vmware® PRESS

FREE

VMware Press eSampler

SECOND EDITION

UPDATED WITH MORE CONTENT

Includes excerpts from newly released and upcoming VMware Press titles

vmware.com/go/vmwarepress

Official Cert Guide
The Official VCP5 Certification
VMware Certified on vSphere 5
PEARSON
Managing and Optimizing VMware vSphere Deployments
IT BEST PRACTICES
Sean Crookston
Harley Stagner
Automating vSphere with VMware vCenter Orchestrator
TECHNOLOGY HANDS-ON
Cody Bunch
ge Implementation n vSphere 5.0
TECHNOLOGY DEEP DIVE
Mostafa Khalil
are vSphere® 5 a Virtual Datacenter
BUSINESS IT
Eric Maillé
ené-Francois Mennecier

SHARE WITH OTHERS

vmware® PRESS

CHAPTERS IN THIS eSAMPLER ARE DRAFT MANUSCRIPTS

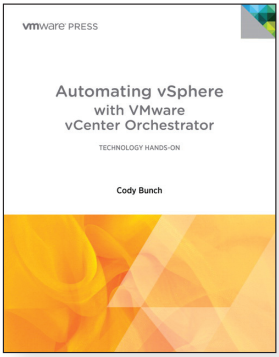
NOT AVAILABLE FOR RESALE

This manuscript has been provided by Pearson Education at this early stage to create awareness for this upcoming book. It has not been copy edited or proofread yet; we trust that you will judge this book on technical merit, not on grammatical and punctuation errors that will be fixed at a later stage.

vmware® PRESS

FREE
VMware Press eSampler
SECOND EDITION

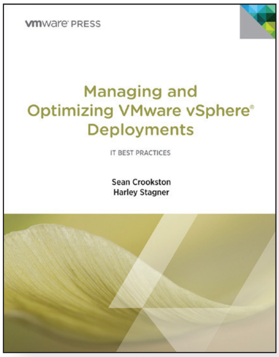
TABLE OF CONTENTS


vmware® PRESS

Automating vSphere
with VMware
vCenter Orchestrator

TECHNOLOGY HANDS-ON

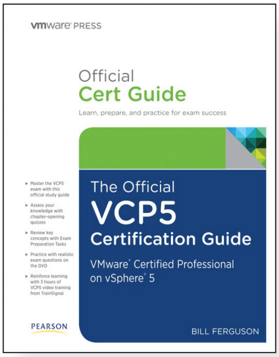
Cody Bunch


vmware® PRESS

Managing and
Optimizing VMware vSphere®
Deployments

IT BEST PRACTICES

Sean Crookston
Harley Stagner


vmware® PRESS

Official
Cert Guide

Learn, prepare, and practice for exam success

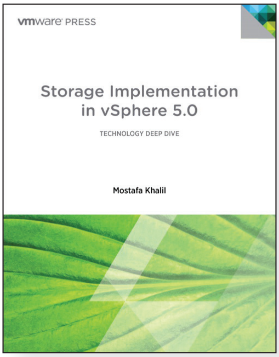
The Official
VCP5
Certification Guide

VMware® Certified Professional
on vSphere® 5

- Master the VCP5 exam with this official study guide
- Assess your knowledge with chapter-opening quizzes
- Review key concepts with Exam Preparation Tasks
- Practice with realistic exam questions on the DVD
- Reinforce learning with 3 hours of VCP5 video training from TrainSignal

PEARSON

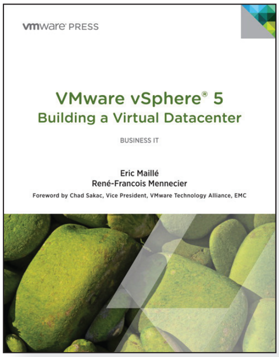
BILL FERGUSON


vmware® PRESS

Storage Implementation
in vSphere 5.0

TECHNOLOGY DEEP DIVE

Mostafa Khalil


vmware® PRESS

VMware vSphere® 5
Building a Virtual Datacenter

BUSINESS IT

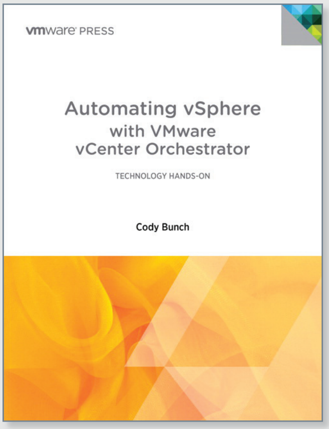
Eric Maillé
René-Francois Mennecier
Foreword by Chad Sakac, Vice President, VMware Technology Alliance, EMC

SHARE WITH OTHERS

vmware.com/go/vmwarepress

ALWAYS LEARNING

PEARSON


vmware PRESS

Automating vSphere with VMware vCenter Orchestrator

TECHNOLOGY HANDS-ON

Cody Bunch

CHAPTER 3 Configuring vCenter Orchestrator

MARCH 2012

Available in Print and eBook formats and through SAFARI BOOKS ONLINE

SHARE WITH OTHERS

vmware PRESS

BUY NOW

Automating vSphere with VMware vCenter Orchestrator

BY CODY BUNCH

Table of Contents

ISBN: 9780321799913

vmware.com/go/vmwarepress

ALWAYS LEARNING

PEARSON

Chapter 3

Configuring vCenter Orchestrator

As you saw in Chapter 2, “Installing vCenter Orchestrator,” vCO will not even let you log in before it is configured. So, configure it we must. There is no shortage of configuration options for vCO, either. The basic configuration we walk through in this chapter leaves you with a working configuration on which we build the remainder of our workflows. Our basic configuration covers the following areas:

- Starting the vCO Web Configuration service
- Changing the default vCO configuration service password
- Configuring networking
- Configuring LDAP
- Database configuration
- Configuring SSL
- Licensing the vCO server
- Plug-in configuration
- Installing vCO as a service
- Backing up the configuration

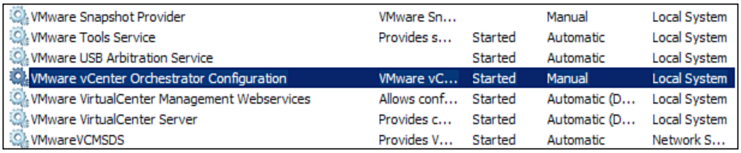
It might not look complicated, but if this is your first time configuring vCO, some of these items can be a bit obtuse when you begin to configure them. With that, here we go.

Starting the vCO Web Configuration Service

This step is needed only if you are using the vCO service that was installed when you installed vCenter Server. This is because when installed alongside vCenter Server, the vCO service is set to manual. Here is the procedure to start this service:

1. Click Start.
2. Click Run.
3. Type **services.msc**.
4. Locate vCenter Orchestrator Web Configuration Service.
5. Click Start.

Figure 3.1 shows what the service looks like in the services list once started.



VMware Snapshot Provider	VMware Sn...		Manual	Local System
VMware Tools Service	Provides s...	Started	Automatic	Local System
VMware USB Arbitration Service		Started	Automatic	Local System
VMware vCenter Orchestrator Configuration	VMware vC...	Started	Manual	Local System
VMware VirtualCenter Management Webservices	Allows conf...	Started	Automatic (D...	Local System
VMware VirtualCenter Server	Provides c...	Started	Automatic (D...	Local System
VMwareVCMSDS	Provides V...	Started	Automatic	Network S...

Figure 3.1 Starting the vCO Web Configuration service

NOTE

Because you will not be logging in every day to configure the vCenter Ops service, I recommend leaving this set to Manual. This keeps the resources consumed by this service from affecting the remainder of your environment. More important, it also helps reduce the attack surface of the vCO server. You will, however, need to start it again should you need to make changes to vCO (perhaps to add a second vCenter Server or add a plug-in, for example).

Changing the Default vCO Configuration Service Password

Although you could skip this step, I strongly recommend against it and suggest instead that you pick a suitably long, random password for the vCO Web Configuration service. There are several web tools available to do this for you. Setting a long, random password helps reduce the attack surface of the box by having one less easily guessed password.

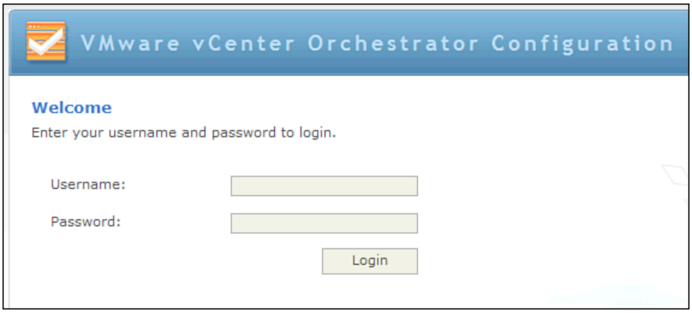
NOTE

The default username is vmware and cannot be changed. This makes it that much more important that you choose a strong password, as with a fixed username half of the battle for guessing is done for an attacker.

The procedure for changing the default configuration password is as follows:

1. Open a browser and browse to http://<vCO_Hostname>:8282.
2. Log in with vmware/vmware.
3. In the right pane, click Change Password.
4. Specify both the old password and new password.
5. Apply the changes.

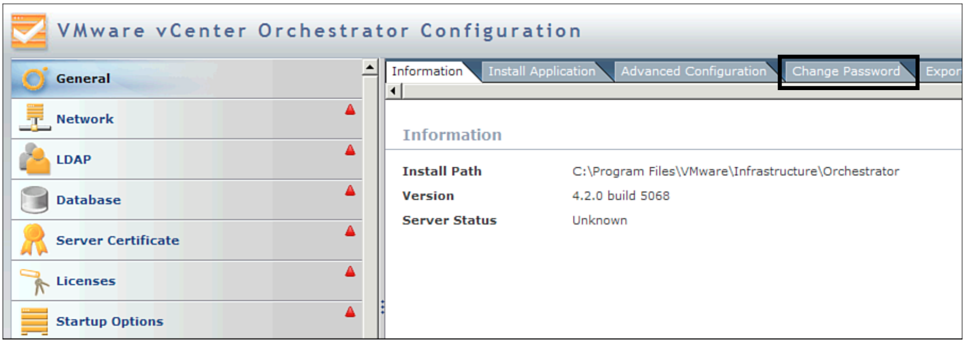
Figures 3.2 through 3.4 show the password-reset process.



Welcome

Enter your username and password to login.

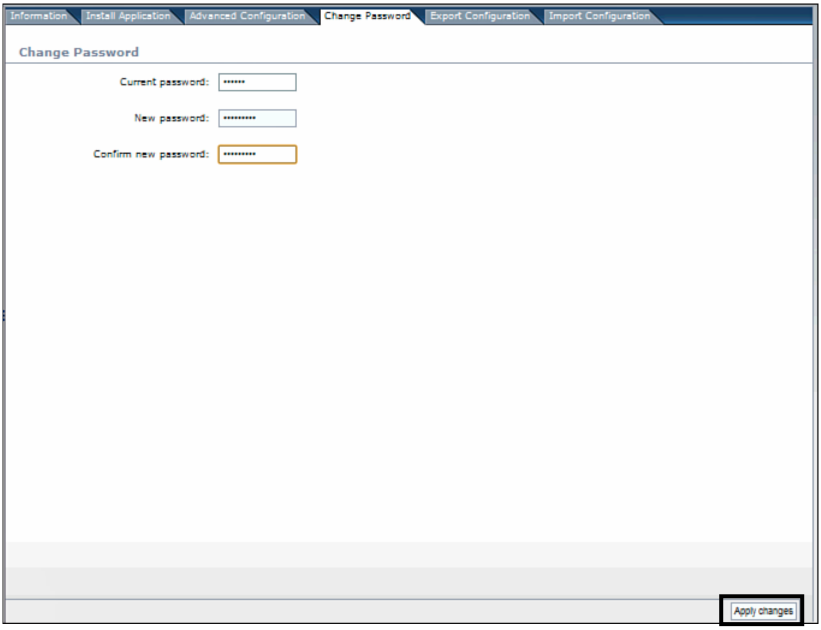
Figure 3.2 The vCO Web Configuration Login screen



Information

Install Path	C:\Program Files\VMware\Infrastructure\Orchestrator
Version	4.2.0 build 5068
Server Status	Unknown

Figure 3.3 Click Change Password



Change Password

Current password: ••••••

New password: •••••••••

Confirm new password: •••••••••

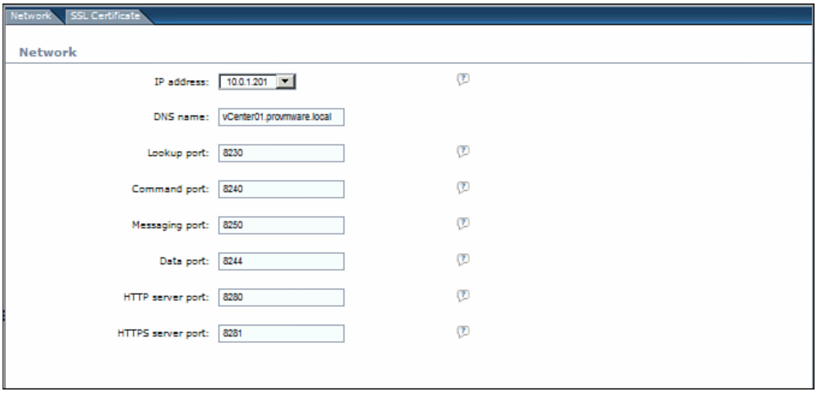
Figure 3.4 Specify passwords and click Apply Changes

Configuring Networking

In most instances, the server on which you run vCO will have multiple network interface cards (NICs) assigned to it. These include a management interface, a backup interface, and other networks vCO may need access to. Some common additional networks vCO might need access into would include things such as your storage network, management network, and Microsoft Active Directory network. In smaller environments, a number of these networks will be consolidated into one. However, you still need to make vCO aware of which network to bind to, or listen on. The steps needed to do this are as follows:

1. Select Network.
2. For the IP address, pick the network you want to use.
3. Confirm the communications ports.
4. Apply the changes.

In Chapter 2, as we were discussing the vCO requirements, we recommended both a static IP as well as a DNS entry. Now that we are configuring vCO, this becomes even more important because having these ensures you can consistently access vCO. It also makes administering vCO, and any required network access control lists, easier. vCO automatically enumerates the IP addresses assigned to the server for you to select in the drop-down, as shown in Figure 3.5.



Network

IP address: 10.0.1.201

DNS name: vCenter01.provmware.local

Lookup port: 8230

Command port: 8240

Messaging port: 8250

Data port: 8244

HTTP server port: 8280

HTTPS server port: 8281

Figure 3.5 IP Address drop-down on the Network page

Also of note in this section are the default communications ports. Rather than list them all here, I refer you to the VMware-specific documentation for them. My recommendation is that unless you have specific organizational requirements to change these from non-default ports, or have a conflict with another application, you want to leave these as default.

Configuring LDAP

LDAP, or Lightweight Directory Access Protocol, is how vCO proxies user authentication back into your existing environment. Be that Microsoft Active Directory, or Novell eDirectory, vCO relies on a third-party authentication source. To perform this configuration step, you need the LDAP path for the OU, or organizational unit, that contains your users. Lost? Well, I show you how to get that path for Microsoft's Active Directory. The procedure to configure vCO to use LDAP is as follows:

1. Find your LDAP path.
2. Fill in the form.
3. Click Apply.

As part of your LDAP configuration, you need to provide vCO with an LDAP group to identify which users are members of the vCO Admins group. In the example, I created a specific group for this. You want to configure this as is appropriate for your environment and its security requirements.

NOTE

A working LDAP configuration is critical to the operation of your vCO environment. This is because vCO uses LDAP groups to establish permissions for its various objects and workflows.

Find Your LDAP Path

Of the tasks required to set up LDAP for vCO, finding the LDAP path for your OU is going to be the one that is not the most straightforward if it's your first time. Of the many ways to obtain this information, we review two examples. First, we obtain this information using the DSQuery command-line tool to search for a specific user or OU. The second example uses the search boxes built in to the vCO interface. For these examples, we will use the AD hierarchy in Figure 3.6.

NOTE

In some cases, if you have multiple LDAP servers and one is not accessible, the configuration will time out and yet still display as if it's correctly configured even though it is not.

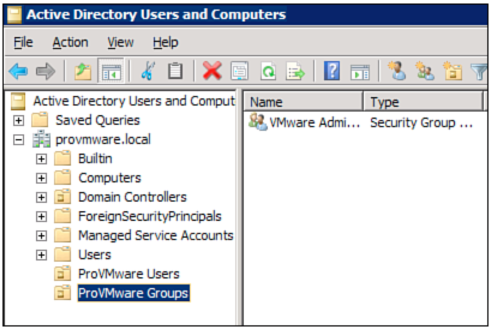


Figure 3.6 Active Directory hierarchy

DSQuery

In this example, there are two OUs in which the users and groups are kept: ProVMware Users and ProVMware Groups. These correlate to the User lookup base and Group lookup base in vCO. We also need to identify the location of VMware Administrators group to use as the vCO Admin group in vCO. To do that, we use the following DSQuery commands:

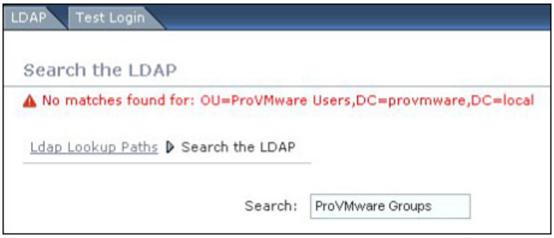
```
PS C:\> DSquery OU -name "ProVMware Users"
"OU=ProVMware Users,DC=provmware,DC=local"

PS C:\> DSquery group -name "VMware Administrators"
"CN=VMware Administrators,OU=ProVMware Groups,DC=provmware,DC=local"
```

This method gives you information that can be copied and pasted right back into the LDAP configuration page. The next method performs a search directly within the interface.

> **NOTE**
>
> The DSQuery tool is only installed by default on Windows 2008 servers that also have the AD Domain Services role installed. To obtain this tool on a server that does not have this role installed, you can download it from the Microsoft website.

vCO Interface LDAP Search

Using the same AD hierarchy from Figure 3.6, we use the vCO interface to fill the remaining values (see Figure 3.7).



Search the LDAP

⚠ No matches found for: OU=ProVMware Users,DC=provmware,DC=local

Ldap Lookup Paths ▶ Search the LDAP

Search: ProVMware Groups

Figure 3.7 The Search dialog

When your search is complete, it should only be a click away from populating the field you searched for.

Fill In the Form

Rather than call out each specific field in the form (because most of them are straightforward), I note some specific fields that warrant some special attention, as follows:

- Root
- Use SSL
- Use Global Catalog
- User/Password

The Root section is called out because this field is also expecting an LDAP path. To obtain this LDAP path, you want to grab the DC= components from your other LDAP fields. This provides vCO a root location from which to start its searching.

Called out next is Use SSL. My recommendation here is to use this setting because it provides SSL encryption for LDAP traffic between vCO and your LDAP server. This additional layer of security further hardens your vCO instance and helps reduce the attack surface by preventing potential information leakage. You cannot set this field, however, until you import the SSL certificate on the Network SSL tab. We cover this in the "Configure SSL" section, later in this chapter.

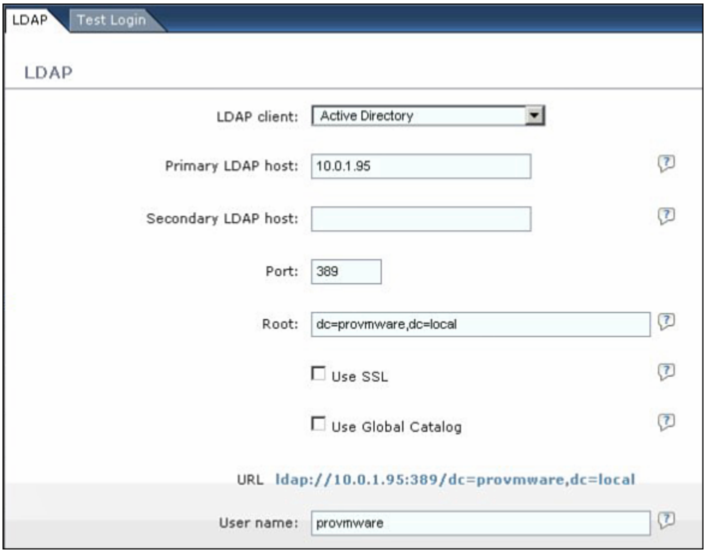
Next on our list of special fields is Use Global Catalog. Checking this option helps vCO in its lookups of Active Directory objects. This brings us to the final field, User/Password. My recommendation for this field is to use a specially created AD account for the vCO service, with a suitably long password.

Figure 3.8 summarizes the options chosen and the entries in the fields.

When you have finished filling out the form, click Apply Changes at the bottom of the page. vCO then runs a validation and calls out any errors in the log pane at the bottom of your window.

There are a few other things to note when configuring vCO to use LDAP that will save you some headaches down the road:

- Use an LDAP server that is physically close to your vCO installation. This reduces the latency of the queries that vCO must make and keeps performance snappy.
- Set the user and group lookups to the narrowest LDAP path possible. As in, don't target your entire directory because doing so will result in huge queries and slow down the entire vCO system.



LDAP

LDAP client: Active Directory

Primary LDAP host: 10.0.1.95

Secondary LDAP host:

Port: 389

Root: dc=provmware,dc=local

□ Use SSL

□ Use Global Catalog

URL ldap://10.0.1.95:389/dc=provmware,dc=local

User name: provmware

Figure 3.8 Screenshot of the options chosen and the filled-in fields

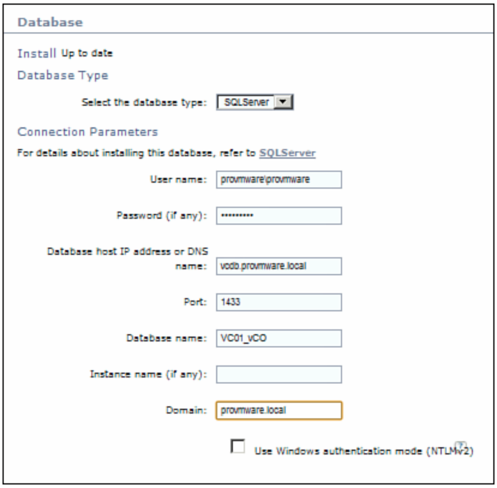
Database Configuration

Another critical component of vCO is the vCO Database. The vCO Database is used to maintain job or workflow execution state among other items. Thus, the database is key to vCO operating successfully. To configure a database for use with vCO, you must perform the following steps:

1. Create a database service account.
2. Create the database.
3. Select the database type.
4. Fill in the form.
5. Install the database.

The first two tasks are done on servers other than where your vCO server is. In fact, they may need to be done by another group or individual within your organization. If you are the domain admin, you first need to create an AD user, with a suitably secure password. This is used as a service account to connect from the vCO server to your database server. You then need to set this login up on your database server, as well, and grant it permissions to your specific database. The specifics of these activities are beyond our scope, however.

Next up, we fill in the form and create the database. In this example, we use the SQLServer database type. You want to refer to the vCO documentation for additional supported database types. After selecting the database type, you are presented with the form in Figure 3.9.



Database

Install Up to date

Database Type

Select the database type: SQLServer

Connection Parameters

For details about installing this database, refer to SQLServer

User name: provmware\provmware

Password (if any): ••••••••••

Database host IP address or DNS name: vcdb.provmware.local

Port: 1433

Database name: VC01_vCO

Instance name (if any):

Domain: provmware.local

□ Use Windows authentication mode (NTLMv2)

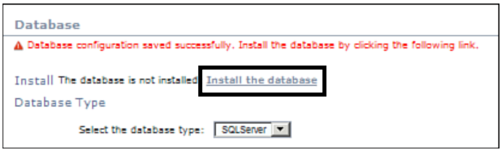
Figure 3.9 Configure database form

There are a couple of "gotchas" to look out for here. The first is to make sure that DNS is working and that the server name you are using is actually in DNS. The second is to ensure that your SQL server is configured for Windows Authentication; otherwise, you receive an error that looks like this:

```
Cannot connect to jdbc:jtds:sqlserver://vcdb.provmware.local:1433/VC01_
vCO;domain=provmware.local. Connection error was: Login failed. The login
is from an untrusted domain and cannot be used with Windows authentication.
```

After you've filled in the form, click Apply and vCO validates your inputs. Once you've corrected any errors, you are presented with the dialog shown in Figure 3.10.

When it's finished, the resulting window looks like Figure 3.11.



Database

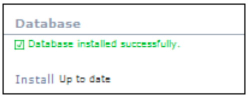
⚠ Database configuration saved successfully. Install the database by clicking the following link.

Install The database is not installed Install the database

Database Type

Select the database type: SQLServer

Figure 3.10 Install the database tables



Database

☑ Database installed successfully.

Install Up to date

Figure 3.11 vCO tables installed

Configuring SSL

vCO uses SSL in two different ways. First, vCO uses SSL to secure communications between vCO and vCenter Server instances. The second way vCO uses SSL is to digitally sign exports from your vCO server. We break this section then into the following two parts:

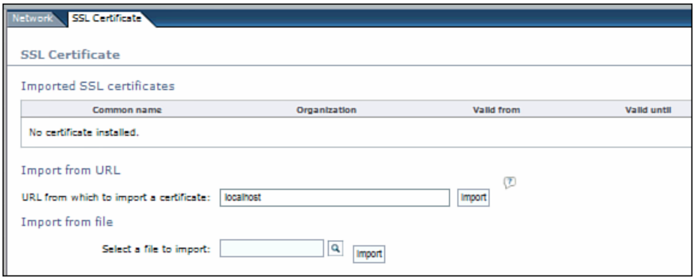
- Secure vCenter communications
- Configure the vCO host certificate

Secure vCenter Communications

Configuring vCO to use SSL communications with vCenter Server consists of the following steps:

1. Open the Networking control panel.
2. Select the SSL tab.
3. Provide an address and select Import from URL.
4. Click Import.
5. Under Startup Options, restart vCO.

Figure 3.12 shows the SSL Certificate tab and the Import from URL section. You need to repeat steps 2 and 3 for each vCenter Server that your vCO server will communicate with.



SSL Certificate

Imported SSL certificates

Common name	Organization	Valid from	Valid until
No certificate installed.			

Import from URL

URL from which to import a certificate: localhost

Import from file

Select a file to import:

Figure 3.12 Import from URL

NOTE

In the URL dialog, you only need to use the server's name or IP address. In fact, at the time of this writing, using https:// causes the interface to error out.

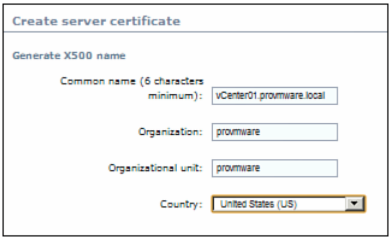
Configure the vCO Host Certificate

As mentioned at the beginning of this section, vCO can use SSL to sign objects that you export from vCO. Most commonly, these are packages. However, before we can sign a package, we need to make sure vCO has an SSL certificate to sign them with. There are three methods for doing this:

- Import an existing certificate.
- Create a self-signed certificate.
- Obtain a third-party certificate.

This section covers the middle option, using the vCO interface to create a self-signed certificate. Self-signed certificates are a good option when you are either resource constrained or do not otherwise run your own PKI (Private Key Infrastructure).

To create and install the self-signed certificate, you need to open the Server Certificate page of the vCO interface. Once on that page, fill in the relevant fields and click Create. Figure 3.13 shows a completed form.



Create server certificate

Generate X500 name

Common name (6 characters minimum): vCenter01.provmware.local

Organization: provmware

Organizational unit: provmware

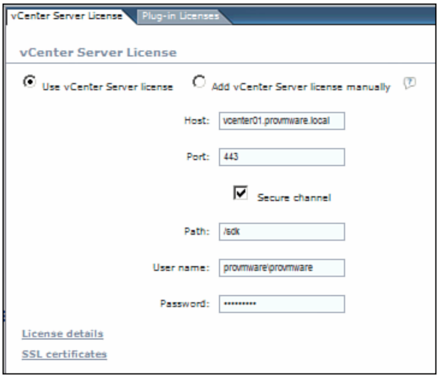
Country: United States (US)

Figure 3.13 A completed server certificate form for a self-signed certificate

Licensing the vCO Server

Next on our list of configuration tasks is the licensing of the vCO server. vCenter Orchestrator, starting at version 4.0.x to the present, uses the same license as your vCenter Server. From this license, it establishes which edition of vCenter Server you have and configures itself accordingly. It is important at this stage to state that you need to have completed the vCenter SSL configuration in the preceding section before performing the configuration in this section.

On this screen, you have two options: have vCO retrieve the key from vCenter or enter the key by hand. Because this book is about automating things, we let vCO grab the license automatically, as shown in Figure 3.14.



vCenter Server License

● Use vCenter Server license ○ Add vCenter Server license manually

Host: vcenter01.provmware.local

Port: 443

☑ Secure channel

Path: /sdk

User name: provmware\provmware

Password: ••••••••••

Figure 3.14 Licensing vCO from vCenter

Plug-In Configuration

This gets to be the second-to-last step in our configuration but is no less important for it. This is because, as we discussed in Chapter 1, "Introducing VMware vCenter Orchestrator," vCO is an orchestration engine that derives its power from its open plug-in architecture. Our job in this section is to configure both the credentials these plug-ins will use and which plug-ins will be enabled when the vCO service starts. To get vCO up and running, our configuration tasks are as follows:

- Provide plug-in credentials.
- Enable plug-ins for vCO startup.
- Install additional plug-ins.

Provide Plug-In Credentials

In providing the plug-in credentials, I recommend the same course of action as we took with the other required accounts in this chapter, and that is to create and provide credentials for a specific service account. This helps further harden vCO installation by providing an auditable user account for workflows, which have as few privileges as is needed to perform the actions defined for that plug-in. The user accounts you create for each plug-in must also be part of the vCO Admin group you specified when configuring LDAP.

NOTE

Each plug-in requires different vCenter Server permissions to perform various actions. The number and type of permissions required will vary based on which plug-ins you enable for your environment and therefore are beyond the scope of this book. However, you can seek guidance for role creation in the VMware Administrators Guide on the VMware website.

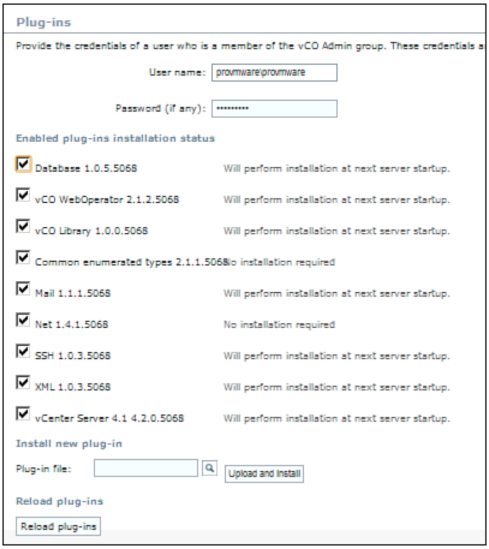
Enable Plug-Ins for vCO Startup

At the time of this writing, nine plug-ins that ship with vCO can be enabled out of the box:

- Database
- vCO Web Operator
- vCO Library
- vCO Enumerated Types

- Mail
- Net
- SSH
- XML
- vCenter Server

As you can see in Figure 3.15, each plug-in is enabled by default. For the installation used in this book, we stick with the defaults because doing so will help us work through our examples.



Plug-ins

Provide the credentials of a user who is a member of the vCO Admin group. These credentials a

User name: provmware\provmware

Password (if any): •••••••••

Enabled plug-ins installation status

☑ Database 1.0.5.5068	Will perform installation at next server startup.
☑ vCO WebOperator 2.1.2.5068	Will perform installation at next server startup.
☑ vCO Library 1.0.0.5068	Will perform installation at next server startup.
☑ Common enumerated types 2.1.1.5068	No installation required
☑ Mail 1.1.1.5068	Will perform installation at next server startup.
☑ Net 1.4.1.5068	No installation required
☑ SSH 1.0.3.5068	Will perform installation at next server startup.
☑ XML 1.0.3.5068	Will perform installation at next server startup.
☑ vCenter Server 4.1 4.2.0.5068	Will perform installation at next server startup.

Install new plug-in

Plug-in file:

Reload plug-ins

Figure 3.15 The vCO Plug-Ins Configuration page

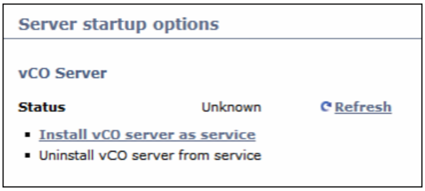
Install Additional Plug-Ins

We have discussed a few times already that a good amount of vCenter Orchestrator comes from the plug-ins that are available for it. In addition to the base plug-ins, you can find more on the vCO Product page (www.vmware.com/products/vcenter-orchestrator/overview.html) and from individual vendors such as EMC and NetApp. At this time, we will not be installing any additional plug-ins.

To complete the installation and enabling of plug-ins, restart your vCO-related services.

Installing vCO as a Service

This particular step is singlehandedly just as critical as the rest of the steps because it configures the vCO server to run as a system service. This means vCO will be available to configure and execute workflows around the clock. After all, what good is orchestration engine if you cannot run disruptive workflows after hours? To install vCO as a service, select the Startup Options configuration item and then Install vCO Server as a Service, as shown in Figure 3.16.



Server startup options

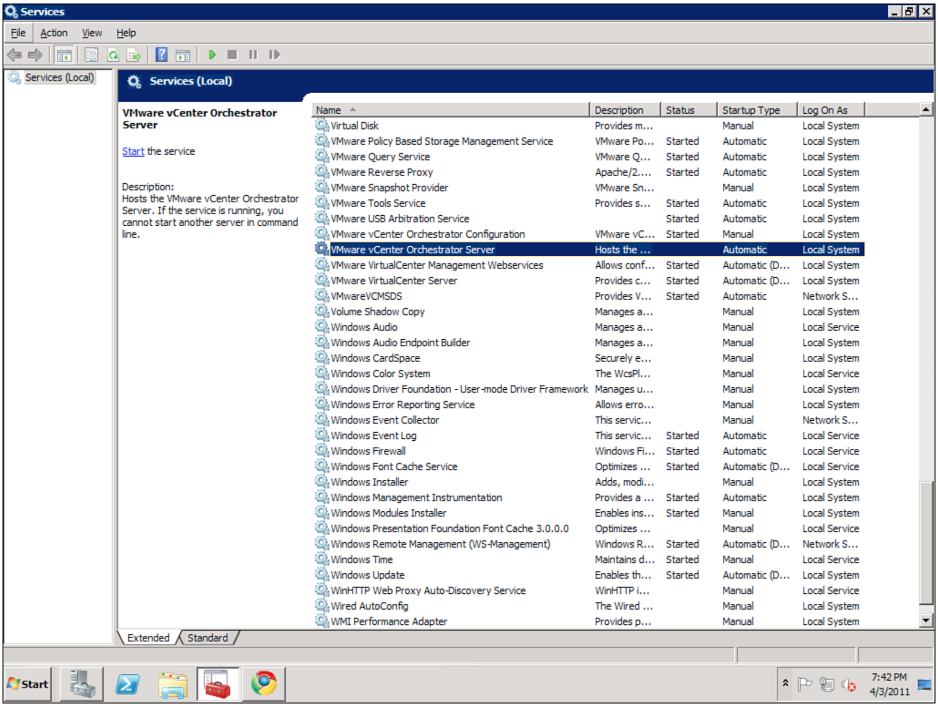
vCO Server

Status Unknown

- Install vCO server as service
- Uninstall vCO server from service

Figure 3.16 Choosing to install vCO as a service

You can also use this interface to monitor and control the vCO server status. Figure 3.17 shows the service now listed as a Windows service.



VMware vCenter Orchestrator Server

Start the service

Description:
Hosts the VMware vCenter Orchestrator Server. If the service is running, you cannot start another server in command line.

Name	Description	Status	Startup Type	Log On As
Virtual Disk	Provides m...		Manual	Local System
VMware Policy Based Storage Management Service	VMware Po...	Started	Automatic	Local System
VMware Query Service	VMware Q...	Started	Automatic	Local System
VMware Reverse Proxy	Apache/2....	Started	Automatic	Local System
VMware Snapshot Provider	VMware Sn...		Manual	Local System
VMware Tools Service	Provides s...	Started	Automatic	Local System
VMware USB Arbitration Service		Started	Automatic	Local System
VMware vCenter Orchestrator Configuration	VMware vC...	Started	Manual	Local System
VMware vCenter Orchestrator Server	Hosts the ...		Automatic	Local System
VMware VirtualCenter Management Webservices	Allows conf...	Started	Automatic (D...	Local System
VMware VirtualCenter Server	Provides c...	Started	Automatic (D...	Local System
VMwareVCMSDS	Provides V...	Started	Automatic	Network S...
Volume Shadow Copy	Manages a...		Manual	Local System
Windows Audio	Manages a...		Manual	Local Service
Windows Audio Endpoint Builder	Manages a...		Manual	Local System
Windows CardSpace	Securely e...		Manual	Local System
Windows Color System	The WcsPl...		Manual	Local Service
Windows Driver Foundation - User-mode Driver Framework	Manages u...		Manual	Local System
Windows Error Reporting Service	Allows erro...		Manual	Local System
Windows Event Collector	This servic...		Manual	Network S...
Windows Event Log	This servic...	Started	Automatic	Local Service
Windows Firewall	Windows Fi...	Started	Automatic	Local Service
Windows Font Cache Service	Optimizes ...	Started	Automatic (D...	Local Service
Windows Installer	Adds, modi...		Manual	Local System
Windows Management Instrumentation	Provides a ...	Started	Automatic	Local System
Windows Modules Installer	Enables ins...	Started	Manual	Local System
Windows Presentation Foundation Font Cache 3.0.0.0	Optimizes ...		Manual	Local Service
Windows Remote Management (WS-Management)	Windows R...	Started	Automatic (D...	Network S...
Windows Time	Maintains d...	Started	Manual	Local Service
Windows Update	Enables th...	Started	Automatic (D...	Local System
WinHTTP Web Proxy Auto-Discovery Service	WinHTTP i...		Manual	Local Service
Wired AutoConfig	The Wired ...		Manual	Local System
WMI Performance Adapter	Provides p...		Manual	Local System

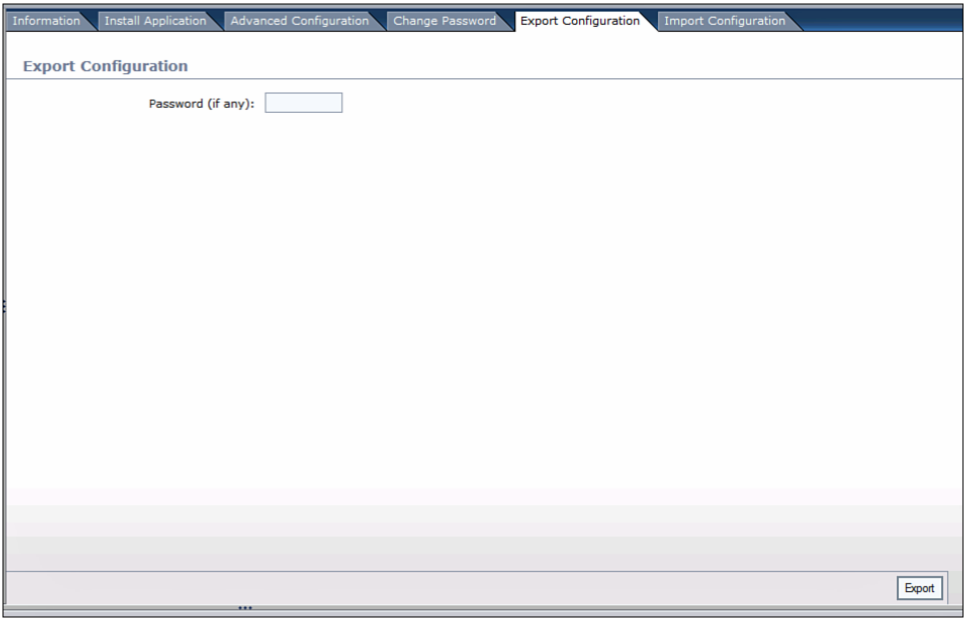
Figure 3.17 vCO installed as a Windows service

Backing Up the vCO Configuration

Now that you have vCO configured, it is a good thing to back up this configuration. This helps you get back to a working state should vCO become unusable in the future, with change control against future changes, and helps you deploy a new vCO engine later. To back up the vCO configuration, do the following:

1. Open the vCO configuration page at http://<vco server>:8282.
2. Log in.
3. Select Genera.
4. Click the Export Configuration tab.
5. Click the Export button.
6. Choose a location to save the file.

Figure 3.18 shows the Export Configuration page.

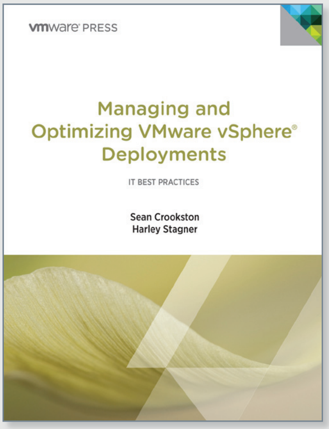


Password (if any):

Figure 3.18 The Export Configuration page

Summary

In this chapter, we covered all the steps necessary to bring a freshly installed vCenter Orchestrator system to a functional status. This included setting a service password, configuring vCO networking, LDAP authentication, installing the vCO Database, vCenter SSL communication, SSL signing for vCO packages, licensing the vCO server, and finally configuring vCO to run as a system service. It is now with a working vCO installation that we move into Part II of this book, “Working with vCenter Orchestrator,” and discuss the various vCO concepts.



vmware PRESS

Managing and Optimizing VMware vSphere® Deployments

IT BEST PRACTICES

Sean Crookston
Harley Stagner

CHAPTER 3 Operating the Environment

JULY 2012

Available in Print and eBook formats and through SAFARI BOOKS ONLINE

SHARE WITH OTHERS

vmware® PRESS

BUY NOW

Managing and Optimizing VMware vSphere Deployments

BY HARLEY STAGNER & SEAN CROOKSTON

Table of Contents

ISBN: 9780321820471

vmware.com/go/vmwarepress

ALWAYS LEARNING

PEARSON

Chapter 3

Operating the Environment

This chapter focuses on maintaining and monitoring an active environment. At this point, you might or might not have designed an optimal environment. The environment also might not have been implemented to your standards. After all, sometimes you can't entirely fix what is currently broken and must deal with it for a period of time.

In the field, we see the excitement in customers' eyes at the power that VMware brings to their infrastructures. Cost savings through hardware, high availability, and ease of management are the main things they are eager to take advantage of. However, this excitement sometimes leads to a lack of focus on some of the new things that must be considered with a virtual infrastructure. A lack of maintenance and insufficient or no monitoring are two huge issues that must be considered. Before delving into maintaining and monitoring a virtual infrastructure, this chapter talks about some other operational items that you might not have considered in the design.

Backups

A virtual infrastructure can pose different challenges for backups in terms of a technical understanding of the environment. This is the main reason we see that backups are not being adequately performed. Every organization has its own set of requirements for backups, but consider the following as important items for a backup strategy:

- An appropriate recovery point objective (RPO) or the ability to roll back to a period of time from today
- An appropriate retention policy, or the number of copies of previous periods of times retained

- An appropriate recovery time objective (RTO) or the ability to restore the appropriate backups in a set time
- An appropriate location of both onsite and offsite backups to enable recovery of data in the event of a complete disaster, while still allowing for a quick restore onsite where needed
- The ability to properly verify the validity of your backup infrastructure through regular testing and verification

Furthermore, outside of a technical understanding of the virtual infrastructure, virtualization poses no other significant challenges to maintaining a backup strategy. In fact, it will actually enable easier and quicker restores if properly designed.

When considering your backup strategy, you need to consider your RTO and RPO. You also need to consider your retention policy and proper offsite storage of backups. Properly storing offsite copies of backups is not just about keeping copies offsite that allow a quick restore to a recent restore point. It is also about considering what to also keep onsite so that simple restores are just that. Beyond that, you need to make sure you have all the small details that make up your infrastructure. This includes credentials, phone numbers for individuals and vendors, documentation, and redundancy in each of these contacts and documentation locations.

When considering backups, you need to determine the proper mix of file-level backups or virtual machine–level backups. Some organizations continue to do backups from within the guest that can provide a bare-metal restore. This is still a good option, and it might be your only option because of the software you presently use for backups; however, it will not be as quick to restore as a backup product that uses the VMware vStorage APIs to provide a complete virtual machine restore.

Let’s take a moment to talk about the verification and monitoring of your backups. Taking backups is not the solution to the task of creating a backup strategy. The solution is the ability to restore the missing or corrupted data to a point in time and within a certain time as dictated by your businesses requirements. Therefore, it is always important to regularly test restoration practices and abilities as well as monitor for issues with backup jobs. Your backup product should be able to verify the data was backed up and not corrupted; however, you should also schedule regular tests to verify this.

And, finally, let’s talk about snapshots. Snapshots are not backups, but in some environments they are used in that fashion. Snapshots are useful when performing updates on a virtual machine as a means of quick rollback; however, they should not be used long term. We’ve witnessed two main things that occur as a result of snapshots being left behind.

For starters, they result in data needing to be written multiple times. If you have three snapshots, any new data is written to all three. As you can see in Figure 3.1, blocks of data that need to be written are written to each snapshot file, resulting in a performance hit as well as increased space utilization. Multiply this by several virtual machines and possibly even worse by multiple nested snapshots, and it is no wonder that we see datastores fill up because of old snapshots. This can bring virtual machines to their knees and makes rectifying the situation complex. When consolidating snapshots, you need to have space available to write the data to the original virtual machine disk. In this case, you would not have that available, requiring the migration of virtual machines to other datastores.

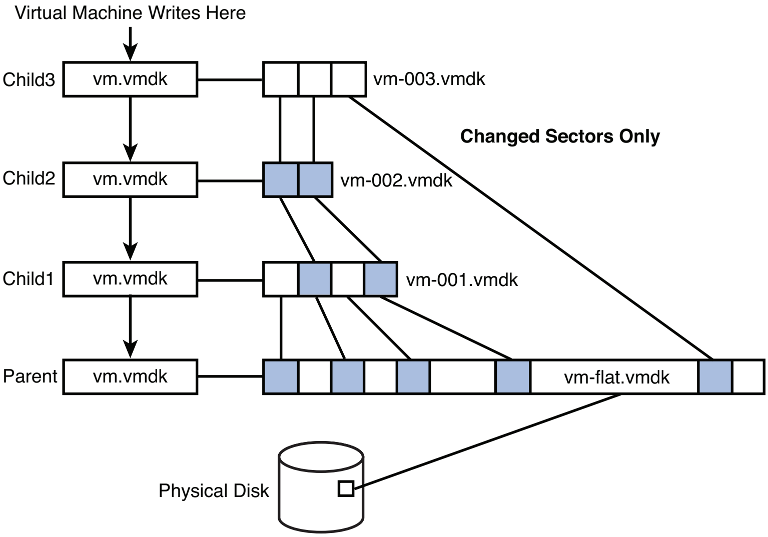


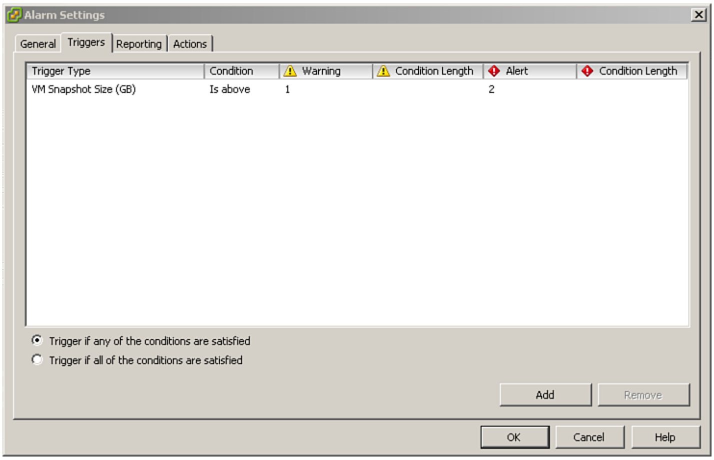
Figure 3.1 Snapshot Disk Chain

A second problem we have seen many times is often caused by full datastores. Snapshot corruption can occur as a result, leading to the disappearance of any data since the time of the snapshot creation. For example, assume a single snapshot was taken six months ago, right after you installed Windows for your new Exchange server. If that snapshot is corrupted, you will likely be able to repoint to the original VMware Disk (VMDK) file; however, you'll be left with a bare Windows virtual machine. Full datastores are not the only time snapshots can be corrupted. This can also occur as a result of problems during snapshot consolidation or manipulating the original virtual machine disk file from the command line while snapshots are present.

It is important to note that a snapshot itself contains only the changes that occur after the snapshot was taken. If the original virtual machine disk is corrupted, you will lose all of your data. Snapshots are dependent on the virtual machine disk.

VMware's Knowledge Base (KB) article 1025279 discusses in detail the best practices when using snapshots. In general, we recommend using snapshots only as needed and for short periods of time. We recommend configuring alarms within vCenter to notify of snapshot creation and regularly checking for snapshots in your environment. There are many PowerShell scripts available that will accomplish this; however, a great tool to have that includes snapshot reporting is PowerGUI (see Appendix A, "Additional Resources," for reference).

Within vCenter, no default alarms exist to alarm for snapshots. You can, however, create a virtual machine alarm with the following trigger to alarm for snapshots, as shown in Figure 3.2. This will help you with snapshots that have been left behind for some time and have grown to 1GB or larger; however, it will not help until the total amount of snapshot data written for a virtual machine totals 1GB. This chapter discusses alarms later, but you can also check out VMware Knowledge Base article 1018029 for a detailed video demonstration of creating an alarm like this one (see Appendix A for a link).



Trigger Type	Condition	Warning	Condition Length	Alert	Condition Length
VM Snapshot Size (GB)	Is above	1		2	

● Trigger if any of the conditions are satisfied
○ Trigger if all of the conditions are satisfied

Figure 3.2 Configuring Snapshot Alarms

Data Recovery

Like many products that use the VMware vStorage APIs, VMware's Data Recovery provides the ability to overcome backup windows. That is not to say you might not want to consider backup windows because you also must consider the traffic that will occur on the network during backups; however, backup windows are of less concern for a few reasons. For starters, Data Recovery provides block-based deduplication and only copies the incremental changes. This occurs from a snapshot copy of the virtual machine that enables virtual machines to continue running while Data Recovery performs the backup from that snapshot copy.

Data Recovery is not going to be the end-all solution to your backup strategy, though. Its intention is to provide disk-based backup storage for your local storage and there is not a native method built in to transfer these backups to tape or other media. Therefore, VMware Data Recovery is best thought of as a complementing product to an existing backup infrastructure. With that said, let's talk about some of the capabilities the product has.

The process to get backups up and running is straightforward:

- Install Data Recovery.
- Define a shared repository location.
- Define a backup job.

Installing Data Recovery

The first thing you need to verify is whether the product will meet your needs. Some of the more common things to consider when implementing Data Recovery are as follows:

- As previously mentioned, Data Recovery is intended to provide a quick method for onsite restores and does not provide offsite capabilities.
- Furthermore, you need to be sure all of your hosts are running ESX or ESXi 4.0 or later.
- Make note that each appliance supports 100 virtual machines with eight simultaneous backups. There is also a maximum of ten appliances per vCenter installation.
- The deduplication store requires a minimum of 10GB of free disk space. When using CIFS, the maximum supported size is 500GB. When using RDM or VMDK deduplication stores, the maximum supported size is 1TB.
- There is a maximum of two deduplication stores per backup appliance.
- Data Recovery will not protect machines with fault tolerance (FT) enabled or virtual machines disks that are marked as Independent.

For a complete list of supported configurations, refer to the *VMware Data Recovery Administration Guide*.

There are two steps to get the appliance installed. First, install the vSphere Client plug-in. Second, import the OVF, which will guide you through where you want to place the appliance. Once completed, you might want to consider adding an additional hard disk, which can be used to store backups.

Defining a Shared Repository

As discussed, each appliance will be limited to two shared repositories and depending on the type of repository, you will be limited to either 500GB (CIFS) or 1TB (virtual hard disk or RDM). You have the following options when choosing to define a shared repository:

- Create an additional virtual hard drive (1TB or less).
- Create a CIFS repository (500GB or less).
- Use a RDM (1TB or less).

If you choose to create and attach an additional virtual hard disk, you need to consider where you are placing it. As mentioned previously, the intention of Data Recovery is to deliver the capability of a quick onsite restore. The use of virtual hard drives provides for the best possible performance. If you use a virtual hard disk, though, you will be storing the backups within the environment they are protecting, so you must consider this carefully. You could store the virtual hard disk on the plentiful local storage that may be present on one of the hosts. You could also store the virtual hard disk on any IP-based or Fibre Channel datastore.

Our recommendation in this case is to use the local storage of one of the hosts if it is available. When given the choice between the two, consider the likelihood of your shared storage failing versus the local storage of a server failing. Additionally, consider the repercussion of each of those failing. If your shared storage were to fail with the backups on them, you would have to use your other backup infrastructure to restore them, which can be quite time consuming. If the local server with your backups on them were to fail, then if a complete disaster occurs you are still going to have the production copies running on shared storage. If you do have a complete site failure, then you are going to need to deploy your disaster recovery strategy. This is discussed further shortly.

Another option is to use a Raw Device Mapping (RDM). If you are using the same storage as your virtual infrastructure, you are taking the same risks. The only way to mitigate such risks is to use storage dedicated for the purposes of backups. Just like the option of using virtual disks, think about where you are going to restore that data to if a disaster occurs. If your storage device is gone, you are going to have to initiate your disaster recovery strategy.

Another option is to use a CIFS share. Remember that CIFS shares are limited to 500GB, so each appliance can only support 1TB of CIFS repositories with its two-repository limit. Although the product lets you configure a CIFS share greater than 500GB, it warns you not to do so. We recommend that you listen to the warning because testing of the product has proven that creating a large CIFS repository can cause Data Recovery to fail to finish its integrity checking, which in turn causes backups to not run.

Another consideration for CIFS is that the share you are sharing out, and for that matter the disk that is being used, should not be used for any other function. Remember that Data Recovery provides for block-based data deduplication. If other data exists on the back-end disk, this can also cause a failure in integrity checking and, in turn, a failure of backup jobs running.

Defining a Backup Job

Now that the appliance is set up and you have set up one or two repositories, it is time to create the backup jobs. Backup jobs entail choosing the following:

- Which virtual machines will be backed up
- The backup destination
- The backup window
- The retention policy

Choosing Which Virtual Machines to Back Up

The virtual machines you choose to back up can be defined by an individual virtual machine level or from vCenter, datacenter, cluster, folder, or resource pool levels as well. Note that when you choose a virtual machine based on the entity it is in, if it is moved it will no longer be backed up by that job.

Choosing a Backup Destination

Your choice of a destination might or might not matter based on the size of your infrastructure or your backup strategy. For sizing purposes, consider that you could exceed the capacity of the deduplication store if you put too many virtual machines on the same destination. For purposes of restoring data, consider the placement of the backups and where it is in your infrastructure.

Defining a Backup Window

Backup windows dictate when the jobs are allowed to run; however, they do not have a direct correlation to the exact time they will execute. By default, jobs are set from 6:00

a.m. to 6:00 p.m. Monday through Friday and all day Saturday and Sunday. Consider staggering the jobs so that multiple jobs do not run simultaneously if you are concerned with network throughput.

Defining a Retention Policy

When choosing a retention policy, you have the option of few, more, many, or custom. Custom allows specifying the retention of as many recent and older backups as required. The other options have their defaults set, as shown in Table 3.1.

Table 3.1 VMware Data Recovery Retention Policies

Retention Policy	Recent Backups	Weekly	Monthly	Quarterly	Yearly
Few	7	4	3	0	0
More	7	8	6	4	1
Many	15	8	3	8	3

Changing any one of the settings for these policies will result in the use of a custom policy. When choosing your retention policy, ensure you have the capability to restore data from as far back as you need, but within the confines of the storage you have to use for backups.

At this point, your backups are up and running. You can either initiate a backup now or wait until the backup window has been entered for backups to begin. Once you've seen your first successful backup, you still have a few other items to consider.

Restoring Data (Full, File, Disks) Verification

When restoring data, you have two key things to consider. When choosing to restore data, you first need to choose your source. A virtual machine can be part of multiple backup jobs, so in addition to having a different set of restore points, you might also have a set of restore points that are also located on a different backup repository. Second, you need to consider where you want to restore the data.

For the purposes of testing the capability to restore, you can perform a restore rehearsal by doing the following from within the Data Recovery interface by right-clicking a virtual machine and then clicking the Restore Rehearsal from Last Backup option. To fully test a restore or to perform an actual restore, you have much more to consider because this option chooses the most recent restore and restores the virtual machine without networking attached. The following sections discuss those considerations further.

Choosing Backup Source

When restoring, you have the option to restore at any level in the tree, so you can restore entire clusters, datacenters, folders, resource pools, or everything under a vCenter server. When looking at the restore of an individual virtual machine, you can restore the entire virtual machine or just specific virtual disks. You may also restore individual virtual machines from the virtual machine backup, which is discussed shortly.

Choosing Restore Destination

When restoring, you have several options during the restore, including choosing where to restore the data. When considering restoring an entire virtual machine, you have the following options to consider:

- Restore the VM to a specific datastore.
- Restore the virtual disk(s) to a specific datastore(s).
- Restore the virtual disk(s) and attach to another virtual machine.
- Choose the Virtual Disk Node.
- Restore the VM configuration (yes/no).
- Reconnect the NIC (yes/no).

When restoring, the default setting is to restore the virtual disk in place, so be careful to consider this if it is your intended result. If possible, in all situations you should restore to another location to retain the set of files that is currently in place if further restore efforts are needed on those files.

File Level Restores

In addition to restoring complete virtual machines or specific disks, you may also restore individual files. File Level Restore (FLR) allows for individual file restoration with an in-guest installed software component. The FLR client is available for both Windows and Linux guests and must be copied off the Data Recovery media locally where it will run. By default, Data Recovery only allows the restoration for files from a virtual machine for which the client is being run; however, if you run the client in Advanced mode, you can restore files from any of the virtual machines being backed up. Note that although you are able to mount Linux or Windows virtual machines regardless of the operating system you are running, you might not be able to read the volumes themselves.

Once mounted, you have the ability to copy files and restore them to locations manually or look through them to find the version you are looking for. The mounted copies are read-only versions of the files, and any changes made will not be saved, so make sure to copy the files to a desired location before making any changes.

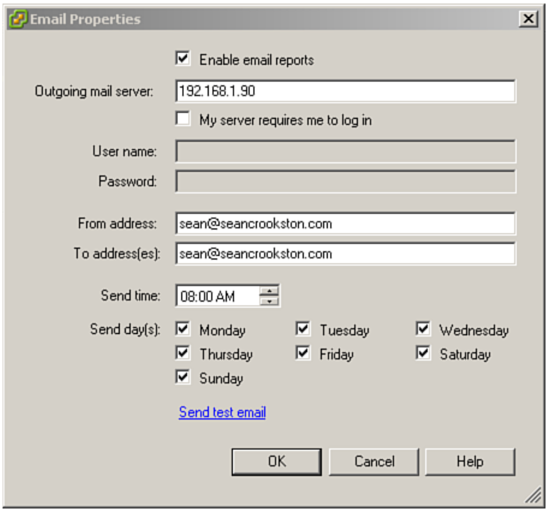
One last note on the use of FLR when using Data Recovery: It is not recommended and Data Recovery should be configured so that File Level Restores are not possible. This is done by configuring the VMware Data Recovery .ini file and setting EnableFileRestore to 0.

Site Disaster

As mentioned previously, the intended use of this product is for quick restores and is not intended to be your disaster recovery plan. If you were to lose a vCenter server and needed to recover another machine, you would have to stand up a new vCenter server and install the plug-in to use Data Recovery to restore the virtual machine. Additionally, if you lose the appliance itself, you must install a new one and import the repository. Be aware that this can take a long time if a full integrity check is required.

Monitoring Backup Jobs

Data Recovery allows the configuration of an email notification that can be sent as often as once a day at a specified time. There isn't much to configure with email notification, as shown in Figure 3.3. The important thing is to make sure the appropriate individuals are being notified and that mail is being relayed from the outgoing mail server specified. Remember the server that needs to be authorized is not the vCenter server but rather the Data Recovery appliance itself.



Email Properties

☑ Enable email reports

Outgoing mail server: 192.168.1.90

□ My server requires me to log in

User name:

Password:

From address: sean@seancrookston.com

To address(es): sean@seancrookston.com

Send time: 08:00 AM

Send day(s): ☑ Monday ☑ Tuesday ☑ Wednesday ☑ Thursday ☑ Friday ☑ Saturday ☑ Sunday

Figure 3.3 Configuring Data Recovery Email Notifications

Managing the Data Recovery Repository

The maintenance tasks that run will check the integrity of the data in the repositories and reclaim space in the deduplication stores. By default, Data Recovery is set to be able to run maintenance at any time. This might not be a problem for your environment; however, when integrity check operations are occurring, backups cannot. Therefore, you should change the maintenance window so that it is set to run during a specified period of time. This ensures backups will always have the time to run each day.

When the deduplication store is using less than 80% of the repository, the retention policy is checked weekly to remove any restore points that are outside the specifications. This means that you might have many more restore points than expected as a result. Once 80% of the repository is utilized, the retention policy is checked daily. In the case of the repository filling up, the retention policy is run immediately if it has not been executed in the last 12 hours.

Disaster Recovery

Disaster recovery is an area of many organizations that has at least some, if not a lot, of room for improvement. When looking at a disaster recovery plan, the following things are important to consider:

- Data is available with an RTO that meets the business's requirements to operate and the data is from a point in time that meets the RPO of the organization.
- Data has been verified as being valid.
- A runbook has been defined for how to and in what order to restore.

The first point is present in most organizations, whereas the second and third are not. It should not be surprising because a failure warranting declaring a disaster is not often needed. Nonetheless, a solid runbook should be defined for your infrastructure. A runbook for restoration for your virtual infrastructure is crucial; however, consider the back-end networking infrastructure first as your virtual machines will be of no use without it.

When looking at recovering your virtual infrastructure, the ideal setup is to replicate among your storage devices and use VMware's Site Recovery Manager (SRM) to automate your restore. Site Recovery Manager is further discussed later in this chapter; however, for those not familiar, it assists in automating the recovery of virtual machine environments during a disaster.

You may also use the set of replicated data to manually configure the virtual machines and power them on at your disaster recovery location. Additionally, you can choose another method of manual restoration. This could be using a copy of the virtual machine files from some other mechanism or using a backup product to perform a bare-metal copy of the machine and restoring it to a newly configured virtual machine. For the purposes of this discussion, we talk in detail about the use of VMware's Site Recovery Manager as it provides the best mechanism. Before doing that, though, the following sections talk briefly about the other options.

Manual Disaster Recovery

When looking at implementing a manual data recovery plan, you need to ensure you are doing a few things that Site Recovery Manager would be automatically handling or assisting with. Many times, the use of manual methods is the result of a lack of sponsorship of the initiative in terms of funding; however, that does not mean the process cannot work. If you are creating a manual data recovery plan consider the following.

- Ensure data is being replicated/copied and is current with your RPO.
- Ensure your processes for restoration meet your RTO.
- Ensure the Disaster Recovery (DR) site hardware is supported and will support the load in the event of a disaster.
- Ensure the recovery processes work by performing regular DR tests.
- Ensure the runbook is updated regularly as network, application, and other requirements change.

By keeping these points in mind, your disaster recovery efforts will be successful; however, you will have to perform many of the steps manually.

Whereas storage replication was previously a condition for using Site Recovery Manager, the latest version now supports host-based replication. If you were previously unable to use Site Recovery Manager because of the storage replication requirement, you should reevaluate the product with host-based replication.

Site Recovery Manager

VMware offers a product called Site Recovery Manager that helps automate most of the process of recovering virtual machines during a disaster. The product allows for isolated testing to ensure recovery is possible in the event of an actual disaster as well as the ability to failback in version 5.0.

When installed at both the production and disaster recovery locations, the product provides for a centralized approach to defining replication and recovery plans. In prior versions, SRM relied on the storage itself to perform the replications and integrated with the storage using a supported Storage Replication Adapter (SRA). This limited the product for some entities with supported storage. Even those with supported storage devices in both locations might not have had matching storage solutions and, hence, no supported replication infrastructure in place.

SRM 5.0, however, has expanded its market base with the introduction of vSphere Replication (VR). This allows replication from one location to another, regardless of the type of storage on both ends. One or both ends can even be local or directly attached storage. SRM is also protocol independent so you can replicate among Fibre Channel, iSCSI, or NFS storage.

For more information on Site Recovery Manager, check out *Administering VMware Site Recovery Manager 5.0* by Mike Laverick. This book provides an in-depth discussion of the product, using it in a number of scenarios, and is a great read when defining a disaster recovery solution in a virtualized environment.

Physical to Virtual Conversions

When a virtualization infrastructure is implemented, the first virtual machine installed is typically going to be brand-new installations of Windows for your vCenter and SQL servers. Many times, a project like this also serves as a good time to clean up and move toward the latest server operating systems. Regardless at some point, you need to begin moving some of the existing physical workloads over and retaining their configurations.

VMware provides a free download for a product that will assist in this migration. VMware vCenter Converter Standalone 5.0 allows the conversion of physical systems as well as systems that are already virtualized. When performing physical to virtual conversions, you should be aware of the following things.

For all systems, you should do the following:

- **Prior to conversion**
    - Perform a survey of the server and its applications.
    - Identify server and application owners for approval and verification testing upon completion.
    - Identify performance and configuration of CPU/memory/disk versus actual usage.
    - Identify destination for virtual machine (host placement and storage placement).

    - Identify and record network configurations.
    - Identify downtime and schedule for system(s).

- **During conversion**
    - Place virtual machine on host and storage per design.
    - Adjust configurations of CPU/memory/disk as appropriate.

- **After conversion**
    - Remove nonpresent devices.
    - Remove legacy software.
    - Install VMware Tools.
    - Reconfigure the network.
    - Perform basic testing.
    - Verify functionality with server and application owners.
    - Fully uncable and remove decommissioned physical servers.

Issues and Troubleshooting

Physical to virtual conversions can fail to start. Typically, we have found this is because of one of the following reasons:

- No Permissions Admin$
- Firewall exists between server to be converted and vCenter
- Incorrect DNS configurations

In some cases, administrators have removed the admin$ share on a server, which is required for the vCenter Converter agent installation when installing remotely. You can install the client locally or re-create the share to resolve these issues.

When a firewall exists, it can cause a failure if certain ports are not reachable. VMware Knowledge Base article 1010056 details the required ports to be allowed through the firewall (see Appendix A for a link).

If DNS configurations are not correct on a source virtual machine, this can also cause failures. Ensure DNS is correct or update DNS so that the system to be converted can reach the vCenter server.

Besides these basic considerations, you should also consider a few special cases:

- Domain controllers
- Windows Server with OEM installations
- SQL, Exchange, and other applications servers
- Linux
- Virtual to virtual conversions (V2V)

Domain Controllers

Active Directory domain controllers have special considerations to take when looking at physical to virtual conversions. Although there are methods to perform a P2V conversion, it is our recommendation you don't and instead create new virtual servers. Domain controllers are extremely sensitive to hardware changes and a failure in following the P2V process at any step can cause major replication issues that will be visible throughout your infrastructure. Creating new servers allows for a clean and safer migration. The recommended process is as follows. Note this may vary depending on how many Active Directory domain controllers you have, their location, and whether they remain physical or not.

- Create one or more brand-new Windows virtual machines.
- Promote these two domain controllers using dcpromo.
- Ensure replication via the command line using the repadmin tool.
- Transfer FSMO roles.
- Transfer DHCP servers.
- Verify role transfer.
- If the domain controller is also a DNS server, ensure DNS is running on new systems.
- Power down physical domain controller and verify functionality. If no issues exist, power back on and demote existing physical systems using dcpromo (FSMO role will be transferred there also).
- Decommission physical hosts.
- Reconfigure DNS settings for servers/workstations to reflect any new IP addresses.

Windows Server with OEM Installations

Another case for consideration is any physical Windows servers that were originally installed with OEM media. Per Microsoft's licensing agreement, these licenses are tied to the original physical hardware and as a result the right to continue using the installed operating system does not carry over. Microsoft licensing is outside the scope of this book; however, two things can be drawn from this situation:

1. You are not in compliance with your licensing. You need to have or purchase an additional Microsoft server license.
2. You are running a version of Windows you are not licensed for. You need to either install a fresh instance of Windows using volume licensed media or perform an in-place upgrade using the volume licensed media.

If you have ignored these recommendations, you will find that once you bring the newly converted virtual machine online, it will require activation. With OEM installation, the key that is required to activate is not necessarily the one that was on the sticker on the box. With that said, you might be able to activate the software installation; however, you need to ensure you are in compliance with Microsoft licensing as soon as possible thereafter.

Although vendor background screens are typically a clear indication that OEM installations are present, you can also check the product ID for OEM to verify. If it is not present, you are fine. You can script this to check for multiple servers in your environments using various product and key software products that reveal this information. Additionally, you can use code like this PowerShell snippet to gather this information. The following code sets $Prod_ID to the ProductID of the machine it is run on:

```
$Prod_ID=(get-item 'HKLM:\Software\Microsoft\Windows NT\
  CurrentVersion').getvalue('ProductID')
```

SQL, Exchange, and Other Applications Servers

Although vCenter Converter will make multiple passes through your data, it is important to consider what could be lost in the transition time in between bringing the physical host down and the last time the data was copied. As a result, it is recommended that you stop all application services.

In addition, consider file shares. You may declare in an email that the server will be down this Saturday, but that doesn't stop someone from changing data. This data could end up being changed at a time of transition and would in effect be lost. Therefore, it is recommended that you unshare any file shares during the conversion process. Additionally, you should remove any temporary files or any data that is no longer needed. This greatly decreases the time involved when converting and optimizes the amount of physical storage being used.

Linux

This chapter has talked a lot about virtualizing Windows servers, but now the focus shifts to Linux servers. Linux servers follow a slightly different process and, as a result, there are some things you should be aware of. For starters, the process for Linux does not deploy an agent, but instead a helper virtual machine is deployed on the destination vSphere host. This helper machine will ultimately become the production virtual machine once it has copied all of the data from the physical Linux machine.

You should consider a few important things:

- You must ensure you have SSH access to the Linux machine when doing an online conversion and you must have root access when doing so.
- Only certain flavors of Linux are supported for online conversions. Presently, these are certain versions of Red Hat, SUSE, and Ubuntu.
- Customization during the conversion process is not supported for Linux guest operating systems.

V2V and Other Methods

Physical machines are not the only machines that may need to be brought into a new infrastructure. For example, you may have existing virtual machines running on storage not accessible to the new environment. You may also have these virtual machines running in a Hyper-V environment. Additionally, you may choose to do an offline conversion by using an imaging product. Regardless of the source, vCenter Converter allows for all of these options when using any of the following supported methods.

The following virtual machine formats are supported for cold conversion:

- Microsoft Hyper-V
- Microsoft Virtual Server
- Microsoft Virtual PC
- Parallels Desktop
- VMware Workstation, GSX Server, Player, Server, Fusion, ESX

The following image formats are supported:

- Symantec Backup Exec System Recovery
- Norton Ghost
- Acronis
- StorageCraft

Additionally, during the conversion process, you can convert any running Windows virtual machine by specifying a powered-on machine as the source.

Offline Boot Disc

When an offline conversion is desired in addition to the mentioned image formats, you can use the VMware vCenter Converter Boot CD. The boot CD is no longer provided as of vCenter Converter 4.3; however, it is still available for download with valid support for a vSphere 4.x Enterprise Edition or greater license. At the time of this writing, there are no current plans to release a vCenter Converter Boot CD for vSphere 5.0; however, version 4.0.1 build 16134 is the latest version and is supported for conversions from a source to a vCenter 5 infrastructure.

The offline boot disc is based on Windows PE and allows the import of network drivers to build a new image if required. The lack of network drivers is the most common reason the offline boot disc does not work.

Maintenance

Maintaining a vSphere-based virtual infrastructure is very important. After all, you have a large number of operating systems now running collectively on a much lesser amount of physical hardware in most cases. A failure to update for and then be exposed to a potential flaw may now put your entire infrastructure at risk instead of only some servers.

Why do organizations not properly maintain their vSphere environments? Everyone agrees with the criticality of maintaining servers whether it is through patches or regular release updates, but still it remains a large problem in many environments. In large part, the main driving force to perform any update is a result of an enhancement release that has added additional features.

