WHITE PAPER

VMware High Availability

Concepts, Implementation, and Best Practices

vmware®

Introduction

VMware Infrastructure 3 is the first full infrastructure virtualization suite to empower enterprises and small businesses alike to transform, manage, and optimize their IT infrastructure through virtualization. VMware Infrastructure 3 delivers comprehensive virtualization, management, resource optimization as well as application availability and operational automation capabilities in an integrated offering. VMware High Availability (VMware HA) helps customers improve service levels for any application by implementing a cost-effective virtualization-based high-availability solution that is easy to deploy and manage.

This paper provides an architectural and conceptual overview of VMware HA and describes how you can use VMware HA to provide high availability for any applications running in virtual machines at lower cost than would be possible with static, physical infrastructure. When you use VMware HA, you can automatically restart virtual machines in the event of hardware failure without investing in costly one-to-one mapping of production and backup hardware.

This paper is intended for VMware partners, resellers, and customers who want to implement virtual infrastructure solutions and want to know how to use distributed infrastructure services such as VMware HA.

Introduction to VMware Infrastructure and VMware HA

With VMware Infrastructure 3, VMware extends the evolution of virtual infrastructure and virtual machines that began with VMware ESX Server 1.0. VMware Infrastructure 3 brings a revolutionary new set of infrastructure wide services for resource optimization, high availability, and data protection that deliver capabilities that previously required complex or expensive solutions to implement using only physical machines.

Use of these services provides significantly higher hardware utilization and better alignment of IT resources with business goals and priorities.

VMware Infrastructure introduces two new concepts:

- Clusters that aggregate and manage the combined resources of multiple hosts as a single collection.
- Resource pools that simplify control over the resources of a host or a cluster.

VMware Infrastructure virtualizes and aggregates industry-standard servers (processors, memory, their attached network and storage capacity) into logical resource pools (from a single ESX Server host or from a VMware cluster) that can be allocated to virtual machines on demand.

Resource pools can also be nested and organized hierarchically so that the IT environment matches company organization. Individual business units can receive dedicated infrastructure while still profiting from the efficiency of resource pooling.

A set of virtualization-based distributed infrastructure services provides virtual machine monitoring and management to automate and simplify provisioning, optimize resource allocation, and provide operating system and application-independent high availability to applications at lower cost and without the complexity of solutions used with static, physical infrastructure. One of these distributed services, VMware HA, provides easy-to-manage, cost-effective high availability for all applications running in virtual machines. In the event of server hardware failure, affected virtual machines are automatically restarted on other physical servers. VMware HA minimizes downtime and IT service disruption while eliminating the need for dedicated standby hardware and installation of additional software. VMware HA also protects against multiple server failures in the cluster. Virtual machines are protected as long as there is spare capacity in the cluster.

VMware Infrastructure 3 reduced the frequency of outages for core infrastructure services by 100 percent, saving 416 man hours per year in unplanned maintenance. VMware Distributed Resource Scheduler and VMware High Availability are fully automated and performing flawlessly.

VMware HA is really a simpler and cost-effective alternative to complex, traditional clustering technologies.

- Faan DeSwardt, Director of Enterprise Architecture, Wyse Technology

While we initially chose VMware virtual infrastructure to address development hardware problems by reducing hardware costs and decreasing server deployment time, we soon discovered additional benefits to adopting the technology, including server portability, protection, and availability. All the advantages provided by VMware software have created an ideal work environment for our developers, allowing them to develop applications more efficiently. With virtual machines, developers avoid downtime and bring products to the market faster. We also envision that the flexible and efficient virtual infrastructure will allow even greater disaster recovery options, virtual labs, training, and desktop portability.

-Keith Leahy, Vice President, Merrill Lynch

VMware HA Architecture and Conceptual Overview

The following sections provide basic information about VMware Infrastructure and describe some of the key elements with which VMware distributed services such as VMware HA interact.

VMware Infrastructure

At the core of VMware Infrastructure, VMware ESX Server is the foundation for delivering virtualization-based distributed services to IT environments. ESX Server provides a robust virtualization layer that abstracts processor, memory, storage, and networking resources into multiple virtual machines that run side-by-side on the same physical server.

ESX Server installs directly on the server hardware, or "bare metal," and inserts a robust virtualization layer between the hardware and the operating system. ESX Server partitions a physical server into multiple secure and portable virtual machines that run on the same physical server. Each virtual machine represents a complete system — with processors, memory, networking, storage and BIOS — so Windows, Linux, Solaris, and NetWare operating systems and software applications run in virtualized machines without any modification.

Another key building block of VMware Infrastructure, VirtualCenter, is used to manage all ESX Server hosts and virtual machines. VirtualCenter Management Server also provides critical services such as:

- Centralized server and virtual machine management
- Virtual machine provisioning
- Performance monitoring
- Operational automation
- Secure access control
- Migration of live virtual machines

Figure 1 shows the architecture and typical configuration of VMware Infrastructure.

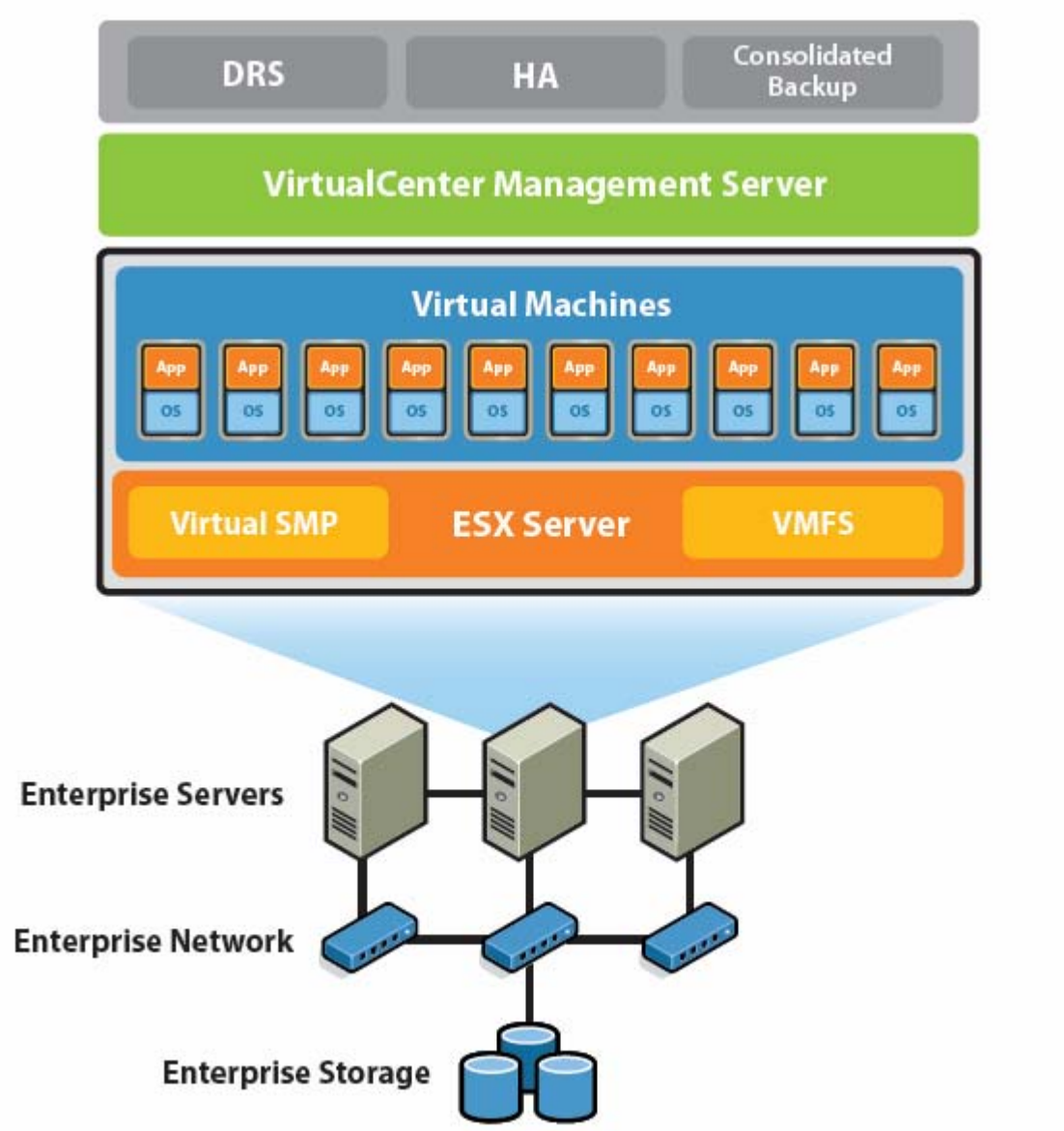

Figure 1: VMware Infrastructure configuration

VMware Infrastructure simplifies management with a single client called the Virtual Infrastructure Client (VI Client) that you can use to perform all tasks. Every ESX Server configuration task, from configuring storage and network connections to managing the service console, can be accomplished centrally through the VI Client.

