EMC²
where information lives®

Using EMC CLARiiON Storage with VMware vSphere and VMware Infrastructure

Version 4.0

- Layout and Configuration
- Cloning and Backup
- Disaster Restart and Recovery

Sheetal Kochavara

Part number h2197.5

Contents

Preface

Chapter 1 Introduction to VMware Technology

VMware vSphere and VMware Infrastructure virtualization platforms 14
VMware vSphere and VMware Infrastructure data centers 18
Distributed services in VMware vSphere and VMware Infrastructure 30
Backup and recovery solutions with VMware vSphere and VMware Infrastructure 35
VMware Data Recovery 35
VMware Consolidated Backup 37
VMware vCenter Site Recovery Manager 39
Key benefits of VMware SRM 40
VMware View 42
Key benefits of VMware View 42
Components of the VMware View solution 43
VMware vCenter Converter 45
Migration with vCenter Converter 46

Chapter 2 EMC Foundation Products

Overview 48
EMC CLARiiON 52
EMC CLARiiON Navisphere Management Tools 54
Manager 54
Command line interface (CLI) 55
Host utilities 56

Analyzer 56
CLARiiON metaLUNs 57
Stripe expansion 57
Concatenate expansion 58
CLARiiON Virtual LUN technology 61
Virtualization-Aware Navisphere 62
Virtual Provisioning 69
Fully Automated Storage Tiering (FAST) LUN Migrator 73
LUNanalyze 73
LUNassist 75
Navisphere Quality of Service Manager 76
EMC SnapView 77
SnapView clones 77
SnapView snapshots 86
EMC SAN Copy 91
SAN Copy requirements 93
EMC MirrorView 94
Configuring MirrorView 94
MirrorView consistency groups 100
Using Snapshots and clones with MirrorView 101
MirrorView Insight for VMware (MVIV) 101
EMC RecoverPoint 102
CLARiiON splitter support 102
VMware affinity 102
EMC PowerPath 104
EMC Replication Manager 107
EMC StorageViewer 109

Chapter 3 VMware ESX/ESXi and EMC CLARiiON

Configuring VMware ESX version 4, 3.x, and ESXi 116
Configuring swap space 117
Configuring the ESXkernel 117
Persistent binding 121
Multipathing and failover in ESX version 3 and ESX3i 122
Multipathing and failover in ESX version 4 and ESX4i 122
Using CLARiiON with ESX Server version 4, 3.x, and ESXi 123
Fibre HBA driver configuration 123
ESX iSCSI HBA and NIC drivers 124
Adding and removing CLARiiON devices 124
Creating VMFS volumes 127
Tuning ESX iSCSI HBA and NIC 135

Using Navisphere in virtualized environments 136
Navisphere Agent, Server Utility and CLI 136
Integration of host utilities with ESX Server 140
Virtual Provisioning with ESX Server 141
Navisphere QoS with ESX Server 145
Mapping a VMware file system to CLARiiON devices 145
Mapping RDM to EMC CLARiiON devices 150
Optimizing VI infrastructure and CLARiiON 151
Storage considerations for 4, 3.x, and ESXi servers 151
Path management 159

Chapter 4 Cloning of Virtual Machines

Overview 177
Copying virtual machines after shutdown 178
Using SnapView clones with ESX Servers 178
Using SnapView snapshots with ESX servers 188
Copying virtual machines with RDMs using SnapView snapshots 191
Using EMC to copy running virtual machines 193
Using SnapView clones with ESX servers 193
Using SnapView snapshots with ESX servers 197
Transitioning disk copies to cloned virtual machines 200
Cloning VMs on VMware file systems in VMware Infrastructure 3 200
Cloning VMs on VMware file systems (VMFS-3) with VMware vSphere 4 212
Cloning virtual machines using RDM in VMware vSphere 4 and Virtual Infrastructure 3 environments .. 219
Choosing a VM cloning methodology 221

Chapter 5 Backup and Restore of Virtual Machines

Recoverable versus restartable copies of data 225
Using recoverable disk copies 225
Using restartable disk copies 226
Backing up using copies of Virtual Infrastructure data 227
Using the ESX Server version 3.x service console 227
Using the ESX server version 4.x service console 236
Using cloned VMs in VMware vSphere 4 and VMware Infrastructure 3 240
Using RDM 240
Restoring VMs data using disk-based copies 242

SnapView copies for VMs with VMFS-hosted virtual disks 242
Restoring all virtual machines hosted on VMware file system with ESX version 4 251
SnapView copies for VMs with RDMs 252
Using backup-to-disk copies 253

Chapter 6 VMware ESX/ESXi in a Disaster Restart Solution

Integration of guest operating environments with EMC technologies and VMware ESX/ESXi Definitions 261
Dependent-write consistency 261
Disaster restart 261
Disaster recovery 262
Roll-forward recovery 262
Design considerations for disaster recovery 263
Recovery point objective 263
Recovery time objective 263
Operational complexity 264
Source server activity 265
Production impact 265
Target server activity 265
Number of copies of data 266
Distance for the solution 266
Bandwidth requirements 266
Federated consistency 266
Testing the solution 267
Cost 267
Protecting physical infrastructure with Virtual Infrastructure 269
Physical-to-Virtual Infrastructure 270
Remotely managing application data LUNs 271
Business continuity with virtual to virtual infrastructure . 272
Tape-based solutions 272
MirrorView consistency groups 273
MirrorView/S from CLARiiON to CLARiiON 275
MirrorView/A from CLARiiON to CLARiiON 279
Configuring remote sites for VMs using VMFS-3 on Virtual Infrastructure 3 284
Configuring remote sites for VMware Infrastructure 3 and VMware vSphere 4 VMs with RDM 293
Creating SRM Protection Groups at the protected site 303

Chapter 7 Data Vaulting and Migration of VMware Virtual Infrastructure

SAN Copy interoperability with VMware file systems..... 313
SAN Copy interoperability with VMs using RDM 315
Using SAN Copy for data vaulting 316
Data vaulting of VMware file system using SAN Copy..... 317
Data vaulting of VMs configured with RDMs using SAN Copy 322
Transitioning disk copies to cloned virtual machines..... 323
Configuring remote sites for VMs using VMFS-3 for VMware Infrastructure 3 323
Configuring remote sites for VMs using VMFS-3 for vSphere 4.x..... 326
Configuring remote sites for VMware Infrastructure 3 and VMware vSphere 4 VMs with RDM 328
SAN Copy for data migration from CLARiiON arrays..... 330
Migration of a VMware file system in ESX version 3 . 330
Migration of a VMware file system in ESX version 4.. 337
Migration of devices used as RDM 337
SAN Copy for data migration to CLARiiON arrays..... 339
Migration of a VMware file system in ESX version 3 . 341
Migration of a VMware file system in ESX version 4.. 342

Appendix A Nondisruptive Expansion of a MetaLUN

Introduction 346
Expanding CLARiiON LUNs 347
Growing VMware file system 3 in Virtual Infrastructure 3 348
Growing VMware file system 3 in vSphere 4 353
Growing RDMs in vSphere 4 and Virtual Infrastructure 3 356

Preface

As part of an effort to improve and enhance the performance and capabilities of its product lines, EMC periodically releases revisions of its hardware and software. Therefore, some functions described in this document may not be supported by all versions of the software or hardware currently in use. For the most up-to-date information on product features, refer to your product release notes.

This TechBook describes how the VMware ESX/ESXi hosts work with EMC CLARiiON storage systems and software technologies. This document focuses on the integration of VMware ESX/ESXi hosts with CLARiiON disk arrays, EMC MirrorView, EMC SnapView, and EMC SAN Copy.

Audience

This document is part of the CLARiiON documentation set, and is intended for use by storage administrators, system administrators, and VMware ESX/ESXi hosts administrators. This document can also be used by individuals who are involved in acquiring, managing, or operating EMC CLARiiON storage arrays and host devices.

Readers of this document are expected to be familiar with the following topics:

- EMC CLARiiON system operation
- EMC MirrorView, EMC SnapView, EMC SAN Copy, and Navisphere
- VMware ESX/ESXi hosts operation

Related documentation

These related documents are available on EMC Powerlink:

- *EMC CLARiiON Integration with VMware ESX Server* white paper
- *EMC Replication Manager with CLARiiON and VMware ESX server* white paper
- *Implementing Virtual Provisioning on EMC CLARiiON and Celerra with VMware Infrastructure* white paper
- *Using Celerra Storage with VMware vSphere and VMware Infrastructure* TechBook

Conventions used in this document

EMC uses the following conventions for special notices.

Note: A note presents information that is important, but not hazard-related.

CAUTION

A caution contains information essential to avoid data loss or damage to the system or equipment.

IMPORTANT

An important notice contains information essential to operation of the software or hardware.

WARNING

A warning contains information essential to avoid a hazard that can cause severe personal injury, death, or substantial property damage if you ignore the warnTypographical conventions

EMC uses the following type style conventions in this document:

Normal	Used in running (nonprocedural) text for: • Names of interface elements (such as names of windows, dialog boxes, buttons, fields, and menus) • Names of resources, attributes, pools, Boolean expressions, buttons, DQL statements, keywords, clauses, environment variables, functions, utilities • URLs, pathnames, filenames, directory names, computer names, filenames, links, groups, service keys, file systems, notifications
Bold	Used in running (nonprocedural) text for: • Names of commands, daemons, options, programs, processes, services, applications, utilities, kernels, notifications, system calls, man pages Used in procedures for: • Names of interface elements (such as names of windows, dialog boxes, buttons, fields, and menus) • What user specifically selects, clicks, presses, or types
Italic	Used in all text (including procedures) for: • Full titles of publications referenced in text • Emphasis (for example a new term) • Variables
`Courier`	Used for: • System output, such as an error message or script • URLs, complete paths, filenames, prompts, and syntax when shown outside of running text
`Courier bold`	Used for: • Specific user input (such as commands)
`Courier italic`	Used in procedures for: • Variables on command line • User input variables
< >	Angle brackets enclose parameter or variable values supplied by the user
[ ]	Square brackets enclose optional values
|	Vertical bar indicates alternate selections - the bar means “or”
{ }	Braces indicate content that you must specify (that is, x or y or z)
...	Ellipses indicate nonessential information omitted from the example

The author of this Techbook

This TechBook was written by Sheetal Kochavara, an employee of EMC with eight years of experience in Systems Engineering.

We'd like to hear from you!

Your feedback on our TechBooks is important to us! We want our books to be as helpful and relevant as possible, so please feel free to send us your comments, opinions and thoughts on this or any other TechBook:

TechBooks@emc.com

1 Introduction to VMware Technology

This chapter presents these topics:

- VMware vSphere and VMware Infrastructure virtualization platforms 14
- Distributed services in VMware vSphere and VMware Infrastructure 30
- Backup and recovery solutions with VMware vSphere and VMware Infrastructure 35
- VMware vCenter Site Recovery Manager 39
- VMware View 42
- VMware vCenter Converter 45

VMware vSphere and VMware Infrastructure virtualization platforms

VMware vSphere and VMware Infrastructure are two virtualization platforms from VMware. VMware Infrastructure 3.5 is the previous major release of the platform, whereas VMware vSphere 4 is the next generation of the platform that VMware recently released. VMware vSphere and VMware Infrastructure virtualization platforms consist of various components including ESX/ESXi hosts and VMware vCenter Server. In addition, VMware vSphere and VMware Infrastructure offer a set of services like distributed resource scheduling, high availability, and backup. The relationship between the various components within the VMware vSphere platform is shown in Figure 1 on page 15.

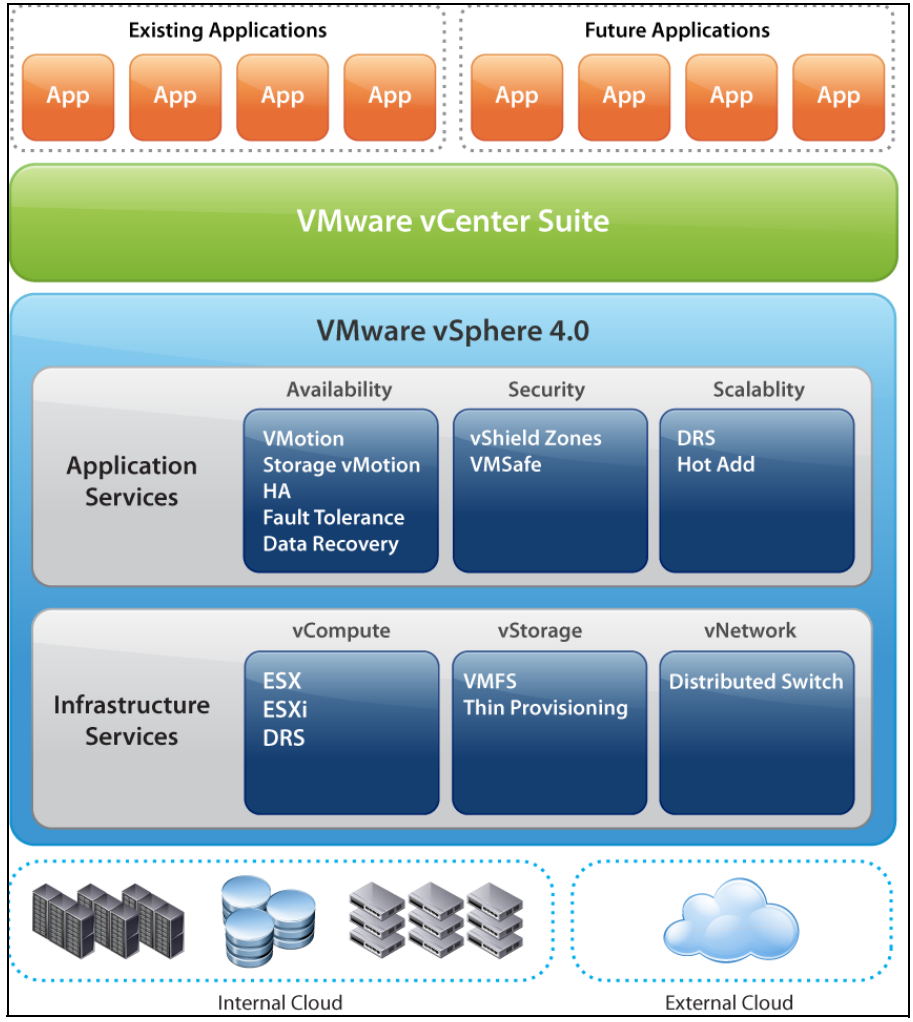


Figure 1 VMware vSphere architecture

ESX and ESXi — ESX and ESXi are the foundation to deliver virtualization-based distributed services to IT environments. As a core building block of VMware vSphere and VMware Infrastructure, both ESX and ESXi form a production-proven virtualization layer that abstracts processor, memory, storage, and networking resources into multiple virtual machines running side-by-side on the same physical

server. Sharing hardware resources across a large number of virtual machines increases hardware utilization and decreases capital and operating costs.

The two versions of ESX available are:

- **VMware ESX —** Contains a built-in service console, which is installed as the first component and is used to bootstrap the ESX server installation. Using a command line interface, the service console can then be used to configure ESX. ESX is available as an installable DVD-ROM boot image. The service console is a virtual machine consisting of a Red Hat Linux kernel. It runs inside the ESX server and can be used to run local commands or scripts agents within it.
- **VMware ESXi —** VMware ESXi does not contain a service console. It is available in two forms: VMware ESXi Embedded and VMware ESXi Installable. ESXi Embedded is a firmware that is built into a server's physical hardware or as an internal USB drive. ESXi Installable is a software that is available as an installable CD-ROM boot image. The ESXi Installable software can be installed on a server's hard drive or on an external USB drive.

vCenter Server — vCenter Server delivers centralized management, operational automation, resource optimization, and high availability to IT environments. Virtualization-based distributed services provided by VMotion, VMware Distributed Resource Scheduler (DRS), and VMware High Availability (HA) equip the dynamic data center with unprecedented levels of serviceability, efficiency, and reliability. Automated resource optimization with VMware DRS aligns available resources with predefined business priorities while streamlining labor-intensive and resource-intensive operations. Migration of live virtual machines with VMotion makes the maintenance of IT environments nondisruptive. VMware HA enables cost-effective application availability independent of hardware and operating systems.

VMware Virtual Machine — A virtual machine is a representation of a physical machine by software. A virtual machine as an entity exists as a series of files on the disk. For example, there is a file for the hard drives, a file for memory swap space, and for virtual machine configuration. A virtual machine has its own set of virtual hardware (such as RAM, CPU, NIC, and hard disks) upon which an operating system and an application is loaded. The operating system sees a consistent and normalized set of hardware regardless of the actual physical hardware

components. VMware virtual machines use advanced hardware features such as 64-bit computing and virtual symmetric multiprocessing.

VMware vSphere Client and VMware Infrastructure Client — Interfaces that allow administrators and users to connect remotely to vCenter Server or ESX/ESXi from any Windows machine.

VMware vSphere Web Access and VMware Infrastructure Web Access — Web interfaces for virtual machine management and remote console access.

Some of the optional components of VMware vSphere 4 and VMware Infrastructure are:

VMware VMotion — VMware VMotion enables the live migration of running virtual machines from one physical server to another.

VMware Storage VMotion — Storage VMotion enables the migration of virtual machine files from one datastore to another, even across storage arrays, without service interruption.

VMware High Availability (HA) —VMware HA provides high availability for applications running on virtual machines. In the event of a server failure, affected virtual machines are automatically restarted on other production servers with spare capacity.

VMware Distributed Resource Scheduler (DRS) —VMware DRS leverages VMotion to dynamically allocate and balance computing capacity across a collection of hardware resources aggregated into logical resource pools.

VMware Fault Tolerance (FT) — When VMware FT is enabled for a virtual machine, a secondary copy of the original (or primary) virtual machine is created in the same data center. All actions completed on the primary virtual machine are also applied to the secondary virtual machine. If the primary virtual machine becomes unavailable, the secondary machine becomes active and provides continuous availability. VMware Fault Tolerance is unique to VMware vSphere.

vNetwork Distributed Switch — This feature, which is also unique to VMware vSphere, includes a distributed virtual switch, which is created and maintained by vCenter Server and spans many ESX/ESXi hosts, enabling significant reduction of ongoing network maintenance activities and increasing network capacity. This allows virtual machines to maintain consistent network configuration and advanced network features and statistics as they migrate across multiple hosts.

VMware Consolidated Backup (VCB) — This feature provides a centralized facility for agent-free backup of virtual machines with VMware Infrastructure. It simplifies backup administration and reduces the impact of backups on ESX/ESXi performance.

VMware Data Recovery — A backup and recovery product for VMware vSphere environments that provides quick and complete data protection for virtual machines. VMware Data Recovery is a disk-based solution that is built on the VMware vStorage API for data protection and is fully integrated with vCenter Server.

Pluggable Storage Architecture (PSA) — A modular partner plug-in storage architecture that enables greater array certification flexibility and improved array-optimized performance. PSA is a multipath I/O framework that allows storage partners to enable array compatibility asynchronously to ESX release schedules. VMware partners can deliver performance-enhancing multipath load-balancing behaviors that are optimized for each array.

VMware vSphere Software Development Kit (SDK) and VMware Infrastructure SDK — SDKs that provide a standard interface for VMware and third-party solutions to access VMware vSphere and VMware Infrastructure.

vStorage APIs for data protection — This API leverages the benefits of Consolidated Backup and makes it significantly easier to deploy, while adding several new features that deliver efficient and scalable backup and restore of virtual machines.

Like Consolidated Backup, this API offloads backup processing from ESX servers, thus ensuring that the best consolidation ratio is delivered, without disrupting applications and users. This API enables backup tools to directly connect the ESX servers and the virtual machines running on them, without any additional software installation. The API enables backup tools to do efficient incremental, differential, and full-image backup and restore of virtual machines.

VMware vSphere and VMware Infrastructure data centers

VMware vSphere and VMware Infrastructure virtualize the entire IT infrastructure including servers, storage, and networks. VMware vSphere and VMware Infrastructure aggregate these resources and present a uniform set of elements in the virtual environment. With

VMware vSphere and VMware Infrastructure, IT resources can be managed like a shared utility and resources can be dynamically provisioned to different business units and projects.

A typical VMware vSphere or VMware Infrastructure data center consists of basic physical building blocks such as x86 virtualization servers, storage networks and arrays, IP networks, a management server, and desktop clients.

The physical topology of a VMware vSphere data center is illustrated in Figure 2 on page 20.

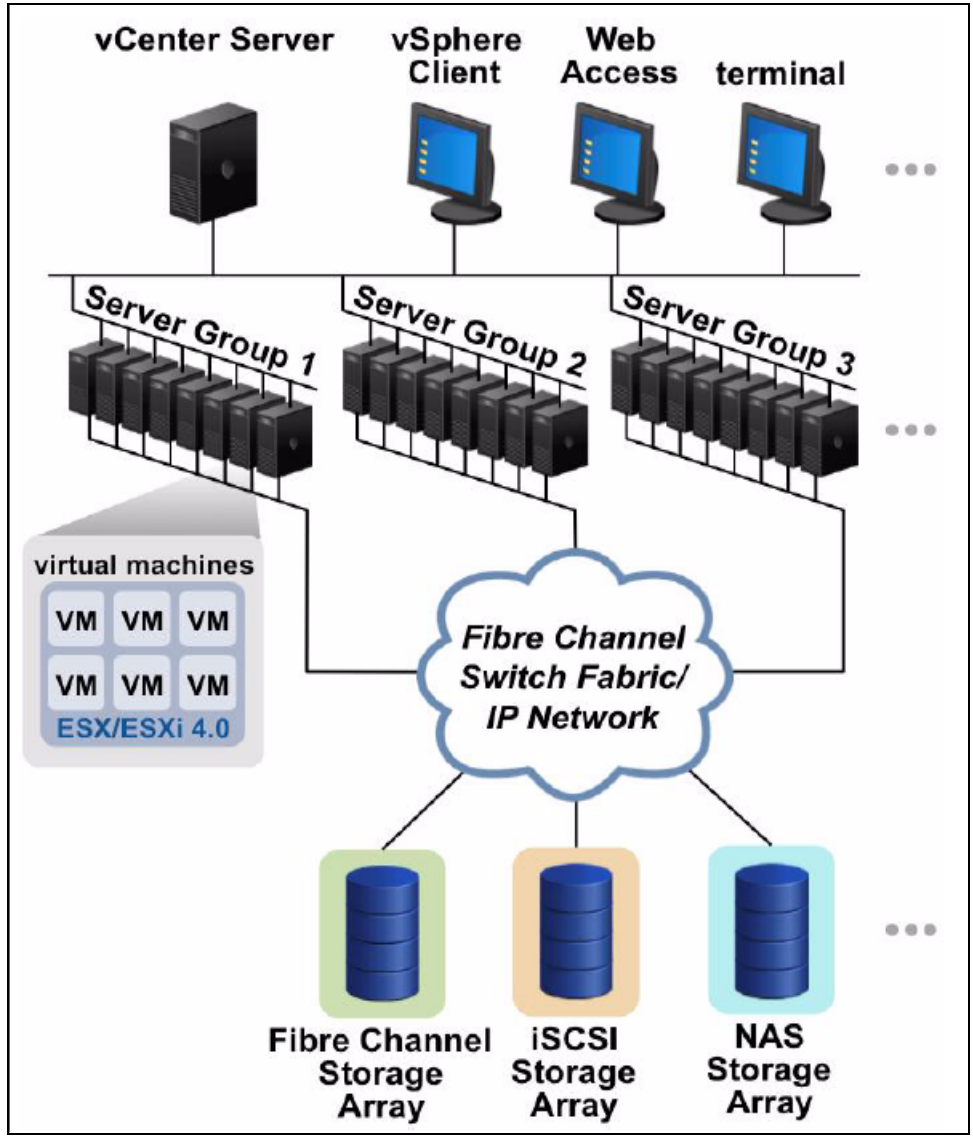


Figure 2 VMware vSphere data center physical topology

Network architecture

The virtual environment provides similar networking elements as the physical world: virtual network interface cards (vNIC), virtual switches (vSwitch), and port groups. VMware vSphere introduced a new type of switch architecture, called vNetwork Distributed Switch, that expands this network architecture.

The network architecture is depicted in Figure 3.

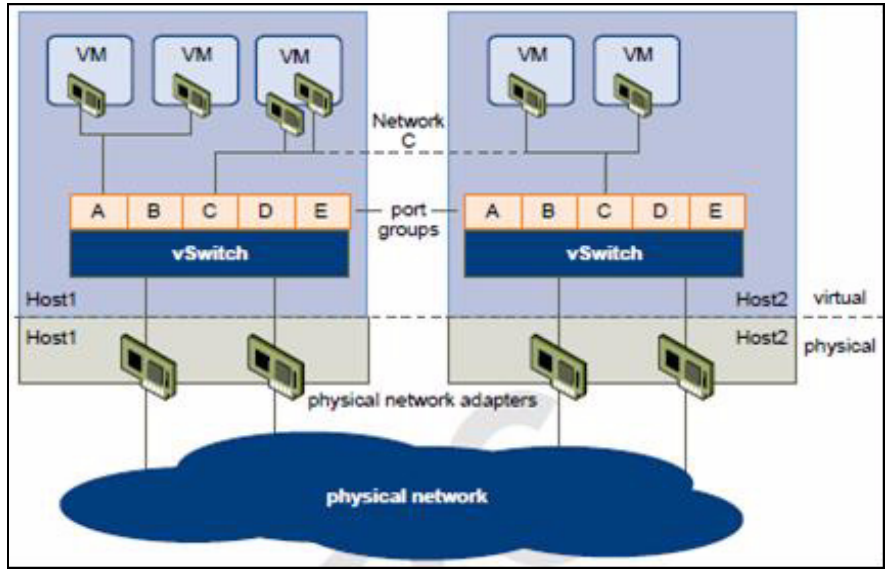


Figure 3 vNIC, vSwitch, and port groups

Like a physical machine, each virtual machine has one or more vNICs. The guest operating system and applications communicate with the vNIC through a standard device driver or a VMware optimized device driver in the same way as a physical NIC. Outside the virtual machine, the vNIC has its own MAC address and one or more IP addresses, and responds to the standard Ethernet protocol in the same way as a physical NIC. An outside agent cannot detect that it is communicating with a virtual machine.

VMware vSphere 4 offers two types of switch architecture: vSwitch and vNetwork Distributed Switch. A vSwitch works like a layer 2 physical switch. Each ESX host has its own vSwitch. One side of the vSwitch has port groups that connect to virtual machines. The other side has uplink connections to physical Ethernet adapters on the server where the

vSwitch resides. Virtual machines connect to the outside world through the physical Ethernet adapters that are connected to the vSwitch uplinks. A vSwitch can connect its uplinks to more than one physical Ethernet adapter to enable NIC teaming. With NIC teaming, two or more physical adapters can be used to share the traffic load or provide passive failover in the event of a physical adapter hardware failure or a network outage. With VMware Infrastructure, only vSwitch is available.

Port group is a unique concept in the virtual environment. A port group is a mechanism for setting policies that govern the network connected to it. A vSwitch can have multiple port groups. Instead of connecting to a particular port on the vSwitch, a virtual machine connects its vNIC to a port group. All virtual machines that connect to the same port group belong to the same network inside the virtual environment, even if they are on different physical servers.

The vNetwork Distributed Switch is a distributed network switch that spans many ESX hosts and aggregates networking to a centralized cluster level. Therefore, vNetwork Distributed Switches are available at the data center level of the vCenter Server inventory. vNetwork Distributed Switches abstract configuration of individual virtual switches and enables centralized provisioning, administration, and monitoring through VMware vCenter Server. Figure 4 on page 23 illustrates a vNetwork Distributed Switch that spans between ESX server hosts.

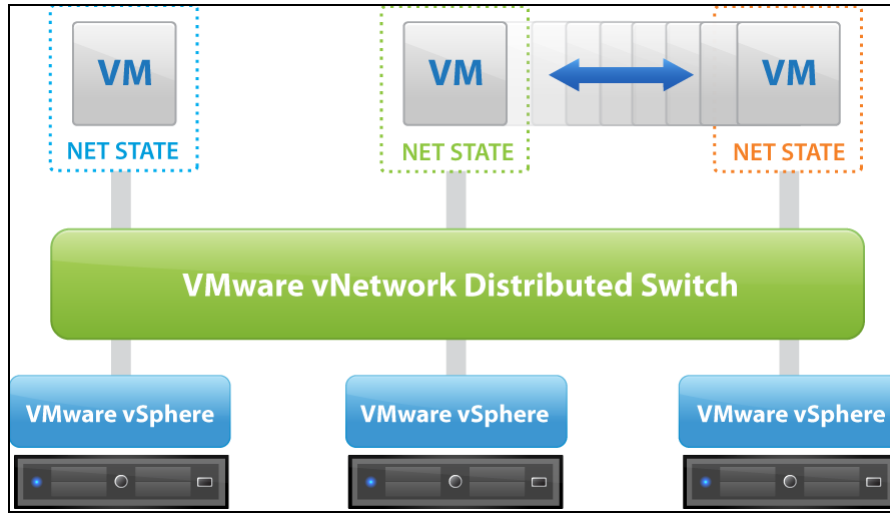


Figure 4 VMware vNetwork Distributed Switch

Storage architecture

The VMware vSphere and VMware Infrastructure storage architecture consists of abstraction layers to manage the physical storage subsystems. The key layer in the architecture is the datastores layer.

Figure 5 on page 24 shows the storage architecture.

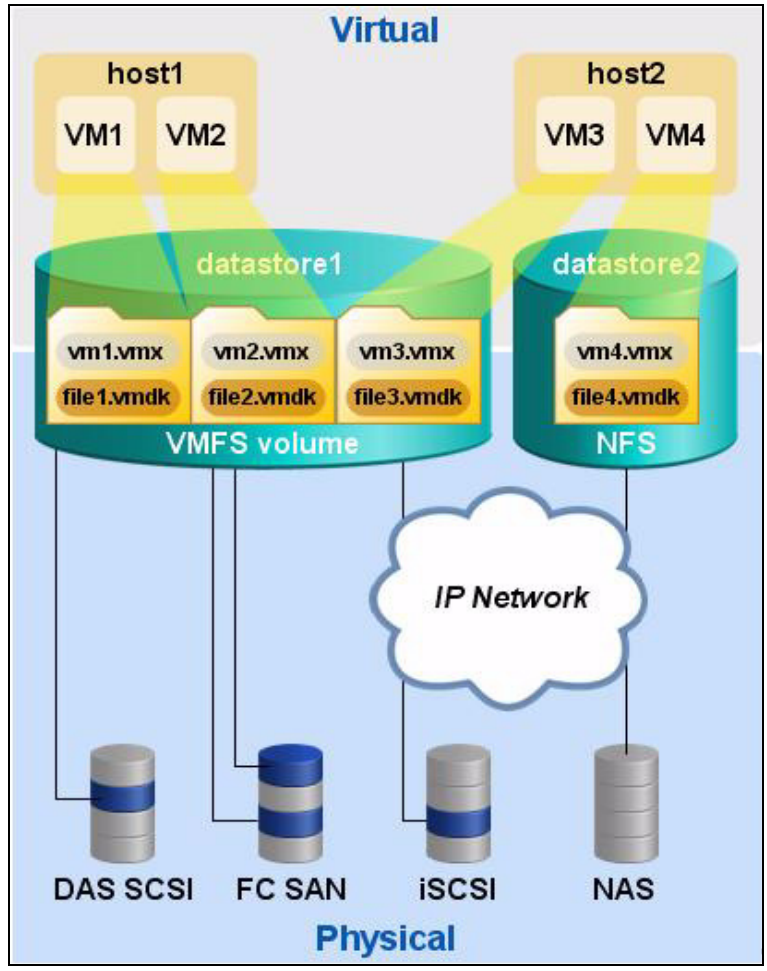


Figure 5 VMware vSphere and VMware Infrastructure storage architecture

A datastore is like a storage appliance that allocates storage space for virtual machines across multiple physical storage devices. The datastore provides a model to allocate storage space to the individual virtual machines without exposing them to the complexity of the physical storage technologies, such as Fibre Channel SAN, iSCSI SAN, direct-attached storage, or NAS.

A virtual machine is stored as a set of files in a datastore directory. A virtual disk, inside each virtual machine, is also a set of files in the directory. Therefore, operations such as copy, move, and backup can be performed on a virtual disk just like with a file. New virtual disks can be hot-added to a virtual machine without powering it down. In such a case, either a virtual disk file (.vmdk) is created in a datastore to provide new storage space for the hot-added virtual disk or an existing virtual disk file is added to a virtual machine.

The two types of datastores available in this storage architecture are vStorage VMFS and NAS. A VMFS datastore is a clustered file system built across one or more physical volumes (LUNs) originating from block storage systems. A NAS datastore is a NFS volume on a file storage system. In this case, the storage is managed entirely by the file storage system.

VMFS datastores can span multiple physical storage subsystems. A single VMFS volume can contain one or more LUNs from a local SCSI disk array on a physical host, a Fibre Channel SAN disk farm, or an iSCSI SAN disk farm. New LUNs added to any of the physical storage subsystems are detected and can be made available to all existing or new datastores. The storage capacity on a previously created VMFS datastore (volume) can be hot-extended by adding a new physical LUN from any of the storage subsystems that are visible to it as long as the VMFS volume extent has not reached the 2 TB minus 1 MB limit. Alternatively, a VMFS volume can be extended (Volume Grow) within the same LUN. With VMware vSphere, this can be done without powering off physical hosts or storage subsystems. If any of the LUNs (except for the LUN which has the first extent of the spanned volume) within a VMFS volume fails or becomes unavailable, only virtual machines that interact with that LUN are affected. All other virtual machines with virtual disks residing in other LUNs continue to function as normal.

Furthermore, a VMFS datastore can be configured to be mapped to a physical volume on a block storage system. To achieve this, the datastore can be configured with virtual disks that map to a physical volume on a block storage system. This functionality of vStorage VMFS is called Raw Device Mapping (RDM). RDM is illustrated in Figure 6 on page 26.

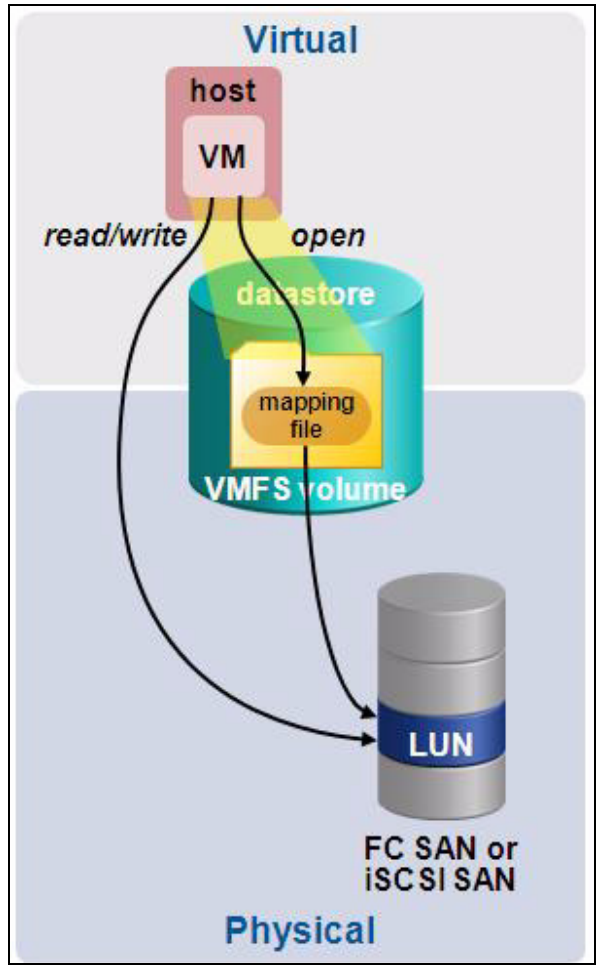


Figure 6 Raw device mapping

With RDM functionality, a virtual machine can be given direct access to a physical LUN in the storage system. This is helpful in various use cases in which the guest OS or the applications within the virtual machine require direct access to the physical volume. One example of such a use case is a physical-to-virtual clustering between a virtual machine and a physical server.

New VMware vSphere 4 storage-related features

The key storage-related features that are new and available with VMware vSphere 4 are:

- **Virtual disk thin provisioning** — VMware vSphere offers an option to create thin provisioned virtual disks when deploying or migrating virtual machines. VMware vCenter Server has also been updated with new management screens and capabilities such as raising alerts, alarms, and improved datastore utilization reports to enable the management of over-provisioned datastores. Virtual disk thin provisioning increases the efficiency of storage utilization for virtualization environments by using only the amount of underlying storage resources needed for that virtual disk. In the past, thin provisioning was the default format for only virtual disks created on NAS datastores in VMware Infrastructure. However, VMware has integrated the management of virtual disk thin provisioning and now fully supports this format for all virtual disks with the release of vSphere. Virtual disk thin provisioning should not be confused with thin provisioning capabilities that an array vendor might offer. In fact, with vSphere, it is even possible to thin provision a virtual disk at the datastore level that resides on a thinly provisioned device on the storage array.
- **Storage VMotion** — This technology performs the migration of the virtual machine while the virtual machine is active. With VMware vSphere, Storage VMotion can be administered through vCenter Server and works across all storage protocols including NFS (in addition to Fibre Channel and iSCSI). In addition, Storage VMotion allows the user to move between different provisioning states. For example, from a thick to thin virtual disk.
- **VMFS Volume Grow** — VMFS Volume Grow offers a new way to increase the size of a datastore that resides on a VMFS volume. It complements the dynamic LUN expansion capability that exists in many storage array offerings today. If a LUN is increased in size, then the VMFS Volume Grow enables the VMFS volume extent to dynamically increase in size as well (up to the standard 2 TB minus 1 MB limit).
- **Pluggable Storage Architecture (PSA)** — In vSphere, leveraging third-party storage vendor multipath software capabilities is introduced through a modular storage architecture that allows storage partners to write a plug-in for their specific capabilities. These modules communicate with the intelligence running in the storage array to determine the best path selection and leverage

parallel paths to increase performance and reliability of the I/O from the ESX to the storage array. Typically the native multipath driver (NMP) supplied by VMware will be used. It can be configured to support round-robin multipath as well. However, if the storage vendor module is available, it can be configured to manage the connections between the ESX and the storage. EMC PowerPath®/VE is an excellent example of such a storage vendor module.

- **Datastore alarms** — Datastore alarms track and warn users on potential resource over-utilization or event conditions for datastores. With the release of vSphere, alarms can be set to trigger on events and notify the administrator when critical error conditions occur.
- **Storage reports and maps** — Storage reports help monitor storage information like datastore, LUNs, virtual machines on datastore, and host access to datastore. Storage maps help to visually represent and understand the relationship between a vSphere datacenter inventory object and the virtual and physical storage resources available for this object. Figure 7 on page 29 shows a storage map that includes both NFS and iSCSI storage resources from EMC Celerra®.

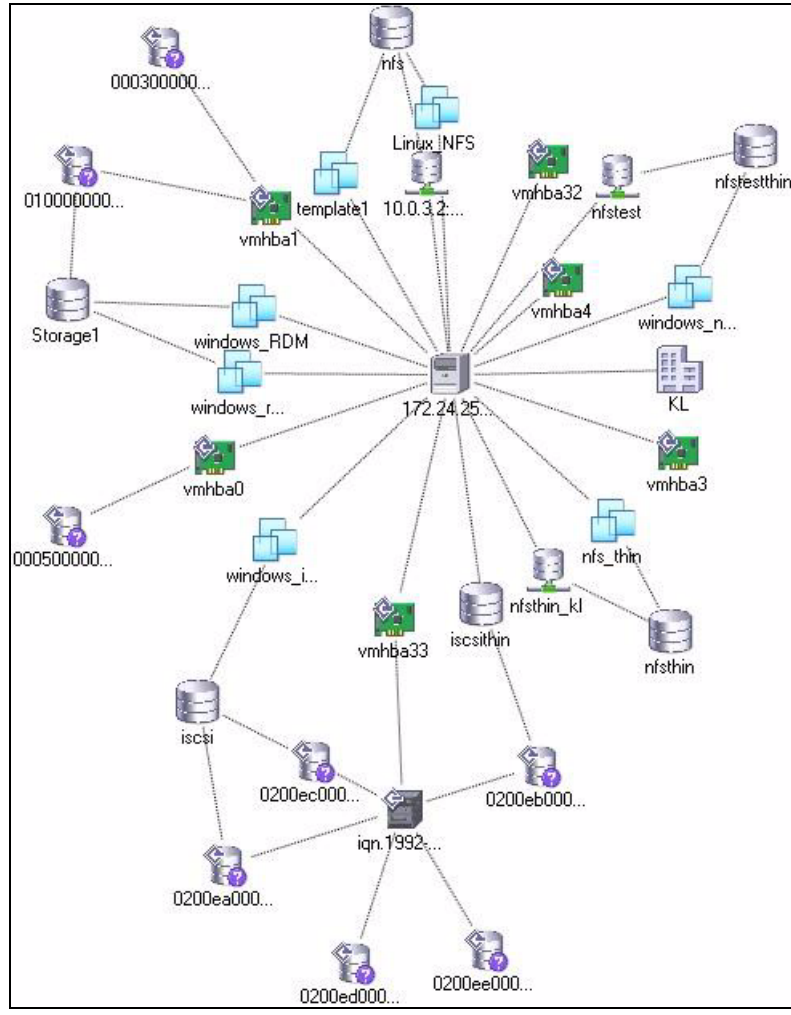


Figure 7 Storage map of vSphere inventory objects

Distributed services in VMware vSphere and VMware Infrastructure

VMware vSphere and VMware Infrastructure include distributed services that enable efficient and automated resource management and high availability of virtual machines. These services include VMware VMotion, VMware Storage VMotion, VMware DRS, and VMware HA. The VMware vSphere platform introduced a new distributed service, VMware Fault Tolerance (FT). This section describes these services and illustrates their functionality.

Shared storage, such as EMC Celerra, EMC CLARiiON®, and EMC Symmetrix®, is required to use these services.

VMware VMotion

Virtual machines run on and consume resources from ESX/ESXi. VMotion enables the migration of running virtual machines from one physical server to another without service interruption, as shown in Figure 8 on page 30. VMotion can help perform maintenance activities such as upgrade or security patches on ESX hosts without any downtime. VMotion is also the foundation for DRS.

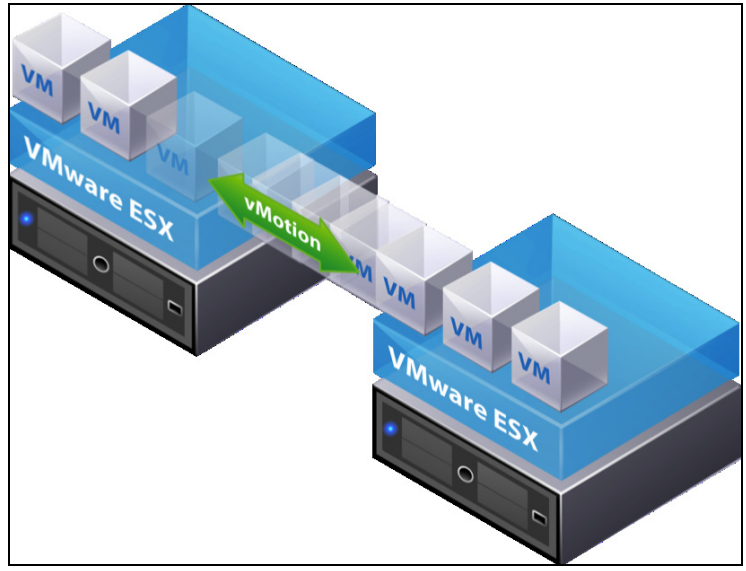


Figure 8 VMware VMotion

Storage VMotion

Storage VMotion enables the migration of virtual machines from one datastore to another datastore without service interruption, as shown in Figure 9 on page 31. This allows administrators, for example, to offload virtual machines from one storage array to another to perform maintenance, reconfigure LUNs, resolve out-of-space issues, and upgrade VMFS volumes. Administrators can also use Storage VMotion to optimize the storage environment for improved performance by seamlessly migrating virtual machine disks. With VMware vSphere, Storage VMotion is supported across all available storage protocols, including NFS. Furthermore, with VMware vSphere, Storage VMotion is fully integrated into vCenter Server and does not require any CLI execution.

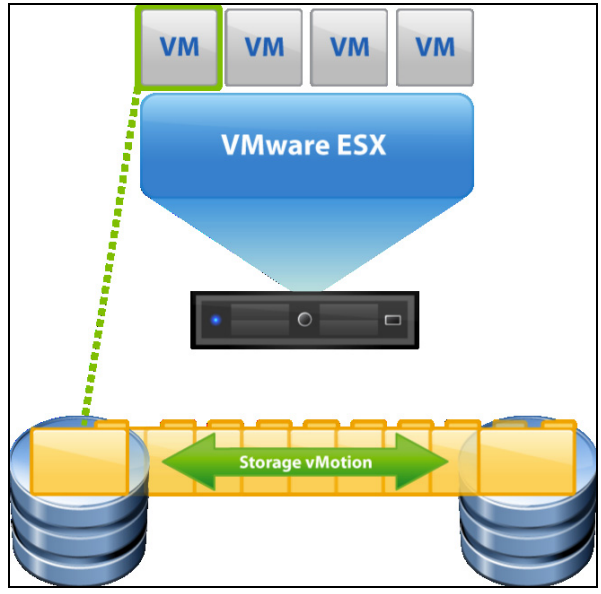


Figure 9 Storage VMotion

VMware Distributed Resource Scheduler (DRS)

VMware DRS helps to manage a cluster of physical hosts as a single compute resource. A virtual machine can be assigned to a cluster. DRS will then find an appropriate host on which the virtual machine will run. DRS places virtual machines in such a way that the load across the cluster is balanced, and cluster-wide resource allocation policies (such as reservations, priorities, and limits) are enforced. When a virtual machine is powered on, DRS performs an initial placement of the

virtual machine on a host. As cluster conditions change (such as the load and available resources), DRS migrates virtual machines (leveraging VMotion) to other hosts as necessary. When a new physical server is added to a cluster, DRS enables virtual machines to immediately and automatically take advantage of the new resources because it distributes the running virtual machines by way of VMotion.

Figure 10 on page 32 shows the DRS.

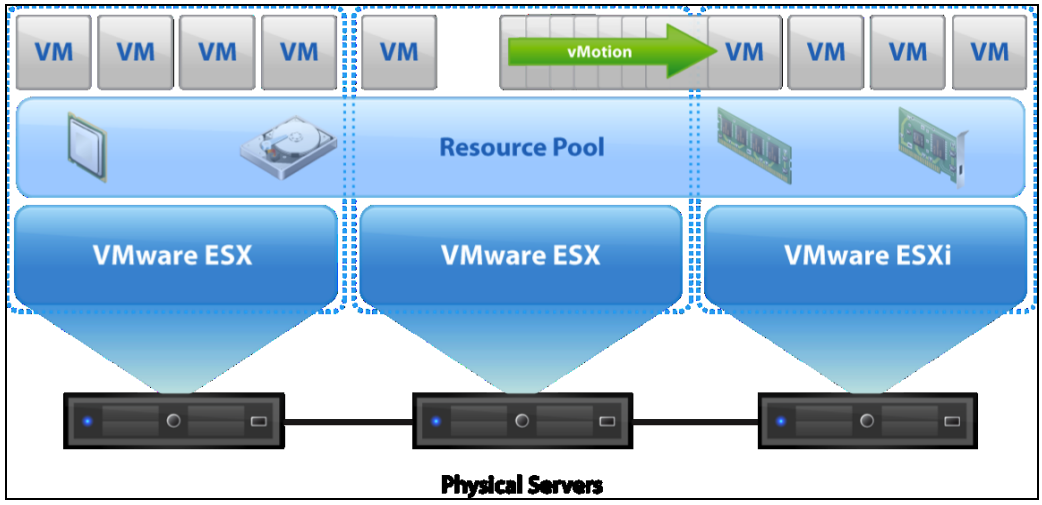


Figure 10 VMware DRS

DRS can be configured to automatically execute virtual machine placement, virtual machine migration, and host power actions, or to provide recommendations, which the data center administrator can assess and manually act upon. For host power actions, DRS leverages the VMware Distributed Power Management (DPM) feature. DPM allows a DRS cluster to reduce its power consumption by powering hosts on and off based on cluster resource utilization.

VMware High Availability (HA)

If a host or virtual machines fail, VMware HA automatically restarts the virtual machines on a different physical server within a cluster. All applications within the virtual machines have the high availability benefit through application clustering.

HA monitors all physical hosts and virtual machines in a cluster and detects failure of hosts and virtual machines. An agent placed on each physical host maintains a heartbeat with the other hosts in the resource

pool. Loss of a heartbeat initiates the process of restarting all affected virtual machines on other hosts. VMware tools help HA check the health of virtual machines. Figure 11 on page 33 gives an example of VMware HA.

HA ensures that sufficient resources are available in the cluster at all times to restart virtual machines on different physical hosts in the event of a host failure.

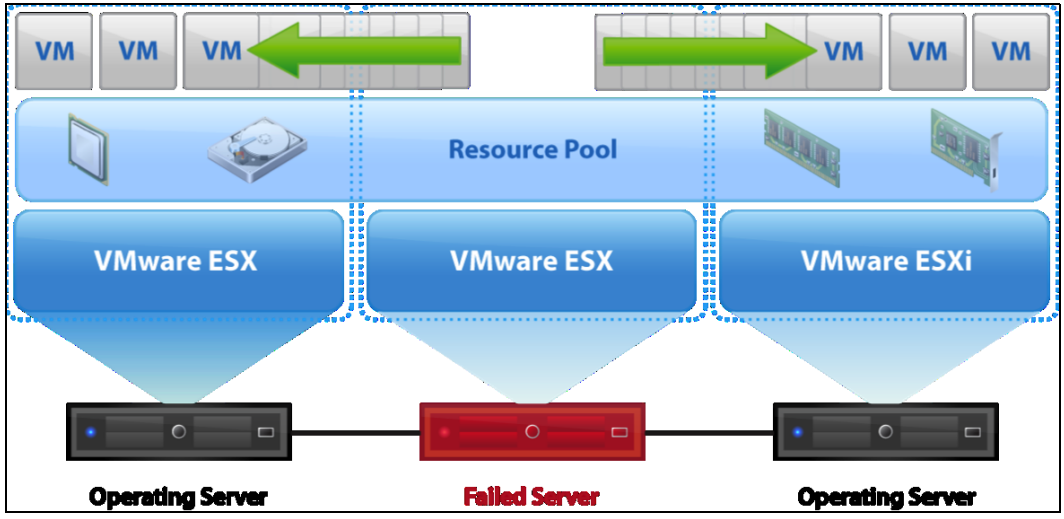


Figure 11 VMware HA

VMware Fault Tolerance (FT)

VMware FT, which was introduced in VMware vSphere, provides continuous availability by protecting a virtual machine (the primary virtual machine) with a shadow copy (secondary virtual machine) that runs in virtual lockstep on a separate host. Figure 12 on page 34 shows an example of VMware FT.

It is worth noting that at this time FT is provided as an initial release that is supported in a limited configuration. VMware vSphere documentation provides further details on the configuration supported for FT.

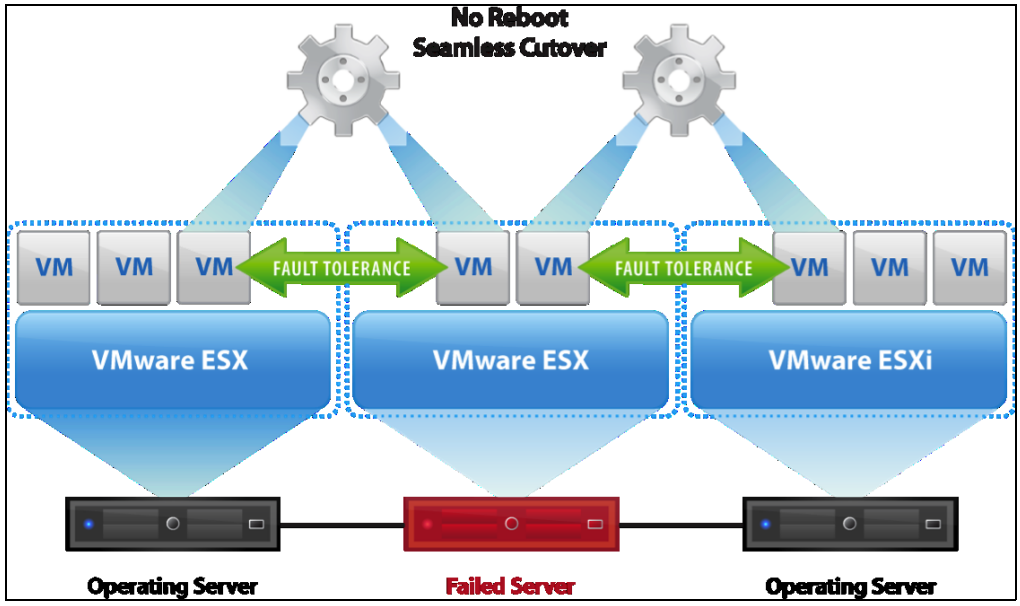


Figure 12 VMware Fault Tolerance

Inputs and events performed on the primary virtual machine are recorded and replayed on the secondary virtual machine to ensure that the two remain in an identical state. Actions such as mouse-clicks and keystrokes that are recorded on the primary virtual machine are replayed on the secondary virtual machine. Because the virtual machine is in virtual lockstep with the primary virtual machine, it can take over execution at any point without interruption or loss of data.

Backup and recovery solutions with VMware vSphere and VMware Infrastructure

VMware vSphere and VMware Infrastructure platforms include a backup and recovery solution for virtual machines that resides in the data center. VMware Consolidated Backup (VCB) is such a solution for VMware Infrastructure environments. VMware Data Recovery is a backup application for VMware vSphere environments that is based on the VMware vStorage for Data Protection API. The following two sections provide further details on these two solutions.

VMware Data Recovery

VMware Data Recovery is a new backup and recovery solution for VMware vSphere. VMware Data Recovery, distributed as a VMware virtual appliance, creates backups of virtual machines without interrupting their use or the data and services they provide. VMware Data Recovery manages existing backups and removes backups as they become older. It also supports target-based deduplication to remove redundant data. VMware Data Recovery supports the Microsoft Windows Volume Shadow Copy Service (VSS), which provides the backup infrastructure for certain Windows operating systems. VMware Data Recovery is built on the VMware vStorage for Data Protection API. It is integrated with VMware vCenter Server and enables centralized scheduling of backup jobs. Integration with vCenter Server also enables virtual machines to be backed up, even when they are moved using VMware VMotion or VMware DRS.

Figure 13 on page 36 illustrates how VMware Data Recovery works.

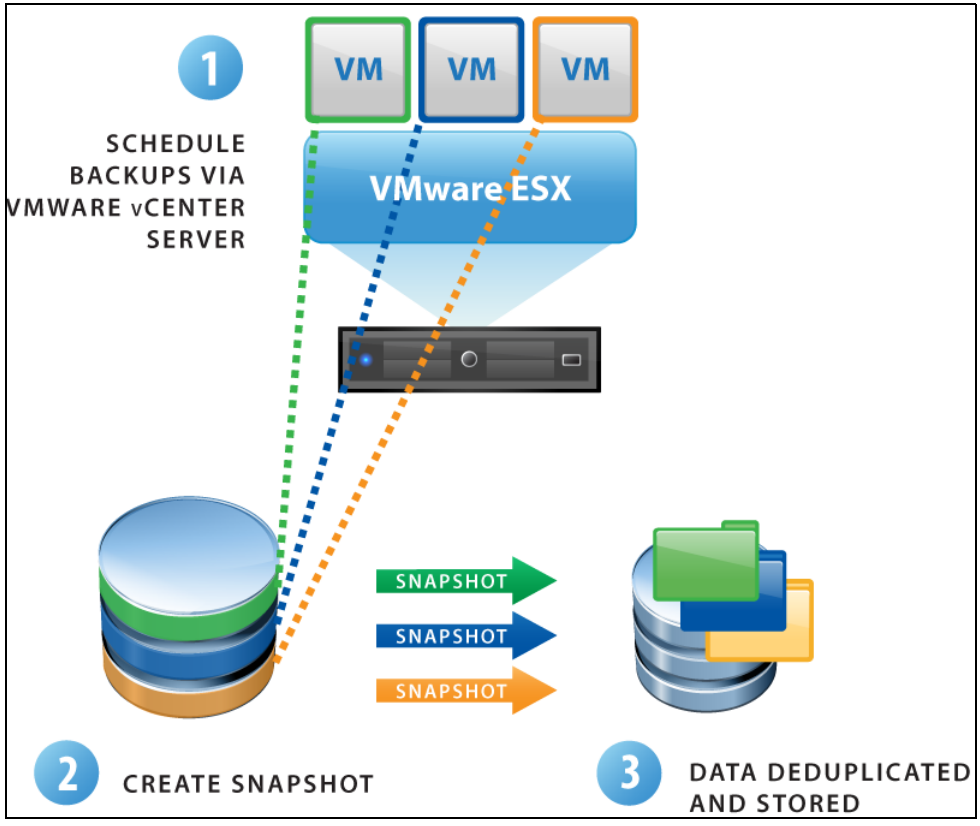


Figure 13 VMware Data Recovery

Backups can be stored on any virtual disk supported by virtual machines hosted on VMware ESX, including SANs, NAS devices, or Common Internet File System (CIFS) based storage such as SAMBA. All backed-up virtual machines are stored in a deduplicated store.

Benefits of deduplication store

VMware deduplication store technology used by VMware Data Recovery provides tight integration, evaluating patterns to be saved to restore points and performing checks to see if identical sections have already been saved. To maximize deduplication rates, ensure that similar virtual machines are backed up to the same destination because VMware supports storing the results of multiple backup jobs to use the same deduplication store. While backing up similar virtual machines to

the same deduplication store may increase space savings, similar virtual machines do not need to be backed up during the same job. (Deduplication is evaluated for all virtual machines stored, even if some are not currently being backed up.) VMware Data Recovery is designed to support deduplication stores that are up to 1 TB in size and each backup appliance is designed to support the use of two deduplication stores. VMware Data Recovery does not impose limits on the size of deduplication stores or the number of deduplication stores. But, if more than two stores are used or if the size of a store exceeds 1 TB, the performance may be affected.

VMware Consolidated Backup

VCB integrates with third-party software to perform backups of virtual machine disks with VMware Infrastructure.

The following are the key features of VCB:

- Integrate with most major backup applications to provide a fast and efficient way to back up data in virtual machines.
- Eliminates the need for a backup agent in a virtual machine (for crash-consistent backup only).
- Reads virtual disk data directly from the SAN storage device by using Fibre Channel or iSCSI, or by using a network connection to an ESX server host.
- Can run in a virtual machine to back up virtual machines that reside on a storage device accessed over a network connection.
- When used with iSCSI, VCB can run in a virtual machine.
- Supports file-level full and incremental backup for virtual machines running Microsoft Windows operating system and image-level backup for virtual machines running any operating system.
- Can be used with a single ESX/ESXi host or with a vCenter Server.
- Supports the Volume Shadow Copy Service (VSS), which provides the backup infrastructure for certain Windows operating systems running inside ESX 3.5 update 2 and later.

How VCB works

VCB consists of a set of utilities and scripts that work in conjunction with third-party backup software. To ensure that VCB works with specific backup software, either VMware or the backup software vendor provides integration modules that contain the required pre-backup and post-backup scripts.

The third-party backup software, integration module, and VCB run on the VCB proxy, which is either a physical or a virtual machine that has Microsoft Windows operating system installed.

VMware vCenter Site Recovery Manager

VMware vCenter Site Recovery Manager (SRM) delivers advanced capabilities for disaster recovery management, nondisruptive testing, and automated failover. VMware SRM can manage the failover from production data centers to disaster recovery sites, as well as the failover between two sites with active workloads. Multiple sites can even recover into a single shared recovery site. VMware SRM can also help with planned data center failovers such as data center migrations.

VMware SRM is integrated with a range of storage replication technologies including EMC SRDF® for Symmetrix, EMC MirrorView™ for CLARiiON, EMC Celerra Replicator™, and EMC RecoverPoint.

VMware SRM 4 introduces NFS storage replication support, many-to-one failover using shared recovery sites, and full integration with VMware vSphere 4.

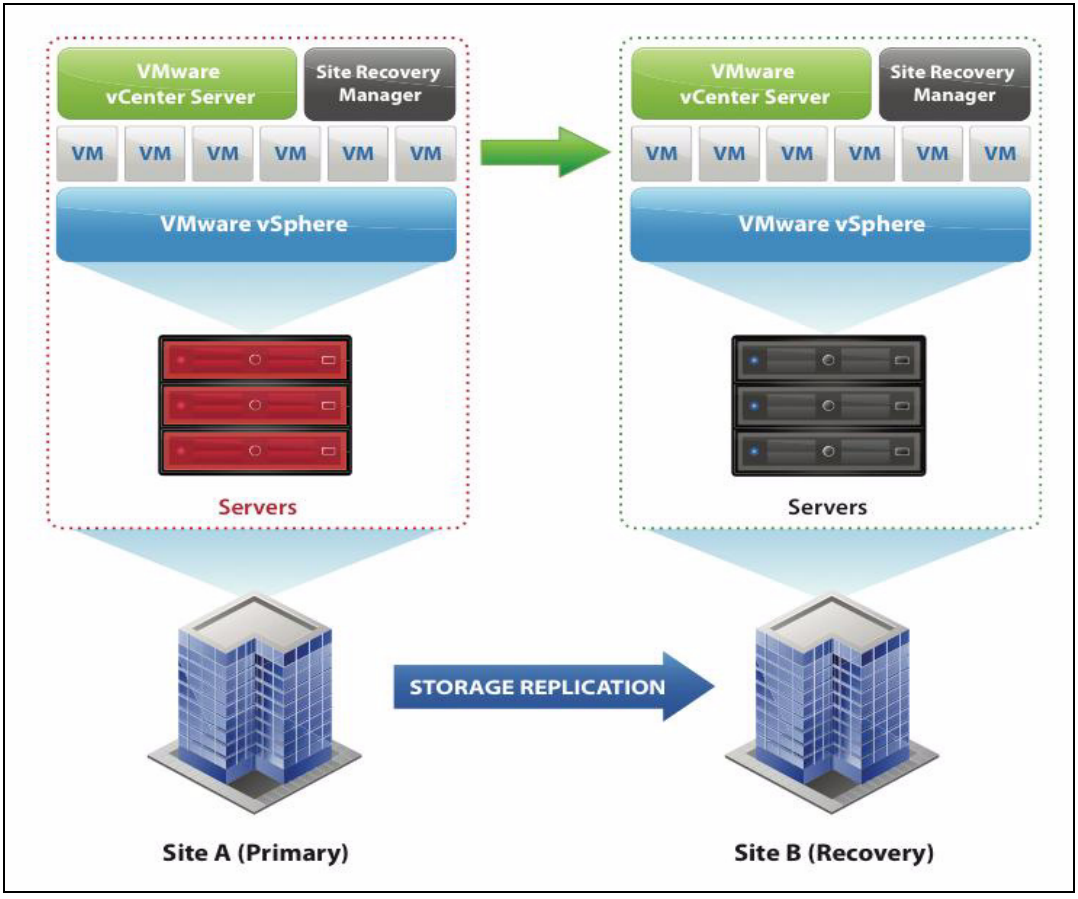


Figure 14 Site Recovery Manager

Key benefits of VMware SRM

VMware SRM provides capabilities to do the following:

Disaster recovery management

- Create and manage recovery plans directly from VMware vCenter Server. These recovery plans can be extended with custom scripts. Access to these recovery plans can be controlled with granular role-based access controls.

- Discover and display virtual machines protected by storage replication using integration certified by storage vendors.
- Monitor the availability of remote sites and alert users of possible site failures.
- Store, view, and export results of test and failover execution from VMware vCenter Server.
- Leverage iSCSI, Fibre Channel, or NFS-based storage replication solutions.
- Recover multiple sites into a single shared recovery site.

Nondisruptive testing

- Use storage snapshot capabilities to perform recovery tests without losing replicated data.
- Connect virtual machines to an existing isolated network for testing purposes.
- Automate the execution of recovery plan tests. Customize the execution of tests for recovery plan scenarios. Automate the cleanup of testing environments after completing tests.

Automated failover

- Initiate the recovery plan execution from VMware vCenter Server with a single button. Manage and monitor the execution of recovery plans within VMware vCenter Server.
- Automate the promotion of replicated datastores for recovery by using adapters created by leading storage vendors for their replication platforms.
- Execute user-defined scripts and pauses during recovery.
- Reconfigure virtual machine IP addresses to match the network configuration at the failover site.

VMware View

VMware View is an end-to-end desktop virtualization solution that leverages VMware vSphere or VMware Infrastructure to enable customers to manage and secure virtual desktops across the enterprise from within the data center.

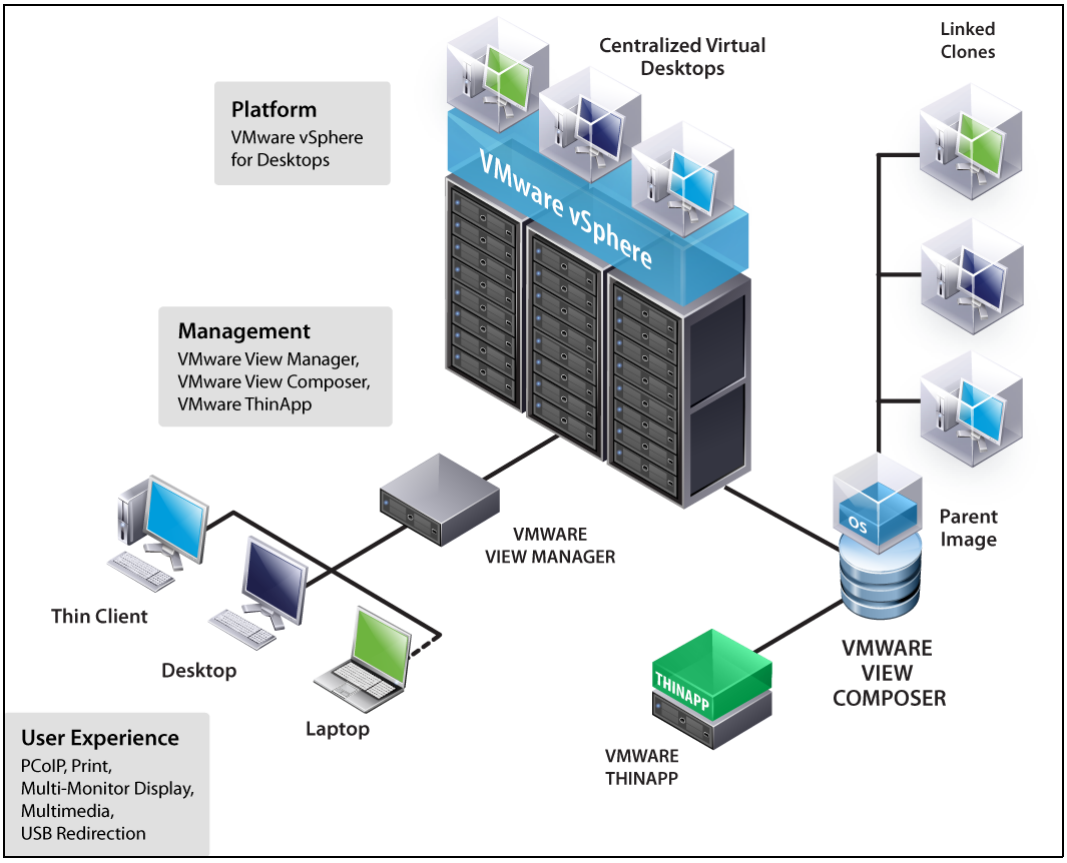


Figure 15 VMware View with VMware vSphere 4

Key benefits of VMware View

VMware View provides the capabilities to do the following:

- Get control and manageability in a single solution — VMware View is a comprehensive solution that provides the functionality that most organizations need to connect and manage their remote clients and centralized virtual desktops while keeping data safe and secure

in the data center. Designed for desktop administrators, VMware View offers an intuitive web-based management interface with Microsoft Active Directory (AD) integration for user authentication and policy enforcement. Centralized administration of all desktop images helps simplify upgrades, patches, and desktop maintenance, and enables the use of VMware View to manage connections between remote clients and their centralized virtual desktop.

- Support remote users without sacrificing security — Since all the data is maintained within the corporate firewall, VMware View minimizes overall risk and data loss. Built-in SSL encryption provides secure tunneling to virtual desktops from unmanaged devices. Furthermore, optional integration with RSA SecurID enables two-factor authentication.
- Provide end users with a familiar desktop experience — With VMware View, end users get the same desktop experience that they would have with a traditional desktop. The VMware View display protocol, PC over IP (PCoIP), provides a superior end-user experience over any network on up to four different displays. Adaptive technology ensures an optimized virtual desktop delivery on both the LAN and the WAN and addresses the broadest list of use cases and deployment options with a single protocol. Personalized virtual desktops, complete with applications and end-user data and settings, can be accessed anywhere and anytime with VMware View.
- Extend the power of VMware vSphere to the desktop — VMware View is built on VMware vSphere 4 and can automate desktop backup and recovery of business processes in the data center.

Components of the VMware View solution

The components of the VMware View solution are:

- **VMware View Manager** — VMware View Manager is an enterprise-class desktop management solution that streamlines the management, provisioning, and deployment of virtual desktops.
- **VMware View Composer** — VMware View Composer is an optional tool that uses VMware Linked Clone technology to rapidly create desktop images that share virtual disks by using a master image. This conserves disk space and streamlines management.
- **VMware ThinApp** — VMware ThinApp is an optional application virtualization software that decouples applications from operating systems and packages them into an isolated and encapsulated file.

This allows multiple versions of applications to execute on a single desktop without conflict, or the same version of an application to run on multiple operating systems without modification.

- **Offline Desktop (experimental)** — Offline Desktop is a technology that allows complete virtual desktops to be moved between the data center and the physical desktop devices, with the security policies intact. Changes to the virtual desktop are intelligently synchronized between the data center and the physical desktop devices.

VMware vCenter Converter

VMware vCenter Converter is an optional module of VMware vCenter Server to import, export, or reconfigure source physical machines, virtual machines, or system images of VMware virtual machines.

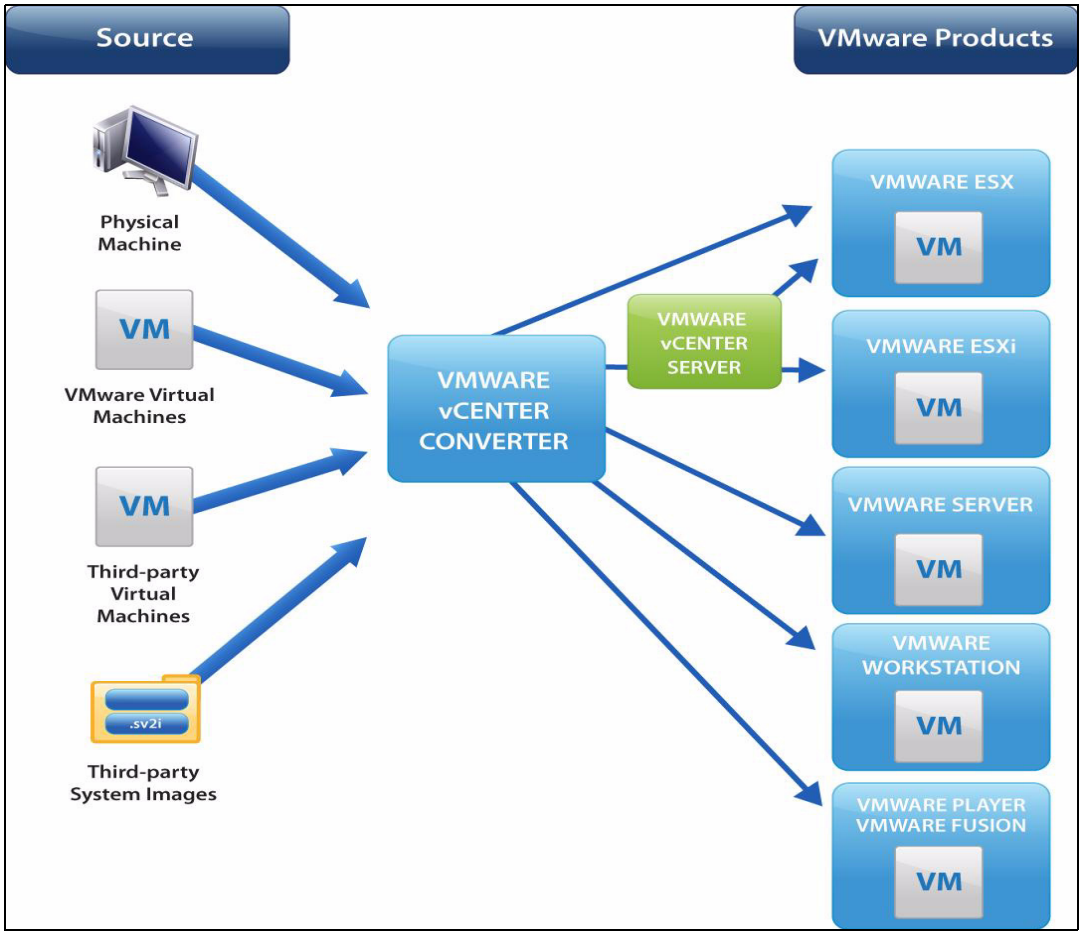


Figure 16 VMware vCenter Converter

Migration with vCenter Converter

Migration with vCenter Converter involves cloning a source machine or image and encapsulating, configuring virtual hardware, and registering it with the destination. The tool allows the conversion of virtual machines, which are managed by vCenter Server, to different VMware virtual machine formats and exports those virtual machines for use with other VMware products.

vCenter Converter can be used to perform the following tasks:

- Convert running remote physical machines to virtual machines and import the virtual machines to ESX/ESXi or ESX/ESXi hosts that are managed by vCenter Server.
- Convert and import virtual machines, such as those created with VMware Workstation or Microsoft Virtual Server 2005, to ESX/ESXi hosts that are managed by vCenter Server.
- Convert third-party backup or disk images to ESX/ESXi hosts that are managed by vCenter Server.
- Restore VCB images to ESX/ESXi hosts that are managed by vCenter Server.
- Export virtual machines managed by vCenter Server hosts to other VMware virtual machine formats.
- Reconfigure virtual machines managed by vCenter Server hosts so that they are bootable.
- Customize virtual machines in the vCenter Server inventory (for example, to change the hostname or to update network settings).

It is important to note that vCenter Converter does not support creating thin provisioned target disks on ESX 4 and ESXi 4. However, this can be achieved by performing a Storage VMotion migration after the virtual machines have been imported using vCenter Converter. Furthermore, thin provisioned virtual disks are supported using the standalone edition of this tool, VMware vCenter Convertor Standalone. This edition runs separately from vCenter Server.

Depending on the vCenter Converter component installed, perform hot or cold cloning by using a command line interface, or with the vCenter Converter Import, Export, or Reconfigure wizard available in the VMware vSphere Client.

2 EMC Foundation Products

This chapter presents these topics:

- Overview 48
- EMC CLARiiON 52
- EMC CLARiiON Navisphere Management Tools 54
- CLARiiON metaLUNs 57
- CLARiiON Virtual LUN technology 61
- Virtualization-Aware Navisphere 62
- Virtual Provisioning 69
- Fully Automated Storage Tiering (FAST) LUN Migrator 73
- EMC SnapView 77
- EMC SAN Copy 91
- EMC MirrorView 94
- EMC RecoverPoint 102
- EMC PowerPath 104
- EMC Replication Manager 107
- EMC StorageViewer 109

Overview

EMC provides many hardware and software products that support application environments on CLARiiON arrays. This chapter provides a technical overview of the EMC products used in this guide. The following products are used and tested with the solutions discussed in this guide:

- EMC CLARiiON Storage Systems — The EMC CLARiiON family of storage systems offer performance, capacity, and advance capabilities, such as local and remote replication for a wide range of application and business requirements. The CLARiiON family includes the entry-level CLARiiON AX4 and scales to the CX4-960, the industry's most powerful midrange system.
- EMC Navisphere® Management Suite — Navisphere Management suite includes a graphical user interface and a command line interface that allow a storage manager to discover, monitor, and provision capacity on one or more CLARiiON storage systems at local and remote locations thru a secure IP connection.
- CLARiiON MetaLUN Technology — MetaLUNs are a configuration option that allows the storage administrator to expand the capacity of existing CLARiiON LUNs when the LUN is online. MetaLUNs potentially provide additional performance benefits by spreading the workload across more resources.
- CLARiiON Virtual LUN Technology — A Virtual LUN technology allows the storage administrator to migrate data between LUNs without disruption to applications. LUNs can be migrated to the same or different RAID levels and disk types. In addition, a LUN can be migrated to a target LUN of larger size. The Virtual LUN technology provides an easy to use methodology for implementing Information Lifecycle Management by migrating LUNs to the appropriate storage based on changing performance and availability requirements.
- Virtualization-Aware Navisphere — Virtualization-aware Navisphere Manager simplifies storage administration tasks by integrating with VMware vCenter APIs and correlating virtual machines and their storage resources. Makes common tasks faster and easier, such as identifying root cause of virtual machine performance problems and optimizing capacity utilization.
- CLARiiON Virtual Provisioning™ — 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.

- EMC Navisphere Quality of Service Manager (NQM) — NQM allows the storage administrator to monitor application performance, and schedule user-defined policies that allocate system resources dynamically to meet required service levels. This allows an organization to achieve the benefit of consolidation while optimizing the performance of mission-critical applications.
- CLARiiON Fully Automated Storage Tiering (FAST) LUN Migrator — CLARiiON FAST LUN Migrator assists in the identification and movement of LUNs from Fibre Channel drives to Flash drives and SATA drives to achieve increased performance and/or lower TCO. This consists of two phases: the analysis phase and the automated LUN migration phase.
- EMC SnapView™ — SnapView allows the user to create local point-in-time copies of data at the LUN level for testing, backup, and recovery operations. The SnapView family includes two flexible options: pointer-based, space saving snapshots, and highly functional, full-volume clones.
  - SnapView clones create full-image copies of a source LUN that can be established, synchronized, fractured, and presented to a different host for backup or other applications. Because SnapView tracks changes, subsequent establish operations only require copying of the changed tracks. A single clone source LUN can have up to eight simultaneous target clones
  - SnapView snapshots use a pointer-based technique to create an image of a source LUN. Because snapshots use a copy-on-first-writer technique, the target devices is available immediately after creation and a snapshot image typically only requires a fraction of disk space of the source LUN. A single source LUN can have as many as eight snapshots, each reflecting a different point-in-time view for the source.
  - Both snapshot and clones leverage consistency technology for control operations that result in consistent point-in-time data image when a data set spans multiple LUNs.

- EMC MirrorView — MirrorView is a business continuity solution that maintains a block level image of a LUN in a remote CLARiiON storage system. The MirrorView family has two replication options: synchronous and asynchronous.
  - MirrorView/Synchronous (MirrorView/S) provides a block-for-block mirror image of a production LUN to and is appropriate when the application requires a Recovery Point Objectives (RPO) of zero data loss.
  - MirrorView/Asynchronous (MirrorView/A) provides a point-in-time consistent image of a LUN in a remote CLARiiON by periodically updating the secondary image in the remote CLARiiON. MirrorView/A is appropriate for applications that require a RPO from minutes to hours.

Both MirrorView replication options support consistency group technology to maintain write-order consistency when a data set spans multiple LUNs.

- EMC RecoverPoint — EMC RecoverPoint brings you continuous data protection and continuous remote replication for on-demand protection and recovery to any point in time. RecoverPoint's advanced capabilities include policy-based management, application integration, and bandwidth reduction.
- EMC Open Replicator — Open Replicator provides the ability to copy the contents of a LUN to a different EMC or non-EMC storage system, or both. The support for a heterogeneous storage environment makes Open Replicator an ideal solution for content distribution and data migration applications. Open Replicator may be configured for full copy pull or push operations, and/ incremental push operations where only the changed data is copied to the remote storage system. When Open Replicator is part of a disaster recovery environment, it is better suited for content distribution as it does not maintain point-in-time consistent image on the remote storage system while synchronization is in progress.
- EMC PowerPath® — PowerPath is host-based software that provides I/O path management. PowerPath operates with several storage systems on different enterprise operating systems. It provides failover and load balancing transparent to the host application and database.
- EMC Connectrix® — Connectrix is a Fibre Channel director or switch that moves information throughout the SAN environment, enabling the networked storage solution.

- EMC Replication Manager — EMC Replication Manager is software that creates replicas of mission-critical databases on disk arrays with traditional tape media. Replication Manager can create a disk replica of data simply, quickly, and automatically. It automates all tasks and procedures related to data replication, as well as reducing the amount of time, resources, and expertise involved with integrating and managing disk-based replication technologies.
- EMC StorageViewer — EMC Storage Viewer provides a convenient solution for managing EMC storage devices from within Virtual Infrastructure Client. It enables administrators to have an in-depth understanding of the back end of their storage without leaving the confines of the Virtual Infrastructure Client, allowing for a higher, more efficient level of management and control.

EMC CLARiiON

EMC CLARiiON is a highly available storage system designed for no-single-points-of-failure, and delivers industry-leading performance for mission-critical applications and databases. CLARiiON storage systems provide both iSCSI and Fibre Channel connectivity options for Open Systems hosts, and supports advanced data replication capabilities. The core software that runs on the CLARiiON, called FLARE®, provides a robust set of functions including data protection, host connectivity, and local and remote data replication such as RecoverPoint and MirrorView.

CLARiiON uses a modular architecture that allows the system to grow nondisruptively as business requirements change. The two major components are the storage processor enclosure (SPE) and the disk-array enclosure (DAE). The SPE contains two independent high performance storage processors that provide front-end connectivity, read and write cache, and connectivity to the back end. The DAE provides the back-end storage and each DAE can house up to 15 disk drive modules. Multiple DAEs can be interconnected across multiple back-end loops to meet capacity and performance requirements.

CX4 is the current generation of CLARiiON; it utilizes UltraFlex™ technology that uses cut-thru-switch technology and full 4 Gb/s back-end disk drives with 8 GB/s and 4 Gb/s Fibre Channel front-end connections. 10 Gb/s and 1 Gb/s front-end iSCSI connections are also available. The UltraScale™ architecture provides high performance and reliability, with advanced fault-detection and isolation capabilities. High-performance Flash and Fibre Channel drives, and low-cost, high-capacity SATA drives can be deployed within the same storage system; this enables tiered storage solutions within a single system.

CLARiiON implements a LUN ownership model where I/O operations for a LUN are serviced by the owned storage processor. Because physical disk drives are shared by both storage processors, in the event of a path failure, the LUN ownership can be moved (trespassed) to the peer storage processor, allowing the I/O operation to proceed. This ownership model provides high availability and performance, by balancing the workload across processing resources. With release 26 of the FLARE Operating Environment, the Asymmetric Logical Unit Access (ALUA) standard is supported. ALUA provides asymmetric active/ active LUN ownership for the CLARiiON. With ALUA, either storage processor can accept an I/O operation and will forward it to the owner Storage Processor through the internal high-speed messaging interface. This capability requires that the path management software

support the ALUA standard. EMC PowerPath leverages the ALUA architecture to optimized performance and provides advanced failover intelligence for the CLARiiON. VMware vSphere 4 supports ALUA connectivity to CLARiiON.

CLARiiON arrays provide the flexibility to configure data protection levels appropriate for the application performance and availability requirements. A mixture of RAID 0, 1, 3, 1/0, 5, and 6 can be configured within the same system. Additional availability features include nondisruptive software and hardware upgrades, proactive diagnostics, alerts, and phone-home capabilities. CLARiiON supports global hot sparing which provides automatic, online rebuilds of redundant RAID groups if any of the group's disk drives fail.

The current CX4 family includes the midrange CX4-960, CX4-480, CX4-240, and CX4-120. The AX4 is an entry-level storage system with a similar architecture and many of the same features and interfaces as the CX4 family.

Compatibility and interoperability between CLARiiON systems enable customers to perform data-in-place upgrades of their storage solutions from one generation to the next, protecting their investment as their capacity and connectivity demands increase.

EMC CLARiiON Navisphere Management Tools

Navisphere is the CLARiiON suite of management tools that provide centralized management of one or more CLARiiON storage systems. It includes both a web-based GUI and a command line interface (CLI). Navisphere provides an easy-to-use and secure interface for all monitoring, configuration, and control operations.

Manager

Navisphere Manager is a browser-based interface that allows users to discover, monitor and provision storage securely on one or more CLARiiON storage systems from any location. Navisphere Manager includes a number of wizards that simplify complex management tasks and thus reduces the risk that is normally associated with change. Navisphere Manager fully integrates the configuration and control of optional software packages including Navisphere Analyzer, SnapView, MirrorView, and Open Replicator within a single management interface.

Access Logix™ is an integrated feature of Navisphere Manager that provides an intuitive, interface for performing LUN masking in a shared storage environment. Access Logix configuration object is storage group. A storage group is a collection of one or more LUNs to which the user connect to one or more servers. A server only has access to the LUNs in the storage group to which it is connected. Normally, only a single host is connected to a storage group, and LUNs belong to only single storage group. However, in the case of a clustered environment, multiple hosts can be connected to a single storage group or the same LUNs can be added to multiple storage groups.

Navisphere management domains

A Navisphere management domain is a collection of one or more CLARiiON arrays that are managed together and share user authentication and authorization information. Through a single interface, Navisphere Manager can perform all management functions without having to log into individual arrays. Domain membership can be managed by dynamically adding or removing CLARiiON systems. Each storage system can belong to only one domain; however, it is possible to manage multiple domains within a single Navisphere Manager instance.

The storage systems within a domain communicate over an IP network. Each domain has a master node (master storage system) that maintains the master copy of the domain user account information. All account

information is duplicated on all systems in the domain, so if the master system fails, the information is retained and surviving systems in the domain remain accessible. While the domain master is unavailable, no updates are allowed to the global accounts until the system is replaced or a new master is designated.

Event notification

Navisphere includes an event-monitoring facility that allows a storage administrator to define events of interest and actions to take if any event occurs. Standard notifications methods are built in, including email notification, paging and asynchronous notification using SNPM traps. Custom responses can also be configured. An example of a custom response would be the execution of a script to perform corrective action or custom notification.

Command line interface (CLI)

Navisphere CLI provides a set of commands for performing all CLARiiON monitoring, configuration and control operations. There are three types of command line interfaces: Classic CLI, Java CLI, and Secure CLI.

- Classic CLI is the legacy command line interface and uses a privileged user list to control access. Because of the limited security implied with the Classic CLI, applications and scripts should be migrated to the newer Secure CLI as the support of the Classic CLI will eventually be dropped in a future release.
- Java CLI is implemented as a Java script (jar navicli.jar) and requires the installation of JRE. The Java CLI is more secure than the Classic CLI as it is integrated with Navisphere security. A user is authenticated and authorized to access by typing a valid Navisphere domain username, password, and scope when executing a Navisphere Java CLI command. The Java Navisphere CLI does not support the full set of Navisphere functionality and has been replaced with the new Navisphere Secure CLI.
- Navisphere Secure CLI is a comprehensive command line interface that is fully integrated into the Navisphere security. Like the Java CLI, the Navisphere Secure CLI uses SSL-based data encryption and requires a valid username, password and scope to authenticate and authorize access to all Navisphere configuration and control operations. All new applications should be written using the

Navisphere Secure CLI facility. Furthermore, applications written using the classic or Java-based CLI should be migrated to the Navisphere Secure CLI.

