vSphere Availability

Update 1
ESXi 6.0
vCenter Server 6.0

This document supports the version of each product listed and supports all subsequent versions until the document is replaced by a new edition. To check for more recent editions of this document, see http://www.vmware.com/support/pubs.

EN-001810-02

vmware®

You can find the most up-to-date technical documentation on the VMware Web site at:

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

The VMware Web site also provides the latest product updates.

If you have comments about this documentation, submit your feedback to:

docfeedback@vmware.com

VMware, Inc.
3401 Hillview Ave.
Palo Alto, CA 94304
www.vmware.com

Contents

About vSphere Availability

vSphere Availability describes solutions that provide business continuity, including how to establish vSphere® High Availability (HA) and vSphere Fault Tolerance.

Intended Audience

This information is for anyone who wants to provide business continuity through the vSphere HA and Fault Tolerance solutions. The information in this book is for experienced Windows or Linux system administrators who are familiar with virtual machine technology and data center operations.

Updated Information

This *vSphere Availability* is updated with each release of the product or when necessary.

This table provides the update history of the *vSphere Availability*.

Revision	Description
EN-001810-02	Change to wording about dedicated FT network under Fault Tolerance Requirements. See "Fault Tolerance Requirements, Limits, and Licensing," on page 46.
EN-001810-01	New note about ESXi host version needed for VM Component Protection feature. See "VM Component Protection," on page 19.
EN-001810-00	Initial release.

Business Continuity and Minimizing Downtime 1

Downtime, whether planned or unplanned, brings with it considerable costs. However, solutions to ensure higher levels of availability have traditionally been costly, hard to implement, and difficult to manage.

VMware software makes it simpler and less expensive to provide higher levels of availability for important applications. With vSphere, organizations can easily increase the baseline level of availability provided for all applications as well as provide higher levels of availability more easily and cost effectively. With vSphere, you can:

- Provide higher availability independent of hardware, operating system, and applications.
- Reduce planned downtime for common maintenance operations.
- Provide automatic recovery in cases of failure.

vSphere makes it possible to reduce planned downtime, prevent unplanned downtime, and recover rapidly from outages.

This chapter includes the following topics:

Reducing Planned Downtime

Planned downtime typically accounts for over 80% of data center downtime. Hardware maintenance, server migration, and firmware updates all require downtime for physical servers. To minimize the impact of this downtime, organizations are forced to delay maintenance until inconvenient and difficult-to-schedule downtime windows.

vSphere makes it possible for organizations to dramatically reduce planned downtime. Because workloads in a vSphere environment can be dynamically moved to different physical servers without downtime or service interruption, server maintenance can be performed without requiring application and service downtime. With vSphere, organizations can:

- Eliminate downtime for common maintenance operations.
- Eliminate planned maintenance windows.
- Perform maintenance at any time without disrupting users and services.

The vSphere vMotion® and Storage vMotion functionality in vSphere makes it possible for organizations to reduce planned downtime because workloads in a VMware environment can be dynamically moved to different physical servers or to different underlying storage without service interruption. Administrators can perform faster and completely transparent maintenance operations, without being forced to schedule inconvenient maintenance windows.

Preventing Unplanned Downtime

While an ESXi host provides a robust platform for running applications, an organization must also protect itself from unplanned downtime caused from hardware or application failures. vSphere builds important capabilities into data center infrastructure that can help you prevent unplanned downtime.

These vSphere capabilities are part of virtual infrastructure and are transparent to the operating system and applications running in virtual machines. These features can be configured and utilized by all the virtual machines on a physical system, reducing the cost and complexity of providing higher availability. Key availability capabilities are built into vSphere:

- Shared storage. Eliminate single points of failure by storing virtual machine files on shared storage, such as Fibre Channel or iSCSI SAN, or NAS. The use of SAN mirroring and replication features can be used to keep updated copies of virtual disk at disaster recovery sites.
- Network interface teaming. Provide tolerance of individual network card failures.
- Storage multipathing. Tolerate storage path failures.

In addition to these capabilities, the vSphere HA and Fault Tolerance features can minimize or eliminate unplanned downtime by providing rapid recovery from outages and continuous availability, respectively.

vSphere HA Provides Rapid Recovery from Outages

vSphere HA leverages multiple ESXi hosts configured as a cluster to provide rapid recovery from outages and cost-effective high availability for applications running in virtual machines.

vSphere HA protects application availability in the following ways:

- It protects against a server failure by restarting the virtual machines on other hosts within the cluster.
- It protects against application failure by continuously monitoring a virtual machine and resetting it in the event that a failure is detected.
- It protects against datastore accessibility failures by restarting affected virtual machines on other hosts which still have access to their datastores.
- It protects virtual machines against network isolation by restarting them if their host becomes isolated on the management or Virtual SAN network. This protection is provided even if the network has become partitioned.

Unlike other clustering solutions, vSphere HA provides the infrastructure to protect all workloads with the infrastructure:

- You do not need to install special software within the application or virtual machine. All workloads are protected by vSphere HA. After vSphere HA is configured, no actions are required to protect new virtual machines. They are automatically protected.
- You can combine vSphere HA with vSphere Distributed Resource Scheduler (DRS) to protect against failures and to provide load balancing across the hosts within a cluster.

vSphere HA has several advantages over traditional failover solutions:

Minimal setup	After a vSphere HA cluster is set up, all virtual machines in the cluster get failover support without additional configuration.
Reduced hardware cost and setup	The virtual machine acts as a portable container for the applications and it can be moved among hosts. Administrators avoid duplicate configurations on multiple machines. When you use vSphere HA, you must have sufficient resources to fail over the number of hosts you want to protect with vSphere HA. However, the vCenter Server system automatically manages resources and configures clusters.
Increased application availability	Any application running inside a virtual machine has access to increased availability. Because the virtual machine can recover from hardware failure, all applications that start at boot have increased availability without increased computing needs, even if the application is not itself a clustered application. By monitoring and responding to VMware Tools heartbeats and restarting nonresponsive virtual machines, it protects against guest operating system crashes.
DRS and vMotion integration	If a host fails and virtual machines are restarted on other hosts, DRS can provide migration recommendations or migrate virtual machines for balanced resource allocation. If one or both of the source and destination hosts of a migration fail, vSphere HA can help recover from that failure.

vSphere Fault Tolerance Provides Continuous Availability

vSphere HA provides a base level of protection for your virtual machines by restarting virtual machines in the event of a host failure. vSphere Fault Tolerance provides a higher level of availability, allowing users to protect any virtual machine from a host failure with no loss of data, transactions, or connections.

Fault Tolerance provides continuous availability by ensuring that the states of the Primary and Secondary VMs are identical at any point in the instruction execution of the virtual machine.

If either the host running the Primary VM or the host running the Secondary VM fails, an immediate and transparent failover occurs. The functioning ESXi host seamlessly becomes the Primary VM host without losing network connections or in-progress transactions. With transparent failover, there is no data loss and network connections are maintained. After a transparent failover occurs, a new Secondary VM is respawned and redundancy is re-established. The entire process is transparent and fully automated and occurs even if vCenter Server is unavailable.

Creating and Using vSphere HA Clusters

2

vSphere HA clusters enable a collection of ESXi hosts to work together so that, as a group, they provide higher levels of availability for virtual machines than each ESXi host can provide individually. When you plan the creation and usage of a new vSphere HA cluster, the options you select affect the way that cluster responds to failures of hosts or virtual machines.

