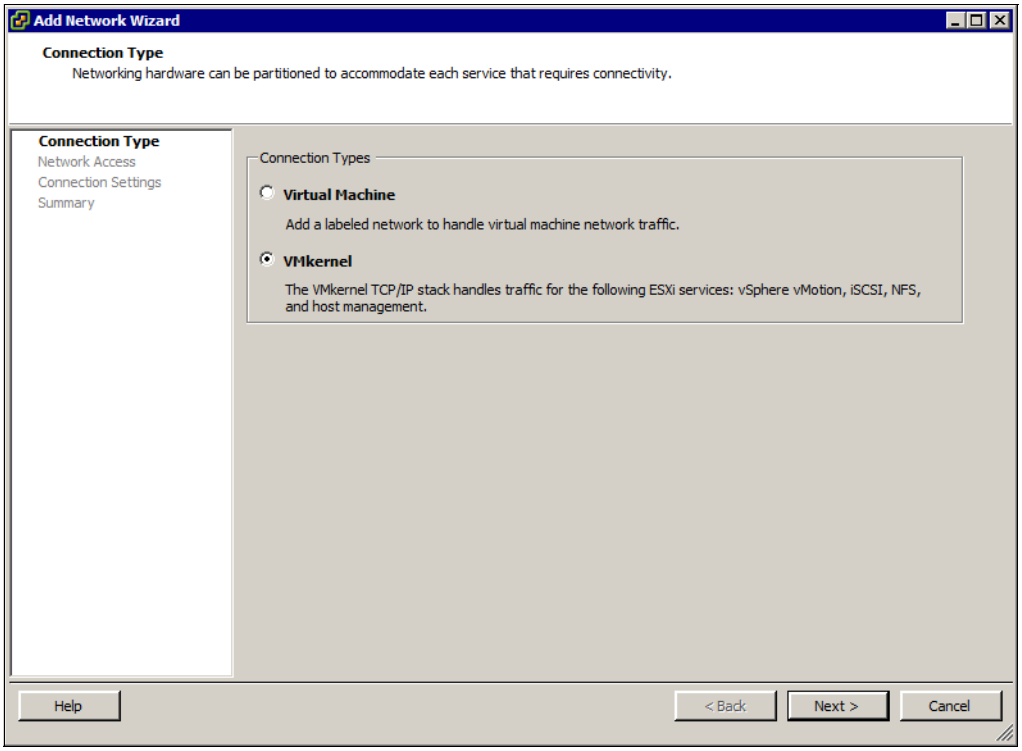

Figure 5-50 Network connection type

3. Select **Create a vSphere standard switch**. Select only one network adapter, which you have planned for iSCSI, to add to the vSphere standard switch as shown in Figure 5-51 and click **Next**. (We add the second network adapter later.)

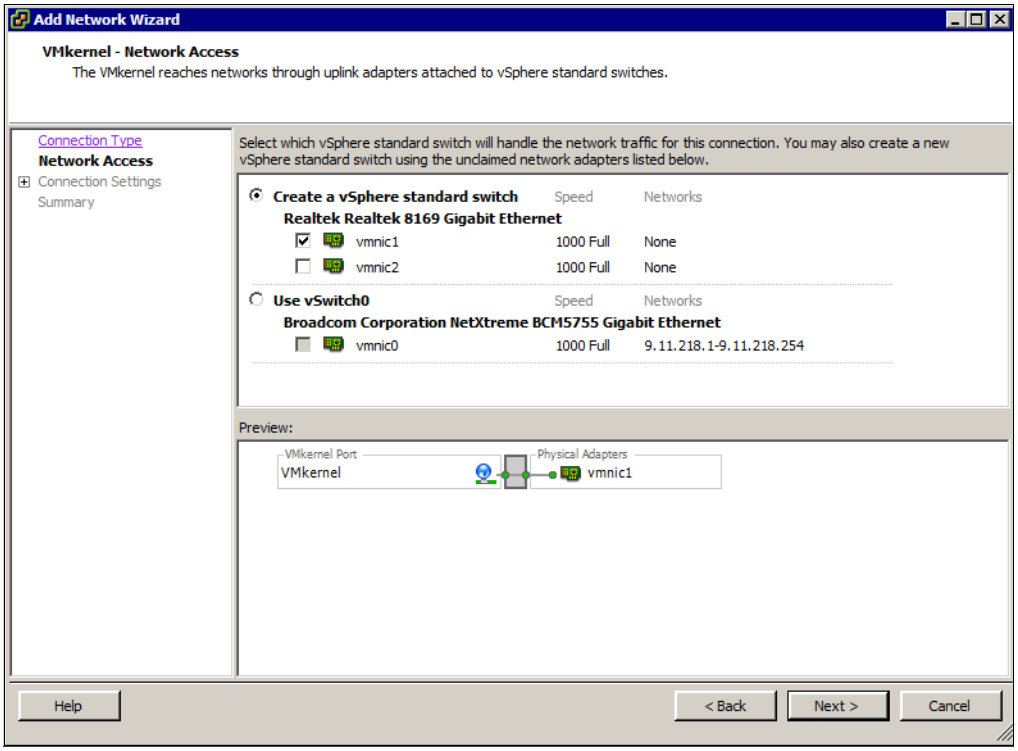

Figure 5-51 VMkernel Network Access

4. Label the VMkernel adapter as shown in Figure 5-52 on page 100 and click **Next**.

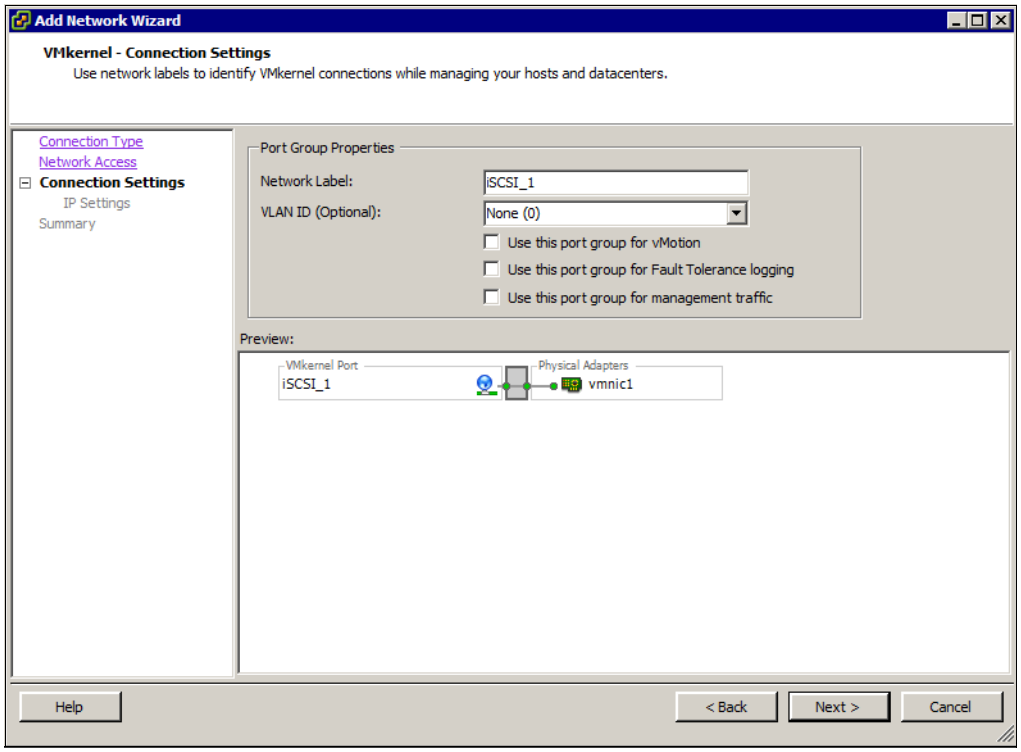

Figure 5-52 VMkernel label

5. Assign the IP address and the subnet mask, which is defined for your iSCSI network, as shown in Figure 5-53 and click **Next**.

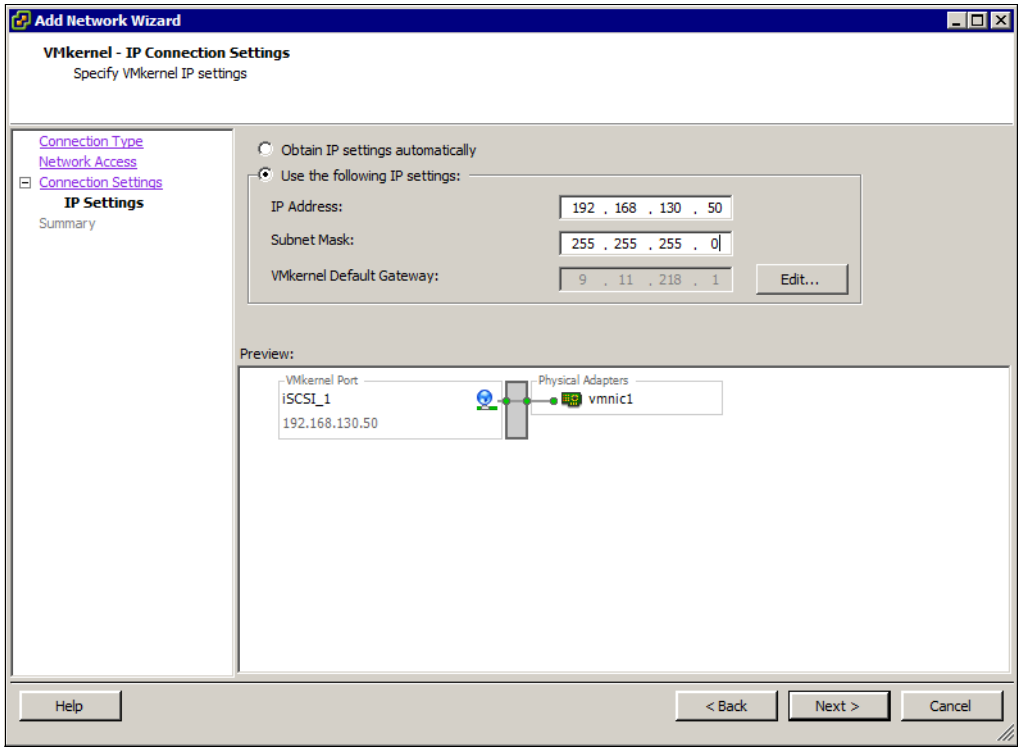

Figure 5-53 VMkernel IP configuration

6. On the final panel shown in Figure 5-54, check that all your settings are correct and click **Finish**.

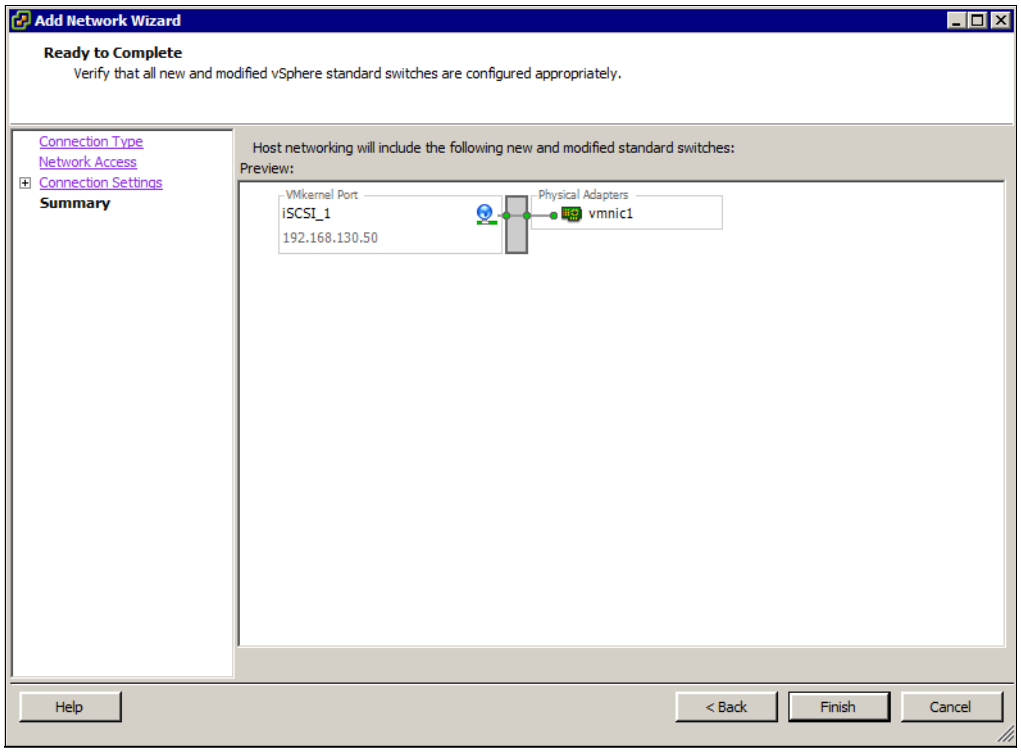

Figure 5-54 VMkernel overview panel

7. You can see the newly created vSphere standard switch and the VMkernel interface that was added to it. Click **Properties** as shown in Figure 5-55.

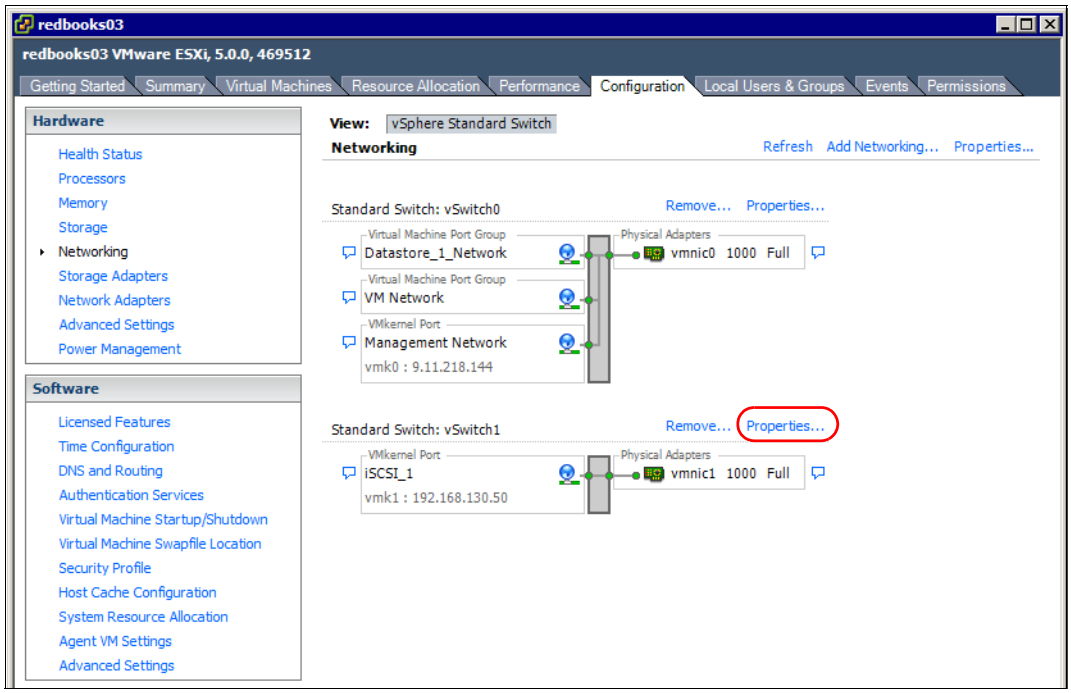

Figure 5-55 New vSphere standard switch added

8. Select the **Network Adapters** tab and click **Add** to add the second network adapter to the vSphere standard switch as shown in Figure 5-56.

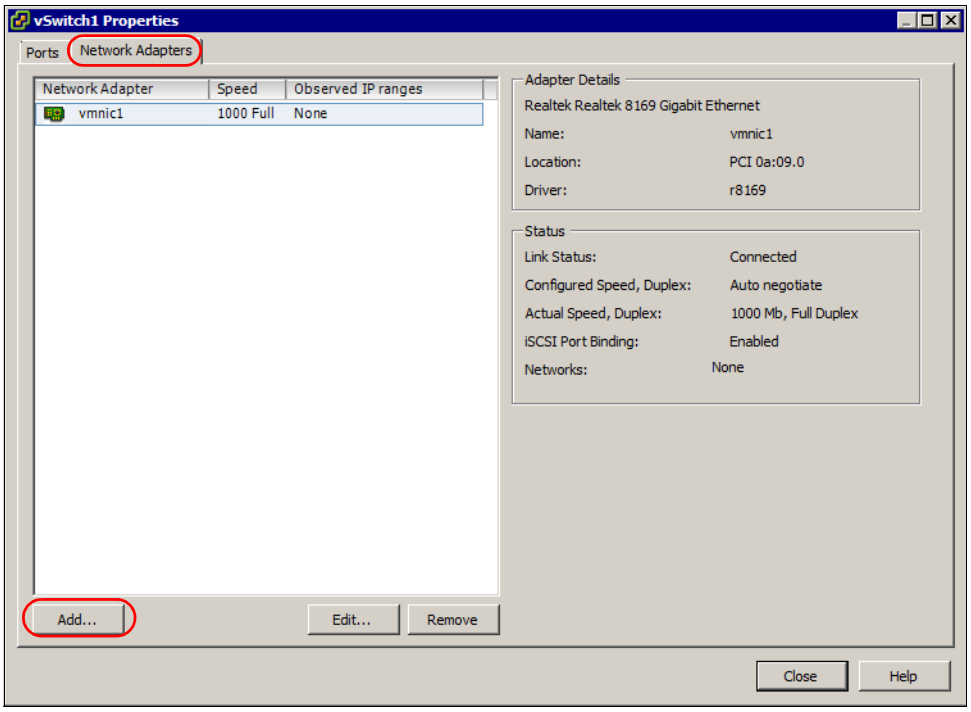

Figure 5-56 Adding network adapters to a vSphere standard switch

9. Select the other network adapter that is planned for iSCSI and click **Next** as shown in Figure 5-57.

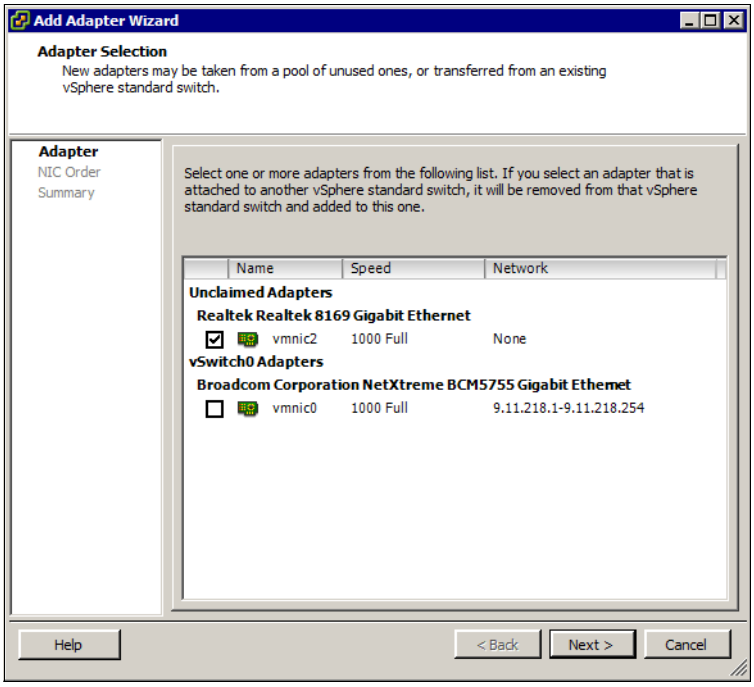

Figure 5-57 Adapter Selection panel

10. Leave the default policy failover order because the network adapters will each be assigned to a separate VMkernel. Click **Next** as shown in Figure 5-58.

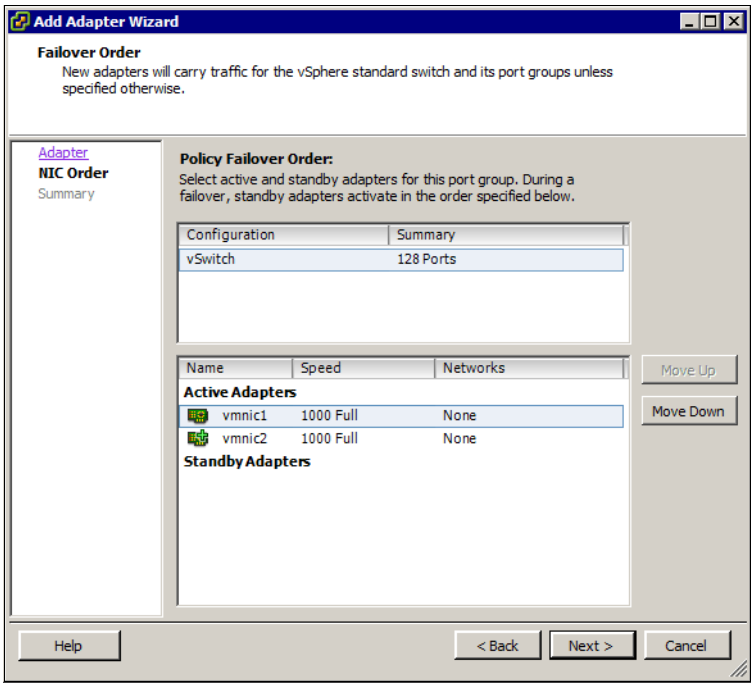

Figure 5-58 Failover Order panel

11. On the Summary panel that is shown in Figure 5-59, click **Finish**.

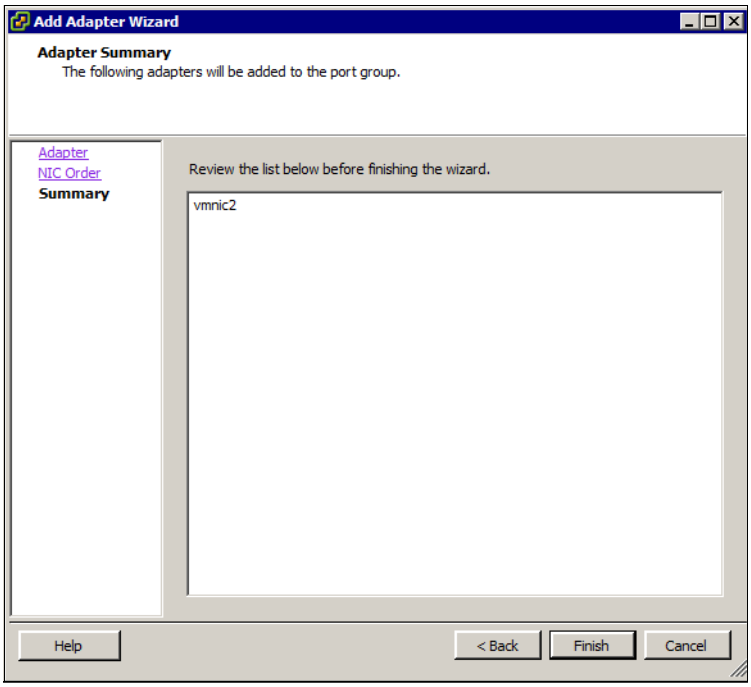

Figure 5-59 Network Adapter Summary panel

12. Now, two network adapters are assigned to a vSphere standard switch as shown in Figure 5-60 on page 104.

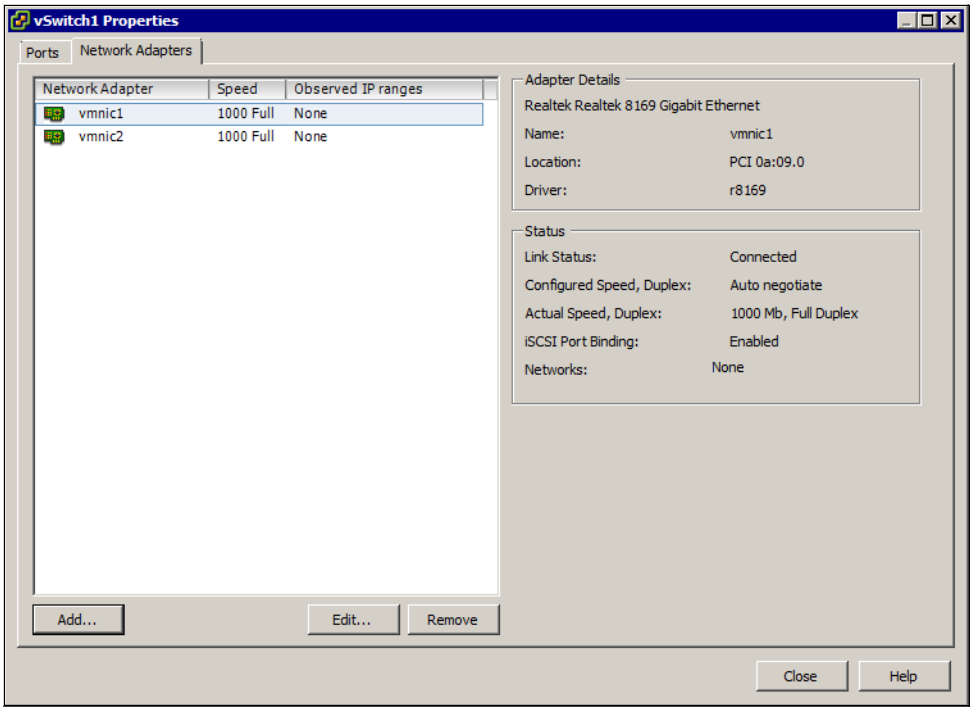

Figure 5-60 vSphere standard switch network adapters

13. Select the **Ports** tab and click **Add** as shown in Figure 5-61.

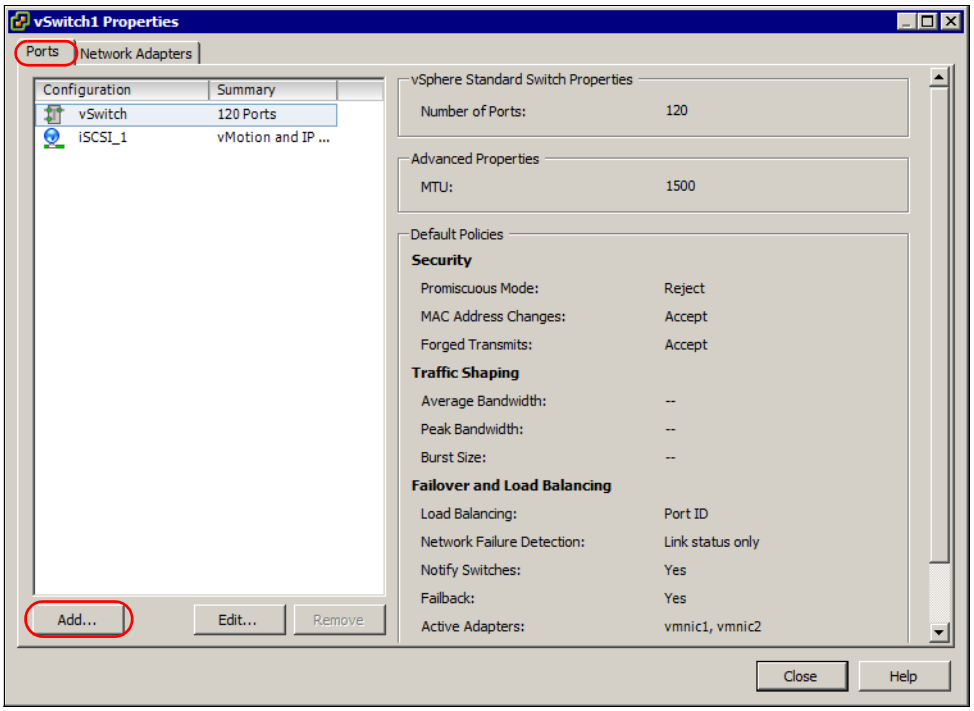

Figure 5-61 Adding ports to a vSphere standard switch

14. In the initial panel of the Add Network Wizard, select **VMkernel** for the connection type as shown in Figure 5-62 on page 105 and click **Next**.

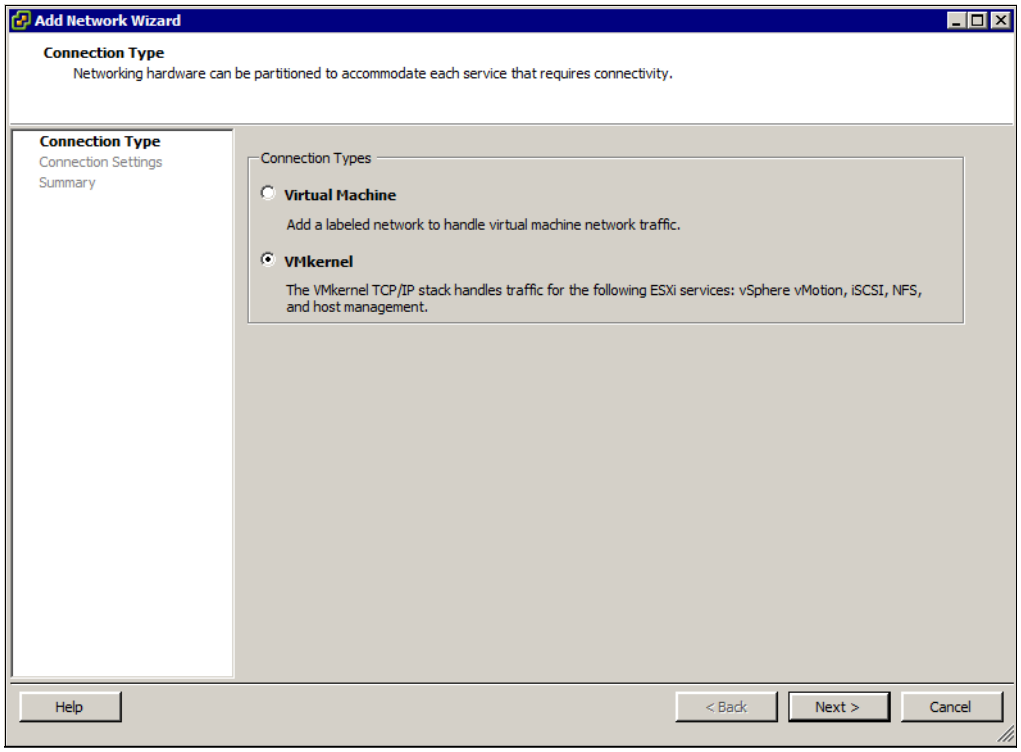

Figure 5-62 Network Connection Type panel

15. Label the VMkernel adapter as shown in Figure 5-63 and click **Next**.

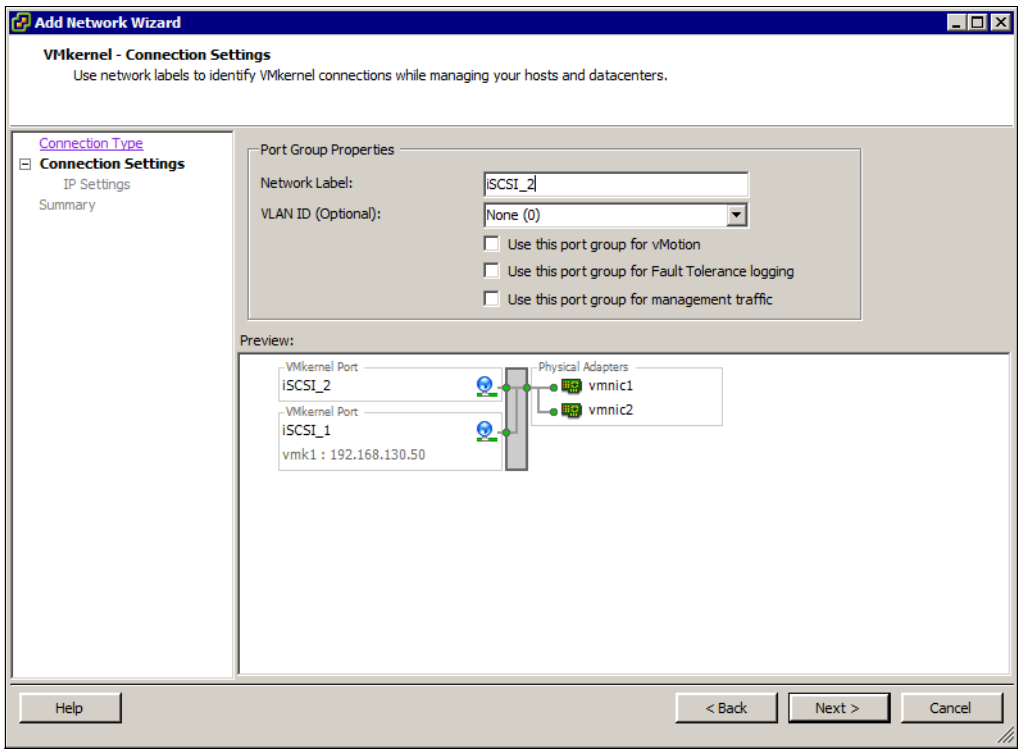

Figure 5-63 VMkernel label

16. Assign the IP address (for the second adapter) and the subnet mask, which is defined for your iSCSI network, as shown in Figure 5-64 and click **Next**.

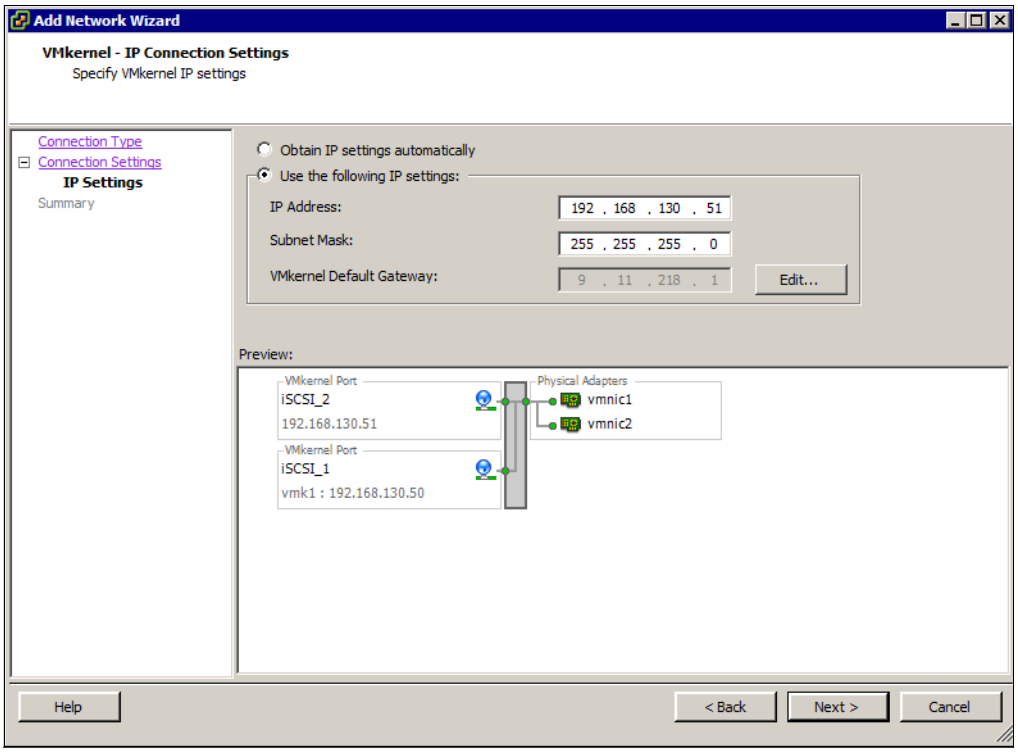

Figure 5-64 VMkernel IP configuration

17. On the final panel, which is shown in Figure 5-65 on page 107, check that all your settings are correct and click **Finish**.

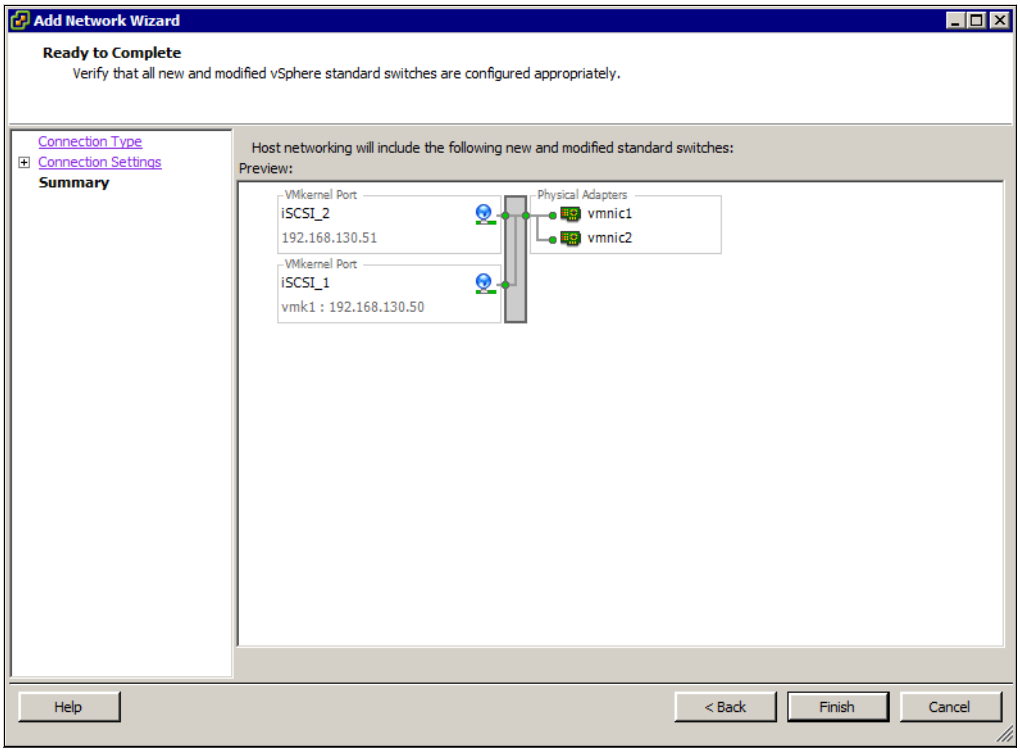

Figure 5-65 VMkernel overview panel

18. Select one of the VMkernels that you created for iSCSI and click **Edit** as shown in Figure 5-66.

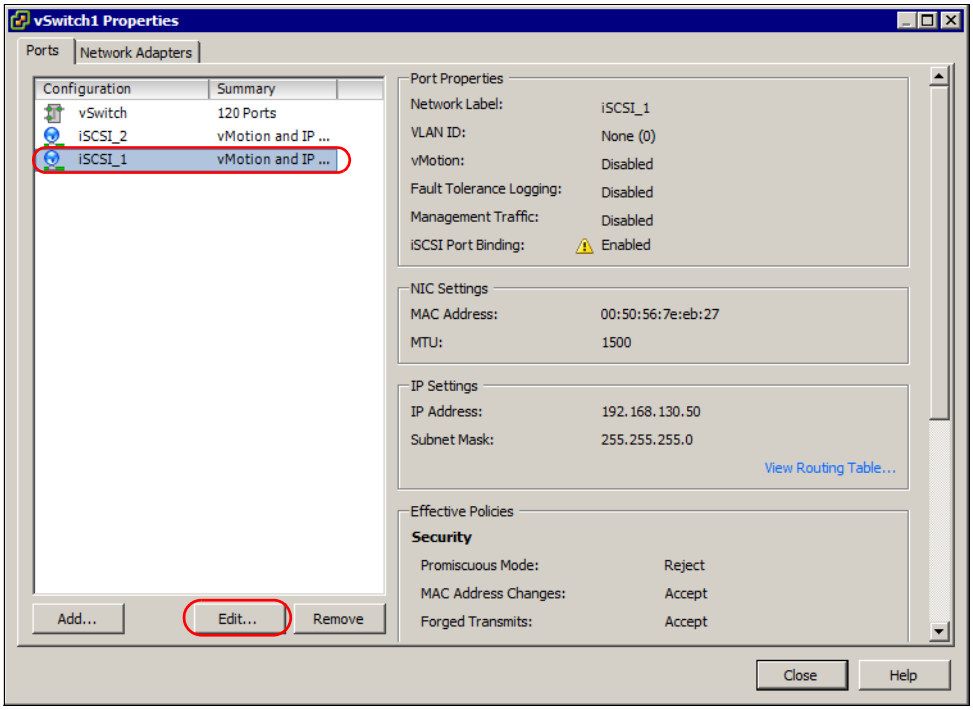

Figure 5-66 vSphere standard switch with two VMkernels for iSCSI

19. Click the **NIC Teaming** tab, select **Override switch failover order**, and move one of the adapters to the Unused Adapters port group by using **Move Down** as shown in Figure 5-67. Click **OK** to exit.

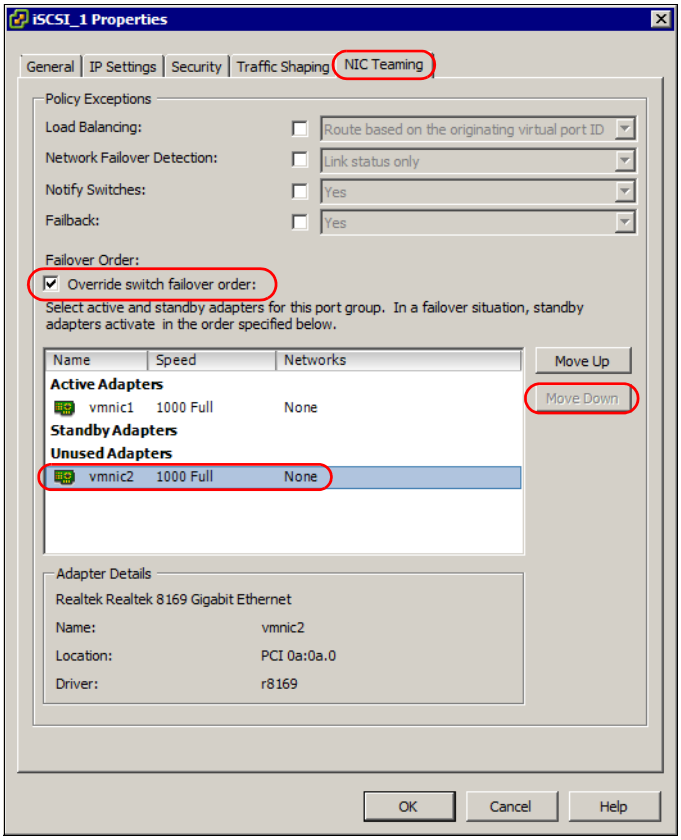

Figure 5-67 NIC Teaming tab

20. A confirmation window opens as shown in Figure 5-68. Click **Yes** to apply the settings.

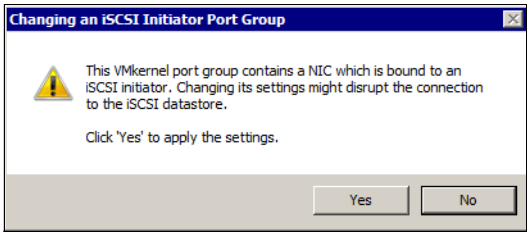

Figure 5-68 NIC Teaming confirmation

21. Select the second VMkernel that was created for iSCSI and click **Edit** as shown in Figure 5-69 on page 109.

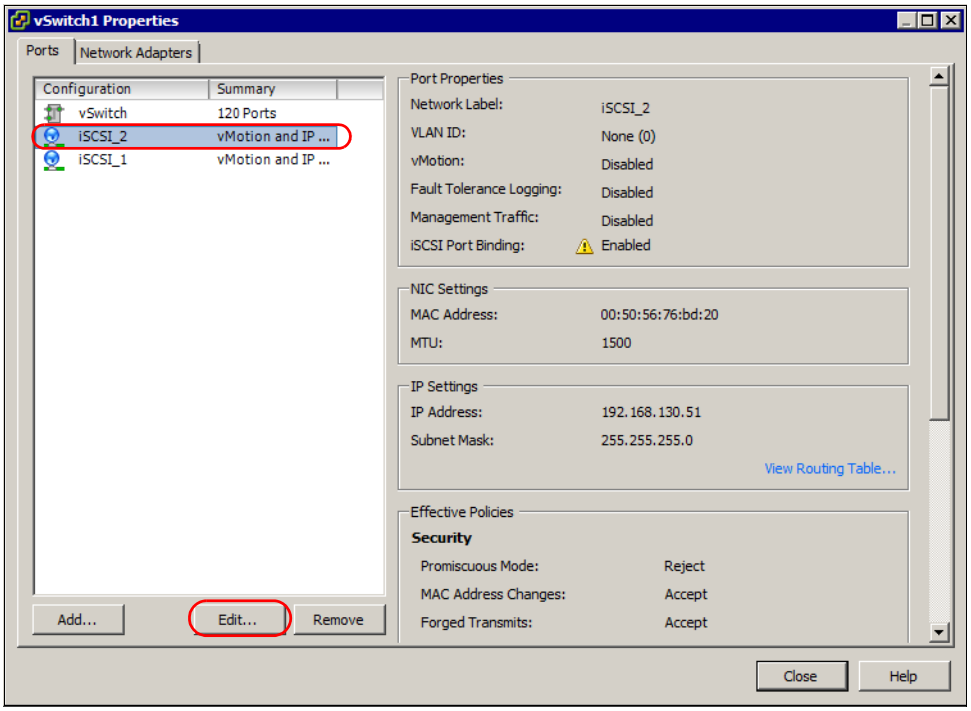

Figure 5-69 vSphere standard switch with two VMkernels for iSCSI

22. Click the **NIC Teaming** tab, select **Override switch failover order,** and move the second adapter to the Unused Adapters port group by using **Move Down** as shown in Figure 5-70 on page 110. Click **OK** to exit.

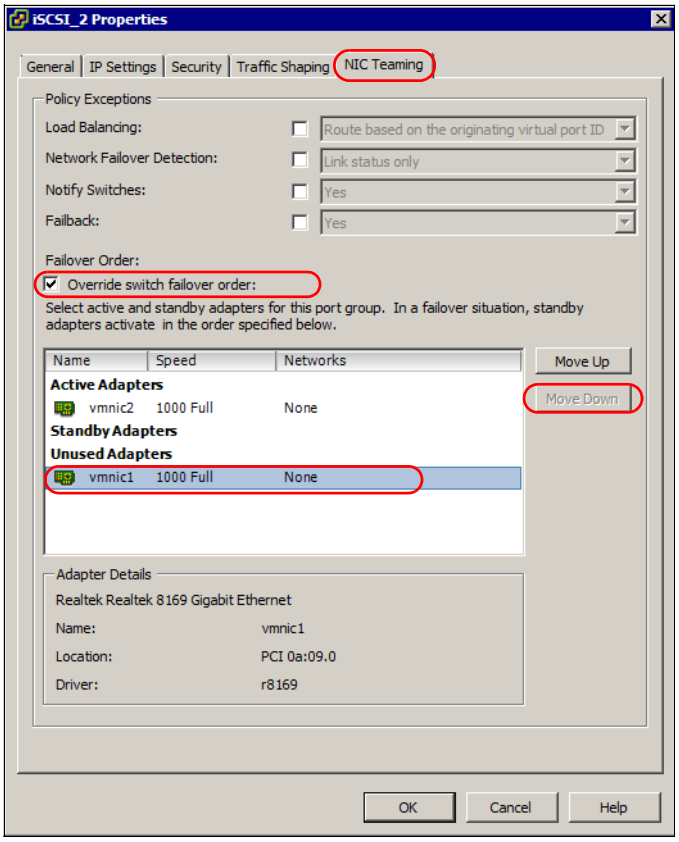

Figure 5-70 NIC Teaming tab

23. A confirmation window opens as shown in Figure 5-68 on page 108. Click **Yes** to apply the settings.

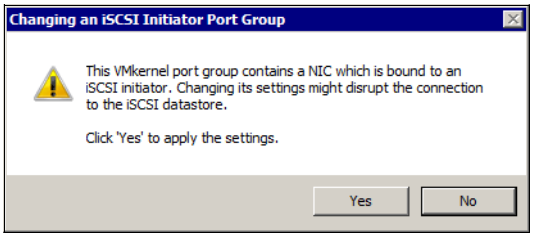

Figure 5-71 NIC Teaming confirmation

24. Each of the VMkernels is now bound to a separate adapter as shown in Figure 5-72 on page 111 and Figure 5-73 on page 111. Click **Close** to exit.

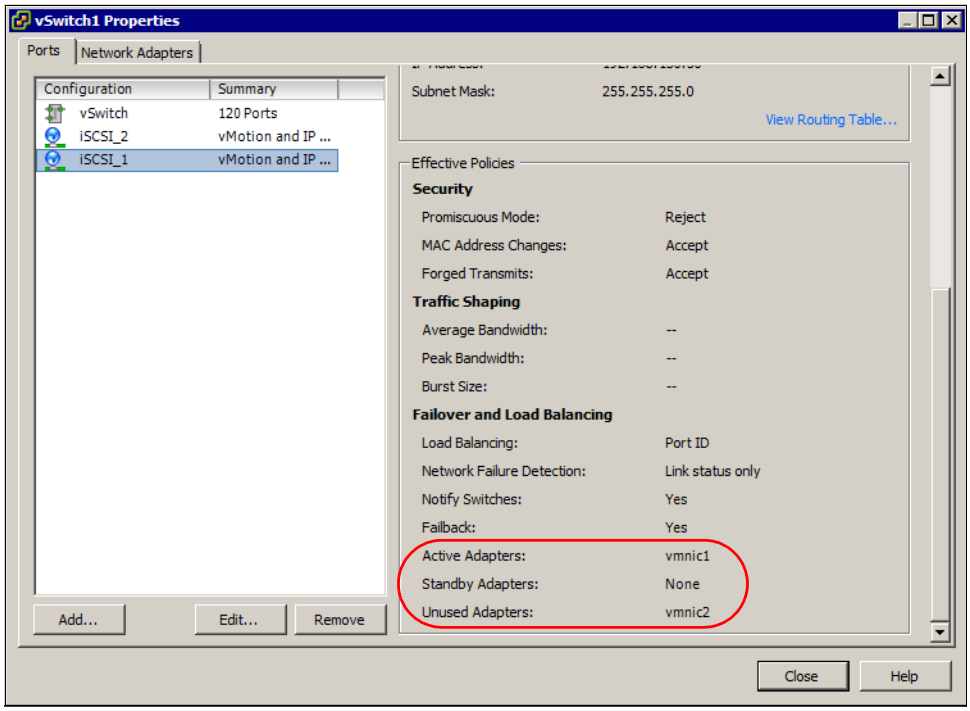

Figure 5-72 VMkernel to network adapter binding for the first VMkernel

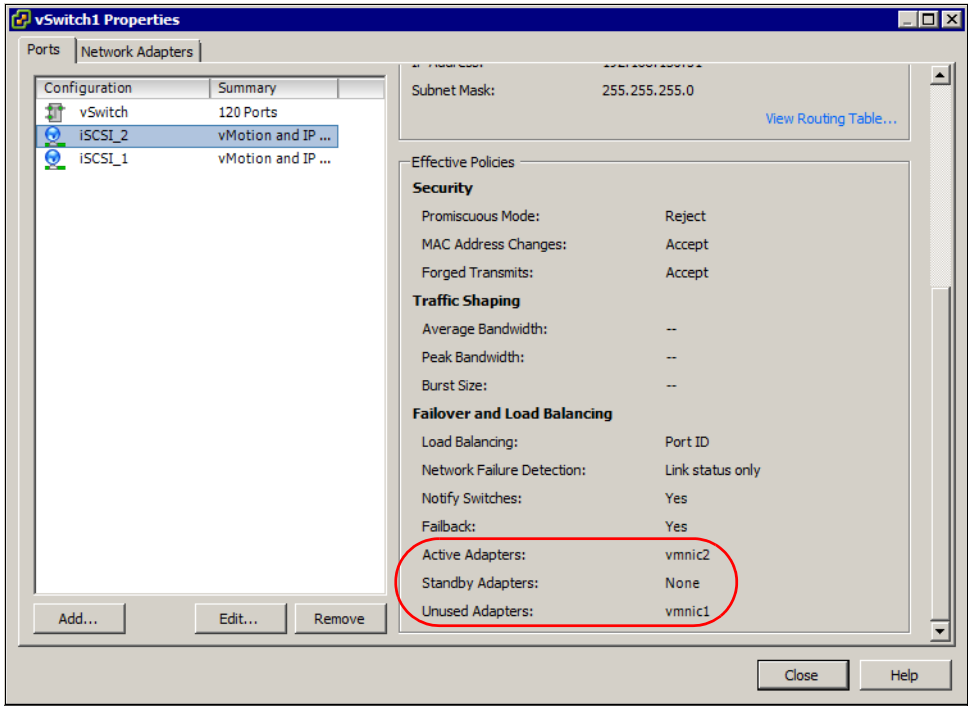

Figure 5-73 VMkernel to network adapter binding for the second VMkernel

25. Now, we see the network configuration with two VMkernel ports that are assigned to two network adapters as shown in Figure 5-74 on page 112.

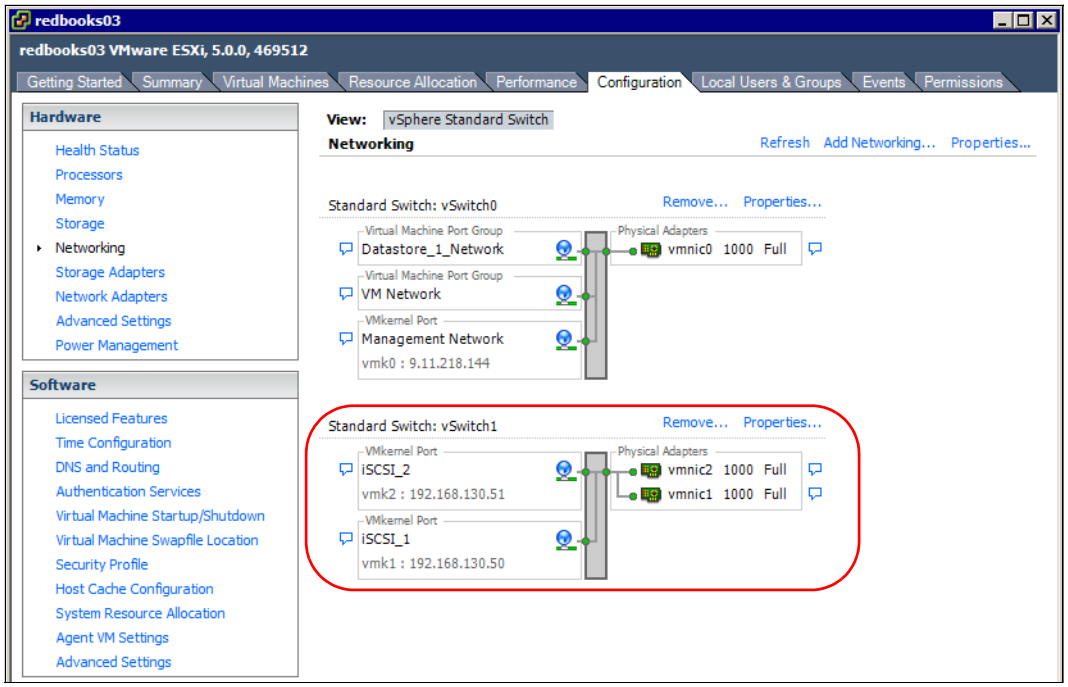

Figure 5-74 iSCSI network configuration

Configuring iSCSI discovery addresses

Follow the procedure to configure the iSCSI discovery addresses:

1. Go back to the Storage Adapters panel, select the iSCSI Software Adapter, and click **Properties** as shown in Figure 5-75 on page 113.

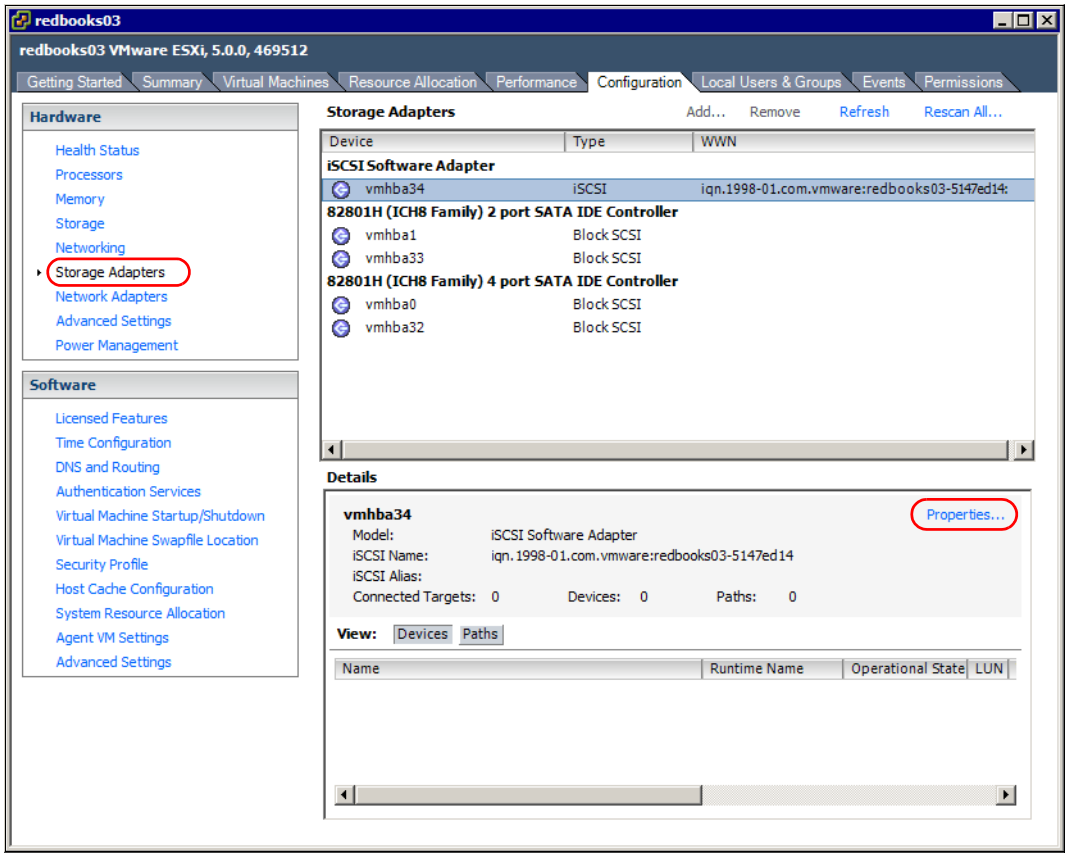

Figure 5-75 vSphere Storage Adapters

2. Click the **Network Configuration** tab and then click **Add** as shown in Figure 5-76 on page 114 to add the VMkernel ports to the iSCSI Software Adapter.

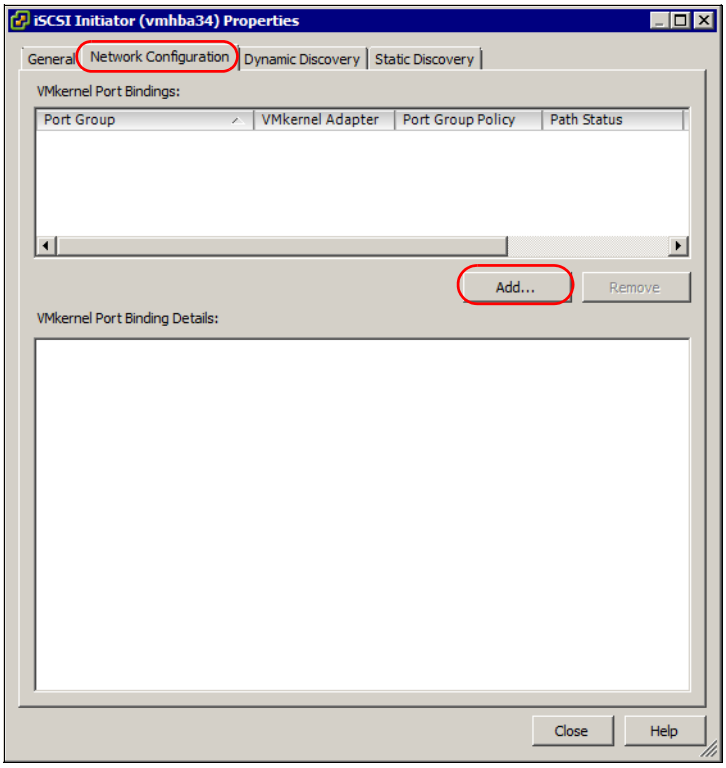

Figure 5-76 iSCSI Software Adapter Network Configuration tab

3. Select one of the VMkernels assigned for iSCSI as shown in Figure 5-77 and then click **OK**.

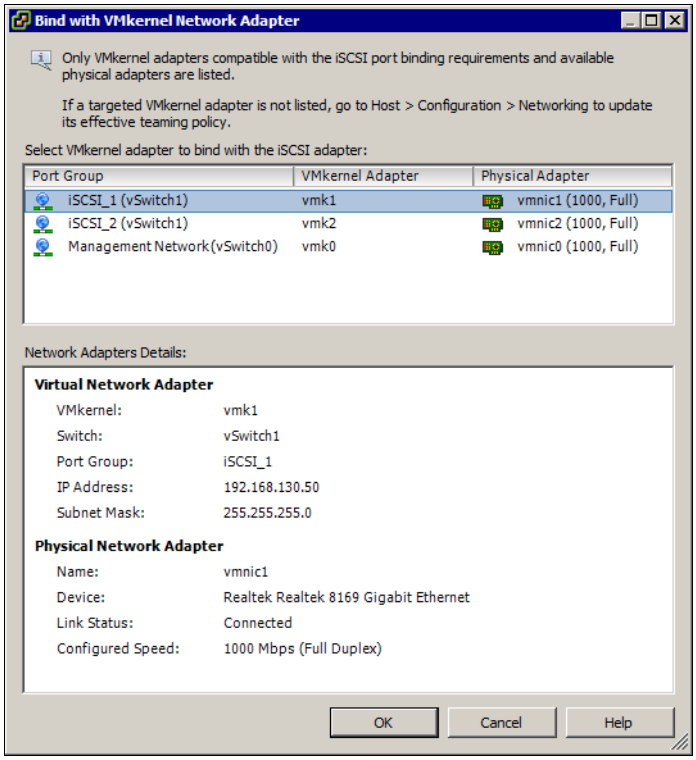

Figure 5-77 Adding VMkernel ports to the iSCSI Software Adapter

4. Repeat Step 2 on page 113 and Step 3 on page 114 for the second VMkernel port. After you add the second VMkernel port to the iSCSI Software Adapter, your configuration looks similar to the configuration that is shown in Figure 5-78.

