EMC²
where information lives®

White Paper

EMC CLARiiON Integration with VMware ESX Server

Applied Technology

Abstract

This white paper provides an overview of how VMware ESX Server integrates with EMC® CLARiiON® storage systems. It introduces the Navisphere® VM-aware feature, a feature that automatically discovers virtual machines managed under VMware vCenter Server and provides end-to-end, virtual-to-physical mapping information. This white paper also discusses the VMotion, VMware HA, and Distributed Resource Scheduling capabilities of VMware ESX Server, as well as clustering of virtual machines when connected to a CLARiiON storage system.

February 2010

Part number H1416.10

Table of Contents

Executive summary

EMC is aggressively expanding product sets from high-end to midtier markets. Through VMware—the industry leader of x86 server-virtualization software—and EMC® CLARiiON®, which offers the best performance in midtier storage, EMC is integrating cutting-edge virtualization technology into its core storage business.

Our latest enhancement, available on the CX4 series, is the Navisphere® *VM-aware* feature. This feature eliminates the painstaking task of manually mapping out the virtual infrastructure, and simplifies common administrative activities such as troubleshooting and capacity planning in virtualized environments. The "Navisphere's new VM-aware feature" section has more information.

Introduction

This white paper outlines the benefits of using VMware virtualization products with the CLARiiON storage system and how to combine features to complement one another. This paper also discusses the connectivity aspect of attaching VMware ESX Server to the CLARiiON storage system.

Audience

This paper is intended for customers, partners, and EMC field personnel requiring information about the features, parameters, and configuration of VMware ESX Server. It includes information about how these features integrate with the CLARiiON storage system. It is assumed that the audience is familiar with CLARiiON hardware and software products, and has a general idea of how VMware ESX Server works.

CLARiiON terminology

CLARiiON LUN — Logical subdivisions of RAID groups in a CLARiiON storage system.

MetaLUNs — These are LUN objects created from multiple CLARiiON LUNs. MetaLUNs provide dynamic LUN expansion and also distribute storage across a very large number of drives.

MirrorView™ — Software designed for disaster recovery solutions by mirroring local production data to a remote disaster recovery site. It offers two complementary remote mirroring products: MirrorView/Synchronous and MirrorView/Asynchronous.

RAID groups — One or more disks grouped together under a unique identifier in a CLARiiON storage system.

SAN Copy™ — Data mobility software that runs on the CLARiiON.

SnapView™ — Software used to create replicas of the source LUN. These point-in-time replicas can be pointer-based snapshots or full binary copies called *clones* or *BCVs*.

Storage Pool — A general term used to describe RAID groups and thin pools. In the Navisphere® Manager GUI, the storage pool node contains RAID groups and thin pool nodes.

Thin LUN — A logical unit of storage where physical space allocated on the storage system may be less than the user capacity seen by the host server.

VMware terminology

Cluster — A cluster is a collection of ESX Server hosts and associated virtual machines that share resources and a management interface.

ESXi — VMware ESXi is a thin, embedded, version of the ESX server that does not have a service console. It moves the server kernels to a dedicated hardware device.

ESX Server — VMware's high-end server product that installs directly on the physical hardware and therefore offers the best performance. ESX Server supports more virtual machines per physical CPU than its other virtualization products such as VMware Server (previously called GSX server).

Farm or Data Center— The primary organizational structure used in VMware vCenter, which contains hosts and virtual machines. The term *Farm* is used with VMware vCenter 1.x while the term *Data Center* is used with vCenter.4.0 or 2.x.

Guest operating system — An operating system that runs on a virtual machine.

Hypervisor - Virtualization software that allows multiple operating systems to run concurrently on a host computer.

ISO image — A CD or DVD image that can be downloaded and burnt on a CD-ROM or DVD-ROM or, mounted as a loopback device.

Management User Interface (MUI) — A web-based graphical interface for VMware that manages a VMware ESX 2.5.x server.

Mapping file — A VMFS file containing metadata used to map and manage a raw device.

Network label — A unique name given to a virtual switch on the ESX server.

Port Groups —Port groups define how a connection is made through the vSwitch to the network. Network services connect to virtual switches through port groups. Usually one or more port group is associated with a single vSwitch.

Raw device mapping (RDM) – Raw device mapping volumes consists of a pointer in a .vmdk file and a physical raw device. The pointer in the .vmdk points to the physical raw device. The .vmdk file resides on a VMFS volume, which must reside on shared storage.

Service console (COS) — The modified Linux kernel that serves as the management interface to the ESX server. Not to be confused with VMkernel.

Templates —A means to import virtual machines and store them as templates that can be deployed at a later time to create new virtual machines.

VMware vCenter — A virtual infrastructure management product that manages and provide valuable services for virtual machines and underlying virtualization platforms from a central, secure location.

VMware vSphere Client — An interface that allows you to connect any Windows PC remotely to a vCenter Server or ESX/ESXi.

Virtual machine — A virtualized x86 PC environment on which a guest operating system and associated application software can run. Multiple virtual machines can operate on the same physical machine concurrently.

Virtual machine configuration file — A file containing a virtual machine configuration that is created by the Configuration Wizard or the Configuration Editor. VMware ESX Server uses this file to identify and run a specific virtual machine. It usually has a .vmx extension.

Virtual switch — A switch that allows virtual network interface cards (NICs) to communicate with one another. Additionally, you can connect one or more physical network adapters to a virtual switch, so that virtual machines can communicate with the outside world.

VMFS — A clustered file system that stores virtual disks and other files that are used by virtual machines.

VMkernel — A kernel that controls the server hardware and schedules virtual machine computations and I/O operations.

ESX overview

ESX consists of virtualization software that provides server consolidation by allowing several instances of similar and dissimilar operating systems to run as virtual machines on one physical machine. This cost-effective, highly scalable virtual machine platform offers advanced resource management capabilities. ESX minimizes the total cost of ownership (TCO) of computing infrastructure by:

- Increasing resource utilization.
- Decreasing the number of servers and all associated costs.
- Maximizing server manageability.

Figure 1 shows the architecture of two ESX servers (ESX 1 and ESX 2) with virtual machines containing guest operating systems that sit on top of the virtualization layer.

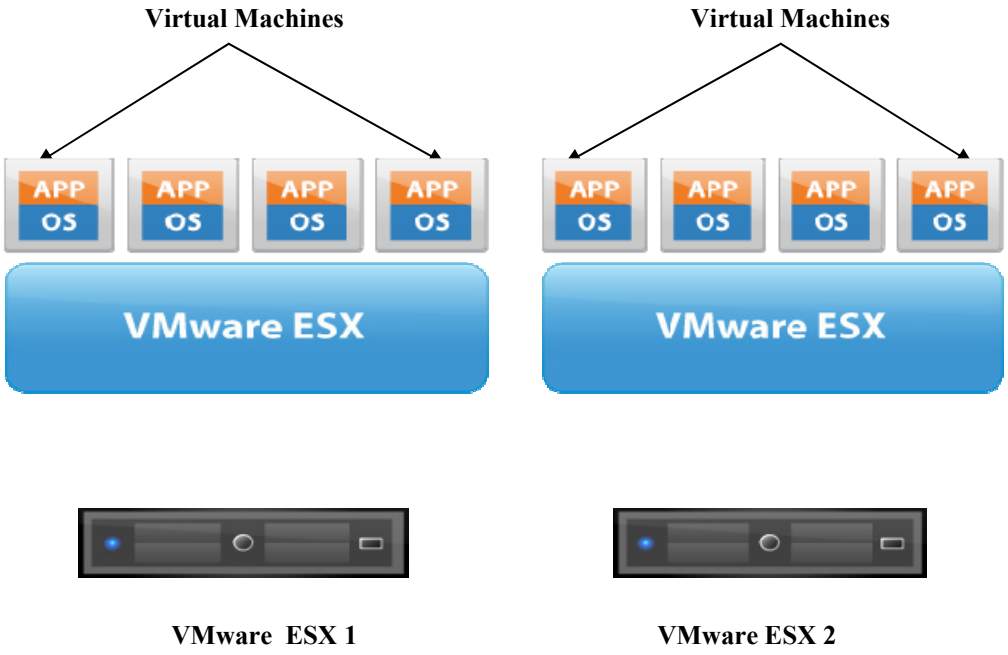


Figure 1. Architecture of VMware ESX Server

ESX Server runs directly on the hardware, and allows virtual machines to run on a hypervisor or on top of the virtualization layer provided by VMware. The combination of an operating system and applications is referred to as a virtual machine. You use the Management User Interface (MUI) to manage ESX 2.x servers. A VMware vCenter server allows you to manage a number of ESX servers and to perform operations such as VMotion. VMware vCenter 1.x is used to manage one or more ESX 2.x servers. No MUI is available to manage ESX 4.0 and 3.x. The VMware vSphere client can only install and manage a single ESX server, while a VMware vCenter server can install and manage one or multiple ESX 4.x and/or ESX 3.x servers.

ESX 4.0 and ESX 4i were introduced in May 2009. VMware ESX 4.0 is similar to the previous versions of ESX and includes a service console (COS) to boot ESX Server. VMware ESXi is a "thin" version that does not include a service console. The management of the VMware ESXi version is done using a vCenter server or a client whose operating system is embedded in the hardware or can be installed on a hard disk. On Windows and Linux platforms, you can also use RemoteCLI to issue commands directly to the VMware ESXi server. For more information on ESX 4.0/ESXi, please see www.vmware.com.

Features

ESX has several features that work with the CLARiiON storage system. The features discussed in this paper are VMotion, Distributed Resource Scheduling and VMware HA, VMware Clustering, and Consolidated Backup technology. The Distributed Resource Scheduling, VMware HA, and Consolidated Backup features are enhanced in ESX 4.0/ESX 4i.

VMware VMotion

ESX Server version 2.0.1 was the first platform to support VMotion. With VMotion, VMware administrators can move virtual machine partitions from machine to machine while real workloads run in the partitions. System administrators can use VMotion, a systems management and provisioning product that works through VMware vCenter, to quickly provision and reprovision servers with any number of virtual machines.

VMotion technology provides the ability to migrate a running virtual machine from one physical ESX server to another—without application service interruption—allowing for fast reconfiguration and optimization of resources without impacting users. With VMotion, VMware allows administrators to move virtual machine partitions from machine to machine on the fly, as real workloads run in the partitions. This allows administrators to do hardware maintenance without interrupting applications and users. It also allows the administrator to do dynamic load balancing to maintain high utilization and performance.

VMware Storage VMotion

This feature, available in ESX 4.0, ESX 3.x, and ESXi, allows you to migrate a virtual machine from one storage system to another while the virtual machine is up and running. For example, using either the VCenter GUI (available in vSphere 4.0) or RemoteCLI interface, virtual machine files can be moved from one FC LUN to another FC LUN without taking the virtual machine offline. Storage VMotion is supported for VMFS volumes and qualified for FC-to-FC storage or FC-to-iSCSI storage and vice versa. For more information on VMFS, see the "LUN partitioning" section.

Distributed Resource Scheduling and VMware High Availability

The VMware Distributed Resource Scheduling (DRS) feature improves resource allocation across all hosts by collecting resource (such as CPU and memory) usage information for all hosts and virtual machines in the cluster and generating recommendations for virtual machine placement. These recommendations can be applied automatically or manually. Depending on the configured DRS automation level, DRS can display or automatically implement recommendations. The result is a self-managing, highly-optimized, highly-efficient computer cluster with built-in resource and load balancing. In ESX 4.0/ESX 3.x/ESXi, VMware's Distributed Power Management reduces power consumption by intelligently balancing a data center's workload. Distributed Power Management, which is part of VMware DRS, automatically powers off servers whose resources are not immediately required and returns power to these servers when they are needed.

VMware High Availability (HA) detects ESX hardware failures and automatically restarts virtual machines and their resident applications and services on alternate ESX hardware, enabling servers to recover more rapidly and deliver a higher level of availability. Using VMware HA and DRS together combines automatic failover with load balancing. This combination results in a fast rebalancing of virtual machines after HA has moved virtual machines to different hosts. In VMware vSphere 4, enhanced HA provides support for monitoring individual virtual machine failures, irrespective of the state of the underlying ESX host. VMware HA can now be configured to restart the failed virtual machine or send a notification to the administrator.

VMware fault tolerance

The fault-tolerance features introduced with ESX 4.0 provide higher availability than VMware HA. The fault-tolerance features allow you to protect any virtual machine from losing data, transactions, or connections when there is a host failure. A duplicate virtual machine called the *secondary* VM of the production (or *primary*) VM is created on a different host. The VMware vLockStep method captures inputs

and events from the primary VM and sends them to the secondary VM, so the VMs are identical. The secondary VM can take over execution from the primary VM, thus providing another level of fault tolerance.

All hosts must have access to the primary VMs datastores and networks through a distributed switch. When a VM is configured to be fault tolerant, the DRS feature is automatically disabled. For more details, please see the *vSphere Availability Guide* available on www.vmware.com.

VMware clustering

The ESX server can be clustered at a virtual machine level within a single ESX server (referred to as an *in-the-box-cluster*) or between two or more ESX servers (referred to as an *outside-the-box-cluster*). The cluster setup within a box is useful for providing high availability when software or administrative errors are the likely causes of failure. Users who want a higher level of protection in the event of hardware failures, as well as software/logical failures, benefit from clustering outside the box.

VMware N_Port ID Virtualization

N_Port ID Virtualization (NPIV) within the Fibre Channel protocol allows multiple virtual N_Port IDs to share a single physical N_Port. In other words, you can define multiple virtual initiators through a single initiator. This feature, available in ESX/ESXi, enables SAN tools that provide QoS at the storage-system level to guarantee service levels for VM applications.

Within VMware ESX, NPIV is enabled for each virtual machine, so that physical HBAs on the ESX server can assign virtual initiators to each virtual machine. As a result, within ESX Server, a virtual machine has virtual initiators (WWNs) available for each HBA. These initiators can log in to the storage like any other host. VMware NPIV support is limited to RDM volumes. For more details about this configuration, please see the "CLARiiON and VMware NPIV support" section.

VMware Site Recovery Manager (SRM)

VMware Site Recovery Manager (SRM) integrates various EMC replication software products (such as MirrorView/S) to automate the failover process for virtual machines. SRM centralizes the creation and management of the disaster recovery strategies that are implemented at the secondary site. SRM uses EMC's array-based snapshot technologies to test the failover process, and to ensure that the recovery image is consistent.

SRM requires that the protected (primary) site and the recovery (secondary) site each have two independent virtual infrastructure servers to facilitate the failover process. Array-based Site Recovery Adapters (SRAs) are also installed at both sites; these SRAs communicate with the storage systems (arrays). For more information about using SRM with CLARiiON storage systems, please see the "CLARiiON and VMware Site Recovery Manager (SRM)" section.

EMC CLARiiON overview

The EMC CLARiiON family of networked storage systems brings best-in-class performance to the mid-tier with a wide range of storage solutions—all based on the powerful, proven, eight generations of CLARiiON architecture. They provide multiple tiers of storage (both Fibre Channel and SATA) in a single storage system, which significantly reduces acquisition costs and management costs by allowing multiple tiers to be managed with a single management interface. The next-generation CLARiiON systems, called the CX4 series with UltraFlex™ technology, deliver storage systems that you can easily customize by populating your I/O slots with either Fibre Channel or iSCSI I/O modules. Products with multiple back ends such as the CX4-240, CX4-480, and CX4-960 can support disks operating at both 2 Gb/s and 4 Gb/s simultaneously.

CLARiiON storage systems address a wide range of storage requirements by providing flexible levels of capacity, functionality, and performance. The AX4-5 is an entry-level system that consists of single-controller and dual-controller models. It supports both Serial Attached SCSI (SAS) and SATA drives and connectivity for up to 64 high availability (HA) connected hosts. The CX4 Model 120 supports up to 120 drives and connectivity for up to 128 HA hosts. The CX4 Model 240 storage system expands the family, supporting up to 256 HA hosts and up to 240 drives. The CX4 Model 480 further expands the CX4 family by supporting 256 HA hosts and 480 drives. The high-end CX4 Model 960 adds even more capability, supporting up to 512 HA hosts and up to 960 drives. Table 1 and Table 2 summarize the basic features for the CLARiiON CX4 and AX4 storage systems.

Table 1. CLARiiON CX4 storage systems feature summary

Feature	CX4-120	CX4-240	CX4-480	CX4-960
Maximum disks	120	240	480	960
Storage processors (SP)	2	2	2	2
Physical memory per SP	3 GB	4 GB	8 GB	16 GB
Max write cache	600 MB	1.264 GB	4.5 GB	10.764 GB
Max initiators per system	256	512	1024	4096
High-availability hosts	128	256	512	2048
Minimum form factor size	6U	6U	6U	9U
Maximum standard LUNs	1024	1024	4096	4096
SnapView snapshots	Yes	Yes	Yes	Yes
SnapView clones	Yes	Yes	Yes	Yes
SAN Copy	Yes	Yes	Yes	Yes
MirrorView/S	Yes	Yes	Yes	Yes
MirrorView/A	Yes	Yes	Yes	Yes

Table 2. CLARiiON AX4-5 storage system feature summary

Feature	AX4-5	AX4-5i
Maximum disks	60	60
Storage processors (SP)	1 or 2	1 or 2
Front-end FC ports/SP	2 @ 4 Gb/s	N/A
Front-end iSCSI ports/SP	N/A	2 @ 1 Gb/s
Back-end FC ports/SP	1 @ 4 Gb/s	1 @ 2 Gb/s
Cache	2 GB	2 GB
High-availability hosts	10/64*	10/64*
Minimum physical size	2U	2U
Maximum standard LUNs	512	512
SnapView snapshots	Yes	Yes
SnapView clones	Yes	Yes
SAN Copy	Yes	N/A

MirrorView/S	Yes	N/A
MirrorView/A	Yes	N/A

*Support for 10 hosts with the base pack and 64 hosts with the expansion enabler

Why use CLARiiON with VMware ESX Server?

CLARiiON and VMware complement each other with the features that they provide. Some of the reasons CLARiiON is an ideal fit for VMware in the midrange storage market include:

- CLARiiON provides a family of storage systems of varying specifications with Fibre Channel and iSCSI connectivity. This allows the user to make the optimal choice of a storage system based on capacity, performance, and cost.
- CLARiiON storage systems can scale quickly to manage anticipated data growth, especially as the storage need for virtual machines increases on the VMware ESX server.
- CLARiiON Virtual (or thin) Provisioning improves storage capacity utilization and simplifies storage management by presenting a virtual machine with sufficient capacity for an extended period of time.
- CLARiiON storage can be shared across multiple ESX servers allowing storage consolidation to provide efficient use of storage resources, which is valuable for clustering and VMotion.
- Software capabilities like VM-aware Navisphere and EMC Storage Viewer give storage and VMware administrators efficient tools to validate changes, plan capacity, and diagnose problems on the ESX and CLARiiON side.
- With EMC Replication Manager support for VMware ESX, customers have a single, easy-to-use interface for provisioning and managing application-consistent replicas running inside a virtual machine that is attached to the CLARiiON storage system.
- Virtual machine applications running on CLARiiON storage systems enhance performance and therefore maximize functionality, reliability, and efficiency of the VMware ESX server as opposed to internal server storage.
- Navisphere Manager suite provides web-based centralized control of global disk space, availability, security, quality-of-service, replication, and reporting for virtual machines provisioned by the CLARiiON storage system.
- The redundant architecture of the CLARiiON storage system provides no single point of failure, thereby reducing application downtime and minimizing business impact for storage upgrades.
- The CLARiiON storage system's modular architecture allows a mixture of EFDs, FC, and SATA drives. EFDs and FC drives can be used for I/O intensive applications, while SATA drives are used for backup and offloading old data, among other things.

CLARiiON configuration with VMware

This section discusses how CLARiiON hardware and software technologies work with VMware ESX Server. It includes topics such as booting from SAN, CLARiiON array-based software implementation, multipathing, and failover software from VMware.

Basic connectivity

Connecting ESX to the CLARiiON storage system requires LUN masking to be enabled (Access Logix™ or Storage Groups) in the **SP Properties** dialog box on the CLARiiON storage system. Access Logix ensures that hosts only have access to "their" LUNs on the CLARiiON. In most customer environments, CLARiiON assigns storage (Fibre Channel or iSCSI) to ESX and not to individual virtual machines. In such a configuration, LUNs presented to the virtual machines are typically transparent to the guest operating system. These LUNs assigned to ESX are not automatically assigned to the virtual machines; the

VMware vCenter server or the Management User Interface assigns a LUN or part of a LUN to the individual virtual machines.

The VMware ESX/ESXi server allows the user to add virtual disks to a virtual machine without powering the virtual machine down. This functionality is provided in the Add dialog box in VMware vCenter 2.0 and later, which is shown in Figure 2.

When connecting a CLARiiON (CX4 or AX4) Fibre Channel storage system to ESX both direct and FC-SW connections are supported. For specific versions of VMware ESX that support direct and FC-SW connect, consult the E-Lab™ Navigator on Powerlink®.

When connecting CLARiiON iSCSI storage to ESX, both the software and hardware initiators are supported. The drivers for the software and hardware initiator are installed by default during the installation of the ESX operating system. The software initiators for network cards and HBA hardware initiators are configured through the VMware vCenter interface. Please see VMware ESX/ESXi documentation for more details.

VMware ESX /ESXi support both Fibre Channel and iSCSI storage. However, VMware and EMC do not support connecting VMware ESX/ESXi servers to CLARiiON Fibre Channel and iSCSI devices on the same array simultaneously.

2 Gb/s, 4 Gb/s, and 8 Gb/s Fibre Channel connections are supported with ESX Server when connected to the CLARiiON CX4 series storage systems.

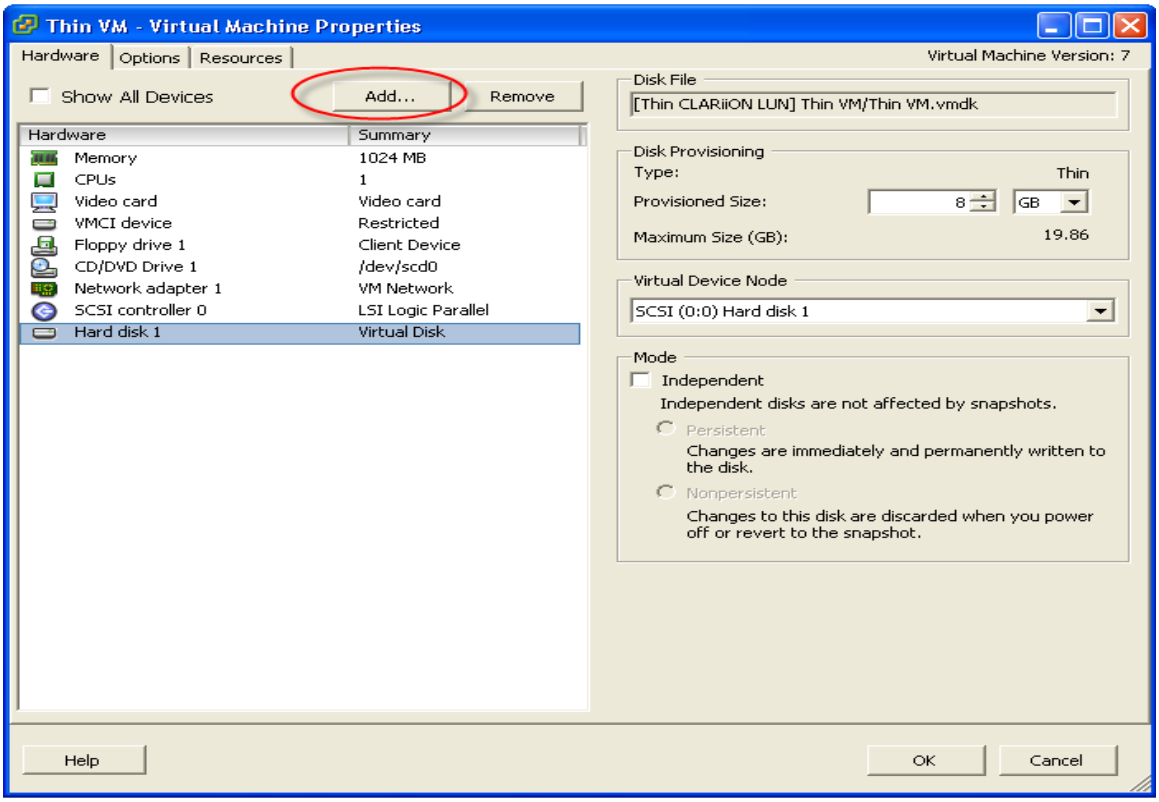


Figure 2. Adding a CLARiiON LUN to a virtual machine (guest operating system) using VMware vCenter for the VMware ESX /ESXi server

Booting from a CLARiiON storage system

This section discusses the procedure for booting the VMware ESX server and guest operating systems—Windows, Linux, NetWare, and Solaris—from CLARiiON LUNs.