The VI Client connects to ESX Server hosts, even those not under VirtualCenter management, and lets you remotely connect to any virtual machine for console access. There is a Windows version of the VI Client, and for access from any networked device, a Web browser application provides virtual machine management and VMware Console access. The browser version of the client, Virtual Infrastructure Web Access, makes it as easy to give a user access to a virtual machine as sending a bookmark URL.

VirtualCenter user access controls provide customizable roles and permissions, so you create your own user roles by selecting from an extensive list of permissions to grant to each role.

Responsibilities for specific VMware Infrastructure components such as resource pools can be delegated based on business organization or ownership. VirtualCenter also provides full audit tracking to provide a detailed record of every action or operation performed on the virtual infrastructure and who did it

Users can also access virtualization-based distributed services provided by VMotion™, VMware DRS, and VMware HA directly through VirtualCenter and the VI Client. In addition, VirtualCenter exposes a rich programmatic Web Service interface for integration with third-party system management products and extension of the core functionality.

- VMware VMotion enables the live migration of running virtual machines from one physical server to another. Live migration of virtual machines enables companies to perform hardware maintenance without scheduling downtime and disrupting business operations. VMotion also allows the mapping of virtual machines to hosts to be continuously and automatically optimized within clusters for maximum hardware utilization, flexibility, and availability.
- VMware DRS works with VMotion to provide automated resource optimization and virtual machine placement and migration to help align available resources with predefined business priorities while maximizing hardware utilization.
- VMware HA enables broad-based, cost-effective application availability, independent of specific hardware and operating systems.
- VMware Consolidated Backup provides an easy-to-use, centralized facility for LAN-free backup of virtual machines. Full and incremental file-based backup is supported for virtual machines running Microsoft Windows operating systems. Full image backup for disaster recovery scenarios is available for all virtual machines regardless of guest operating system.

VMware Clusters

Clusters, a new concept in virtual infrastructure management, give you the power of multiple hosts with the simplicity of managing a single entity. New cluster support in VMware Infrastructure 3 reduces management complexity by combining standalone hosts into a single cluster with pooled resources and inherently higher availability.

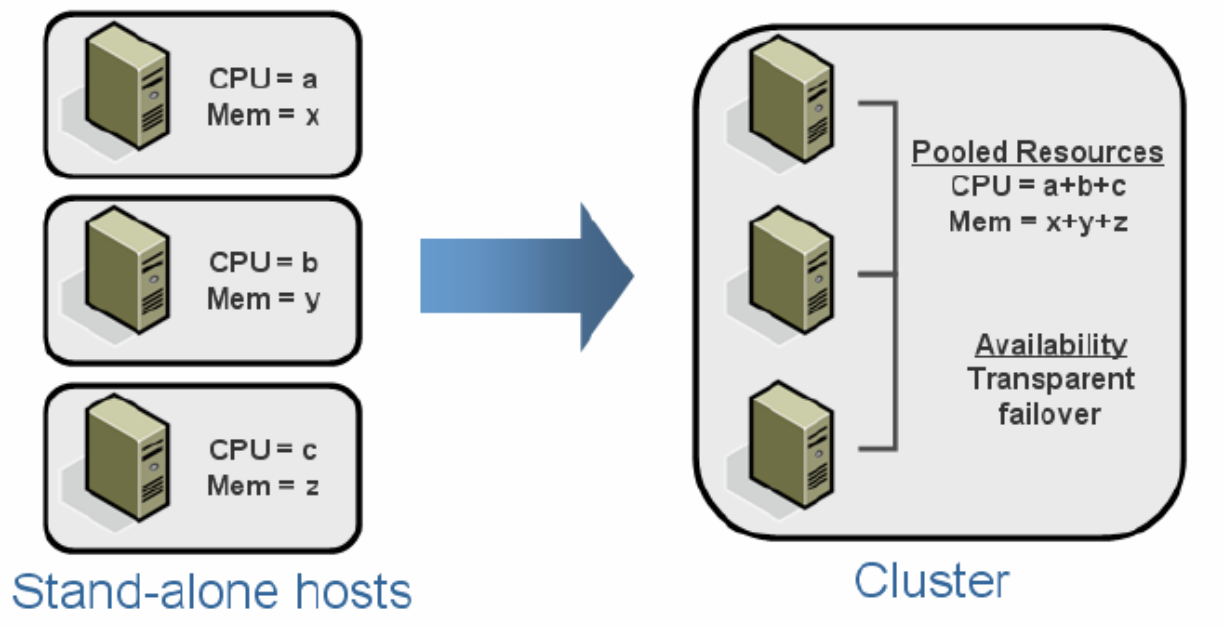

Figure 2: Resource aggregation in VMware clusters

VMware clusters let you aggregate the hardware resources of individual ESX Server hosts but manage the resources as if they resided on a single host. Now, when you power on a virtual machine, it can be given resources from anywhere in the cluster, rather than be tied to a specific ESX Server host.

VMware Infrastructure 3 provides two services to help with the management of VMware clusters, VMware HA, and VMware DRS. VMware HA allows virtual machines running on specific hosts to be restarted automatically using other host resources in the cluster in the case of host machine failures. VMware DRS provides automatic initial virtual machine placement and makes automatic resource relocation and optimization decisions as hosts are added or removed from the cluster or the load on individual virtual machines goes up or down. VMware DRS also makes clusterwide resource pools possible.

Note: For more information about resource pools and using VMware DRS to manage operations such as virtual machine placement and providing dynamic resource allocation for virtual machines running on VMware cluster hosts, see the VMware Infrastructure 3 white paper titled "Resource Management with VMware DRS."

VMware HA

Traditional High Availability and Failover Solutions

Both VMware HA and traditional clustering and high availability solutions support automatic recovery from host failures. Fundamentally, VMware HA provides simpler, cost-effective high

availability across a broader range of your x86 IT infrastructure and workloads. VMware HA can be complementary to traditional application-based and operating-system–based clustering solutions, but they differ in hardware and software requirements, capital and operating expenses, time to recovery, and the degree to which they incorporate application and operating system awareness.

Capital and Operational Costs of Clustering

Traditional clustering solutions aim to provide immediate recovery with minimal downtime for applications in case of host or software failure. To achieve this, the IT infrastructure must be set up such that each machine (or virtual machine) has a mirror virtual machine. The machine (or the virtual machine and its host) are set up to mirror each other using the clustering software. Often the hardware must be identical across the cluster. Once set up, clustering software can then send constant heartbeats between mirrors. Application-aware clustering agents can monitor application services running on each virtual or physical machine. In case of failure on the primary host, the mirror takes over operations after going through a failover sequence.

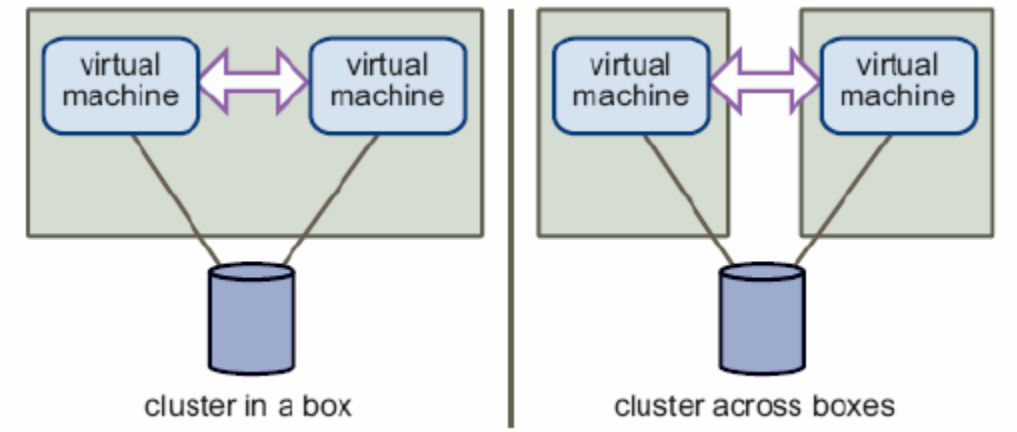

Figure 3: Traditional clustering solutions with virtual machines

Figure 3 shows the typical host setup for virtual machines using a traditional clustering approach. Setup and maintenance of traditional clustering solutions is expensive and expertise-intensive. Each time you add a new virtual machine, you may need additional hosts and additional virtual machines for failover. Traditional clustering tends to require complex systems and networking setup. It is generally resource-intensive both in terms of labor and systems requirements, because of the need for identical hardware. Traditional clustering expertise is hard to find and requires vast experience with application services, cluster configuration, and maintenance, because traditional cluster configurations often need recalibration because they tend to degrade over time.

VMware HA, which builds upon the proven consolidation benefits of VMware Infrastructure 3, automatically lowers the cost of high availability. VMware HA does not require dedicated cluster administrators or standby hardware. VMware HA makes clustering simple using a series of straightforward GUI-based selections within the VI Client to enable the cluster configuration. It is agnostic to the applications and operating system that run in the virtual machine. Failover is quick and seamless with recovery times typically of a few minutes.