Update Manager

One reason many administrators do not update their infrastructures is due to a lack of understanding of the process. Maybe they are new to VMware and never bothered to even install Update Manager with vCenter. Update Manager is not a requirement to patch systems but the process does become much more involved when using the command-line interface to do so. An administrator must download the update bundle and transfer it to each of the hosts. Then a command-line process must be invoked from each of the hosts. In the days before Update Manager, it is no wonder why some administrators might have chosen to patch less frequently or not at all.

vSphere is a hardened hypervisor and, as a result, needs much less patching and updating for vulnerabilities than a typical operating system. Many administrators, though, take this as a reason not to patch at all.

Some also entirely understand the advantages of Update Manager and have it installed and running. They realize how the effect of an issue with their vSphere infrastructures could now affect all their operating systems instead of just a handful. As a result, they view this increased impact of any updates as possibly negative. This may be the proper viewpoint as certain vulnerabilities may not be a high risk for their environments. They are further justified in their decision in knowing that the impact of any issues that occur in a virtual infrastructure can be huge if not properly planned. Perhaps the feature that is affected is also not something they are using. Being cautious and properly planning and testing for updates is certainly the way to go. To date, I have never worked directly with anyone who has been exploited by a VMware vulnerability. This is a true testament to the ability to harden the hypervisor and keep ahead of the curve with security exploits.

Again, that does not justify not patching. With the ever-increasing deployments of vSphere, it seems pretty reasonable to think the focus will continue to shift toward attacking these consolidated infrastructures powered by VMware. After all, wouldn't it be easier to bring down 10 vSphere hosts running 200 servers than to try to bring them down individually?

Update Manager is a patch-management solution provided by VMware with all versions of vCenter Server. It helps to automate the deployment of patches and updates and provides a means to maintain compliancy among your entire infrastructure. Its capabilities are not just limited to vSphere ESXi hosts either, as you can now patch many virtual appliances as well as extensions such as the Cisco Nexus 1000V and EMC PowerPath.

Formerly, Update Manager was capable of remediating Windows guest virtual machines by providing operating system and application patches. As of vSphere 5, however, this capability is no longer included. Interestingly, they licensed the technology from Shavlik, which they recently acquired. VMware now offers several other products that offer comparable capabilities. vCenter Protect Essentials and VMware Go both offer abilities to patch and manage guest operating systems. vCenter Protect Essentials provides for an on-premise solution that will patch and manage virtual machines. In the case of VMware Go, it is a cloud-based solution that also offers capabilities for help desk end-user portals. Both products also provide asset and configuration management capabilities that are geared toward the small to medium business market.

A major selling point of utilizing Update Manager to patch your vSphere servers is that when set up and used properly, it requires zero downtime to any of your virtual machines. There is no need to worry about having to have downtime twice for virtual machines for both the vSphere and guest patching. Utilizing DRS in conjunction with Maintenance mode, an administrator can deploy patches to a host with zero downtime to any of the virtual machines in the entire infrastructure.

Update Manager also allows the scheduling of updates. Simply create or attach a baseline to a set of hosts and choose a date and time to run the updates. These baselines can be assigned at the vCenter level or at the datacenter, cluster, or host level. Another useful ability of Update Manager is to stage and schedule virtual machine hardware and tools upgrades.

Patching Hosts Using Update Manager

With a DRS-enabled cluster and the use of Maintenance mode, patching hosts using Update Manager is a straightforward process; however, the following key areas are often overlooked:

- **Sizing the patch repository**—The patch repository can become quite large depending on the versions of vSphere you choose to implement over time. As a result, it is best practice to configure a shared repository outside of the vCenter server or server where Update Manager is installed when separated. VMware offers the vSphere Update Manager Sizing Estimator for download, which will aid in sizing not only the shared repository, but also the database itself.

- **Notification of new patches**—You will have a hard time knowing when to install updates if you do not know when they come out. The easiest way to be notified of new patches is by configuring email notifications under the Download Schedule of Update Manager.

- **Failure to consider compatibility and support**—There is a lot to consider when choosing to install updates. If you are running a solution where the vendor will only support virtual machines on a certain revision of the software, then you should clarify how these support policies are affected by updates. This is a rarity these days as solutions such as Cisco's unified communications on top of UCS software are fully supported by all updates at the time of release.

- **Failure to disable HA during an update**—If you have a smaller cluster of hosts, you might run into a failure if you do not disable HA during a host update. By default, this is not set but can be if you are going to run into this issue. Without doing so, if your cluster cannot support HA and you attempt a remediation, it will fail.
- **Failure to properly configure DNS**—If DNS is not properly configured, you will spend a lot of time troubleshooting why Update Manager is not working. It is highly dependent on DNS to be configured properly on both the vCenter and vSphere hosts. Failing to do so causes Update Manager to fail during the remediation.

Upgrading Hosts

VMware periodically releases new versions of vSphere that require an upgrade to vSphere. If an environment is healthy and no issues exist, we recommend using Upgrade Manager to upgrade the hosts in place. If, however, there are issues with your environment, consider wiping away each host and starting fresh.

You should also consider downtime in your environment for the upgrade. If your virtual machines are on shared storage backed by hosts that can vMotion among one another, you will be able to have much less downtime than an environment with virtual machines on local storage, for example. Different circumstances warrant different paths, so let's talk about some of the key items to consider when planning your upgrade.

Planning for vSphere Upgrades

Planning for vSphere upgrades requires investigating your environment from top to bottom to ensure you are presently free of any issues and have the appropriate pieces to perform a successful upgrade of your environment. In fact, before an upgrade is the perfect time for a health check to be performed by a VMware authorized partner. Although this section describes the important steps to consider, you might also want to take a look at the *vSphere Upgrade Guide* provided by VMware.

Upgrade Entitlement

Before you get too far along, you need to ensure you are eligible to upgrade. Upgrades are not at an additional cost when you have a valid support contract with VMware for your purchased licenses. If you have an eligible support contract, you can find both the software and licenses available in your VMware software and licensing portals. If you don't have an eligible support contract, you need to either renew your support or purchase additional licenses. In addition to support for vSphere, this is also a good time to ensure your hardware is still supported by your vendor before proceeding with an upgrade.

Feature Changes

Another consideration is changes that might have occurred between the old and new versions of vSphere. An example of this is an organization using Update Manager as part of vSphere 4 to patch its Windows guest. This functionality is no longer included as of vSphere 5; however, it can be acquired as part of vCenter Protect Essentials or Essentials Plus.

Hardware Compatibility

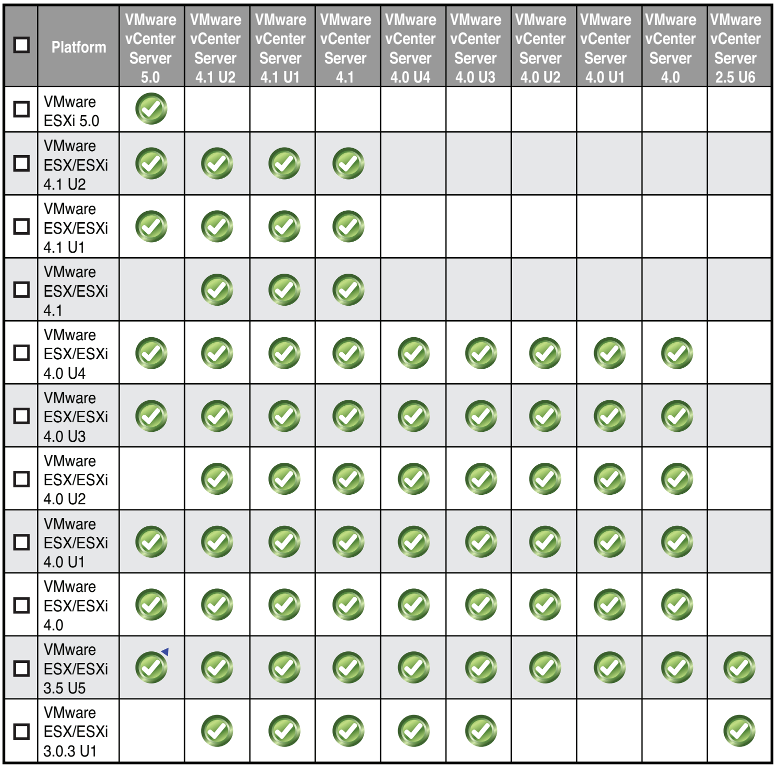
If your hardware is older, there is a chance that the hosts, storage, or IO devices might not be supported with the new release. Regardless of how new your equipment is, you should reference the *VMware Compatibility Guide* online to ensure the hardware will be supported after an upgrade. Although some people might not be concerned with hardware being fully supported, they should be advised that if it is not supported, there is a chance it will not work in some fashion. Be sure to check compatibility for your specific host. You will find that there may be several versions and revisions for popular brand models. Additionally, be sure to check your I/O devices and storage as well.

Database Compatibility

You need to be sure your database is supported with the new version of vCenter to which you will be upgrading. Many versions of SQL 2000 and 2005 are no longer supported despite their use, and you should consider upgrading the database servers if yours are not supported. Check out the VMware Product Interoperability Matrix online to verify your database software support before proceeding with any vCenter upgrades.

vCenter Support

You also need to ensure your vCenter server will support any older versions for hosts that are going to run for any period of time on an older version as part of that vCenter server. You can again check the VMware Product Interoperability Matrix online to verify this information. Pay close attention to the matrix, as shown in Figure 3.4. At the time of this writing, there is a known issue with VMware 4.0 U2 that does not allow it to be managed by a VMware vCenter server. This is a good example where a lot of people continue to make assumptions of the support to find out later that it does not work correctly.

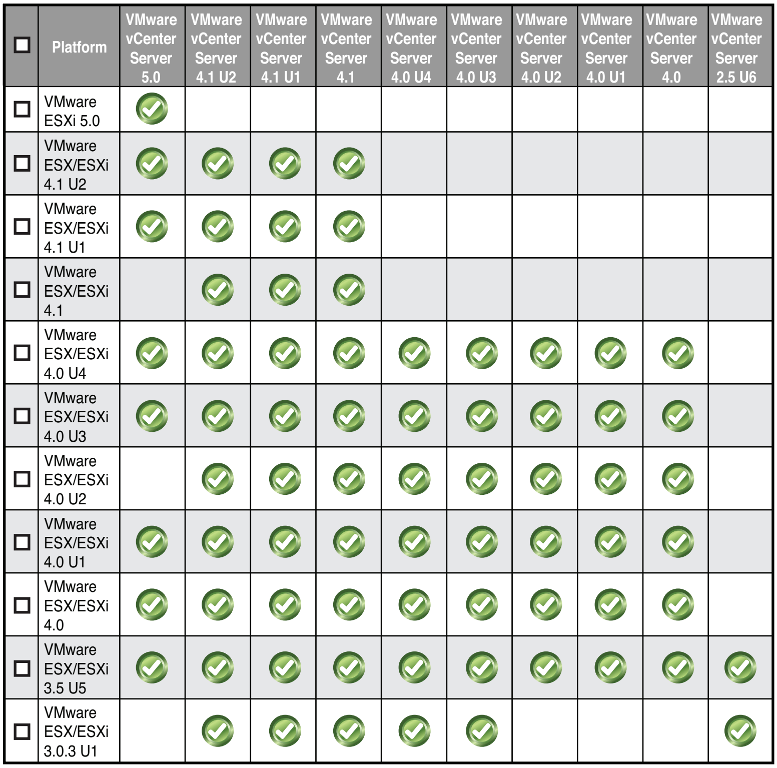


□	Platform	VMware vCenter Server 5.0	VMware vCenter Server 4.1 U2	VMware vCenter Server 4.1 U1	VMware vCenter Server 4.1	VMware vCenter Server 4.0 U4	VMware vCenter Server 4.0 U3	VMware vCenter Server 4.0 U2	VMware vCenter Server 4.0 U1	VMware vCenter Server 4.0	VMware vCenter Server 2.5 U6
□	VMware ESXi 5.0	✓									
□	VMware ESX/ESXi 4.1 U2	✓	✓	✓	✓						
□	VMware ESX/ESXi 4.1 U1	✓	✓	✓	✓						
□	VMware ESX/ESXi 4.1		✓	✓	✓						
□	VMware ESX/ESXi 4.0 U4	✓	✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 4.0 U3	✓	✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 4.0 U2		✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 4.0 U1	✓	✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 4.0	✓	✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 3.5 U5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
□	VMware ESX/ESXi 3.0.3 U1		✓	✓	✓	✓	✓				✓

Figure 3.4 vCenter Compatibility and Support Matrix

Additionally, you need to make sure vCenter is installed on a 64-bit operating system. If your existing vCenter server is older, you might not be able to directly upgrade anyway; however, if it is installed on a 32-bit operating system, you definitely need to install a fresh operating system.

In addition to support, you must also see if an upgrade is possible. As shown in Figure 3.5, you can see that, in general, there is direct upgrade available from 4.0 U1 up to vCenter 5.0 with the exception of 4.0 U4. Both 4.0 and 4.0 U4, and even 2.4 U6, however, can be upgraded to 4.1 U2. Once at 4.1 U2, they can be updated to vCenter 5.0 directly. Always check the VMware Product Interoperability Matrixes and *vSphere Upgrade Guide* for the most up-to-date support information. Again, remember there is a 64-bit requirement, so if you don't have a 64-bit server, you need to install a new version of Windows to support your new vCenter Server installation.



□	Platform	VMware vCenter Server 5.0	VMware vCenter Server 4.1 U2	VMware vCenter Server 4.1 U1	VMware vCenter Server 4.1	VMware vCenter Server 4.0 U4	VMware vCenter Server 4.0 U3	VMware vCenter Server 4.0 U2	VMware vCenter Server 4.0 U1	VMware vCenter Server 4.0	VMware vCenter Server 2.5 U6
□	VMware ESXi 5.0	✓									
□	VMware ESX/ESXi 4.1 U2	✓	✓	✓	✓						
□	VMware ESX/ESXi 4.1 U1	✓	✓	✓	✓						
□	VMware ESX/ESXi 4.1		✓	✓	✓						
□	VMware ESX/ESXi 4.0 U4	✓	✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 4.0 U3	✓	✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 4.0 U2		✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 4.0 U1	✓	✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 4.0	✓	✓	✓	✓	✓	✓	✓	✓	✓	
□	VMware ESX/ESXi 3.5 U5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
□	VMware ESX/ESXi 3.0.3 U1		✓	✓	✓	✓	✓				✓

Figure 3.5 vCenter Upgrade Compatibility and Support Matrix

Dependencies

Outside of the core functionality in the vCenter server and the vSphere hosts, there exist some other pieces that need consideration as well. These are just some examples and you need to also consider any additional software or plug-ins that are used in your environment. Make sure to consider these pieces by verifying support by the vendor or within the *vSphere Upgrade Guide*:

- vCenter Update Manager
- vCenter License Server
- VMware View
- VMware Data Recovery
- Site Recovery Manager
- Third-party plug-ins like PowerPath

- Use of Nexus 1000V
- Any PowerShell or other scripting used for troubleshooting and reporting

Upgrade Paths

vSphere 5 is the first version of vSphere that has been released in only the ESXi flavor, so there is only one destination when upgrading to vSphere 5. You must also consider the source of the server and whether you have the option to upgrade.

The following is true about upgrading older versions of ESX and ESXi to vSphere 5.0. Note there are conditions where these items might not apply, so be sure to check the VMware Product Interoperability Matrixes and *vSphere Upgrade Guide* for the most up-to-date support information.

ESX & ESXi 3.5

- No direct upgrade available
- Upgrade to 4.x first
- Note that the partition layout might be incompatible with vSphere 5, so this can prohibit such an upgrade to 5.0

ESX & ESXi 4.0

- Direct upgrade available with Update Manager, interactively, or scripted
- Might not be compatible with all environments
- For example, ESX 4 hosts on SAN/SSD might not have optimal partitions and disks with multiple VMFS partitions cannot be upgraded
- Additionally, note that a host with any third-party vSphere Installation Bundles (VIB) may require using the ESXi Image Builder CLI to create a customized ESXi install ISO

And one last note on upgrading hosts. As of vSphere 5, the advanced version no longer exists and any customers with active support agreements for vSphere 4 Advanced are entitled to vSphere Enterprise.

Order of Operations

When laying out your plan for an upgrade, you must consider the order in which you are going to do so. Outside of the vCenter and vSphere hosts themselves, you need to make sure you upgrade to supported code and firmware for your storage and other devices ahead of time. Additionally, be sure you have proper backups of the necessary components. For vCenter, you need at minimum a backup copy of the database as well as Secure Socket Layer (SSL) certificates from the server. For the hosts themselves, you need to have good documentation on their configuration as well as a backup copy of all virtual machines. This holds especially true if you are upgrading a host with virtual machines running on local storage. For virtual machines on shared storage, you need to ensure backups exist as you will be upgrading our virtual hardware and VMware Tools later on.

In general, follow these steps to perform an upgrade:

1. Run the vCenter Host Agent Pre-Upgrade Checker. This can be found on the vCenter installation media and is a great verification tool to ensure the likelihood of a successful upgrade.
2. Upgrade or install a new vCenter server.
3. Upgrade or install a new Update Manager.
4. Upgrade or install other plug-ins and third-party packages.
5. Upgrade or install vSphere on hosts.
6. Upgrade VMFS.
7. Upgrade virtual machine tools and hardware.

Methods for Upgrading vSphere

As discussed previously, to perform your vCenter upgrade, you can either upgrade the software in place if supported or install a fresh vCenter server. You can then choose to either start completely fresh, redefine roles and other vCenter configurations, or import the database and continue from there.

For vSphere hosts, you not only have the option of upgrading or starting fresh, but you also have several methods to perform the upgrade. When possible, we recommend building new hosts and bringing configurations over.

In previous versions of vSphere, the Host Update Utility was included on the vCenter installation media for performing host upgrades on a host-by-host basis. Note that this is no longer the case and you must upgrade your hosts by either using vSphere 5 media or through Update Manager.

Manual Upgrade

You may manually perform an upgrade to a host using the ESXi installation media by performing an interactive or scripted upgrade. It is recommended you disconnect all storage from the host as this greatly reduces the amount of time required for the upgrade.

When upgrading a host, you have three options:

- Upgrade ESXi, Preserve VMFS Datastore or Force Migrate ESXi, Preserve VMFS Datastore
- Install ESXi, Preserve VMFS Datastore
- Install ESXi, Overwrite VMFS Datastore

The first option will vary if any custom VIBs are not included with the vSphere 5 media. If that is the case, Force Migrate ESXi replaces Upgrade ESXi. Make sure to back up any items on the local VMFS datastore beforehand and especially when choosing to overwrite the VMFS datastore.

In addition to performing an interactive upgrade, you may also choose to perform a scripted installation. For full details on creating a scripted installation, including adding custom drivers and third-party VIBs, check out the *vSphere Upgrade Guide*.

Update Manager

When using Update Manager to upgrade hosts, an orchestrated host upgrade can occur that allows not only for vSphere host installation, but the installation of VMware Tools and the upgrade of virtual hardware.

Update Manager does have some limitations that you may encounter. Recall from the earlier discussion of upgrade paths that there are some limitations even when following a supported path. Update Manager cannot be used to upgrade an ESX 4.x host if it was previously upgraded from 3.x as a result of insufficient space in the /boot partition. This problem is not unique as it is possible an ESX 4.x host may also not have the proper amount of space.

If you are not installing a fresh version of vSphere, it is recommended to use Update Manager because it greatly eases the upgrade process. The use of Update Manager does a better job of preventing erroneous actions and disallows things such as upgrading the virtual machine hardware before installing VMware Tools.

Host Upgrades Upgrading a host requires the creation of a host upgrade baseline. Additionally, you are required to import the ESXi image to be used for upgrades.

You may choose to have separate baselines and separate images in the repository. For example, you may have different images based on hardware for the hosts, which may be of different vendor types and contain different third-party VIBs.

You cannot roll back to the previous version of ESX/ESXi when upgrading with Update Manager, so, as always, make sure you have the configuration of your host documented and the proper backups of all virtual machines in place before proceeding with any upgrades.

Virtual Machine Upgrades Upgrading virtual machines after an upgrade requires using an existing baseline or the creation of a baseline group. You cannot upgrade VM hardware until the virtual machine is running the latest version of VMware Tools. Update Manager makes sure this happens to avoid these operations not occurring in sequence.

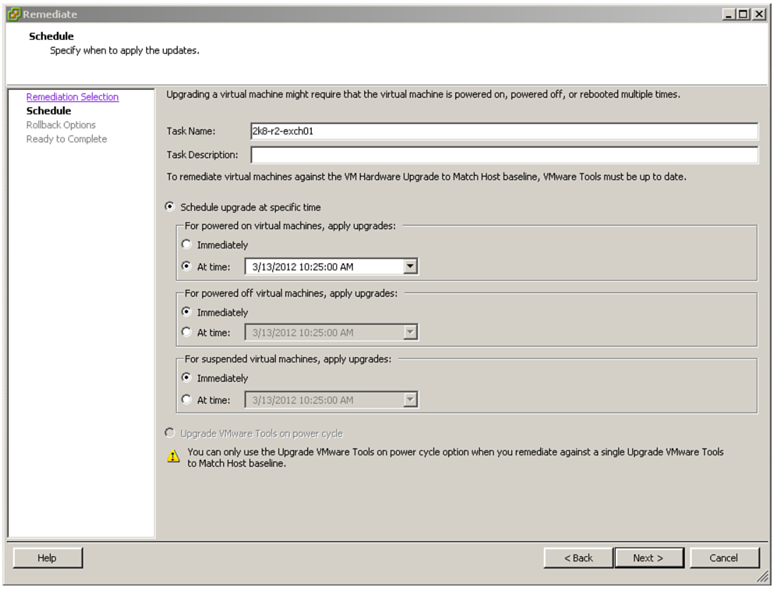
By default, the following two baselines are created:

- VMware Tools Upgrade to Match Host
- VM Hardware Upgrade to Match Host

When scheduling the update, you can granularly schedule separate virtual machines depending on the following power states. For example, you might want to schedule any powered-on machine later because they will require downtime, as shown in Figure 3.6. Your options for scheduling virtual machine updates include the following:

- Powered On
- Powered Off
- Suspended

An orchestrated upgrade of virtual machines is not required but greatly reduces the time it takes to remediate a large number of virtual machines at the same time. If you would rather manually remediate the virtual machines, simply upgrade VMware Tools on each virtual machine and then power off the virtual machine to perform the virtual hardware upgrade.



Schedule
Specify when to apply the updates.

Upgrading a virtual machine might require that the virtual machine is powered on, powered off, or rebooted multiple times.

Task Name: 2k8-r2-exch01
Task Description:

To remediate virtual machines against the VM Hardware Upgrade to Match Host baseline, VMware Tools must be up to date.

● Schedule upgrade at specific time

For powered on virtual machines, apply upgrades:
○ Immediately
● At time: 3/13/2012 10:25:00 AM

For powered off virtual machines, apply upgrades:
● Immediately
○ At time: 3/13/2012 10:25:00 AM

For suspended virtual machines, apply upgrades:
● Immediately
○ At time: 3/13/2012 10:25:00 AM

○ Upgrade VMware Tools on power cycle

⚠ You can only use the Upgrade VMware Tools on power cycle option when you remediate against a single Upgrade VMware Tools to Match Host baseline.

Figure 3.6 Scheduling Update Manager Updates

Monitoring

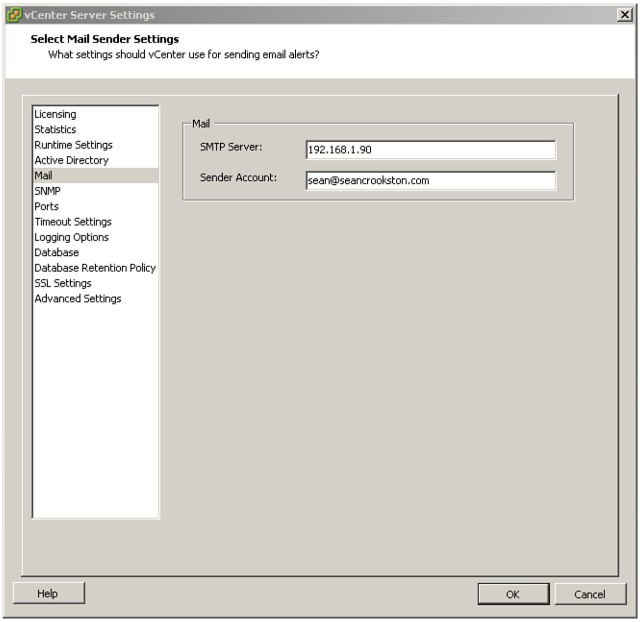
Like maintenance, monitoring is sometimes forgotten with a virtual infrastructure. Many organizations continue their monitoring of their guest virtual machines without a consideration for the hosts themselves. Others consider the hosts but don't have the proper monitoring software, licensing, or understanding of how or what to monitor in the virtual infrastructure. Regardless of the reason, the need to monitor the underlying components of a virtual infrastructure remains high.

Alerting

I once had a customer contact me who did not understand why he didn't receive an email notification that one of his storage paths had lost redundancy. He had logged in to his vCenter server and noticed the down host, which had been offline for two days. Although this showed off how well the cluster handled the failure of the host, it was a major point of concern for him because he didn't know the host had failed. In this case, the customer had not fully configured the alarms in vCenter. This section discusses the process required to set up alarms as well as some common issues encountered.

For starters, you need to configure the mail setting in vCenter Server.

To do this, go to Administration, vCenter Server Settings from the vSphere Client. Next, configure the SMTP server and appropriate sender account, as shown in Figure 3.7.



Select Mail Sender Settings

What settings should vCenter use for sending email alerts?

SMTP Server: 192.168.1.90

Sender Account: sean@seancrookston.com

Figure 3.7 Configuring vCenter Email Settings

You need to configure both an SMTP and a sending account. Additionally, you need to ensure your SMTP server can accept relayed messages from your vCenter server.

This is a step that nearly everyone configures during the default install. A common problem, though, is this is where many people stop. By default, vCenter 5 has 54 alarms defined; however, to set up any type of SNMP or email alerting, actions must be individually defined for each alarm.

Defining Actions for Alarms

For most alarms, only three actions can be defined. You may define an action once or multiple times for each alarm, and you may define multiple types of actions for a single alarm. The actions that are available to be configured are as follows.

- Send a Notification Email
- Send a Notification Trap
- Run a Command

Two monitor types, however, have the capability of performing specific actions. The Alarm Type Monitor for Virtual Machines may take the following actions in addition to sending an email, sending an SNMP trap, or running a command:

- Enter Maintenance Mode
- Exit Maintenance Mode
- Enter Standby
- Exit Standby
- Reboot Host
- Shutdown Host

The Alarm Type Monitor for Hosts may take the following actions in addition to the three actions mentioned—sending an email, sending an SNMP trap, or running a command:

- Power On VM
- Power Off VM
- Suspend VM
- Reset VM
- Migrate VM
- Reboot Guest On VM
- Shutdown Guest On VM

For the following Alarm Type Monitors, the only three actions are to send a notification email, send a notification trap, or run a command:

- Clusters
- Datacenters
- Datastores
- vSphere Distributed Switches
- Distributed Port Groups

- Datastore Clusters
- vCenter Server

The process for defining actions for alarms is pretty straightforward; however, there are a few things to be aware of.

First, as mentioned, 54 alarms are defined by default. Defining all 54 alarms individually would take a long time and would likely result in a few of them being configured incorrectly due to an occasional keystroke error. Don't worry, though, because PowerShell can be used to automate the creation of these actions and is discussed shortly.

Second, when you are defining actions, you must define when the action will occur and how often notification will occur for issues that persist. By default, you receive an email notification only when going from a yellow to a red state. There are four configurable options to consider:

- Green→Yellow
- Yellow→Red
- Red→Yellow
- Yellow→Green

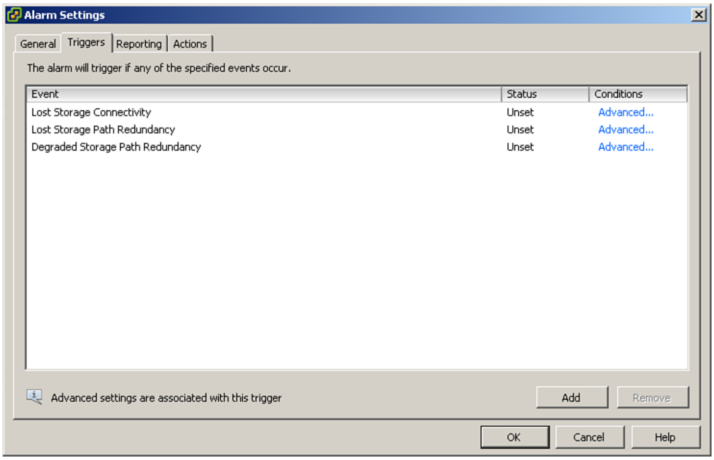
Let's stop for a moment to talk about which of these four you will want to be notified of. If you are relying on SNMP traps being sent to your existing monitoring software, you may choose to have very little to no email notifications. Many smaller environments do not rely on SNMP notifications or still may require email notifications outside of their existing monitoring solutions. For environments with no other monitoring, it is best to configure all of the default alarms and some additional ones as well. These additional recommendations as well as automating the process are discussed in just a bit.

So you now have defined actions for all of your desired alarms as well as the severity changes you would like to be notified of and the amount of times you would like to be notified if the issue persists. That brings us to another common thing to consider for a new implementation.

We have witnessed some environments that simply forgot to allow the vCenter server to use the mail server as a relay. After all, the vCenter server may be a new addition to an environment and would not have been previously configured to relay email messages from the SMTP server. If you are unsure if the mail server is allowing relay for the host and do not have access to the email server to check, you may try the following:

```
telnet mailservername.vmware.com 25
helo vmware.com
```

There is still one more thing to be aware of. Even after all of this, you might find you are not being notified of some issues, for example when storage path redundancy is lost. This is because some triggers are left unset by default, as shown in Figure 3.8. When set to Unset, alarms do not show in vCenter; however, they are sent to email or as SNMP traps if configured. As you can see for the case of lost storage path redundancy, the status for each event is not set.



The alarm will trigger if any of the specified events occur.

Event	Status	Conditions
Lost Storage Connectivity	Unset	Advanced...
Lost Storage Path Redundancy	Unset	Advanced...
Degraded Storage Path Redundancy	Unset	Advanced...

Advanced settings are associated with this trigger

Figure 3.8 Unset vCenter Alarms

The following is a list of the other default alarms that are not set up:

- Unmanaged workload detected on Storage I/O Control (SIOC)-enabled datastore (this is disabled by default)
- VMkernel NIC not configured correctly
- Network uplink redundancy degraded
- Health Status Changed Alarm
- License Error
- Exit Standby Error
- Migration Error
- Host Connection Failure
- Virtual Machine Error
- Host Error

- No Compatible Host for Secondary VM
- Timed Out Starting Secondary VM

Two of the default alarms also are not configurable. These alarms are triggered via the vSphere API and can only be modified as such:

- Datastore Capability Alarm
- Thin-Provisioned LUN Capacity Exceeded

When creating actions, you just need to select an SNMP action in addition to or instead of an email notification so that a trap is sent. You may also enable SNMP traps for each individual host if desired. This may be beneficial in the event of a vCenter server outage as the individual hosts themselves will not communicate any status back otherwise.

Considerations for Tweaking Default Alarms

Some of the default alarms may have some notification options that are less than desirable for your environment. For example, you may have an environment that is strictly testing for internal IT staff. You may decide you still want all the alarms but fully accept that the vSphere hosts in question will likely be pegged pretty hard in terms of memory at certain times of the day. After all, this may be older hardware with lesser memory. You still, however, want to know if there is a consistent condition where memory is steady at 95% or greater for 30 minutes or more.

In this case, by default, the Host Memory Usage alarm warning triggers a warning when host memory usage is above 90% for 5 minutes. Also by default, an alert triggers when host memory usage is above 95% for 5 minutes. By setting both values to 5% higher and to lengths of 30 minutes, you do not get repeated alerts for expected high memory conditions, but do get notified when the issue becomes persistent enough where it may warrant finding additional memory for these hosts.

In closing, you can see that there is a lot to consider even when looking specifically at just vCenter alarms. Walking away from this discussion on alarms, remember the following key points:

- Consider that the alarms can be defined at many levels. Depending on your infrastructure, you might want to define alarms at the vCenter, datacenter, cluster, or individual host level. For that matter, you may also want to get even more granular and enable alarms on specific virtual machines, datastores, datastore clusters, and virtual distributed switches.

- Consider that triggers may have multiple actions that trigger based on both actions happening or one or the other.
- Consider how often you want to be notified and of what state changes you would like to be notified. Too many alerts can become just as big of a problem as not enough alerts at times if you begin tuning them out.

Before moving on to the next section, some assistance in setting up these alarms using PowerShell was promised. With just a few modifications, the provided PowerShell script allows you to easily set up all or as many of the default alarms as you would like. Note that you need to configure the alarms mentioned that are not configured by default to your liking for your environment. Although this still leaves some manual configuration, you no longer have to enter an email address for any of the alarms. It is our recommendation that you start by configuring all vCenter alarms and remove alarms that are not necessary for your environment.

You can download this script from http://www.seancrookston.com/set_alarms.ps1 (see Appendix A for a link).

Verifying Configurations

Another important component of operating a vSphere infrastructure is configuration management. When talking about configuration management, the concern is with ensuring configurations are not unknowingly changed or drift from their intended configurations. You want configurations to match their intended configuration and be consistent across the environment. For example, you want your hosts to be running on a certain build of vSphere and to be consistent with the other hosts within the same cluster.

vCenter Operations Enterprise versions include vCenter Configuration Manager, which provides the ability to monitor configuration virtual infrastructure configurations. vCenter Operations are discussed in further detail in Chapter 4, “Managing the Environment.”

Even with a product like vCenter Operations, you still need to implement the most important part of a solid configuration management strategy. Policies and procedures for documenting configuration are the foundation to maintaining an environment with consistent and desired configurations.

When thinking about the configurations, the goal is to maintain many items that might not seem obvious initially. The following is a list of some of the pieces in your virtual infrastructure that might have configurations—in terms of software or firmware—to track and ensure are desired and consistent. Keep in mind this list is brief and we could easily drill even deeper.

- vCenter Server configuration
    - Cluster configuration
    - High Availability configuration
    - DRS configuration
    - Update Manager configuration
- vSphere host configuration
    - vSphere drivers and operating software
    - HBA drivers and firmware
    - NIC drivers and firmware
- IP network configuration
- Storage network configuration
- Storage firmware and software versions
- Virtual guest configuration

Host Profiles

If a product like Operations Manager is not a fit, you may also use Host Profiles. This feature is included with the Enterprise Plus level of vCenter licensing and is of great assistance with managing the delivery of consistent configurations. Additionally, you may use Scheduled Tasks in vCenter to define a scheduled compliance check that will notify you daily of any configuration drift.

After you've configured your first host to the desired gold state, you can simply create a profile using this host as a reference host. Then you can take your baseline profile and apply it to other hosts or clusters. You will be prompted to enter dynamic information, such as network information during the application, but other configuration settings will be applied consistently to your hosts.

At any time, you can check the host's compliance against the profile or receive notification via email when a drift in configuration occurs. When the time comes to make a change to your standard configuration, the process is just as easy. Simply update your reference host's configuration and then update the profile and reapply the configuration to your other hosts.

Even if you do not have Enterprise Plus licensing, you should consider using Host Profiles during your setup as part of your 60-day evaluation licensing.

Health Check

So far, this chapter has discussed ways to operationally maintain the environment through updates and alerts. Another important operational step is to perform regular health checks of your environment. This may consist of a physical inspection as well as checking configurations. You may also be ensuring your configured alarms are configured as expected and manually checking for issues just in case. You may also be looking for drifts in configuration based on your organization's standardized configuration.

These are all important things to do and there are many community resources that can assist in these efforts. One such resource is a daily health check script developed by Alan Renouf called vCheck, detailed further in Appendix A.

This script creates a daily report that gives a great report of the environment, including items such as snapshots and new virtual machines that have been created. The setup process has been made easy with an install script, and a great demo video is included on the site for guidance in setting the script up.

Continuing the discussion of performing health checks, another reason to do a health check might be to get a new perspective on the current state of the environment. You might think to yourself, "Well, nothing has changed in this environment in the last three months." Considering that perhaps nothing has changed in the environment, you also need to consider what has changed externally to your environment. This doesn't strictly refer to the storage or networking attached to your vSphere hosts, although checking on these is equally important. Technology is often updated or at times has vulnerabilities due to security flaws in the product.

Bugs, workarounds, patches, and best practices are regularly released and updated. Many individuals barely have the time to perform their regular day-to-day duties, and this information can be difficult to find at times. This is where the aid of someone focused on vSphere technologies is of great advantage.

VMware's Health Check Delivery

VMware offers a Health Check service that can greatly aid in this need. Any of the information that is used during this process is available to anyone and you could use scripts like the ones mentioned to verify much of the same information. The time to do so could be substantial and unless you have significant experience across many environments, there may be the risk that you are missing something. The health check delivery has many big advantages, such as the following:

- Consultants will add in their experiences recently as well as perform additional checks.
- Consultants will have at minimum a VCP.
- Quick collection of data for analysis will be performed by an expert.

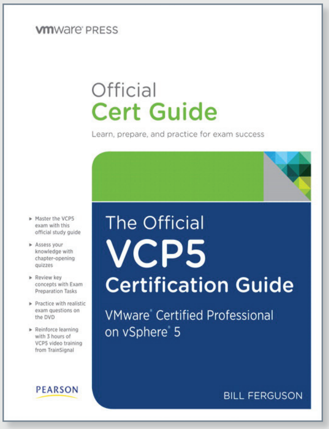
The result of the engagement is a report and in-depth analysis of the environment with suggestions and remediation. The suggestions are based on best-practice configuration and known issues across a wide range of industries and environments.

Operating the Environment Summary

This chapter focused on the operating phase of a virtualization infrastructure. You learned about many tools and methods to operate and monitor on a daily basis in addition to best practices and methods for continuing to bring existing physical workloads into your virtual infrastructure.

This chapter also discussed methods for monitoring and alerting of issues in the virtual infrastructure. From an operational perspective, you have covered the grounds of day-to-day management of your virtual infrastructure.

Moving forward, you need to monitor your environment's performance and capacity for growth. This is discussed in the following chapter.



vmware PRESS

Official
Cert Guide

Learn, prepare, and practice for exam success

- Master the VCP5 exam with this official study guide
- Assess your knowledge with chapter-opening quizzes
- Review key concepts with Exam Preparation Tasks
- Practice with realistic exam questions on the DVD
- Reinforce learning with 3 hours of VCP5 video training from TrainSignal

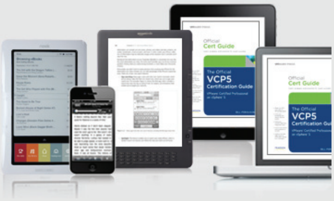
The Official
VCP5
Certification Guide

VMware® Certified Professional on vSphere® 5

PEARSON

BILL FERGUSON

CHAPTER 2 Planning and Configuring vSphere Networking



AUGUST 2012

Available in Print and eBook formats and through SAFARI BOOKS ONLINE

SHARE WITH OTHERS

vmware® PRESS

BUY NOW

The Official VCP5 Certification Guide

BY BILL FERGUSON

Table of Contents

ISBN: 9780789749314

vmware.com/go/vmwarepress

ALWAYS LEARNING

PEARSON

CHAPTER 2

Planning and Configuring vSphere Networking

In our discussion on vSphere networking, I will address many topics such as vSphere standard switches (vSS), vSphere distributed switches (vDS), port groups, and the properties for all of these. It's easy to get overwhelmed in all the terminology, especially when most of the components are not something that you can see or hold in your hand. To keep from becoming overwhelmed with the technology, I want you to focus on two primary questions. The first question is, "What type of connections can I create and what do they do?" The second is, "Where does the 'virtual world' meet the 'physical world,' and how is that point of reference defined?" If you just focus on these two questions, I believe that the rest of the picture will come to your mind.

That said, this section covers configuring vSSs, configuring vDSs, and configuring vSS and vDS policies. In each section, I will explain why these should be configured, and then I will discuss how you can configure them. In addition, I will walk you through the steps to configure each of these settings.

"Do I Know This Already?" Quiz

The "Do I Know This Already?" quiz allows you to assess whether you should read this entire chapter or simply jump to the "Exam Preparation Tasks" section for review. If you are in doubt, read the entire chapter. Table 2-1 outlines the major headings in this chapter and the corresponding "Do I Know This Already?" quiz questions. You can find the answers in Appendix A, "Answers to the 'Do I Know This Already?' Quizzes and Chapter Review Questions."

Table 2-1 "Do I Know This Already?" Section-to-Question Mapping

Foundations Topics Section	Questions Covered in This Section
Configuring vSphere Standard Switches	1–3
Configuring vSphere Distributed Switches	4–6
Configuring vSS and vDS Policies	7–10

1. Which of following will result if you choose **Add Networking** on the Networking link of your ESXi host?
   a. You can add a new VMkernel port to an existing switch.
   b. You can add a new VM port to an existing switch.
   c. You will be creating a new vSS.
   d. You can add a new vmnic to an existing switch.

2. Which of the following is *not* a common use of a VMkernel port?
   a. IP storage
   b. Storage vMotion
   c. vMotion
   d. Management

3. Which of the following is true about switch and port group policies on a vSS?
   a. Switch settings override port group settings.
   b. You cannot configure port group settings different from switch settings.
   c. There are no switch settings on a vSS.
   d. Port group settings override switch settings for the VMs on the port group.

4. What is the maximum number of hosts that can be connected to a single vDS?
   a. 32
   b. 1000
   c. 350
   d. 100

5. Which of the following is the minimum license requirement to create a vDS?
   a. Enterprise Plus
   b. Enterprise
   c. Advanced
   d. Essentials

6. Which view should you be in to add a host to an existing vDS?
   - a. Hosts and Clusters
   - b. Networking
   - c. vSphere
   - d. VMs and Templates

7. Which of the following is *not* a common policy for vSS switch and port groups?
   - a. Traffic shaping
   - b. NIC teaming
   - c. Permissions
   - d. Security

8. Which of the following is true about vDS policies?
   - a. Policies set at the port group level override those are the port level.
   - b. Policies cannot be set at the port level.
   - c. Policies are always set at the port level.
   - d. Policies set at the port level override policies set at the port group level.

9. Which of the following is *not* a load-balancing option in vSphere?
   - a. Route based on the originating virtual port ID
   - b. Beacon probing
   - c. Route based on source MAC hash
   - d. Route based on IP hash

10. Which of the following is *not* a type of private VLAN?
   - a. Isolated
   - b. Trunking
   - c. Promiscuous
   - d. Community

Foundation Topics

Configuring vSphere Standard Switches

A vSphere standard switch (vSS) is a logical construct within one ESXi host that connects virtual machines (VMs) to other VMs on the same switch. In addition, using connections called uplinks, it can connect VMs to other virtual or physical machines on other ESX/ESXi hosts, other vSSs in the same host, or anywhere in the physical environment. In this section I will discuss vSS capabilities and how to create and delete them. In addition, I will cover adding, configuring, and removing vmnics; configuring VMkernel ports and services; adding and removing port groups; and determining use cases for a vSS.

Identifying vSphere Standard Switch (vSS) Capabilities

A vSS models a simple Layer 2 switch that provides networking for the VMs connected to it. It can direct traffic between VMs on the switch as well as link them to external networks. Figure 2-1 shows a diagram of a vSS. I'm sorry that I don't have a photograph, but remember that they only exist in a software state. Note that there are actually two VMkernel ports on the vSS in this ESXi host. One is for management (management network), and the other is for other purposes that I will describe later in this section).

Key Topic

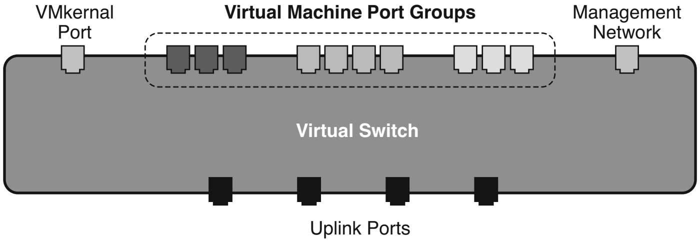


Figure 2-1 A Diagram of a vSphere Standard Switch

As I mentioned earlier, a vSS models an Ethernet Layer 2 switch on which a virtual machine network interface card (vNIC) can connect to its port and thereby be connected to other machines on the same switch; or off of the switch by way of an uplink to the physical world. Each uplink adapter also uses a port on a vSS. As I said before, one of the main questions to ask yourself is, "What type of connections can I create?" So, now I will discuss connections on vSSs.

You can create two main types of connections that you can create on vSSs; VMkernel ports and VM ports. The difference between these two types of connections is dramatic. It is important to understand how each type of connection is used.

VMkernel ports are used to connect the VMkernel to services that it controls. There is only one VMkernel on an ESXi host (also called the hypervisor), but there can be many VMkernel ports. In fact, it is best practice to use a separate VMkernel port for each type of VMkernel service. There are four main types of VMkernel services that require the use of a VMkernel port, as follows:

Key Topic

- **IP storage:** iSCSI or networked-attached storage (NAS). (Chapter 3, "Planning and Configuring vSphere Storage," covers these in more detail.)
- **vMotion:** A VMkernel port is required and a separate network is highly recommended. (Chapter 5, "Establishing and Maintaining Service Levels," covers vMotion in more detail.)
- **Management:** Because ESXi does not have a service console, or service console ports, management is performed through a specially configured VMkernel port.
- **Fault-tolerant logging:** A feature in vSphere that allows a high degree of hardware fault tolerance for the VMs involved, but also requires a separate and distinct VMkernel port. (Chapter 5 covers fault-tolerant logging in greater detail.)

VM port groups, however, are only used to connect VMs to the virtual switches. They are primarily a Layer 2 connection that does not require any configuration other than a label to identify a port group, such as Production. A VLAN can be configured for a port group, but that is optional as well. You can have multiple VM port groups on a single switch and use them to establish different polices, such as security, traffic shaping, and NIC teaming for various types of VMs. You will learn more about these in the section "Configuring vSS and vDS Policies," later in this chapter.

Creating / Deleting a vSphere Standard Switch

The first question that you might want to ask yourself is, "Do I really need a new vSS?" The answer to this question might not be as straightforward as you think. You do not necessarily need a new vSS for every new port or group of ports, because you can also just add components to the vSS that you already have. In fact, you might make better use of your resources by adding to a vSS that you already have, instead of creating a new one. Later in this chapter, in the section "Adding/Editing/Removing Port Groups on a vNetwork Standard Switch," I will discuss the power of using port groups and policies. In this section, I will discuss how to create a new vSS and how to delete a vSS that you no longer require.

If you decide to create a new vSS, you should select **Add Networking** from the Networking link and follow the wizard from there. The main thing to remember is that when you select Add Networking you are always creating a new vSS, not just adding networking components to an existing vSS. For example, if you want to create a new vSS for a VMkernel port used for vMotion, follow the steps outlined in Activity 2-1.

Key Topic

Activity 2-1 Creating a New vSphere Standard Switch

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the ESX host on which you want to create the new vSS and then open the Configuration tab.
4. Click the **Networking** link under Hardware.
5. In the upper-right corner, click the **Add Networking** link, as shown in Figure 2-2.

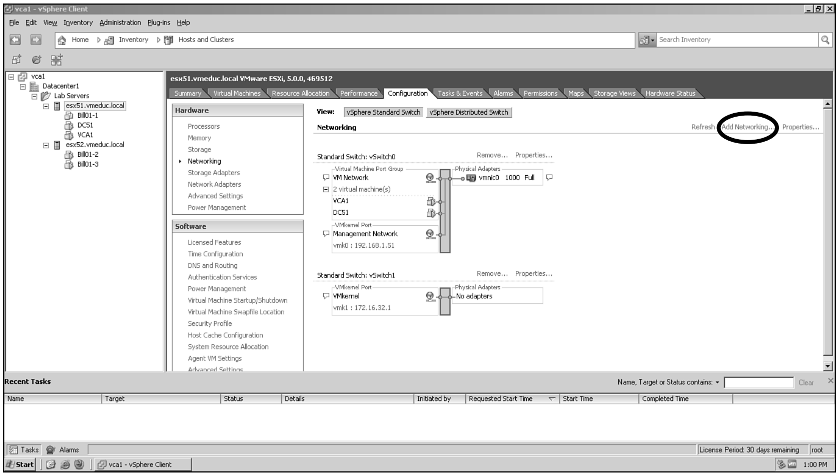
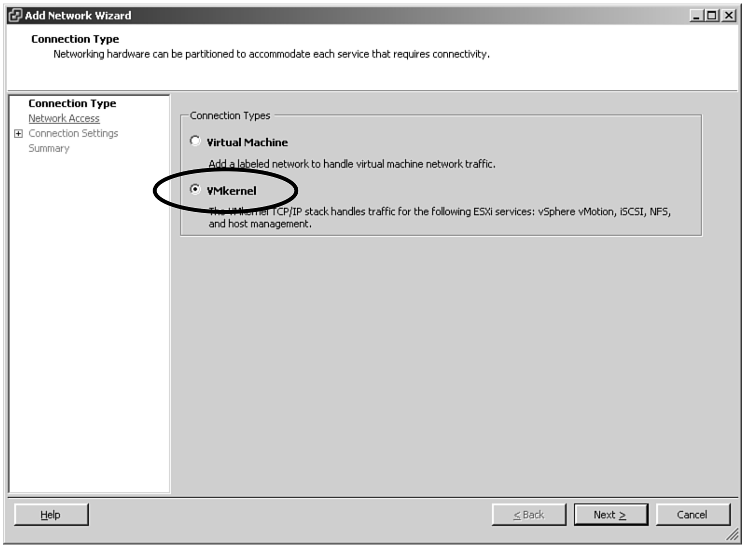


Figure 2-2 The Add Networking Link on a vSS

6. On the Connection Type of the Add Network Wizard, select **VMkernel** and click **Next**, as shown in Figure 2-3.



Connection Type

Networking hardware can be partitioned to accommodate each service that requires connectivity.

Connection Types

○ Virtual Machine

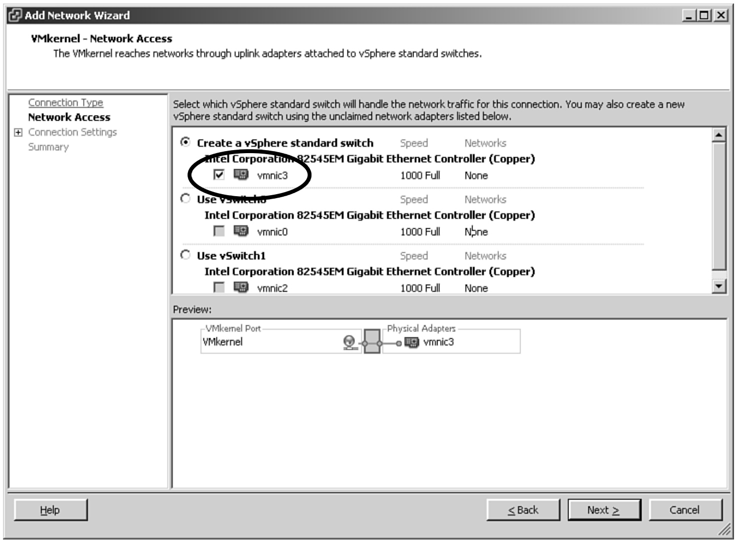
Add a labeled network to handle virtual machine network traffic.

● VMkernel

The VMkernel TCP/IP stack handles traffic for the following ESXi services: vSphere vMotion, iSCSI, NFS, and host management.

Figure 2-3 Selecting the VMkernel Connection Type

7. In VMkernel - Network Access, select the vmnic that you will use for the VMkernel port and click **Next**, as shown in Figure 2-4.



VMkernel - Network Access

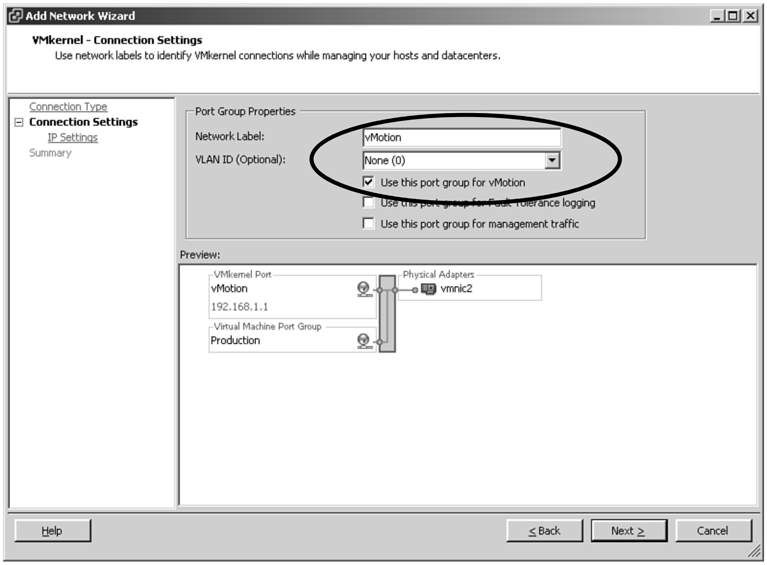
The VMkernel reaches networks through uplink adapters attached to vSphere standard switches.

Select which vSphere standard switch will handle the network traffic for this connection. You may also create a new vSphere standard switch using the unclaimed network adapters listed below.

	Speed	Networks
● Create a vSphere standard switch		
Intel Corporation 82545EM Gigabit Ethernet Controller (Copper)		
☑ vmnic3	1000 Full	None
○ Use vSwitch0		
Intel Corporation 82545EM Gigabit Ethernet Controller (Copper)		
□ vmnic0	1000 Full	None
○ Use vSwitch1		
Intel Corporation 82545EM Gigabit Ethernet Controller (Copper)		
□ vmnic2	1000 Full	None

Figure 2-4 Selecting a vmnic

8. In VMkernel - Connection Settings, enter the Network Label and optionally the VLAN, as shown in Figure 2-5. (The Network Label should generally indicate the purpose of the switch or port group. In this case, you might use vMotion, and then enable it for vMotion.) Click **Next**.



Add Network Wizard

VMkernel - Connection Settings

Use network labels to identify VMkernel connections while managing your hosts and datacenters.

Port Group Properties

Network Label: vMotion

VLAN ID (Optional): None (0)

☑ Use this port group for vMotion

☐ Use this port group for Fault Tolerance logging

☐ Use this port group for management traffic

Preview:

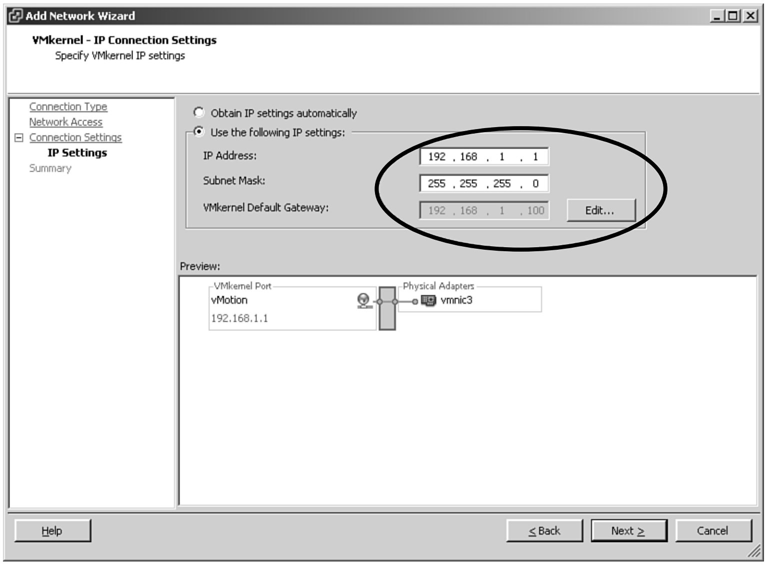
VMkernel Port: vMotion, 192.168.1.1

Virtual Machine Port Group: Production

Physical Adapters: vmnic2

Figure 2-5 Selecting the VMkernel Connection Type

9. In VMkernel - IP Connection Settings, enter the IP address, subnet mask, and VMkernel Default Gateway to be used for the switch, as shown in Figure 2-6, and then click **Next**. (I will discuss these settings in greater detail later in this chapter in the section “Creating/Configuring/Removing Virtual Adapters.”)



Add Network Wizard

VMkernel - IP Connection Settings

Specify VMkernel IP settings

○ Obtain IP settings automatically

● Use the following IP settings:

IP Address: 192 . 168 . 1 . 1

Subnet Mask: 255 . 255 . 255 . 0

VMkernel Default Gateway: 192 . 168 . 1 . 100

Preview:

VMkernel Port: vMotion, 192.168.1.1

Physical Adapters: vmnic3

Figure 2-6 Entering IP Information

10. In Ready to Complete, review your configuration settings and click **Finish**.

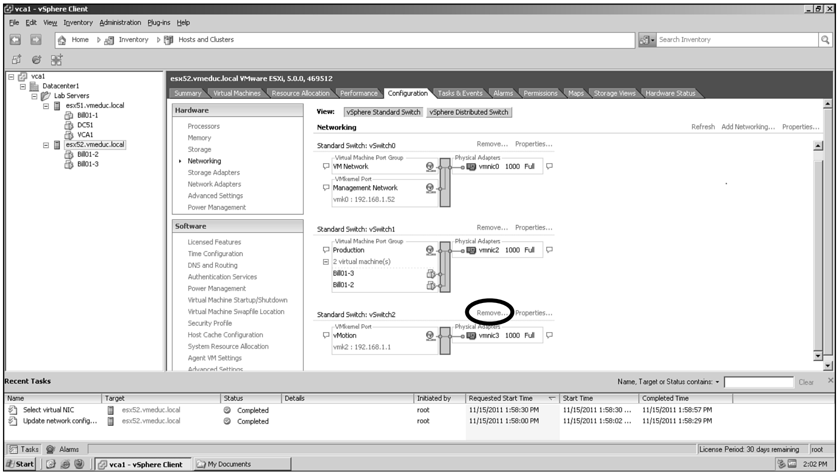
Deleting a vSphere Standard Switch

There might come a time when you no longer require a vSS that you have in your inventory. This might be because you have chosen to upgrade to a vSphere distributed switch (vDS) or because you are changing the networking on each of the hosts to provide consistency across the hosts, which is a very good idea. In this case, follow the steps outlined in Activity 2-2.

Activity 2-2 Deleting a vSphere Standard Switch

Key Topic

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the ESX host on which you want to delete the vSS, and then open the Configuration tab.
4. Click the **Networking** link under Hardware.
5. Click the **Remove** link next to the switch that you want to remove and then confirm your selection by clicking **Yes**, as shown in Figure 2-7. (There is a Remove link for each switch, so take care to select the right one.)



Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Select virtual NIC	esx52.vmeduc.local	Completed		root	11/15/2011 1:58:30 PM	11/15/2011 1:58:30 ...	11/15/2011 1:58:57 PM
Update network config...	esx52.vmeduc.local	Completed		root	11/15/2011 1:58:00 PM	11/15/2011 1:58:02 ...	11/15/2011 1:58:29 PM

Figure 2-7 Deleting a vSphere Standard Switch

Adding / Configuring / Removing vmnics on a vSphere Standard Switch

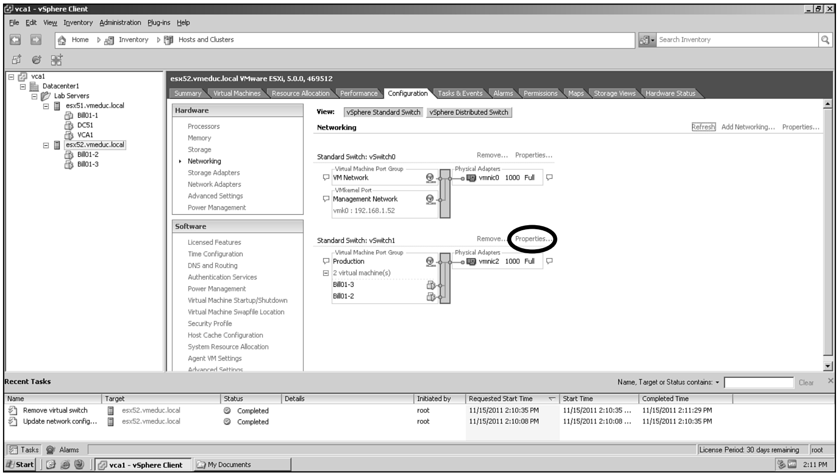
As I mentioned earlier, you might not want to create a new switch every time you need a new connection. In fact, you will make better use of your resources by adding to a current switch and thereby leveraging NIC teaming. In this section, I will discuss how to add new vmnics to a switch that you already have. In addition, I will discuss configuring vmnics and VMkernel ports on switches, including changing the IP address, VLAN, and so on. Finally, you will learn how to remove a vmnic from a switch if you no longer require it.

To add a new vmnic to an existing switch, you should *not* click Add Networking! As you might remember, clicking Add Networking takes you into a wizard that adds a new switch, not just into the networking properties of a switch you already have. So if you don't click Add Networking, what do you do? Well, if you think about it, what you really want to do is edit the configuration of a switch. For example, if you want to add a new vmnic to an existing switch to be used for vMotion, follow the steps outlined in Activity 2-3.

Key Topic

Activity 2-3 Adding a vmnic to a switch

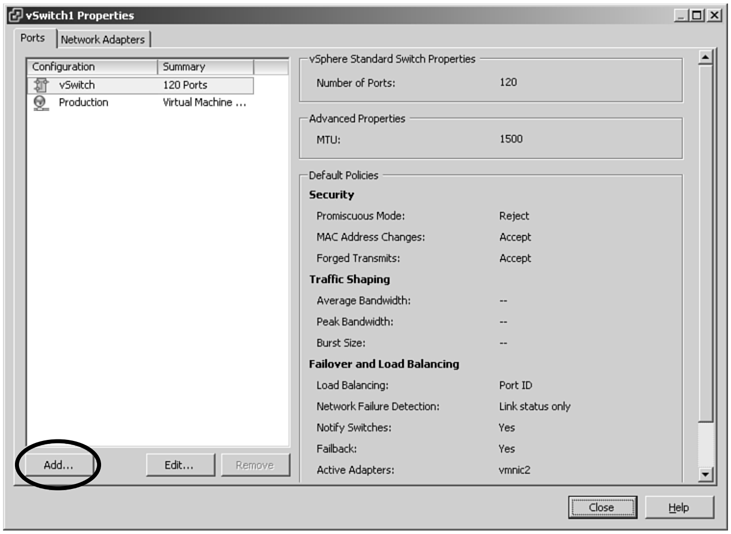
1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the ESX host on which you want to edit the vSS.
4. Click the **Networking** link under Hardware.
5. Click the **Properties** link next to the switch that you want to edit, as shown in Figure 2-8.



Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Remove virtual switch	esx52.vmeduc.local	Completed		root	11/15/2011 2:10:35 PM	11/15/2011 2:10:35 ...	11/15/2011 2:11:29 PM
Update network config...	esx52.vmeduc.local	Completed		root	11/15/2011 2:10:08 PM	11/15/2011 2:10:08 ...	11/15/2011 2:10:35 PM

Figure 2-8 The Properties Link on a vSS

6. On the Properties dialog box for the switch, click **Add**, as shown in Figure 2-9.



vSwitch1 Properties

Configuration	Summary
vSwitch	120 Ports
Production	Virtual Machine ...

vSphere Standard Switch Properties

Number of Ports:	120

Advanced Properties

MTU:	1500

Default Policies

Security

Promiscuous Mode:	Reject
MAC Address Changes:	Accept
Forged Transmits:	Accept

Traffic Shaping

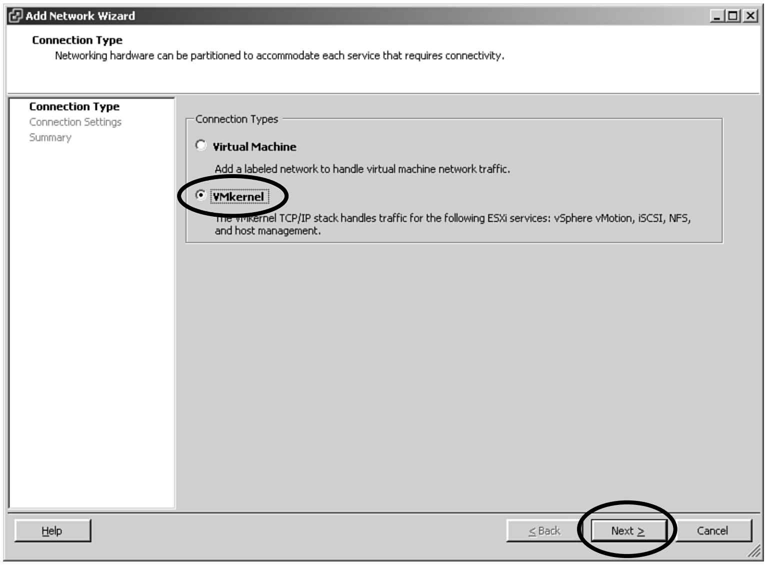
Average Bandwidth:	--
Peak Bandwidth:	--
Burst Size:	--

Failover and Load Balancing

Load Balancing:	Port ID
Network Failure Detection:	Link status only
Notify Switches:	Yes
Failback:	Yes
Active Adapters:	vmnic2

Figure 2-9 Adding a vmnic to a Switch

7. On the Connection Type of the Add Network Wizard, select **VMkernel** and click **Next**, as shown in Figure 2-10.



Add Network Wizard

Connection Type

Networking hardware can be partitioned to accommodate each service that requires connectivity.

Connection Types

○ **Virtual Machine**

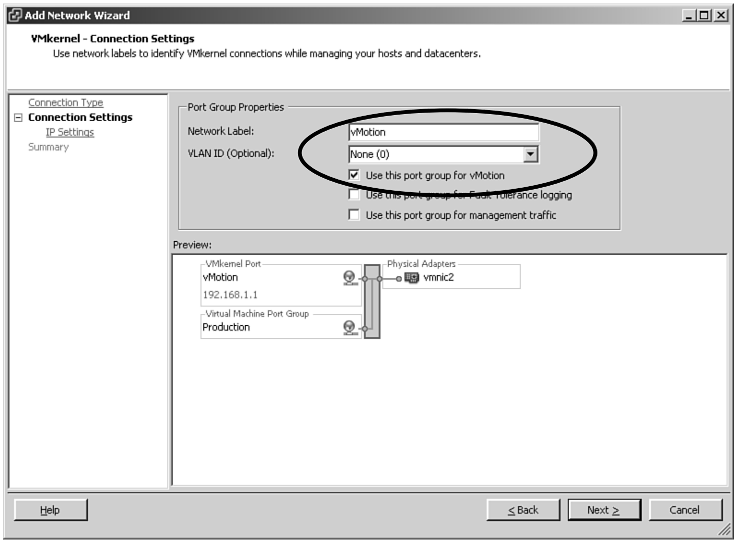
Add a labeled network to handle virtual machine network traffic.

● **VMkernel**

The VMkernel TCP/IP stack handles traffic for the following ESXi services: vSphere vMotion, iSCSI, NFS, and host management.

Figure 2-10 Selecting the VMkernel Connection Type

8. From **VMkernel > Connection Settings**, enter the Network Label and optionally the VLAN, as shown in Figure 2-11. (The Network Label should generally indicate the purpose of the switch or port group. In this case, you might use "vMotion" and enable it for vMotion.) Click Next.


Add Network Wizard

VMkernel - Connection Settings
Use network labels to identify VMkernel connections while managing your hosts and datacenters.

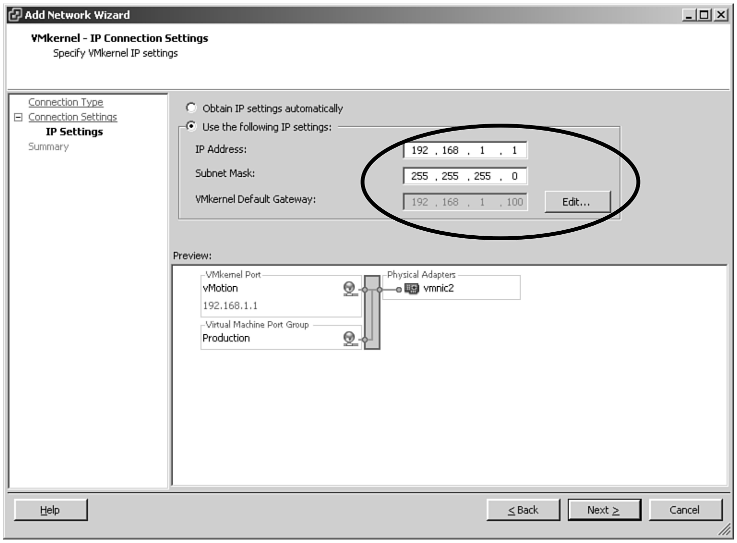
Connection Type
Connection Settings
IP Settings
Summary

Port Group Properties
Network Label: vMotion
VLAN ID (Optional): None (0)
☑ Use this port group for vMotion
□ Use this port group for Fault Tolerance logging
□ Use this port group for management traffic

Preview:
VMkernel Port
vMotion
192.168.1.1
Virtual Machine Port Group
Production
Physical Adapters
vmnic2

Figure 2-11 Entering a Network Label

9. From **VMkernel > IP Connection Settings**, enter the IP address, subnet mask, and VMkernel default gateway to be used for the switch, as shown in Figure 2-12, and then click **Next**.


Add Network Wizard

VMkernel - IP Connection Settings
Specify VMkernel IP settings

Connection Type
Connection Settings
IP Settings
Summary

○ Obtain IP settings automatically
● Use the following IP settings:
IP Address: 192 . 168 . 1 . 1
Subnet Mask: 255 . 255 . 255 . 0
VMkernel Default Gateway: 192 . 168 . 1 . 100

Preview:
VMkernel Port
vMotion
192.168.1.1
Virtual Machine Port Group
Production
Physical Adapters
vmnic2

Figure 2-12 Entering IP Information

10. In Ready to Complete, review your configuration settings and click **Finish**.

> **NOTE** As you might have noticed, after you select Add and Edit, the rest of the steps are very much the same whether you are creating a new switch for the port or just adding a port to an existing switch. This is not just a coincidence in this case, it is always true.

You will sometimes need to change the settings of a vmnic that you have already configured for a vSS. For example, you might want to edit the physical configuration such as the speed and duplex settings to match those of a physical switch to which your ESXi host is connected. To edit the physical configuration of the vmnic, follow the steps outlined in Activity 2-4.

Activity 2-4 Configuring the physical aspects of a vmnic

Key Topic

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the ESXi host on which you want to edit the vSS.
4. Click the **Networking** link under Hardware.
5. Click the **Properties** link next to the switch that you want to edit, as shown in Figure 2-13.