Host utilities

The Navisphere suite also includes several optional host utilities that provide significant value in managing CLARiiON-based storage environments.

- Navisphere Agent provides a communication bridge between the CLARiiON system and attached hosts. The two primary functions of the host agent is to register the host and HBA connections and to provide Navisphere Manager with mapping information that identify how a LUN is used by the host system. For example, the host agent maps the LUN ID to the host physical device name or drive letter. The Navisphere Server Utility provides a similar function but only runs when invoked by a user. The Server Utility is recommended for environments where it is not possible to run the Host agent.
- Navisphere Array initialization utility is host-based software used to initialize Storage Processor IP address information and domain security. The initialization utility requires that the host and storage processor be connected to the same subnet. This utility has significantly reduced the complexity of array installation.
- Navisphere service toolbar is a standalone utility for performing CLARiiON software installation and upgrades.

Analyzer

Navisphere Analyzer is a performance reporting tool that is fully integrated with Navisphere Manager and CLI. It reports key performance metrics that enables users understand usage patterns trends in normal operation mode, as well as identify potential performance bottlenecks, which can be addressed by tuning system parameters, reallocation of resources, or by adding hardware components, such as cache memory or disks. Navisphere Analyzer can be used to continually monitor and analyze performance and to fine tune storage-system for maximum performance and utilization.

CLARiiON metaLUNs

CLARiiON metaLUNs allow users to expand the capacity while the LUN is online and available to the host. MetaLUNs are comprised of two or more individual LUNs that are logically connected and presented to a host or application as a single LUN. An additional benefit of metaLUNs is the potential performance improvement that can be achieved by spreading the capacity and workload across more disk resources on the back end.

MetaLUNs can be created from the existing LUNs by using either striped or concatenated expansion."Stripe expansion," on page 57 and "Concatenate expansion," on page 58 describe in detail the two types of expansion that is possible.

Stripe expansion

Figure 17 on page 58 shows a striped metaLUN before and after expansion. This is represented both graphically, and as it appears in Navisphere Manager. The graphical depiction shows that after the striping process, the data—which was originally contained on LUN 27—is now spread evenly across all three of the LUNs in the component. Since the process depicted is a striped expansion, there is only one component, component 0, that includes all members of the new metaLUN.

Distributing the data in this manner spreads the workload across multiple physical spindles and potentially provides performance benefits. When expanding a LUN using striped expansion, all member LUNs must be of the same size, protection scheme, and physical drive type. The additional capacity of the added LUNs is available to the host only after the restriping process has completed. Expansion rate can be configured to minimize the effect on other concurrent workloads.

The rules for conducting striped LUN expansion are:

- All FLARE LUNs in the striped component must be of the same RAID type.
- All FLARE LUNs in the striped component must be of the same user capacity.
- All FLARE LUNs in a metaLUN must reside on the same disk type—either all Fibre Channel or all ATA.

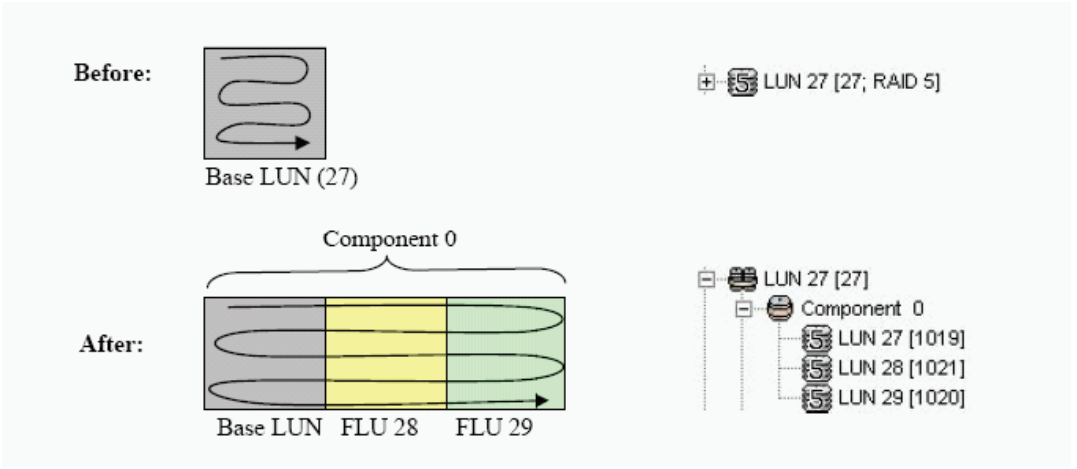


Figure 17 MetaLUN expansion using the striping method

Concatenate expansion

Figure 18 on page 59 shows a concatenated metaLUN before and after expansion. In concatenate expansion, data residing on the base LUN remains in place. Additional capacity is added by simply appending to the end of the addressable space. The advantage of concatenated expansion is that access to the additional capacity is available immediately as no data reorganization is required as is the case with striped expansion.

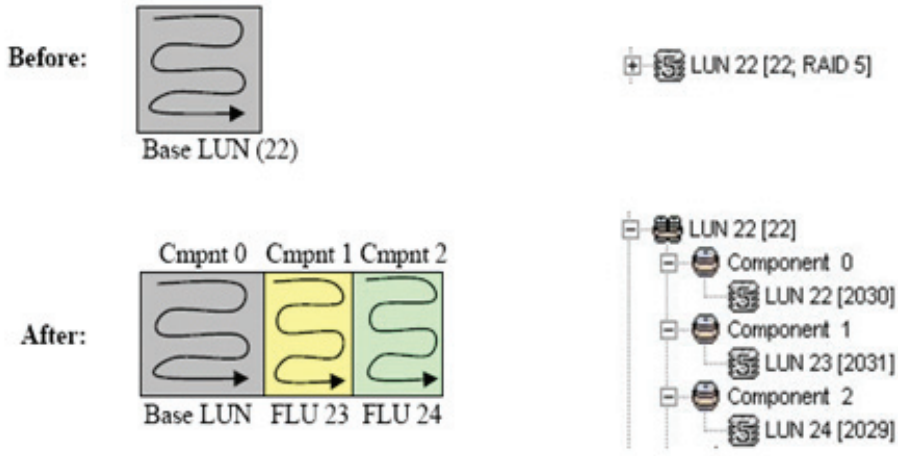


Figure 18 MetaLUN expansion using the concatenation method

Concatenate expansion also offers more flexibility as the members of the metaLUN can be of different RAID types and capacities. This provides the ability for a storage administrator to use noncontiguous disk space and maximize capacity utilization.

The following are expansion rules for concatenated expansion:

- All LUNs in a concatenated metaLUN must be either protected (parity or mirrored) or unprotected. RAID types within a metaLUN can be mixed. For example, a RAID 1/0 LUN can be concatenated with a RAID 5 LUN. A RAID 0 can be concatenated with another RAID 0, but not with a RAID 5 LUN.
- All LUNs in a concatenated metaLUN must reside on the same disk-drive type—either all Fibre Channel or all ATA.

Navisphere Manager includes an easy-to-use Wizard that takes the user through the process for creating metaLUNs. Figure 19 on page 60 shows the first dialog for creating a metaLUN.

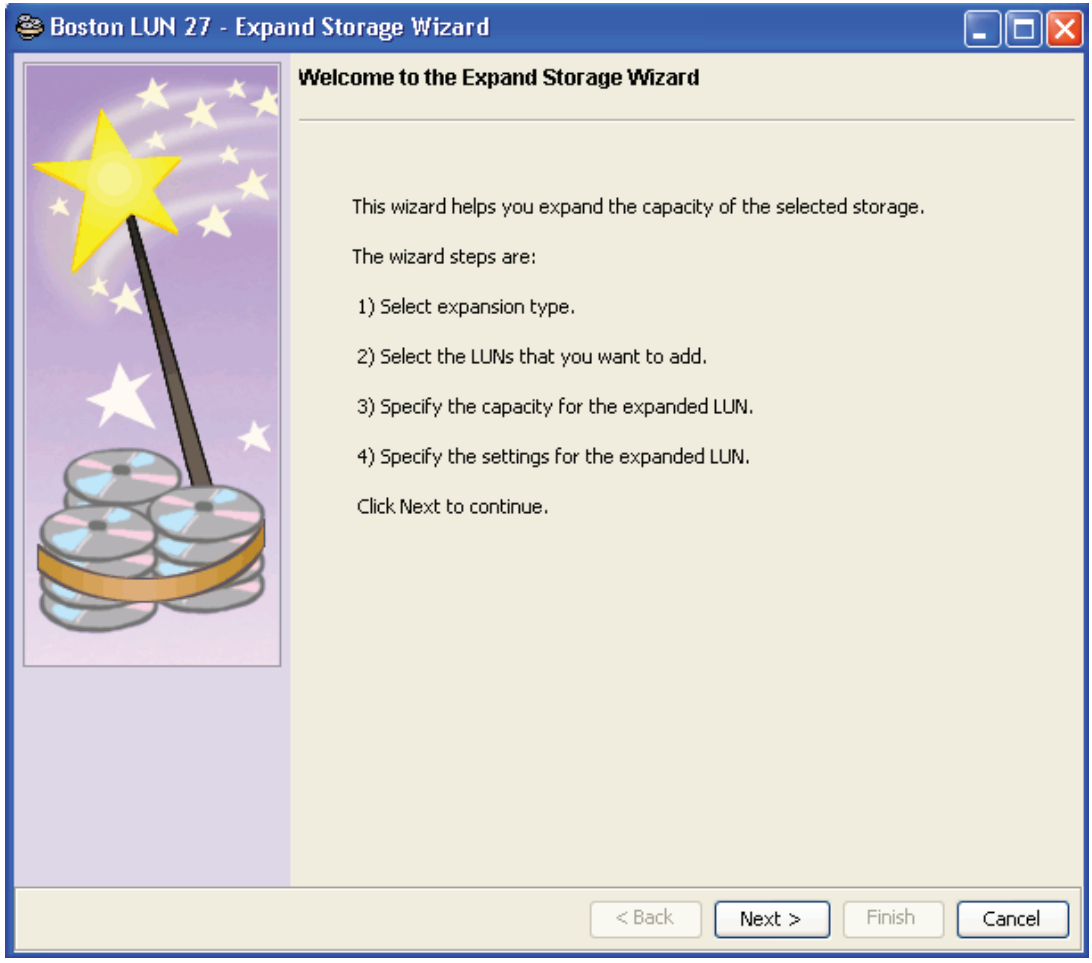


Figure 19 Expand Storage Wizard

CLARiiON Virtual LUN technology

CLARiiON Virtual LUN technology allows users to reconfigure the backend storage allocated to a LUN. The reconfiguration can be performed while the LUN is online and without disruption to the host application. Usage examples include migration from one RAID group to another; optionally changing the data protection scheme; or even changing the physical drive type. This feature is often used in implementing information lifecycle management (ILM) as the back-end characteristics can be changed as performance and availability requirements change. Virtual LUN technology is implemented by performing a block-by-block migration from a source LUN to a target LUN. When the migration is complete, the target LUN retains all the attributes of the source LUN including the world wide name (WWN), LUN ID and logical unit number, making the migration completely transparent to the attached host. At the successful completion of the migration, the original source LUN is unbound, freeing up the disk space for other uses. Additionally, a source LUN can be migrated to a larger target to increase the capacity.

LUN migration is managed through Navisphere Manager. The Migrate option invokes a dialog that allows the user to select the destination and rate of the migration for the session. The migration rate and the number of concurrent migration sessions can be set to minimize performance impact. A set of CLI commands are also available for managing migrations.

The following migration rules apply:

- Source and destination LUNs must reside on the same storage system.
- Source and destination LUN can be in the same or different RAID groups.
- Any public LUN or metaLUN can migrate to any LUN or metaLUN of equal or larger size.
- RAID type of source and target LUN can be different.
- The source and target LUNs can be on disks of different type (Fibre Channel or ATA, 2 Gb/s or 4 Gb/s disk speeds)

Virtualization-Aware Navisphere

The Navisphere VM-aware 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.

It 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 (as shown in Figure 20 on page 63).

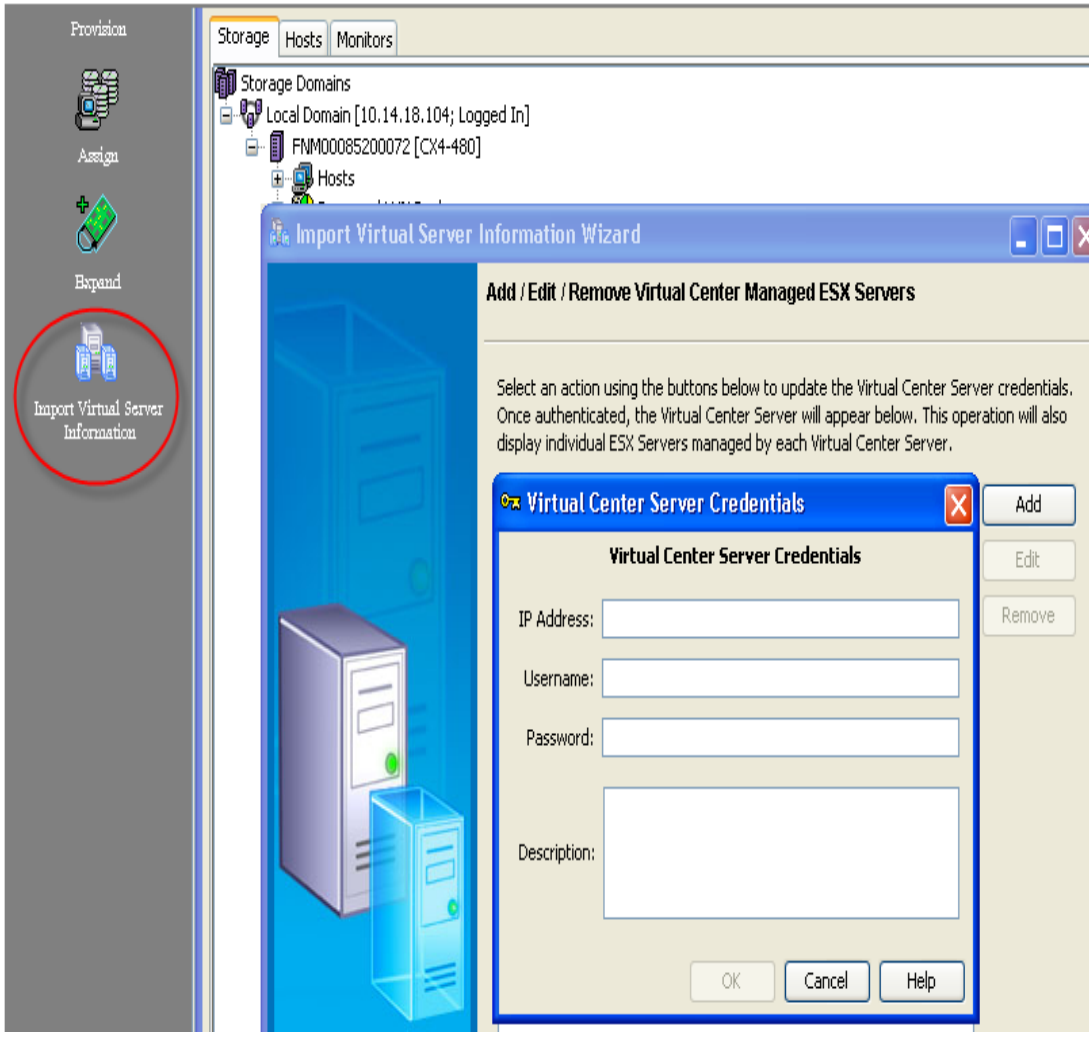


Figure 20 Import Virtual Server Information Wizard

The Storage group tab will display VMFS datastore name and Raw Mapped LUN information as shown in Figure 21 on page 64.

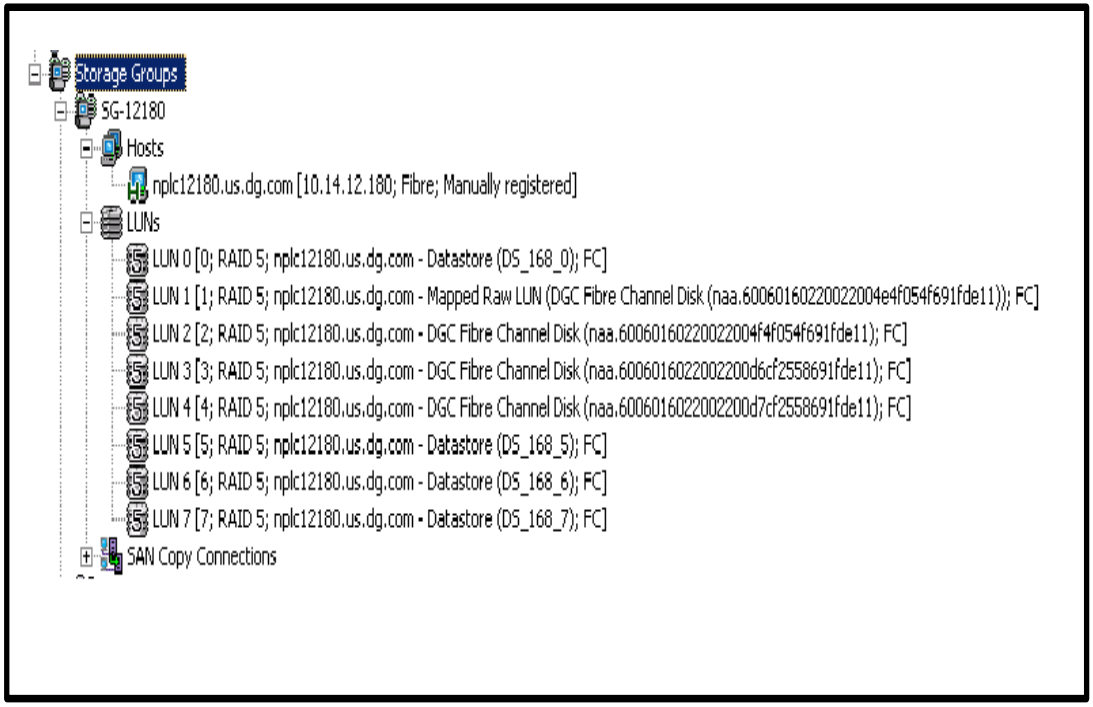


Figure 21 ESX 4 Filesystem information display

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

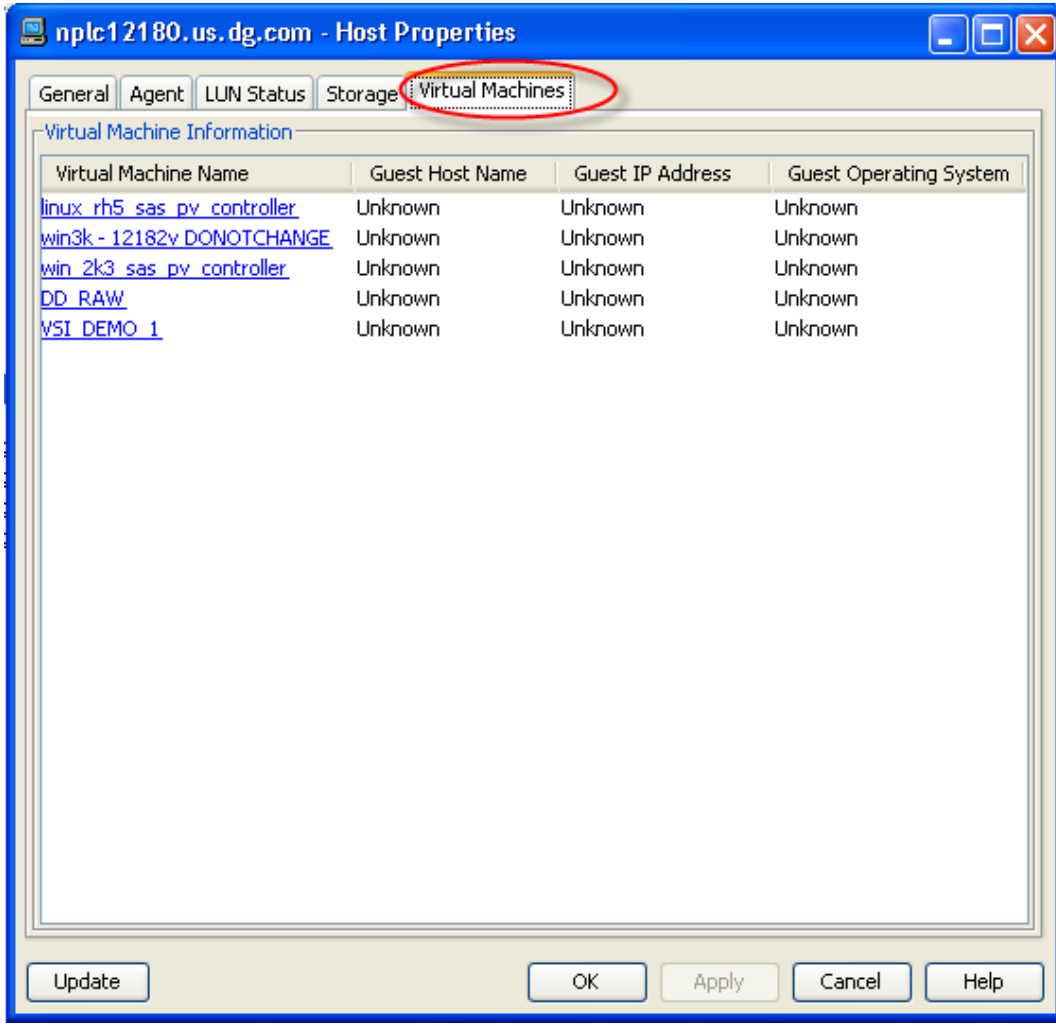


Figure 22 Virtual Machines tab available under ESX server

Clicking on one of the virtual machines in Figure 22 on page 65 opens the Virtual Machine Properties dialog box shown in Figure 23 on page 66. 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.

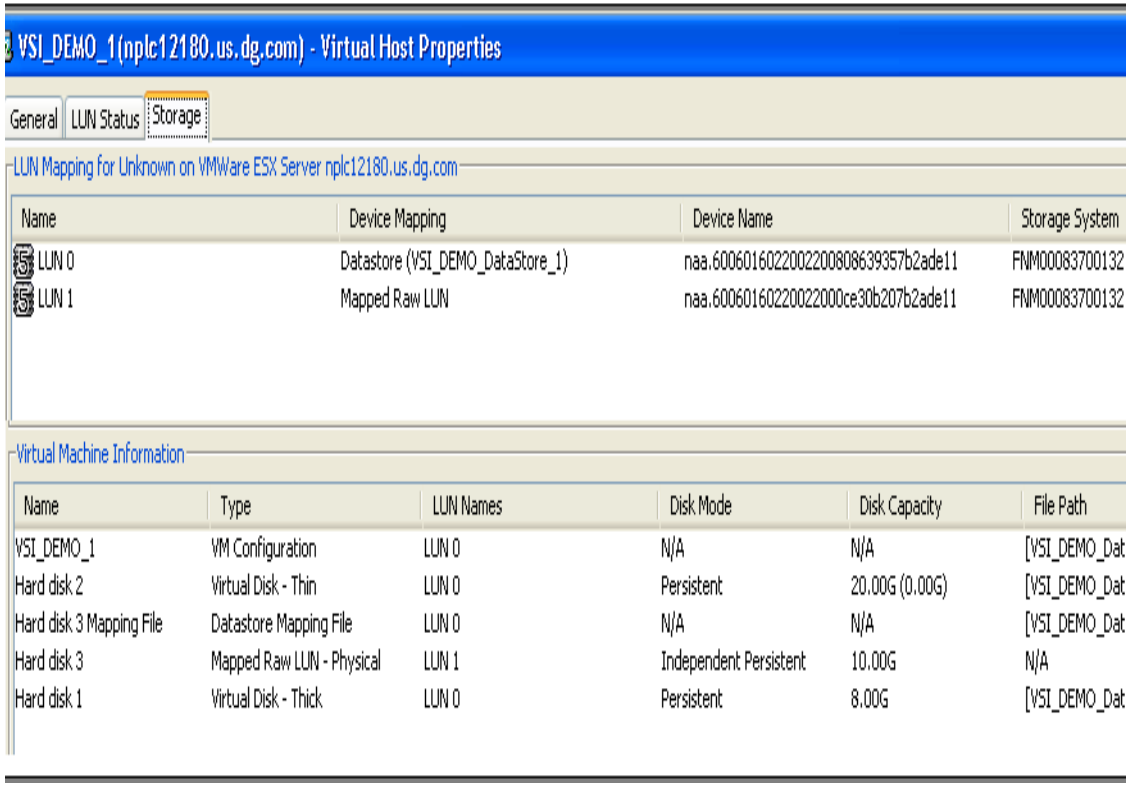


Figure 23 Virtual Host Properties dialog box

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.

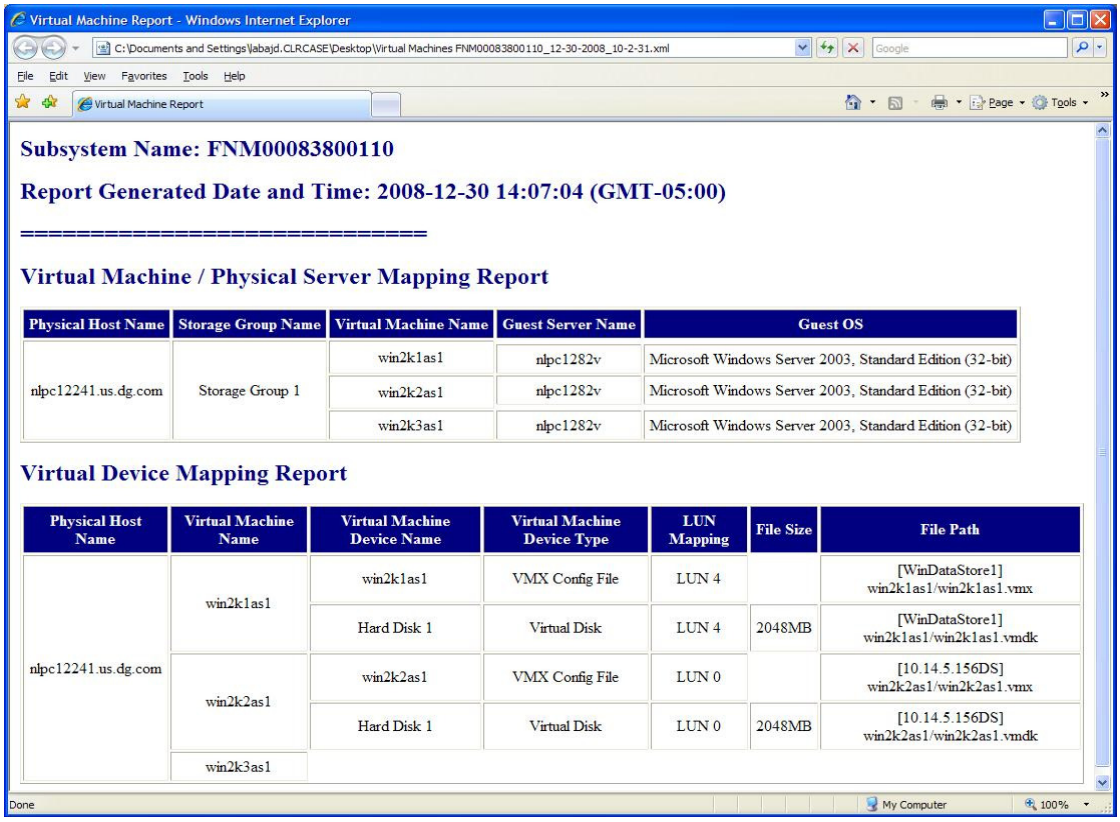


Figure 24 Virtual Machine report

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 Figure 25 on page 68.

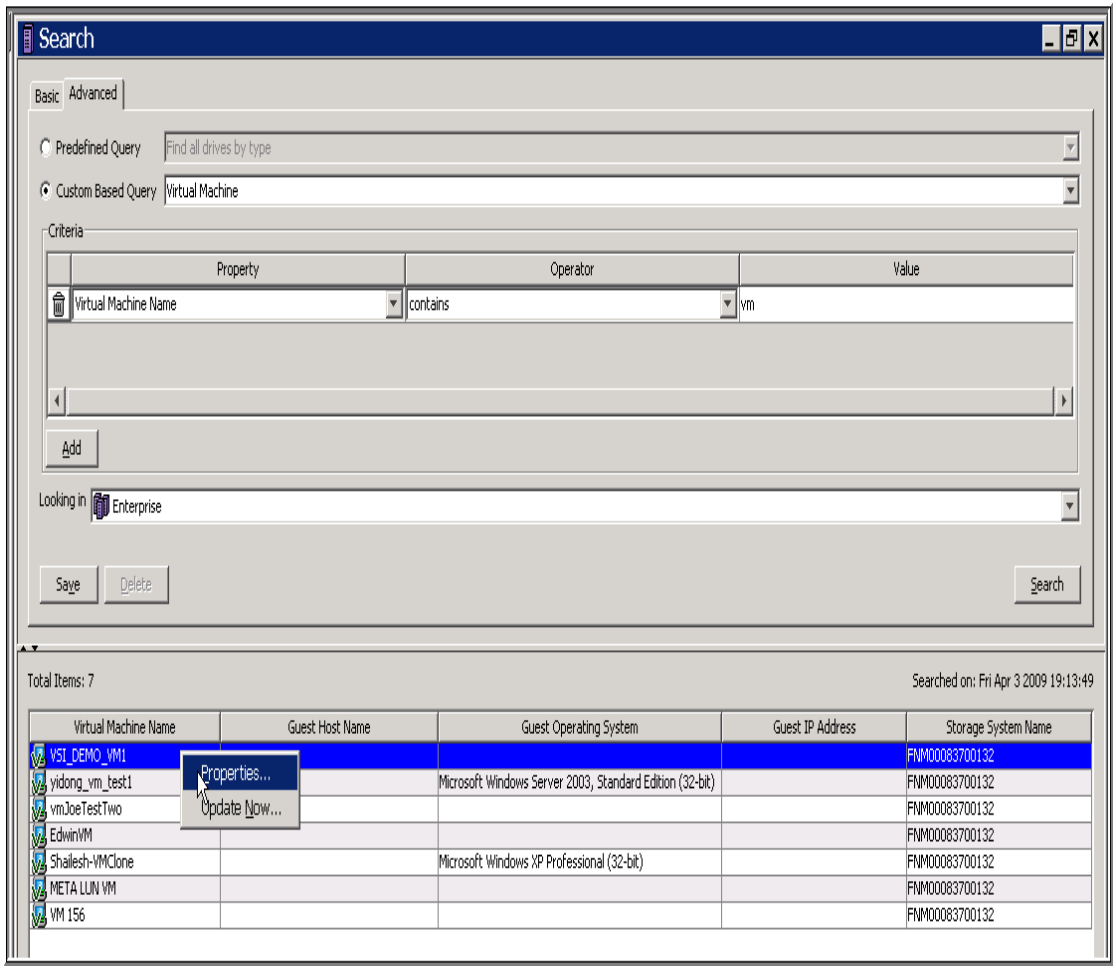


Figure 25 Search for a virtual machine using the Advanced Search function in Navisphere

Virtual Provisioning

Storage provisioning is the process of assigning storage resources to meet the capacity, availability, and performance needs of applications. With traditional provisioning, the amount of storage allocated to an application is equal to the amount of physical storage that is actually allocated for that application on the storage system.

With virtual or thin provisioning, *user capacity* (storage perceived by the application) is *larger* than the actual allocated space on the storage system. This simplifies the creation and allocation of storage capacity. The provisioning decision is not bound by currently available physical storage, but is assigned to the server in a capacity-on-demand fashion from a shared storage pool. The storage administrator monitors and replenishes each storage pool, not each LUN.

A thin storage pool must be created first by selecting a collection of disks (as shown in Figure 26 on page 70) using Navisphere Manager. As shown, you can set parameters in this dialog box and create alerts to help administrators monitor storage utilization for a given pool. For example, the *usable capacity* is the total physical capacity available to all LUNs in the pool. *Allocated capacity* is the total physical capacity currently assigned to all thin LUNs. *Subscribed capacity* is the total host reported capacity supported by the pool.

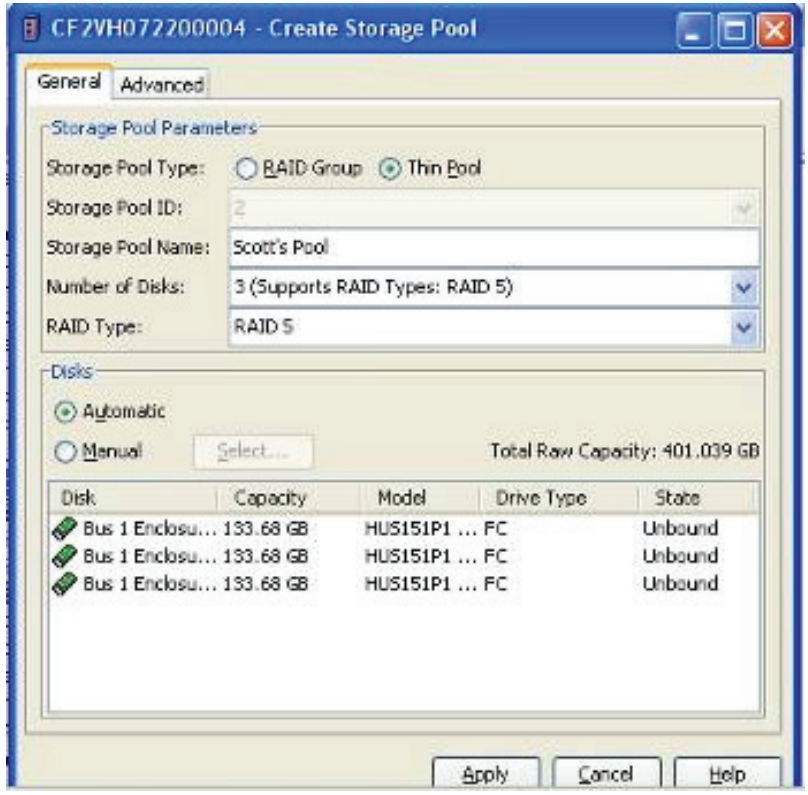


Figure 26 Create Storage Pool dialog box in Navisphere Manager

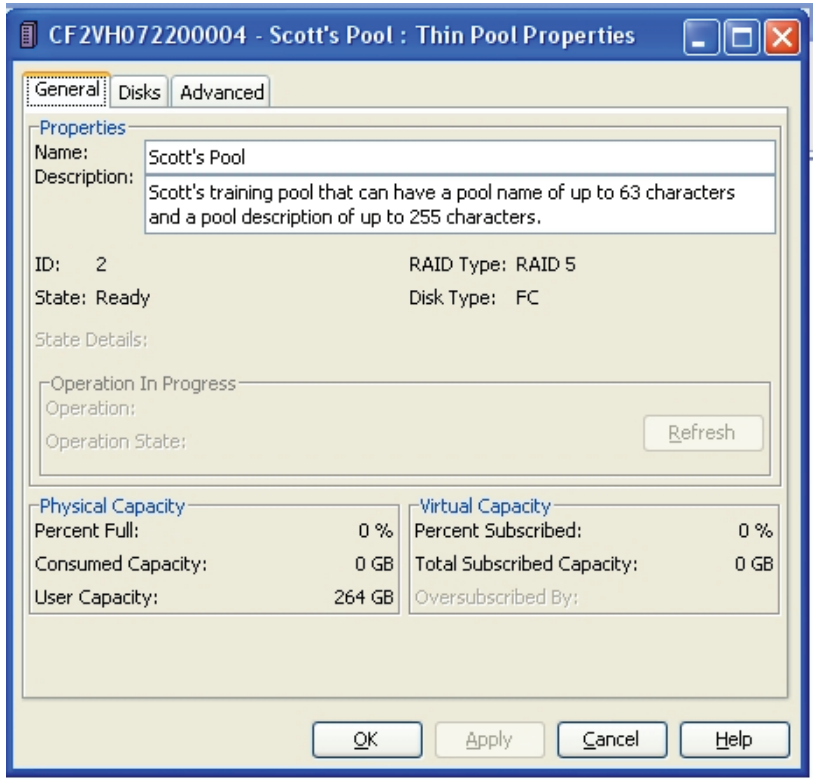


Figure 27 Thin Pool Properties dialog box

Next, a thin LUN can be created from the thin storage pool as shown in Figure 28 on page 72. This thin LUN can then be provisioned to a host. The primary difference between traditional and thin LUNs is that thin LUNs consume less physical space on the storage system. Other features such as Proactive Sparing are also available on thin LUNs.

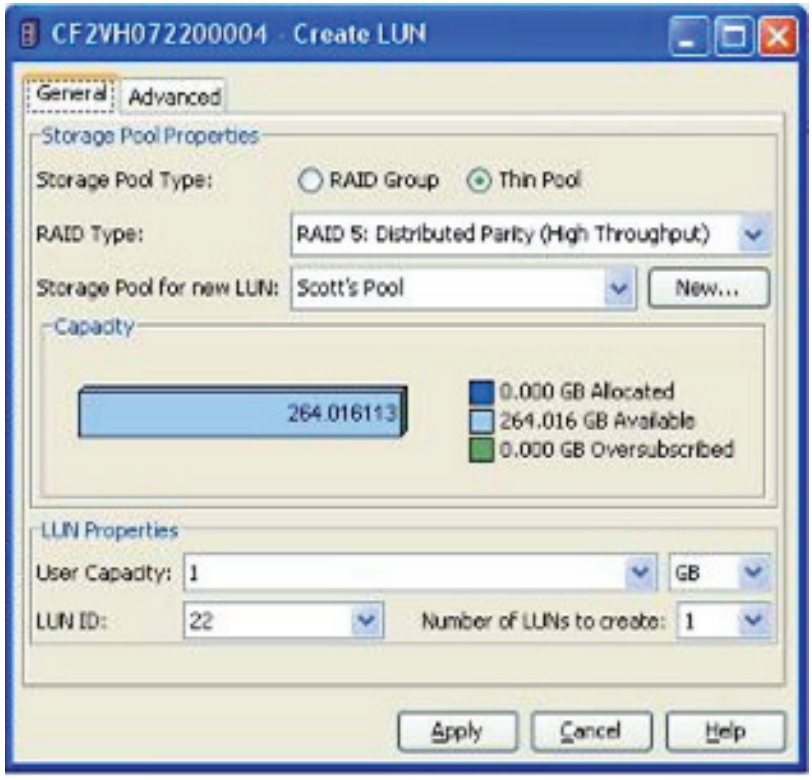


Figure 28 Creating a thin LUN using Navisphere Manager

Fully Automated Storage Tiering (FAST) LUN Migrator

The CLARiiON storage system offers support for three tiers of storage: Flash, Fibre Channel (FC) and SATA drives.

FAST LUN Migrator is a host-based application with two main components. The first component analyzes FC LUNs to make recommendations as to whether those LUNs should be migrated up to Flash drives, down to SATA drives or remain on FC drives. Upon acceptance of these recommendations, the second component automates the migration of LUNs to the appropriate tier.

The steps for using FAST LUN Migrator are as follows:

- Use Navisphere Analyzer to create a NAR (Navisphere Analyzer archive) file that is representative of your system's activity.
- Use FAST LUN Migrator to analyze the NAR file - execute the lunanalyze command to create an XML file and a CSV file. The XML file is a report of the analysis. The CSV file contains the information about which LUNs will be migrated.
- Review the XML report and edit the CSV file as desired.
- Use FAST LUN Migrator to perform the migrations specified in the CSV file - using the CSV file as input, execute the **lunassist** command to perform the migrations.

LUNanalyze

The lunanalyze command takes a Navisphere Analyzer archive (NAR) file as input and produces two outputs.

- An XML report that describes each LUN recommended for migration. This report can be viewed in a browser window.
- A CSV file listing the recommendations found in the report. You can edit this file in Excel or any other program that can edit a CSV file. The edited CSV is used as the input for the **lunassist** command.

Flash recommendations are broken into High, Strong, and Possible. "High" recommendations make excellent Flash candidates. They meet all the criteria of an ideal Flash candidate and reside on heavily utilized drives. "Strong" recommendations are very good candidates for Flash, but these candidates reside on less-utilized FC drives. Flash candidates listed as "possible" are highly utilized candidates that met some of the Flash criteria, but not all. The ways in which they do not meet the

criteria are listed in the report. These candidates can be added or removed from the CSV migration list to round out Flash RAID groups as desired.

SATA recommendations are broken into “Random” and “Sequential”. Random SATA candidates are LUNs with a random I/O pattern and very low LUN and disk utilization. LUNs on the sequential list are LUNs that take advantage of CLARiiON's optimization of large and sequential data streams aligning them with SATA characteristics.

The **lunanalyze** command can be run on any CX4 array running FLARE release 28 and later. Output of the LUN analyze command is show below, depicting analysis of the NAR file and making recommendations for the best storage tier for each LUN.

Version #	1
Array Name	Test
Array Serial #	FNM00083700132
Report Name	C:\EMC\FAST LUN Migrator\fastlun-1.xml
Report Time	Thu 1/14/2010 3:07:28 PM

Source LUN Name	Source LUN ID	Source RAID Group ID	Source LUN Capacity	Destination Recommendation	Destination Storage Pool ID
LUN 0	0	0	125.00 GB	EFD Highest	
LUN 1	1	0	1.25 GB	EFD Strong	
LUN 5	5	2	20.00 GB	EFD Possible	
LUN 2	2	0	2.50 GB	SATA Random	
LUN 3	3	0	2.50 GB	SATA Sequential	

Figure 29 Output of the LUN analyze command

LUNassist

The **lunassist** command takes the edited CSV file as input and sequentially migrates each LUN listed in the CSV automatically. This command can be run on any CX4 array running FLARE release 29 and later. Output of the **lunassist** command lists the migration status of the various LUNs.

Text (default)

```
Source LUN: 0
Destination Storage Pool: RAID Group 1
State: MIGRATED

Source LUN: 3
Destination Storage Pool: RAID Group 1
State: FAULTED –Error: The LUN is already in migration state


Source LUN: 34
Destination Storage Pool: RAID Group 1
State: MIGRATING - 25% complete, 11 minute(s) remaining


Source LUN: 99
Destination Storage Pool: Thin Pool 0
State: PENDING
```

Figure 30 Ouput of the lunassist command

For more details on the CLARiiON FAST LUN Migrator, please see the *EMC FAST LUN Migrator - Best Practice Planning* white paper available on Powerlink.

Navisphere Quality of Service Manager

Navisphere Quality of Service Manager (NQM) is a tool that runs on the CLARiiON storage system and measures and reports on storage system performance characteristics, and provides users the ability to set performance targets for high-priority applications or performance limits for lower priority applications.

The following outlines the configuration steps in setting up NQM:

1. **Creating I/O classes:** I/O classes are proxies for application profiles on the storage system. An I/O class can be specified on an application-by-application basis (LUNs or metaLUNs), I/O size (over 2 MB) or I/O type (read or write).
2. **Monitoring applications:** NQM monitors the current application service levels on the storage system and how applications are performing in relation to their service level requirements and overall storage system
3. **Setting goals and limits:** Service goals can be set for each I/O class. NQM monitors three key application characteristics throughput (IOPS), bandwidth (MB/s) and response time (ms). The service goal defined should be reasonable and match the performance characteristics of the application. A control method identifies how NQM enforces the service goals for an I/O class. The control method can be defined by a performance target for a high priority application or a limit to ensure the application does not exceed a certain service level.
4. **Creation of policies:** A policy is a group of I/O classes for which service level goals are enforced using a control method. All I/O classes within the policy must use the same control method. Multiple polices can be scheduled to enforce at different times, so that users can set different performance targets for the same application based on the time of day.

Once configured, NQM helps a user achieve application performance service level goals, and provides an application centric-view of performance on the storage system.

EMC SnapView

EMC SnapView provides two options for creating local replicas of LUNs:

- SnapView clones — Point-in-time full copies of a source LUN, synchronized incrementally after initial creation.
- SnapView snapshots — Pointer-based copies that store pointers to the original data changed from the source LUN, rather than a full copy of the data.

Clones and snapshots provide flexible options to meet different business needs. Depending on application requirements, a mixture of clones and snapshots may be deployed.

SnapView clones

A clone is a complete copy of a source LUN. To configure a clone, the user creates a clone group specifying the source LUN. When a LUN is added to the clone group, a block-by-block copy of the source LUN begins to clone in the clone group. While the clone is part of the clone group, it is not accessible to a secondary server. Any writes to the source LUN are also copied to the clone. Once the clone is fully synchronized, it can be fractured. Fracturing the clone suspends the relationship with the source LUN and makes it available to a secondary server. The source and clone must be equal in size to the source LUN. SnapView supports up to eight simultaneous clones of a source LUN.

Allocating a clone private LUN

When a clone is fractured from the source, clone private LUNs are used to persistently track changes to the source and clones, thus allowing incremental resynchronization. When configuring SnapView clone, the user must allocate one clone private LUN for each storage processor. Figure 31 on page 78 depicts the configuration of the clone private LUN using Navisphere Manager.

Note: Clone Private LUNs must be a minimum size of 250,000 blocks.

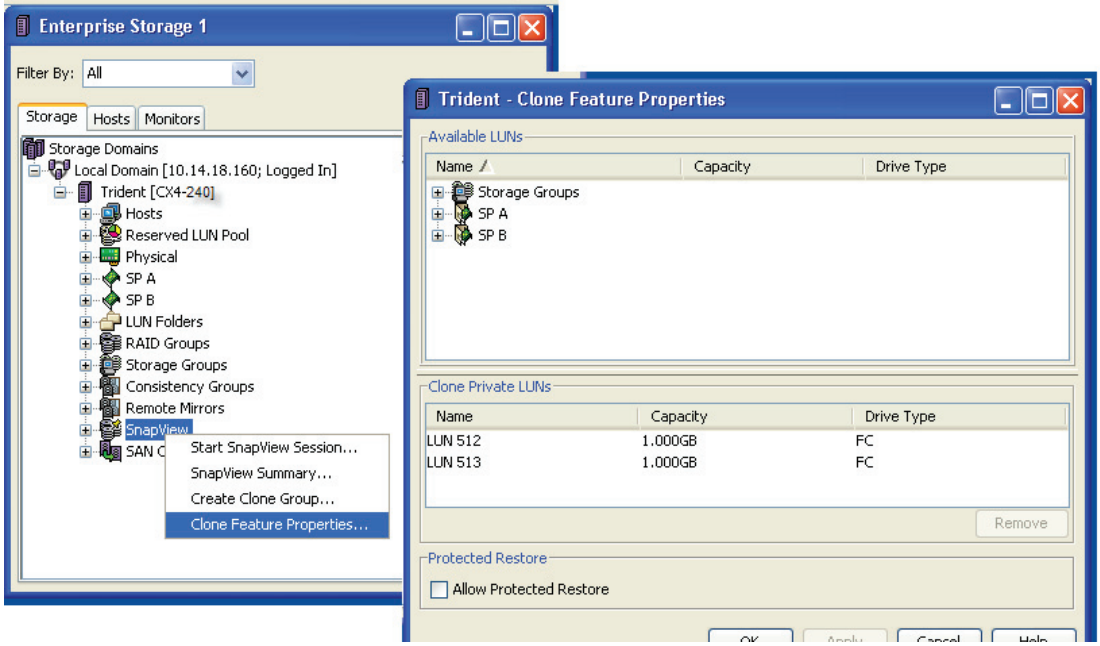


Figure 31 Allocation of a clone private LUN using Navisphere Manager

Clone creation

A clone source and associated clones are configured in clone groups. When configuring clones, the source LUN is identified and a clone group is created. Clones are then added to the clone group and data is synchronized from the source to the clone. Each source can have up to eight clones. A clone can be removed from a clone group as needed. When a clone is removed from the clone group, it retains the current data; however, all metadata used to track changes is removed from the Clone Private LUNs thus preventing incremental resynchronization. If a clone is removed from a clone group and later re-added, a full resynchronization is required.

Figure 32 on page 79 and Figure 33 on page 79 show the process for creating a clone group and adding clones to the clone group.

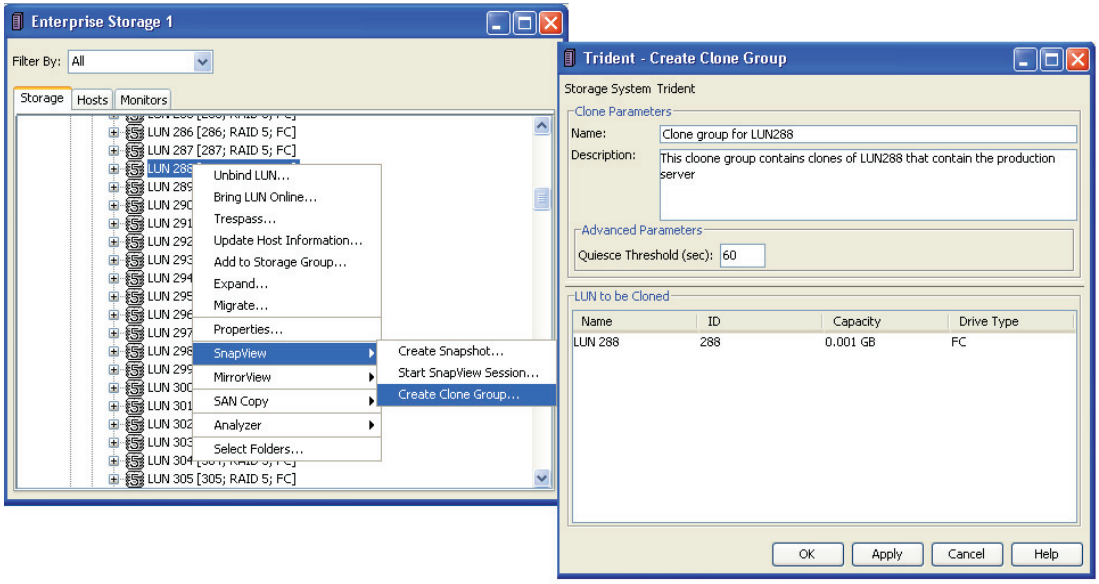


Figure 32 Creating a clone group using Navisphere Manager

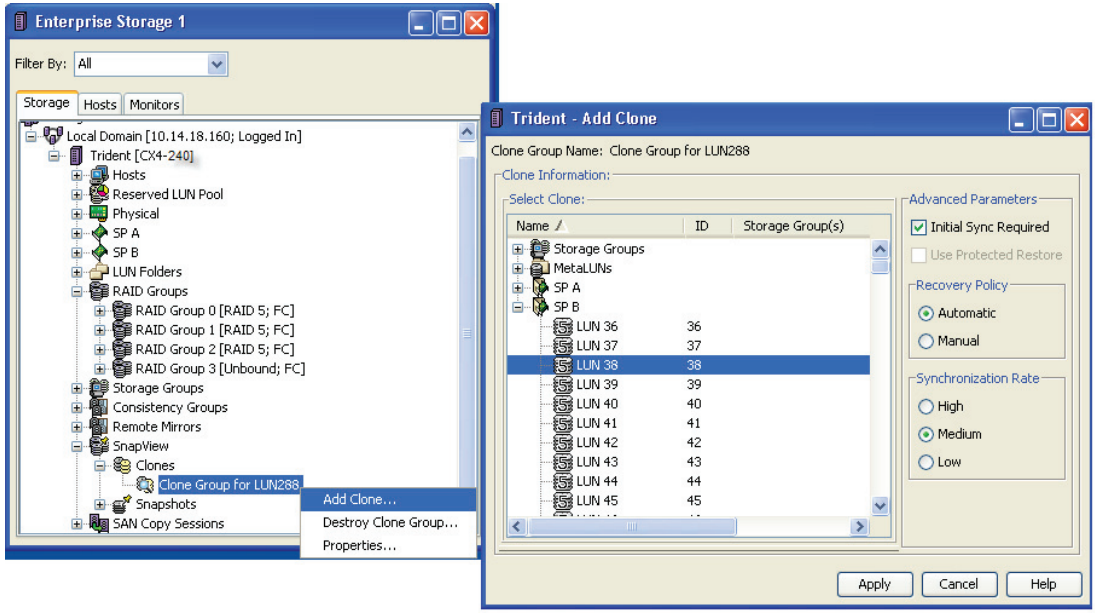


Figure 33 Adding clones to a clone group using Navisphere Manager

Clone synchronization operation

When a clone is added to a clone group, a full block-by-block synchronization is initiated. The synchronization rate can be specified to minimize the performance impact. Full copy is only required for the initial synchronization. Subsequent synchronizations involve only copying the data that has been changed since the clone was fractured. During synchronization, host I/Os to the source continue as usual. The clone, however, is not accessible to any host until synchronization completes and it is fractured from the source. Figure 34 on page 80 shows the properties view of a clone group while synchronization is in progress.

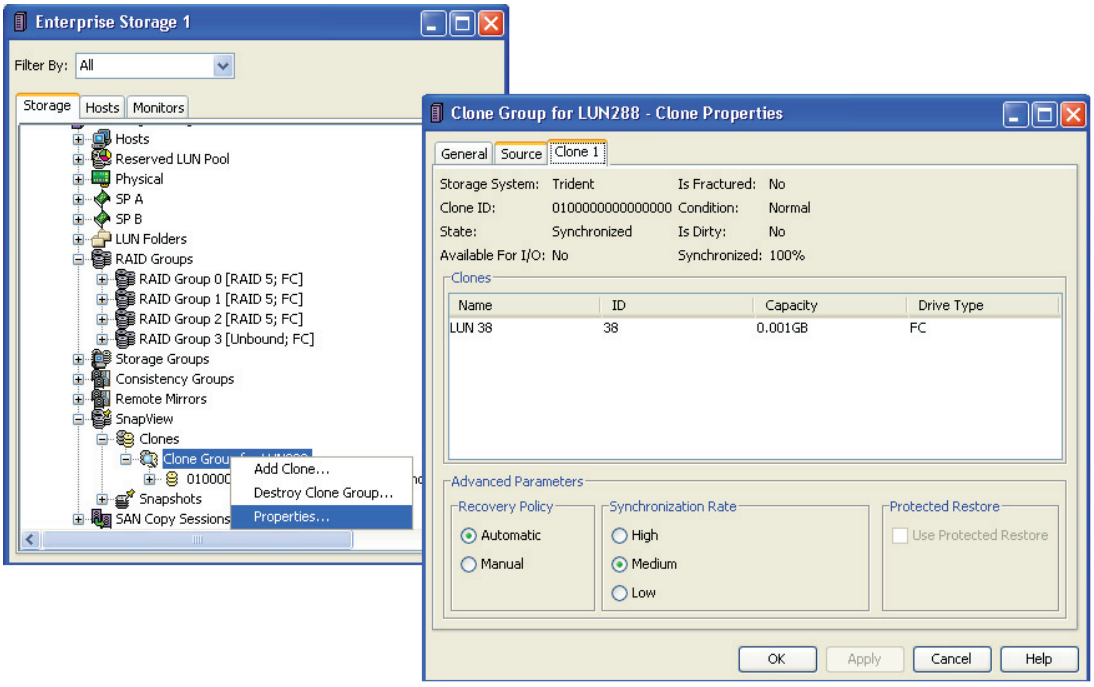


Figure 34 Properties view showing progress of clone synchronization

Clone fracture operation

Fracturing a clone detaches the clone from its source LUN and makes the clone available for host access. When fracturing a clone, I/O is briefly suspended to the source. To make the clone available for I/O by a secondary host, it is added to a storage group in the same way as any other LUN. Figure 35 on page 81 shows the clone fracture operation.

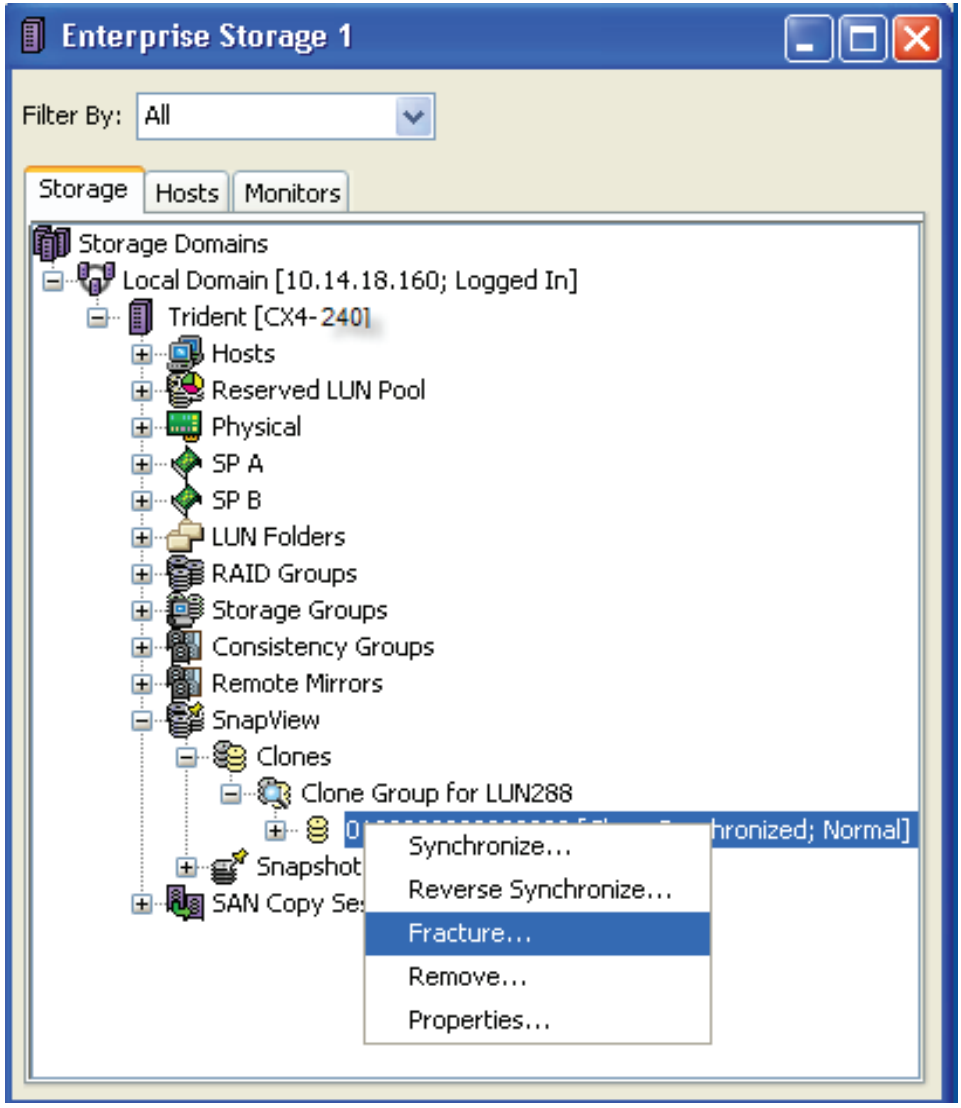


Figure 35 Fracturing a clone using Navisphere Manager

Clone consistent fracture operation

When a data set spans multiple LUNs, to ensure a consistent and restartable image of application data, all LUNs must reflect the same point-in-time. Clone-consistent fracture allows users to fracture multiple LUNs in a single operation with each reflecting the same point-in-time. Consistent fracture avoids inconsistencies and restart problems that can occur when fracturing multiple clones without quiescing or halting the application. During a consistent fracture operation, all writes to the source LUNs are appended until all clones are fractured. Figure 36 on page 83 illustrates a consistent fracture operation.

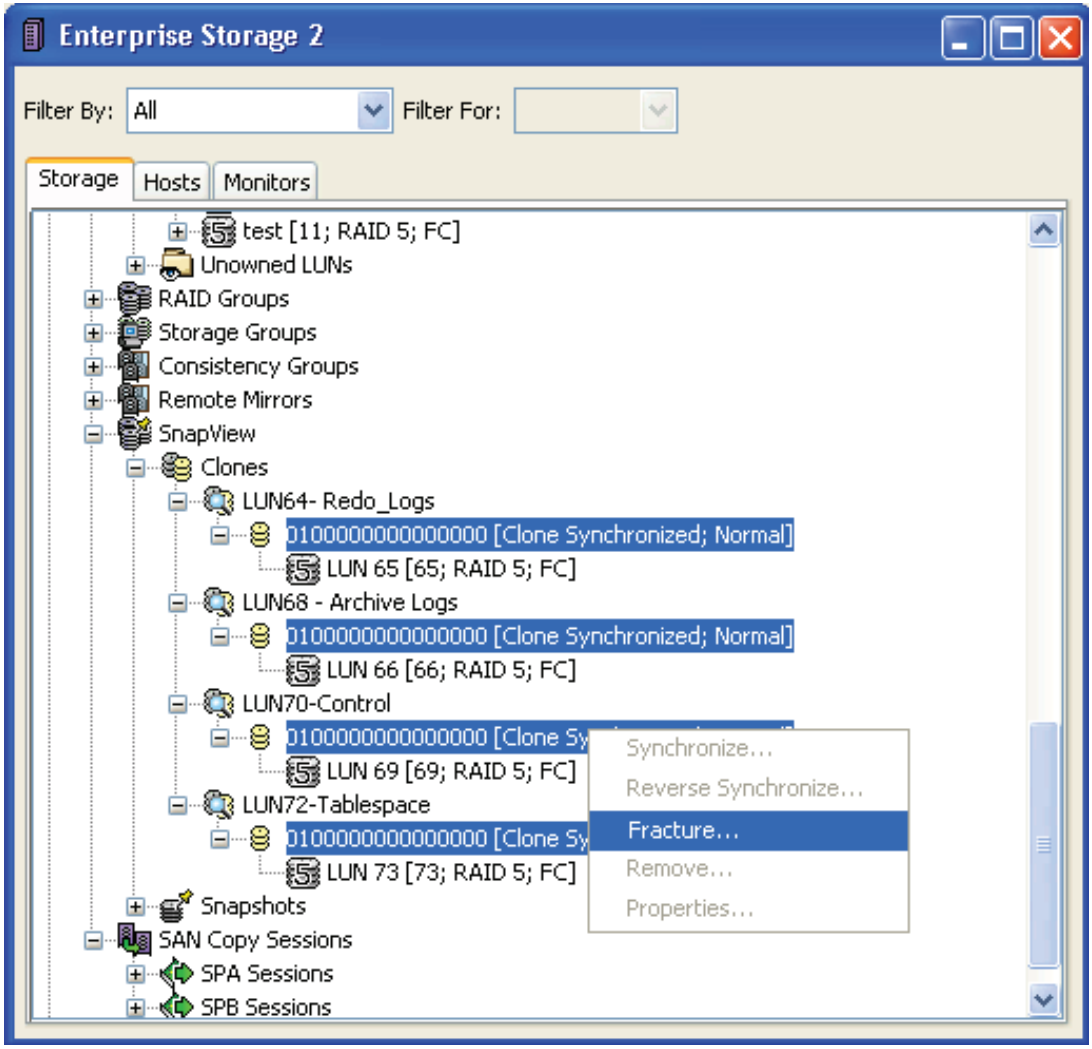


Figure 36 Consistent fracture operation using Navisphere Manager

Clone reverse synchronization

A benefit for clones is the rapid restore of the source LUN in the event of data corruption or accidental deletion. Clones are also useful in refreshing data in a test environment. A reverse synchronization operation incrementally copies the contents of the clone to the source

LUN. The incremental feature only copies tracks that have changed on the source and the clone since the clone was fractured. Figure 37 on page 84 illustrates a reverse synchronization operation.

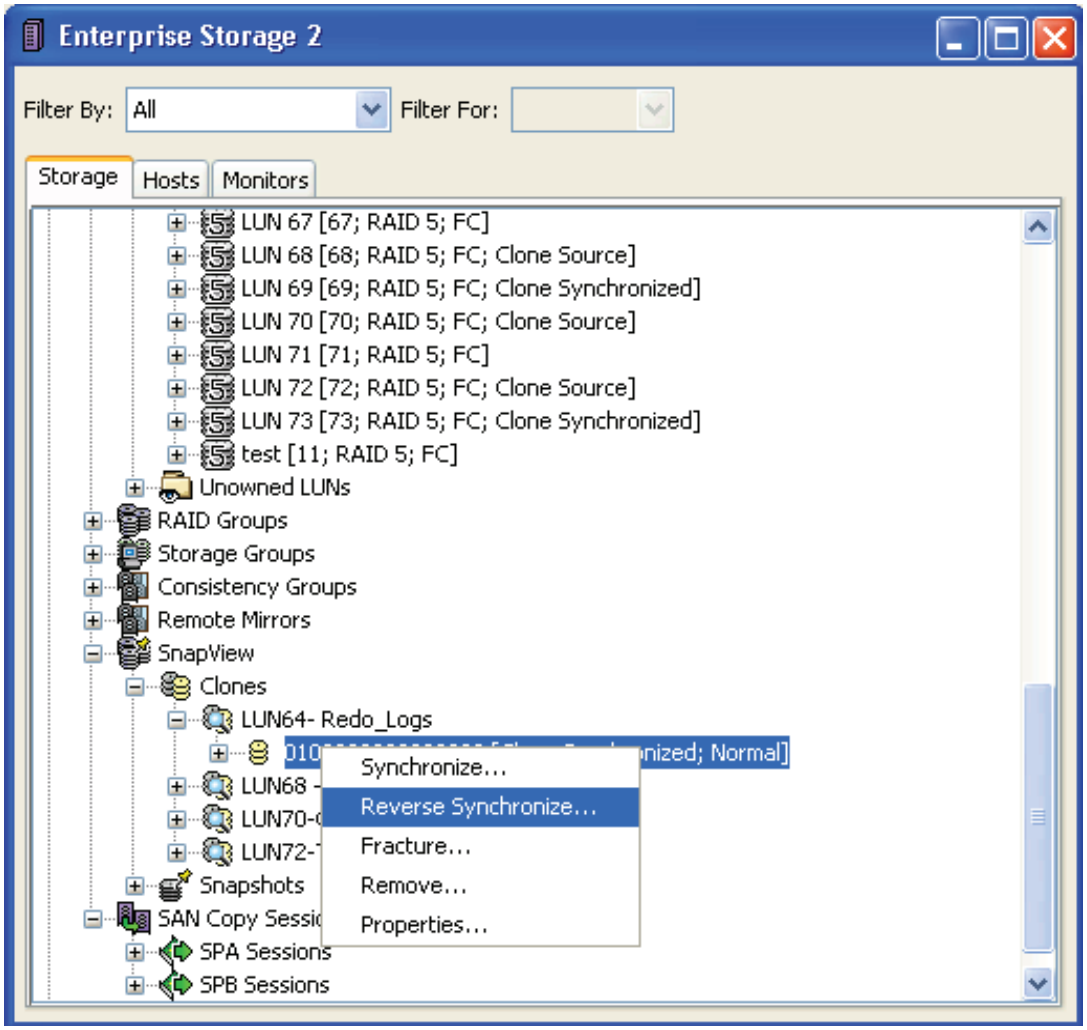


Figure 37 Clone reverse synchronization using Navisphere Manager

Note: An incremental restore of a clone volume to a source LUN is only possible when the two volumes have an existing clone relationship. To restore a clone to a different LUN requires that the clone first be removed from the clone group and a new clone group be created using the clone as the source.

Please note that most of the clone configuration steps mentioned above can be accomplished using the SnapView clone wizard available within Navisphere Manager, as shown in Figure 38 on page 85.

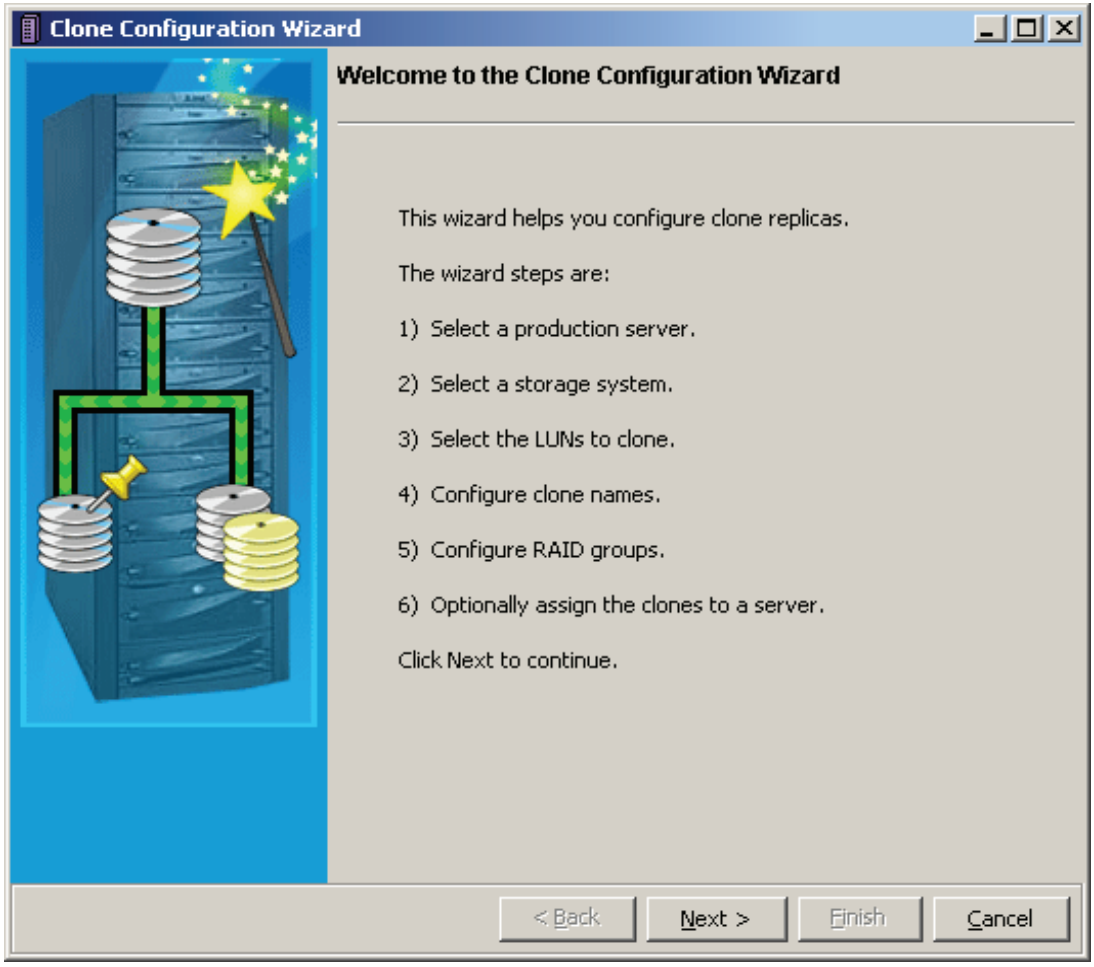


Figure 38 Configuring SnapView clones using the Clone Configuration Wizard

SnapView snapshots

SnapView snapshots provide a point-in-time view of a source LUN without the cost of a full LUN copy. Instead, SnapView snapshots use a copy-on-first-write technique to create a point-in-time view of a source LUN. There are two objects associated with a snapshot:

- Snapshot session — the snapshot session is a data structure that contains pointers that represent the point-in-time view of the source LUN at the time the session was started. The pointers either point to unmodified data on the source device or modified data that resides in the reserved LUN pool.
- Snapshot device — the snapshot device is a virtual device that can be added to a storage group and made visible to a host like any other LUN. However, there is no storage on the back end associated with a snapshot device. When a snapshot is activated, a snapshot session is associated with the snapshot device. Thus, the point-in-time image of the source LUN at the time the session was created is presented to the host.

An activated snapshot is available for I/O like any other LUN. However, SnapView maintains the original view of the data at the point-in-time the session was started. When the snapshot is deactivated, any changes made to the snapshot are discarded and the session once more represents the view of the source LUN at the time the session was created. When a session is stopped, all copy-on-first-write data in the reserve LUN pool is released.

SnapView maintains up to eight sessions per source LUN, each reflecting a different point-in-time. By activating, deactivating, and then reactivating using a different session, different point-in-time images of the source LUN can be presented to a host. In addition, up to eight snapshot devices could be created and each added to a storage group and presented to a different host. A session can only be activated by a single snapshot at a time. SnapView snapshot operations can be controlled using Navisphere Manager, Navisphere CLI, or the **admsnap** host command.

Reserved LUN pool

Before snapshot operations can be performed, a reserved LUN pool must be setup. The reserved LUN pool is used to store the copy-on-first write data. When a session is first started, a reserved LUN from the pool is assigned to source LUN. As data is modified on the source LUN,

the original view of the data is copied to the reserved LUN pool before modifying the data on the source. Additional LUNs from the reserved LUN pool can be allocated to the source LUN as required.

Typically, the reserved LUN pool contains multiple LUNs that are a fraction of the size of the source. Actual size of the reserved LUN pool depends on the write activity and the duration of the session. However, a good rule of thumb is for every source LUN is to create two reserved LUNs that are 10% of the size of the source LUN.

Note: The reserved LUN pool is also shared with incremental SAN Copy and MirrorView/Asynchronous. If these software technologies are to be deployed, additional planning is required to properly size the reserved LUN pool.

With release 24 of the FLARE code, the reserved LUN pool is global and shared by both storage processors. Figure 39 on page 87 shows the process for configuring the reserved LUN pool:

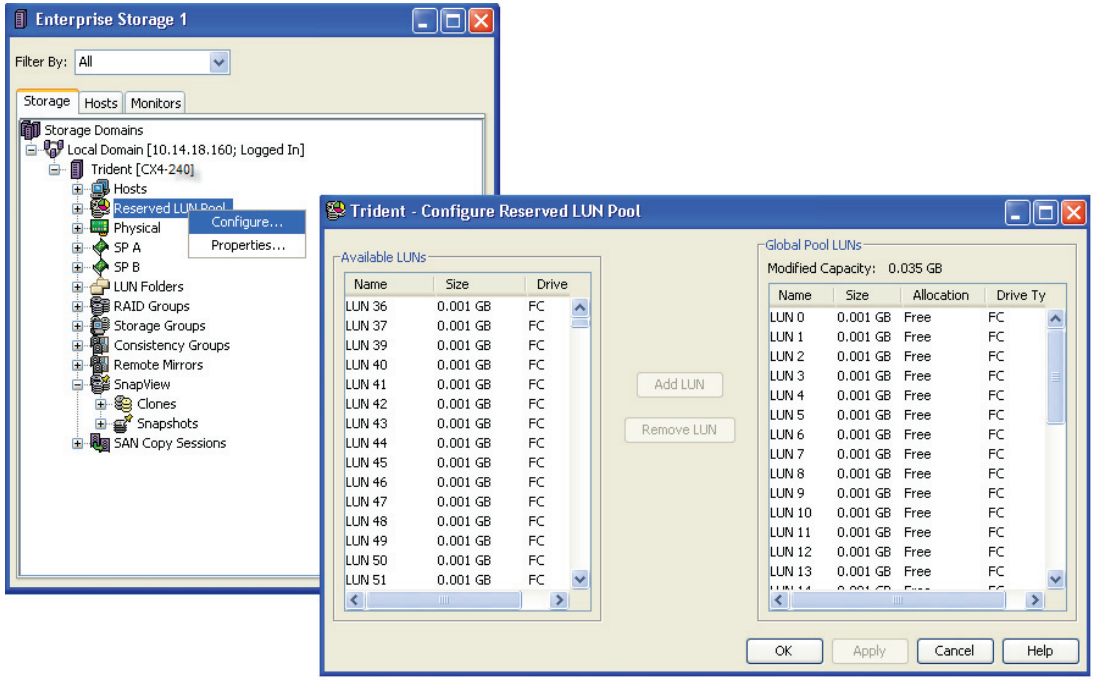


Figure 39 Configuring the reserved LUN pool using Navisphere Manager

Starting SnapView snapshot sessions

Creating a SnapView session captures a point-in-time image of the source LUN and begins the copy-on-first write operations. In the original implementation of SnapView, the user was given the option of creating non-persistent sessions. Using this option retained the pointer-information in the memory of the storage processor. Any disruption in the operation of the storage processor resulted in the loss of the snapshot session. In the latest implementation of SnapView, all sessions are persistent and the pointers are saved in the reserve LUN pool along with the copy-on-first-write data. In addition, SnapView software also provides the consistency option. This option is utilized when the application data set spans multiple source LUNs. When the consistency option is invoked when creating the SnapView snapshot session, the SnapView driver delays any I/O requests to the set of source LUNs until the session has started on all LUNs. This operation thus creates a dependent-write point-in-time of the data that can be restarted by the application. Figure 40 on page 88 shows the create session dialog.

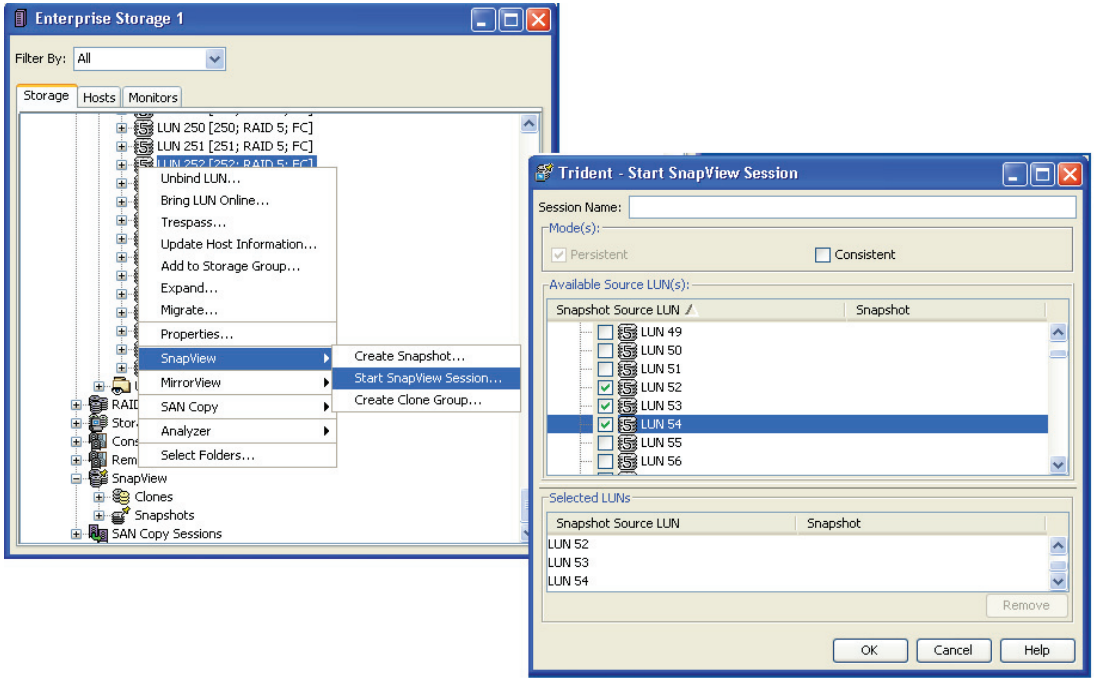


Figure 40 Creating a consistent SnapView session using Navisphere Manager

Restore (rollback) operations

A rollback operation allows the user to restore the snapshot session to the source LUN(s). If a snapshot is activated using the session, any changes made by the host will also be restored to the source. If there are no active snapshots using the session, the data on the source LUN is rolled back to the point-in-time that the session was first started.

Before starting the rollback operation, the user takes the source LUN offline momentarily to maintain data consistency. Once the rollback is started, the source LUN can be brought back online and it may be accessed as the restore takes place in background.

When rollback operation is initiated, the user is given the option of starting a recovery session. A recovery session allows the user to undo the rollback operation. The recovery session contains the point-in-time view of the data on source LUN(s) before the rollback is started and provides the option to rolling the image forward again to the state it was before to the rollback.

Please note that most of the SnapView snapshot configuration steps mentioned above can be accomplished using the SnapView Snapshot wizard available within Navisphere Manager as shown in Figure 41 on page 90.

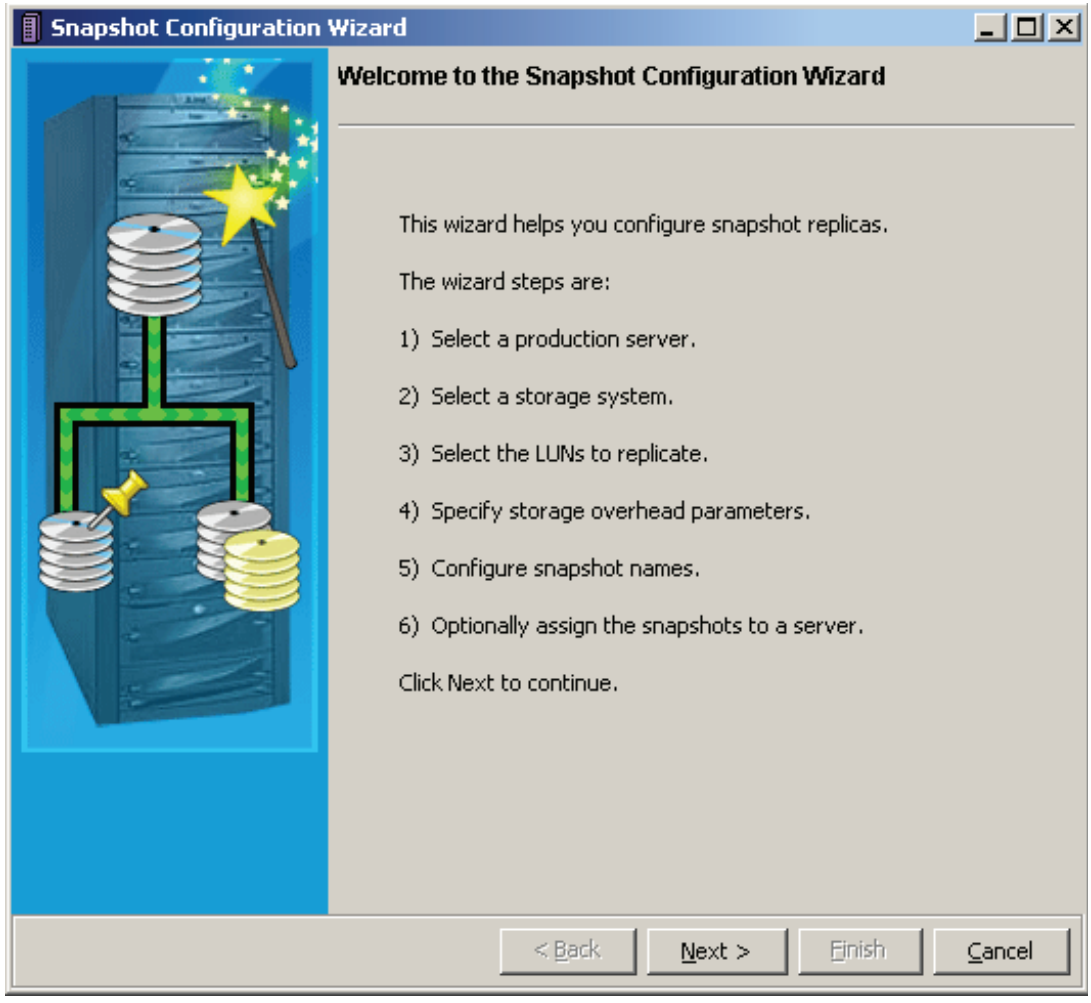


Figure 41 Configuring SnapView snapshots using the Snaphot Confiuguration wizard

EMC SAN Copy

EMC SAN Copy is CLARiiON-based software that enables LUNs to be copied to or from the same or different CLARiiON, Symmetrix, or qualified non-EMC storage systems. After the copy session is configured, the data movement is directly between storage systems and no host resources are involved. The communications between arrays is through a Fibre Channel or IP SAN, and the distance between the source and destination is limited only by SAN connectivity. SAN Copy is an ideal solution for data mobility and data migration applications due to the heterogeneous storage systems support.

SAN Copy software is installed on either the source or the target array. The underlying architecture is based on the standard SCSI initiator and target model. When a CLARiiON is configured for SAN Copy, a front-end port on the storage processor is configured to emulate a SCSI initiator device. When doing a push operation, the CLARiiON acting like an initiator, reads from a local LUN and writes to a remote LUN, which is the target. When doing a pull operation, the CLARiiON reads from a remote LUN and writes locally. SAN Copy appears like a host system to the remote storage system.

Setting up a SAN Copy environment is similar to configuring host connectivity. There must be a physical connection between the SAN Copy CLARiiON front-end port and the front-end port of the remote storage system. This is configured through direct cable connections or through a storage fabric (Fibre Channel or IP). In addition, appropriate operations need to be performed on the remote storage array to ensure that the remote LUN is accessible to the SAN Copy front-end port on the CLARiiON. If the remote LUN is also a CLARiiON, this is accomplished by creating a storage group, adding the target LUN to the storage group, and connecting the WWPN or IQN of the SAN Copy port to the storage group.

The relationship between the source storage array and LUN, and the remote storage array and LUN is defined by creating a SAN Copy session. Navisphere Manager has an easy to use wizard that steps a user through the process. Figure 42 on page 92 shows the opening dialog for the wizard to configure a SAN Copy session.