Breadth of Application Coverage

Given the complexity, expertise requirements, and costs of traditional clustering software, it is typically focused on protecting only a few narrow bands of critical tier 1 applications such as enterprise databases, enterprise messaging, OLTP applications, and ERP applications in mission

critical production environments. This often leaves a vast majority of the infrastructure workloads (including middleware, home-grown applications, smaller local databases, file, print, fax authentication, DNS, and DHCP servers) exposed or unprotected. Many of these applications act as direct dependencies to your tier 1 applications. If they are offline, you may be unable to meet your SLA commitments for key applications and services.

VMware HA is independent of applications and operating systems and does not require application agents or manual scripting. It is simple and robust enough to protect all applications and x86 workloads that are fit for virtualization, not just a favored few applications. It is part of VMware Infrastructure 3 Enterprise, and the high availability capability builds on the core consolidation benefits available from the VMware Infrastructure platform.

Manageability and Flexibility

Traditional clusters are rigid, and usually management of a cluster requires taking applications offline whenever you need to reconfigure or maintain the cluster. Also, these clustering solutions are active-passive or active-active configurations in which standby server nodes in a cluster are idle, or they rigidly tie the cluster system to a particular application and operating system. VMware HA provides availability minus the complexity of traditional solutions. VMware HA delivers a highly manageable and flexible HA architecture that accommodates different application and operating system combinations, allows for N-N (any-to-any) failovers that are intelligently managed and can deliver online maintenance of physical server configurations without any application disruption. Moreover, VMware HA does not need any additional management tools. You manage it using the standard VI Client interface, making it easily manageable.

VMware Infrastructure: a Cost-Effective and Reliable Production Platform

To summarize, traditional clustering solutions can protect against application failures, but they are resource- and labor-intensive in addition to typically being application- and operating-system–dependent. Because of the cost and complexity of clustering solutions, they typically are used for a small percentage of enterprise applications, leaving the vast majority of applications without any failover protection, posing a potential risk to important business-critical services.

VMware HA, which is implemented at the level of ESX Server on the physical host, offers a seamless and simple solution to protect against hardware failures. Because the protection is offered to all virtual machines running in the ESX Server cluster, VMware HA "democratizes" high availability by making it available and cost-justifiable for any application.

Additionally, VMware Infrastructure is a mature production-proven platform and has rapidly evolved to provide robust availability capabilities that protect against a broad variety of planned and unplanned failure scenarios from component-level to system-level, data-level, and even site-level failover (see Figure 4).

<table>
<tr><th></th><th>Avoid Planned Outages</th><th>Quick Recovery from Unplanned Outages</th></tr>
<tr><td>Component</td><td colspan="2">NIC Teaming, Multipathing</td></tr>
<tr><td>Server</td><td>VMotion,
DRS + Maintenance Mode</td><td>VMware HA</td></tr>
<tr><td>Storage</td><td>Storage VMotion</td><td>Encapsulation,
VMware Consolidated Backup</td></tr>
<tr><td>Data</td><td>NA</td><td>Encapsulation,
VMware Consolidated Backup</td></tr>
<tr><td>Site</td><td colspan="2">VMware Site Recovery Manager</td></tr>
</table>

Figure 4: VMware Infrastructure components for business continuity

Planned and Unplanned Server Downtime

VMware HA protects applications contained in virtual machines from unplanned server failures. With VMware Distributed Resource Scheduling (VMware DRS) and VMotion technology, virtual machines can be live migrated to alternative hosts when you need to perform server maintenance, with zero application or service downtime.

Data Protection

VMware Consolidated Backup, a part of VMware Infrastructure, provides the required framework to protect and back up virtual machine files and images for longer term data protection and recovery, using a centralized host-free and LAN-free approach.

Site Recovery

VMware Site Recovery Manager, a new VMware product, works with VMware Infrastructure to provide automated disaster recovery solutions. Combined with VMware Infrastructure, it makes recovery faster, reliable, affordable, and manageable by making it possible for you to program, automate, and test disaster recovery plans that cover your entire production site. Site Recovery Manager is expected to be available in the first half of 2008.

How Does VMware HA Work?

VMware HA continuously monitors all ESX Server hosts in a cluster and detects failures. The VMware HA agent placed on each host maintains a heartbeat with the other hosts in the cluster using the service console network. Each server sends heartbeats to the others servers in the cluster at five-second intervals. If any servers lose heartbeat over three consecutive heartbeat intervals, VMware HA initiates the failover action of restarting all affected virtual machines on other hosts.

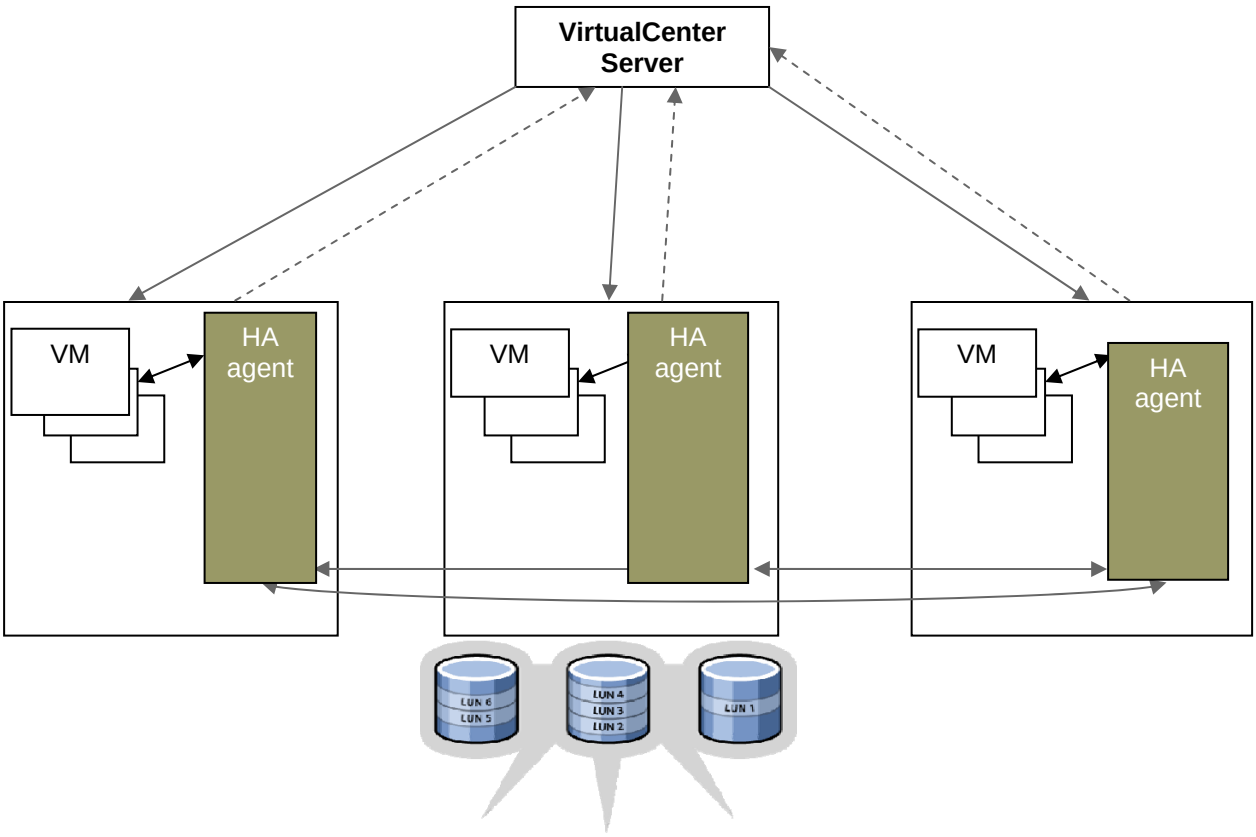

Figure 5: Host failover using VMware HA

VMware HA also monitors whether sufficient resources are available in the cluster at all times in order to be able to restart virtual machines on different physical host machines in the event of host failure. Safe restart of virtual machines is made possible by the locking technology in the ESX Server storage stack, which allows multiple ESX Server hosts to have simultaneous access to the same virtual machine files.

Planning VMware HA Clusters

When planning the size of VMware HA clusters to provide the desired levels of failover capacity, keep in mind that each host requires some overhead memory and CPU resources and that each virtual machine must be guaranteed its CPU and memory reservations. VMware HA factors in the worst-case failure scenarios when deciding to allow new virtual machines to be powered up. When computing required failover capacity, VMware HA first considers the host with the largest capacity to run virtual machines with the highest resource requirements. VMware HA might therefore be quite conservative in its estimates if the hosts in your cluster vary widely in the individual resources they provide.

Configuring for Redundancy

Real high availability is achieved with redundancy. Redundancy concepts do not apply only to key resources such as servers, networks, and storage. Management is also an important part of the high-availability configuration.

Using proven systems for servers is an important first step in implementing high availability. You can make servers more resilient to failure by proving better cooling, redundant power supplies, and the like. To make shared storage more fail-proof, you can implement multi-pathing from servers to storage arrays. Mirroring storage volumes also offers additional protection against storage failures.