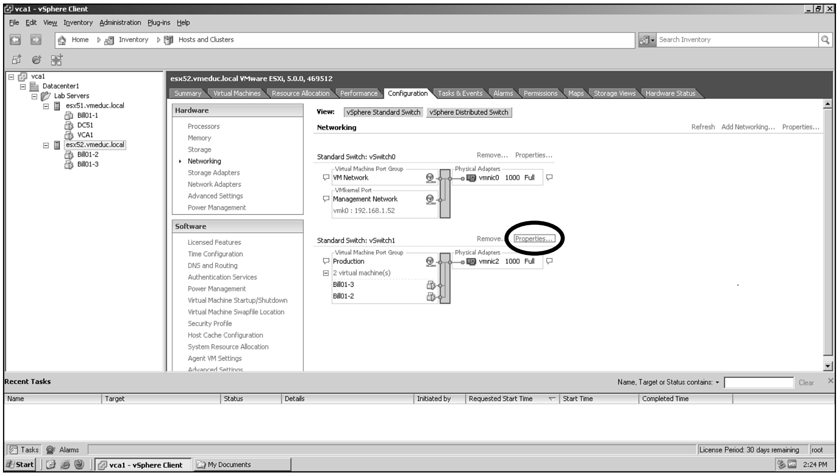
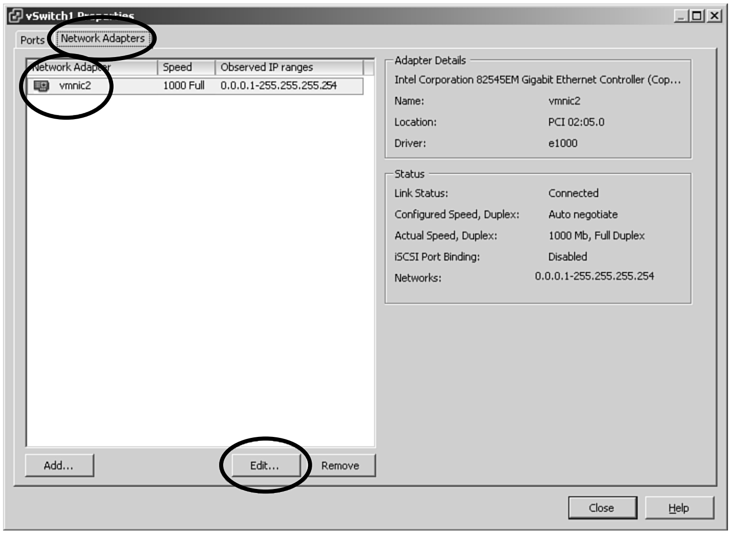

vca1 - vSphere Client
File Edit View Inventory Administration Plug-ins Help
Home Inventory Hosts and Clusters
Search Inventory
vca1
Datacenter1
Lab Servers
esx51.vmeduc.local
Bill01-1
DC51
VCA1
esx52.vmeduc.local
Bill01-2
Bill01-3
esx52.vmeduc.local VMware ESXi, 5.0.0, 469512
Summary Virtual Machines Resource Allocation Performance Configuration Tasks & Events Alarms Permissions Maps Storage Views Hardware Status
Hardware
Processors
Memory
Storage
Networking
Storage Adapters
Network Adapters
Advanced Settings
Power Management
Software
Licensed Features
Time Configuration
DNS and Routing
Authentication Services
Power Management
Virtual Machine Startup/Shutdown
Virtual Machine Swapfile Location
Security Profile
Host Cache Configuration
System Resource Allocation
Agent VM Settings
Advanced Settings
View: vSphere Standard Switch vSphere Distributed Switch
Networking
Refresh Add Networking... Properties...
Standard Switch: vSwitch0
Remove... Properties...
Virtual Machine Port Group
VM Network
VMkernel Port
Management Network
vmk0 : 192.168.1.52
Physical Adapters
vmnic0 1000 Full
Standard Switch: vSwitch1
Remove... Properties...
Virtual Machine Port Group
Production
2 virtual machine(s)
Bill01-3
Bill01-2
Physical Adapters
vmnic2 1000 Full
Recent Tasks
Name, Target or Status contains:
Clear
Name Target Status Details Initiated by Requested Start Time Start Time Completed Time
Tasks Alarms
License Period: 30 days remaining root
Start vca1 - vSphere Client My Documents
2:24 PM

Figure 2-13 The Properties Link

6. On the Properties dialog box for the switch, open the Network Adapters tab and select the vmnic that you want to configure, as shown in Figure 2-14.



Network Adapter	Speed	Observed IP ranges
vmnic2	1000 Full	0.0.0.1-255.255.255.254

Adapter Details

Intel Corporation 82545EM Gigabit Ethernet Controller (Cop...

Name: vmnic2

Location: PCI 02:05.0

Driver: e1000

Status

Link Status: Connected

Configured Speed, Duplex: Auto negotiate

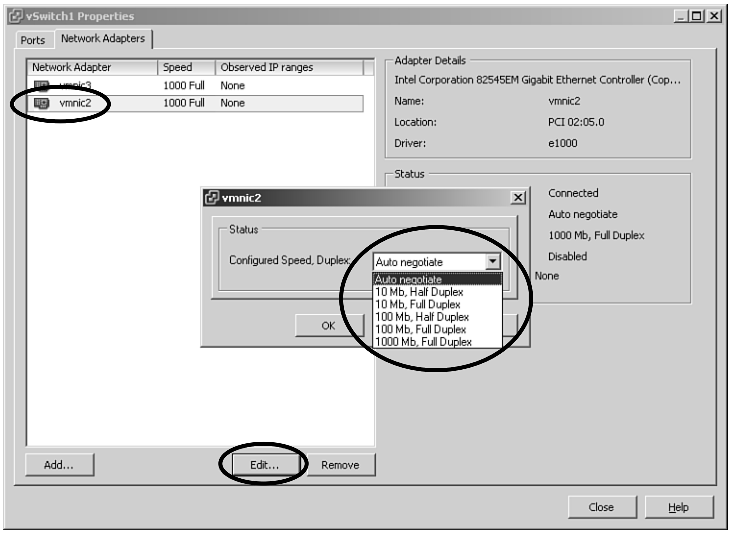
Actual Speed, Duplex: 1000 Mb, Full Duplex

iSCSI Port Binding: Disabled

Networks: 0.0.0.1-255.255.255.254

Figure 2-14 The Network Adapters Tab

7. Click **Edit**, and then select the speed and duplex that matches the physical switch to which the ESXi host is connected, as shown in Figure 2-15, and click **OK**.



Network Adapter	Speed	Observed IP ranges
vmnic3	1000 Full	None
vmnic2	1000 Full	None

Adapter Details

Intel Corporation 82545EM Gigabit Ethernet Controller (Cop...

Name: vmnic2

Location: PCI 02:05.0

Driver: e1000

Status

Connected

Auto negotiate

1000 Mb, Full Duplex

Disabled

None

vmnic2 — Status

Configured Speed, Duplex: Auto negotiate

Figure 2-15 Configuring Physical Aspects of a vmnic

8. Click **Close** to exit the Properties dialog box.

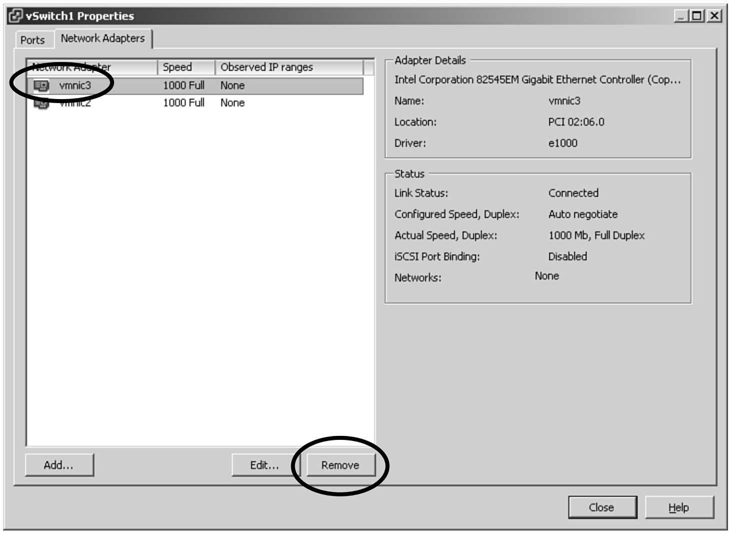
NOTE Auto Negotiate is the default, but is not always considered a best practice when more than one vendor is involved. This is because the result will often be less than the desired setting (such as 100Mb Half Duplex). If you use Auto Negotiate, verify that the resulting setting is what you expected.

There might come a time when you need to remove a vmnic from a switch. This might happen if you are changing network settings to provide consistency or if you intend to use the vmnic on a new switch. If you need to remove a vmnic from a vSS, follow the steps outlined in Activity 2-5.

Activity 2-5 Removing a vmnic from a vSphere Standard Switch

Key Topic

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the ESX host on which you want to remove the vmnic.
4. Click the **Networking** link under Hardware.
5. Click the **Properties** link next to the switch that contains the vmnic that you want to remove.
6. On the Properties dialog box for the switch, open the Network Adapters tab, select the vmnic that you want to remove, select **Remove**, and confirm your selection by clicking **Yes**, as shown in Figure 2-16.



Network Adapter	Speed	Observed IP ranges
vmnic3	1000 Full	None
vmnic2	1000 Full	None

Adapter Details
Intel Corporation 82545EM Gigabit Ethernet Controller (Cop...
Name: vmnic3
Location: PCI 02:06.0
Driver: e1000

Status
Link Status: Connected
Configured Speed, Duplex: Auto negotiate
Actual Speed, Duplex: 1000 Mb, Full Duplex
iSCSI Port Binding: Disabled
Networks: None

Figure 2-16 Removing a vmnic

Configuring VMkernel Ports for Network Services

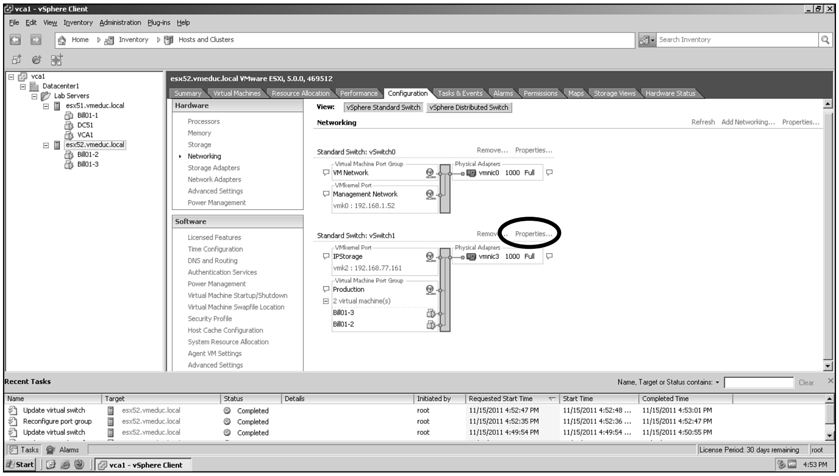
As I mentioned earlier, there are only four reasons that you would create a VMkernel port: management, IP storage, fault-tolerant logging, and vMotion. I will discuss each of these in much greater detail in the chapters that follow, but for now you should understand that they all share the same configuration requirements for network services (namely, an IP address and subnet mask). In addition, you should know that all VMkernel ports will share the same default gateway. You might also want to configure a VLAN, and you will want to enable the port with the services for which it was created (such as vMotion, management, or fault-tolerant logging).

To configure a VMkernel port with network service configuration, you should configure the IP settings of the port group to which is it assigned. I will discuss port group configuration in much greater detail later in this chapter. For now, if you want to configure the IP settings of a VMkernel port, follow the steps outlined in Activity 2-6.

Key Topic

Activity 2-6 Configuring a VMkernel port for Network Services

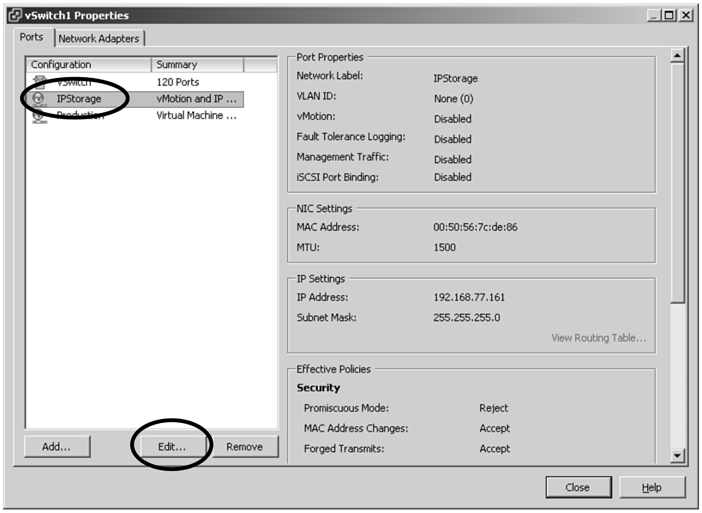
1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the ESX host on which you want to configure the VMkernel port.
4. Click the **Networking** link under Hardware.
5. Click the **Properties** link next to the switch that contains the port, as shown in Figure 2-17.



Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Update virtual switch	esx52.vmeduc.local	Completed		root	11/15/2011 4:52:47 PM	11/15/2011 4:52:48 ...	11/15/2011 4:53:01 PM
Reconfigure port group	esx52.vmeduc.local	Completed		root	11/15/2011 4:52:35 PM	11/15/2011 4:52:36 ...	11/15/2011 4:52:47 PM
Update virtual switch	esx52.vmeduc.local	Completed		root	11/15/2011 4:49:54 PM	11/15/2011 4:49:54 ...	11/15/2011 4:50:55 PM

Figure 2-17 Properties Link for vSS

6. On the Properties dialog box for the switch, on the Ports tab, select the port group to which the VMkernel port is assigned and click **Edit**, as shown in Figure 2-18.



Configuration	Summary
vSwitch	120 Ports
IPStorage	vMotion and IP ...
Production	Virtual Machine ...

Port Properties
Network Label: IPStorage
VLAN ID: None (0)
vMotion: Disabled
Fault Tolerance Logging: Disabled
Management Traffic: Disabled
iSCSI Port Binding: Disabled

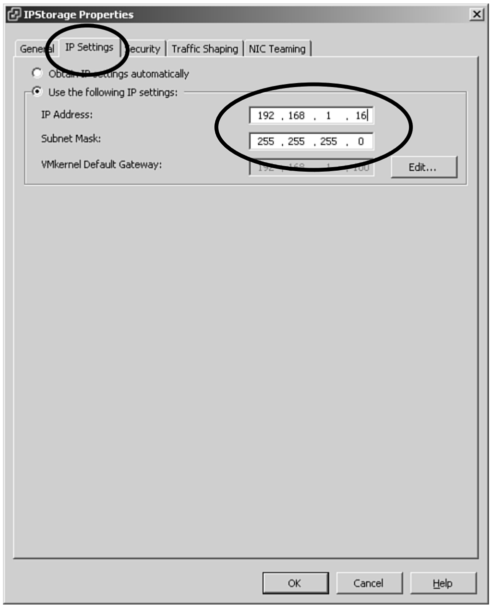
NIC Settings
MAC Address: 00:50:56:7c:de:86
MTU: 1500

IP Settings
IP Address: 192.168.77.161
Subnet Mask: 255.255.255.0

Effective Policies
Security
Promiscuous Mode: Reject
MAC Address Changes: Accept
Forged Transmits: Accept

Figure 2-18 Editing a Port Group

7. Open the IP Settings tab, and enter the IP information for your network, as shown in Figure 2-19, and click **OK**.



○ Obtain IP settings automatically
● Use the following IP settings:
IP Address: 192 . 168 . 1 . 16
Subnet Mask: 255 . 255 . 255 . 0
VMkernel Default Gateway: 192 . 168 . 1 . 100

Figure 2-19 Editing IP Information

8. If you want to configure a VLAN for the port group, open the General tab and enter the VLAN information directly under the Network Label.

9. On the General tab, you can also enable the vmnic for the specific services for which it was created, such as vMotion, FT Logging, or Management. If the port was only created for IP storage, you do not need to check any of the Enabled boxes.

10. Finally, if appropriate you can change the maximum transmission unit (MTU) for the vmnic (for example, if you are using jumbo frames for iSCSI storage). (Chapter 3 covers storage options in greater detail.) Click **OK** to close the Properties dialog box and save your settings.

Adding / Editing / Removing Port Groups on a vSphere Standard Switch

The main reason to use port groups is to get more than one function out of each switch. This is possible because port group configuration supersedes switch configuration. Because of this, you can have policies for security, traffic shaping, NIC teaming, and so on that apply to the switch but also have a separate policy for each that applies to any port group on which the settings differ from those of the switch. This tremendously improves your flexibility and gives you options such as those security options discussed in Chapter 1, "Planning, Installing, Configuring, and Upgrading vCenter Server and VMware ESXi." In this section, I will discuss adding, editing, and removing port groups on a vSS.

Suppose you decide to add a new group of VMs on which you will test software and monitor performance. Furthermore, suppose you decide that you will not create a new switch but that you will instead add the VMs to a switch that you already have in your inventory. However, suppose the VMs that are already on the switch are not for testing and development but are actually in production. Chances are good that you do not want to "mix them in" with the new testing VMs, but how can you keep them separate without creating a new vSS?

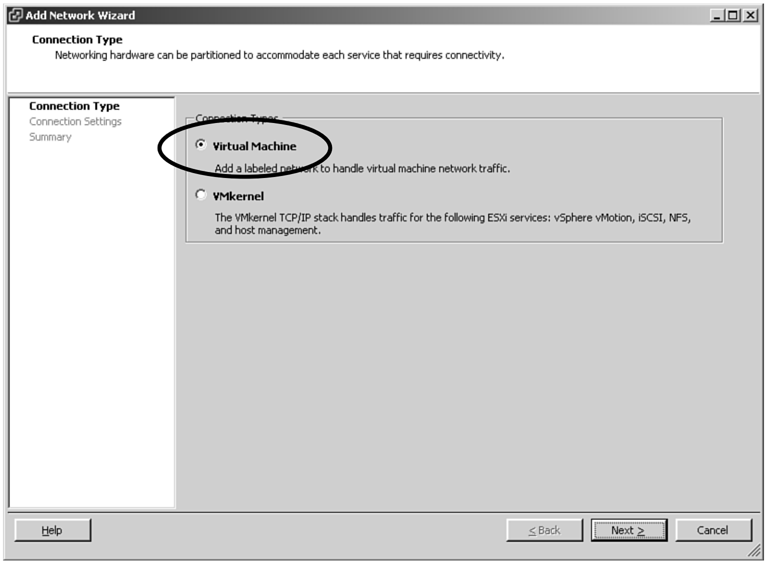
Well, if you create a new port group and assign a different vmnic to it, you can manage the new testing VMs completely separate from the production VMs, even though they are both on the same vSS. In this case, you might want to label your existing port group Production and label your new port group Test-Dev. It does not matter what label you use, but it is a best practice to relate it to the function of the port group, which is generally related to the function of the VMs that will be on it. Also, you should strive for consistency across all of your ESXi hosts in a small organization or at least across all of the hosts in the same cluster in a medium-sized or large organization. (Chapter 5 covers clusters in greater detail.)

So, what was the purpose of all of that labeling? Well, after you have done that, you will have a set of five tabs on the Properties link of the port group that only apply to that port group. You can make important changes to port group policies such as security, traffic shaping, and NIC teaming that will override any settings on the vSS properties tabs. I will discuss the details of these port group policies later in this chapter in "Configuring vSS and vDS Policies." For now, if you want to add a new VM port group to an existing vSS, follow the steps outlined in Activity 2-7.

Activity 2-7 Adding a Port Group to a vSphere Standard Switch

Key Topic

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the ESX host on which you want to add the port group.
4. Click the **Networking** link under Hardware.
5. Click the **Properties** link next to the switch on which you want to add the port group.
6. On the Ports tab, click **Add**, and then choose **Virtual Machine**, as shown in Figure 2-20. Click **Next**.



Connection Type

Networking hardware can be partitioned to accommodate each service that requires connectivity.

Connection Types

● Virtual Machine

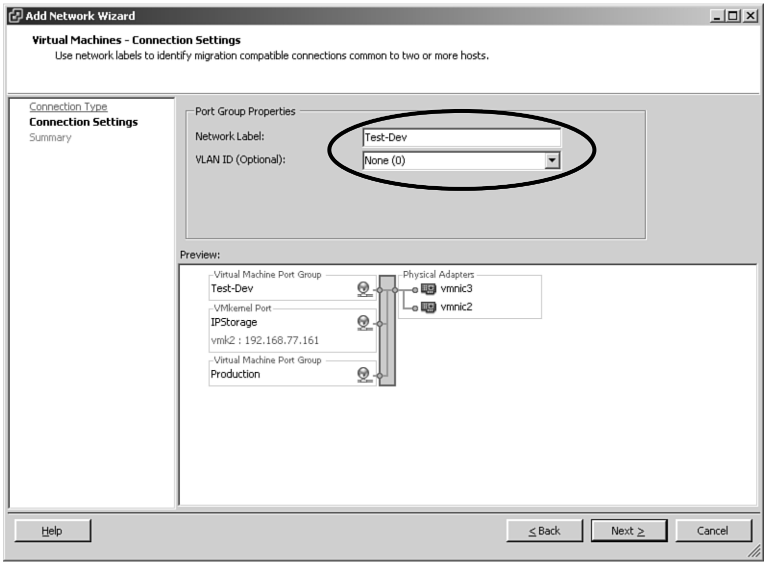
Add a labeled network to handle virtual machine network traffic.

○ VMkernel

The VMkernel TCP/IP stack handles traffic for the following ESXi services: vSphere vMotion, iSCSI, NFS, and host management.

Figure 2-20 Adding a Virtual Machine Port Group

7. From **Virtual Machines > Connection Settings**, enter the label that you want to use (such as Test-Dev) and the VLAN if you are using a VLAN, as shown in Figure 2-21. Click **Next**.



Add Network Wizard

Virtual Machines - Connection Settings

Use network labels to identify migration compatible connections common to two or more hosts.

Port Group Properties

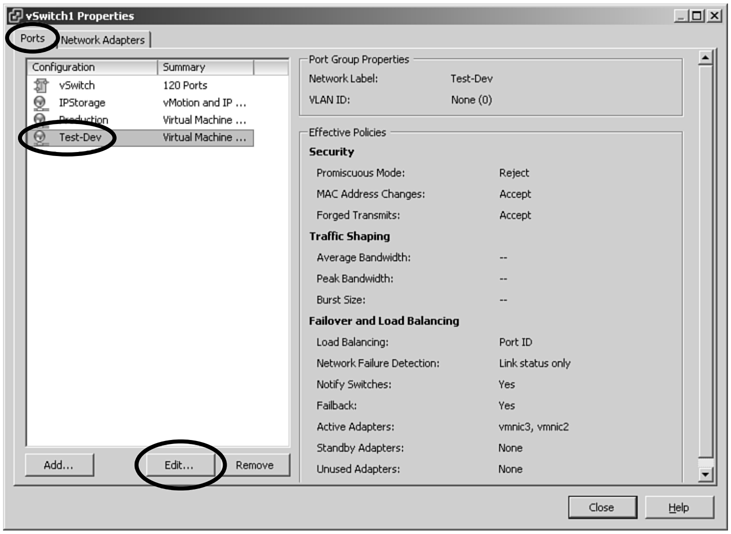
Network Label: Test-Dev

VLAN ID (Optional): None (0)

Figure 2-21 Entering and Network Label

8. On Ready to Complete, review your configuration settings and click **Finish**.

Your new port group should now appear in the Properties dialog box under Configuration. This new port group is now completely configurable and will have its own set of five tabs for you to configure. Just click the port group under Configuration and select **Edit**, as shown in Figure 2-22. I will discuss the configuration of port group policies in detail later in this chapter in the section "Configuring vSS and vDS Policies."



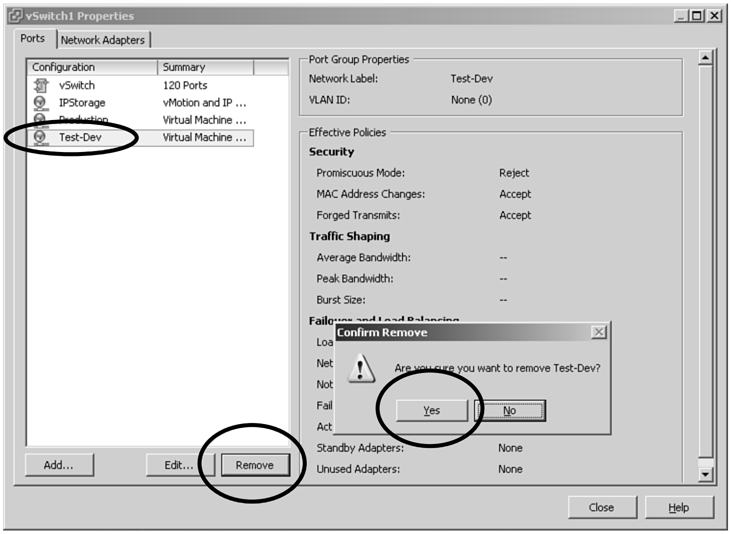
Configuration	Summary
vSwitch	120 Ports
IPStorage	vMotion and IP ...
Production	Virtual Machine ...
Test-Dev	Virtual Machine ...

Port Group Properties
Network Label: Test-Dev
VLAN ID: None (0)

Effective Policies
Security
Promiscuous Mode: Reject
MAC Address Changes: Accept
Forged Transmits: Accept
Traffic Shaping
Average Bandwidth: --
Peak Bandwidth: --
Burst Size: --
Failover and Load Balancing
Load Balancing: Port ID
Network Failure Detection: Link status only
Notify Switches: Yes
Failback: Yes
Active Adapters: vmnic3, vmnic2
Standby Adapters: None
Unused Adapters: None

Figure 2-22 Port Group Configuration

Finally, you might want to remove a port group that you no longer need. This might happen because you are reorganizing your network or because you are no longer using the VMs to which the port group was associated. To remove the port group, click the port group, select **Remove**, and confirm your selection by clicking **Yes**, as shown in Figure 2-23.



Configuration	Summary
vSwitch	120 Ports
IPStorage	vMotion and IP ...
Production	Virtual Machine ...
Test-Dev	Virtual Machine ...

Port Group Properties
Network Label: Test-Dev
VLAN ID: None (0)

Effective Policies
Security
Promiscuous Mode: Reject
MAC Address Changes: Accept
Forged Transmits: Accept
Traffic Shaping
Average Bandwidth: --
Peak Bandwidth: --
Burst Size: --
Failover and Load Balancing
Loa
Net
Not
Fail
Act
Standby Adapters: None
Unused Adapters: None

Are you sure you want to remove Test-Dev?

Figure 2-23 Removing a Port Group

Determining Use Cases for a vSphere Standard Switch

Now that I have discussed how you would create and manage a vSS, let's talk about why you would want one in the first place. In other words, what would cause you to use a vSS instead of a vDS? One very practical reason might be that you do not have the appropriate license to use a vDS. As I previously discussed in Chapter 1, in the section, "Installing and Configuring vCenter Server," creating a vDS requires an Enterprise Plus license. Another reason might be that you have a small to medium-size organization and therefore the settings on a vSS are sufficient for your needs. Your organization can have many hosts and those hosts can communicate to each other using vSSs.

The main point to consider is how you can keep the networking that is inside of each ESXi host consistent with the networking that is inside the other hosts, or at least all the hosts in the same cluster. If possible, you should have the same number of vSSs in each of your hosts and the same port groups on each of them as well (at least the ones that are in the same clusters). In fact, the consistent spelling of the port group names is even important. In addition, to leverage the power of port groups, you should have as few vSSs on each host as possible while still maintaining consistency across the hosts. If you balance these two factors in your organization as much as possible, you will be on the right track.

Configuring vSphere Distributed Switches

Now that you understand what a virtual switch does and understand that consistency of configuration is a key component, what if I were to tell you that there is a way to guarantee consistency by associating a virtual switch to more than one host at the same time? Well, that's what a vDS does.

A vDS is the same as a vSS in many ways except that it can be connected to more than one host at the same time, which makes a radical difference. I know what you're thinking, "Is it similar to a vSS or radically different?" Well, in a word, "Yes." It's similar in that it uses the same types of connections (namely, VMkernel and VMs). It's also similar in that the point at which the virtual world meets the physical world is an important thing to know and understand. However, it's radically different because it is managed centrally in the vCenter and can be connected to multiple hosts at the same time. In fact, a single vDS can be connected to as many as 350 hosts. Because of this difference, vDSs come with a whole new set of terms to understand.

In this section, I will discuss the capabilities of a vDS versus those of a vSS. I will also discuss creating and deleting a vDS and adding and removing ESXi hosts. In addition, I will cover adding, configuring, and removing dvPort groups, dvUplinks (new terms in vDSs). A vDS also has virtual adapters just like a vSS, except that they can be connected to more than one host. I will discuss creating, configuring, migrating, and removing virtual adapters. I will also cover migrating VMs to and from a vDS. In addition, you will learn how to determine a use case for a vSphere distributed switch.

Identifying vSphere Distributed Switch Capabilities

If I were you, what I would want to know is what vDSs can do that vSSs cannot do. In other words, "Why should I consider using one instead of the other?" In fact, there is quite a large list of features that are specific to a vDS, but to really understand them you need to see what they both can do and then what only the vDS can do. Table 2-2 illustrates the features that are common between vSSs and vDSs and then those that are unique to vDSs.

Key Topic

Table 2-2 vSS Capabilities Versus vDS Capabilities

	vSS	**vDS**
Layer 2 switch	X	X
VLAN segmentation	X	X
802.1Q tagging	X	X
NIC teaming	X	X
Outbound traffic shaping	X	X
Inbound traffic shaping		X
VM network port block		X
Private VLANs		X
Load-based teaming		X
Datacenter-level management		X
Network vMotion		X
vSphere switch APIs		X
Per-port policy settings		X
Port state monitoring		X
Link Layer Discovery Protocol (LLDP)		X
User-defined network I/O control		X
NetFlow		X
Port mirroring		X

The following is a brief description of each of the features available on a vDS that are not available on a vSS.

Key Topic

- **Inbound traffic shaping:** A port group setting that can throttle the aggregate bandwidth inbound to the switch. This might be useful for a port group containing VMs that are being used a web servers.
- **VM network port block:** Specific ports can be configured as "blocked" for a specified VMs use. This might be helpful for troubleshooting or for advanced configurations.
- **Private VLANs:** This is a vSphere implementation of a VLAN standard that is available on the latest physical switches. With regard to vSphere, private virtual local-area networks (PVLANs) can be created in the vSphere that are only used in the vSphere and not on your external network. In essence, a PVLAN is a VLAN within a VLAN. In addition, the PVLANs in your vSphere can be kept from seeing each other. Later in this chapter, the section "Configuring vSS and vDS Policies" covers PVLANs in greater depth.
- **Load-based teaming:** You can configure network load balancing in a much more intelligent fashion than with vSSs, by enabling the system to recognize the current load on each link before making frame forwarding decisions. This could be useful if the loads that are on each link vary considerably over time.
- **Datacenter-level management:** A vDS is managed from the vCenter as a single switch from the control plane, even though many hosts are connected to each other at the I/O plane. This provides a centralized control mechanism and guarantees consistency of configuration.
- **Network vMotion:** Because a port group that is on a vDS is actually connected to multiple hosts, a VM can migrate from one host to another without changing ports. The positive effect of this is that the attributes assigned to the port group (such as security, traffic shaping, and NIC teaming) will migrate as well.
- **vSphere switch APIs:** Third-party switches have been and are being created that can be installed in the control plane. On switches such as the Cisco Nexus 1000v, the true essence of the switch is installed into the vCenter as a virtual appliance (VA).
- **Per-port policy settings:** Most of the configuration on a vDS is at the port group level, but it can be overridden at the individual port level. This allows you tremendous flexibility with regard to port settings such as security, traffic shaping, and so on.

- **Port state monitoring:** Each port on vDS can be managed and monitored independently of all other ports. This means that you can quickly identify an issue that relates to a specific port.
- **Link Layer Discovery Protocol:** Similar to Cisco Discovery Protocol (CDP), Link Layer Discovery Protocol (LLDP) enables vDSs to discover other devices such as switches and routers that are directly connected to them. The advantage of LLDP is that it is an open protocol which is not proprietary to Cisco.
- **User-defined network I/O control:** You can set up a quality of service (QoS) (of a sort), but instead of defining traffic paths by protocols, you can define the traffic paths by types of VMware traffic. In earlier versions of vDSs, you could define traffic as vMotion, Management, and others, but now you can define your own categories. This adds to flexibility in network control and design.
- **NetFlow:** You can use the standard for traffic monitoring, NetFlow, to monitor, analyze, and log traffic flows in your vSphere. This enables you to easily monitor virtual network flows with the same tools that you use to monitor traffic flows in the physical network. Your vDS can forward NetFlow information to a monitoring machine in your external network.
- **Port mirroring:** Most commonly used with intrusion detection systems (IDSs) and intrusion prevention systems (IPSs), port mirroring provides for a copy of a packet to be sent to a monitoring station so that traffic flows can be monitored without the IPS/IDS skewing the data. Port mirroring is new to vSphere 5.0 vDSs.

NOTE As you might remember, one of the main goals with vSSs was consistency of networking between hosts that are in the same clusters. Likewise, one of the main benefits of vDSs is that they “force” this consistency, because multiple hosts are connected to the same virtual switch.

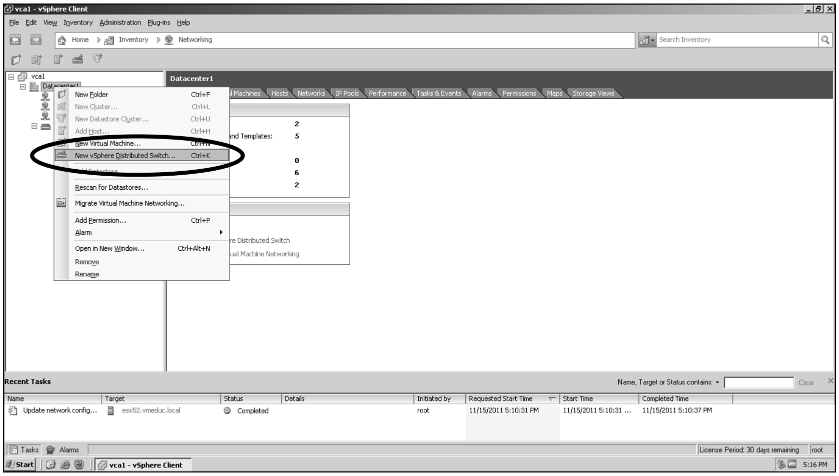
Creating/Deleting a vSphere Distributed Switch

The first thing to consider if you want to create a vDS is your license level, because they can be created only with an Enterprise Plus license. Oh, I suppose you could create them with the 60-day evaluation license, but you would then need to purchase an Enterprise Plus license before the evaluation period expires; otherwise, your switch would cease to function. You also need to consider the level of hosts that you have in the datacenter onto which you are adding the switch, because this will have an impact on the version of the switch that you create. That said, to begin to create a new vDS, follow the steps outlined in Activity 2-8.

Key Topic

Activity 2-8 Creating an New vSphere Distributed Switch

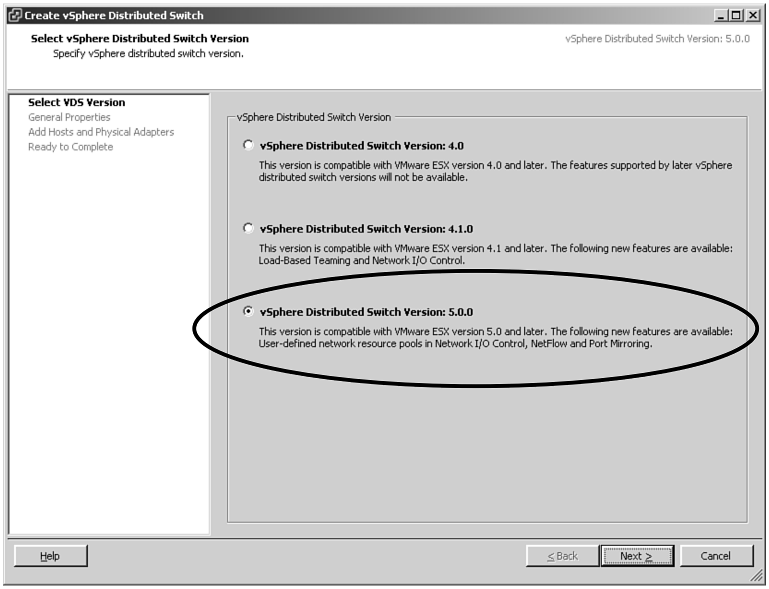
1. Log on to your vSphere Client.
2. Select **Home** and then **Networking**.
3. Right-click your datacenter and then select **New vSphere Distributed Switch**, as shown in Figure 2-24.


vca1 - vSphere Client
File Edit View Inventory Administration Plug-ins Help
Home Inventory Networking
Search Inventory
vca1
Datacenter1
Datacenter1
Virtual Machines Hosts Networks IP Pools Performance Tasks & Events Alarms Permissions Maps Storage Views
New Folder Ctrl+F
New Cluster... Ctrl+L
New Datastore Cluster... Ctrl+U
Add Host... Ctrl+H
New Virtual Machine... Ctrl+N
New vSphere Distributed Switch... Ctrl+K
Rescan for Datastores...
Migrate Virtual Machine Networking...
Add Permission... Ctrl+P
Alarm
Open in New Window... Ctrl+Alt+N
Remove
Rename
and Templates:
2
5
0
6
2
re Distributed Switch
ual Machine Networking
Recent Tasks
Name, Target or Status contains:
Clear
Tasks Alarms
License Period: 30 days remaining
root
Start
vca1 - vSphere Client
5:16 PM

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Update network config...	esx52.vmeduc.local	Completed		root	11/15/2011 5:10:31 PM	11/15/2011 5:10:31 ...	11/15/2011 5:10:37 PM

Figure 2-24 Creating a New vDS

4. On Select vDS Version, choose the level of switch that fits your datacenter based on the hosts that you have in it. For example, if all of your hosts are ESXi 5.0, you can use a Version 5.0.0 switch. However, if you have hosts that are older than ESXi 5.0, you want to choose the version corresponding to the earliest version in your datacenter. This will, of course, affect the list of features that you will have on your switch, as shown in Figure 2-25. Click **Next**.



Select vSphere Distributed Switch Version
Specify vSphere distributed switch version.

vSphere Distributed Switch Version: 5.0.0

vSphere Distributed Switch Version

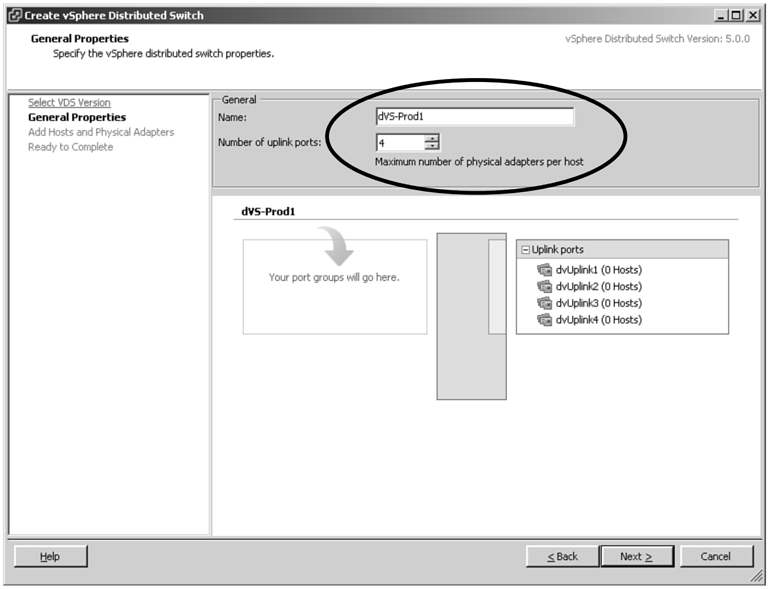
○ **vSphere Distributed Switch Version: 4.0**
This version is compatible with VMware ESX version 4.0 and later. The features supported by later vSphere distributed switch versions will not be available.

○ **vSphere Distributed Switch Version: 4.1.0**
This version is compatible with VMware ESX version 4.1 and later. The following new features are available: Load-Based Teaming and Network I/O Control.

● **vSphere Distributed Switch Version: 5.0.0**
This version is compatible with VMware ESX version 5.0 and later. The following new features are available: User-defined network resource pools in Network I/O Control, NetFlow and Port Mirroring.

Figure 2-25 vDS Versions

5. On General Properties, type a name for your new switch that implies what it does and select the maximum number of uplinks that you will want to use per host for this switch, as shown in Figure 2-26. Click **Next**.


Create vSphere Distributed Switch
Select VDS Version
General Properties
Add Hosts and Physical Adapters
Ready to Complete
Your port groups will go here.
Help
< Back
Next >
Cancel

General Properties
Specify the vSphere distributed switch properties.

vSphere Distributed Switch Version: 5.0.0

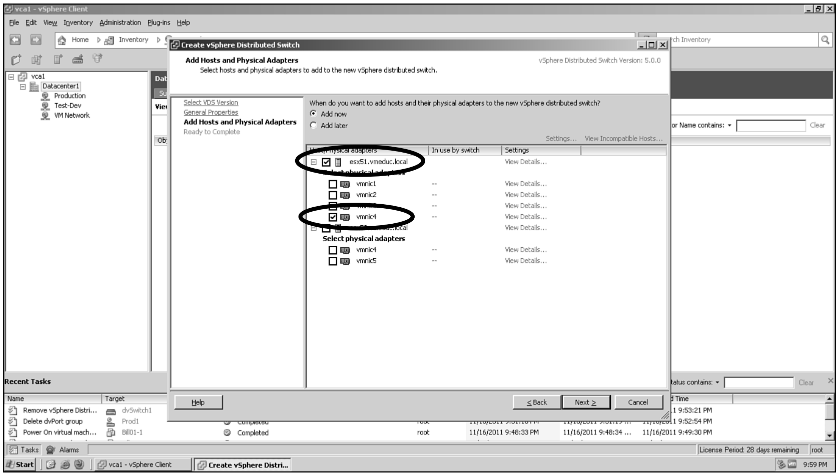
General
Name: dVS-Prod1
Number of uplink ports: 4
Maximum number of physical adapters per host

dVS-Prod1

Uplink ports
dvUplink1 (0 Hosts)
dvUplink2 (0 Hosts)
dvUplink3 (0 Hosts)
dvUplink4 (0 Hosts)

Figure 2-26 General Properties for a vDS

6. From Add Hosts and Physical Adapters, select the hosts that you want to add and the vmnic that you will connect to on each host, as shown in Figure 2-27. (I am selecting just one host for now.) Click **Next**.



Create vSphere Distributed Switch

Add Hosts and Physical Adapters
Select hosts and physical adapters to add to the new vSphere distributed switch.

vSphere Distributed Switch Version: 5.0.0

When do you want to add hosts and their physical adapters to the new vSphere distributed switch?
● Add now
○ Add later

Host/Physical adapters	In use by switch	Settings
☑ esx51.vmeduc.local		View Details...
Select physical adapters		
□ vmnic1	--	View Details...
□ vmnic2	--	View Details...
□	--	View Details...
☑ vmnic4	--	View Details...
□		View Details...
Select physical adapters		
□ vmnic4	--	View Details...
□ vmnic5	--	View Details...

Figure 2-27 Adding Hosts while Creating a vDS

7. On Ready to Complete, review your configuration settings and click **Finish**.

Deleting a vDS

You might assume that deleting a vDS would just be a matter of right-clicking it and selecting to remove it. This is almost true. However, you first need to remove the hosts and the port groups from the vDS. Then you can right-click it and select to remove it. In the next two sections, I will discuss (among other topics) removing hosts and port groups from a vDS. Once you know how to do that, deleting the vDS is as simple as right-clicking and selecting **Remove**.

Adding/Removing ESXi Hosts from a vSphere Distributed Switch

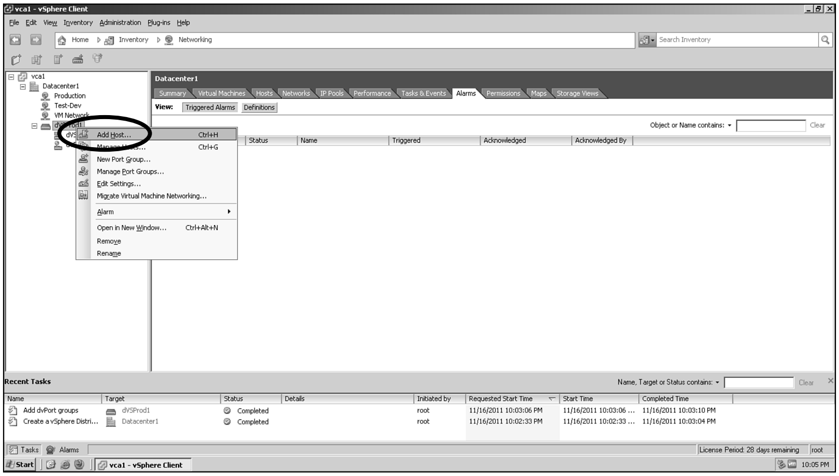
As you observed earlier, you can add hosts to a vDS when you create the switch in the first place, but you certainly do not have to recreate the switch to change the number of hosts that are connected to it. Instead, you can use the tools provided by the vCenter to make the modifications with relative ease. In the following activities, I will first illustrate how to add a host to an existing vDS, and then I will show you how to remove a host from an existing vDS.

To add a host to an existing vDS, follow the steps outlined in Activity 2-9.

Key Topic

Activity 2-9 Adding a Host to a vSphere Distributed Switch

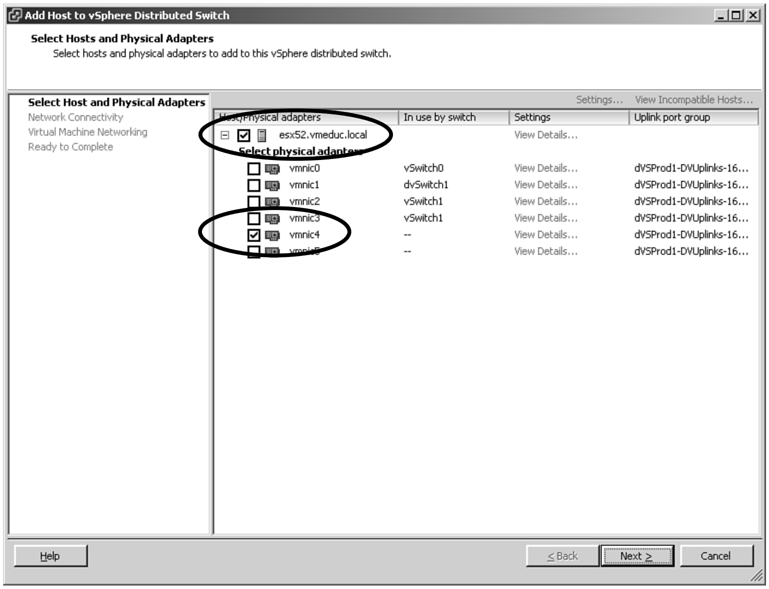
1. Log on to your vSphere Client.
2. Select **Home** and then **Networking**.
3. Right-click the vDS on which you want to add a host and click **Add Host**, as shown in Figure 2-28.



Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Add dvPort groups	dVSProd1	Completed		root	11/16/2011 10:03:06 PM	11/16/2011 10:03:06 ...	11/16/2011 10:03:10 PM
Create a vSphere Distri...	Datacenter1	Completed		root	11/16/2011 10:02:33 PM	11/16/2011 10:02:33 ...	11/16/2011 10:03:04 PM

Figure 2-28 Adding Hosts After Creating a vDS

4. From Select Hosts and Physical Adapters, choose the host and the vmnic to which you want to connect, as shown in Figure 2-29. (You can determine the appropriate vmnic by clicking **View Details** and examining the properties of the card and the CDP information associated with it. Take care that you know whether it is in use by another switch.) Click **Next**.



Add Host to vSphere Distributed Switch

Select Hosts and Physical Adapters

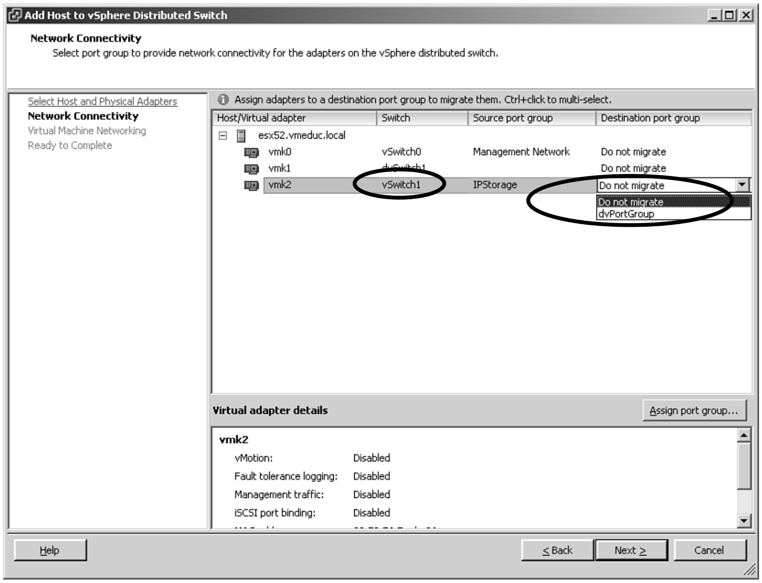
Select hosts and physical adapters to add to this vSphere distributed switch.

Select Host and Physical Adapters
Network Connectivity
Virtual Machine Networking
Ready to Complete
Settings...
View Incompatible Hosts...

Host/Physical adapters	In use by switch	Settings	Uplink port group
⊟ ☑ esx52.vmeduc.local		View Details...	
Select physical adapters			
☐ vmnic0	vSwitch0	View Details...	dVSProd1-DVUplinks-16...
☐ vmnic1	dvSwitch1	View Details...	dVSProd1-DVUplinks-16...
☐ vmnic2	vSwitch1	View Details...	dVSProd1-DVUplinks-16...
☐ vmnic3	vSwitch1	View Details...	dVSProd1-DVUplinks-16...
☑ vmnic4	--	View Details...	dVSProd1-DVUplinks-16...
☐ vmnic5	--	View Details...	dVSProd1-DVUplinks-16...

Figure 2-29 Connecting vmnics on a vDS

5. On Network Connectivity, select the port group that will provide network connectivity for the host. If you choose a VMkernel port group that is already associated to a vmnic on another switch, you must migrate the VMkernel port to the vDS or choose another VMkernel port group, as shown in Figure 2-30. Click **Next**.



Add Host to vSphere Distributed Switch

Network Connectivity

Select port group to provide network connectivity for the adapters on the vSphere distributed switch.

Assign adapters to a destination port group to migrate them. Ctrl+click to multi-select.

Host/Virtual adapter	Switch	Source port group	Destination port group
⊟ esx52.vmeduc.local			
vmk0	vSwitch0	Management Network	Do not migrate
vmk1	dvSwitch1		Do not migrate
vmk2	vSwitch1	IPStorage	Do not migrate Do not migrate dvPortGroup

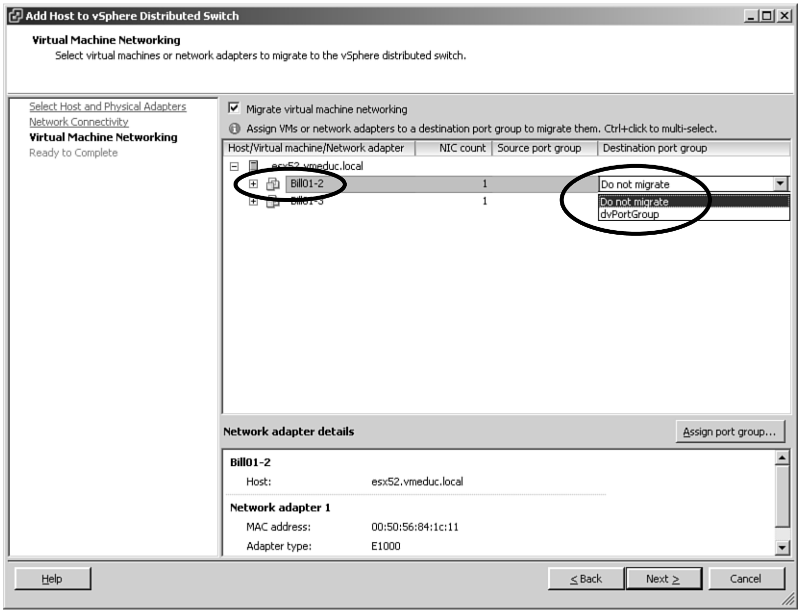
Virtual adapter details

vmk2

vMotion:	Disabled
Fault tolerance logging:	Disabled
Management traffic:	Disabled
iSCSI port binding:	Disabled

Figure 2-30 Migrating VMkernel Port Groups to a vDS

6. From Virtual Machine Networking, select the VMs on the vSSs of the host that you want to migrate to the vDS and the port group that you want to migrate them to, as shown in Figure 2-31. Click **Next**.



Add Host to vSphere Distributed Switch

Virtual Machine Networking

Select virtual machines or network adapters to migrate to the vSphere distributed switch.

☑ Migrate virtual machine networking

Assign VMs or network adapters to a destination port group to migrate them. Ctrl+click to multi-select.

Host/Virtual machine/Network adapter	NIC count	Source port group	Destination port group
esx52.vmeduc.local			
Bill01-2	1		Do not migrate
Bill01-3	1		

Network adapter details

Bill01-2

Host: esx52.vmeduc.local

Network adapter 1

MAC address: 00:50:56:84:1c:11

Adapter type: E1000

Figure 2-31 Migrating VM Networking to a VDS

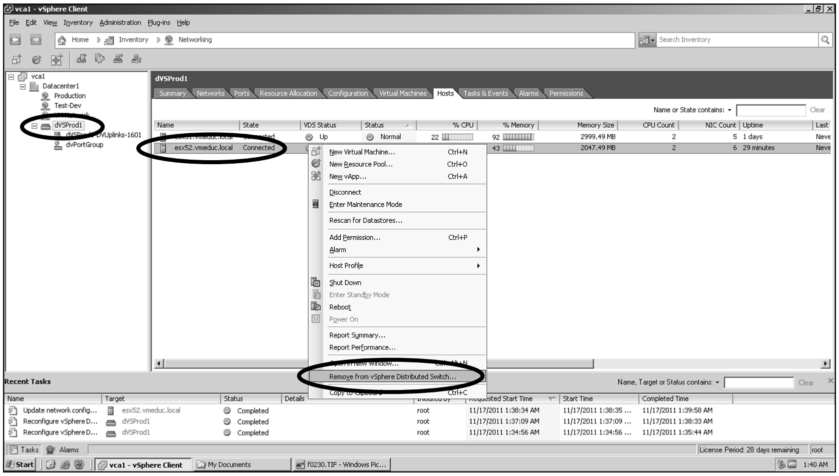
7. From Ready to Complete, review your configuration settings and click **Finish**.

To remove a host from an existing vDS, follow the steps outlined in Activity 2-10.

Activity 2-10 Removing a Host from a vSphere Distributed Switch

Key Topic

1. Log on to your vSphere Client.
2. Select **Home** and then **Networking**.
3. Click the vDS on which you want to remove a host and open the Hosts tab.
4. On the Hosts tab, right-click the host that you want to remove from the switch and select **Remove from vSphere Distributed Switch**, as shown in Figure 2-32.



Name	State	VDS Status	Status	% CPU	% Memory	Memory Size	CPU Count	NIC Count	Uptime	Last
esx51.vmeduc.local	Connected	Up	Normal	22	92	2999.49 MB	2	5	1 days	Neve
esx52.vmeduc.local	Connected				43	2047.49 MB	2	6	29 minutes	Neve

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Update network config...	esx52.vmeduc.local	Completed		root	11/17/2011 1:38:34 AM	11/17/2011 1:38:35 ...	11/17/2011 1:39:58 AM
Reconfigure vSphere D...	dVSProd1	Completed		root	11/17/2011 1:37:09 AM	11/17/2011 1:37:09 ...	11/17/2011 1:38:33 AM
Reconfigure vSphere D...	dVSProd1	Completed		root	11/17/2011 1:34:56 AM	11/17/2011 1:34:56 ...	11/17/2011 1:35:44 AM

Figure 2-32 Removing a Host from a VDS

5. On the warning screen, confirm your selection by clicking **Yes**. (You should ensure that that there are no VM resources on the switch; otherwise, the removal will fail.)

NOTE It is necessary to remove a host from a vDS before you can remove the host from vCenter.

Adding/Configuring/Removing dvPort Groups

As you might remember, I said earlier that port groups allow you to get more than one set of attributes out of the same switch. This is especially true with vDS port groups. The port groups that you create on a vDS are connected to all of the hosts to which the vDS is connected; hence they are called *dvPort groups*. Because a vDS can be connected to up to 350 hosts, the dvPort groups can become very large and powerful indeed. After you create port groups on a vDS, you can migrate your VMs to the dvPort groups. In the following activities, I will illustrate how to add, configure, and remove dvPort groups on vDSs.

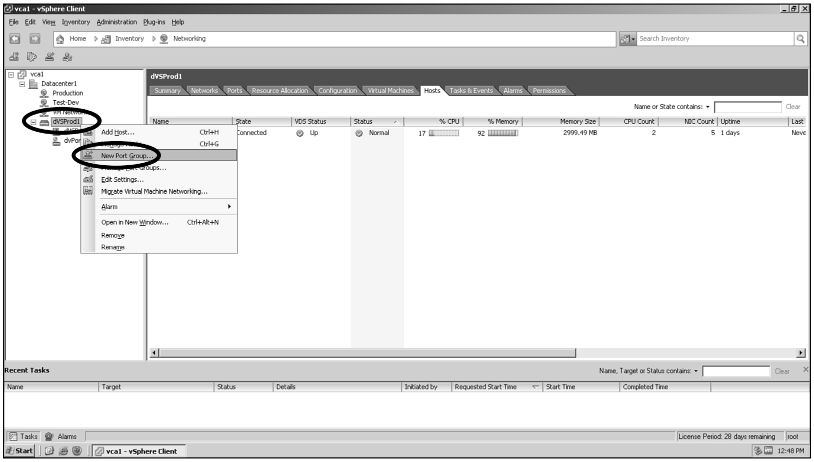
To add a port group to a vDS, follow the steps outlined in Activity 2-11.

Key Topic

Activity 2-11 Adding a Port Group to a vSphere Distributed Switch

1. Log on to your vSphere Client.
2. Select **Home** and then **Networking**.

3. Right-click the vDS on which you want to add the port group and select **New Port Group**, as shown in Figure 2-33.

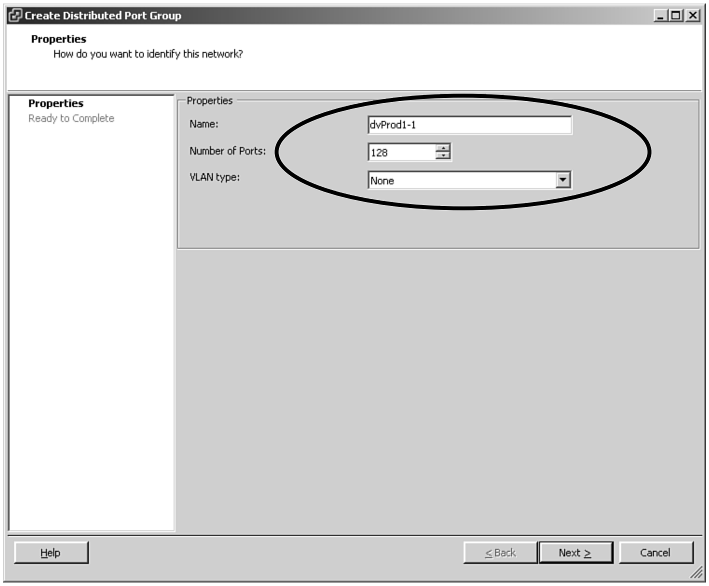

vca1 - vSphere Client
File Edit View Inventory Administration Plug-ins Help
Home Inventory Networking
Search Inventory
vca1
Datacenter1
Production
Test-Dev
VM Network
dVSProd1
dvPor
Add Host... Ctrl+H
Ctrl+G
New Port Group...
Edit Settings...
Migrate Virtual Machine Networking...
Alarm
Open in New Window... Ctrl+Alt+N
Remove
Rename
dVSProd1
Summary Networks Ports Resource Allocation Configuration Virtual Machines Hosts Tasks & Events Alarms Permissions
Name or State contains:
Clear
Recent Tasks
Name, Target or Status contains:
Clear
Tasks Alarms
License Period: 28 days remaining
root
Start
vca1 - vSphere Client
12:48 PM

Name	State	VDS Status	Status	% CPU	% Memory	Memory Size	CPU Count	NIC Count	Uptime	Last
	Connected	Up	Normal	17	92	2999.49 MB	2	5	1 days	Neve

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time

Figure 2-33 Adding a Port Group to a vDS

4. Type a name for your new port group that will help you identify the types of VMs that you will place on that port group, choose the number of ports that will be assigned to this port group (0–8192), and choose the VLAN type (I will discuss VLANs later in this chapter in the section "Configuring VLAN Settings"), as shown in Figure 2-34.


Create Distributed Port Group
Properties
How do you want to identify this network?
Properties
Ready to Complete
Properties
Help
< Back
Next >
Cancel

Name: dvProd1-1

Number of Ports: 128

VLAN type: None

Figure 2-34 Naming a dvPort Group

5. On Ready to Complete, confirm your selections by clicking **Finish**.

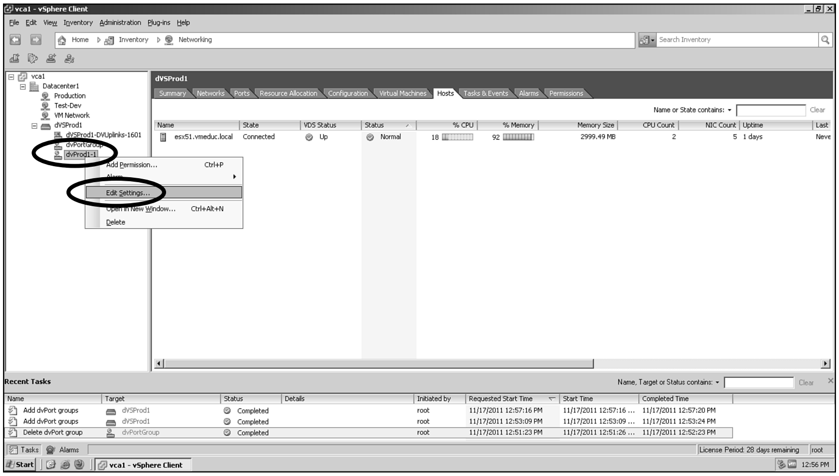
> **NOTE** One port group, named dvPortGroup, is created by default when you create the switch. For your first port group, you could just rename that one.

I will discuss configuring port groups in great detail in the next major section of this chapter, which covers configuring vSS and vDS polices. For now, I will just point out the steps involved in accessing the area in which you can configure the policies of port groups on vDSs. To begin to configure a port group on a vDS, follow the steps outlined in Activity 2-12.

Key Topic

Activity 2-12 Configuring Port Groups on a vSphere Distributed Switch

1. Log on to your vSphere Client.
2. Select **Home** and then **Networking**.
3. Expand on the vDS on which you want to configure the port group, right-click the port group that you want to configure, and select **Edit Settings**, as shown in Figure 2-35.



Name	State	VDS Status	Status	% CPU	% Memory	Memory Size	CPU Count	NIC Count	Uptime	Last
esx51.vmeduc.local	Connected	Up	Normal	18	92	2999.49 MB	2	5	1 days	Neve

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Add dvPort groups	dVSProd1	Completed		root	11/17/2011 12:57:16 PM	11/17/2011 12:57:16 ...	11/17/2011 12:57:20 PM
Add dvPort groups	dVSProd1	Completed		root	11/17/2011 12:53:09 PM	11/17/2011 12:53:09 ...	11/17/2011 12:53:24 PM
Delete dvPort group	dvPortGroup	Completed		root	11/17/2011 12:51:23 PM	11/17/2011 12:51:26 ...	11/17/2011 12:52:23 PM

Figure 2-35 Configuring a dvPort Group

4. In the warning box, confirm your selection by choosing **Yes**.

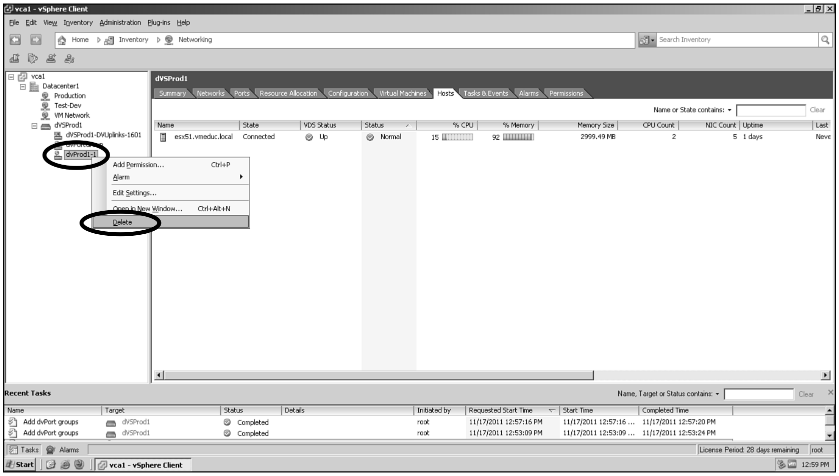
Over time, your networking needs will change, and you might decide to reorganize by removing some port groups. Take care not to “orphan” the VMs by removing the port group while they are still assigned to it. Instead, carefully consider your options and plans, and simply migrate the VMs to another port group as part of your plan. I will discuss your options with regard to migrating VMs later in Chapter 5. For now, I’ll just point out how you would go about removing a port group after you have migrated the VMs.

To remove a port group that you no longer are using, follow the steps outlined in Activity 2-13.

Activity 2-13 Removing a Port Group from a vSphere Distributed Switch

Key Topic

1. Log on to your vSphere Client.
2. Select **Home** and then **Networking**.
3. Click the vDS on which you want to remove the port group.
4. Right-click the port that you want to remove and select **Delete**, as shown in Figure 2-36.



Name	State	VDS Status	Status	% CPU	% Memory	Memory Size	CPU Count	NIC Count	Uptime	Last
esx51.vmeduc.local	Connected	Up	Normal	15	92	2999.49 MB	2	5	1 days	Neve

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Add dvPort groups	dVSProd1	Completed		root	11/17/2011 12:57:16 PM	11/17/2011 12:57:16 ...	11/17/2011 12:57:20 PM
Add dvPort groups	dVSProd1	Completed		root	11/17/2011 12:53:09 PM	11/17/2011 12:53:09 ...	11/17/2011 12:53:24 PM

Figure 2-36 Removing a dvPort Group

5. On the warning screen, confirm your selection by clicking **Yes**.

Adding/Removing Uplink Adapters to dvUplink Groups

As shown in Figure 2-37, dvUplink groups connect your vDS to the hidden switches that are contained in your hosts and then from there to the physical world. This allows you to control networking at the control plane on the vDS while the actual input/out (I/O) is still passing from host to host at the I/O plane. Each host keeps its own network configuration in its hidden switch that is created when you add a host to a vDS. This ensures that the network will continue to function even if your vCenter fails or is not available.

I know that's a lot more terminology all of the sudden, but as you might remember, I said that one of the main things to understand was where the virtual meets the physical. Well, you should know that the dvUplink groups are virtual but the uplink adapters are physical. Connecting multiple uplink adapters to a dvUplink group opens up the possibilities of load balancing and fault tolerance, which I discuss in detail later in the section "Configuring Load Balancing and Failover Policies." For now, I will show you how to add and remove uplink adapters.

Key Topic

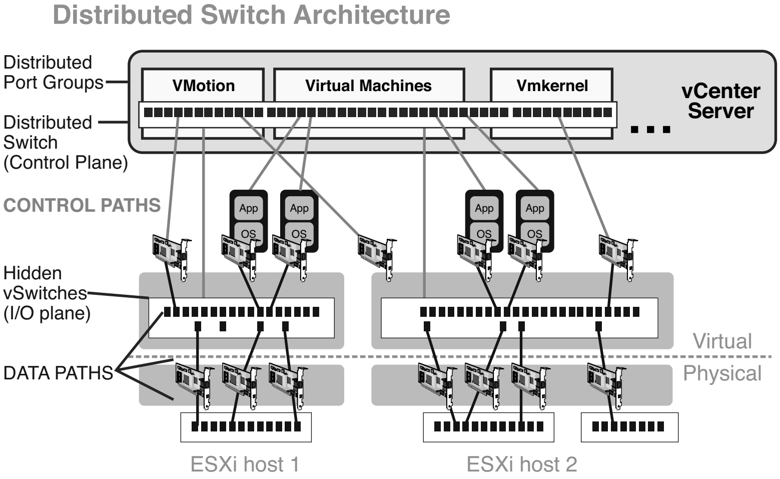


Figure 2-37 Distributed Switch Architecture

To add uplink adapters to a dvUplink group, follow the steps outlined in Activity 2-14.

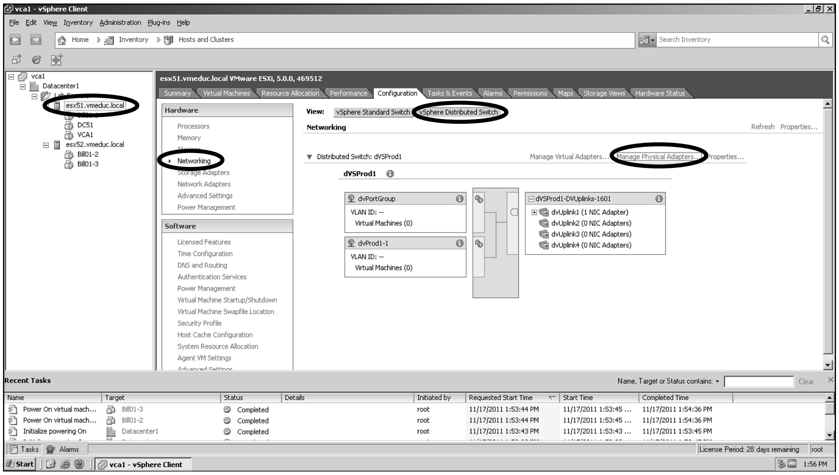
Key Topic

Activity 2-14 Adding an Uplink Adapter to a dvUplink Group

1. Log on to your vSphere Client.
2. Select Home and then **Hosts and Clusters**.

3. Select the host on which you want to configure an uplink and open the Configuration tab.

4. Choose **Networking**, click the **vSphere Distributed Switch** link, and then click the **Manage Physical Adapters** link, as shown in Figure 2-38.


vca1 - vSphere Client
File Edit View Inventory Administration Plug-ins Help
Home Inventory Hosts and Clusters
Search Inventory
vca1
Datacenter1
esx51.vmeduc.local
DC51
VCA1
esx52.vmeduc.local
Bill01-2
Bill01-3
esx51.vmeduc.local VMware ESXi, 5.0.0, 469512
Summary Virtual Machines Resource Allocation Performance Configuration Tasks & Events Alarms Permissions Maps Storage Views Hardware Status
Hardware
Processors
Memory
Networking
Storage Adapters
Network Adapters
Advanced Settings
Power Management
Software
Licensed Features
Time Configuration
DNS and Routing
Authentication Services
Power Management
Virtual Machine Startup/Shutdown
Virtual Machine Swapfile Location
Security Profile
Host Cache Configuration
System Resource Allocation
Agent VM Settings
View: vSphere Standard Switch vSphere Distributed Switch
Networking
Refresh Properties...
Distributed Switch: dVSProd1
Manage Virtual Adapters... Manage Physical Adapters... Properties...
dVSProd1
dvPortGroup
VLAN ID: --
Virtual Machines (0)
dvProd1-1
VLAN ID: --
Virtual Machines (0)
dVSProd1-DVUplinks-1601
dvUplink1 (1 NIC Adapter)
dvUplink2 (0 NIC Adapters)
dvUplink3 (0 NIC Adapters)
dvUplink4 (0 NIC Adapters)
Recent Tasks
Name, Target or Status contains:
Clear
Tasks Alarms
License Period: 28 days remaining
root
Start
vca1 - vSphere Client
1:56 PM

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Power On virtual mach...	Bill01-3	Completed		root	11/17/2011 1:53:44 PM	11/17/2011 1:53:45 ...	11/17/2011 1:54:36 PM
Power On virtual mach...	Bill01-2	Completed		root	11/17/2011 1:53:44 PM	11/17/2011 1:53:45 ...	11/17/2011 1:54:36 PM
Initialize powering On	Datacenter1	Completed		root	11/17/2011 1:53:43 PM	11/17/2011 1:53:43 ...	11/17/2011 1:53:45 PM

Figure 2-38 Adding an Uplink Adapter to a dvUplink Group

5. From Manage Physical Adapters, choose the uplinks group to which you want to add a physical adapter, and select **<Click to Add NIC>**, as shown in Figure 2-39.

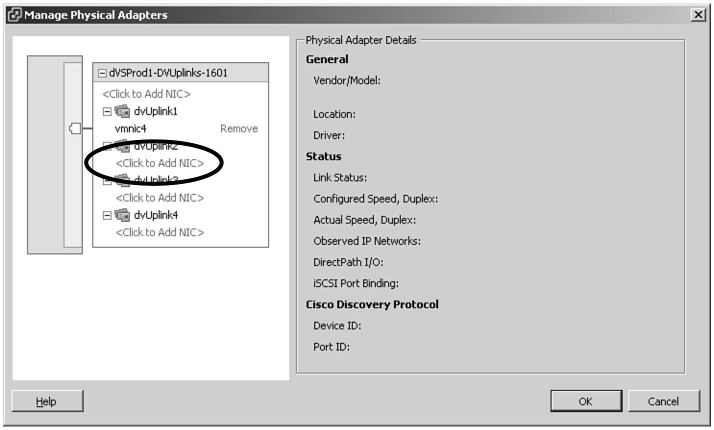
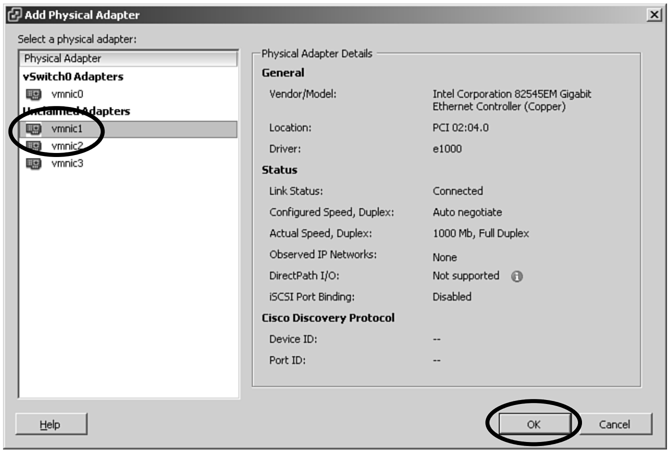


Figure 2-39 Selecting the Physical NIC

6. From Add Physical Adapter, choose the appropriate vmnic based on your network topology, and click **OK**. You can view the status information about the NIC, as shown in Figure 2-40.