Booting ESX 4.0, 3.x, and ESX 2.5.x from CLARiiON LUNs with ESX

In ESX 4.0, 3.x, and 2.5 containing Fibre Channel HBAs zoned to a CLARiiON storage system, the service console can boot from a CLARiiON LUN through the HBAs. To boot an ESX 2.5.x server from a LUN, the HBAs must be shared between the service console and the virtual machines. In ESX 4.0 and 3.x, this is not necessary because there is no concept of shared or dedicated HBAs. ESX 4.0 and 3.x support SAN boot through Fibre Channel and iSCSI HBAs. SAN storage can be connected to VMware ESX 4.0 or 3.x directly or through a switch.

There is no boot-from-SAN support for the VMware ESXi operating system image.

VMware ESX Server accesses LUNs through either a Fibre Channel or iSCSI HBA. If the ESX Server machine has more than one HBA, all its HBAs must be the same model. The HBAs can be single or dual-ported.

To boot through a LUN you must:

- Configure the BIOS settings for the Fibre Channel HBA to select the CLARiiON LUN as the boot device.
- With ESX version 2.5, make sure that the boot LUN is **`/dev/sda`** and **ID 0**—the lowest-numbered LUN visible to the ESX server. This is not necessary for ESX 4.0 and 3.x, since it uses the device UUID for LUN identification.
- The internal SCSI de vice (controller) must be disabled for the CLARiiON LUN to map as **`/dev/sda`**. Since the CLARiiON storage system consists of active/passive path configuration, the lowest-numbered path to the boot LUN must be the active path.
- For ESX 2.5, install the VMware ESX Server software on the CLARiiON LUN using the **boot-from-SAN** option. For ESX 3.x, when installing the server software, select the CLARiiON LUN from which the operating system will boot.

  Note: VMware ESX Server version 2.5 or later is supported for booting ESX over the SAN.

Booting guest operating systems on CLARiiON LUNs

Virtual machines can run on CLARiiON LUNs, as well as on internal disks. Virtual machines can boot using both Fibre Channel and iSCSI CLARiiON storage. Booting virtual machines from shared storage is also a requirement for VMware VMotion.

When booting virtual machines from a CLARiiON storage system, the LUNs are first presented to ESX.

LUNs presented to virtual machines are presented as “virtual” disks; to the virtual machine it appears that it is accessing a physical disk. These disks are created when the virtual machine is created using a VMware vCenter server or the Management User Interface (MUI). The operating system can be installed on these “virtual” disks using a CD or ISO image. When the virtual machine is created, a configuration file with a .vmx extension is also generated. This file contains the location of virtual disks, memory size, and some basic hardware setup information (CD-ROM drive, floppy drive, network connections) for the virtual machine.

Once the virtual machine is up and running, it is highly recommended that VMware Tools be installed on each virtual machine. VMware Tools will optimize the use of the VMware ESX Server resources. VMware Tools also provide a VGA device driver and a heartbeat mechanism for the virtual machine to communicate with the VMkernel.

The procedure for connecting VMware ESX Server to the EMC CLARiiON storage system is found in the *Host Connectivity Guide for VMware ESX Server*.

VMware does not support booting ESX Server over iSCSI storage using the software initiator; however, it does support booting VMs residing on iSCSI LUNs, which is a requirement for VMotion.

In addition, a virtual machine can install an iSCSI software initiator and connect directly to iSCSI ports on the CLARiiON storage system. Thus, a VMware ESX server can be connected (via Fibre Channel or iSCSI) to a CLARiiON, and a VM on the ESX server can also be connected (via iSCSI) to the same CLARiiON.

With the FLARE® 29 release, the number of initiators that login to the CLARiiON has been increased; this allows more virtual machines to directly connect to CLARiiON systems using iSCSI within the VM or NPIV technology (discussed later), thus providing scalability improvements.

Navisphere management

Navisphere Agent (for CX, CX3, CX4) arrays) or the Server Utility should be installed on the ESX service console to register ESX 3.x servers with the CLARiiON storage system. The VMware Navisphere Agent installed on the ESX provides device mappings information and allows path registration with the storage system. It does not provide the device mapping information from the virtual machines since the agent is installed on the ESX.

Navisphere Agent and Server Utility software packages are not supported on ESX 4.0, instead the CLARiiON storage system initiator records are automatically registered when ESX reboots or when a rescan of the ESX 4.0 server occurs. The same thing happens on ESXi, which does not have a service console to install or run the host agent or server utility. For this reason, manual registration is not necessary. It is important to make sure that the ESX host is properly configured with an IP address and a hostname to ensure proper registration with the CLARiiON storage system. If you have multiple service console NICs configured, ensure they have a valid IP address. Check the /etc/hosts file on the ESX server to see if the NICs are properly configured and do not have any 127.0.0.1 entries.

Navisphere CLI and array initialization software for the CX and AX4 series storage systems can run on the ESX Server console for ESX 4.0 and ESX 3.x servers, as well as the individual virtual machines.

For Navisphere Agent/CLI to work with ESX 3.x and for Navisphere CLI to work on ESX 4.0 when connected to a CLARiiON storage system, the ports for agent and/or CLI need to be opened. This can be done by executing the following command on the ESX service console:

```
# esxcfg-firewall –o --openPort <port,tcp|udp,in|out,name>
```

For example:

```
esxcfg-firewall –o 6389,tcp,in,naviagent
```

There is also a shell script that automatically opens the firewall ports for Navisphere Agent. For detailed information on which ports to open, see the *CLARiiON Server Support Products for Linux and VMware ESX Server Installation Guide* available on Powerlink®, EMC's password-protected extranet for customers and partners. When Navisphere Agent is installed on ESX 3.x servers, rescan the VMware ESX server and then restart the agent so that it communicates with the storage system and sends updated information.

When Navisphere CLI is installed on virtual machines configured on ESX 3.x, some commands (for example, **lunmapinfo or volmap**) that require Navisphere Agent must be directed to the ESX service console and not to the virtual machines. Check the Navisphere Agent/CLI release notes on Linux and VMware for more details. Figure 3 shows the device mapping information that is listed when the **lunmapinfo** command is issued from Navisphere CLI on a Windows virtual machine. This command is directed to the agent residing on the ESX service console. For ESX 4.0 or ESXi servers, the **volmap** command is used to get device information using the following command.

```
# naviseccli –h 10.14.15.16 server –volmap –host 10.14.15.140
```

```
C:\Program Files\EMC\Navisphere CLI>navicli -h 10.14.17.73 lunmapinfo
Logical Drives:             vmhba0:1:0, /boot, /
Physical Device:            sda

Logical Drives:             vmhba1:0:0
Physical Device:            sdc

Logical Drives:             vmhba1:0:1
Physical Device:            sdd

Logical Drives:             vmhba1:0:2
Physical Device:            sde

No storage systems were found.  Certain fields could not be displayed.
```

Figure 3. Executing the lunmapinfo command issued on a Windows virtual machine and sent to the Navisphere Agent on the ESX service console

You can use the symm **inq** utility to get device mapping information from the virtual machine to the CLARiiON LUN level. The virtual disks assigned to the virtual machine must be configured as raw mapped LUN for this to work correctly. Figure 4 shows output for the inq –clar_wwn command.

```
C:\>
C:\>inq -clar_wwn
Inquiry utility, Version V7.3-623 (Rev 0.0)      (SIL Version V6.0.0.0 (Edit Lev
el 623)
Copyright (C) by EMC Corporation, all rights reserved.
For help type inq -h.

...

------------------------------------------------------------------------------
-----------------
CLARiiON Device     Array Serial #   SP IP Address      LUN      WWN (all 32 hex
 digits required)
------------------------------------------------------------------------------
-----------------
\\.\PHYSICALDRIVE0  WRE00022100934   A  10.14.17.78     0x0006   60060160b7a5080
03ec2197bffc2d911
\\.\PHYSICALDRIVE1  WRE00022100934   B  10.14.17.79     0x000a   60060160b7a5080
07df535167eedd911
\\.\PHYSICALDRIVE2  WRE00022100934   A  10.14.17.78     0x0003   60060160b7a5080
03bc2197bffc2d911
```

Figure 4. Output for the inq –clar_wwn command that provides device mapping information from the virtual machine level to the CLARiiON LUN level

Navisphere Server Utility software can determine the ESX server configuration and check to see if the VMware ESX configuration is a high-availability environment. Support for VMware ESX with the Navisphere Server Utility is available with FLARE 28.

The Navisphere Server Utility must be installed on a Windows server to communicate with the VMware ESX 4.0, ESX 3.x, and ESXi or VMware vCenter server. Figure 5 shows how to enter the credentials for VMware vCenter using the Navisphere Server Utility. You can now view the report generated by the server utility for a particular ESX server.

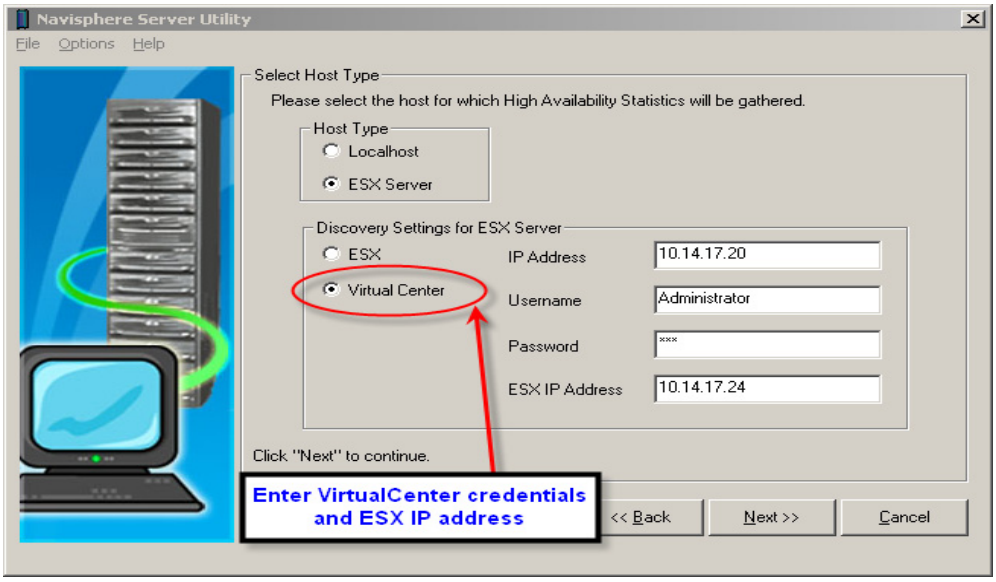
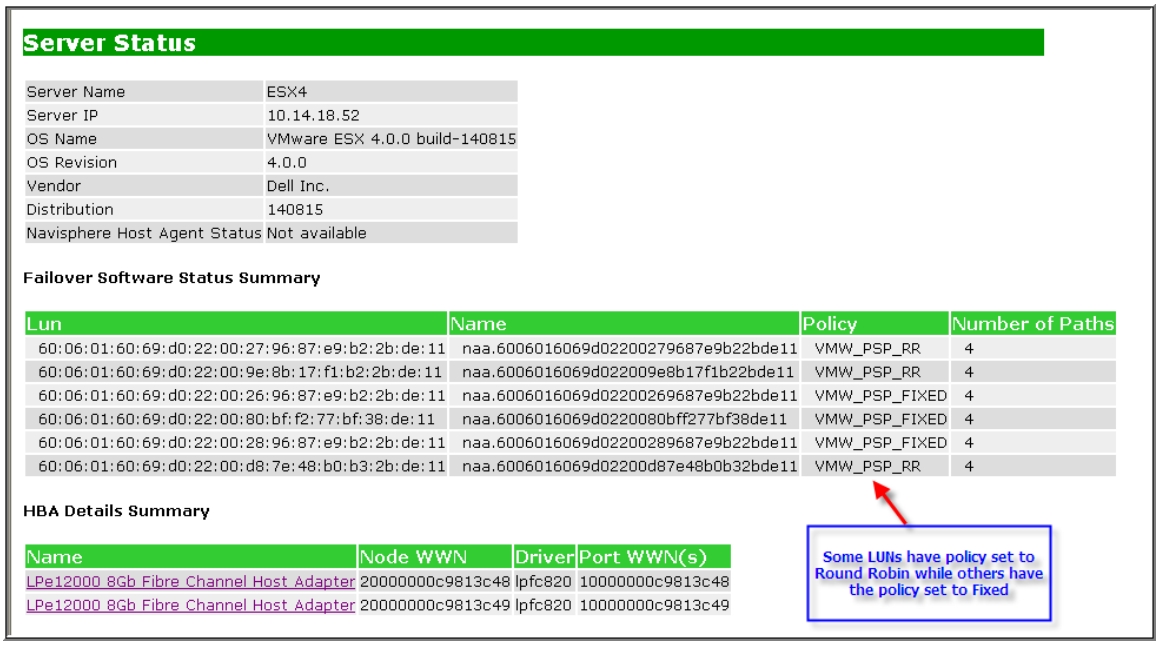


Figure 5. Using the Navisphere Server Utility with a VMware ESX 3.x/3i server

Figure 6, Figure 7, and Figure 8 are examples for reports generated by the server utility for ESX 4.0 and 3.5. In Figure 7, the server utility reports the policies configured for the LUNs.



Server Status

Server Name	ESX4
Server IP	10.14.18.52
OS Name	VMware ESX 4.0.0 build-140815
OS Revision	4.0.0
Vendor	Dell Inc.
Distribution	140815
Navisphere Host Agent Status	Not available

Failover Software Status Summary

Lun	Name	Policy	Number of Paths
60:06:01:60:69:d0:22:00:27:96:87:e9:b2:2b:de:11	naa.6006016069d02200279687e9b22bde11	VMW_PSP_RR	4
60:06:01:60:69:d0:22:00:9e:8b:17:f1:b2:2b:de:11	naa.6006016069d022009e8b17f1b22bde11	VMW_PSP_RR	4
60:06:01:60:69:d0:22:00:26:96:87:e9:b2:2b:de:11	naa.6006016069d02200269687e9b22bde11	VMW_PSP_FIXED	4
60:06:01:60:69:d0:22:00:80:bf:f2:77:bf:38:de:11	naa.6006016069d0220080bff277bf38de11	VMW_PSP_FIXED	4
60:06:01:60:69:d0:22:00:28:96:87:e9:b2:2b:de:11	naa.6006016069d02200289687e9b22bde11	VMW_PSP_FIXED	4
60:06:01:60:69:d0:22:00:d8:7e:48:b0:b3:2b:de:11	naa.6006016069d02200d87e48b0b32bde11	VMW_PSP_RR	4

HBA Details Summary

Name	Node WWN	Driver	Port WWN(s)
LPe12000 8Gb Fibre Channel Host Adapter	20000000c9813c48	lpfc820	10000000c9813c48
LPe12000 8Gb Fibre Channel Host Adapter	20000000c9813c49	lpfc820	10000000c9813c49

Figure 6. Report generated by the Navisphere Server Utility showing configuration information for ESX 4.0 when using VMware's native multipathing

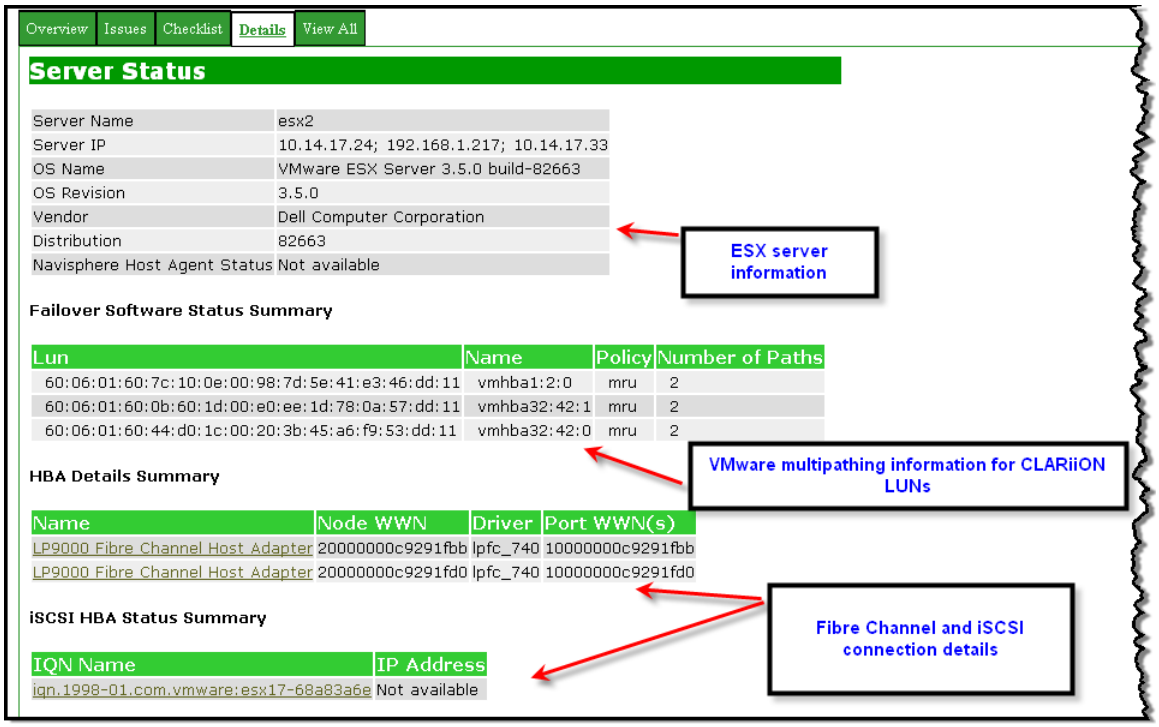


Figure 7. Report generated by the Navisphere Server Utility showing configuration information for ESX 3.x

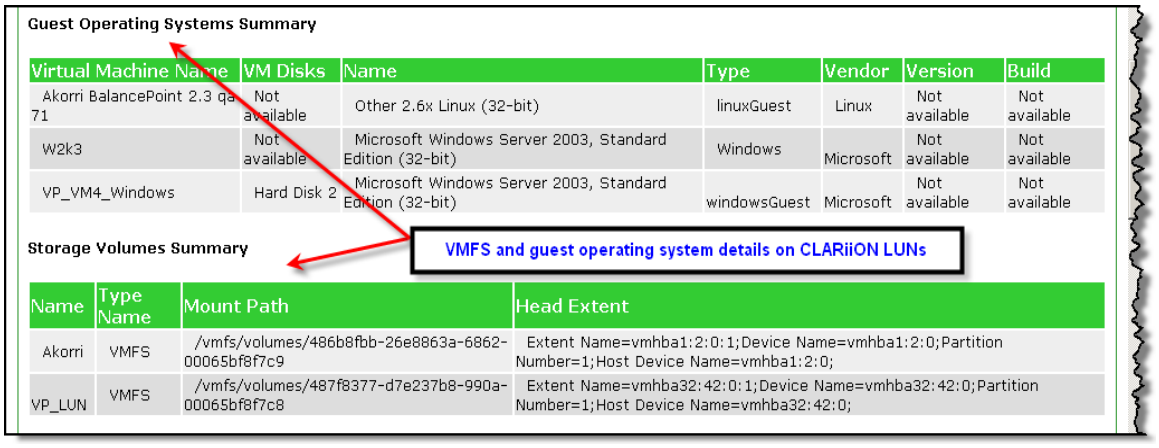


Figure 8. Report generated by the Navisphere Server Utility showing guest OS and VMFS volume information

Multipathing and failover with ESX on CLARiiON

Multipathing and load balancing increase the level of availability for applications running on ESX servers. CLARiiON storage systems also support the nondisruptive upgrade (NDU) operation for VMware's native failover software and EMC PowerPath®. E-Lab Navigator has a list of ESX Server versions for which NDU operations are supported.

We recommend that you disable the auto-assign parameter on the CLARiiON LUN. You should only enable auto assign if the host does *not* use failover software. In this situation, the failover software (instead

of auto assign) controls ownership of the LUN in a storage system with two SPs. For more information about the auto-assign LUN, please see Knowledgebase case emc165941.

VMware native multipathing and failover on ESX 4.0 with CLARiiON

VMware ESX 4.0 contains its own native multipathing software that is built into its kernel. This failover software, called Native Multipathing Plugin (NMP), has three policies:

- FIXED policy
- Round Robin policy
- Most Recently Used (MRU) policy

On VMware 4.0 servers with CX4 arrays, the FIXED or Round Robin policy is supported.
The FIXED policy on the CX4 provides failback capability. To use the FIXED policy, you must be running FLARE release 28 version 04.28.000.5.704 or later. Also, the failovermode mode must be set to 4 (ALUA mode or Asymmetric Active/Active mode). The default failovermode for ESX 4.0 is 1. Use the Failover Setup Wizard within Navisphere to change the failovermode from 1 to 4.

When using the FIXED policy, the auto-restore or failback capability distributes the LUNs to their respective storage processors (SPs) after an NDU operation. This prevents the LUNs from all being on a single storage processor after an NDU. When using the FIXED policy, ensure the preferred path setting is configured to be on the same storage processor for all ESX hosts accessing a given LUN.

For more details on the benefits of using the Asymmetric Active/Active mode with CLARiiON storage systems, please see *EMC CLARiiON Asymmetric Active/Active Feature (ALUA)* available on Powerlink.

With the FIXED policy, there is some initial setup that is required to select the preferred path; the preferred path should also be the optimal path when using the ALUA mode. If set up properly, there should not be any performance impact when using failovermode 4 (ALUA). Note that FIXED sends I/O down only a single path. However, if you have multiple LUNs in your environment, you could very well choose a preferred path for a given LUN that is different for other LUNs and achieve static I/O load balancing. FIXED performs an automatic restore, hence LUNs won't end up on a single SP after an NDU.

When using Round Robin there is no auto-restore functionality, hence after an NDU all LUNs will end up on a single SP. A user would need to manually trespass some LUNs to the other SP in order to balance the load. The benefit of Round Robin is that not too many manual setups are necessary during intial setup; by default it uses the optimal path and does primitive load balancing (however, it still sends I/O down only a single path at a time). If multiple LUNs are used in the environment, you might see some performance boost. If you had a script that takes care of the manual trespass issue, then Round Robin would be the way to avoid manual configuration.