Heartbeats are a critical element of VMware HA configuration. Redundant paths for heartbeats are thus essential for highest reliability of VMware HA. You can choose from many options to create multiple paths for heartbeats. Whichever way you create redundancy for heartbeats, testing failure scenarios is essential. You must also plan and test resiliency for the networking and storage switches which are part of the infrastructure. Information on how to create redundant network paths is included in the best practices section of this paper.

Even though VMware HA agents operate independently of VirtualCenter and can perform failovers without accessing VirtualCenter first, centralized management views of the infrastructure are essential for smooth functioning of any business. It is thus critical to protect VirtualCenter for resiliency.

Designating Failover Capacity

When you enable a cluster for VMware HA, the New Cluster Wizard prompts you to specify the maximum number of host failures you want to protect against. This number is shown as the Configured Failover Capacity in the Virtual Infrastructure Client. VMware HA uses this number as it continuously monitors whether there are enough resources to power on virtual machines in the cluster. You need to specify only the number of hosts for which you want failover capacity.

VMware HA computes the resources that it requires to fail over virtual machines with the specified failover capacity. This resource determination is based on the virtual machines' configured CPU and memory resource reservations and capability to handle the failure of the largest hosts in the cluster. It helps to have more-uniform hosts in the cluster, for example, to avoid situations in which virtual machines do not have enough resources to be restarted on new hosts. When the number of host failures exceeds configured spare capacity, virtual machines with the highest priorities are failed over first.

The figure below shows two host failure scenarios. Notice that on the there is only enogh capacity available to restart all virtual machines running on only one server. If more than one server fails, some low-priority virtual machines will not restart. It is thus important to carefully consider resources when planning for HA clustering.

Note: You can choose to allow the cluster to power on virtual machines even when they violate availability constraints; however, this means that failover guarantees may no longer be valid.

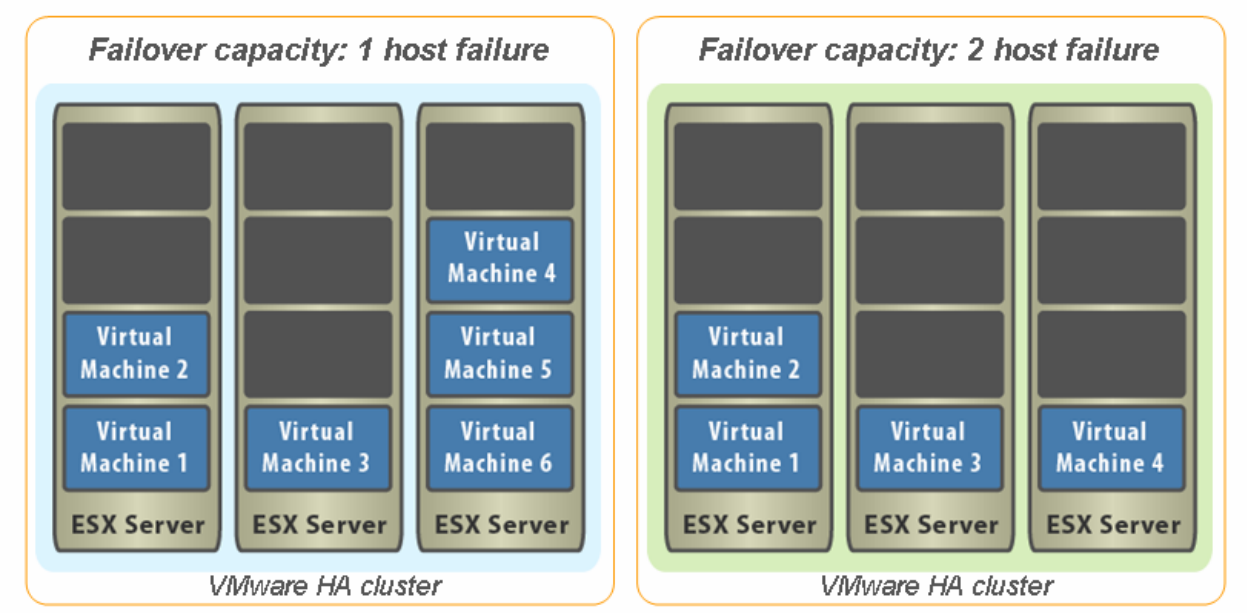

Figure 6: Failover capacity

Clusters and VirtualCenter Failure

You create and manage clusters using VirtualCenter. The VirtualCenter Management Server places an agent on each host in the cluster so each host can communicate with other hosts to maintain state information and know what to do in case of another host's failure. (The VirtualCenter Management Server does not provide a single point of failure.) If the VirtualCenter Management Server host goes down, VMware HA functionality changes as follows: VMware HA clusters can still restart virtual machines on other hosts in case of failure; however, the information about what extra resources are available is based on the state of the cluster before the VirtualCenter Management Server went down.

Note: If you are also using DRS, the virtual machines running on VMware cluster hosts continue running using available resources. However because VirtualCenter Server is not running, it cannot provide further recommendation for resource optimization.

Using VMware HA

This section describes some of the setup and operation tasks you can perform using VMware HA and VirtualCenter — such as creating VMware HA clusters, adding or removing hosts from clusters, planning failover capacity, and setting properties.

Enabling VMware HA

VMware HA is included as an integrated component in VMware Infrastructure 3 Enterprise. It is also available as an add-on license option to VMware Infrastructure 3 Starter and VMware Infrastructure 3 Standard. To enable VMware HA when you create a VMware cluster, you need to set the **Enable VMware HA** option.

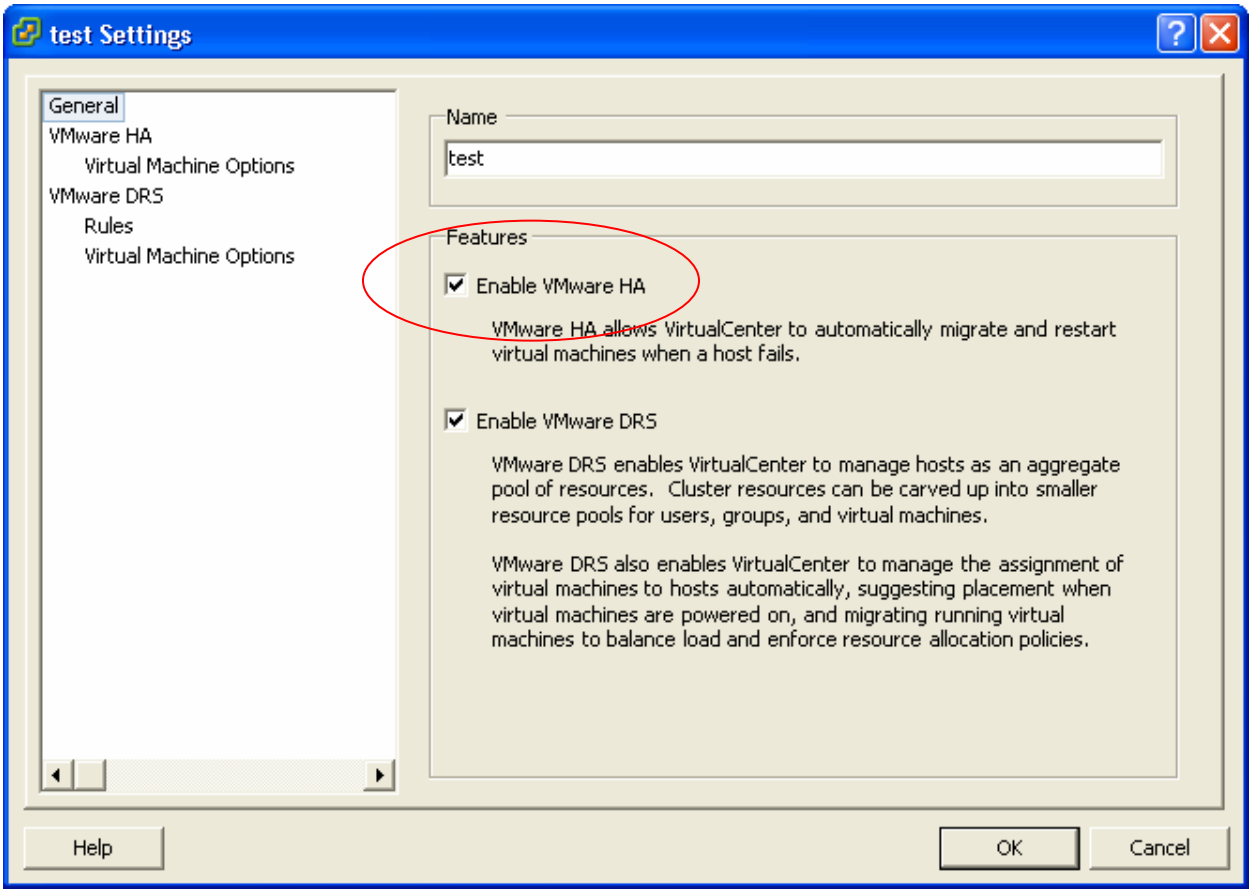

Figure 7: Configuring VMware HA

For clusters enabled for VMware HA, the resources of all included hosts are assigned to the cluster. If clusters are also enabled for VMware DRS, you can use VMware DRS to provide dynamic and intelligent resource allocation, optimization, and load-balancing of virtual machines, after failover.