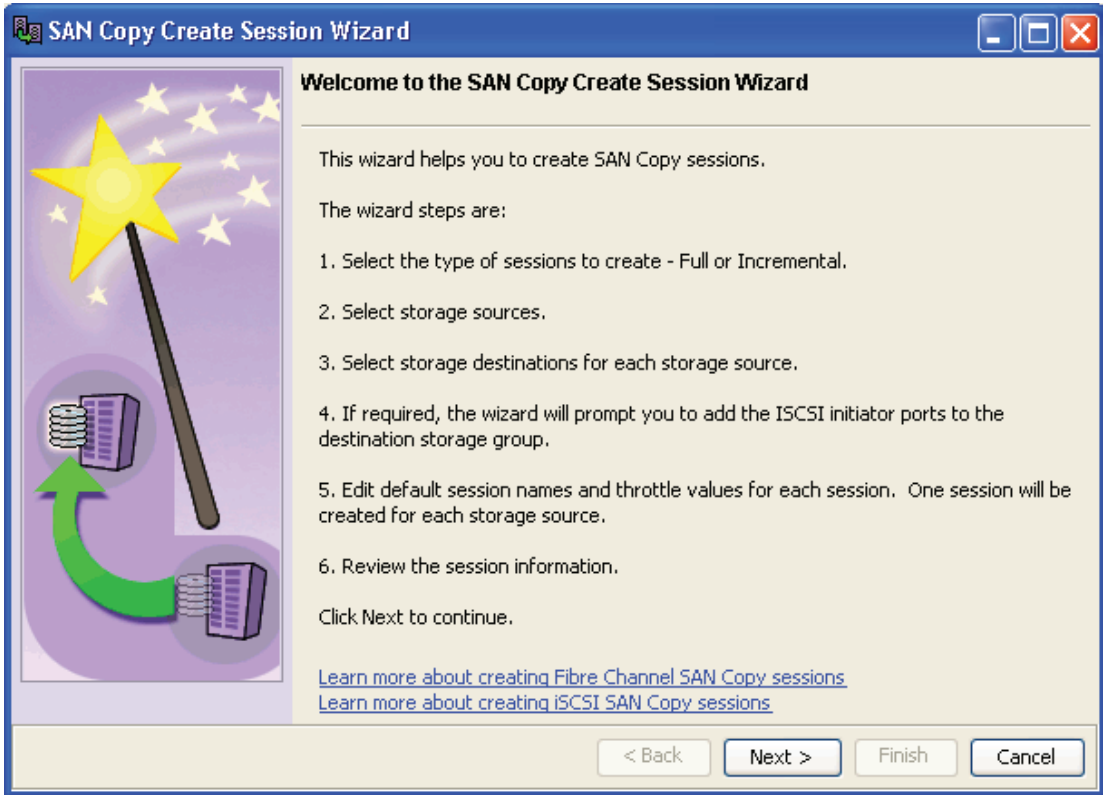


Figure 42 SAN Copy Create Session Wizard

SAN Copy supports full copy push or pull operations and incremental push operations. During a full SAN Copy session, the data on the source LUN must not change or the resulting copy on the remote storage system will be inconsistent. The best practice is to take either the source LUN offline, or uses SnapView to make an image of the source and use it as the SAN Copy source LUN. With incremental push operations, SAN Copy automatically creates a snapshot session and

uses it as the source of the copy operation. In both incremental and full SAN Copy, the target device is not a complete and consistent copy and should not be accessed until the copy session completes.

SAN Copy sessions are managed using Navisphere Manager or CLI. In addition, the `admhost` utility can be leveraged in a Windows environment to perform prerequisite tasks that are necessary to ensure a complete and consistent copy of the data. The `admhost` utility is used to activate and deactivate LUNs by assigning and removing drive letters and for flushing host buffers.

A SAN Copy source and target LUN could be within the same CLARiiON system. This might be useful for migrating a LUN between different RAID groups, protection types, drive architecture, expanding the LUN size or creating copies. However, the Navisphere Virtual LUN or SnapView technology discussed in "CLARiiON Virtual LUN technology," on page 61 and "EMC SnapView," on page 77 respectively, provides a more appropriate technique.

SAN Copy requirements

To copy LUNs between CLARiiON storage systems, or between CLARiiON and Symmetrix or third-party storage systems, the following requirements must be met:

- Either the source LUN, the destination LUN, or both must reside on a CLARiiON System with the SAN Copy software feature enabled.
- The destination LUN must be equal or greater in size than the source LUN.
- If the remote storage system is a CLARiiON, a storage group must be created, target LUNs added, and the SAN Copy front-end ports connected to allow SAN Copy full read or write access.
- If the remote array is a Symmetrix or third-party storage system, the remote LUN must be LUN masked to allow the front-end port on the SAN Copy CLARiiON full read or write access.
- Connectivity between the SAN Copy front-end port and the front-end port on the remote array must be configured. Typically, this involves physical cabling and/or fabric zoning.

For incremental copy sessions, the reserved LUN pool must be configured as incremental SAN Copy leverages SnapView snapshot sessions to allow continuous access to the source device during the copy operation.

EMC MirrorView

EMC MirrorView is a CLARiiON business continuity solution that provides LUN-level data replication to a remote CLARiiON. The copy of the data on the production CLARiiON is called the primary image whereas the copy at the recovery site is called the secondary image. During normal operations, the primary images are online and available for read or write operations, and the secondary image is not ready. The write operations to the primary image are mirrored to the secondary. MirrorView provides synchronous and asynchronous replication options. These are separately licensed features:

- **MirrorView/Synchronous (MirrorView/S)** — provides real-time mirroring of data between the primary CLARiiON systems and the secondary CLARiiON systems. Data must be successfully stored in both the local and remote CLARiiON units before an acknowledgment is sent to the local host. This mode is used mainly for campus or metropolitan area network distances of less than 200 km.
- **MirrorView/Asynchronous (MirrorView/A)** — maintains a dependent write-consistent copy of data at all times across any distance with no host application impact except during the update. MirrorView/A transfers data to the secondary storage system in predefined timed cycles or delta sets. Before each update, a point-in-time copy of the secondary image is automatically created. In the event that an update is unable to complete, the secondary image can be rolled back to the last consistent state. Cycle times are dependent on write activity and network bandwidth but typically can be configured to meet Recovery Point Objectives in minutes to hours.

Configuring MirrorView

MirrorView/S and MirrorView/A are configured and managed using Navisphere Manager and Navisphere CLI. While MirrorView/A and MirrorView/S are separate products, they are configured in a similar manner. Note that most of the configuration steps mentioned below can be accomplished using the MirrorView wizard available within Navisphere Manager as shown in Figure 43 on page 95.

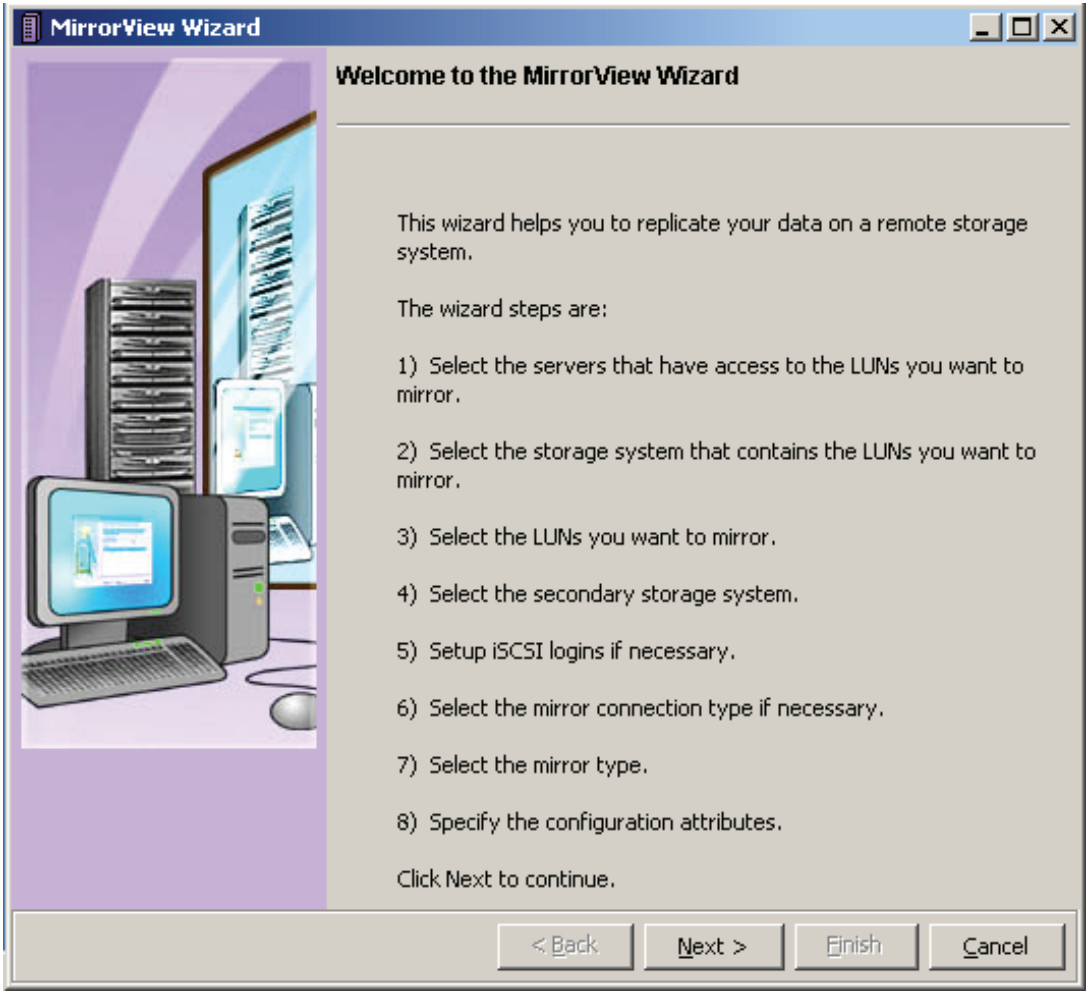


Figure 43 Configuring MirrorView using the Navisphere Manager MirrorView wizard

However, this document highlights the configuration steps without using the MirrorView wizard. The first step is to ensure that the MirrorView feature is enabled on both the primary and secondary array, and there is connectivity between the CLARiiON front-end ports. Normally, the high number port on both SP-A and SP-B are used for MirrorView. The physical connection could be a direct cable between the ports or through a SAN fabric that have been zoned to enable the storage processor ports to communicate.

Next, a MirrorView Connection must be defined. Figure 44 on page 96 is an example of the Manage Mirror Connection dialog. On the right is a list of CLARiiON systems that have MirrorView software enabled and have connectivity. To enable a remote CLARiiON for MirrorView, select it and click Enable. It will move to the panel on the right.

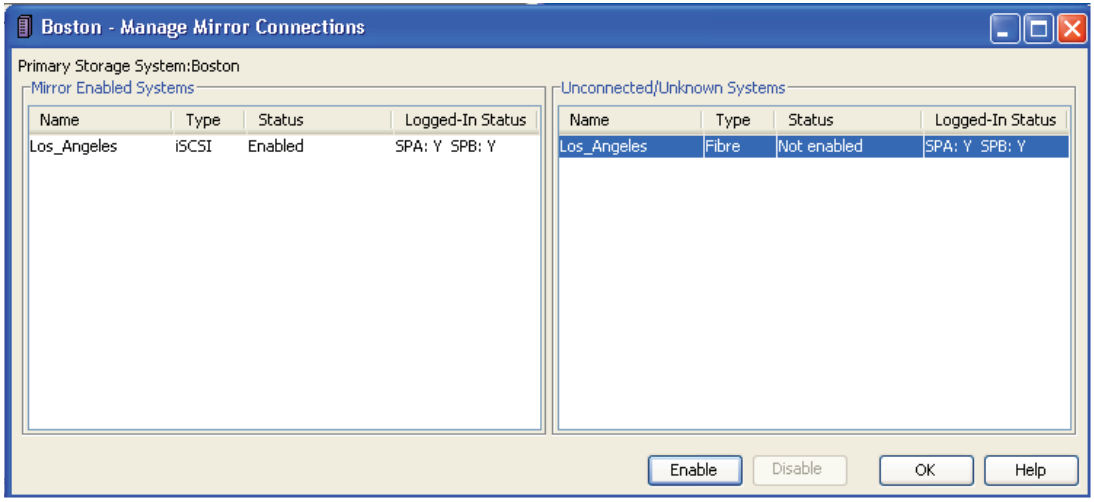


Figure 44 Manage Mirror Connections dialog box

When changes to the primary image cannot be propagated to the secondary image, the secondary image is considered to be in a fractured state. This state could be because of an administrative action or because of communication failure. Either way, changes that are not propagated are tracked. The SnapView snapshot session technology is used to track changes when MirrorView/A is deployed. With MirrorView/S, either a memory resident fracture log or a disk resident write intent log (WIL) is used. The best practice is to use the write intent log, as this is persistent across storage processor failures. write intent logs are configured by designating a LUN for that purpose. Figure 45 on page 97 shows the process for creating the write intent log.

Note: Two WIL LUNs each at least 128 MB should be allocated, one for each storage processor.

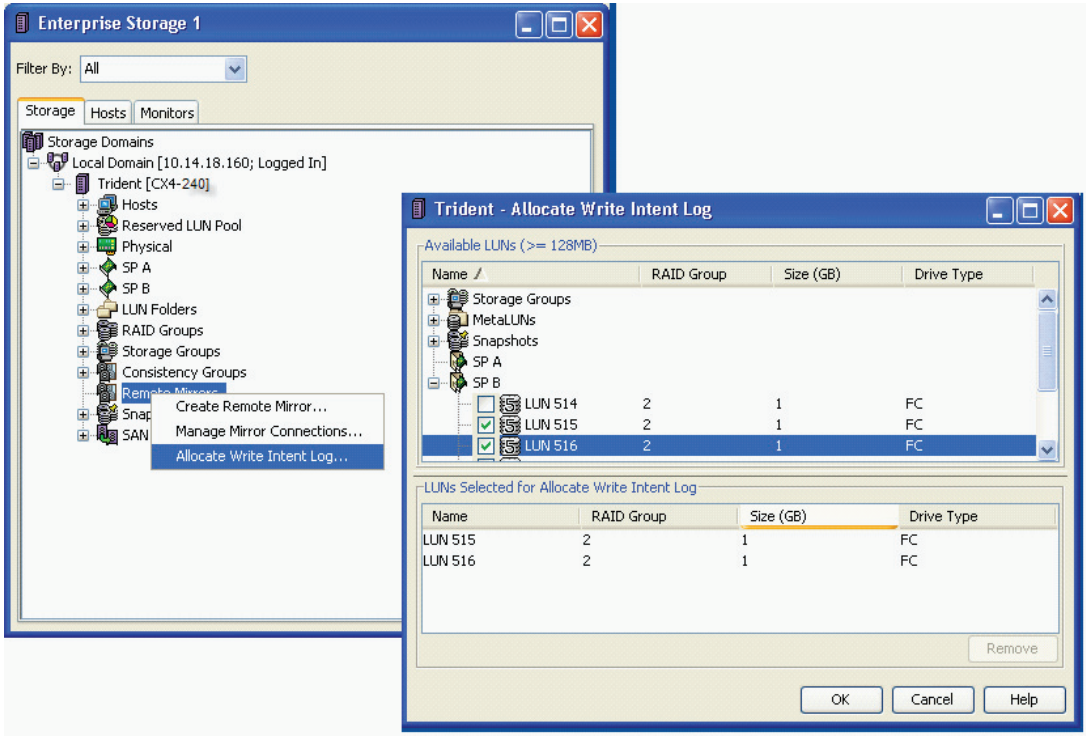


Figure 45 Allocate Write Intent Log dialog box

The next step in the process is to designate a LUN as a primary image. This makes a regular LUN capable of having a secondary image. Figure 46 on page 99 depicts the process for creating a primary image. When creating a remote mirror, specify the mode of operation – synchronous or asynchronous replication. For synchronous, also specify whether to use the write intent log.

The final step is to add a secondary image to remote mirror. With MirrorView/A, a primary image can only have one secondary image. With MirrorView/S, it is possible to have one or two secondary images. Figure 46 on page 99 shows the dialog for adding a remote mirror.

Note: The secondary image may be of any protection type but must be exactly the same size as the primary image. The initial synchronization will consume system resources, the performance impact can be minimized by selecting the

Low option for the Synchronization Rate (see Figure 18 on page 59). After intial sync, users should change the sync rate to medium or high to transfer updates quickly at the secondary site.

After the secondary Images are added to the remote mirror, it will immediately begin synchronization. The initial synchronization will be a full LUN copy.

Mirror states

A remote mirror will be in one of the following states:

- *Synchronizing* – A data copy is in progress from the primary image to the secondary image.
- *Synchronized* – Secondary image is identical to the primary image.
- *Consistent* – Secondary image is identical to the primary image or to some previous instance of the primary image. This means the secondary image is available for recovery when a user promotes it.
- *Out-of-Sync* – The secondary image needs synchronization with the primary image. The secondary image is not available for recovery.

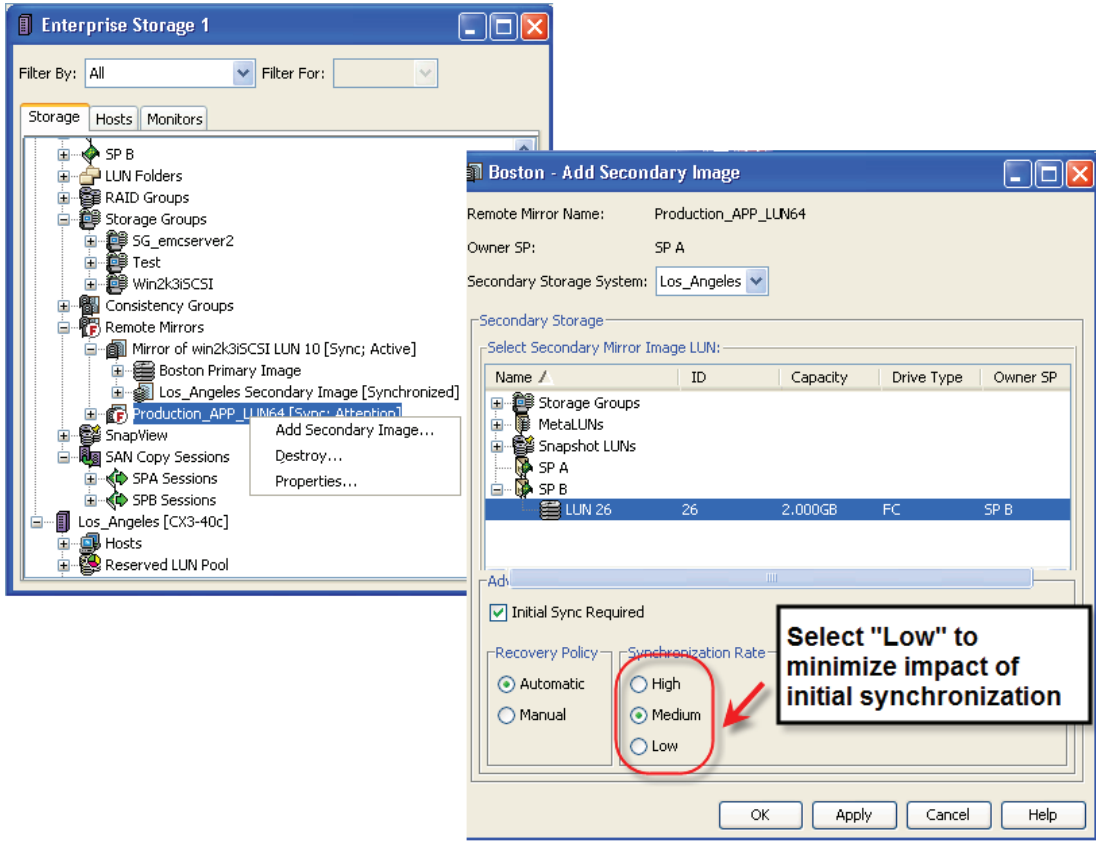


Figure 46 Adding a secondary image using Navisphere Manager

MirrorView operations

Operations that can be performed on a remote mirror include synchronize, promote, remove, and fracture. The following describes each of these operations.

- *Synchronization Operation:* Initial synchronization occurs when a secondary image is added to a remote image. During normal operations, changes to the primary image will either synchronously or asynchronously be copied to the secondary image. If the primary image loses contact with the secondary image or an administrator fractures the relationship. Normal mirroring operations can be resumed by performing a Synchronize operation.

- *Promote Operation:* The promote operation is performed in the event of the disaster or when it is necessary to move the workload to the remote site. The promote operation swaps the role of the primary and secondary image. The secondary image is promoted to be the primary image, and the primary image becomes the secondary image. The old primary is made not ready (NR), the new primary (the old secondary) becomes available for read or write operations. Furthermore, the direction of synchronization is reversed. A second promote operation restores the original mirror relationship.
- *Fracture Operation:* A fracture operation suspends the mirror relationship. With MirrorView/S changes to the primary image are tracked in the fracture log or write intent log. SnapView snapshot session technology is leveraged to track changes when MirrorView/A technology is used. When a remote mirror is in a fractured state, a synchronize operation returns the relationship to normal.
- *Remove Operation*. A remove operation converts a secondary image into a regular LUN. All tracking information is discarded. A full synchronization is required to reestablish the relationship between the same pair of LUNs.

Note: MirrorView/A leverages SnapView snapshots and clones and SAN Copy technologies. The reserved snapshots and SAN Copy sessions used by MirrorView are displayed when viewing SnapView and SAN Copy within Navisphere Manager.

MirrorView consistency groups

When an application or a group of related applications span multiple LUNs, it is critical that all remote images of LUNs reflect the same point-in-time to maintain write-order consistency. This is critical to not only ensure restart of the application but also integrity of the business process. MirrorView consistency groups allow an administrator to logically group remote mirrors together and perform operations on all images in a single operation. In addition, if a primary LUN in the consistency group cannot propagate changes to its corresponding secondary LUN, MirrorView suspends data propagation from all LUNs in the consistency group. This suspension ensures a business process consistent, dependent write-consistent copy of the data on the secondary storage system.

Using snapshots and clones with MirrorView

When the secondary image is not ready to the attached host, SnapView can be used to create replicas of the MirrorView secondary image, which can be used to perform backup, data verification, or other parallel processing tasks. Point-in-time replicas of MirrorView primary images can also created using SnapView snapshots and clones.

MirrorView Insight for VMware (MVIV)

MirrorView insight for VMware (MVIV) is a new tool bundled with the MirrorView Site Recovery Adapter (SRA) that complements the Site Recovery Manager framework by providing failback capability for test purposes. MVIV also provides detailed mapping of VMware filesystems and their replication relationships. More details for using MVIV with SRM is provided in Chapter 6 of this TechBook.

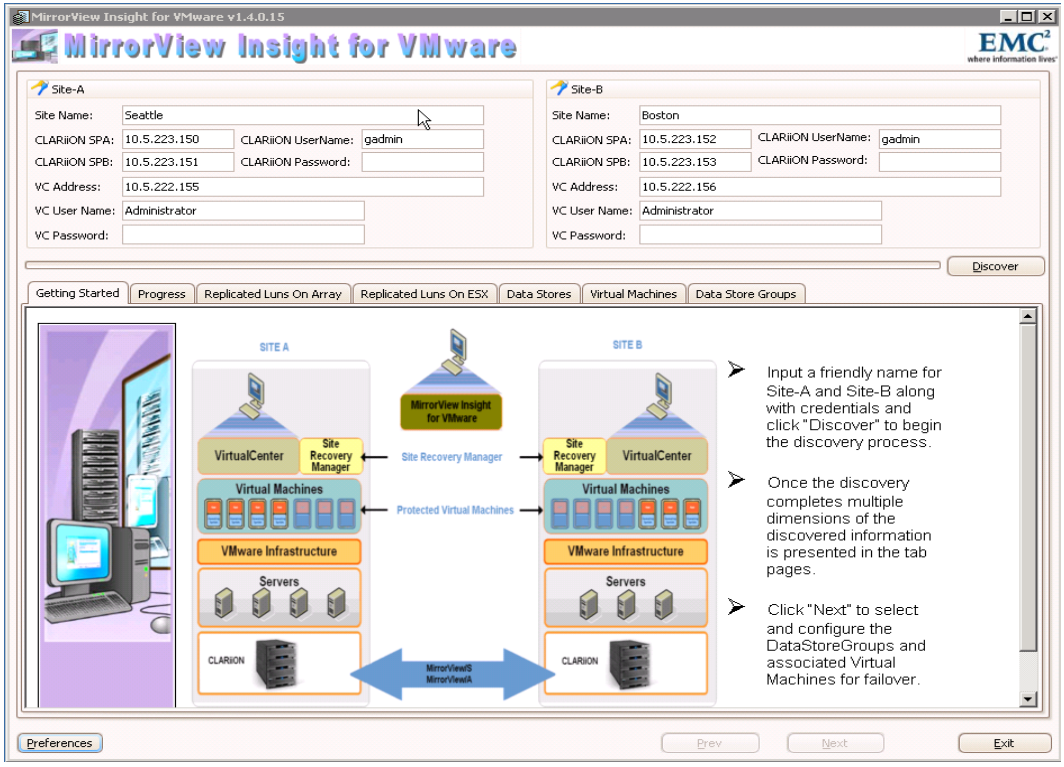


Figure 47 MirrorView insight for VMware (MVIV) framework

EMC RecoverPoint

RecoverPoint is an appliance-based DR solution. RecoverPoint supports both local and remote replication. It supports local replication with full binary copies. For remote replication it offers a zero data loss synchronous option as well as an asynchronous option. RecoverPoint should be considered when:

- A large number of replicas are required (100s-1000s)
- Heterogeneous storage system support is required
- The ability to switch between sync and async replication on the fly is required
- Only seconds of data loss is required for asynch replication

RecoverPoint offers many other benefits such as supporting multiple points in time for recovery and WAN optimization features such as compression and deduplication. RecoverPoint also provides a scripting interface that enables business applications to integrate RecoverPoint into existing application protection and recovery processes.

CLARiiON splitter support

The RecoverPoint write-splitting model can reside on a switch, host or CLARiiON CX4 or CX3 storage system. The array-based splitter runs in each storage processor and will split (i.e. mirror) all writes on the CLARiiON LUN, sending one copy to the target and one copy to the RecoverPoint appliance.

VMware affinity

RecoverPoint integrates with VMware vCenter and provides discovery of ESX servers and their associated virtual machines. It will also display the mapping of virtual machine storage resources to RecoverPoint consistency groups. The RecoverPoint protection status of each virtual machine is displayed in the GUI as shown in Figure 48 on page 103 and is also available via command line utilities. An alert is generated with the protection status of a virtual machine changing.

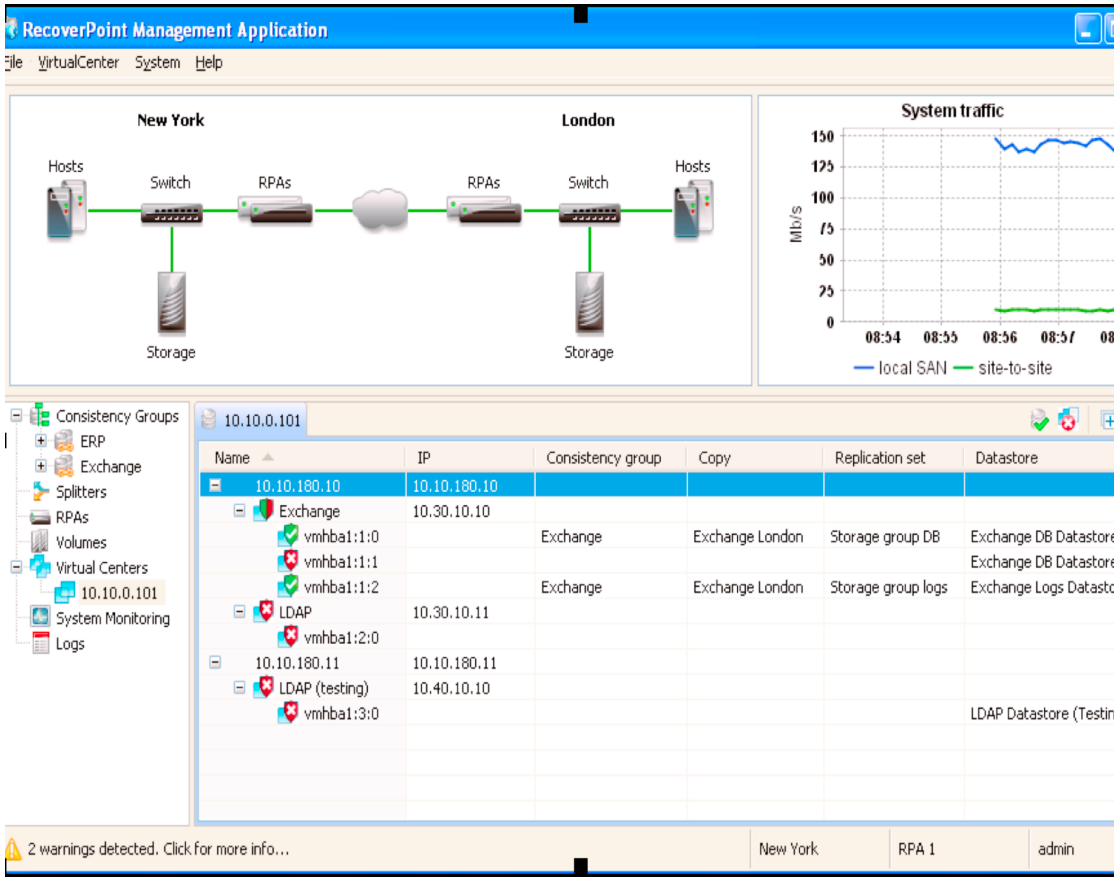


Figure 48 RecoverPoint vCenter integration

For more information see the *EMC RecoverPoint Family Overview* white paper on Powerlink.

EMC PowerPath

EMC PowerPath is host-based software that works with networked storage systems to intelligently manage I/O paths. PowerPath manages multiple paths to a storage array. Supporting multiple paths enables recovery from path failure because PowerPath automatically detects path failures and redirects I/O to other available paths. PowerPath also uses sophisticated algorithms to provide dynamic load balancing for several kinds of path management policies that the user can set. With the help of PowerPath, systems administrators are able to ensure that applications on the host have highly available access to storage and perform optimally at all times.

A key feature of path management in PowerPath is dynamic and multipath load balancing. Without PowerPath, an administrator must load-balance paths to logical devices statically to improve performance. For example, based on current usage, the administrator might configure three heavily used logical devices on one path, seven moderately used logical devices on a second path, and 20 lightly used logical devices on a third path. As I/O patterns change, these statically configured paths may become unbalanced, causing performance to suffer. The administrator must then reconfigure the paths, and continue to reconfigure them as I/O traffic between the host and the storage system shifts in response to use changes.

Designed to use all paths concurrently, PowerPath distributes I/O requests to a logical device across all available paths, rather than requiring a single path to handle the entire I/O operations. PowerPath can distribute the I/O for all logical devices over all paths shared by those logical devices, so that all paths are equally burdened. PowerPath load-balances I/O on a host-by-host basis, and maintains statistics on all I/O for all paths. For each I/O request, PowerPath intelligently chooses the least-burdened available path, depending on the load-balancing and failover policy in effect. In addition to improving I/O performance, dynamic load balancing reduces management time and downtime, because administrators no longer need to manage paths across logical devices. With PowerPath, configurations of paths and policies for an individual device can be changed dynamically, taking effect immediately, without any disruption to the applications.

PowerPath provides the following features and benefits:

- Multiple paths, for higher availability and performance — PowerPath supports multiple paths between a logical device and a host bus adapter (HBA, a device through which a host can issue I/O requests). Multiple paths enable the host to access a logical device even if a specific path is unavailable. Also, multiple paths can share the I/O workload to a given logical device.
- Dynamic multipath load balancing — through continuous I/O balancing, PowerPath improves a host's ability to manage heavy I/O loads. PowerPath dynamically tunes paths for performance as workloads change, eliminating the need for repeated static reconfigurations.
- PowerPath has an intuitive command line interface (CLI) that provides end-to-end viewing and reporting for the host storage resources, including HBAs all the way to the storage system. PowerPath eliminates the need to manually change the load-balancing policy on a per-device basis.
- Proactive I/O path-testing and automatic path recovery — PowerPath periodically tests failed paths to determine if they are available. A path is restored automatically when available, and PowerPath resumes sending I/O to it. PowerPath also periodically tests available but unused paths to ensure they are operational.
- Automatic path failover — PowerPath automatically redirects data from a failed I/O path to an alternate path. This eliminates application downtime; failovers are transparent and nondisruptive to applications.
- Enhanced high availability cluster support — PowerPath is particularly beneficial in cluster environments, because it can prevent interruptions to operations and costly downtime. PowerPath's path failover capability avoids node failover, maintaining uninterrupted application support on the active node in the event of a path disconnect (as long as another path is available).

PowerPath/VE (Virtual Edition) only works with vSphere (ESX 4 and ESX4i) and is supported with all CLARiiON CX-series arrays configured with failovermode=4 (ALUA mode or Asymmetric Active/Active mode).

It plugs into the vSphere I/O framework to bring advanced multipathing capabilities of PowerPath - dynamic load balancing and automatic failover. PowerPath/VE gets installed using RemoteCLI and uses the command set of the rpowermt utility to monitor, manage and configure PowerPath devices in vSphere.

Figure 49 on page 106 shows CLARiiON LUNs controlled by EMC PowerPath.

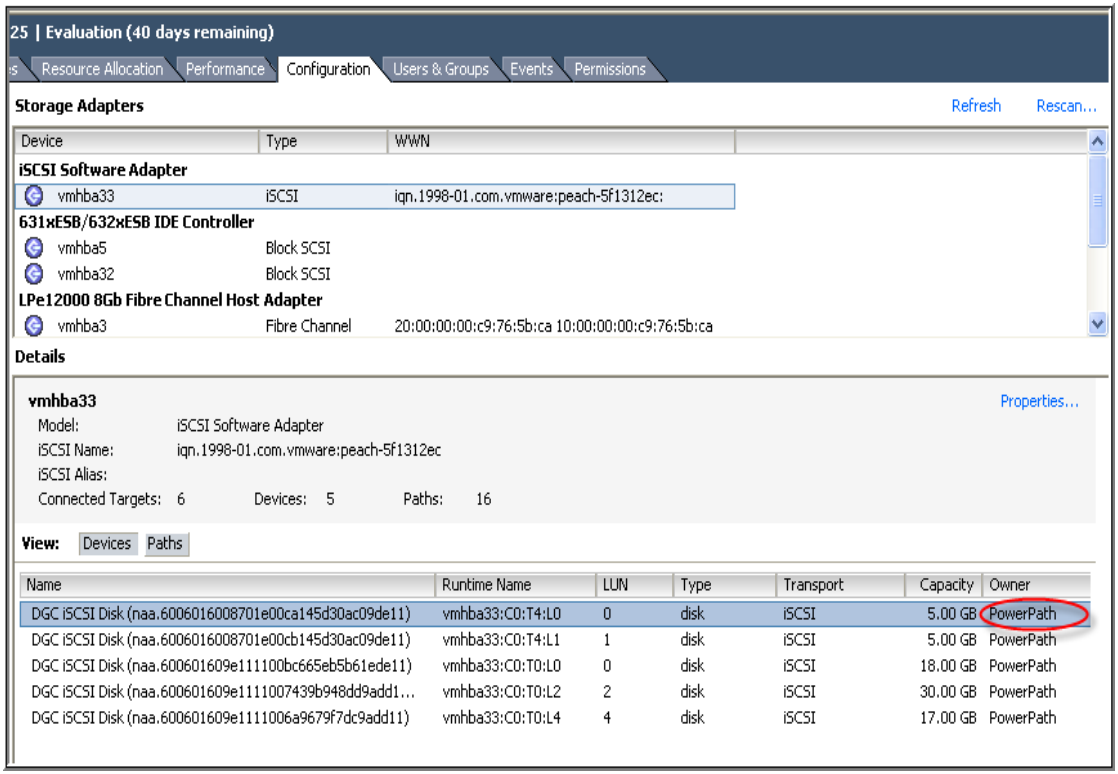


Figure 49 PowerPath/VE configured on vSphere 4.0 connected to CLARiiON

EMC Replication Manager

EMC Replication Manager is an EMC software application that dramatically simplifies the management and use of disk-based replications to improve the availability of user's mission-critical data and rapid recovery of that data in case of corruption.

Replication Manager helps the user to manage replicas as if they were tape cartridges in a tape library unit. Replicas may be scheduled or created on demand, with predefined expiration periods and automatic mounting to alternate hosts for backups or scripted processing. Individual users with different levels of access ensure system and replica integrity. In addition to these features, Replication Manager is fully integrated with many critical applications, such as DB2 LUW, Oracle, and Microsoft Exchange.

Replication Manager makes it easy to create point-in-time, disk-based replicas of applications, file systems, or logical volumes residing on existing storage arrays. It can create replicas of information stored in the following environments:

- Oracle databases
- DB2 LUW databases
- Microsoft SQL Server databases
- Microsoft Exchange databases
- UNIX file systems
- Windows file systems

The software utilizes a Java-based client or server architecture. Replication Manager can:

- Create point-in-time replicas of production data in seconds
- Facilitate quick, frequent, and non-destructive backups from replicas.
- Mount replicas to alternate hosts to facilitate offline processing (for example, decision-support services, integrity checking, and offline reporting)
- Restore deleted or damaged information quickly and easily from a disk replica.
- Set the retention period for replicas so that storage is made available automatically.

Replication Manager has a generic storage technology interface that allows it to connect and invoke replication methodologies available on:

- EMC Symmetrix arrays
- EMC CLARiiON arrays
- EMC Celerra® arrays
- HP StorageWorks arrays

Replication Manager uses SYMAPI Solutions Enabler software and interfaces to the storage array's native software to manipulate the supported disk arrays. Replication Manager automatically controls the complexities associated with creating, mounting, restoring, and expiring replicas of data. Replication Manager performs all of these tasks and offers a logical view of the production data and corresponding replicas. Replicas are managed and controlled with the easy-to-use Replication Manager console.

EMC StorageViewer

The EMC Storage Viewer provides simple, read-only storage mapping functionality for the various storage- related entities that exist within Virtual Infrastructure Client, including datastores, LUNs and SCSI targets. The storage information displayed through the EMC Storage Viewer allows the distinction between the types of storage used, the specific arrays and devices presented, the paths that are used for the storage, and individual characteristics of the existing storage.

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. If running StorageViewer 2.1 or later a license for Solutions Enabler is not needed.
  - 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 as shown in Figure 50 on page 110.

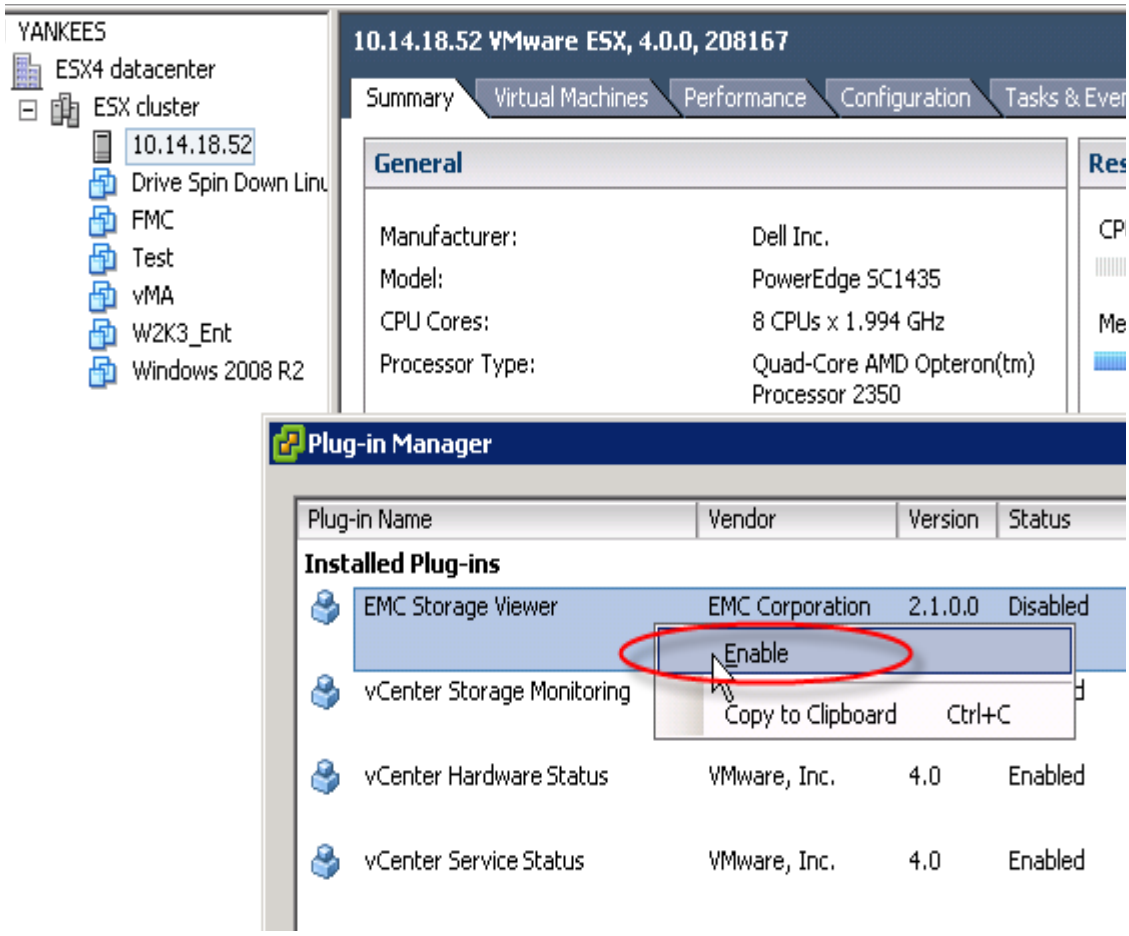


Figure 50 Enabling EMC Storage Viewer in the Virtual Infrastructure Client

The Arrays tab in StorageViewer now allows you to discover, update, and synchronize CLARiiON storage system information as shown in Figure 51.

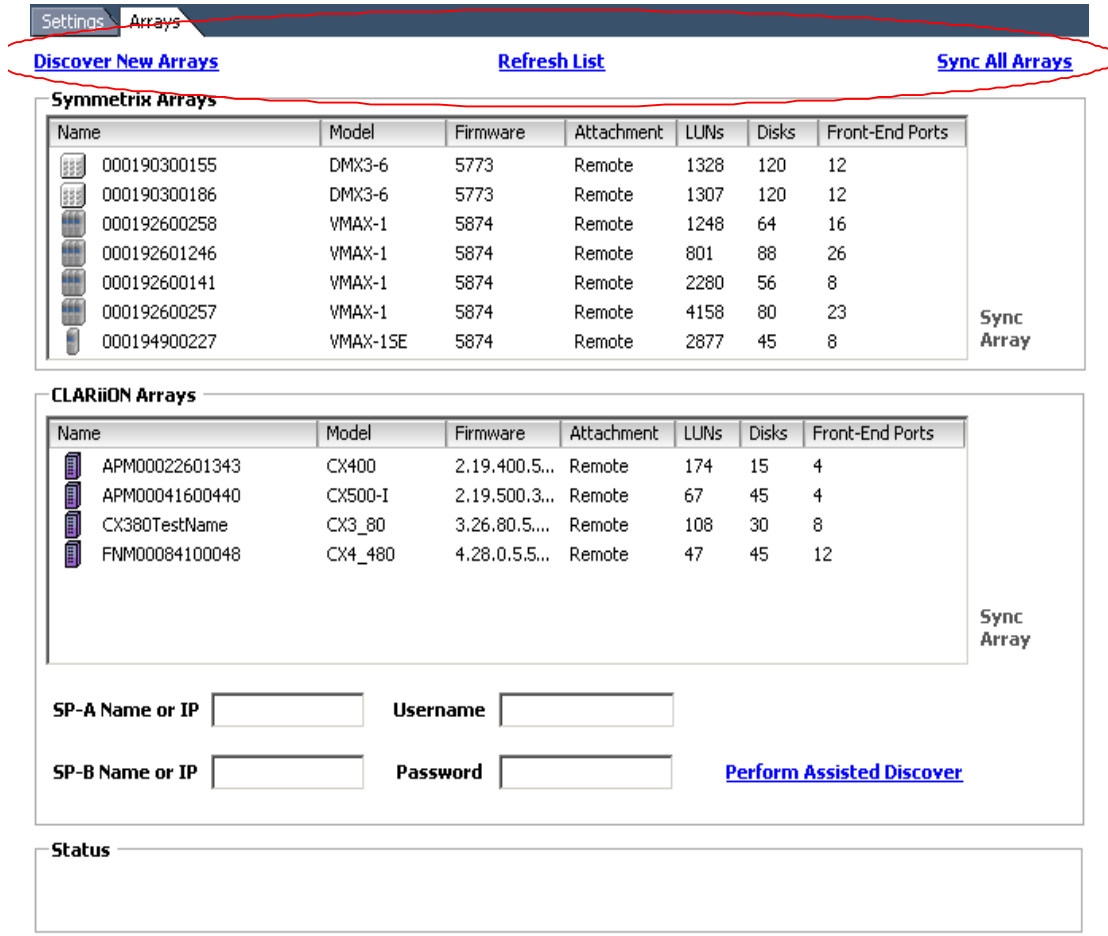


Figure 51 Arrays tab now allows you to discover, refresh, and synchronize arrays

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 three figures.

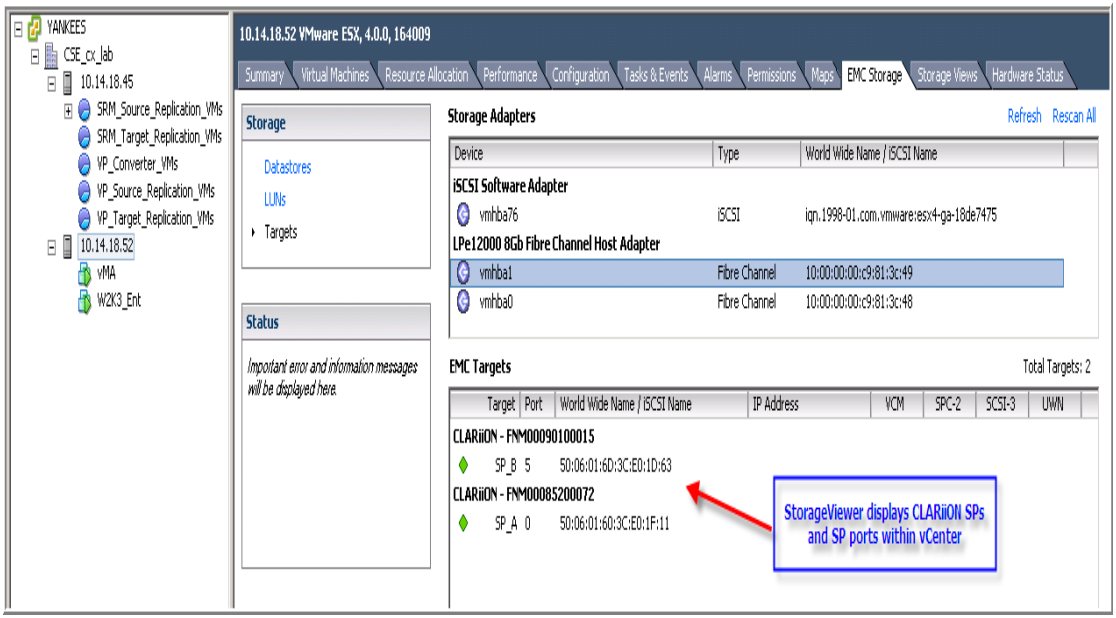


Figure 52 CLARiiON SP and ports information within VMware vCenter

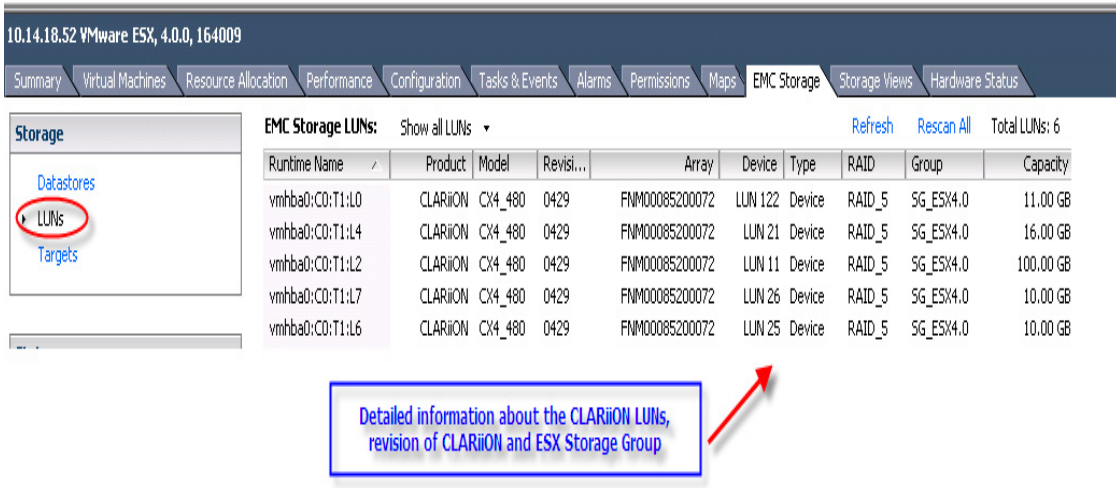


Figure 53 CLARiiON detailed LUN information visible within VMware vCenter

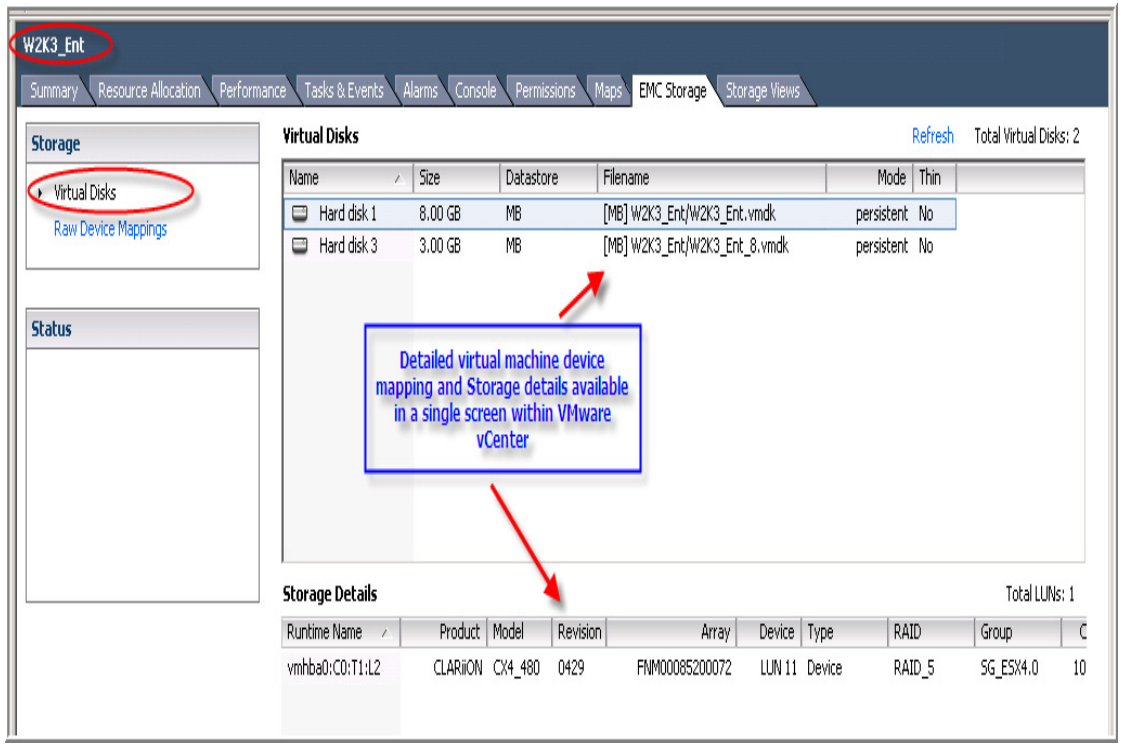


Figure 54 Virtual machine and corresponding CLARiiON detailed LUN information visible within VMware vCenter

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

3
VMware ESX/ESXi and EMC CLARiiON

This chapter presents these topics:

- Configuring VMware ESX version 4, 3.x, and ESXi 116
- Using CLARiiON with ESX Server version 4, 3.x, and ESXi 123
- Using Navisphere in virtualized environments 136
- Mapping a VMware file system to CLARiiON devices..... 145
- Optimizing VI infrastructure and CLARiiON 151

A VMware ESX/ESXi host virtualizes IT assets into a flexible, cost-effective pool of compute, storage, and networking resources. These resources can then be mapped to specific business needs by creating virtual machines. EMC CLARiiON storage arrays are loosely coupled parallel processing machines that handle various workloads from disparate hardware and operating systems simultaneously. When VMware ESX/ESXi host is used with EMC CLARiiON storage arrays, it is critical to ensure proper configuration of both the storage array and the ESX Server to ensure optimal performance and availability.

This chapter addresses the following topics:

- Configuration of EMC CLARiiON arrays when used with VMware ESX/ESXi hosts
- Discovering and using EMC CLARiiON devices in VMware ESX/ESXi hosts
- Optimizing EMC CLARiiON storage array and VMware ESX/ESXi host for interoperability

Detailed information on configuring and using VMware ESX/ESXi hosts in an EMC FC, NAS and iSCSI environment can also be found in the ***EMC Host Connectivity Guide for VMware ESX.*** This is the authoritative guide for connecting VMware ESX/ESXi hosts to EMC CLARiiON storage arrays and should be consulted for the most current information.

Configuring VMware ESX version 4, 3.x, and ESXi

VMware ESX/ESXi hosts can be booted off an EMC CLARiiON storage array or internal disks. VMware and EMC fully support booting the VMware ESX/ESXi 4.x and 3.x hosts from EMC CLARiiON storage arrays when using either QLogic or Emulex HBAs.

Booting the VMware ESX/ESXi hosts from the SAN enables the physical servers to be treated as an appliance allowing for easier upgrades and maintenance. Furthermore, booting VMware ESX/ESXi hosts from the SAN can simplify the processes for providing disaster restart protection of the virtual infrastructure. Specific considerations when booting VMware ESX/ESXi hosts are beyond the scope of this document. The ***EMC Support Matrix***, available at emc.com, and appropriate VMware documentation should be consulted for further details.

When performing the installation of ESX server, neither VMware and EMC do not recommends creating VMware file system partitions using the installer. The VMware ESX/ESXi hosts installer does not create aligned partitions. Alignment of partitions and VMware file systems on EMC CLARiiON storage array track is discussed in "Single vSwitch iSCSI configuration," on page 170.

Detailed installation procedures can be found in the product documentation available on the VMware website.

Configuring swap space

The service console operating system for ESX 4.x and 3.x are provided with a fixed amount of memory. This is independent of the maximum number of virtual machines that is anticipated to be simultaneously executing on the physical server. The VMware ESX/ESXi hosts installer automatically provisions the correct service console swap space if the boot disk can accommodate it.

In VMware ESX/ESXi version 4.x and 3.x and ESXi versions, a swap file is automatically configured and maintained with each virtual machine on a VMware file system. This change in the architecture has two different repercussions:

1. The swap file resides on the SAN-attached storage array. This could have potential performance implications on a heavily loaded VMware ESX/ESXi hosts.
2. The available storage on a VMware file system to store virtual disks can be severely impacted by the presence of the swap file. Proper planning and design of the storage array, VMware file systems and virtual machine memory is critical to ensure optimal utilization of the resources.

Note: The size of the virtual machine swap file is equivalent to the maximum amount of memory configured for that virtual machine minus the reservation.

Configuring the ESXkernel

With VMware ESX/ESXi version 4 and 3, the Virtual Infrastructure Client is used to manage a VMware ESX/ESXi host directly or through a vCenter Infrastructure server. An example of the splash screen that is provided when Virtual Infrastructure client is started is shown in Figure 55 on page 118.

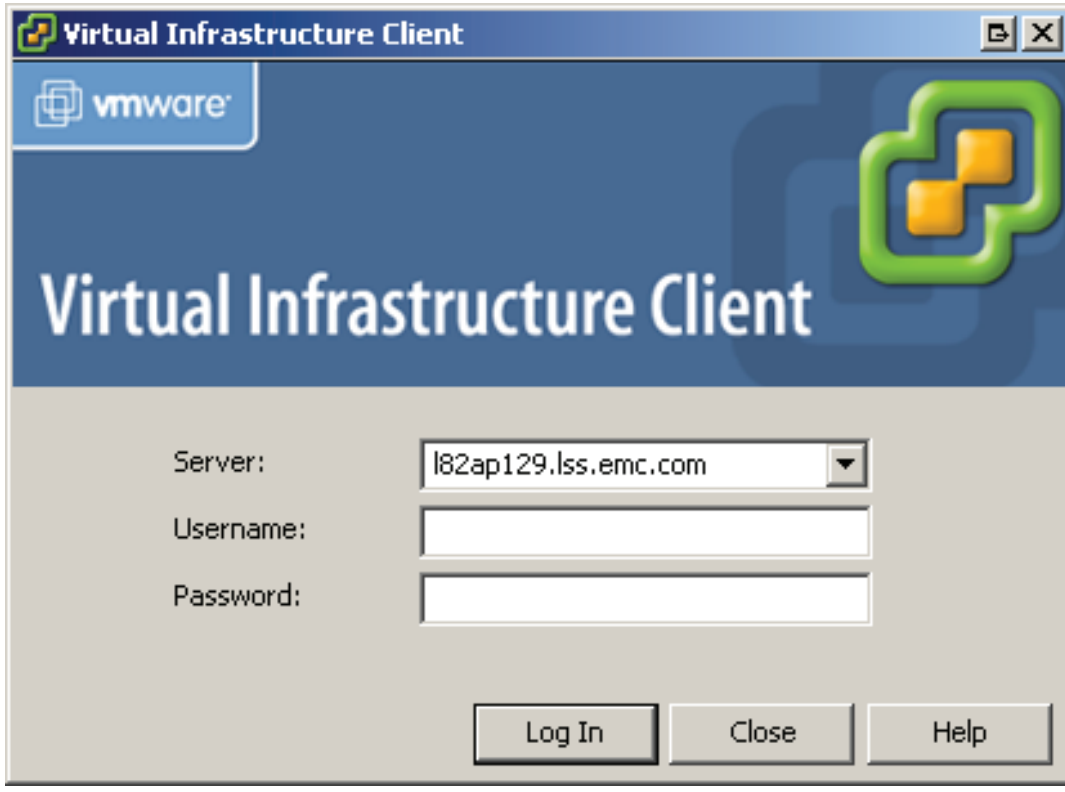


Figure 55 VMware vClient login screen

The hostname of the VMware ESX/ESXi host or the vCenter management server should be entered in the server row shown in Figure 55 on page 118. Providing proper credentials to the VMware Virtual Infrastructure client enables the client to authenticate against the server listed. A window as shown in Figure 56 on page 119 is displayed on successful authentication. The customization of the VMware ESX/ESXi version 3 kernel for use with an EMC CLARiiON storage array needs to be performed utilizing this interface.

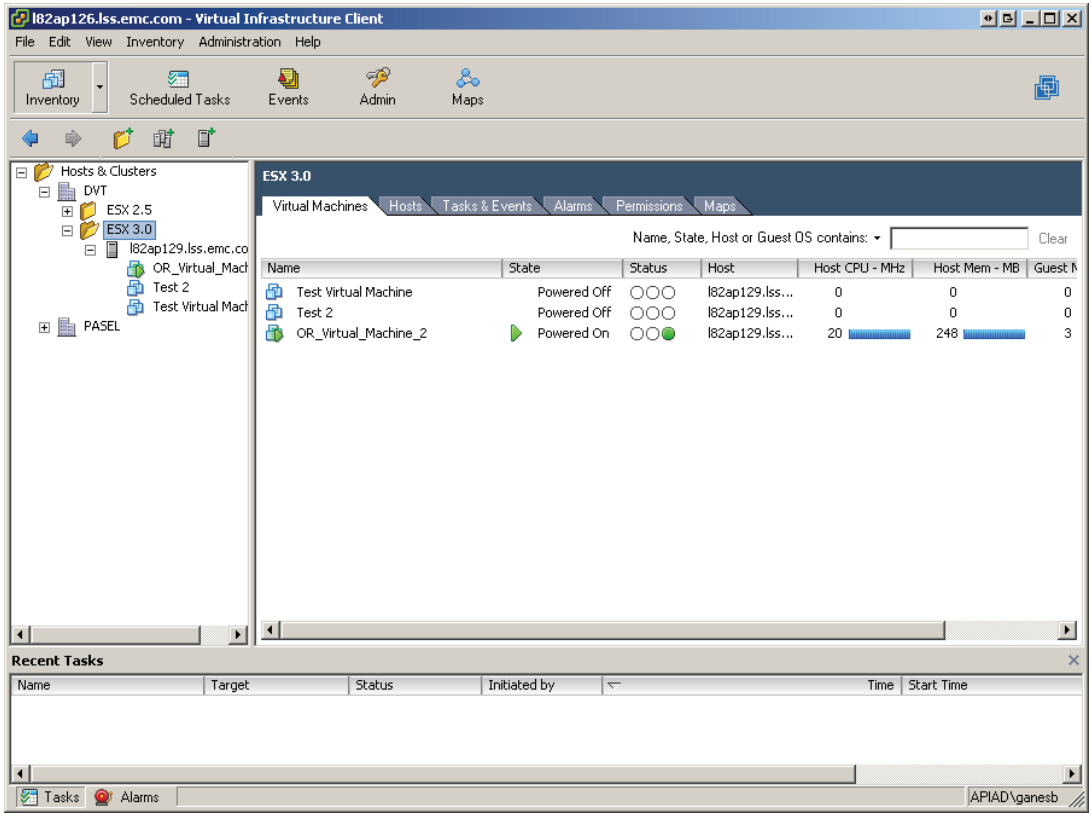


Figure 56 Virtual Infrastructure Status Monitor screen

VMware ESX/ESXi version 4.x, 3.x and ESXi support a maximum of 256 SCSI devices. These devices can be a combination of internal and SAN-attached disks and tape drives.

The parameters, `Disk.MaxLUN and Disk.SupportSparseLUN`, should be set to 256 and 1, respectively, for VMware ESX/ESXi versions 3 and 4, and ESXi. These are usually the default values for these parameters. You can modify the **Disk.MaxLUN** parameter to improve LUN discovery speed; you can also modify the **Disk.SupportSparseLUN** parameter to decrease the time VMware ESX/ESXi takes to scan LUNs. To change the parameters cited above, the following steps should be performed:

1. In the VMware Virtual Infrastructure Client screen, select the VMware ESX/ESXi host that you need to configure.

2. Select the Configuration tab on the right hand pane.
3. Select the Advanced Setting link on the left hand bottom corner of the pane.
4. This opens a new window. In the new window, select the Disk option.
5. Scroll through the parameters and modify the parameters `Disk.MaxLUN` and `Disk.SupportSparseLUN` as recommended above.

Steps 1–3 of the procedure listed above are shown in Figure 57 on page 120. Figure 58 on page 121 pictorially depicts the steps 4 and 5 documented in the procedure above.

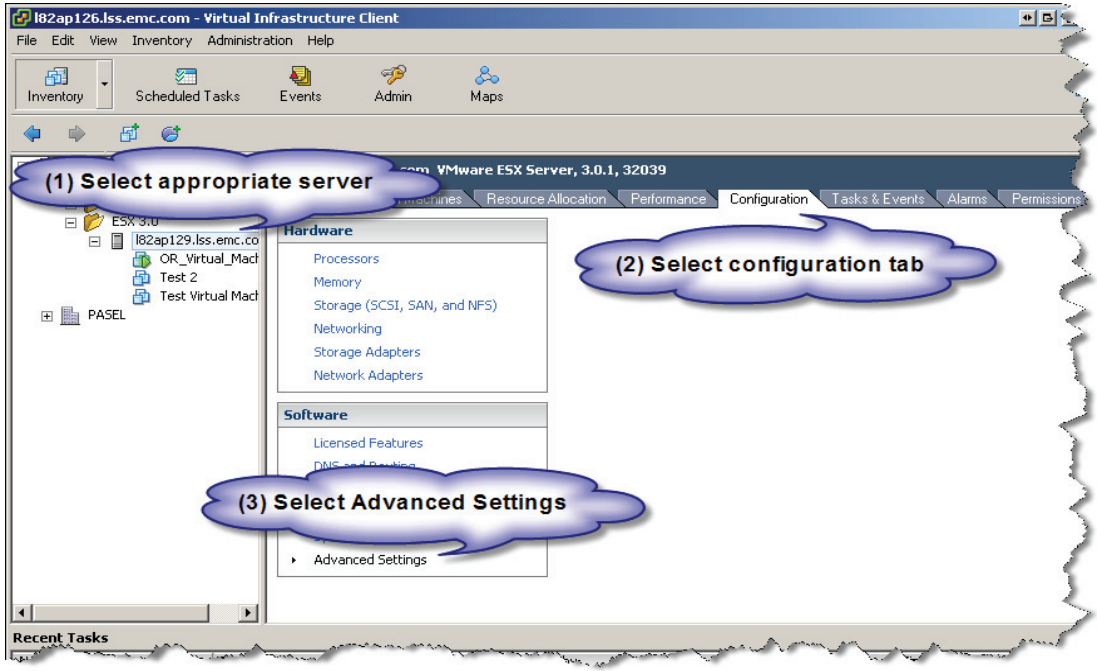


Figure 57 Configuring the VMware ESX and ESXi kernel for EMC CLARiiON arrays

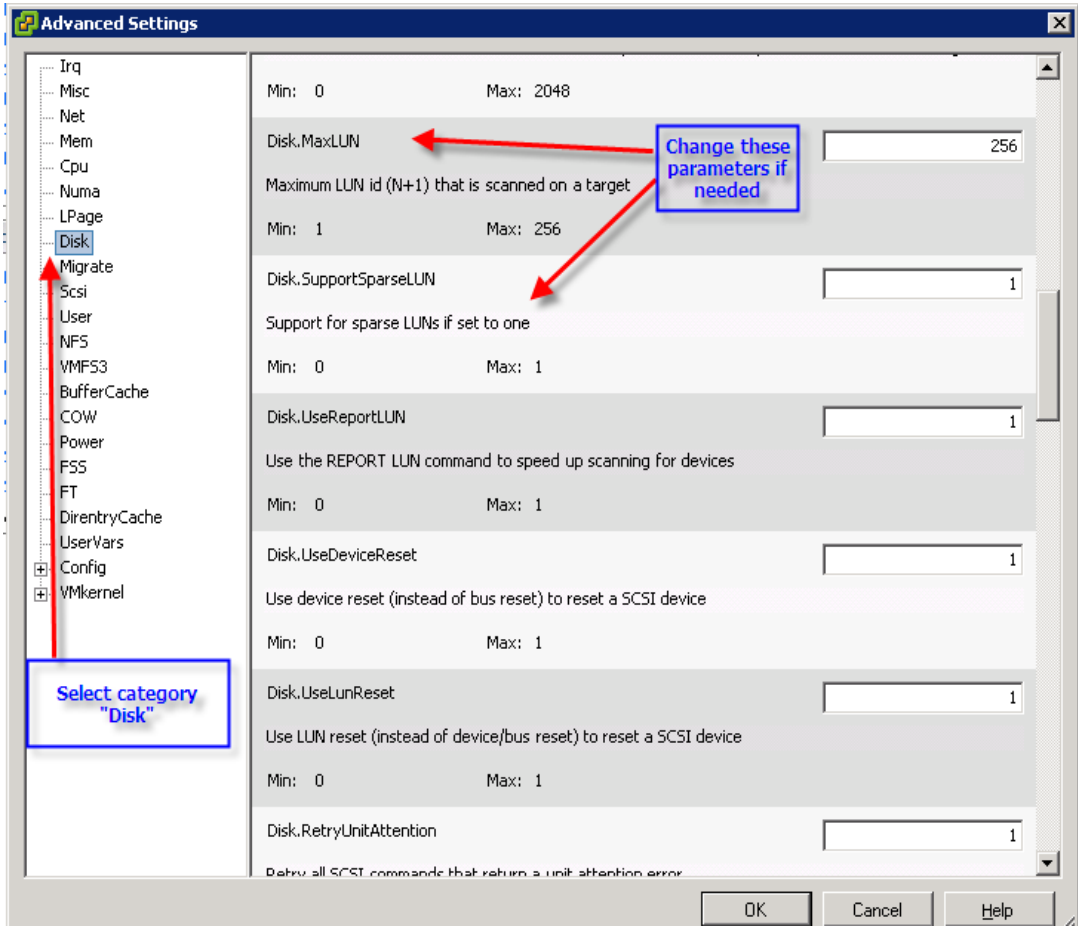


Figure 58 Changing Advanced settings for disks in VMware ESX/ESXi 4 host

Persistent binding

VMware ESX/ESXi version 4 and 3 do not need persistent binding. The virtualization architecture accommodates dynamic changes in the storage area networks transparently to the end user. This is achieved by using two pieces of information that are unique to the storage device—one is provided by the SCSI standard specification (unique

identifier); the other is generated by the VMware kernel when the device is initialized for use with VMware logical volume manager (LVM).

Multipathing and failover in ESX version 3 and ESX3i

The multipathing and failover functionality has not changed in VMware ESX/ESXi version 3. Similar to VMware ESX hosts version 2.x, the multipathing and failover software does not provide dynamic load balancing. On active-passive storage arrays, such as EMC CLARiiON, the MRU (Most Recently Used) policy must be configured. If the policy is set to MRU, the **preferred** mode, although displayed, is not used. Static load balancing can be achieved by balancing CLARiiON LUNs across the two SPs and HBAs. The ***Fibre Channel and iSCSI SAN Configuration Guide*** for ESX Server discusses the process using the vClient.

Note: Dynamic load balancing software, such as EMC PowerPath, is not supported on the VMware ESX/ESXi host service console or in the virtual machines. VMware ESX/ESXi host multipathing must be used to provide access to multiple paths to a storage device.

Multipathing and failover in ESX version 4 and ESX4i

VMware ESX 4 contains its own native multipathing software that is built into its kernel. This failover software, called Native Multipathing Plugin (NMP), contains three policies:

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

In VMware ESX 4, the NMP module supports basic load balancing when the Round Robin policy is used with the CLARiiON arrays. FIXED and MRU policies do not support load balancing. However, static load balancing can be achieved by balancing CLARiiON LUNs across the two SPs and HBAs. The ***Fibre Channel and iSCSI SAN Configuration Guide*** for ESX Server discusses the process using the vClient.

PowerPath/VE is also supported with ESX version 4 and 4i and is installed using Remote CLI. PowerPath is supported in FC and iSCSI (software and hardware initiators) configurations. It supports dynamic load balancing and failback options.

Using CLARiiON with ESX Server version 4, 3.x, and ESXi

Configuring VMware ESX/ESXi hosts version 3 enables the VMware ESX/ESXi host to discover and use EMC CLARiiON fibre and iSCSI devices. The CLARiiON storage array also needs to be configured for proper communications between the VMware ESX/ESXi hosts and the SAN. The following settings on the CLARiiON initiators records are needed for both Fibre Channel and iSCSI connections:

- arraycompath = enabled
- failovermode = 1
- Access Logix enabled

In order for the VMware ESX/ESXi 4 host to discover and use EMC CLARiiON fibre and iSCSI devices, the following settings on the CLARiiON initiators records are recommended when using the FIXED or Round Robin policy with NMP or PowerPath/VE for both Fibre Channel and iSCSI connections with CX4 storage systems.

- arraycompath = enabled
- failovermode = 4
- Access Logix enabled

If you use NMP's MRU policy or the Round Robin policy on CX3 or CX-series [CX700, CX500, CX300] arrays connected to VMware ESX/ESXi 4 hosts, the failover mode must set to **failover=1.** If you use PowerPath/VE with CX3- or CX-series arrays, you can set the failover mode to **failovermode=4** or **failovermode=1** for the host iniatiator records. See the Section , "Path management," on page 159 for more details.

Note: The ***EMC Host Connectivity Guide for VMware ESX Server*** provides up-to-date listings of the initiator settings.

Fibre HBA driver configuration

The drivers provided by VMware as part of the VMware ESX/ESXi distribution should be utilized when connecting VMware ESX/ESXi hosts to EMC CLARiiON storage using Fibre Channel. However, EMC

E-Lab does perform extensive testing to ensure the BIOS, BootBIOS and the VMware supplied drivers work together properly with EMC storage arrays. The results from the qualification tests are reported in the ***EMC Support Matrix*** available on the EMC website.

ESX iSCSI HBA and NIC drivers

The iSCSI HBA and NIC drivers provided by VMware as part of the VMware ESX/ESXi 3 and 4 distribution should be utilized when connecting VMware ESX/ESXi hosts to EMC CLARiiON storage using iSCSI protocol. The ***Fibre Channel and iSCSI SAN Configuration Guide*** available on the VMware website provides information on configuring the HBA and NIC drivers for connectivity to the CLARiiON storage system.

Adding and removing CLARiiON devices

The addition or removal of EMC CLARiiON devices to and from VMware ESX/ESXi version 3, 4, and ESXi is a two-step process:

1. In the first step, appropriate changes need to be made to the EMC CLARiiON storage array configuration. This may include, in addition to LUN masking, creation and assignment of EMC CLARiiON LUNs and metaLUNs to the Fibre Channel ports utilized by the VMware ESX/ESXi hosts. The configuration changes can be performed using Navisphere Manager or CLI from an independent storage management host.
2. The second step of the process forces the VMware kernel to rescan the Fibre Channel bus to detect changes in the environment. This can be achieved by the same three steps discussed in "Adding and removing CLARiiON devices," on page 124—restarting the VMware ESX/ESXi hosts, using the graphical user interface, or using the command line utilities. The process to discover changes to the storage environment using these tools are discussed in the next two subsections.

Using the vCenter client

Changes to the storage environment can be detected using the vCenter client by using the following process:

1. Select the VMware ESX/ESXi host on which you need to detect the changes.
2. Click on the Configuration tab to highlight it.

3. The current storage environment is displayed on the right hand pane by selecting Storage Adapters.
4. Highlight any storage adapter in the pane.
5. Click on Rescan to initiate the rescan process on the VMware ESX/ESXi hosts.

The steps listed above are captured and displayed in Figure 59 on page 125.

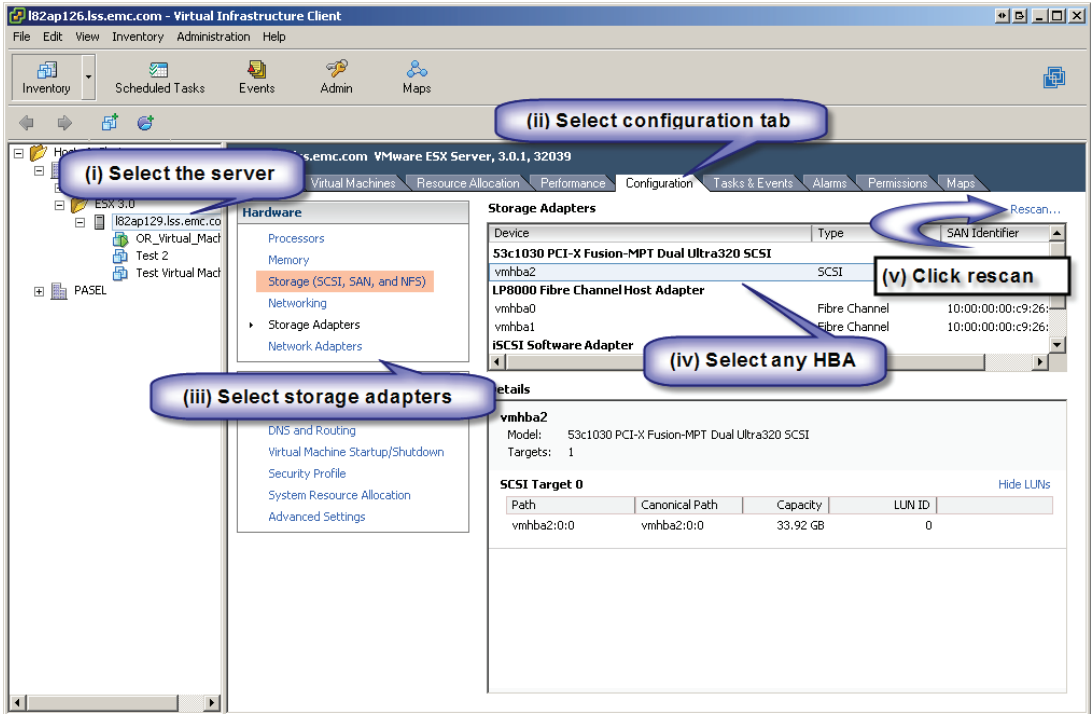


Figure 59 Using Virtual Infrastructure Client to detect changes to storage environment

With VMware ESX version 4,3 and VMware ESXi initiating the rescan process using Virtual Infrastructure client results in a new window providing users with two options. An example of the pop up window is shown in Figure 60 on page 126. The two options allow users to customize the rescan to either detection of changes to the storage area network, or to the changes in the VMFS volumes. The process to scan the storage area network is much slower than the process to scan for changes to VMFS volumes. The storage area network should be

scanned only if there are known changes to the environment. Similarly, the checkbox to scan for VMFS volumes should not be selected if uninitializeddevices (with no filesystem) are being added to the VMware environment. Also note that in vSphere 4.x, the rescan can also be performed at the cluster and data center level.

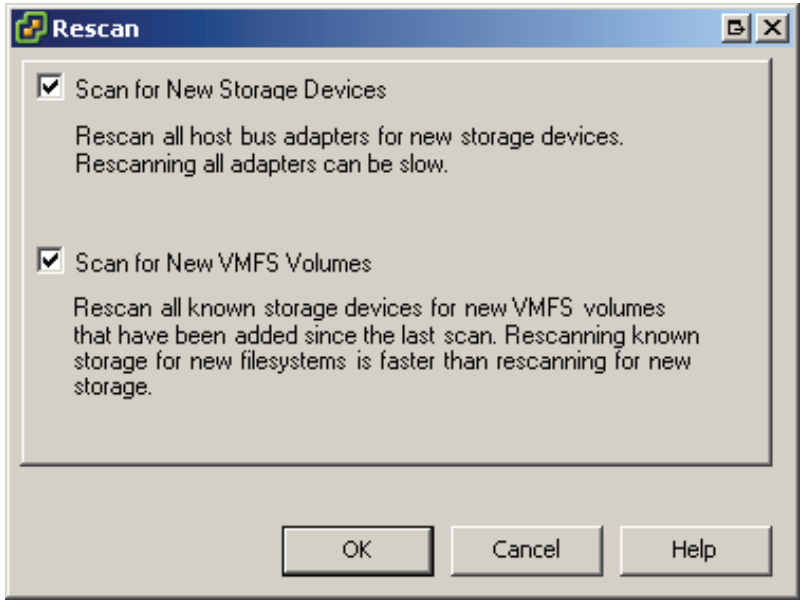


Figure 60 Rescanning options in a Virtual Infrastructure 3 environment

Using VMware ESX version 4 or 3.x command line utilities

In both, VMware ESXversion 4 and 3.x the service console utility, `vicfg-rescan or esxcfg-rescan`, detects changes to the storage environment. The utility, `esxcfg-rescan`, takes the VMkernel SCSI adapter name (vmhbax) as an argument. This utility should be executed on all relevant VMkernel SCSI adapters if EMC CLARiiON devices are presented to the VMware ESX on multiple paths. Figure 61 on page 127 displays an example using `esxcfg-rescan`.

The management of the VMware ESXi version (whose operating system is embedded in the hardware or can be installed on a hard disk) is accomplished using a vCenter server or a client. On Windows and Linux platforms, you can also use Remote CLI to issue commands directly to the VMware ESXi server.

Creating VMFS volumes

Both VMware vSphere 4 and Virtual Infrastructure 3 provide a beneficial change that reduces the complexity of managing the storage environment while offering potential performance and scalability benefits. VMware file system volumes created utilizing Virtual Infrastructure client are automatically aligned on 64 KB boundaries. A manual process is unnecessary on VMware ESX version 4, 3.x or VMware ESXi as long as the volume is created utilizing Virtual Infrastructure client. Therefore, EMC strongly recommends utilizing the Virtual Infrastructure client to create and format VMware file system volumes.

Note: A detailed description of track and sector alignment in x86 environments is presented in "Single vSwitch iSCSI configuration," on page 170.

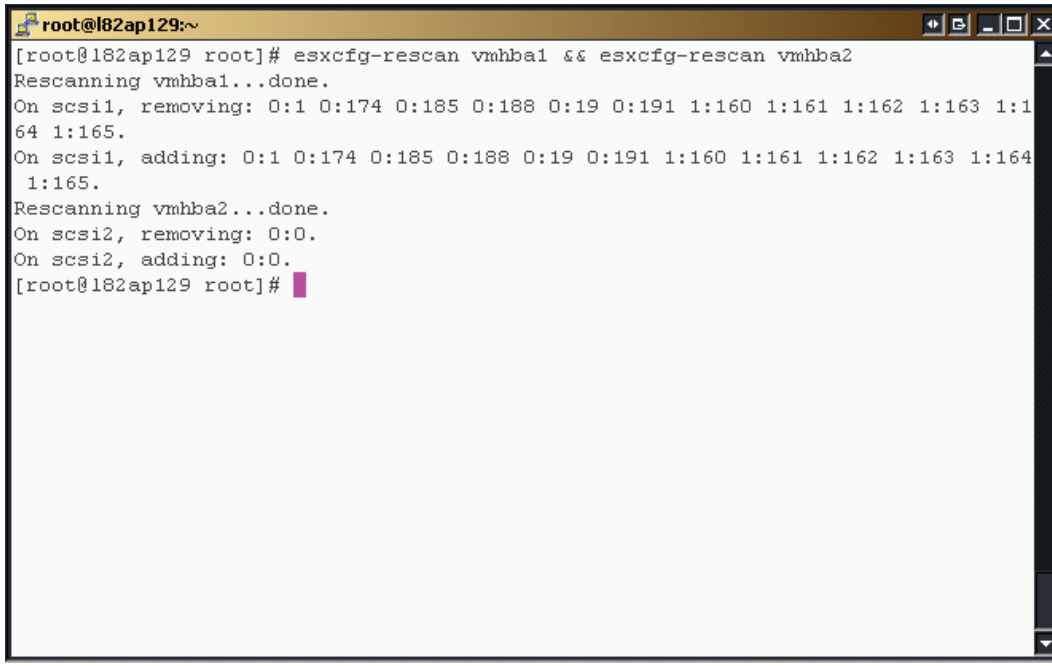


Figure 61 Using VMware ESX version 3.x service console utilities to rescan SAN

Creating a datastore using the vCenter client

The Virtual Infrastructure client does not distinguish between creation of VMware file system volume and a VMware file system. Therefore, the vCenter client offers a single process to create an aligned VMware file system.

Note: A datastore in VMware vSphere or Virtual Infrastructure 3 environment can be either a NFS file system or a VMware file system. Therefore the term, datastore, is utilized in the rest of the document. Furthermore, a group of VMware ESX/ESXi hosts sharing a set of datastores is referred to as a cluster or a cluster group.

The storage (SCSI, SAN, and NAS) object on the right hand pane (see the area highlighted in orange in Figure 59 on page 125) provides the path to create a new datastore. As seen in Figure 62 on page 128, selecting this object displays all available datastores on the VMware ESX/ESXi hosts. In addition to the current state information, the pane also provides the options to manage the datastore information and create new datastore. The wizard to create a new datastore can be launched by clicking on Add storage on the top right-hand corner of the storage pane (see Figure 62 on page 128).

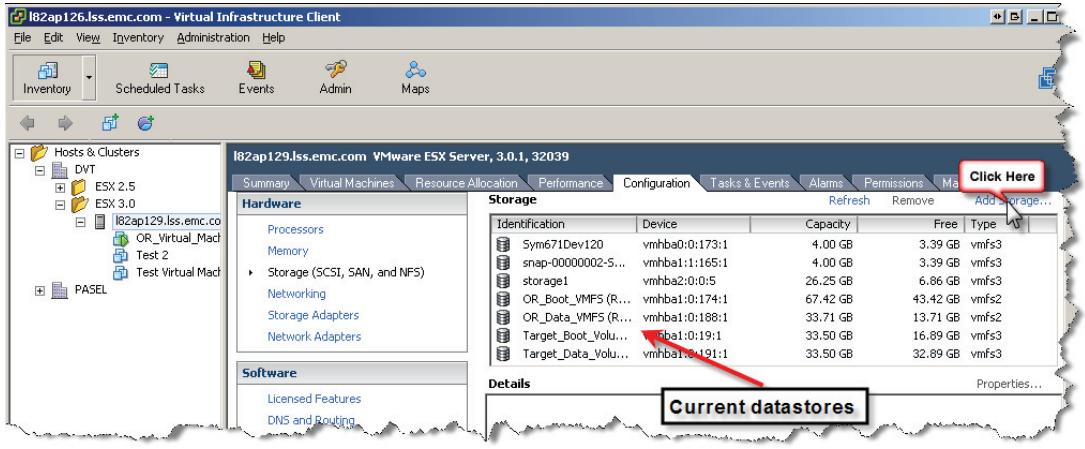


Figure 62 Displaying and managing datastores in a Virtual Infrastructure 3 environment

The Add Storage wizard on start up, as seen in Figure 63 on page 130, presents a summary of the required steps to provision a new datastore. The Disk/LUN option should be selected to provision a datastore on Fibre Channel or iSCSI-attached EMC CLARiiON storage array.

Clicking Next in the wizard presents all viable FC, iSCSI, or SCSI attached devices. Devices that have existing VMware file systems (VMFSor NFS) are not presented on this screen. This is independent of whether or not that device contains free space. However, devices with existing non-VMFS formatted partitions but with free space are visible in the wizard. An example of this is exhibited in Figure 63 on page 130.

Virtual Infrastructure client allows only one datastore on a device. EMC CLARiiON storage arrays support nondisruptive expansion of storage LUNs. The excess capacity available after expansion can be utilized to expand the existing datastore on the LUN. Appendix A, "Nondisruptive Expansion of a MetaLUN" focuses on this feature of EMC CLARiiON storage arrays.

The next step in the process involves selecting the appropriate device in the list provided by the wizard and clicking Next. The user is then presented with either a summary screen or with a screen with two options depending on the configuration of the selected device. If the selected device has no existing partition, the wizard presents a summary screen detailing the proposed layout on the selected device. Devices with existing partitions (as is the case in the example detailed in Figure 63 on page 130) are prompted with the option of either deleting the existing partition or creating a VMFS volume on the free space available on the device. After selecting the appropriate option (if applicable), clicking Next on the wizard enables the user to provide a name for the datastore (see Figure 63 on page 130).

The final step in the wizard is the selection of options for formatting the device with the VMware file system version 3. As seen in Figure 64 on page 131, the wizard automatically selects the appropriate formatting option. The block size of the VMware file system influences the maximum size of a single file on the file system. The default block size (1 MB) should not be changed unless a virtual disk larger than 256 GB has to be created on that file system. However, unlike other file systems, VMFS-3 is a self-tuning file system that changes the allocation unit depending on the size of the file that is being created. This approach reduces wasted space commonly found in file systems with average file size smaller than the block size.