Physical Adapter Details

General

Vendor/Model:	Intel Corporation 82545EM Gigabit Ethernet Controller (Copper)
Location:	PCI 02:04.0
Driver:	e1000

Status

Link Status:	Connected
Configured Speed, Duplex:	Auto negotiate
Actual Speed, Duplex:	1000 Mb, Full Duplex
Observed IP Networks:	None
DirectPath I/O:	Not supported
iSCSI Port Binding:	Disabled

Cisco Discovery Protocol

Device ID:	--
Port ID:	--

Figure 2-40 Viewing NIC Status

NOTE If you choose a vmnic that is already assigned, it will be removed from its current assignment. For this reason, take care not to remove the vmnic that is assigned to the management network that you are using to manage the host. This would be like "cutting off the limb that you are sitting on", and would cause you to lose control of the host until you could gain local access and restore the link.

When you reorganize, you might want to remove an uplink from a dvUplink group. Activity 2-15 outlines the process to remove the uplink.

Key Topic

Activity 2-15 Removing an Uplink Adapter from an dvUplink Group

1. Log on to your vSphere Client.

2. Select **Home** and then **Hosts and Clusters**.

3. Select the host on which you want to configure an uplink and open the Configuration tab.

4. Choose **Networking**, click the **vSphere Distributed Switch** link, and then click the **Manage Physical Adapters** link.

5. From the Manage Physical Adapters dialog box, choose the uplinks group to which you want to remove a physical adapter, and select **Remove** next to the uplink that you want to remove, as shown in Figure 2-41.

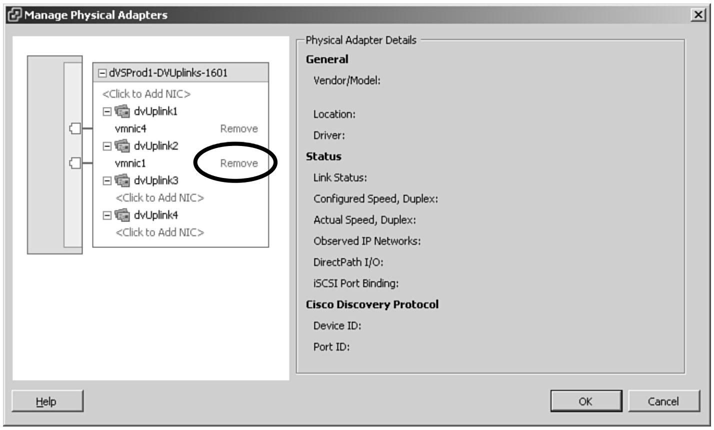


Figure 2-41 Removing an Uplink from a vDS

6. In the warning box, confirm that you want to remove the uplink by clicking **Yes**.

Creating/Configuring/Removing Virtual Adapters

Prior to vSphere 5.0 and ESXi 5.0, virtual adapters on vDSs included service console ports as well as VMkernel ports. In fact, if you are still using ESX hosts in your virtual datacenter, you must take into account that they will require a service console port on either a vSS or a vDS, for the purpose of connecting to and managing the switch from the physical world. Because ESXi 5.0 hosts do not have service consoles, they also do not have service console ports, so with regard to this topic I will limit the discussion of virtual adapters to VMkernel ports. That said, this section covers creating, configuring, and removing virtual adapters.

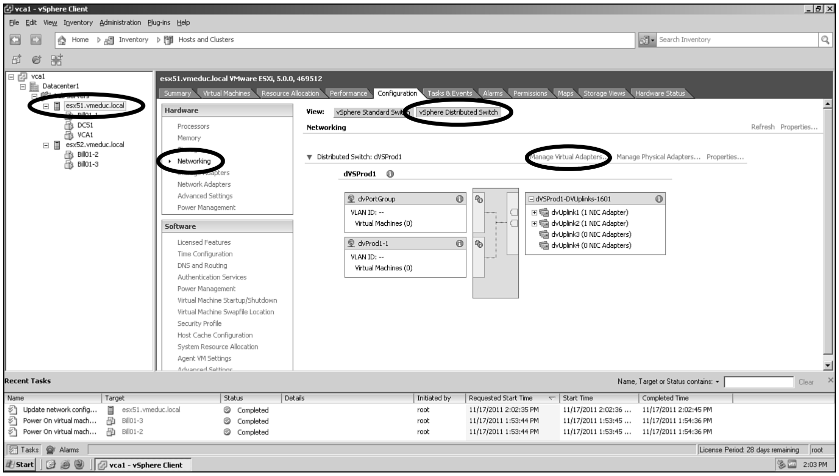
As you might remember, I discussed the fact that we create VMkernel ports for one of four reasons: IP storage, management, vMotion, or FT logging. There is only one VMkernel on the ESXi host, which is the hypervisor, but there can be many VMkernel ports. To create a new VMkernel port on a vDS, you simply create and configure a virtual adapter.

To create a virtual adapter, follow the steps outlined in Activity 2-16.

Key Topic

Activity 2-16 Creating a Virtual Adapter

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the host on which you want to configure a virtual adapter and open the Configuration tab.
4. Choose **Networking,** click the **vSphere Distributed Switch** link, and then click the **Manage Virtual Adapters** link, as shown in Figure 2-42.


vca1 - vSphere Client
File Edit View Inventory Administration Plug-ins Help
Home Inventory Hosts and Clusters
Search Inventory
vca1
Datacenter1
esx51.vmeduc.local
DC51
VCA1
esx52.vmeduc.local
Bill01-2
Bill01-3
esx51.vmeduc.local VMware ESXi, 5.0.0, 469512
Summary Virtual Machines Resource Allocation Performance Configuration Tasks & Events Alarms Permissions Maps Storage Views Hardware Status
Hardware
Processors
Memory
Networking
Network Adapters
Advanced Settings
Power Management
Software
Licensed Features
Time Configuration
DNS and Routing
Authentication Services
Power Management
Virtual Machine Startup/Shutdown
Virtual Machine Swapfile Location
Security Profile
Host Cache Configuration
System Resource Allocation
Agent VM Settings
View: vSphere Standard Switch vSphere Distributed Switch
Networking
Refresh Properties...
Distributed Switch: dVSProd1
Manage Virtual Adapters... Manage Physical Adapters... Properties...
dVSProd1
dvPortGroup
VLAN ID: --
Virtual Machines (0)
dvProd1-1
VLAN ID: --
Virtual Machines (0)
dVSProd1-DVUplinks-1601
dvUplink1 (1 NIC Adapter)
dvUplink2 (1 NIC Adapter)
dvUplink3 (0 NIC Adapters)
dvUplink4 (0 NIC Adapters)
Name, Target or Status contains:
Clear
Tasks Alarms
License Period: 28 days remaining root
Start vca1 - vSphere Client 2:03 PM

Recent Tasks

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Update network config...	esx51.vmeduc.local	Completed		root	11/17/2011 2:02:35 PM	11/17/2011 2:02:36 ...	11/17/2011 2:02:45 PM
Power On virtual mach...	Bill01-3	Completed		root	11/17/2011 1:53:44 PM	11/17/2011 1:53:45 ...	11/17/2011 1:54:36 PM
Power On virtual mach...	Bill01-2	Completed		root	11/17/2011 1:53:44 PM	11/17/2011 1:53:45 ...	11/17/2011 1:54:36 PM

Figure 2-42 The Manage Virtual Adapters Link

5. From Manage Virtual Adapters, click **Add**, as shown in Figure 2-43.

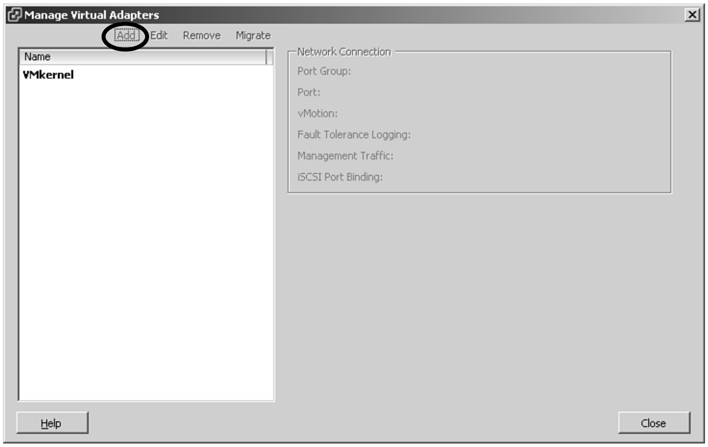
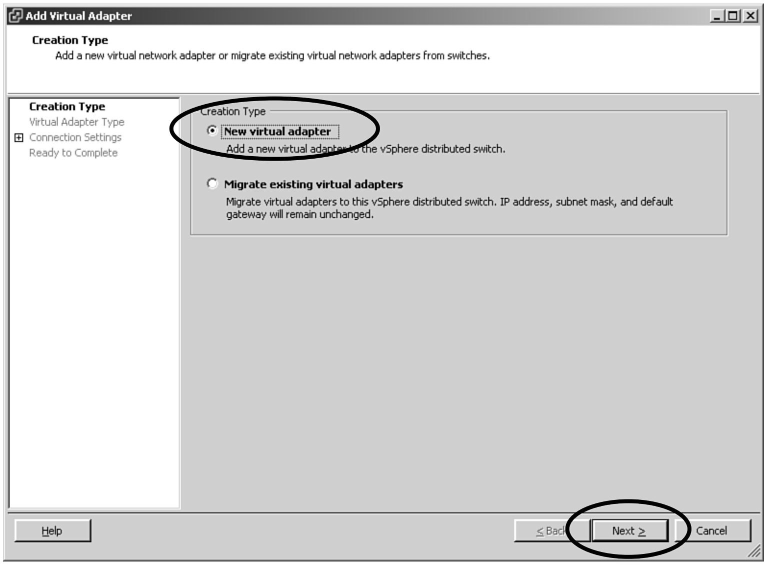


Figure 2-43 Creating a Virtual Adapter

6. From Add Virtual Adapter, choose **New Virtual Adapter** and click **Next**, as shown in Figure 2-44.



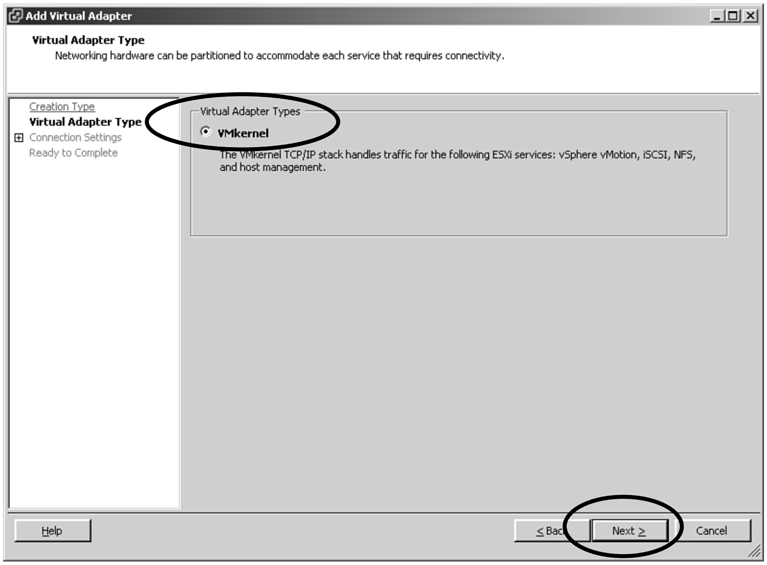
Add a new virtual network adapter or migrate existing virtual network adapters from switches.

● New virtual adapter
Add a new virtual adapter to the vSphere distributed switch.

○ Migrate existing virtual adapters
Migrate virtual adapters to this vSphere distributed switch. IP address, subnet mask, and default gateway will remain unchanged.

Figure 2-44 Adding a Virtual Adapter

7. Select **VMkernel** (the only selection on ESXi hosts) and click **Next**, as shown in Figure 2-45.

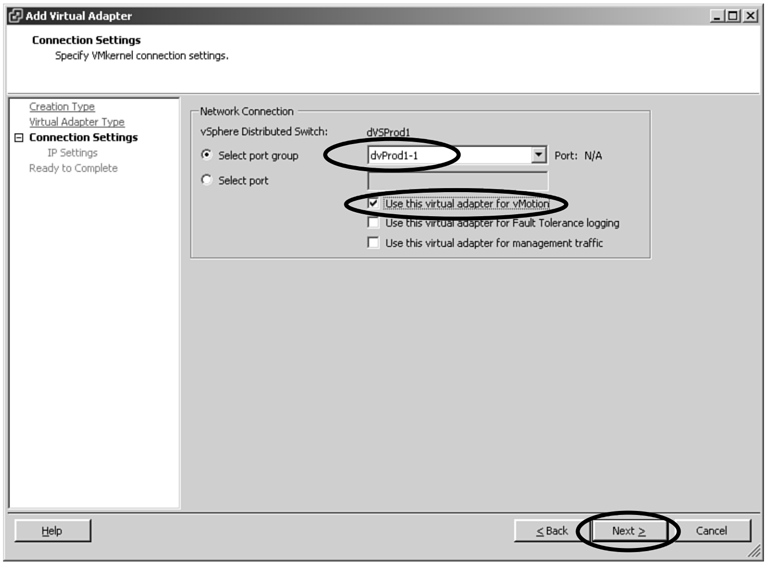


Networking hardware can be partitioned to accommodate each service that requires connectivity.

● VMkernel
The VMkernel TCP/IP stack handles traffic for the following ESXi services: vSphere vMotion, iSCSI, NFS, and host management.

Figure 2-45 Choosing Virtual Adapter Type

8. Select the port group or port to which the VMkernel port will be added and select the box indicating the function of the VMkernel port, as shown in Figure 2-46, and then click **Next**.



Add Virtual Adapter

Connection Settings

Specify VMkernel connection settings.

Network Connection

vSphere Distributed Switch: dVSProd1

● Select port group dvProd1-1 Port: N/A

○ Select port

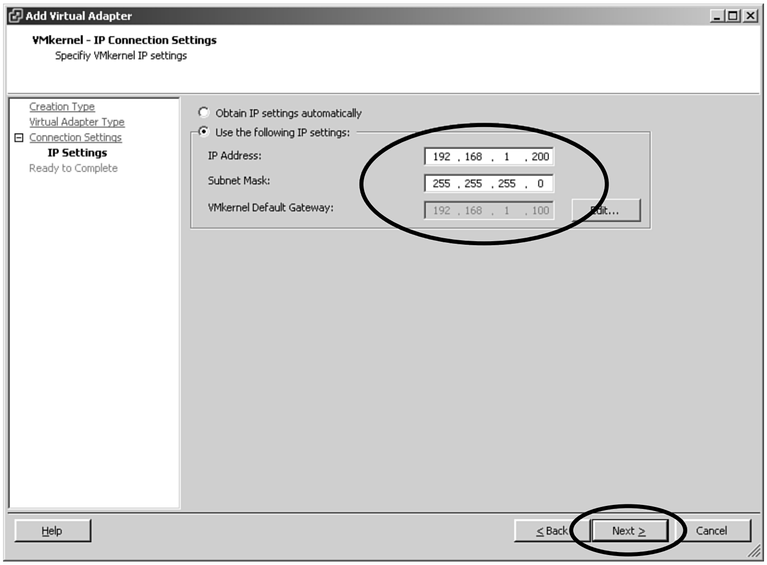
☑ Use this virtual adapter for vMotion

□ Use this virtual adapter for Fault Tolerance logging

□ Use this virtual adapter for management traffic

Figure 2-46 Connecting a Virtual Adapter to a vDS

9. Enter the IP address, subnet mask, and default gateway to be used for the VMkernel port; based on your network topology, as shown in Figure 2-47. Click **Next**. Note that once the default gateway is assigned for the first VMkernel port, the rest use the same default gateway. You can change this setting for all VMkernel ports by selecting **Edit**, but most of the time you do not need to change it.



Add Virtual Adapter

VMkernel - IP Connection Settings
Specifiy VMkernel IP settings

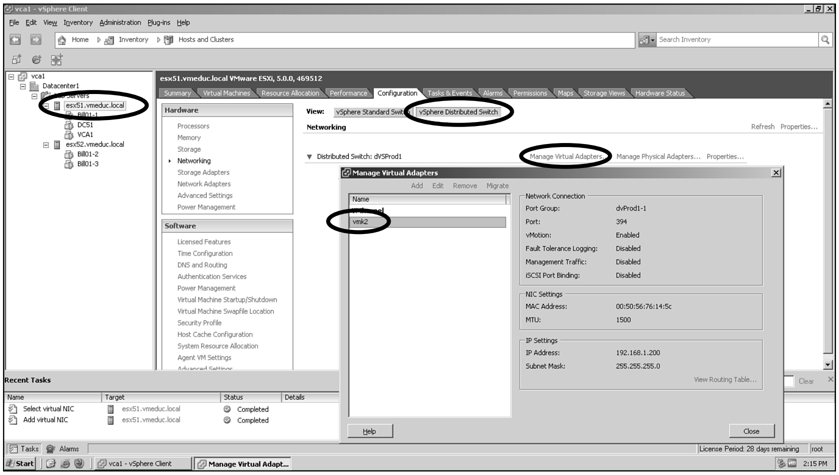
○ Obtain IP settings automatically
● Use the following IP settings:

IP Address: 192 . 168 . 1 . 200
Subnet Mask: 255 . 255 . 255 . 0
VMkernel Default Gateway: 192 . 168 . 1 . 100

Figure 2-47 Entering IP Information for Virtual Adapter

10. On Ready to Complete, confirm your selections and click **Finish**.

After you have finished configuring it, you can check the setting of your virtual adapter by coming back to Manage Virtual Adapters and clicking the adapter, as shown in Figure 2-48. To make changes to those configuration settings, you can simply elect to edit the properties of the virtual adapter.



Manage Virtual Adapters

Network Connection
Port Group: dvProd1-1
Port: 394
vMotion: Enabled
Fault Tolerance Logging: Disabled
Management Traffic: Disabled
iSCSI Port Binding: Disabled

NIC Settings
MAC Address: 00:50:56:76:14:5c
MTU: 1500

IP Settings
IP Address: 192.168.1.200
Subnet Mask: 255.255.255.0

Recent Tasks

Name	Target	Status	Details
Select virtual NIC	esx51.vmeduc.local	Completed	
Add virtual NIC	esx51.vmeduc.local	Completed	

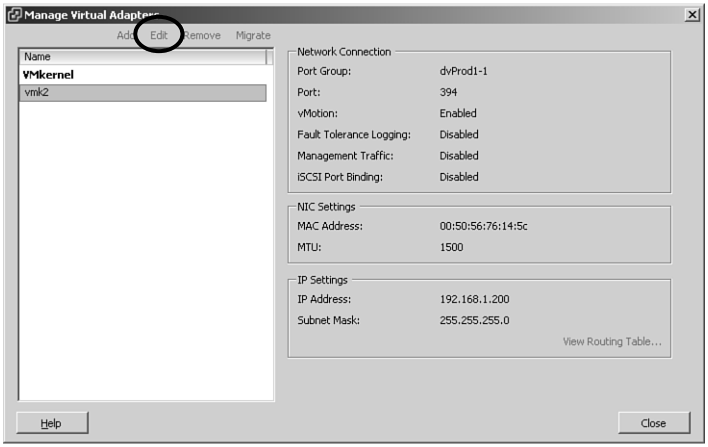
Figure 2-48 Viewing Adapter Settings

To configure a virtual adapter, follow the steps outlined in Activity 2-17.

Key Topic

Activity 2-17 Configuring a Virtual Adapter

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the host on which you want to configure a virtual adapter and open the Configuration tab.
4. Choose **Networking**, click the **vSphere Distributed Switch** link, and then click the **Manage Virtual Adapters** link.
5. From Manage Virtual Adapters, click **Edit**, as shown in Figure 2-49.



Network Connection

Port Group: dvProd1-1

Port: 394

vMotion: Enabled

Fault Tolerance Logging: Disabled

Management Traffic: Disabled

iSCSI Port Binding: Disabled

NIC Settings

MAC Address: 00:50:56:76:14:5c

MTU: 1500

IP Settings

IP Address: 192.168.1.200

Subnet Mask: 255.255.255.0

Figure 2-49 Configuring a Virtual Adapter

6. From the General tab, you can make changes to the port group or the port as well as the MTU settings (useful for jumbo frames). On the IP Settings tab, you can change the IP address, subnet mask, and default gateway if necessary.
7. Click **OK** to confirm and save all of your changes.

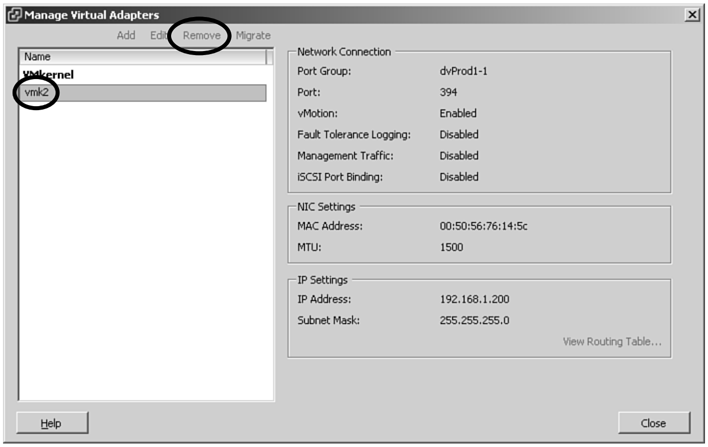
When things change and you no longer need the service that the VMkernel port was providing, you can free up the vmnic by removing it from the virtual adapter.

To remove a vmnic from a virtual adapter, follow the steps outlined in Activity 2-18.

Activity 2-18 Removing a Virtual Adapter

Key Topic

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the host on which you want to remove a virtual adapter and open the Configuration tab.
4. Choose **Networking**, click the **vSphere Distributed Switch** link, and then click the **Manage Virtual Adapters** link.
5. From Manage Virtual Adapters, click the virtual adapter that you want to remove and select **Remove**, as shown in Figure 2-50.



Network Connection
Port Group: dvProd1-1
Port: 394
vMotion: Enabled
Fault Tolerance Logging: Disabled
Management Traffic: Disabled
iSCSI Port Binding: Disabled

NIC Settings
MAC Address: 00:50:56:76:14:5c
MTU: 1500

IP Settings
IP Address: 192.168.1.200
Subnet Mask: 255.255.255.0

Figure 2-50 Removing a Virtual Adapter

6. In the warning box, confirm your selection by clicking **Yes**.

Migrating Virtual Adapters to/from a vSphere Standard Switch

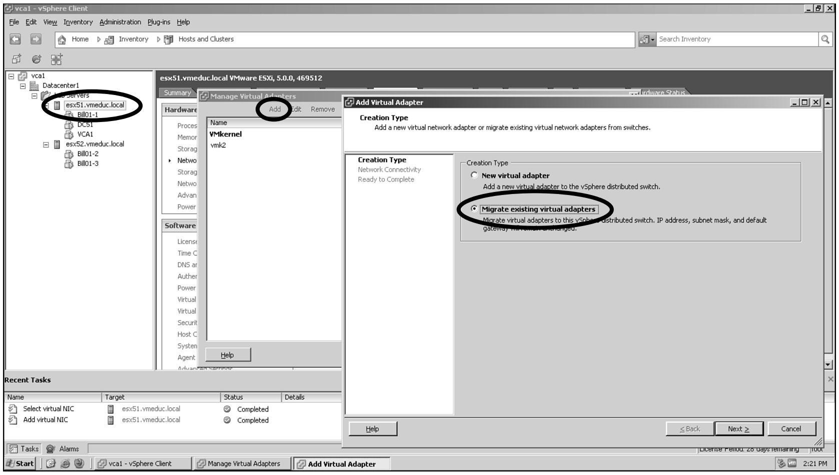
You do not necessarily have to migrate virtual adapters from your vSSs to your vDSs, but you might want to, especially if your ultimate goal is to do away with the vSS altogether. In that case, make sure that all the VMkernel ports that you have been using on your vSSs are successfully migrated to your vDSs. This section shows how you can use the tools provided by the vCenter to easily migrate VMkernel ports from vSSs to vDSs.

To migrate virtual adapters from a vSS to a vDS, follow the steps outlined in Activity 2-19.

Key Topic

Activity 2-19 Migrating Virtual Adapters from a vSS to a vDS

1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Select the host on which you want to migrate a virtual adapter and open the Configuration tab.
4. Choose **Networking**, click the **vSphere Distributed Switch** link, and then click the **Manage Virtual Adapters** link.
5. On Manage Virtual Adapters, click **Add**.
6. On Creation Type, choose **Migrate Existing Virtual Adapters**, as shown in Figure 2-51, and then click **Next**.



Add Virtual Adapter

Creation Type

Add a new virtual network adapter or migrate existing virtual network adapters from switches.

Creation Type
Network Connectivity
Ready to Complete

Creation Type

○ **New virtual adapter**

Add a new virtual adapter to the vSphere distributed switch.

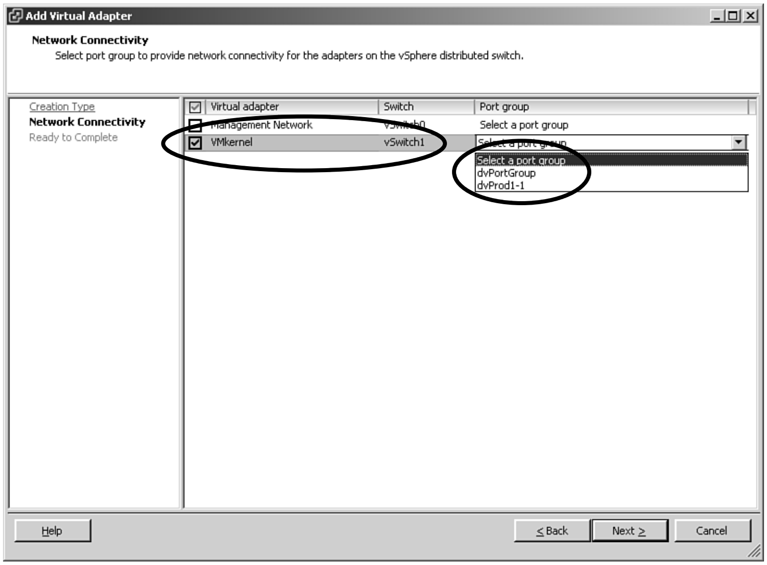
● **Migrate existing virtual adapters**

Migrate virtual adapters to this vSphere distributed switch. IP address, subnet mask, and default gateway will remain unchanged.

Name	Target	Status	Details
Select virtual NIC	esx51.vmeduc.local	Completed	
Add virtual NIC	esx51.vmeduc.local	Completed	

Figure 2-51 Migrating Virtual Adapters

7. Select the virtual adapter that you want to migrate and the port group to which you want to migrate it, as shown in Figure 2-52. Click **Next**.



Add Virtual Adapter

Network Connectivity

Select port group to provide network connectivity for the adapters on the vSphere distributed switch.

☑ Virtual adapter	Switch	Port group
☐ Management Network	vSwitch0	Select a port group
☑ VMkernel	vSwitch1	Select a port group

Figure 2-52 Choosing a Port Group

8. From Ready to Complete, review your settings and then confirm by clicking **Finish**.

NOTE If you have redundant links on both the VMkernel and the VM ports, it is possible to migrate your virtual adapters during production time with no loss of service. If you choose to go down this path, however, take great care as to the order in which you migrate the links so as to ensure users always retain the links that they need.

Migrating Virtual Machines to/from a vSphere Distributed Switch

As I mentioned earlier, the purpose of port groups is to get more than one function from a switch. In other words, port groups give you options on which to connect your VMs. You can configure different policies on port groups that are specific to the VMs that you will connect to them. In this regard, port groups on vDSs are no different from vSSs; they both give you more options for your VMs.

To help you understand the concept of migrating the VMs from a vSS to a vDS, let's pretend for a moment that the switches are physical. You walk into your network closet and you have some switches that have been there for years. They are old and noisy, and they have a limited set of features compared to new switches available today. Well, as luck would have it, you have received some money in the budget to buy a shiny new switch that has lots of features that the old noisy switches do not have. You have racked the switch and powered it up for testing, and you are now ready to start moving the cables that the computers are using from the old switch to the new switch.

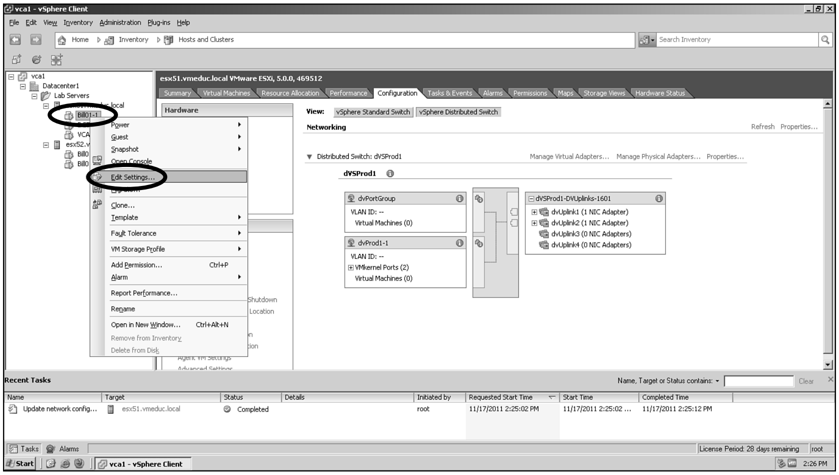
In essence, this is the opportunity that you have when you create a new vDS. You can take advantage of all of the new features of vDS, but only after you have actually moved the VMs over to the vDS. You could do this one at time, much like you would be forced to do in the physical world, but there are tools in vSphere that make it much faster and easier to move multiple VMs at the same time. In this section, I will first discuss how you would move an individual VM from a vSS to a vDS or vice versa, and then I will show you how to use the tools provided by vSphere to move multiple VMs at the same time. In both cases, the focus will be on the VM port group, which you might remember is one of the connection types that I said were very important.

To migrate a single VM to/from and vDS, follow the steps outlined in Activity 2-20.

Key Topic

Activity 2-20 Migrating a Single VM to/from a vDS

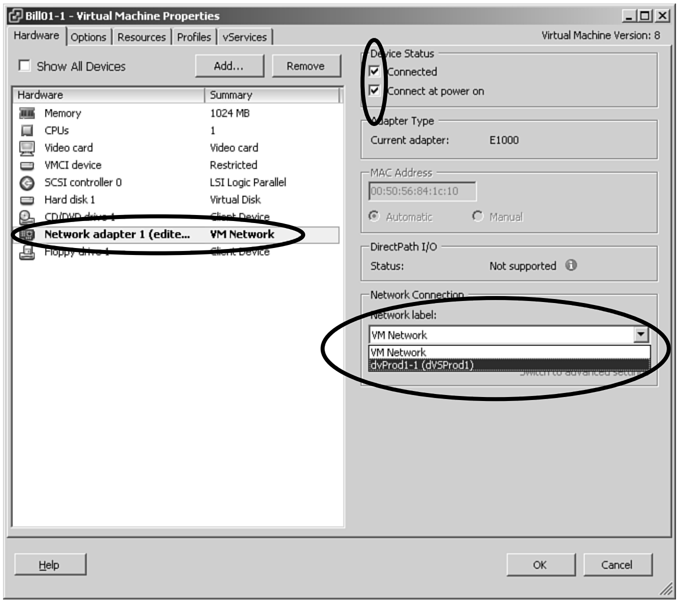
1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Right-click the VM that you want to migrate and select **Edit Settings**, as shown in Figure 2-53.



Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Update network config...	esx51.vmeduc.local	Completed		root	11/17/2011 2:25:02 PM	11/17/2011 2:25:02 ...	11/17/2011 2:25:12 PM

Figure 2-53 Migrating a VM to/from a vDS

4. Choose the network adapter that you want to migrate and the Network label (port group) to which you want to migrate it and ensure the **Connected** box in the upper-right corner is checked, as shown in Figure 2-54.



□ Show All Devices

Hardware	Summary
Memory	1024 MB
CPUs	1
Video card	Video card
VMCI device	Restricted
SCSI controller 0	LSI Logic Parallel
Hard disk 1	Virtual Disk
CD/DVD drive 1	Client Device
Network adapter 1 (edite...	**VM Network**
Floppy drive 1	Client Device

Device Status
☑ Connected
☑ Connect at power on

Adapter Type
Current adapter: E1000

MAC Address
00:50:56:84:1c:10
● Automatic ○ Manual

DirectPath I/O
Status: Not supported

Network Connection
Network label:
VM Network
VM Network
dvProd1-1 (dVSProd1)

Figure 2-54 Choosing the Port Group

5. Click **OK** to confirm and save your settings.

6. If the port group is on a different IP subnet, it might be necessary to release and renew the IP address of the VM or restart the OS.

If you only have a few VMs to move, this might be an attractive option for you and your organization. However, if you have many VMs to move, you might want a better way that will allow you to move many VMs at once.

To migrate multiple VMs from one port group to another simultaneously, follow the steps outlined in Activity 2-21.

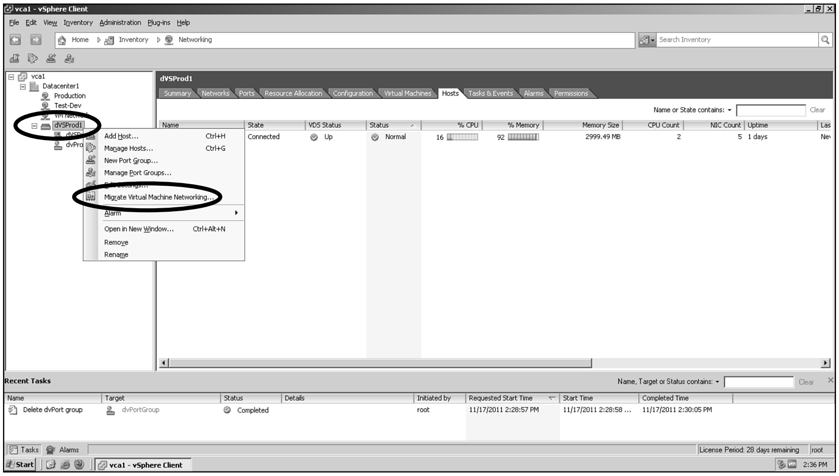
Activity 2-21 Migrating Multiple VMs Using vSphere

Key Topic

1. Log on to your vSphere Client.

2. Select **Home** and then **Networking**.

3. Right-click the vDS to which you want to migrate VMs and choose **Migrate Virtual Machine Networking**, as shown in Figure 2-55.

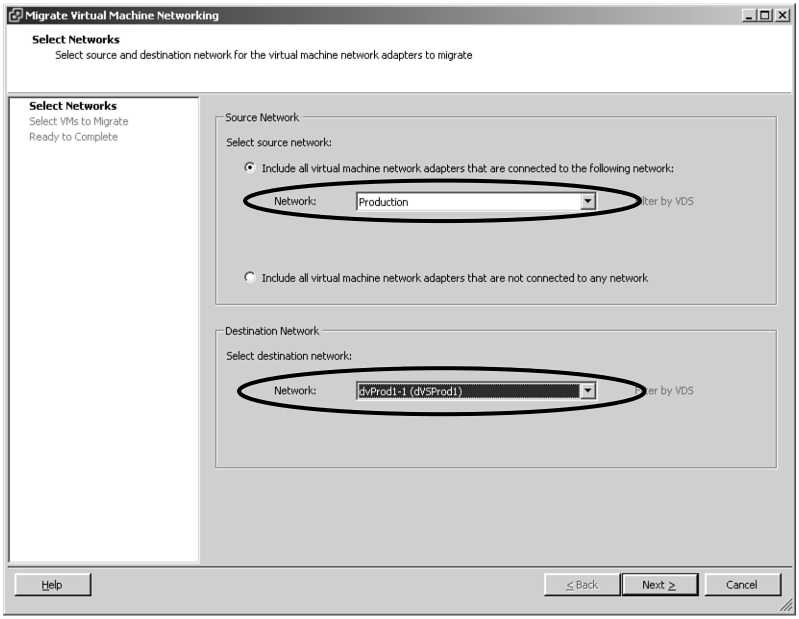


Name	State	VDS Status	Status	% CPU	% Memory	Memory Size	CPU Count	NIC Count	Uptime	Las
	Connected	Up	Normal	16	92	2999.49 MB	2	5	1 days	Ne\

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Delete dvPort group	dvPortGroup	Completed		root	11/17/2011 2:28:57 PM	11/17/2011 2:28:58 ...	11/17/2011 2:30:05 PM

Figure 2-55 Migrate Virtual Machine Networking Tool

4. From Migrate Virtual Machine Networking, choose the source network that indicates where the VMs are currently connected and the destination network to which you want to migrate them, as shown in Figure 2-56. Sources will include port groups on vSSs as well as on vDSs. (You can filter your search by network or even by vDS if you have multiple vDSs in your vSphere.)



Select Networks

Select source and destination network for the virtual machine network adapters to migrate

Source Network

Select source network:

● Include all virtual machine network adapters that are connected to the following network:

Network: Production ▾ Filter by VDS

○ Include all virtual machine network adapters that are not connected to any network

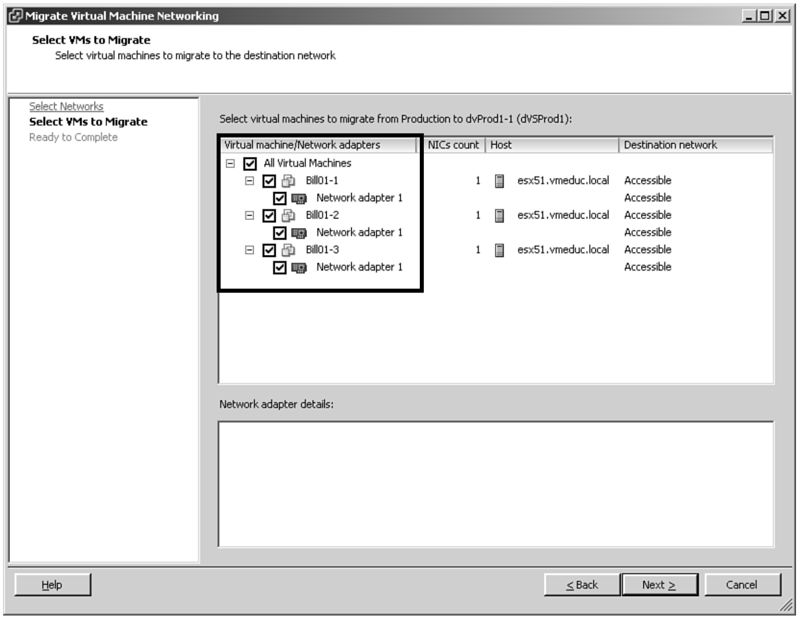
Destination Network

Select destination network:

Network: dvProd1-1 (dVSProd1) ▾ Filter by VDS

Figure 2-56 Choosing the Source and Destination

5. From Select VMs to Migrate, choose the VMs that you want to migrate from the results of the search, as shown in Figure 2-57. Click **Next**.



Migrate Virtual Machine Networking

Select VMs to Migrate

Select virtual machines to migrate to the destination network

Select Networks

Select VMs to Migrate

Ready to Complete

Select virtual machines to migrate from Production to dvProd1-1 (dVSProd1):

Virtual machine/Network adapters	NICs count	Host	Destination network
☑ All Virtual Machines			
☑ Bill01-1	1	esx51.vmeduc.local	Accessible
☑ Network adapter 1			Accessible
☑ Bill01-2	1	esx51.vmeduc.local	Accessible
☑ Network adapter 1			Accessible
☑ Bill01-3	1	esx51.vmeduc.local	Accessible
☑ Network adapter 1			Accessible

Network adapter details:

Figure 2-57 Selecting the VMs to Migrate

6. From Ready to Complete, review your settings and then confirm by clicking **Finish**.

Determining Use Cases for a vSphere Distributed Switch

As I mentioned earlier, if you decide that you are going to use a vDS in your vSphere, you first need to obtain an Enterprise Plus license. Of course, the Enterprise Plus license gives you many other features in addition to those that relate to networking, but this section focuses on networking features and ways that they might benefit a medium-sized to large-sized organization.

As outlined in Table 2-2, many features are available only on vDSs. These include features such as inbound traffic shaping, private VLANs, more granular port control for blocking, mirroring, and so on. These features can benefit your organization by giving you greater flexibility, tighter control, and enhanced security in your network. How you use them will likely vary based on what you are currently using in the physical world.

One of the nice things about this decision is that it does not have to be an "all or nothing" one. In other words, you can continue to use vSSs and begin to incorporate vDSs as well, as long as you have an Enterprise Plus license. You can leave your VMkernel ports or even service console ports (on ESX hosts) on the vSSs and use only VM port groups on the vDSs if you so desire. It's really up to you to decide what will be best for your virtual networking now and into the future and how to best use the features that VMware provides. The flexibility is there, and it's your decision as to its implementation in your virtual network and its connection to your physical network.

Configuring vSS and vDS Policies

In previous sections of this chapter, I said that we would discuss many things in greater detail later in the chapter. Well, now we are getting toward the end of this chapter, so we had better start getting into some details. In this section, I will identify common vSS and vDS policies and discuss how you can configure them on your port groups. In addition, I will discuss TCP Segmentation Offload support for VMs, jumbo frames support, and VLAN configuration.

Identifying Common vSS and vDS policies

Policies are configuration settings that enable you to customize your switches and port groups with regard to traffic control, security, and so on. In general, you can set a policy that applies to a larger network object and then "tweak" the policy to establish new settings for a smaller network object within the larger network object. The biggest difference between how this applies to vSSs versus vDSs is the network objects that are used for the large and small configurations.

With regard to vSSs, policies can be set at the switch level or they can be set at the port group level. Policies that are set at the switch level will apply to all of the ports on the switch, unless overridden by policies set the port group level. In other words, policies that are set at the port group level override any policies that are set at the switch level. This allows you to get the "best of both worlds." For example, you could set strong security policies for the switch, but then allow a "weakening" of the security policies on one port group to be used for testing and development.

Key Topic

There are three main polices for vSSs:

- Security
- Traffic shaping
- NIC teaming

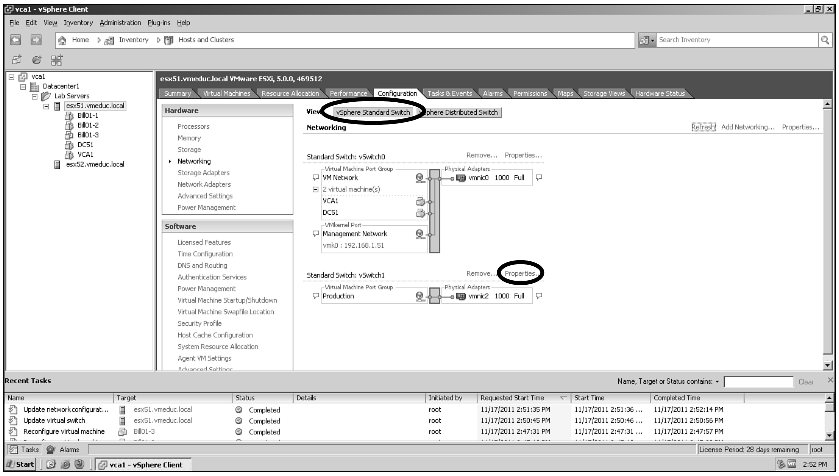
Each of these can be set at the switch level and overridden at the port group level if necessary. You can set these polices in the properties of the switch and/or port group.

To identify and configure switch and port group settings, follow the steps outlined in Activity 2-22.

Activity 2-22 Identifying Common vSS Policies

Key Topic

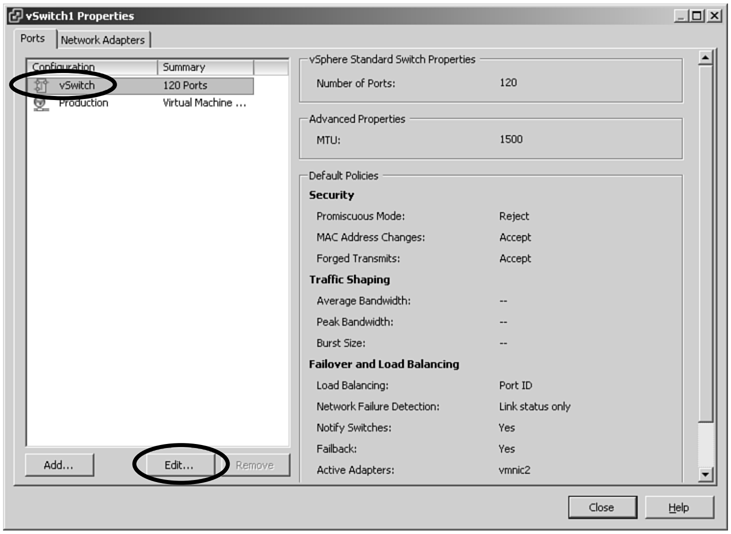
1. Log on to your vSphere Client.
2. Select **Home** and then **Hosts and Clusters**.
3. Ensure that vSphere Standard Switch is selected next to View, and then click the **Properties** link next to the switch with the policies that you want to identify and configure, as shown in Figure 2-58.



Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Update network configurat...	esx51.vmeduc.local	Completed		root	11/17/2011 2:51:35 PM	11/17/2011 2:51:36 ...	11/17/2011 2:52:14 PM
Update virtual switch	esx51.vmeduc.local	Completed		root	11/17/2011 2:50:45 PM	11/17/2011 2:50:46 ...	11/17/2011 2:50:56 PM
Reconfigure virtual machine	Bill01-3	Completed		root	11/17/2011 2:47:31 PM	11/17/2011 2:47:31 ...	11/17/2011 2:47:57 PM

Figure 2-58 Properties Link on vSS

4. Under the Configuration column, click the switch, and then click **Edit**, as shown in Figure 2-59.



Configuration	Summary
vSwitch	120 Ports
Production	Virtual Machine ...

vSphere Standard Switch Properties
Number of Ports: 120

Advanced Properties
MTU: 1500

Default Policies

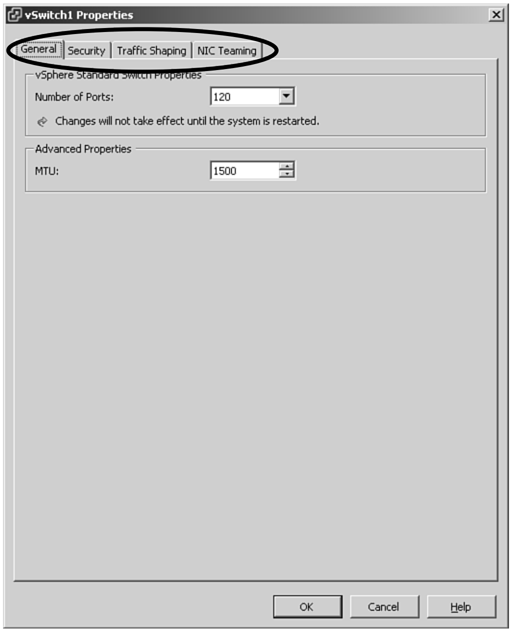
Security
Promiscuous Mode: Reject
MAC Address Changes: Accept
Forged Transmits: Accept

Traffic Shaping
Average Bandwidth: --
Peak Bandwidth: --
Burst Size: --

Failover and Load Balancing
Load Balancing: Port ID
Network Failure Detection: Link status only
Notify Switches: Yes
Failback: Yes
Active Adapters: vmnic2

Figure 2-59 Editing vSS Policies

5. Note the tabs for General, Security, Traffic Shaping, and NIC Teaming shown in Figure 2-60.



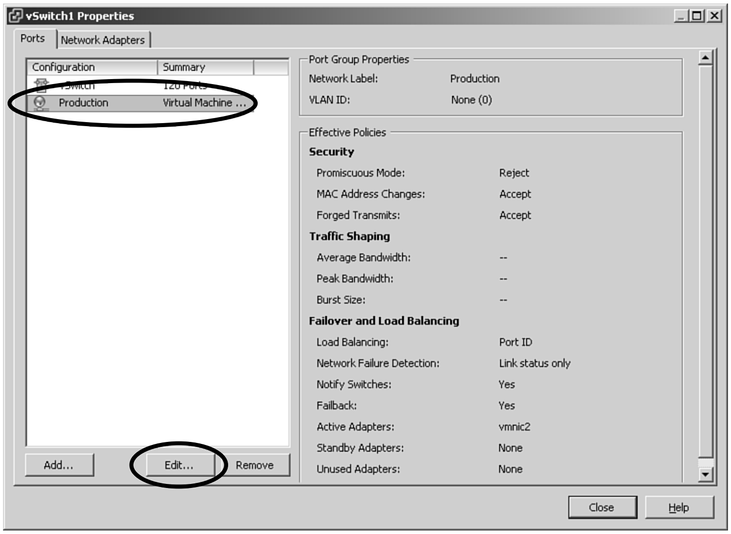
vSphere Standard Switch Properties
Number of Ports: 120
Changes will not take effect until the system is restarted.

Advanced Properties
MTU: 1500

Figure 2-60 Policies for vSS Switches and Port Groups

6. Later in this chapter, in the section "Configuring vSS and vDS policies," we discuss many of these in more detail. For now, just browse the settings, and then click **OK** or **Cancel** when you have finished.

7. Click a port group within the switch and click **Edit**, as shown in Figure 2-61.



Configuration | Summary
vSwitch | 120 Ports
Production | Virtual Machine ...

Port Group Properties
Network Label: Production
VLAN ID: None (0)

Effective Policies

Security
Promiscuous Mode: Reject
MAC Address Changes: Accept
Forged Transmits: Accept

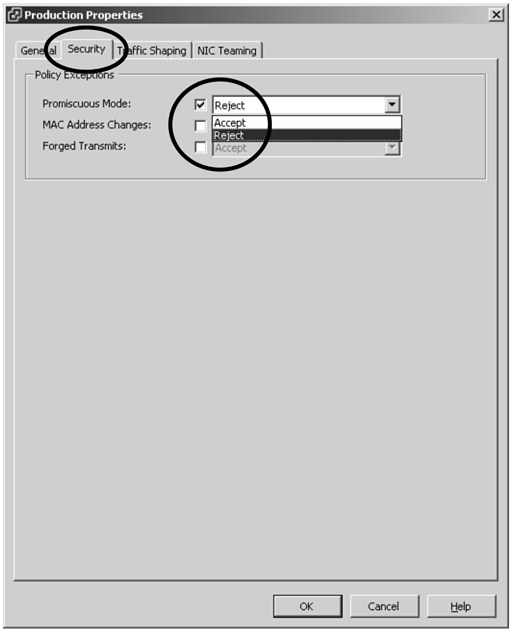
Traffic Shaping
Average Bandwidth: --
Peak Bandwidth: --
Burst Size: --

Failover and Load Balancing
Load Balancing: Port ID
Network Failure Detection: Link status only
Notify Switches: Yes
Failback: Yes
Active Adapters: vmnic2
Standby Adapters: None
Unused Adapters: None

Figure 2-61 Editing Port Group Policies

8. Note the tabs for General, IP Settings (if VMkernel), Security, Traffic Shaping, and NIC Teaming.

9. Open the Security, Traffic Shaping, and NIC Teaming tabs and note the difference, especially the white "override box." This is the setting that will cause the port group to override the switch. Figure 2-62 shows an example for Security.



Policy Exceptions

Promiscuous Mode: ☑ Reject

MAC Address Changes: □ Accept

Forged Transmits: □ Accept

Figure 2-62 Default Security Settings for a vSS and its Port Groups

10. Click **OK** or **Cancel** when you have finished.

So, now that you've seen the policies available for vSSs, you might wonder how the policies differ for vDSs. As I mentioned earlier, the main difference is between "what overrides what." As you have now seen, in vSSs most the settings are on the switch level with the port group settings occasionally overriding those of the switch. If you think about it, this cannot really apply in a vDS because the vDS could span multiple hosts (up to 350) and be connected to a huge virtual network that would have very different settings in each of its individual segments or locations. For this reason, only a few settings apply to a vDS on the switch level (which I will discuss later in the section "Configuring vSS and vDS Policies"). Instead, most policies are applied at the port group level. Now, before you start thinking that this will give you less flexibility, you should know that these policies can be overridden at the individual port level. In other words, there is even more flexibility in vDSs than there is in vSSs.

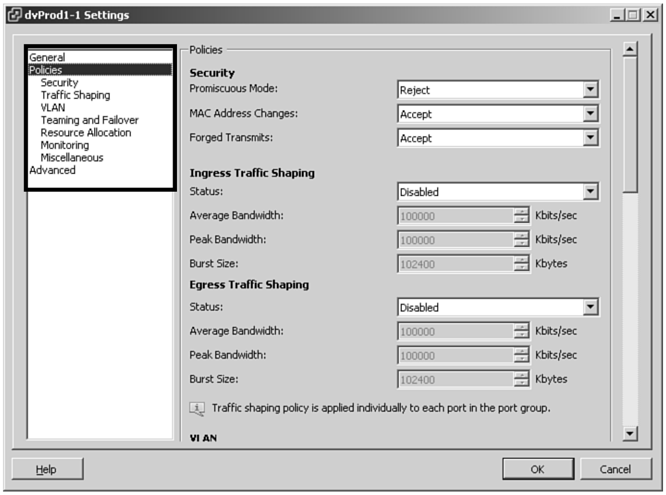
Policies that can be set at the port group level on a vDS and be overridden at the port level include Security, Traffic Shaping, VLAN, Teaming and Failover, Resource Allocation, Monitoring, Miscellaneous (port blocking), Advanced (override settings).

To identify these policy settings for a particular port group, follow the steps outlined in Activity 2-23.

Activity 2-23 Identifying Common vDS Port Group Policies

Key Topic

1. Log on to your vSphere Client.
2. Select **Home** and then **Networking**.
3. Right-click the port group that you want to examine and select **Edit Settings**.
4. Note the list of settings under the general category of Policies, as shown in Figure 2-63. Also note that the Policies category is a dialog box that gives an overview of each setting. We discuss many of these settings later in the section "Configuring vSS and vDS policies."



Policies

Security

Promiscuous Mode: Reject

MAC Address Changes: Accept

Forged Transmits: Accept

Ingress Traffic Shaping

Status: Disabled

Average Bandwidth: 100000 Kbits/sec

Peak Bandwidth: 100000 Kbits/sec

Burst Size: 102400 Kbytes

Egress Traffic Shaping

Status: Disabled

Average Bandwidth: 100000 Kbits/sec

Peak Bandwidth: 100000 Kbits/sec

Burst Size: 102400 Kbytes

Traffic shaping policy is applied individually to each port in the port group.

VLAN

Figure 2-63 Port Group Policies on a vDS

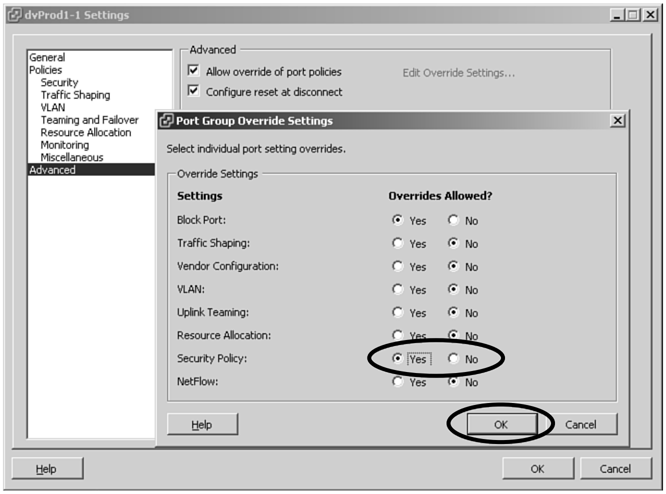
5. View each of the settings noting their features versus those of vSSs and also those settings that exist here that are not on vSSs, such as inbound (Ingress) traffic shaping.
6. When you have finished, click **OK** or **Cancel** to close.

Now you might be wondering how you can override these settings at each port. Well, it's a two-step process. First, you have to configure the port group to allow changes to a particular setting at port level and then you have to locate the port that you want to configure. For example, suppose that you wanted to configure a setting for security on a specific port that will override your security settings for the port group. In that case, follow the steps outlined in Activity 2-24.

Key Topic

Activity 2-24 Overriding vDS Port Group Policies at the Port Level

1. Log on to your vSphere Client.
2. Select **Home** and then **Networking**.
3. Right-click the port group that you want to configure and click **Edit Settings**.
4. Select **Advanced** from the bottom of the list of Policies, ensure that **Allow Override of Port Policies** is selected and click the **Edit Override Settings** link.
5. Note all the settings that you can override, and select **Yes** next to Security Policy, and then click **OK** to confirm, as shown in Figure 2-64.



Advanced

☑ Allow override of port policies

☑ Configure reset at disconnect

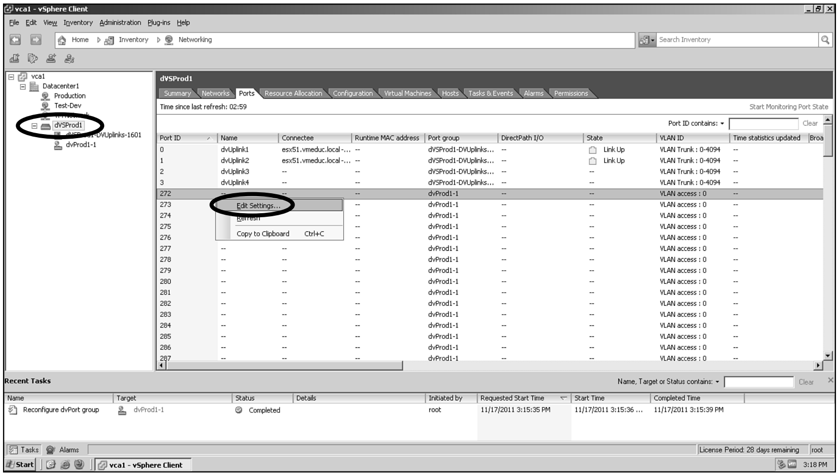
Select individual port setting overrides.

Override Settings

Settings	Overrides Allowed?
Block Port:	● Yes ○ No
Traffic Shaping:	○ Yes ● No
Vendor Configuration:	○ Yes ● No
VLAN:	○ Yes ● No
Uplink Teaming:	○ Yes ● No
Resource Allocation:	○ Yes ● No
Security Policy:	● Yes ○ No
NetFlow:	○ Yes ● No

Figure 2-64 Editing Override Settings on a vDS Port Group

6. Later in the section, "Configuring vSS and vDS Policies," we discuss many of these in more detail. For now, just browse the settings, and then click **OK** or **Cancel** when you have finished.
7. Click **OK** to change the port group settings.
8. In the console pane (on the left), click the vDS and then on the Ports tab, right-click the individual port that you want to configure, and click **Edit Settings**, as shown in Figure 2-65.


vca1 - vSphere Client
File Edit View Inventory Administration Plug-ins Help
Home Inventory Networking
Search Inventory
vca1
Datacenter1
Production
Test-Dev
dVSProd1
dVSProd1-DVUplinks-1601
dvProd1-1
dVSProd1
Summary Networks Ports Resource Allocation Configuration Virtual Machines Hosts Tasks & Events Alarms Permissions
Start Monitoring Port State
Port ID contains:
Clear
Edit Settings...
Refresh
Copy to Clipboard Ctrl+C
Recent Tasks
Name, Target or Status contains:
Clear
Tasks Alarms
Start
vca1 - vSphere Client
3:18 PM

Time since last refresh: 02:59

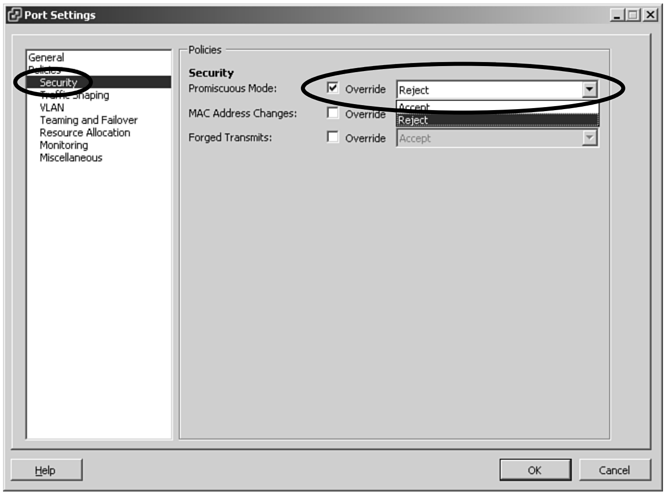
Port ID	Name	Connectee	Runtime MAC address	Port group	DirectPath I/O	State	VLAN ID	Time statistics updated	Broa
0	dvUplink1	esx51.vmeduc.local -...	--	dVSProd1-DVUplinks...	--	Link Up	VLAN Trunk : 0-4094	--	
1	dvUplink2	esx51.vmeduc.local -...	--	dVSProd1-DVUplinks...	--	Link Up	VLAN Trunk : 0-4094	--	
2	dvUplink3	--	--	dVSProd1-DVUplinks...	--	--	VLAN Trunk : 0-4094	--	
3	dvUplink4	--	--	dVSProd1-DVUplinks...	--	--	VLAN Trunk : 0-4094	--	
272	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
273			--	dvProd1-1	--	--	VLAN access : 0	--	
274			--	dvProd1-1	--	--	VLAN access : 0	--	
275			--	dvProd1-1	--	--	VLAN access : 0	--	
276			--	dvProd1-1	--	--	VLAN access : 0	--	
277	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
278	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
279	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
280	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
281	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
282	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
283	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
284	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
285	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
286	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	
287	--	--	--	dvProd1-1	--	--	VLAN access : 0	--	

Name	Target	Status	Details	Initiated by	Requested Start Time	Start Time	Completed Time
Reconfigure dvPort group	dvProd1-1	Completed		root	11/17/2011 3:15:35 PM	11/17/2011 3:15:36 ...	11/17/2011 3:15:39 PM

License Period: 28 days remaining | root

Figure 2-65 Editing Port Settings on a vDS Port Group

9. From Port Settings, select **Security**, and then select **Override** next to the security settings that you want to change for this port only. Change to your desired settings, as shown in Figure 2-66.



Security

Promiscuous Mode: ☑ Override Reject

MAC Address Changes: ☐ Override

Forged Transmits: ☐ Override Accept

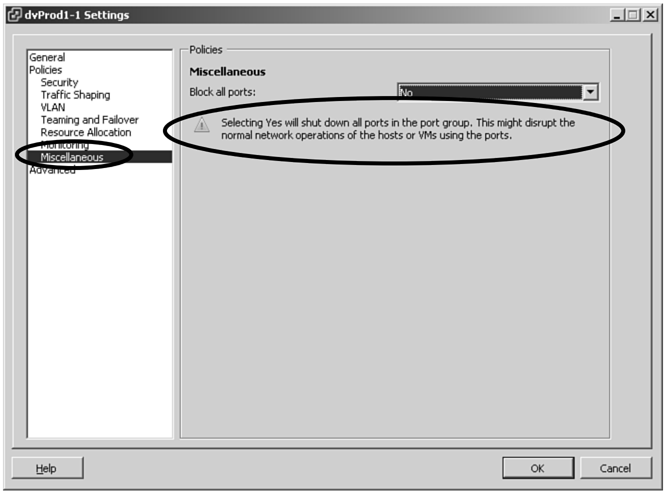
Figure 2-66 Override Settings for Ports on a vDS Port Group

10. Click **OK** to confirm and save your changes.

NOTE To stay as close the test blueprint as possible, I am focusing my configuration discussion on those topics that are specified on the blueprint.

Configuring dvPort Group Blocking Policies

You might have noticed that I included Miscellaneous in the list of port group policies and specified that it involves port group blocking. That is because that's what it says on the dialog box. Interestingly enough, as shown in Figure 2-67, it also says "Selecting Yes will shut down all ports in a port group. This might disrupt the normal network operations of the hosts or VMs using the ports." Gee, ya think?



dvProd1-1 Settings

Policies

Miscellaneous

Block all ports: No

Selecting Yes will shut down all ports in the port group. This might disrupt the normal network operations of the hosts or VMs using the ports.

Figure 2-67 Configuring dvPort Group Blocking

Based on this, why would anyone want to select Yes? Well, this isn't your everyday setting. It's more of a "one-off" scenario setting that can come in handy if you know how to use it. Suppose that you do want to isolate all machines on a port group from the network for a period of time while you are making a software change. After the change, you want to connect them all again.

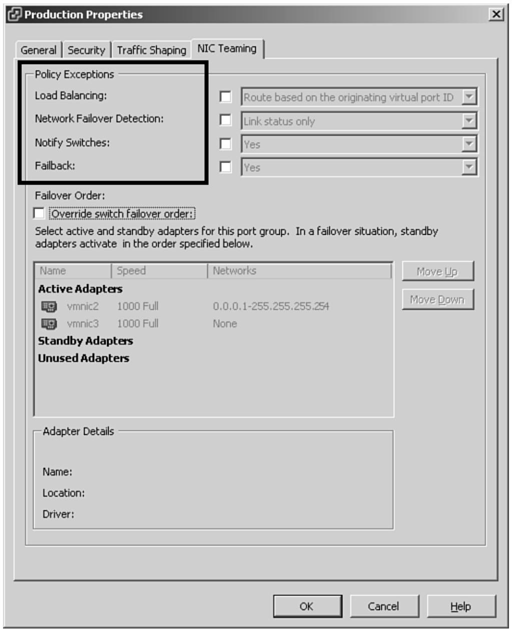
You could remove the vmnics from the port group, but what about any internal links? Also, you could disconnect each of the vNICs on the individual VMs, but what if there are many VMs, and what if you miss one or two of them? With the option of dvPort port group blocking, you can throw a "master switch" that disables networking for all VMs on that port group, no matter where they are connected. Before you throw that switch, though, make sure that you are on the right port group and make sure that the VMs that are on the port group are the ones that you want to isolate!

Configuring Load Balancing and Failover Policies

If you assign more than one vmnic (physical NIC) to a switch or port group, you can configure load balancing and failover policies using the vmnics that you assign. This is the concept of *NIC teaming*, which you should clearly understand is not using more than one vNIC on a VM, but instead using more than one vmnic on a switch or port group. In this section, I discuss configuring load balancing and failover policies, first on vSSs and then on vDSs.

On a vSS, as you might remember, NIC teaming is one of the three policies that you can configure at the switch level or at the port group level. As discussed, any policy setting that you configure at the port group level will override the settings at the switch level. So, now I will discuss the policies that you can configure at the switch level and override at the port group level.

On the NIC Teaming tab of vSS, or a port group on a vSS, you will find a list of policy exceptions, as shown in Figure 2-68. They are called *exceptions* because they each have a default setting, but that setting can be changed if necessary. I will now discuss each of these settings and the options that you have from which to choose.



Policy Exceptions

Load Balancing: □ Route based on the originating virtual port ID

Network Failover Detection: □ Link status only

Notify Switches: □ Yes

Failback: □ Yes

Failover Order:

□ Override switch failover order:

Select active and standby adapters for this port group. In a failover situation, standby adapters activate in the order specified below.

Name	Speed	Networks
Active Adapters		
vmnic2	1000 Full	0.0.0.1-255.255.255.254
vmnic3	1000 Full	None
Standby Adapters		
Unused Adapters		

Adapter Details

Name:

Location:

Driver:

Figure 2-68 Policy Exceptions on dvPort Groups

Load Balancing

There are four load balancing options from which you can choose:

- **Route based on the originating virtual port ID:** The physical NIC is determined by the ID of the virtual port to which the VM is connected. This option has the lowest overhead and is the default option for vSSs and port groups on vSSs.
- **Route based on source MAC hash:** All of each VM's outbound traffic is mapped to specific physical NIC that is based on the MAC address associated with the VM's virtual network interface card (vNIC). This method has relatively low overhead and is compatible with all switches; even those that do not support the 802.3ad protocol.
- **Route based on IP hash:** The physical NIC for each outbound packet is chosen based on a hash of the source and destination addresses contained in the packet. This method has the disadvantage of using more CPU resources; however, it can provide better distribution of traffic across the physical NICs. This method also requires the 802.3ad link aggregation support or EtherChannel on the switch.
- **Use explicit failover order:** The switch will always choose from its list of active adapters the highest order uplink that is not currently in use.

You should make these choices based on your virtual networking needs and based on how your virtual network connects to your physical network.

Network Failover Detection

As I discussed earlier, one of the reasons that you might want to assign more than one vmnic to a switch or port group is that you will have redundancy so that if one physical NIC fails another one can take over. That said, how will you know whether your redundancy is still intact? The following are your two options with regard to network failure detection and a brief description of each option:

- **Link Status Only:** This option relies solely on the link status that the network adapter provides. In other words, "Do I feel electricity?" or with fiber "Do I see a light?" This option detects cable pulls and physical switch failures, but it does not detect configuration errors that are beyond the directly connected switch. This method has no overhead and is the default.

- **Beacon Probing:** This option listens for link status but also sends out beacon packets from each physical NIC that it expects to be received on the other physical NIC. In this way, physical issues can be detected as well as configuration errors such as improper settings on Spanning Tree Protocol (STP) or VLANs. Also, you should not use Beacon Probing with IP-hash load balancing because the way the beacon traffic is handled does not work well with this option and can cause a "network flapping" error.

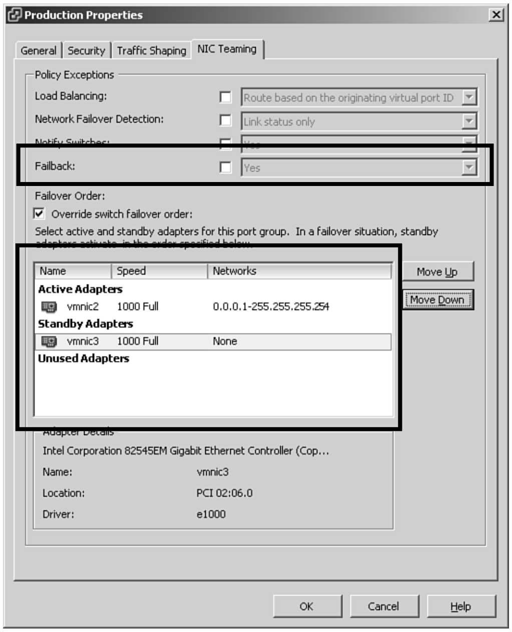
Notify Switches

The main job of a physical switch is to learn the MAC addresses of the computers and other devices on the network to which it is connected. If these change, its job is to make the change in its MAC address table. In most cases, you want to notify the physical switch of any changes in your virtual network that affect the MAC address table of the physical switch, but not always. The following are your two simple options with regard to notifying switches:

- **Yes:** If you select this option, the switch is notified whenever a VM's traffic will be routed over a different physical NIC because of a failover event. In most cases, this is the setting that you want to configure because it offers the lowest latency for failover occurrence and for vMotion migrations.
- **No:** If you select this option, the switch will not be notified and will not make the changes to its MAC address table. You should only select this option if you are using Microsoft Network Load Balancing (NLB) in unicast mode because a selection of **Yes** prevents the proper function of Microsoft Network Load Balancing in unicast mode

Failback

On each switch and/or port group, you can assign vmnics (physical NICs) as Active, Standby, or Unused, as shown in Figure 2-69.



Policy Exceptions

Load Balancing:	□	Route based on the originating virtual port ID
Network Failover Detection:	□	Link status only
Notify Switches:	□	Yes
Failback:	□	Yes

Failover Order:

☑ Override switch failover order:

Select active and standby adapters for this port group. In a failover situation, standby adapters activate in the order specified below.

Name	Speed	Networks
Active Adapters		
vmnic2	1000 Full	0.0.0.1-255.255.255.254
Standby Adapters		
vmnic3	1000 Full	None
Unused Adapters		

Adapter Details

Intel Corporation 82545EM Gigabit Ethernet Controller (Cop...