Before you create a vSphere HA cluster, you should know how vSphere HA identifies host failures and isolation and how it responds to these situations. You also should know how admission control works so that you can choose the policy that fits your failover needs. After you establish a cluster, you can customize its behavior with advanced options and optimize its performance by following recommended best practices.

NOTE You might get an error message when you try to use vSphere HA. For information about error messages related to vSphere HA, see the VMware knowledge base article at http://kb.vmware.com/kb/1033634.

This chapter includes the following topics:

How vSphere HA Works

vSphere HA provides high availability for virtual machines by pooling the virtual machines and the hosts they reside on into a cluster. Hosts in the cluster are monitored and in the event of a failure, the virtual machines on a failed host are restarted on alternate hosts.

When you create a vSphere HA cluster, a single host is automatically elected as the master host. The master host communicates with vCenter Server and monitors the state of all protected virtual machines and of the slave hosts. Different types of host failures are possible, and the master host must detect and appropriately deal with the failure. The master host must distinguish between a failed host and one that is in a network partition or that has become network isolated. The master host uses network and datastore heartbeating to determine the type of failure.

Sphere HA Clusters (http://link.brightcove.com/services/player/bcpid2296383276001?bctid=ref:vSphereHAClusters)

Master and Slave Hosts

When you add a host to a vSphere HA cluster, an agent is uploaded to the host and configured to communicate with other agents in the cluster. Each host in the cluster functions as a master host or a slave host.

When vSphere HA is enabled for a cluster, all active hosts (those not in standby or maintenance mode, or not disconnected) participate in an election to choose the cluster's master host. The host that mounts the greatest number of datastores has an advantage in the election. Only one master host typically exists per cluster and all other hosts are slave hosts. If the master host fails, is shut down or put in standby mode, or is removed from the cluster a new election is held.

The master host in a cluster has a number of responsibilities:

- Monitoring the state of slave hosts. If a slave host fails or becomes unreachable, the master host identifies which virtual machines need to be restarted.
- Monitoring the power state of all protected virtual machines. If one virtual machine fails, the master host ensures that it is restarted. Using a local placement engine, the master host also determines where the restart should be done.
- Managing the lists of cluster hosts and protected virtual machines.
- Acting as vCenter Server management interface to the cluster and reporting the cluster health state.

The slave hosts primarily contribute to the cluster by running virtual machines locally, monitoring their runtime states, and reporting state updates to the master host. A master host can also run and monitor virtual machines. Both slave hosts and master hosts implement the VM and Application Monitoring features.

One of the functions performed by the master host is to orchestrate restarts of protected virtual machines. A virtual machine is protected by a master host after vCenter Server observes that the virtual machine's power state has changed from powered off to powered on in response to a user action. The master host persists the list of protected virtual machines in the cluster's datastores. A newly elected master host uses this information to determine which virtual machines to protect.

NOTE If you disconnect a host from a cluster, all of the virtual machines registered to that host are unprotected by vSphere HA.

Host Failure Types and Detection

The master host of a vSphere HA cluster is responsible for detecting the failure of slave hosts. Depending on the type of failure detected, the virtual machines running on the hosts might need to be failed over.

In a vSphere HA cluster, three types of host failure are detected:

- Failure- A host stops functioning.
- Isolation- A host becomes network isolated.
- Partition- A host loses network connectivity with the master host.

The master host monitors the liveness of the slave hosts in the cluster. This communication is done through the exchange of network heartbeats every second. When the master host stops receiving these heartbeats from a slave host, it checks for host liveness before declaring the host to have failed. The liveness check that the master host performs is to determine whether the slave host is exchanging heartbeats with one of the datastores. See "Datastore Heartbeating," on page 21. Also, the master host checks whether the host responds to ICMP pings sent to its management IP addresses.

If a master host is unable to communicate directly with the agent on a slave host, the slave host does not respond to ICMP pings, and the agent is not issuing heartbeats it is considered to have failed. The host's virtual machines are restarted on alternate hosts. If such a slave host is exchanging heartbeats with a datastore, the master host assumes that it is in a network partition or network isolated and so continues to monitor the host and its virtual machines. See "Network Partitions," on page 21.

Host network isolation occurs when a host is still running, but it can no longer observe traffic from vSphere HA agents on the management network. If a host stops observing this traffic, it attempts to ping the cluster isolation addresses. If this also fails, the host declares itself as isolated from the network.

The master host monitors the virtual machines that are running on an isolated host and if it observes that they power off, and the master host is responsible for the virtual machines, it restarts them.

NOTE If you ensure that the network infrastructure is sufficiently redundant and that at least one network path is available at all times, host network isolation should be a rare occurrence.

Determining Responses to Host Issues

If a host fails and its virtual machines must be restarted, you can control the order in which the virtual machines are restarted with the VM restart priority setting. You can also configure how vSphere HA responds if hosts lose management network connectivity with other hosts by using the host isolation response setting. Other factors are also considered when vSphere HA restarts a virtual machine after a failure.

The following settings apply to all virtual machines in the cluster in the case of a host failure or isolation. You can also configure exceptions for specific virtual machines. See "Customize an Individual Virtual Machine," on page 40.

VM Restart Priority

VM restart priority determines the relative order in which virtual machines are allocated resources after a host failure. Such virtual machines are assigned to hosts with unreserved capacity, with the highest priority virtual machines placed first and continuing to those with lower priority until all virtual machines have been placed or no more cluster capacity is available to meet the reservations or memory overhead of the virtual machines. A host then restarts the virtual machines assigned to it in priority order. If there are insufficient resources, vSphere HA waits for more unreserved capacity to become available, for example, due to a host coming back online, and then retries the placement of these virtual machines. To reduce the chance of this situation occurring, configure vSphere HA admission control to reserve more resources for failures. Admission control allows you to control the amount of cluster capacity that is reserved by virtual machines, which is unavailable to meet the reservations and memory overhead of other virtual machines if there is a failure.

The values for this setting are Disabled, Low, Medium (the default), and High. The Disabled setting is ignored by the vSphere HA VM/Application monitoring feature because this feature protects virtual machines against operating system-level failures and not virtual machine failures. When an operating system-level failure occurs, the operating system is rebooted by vSphere HA, and the virtual machine is left running on the same host. You can change this setting for individual virtual machines.

NOTE A virtual machine reset causes a hard reboot of the guest operating system, but does not power cycle the virtual machine.

The restart priority settings for virtual machines vary depending on user needs. Assign higher restart priority to the virtual machines that provide the most important services.

For example, in the case of a multitier application, you might rank assignments according to functions hosted on the virtual machines.

- High. Database servers that provide data for applications.

- Medium. Application servers that consume data in the database and provide results on web pages.
- Low. Web servers that receive user requests, pass queries to application servers, and return results to users.

If a host fails, vSphere HA attempts to register to an active host the affected virtual machines that were powered on and have a restart priority setting of Disabled, or that were powered off.

Host Isolation Response

Host isolation response determines what happens when a host in a vSphere HA cluster loses its management network connections, but continues to run. You can use the isolation response to have vSphere HA power off virtual machines that are running on an isolated host and restart them on a nonisolated host. Host isolation responses require that Host Monitoring Status is enabled. If Host Monitoring Status is disabled, host isolation responses are also suspended. A host determines that it is isolated when it is unable to communicate with the agents running on the other hosts, and it is unable to ping its isolation addresses. The host then executes its isolation response. The responses are Power off and restart VMs or Shutdown and restart VMs. You can customize this property for individual virtual machines.