On a CX3 or earlier CLARiiON storage systems, the Most Recently Used (MRU) or Round Robin policy must be used with failovermode=1.
Note that the Most Recently Used (MRU) and Round Robin policies do *not* provide failback capability. Furthermore, the Round Robin policy does not provide true dynamic load balancing; it sends I/O down one chosen path at a time alternating through paths on the owning SP. The path selected for I/O is controlled by the Round Robin algorithm. The FIXED and MRU policies also send I/O down only a single selected path for a given LUN, unless that path becomes unavailable.

Figure 9 shows how the FIXED policy is configured with the CX4 storage system using VMware's NMP software.

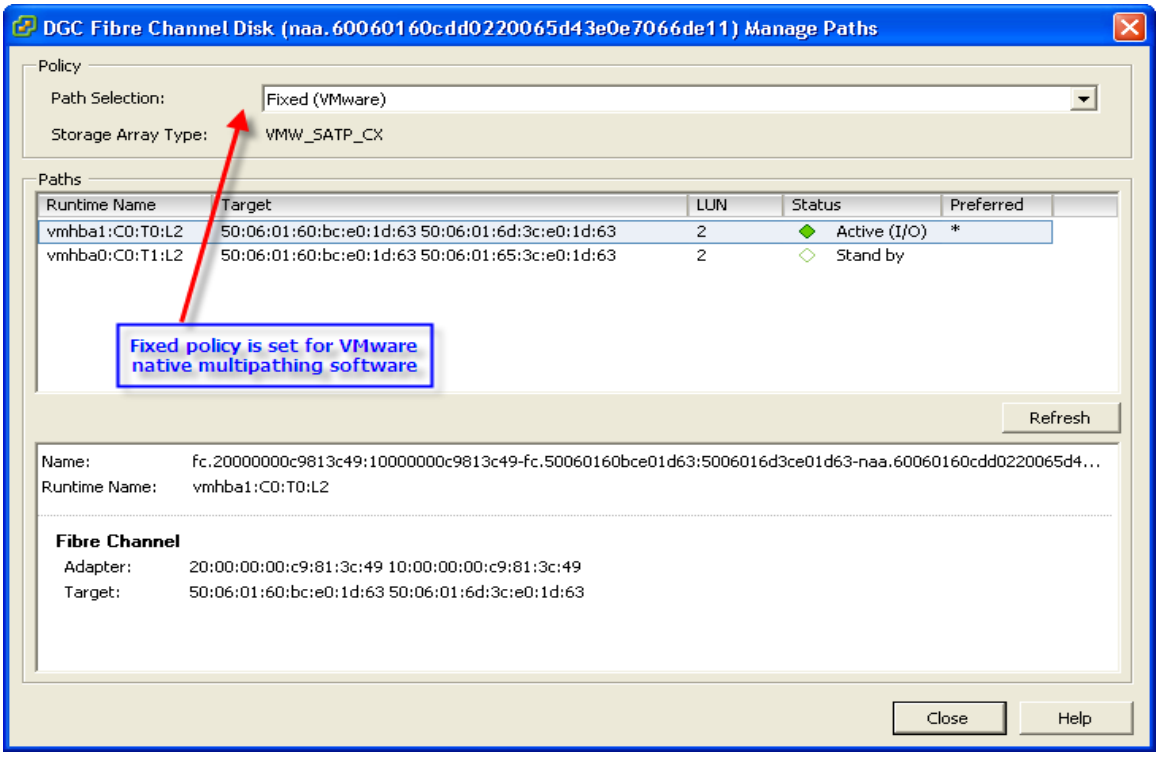


Figure 9. VMware's native multipathing software on ESX Server 4.0 configured with the FIXED policy setting for CLARiiON storage systems

EMC PowerPath multipathing and failover on ESX 4.0 with CLARiiON

EMC PowerPath software is supported on the ESX 4.0 server and is installed using RemoteCLI. RemoteCLI is a software package available for remotely managing the ESX server. PowerPath can co-exist with VMware's native failover such that some LUNs can be controlled by PowerPath on one array while some LUNs from a different array are under the control of VMware's NMP software. PowerPath is supported in FC and iSCSI (software and hardware initiators) configurations. Some of the benefits of using PowerPath with ESX 4.0 are as follows:

- PowerPath on ESX 4.0 is supported with all CLARiiON CX-series arrays configured with failovermode=4 (ALUA mode or Asymmetric Active/Active mode).
- PowerPath has an intuitive CLI that provides an end-to-end view and reporting of the host storage resources including HBAs all way to the storage system.
- PowerPath eliminates the need to manually change the load-balancing policy on a per-device basis.
- PowerPath's auto-restore capability automatically restores LUNs to default SPs when an SP recovers, ensuring balanced load and performance.

Figure 10 depicts the CLARiiON LUNs controlled by EMC PowerPath software.

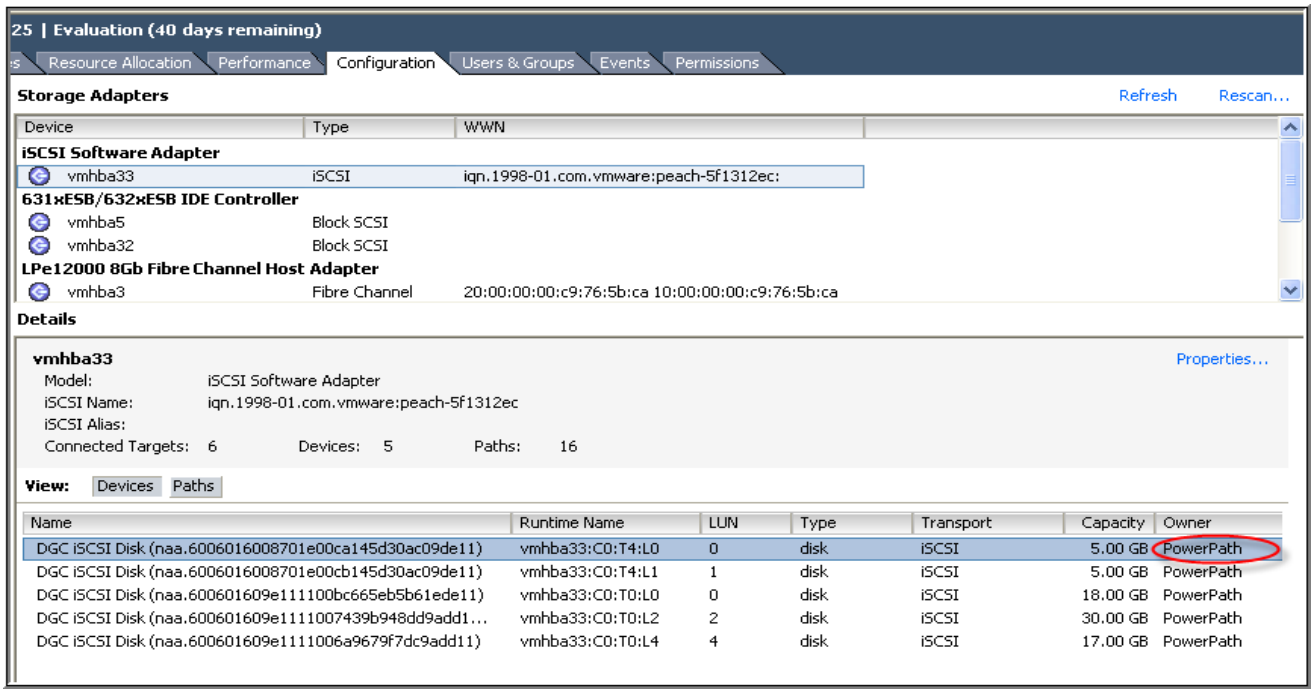


Figure 10. EMC PowerPath software configured on ESX 4.0 connected to a CLARiiON storage system

iSCSI configurations and multipathing with ESX 4.0

Figure 11 shows how to configure the iSCSI software on a CLARiiON storage system. Note that the iSCSI hardware-initiator configuration is similar to the Fibre Channel HBA configuration, and is not covered in this section.

Two virtual switches (vSwitches), each containing one or more NICs, can be configured on ESX 4.0 as shown in Figure 12. The two NICs or vmkernel ports should be on different subnets. The SP ports should be split across subnets in order to spread the network load across NICS, subnets, and SP ports.

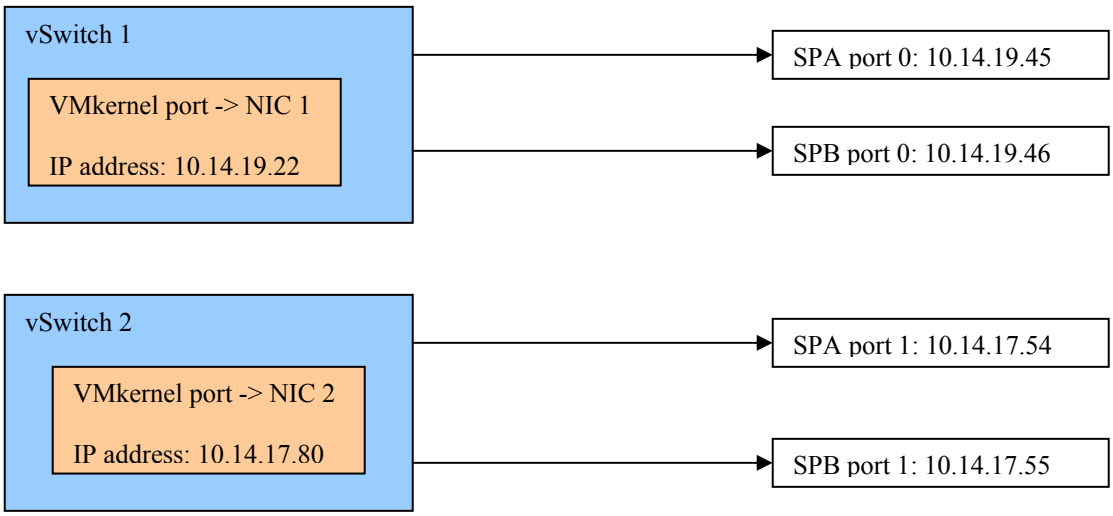


Figure 11. Dual virtual switch iSCSI configuration

With **port binding** enabled, a single vSwitch with two NICs can be configured so that each NIC is bound to one vmkernel port. The two NICs, or vmkernel ports, should be on different subnets. Also, the two SPs for a storage processor should be on different subnets, as shown in Figure 12. No gateways should be configured on the NICs or SP ports.

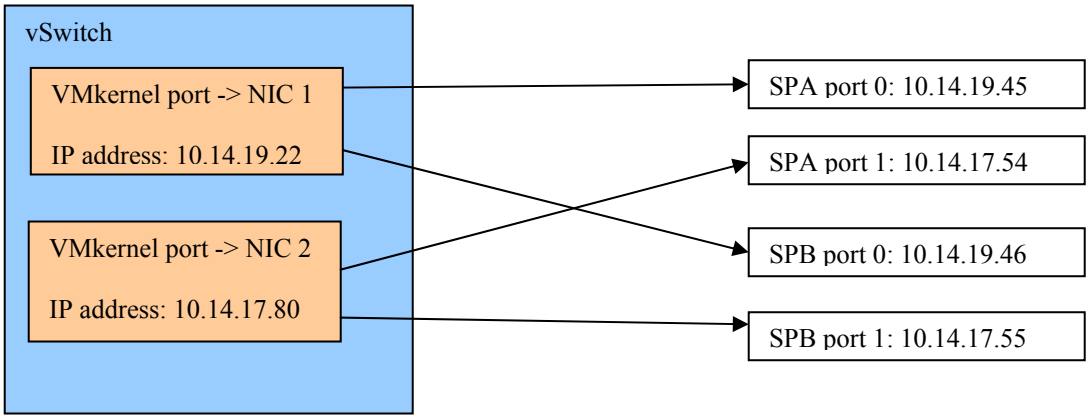


Figure 12. Single vSwitch iSCSI configuration

Multipathing and failover on ESX 3.x and ESX 2.x with CLARiiON

The CLARiiON storage system supports VMware ESX Server's built-in failover mechanism; this mechanism is available for ESX 3.x and 2.x servers, and provides failover but not active I/O-load balancing.

PowerPath multipathing and failover software is not supported on VMware ESX 3.x and 2.x servers.

The native failover software provides a listing of the paths—whether active or passive—from VMware ESX Server to the CLARiiON storage system. The **exscfg-mpath** command in ESX 3.x provides details on all devices (Fibre Channel, iSCSI, and local) and the number of paths attached to that device. With

VMware ESXi, you can use VMware vCenter or the RemoteCLI package to see the number of paths to a given LUN. If you are using an ESX version with a service console, type:

esxcfg-mpath -l

```
[root@esx3 /]# esxcfg-mpath -l
Disk vmhba0:0:0 /dev/sda (2048MB) has 2 paths and policy of Most Recently Used
 FC 1:8.0 10000000c92930e7<->500601601060174b vmhba0:0:0 On active preferred
 FC 1:8.0 10000000c92930e7<->500601681060174b vmhba0:1:0 Standby

Disk vmhba0:0:1 /dev/sdb (5120MB) has 2 paths and policy of Most Recently Used
 FC 1:8.0 10000000c92930e7<->500601601060174b vmhba0:0:1 Standby  preferred
 FC 1:8.0 10000000c92930e7<->500601681060174b vmhba0:1:1 On active

Disk vmhba0:0:3 /dev/sdc (5120MB) has 2 paths and policy of Most Recently Used
 FC 1:8.0 10000000c92930e7<->500601601060174b vmhba0:0:3 Standby  preferred
 FC 1:8.0 10000000c92930e7<->500601681060174b vmhba0:1:3 On active

Disk vmhba0:0:5 /dev/sdd (1024MB) has 2 paths and policy of Most Recently Used
 FC 1:8.0 10000000c92930e7<->500601601060174b vmhba0:0:5 On active preferred
 FC 1:8.0 10000000c92930e7<->500601681060174b vmhba0:1:5 Standby

Disk vmhba0:2:0  (0MB) has 2 paths and policy of Most Recently Used
 FC 1:8.0 10000000c92930e7<->5006016010208b46 vmhba0:2:0 On active preferred
 FC 1:8.0 10000000c92930e7<->5006016810208b46 vmhba0:3:0 On

Disk vmhba2:0:0 /dev/sdq (17366MB) has 1 paths and policy of Fixed
 Local 5:6.0 vmhba2:0:0 On active preferred

Processor Device vmhba2:6:0  (0MB) has 1 paths and policy of Fixed
 Local 5:6.0 vmhba2:6:0 On active preferred
```

Figure 13. VMware ESX Server 3.0 path information for Fibre Channel devices

Figure 13 shows the seven devices attached to the CLARiiON storage system. The **vmhba0:x:x** devices are Fibre Channel devices. All Fibre Channel devices have paths to both SP A and SP B. The **active** mode for each path shows the path the ESX server uses to access the disk. The **preferred** mode, although it is displayed, is not honored (it is ignored) since the policy is set to **Most Recently Used** (MRU). Device **vmhba2:0:0** is the internal boot device and has a single path.

Figure 14 shows the three devices attached to the CLARiiON storage system. The **vmhba40:0:x** devices are iSCSI devices. All iSCSI devices have paths going to both SP A and SP B. If using VMware NIC teaming, the NICs must be on the same subnet and use the same IP address for failover to work between multiple iSCSI NICs (uplink adapters). As a best practice, create dedicated virtual switches for iSCSI traffic.

With ESX 3.5/ESXi, VMware supports the configuration of two virtual switches on separate subnets that go to different network switches if you use the iSCSI software initiator that is built in to VMware ESX Server.

```
Disk vmhba2:0:0 /dev/sdq (17366MB) has 1 paths and policy of Fixed
 Local 5:6.0 vmhba2:0:0 On active preferred

Processor Device vmhba2:6:0  (0MB) has 1 paths and policy of Fixed
 Local 5:6.0 vmhba2:6:0 On active preferred

Disk vmhba40:0:0 /dev/sda (10240MB) has 2 paths and policy of Most Recently Used
 iScsi sw iqn.1998-01.com.vmware:esx3-1eedf183<->iqn.1992-04.com.emc:cx.apm00042102262.a0 vmhba40:0:0 Standby  preferred
 iScsi sw iqn.1998-01.com.vmware:esx3-1eedf183<->iqn.1992-04.com.emc:cx.apm00042102262.b0 vmhba40:1:0 On active

Disk vmhba40:0:1 /dev/sdb (10240MB) has 2 paths and policy of Most Recently Used
 iScsi sw iqn.1998-01.com.vmware:esx3-1eedf183<->iqn.1992-04.com.emc:cx.apm00042102262.a0 vmhba40:0:1 Standby  preferred
 iScsi sw iqn.1998-01.com.vmware:esx3-1eedf183<->iqn.1992-04.com.emc:cx.apm00042102262.b0 vmhba40:1:1 On active

Disk vmhba40:0:2 /dev/sdc (10240MB) has 2 paths and policy of Most Recently Used
 iScsi sw iqn.1998-01.com.vmware:esx3-1eedf183<->iqn.1992-04.com.emc:cx.apm00042102262.a0 vmhba40:0:2 Standby  preferred
 iScsi sw iqn.1998-01.com.vmware:esx3-1eedf183<->iqn.1992-04.com.emc:cx.apm00042102262.b0 vmhba40:1:2 On active

Disk vmhba40:0:3 /dev/sdd (14336MB) has 2 paths and policy of Most Recently Used
 iScsi sw iqn.1998-01.com.vmware:esx3-1eedf183<->iqn.1992-04.com.emc:cx.apm00042102262.a0 vmhba40:0:3 Standby  preferred
 iScsi sw iqn.1998-01.com.vmware:esx3-1eedf183<->iqn.1992-04.com.emc:cx.apm00042102262.b0 vmhba40:1:3 On active
```

Figure 14. VMware ESX Server 3.0 path information for iSCSI devices

The **vmkmultipath** command, when issued on an ESX 2.x server, provides details about the devices and the number of paths attached to each device. At the service console of the VMware ESX server, type:

```
# vmkmultipath -q
```

```
[root@Vmware2 root]# vmkmultipath -q
Disk and multipath information follows:

Disk vmhba0:0:0 (10,236 MB) has 2 paths.  Policy is mru.
      vmhba0:0:0       on  (active, preferred)
      vmhba1:0:0       on

Disk vmhba0:1:1 (40,954 MB) has 2 paths.  Policy is mru.
      vmhba0:1:1       on  (active, preferred)
      vmhba1:1:1       on

Disk vmhba2:0:0 (34,726 MB) has only 1 path.
```

Figure 15. VMware ESX Server 2.x path information through the native failover software

The most recently used MRU policy is the default policy for active/passive storage devices in ESX 2.x and 3.0. The policy for the path should be set to MRU for CLARiiON storage systems to avoid path thrashing. When using the MRU policy, there is no concept of preferred path; in this case, the preferred path can be disregarded. The MRU policy uses the most recent path to the disk until this path becomes unavailable. As a result, ESX Server does not automatically revert to the original path until a manual restore is executed.

If you connect two ESX servers with path one from HBA1 to SPA, and path two from HBA0 to SPB, a single LUN configured as a VMFS volume can be accessed by multiple ESX servers; in this example a LUN can be accessed by both ESX servers.

If the HBA1-SPA path on ESX1 fails, it issues a trespass command to the array, and SPB takes ownership of the LUN. If the path from HBA1-SPB on ESX2 then fails, the LUN will trespass back and forth between the SPs, which could result in performance degradation.

When the CLARiiON LUN policy is set to MRU and an ESX server with two HBAs is configured so that each HBA has a path to both storage processors, VMware ESX Server accesses all LUNs through one HBA and does not use the second HBA. You can edit the path configuration settings so the other HBA is the

active path for some LUNs; however, this configuration is not persistent across reboots. After a reboot, the LUNs will be on a single HBA. The advantage of this configuration is it prevents unnecessary trespasses of LUNs in the case of failure.

The failover time can be adjusted at the HBA, ESX, and virtual machine levels. The *Fibre Channel SAN Configuration Guide and iSCSI SAN Configuration Guide* found on www.vmware.com provide recommendations for setting the failover time at the HBA and virtual machine level.

VMware ESX Server periodically evaluates the state of each path. The default evaluation period is 300 seconds. This can be changed by modifying the **`/proc/vmware/config/disk/PathEvalTime`** **`vmkernel`** **config** value. This can also be done thru the MUI for ESX 2.x by going to **Advanced Setting** and changing the **Disk.PathEvalTime** parameter. The evaluation period can be set to any value between 30 and 1500 seconds. Note that reducing the **PathEvalTime** causes path evaluation to run more frequently. This puts a slightly higher CPU load on the system. Reducing this value (to 90 for example) will help improve failover time (keeping in mind the preceding caveat).

Table 3 shows the failovermode policies supported for ESX 3.5 and ESX 4.0 with CLARiiON storage systems.

Table 3. Failovermode policies

ESX version	ALUA PowerPath (Failovermode = 4)	ALUA Native (Failovermode =4)	PNR mode PowerPath (Failovermode =1)	PNR mode Native (Failovermode =1)
ESX 4.0	Yes (CX arrays running FLARE 26 or later)	Yes (Fixed or Round Robin) (CX4 arrays running FLARE 28 version 04.28.000.5.704 or later)	Yes (CX arrays running FLARE 22 or later)	Yes (Round Robin or MRU) (CX arrays running FLARE 22 or later)
ESX 3.5	No	No	No	MRU (CX arrays running FLARE 22 or later)

LUN partitioning

LUNs presented to the ESX server are ultimately presented to the virtual machines. Any storage device presented to any virtual machine is represented as a virtual disk. To the virtual machine, the virtual disk appears to be a physical disk. A virtual machine can have multiple virtual disks of different/multiple virtual disk types located on multiple SCSI controllers. A CLARiiON LUN presented to the ESX server can be partitioned using one of the three methods:

- Raw disks
- VMFS volumes
- Raw device mapping

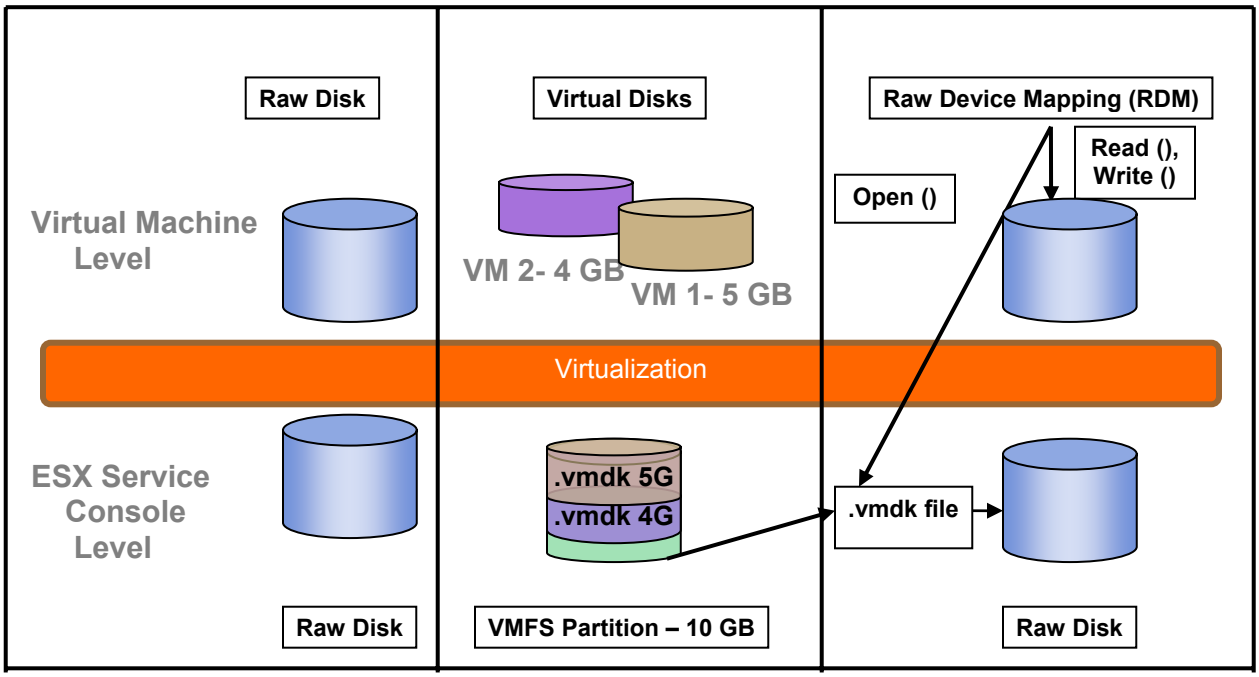


Figure 16. Partitioning a CLARiiON LUN

Raw disks

For raw disks, an entire CLARiiON LUN is presented to a single virtual machine without being partitioned at the ESX service console level. When a virtual machine is configured to use a raw disk, VMware directly accesses the local disk/partition as a raw device. Raw devices are available in ESX 2.5, but it is recommended that the LUN be configured as a raw device mapping device (RDM) instead. RDMs are very similar to raw disks except that RDMs are compatible with VMotion.