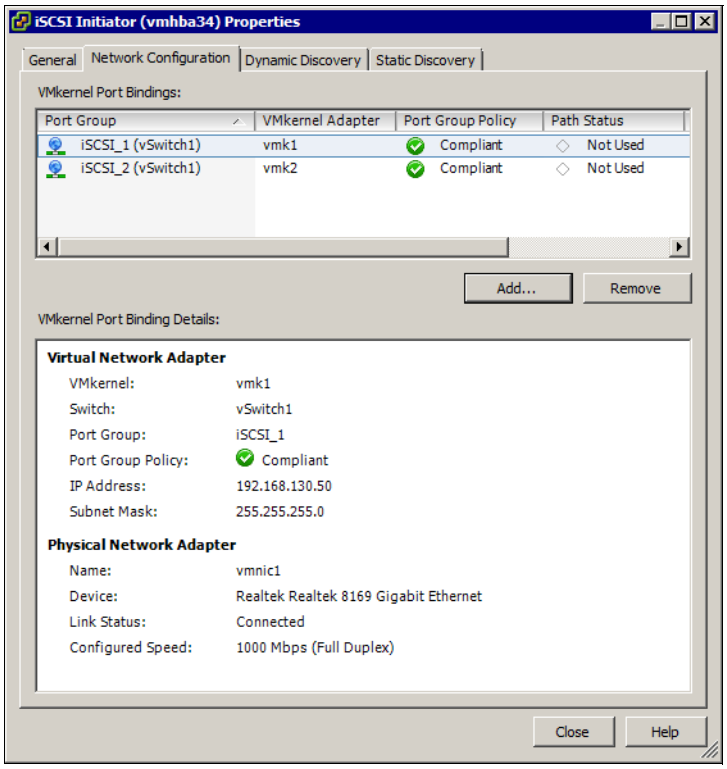

Figure 5-78 iSCSI Software Adapter Network Configuration tab with two VMkernel ports

5. Click the **Dynamic Discovery** tab and click **Add** as shown in Figure 5-79 on page 116.

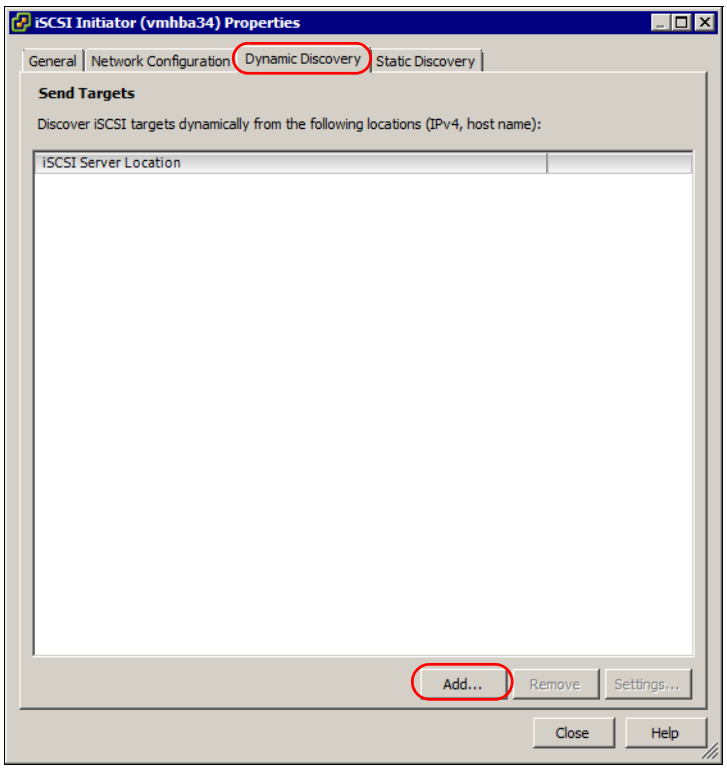

Figure 5-79 iSCSI Dynamic Discovery tab

6. Type the iSCSI IP address of one of the storage subsystem controllers as shown in Figure 5-80. The iSCSI Storage Adapter automatically assigns the iSCSI IP address of the second storage subsystem controller because both IP addresses are assigned to the same iSCSI target name.

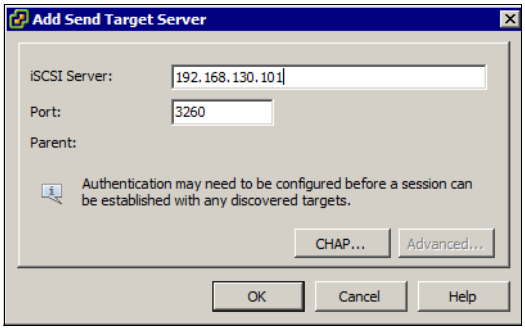

Figure 5-80 iSCSI Add Send Target Server

7. If you are using CHAP authentication for iSCSI communication, click **Use CHAP** as shown in Figure 5-80 and complete the CHAP credentials that are shown in Figure 5-81 on page 117. Click **OK** to close the CHAP credentials window.

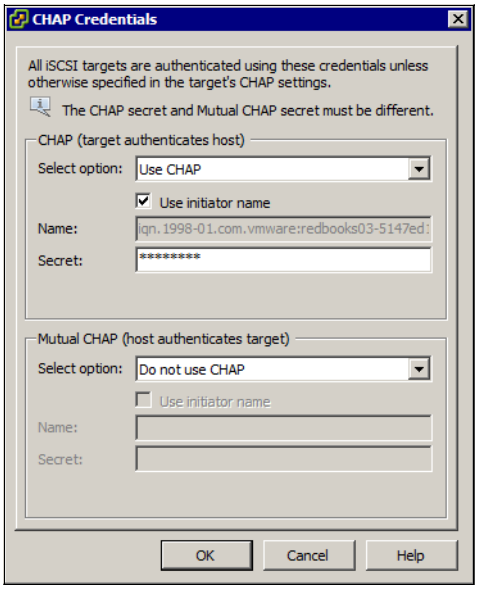

Figure 5-81 iSCSI CHAP authentication

8. Click **OK** to confirm adding the iSCSI target. After adding the iSCSI target and completing the CHAP authentication (if necessary), the iSCSI target appears in the list of Send Targets as shown in Figure 5-82.

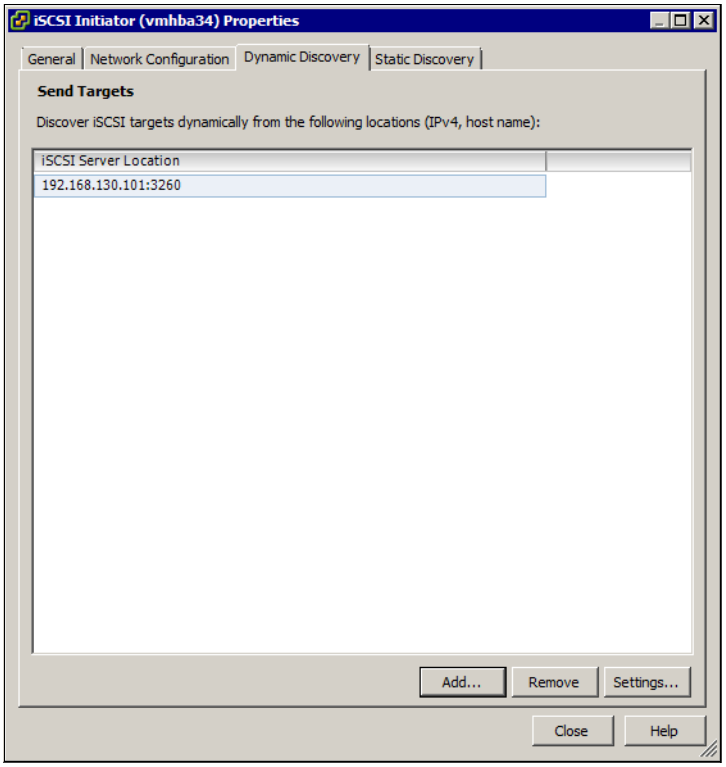

Figure 5-82 iSCSI target added

9. Click **Close** to close the iSCSI Software Adapter properties window. Because we changed the configuration of the iSCSI Software Adapter, we recommend a host bus adapter rescan (Figure 5-83 on page 118). Click **Yes** to perform the rescan.

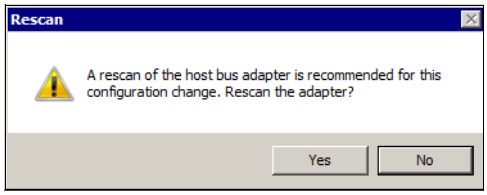

Figure 5-83 Adapter rescan

10. You now see all the LUNs that are mapped to this host as shown in Figure 5-84.

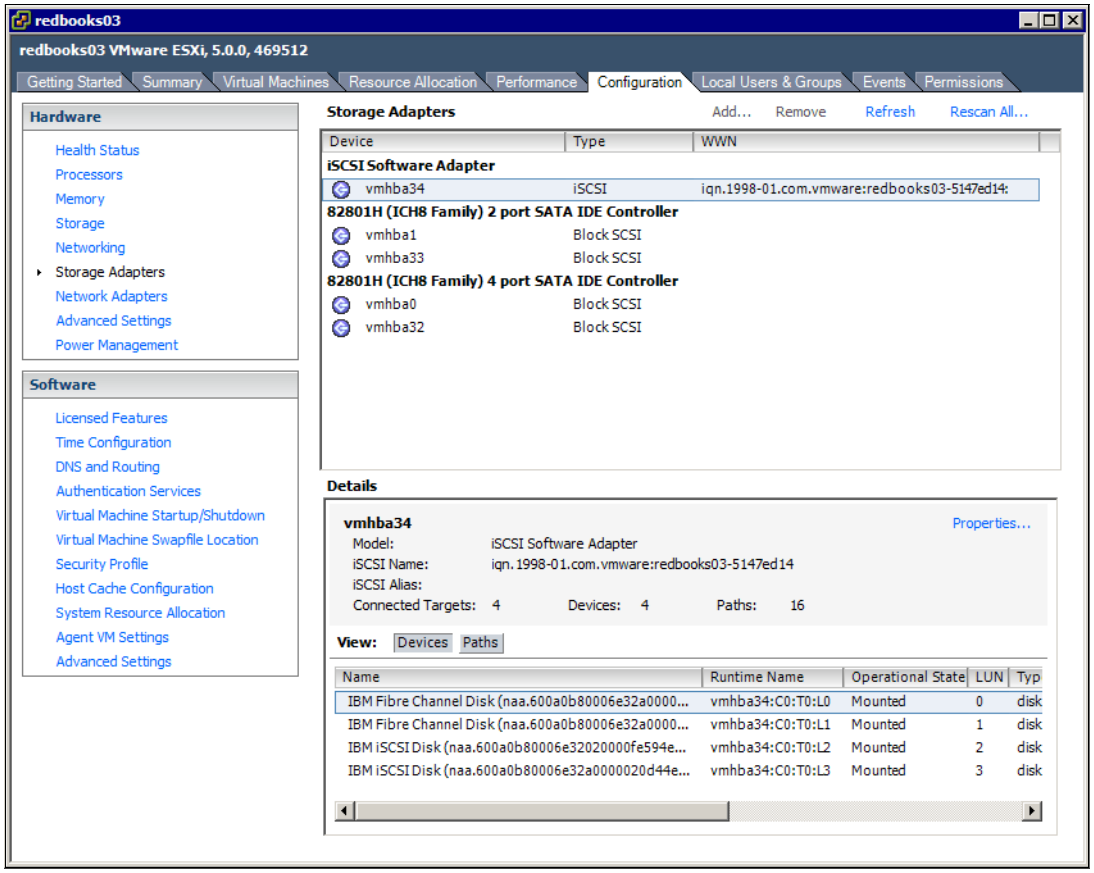

Figure 5-84 iSCSI discovered LUNs

11. To see all the paths that lead to these LUNs, click **Paths** as shown in Figure 5-85 on page 119. In this configuration, we have four LUNs and four paths per LUN, which is a total of 16 paths. Each LUN has two Active and two Stand by paths.

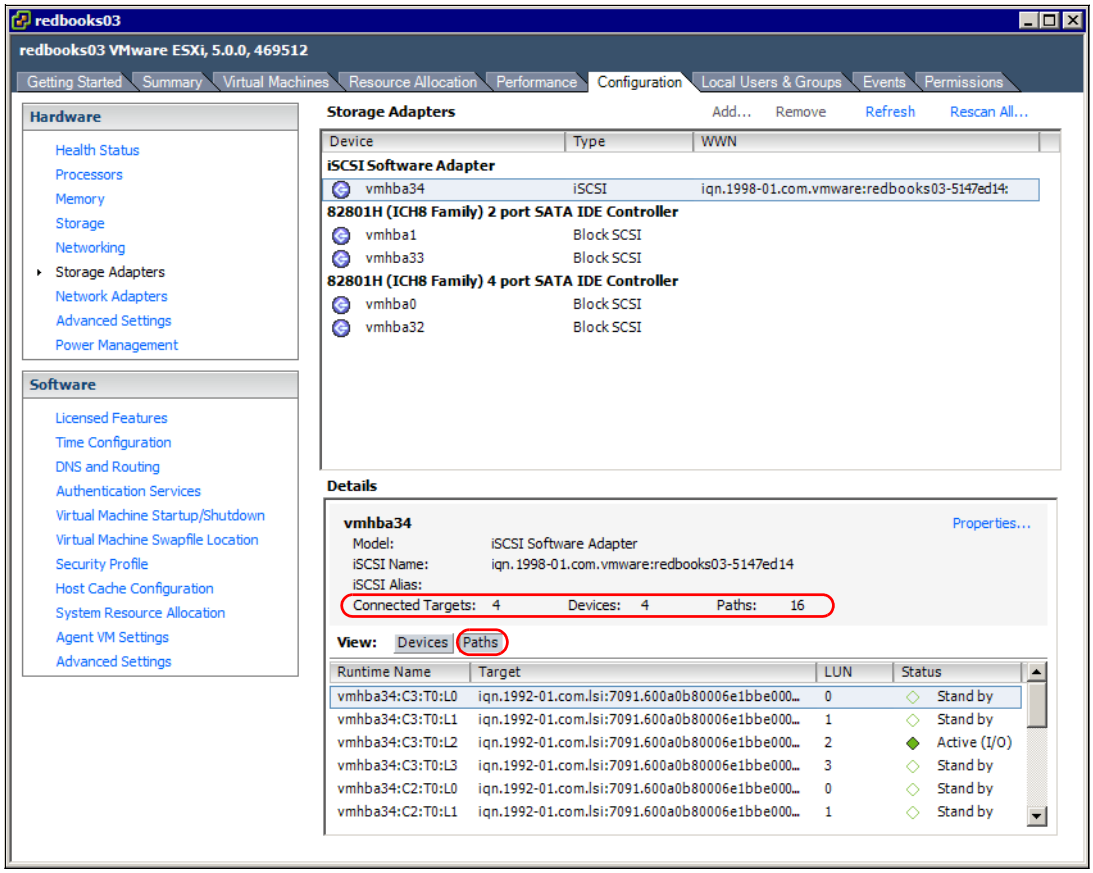

Figure 5-85 iSCSI discovered paths

5.2.7 Configuring VMware ESXi Server Storage

The following procedure demonstrates a basic configuration of FC or iSCSI storage for a VMware ESXi Server guest VM. This configuration might differ depending on your specific setup, for example, clustered or shared. For more information, see the VMware documentation that is listed in “Related publications” on page 255.

> **Note:** Although we show how to configure FC storage in this section, the procedure to configure iSCSI storage is identical.

Follow these steps:

1. Use the VMware vSphere Client to connect to the new VMware ESXi Server (log in as `root`).
2. On the Configuration tab, click **Storage** in the Hardware section as shown in Figure 5-86 on page 120.

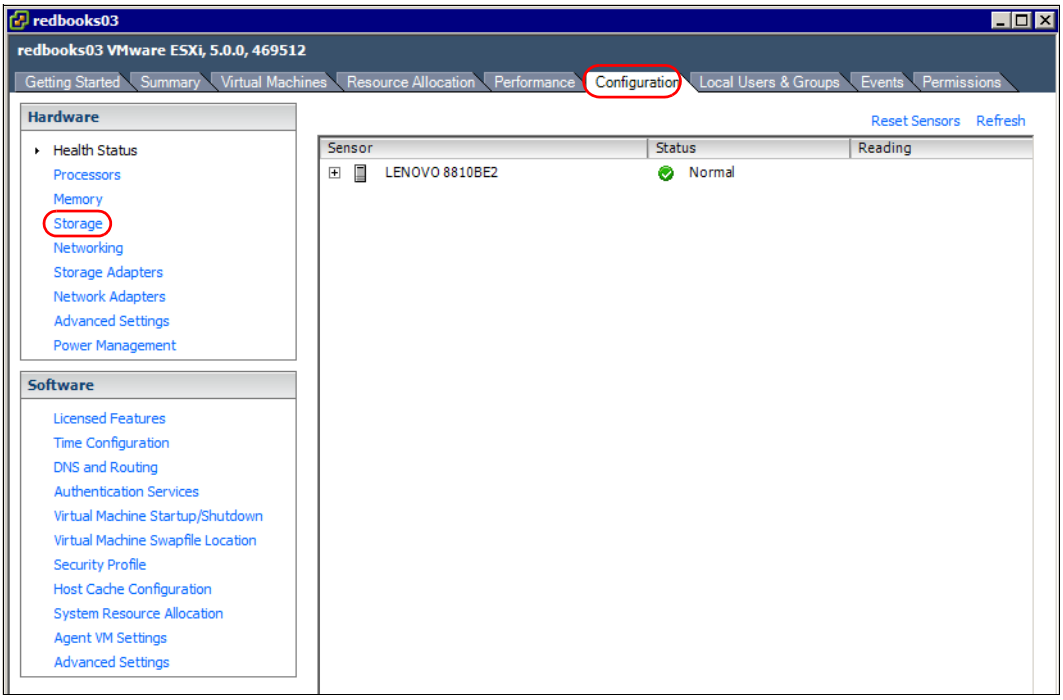

Figure 5-86 vSphere initial panel

3. In the Storage panel, click **Add Storage** as shown in Figure 5-87.

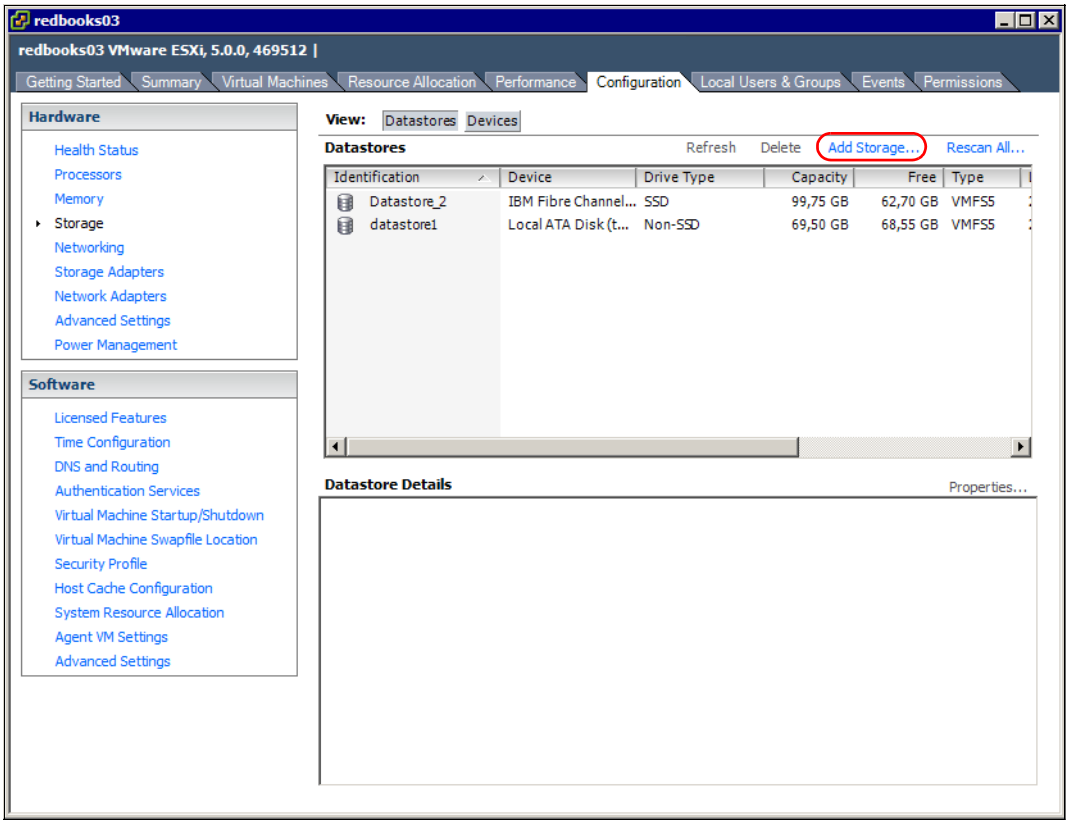

Figure 5-87 vSphere Client adding storage

4. You are presented with the Storage Type selection panel. Select **Disk/LUN** to create a datastore on the Fibre Channel SAN drives, as shown in Figure 5-88. Click **Next**.

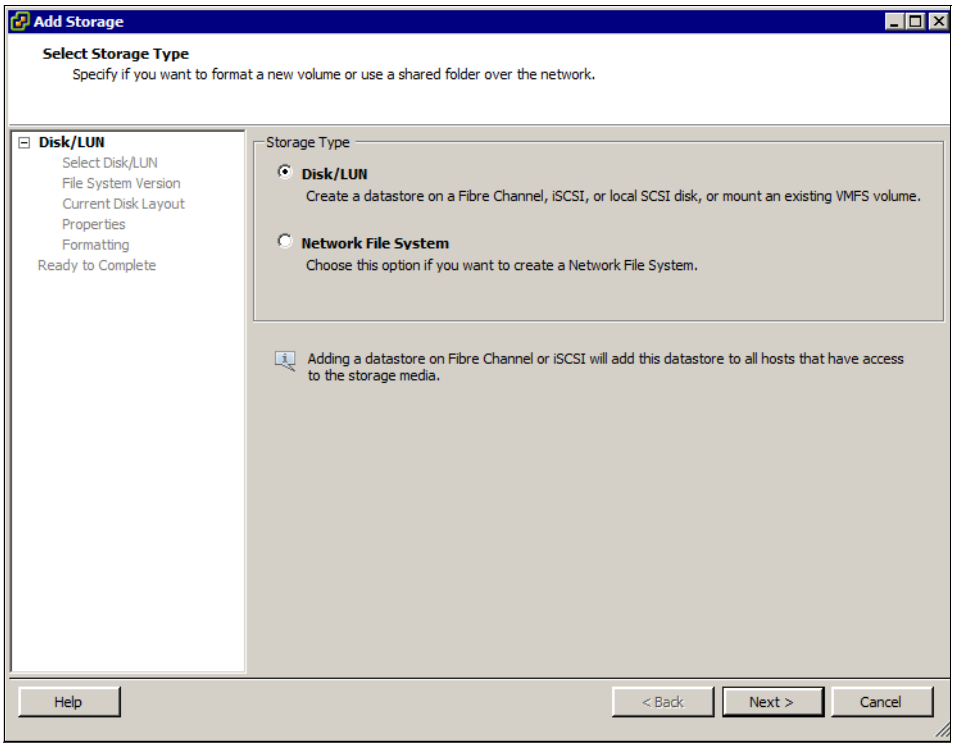

Figure 5-88 Storage Type selection

5. On Figure 5-89 on page 122, select the SAN Disk/LUN on which you want to create a datastore VMFS partition. Click **Next**.

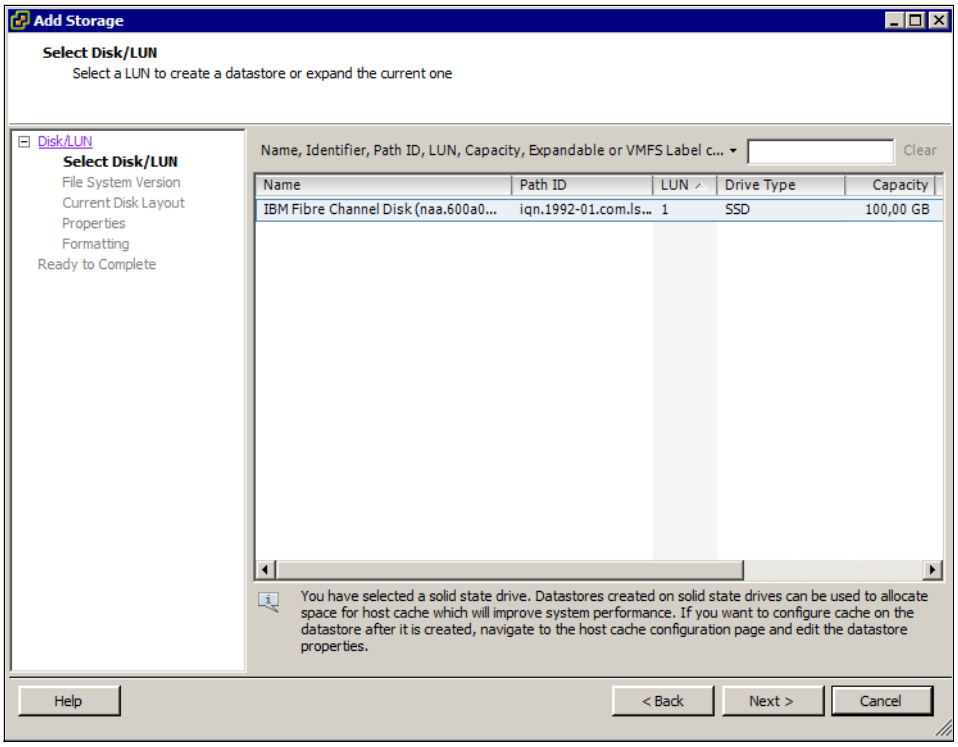

Figure 5-89 Select a LUN

6. Select the File System Version as shown in Figure 5-90.

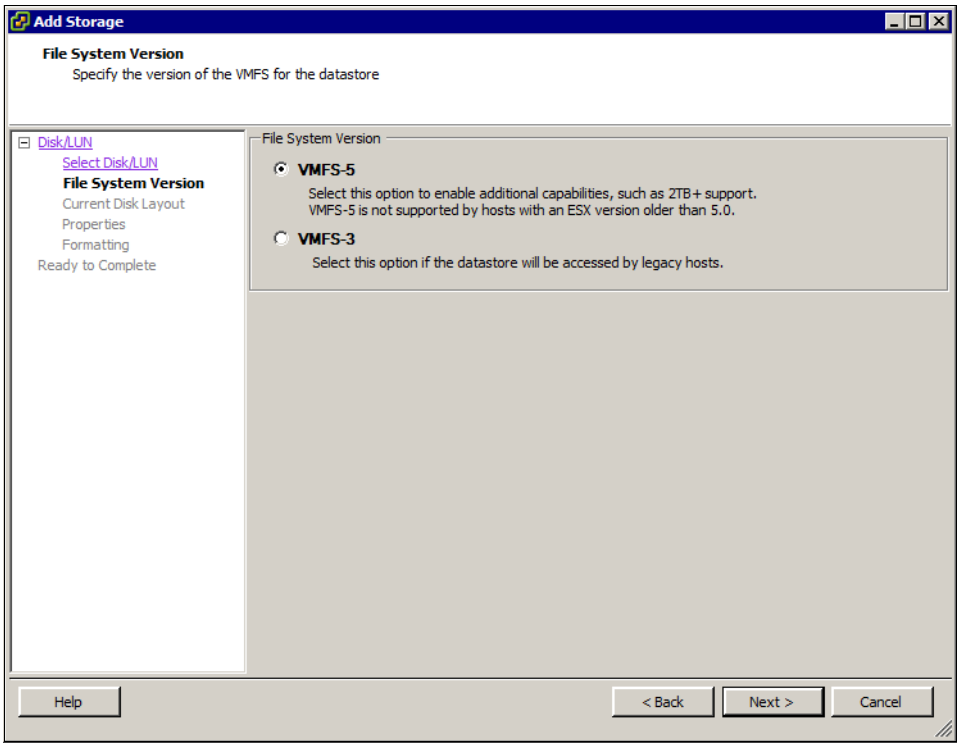

Figure 5-90 File System Version

7. Figure 5-91 on page 123 shows the disk layout of the LUN. Click **Next**.

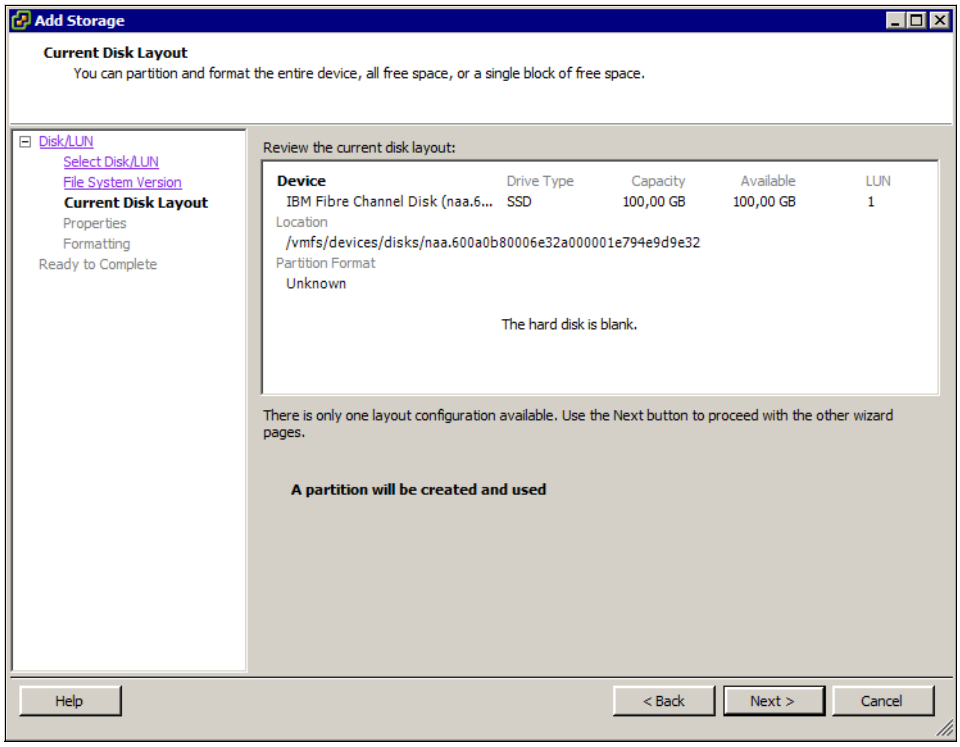

Figure 5-91 Disk layout

8. Enter a descriptive name for the datastore, and click **Next** as shown in Figure 5-92.

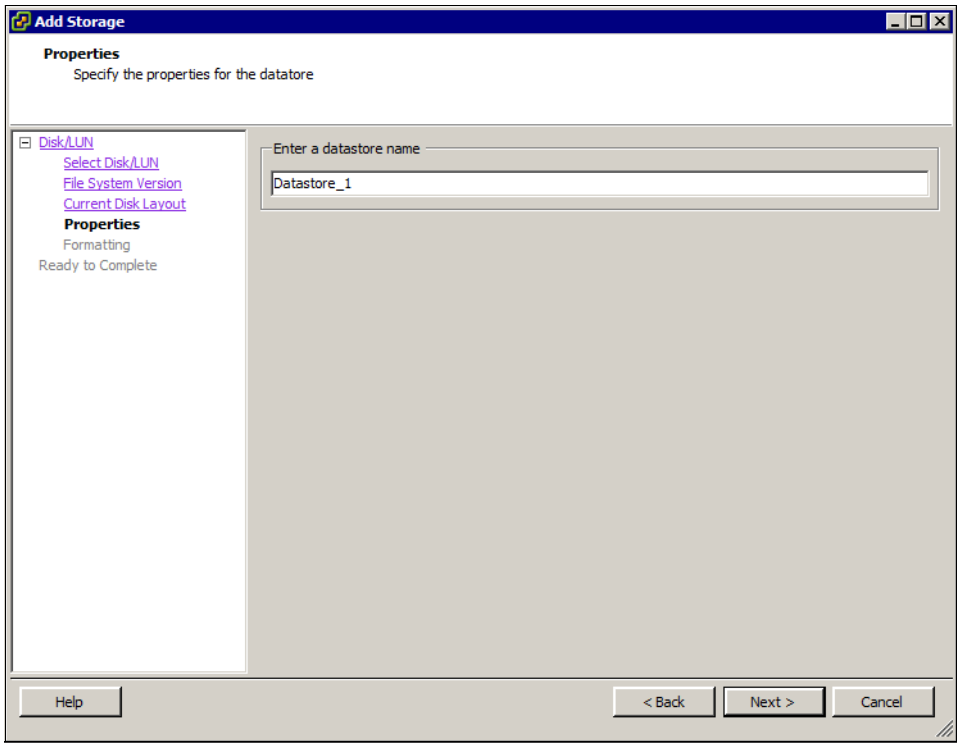

Figure 5-92 Datastore name

9. On Figure 5-93 on page 124, select the appropriate LUN capacity, and click **Next**.

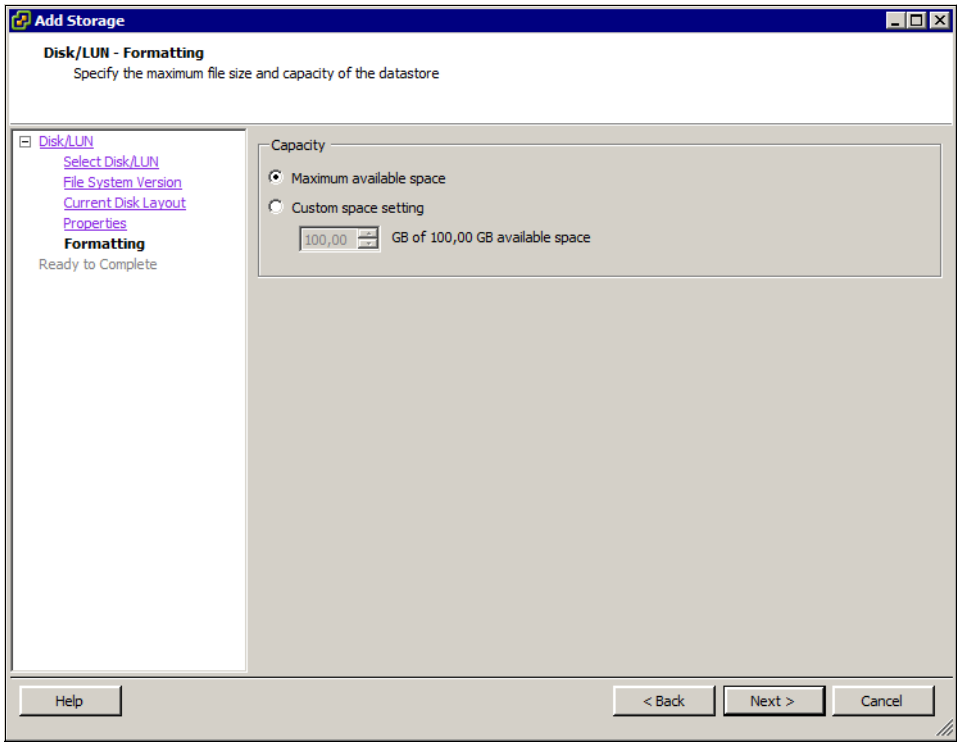

Figure 5-93 LUN capacity

10. Figure 5-94 is a summary panel for adding the storage. Click **Finish** to proceed.

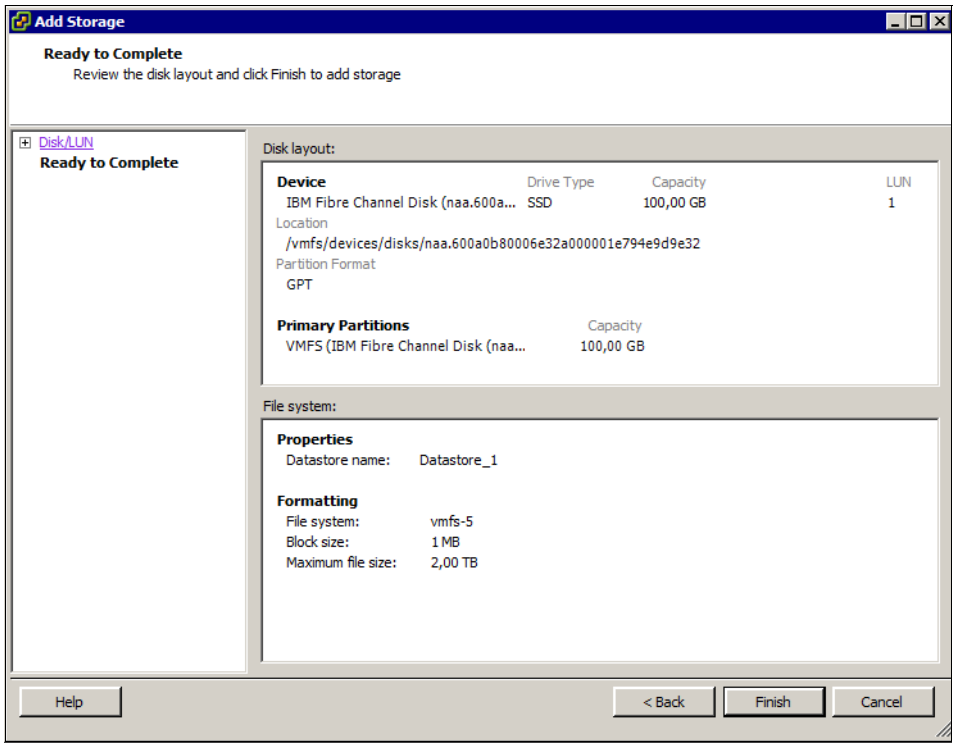

Figure 5-94 Adding storage summary panel

11. Click **Refresh** to show the newly created datastore as shown in Figure 5-95 on page 125.

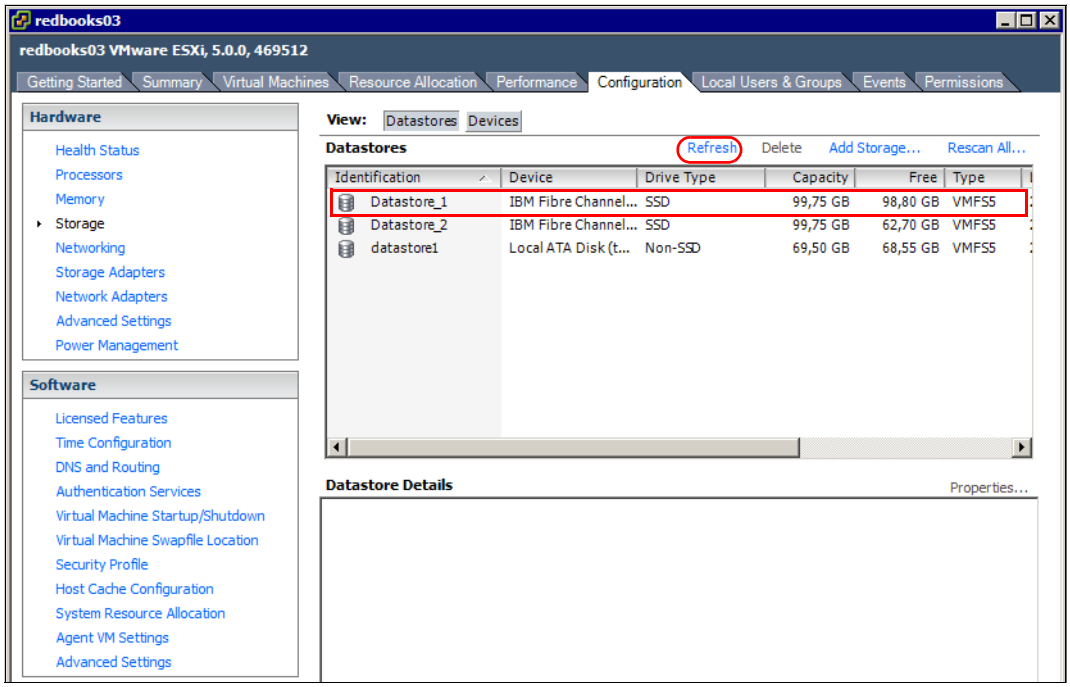

Figure 5-95 vSphere Client Storage new datastore

12. Repeat the same task for any additional SAN Fibre Channel or iSCSI LUNs.

5.2.8 Verifying the multipathing policy for Fibre Channel LUNs

This procedure describes how to set up and verify the multipathing policy for your Fibre Channel LUNs by using VMware ESXi Server.

> **Important:** The VMware Path Selection policy must be set to **Most Recently Used** for all DS5000 LUNs.

Follow these steps:

1. Use the VMware vSphere Client to connect to the VMware ESXi Server (log in as `root`).
2. Click the **Configuration** tab and select **Storage** as shown in Figure 5-96 on page 126.

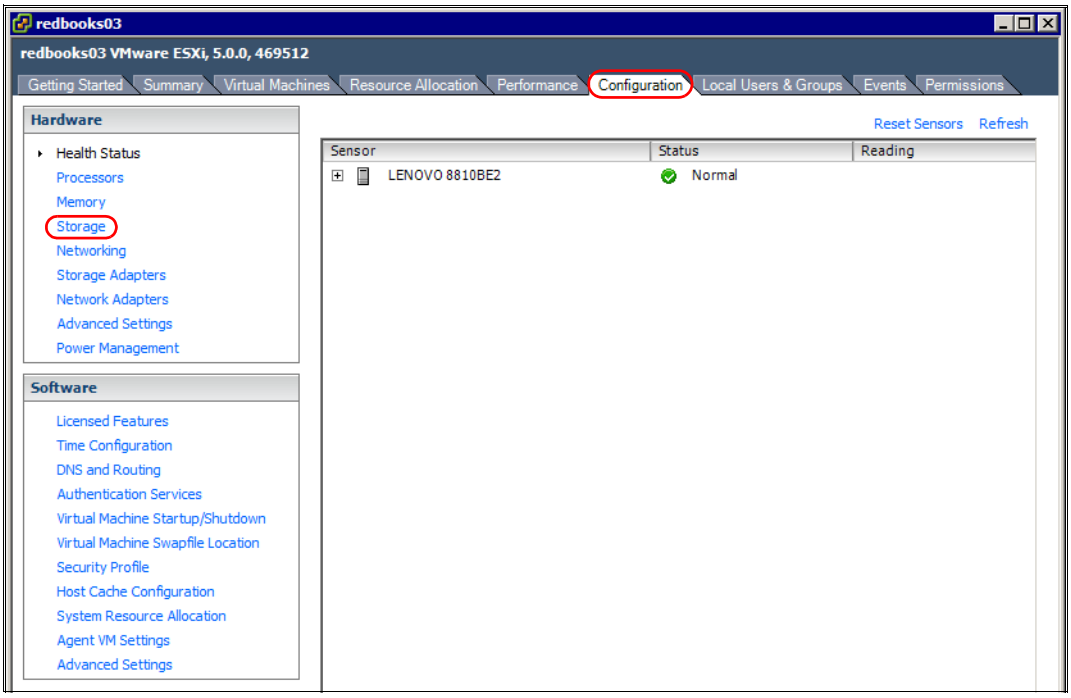

Figure 5-96 vSphere Client initial panel

3. Select your datastore and click **Properties** as shown in Figure 5-97.

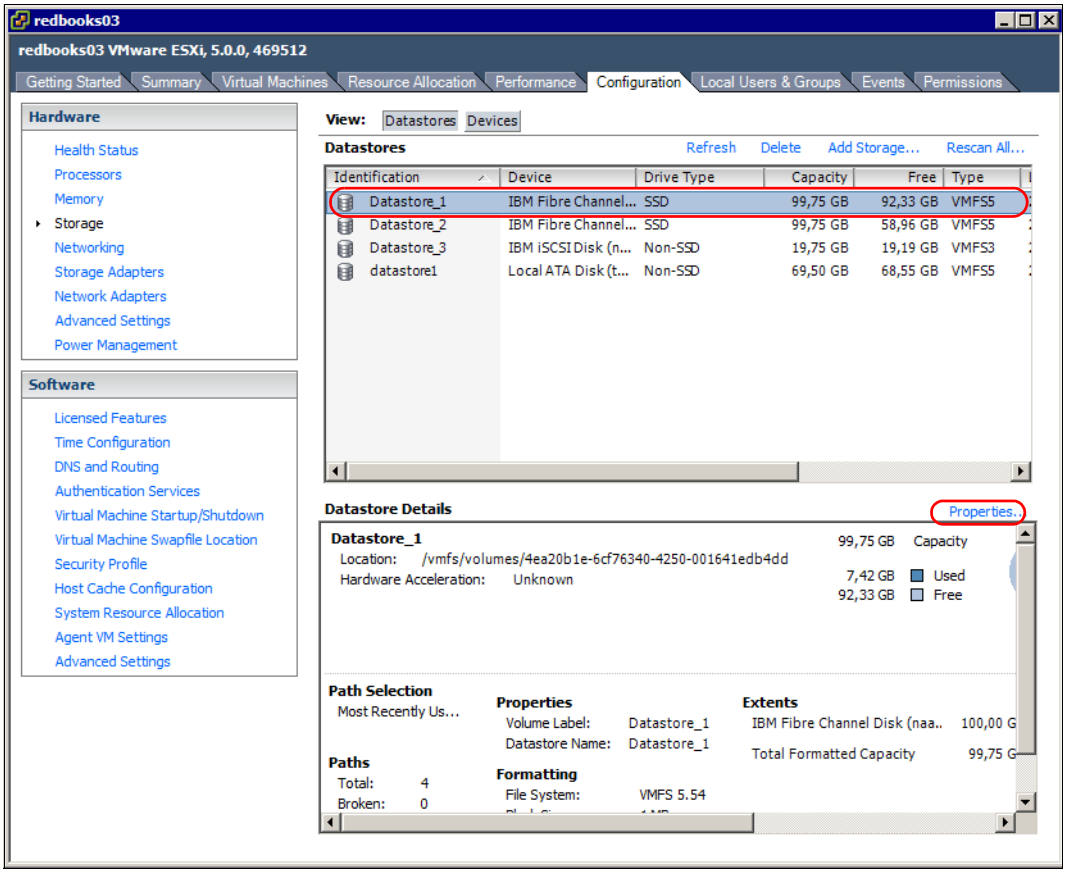

Figure 5-97 VMware Storage Datastores

4. Click **Manage Paths** as shown in Figure 5-98.

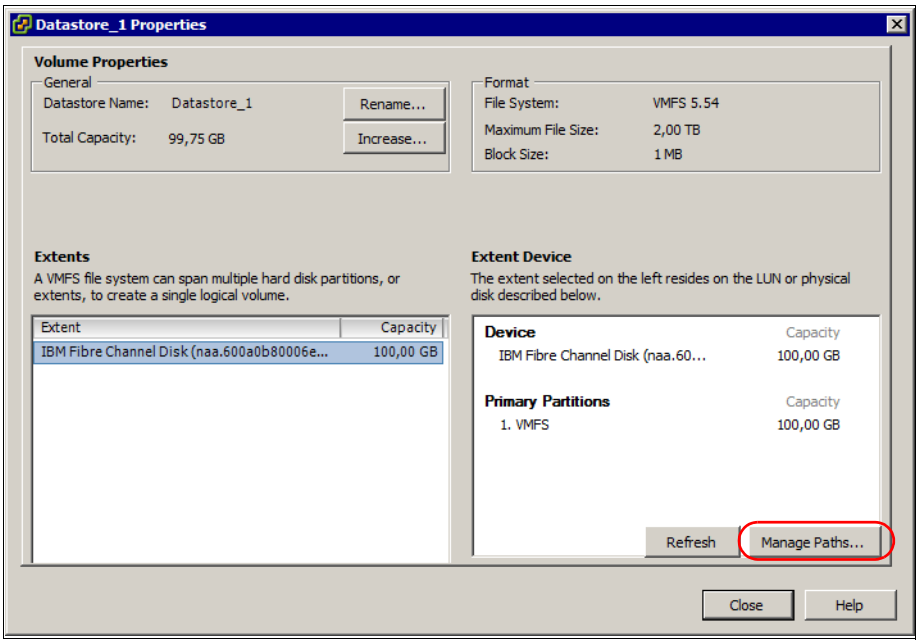

Figure 5-98 VMware datastore properties

5. If everything is zoned correctly as described in 3.4.11, “Zoning” on page 52, the VMware ESXi Server has four paths to each LUN as shown in Figure 5-99. Two paths are in `Stand by` status, and two paths are in `Active` status. Path selection must be set to **Most Recently Used (VMware)**.

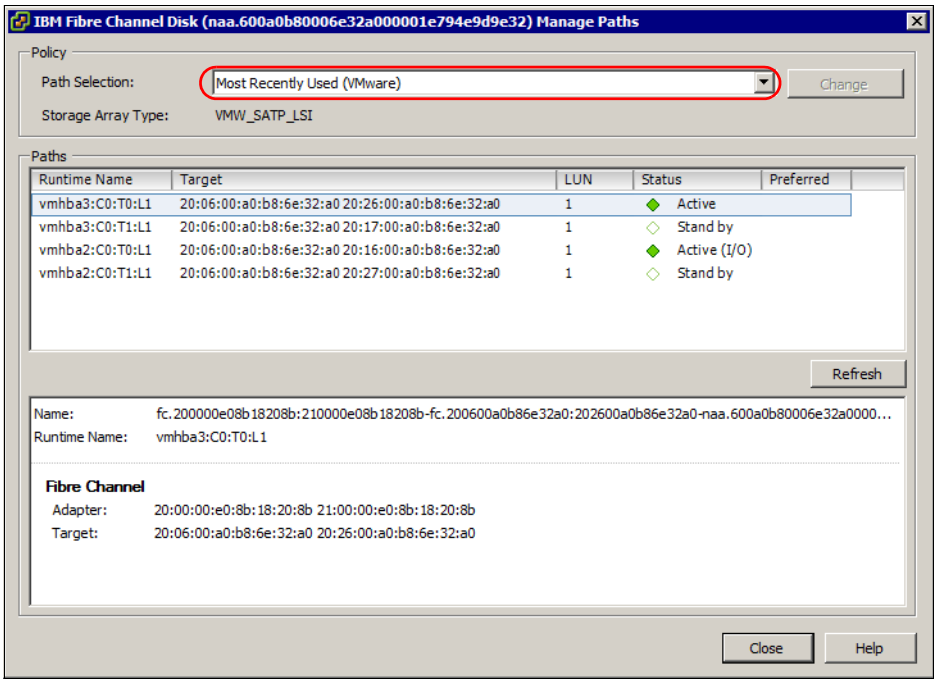

Figure 5-99 Manage Paths

5.2.9 Creating virtual machines

We explain how to create a virtual machine. How you configure your virtual machines is dictated by your requirements (guest operating system, virtual hardware requirements, function, and so on). For our example, we selected the creation of a virtual machine running Novell SUSE Linux Enterprise 11 (32-bit).

Follow these steps to create a virtual machine:

1. Use the VMware vSphere Client to connect to the VMware ESXi Server (log in as `root`).
2. Click **File** → **New** → **Virtual Machine** as shown in Figure 5-100.

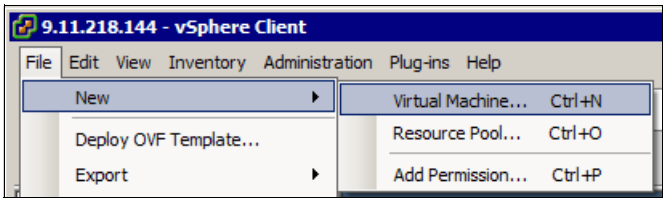

Figure 5-100 New Virtual Machine

3. On the “Select the configuration for the virtual machine” panel, select **Custom**, as shown in Figure 5-101. Click **Next**.

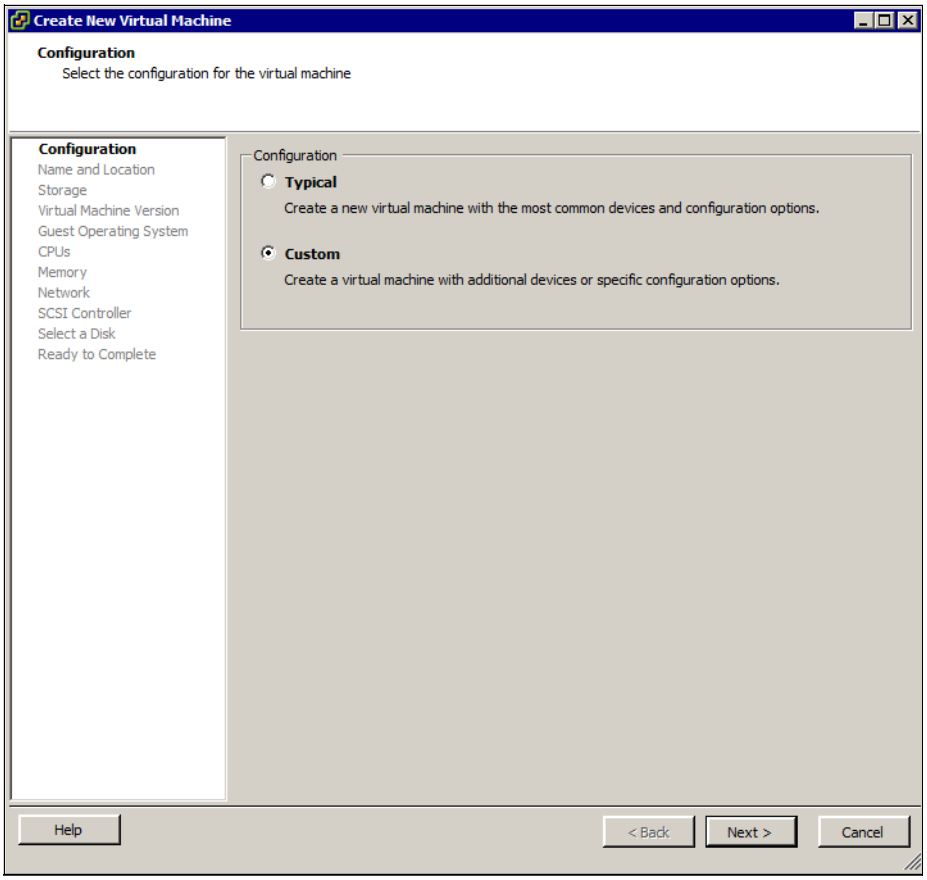

Figure 5-101 Configuration type

4. In the Virtual Machine Name field, type the name of the virtual machine, as shown in Figure 5-102 on page 129. Click **Next**.

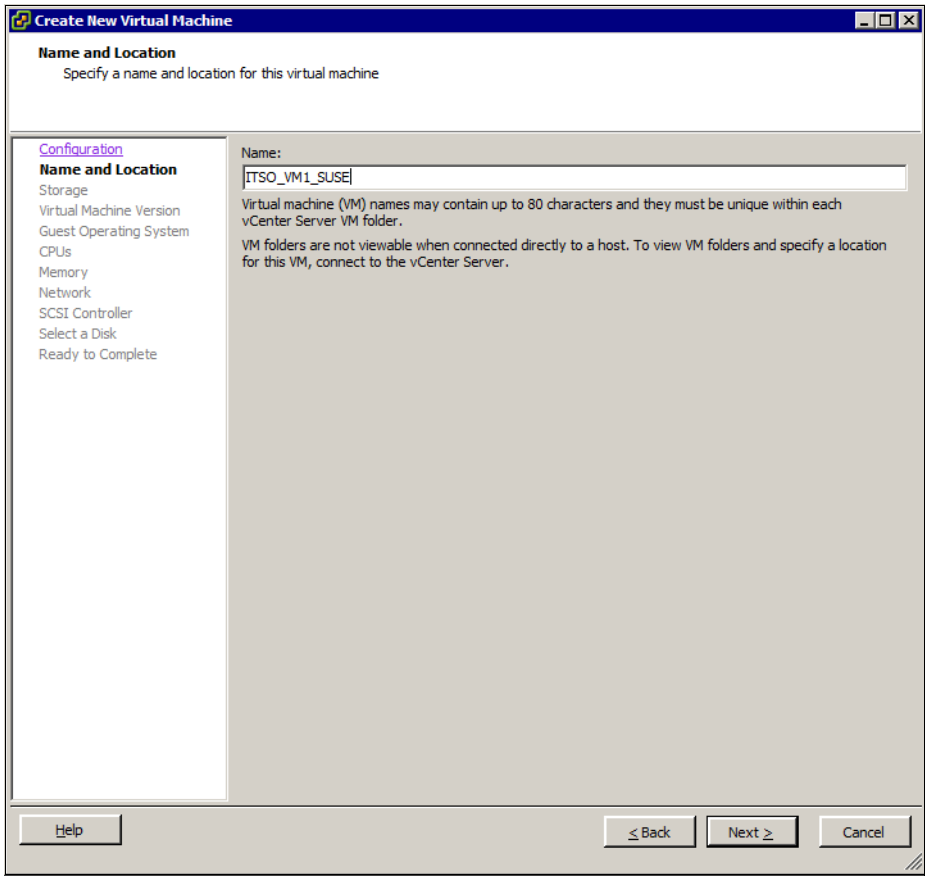

Figure 5-102 Virtual Machine Name

5. Select the datastore VMFS partition where the guest files will reside, which can be the Datastore_1 partition that we created in the earlier steps. All configuration files, including the disk files, are in that location. See Figure 5-103 on page 130. Click **Next**.

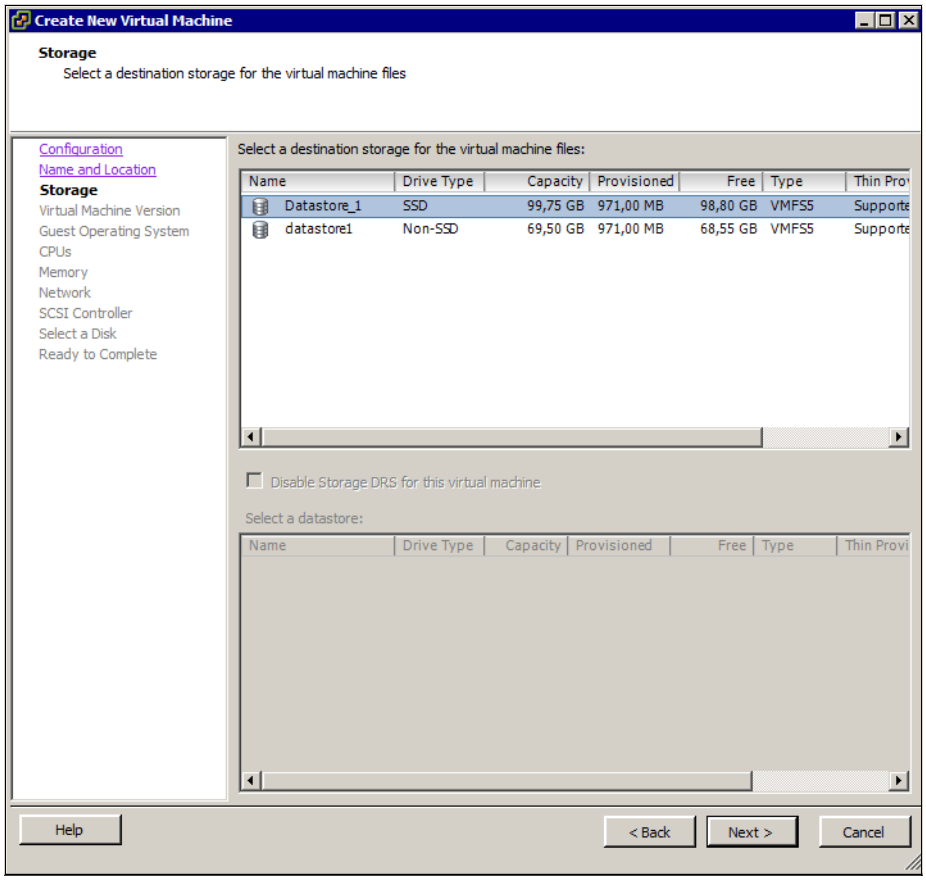

Figure 5-103 Datastore VMFS partition

6. Select the Virtual Machine Version, based on your requirements, as shown in Figure 5-104 on page 131, and click **Next**.