NOTE If a virtual machine has a restart priority setting of Disabled, no host isolation response is made.

To use the Shutdown and restart VMs setting, you must install VMware Tools in the guest operating system of the virtual machine. Shutting down the virtual machine provides the advantage of preserving its state. Shutting down is better than powering off the virtual machine, which does not flush most recent changes to disk or commit transactions. Virtual machines that are in the process of shutting down take longer to fail over while the shutdown completes. Virtual Machines that have not shut down in 300 seconds, or the time specified in the advanced option `das.isolationshutdowntimeout`, are powered off.

After you create a vSphere HA cluster, you can override the default cluster settings for Restart Priority and Isolation Response for specific virtual machines. Such overrides are useful for virtual machines that are used for special tasks. For example, virtual machines that provide infrastructure services like DNS or DHCP might need to be powered on before other virtual machines in the cluster.

A virtual machine "split-brain" condition can occur when a host becomes isolated or partitioned from a master host and the master host cannot communicate with it using heartbeat datastores. In this situation, the master host cannot determine that the host is alive and so declares it dead. The master host then attempts to restart the virtual machines that are running on the isolated or partitioned host. This attempt succeeds if the virtual machines remain running on the isolated/partitioned host and that host lost access to the virtual machines' datastores when it became isolated or partitioned. A split-brain condition then exists because there are two instances of the virtual machine. However, only one instance is able to read or write the virtual machine's virtual disks. VM Component Protection can be used to prevent this split-brain condition. When you enable VMCP with the aggressive setting, it monitors the datastore accessibility of powered-on virtual machines, and shuts down those that lose access to their datastores.

To recover from this situation, ESXi generates a question on the virtual machine that has lost the disk locks for when the host comes out of isolation and cannot reacquire the disk locks. vSphere HA automatically answers this question, allowing the virtual machine instance that has lost the disk locks to power off, leaving just the instance that has the disk locks.

Factors Considered for Virtual Machine Restarts

After a failure, the cluster's master host attempts to restart affected virtual machines by identifying a host that can power them on. When choosing such a host, the master host considers a number of factors.

File accessibility	Before a virtual machine can be started, its files must be accessible from one of the active cluster hosts that the master can communicate with over the network
Virtual machine and host compatibility	If there are accessible hosts, the virtual machine must be compatible with at least one of them. The compatibility set for a virtual machine includes the effect of any required VM-Host affinity rules. For example, if a rule only permits a virtual machine to run on two hosts, it is considered for placement on those two hosts.
Resource reservations	Of the hosts that the virtual machine can run on, at least one must have sufficient unreserved capacity to meet the memory overhead of the virtual machine and any resource reservations. Four types of reservations are considered: CPU, Memory, vNIC, and Virtual flash. Also, sufficient network ports must be available to power on the virtual machine.
Host limits	In addition to resource reservations, a virtual machine can only be placed on a host if doing so does not violate the maximum number of allowed virtual machines or the number of in-use vCPUs.
Feature constraints	If the advanced option has been set that requires vSphere HA to enforce VM to VM anti-affinity rules, vSphere HA does not violate this rule. Also, vSphere HA does not violate any configured per host limits for fault tolerant virtual machines.

If no hosts satisfy the preceding considerations, the master host issues an event stating that there are not enough resources for vSphere HA to start the VM and tries again when the cluster conditions have changed. For example, if the virtual machine is not accessible, the master host tries again after a change in file accessibility.

Limits for Virtual Machine Restart Attempts

If the vSphere HA master agent's attempt to restart a VM, which involves registering it and powering it on, fails, this restart is retried after a delay. vSphere HA attempts these restarts for a maximum number of attempts (6 by default), but not all restart failures count against this maximum.

For example, the most likely reason for a restart attempt to fail is because either the VM is still running on another host, or because vSphere HA tried to restart the VM too soon after it failed. In this situation, the master agent delays the retry attempt by twice the delay imposed after the last attempt, with a 1 minute minimum delay and a 30 minute maximum delay. Thus if the delay is set to 1 minute, there is an initial attempt at T=0, then additional attempts made at T=1 (1 minute), T=3 (3 minutes), T=7 (7 minutes), T=15 (15 minutes), and T=30 (30 minutes). Each such attempt is counted against the limit and only six attempts are made by default.

Other restart failures result in countable retries but with a different delay interval. An example scenario is when the host chosen to restart virtual machine loses access to one of the VM's datastores after the choice was made by the master agent. In this case, a retry is attempted after a default delay of 2 minutes. This attempt also counts against the limit.

Finally, some retries are not counted. For example, if the host on which the virtual machine was to be restarted fails before the master agent issues the restart request, the attempt is retried after 2 minutes but this failure does not count against the maximum number of attempts.

Virtual Machine Restart Notifications

vSphere HA generates a cluster event when a failover operation is in progress for virtual machines in the cluster. The event also displays a configuration issue in the **Cluster Summary** tab which reports the number of virtual machines that are being restarted. There are four different categories of such VMs.

- VMs being placed: vSphere HA is in the process of trying to restart these VMs
- VMs awaiting a retry: a previous restart attempt failed, and vSphere HA is waiting for a timeout to expire before trying again.
- VMs requiring additional resources: insufficient resources are available to restart these VMs. vSphere HA retries when more resources become available, for example a host comes back online.
- Inaccessible Virtual SAN VMs: vSphere HA cannot restart these Virtual SAN VMs because they are not accessible. It retries when there is a change in accessibility.

These virtual machine counts are dynamically updated whenever a change is observed in the number of VMs for which a restart operation is underway. The configuration issue is cleared when vSphere HA has restarted all VMs or has given up trying.

In vSphere 5.5 or earlier, a per-VM event is triggered for an unsuccessful attempt to restart the virtual machine. This event is disabled by default in vSphere 6.x and can be enabled by setting the vSphere HA advanced option `das.config.fdm.reportfailoverfailevent` to 1.

VM and Application Monitoring

VM Monitoring restarts individual virtual machines if their VMware Tools heartbeats are not received within a set time. Similarly, Application Monitoring can restart a virtual machine if the heartbeats for an application it is running are not received. You can enable these features and configure the sensitivity with which vSphere HA monitors non-responsiveness.

When you enable VM Monitoring, the VM Monitoring service (using VMware Tools) evaluates whether each virtual machine in the cluster is running by checking for regular heartbeats and I/O activity from the VMware Tools process running inside the guest. If no heartbeats or I/O activity are received, this is most likely because the guest operating system has failed or VMware Tools is not being allocated any time to complete tasks. In such a case, the VM Monitoring service determines that the virtual machine has failed and the virtual machine is rebooted to restore service.

Occasionally, virtual machines or applications that are still functioning properly stop sending heartbeats. To avoid unnecessary resets, the VM Monitoring service also monitors a virtual machine's I/O activity. If no heartbeats are received within the failure interval, the I/O stats interval (a cluster-level attribute) is checked. The I/O stats interval determines if any disk or network activity has occurred for the virtual machine during the previous two minutes (120 seconds). If not, the virtual machine is reset. This default value (120 seconds) can be changed using the advanced option `das.iostatsinterval`.

To enable Application Monitoring, you must first obtain the appropriate SDK (or be using an application that supports VMware Application Monitoring) and use it to set up customized heartbeats for the applications you want to monitor. After you have done this, Application Monitoring works much the same way that VM Monitoring does. If the heartbeats for an application are not received for a specified time, its virtual machine is restarted.