VMFS volumes

When a CLARiiON LUN is configured as a VMFS volume, this volume can be partitioned and presented to a number of virtual machines. For example, if you present a 10 GB CLARiiON LUN to your ESX server, a VMFS file system can be created on that LUN. New VMFS-3 volumes created with 3.5/ESXi must be 1,200 MB or larger. For previous versions of ESX Server, the VMFS-3 requirement was 600 MB. The user has the option of presenting this entire VMFS volume to an individual virtual machine or presenting portions of this volume to a number of virtual machines. In Figure 16, the VMFS volume is used to create two virtual disks (.vmdk files)—one is 5 GB and the other is 4 GB. Each of these virtual disks is presented to a different virtual machine. It is also possible to create a virtual disk on an entire VMFS volume and assign this virtual disk to a single virtual machine.

In ESX 4.0/3.x/ESXi, the swap files, NVRAM files, and configuration (.vmx) files for a virtual machine reside on a VMFS-3 volume. On ESX 2.0, these files reside on an ext3 file system on the service console.

ESX 2.x supports an undoable disk mode that allows you to keep or discard changes to a virtual disk using snapshot technology. Snapshot technology on the ESX server is supported for VMFS-3 and VMFS-2 volumes. In ESX 4.0/3.x/ESXi, the snapshot technology allows all virtual disks within a VM configured as VMFS-3 volumes to be snapshot together along with VM memory, processor, and other states using the consolidated backup solution.

Raw device mapping (RDM)

VMware ESX 2.5 introduced a new technology called raw device mapping (RDM); this is also called a *mapped raw LUN* when assigned to a virtual machine. This technology has a **SCSI** pass-through mode that allows virtual machines to pass SCSI commands directly to the physical hardware. Utilities like `admsnap` and `admhost`, when installed on virtual machines, can directly access the virtual disk when the virtual disk is in physical compatibility mode. In virtual compatibility mode, a raw device mapping volume looks like a virtual disk in a VMFS volume. This streamlines the development process by providing advance file locking data protection and VMware snapshots. In RDM virtual compatibility mode certain advanced storage-based technologies, such as expanding an RDM volume at the virtual machine level using metaLUNs, do not work.

Using a raw CLARiiON LUN, a user can create a raw device mapping volume by creating a mapping file on a VMFS volume. This mapping file, which contains a `.vmdk` extension, points to the raw device, as shown in Figure 16. The mapping file is created when the raw device is ready to be assigned to a virtual machine. The entire CLARiiON LUN is presented to an individual virtual machine. The virtual machine opens the mapping file information from the VMFS volume and can directly access the raw device mappings volume for reading and writing.

For more information on configuring VMFS and raw device mapping volumes, refer to the *ESX 4.0 Basic System Administration* guide.

LUN layout recommendations

This section discusses some of the best practices for optimal capacity when designing and implementing the LUN layout for VMware ESX servers connected to CLARiiON storage systems.

OS images and application data images of virtual machines can reside on CLARiiON LUNs. Since VMFS is a clustered file system, when LUNs are configured as VMFS volumes, many ESX servers can share different virtual disks on the same LUN (VMFS) volume. Hence, the number of virtual machines images installed on that particular LUN, and the workload on those virtual machines and the ESX servers that are accessing the LUN, will dictate the number of spindles that need to be assigned to that particular LUN (VMFS volume).

We recommend that you use striped metaLUNs to distribute the load across different RAID groups when booting a number of OS images on a given LUN (VMFS volume), since most users assign larger LUNs to their VMware ESX servers. When installing a guest operating on a CLARiiON LUN, configure the LUN to use RAID 1/0 or RAID 5. Choose RAID 1/0 instead of RAID 5 to reduce rebuild times if there is a disk failure, and to reduce required drive counts when workloads are performance-bound as opposed to capacity-limited. Choose RAID 5 to provide the best efficiency of RAW storage for VMs that are capacity-bound as opposed to performance-limited.

For I/O-intensive application data volumes, it is best to separate OS images from application data. In this case, EMC recommends that you use either RDM or a single virtual disk configured on a VMFS volume; since they are dedicated to only one virtual machine (that is, the entire LUN is presented to the virtual machine), replication and backup of applications are almost similar to that of a physical server. The management complexity might increase if multiple RDM volumes are created on the servers. A mix of VMFS and raw device mapping volumes are allowed on an ESX server. However, note that ESX Server 2.x has a limit of 128 SCSI disks. This limit includes both local devices and SAN LUNs. With VMware ESX 4.0, 3.x/ESXi, the limit is increased to 256 SCSI disks.

Also, because of the use of VMware redo logs, EMC recommends that you use separate disks for test and development applications, virtual machine templates (because of sequential I/O intensity), and production LUNs. Virtual machine templates, ISO images, and archived VMs are good candidates for SATA/ATA drives.

VMware ESX 4.0/3.x/ESXi provide performance and reliability improvements where a single swap file is available for each virtual machine. These swap files and NVRAM files for a given VM can reside on a

VMFS-3 volume. Ensure that the VMFS-3 volume has enough space to accommodate the swap files. With ESX 2.x, a single swap file is used for all virtual machines.

Application data disks residing on virtual machines should be aligned with the CLARiiON disk stripe, just as they are on physical servers. When aligning RDMs, align them at the virtual machine level. For Windows virtual machines, use diskpart from Windows 2003 SP1 to perform the alignment.

For VMFS-2 volumes, the alignment can be done at the ESX Server level and virtual machine level using **`fdisk`**. VMFS-3 volumes are already aligned to 64KB during creation; however, for Intel-based systems the virtual disks from a VMFS-3 volume need to be aligned at the virtual machine level. OS disks are difficult to align, however they can be aligned using native or specialized software if needed.

Align app/data disks to a 64k. (This step will not be required on Windows 2008, Vista, or Windows 7, because in these newer operating systems partitions are created on 1MB boundaries by default.) When formatting the app/data NTFS partitions for Windows virtual machines:

- If you are running applications, follow the recommended allocation unit size if there is one.
- If there are no allocation unit recommendation, or if this is just a file share, use 8K or multiples of 8K.

For best performance, use VI Client or Virtual Infrastructure Web Access to set up your VMFS-3 partitions instead of using the ESX 4.0 or 3.x service console. Using VI Client or VI Web Access ensures that the starting sectors of partitions are 64K-aligned, which improves storage performance. Please review the VMware white paper on alignment, available on vmware.com, for details. This white paper is at http://www.vmware.com/pdf/esx3_partition_align.pdf.

Using CLARiiON metaLUNs, LUN migration, and Virtual Provisioning technology with VMware ESX 4.0/3.x/ESXi and 2.x

CLARiiON virtual LUN technology provides an additional layer of abstraction between the host and back-end disks. This technology consists of two features: CLARiiON metaLUNs and the CLARiiON LUN migration that is available on the CLARiiON storage system. This section explains how CLARiiON metaLUNs and CLARiiON LUN migration work with VMware ESX Server.

CLARiiON metaLUNs are a collection of individual LUNs. They are presented to a host or application as a single storage entity. MetaLUNs allow users to expand existing volumes on the fly using the stripe or concatenation method.

CLARiiON LUN migration allows users to change performance and other characteristics of existing LUNs without disrupting host applications. It moves data—with the change characteristics that the user selects—from a source LUN to a destination LUN of the same or larger size. LUN migration can also be used on a metaLUN.

Virtual Provisioning, generally known in the industry as *thin provisioning,* increases capacity utilization for certain applications and workloads. It allows more storage to be presented to an application than is physically available. More importantly, Virtual Provisioning allocates physical storage only when the storage is actually written to. This allows more flexibility and can reduce the inherent waste in overallocation of space and administrative management of storage allocations. For more details on CLARiiON Virtual Provisioning, please see the *EMC CLARiiON Virtual Provisioning* white paper available on EMC.com and Powerlink.

CLARiiON metaLUNs, LUN migration, and Virtual Provisioning are supported with both VMFS and RDM volumes.

Expanding and migrating LUNs used as raw device mapping

A LUN presented to VMware ESX Server (ESX 4.0, 3.x, or ESX 2.x) can be expanded with metaLUNs using the striping or concatenation method. After the CLARiiON completes the expansion, rescan the HBAs using either VMware vCenter for VMware ESX 3.x/ESXi or the Management User Interface for

VMware ESX 2.x to ensure the ESX service console and VMkernel see the additional space. Since the LUN is presented to the virtual machine, expansion must take place at the virtual machine level. Use the native tools available on the virtual machine to perform the file system expansion at the virtual machine level.

CLARiiON LUN migration conducted on VMFS or RDM volumes is transparent to the guest OS. For RDM volumes, if the destination LUN is larger than the source LUN after the migration process completes, use the procedure previously outlined to rescan the HBAs, and then expand the disk at the virtual machine level. Note that RDM volumes must use the **physical compatibility** mode for expansion when using the CLARiiON metaLUN technology.

Expanding and migrating LUNs used as VMFS volumes

VMware ESX supports the volume management functions where VMFS volumes can be concatenated together as a single volume. This procedure is also called VMFS spanning in ESX 2.x and is done at the ESX Server level. VMware ESX 4.0/3.x/ESXi also provide volume management functionality for VMFS volumes through a process called Adding Extents with ESX 3.x or the "Increase" function in ESX 4 within VMware vCenter. The difference between VMFS-2 spanning within ESX 2.x and volume management using VMFS-3 within VMware ESX 3.x/ESXi is:

- Unlike VMFS-2, a VMFS-3 volume can be extended while in use.
- With VMFS-2, loss of any partition renders the whole volume inaccessible. For VMFS-3, except for the head partition, loss of a partition renders only the data on that partition inaccessible.

The first option to expand a CLARiiON LUN configured as a VMFS-2 or VMFS-3 volume is to add a new CLARiiON LUN and concatenate the two LUNs using the VMFS spanning process in ESX 2.x, or by adding extents in VMware ESX 3.x/ESXi or by using the "Increase" function in ESX 4.0. To expand the virtual disk presented to the virtual machine, use the **vmkfstools** utility available on ESX Server.

The other option is to expand a VMFS-3 volume using CLARiiON metaLUNs and span, add an extent, or increase the size of the VMFS datastore using the additional space in the original VMFS volume available before expansion. Steps required to expand a VMFS datastore in ESX 4.0 are as follows:

1) Expand the CLARiiON LUN on the CLARiiON storage system to the desired LUN size.
2) Issue a rescan at the ESX level.
3) Select the **Properties** tab of the datastore and select the **Increase** option as shown in Figure 17.
4) After selecting the increased original CLARiiON LUN through the "Increase" wizard, the VMFS datastore size automatically increases as needed.
5) With ESX 4.0, the VMFS datastore size can be increased to almost a 2 TB limit while the virtual machines are powered on while the "Increase" wizard is executed. This is not true for ESX 3.x/ESX3i since the virtual machines must be powered off before increasing the size of the VMFS datastore.

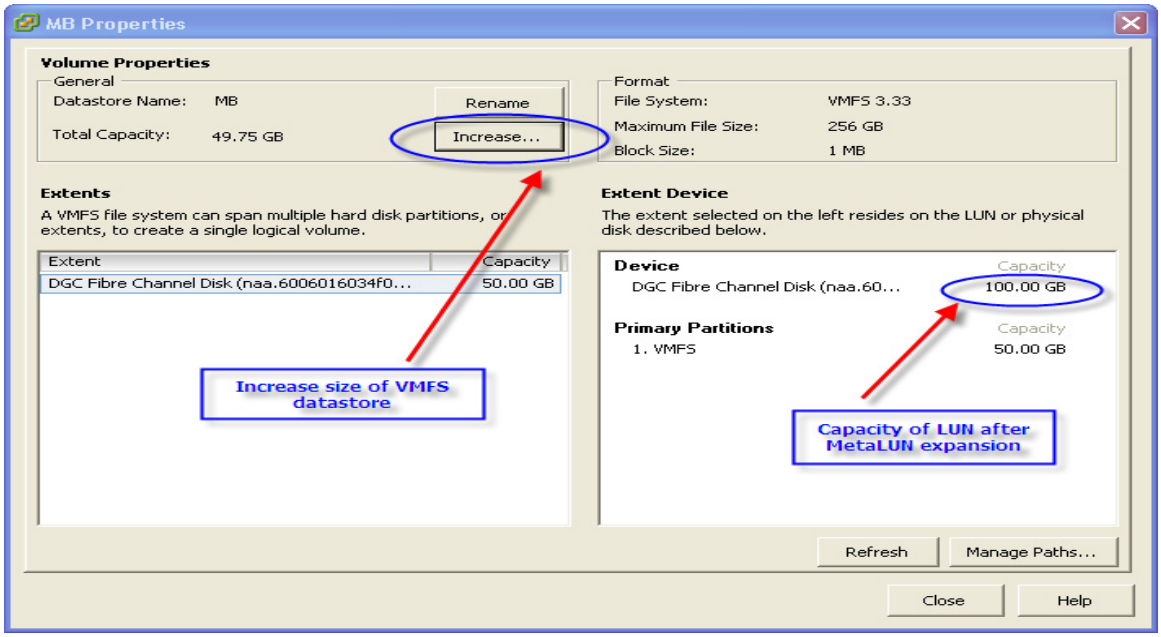


Figure 17. VMFS datastore "Properties" dialog for LUN expansion on ESX 4.0

With ESX 3.5, after expanding the VMFS volume, you can expand the individual virtual disk given to the virtual machine by using the **vmkfstools –extendvirtualdisk** option, but first you must power off the virtual machine that uses the virtual disk

With ESX 4.0, hot virtual disk (.vmdk) expansion is supported; you can use the **Virtual Machine Properties** dialog box to expand the volume *without* powering off the virtual machine that uses the virtual disk. However, the virtual disk must be in persistent mode and not have any snapshots associated with it. After the virtual disk is expanded, a guest OS rescan should show the additional space. As a best practice, always have a backup copy in place before performing any of these procedures.

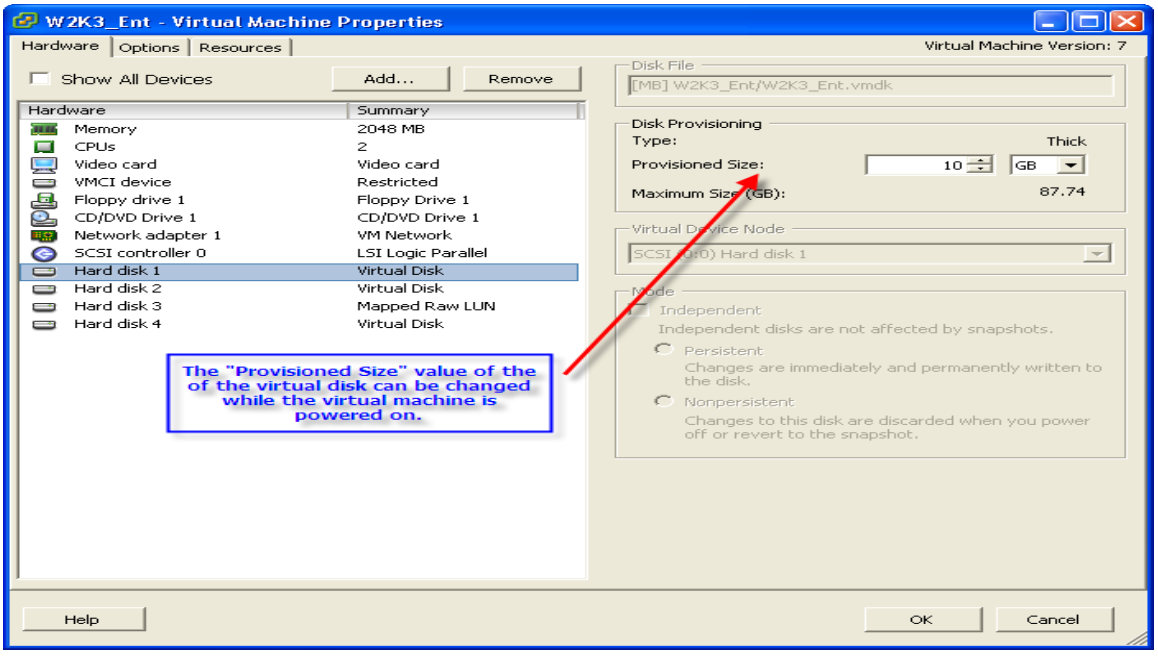


Figure 18. Hot virtual disk expansion through "Edit settings" of the virtual machine

When CLARiiON LUN migration is used to migrate to a larger LUN, after the migration completes, and a rescan is performed on the VMware ESX server, additional space for the LUN is visible. Use the procedure for expanding the VMFS volume for CLARiiON metaLUNs discussed previously.

CLARiiON Virtual Provisioning with VMFS and RDM volumes

A CLARiiON thin pool can contain multiple thin LUNs that can be assigned to multiple hosts. The space assigned to these thin LUNs is the space that the VMware ESX server sees. This does not mean that the space is fully allocated to the thin LUN from the thin pool. As the host writes to the thin LUN, space is allocated on the fly from the thin pool.

A thin LUN created on a thin pool can be used to create a VMware file system (VMFS), or assigned exclusively to a virtual machine as a raw disk mapping (RDM). Testing has shown that the VMFS datastore is *thin friendly*, meaning when a VMware file system is created on Virtual Provisioning (thin) LUNs, a minimal number of thin extents is allocated from the thin pool. Furthermore, a VMFS datastore reuses previously allocated blocks, thus benefiting from Virtual Provisioning LUNs. When using RDM volumes, the file system or device created on the guest OS will dictate whether the RDM volume will be thin friendly. Table 4 lists the allocation polices when creating new virtual disks.

Table 4. Allocation policies when creating new virtual disks on a VMware datastore

Allocation mechanism (Virtual Disk format)	VMware kernel behavior
Zeroedthick	All space is allocated at creation but is not initialized with zeroes. However, the allocated space is wiped clean of any previous contents of the physical media. All blocks defined by the block size of the VMFS datastore are initialized on the first write. This is the default policy when creating new virtual disks.
Eagerzeroedthick	This allocation mechanism allocates all of the space and initializes all of the blocks with zeroes. This allocation mechanism performs a write to every block of the virtual disk, and hence results in equivalent storage use in the thin pool.
Thick (not available with ESX 4.0)	A thick disk has all the space allocated at creation time. If the guest operating system performs a read from a block before writing to it, the VMware kernel may return stale data if the blocks are reused.
Thin	This allocation mechanism does not reserve any space on the VMware file system on creation of the virtual disk. The space is allocated and zeroed on demand.
Rdm	The virtual disk created in this mechanism is a mapping file that contains the pointers to the blocks of SCSI disk it is mapping. However, the SCSI INQ information of the physical media is virtualized. This format is commonly known as the “Virtual compatibility mode of raw disk mapping”.
Rdmp	This format is similar to the rdm format. However, the SCSI INQ information of the physical media is not virtualized. This format is commonly known as the “Pass-through raw disk mapping”.
Raw	This mechanism can be used to address all SCSI devices supported by the kernel except for SCSI disks.
2gbsparse	The virtual disk created using this format is broken into multiple sparsely allocated extents (if needed), with each extent no more than 2 GB in size.

For ESX 3.x, the **zeroedthick** (default) should be used when you create virtual disks on VMFS datastores, since this option does not initialize or zero all blocks and claim all the space during creation. RDM volumes are formatted by the guest operating system, hence virtual disk options like zeroedthick, thin, and eagerzeroedthick only apply to VMFS volumes.

When the zeroedthick option is selected for virtual disks on VMFS volumes, the guest operating file system (or writing pattern of the guest OS device) has an impact on how the space is allocated; if the guest

operating file system initializes all blocks, the virtual disk will need all the space to be allocated up front. Note that when the first write is triggered on a zeroedthick virtual disk, it will write zeroes on the region defined by the VMFS block size and not just the block that was written to by the application. This behavior will impact performance of array-based replication software since more data needs to be copied based on the VMFS block size than needed. If the thick option is used (as shown in the table) when using array-based replication software, only the block that it is written to is consumed. However there is a possibility that stale data might be returned to the user if the blocks are reused.

In ESX 4.0, a virtually provisioned CLARiiON LUN can be configured as zeroedthick or thin. When using the thin virtual disk format, the VMFS datastore is aware of the space consumed by the virtual machine, as shown in Figure 19. When using the virtual disk thin option, the VMware admin needs to monitor the VMFS datastore consumed capacity; vSphere provides a simple alert when datastore thresholds are reached.