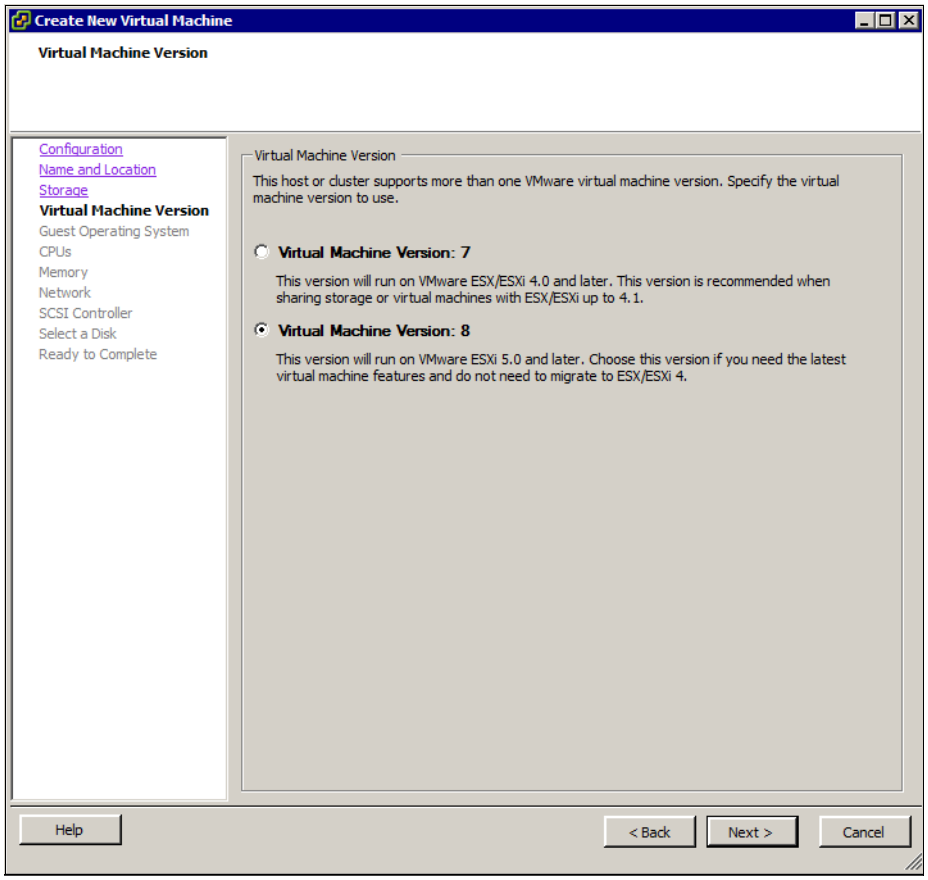

Figure 5-104 Virtual Machine Version

7. Select the Guest Operating System and the corresponding Version. In our example, we use Novell SUSE Linux Enterprise 11 (32-bit), as shown in Figure 5-105 on page 132. Click **Next**.

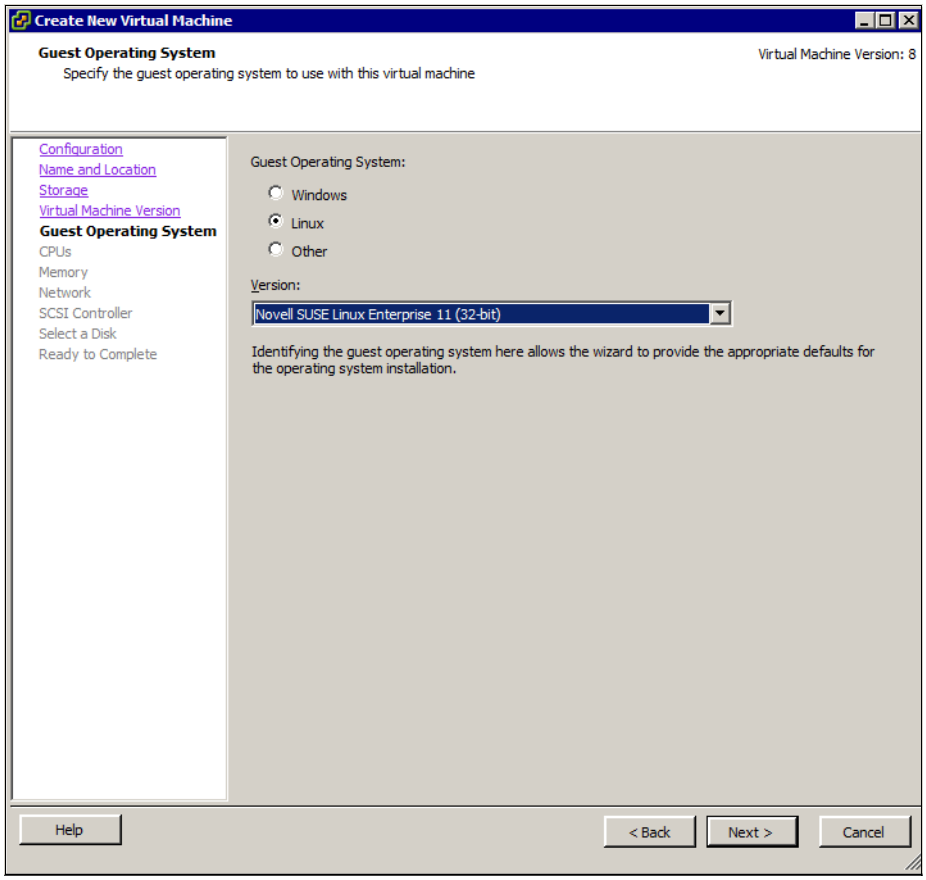

Figure 5-105 Guest Operating System selection

8. Select the number of virtual sockets (CPUs) and a number of cores per virtual socket needed for your operating environment, as shown in Figure 5-106 on page 133. Click **Next**.

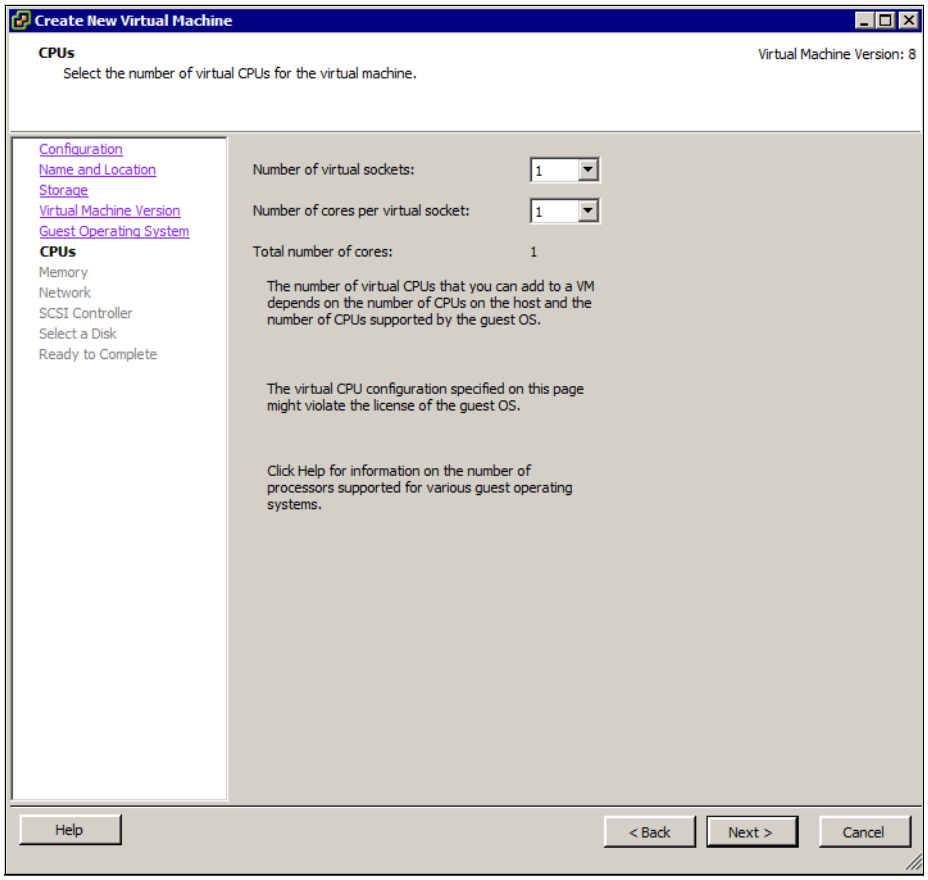

Figure 5-106 vCPU selection

9. On the Memory allocation panel, Figure 5-107 on page 134, provide the guest with the necessary amount of required memory. Click **Next**.

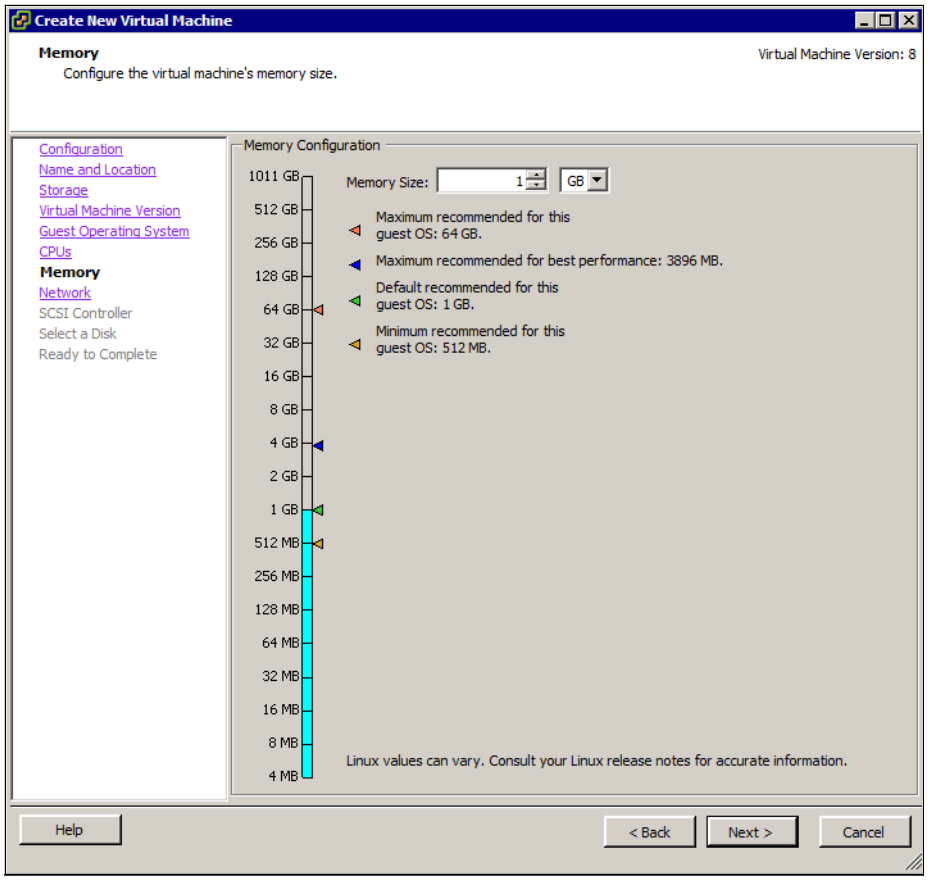

Figure 5-107 Memory allocation

10. On the Choose Networks panel, Figure 5-108 on page 135, select the appropriate number of Network Adapters with which the guest will operate (Default is 1). Choose the appropriate Network Label (ensure that the host is not overloading one particular network), which can be the VM network that we defined in 5.2.5, “Creating additional virtual switches for guests’ connectivity” on page 90. Click **Next**.

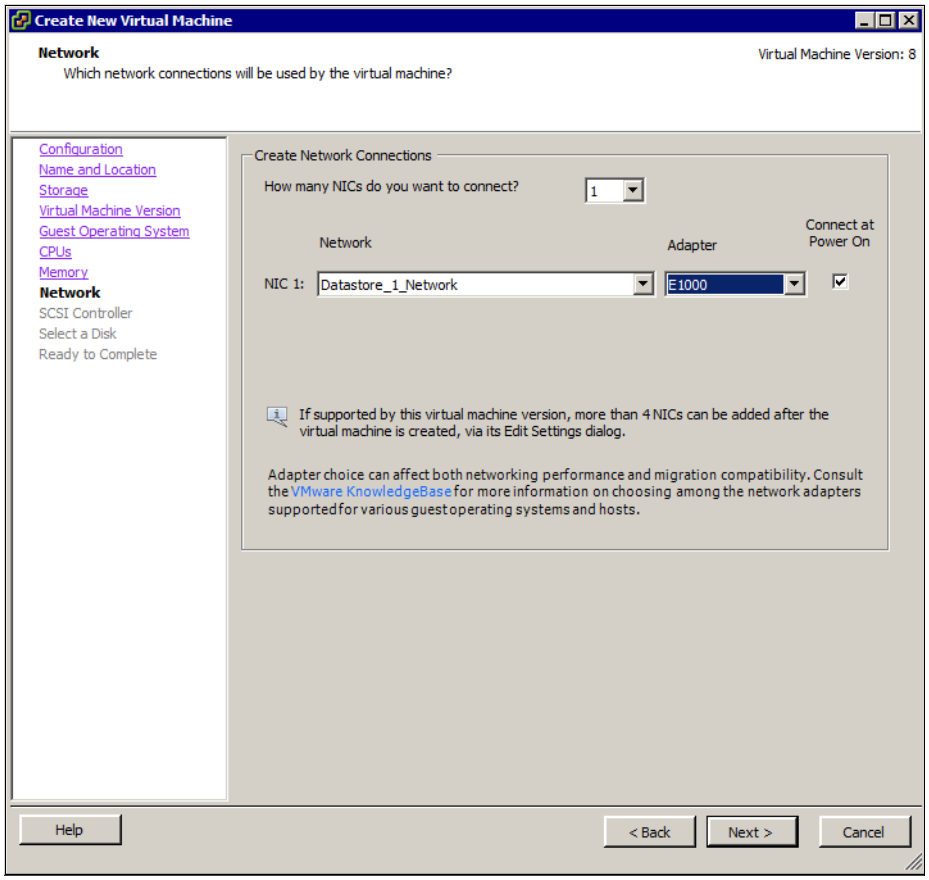

Figure 5-108 Network selection

11. On the SCSI Controller types panel, Figure 5-109 on page 136, select the controller that is based on the OS requirement. In our example, we select the LSI Logic SAS SCSI controller. Click **Next**.

    For additional information about the types of available SCSI controllers, see the *VMware Administration Guide* and the *Guest Operating System Installation Guide*:

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

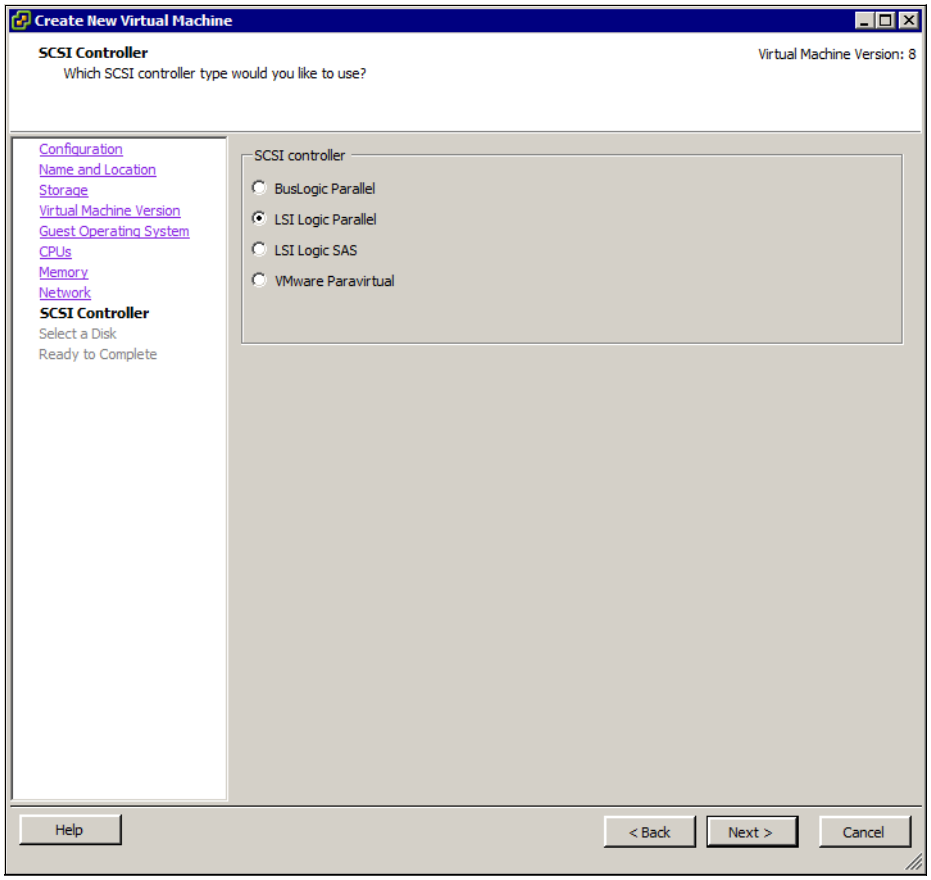

Figure 5-109 VM SCSI Controller type

12. On the Select a Disk panel, Figure 5-110 on page 137, select one of these options:
    - Create a new virtual disk: Use this option if there is no existing disk.
    - Use an existing virtual disk: If you are connecting the guest to a previously built `.vmdk` file.
    - Raw Device Mappings: Direct access to Fibre Channel SAN disks.

    Click **Next** as shown in Figure 5-110 on page 137.

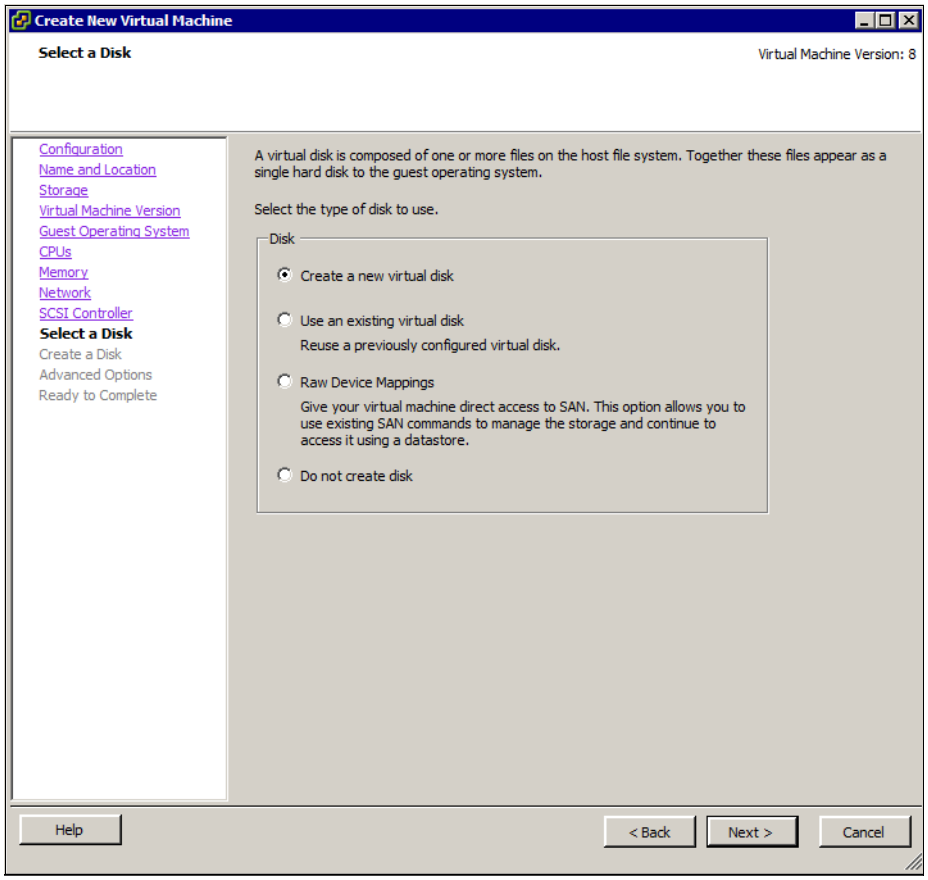

Figure 5-110 Select a Disk

13. If you selected the “Create a new virtual disk” option, allocate the Disk Size. This size is the size of the `.vmdk` file that represents the hardware disk in the virtual machine’s configuration. See Figure 5-111 on page 138. Click **Next**.

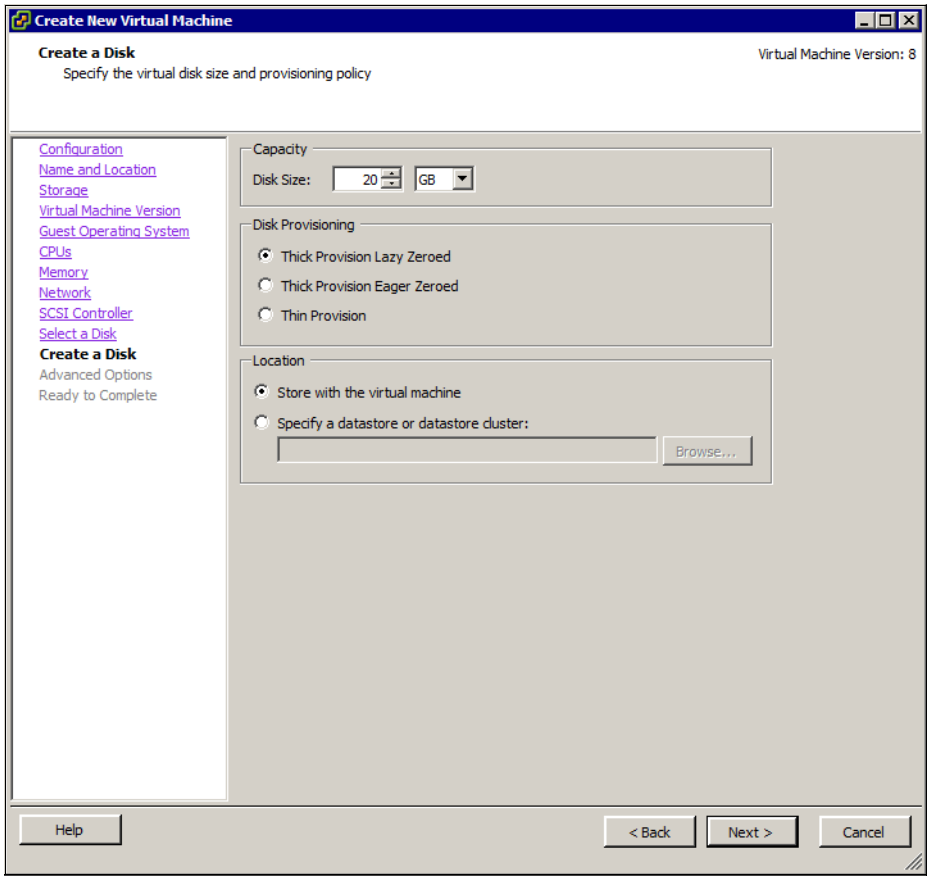

Figure 5-111 Create a Disk

14. On the Advanced Options panel, we continue with the default options and click **Next** as shown in Figure 5-112 on page 139.

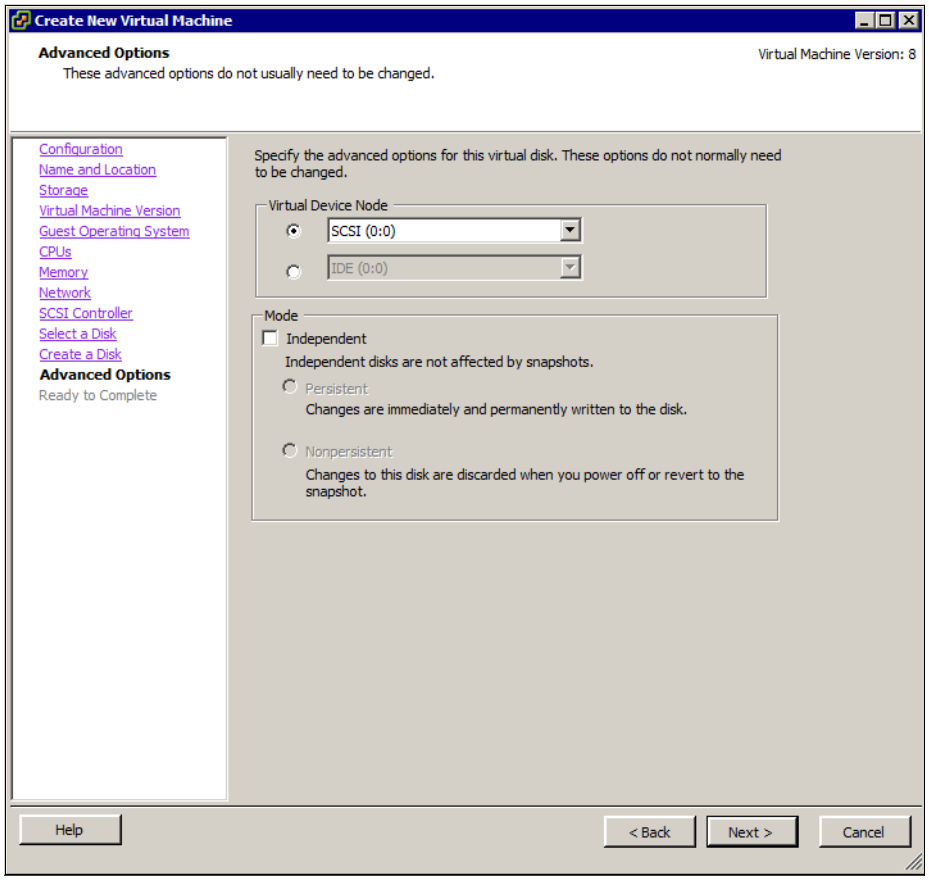

Figure 5-112 Advanced Options

15. On the Ready to Complete summary panel, Figure 5-113 on page 140, click **Finish**. You can see the progress in the Recent Tasks pane at the bottom of the vCenter GUI.

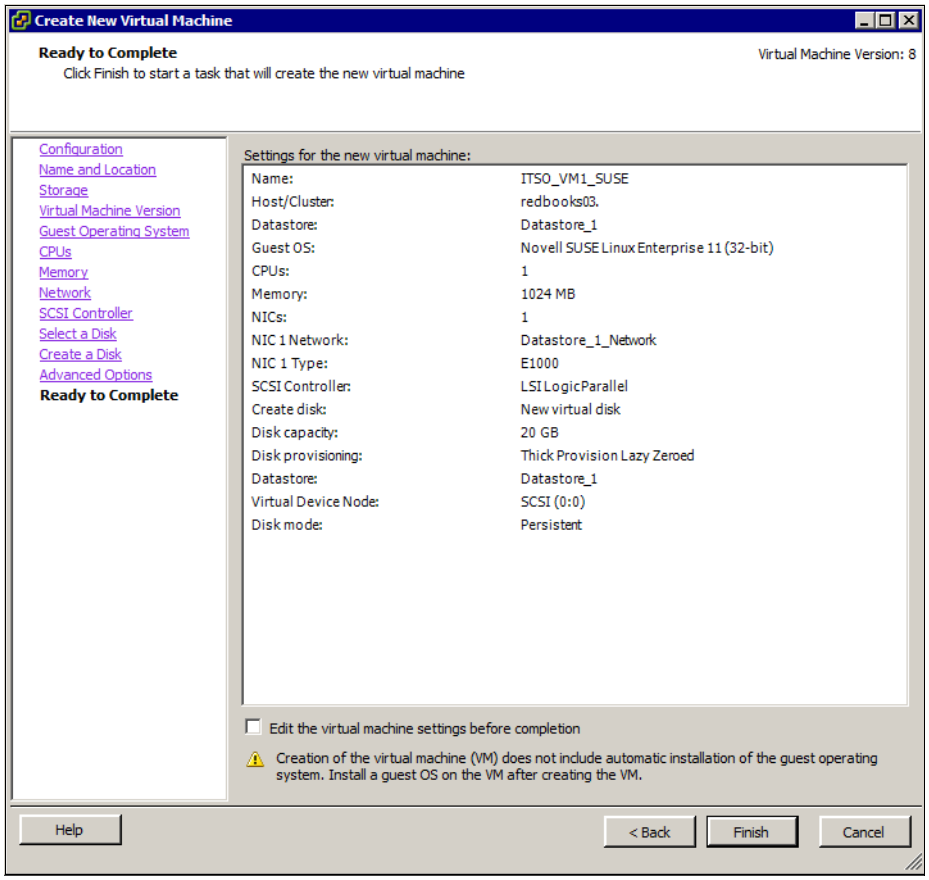

Figure 5-113 Summary panel

16. You are now ready to perform the Guest Operating System installation.

5.2.10 Additional VMware ESX Server Storage configuration

This section shows how to set the VMware ESXi Server Advanced options. They are recommended to maintain the normal operation of the VMware ESXi Server with IBM Midrange Storage Subsystem and to help with troubleshooting if necessary. Follow these steps:

1. Use the VMware vSphere Client to connect to the VMware ESXi Server (log in as `root`).
2. Click the **Configuration** tab and select **Advanced Settings** in the Software section as shown in Figure 5-114 on page 141.

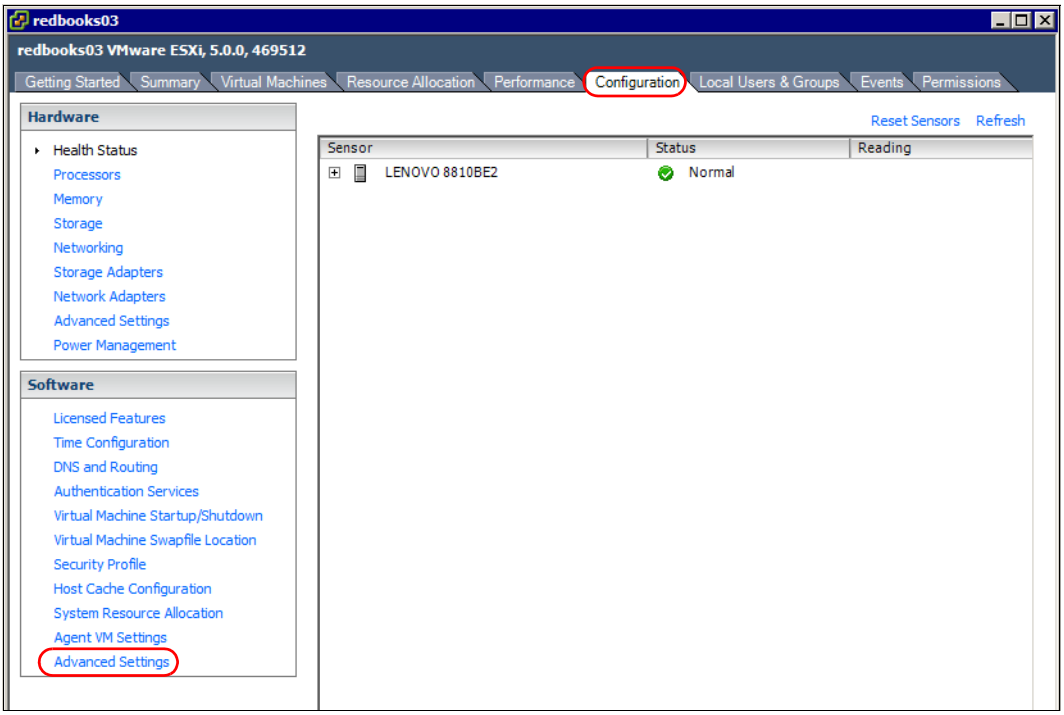

Figure 5-114 vSphere Configuration panel

3. Click **Disk** in the left section and set these options, as shown in Figure 5-115:
   - Disk.UseDeviceReset = 0
   - Disk.UseLunReset = 1

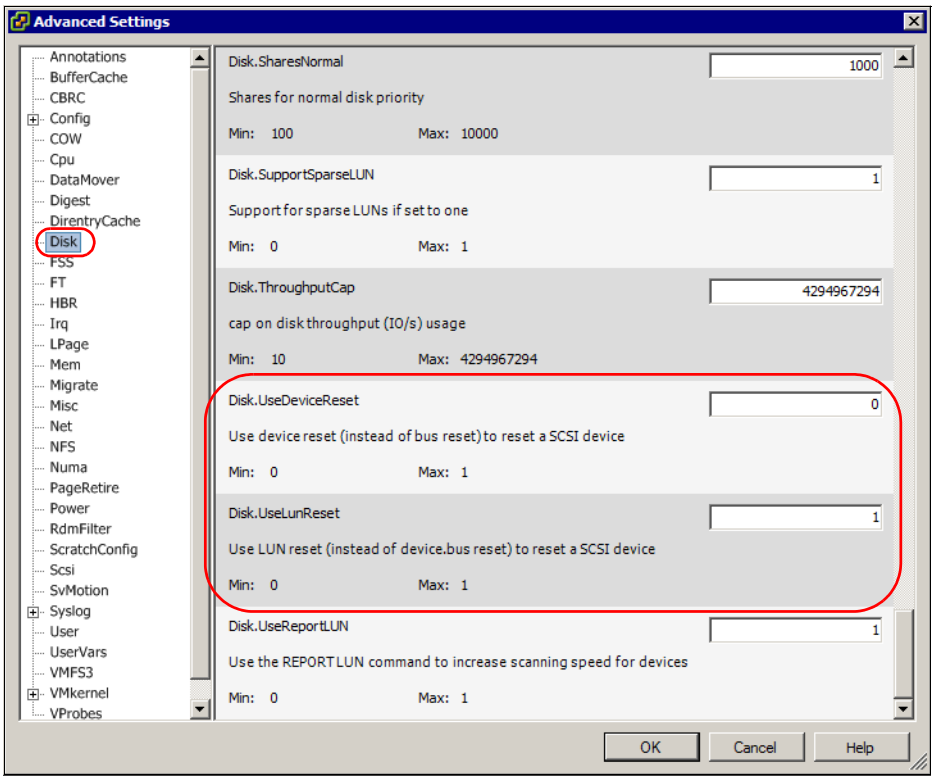

Figure 5-115 VMware ESXi Advanced Settings

4. Enable logging on VMware ESXi Server by enabling these options (if logging is not on by default), as shown in Figure 5-116, and click **OK**:
   - Scsi.LogCmdErrors = 1
   - Scsi.LogMPCmdErrors = 1

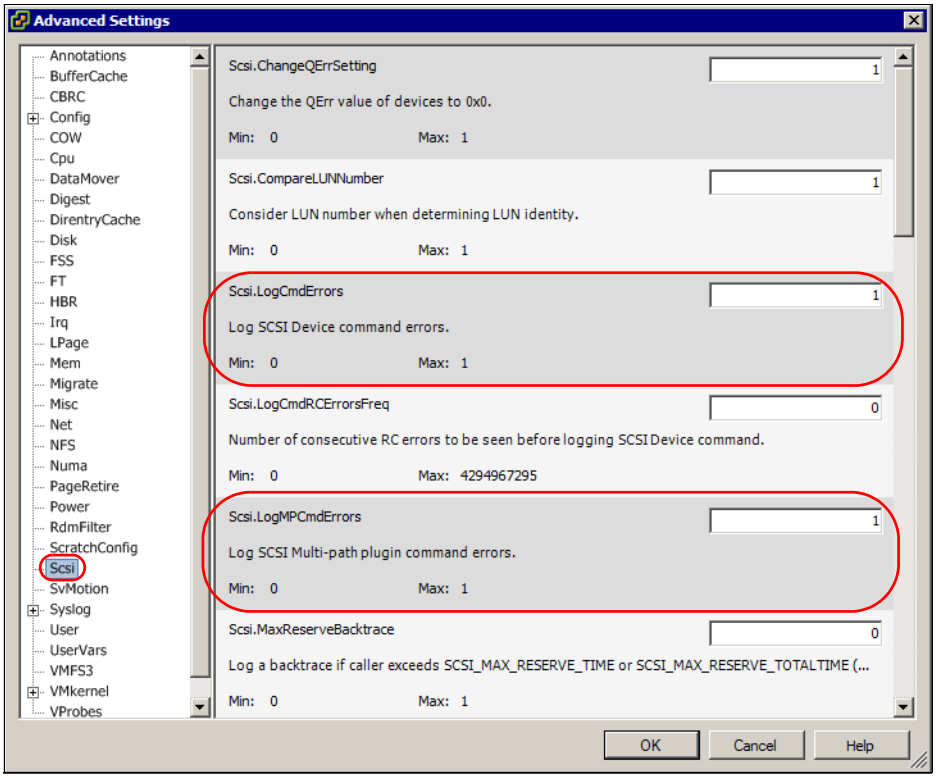

Figure 5-116 VMware ESXi logging settings

6

VMware command-line tools for configuring vSphere ESXi storage

In this chapter, we outline the process to configure SAN storage by using the iSCSI software initiator and Fibre Channel protocol. We also describe the necessary settings to connect to DS5000 storage subsystems.

6.1 Introduction to command-line tools

vSphere supports several command-line interfaces for managing your virtual infrastructure:

- vSphere command-line interface (vCLI)
- ESXi Shell commands (**`esxcli`** - **`vicfg`**)
- PowerCLI

We use ESXi Shell commands to configure iSCSI and FC SAN storage.

6.1.1 Enabling ESXi Shell from DCUI

The ESXi Shell commands feature is natively included in the local support consoles, but this feature is not enabled by default.

Follow these steps from the Direct Console User Interface (DCUI):

1. At the direct console of the ESXi host, press F2 and provide credentials when prompted.
2. Scroll to **Troubleshooting Mode Options** and press Enter.
3. Choose **Enable ESXi Shell** and press Enter.
4. The "`ESXi Shell is Enabled`" message is displayed on the right side of the window, as shown in Figure 6-1.

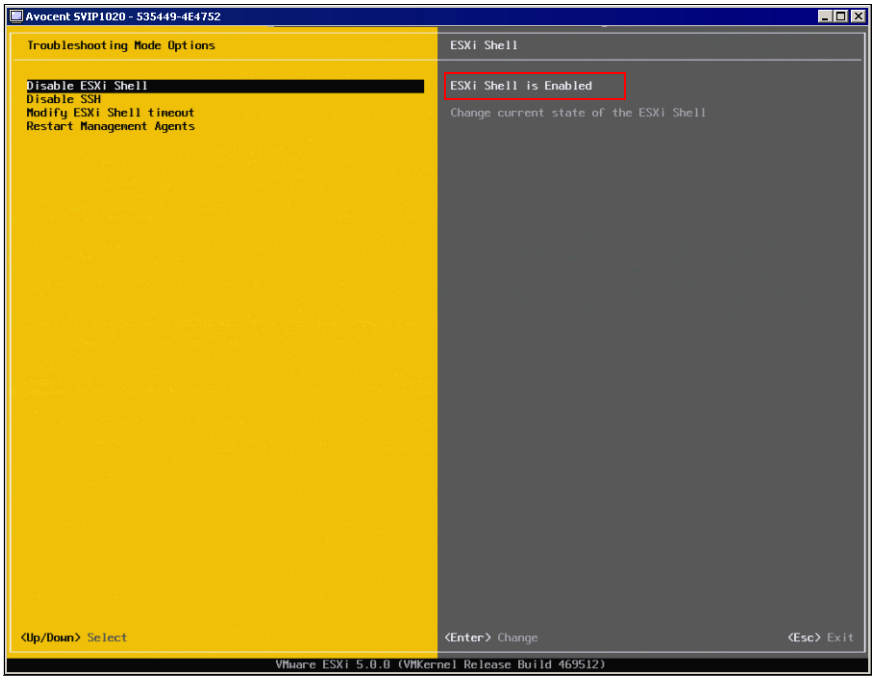

Figure 6-1 Enabling ESXi Shell from the DCUI

5. Press Esc until you return to the main direct console panel. Click **Saving the Configuration Changes**.

6.1.2 Enabling ESXi Shell with the vSphere Client

Follow these steps to enable ESXi Shell with the vSphere Client:

1. Log in to a vCenter Server system by using the vSphere Client.
2. Select the host in the inventory panel.
3. Click the **Configuration** tab and click **Security Profile**.
4. In the Services section, click **Properties**.
5. Select **ESXi Shell** from this list as shown in Figure 6-2.

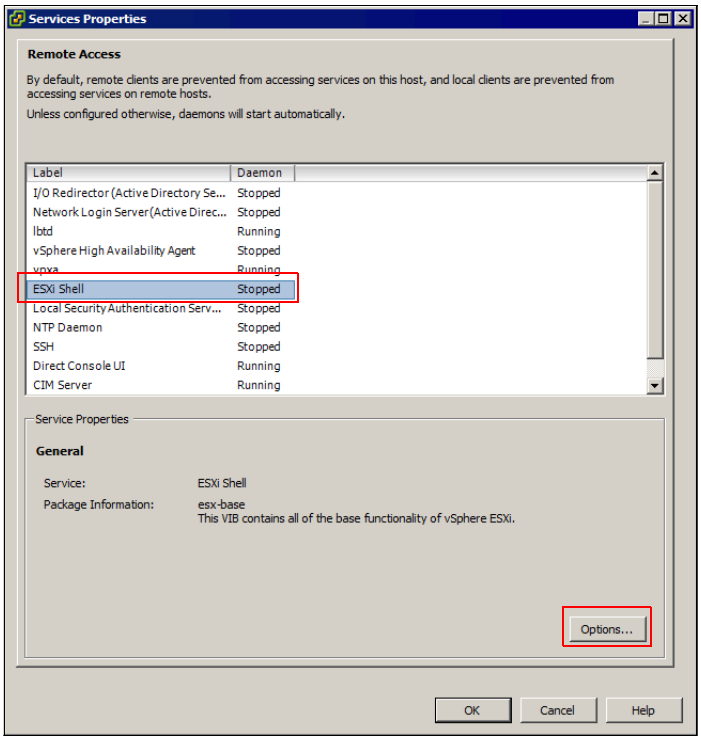

Figure 6-2 Checking Services from vSphere Client

6. Click **Options** to open the ESXi Shell Options window.
7. From the ESXi Shell Options window, select the required Startup Policy. Click **Start** to enable the service as shown in Figure 6-3 on page 146.

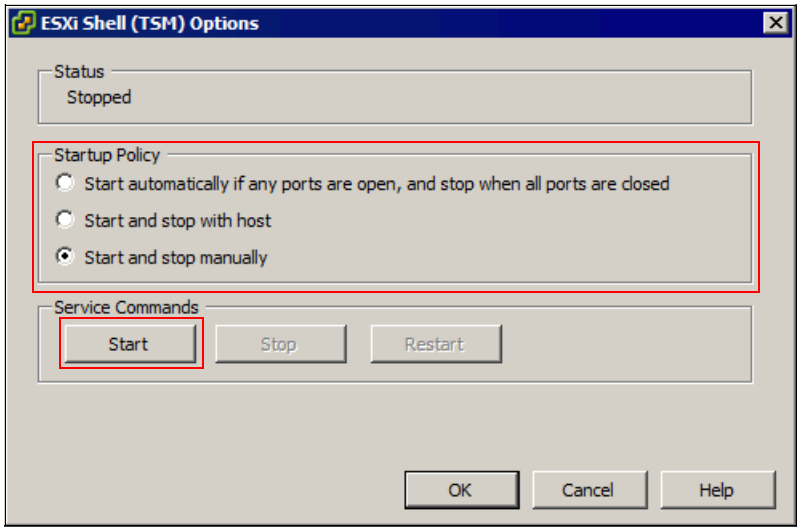

Figure 6-3 Starting services

8. Repeat steps 5 - 7 to enable the Secure Shell (SSH) service.

6.1.3 Running ESXi Shell commands

vSphere ESXi supports the execution of ESXi Shell commands in different ways:

- Locally executed from the DCUI console
- Remotely executed by using SSH through the local support console
- Remotely using the vMA appliance
- Remotely using the vSphere CLI

For this example, we run ESXi Shell commands remotely by using vSphere CLI. We can install the vSphere CLI command set on a supported Linux or Microsoft Windows system. The installation package and deployment procedure are available at the following link:

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

The vSphere CLI command is, by default, available by clicking **Start** → **Programs** → **VMware** → **VMware vSphere CLI**.

The basic usage is formatted this way:

```
esxcli --server <vc_server> --username <privileged_user> --password <pw> --vihost
<esx<namespace> [<namespace]...> <command> --<option_name=option_value>
```

Later, we describe the basic command-line syntax. For more details about available ESXi Shell commands, see the main reference support document:

http://pubs.vmware.com/vsphere-50/topic/com.vmware.vcli.getstart.doc_50/cli_about.html

6.1.4 Saving time running ESXi Shell commands

To avoid the redundancy of adding the connection information on the command line, you can create a connection document to use every time that you run a command.

The following example illustrates the contents of the configuration file that we have saved as `esxcli.config`:

```
VI_SERVER = XX.XXX.XXX.XX
VI_USERNAME = root
VI_PASSWORD = my_password
VI_PROTOCOL = https
VI_PORTNUMBER = 443
```

Replace this information with your environment data for a useful tool for running ESXi Shell commands.

Note: Save the configuration file in the same location/path of your ESXi Shell command to avoid syntax errors. The following locations are the default locations for the ESXi Shell command in Windows OS:

- 32-Bit OS:

  `C:\Program Files\VMware\VMware vSphere CLI\`

- 64-Bit OS:

  `C:\Program Files (x86)\VMware\VMware vSphere CLI\`

VMware provides many resources and an active user community forum at the ESXi Shell main page:

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

6.2 Connecting to SAN storage by using iSCSI

You can attach the DS Storage Systems to your hosts by using iSCSI interfaces. We show how to configure your vSphere ESXi hosts to use a regular Ethernet network interface card (NIC) and the native software iSCSI Initiator to connect to a DS5300 system with iSCSI host interface cards (HICs).

Our implementation example uses vSphere ESXi 5.0, two Ethernet network cards connected to different Ethernet switches. The traffic is isolated on a dedicated private network where the DS5300 iSCSI controllers reside.

The DS Storage System iSCSI ports are defined in the following way:

- `192.168.130.101` - iSCSI Controller A
- `192.168.130.102` - iSCSI Controller B

The following procedure explains how to connect to your storage by using iSCSI. A software iSCSI adapter is part of the VMware code.

Configuring the iSCSI software initiator takes several steps:

1. Activate the software iSCSI adapter.
2. Configure networking for iSCSI.
3. Configure iSCSI discovery addresses.
4. Enable security (CHAP).

6.2.1 Activating the software iSCSI adapter

To activate the software iSCSI adapter, click **Start** → **Programs** → **VMware** → **VMware vSphere CLI** → **Command Prompt** and enter the following commands:

1. Enable the iSCSI software initiator:

   ```
   esxcli --config esxcli.config iscsi software set --enabled=true
   ```

2. Check the iSCSI software initiator status:

   ```
   esxcli --config esxcli.config iscsi software get
   ```

> **Note:** The system prints correctly (true) if software iSCSI is enabled, or false if it is not enabled.

Now that the iSCSI software initiator is enabled on your system, you can obtain the iSCSI host bus adapter (HBA) name and its iSCSI qualified name (IQN).

To discover the available adapters and get the iSCSI IQN name, run the command that is shown in Example 6-1.

Example 6-1 Discovering available adapters

```
C:\Program Files\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage core
adapter list
HBA Name  Driver     Link State  UID                                       Description
--------  ---------  ----------  ----------------------------------------
--------------------------------------------------------------------------
vmhba0    ata_piix   link-n/a    sata.vmhba0                               (0:0:31.2)
Intel Corporation 82801H (ICH8 Family) 4 port SATA IDE Controller
vmhba1    ata_piix   link-n/a    sata.vmhba1                               (0:0:31.5)
Intel Corporation 82801H (ICH8 Family) 2 port SATA IDE Controller
vmhba32   ata_piix   link-n/a    sata.vmhba32                              (0:0:31.2)
Intel Corporation 82801H (ICH8 Family) 4 port SATA IDE Controller
vmhba33   ata_piix   link-n/a    sata.vmhba33                              (0:0:31.5)
Intel Corporation 82801H (ICH8 Family) 2 port SATA IDE Controller
vmhba34   iscsi_vmk  online      iqn.1998-01.com.vmware:redbooks03-5147ed14  iSCSI Software
Adapter
```

6.2.2 Configuring networking for iSCSI

We use two network adapters for iSCSI connection to the storage subsystem. They need to be added to a separate virtual switch, and they both need to be assigned a separate IP address. This procedure shows the necessary steps:

From the menu, click **Start** → **Programs** → **VMware** → **VMware vSphere CLI** → **Command Prompt**, and enter the following commands:

1. Create a Virtual Standard Switch (VSS) named `vSwitch_iSCSI`:

   ```
   esxcli --config esxcli.config network vswitch standard add
   -vswitch-name=vSwitch_iSCSI
   ```

2. Add a portgroup to my standard vswitch `vSwitch_iSCSI`:

   ```
   esxcli --config esxcli.config network vswitch standard portgroup add -p iSCSI-1
   -v vSwitch_iSCSI
   ```

3. Add a secondary portgroup to `vSwitch_iSCSI`:

   ```
   esxcli --config esxcli.config network vswitch standard portgroup add -p iSCSI-2
   -v vSwitch_iSCSI
   ```

After we create the virtual switch and add the portgroups, the next step is to configure the portgroups adding VMkernel interfaces.

Note: In this example, we assume that you have one VMkernel interface that is created for the use of the vSwitch0 → Management Network (`vmk0`). We add two more VMkernel ports using `vmk1` and `vmk2` as default names.

4. Add a VMkernel interface (`vmk1`) to the `iSCSI-1` portgroup:

```
esxcli --config esxcli.config network ip interface add -i vmk1 -p iSCSI-1
```

5. Repeat the process to add a VMkernel interface (`vmk2`) to the `iSCSI-2` portgroup:

```
esxcli --config esxcli.config network ip interface add -i vmk2 -p iSCSI-2
```

The following lines cover the network configuration of the recently created VMkernel ports `vmk1` and `vmk2`. The IP addresses to be used need to be in the same network/VLAN where you configured in your DS Subsystem Storage iSCSI adapters. Follow these steps:

1. Set the static IP addresses on both VMkernel NICs as part of the iSCSI network:

```
esxcli --config esxcli.config network ip interface ipv4 set -i vmk1 -I
192.168.130.50 -N 255.255.255.0 -t static
```

2. Repeat the process to configure the secondary VMkernel interface `vmk2`:

```
esxcli --config esxcli.config network ip interface ipv4 set -i vmk2 -I
192.168.130.51 -N 255.255.255.0 -t static
```

Now, add uplinks to our `vSwitch_iSCSI` virtual switch:

1. Add a primary uplink adapter:

```
esxcli --config esxcli.config network vswitch standard uplink add
--uplink-name=vmnic1 --vswitch-name=vSwitch_iSCSI
```

2. Repeat the process to add a secondary uplink adapter:

```
esxcli --config esxcli.config network vswitch standard uplink add
--uplink-name=vmnic2 --vswitch-name=vSwitch_iSCSI
```

Note: To check the available vmnics, use the following command line:

```
esxcli --config esxcli.config network nic list
```

Set the manual override failover policy so that each iSCSI VMkernel portgroup has one active physical vmnic and one vmnic that is configured as “unused”:

1. Change the default failover policy for the `iSCSI-1` portgroup:

```
esxcli --config esxcli.config network vswitch standard portgroup policy
failover set -p iSCSI-1 -a vmnic1 -u vmnic2
```

2. Repeat the process for changing the default failover policy for the `iSCSI-2` portgroup:

```
esxcli --config esxcli.config network vswitch standard portgroup policy
failover set -p iSCSI-2 -a vmnic2 -u vmnic1
```

3. Configure the policy failover at the virtual switch level:

```
esxcli --config esxcli.config network vswitch standard policy failover set -v
vSwitch_iSCSI -a vmnic1,vmnic2
```

At this point, we have created a virtual switch (vSwitch). To check the vSwitch configuration parameters, execute the following command as shown in Example 6-2 on page 150.

Example 6-2 Checking virtual switch configuration parameters

```
C:\Program Files\VMware\VMware vSphere CLI>esxcli --config esxcli.config network
 vswitch standard list -v vSwitch_iSCSI
vSwitch_iSCSI
   Name: vSwitch_iSCSI
   Class: etherswitch
   Num Ports: 128
   Used Ports: 5
   Configured Ports: 128
   MTU: 1500
   CDP Status: listen
   Beacon Enabled: false
   Beacon Interval: 1
   Beacon Threshold: 3
   Beacon Required By:
   Uplinks: vmnic2, vmnic1
   Portgroups: iSCSI-2, iSCSI-1
```

6.2.3 Configuring iSCSI discovery addresses

Before we proceed with the discovery process, we configure the iSCSI initiator by adding the `vmk1` and `vmk2` ports as *binding ports*:

1. Bind each of the VMkernel NICs to the software iSCSI HBA:

   ```
   esxcli --config esxcli.config iscsi networkportal add -A vmhba34 -n vmk1
   esxcli --config esxcli.config iscsi networkportal add -A vmhba34 -n vmk2
   ```

2. Now, we discover the targets by using the IP addresses of the IBM DS Storage Subsystems. We have two iSCSI interfaces on the DS5300 that use the `192.168.130.101` and `192.168.130.102` IP addresses.

   Add the IP address of your iSCSI array or SAN as a dynamic discovery:

   ```
   esxcli --config esxcli.config iscsi adapter discovery sendtarget add -A vmhba34
   -a 192.168.130.101
   ```

3. Repeat the process for the secondary iSCSI array IP address:

   ```
   esxcli --config esxcli.config iscsi adapter discovery sendtarget add -A vmhba34
   -a 192.168.130.102
   ```

4. Rescan your software iSCSI HBA to discover volumes and Volume Manager File Systems (VMFS) datastores:

   ```
   esxcli --config esxcli-config storage core adapter rescan --adapter vmhba34
   ```

5. To list the available filesystem, run the following command as shown in Example 6-3.

Example 6-3 Listing the available storage file system from the command line

```
C:\Program Files\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage filesystem
list
Mount Point                                        Volume Name  UUID
Mounted  Type            Size         Free
-------------------------------------------------  -----------
-----------------------------------  -------  ------  ------------  -----------
/vmfs/volumes/4e9ddd95-696fcc42-fa76-0014d126e786  datastore1
4e9ddd95-696fcc42-fa76-0014d126e786     true  VMFS-5   74625056768  73606889472
/vmfs/volumes/4e9f531f-78b18f6e-7583-001641edb4dd  Datastore_2
4e9f531f-78b18f6e-7583-001641edb4dd     true  VMFS-5  107105746944  63313018880
```

```
/vmfs/volumes/4ea20b1e-6cf76340-4250-001641edb4dd  Datastore_1
4ea20b1e-6cf76340-4250-001641edb4dd     true  VMFS-5  107105746944  99139715072
/vmfs/volumes/4e9ddd95-f1327d50-b7fc-0014d126e786
4e9ddd95-f1327d50-b7fc-0014d126e786     true  vfat      4293591040   4280156160
/vmfs/volumes/b0b41f71-1bc96828-21df-6548ab457c03
b0b41f71-1bc96828-21df-6548ab457c03     true  vfat       261853184    128225280
/vmfs/volumes/1f1e5f79-ce9138bf-c62c-3893b933397e
1f1e5f79-ce9138bf-c62c-3893b933397e     true  vfat       261853184    261844992
/vmfs/volumes/4e9ddd8d-b69852dc-3d8b-0014d126e786
4e9ddd8d-b69852dc-3d8b-0014d126e786     true  vfat       299712512    114974720
```

6.2.4 Enabling security (CHAP)

The best practice CHAP security configuration is recommended. To enable basic CHAP authentication, run the following command:

```
esxcli --config esxcli.config iscsi adapter auth chap set --adapter vmhba34
--authname iqn.1998-01.com.vmware:redbook s03-5147ed14 --direction uni --level
preferred --secret ITSO2011_Secured
```

Security recommendations: Use strong passwords for all accounts. Use CHAP authentication because it ensures that each host has its own password. Mutual CHAP authentication is also recommended.

For more information: We assume that your DS Storage System is already configured for using CHAP authentication. For more information about iSCSI configuration at the DS Storage System level, see *IBM System Storage DS5000 Series Implementation and Best Practices Guide*, SG24-8024.

6.3 Connecting to SAN storage by using Fibre Channel (FC)

Unlike iSCSI, FC configuration is relatively simple. In the next example, we use two HBAs that are connected to different SAN fabric switches. We have our own zone defined on both fabric switches to separate the traffic for stability and improve the management. The IBM DS5300 has two controllers that are defined as Controller A and Controller B, and both controllers are also physically connected to different SAN fabric switches. Based on the VMware Native Multipathing Plug-in (NMP) or failover driver that is implemented at the ESXi level (natively provided by hypervisor) and the proposed cabling connections, the vSphere ESXi host can access the SAN-attached storage by using alternatives paths for redundancy. The Most Recent Used (MRU) policy is the recommended path policy.

As shown in Example 6-4, two HBAs cards are physically installed in our vSphere ESXi hosts.

Example 6-4 Discovering available adapters

```
C:\Program Files\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage core
adapter list
HBA Name  Driver    Link State  UID                                   Description
--------  --------  ----------  ------------------------------------
--------------------------------------------------------------------------------
vmhba0    ata_piix  link-n/a    sata.vmhba0                           (0:0:31.2) Intel
Corporation 82801H (ICH8 Family) 4 port SATA IDE Controller
```

```
vmhba1    ata_piix  link-n/a    sata.vmhba1                          (0:0:31.5) Intel
Corporation 82801H (ICH8 Family) 2 port SATA IDE Controller
vmhba2    qla2xxx   link-n/a    fc.200000e08b892cc0:210000e08b892cc0  (0:10:9.0) QLogic
Corp QLA2340-Single Channel 2Gb Fibre Channel to PCI-X HBA
vmhba3    qla2xxx   link-n/a    fc.200000e08b18208b:210000e08b18208b  (0:10:10.0) QLogic
Corp QLA2340-Single Channel 2Gb Fibre Channel to PCI-X HBA
vmhba32   ata_piix  link-n/a    sata.vmhba32                         (0:0:31.2) Intel
Corporation 82801H (ICH8 Family) 4 port SATA IDE Controller
vmhba33   ata_piix  link-n/a    sata.vmhba33                         (0:0:31.5) Intel
Corporation 82801H (ICH8 Family) 2 port SATA IDE Controller
```

The following steps show the basic SAN storage tasks by using Fibre Channel (FC). In Example 6-5, we show the SAN-attached disks and their configuration.