You can configure the level of monitoring sensitivity. Highly sensitive monitoring results in a more rapid conclusion that a failure has occurred. While unlikely, highly sensitive monitoring might lead to falsely identifying failures when the virtual machine or application in question is actually still working, but heartbeats have not been received due to factors such as resource constraints. Low sensitivity monitoring results in longer interruptions in service between actual failures and virtual machines being reset. Select an option that is an effective compromise for your needs.

The default settings for monitoring sensitivity are described in Table 2-1. You can also specify custom values for both monitoring sensitivity and the I/O stats interval by selecting the **Custom** checkbox.

Table 2-1. VM Monitoring Settings

Setting	Failure Interval (seconds)	Reset Period
High	30	1 hour
Medium	60	24 hours
Low	120	7 days

After failures are detected, vSphere HA resets virtual machines. The reset ensures that services remain available. To avoid resetting virtual machines repeatedly for nontransient errors, by default, virtual machines will be reset only three times during a certain configurable time interval. After virtual machines have been reset three times, vSphere HA makes no further attempts to reset the virtual machines after subsequent failures until after the specified time has elapsed. You can configure the number of resets using the **Maximum per-VM resets** custom setting.

NOTE The reset statistics are cleared when a virtual machine is powered off then back on, or when it is migrated using vMotion to another host. This causes the guest operating system to reboot, but is not the same as a 'restart' in which the power state of the virtual machine is changed.

If a virtual machine has a datastore accessibility failure (either All Paths Down or Permanent Device Loss), the VM Monitoring service suspends resetting it until the failure has been addressed.

VM Component Protection

If VM Component Protection (VMCP) is enabled, vSphere HA can detect datastore accessibility failures and provide automated recovery for affected virtual machines.

VMCP provides protection against datastore accessibility failures that can affect a virtual machine running on a host in a vSphere HA cluster. When a datastore accessibility failure occurs, the affected host can no longer access the storage path for a specific datastore. You can determine the response that vSphere HA will make to such a failure, ranging from the creation of event alarms to virtual machine restarts on other hosts.

NOTE When you use the VM Component Protection feature, your ESXi hosts must be version 6.0 or higher.

Types of Failure

There are two types of datastore accessibility failure:

PDL PDL (Permanent Device Loss) is an unrecoverable loss of accessibility that occurs when a storage device reports the datastore is no longer accessible by the host. This condition cannot be reverted without powering off virtual machines.

APD APD (All Paths Down) represents a transient or unknown accessibility loss or any other unidentified delay in I/O processing. This type of accessibility issue is recoverable.

Configuring VMCP

VM Component Protection is enabled and configured in the vSphere Web Client. To enable this feature, you must select the **Protect against Storage Connectivity Loss** checkbox in the edit cluster settings wizard. The storage protection levels you can choose and the virtual machine remediation actions available differ depending on the type of database accessibility failure.

PDL failures

A virtual machine is automatically failed over to a new host unless you have configured VMCP only to **Issue events**.

APD events

The response to APD events is more complex and accordingly the configuration is more fine-grained.

After the user-configured **Delay for VM failover for APD** period has elapsed, the action taken depends on the policy you selected. An event will be issued and the virtual machine is restarted conservatively or aggressively. The conservative approach does not terminate the virtual machine if the success of the failover is unknown, for example in a network partition. The aggressive approach does terminate the virtual machine under these conditions. Neither approach terminates the virtual machine if there are insufficient resources in the cluster for the failover to succeed.

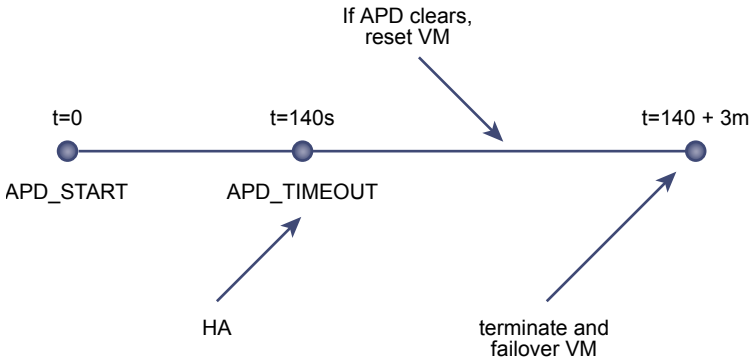
If APD recovers before the user-configured **Delay for VM failover for APD** period has elapsed, you can choose to reset the affected virtual machines, which recovers the guest applications that were impacted by the IO failures.

NOTE If either the Host Monitoring or VM Restart Priority settings are disabled, VMCP cannot perform virtual machine restarts. Storage health can still be monitored and events can be issued, however.

For more information on configuring VMCP, see "Configure Virtual Machine Responses," on page 35.

VMCP Recovery Timeline

The following timeline graphically demonstrates how VMCP recovers from a storage failure.



- `T=0s`: A storage failure is detected. vSphere HA starts the recovery process. For a PDL event, the workflow immediately starts and VMs are restarted on healthy hosts in the cluster. If the storage loss is due to an APD event, the APD Timeout timer starts (the default is 140 seconds).
- `T=140s`: The host declares an APD Timeout and begins to fail non-VM I/O to the unresponsive storage device.
- `Between T=140s and 320s`: This is the time period defined by the **Delay for VM failover for APD**, which is 3 minutes by default. The guest applications might become unstable after losing access to storage for an extended period of time. If an APD is cleared in this time period, the option to reset the VMs is available.

- `T=320s`: vSphere HA now starts the APD recovery response after the **Delay for VM failover for APD** elapses (3 minutes after the APD Timeout is reached).

Network Partitions

When a management network failure occurs for a vSphere HA cluster, a subset of the cluster's hosts might be unable to communicate over the management network with the other hosts. Multiple partitions can occur in a cluster.

A partitioned cluster leads to degraded virtual machine protection and cluster management functionality. Correct the partitioned cluster as soon as possible.

- Virtual machine protection. vCenter Server allows a virtual machine to be powered on, but it can be protected only if it is running in the same partition as the master host that is responsible for it. The master host must be communicating with vCenter Server. A master host is responsible for a virtual machine if it has exclusively locked a system-defined file on the datastore that contains the virtual machine's configuration file.
- Cluster management. vCenter Server can communicate with the master host, but only a subset of the slave hosts. As a result, changes in configuration that affect vSphere HA might not take effect until after the partition is resolved. This failure could result in one of the partitions operating under the old configuration, while another uses the new settings.

Datastore Heartbeating

When the master host in a vSphere HA cluster can not communicate with a slave host over the management network, the master host uses datastore heartbeating to determine whether the slave host has failed, is in a network partition, or is network isolated. If the slave host has stopped datastore heartbeating, it is considered to have failed and its virtual machines are restarted elsewhere.

vCenter Server selects a preferred set of datastores for heartbeating. This selection is made to maximize the number of hosts that have access to a heartbeating datastore and minimize the likelihood that the datastores are backed by the same LUN or NFS server.