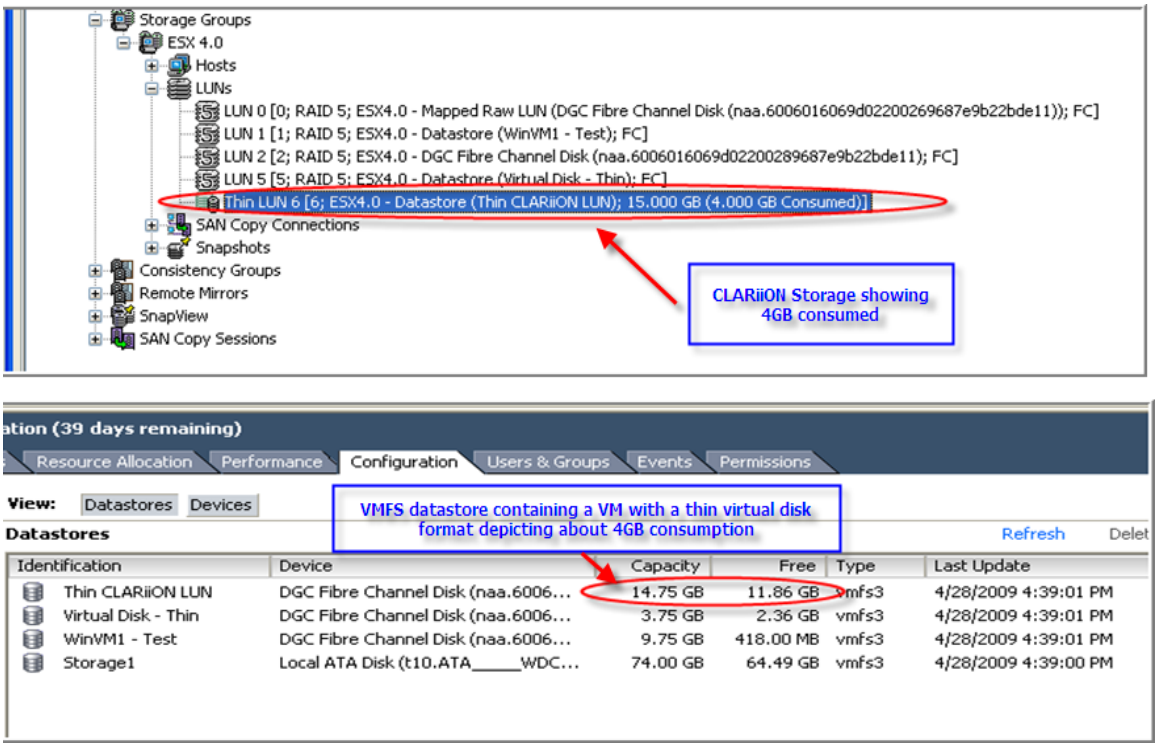


Figure 19. View VMFS datastore and CLARiiON LUN consumption when using the virtual disk "thin" format

In addition, with ESX 4.0, when using the vCenter features like Cloning, Storage VMotion, Cold Migration, and Deploying a template, the zeroedthick or thin format remains intact on the destination datastore. In other words, the consumed capacity of the source virtual disk is preserved on the destination virtual disk and not fully allocated. This is not the case with ESX 3.x/ESXi where the zeroedthick or thick format is changed to eagerzeroedthick format when operations like Cloning, Storage VMotion, and others are performed on the source virtual disk resulting in a fully allocated destination virtual disk.

Using CLARiiON replication software with VMware ESX 4.0/3.x/ESX ESXi and 2.5.x

CLARiiON replication software products including SnapView, MirrorView, and SAN Copy are supported with VMware ESX Server using both VMFS and RDM volumes. The OS image and the application/data can be replicated using CLARiiON replication software. The following considerations apply to iSCSI and

FC storage systems. Please note that remote replication software (MirrorView and SAN Copy) is supported on CLARiiON iSCSI storage systems:

CLARiiON replication considerations with VMware ESX Server

Please note that:

- Use of RDM volumes for replication is not supported when an ESX 2.5.x server is booted from a SAN LUN. In other words, when the Fibre Channel HBAs are shared between the service console and the virtual machines, RDM cannot be configured on an ESX 2.5.x server. There is no such restriction with VMware ESX 4.0/3.x/ESXi.
- `admsnap` and `admhost` must be installed on the virtual machines and not the ESX Server service console.
- With ESX 2.5.x, ensure that a CLARiiON snapshot, clone, or mirror is not in a **device not ready** state (snapshot not activated, session not started, clone not fractured, or secondary mirror not promoted) when it is assigned to an ESX server. The ESX service console does not create a device file for a LUN in this state. This restriction only applies at the ESX service console level and not the virtual machine level. Users can execute activate and deactivate operations at the virtual machine level using `admsnap` and `admhost` utilities after the ESX server sees the device the first time. For an AX100 system running Navisphere Express and connected to an ESX 2.5.x server, the replica must be presented to a secondary physical server (and not an ESX server) since the replica cannot be activated through the Navisphere Express GUI. There is no such restriction with VMware ESX 4.0/3.x/ESXi; it sees the snapshot, clone, or mirror in spite of the device's not ready condition.

CLARiiON replication software considerations when using VMFS volumes

Please note that:

- The virtual disks in a VMFS-2 volume must be in **persistent** mode during the replication process.
- When using VMFS-2 volumes, do not present two copies of the same VMFS volume to the same ESX server. For example, an ESX server participating in VMotion with the primary ESX server has access to the original source LUNs. Hence, this ESX server should not be a target when a replica is presented.

  With VMFS-3 volumes on ESX 3.x/ESXi, the replica can be presented to the same ESX server or a standby ESX server. This can be done using VMware vCenter by enabling the **LVM.EnableResignature** parameter in the ESX server. After a rescan, the replica is resignatured and can be assigned to a virtual machine.

  See the "Automatic Volume Resignaturing" section in the *Fibre Channel SAN Configuration Guide* on www.vmware.com.

  With VMFS-3 volumes on ESX 4.0/4i, the replica can be presented to the same ESX server or a standby ESX server. ESX 4.0 supports selective resignaturing at an individual LUN level and not at the ESX level. After a rescan, the user can either keep the existing signature of the replica (LUN) or can resignature the replica (LUN) if needed. See the "Managing Duplicate VMFS Datastores" section in the *Fibre Channel SAN Configuration Guide* on www.vmware.com.
- When replicating an entire VMFS (VMFS-2 and VMFS-3) volume that contains a number of virtual disks on a single CLARiiON LUN, the granularity of replication is the entire LUN with all its virtual disks.
- CLARiiON VSS Provider is not supported on VMFS volumes.
- When making copies of VMFS volumes that span multiple CLARiiON LUNs, use the array-based consistency technology.
- ESX Server-based VM snapshot copies should not be used in conjunction with CLARiiON replication software copies on the same VMFS volume for ESX 2.5.x since VM snapshot copies require the virtual disk to be in nonpersistent mode.
- Most of the **admsnap** and **admhost** commands when issued on VMFS volumes will fail since VMFS volumes do not support SCSI pass-through commands to communicate with the CLARiiON storage

system. Use Navisphere Manager or Navisphere CLI instead. The only commands that will work are **admsnap flush** and **admhost flush**.

- VMFS volumes are not supported when replicating application data images from a physical (native) server to an ESX server.
- For ESX 2.5.x, ensure that a CLARiiON snapshot, clone, or mirror of a VMFS volume is not in a **device not ready** state (snapshot not activated, session not started, clone not fractured, or secondary mirror not promoted) when it is assigned to an ESX server.

Since VMFS-3 volumes may contain VM configuration files, swap files, and NVRAM files, these files can be replicated using CLARiiON replication software.

CLARiiON replication software considerations when using RDM volumes

Please note that:

- You should configure the LUNs to use the **physical compatibility mode** option when replicating RDM volumes using CLARiiON replication software. Otherwise, **admsnap** and **admhost** will not work on the virtual machine.
- VMware ESX servers do not write a *signature* on RDM volumes. Hence, replicas can be presented back to the same VMware ESX server for use. The copies cannot be used on the source virtual machines unless the guest OS supports this feature. However, they can be assigned as raw devices to another virtual machine.
- RDM volumes are supported on the ESX server when replicating application data images from a physical (native) server to an ESX server.

When replicating OS disks while the virtual machine is powered on, the replica or copy will be in a crash-consistent state since there is no mechanism available to quiesce a boot image. For application data disks, native tools available with the application can be deployed to quiesce the application to get a consistent replica. The array-based consistency technology can be used when applications span multiple LUNs or for write-order dependent applications such as databases. For automation, scripts may need to be developed to integrate CLARiiON replication software with the different applications running on the virtual machines. EMC Replication Manager automates this process by integrating with virtual machine applications, vCenter/ESX, and CLARiiON replication software.

Using EMC Replication Manager with VMFS and RDM volumes

EMC Replication Manager supports the replication of VMFS and RDM volumes. Replication Manager can replicate the guest OS and the application data volumes. Following are some things to consider when using Replication Manager with VMFS and RDM volumes.

Using VMFS volumes

- If the VMFS volume contains multiple virtual disks assigned to one or more VM, the entire VMFS volume (OS or application data) can be replicated with Replication Manager. Replication Manager automates the process of presenting the VMFS replica to a secondary ESX server. Then the VMware administrator must assign the replica to an existing virtual machine or create new virtual machines.
- If a single virtual disk is configured on an entire VMFS volume and is presented to a VM as an application data disk, Replication Manager can quiesce, freeze the application and automatically assign the application's quiesced data-disk (virtual disk) replica to a secondary ESX server. Replication Manager also can assign or remove the virtual disk or replica (application data disk) on an individual virtual machine.

Using RDM volumes

- If using RDM volumes, only application data disk on a VM can be replicated. If replicating a RDM volume presented as an application data disk to a VM, Replication Manager enables the application data disk (virtual disk) to be quiesced and automatically assigns the application-

consistent data-disk replica of the RDM volume to a virtual machine connected directly (via iSCSI) to a CLARiiON storage system or a physical server. If the application data disk replica needs to be assigned to the VMware ESX server, the LUN must be manually placed in the VMware ESX server storage group, and the VMware administrator must then assign the replica (application data disk) to an existing virtual machine.

For more information on using Replication Manager with VMware and CLARiiON storage systems, refer to the *EMC Replication Manager Administrator's Guide* available on Powerlink.

CLARiiON and VMotion

Migration with VMotion allows you to move a virtual machine between two ESX servers while the virtual machine is powered on and performing transactions. When a migration with VMotion is performed, the operations of the virtual machine can continue uninterrupted. The virtual machine must reside on a SAN LUN accessible to both source and destination hosts. VMotion only moves the virtual machine configuration file and memory contents to the alternate host. Any disks assigned to the VM are moved by transferring their ownership.

The conditions for VMotion to work are:

- The VMware vCenter server and client must be installed on a Windows system.
- A Gigabit Ethernet connection is required between the two ESX Server hosts participating in VMotion migration.
- The guest operating system must boot from a CLARiiON LUN. The virtual machine boot LUN and its associated data disks must be shared between the source and destination hosts.
- VMware VMotion is supported with Fibre Channel and iSCSI connectivity to CLARiiON storage systems. Furthermore, VMware VMotion is supported in a configuration where both of the following occur:
  - A single LUN is presented to two VMware ESX Server nodes.
  - The ESX Server nodes are in a cluster in which one node accesses the LUN via FC and the other node accesses the LUN via iSCSI.

  CLARiiON storage systems preserve the LUN HLU number across both protocols. Therefore, when a LUN is presented via FC to one host and iSCSI to another, a false snapshot is *not* detected.
- VMFS and RDM volumes are supported with VMotion.
- Both hosts must have identical CPUs (both CPUs are P4, for instance) unless VMware's Enhanced VMotion Compatibility mode is used. The CPUs must also be manufactured by the same vendor.

Migration with VMotion cannot occur when virtual machines are in a raw, clustered, or nonpersistent mode. Figure 20 shows two ESX servers with three NICs each. The recommended number of NICs is three; two is the minimum requirement. A crossover or gigabit LAN connection should exist between the two gigabit NICs on the two VMware ESX servers. For ESX 2.x, a virtual switch must be created with the same network label on both servers for the two network cards to communicate. In VMware ESX 3.x/ESXi, a VMotion VMkernel port group on a vSwitch must be created on both ESX servers for VMotion migration.

VMotion with VMFS volumes

In order to perform VMotion with VMFS volumes, the following prerequisites must be met:

- All VMFS volumes assigned to the virtual machine that is to be migrated must be shared by both ESX servers.
- VMFS volumes must be in public access mode.

VMware ESX Server identifies VMFS-2 volumes by their label information. For VMFS-3 volumes, they are identified by the Host LUN number. Hence, as a best practice (unless doing boot from SAN for the

ESX hosts) for VMotion with VMFS volumes create a single storage group for all ESX servers in a farm or cluster since LUNs may be assigned different Host LUN numbers if separate storage groups are created. For additional information, see EMC Knowledgebase cases emc151686 and emc153719.

Starting with the VMware ESX Server 3.5 latest patch version, LUNs are identified by their LUN NAA (WWNs), thus the requirement that the Host LUN numbers (HLU) must match across storage groups for VMotion is no longer valid.

VMotion with RDM volumes

In order to perform VMotion with RDM volumes, the following prerequisites must be met:

- Both the RDM LUN and the VMFS volume that contains the mapping file must be shared by both ESX servers.
- VMFS volumes must be in public access mode. This mode allows the volume to be accessed by multiple ESX servers.
- The RDM LUN must have the same host LUN number on the source and destination ESX servers. This can be set when adding LUNs to the storage group in Navisphere Manager.
- RDM volumes in both physical and virtual compatibility mode are supported with VMotion.

When using RDM volumes, the recommendation is to create a single storage group since LUNs may be assigned different Host LUN numbers if separate storage groups are created.

Initially, swap files for virtual machines had to be stored on shared storage for VMotion. VMware Infrastructure 3 (ESX Server 3.5 hosts and VMware vCenter 2.5 or later) now allows swap files to be stored on local storage during VMotion migrations for virtual machines. For virtual machines with swap files on local storage, when local storage is the destination during a VMotion migration or a failover, the virtual machine swap file is re-created. The creation time for the virtual machine swap file depends on either local disk I/O, or on how many concurrent virtual machines are starting due to an ESX Server host failover with VMware HA.

If you have an AX4 system running Navisphere Express attached to a VMware ESX server (not VMware ESX Server 3.5 and later), you can use Navisphere CLI to implement VMotion. Ensure that the Host LUN number is consistent across all ESX servers using the **storagegroup** command. Your other option is to set the LVM.Disallowsnapshot and LVM.EnableResignature parameter to 0 on all the ESX servers or LUNs participating in VMotion/DRS/VMware HA for ESX 4.0 and/or 3.x servers.

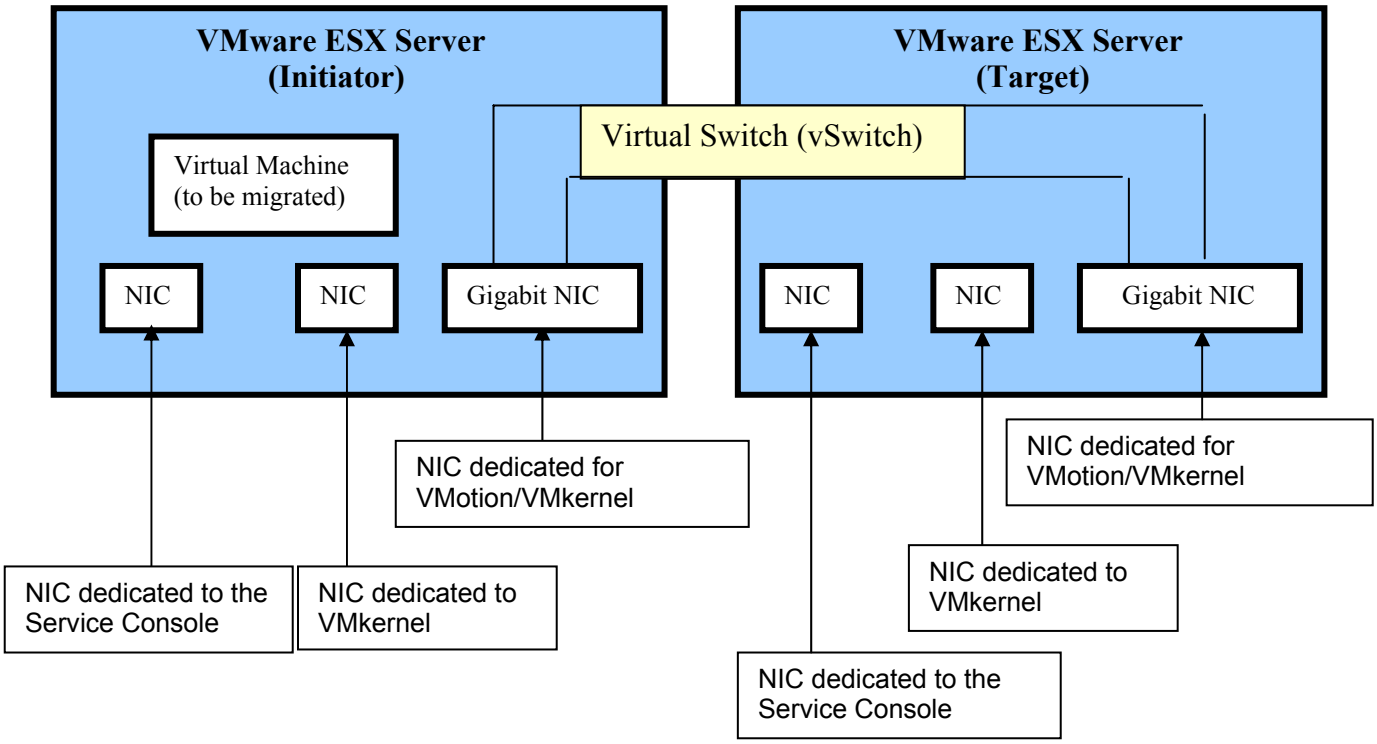


Figure 20. Two VMware ESX servers ready for VMotion migration

Using the VMware vCenter console, you can initiate the migration process by right-clicking the virtual machine for the initiator host, as shown in Figure 21. After the migration request has been initiated, a wizard opens requesting initiator and target information. VMotion migration takes place through the gigabit network connection setup between the two ESX servers. After the VMotion migration process completes, the virtual machine is automatically resumed on the target VMware ESX server with the same name and characteristics as the initiator virtual machine.

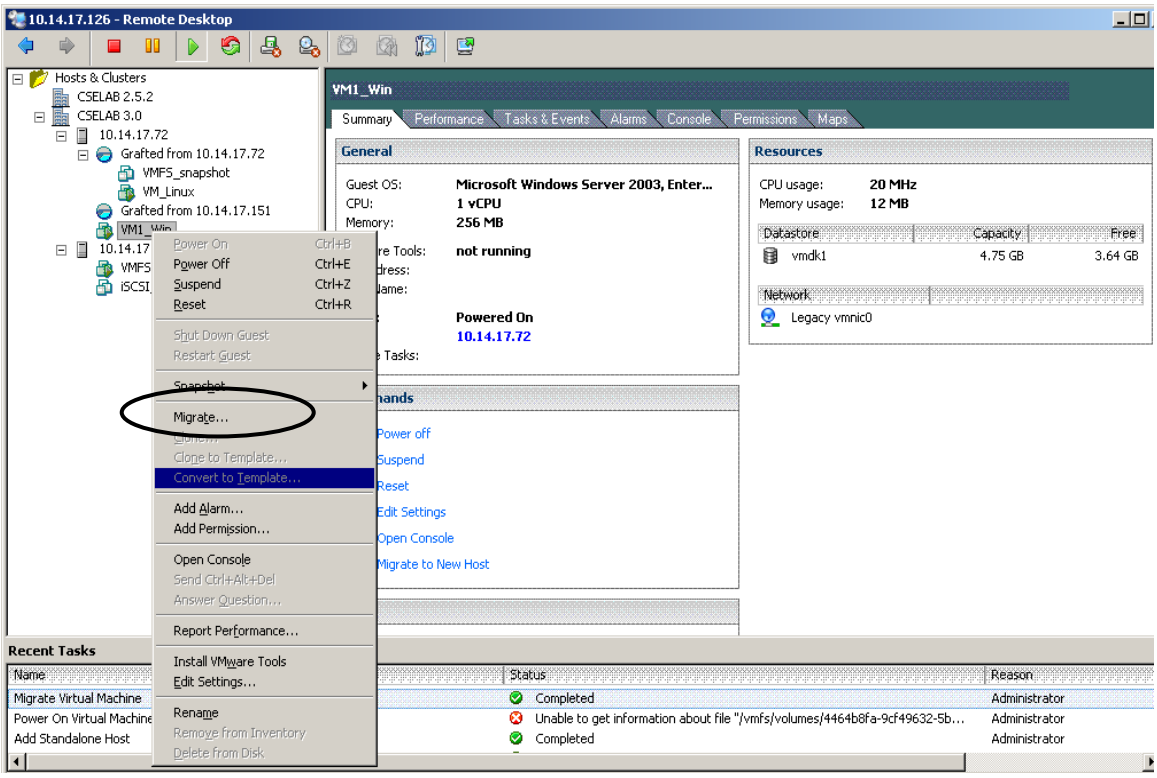


Figure 21. VMware vCenter 4.0 management screen showing how VMotion migration is initiated and completed

CLARiiON with VMware Distributed Resource Scheduling and High Availability

Both VMware Distributed Resource Scheduling (DRS) and VMware High Availability (HA), when used with VMotion technology, provide load balancing and automatic failover for virtual machines with VMware ESX 4.0/3.x/ESXi. To use VMware DRS and HA, a cluster definition must be created using VMware vCenter 2.0. The ESX hosts in a cluster share resources including CPU, memory, and disks. All virtual machines and their configuration files on ESX servers in a cluster must reside on CLARiiON storage, so that you can power on the virtual machines from any host in the cluster. Furthermore, the hosts must be configured to have access to the same virtual machine network so VMware HA can monitor heartbeats between hosts on the console network for failure detection.

The conditions for VMware DRS and HA to work are as follows:

- The guest operating system must boot from a CLARiiON LUN. The virtual machine boot LUN, its associated data disks, and configuration files must be shared among all ESX servers in a cluster.
- VMware HA and DRS are supported with both Fibre Channel and iSCSI CLARiiON storage systems.
- Both VMFS and RDM volumes are supported and are configured exactly as they would be for VMotion. See the “CLARiiON and VMotion” section on page 34.
- DRS is based on the VMotion technology, therefore, ESX servers configured for a DRS cluster must pass the CPU type check (identical CPUs, same manufacturer). VMware HA does not depend on CPU type check.

CLARiiON and virtual machine clustering

Clustering refers to providing services through a group of servers to achieve high availability and/or scalability. Clustering with VMware ESX Server is supported at the virtual machine level. Refer to the E-Lab Navigator for the cluster software products that are supported on virtual machines. To implement clustering at the virtual machine level, the virtual machines must boot from local disks and not CLARiiON disks. There are two types of cluster configuration at the virtual machine level:

- In-the-box cluster
- Out-of-the-box cluster
  - Virtual to virtual
  - Virtual to physical

In-the-box cluster

This section outlines clustering virtual machines on a CLARiiON storage system between virtual machines on the same VMware ESX server. This provides simple clustering to protect against software crashes or administrative errors. The cluster consists of multiple virtual machines on a single ESX server.

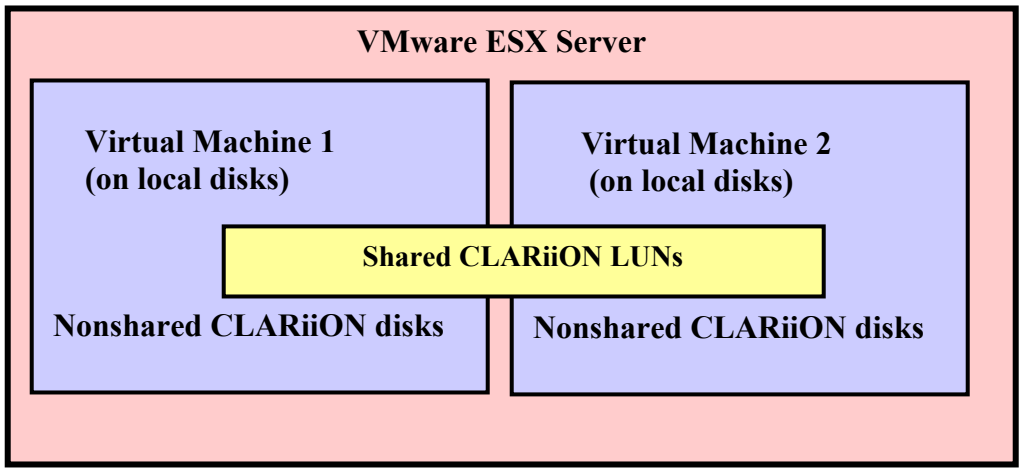


Figure 22. In-the-box cluster configuration

In Figure 22, a single VMware ESX server consists of two virtual machines. The quorum device and/or clustered applications are shared between the two virtual machines. Each virtual machine can have virtual disks that are local to the virtual machine, that is, virtual disks not shared between virtual machines. The device containing the clustered applications is added to the storage group of the VMware ESX server and is assigned to both virtual machines using the VMware MUI for VMware ESX Server 2.5.x or VMware vCenter for VMware ESX 4.0/3.x/ESXi. Additional CLARiiON disks can be assigned to each virtual machine for running other non-clustered applications. Only virtual disks on VMFS volumes are supported with this configuration for ESX 2.5.x. With ESX Server 4.0/3.x, RDM volumes are also supported with an in-the-box cluster.