From the menu, click **Start** → **Programs** → **VMware** → **VMware vSphere CLI**. At the command prompt, enter the following commands:

- List all devices with their corresponding paths, state of the path, adapter type, and other information:

  ```
  esxcli --config esxcli.config storage core path list
  ```

- Limit the display to only a specified path or device:

  ```
  esxcli --config esxcli.config  storage core path list --device vmhba2
  ```

- List detailed information for the paths for the device specified with `--device`.

  ```
  esxcli --config esxcli.config storage core path list -d <naa.xxxxxx>
  ```

- Rescan all adapters:

  ```
  esxcli --config esxcli.config storage core adapter rescan
  ```

Example 6-5 Showing discovered FC SAN attach through the command line

```
C:\Program Files\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage
 core device list
naa.600a0b80006e32a000001e764e9d9e1d
   Display Name: IBM Fibre Channel Disk (naa.600a0b80006e32a000001e764e9d9e1d)
   Has Settable Display Name: true
   Size: 102400
   Device Type: Direct-Access
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/disks/naa.600a0b80006e32a000001e764e9d9e1d
   Vendor: IBM
   Model: 1818      FAStT
   Revision: 0730
   SCSI Level: 5
   Is Pseudo: false
   Status: on
   Is RDM Capable: true
   Is Local: false
   Is Removable: false
   Is SSD: true
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
   Attached Filters:
   VAAI Status: unknown
   Other UIDs: vml.0200000000600a0b80006e32a000001e764e9d9e1d313831382020

naa.600a0b80006e32020000fe594ea59de0
   Display Name: IBM iSCSI Disk (naa.600a0b80006e32020000fe594ea59de0)
```

```
   Has Settable Display Name: true
   Size: 20480
   Device Type: Direct-Access
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/disks/naa.600a0b80006e32020000fe594ea59de0
   Vendor: IBM
   Model: 1818      FASt T
   Revision: 0730
   SCSI Level: 5
   Is Pseudo: false
   Status: on
   Is RDM Capable: true
   Is Local: false
   Is Removable: false
   Is SSD: false
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
   Attached Filters:
   VAAI Status: unknown
   Other UIDs: vml.0200020000600a0b80006e32020000fe594ea59de0313831382020

t10.ATA_____WDC_WD800JD2D08MSA1_____WD2DWMAM9ZY50888
   Display Name: Local ATA Disk (t10.ATA_____WDC_WD800JD2D08MSA1_____
__________WD2DWMAM9ZY50888)
   Has Settable Display Name: true
   Size: 76324
   Device Type: Direct-Access
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/disks/t10.ATA_____WDC_WD800JD2D08MSA1_____
____________WD2DWMAM9ZY50888
   Vendor: ATA
   Model: WDC WD800JD-08MS
   Revision: 10.0
   SCSI Level: 5
   Is Pseudo: false
   Status: on
   Is RDM Capable: false
   Is Local: true
   Is Removable: false
   Is SSD: false
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
   Attached Filters:
   VAAI Status: unknown
   Other UIDs: vml.010000000020202020202057442d574d414d395a5935303838385744432057 4

mpx.vmhba32:C0:T0:L0
   Display Name: Local HL-DT-ST CD-ROM (mpx.vmhba32:C0:T0:L0)
   Has Settable Display Name: false
   Size: 0
   Device Type: CD-ROM
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/cdrom/mpx.vmhba32:C0:T0:L0
   Vendor: HL-DT-ST
   Model: CDRW/DVD GCCH10N
   Revision: C103
   SCSI Level: 5
```

```
   Is Pseudo: false
   Status: on
   Is RDM Capable: false
   Is Local: true
   Is Removable: true
   Is SSD: false
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
   Attached Filters:
   VAAI Status: unsupported
   Other UIDs: vml.0005000000766d68626133323a303a30

naa.600a0b80006e32a000001e794e9d9e32
   Display Name: IBM Fibre Channel Disk (naa.600a0b80006e32a000001e794e9d9e32)
   Has Settable Display Name: true
   Size: 102400
   Device Type: Direct-Access
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/disks/naa.600a0b80006e32a000001e794e9d9e32
   Vendor: IBM
   Model: 1818      FAStT
   Revision: 0730
   SCSI Level: 5
   Is Pseudo: false
   Status: on
   Is RDM Capable: true
   Is Local: false
   Is Removable: false
   Is SSD: true
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
   Attached Filters:
   VAAI Status: unknown
   Other UIDs: vml.0200010000600a0b80006e32a000001e794e9d9e32313831382020
```

6.4 Managing multipath policies

We describe how to show and modify the multipath policies that are natively supported by the Hypervisor through the Path Selection Plug-Ins (PSPs).

Example 6-6 on page 155 shows two LUNs that are presented to the ESXi host with four paths. Two paths act as *Active* paths and two paths are in *Standby* mode. The multipath policy is set to the Most recently used (MRU) policy.

From the menu, click **Start** → **Programs** → **VMware** → **VMware vSphere CLI**. At the command prompt, enter the following commands:

- List all devices with their corresponding paths, state of the path, adapter type, and other information:

  ```
  esxcli --config esxcli.config storage core path list
  ```

- Limit the display to only a specified path or device:

  ```
  esxcli --config esxcli.config  storage core path list --device vmhba2
  ```

- List detailed information for the paths for the device that is specified with `--device`:

```
esxcli --config esxcli.config storage core path list -d <naa.xxxxxx>
```

Device ID: To limit the Display output, we use the `naa.600a0b80006e16a80000e71950111711` device ID.

Example 6-6 List storage devices

```
C:\Program Files (x86)\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage core
path list -d  naa.600a0b80006e16a80000e71950111711
fc.200000e08b892cc0:210000e08b892cc0-fc.200600a0b86e16a8:202700a0b86e16a8-naa.600a0b80006e1
6a80000e71950111711
   UID:
fc.200000e08b892cc0:210000e08b892cc0-fc.200600a0b86e16a8:202700a0b86e16a8-naa.600a0b80006e1
6a80000e71950111711
   Runtime Name: vmhba2:C0:T3:L0
   Device: naa.600a0b80006e16a80000e71950111711
   Device Display Name: IBM Fibre Channel Disk (naa.600a0b80006e16a80000e71950111711)
   Adapter: vmhba2
   Channel: 0
   Target: 3
   LUN: 0
   Plugin: NMP
   State: standby
   Transport: fc
   Adapter Identifier: fc.200000e08b892cc0:210000e08b892cc0
   Target Identifier: fc.200600a0b86e16a8:202700a0b86e16a8
   Adapter Transport Details: WWNN: 20:00:00:e0:8b:89:2c:c0 WWPN: 21:00:00:e0:8b:89:2c:c0
   Target Transport Details: WWNN: 20:06:00:a0:b8:6e:16:a8 WWPN: 20:27:00:a0:b8:6e:16:a8

fc.200000e08b892cc0:210000e08b892cc0-fc.200600a0b86e16a8:201700a0b86e16a8-naa.600a0b80006e1
6a80000e71950111711
   UID:
fc.200000e08b892cc0:210000e08b892cc0-fc.200600a0b86e16a8:201700a0b86e16a8-naa.600a0b80006e1
6a80000e71950111711
   Runtime Name: vmhba2:C0:T2:L0
   Device: naa.600a0b80006e16a80000e71950111711
   Device Display Name: IBM Fibre Channel Disk (naa.600a0b80006e16a80000e71950111711)
   Adapter: vmhba2
   Channel: 0
   Target: 2
   LUN: 0
   Plugin: NMP
   State: standby
   Transport: fc
   Adapter Identifier: fc.200000e08b892cc0:210000e08b892cc0
   Target Identifier: fc.200600a0b86e16a8:201700a0b86e16a8
   Adapter Transport Details: WWNN: 20:00:00:e0:8b:89:2c:c0 WWPN: 21:00:00:e0:8b:89:2c:c0
   Target Transport Details: WWNN: 20:06:00:a0:b8:6e:16:a8 WWPN: 20:17:00:a0:b8:6e:16:a8

fc.200000e08b892cc0:210000e08b892cc0-fc.200600a0b86e16a8:202600a0b86e16a8-naa.600a0b80006e1
6a80000e71950111711
   UID:
fc.200000e08b892cc0:210000e08b892cc0-fc.200600a0b86e16a8:202600a0b86e16a8-naa.600a0b80006e1
6a80000e71950111711
   Runtime Name: vmhba2:C0:T1:L0
   Device: naa.600a0b80006e16a80000e71950111711
   Device Display Name: IBM Fibre Channel Disk (naa.600a0b80006e16a80000e71950111711)
   Adapter: vmhba2
   Channel: 0
```

```
   Target: 1
   LUN: 0
   Plugin: NMP
   State: active
   Transport: fc
   Adapter Identifier: fc.200000e08b892cc0:210000e08b892cc0
   Target Identifier: fc.200600a0b86e16a8:202600a0b86e16a8
   Adapter Transport Details: WWNN: 20:00:00:e0:8b:89:2c:c0 WWPN: 21:00:00:e0:8b:89:2c:c0
   Target Transport Details: WWNN: 20:06:00:a0:b8:6e:16:a8 WWPN: 20:26:00:a0:b8:6e:16:a8

fc.200000e08b892cc0:210000e08b892cc0-fc.200600a0b86e16a8:201600a0b86e16a8-naa.600a0b80006e1
6a80000e71950111711
   UID:
fc.200000e08b892cc0:210000e08b892cc0-fc.200600a0b86e16a8:201600a0b86e16a8-naa.600a0b80006e1
6a80000e71950111711
   Runtime Name: vmhba2:C0:T0:L0
   Device: naa.600a0b80006e16a80000e71950111711
   Device Display Name: IBM Fibre Channel Disk (naa.600a0b80006e16a80000e71950111711)
   Adapter: vmhba2
   Channel: 0
   Target: 0
   LUN: 0
   Plugin: NMP
   State: active
   Transport: fc
   Adapter Identifier: fc.200000e08b892cc0:210000e08b892cc0
   Target Identifier: fc.200600a0b86e16a8:201600a0b86e16a8
   Adapter Transport Details: WWNN: 20:00:00:e0:8b:89:2c:c0 WWPN: 21:00:00:e0:8b:89:2c:c0
   Target Transport Details: WWNN: 20:06:00:a0:b8:6e:16:a8 WWPN: 20:16:00:a0:b8:6e:16:a8
```

List the detailed information for the paths for the device that is specified with `--device` as shown in Example 6-7.

Example 6-7 List the detailed information for the paths

```
C:\Program Files (x86)\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage nmp
device list -d naa.600a0b80006e16a80000e71950111711
naa.600a0b80006e16a80000e71950111711
   Device Display Name: IBM Fibre Channel Disk (naa.600a0b80006e16a80000e71950111711)
   Storage Array Type: VMW_SATP_LSI
   Storage Array Type Device Config: SATP VMW_SATP_LSI does not support device
configuration.
   Path Selection Policy: VMW_PSP_MRU
   Path Selection Policy Device Config: Current Path=vmhba2:C0:T1:L0
   Path Selection Policy Device Custom Config:
   Working Paths: vmhba2:C0:T1:L0
```

List the path selection policies that are available on the system. Check the values that are valid for the `--psp` option as shown in Example 6-8.

Example 6-8 List the path selection policies available

```
C:\Program Files (x86)\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage core
plugin registration list --plugin-class="PSP"
Module Name    Plugin Name    Plugin Class  Dependencies  Full Path
-------------  -------------  ------------  ------------  ---------
vmw_psp_lib    None           PSP
vmw_psp_mru    VMW_PSP_MRU    PSP           vmw_psp_lib
vmw_psp_rr     VMW_PSP_RR     PSP           vmw_psp_lib
```

```
vmw_psp_fixed  VMW_PSP_FIXED  PSP          vmw_psp_lib
```

Set the Round Robin path policy and list detailed information for the paths as shown in Example 6-9.

Example 6-9 Set path policy

```
C:\Program Files (x86)\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage nmp
device set --device naa.600a0b80006e16a80000e71950111711 --psp VMW_PSP_RR

C:\Program Files (x86)\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage nmp
device list -d naa.600a0b80006e16a80000e71950111711
naa.600a0b80006e16a80000e71950111711
   Device Display Name: IBM Fibre Channel Disk (naa.600a0b80006e16a80000e71950111711)
   Storage Array Type: VMW_SATP_LSI
   Storage Array Type Device Config: SATP VMW_SATP_LSI does not support device
configuration.
   Path Selection Policy: VMW_PSP_RR
   Path Selection Policy Device Config:
{policy=rr,iops=1000,bytes=10485760,useANO=0;lastPathIndex=1: NumIOsPending=0,nu
mBytesPending=0}
   Path Selection Policy Device Custom Config:
   Working Paths: vmhba2:C0:T0:L0, vmhba2:C0:T1:L0
```

Note: After you complete these modifications, do not forget to set your configuration back by using the MRU preferred path policy.

6.5 Matching DS logical drives with VMware vSphere ESXi devices

After the host is installed and configured, we can identify the SAN attach space that we have assigned. We assume that you already assign some space to your host on the DS Storage System side by using DS Storage Manager. Also, before you start to recognize these volumes on your vSphere ESXi host, ensure that the SAN zoning is set up correctly, if you work in an FC environment, according to your planned configuration. For specific steps to configure SAN FC zoning, see *Implementing an IBM/Brocade SAN with 8 Gbps Directors and Switches*, SG24-6116, and *IBM System Storage DS5000 Series Hardware Guide*, SG24-8023.

For iSCSI attachment, ensure that the network that is used is configured correctly (IP, VLANs, frame size, and so on), and has enough bandwidth to provide storage attachment. You have to analyze and understand the impact of the network into which an iSCSI target is to be deployed before the actual installation and configuration of an IBM DS5000 storage system. See the “iSCSI” sections of the *IBM System Storage DS5000 Series Hardware Guide*, SG24-8023.

First, we must discover the SAN space that is attached to our ESXi host. To obtain this information, run the command that is shown in Example 6-10.

As an example, we use the first discovered device, which is a 100 GB LUN that is attached (LUN ID `60:0a:0b:80:00:6e:32:a0:00:00:1e:76:4e:9d:9e:1d`).

Example 6-10 Matching LUNs on DS Storage Manager

```
C:\Program Files\VMware\VMware vSphere CLI>esxcli --config esxcli.config storage core
device list
naa.600a0b80006e32a000001e764e9d9e1d
```

```
   Display Name: IBM Fibre Channel Disk (naa.600a0b80006e32a000001e764e9d9e1d)
   Has Settable Display Name: true
   Size: 102400
   Device Type: Direct-Access
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/disks/naa.600a0b80006e32a000001e764e9d9e1d
   Vendor: IBM
   Model: 1818      FAStT
   Revision: 0730
   SCSI Level: 5
   Is Pseudo: false
   Status: on
   Is RDM Capable: true
   Is Local: false
   Is Removable: false
   Is SSD: true
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
Dan J Attached Filters:
   VAAI Status: unknown
   Other UIDs: vml.0200000000600a0b80006e32a000001e764e9d9e1d313831382020

naa.600a0b80006e32020000fe594ea59de0
   Display Name: IBM iSCSI Disk (naa.600a0b80006e32020000fe594ea59de0)
   Has Settable Display Name: true
   Size: 20480
   Device Type: Direct-Access
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/disks/naa.600a0b80006e32020000fe594ea59de0
   Vendor: IBM
   Model: 1818      FAStT
   Revision: 0730
   SCSI Level: 5
   Is Pseudo: false
   Status: on
   Is RDM Capable: true
   Is Local: false
   Is Removable: false
   Is SSD: false
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
   Attached Filters:
   VAAI Status: unknown
   Other UIDs: vml.0200020000600a0b80006e32020000fe594ea59de0313831382020

t10.ATA_____WDC_WD800JD2D08MSA1_____________________________WD2DWMAM9ZY50888
   Display Name: Local ATA Disk
(t10.ATA_____WDC_WD800JD2D08MSA1_____________________________WD2DWMAM9ZY50888)
   Has Settable Display Name: true
   Size: 76324
   Device Type: Direct-Access
   Multipath Plugin: NMP
   Devfs Path:
/vmfs/devices/disks/t10.ATA_____WDC_WD800JD2D08MSA1_____________________________WD2DWMAM9ZY50
888
   Vendor: ATA
   Model: WDC WD800JD-08MS
   Revision: 10.0
```

```
   SCSI Level: 5
   Is Pseudo: false
   Status: on
   Is RDM Capable: false
   Is Local: true
   Is Removable: false
   Is SSD: false
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
   Attached Filters:
   VAAI Status: unknown
   Other UIDs: vml.0100000000202020202057442d574d414d395a5935303838385744432057 44

mpx.vmhba32:C0:T0:L0
   Display Name: Local HL-DT-ST CD-ROM (mpx.vmhba32:C0:T0:L0)
   Has Settable Display Name: false
   Size: 3020
   Device Type: CD-ROM
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/cdrom/mpx.vmhba32:C0:T0:L0
   Vendor: HL-DT-ST
   Model: CDRW/DVD GCCH10N
   Revision: C103
   SCSI Level: 5
   Is Pseudo: false
   Status: on
   Is RDM Capable: false
   Is Local: true
   Is Removable: true
   Is SSD: false
   Is Offline: false
   Is Perennially Reserved: false
   Thin Provisioning Status: unknown
   Attached Filters:
   VAAI Status: unsupported
   Other UIDs: vml.0005000000766d68626133323a303a30
```

Then, we show how to match the path to the specific DS Storage System controller. Open the DS Storage Manager and select the storage subsystem to be managed. Then, go to the **Mappings** tab to identify the LUNs that we have assigned to the Host Group. For this example, we use **Host VMware_5**. As shown in Figure 6-5 on page 160, we have three logical drives.

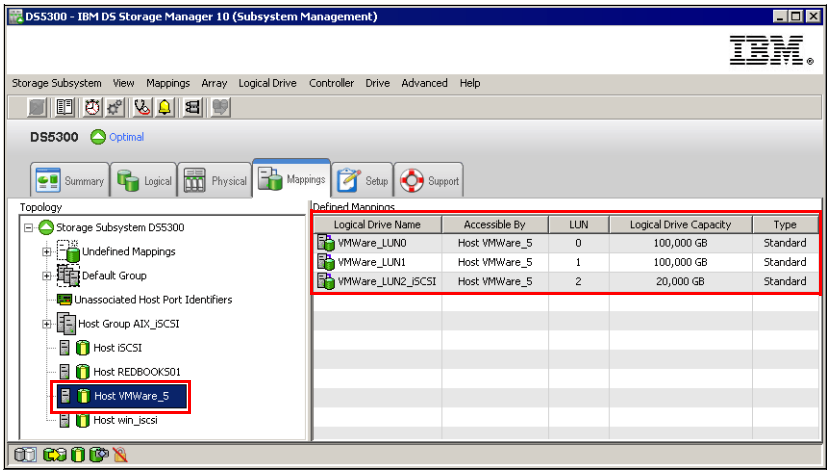

Figure 6-4 Identifying logical drives

Now, we need to obtain the LUN ID. Go to the **Logical** tab and select **VMware_LUN0** as shown in Figure 6-5.

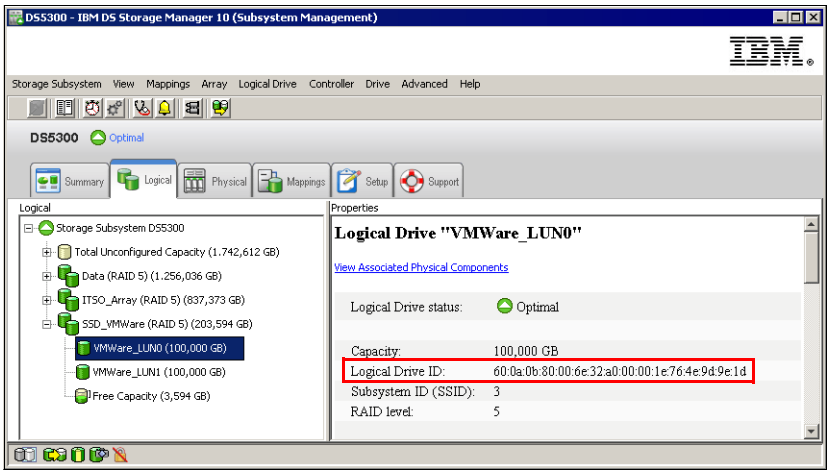

Figure 6-5 Getting the LUN ID from DS Manager

7

IBM System Storage Plug-In for VMware vCenter

We describe how to download, install, and use the common features with the IBM System Storage Plug-in for VMware vCenter for IBM DS3000, DS4000, and DS5000 series storage.

7.1 Overview

IBM System Storage Plug-In for VMware vCenter was designed to provide integrated management for IBM System Storage DS3000, DS4000, and DS5000 from within the same interface as VMware vSphere Client. It simplifies storage management without requiring you to learn a new tool while it delivers a centralized console for virtual environments.

The vCenter plug-in uses an application server as the interface between the vSphere Client and the DS3000, DS4000, and DS5000 series storage. The access level is based on the privileges that are assigned to the logged-in user.

We show the most common features. To obtain more information about the premium and advanced features, see the following website to download specific documentation about this plug-in:

http://www.ibmvcenterplugin.com

This plug-in allows the following tasks to be performed from within the vSphere Client:

- Configure ESX and ESXi hosts to DS3000, DS4000, and DS5000 series storage subsystems.
- Create and delete logical drives.
- Map logical drives from DS3000, DS4000, and DS5000 subsystems to the ESXi hosts.
- View vCenter datastores on DS3000, DS4000, and DS5000 series logical drives.
- With the premium features enabled on the storage system, this plug-in allows you to create and manage flashcopies, logical-drive copies, and Enhanced Remote Mirroring (ERM).

Note: For more information about premium features, see the *IBM System Storage DS Storage Manager Copy Services Guide*, SG24-7822.

7.2 Download

To obtain the latest version of the plug-in, go to the following website:

http://www.ibmvcenterplugin.com

You must enter a user name and password combination or register if you do not have a user name and password (as shown in Figure 7-1 on page 163).

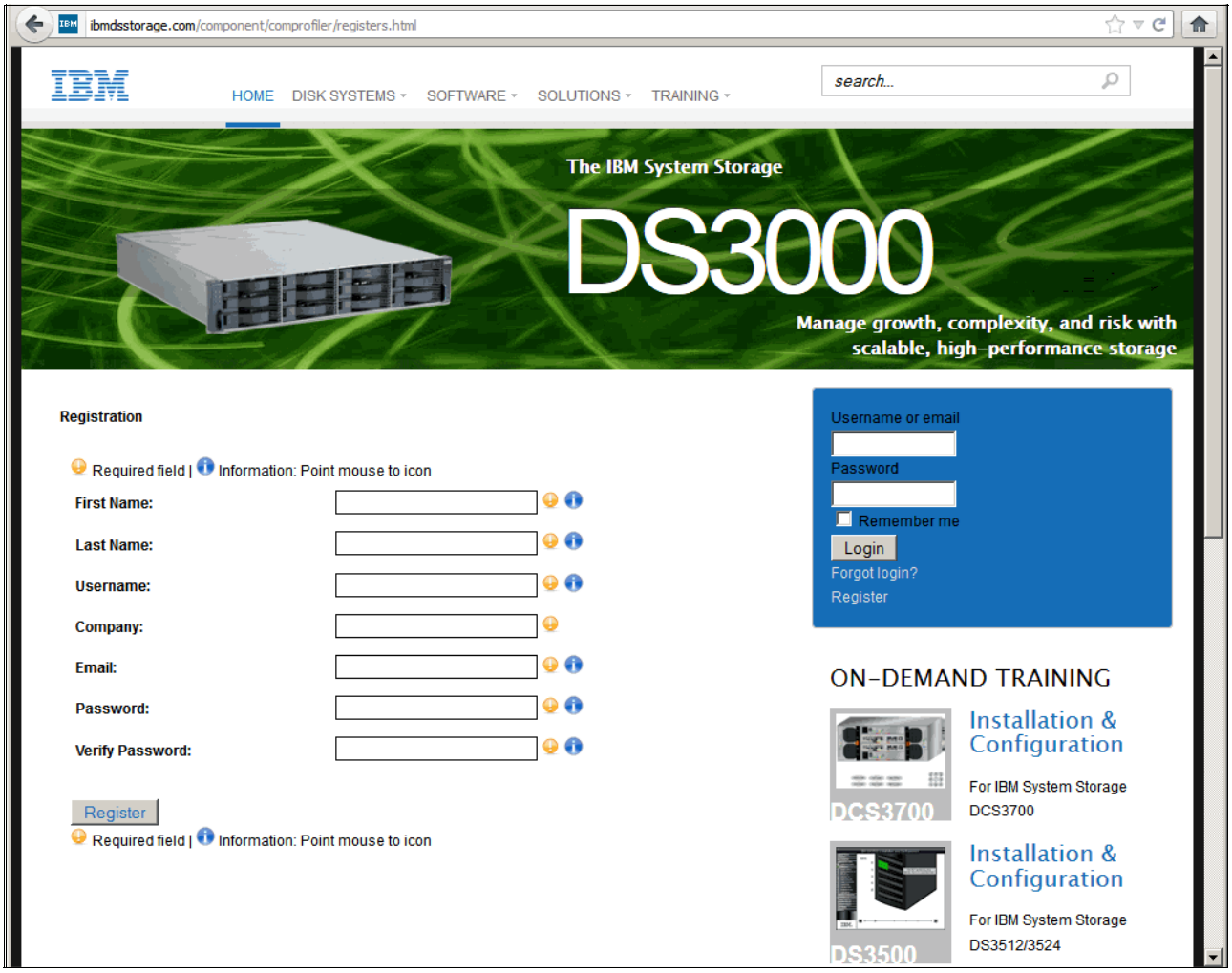

Figure 7-1 Register to download plug-in

> **Note:** The IBM support user name cannot be used on this website. You must create a user name if you do not have one that is registered.

After the registration process and after you log in to the website, you must complete the information about your company and current infrastructure, before you download, as shown on Figure 7-2 on page 164.

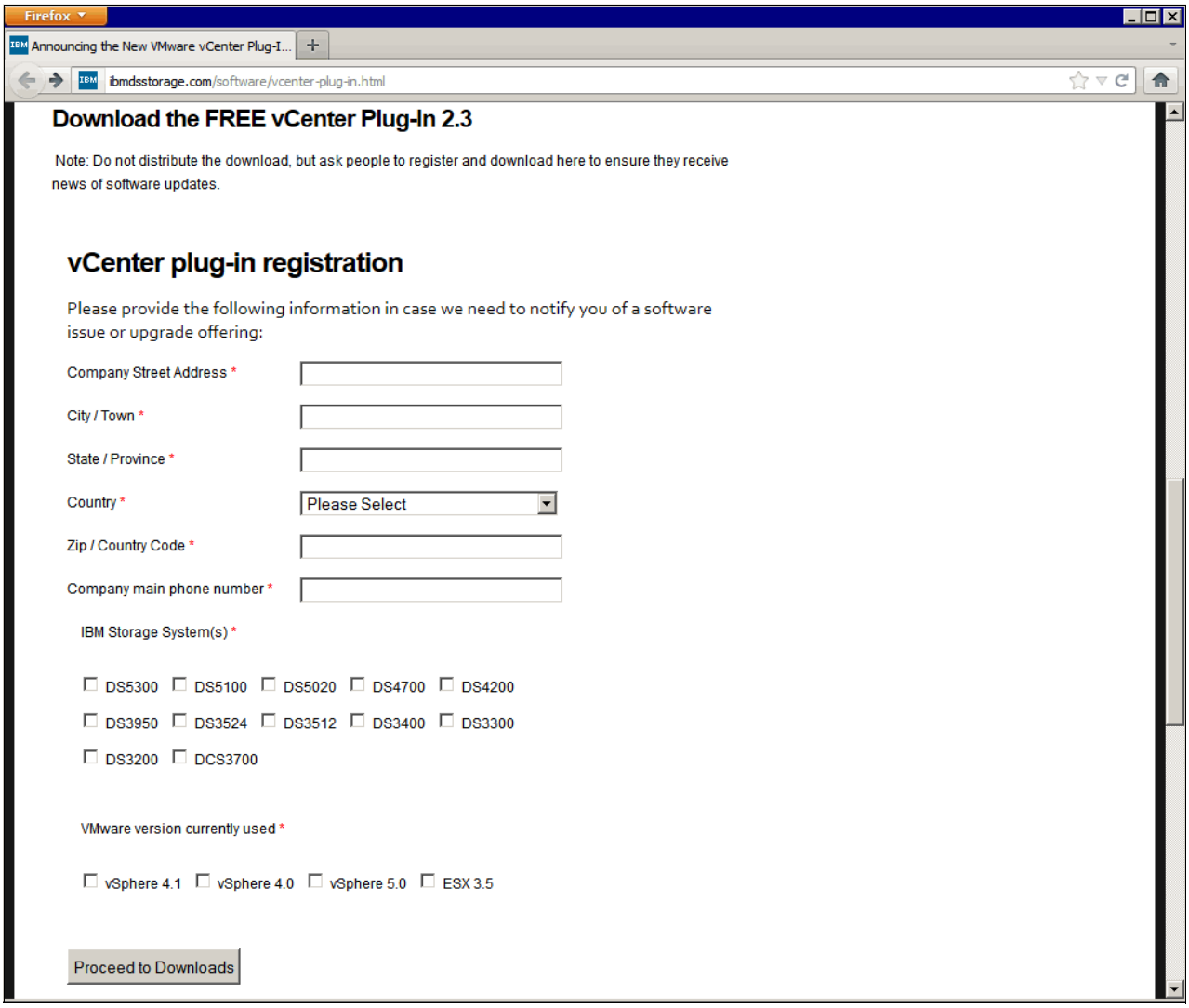

Figure 7-2 Complete the form before you download the plug-in

After you complete the form, click **Proceed to Downloads** where you have two options for the vCenter plug-in:

- SMIA-vCenter-WSX64-05.23.3550.0003.exe for the Microsoft Windows x64 version
- SMIA-vCenter-WS32-05.23.3550.0003.exe for the Windows x86 version

Note: At the time of writing this book, the latest version of the plug-in is 2.3, which is the first version to add support for vSphere 5.

7.3 vCenter plug-in installation

We describe the vCenter plug-in installation procedure.

7.3.1 Prerequisites

Before you proceed with the vCenter plug-in installation, ensure that your system satisfies all prerequisites that are listed:

- VMware vCenter 4.x Server is installed

  Note: The vCenter plug-in requires that a vCenter Server is installed on the environment. The vCenter plug-in does not function in a configuration with only a vSphere Client and an ESX host configuration.

- One of the following servers is available to be the vCenter plug-in application server:
  - Microsoft Windows Server 2003
  - Microsoft Windows Server 2003 R2
  - Microsoft Windows Server 2008 32-bit
  - Microsoft Windows Server 2008 64-bit
  - Microsoft Windows Server 2008 R2
- DS3000, DS4000, and DS5000 series storage with VMware certified firmware. The vCenter plug-in requires one of the following firmware versions: 7.35, 7.60, 7.70, 7.75, 7.77 or 7.80.
- A VMware vSphere Client installation is recommended with the vCenter Server version. For more information and to download this software, see this website:

  https://my.vmware.com/web/vmware/downloads

  Note: The vCenter plug-in is not a direct replacement for the *Santricity ES Storage Management Software*, which is still required to perform certain storage-administration tasks.

7.3.2 Installing the vCenter plug-in

Install the vCenter Application Server in a separate server from VMware vCenter Server.

Important: It is possible to install the vCenter Application Server plug-in on the same server as VMware vCenter Server if the server has enough resources. It is important to observe the port assignments during the setup process as shown in Figure 7-9 on page 168.

After you download the correct package, as described in 7.2, “Download” on page 162, proceed with the installation wizard on the chosen vCenter Application Server. We used a Windows Server 2008 R2 x64.

Follow these steps:

1. Choose an installation language in the splash panel as shown in Figure 7-3 on page 166. English was used for this demonstration because it was our Operational System default. Click **OK** to proceed.

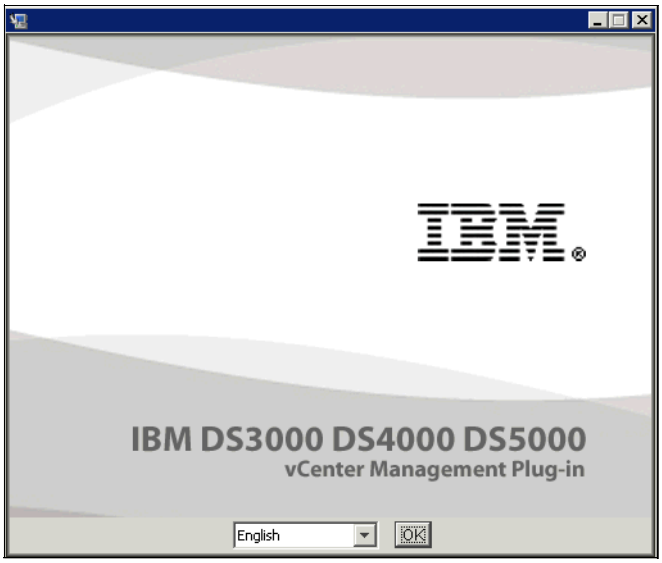

Figure 7-3 Installation wizard splash panel

2. Read and accept the license agreement (Figure 7-5). If you do not click **Accept**, you see the Warning message that is shown in Figure 7-4.

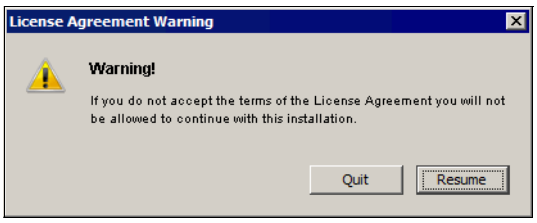

Figure 7-4 License Agreement warning

3. Click **Next**.

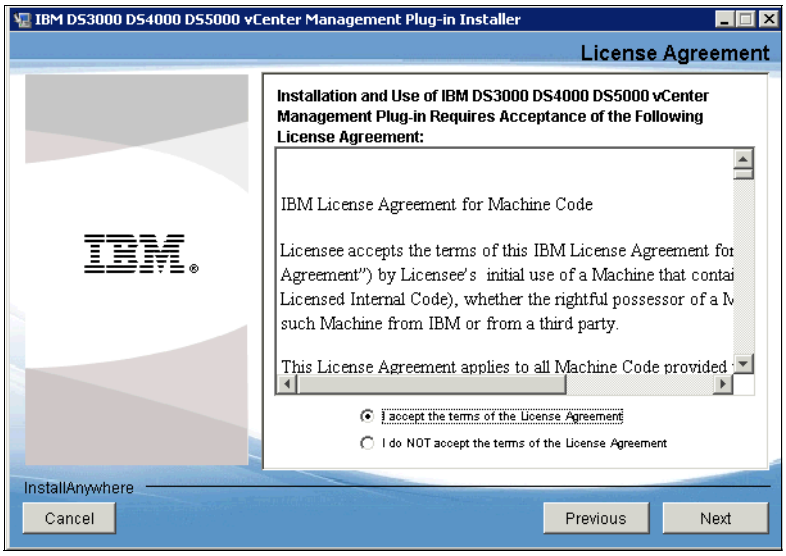

Figure 7-5 License Agreement

4. Decide where on the Application Server (reserve at least 500 MB of free disk space) to install the plug-in files. By default, the plug-in files are installed at `%SystemDrive%\Program Files (x86)\IBM DS3000 DS4000 DS5000 vCenter Management Plug-in`, as shown in Figure 7-6. Click **Next**.

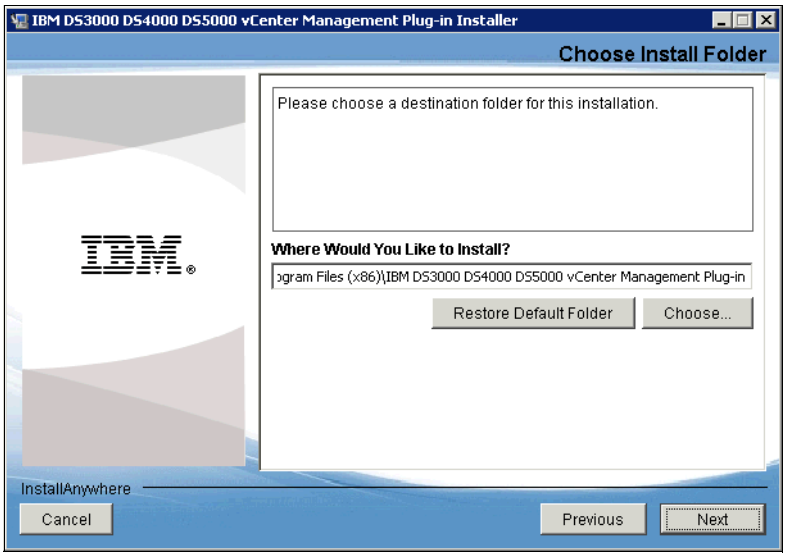

Figure 7-6 Installation path

5. The Pre-installation Summary window (Figure 7-7) opens. You can review the options that were selected in earlier steps. Click **Install** to begin the process.

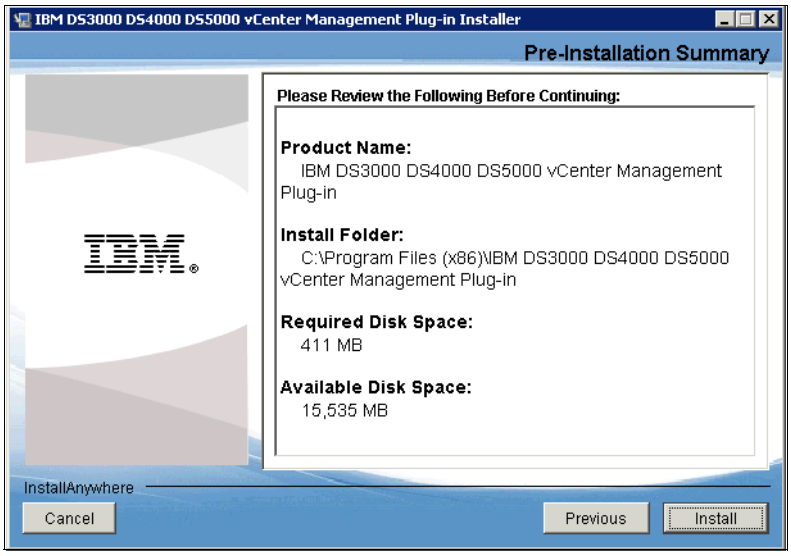

Figure 7-7 Pre-installation Summary window

6. It might take some time to finish this process. A progress window opens similar to Figure 7-8 on page 168.

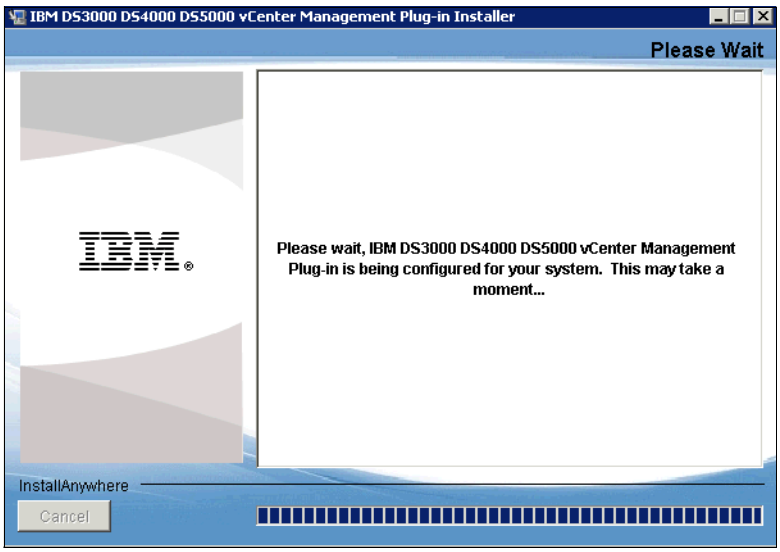

Figure 7-8 Installation progress window

7. Configure the Jetty Secure Sockets Layer (SSL) and non-SSL port numbers (Figure 7-9). Click **Next** to continue.

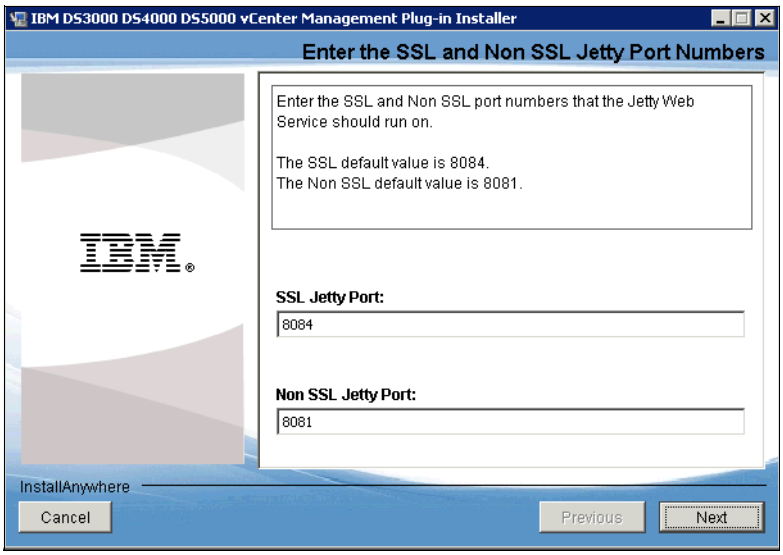

Figure 7-9 Configuring Jetty SSL and non-SSL port numbers

Important: If this installation is performed on the VMware vCenter Server and not on a separate server, you must change these ports to port numbers that are *not in use* by other components. If you fail to change them, you might break your current VMware vCenter.

8. Provide the IP address of the application server that you are installing with the plug-in as shown in Figure 7-10 on page 169. Press **Next** to continue.

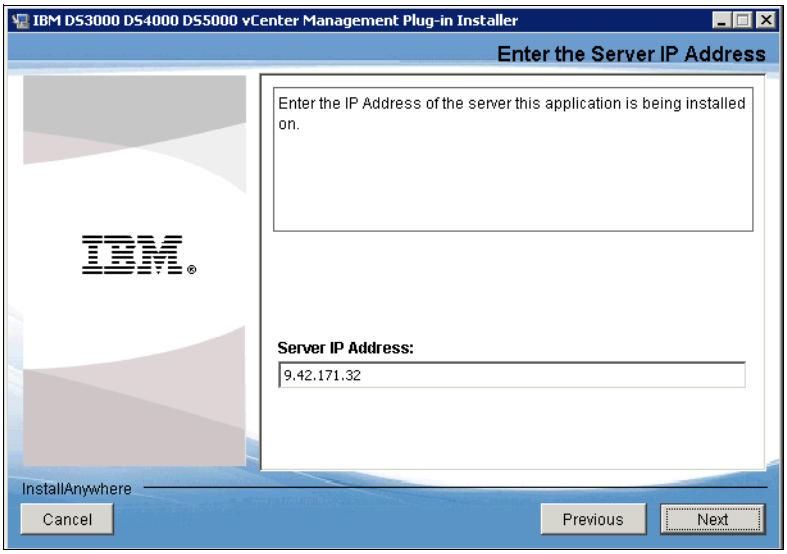

Figure 7-10 Application Server IP address window

9. The vCenter Host IP address must be provided to register the new plug-in (Figure 7-11). Click **Next** to proceed.

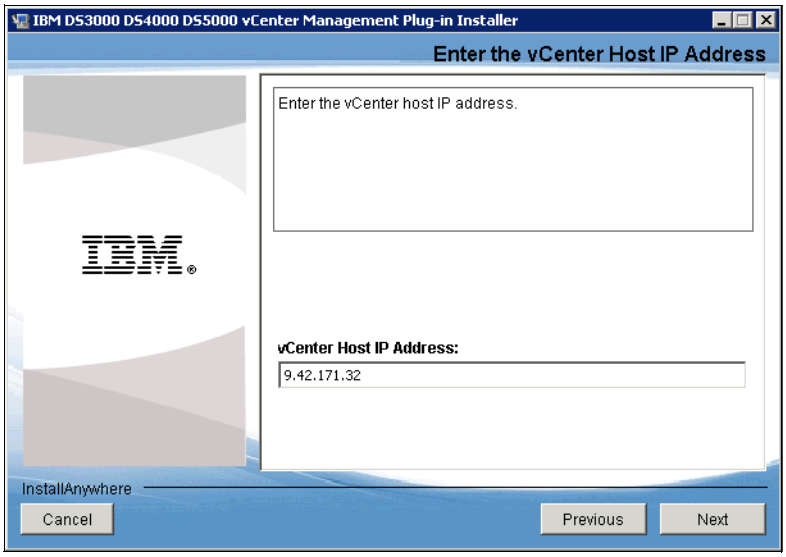

Figure 7-11 Enter the vCenter Host IP Address

10. Provide an administrative credential to authenticate and register the newly installed plug-in, as shown in Figure 7-12 on page 170. Click **Next** to proceed.

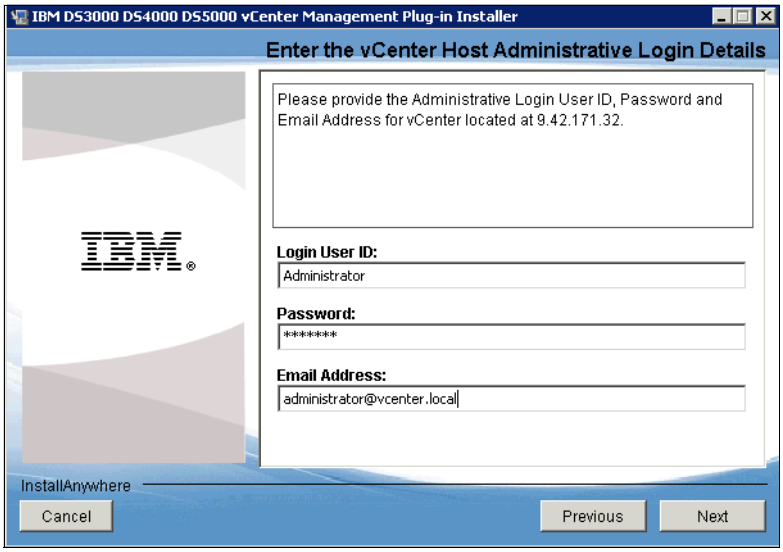

Figure 7-12 vCenter administrative authentication required

11. The Install Complete window opens after a successful installation (Figure 7-13). Click **Done** to start using the plug-in.

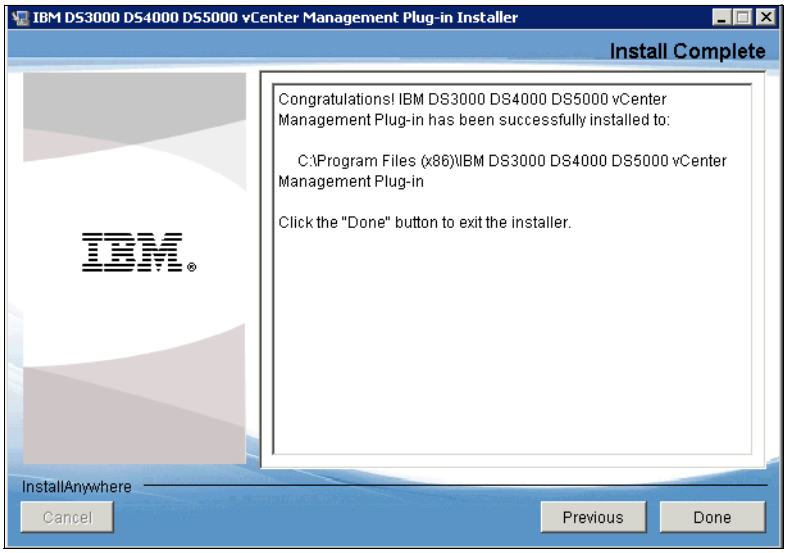

Figure 7-13 Installation complete

7.4 Configuring the vCenter plug-in

To verify the plug-in installation and availability, open your vCenter Infrastructure Client **Home** tab (as shown on Figure 7-14 on page 171). Under the Solutions and Applications section, look for the IBM DS3000/DS4000/DS5000 vCenter plug-in.

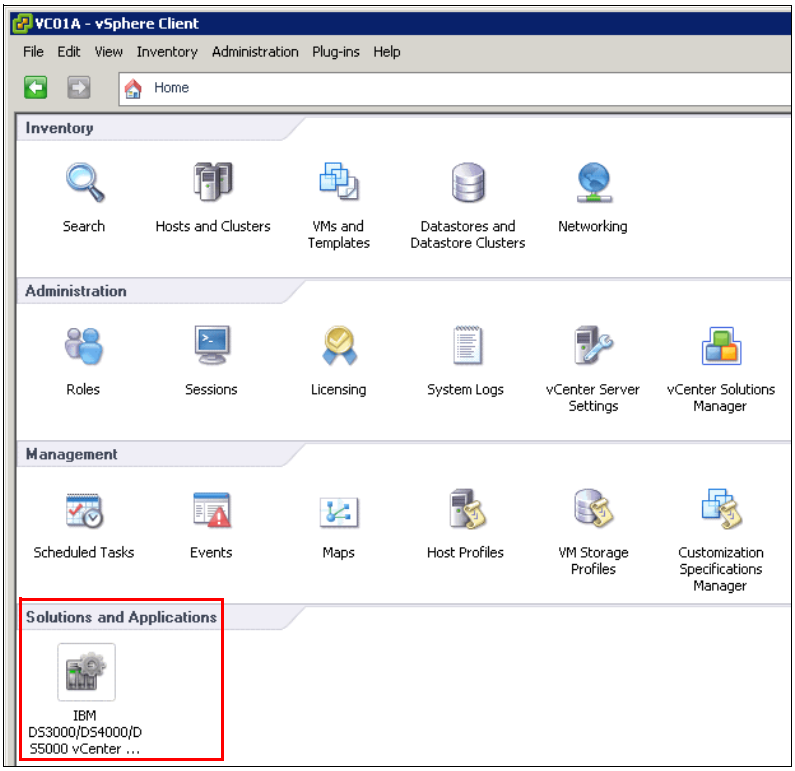

Figure 7-14 Plug-in installed and available

If you try to access the plug-in immediately after the installation, even by using an administrative account, a message displays, as shown in Figure 7-15. By default, all defined vCenter user IDs have no rights to DS3000, DS4000, and DS5000 series storage subsystems. A user requires either read permissions or read/write permissions to access the vCenter plug-in. The user's role must be modified to permit access to the vCenter plug-in.

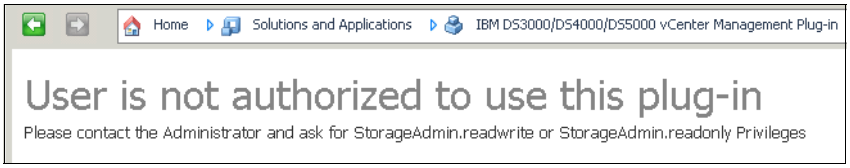

Figure 7-15 User is not authorized to use the plug-in by default

To allow user access to the plug-in, a role must be changed or created as shown in Figure 7-17 on page 172 to include the correct permissions.

Follow these steps:

1. On the vSphere Client Home page, under the Administration group, click **Roles** to access the window that is shown on Figure 7-16 on page 172.

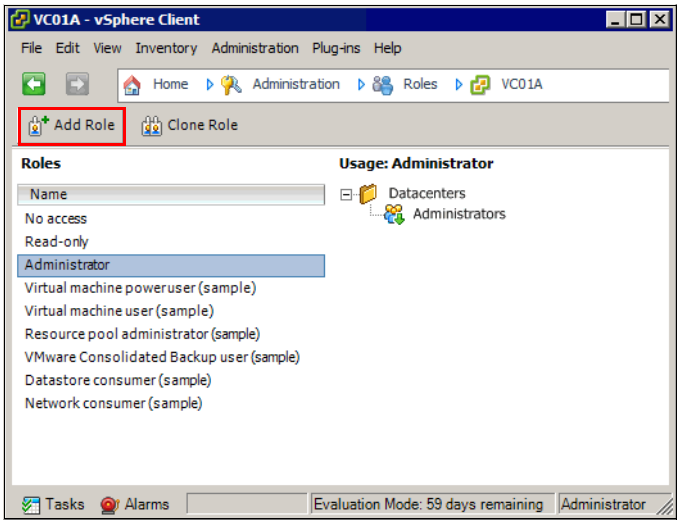

Figure 7-16 Add a role

2. Click **Add Role** to open a new window where you set the privileges (Figure 7-17).

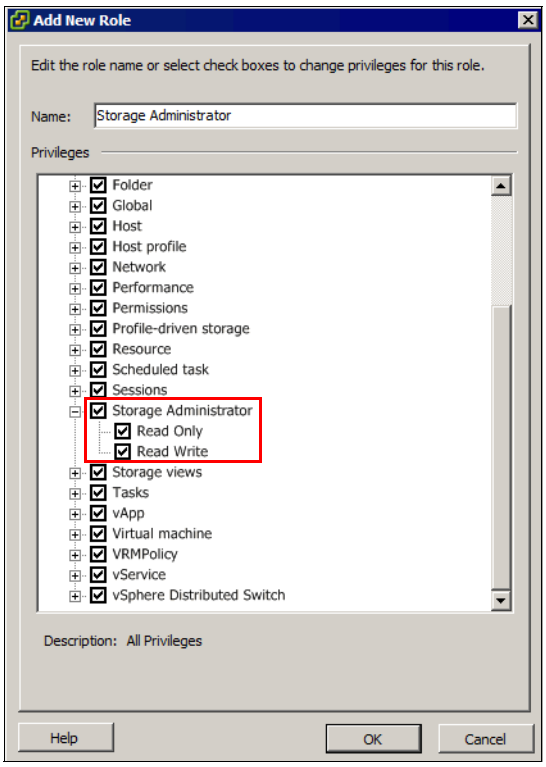

Figure 7-17 Privileges for the new Storage Administrator role

3. After you create the role, you assign permissions. Go to **Home** → **Hosts and Clusters**. Click the **Permissions** tab. Right-click the window and click **Add Permission** as shown in Figure 7-18 on page 173.

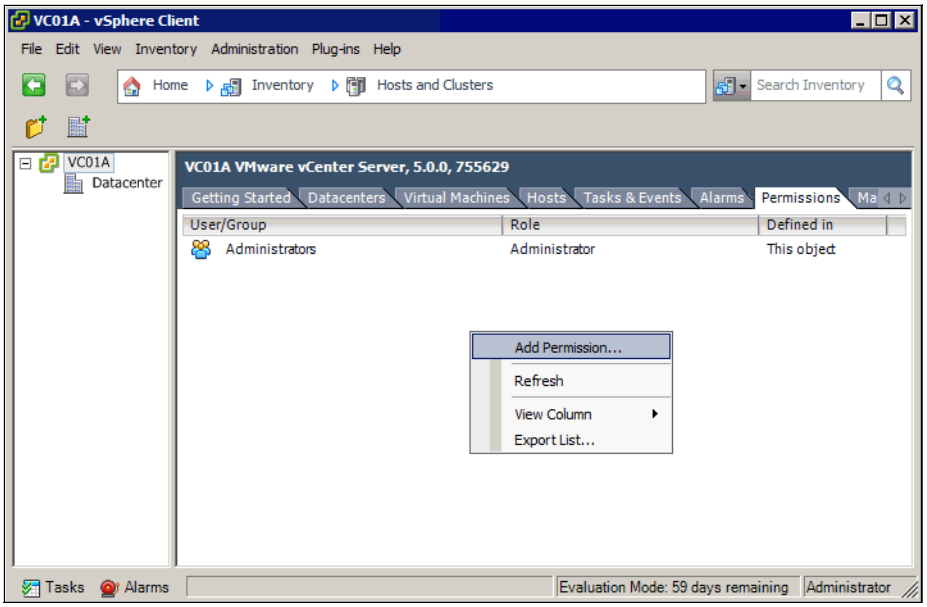

Figure 7-18 Add Permission

Note: The administrator role is not editable. If the administrator account is used to manage storage, create a role with all the privileges. The administrator account must then be added to this new role, as shown in this example.

4. The window on Figure 7-19 shows how to assign the new role to the correct account. Click **OK** to finish.

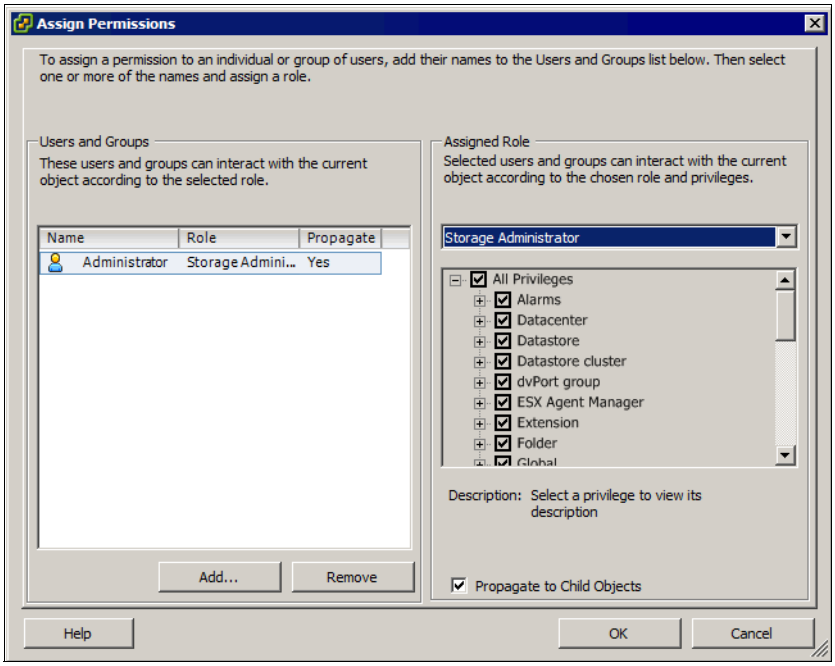

Figure 7-19 Assign Permissions window

Important: Close and reopen the vSphere Client to refresh permissions.

5. After you have the correct permission set, add the first DS3000, DS4000, or DS5000 series storage subsystem. You have to obtain the Domain Name System (DNS) names or IP addresses of both controllers to proceed. A password for the storage subsystem is also required.
6. On the vSphere Client, open **Solutions and Applications** → **IBM DS3000/DS4000/DS5000 vCenter Management Plug-in**. Click **Add Subsystem** as shown in Figure 7-20. You must enter both of the DNS names or IP addresses of the subsystem and its password before you click **Add**.

Note: For this demonstration, we used a DS5100 Storage Subsystem.

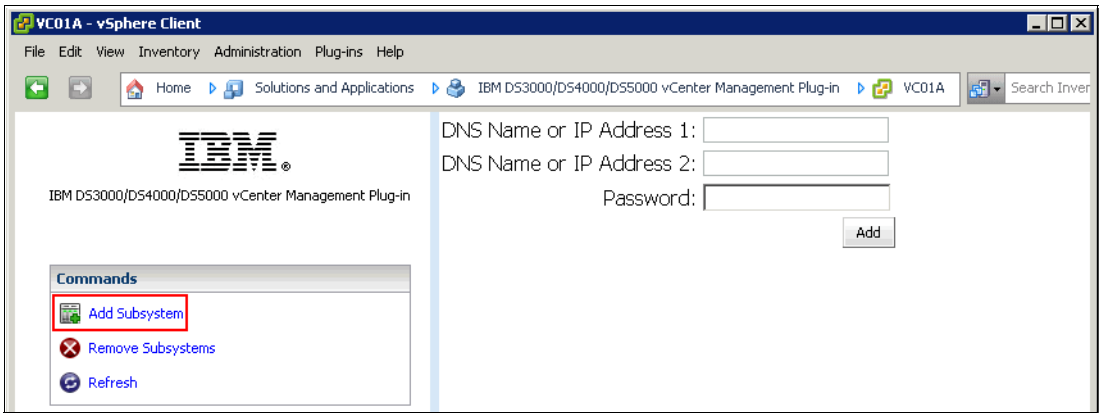

Figure 7-20 Add Subsystem

If all the information is entered correctly before you click **Add**, a new storage subsystem appears on the left side of the pane as shown in Figure 7-21.

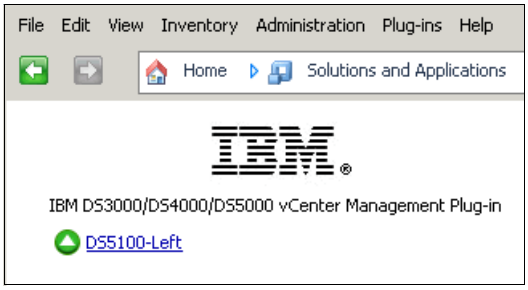

Figure 7-21 New storage subsystem added

7.5 Using the vCenter plug-in

We briefly describe the most common features that can be used to manage your DS3000, DS4000, and DS5000 storage subsystem through the vCenter plug-in.

Note: To obtain more details about these functions, see this website to download the latest technical documentation about this plug-in:

http://www.ibmvcenterplugin.com

The Summary tab shows an overview of the entire storage subsystem as shown in Figure 7-22.

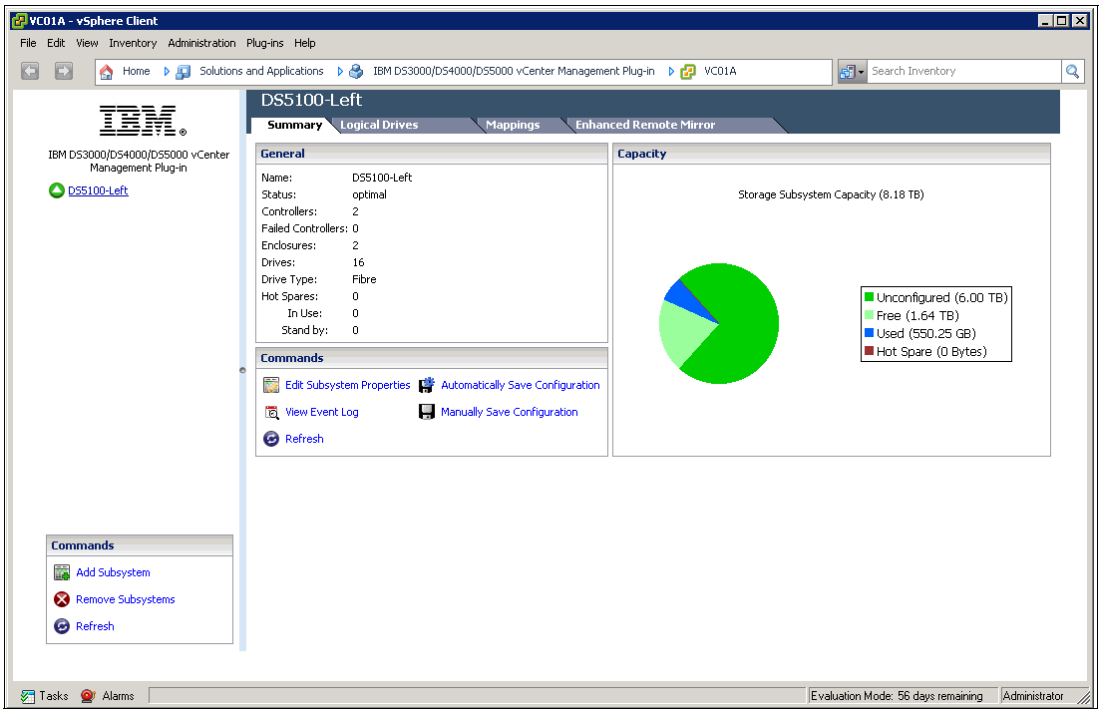

Figure 7-22 Summary tab

From this Summary tab, you can edit the subsystem properties, view the event log (Figure 7-23), and choose to automatically save the configuration schedule or manually download it.