Creating a VMware Cluster

A cluster is a collection of ESX Server hosts and associated virtual machines with shared resources and a shared management interface. When you add a host to a cluster, the host's resources become part of the cluster's resources. When you create a cluster, you can enable it for VMware DRS, VMware HA, or both. If you enable VMware DRS, the cluster supports shared resource pools

and performs placement and dynamic load balancing for virtual machines in the cluster. If you enable VMware HA, the cluster supports failover. When a host fails, VMware HA automatically restarts virtual machines on a different host. If you enable clusters for both VMware DRS and VMware HA, VMware DRS optimizes host placement and balanced resource allocation after failover and restart of virtual machines on new hosts.

Your system must also meet certain prerequisites to use VMware cluster features successfully. See VMware HA Requirements and Best Practices, later in this paper, for more specific requirements and recommendations.

VirtualCenter provides a New Cluster Wizard to take you through the steps of creating a new cluster. When you first invoke the wizard, it prompts you to choose whether to create a cluster that supports VMware DRS, VMware HA, or both. Following that, the wizard prompts you for the corresponding configuration information.

Note: When you create a cluster, it initially does not include any hosts or virtual machines. If you use VMware HA and VMware DRS together, when VMware HA performs failover and restarts virtual machines on different hosts, its first priority is immediate availability of all virtual machines. After the virtual machines have been restarted, those hosts on which they were powered on are usually heavily loaded, while other hosts are comparatively lightly loaded.

Using VMware HA and VMware DRS together combines automatic failover with load balancing. This combination can result in a fast rebalancing of virtual machines after VMware HA has moved virtual machines to different hosts. You can set up affinity and anti-affinity rules to start two or more virtual machines preferentially on the same host (affinity) or on different hosts.

Note: For more information about resource pools and using VMware DRS to manage operations such as virtual machine placement and providing dynamic resource allocation for virtual machines running on VMware cluster hosts, see the VMware Infrastructure 3 best practices paper titled "Resource Management with VMware DRS."

Selecting High Availability Options

If you enable VMware HA, the New Cluster Wizard allows you to set the following options.

Option	Description
Host failures	Specifies the number of host failures (or failure capacity) for which you want to guarantee failover of virtual machines.
Admission control	Offers two choices about how decisions are made to allow new virtual machines to be powered up: ▪ Do not power on virtual machines if they violate availability constraints and enforce the specified failover capacity limits. ▪ Allow virtual machines to be powered on even if they violate availability constraints. This allows you to power on virtual machines even if failover of the number of specified hosts can no longer be guaranteed. (A warning is issued.)

After initial creation of the cluster, you can add hosts and virtual machines to the cluster or specify additional cluster customization such as setting the priority for individual virtual machines. VMware HA uses virtual machine priority to decide order of restart in case of a red cluster (when configured failover capacity exceeds current failover capacity).

Note: If you are using a cluster enabled for VMware HA, that cluster might be marked with a red warning icon until you have added enough hosts to satisfy the specified failover capacity. See the Cluster Status Information section in this paper.

Adding Hosts to a VMware HA Cluster

The VirtualCenter inventory panel displays all clusters and hosts managed by the VirtualCenter Management Server to which the VI Client is connected. Adding managed hosts to a VMware HA cluster is as simple as selecting and dragging a host machine to the desired target cluster.

Note: You can also add unmanaged hosts by selecting the **Add Host** option and specifying the unmanaged host name, user name, and password. Adding a host to the cluster spawns a system task called Configuring HA on the host. After you complete this task, the host is included in the VMware HA service, and virtual machines deployed to the host become part of the cluster.

When you add a new host to the cluster:

- The resources for that host immediately become available to the cluster for use in the cluster's root resource pool.
- Unless the cluster is also enabled for VMware DRS, all resource pools are collapsed into the cluster's top-level (invisible) resource pool.
- Any capacity on the host beyond what is required or guaranteed for each running virtual machine becomes available as spare capacity in the cluster pool. This spare capacity can be used for starting virtual machines on other hosts in case of a host failure.
- If you add a host with several running virtual machines, and the cluster no longer fulfills its failover requirements because of that addition, a warning appears and the cluster's status is changed to invalid (red).
- By default, all virtual machines on hosts you add to the cluster are given a restart priority of medium. You can change the priority and specify other VMware HA customization options to tailor individual virtual machine priorities and other settings. See the section Customizing Virtual Machine HA Options in this paper for more information.
- The system also monitors the status of the VMware HA service on each host and displays information about configuration issues on the Summary page.

Viewing Cluster Information

When you select a cluster from the VirtualCenter inventory panel, the Summary page displays high-level information about the selected cluster.

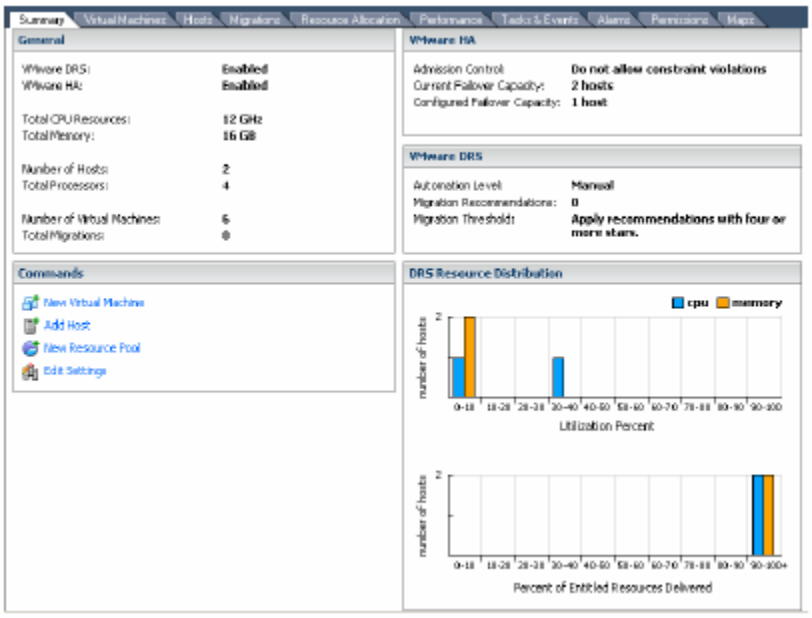

Figure 8: Viewing cluster information

The Summary pane of the VI Client provides various information about the cluster and virtual machines assigned to the cluster. VMware HA configuration details on the pane include the cluster's admission control setting, current failover capacity, and configured failover capacity for clusters enabled for VMware HA.

Note: For clusters enabled for VMware DRS, the Summary pane also displays automation level and migration threshold settings, outstanding migration recommendations, and real-time histograms of utilization percent and percent of entitled resources delivered, to show how balanced the cluster is.

The Summary pane updates the current failover capacity whenever a host has been added to or removed from the cluster or when virtual machines have been powered on or powered off.

Cluster Status Information

As hosts and virtual machines are added or removed, clusters can become overcommitted or invalid because of VMware HA or VMware DRS violations. Messages displayed on the Summary pane indicate the status of the currently selected cluster.

The VI Client indicates whether a cluster is valid (green), overcommitted (yellow), or invalid (red).

Valid Clusters

A cluster is considered valid unless something happens that makes it overcommitted or it no longer satisfies failover capacity requirements. For example, a VMware HA cluster becomes invalid if the current failover capacity is lower than the configured failover capacity.

If a cluster is labeled green (valid), this indicates available resources can meet all reservations and support all running virtual machines. In addition, at least one host has enough resources to run

each virtual machine assigned to the cluster. If you use a particularly large virtual machine — for example, a virtual machine with a 16GB memory reservation — you must have at least one host with that much memory. It is not enough if two hosts together fulfill the requirement.

Overcommitted Clusters

A cluster becomes yellow (overcommitted) when capacity is removed from the cluster, for example, because a host fails or is removed and there are no longer enough resources to support all requirements.

Invalid Clusters

A cluster enabled for VMware HA becomes red (invalid) when the number of virtual machines powered on exceeds the requirements of strict failover — that is, current failover capacity is smaller than the configured failover capacity. This can happen, for example, if you first check **Allow virtual machines to be started even if they violate availability constraints** for that cluster and later power on so many virtual machines that the cluster no longer has sufficient resources to guarantee failover for the specified number of hosts. A cluster can also become red if you power on a virtual machine or perform other operations directly on the host that over-utilizes the server capacity.

A cluster can also become red, for example, if VMware HA is set up for two-host failure in a four-host cluster and one host fails. The remaining three hosts might no longer be able to satisfy a two-host failure.