Out-of-the-box cluster

An out-of-the-box cluster consists of virtual machines on multiple physical machines. The virtual disks are stored on shared physical disks, so all virtual machines can access them. Using this type of cluster, you can protect against the crash of a physical machine.

Virtual-to-virtual clustering

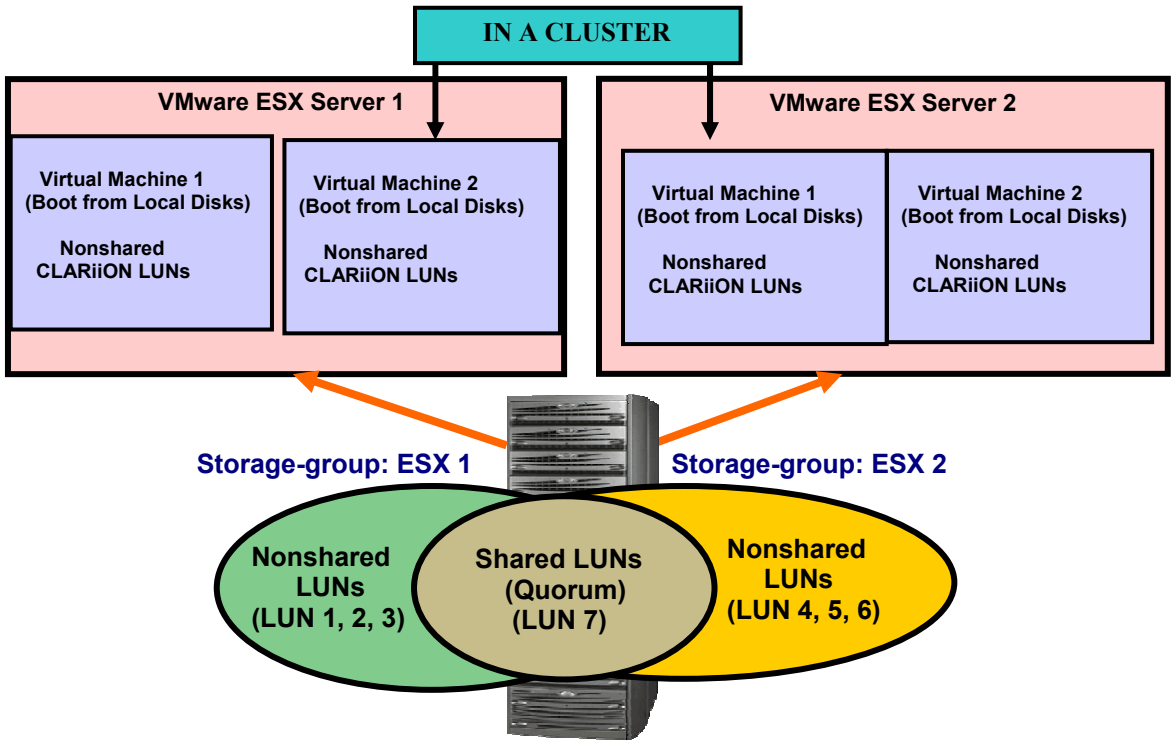


Figure 23. Out-of-the-box cluster configuration

Figure 23 shows two VMware ESX servers configured with two virtual machines each. The virtual machine boot image must reside on local disks to cluster virtual machines. The bus sharing must be set to **physical**. The quorum device and/or clustered applications residing on CLARiiON disks are shared at the CLARiiON level by assigning the respective LUNs to both storage groups. In this case, LUN 7 is assigned to both storage groups and is a shared LUN. This device is then assigned to **Virtual Machine 2** for ESX Server 1 and **Virtual Machine 1** for ESX Server 2, using VMware vCenter for VMware ESX 3.x/ESXi or the MUI for ESX 2.5.x. Additional CLARiiON disks can be assigned to each virtual machine for running non-clustered applications.

The shared resource can be configured as raw disks, VMFS, and/or RDM volumes for a virtual-to-virtual cluster configuration with ESX 2.5.x. If using VMFS volumes for this configuration, the access mode for VMFS volumes must be **shared.** For VMware ESX 4.0/3.x/ESXi, only RDM volumes (set to physical and virtual compatibility mode) are supported for a virtual-to-virtual cluster configuration. For RDM volumes, both the RDM volume and the VMFS volume that contain the mapping file must be shared by both ESX servers. The VMFS volume that contains the mapping file must be in **public** access mode.

Physical-to-virtual clustering

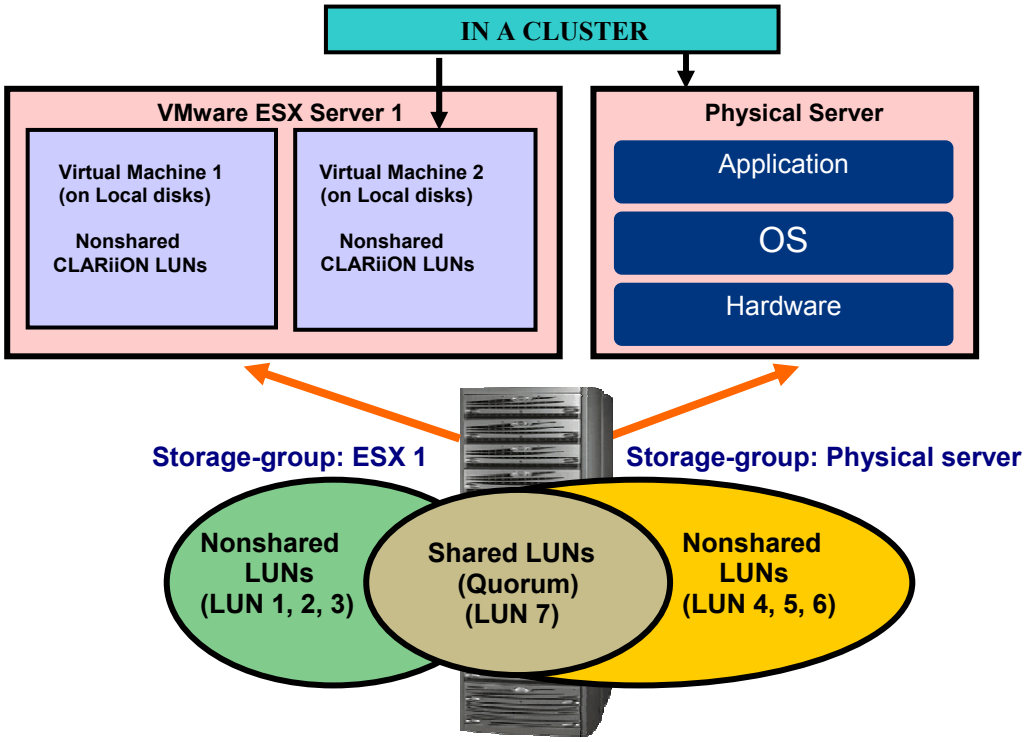


Figure 24. Out-of-the-box cluster configuration (physical to virtual)

Figure 24 shows a VMware ESX server configured with two virtual machines and a physical server. The virtual machine boot image for the VMware ESX server must reside on local disks to cluster virtual machines. The quorum device and/or clustered applications residing on CLARiiON disks are shared at the CLARiiON level by assigning the respective LUNs to both storage groups. In this case, LUN 7 is assigned to both storage groups and hence is a shared LUN. This device is then assigned to virtual machine 2 using VMware vCenter for VMware ESX 4.0/3.x/ESXi or MUI for ESX 2.5.x. Additional CLARiiON disks can be assigned to each virtual machine for running other non-clustered applications.

The shared resource can be configured as raw disks (ESX 2.x) and/or RDM (physical compatibility mode) volumes, and are supported for a virtual-to-physical cluster configuration. For RDM volumes, only the RDM volume needs to be assigned to both servers. The VMFS volume that contains the mapping file can be in **public** access mode.

MirrorView/Cluster Enabler (MV/CE) and VMware support

EMC MirrorView/Cluster Enabler (MV/CE) is a replication-automation product from EMC. MV/CE is host-based software that integrates with Microsoft Failover Cluster software to allow geographically dispersed cluster nodes across MirrorView links. MV/CE seamlessly manages all the storage system processes necessary to facilitate cluster-node failover. MirrorView/A replication and MirrorView/S replication are supported.

MirrorView/CE also supports *guest clustering* for VMware ESX servers. Guest clustering enables high availability for services and applications running in the virtual layer. This is done by installing Failover Clustering on several virtual machines, then clustering them as if they were physical nodes.

MV/CE and guest clustering using RDMs

In this type of guest clustering, RDM volumes are assigned to the virtual machines and failover-cluster software, and MV/CE software is installed and configured on these virtual machines. The following restrictions apply when working with this configuration:

- Only physical compatibility mode RDMs can be used for cluster shared storage.
- Virtual-to-physical clustering (for example, standby host clustering) is not supported.
- MV/CE 3.1.0.14 (see EMC Knowledgebase emc213675) must be used.
- CLARiiON storage system must be running FLARE release 26 or later.

MirrorView/CE and guest clustering using direct iSCSI disks

In this type of guest clustering, the failover cluster software and MirrorView/CE software are installed on the virtual machines and direct iSCSI disks are configured. In other words, the Microsoft software initiator must be running on the individual guest operating systems, and must be connected directly to the CLARiiON storage system via iSCSI. You can configure virtual machines as cluster nodes in the **Failover cluster wizard** provided by Microsoft.

For more information, please see the following resources on Powerlink:

- *EMC MirrorView/Cluster Enabler (MV/CE) – A Detailed Review* white paper
- *EMC MirrorView/Cluster Enabler Product Guide*
- *EMC MirrorView/Cluster Enabler Release Notes*

Setup for Microsoft Cluster Service is available at http://www.vmware.com/pdf/vi3_35/esx_3/r35u2/vi3_35_25_u2_mscs.pdf.

CLARiiON and VMware NPIV support

VMware ESX 4.0 and 3.5/ESXi support NPIV, which allows individual virtual machines to have their own virtual WWNs. This allows SAN vendors to implement their QoS tools to distinguish I/O from an ESX server and a virtual machine. VMware NPIV is still primitive with a lot of restrictions; however, this section gives the user an idea on how to configure VMware NPIV with CLARiiON storage systems,

To configure VMware NPIV, the HBAs (within the ESX server) and the FC switches must support NPIV, and NPIV must be enabled on each virtual machine. In order to enable NPIV on a virtual machine, at least one RDM volume must be assigned to the virtual machine. In addition, to use the NPIV feature within VMware, LUNs must be masked to both VMware ESX Server and the virtual machine that is NPIV enabled.

Figure 25 shows how to enable NPIV for a virtual machine. As mentioned, for the NPIV feature to be enabled, an RDM volume must be presented through the ESX server to the virtual machine. Note that once NPIV is enabled virtual WWNs are assigned to that virtual machine.

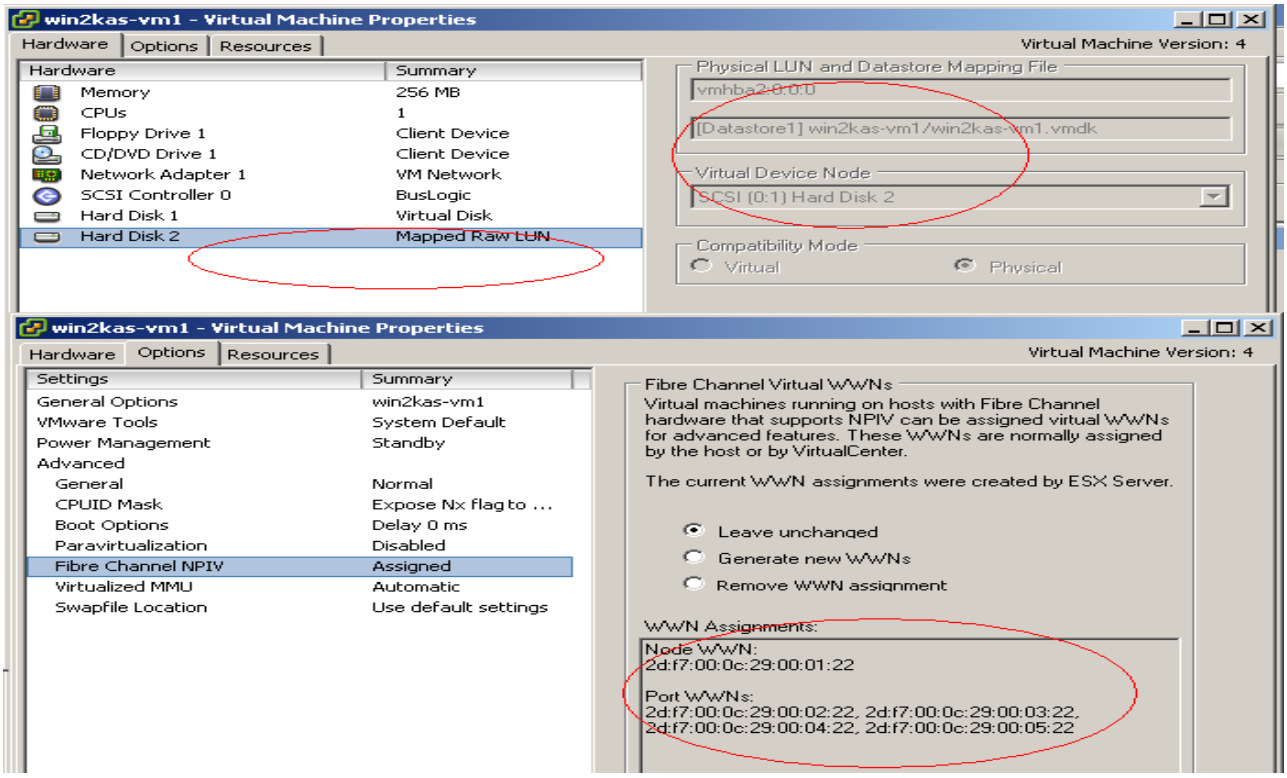


Figure 25. Enable NPIV for a virtual machine after adding a RDM volume

For some switches, the virtual WWNs' names must be entered manually within the switch interface and then zoned to the storage system. The CLARiiON storage system can then see the initiator records for the virtual machine (virtual WWNs). These initiator records need to be manually registered as shown in Figure 26. A separate storage group can be created for each virtual machine that is NPIV enabled; however, additional LUNs to that virtual machine must be masked to both the ESX server and the virtual machine that is NPIV enabled.

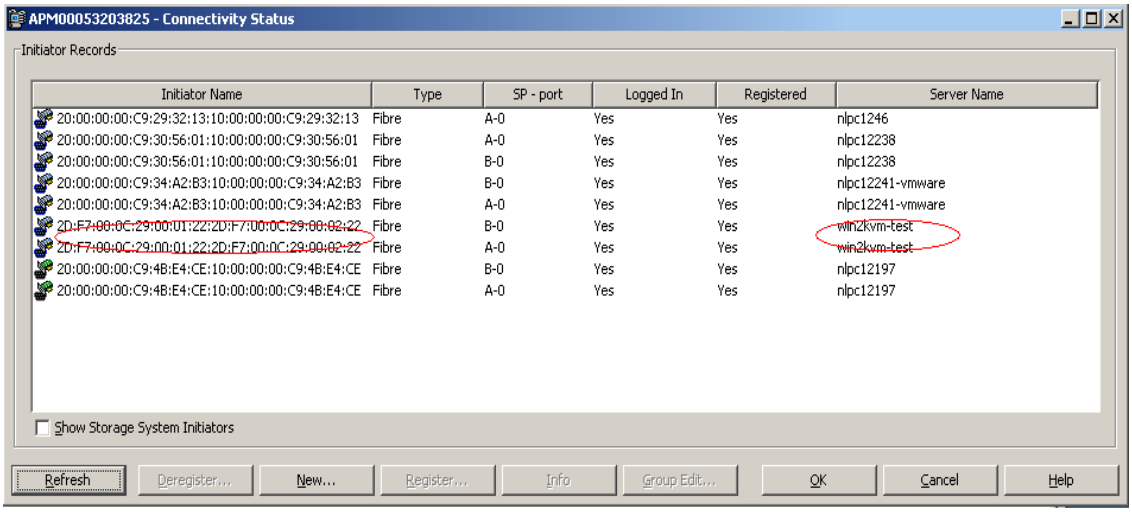


Figure 26. Manually register virtual machine (virtual WWN) initiator records

The following points summarize the steps required to configure NPIV:

1. Ensure the HBA, switch, and ESX version support NPIV.

2. Assign a RDM volume to the ESX server and then to the virtual machine.

3. Enable NPIV for that virtual machine to create virtual WWNs.

4. Manually enter the virtual WWNs within the switch interface. Some switches might be able to view the WWNs without manual entry of the virtual machine virtual WWNs.

5. Zone the virtual WWNs to the CLARiiON storage systems using the switch interface. Add them to the same zone containing the physical HBA initiator and CLARiiON storage ports.

6. Manually register the initiator records for that virtual machine using Navisphere using the same failovermode setting of that of the ESX server.

7. Add the virtual machine (host) to the same storage group as the ESX server or assign them to a different storage group.

8. To add LUNs to the virtual machine ensure that:

    a. LUNs are masked to the ESX server and the virtual machine storage group.
    b. LUNs have the same host LUN number (HLU) as the ESX server.
    c. VMs are defined in different storage groups.
    d. LUNs must be assigned as RDMs to each virtual machine

CLARiiON and VMware Site Recovery Manager (SRM)

VMware Site Recovery Manager (SRM) provides a standardized framework to automate site failover in conjunction with Storage Replication Adapters (SRAs) provided by storage vendors. CLARiiON has an SRA for MirrorView that works within the SRM framework to automate most of the steps required for a site failover operation. The EMC CLARiiON SRA supports MirrorView/S and MirrorView/A software for FC and iSCSI connections.

MirrorView/S and MirrorView/A run on all available storage system platforms, including CX4, CX3, and AX-4 systems. SRM added support for AX4 storage systems. MirrorView does not consume server resources to perform replication. Please consult the white paper *MirrorView Knowledgebook* on Powerlink for further information about MirrorView/S and MirrorView/A.

RecoverPoint SRA is also supported with VMware SRM; CLARiiON LUNs can be replicated using the CLARiiON splitter or the RecoverPoint appliance. For more details, please refer to the RecoverPoint documentation available on Powerlink.

SRM requires that the protected (primary) site and the recovery (secondary) site each has two independent virtual infrastructure servers. To use the MirrorView SRA, mirrors need to be created, and secondary LUNs need to be added. MirrorView SRA 1.3.0.4 and later can work with individual mirrors as well as consistency groups. But as a best practice, if a datastore spans multiple mirrors, it should be put in a consistency group and placed in a MirrorView consistency group. To leverage the test functionality within SRM, SnapView snapshots of the mirrors must exist at the recovery site within the proper CLARiiON Storage Group. (We also recommend that you create snapshots for the mirrors at the protected site, in case a failback is necessary). For installation and configuration information please see the *EMC MirrorView Adapter for VMware Site Recovery Manager Version 1.3 Release Notes*.

VMware Site Recovery Manager for vSphere 4.0 is currently in beta and thus not generally available.

The following steps outline the process for initializing an SRM environment using Navisphere Manager and/or Navisphere SecureCLI. The commands must be issued from a management host that is network connected to the production CLARiiON storage array. Note that all of these commands can be performed in the Navisphere Manager GUI or in CLI.

Using Navisphere Manager to configure MirrorView

To configure sync or async mirrors, open the wizard and follow the instructions in the wizard.