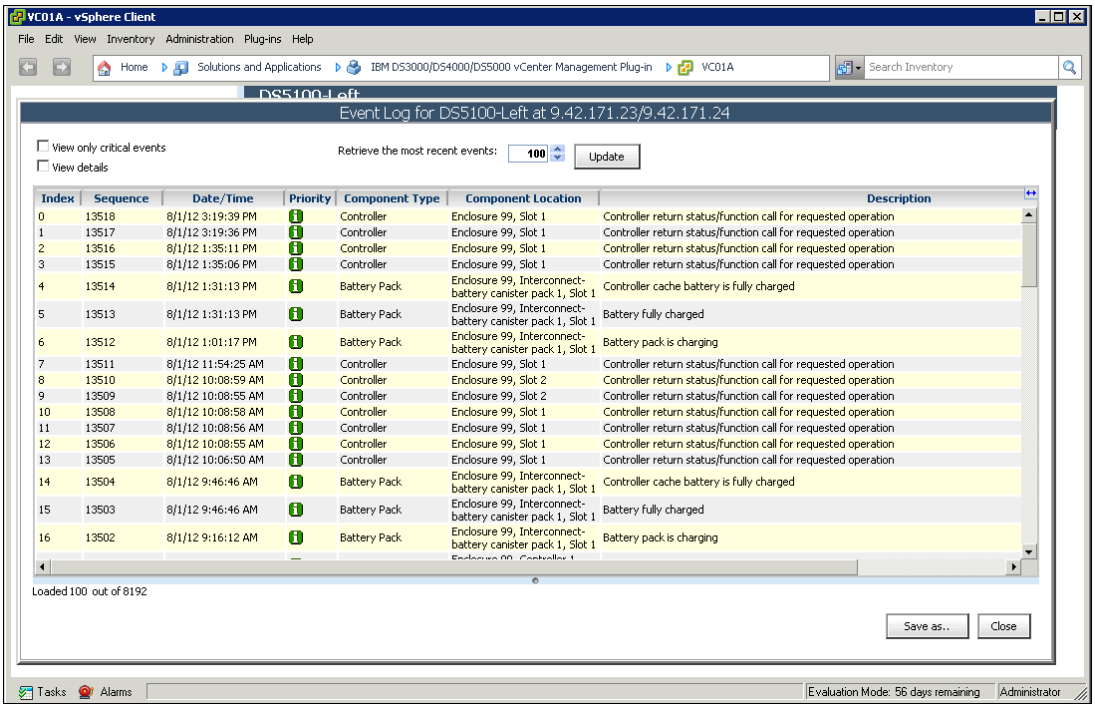

Figure 7-23 Event Log

The Logical Drives tab allows you to create new arrays and logical drives as shown in Figure 7-24.

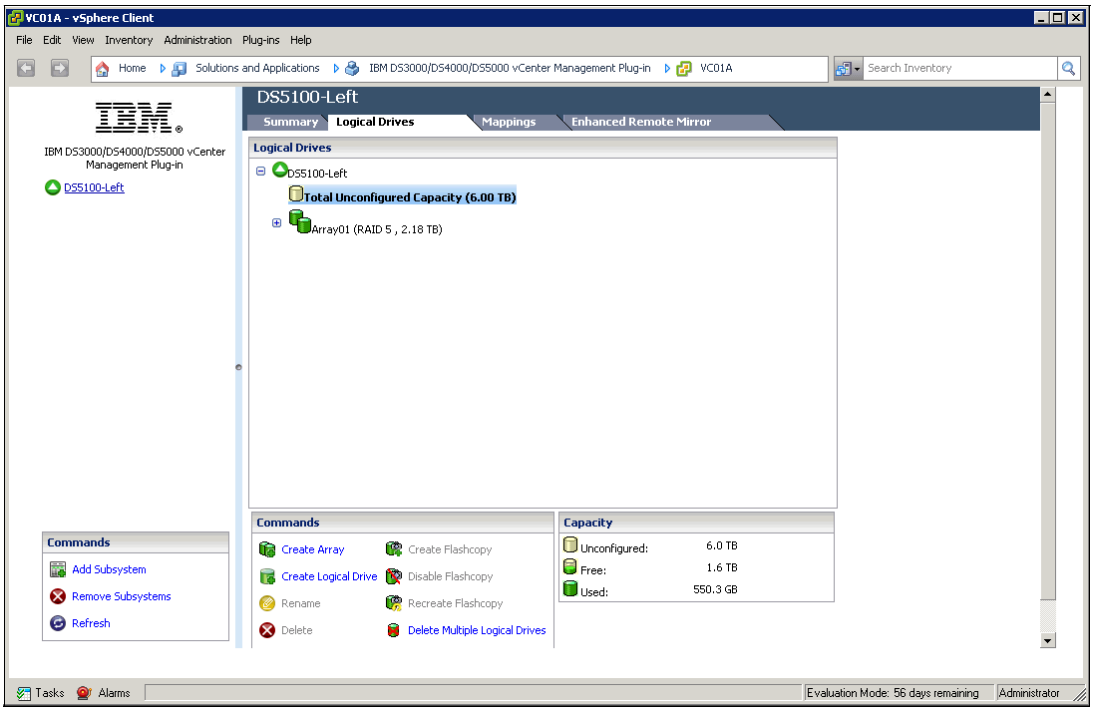

Figure 7-24 Logical Drives tab

The commands for Flashcopy on this tab, as well the Logical Drive copies and Enhanced Remote Mirror, are only available when those premium features are enabled on the storage subsystem level. For more information, see the *IBM System Storage DS5000 Series Implementation and Best Practices Guide*, SG24-8024.

Use the Mappings tab (Figure 7-25 on page 177) to add or delete hosts; create, remove, and manage host groups; and manage logical drive mapping, as well.

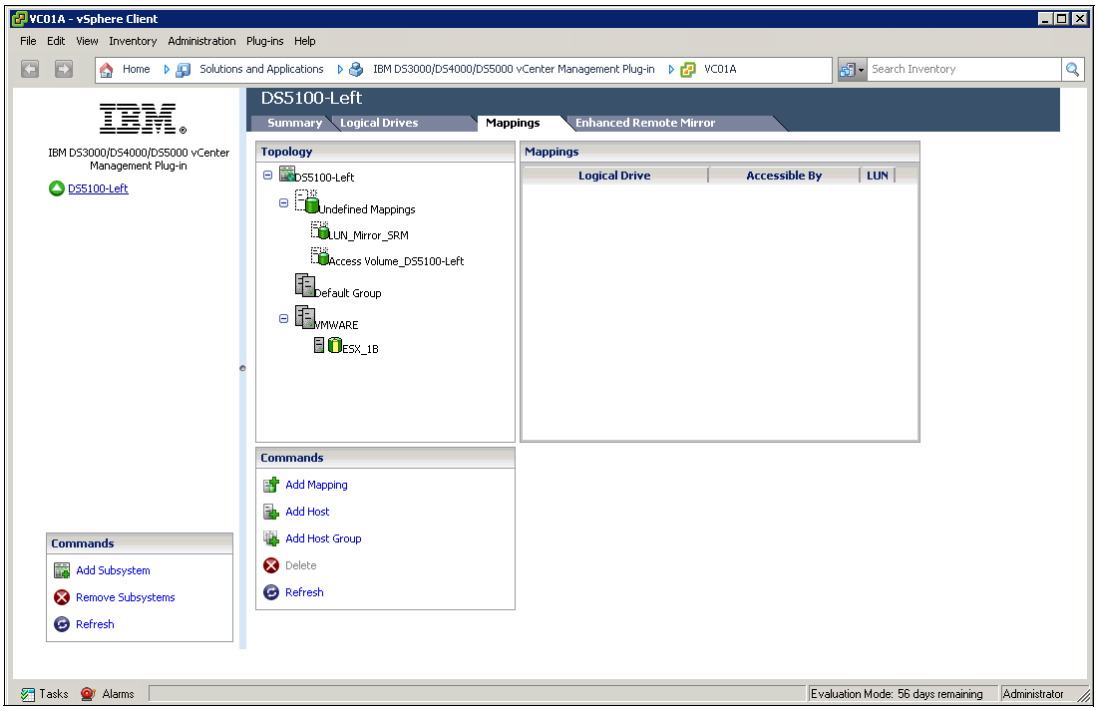

Figure 7-25 Mappings tab

8

VMware Site Recovery Manager (SRM5) implementation

In this chapter, we outline the process to implement VMware Site Recovery Manager 5 (SRM5) by using array-based replication that is provided by the IBM Enhanced Remote Mirroring (ERM) premium feature.

8.1 Planning for the installation

We present the steps that you need to perform to prepare your IBM Storage System and your Virtual Infrastructure for using IBM ERM and VMware vCenter Site Recovery Manager 5.

8.1.1 Basic solution components and layout

For our setup, we used the following hardware and software components:

- Two HS23 BladeServer with QLogic 8Gb Fibre Channel Expansion Card (CIOv)
- Two IBM BladeCenter® Fibre Channel pass-thru modules
- Two DS5300 Storage Subsystems + Expansions
- Two Brocade 6510 FC Switches
- VMware vSphere ESXi 5
- VMware vSphere vCenter 5
- VMware vSphere vCenter Site Recovery Manager 5
- IBM Storage Replication Adapter Modules (IBM_SRA_05.00.3550.0017)
- Licensed IBM Enhanced Remote Premium feature (IBM ERM)

Figure 8-1 shows the SRM implementation layout that is described in this chapter.

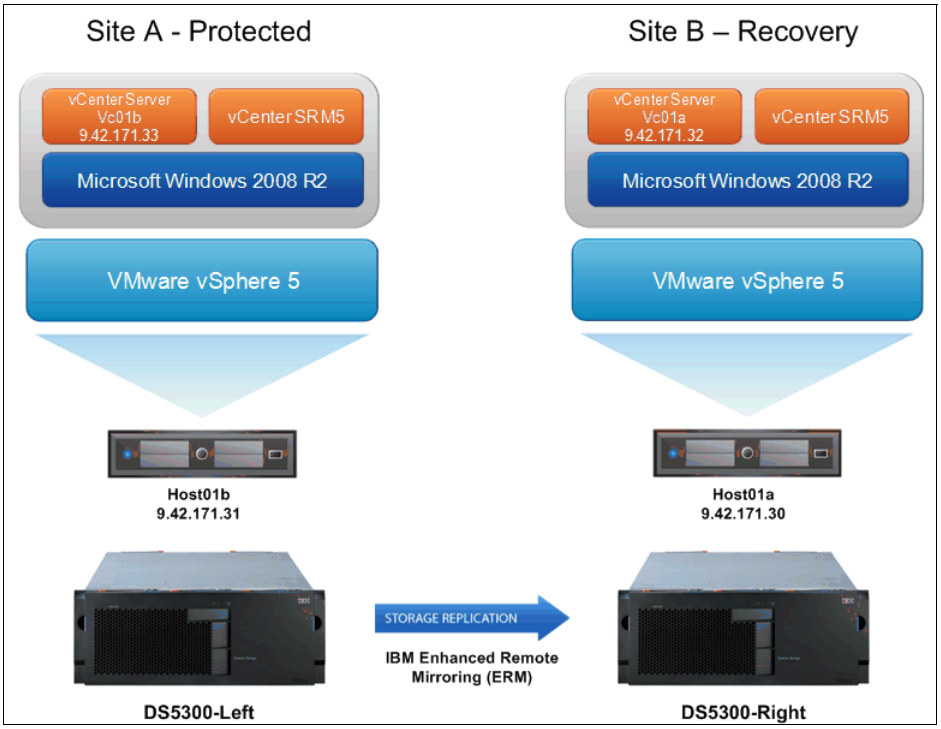

Figure 8-1 SRM implementation layout

Regarding *zoning configuration,* ensure that the SAN zoning is correctly set up according to your planned configuration. For more information about SAN FC Zoning, see *Implementing an IBM/Brocade SAN with 8 Gbps Directors and Switches*, SG24-6116.

For information about the HS23 BladeServer, components, and standard options, see *IBM BladeCenter HS23E*, TIPS0887.

8.1.2 SRM prerequisites

> **Note:** As part of the vSphere setup, SRM requires a plug-in to work. This plug-in is delivered only in vSphere Client.

vCenter Site Recovery Manager has the following prerequisites:

- Each site must have at least one data center. The SRM server operates as an extension to the vCenter Server at a site. Because the SRM server depends on vCenter for some services, you must install and configure vCenter Server at the protected site and at the recovery site.
- Pre-configured array-based replication: If you are using array-based replication, identical replication technologies, such as IBM ERM, must be available at both sites.
- A supported database engine that uses Open Database Connectivity (ODBC) for connectivity in the protected site and in the recovery site.
- An SRM license that is installed on the vCenter Server at the protected site and the recovery site. Additionally, vSphere needs to be licensed sufficiently for SRM to protect and recover virtual machines.
- The recovery site must have hardware, network, and storage resources that can support and guarantee the virtual machine workloads that are planned on the protected site.
- The sites must be connected by a reliable IP network. If you use array-based replication, ensure that your network connectivity meets the arrays' network requirements.
- The recovery site must have access to comparable networks (public and private) as the protected site.

8.1.3 System requirements

The system on which SRM is installed has the following hardware requirements:

- Processor: 2.0 GHz or higher Intel or AMD x86 processor
- Memory: 2 GB minimum
- Disk storage: 5 GB minimum
- Networking: Gigabit recommended

For current information about supported platforms and databases, see the Site Recovery Manager Compatibility Matrixes:

`http://www.vmware.com/support/pubs/srm_pubs.html`

8.1.4 Database requirements

The SRM server requires its own database, which it uses to store data, such as recovery plans and inventory information. The SRM database is a critical part of any SRM installation. The database must be created and a database connection must be established before you can install SRM.

The SRM database at each site holds information about virtual machine protection groups and recovery plans. SRM cannot use the vCenter database because it has different database

schema requirements, although, you can use the vCenter database server to create and support the SRM database.

SRM supports many databases that are provided by Microsoft, Oracle, and IBM among other vendors.

Our example uses *Microsoft SQL Server 2008 Standard (R2) - 64-bit* to support our SRM implementation. As shown in Figure 8-2, Microsoft SQL 2008 R2 Standard version is supported according to the VMware Product Interoperability Matrixes.

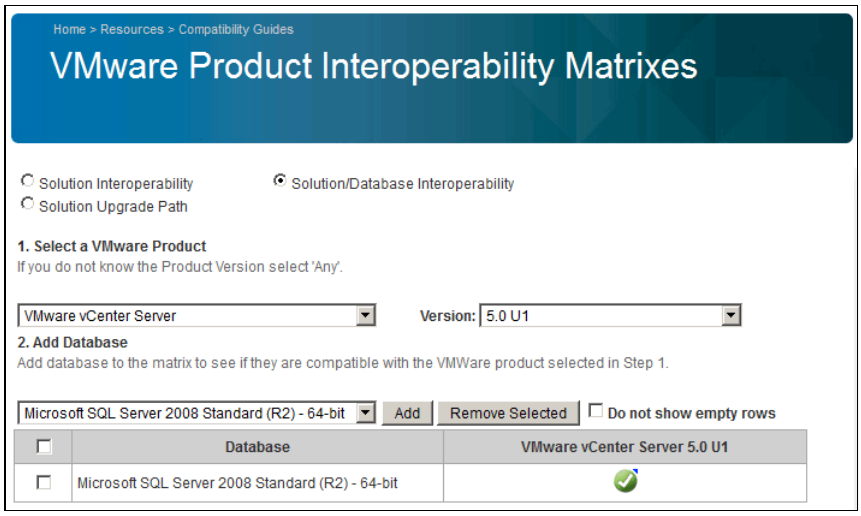

Figure 8-2 VMware Product Interoperability Matrixes - DB Engines

For a detailed Microsoft SQL Server configuration for SRM, see the *Site Recovery Manager Administration Guide*:

http://www.vmware.com/pdf/srm_admin_5_0.pdf

For more information about supported databases, see VMware Product Interoperability Matrixes:

http://partnerweb.vmware.com/comp_guide2/sim/interop_matrix.php?

8.1.5 Storage requirements

Note: You can use the vSphere Replication feature to configure SRM. The vSphere Replication feature implementation is out of the scope of this book. We only cover SRM by using IBM ERM as the array replication method.

SRM requires array-based replication as a prerequisite to support features such as a disaster recovery scenario and reprotection. The scenario that we present in this chapter uses array-based replication that is provided by the Storage Replication Adapter (SRA) extensions and the IBM ERM premium feature.

For this example (as shown in Figure 8-3 on page 183), we defined two datastores that are named LUN_Mirror_SRM that are replicated by using the ERM feature.

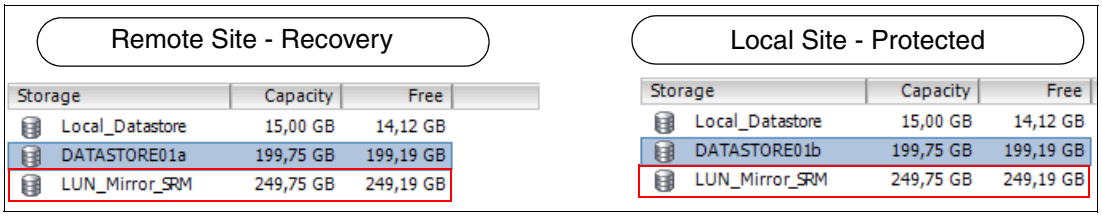

Figure 8-3 Mirrored sites

The detailed instructions and procedures for setting up and maintaining environments by using ERM are described in *IBM System Storage DS Storage Manager Copy Services Guide*, SG24-7822.

8.2 Installing Site Recovery Manager 5 (SRM)

Before we proceed with the SRM installation, we assume that the following prerequisites are met:

- The fully qualified domain name (FQDN) or IP address of the site's vCenter Server.
- A dedicated database for the exclusive use of the SRM service.
- User name and password for the SRM database.
- User name and password of the vCenter administrator.
- Provide a System DNS for the connection to the SRM database.
- Certificate information: If you plan to use certificate-based authentication, set the path name to an appropriate certificate file.

8.2.1 Installation procedure

Follow these steps:

1. Log in to the machine on which you are installing SRM (protected site). Use an account with sufficient privileges. This account is often an Active Directory domain administrator, but it can also be a local administrator.
2. Download the SRM installation file to a folder on the machine, or open a folder on the network that contains this file.
3. Double-click the SRM installer icon.
4. Select the language and click **OK** as shown in Figure 8-4.

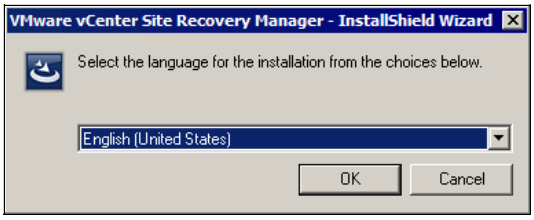

Figure 8-4 Select language

5. Click **Next** as shown in Figure 8-5 on page 184.

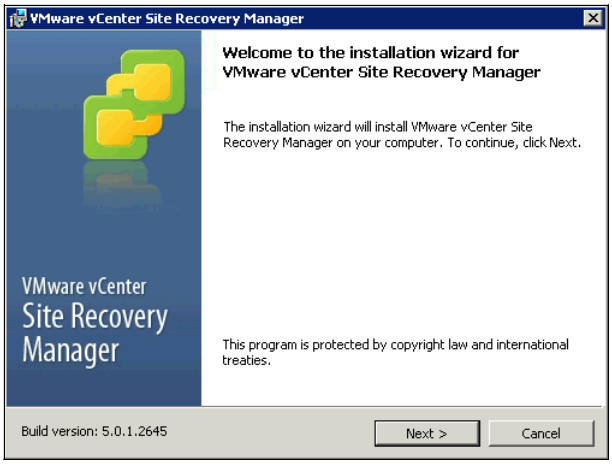

Figure 8-5 Installation wizard

6. Click **Next** on the VMware Patents window as shown in Figure 8-6.

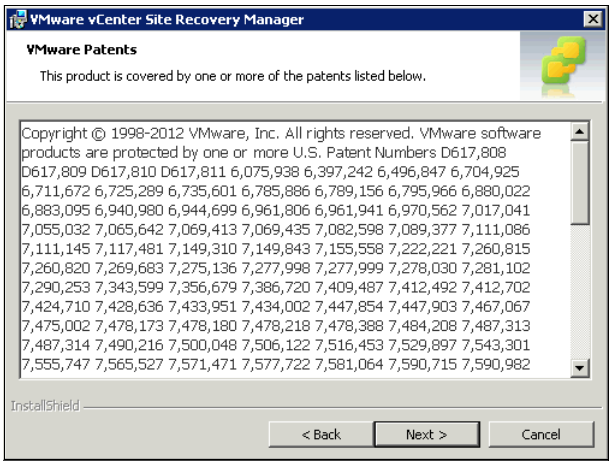

Figure 8-6 VMware Patents

7. Select **I agree to the terms in the license agreement** and click **Next** as shown in Figure 8-7.

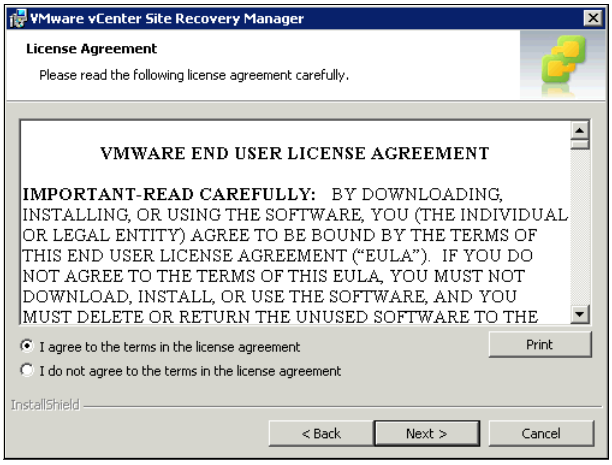

Figure 8-7 License Agreement

8. Select the folder in which to install SRM and click **Next** as shown in Figure 8-8.

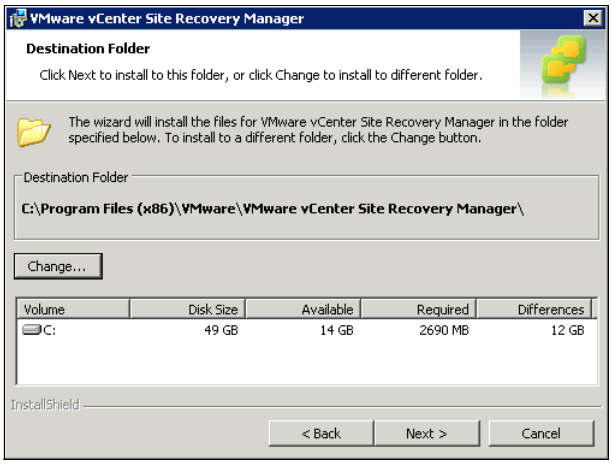

Figure 8-8 Destination Folder

9. Select whether to install vSphere Replication functions as shown in Figure 8-9. For this example, select **Do not install vSphere Replication**. Click **Next**.

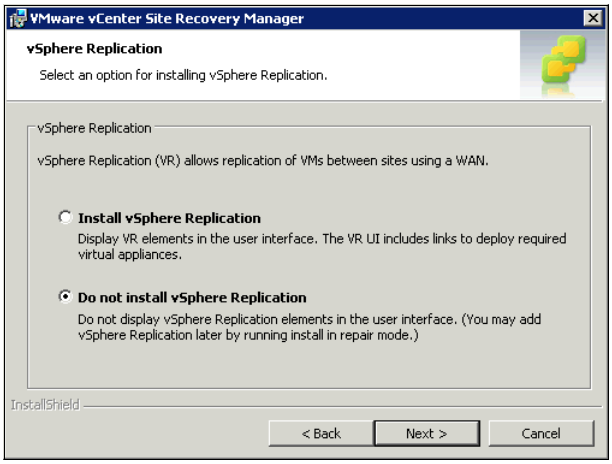

Figure 8-9 vSphere Replication

Installing vSphere Replication enables additional functions. You can install vSphere Replication later by using the Repair Installation option.

10. Enter information about the vCenter server at the site where you are installing SRM and click **Next** as shown in Figure 8-10 on page 186.

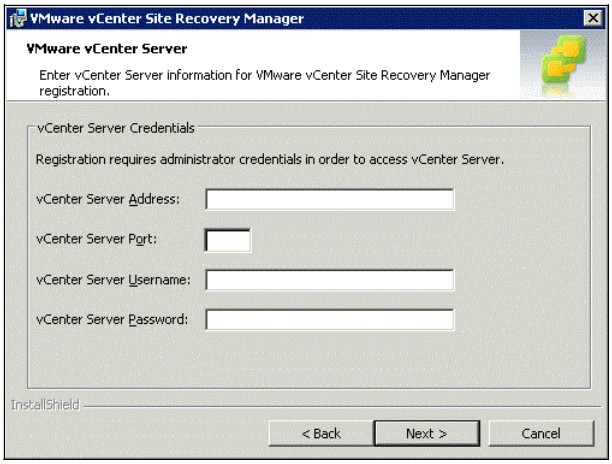

Figure 8-10 vCenter Server Credentials

11. Select an authentication method as shown in Figure 8-11:
    - To use credential-based authentication, select **Automatically generate a certificate** and click **Next**. Type text values for your organization and organization unit, which typically are your company name and the name of your group within the company.
    - To use certificate-based authentication, select **Use a PKCS#12 certificate file** and click **Next**. Type the path to the certificate file. The certificate file must contain exactly one certificate with exactly one private key that matches the certificate. Type the certificate password.

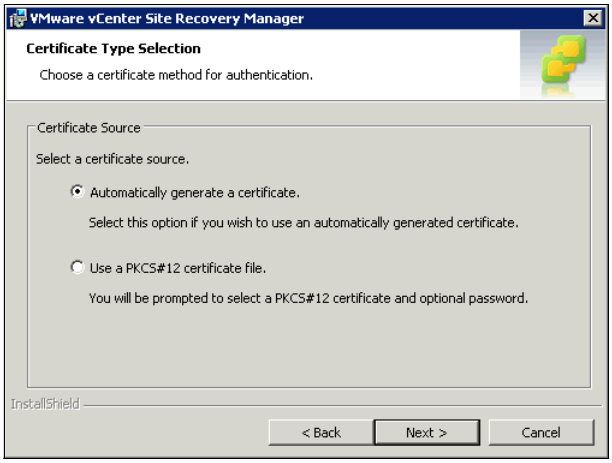

Figure 8-11 Certificate Type Selection

12. We plan to generate the certificate automatically. Type text values for your organization and organization unit as shown Figure 8-12 on page 187**.**

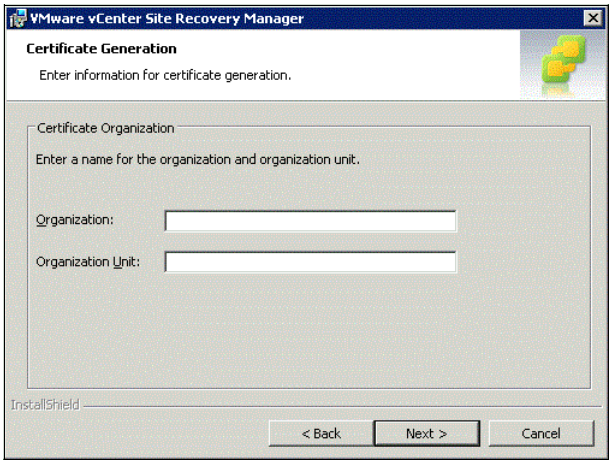

Figure 8-12 Certificate Generation

13. Type the administrator email and local host port configuration information as shown Figure 8-13. Click **Next**.

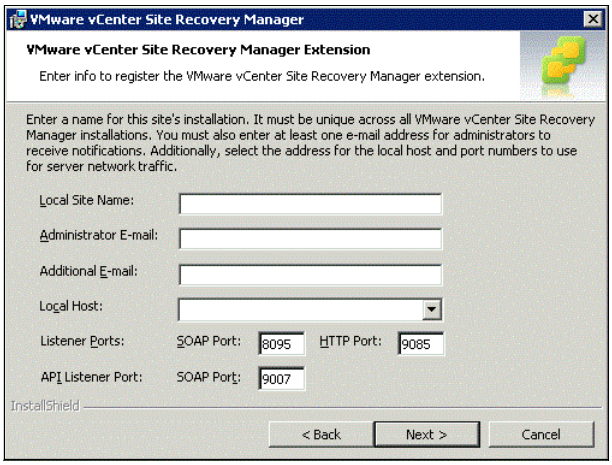

Figure 8-13 SRM extension

14. Type the SRM database configuration information and click **Next** as shown in Figure 8-14 on page 188. (We assume that you created the SRM database already and configured the DSN System correctly.)

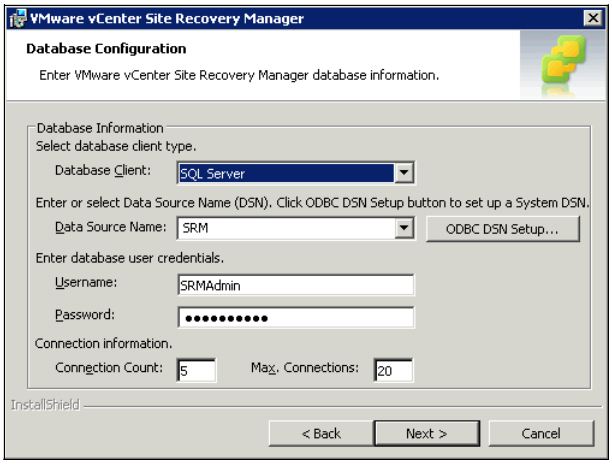

Figure 8-14 Database Configuration

15. Click **Install** as shown in Figure 8-15.

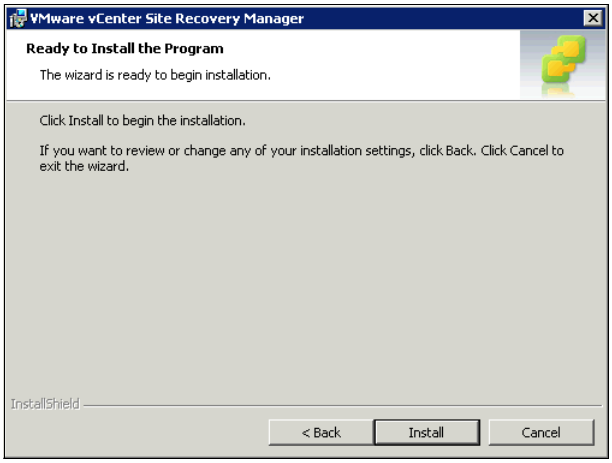

Figure 8-15 Ready to Install the Program

16. When the installation is finished, click **Finish** as shown in Figure 8-16.

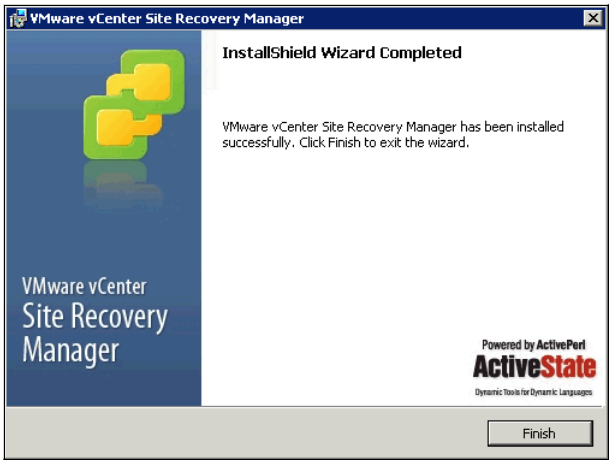

Figure 8-16 Installation completed

17. After the SRM is installed on the primary site (protected site), we need to proceed to repeat step 1 on page 183 through step 16 on page 188 on the secondary site (recovery site).

For more information, visit the SRM documentation support page:

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

8.2.2 Installing the VMware vCenter Site Recovery Manager Extension

When you install the SRM server, the VMware vCenter Site Recovery Manager Extension (SRM client plug-in) becomes available as a download from the vCenter server that the Site Recovery Manager server installation extends. SRM is inaccessible until the plug-in is installed.

You can download, install, and enable the SRM client plug-in on any host where a vSphere Client is installed.

Follow these steps to install the SRM plug-in:

1. Start the vSphere Client and connect to vCenter Server at the protected or recovery site.
2. Select **Plug-ins** → **Manage Plug-ins**.
3. In the Available Plug-ins area, right-click **VMware vCenter Site Recovery Manager Extension** and click **Download and Install** as shown in Figure 8-17.

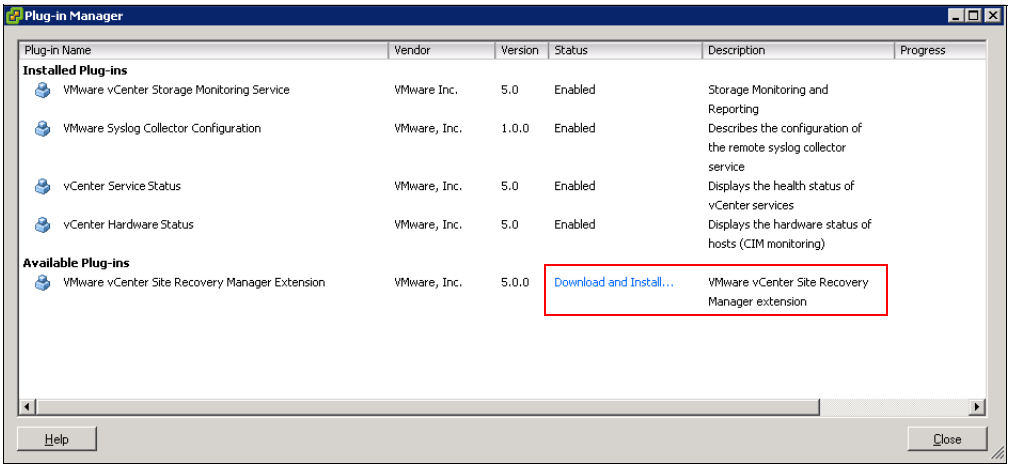

Figure 8-17 Plug-in Manager

4. After the download finishes, click **Next** to start the wizard.
5. Click **I accept the terms** in the license agreement, and click **Next**.
6. Click **Install**.
7. When the installation completes, click **Finish.**

Note: If the installation replaced any open files, you might be prompted to shut down and restart Microsoft Windows.

8. To validate the correct installation of the SRM modules and plug-in, open a vSphere Client connection to one of the vCenter Servers. Click **Home**. Click the **Site Recovery** icon under the Solutions and Applications section as shown in Figure 8-18 on page 190.

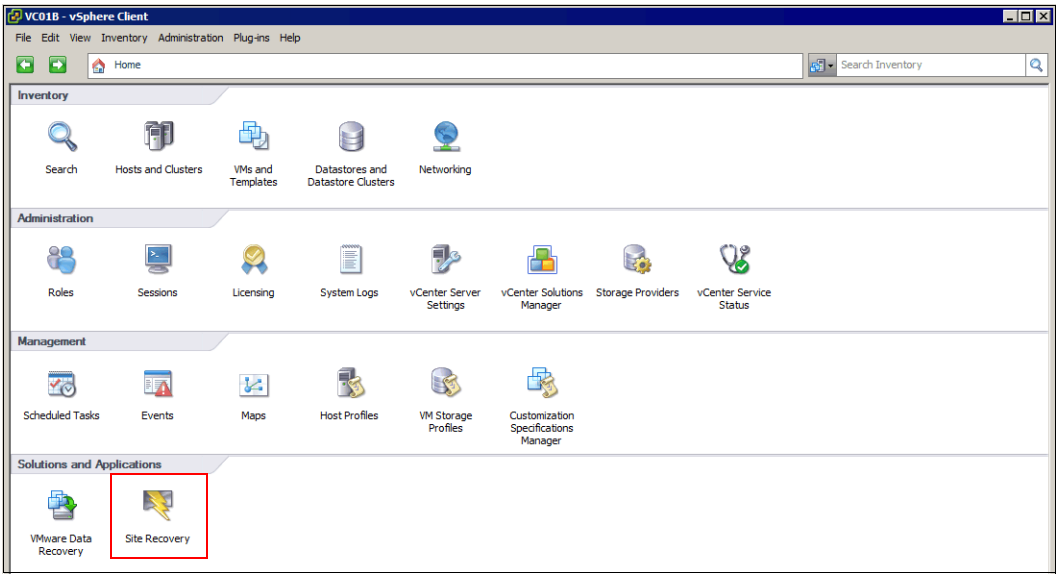

Figure 8-18 Validate the site recovery installation

8.2.3 Pairing sites

To implement disaster recovery (DR) protection between our two data centers, we need to establish reciprocity or “pair” the SRM servers that run at each site.

> **Note:** Before performing any configuration activity such as site pairing, ensure that you have installed an SRM server at both sites, and that you have installed the SRM plug-in at a vSphere Client from which you want to administer SRM.

Follow these steps to pair the sites:

1. Click **Sites** in the left pane and click **Configure Connection** as shown in Figure 8-19 on page 191.

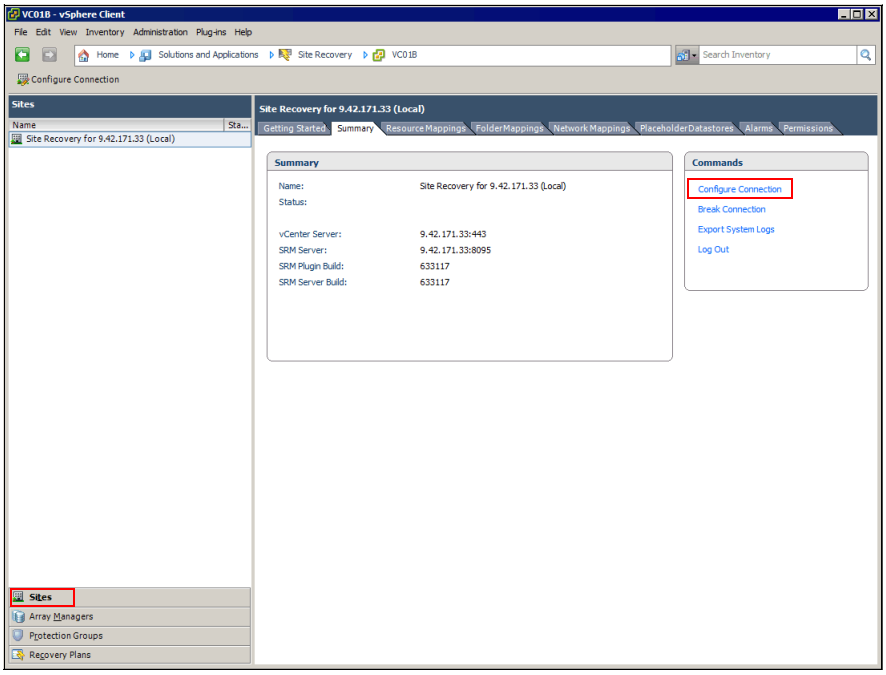

Figure 8-19 Configure a connection

2. On the Remote Site Information page, type the IP address or host name of the vCenter server at the recovery site and the port to which to connect and click **Next** as shown in Figure 8-20.

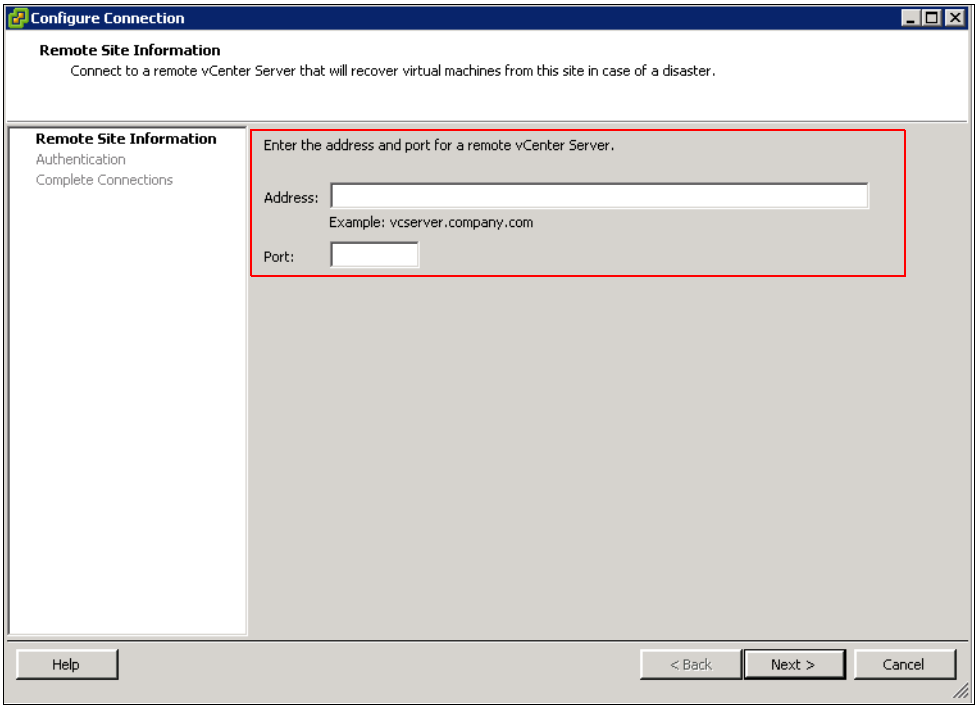

Figure 8-20 Remote Site Information

3. On the vCenter Server Authentication page, provide the vCenter administrator user name and password for the remote site as shown in Figure 8-21 on page 192 and click **Next**.

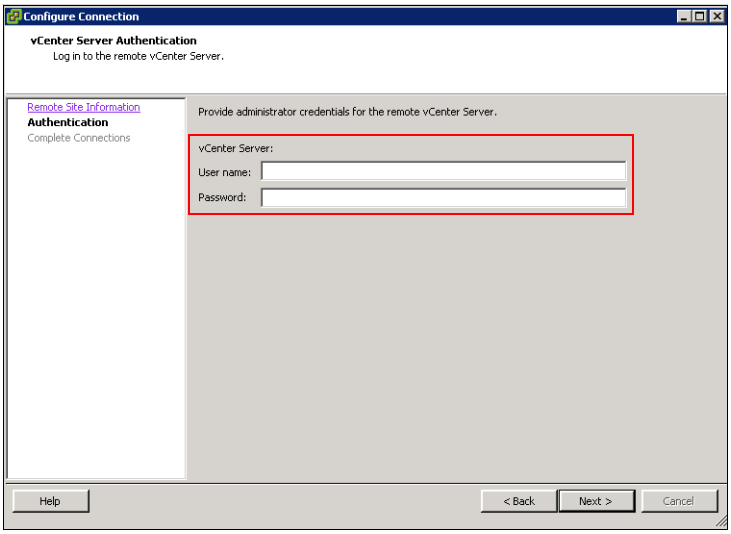

Figure 8-21 Authentication

4. On the Complete Connections page, click **Finish** after all of the site pairing steps complete successfully as shown in Figure 8-22.

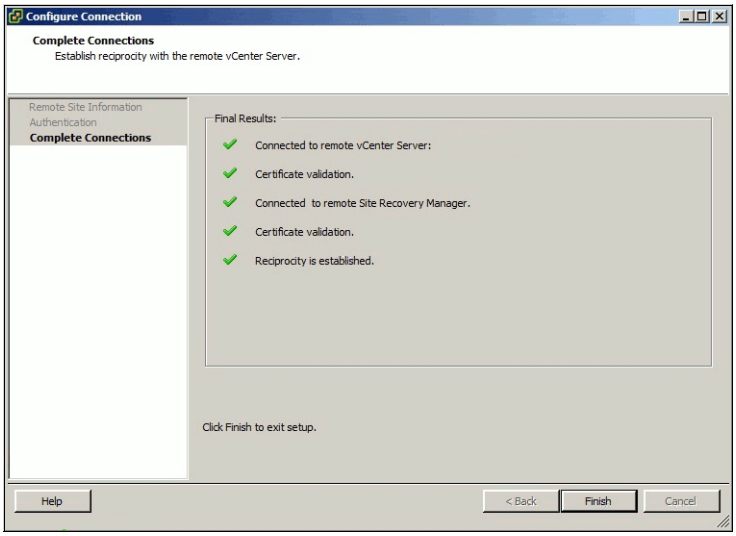

Figure 8-22 Complete Connections

5. After the pairing task completes, the new site (the Recovery site) is listed under the Sites section. On the Summary tab, the Status is *Connected* as shown in Figure 8-23 on page 193.

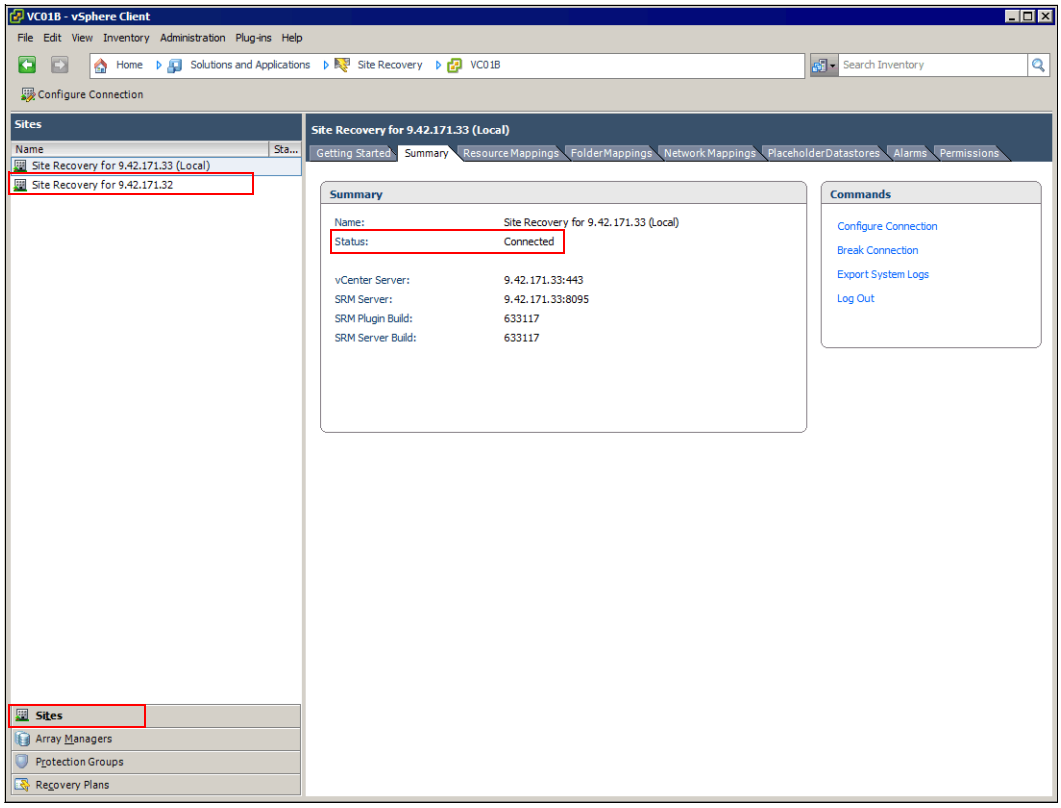

Figure 8-23 Pairing status

8.3 Storage Replication Adapters (SRAs)

The *Storage Replication Adapters* (SRAs) are software modules for Site Recovery Manager that are developed and supported by storage partners of VMware. The SRAs are distributed by VMware with the permission of the storage partners. For the instructions to download the SRAs, see 8.3.1, “Downloading SRAs” on page 193.

The SRA for DS3000, DS4000, and DS5000 uses both the ERM premium feature and the IBM FlashCopy premium feature to facilitate failover and to test failover between storage subsystems.

8.3.1 Downloading SRAs

Follow these steps to download the SRAs:

1. Go to the VMware Download home page:

   `https://my.vmware.com/web/vmware/downloads`

2. Select **All Downloads**.
3. On the Product Downloads tab, select **VMware vCenter Site Recovery Manager** and click **View Download** as shown in Figure 8-24 on page 194.

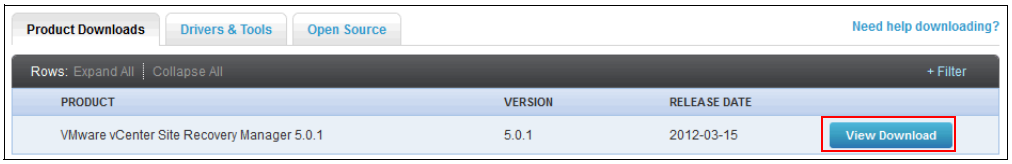

Figure 8-24 SRM components download: SRAs

4. Browse the SRM components list. Download the **IBM_SRA_05.00.3550.0017.zip** file as shown in Figure 8-25.

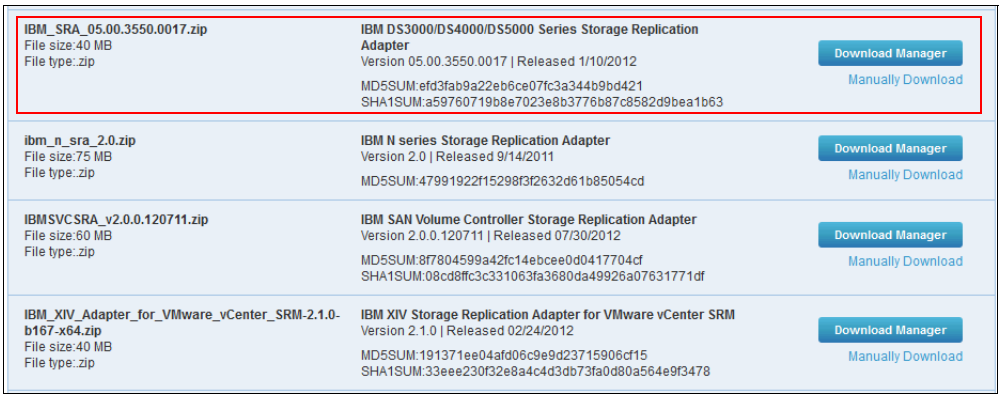

Figure 8-25 Downloading SRAs

Note: A *My VMware* account is required to download any VMware software package. For more information, review the instructions on the vendor's home page.

5. After the installation package is downloaded, extract the source file from a `.zip` file.

8.3.2 Installing SRAs

Before you proceed with the SRA installation, ensure that the SRM Server is installed at both the protected site and the recovery site in order to participate in disaster recovery. Follow these steps:

1. At each site on the SRM server, run the **SRAInstaller-05.00.3050.00*xx*.exe** package.
2. Select the installation language and click **OK** as shown in Figure 8-26 on page 195.

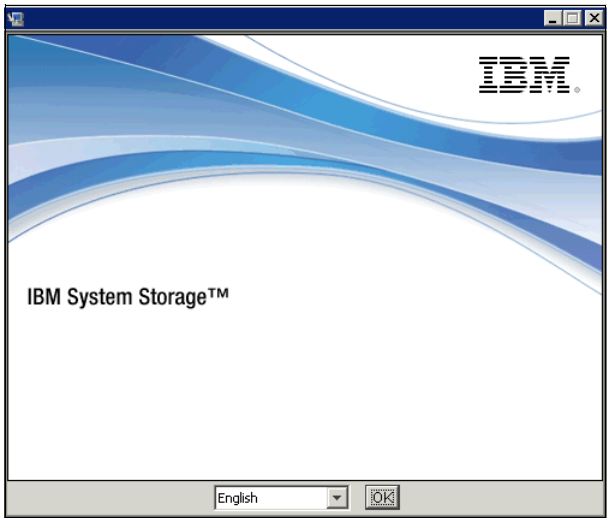

Figure 8-26 Language selection

3. Review the introduction information and click **Next** as shown in Figure 8-27.

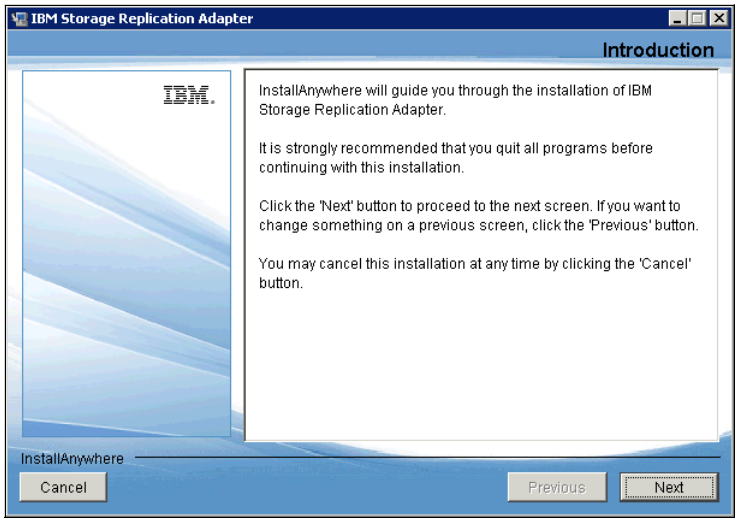

Figure 8-27 Introduction information

4. Review the copyright statement and click **Next** as shown in Figure 8-28 on page 196.

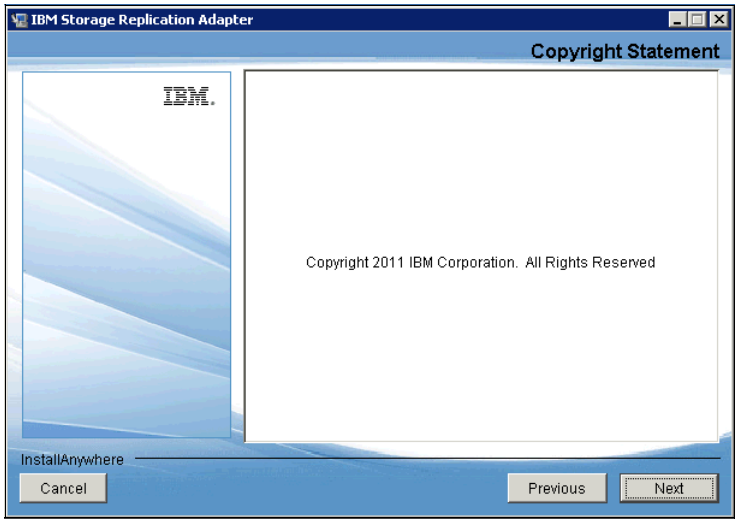

Figure 8-28 Copyright Statement

5. Review and accept the license agreement and click **Next** as shown in Figure 8-29.

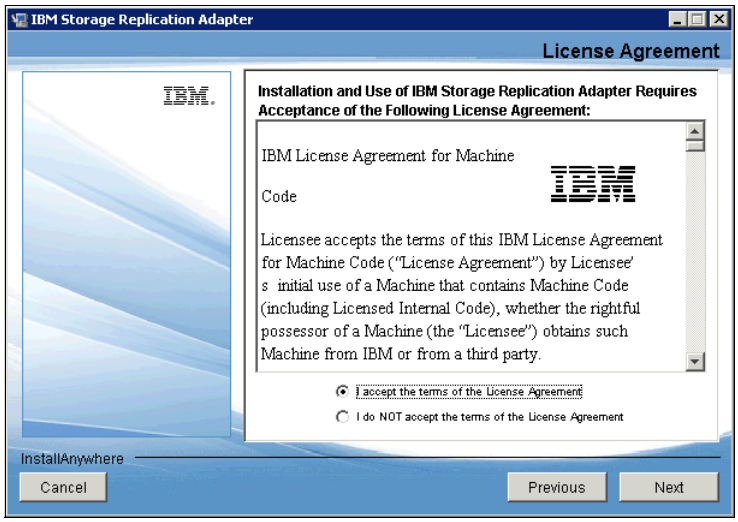

Figure 8-29 License Agreement

6. Verify that the SRA is installed to the `C:\Program Files (x86)\VMware\VMware vCenter Site Recovery Manager\storage\sra\IBM` directory as shown in Figure 8-30 on page 197. Click **Install** to begin the installation process.

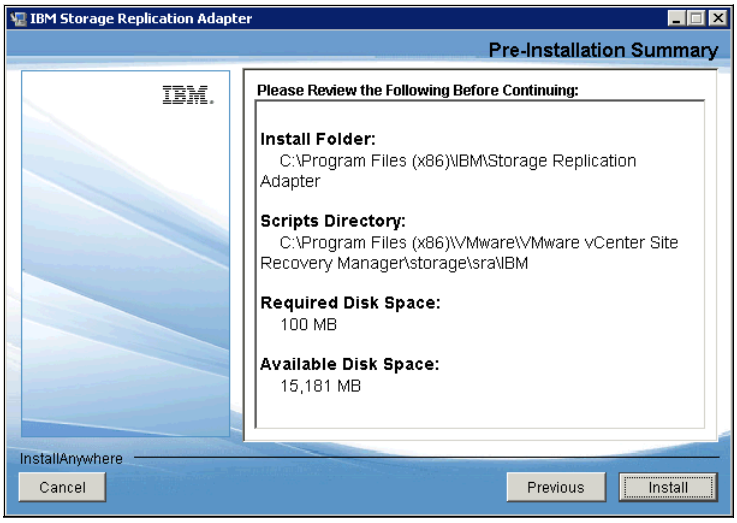

Figure 8-30 Review the installation parameters

7. After the installation completes, click **Done** as shown in Figure 8-31.

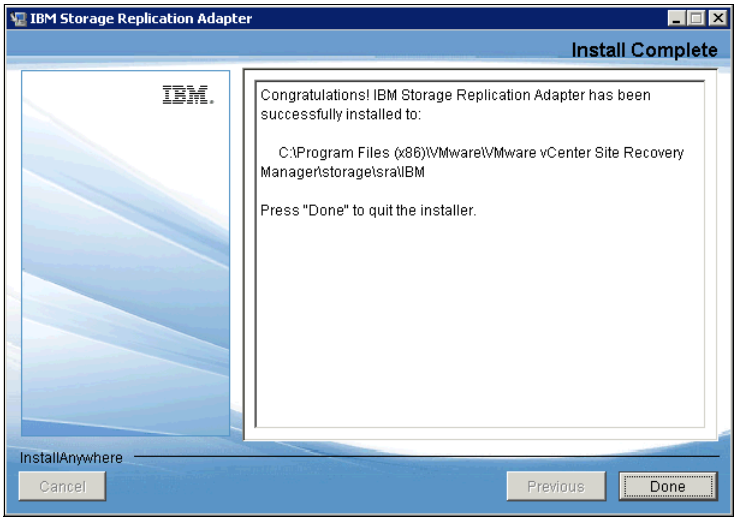

Figure 8-31 Completed SRA installer

8.3.3 Configuring SRAs

After the SRA is installed at both locations, open a vSphere Client connection to one of the vCenter Servers to configure the SRA. Follow these steps:

1. Click **Site Recovery** on the Home window to access the Site Recovery Manager interface.
2. Click **Array Managers** and then select the **SRAs** tab. Click **Rescan SRAs** to load the SRA into SRM as shown in Figure 8-32 on page 198.

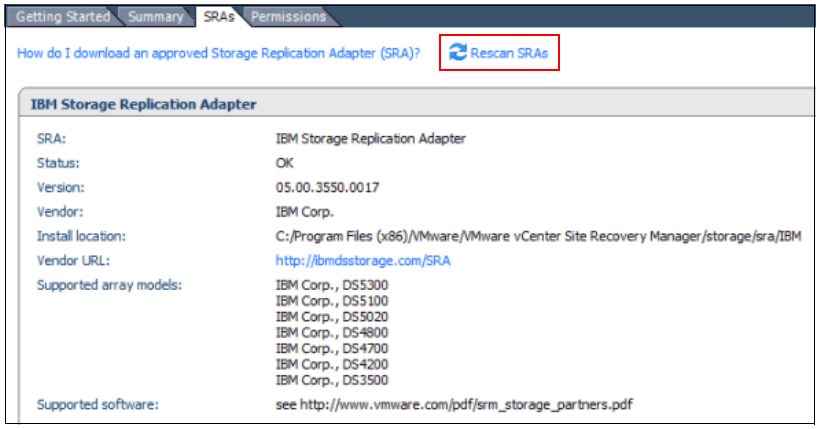

Figure 8-32 Rescan SRAs

3. Select the Site Recovery folder for the peer location and repeat step 2 on page 197.
4. After the rescan completes, the Storage Replication Adapter for DS3000, DS4000, and DS5000 is displayed in the SRM window. We can review the status information as shown in Figure 8-33.