If a cluster enabled for VMware HA becomes red, it can no longer guarantee failover for the specified number of hosts, but it does continue performing failover. In case of host failure, VMware HA first fails over the virtual machines of one host in order of priority, then the virtual machines of the second host in order of priority, and so on.

Customizing Virtual Machine HA Options

Reconfiguring VMware HA can mean turning it off or reconfiguring its options. To turn off VMware HA, select the cluster and deselect the VMware HA check box from the Edit Settings panel.

To customize VMware HA behavior for individual virtual machines select the cluster and select HA Services from the **Edit Settings** > **Cluster Settings** dialog box.

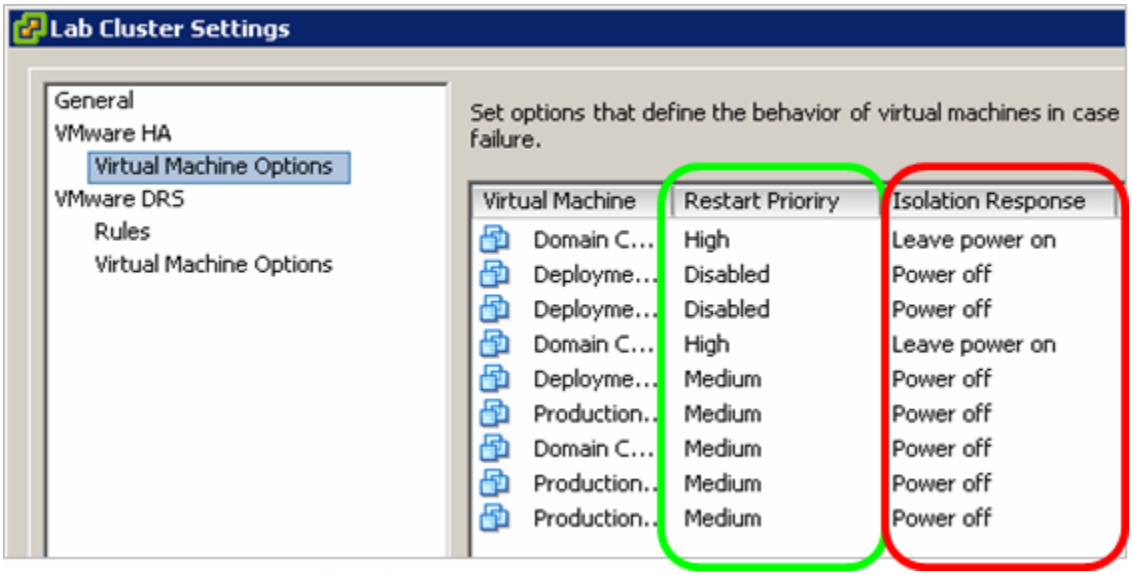

Figure 9: Customizing virtual machine options

This dialog box lets you make changes to the number of host failovers or the admission control behavior.

You can customize VMware HA for restart priority and isolation response.

Restart priority determines the order in which virtual machines are restarted upon host failure. Higher priority virtual machines are started first. Restart priority is always considered, but it is especially important in the following cases:

- If you set host failure capacity to a certain number of hosts and more than that number of hosts actually fails.
- If you turn off strict admission control and have started more virtual machines than VMware HA has been set up to support.

Note: This priority applies only on a per-host basis. If multiple hosts fail, VMware HA first migrates all virtual machines from the first host, in order of their priority, then all virtual machines from the second host in order of priority, and so on.

Isolation Response Configuration

A host in a VMware HA cluster might lose its console network connectivity. Such a host is isolated from other hosts in the cluster. Isolation is a special case in which an ESX Server host has not actually failed, but its service console network is broken (for example, due to switch failure, Ethernet adapter failure, or some similar cause). Isolation is handled as a special case of failure in VMware HA. The isolated host intentionally shuts down the virtual machines running on that host, so they can be restarted on hosts in the cluster that are not isolated.

By default, about 12 seconds after heartbeats have ceased arriving, the isolated host itself starts a special procedure called isolation response. This typically involves shutting down its virtual machines. About 15 seconds after the start of the isolation event, other hosts in the cluster consider that isolated host as failed and attempt to restart the virtual machines. You can change both of these timeout values from their defaults using VMware HA advanced options in VirtualCenter.

If a virtual machine is configured to continue to run on the isolated host, VMFS disk locking prevents it from being powered on elsewhere, avoiding a "split-brain" condition.

If virtual machines share the network adapter that failed, they do not have access to the network. It might be advisable to start the virtual machine on another host.

By default, virtual machines are powered off in case of a host isolation incident. This releases their shared storage locks, which allows the virtual machines to be started up on other hosts. Carefully review networking best practices explained later in the paper, for avoiding Isolation. You can change the default behavior for individual virtual machines in case of Isolation. Using the advanced option `das.poweroffonisolation`, you can change the default isolation response behavior of the cluster. Setting this parameter to `false` indicates that the virtual machines on isolated hosts should continue running even if the host can no longer communicate with other hosts in the cluster. This change must be carefully considered. If you choose to do this and the virtual machines cannot communicate over the network, they may become unavailable. If you choose to use this option with network-based storage, such as NAS or iSCSI, you run the risk that virtual machines might be unable to access their storage if they lose access to the network.

Note: If you add a host to a cluster, all virtual machines in the cluster default to a restart priority of medium and an isolation response of power off.

Powering On Virtual Machines in a Cluster

When you power on a virtual machine on a host that is part of a cluster, the resulting VirtualCenter behavior depends on the type of cluster.

If you power on a virtual machine and VMware HA is enabled, VirtualCenter first checks whether there are enough resources to continue supporting the specified number of host failovers if you power on the virtual machine.

If there are enough resources, the virtual machine is powered on.

If there are not enough resources and you are using strict admission control (the default), a message informs you that the virtual machine cannot be powered on. If you are not using strict admission control, a message informs you that there are no longer enough resources to guarantee failover for all hosts. The virtual machine is powered on but the cluster turns red.

Host Removal and Virtual Machines

Both stand-alone hosts and hosts within a cluster support maintenance mode, which restricts virtual machine operations on the host to allow the user to shut down running virtual machines in preparation for shutting down the host. While in maintenance mode, the host does not allow you to deploy or power on new virtual machines. Virtual machines that are already running on the host continue to run normally. You can either migrate them to another host or shut them down.

When there are no more running virtual machines on the host or cluster, its icon changes and its summary indicates the new state. In addition, menu and command options involving virtual machine deployment are disabled when this host or cluster is selected.

Because a host must be in maintenance mode before you can remove it from a cluster, all virtual machines must be powered off first — unless VMware DRS is also enabled, in which case virtual machines are automatically removed from the host. When you then remove the host from the cluster, the virtual machines that are currently associated with the host are also removed from the cluster.

Removing Hosts with Virtual Machines from a Cluster

If you remove a host from a cluster, the available resources for the cluster decrease. When you remove a host with virtual machines from a cluster, all that host's virtual machines are removed as well. You can remove a host only if it is in maintenance mode or disconnected.

Note: If a cluster enabled for VMware HA loses so many resources that it can no longer fulfill its failover requirements, a message appears and the cluster turns red. The cluster will fail over virtual machines in case of host failure but is not guaranteed to have enough resources available to fail over all virtual machines.

Removing Virtual Machines from a Cluster

You can remove virtual machines by migrating them out of a cluster or removing a host with virtual machines from the cluster.

You can migrate a virtual machine from a cluster to a standalone host or from a cluster to another cluster, using the standard drag-and-drop method or selecting Migrate from the virtual machine's right-button menu or the VirtualCenter menu bar. If the cluster is also enabled for VMware DRS and the virtual machine is a member of a VMware DRS cluster affinity group, VirtualCenter displays a warning before it allows the migration to proceed. The warning indicates that dependent virtual machines are not migrated automatically, so you have to acknowledge the warning before migration can proceed.

VMware HA Best Practices

Use the VMware HA best practices in this section that are applicable to your ESX Server implementation and networking architecture.

Networking Best Practices

The configuration of ESX Server host networking and name resolution, as well as the networking infrastructure external to ESX Server hosts (switches, routers, and firewalls), is critical to optimizing VMware HA setup. The following suggestions are best practices for configuring these components for improved HA performance:

- If your switches support the PortFast (or an equivalent) setting, enable it on the physical network switches connecting servers. This helps avoid spanning tree isolation events. For more information on this option, refer to the documentation provided by your networking switch vendor.
- Ensure that the following firewall ports are open for communication by the service console for all ESX Server 3 hosts:

  Incoming Port: TCP/UDP 8042-8045

  Outgoing Port: TCP/UDP 2050-2250
- For better heartbeat reliability, configure end-to-end dual network paths between servers for service console networking. You should also configure shorter network paths between the

servers in a cluster. Routes with too many hops can cause networking packet delays for heartbeats.