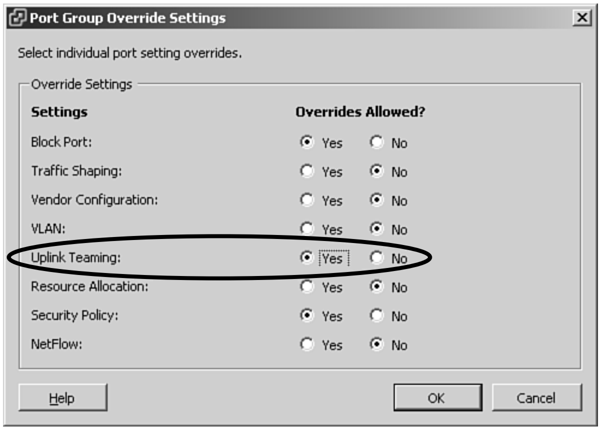
Name:	vmnic3
Location:	PCI 02:06.0
Driver:	e1000

Figure 2-69 Active/Standby NICs and Failback

If a vmnic is listed as Active, it will be used unless it fails. If a vmnic fails, the first vmnic in the Standby list will be used. Now, what if the first vmnic should come back? Then, should you go back to it immediately or should you stay with the vmnic that is currently working fine? The following are your two simple options for Failback settings:

- **Yes:** If you select this option, a failed adapter that has recovered will be returned to Active status immediately after its recovery, thereby replacing the vmnic that is working fine. This might be an advantage if the primary adapter is somehow superior to the secondary one, such as a faster speed or other features. The disadvantage of this option is that a "flapping" connection could cause the system to play "ping pong" with itself, continually changing between adapters.
- **No:** If you select this option and an adapter fails and then recovers, the adapter that took its place when it failed will continue to be used. This "if it ain't broke, don't fix it" approach avoids the "ping pong" of the other option, but might leave the traffic on a slower or less desirable adapter.

On a vDS, many of the settings are reasoned in the same way, but the dialog boxes are a little different. In addition, as I mentioned before, the settings are typically configured at the port group level and can be overridden at the port level. As you might recall from earlier in this chapter, Uplink Teaming is one of the Policy settings that you can edit for port groups, as shown in Figure 2-70.



Port Group Override Settings

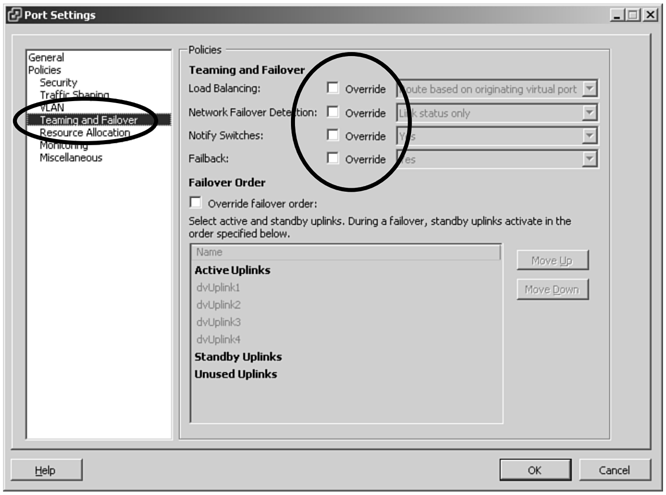
Select individual port setting overrides.

Override Settings

Settings	Overrides Allowed?	
Block Port:	● Yes	○ No
Traffic Shaping:	○ Yes	● No
Vendor Configuration:	○ Yes	● No
VLAN:	○ Yes	● No
Uplink Teaming:	● Yes	○ No
Resource Allocation:	○ Yes	● No
Security Policy:	● Yes	○ No
NetFlow:	○ Yes	● No

Figure 2-70 Uplink Teaming Override Settings on dvPort Groups

Once you are on the right dialog box, you will notice that the settings are exactly the same, except that they can be overridden at the individual port level, as shown on Figure 2-71.



Port Settings

Policies

Teaming and Failover

Load Balancing: □ Override Route based on originating virtual port

Network Failover Detection: □ Override Link status only

Notify Switches: □ Override Yes

Failback: □ Override Yes

Failover Order

□ Override failover order:

Select active and standby uplinks. During a failover, standby uplinks activate in the order specified below.

Name

Active Uplinks

dvUplink1

dvUplink2

dvUplink3

dvUplink4

Standby Uplinks

Unused Uplinks

Figure 2-71 Override Settings for Uplink Teaming at Port Level

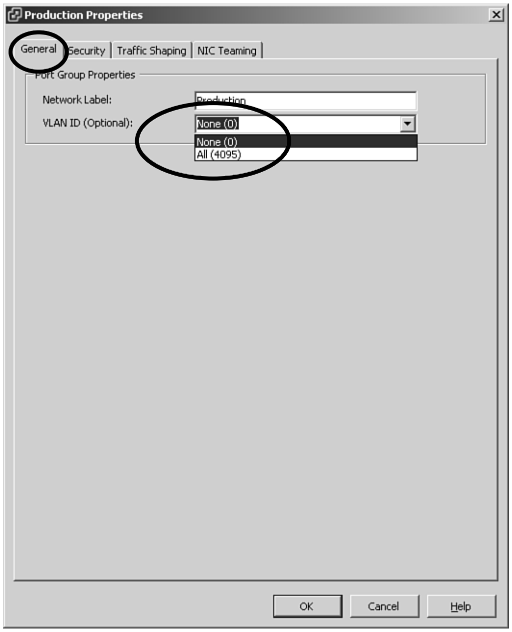
Configuring VLAN Settings

Virtual local-area networks (VLANs) are commonly used in today's networks to create and manage subnets in networks that contain many switches. They offer a high degree of flexibility and security and are useful for carrying many subnets on one or a few cables using a packet marking method called tagging. vSphere fully supports IEEE 802.1Q tagging.

Because this is not a Cisco discussion, or a Cisco test, you don't need to know all of the details of VLAN configuration, but you should know how to configure your port group properties or individual port properties to work with the VLANs that you already have in your organization. The bottom line is that if you want to bring more subnets in and out of your virtual network than you want to use physical NICs to carry, you will need to use VLANs and 802.1Q tagging. VLANs will also give you the flexibility to use the load balancing and fault tolerance options of which we've spoken, in more creative ways.

I will first discuss the configuration of VLANs on a vSS and then on a vDS. For each type of switch, I will cover your options and the impact of your decisions. In addition, I will discuss the advantages of using a vDS versus a vSS with regard to VLANs.

On vSS port groups, you can configure the VLAN setting on the General tab of the properties for the port group as shown in Figure 2-72. You can do so by typing the VLAN number from your network in the box labeled VLAN ID (Optional). If you have VMs that need to receive packets from more than one subnet and provide their own tagging for more than one subnet, you should select **All** (4095).



Port Group Properties

Network Label: Production

VLAN ID (Optional): None (0)

None (0)
All (4095)

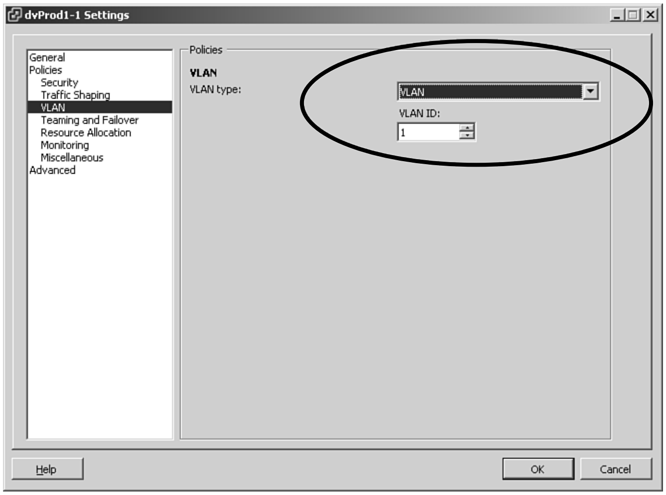
Figure 2-72 VLAN Configuration on a vSS Port Group

NOTE The All (4095) setting configures the port group to receive all VLANs (0–4094). It is only necessary if a VM that is on the port group is actually creating its own tagging and needs to be connected to port groups on other VLANs as well. The All (4095) setting is not necessary to establish a trunk between the vSS and the physical switch; however, the interface on the physical switch should be set to trunk. The All (4095) setting is rarely used.

On vDS port groups, you can configure the VLAN in a much more granular fashion. You might have noticed that the VLAN setting is one of options under Polices. On this setting, you have three options from which to choose: VLAN, VLAN Trunking, and Private VLAN. In this section, I will discuss each of these options briefly and illustrate how you would configure them.

Configuring VLAN Policy Settings on a VDS

If you select VLAN, the screen changes and you are presented with a simple box in which you can input a number, as shown in Figure 2-73. This number should be an actual VLAN number that you are using on your physical network and that you want to incorporate into your virtual network as well, on this port group. Your range of choices is from 1– 4094.



dvProd1-1 Settings

Policies

VLAN

VLAN type: VLAN

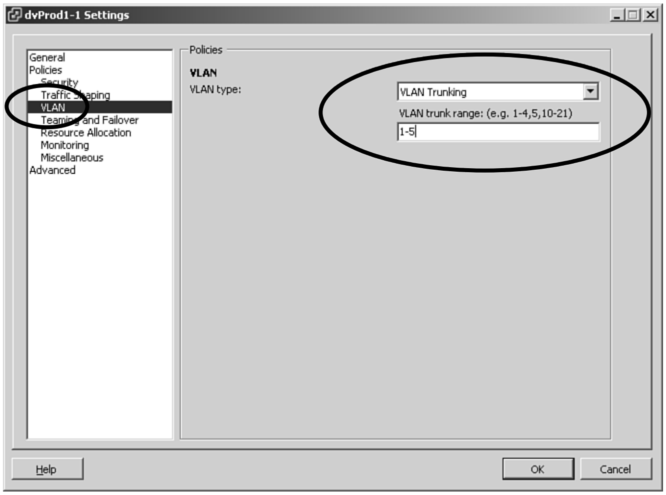
VLAN ID: 1

Figure 2-73 VLAN Configuration on a vDS

NOTE VLAN 1 is often used as a management VLAN for management traffic, including CDP, so it may not be a valid choice in your network.

Configuring VLAN Trunking Policies on a VDS

This option establishes the port group as a trunk that can carry multiple VLANs to VMs that are connected to it. However, rather than having to carry all 4094 VLANs just to have more than one, on vDSs this setting can be pruned to carry only the VLANs or range of VLANs that you specify, as shown in Figure 2-74.



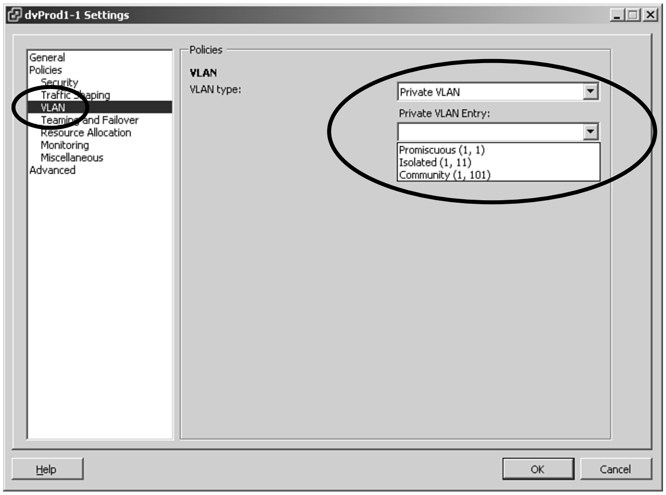
VLAN type: VLAN Trunking

VLAN trunk range: (e.g. 1-4,5,10-21) 1-5

Figure 2-74 VLAN Trunking Configuration on a vDS

Configuring Private VLAN Policy Settings on a vDS

This setting, shown in Figure 2-75, allows you to use a VLAN that you have created on the vDS that can only be used by your vSphere environment and not by your external network.



VLAN

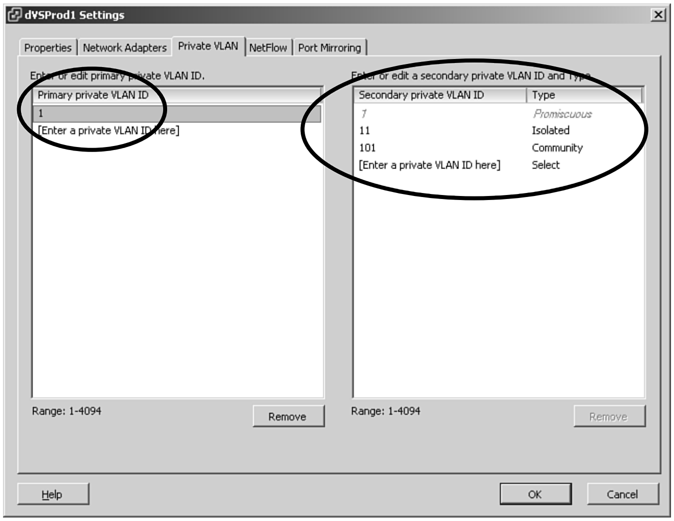
VLAN type: Private VLAN

Private VLAN Entry:

Promiscuous (1, 1)
Isolated (1, 11)
Community (1, 101)

Figure 2-75 BIOS Chips and CMOS Batteries on Typical Motherboards

To create a private VLAN, you in essence further segment a VLAN that you are already receiving into the switch. You must first create these on the vDS, as shown in Figure 2-76.



Enter or edit primary private VLAN ID.

Primary private VLAN ID
1
[Enter a private VLAN ID here]

Range: 1-4094

Enter or edit a secondary private VLAN ID and Type.

Secondary private VLAN ID	Type
1	Promiscuous
11	Isolated
101	Community
[Enter a private VLAN ID here]	Select

Range: 1-4094

Figure 2-76 Private VLAN Creation on a vDS

There are three types of private VLANs that you can create and use in your vSphere:

- **Promiscuous:** This is named (numbered) by the primary VLAN that you chose from your physical network. It is the remaining piece that is not separated from the primary VLAN. VMs on this VLAN are reachable and can be reached by any VM in the same primary VLAN.
- **Isolated:** This is a private VLAN used to create a separate network for one VM in your virtual network that is not used at all in physical world. It can be used to isolate a highly sensitive VM, for example. If a VM is in an isolated VLAN, it will not communicate with any other VMs in other isolated VLANs or in other community VLANs. It can communicate with promiscuous VLANs.
- **Community:** This a private VLAN used to create a separate network to be shared by more than one VM. This VLAN is also only used in your virtual network and is not used in your physical network. VMs on community VLANs can communicate only to other VMs on the same community or to VMs on a promiscuous VLAN.

NOTE To use private VLANs between your host and the rest of your physical network, the physical switch connected to your host needs to be private VLAN capable and configured with the VLAN IDs being used by ESXi for the private VLAN functionality. The precise configuration of your physical switch is beyond the scope of this book.

Configuring Traffic Shaping Policies

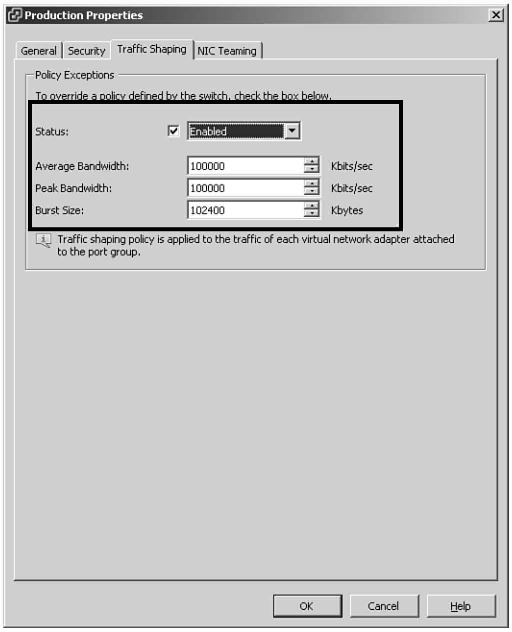
By default, all of the VMs on a port group have an unlimited share of the bandwidth assigned to that port group, and all of the port groups have an unlimited share of the bandwidth that is provided by the uplinks on the virtual switch. This is true on vSSs and on vDSs. In other words, by default, it is an "all you can eat buffet" for all of the VMs on the switch, regardless of which port group.

When you decide to use traffic shaping, your goal should be to free up available bandwidth by limiting the bandwidth usage on port groups that contain VMs that can function with less bandwidth. This might not be as straightforward as it first seems. You might be thinking that there are some "bandwidth hogs" on your network that you want to traffic shape right away. Well, if those have anything to do with Voice over Internet Protocol (VoIP) or video, you might want to reconsider your options. In fact, you might want to traffic shape port groups that hold VMs that are file and print servers first, because they can take the bandwidth reduction hit, and thereby give the VoIP and video VMs more available bandwidth. That said, traffic shaping should never be done without first studying your virtual network to determine what you want to accomplish and to find out if you have the resources to accomplish it.

Your options for traffic shaping are very different on vSSs versus vDSs. In addition, the tools that you use to configure them are very different as well. In this section, I will first discuss your traffic shaping options on vSSs and then I will examine your additional options on vDSs.

Traffic Shaping Policies for vSphere Standard Switches

On vSSs, all traffic shaping is for outbound traffic only. This is the case regardless of which version vSS you are using. You might have heard that inbound traffic shaping is available in vSphere. This is true, but only with vDS port groups (which I will discuss next). As with other policies, you can configure traffic shaping on a vSS at the switch level and then you can override it at the port group level. After you enable traffic shaping, you can configure three main settings for outbound traffic on vSS switches and port groups, as shown with their default settings (not configured yet) in Figure 2-77.



Policy Exceptions

To override a policy defined by the switch, check the box below.

Status: ☑ Enabled

Average Bandwidth: 100000 Kbits/sec

Peak Bandwidth: 100000 Kbits/sec

Burst Size: 102400 Kbytes

Traffic shaping policy is applied to the traffic of each virtual network adapter attached to the port group.

Figure 2-77 Traffic Shaping on vSS

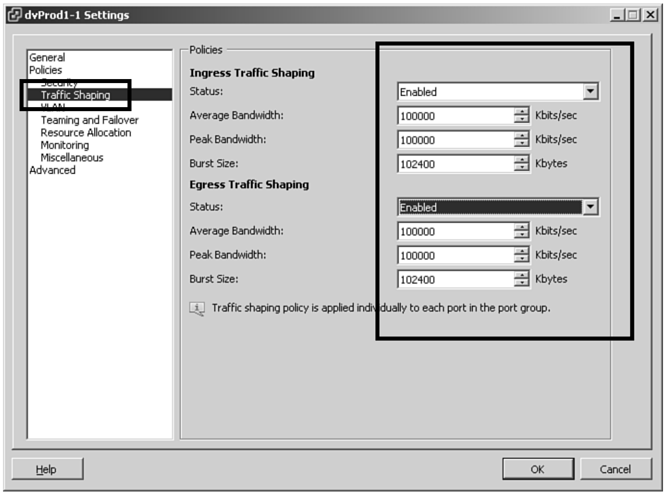
Key Topic

The following is a brief description of each of these settings:

- **Average Bandwidth:** This establishes the number of kilobits per second to allow across a port, averaged over time. It should be an amount based on that which you have observed or monitored in the past.
- **Peak Bandwidth:** This is the maximum aggregate traffic measured in kilobits per second that will be allowed for a port group or switch. It should be an amount that will not hamper the effective use of the VMs connected to the port group or switch.
- **Burst Size:** This is maximum number of bytes to be allowed in a burst. A burst is defined as exceeding the average bandwidth. This setting determines how long the bandwidth can exceed the average as a factor of how far it has exceeded the average. The higher it goes, the less time it can spend there. In other words, this setting is a factor of "bandwidth X time".

Traffic Shaping Policies for vSphere Distributed Switches

On vDSs, traffic shaping can be configured for the port group and overridden if necessary at the individual port level; just as with other policies. The biggest difference from that of vSSs being that it can be configured for both inbound (ingress) and outbound (egress) traffic. You might have noticed that traffic shaping is listed under the policies of a vDS port group and/or individual port, as shown in Figure 2-78.



Ingress Traffic Shaping

Status: Enabled

Average Bandwidth: 100000 Kbits/sec

Peak Bandwidth: 100000 Kbits/sec

Burst Size: 102400 Kbytes

Egress Traffic Shaping

Status: Enabled

Average Bandwidth: 100000 Kbits/sec

Peak Bandwidth: 100000 Kbits/sec

Burst Size: 102400 Kbytes

Traffic shaping policy is applied individually to each port in the port group.

Figure 2-78 Traffic Shaping on vDS Port Groups

You can choose to enable ingress, egress, neither, or both. The other settings are very much the same as those for vSSs. You can use ingress traffic to control the amount of bandwidth that hit a port group in a given period of time. This might be useful for web servers as an additional throttling mechanism.

Enabling TCP Segmentation Offload support for a Virtual Machine

TCP Segmentation Offload (TSO) enhances the networking performance of VMs by allowing the TCP stack to emit very large frames (up to 64KB) even though the maximum transmission unit (MTU) of the interface is much smaller. The network adapter will then separate the large frames into MTU sized frames and prepend an adjusted copy of the original TCP/IP headers. In other words, you can send more data through the network in a given time and the vnic on the VM can “take sips from the fire hose.” This is especially useful for VMkernel ports that are being used for iSCSI. TSO is enabled by default for the VMkernel port, but must be enabled on the VM.

As you can imagine, not just any vNIC can handle TSO. In fact, not just any OS can handle it either. If you want to use TSO, you must install an enhanced vmxnet adapter. You can enable TSO support on the VMs that run the following guest OSs:

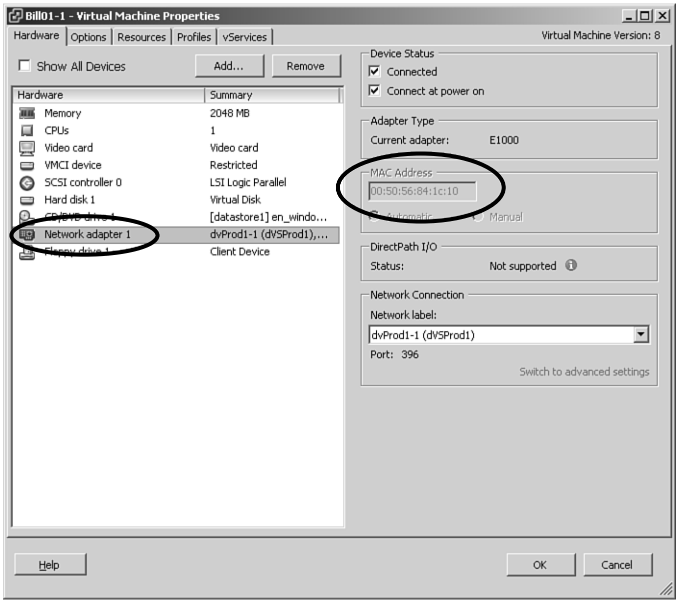
- Microsoft Windows Server 2003 Enterprise Edition with Service Pack 2 (32 bit and 64 bit)
- Red Hat Enterprise Linux 4 (64 bit)
- Red Hat Enterprise Linux 5 (32 bit and 64 bit)
- SUSE Linux Enterprise Server 10 (32 bit and 64 bit)

To enable replace the existing adapter and enable TSO, follow the steps outlined in Activity 2-25.

Activity 2-25 Enabling TSO on a VM

Key Topic

1. Log on to your vCenter Server through your vSphere Client.
2. Locate the VM that you want to configure.
3. Right-click the VM and click **Edit Settings**.
4. Select the network adapter from the hardware list.
5. Record the network settings and the MAC address that the network adapter is using, as shown in Figure 2-79.



Virtual Machine Version: 8

□ Show All Devices

Hardware	Summary
Memory	2048 MB
CPUs	1
Video card	Video card
VMCI device	Restricted
SCSI controller 0	LSI Logic Parallel
Hard disk 1	Virtual Disk
CD/DVD drive 1	[datastore1] en_windo...
Network adapter 1	dvProd1-1 (dVSProd1),...
Floppy drive 1	Client Device

Device Status
☑ Connected
☑ Connect at power on

Adapter Type
Current adapter: E1000

MAC Address
00:50:56:84:1c:10
○ Automatic ○ Manual

DirectPath I/O
Status: Not supported

Network Connection
Network label: dvProd1-1 (dVSProd1)
Port: 396
Switch to advanced settings

Figure 2-79 Enabling TSO on a VM

6. Click **Remove** to remove the network adapter from the VM.
7. Click **Add**, select **Ethernet Adapter**, and click **Next**.
8. In the Adapter Type group, select **Enhanced vmxnet**.
9. Select the network setting and MAC address that the old network adapter was using and click **Next**.
10. Click **Finish**, and then click **OK**.
11. If the VM is not set to upgrade the VMware tools when powered on, upgrade the VMware tools now; otherwise just restart the VM to upgrade the tools.

NOTE TSO is enabled by default on a VMkernel interface. If TSO becomes disabled, the only way to enable it if to delete the VMkernel interface and re-create it with TSO enabled.

Enabling Jumbo Frames Support on Appropriate Components

Another way to enhance network performance and reduce CPU load is through the use of jumbo frames. Enabling jumbo frame support on your ESXi host and VMs allows them to send out much larger frames than normal into the network (9000 bytes versus 1518 bytes). If you are going to send the larger frames, the physical network to which are sending them must be enabled for jumbo frames as well. Before you enable jumbo frames on your ESXi host and VMs, check your vendor documentation to ensure that your physical adapter supports them.

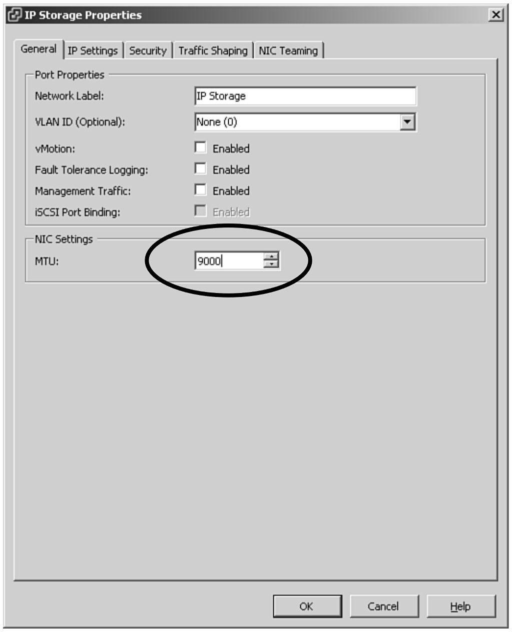
You can then enable jumbo frames for the VMkernel interfaces and for the VMs. Of course, the actual steps to enable jumbo frame support for vSSs are different from those for vDSs. This section first covers enabling jumbo frames on vSSs, then on vDSs, and finally on VMs.

Enabling Jumbo Frames for VMkernel Interface on a vSS

To enable jumbo frames on a VMkernel interface, you only need to change the MTU for the interface. On a vSS, you can make this change in the properties of the switch as outlined in the steps in Activity 2-26.

Activity 2-26 Enabling Jumbo Frames for a VMkernel Interface on a vSS

1. Log on to your vCenter Server through your vSphere Client.
2. Click **Home** and then **Hosts and Clusters**.
3. Select the ESXi host that contains the VMkernel port, and open the Configuration tab.
4. Click the **Networking**, link and then click the **Properties** link next to the switch that contains the VMkernel port that you want to configure.
5. On the Ports tab, select the VMkernel interface and click **Edit**.
6. Set the MTU to **9000**, as shown in Figure 2-80, and click **OK** to confirm and save.



Port Properties

Network Label: IP Storage

VLAN ID (Optional): None (0)

vMotion: □ Enabled

Fault Tolerance Logging: □ Enabled

Management Traffic: □ Enabled

iSCSI Port Binding: □ Enabled

NIC Settings

MTU: 9000

Figure 2-80 Enabling Jumbo Frames on a vSS

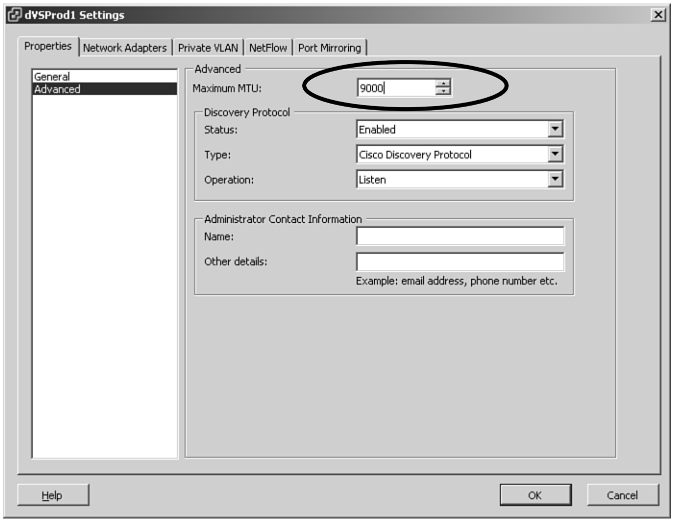
Enabling Jumbo Frames on a vDS

You can enable jumbo frames for an entire vDS. Just as with the VMkernel port on the vSS, you must increase the MTU. Activity 2-27 outlines the steps that you should take.

Key Topic

Activity 2-27 Enabling Jumbo Frames on a vDS

1. Log on to your vCenter Server through your vSphere Client.
2. Click **Home** and then **Networking**.
3. Right-click the vDS that you want to configure and click **Edit**.
4. Click **Advanced**, and then set Maximum MTU to **9000**, as shown in Figure 2-81.
5. Click **OK** to confirm your change and save.



Advanced
Maximum MTU: 9000

Discovery Protocol
Status: Enabled
Type: Cisco Discovery Protocol
Operation: Listen

Administrator Contact Information
Name:
Other details:
Example: email address, phone number etc.

Figure 2-81 Enabling Jumbo Frames on a vDS

Enabling Jumbo Frame Support on Virtual Machines

After you have enabled jumbo frames on your physical network and your virtual switches, configuring the VMs to work with them is simply a matter of installing the proper vnic on the VM and configuring the guest OS to use it. You might think that the best vNIC to use would be the vmxnet3. Actually, there is a known issue with the vmxnet3, so you should choose either the vmxnet2 (enhanced vmxnet) or the e1000 vnic, whichever is best for the OS on the VM. (The precise configuration of the guest OS to use jumbo frames will vary by guest OS and is beyond the scope of this text.)

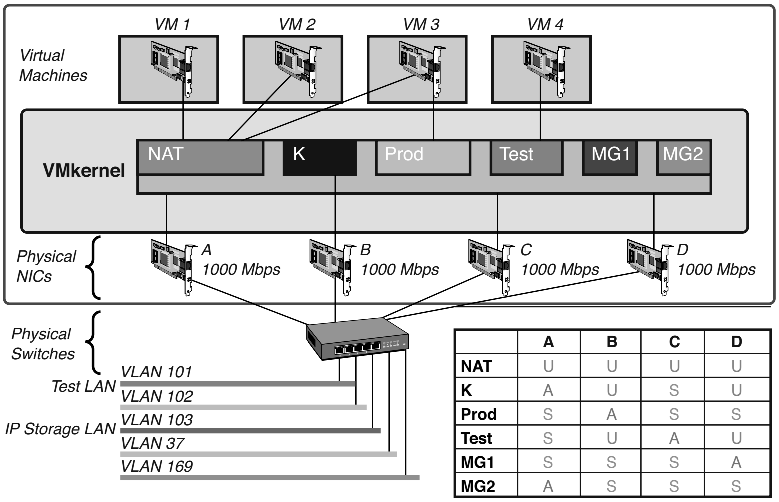
Determining Appropriate VLAN Configuration for a vSphere Implementation

VLANs can be a very powerful tool if used correctly in your network. You can create and control multiple subnets using the same vmnic, or your group of subnets with a small group or vmnics, and provide for load balancing and fault tolerance as well. Using multiple subnets enhances the flexibility and the security of your network because the subnets can indicate a defined purpose for that part of the network, such as iSCSI storage, vMotion, management, NFS datastores, and so on (all of which are covered in Chapter 3). VLANs provide some security if configured properly, but they can also be susceptible to a VLAN hopping attack whereby a person who has access to one VLAN can gain access to the others on the same cable. This is not a good scenario, especially if one of the other networks is the management network. This can typically be prevented by proper configuration of your physical switches and keeping them up to date with firmware and patches.

Still, an even more defined way to separate the components of your network is by using a new vmnic for each one. This is referred to as *physical separation*. It is considered even more secure than VLANs because it avoids the VLAN hopping attack. It is a best practice to use a separate vmnic for each type of network that you create. For example, you should use a separate vmnic for VM port groups than you do for management VMkernel ports. Also, you should separate vmnics for each type of VMkernel port whenever possible. For example, it would be best to have a different vmnic for each of your VMkernel ports for vMotion, FT logging, iSCSI storage, NFS datastores, and management.

So to recap, you are supposed to make use of the VLANs while at the same time using a separate vmnic for just about everything. How can these two best practices possibly coexist? Figure 2-82 illustrates an example whereby a different vmnic is used as the active adapter for each important service, but at the same time VLANs are used to allow those adapters to be a standby adapter for another service. (In the table, *A* stands for active, *S* for standby, and *U* for unused.)

This is just one scenario on one ESXi host's vSS, but it shows the power of what can be done when you begin to use all of your options. Hopefully, you can take the principles learned from this scenario and apply them to your own virtual network. Examine Figure 2-82 to determine what you like about it and what you want to change. For example, adding another physical switch would definitely be a good practice. What else do you see? Again, this is just one scenario of many. So, see whether you can take what I have discussed and build your own scenarios.



	A	B	C	D
NAT	U	U	U	U
K	A	U	S	U
Prod	S	A	S	S
Test	S	U	A	U
MG1	S	S	S	A
MG2	A	S	S	S

Figure 2-82 An Example of VLAN Configuration on a vSS

Summary

The main topics covered in this chapter are the following:

- I began this chapter by identifying the capabilities of vSSs and the creation, configuration, editing, and removal of vSSs and the port groups they contain.
- I then discussed the creation, configuration, management, and removal of vDSs and the port groups that they contain, including comparing and contrasting their features with those of vSSs.
- Finally, I covered the configuration of port groups on both vSSs and vDSs, including policies such as port group blocking, load balancing, failover, traffic shaping, TCP Segmentation Offload, and jumbo frames.

Exam Preparation Tasks

Review All of the Key Topics

Review the most important topics from inside the chapter, noted with the Key Topic icon in the outer margin of the page. Table 2-3 lists these key topics and the page numbers where each is found. Know the main differences between vSSs and vDSs and the port groups on each. Understand how to create, configure, edit, and delete these components and policies.

Table 2-3 Key Topics for Chapter 2

Key Topic Element	Description	Page Number
Figure 2-1	A Diagram of a vSphere Standard Switch	76
Bullet List	Uses of VMkernel Ports	77
Activity 2-1	Creating a New vSphere Standard Switch	78
Activity 2-2	Deleting a vSphere Standard Switch	81
Activity 2-3	Adding a vmnic to a Switch	82
Activity 2-4	Configuring the Physical Aspects of a vmnic	85
Activity 2-5	Removing a vmnic from a vSphere Standard Switch	87
Activity 2-6	Configuring a VMkernel Port for Network Services	88
Activity 2-7	Adding a Port Group to a vSphere Standard Switch	91
Table 2-2	vSS Capabilities Versus vDS Capabilities	95
Bullet List	vDS Capabilities	96
Activity 2-8	Creating a New vSphere Distributed Switch	98
Activity 2-9	Adding a Host to a vSphere Distributed Switch	101
Activity 2-10	Removing a Host from a vSphere Distributed Switch	103
Activity 2-11	Adding a Port Group to a vSphere Distributed Switch	104
Activity 2-12	Configuring Port Groups on a vSphere Distributed Switch	106
Activity 2-13	Removing a Port Group from a vSphere Distributed Switch	107
Figure 2-37	Distributed Switch Architecture	108
Activity 2-14	Adding an Uplink Adapter to a dvUplink Group	108

Key Topic Element	Description	Page Number
Activity 2-15	Removing an Uplink Adapter from a dvUplink Group	110
Activity 2-16	Creating a Virtual Adapter	112
Activity 2-17	Configuring a Virtual Adapter	116
Activity 2-18	Removing a Virtual Adapter	117
Activity 2-19	Migrating Virtual Adapters from a vSS to a vDS	118
Activity 2-20	Migrating a Single VM to/from a vDS	120
Activity 2-21	Migrating Multiple VMs Using vSphere	121
List	Three Main Polices for vSSs	124
Activity 2-22	Identifying Common vSS Policies	125
Activity 2-23	Identifying Common vDS Port Group Policies	129
Activity 2-24	Overriding vDS Port Group Policies at the Port Level	130
Bullet List	Settings for Traffic Shaping	144
Activity 2-25	Enabling TSO on a VM	145
Activity 2-26	Enabling Jumbo Frames for a VMkernel Interface on a vSS	147
Activity 2-27	Enabling Jumbo Frames on a vDS	148

Review Questions

The answers to these review questions are in Appendix A.

1. Which of the following is not a valid use for a VMkernel port?
   a. IP storage
   b. vMotion
   c. FT logging
   d. Service console port

2. Which of the following are types of connections on a vSS on an ESXi 5.0 host? (Choose two.)
   a. Service console
   b. VM
   c. Host bus adapter
   d. VMkernel

3. If you configure policies on a specific port group that conflict with polices on the switch, which of the following will result?
   a. The port group policies on the switch always override those on a port group.
   b. The port group policies override the switch policies for the VMs on that port group.
   c. The port group policies override those on the switch and will be applied to all VMs on all port groups.
   d. A configuration error will be indicated.

4. What is the maximum number of hosts that you can connect to a vDS?
   a. 100
   b. 10
   c. 350
   d. 32

5. Which of the following is the correct traffic shaping metric for average bandwidth, peak bandwidth, and burst size, respectively?
   a. Kbps, Kbps, KB
   b. KB, KB, KB
   c. Kbps, Kbps, Kbps
   d. KB, KB, Kbps

6. What should you change to enable TSO on a VMkernel interface on ESXi 5.0?
   a. You must place a check mark in the correct configuration parameter.
   b. TSO is enabled by default on all VMkernel interfaces on ESXi 5.0.
   c. You need more than one vmnic assigned to the port.
   d. You must enable the interface for IP storage as well.

7. Which of the following is a capability of a vDS but not of a vSS?
   a. Outbound traffic shaping
   b. Network vMotion
   c. VLAN segmentation
   d. NIC teaming

8. If you have a vDS network policy configured for a port group and a conflicting network policy configured for a specific port within the port group, which of the following will result?

   a. The port group policy overrides the specific port policy.

   b. A configuration error will be indicated.

   c. The specific port setting overrides the port group setting for the VM on that port.

   d. The specific port setting is applied to all VMs connected to the port group.

9. Which of the following load balancing policies requires 802.3ad or EtherChannel on the switch?

   a. Route based on IP hash

   b. Route based on MAC hash

   c. Route based on originating virtual port ID

   d. Use explicit failover order

10. To what should you configure the MTU setting on a vSS or vDS to allow for jumbo frames?

    a. 1500

    b. 15000

    c. 150

    d. 9000

11. Which of the following is not a part of the configuration of a VMkernel port?

    a. IP address

    b. Subnet mask

    c. Default gateway

    d. MAC address

12. Which two types of ports can you create on a vSphere 5.0 vSS when you select Add Networking? (Choose two.)

    a. Service console

    b. VM

    c. Host bus adapter

    d. VMkernel

13. Which feature is available with a vDS but not with a vSS?

- **a.** VLAN segmentation
- **b.** Network vMotion
- **c.** 802.1Q tagging
- **d.** NIC teaming

14. Which of the following is required for VM port group configuration?

- **a.** IP address
- **b.** MAC address
- **c.** Label
- **d.** Uplink

15. Which of the following is not a requirement when configuring a VMkernel port?

- **a.** MAC address
- **b.** IP address
- **c.** Subnet mask
- **d.** Default gateway

16. Which of the following tools allows you to migrate multiple VMs from a vSS onto a vDS?

- **a.** vMotion
- **b.** The Migrate Virtual Machine Wizard
- **c.** Storage vMotion
- **d.** DRS

17. Which of the following is *not* one of the three main policies on a vSS?

- **a.** Security
- **b.** IP storage
- **c.** Traffic shaping
- **d.** NIC teaming

18. Which of the following best describes NIC teaming?

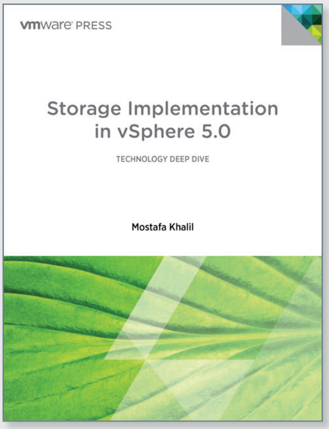
- **a.** NIC teaming is using more than one vmnic on a VM.
- **b.** NIC teaming is using more than one vNIC on a VM.
- **c.** NIC teaming is using more than one vmnic on a switch or port group.
- **d.** NIC teaming is connecting more than one virtual switch to the same vmnic.

19. Which of the following load balancing policies is the default for a vSphere vSS?

- **a.** Route based on originating virtual port ID
- **b.** Route based on MAC hash
- **c.** Route based on IP Hash
- **d.** Use explicit failover order

20. Which VLAN setting should you use on a port group if a VM within it is going to create its own "tagging" and needs to be connected to port groups on other VLANs as well?

- **a.** 1500
- **b.** 1111
- **c.** 1
- **d.** 4095



vmware PRESS

Storage Implementation in vSphere 5.0

TECHNOLOGY DEEP DIVE

Mostafa Khalil

BUY NOW

Storage Implementation in vSphere 5.0

BY MOSTAFA KHALIL

CHAPTER 5 VMware Pluggable Storage Architecture (PSA)

AUGUST 2012

Available in Print and eBook formats and through SAFARI BOOKS ONLINE

Table of Contents

ISBN: 9780321799937

SHARE WITH OTHERS

vmware® PRESS

vmware.com/go/vmwarepress

ALWAYS LEARNING

PEARSON

Chapter 5

vSphere Pluggable Storage Architecture (PSA)

vSphere 5.0 continues to utilize the Pluggable Storage Architecture (PSA) which was introduced with ESX 3.5. The move to this architecture modularizes the storage stack, which makes it easier to maintain and to open the doors for storage partners to develop their own proprietary components that plug into this architecture.

Availability is critical, so redundant paths to storage are essential. One of the key functions of the storage component in vSphere is to provide multipathing (if there are multiple paths, which path should a given I/O use) and failover (when a path goes down, I/O failovers to using another path).

VMware, by default, provides a generic Multipathing Plugin (MPP) called Native Multipathing (NMP).

Native Multipathing

To understand how the pieces of PSA fit together, Figures 5.1, 5.2, 5.4, and 5.6 build up the PSA gradually.

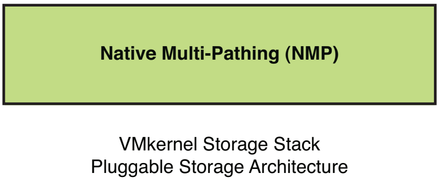


Figure 5.1 Native MPP

NMP is the component of vSphere 5 vmkernel that handles multipathing and failover. It exports two APIs: Storage Array Type Plugin (SATP) and Path Selection Plugin (PSP), which are implemented as plug-ins.

NMP performs the following functions (some done with help from SATPs and PSPs):

- Registers logical devices with the PSA framework
- Receives input/output (I/O) requests for logical devices it registered with the PSA framework
- Completes the I/Os and posts completion of the SCSI command block with the PSA framework, which includes the following operations:
    - Selects the physical path to which it sends the I/O requests
    - Handles failure conditions encountered by the I/O requests
- Handles task management operations—for example, Aborts/Resets

PSA communicates with NMP for the following operations:

- Open/close logical devices.
- Start I/O to logical devices.
- Abort an I/O to logical devices.
- Get the name of the physical paths to logical devices.
- Get the SCSI inquiry information for logical devices.

Storage Array Type Plug-in (SATP)

Figure 5.2 depicts the relationship between SATP and NMP.

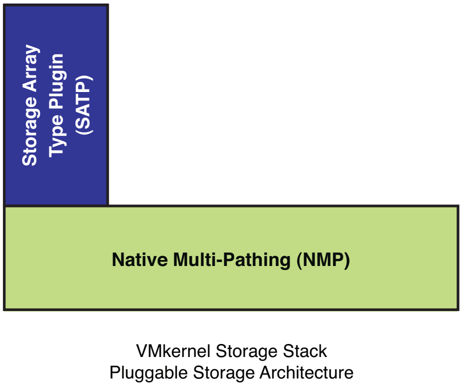

Storage Array Type Plugin (SATP)
Native Multi-Pathing (NMP)
VMkernel Storage Stack
Pluggable Storage Architecture

Figure 5.2 SATP

SATPs are PSA plug-ins specific to certain storage arrays or storage array families. Some are generic for certain array classes—for example, Active/Passive, Active/Active, or ALUA-capable arrays.

SATPs handle the following operations:

- Monitor the hardware state of the physical paths to the storage array
- Determine when a hardware component of a physical path has failed
- Switch physical paths to the array when a path has failed

NMP communicates with SATPs for the following operations:

- Set up a new logical device—claim a physical path
- Update the hardware states of the physical paths (for example, Active, Standby, Dead)
- Activate the standby physical paths of an active/passive array (when Active paths state is dead or unavailable)
- Notify the plug-in that an I/O is about to be issued on a given path
- Analyze the cause of an I/O failure on a given path (based on errors returned by the array)

Examples of SATPs are listed in Table 5.1:

Table 5.1 Examples of SATPs

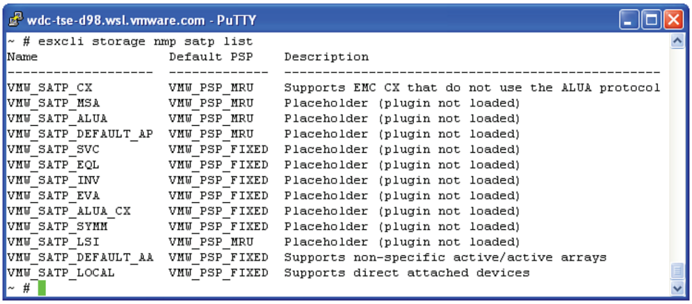
SATP	Description
VMW_SATP_CX	Supports EMC CX that do not use the ALUA protocol
VMW_SATP_ALUA_CX	Supports EMC CX that use the ALUA protocol
VMW_SATP_SYMM	Supports EMC Symmetrix array family
VMW_SATP_INV	Supports EMC Invista array family
VMW_SATP_EVA	Supports HP EVA arrays
VMW_SATP_MSA	Supports HP MSA arrays
VMW_SATP_EQL	Supports Dell Equalogic arrays
VMW_SATP_SVC	Supports IBM SVC arrays
VMW_SATP_LSI	Supports LSI arrays and others OEMed from it (for example, DS4000 family)
VMW_SATP_ALUA	Supports non-specific arrays that support ALUA protocol
VMW_SATP_DEFAULT_AA	Supports non-specific active/active arrays
VMW_SATP_DEFAULT_AP	Supports non-specific active/passive arrays
VMW_SATP_LOCAL	Supports direct attached devices

How to List SATPs on an ESXi 5 Host

To obtain a list of SATPs on a given ESXi 5 host, you may run the following command directly on the host or remotely via an SSH session, a vMA appliance, or ESXCLI:

```
# esxcli storage nmp satp list
```

An example of the output is shown in Figure 5.3.



```
~ # esxcli storage nmp satp list
Name                 Default PSP    Description
-------------------  -------------  -------------------------------------------------
VMW_SATP_CX          VMW_PSP_MRU    Supports EMC CX that do not use the ALUA protocol
VMW_SATP_MSA         VMW_PSP_MRU    Placeholder (plugin not loaded)
VMW_SATP_ALUA        VMW_PSP_MRU    Placeholder (plugin not loaded)
VMW_SATP_DEFAULT_AP  VMW_PSP_MRU    Placeholder (plugin not loaded)
VMW_SATP_SVC         VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_EQL         VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_INV         VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_EVA         VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_ALUA_CX     VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_SYMM        VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_LSI         VMW_PSP_MRU    Placeholder (plugin not loaded)
VMW_SATP_DEFAULT_AA  VMW_PSP_FIXED  Supports non-specific active/active arrays
VMW_SATP_LOCAL       VMW_PSP_FIXED  Supports direct attached devices
~ #
```

Figure 5.3 Listing SATPs

Notice that each SATP is listed in association with a specific PSP. The output shows the default configuration of a freshly installed ESXi 5 host. To modify these associations, refer to the "Modifying PSA Plug-in Configurations Using the UI" section later in this chapter.

If you installed third-party SATPs, they are listed along with the SATPs shown in Table 5.1.

NOTE

ESXi 5 only loads the SATPs matching detected storage arrays based on the corresponding claim rules. See the "Claim Rules" section later in this chapter for more about claim rules. Otherwise, you see them listed as (Plugin not loaded) similar to the output shown in Figure 5.3.

Path Selection Plugin (PSP)

Figure 5.4 depicts the relationship between SATP, PSP, and NMP.

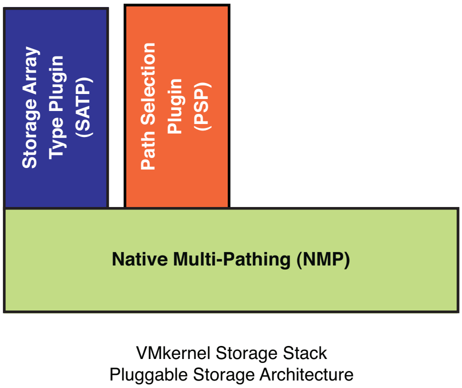

Storage Array Type Plugin (SATP)
Path Selection Plugin (PSP)
Native Multi-Pathing (NMP)
VMkernel Storage Stack
Pluggable Storage Architecture

Figure 5.4 PSP

PSPs are PSA plug-ins that handle path selection policies and are replacements of failover policies used by the Legacy-MP (or Legacy Multipathing) used in releases prior to vSphere 4.x.

PSPs handle the following operations:

- Determine on which physical path to issue I/O requests being sent to a given storage device. Each PSP has access to a group of paths to the given storage device and has knowledge of the paths' states—for example, Active, Standby, Dead, as well as Asymmetric Logical Unit Access (ALUA), Asymmetric Access States (AAS) such as Active optimized Active non-optimized, and so on. This knowledge is obtained from what SATPs report to NMP. Refer to Chapter 6, "ALUA," for additional details about ALUA.
- Determine which path to activate next if the currently working physical path to storage device fails.

NOTE

PSPs do not need to know the actual storage array type (this function is provided by SATPs). However, a storage vendor developing a PSP may choose to do so (see Chapter 8, "Third-Party Multipathing I/O Plug-ins").

NMP communicates with PSPs for the following operations:

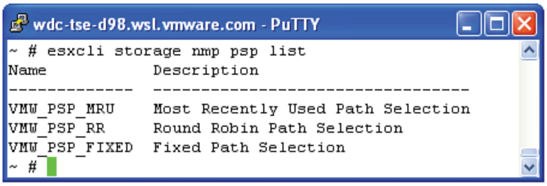
- Set up a new logical storage device and claim the physical paths to that device.
- Get the set of active physical paths currently used for path selection.
- Select a physical path on which to issue I/O requests for a given device.
- Select a physical path to activate when a path failure condition exists.

How to List PSPs on an ESXi 5 Host

To obtain a list of PSPs on a given ESXi 5 host, you may run the following command directly on the host or remotely via an SSH session, a vMA appliance, or ESXCLI:

```
# esxcli storage nmp psp list
```

An example of the output is shown in Figure 5.5.



```
~ # esxcli storage nmp psp list
Name           Description
-------------  ---------------------------------
VMW_PSP_MRU    Most Recently Used Path Selection
VMW_PSP_RR     Round Robin Path Selection
VMW_PSP_FIXED  Fixed Path Selection
~ #
```

Figure 5.5 Listing PSPs

The output shows the default configuration of a freshly installed ESXi 5 host. If you installed third-party PSPs, they are also listed.

Third-Party Plug-ins

Figure 5.6 depicts the relationship between third-party plug-ins, NMP, and PSA.

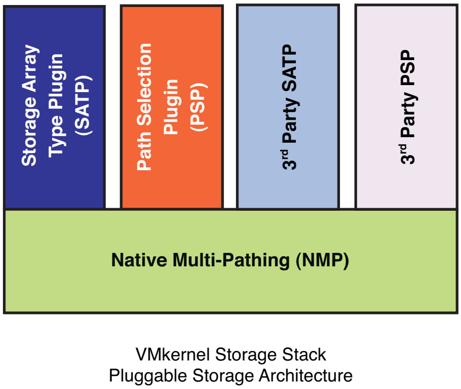

Storage Array Type Plugin (SATP)
Path Selection Plugin (PSP)
3rd Party SATP
3rd Party PSP
Native Multi-Pathing (NMP)
VMkernel Storage Stack
Pluggable Storage Architecture

Figure 5.6 Third-party plug-ins

Because PSA is a modular architecture, VMware provided APIs to its storage partners to develop their own plug-ins. These plug-ins can be SATPs, PSPs, or MPPs.

Third-party SATPs and PSPs can run side by side with VMware-provided SATPs and PSPs.

The third-party SATPs and PSPs providers can implement their own proprietary functions relevant to each plug-in that are specific to their storage arrays. Some partners implement only multipathing and failover algorithms, whereas others implement load balancing and I/O optimization as well.

Examples of such plug-ins in vSphere 4.x that are also planned for vSphere 5 are

- **DELL_PSP_EQL_ROUTED**—Dell EqualLogic PSP that provides the following enhancements:
  - Automatic connection management
  - Automatic load balancing across multiple active paths
  - Increased bandwidth
  - Reduced network latency
- **HTI_SATP_HDLM**—Hitachi ported their HDLM MPIO (Multipathing I/O) management software to an SATP. It is currently certified for vSphere 4.1 with most of the USP family of arrays from Hitachi and HDS. A version is planned for vSphere 5 as well for the same set of arrays. Check with VMware HCL for the current list of certified arrays for vSphere 5 with this plug-in.

See Chapter 8 for further details.

Multipathing Plugins (MPPs)

Figure 5.7 depicts the relationship between MPPs, NMP, and PSA.

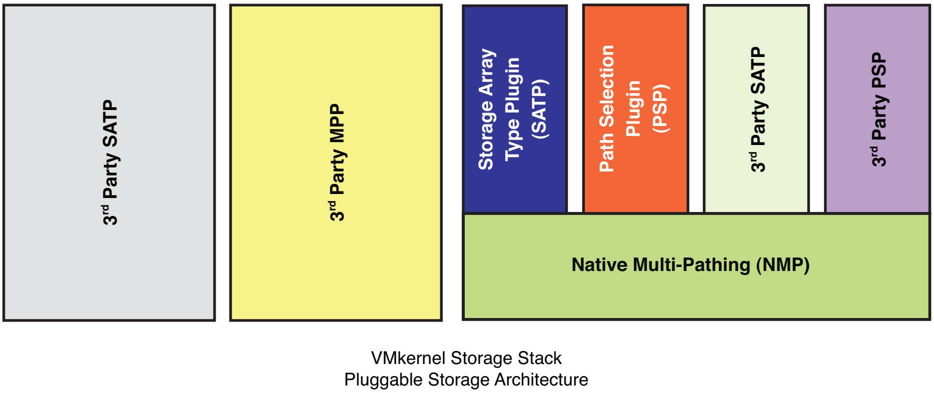

3rd Party SATP
3rd Party MPP
Storage Array Type Plugin (SATP)
Path Selection Plugin (PSP)
3rd Party SATP
3rd Party PSP
Native Multi-Pathing (NMP)
VMkernel Storage Stack
Pluggable Storage Architecture

Figure 5.7 MPPs, including third-party plug-ins

MPPs that are not implemented as SATPs or PSPs can be implemented as MPPs instead. MPPs run side by side with NMP. An example of that is EMC PowerPath/VE. It is certified with vSphere 4.x and is planned for vSphere 5.

See Chapter 8 for further details.

Anatomy of PSA Components

Figure 5.8 is a block diagram showing the components of PSA framework.

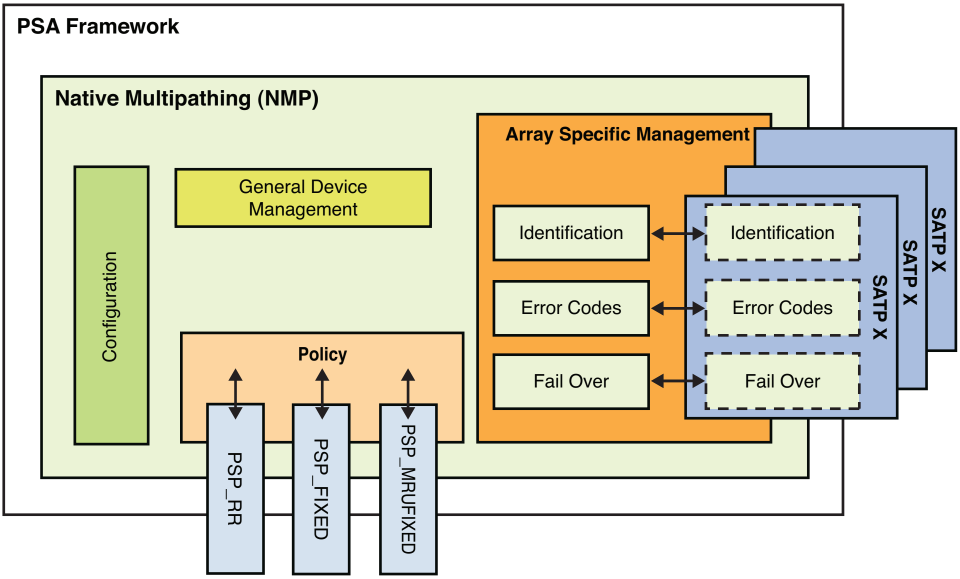


Figure 5.8 NMP components of PSA framework

Now that we covered the individual components of PSA framework, let's put its pieces together. Figure 5.8 shows the NMP component of the PSA framework. NMP provides facilities for configuration, general device management, array-specific management, and path selection policies.

The configuration of NMP-related components can be done via ESXCLI or the user interface (UI) provided by vSphere Client. Read more on this topic in the "Modifying PSA Plug-in Configurations Using the UI" section later in this chapter.

Multipathing and failover policy is set by NMP with the aid of PSPs. For details on how to configure the PSP for a given array, see the “Modifying PSA Plug-in Configurations Using the UI” section later in this chapter.

Arrray-specific functions are handled by NMP via the following functions:

- **Identification**—This is done by interpreting the response data to various inquiry commands (Standard Inquiry and Vital Product Data (VPD) received from the array/storage. This provides details of device identification which include the following:

    - Vendor
    - Model
    - LUN number
    - Device ID—for example, NAA ID, serial number
    - Supported mode pages—for example, page 80 or 83

    I cover more detail and examples of inquiry strings in Chapter 7, “Multipathing and Failover” in, the “LUN Discovery and Path Enumeration” section.

- **Error Codes**—NMP interprets error codes received from the storage arrays with help from the corresponding SATPs and acts upon these errors. For example, an SATP can identify a path as dead.

- **Failover**—After NMP interprets the error codes, it reacts in response to them. Continuing with the example, after a path is identified as dead, NMP instructs the relevant SATP to activate standby paths and then instructs the relevant PSP to issue the I/O on one of the activated paths. In this example, there are no active paths remaining, which results in activating standby paths (which is the case for Active/Passive arrays).

I/O Flow Through PSA and NMP

In order to understand how I/O sent to storage devices flows through the ESXi storage stack, you first need to understand some of the terminology relevant to this chapter.

Classification of Arrays Based on How They Handle I/O

Arrays can be one of the following types:

- **Active/Active**—This type of array would have more than one Storage Processor (SP) (also known as Storage Controller) that can process I/O concurrently on all SPs (and SP ports) with similar performance metrics. This type of array has no concept of logical unit number (LUN) ownership because I/O can be done on any LUN via any SP port from initiators given access to such LUNs.
- **Active/Passive**—This type of array would have two SPs. LUNs are distributed across both SPs in a fashion referred to as LUN ownership in which one of the SPs owns some of the LUNs and the other SP owns the remaining LUNs. The array accepts I/O to given LUN via ports on that SP that "owns" it. I/O sent to the non-owner SPs (also known as Passive SP) is rejected with a SCSI check condition and a sense code that translates to ILLEGAL REQUEST. Think of this like the No Entry sign you see at the entrance of a one-way street in the direction opposite to the traffic. For more details on sense codes, see Chapter 7 's "LUN Discovery and Path Enumeration" section.

NOTE

Some older firmware versions of certain arrays, such as HP MSA, are a variety of this type where one SP is active and the other is standby. The difference is that all LUNs are owned by the active SP and the standby SP is only used when the active SP fails. The standby SP still responds with a similar sense code to that returned from the passive SP described earlier.

- **Asymmetric Active/Active or AAA (AKA Pseudo Active/Active)**—LUNs on this type of arrays are owned by either SP similarly to the Active/Passive Arrays concept of LUN ownership. However, the array would allow concurrent I/O on a given LUN via ports on both SPs but with different I/O performance metrics as I/O is sent via proxy from the non-owner SP to the owner SP. In this case, the SP providing the lower performance metric accepts I/O to that LUN without returning a check condition. You may think of this as a hybrid between Active/Passive and Active/Active types. This can result in poor I/O performance of all paths to the owner SP that are dead, either due to poor design or LUN owner SP hardware failure.
- **Asymmetrical Logical Unite Access (ALUA)**—This type of array is an enhanced version of the Asymmetric Active/Active arrays and also the newer generation of some of the Active/Passive arrays. This technology allows initiators to identify the ports on the owner SP as one group and the ports on the non-owner SP as a

different group. This is referred to as Target Port Group Support (TPGS). The port group on the owner SP is identified as Active Optimized port group with the other group identified as Active Non-Optimized port group. NMP would send the I/O to a given LUN via a port in the ALUA optimized port group only as long as they are available. If all ports in that group are identified as dead, I/O is then sent to a port on the ALUA non-optimized port group. When sustained I/O is sent to the ALUA non-optimized port group, the array can transfer the LUN ownership to the non-owner SP and then transition the ports on that SP to ALUA optimized state. For more details on ALUA see Chapter 6.

Paths and Path States

From a storage perspective, the possible routes to a given LUN through which the I/O may travel is referred to as *paths*. A path consists of multiple points that start from the initiator port and end at the LUN.

A path can be in one of the states listed in Table 5.2.

Table 5.2 Path States

Path State	Description
Active	A path via an Active SP. I/O can be sent to any path in this state.
Standby	A path via a Passive or Standby SP. I/O is not sent via such a path.
Disabled	A path that is disabled usually by the vSphere Administrator.
Dead	A path that lost connectivity to the storage network. This can be due to an HBA (Host Bus Adapter), Fabric or Ethernet switch, or SP port connectivity loss. It can also be due to HBA or SP hardware failure.
Unknown	The state could not be determined by the relevant SATP.

Preferred Path Setting

A preferred path is a setting that NMP honors for devices claimed by VMW_PSP_FIXED PSP only. All I/O to a given device is sent over the path configured as the Preferred Path for that device. When the preferred path is unavailable, I/O is sent via one of the surviving paths. When the preferred path becomes available, I/O fails back to that path. By default, the first path discovered and claimed by the PSP is set as the preferred path. To change the preferred path setting, refer to the "Modifying PSA Plug-in Configurations Using the UI" section later in this chapter.

Figure 5.9 shows an example of a path to LUN 1 from host A (interrupted line) and Host B (interrupted line with dots and dashes). This path goes through HBA0 to target 1 on SPA.

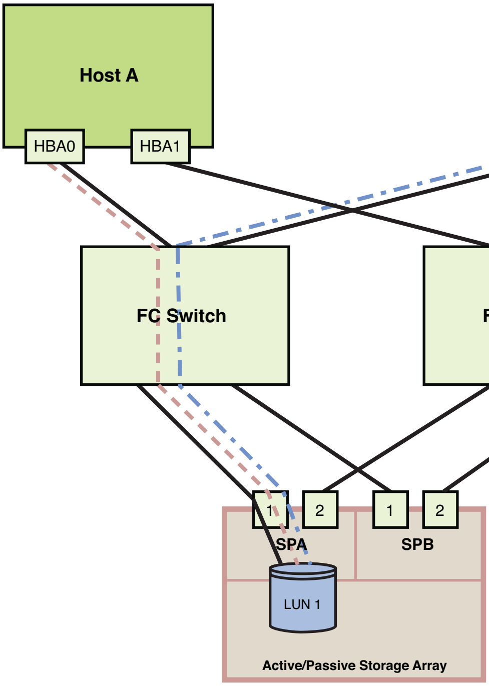


Figure 5.9 Paths to LUN1 from two hosts

Such a path is represented by the following Runtime Name naming convention. (Runtime Name is formerly known as Canonical Name.) It is in the format of HBAx:Cn:Ty:Lz—for example, vmhba0:C0:T0:L1—which reads as follows:

vmhba0, Channel 0, Target 0, LUN1

It represents the path to LUN 0 broken down as the following:

- **HBA0**—First HBA in this host. The vmhba number may vary based on the number of storage adapters installed in the host. For example, if the host has two RAID controllers installed which assume vmhba0 and vmhba1 names, the first FC HBA would be named vmhba2.
- **Channel 0**—Channel number is mostly zero for Fiber Channel (FC)- and Internet Small Computer System Interface (iSCSI)-attached devices to target 0, which is the

first target. If the HBA were a SCSI adapter with two channels (for example, internal connections and an external port for direct attached devices), the channel numbers would be 0 and 1.

- **Target 0**—The target definition was covered in Chapters 3, "FCoE Storage Connectivity," and 4, "iSCSI Storage Connectivity." The target number is based on the order in which the SP ports are discovered by PSA. In this case, SPA-Port1 was discovered before SPA-Port2 and the other ports on SPB. So, that port was given "target 0" as the part of the runtime name.

NOTE

Runtime Name, as the name indicates, does not persist between host reboots. This is due to the possibility that any of the components that make up that name may change due to hardware or connectivity changes. For example, a host might have an additional HBA added or another HBA removed, which would change the number assumed by the HBA.

Flow of I/O Through NMP

Figure 5.10 shows the flow of I/O through NMP.

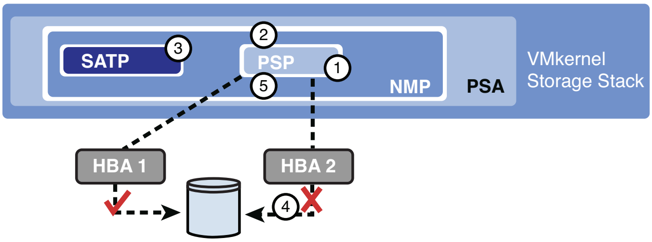


Figure 5.10 I/O flow through NMP

The numbers in the figure represent the following steps:

1. NMP calls the PSP assigned to the given logical device.
2. The PSP selects an appropriate physical path on which to send the I/O. If the PSP is VMW_PSP_RR, it load balances the I/O over paths whose states are Active or, for ALUA devices, paths via a target port group whose AAS is Active/Optimized.

3. If the array returns I/O error, NMP calls the relevant SATP.
4. The SATP interprets the error codes, activates inactive paths, and then fails over to the new active path.
5. PSP selects new active path to which it sends the I/O.

Listing Multipath Details

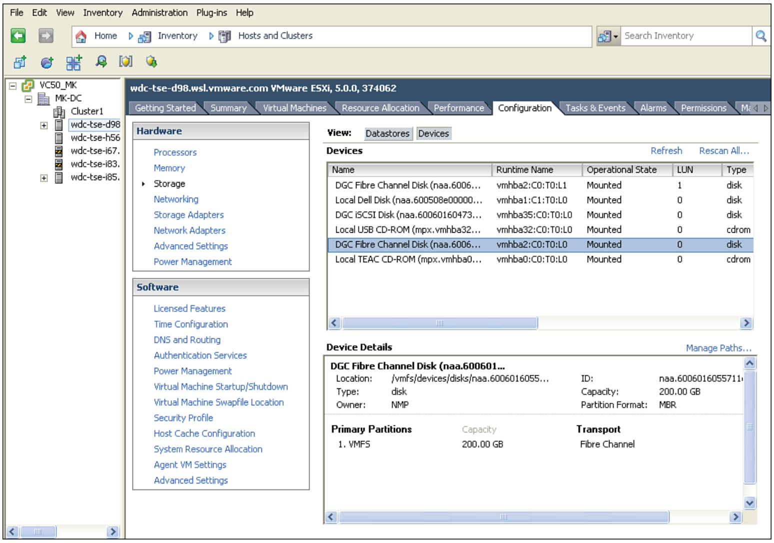
There are two ways by which you can display the list of paths to a given LUN, each of which are discussed in this section:

- Listing paths to a LUN using the UI
- Listing paths to a LUN using the CLI

Listing Paths to a LUN Using the UI

To list all paths to a given LUN in the vSphere 5.0 host, you may follow this procedure, which is similar to the procedure for listing all targets discussed earlier in Chapter 2, "Fibre Channel Storage Connectivity" Chapter 3 and Chapter 4:

1. Log on to the vSphere 5.0 host directly or to the vCenter server that manages the host using the VMware vSphere 5.0 Client as a user with Administrator privileges.
2. While in the Inventory—Hosts and Clusters view, locate the vSphere 5.0 host in the inventory tree and select it.
3. Navigate to the **Configuration** tab.
4. Under the Hardware section, select the **Storage** option.
5. Under the **View** field, click the **Devices** button.
6. Under the Devices pane, select one of the SAN LUNs (see Figure 5.11). In this example, the device name starts with DGC Fibre Channel Disk.



wdc-tse-d98.wsl.vmware.com VMware ESXi, 5.0.0, 374062

Devices

Name	Runtime Name	Operational State	LUN	Type
DGC Fibre Channel Disk (naa.6006...	vmhba2:C0:T0:L1	Mounted	1	disk
Local Dell Disk (naa.600508e00000...	vmhba1:C1:T0:L0	Mounted	0	disk
DGC iSCSI Disk (naa.60060160473...	vmhba35:C0:T0:L0	Mounted	0	disk
Local USB CD-ROM (mpx.vmhba32...	vmhba32:C0:T0:L0	Mounted	0	cdrom
DGC Fibre Channel Disk (naa.6006...	vmhba2:C0:T0:L0	Mounted	0	disk
Local TEAC CD-ROM (mpx.vmhba0...	vmhba0:C0:T0:L0	Mounted	0	cdrom

Device Details

DGC Fibre Channel Disk (naa.600601...

Location:	/vmfs/devices/disks/naa.6006016055...	ID:	naa.6006016055711
Type:	disk	Capacity:	200.00 GB
Owner:	NMP	Partition Format:	MBR

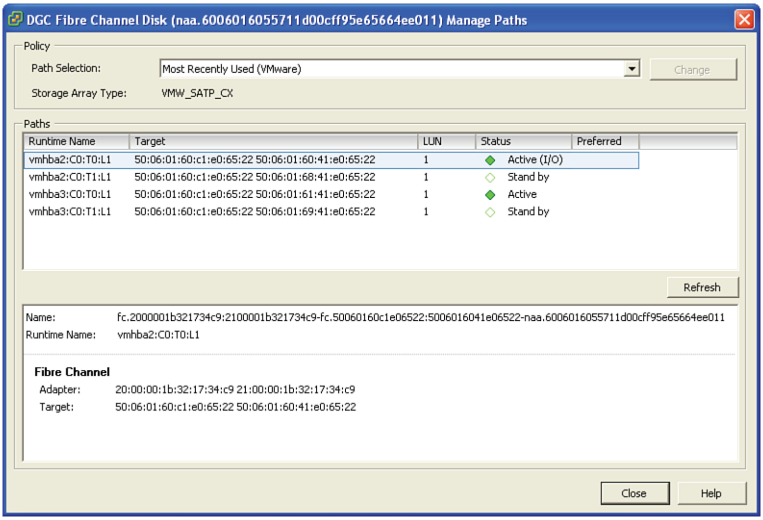
Primary Partitions	Capacity	Transport
1. VMFS	200.00 GB	Fibre Channel

Figure 5.11 Listing storage devices

7. Select **Manage Paths** in the **Device Details** pane.

8. Figure 5.12 shows details for an FC-attached LUN. In this example, I sorted on the Runtime Name column in ascending order. The **Paths** section shows all available paths to the LUN in the format:

   - **Runtime Name**—vmhbaX:C0:Ty:Lz where X is the HBA number, y is the target number, and z is the LUN number. More on that in the "Preferred Path Setting" section later in this chapter.
   - **Target**—The WWNN followed by the WWPN of the target (separated by a space).
   - **LUN**—The LUN number that can be reached via the listed paths.
   - **Status**—This is the path state for each listed path.



DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011) Manage Paths

Policy

Path Selection: Most Recently Used (VMware)

Storage Array Type: VMW_SATP_CX

Paths

Runtime Name	Target	LUN	Status	Preferred
vmhba2:C0:T0:L1	50:06:01:60:c1:e0:65:22 50:06:01:60:41:e0:65:22	1	◆ Active (I/O)	
vmhba2:C0:T1:L1	50:06:01:60:c1:e0:65:22 50:06:01:68:41:e0:65:22	1	◇ Stand by	
vmhba3:C0:T0:L1	50:06:01:60:c1:e0:65:22 50:06:01:61:41:e0:65:22	1	◆ Active	
vmhba3:C0:T1:L1	50:06:01:60:c1:e0:65:22 50:06:01:69:41:e0:65:22	1	◇ Stand by	

Name: fc.2000001b321734c9:2100001b321734c9-fc.50060160c1e06522:5006016041e06522-naa.6006016055711d00cff95e65664ee011

Runtime Name: vmhba2:C0:T0:L1

Fibre Channel

Adapter: 20:00:00:1b:32:17:34:c9 21:00:00:1b:32:17:34:c9

Target: 50:06:01:60:c1:e0:65:22 50:06:01:60:41:e0:65:22

Figure 5.12 Listing paths to an FC-attached LUN

9. The Name field in the lower pane is a permanent one compared to the Runtime Name listed right below it. It is made up of three parts: HBA name, Target Name, and the LUN's device ID separated by dashes (for FC devices) or commas (for iSCSI devices). The HBA and Target names differ by the protocol used to access the LUN.

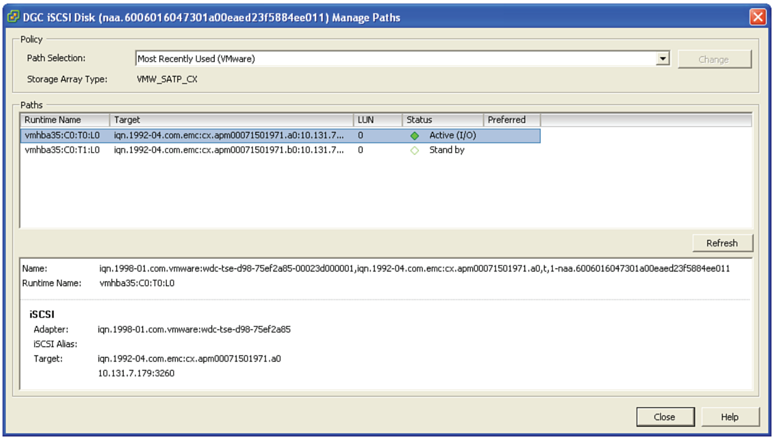
   Figure 5.12 shows the FC-based path Name, which is comprised of

   - **Initiator Name**—Made up from the letters FC followed by a period and then the HBA's WWNN and WWPN. The latter two are separated by a colon (these are discussed in Chapter 3).
   - **Target Name**—Made up from the target's WWNN and WWPN separated by a colon.
   - **LUN's Device ID**—In this example the NAA ID is naa.6006016055711d00cff95e65664ee011, which is based on the Network Address Authority naming convention and is a unique identifier of the logical device representing the LUN.

   Figure 5.13 shows the iSCSI-based path Name which is comprised of

   - **Initiator Name**—This is the iSCSI iqn name discussed in Chapter 4.

- **Target Name**—Made up from the target's iqn name and target number separated by colons. In this example, the target's iqn names are identical while the target numbers are different—such as t,1 and t,2. The second target info is not shown here, but you can display them by selecting one path at a time in the paths, pane to display the details in the lower pane.
- **LUN's Device ID**—In this example the NAA ID is naa.6006016047301a00eaed23f5884ee011, which is based on the Network Address Authority naming convention and is a unique identifier of the logical device representing the LUN.



DGC iSCSI Disk (naa.6006016047301a00eaed23f5884ee011) Manage Paths

Policy

Path Selection: Most Recently Used (VMware)

Storage Array Type: VMW_SATP_CX

Paths

Runtime Name	Target	LUN	Status	Preferred
vmhba35:C0:T0:L0	iqn.1992-04.com.emc:cx.apm00071501971.a0:10.131.7...	0	◆ Active (I/O)	
vmhba35:C0:T1:L0	iqn.1992-04.com.emc:cx.apm00071501971.b0:10.131.7...	0	◇ Stand by	

Name: iqn.1998-01.com.vmware:wdc-tse-d98-75ef2a85-00023d000001,iqn.1992-04.com.emc:cx.apm00071501971.a0,t,1-naa.6006016047301a00eaed23f5884ee011

Runtime Name: vmhba35:C0:T0:L0

iSCSI

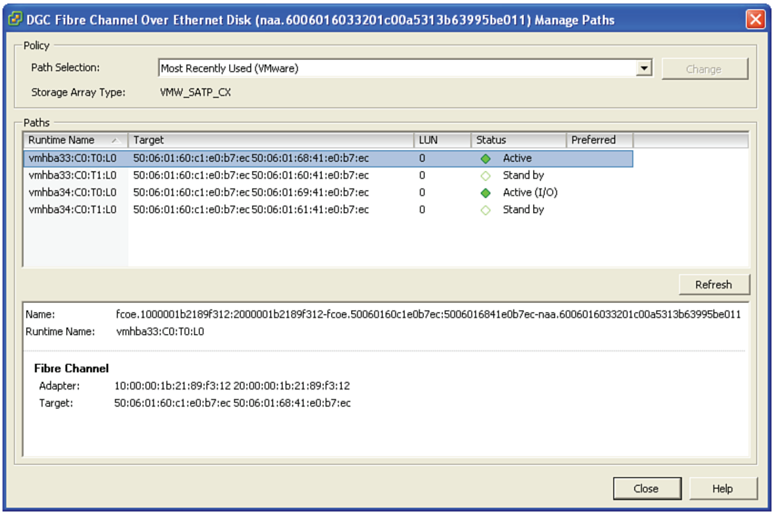
Adapter: iqn.1998-01.com.vmware:wdc-tse-d98-75ef2a85

iSCSI Alias:

Target: iqn.1992-04.com.emc:cx.apm00071501971.a0
10.131.7.179:3260

Figure 5.13 Listing paths to an iSCSI-attached LUN

Figure 5.14 shows a Fibre Channel over Ethernet (FCoE)-based path name, which is identical to the FC-based pathnames. The only difference is that fcoe is used in place of fc throughout the name.



DGC Fibre Channel Over Ethernet Disk (naa.6006016033201c00a5313b63995be011) Manage Paths

Policy

Path Selection: Most Recently Used (VMware)

Storage Array Type: VMW_SATP_CX

Paths

Runtime Name	Target	LUN	Status	Preferred
vmhba33:C0:T0:L0	50:06:01:60:c1:e0:b7:ec 50:06:01:68:41:e0:b7:ec	0	◆ Active	
vmhba33:C0:T1:L0	50:06:01:60:c1:e0:b7:ec 50:06:01:60:41:e0:b7:ec	0	◇ Stand by	
vmhba34:C0:T0:L0	50:06:01:60:c1:e0:b7:ec 50:06:01:69:41:e0:b7:ec	0	◆ Active (I/O)	
vmhba34:C0:T1:L0	50:06:01:60:c1:e0:b7:ec 50:06:01:61:41:e0:b7:ec	0	◇ Stand by	

Name: fcoe.1000001b2189f312:2000001b2189f312-fcoe.50060160c1e0b7ec:5006016841e0b7ec-naa.6006016033201c00a5313b63995be011

Runtime Name: vmhba33:C0:T0:L0

Fibre Channel

Adapter: 10:00:00:1b:21:89:f3:12 20:00:00:1b:21:89:f3:12

Target: 50:06:01:60:c1:e0:b7:ec 50:06:01:68:41:e0:b7:ec

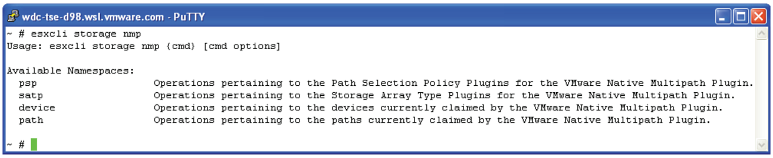
Figure 5.14 Listing paths to an FCoE-attached LUN

Listing Paths to a LUN Using the Command-Line Interface (CLI)

ESXCLI provides similar details to what is covered in the preceding section. For details about the various facilities that provide access to ESXCLI, refer to the “Locating HBA’s WWPN and WWNN in vSphere 5 Hosts” section in Chapter 2.

The namespace of ESXCLI in vSphere 5.0 is fairly intuitive! Simply start with esxcli followed by the area of vSphere you want to manage—for example, esxcli network, esxcli software, esxcli storage—which enables you to manage Network, ESXi Software, and Storage, respectively. For more available options just run `esxcli -help`. Now, let’s move on to the available commands:

Figure 5.15 shows the `esxcli storage nmp` namespace.



wdc-tse-d98.wsl.vmware.com - PuTTY

```
~ # esxcli storage nmp
Usage: esxcli storage nmp {cmd} [cmd options]

Available Namespaces:
  psp                     Operations pertaining to the Path Selection Policy Plugins for the VMware Native Multipath Plugin.
  satp                    Operations pertaining to the Storage Array Type Plugins for the VMware Native Multipath Plugin.
  device                  Operations pertaining to the devices currently claimed by the VMware Native Multipath Plugin.
  path                    Operations pertaining to the paths currently claimed by the VMware Native Multipath Plugin.

~ #
```

Figure 5.15 esxcli storage nmp namespace

The namespace of `esxcli storage nmp` is for all operations pertaining to native multipathing, which include psp, satp, device, and path.

I cover all these namespaces in detail later in the “Modifying PSA Plug-in Configurations Using the UI” section. The relevant operations for this section are

- `esxcli storage nmp path list`
- `esxcli storage nmp path list -d <device ID e.g. NAA ID>`

The first command provides a list of paths to *all* devices regardless of how they are attached to the host or which protocol is used.

The second command lists the paths to the device specified by the device ID (for example, NAA ID) by using the `-d` option.

The command in this example is

```
esxcli storage nmp path list -d naa.6006016055711d00cff95e65664ee011
```

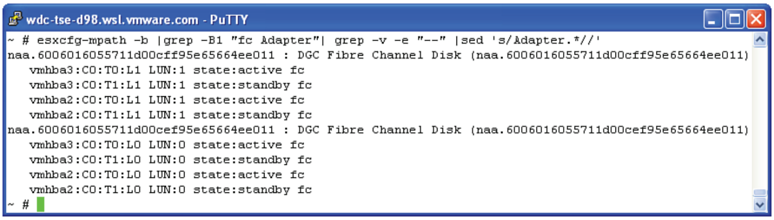
You may also use the verbose command option `--device` instead of `-d`.

You can identify the NAA ID of the device you want to list by running a command like this:

```
esxcfg-mpath -b |grep -B1 “fc Adapter”| grep -v -e “--” |sed ‘s/
Adapter.*//’
```

You may also use the verbose command option `--list-paths` instead of `-b`.

The output of this command is shown in Figure 5.16.



```
~ # esxcfg-mpath -b |grep -B1 "fc Adapter"| grep -v -e "--" |sed 's/Adapter.*//'
naa.6006016055711d00cff95e65664ee011 : DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011)
   vmhba3:C0:T0:L1 LUN:1 state:active fc
   vmhba3:C0:T1:L1 LUN:1 state:standby fc
   vmhba2:C0:T0:L1 LUN:1 state:active fc
   vmhba2:C0:T1:L1 LUN:1 state:standby fc
naa.6006016055711d00cef95e65664ee011 : DGC Fibre Channel Disk (naa.6006016055711d00cef95e65664ee011)
   vmhba3:C0:T0:L0 LUN:0 state:active fc
   vmhba3:C0:T1:L0 LUN:0 state:standby fc
   vmhba2:C0:T0:L0 LUN:0 state:active fc
   vmhba2:C0:T1:L0 LUN:0 state:standby fc
~ #
```

Figure 5.16 Listing paths to an FC-attached LUN via the CLI

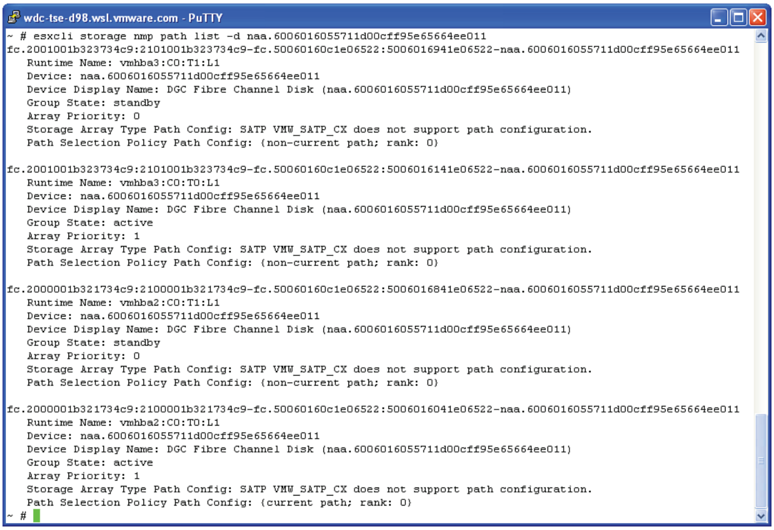
This output shows all FC-attached devices. The Device Display Name of each device is listed followed immediately by the Runtime Name (for example, vmhba3:C0:T0:L1) of all paths to that device. This output is somewhat similar to the lagacy multipathing outputs you might have seen with ESX server release 3.5 and older.

The Device Display Name is actually listed after the device NAA ID and a colon.

From the runtime name you can identify the LUN number and the HBA through which they can be accessed. The HBA number is the first part of the Runtime Name, and the LUN number is the last part of that name.

All block devices conforming to the SCSI-3 standard have an NAA device ID assigned, which is listed at the beginning and the end of the Device Display Name line in the preceding output.

In this example, FC-attached LUN 1 has NAA ID `naa.6006016055711d00cff95e65664ee011` and that of LUN0 is `naa.6006016055711d00cef95e65664ee011`. I use the device ID for LUN 1 in the output shown in Figure 5.17.



```
~ # esxcli storage nmp path list -d naa.6006016055711d00cff95e65664ee011
fc.2001001b323734c9:2101001b323734c9-fc.50060160c1e06522:5006016941e06522-naa.6006016055711d00cff95e65664ee011
   Runtime Name: vmhba3:C0:T1:L1
   Device: naa.6006016055711d00cff95e65664ee011
   Device Display Name: DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011)
   Group State: standby
   Array Priority: 0
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {non-current path; rank: 0}

fc.2001001b323734c9:2101001b323734c9-fc.50060160c1e06522:5006016141e06522-naa.6006016055711d00cff95e65664ee011
   Runtime Name: vmhba3:C0:T0:L1
   Device: naa.6006016055711d00cff95e65664ee011
   Device Display Name: DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011)
   Group State: active
   Array Priority: 1
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {non-current path; rank: 0}

fc.2000001b321734c9:2100001b321734c9-fc.50060160c1e06522:5006016841e06522-naa.6006016055711d00cff95e65664ee011
   Runtime Name: vmhba2:C0:T1:L1
   Device: naa.6006016055711d00cff95e65664ee011
   Device Display Name: DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011)
   Group State: standby
   Array Priority: 0
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {non-current path; rank: 0}

fc.2000001b321734c9:2100001b321734c9-fc.50060160c1e06522:5006016041e06522-naa.6006016055711d00cff95e65664ee011
   Runtime Name: vmhba2:C0:T0:L1
   Device: naa.6006016055711d00cff95e65664ee011
   Device Display Name: DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011)
   Group State: active
   Array Priority: 1
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {current path; rank: 0}
~ #
```

Figure 5.17 Listing pathnames to an FC-attached device

You may use the verbose version of the command shown in Figure 5.17 by using `--device` instead of `-d`.

From the outputs of Figure 5.16 and 5.17, LUN 1 has four paths.

Using the Runtime Name, the list of paths to LUN1 is

- `vmhba3:C0:T1:L1`
- `vmhba3:C0:T0:L1`

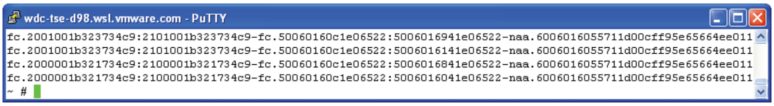
- `vmhba2:C0:T1:L1`
- `vmhba2:C0:T0:L1`

This translates to the list shown in Figure 5.18 based on the physical pathnames. This output was collected using this command:

```
esxcli storage nmp path list -d naa.6006016055711d00cff95e65664ee011 |grep
fc
```

Or the verbose option using the following:

```
esxcli storage nmp path list --device naa.6006016055711d00cff95e65664ee011
|grep fc
```



```
fc.2001001b323734c9:2101001b323734c9-fc.50060160c1e06522:5006016941e06522-naa.6006016055711d00cff95e65664ee011
fc.2001001b323734c9:2101001b323734c9-fc.50060160c1e06522:5006016141e06522-naa.6006016055711d00cff95e65664ee011
fc.2000001b321734c9:2100001b321734c9-fc.50060160c1e06522:5006016841e06522-naa.6006016055711d00cff95e65664ee011
fc.2000001b321734c9:2100001b321734c9-fc.50060160c1e06522:5006016041e06522-naa.6006016055711d00cff95e65664ee011
~ #
```

Figure 5.18 Listing physical pathnames of an FC-attached LUN

This output is similar to the aggregate of all paths that would have been identified using the corresponding UI procedure earlier in this section.

Using Table 2.1, "Identifying SP port association with each SP," in Chapter 2, we can translate the targets listed in the four paths as shown in Table 5.3:

Table 5.3 Identifying SP Port for LUN Paths

Runtime Name	Target WWPN	Sp Port Association
vmhba3:C0:T1:L1	5006016941e06522	SPB1
vmhba3:C0:T0:L1	5006016141e06522	SPA1
vmhba2:C0:T1:L1	5006016841e06522	SPB0
vmhba2:C0:T0:L1	5006016041e06522	SPA0

Identifying Path States and on Which Path the I/O Is Sent—FC

Still using the FC example (refer to Figure 5.17), two fields are relevant to the task of identifying the path states and the I/O path: Group State and Path Selection Policy Path Config. Table 5.4 shows the values of these fields and their meanings.

Table 5.4 Path State Related Fields

Runtime Name	Group State	PSP Path Config	Meaning
vmhba3:C0:T1:L1	Standby	non-current path; rank: 0	Passive SP—no I/O
vmhba3:C0:T0:L1	Active	non-current path; rank: 0	Active-SP—no I/O
vmhba2:C0:T1:L1	Standby	non-current path; rank: 0	Passive SP—no I/O
vmhba2:C0:T0:L1	Active	current path; rank: 0	Active SP—I/O

Combining the last two tables, we can extrapolate the following:

- The LUN is currently owned by SPA (therefore the state is Active).
- The I/O to the LUN is sent via the path to SPA Port 0.

NOTE

This information is provided by the PSP path configuration because its function is to "Determine on which physical path to issue I/O requests being sent to a given storage device" as stated under the PSP section.

The rank configuration listed here shows the value of 0. I discuss the ranked I/O in Chapter 7.

Example of Listing Paths to an iSCSI-Attached Device

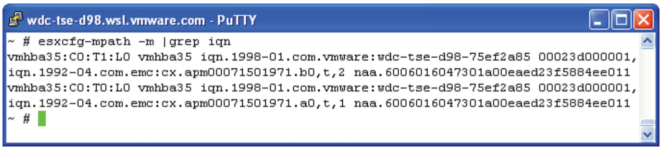
To list paths to a specific iSCSI-attached LUN, try a different approach for locating the device ID:

```
esxcfg-mpath -m |grep iqn
```

You can also use the verbose command option:

```
esxcfg-mpath --list-map |grep iqn
```

The output for this command is shown in Figure 5.19.



```
~ # esxcfg-mpath -m |grep iqn
vmhba35:C0:T1:L0 vmhba35 iqn.1998-01.com.vmware:wdc-tse-d98-75ef2a85 00023d000001,
iqn.1992-04.com.emc:cx.apm00071501971.b0,t,2 naa.6006016047301a00eaed23f5884ee011
vmhba35:C0:T0:L0 vmhba35 iqn.1998-01.com.vmware:wdc-tse-d98-75ef2a85 00023d000001,
iqn.1992-04.com.emc:cx.apm00071501971.a0,t,1 naa.6006016047301a00eaed23f5884ee011
~ #
```

Figure 5.19 Listing paths to an iSCSI-attached LUN via the CLI

In the output, the lines wrapped. Each line actually begins with vmhba35 for readability. From this ouput, we have the information listed in Table 5.5.

Table 5.5 Matching Runtime Names with Their NAA IDs

Runtime Name	NAA ID
vmhba35:C0:T1:L0	naa.6006016047301a00eaed23f5884ee011
vmhba35:C0:T0:L0	naa.6006016047301a00eaed23f5884ee011

This means that these two paths are to the same LUN 0 and the NAA ID is `naa.6006016047301a00eaed23f5884ee011`.

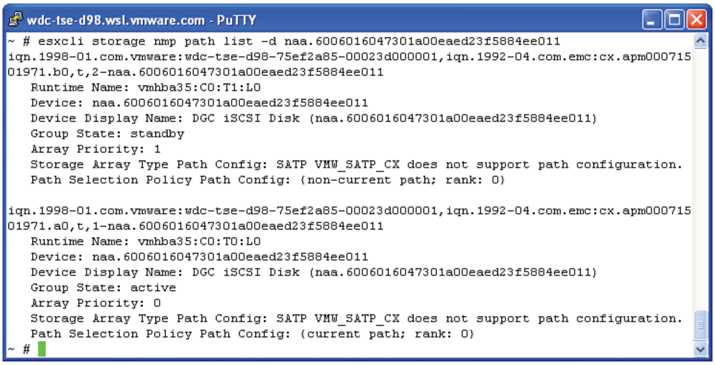
Now, get the pathnames for this LUN. The command is the same as what you used for listing the FC device:

```
esxcli storage nmp path list -d naa.6006016047301a00eaed23f5884ee011
```

You may also use the verbose version of this command:

```
esxcli storage nmp path list --device naa.6006016047301a00eaed23f5884ee011
```

The output is shown in Figure 5.20.



```
~ # esxcli storage nmp path list -d naa.6006016047301a00eaed23f5884ee011
iqn.1998-01.com.vmware:wdc-tse-d98-75ef2a85-00023d000001,iqn.1992-04.com.emc:cx.apm000715
01971.b0,t,2-naa.6006016047301a00eaed23f5884ee011
   Runtime Name: vmhba35:C0:T1:L0
   Device: naa.6006016047301a00eaed23f5884ee011
   Device Display Name: DGC iSCSI Disk (naa.6006016047301a00eaed23f5884ee011)
   Group State: standby
   Array Priority: 1
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {non-current path; rank: 0}

iqn.1998-01.com.vmware:wdc-tse-d98-75ef2a85-00023d000001,iqn.1992-04.com.emc:cx.apm000715
01971.a0,t,1-naa.6006016047301a00eaed23f5884ee011
   Runtime Name: vmhba35:C0:T0:L0
   Device: naa.6006016047301a00eaed23f5884ee011
   Device Display Name: DGC iSCSI Disk (naa.6006016047301a00eaed23f5884ee011)
   Group State: active
   Array Priority: 0
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {current path; rank: 0}
~ #
```

Figure 5.20 Listing paths to an iSCSI-attached LUN via CLI

Note that the path name was wrapped for readability.

Similar to what you observed with the FC-attached devices, the output is identical except for the actual path name. Here, it starts with iqn instead of fc.

The Group State and Path Selection Policy Path Config shows similar content as well. Based on that, I built Table 5.6.

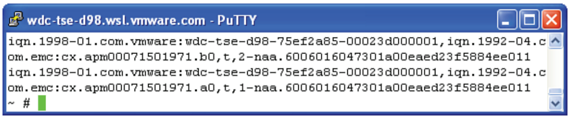
Table 5.6 Matching Runtime Names with Their Target IDs and SP Ports

Runtime Name	Target IQN	Sp Port Association
vmhba35:C0:T1:L0	iqn.1992-04.com.emc:cx.apm00071501971.b0	SPB0
vmhba35:C0:T0:L0	iqn.1992-04.com.emc:cx.apm00071501971.a0	SPA0

To list only the pathnames in the output shown in Figure 5.20, you may append `|grep iqn` to the command.

The output of the command is listed in Figure 5.21 and was wrapped for readability. Each path name starts with `iqn`:

```
esxcli storage nmp path list --device naa.6006016047301a00eaed23f5884ee011
|grep iqn
```



```
iqn.1998-01.com.vmware:wdc-tse-d98-75ef2a85-00023d000001,iqn.1992-04.c
om.emc:cx.apm00071501971.b0,t,2-naa.6006016047301a00eaed23f5884ee011
iqn.1998-01.com.vmware:wdc-tse-d98-75ef2a85-00023d000001,iqn.1992-04.c
om.emc:cx.apm00071501971.a0,t,1-naa.6006016047301a00eaed23f5884ee011
~ #
```

Figure 5.21 Listing pathnames of iSCSI-attached LUNs

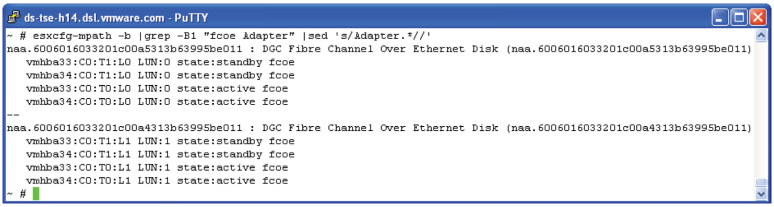
Identifying Path States and on Which Path the I/O Is Sent—iSCSI

The process of identifying path states and I/O path for iSCSI protocol is identical to that of the FC protocol listed in the preceding section.

Example of Listing Paths to an FCoE-Attached Device

The process of listing paths to FCoE-attached devices is identical to the process for FC except that the string you use is `fcoe Adapter` instead of `fc Adapter`.

A sample output from an FCoE configuration is shown in Figure 5.22.



```
~ # esxcfg-mpath -b |grep -B1 "fcoe Adapter" |sed 's/Adapter.*//'
naa.6006016033201c00a5313b63995be011 : DGC Fibre Channel Over Ethernet Disk (naa.6006016033201c00a5313b63995be011)
   vmhba33:C0:T1:L0 LUN:0 state:standby fcoe
   vmhba34:C0:T1:L0 LUN:0 state:standby fcoe
   vmhba33:C0:T0:L0 LUN:0 state:active fcoe
   vmhba34:C0:T0:L0 LUN:0 state:active fcoe
--
naa.6006016033201c00a4313b63995be011 : DGC Fibre Channel Over Ethernet Disk (naa.6006016033201c00a4313b63995be011)
   vmhba33:C0:T1:L1 LUN:1 state:standby fcoe
   vmhba34:C0:T1:L1 LUN:1 state:standby fcoe
   vmhba33:C0:T0:L1 LUN:1 state:active fcoe
   vmhba34:C0:T0:L1 LUN:1 state:active fcoe
~ #
```

Figure 5.22 List of runtime paths of FCoE-attached LUNs via CLI

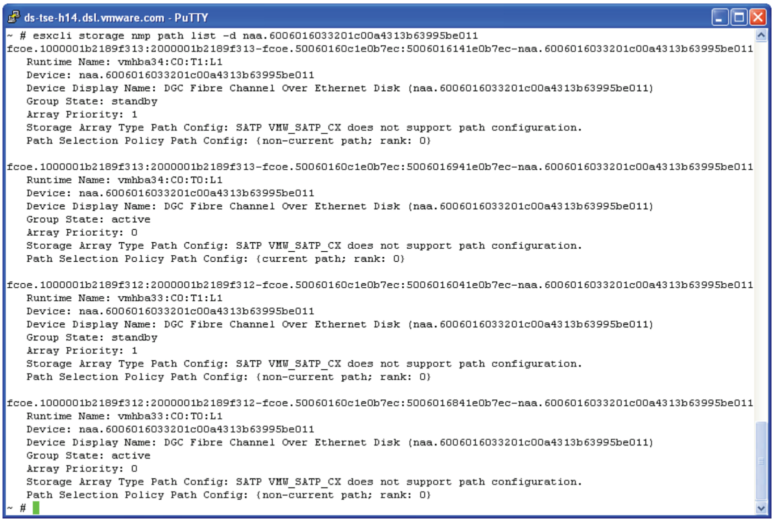
The command used is the following:

```
esxcfg-mpath -b |grep -B1 “fcoe Adapter” |sed ‘s/Adapter.*//’
```

You may also use the verbose command:

```
esxcfg-mpath --list-paths |grep -B1 “fcoe Adapter” |sed ‘s/Adapter.*//’
```

Using the NAA ID for LUN 1, the list of pathnames is shown in Figure 5.23.



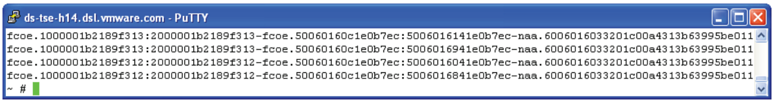
ds-tse-h14.dsl.vmware.com - PuTTY

```
~ # esxcli storage nmp path list -d naa.6006016033201c00a4313b63995be011
fcoe.1000001b2189f313:2000001b2189f313-fcoe.50060160c1e0b7ec:5006016141e0b7ec-naa.6006016033201c00a4313b63995be011
   Runtime Name: vmhba34:C0:T1:L1
   Device: naa.6006016033201c00a4313b63995be011
   Device Display Name: DGC Fibre Channel Over Ethernet Disk (naa.6006016033201c00a4313b63995be011)
   Group State: standby
   Array Priority: 1
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {non-current path; rank: 0}

fcoe.1000001b2189f313:2000001b2189f313-fcoe.50060160c1e0b7ec:5006016941e0b7ec-naa.6006016033201c00a4313b63995be011
   Runtime Name: vmhba34:C0:T0:L1
   Device: naa.6006016033201c00a4313b63995be011
   Device Display Name: DGC Fibre Channel Over Ethernet Disk (naa.6006016033201c00a4313b63995be011)
   Group State: active
   Array Priority: 0
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {current path; rank: 0}

fcoe.1000001b2189f312:2000001b2189f312-fcoe.50060160c1e0b7ec:5006016041e0b7ec-naa.6006016033201c00a4313b63995be011
   Runtime Name: vmhba33:C0:T1:L1
   Device: naa.6006016033201c00a4313b63995be011
   Device Display Name: DGC Fibre Channel Over Ethernet Disk (naa.6006016033201c00a4313b63995be011)
   Group State: standby
   Array Priority: 1
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {non-current path; rank: 0}

fcoe.1000001b2189f312:2000001b2189f312-fcoe.50060160c1e0b7ec:5006016841e0b7ec-naa.6006016033201c00a4313b63995be011
   Runtime Name: vmhba33:C0:T0:L1
   Device: naa.6006016033201c00a4313b63995be011
   Device Display Name: DGC Fibre Channel Over Ethernet Disk (naa.6006016033201c00a4313b63995be011)
   Group State: active
   Array Priority: 0
   Storage Array Type Path Config: SATP VMW_SATP_CX does not support path configuration.
   Path Selection Policy Path Config: {non-current path; rank: 0}
~ #
```

Figure 5.23 List of pathnames of an FCoE-attached LUN

You may also use the verbose version of the command shown in Figure 5.23 by using `--device` instead of `-d`.

This translates to the physical pathnames shown in Figure 5.24.



ds-tse-h14.dsl.vmware.com - PuTTY

```
fcoe.1000001b2189f313:2000001b2189f313-fcoe.50060160c1e0b7ec:5006016141e0b7ec-naa.6006016033201c00a4313b63995be011
fcoe.1000001b2189f313:2000001b2189f313-fcoe.50060160c1e0b7ec:5006016941e0b7ec-naa.6006016033201c00a4313b63995be011
fcoe.1000001b2189f312:2000001b2189f312-fcoe.50060160c1e0b7ec:5006016041e0b7ec-naa.6006016033201c00a4313b63995be011
fcoe.1000001b2189f312:2000001b2189f312-fcoe.50060160c1e0b7ec:5006016841e0b7ec-naa.6006016033201c00a4313b63995be011
~ #
```

Figure 5.24 List of paths names of an FCoE LUN

The command used to collect the ouput shown in Figure 5.24 is

```
esxcli storage nmp path list -d 6006016033201c00a4313b63995be011 |grep fcoe
```

Using Table 2.1, “Identifying SP Port Association with Each SP,” in Chapter 2, you can translate the targets listed in the returned four paths as shown in Table 5.7.

Table 5.7 Translation of FCoE Targets

Runtime Name	Target WWPN	SP Port Association
vmhba34:C0:T1:L1	5006016141e0b7ec	SPA1
vmhba34:C0:T0:L1	5006016941e0b7ec	SPB1
vmhba33:C0:T1:L1	5006016041e0b7ec	SPA0
vmhba33:C0:T0:L1	5006016841e0b7ec	SPB0

Identifying Path States and on Which Path the I/O Is Sent—FC

Still following the process as you did with the FC example (refer to Figure 5.17), two fields are relevant to the task of identifying the path states and the I/O path: Group State and Path Selection Policy Path Config. Table 5.8 shows the values of these fields and their meaning.

Table 5.8 Interpreting Path States—FCoE

Runtime Name	Group State	PSP Path Config	Meaning
vmhba34:C0:T1:L1	Standby	non-current path; rank: 0	Passive SP — no I/O
vmhba34:C0:T0:L1	Active	current path; rank: 0	Active-SP — I/O
vmhba33:C0:T1:L1	Standby	non-current path; rank: 0	Passive SP — no I/O
vmhba33:C0:T0:L1	Active	non-current path; rank: 0	Active SP — no I/O

Combining the last two tables, we can extrapolate the following:

- The LUN is currently “owned” by SPB (hence the state is Active).
- The I/O to the LUN is sent via the path to SPB Port 1.

Claim Rules

Each storage device is managed by one of the PSA plug-ins at any given time. In other words, a device cannot be managed by more than one PSA plug-in.

For example, a host that has a third-party MPP installed alongside with NMP, devices managed by the third-party MPP cannot be managed by NMP unless the configuration is changed to assign these devices to NMP. The process of associating certain devices with certain PSA plug-ins is referred to as *claiming* and is defined by Claim Rules. These rules define the correlation between a device and NMP or MPP. NMP has additional association between the claimed device and a specific SATP and PSP.

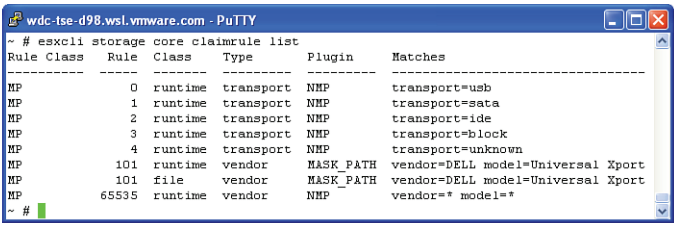
This section shows you how to list the various claim rules. The next section discusses how to change these rules.

Claim rules can be defined based on one or a combination of the following:

- **Vendor String**—In response to the standard inquiry command, the arrays return the standard inquiry response, which includes the Vendor string. This can be used in the definition of a claim rule based on the exact match. A partial match or a string with padded spaces does not work.
- **Model String**—Similar to the Vendor string, the Model string is returned as part of the standard inquiry response. Similar to the Vendor string, a claim rule can be defined using the exact match of the Model string and padded spaces are not supported here.
- **Transport**—Defining a claim rule based on the transport type, Transport facilitates claiming of all devices that use that transport. Valid transport types are block, fc, iscsi, iscsivendor, ide, sas, sata, usb, parallel, and unknown.
- **Driver**—Specifying a driver name as one of the criteria for a claim rule definition allows all devices accessible via such a driver to be claimed. An example of that is a claim rule to mask all paths to devices attached to an HBA that uses mptscsi driver.

MP Claim Rules

The first set of claim rules defines which MPP claims which devices. Figure 5.25 shows the default MP claim rules.



```
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

Figure 5.25 Listing MP Claim Rules

The command to list these rules is

```
esxcli storage core claimrule list
```

The namespace here is for the Core Storage because the MPP definition is done on the PSA level. The output shows that this rule class is MP, which indicates that these rules define the devices' association to a specific multipathing plug-in.

There are two plugins specified here: NMP and MASK_PATH. I have already discussed NMP in the previous sections. The MASK_PATH plug-in is used for masking paths to specific devices and is a replacement for the deprecated Legacy Multipathing LUN Masking vmkernel parameter. I provide some examples in the "Modifying PSA Plug-in Configurations Using the UI" section.

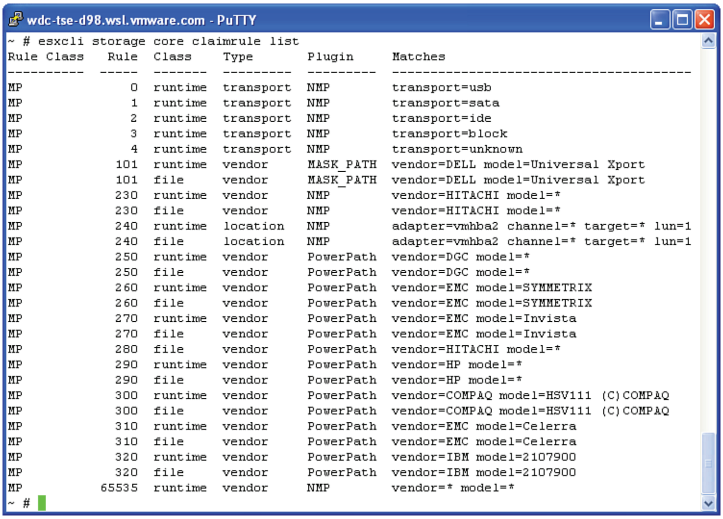
Table 5.9 lists each column name in the ouput along with an explanation of each column.

Table 5.9 Explanation of Claim Rules Fields

Column Name	Explanation
Rule Class	The plugin class for which this claim rule set is defined. This can be MP, Filter, or VAAI.
Rule	The rule number. This defines the order the rules are loaded. Similar to firewall rules, the first match is used and supersedes rules with larger numbers.
Class	The value can be `runtime` or `file`. A value of `file` means that the rule definitions were stored to the configuration files (more on this later in this section). A value of `Runtime` means that the rule was read from the configuration files and loaded into memory. In other words, it means that the rule is active. If a rule is listed as `file` only and no `runtime`, the rule was just created but has not been loaded yet. Find out more about loading rules in the next section.
Type	The type can be `vendor`, `model`, `transport`, or `driver`. See the explanation in the "Claim Rules" section.
Plugin	The name of the plug-in for which this rule was defined.
Matches	This is the most important field in the rule definition. This column shows the "Type" specified for the rule and its value. When the specified type is `vendor`, an additional parameter, `model`, must be used. The `model` string must be an exact string match or include an `*` as a wild card. You may use a `^` as "begins with" and then the string followed by an `*`—for example, `^OPEN-*`.

The highest rule number in any claim rules set is 65535. It is assigned here to a Catch-All rule that claims devices from “any” vendor with “any” model string. It is placed as the last rule in the set to allow for lower numbered rules to claim their specified devices. If the attached devices have no specific rules defined, they get claimed by NMP.

Figure 5.26 is an example of third-party MP plug-in claim rules.



```
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            230  runtime  vendor     NMP        vendor=HITACHI model=*
MP            230  file     vendor     NMP        vendor=HITACHI model=*
MP            240  runtime  location   NMP        adapter=vmhba2 channel=* target=* lun=1
MP            240  file     location   NMP        adapter=vmhba2 channel=* target=* lun=1
MP            250  runtime  vendor     PowerPath  vendor=DGC model=*
MP            250  file     vendor     PowerPath  vendor=DGC model=*
MP            260  runtime  vendor     PowerPath  vendor=EMC model=SYMMETRIX
MP            260  file     vendor     PowerPath  vendor=EMC model=SYMMETRIX
MP            270  runtime  vendor     PowerPath  vendor=EMC model=Invista
MP            270  file     vendor     PowerPath  vendor=EMC model=Invista
MP            280  file     vendor     PowerPath  vendor=HITACHI model=*
MP            290  runtime  vendor     PowerPath  vendor=HP model=*
MP            290  file     vendor     PowerPath  vendor=HP model=*
MP            300  runtime  vendor     PowerPath  vendor=COMPAQ model=HSV111 (C)COMPAQ
MP            300  file     vendor     PowerPath  vendor=COMPAQ model=HSV111 (C)COMPAQ
MP            310  runtime  vendor     PowerPath  vendor=EMC model=Celerra
MP            310  file     vendor     PowerPath  vendor=EMC model=Celerra
MP            320  runtime  vendor     PowerPath  vendor=IBM model=2107900
MP            320  file     vendor     PowerPath  vendor=IBM model=2107900
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

Figure 5.26 Listing EMC PowerPath/VE claim rules.

Here you see that rules number 250 through 320 were added by PowerPath/VE, which allows PowerPath plug-in to claim all the devices listed in Table 5.10.

Table 5.10 Arrays Claimed by PowerPath

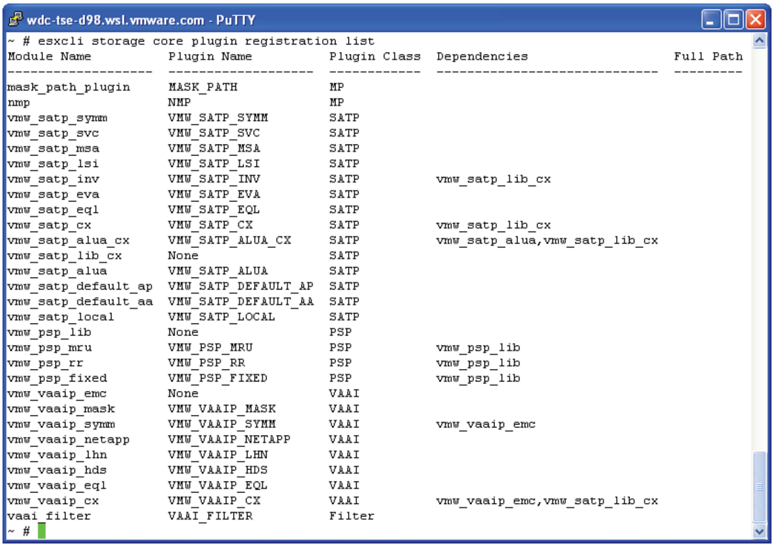
Storage Array	Vendor	Model
EMC CLARiiON Family	DGC	Any (* is a wild card)
EMC Symmetrix Family	EMC	SYMMETRIX
EMC Invista	EMC	Invista
HITACHI	HITACHI	Any
HP	HP	Any
HP EVA HSV111 family (Compaq Branded)	HP	HSV111 (C) COMPAQ
EMC Celerra	EMC	Celerra
IBM DS8000 family	IBM	2107900

NOTE

There is currently a known limitation with claim rules that use a partial match on the model string. So, older versions of PowerPath/VE that used to have rules stating model=OPEN may not claim the devices whose model string is something such as OPEN-V, OPEN-10, and so on. As evident from Figure 5.26, version 5.7 no longer uses partial matches. Instead, partial matches have been replaced with an `*`.

Plug-in Registration

New to vSphere 5 is the concept of *plug-in registration*. Actually this existed in 4.x but was not exposed to the end user. When a PSA plug-in is installed, it gets registered with the PSA framework along with their dependencies, if any, similar to the output in Figure 5.27.



```
~ # esxcli storage core plugin registration list
Module Name          Plugin Name          Plugin Class  Dependencies                   Full Path
-------------------  -------------------  ------------  -----------------------------  ---------
mask_path_plugin     MASK_PATH            MP
nmp                  NMP                  MP
vmw_satp_symm        VMW_SATP_SYMM        SATP
vmw_satp_svc         VMW_SATP_SVC         SATP
vmw_satp_msa         VMW_SATP_MSA         SATP
vmw_satp_lsi         VMW_SATP_LSI         SATP
vmw_satp_inv         VMW_SATP_INV         SATP          vmw_satp_lib_cx
vmw_satp_eva         VMW_SATP_EVA         SATP
vmw_satp_eql         VMW_SATP_EQL         SATP
vmw_satp_cx          VMW_SATP_CX          SATP          vmw_satp_lib_cx
vmw_satp_alua_cx     VMW_SATP_ALUA_CX     SATP          vmw_satp_alua,vmw_satp_lib_cx
vmw_satp_lib_cx      None                 SATP
vmw_satp_alua        VMW_SATP_ALUA        SATP
vmw_satp_default_ap  VMW_SATP_DEFAULT_AP  SATP
vmw_satp_default_aa  VMW_SATP_DEFAULT_AA  SATP
vmw_satp_local       VMW_SATP_LOCAL       SATP
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
vmw_vaaip_emc        None                 VAAI
vmw_vaaip_mask       VMW_VAAIP_MASK       VAAI
vmw_vaaip_symm       VMW_VAAIP_SYMM       VAAI          vmw_vaaip_emc
vmw_vaaip_netapp     VMW_VAAIP_NETAPP     VAAI
vmw_vaaip_lhn        VMW_VAAIP_LHN        VAAI
vmw_vaaip_hds        VMW_VAAIP_HDS        VAAI
vmw_vaaip_eql        VMW_VAAIP_EQL        VAAI
vmw_vaaip_cx         VMW_VAAIP_CX         VAAI          vmw_vaaip_emc,vmw_satp_lib_cx
vaai_filter          VAAI_FILTER          Filter
~ #
```

Figure 5.27 Listing PSA plug-in registration

This output shows the following:

- **Module Name**—The name of the plug-in kernel module; this is the actual plug-in software binary as well as required libraries, if any, that get plugged into vmkernel.

- **Plugin Name**—This is the name by which the plug-in is identified. This is the exact name to use when creating or modifying claim rules.
- **Plugin class**—This is the name of the class to which the plug-in belongs. For example, the previous section covered the MP class of plug-ins. The next sections discuss SATP and PSP plug-ins and later chapters cover VAAI and VAAI_Filter classes.
- **Dependencies**—These are the libraries and other plug-ins which the registered plug-ins require to operate.
- **Full Path**—This is the full path to the files, libraries, or binaries that are specific to the registered plug-in. This is mostly blank in the default registration.

SATP Claim Rules

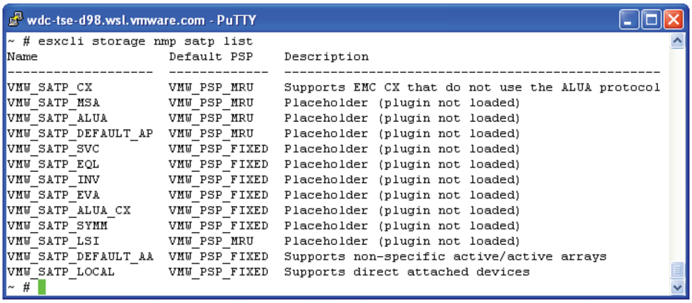
Now that you understand how NMP plugs into PSA, it's time to examine how SATP plugs into NMP.

Each SATP is associated with a default PSP. The defaults can be overridden using SATP claim rules. Before I show you how to list these rules, first review the default settings.

The command used to list the default PSP assignment to each SATP is

```
esxcli storage nmp satp list
```

The output of this command is shown in Figure 5.28.



```
~ # esxcli storage nmp satp list
Name                 Default PSP    Description
-------------------  -------------  -------------------------------------------------
VMW_SATP_CX          VMW_PSP_MRU    Supports EMC CX that do not use the ALUA protocol
VMW_SATP_MSA         VMW_PSP_MRU    Placeholder (plugin not loaded)
VMW_SATP_ALUA        VMW_PSP_MRU    Placeholder (plugin not loaded)
VMW_SATP_DEFAULT_AP  VMW_PSP_MRU    Placeholder (plugin not loaded)
VMW_SATP_SVC         VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_EQL         VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_INV         VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_EVA         VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_ALUA_CX     VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_SYMM        VMW_PSP_FIXED  Placeholder (plugin not loaded)
VMW_SATP_LSI         VMW_PSP_MRU    Placeholder (plugin not loaded)
VMW_SATP_DEFAULT_AA  VMW_PSP_FIXED  Supports non-specific active/active arrays
VMW_SATP_LOCAL       VMW_PSP_FIXED  Supports direct attached devices
~ #
```

Figure 5.28 Listing SATPs and their default PSPs

The name space is Storage, NMP, and finally SATP.

NOTE

VMW_SATP_ALUA_CX plug-in is associated with VMW_PSP_FIXED. Starting with vSphere 5.0, the functionality of VMW_PSP_FIXED_AP has been rolled into VMW_PSP_FIXED. This facilitates the use of the Preferred Path option with ALUA arrays while still handling failover triggering events in a similar fashion to Active/Passive arrays. Read more on this in Chapter 6.

Knowing which PSP is the default policy for which SATP is half the story. NMP needs to know which SATP it will use with which storage device. This is done via SATP claim rules that associate a given SATP with a storage device based on matches to Vendor, Model, Driver, and/or Transport.

To list the SATP rule, run the following:

```
esxcli storage nmp satp rule list
```

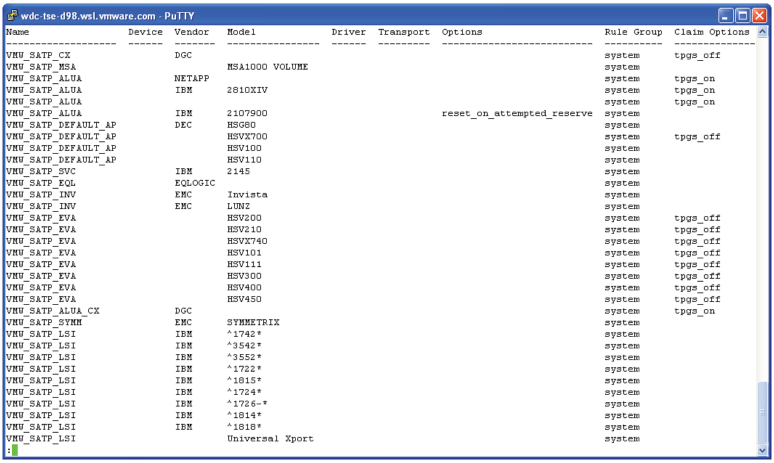
The output of the command is too long and too wide to capture in one screenshot. I have divided the output to a set of images in which I list a partial output then list the text of the full output in a subsequent table. Figures 5.29, 5.30, 5.31, and 5.32 show the four quadrants of the output.

TIP

To format the output of the preceding command so that the text is arranged better for readability, you can pipe the output to `less -S`. This truncates the long lines and aligns the text under their corresponding columns.

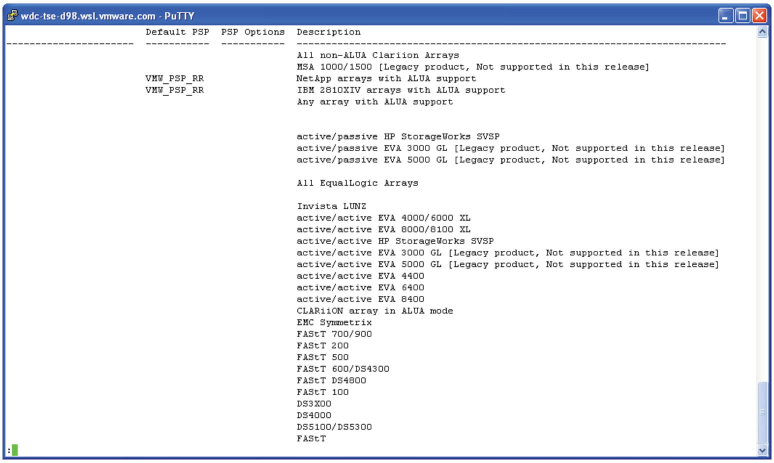
So, the command would look like this:

```
esxcli storage nmp satp list | less -S
```



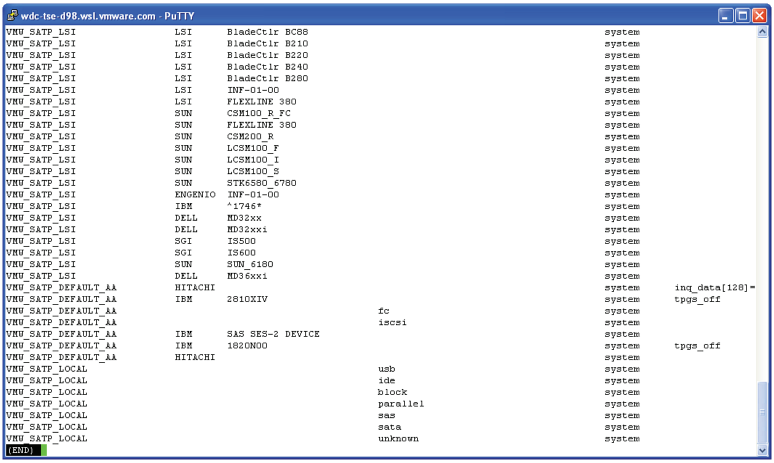
```
Name                 Device  Vendor   Model            Driver  Transport  Options                     Rule Group  Claim Options
-------------------  ------  -------  ----------------  ------  ---------  --------------------------  ----------  --------------
VMW_SATP_CX                  DGC                                                                       system      tpgs_off
VMW_SATP_MSA                          MSA1000 VOLUME                                                   system
VMW_SATP_ALUA                NETAPP                                                                    system      tpgs_on
VMW_SATP_ALUA                IBM      2810XIV                                                          system      tpgs_on
VMW_SATP_ALUA                                                                                          system      tpgs_on
VMW_SATP_ALUA                IBM      2107900                      reset_on_attempted_reserve  system
VMW_SATP_DEFAULT_AP          DEC      HSG80                                                            system
VMW_SATP_DEFAULT_AP                   HSVX700                                                          system      tpgs_off
VMW_SATP_DEFAULT_AP                   HSV100                                                           system
VMW_SATP_DEFAULT_AP                   HSV110                                                           system
VMW_SATP_SVC                 IBM      2145                                                             system
VMW_SATP_EQL                 EQLOGIC                                                                   system
VMW_SATP_INV                 EMC      Invista                                                          system
VMW_SATP_INV                 EMC      LUNZ                                                             system
VMW_SATP_EVA                          HSV200                                                           system      tpgs_off
VMW_SATP_EVA                          HSV210                                                           system      tpgs_off
VMW_SATP_EVA                          HSVX740                                                          system      tpgs_off
VMW_SATP_EVA                          HSV101                                                           system      tpgs_off
VMW_SATP_EVA                          HSV111                                                           system      tpgs_off
VMW_SATP_EVA                          HSV300                                                           system      tpgs_off
VMW_SATP_EVA                          HSV400                                                           system      tpgs_off
VMW_SATP_EVA                          HSV450                                                           system      tpgs_off
VMW_SATP_ALUA_CX             DGC                                                                       system      tpgs_on
VMW_SATP_SYMM                EMC      SYMMETRIX                                                        system
VMW_SATP_LSI                 IBM      ^1742*                                                           system
VMW_SATP_LSI                 IBM      ^3542*                                                           system
VMW_SATP_LSI                 IBM      ^3552*                                                           system
VMW_SATP_LSI                 IBM      ^1722*                                                           system
VMW_SATP_LSI                 IBM      ^1815*                                                           system
VMW_SATP_LSI                 IBM      ^1724*                                                           system
VMW_SATP_LSI                 IBM      ^1726-*                                                          system
VMW_SATP_LSI                 IBM      ^1814*                                                           system
VMW_SATP_LSI                 IBM      ^1818*                                                           system
VMW_SATP_LSI                          Universal Xport                                                  system
:
```

Figure 5.29 Listing SATP claim rules—top-left quadrant of output.



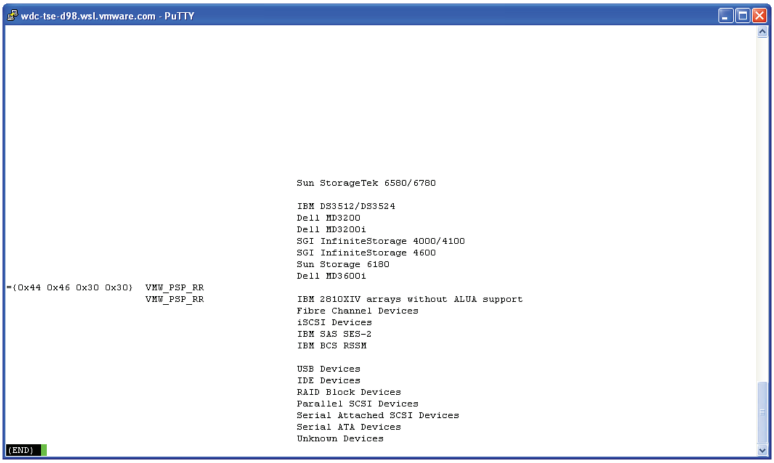
```
                        Default PSP  PSP Options  Description
----------------------  -----------  -----------  --------------------------------------------------------------------------
                                                  All non-ALUA Clariion Arrays
                                                  MSA 1000/1500 [Legacy product, Not supported in this release]
                        VMW_PSP_RR                NetApp arrays with ALUA support
                        VMW_PSP_RR                IBM 2810XIV arrays with ALUA support
                                                  Any array with ALUA support


                                                  active/passive HP StorageWorks SVSP
                                                  active/passive EVA 3000 GL [Legacy product, Not supported in this release]
                                                  active/passive EVA 5000 GL [Legacy product, Not supported in this release]

                                                  All EqualLogic Arrays

                                                  Invista LUNZ
                                                  active/active EVA 4000/6000 XL
                                                  active/active EVA 8000/8100 XL
                                                  active/active HP StorageWorks SVSP
                                                  active/active EVA 3000 GL [Legacy product, Not supported in this release]
                                                  active/active EVA 5000 GL [Legacy product, Not supported in this release]
                                                  active/active EVA 4400
                                                  active/active EVA 6400
                                                  active/active EVA 8400
                                                  CLARiiON array in ALUA mode
                                                  EMC Symmetrix
                                                  FAStT 700/900
                                                  FAStT 200
                                                  FAStT 500
                                                  FAStT 600/DS4300
                                                  FAStT DS4800
                                                  FAStT 100
                                                  DS3X00
                                                  DS4000
                                                  DS5100/DS5300
                                                  FAStT
:
```

Figure 5.30 Listing SATP claim rules—top-right quadrant of output.



```
VMW_SATP_LSI                 LSI      BladeCtlr BC88                                                   system
VMW_SATP_LSI                 LSI      BladeCtlr B210                                                   system
VMW_SATP_LSI                 LSI      BladeCtlr B220                                                   system
VMW_SATP_LSI                 LSI      BladeCtlr B240                                                   system
VMW_SATP_LSI                 LSI      BladeCtlr B280                                                   system
VMW_SATP_LSI                 LSI      INF-01-00                                                        system
VMW_SATP_LSI                 LSI      FLEXLINE 380                                                     system
VMW_SATP_LSI                 SUN      CSM100_R_FC                                                      system
VMW_SATP_LSI                 SUN      FLEXLINE 380                                                     system
VMW_SATP_LSI                 SUN      CSM200_R                                                         system
VMW_SATP_LSI                 SUN      LCSM100_F                                                        system
VMW_SATP_LSI                 SUN      LCSM100_I                                                        system
VMW_SATP_LSI                 SUN      LCSM100_S                                                        system
VMW_SATP_LSI                 SUN      STK6580_6780                                                     system
VMW_SATP_LSI                 ENGENIO  INF-01-00                                                        system
VMW_SATP_LSI                 IBM      ^1746*                                                           system
VMW_SATP_LSI                 DELL     MD32xx                                                           system
VMW_SATP_LSI                 DELL     MD32xxi                                                          system
VMW_SATP_LSI                 SGI      IS500                                                            system
VMW_SATP_LSI                 SGI      IS600                                                            system
VMW_SATP_LSI                 SUN      SUN_6180                                                         system
VMW_SATP_LSI                 DELL     MD36xxi                                                          system
VMW_SATP_DEFAULT_AA          HITACHI                                                                   system      inq_data[128]=
VMW_SATP_DEFAULT_AA          IBM      2810XIV                                                          system      tpgs_off
VMW_SATP_DEFAULT_AA                                              fc                                    system
VMW_SATP_DEFAULT_AA                                              iscsi                                 system
VMW_SATP_DEFAULT_AA          IBM      SAS SES-2 DEVICE                                                 system
VMW_SATP_DEFAULT_AA          IBM      1820N00                                                          system      tpgs_off
VMW_SATP_DEFAULT_AA          HITACHI                                                                   system
VMW_SATP_LOCAL                                                   usb                                   system
VMW_SATP_LOCAL                                                   ide                                   system
VMW_SATP_LOCAL                                                   block                                 system
VMW_SATP_LOCAL                                                   parallel                              system
VMW_SATP_LOCAL                                                   sas                                   system
VMW_SATP_LOCAL                                                   sata                                  system
VMW_SATP_LOCAL                                                   unknown                               system
(END)
```

Figure 5.31 Listing SATP claim rules—bottom-left quadrant of output

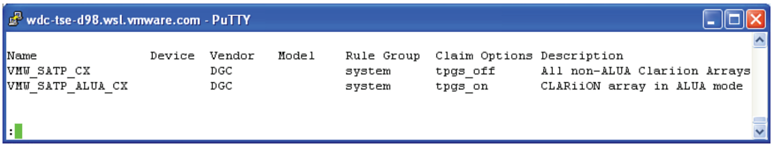


```
                                                  Sun StorageTek 6580/6780

                                                  IBM DS3512/DS3524
                                                  Dell MD3200
                                                  Dell MD3200i
                                                  SGI InfiniteStorage 4000/4100
                                                  SGI InfiniteStorage 4600
                                                  Sun Storage 6180
                                                  Dell MD3600i
={0x44 0x46 0x30 0x30}  VMW_PSP_RR
                        VMW_PSP_RR                IBM 2810XIV arrays without ALUA support
                                                  Fibre Channel Devices
                                                  iSCSI Devices
                                                  IBM SAS SES-2
                                                  IBM BCS RSSM

                                                  USB Devices
                                                  IDE Devices
                                                  RAID Block Devices
                                                  Parallel SCSI Devices
                                                  Serial Attached SCSI Devices
                                                  Serial ATA Devices
                                                  Unknown Devices
(END)
```

Figure 5.32 Listing SATP claim rules—bottom-right quadrant of output

To make things a bit clearer, let's take a couple of lines from the output and explain what they mean.

Figure 5.33 shows the relevant rules for CLARiiON arrays both non-ALUA and ALUA capable. I removed three blank columns (Driver, Transport, and Options) to fit the content on the lines.

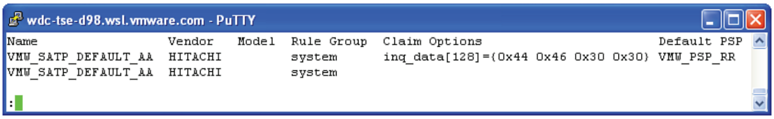


```
Name               Device  Vendor   Model    Rule Group  Claim Options Description
VMW_SATP_CX                DGC               system      tpgs_off      All non-ALUA Clariion Arrays
VMW_SATP_ALUA_CX           DGC               system      tpgs_on       CLARiiON array in ALUA mode

:
```

Figure 5.33 CLARiiON Non-ALUA and ALUA Rules

The two lines show the claim rules for EMC CLARiiON CX family. Using this rule, NMP identifies the array as CLARiiON CX when the Vendor string is DGC. If NMP stopped at this, it would have used VMW_SATP_CX as the SATP for this array. However, this family of arrays can support more than one configuration. That is the reason the value `Claim Options` column comes in handy! So, if that option is `tpgs_off`, NMP uses the VMW_SATP_CX plug-in, and if the option is `tpgs_on`, NMP uses VMW_SATP_ALUA_CX. I explain what these options mean in Chapter 6.

Figure 5.34 shows another example that utilizes additional options. I removed the Device column to fit the content to the display.



```
Name                 Vendor   Model  Rule Group  Claim Options                       Default PSP
VMW_SATP_DEFAULT_AA  HITACHI         system      inq_data[128]={0x44 0x46 0x30 0x30} VMW_PSP_RR
VMW_SATP_DEFAULT_AA  HITACHI         system

:
```

Figure 5.34 Claim rule that uses Claim Options

In this example, NMP uses VMW_SATP_DEFAULT_AA SATP with all arrays returning `HITACHI` as a model string. However, the default PSP is selected based on the values listed in the Claim Options column:

- If the column is blank, the default PSP (which is VMW_PSP_FIXED and is based on the list shown earlier in this section in Figure 5.28) is used. In that list, you see that VMW_SATP_DEFAULT_AA is assigned the default PSP named VMW_PSP_FIXED.
- If the column shows `inq_data[128]={0x44 0x46 0x30 0x30}`, which is part of the data reported from the array via the Inquiry String, NMP overrides the default PSP configuration and uses VMW_PSP_RR instead.

Modifying PSA Plug-in Configurations Using the UI

You can modify PSA plug-ins' configuration using the CLI and, to a limited extent, the UI. Because the UI provides far fewer options for modification, let me address that first to get it out of the way!

Which PSA Configurations Can Be Modified Using the UI?

You can change the PSP for a given device. However, this is done on a LUN level rather than the array.

Are you wondering why you would want to do that?

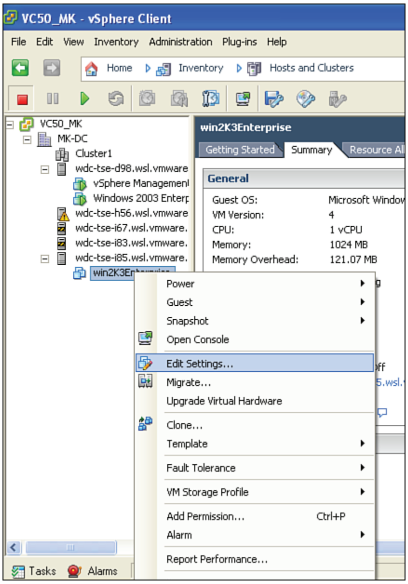
Think of the following scenario:

You have Microsoft Clustering Service (MSCS) cluster nodes in Virtual Machines (VMs) in your environment. The cluster's shared storage is Physical Mode Raw Device Mappings (RDMs), which are also referred to as (Passthrough RDMs). Your storage vendor recommends using Round-Robin Path Selection Policy (VMW_PSP_RR). However, VMware does not support using that policy with the MSCS clusters in shared RDMs.

The best approach is to follow your storage vendor's recommendations for most of the LUNs, but follow the procedure listed here to change just the RDM LUNs' PSP to their default PSPs.

Procedure to Change PSP via UI

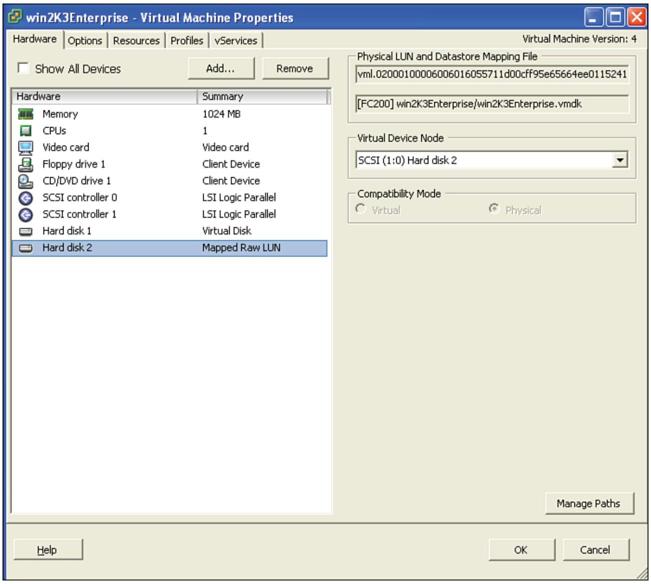
1. Use the vSphere client to navigate to the MSCS node VM and right-click the VM in the inventory pane. Select **Edit Settings** (see Figure 5.35).



Guest OS: Microsoft Windov
VM Version: 4
CPU: 1 vCPU
Memory: 1024 MB
Memory Overhead: 121.07 MB

Figure 5.35 Editing VM's settings via the UI

The resulting dialog is shown in Figure 5.36.



Virtual Machine Version: 4

□ Show All Devices

Hardware	Summary
Memory	1024 MB
CPUs	1
Video card	Video card
Floppy drive 1	Client Device
CD/DVD drive 1	Client Device
SCSI controller 0	LSI Logic Parallel
SCSI controller 1	LSI Logic Parallel
Hard disk 1	Virtual Disk
Hard disk 2	Mapped Raw LUN

Physical LUN and Datastore Mapping File

vml.020001000060060160557 11d00cff95e65664ee0115241

[FC200] win2K3Enterprise/win2K3Enterprise.vmdk

Virtual Device Node

SCSI (1:0) Hard disk 2

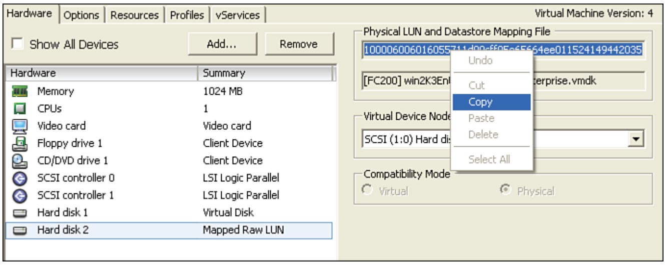
Compatibility Mode

○ Virtual ● Physical

Figure 5.36 Virtual Machine Properties dialog

2. Locate the RDM listed in the Hardware tab. You can identify this by the summary column showing Mapped Raw LUN. On the top right-hand side you can locate the Logical Device Name, which is prefixed with vml in the field labeled Physical LUN and Datastore Mapping File.

3. Double-click the text in that field. Right-click the selected text and click **Copy** (see Figure 5.37).



Virtual Machine Version: 4

□ Show All Devices

Hardware	Summary
Memory	1024 MB
CPUs	1
Video card	Video card
Floppy drive 1	Client Device
CD/DVD drive 1	Client Device
SCSI controller 0	LSI Logic Parallel
SCSI controller 1	LSI Logic Parallel
Hard disk 1	Virtual Disk
Hard disk 2	Mapped Raw LUN

Physical LUN and Datastore Mapping File

1000060060160557 64ee011524149442035

[FC200] win2K3En terprise.vmdk

Virtual Device Node

SCSI (1:0) Hard di

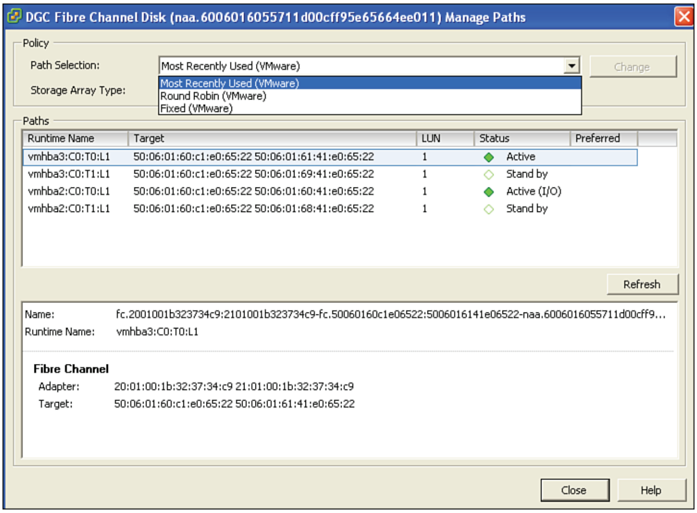
Compatibility Mode

○ Virtual ● Physical

Figure 5.37 Copying RDM's VML ID (Logical Device Name) via the UI

4. I use the copied text to follow Steps 4 and 5 of doing the same task via the CLI in the next section. However, for this section, click the **Manage Paths** button in the dialog shown in Figure 5.37.

   The resulting Manage Paths dialog is shown in Figure 5.38.



DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011) Manage Paths

Policy

Path Selection: Most Recently Used (VMware)

Storage Array Type:

Paths

Runtime Name	Target	LUN	Status	Preferred
vmhba3:C0:T0:L1	50:06:01:60:c1:e0:65:22 50:06:01:61:41:e0:65:22	1	◆ Active	
vmhba3:C0:T1:L1	50:06:01:60:c1:e0:65:22 50:06:01:69:41:e0:65:22	1	◇ Stand by	
vmhba2:C0:T0:L1	50:06:01:60:c1:e0:65:22 50:06:01:60:41:e0:65:22	1	◆ Active (I/O)	
vmhba2:C0:T1:L1	50:06:01:60:c1:e0:65:22 50:06:01:68:41:e0:65:22	1	◇ Stand by	

Name: fc.2001001b323734c9:2101001b323734c9-fc.50060160c1e06522:5006016141e06522-naa.6006016055711d00cff9...