Figure 8-33 Rescan SRA status

8.3.4 Adding the array manager

Add the array managers to your configuration. *Array managers* are the storage that we plan to use in the configuration to support the SRM implementation. As a prerequisite, we need the IP addresses of the controllers (array management controller management information) in advance. Follow these steps:

1. Click **Array Managers**, and select the folder in which you want to configure array managers.
2. Click **Add Array Manager** as shown in Figure 8-34 on page 199.

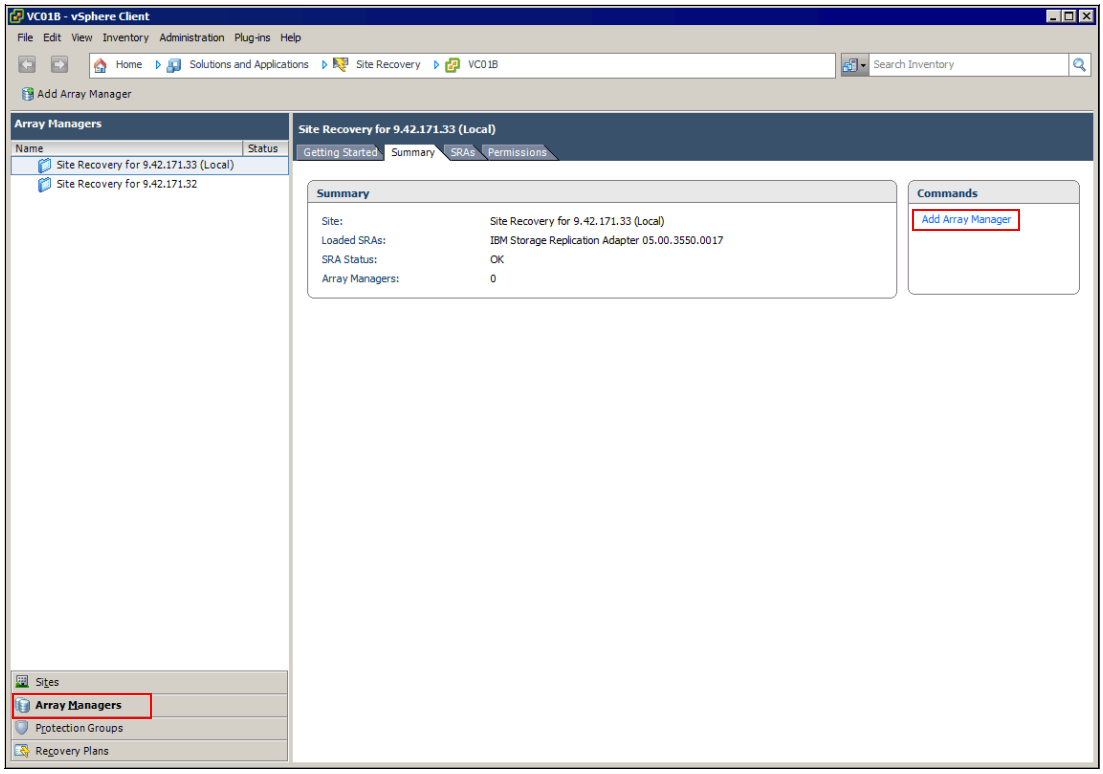

Figure 8-34 Add Array Manager

3. Provide an SRA name and select an adapter type as shown in Figure 8-35.

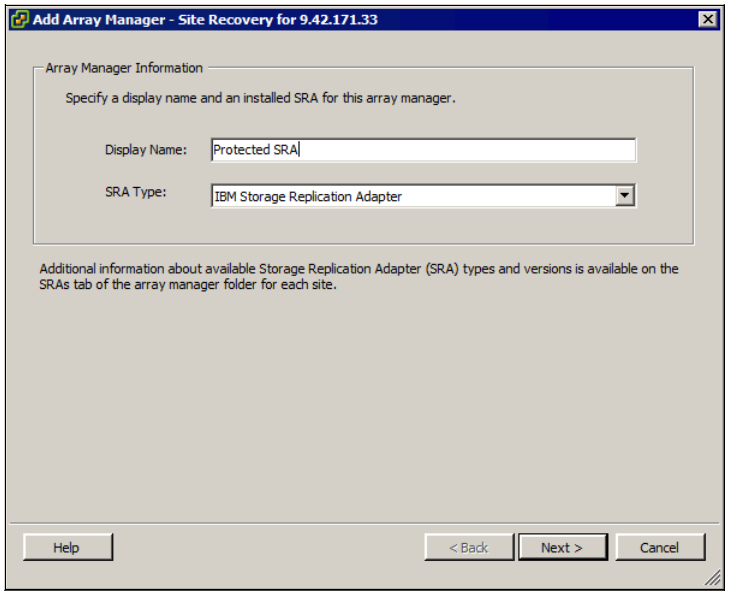

Figure 8-35 Array Manager Information

4. Provide the Storage Subsystem connection parameters and the peer Storage Subsystem connection parameters for the selected SRA as shown in Figure 8-36 on page 200.

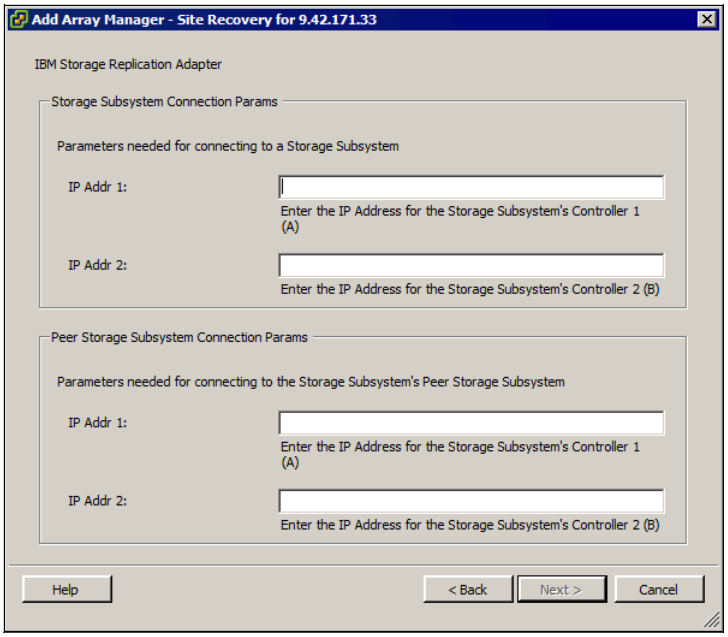

Figure 8-36 Storage Subsystem Connection Parameters

5. After you add the array manager successfully, click **Finish** as shown in Figure 8-37.

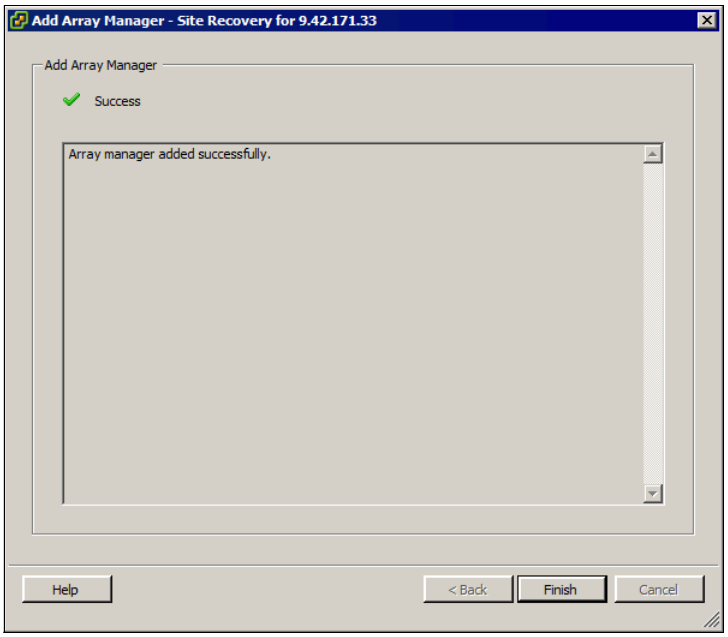

Figure 8-37 Finish the wizard

6. Repeat step 1 on page 198 through step 5 to configure an array manager for the recovery site.
7. Review the Array Managers section to validate the configuration for both sites as shown in Figure 8-38 on page 201.

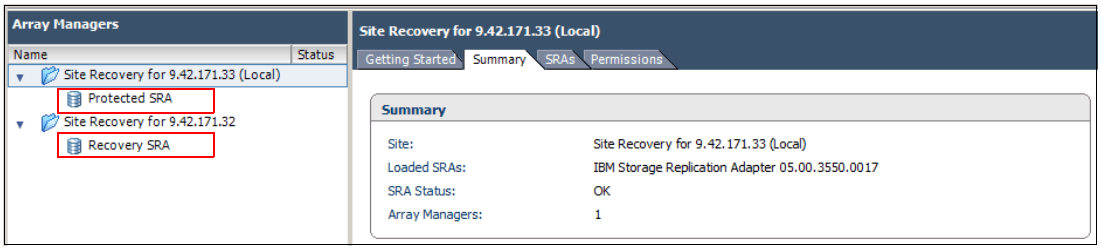

Figure 8-38 Checking array managers

8. On the Array Pairs tab, select an array pair, and click **Enable** as shown in Figure 8-39.

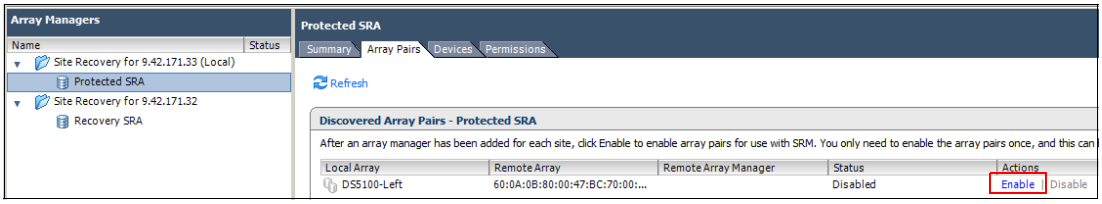

Figure 8-39 Enable Array Pairs

9. Navigate to the **Devices** tab on both the protected and recovery sites to verify the array pair directions as shown in Figure 8-40 and Figure 8-41 on page 202.

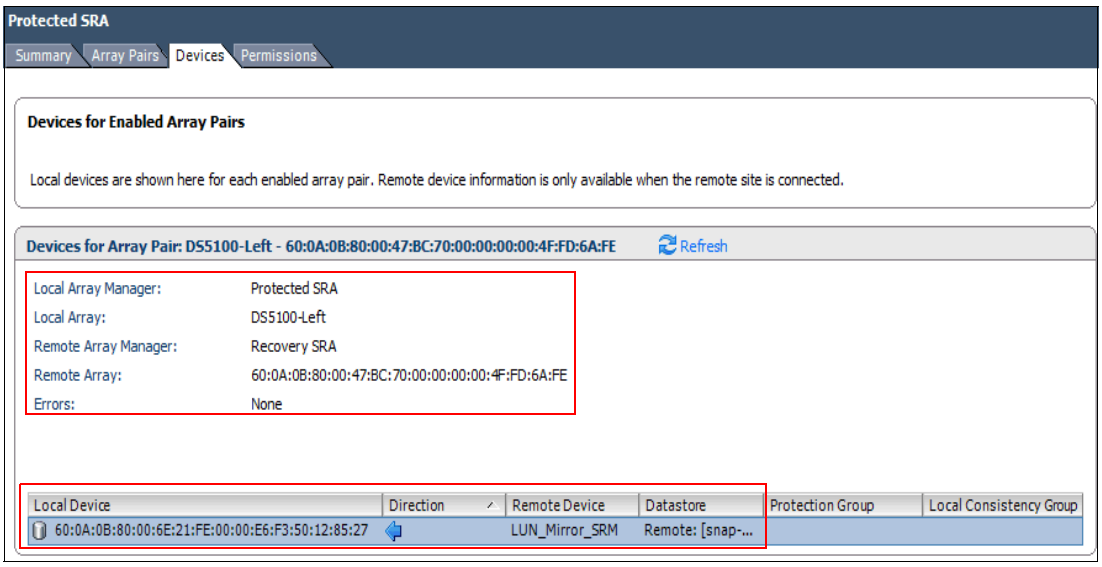

Figure 8-40 Array pair direction: Protected site

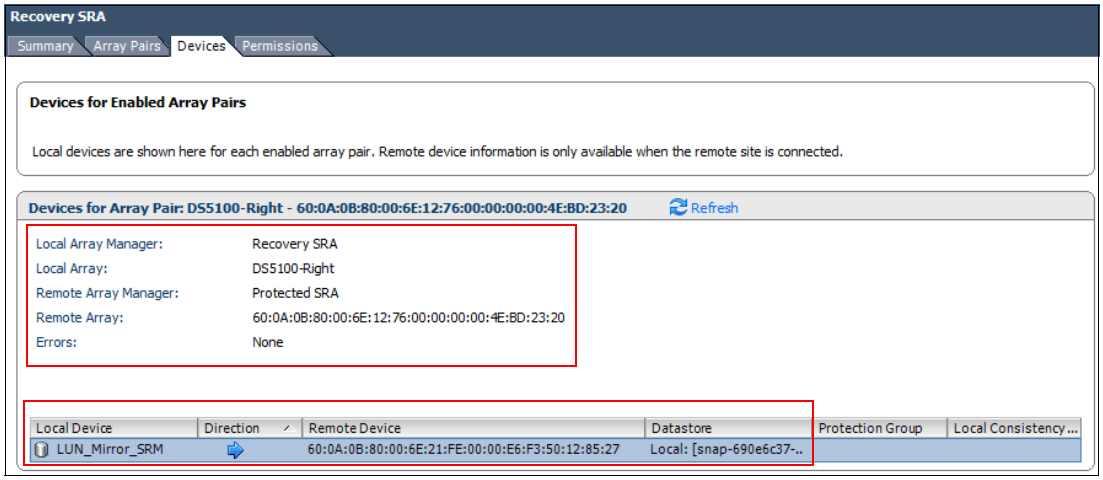

Figure 8-41 Array pair direction: Recovery site

Note: If array managers are added, but no array pairs are visible, click **Refresh** to collect the latest information about array pairs.

For more information about the SRA installation and implementation best practices, see the documentation that is included with the installation package (Figure 8-42).

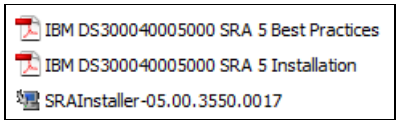

Figure 8-42 SRA documentation

8.4 Configuring Site Recovery Manager 5 (SRM)

We describe the following topics:

- Configuring mappings
- Configuring placeholder datastores
- Configuring a protection group
- Planning recovering plans

8.4.1 Configuring mappings

Mappings provide default locations and networks for use when placeholder virtual machines are initially created on the recovery site. *Inventory mappings* provide a convenient way to specify how resources at the protected site are mapped to resources at the recovery site. These mappings are applied to all members of a protection group when the group is created, and can be reapplied as needed. These mappings can be manually configured on each protected virtual machine as well.

Configuring resource mappings

Follow these steps to configure resource mappings:

1. Ensure that the protected site **9.42.171.33 - vc01b** is selected.
2. Click the **Resource Mapping** tab.
3. Expand the Protected Site Resource inventory and select the **Cluster A - Local** cluster as shown in Figure 8-43. Click the **Configure Mapping** action link.

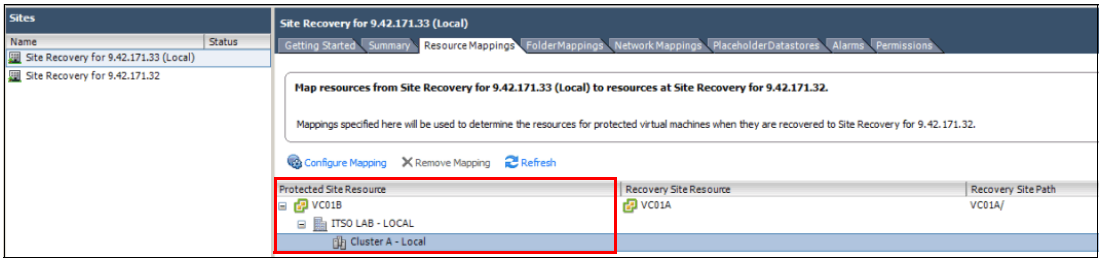

Figure 8-43 Configuring resource mappings

4. A new window opens for you to select the recovery site resource pool as shown in Figure 8-44. Select **9.42.171.30 - vc01a**. Click **OK** to continue.

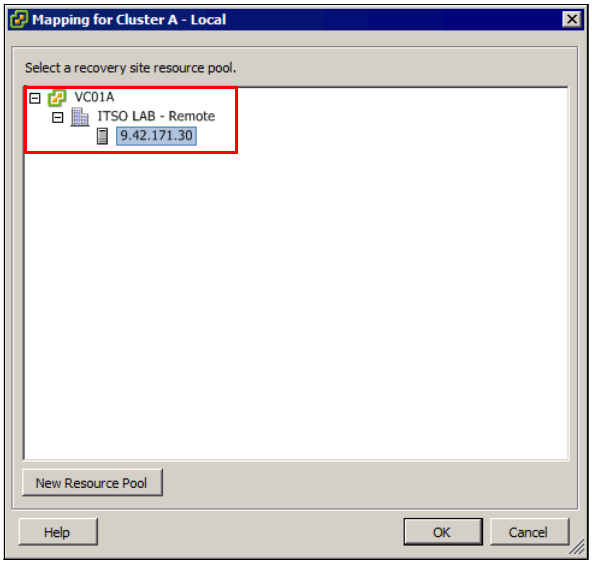

Figure 8-44 Configuring resource mappings

By configuring this mapping, you instruct SRM to use the `9.42.171.30 - vc01a` resource pool (the host in this case) during VM recovery at the secondary site.

8.4.2 Configuring placeholder datastores

For every virtual machine in a protection group, SRM creates a placeholder virtual machine at the recovery site. You must identify a datastore on the recovery site in which SRM can store the placeholder virtual machines.

After you select the datastore to contain the placeholder virtual machines, SRM reserves a place for protected virtual machines in the inventory on the recovery site. SRM creates a set of virtual machine files on the specified datastore at the recovery site and uses that subset to register the placeholder virtual machine with vCenter Server on the recovery site.

When using array-based replication as we do in this scenario to enable planned migration and reprotect, you must select placeholder datastores at both sites.

Placeholder datastores must meet certain criteria:

- For clusters, the placeholder datastores must be visible to all of the hosts in the cluster.
- You cannot replicate placeholder datastores.

For this example, we have two small datastores, `Recovery_SRM_Placeholder` and `Protected_SRM_Placeholder`, reserved on each site.

Follow these steps:

1. Click **Sites** in the left pane, and select a site.
2. Select the **Placeholder Datastores** tab and click **Configure Placeholder Datastore** as shown in Figure 8-45.

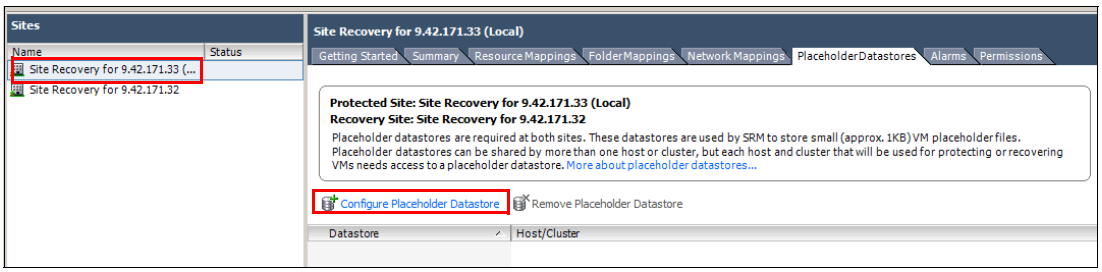

Figure 8-45 Configure Placeholder Datastore in the protected site

3. Expand the folders to find a datastore as shown in Figure 8-46 to designate as the location for placeholder datastores, click the datastore, and click **OK**.

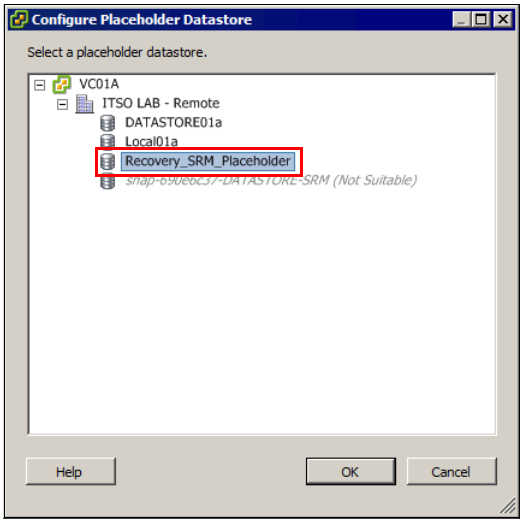

Figure 8-46 Adding a placeholder datastore

4. Review the recently added placeholder datastore on the Placeholder Datastores tab as shown in Figure 8-47 on page 205.

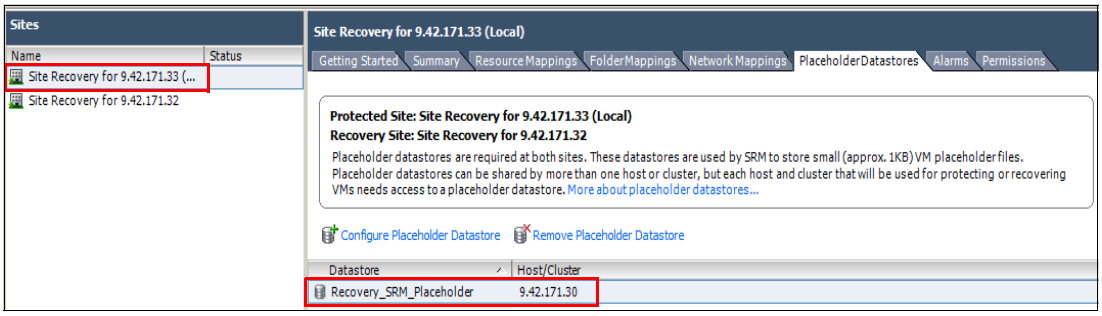

Figure 8-47 Checking placeholder datastore in the protected site

5. Repeat step 1 through step 4 to configure a placeholder datastore for the recovery site. The final configuration is similar to Figure 8-48.

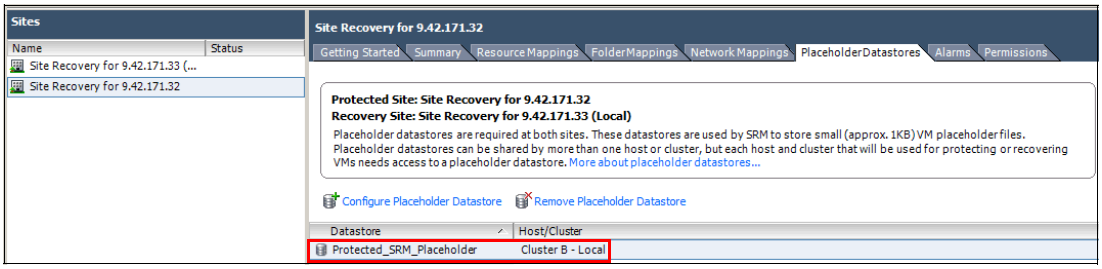

Figure 8-48 Checking placeholder datastore in the recovery site

8.4.3 Configuring a protection group

SRM organizes datastore groups to collect all files that are associated with protected virtual machines. You then associate these datastore groups with protection groups. All virtual machines in a datastore group replicate their files together, and all fail over together.

For this exercise, Site A (9.42.171.33 - vc01b) is the protected site. Array-based replication (SAN), which is provided by IBM ERM, will be used between the two sites. All the virtual machines that are stored on the replicated datastore can be included as part of the protection group.

Follow these steps:

1. Click **Protection Groups** and click **Create Protection Group** as shown in Figure 8-49.

Figure 8-49 Creating a protection group

2. On the Select Site and Protection Group Type page, select which site will be protected and select Array-Based Replication as shown in Figure 8-50 on page 206.

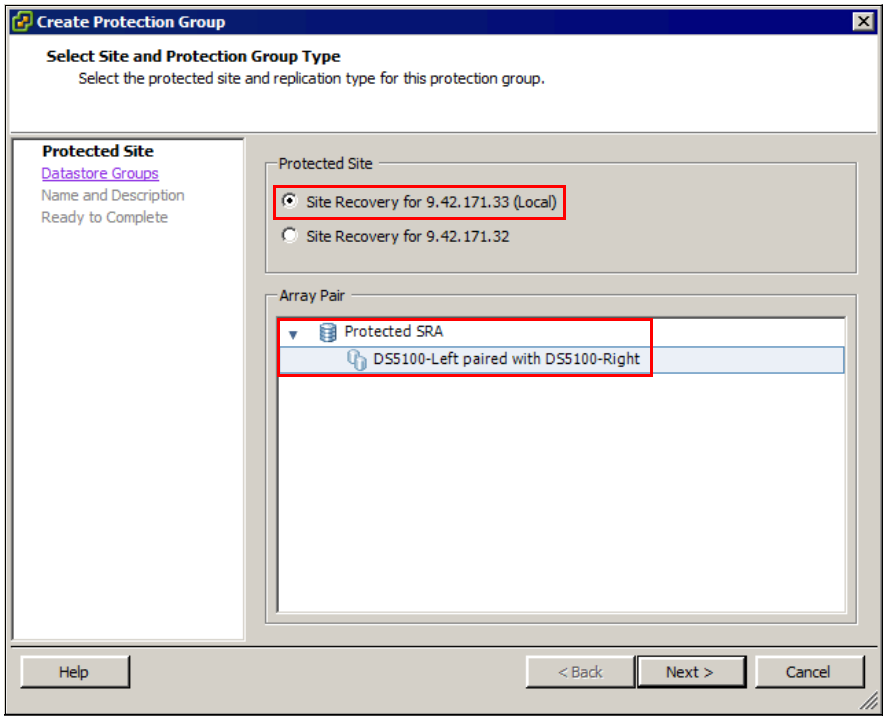

Figure 8-50 Protected Site

3. Select a datastore group from the list, and click **Next** as shown in Figure 8-51.

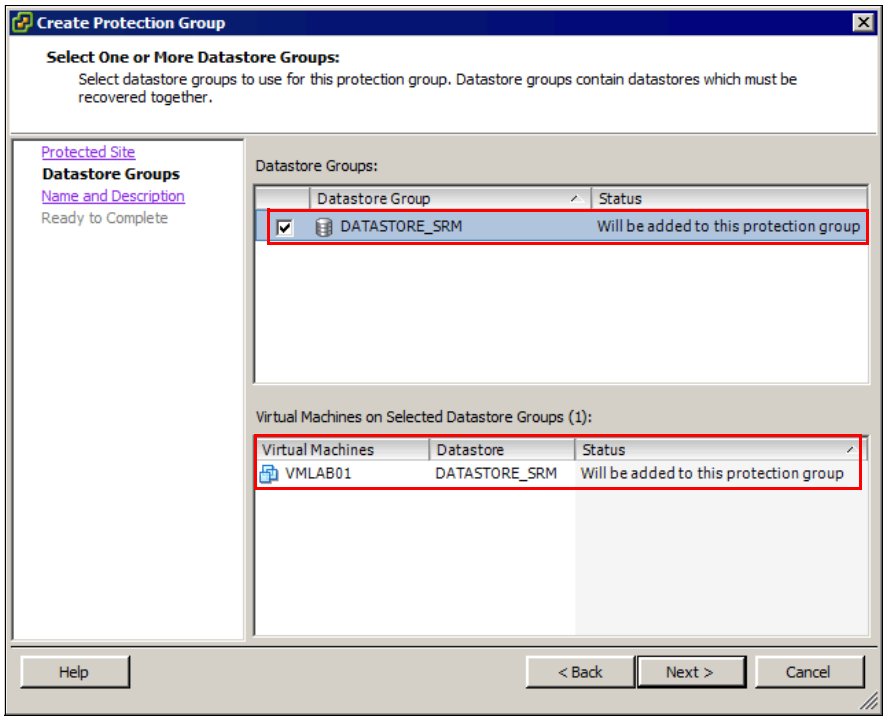

Figure 8-51 Datastore Groups

4. Type a name and optional description for the protection group, and click **Next** as shown in Figure 8-52 on page 207.

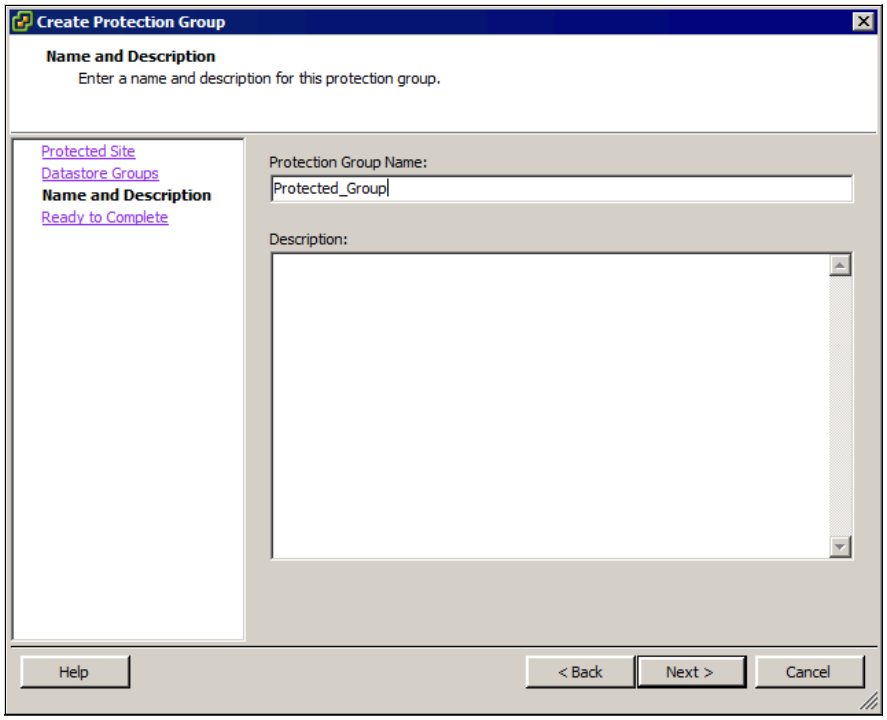

Figure 8-52 Name and Description

5. Click **Finish** to create the protection group and begin the automatic protection of the specified virtual machines as shown in Figure 8-53.

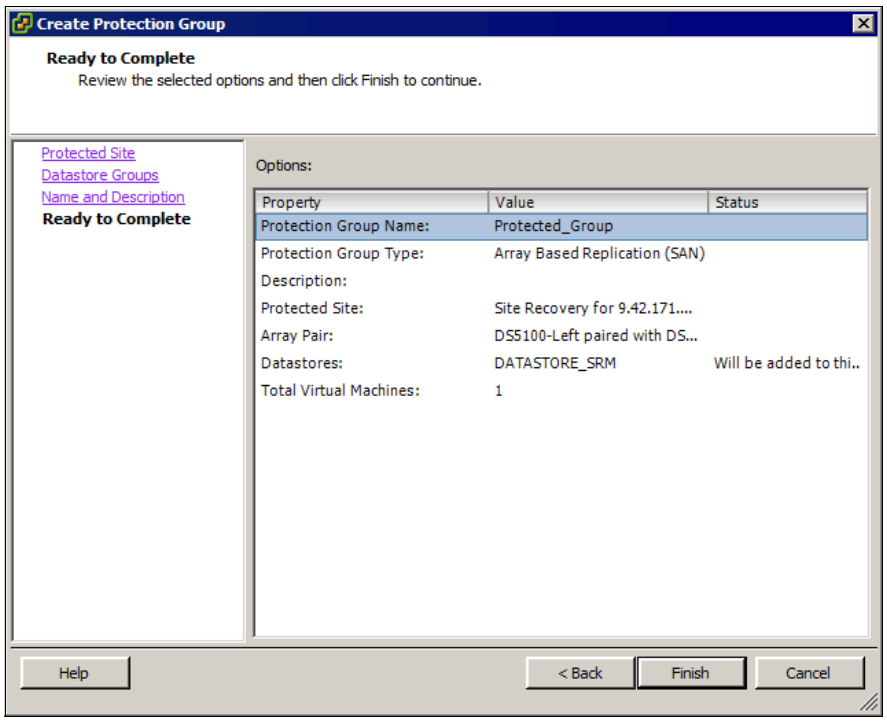

Figure 8-53 Ready to Complete: Review parameters

6. Review the recently added protected group. Select the **Virtual Machines** tab to check the protected virtual machines and status as shown in Figure 8-54 on page 208.

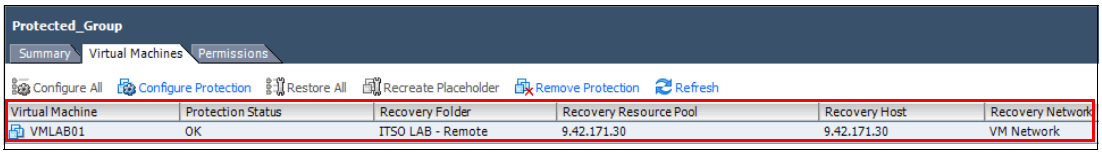

Figure 8-54 Virtual Machine Protection Status

SRM creates protection groups that you can use to protect virtual machines. After protection is established, placeholders are created and inventory mappings are applied for each virtual machine in the group. If a virtual machine cannot be mapped to a folder, network, and resource pool on the recovery site, it is listed with a status of Mapping Missing, and a placeholder is not created for it. Wait to ensure that the operations complete as expected. Ensure that the protection group was created and virtual machines were protected (as shown in Figure 8-54).The progress of those tasks can be monitored in the Recent Tasks panel of the vSphere Client.

8.4.4 Planning recovering plans

The following steps show you how to create a recovery plan to manage the process of moving our protected virtual machine (`VMLAB01`) workloads from the primary site (Site A - Protected) to the secondary site (Site B - Recovery).

Recovery plans are different from protection groups. By creating a recovery plan, you can manage the process of moving our protected VM workloads from your primary data center to the secondary data center and you can manage how virtual machines are recovered.

Recovery plans can be organized in four different stages or steps:

- **Test**: When you test a recovery plan, you use a test network and a temporary snapshot of replicated data at the recovery site. No operations are disrupted at the protected site. Testing a recovery plan runs all the steps in the plan except for powering down virtual machines at the protected site and forcing devices at the recovery site to assume mastership of replicated data.
- **Cleanup**: After you run a recovery plan to test, a clean-up task needs to be performed to power off each recovered/tested virtual machine. The clean-up task replaces recovered virtual machines with placeholders that preserve their identity and configuration information and cleans up replicated storage snapshots that were used by the recovered virtual machines during the test.
- **Recovery**: When you run a recovery plan, all virtual machines in the recovery plan are migrated to the recovery site. The corresponding virtual machines in the protected site are shut down. When you perform a planned migration, SRM attempts to replicate all virtual machines and to gracefully shut down the protected machines.
- **Reprotect**: After a recovery, the recovery site becomes the new production site and is not protected. If a disaster occurs at the new production site, no other site is available to fail over to. A best practice is to protect the new production site to some other site immediately after a recovery. If the original production site is operational, you can use the original production site as a new recovery site to protect the new production site, effectively reversing the direction of protection. Reestablishing protection in the opposite direction by re-creating all protection groups and recovery plans is time-consuming and error-prone.

Running a test

Follow these steps:

1. Click **Recovery Plans** in the left pane.
2. Click the recovery plan to test, and click **Test** as shown in Figure 8-55.

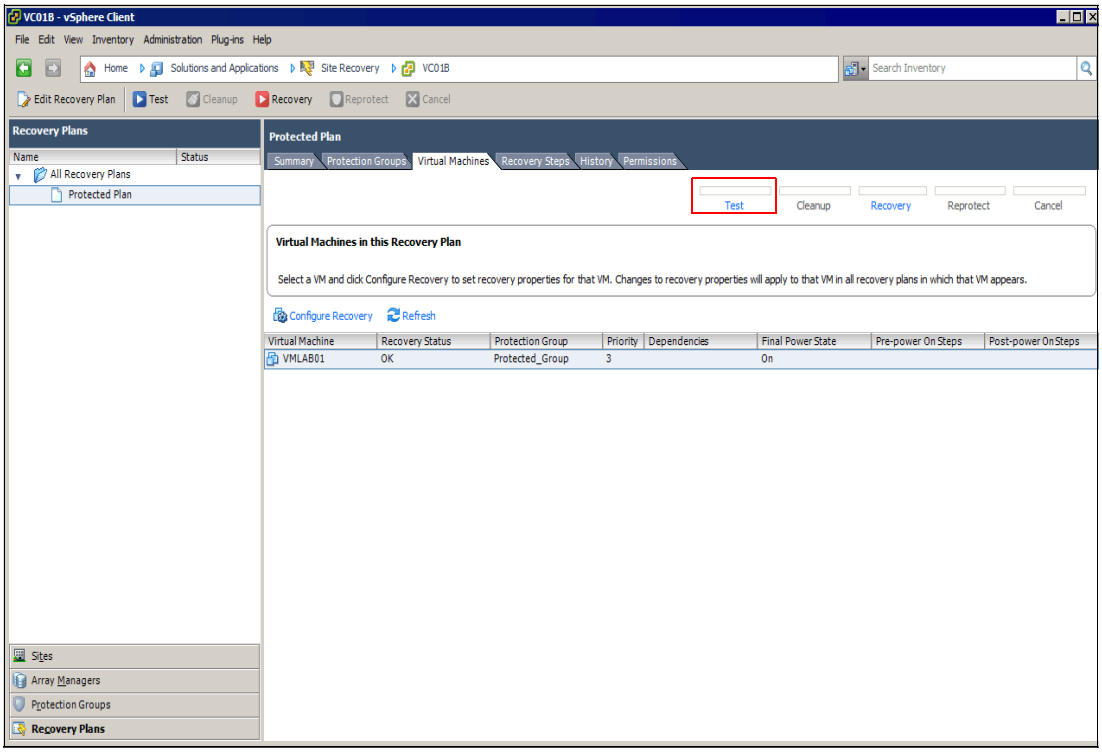

Figure 8-55 Starting a test recovery plan

3. Determine whether you want to select **Replicate recent changes to recovery site**. Click **Next** as shown in Figure 8-56 on page 210.

   Enabling this option ensures that the recovery site has the latest copy of protected virtual machines, but the synchronization might take additional time.

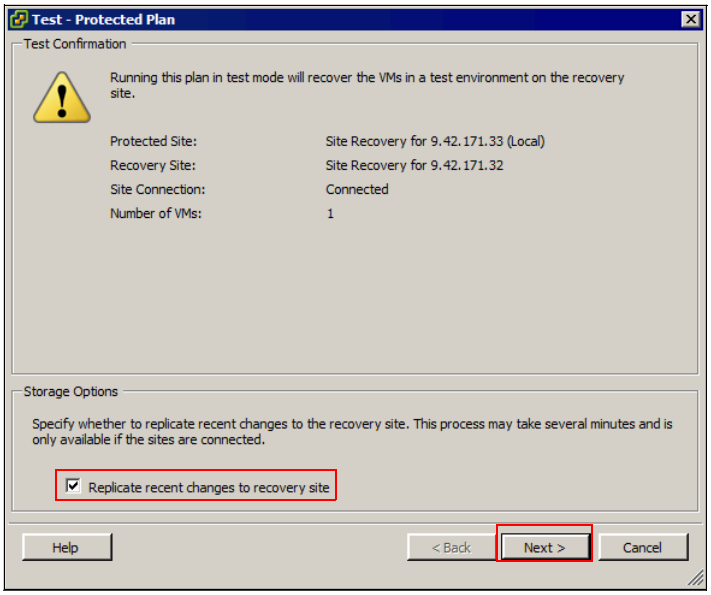

Figure 8-56 Test Confirmation

4. Review the confirmation window as shown Figure 8-57 and click **Start**.

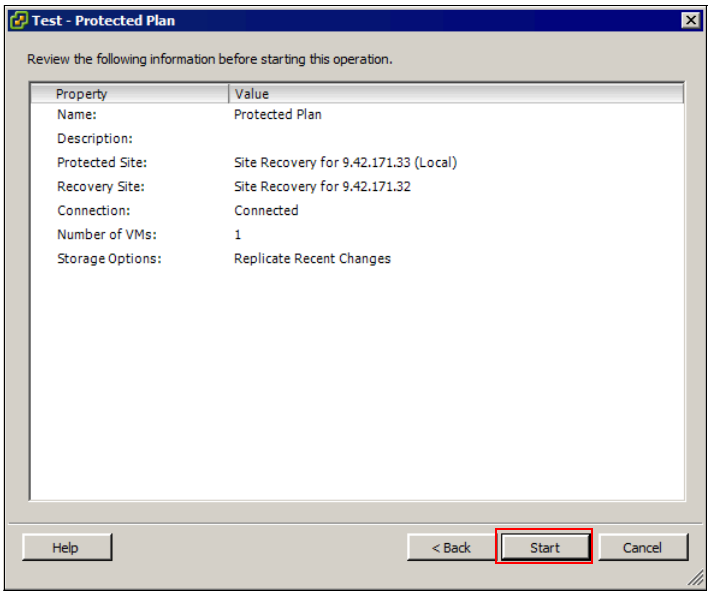

Figure 8-57 Reviewing information

5. The wizard closes and the recovery plan test begins.
6. Click the **Recovery Steps** tab to monitor the progress of the test and respond to messages as shown in Figure 8-58 on page 211.

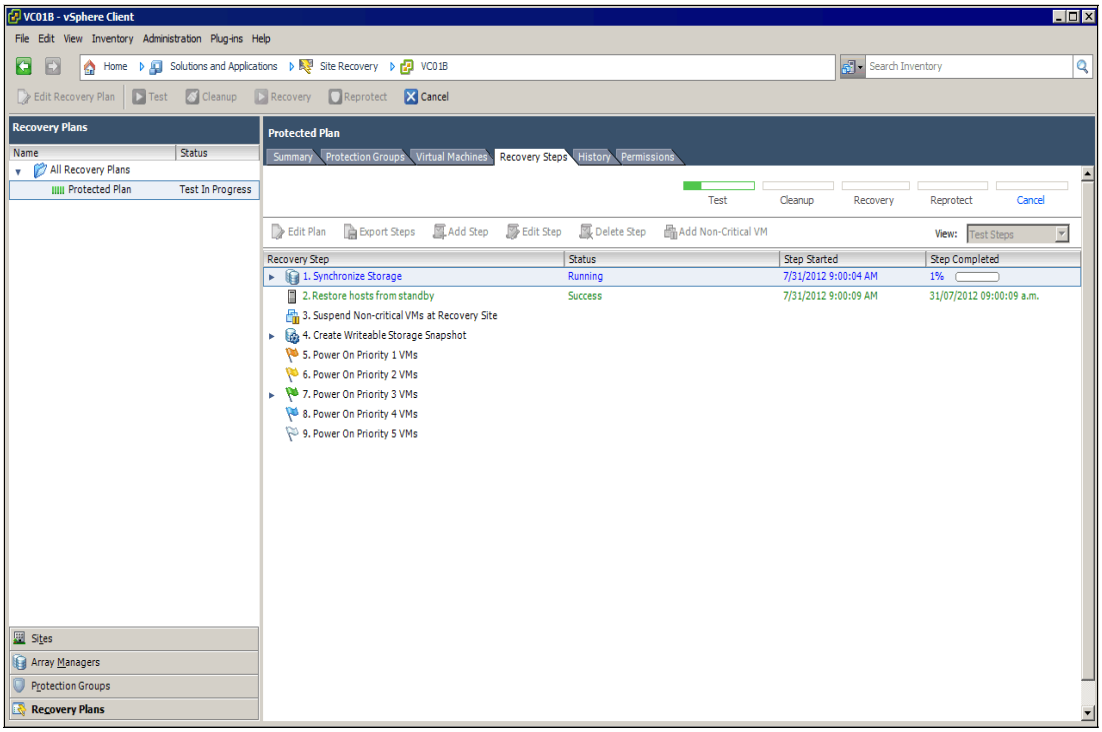

Figure 8-58 Monitoring the test

7. When the recovery plan test completes as shown in Figure 8-59, click **Cleanup**.

   Running the cleanup returns the protected virtual machines to their initial state and resets the recovery plan to the Ready state.

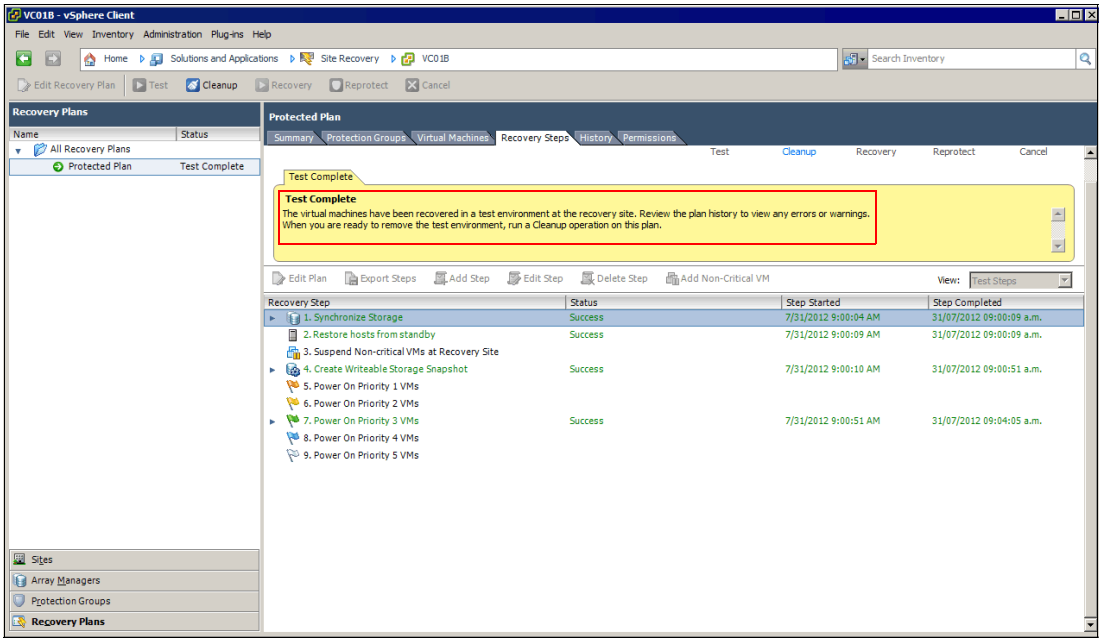

Figure 8-59 Review the completed test

Important: If the Cleanup process encounters errors, run the Cleanup process again and select the Force Cleanup option.

Running a recovery

Follow these steps:

1. Click **Recovery Plans** in the left pane and click the recovery plan that you want to run.
2. In the Commands area, click **Planned Migration**.
3. Review the information in the confirmation prompt, and select **I understand that this process will permanently alter the virtual machines and infrastructure of both the protect and recovery datacenters**.
4. Optional: If you enabled the forced failover function, you can select the “Forced Recovery - recovery site operations only” check box. Click **Next** as shown in Figure 8-60.

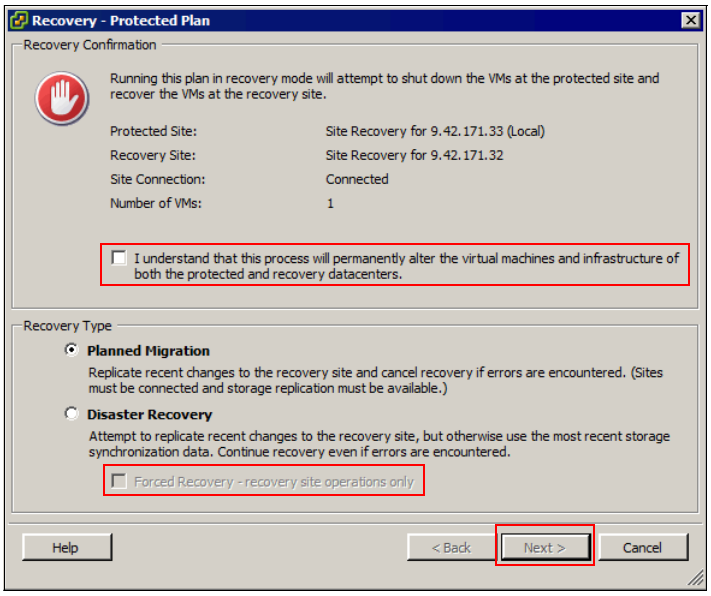

Figure 8-60 Test confirmation

5. Click **Start** to run the recovery plan as shown in Figure 8-61.

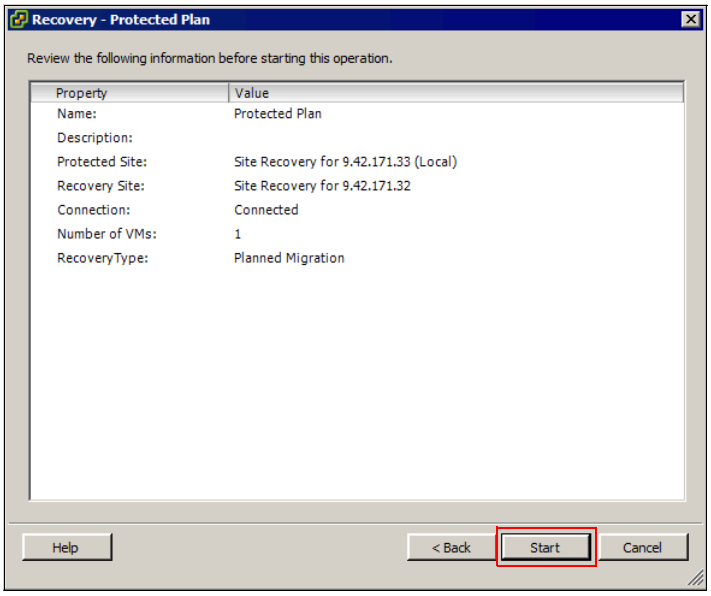

Figure 8-61 Review information

6. Click the **Recovery Steps** tab as shown in Figure 8-62.

   The Recovery Steps tab displays the progress of the individual steps. The recent tasks area reports the progress of the overall plan.

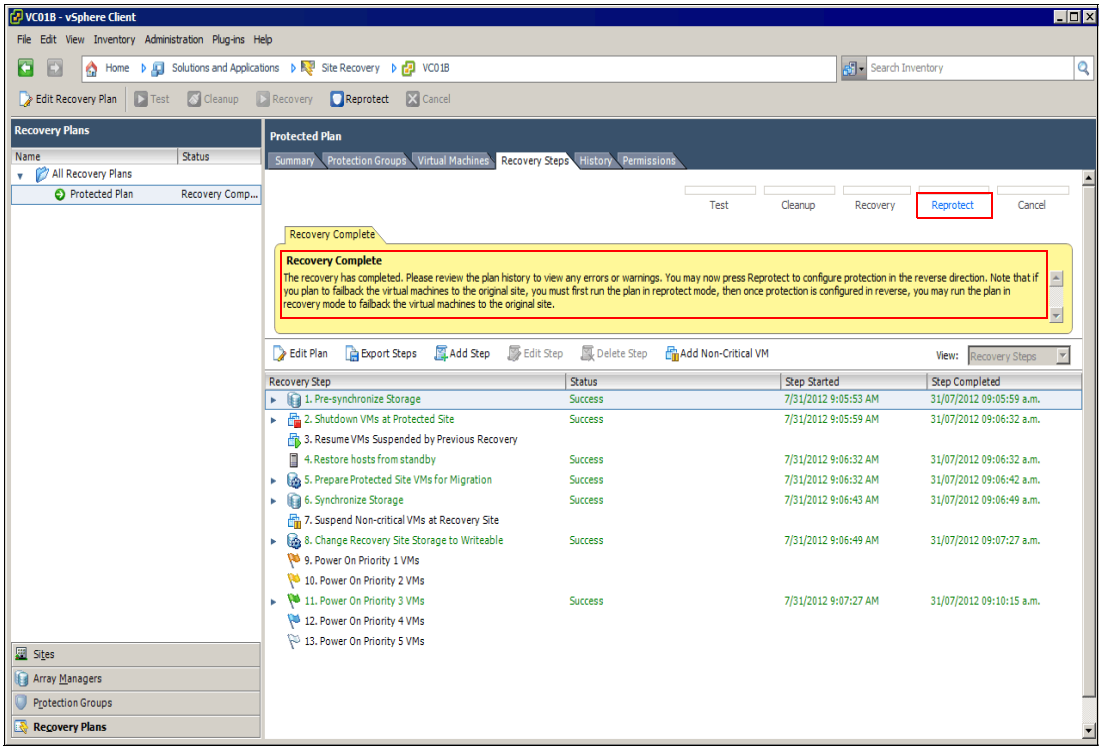

Figure 8-62 Review completed test

7. Click **Recovery Plans** in the left pane and click the recovery plan to run.
8. In the Commands area, click **Planned Migration**.
9. Review the information in the confirmation prompt, and select **I understand that this process will permanently alter the virtual machines and infrastructure of both the protect and recovery datacenters**.

Reprotecting virtual machines

Reprotection is available only in non-catastrophic failures. This means that the original vCenter servers, ESX Servers, SRM servers, and corresponding databases must be eventually recoverable.

For reprotect to be available, the following steps must first occur:

1. A recovery must be completed in which all steps finish successfully. If errors occurred during the recovery, the user must resolve the problems that caused the errors and rerun the recovery. When you rerun a recovery, operations that succeeded previously are skipped. For example, successfully failed over virtual machines are not failed over again and they continue running without interruption.
2. The original site must be available and SRM servers at both sites must be in a connected state.
3. The recovery plan runs in the reprotection mode to reverse the replication direction for the underlying arrays.

Important: Reprotect is supported only for array-based replication. The scenario that we present in this chapter uses array-based replication that is provided by the SRA extensions and IBM ERM premium feature.

Reprotecting process

Follow these steps:

1. Validate the existence of the virtual machine `VMLAB01` on the protected site as shown in Figure 8-63.

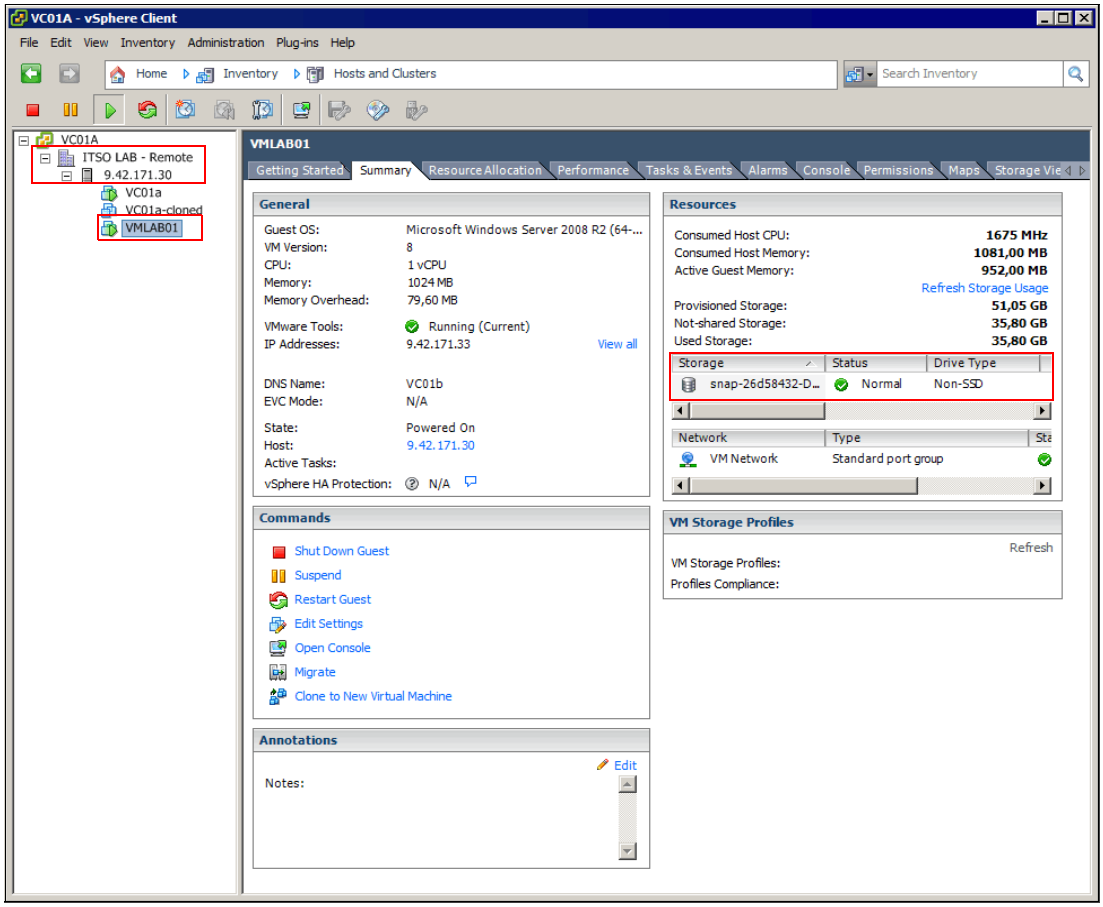

Figure 8-63 Validate the existence of the virtual machine on the recovery site

2. Click Recovery Plans in the left pane and click the recovery plan to run.
3. In the Commands area, click **Reprotect** as shown in Figure 8-62 on page 213.
4. Review the information in the confirmation prompt, and select **I understand that this operation cannot be undone** as shown in Figure 8-64 on page 215. Click **Next**.