You can use the advanced option `das.heartbeatdsperhost` to change the number of heartbeat datastores selected by vCenter Server for each host. The default is two and the maximum valid value is five.

vSphere HA creates a directory at the root of each datastore that is used for both datastore heartbeating and for persisting the set of protected virtual machines. The name of the directory is `.vSphere-HA`. Do not delete or modify the files stored in this directory, because this can have an impact on operations. Because more than one cluster might use a datastore, subdirectories for this directory are created for each cluster. Root owns these directories and files and only root can read and write to them. The disk space used by vSphere HA depends on several factors including which VMFS version is in use and the number of hosts that use the datastore for heartbeating. With vmfs3, the maximum usage is approximately 2GB and the typical usage is approximately 3MB. With vmfs5 the maximum and typical usage is approximately 3MB. vSphere HA use of the datastores adds negligible overhead and has no performance impact on other datastore operations.

vSphere HA limits the number of virtual machines that can have configuration files on a single datastore. See *Configuration Maximums* for updated limits. If you place more than this number of virtual machines on a datastore and power them on, vSphere HA protects a number of virtual machines only up to the limit.

NOTE A Virtual SAN datastore cannot be used for datastore heartbeating. Therefore, if no other shared storage is accessible to all hosts in the cluster, there can be no heartbeat datastores in use. However, if you have storage that can be reached by an alternate network path that is independent of the Virtual SAN network, you can use it to set up a heartbeat datastore.

vSphere HA Security

vSphere HA is enhanced by several security features.

Select firewall ports opened

vSphere HA uses TCP and UDP port 8182 for agent-to-agent communication. The firewall ports open and close automatically to ensure they are open only when needed.

Configuration files protected using file system permissions

vSphere HA stores configuration information on the local storage or on ramdisk if there is no local datastore. These files are protected using file system permissions and they are accessible only to the root user. Hosts without local storage are only supported if they are managed by Auto Deploy.

Detailed logging

The location where vSphere HA places log files depends on the version of host.

- For ESXi 5.x hosts, vSphere HA writes to syslog only by default, so logs are placed where syslog is configured to put them. The log file names for vSphere HA are prepended with `fdm`, fault domain manager, which is a service of vSphere HA.
- For legacy ESXi 4.x hosts, vSphere HA writes to `/var/log/vmware/fdm` on local disk, as well as syslog if it is configured.
- For legacy ESX 4.x hosts, vSphere HA writes to `/var/log/vmware/fdm`.

Secure vSphere HA logins

vSphere HA logs onto the vSphere HA agents using a user account, **vpxuser**, created by vCenter Server. This account is the same account used by vCenter Server to manage the host. vCenter Server creates a random password for this account and changes the password periodically. The time period is set by the vCenter Server `VirtualCenter.VimPasswordExpirationInDays` setting. Users with administrative privileges on the root folder of the host can log in to the agent.

Secure communication

All communication between vCenter Server and the vSphere HA agent is done over SSL. Agent-to-agent communication also uses SSL except for election messages, which occur over UDP. Election messages are verified over SSL so that a rogue agent can prevent only the host on which the agent is running from being elected as a master host. In this case, a configuration issue for the cluster is issued so the user is aware of the problem.

Host SSL certificate verification required

vSphere HA requires that each host have a verified SSL certificate. Each host generates a self-signed certificate when it is booted for the first time. This certificate can then be regenerated or replaced with one issued by an authority. If the certificate is replaced, vSphere HA needs to be reconfigured on the host. If a host becomes disconnected from vCenter Server after its certificate is updated and the ESXi or ESX Host agent is restarted, then vSphere HA is automatically reconfigured when the host is reconnected to vCenter Server. If the disconnection does not occur because vCenter Server host SSL certificate verification is disabled at the time, verify the new certificate and reconfigure vSphere HA on the host.

vSphere HA Admission Control

vCenter Server uses admission control to ensure that sufficient resources are available in a cluster to provide failover protection and to ensure that virtual machine resource reservations are respected.

Three types of admission control are available.

Host	Ensures that a host has sufficient resources to satisfy the reservations of all virtual machines running on it.
Resource Pool	Ensures that a resource pool has sufficient resources to satisfy the reservations, shares, and limits of all virtual machines associated with it.
vSphere HA	Ensures that sufficient resources in the cluster are reserved for virtual machine recovery in the event of host failure.

Admission control imposes constraints on resource usage and any action that would violate these constraints is not permitted. Examples of actions that could be disallowed include the following:

- Powering on a virtual machine.
- Migrating a virtual machine onto a host or into a cluster or resource pool.
- Increasing the CPU or memory reservation of a virtual machine.

Of the three types of admission control, only vSphere HA admission control can be disabled. However, without it there is no assurance that the expected number of virtual machines can be restarted after a failure. Do not permanently disable admission control, however you might need to do so temporarily, for the following reasons:

- If you need to violate the failover constraints when there are not enough resources to support them--for example, if you are placing hosts in standby mode to test them for use with Distributed Power Management (DPM).
- If an automated process needs to take actions that might temporarily violate the failover constraints (for example, as part of an upgrade or patching of ESXi hosts as directed by vSphere Update Manager).
- If you need to perform testing or maintenance operations.

Admission control sets aside capacity, but when a failure occurs vSphere HA uses whatever capacity is available for virtual machine restarts. For example, vSphere HA places more virtual machines on a host than admission control would allow for user-initiated power ons.

Note When vSphere HA admission control is disabled, vSphere HA ensures that there are at least two powered-on hosts in the cluster even if DPM is enabled and can consolidate all virtual machines onto a single host. This is to ensure that failover is possible.

Host Failures Cluster Tolerates Admission Control Policy

You can configure vSphere HA to tolerate a specified number of host failures. With the Host Failures Cluster Tolerates admission control policy, vSphere HA ensures that a specified number of hosts can fail and sufficient resources remain in the cluster to fail over all the virtual machines from those hosts.

With the Host Failures Cluster Tolerates policy, vSphere HA performs admission control in the following way:

1 Calculates the slot size.

   A slot is a logical representation of memory and CPU resources. By default, it is sized to satisfy the requirements for any powered-on virtual machine in the cluster.

2 Determines how many slots each host in the cluster can hold.

3 Determines the Current Failover Capacity of the cluster.

   This is the number of hosts that can fail and still leave enough slots to satisfy all of the powered-on virtual machines.

4 Determines whether the Current Failover Capacity is less than the Configured Failover Capacity (provided by the user).

   If it is, admission control disallows the operation.

NOTE You can set a specific slot size for both CPU and memory in the admission control section of the vSphere HA settings in the vSphere Web Client.

Slot Size Calculation

vSphere HA Slot Size and Admission Control
(http://link.brightcove.com/services/player/bcpid2296383276001?bctid=ref:video_vsphere_slot_admission_control)

Slot size is comprised of two components, CPU and memory.

- vSphere HA calculates the CPU component by obtaining the CPU reservation of each powered-on virtual machine and selecting the largest value. If you have not specified a CPU reservation for a virtual machine, it is assigned a default value of 32MHz. You can change this value by using the `das.vmcpuminmhz` advanced option.)
- vSphere HA calculates the memory component by obtaining the memory reservation, plus memory overhead, of each powered-on virtual machine and selecting the largest value. There is no default value for the memory reservation.

If your cluster contains any virtual machines that have much larger reservations than the others, they will distort slot size calculation. To avoid this, you can specify an upper bound for the CPU or memory component of the slot size by using the `das.slotcpuinmhz` or `das.slotmeminmb` advanced options, respectively. See "vSphere HA Advanced Options," on page 38.

You can also determine the risk of resource fragmentation in your cluster by viewing the number of virtual machines that require multiple slots. This can be calculated in the admission control section of the vSphere HA settings in the vSphere Web Client. Virtual machines might require multiple slots if you have specified a fixed slot size or a maximum slot size using advanced options.