Runtime Name: vmhba3:C0:T0:L1

Fibre Channel

Adapter: 20:01:00:1b:32:37:34:c9 21:01:00:1b:32:37:34:c9

Target: 50:06:01:60:c1:e0:65:22 50:06:01:61:41:e0:65:22

Figure 5.38 Modifying PSP selection via the UI

5. Click the pull-down menu next to the Path Selection field and change it from Round Robin (VMware) to the default PSP for your array. Click the **Change** button. To locate which PSP is the default, check VMware HCL. If the PSP listed there is Round Robin, follow the examples listed in the previous section, “SATP Claim Rules,” to identify which PSP to select.

6. Click **Close**.

Modifying PSA Plug-ins Using the CLI

The CLI provides a range of options to configure, customize, and modify PSA plug-in settings. I provide the various configurable options and their use cases as we go.

Available CLI Tools and Their Options

New to vSphere 5.0 is the expansion of using esxcli as the main CLI utility for managing ESXi 5.0. The same binary is used whether you log on to the host locally or remotely via

SSH. It is also used by vMA or vCLI. This simplifies administrative tasks and improves portability of scripts written to use esxcli.

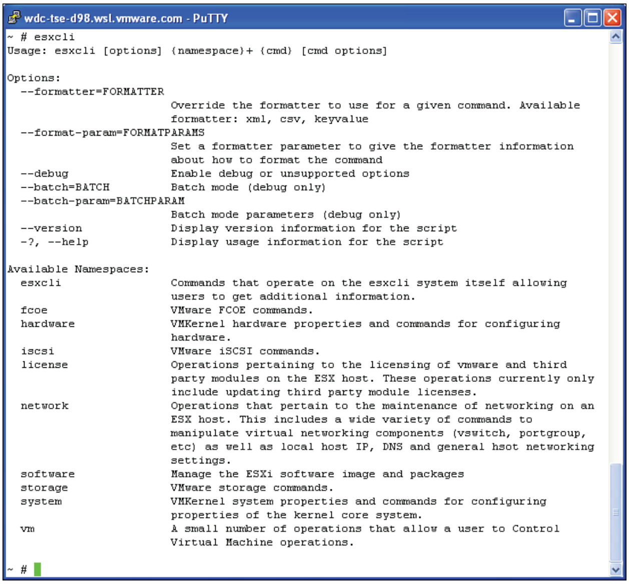
TIP

The only difference between the tools used locally or via SSH compared to those used in vMA and Remote CLI is that the latter two require providing the server name and the user's credentials on the command line. Refer to Chapter 3 in which I covered using the FastPass (fp) facility of vMA and how to add the users' credentials to the CREDSTORE environment variable on vCLI.

Assuming that the server name and user credentials are set in the environment, the command-line syntax in all the examples in this book is identical regardless of where you use them.

ESXCLI Namespace

Figure 5.39 shows the command-line help for esxcli.

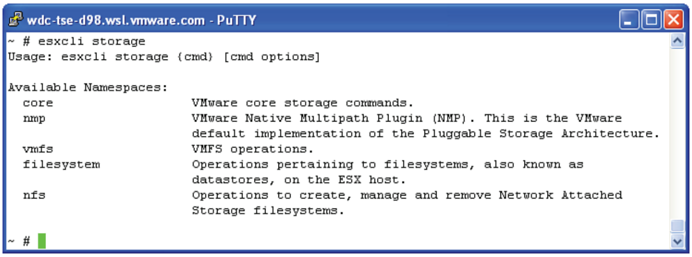


```
~ # esxcli
Usage: esxcli [options] {namespace}+ {cmd} [cmd options]

Options:
  --formatter=FORMATTER
                        Override the formatter to use for a given command. Available
                        formatter: xml, csv, keyvalue
  --format-param=FORMATPARAMS
                        Set a formatter parameter to give the formatter information
                        about how to format the command
  --debug               Enable debug or unsupported options
  --batch=BATCH         Batch mode (debug only)
  --batch-param=BATCHPARAM
                        Batch mode parameters (debug only)
  --version             Display version information for the script
  -?, --help            Display usage information for the script

Available Namespaces:
  esxcli                Commands that operate on the esxcli system itself allowing
                        users to get additional information.
  fcoe                  VMware FCOE commands.
  hardware              VMKernel hardware properties and commands for configuring
                        hardware.
  iscsi                 VMware iSCSI commands.
  license               Operations pertaining to the licensing of vmware and third
                        party modules on the ESX host. These operations currently only
                        include updating third party module licenses.
  network               Operations that pertain to the maintenance of networking on an
                        ESX host. This includes a wide variety of commands to
                        manipulate virtual networking components (vswitch, portgroup,
                        etc) as well as local host IP, DNS and general hsot networking
                        settings.
  software              Manage the ESXi software image and packages
  storage               VMware storage commands.
  system                VMKernel system properties and commands for configuring
                        properties of the kernel core system.
  vm                    A small number of operations that allow a user to Control
                        Virtual Machine operations.

~ #
```

Figure 5.39 Listing esxcli namespace

The relevant namespace for this chapter is `storage`. This is what most of the examples use. Figure 5.40 shows the command-line help for the `storage` namespace:

```
esxcli storage
```



```
~ # esxcli storage
Usage: esxcli storage {cmd} [cmd options]

Available Namespaces:
  core                    VMware core storage commands.
  nmp                     VMware Native Multipath Plugin (NMP). This is the VMware
                          default implementation of the Pluggable Storage Architecture.
  vmfs                    VMFS operations.
  filesystem              Operations pertaining to filesystems, also known as
                          datastores, on the ESX host.
  nfs                     Operations to create, manage and remove Network Attached
                          Storage filesystems.

~ #
```

Figure 5.40 Listing esxcli `storage` namespace

Table 5.11 lists ESXCLI namespaces and their usage.

Table 5.11 Available Namespaces in the `storage` Namespace

Name Space	Usage
core	Use this for anything on the PSA level like other MPPs, PSA claim rules, and so on.
nmp	Use this for NMP and its "children," such as SATP and PSP.
vmfs	Use this for handling VMFS volumes on snapshot LUNs, managing extents, and upgrading VMFS manually.
filesystem	Use this for listing, mounting, and unmounting supported datastores.
nfs	Use this to mount, unmount, and list NFS datastores.

Adding a PSA Claim Rule

PSA claim rules can be for MP, Filter, and VAAI classes. I cover the latter two in Chapter 6.

Following are a few examples of claim rules for the MP class.

Adding a Rule to Change Certain LUNs to Be Claimed by a Different MPP

In general, most arrays function properly using the default PSA claim rules. In certain configurations, you might need to specify a different PSA MPP.

A good example is the following scenario:

You installed PowerPath/VE on your ESXi 5.0 host but then later realized that you have some MSCS cluster nodes running on that host and these nodes use Passthrough RDMs (Physical compatibility mode RDM). Because VMware does not support third-party MPPs with MSCS, you must exclude the LUNs from being managed by PowerPath/VE.

You need to identify the device ID (NAA ID) of each of the RDM LUNs and then identify the paths to each LUN. You use these paths to create the claim rule.

Here is the full procedure:

1. Power off one of the MSCS cluster nodes and locate its home directory. If you cannot power off the VM, skip to Step 6.

   Assuming that the cluster node is located on Clusters_Datastore in a directory named node1, the command and its output would look like Listing 5.1.

Listing 5.1 Locating the RDM Filename

```
#cd /vmfs/volumes/Clusters_datastore/node1

#fgrep scsi1 *.vmx |grep fileName

scsi1:0.fileName = “/vmfs/volumes/4d8008a2-9940968c-04df-001e4f1fbf2a/
node1/quorum.vmdk”

scsi1:1.fileName = “/vmfs/volumes/4d8008a2-9940968c-04df-001e4f1fbf2a/
node1/data.vmdk”
```

   The last two lines are the output of the command. They show the RDM filenames for the node’s shared storage, which are attached to the virtual SCSI adapter named `scsi1`.

2. Using the RDM filenames, including the path to the datastore, you can identify the logical device name to which each RDM maps as shown in Listing 5.2.

Listing 5.2 Identifying RDM’s Logical Device Name Using the RDM Filename

```
#vmkfstools --queryrdm /vmfs/volumes/4d8008a2-9940968c-04df-001e4f1fbf2a/
node1/quorum.vmdk

Disk /vmfs/volumes/4d8008a2-9940968c-04df-001e4f1fbf2a/node1/quorum.vmdk is
a Passthrough Raw Device Mapping
Maps to: vml.02000100006006016055711d00cff95e65664ee011524149442035
```

You may also use the shorthand version using `-q` instead of `--queryrdm`.

This example is for the `quorum.vmdk`. Repeat the same process for the remaining RDMs. The device name is prefixed with vml and is highlighted.

3. Identify the NAA ID using the vml ID as shown in Listing 5.3.

Listing 5.3 Identifying NAA ID Using the Device vml ID

```
#esxcfg-scsidevs --list --device vml.02000100006006016055711d00cff95e65664
ee011524149442035 |grep Display


Display Name: DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011)
```

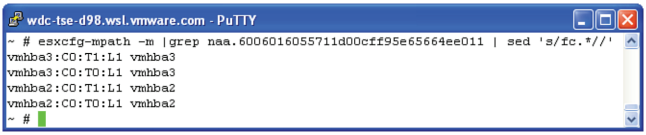
You may also use the shorthand version:

```
#esxcfg-scsidevs -l -d vml.02000100006006016055711d00cff95e65664
ee011524149442035 |grep Display
```

4. Now, use the NAA ID (highlighted in Listing 5.3) to identify the paths to the RDM LUN.

   Figure 5.41 shows the output of command:

   ```
   esxcfg-mpath -m |grep naa.6006016055711d00cff95e65664ee011 | sed 's/
   fc.*//'
   ```



```
~ # esxcfg-mpath -m |grep naa.6006016055711d00cff95e65664ee011 | sed 's/fc.*//'
vmhba3:C0:T1:L1 vmhba3
vmhba3:C0:T0:L1 vmhba3
vmhba2:C0:T1:L1 vmhba2
vmhba2:C0:T0:L1 vmhba2
~ #
```

Figure 5.41 Listing runtime pathnames to an RDM LUN

You may also use the verbose version of the command:

```
esxcfg-mpath --list-map |grep naa.6006016055711d00cff95e65664ee011 |
sed 's/fc.*//'
```

This truncates the output beginning with "`fc`" to the end of the line on each line. If the protocol in use is not FC, replace that with "`iqn`" for iSCSI or "`fcoe`" for FCoE.

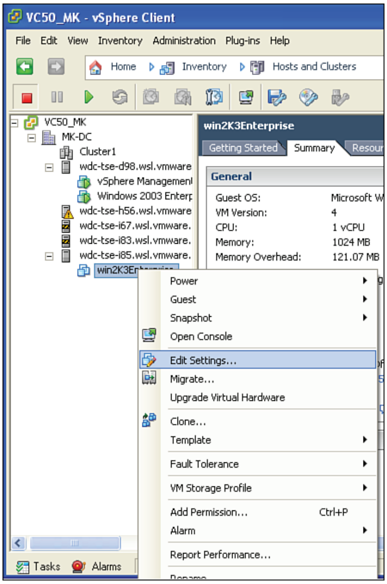
The output shows that the LUN with the identified NAA ID is LUN 1 and has four paths shown in Listing 5.4.

Listing 5.4 RDM LUN's Paths

```
vmhba3:C0:T1:L1
vmhba3:C0:T0:L1
vmhba2:C0:T1:L1
vmhba2:C0:T0:L1
```

If you cannot power off the VMs to run Steps 1–5, you may use the UI instead.

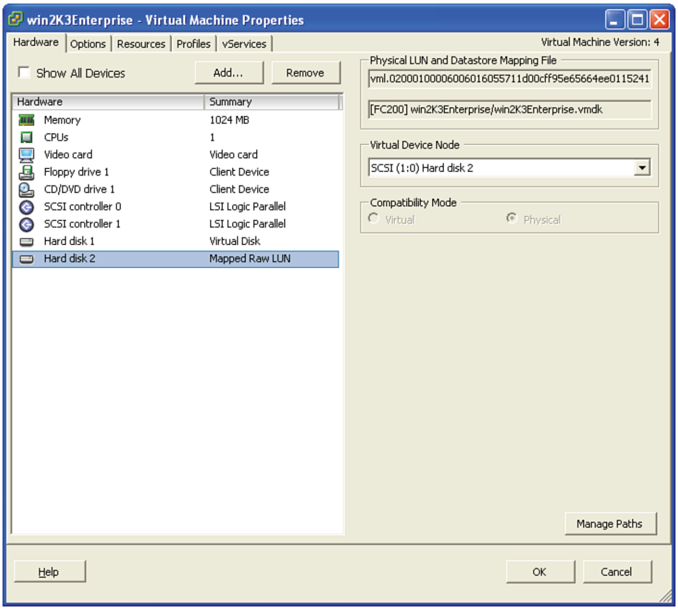
5. Use the vSphere client to navigate to the MSCS node VM. Right-click the VM in the inventory pane and then select **Edit Settings** (see Figure 5.42).



Guest OS: Microsoft W
VM Version: 4
CPU: 1 vCPU
Memory: 1024 MB
Memory Overhead: 121.07 MB

Figure 5.42 Editing VM's settings via the UI

6. In the resulting dialog (see Figure 5.43), locate the RDM listed in the Hardware tab. You can identify this by the summary column showing Mapped Raw LUN. On the top right-hand side you can locate the Logical Device Name, which is prefixed with `vml` in the field labeled Physical LUN and Datastore Mapping File.



Virtual Machine Version: 4

☐ Show All Devices

Hardware	Summary
Memory	1024 MB
CPUs	1
Video card	Video card
Floppy drive 1	Client Device
CD/DVD drive 1	Client Device
SCSI controller 0	LSI Logic Parallel
SCSI controller 1	LSI Logic Parallel
Hard disk 1	Virtual Disk
Hard disk 2	Mapped Raw LUN

Physical LUN and Datastore Mapping File

vml.02000100006006016055711d00cff95e65664ee0115241

[FC200] win2K3Enterprise/win2K3Enterprise.vmdk

Virtual Device Node

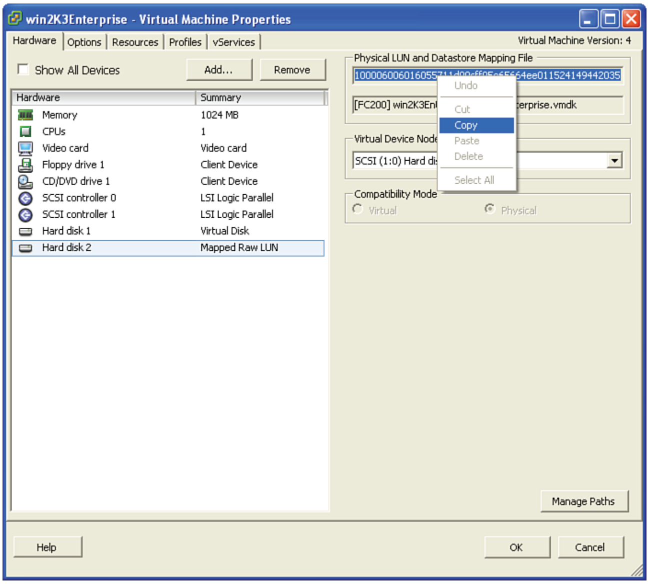
SCSI (1:0) Hard disk 2

Compatibility Mode

○ Virtual ● Physical

Figure 5.43 Virtual machine properties dialog

7. Double-click the text in that field. Right-click the selected text and click **Copy** as shown in Figure 5.44.



Virtual Machine Version: 4

☐ Show All Devices

Hardware	Summary
Memory	1024 MB
CPUs	1
Video card	Video card
Floppy drive 1	Client Device
CD/DVD drive 1	Client Device
SCSI controller 0	LSI Logic Parallel
SCSI controller 1	LSI Logic Parallel
Hard disk 1	Virtual Disk
Hard disk 2	Mapped Raw LUN

Physical LUN and Datastore Mapping File

100006006016055711d00cff95e65664ee011524149442035

[FC200] win2K3En terprise.vmdk

Virtual Device Node

SCSI (1:0) Hard di

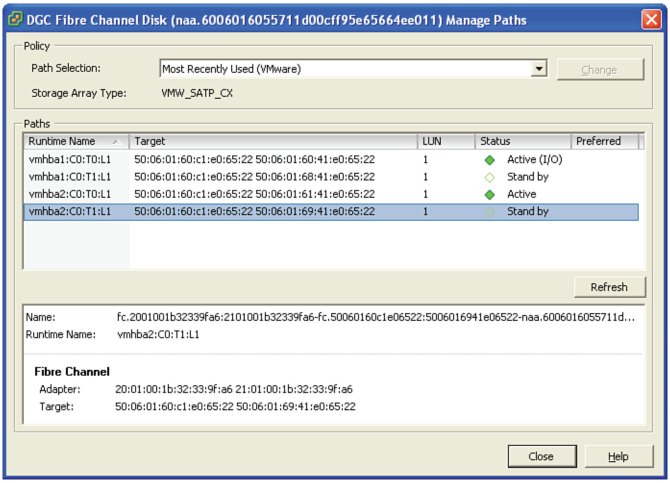
Compatibility Mode

○ Virtual ● Physical

Figure 5.44 Copying RDM’s VML ID (Logical Device Name) via the UI

8. You may use the copied text to follow Steps 4 and 5. Otherwise, you may instead get the list of paths to the LUN using the **Manage Paths** button in the dialog shown in Figure 5.44.

9. In the Manage Paths dialog (see Figure 5.45), click the Runtime Name column to sort it. Write down the list of paths shown there.



DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011) Manage Paths

Policy

Path Selection: Most Recently Used (VMware)

Storage Array Type: VMW_SATP_CX

Paths

Runtime Name	Target	LUN	Status	Preferred
vmhba1:C0:T0:L1	50:06:01:60:c1:e0:65:22 50:06:01:60:41:e0:65:22	1	◆ Active (I/O)	
vmhba1:C0:T1:L1	50:06:01:60:c1:e0:65:22 50:06:01:68:41:e0:65:22	1	◇ Stand by	
vmhba2:C0:T0:L1	50:06:01:60:c1:e0:65:22 50:06:01:61:41:e0:65:22	1	◆ Active	
vmhba2:C0:T1:L1	50:06:01:60:c1:e0:65:22 50:06:01:69:41:e0:65:22	1	◇ Stand by	

Name: fc.2001001b32339fa6:2101001b32339fa6-fc.50060160c1e06522:5006016941e06522-naa.6006016055711d...

Runtime Name: vmhba2:C0:T1:L1

Fibre Channel

Adapter: 20:01:00:1b:32:33:9f:a6 21:01:00:1b:32:33:9f:a6

Target: 50:06:01:60:c1:e0:65:22 50:06:01:69:41:e0:65:22

Figure 5.45 Listing the runtime pathnames via the UI

10. The list of paths shown in Figure 5.45 are

```
vmhba1:C0:T0:L1
vmhba1:C0:T1:L1
vmhba2:C0:T0:L1
vmhba2:C0:T1:L1
```

NOTE

Notice that the list of paths in the UI is different from that obtained from the command line. The reason can be easily explained; I used two different hosts for obtaining the list of paths. If your servers were configured identically, the path list should be identical as well.

However, this is not critical because the LUN's NAA ID is the same regardless of paths used to access it. This is what makes NAA ID the most unique element of any LUN, and that is the reason ESXi utilizes it for uniquely identifying the LUNs. I cover more on that topic later in Chapter 7.

11. Create the claim rule.

I use the list of paths obtained in Step 5 for creating the rule from the ESXi host from which it was obtained.

The Ground Rules for Creating the Rule

- The rule number must be lower than any of the rules created by PowerPath/VE installation. By default, they are assigned rules 250–320 (refer to Figure 5.26 for the list of PowerPath claim rules).
- The rule number must be higher than 101 because this is used by the Dell Mask Path rule. This prevents claiming devices masked by that rule.
- If you created other claim rules in the past on this host, use a rule number that is different from what you created in a fashion that the new rules you are creating now do not conflict with the earlier rules.
- If you must place the new rules in an order earlier than an existing rule but there are no rule numbers available, you may have to move one of the lower-numbered rules higher by the number of rules you plan on creating.

  For example, you have previously created rules numbered 102–110 and that rule 109 cannot be listed prior to the new rules you are creating. If the new rules count is four, you need to assign them rule numbers 109–112. To do that, you need to move rules 109 and 110 to numbers 113 and 114. To avoid having to do this in the future, consider leaving gaps in the rule numbers among sections.

  An example of moving a rule is

  ```
  esxcli storage core claimrule move --rule 109 --new-rule 113
  esxcli storage core claimrule move --rule 110 --new-rule 114
  ```

  You may also use the shorthand version:

  ```
  esxcli storage core claimrule move -r 109 -n 113
  esxcli storage core claimrule move -r 110 -n 114
  ```

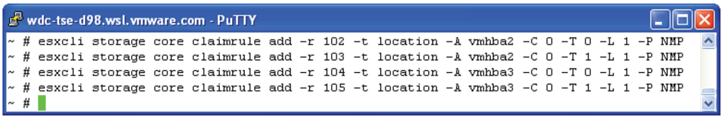
Now, let’s proceed with adding the new claim rules:

1. The set of four commands shown in Figure 5.46 create rules numbered 102–105. The rules criteria are
   - The claim rule type is “location” (`-t location`).
   - The location is specified using each path to the same LUN in the format:
     - `–A` or `--adapter vmhba(x)` where `X` is the vmhba number associated with the path.

    - `-C` or `--channel (Y)` where `Y` is the channel number associated with the path.
    - `–T` or `--target (Z)` where `Z` is the target number associated with the path.
    - `–L` or `--lun (n)` where `n` is the LUN number.
- The plug-in name is NMP, which means that this claim rule is for NMP to claim the paths listed in each rule created.

NOTE

It would have been easier to create a single rule using the LUN's NAA ID by using the `--type device` option and then using `--device <NAA ID>`. However, the use of device as a rule type is not supported with MP class plug-ins.

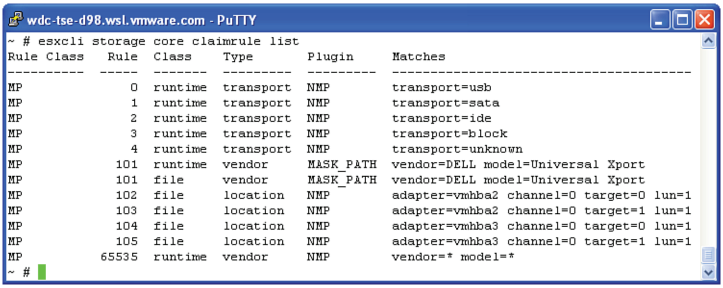


```
~ # esxcli storage core claimrule add -r 102 -t location -A vmhba2 -C 0 -T 0 -L 1 -P NMP
~ # esxcli storage core claimrule add -r 103 -t location -A vmhba2 -C 0 -T 1 -L 1 -P NMP
~ # esxcli storage core claimrule add -r 104 -t location -A vmhba3 -C 0 -T 0 -L 1 -P NMP
~ # esxcli storage core claimrule add -r 105 -t location -A vmhba3 -C 0 -T 1 -L 1 -P NMP
~ #
```

Figure 5.46 Adding new MP claim rules

2. Repeat Step 1 for each LUN you want to reconfigure.
3. Verify that the rules were added successfully. To list the current set of claim rules, run the command shown in Figure 5.47:

   `esxcli storage core claimrule list.`



```
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            102  file     location   NMP        adapter=vmhba2 channel=0 target=0 lun=1
MP            103  file     location   NMP        adapter=vmhba2 channel=0 target=1 lun=1
MP            104  file     location   NMP        adapter=vmhba3 channel=0 target=0 lun=1
MP            105  file     location   NMP        adapter=vmhba3 channel=0 target=1 lun=1
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

Figure 5.47 Listing added claim rules

Notice that the four new rules are now listed, but the `Class` column shows them as file. This means that the configuration files were updated successfully but the rules were not loaded into memory yet.

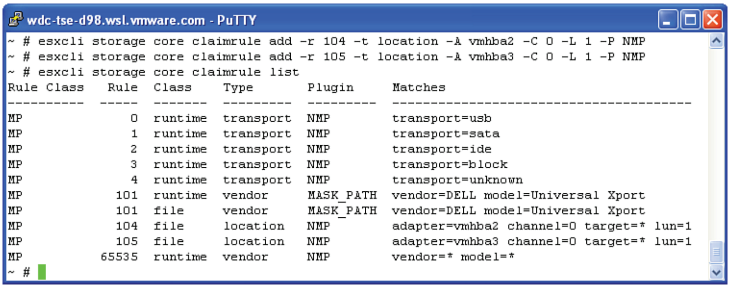
NOTE

I truncated the PowerPath rules in Figure 5.47 for readability. Also note that using the Location type utilizes the current runtime names of the devices, and they may change in the future. If your configuration changes—for example, adding new HBAs or removing existing ones—the runtime names change, too. This results in these claim rules claiming the wrong devices. However, in a static environment, this should not be an issue.

TIP

To reduce the number of commands used and the number of rules created, you may omit the `-T` or `--target` option, which assumes a wildcard. You may also use the `-u` or `--autoassign` option to auto-assign the rule number. However, the latter assigns rule numbers starting with 5001, which may be higher than the existing claim rules for the device hosting the LUN you are planning to claim.

Figure 5.48 shows a sample command line that implements a wildcard for the target. Notice that this results in creating two rules instead of four and the "target" match is `*`.



```
~ # esxcli storage core claimrule add -r 104 -t location -A vmhba2 -C 0 -L 1 -P NMP
~ # esxcli storage core claimrule add -r 105 -t location -A vmhba3 -C 0 -L 1 -P NMP
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            104  file     location   NMP        adapter=vmhba2 channel=0 target=* lun=1
MP            105  file     location   NMP        adapter=vmhba3 channel=0 target=* lun=1
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

Figure 5.48 Adding MP claim rules using a wildcard

4. Before loading the new rules, you must first unclaim the paths to the LUN specified in that rule set. You use the NAA ID as the device ID:

```
esxcli storage core claiming unclaim --type device --device naa.600601
6055711d00cff95e65664ee011
```

You may also use the shorthand version:

```
esxcli storage core claiming unclaim -t device -d naa.6006016055711d00
cff95e65664ee011
```

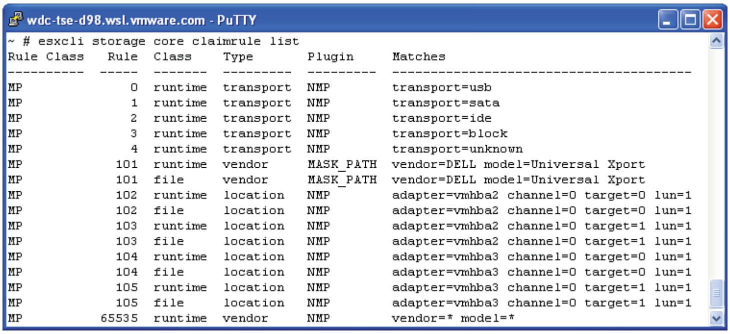
5. Load the new claim rules so that the paths to the LUN get claimed by NMP:

   ```
   esxcli storage core claimrule load
   ```

6. Use the following command to list the claim rules to verify that they were successfully loaded:

   ```
   esxcli storage core claimrule list
   ```

   Now you see that each of the new rules is listed twice—once with file class and once with runtime class—as shown in Figure 5.49.



```
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            102  runtime  location   NMP        adapter=vmhba2 channel=0 target=0 lun=1
MP            102  file     location   NMP        adapter=vmhba2 channel=0 target=0 lun=1
MP            103  runtime  location   NMP        adapter=vmhba2 channel=0 target=1 lun=1
MP            103  file     location   NMP        adapter=vmhba2 channel=0 target=1 lun=1
MP            104  runtime  location   NMP        adapter=vmhba3 channel=0 target=0 lun=1
MP            104  file     location   NMP        adapter=vmhba3 channel=0 target=0 lun=1
MP            105  runtime  location   NMP        adapter=vmhba3 channel=0 target=1 lun=1
MP            105  file     location   NMP        adapter=vmhba3 channel=0 target=1 lun=1
MP          65535  runtime  vendor     NMP        vendor=* model=*
```

Figure 5.49 Listing MP claim rules

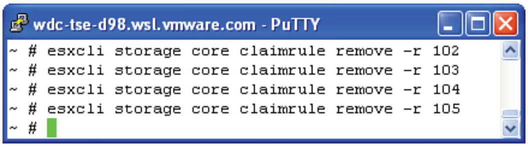
How to Delete a Claim Rule

Deleting a claim rule must be done with extreme caution. Make sure that you are deleting the rule you intend to delete. Prior to doing so, make sure to collect a "vm-support" dump by running `vm-support` from a command line at the host or via SSH. Alternatively, you can select the menu option Collect Diagnostics Data via the vSphere client.

To delete a claim rule, follow this procedure via the CLI (locally, via SSH, vCLI, or vMA):

1. List the current claim rules set and identify the claim rule or rules you want to delete. The command to list the claim rules is similar to what you ran in Step 6 and is shown in Figure 5.49.

2. For this procedure, I am going to use the previous example and delete the four claim rules I added earlier which are rules 102–105. The command for doing that is in Figure 5.50.



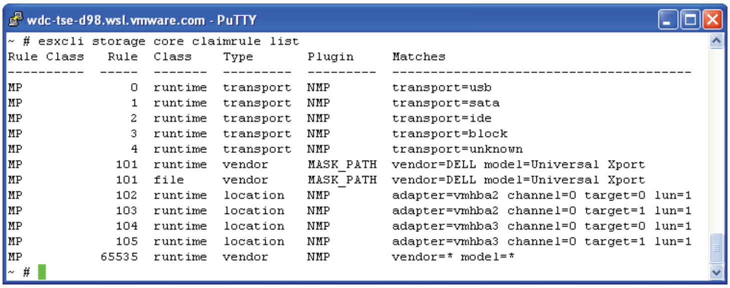
```
~ # esxcli storage core claimrule remove -r 102
~ # esxcli storage core claimrule remove -r 103
~ # esxcli storage core claimrule remove -r 104
~ # esxcli storage core claimrule remove -r 105
~ #
```

Figure 5.50 Removing claim rules via the CLI

You may also run the verbose command:

```
esxcli storage core claimrule remove --rule <rule-number>
```

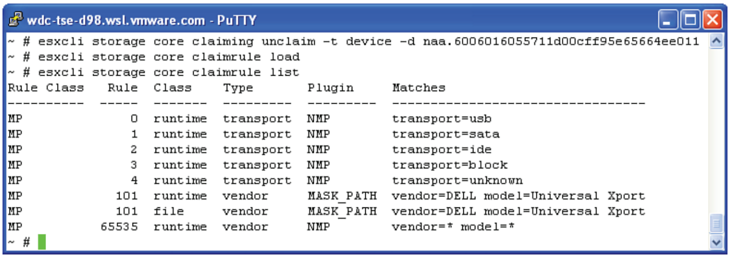
3. Running the `claimrule` list command now results in an output similar to Figure 5.51. Observe that even though I just deleted the claim rules, they still show up on the list. The reason for that is the fact that I have not loaded the modified claim rules. That is why the deleted rules show runtime in their `Class` column.



```
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            102  runtime  location   NMP        adapter=vmhba2 channel=0 target=0 lun=1
MP            103  runtime  location   NMP        adapter=vmhba2 channel=0 target=1 lun=1
MP            104  runtime  location   NMP        adapter=vmhba3 channel=0 target=0 lun=1
MP            105  runtime  location   NMP        adapter=vmhba3 channel=0 target=1 lun=1
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

Figure 5.51 Listing MP claim rules

5. Because I know from the previous procedure the device ID (NAA ID) of the LUN whose claim rules I deleted, I ran the `unclaim` command using the `-t` device or `--type` option and then specified the `-d` or `--device` option with the NAA ID. I then loaded the claim rules using the load option. Notice that the deleted claim rules are no longer listed see Figure 5.52.



```
~ # esxcli storage core claiming unclaim -t device -d naa.6006016055711d00cff95e65664ee011
~ # esxcli storage core claimrule load
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

Figure 5.52 Unclaiming a device using its NAA ID and then loading the claim rules

You may also use the verbose command options:

```
esxcli storage core claiming unclaim --type device --device <Device-ID>
```

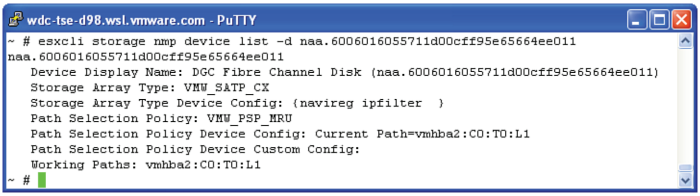
You may need to claim the device after loading the claim rule by repeating the claiming command using the “claim” instead of the “unclaim” option:

```
esxcli storage core claiming claim -t device -d <device-ID>
```

How to Mask Paths to a Certain LUN

Masking a LUN is a similar process to that of adding claim rules to claim certain paths to a LUN. The main difference is that the plug-in name is MASK_PATH instead of NMP as used in the previous example. The end result is that the masked LUNs are no longer visible to the host.

1. Assume that you want to mask LUN 1 used in the previous example and it still has the same NAA ID. I first run a command to list the LUN visible by the ESXi host as an example to show the before state (see Figure 5.53).

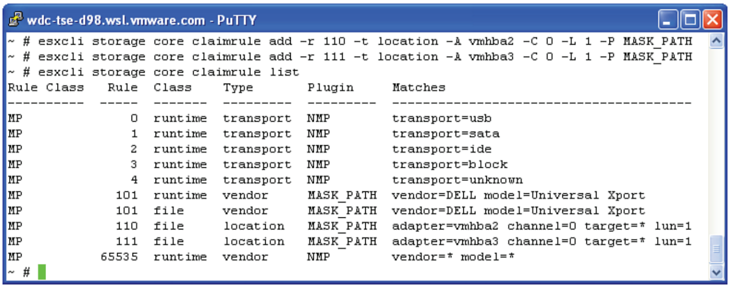


```
~ # esxcli storage nmp device list -d naa.6006016055711d00cff95e65664ee011
naa.6006016055711d00cff95e65664ee011
   Device Display Name: DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011)
   Storage Array Type: VMW_SATP_CX
   Storage Array Type Device Config: {navireg ipfilter  }
   Path Selection Policy: VMW_PSP_MRU
   Path Selection Policy Device Config: Current Path=vmhba2:C0:T0:L1
   Path Selection Policy Device Custom Config:
   Working Paths: vmhba2:C0:T0:L1
~ #
```

Figure 5.53 Listing LUN properties using its NAA ID via the CLI

You may also use the verbose command option `--device` instead of `-d`.

2. Add the MASK_LUN claim rule, as shown in Figure 5.54.

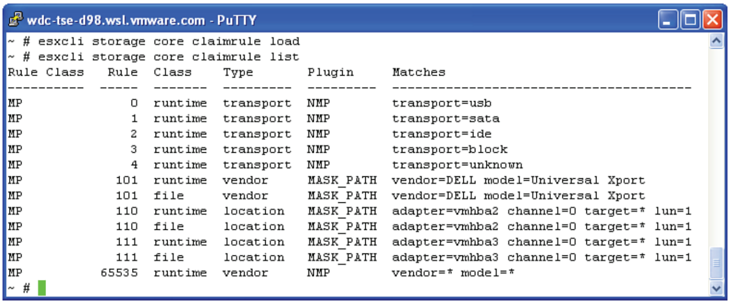


```
~ # esxcli storage core claimrule add -r 110 -t location -A vmhba2 -C 0 -L 1 -P MASK_PATH
~ # esxcli storage core claimrule add -r 111 -t location -A vmhba3 -C 0 -L 1 -P MASK_PATH
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            110  file     location   MASK_PATH  adapter=vmhba2 channel=0 target=* lun=1
MP            111  file     location   MASK_PATH  adapter=vmhba3 channel=0 target=* lun=1
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

Figure 5.54 Adding Mask Path claim rules

As you see in Figure 5.54, I added rule numbers 110 and 111 to have MASK_PATH plug-in claim all targets to LUN1 via vmhba2 and vmhba3. The claim rules are not yet loaded, hence the file class listing and no runtime class listings.

3. Load and then list the claim rules (see Figure 5.55).

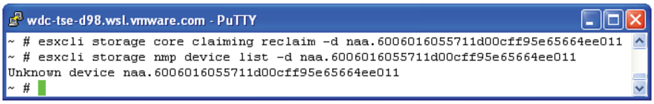


```
~ # esxcli storage core claimrule load
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            110  runtime  location   MASK_PATH  adapter=vmhba2 channel=0 target=* lun=1
MP            110  file     location   MASK_PATH  adapter=vmhba2 channel=0 target=* lun=1
MP            111  runtime  location   MASK_PATH  adapter=vmhba3 channel=0 target=* lun=1
MP            111  file     location   MASK_PATH  adapter=vmhba3 channel=0 target=* lun=1
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

Figure 5.55 Loading and listing claim rules after adding Mask Path rules

Now you see the claim rules listed with both file and runtime classes.

4. Use the reclaim option to unclaim and then claim the LUN using its NAA ID. Check if it is still visible (see Figure 5.56).



```
~ # esxcli storage core claiming reclaim -d naa.6006016055711d00cff95e65664ee011
~ # esxcli storage nmp device list -d naa.6006016055711d00cff95e65664ee011
Unknown device naa.6006016055711d00cff95e65664ee011
~ #
```

Figure 5.56 Reclaiming the paths after loading the Mask Path rules

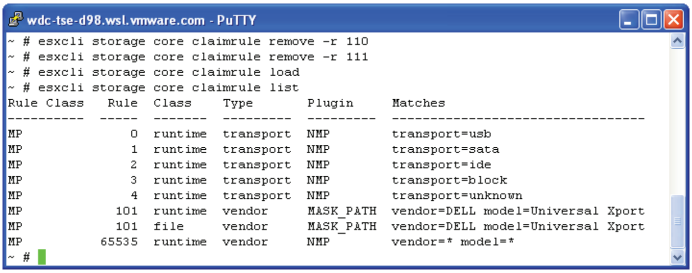
You may also use the verbose command option `--device` instead of `-d`.

Notice that after reclaiming the LUN, it is now an Unknown device.

How to Unmask a LUN

To unmask this LUN, reverse the preceding steps and then reclaim the LUN as follows:

1. Remove the MASK_PATH claim rules (numbers 110 and 111) as shown in Figure 5.57.



```
~ # esxcli storage core claimrule remove -r 110
~ # esxcli storage core claimrule remove -r 111
~ # esxcli storage core claimrule load
~ # esxcli storage core claimrule list
Rule Class   Rule  Class    Type       Plugin     Matches
----------  -----  -------  ---------  ---------  ---------------------------------
MP              0  runtime  transport  NMP        transport=usb
MP              1  runtime  transport  NMP        transport=sata
MP              2  runtime  transport  NMP        transport=ide
MP              3  runtime  transport  NMP        transport=block
MP              4  runtime  transport  NMP        transport=unknown
MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
MP          65535  runtime  vendor     NMP        vendor=* model=*
~ #
```

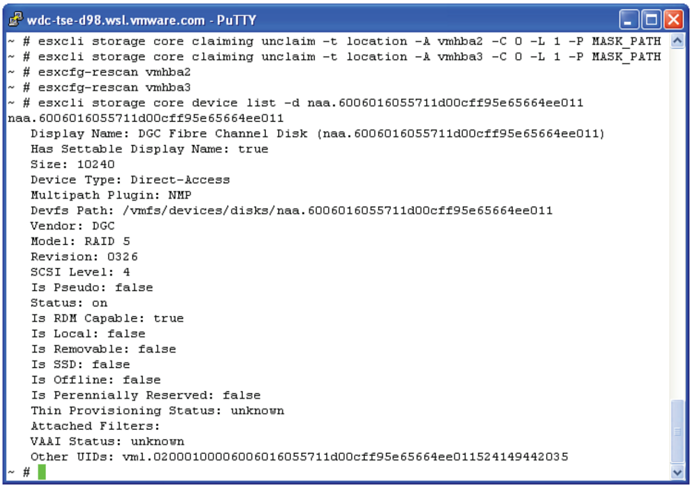
Figure 5.57 Removing the Mask Path claim rules

You may also use the verbose command options:

```
esxcli storage core claimrule remove --rule <rule-number>
```

2. Unclaim the paths to the LUN in the same fashion you used while adding the MASK_PATH claim rules—that is, using the `-t` location and omitting the `-T` option so that the target is a wildcard.
3. Rescan using both HBA names.
4. Verify that the LUN is now visible by running the list command.

   Figure 5.58 shows the outputs of Steps 2–4.



```
~ # esxcli storage core claiming unclaim -t location -A vmhba2 -C 0 -L 1 -P MASK_PATH
~ # esxcli storage core claiming unclaim -t location -A vmhba3 -C 0 -L 1 -P MASK_PATH
~ # esxcfg-rescan vmhba2
~ # esxcfg-rescan vmhba3
~ # esxcli storage core device list -d naa.6006016055711d00cff95e65664ee011
naa.6006016055711d00cff95e65664ee011
   Display Name: DGC Fibre Channel Disk (naa.6006016055711d00cff95e65664ee011)
   Has Settable Display Name: true
   Size: 10240
   Device Type: Direct-Access
   Multipath Plugin: NMP
   Devfs Path: /vmfs/devices/disks/naa.6006016055711d00cff95e65664ee011
   Vendor: DGC
   Model: RAID 5
   Revision: 0326
   SCSI Level: 4
   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.02000100006006016055711d00cff95e65664ee011524149442035
~ #
```

Figure 5.58 Unclaiming the Masked Paths

You may also use the verbose command options:

```
esxcli storage core claiming unclaim --type location --adapter vmhba2
--channel 0 --lun 1 --plugin MASK_PATH
```

Changing PSP Assignment via the CLI

The CLI enables you to modify the PSP assignment per device. It also enables you to change the default PSP for a specific storage array or family of arrays. I cover the former use case first because it is similar to what you did via the UI in the previous section. I follow with the latter use case.

Changing PSP Assignment for a Device

To change the PSP assignment for a given device, you may follow this procedure:

1. Log on to the ESXi 5 host locally or via SSH as root or using vMA 5.0 as vi-admin.
2. Identify the device ID for each LUN you want to reconfigure:

   ```
   esxcfg-mpath -b |grep -B1 “fc Adapter”| grep -v -e “--” |sed ‘s/
   Adapter.*//’
   ```

   You may also use the verbose version of this command:

   ```
   esxcfg-mpath --list-paths grep -B1 “fc Adapter”| grep -v -e “--“ | sed
   ‘s/Adapter.*//’
   ```

   Listing 5.5 shows the output of this command.

Listing 5.5 Listing Device ID and Its Paths

```
naa.60060e80052751000000275100000011a : HITACHI Fibre Channel Disk (naa.6006
0e80052751000000275100000011a)
     vmhba2:C0:T0:L1 LUN:1 state:active fc
     vmhba2:C0:T1:L1 LUN:1 state:active fc
     vmhba3:C0:T0:L1 LUN:1 state:active fc
     vmhba3:C0:T1:L1 LUN:1 state:active fc
```

From there, you can identify the device ID (in this case, it is the NAA ID). Note that this output was collected using a Universal Storage Platform®V (USP V), USP VM, or Virtual Storage Platform (VSP).

This output means that LUN1 has device ID `naa.60060e800527510000002751 0000011a`.

3. Using the device ID you identified, run this command:

```
esxcli storage nmp device set -d <device-id> --psp=<psp-name>
```

You may also use the verbose version of this command:

```
esxcli storage nmp device set --device <device-id> --psp=<psp-name>
```

For example:

```
esxcli storage nmp device set -d naa.60060e80052751000000275100000011a
--psp=VMW_PSP_FIXED
```

This command sets the device with ID `naa.60060e8005275100000027510000011a` to be claimed by the PSP named VMW_PSP_FIXED.

Changing the Default PSP for a Storage Array

There is no simple way to change the default PSP for a specific storage array unless that array is claimed by an SATP that is specific for it. In other words, if it is claimed by an SATP that also claims other brands of storage arrays, changing the default PSP affects *all* storage arrays claimed by the SATP. However, you may add an SATP claim rule that uses a specific PSP based on your storage array's Vendor and Model strings:

1. Identify the array's Vendor and Model strings. You can identify these strings by running

```
esxcli storage core device list -d <device ID> |grep 'Vendor\|Model'
```

Listing 5.6 shows an example for a device on an HP P6400 Storage Array.

Listing 5.6 Listing Device's Vendor and Model Strings

```
esxcli storage core device list -d naa.600508b4000f02cb0001000001660000
|grep 'Model\|Vendor'
   Vendor: HP
   Model: HSV340
```

In this example, the Vendor String is HP and the Model is HSV340.

2. Use the identified values in the following command:

```
esxcli storage nmp satp rule add --satp <current-SATP-USED> --vendor
<Vendor string> --model <Model string> --psp <PSP-name> --description
<Description>
```

TIP

It is always a good practice to document changes manually made to the ESXi host configuration. That is why I used the `--description` option to add a description of the rules I add. This way other admins would know what I did if they forget to read the change control record that I added using the company's change control software.

In this example, the command would be like this:

```
esxcli storage nmp satp rule add --satp VMW_SATP_EVA --vendor HP
--model HSV340 --psp VMW_PSP_FIXED --description "Manually added to
use FIXED"
```

It runs silently and returns an error if it fails.

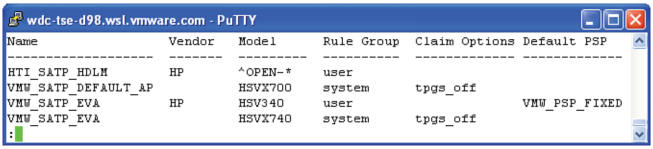
Example of an error:

```
"Error adding SATP user rule: Duplicate user rule found for SATP VMW_
SATP_EVA matching vendor HP model HSV340 claim Options PSP VMW_PSP_
FIXED and PSP Options"
```

This error means that a rule already exists with these options. I simulated this rule by first adding it and then rerunning the same command. To view the existing SATP claim rules list for all HP storage arrays, you may run the following command:

```
esxcli storage nmp satp rule list |less -S |grep 'Name\|---\|HP'|less
-S
```

Figure 5.59 shows the output of this command (I cropped some blank columns, including Device, for readability):



```
Name                 Vendor  Model      Rule Group  Claim Options Default PSP
-------------------- ------- ---------- ----------- ------------- -------------
HTI_SATP_HDLM        HP      ^OPEN-*    user
VMW_SATP_DEFAULT_AP          HSVX700    system      tpgs_off
VMW_SATP_EVA         HP      HSV340     user                      VMW_PSP_FIXED
VMW_SATP_EVA                 HSVX740    system      tpgs_off
:
```

Figure 5.59 Listing SATP rule list for HP devices

You can easily identify non-system rules where the `Rule Group` column value is `user`. Such rules were added by a third-party MPIO installer or manually added by an ESXi 5 administrator. The rule in this example shows that I had already added VMW_PSP_FIXED as the default PSP for VMW_SATP_EVA when the matching vendor is HP and Model is HSV340.

I don't mean to state by this example that HP EVA arrays with HSV340 firmware should be claimed by this specific PSP. I am only using it for demonstration purposes. You *must* verify which PSP is supported by and certified for your specific storage array from the array vendor.

As a matter of fact, this HP EVA model happens to be an ALUA array and the SATP must be VMW_SATP_ALUA see Chapter 6. How did I know that? Let me explain!

- Look at the output in Figures 5.29–5.32. There you should notice that there are no listings of HP EVA arrays with Claim Options value of `tpgs_on`. This means that they were not claimed by any specific SATP explicitly.
- To filter out some clutter from the output, run the following command to list all claim rules with a match on Claim Options value of `tpgs_on`.

```
esxcli storage nmp satp rule list |grep ‘Name\|---\|tpgs_on’ |less -S
```

Listing 5.7 shows the output of that command:

Listing 5.7 Listing SATP Claim Rules List

```
Name                 Device  Vendor   Model     Rule Group Claim Options
-------------------- ------  -------  --------  ----------  ------------
VMW_SATP_ALUA                NETAPP             system      tpgs_on
VMW_SATP_ALUA                IBM      2810XIV   system      tpgs_on
VMW_SATP_ALUA                                   system      tpgs_on
VMW_SATP_ALUA_CX             DGC                system      tpgs_on
```

I cropped some blank columns for readability.

Here you see that there is a claim rule with a blank vendor and the Claim Options is `tpgs_on`. This claim rule claims *any* device with *any* vendor string as long as its Claim Options is `tpgs_on`.

Based on this rule, VMW_SATP_ALUA claims *all* ALUA-capable arrays including HP storage arrays based on a match on the Claim Options value of `tpgs_on`.

What does this mean anyway?

It means that the claim rule that I added for the HSV340 is wrong because it will force it to be claimed by an SATP that does not handle ALUA. I must remove the rule that I added then create another rule that does not violate the default SATP assignment:

1. To remove the SATP claim rule, use the same command used to add, substituting the add option with remove:

   ```
   esxcli storage nmp satp rule remove --satp VMW_SATP_EVA --vendor HP
   --model HSV340 --psp VMW_PSP_FIXED
   ```

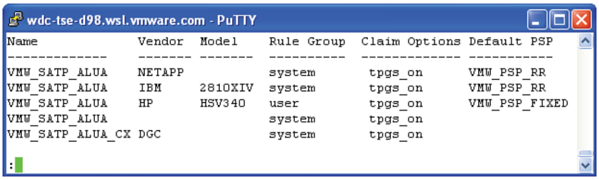
2. Add a new claim rule to have VMW_SATP_ALUA claim the HP EVA HSV340 when it reports Claim Options value as `tpgs_on`:

   ```
   esxcli storage nmp satp rule add --satp VMW_SATP_ALUA --vendor
   HP --model HSV340 --psp VMW_PSP_FIXED --claim-option tpgs_on
   --description "Re-added manually for HP HSV340"
   ```

3. Verify that the rule was created correctly. Run the same command used in Step 2 in the last procedure:

   ```
   esxcli storage nmp satp rule list |grep 'Name\|---\|tpgs_on' |less -S
   ```

   Figure 5.60 shows the output.



```
Name             Vendor  Model    Rule Group  Claim Options Default PSP
-------------    ------- -------  ----------  ------------- -----------
VMW_SATP_ALUA    NETAPP           system       tpgs_on      VMW_PSP_RR
VMW_SATP_ALUA    IBM     2810XIV  system       tpgs_on      VMW_PSP_RR
VMW_SATP_ALUA    HP      HSV340   user         tpgs_on      VMW_PSP_FIXED
VMW_SATP_ALUA                     system       tpgs_on
VMW_SATP_ALUA_CX DGC              system       tpgs_on

:
```

Figure 5.60 SATP rule list after adding rule

Notice that the claim rule has been added in a position prior to the catch-all rule described earlier. This means that this HP EVA HSV340 model will be claimed by VMW_SATP_ALUA when the Claim Options value is `tpgs_on`.

NOTE

If you had manually set certain LUNs to a specific PSP previously, the preceding command will not affect that setting.

To reset such a LUN to use the current default PSP, use the following command:

```
esxcli storage nmp device set --device <device-ID> --default
```

For example:

```
esxcli storage nmp device set --device naa.6006016055711d00cef95e65
664ee011 –-default
```

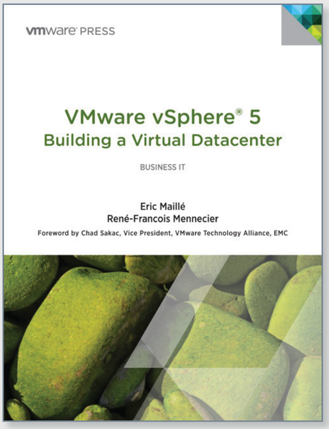
NOTE

All EVAs today have the `tpgs_on` option enabled by default, and it CANNOT be changed by the user. So adding an EVA claim rule would only be useful in the context of trying to use a different PSP by default for all EVA LUNs or assigning PSP defaults to EVA different from other ALUA-capable arrays using the default SATP_ALUA.

Summary

This chapter covered PSA (VMware Pluggable Storage Architecture) components. I showed you how to list PSA plug-ins and how they interact with vSphere ESXi 5. I also showed you how to list, modify, and customize PSA claim rules and how to work around some common issues.

It also covered how ALUA-capable devices interact with SATP claim rules for the purpose of using a specific PSP.



vmware PRESS

VMware vSphere® 5
Building a Virtual Datacenter

BUSINESS IT

Eric Maillé
René-Francois Mennecier
Foreword by Chad Sakac, Vice President, VMware Technology Alliance, EMC

BUY NOW

VMware vSphere® 5 Building a Virtual Datacenter

BY ERIC MAILLE & RENÉ-FRANCOIS MENNECIER

CHAPTER 3 Storage in vSphere 5

Table of Contents

ISBN: 9780321832214

AUGUST 2012

Available in Print and eBook formats and through SAFARI BOOKS ONLINE

SHARE WITH OTHERS

vmware PRESS

vmware.com/go/vmwarepress

ALWAYS LEARNING

PEARSON

Chapter 3

Storage in vSphere 5

Storage is usually the most essential component of virtualized architecture, playing a major role in your system's performance and extensibility. It must be able to support the activity of hosted VMs and be upgradeable to meet future needs. In some projects, the time devoted to designing storage architecture can represent up to 60% of all work. Therefore, the best solution must be chosen according to your business constraints, goals, and allocated budget, because costs can vary significantly among the different storage solutions available.

Storage Representation

Because vSphere 5 offers a wide variety of storage options, it is important to know what features are offered and to understand the interactions between traditional storage in the physical world and the integration of vSphere into such an environment (see Figure 3.1).

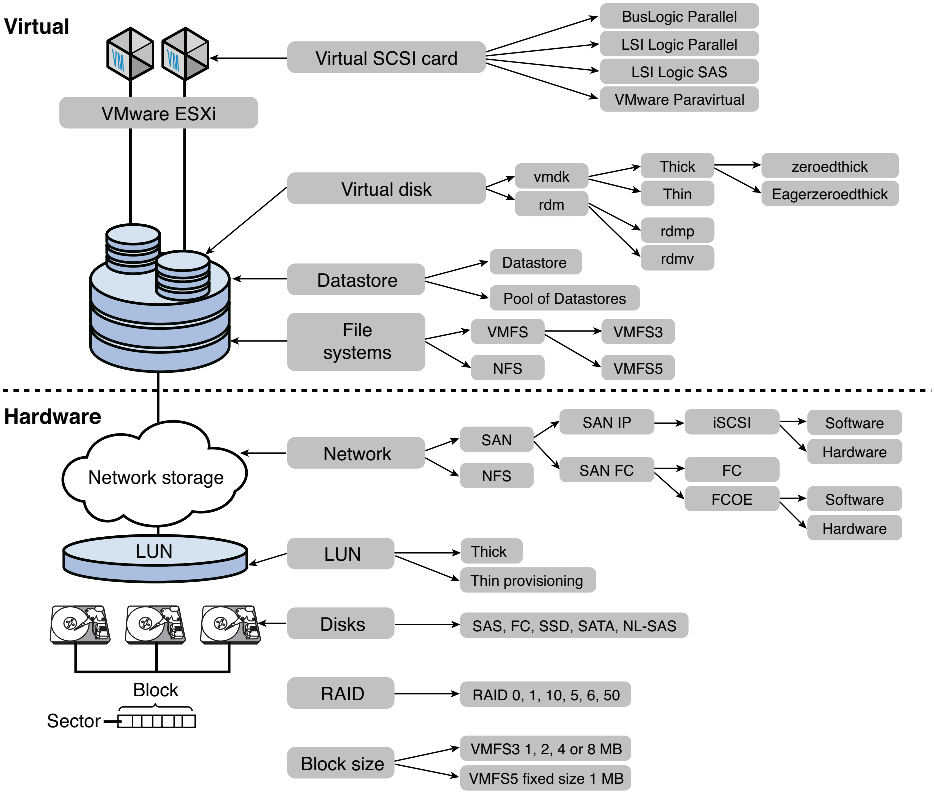


Figure 3.1 How material objects traditionally manipulated by storage administrators (bottom) interact with those manipulated by VMware administrators (top).

Available Storage Architectures

VMware supports several storage protocols, which can make it difficult for companies to know which option best suits their needs. Although this flexibility and freedom can be a good thing, too many options can make decision making difficult or even overwhelming. Although a few years ago the only viable option for production environments was *storage-area network (SAN) Fibre Channel (FC)*, today the differences between protocols are of less importance, and several criteria must be taken into account. Figure 3.2 shows the supported protocols.

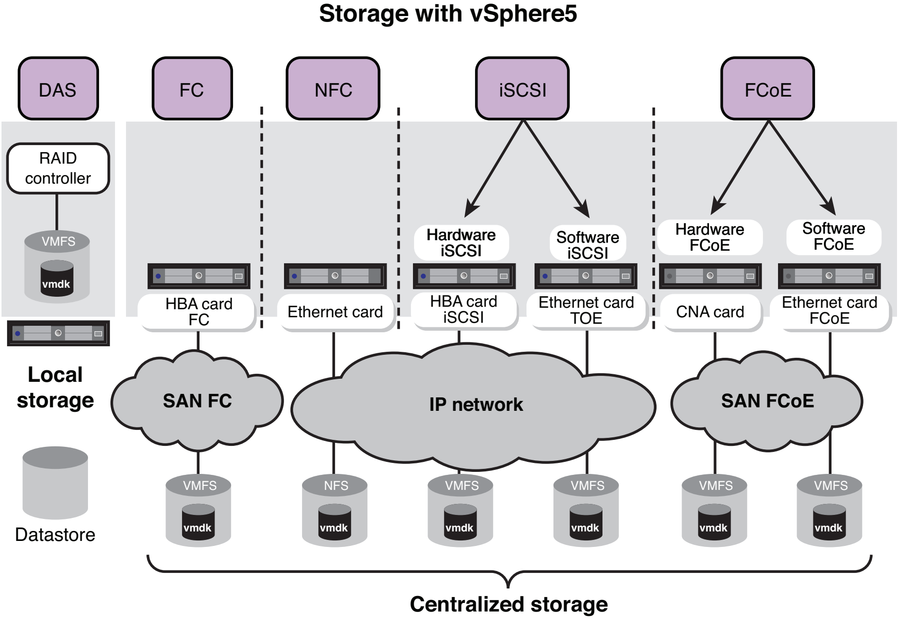
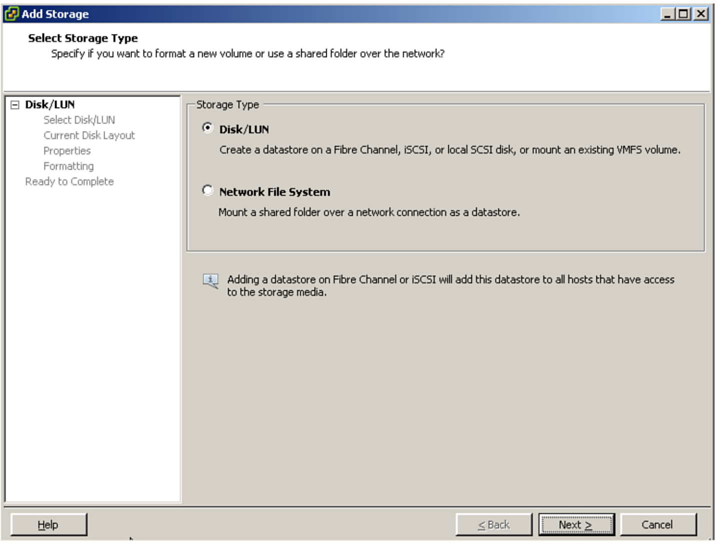


Figure 3.2 Local versus centralized storage architectures.

The following storage options, shown in Figure 3.3, are available in virtual environments:

- **Local storage**: Hard drives are directly connected within the server or as *direct-attached storage (DAS)*, which are disk arrays directly attached to the server.
- **Centralized storage**: Storage is external from the server. The following protocols are supported by ESX:
  - *Fibre Channel (FC)*
  - *Internet Small Computer System Interface (iSCSI)* software or hardware initiator
  - *Network File System (NFS)* used by *network-attached storage (NAS)*
  - *Fibre Channel over Ethernet (FCoE)* software or hardware initiator



Select Storage Type

Specify if you want to format a new volume or use a shared folder over the network?

Storage Type

● Disk/LUN

Create a datastore on a Fibre Channel, iSCSI, or local SCSI disk, or mount an existing VMFS volume.

○ Network File System

Mount a shared folder over a network connection as a datastore.

Adding a datastore on Fibre Channel or iSCSI will add this datastore to all hosts that have access to the storage media.

Figure 3.3 The type of storage must be chosen when creating the datastore.

Local Storage

Local storage is commonly used when installing ESXi Hypervisor. When an ESXi server is isolated and not in a cluster, this storage space can be used for operating system image files (provided as ISO files) or noncritical test and development VMs. Because by definition local storage is (usually) not shared, placement of critical-production VMs should be avoided because the service levels are too low. Features such as vMotion, *Distributed Resource Scheduler (DRS)*, *High Availability (HA)*, *and Fault Tolerance (FT)* are not available except when the vSphere Storage Appliance is used.

Centralized Storage

In a centralized architecture, vSphere can be made to work in clusters, increasing service levels by using advanced features such as vMotion, DRS, HA, FT, and *Site Replication Manager (SRM)*. Moreover, these types of architectures provide excellent performance, and the addition of the *vStorage APIs for Array Integration (VAAI)* relieves the host server from some storage-related tasks by offloading it to a storage array.

NAS storage servers are based on a client/server architecture that accesses data at the NFS level. This protocol, called *file mode*, uses the company's standard Ethernet network. Network cards are available in 1 GbE (1 Gbps) or 10 GbE(10 Gbps).

Other protocols provide a direct I/O access (also called *block mode*) between host servers and storage by using SCSI commands in a dedicated network called a *storage-area network* (*SAN*). With VMware, the advantage of block mode over file mode is that *Raw Device Mapping (RDM)* volumes can be attributed to VMs. VMware uses the *Virtual Machine File System (VMFS)* in this architecture.

NOTE

In VMware, one notable difference exists between NFS and VMFS. With NFS, the NAS server manages the file system and relies on the ESXi network layer (issues are resolved by network teams), whereas VMFS is managed directly by the ESXi storage layer.

There are different types of SANs, both IP-based SANs and FC-based SANs:

- **SAN IP (called iSCSI):** Encapsulates SCSI commands through the TCI/IP network (SCSI over IP). You can access the iSCSI network by using either a software initiator coupled with a standard network card, or a dedicated hardware *host bus adapter (HBA)*.
- **SAN FC:** Dedicated Fibre Channel high-performance storage network for applications requiring high I/O to access data directly and sequentially. The FC protocol encapsulates the SCSI frames. This protocol has very little overhead because SCSI packets are sent natively. The server uses a Fibre Channel HBA to access the SAN.
- **SAN FCoE (rarely used in 2012):** Convergence of two worlds: IP and FC networks. FCoE uses Fibre Channel technology, but on a converged FCoE network. A *converged network adapter (CNA)* is the type of card used.

As shown in Figure 3.4, SCSI commands are encapsulated in different layers depending on the protocol used. The more layers used, the more overhead at the host level.

NOTE

Companies sometimes ask which protocol is best in a VMware environment. Obviously, this is a difficult question to answer. It is like asking you to pick the best means of travel between two points without having any context. Obviously, the best mode of transportation for going grocery shopping will not be the same as for going on a vacation abroad. Therefore, before answering the question about the best protocol, you need to know the general context as well as information about the infrastructure, the IT team skills in place, the type of applications (critical, noncritical) to virtualize, performance expectations, financial concerns, and so on.

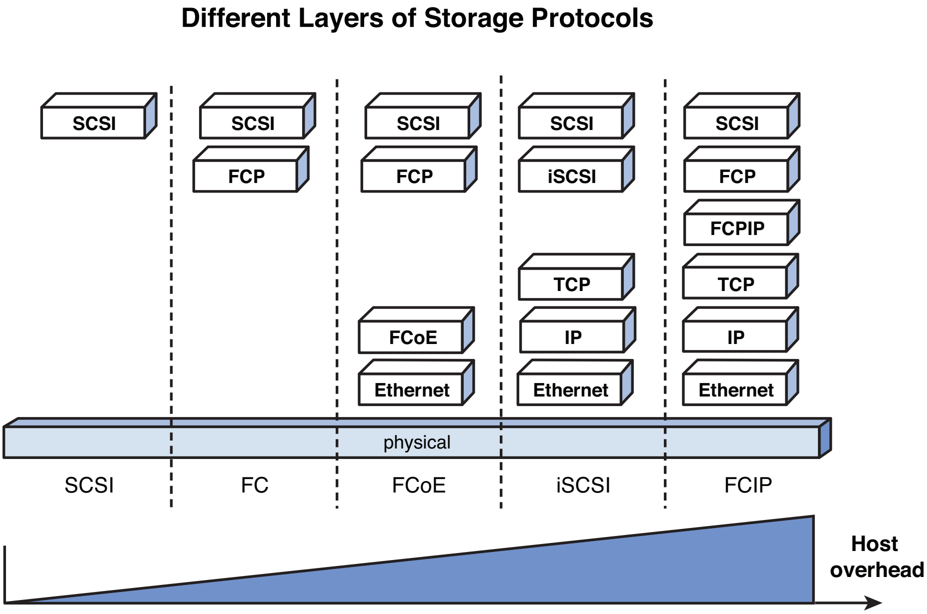


Figure 3.4 Layers of SCSI commands in the different protocols.

Storage Networks

As explained in the preceding section, two networks may be used: the IP Ethernet network (in NAS or iSCSI modes) and the FC network (FC or FCoE).

IP Storage Network

This type of network was not originally designed to offer high-performance storage but rather to carry information between the network's various active elements. Therefore, it is not adapted for applications requiring high performance, such as database applications. The IP network is located at Level 3 of the OSI layer, so it is routable, which favors network interconnectivity over long distances. The FC network is found at Level 2 and, therefore, not routable. Today, throughputs reach 10 GbE, and the future promises 40 GbE and 100 GbE.

The problem with IP networks is that an IP network experiences *packet loss* because of the following factors:

- Signal degradation on the physical line

- Routing errors
- Buffer overflow (when the receptor cannot absorb the incoming flow)

The TCI/IP protocol allows the retransmission of lost packets (if sent data is not acknowledged by the receiver), but this has a dramatic impact on performance.

Another issue is that only a limited quantity of data, called *maximum transmission unit (MTU)*, can be sent in an IP packet. This quantity, called the *payload*, is set at 1500 bytes for an Ethernet packet. Data beyond 1500 bytes must be fragmented before it is sent. Each time a packet is received by the network card, the card sends an interrupt to the host to confirm reception. This adds to the overload at the host and CPU cycle level (called *overhead*). As the number of sent packets increases, routing becomes more complicated and time-consuming.

NOTE

In consolidated virtual environments, the overhead must be taken into account; it should not deteriorate the performance of the host server, which should be entirely dedicated to applications.

To reduce this frame fragmentation, *jumbo frames* were created These allow the transmission of packets larger than 1500 bytes (up to an MTU of 9000 bytes). The jumbo frames play a significant role in improving efficiency, and some studies have shown reductions of 50% in CPU overhead. The MTU must be activated and compatible from the *beginning to the end* of the chain, including physical switches, cards, cables, and so on.

NOTE

A 9000-byte jumbo frame replaces 6 * 1500 bytes from a standard Ethernet packet, generating five times fewer CPU cycles by the host.

Exercise caution. If a problem occurs, the higher the MTU between the source and the target, the larger the packets to retransmit will be, which decreases performance and increases latency. To make the most of jumbo frames, the network must be robust and well implemented.

IP storage networks have the advantage of being less expensive than SAN FC equipment. Ethernet networks are already in place, so in some cases, less implementation is required, making them easier to use. Furthermore, IT teams have used the technology for several years.

iSCSI in VMware

In the VMware environment, the iSCSI protocol has been supported only since 2006. If deployed in an optimal fashion, this protocol offers very good performance. The IP network is administered by a team other than the storage team.

Advantages**:** iSCSI has been adopted in many activity sectors because it uses the company's TCP/IP network for access in block mode, without the need to invest in FC equipment. For this reason, it is an ideal solution in certain environments because it is much easier to set up. Using the traditional Ethernet network means greater distances can be covered before requiring special conversion equipment (such as FC to IP converters)—for example, for replication. The skills necessary are network skills rather than advanced storage skills.

Disadvantages**:** Tests have proven iSCSI is the protocol that uses the most CPU resources. Therefore, monitoring CPU use is important and should be taken into account when provisioning networks.

The following best practices are recommended:

- Using iSCSI is worthwhile only if the architecture can take full advantage of this protocol by activating jumbo frames (MTU 9000), which provides excellent performance. This activation must exist from one end of the chain to the other.
- Using iSCSI HBA cards becomes essential when using 10-GB connections, and links should be aggregated wherever possible for high performance and redundancy in case of failure.
- It is advisable to physically separate the iSCSI storage network from the standard IP network. If this is not possible, streams should be isolated using *virtual local-area networks (VLANs)*.
- Use cards with *TCP/IP offline engine (TOE)* functionality to unload the host from some instructions related to the iSCSI overlay and to reduce the overhead.
- Implement *quality of service (QoS)* by putting the priority on streams. Using vSphere, this can be done using the *Storage I/O Control (SIOC)* functionality.
- Network packet loss is one of the main challenges to achieving good iSCSI network performance. Packet loss can be caused by faulty network configuration or the wrong quality of wiring (for example, using Category 5 cables rather than Category 6 for gigabit links).

NFS in VMware

Network File System (NFS) is a protocol used by NAS and supported by ESX since 2006. It provides storage sharing through the network at the file-system level. *VMware supports* NFS version 3 over TCP. Contrary to what is sometimes believed, tests show good performance if this protocol is implemented properly. Therefore, it is possible to use this protocol under certain conditions for virtual environments. Activation of jumbo frames (MTU 9000) allows the transmission of 8192 (8 KB) NFS data blocks, which are well suited for the protocol. By default, 8 NFS mounts per ESXi host are possible. This can be extended to 64 NFS mounts.

NOTE

By default, thin provisioning is the format used when a virtual disk is created on NFS datastores.

Advantages**:** Like iSCSI, NFS uses the standard TCP/IP network and is very easy to implement without the need for a dedicated storage infrastructure. It is the least expensive solution, and it does not require particular storage skills. Very often, NAS offers de-duplication, which can reduce the amount of storage space required.

Disadvantages**:** It offers the lowest performance of the described solutions, but it is close to iSCSI's. It makes more use of the host server's CPU than the FC protocol, but less than iSCSI software. Therefore, it could conceivably be used in a production environment with VMs requiring average performance for Tier 2 and Tier 3 applications.

NOTE

With vSphere 5, this protocol does not support NFS boot or the use of RDM.

The following best practices are recommended:

- Use 100 to 400 vmdk files per NFS volume. The maximum possible *logical unit number (LUN)* is 256 for a maximum size of 64 TB per NFS volume. (Manufacturers can provide information about the limit supported by file systems, usually 16 TB.)
- Separate the network dedicated to storage from the Ethernet network by using dedicated switches or VLANs.
- Activate flow control.
- Activate jumbo frames by using dedicated switches with large per-port buffers.

- Activate the Spanning Tree Protocol.
- Use a 10-Gb network (strongly recommended).
- Use full TOE cards to unload ESXi host servers.

Fibre Channel Network

Essentially, the Fibre Channel network is dedicated to storage that offers direct *lossless* access to block mode data. This network is designed for high-performance storage with very low latency, through advanced mechanisms such as *buffer credits* (a kind of buffer memory used to regulate streams in a SAN). The FC protocol (FC) encapsulates SCSI packets through a dedicated Fibre Channel network. Speeds are 1, 2, 4, 8, or 16 Gbps. FC packets carry a payload of 2112 bytes.

SAN FC in VMware

Fibre Channel (FC) is the most advanced of the protocols supported by VMware. This is why it is most often the one used by clients in their production environment.

Advantages**:** It seems to be a given today that FC is the most high-performing protocol, as well as the one using the least host-server CPU resources when compared to NFS and iSCSI. High performance levels can be reached, and because the technology is lossless, the network is predictive. This protocol works for all popular applications and is ideal for those that are I/O intensive such as databases or *enterprise resource planning (ERP)* applications.

Disadvantages**:** FC is the most expensive solution because it involves building a specialized storage architecture and requiring an investment in HBA cards, FC switches, *small form-factor pluggable (SFP)* ports, and cables. Moreover, implementing this solution is more complex and requires specialized storage skills. Training is required, and so is learning new terminology to manage the SAN, such as *LUN masking*, *zoning WWN*, and *fabrics*.

The following best practices are recommended:

- To reduce broken links, insert several HBA cards in each server and use numerous storage array access paths.
- Use load-balancing software, such as native ESXi Round Robin, or EMC PowerPath/Virtual Edition, to optimize path management between servers and storage.
- Use ALUA-compliant storage arrays that are compatible with VMware's VAAI APIs.