- If redundant service consoles are on separate subnets, specify "isolation address" for each service console that is on its subnet. By default, gateway address for the network is used as isolation address.
- Disable VMware HA (using VirtualCenter, deselect the **Enable VMware HA** check box in the cluster's Settings dialog box) when performing any networking maintenance that might disable all heartbeat paths between hosts.
- Use DNS for name resolution rather than the error–prone method of manually editing the local `/etc/hosts` file on ESX Server hosts. If you do edit `/etc/hosts`, you must include both long and short names.
- Use consistent port names on VLANs for public networks on all ESX servers in the cluster. Port names are used to reconfigure access to the network by virtual machines. If the names are used on the original server and the failover server are inconsistent, virtual machines are disconnected from their networks after failover.

Setting Up Networking Redundancy

Networking redundancy between cluster nodes is important for VMware HA reliability. Redundant service console networking on ESX Server 3 (or VMkernel networking on ESX Server 3i) allows the reliable detection of failures and prevents isolation conditions from occurring, because heartbeats can be sent over multiple networks.

You can implement network redundancy at the NIC level or at the service console or VMkernel port level. In most implementations, NIC teaming provides sufficient redundancy, but you can use or add service console or port redundancy if you need additional redundancy.

NIC Teaming

As shown in Figure 10, using a team of two NICs connected to separate physical switches can improve the reliability of a service console (or, in ESX Server 3i, VMkernel) network. Because servers connected to each other through two NICs (and through separate switches) have two independent paths for sending and receiving heartbeats, the cluster is more resilient.

To configure a NIC team for the service console, configure the vNICs in vSwitch configuration for the ESX Server host, for Active/standby configuration. The recommended parameters for the vNICs are:

- Rolling Failover = Yes
- Default Load Balancing = route based on originating port ID

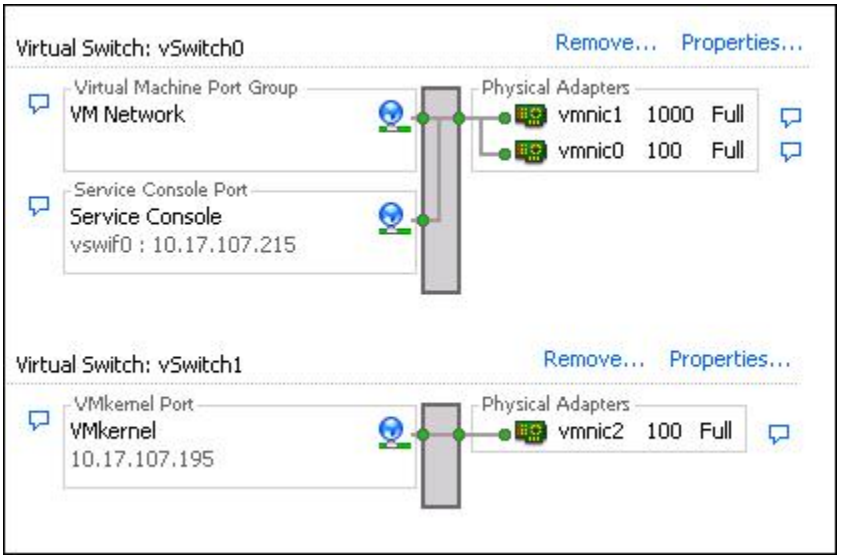

Figure 10: Service console redundancy using NIC teaming

The following example illustrates the use of a single service console network with NIC teaming for network redundancy:

- You assume some risk if you configure hosts in the cluster with only one service console network (subnet 10.20.XX.XX), so this example uses two teamed NICs to protect against NIC failure.
- The default timeout is increased to 60 seconds (`das.failuredetectiontime = 60000`).

Secondary Service Console Network

As an alternative to NIC teaming for providing redundancy for heartbeats, you can create a secondary service console network (or VMkernel port for ESX Server 3i), then attach that port to a separate virtual switch. The primary service console network is still used for network and management purposes. When you create the secondary service console network, VMware HA sends heartbeats over both the primary and secondary service console networks. If one path fails, VMware HA can still send and receive heartbeats over the other path.

By default, the gateway IP address specified in each ESX Server host's service console network configuration is used as the isolation address. Each service console network should have one isolation address it can reach. When you set up service console network redundancy, you must specify an additional isolation response address (`das.isolationaddress2`) for the secondary service console network. When you specify this secondary isolation address, VMware also recommends that you increase the `das.failuredetectiontime` setting to 20000 milliseconds or greater.

Also, make sure you configure isolation addresses properly for the redundant service console network that you create. Follow the networking best practices when designating isolation addresses.

A further optimization you can make (if you have already configured a VMotion network) is to add a secondary service console network to the VMotion vSwitch. As shown in Figure 11, a virtual switch can be shared between VMotion networks and a secondary service console network.

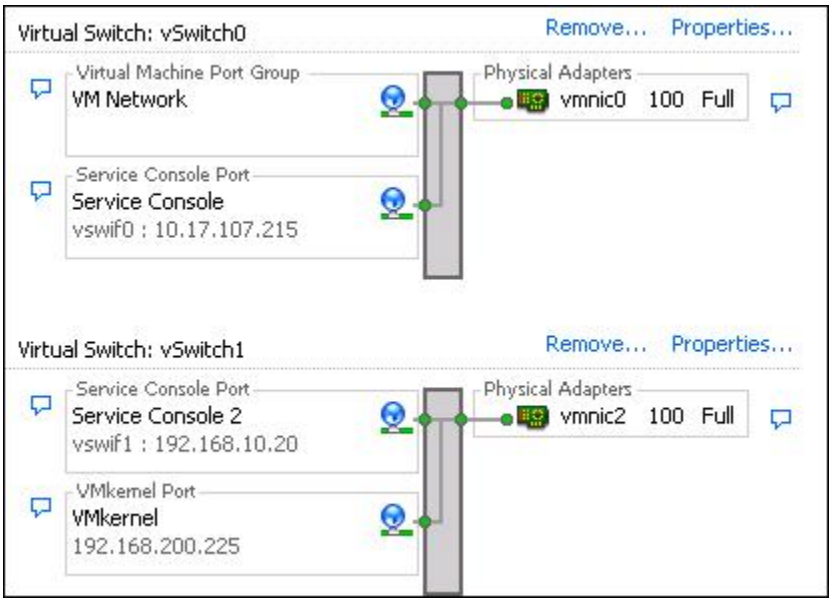

Figure 11: Network redundancy with a secondary service console

As the figure shows, each host in the cluster is configured with two service consoles. Each of these service console networks is connected to a separate physical NIC. The two networks are also on different subnets.

Use the default gateway for the first network and specify `das.isolationaddress2 = 192.168.1.103` as the additional isolation address for the second network. Increase the default timeout to 20 seconds (`das.failuredetectiontime = 20000`).

Other HA Cluster Considerations

Other considerations for optimizing the performance of your HA cluster include:

- Use larger groups of homogenous servers to allow higher levels of utilization across an HA-enabled cluster (on average).
  - More nodes per cluster can tolerate multiple host failures while still guaranteeing failover capacities.
  - Admission control heuristics are conservatively weighted so that large servers with many virtual machines can fail over to smaller servers.
- To define the sizing estimates used for admission control, set reasonable reservations for the minimum resources needed.
  - Admission control exceeds failover capacities when reservations are not set; otherwise, VMware HA uses the largest reservation specified for a virtual machine in the cluster when deciding failover capacity.
  - At a minimum, set reservations for a few virtual machines considered average.
- Admission control may be too conservative when host and virtual machine sizes vary widely. You may choose to do your own capacity planning by choosing **Allow virtual machines to be powered on**

even if they violate availability constraints. VMware HA will still try to restart as many virtual machines as it can.

Troubleshooting

Typically, troubleshooting VMware HA involves following the steps below and taking corrective action to see if the problem is fixed:

- Verify the network is working properly. At a minimum, your VirtualCenter Management Server must be able to reach the ESX Server hosts, and the hosts must be able to ping the gateway on the service console networks. If you are using multiple service console networks for redundancy, each of the service consoles must be able to ping the gateway IP address on that network. For information on networking troubleshooting, refer to the *ESX Server Configuration* guide.
- Verify shared storage is accessible from all hosts in the cluster. Carefully verify connectivity at server, storage, and switch levels.
- Pay attention to cluster warnings, task details, and events. Task details, configuration issue error messages, and event histories in VirtualCenter are useful places to look for information related to an error.
- Check logs for clues. Logs on the ESX Server hosts are the best place to start troubleshooting VMware HA. Always start by looking for ESX Server service console networking errors, followed by VMware HA agent problems.
- Use the **Reconfigure for HA** menu in the VI Client as a last resort to reconfigure a host when the VMware HA agent on the host appear unresponsive or have encountered configuration error. This option initiates configuration and restart of the VMware HA agent on the selected ESX Server host.