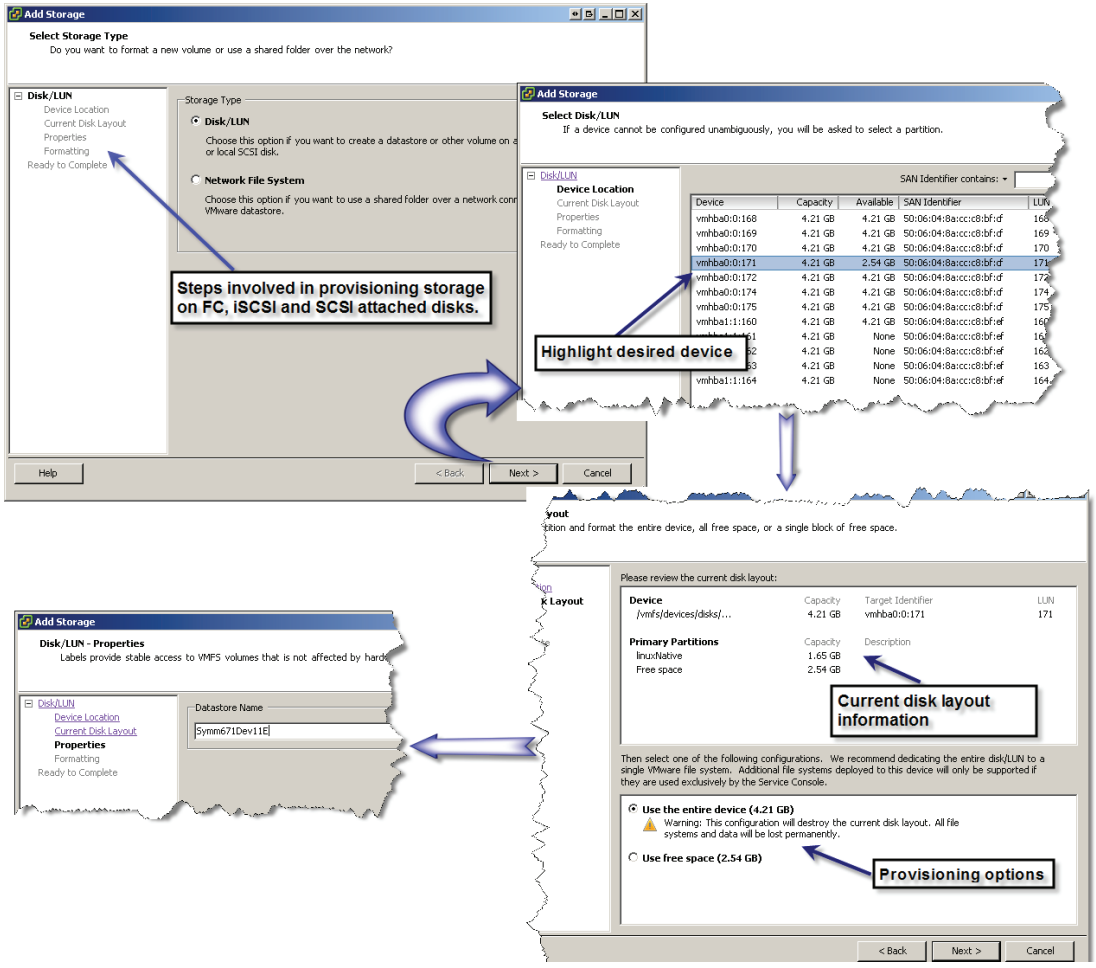


Figure 63 Provisioning a new datastore in a Virtual Infrastructure 3 environment

The wizard, as seen in Figure 63 on page 130, also offers the opportunity to allocate a part of the selected SCSI device for the datastore. This option should not be used unless a second datastore needs to be created utilizing command line utilities. However, configuring multiple datastores on a single device is not recommended by VMware or EMC.

Clicking Next and Finish results in the creation of a datastore on the selected device.

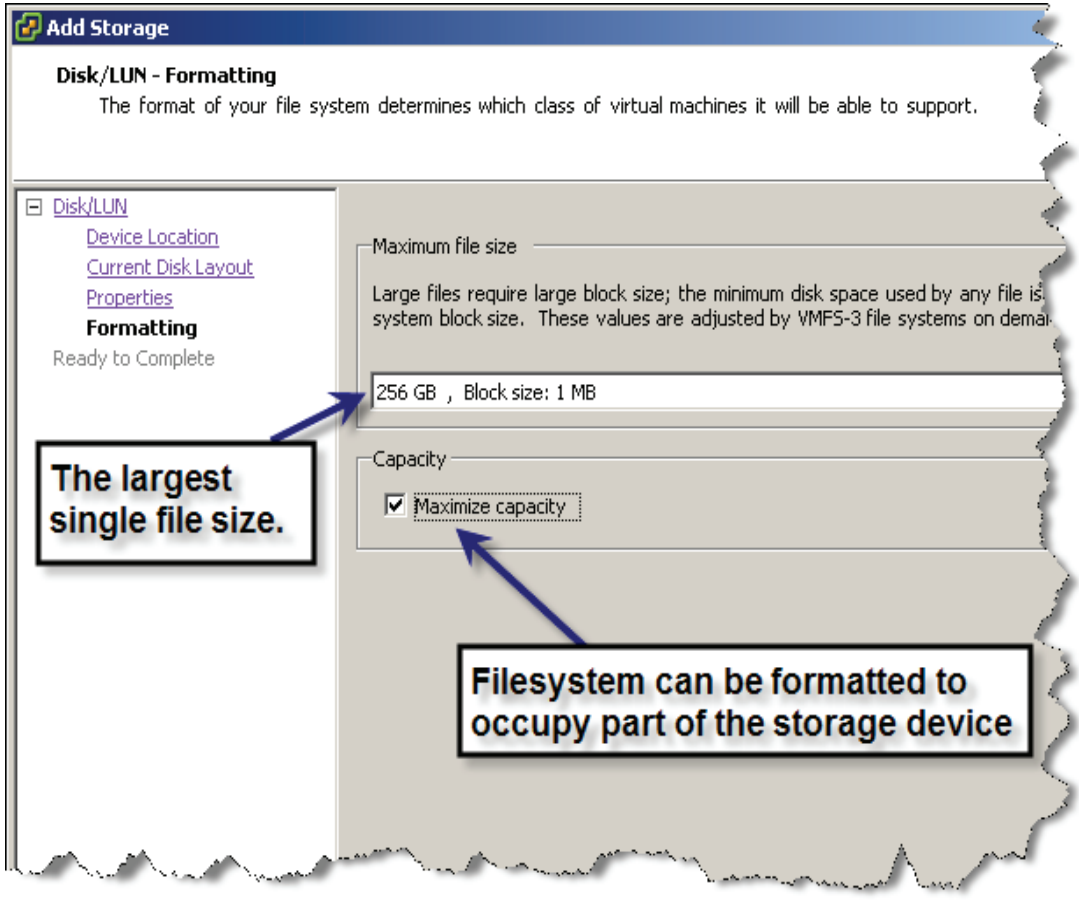


Figure 64 Options for formatting VMFS volumes with VMFS version 3

Creating a datastore using command line utilities

Both VMware ESX 4 andVMware ESX 3 provide a command line utility, `vmkfstools`, to create VMware file system on VMFS volumes. The VMFS volume can be created on either FC or iSCSI attached to EMC CLARiiON storage devices by utilizing fdisk. Due to the complexity involved in utilizing command line utilities, VMware and EMC recommends use of the vCenter client to create a datastore on EMC CLARiiON devices.

Creating RDM volumes on VMware ESX version 4, 3 or VMware ESXi

RDM volumes have an 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.

The creation of RDM volumes in VMware ESX 4, 3 and ESXi is accomplished by presenting CLARiiON LUNs to the ESX server and then adding the raw LUN through the virtual machine Edit settings interface. Using the Add button and selecting the Add Hard Disk wizard allows users to add Raw Device Mapping to a virtual machine shown in Figure 65 on page 133. Note that these are raw devices– no VMware file system exists on these LUNs.

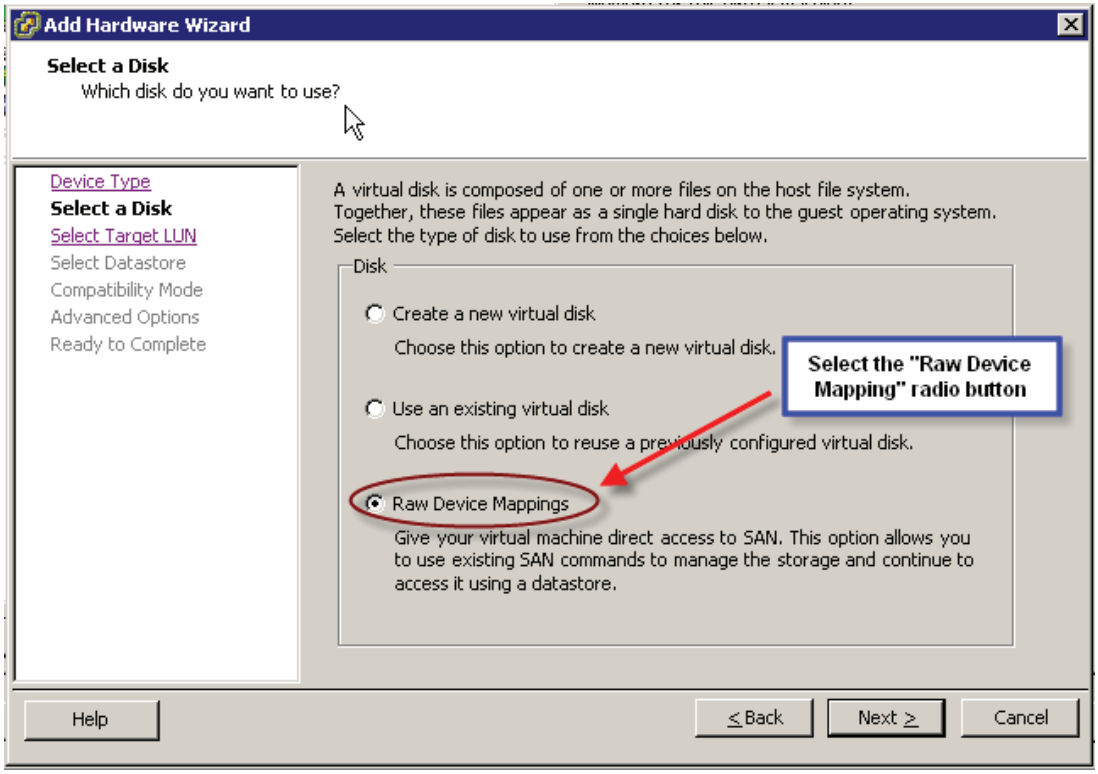


Figure 65 Selecting a Raw Device Mapping volume

The VMware file system hosting the mapping file for the RDM volume is selected as part of the Add Hardware Wizard process. This is shown in Figure 66 on page 134.

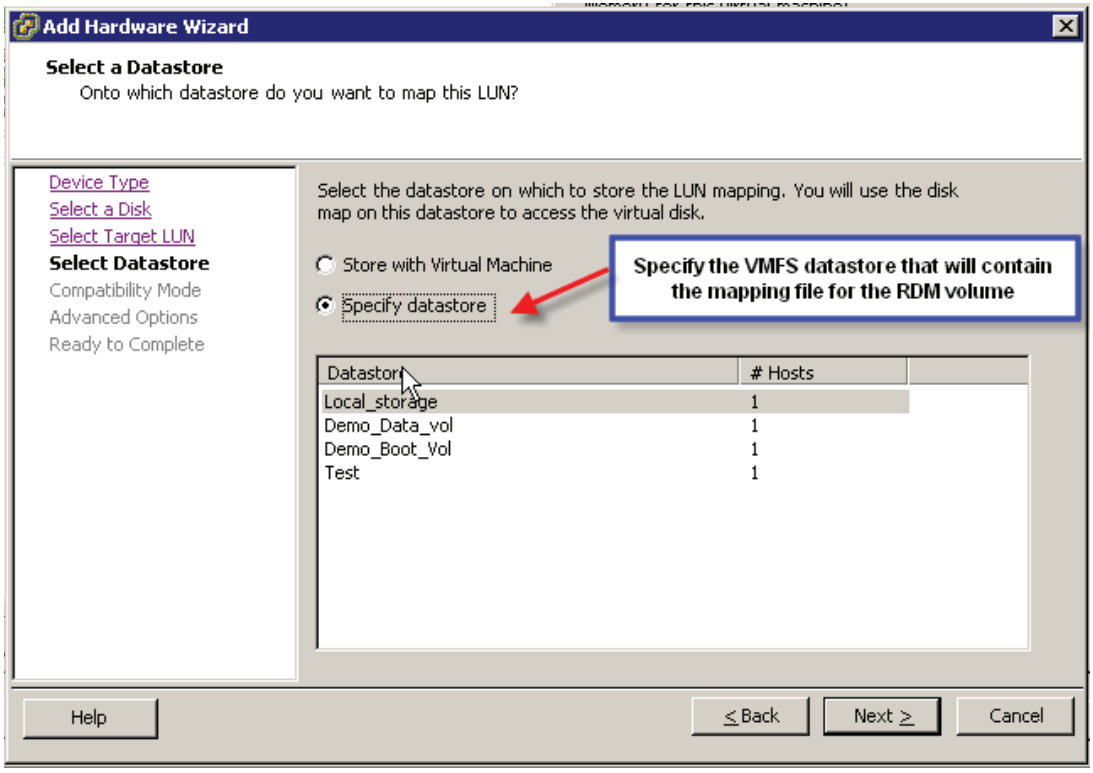


Figure 66 Specifying a VMFS volume that would contain the mapping file for a RDM volume

Fibre HBA performance and tuning on VMware ESX/ESXi hosts

The VMkernel for both VMware ESX 4 and 3 uses a default queue depth of 16 and 30 for the QLogic and Emulex Fibre Channel adapters, respectively. Furthermore, to prevent a single virtual machine from monopolizing a target, the kernel controls the maximum number of outstanding SCSI commands from a virtual machine to a LUN by using the contents of the parameter `Disk.SchedNumReqOutstanding`. The default value for this parameter is 16. Simple tests performed with a single and multiple virtual machine connected to EMC CLARiiON storage devices have shown that changing the HBA queue depth can provide significant improvement in throughput performance under a heavy random read workload.

CAUTION

Changing queue depths can significantly impact response time. Furthermore, increasing the queue depth from the default value will impact the heap usage in VMkernel. Since the performance impact of changing VMware ESX/ESXi hosts parameters depends on the workload, EMC does not recommend changing the default values of queue depth and Disk.SchedNumReqOutstanding. The procedures described in this section are provided as guidance for those customers able to tune the performance of the VMware ESX/ESXi hosts for their workload.

The throughput and performance characteristics of the EMC CLARiiON storage arrays can be influenced significantly by changing the default queue depth. The ***Fibre Channel and iSCSI SAN Configuration guide*** provides a step-by-step process to change the queue depth. However, as stated previously, changes to the queue depth can negatively impact average response time. The queue depth should not be changed unless the storage devices exhibit unsatisfactory performance.

Tuning ESX iSCSI HBA and NIC

The VMware iSCSI HBA driver for hardware iSCSI initiator supports jumbo frames that reduce CPU overhead for processing the iSCSI packets. As a result, the ESX iSCSI HBA performs better than the VMware ESX/ESXi iSCSI software initiator. If needed, the queue depth for the iSCSI HBA can be changed. The procedure for changing queue depth for iSCSI hardware initiator is available from VMware.

In addition to the configuration listed above, EMC also supports the ability where an ESX software initiator driver interacts with the ESX hosts to connect to the iSCSI target through an existing Ethernet adapter.

Guest OS iSCSI initiators are thirty-party software iSCSI initiators available for download, and based on their minimum requirements, can be successfully installed to a supported guest operating system running in a virtual machine.

Tests conducted by VMware have shown that the performance of the Microsoft software initiator running inside a VM is almost equal to running the software initiator within a physical server. Configuring the

virtual machine to use Microsoft iSCSI initiator enables the virtual machines to access CLARiiON iSCSI LUNs directly. This simplifies replication of virtual machine data using software, such as Replication Manager. The white paper ***EMC Replication Manager with CLARiiON and VMware ESX - Best Practices Planning*** details some of these solutions.

Using Navisphere in virtualized environments

Navisphere Agent, Server Utility and CLI

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 mapping 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. For Navisphere Agent/CLI to work with a VMware ESX 3.x server, when connected to a CLARiiON storage system, the ports for agent and 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
```

Alternatively, the ESX_install.sh script can be used to install Navisphere Agent/CLI packages. The ESX_install.sh automatically opens the ports needed for Navisphere Agent/CLI. The *CLARiiON Server Support* ***Products for Linux and VMware ESX Server Installation Guide,*** available on EMC Powerlink, provides detailed information on which ports to open.

If Navisphere Agent is installed, after a rescan of the VMware ESX host, restart the agent so that it communicates with the storage system and sends updated information.

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 Server Utility software can also 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 67 on page 138 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.

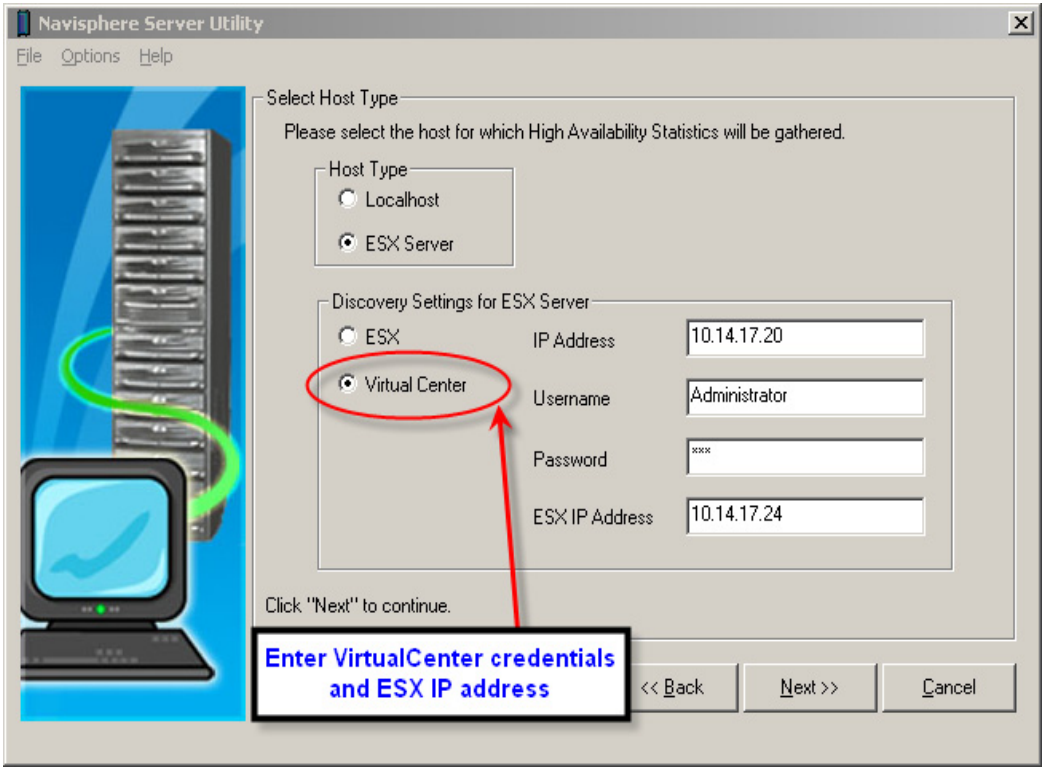
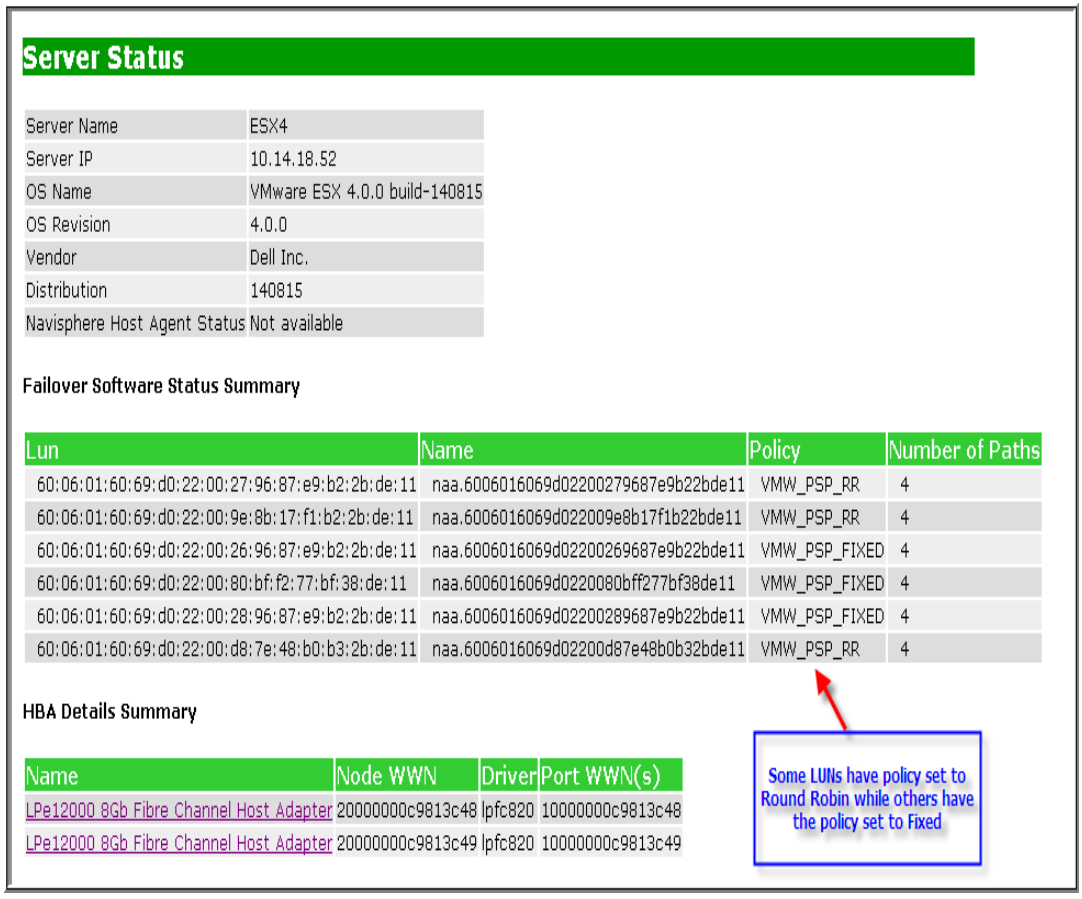


Figure 67 Using the server utility which is installed on a Windows host

Figure 68 on page 139 and Figure 69 on page 140 are examples for reports generated by the server utility for ESX 4.0 environments.



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 68 Report generated by the server utility showing ESX 4.0 NMP configuration information

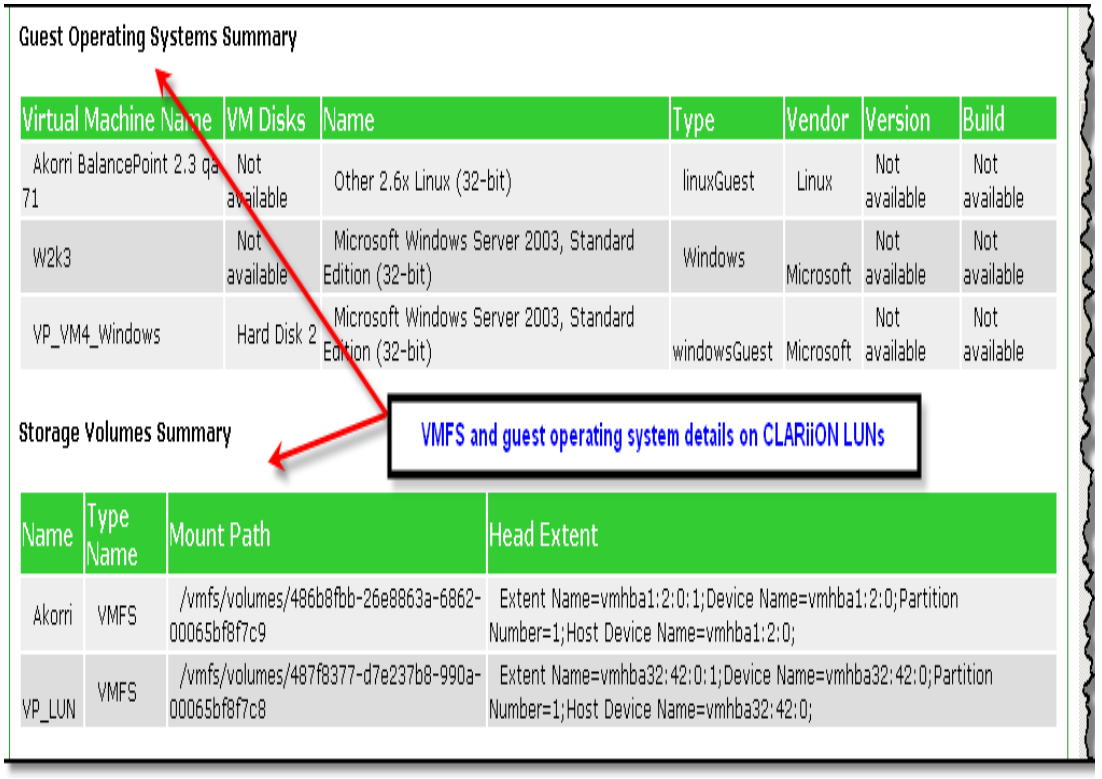


Figure 69 Report generated by the server utility showing guest OS and VMFS datastore information

Navisphere CLI can also be installed on virtual machines; 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.

Integration of host utilities with ESX Server

Navisphere Agent or the Server utility if supported must be installed on the ESX service console to register the ESX server with the CLARiiON storage system.

Navisphere CLI and array initialization software for the CX4, CX3, CX, and AX series storage systems can run on the ESX Server console as well as the individual virtual machines.

Navisphere Off-array for Windows is now supported to run on a Windows virtual machine.

Virtual Provisioning with ESX Server

A thin LUN can be used to create a VMware file system (VMFS), or assigned exclusively to a virtual machine as a raw disk mapping (RDM).

The VMFS datastore is *thin friendly*, meaning that it works well with thin LUNs. For one thing, when a VMware file system is created on Virtual Provisioning (thin) LUNs, the 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.

EMC recommends that you select the **zeroedthick** option when you create virtual disks on VMFS datastores, since this option does not initialize all blocks and claim all the space. Note that 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 upfront.

For RDM volumes, the file system or device created on the guest OS will dictate whether the RDM volume will be thin friendly.

Figure 70 on page 142 lists the allocation polices when creating new virtual disks.

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.

Figure 70 Allocation policies when creating new virtual disks on VMFS datastore

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 71 on page 144. 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.

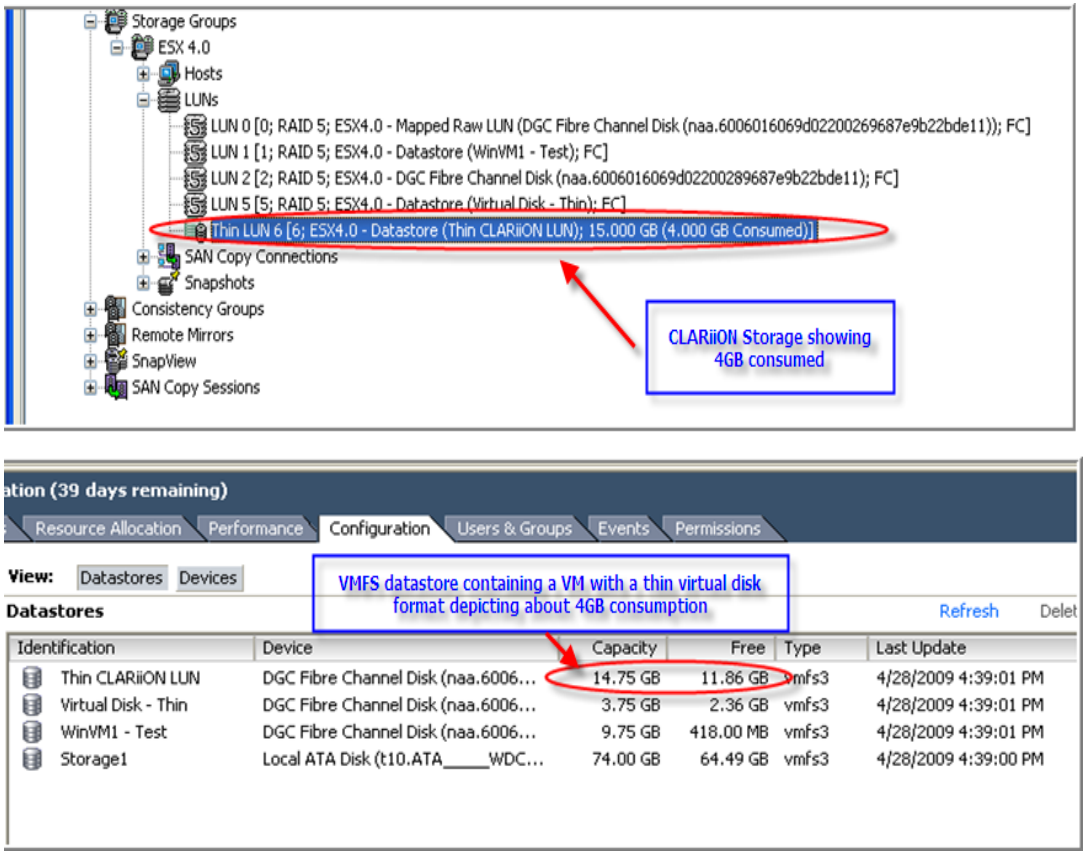


Figure 71 Using virtual disk "thin" format on CLARiiON thin LUNs

In VMware Infrastructure 3, VMware vCenter Converter can be used to clone virtual machines or migrate source LUNs from fully provisioned to virtually provisioned LUNs. The VMware vCenter Converter product is thin friendly. DRS, VMotion, and "cold" VM migration are unaffected. VM Clones Templates,Cold Migration and Storage VMotion are not thin-friendly. VM Cloning fully initializes and allocates all blocks. VMware Templates also allocate all blocks. The workaround is to shrink VMDKs before creating a template and use the "Compact" option. However deploying new virtual machines using templates initializes all blocks assigned to it.

Note: A future update to vCenter and ESX version 3.x will fix this issue.

In vSphere 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.

Navisphere QoS with ESX Server

Navisphere quality of service (QoS) functionality allows virtual machines configured on CLARiiON LUNs to achieve certain service levels based on the priority of the application running on these virtual machines.

Although QoS works at a LUN level, in a VMware environment the user can have LUNs configured as VMFS volumes, which allow multiple virtual machines to reside on the same LUN. Hence, to use the QoS functionality in a VMware environment, EMC and VMware recommend separation of higher priority virtual machines from the lower priority virtual machines on separate VMware file system created on disparate LUNs.

In addition, the user can create LUNs configured as VMFS or RDM volumes and dedicate them to an individual VM in order for QoS to provide the service level needed for the LUN and virtual machine.

Navisphere QoS (at the storage LUN level) when used in conjunction with VMware DRS (at the host CPU and memory level) provides an end-to-end service level protection for virtual machines. The details are presented in the white paper ***Maintaining End-to-End Service Levels for VMware Virtual Machines Using VMware DRS and EMC Navisphere QoS - Applied Technology.***

Mapping a VMware file system to CLARiiON devices

The mapping of the components of a VMware file system to the CLARiiON devices is a critical component when using EMC CLARiiON based storage software.

A new feature called *VM-aware Navisphere* 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. This feature is supported with ESX 4, 3.x and ESXi versions.

Figure 72 on page 146 displays VMFS datastore name information which helps to identify the LUNs, in addition right-clicking on the LUN will show all the virtual machine disks that are configured on this LUN.

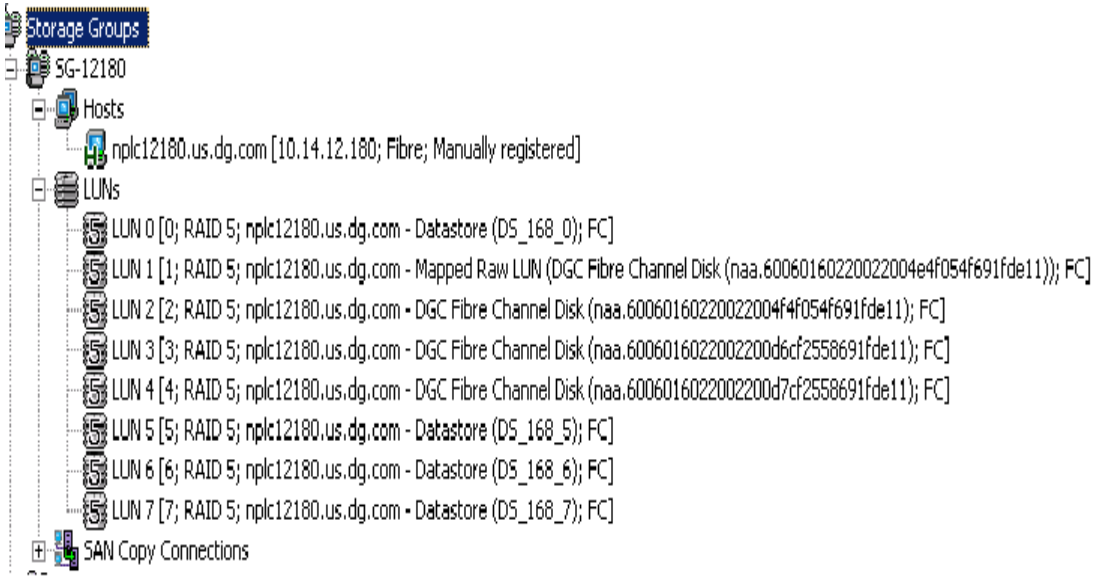


Figure 72 ESX 4 filesystem information display within Navisphere

If the CLARiiON is not running release 29 of the FLARE code, Navisphere Agent and CLI can be used to provide the canonical name of the CLARiiON device. Figure 73 on page 147 shows the device mapping information that is listed when the `lunmapinfo` command is issued from Navisphere CLI on the ESX Server service console. This command is directed to the agent residing on the ESX service console. The `lunmapinfo` command used for ESX 3.x servers can also run from a virtual machine running the Navisphere CLI software and use the IP address of the ESX service console that has the Navisphere agent software installed.

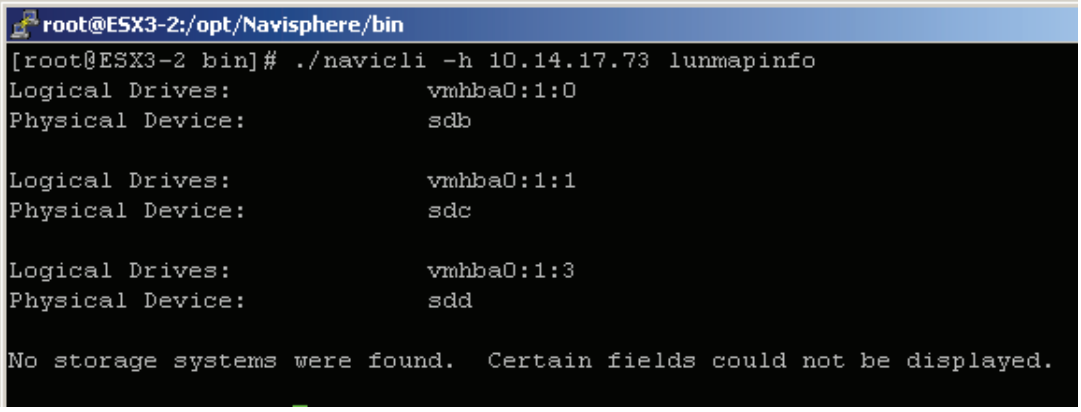


Figure 73 Executing the `lunmapinfo` command on the ESX Server console

For ESX 4.0 or ESXi servers, the **volmap** command is used to get device information as shown:

naviseccli -h 10.14.15.16 server -volmap -host 10.14.15.140

The canonical name and CLARiiON LUN co-relation are available through Navisphere Manager as shown in Figure 74 on page 148.

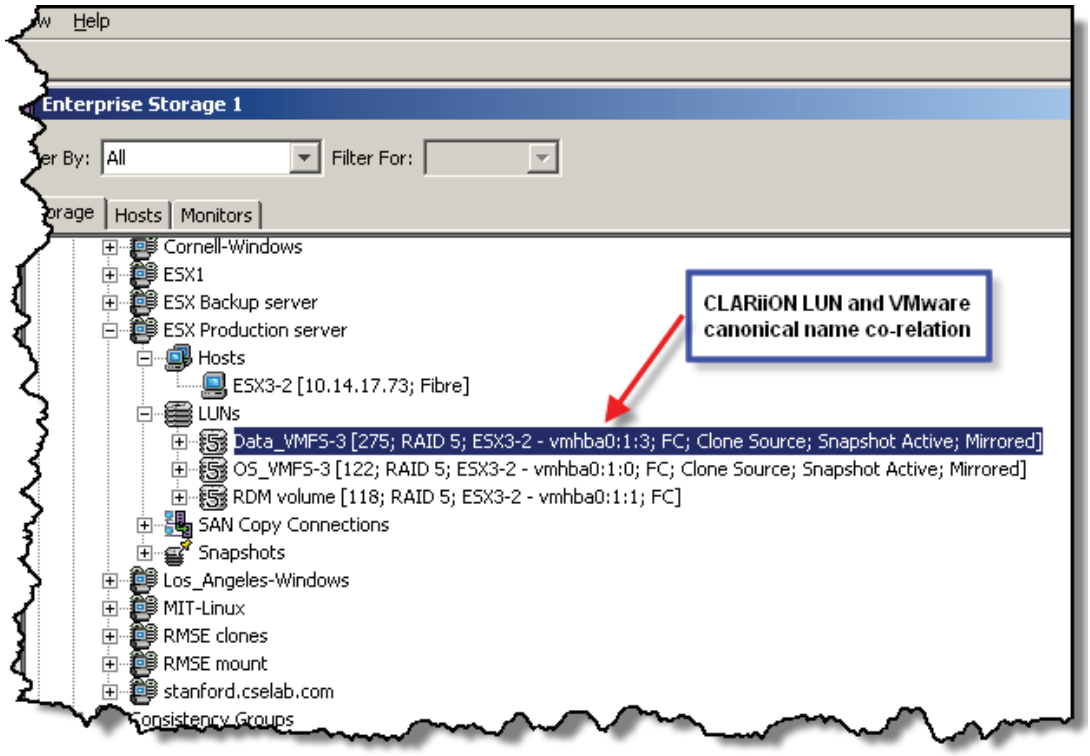


Figure 74 CLARiiON LUN and VMware canonical name co-relation in Navisphere Manager

From the canonical name `(vmhbax:y:z)`, users can obtain the VMware file system volume label information by interrogating the /vmfs/volumes structure on the VMware ESX/ESXi hosts. The command line utility, `vmkfstools`, can also be used to obtain the relationship between the canonical name and the VMware file system label (see Figure 75 on page 149).

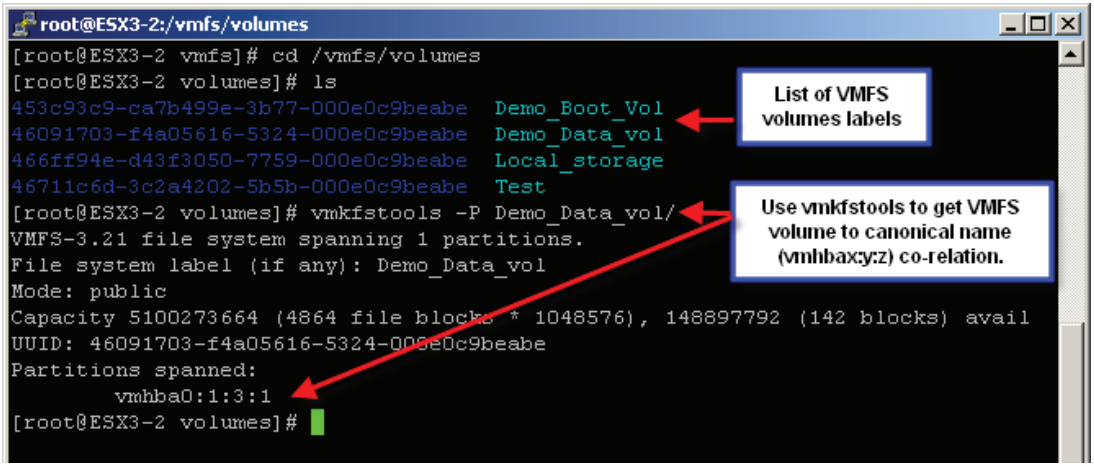


Figure 75 Using vmkfstools to determine mapping between a VMFS label and canonical name

The canonical name and VMFS volume label co-relation are also available through vCenter as shown in Figure 76 on page 150.

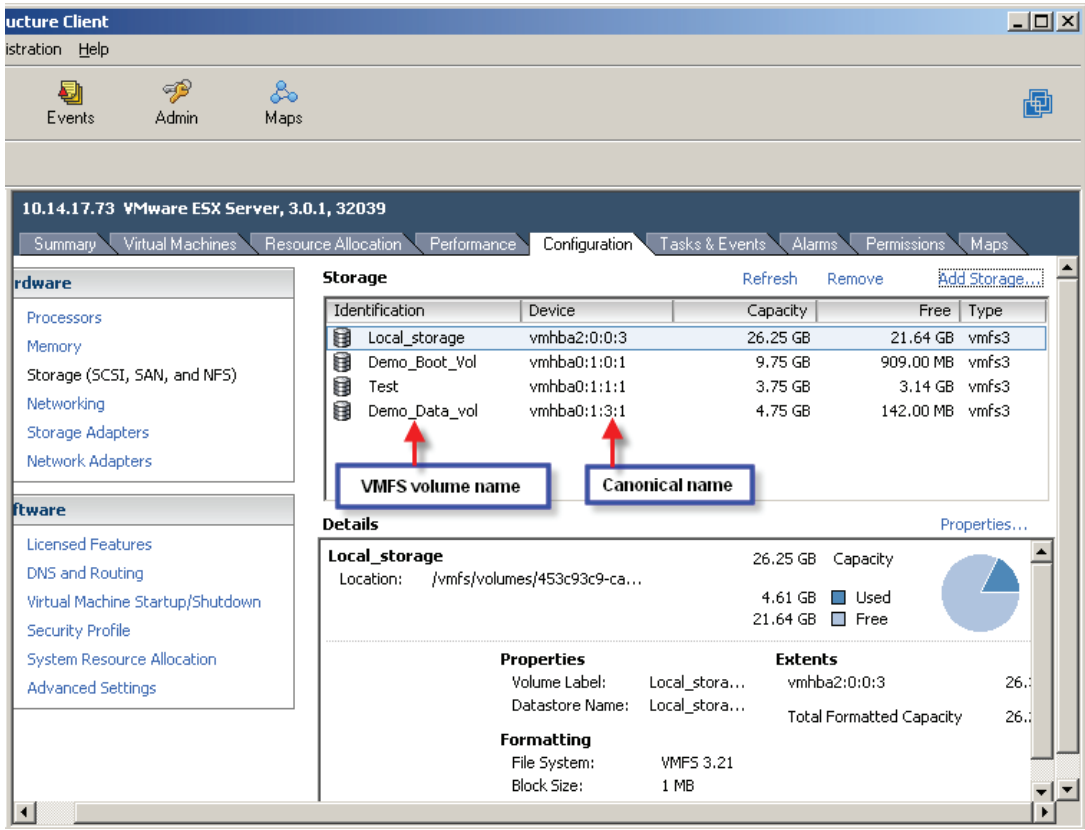


Figure 76 Using vCenter to determine the relationship between a VMFS label and canonical name

Mapping RDM to EMC CLARiiON devices

As mentioned earlier, Release 29 of FLARE allows Navisphere to display filesystem and virtual machine information. As shown in Figure 72 on page 146 , RDM volumes are also displayed within Navisphere. By right-clicking on a given RDM volume configured on a LUN, the virtual machine using the RDM volume can be determined.

If the CLARiiON storage system is not running FLARE 29, the EMC SCSI inquiry utility, inq, running on the virtual machine can be used to get device mapping information from the virtual machine to the

CLARiiON LUN. The virtual disks have to be configured as RDMs in physical compatibility mode for this to work correctly (see Figure 77 on page 151).

```
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 77 Using SCSI inquiry utility, `inq`, to map virtual machine RDM to CLARiiON LUN number

Optimizing VI infrastructure and CLARiiON

The EMC CLARiiON product line includes the CX4 series using UltraFlex technology and the AX4 series. EMC CLARiiON has a fully redundant, high-availability storage processor providing nondisruptive component replacements and code upgrades. The CLARiiON system features high levels of performance, data integrity, reliability, and availability. Configuring the CLARiiON storage array appropriately for a VMware ESX/ESXi host environment is critical to ensure a scalable, high-performance architecture. This section briefly discusses these best practices.

Storage considerations for 4, 3.x, and ESXi servers

Physical disk size and data protection

EMC CLARiiON storage arrays offer customers a wide choice of physical drives to meet different workloads. These include best performance drives (73-GB, 200-GB, and 400-GB Flash drives); better performance drives (73-GB 15k-rpm Fibre Channel drives); and cost-effective drives (600-GB 7200-rpm SATA-II drives) that are targeted

for less demanding workloads . Various drive sizes can be intermixed on the same storage array to allow customers with the option of providing different applications with the appropriate service level.

In addition to the different physical drives, EMC also offers various protection levels on EMC CLARiiON storage array. The physical drives can be configured as RAID 1, RAID 3, RAID 10, RAID 5 or RAID 6 groups. The RAID protection type can be mixed in the same storage array.

The flexibility provided by the EMC CLARiiON storage array enables customers to provide different service levels to the virtual machines using the same storage array. However, to configure appropriate storage for virtual infrastructure, a prior knowledge of the anticipated I/O workload is required. If this information is not available, the following general guidelines can be used to architect the storage:

1. Virtual machines boot volume is generally subject to low I/O rates. The boot volume can be on RAID 5 protected devices on large Fibre Channel drives, such as a 450 GB or 650 GB , 10k rpm drive.
2. If a separate virtual disk is provided for applications (binaries, application log, and so on), the virtual disk can be configured to use RAID 5 protected devices on large Fibre Channel drives. However, if the application performs extensive logging (for example, financial applications), a RAID 10 protected device may be more appropriate.
3. Infrastructure servers, such as DNS, perform a vast majority of their activity utilizing CPU and RAM. Therefore, low I/O activity is expected from virtual machines supporting the enterprise infrastructure functions. These servers should be provided with RAID 5 protected devices on medium size Fibre Channel drives.
4. Virtual machines that are anticipated to have a write-intensive workload should use RAID 10 protected devices on Flash drives or use medium size, fast Fibre Channel drives, such as 73 GB or 146 GB, 15k drives.
5. The log devices of databases should be on RAID 10 protected devices. OLTP databases should utilize small and fast drives. Furthermore, if database or application logs are mirrored, they should be on separate set of disks (and VMware file system, if applicable).
6. The virtual machines that generate high small block random I/O read workloadshould be allocated RAID 10 protected volumes. The use of RDM should be evaluated for these virtual machines.

7. Large file servers with vast majority of the storage consumed by static files can be provided with RAID 5 protected devices since the I/O activity is anticipated to be low. Medium size Fibre Channel drives, such as the 146 GB, 15k rpm drive, may be appropriate for these virtual machines. Microsoft technologies such as DFS, should be considered. Adoption of DFS, for example, enables tiering of storage while presenting a single namespace to the end user.
8. The 1 TB SATA-II drives should be considered for virtual machines that are used for storing archived data. The SATA-II could be RAID 5 or RAID 1 protected.

LUNs presented to the VMware ESX 4 or ESX 3 or ESXi cluster

The most common configuration of a VMware ESX 4, VMware ESX 3, or VMware ESXi cluster presents the storage to the virtual machines as flat files in a VMware file system (VMFS). It is, therefore, tempting to present the storage requirement for the VMware ESX/ESXi hosts as one large LUN. However, this can be detrimental to the scalability and performance characteristics of the environment.

Note: As discussed in "Creating VMFS volumes," on page 127 and "Using the vCenter client," on page 124 a VMware ESX 4, 3, or a VMware ESXi cluster is a logical grouping of VMware ESX/ESXi hosts sharing access to a set of VMFS configured in the public access mode.

Presenting the storage as one large LUN, forces the VMkernel to serially queue I/Os from all of the virtual machines utilizing the LUN. The VMware parameter, `Disk.SchedNumReqOutstanding` prevents one virtual machine from monopolizing the Fibre Channel queue for the LUN. Nevertheless, unpredictable elongation of response time results when there is a long queue against the LUN.

This problem can be further exacerbated in configurations that allow multiple VMware ESX/ESXi hosts to share a single LUN. In this configuration, the I/Os from all of the VMware ESX/ESXi hosts sharing the LUN queue on the EMC CLARiiON storage array Fibre Channel port. In a large farm or cluster with multiple active virtual machines, it is easy to overrun the queue on the EMC CLARiiON storage array front-end port. When such an event occurs, the benefits of moderate queuing are lost.

The potential response time elongation and performance degradation can be addressed by presenting the storage requirements for a VMware ESX/ESXi host cluster as a number of small LUNs. However, this imposes overhead for managing the virtual infrastructure.

Furthermore, VMFS datastores provides on-disk locking that prevents other ESX servers from accessing a given virtual machine at the same time. Also, to coordinate or update internal file information (metadata) for the VMFS datastore, ESX issues SCSI reservations on the entire LUN. Operations such as powering on or powering off a VM, creating and managing VM snapshots, VMotion, and so forth also require the VMFS datastore to lock the entire for a short period of time.

If such operations occur frequently on a VMFS datastore that is shared by multiple ESX servers, a user might see some performance degration. Hence, it is best to reduce the virtual machines residing on a single VMFS datastore that is accessed by multiple ESX servers. The total number of SCSI devices supported by vSphere 4 and ESX version 3 of VMkernel has increased from 128 to 256 , the overhead of managing large number of small LUNs can be prohibitive.

Table 1 on page 155 compares the advantages and disadvantages of presenting the storage to a VMware ESX cluster as a single or multiple LUNs. The table shows that the benefits of presenting storage as multiple LUNs overcome the disadvantages.

The anticipated I/O activity influences the maximum size of the LUN that can be presented to the VMware ESX/ESXi hosts. The storage stack on VMware ESXversion 4, 3 and VMware ESXi has a number of enhancements that is anticipated to provide better performance. However, due to greater scalability provided by VMware vSphere and Virtual Infrastructure 3 environments (for example, support for more virtual CPUs and RAM in each virtual machine), these environments are anticipated to handle more I/O intensive workload. Typically, most environments configure the maximum LUN size of 300-500 GB in a VMware ESX/ESXi 4 or 3 cluster.

Note: The EMC CLARiiON storage arrays support nondisruptive expansion of LUNs. This functionality can be exploited to grow VMware ESX/ESXi host LUNs over the common size of 300 to 500 GB. If the performance characteristics of the virtual infrastructure can support larger LUNs, the aforementioned technique of nondisruptive LUN expansion using metaLUNs can be used to provide larger LUNs to the virtual infrastructure. Furthermore, as discussed in "Spanned VMware file system," on page 156 the spanning functionality of VMware file system can be used to present a single VMware file system using the expanded LUN. Appendix A presents a procedure for using the nondisruptive expansion of LUNs in VMware virtual infrastructure.

Table 1 Comparing different approaches for presenting storage to VMware ESX/ESXi hosts

Category	Storage as single LUN	Storage as multiple LUNs
Management	Easier management. Storage can be under-provisioned. One VMFS to manage	Small management overhead. Storage provisioning has to be on demand. One VMFS to manage (spanned)
Performance	Can result in poor response time. No opportunity to perform manual load balancing	Multiple queues to storage ensure minimal response times. Opportunity to perform manual load balancing
Scalability	Limits number of virtual machines due to response time elongation. Limits number of I/O-intensive virtual machines	Multiple VMFS allow more virtual machines per ESX Server. Response time of limited concern (can optimize)
Functionality	All virtual machines share one LUN. Cannot leverage all available storage functionality	Multiple VMFS allow more virtual machines per ESX Server. Response time of limited concern (can optimize)

Number of VMware file systems (VMFS) in a ESX Server 4 or 3 cluster

Virtualization enables better utilization of IT assets. However, the fundamentals for managing information in the virtualized environment are no different from a physical environment. EMC recommends the following best practices for a virtualized infrastructure:

- A VMware file system to store virtual machine boot disks. In most modern operating systems, there is minimal I/O to the boot disk. Furthermore, most of the I/O to boot disk tend to be paging activity that is sensitive to response time. By separating the boot disks from application data, the risk of response time elongation due to application related I/O activity is mitigated.
- Data managers, such as Microsoft SQL Server or Oracle, use an active log and/or recovery data structure that track changes to the data. In case of an unplanned application or operating system disruption, the active log or the recovery data structure is critical to ensure proper recovery and data consistency. Since the recovery structures are a critical component, any virtual machine that

supports data managers should be provided a separate VMware file system for storing active log files and other structures critical for recovery. Furthermore, if mirrored recovery structures are employed, the copy should be stored in a separate VMware file system.

- Application data, including database files, should be stored in a separate VMware file system. Furthermore, this file system should not contain any structures that are critical for application and/or database recovery.
- As discussed in "Physical disk size and data protection," on page 151, VMware ESX/ESXi hosts serialize and queue all I/Os scheduled for a SCSI target. The average response time from the disk depends on the average queue length and residency in the queue. As the utilization rate of the disks increases, the queue length and hence the response time, increases nonlinearly. Therefore, applications requiring high performance or predictable response time should be provided their own VMware file systems. Multiple VMware file systems may be needed to meet the performance requirements.
- VMware ESX version 4, VMware ESX 3, and VMware ESXi do not provide a sophisticated mechanism to control access to slowest component in modern computing—the disk subsystem. Due to this limitation, if a VMware file system is shared across all virtual machines, it is easy for noncritical servers to impact the performance of business-critical servers. Hence, virtual machines with different service-level requirements should be separated on their own VMware file system.
- It is recommended that the VMFS volumes be about 80% or less full. This would allow administrators to accommodate space for user data as quickly as possible as well as accommodate space for VMware snapshots for making copies of the virtual machines.

Spanned VMware file system

Enterprise VMware ESX 4, VMware ESX 3, and VMware ESXi clusters contain several VMware ESX/ESXi hosts sharing a common group of VMware file systems and SAN storage. Recommendations listed in "LUNs presented to the VMware ESX 4 or ESX 3 or ESXi cluster," on page 153 result in architecture with approximately 10 to 20 EMC CLARiiON LUNs configured for a VMware ESX 4, VMware ESX 3, or VMware ESXi cluster. Also, as discussed in "Number of VMware file systems (VMFS) in a ESX Server 4 or 3 cluster," on page 155, most common configurations of virtual infrastructure results in

approximately six VMware file systems; but this is directly proportional to the number of VMware ESX/ESXi hosts and the virtual machines that are hosted on these VMware ESX/ESXi hosts.

The simplest virtual infrastructure architecture creates a VMware file system on every SCSI disk in the VMware ESX/ESXi hosts cluster. However, this approach can result in inefficient use of storage and management overhead. A VMware file system that uses more than one EMC CLARiiON LUNs is required to reconcile the recommendations listed in "LUNs presented to the VMware ESX 4 or ESX 3 or ESXi cluster," on page 153 and "Number of VMware file systems (VMFS) in a ESX Server 4 or 3 cluster," on page 155.

Version 3 of VMware file system (VMFS-3) supports concatenation of multiple SCSI disks to create a single file system. VMware ESX 3.x and vSphere 4.x support the VMFS-3 filesystem. Allocation schemes used in VMware file system version 3 spread the data across all LUNs supporting the file system thus exploiting all available spindles. EMC recommends using this functionality while using VMware ESX/ESXi hosts with EMC CLARiiON storage systems.

CAUTION

The spanning functionality in VMFS-3 was enhanced from theVMFS-2 volume. If a member of a spanned VMFS-3 volume is unavailable, the datastore will be still available for use, except the data from the missing extent. An example of this situation is shown in Figure 78 on page 158.

Although the loss of a physical extent is not of great concern in the EMC CLARiiON storage systems, good change control mechanisms are required to prevent inadvertent loss of access.

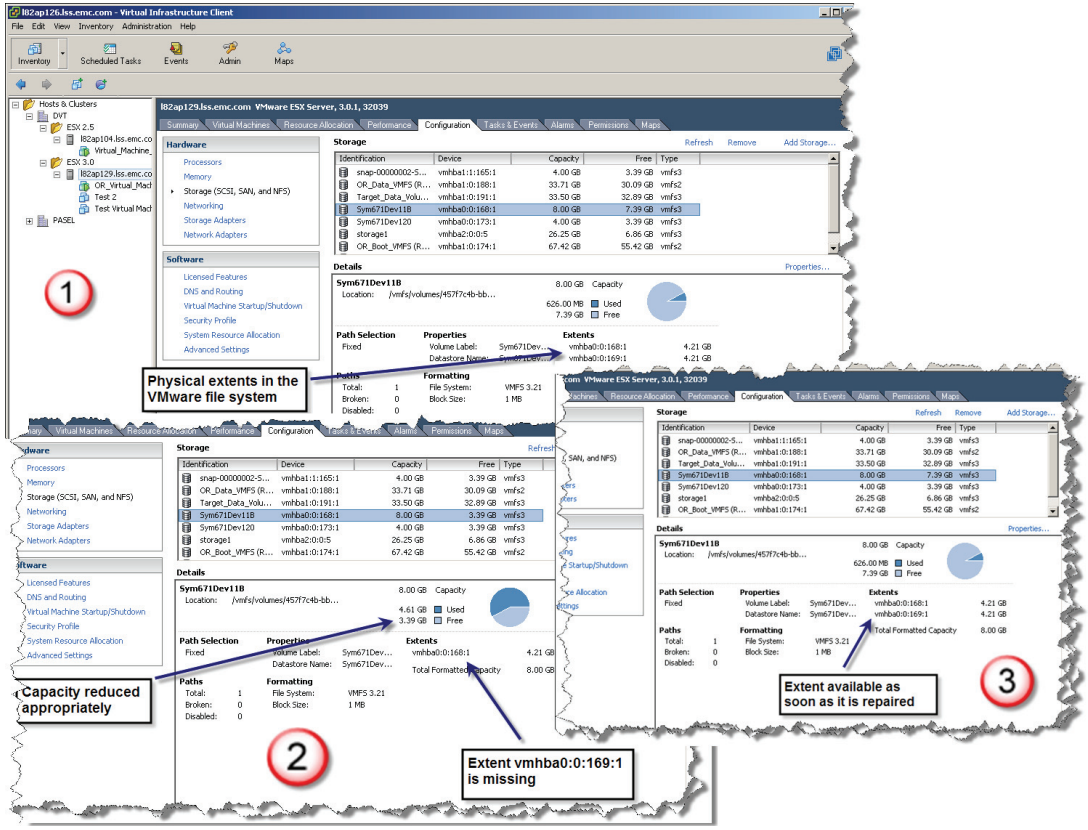


Figure 78 Spanned VMFS-3 tolerance to missing physical extent

Use of CLARiiON metaLUNs

The use of metaLUNs is recommended for applications that have higher bandwidth needs where multiple disks are working simultaneously. MetaLUNs are also recommended when creating large LUNs in order to spread the load across multiple disks. Always balance your LUNs or metaLUNs across the two CLARiiON storage processors for better performance.

CLARiiON metaLUNs can be used in conjunction with VMFS spanning, where two or more LUNs could be striped at the CLARiiON level and then concatenated at the VMFS volume level. This would help spread the I/O load across all the disks.

VMFS metadata, which is frequently accessed information, is spread across the volume. Hence the metadata information does not reside on one particular section of drives when you create multiple LUNs on a RAID group, metaLUNs across more than one RAID group, or LUNs in a storage pool. For additional details on using CLARiiON metaluns with VMFS and RDM volumes, see Appendix A.

Using VMware's Paravirtualized SCSI adapter in ESX 4

Paravirtualized SCSI (PVSCI) adapter is a new adapter that can be used with virtual disks. It is recommended to use PVSCSI adapters as this offers improved I/O performance with as much as 18 percent reduction in ESX 4 host CPU usage. It also reduces the cost of virtual interrupts and batches the processing of I/O requests. With vSphere Update 1, PVSCSI adapter is supported for both boot and data virtual disks. Using PVSCSI with Windows 2003 and 2008 guest OS has been found to improve the virtual machine resiliency during Celerra and storage network failure events.

To configure the PVSCSI:

- In the vSphere Client, right-click the virtual machine icon and click **Edit Settings**. Under the **Hardware** tab click **add Hard disk** and choose a Virtual Device Node between SCSI (1:0) to SCSI (3:15).
- Select the newly created controller for the disk as VMware **Paravirtual.**

Path management

Path failover and load balancing

VMware ESX versions 4, VMware ESX 3, and VMware ESXi provide native channel failover capabilities using VMware's own native multipathing software or EMC PowerPath/VE.

CLARiiON storage systems support nondisruptive upgrade (NDU) operations for VMware's native failover software and EMC PowerPath/VE. E-Lab Navigator has a list of ESX Server versions for which NDU operations are supported. We recommend that you keep the **auto-assign** parameter disabled (this is the default setting) 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 primus case emc16594.

Configuring failover on VMware ESX version 3.x and VMware ESX3i

The path management in VMware ESX/ESXi version 3 supports both FC and iSCSI devices from the CLARiiON. . VMware ESX/ESXi version 3.x introduced a new command, `esxcfg-mpath`, to view the configuration and status of the paths of the devices. As shown in Figure 79 on page 160, the native failover software provides a listing of the paths—whether active or passive—from the VMware ESX host to the CLARiiON storage system. In VMware ESXi, you can achieve the same output using the remotecli software installed on Windows and Linux machines.

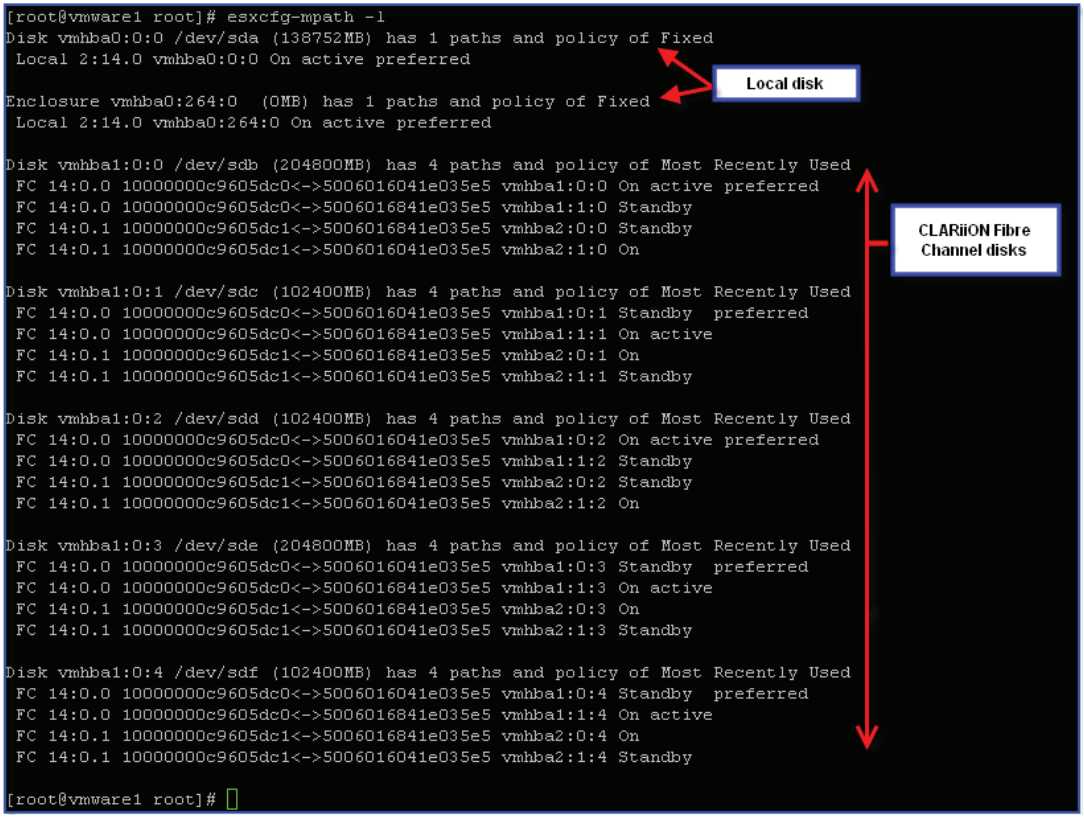


Figure 79 Output of the esxcfg-mpath command for displaying path information on ESX Server 3

Figure 79 on page 160 shows five LUNs attached to the CLARiiON storage system. The `vmhba1:x:x` are Fibre Channel LUNs. Each Fibre Channel LUN has 4 paths, two paths to each SP. The active label displays the path that is used by the ESX Server to access the disk. The preferred label displays the preferred path and is ignored since the policy is set to Most Recently Used (MRU). Device `vmhba0:0:0` and `vmhba0:264:0` are internal devices that have a single path.

Figure 80 on page 161 shows four LUNs attached to the CLARiiON storage system. The vmhba40:0:x are iSCSI LUNs attached to the ESX Server using the iSCSI software initiator. All iSCSI devices have paths going to both storage processors. The network adapters supporting the iSCSI software initiators need to be connected to the same subnet for transparent path failover if NIC teaming is configured between the network adapters. Furthermore, it is appropriate to isolate the iSCSI traffic from other IP traffic by creating dedicated virtual switches for iSCSI traffic. The output shown in Figure 80 on page 161 would be similar if hardware iSCSI initiators instead of the software iSCSI initiators are used.

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

Figure 80 Output of esxcfg-mpath for iSCSI LUNs

In 3.x, the most recently used MRU policy is the default policy for active/passive storage devices. 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 case 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 HBA1-SPB path on ESX2 then fails, the LUN will trespass back and forth between the SPs, which could result in performance degradation. In addition, if the ESX server reboots, all LUNs will end up of a single SP. Hence, EMC recommends that you always have four connections to the ESX server with each HBA having access to both SPs.

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 the ESX/ESXi host 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 that 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.

Configuring failover on VMware ESX version 4 and VMware ESXi 4.x

With ESX 4 and ESXi versions, CLARiiON supports two Multipath plug-ins (MPP) including VMware's native multipathing plugin (NMP) and PowerPath/VE. PowerPath/VE is a third-party MPP. If multiple MPPs are installed on the ESX 4 server, these MPPs cannot manage the same storage LUNs, so claim rules allow you to designate which MPP is assigned to which storage LUN.

Claim rules are used to assign storage devices to either PowerPath or NMP devices. Claim rules are defined within /etc/vmware/esx.conf on each ESX host and can be managed via the vSphere CLI.

NMP- Native Multi-Pathing Plug-in provides native multipathing on the ESX 4 server and is built in to the VMware kernel.

SATP- Storage Array Type Plug-in monitors path health, reports changes in those paths, and enables inactive paths in storage failover situations.

PSP- Path Selection Plug-in allows users to select a load-balancing policy within MPP. For example, Fixed, MRU and Round Robin are the native default path selections

VMware native multipathing and failover on ESX 4.x or ESXi 4.x with CLARiiON

VMware ESX 4.x contains its own native multipathing software that is built into its kernel. This failover software, called Native Multipathing Plugin (NMP), has three policies or PSP (Path Selection Plug-ins):

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

On VMware ESX 4.x or ESXi 4.x with CX4 arrays, the FIXED or Round Robin policy is supported.

VMware's native failover with vSphere allows you to the Robin Round or FIXED policy in ALUA mode. MRU policy only works with failovermode 1, and is not a recommended when attaching CX4 arrays to VMware vSphere 4.

The FIXED policy on the CX4 provides failback capability. To use the FIXED POLICY, you must be running FLARE release 28 with the latest patch (702 or higher). Also, failovermode mode must be set to 4 (ALUA mode or Asymmetric Active/Active mode). 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.

The default failovermode for ESX 4.x is 1. Use the **Failover Setup Wizard** within Navisphere to change the failovermode from 1 to 4. You need to reboot the ESX server after changing the failovermode on the

array due to the claim rules configured within the ESX SCSI architecture. In addition, the failover wizard should be used within Navisphere Manager to change the failovermode.

With the FIXED policy, which is sometimes referred to as FiXED, there is initial setup in which you select the *preferred* (or optimal) path. If you set this up properly, you will not see a performance hit with ALUA. Note that FIXED sends I/O down a single path. However, if you have multiple LUNs in your environment, you can choose different preferred paths for different LUNs to achieve static I/O load balancing. FIXED performs an automatic restore, so LUNs do not end up on a single SP after an NDU. Figure 81 on page 166 shows how the FIXED is configured with the CX4 storage system using VMware's NMP software.

When using Round Robin there is no autorestore. As a result, all LUNs end up on a single SP after an NDU A, and you need to manually trespass some LUNs to the other SP balance the load. The benefit of Robin Round is that not too many manual setups are necessary, because it uses the optimal path by default,and it “does primitive load” balancing (however 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.

NMP Round Robin will only alternate I/O between the NMP Active paths. Moreover, in ALUA mode the NMP RR knows the difference between the Active and Active-unoptimized paths, and it only uses Active paths for the I/O.

If you had a script that takes care of the manual trespass issue, then Robin Round would be the way to go to avoid manual configuration. The other option would be to issue the following commands on the CLARiiON storage system after an NDU.

Naviseccli -h <SPA> trespass mine

Naviseccli -h <SPB> trespass mine

By executing this commands, all LUNs will then trespass to their default owner. In addition, there is an issue with trespasses when using the NMP Round Robin policy is used with ALUA (failovermode=4) after an ESX reboots.

Issue:

If a user uses the vClient GUI or the **esxcli nmp device setpolicy** command to set the policy for each device, after an ESX reboot the vmkernel creates the device with the default configuration specified by

the claim-rules (for example PSP_FIXED) first. Only after that vmkernel can apply the device-specific configuration (PSP_RR) to the device. In this particular case having PSP_FIXED claim the device for a brief period of time is enough to cause the unwanted path failovers.

Workaround:

Change the default PSP for SATP_ALUA_CX to be VMW_PSP_RR instead of VMW_PSP_FIXED using the following command

esxcli nmp satp setdefaultpsp --satp=VMW_SATP_ALUA_CX --psp=VMW_PSP_RR

If you want some of your devices to be managed by PSP_FIXED or CLI PSP_MRU, you can configure these devices separately using the GUI or "esxcli nmp device setpolicy" command.

In this case all devices for CLARiiON in ALUA mode will first be claimed by the VMW_PSP_RR, and then some of them will be switched to FIXED or MRU. For additional details see article emc232355.

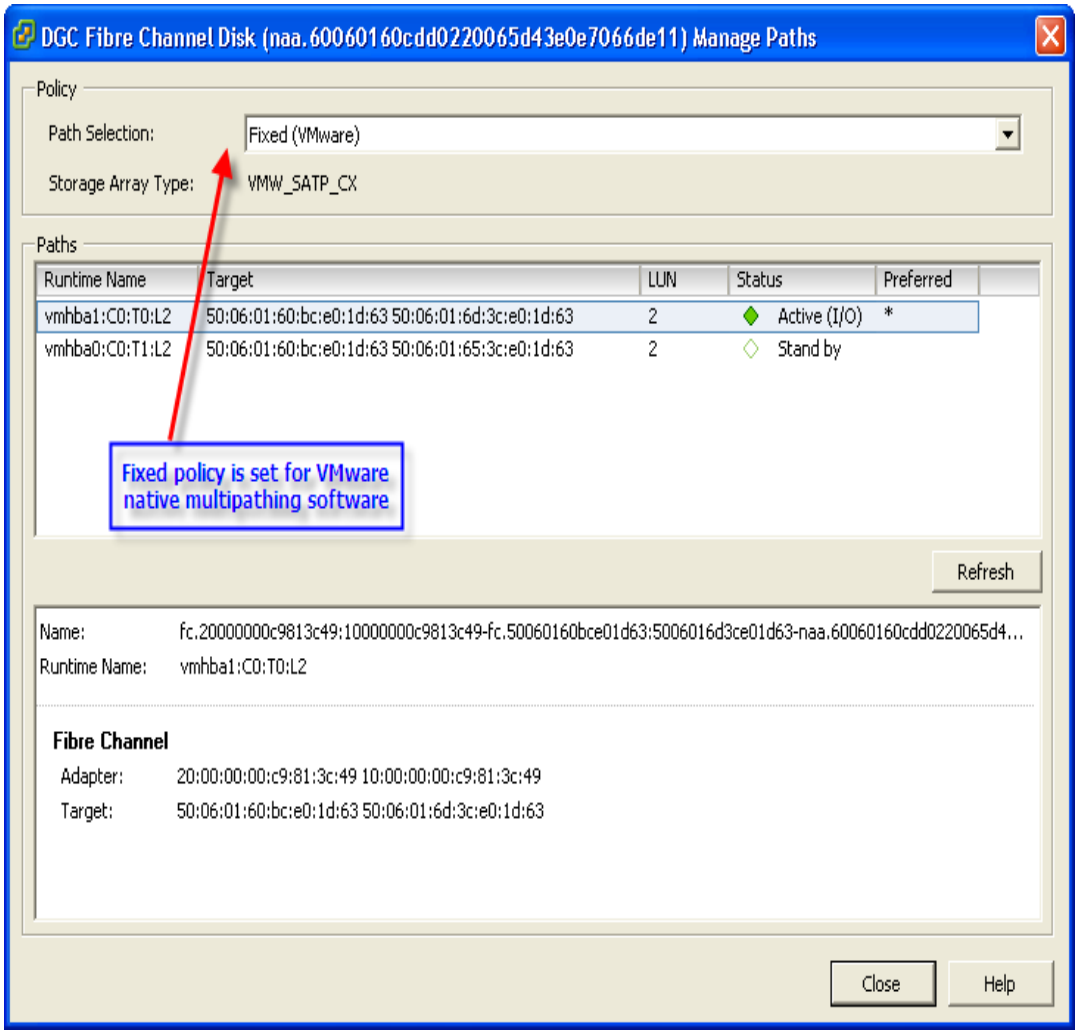


Figure 81 NMP software configured with the FIXED policy for a CLARiiON system

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. The path selected for I/O is controlled by the round robin algorithm. The FIXED and MRU policy also send I/O down only a single selected path for a given LUN, unless that path becomes unavailable.

EMC PowerPath multipathing and failover on ESX 4.x with CLARiiON

EMC PowerPath software is supported on the ESX 4.x 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. For iSCSI both the software and hardware initiators are supported with PowerPath. Some of the benefits of using PowerPath with ESX 4.0 are as follows:

- PowerPath on ESX 4.x 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 82 on page 168 depicts the CLARiiON LUNs controlled by EMC PowerPath software.

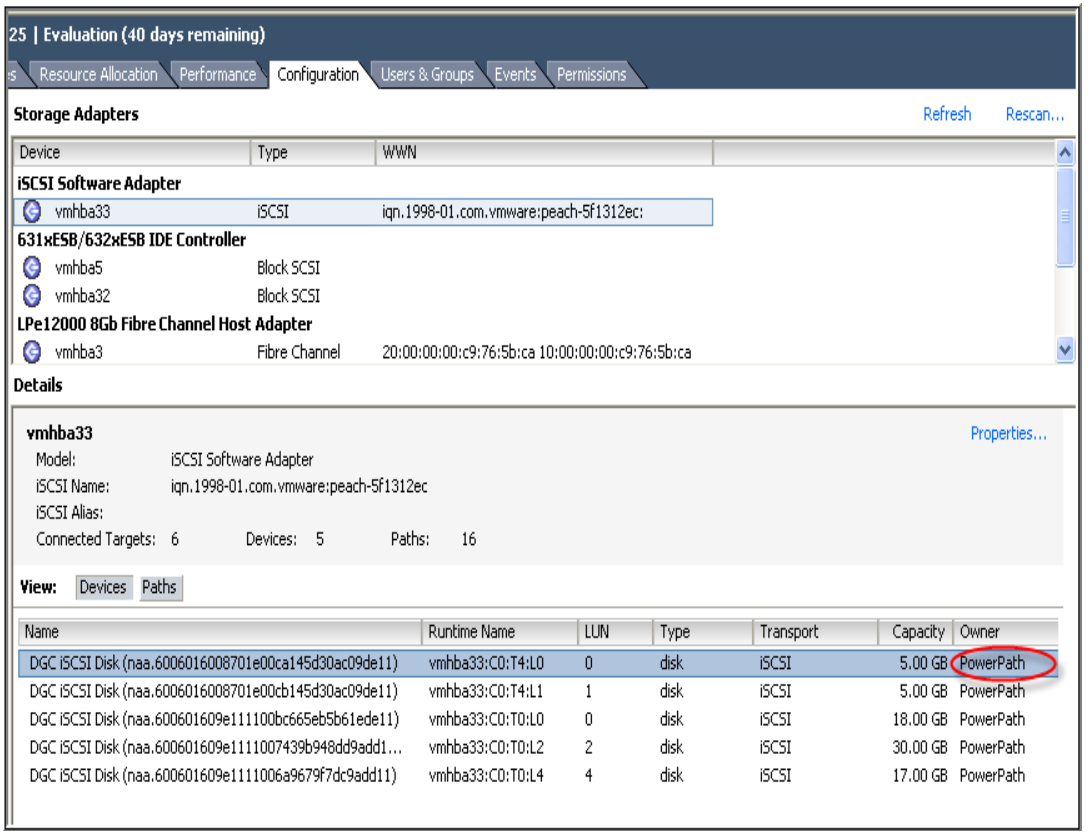


Figure 82 EMC PowerPath/VE software configured on ESX 4.0 connected to a CLARiiON system

iSCSI configurations and multipathing with ESX 4.0

This section 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 83 on page 169. The two NICs or vmkernel ports should be on different subnets. Also, ensure the SP ports for a storage processor are on different subnets.

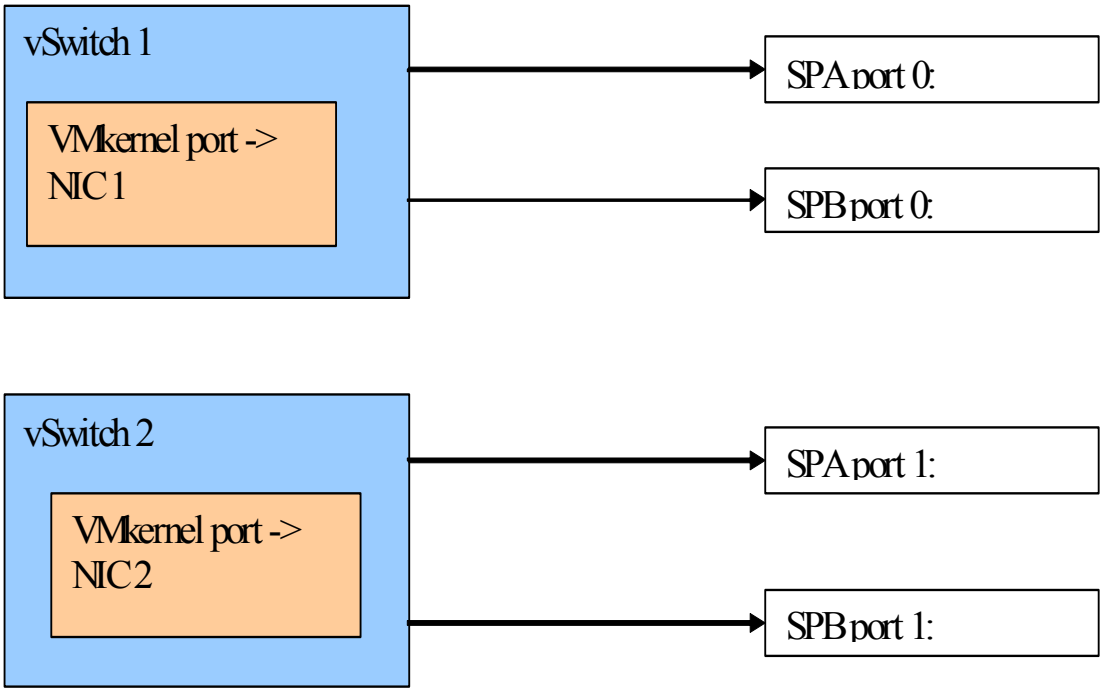


Figure 83 Dual virtual switch iSCSI configuration

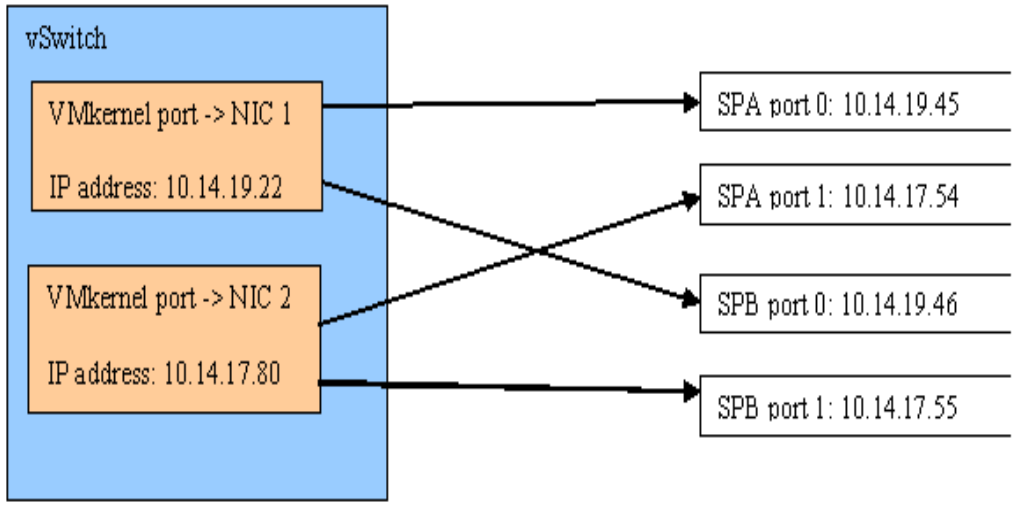


Figure 84 Single vSwitch iSCSI configuration

Table 2 shows the failover mode and policies supported on ESX servers.

Table 2 Failoverrmode and policies

ESX version	ALUA PowerPath (failovermode =4)	ALUA Native (failovermode=4)	PNR mode PowerPath (failovermode=1)	PNR mode Native (failovermode=1)
ESX 4.x	Yes (CX arrays running R26 or higher)	Yes (FIXED or Round Robin) (CX4 systems running R28 version 04.28.000.5.704 or later)	Yes (CX arrays running R22 or higher)	Yes (Round Robin or MRU) (CX arrays running R22 or higher)
ESX 3.x	No	No	No	Yes. - MRU (CX arrays running R22 or higher)

Partition alignment

Modern hard disk systems use the logical block address (LBA) to position the head. This is true for both SCSI and IDE disks. However, older disks systems used a different addressing scheme called CHS (cylinder, head, and sectors) to describe the geometry of the drive. Hard disks using this addressing scheme expect three numbers to position the disk head accurately. Various specifications for IDE and BIOS have evolved over the years to accommodate larger disk storage capacities. These standards provide various combinations for the maximum value for CHS. These range from 1024-65536 for cylinders, 16-255 for heads, and 1-255 sectors per track.

The BIOS of all x86-based computers still supports CHS addressing. The BIOS also provides a mechanism that maps LBA addresses to CHS addresses using the geometry information provided by the disks. Modern operating systems, such as Linux and VMware ESX/ESXi , do not normally use the mapping information provided by the BIOS to access the disk. However, these operating systems need the geometry information when communicating with the BIOS or with other operating systems that use CHS mapping information, such as DOS or Microsoft Windows.

The first cylinder of all hard disks contains a reserved area called the master boot record (MBR). When an IBM compatible system is started, the BIOS reads the MBR from the first available disk. The bootstrap loader code found at this location is used to load the operating system. The MBR also contains critical partition table information for four entries describing the location of the primary data partitions on the disk. The partition table structure resembles:

```
struct partition {
char active;    /* 0x80: bootable, 0: not bootable */
char begin[3];  /* CHS for first sector */
char type;
char end[3];    /* CHS for last sector */
int start;     /* 32 bit sector number (counting from 0) */
int length;     /* 32 bit number of sectors */
};
```

The information in the structure is redundant– the location of a partition is given both by the 24-bit begin and end fields, and by the 32-bit start and length fields. Only one of the two sets of field is needed to describe the location of the partition. VMware ESX/ESXi hosts uses

the start and length fields of the partition table structure. By default, the VMware ESX/ESXi host creates the first data partition starting at the first available LBA after the area reserved for the MBR.

Assuming the default stripe element size of EMC CLARiiON storage arrays all I/O of 64 KB will cause disk crossing. Therefore, with using the default configuration for disk partitions results in inefficient use of storage components.

Prior experience with misaligned Windows partitions and file systems has shown as much as 20 to 30 percent degradation in performance. Studies by VMware have shown similar impact in a VMware ESX/ESXi version 3.x environment. The VMware technical paper, ***Recommendations for Aligning VMFS Partitions,*** provides further details. Aligning the data partitions on 64 KB boundary results in positive improvements in overall I/O response time experienced by all hosts connected to the shared storage array.

Alignment for VMs using VMFS volumes

VMware ESX/ESXi hosts, by default, create VMware file system on the data partition using one MB block size. Since the block size is a multiple of the stripe element size, file allocations is in even multiples of the stripe element size. Thus, virtual disks created on the partitions normally created by VMware ESX/ESXi hosts are always track-misaligned.

A virtual disk created on VMware file system is presented to the guest operating system with geometry of 63 sectors. The Phoenix BIOS used by the virtual machine reserves one track of the virtual disk for storing the MBR. The data partition created on the virtual disk starts at sector 64. The guest operating-system layout exacerbates the track-misalignment problem created by the VMware ESX/ESXi host—the I/Os generated by the guest operating system is sector misaligned.

EMC recommends a two-step process to address the sector misalignment issue. Aligning both the VMware file system and the virtual disk on a track boundary ensures the optimal performance from the storage subsystem. Furthermore, EMC recommends aligning the partitions on 64 KB boundaries. This ensures optimal performance on all EMC storage platforms.

CAUTION

The benefits of aligning boot partitions are generally marginal. It is more important to align the heaviest I/O workloads located in app/data disk partitions. The partition alignment discussed in this section applies only to volumes containing application data.

Creating track-aligned VMFS on ESX Server version 4, 3 and ESXi

VMFS volumes created utilizing the Virtual Infrastructure client are automatically aligned on a 64 KB boundary. EMC recommends the use of vClient to create VMFS volumes and VMware file systems in vSphere andVirtual Infrastructure 3 environments.

Track-aligned virtual disks in vSphere 4, Virtual Infrastructure 3 and VMware ESXi

The virtual disks created on a VMware file system are presented to the guest OS with a geometry of 63 sectors. It is critical to align the virtual disks on a track boundary in addition to aligning the VMware file system. VMware recommends this for VMFS data partitions to reduce latency and increase throughput. The process of aligning virtual disks should be performed in the virtual machine, and is the same as the one used for physical servers.

Microsoft recommends aligning virtual machine partitions to 1 MB track boundaries for most Windows systems when using shared storage such as CLARiiON (see Microsoft TechNet article 92949). In addition, when formatting an NTFS volume, set the allocation size to a multiple of 8KB, otherwise set it to the value that the application recommends.

In additon, for Linux virtual machines, the disk can also be aligned to 1 MB. Use the fdisk command to create an aligned partition. For details procedure to align Windows and Linux virtual machines, see the ***Host Connectivity Guide for Windows*** and ***Host Connectivity Guide for Linux*** available on Powerlink.

Alignment for VMs using RDM

EMC CLARiiON devices accessed by virtual machines using RDM do not contain VMware file systems. In this configuration, the alignment problem is the same as that seen on physical servers. The process employed for aligning partitions on physical servers needs to be used in the virtual machines. "Creating track-aligned VMFS on ESX Server version 4, 3 and ESXi," on page 173, outlines the process that is required for both Microsoft Windows and Linux operating systems.