- Use the same number of paths between all members of the cluster, and all host servers within a cluster should see the same volumes.
- Comply with the connection compatibility matrix between the members of the ESXi cluster and storage.

NOTE

We have met administrators who connected their servers directly to controller storage arrays to save on the cost of switches! Such a practice cancels the redundancy benefit offered by a SAN FC and is therefore discouraged.

- Use the same speed of switches in all connections to avoid the creation of contention points in the tSAN.

EXAMPLE

A company had new blade servers with FC 8-Gb ports. The SAN's core switch was at 4 Gbps. A significant contention appeared on the core switch, which had an impact on all connected elements. The corrective action was to force the new equipment to adopt the fabric speed at 4 Gbps.

- Check firmware levels on FC switches and HBAs, and follow instructions from the storage array manufacturer.

SAN FCoE in VMware

Fibre Channel over Ethernet (FCoE) represents the convergence of various fields: Ethernet for the network (TCP/IP), SAN for storage (SAN FC), and InfiniBand for clustering (IPC). This means a single type of card, switch cables, and management interface can now be used for these various protocols. FC frames are encapsulated in Ethernet frames that provide transport in a more efficient manner than TCP/IP.

FCoE frames carry a payload of 2500 bytes. The goal is to render Ethernet lossless like FC. This is achieved by making the physical network more reliable and by making a number of improvements, especially with regard to QoS. Dedicated equipment is required, as is the activation of jumbo frames (2180 bytes). Congestion is eliminated through stream-management mechanisms.

Because FCoE remains relatively uncommon in 2012, we lack practical experience regarding the advantages and disadvantages of this type of protocol in a VMware environment.

Which Protocol Is Best for You?

In our experience, SAN FC is the protocol administrators prefer for virtual production environments. An estimated 70% of customers use SAN FC for production in VMware environments. The arrival of 10 GbE with jumbo frames, however, allows the easy implementation of a SAN IP infrastructure while maintaining a level of performance that can suffice in some cases. Aside from technical criteria, the optimal choice is based on existing architectures and allocated budgets.

To summarize

- SAN FC should be favored for applications that require high performance (Tier 1 and Tier 2), such as database applications.
- iSCSI can be used for Tier 2 applications. Some businesses use IP in iSCSI for remote data replication, which works well and limits costs.
- NAS can be used for network services such as infrastructure VMs—domain controller, DNS, file or noncritical application servers (Tier 3 applications)—as well as for ISO image, template, and VM backup storage.

VMFS

Virtual Machine File System (VMFS) is a file system developed by VMware that is dedicated and optimized for clustered virtual environments and the storage of large files. The structure of VMFS makes it possible to store VM files in a single folder, simplifying VM administration.

Advantages**:** Traditional file systems authorize only a single server to obtain read/write access to a storage resource. VMFS is a so-called *clustered* file system—it allows read/write access to storage resources by several ESXi host servers simultaneously. To ensure that several servers do not simultaneously access the same VM, VMFS provides a system called *on-disk locking*. This guarantees that a VM works with only a single ESXi server at a time. To manage access, ESXi uses a *SCSI reservation* technique that modifies metadata files. This very short locking period prevents I/O on the entire LUN for any ESXi server and for VMs. This is why it is important not to have frequent SCSI reservations, because they could hinder performance.

The SCSI reservation is used by ESXi when

- Creating a VMFS datastore
- Expanding a VMFS datastore onto additional extends
- Powering on a VM
- Acquiring a lock on a file
- Creating or deleting a file
- Creating a template
- Deploying a VM from a template
- Creating a new VM
- Migrating a VM with vMotion
- Growing a file (for example a*VMFS)* snapshot file or a thin provisioned virtual disk)
- HA functionality is used (if a server fails, disk locking is released, which allows another ESXi server to restart VMs and use disk locking for its own purposes)

NOTE

One particular VAAI feature, hardware-assisted locking reduces SCSI reservations. This API unloads locking activity directly through to the storage array controllers.

VMFS-5 Specifications

vSphere 5 introduces VMFS5 with a maximum size of 64 TB. Table 3.1 outlines the evolution of VMFS from version 3 to version 5.

Table 3.1 VMFS-3 Versus VMFS-5

Functionalities	VMFS3	VMFS-5
Maximum volume	2 TB	64 TB
Block size	1, 2, 4, or 8 MB	1 MB
Sub-blocks	64 Kb	8 Kb
Small files	No	1 Kb

VMFS-5 offers higher limits than VMFS-3 because the addressing table was redeveloped in 64 bits. (VMFS-3 offered a 32-bit table and was limited to 256,000 blocks of 8 MB [or 2 TB].) With VMFS-5, blocks have a fixed size of 1 MB, and the maximum volume is 64 TB. With VMFS-3, blocks vary in size between 1 MB and 8 MB, which can cause virtual disk maximum size issues if the block size is too low. (For example, 1 MB blocks are limited to 256 GB vmdk files, and the volume must be reformatted using the right size of blocks for a larger file size.) Sub-blocks go from 64 KB to 8 KB, with the possibility of managing files as small as 1 KB.

You should also note the following:

- A single VMFS datastore must be created for each LUN.
- VMFS keeps an event log. This preserves data integrity and allows quick restoration should problems arise.

Upgrading VMFS-3 to VMFS-5

VMFS-3 is compatible with vSphere 5. The upgrade from VMFS-3 to VMFS-5 is supported and occurs without service interruption while VMs are running. Creating a new VMFS volume is preferable, however, because the VMFS-3 to VMFS-5 upgrade carries the following limitations:

- Blocks keep their initial size (which can be larger than 1 MB). Copy operations between datastores with different block sizes will not benefit from the VAAI feature full copy.
- Sub-blocks remain at 64 KB.
- When a new VMFS-5 volume is created, the maximum number of files remains unchanged at 30,720 instead of a maximum of 100,000 files.
- The use of a *master boot record (MBR)* type partition remains, but it is automatically changed to a *GUID partition table (GPT)* when volumes are larger than 2 TB.

Signature of VMFS Datastores

Each VMFS datastore has a *universal unique identifier (UUID)* to identify on which LUN the VMFS datastore is located. This UUID must be unique. If two VMFSs are simultaneously mounted with the same UUID, ESXi does not know on which volume to perform read/write operations (it will send at random to each volume), which can lead to data corruption. vSphere detects this situation and prevents it.

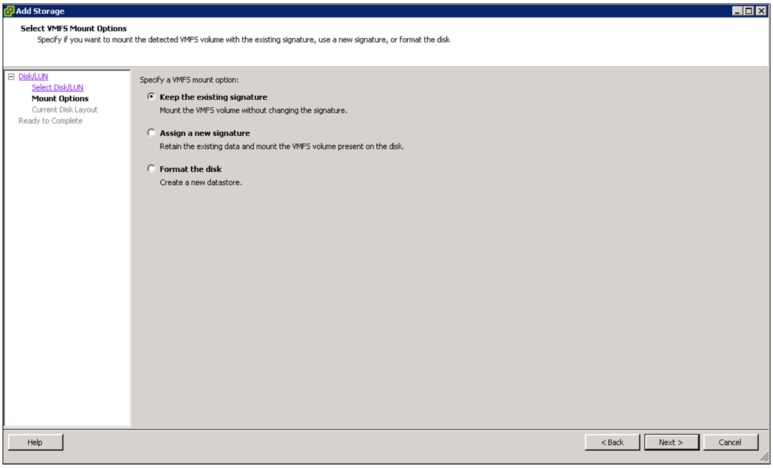
NOTE

The UUID is stored in the header of the VMFS file system and is generated from four variables: the date, time, part of an ESXi MAC address, and LUN identifier in the storage array. This ensures a unique value within the environment and forms part of the VMFS volume's metadata.

When a VMFS LUN is replicated, snapshot, or cloned, the VMFS LUN created is 100% identical to the original, including the UUID. To exploit this new VMFS LUN, it is possible to assign a new signature or to keep the same signature as the original under certain conditions, using the following options (shown in Figure 3.5):

NOTE

Volume signature applies to VMFS file systems in FC or iSCSI, but not to NFS volumes.



Add Storage

Select VMFS Mount Options

Specify if you want to mount the detected VMFS volume with the existing signature, use a new signature, or format the disk

Specify a VMFS mount option:

● **Keep the existing signature**
Mount the VMFS volume without changing the signature.

○ **Assign a new signature**
Retain the existing data and mount the VMFS volume present on the disk.

○ **Format the disk**
Create a new datastore.

Figure 3.5 Options offered when remounting a replicated or snapshot LUN.

- **Keep the Existing Signature:** This option enables the preservation of the same signature and mounting of the replicated datastore. To avoid UUID conflicts, such mounting can be performed only in cases where the source VMFS LUN is unmounted (or removed).
- **Assign a New Signature:** When re-signaturing a VMFS, ESXi assigns a new UUID and name to the LUN copy. This enables the simultaneous mount of both

VMFS datastores (the original volume and its copy) with two distinct identifiers. Note that re-signaturing is irreversible. Remember to perform a datastore rescan to update the LUNs introduced to the ESXi.

NOTE

Re-signaturing of a VMFS datastore has consequences if the datastore contains VMs. Indeed, each VM's configuration files (vmx, vmsd, and vmdk files) specify on which datastores the VM's virtual disks are located, based on the UUID value. In the case of a volume re-signaturing, the UUID values in these files are no longer correct because they point to the former VMFS with its former UUID. VMs must be re-registered in vCenter to integrate the new UUID, and the datacenter, resource pools, network mappings,

- **Format the Disk:** This option entirely reformats the volume.

Re-Signature of a VMFS Volume as Part of a DRP

A new UUID is generated when implementing a *disaster recovery plan (DRP)* and in cases where the replicated volume changes signature. The vmx and vmdk configuration files from the VMs recorded on the volume point to the former UUID rather than to the new volume. Therefore, all VMs part of the DRP must be manually removed from the vCenter's inventory and re-recorded to recuperate the new UUID. This can be a cumbersome process and can lead to handling errors when these operations are performed manually.

One of the valuable propositions offered by *Site Recovery Manager 5 (SRM5)* is the automation of this workflow to simplify the process and avoid errors. With SRM5, the replicated volume is re-signatured on the backup site, and configuration files are automatically referenced with the proper UUID, pointing the VMs to the new replicated volume. Each protected VM is associated with the virtual disks assigned to it.

NOTE

Manual operations are further complicated when RDM volumes are used because the RDM's VMFS pointer no longer exists. SRM also allows the automatic remapping of these volumes with new reinventoried VMs.

Technical Details

Within the environment, a VMFS volume is represented in the following ways:

- By its UUID (for example, 487788ae-34666454-2ae3-00004ea244e1).
- By a network address authority (NAA) ID (for example, naa.5000.xxx). vSphere uses NAA ID to detect the UUID with which the LUN ID is associated.
- By a label name seen by ESXi and a datastore name seen by vCenter Server (for example, myvmfsprod). This name is provided by the user and is only an alias pointing to the VMFS UUID, but it makes it easier to find your way.
- By a VMkernel device name, called *runtime name* in vCenter *(for* example, vmhba1:0:4).

When re-signaturing a VMFS, ESXi assigns a new UUID and a new label name for the copy and mounts the copied LUN like an original. The new associated name adopts the format type snap—for example, *snapID-oldLabel*, where *snapID* is an integer and *oldLabel* is the datastore's former name.

Besides snapshots and replication, other operations performed on a datastore are seen by ESXi as a copy of the original and, therefore, require action from the administrator:

- **LUN ID change:** When changing a LUN ID, vSphere detects that the UUID is now associated with a new device.
- **Change of SCSI type:** For example, going from SCSI-2 to SCSI-3.
- **Activation of SPC-2 compliance for some systems:** For example, EMC Symmetrix requires this activation.

Rescanning the Datastore

After each storage-related modification at the ESXi or storage level, it is necessary to rescan storage adapters to take the new configuration into account. This allows updating of the list of visible datastores and related information.

Rescanning is required each time the following tasks are performed:

- Changing zoning at the SAN level, which has an impact on ESXi servers
- Creating a new LUN within the SAN or performing a re-signature
- Changing the LUN masking within the storage array
- Reconnecting a cable or fiber
- Changing a host at the cluster level

By default, VMkernel scans LUNs from 0 to 255. (Remember, the maximum number of LUNs that can be introduced to a host is 256.) To accelerate the scanning process, it is possible to specify a lower value in the advanced parameters: Disk.MaxLUN (for example, 64 in Figure 3.6).

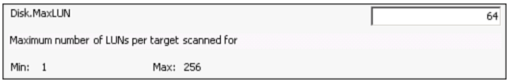

Disk.MaxLUN 64
Maximum number of LUNs per target scanned for
Min: 1 Max: 256

Figure 3.6 Performing a datastore scan.

> **NOTE**
>
> You can also start a rescan of datastores by right-clicking the datacenter, cluster, or folder containing the relevant hosts.

Alignment

Alignment is an important issue to take into account. Stacking up various layers can create nonaligned partitions, as shown in Figure 3.7. Contrast this with an example of aligned partitions, shown in Figure 3.8.

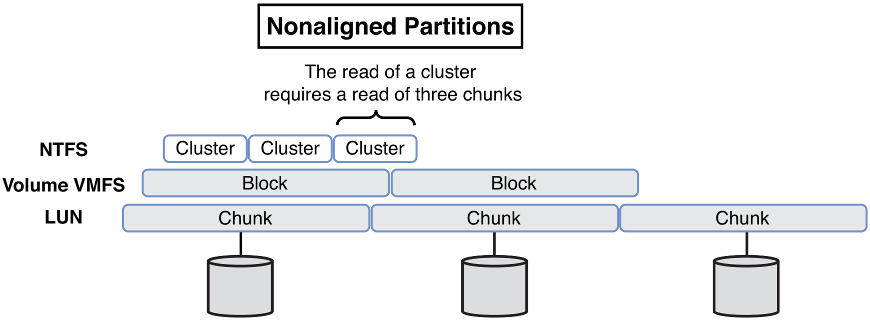


Figure 3.7 Nonaligned partitions.

The smallest unit in a RAID stack is called a *chunk*. Under it is the VMFS, which uses 1-MB blocks. Above it is the formatted NTFS using blocks of 1 KB to 64 KB (called the *disk cluster*). If these layers are not aligned, reading a cluster can mean reading two blocks overlapping three chunks on three different hard drives, which can offset writing and, thus, decrease performance.

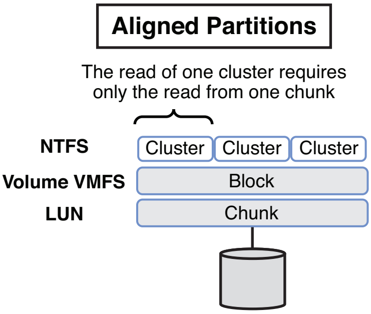


Figure 3.8 Aligned partitions.

When the partition is aligned, a cluster should read a single block, itself aligned with a chunk. This alignment is crucial, and in a VMware environment, nonalignment can cause a 40% drop in performance.

In a Microsoft environment, Windows Server 2008 is automatically aligned, whereas older operating systems must be aligned using the Diskpart utility. See the software publisher's instructions.

Increasing Volume

Volume Grow allows the dynamic extension of an existing VMFS without shutting down the VMs (up to 32 extensions). When a physical storage space is added to a LUN, the existing datastore can be extended without shutting down the server or the associated storage. This complements storage array options, which allow the dynamic extension of LUNs. Extending the storage space of a virtual disk (vmdk) is also possible in persistent mode without snapshots, using *Hot VMDK Extend*. It is recommended that extensions be put on disks with the same performance.

The vmdk extension and the visibility of the disk's free space depend on the OS mechanism and its file system. *Depending on the OS version, third-party tools might be required to extend a system partition, as is the case with Windows 2003. To find out more, refer to VMware's Knowledge Base: KB 1004071.*

Can a Single Large 64-TB Volume Be Created to Host All VMs?

With vSphere 5, the maximum size for a LUN VMFS-5 is 64 TB. Theoretically, a single, very large 64-TB volume could be created. Because the storage arrays integrate VMware's APIs (VAAI), they offer excellent volume-access performance. However, *we do not recommend adopting this approach*, for the following reasons:

- Separating environments is absolutely essential; production, testing, receipt, and backup should each have its own dedicated environment and LUN. It is important not to mix I/O profiles when these are known (random versus sequential access, for example) and not to balance the load based on the VMs' activities (even though Storage DRS allows load balancing).
- During migrations, because migrating a large volume is more complex than migrating several small volumes, which can be performed in stages.
- When a large volume gets corrupted, the impact is more significant than if the volume is smaller, containing fewer VMs.

Because of the preceding issues, creating separate LUNs is the preferred approach. It also makes replication easier (for example, by allowing protection to apply only to the critical environment).

Best Practices for VMFS Configuration

The following best practices are recommended:

- Generally, you should create VMFS volumes between 600 GB and 1 TB and use 15 to 20 active vmdks per volume (no more than 32). (A VM can have several active vmdks.)
- For environments that require high levels of performance, such as Oracle, Microsoft SQL, and SAP, the RDM mode is preferable.
- VMware recommends the use of VMFS over NFS because VMFS offers the complete set of capabilities and allows the use of RDM volumes for I/O-intensive applications.
- To avoid significant contentions, avoid connecting more than eight ESX servers to a LUN.
- Avoid placing several VMs with snapshots on the same VMFS.
- Avoid defining the DRS as aggressive because this will trigger frequent VM migration from one host server to another and, therefore, frequent SCSI reservations.
- Separate production LUNs from test LUNs, and store ISO files, templates, and backups on dedicated LUNs.
- Align vmdk partitions after the OS is configured for new disks.

- Avoid grouping several LUNs to form a VMFS because the different environments cannot be separated (the production environment, test environment, and templates), which increases the risk of contentions, with more frequent reservations.
- Avoid creating one VMFS per VM because it increases the number of LUNs and makes management more complex while limiting expansions to 256 LUNs or 256 VMs.

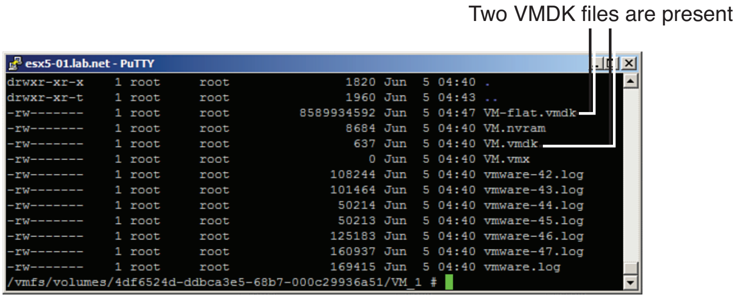
Virtual Disk

Just like a traditional hard drive, the virtual disk contains the OS, applications, and data. A VM's virtual disk is represented by a vmdk file or by an RDM volume.

VMDKs

The vmdks are the most important files because they are the VM's virtual disks, so they must be protected and secured. In vSphere 5, the vmdk's maximum size is 2 TB (more precisely, 2 TB minus 512 bytes). Two files make up a virtual disk: a descriptor bearing the extension .vmdk, and a file containing data, using the extension –flat.vmdk, which you can see in the command-line interface (Figure 3.9) or in the graphical user interface (Figure 3.10).

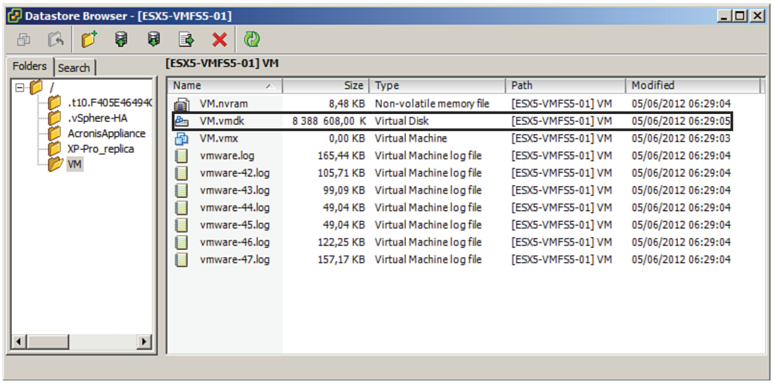
- The vmdk file corresponds to a metadata file, which is the virtual disk's description (editable file in some support maintenance needs). This file provides the link to the –flat.vmdk file and contains information regarding the UUID. (See the section "Re-signature of a VMFS Volume as Part of a DRP," earlier in this chapter.)
- The -flat.vmdk file corresponds to the virtual disk with its content.



Two VMDK files are present

```
drwxr-xr-x    1 root     root              1820 Jun  5 04:40 .
drwxr-xr-t    1 root     root              1960 Jun  5 04:43 ..
-rw-------    1 root     root        8589934592 Jun  5 04:47 VM-flat.vmdk
-rw-------    1 root     root              8684 Jun  5 04:40 VM.nvram
-rw-------    1 root     root               637 Jun  5 04:40 VM.vmdk
-rw-------    1 root     root                 0 Jun  5 04:40 VM.vmx
-rw-------    1 root     root            108244 Jun  5 04:40 vmware-42.log
-rw-------    1 root     root            101464 Jun  5 04:40 vmware-43.log
-rw-------    1 root     root             50214 Jun  5 04:40 vmware-44.log
-rw-------    1 root     root             50213 Jun  5 04:40 vmware-45.log
-rw-------    1 root     root            125183 Jun  5 04:40 vmware-46.log
-rw-------    1 root     root            160937 Jun  5 04:40 vmware-47.log
-rw-------    1 root     root            169415 Jun  5 04:40 vmware.log
/vmfs/volumes/4df6524d-ddbca3e5-68b7-000c29936a51/VM_1 #
```

Figure 3.9 Command-line interface showing the vmdk files.



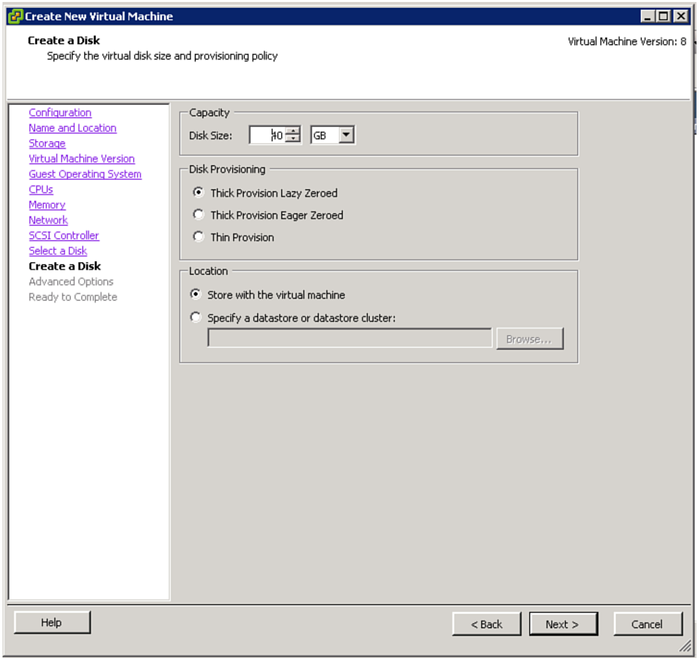
[ESX5-VMFS5-01] VM

Name	Size	Type	Path	Modified
VM.nvram	8,48 KB	Non-volatile memory file	[ESX5-VMFS5-01] VM	05/06/2012 06:29:04
VM.vmdk	8 388 608,00 K	Virtual Disk	[ESX5-VMFS5-01] VM	05/06/2012 06:29:05
VM.vmx	0,00 KB	Virtual Machine	[ESX5-VMFS5-01] VM	05/06/2012 06:29:03
vmware.log	165,44 KB	Virtual Machine log file	[ESX5-VMFS5-01] VM	05/06/2012 06:29:04
vmware-42.log	105,71 KB	Virtual Machine log file	[ESX5-VMFS5-01] VM	05/06/2012 06:29:04
vmware-43.log	99,09 KB	Virtual Machine log file	[ESX5-VMFS5-01] VM	05/06/2012 06:29:04
vmware-44.log	49,04 KB	Virtual Machine log file	[ESX5-VMFS5-01] VM	05/06/2012 06:29:04
vmware-45.log	49,04 KB	Virtual Machine log file	[ESX5-VMFS5-01] VM	05/06/2012 06:29:04
vmware-46.log	122,25 KB	Virtual Machine log file	[ESX5-VMFS5-01] VM	05/06/2012 06:29:04
vmware-47.log	157,17 KB	Virtual Machine log file	[ESX5-VMFS5-01] VM	05/06/2012 06:29:04

Figure 3.10 vCenter's GUI showing a single vmdk file, the same size as the virtual disk.

Disk Types

When a VM is created, the following disk types can be used: *thick disk* (*lazy zeroed* or *eager zeroed*) or *thin disk*, depending on the option you select, as shown in Figure 3.11. Table 3.2 compares the advantages of these disk types.



Create a Disk
Specify the virtual disk size and provisioning policy

Virtual Machine Version: 8

Capacity

Disk Size: 40 GB

Disk Provisioning

● Thick Provision Lazy Zeroed
○ Thick Provision Eager Zeroed
○ Thin Provision

Location

● Store with the virtual machine
○ Specify a datastore or datastore cluster:

Figure 3.11 Optional disk types.

Table 3.2 Disk Types and Their Respective Advantages

	Advantage	When to Use
Zeroed thick	Creation is faster, but performance is lower for the first writes.	Standard when creating a VM.
Eager zeroed thick	Longer to create, but performance is better during the first writes.	Cloning a VM or deploying a VM from a template uses this mode.
Thin	Very rapidly created, but write performance is not as good as in other modes.	The NFS datastore uses this mode by default.

Thick Disks

Thick disks are easier to administer because, after they are provisioned, verifying the space available for the VM is no longer required. However, this means additional costs because the disk space is not optimized. This type of disk supports the *Fault Tolerance (FT)* feature.

With a thick disk, the size of the vmdk file is equal to the size of the disk configured when creating the VM.

Thick disks have two formats:

- **Lazy zeroed (also called zeroed):** This is the default format. All disk space is allocated, but the data previously written at disk level is not deleted. Existing data in the storage space is not deleted but remains on the physical disk. Erasing data and zeroing out blocks (formatting) is done only when first writing on the disk, which somewhat deteriorates performance. This performance degradation is greatly reduced by the VAAI feature Block Zero (utilizing the SCSI command write same).
- **Eager Zeroed:** All disk space is reserved; data is entirely deleted from the disk, and blocks are zeroed out (formatted) when the disk is created. Creating such a disk takes longer, but security is improved because previous data is removed and deleted. Compared with zeroed thick, it offers much better performance when writing on the disk.

The thick disk format is recommended for applications that require high levels of performance. A simple way to use this mode is to select Support Clustering Features Such As Fault Tolerance when configuring VM disks.

It is always faster to create a new VM than to create one from a clone or to deploy a template.

Thin Disks

Some studies show that 40% to 60% of disk space is allocated but never used. In cases where the thin disk option (called thin provisioning) is used, the reserved space on VMFS equals the space actually used on the disk. The size of this space increases dynamically so that storage space is optimized.

EXAMPLE

A 20-GB file is created but only 6 GB are used

With thin disks, the space taken up by the vmdk file in the storage space is 6 GB, whereas with thick disks, the vmdk file uses 20 GB of storage space.

In thin mode, performance is inferior because the space is allocated dynamically upon request and disk blocks must be zeroed out. Thin disks are useful for avoiding wasted storage space, but they require particular care and supervision to ensure that there is no shortage of storage space. The Out of Space API allows proactive monitoring and alerting to prevent this situation.

NOTE

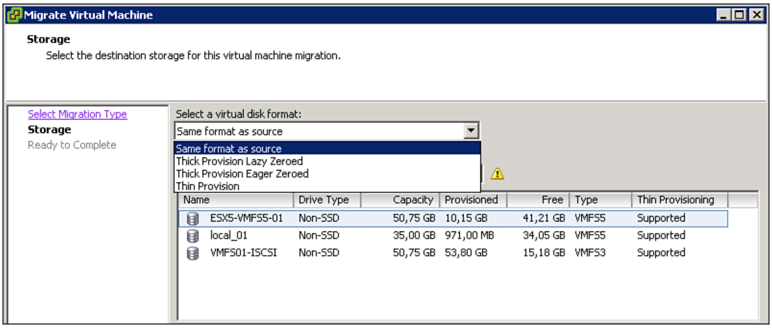
Using a thin LUN is very useful when replication is implemented because the first synchronization replicates only the data used on the disk. For a thick LUN, all data must be replicated, even if the blocks are empty. Initial synchronization is greatly reduced with a thin LUN.

NOTE

Avoid using storage array based thin provisioning in conjunction with vmdk disks in thin mode because keeping things straight becomes very difficult, and it's easy to make interpretation errors.

You can convert a disk from thin to thick by using either of the following methods:

- Use the Inflate option in the Datastore Browser.
- Use Storage vMotion to change the disk type to thick, as shown in Figure 3.12.



Migrate Virtual Machine

Storage

Select the destination storage for this virtual machine migration.

Select a virtual disk format: Same format as source

Name	Drive Type	Capacity	Provisioned	Free	Type	Thin Provisioning
ESX5-VMFS5-01	Non-SSD	50,75 GB	10,15 GB	41,21 GB	VMFS5	Supported
local_01	Non-SSD	35,00 GB	971,00 MB	34,05 GB	VMFS5	Supported
VMFS01-ISCSI	Non-SSD	50,75 GB	53,80 GB	15,18 GB	VMFS3	Supported

Figure 3.12 Using Storage VMotion interface to change disk type.

Modes

There are three modes for virtual disks, as follows:

- **Independent persistent:** All disk writes by the VM are actually written on disk (in the vmdk file). Even after rebooting, the modifications are preserved. This mode offers the best performance.
- **Independent nonpersistent:** All changes made since the VM was started up are destroyed when it is shut down. Modifications are written in a file that logs all changes at the VM's file system level. In this mode, rebooting the VM means going back to a reference VM. Performance is not as good.
- **Snapshot** This enables the return to a previous state.

> **NOTE**
>
> Following security rules and associated good practices, avoid nonpersistent disks. When the VM is rebooted, these make it impossible to analyze the logs because everything is put back into its initial state. This prevents investigations and corrective actions in case of security problems.

Raw Device Mapping

Using the *Raw Device Mapping (RDM)* format, raw storage volumes can be introduced to ESX servers. This mode is mainly used in the following situations:

- When a Microsoft cluster (MSCS) is used (the only supported mode)

- When array-based snapshots are taken
- When introducing volumes directly to VMs for high performance (database-type)
- When introducing big SAN volumes to a VM (from 300 TB), avoiding the long P2V volume conversion to vmdk

RDM takes the form of a file (a kind of pointer) stored in a VMFS datastore that acts as a proxy for the LUN volume.

Figure 3.13 illustrates the difference between vmdk and RDM.

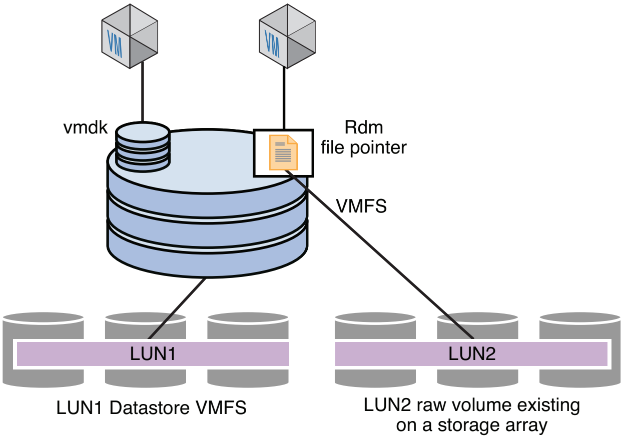


Figure 3.13 vmdk versus RDM format.

The RDM format exists in two modes: RDMv (virtual compatibility mode) and RDMp (physical compatibility mode).

RDMv Disks

The maximum size of an RDMv disk is 2 TB (precisely, 2 TB minus 512 bytes). RDMv is mainly used for large volumes. Beyond 300 GB, introducing dedicated LUNs to the VM can be interesting. Indeed, the vmdk is a file that can be easily moved, but when it is a large file, moving it can be more complex. In this case, the better practice is to introduce the raw volume and use storage array functionalities to move volumes.

RDMv creates a file on the VMFS that acts as a proxy between the VMFS and the LUN in direct link with the VM. It allows the hypervisor to intercept I/O and logs them at need. RDMv authorizes VM snapshots (but not storage array snapshots) as well as the creation of clones and templates.

RDMp Disks

The maximum size of RDMp disks is 64 TB. This type of disk does not allow the hypervisor to intercept I/O. *This means that VM snapshots cannot be taken* (but array-based snapshots are possible), and creating clones or templates is not possible.

In general, RDMp disks are used to introduce to test servers the same data that is in the production databases, using the storage array snapshot functionality. They are also used for MSCS clustering. When using MSCS, the shared disks must not share the virtual controller of the OS.

Some companies might be hesitant about migrating their applications to virtual environments. With RDMp, the change can be done slowly and confidently because the company is free to return to a physical environment if tests are not conclusive in the virtual environment. For applications not officially supported in a virtual environment (for example, older versions of Oracle), RDMp can be used to provide a simple means of replicating the problem in a physical environment that is supported by the software publisher.

RDMp disks cannot be backed up like traditional VMs. The capabilities offered by the two modes are summarized in Table 3.3.

Table 3.3 RDMv Versus RDMp Disks

RDM Type	vMotion	Storage vMotion	Filename	VM Snapshot	Snapshot at Storage Array Level
Rdmv	Yes	Yes	rdm.vmdk	Yes	Not recommended
Rdmp	Yes	Yes	rdmp.vmdk	No	Yes

OVF Format

Current virtual disk formats are vmdk (used by VMware) and *Virtual Hard Disk (vhd)* (used by Microsoft Hyper-V and Citrix XenServer).

Open Virtual Machine Format (OVF) is not a virtual disk format; it is a file format whose particularities facilitate interoperability between the various virtualization and hypervisor platforms. An OVF file includes parameters and metadata such as virtual hardware settings, prerequisites, and security attributes. Provided OVF packages are not limited to one VM and can contain several. An OVF file can be encrypted and compressed.

The OVF template is made up of the following files:

- **MF:** A manifest file, serving to verify the integrity of the OVF template and determine whether it was modified.
- **OVF:** An XML file containing the information on the virtual disk.

- **vmdk:** A virtual disk in VMware, but this file can use a different format to facilitate the interoperability of hypervisors. VMware specifications authorize different types of virtual disks.

NOTE

To simplify moving and manipulating various items for the export of OVF files, it is possible to use the *Open Virtualization Appliance (OVA)* format that groups multiple files into a single file. An OVA file is identical to a TAR file and can actually be renamed to use the .tar extension instead of the .ova extension so that the contents can be seen using a typical archive application.

You can download preconfigured virtual appliances containing an OVF operating system and application solution from https://solutionexchange.vmware.com/store/category_groups/19.

The Datastore

In VMware, the storage space is conceived as a *datastore*. The datastore is a virtual representation of the storage resources on which VMs, templates, or ISO images are stored. The datastore hides the complexity of the different technologies and storage solutions by offering the ESX server a uniform model no matter what storage has been implemented. Datastore types are VMFS and NFS.

NOTE

VMware's best practices recommend the proper separation of datastores used to store templates or ISO images from the datastores used for VMs. We also recommend monitoring the available space of datastores. These must always have at least 25% to 30% available space. This space is required for snapshot or backup operations or for VM swaps. Lack of space can have significant negative consequences and can have an impact on the global performance of the virtual environment.

A *datastore cluster*, also called a *pool of datastores* (*POD*), is a collection of datastores grouped to form a single entity as illustrated in Figure 3.14. When a datastore cluster is created, Storage DRS can be used.

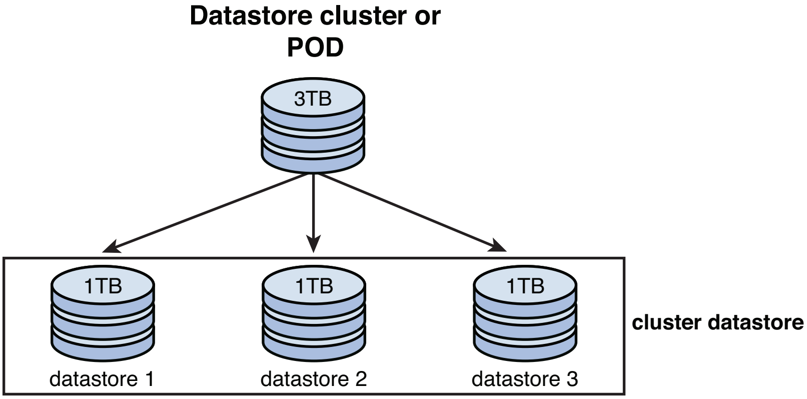


Figure 3.14 *Pool of datastores* (*POD*).

A datastore cluster can consist of volumes from different storage arrays (in terms of performance and volume), and different VMFS can be mixed (VMFS-3and VMFS-5) but this is not generally recommended. The mix of VMFS and NFS volumes in a datastore cluster is not supported.

Storage vMotion

Storage vMotion allows the hot migration of VM virtual disks between two different storage spaces. All files forming the VM migrate from one datastore to another in the same storage array or in another storage array without service interruption. The storage arrays can be from different manufacturers.

NOTE

vMotion migrates a VM from one physical server to another but without moving the files that make up the VM. Storage vMotion moves virtual disks. These two operations cannot be run concurrently on the same VM unless that VM is powered off.

When to Use Storage vMotion

Storage vMotion is used for preventive maintenance operations for storage arrays. It can also be very useful when purchasing a new storage array because it does not require service interruption. Migration is performed very easily, in a fully transparent manner. This relieves administrators from this task, often a cumbersome and sensitive one in a traditional physical environment. Storage vMotion allows the administrator to switch storage array manufacturers and to migrate VMs without the need for a complex compatibility matrix.

NOTE

Using Storage vMotion when there is little storage-level activity is preferable. Before performing migrations with Storage vMotion, it is necessary to confirm that sufficient storage bandwidth exists between the source and destination ESXi servers.

How Storage vMotion Works

A number of improvements were made to Storage vMotion with vSphere 5. Several technologies have been used in the past. In vSphere 4.1, Dirty Block Tracking is used to copy disk blocks between source and destination: full copy, followed by the sending of modified blocks only to the target. (Dirty Block Tracking is a form of *Changed Block Tracking [CBT]* mode.) Issues with this technique were the duration of the switch to the target VM and the malfunction risk in cases of significant I/O chargen in the source VM. In vSphere 5, as shown in Figure 3.15, Storage vMotion makes a full copy of the VM and then uses a mirror driver to split write-modified blocks between the source and destination VMs.

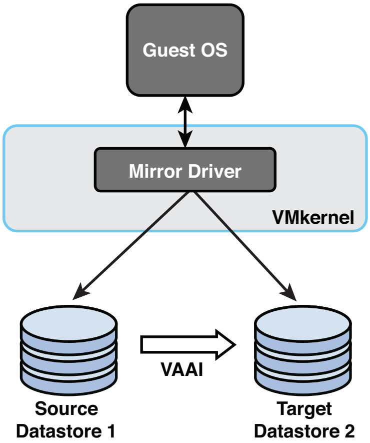


Figure 3.15 Storage vMotion using a mirror driver.

This I/O mirroring is preferable to a succession of disk copies because it has the advantage of guaranteeing migration success even if the destination VM is slow. Migration will be shorter and more predictable.

The following occurs when using Storage vMotion:

1. The VM's working folder is copied to the destination datastore.
2. An image of the VM, called a *shadow VM*, starts on the destination datastore by using the copied files. The shadow VM is paused.
3. Storage vMotion activates a driver (called a *mirror driver*) to write a mirror of the blocks already copied to the destination.
4. The copy of the VM's disk files to the destination datastore is completed while the I/O is being mirrored.
5. Storage vMotion pauses the source VM and transfers the source VM being executed to the shadow VM.
6. The old folder and the VM's disk files are deleted from the source datacenter.

NOTE

The original file is deleted only after the destination file is correctly written and an acknowledgment is sent, which ensures the operation succeeded.

Storage vMotion, available in the Enterprise version, works with VMs that have snapshots and also supports the migration of linked clones.

Storage DRS

Storage DRS (SDRS) allows the automation of the choice of datastore to use for VMs. It makes for more balanced performance and more efficient storage space. This frees up time for administrators, who no longer have to spend time choosing which datastore to use. To function, datastores are grouped in datastore clusters.

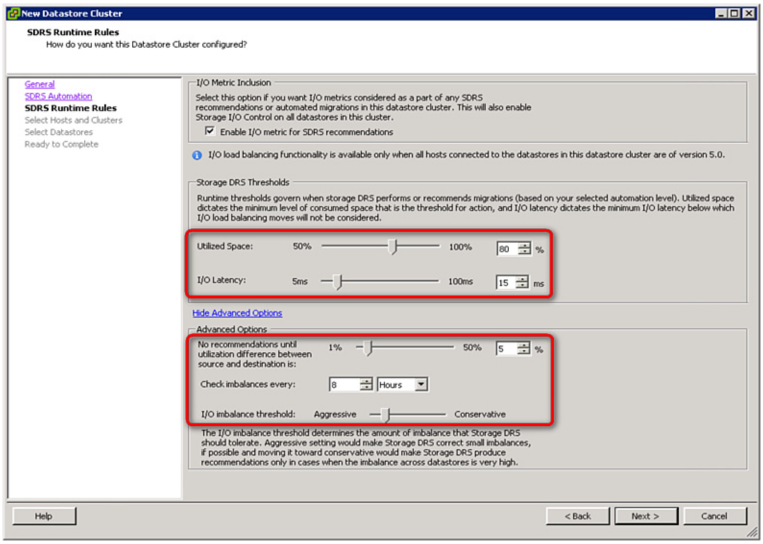
SDRS takes care of the following operations:

- Initial placement of a VM
- Balancing the load between datastores according to the following factors:
    - The use of storage space
    - The I/O chargen based on latency

Initial placement occurs when a VM is created, moved, or cloned. Based on the space used and I/O charge of the cluster's datastores, SDRS provides a particular datastore to store the vmdk.

Datastore Load Balancing

The load balancing is performed every 2 hours for used space and every 8 hours for the I/O load based on the history of the last 24 hours. As shown in Figure 3.16, SDRS makes migration recommendations when a datastore passes the thresholds defined by the user based on the percentage of disk space used (80% by default) and/or the I/O latency (15 ms by default).



SDRS Runtime Rules

How do you want this Datastore Cluster configured?

I/O Metric Inclusion

Select this option if you want I/O metrics considered as a part of any SDRS recommendations or automated migrations in this datastore cluster. This will also enable Storage I/O Control on all datastores in this cluster.

☑ Enable I/O metric for SDRS recommendations

I/O load balancing functionality is available only when all hosts connected to the datastores in this datastore cluster are of version 5.0.

Storage DRS Thresholds

Runtime thresholds govern when storage DRS performs or recommends migrations (based on your selected automation level). Utilized space dictates the minimum level of consumed space that is the threshold for action, and I/O latency dictates the minimum I/O latency below which I/O load balancing moves will not be considered.

Utilized Space: 50% – 100% 80 %

I/O Latency: 5ms – 100ms 15 ms

Hide Advanced Options

Advanced Options

No recommendations until utilization difference between source and destination is: 1% – 50% 5 %

Check imbalances every: 8 Hours

I/O imbalance threshold: Aggressive – Conservative

The I/O imbalance threshold determines the amount of imbalance that Storage DRS should tolerate. Aggressive setting would make Storage DRS correct small imbalances, if possible and moving it toward conservative would make Storage DRS produce recommendations only in cases when the imbalance across datastores is very high.

Figure 3.16 SDRS interface showing load balancing information.

Several levels of automation are possible:

- Manual (default).
- Automatic.
- Planned (scheduled). The planning mode is interesting during a backup period, for example; it is not necessary to move the virtual disks. It is therefore possible to disable SDRS during the backup operations.
- Datastore maintenance mode. A datastore's *maintenance mode* removes all vmdks from the datastore and distributes them to the cluster's other datastores

At this point, you might ask, "How does SDRS detect the datastores' I/O load?"

SDRS uses the SIOC *functionality and an injector mechanism to choose the best target datastore to use.* The injector is used to determine each datastore's characteristics by randomly "injecting" I/O. This allows it to determine the response time and latency associated with each datastore.

Affinity Rules

As illustrated in Figure 3.17, several affinity rules can be applied:

- **Intra-VM vmdk affinity:** All the same vmdk VMs are placed on the same datastore.
- **Intra-VM vmdk anti-affinity:** This rule can be applied to ensure that the vmdks are placed on different datastores. This rule is useful, for instance, to separate a database VM's log disks from its data disks. The rule applies to all or part of the disks within a VM.
- **VM-VM anti-affinity:** Different VM are placed on separate datastores. This offers redundancy for VMs in the event of failure of a datastore.

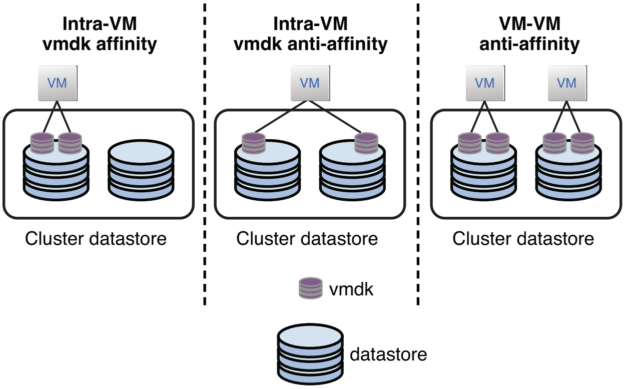


Figure 3.17 Affinity rules.

The current limitations of SDRS are as follows:

- SDRS is not supported with SRM.
- SDRS works only with hosts using ESXi5 or later.

Profile-Driven Storage

Profile-Driven Storage preserves the compliance of VMs with the defined storage needs. This functionality eliminates initial placement errors and simplifies day-to-day management for administrators by automating these tasks. Administrators create profiles that contain the characteristics of the storage. These can be enforced using *vSphere Storage storage-detection APIs (VASA)* or associated with indicators defined by the user (for example, Gold, Silver, Bronze).

A VM's profile is associated during provisioning, creation, migration, cloning, and so on. If the VM is placed in a storage space offering the capacities defined in the VM's storage profile, such storage is deemed compliant. Profile-Driven Storage complements SDRS for initial placement and the automatic migration of vmdk.

Storage I/O Control

Resource sharing creates new challenges. Noncritical VMs should not monopolize available resources. Disk share addresses only part of the issue because sharing is established only with regard to a single ESXi host and is used only when contention occurs at the ESXi host level. This latter scenario is not relevant because VMs located on another ESXi server can use a larger share while being less of a priority. Figure 3.18 illustrates storage sharing with and without *Storage I/O Control (SIOC)*.

For the management of I/O resource allocation to be efficient, it must be independent from the VM's location. The issue must be addressed by sharing the datastore's access resources at the level of the ESXi cluster. This is what SIOC does by placing sharing (QoS) at the cluster level instead of at the ESX level. SIOC monitors the I/O latency of a datastore. When latency reaches a threshold (set at 30 ms by default), the datastore is deemed *congested* and SIOC intervenes to distribute the available resources following sharing rules defined in each VM. Lower-priority VMs have reduced I/O queues. Sharing occurs if and only if contentions appear in storage I/O for access to datastores. Using SIOC guarantees that the most important VMs will have adequate resources no matter what happens, even in the case of congestion.

Using this QoS regarding VM access to the datastore, administrators can confidently consolidate the environment. Even in times of high activity, the most critical VMs will have the necessary resources.

SIOC activation is found in the properties dialog (Figure 3.19) of the datastore. Note that at this time SIOC does not support datastore with multiple extents and RDM disks.

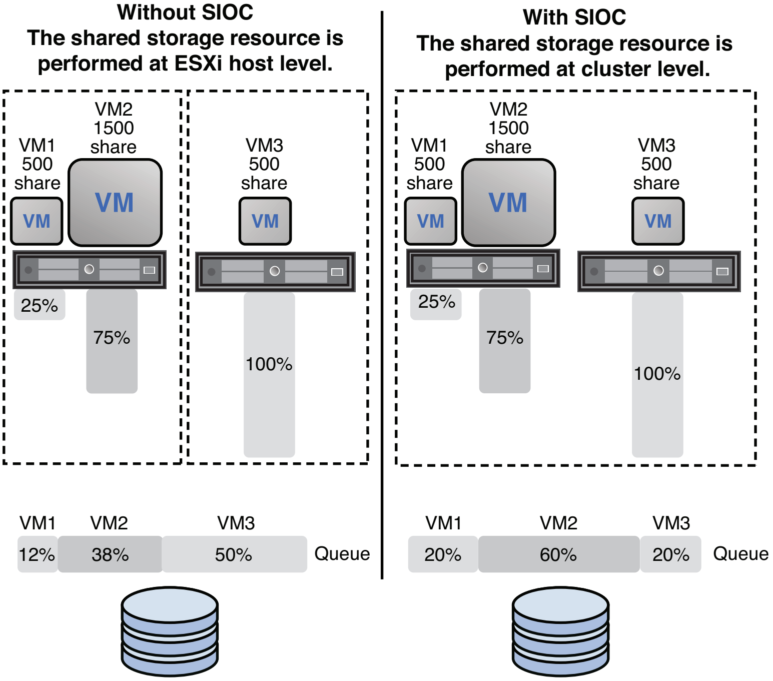
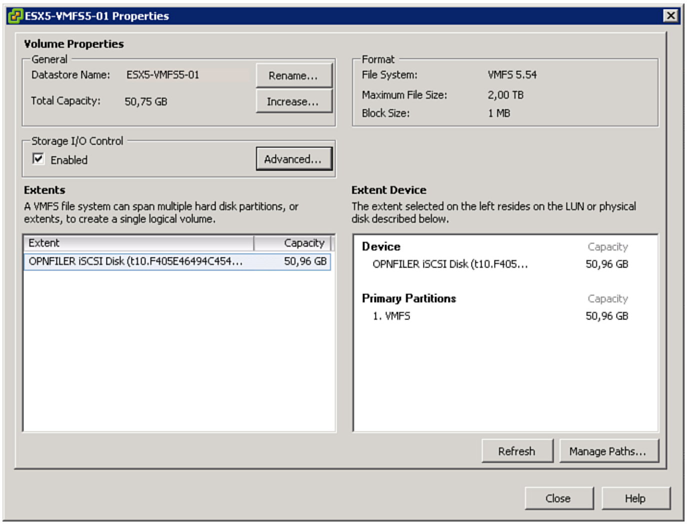

Without SIOC
The shared storage resource is performed at ESXi host level.
With SIOC
The shared storage resource is performed at cluster level.
VM1
500
share
VM2
1500
share
VM3
500
share
VM
25%
75%
100%
VM1
VM2
VM3
12%
38%
50%
Queue
20%
60%
20%
Queue

Figure 3.18 Storage sharing with and without SIOC.



ESX5-VMFS5-01 Properties

Volume Properties

General

Datastore Name: ESX5-VMFS5-01

Total Capacity: 50,75 GB

Format

File System: VMFS 5.54

Maximum File Size: 2,00 TB

Block Size: 1 MB

Storage I/O Control

☑ Enabled

Extents

A VMFS file system can span multiple hard disk partitions, or extents, to create a single logical volume.

Extent	Capacity
OPNFILER iSCSI Disk (t10.F405E46494C454...	50,96 GB

Extent Device

The extent selected on the left resides on the LUN or physical disk described below.

Device	Capacity
OPNFILER iSCSI Disk (t10.F405...	50,96 GB
Primary Partitions	Capacity
1. VMFS	50,96 GB

Figure 3.19 Datastore properties dialog with SIOC enabled.

VMware recommends different threshold values for different disk types:

- **Fibre Channel:** 20 ms to 30 ms
- **Serial Attached SCSI (SAS):** 20 ms to 30 ms
- **Solid State Drive (SSD):** 10 ms to 15 ms
- **Serial ATA (SATA):** 30 ms to 50 ms

vSphere Storage Appliance

vSphere Storage Appliance (VSA) is an appliance designed for *small to medium businesses (SMBs)* (20 to 35 VMs), allowing access to shared storage at a lower cost by using advanced features such as HA, DRS, DPM, FT, and vMotion. The appliance, available only for vSphere 5, is deployed as a VM on each ESXi server (distributed as a 3-GB OVF file). VSAs occupy the available space on the local disks of ESXi servers and show an NFS volume replicated by the ESXi server.

The replication of local storage on another ESXi server ensures redundancy if a host server is out of service. When a VSA node is out of service, the VSA Manager switches IP addresses and shared storage to the replicated VSA. This is done without service interruption for that datastore's VMs.

VSA supports 2 or 3 ESXi servers in a cluster, and up to 25 VMs in a two-node configuration or up to 35 VMs in a three-node configuration.

Therefore, there are two deployment configurations for VSA: Two ESXi servers with one VSA and the VSA cluster service installed on vCenter, or, as illustrated in Figure 3.21, three ESXi servers with one VSA.

VSA Manager (installed as a plug-in in vCenter Server) is the administrative interface of the VSA cluster. It enables monitoring of the cluster's state and allows maintenance and VSA-node replacement operations.

The VSA appliance has the following minimum requirements:

- 6 GB RAM
- 4, 6, or 8 identical disks (same size, same characteristics), configured in RAID 5, RAID 6, or RAID 10
- 4 1-GB network cards
- 2 VLANs configured on physical switches

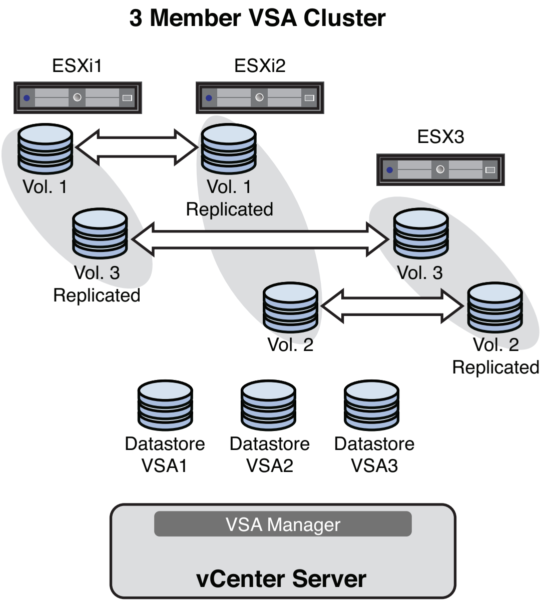


Figure 3.20 VSA deployment configuration.

Because vCenter Server provides VSA management, it must be placed outside the VSA cluster—either in a VM outside the cluster or on a dedicated physical server. Note that this is the only case in which VMware recommends putting vCenter Server on a physical server.

Installing the appliance is rather simple and quick (approximately 10 minutes) and requires no particular storage skill.

VMware Storage APIs

The APIs provided by VMware allow administrators and publishers to extend the functionality of vSphere 5.

vStorage API for Array Integration

The *vStorage API for Array Integration (VAAI)* is a set of application program interfaces allowing interoperability between VMware and storage array manufacturers to communicate with VMware in a smarter manner. Some tasks can be offloaded to the storage array, which lightens the load of ESXi hosts.

NOTE

Processor manufacturers have already integrated Intel VT and AMD V instructions into their chips to reduce high-consuming CPU interceptions. What processor manufacturers have done for servers, VAAIs do for storage arrays. These APIs now seem indispensable to obtain high levels of consolidation.

Table 3.4 lists the VAAI in vSphere 4.1 and VAAI 2 in vSphere 5.

Table 3.4 VAAI Functionality: vSphere 4.1 Versus vSphere 5

VAAI vSphere 4.1	VAAI 2 vSphere 5
Block	
Hardware Assisted Locking	Out of Space
Hardware Accelerated Zero	Space Reclaim
Hardware Accelerated Copy	
NAS	
Not available	Full Clone
	Extended Stats
	Space Reservation

Following is a brief description of each of the features listed in Table 3.4:

- **Hardware Assisted Locking:** Without the API, SCSI reservations are done at the global LUN level. With the API, the work is done at the block level instead of the LUN level, which causes fewer issues related to SCSI reservations and reduces VM startup time, in particular for virtual desktop infrastructure (VDI) projects.
- **Hardware Accelerated Zero:** Without the API, when creating a datastore, the "zero write" is done by the server, which sends SCSI commands to the storage array. With the API, a single command is initiated by the ESX server, and the storage array is responsible for repeating the operation and informing the ESX server when the operation is finished. This reduces traffic between the ESXi server and the storage array.
- **Hardware Accelerated Copy:** Without the API, copy operations are performed from the ESX server toward the storage array. With the API, data is moved within the array by the storage array without going through the server. This reduces the load of the ESXi server and the time required for data migration.

In vSphere 5, new primitives have been defined for VAAI 2:

- **Dead Space Reclaim:** Allows the recovery of spaces that are no longer used when a virtual disk is deleted or after the migration of a virtual disk from one datastore to another by using Storage vMotion on a provisioned thin LUN. ESXi 5.0 transmits the information about the freed-up blocks to the storage system via VAAI through commands, and the storage system recovers the blocks.
- **Thin Provisioning Out of Space API:** Guards against storage-space problems on thin LUNs.
    - **Thin Provisioning LUN Reporting:** Enables the identification of the storage array use in vCenter.
    - **Quota Exceeded:** Displays an alert in vCenter when a capacity threshold is passed within a datastore.
    - **Out of Space Behavior:** The VM determines whether the space is available before the write. If storage space is full, an alert message is displayed in vCenter and this VM is paused (while the other VMs continue running).

The following primitives are defined for NAS VAAI storage:

- **Full-File Clone:** Enables cloning and snapshot operations of vmdk files to be performed by the NAS in a cold manner, similar to VMFS block cloning (Full Copy)
- **Extended Stats:** Enables the visibility of consumed spaces on NFS datastores
- **Space Reservation:** Allows the creation of vmdk files in thick-provisioning mode for NAS storage

vSphere Storage API: Storage Awareness

The *vStorage API for Storage Awareness (VASA)* is a storage-detection API. It allows the visualization, straight from vCenter, of the information related to storage arrays, such as replication, RAID type, compression, de-duplication, thin or thick format, disk type, snapshot state, and performance (IOPS/MBps). Among other things, the vStorage API is used for Profile-Driven Storage.

Multipathing

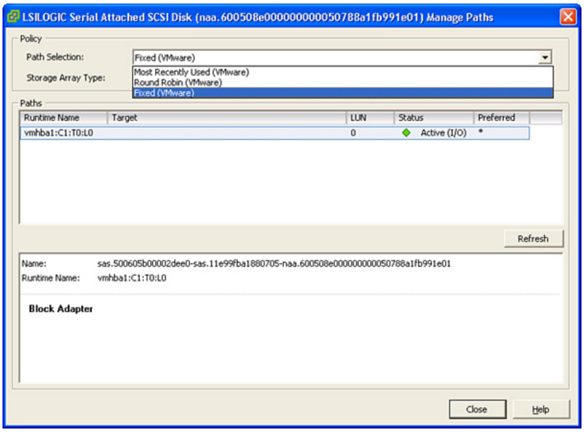
Multipathing can be defined as a solution that uses redundant components, such as adapters and switches, to create logical paths between a server and a storage device.

Pluggable Storage Architecture

Pluggable Storage Architecture (PSA) is a collection of APIs that allows storage manufacturers to insert code directly into the VMkernel layer. Third-party software (for example, EMC PowerPath VE) can thus be developed to offer more advanced load-balancing functionalities in direct relation to their storage array's technology. VMware, however, offers standard basic multipathing mechanisms, called *native multipathing* (*NMP*), divided into the following APIs: *Storage Array Type Plug-in (SATP)*, which is in charge of communicating with the storage array; and " *Path Selection Plug-in (PSP)*, which provides access to load balancing between paths.

As shown in Figure 3.21, VMware offers three PSPs:

- **Most Recently Used (MRU):** Selects the first path discovered upon the boot of ESXi. If this path becomes inaccessible, ESXi chooses an alternate path.
- **Fixed:** Uses a dedicated path designated as the preferred path. If configured otherwise, it uses the path found at boot. When it can no longer use this path, it selects another available path at random. When it becomes available again, ESXi uses the fixed preferred path again.
- **Round Robin (RR):** Automatically selects all available paths and sends the I/O to each in a circular fashion, which allows basic load balancing. PSA coordinates NMP operations, and third-party software coordinates the *multipathing plug-in (MPP)* software."



LSILOGIC Serial Attached SCSI Disk (naa.600508e000000000050788a1fb991e01) Manage Paths

Policy

Path Selection: Fixed (VMware)

Storage Array Type:

Paths

Runtime Name	Target	LUN	Status	Preferred
vmhba1:C1:T0:L0		0	● Active (I/O)	*

Name: sas.500605b00002dee0-sas.11e99fba1880705-naa.600508e000000000050788a1fb991e01

Runtime Name: vmhba1:C1:T0:L0

Block Adapter

Figure 3.21 PSPs offered by VMware.

The NMP Round Robin path-selection policy has a parameter known as the I/O operation limit, which controls the number of I/Os sent down each path before switching to the next path. The default value is 1000; therefore, NMP defaults to switching from one path to another after sending 1000 I/Os down any given path. Tuning the Round Robin I/O operation limit parameter can significantly improve the performance of certain workloads (such as *online transaction processing [OLTP]*). In environments that have random and OLTP workloads, setting the Round Robin parameter to a lower number yields the best throughput, but lowering the value does not improve performance as significantly as it does for sequential workloads. For these reasons, some hardware storage companies recommend that the NMP Round Robin I/O Operation parameter should be lower (can be set to 1).

Third-party software solutions use more advanced algorithms because a limitation of Round Robin is that it performs an automatic distribution without taking into account the actual activity at path level. Some software establishes dynamic load balancing and is designed to use all paths at all times rather than Round Robin, which uses only a single path at a time to bear the entire I/O burden.

Modes

Access to data stored on shared storage space is fundamental in a virtual environment. VMware strongly recommends implementing several access paths to the LUN. Two paths is a minimum, but VMware recommends using four. Multipathing reduces service interruptions by offering a redundancy of access paths to the LUNs. When a path is not available, another is used—without service interruption. These switch mechanisms are called *multipath I/O (MPIO)*.

In VMware, as shown in Figure 3.22, storage can adopt various modes:

- **Active/active:** At a given moment, a LUN is connected to several storage controllers at the same time. The I/O can arrive from several controllers simultaneously.
- **Active/passive:** At a given moment, a single controller owns one LUN (owned LUN). No other controller can send I/O to this LUN as long as it is linked to a controller.
- **ALUA:** Access to a LUN is not direct (nonoptimized) but occurs in an asymmetrical manner, going through the secondary controller.

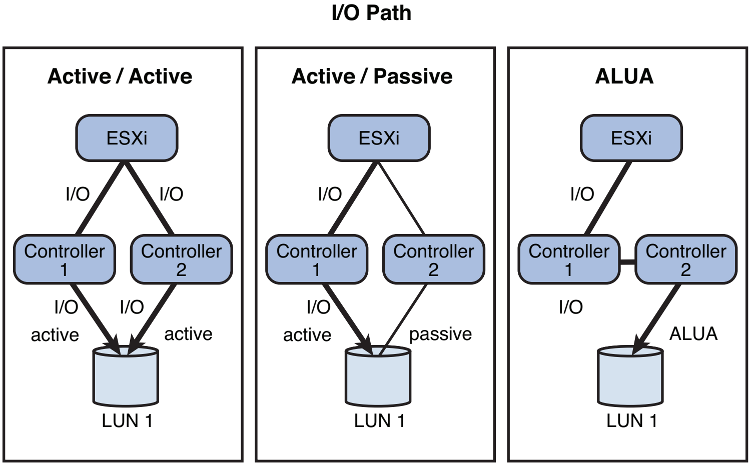


Figure 3.22 Storage modes.

Disk Technology Considerations

This section examines a number of the factors to consider when deciding on the disk technology to use in your environment.

Supported Disk Types

As you have seen, storage architecture is important, and the disk technology plays an important part. ESXi supports a variety of disks, including SSD, SAS, FC, SATA, NL-SAS, IDE, USB, and SCSI.

Many options are available, making it possible to adapt the technology according to several criteria. As shown in Table 3.5, in terms of disk technology, many parameters are to be considered: speed expressed in revolutions per minute and in *I/O per second (IOPS)*, as well as bandwidth transfers.

Table 3.5 Average Speed Parameters for Disk Types (May Vary)

Disks	RPM	IOPS
SSD	N/A	3000
SAS	15 K	180
SAS	10 K	130
NL-SAS	7.2 K	100
SATA	5.4	50

Solid-State Drives (SSDs) are high-performance drives composed of flash memory. These disks are nonmechanical. They are less likely to experience failures, they consume less energy, and they heat up much less than traditional disks. Their access time is low, with very high IOPS (3000 IOPS). They are ideal for reading but not well adapted to a large quantity of writes.

These disks are typically used for log files (for example, for databases). They are often used to extend the cache memory of controllers. (EMC calls them Fast Cache disks, and Netapp calls them Flash Cache disks.) In a VMware environment, these high-performance disks are ideal for storing the swap memory of VMs. They are also very useful for absorbing the charge when activity spikes appear—for example, in a VDI environment, when all VMs boot simultaneously. (This phenomenon is called a *boot storm*.) Disk sizes currently available are 100 GB, 200 GB, and 400 GB. Soon, 800 GB will also be available.

Serial Attached SCSI (SAS) disks replace Fibre Channel disks. These disks are directly connected to the controller, point to point. Revolution speeds are high—10,000 RPM or 15,000 RPM. They are ideal for applications with random access, and they can process small-size I/O of 8 bytes to 16 bytes, typically databases. The stream is bidirectional. Current disk sizes are 300 GB, 600 GB, and 900 GB.

Today, SAS disks are best adapted to virtual environments, and they offer the best price–performance ratio. Although FC disks are still widely found in production environments, the trend is to replace them with SAS disks.

Near-Line SAS (NL-SAS) disks use the mechanics of SATA disks mounted on SAS interfaces. Their advantage over SATA is that they transmit data in full duplex. Read and write can be performed simultaneously, contrary to SATA, which allows only a single read or write at a time. These disks offer features that allow the interpretation of SCSI commands, such as command queuing (to reduce read-head movements), and they provide better control of errors and reporting than SATA disks.

Serial-ATA (SATA) disks allow the management of a large capacity—2 TB, 3 TB, and soon 4 TB. They are recommended for the sequential transfer of large files (for example, backups, video files), but are not suitable for applications with elevated random I/O-like databases (for example, Oracle, Microsoft SQL, MySQL). They are unidirectional and allow a single read or write at a time. Depending on storage array manufacturers, SATA may or may not be recommended for critical-production VMs. Find out from the manufacturer. SATA disks are always well-suited for test VMs or for ISO image, template, or back-up storage.

RAID

Table 3.6 lists recommendations for RAID types and associated traditional uses.

Table 3.6 RAID Types and Traditional Uses

	Write	Read	Use	Protection
RAID0	Excellent	Excellent	Real-time workstation	None (striping)
RAID1	Excellent	Excellent	DB log file, operating system , ESXi Hypervisor	Mirror
RAID5	Good	Very good	DB, ERP, web server, file server, mail	Parity
RAID6	Average	Very good	Archiving, backup, file server	Double parity
RAID10	Excellent	Excellent	Large DB , application servers	Striping + mirror

Storage Pools

In a physical environment, a LUN is dedicated to a server and, thus, to a specific application. In this case, parameters can be set to adapt RAID levels to the application, either sequential or random. This method is not well adapted to a virtual environment. Indeed, because of the dynamic nature of a virtual environment, keeping the same LUN-attribution logic based on the application becomes difficult. VMs are mobile and move from one datastore to another. RAID levels risk not remaining the same. Instead of using dedicated RAID levels, some manufacturers suggest using storage pools. This method is preferable because it offers excellent performance and simplifies management.

Automatic Disk Tiering

Only 20% of a LUN's data is frequently accessed. Statistics also show that 80% of data is unused after two weeks. Through automatic tiering, frequently used data is automatically placed on high-performance SSD or SAS disks, while less frequently used data is stored on lower-performance disks such as SATA or NL-SAS.

Performance

In virtual environments, monitoring performance is complex because of resource pooling and the various layers (for example, applications, hypervisor, storage array). Speeds measured in IOPS and bandwidths in MBps depend on the type and number of disks on which the datastore is hosted. Storage activity should be monitored to determine whether waiting queues form on either of these criteria (*queue length*). At the hypervisor or vCenter level, the most reliable and simplest performance indicator for identifying contentions is the device access time.

Access time through all HBAs should be below 20 ms in read and write. Another indicator that should be monitored and that shows a contention by highlighting an activity that cannot be absorbed by the associated storage is the *Stop Disk* value. This value should always be set to 0. If its value is higher than 0, the load should be rebalanced. There are usually two causes:

- VM activity is too high for the storage device.
- Storage has not been properly configured. (For example, make sure there is no zoning issue, that all paths to storage are available, that the activity is well distributed among all paths, and that the storage cache is not set to forced flush.)

Additional Recommendations

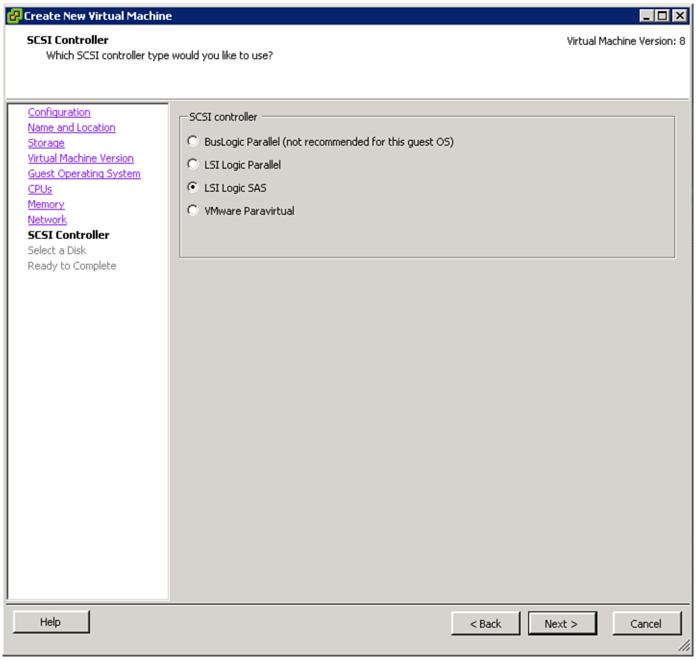
Following are additional recommendations that can improve disk performance:

- Using solid-state cache allows a significant number of I/O to disk. The cache serves as leverage because the major part of the read and write I/O activity occurs in the cache. Databases require very high I/O operations in 4-byte, 8-byte, or 16-byte random access, while video-file backup servers require high speeds with large block sizes (32, 64, 128, or 256 bytes).
- Sequential access and random access should not be mixed on the same disk. If possible, I/Os should be separated by type (read, write, sequential, random). For example, three VMs hosting one transactional DBMS-type database should each have three datastores:
    - One datastore for the OS in RAID 5. Separating the OS means a VM can be booted without drawing from the database's available I/Os from RAID 5.
    - One datastore for the RAID 5 database if the read/write ratio is 70%/30%. If not, the RAID type should be changed A database generally uses 70% of random read-type transactions.
    - One datastore for logs in RAID 1 because writes are sequential (or RAID 10 for large databases with a high write rate).

Device Drivers

Figure 3.23 displays the standard SCSI controllers that ESXi makes available to the guest OS.

Among the available options, BusLogic offers a greater compatibility with older-generation operating systems. LSI Logic Parallel offers a higher performance on some operating systems, but a driver must be added.



SCSI Controller
Which SCSI controller type would you like to use?

Virtual Machine Version: 8

SCSI controller

○ BusLogic Parallel (not recommended for this guest OS)
○ LSI Logic Parallel
● LSI Logic SAS
○ VMware Paravirtual

Figure 3.23 Standard SCSI controller options.

The VM VMware Paravirtual device driver is used mainly for heavy workloads. It directly accesses the virtualization layer. This driver is said to be paravirtualized because it interprets requests from the guest OS and sends them directly to the hypervisor. This reduces request interceptions by the *Virtual Machine Monitor (VMM)*, which improves performance. This option works only with Windows Server 2003, Windows Server 2008, and RHEl 5.

Another available controller, VMDirectPath I/O, is a storage I/O driver that allows direct access to the physical device without going through ESXi's virtualization layer to benefit from all the driver's native functionality and performance within the VM.

Storage Is the Foundation

In any architecture, having a strong foundation is key, and this is especially true in virtualization. Many options are available, allowing virtualization to thrive in remote offices or the highest-performance data center. VSphere 5 raises the bar by providing additional array awareness, ensuring that the hypervisor can communicate with the storage. vSphere can use that information intelligently, moving VMs off a poorly performing volume. With so many options, you might think that the configuration will be very difficult, but vSphere provides the tools to ensure that the configuration is optimal.