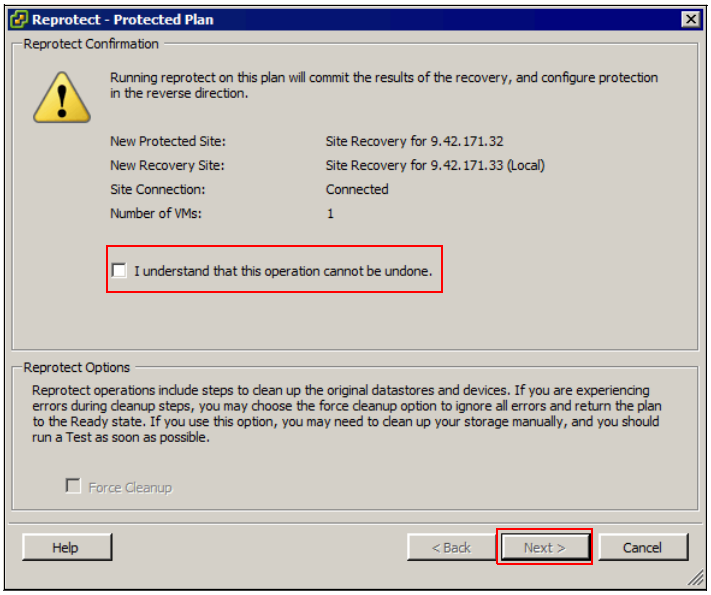

Figure 8-64 Reprotect Confirmation

5. Click **Start** to run the recovery plan as shown in Figure 8-65.

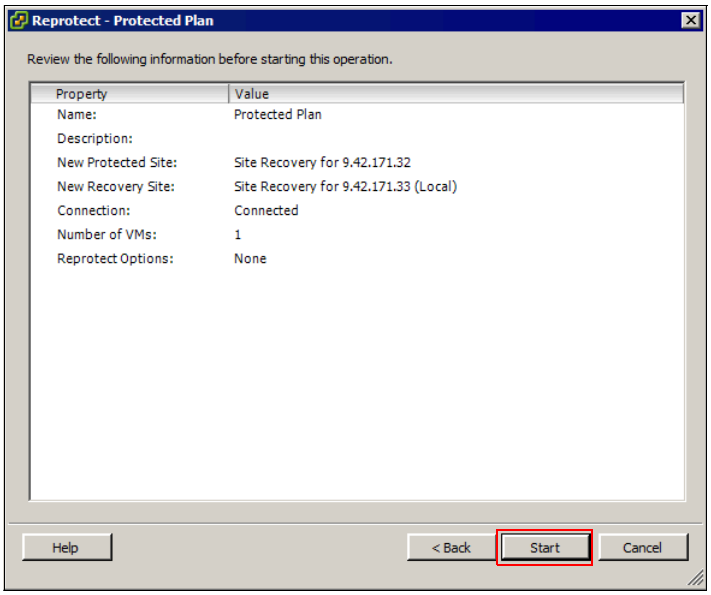

Figure 8-65 Review information

6. We can follow the entire reprotect process on Figure 8-66, Figure 8-67 on page 216, and Figure 8-68 on page 216.

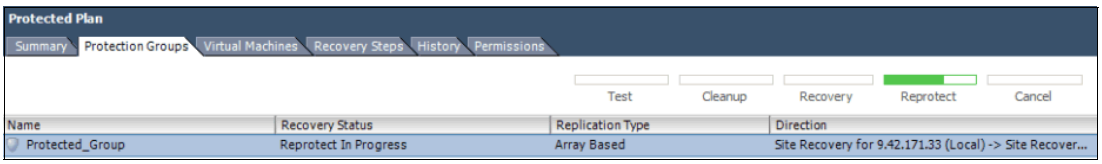

Figure 8-66 Reprotect status: Protection Groups tab

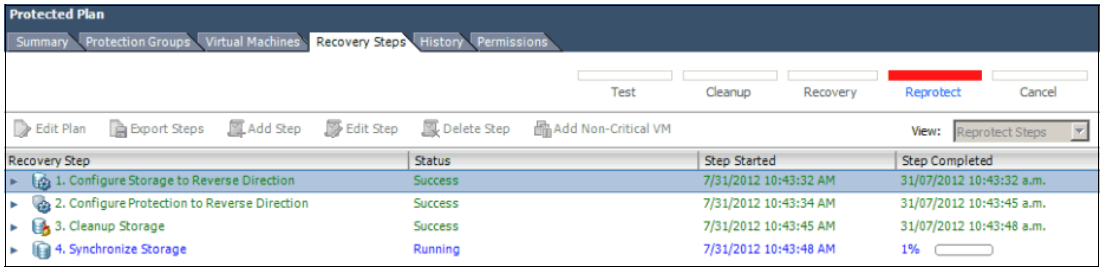

Figure 8-67 Reprotect status: Recovery Steps tab

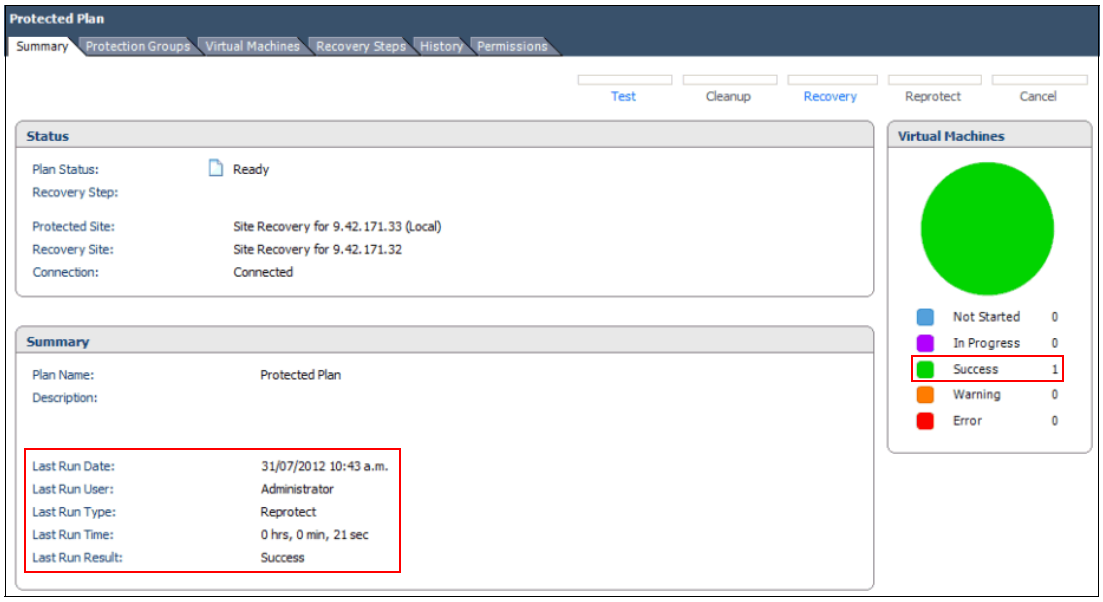

Figure 8-68 Reprotect status: Summary tab

9

VMware vSphere ESXi 5 advanced configuration and new features

This chapter describes the advanced settings and new features of VMware vSphere ESXi 5.0 to enhance your virtual environment with IBM Midrange Storage solutions.

VMware vSphere ESXi 5.0 brings many new capabilities to extend the benefits of previous versions. These new features and enhancements to core capabilities provide more performance optimization and easier provisioning, monitoring, and troubleshooting.
This chapter focuses on the storage-specific features and enhancements that are available in ESXi 5.0 and provides an overview of how they optimize storage utilization, ease monitoring, and increase operational efficiency. Wherever possible, we also provide use cases and requirements that might apply to these new functions.

The following topics are covered in this chapter:

- vSphere VMFS-5
- Profile Driven Storage (storage profiles)
- vSphere Storage vMotion (Storage vMotion)
- vSphere Storage Distributed Resource Scheduler (Storage DRS)
- vSphere Storage I/O Control

This chapter provides a technical overview of new capabilities and enhancements, as well as links to additional information about each of these new storage features.

9.1 vSphere VMFS-5

vSphere VMFS-5 was introduced in vSphere 5.0.

9.1.1 VMFS-5 enhancements

vSphere 5.0 introduces a new version of the vSphere Virtual Machine File System (VMFS), VMFS-5, which contains many important architectural changes that enable greater scalability and performance while reducing complexity. Although numerous fundamental changes exist, the following enhancements are significant from operational and architectural aspects:

- Unified 1 MB file block size: Previous versions of VMFS used 1 MB, 2 MB, 4 MB, or 8 MB file blocks. These larger blocks were needed to create large files (>256 GB). These large blocks are no longer needed for large files on VMFS-5. Very large files can now be created on VMFS-5 by using 1 MB file blocks.
- Large single extent volumes: In previous versions of VMFS, the largest single extent was 2 TB. With VMFS-5, this limit is increased to approximately 60 TB.
- Smaller subblock: VMFS-5 introduces a smaller subblock, which is now 8 KB rather than the 64 KB in previous versions. Now, small files < 8 KB (but > 1 KB) only consume 8 KB rather than 64 KB, which reduces the amount of disk space that is stranded by small files.
- Small file support: VMFS-5 introduces support for very small files. For files less than or equal to 1 KB, VMFS-5 uses the file descriptor location in the metadata for storage rather than file blocks. When they grow larger than 1 KB, these files then start to use the new 8 KB subblocks. Using the new 8 KB subblocks again reduces the amount of disk space that is stranded by very small files.
- Increased file count: VMFS-5 introduces support for greater than 100,000 files, which is a three-fold increase on the number of files supported on VMFS-3, which was approximately 30,000.
- ATS enhancement: This Hardware Acceleration primitive, Atomic Test and Set (ATS), is now used throughout VMFS-5 for file locking. ATS is part of the vSphere Storage APIs for Array Integration (VAAI). This enhancement improves the file-locking performance over previous versions of VMFS.

Table 9-1depicts the most significant architectural changes for VMFS-5 in comparison to VMFS-3.

Table 9-1 VMFS-3 and VMFS-5 comparison

Feature	VMFS-3	VMFS-5
Support for larger single-extent VMFS volumes up to 64 TB	No	Yes
Support for pass-through RDMs larger than 2 TB	No	Yes
Subblock for space efficiency	Yes (64 KB, maximum of approximately 3K)	Yes (8 KB, maximum of approximately 30K)
Small-file support	No	Yes (1 KB)
Unified block size	No	Yes

9.1.2 VMFS-5 considerations

Consider these points before you migrate the VMFS-3 datastore to the new VMFS-5 file system:

- The process to upgrade from VMFS-3 to VMFS-5 is an *online and nondisruptive* upgrade operation; therefore, the virtual machines can continue to run on the datastore.
- Upgraded VMFS-5 can use the new 1 KB small-files feature.
- Upgraded VMFS-5 can grow to approximately 60 TB, which is the same as a new VMFS-5.
- Upgraded VMFS-5 has all the VAAI ATS improvements of a new VMFS-5.

Be aware of the differences between *new* datastores and *upgraded* datastores:

- A VMFS-5 that is upgraded from VMFS-3 continues to use the previous file block size, which might be larger than the unified 1 MB file block size.
- A VMFS-5 that is upgraded from VMFS-3 continues to use 64 KB subblocks and not new 8 KB subblocks.
- A VMFS-5 that is upgraded from VMFS-3 continues to have a file limit of 30,720 rather than the new file limit of greater than 100,000 files for the new VMFS-5.
- A VMFS-5 that is upgraded from VMFS-3 continues to use the Master Boot Record (MBR) partition type. When the VMFS-5 volume is grown above 2 TB, it automatically and seamlessly switches from MBR to GUID Partition Table (GPT) with no impact to the running virtual machines.
- A VMFS-5 that is upgraded from VMFS-3 continues to start its partition on sector 128. New VMFS-5 partitions start their partition at sector 2048.

Note: We recommend creating a VMFS-5 file system rather than upgrading VMFS-3 to VMFS-5. Storage vMotion operations can then be used to seamlessly move your virtual machines to the new VMFS-5. This way, you enjoy all the benefits of VMFS-5.

Also, if you use RDM disks, consider this information:

- There is now support for pass-through RDMs of up to approximately 60 TB.
- Non-pass-through RDMs are still limited to 2 TB minus 512 bytes.
- Both upgraded VMFS-5s and new VMFS-5s support the larger pass-through RDMs.

For more information about the maximum values for VMFS-5, see this link:

http://www.vmware.com/pdf/vsphere5/r50/vsphere-50-configuration-maximums.pdf

9.1.3 Creating a VMFS-5 datastore step-by-step

In this section, we show how to create a VMFS-5 datastore in a few steps. You are required to use an existing LUN to start the creation process.

Follow these steps:

1. Use the vSphere user interface, select the ESXi host, click the **Configuration** tab, and click **Add Storage** as shown in Figure 9-1 on page 220.

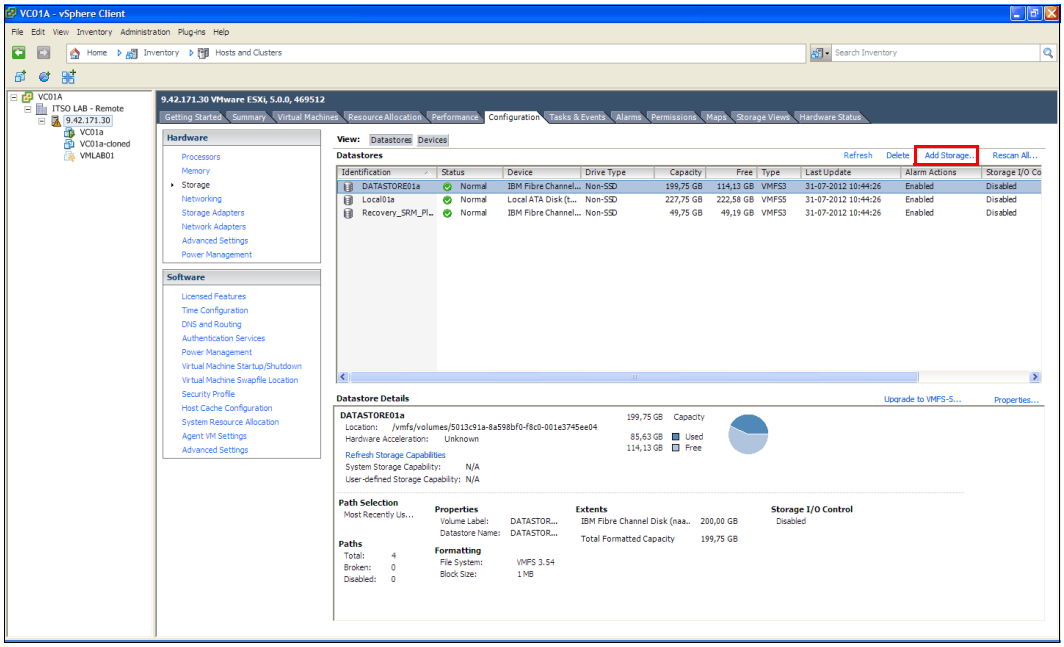

Figure 9-1 Preparing to add a VMFS-5 datastore

2. In the next window (Figure 9-2), we select the storage type to add. The options are Disk/LUN (SAN) or Network File System (NAS).

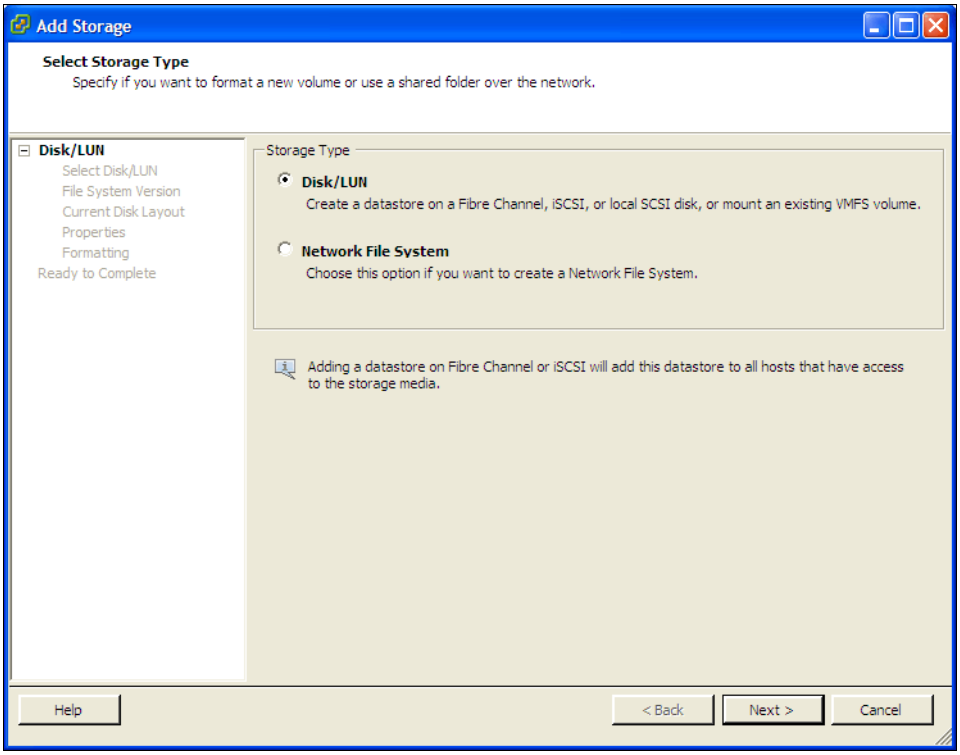

Figure 9-2 Selecting the type of storage to add

3. In this case, we select **Disk/LUN** and click **Next**.
4. Select the LUN (that was created before) to add to your environment and click **Next** (Figure 9-3 on page 221).

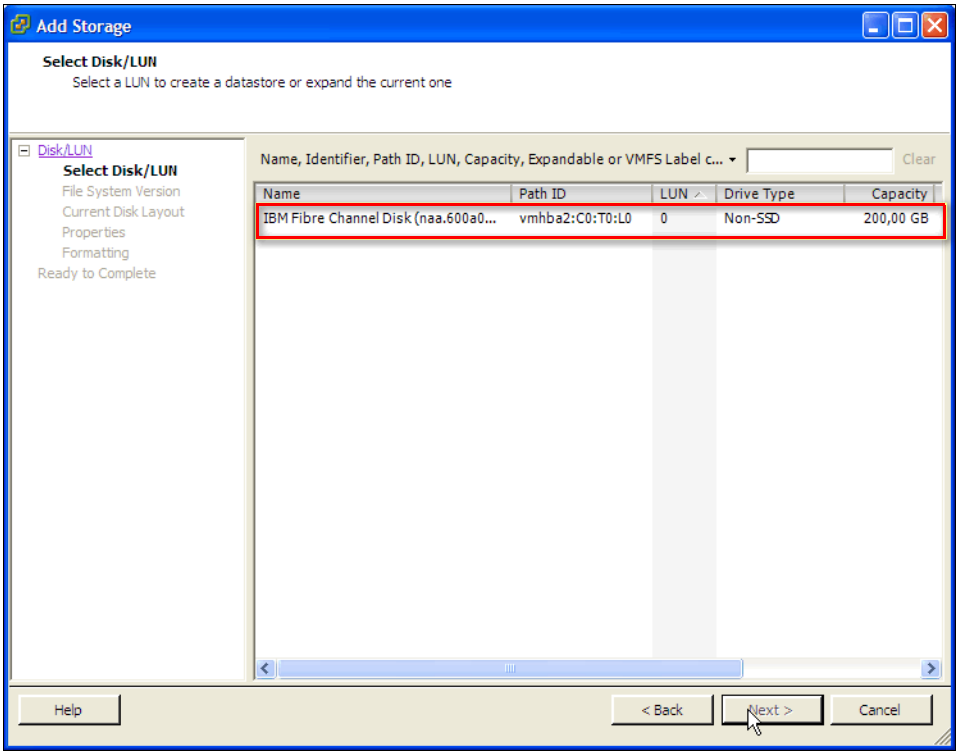

Figure 9-3 Selecting a LUN

5. Select the file system version (as shown in Figure 9-4). We chose **VMFS-5** and click **Next**.

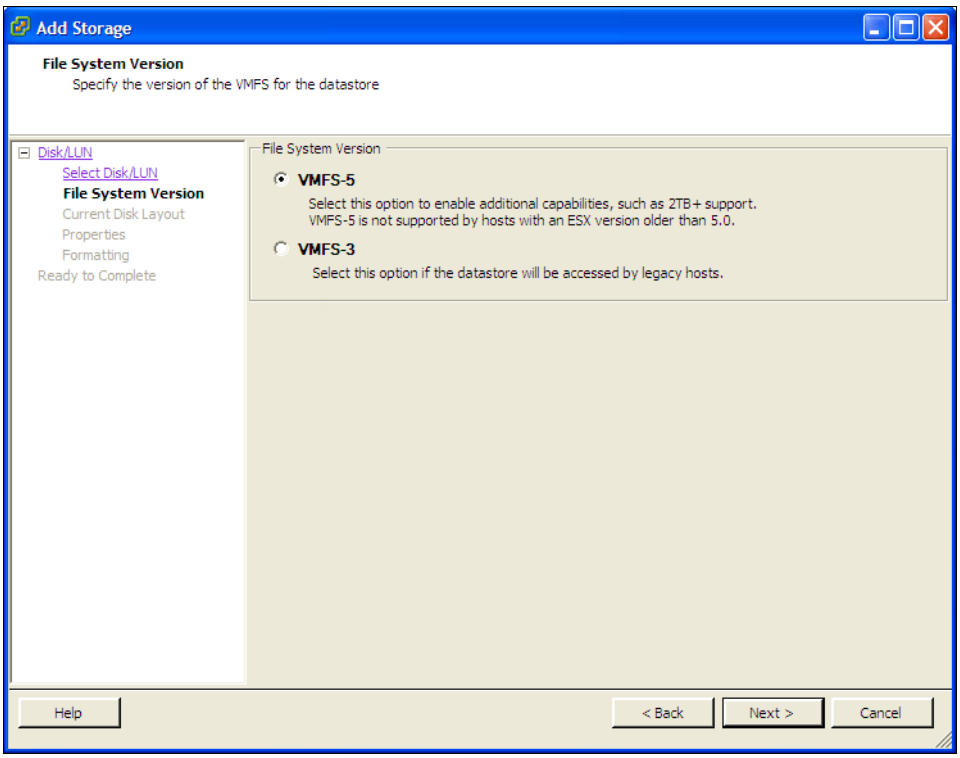

Figure 9-4 Selecting the file system version

6. On the Current Disk Layout window (Figure 9-5), verify the LUN that was chosen in the previous step. Click **Next**.

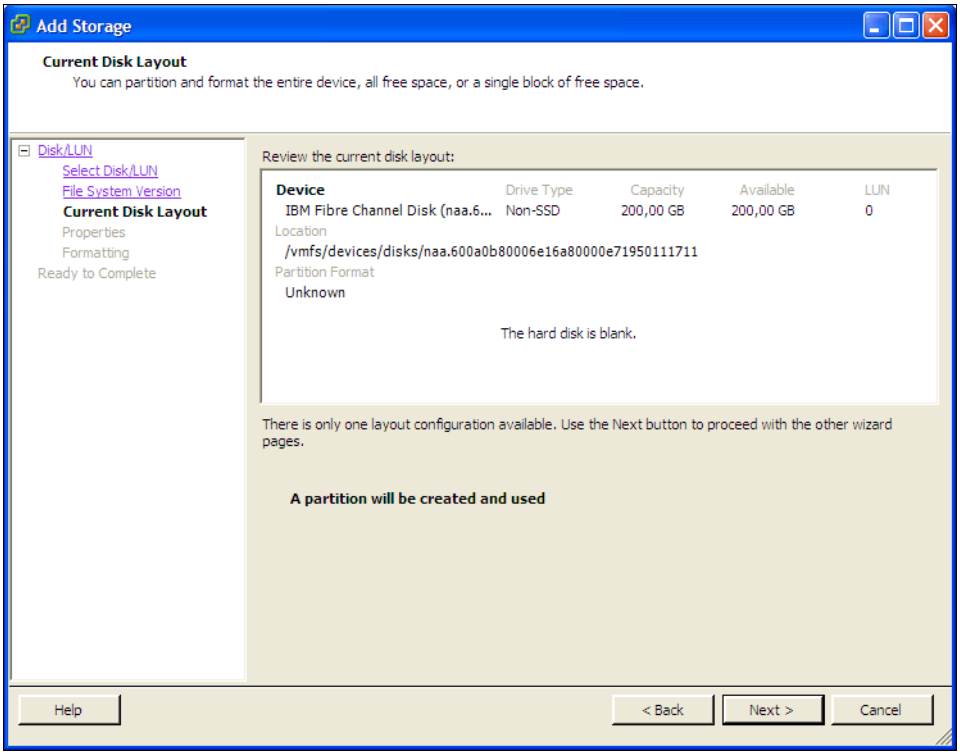

Figure 9-5 Current Disk Layout

7. Assign a name to the new VMFS-5 datastore and click **Next** as shown in Figure 9-6 on page 223.

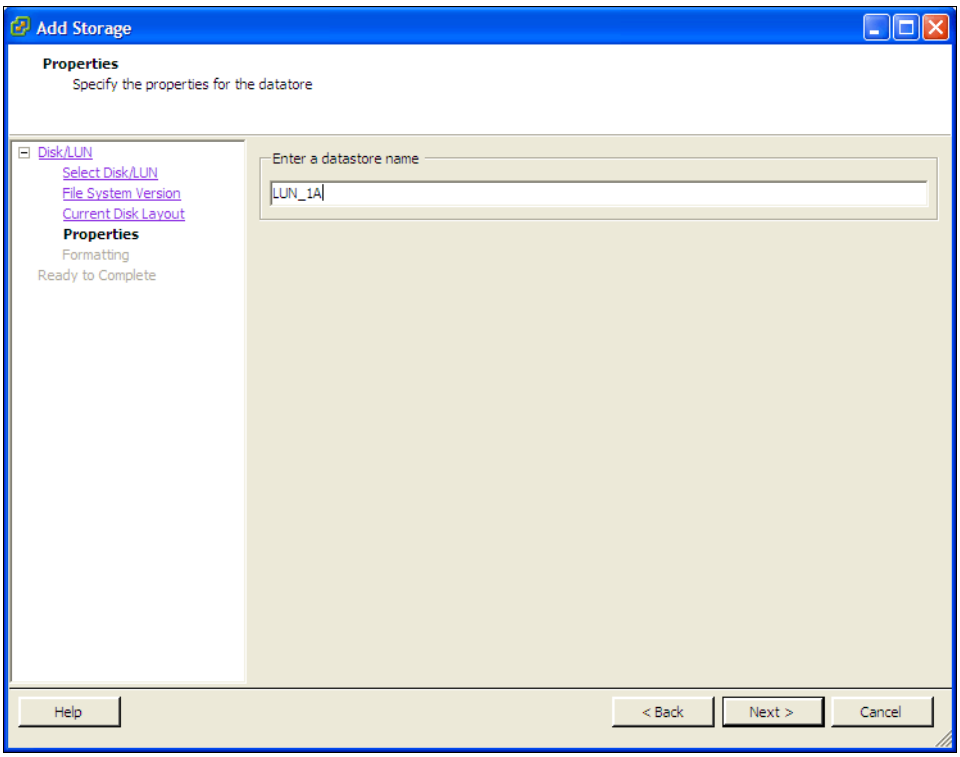

Figure 9-6 Naming the new VMFS-5 datastore

8. Specify the capacity for the datastore (Figure 9-7). Choose between “Maximum available space” (to use all the assigned space) or “Custom Space Setting” (to use part of the assigned space), and click **Next**.

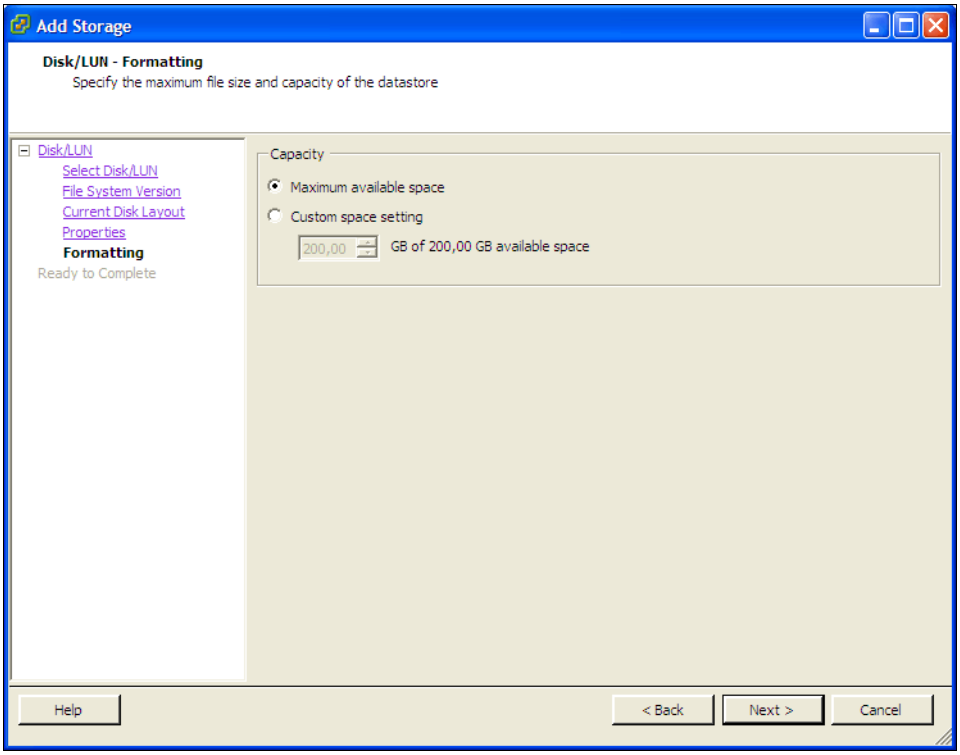

Figure 9-7 Choosing capacity for the datastore

9. A summary window (Figure 9-8) opens that shows all the selections. Click **Finish**. The LUN is formatted for VMFS-5 to use the maximum assigned space.

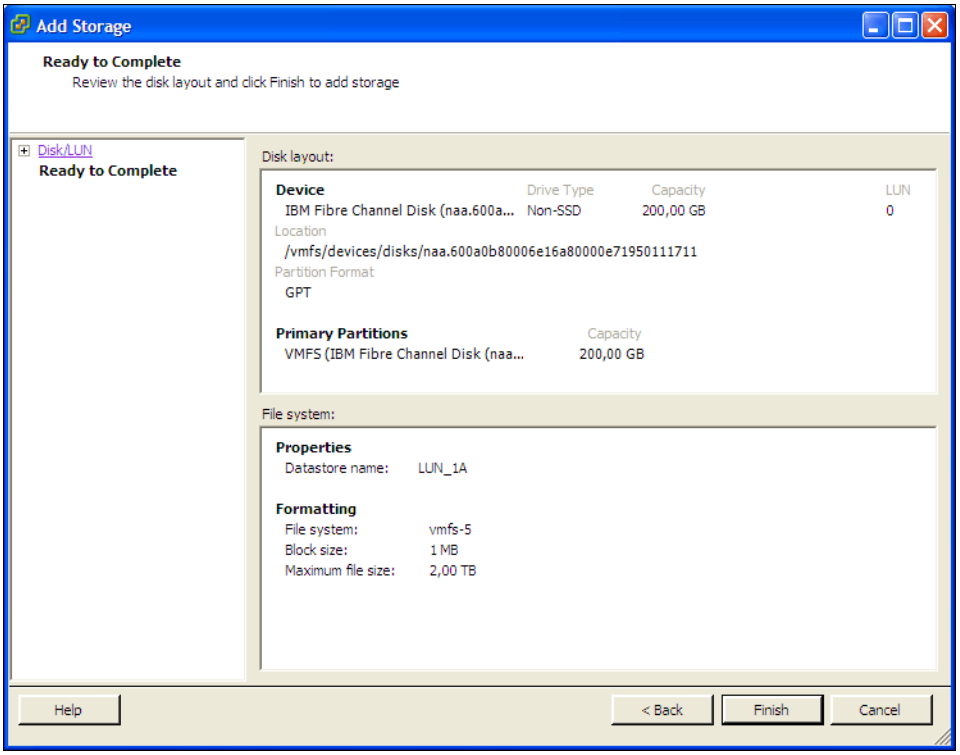

Figure 9-8 Summary window before formatting process

10. A new VMFS-5 datastore is ready for use in your virtual environment.

9.1.4 Prerequisites to upgrade a datastore to VMFS-5

Consider these prerequisites to upgrade a datastore to VMFS-5:

- If you use a VMFS-2 datastore, you must first upgrade it to VMFS-3.
- All hosts that access the datastore must support VMFS-5. All hosts need to run ESX(i) 5 or higher.
- Verify that the volume to upgrade has at least 2 MB of free blocks available and one free file descriptor.

9.1.5 Upgrading a datastore to VMFS-5 step-by-step

We can upgrade to VMFS-5 at anytime since this operation is online and nondisruptive.

Follow these steps to upgrade the datastore to VMFS-5:

1. Using the vSphere user interface, select the correct ESXi host, click the **Configuration** tab, and select the datastore to upgrade (as shown in Figure 9-9 on page 225). Then, click **Upgrade to VMFS-5**.

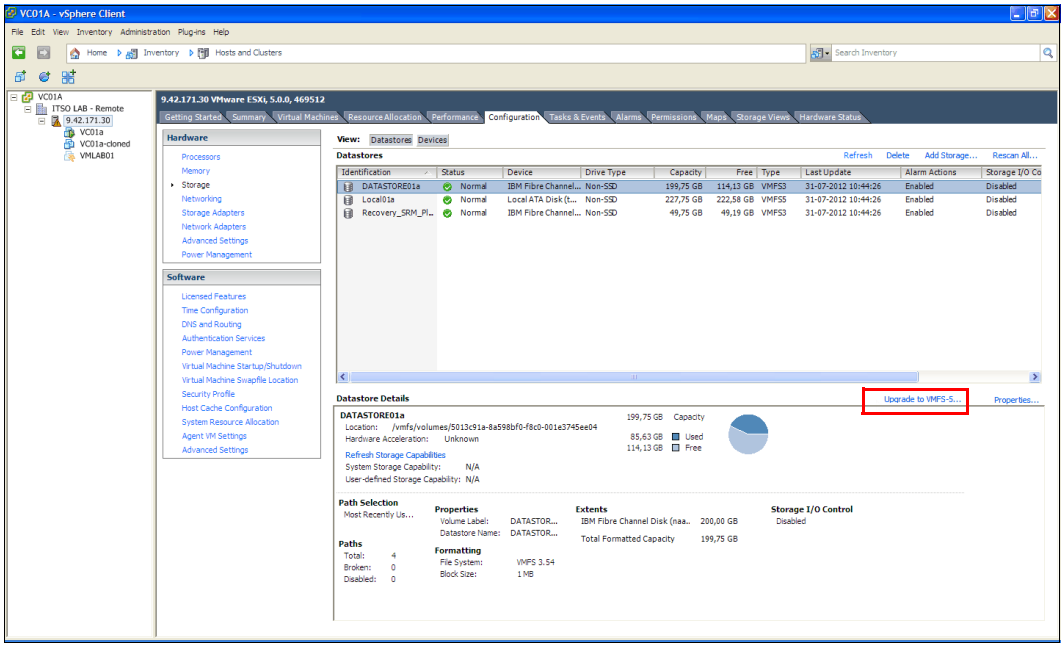

Figure 9-9 Selecting datastore to be upgraded

2. The window that is shown in Figure 9-10 opens to notify you whether your datastore can be upgraded to VMFS-5. Click **OK**.

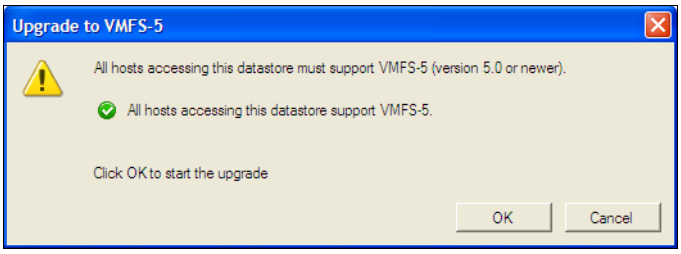

Figure 9-10 VMFS-5 support confirmation

3. After a few minutes (depending of the size and usage of the datastore), the upgrade process completes. Verify the completion by reviewing the datastore properties as shown in Figure 9-11 on page 226.

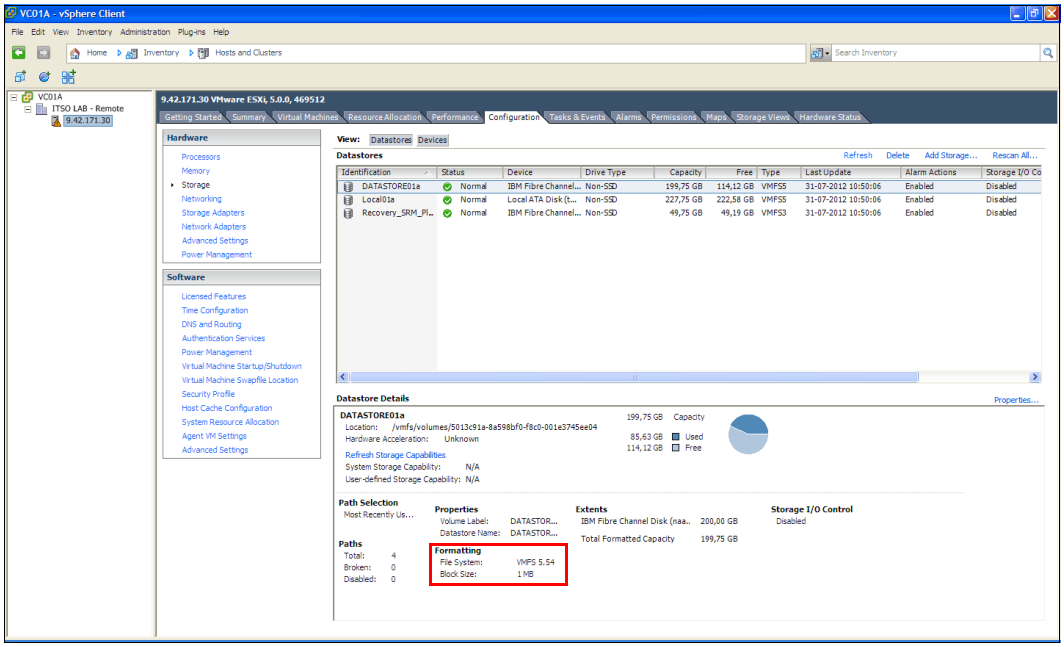

Figure 9-11 Verifying VMFS-5

9.2 Profile Driven Storage (storage profiles)

vSphere 5.0 introduced a new feature, Profile Driven Storage, which enables rapid and intelligent placement of virtual machines based on service level agreement (SLA), availability, performance, or other requirements and provided storage capabilities.

Using Profile Driven Storage, various storage characteristics, typically defined as a *tier*, can be requested in a virtual machine storage profile. These profiles are used during provisioning, cloning, and Storage vMotion to ensure that only those datastores or datastore clusters that are compliant with the virtual machine storage profile are made available. The virtual machine storage profile can also help select a similar type of datastores when creating a VMware Distributed Resource Scheduler (DRS) datastore cluster. Profile Driven Storage reduces the amount of manual administration required for virtual machine placement while improving virtual machine SLA storage compliance. See Figure 9-12 on page 227.

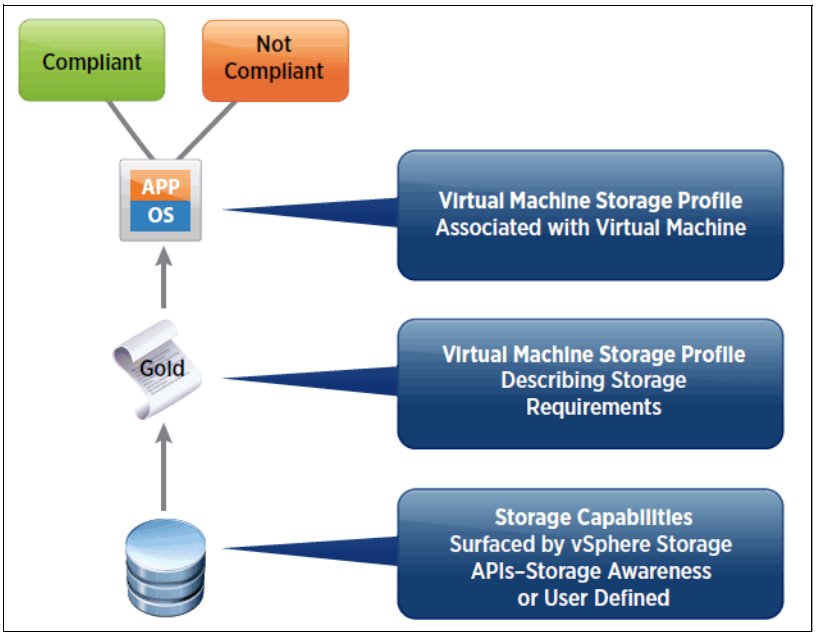

Figure 9-12 Profile Drive Storage overview

9.2.1 Profile Driven Storage benefits

Profile Driven Storage delivers these benefits:

- Offering full integration with vSphere Storage APIs – Storage Awareness, enabling usage of storage characterization supplied by storage vendors.
- Supporting Network File System (NFS), iSCSI, and Fibre Channel (FC) storage, and all storage arrays on the hardware compatibility list (HCL).
- Enabling the vSphere administrator to tag storage based on customer-specific or business-specific descriptions.
- Using storage characterizations and administrator-defined descriptions to create virtual machine placement rules in the form of storage profiles.
- Providing an easy means to check a virtual machine's compliance with these rules. This ensures that the virtual machine is not deployed or migrated to an incorrect type of storage without the administrator's knowledge.

Note: Profile Driven Storage does not support RDMs.

9.2.2 Storage capabilities

A *storage capability* outlines the quality of service that a storage system can deliver. It is a guarantee that the storage system can provide a specific set of characteristics for capacity, performance, availability, redundancy, and so on.

If a storage system uses Storage APIs - Storage Awareness, it informs vCenter Server that it can guarantee a specific set of storage features by presenting them as a storage capability. vCenter Server recognizes the capability and adds it to the list of storage capabilities in the Manage Storage Capabilities dialog box. Such storage capabilities are system-defined. vCenter Server assigns the system-defined storage capability to each datastore that you create from that storage system.

You can create user-defined storage capabilities and associate them with datastores. You should associate the same user-defined capability with datastores that guarantee the same level of storage capabilities. You can associate a user-defined capability with a datastore that already has a system-defined capability. A datastore can have only one system-defined and only one user-defined capability at a time.

You define storage requirements for virtual machines and virtual disks by adding storage capabilities to virtual machine storage profiles.

9.2.3 Working with Profile Driven Storage

There are a number of steps to follow in order to successfully use Profile Driven Storage. Before building a storage profile, the storage devices on your host must have capabilities associated with them. For instance, you might like to use user-defined business tags for your storage, such as Bronze and Gold.

Step 1 - Creating user-defined storage capabilities

From the vSphere user interface, click the icon labeled **VM Storage Profiles** (as shown in Figure 9-13).

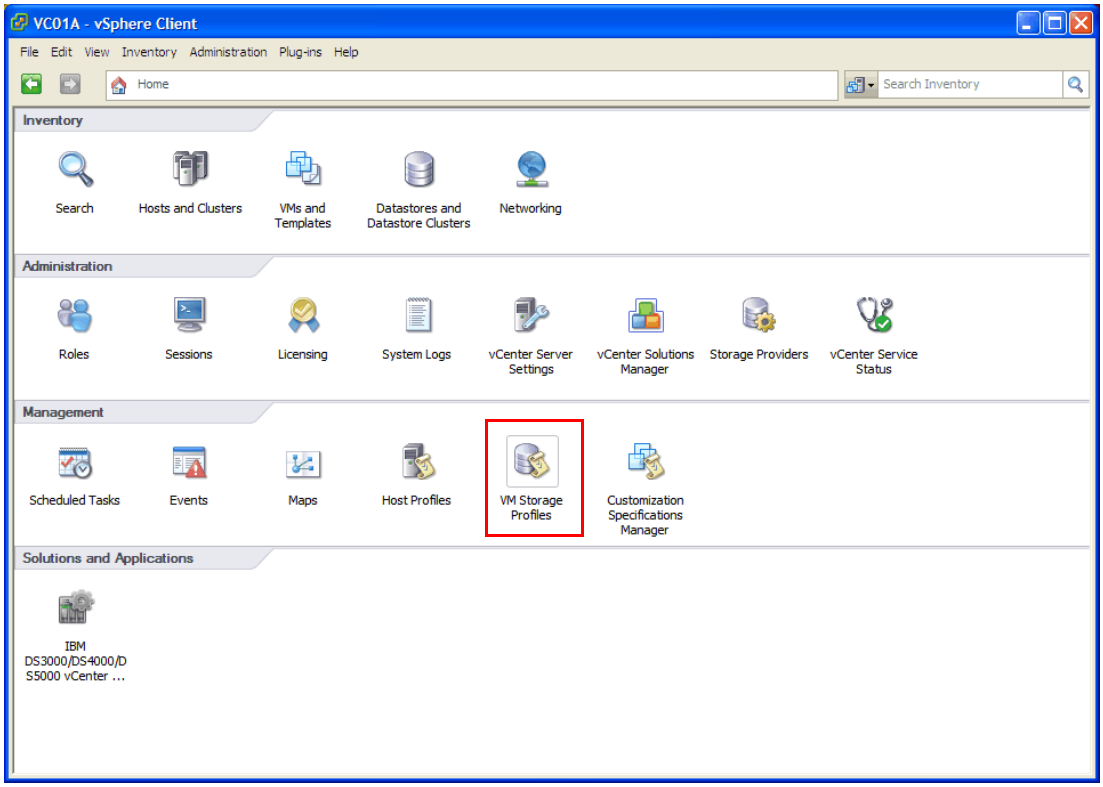

Figure 9-13 Selecting VM Storage Profiles

The VM Storage Profiles view (Figure 9-14 on page 229) opens.

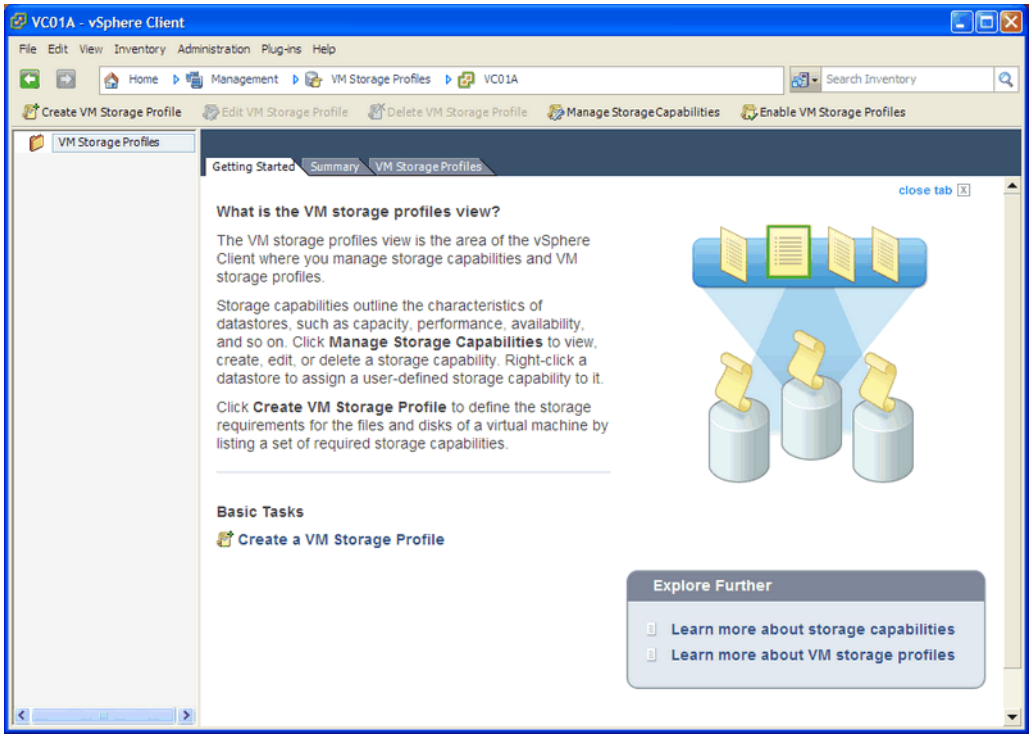

Figure 9-14 VM Storage Profile view

Add the user-defined storage capabilities (or business tags) by selecting **Manage Storage Capabilities** (Figure 9-15) and adding them.

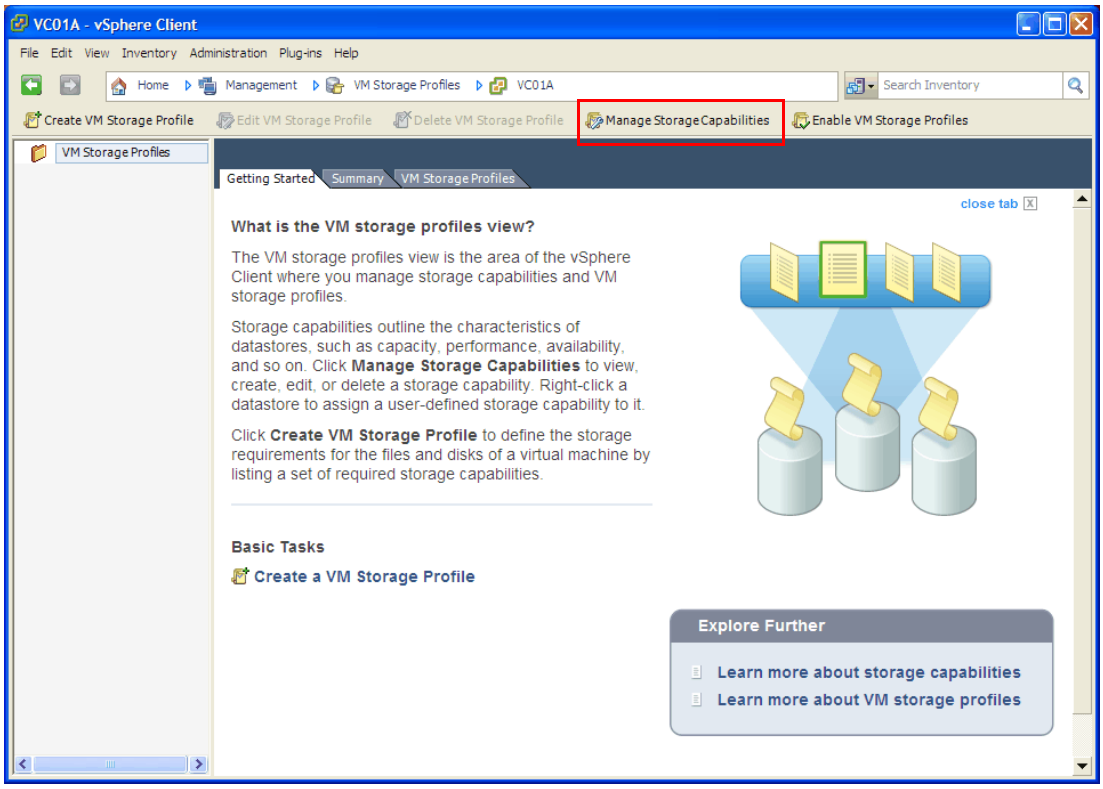

Figure 9-15 Selecting Manage Storage Capabilities

For the gold/silver/bronze example, Figure 9-16 shows how to create a Bronze user-defined storage capability.

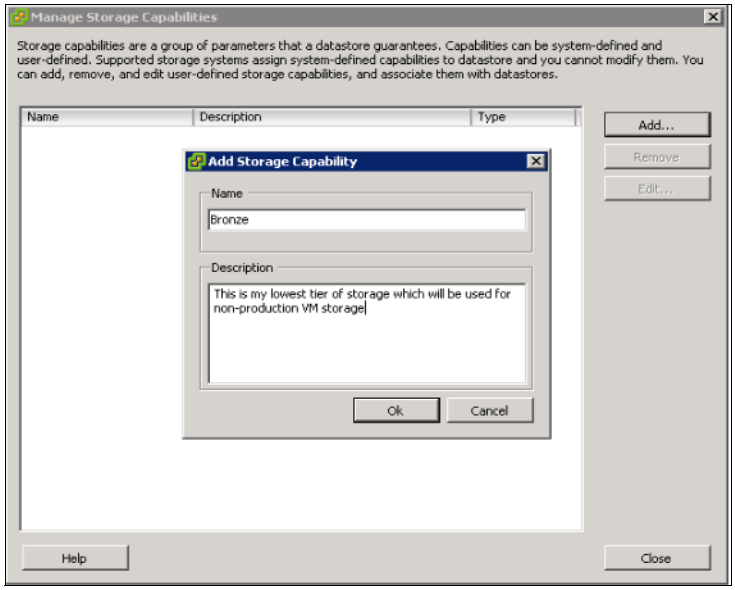

Figure 9-16 Creating Storage Capabilities

If you continue to create additional storage capabilities (Figure 9-17), you can use them to classify many different types of storage.

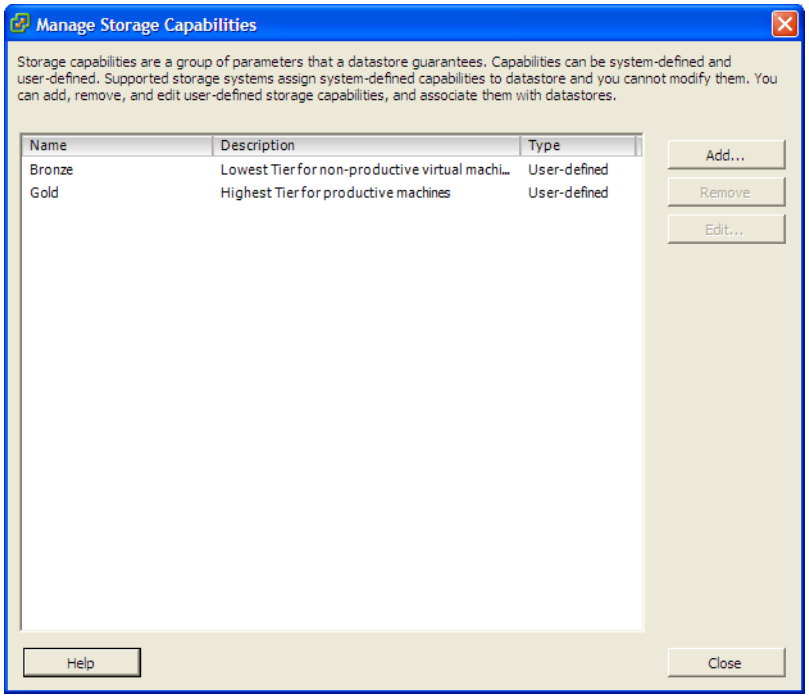

Figure 9-17 Defined storage capabilities

Step 2 - Creating VM storage profiles

The user-defined storage capabilities are created. The next step is to create a storage profile. To create a profile, select **Create VM Storage Profile** in the VM Storage Profiles view as shown in Figure 9-18 on page 231.

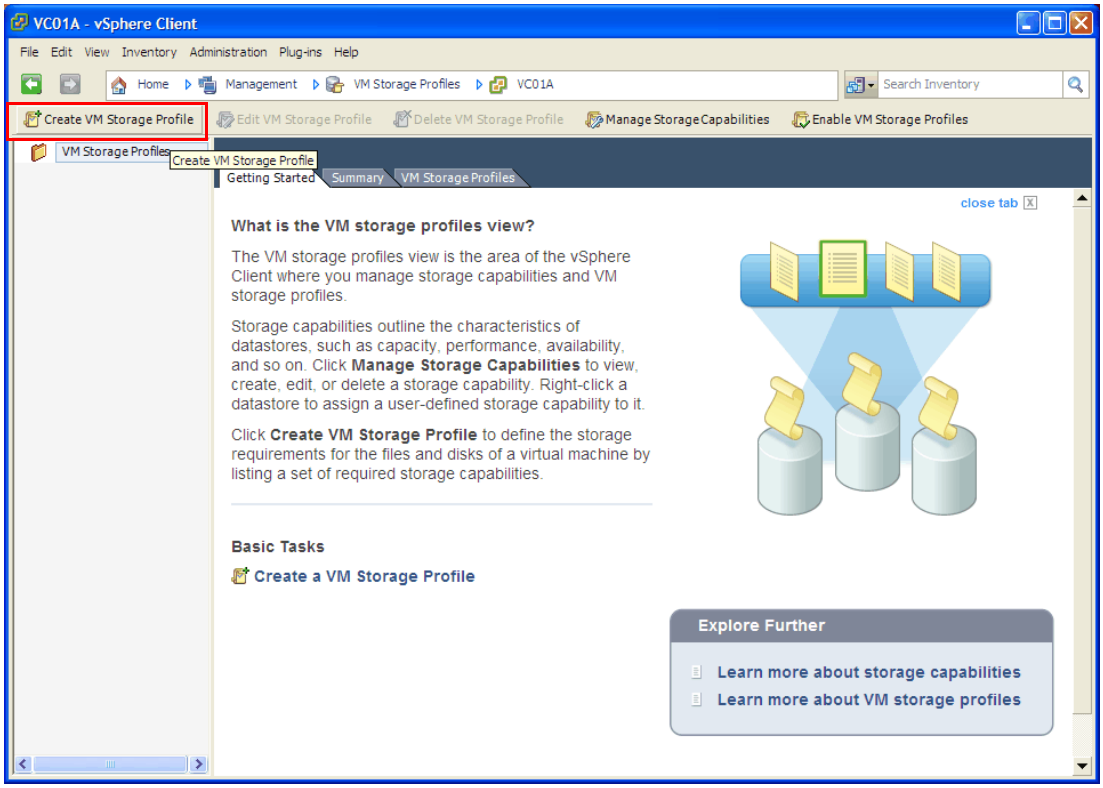

Figure 9-18 Create a VM storage profile

Enter a name and description (Figure 9-19).

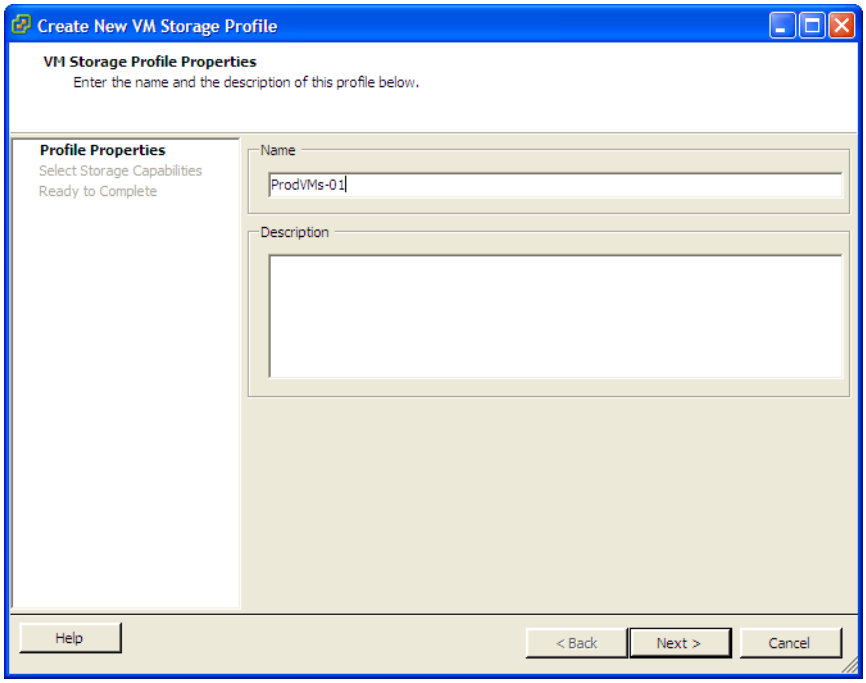

Figure 9-19 Naming your VM storage profile

Then, select the storage capabilities for that profile (Figure 9-20 on page 232).

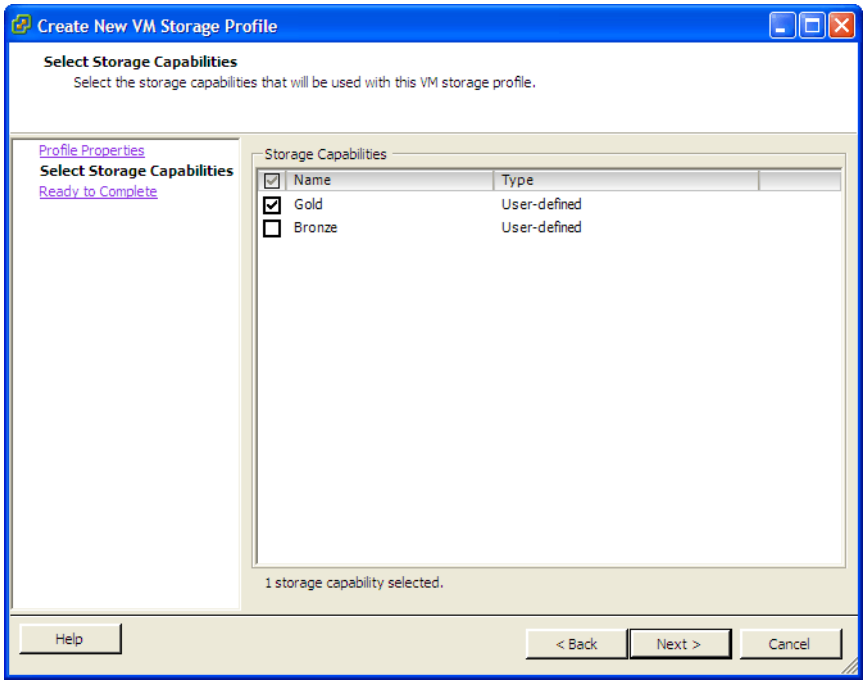

Figure 9-20 Assigning storage capabilities for your VM storage profile

You can create several profiles. For this example (Figure 9-21), we created two profiles, one for each tier of storage. Each profile contains a different capability (Bronze or Gold).

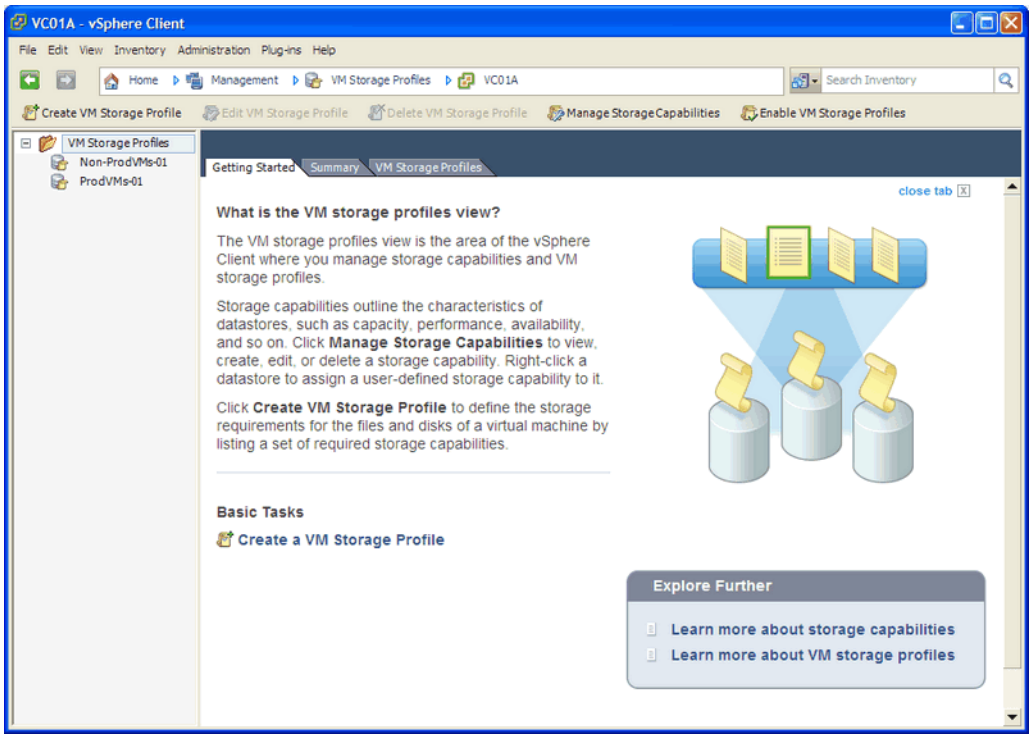

Figure 9-21 Storage profiles used as examples

Step 3 - Adding the user-defined capability to the datastore

The capabilities are now defined and the VM storage profiles are created. The next step is to add the capabilities to the datastores. This task is a simple point and click task. Right-click the desired datastore and select **Assign User-Defined Storage Capability** as shown in Figure 9-22.