4
Cloning of Virtual Machines

This chapter presents the following topics:

- Overview 177
- Copying virtual machines after shutdown..... 178
- Using EMC to copy running virtual machines 193
- Transitioning disk copies to cloned virtual machines..... 200
- Choosing a VM cloning methodology 221

VMware ESX/ESXi virtualizes IT assets into a flexible, cost-effective pool of compute, storage, and networking resources. These resources can then be mapped to specific business needs by creating virtual machines. VMware ESX/ESXi provides several utilities to manage the environment. This includes utilities to clone, back up, and restore virtual machines. All these utilities use host CPU resources to perform the functions. Furthermore, the utilities cannot operate on data not residing in the VMware infrastructure.

Large enterprises have line-of-business operations that interact and operate with data on disparate set of applications and operating systems. These enterprises can benefit by leveraging technology offered by storage array vendors to provide alternative methodologies to protect and replicate the data. The same storage technology can also be used for presenting various organizations in enterprises with a point-in-time view of their data without any disruption or impact to the production workload.

VMware ESX/ESXi hosts can be used in conjunction with SAN-attached EMC CLARiiON storage arrays and the advanced storage functionality they offer. The configuration of virtualized servers when used in conjunction with EMC CLARiiON storage array functionality is not very different from the setup used if the applications were running on a physical server. However, it is critical to ensure proper configuration of both the storage array and VMware ESX/ESXihosts so applications in the virtual environment can exploit storage array functionality. The focus of this chapter is the use of EMC SnapView with VMware ESX/ESXi to clone virtual machines and their data.

Overview

The EMC SnapView family of products provides different technologies to enable users to nondisruptively create and manage local point-in-time copies of data. The copies of the data can be used to offload operational processes, such as backup, reporting, and application testing, from the production environment. The creation of the copies is performed independent of the source application without impacting performance or availability. The SnapView family includes two different products, SnapView clones and SnapView snapshots. Detailed description of these products is available in "EMC SnapView," on page 77 and on EMC Powerlink.

SnapView products run on the EMC CLARiiON storage array. However, the management of the functionality is performed using either Navisphere Manager or Navisphere command line interface (navicli). Using SnapView on VMware file system requires that all extents of the file system be replicated. If proper planning and procedures are not followed, this requirement forces replication of all data that is present on a VMware file system, including virtual disk images not needed. Therefore, EMC recommends separation of virtual machines that require use of storage-array-based replication on one or more VMware file systems. This does not completely eliminate replication of unneeded data, but minimizes the storage overhead. The storage overhead can be eliminated by the use of RDMs on virtual machines that require storage-array-based replication.

VMware ESX/ESXi hosts allow creation of VMFS on partitions of the physical devices. It is possible to create up to 15 partitions on each physical device and a separate VMFS on each partition. Furthermore, since VMFS supports spanning of the file systems across partitions, it is possible, for example, to create a VMware file system with part of the file system on partition 1 of a disk and partition 10 of another. Such designs complicate the management of the environment. If the use of EMC SnapView technology with VMFS is desired, EMC recommends creating only one partition per physical disk.

Each software product mentioned in the previous paragraphs has different performance and availability characteristics. A thorough understanding of the options is important to deploy the optimal replication solution. The following sections present procedures to clone VMware virtual machines using different SnapView product sets. Advantages and disadvantages of each solution are presented to help the reader select an appropriate product for their environment.

Copying virtual machines after shutdown

Ideally, virtual machines should be shut down before the metadata and virtual disks associated with the virtual machines are copied. Copying virtual machines after shutdown ensures a clean copy of the data that can be used for backup or quick start up of the cloned virtual machine.

Using SnapView clones with ESX servers

EMC SnapView Clone provides the flexibility of copying any source device on the CLARiiON storage array to another device of equal size in the storage array. When the SnapView Clone relationship between the source and target devices is created, the target devices are presented as not ready (NR) to any host that is accessing the volumes. Therefore, EMC recommends unmounting targets of the SnapView Clone operation from the hosts accessing them before performing a synchronize operation. Unfortunately, VMware ESX/ESXi hosts do not provide a mechanism to unmount devices. However, since VMware ESX/ESXi hosts do not cache any information for mounted VMFS volumes, the presence of VMware file system in the I/O path is unimportant. The synchronize operation makes the VMware file system on the target devices of the clone operation unavailable to the VMware ESX/ESXi host cluster. Therefore, the recommendation of unmounting the target devices in the clone operation applies directly to the virtual machines that are impacted by the absence of the VMware file system on the target devices.

Note: When the synchronize operation is performed, the VMkernel loses access to the target devices involved in the cloning operation. The VMkernel, in this case, may log error messages to indicate the change in the status of the target devices.

A number of organizations use a cloned virtual machine image for different purposes. For example, a cloned virtual machine may be configured for reporting activities during the day and for backups in the night. When the SnapView synchronize operation is in progress, the target device of the operation is unavailable to the VMware ESX/ESXi hosts, and any virtual machine using that VMFS volume cannot be powered on. This restriction must be considered when designing virtual infrastructure that re-provision cloned virtual machines. The same restriction also applies to virtual machines on VMware ESX/ESXi hosts version 4.0 and 3.x using RDMs.

Virtual machines running on VMware ESX/ESXi hosts version 4.0, 3.x and VMware ESXi can be powered on with RDMs that map to devices that are in a "not ready" state. However, the VMware file system holding the configuration file, the metadata information about the virtual machine and the virtual disk mapping files has to be available for the power-on operation.

Copying virtual machines on VMware file systems using SnapView clones

Clones and their associated source devices are grouped in clone groups. Clone groups are created using the source devices. The target devices, such as clones, become members of the clone group as they are added to the source devices. Navisphere Manager or CLI can be used to create clone groups and manage the copying process.

The following explains the steps required to clone a group of virtual machines utilizing EMC SnapView Clone technology:

1. The LUN number of the CLARiiON volumes used by the VMware file system needs to be identified. Installing the Navisphere Agent or Navisphere CLI on the ESX Server console, or using the VM-aware Navisphere feature available with Release 29, provides the mapping information on the Navisphere Manager console. An example of this is shown in Figure 71 on page 180.

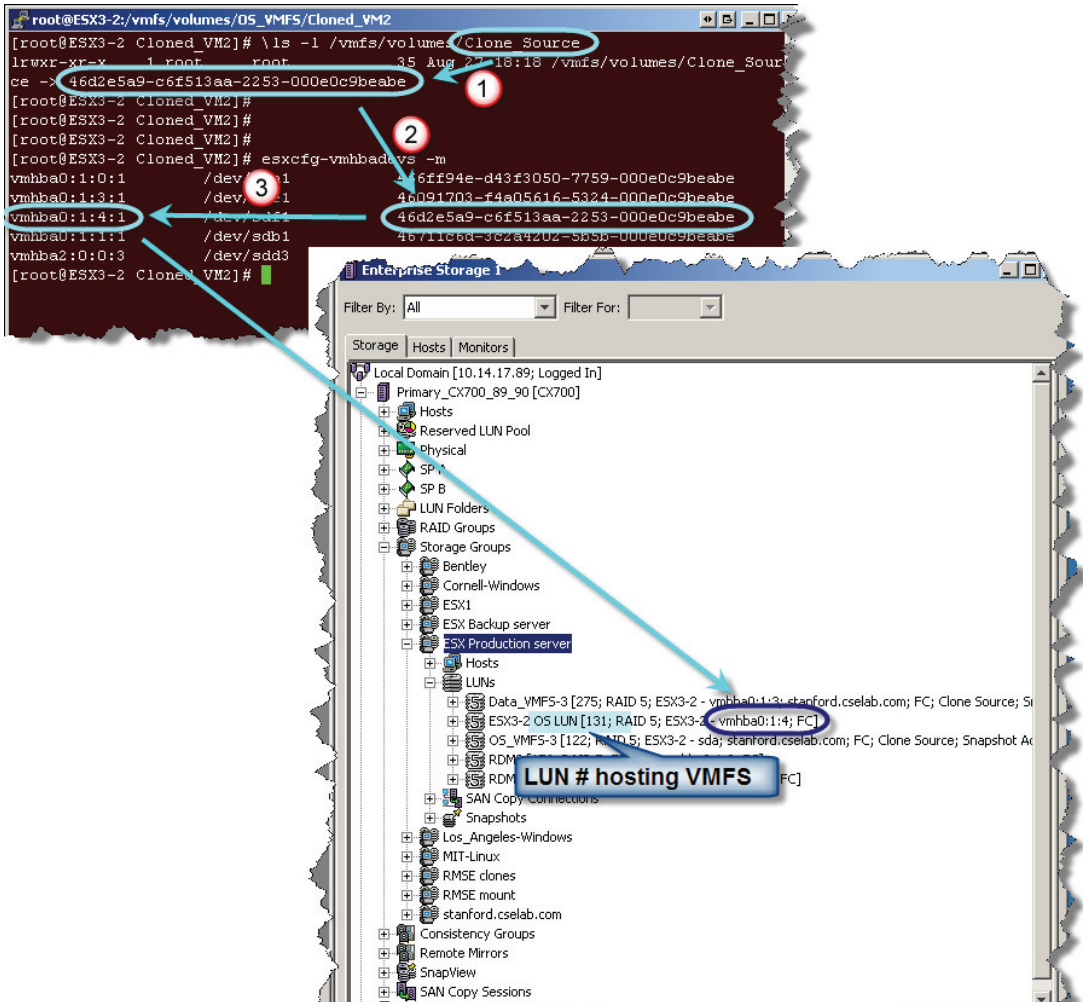


Figure 71 Determining the CLARiiON LUN hosting a VMware file system

A clone group containing the members of the VMware file system should be created. The creation of the appropriate clone group is shown in Figure 72 on page 181 and Figure 73 on page 182.

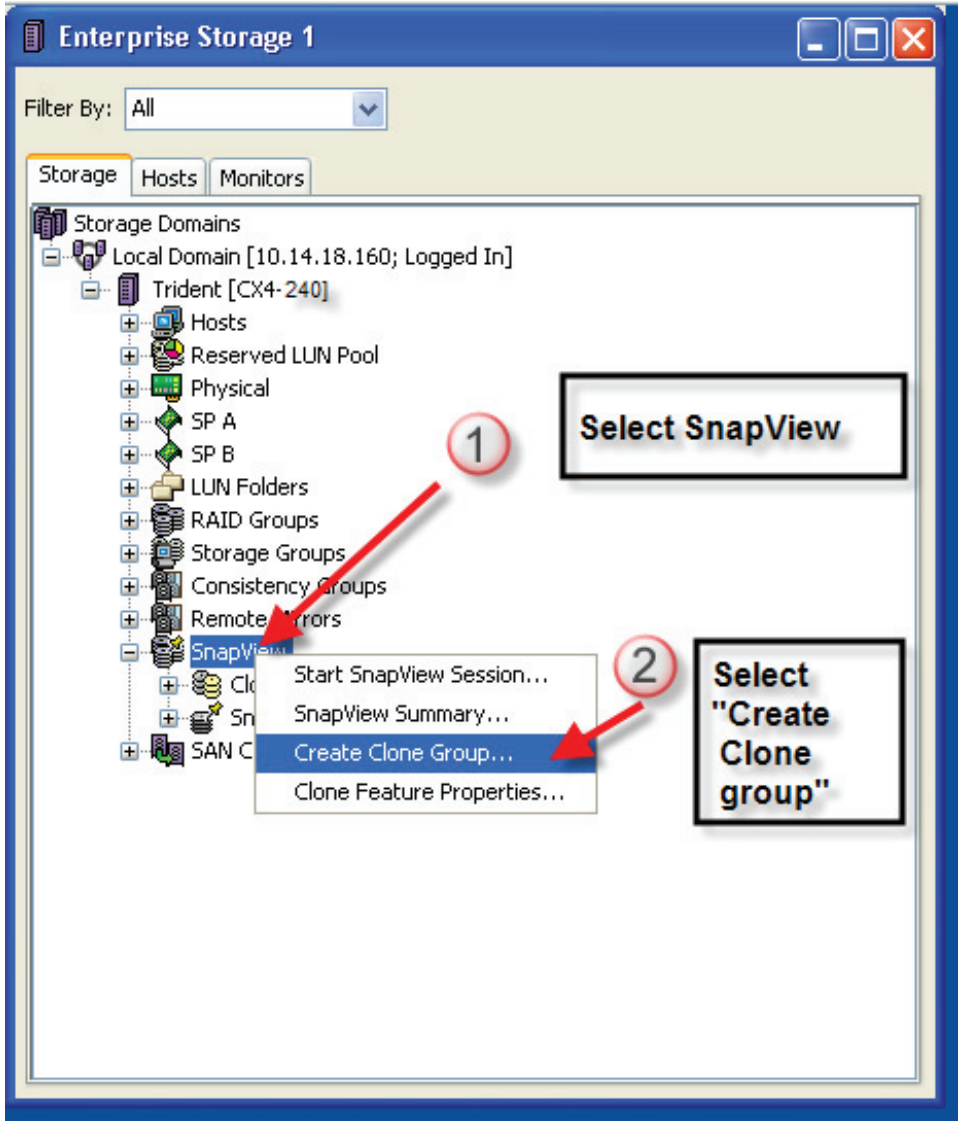


Figure 72 Creating a clone group using Navisphere Manager

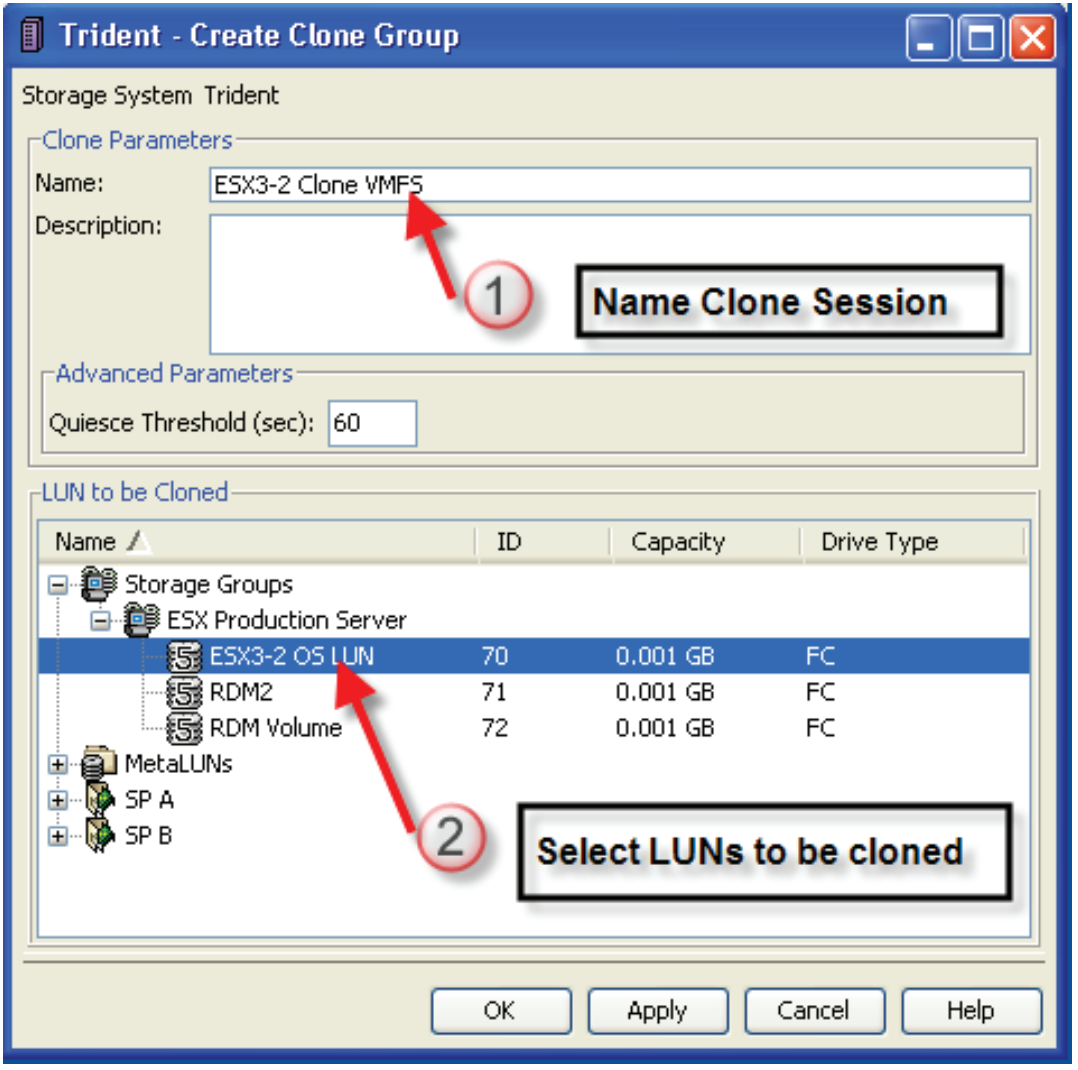


Figure 73 Naming a clone group and selecting advanced parameters

2. The LUNs that would hold the copy of the source data need to be added or associated with the clone group using the Add Clone property dialog box within Navisphere Manager as shown in Figure 74 on page 183.

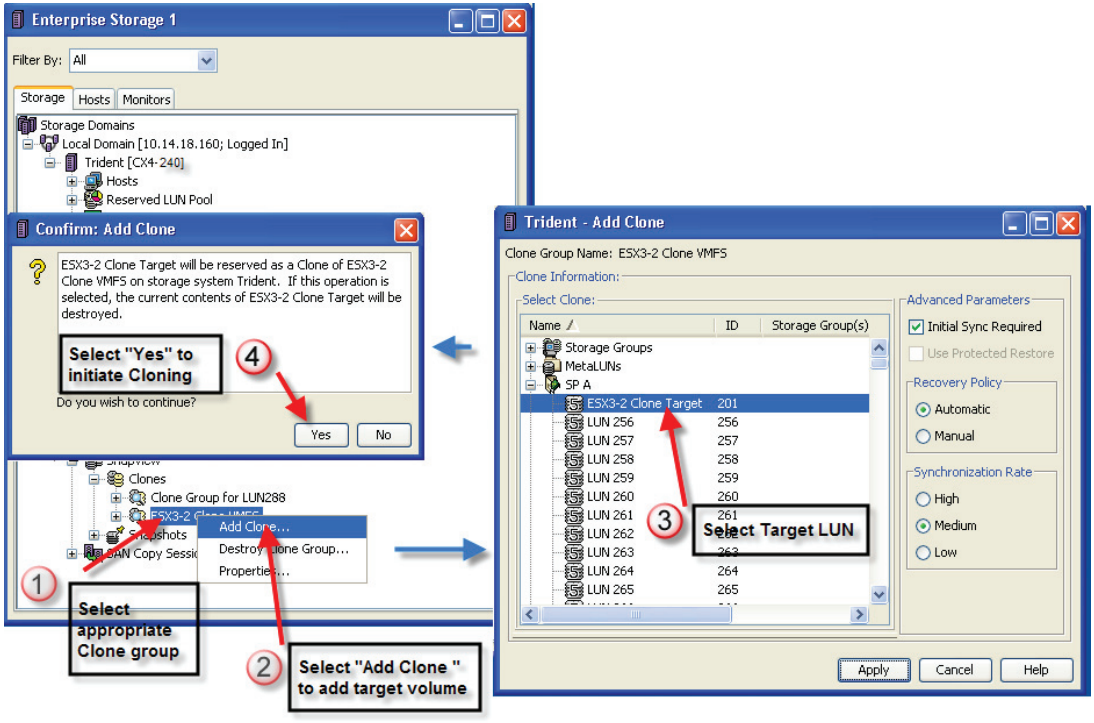


Figure 74 Adding clone target volumes to a clone group

3. The Add Clone wizard shown in Figure 74 on page 183 responds with a question asking the user to verify the clone target. If the default parameters are in effect, selecting Yes in response initiates the copy process. The copy process from the source LUN to target LUN can be set to either automatic or manual. The default is automatic.

   Subsequent requests for synchronization between source and target LUNs are performed incrementally.

4. After the target LUNs in the clone group are synchronized to the source LUNs, the virtual machines can be shut down to make a "cold" copy of the virtual machine data. The virtual machines can

be either shut down using the vCenter client or the service console. The process to shut down the virtual machines using vCenter client is shown in Figure 75 on page 184. The command line utility, `vmware-cmd`, may be the most appropriate tool if a number of virtual machines need to be powered down before the clones are fractured from the source LUNs.

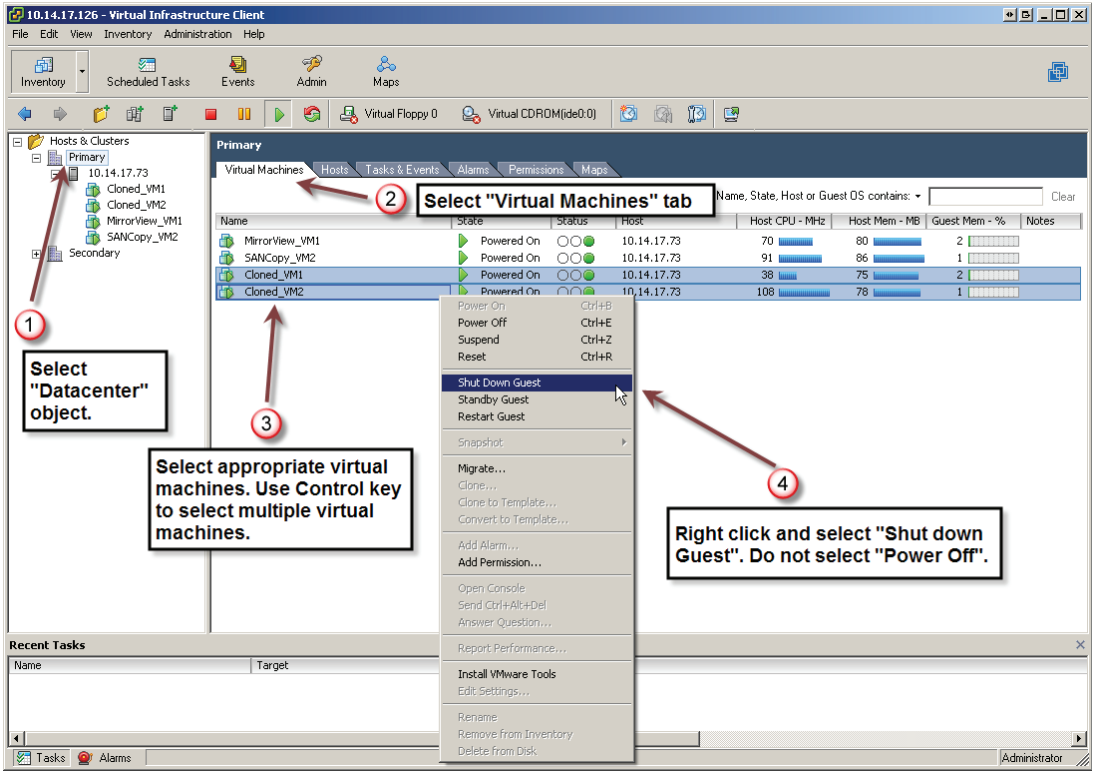


Figure 75 Shutting down virtual machines to create a “cold” copy of data

5. After all of the virtual machines accessing the VMware file system have been shut down, the cloned LUNs can be fractured from the source LUNs as shown in Figure 76 on page 185.

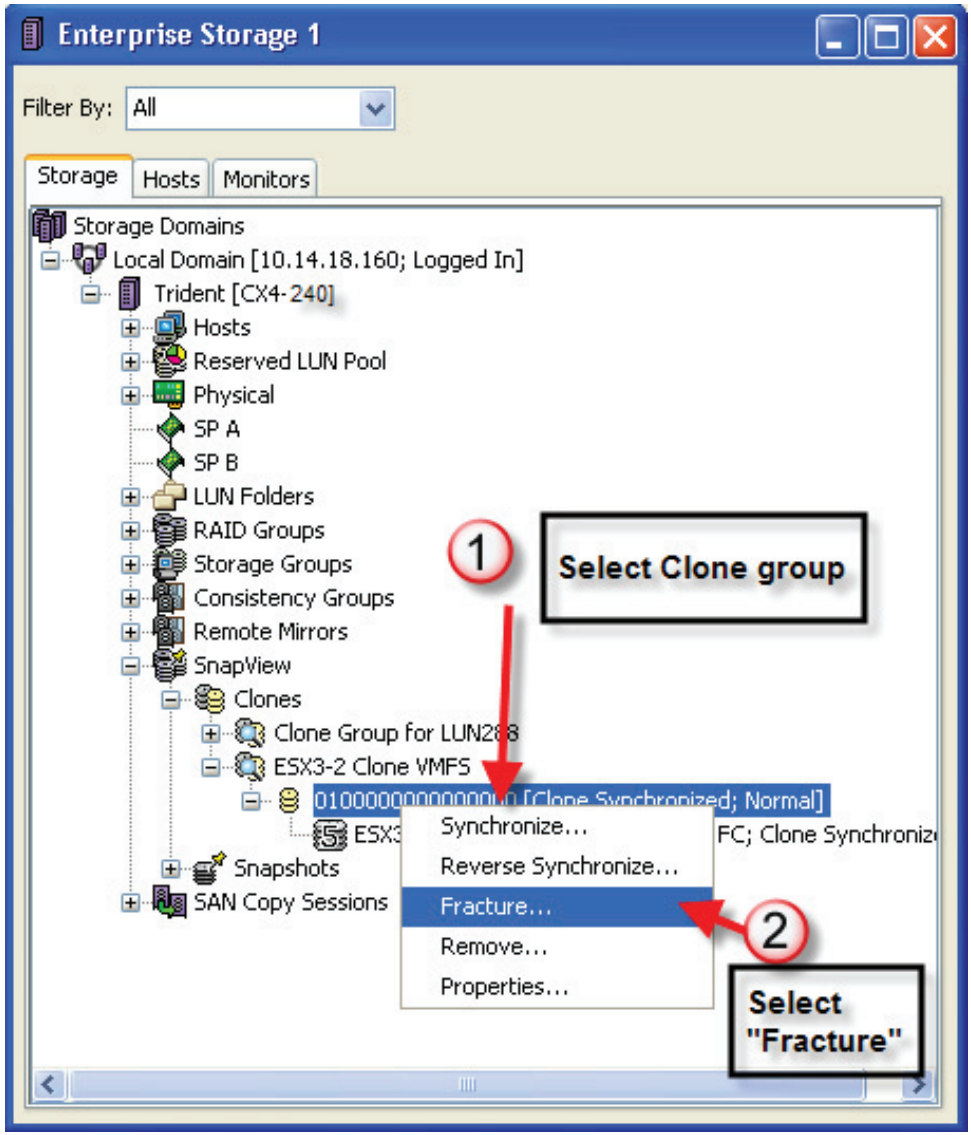


Figure 76 Fracturing a target LUN from a source LUN

6. The virtual machines accessing the VMware file system on the source devices can be powered on and made available to the users. Similar to the shutdown process, graphical user interface (MUI or vCenter client) or command line utility can be used for this.

Figure 77 on page 186 pictorially depicts the steps discussed.

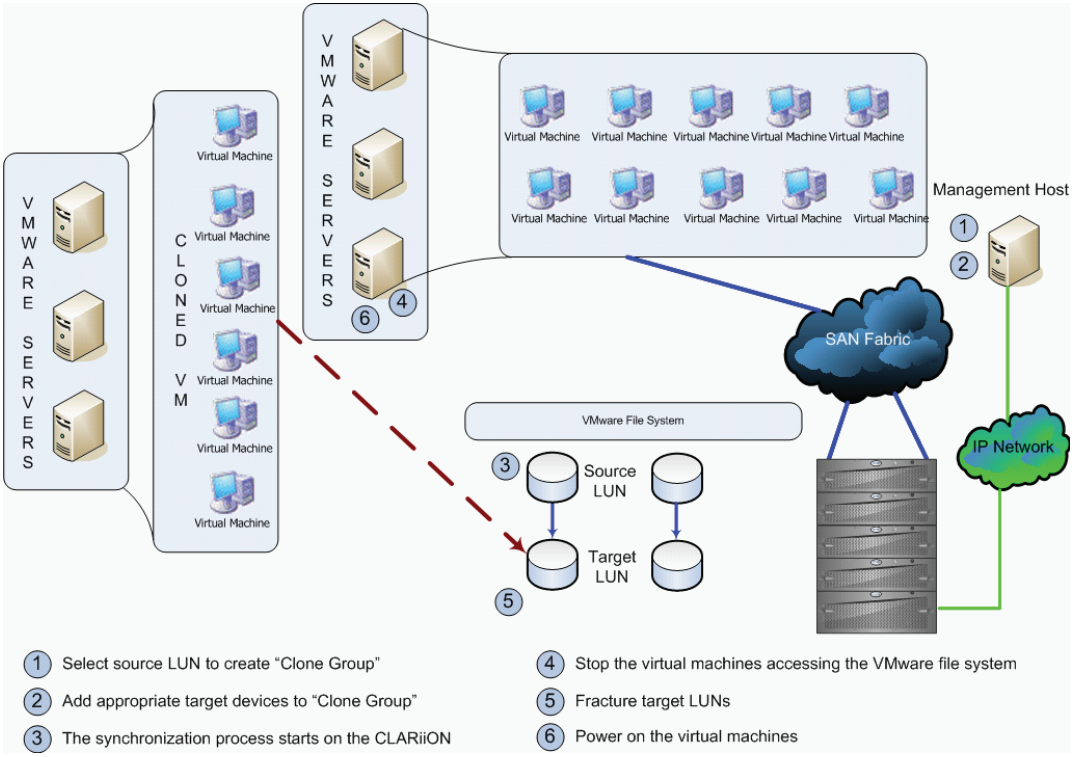


Figure 77 Copying shutdown virtual machines using EMC SnapView clones

Copying virtual machines with RDMs using SnapView clones

The first step in using SnapView Clone technology to copy virtual machines that access disks as Raw Device Mapping (RDM) is identifying the CLARiiON LUN numbers associated with the virtual machines. This can be accomplished by using the SCSI INQUIRY utility, inq, provided by EMC or by running the Navisphere Agent/CLI in the virtual machine. Figure 78 on page 187 depicts the process of using inq to determine the CLARiiON LUN numbers allocated as RDMs to a virtual machine. Starting with Release 29 of FLARE, the RDM volumes

will be displayed with Navisphere using the VM-aware Navisphere feature.

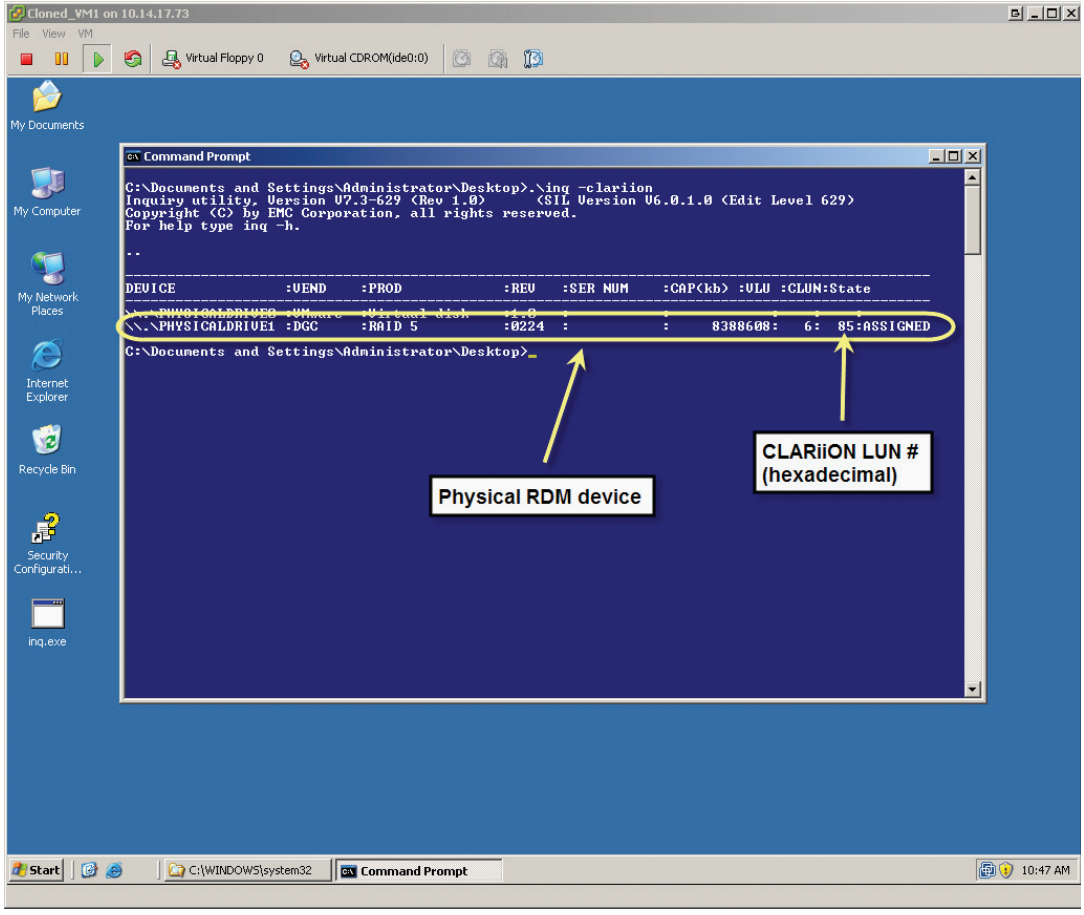


Figure 78 Using INQ to determine the CLARiiON LUN number presented as RDM to a virtual machine

After the CLARiiON LUN numbers used by the virtual machines are identified, the process to create a copy of a virtual machine that is utilizing Raw Device Mapping (RDM) is identical to the one presented in "Copying virtual machines on VMware file systems using SnapView clones," on page 179.

Using SnapView snapshots with ESX servers

SnapView snapshots enable users to create what appears to be a complete copy of their data while consuming only a fraction of the disk space required by the original copy. This is achieved by using a snapshot device as the target of the process. A snapshot device is a construct inside the CLARiiON storage array with minimal physical storage associated with it. Therefore, the snapshot devices are normally presented in a not-ready state to any host accessing it.

When a SnapView session is created, a point-in-time copy of source LUN is generated. A reserved LUN from the reserved LUN pool is assigned to the source LUN. The point-in-time copy (SnapView session) is accessed by using a snapshot device associated with the source LUN. After the snapshot device is created, it is presented in a not-ready (NR) to any host that is accessing the volumes. When the snapshot is activated to a particular session associated with the same source LUN, the point-in-time copy can be accessed through the snapshot device.

Data changed by either the hosts accessing the source device or the snapshot device is stored in the reserved LUN pool area. The amount of data saved depends on the write activity on the source LUN, the snapshot device, and the duration for which the SnapView session remains active.

Copying virtual machines on VMware file systems using SnapView snapshots

SnapView snapshots are managed using Navisphere Manager or CLI. Figure 79 on page 189 depicts the necessary steps to make a copy of powered off virtual machines using the SnapView snapshot technology.

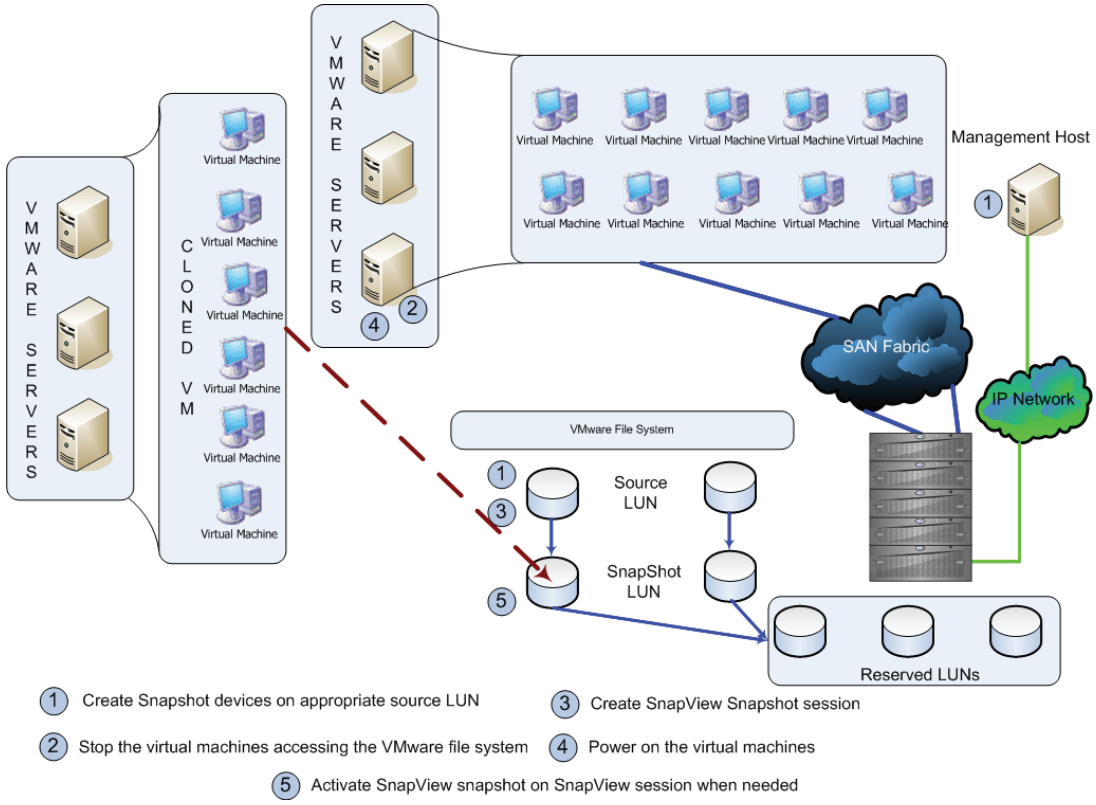


Figure 79 Copying inactive VMware file systems with SnapView snapshots

1. The source devices to be snapped first need to be identified. The process shown in Figure 71 on page 180 can be used for this.

2. A SnapView snapshot of the source devices can be created using Navisphere Manager. An example of this is shown in Figure 80 on page 190.

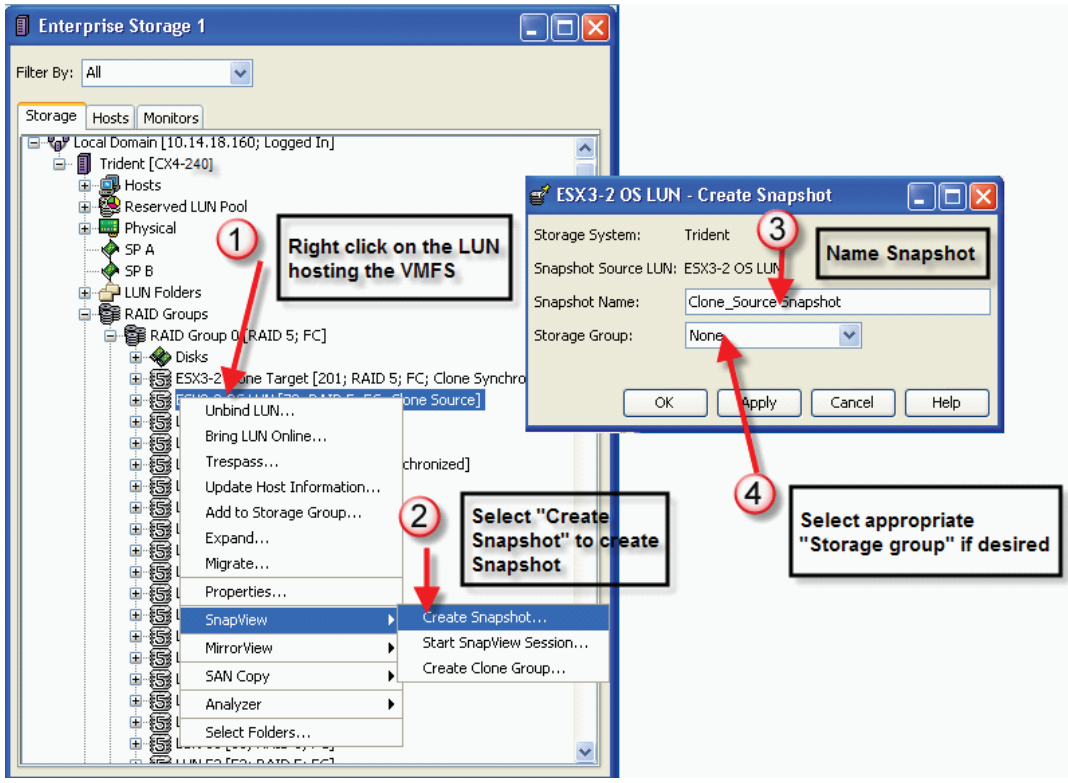


Figure 80 Creating SnapView snapshot devices for source devices

3. After the SnapView snapshot device has been created, the virtual machines accessing the source VMware file system must be shut down to make a cold copy of the VMware file system. The VMware infrastructure tools (Virtual Infrastructure client or `vmware-cmd`) can be used to perform this function. Figure 75 on page 184 shows an example of using a Virtual Infrastructure client to shut down virtual machines.

4. With the virtual machines in a powered-off state, a SnapView session must be started on the source device using either Navisphere Manager or CLI (see Figure 81 on page 191). The SnapView session can be accessed by activating the session on the SnapView snapshot device created in step 2.

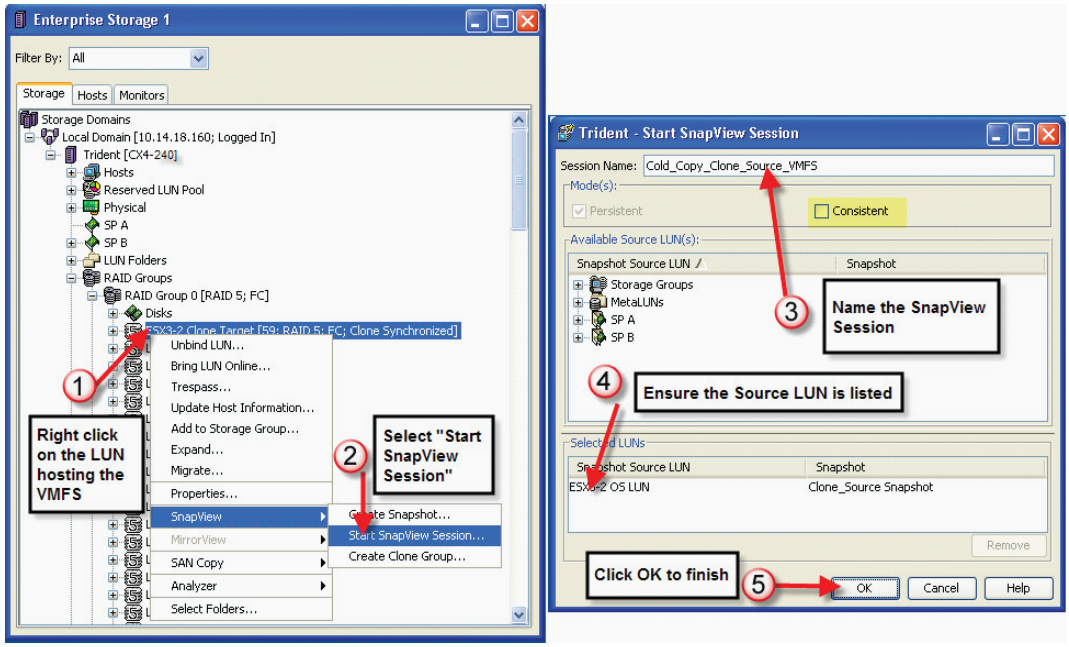


Figure 81 Creating a SnapView session to create a cold copy of a VMware file system

5. The virtual machines using the source VMware file systems can now be powered on. The same VMware infrastructure tool that was utilized to shut down the virtual machines can be used to power them on.

Copying virtual machines with RDMs using SnapView snapshots

The process for using SnapView snapshots technology to copy virtual machines that access disks as Raw Device Mapping (RDM) is no different from that discussed for SnapView clones. The first step in using SnapView snapshot technology to copy virtual machine data that access disks as Raw Device Mapping (RDM) is identifying the CLARiiON LUN numbers associated with the virtual machines. This

can be accomplished by using the SCSI INQUIRY utility, `inq`, provided by EMC or by using the VM-aware Navisphere feature or by running the Navisphere Agent or CLI in the virtual machine.

After the CLARiiON devices used by the virtual machines are identified, the process to create a copy of a virtual machine that is utilizing Raw Device Mapping (RDM) is identical to the one presented in the section "Copying virtual machines with RDMs using SnapView clones," on page 186.

Using EMC to copy running virtual machines

"Copying virtual machines after shutdown," on page 178 discussed use of the SnapView family of products to clone virtual machines that have been shut down. Although this is the ideal way to obtain a copy of the data, it is impractical in most production environments. For these environments, EMC consistency technology can be leveraged to create a copy of the virtual machines while it is servicing applications and users. Using the consistency technology enables a group of active virtual machines on different LUNs to be copied in an instant. The image created in this way is a dependent-write consistent data state and can be utilized as a restartable copy of the virtual machine.

Virtual machines running modern operating systems, such as Microsoft Windows and database management systems, enforce the principle of dependent-write I/O. That is, no dependent write is issued until the predecessor write it is dependent on has completed. For example, Microsoft Windows does not update the contents of a file on a NT file system (NTFS) until an appropriate entry in the file system journal is made. This technique enables the operating system to quickly bring NTFS to a consistent state when recovering from an unplanned outage such as power failure.

Note: Microsoft Window NT file system is a journal file system and not a logged file system. When recovering from an unplanned outage the contents of the journal may not be sufficient to recover the file system. A full check of the file system using `chkdsk` is needed for these situations.

Using the EMC consistency technology option during the virtual machine, copying process also creates a copy with a dependent-write consistent data state. "EMC Foundation Products," on page 47, has a detailed discussion on EMC consistency technology. The following sections describe how to copy a group of live virtual machines using SnapView and EMC consistency technology.

Using SnapView clones with ESX servers

EMC SnapView Clone copies data using the resources available on the CLARiiON storage array. A relationship between the source volume and the clone is created by using clone groups. The clone added to the clone group must have the same configuration and size as the source volume. However, the underlying protection mechanism (RAID 1, RAID 3, RAID 5, or RAID 6) of the source and target volume can be different.

When the clone devices are synchronizing or in a synchronized state with the source volumes, the clone devices are presented in a not-ready (NR) to any host that is accessing the volumes. Therefore, in a VMware environment is important to ensure the virtual machines accessing the copy of the production data are in a powered-off state before the clone devices are synchronized with the source devices. The clone devices can be accessed by VMware ESX/ESXi hosts as soon as the clones are fractured from the source devices.

Using SnapView clones to copy virtual machines running on VMware file systems

SnapView clones are managed using Navisphere Manager or CLI. If multiple devices are being cloned, a SnapView Clone for each device should be created separately. Figure 82 on page 194 depicts the steps necessary to make a copy of the group of running virtual machines using EMC SnapView clones and EMC consistency technology.

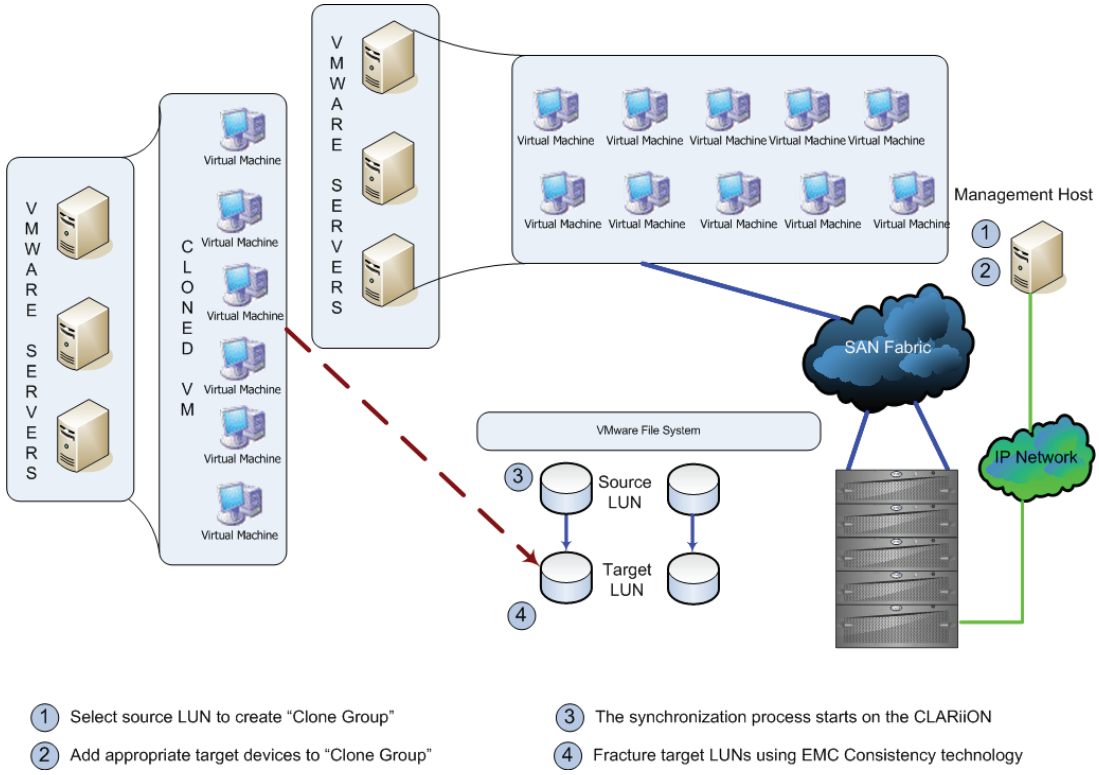


Figure 82 Copying running virtual machine data using SnapView Clone technology

1. The LUN number of the CLARiiON volumes used by the VMware file system needs to be identified. The members of the VMware file system, which include the different virtual machines, need to be identified. The process described in step 1 of "Using SnapView clones with ESX servers," on page 178 can be used to determine the devices that need to be cloned.
2. A clone group needs to be defined for each source volume that holds the data that needs to be copied. Figure 72 on page 181 depicts the process to create a clone group using Navisphere Manager.
3. The target devices that will hold a copy of the source devices needs to be added to each clone group created in the previous step. The addition of the target devices, by default, automatically starts the synchronization process.
4. Once the clone volumes are synchronized with the source volume, they can now be fractured from the source volume when a point in copy is desired.

If multiple CLARiiON LUNs are involved in the cloning process, as seen in Figure 83 on page 196, a consistent fracture operation can be performed by selecting multiple clone sessions using the Control key in Navisphere Manager. When using Navisphere CLI, the consistent switch can be specified to perform a consistent fracture.

Any VMware ESX/ESXi hosts with access to the clone volumes is presented with a consistent point-in-time read-write copy of the source volumes at the moment of fracture.

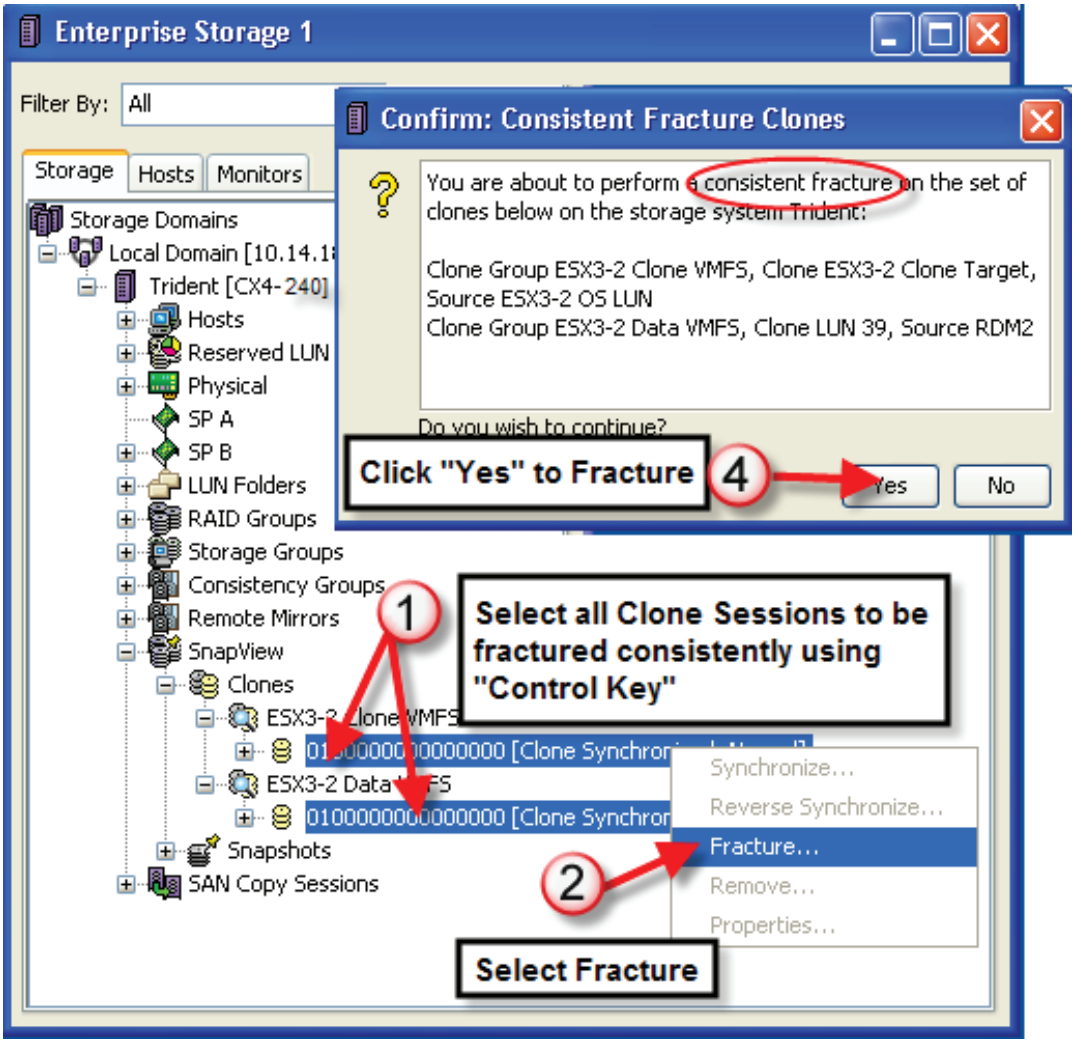


Figure 83 Using Navisphere Manager to consistently fracture SnapView clone groups

Copying running virtual machines with RDMs using SnapView clones

The first step in copying running virtual machines that access disks as Raw Device Mapping (RDM) is identifying the CLARiiON LUN numbers associated with the virtual machine. This can be accomplished by using the SCSI INQUIRY utility, `inq`, provided by EMC.using the VM-aware Navisphere feature available with R29, or by running the Navisphere Agent/CLI in the virtual machine.

After the CLARiiON LUNs used by the virtual machines are identified, the process to create a copy of a virtual machine that is utilizing raw disks or Raw Device Mapping (RDM) is identical to the one presented in Section Using SnapView clones to copy virtual machines running on VMware file systems .

Using SnapView snapshots with ESX servers

SnapView snapshots enable users to create a complete copy of their data while consuming only a fraction of the disk space required by the original copy. This is achieved by use of snapshot device as the target of the process. A snapshot device is a construct inside the CLARiiON storage array with minimal physical storage associated with it. Therefore, the “snapshot” devices are normally presented in a not-ready state to any host accessing it.

When a SnapView session is created, a point-in-time copy of source LUN is generated. A reserved LUN from the reserved LUN pool is assigned to the source LUN. The point-in-time copy (SnapView session) is accessed by using a snapshot device associated with the source LUN. After the snapshot device is created, it is presented in a not-ready (NR) to any host that is accessing the volumes. When the snapshot is activated to a particular session associated with the same source LUN, the point-in-time copy can be accessed through the snapshot device.

Data changed by either the hosts accessing the source device or the snapshot device is stored in the reserved LUN pool area. The amount of data saved depends on the write activity on the source LUN, the snapshot device, and the duration for which the SnapView session remains active.

Copying running virtual machines on VMware file systems using SnapView snapshots

SnapView snapshots are managed using Navisphere Manager or CLI. Figure 84 on page 198 depicts the steps necessary to make a copy of group of running virtual machines using EMC SnapView clones and EMC consistency technology.

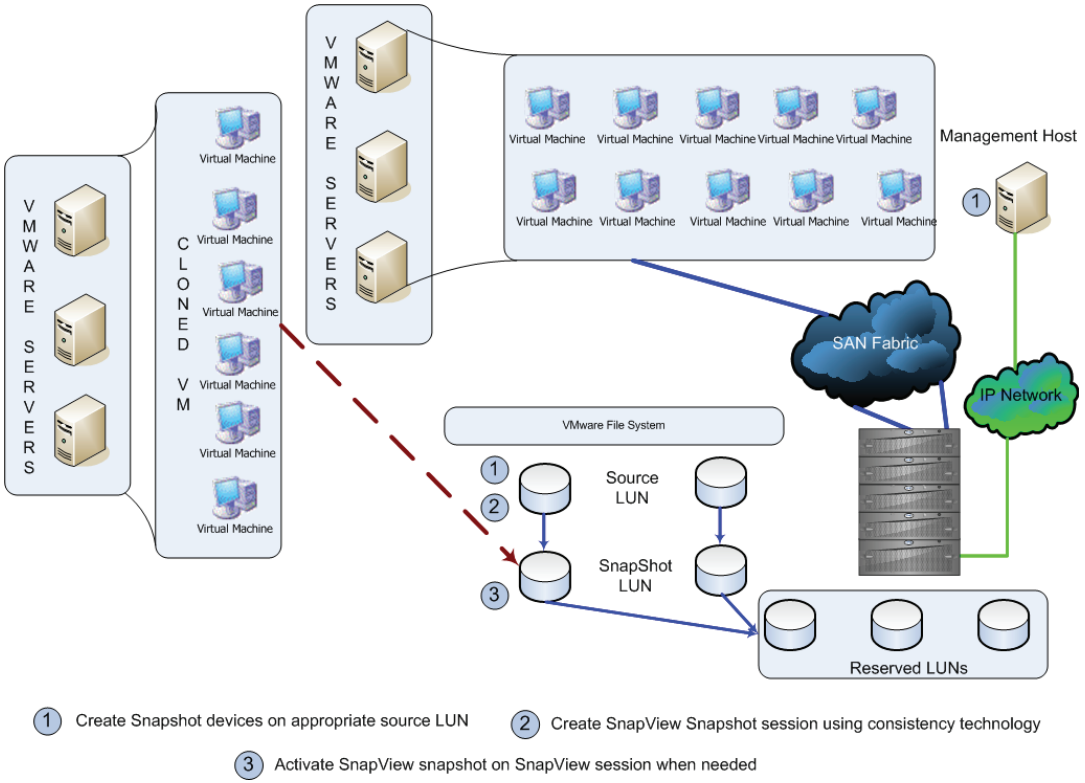


Figure 84 Copying active VMware file systems with SnapView snapshots

1. The source devices to be snapped need to be identified. The process shown in Figure 71 on page 180 can be used for this.
2. A SnapView snapshot of the source devices can be created using Navisphere Manager. An example of this is shown in Figure 80 on page 190. When desired, a SnapView session must be started on the source device using either Navisphere Manager or CLI. To ensure dependent-write consistent image of the virtual infrastructure data is capture, select the Consistent flag when creating the session (see

the area highlighted in yellow in Figure 81 on page 191). The SnapView session can be accessed by activating the session on the SnapView snapshot device created in step 2.

Copying virtual machines with RDMs using SnapView snapshots

The process for using SnapView snapshots technology to copy running virtual machines that access disks as Raw Device Mapping (RDM) is no different from that discussed for SnapView clones in "Using SnapView clones with ESX servers," on page 178. This can be accomplished by using the SCSI INQUIRY utility, `inq`, provided by EMC.using the VM-aware Navisphere feature available with R29, or by running the Navisphere Agent/CLI in the virtual machine.

After the CLARiiON LUNs used by the virtual machines are identified, the process to create a copy of a virtual machine that is utilizing raw disks or Raw Device Mapping (RDM) is identical to the one presented in "Copying virtual machines on VMware file systems using SnapView clones," on page 179.

Transitioning disk copies to cloned virtual machines

This section discusses how to use the copy of the data to create cloned virtual machines. The cloned virtual machines can be deployed to support ancillary business processes such as development and testing. The methodology deployed in creating the copy of the data influences the type of supporting business operations that it can support. VMware ESX/ESXi virtualizes IT assets into a flexible, cost-effective pool of compute, storage, and networking resources. These resources can then be mapped to specific business needs by creating virtual machines. VMware ESX/ESXi provides several utilities to manage the environment. This includes utilities to clone, back up, and restore virtual machines. All these utilities use host CPU resources to perform the functions. Furthermore, the utilities cannot operate on data not residing in the VMware infrastructure. discusses the application of the cloned virtual machines for backup and recovery purposes.

Cloning VMs on VMware file systems in VMware Infrastructure 3

VMware ESX version 3 and VMware ESXi3.x assign a unique signature to all VMFS-3 volumes when they are formatted with the VMware file system. Furthermore, if the VMware file system is labeled that information is also stored on the device. The signature is generated using the unique ID (UID) of the device and the LUN number at which the device is present.

Since storage array technologies create exact replicas of the source volumes, all information including the unique signature (and label, if applicable) is replicated. If a copy of a VMFS-3 volume is presented to any VMware ESX version 3 or VMware ESXi host or cluster group, the VMware ESX/ESXi hosts, by default, automatically mask the copy. The device holding the copy is determined by comparing the signature stored on the device with the computed signature. Clones, for example, have a different unique ID from the source device it is associated with it. Therefore, the computed signature for a clone device always differs from the one stored on it. This enables the VMware ESX/ESXi hosts to always identify the copy correctly.

VMware ESX version 3 and VMware ESXiprovide two different mechanisms to access copies of VMFS-3 volumes. The advanced configuration parameters, `LVM.DisallowSnapshotLun` or `LVM.EnableResignature`, control the behavior of the VMkernel when presented with copies of a VMware file system.

- If `LVM.DisallowSnapshotLun` is set to 0, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts that have access to both source and target devices, the parameter has no effect since VMware ESX/ESXi hosts never presents a copy of the data if there are signature conflicts. The default value for this parameter is 1.
- If `LVM.EnableResignature` is set to 1, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include "snap-*x*", where *x* is a hexadecimal number that can range from 0x2 to 0xFFFFFFFF. The default value for this parameter is 0. If this parameter is changed to 1, the advanced parameter, `LVM.DisallowSnapShotLun`, is ignored.

By using the proper combination of the advanced configuration parameter, copies of VMFS-3 can be used to clone source virtual machines. The following paragraphs discuss the process to clone virtual machines in a Virtual Infrastructure 3 environment.

Cloning Virtual Infrastructure 3 virtual machines using LVM.DisallowSnapshotLun

A separate cluster, of VMware ESX version 3.x or VMware ESXi that has no access to the source volumes, is required to access the copy of virtual machine data using the `LVM.DisallowSnapshotLun` parameter. The parameter can be turned off (set to value 0) either from the command line or using the Virtual Infrastructure client. By selecting the Advanced Setting link on the Virtual Infrastructure client (see Figure 85 on page 202), the screen shown in Figure 85 on page 202 appears. The parameter, `LVM.DisallowSnapshotLun`, can be changed as indicated in the figure. The parameter can also be changed using the service console. The process is shown in Figure 86 on page 203. VMware and EMC recommend the use of Virtual Infrastructure client whenever possible.

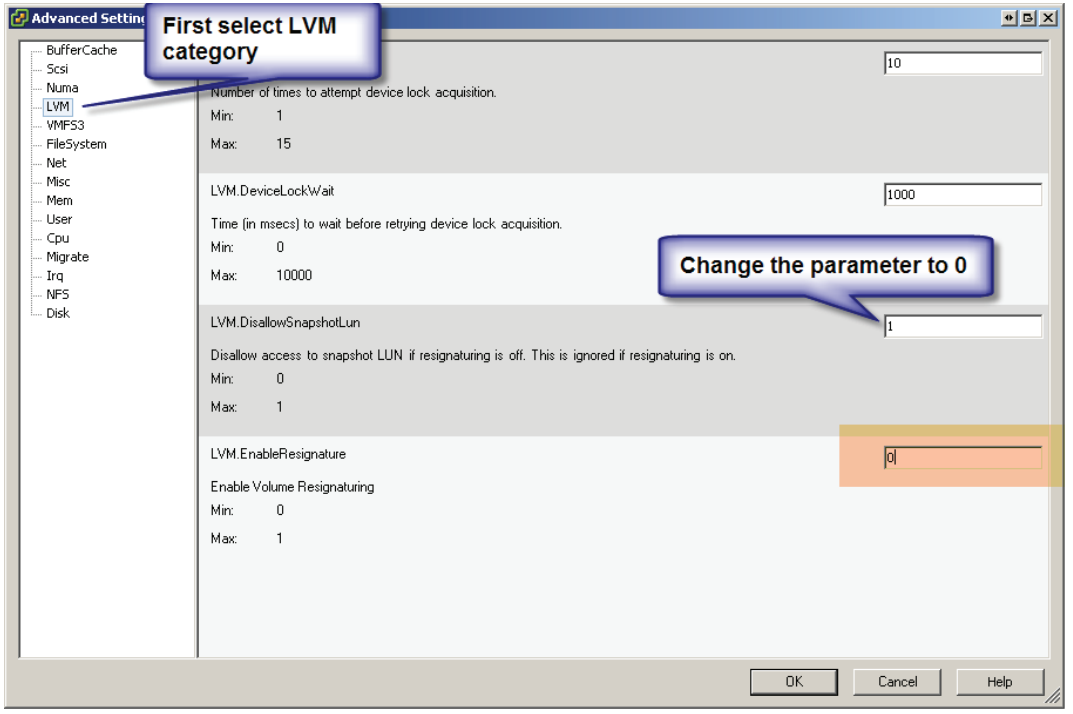


Figure 85 Changing the LVM.DisallowSnapshotLun parameter using the Virtual Infrastructure client

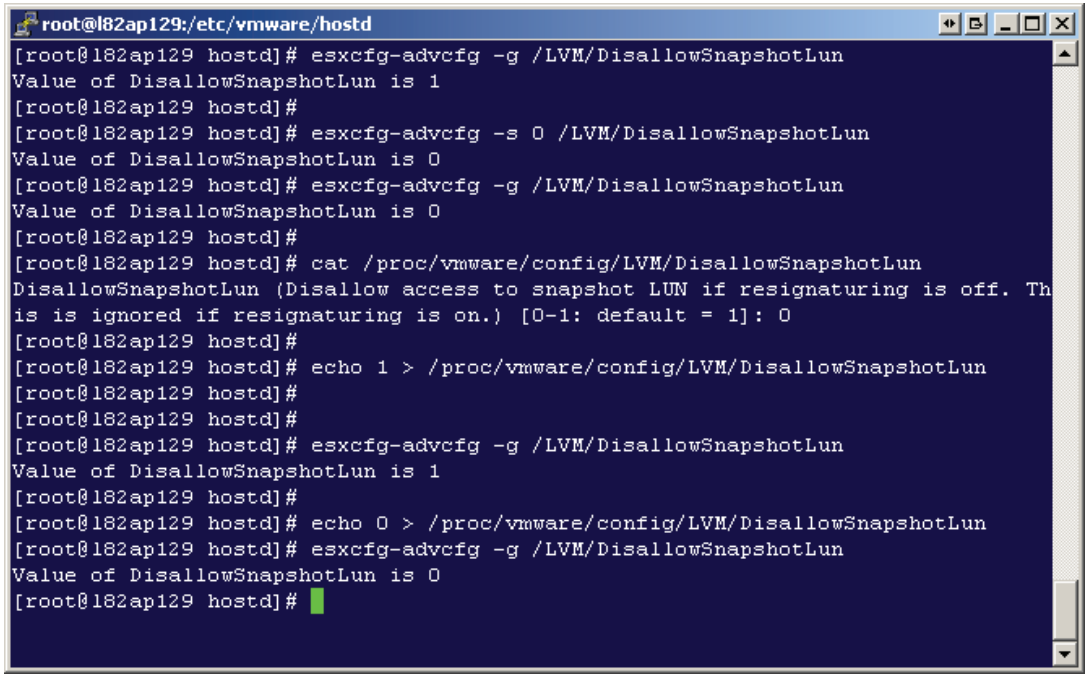


Figure 86 Changing the LVM.DisallowSnapshotLun parameter using the service console

The following steps are required to clone virtual machines using copied data:

1. After changing the `LUN.DisallowSnapshotLun` parameter to 0, the SCSI bus should be rescanned using the service console or the Virtual Infrastructure client. The devices that hold the copy of the VMware file system are displayed on the target VMware ESX/ESXi hosts if the source devices are not present. The process of discovering the target devices is shown in Figure 87 on page 204.
2. Figure 87 on page 204 shows that by setting the parameter, `LVM.DisallowSnapshotLun`, to 0, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts that have access to both source and target devices, the parameter has no effect since VMware ESX/ESXi never present a copy of the data if there are signature conflicts.

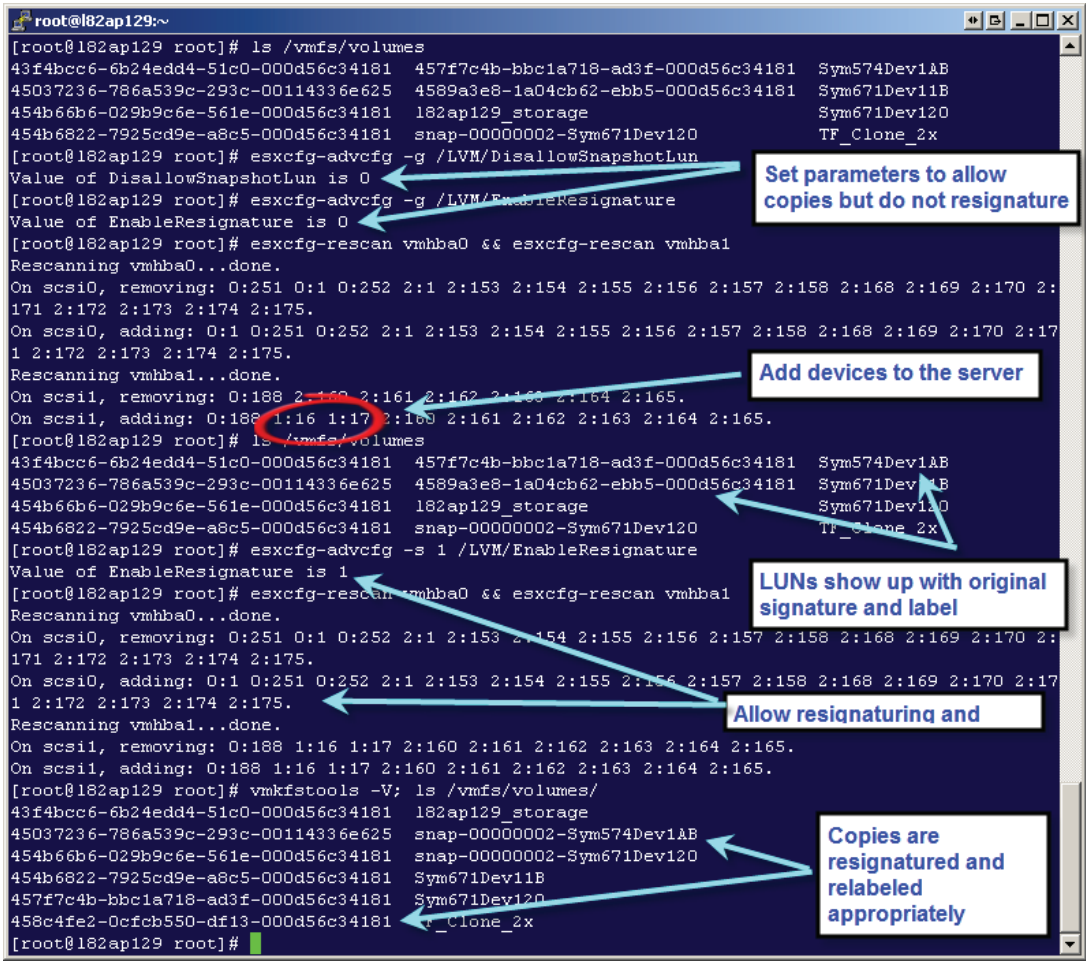


Figure 87 Using LVM.DisallowSnapshotLun parameter to access copies of data

When a virtual machine is created in a Virtual Infrastructure 3 environment, all files related to the virtual machine are stored in a directory on a Virtual Infrastructure 3 datastore. This includes the configuration file and, by default, the virtual disks associated with a virtual machine. Thus, the configuration files automatically get replicated with the virtual disks when storage array based copying technologies are leveraged. Figure 88 on page 206 shows a listing of all configuration files that were copied as part of the replication

from the source device to the target device. The registration of the virtual machines from the target device can be performed using Virtual Infrastructure client or the service console. The registration using the service console utility, `vmware-cmd`, is shown in Figure 88 on page 206. The registration of cloned virtual machines is not required every time the target devices are refreshed with the latest copy of the data from the source device.

Virtual Infrastructure 3 tightly integrates the vCenter infrastructure and the VMware ESX/ESXi hosts version 3 or VMware ESXi. vCenter infrastructure does not allow duplication of objects in a vCenter data center. Therefore, when registering the copy of virtual machines using vCenter client, the cloned virtual machines should be provided with a unique name. The display name for the cloned virtual machines registered using vmware-cmd is automatically renamed by vCenter if the target VMware ESX/ESXi hosts cluster is in the same vCenter data center as the source cluster. The change in the cloned virtual machine name, however, does not impact the operations that can be performed.

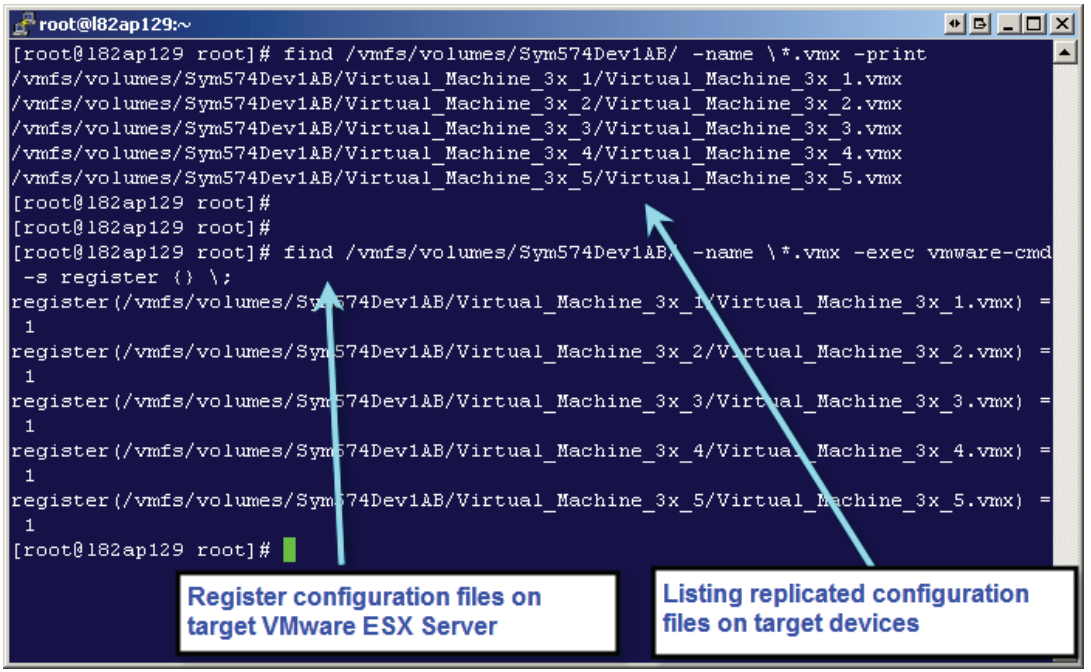


Figure 88 Listing and registering virtual machines on target devices

3. The cloned virtual machines can be started on the target VMware ESX/ESXi hosts without any modification if the following requirements are met:

   - The target VMware ESX/ESXi hosts have the same virtual network switch configuration— that is, the name and number of virtual switches should be duplicated from the source VMware ESX/ESXi host cluster group.
   - All VMware file systems used by the source virtual machines are replicated. Furthermore, the VMFS labels should be unique on the target VMware ESX/ESXi hosts.
   - The minimum memory and processor resource reservation requirements of all cloned virtual machines can be supported on the target VMware ESX/ESXi hosts. For example, if ten source virtual machines, each with a memory resource reservation of

256 MB needs to be cloned and used simultaneously, the target VMware ESX/ESXi host cluster should have at least 2.5 GB of physical RAM allocated to the VMkernel.

- Virtual devices, such as CD-ROM and floppy drives, are attached to physical hardware, or are started in a disconnected state when the virtual machines are powered on.
- The cloned virtual machines can be powered on using the vCenter client or command line utilities, as shown in Figure 89 on page 207.

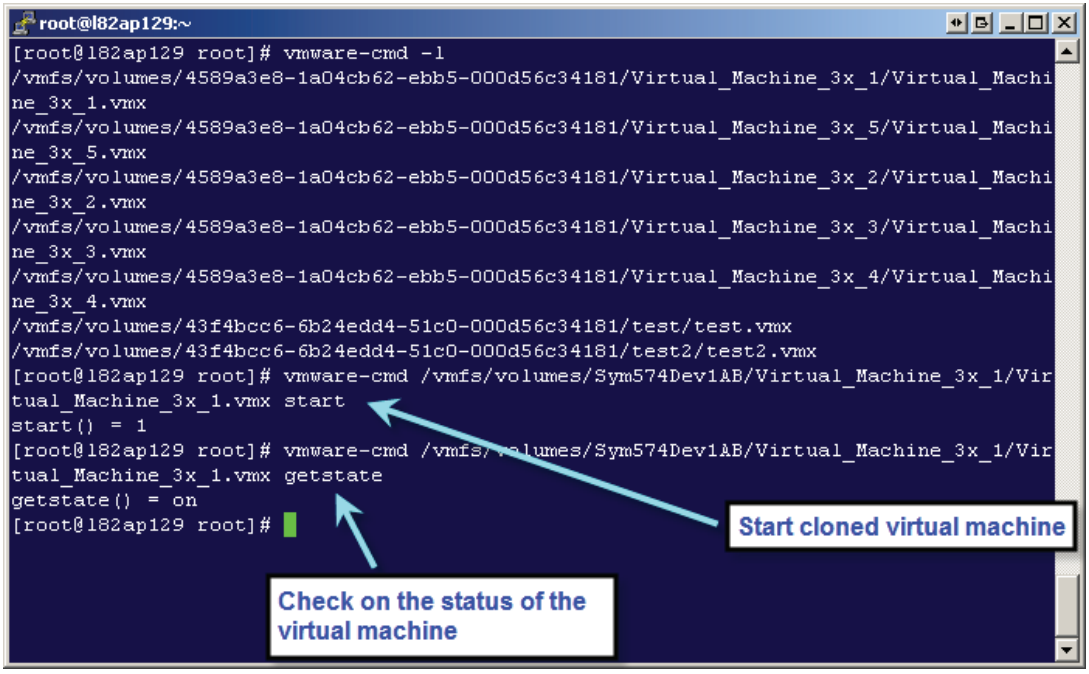


Figure 89 Powering on a cloned virtual machine using service console utility

Cloning Virtual Infrastructure 3 virtual machines using `LVM.EnableResignature`

The parameter, `LVM.EnableResignature`, when enabled allows a VMware ESX 3.x cluster to present both source and target devices simultaneously. The parameter, `LVM.EnableResignature`, can be turned on (set to value 1) either from the command line or using the Virtual Infrastructure client. The parameter is highlighted in orange in Figure 85 on page 202. The parameter can also be changed using the service console, replacing `DisallowSnapshotLun` by

`EnableResignature` (see the process shown in Figure 86 on page 203). VMware and EMC recommend the use of Virtual Infrastructure client whenever possible.

The following steps are required to clone virtual machines using copied data:

1. After changing the `LVM.EnableResignature` parameter to 1, the SCSI bus should be rescanned. This can be done either using the service console or the vCenter client. The devices holding the copy of the VMware file system are resignatured, relabeled, and displayed on the target VMware ESX/ESXi hosts. The process of discovering the target devices is shown in Figure 90 on page 209.
2. When the `LVM.EnableResignature` parameter is set to 1, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include "snap-*x*", where *x* is a hexadecimal number that can range from 0x2 to 0xFFFFFFFF.
3. When a virtual machine is created in a Virtual Infrastructure 3 environment, all files related to the virtual machine are stored in a directory on a Virtual Infrastructure 3 datastore. This includes the configuration file and, by default, the virtual disks associated with a virtual machine. Thus, the configuration files automatically get replicated with the virtual disks when storage array based copying technologies are leveraged. Therefore, unlike a VMware ESX/ESXi host 2.x environment, there is no need to manually copy configuration files.

   Figure 91 on page 211 shows a listing of all configuration files copied as part of the replication from the source device to the target device. The registration of the virtual machines from the target device can be performed using Virtual Infrastructure client or the service console. The registration using the service console utility, `vmware-cmd`, is shown in Figure 91 on page 211.The registration of cloned virtual machines is not required every time the target devices are refreshed with the latest copy of the data from the source device.

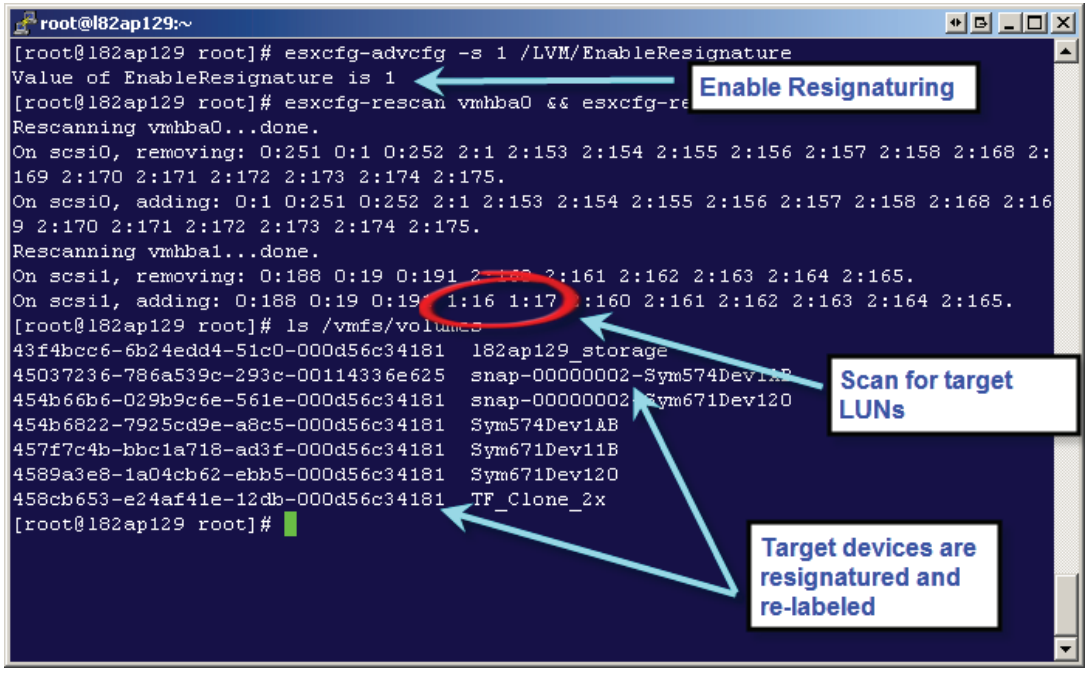


Figure 90 Discovering target devices with LVM.EnableResignature enabled

As discussed in "Cloning Virtual Infrastructure 3 virtual machines using LVM.EnableResignature," on page 207, there is a tight integration between the vCenter infrastructure and VMware ESXversion 3 or VMware ESXi that does not allow duplicate object names. Therefore, when registering the copy of virtual machines using the vCenter client, the cloned virtual machines should be provided with a unique name. The display name for the cloned virtual machines registered using `vmware-cmd` is automatically renamed by the vCenter management server if the target VMware ESX/ESXi hosts cluster is in the same vCenter data center as the source cluster group. The change in the cloned virtual machine name, however, does not impact the operations that can be performed.

4. The cloned virtual machines can be started on the target VMware ESX/ESXi hosts without any modification if the following requirements are met:

- The target VMware ESX/ESXi hosts have the same virtual network switch configuration— that is, the name and number of virtual switches should be duplicated from the source VMware ESX/ESXi hosts cluster group.
- The virtual disks allocated to the source virtual machine are contained in the same VMware file system that contains the configuration file.
- The minimum memory and processor resource reservation requirements of all cloned virtual machines can be supported on the target VMware ESX/ESXi hosts. For example, if ten source virtual machines, each with a memory resource reservation of 256 MB need to be cloned, the target VMware ESX/ESXi host cluster should have at least 2.5 GB of physical RAM allocated to the VMkernel.
- Devices, such as CD-ROM and floppy drives, are attached to physical hardware, or are started in a disconnected state when the virtual machines are powered on.

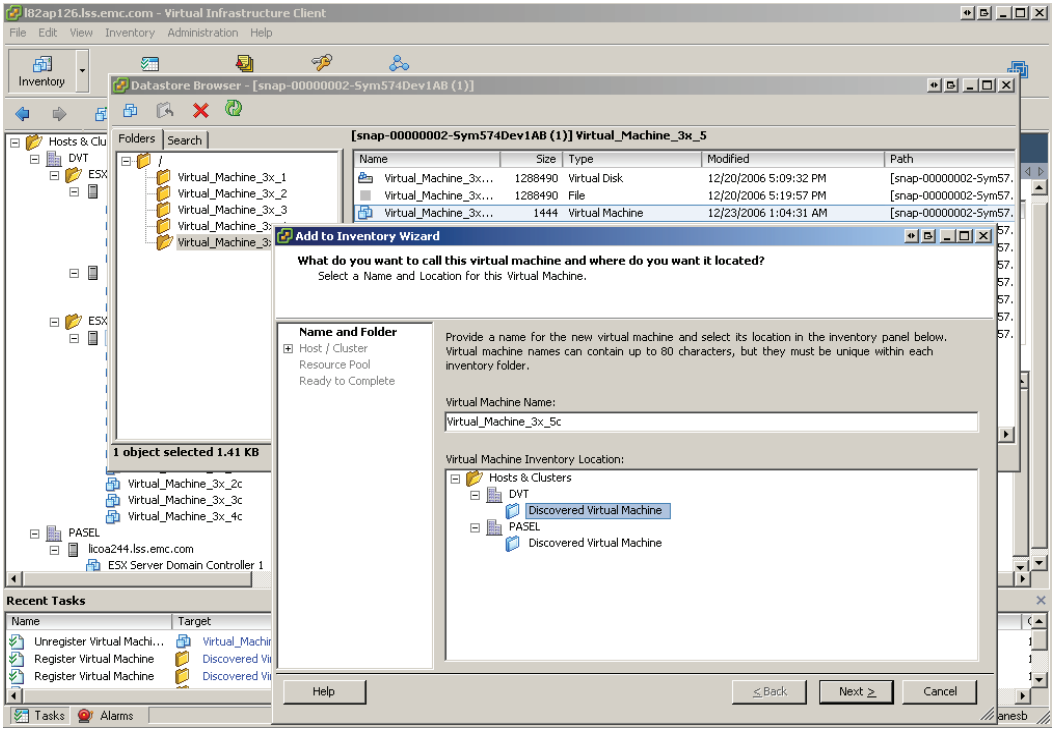


Figure 91 Registering virtual machines using resignatured volumes

5. The cloned virtual machines can be powered on using Virtual Infrastructure client or command line utilities, as shown in Figure 96 on page 219. The use of volume resignaturing introduces complexity into the cloning process. This is particularly true if the source virtual machines are provided with virtual disks from different VMware file systems in line with the best practices recommendations in "Number of VMware file systems (VMFS) in a ESX Server 4 or 3 cluster," on page 155. In this case, changes to the configuration of the cloned virtual machine are required before the machine can be powered on. For this reason, the use of `LVM.EnableResignature` should be limited to environments that cannot provide dedicated VMware ESX/ESXi hosts to run the cloned virtual machines.