Using Slots to Compute the Current Failover Capacity

After the slot size is calculated, vSphere HA determines each host's CPU and memory resources that are available for virtual machines. These amounts are those contained in the host's root resource pool, not the total physical resources of the host. The resource data for a host that is used by vSphere HA can be found on the host's **Summary** tab on the vSphere Web Client. If all hosts in your cluster are the same, this data can be obtained by dividing the cluster-level figures by the number of hosts. Resources being used for virtualization purposes are not included. Only hosts that are connected, not in maintenance mode, and that have no vSphere HA errors are considered.

The maximum number of slots that each host can support is then determined. To do this, the host's CPU resource amount is divided by the CPU component of the slot size and the result is rounded down. The same calculation is made for the host's memory resource amount. These two numbers are compared and the smaller number is the number of slots that the host can support.

The Current Failover Capacity is computed by determining how many hosts (starting from the largest) can fail and still leave enough slots to satisfy the requirements of all powered-on virtual machines.

Advanced Runtime Info

When you select the Host Failures Cluster Tolerates admission control policy, the **Advanced Runtime Info** pane appears in the vSphere HA section of the cluster's **Monitor** tab in the vSphere Web Client. This pane displays the following information about the cluster:

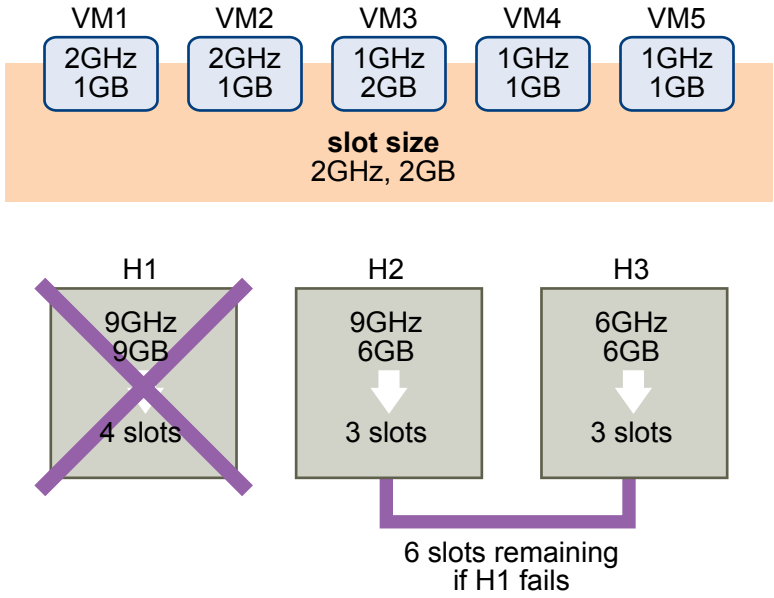
- Slot size.
- Total slots in cluster. The sum of the slots supported by the good hosts in the cluster.
- Used slots. The number of slots assigned to powered-on virtual machines. It can be more than the number of powered-on virtual machines if you have defined an upper bound for the slot size using the advanced options. This is because some virtual machines can take up multiple slots.
- Available slots. The number of slots available to power on additional virtual machines in the cluster. vSphere HA reserves the required number of slots for failover. The remaining slots are available to power on new virtual machines.
- Failover slots. The total number of slots not counting the used slots or the available slots.
- Total number of powered on virtual machines in cluster.
- Total number of hosts in cluster.
- Total good hosts in cluster. The number of hosts that are connected, not in maintenance mode, and have no vSphere HA errors.

Example: Admission Control Using Host Failures Cluster Tolerates Policy

The way that slot size is calculated and used with this admission control policy is shown in an example. Make the following assumptions about a cluster:

- The cluster is comprised of three hosts, each with a different amount of available CPU and memory resources. The first host (H1) has 9GHz of available CPU resources and 9GB of available memory, while Host 2 (H2) has 9GHz and 6GB and Host 3 (H3) has 6GHz and 6GB.
- There are five powered-on virtual machines in the cluster with differing CPU and memory requirements. VM1 needs 2GHz of CPU resources and 1GB of memory, while VM2 needs 2GHz and 1GB, VM3 needs 1GHz and 2GB, VM4 needs 1GHz and 1GB, and VM5 needs 1GHz and 1GB.
- The Host Failures Cluster Tolerates is set to one.

Figure 2-1. Admission Control Example with Host Failures Cluster Tolerates Policy



1 Slot size is calculated by comparing both the CPU and memory requirements of the virtual machines and selecting the largest.

   The largest CPU requirement (shared by VM1 and VM2) is 2GHz, while the largest memory requirement (for VM3) is 2GB. Based on this, the slot size is 2GHz CPU and 2GB memory.

2 Maximum number of slots that each host can support is determined.

   H1 can support four slots. H2 can support three slots (which is the smaller of 9GHz/2GHz and 6GB/2GB) and H3 can also support three slots.

3 Current Failover Capacity is computed.

   The largest host is H1 and if it fails, six slots remain in the cluster, which is sufficient for all five of the powered-on virtual machines. If both H1 and H2 fail, only three slots remain, which is insufficient. Therefore, the Current Failover Capacity is one.

The cluster has one available slot (the six slots on H2 and H3 minus the five used slots).

Percentage of Cluster Resources Reserved Admission Control Policy

You can configure vSphere HA to perform admission control by reserving a specific percentage of cluster CPU and memory resources for recovery from host failures.

With the Percentage of Cluster Resources Reserved admission control policy, vSphere HA ensures that a specified percentage of aggregate CPU and memory resources are reserved for failover.

With the Cluster Resources Reserved policy, vSphere HA enforces admission control as follows:

1 Calculates the total resource requirements for all powered-on virtual machines in the cluster.

2 Calculates the total host resources available for virtual machines.

3 Calculates the Current CPU Failover Capacity and Current Memory Failover Capacity for the cluster.

4 Determines if either the Current CPU Failover Capacity or Current Memory Failover Capacity is less than the corresponding Configured Failover Capacity (provided by the user).

   If so, admission control disallows the operation.

vSphere HA uses the actual reservations of the virtual machines. If a virtual machine does not have reservations, meaning that the reservation is 0, a default of 0MB memory and 32MHz CPU is applied.

NOTE The Percentage of Cluster Resources Reserved admission control policy also checks that there are at least two vSphere HA-enabled hosts in the cluster (excluding hosts that are entering maintenance mode). If there is only one vSphere HA-enabled host, an operation is not allowed, even if there is a sufficient percentage of resources available. The reason for this extra check is that vSphere HA cannot perform failover if there is only a single host in the cluster.

Computing the Current Failover Capacity

The total resource requirements for the powered-on virtual machines is comprised of two components, CPU and memory. vSphere HA calculates these values.

- The CPU component by summing the CPU reservations of the powered-on virtual machines. If you have not specified a CPU reservation for a virtual machine, it is assigned a default value of 32MHz (this value can be changed using the `das.vmcpuminmhz` advanced option.)
- The memory component by summing the memory reservation (plus memory overhead) of each powered-on virtual machine.

The total host resources available for virtual machines is calculated by adding the hosts' CPU and memory resources. These amounts are those contained in the host's root resource pool, not the total physical resources of the host. Resources being used for virtualization purposes are not included. Only hosts that are connected, not in maintenance mode, and have no vSphere HA errors are considered.