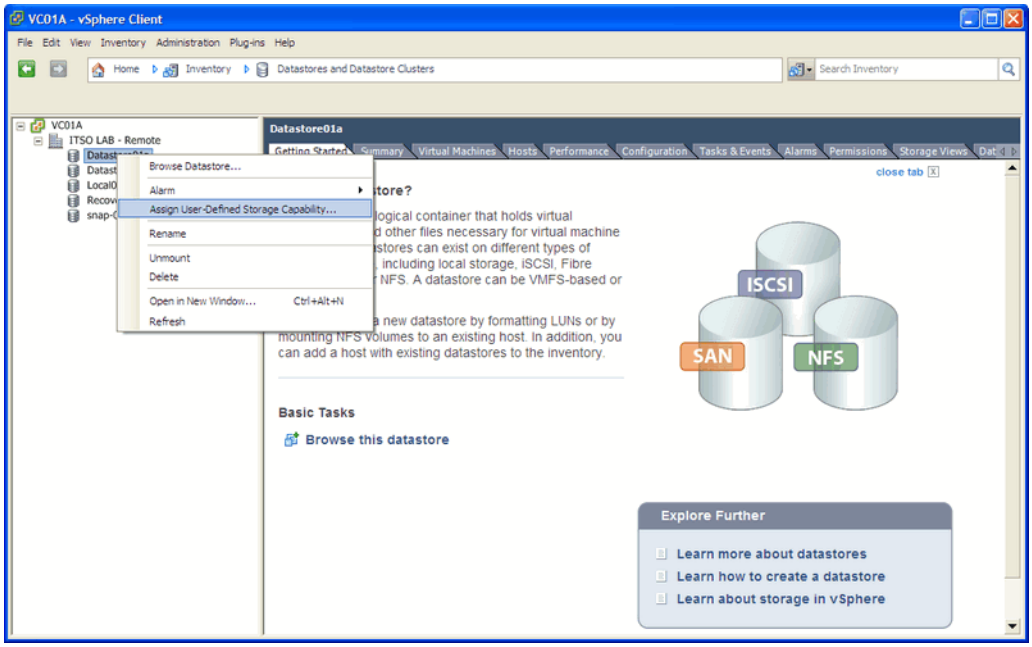

Figure 9-22 Adding the user-defined capability to the datastore

Select the storage capabilities that you want for your datastore (Figure 9-23).

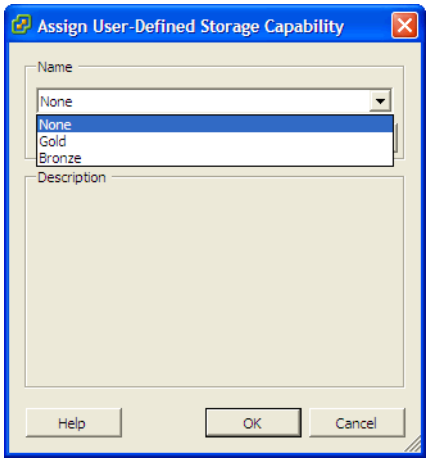

Figure 9-23 Selecting storage capability

In the Summary tab of the datastore, a new window called Storage Capabilities now displays user-defined storage capabilities (as shown in Figure 9-24 on page 234). The bubble icon next to the capability displays additional details (Figure 9-25 on page 234).

Figure 9-24 Storage capability window

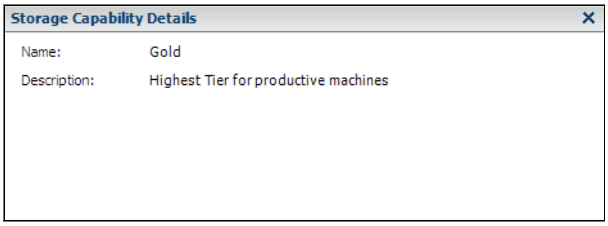

Figure 9-25 Bubble icon details

Step 4 - Using the VM storage profile

The profile is created and the user-defined capabilities are added to the datastore. Use the profile to select the correct storage for the virtual machine. The profile is automatically attached to the virtual machine during the deployment phase (as shown in Figure 9-26).

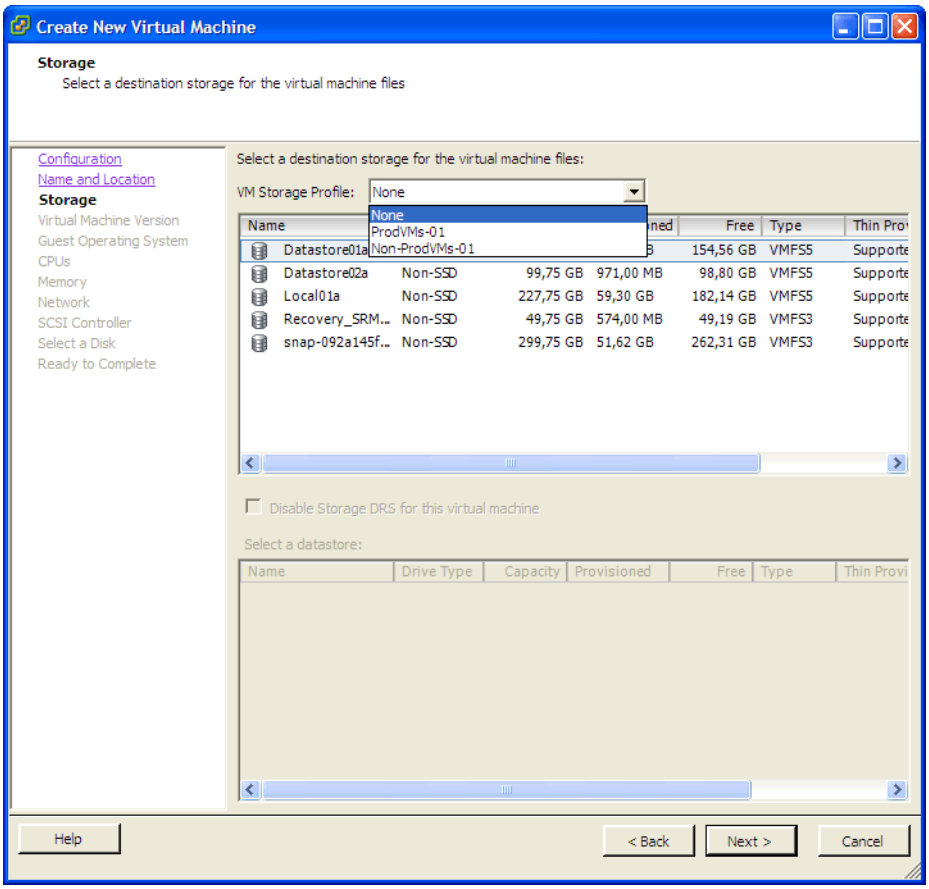

Figure 9-26 Selecting VM storage profile

Later, we can check whether the datastore on which the virtual machine is placed has the same capabilities as the profile. If it does, the virtual machine is *compliant*. If they do not, the virtual machine is *non-compliant* (Figure 9-27).

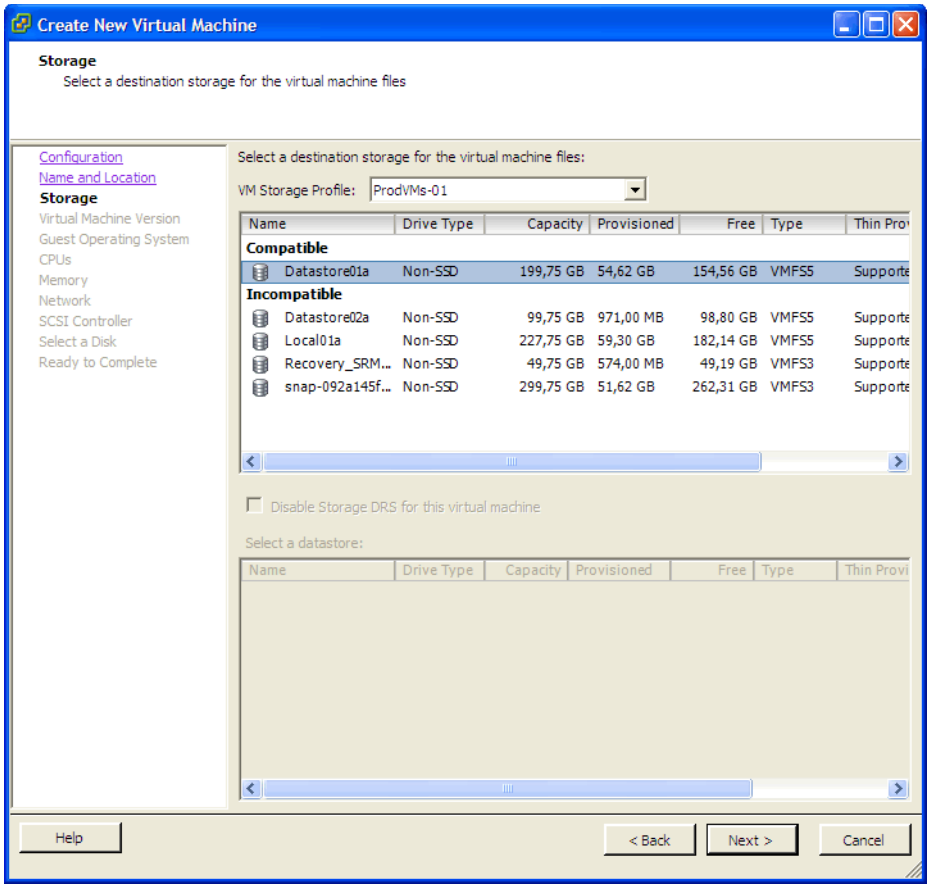

Figure 9-27 Checking capabilities for the virtual machine

Note: VM storage profiles can be used during deployment or during migrations, or they can be attached dynamically.

The datastores are now split into *Compatible* and *Incompatible*. The Compatible datastores have the same storage capabilities as those defined in the profile called `ProdVMs-01`. Only one datastore (`Datastore01a`) has this capability; none of the other datastores have that capability. However, you can still choose to deploy this virtual machine onto one of the Incompatible datastores if you want. The virtual machine shows up as Incompatible in the vSphere user interface when checked.

Step 5 - Checking compatibility

We have seen how to associate the virtual machine with the correct storage at the initial deployment time by using VM storage profiles. But during the lifecycle of a virtual machine, it can be migrated to other storage. A number of built-in mechanisms exist for checking the compliance of individual virtual machines or multiple virtual machines.

To check individual virtual machines, go to the **Summary** tab of the virtual machine. You can see a new VM Storage Profiles window, which indicates whether the virtual machine is compliant or not as shown in Figure 9-28 on page 236.

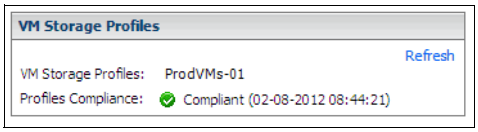

Figure 9-28 Checking compatibility on individual virtual machines

However, it is tedious to check all of these virtual machines individually. Therefore, if you go back to the VM Storage Profiles view, you can check all virtual machines per profile in one place as shown in Figure 9-29.

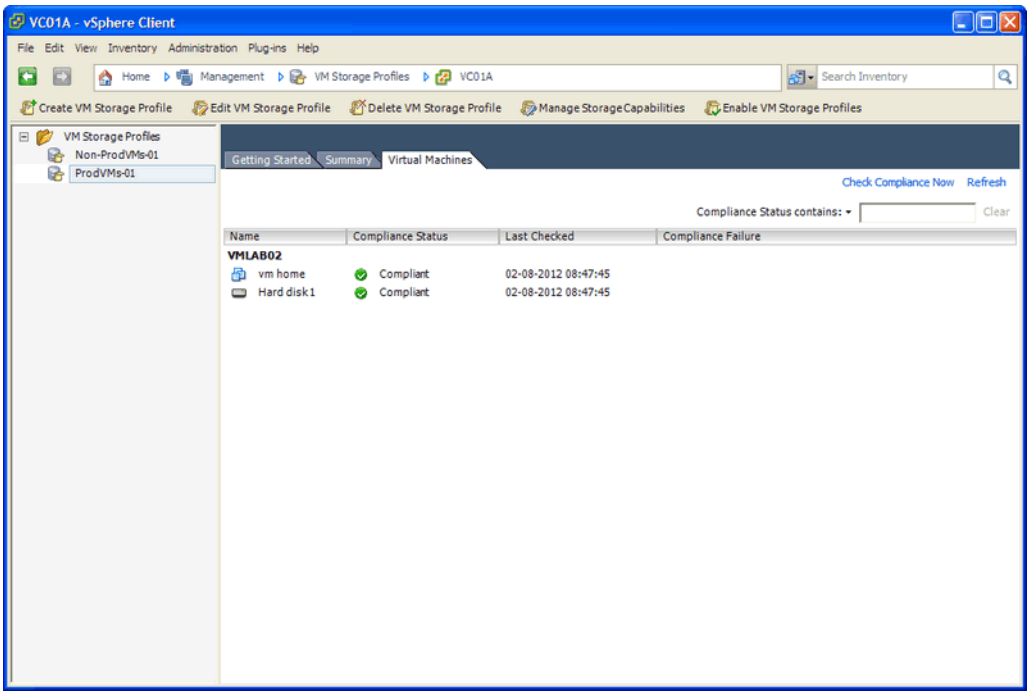

Figure 9-29 Checking compatibility on multiple virtual machines

9.3 vSphere Storage vMotion (Storage vMotion)

vSphere Storage vMotion (Storage vMotion) enables live migration for running virtual machine disk files from one storage location to another with no downtime or service disruption. With vSphere 5.0, multiple enhancements increase the efficiency of the Storage vMotion process, improve overall performance, and enhance supportability. Storage vMotion in vSphere 5.0 now also supports the migration of virtual machines with a vSphere snapshot and the migration of linked clones.

9.3.1 Storage vMotion enhancements

The enhancements to Storage vMotion include a new and more efficient migration process through the use of a new feature called *Mirror Mode*. Mirror Mode enables a single-pass block copy of the source disk to the destination disk by mirroring I/Os of copied blocks. Not only has the efficiency of Storage vMotion increased, but also the migration time predictability, making it easier to plan migrations and reducing the elapsed time per migration.

The mirror driver is enabled on a per-virtual machine basis and resides within the VMkernel. When the guest operating system of the virtual machine that is undergoing the process using Storage vMotion initiates a write to an already copied block, the mirror driver (Figure 9-30) synchronously mirrors this write and waits for both acknowledgements before communicating this information to the guest operating system.

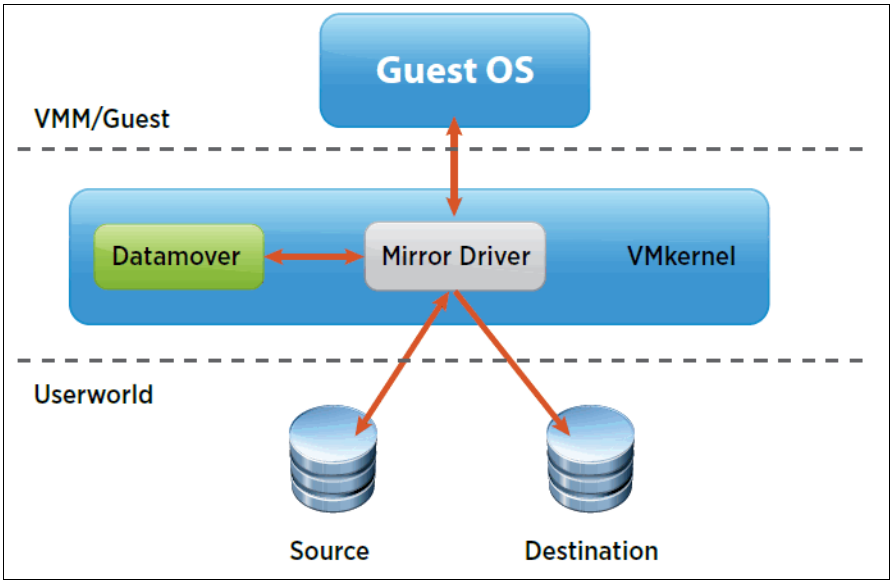

Figure 9-30 Data flow with Storage vMotion with Mirror Mode

9.3.2 Storage vMotion considerations

By using Storage vMotion, you can migrate a virtual machine and its disk files from one datastore to another while the virtual machine is running.

You can choose to place the virtual machine and all its disks in a single location, or select separate locations for the virtual machine configuration file and each virtual disk. The virtual machine does not change the execution host during a migration with Storage vMotion.

During a migration with Storage vMotion, you can transform virtual disks from thick-provisioned to thin-provisioned or from thin-provisioned to thick-provisioned.

Storage vMotion has many uses in administering virtual infrastructure, including the following examples:

- Upgrading datastores without virtual machine downtime. You can migrate running virtual machines from a VMFS-2 datastore to a VMFS-3 datastore, and upgrade the VMFS-2 datastore by using a two-step process. You can then use Storage vMotion to migrate virtual machines back to the original datastore without any virtual machine downtime.
- Performing storage maintenance and reconfiguration. You can use Storage vMotion to move virtual machines off a storage device to allow maintenance or reconfiguration of the storage device without virtual machine downtime.
- Redistributing storage load. You can use Storage vMotion to manually redistribute virtual machines or virtual disks to different storage volumes to balance capacity or improve performance.

9.3.3 Storage vMotion requirements

A virtual machine and its host must meet resource and configuration requirements for the virtual machine disks to be migrated with Storage vMotion.

Storage vMotion is subject to the following requirements and limitations:

- Virtual machine disks must be in persistent mode or be raw device mappings (RDMs). For virtual compatibility mode RDMs, you can migrate the mapping file or convert to thick-provisioned or thin-provisioned disks during migration as long as the destination is not an NFS datastore. If you convert the mapping file, a new virtual disk is created and the contents of the mapped LUN are copied to this disk. For physical compatibility mode RDMs, you can migrate the mapping file only.
- Migration of virtual machines during VMware Tools installation is not supported.
- The host on which the virtual machine is running must have a license that includes Storage vMotion.
- The host on which the virtual machine is running must have access to both the source and target datastores.

9.3.4 Using Storage vMotion step-by-step

We show you how to use the Storage vMotion feature to move a virtual machine from one datastore to another datastore.

Consider these points before you start the Storage vMotion process:

- Ensure that you have enough space on the destination datastore.
- Storage vMotion works with virtual machines that have snapshots/linked clones.
- Storage vMotion allows VMware to implement a new balancing technique for virtual machines based on storage usage and load. This feature is called Storage DRS.
- You can change both host and destination datastore *only* if the virtual machine is powered off.

Follow these steps to use the Storage vMotion feature:

1. Use the vSphere user interface to select the virtual machine to move by using Storage vMotion, right-click it, and select **Migrate** as shown in Figure 9-31 on page 239.

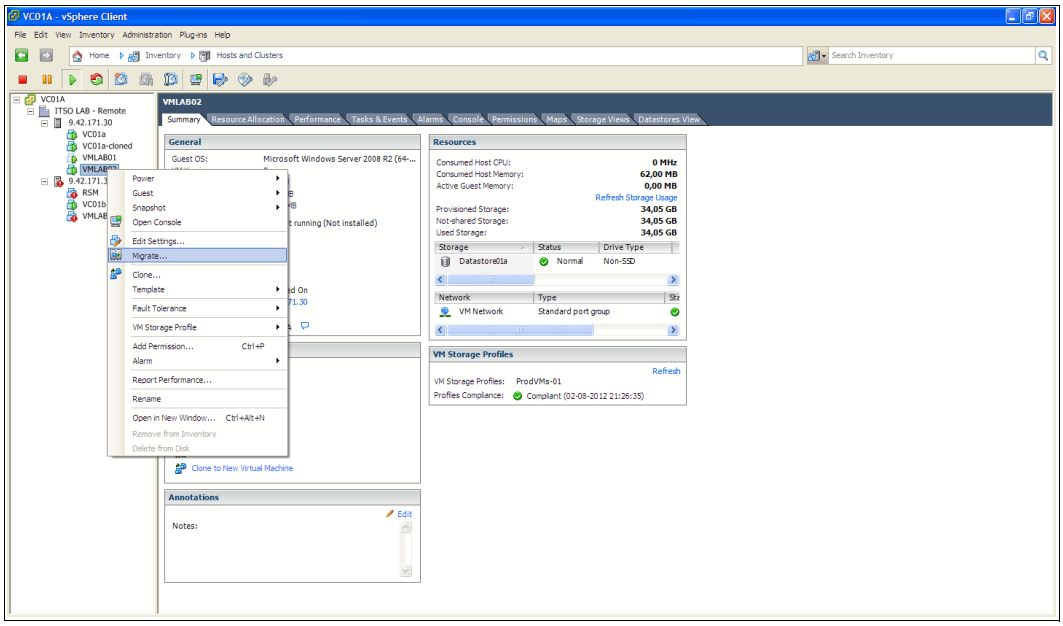

Figure 9-31 Selecting virtual machine

2. In Figure 9-32 on page 240, choose the type of migration and click **Next**. Choose one of these options:
   - To move your virtual machine from one host to another host while keeping the virtual machine files on the same datastore, use “Change host” (which is frequently called vMotion).
   - To move your virtual machine from one datastore to another datastore while keeping the virtual machine on the same host, use “Change datastore” (Storage vMotion).
   - To move your virtual machine from the datastore and host at the same time to another datastore and host, use “Change both host and datastore”.

> **Note:** You can change both host and destination datastore *only* if the virtual machine is powered off.

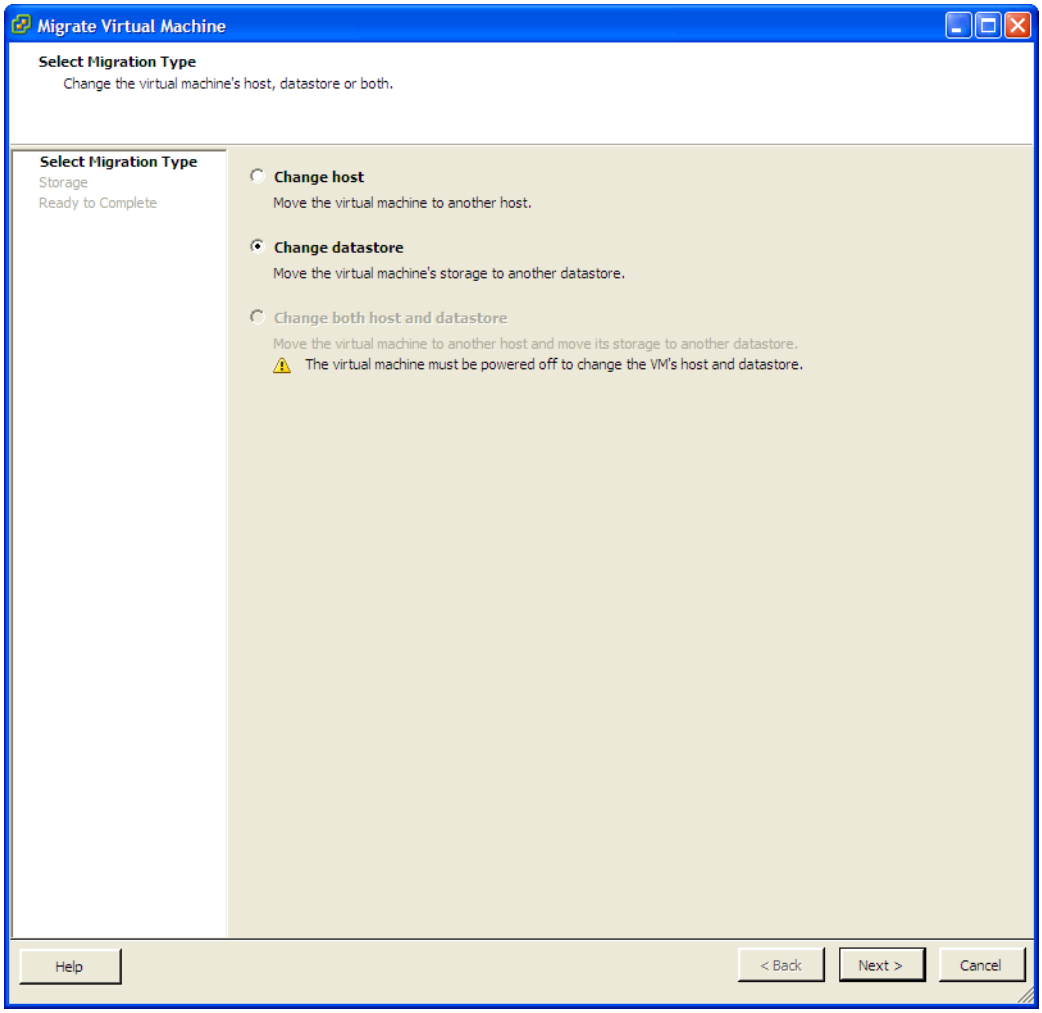

Figure 9-32 Choosing migration type

3. Choose the destination datastore as shown in Figure 9-33 on page 241.

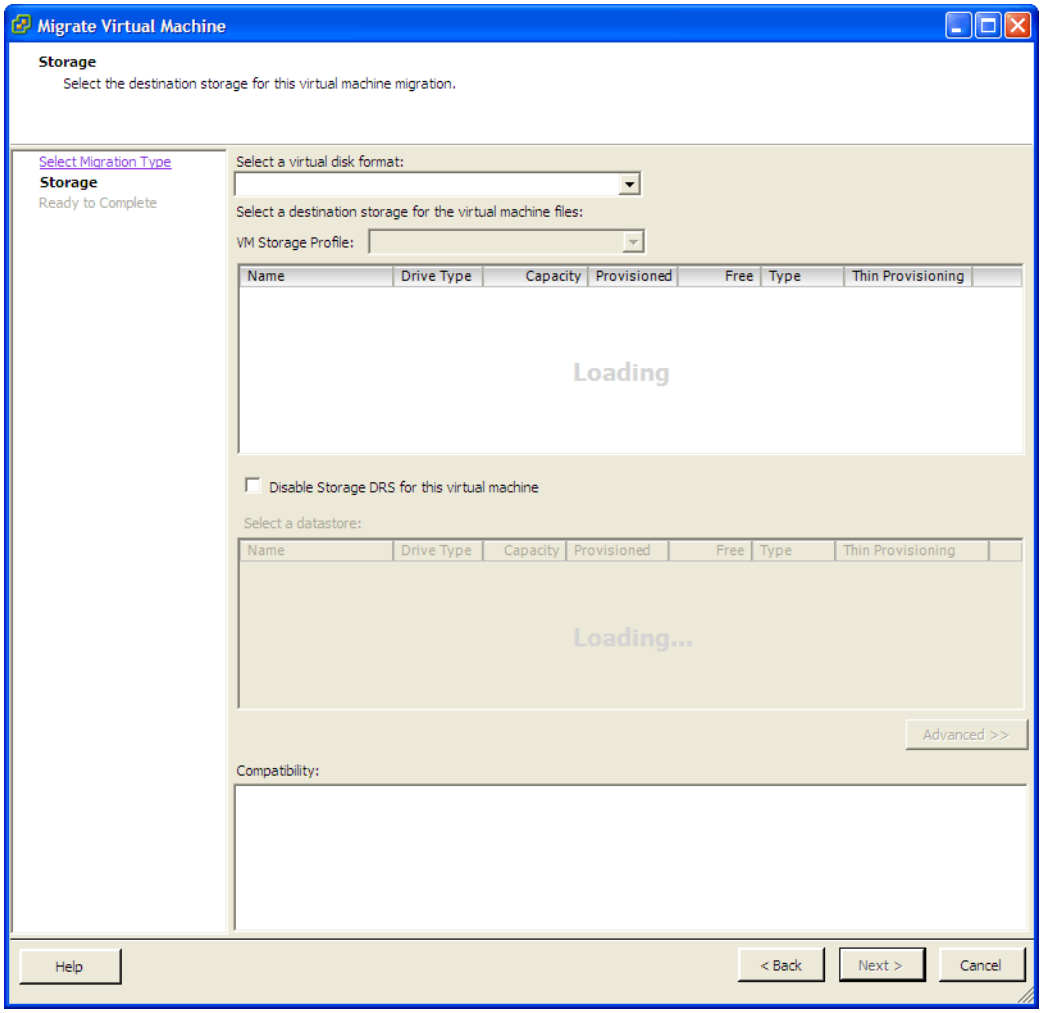

Figure 9-33 Selecting destination datastore

Make the following decisions in this step:

– Select the virtual disk format: In this pull-down menu, we have the options that are shown in Figure 9-34.

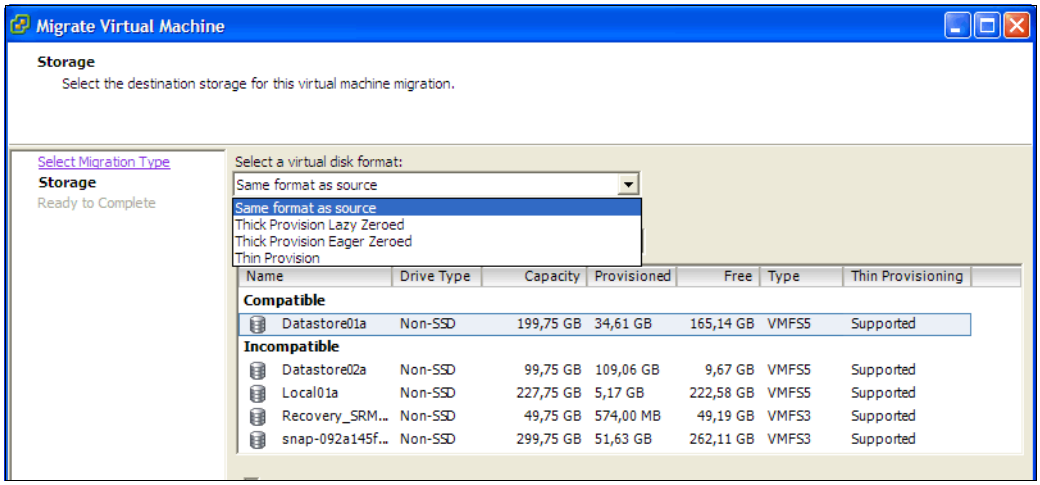

Figure 9-34 Selecting a virtual disk format

Select an option:

- Same format as source

  Uses the format of the original virtual disk. If you select this option for an RDM disk in physical compatibility mode, only the mapping file is migrated. If you select this option for an RDM disk in virtual compatibility mode, the RDM is converted to a virtual disk.

- Thick Provisioning Lazy Zeroed (new in vSphere 5.0)

  Creates a virtual disk in a default thick format. Space required for the virtual disk is allocated when the virtual disk is created. Data remaining on the physical device is not erased during creation, but it is zeroed out on demand at a later time on the first write from the virtual machine. Using the default flat virtual disk format does not zero out or eliminate the possibility of recovering deleted files or restoring old data that might be present on this allocated space. You cannot convert a flat disk to a thin disk.

- Thick Provisioning Eager Zeroed (new in vSphere 5.0)

  A type of thick virtual disk that supports clustering features, such as fault tolerance. Space required for the virtual disk is allocated at creation time. In contrast to the flat format, the data remaining on the physical device is zeroed out when the virtual disk is created. It might take much longer to create disks in this format than to create other types of disks.

- Thin provisioning

  Use this format to save storage space. For the thin disk, you provision as much datastore space as the disk would require based on the value that you enter for the disk size. However, the thin disk starts small and, at first, uses only as much datastore space as the disk needs for its initial operations.

– Destination storage for the virtual machine files: In this pull-down menu, we have the following options that are shown in Figure 9-35.

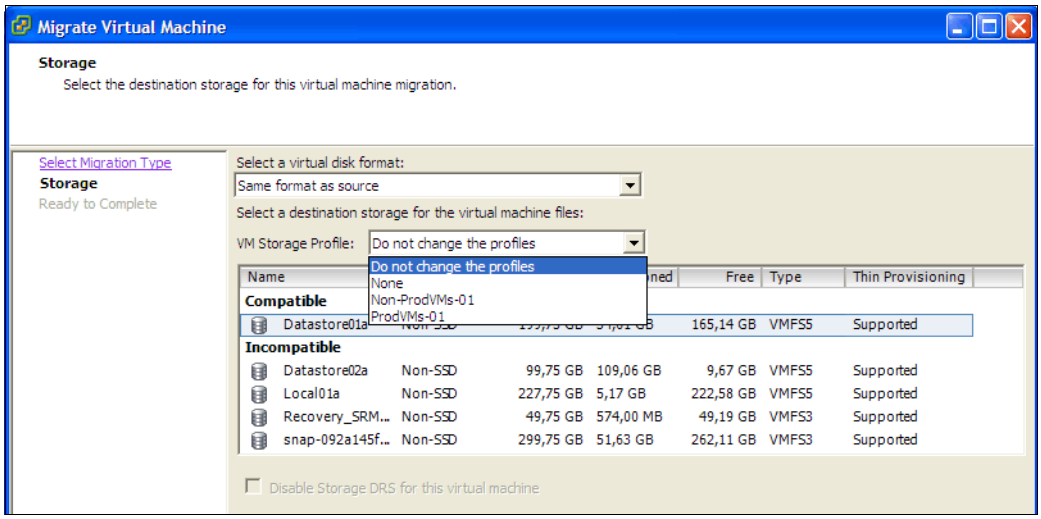

Figure 9-35 Selecting destination storage

Select an option:

- Because this virtual machine already has a VM Storage Profile called ProdVMs-01 that is associated with it, the “Do not change the profiles” option is identical to selecting the ProdVMs-01 option from the pull-down menu. This is also identical to

the Storage window that you see during the initial creation of a virtual machine, so the correct compatible storage can be chosen for the virtual machine right from the start.

- By choosing the compatible datastore from the list, when our migration is done, we know that the virtual machine resides on a datastore that has the same storage capabilities as those defined in the VM Storage Profile, that is, ProdVMs-01.
- If the option “None” is selected, you can choose any datastore that is available to the ESXi host as shown on Figure 9-36.

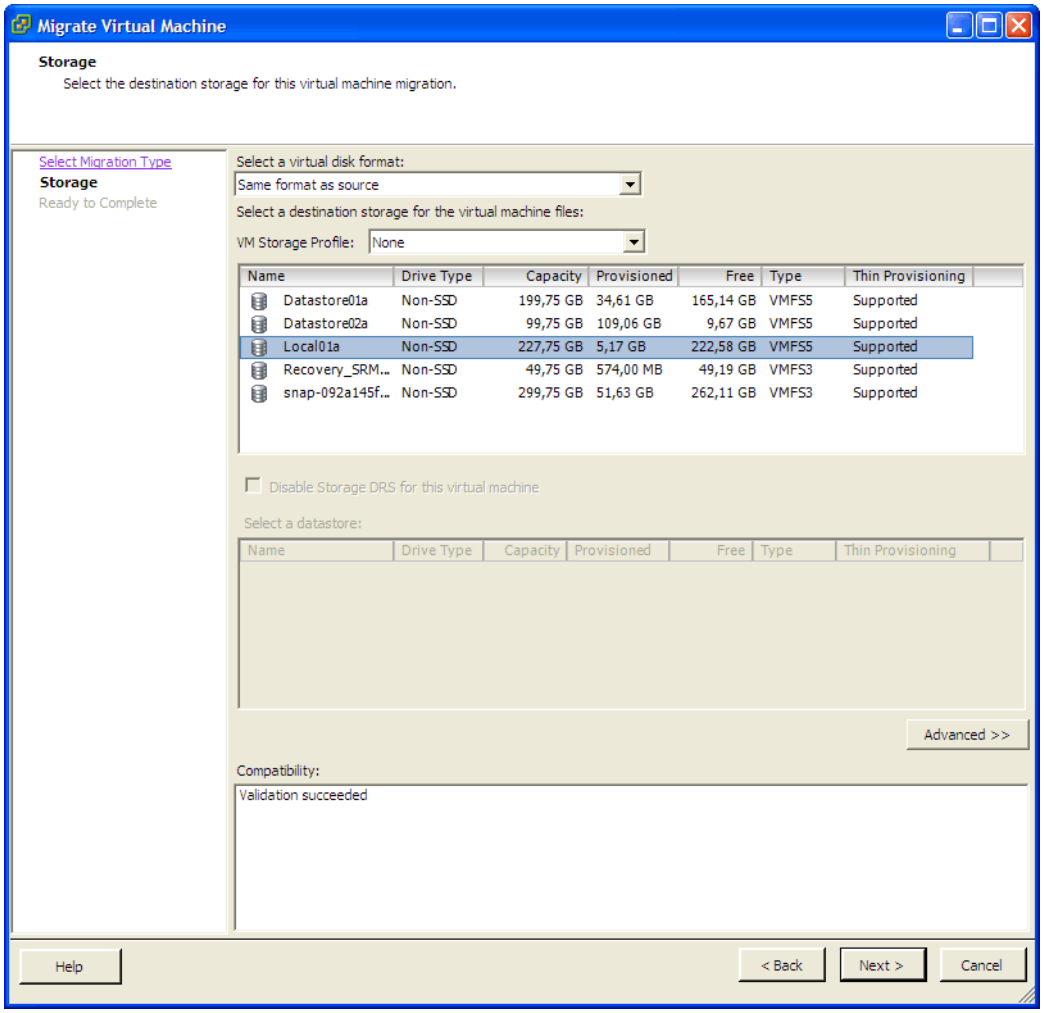

Figure 9-36 Selecting the None option

A summary of the selections shows before you start the migration process (as shown in Figure 9-37 on page 244).

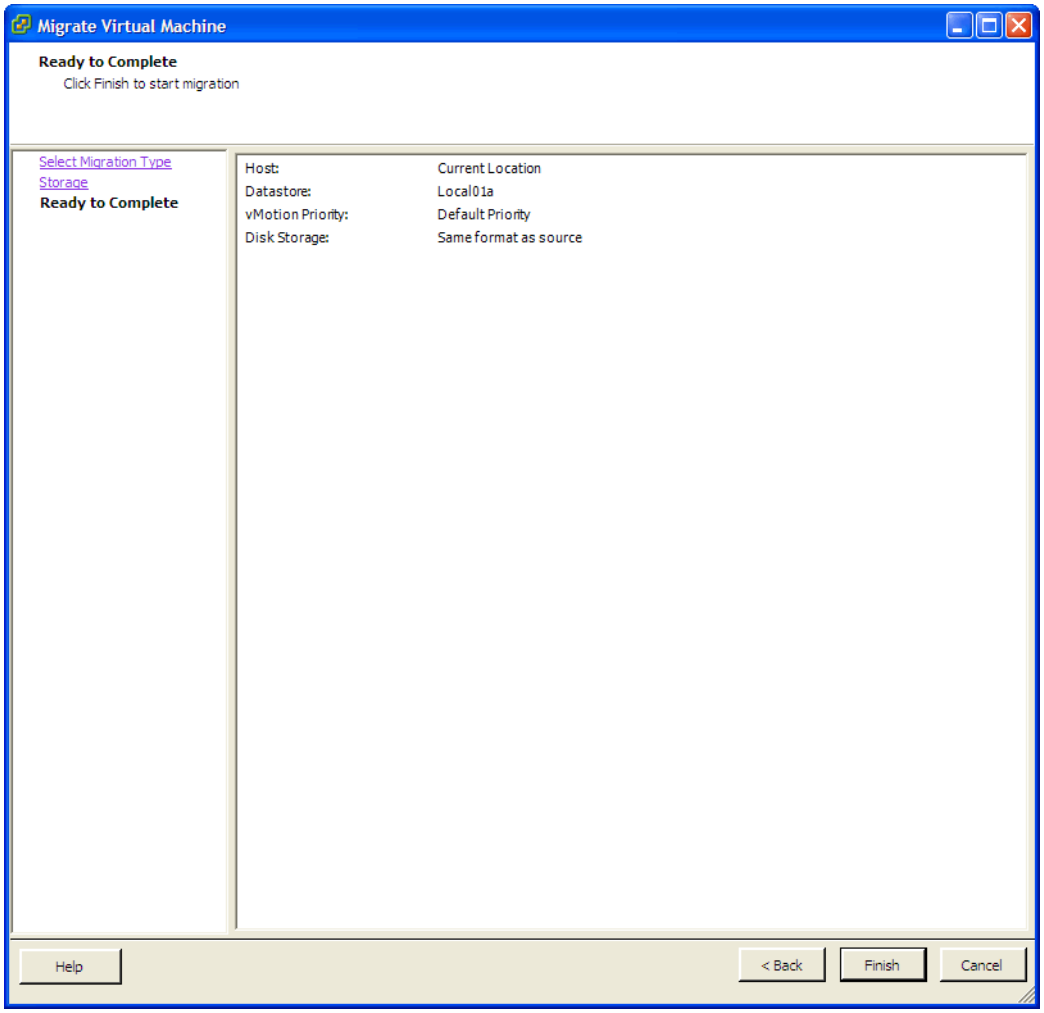

Figure 9-37 Summary selections before migration process

9.4 vSphere Storage Distributed Resource Scheduler (Storage DRS)

vSphere Storage DRS continuously monitors storage space and I/O utilization across a pre-assigned pool of storage volumes and intelligently aligns storage resources to meet your business growth objectives:

- Specify how storage resources are allocated to virtual machines with rules and policies.
- Give IT autonomy to business organizations by assigning dedicated storage infrastructure to business units while still achieving higher storage utilization through pools of storage volumes.
- Empower business units to build and manage virtual machines within their storage while giving central IT control over storage resources.

VMware Storage DRS continuously balances storage space usage and storage I/O load and avoids resource bottlenecks to meet application service levels, and increases manageability of storage at scale.

VMware Storage DRS allows you to perform these functions:

- Easily deploy additional storage capacity

  VMware Storage DRS seamlessly takes advantage of additional capacity when storage is added to a pool of storage volumes by balancing storage space usage and I/O load without disruption to running workloads.

- Maintain storage volumes in a nondisruptive manner

  VMware Storage DRS can evacuate a storage volume when in maintenance mode while balancing storage space usage and I/O load within the pool of storage volumes.

- Improve service levels for all applications

  VMware Storage DRS continuously balances storage I/O load between volumes, ensuring that congestion is avoided. In conjunction with Storage I/O Control (SIOC), it ensures that during congestion, critical virtual machines and applications have access to appropriate storage resources at any point in time.

- Increase the vSphere administrator's productivity

  Enable vSphere administrators to monitor and effectively manage more IT infrastructure.

9.4.1 Storage DRS features

We describe the key features of VMware Storage DRS.

New Virtual Center object

Storage DRS introduces a new vCenter Server object that is called "*datastore cluster*", which is primarily a pool of datastores. The user can create a datastore cluster by grouping similar datastores. Storage DRS can be enabled or disabled per "datastore cluster" or per virtual machine.

These clusters form the basis of Storage DRS. When one is created, Storage DRS can manage the storage resources in a comparable manner to how it manages compute resources in a cluster. As with a cluster of hosts, a datastore cluster is used to aggregate storage resources, enabling smart and rapid placement of the virtual disk files of a virtual machine and the load balancing of existing workloads. Figure 9-38 depicts a datastore cluster of 12 TB formed by four 3 TB datastores.

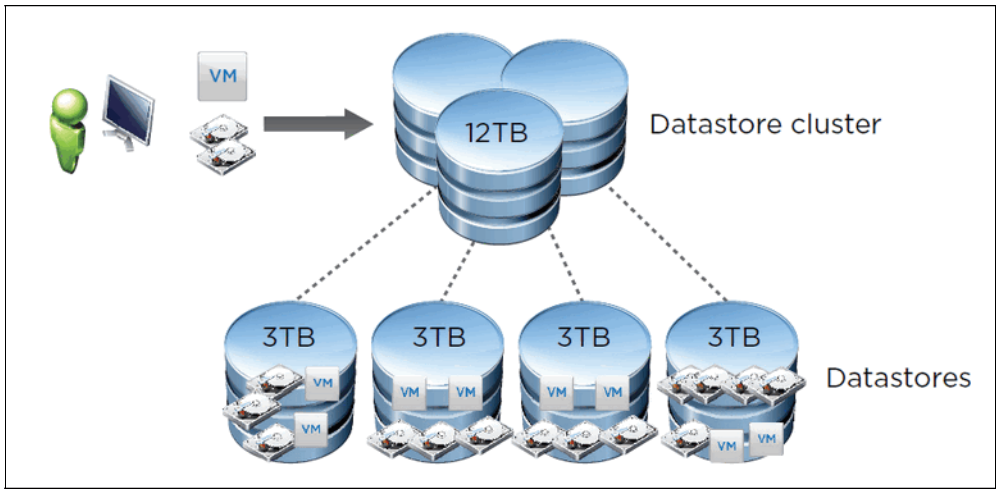

Figure 9-38 Datastore cluster example

Initial placement of virtual machines and VMDK files

In workflows for virtual machine create, clone, and relocate operations, a user can select a datastore cluster instead of a datastore.

In a case where a datastore cluster is selected as the destination, Storage DRS selects a member datastore based on space utilization and I/O load. Storage DRS provides recommendations for the placement of virtual machines and VMDK files.

Ongoing balancing between datastores within a datastore cluster

The SDRS algorithm issues migration recommendations in these situations:

- Space utilization and I/O response time thresholds of a datastore have been exceeded and there is a significant space or I/O imbalance between datastores within the datastore cluster.
- Storage DRS collects statistics on space utilization of all datastores within the datastore cluster every two hours and checks it once every eight hours. The I/O load history of past 24 hours is checked every eight hours.

Both manual enforcement and automatic enforcement of recommendations are supported. The user can select the desired automation level for a datastore cluster.

Affinity rules and maintenance mode

Storage DRS *affinity rules* enable you to control which virtual disks should or should not be placed on the same datastore within a datastore cluster. By default, a virtual machine's virtual disks are kept together on the same datastore. Storage DRS offers three types of affinity rules (Figure 9-39):

- VMDK Anti-Affinity

  Virtual disks of a virtual machine with multiple virtual disks are placed on different datastores.
- VMDK Affinity

  Virtual disks are kept together on the same datastore.
- VM Anti-Affinity

  Two specified virtual machines, including associated disks, are placed on different datastores.

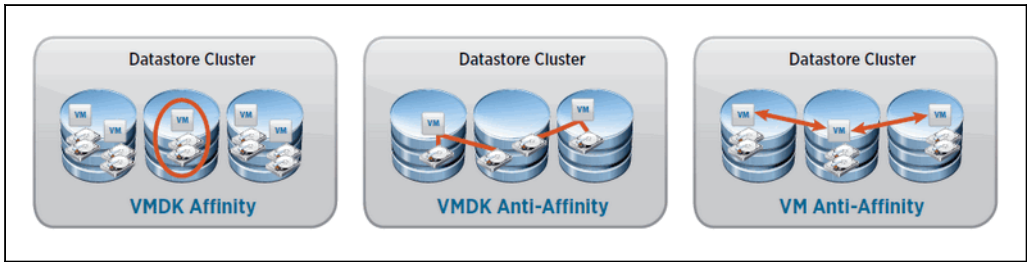

Figure 9-39 Affinity rules

Placement recommendations

Storage DRS provides initial placement and ongoing balancing recommendations, helping vSphere administrators make placement decisions based on space and I/O capacity. During the provisioning of a virtual machine, a datastore cluster can be selected as the target destination for this virtual machine or virtual disk, after which a recommendation for initial placement is made based on space and I/O capacity. Initial placement in a manual

provisioning process is very complex in most environments. As a result, crucial provisioning factors, such as current space utilization and I/O load, are often ignored. Storage DRS ensures that initial placement recommendations are made in accordance with space constraints and with respect to the goals of space and I/O load balancing. These goals aim to minimize the risk of storage I/O bottlenecks and the performance impact on virtual machines.

Ongoing balancing recommendations are made when one or more datastores in a datastore cluster exceed the user-configurable space utilization or I/O latency thresholds. These thresholds are typically defined during the configuration of the datastore cluster. Storage DRS applies the datastore utilization reporting mechanism of VMware vCenter Server to make recommendations whenever the configured utilized space threshold is exceeded. The I/O load is evaluated by default every eight hours. When the configured maximum space utilization or the I/O latency threshold (15 ms by default) is exceeded, Storage DRS calculates all possible moves to balance the load accordingly while considering the cost and the benefit of the migration.

Datastore maintenance mode

When a member datastore of a datastore cluster is put in maintenance mode, Storage DRS evacuates all virtual machines and VMDK files from the selected datastore while ensuring that space and I/O load are balanced across the remaining datastores.

In addition, Storage DRS offers *Datastore Maintenance Mode*, which automatically evacuates all virtual machines and virtual disk drives from the selected datastore to the remaining datastores in the datastore cluster.

Note: Storage DRS works with both VMFS-based and NFS-based datastores. Mixing NFS and VMFS datastores in a single datastore cluster is not supported.

9.5 vSphere Storage I/O Control

vSphere 5.0 extends Storage I/O Control to provide clusterwide I/O shares and limits for NFS datastores. This means that no single virtual machine should be able to create a bottleneck in any environment regardless of the type of shared storage that is used. Storage I/O Control automatically throttles a virtual machine that is consuming a disparate amount of I/O bandwidth when the configured latency threshold has been exceeded. This enables other virtual machines using the same datastore to receive their fair share of I/O. Storage DRS and Storage I/O Control are the perfect partners for preventing deprecation of service-level agreements while providing long-term and short-term I/O distribution fairness.

With VMware Storage I/O Control (Figure 9-40 on page 248), you can configure rules and policies to specify the business priority of each virtual machine. When I/O congestion is detected, Storage I/O Control dynamically allocates the available I/O resources to virtual machines according to your rules, helping you to realize these benefits:

- Improve service levels for critical applications
- Virtualize more types of workloads, including I/O-intensive business-critical applications
- Ensure that each cloud tenant gets its fair share of I/O resources
- Increase administrator productivity by reducing the amount of active performance management that is required
- Increase flexibility and agility of your infrastructure by reducing your need for storage volumes that are dedicated to a single application

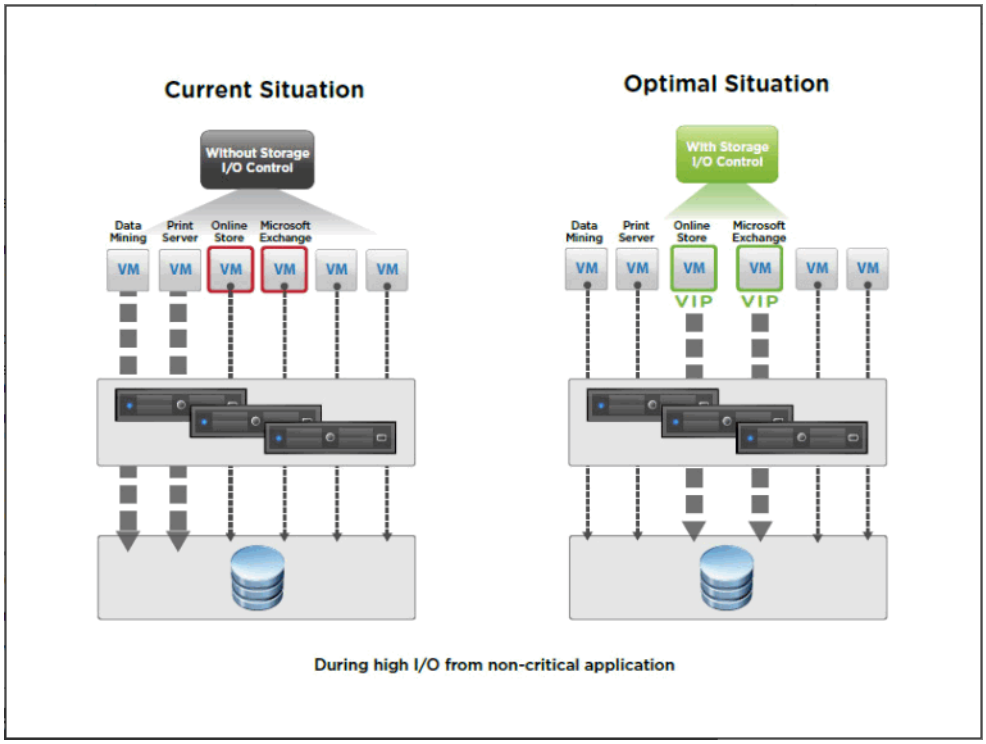

Figure 9-40 Storage I/O Control overview

9.5.1 vSphere Storage I/O Control features

The following list shows the key features of VMware Storage I/O Control:

Management:

- Set, view, and monitor storage resource shares and limits.
- Storage priorities (per VM) can be set and enforced across a cluster.

Automation:

- Enabling Storage I/O Control on a datastore triggers the monitoring of device latency that hosts observe when communicating with that datastore.
- When latency exceeds a set threshold, the feature engages automatically as the datastore is considered to be congested. Each virtual machine that accesses that datastore is then allocated I/O resources in proportion to their shares.

A

VMware ESXi Fibre Channel configuration checklist

In this appendix, we summarize the best practices and configuration steps needed to configure your VMware ESXi Server to work with DS5000 storage subsystems in Fibre Channel environments. For details about the settings explained here, check Chapter 3, “Planning the VMware vSphere storage system design” on page 29, Chapter 4, “Planning the VMware vSphere Server design” on page 55, and Chapter 5, “VMware ESXi Server and storage configuration” on page 69. Follow these guidelines to maintain the best performance and normal operation of your VMware ESXi environment. You can print them out to help you with the VMware ESXi Server implementation and configuration, or assist with troubleshooting.

Hardware, cabling, and zoning best practices

This section describes hardware, cabling, and zoning best practices to use with VMware ESXi environments using Fibre Channel connections.

Hardware

- ☐ Two identical Host Bus Adapters (HBAs) for each VMware ESXi Server:
  - ☐ Identical brand
  - ☐ Identical firmware
  - ☐ Identical settings

> **Important:** Single HBA configurations are allowed, but a single HBA configuration might result in loss of data access in a path failure.

QLogic HBA settings

The following settings should be set in QLogic HBA BIOS.

Adapter settings

- ☐ Host Adapter BIOS: Disabled (set it to enabled only if booting from SAN)
- ☐ Fibre Channel Tape Support: Disabled
- ☐ Data Rate: Set to fixed rate, which is supported by the HBA and the SAN switch

Advanced adapter settings

- ☐ Enable LIP reset: No
- ☐ Enable LIP Full Login: Yes
- ☐ Enable Target Reset: Yes

Cabling

- ☐ Each DS5000 controller should have connections to two SAN fabrics.
- ☐ Each HBA should be cabled into its own SAN fabric.
- ☐ Disk and tape traffic on separate HBAs.

SAN zoning

- ☐ Zone each HBA to see both DS5000 controllers (two paths per HBA - four paths per LUN)
- ☐ Use 1-to-1 zoning: In each SAN zone, there should only be one HBA and one DS5000 controller port, for example:
  - Zone 1: HBA1 with controller A, port 1
  - Zone 2: HBA1 with controller B, port 1
  - Zone 3: HBA2 with controller A, port 2
  - Zone 4: HBA2 with controller B, port 2

DS5000 settings

This section describes the settings that need to be defined in the DS5000 storage subsystem for it to work correctly with VMware ESXi environments.

Host type

- ☐ Host type needs to be set to VMware. All the necessary NVSRAM settings are included with that host type.
- ☐ Enable Auto Volume Transfer (AVT) on systems running firmware 7.83 or higher. Host type can be configured as VMwareTPGSALUA.

> **Note:** DS5000 running 7.83 or higher firmware can use ALUA where AVT feature needs to be enabled on the storage controllers and the Host Type needs to be configured as VMwareTPGSALUA.

LUN settings

- ☐ LUN numbering must be the same for each DS5000 LUN on each VMware ESXi Server.
- ☐ LUN numbering must start with 0, and increase consecutively with no gaps.
- ☐ Default LUN ID 31 (Access Logical Drive) is not supported and must be removed from the mappings list for each VMware ESXi host and host group.

Segment size

- ☐ Set the segment size to 256 KB.

VMware ESXi Server settings

This section describes the settings that need to be defined in VMware ESXi Server for it to work correctly with DS5000 storage subsystems.

Multipathing policy

- Path Selection: Most Recently Used

  Four paths for each LUN, two showing as Active and two showing as Stand by (each HBA has two paths to each DS5000 controller)

Advanced settings

To set these settings, open vSphere Client, then click **Configuration** → **Advanced Settings** (under Software).

- Disk.UseDeviceReset =0
- Disk.UseLunReset =1

Restrictions

This section describes the restrictions in VMware ESXi Server and DS5000 storage subsystem environments.

Controller firmware upgrade

Concurrent controller firmware download is not supported in storage subsystem environment with VMware ESXi Server host attached.

SAN and connectivity

VMware ESXi Server hosts support host-agent (out-of-band) managed storage subsystem configurations only. Direct-attached (in-band) management configurations are not supported.

VMware ESXi Server hosts can support multiple host bus adapters (HBAs) and DS5000 devices. However, there is a restriction on the number of HBAs that can be connected to a single storage subsystem. You can configure up to two HBAs per partition and up to two partitions per storage subsystem. Additional HBAs can be added for additional storage subsystems and other SAN devices, up to the limits of your specific storage subsystem platform.

Other

Dynamic Volume Expansion is not supported for VMFS-formatted LUNs.

Do not boot your system from a SATA device.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *IBM System Storage DS4000 and Storage Manager V10.30*, SG24-7010
- *IBM Midrange System Storage Hardware Guide*, SG24-7676
- *IBM System Storage DS3000: Introduction and Implementation Guide*, SG24-7065
- *IBM Midrange System Storage Implementation and Best Practices Guide*, SG24-6363
- *IBM System Storage DS5000 Series Hardware Guide*, SG24-8023
- *IBM System Storage DS5000 Series Implementation and Best Practices Guide*, SG24-8024
- *IBM System Storage DCS3700 Introduction and Implementation Guide*, SG24-8037
- *Implementing an IBM/Cisco SAN*, SG24-7545
- *Implementing an IBM b-type SAN with 8 Gbps Directors and Switches*, SG24-6116
- *IBM System Storage DS Storage Manager Copy Services Guide*, SG24-7822

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- IBM System Storage Interoperation Center (SSIC)

  http://www-03.ibm.com/systems/support/storage/ssic/interoperability.wss
- VMware vCenter Plug-In

  http://www.ibmvcenterplugin.com
- VMware vSphere 5.1 Documentation Center

  http://pubs.vmware.com/vsphere-51/index.jsp
- VMware ESX 4.0 and vCenter Server 4.0 Edition - Online Library

  http://pubs.vmware.com/vsp40

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

VMware Implementation with IBM System Storage DS5000

(0.5” spine)
0.475”<->0.873”
250 <-> 459 pages

VMware Implementation with IBM System Storage DS5000

Introduction to VMware design and architecture

VMware and storage planning and configuration

VMware Site Recovery Manager implementation

In this IBM Redbooks publication, we compiled best practices for planning, designing, implementing, and maintaining IBM Midrange Storage Solutions. In this publication, we use IBM System Storage DS5000 storage subsystem for the implementation procedures, and the same procedures can be used for implementations with DCS3700 or DS3500 storage subsystems. We also compiled configurations for a VMware ESX and VMware ESXi Server host environment.

Setting up an IBM Midrange Storage Subsystem is a challenging task. Our principal objective in this book is to provide you with a sufficient overview to effectively enable storage area network (SAN) storage and VMware. There is no single configuration that is satisfactory for every application or situation. However, the effectiveness of the VMware implementation is enabled by careful planning and consideration. Although the compilation of this publication is derived from an actual setup and verification, we did not stress test or test for all possible use cases that are used in a limited configuration assessment.

Because of the highly customizable nature of a VMware ESXi host environment, you must consider your specific environment and equipment to achieve optimal performance from an IBM Midrange Storage Subsystem. When you are weighing the recommendations in this publication, you must start with the first principles of input/output (I/O) performance tuning. Each environment is unique and the correct settings that are used depend on the specific goals, configurations, and demands for the specific environment.

This publication is intended for technical professionals who want to deploy VMware ESXi and VMware ESX Servers with IBM Midrange Storage Subsystems.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-8052-00 ISBN 073843826X