Cloning VMs on VMware file systems (VMFS-3) with VMware vSphere 4

VMware ESX 4.0/4i also assigns a unique signature to all VMFS-3 volumes when formatted with the VMware file system. Furthermore, if the VMware file system is labeled, that information is also stored on the device. Since storage array technologies create exact replicas of the source volumes, all information, including the unique signature (and label, if applicable) is replicated.

If a copy of a VMFS-3 volume is presented to any VMware ESX version 4 or VMware ESXi 4.x host or cluster group, the VMware ESX/ESXi hosts, by default, automatically mask the copy. The device holding the copy is determined by comparing the signature stored on the device with the computed signature. Clones, for example, have a different unique ID from the source device it is associated with it. Therefore, the computed signature for a clone device always differs from the one stored on it. This enables the VMware ESX/ESXi hosts to always identify the copy correctly.

VMware ESX version 4 and VMware ESXi 4.x provide selective re-signaturing 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 re-signature the replica (LUN) if needed.

- If the **Keep the existing signature** option is chosen, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts that have access to both source and target devices, the parameter has no effect since VMware ESX/ESXi never presents a copy of the data if there are signature conflicts.
- If the **Assign a new signature** option is chosen, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include "snap-*x*", where *x* is a hexadecimal number that can range from 0x2 to 0xFFFFFFFF.

Cloning Virtual Infrastructure 4 VMs using the "keep the existing signature" option

A separate cluster, of VMware ESX version 4.x or VMware ESXi 4.x that has no access to the source volumes, is required to access the copy of virtual machine data using the `LVM.DisallowSnapshotLun` parameter.

The following steps are required to clone virtual machines using copied data:

1. When a virtual machine is created in a vSsphere 4.0, all files related to the virtual machine are stored in a directory on a Virtual Infrastructure 4 datastore. This includes the configuration file and, by default, the virtual disks associated with a virtual machine. Thus, the configuration files automatically get replicated with the virtual disks when storage array based copying technologies are leveraged. The configuration files for the source virtual machines can be used without any modification for the cloned virtual machines if the following requirements are met:

   - The target VMware ESX/ESXi hosts have the same virtual network switch configuration— that is, the name and number of virtual switches should be duplicated from the source VMware ESX/ESXi cluster.
   - Devices used as raw mapped LUN disks should have the same canonical names on both source and target VMware ESX/ESXi hosts.
   - All VMware file systems used by the source virtual machines are replicated. Furthermore, the VMFS labels should be unique on the target VMware ESX/ESXi hosts.
   - The minimum memory and processor requirements of all cloned virtual machines can be supported on the target VMware ESX/ESXi hosts. For example, if 10 source virtual machines, each with a minimum memory allocation of 256 MB needs to be cloned and used simultaneously, the target VMware ESX/ESXi host cluster should have at least 2.5 GB of physical RAM allocated to the VMkernel.
   - Virtual devices such as CD-ROM and floppy drives are attached to physical hardware, or are started in a disconnected state when the virtual machines are powered on.

2. The copy of the virtual machine data created using the process listed in "Copying virtual machines after shutdown," on page 178 or "Using EMC to copy running virtual machines ," on page 193 should be presented to the target VMware ESX/ESXi host. After the clones are fractured from the standard devices (or the snapshot devices have been activated on a SnapView session), a rescan of the SCSI bus should be performed. The vCenter Client or the command line utility, esxcfg-`rescan.sh`, can be used for this purpose. The devices that hold the copy of the VMware file system are displayed on the target VMware ESX/ESXi host. At this point, the user has the option to select "Keep the existing signature" for that individual

LUN using the "Add Storage" wizard available with vCenter client. The step for discovering the target LUNs is shown in Figure 92 on page 214.

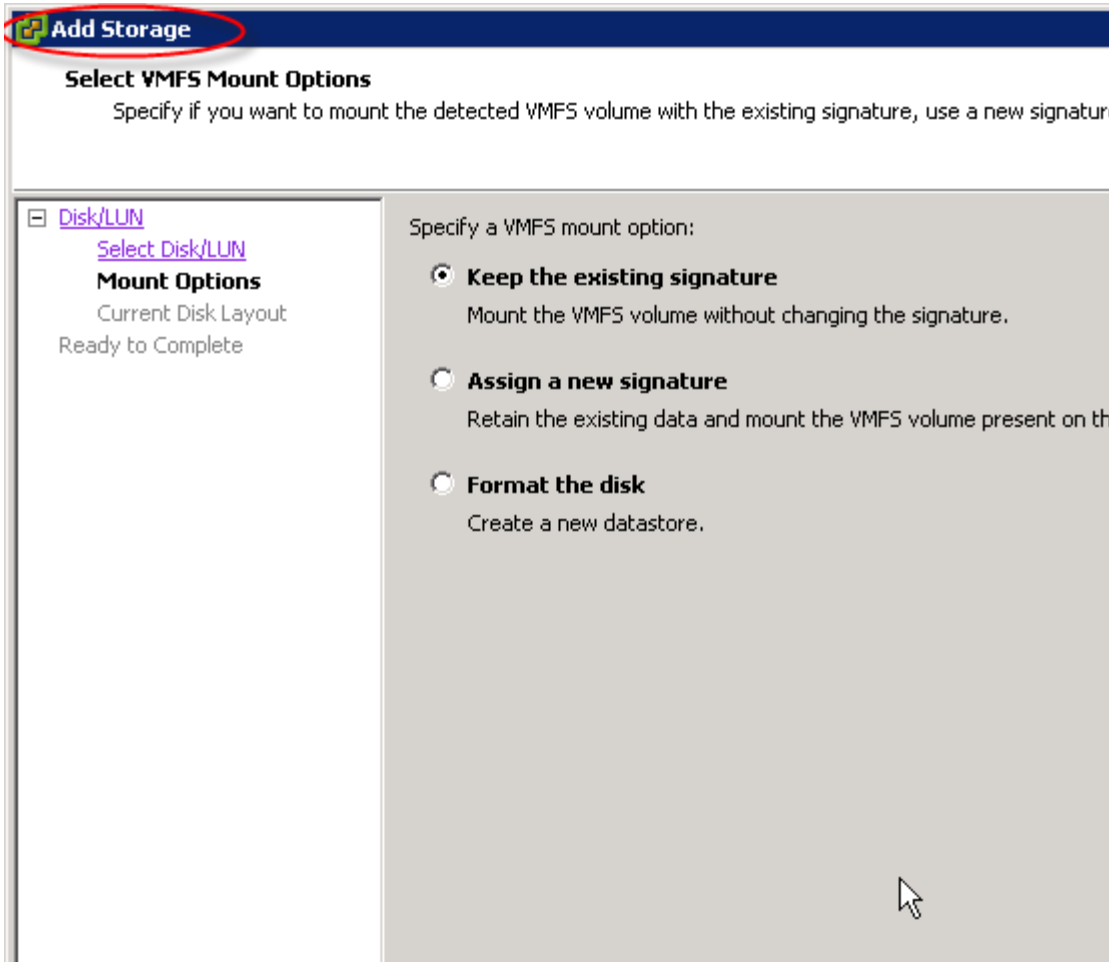


Figure 92 Selecting the "Keep the existing signature" option for a given LUN replica

3. The registration of the virtual machines from the target device can be performed using Virtual Infrastructure client or the service console and is similar to the Virtual Infrastructure 3 registration.

4. The cloned virtual machines can be powered on by utilizing the vCenter client or the `vmware-cmd` utility on the service console. Figure 93 on page 215 shows one of the cloned virtual machine powered on, on the target VMware ESX/ESXi host.

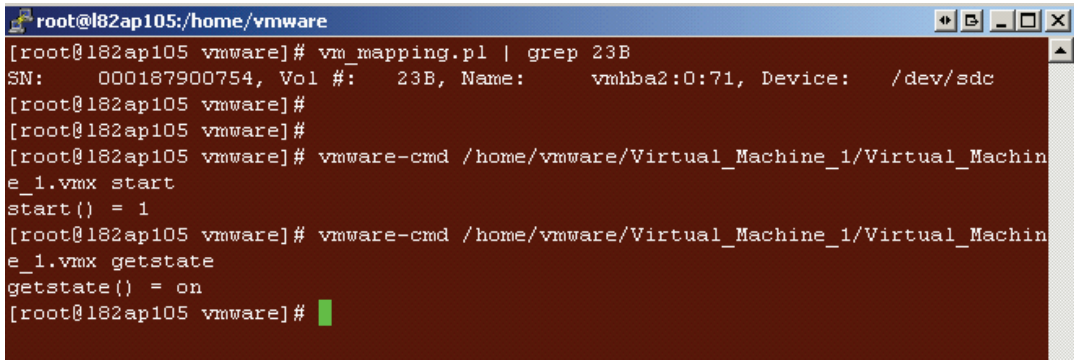


Figure 93 Powering on cloned virtual machines on the target VMware ESX

Cloning Virtual Infrastructure 4 VMs using the "Assign a new signature" option

The following steps are required to clone virtual machines using copied data:

1. Once the replica is presented to the target ESX server or same ESX server as the source LUN, the SCSI bus should be rescanned. This can be done either using the service console or the vCenter client. The devices holding the copy of the VMware file system are displayed in the "Add storage" wizard in vSphere Client. When the Assign a new signature option is used for a LUN within aVMware ESX 4.0 cluster as shown in Figure Figure 94, it allows the ESX 4.x

cluster to present both source and target devices simultaneously. The LUN is resignatured, relabeled, and displayed on the target VMware ESX/ESXi host.

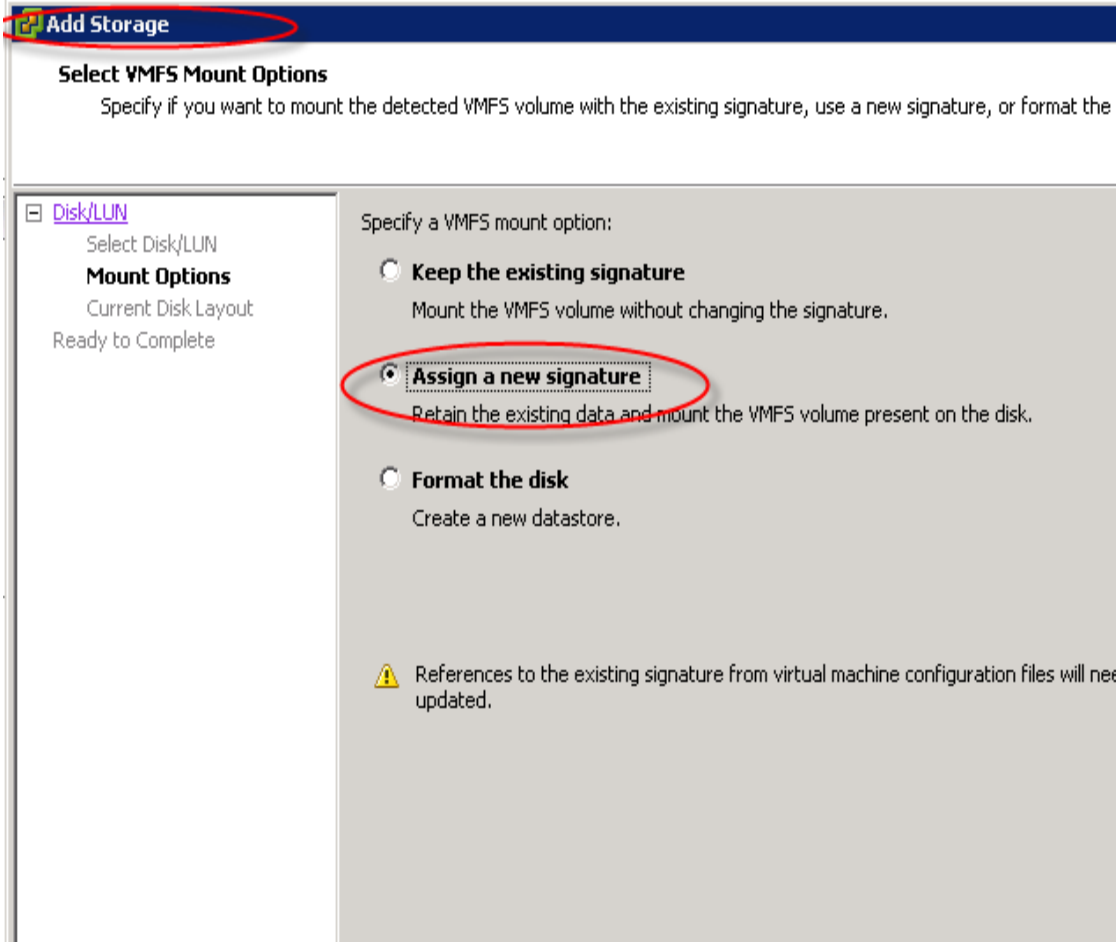


Figure 94 Selecting the "Assign a new signature" option for a given LUN replica

2. When a virtual machine is created in a vSphere 4 environment, all files related to the virtual machine are stored in a directory on a datastore. This includes the configuration file and, by default, the virtual disks associated with a virtual machine. Thus, the configuration files automatically get replicated with the virtual disks when storage array based copying technologies are leveraged.

3. The registration of the virtual machines from the target device can be performed using Virtual Infrastructure client or the service console. The registration using the service console utility, `vmware-cmd`, is shown in Figure 91 on page 211.The registration of cloned virtual machines is not required every time the target devices are refreshed with the latest copy of the data from the source device.

   There is a tight integration between the vCenter infrastructure and VMware ESX version 4 or VMware ESXi that does not allow duplicate object names. Therefore, when registering the copy of virtual machines using the vCenter client, the cloned virtual machines should be provided with a unique name. The display name for the cloned virtual machines registered using `vmware-cmd` is automatically renamed by the vCenter management server if the target VMware ESX/ESXi host cluster is in the same vCenter data center as the source cluster group. The change in the cloned virtual machine name, however, does not impact the operations that can be performed.

4. The cloned virtual machines can be started on the target VMware ESX/ESXi host without any modification if the following requirements are met:

   - The target VMware ESX/ESXi host have the same virtual network switch configuration— that is, the name and number of virtual switches should be duplicated from the source VMware ESX/ESXi host cluster group.
   - The virtual disks allocated to the source virtual machine are contained in the same VMware file system that contains the configuration file.
   - The minimum memory and processor resource reservation requirements of all cloned virtual machines can be supported on the target VMware ESX/ESXi host. For example, if 10 source virtual machines, each with a memory resource reservation of 256 MB, need to be cloned, the target VMware ESX/ESXi host cluster should have at least 2.5 GB of physical RAM allocated to the VMkernel.

- Devices, such as CD-ROM and floppy drives, are attached to physical hardware, or are started in a disconnected state when the virtual machines are powered on.

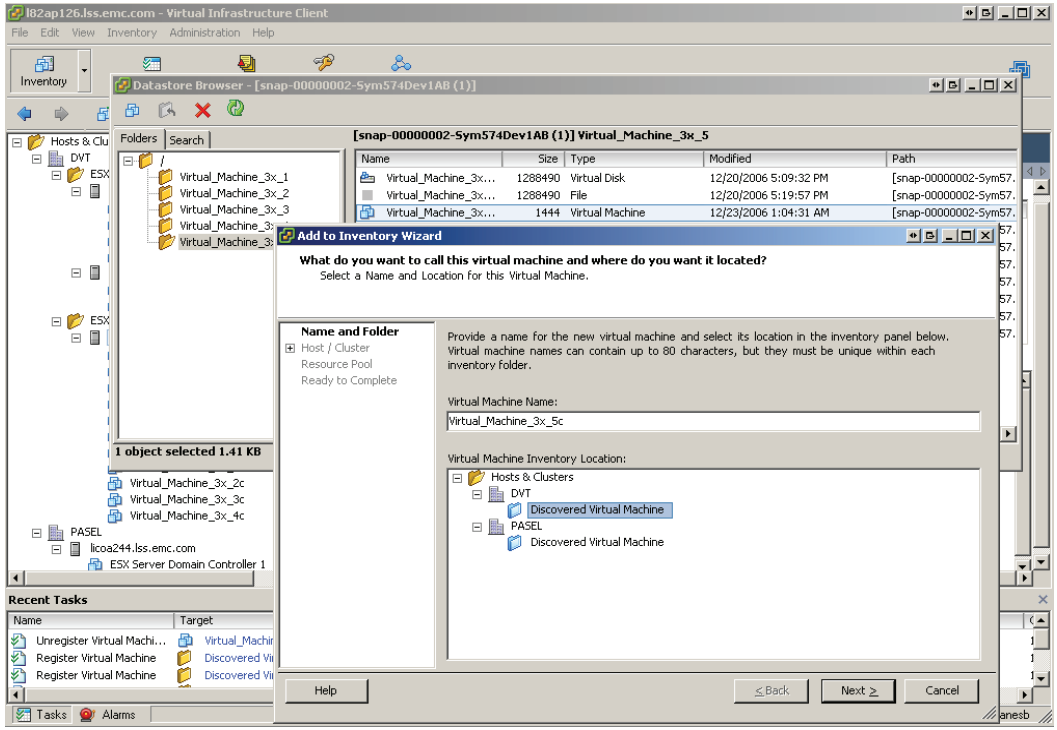


Figure 95 Registering virtual machines using resignatured volumes

The cloned virtual machines can be powered on using Virtual Infrastructure client or command line utilities, as shown in Figure 96 on page 219. The use of volume resignaturing introduces complexity into the cloning process. This is particularly true if the source virtual machines are provided with virtual disks that are on other VMware file systems; this is explained in the best practices recommendations in "Number of VMware file systems (VMFS) in a ESX Server 4 or 3 cluster," on page 155. In this case, changes to the configuration of the cloned virtual machine are required before the machine can be powered on. For this reason, the use of the **`Assign a new resignature`** option should be limited to environments that cannot provide dedicated VMware ESX/ESXi hosts to run the cloned virtual machines.

Cloning virtual machines using RDM in VMware vSphere 4 and Virtual Infrastructure 3 environments

The configuration file located in the VMFS-3 volumes can be used to clone virtual machines provided with storage using RDM. However, in the VMware vSphere 4 andVirtual Infrastructure 3 environments, it is easier to use copies of the configuration files on the target VMware ESX/ESXi hosts.

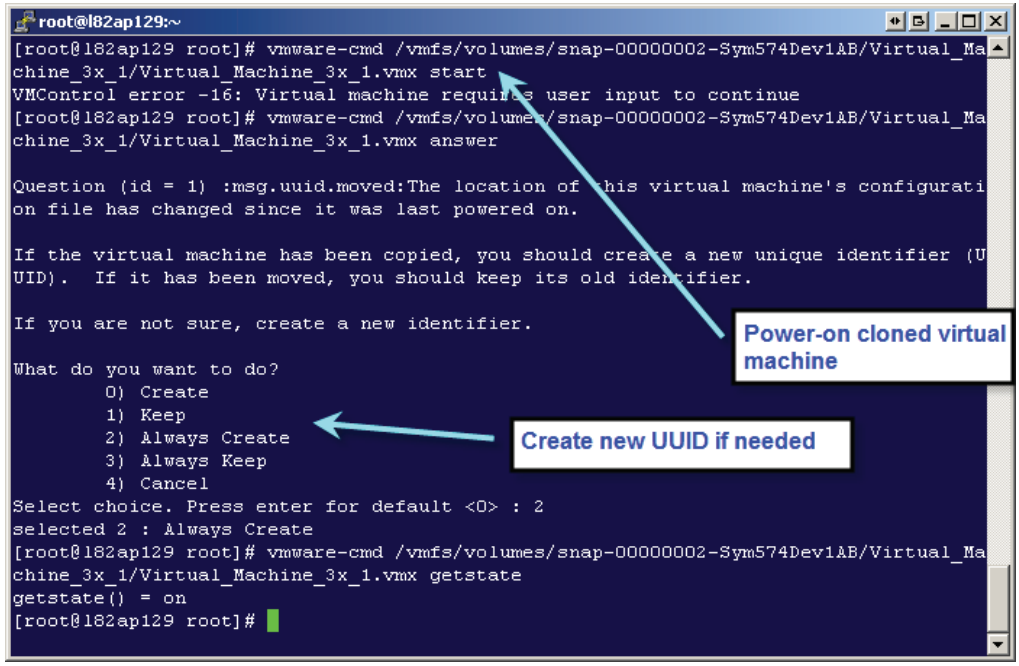


Figure 96 Power on cloned virtual machines on a resignatured target volume

When a RDM is generated, a file is created on a VMware file system that points to the physical device that is mapped. The file that provides the mapping also includes the unique ID and LUN number of the device it is mapping. The configuration file for the virtual machine using the RDM contains an entry that includes the label of the VMware file system holding the RDM and its name. If the VMware file system holding the information for the virtual machines is replicated and presented on the target VMware ESX/ESXi host, the virtual disks that provide the mapping is also available in addition to the configuration files. However, the mapping file cannot be used on the target VMware ESX/ESXihost since the cloned virtual machines need to be provided

with access to the devices holding the copy of the data. Therefore, EMC recommends using a copy of the source virtual machine's configuration file instead of replicating the VMware file system. The following steps clone virtual machines using RDMs in a Virtual Infrastructure 3 environment:

1. On the target VMware ESX/ESXihost, create a directory on a datastore (VMware file system or NAS storage) that holds the files related to the cloned virtual machine. A VMware file system on internal disk, un-replicated SAN-attached disk or NAS-attached storage should be used for storing the files for the cloned virtual disk. This step has to be performed once.
2. Copy the configuration file for the source virtual machine to the directory created in step 1. The command line utility, `scp`, can be used for this purpose. This step has to be repeated only if the configuration of the source virtual machine changes.
3. Register the cloned virtual machine using the vCenter client or the service console. This step does not need to be repeated.
4. Generate RDMs on the target VMware ESX/ESXi host in the directory created in step 1. The RDMs should be configured to address the target devices.
5. The cloned virtual machine can be powered on using either the Virtual Infrastructure client or the service console.

Note: The process listed in this section assumes that the source virtual machine does not have a virtual disk on a VMware file system. The process to clone virtual machines with a mix of RDMs and virtual disks is complex and beyond the scope of this document. Readers are requested to contact the authors at Kochavara_sheetal@emc.com or Ganeshan_bala@emc.com if such requirements arise.

Choosing a VM cloning methodology

The replication techniques described in the previous sections have pros and cons with respect to their applicability to solve a given business problem. The matrix in Table 3 on page 221 provides a comparison of the different replication methods to use and the differing attributes of those methods.

Table 3 A comparison of storage array based virtual machine cloning technologies

	Snapshots	Clones
Maximum number of copies per source LUN	8	8
Production impact	COFW	COFW
VM clone needed a long time	Not Recommended	Recommended
High write usage to VM clone	Not Recommended	Recommended

COFW = Copy on First Write

Table 4 on page 222 shows examples of the choices a VMware or storage administrator might make for cloning virtual machines based on the matrix presented in Table 3 on page 221.

Table 4 Virtual machine cloning requirements and solutions

System Requirements	Replication Choice
The application on the source volumes is performance sensitive, and the slightest degradation cause responsiveness of the system to miss SLAs.	SnapView Clone
Space and economy are a concern. Multiple copies are needed and retained only for a short time, with performance not critical.	Snap View snapshots

5
Backup and Restore of Virtual Machines

This chapter presents the following topics:

- Recoverable versus restartable copies of data 225
- Backing up using copies of Virtual Infrastructure data 227
- Restoring VMs data using disk-based copies 242

All IT environments create backup procedures to protect their critical data. The backup processes run one or more times per day to protect the data in the event of user error, data loss, system outage, or catastrophic event. Modern environments require backup processes to complete while all business applications remain in service. Furthermore, the backup processes are expected to have minimum impact on the performance of the most critical systems.

This chapter describes how the IT personnel can leverage EMC technologies to:

- Reduce production impact of backups.
- Create consistent point-in-time backup images.
- Provide alternate methodology for file-level restore using disk-based copies.
- Enhance recovery times in case of catastrophic failures.

Recoverable versus restartable copies of data

The CLARiiON-based replication technologies can generate a restartable or recoverable copy of the data. The difference between the two types of copies can be confusing; a clear understanding of the differences between the two is critical to ensure that the recovery goals for a Virtual Infrastructure environment can be met.

Using recoverable disk copies

A recoverable copy of the data is one in which the application (if it supports it) can apply logs and roll the data forward to an arbitrary point in time after the copy was created. The recoverable copy is most relevant in the database realm where database administrators use it frequently to create backup copies of database. In the event of a failure to the database, the ability to recover the database not only to a point-in-time when the last backup was taken, but also to roll forward subsequent transactions up to the point of failure is critical to most business applications. Without that capability, in an event of a failure, there will be an unacceptable loss of all transactions that occurred since the last backup.

Creating recoverable images of applications running inside virtual machines using EMC replication technology requires that the application or the virtual machine be shut down when it is copied. A recoverable copy of an application can also be created if the application supports a mechanism to suspend writes when the copy of the data is created. Most database vendors provide functionality in their RDBMS engine to suspend writes. This functionality has to be invoked inside the virtual machine when EMC technology is deployed to ensure a recoverable copy of the data is generated on the target devices.

Using restartable disk copies

If a copy of a running virtual machine is created using EMC consistency technology without any action inside the virtual machines, the copy is normally a restartable image of the virtual machine. This means that when the data is used on cloned virtual machines, the operating system and the application enter into crash recovery. The exact implications of crash recovery in a virtual machine depend on the application that the machine supports:

- If the source virtual machine is a file server or runs an application that uses flat files, the operating system performs a file-system check and fixes any inconsistencies in the file system. Modern file systems such as Microsoft NTFS use journals to accelerate the process
- When the virtual machine is running any database or application with a log-based recovery mechanism, the application uses the transaction logs to bring the database or application to a point of consistency. The process deployed varies depending on the database or application, and is beyond the scope of this document.

Most applications and databases cannot perform roll-forward recovery from a restartable copy of the data. Therefore, a restartable copy of data created from a virtual machine that is running a database engine is inappropriate for performing backups. However, applications that use flat files or virtual machines that act as file servers can be backed up from a restartable copy of the data. This is possible since none of the file systems provide logging mechanism that enable roll forward recovery.

Note: Without additional steps, VMware Consolidated Backup (VCB) creates a restartable copy of virtual disks associated with virtual machines. The quiesced copy of the virtual disks created by VCB is similar to the copy created using EMC consistency technology.

Backing up using copies of Virtual Infrastructure data

The cloned virtual machines can be utilized for supporting business processes such as reporting, QA, testing, and development. As discussed in "Recoverable versus restartable copies of data," on page 225, depending on the application running on the source virtual machines and the type of copy on the target devices, the copy can be used for performing backups.

The next few sections discusses various options available inVMware ESX 4.x and 3.x environments to back up the virtual infrastructure data. All backup strategies discussed therein optimize the utilization of the IT resources by offloading the backup process from production servers to a dedicated backup environment.

Using the ESX Server version 3.x service console

The service console of the target VMware ESX in a Virtual Infrastructure 3 environment can be used to back up the virtual disks to tape. However, for the copy of the data to be accessible on the target VMware ESX cluster, either the `LVM.DisallowSnapshotLun` should be disabled or `LVM.EnableResignature` should be enabled. "Cloning VMs on VMware file systems in VMware Infrastructure 3," on page 200 provides detailed discussion about the differences between the two parameters and its use with SnapView family.

- A backup agent, such as an NetWorker® client agent, can be installed on the service console of the target VMware ESX host. The virtual disks can be backed up as individual files to tapes on a storage node over the IP network.
- The target VMware ESX can also be configured to be a storage node. In this configuration, the target VMware ESX can perform backups to a local disk or a tape drive.
- Since the granularity of restore is limited to the whole virtual disk, this approach is appropriate when a tape based remote disaster recovery solution is desired. The solution is also appropriate for protection against catastrophic failures in the data center.
- Virtual Infrastructure version 3 includes a new paradigm for backing up virtual machines with Microsoft Windows as the guest operating system. The new product, VMware Consolidated Backup (VCB), enables off-host backup of virtual machines thus eliminating backup load from the VMware ESX hosts. Furthermore, the product

includes integration modules jointly developed with major backup vendors. The integration modules allow virtual machines running Microsoft Windows operating system to be backed up on a Windows proxy server using techniques similar to those deployed for physical servers.

- VMware Consolidated Backup provides an excellent mechanism to back up virtual machines. However, the proxy hosts continue to access the production volumes to perform the backups. The backup activity can thus impact production workloads. The VCB framework has not been integrated with storage array based replication products such as EMC SnapView. However, VCB provides tools to offload all backup activities from the production environment.

The following steps can be followed to create a backup using `vcbMounter` and a copy of virtual machine data:

1. Virtual Infrastructure 3 provides tight integration between various components. This integration includes VMware Consolidated Backup. To perform backups using `vcbMounter` while maintaining a secure virtual infrastructure, a new role called Backup Operator should be created. This role, cloned from the Read-Only role defined by default in the vCenter management server, has

permissions shown in Figure 97 on page 229. The limited role ensures that the backup operator does not get unnecessary permissions to the Virtual Infrastructure 3 environment.

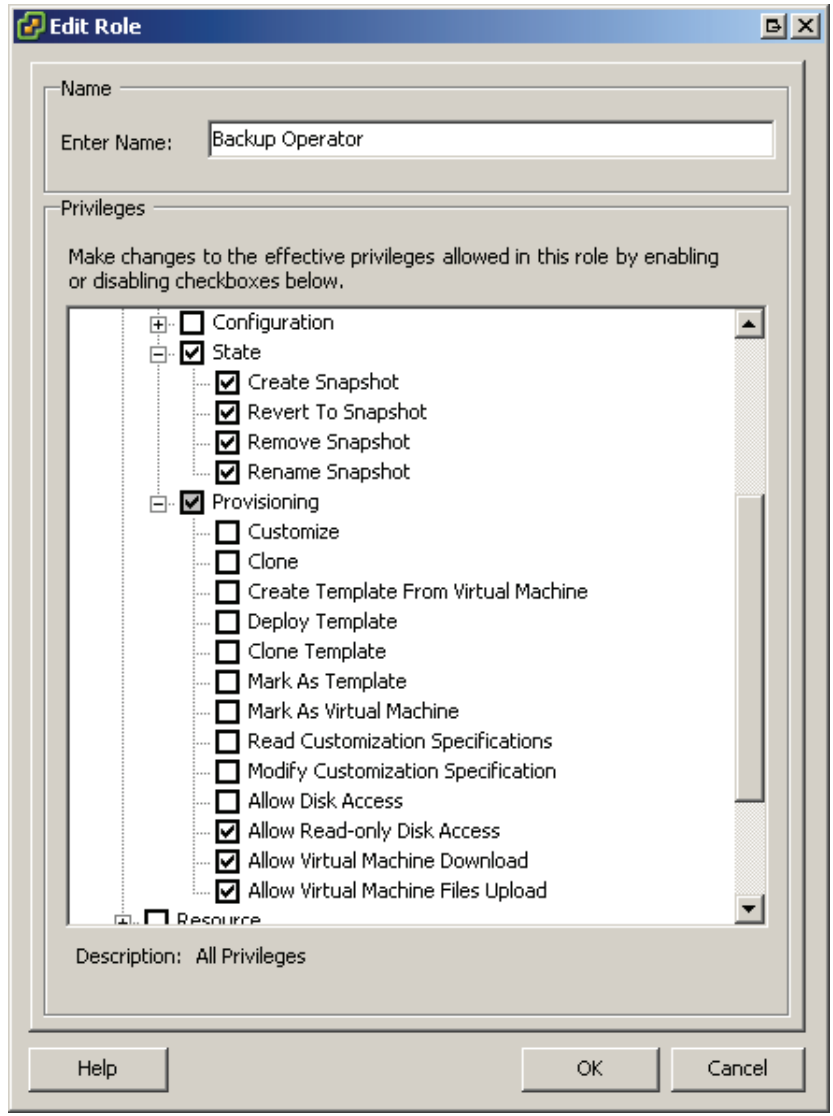


Figure 97 Defining a Backup Operator role in vCenter

2. The Backup Operator role should be associated with a user or group. This enables that a user or members of the group to initiate backup on the target VMware ESX utilizing `vcbMounter`. This can be seen in Figure 98 on page 230, where the domain user, APIAD\backup, is provided with permissions to back up the virtual infrastructure data using `vcbMounter`.

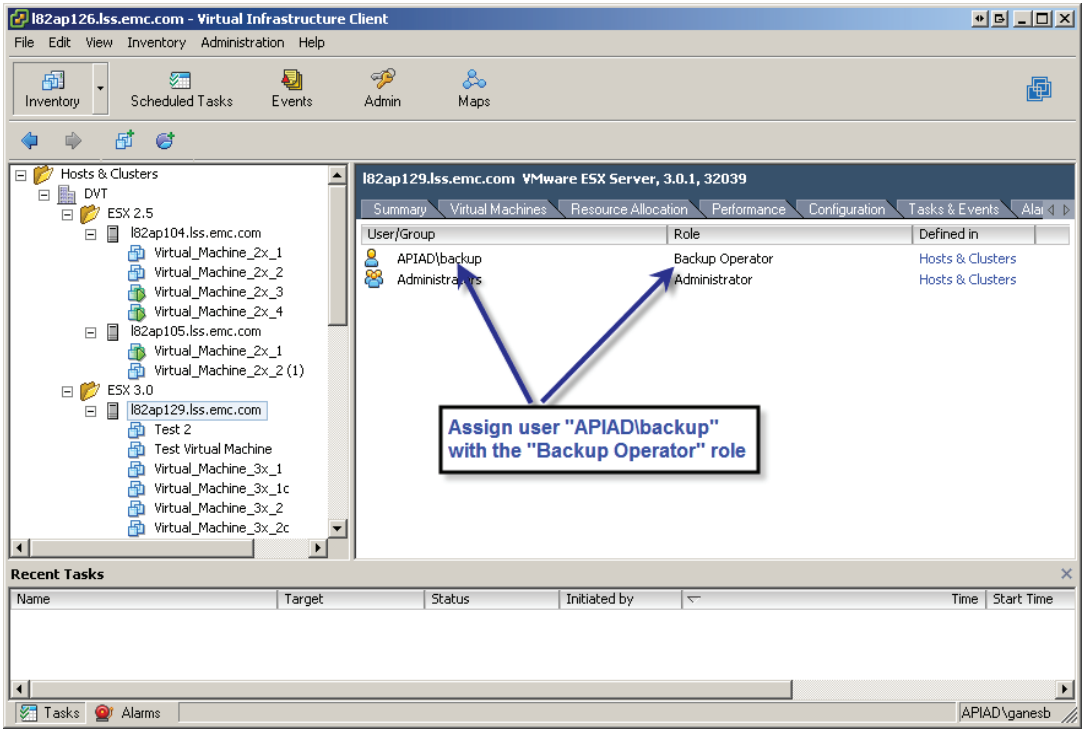


Figure 98 Assigning a Backup Operator role to a domain user

3. Figure 99 on page 232 shows an example where `vcbMounter` is used to back up a cloned virtual machine. In the example, `vcbMounter` is used to communicate with the vCenter management server (l82ap126.lss.emc.com) using the user ID, APIAD\backup, that has appropriate permissions. The virtual machine to be backed up is also provided as an option to the `vcbMounter` command. The backup is performed to a local directory, /backups, on the target VMware ESX. The file system, /backups, is an EXT3 file system. Unlike `vmsnap.pl` or `vmsnap_all`, `vcbMounter` does not require the cloned virtual

machines to be powered on before backups can be performed. The ***VMware Virtual Machine Backup Guide*** available at the VMware website provides further details.

Note: Backup vendors use different terminology to describe similar components in the backup infrastructure. For example, storage node used by EMC NetWorker is equivalent to the media server used by Symantec NetBackup. This document uses the terms defined by EMC NetWorker when appropriate. The ***VMware Virtual Machine Backup Guide*** available at the VMware website provides information to determine the equivalent component.

The backup process described in this section is a convenient mechanism to implement a backup-to-disk philosophy in a Virtual Infrastructure 3 environment. As stated earlier, the service console of the VMware ESXcan be designated as a storage node. The node can be configured to use local disks as the target for the backups of the cloned virtual machine. The local disks can actually be SATA drives on a CLARiiON or DMX storage array.

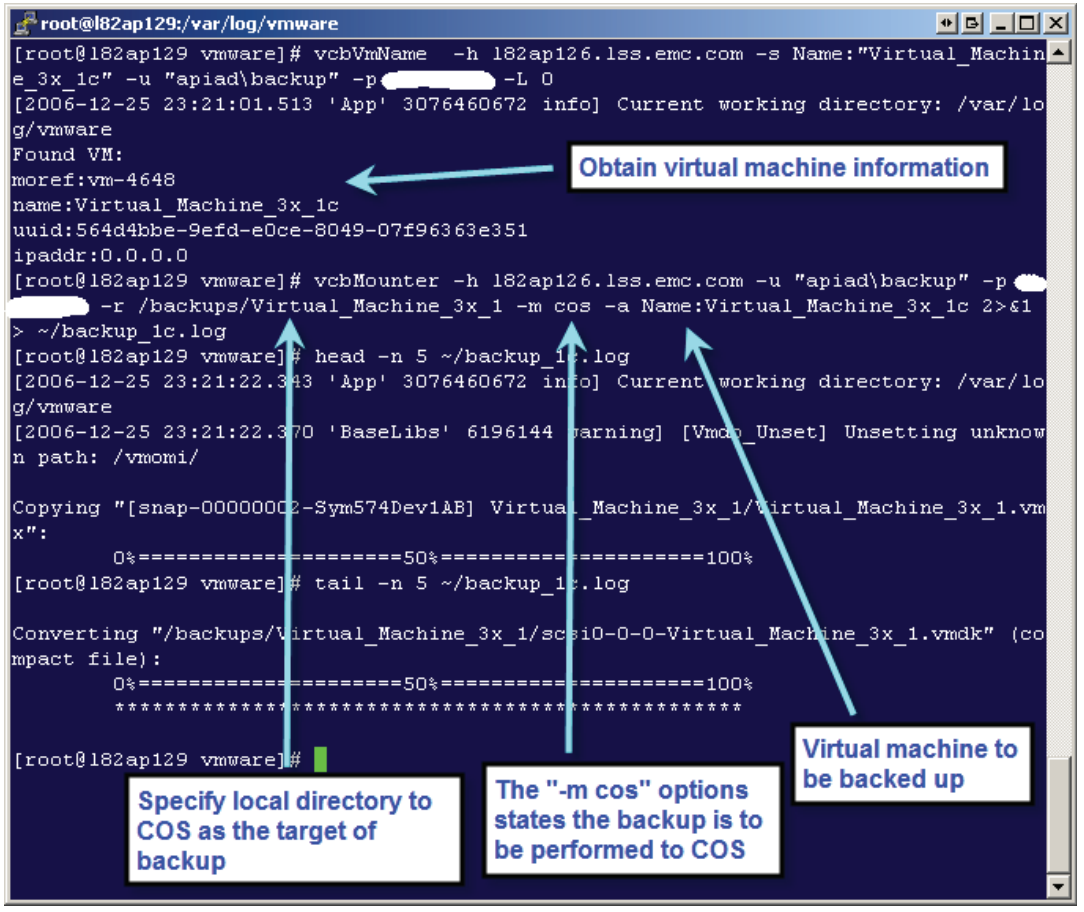


Figure 99 Using vcbMounter to perform backups using the VMware ESX 3.x service console

- EMC Replication Manager (RM) integrates with Virtual Infrastructure version 3 and has the ability to replicate VMFS and RDM volumes. For 7 VMFS volumes, RM supports the replication of an individual virtual disk on a dedicated VMFS datastore presented to a single virtual machine, or the replication of the entire VMFS datastore containing multiple VMs.

The following steps outline how Replication Manager replicates VMFS volumes:

1. For the replication of an individual virtual disk assigned to a single virtual machine, the RM agent is installed on the virtual machine itself. This gives it the ability to freeze/thaw the application and create either snaps or clone replicas on the CLARiiON storage system as shown in Figure 100 on page 233.

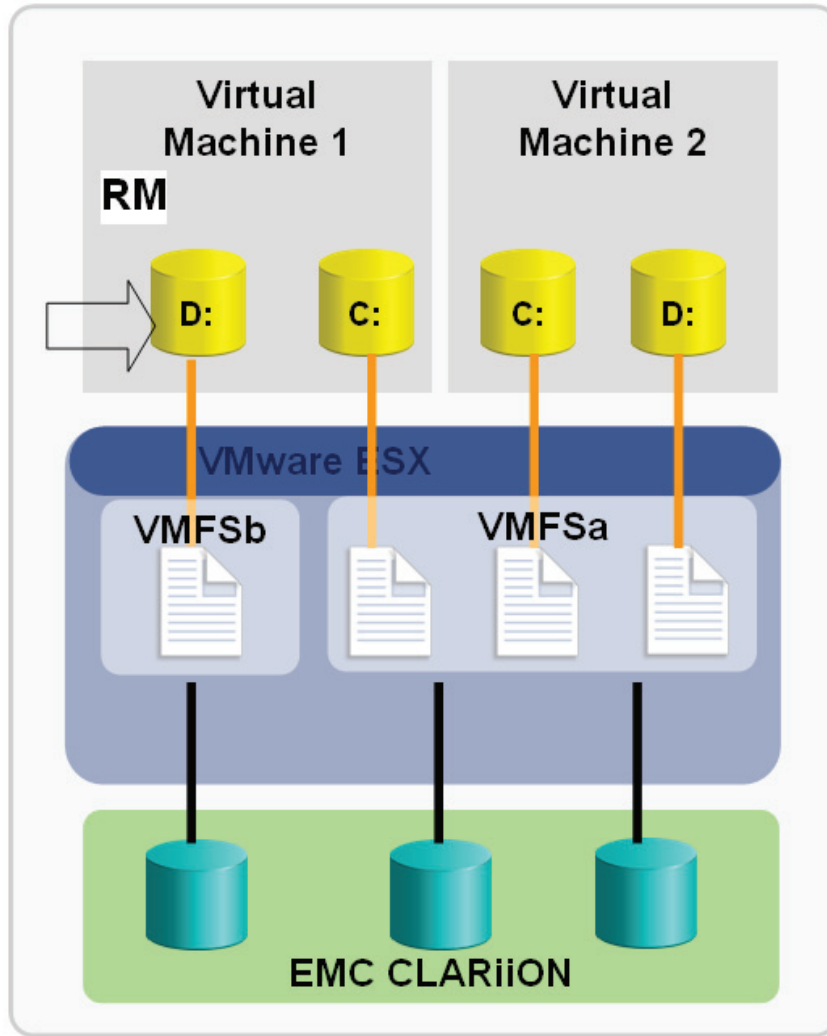


Figure 100 Replication Manager Backup for individual virtual disk

2. For the replication of an entire VMFS datastore on single or multiple LUNs containing multiple virtual machines, a Replication Manager Windows proxy host is needed. This proxy host schedules the replication of the VMFS datastores, and can be a virtual or physical machine. Since no agent is installed on the virtual machine to freeze/thaw applications, the replicas are crash-consistent when the VMs are running. This is shown in Figure 101 on page 235.

   Please note that the parameter **`LVM.EnableResignature`** must be set to **ON** in VMware ESX/ESXi hosts.

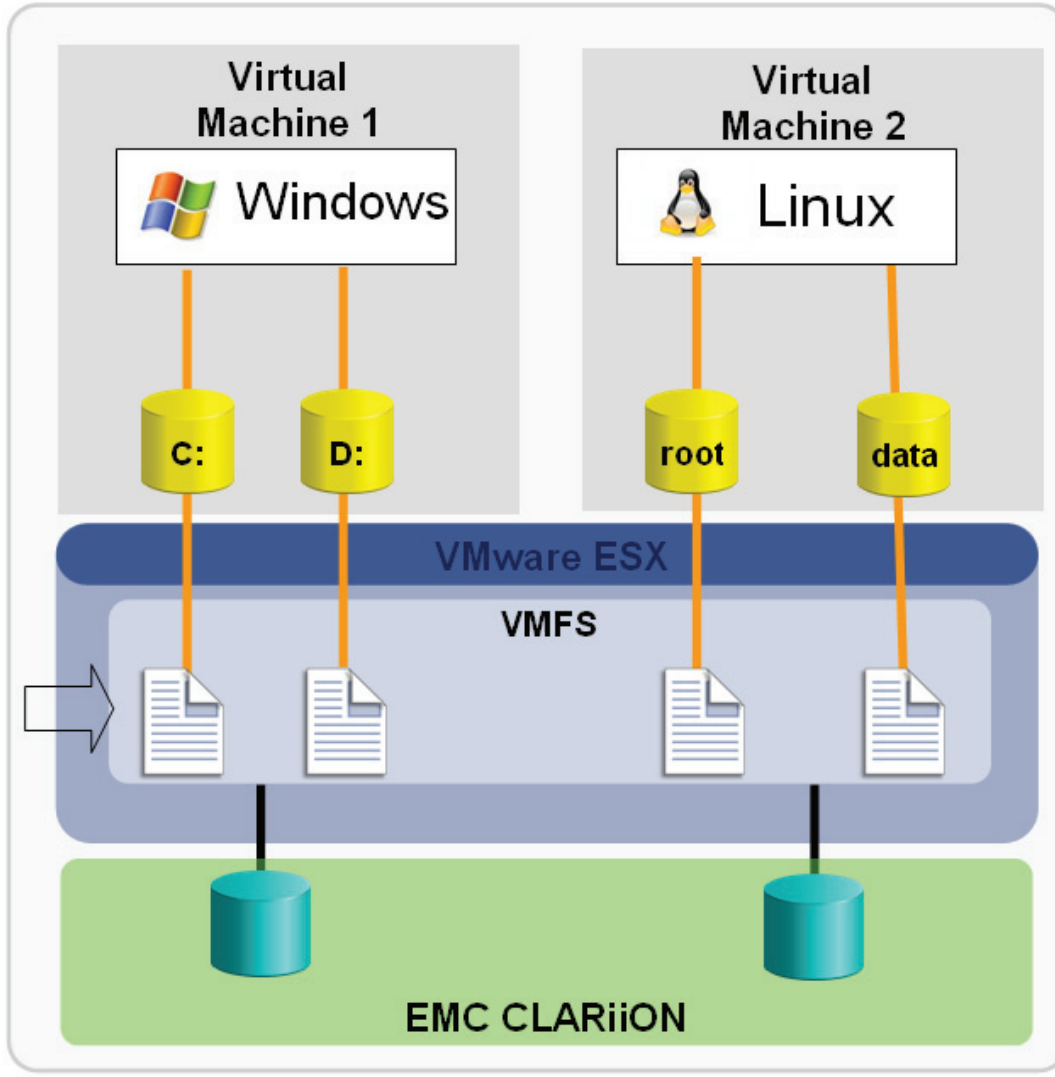


Figure 101 Replication Manager backup for an entire VMFS datastore

Using the ESX server version 4.x service console

The service console of the target VMware ESX/ESXi hosts in a VMware vSphere 4 environment can be used to back up the virtual disks to tape. However, for the copy of the data to be accessible on the target VMware ESX/ESXi host cluster, either **Keep existing signature** or **Assign a new signature** should be set on the individual LUN. "Cloning VMs on VMware file systems (VMFS-3) with VMware vSphere 4," on page 212 describes the differences between the two parameters, and how they are used with the SnapView family.

- A backup agent, such as an NetWorker client agent, can be installed on the service console of the target ESX server. The virtual disks can be backed up as individual files to tapes on a storage node over the IP network.
- The target VMware ESX/ESXi hosts can also be configured to be a storage node. In this configuration, the target VMware ESX/ESXi hosts can perform backups to a local disk or a tape drive.
- Since the granularity of restore is limited to the whole virtual disk, this approach is appropriate when a tape-based remote disaster recovery solution is desired. The solution is also appropriate for protection against catastrophic failures in the data center.
- VMware vSphere 4 includes a new paradigm for backing up virtual machines. The new product, VMware Data Recovery, enables disk-based backup of virtual machines directly through vCenter. Furthermore, the product utilizes built-in data deduplication technology to save significant disk space.

The following steps can be followed to create backups using virtual machine data:

1. VMware Data Recovery installs as a plug-in with VMware vCenter to back up virtual machines. In addition, a backup appliance running as a virtual machine must be configured on the vCenter. The VMware Data Recovery backup appliance can then connect to

the vCenter server as shown in Figure 102 on page 237. It automatically identifies all the virtual machines under VMware vCenter.

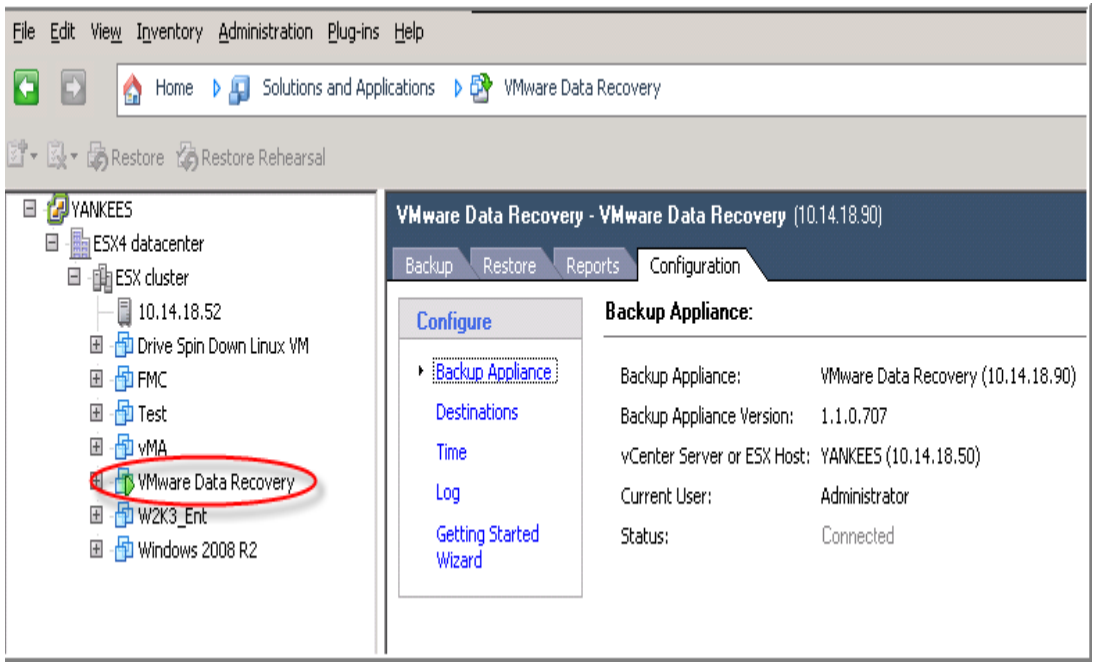


Figure 102 Creating and connecting the VMware Data Recovery backup appliance

2. After the backup appliance is configured, a Backup job can be scheduled using a wizard by
   - Selecting the VM machines that need to be backed-up
   - Selecting the Destination or target storage disk on which the data will be stored
   - Selecting the schedule

This can be seen in Figure 98 on page 230, where the virtual machines, destination storage, schedule and retention policy for a particular VM are selected.

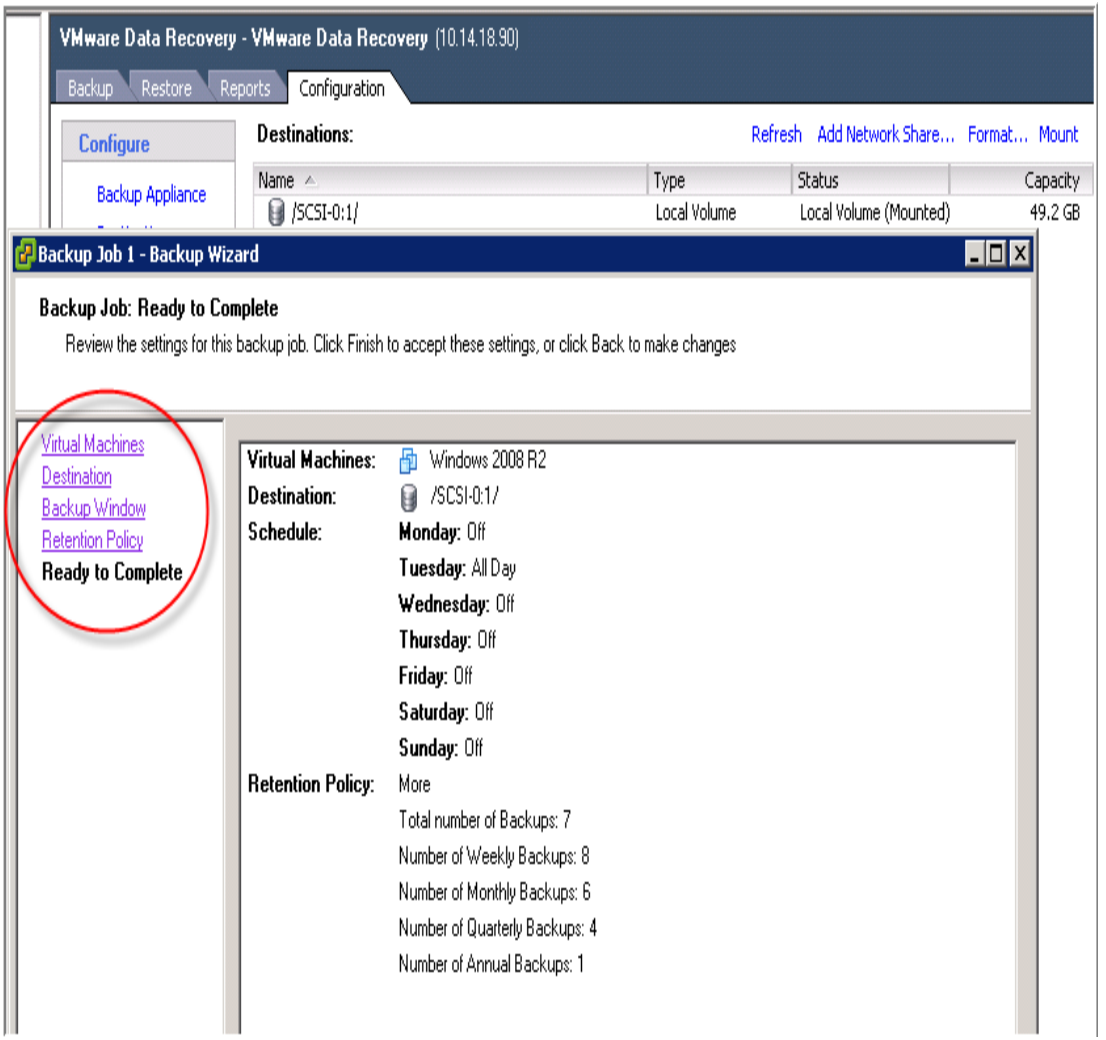


Figure 103 Backup job wizard within VMware Data Recovery appliance

3. As soon as the backup job is executed depending on the schedule selected, VMware Data Recovery creates a snapshot of the VM and preforms deduplication to reduce the amount of storage on the target destination. The ***VMware Data Recovery Administration Guide*** available at the VMware website provides further details.

Note: Backup vendors use different terminology to describe similar components in the backup infrastructure. For example, a storage node used by EMC NetWorker is equivalent to the media server used by Symantec NetBackup. This document uses the terms defined by EMC NetWorker when appropriate. The ***VMware Virtual Machine Backup Guide*** available at the VMware website provides information to determine the equivalent component.

The backup process described in this section is a convenient mechanism to implement a backup-to-disk philosophy in a VMware vSphere 4 environment. As stated earlier, the service console of the VMware ESX can be designated as a storage node. The node can be configured to use local disks as the target for the backups of the cloned virtual machine. The local disks can actually be SATA drives on a CLARiiON or DMX storage array or LC-FC drives on the DMX storage arrays.

EMC Replication Manager (RM) also integrates with VMware vSphere 4 and has the ability to replicate VMFS and RDM volumes. As mentioned previously, RM supports the replication of an individual virtual disk on a dedicated VMFS datastore presented to a single virtual machine, or the replication of the entire VMFS datastore containing multiple VMs.

The following steps outline how Replication Manager replicates VMFS volumes:

1. For the replication of an individual virtual disk assigned to a single virtual machine, the RM agent is installed on the virtual machine itself. This gives it the ability to freeze/thaw the application and create either snaps or clone replicas on the CLARiiON storage system.
2. For the replication of an entire VMFS datastore on single or multiple LUNs containing multiple virtual machines, a Replication Manager Windows proxy host is needed. This proxy host schedules the replication of the VMFS datastores, and can be a virtual or physical machine. Since no agent is installed on the virtual machine to freeze/thaw applications, the replicas are crash-consistent when the VMs are running.

Replication Manager automaticallyresignatures the individual LUN for VMware vSphere environments, hence no manual configuration is required.

Using cloned VMs in VMware vSphere 4 and VMware Infrastructure 3

For both ESX 3.x and 4.x servers, backup agent can be installed on the source virtual machines with a policy of client-initiated backup. The cloned virtual machines can then be used to back up the virtual machines to a pre-determined storage node using the backup network infrastructure. This approach is similar to the traditional backup mechanism deployed to back up physical servers.

Within VMware vSphere the VMware Data Recovery product provides acentralized mechanism to back up VMs directly from VMware vCenter Server while VCB offloads the backups to be a proxy server in VMware Infrastructure 3 environments. However, both VMware Data Recovery and VCB use the production volumes during backups. The use of cloned virtual machines to perform backups should be considered in VMware vSphere environment. The use of cloned virtual machines for backups ensures there is no impact to the production infrastructure thus enabling extended backup windows and optimal use of backup resources.

The detailed architecture and processes required to perform backups using cloned virtual machines is beyond the scope of this document.

Using RDM

The offloading of backup activities to a designated group of VMware ESX/ESXi hosts discussed in the previous sections can be used with virtual machines using RDM. However, these virtual machines provide an additional level of optimization not afforded by virtual machines that use storage on VMware file systems.

As discussed earlier, RDM allows virtual machines to directly access the storage devices. Therefore, the storage devices can be presented and used on physical server running the same operating system as the virtual machine. This fact can be exploited to accelerate backup and restore of virtual machine data volumes. Instead of using VMware ESX/ESXi hosts to perform the backups, the copies of the data can be directly presented to a storage node and backed up locally. This approach reduces the amount of complexity and cost by eliminating the need for target VMware ESX/ESXi hosts to perform the backups.

For more information about how to schedule backups of RDM volumes using EMC Replication Manager, please see ***EMC Replication Manager with CLARiiON and VMware ESX Server– Best Practice Planning.***

Restoring VMs data using disk-based copies

The disk-based copy of virtual infrastructure data created using SnapView family can be utilized as a first line of defense if the need to restore data arises. If the techniques discussed in Section Backing up using copies of Virtual Infrastructure data are deployed, the backup copies on the second tier storage can also be utilized to restore data. The next few sections discuss various methods that can be used to restore virtual machines.

Note: The use of disk-based copies of data is ideal when there is a catastrophic failure of virtual infrastructure components. Although restore of individual files can be accomplished using the disk-based copies, traditional backup and restore techniques are much more suited for that business need.

SnapView copies for VMs with VMFS-hosted virtual disks

The copy of the virtual machine data on the clone or snapshot devices created using SnapView can be used to restore individual virtual machines or a group of virtual machines. The restore requirement dictates the required process.

Restoring individual virtual machines in a VMware ESX 3 or VMware ESX3i environment

Virtual Infrastructure 3 environment allows multiple copies of the same VMware file system to be accessed on the VMware ESX/ESXi cluster group, the restore process can be optimized (see "Cloning VMs on VMware file systems in VMware Infrastructure 3," on page 200.)

The process to restore individual virtual machine in a Virtual Infrastructure 3 environment is depicted in Figure 104 on page 243.

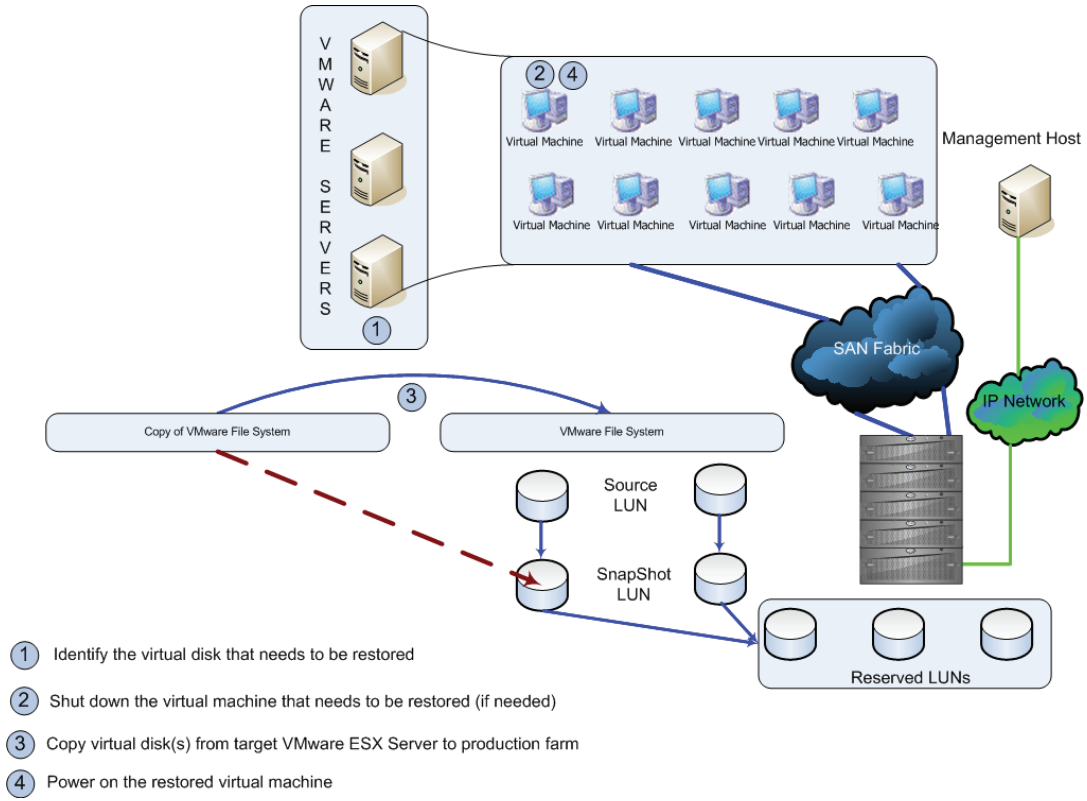


Figure 104 Restoring individual VMs using SnapView copies in Virtual Infrastructure 3

1. The virtual disks that need to be restored should be identified on the production VMware ESX/ESXi cluster group.
2. VMware ESX/ESXi hosts prevent access to the virtual disks when the virtual machines accessing the disks are in a powered-on state. Therefore, the source virtual machine that needs to be restored should be shut down.

3. The virtual disks identified in step 1 should be copied over the SAN network using the service console:

```
vmkfstools -U /vmfs/volumes/<VMFS_label>/<VM dir>/<vm.vmdk>

vmkfstools -i /vmfs/volumes/<snap VMFS_label>/<VM dir>/<vm.vmdk> \

              /vmfs/volumes/<VMFS_label>/<VM dir>/<vm.vmdk>
```

4. The restored virtual machine can be powered on. The state of the virtual machine is restored to the point when the copy of the data was created. Changes to the data that have occurred since the disk-based copy was created are lost.

Note: If the application running on the virtual machine supports roll-forward logging, the restored virtual disk can be rolled forward as long as the copy on the target devices contain a recoverable copy of the data. "Recoverable versus restartable copies of data," on page 225 provides further details.

The restore using SAN instead of the IP network is possible only when the parameter `LVM.EnableResignature` is set to 1. However, enabling the resignaturing updates the signature and the label on the target devices. If the Virtual Infrastructure 3 backup infrastructure uses a separate group of VMware ESX/ESXi hosts for performing backups, the change can negatively impact the process. In addition, resignaturing the target devices negatively impacts the capability of using the target device to restore the VMware file system. "Using EMC to copy running virtual machines ," on page 193 provides further details. Therefore, the procedure described in this section is recommended for environments resignaturing the target devices for ancillary activities.

Restoring individual virtual machines in a VMware vSphere 4 or VMware ESXi 4.x environment

You can use the same process to restore an individual virtual machine in a Virtual Infrastructure 3 environment that you use in a vSphere 4 environment. Like Virtual Infrastructure 3, vSphere 4 allows multiple copies of the same VMware file system to be accessed on the VMware ESX/ESXi cluster group; so the restore process can be optimized as described in "Cloning VMs on VMware file systems (VMFS-3) with VMware vSphere 4," on page 212.

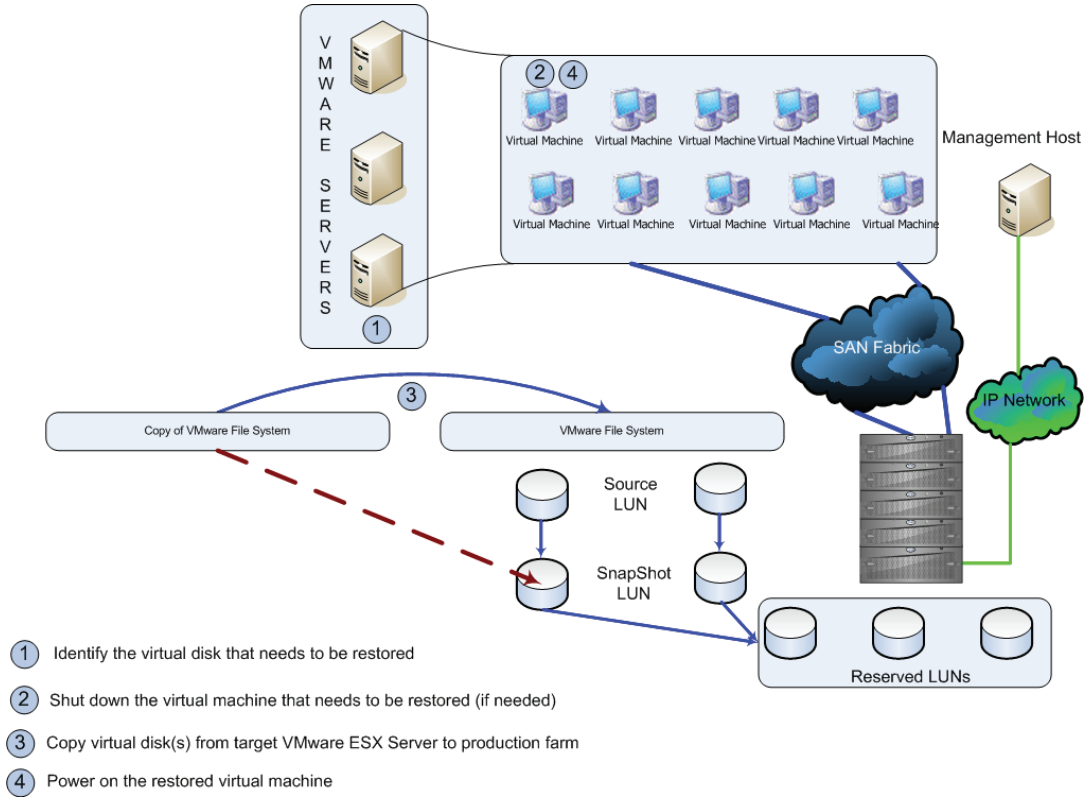


Figure 105 Restoring individual VMs using SnapView copies in VMware vSphere

The process to restore individual virtual machine in a VMware vSphere 4 environment is shown in Figure 105 on page 252 and includes the following steps:

1. Identify the virtual disks that need to be restored in the production VMware ESX/ESXi cluster group.
2. VMware ESX/ESXi hosts prevent access to the virtual disks when the virtual machines accessing the disks are in a powered-on state. Therefore, you need to shut down the source virtual machine that you wish to restore.

3. Copy the virtual disks (identified in step 1) over the SAN network using the service console:

```
vmkfstools –U /vmfs/volumes/<VMFS_label>/<VM
dir>/<vm.vmdk>

vmkfstools –i /vmfs/volumes/<snap VMFS_label>/<VM
dir>/<vm.vmdk> \

/vmfs/volumes/<VMFS_label>/<VM dir>/<vm.vmdk>
```

4. Power the restored virtual machine on. The state of the virtual machine is restored to the point when the copy of the data was created. Changes to the data that have occurred since the disk-based copy was created are lost.

Note: If the application running on the virtual machine supports roll-forward logging, you can roll the restored virtual disk forward as long as the copy on the target devices contain a recoverable copy of the data. "Recoverable versus restartable copies of data," on page 225 provides further details.

To restore using SAN instead of the IP network, the **Assign New Signature** option must be selected for the individual LUNs. As with Virtual Infrastructure 3, enabling resignaturing causes the signature and label on the target devices to be updated when a copy of the VMFS datastore is presented to the ESX host. If the vSphere 4 backup infrastructure uses a separate group of ESX servers for performing backups, the change can negatively impact the process. In addition, resignaturing the target devices negatively impacts the ability to use the target device to restore the VMware file system. "Using EMC to copy running virtual machines," on page 197 provides further details. Therefore, you should use the procedure described in this section for environments that resignature the target devices for auxiliary activities.

Restoring all virtual machines hosted on VMware file system in ESX version 3

Restoring individual virtual machines using the processes described earlier can become cumbersome if you wish to restore all virtual machines hosted on a VMware file system. The SnapView product provides a mechanism to perform incremental restores from the target device back to the production devices.

VMware ESX/ESXi version 3 hosts provide access to copies of VMware file system by either disabling the advanced parameter, `LVM.DisallowSnapshotLun`, or enabling the parameter, `LVM.EnableResignature`.

The restore of all virtual machines on a VMFS-3 volume using a copy created by SnapView is convoluted if the target volumes are accessed by setting `LVM.EnableResignature` to 1. The signature and label of the VMware file system on the target device are updated when the parameter `LVM.EnableResignature` is enabled. In a Virtual Infrastructure 3 environment, the following process needs to be used when utilizing resignatured target volumes:

1. Shut down all virtual machines accessing the production volumes. This can be done using the Virtual Infrastructure client or the service console.
2. Shut down all cloned virtual machines accessing the target volumes.

3. Remove all virtual machines accessing the production volumes from the inventory as shown in Figure 106 on page 248.

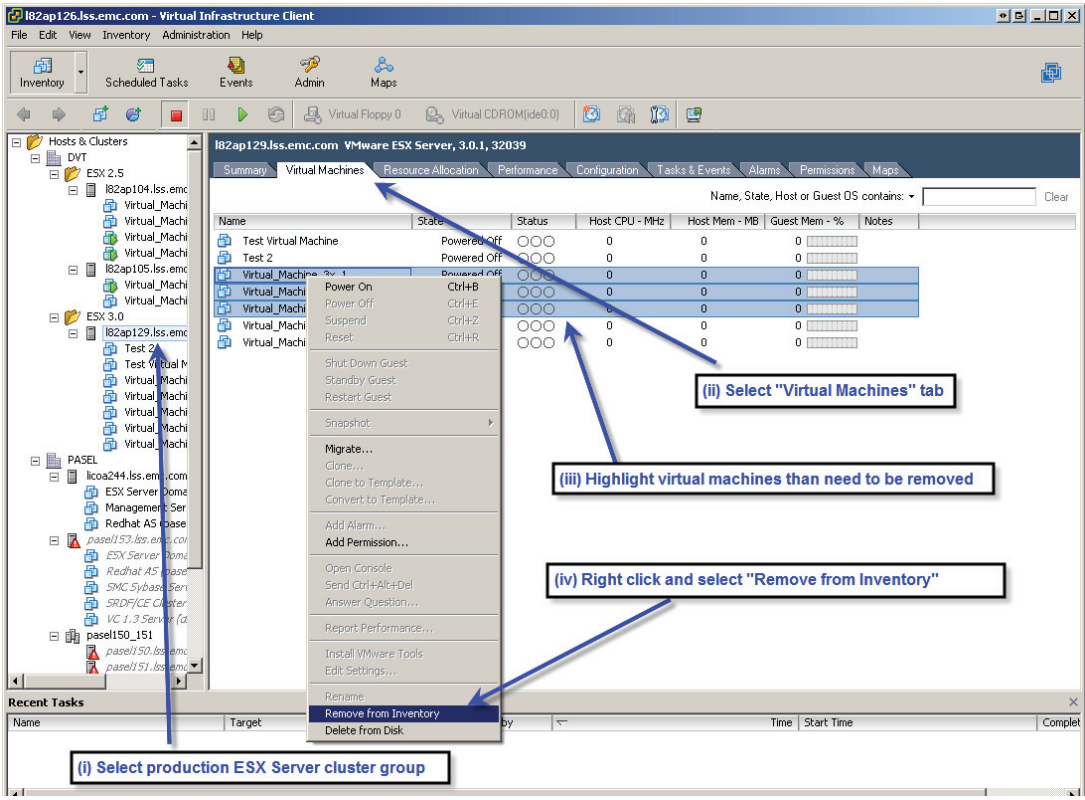


Figure 106 Removing a group of virtual machines from Virtual Infrastructure 3 inventory

4. The datastore associated with the production volumes that are being restored should be removed from the vCenter infrastructure.
5. Ensure the advanced configuration parameter, `LVM.EnableResignature`, is set to 1 on the entire production VMware ESX/ESXi cluster group.
6. Restore the data from the target volumes to the production volumes using the appropriate process:
   a. If the copy was created using SnapView clones, initiate a Reverse Synchronization operation from the clone to the source volume.

If the copy was created using SnapView snapshots, initiate a Start Rollback operation to restore the snapshot session to the source LUN as shown in Figure 107 on page 249.

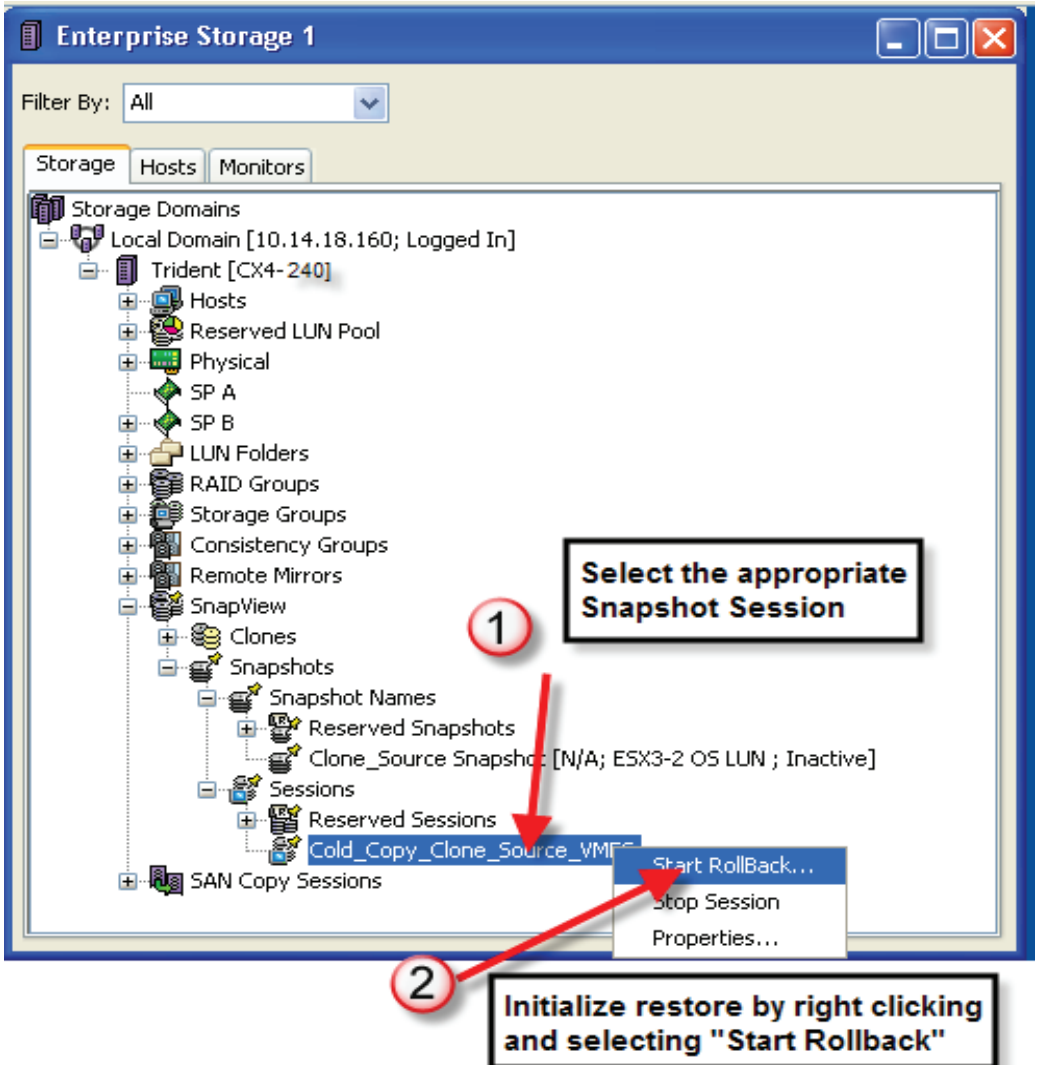


Figure 107 Using Navisphere Manager to restore a production volume from a SnapView snapshot session

7. The SCSI bus should be rescanned using the `esxcfg-rescan` command on the service console, or the Virtual Infrastructure client.
8. The restored production volume is recognized as a snap volume and resignatured and relabeled.
9. The resignatured volume can be relabeled back to the original name if step 4 listed above has been executed properly.
10. The virtual machine information from the restored volume can be added to the vCenter inventory using the service console or the vCenter client. Figure 107 on page 249 shows an example for this.
11. The clone volumes should be fractured from the production volumes if SnapView clones technology was used to create the original copy.

Note: When Replication Manager is used to back up and restore virtual disks on VMFS datastores, steps 6, 7 and 8 are automated during the restore process.

In a Virtual Infrastructure 3 environment, if you deselect the parameter **LVM.DisallowSnapshotLun** to access the copy created by a SnapView product, you must perform the following steps to restore all virtual machines hosted on a VMFS-3 volume:

1. Shut down all virtual machines accessing the production volumes. You can do this using the Virtual Infrastructure client or the service console.
2. Shut down all cloned virtual machines that are accessing the target volumes.
3. Remove all virtual machines that are accessing the production volumes from the inventory.
4. Remove the datastore associated with the production volumes that you wish to restore from the vCenter infrastructure.
5. Ensure the advanced configuration parameter, **`LVM.DisallowSnapshotLun`** is set to 0 on the entire production VMware ESX/ESXi cluster group.
6. Restore the data from the target volumes to the production volumes by following these steps:
   a. If the copy was created using SnapView clones, initiate a **Reverse Synchronization** operation from the clone to the source volume.

a. If the copy was created using SnapView snapshots, initiate a **Start Rollback** operation to restore the snapshot session to the source LUN

After executing these steps, the restored volume will have the same name and label as the previous production volume.

7. Add the virtual machine information from the restored volume to the vCenter inventory using the service console or the vCenter client.

Restoring all virtual machines hosted on VMware file system with ESX version 4

VMware ESX version 4 also provides access to copies of VMware file system if you select either the **Keep the signature** or **Assign a new signature** option for an individual LUN copy. The behavior of VMware ESX/ESXi host version 4.x when **keep the signature** is selected is similar to that of VMware ESX 3.x when the **LVM.DisallowSnapshotLUN** parameter is set to 0. Similarly, the behavior when the **ILVM.EnableResignature** parameter is set to 1 in ESX 3.x environments is similar to when the **Assign a new signature** option is selected in ESX 4.x. *The difference is in ESX 4.x where selective resignaturing is available at the individual LUN level.*

In a vSphere 4 environment, if you use the **Keep existing signature** option to access a copy created with a SnapView product, you can restore all virtual machines hosted on a VMFS-3 volume with the process described in "Restoring all virtual machines hosted on VMware file system in ESX version 3," on page 247.

Restoring all virtual machines on a VMFS-3 volume using a copy created by SnapView is problematic if the target volumes are accessed by selecting the **Assign a new signature** option, because the signature and label of the VMware file system on the target device are updated when the Assign a new signature option is selected. In a VMware vSphere 4 environment, you need to use the following process when utilizing resignatured target volumes:

1. Shut down all virtual machines accessing the production volumes. You can do this with the Virtual Infrastructure client or the service console.
2. Shut down all cloned virtual machines accessing the target volumes.

3. Remove the datastore associated with the production volumes that are being restored from the vCenter infrastructure.
4. Restore the data from the target volumes to the production volumes by following these steps:
   a. If the copy was created using SnapView clones, initiate a Reverse Synchronization operation from the clone to the source volume.
   b. If the copy was created using SnapView snapshots, initiate a Start Rollback operation to restore the snapshot session to the source LUN.
5. Rescan the SCSI bus using the esxcfg-rescan command on the service console, or the Virtual Infrastructure client.
6. The restored production volume will be displayed in the Add storage wizard in vSphere Client. At this point. select the **Assign a new signature** option for that restored production volume.
7. The restored production volume is recognized as a snap volume and is resignatured.
8. The resignatured volume can be relabeled back to the original VMFS datastore name if step 3 (listed above) has been executed properly.
9. The virtual machine information from the restored volume can be added to the vCenter inventory using the service console or the vCenter client.
10. The clone volumes should be fractured from the production volumes if SnapView clones technology was used to create the original copy.

Note: Replication Manager can also be used to back up and restore virtual disks on VMFS datastores; in this case steps 5, 6 and 7 are automated during the restore process with Replication Manager.

SnapView copies for VMs with RDMs

Virtual machines when configured with RDMs have exclusive access to the storage device. The virtual machines are similar to physical servers in this configuration. The restoration process, therefore, is similar to the one used when restoring data on a physical server. The following steps should be used:

1. The production virtual machine that needs to be restored should be powered off.
2. The cloned virtual machine accessing the cloned devices should be powered off.
3. Restore the data from the target volumes to the production volumes using the appropriate process:
   a. If the copy was created using SnapView clones, initiate a Reverse Synchronization operation from the clone to the source volume.
   b. If the copy was created using SnapView snapshots, initiate a Start Rollback operation to restore the snapshot session to the source LUN.
4. Power on the production virtual machines as the restore continues in the background.
5. The clone volumes should be fractured from the production volumes if SnapView clones technology was used to create the original copy.

Note: If Replication Manager is used to back up RDM, steps 3 and 6 are automated during the restore process.

Using backup-to-disk copies

Virtual machines backed up to an EXT3 file system using the service console can be restored quickly. The backup of the virtual disks associated with a virtual machine could have been performed by configuring the target VMware ESX/ESXi host as a storage node or by using VMware provided utilities. The process to restore from such backups depends on the software used to create the backups. The next few sections discuss the procedures to restore virtual machines using disk-based backups.

Restoring virtual machine disks using third-party backup software

Backups to disk created using third-party backup software such as NetWorker can restore virtual machines efficiently. The virtual disks are treated as a monolithic file by the backup software. Therefore, the restore of virtual disks associated with a virtual machine is similar to restoring a file on a physical server. The process to perform the restore depends on the backup software, and is beyond the scope of this document. The readers should consult their backup software vendor to

develop a backup and restore strategy that makes effective use of the second tier disk offering from EMC for backing up and restoring Virtual Infrastructure data.

Restoring virtual machines using VMware ESX/ESXi 3 utilities

Virtual Infrastructure version 3 includes a new paradigm for backing up virtual machines. VMware Consolidated Backup (VCB), enables off-host backup of virtual machines running Microsoft Windows operating systems, thus eliminating backup load from VMware ESX/ESXi hosts. Furthermore, the product includes integration modules jointly developed with major backup vendors. The integration modules allow virtual machines to be backed up on a Windows proxy server using techniques similar to those deployed for physical servers.

VMware Consolidate Backup also provides command-line utilities to back up virtual machines using the service console. The backups created using the VCB utility, `vcbMounter`, can be restored using either `vcbRestore` or `vcbResAll`. Restores of individual virtual machines can be performed using `vcbRestore` whereas, `vcbResAll` is appropriate to restore all virtual machines from a specified directory. As discussed in section Using the ESX Server version 3.x service console, Virtual Infrastructure 3 provides a tightly integrated set of products that provide a secure mechanism to back up and restore virtual machines.

The following steps should be used to restore an individual virtual machine using `vcbRestore` utility:

1. The virtual machines that need to be restored should be powered off using the Virtual Infrastructure client or the service console.
2. Create a new catalog file if the virtual disks are restored to a different datastore from the one used to create the backup. This step is required if the backup was performed using a VMware file system resignatured and relabeled using the `LVM.EnableResignature` parameter.
3. The file system that holds the backups created using `vcbMounter` should be mounted on the production VMware ESX/ESXi host performing the restore. This is possible only if the devices containing the file system is accessible on the production VMware ESX/ESXi cluster group and is not mounted on the target VMware ESX/ESXi host. Figure 108 on page 255 shows the restoration of the virtual machine, Virtual_Machine_3x_1, from a backup created on the production VMware ESX/ESXi cluster group by resignaturing the target devices. The restore utility, `vcbRestore`, has to be

provided with an updated catalog pointing to the production datastore. The -a option, as shown in Figure 108 on page 255, provides the updated catalog.

The restore of the virtual machine can also be performed by denoting the target VMware ESX/ESXi host as the archive server to the vcbRestore utility. This technique is required if the low-cost storage devices holding the backups cannot be presented to the production VMware ESX/ESXi cluster, or if the file system holding the backups is mounted on the target VMware ESX/ESXi host. Performing restores using a remote archive server is inefficient since it uses the IP network to transfer the virtual disks from the archive server to the production VMware ESX/ESXi cluster group.