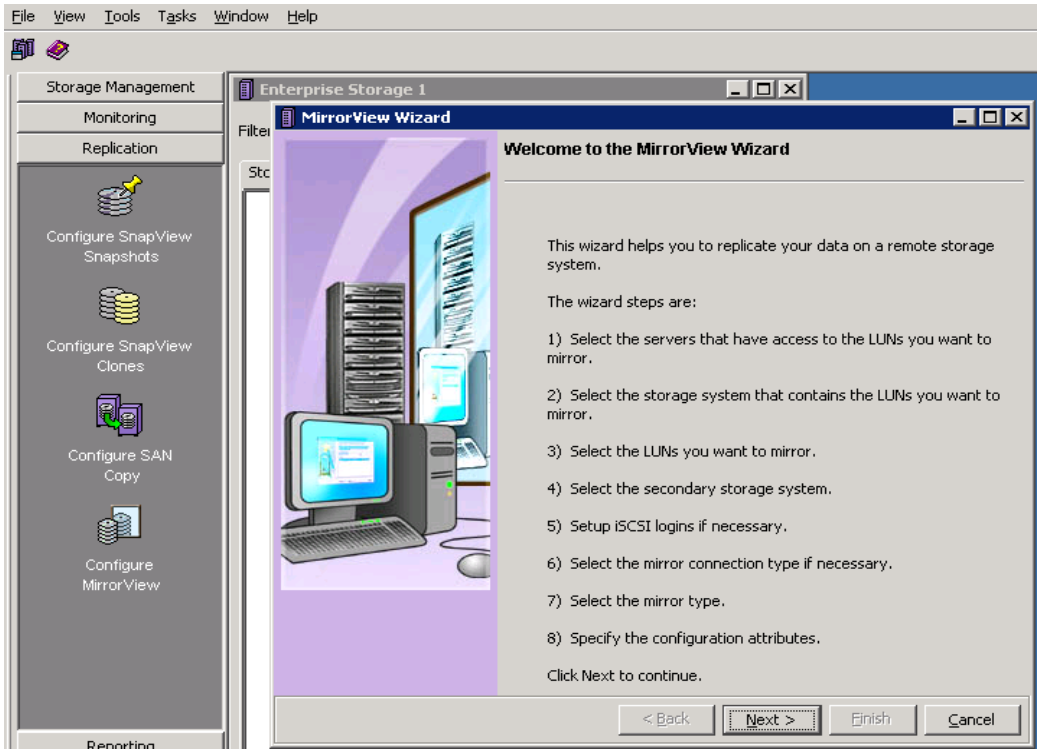


Figure 27. MirrorView Wizard

Configuring sync mirrors via NaviSecCLI

1. If not already established, create a path or paths for remote mirroring between the primary and secondary CLARiiON with this command:

   naviseccli –h SP ipaddress **mirror –sync –enablepath** SPhostname
   [**-connection** type fibre|iscsi]

2. Once you have created mirror paths, create a remote mirror of the LUN(s) that you wish to protect with SRM. The LUN(s) on which the mirror is created becomes the primary image.

   naviseccli –h SP ipaddress **mirror –sync –create –lun** <LUN_Number>

3. The secondary image on the remote CLARiiON can then be added to the primary image. After the secondary image is added, the initial synchronization between the primary and the secondary images is started. The following command assumes that the LUN(s) are already created on the remote CLARiiON storage system.

naviseccli –h SP ipaddress **mirror –sync –addimage –name** <name>
-arrayhost <sp-hostname| sp ipaddress> **-lun** <lunnumber| lun uid>

4. Even if there is only a single LUN being replicated to the secondary site, you still need to create a consistency group for SRM. The following commands show how to create a consistency group and add existing mirrors to the consistency group.

   naviseccli –h SP ipaddress **mirror –sync –creategroup –name** <name>

   naviseccli –h SP ipaddress **mirror –sync –addgroup –name** <name>
   -mirrorname <mirrorname>

5. If for some reason the mirrors are fractured, the **syncgroup** option (shown below), can be used to resynchronize the primary and secondary images:

   naviseccli –h SP ipaddress **mirror –sync –syncgroup –name** <name>

6. While the mirrors are synchronizing or a consistent state, you can add all the LUNs (if you have not already done so) to the ESX Server CLARiiON Storage Group at the protected and recovery site using the following command:

   naviseccli –h SP ipaddress **storagegroup –addhlu –gname** <ESX CLARiiON Storage Group Name> **-hlu** <Host Device ID> **-alu** <Array LUN ID>

Configuring async mirrors using NaviSecCLI

The following steps outline the process for setting up asynchronous replication using Navisphere SecureCLI commands. The commands should be run from a management host that is connected to the production CLARiiON storage array:

1. If not already established, create a path or paths for remote mirroring between the primary and secondary CLARiiON with this command:

   naviseccli –h SP ipaddress **mirror –async –enablepath** SPhostname
   [**-connection** type fibre|iscsi]

2. Once you have created mirror paths, create a remote mirror of the LUN(s) that you wish to protect with SRM. The LUN(s) on which the mirror is created becomes the primary image.

   naviseccli –h SP ipaddress **mirror –async –create –lun** <LUN_Number>

3. The secondary image on the remote CLARiiON can then be added to the primary image. After the secondary image is added, the initial synchronization between the primary and secondary images begins. (The following command assumes that the LUN(s) are already created on the remote CLARiiON storage system).

   naviseccli –h SP ipaddress **mirror –async –addimage –name** <name> **-arrayhost** <hostname| sp ip-address> **-lun** <lunnumber | lunuid>

4. Even if there is only a single LUN being replicated to the secondary site, you still need to create a consistency group for SRM. The following commands create a consistency group and add existing mirrors to the consistency group.

   naviseccli –h SP ipaddress **mirror –async –creategroup –name** <name>

   naviseccli –h SP ipaddress **mirror –async –addgroup –name** <name>

5. If for some reason the mirrors are fractured, the **syncgroup** option (shown next), can be used to resynchronize the primary and secondary images:

   naviseccli –h SP ipaddress **mirror –async –syncgroup –name** <name>

6. While the mirrors are synchronizing or in a consistent state, you can add all the LUNs (if you have not already done so) to the ESX Server CLARiiON Storage Group at the protected and recovery sites using the following command:

   naviseccli –h SP ipaddress **storagegroup –addhlu –gname** <ESX CLARiiON Storage Group Name> **-hlu** <Host Device ID> **-alu** <Array LUN ID>

Using SnapView to configure SnapView snapshots for SRM testing purposes
For SRM testing purposes, you need to create snapshots on the array at the SRM recovery site. Use the wizard to create and configure these snapshots. This wizard will create LUNs automatically to be placed within the Reserved LUN Pool. The default is to allocate 20% storage capacity to the LUN where the snapshot is created. If you have determined that this is not enough for your environment, override the value and select the appropriate percentage. Use the wizard to add the snapshots to the proper CLARiiON Storage Group at the SRM recovery site.

You can also use the **Configure SnapView Snapshot** wizard to create snapshots on the array at the SRM protection site, so that if a failback is necessary, this step has already been performed.

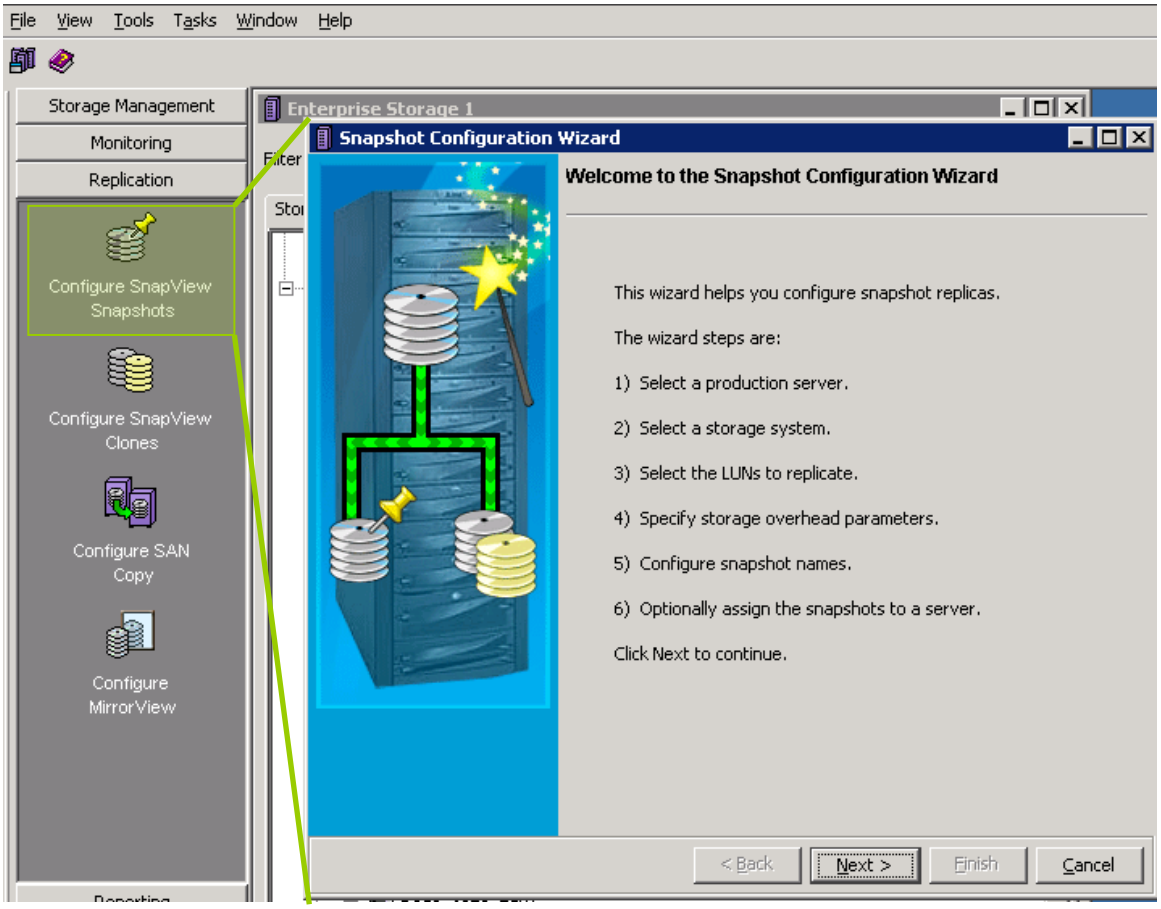


Figure 28. SnapView SnapShot Configuration Wizard

Configuring SnapView snapshots for SRM testing purposes via NaviSecCli

1. Add the LUNs bound for SnapView Sessions into the Reserved LUN Pool.
   naviseccli –h SP ipaddress **reserved –lunpool –addlun** <LUN IDS separated by spaces>

2. Create a snapshot for each LUN at the recovery site, and add the SnapShot to ESX Server's CLARiiON Storage Group at the recovery site.

   (**NOTE:** This snapshot will not be activated until a user tests the SRM failover operation, in which SRM will create a session and activate it with the corresponding snapshot.)

   naviseccli –h SP ipaddress **snapview –createsnapshot** <LUN ID>
   -snapshotname VMWARE_SRM_SNAP[1]_LUNID

   naviseccli –h SP ipaddress **storagegroup –addsnapshot –gname** <ESX CLARiiON Storage Group name> **-snapshotname** <name of snapshot>

For more information about using Navisphere CLI with MirrorView, please see the *MirrorView/Synchronous Command Line Interface (CLI) Reference, MirrorView/Asynchronous Command Line Interface (CLI) Reference* and *SnapView Command Line Interface (CLI) References* available on Powerlink.

EMC recommends that you configure SRM and the CLARiiON MirrorView Adapter with vCenter Infrastructure after the mirrors and consistency groups have been configured. Refer to the *VMware SRM Administration Guide* along with the *EMC MirrorView Adapter for VMware Site Recovery Manager Version 1.3 Release Notes* for installation and configuration instructions.

SRM Protection Groups

A *Protection Group* specifies the items you want to transition to the recovery site in the event of a disaster. A Protection Group may specify things such as virtual machines (VMs), resource pools, datastores, and networks. Protection Groups are created at the primary site. Depending on what the SRM will be protecting, you can define the Protection Group using VMs or based on the application being protected (for example, distributed application across multi-VMs). Usually there is a 1-to-1 mapping between a SRM Protection Group and a CLARiiON consistency group. However, if your CLARiiON model does not support the number of devices being protected within a Protection Group, you can create multiple CLARiiON consistency groups for each Protection Group.

Table 5 shows the maximum number of devices allowed per consistency group.

Table 5. Maximum number of sync mirrors and consistency groups

Parameter	CX4-120	CX4-240	CX4-480	CX4-960
Total mirrors per storage system	128	256	512	512
Total mirrors per consistency group	32	32	64	64
Total consistency groups per storage system	64	64	64	64

[1] The text **VMWARE_SRM_SNAP** must be somewhere in this name for the SRA adapter to function properly.

Table 6. Maximum number of async mirrors and consistency groups

Parameter	CX4-120	CX4-240	CX4-480	CX4-960
Total mirrors per storage system	256	256	256	256
Total mirrors per consistency group	32	32	64	64
Total consistency groups per storage system	64	64	64	64

Note: The maximum allowed number of consistency groups per storage system is 64. Both MirrorView/A and MirrorView/S consistency groups count toward the total.

For the maximum number of sync and async mirrors and consistency groups for the CLARiiON CX3 storage systems, please refer to the *EMC CLARiiON Open Systems Configuration Guide* available on Powerlink.

SRM recovery plan

The SRM recovery plan is the list of steps required to switch operation of the data center from the protected site to the recovery site. Recovery plans are created at the recovery site, and are associated with a Protection Group or multiple Protection Groups created at the protected site. More than one recovery plan may be defined for a Protection Group if different recovery priorities are needed during failover. The purpose of a recovery plan is to ensure priority of a failover. For example, if a database management server needs to be powered on before an application server, the recovery plan can start the database management server, and then start the application server. Once the priorities are established, the recovery plan should be tested to ensure the ordering of activities has been properly aligned for the business to continue running at the recovery site.

Testing the SRM recovery plan

Once the SRM recovery plan is created, it is important to test that the plan performs the operations expected. A recovery plan is shown in Figure 29. To test the plan, click the **Test** button on the menu bar.

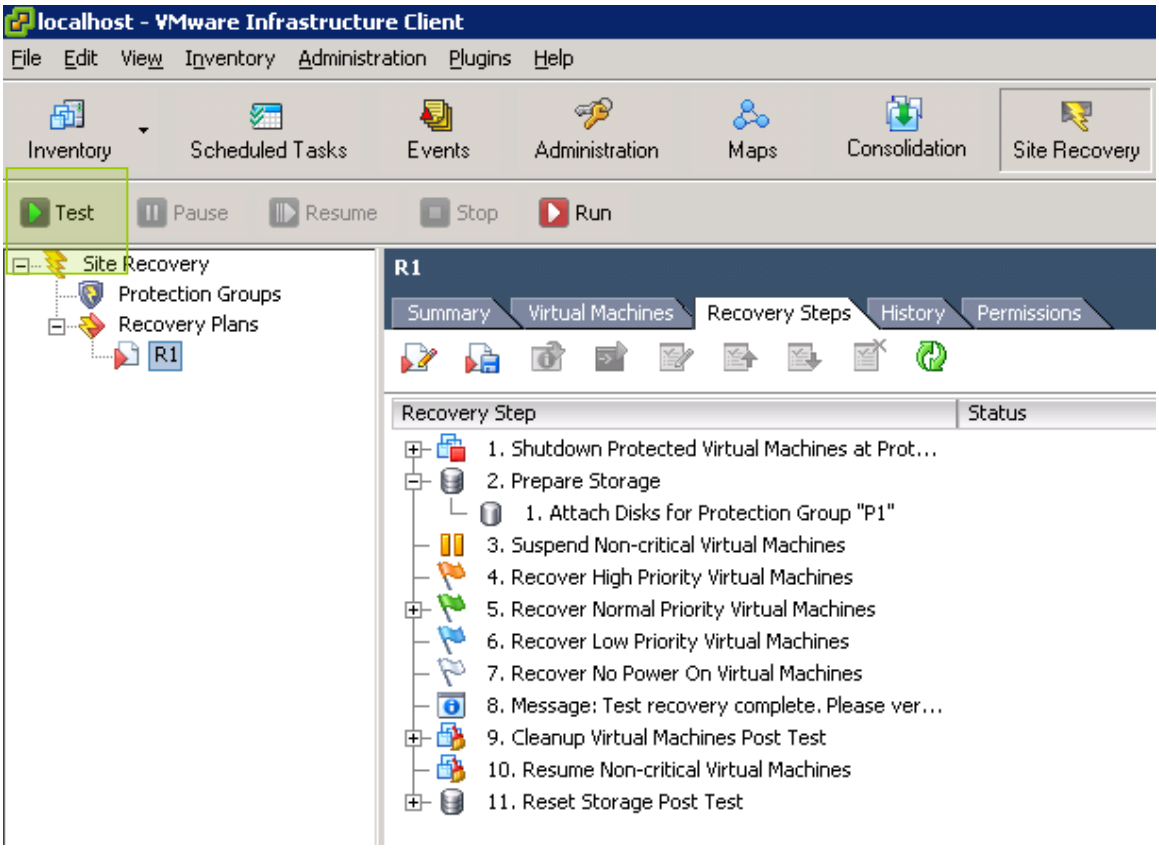


Figure 29. SRM recovery plan

During this test, you would see the following events occur:

1. Production VMs are still up and running
2. CLARiiON SnapView sessions are created and activated against the snapshots created above
3. All resources created within the SRM Protection Group carry over to the recovery site
4. VMs power on in the order defined within the recovery plan

Once all the VMs are powered on according to the recovery plan, SRM will wait for the user to verify that the test works correctly. You verify this by opening a console for the VM started at the recovery site and checking the data. After checking your data, click the **Continue** button, and the environment will revert back to its original production state. For more information concerning SRM recovery plans and Protection Groups, please see the *VMware SRM Administration Guide*.

Executing an SRM recovery plan

Executing an SRM recovery plan is similar to testing the environment with the following differences:

- Execution of the SRM recovery plan is a one-time activity, while running an SRM Test can be done multiple times without user intervention.
- SnapView snapshots are not involved during an executed SRM recovery plan.
- The MirrorView secondary copies are promoted as the new primary LUNs to be used for production operation.
- After executing a recovery plan manual steps are needed to resume operation at the original production site.

You should execute a SRM recovery plan only in the event of a declared disaster, to resume operation at the recovery site.

Failback scenarios

The nature of the disaster, and which components of the data center infrastructure are affected, will dictate what steps are necessary to restore the original production data center. For details on how to address different failback scenarios for MirrorView, please see the white paper *MirrorView Knowledgebook* on Powerlink. For details on how to address these failback scenarios with the MirrorView SRA, please see the *EMC MirrorView Adapter for VMware Site Recovery Manager Version 1.3 Release Notes*.

Navisphere's new VM-aware feature

Navisphere's VM-aware feature automatically discovers virtual machines managed under VMware vCenter Server and provides end-to-end, virtual-to-physical mapping information The VM-aware Navisphere feature, available with FLARE 29, allows you to quickly map from VM to LUNs, or from LUN to VMs. This feature imports ESX-server file system and VM-device mapping information, and is only available in CX4 storage systems running FLARE release 29 or later. The CLARiiON talks directly to the ESX or VCenter APIs, and the communication is out-of-band via IP. This feature is supported with ESX 4, ESX 3.5, and ESXi servers and vCenter version 2.5 and later.

To get detailed information about the VMware environment, use the **Import Virtual Server** option in Navisphere Task Bar and enter user credentials for ESX or vCenter. The **Storage group** tab will display VMFS datastore name and Raw Mapped LUN information as shown in Figure 30.

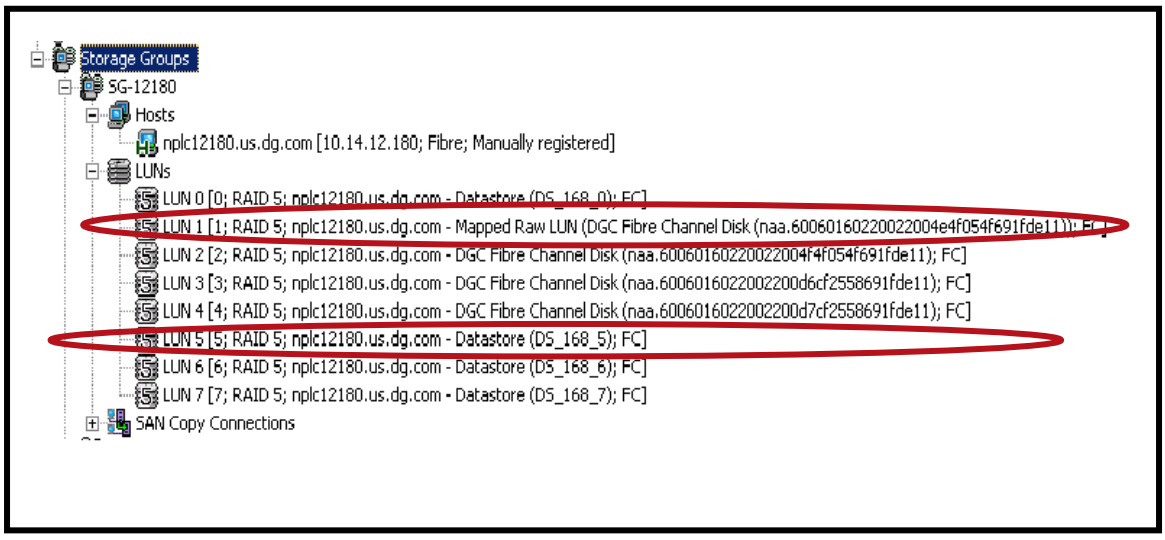


Figure 30. ESX 4 filesystem information display

The **ESX host** dialog box has a **Virtual Machines** tab that lists the virtual machines on the ESX server. To view Guest host name, IP address, and operating system information, VMTools must be installed, configured, and running on the VM.

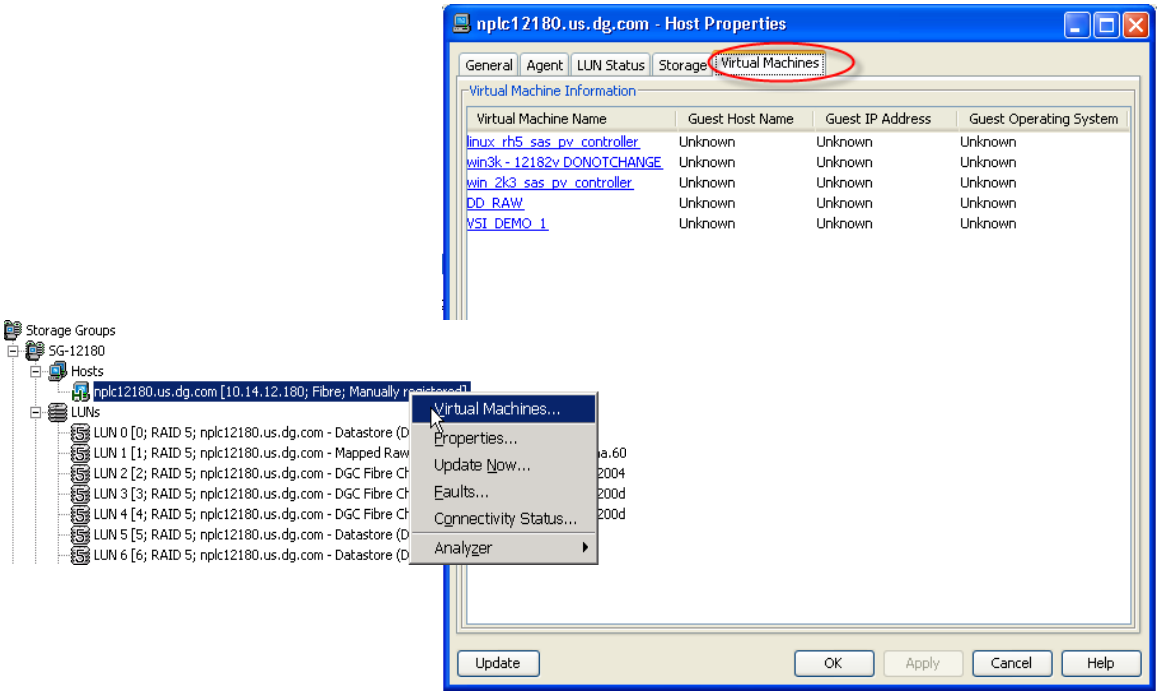


Figure 31. Virtual Machine tab available under ESX server

Clicking on one of the virtual machines in Figure 32 opens the **Virtual Machine Properties** dialog box shown in Figure 33. You can see which LUNs have been assigned to this virtual machine and how they have been configured. In this example, the VM configuration (.vmx) file location, virtual disk properties (thick or thin), and raw mapped volume information are listed for the virtual machine. Allocated *and* consumed capacities are displayed for thin virtual disks.

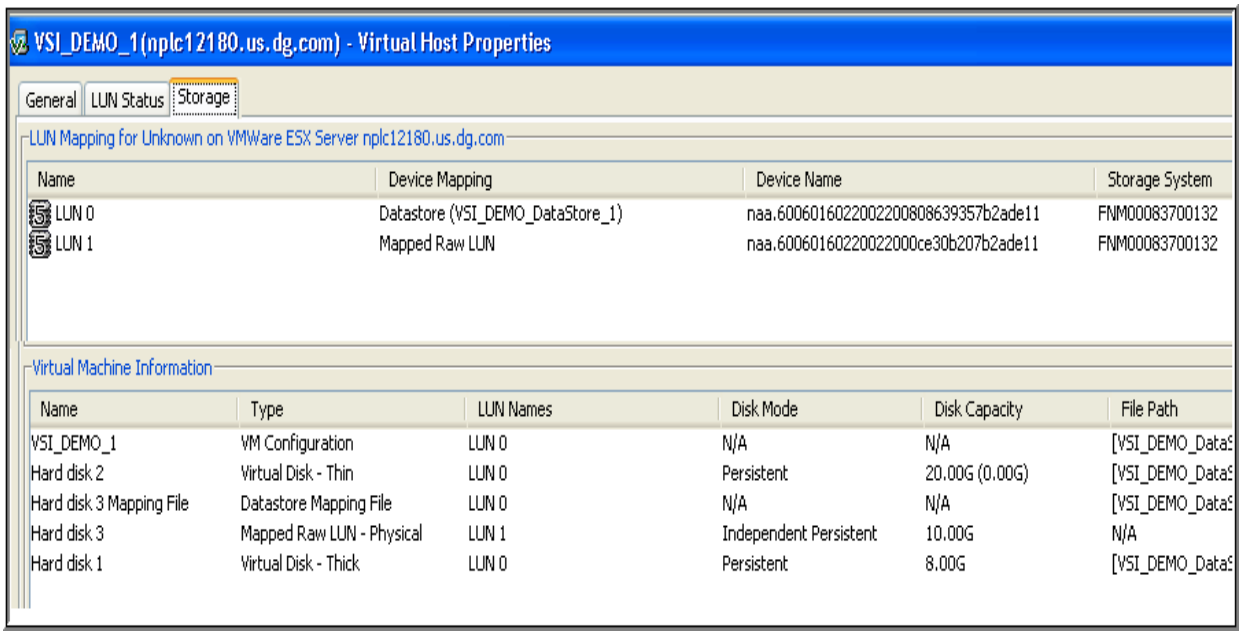


Figure 32. Virtual Machine property dialog box

In addition, a new report called the **Virtual Machine report** can now be generated with the Reporting wizard on the Navisphere Task Bar. This report gives you an overall picture of the LUN-to-VM device mapping.