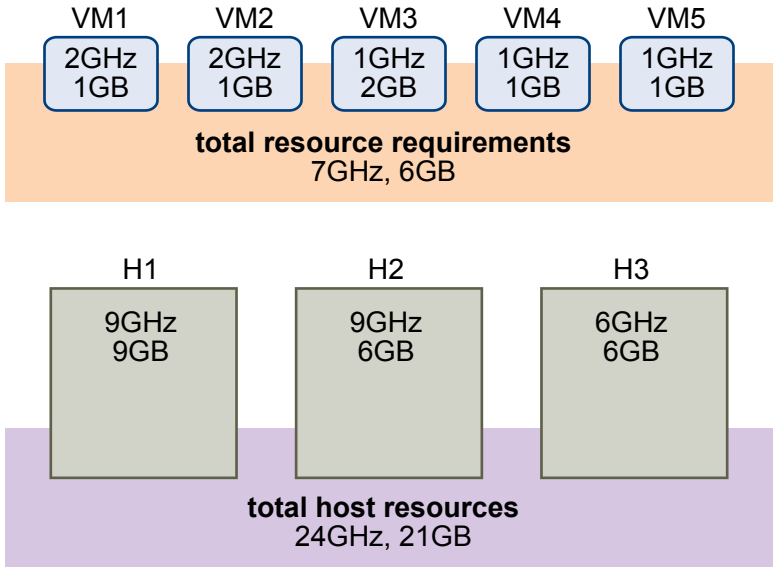
The Current CPU Failover Capacity is computed by subtracting the total CPU resource requirements from the total host CPU resources and dividing the result by the total host CPU resources. The Current Memory Failover Capacity is calculated similarly.

Example: Admission Control Using Percentage of Cluster Resources Reserved Policy

The way that Current Failover Capacity is calculated and used with this admission control policy is shown with an example. Make the following assumptions about a cluster:

- The cluster is comprised of three hosts, each with a different amount of available CPU and memory resources. The first host (H1) has 9GHz of available CPU resources and 9GB of available memory, while Host 2 (H2) has 9GHz and 6GB and Host 3 (H3) has 6GHz and 6GB.
- There are five powered-on virtual machines in the cluster with differing CPU and memory requirements. VM1 needs 2GHz of CPU resources and 1GB of memory, while VM2 needs 2GHz and 1GB, VM3 needs 1GHz and 2GB, VM4 needs 1GHz and 1GB, and VM5 needs 1GHz and 1GB.
- The Configured Failover Capacity for CPU and Memory are both set to 25%.

Figure 2-2. Admission Control Example with Percentage of Cluster Resources Reserved Policy



The total resource requirements for the powered-on virtual machines is 7GHz and 6GB. The total host resources available for virtual machines is 24GHz and 21GB. Based on this, the Current CPU Failover Capacity is 70% ((24GHz - 7GHz)/24GHz). Similarly, the Current Memory Failover Capacity is 71% ((21GB-6GB)/21GB).

Because the cluster's Configured Failover Capacity is set to 25%, 45% of the cluster's total CPU resources and 46% of the cluster's memory resources are still available to power on additional virtual machines.

Specify Failover Hosts Admission Control Policy

You can configure vSphere HA to designate specific hosts as the failover hosts.

With the Specify Failover Hosts admission control policy, when a host fails, vSphere HA attempts to restart its virtual machines on any of the specified failover hosts. If this is not possible, for example the failover hosts have failed or have insufficient resources, then vSphere HA attempts to restart those virtual machines on other hosts in the cluster.

To ensure that spare capacity is available on a failover host, you are prevented from powering on virtual machines or using vMotion to migrate virtual machines to a failover host. Also, DRS does not use a failover host for load balancing.

Note If you use the Specify Failover Hosts admission control policy and designate multiple failover hosts, DRS does not attempt to enforce VM-VM affinity rules for virtual machines that are running on failover hosts.

The Current Failover Hosts appear in the vSphere HA section of the cluster's **Summary** tab. The status icon next to each host can be green, yellow, or red.

- Green. The host is connected, not in maintenance mode, and has no vSphere HA errors. No powered-on virtual machines reside on the host.
- Yellow. The host is connected, not in maintenance mode, and has no vSphere HA errors. However, powered-on virtual machines reside on the host.
- Red. The host is disconnected, in maintenance mode, or has vSphere HA errors.

Choosing an Admission Control Policy

You should choose a vSphere HA admission control policy based on your availability needs and the characteristics of your cluster. When choosing an admission control policy, you should consider a number of factors.

Avoiding Resource Fragmentation

Resource fragmentation occurs when there are enough resources in aggregate for a virtual machine to be failed over. However, those resources are located on multiple hosts and are unusable because a virtual machine can run on one ESXi host at a time. The default configuration of the Host Failures Cluster Tolerates policy avoids resource fragmentation by defining a slot as the maximum virtual machine reservation. The Percentage of Cluster Resources policy does not address the problem of resource fragmentation. With the Specify Failover Hosts policy, resources are not fragmented because hosts are reserved for failover.

Flexibility of Failover Resource Reservation

Admission control policies differ in the granularity of control they give you when reserving cluster resources for failover protection. The Host Failures Cluster Tolerates policy allows you to set the failover level as a number of hosts. The Percentage of Cluster Resources policy allows you to designate up to 100% of cluster CPU or memory resources for failover. The Specify Failover Hosts policy allows you to specify a set of failover hosts.

Heterogeneity of Cluster

Clusters can be heterogeneous in terms of virtual machine resource reservations and host total resource capacities. In a heterogeneous cluster, the Host Failures Cluster Tolerates policy can be too conservative because it only considers the largest virtual machine reservations when defining slot size and assumes the largest hosts fail when computing the Current Failover Capacity. The other two admission control policies are not affected by cluster heterogeneity.

NOTE vSphere HA includes the resource usage of Fault Tolerance Secondary VMs when it performs admission control calculations. For the Host Failures Cluster Tolerates policy, a Secondary VM is assigned a slot, and for the Percentage of Cluster Resources policy, the Secondary VM's resource usage is accounted for when computing the usable capacity of the cluster.

---

vSphere HA Interoperability

vSphere HA can interoperate with many other features, such as DRS and Virtual SAN.

Before configuring vSphere HA, you should be aware of the limitations of its interoperability with these other features or products.

Using vSphere HA with Virtual SAN

You can use Virtual SAN as the shared storage for a vSphere HA cluster. When enabled, Virtual SAN aggregates the specified local storage disks available on the hosts into a single datastore shared by all hosts.

To use vSphere HA with Virtual SAN, you must be aware of certain considerations and limitations for the interoperability of these two features.

For information about Virtual SAN, see *VMware Virtual SAN*.

ESXi Host Requirements

You can use Virtual SAN with a vSphere HA cluster only if the following conditions are met:

- The cluster's ESXi hosts all must be version 5.5 or later.

- The cluster must have a minimum of three ESXi hosts.

Networking Differences

Virtual SAN has its own network. When Virtual SAN and vSphere HA are enabled for the same cluster, the HA interagent traffic flows over this storage network rather than the management network. The management network is used by vSphere HA only when Virtual SAN is disabled. vCenter Server chooses the appropriate network when vSphere HA is configured on a host.

NOTE Virtual SAN can only be enabled when vSphere HA is disabled.