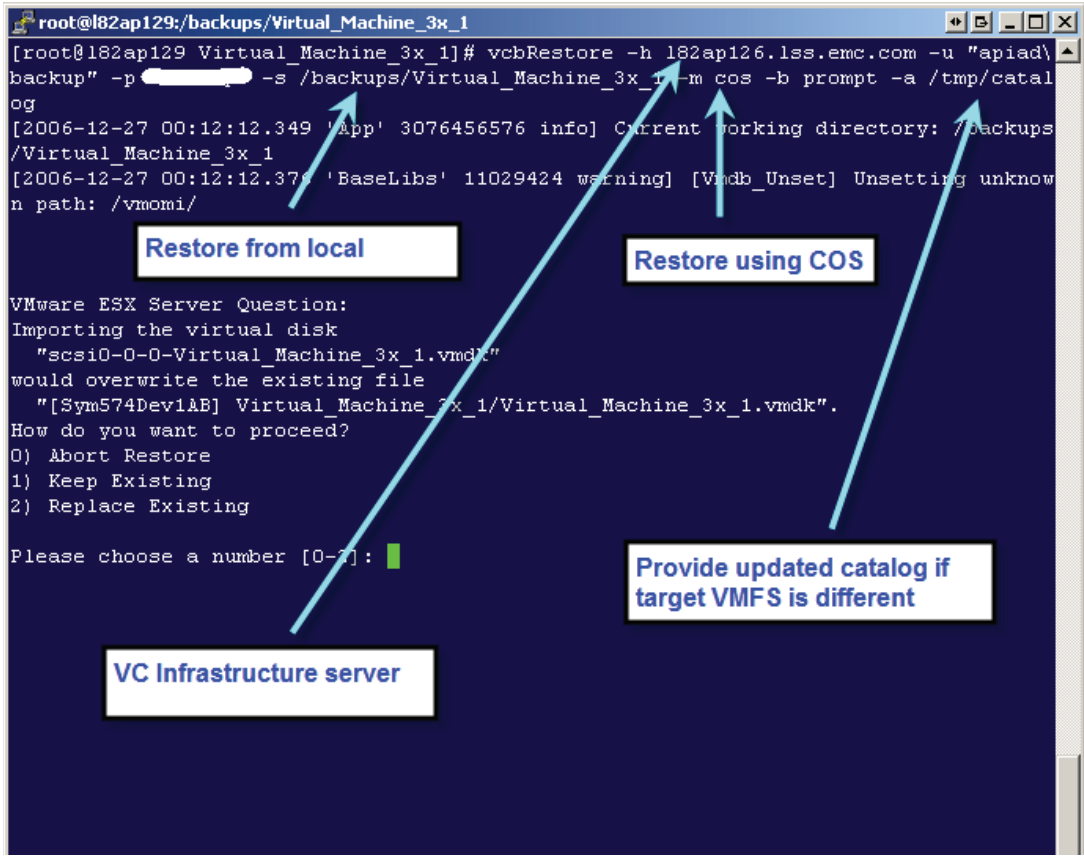


Figure 108 Restoring a VM in a VMware ESX/ESXi hosts 3 environment using vcbRestore

Restoring virtual machines using VMware ESX/ESXi 4.x utilities

VMware vSphere includes a new paradigm for backing up and restoring virtual machines. VMware Data Recovery allows you to perform disk-based backup of virtual machines directly through vCenter. Furthermore, the product utilizes built-in data deduplication technology to save significant disk space. You can restore virtual machines using a vclient connected to a vCenter server.

The following steps should be used to restore an individual virtual machine using the VMware Data Recovery utility:

1. Power off the virtual machines that need to be restored using the Virtual Infrastructure client or the service console.
2. Select the **Restore** tab with VMware Data Recovery and click **Restore** to perform a restore of the virtual machine as shown in Figure 109 on page 257.
3. The Restore Wizard will appear, specify the source of restore (if not already selected) and specify how the restored virtual machine will be configured. Click **Finish** and the restore process with begin.

The restore process with VMware Data Recovery does not require the SnapView copy to be available on the ESX server, VMware Data Recovery can restore the virtual machine directly to the production volume. The ***VMware Data Recovery Administration Guide*** available at the VMware website provides further details.

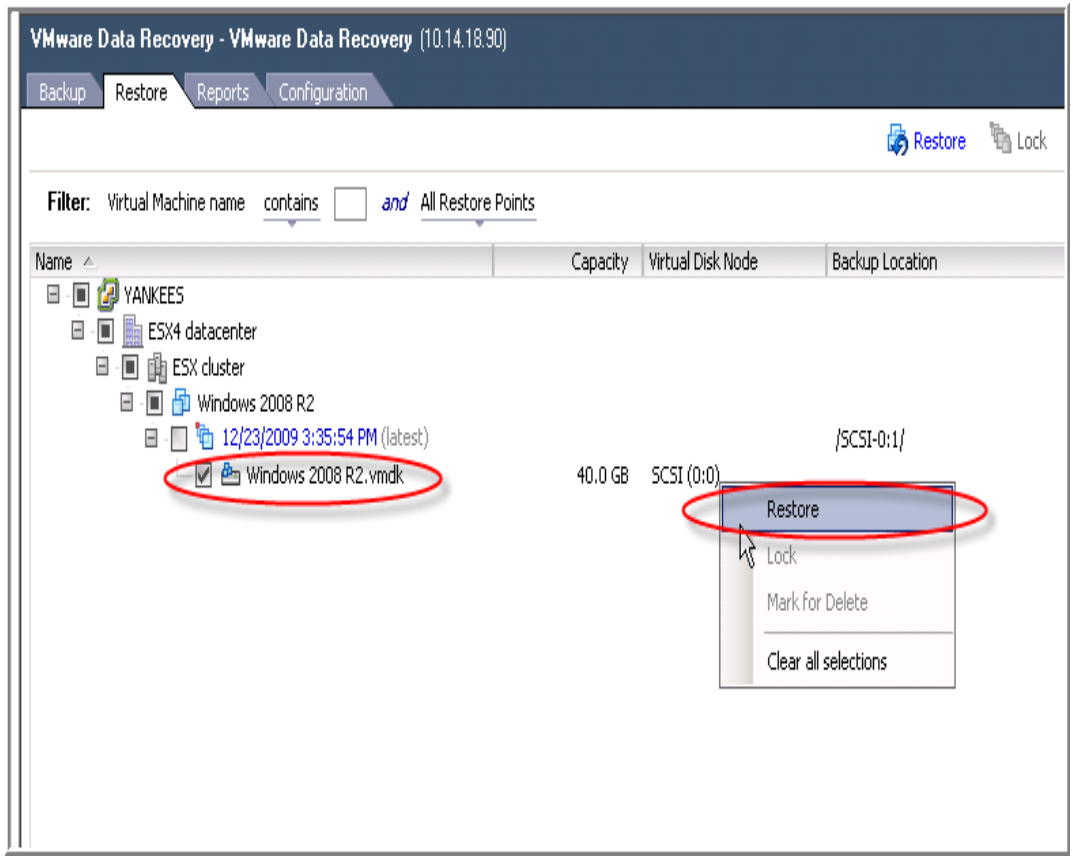


Figure 109 Restore an individual virtual machine using VMware Data Recovery

In VMware ESX/ESXi version 3.x environments, the VCB software provides an excellent mechanism to offload backups from production VMware ESX/ESXi clusters to a proxy host running Microsoft Windows. The software should be leveraged in situations where the offload of backup activities to a dedicated VMware ESX/ESXi cluster group cannot be justified. The proxy server provides integration to various backup software, like EMC NetWorker, that provide the capability of performing file-level restores using a backup client. In VMware vSphere 4x environments, VMware Data Recovery consists of a FLR (File levelRestore) option. FLR addresses these issues by providing a way to access individual files within restore points for Windows virtual machines.

The detailed architecture and processes required to perform restores of individual files using a backup software client,VMware Consolidated Backup, or VMware Data Recovery are beyond the scope of this document.

6
VMware ESX/ESXi in a Disaster Restart Solution

This chapter presents the following topics:

- Integration of guest operating environments with EMC technologies and VMware ESX/ESXi Definitions 261
- Design considerations for disaster recovery 263
- Protecting physical infrastructure with Virtual Infrastructure..... 269
- Business continuity with virtual to virtual infrastructure..... 272

VMware technology virtualizes the x86-based physical infrastructure into a pool of resources. Virtual machines are presented with a virtual hardware environment independent of the underlying physical hardware. This enables organizations to leverage disparate physical hardware in the environment and provide low total cost of ownership.

The virtualization of the physical hardware can also be used to create disaster recovery and business continuity solutions that would have been impractical otherwise. These solutions normally involve a combination of virtual infrastructure at one or more geographically separated data centers and EMC remote replication technology. One example of such architecture has physical servers running various business applications in their primary data center while the secondary data center has limited number of virtualized physical servers. During normal operations, the physical servers in the secondary data center are used for supporting workload such as QA and testing. In case of a disruption in services at the primary data center, the physical servers in the secondary data center run the business applications in a virtualized environment.

The purpose of this chapter is to discuss the following:

- EMC MirrorView configurations and their interaction with VMware ESX/ESXiEMC MirrorView
- VMware ESX/ESXi application-specific considerations

Integration of guest operating environments with EMC technologies and VMware ESX/ESXi definitions

In the next sections, the terms dependent-write consistency, disaster restart, disaster recovery, and roll-forward recovery are used. A clear definition of these terms is required to understand the context of this section.

Dependent-write consistency

A dependent-write I/O is one that cannot be issued until a related predecessor I/O has completed. Dependent-write consistency is a data state where data integrity is guaranteed by dependent-write I/Os embedded in application logic. Database management systems are good examples of the practice of dependent-write consistency.

Database management systems must devise protection against abnormal termination to successfully recover from one. The most common technique used is to guarantee that a dependent write cannot be issued until a predecessor write is complete. Typically, the dependent write is a data or index write, while the predecessor write is a write to the log.

Because the write to the log must be completed before issuing the dependent write, the application thread is synchronous to the log write—it waits for that write to complete before continuing. The result of this kind of strategy is a dependent-write consistent database.

Disaster restart

Disaster restart involves the implicit application of active logs by various databases and applications during their normal initialization process to ensure a transactionally consistent data state.

If a database or application is shut down normally, the process of getting to a point of consistency during restart requires minimal work. If the database or application abnormally terminates, then the restart process takes longer, depending on the number and size of in-flight transactions at the time of termination. An image of the database or application created by using EMC consistency technology while it is running, without any conditioning of the database or application, is in a dependent-write consistent data state, which is similar to that created by a local power failure. This is also known as a *restartable* image. The

restart of this image transforms it to a transactionally consistent data state by completing committed transactions and rolling back uncommitted transactions during the normal initialization process.

Disaster recovery

Disaster recovery is the process of rebuilding a data from a backup image, and then explicitly applying subsequent logs to roll the data state forward to a designated point of consistency. The mechanism to create recoverable copies of the data depends on the database and applications.

Roll-forward recovery

With some databases, it may be possible to take a DBMS restartable image of the database, and apply subsequent archive logs, to roll forward the database to a point in time after the image was created. This means the image created can be used in a backup strategy in combination with archive logs.

Design considerations for disaster recovery

The effect of loss of data or loss of application availability varies from one business type to another. For instance, the loss of transactions for a bank could cost millions of dollars, whereas system downtime may not have a major fiscal impact. In contrast, businesses primarily engaged in web commerce require nonstop application availability to survive. The two factors, loss of data and availability, are the business drivers that determine the baseline requirements for a disaster restart or disaster recovery solution. When quantified, loss of data is more frequently referred to as recovery point objective (RPO), while loss of uptime is known as recovery time objective (RTO).

When evaluating a solution, the RPO and RTO requirements of the business need to be met. In addition, the solution's operational complexity, cost, and its ability to return the entire business to a point of consistency need to be considered. Each of these aspects is discussed in the following sections.

Recovery point objective

The RPO is a point of consistency to which a user wants to recover or restart. It is measured in the amount of time from when the point of consistency was created or captured to the time the disaster occurred. This time equates to the acceptable amount of data loss. Zero data loss (no loss of committed transactions from the time of the disaster) is the ideal goal, but the potentially high cost of implementing such a solution must be weighed against the business impact and cost of a controlled data loss.

Some organizations, like banks, have zero data loss requirements. The transactions entered at one location must be replicated immediately to another location. This can affect application performance when the two locations are far apart. On the other hand, keeping the two locations close to one another might not protect against a regional disaster, such as a typhoon or earthquake.

Defining the required RPO is usually a compromise between the needs of the business, the cost of the solution, and the probability of a particular event happening.

Recovery time objective

The RTO is the maximum amount of time allowed after the declaration of a disaster for recovery or restart to a specified point of consistency. This includes the time taken to:

- Provision power and utilities
- Provision servers with the appropriate software
- Configure the network
- Restore the data at the new site
- Roll forward the data to a known point of consistency
- Validate the data

Some delays can be reduced or eliminated by proactively completing certain tasks *before* disaster strikes, such as having a hot site where servers are preconfigured and on standby. Furthermore, the time that it takes to restore data in a tape restore operation is completely eliminated with storage-based replication.

As with RPO, each solution with varying RTO has a different cost profile. Defining the RTO is usually a compromise between the cost of the solution and the cost to the business when applications are unavailable.

Operational complexity

The operational complexity of a DR solution may be the most critical factor that determines the success or failure of a DR activity. The complexity of a DR solution can be considered as three separate phases.

1. Initial setup of the implementation.
2. Maintenance and management of the running solution.
3. Execution of the DR plan in the event of a disaster.

While initial configuration complexity and running complexity can be a demand on people resources, the third phase—execution of the plan—is where automation and simplicity must be the focus. a disaster is declared, key personnel may be unavailable in addition to the loss of servers, storage, networks, and buildings. If the DR solution is so complex that it requires skilled personnel (with an intimate knowledge of all systems involved) to restore, recover, and validate application and database services, the solution is much more complex and difficult. VMware Site Recovery helps automate the execution of a DR test plan for simple and complex VMware environments.

Multiple database and application environments over time grow organically into complex federated database architectures. In these federated environments, reducing the complexity of DR is critical. Validation of transactional consistency within a business process is

time-consuming, costly, and requires application and database familiarity. One of the reasons for this complexity is the heterogeneous applications, databases and operating systems in these federated environments. Across multiple heterogeneous platforms, it is hard to establish time synchronization, and therefore hard to determine a business point of consistency across all platforms. This business point of consistency has to be created from intimate knowledge of the transactions and data flows.

Source server activity

DR solutions might require additional processing activity on the source servers. The extent of that activity can impact both response time and throughput of the production application. This effect would be true with host or array based replication, hence the additional processing needs to be understood and quantified for any given solution to ensure the impact to the business is minimized. The effect for some solutions is continuous while the production application is running; for other solutions, the impact is sporadic, where bursts of write activity are followed by periods of inactivity.

Production impact

Some DR solutions delay the host activity while taking actions to propagate the changed data to another location. This action only affects write activity. Although the introduced delay may only be of the order of a few milliseconds it can negatively impact response time in a high-write environment. Synchronous solutions introduce a delay into write transactions at the source site; asynchronous solutions can also introduce a delay during an update.

Target server activity

Some DR solutions require a target server at the remote location to perform DR operations. The server has both software and hardware costs and needs personnel with physical access to it for basic operational functions such as power on and power off. Ideally, this server could have some usage such as running development or test databases and applications. Some DR solutions require more target server activity and some require none.

Number of copies of data

DR solutions require replication of data in one form or another. Replication of application data and associated files can be as simple as making a tape backup and shipping the tapes to a DR site or as sophisticated as asynchronous array-based replication. Some solutions require multiple copies of the data to support DR functions. More copies of the data may be required to perform testing of the DR solution in addition to those that support the data replication process.

Distance for the solution

Disasters, when they occur, have differing ranges of impact. For example:

- A fire may be isolated to a small area of the data center or a building.
- An earthquake may destroy a city.
- A hurricane may devastate a region.

The level of protection for a DR solution should address the probable disasters for a given location. This means for protection against an earthquake, the DR site should not be in the same locale as the production site. For regional protection, the two sites need to be in two different regions. The distance associated with the DR solution affects the kind of DR solution that can be implemented.

Bandwidth requirements

One of the largest costs for DR is in provisioning bandwidth for the solution. Bandwidth costs are an operational expense; this makes solutions with reduced bandwidth requirements attractive to customers. It is important to recognize in advance the bandwidth consumption of a given solution to anticipate the running costs. Incorrect provisioning of bandwidth for DR solutions can adversely affect production performance and invalidate the overall solution.

Federated consistency

Databases are rarely isolated islands of information with no interaction or integration with other applications or databases. Most commonly, databases are loosely or tightly coupled to other databases and applications using triggers, database links, and stored procedures. Some databases provide information downstream for other databases and application using information distribution middleware; other

applications and databases receive feeds and inbound data from message queues and EDI transactions. The result can be a complex, interwoven architecture with multiple interrelationships. This is referred to as federated architecture. With federated environments, making a DR copy of a single database regardless of other components invites consistency issues and creates logical data integrity problems. All components in a federated architecture need to be recovered or restarted to the same dependent-write consistent point in time to avoid data consistency problems.

With this in mind, it is possible that point solutions for DR, like host-based replication software, do not provide the required business point of consistency in federated environments. Federated consistency solutions guarantee that all components, databases, applications, middleware, and flat files are recovered or restarted to the same dependent-write consistent point in time.

Testing the solution

A DR solution also requires tested, proven, and documented procedures. Often, the DR test procedures are operationally different from a true disaster set of procedures. Operational procedures need to be clearly documented. In the best-case scenario, companies should periodically execute the actual set of procedures for DR. This could be costly to the business because of the application downtime required to perform such a test, but is necessary to ensure validity of the DR solution.

Cost

The cost of doing DR can be justified by comparing it with the cost of not doing it. What does it cost the business when the database and application systems are unavailable to users? For some companies this is easily measurable, and revenue loss can be calculated per hour of downtime or data loss.

Whatever the business, the DR cost is going to be an additional expense item and, in many cases, with little in return. The costs include, but are not limited to:

- Hardware (storage, servers, and maintenance)
- Software licenses and maintenance

- Facility leasing or purchase
- Utilities
- Network infrastructure
- Personnel

Protecting physical infrastructure with Virtual Infrastructure

Figure 108 on page 270 is a schematic representation of the business-continuity solution that integrates physical infrastructure with virtual infrastructure using CLARiiON MirrorView technology.

The physical infrastructure at the production site can be replicated using MirrorView (using FC or iSCSI protocols); the replica can be presented to a virtual machine on a VMware ESX/ESXi host at the remote site in case of disaster.

The LUNs containing application data on a physical server do not need reconfiguration during the failover or failback process and are thus supported for replication with MirrorView. The replica presented to the virtual machine must be configured as an RDM volume. Ensure the correct application data volumes are presented and assigned to the VMware ESX/ESXi hosts and virtual machines, respectively. This requires proper mapping of the CLARiiON device numbers on the target CLARiiON storage system. Replicating physical machine OS images for conversion to virtual machines on a VMware ESX/ESXi host is currently not supported with MirrorView.

To maintain data consistency for dependent write-order LUNs, EMC recommends using the MirrorView or consistency group feature with the solution. The secondary images on the target CLARiiON storage system are normally presented as read and write disabled, and hence cannot be seen by the VMware ESX/ESXi host unless those images are promoted. Copies of the application data can be obtained on the remote site by replicating secondary images using SnapView snapshots and clones. These copies can be used for ancillary operation processes such as QA or backup. The snapshots must be in an activated state before presenting them to the ESX Server.

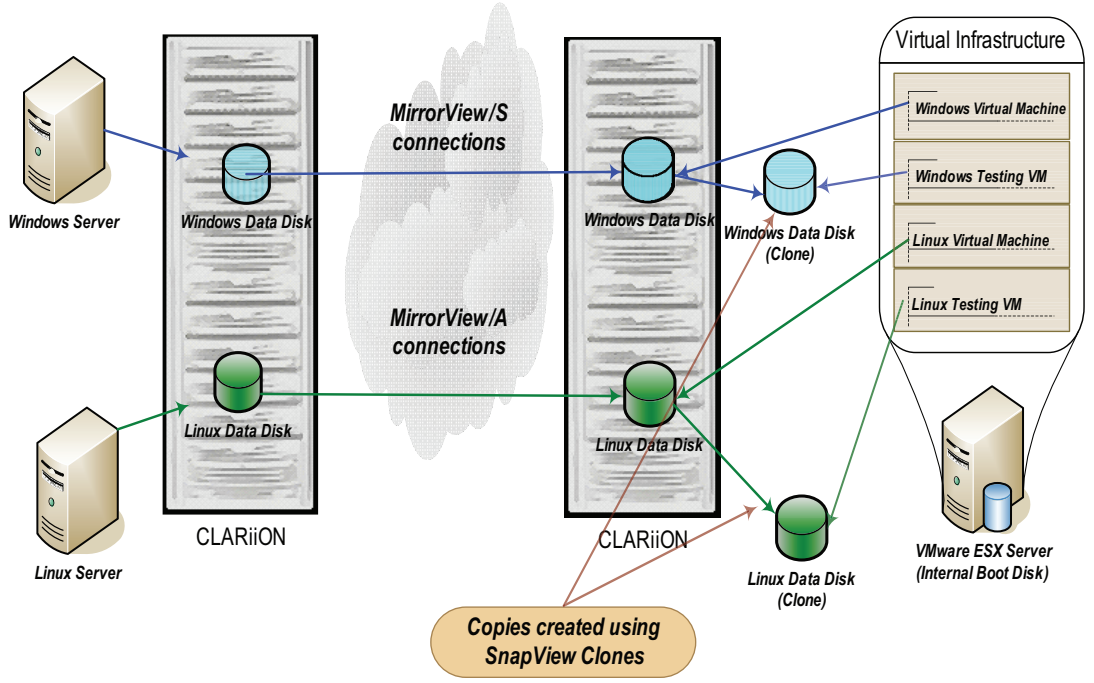


Figure 108 Using MirrorView and Virtual Infrastructure for a business continuity solution

Physical-to-Virtual Infrastructure

Each virtual machine boot disks image on the remote site must be created manually to match the configuration of the physical machine boot disk on the primary site. In addition, the most important consideration in the solution previously presented is to ensure the correct data volumes are presented and assigned to the VMware ESX/ESXi hosts and the virtual machines, respectively. This requires proper mapping of the CLARiiON device numbers on the secondary images to the canonical name assigned by the VMkernel.

After the correct mapping has been determined, it is important to preserve the ordering of the disks presented to the virtual machines on the remote site. For example, consider a physical server on the production site with its three application data disks, \\.\PHYSICALDRIVE2, \\.\PHYSICALDRIVE3, and

\\.\PHYSICALDRIVE4 that correspond to CLARiiON LUNs numbers 2, 3, and 4 respectively. These three CLARiiON LUNs are replicated on the remote CLARiiON LUNs 2, 3, and 4 respectively. Then, the virtual machine that already consists of a boot image disk configured as SCSI target 0:0 on the remote site should be presented on three RDM disks, that is, SCSI disks 0:1, 0:2, and 0:3, respectively.

Remotely managing application data LUNs

Data devices require no reconfiguration. To maintain data consistency, EMC recommends the use of the EMC consistency technology with the solutions. Furthermore, a good disaster recovery plan involves frequent testing. For the disaster restart solutions using MirrorView, a copy of the application data can be made at the remote site by using the SnapView family. The copy of the data can be presented to a virtual machine for testing.

It is ideal to have all virtual machines on the remote site powered off before testing or failing over production operations to the remote site. In addition, for the virtual machines used for test, ensure the guest operating system does not maintain any cache for the copies of the data if they are created on a regular basis. This can be easily done by powering off the virtual machine after the testing is complete.

The MirrorView secondary images are presented as not-ready on the host channel. VMware ESX version 4, 3 and VMware ESXi allow devices in a not-ready state to be assigned to virtual machines as RDM devices. This capability allows virtual machines to be preconfigured with secondary images even when there is active replication of data from the source site. However, to prevent potential issues, EMC recommends leaving all virtual machines in a powered off state until needed.

Business continuity with virtual to virtual infrastructure

The business continuity solution for a production environment with VMware virtual infrastructure is much simpler than the solutions discussed in "Protecting physical infrastructure with Virtual Infrastructure," on page 269. In addition to a tape-based disaster recovery solution, EMC MirrorView can be used as the mechanism to replicate data from the production data center to the remote data center. The copy of the data in the remote data center can be presented to a VMware ESX/ESXi version 4 or 3 cluster group. The Virtual Infrastructure at the remote data center thus provides a business continuity solution.

Tape-based solutions

Tape-based disaster recovery

Traditionally, the most common form of disaster recovery was to make a copy of the database onto a tape, and take the tapes offsite to a hardened facility. In most cases, the database and application needed to be available to users during the backup process. Taking a backup of a running database or application created a fuzzy image of the data on tape, one that required recovery processes after the image was restored. Recovery usually involves application of logs active during the time the backup was in process. These logs had to be archived and kept with the backup image to ensure successful recovery.

With the rapid growth of data over the last two decades, this method has become unmanageable. Making a hot copy of the database is now the standard—but this method has its own challenges. How can a consistent copy of the application data and supporting files be made when they are changing throughout the duration of the backup? What exactly is the content of the tape backup at completion? The reality is that the tape data is a fuzzy image of the disk data, and considerable expertise is required to restore the applications back to a point of consistency.

In addition, the challenge of returning the data to a business point of consistency, where the data from various applications and databases must be recovered to the same point in time, is making this solution less viable.

Tape-based disaster restart

Tape-based disaster restart is a recent development in disaster recovery strategies. It is used to avoid the fuzziness of a backup taken while the database and application are running. A restart copy of the system data is created by locally mirroring the disks that contain the production data, and splitting off the mirrors to create a dependent-write consistent point-in-time image of the disks. The image on the disk is a restartable image as described previously in "Integration of guest operating environments with EMC technologies and VMware ESX/ESXi Definitions," on page 261. Therefore, if this image was restored and the applications brought up, the operating system or the application would perform an implicit recovery to attain transactional consistency. Roll-forward recovery from the image is not normally possible.

The restartable image on the disks can be backed up to tape and moved offsite to a secondary facility. In a VMware ESX/ESXi environment, this can be achieved by deploying the backup methodologies discussed in "Using the ESX Server version 3.x service console," on page 227 and "Using the ESX server version 4.x service console," on page 236. If backup tapes of Virtual Infrastructure data are created and shipped offsite on a daily basis, the maximum amount of data loss could be as high as 48 hours. The amount of time required to restore the data at the remote data center in this solution is significant since restores from tape is typically slow. Consequently, this solution can be effective for customers with longer RPOs and RTOs.

MirrorView consistency groups

MirrorView (both synchronous and asynchronous) includes the storage-system-based consistency groups feature. A storage-system-based consistency group is a collection of mirrors that function together as a unit within a storage system. All operations, such as synchronization, promote, and fracture, occur on all the members of the consistency group. After a mirror is part of a consistency group, most operations on individual members are prohibited. This is to ensure that operations are automatically performed across all the member mirrors.

The members of a consistency group can span across the storage processors on the CLARiiON storage system, but all of the member mirrors must be on the same storage system. A consistency group cannot have mirrors that span across the storage systems. In addition, although consistency groups are supported for both synchronous and

asynchronous mirrors, all components of a consistency group must be protected by same mode of replication (either synchronous or asynchronous).

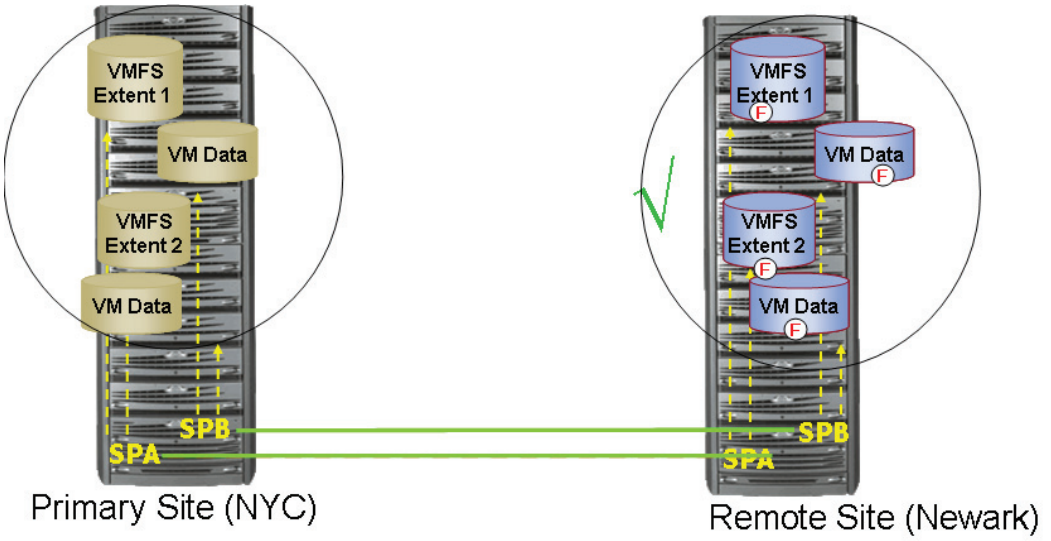


Figure 109 Preserving dependent-write consistency with MirrorView consistency group technology

Figure 109 on page 274 shows an example of a configuration where it is critical to use consistency groups. Consistency groups ensure that if one member of the group is fractured for any reason, then all of the members of the group fracture, and data integrity is preserved across the set of secondary images.

In the example depicted in Figure 109 on page 274, due to communication failure between SP-A of the two arrays, the LUNs replicated by that storage processor become fractured. At the point of disruption, MirrorView fractures the rest of the mirrors in the consistency group. After the secondary images are fractured, updates to the primary volumes are not propagated to the secondary volumes thus preserving the consistency of the data. While MirrorView performs the fracture operation, it briefly holds write I/Os to members of the consistency group until that particular member is fractured. After each corresponding member is fractured, I/O is allowed to continue to that volume.

MirrorView/S from CLARiiON to CLARiiON

Figure 110 on page 276 is a schematic representation of the business continuity solution that integrates VMware virtual infrastructure and MirrorView technology. The solution shows two virtual machines accessing LUNs on the CLARiiON storage arrays as RDMs. An equivalent solution utilizing the VMware file system is depicted in Figure 111 on page 279. The proposed solution provides an excellent opportunity to consolidate the virtual infrastructure at the remote site. It is possible to run VMware virtual machines on any VMware ESX/ESXi in the cluster group. This capability also allows the consolidation of the production VMware ESX/ESXi to fewer VMware ESX/ESXi hosts. However, by doing so, there is a potential for duplicate virtual machine IDs when multiple virtual machines are consolidated in the remote site. If this occurs, the virtual machine IDs can be easily changed at the remote site.

MirrorView/S can be used for replicating production data changes from locations less than 200 km apart round-trip. Synchronous mode replicates writes to the source CLARiiON LUN to the target CLARiiON LUN. The resources of the storage arrays are exclusively used for the copy. The write operation from the virtual machine is not

acknowledged until both CLARiiON arrays have a copy of the data in their cache. The ***MirrorView Knowledgebook*** available on Powerlink provides further information about MirrorView/S.

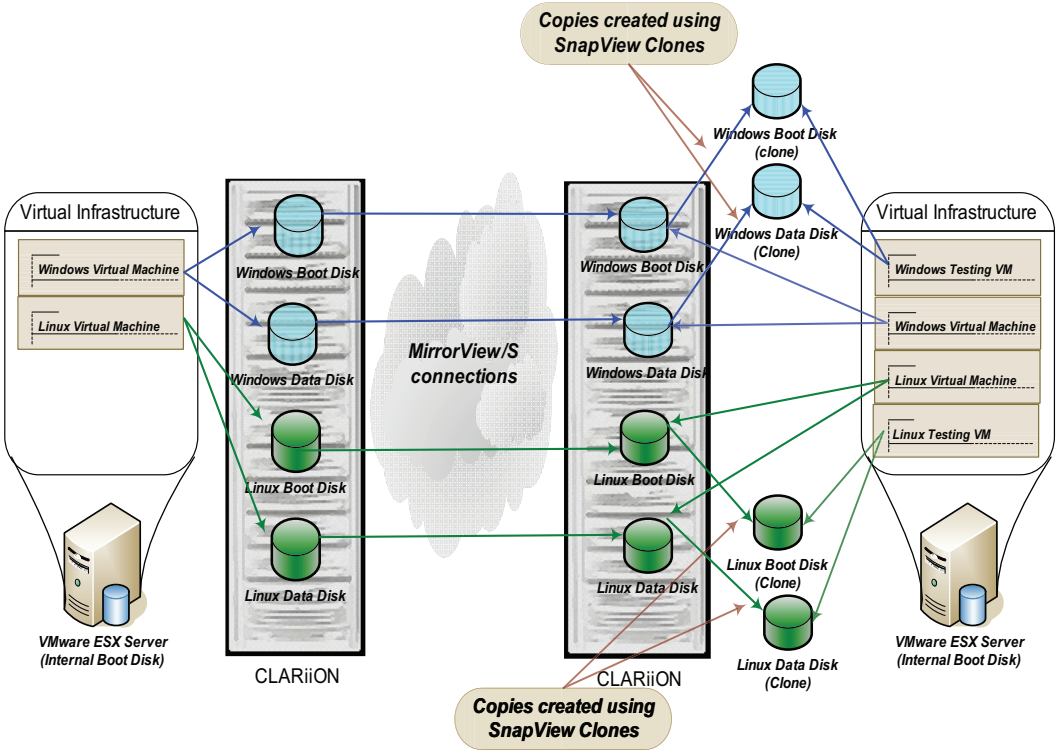


Figure 110 Business continuity solution using MirrorView/S in a virtual infrastructure using RDM

The following steps outline the process of setting up synchronous replication using Navisphere SecureCLI commands. The commands should be run from a management host, which is connected to the production CLARiiON storage array.

1. The first step in creating a disaster restart solution using MirrorView/S is to create a path for remote mirroring between the primary and secondary CLARiiON.

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

2. Identify the LUN that need to be replicated. This can be accomplished by using the techniques discussed in "Mapping a VMware file system to CLARiiON devices," on page 145 and "Using vCenter to determine the relationship between a VMFS label and canonical name," on page 150. This command can be used to create a remote mirror of the LUN:

   ```
   naviseccli -h SP ipaddress mirror -sync -create -lun
   <Lun_number>
   ```

   The LUN on which the mirror was created becomes the primary image.

3. The secondary image on the remote CLARiiON can then be added to the primary image. The following command assume that the LUN is already created on the remote CLARiiON storage system:

   ```
   naviseccli -h SPipaddress mirror -sync -addimage -name
   <name> -arrayhost <sp-hostname| sp ip-address> -lun
   <lunnumber | lunuid>
   ```

4. If multiple related LUNs are protected with MirrorView/S, the user has the option of creating a consistency group and adding the two or more mirrors to this consistency group. 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. After the images are added with or without the consistency group option, the initial synchronization between the primary and secondary images is started. If for some reason the mirrors are fractured, the syncimage option, as shown below, can be used to resynchronize the primary and secondary images:

   ```
   naviseccli -h SPipaddress mirror -sync -syncimage -name
   <name>
   ```

   If consistency groups are used, the syncgroup option can be used to synchronize all mirror images:

   ```
   naviseccli -h SPipaddress mirror -sync -syncgroup -name
   <name>
   ```

6. When the secondary images are in a synchronized or consistent state, consistent point-in-time copies of the secondary image can be created using SnapView clones or snapshots. The process to create clones is similar to the process discussed in "Using EMC to copy running virtual machines ," on page 193.

   To access the secondary image at the remote site, the images can be promoted at the DR site as:

   ```
   naviseccli –h SPipaddress mirror -sync -promoteimage
   -name <name>
   ```

7. If using consistency groups, you can use the promotegroup command to promote all mirror images:

   ```
   naviseccli –h SPipaddress mirror -sync -promotegroup
   -name <name>
   ```

The ***MirrorView/Synchronous Command Line Interface Reference*** available on Powerlink provides additional details on using Navisphere CLI with MirrorView.

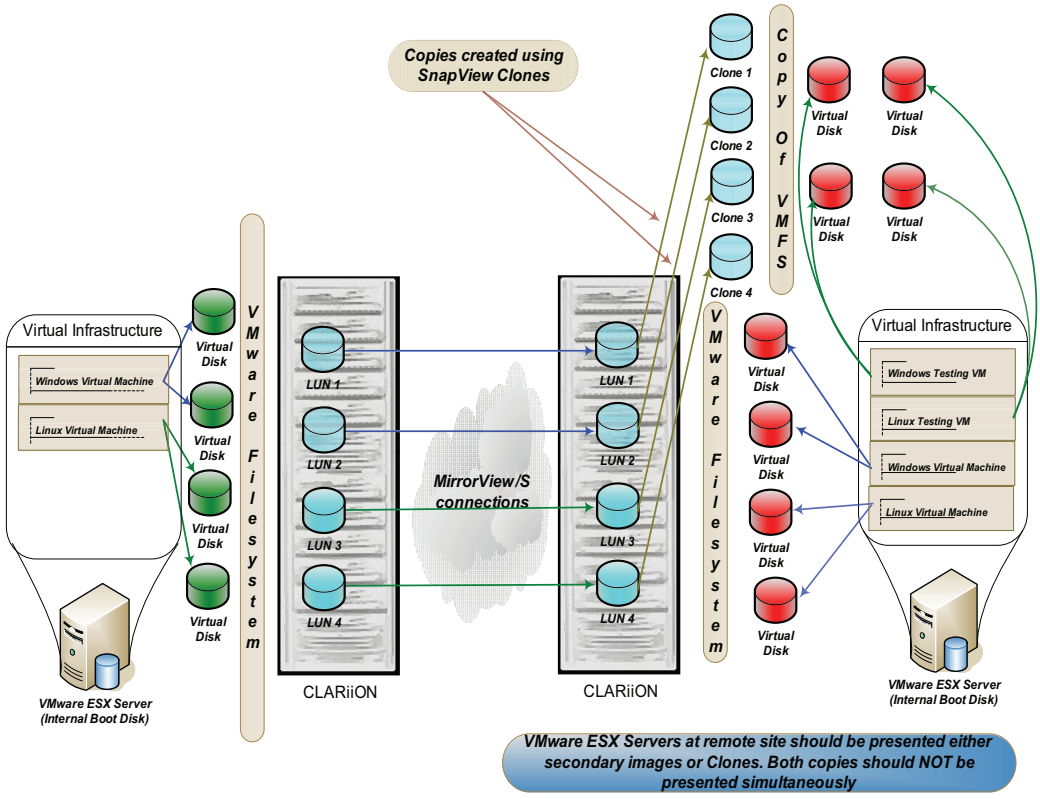


Figure 111 Business continuity solution using MirrorView/S in a virtual infrastructure with VMFS

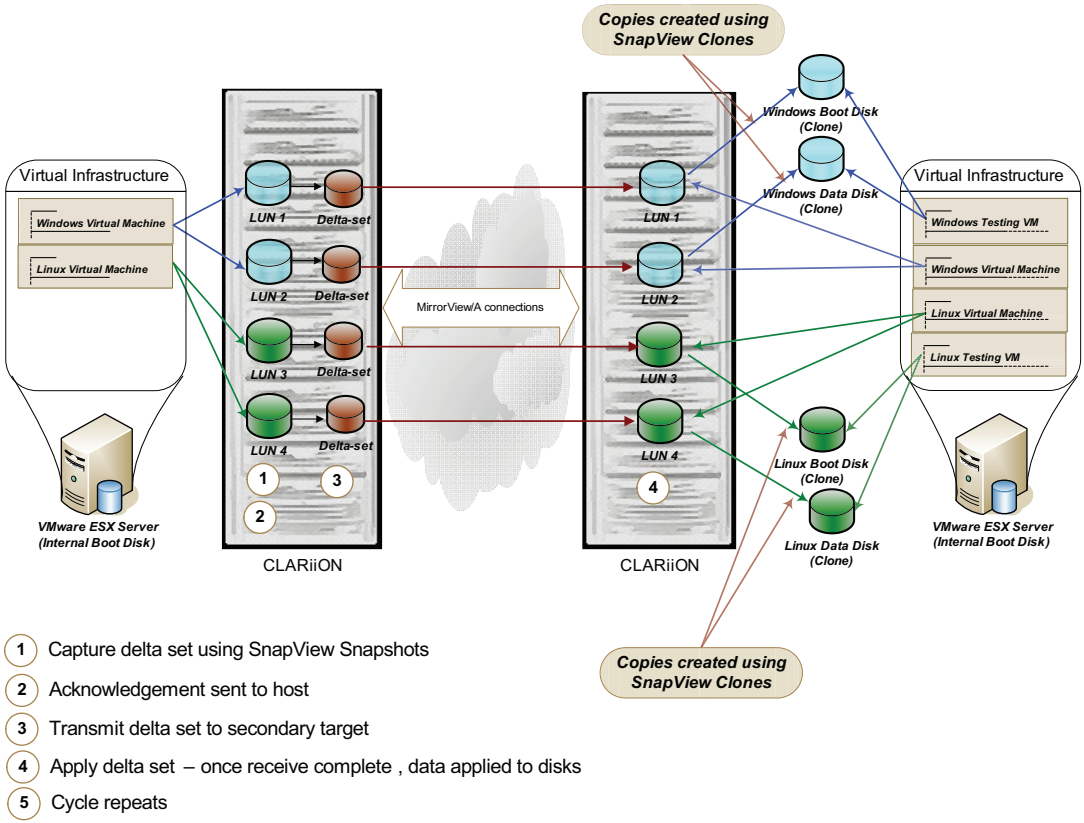
MirrorView/A from CLARiiON to CLARiiON

MirrorView/A is an asynchronous method of replicating production data changes from one CLARiiON to another. The replication is performed using a delta set technology. Delta sets are collection of changed blocks grouped together by a time interval that can be configured at the source site.

The asynchronous nature of replication implies a non-zero RPO. MirrorView/A is designed to provide customers with a RPO greater than or equal to 30 minutes. MirrorView/A replicates delta sets to

create consistent, write ordered point in-time copies of production data on the remote system. The ***MirrorView Knowledgebook*** available on Powerlink provides further information about MirrorView/A.

The distance between the source and target CLARiiON in a MirrorView/A relationship is unlimited since no acknowledgement is required. Furthermore, due to the asynchronous nature of replication there is no host impact. Writes are acknowledged immediately on the source CLARiiON. Figure 112 on page 281 shows the MirrorView/A process as applied to a VMware virtual infrastructure environment using RDM. A similar process, shown in Figure 113 on page 282, can also be used to replicate LUNs formatted using VMware file system.



1. Capture delta set using SnapView Snapshots
2. Acknowledgement sent to host
3. Transmit delta set to secondary target
4. Apply delta set – once receive complete , data applied to disks
5. Cycle repeats

Figure 112 Business continuity solution using MirrorView/A in a virtual infrastructure using RDM

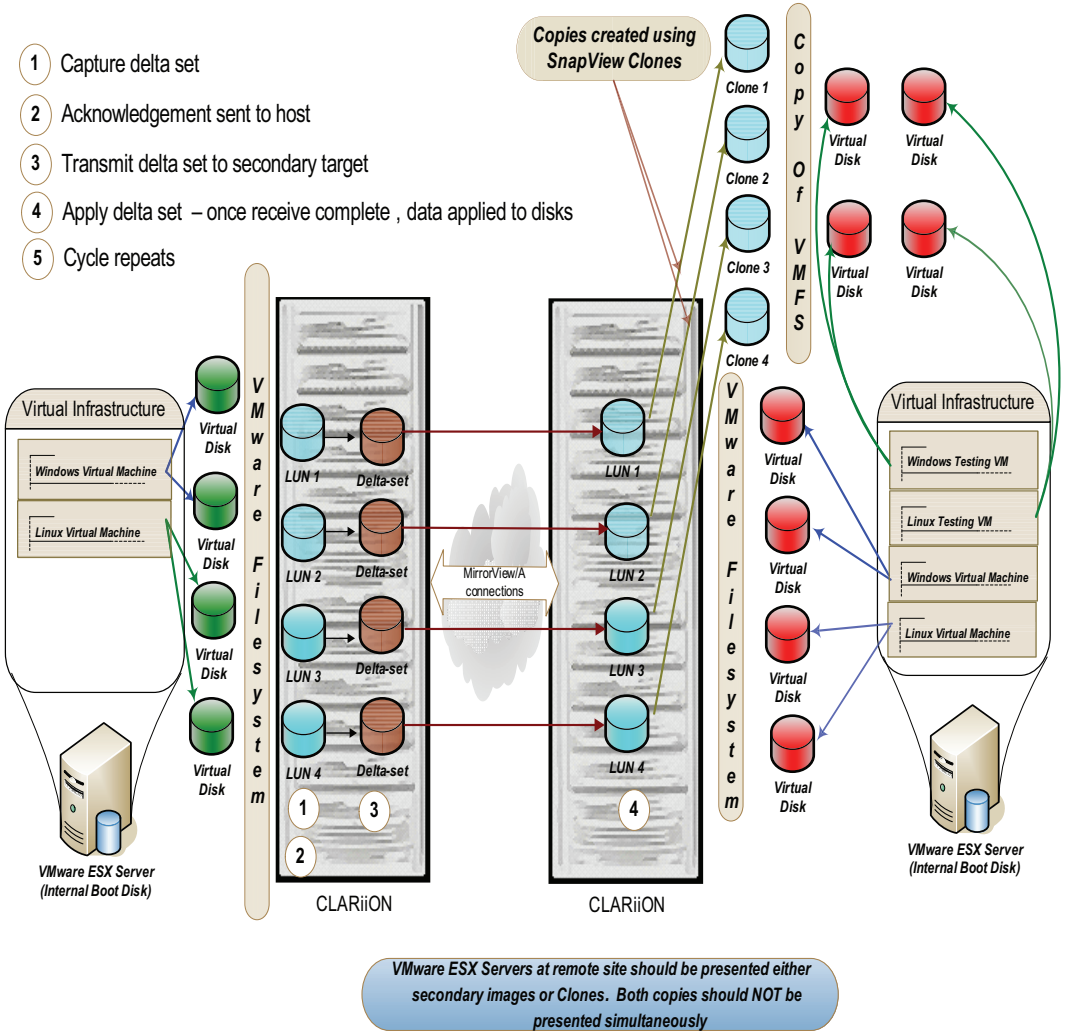


Figure 113 Business continuity solution using MirrorView/A in a virtual infrastructure with VMFS

The following steps outline the process of 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. The first step in creating a disaster restart solution using MirrorView/A is to create a path for remote mirroring between the primary and secondary CLARiiON:

   ```
   naviseccli -h SP ipaddress mirror -async -enablepath
   SPhostname
   ```

2. The next step is to identify the LUNs that need to be replicated. This can be accomplished by using the techniques discussed in "Mapping a VMware file system to CLARiiON devices," on page 145 and "Using vCenter to determine the relationship between a VMFS label and canonical name," on page 150. This command can be used to create a remote mirror of the LUN that needs to be replicated:

   ```
   naviseccli -h SP ipaddress mirror -async -create -lun
   <Lun_number>
   ```

   The LUN on which the mirror was created becomes the primary image.

3. Then, create the secondary image for replication on the remote CLARiiON storage system. The following command assume that the LUN is already created on the remote CLARiiON storage system:

   ```
   naviseccli -h SPipaddress mirror -async -addimage -name
   <name> -arrayhost <sp-hostname| sp ip-address> -lun
   <lunnumber | lunuid>
   ```

4. If multiple related LUNs are protected with MirrorView/A, the user has the option of creating a consistency group and adding the two or more mirrors to this consistency group. The following commands show how to 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> -mirrorname <mirrorname>
   ```

5. The initial synchronization is initiated as soon as the secondary image is added to the replication pair. If for some reason the mirrors are fractured, the syncimage option, as shown below, can be used to resynchronize the primary and secondary images:

   ```
   naviseccli -h SPipaddres mirror -async -syncimage -name
   <name>
   ```

If consistency groups are used, the syncgroup option can be used to synchronize all mirror images:

```
naviseccli –h SPipaddress mirror –async –syncgroup –name
<name>
```

Configuring remote sites for VMs using VMFS-3 on Virtual Infrastructure 3

VMware ESX/ESXi version 3 and VMware ESXi also assign a unique signature to all VMFS-3 volumes when they are formatted with the VMware file system. Furthermore, if the VMware file system is labeled that information is also stored on the device. The signature is generated using the unique ID (UID) of the device and the LUN number at which the device is presented.

Since storage array technologies create exact replicas of the source volumes, all information including the unique signature (and label, if applicable) is replicated. If a copy of a VMFS-3 volume is presented to the any VMware ESX/ESXi version 3 cluster, the VMware ESX/ESXi host automatically masks the copy. The device that holds the copy is determined by comparing the signature stored on the device with the computed signature. Secondary images, for example, have different unique IDs from the primary images with which it is associated. Therefore, the signature stored on the secondary image differs from the computed signature. This enables the VMware ESX/ESXi host to always identify the copy correctly.

VMware ESX version 3.x and VMware ESXi provide two different mechanisms to access copies of VMFS-3 volumes. The advanced configuration parameters, `LVM.DisallowSnapshotLun` or `LVM.EnableResignature`, control the behavior of the VMkernel when presented with copies of a VMware file system.

- If `LVM.DisallowSnapshotLun` is set to 0, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts with access to both source and target devices, the parameter has no effect since VMware ESX/ESXi host never presents a copy of the data if there are signature conflicts. The default value for this parameter is 1.
- If `LVM.EnableResignature` is set to 1, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include “snap-x”, where x is a hexadecimal number that can range

from 0x2 to 0xffffffff. The default value for this parameter is 0. If this parameter is changed to 1, the advanced parameter, `LVM.DisallowSnapShotLun`, is ignored.

When using MirrorView-based remote replication to protect production Virtual Infrastructure 3 environment, EMC recommends setting the `LVM.DisallowSnapshotLun` to 0 on VMware ESX/ESXi version 3 cluster at the remote site. The use of `LVM.EnableResignature` is strongly discouraged since it introduces complexity to the process of starting the virtual machines in case of a disaster. Furthermore, use of `LVM.EnableResignature` parameter on the VMware ESX/ESXi at the remote site makes the failback process from a planned failover event extremely difficult.

CAUTION

The `LVM.EnableResignature` parameter should not be changed for any reason on the VMware ESX/ESXi hosts at the remote site. All volumes that are considered to be copies of the original data will be resignatured if the parameter is enabled. Furthermore, there is no mechanism currently available to undo the resignaturing process. Depending on the state of the infrastructure at the remote site, the resignaturing process can cause havoc with the disaster restart environment.

The following paragraphs discuss the process to create virtual machines at the remote after changing the advanced parameter, `LVM.DisallowSnapshotLun` to 0 (see "Cloning Virtual Infrastructure 3 virtual machines using LVM.DisallowSnapshotLun ," on page 201 for the process to change the parameter).

1. The first step to create virtual machines at the remote site is to enable access to the secondary devices for the VMware ESX/ESXi cluster group at the remote data center.
2. Virtual Infrastructure 3 tightly integrates the VirtualCenter infrastructure and VMware ESX version 3 and VMware ESXi version. VirtualCenter infrastructure does not allow duplication of objects in a VirtualCenter data center. If the same VirtualCenter infrastructure is used to manage the VMware ESX/ESXi host at the production and remote site, the servers should be added to different data center constructs in VirtualCenter.
3. The SCSI bus should be scanned after ensuring the secondary images are promoted when they are in a synchronized or a consistent state. The scanning of the SCSI bus can be done either

using the service console or the VirtualCenter client. on page 291 shows an example of the process. The devices that hold the copy of the VMware file system is displayed on the VMware ESX/ESXi cluster at the remote site.

4. In a Virtual Infrastructure 3 environment, when a virtual machine is created, all files related to the virtual machine are stored in a directory on a Virtual Infrastructure 3 datastore. This includes the configuration file and, by default, the virtual disks associated with a virtual machine. Thus, the configuration files automatically get replicated with the virtual disks when storage array based copying technologies are leveraged.

5. The registration of the virtual machines from the target device can be performed using Virtual Infrastructure client or the service console. The re-registration of cloned virtual machines is not required when the configuration information of the production virtual machine changes. The changes are automatically propagated and used when needed.

   As recommended in step 2, if the VMware ESX/ESXi host at the remote site are added to a separate data center construct, the names of the virtual machines at the remote data center matches those of the production data center.

6. The virtual machines can be started on the VMware ESX/ESXi host at the remote site without any modification if the following requirements are met:

   - The target VMware ESX/ESXi host have the same virtual network switch configuration—that is, the name and number of virtual switches should be duplicated from the source VMware ESX/ESXi cluster group.
   - All VMware file systems used by the source virtual machines are replicated. Furthermore, the VMFS labels should be unique on the target VMware ESX/ESXi host.
   - The minimum memory and processor resource reservation requirements of all cloned virtual machines can be supported on the target VMware ESX/ESXi host. For example, if 10 source virtual machines, each with a memory resource reservation of 256 MB needs to be cloned, the target VMware ESX/ESXi cluster should have at least 2.5 GB of physical RAM allocated to the VMkernel.

- Virtual devices, such as CD-ROM and floppy drives, are attached to physical hardware, or are started in a disconnected state when the virtual machines are powered on.

Starting virtual machines at a remote site in the event of a disaster

The following steps should be performed at the remote site to restart virtual machines using the replicated copy of the data:

1. The first step involves promotion of the secondary images to become read-write enabled. The local or forced promote of the secondary image or the consistency group should be utilized to make the secondary LUN available for read and write operation. Ensure the secondary LUNs are in synchronized or in a consistent state before promoting the LUNs. The command to promote a secondary image participating in a MirrorView/S relationship is:

   ```
   naviseccli –h SPipaddress mirror -sync -promoteimage
   -name <name> type local
   ```

   When utilizing MirrorView/S consistency groups, the local *promotegroup* option can be used to promote all secondary images in the consistency group:

   ```
   naviseccli –h SPipaddress mirror -sync -promotegroup
   -name <name> type local|force
   ```

   For MirrorView/A, the following command can be used to promote an individual secondary image:

   ```
   naviseccli –h SPipaddress mirror -async -promoteimage
   -name <name> -type local
   ```

   The *promotegroup* option, as shown below, can be utilized to promote all secondary images in a MirrorView/A relationship:

   ```
   naviseccli –h SPipaddress mirror -async -promotegroup
   -name <name> -type local
   ```

   The MirrorView/S and MirrorView/A command line interface documentation available on Powerlink provides details on the local and force promote options.

2. Virtual machines used for ancillary business operations at the remote site should be shut down. At this point, if needed, the LUN masking on the CLARiiON storage array should be modified to provide VMware ESX/ESXi with access to the secondary images.

Any copy of the data that is being utilized for other operations should be masked away from the VMware ESX/ESXi cluster group.

3. A subsequent rescan of the SCSI bus makes the secondary images of the VMware file system on the secondary images accessible on the VMware ESX/ESXi host that runs the DR copy of the virtual machines. The VMware file system label created on the source volumes is recognized by the target VMware ESX/ESXi host. This is depicted in Figure 114 on page 288.

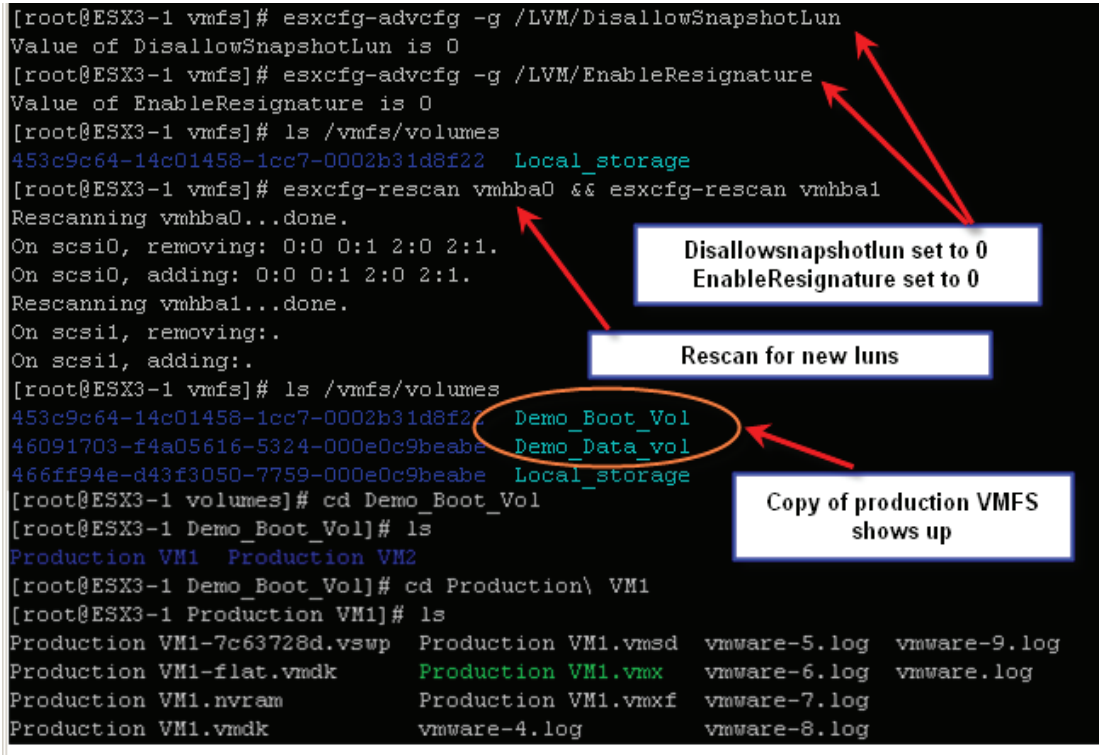


Figure 114 Presenting MirrorView secondary images to ESX 3.x server at the remote site

4. As stated previously, if VMware file system labels are not being used, the virtual machine configuration files needs modification to accommodate changes in the canonical names of the devices.
5. The cloned virtual machines can be powered on using the VirtualCenter client or command line utilities when required.

6. The process to promote the secondary images can be performed using Navisphere Manager.

Starting virtual machines at a remote site in the event of a planned failover

The process to start the virtual environment at the remote site in a planned failover event is slightly different from the one discussed in "Starting virtual machines at a remote site in the event of a disaster," on page 287. In a planned failover, the production CLARiiON is not lost. The dynamic swap capability of MirrorView can be used to provide continuous protection before production workload is started at the remote site.

1. The production environment has to be shut down before the planned failover is initiated. Furthermore, to ensure no loss of data, the secondary images of LUNs in a MirrorView/S relationship should be in a synchronized state before the failover process is started. In addition, since the secondary image in a MirrorView/A pair is always behind the primary image, a manual update of the secondary image should be performed after the applications at the production site have been shut down.
2. The promote command on an individual secondary image or the promotegroup command for a consistency group using the normal option type, swaps the roles of the primary and secondary images. The commands for LUNs in MirrorView/S relationship is shown below:

```
naviseccli –h SPipaddress mirror -sync -promoteimage
-name <name> type normal
naviseccli –h SPipaddress mirror -sync -promotegroup
-name <name> type normal
```

The same effect can be created for LUNs in MirrorView/A relationship by substituting the "-sync" option with "-async" option as shown below:

```
naviseccli –h SPipaddress mirror -async -promoteimage
-name <name> type normal
naviseccli –h SPipaddress mirror -async -promotegroup
-name <name> type normal
```

The two commands shown above make the following changes:

1. The current primary images on the production site are write disabled.
2. The personality of the devices is swapped.

3. The devices at the remote site (now primary) are made available as read-write.
4. The MirrorView link is resumed to allow data to flow from the remote data center to the production data center. (See Figure 115 on page 290.)

The process to power on the virtual machines at the remote site after executing the device failover process is the same as that discussed in "Starting virtual machines at a remote site in the event of a disaster," on page 287.

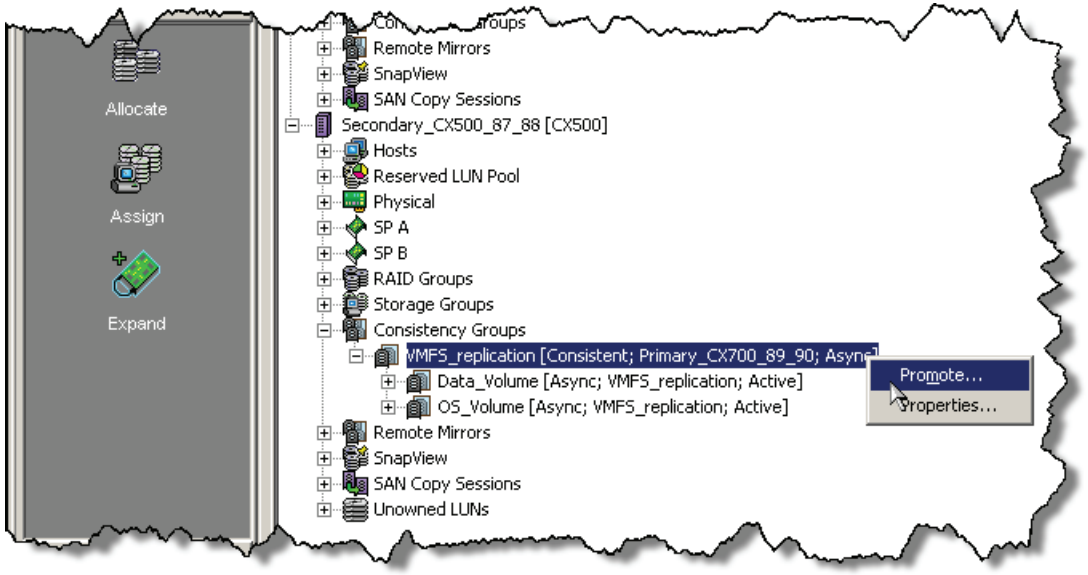


Figure 115 Promoting MirrorView secondary images using Navisphere Manager

5. The cloned virtual machines can be powered on using the VirtualCenter client or command line utilities when required. The process for starting the virtual machines at the remote site after an unplanned or planned event is the same as those discussed in the section "Starting virtual machines at a remote site in the event of a disaster," on page 287.

Configuring remote sites for VMs using VMFS-3 on VMware vSphere 4

Like VMware ESX/ESXi version 3, VMware ESX 4.x also assigns a unique signature to all VMFS-3 volumes when they are formatted with the VMware file system. Furthermore, if the VMware file system is

labeled that information is also stored on the device. The signature is generated using the unique ID (UID) of the device and the Host LUN number at which the device is presented.

VMware ESX version 4 and ESXi 4.x also provide two mechanisms to access copies of VMFS-3 volumes. Selective resignaturing is available for an individual LUN using the **Keep existing signature** or **Assign a new signature** option,

- If **Keep existing signature** is selected for an individual LUN, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts that have access to both source and target devices, the parameter has no effect since VMware ESX/ESXi hosts never present a copy of the data if there are signature conflicts. The default value for this parameter is 1.
- If **Assign a new signature** is selected for an individual LUN, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include snap-x, where x is a hexadecimal number that can range from 0x2 to 0xffffffff.

When using MirrorView for replication of a production VMware vSphere 4 environment, EMC recommends setting the Keep existing signature for the individual LUN on VMware ESX version 4.x or a VMware ESXi 4.x cluster at the remote site. EMC recommends that you *not* use **Assign a new signature** since it introduces complexity to the process of starting the virtual machines if there is a disaster.

The following points discuss the process of creating virtual machines at the remote site by selecting the **Keep existing signature** option.

1. You can promote the secondary images so that they become read-write enabled, so they can be accessed by the VMware ESX cluster group at the remote data center.

   The other option is to create SnapView replicas of the secondary images and present them for access to the VMware ESX cluster group at the remote site.

2. vSphere 4.x tightly integrates the vCenter infrastructure and VMware ESX 4.x or VMware ESXi 4.x version. vCenter infrastructure does not allow duplication of objects in a vCenter data center. If the same vCenter infrastructure is used to manage the

VMware ESX/ESXi hosts at the production and remote site, the servers should be added to different data center constructs in vCenter.

3. The SCSI bus should be scanned after providing the VMware ESX/ESXi hosts at the target site with access to the copy of the remote devices. You can scan the SCSI bus using the service console or the vCenter client.

4. Using the vCenter client **Add storage** wizard will list the devices holding the copy of the VMware file systems replicated from the source devices. Select the **Keeping existing signature** option for each LUN copy. After this option is selected for all LUNs, the VMware filesystems will be displayed under the **Storage** tab of the vClient interface.

5. In a Virtual Infrastructure 4 environment, when a virtual machine is created all files related to the virtual machine are stored in a directory on a Virtual Infrastructure 4 datastore. This includes the configuration file and, by default, the virtual disks associated with a virtual machine. Thus, the configuration files automatically get replicated with the virtual disks when storage array based copying technologies are leveraged.

6. You can perform the registration of the virtual machines from the target device using the vCenter client or the service console. The registration of cloned virtual machines is not required when the configuration information of the production virtual machine changes. The changes are automatically propagated and used when needed.

   As recommended in step 2, if the VMware ESX/ESXi hosts at the remote site are added to a separate data center construct, the names of the virtual machines at the remote data center match those of the production data center.

7. You can start the virtual machines on the VMware ESX/ESXi hosts at the remote site without any modification if the following requirements are met:

   - The target VMware ESX/ESXi host has the same virtual network switch configuration—for example, the name and number of virtual switches are duplicated from the source VMware ESX/ESXi cluster group.

- All VMware file systems that are used by the source virtual machines are replicated. Furthermore, the VMFS labels should be unique on the target VMware ESX/ESXi hosts.
- The minimum memory and processor resource requirements of all cloned virtual machines can be supported on the target VMware ESX/ESXi hosts. For example, if 10 source virtual machines, each with a memory resource reservation of 256 MB, need to be cloned, the target VMware ESX/ESXi hosts cluster should have at least 2.5 GB of physical RAM allocated to the VMkernel.
- Devices, such as CD-ROM and floppy drives, are attached to physical hardware, or are started in a disconnected state when the virtual machines are powered on.

8. You can power on the cloned virtual machines using the VirtualCenter client or command line utilities when required. The process for starting the virtual machines at the remote site after an unplanned or planned event is the same as those discussed in the section "Starting virtual machines at a remote site in the event of a disaster," on page 287.

Configuring remote sites for VMware Infrastructure 3 and VMware vSphere 4 VMs with RDM

When a RDM is generated, a virtual disk is created on a VMware file system pointing to the physical device that is mapped. The virtual disk that provides the mapping also includes the unique ID and LUN number of the device it is mapping. The configuration file for the virtual machine using the RDM contains an entry that includes the label of the VMware file system holding the RDM and the name of the RDM. If the VMware file system holding the information for the virtual machines is replicated and presented on the VMware ESX/ESXi host at the remote site, the virtual disks that provide the mapping is also available in addition to the configuration files. However, the mapping file cannot be used on the VMware ESX/ESXi host at the remote site since they point to non-existent devices. Therefore, EMC recommends using a copy of the source virtual machine's configuration file instead of replicating the VMware file system. The following steps create copies of production virtual machine using RDMs at the remote site:

1. On the VMware ESX/ESXi cluster group at the remote site, create a directory on a datastore (VMware file system or NAS storage) to hold the files related to the cloned virtual machine. A VMware file

system on internal disk, unreplicated SAN-attached disk or NAS-attached storage should be used for storing the files for the cloned virtual machine. This step has to be performed only once.

2. Copy the configuration file for the source virtual machine to the directory created in step 1. The command line utility, `scp`, can be used for this purpose. This step has to be repeated only if the configuration of the source virtual machine changes.
3. Register the cloned virtual machine using the Virtual Infrastructure client or the service console. This step does not need to be repeated.
4. Generate RDMs on the target VMware ESX/ESXi hosts in the directory created in step 1. The RDMs should be configured to use the secondary MirrorView images.

The virtual machine at the remote site can be powered on using either the Virtual Infrastructure client or the service console when needed. The process for starting the virtual machines at the remote site after a unplanned or planned event is the same as those discussed in sections "Starting virtual machines at a remote site in the event of a disaster," on page 287.

Automate continuity with 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 for FC and iSCSI connections.

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

SRM supports the replication of both VMFS and RDM volumes. SRM requires that the protected (primary) site and the recovery (secondary) site each has two independent virtual infrastructure clients. To use the MirrorView SRA, mirrors need to be created, and secondary LUNs need to be added and placed in a MirrorView/S or MirrorView/A 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 1.4 Release Notes***.

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.

Configuring MirrorView/S or MirrorView/A via the Navisphere Manager Configure MirrorView wizard

To configure sync or async mirrors, open the wizard and follow the instructions in the wizard (Figure 116 on page 296).

Note: The MirrorView SRA now supports both MirrorView/S and MirrorView/A mirror types.

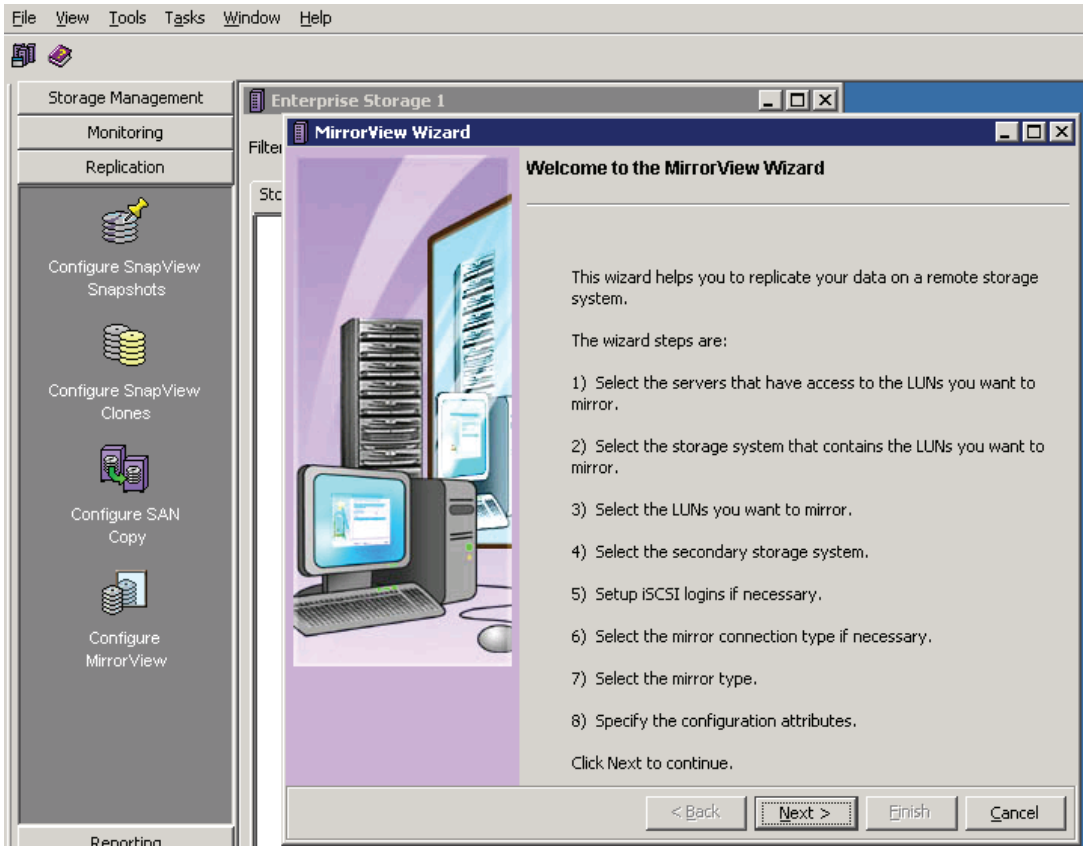


Figure 116 MirrorView Wizard

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>

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>
```

3. It is not mandatory to create a consistency group for the LUNs used by SRM. EMC recommends that you do this if possible, which will depend on the limits on the array. 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>
```

4. 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>
```

5. 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 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 –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 begins between the primary and the secondary images. 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 <sp-hostname| sp ipaddress> -lun
   <lunnumber| lun uid>
   ```

4. It is not mandatory to create a consistency group for the LUNs used by SRM. However, EMC recommends that you do this if it is possible, subject to the limits on the array. The following commands show how to 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> ;-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 –async –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>
   ```

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 30% 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 SnapView Snapshot Configuration Wizard (Figure 117 on page 300) to create snapshots on the array at the SRM protection site, so that if a failback is necessary, this step has already been performed.

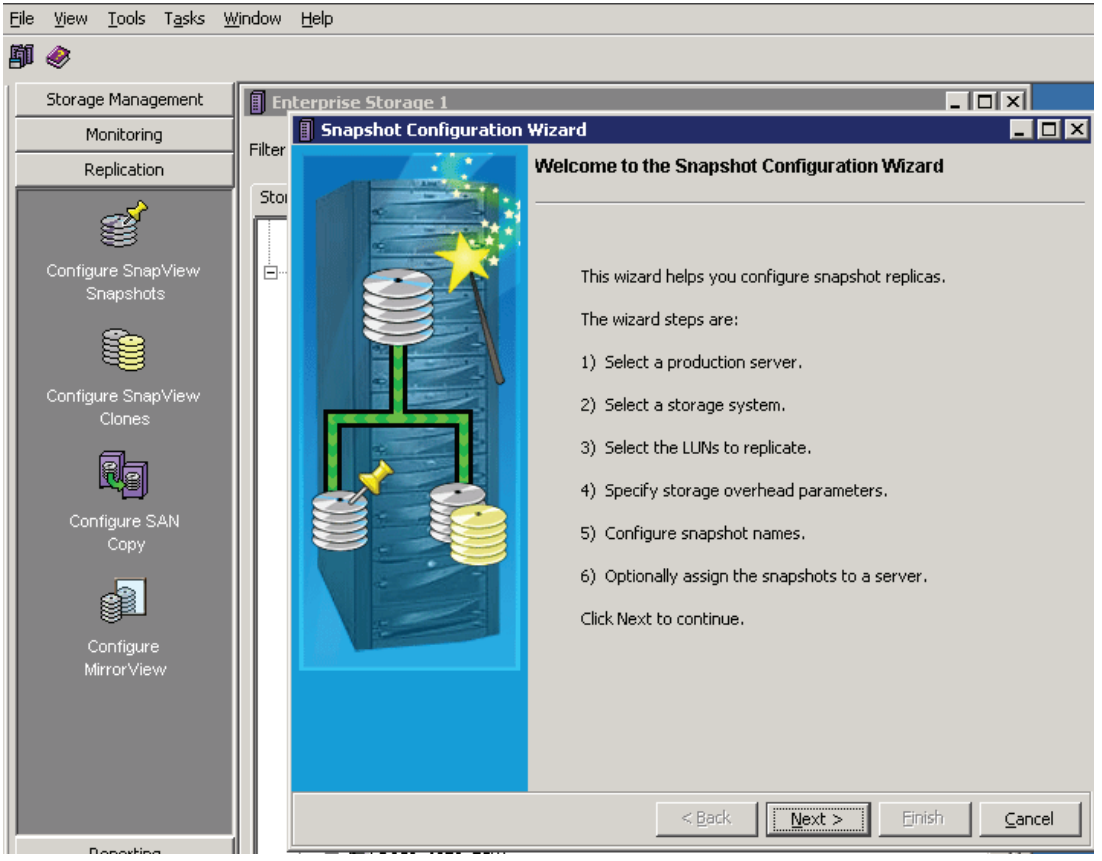


Figure 117 SnapView Snapshot Configuration Wizard

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 the 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 Reference* and *MirrorView/Asynchronous Command Line Interface Reference* available on Powerlink.

EMC recommends that you configure SRM and CLARiiON MirrorView Adapter within vCenter on the protected and recovery sites after the mirrors and consistency groups are configured on the CLARiiON. Refer to the *VMware vCenter SRM Administration Guide,* along with the *EMC MirrorView Adapter for VMware Site Recovery Manager Version 1.4 Release Notes* for installation and configuration instructions.

MirrorView Insight for VMware (MVIV)

After you complete the CLARiiON configuration, the SRM administrator can use the MVIV tool available with the MirrorView SRA to verify and validate the underlying replication configuration of the CLARiiON storage systems, as well as the ESX servers. MVIV also

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

helps troubleshoot the configuration by presenting the entire configuration, such as MirrorView and the VMware ESX/ESXi hosts and datastores on a single screen as shown in Figure 118 on page 302.

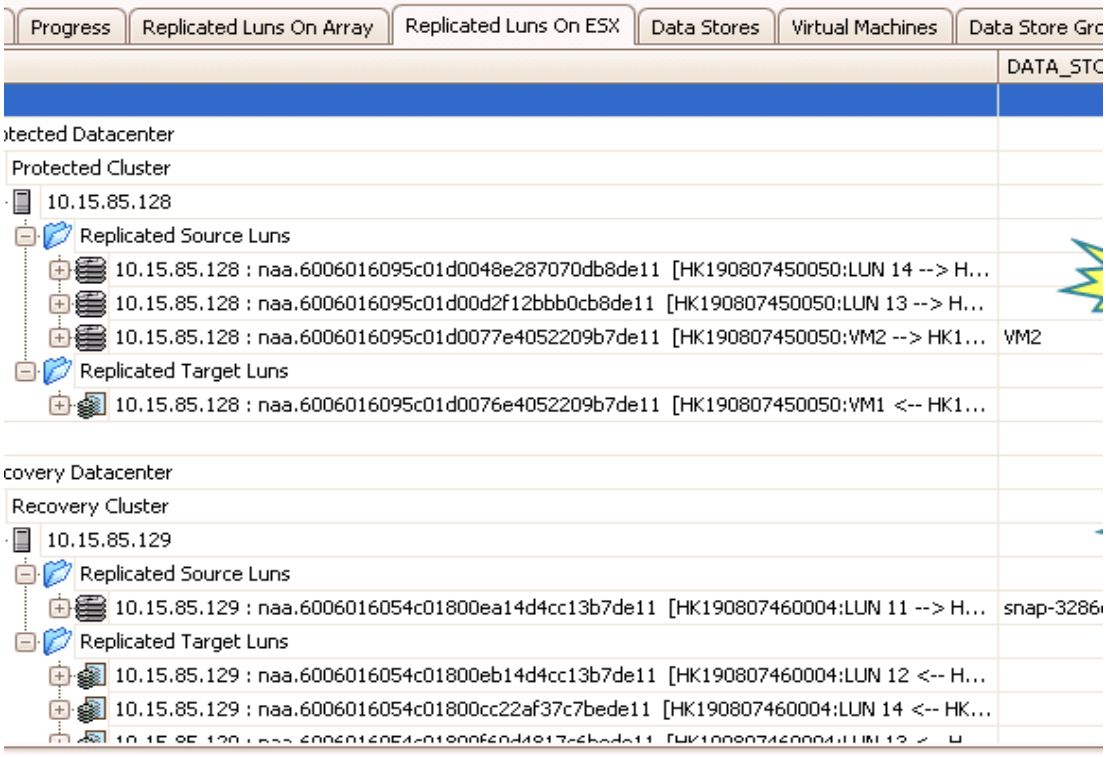


Figure 118 MVIV reporting for SRM environments

For example, MVIV detects if:

- All datastores are replicated and none of the RDMs attached to VMs on the datastores are replicated.
- All RDMs attached to the VMs are replicated and none of the datastores for the VMs are replicated.
- The target LUN is not attached to the any ESX Servers and more.

Refer to the *MirrorView Insight for VMware (MVIV)* technical note available on Powerlink for more details on MVIV.

Creating SRM Protection Groups at the protected site

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. See Table 5 on page 303 and Table 6 on page 303 for relevant maximum mirror limits for storage systems.

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
Total mirrors per consistency group	32	32	64
Total consistency groups per storage system	32	32	64

Table 6 Maximum number of async mirrors and consistency groups

Parameter	CX4-120	CX4-240	CX4-480, CX4-960
Total mirrors per storage system	100	100	100
Total mirrors per consistency group	32	32	64
Total consistency groups per storage system	64	64	64

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

For the maximum number of sync and async mirrors and consistency groups for the CLARiiON CX4 andCX3 storage systems, please refer to the ***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 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 at the recovery site

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 119 on page 305. To test the plan, click the **Test** button on the menu bar.

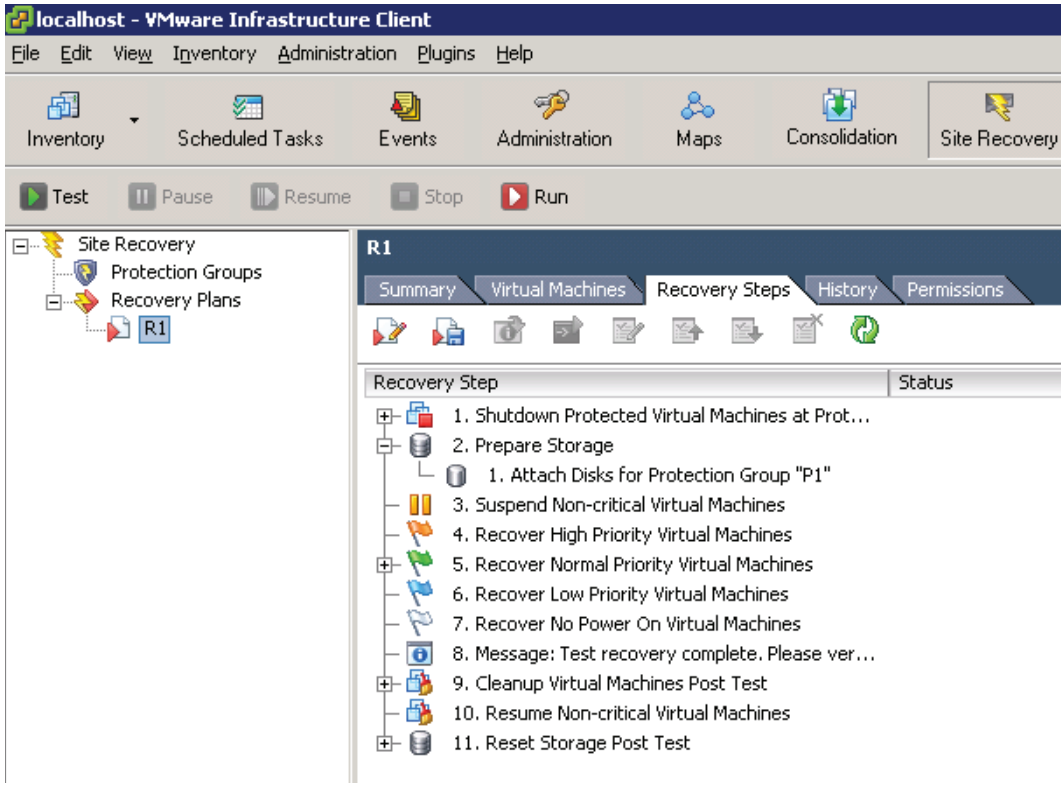


Figure 119 SRM Discovery Plan

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

1. Production VMs shut down
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 vCenter SRM Administration Guide***.

Executing an SRM recovery plan at the recovery site

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.

SRM 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/S and MirrorView/A, please see the white paper ***MirrorView Knowledgebook*** on Powerlink.

Note that you can use either SRM to fail back; however you must set up the array managers, protection groups, and recovery plans from the primary site to the disaster site. For more information on configuring failback, please see the ***VMware vCenter SRM Administration Guide***.

In addition, when SRM is used for failover, you can MVIV to perform a failback; however the support is currently experimental. Failback with MVIV does not require any re-configuration becuase it discovers the existing environment and automatically performs the failback by executing the **Failover** option.

For details on how to address the failback scenarios with the MirrorView SRA, please see the ***MirrorView Insight for VMware (MVIV)*** technical note available on Powerlink.

Recommendations and cautions

The following steps outline recommendations and cautions that we discovered when we tested this integration solution.

- If the VMs to be failed over do not have VMware Tools installed, the recovery plan generates an error when it tries to shut down the production VMs. (This step is annotated in Figure 120 on page 308.) However, if the plan was configured properly, it will still work properly. Please note in this case, this will be flagged as an error in the recovery plan (which is accessed by clicking the **History** tab) even if the VMs fail over successfully.

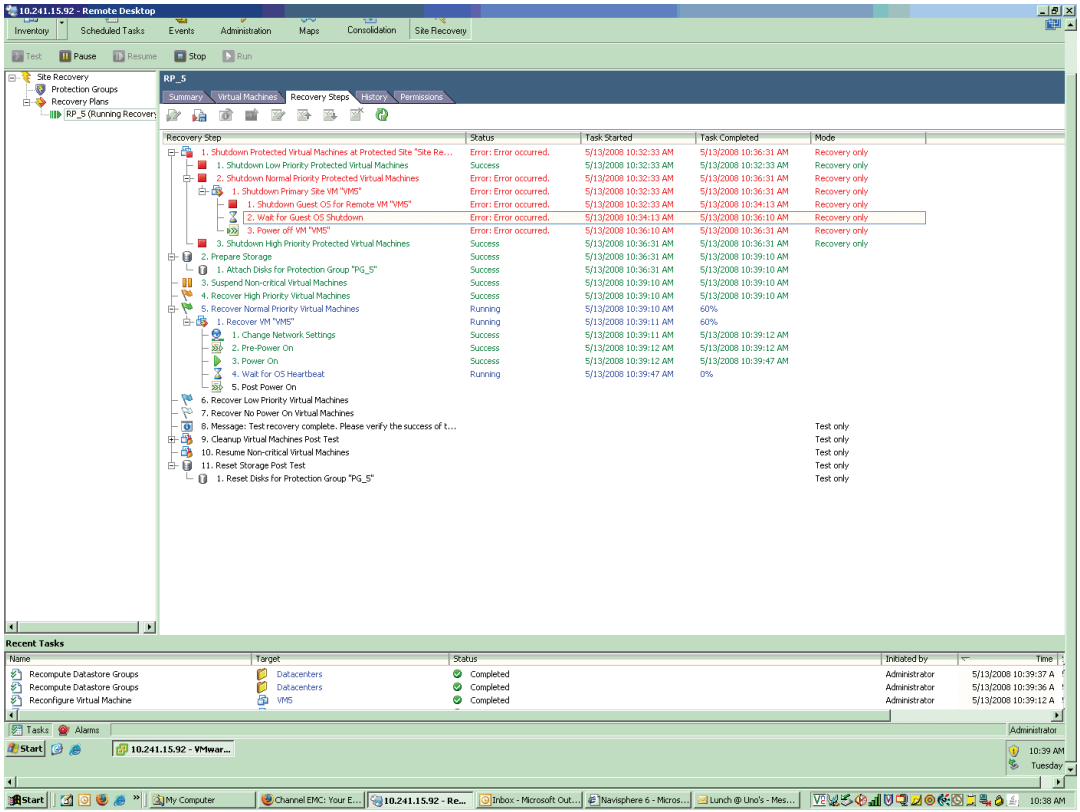


Figure 120 SRM recovery plan

- To test failover and failback, SnapView must be enabled on the arrays and snapshots at both the primary and secondary sites.
- Alarms must be created to announce the creation of new VMs on the datastore, so that mirrors can be configured to include the new VMs in the SRM protection scheme.
- We strongly recommend that CLARiiON-side configurations be completed (setting up MirrorView, creating snapshots, and so on) before installing SRM and SRA.
- If SRM is used for failover, we recommend that you use either SRM or MVIV for failback, since manual failback is cumbersome and requires changing the LVMEnableresignature on ESX 3.x, or

individually selecting each LUN and configuring the **Keep the existing signature** or **Assign a new signature** option within vSphere on the primary ESX servers. By default, SRM changes the LVMEnableresignature to 1in ESX 3.x environments or resignatures the datastore using the **Assign a new signature** option in ESX 4.0 environments and then renames the VMFS datastores.