Troubleshooting Cluster Configuration Errors

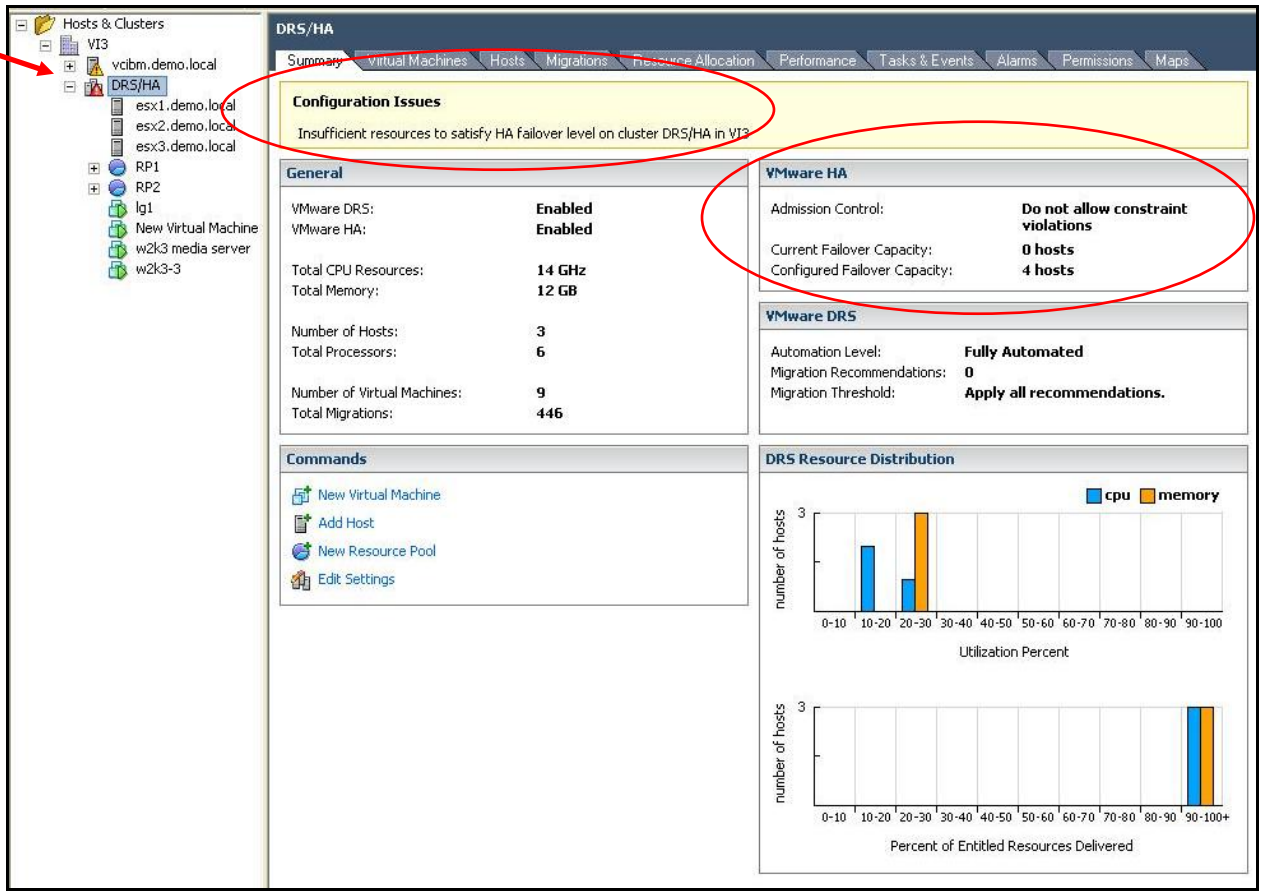

Figure 12: Looking for messeges on screen to help troubleshoot

If you see errors while configuring a VMware HA cluster, they are most likely displayed in the yellow error box that appears in the VirtualCenter clustering screen. As shown in Figure 12, the on-screen message provides enough information to point out the problem. In this case, admission control constraints are turned on, and the cluster does not have enough servers to allow VMware HA to function and meet the admission control criterion.

The VMware HA cluster thus turns red, as indicated by the red triangle in front of the cluster name in the left pane of the VI Client window. The Configuration Issue message in the yellow box indicates the nature of the problem.

If you encounter errors during installation or initial configuration, check the VMware HA installation log that is created on each ESX Server host as it is added to the cluster. It contains all available information about the failures. VMware HA logs are stored on each ESX Server host that has VMware HA enabled. Refer to VMware Basic System Administrator guide for more information on logs.

If installation of VMware HA has finished successfully but VMware HA services are not starting, as a last resort, you can try restarting the services manually on individual hosts or the entire cluster. To do this on individual hosts, select the host in the VI Client. Click the **Summary** tab, then click **Reconfigure for HA**. To reconfigure VMware HA for the entire cluster, select the cluster and choose **Edit Settings**. In the wizard, deselect **VMware HA** and click **OK**. Once reconfiguration is complete, select **VMware HA** again to reconfigure VMware HA.

When VMware HA encounters an error, VirtualCenter shows the error in the Recent Tasks pane of the VI Client. To get more details on the error, select the host on which the error occurred and click the **Tasks and Events** tab.

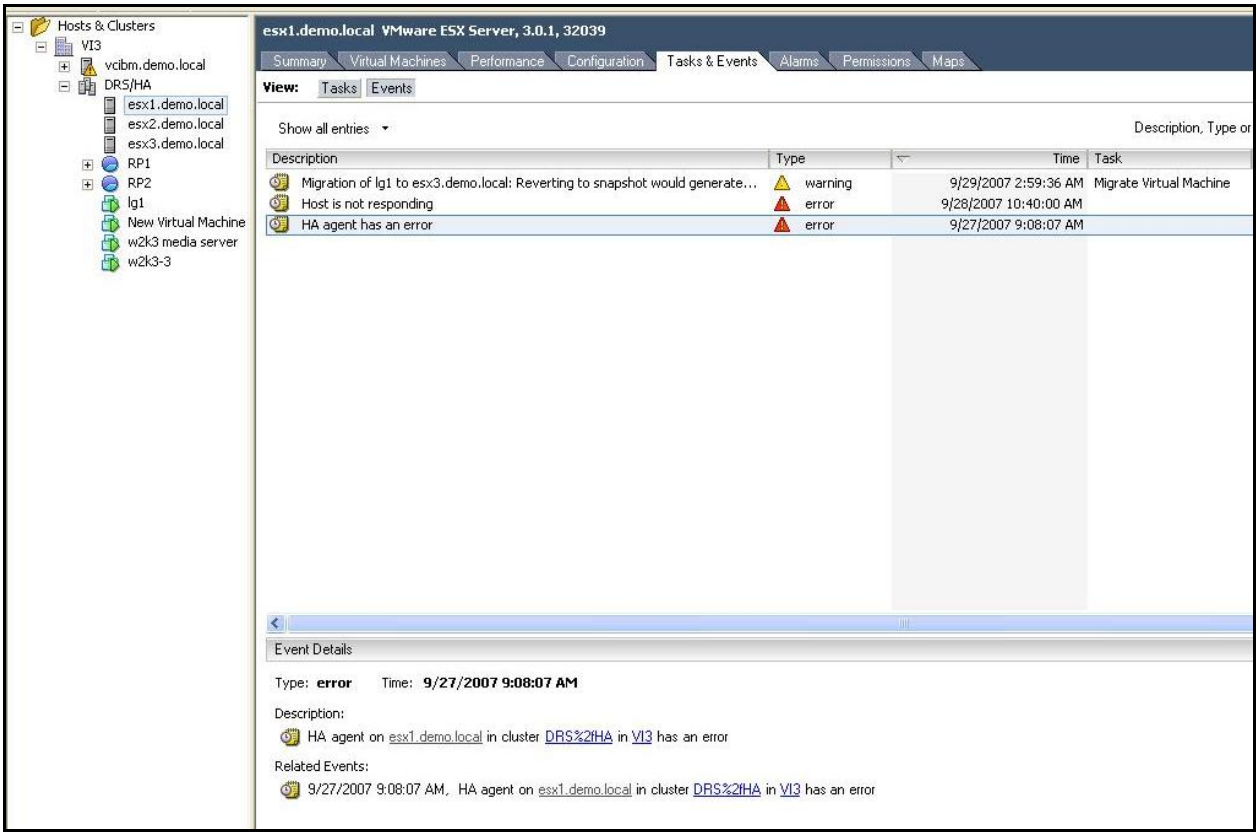

Figure 13: Viewing detailed error messages using the VI Client

Summary

VMware HA — along with new VMware Infrastructure capabilities of clusters and resource pools and tighter integration with other VMware tools such as VirtualCenter, VMotion, and VMware DRS — greatly simplifies virtual machine provisioning, resource allocation, load balancing, and migration, while also providing an easy-to-use, cost-effective high-availability and failover solution for applications running in virtual machines. Using VMware Infrastructure and VMware HA helps to eliminate single points of failure in the deployment of business-critical applications in virtual machines. At the same time, it allows you to maintain other inherent virtualization benefits such as higher system utilization, closer alignment of IT resources with business goals and priorities, and more streamlined, simplified, and automated administration of larger infrastructure installations and systems.

Revision: 20071204 Item: WP-025-ISV-01-02

VMware, Inc. 3401 Hillview Ave. Palo Alto CA 94304 USA Tel 650-475-5000 Fax 650-475-5001 www.vmware.com

vmware®