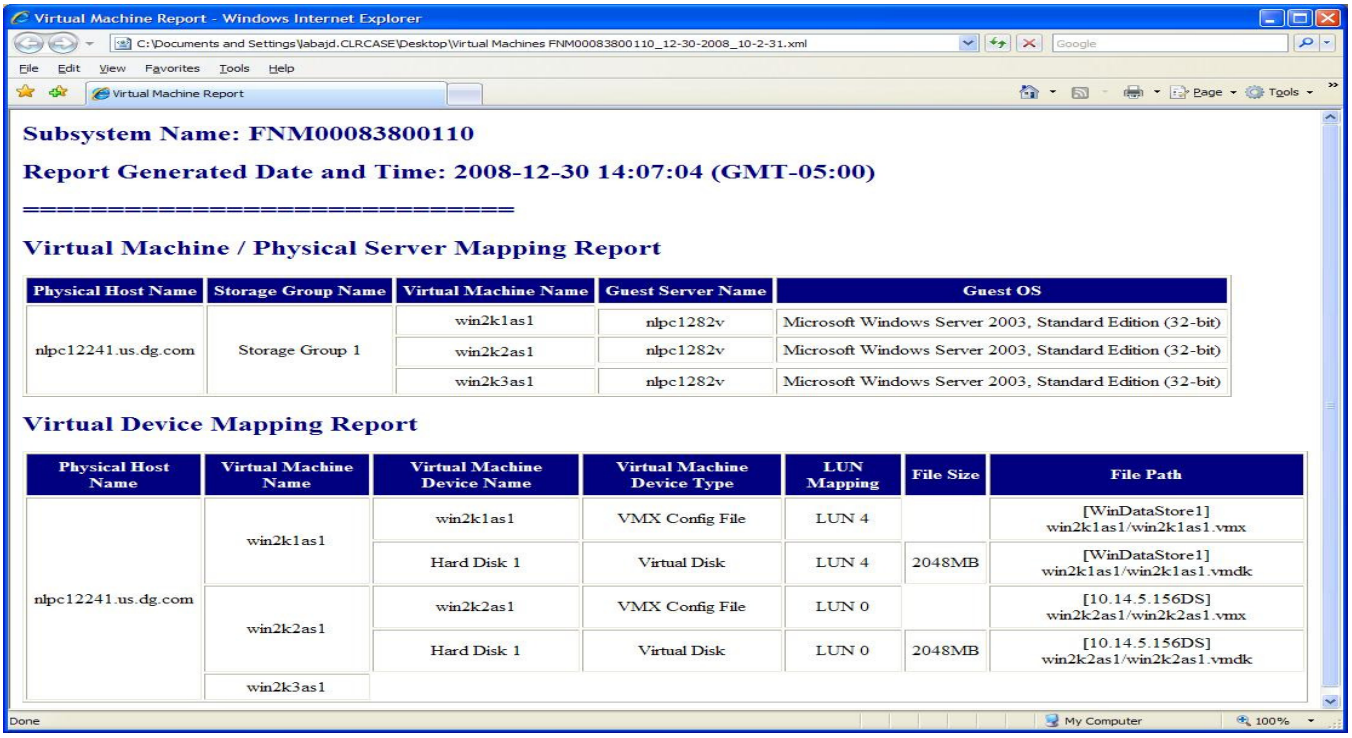


Figure 33. Virtual Machine report

The search function in FLARE release 29 allows you to search for a given virtual machine. When the desired virtual machine is found, you can go directly to the **Virtual Machine Properties** dialog box and get detailed information about the LUNs and their usage on that virtual machine as shown in the next figure.

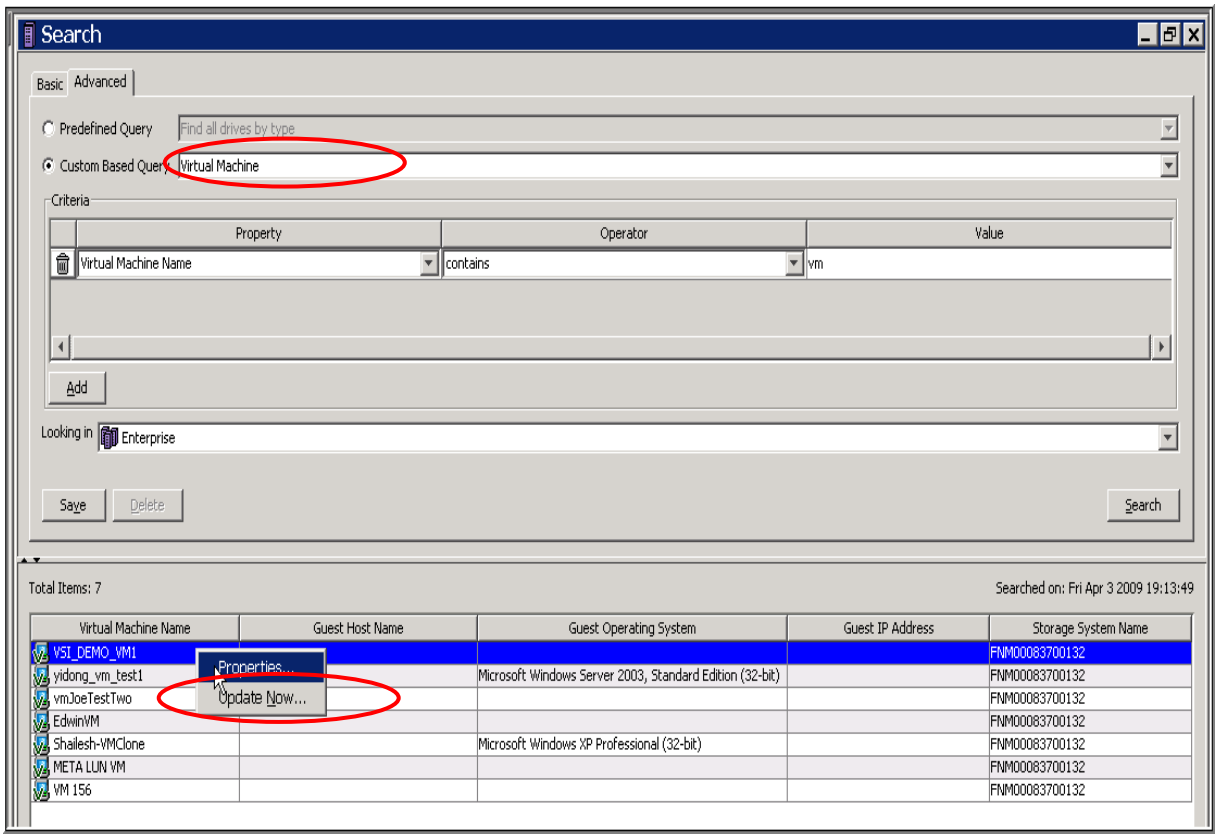


Figure 34. Search for a virtual machine using the "Advanced Search" function in Navisphere

Use Cases

Issue with VM performance -- Find out which LUN a virtual machine is using.

Before VM-aware Navisphere	**After VM-aware Navisphere**
If using VMFS: In the VMware Client GUI, open the **VM "Edit Setting"** dialog box, and find the datastore used by the VM hard disk. Find the ESX device name for that datastore. If using RDM: Use the RDM device to find the datastore containing the RDM mapping file. Then, find the ESX device name for the datastore. Next, in Navisphere, find the LUNs with that ESX device name. Assuming 10 ESX servers each containing 10 VMs, this procedure would require almost 60 clicks using both Navisphere Manager and vCenter	In Navisphere, search for the VM by name or IP. In the **Virtual Machine Properties** dialog box, find which LUNs the VM is using. Assuming 10 ESX servers each containing 10 VMs, this procedure would require 15 clicks using Navisphere Manager

Validate changes made to the VMware environment -- Find and record all VM to all LUNs mapping.

Before VM-aware Navisphere	After VM-aware Navisphere
Generate an ESX device to CLARiiON mapping report. Then the VMware admin must perform multiple steps in VCenter to calculate which VM disks uses which LUN, and put the two reports together side by side. They could also maintain updated Excel spreadsheets for this mapping information. Assuming 10 ESX servers each containing 10 VMs, this procedure would require almost 500 clicks using both Navisphere Manager and vCenter	Generate an end-to-end mapping of VM disk to CLARiiON LUN report. This report lists all VM-to-LUN mappings, and you can examine existing VM storage allocation. The tree view in this report shows the vMotion storage and vMotion changes. You can also use this report to perform VM storage planning for performance, security compliance, and high availability. Assuming 10 ESX servers each containing 10 VMs, this procedure would require 10 clicks using both Navisphere Manager

Capacity Planning (For ESX 4.0 and ESX 4i) -- Determine how much storage has been overcommitted and used by VMs for a given LUN to ensure VMs don't run out of space.

Before VM-aware Navisphere	After VM-aware Navisphere
If multiple virtual disks with a thin virtual disk option are created on a given LUN, you must go through each virtual machine and find its capacity allocation. First, in Navisphere, you need to find the VMware ESX server and ESX device name for the given LUN in Navisphere. Then, in the VMware Client GUI, find the datastore with this ESX device name, if any. For each datastore (VMFS), find all the VMs using this datastore, and find the LUN usages and capacity allocation for hard disks in each VM. Also, you need to browse the VMFS datastore to determine the consumed capacity of the virtual disks on a given datastore. Assuming 10 ESX servers each containing 10 VMs, this procedure would require almost 60 clicks using both Navisphere Manager and vCenter.	In Navisphere, you can simply search for the LUN by name to see what ESX servers and VMs (hard disks) are affected in the **LUN Properties** dialog box. By expanding and selecting all VMs, you can figure out the total promised space to the VMs on this LUN and the total committed space by the VMs on this LUN. **Note:** If you need information about all VMs overcommitted for all LUNs, use the VM-aware Navisphere interface to run two (LUN and virtual machine) reports. Assuming 10 ESX servers each containing 10 VMs, this procedure would require 15 clicks using Navisphere Manager.

EMC Storage Viewer

EMC Storage Viewer is a feature for managing EMC storage devices (Symmetrix and CLARiiON) and storage systems within the VMware vCenter management interface.

The following are requirements for running EMC Storage Viewer in CLARiiON environments:

- Install Solutions Enabler software
  - Ensure you have a license file for Solutions Enabler

- Authenticate your CLARiiON storage system using the symcfg command
- Discover your CLARiiON storage system using the symcfg discover -clariion

- Install the Storage Viewer plug-in
- Navisphere CLI (recommended)
- VMware vCenter 2.5 or later is needed

Once the software stack above has been installed and the CLARiiON storage system has been discovered by Solutions Enabler, enable the Storage Viewer plug-in using the **Managed Plugin** tab within vCenter.

After the Storage Viewer plug-in is enabled, the **EMC Storage** tab is visible and displays detailed information about the CLARiiON storage system, as shown in the next two figures.

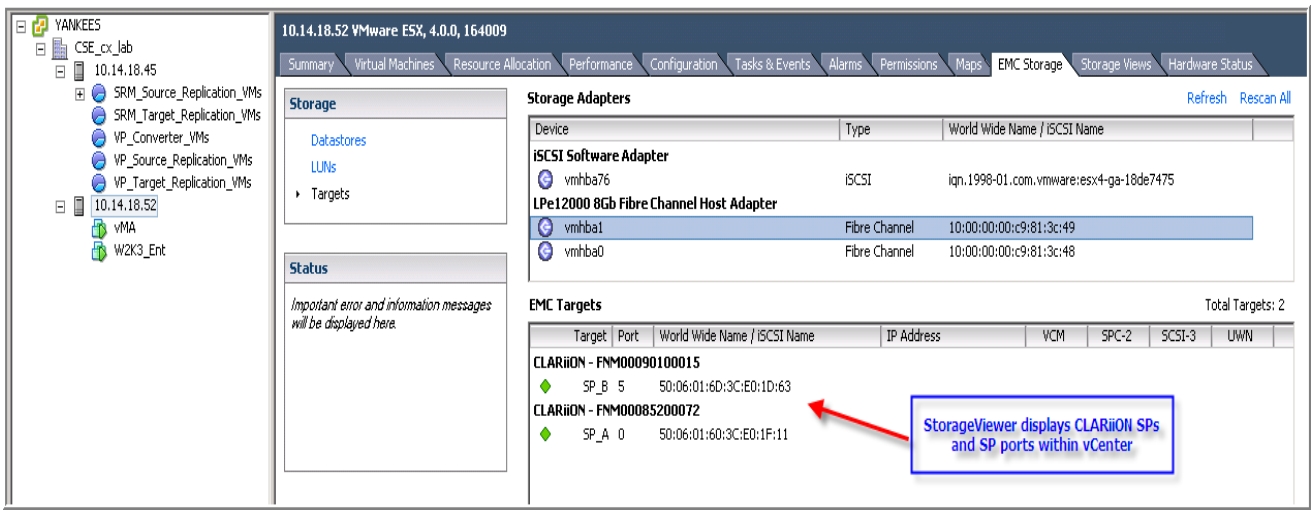


Figure 35. CLARiiON SP and ports information within VMware vCenter

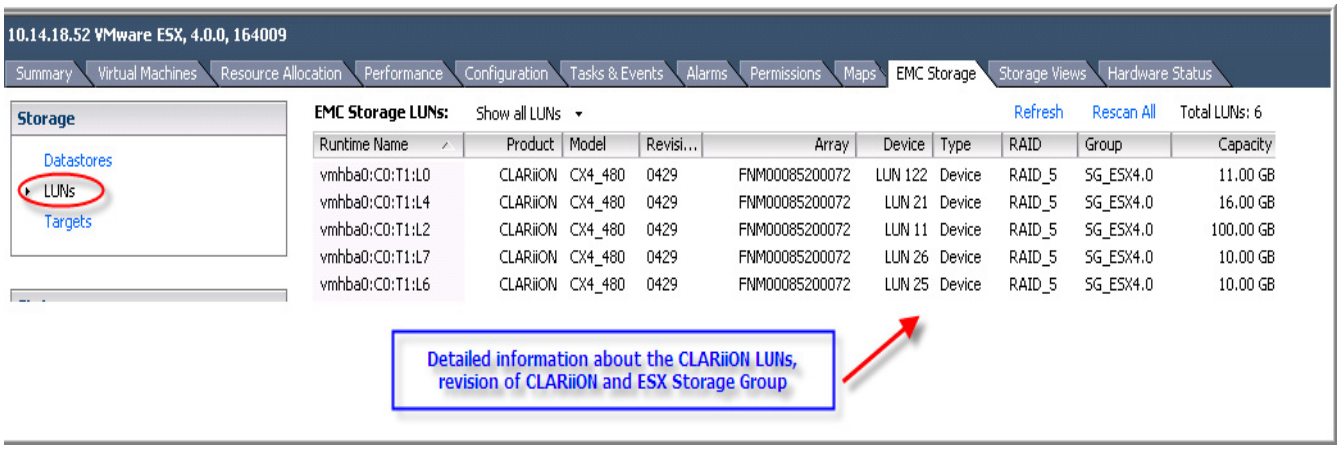


Figure 36. CLARiiON detailed LUN information visible within VMware vCenter

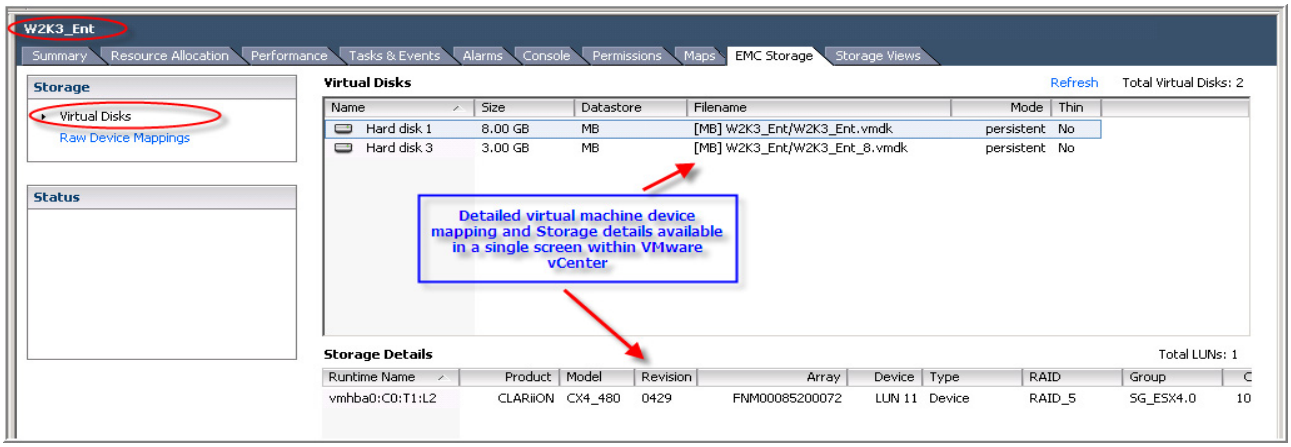


Figure 37. Virtual machine and corresponding CLARiiON detailed LUN information visible within VMware vCenter

For more detailed information on installing and configuring EMC Storage Viewer, see the *Using the EMC Storage Viewer for Virtual Infrastructure Client* white paper available on Powerlink.

Conclusion

EMC CLARiiON and VMware technologies provide the complete Information Lifecycle Management solutions that customers need to consolidate their storage and servers at a low cost. Tools like VMotion, when used with CLARiiON storage, provide online migration of server application workloads without any downtime. VMware HA and Distributed Resource Scheduling coupled with CLARiiON high availability and performance provide reliable and cost-effective solutions. Clustering of virtual machines within the box provides protection against software errors within the cluster.

VMware provides virtualization at the server level while CLARiiON provides protection, performance, and backup at the disk level. Both technologies complement each other, with the high level of functionality and features they provide, to satisfy customer needs.

References

The following documents and resources can be found on Powerlink, EMC's password-protected extranet for partners and customers:

- *EMC Navisphere Manager Administrator's Guide*
- *EMC SnapView for Navisphere Administrator's Guide*
- *EMC SAN Copy for Navisphere Administrator's Guide*
- *EMC MirrorView/Synchronous for Navisphere Administrator's Guide*
- *EMC MirrorView/Asynchronous for Navisphere Administrator's Guide*
- *VMware ESX Server using EMC CLARiiON Storage Systems TechBook*
- *Host Connectivity Guide for VMware ESX Server Version 2.x*
- E-Lab Navigator

The following documents and resources can be found on VMware.com:

- VMware resource documents
  http://www.vmware.com/support/pubs/vs_pages/vsp_pubs_esxi40_e_vc40.html
- Fibre Channel SAN Configuration Guide
  http://www.vmware.com/pdf/vsphere4/r40/vsp_40_san_cfg.pdf

- iSCSI SAN Configuration Guide
  http://www.vmware.com/pdf/vsphere4/r40/vsp_40_iscsi_san_cfg.pdf

Appendix A: Copying data from a VMFS to RDM volume

The export/import CLI command can be used at the ESX Server level to copy data from a VMFS volume to a raw device mapping volume. Consider an example;

The virtual machine resides on a VMFS volume and has the virtual disk name `test.vmdk`. The data on this virtual machine needs to be copied to a RDM volume. Following is an outline of the steps required to move the data.

Store the data (test.vmdk) that resides on a VMFS volume (for example, vmfsprod) to a temporary location, say vmimages, using the vmkfstools export function:

For ESX 2.5.x, execute the following command:

```
vmkfstools –e /vmimages/test1.vmdk /vmfs/vmfsprod/test.vmdk
```

For ESX 3.x, execute the following command:

```
vmkfstools –e /vmimages/test1.vmdk
/vmfs/volumes/vmfsprod/test.vmdk
```

Create a raw device mapping on `vmhba0:0:1:0`, which is a CLARiiON LUN, and the mapping file called `rdm.vmdk` that resides on a VMFS volume (for example, VMFS1):

For ESX 2.5.x, execute the following command:

```
vmkfstools –r vmhba0:0:1:0 /vmfs/vmfs1/rdm.vmdk
```

Import the data from `vmimages` to the mapping file residing on a VMFS (for example, vmfs1), which points to the RDM volume:

```
vmkfstools –i /vmimages/test1.vmdk /vmfs/vmfs1/rdm.vmdk
```

You may see geometry errors after this command is executed. Ignore these errors.

For ESX 3.x, execute the following command:

```
vmkfstools –i /vmimages/test1.vmdk –d
rdm:/vmfs/devices/disks/vmhba0:0:1:0 /vmfs/volumes/vmfs1/rdm.vmdk
```

Note: The import command for ESX 3.x creates a raw device mapping volume and imports the data from VMFS volume. Assign the RDM volume to the virtual machine. Power on the virtual machine to ensure the data on the virtual machine is intact.

Appendix B: Using vm-support on VMware ESX Server

VM support is the command tool used to aid in diagnostics and/or troubleshooting of the ESX server. This service tool is supported on ESX 4.0, 3.x, and 2.x. For VMware ESXi, use VI Client's Export Diagnostics Data option to get vm-support files.

The following procedure outlines the steps executed on the ESX 4.0/3.x service console. Enter the **vm-support** command on the ESX service console. This script generates a .tgz file in the current directory. Extract this file using the following command:

```
tar –zxvf "Vm-support file name"
```

Note that WinZip cannot be used to extract vm-support script output. You have to have a Linux machine to extract these files. Once these files get extracted, a folder with the version of vm-support is created.

Important files to look at within this folder from the storage point of view are as follows:

- /tmp/vmware_xxx.txt – ESX version and patch information
- /var/log/messages – For hardware BIOS versions
- /tmp/chkconfig.*.txt – Confirm naviagent is installed
- /proc/scsi/lpfc or /proc/scsi/qla_2xx – HBA driver versions
- /tmp/esxcfg-swiscsi – Software iSCSI initiator information for ESX 4.0/3.x only
- /tmp/esxcfg-mpath – ESX path information for ESX 4.0/3.x
- /tmp/vmkmultipath – ESX path information for ESX 2.x

Additional information related to the vmkernel configuration can be found at:

- vmkpcidivy – Device configuration information available only on ESX 2.x systems
- /proc/vmware/config – For additional SCSI, disk, and file system configuration information
- /home/vmware/ – Contains VM configuration files information
- var/log/vmkernel and /var/log/vmkernel.1 – For additional troubleshooting information