- Testing a recovery plan only captures snapshots of the MirrorView secondary image, it does not check for connectivity between the arrays or to see if MirrorView is working properly. To verify connectivity between VM consoles, use the SRM connection. To check connectivity between arrays, use SRM Array Manager or Navisphere Manager.
- Ensure that you have enough disk space configured for both the VM and swap file at the secondary site to ensure that the recovery plan test runs successfully and without errors.

7
Data Vaulting and Migration of VMware Virtual Infrastructure

This chapter presents these topics:

- SAN Copy interoperability with VMware file systems 313
- SAN Copy interoperability with VMs using RDM 315
- Using SAN Copy for data vaulting 316
- Transitioning disk copies to cloned virtual machines 323
- SAN Copy for data migration from CLARiiON arrays 330
- SAN Copy for data migration to CLARiiON arrays 339

Every business strives to increase productivity and utilization of its most important resource or information. This asset is critical for finding the right customers, building the right products, and offering the best service. This often requires creating copies of the information and making it available to different business processes in the most cost-effective way. It can also involve migration of the information from one storage array to another as the criticality and the requirements of the business change. Finally, various compliance regulations can impose data vaulting requirements that may require creating additional copies of the data.

The criticality of the information has also imposed demanding availability requirements. Few businesses can afford protracted downtime as the information is copied and moved to various business units. On the other hand, creating copies of data and data migrations can often require extensive manual work, and long and complex planning. Unfortunately, due to the complexity of the processes, they tend to be error-prone as well, posing the risk of data corruption or loss.

VMware ESX/ESXi hosts and related products reduce the total cost of ownership by consolidating computing resources. However, the consolidation process can also result in applications with disparate service level agreements competing with one another for both compute and storage resources. VMware provides excellent technologies, such as VMotion, to minimize the competition for CPU, memory, and network resources. However, there is limited functionality within VMware ESX/ESXi hosts to optimize and migrate storage resources. Furthermore, creating copies of information in a VMware ESX/ESXi hosts environment using native tools requires elongated downtime of virtual machines. EMC offers various technologies to migrate information from one storage array to another, and create copies of the data with minimal impact to the operating environment. The purpose of this chapter is to discuss one such technology—EMC SAN Copy, and its interoperability in VMware ESX/ESXi environments.

SAN Copy interoperability with VMware file systems

A VMware file system is a high-performance clustered file system that is frequently deployed in a VMware virtual infrastructure. The file system can exist on a single CLARiiON device (a single LUN or metaLUN) or span multiple devices. Virtual machines access the abstracted form of physical storage through virtual disks that are represented by large flat files.

SAN Copy can be used to migrate or create copies of VMware file systems. If the VMware file system is a spanned file system with two or more physical extents, all members need to be replicated or migrated automatically. With SAN Copy, this is possible only when all virtual machines accessing a spanned VMware file system needs to be shut down before SAN Copy sessions are started. In lieu, SAN Copy session can be initiated from a SnapView clone of an active VMware file system. This is acceptable since an activated clone target maintains a consistent point-in-time copy of the VMware file system.

VMware ESX version 4, 3 and VMware ESXi behave differently when presented with copies of the same VMware file system. If a copy of a VMFS-3 volume is presented to any VMware ESX/ESXi version 4 or 3 cluster, the VMware ESX/ESXi host automatically masks the copy. The device holding the copy is determined by comparing the signature stored on the device with the computed signature. A clone, for example, has a different unique ID from the source LUN it is associated with it. Hence, the computed signature for a clone always differs from the computed signature of the source LUN. This enables the VMware ESX/ESXi hosts to always identify the copy correctly.

VMware ESXversion 3 and VMware ESX 3i provide two different mechanisms to access copies of VMFS-3 volumes. The advanced configuration parameters, `LVM.DisallowSnapshotLun` or `LVM.EnableResignature`, control the behavior of the VMkernel when presented with copies of a VMware file system.

- If `LVM.DisallowSnapshotLun` is set to 0, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts that have access to both source and target devices, the parameter has no effect since VMware ESX/ESXi hosts never presents a copy of the data if there are signature conflicts. The default value for this parameter is 1.
- If `LVM.EnableResignature` is set to 1, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include snap-x, where x is a hexadecimal number that can range from 0x2 to 0xffffffff. The default value for this parameter is 0. If this parameter is changed to 1, the advanced parameter `LVM.DisallowSnapShotLun` is ignored.

By using the proper combination of the advanced configuration parameter, copies of VMFS-3 can be used to perform ancillary business operations.

VMware ESX version 4 and ESXi 4.x also provide two mechanisms to access copies of VMFS-3 volumes. Selective resignaturing is available for an individual LUN using the **Keep existing signature** option or the **Assign a new signature** option.

- If **Keep existing signature** is selected for an individual LUN, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts that have access to both source and target devices, the parameter has no effect since VMware ESX/ESXi never presents a copy of the data if there are signature conflicts.
- If **Assign a new signature i**s selected for an individual LUN, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include snap-x, where x is a hexadecimal number that can range from 0x2 to 0xffffffff.

SAN Copy interoperability with VMs using RDM

RDM volumes in physical compatibility mode provide direct access to virtual machines created in VMware ESXversion 4.x, VMware 3, and VMware ESXi. When virtual machines are provided direct access to storage, the VMware ESX/ESXi hosts kernel does not participate in any facet of I/Os generated by the virtual machine to the device. This limits some of the advanced functionality that is provided by the VMware ESX/ESXi kernel. However, providing virtual machines with dedicated access to storage devices does provide some advantages.

Storage array-based replication and migration are performed at the CLARiiON LUN level. When virtual machines are provided with raw devices, creating copies of data using storage array software can be done at an individual virtual machines level. Furthermore, since the virtual machines communicate directly with the storage array, storage management commands can be directly executed in the virtual machines. This capability is of limited use when using SAN Copy for data migrations. To ensure migrations do not result in data loss, all virtual machines need to be shut down before the migration is completed. In this state, it is impossible to control the data migration if a virtual machine is used to manage the SAN Copy sessions. EMC, therefore, recommends a separate management host when performing SAN Copy activities.

Using SAN Copy for data vaulting

SAN Copy provides different modes of operation. One such mode is referred to as the incremental mode that can be used for push operations. In incremental mode, SAN Copy ensures propagation of data from the production volume to a volume of equal or larger size on the remote storage array. This mechanism can be leveraged to provide a data vaulting solution in which a copy of the production data can be made available for ancillary business processes on a cost-effective storage platform.

Figure 121 on page 316 is a schematic representation of the data vaulting solution that can be used for environment in which the write I/O rate to the production volumes is not very large. To maintain consistency of the data at the remote location, incremental SAN Copy uses reserved LUNs on the production storage array when the production data not copied to the target array is updated. This behavior results in performance overhead and is not appropriate for environments subject to very large rate of change.

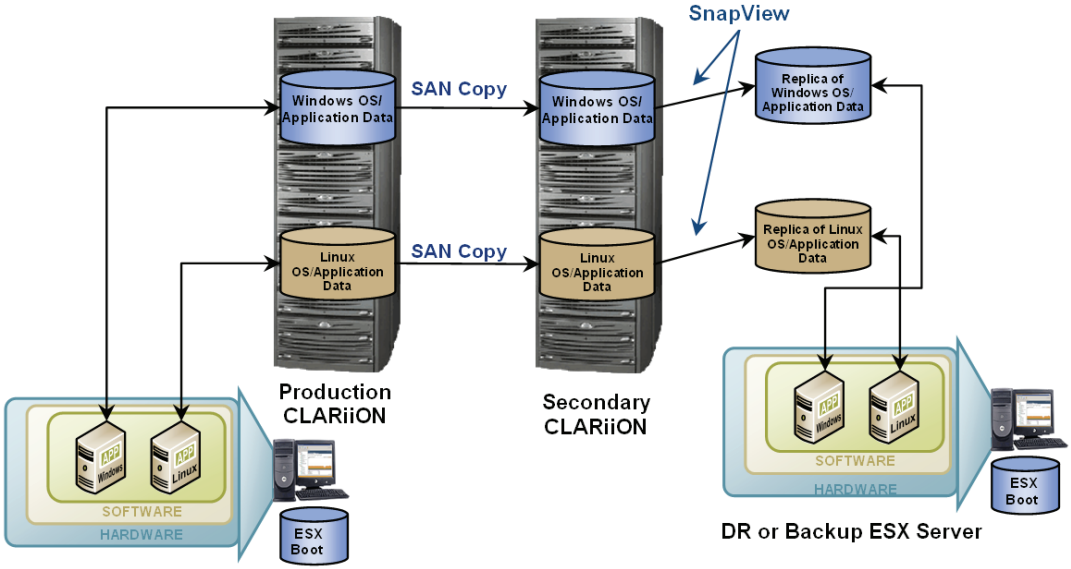


Figure 121 Data vaulting solution using incremental SAN Copy in a virtual infrastructure

The performance penalty can be minimized by modifying the solution presented in Figure 121 on page 316. The solution, shown in Figure 122 on page 317, leverages SnapView Clone technology to create a copy of the production volume, and propagating that copy to the remote storage array utilizing incremental SAN Copy. Both solutions presented in Figure 121 on page 316 and Figure 122 on page 317 depict virtual machines accessing the storage as RDM. However, the solution can be used to replicate VMware file systems.

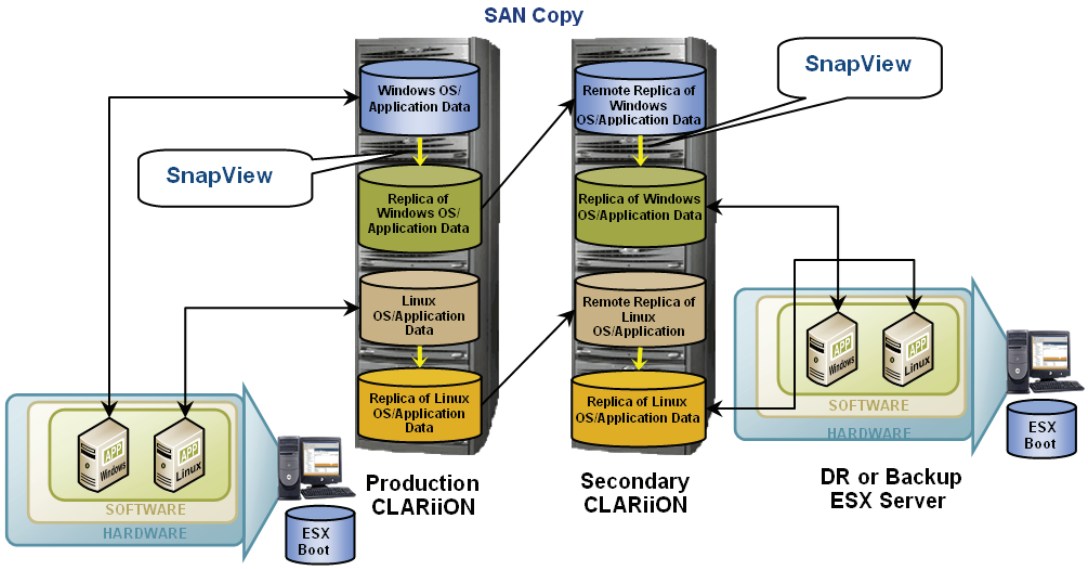


Figure 122 Minimum performance penalty data vaulting solution using incremental SAN Copy

Data vaulting of VMware file system using SAN Copy

The first step in creating a data vaulting solution using SAN Copy involves the identification of the appropriate devices and their canonical names on the VMware ESX/ESXi host environments

constituting the VMware file systems. Figure 123 on page 318 shows an example for utilizing VMware file system version 3. The devices have to be then related to the CLARiiON LUN numbers.

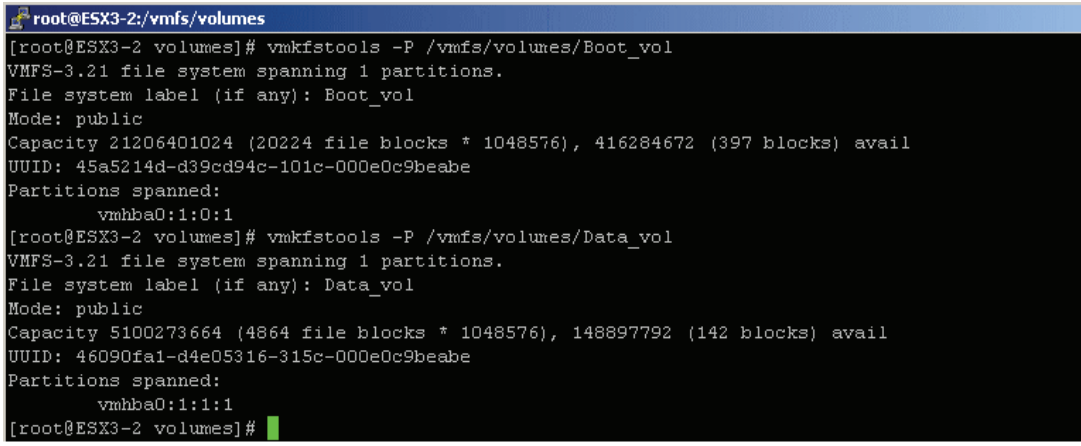


Figure 123 Identifying the canonical name associated with VMware file systems

As shown in Figure 124 on page 319, the Navisphere CLI and agent can then be used to determine the WWN of CLARiiON devices that need to be replicated. In addition, the VMware-aware Navisphere feature available with Release 29 of FLARE also provides mapping between CLARiiON LUNs and VMware filesystems, and helps determine which LUNs need to be replicated.

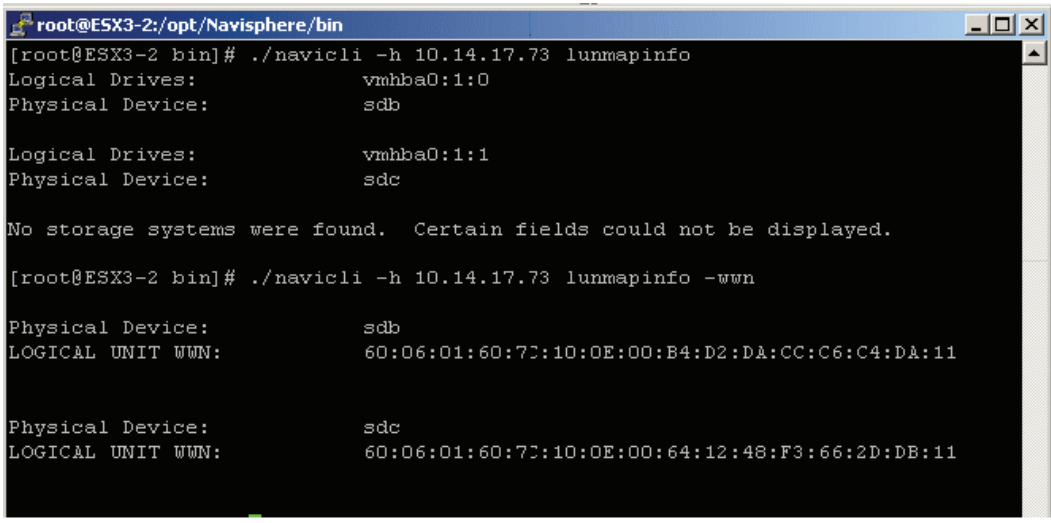


Figure 124 Using Navisphere CLI/Agent to map the canonical name to EMC CLARiiON devices

The next step in the process of creating a data vaulting solution is identifying the WWN of the remote devices. The WWN is a 128-bit number that uniquely identifies any SCSI device. The WWN for a SCSI device can be determined using different techniques. Management software for the storage array can provide the information. Solutions Enabler can be used to obtain the WWN of devices on supported storage arrays (Symmetrix, HDS and HP StorageWorks).

After the CLARiiON devices constituting the VMware file systems and the WWN of the remote devices have been identified, the process to create and manage SAN Copy is the same as the one for physical servers. The following paragraphs describe the steps required to create a data vaulting solution using SAN Copy:

1. In a Fibre Channel storage area network (SAN), for a host initiator to perform I/Os to a storage array port, the host initiator needs to log in to the storage array port. The host initiator, when it logs in to the fabric, is provided by the fabric name service (NS) with the Fibre Channel addresses of all storage array ports it is allowed to access. Zoning is the mechanism used to educate NS with appropriate access information.

2. SAN Copy enables the CLARiiON Storage Processor (SP) ports to act as a host initiator. Therefore, a zone that includes the CLARiiON SP ports and the Fibre Channel ports on the target storage array allows the CLARiiON SP ports to perform I/Os to target storage arrays.

3. Most modern storage arrays do not allow unrestricted access to storage devices. The access to storage devices is enforced by the LUN masking software running on the storage array. The CLARiiON SP ports need to be able to access the remote devices to be able to perform I/Os to the devices and propagate a point-in-time copy of the data from the source devices. The next step in creation of a data vaulting solution is to provide the CLARiiON SP ports with appropriate access to the remote devices. The management software for the storage array should be used to provide the CLARiiON SP ports with access to the appropriate LUNs on the remote storage array.

4. SAN Copy incremental sessions internally communicate with SnapView software to keep track of changes and updates for a SAN Copy session. SnapView needs a set of reserved LUNs that it uses to keep track of the changed data. Therefore, a reserved LUN pool needs to be assigned to SnapView before incremental SAN Copy sessions can be created. The number and size of these LUNs depend on the rate of the change on the source LUN during the SAN Copy update operation.

5. The next step in the process, as shown in Figure 125 on page 321, is to define a SAN Copy session, which involves the LUNs in the relationship. The attributes for the SAN Copy session— SAN Copy session name, WWN of source and destination LUNs, throttle value, latency and bandwidth control— can be specified when the session is created. The bandwidth value is required, but the value for the latency parameter can be left to default, in which case SAN Copy will measure latency by sending test I/O to the target.

6. There is no movement of data when SAN Copy sessions are created. When a session is created, SAN Copy performs a series of tests to validate the configuration. These include checks to ensure the CLARiiON SP ports have access to the remote devices, and that the remote devices are of equal or larger size than the source devices.

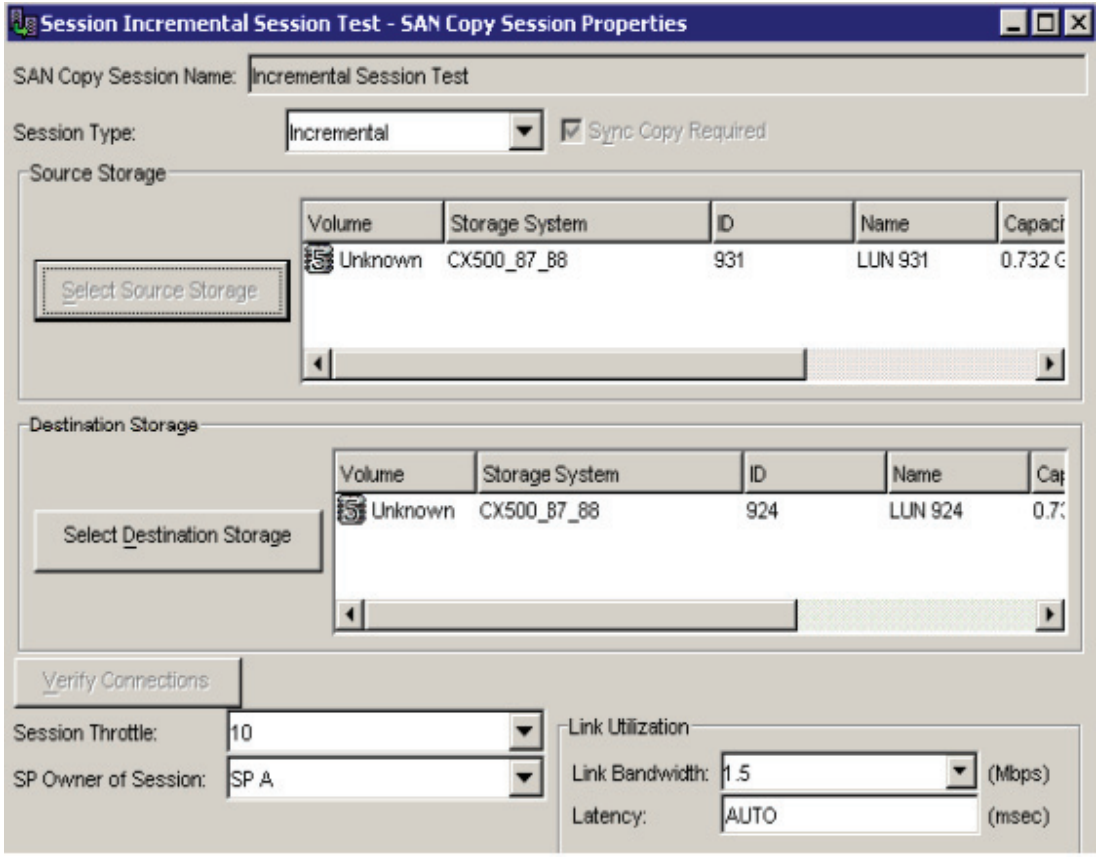


Figure 125 Creating an incremental SAN Copy session

7. Starting or activating the session created in the previous step results in the propagation of a point-in-time copy of the data from the source devices to target devices.

8. SAN Copy provides a parameter (throttle) to control the rate at which data is copied from or to the source and target devices. A throttle value of 10 will cause SAN Copy software to use all available system resources to speed the transfer rate. The throttle value can be changed dynamically after the session is created.
9. After the copy process is complete, the data on the remote devices can be accessed by virtual machines configured on a different VMware ESX/ESXi host. However, EMC does not recommend this. Incremental updates to the target volumes are possible only if the remote devices are not actively accessed by hosts.
10. To access the copy of the data on remote devices, EMC recommends use of snapshot technology native to the target storage array. For example, if the target storage array is an EMC CLARiiON storage array, SnapView snapshots can be leveraged to present a copy of the data to the virtual machines.
11. An incremental update of the remote device can be achieved by restarting the previously created SAN Copy session. Incremental updates can dramatically reduce the amount of data that needs to be propagated from the source volume in cases where the amount of data to be copied is a small fraction of the size of the source volume.

Data vaulting of VMs configured with RDMs using SAN Copy

SAN Copy provides a storage array-based mechanism to copy a consistent point-in-time copy of data on CLARiiON devices to supported third-party storage. When virtual machines are configured with raw devices or RDM, the use of SAN Copy is simplified. Furthermore, by replicating data at the individual virtual machine level, copying of unnecessary data can be eliminated.

Virtual machines configured with RDM in physical compatibility mode are aware of the presence of CLARiiON devices. Navisphere CLI/Agent installed on the virtual can be used to easily determine the devices that need to be replicated using SAN Copy. Once the devices have been identified, the process to create and use SAN Copy with virtual machines using raw devices is the same as that listed in "Data vaulting of VMware file system using SAN Copy," on page 317.

Transitioning disk copies to cloned virtual machines

Configuring remote sites for VMs using VMFS-3 for VMware Infrastructure 3

VMware ESX/ESXi hosts version 3 and VMware ESXi assign a unique signature to all VMFS-3 volumes when they are formatted with the VMware file system. Furthermore, if the VMware file system is labeled that information is also stored on the device. The signature is generated using the unique ID (UID) of the device and the Host LUN number at which the device is presented.

Since storage array technologies create exact replicas of the source volumes, all information including the unique signature (and label, if applicable) is replicated. If a copy of a VMFS-3 volume is presented to any VMware ESX version 3.x or VMware ESXi cluster, the VMware ESX/ESXi host automatically masks the copy. The device that holds the copy is determined by comparing the signature stored on the device with the computed signature. MirrorView secondary images, for example, have different unique IDs from the primary images with which they are associated. Therefore, the computed signature for these secondary images will always differs from the primary images. This enables the VMware ESX/ESXi host to always identify the copy correctly.

VMware ESX version 3.x and VMware ESXi provide two different mechanisms to access copies of VMFS-3 volumes. The advanced configuration parameters, `LVM.DisallowSnapshotLun` or `LVM.EnableResignature`, control the behavior of the VMkernel when presented with copies of a VMware file system.

- If `LVM.DisallowSnapshotLun` is set to 0, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts with access to both source and target devices, the parameter has no effect since VMware ESX/ESXi hosts never presents a copy of the data if there are signature conflicts. The default value for this parameter is 1.
- If `LVM.EnableResignature` is set to 1, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include “snap-x”, where x is a hexadecimal number that can range from 0x2 to 0xffffffff. The default value for this parameter is 0. If this parameter is changed to 1, the advanced parameter `LVM.DisallowSnapShotLun` is ignored.

When using SAN Copy for data vaulting of production Virtual Infrastructure 3 environment, EMC recommends setting the `LVM.DisallowSnapshotLun` to 0 on VMware ESX version 3.x or a VMware ESXi cluster at the remote site. The use of `LVM.EnableResignature` is strongly discouraged since it introduces complexity to the process of starting the virtual machines in case of a disaster.

CAUTION

The `LVM.EnableResignature` parameter should not be changed for any reason on the VMware ESX/ESXi hosts at the remote site. All volumes that are considered to be copies of the original data will be resignatured if the parameter is enabled. Furthermore, there is no mechanism currently available to undo the resignaturing process. Depending on the state of the infrastructure at the remote site, the resignaturing process can cause havoc with the data vaulting environment.

The following paragraphs discuss the process to create virtual machines at the remote after changing the advanced parameter, `LVM.DisallowSnapshotLun` to 0 (see "Cloning Virtual Infrastructure 3 virtual machines using LVM.DisallowSnapshotLun ," on page 201 for the process to change the parameter).

1. Enable access to the copy of the remote devices for the VMware ESX/ESXi cluster group at the remote data center. To preserve the incremental push capabilities of SAN Copy, the remote devices should never be accessed directly by the VMware ESX/ESXi hosts.
2. Virtual Infrastructure 3 tightly integrates the vCenter infrastructure and VMware ESX version 3.x or VMware ESXi version. vCenter infrastructure does not allow duplication of objects in a vCenter data center. If the same vCenter infrastructure is used to manage the VMware ESX/ESXi hosts at the production and remote site, the servers should be added to different data center constructs in vCenter.
3. The SCSI bus should be scanned after providing the VMware ESX/ESXi hosts at the target site with access to the copy of the remote devices. The scanning of the SCSI bus can be done either using the service console or the vCenter client. The devices holding the copy of the VMware file system is displayed on the VMware ESX/ESXi cluster at the remote site.

4. In a Virtual Infrastructure 3 environment, when a virtual machine is created all files related to the virtual machine are stored in a directory on a Virtual Infrastructure 3 datastore. This includes the configuration file and, by default, the virtual disks associated with a virtual machine. Thus, the configuration files automatically get replicated with the virtual disks when storage array based copying technologies are leveraged. Therefore, unlike a VMware ESX 2.x environment, there is no need to manually copy configuration files.

   The registration of the virtual machines from the target device can be performed using the vCenter client or the service console. The registration of cloned virtual machines is not required when the configuration information of the production virtual machine changes. The changes are automatically propagated and used when needed.

   As recommended in step 2, if the VMware ESX/ESXi hosts at the remote site are added to a separate data center construct, the names of the virtual machines at the remote data center matches those of the production data center.

5. The virtual machines can be started on the VMware ESX/ESXi hosts at the remote site without any modification if the following requirements are met:

   - The target VMware ESX/ESXi hosts have the same virtual network switch configuration—i.e., the name and number of virtual switches are duplicated from the source VMware ESX/ESXi cluster group.
   - All VMware file systems that are used by the source virtual machines are replicated. Furthermore, the VMFS labels should be unique on the target VMware ESX/ESXi hosts.
   - The minimum memory and processor resource reservation requirements of all cloned virtual machines can be supported on the target VMware ESX/ESXi hosts. For example, if ten source virtual machines, each with a memory resource reservation of 256 MB needs to be cloned, the target VMware ESX/ESXi cluster should have at least 2.5 GB of physical RAM allocated to the VMkernel.
   - Devices, such as CD-ROM and floppy drives, are attached to physical hardware, or are started in a disconnected state when the virtual machines are powered on.

6. The cloned virtual machines can be powered on using the vCenter client or command line utilities when required.

The process for starting the virtual machines at the remote site is the same as those discussed in "Starting virtual machines at a remote site in the event of a disaster," on page 287.

Configuring remote sites for VMs using VMFS-3 for vSphere 4.x

Like VMware ESX/ESXi version 3, VMware ESX 4.x also assigns a unique signature to all VMFS-3 volumes when they are formatted with the VMware file system. Furthermore, if the VMware file system is labeled that information is also stored on the device. The signature is generated using the unique ID (UID) of the device and the Host LUN number at which the device is presented.

VMware ESX version 4 and ESXi 4.x also provide two mechanisms to access copies of VMFS-3 volumes. Selective resignaturing is available for an individual LUN using the **Keep existing signature** option or **Assign a new signature** option.

- If **Keep existing signature** is selected for an individual LUN, the copy of the data is presented with the same label name and signature as the source device. However, on VMware ESX/ESXi hosts that have access to both source and target devices, the parameter has no effect since VMware ESX/ESXi hosts never present a copy of the data if there are signature conflicts. The default value for this parameter is 1.
- If **Assign a new signature** is selected for an individual LUN, the VMFS-3 volume holding the copy of the VMware file system is automatically resignatured with the computed signature (using the UID and LUN number of the target device). In addition, the label is appended to include snap-x, where x is a hexadecimal number that can range from 0x2 to 0xffffffff.

When using SAN Copy for data vaulting of a production VMware vSphere 4 environment, EMC recommends selecting the **Keep existing signature** for the individual LUN on VMware ESX/ESXi version 4.x or a VMware ESXi 4.x cluster at the remote site. You should not select the **Assign a new signature option** since it introduces complexity to the process of starting the virtual machines if there is a disaster.

The following paragraphs discuss the process of creating virtual machines at the remote site by selecting the **Keep existing signature** option.

1. Enable access to the copy of the remote devices for the VMware ESX/ESXi cluster group at the remote data center. To preserve the incremental push capabilities of SAN Copy, the remote devices should never be accessed directly by the VMware ESX/ESXi hosts.
2. vSphere 4.x tightly integrates the vCenter infrastructure and VMware ESX 4.x or VMware ESXi 4 version. vCenter infrastructure does not allow duplication of objects in a vCenter data center. So, if you use the same vCenter infrastructure to manage the VMware ESX/ESXi hosts at the production and remote sites, you need to add the servers to different data center constructs in vCenter.
3. After providing the VMware ESX/ESXi hosts at the target site with access to the copy of the remote devices, you need to scan the SCSI bus using the service console or the vCenter client.
4. Use the vCenter client Add **storage wizard** to list the devices holding the copy of the VMware file systems replicated from the source devices. Select the **Keeping existing signature** option for each LUN copy. After you select this option for all LUNs, the VMware filesystems are displayed under the **Storage** tab in the vClient interface.
5. In a Virtual Infrastructure 4 environment, when a virtual machine is created all files related to the virtual machine are stored in a directory on a Virtual Infrastructure 4 datastore. This includes the configuration file and, by default, the virtual disks associated with a virtual machine. Thus, the configuration files automatically get replicated with the virtual disks when storage array based copying technologies are leveraged.
6. You can register the virtual machines from the target device using the vCenter client or the service console. The registration of cloned virtual machines is not required when the configuration information of the production virtual machine changes. The changes are automatically propagated and used when needed.

   As mentioned in step 2, if the VMware ESX/ESXi hosts at the remote site are added to a separate data center construct, the names of the virtual machines at the remote data center match those of the production data center.
7. The virtual machines can be started on the VMware ESX/ESXi hosts at the remote site without any modification if the following requirements are met:

- The target VMware ESX/ESXi hosts have the same virtual network switch configuration—for example the name and number of virtual switches are duplicated from the source VMware ESX/ESXi cluster group.
- All VMware file systems that are used by the source virtual machines are replicated. Furthermore, the VMFS labels should be unique on the target VMware ESX/ESXi hosts.
- The minimum memory and processor resource reservation requirements of all cloned virtual machines can be supported on the target VMware ESX/ESXi hosts. For example, if 10 source virtual machines, each with a memory resource reservation of 256 MB needs to be cloned, the target VMware ESX/ESXi cluster should have at least 2.5 GB of physical RAM allocated to the VMkernel.
- Devices, such as CD-ROM and floppy drives, are attached to physical hardware, or are started in a disconnected state when the virtual machines are powered on.

8. The cloned virtual machines can be powered on using the vCenter client or command line utilities when required. The process for starting the virtual machines at the remote site is discussed in the section "Starting virtual machines at a remote site in the event of a disaster," on page 287.

Configuring remote sites for VMware Infrastructure 3 and VMware vSphere 4 VMs with RDM

When a RDM is generated, a virtual disk is created on a VMware file system that points to the physical device that is mapped. The virtual disk that provides the mapping also includes the unique ID and LUN number of the device that it is mapping. The configuration file for the virtual machine using the RDM contains an entry that includes the label of the VMware file system that holds the RDM and the name of the RDM. If the VMware file system, holding the information for the virtual machines is replicated and presented on the VMware ESX/ESXi hosts at the remote site, the virtual disks that provide the mapping is also available in addition to the configuration files. However, the mapping file cannot be used on the VMware ESX/ESXi hosts at the remote site since they point to nonexistent devices. Therefore, EMC recommends using a copy of the source virtual machine's configuration file instead of replicating the VMware file system. The following steps create copies of production virtual machine using RDMs at the remote site:

1. On the VMware ESX/ESXi cluster group at the remote site create a directory on a datastore (VMware file system or NAS storage) that holds the files related to the cloned virtual machine. A VMware file system on internal disk, unreplicated SAN-attached disk or NAS-attached storage should be used for storing the files for the cloned virtual disk. This step has to be performed only once.
2. Copy the configuration file for the source virtual machine to the directory created in step 1. The command line utility, `scp` can be used for this purpose. This step has to be repeated only if the configuration of the source virtual machine changes.
3. Register the cloned virtual machine using the vCenter client or the service console. This step does not need to be repeated.
4. Generate RDMs on the target VMware ESX/ESXi hosts in the directory created in step 1. The RDMs should be configured to address the copy of the remote devices.
5. The virtual machine at the remote site can be powered on using either the vCenter client or the service console when needed.

Note: The process listed in this section assumes the source virtual machine does not have a virtual disk on a VMware file system. The process to clone virtual machines with a mix of RDMs and virtual disks is complex and beyond the scope of this document. Readers are requested to contact the authors at ganeshan_bala@emc.com or Kochavara_sheetal@emc.com if such requirements arise.

The process for starting the virtual machines at the remote site is the same as those discussed in "Starting virtual machines at a remote site in the event of a disaster," on page 287.

SAN Copy for data migration from CLARiiON arrays

VMware ESX/ESXi hosts provide a limited set of tools to perform data migrations. Furthermore, most of the native tools require extensive downtime as the data is migrated from source devices to target devices. The extended downtime is normally unacceptable for critical business applications.

SAN Copy is frequently used for migrating data from CLARiiON storage arrays to other supported storage arrays. One of the major advantages that SAN Copy provides over other data migration technologies is the capability of providing incremental updates. This capability can be leveraged to provide a testing environment before switching production workload to the migrated devices. In addition, the incremental update capability can be used to minimize the outage window when the production workload is switched from the source devices to the migrated devices.

Note: MirrorView can be used for migration of VMware data. SAN Copy in discussed in this chapter since it provides support for heterogeneous arrays. If the migration is between two CLARiiON arrays, MirrorView can be used. The procedures and considerations are similar to those discussed in this section.

Migration of a VMware file system in ESX version 3

The data vaulting solution discussed in "Data vaulting of VMware file system using SAN Copy," on page 317 can be modified to provide a data migration solution when moving production workload from CLARiiON storage arrays to supported storage arrays. In addition to the process listed in "Data vaulting of VMware file system using SAN Copy," on page 317 additional steps are needed to complete the data migration. These steps are listed below:

1. The session created in step 5 "Data vaulting of VMware file system using SAN Copy," on page 317 should be started repeatedly to reduce the amount of data that needs to be migrated. Every time the SAN Copy session is started, an incremental refresh of the data on remote devices with up-to-date data on the source volume is performed. The time required to perform the incremental update should reduce exponentially as the sessions are started.
2. The switchover from the source devices to the remote devices should be planned when the amount of time required to perform incremental update does not show any further reduction.

The switchover process is initiated by first shutting down all virtual machines being migrated. After the virtual machines are shut down, a final incremental push of changed data from the control devices to remote devices should be initiated.

3. The virtual machines being migrated should be removed from the vCenter infrastructure inventory as shown in Figure 126 on page 332.
4. The references to the datastore on the current volumes should be removed.

Note: VMware ESX/ESXi version 3.0.1 or later and vCenter 2.0.1 or later will automatically remove datastores with no relations to other objects. This step is, therefore, unnecessary for these configurations.

---

5. As SAN Copy migrates the final pieces of changed data from the source devices to remote devices, the LUN masking on the CLARiiON storage array and the target storage array should be changed. As discussed earlier, VMware ESX/ESXi hosts should not be presented with two copies of the same VMware file system. The LUN masking database should ensure that the VMware ESX/ESXi hosts have access to the target devices only.

In addition, the zoning information active on the storage area network may need changes to ensure the VMware ESX/ESXi hosts has access to the storage array that hosts the remote devices.

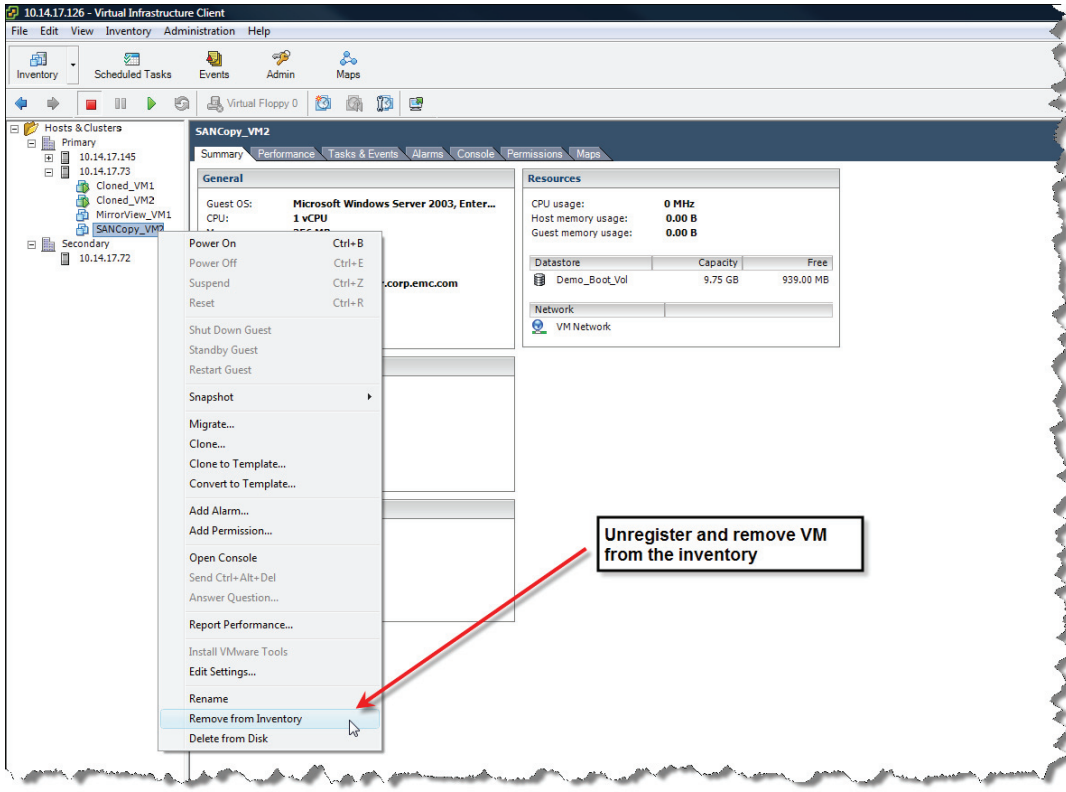


Figure 126 Removing virtual machines from a Virtual Infrastructure 3 environment as part of migration

6. The parameter, `LVM.EnableResignature`, should be set to 1 on one of the VMware ESX/ESXi hosts in the cluster group. As soon the SAN Copy completes the final push of the data from the source devices to the remote devices, a rescan of the SAN environment on those VMware ESX/ESXi hosts should be initiated. The VMware ESX/ESXi host recognizes the changes to the storage environment and resignatures and relabels the target devices as shown in Figure 127 on page 333 and Figure 128 on page 334.

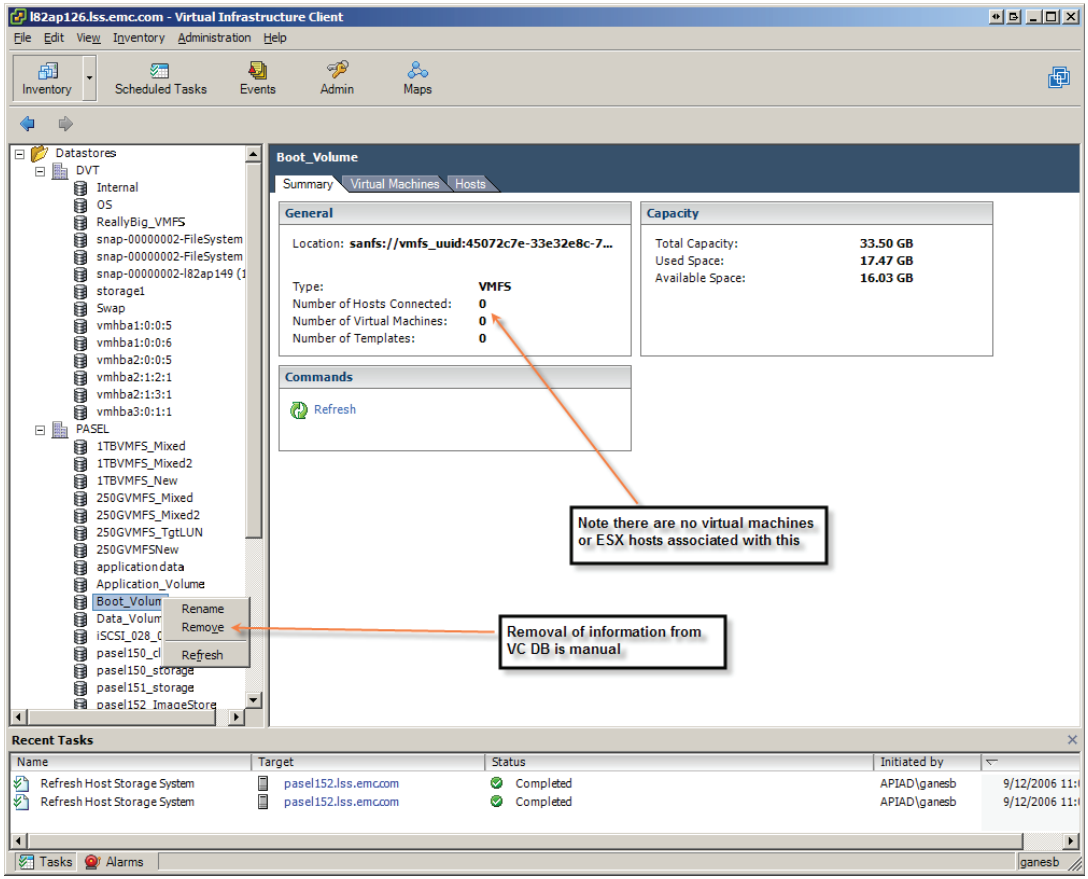


Figure 127 Removing datastore information from a vCenter infrastructure

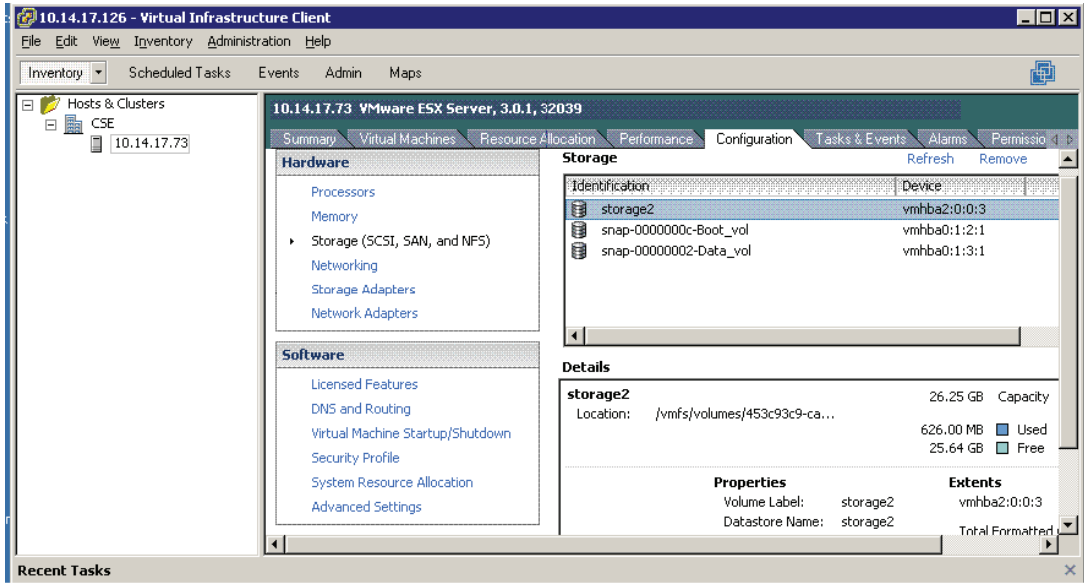


Figure 128 Resignaturing and relabeling of target devices

7. The VMware file system label on the target devices should be relabeled to the original name. This is shown in Figure 129 on page 335, where the name is changed from snap-0000000c-Boot_vol to Boot_vol. The renaming back to the original is possible since the vCenter infrastructure no longer has any reference to that name.

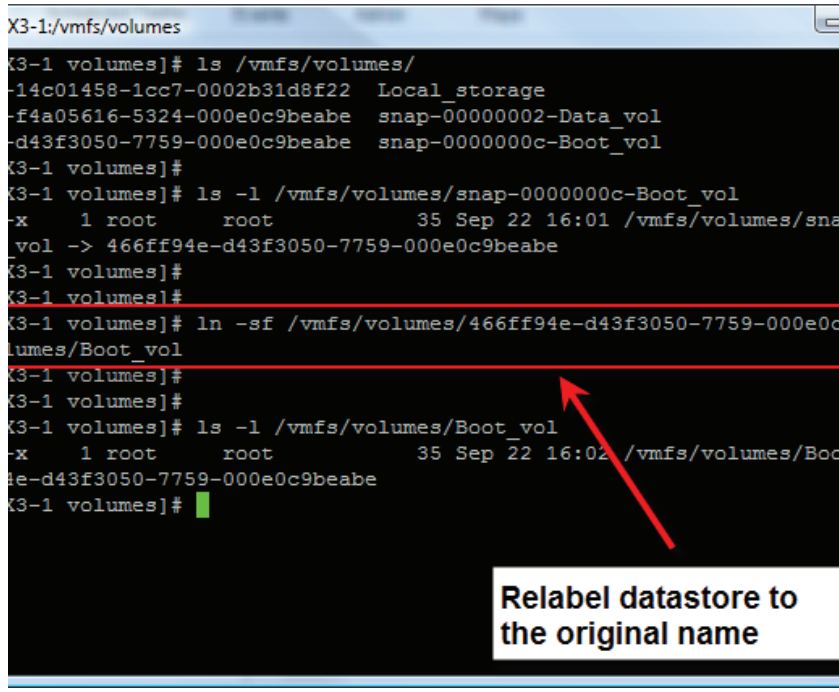


Figure 129 Renaming the relabeled datastore back to the original name

8. The configuration files for the virtual machines on the migrated volume should be used to register the virtual machines. The action taken in step 3 ensures that the virtual machines are reregistered back with their original names. If step 3 is not executed, vCenter automatically adds (1) to the end of the display names for the virtual machines. The process to register the virtual machines is shown in Figure 130 on page 336. The Virtual Infrastructure client can also be used to register the virtual machines.

9. The virtual machines, as shown in Figure 130 on page 336, can be started as soon as the machines are registered.

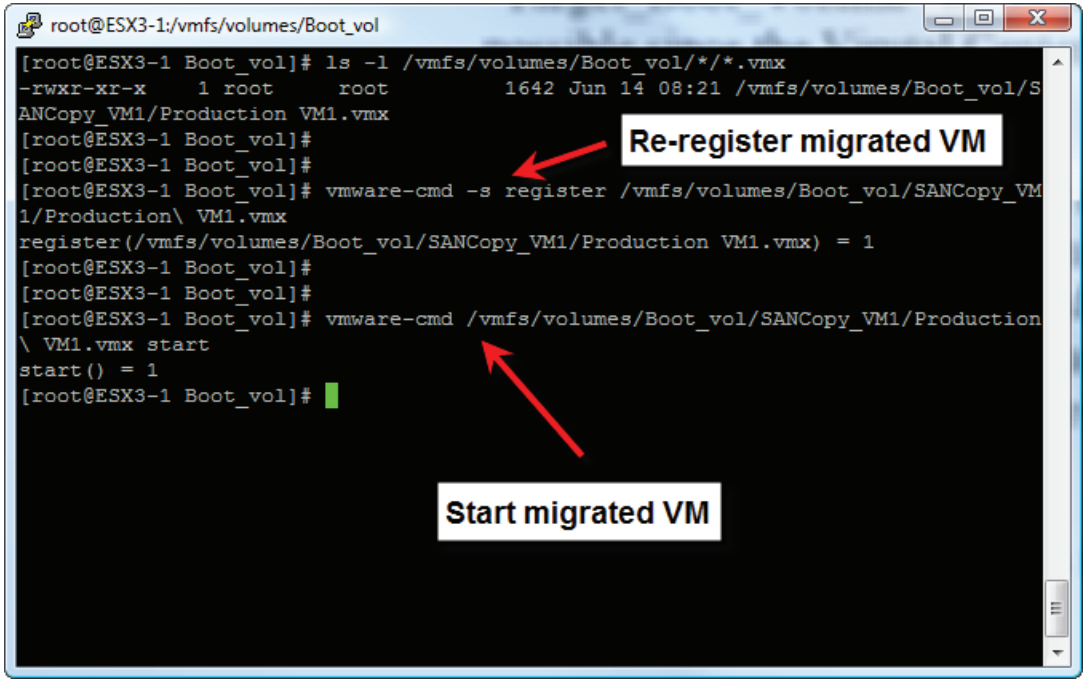


Figure 130 Reregistering and starting virtual machines on migrated volumes

It should be obvious from the steps listed here that the amount of outage when migrating data using SAN Copy is approximately equal to the time required for the final incremental push of data from the control devices to the remote devices. As it can be expected, even for a very large migration that involves terabytes of data, with careful planning the amount of time the virtual machines needs to be shut down can be reduced to a few minutes. Migration using VMware native tools, such as cp and vmkfstools, may require hours of downtime as each virtual machine is migrated from CLARiiON storage array to the target storage arrays.

Migration of a VMware file system in ESX version 4

The process to migrate VMware file system version 4 to a CLARiiON array using SAN Copy is the same process described in "Data vaulting of VMware file system using SAN Copy," on page 317. A few additional steps, which are listed below, are needed to handle the new functionality introduced in the vSphere 4.x environment:

1. After you power off the virtual machines (step 3 in "Data vaulting of VMware file system using SAN Copy," on page 317), remove the virtual machines from the vCenter infrastructure inventory.
2. You need to remove the original datastore from the inventory after you remove the original LUNs from the VMware ESX/ESXi cluster group (step 5 in "Data vaulting of VMware file system using SAN Copy," on page 317). This step is unnecessary if you are using VMware ESX/ESXi hosts version 3.0.1 and vCenter version 2.0.1 or later.
3. You need to rescan the SCSI bus after providing the VMware ESX/ESXi hosts with access to the migrated devices. You can scan the SCSI bus using the service console or the vCenter client.
4. The vCenter client **Add storage** wizard will list the devices holding the copy of the VMware file systems replicated from the source devices. Select the **Keep existing signature** option for each migrated LUN. After this option is selected for all LUNs, each VMware filesystem with the original name will be displayed in the "Storage" tab of the vClient interface. Choosing the **Keep existing signature option** prevents users from manually editing the VM config file (.vmx) for the existing virtual disks on the virtual machine.
5. Use the configuration files on the migrated volumes to add the virtual machines back to the vCenter infrastructure inventory. You can do this using the Virtual Infrastructure client or the service console. Do this after step 5 in "Migration of a VMware file system in ESX version 3 ," on page 330, after relabeling of the VMware file system on the remote devices.

Migration of devices used as RDM

The steps required to perform the migration of data is the same as that listed in "SAN Copy for data migration from CLARiiON arrays," on page 330.

Configuration files of virtual machines that utilize RDM do not need to be updated. However, changes to the mapping files are required to ensure successful migration. When a RDM is generated, a virtual disk is created on a VMware file system. This virtual disk does not contain any data. However, it maintains a list of pointers back to the addresses on the physical disk that it maps. In addition, the virtual disk also contains information about the unique identifier of the physical disk that it maps.

When the data for virtual machines containing RDM is migrated using the process described in "SAN Copy for data migration from CLARiiON arrays," on page 330. The virtual disk denoting the Raw Device Mapping points to a device that does not exist. As a result, the virtual machine cannot be powered on. To ensure this does not occur the following steps are needed:

- The existing RDM should be deleted before the source devices are removed from the VMware ESX/ESXi hosts (step 5 in "Data vaulting of VMware file system using SAN Copy," on page 317). This can be achieved by using the `rm` command on the service console or by using the -U option to `vmkfstools` utility.
- The RDM should be re-created utilizing the canonical name of the remote devices after the devices are discovered on the VMware ESX/ESXi hosts (step 5 in "Data vaulting of VMware file system using SAN Copy," on page 317). The virtual disk created during this process should have the same name as the one deleted in the previous step.

SAN Copy for data migration to CLARiiON arrays

SAN Copy provides various modes of operation. In addition to the incremental copy mode, SAN Copy supports the full copy mode in which data from a supported storage system can be migrated to the CLARiiON. The full copy option requires the source devices to be offline since SAN Copy does not support incremental pull from remote storage arrays.The following process needs to be followed when migrating VMware virtual infrastructure data to EMC CLARiiON arrays from supported storage arrays:

1. The first step in any migration process that uses SAN Copy is the identification of the WWN of the source devices. The management software for the source storage array should be used for this. The device numbers of the CLARiiON LUNs involved in the migration should also be identified.

2. After the appropriate information about the devices has been obtained, a full SAN Copy session of the clone volume on the remote array needs to be created. Figure 131 on page 340 displays the options necessary to create a full SAN Copy session.

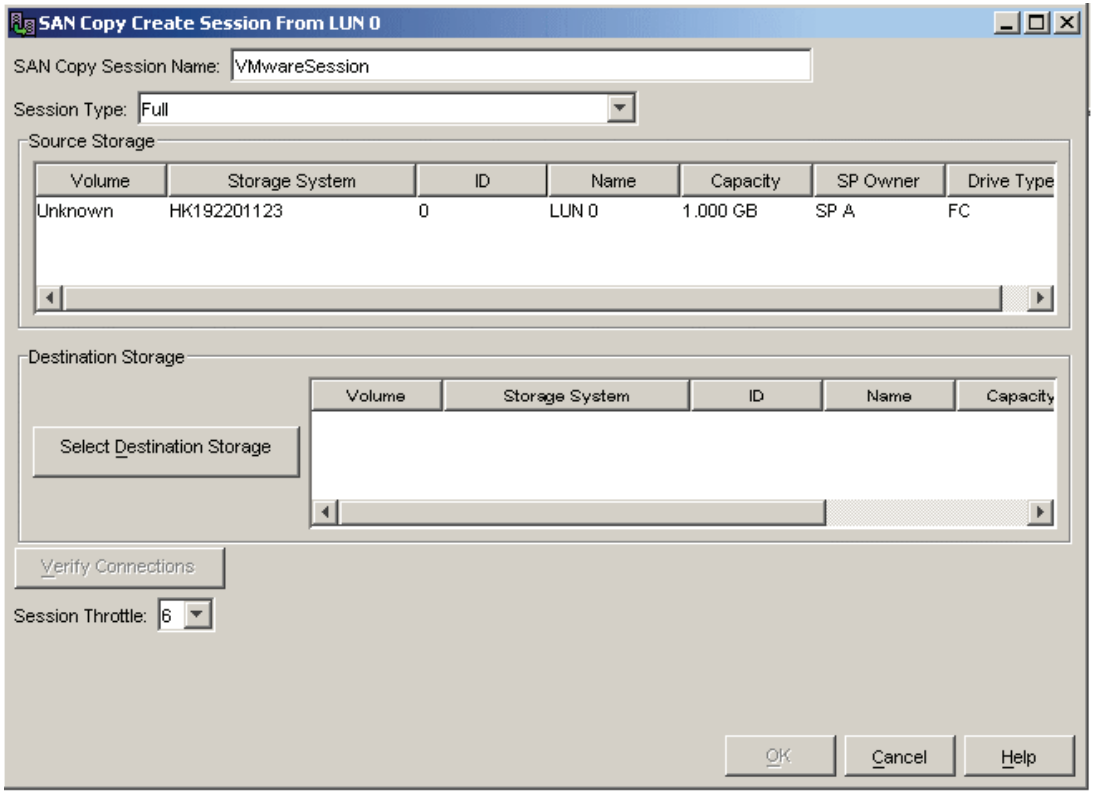


Figure 131 Creating a SAN Copy session for migrating data to a CLARiiON storage array

3. The virtual machines using the devices that are being migrated should be shut down. The SAN Copy session created in the previous step should be started to initiate the data migration from the source devices to the CLARiiON devices.
4. The LUN masking information on both the remote storage array and the CLARiiON array should be modified to ensure the VMware ESX/ESXi hosts have access to just the devices on the CLARiiON.

Note that the zoning information may also need to be updated to ensure the VMware ESX/ESXi hosts have access to the appropriate front-end Fibre Channel ports on the CLARiiON storage system.

5. After the full SAN Copy session completes, a rescan of the fabric on the VMware ESX/ESXi hosts enables the servers to discover the remote devices on the CLARiiON. The VMware ESX/ESXi hosts also update the /vmfs structures automatically.

6. After the remote devices have been discovered, the virtual machines can be restarted. Note that the discussion about virtual machines utilizing unlabeled VMFS or raw devices also applies for the migrations discussed in this section.

When the amount of data being migrated from the remote storage array to a CLARiiON array is significant, SAN Copy provides a convenient mechanism to leverage storage array capabilities to accelerate the migration. Thus, by leveraging SAN Copy, one can reduce the downtime significantly while migrating data to CLARiiON arrays.

Migration of a VMware file system in ESX version 3

The process to migrate VMware file system in ESX version 3 to a CLARiiON array using SAN Copy is the same process that was described in "Data vaulting of VMware file system using SAN Copy," on page 317. A few additional steps are needed to handle the new functionality introduced in the Virtual Infrastructure 3 environment. The process is similar to that discussed in "Data vaulting of VMware file system using SAN Copy," on page 317. Specifically the following steps need to be taken:

1. After the virtual machines are powered off (step 3 in "Data vaulting of VMware file system using SAN Copy," on page 317), the virtual machines should be removed from the vCenter infrastructure inventory.

2. The original datastore should be removed from the inventory after the original LUNs are removed from the VMware ESX/ESXi cluster group (after step 5 in "Data vaulting of VMware file system using SAN Copy," on page 317). This step is unnecessary when the environment involves VMware ESX/ESXi version 3.0.1and vCenter version 2.0.1 or later.

3. The parameter, `LVM.EnableResignature`, should be enabled on one of the VMware ESX/ESXi hosts in the cluster before forcing the ESX server to scan the SAN environment (before step 5 in "Migration of a VMware file system in ESX version 3 ," on page 330).
4. The label on CLARiiON devices to which the data has been migrated should be changed back to the original name (after step 5 in "Data vaulting of VMware file system using SAN Copy," on page 317 and removal of the original datastore name from the vCenter infrastructure).
5. The configuration files on the migrated volumes should be used to add the virtual machines back to the vCenter infrastructure inventory. This can be done using the Virtual Infrastructure client or the service console (after step 5 in "Migration of a VMware file system in ESX version 3 ," on page 330 and relabeling of the VMware file system on the remote devices).

Migration of a VMware file system in ESX version 4

The process to migrate VMware file system version 4 to a CLARiiON array using SAN Copy is the same as the process described in "Data vaulting of VMware file system using SAN Copy," on page 317. A few additional steps are needed to handle the new functionality introduced in the vSphere 4.x environment and are listed below:

1. After you power off the virtual machines (step 3 in "Data vaulting of VMware file system using SAN Copy," on page 317), remove the virtual machines from the vCenter infrastructure inventory.
2. After you remove the original LUNs from the VMware ESX/ESXi cluster group, remove the original datastore from the inventory. (You would do this after step 5 in "Data vaulting of VMware file system using SAN Copy," on page 317.) You do not need to do this if you are using VMware ESX/ESXi version 3.0.1 and vCenter version 2.0.1 or later.
3. Using the service console or the vCenter client, you need to scan the SCSI bus after providing the VMware ESX/ESXi hosts with access to the migrated devices.
4. Use the vCenter client **Add storage** wizard to list the devices holding the copy of the VMware file systems replicated from the source devices. Select the Keep existing signature option for each

migrated LUN. After this option is selected for all LUNs, each VMware filesystems will be will displayed under the **Storage** tab of the vClient interface.

5. Use the configuration files on the migrated volumes to add the virtual machines back to the vCenter infrastructure inventory. You can do this using the Virtual Infrastructure client or the service console. (You should do this after step 5 in "Migration of a VMware file system in ESX version 3 ," on page 330, after the relabeling of the VMware file system on the remote devices.)

Migration of devices used as RDM

If virtual machine access devices are using Raw Device Mapping, the mapping should be deleted and re-created to map to the CLARiiON devices to which the data has been migrated. The deletion of the existing raw disk map should be executed on the service console after the virtual machines are shut down but before the access to the remote volumes is removed. The raw disk map should be re-created before the virtual machines are powered on, pointing to the appropriate CLARiiON devices.

Migration of remote volumes accessed as RDM have the same considerations as discussed in "Migration of devices used as RDM ," on page 337.

A
Nondisruptive Expansion of a MetaLUN

This appendix presents the following topics:

- Introduction 346
- Expanding CLARiiON LUNs 347
- Growing VMware file system 3 in Virtual Infrastructure 3 348
- Growing VMware file system 3 in vSphere 4 353
- Growing RDMs in vSphere 4 and Virtual Infrastructure 3 356

Introduction

CLARiiON storage arrays support expansion of CLARiiON LUNs or metaLUNs. The expansion can be performed by recreating the LUN with additional members nondisruptively while preserving the existing data. Once expanded, the LUN is referred to as a metaLUN. A metaLUN can also be expanded using the same process.

This appendix focuses on the use of the nondisruptive CLARiiON LUNs expansion using the metaLUN technology process to grow VMware file system (VMFS) and RDM volumes.

Expanding CLARiiON LUNs

A CLARiiON LUN or metaLUN can be expanded using the Navisphere Manager or Navisphere CLI. An existing LUN can be extended by executing the following command:

```
naviseccli -h SPipaddress metalun –expand –base
<number|WWN> -lus <lunnumber|WWN> -expansionrate
<low|medium|high> -type <C|S>
```

where, `-base` indicates the LUN or metaLUN that needs to be expanded, `-lus` indicates the LUN number that needs to be added to the LUN or metaLUN. The type "C" or "S" indicates whether the concatenation or striping option should be used for expansion.

There are other options that can be used when expanding LUNs or metaLUNs. The Navisphere Command Line Interface manual available on Powerlink provides more information on executing metaLUN commands using the Navisphere CLI.

Growing VMware file system 3 in Virtual Infrastructure 3

1. This section explains how to expand a VMware file system by utilizing the nondisruptive metaLUN technology to grow an existing LUN presented to VMware ESX/ESXi hosts. The first step in the expansion process is the identification of the CLARiiON LUN that hosts the VMware file system. You can use VM-aware Navisphere or Navisphere Agent/CLI software to obtain the mapping information.
2. The CLARiiON LUN should be expanded with the additional CLARiiON LUNs as shown in Figure 132 on page 348. Navisphere CLI can also be used to expand a CLARiiON LUN or metaLUN.

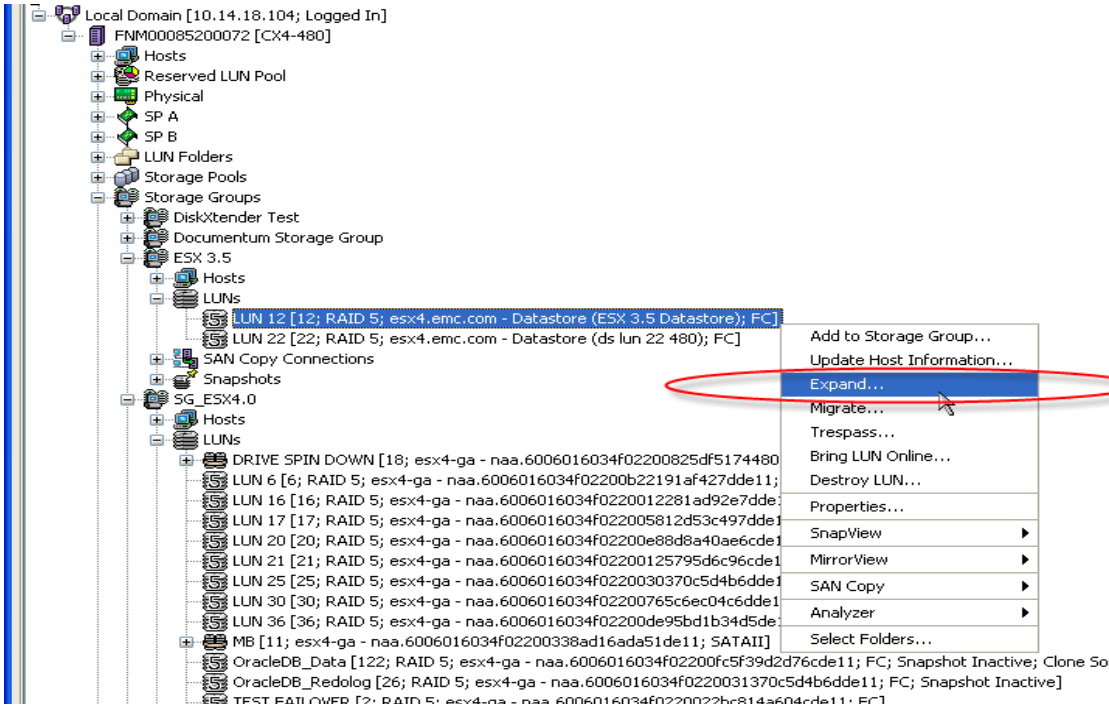


Figure 132 Identifying a CLARiiON LUN from Navisphere Manager to be expanded

3. After the expansion of the LUN completes successfully, as shown in Figure 133 on page 349, the SCSI bus should be rescanned. The extra space available on the storage device can be seen in the figure. Partition 1, which contains the VMware file system Expand_LUN,

does not occupy the whole disk. The extra space left from cylinder number 4402 to cylinder number 4951 is used to grow the VMware file system

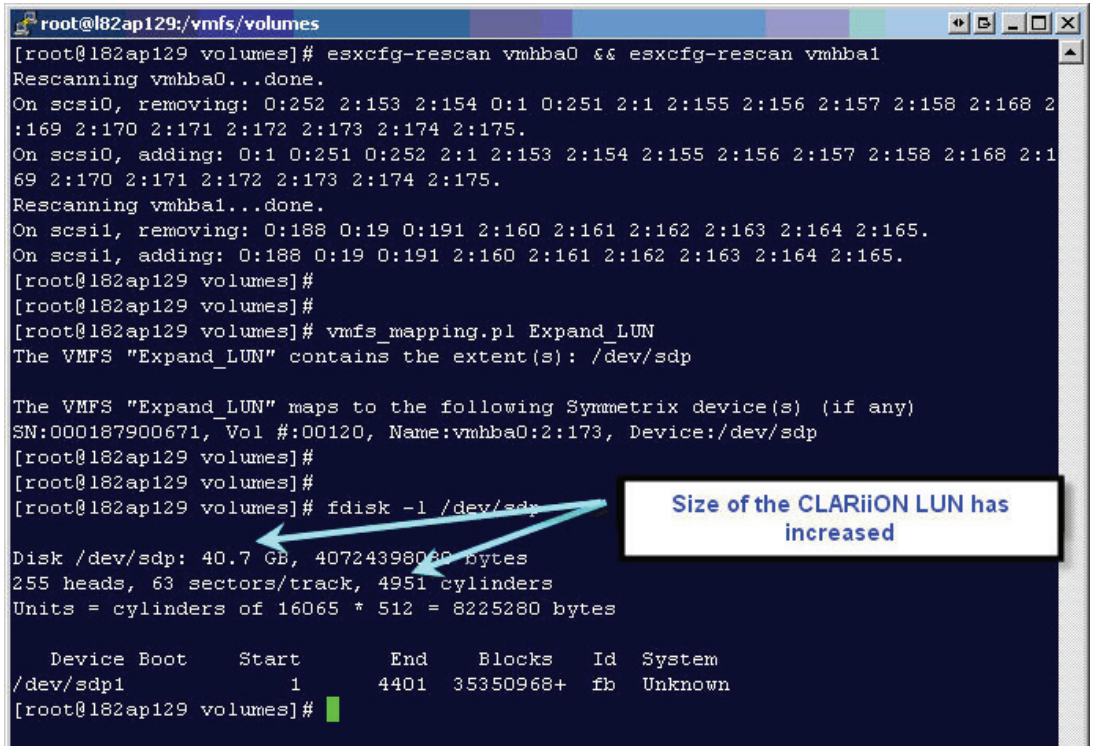


Figure 133 Forcing VMkernel to recognize changes to the device configuration

4. The free space at the end of the disk can be used to grow the VMware file system. The growth of the file system is achieved by creating a new partition on the free space, and adding that partition as a physical extent to the file system. The virtual machines need to be powered off to expand a VMware file system version 3.

5. Figure 135 on page 351 is an exhibit that shows the process listed in this step. Although the figure shows the use of the Virtual Infrastructure client to expand the VMware file system, the same result can be achieved by using the service console. It can be seen from the figure that a new partition (partition number 2) has been added to the CLARiiON LUN. This partition appears as the second physical extent of the VMware file system.

6. Note that the virtual machines on the VMware file system need to be powered off or suspended before expanding the VMware file system. Figure 134 on page 350 shows the error message that willappear when a VMware file system is extended to occupy an extended LUN when virtual machines that are provisioned on that datastore are powered on.

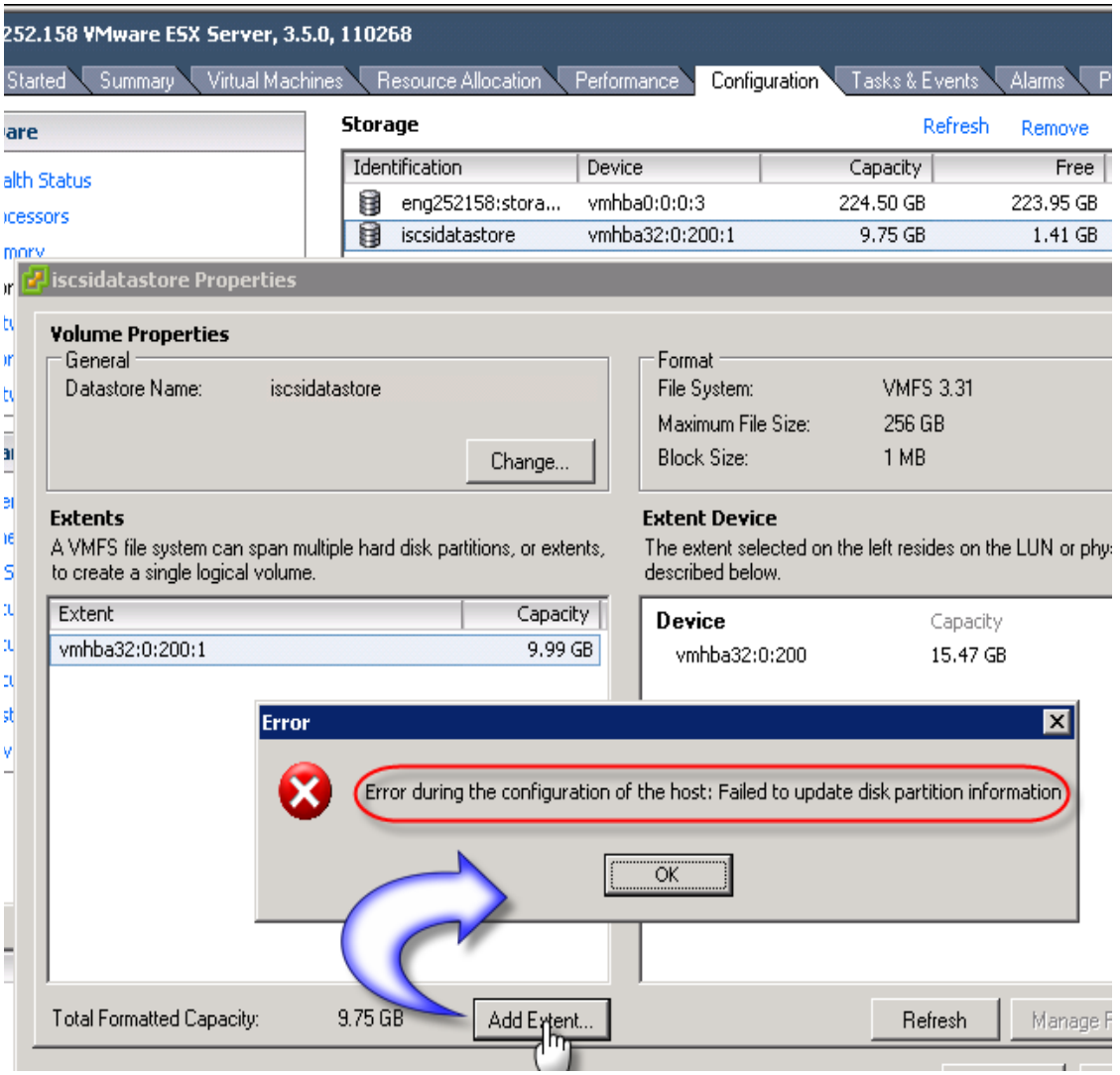


Figure 134 VMware file system error following a CLARiiON LUN expansion with virtual machines powered on

7. Figure 135 on page 351 shows the details of the VMware file system, Expand_LUN. Comparing the details in this figure with those in Figure 133 on page 349 how how the VMware file system has grown in size from approximately 34 GB to 38 GB. The figure also shows that the data originally on the VMware file system was preserved during the expansion process.

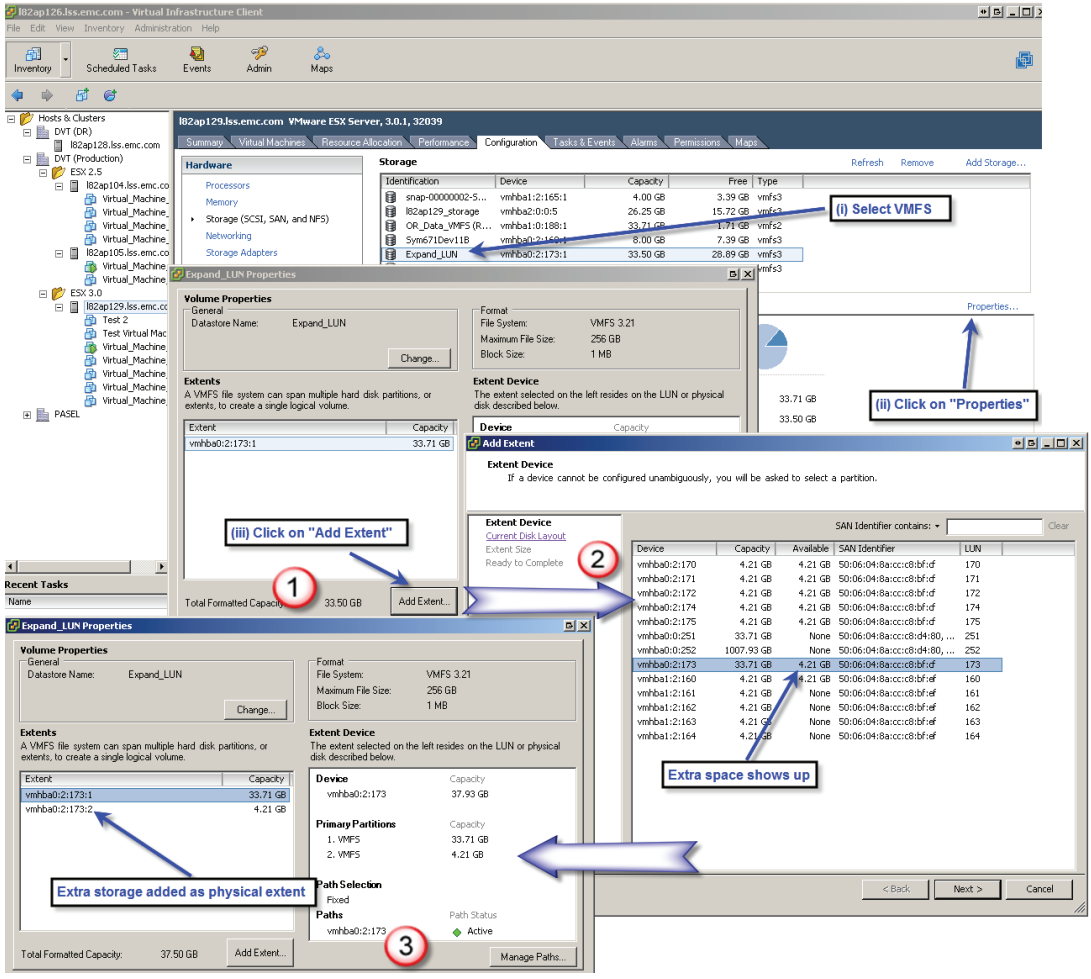


Figure 135 Expanding VMware file system version 3 using the vCenter client

Figure 136 Details of the expanded VMware file system

8. With ESX 3.x, 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

Growing VMware file system 3 in vSphere 4

This section describes the process to extend a VMware file system version 3 in vSphere 4.0, and highlights the differences with the process for VMware Infrastructure 3.

1. The first step in the expansion process is to identify the CLARiiON LUN that hosts the VMware file system. You can use VM-aware Navisphere or Navisphere Agent/CLI software to obtain the mapping information.
2. Expand the CLARiiON LUN with the additional CLARiiON LUNs as shown in Figure 132 on page 348. You can also use Navisphere CLI to expand a CLARiiON LUN or metaLUN.
3. After the expansion of the LUN completes successfully, as shown in Figure 133 on page 349, you need to rescan the SCSI bus using vClient or the esx-cfg-rescan option on the ESX host.
4. To extend the LUN complete the following steps in vClient
   a. Click the Cofiguration tab and click **Storage.**
   b. From the Datastores view, select the datastore to increase, click **Properties,** and then click **Increase a**s shown in Figure 137 on page 354.
   c. Select a device from the list of storage devices and click **Next**.
   d. To expand an existing extent, select the device for which the expandable column reads “Yes”
   e. Select a configuration option from the bottom panel.
   f. Set the capacity for the extent and click **Next.**
   g. Review the proposed layout and the new configuration of the datastore, and click **Finish**.
5. The VMFS datastore size automatically increases as needed. With ESX 4.0, the VMFS datastore size can be increased to almost a 2 TB limit while the virtual machines are powered and 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.
6. 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, as shown in Figure 138 on page 355. 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.

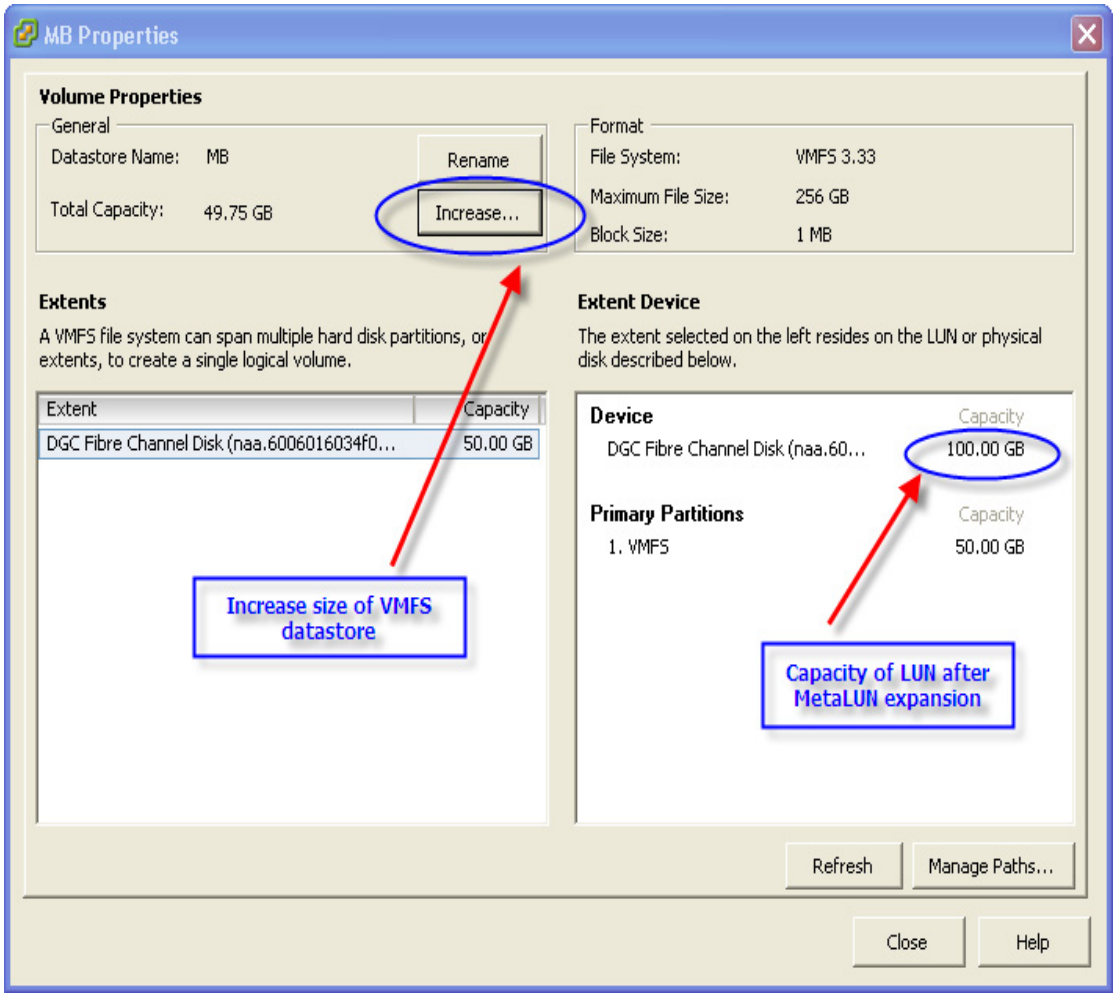


Figure 137 VMFS datastore "Properties" dialog for LUN expansion on ESX 4.0

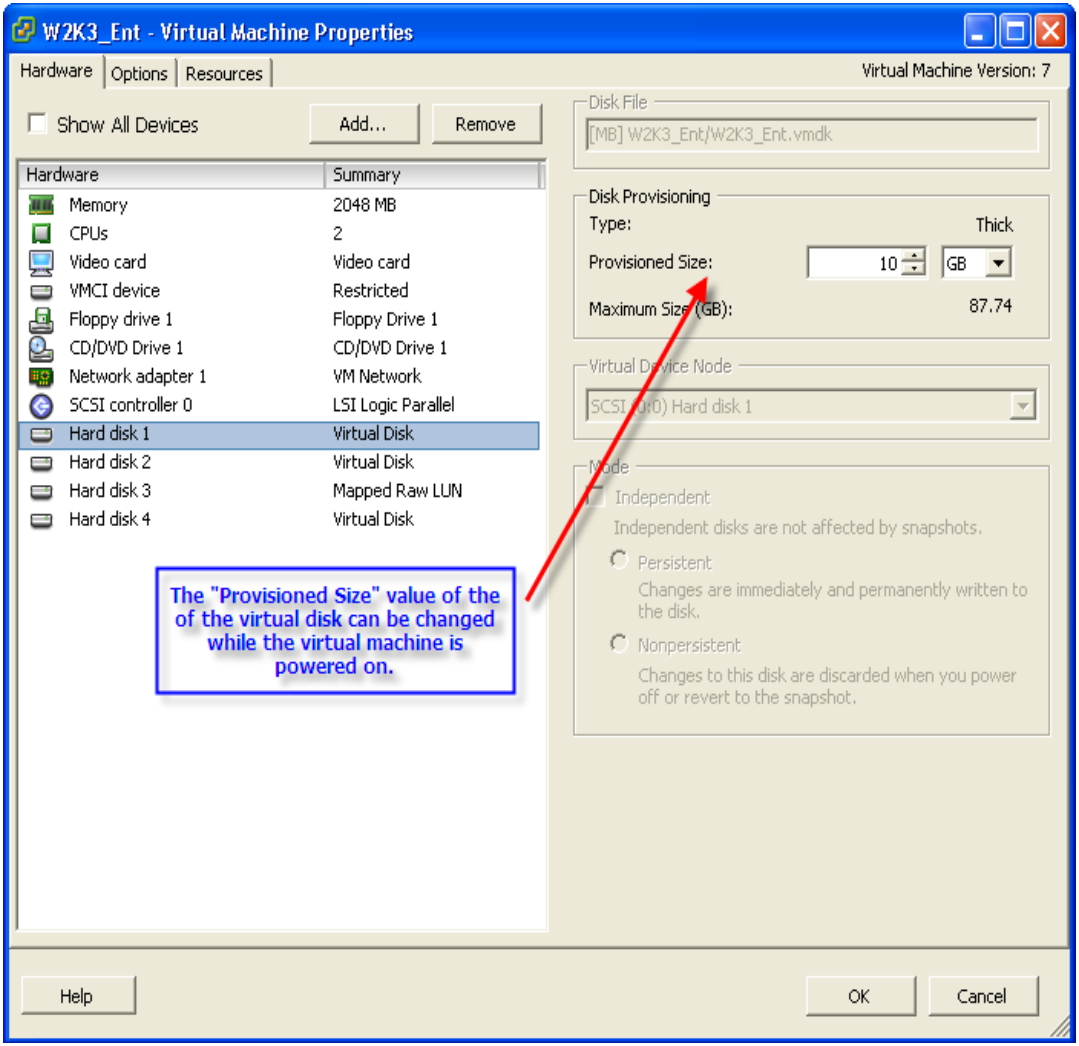


Figure 138 Hot virtual disk expansion through "Edit settings" of the virtual machine

Growing RDMs in vSphere 4 and Virtual Infrastructure 3

A LUN presented as an RDM to a virtual machine on a VMware ESX/ESXi (ESX 4.x or 3.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 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.

Note that RDM volumes must use the physical compatibility mode for expansion when using the CLARiiON metaLUN technology.