If you change the Virtual SAN network configuration, the vSphere HA agents do not automatically pick up the new network settings. So to make changes to the Virtual SAN network, you must take the following steps in the vSphere Web Client:

1 Disable Host Monitoring for the vSphere HA cluster.

2 Make the Virtual SAN network changes.

3 Right-click all hosts in the cluster and select **Reconfigure for vSphere HA**.

4 Re-enable Host Monitoring for the vSphere HA cluster.

Table 2-2 shows the differences in vSphere HA networking when Virtual SAN is used or not.

Table 2-2. vSphere HA networking differences

	Virtual SAN Enabled	Virtual SAN Disabled
Network used by vSphere HA	Virtual SAN storage network	Management network
Heartbeat datastores	Any datastore mounted to > 1 host, but not Virtual SAN datastores	Any datastore mounted to > 1 host
Host declared isolated	Isolation addresses not pingable and Virtual SAN storage network inaccessible	Isolation addresses not pingable and management network inaccessible

Capacity Reservation Settings

When you reserve capacity for your vSphere HA cluster with an admission control policy, this setting must be coordinated with the corresponding Virtual SAN setting that ensures data accessibility on failures. Specifically, the Number of Failures Tolerated setting in the Virtual SAN rule set must not be lower than the capacity reserved by the vSphere HA admission control setting.

For example, if the Virtual SAN rule set allows for only two failures, the vSphere HA admission control policy must reserve capacity that is equivalent to only one or two host failures. If you are using the Percentage of Cluster Resources Reserved policy for a cluster that has eight hosts, you must not reserve more than 25% of the cluster resources. In the same cluster, with the Host Failures Cluster Tolerates policy, the setting must not be higher than two hosts. If less capacity is reserved by vSphere HA, failover activity might be unpredictable, while reserving too much capacity overly constrains the powering on of virtual machines and inter-cluster vMotion migrations.

Using vSphere HA and DRS Together

Using vSphere HA with Distributed Resource Scheduler (DRS) combines automatic failover with load balancing. This combination can result in a more balanced cluster after vSphere HA has moved virtual machines to different hosts.

When vSphere HA performs failover and restarts virtual machines on different hosts, its first priority is the immediate availability of all virtual machines. After the virtual machines have been restarted, those hosts on which they were powered on might be heavily loaded, while other hosts are comparatively lightly loaded. vSphere HA uses the virtual machine's CPU and memory reservation and overhead memory to determine if a host has enough spare capacity to accommodate the virtual machine.

In a cluster using DRS and vSphere HA with admission control turned on, virtual machines might not be evacuated from hosts entering maintenance mode. This behavior occurs because of the resources reserved for restarting virtual machines in the event of a failure. You must manually migrate the virtual machines off of the hosts using vMotion.

In some scenarios, vSphere HA might not be able to fail over virtual machines because of resource constraints. This can occur for several reasons.

- HA admission control is disabled and Distributed Power Management (DPM) is enabled. This can result in DPM consolidating virtual machines onto fewer hosts and placing the empty hosts in standby mode leaving insufficient powered-on capacity to perform a failover.
- VM-Host affinity (required) rules might limit the hosts on which certain virtual machines can be placed.
- There might be sufficient aggregate resources but these can be fragmented across multiple hosts so that they can not be used by virtual machines for failover.

In such cases, vSphere HA can use DRS to try to adjust the cluster (for example, by bringing hosts out of standby mode or migrating virtual machines to defragment the cluster resources) so that HA can perform the failovers.

If DPM is in manual mode, you might need to confirm host power-on recommendations. Similarly, if DRS is in manual mode, you might need to confirm migration recommendations.

If you are using VM-Host affinity rules that are required, be aware that these rules cannot be violated. vSphere HA does not perform a failover if doing so would violate such a rule.

For more information about DRS, see the *vSphere Resource Management* documentation.

vSphere HA and DRS Affinity Rules

If you create a DRS affinity rule for your cluster, you can specify how vSphere HA applies that rule during a virtual machine failover.

The two types of rules for which you can specify vSphere HA failover behavior are the following:

- VM anti-affinity rules force specified virtual machines to remain apart during failover actions.
- VM-Host affinity rules place specified virtual machines on a particular host or a member of a defined group of hosts during failover actions.

When you edit a DRS affinity rule, select the checkbox or checkboxes that enforce the desired failover behavior for vSphere HA.

- **HA must respect VM anti-affinity rules during failover** -- if VMs with this rule would be placed together, the failover is aborted.

- **HA should respect VM to Host affinity rules during failover** --vSphere HA attempts to place VMs with this rule on the specified hosts if at all possible.

NOTE vSphere HA can restart a VM in a DRS-disabled cluster, overriding a VM-Host affinity rules mapping if the host failure happens soon (by default, within 5 minutes) after setting the rule.

Other vSphere HA Interoperability Issues

To use vSphere HA, you must be aware of the following additional interoperability issues.

VM Component Protection

VM Component Protection (VMCP) has the following interoperability issues and limitations:

- VMCP does not support vSphere Fault Tolerance. If VMCP is enabled for a cluster using Fault Tolerance, the affected FT virtual machines will automatically receive overrides that disable VMCP.
- VMCP does not detect or respond to accessibility issues for files located on Virtual SAN datastores. If a virtual machine's configuration and VMDK files are located only on Virtual SAN datastores, they are not protected by VMCP.
- VMCP does not detect or respond to accessibility issues for files located on Virtual Volume datastores. If a virtual machine's configuration and VMDK files are located only on Virtual Volume datastores, they are not protected by VMCP.
- VMCP does not protect against inaccessible Raw Device Mapping (RDM)s.

IPv6

vSphere HA can be used with IPv6 network configurations, which are fully supported if the following considerations are observed:

- The cluster contains only ESXi 6.0 or later hosts.
- The management network for all hosts in the cluster must be configured with the same IP version, either IPv6 or IPv4. vSphere HA clusters cannot contain both types of networking configuration.
- The network isolation addresses used by vSphere HA must match the IP version used by the cluster for its management network.
- IPv6 cannot be used in vSphere HA clusters that also utilize Virtual SAN.

In addition to the previous restrictions, the following types of IPv6 address types are not supported for use with the vSphere HA isolation address or management network: link-local, ORCHID, and link-local with zone indices. Also, the loopback address type cannot be used for the management network.

NOTE To upgrade an existing IPv4 deployment to IPv6, you must first disable vSphere HA.

Creating and Configuring a vSphere HA Cluster

vSphere HA operates in the context of a cluster of ESXi (or legacy ESX) hosts. You must create a cluster, populate it with hosts, and configure vSphere HA settings before failover protection can be established.

When you create a vSphere HA cluster, you must configure a number of settings that determine how the feature works. Before you do this, identify your cluster's nodes. These nodes are the ESXi hosts that will provide the resources to support virtual machines and that vSphere HA will use for failover protection. You should then determine how those nodes are to be connected to one another and to the shared storage where your virtual machine data resides. After that networking architecture is in place, you can add the hosts to the cluster and finish configuring vSphere HA.

You can enable and configure vSphere HA before you add host nodes to the cluster. However, until the hosts are added, your cluster is not fully operational and some of the cluster settings are unavailable. For example, the Specify a Failover Host admission control policy is unavailable until there is a host that can be designated as the failover host.

NOTE The Virtual Machine Startup and Shutdown (automatic startup) feature is disabled for all virtual machines residing on hosts that are in (or moved into) a vSphere HA cluster. Automatic startup is not supported when used with vSphere HA.

vSphere HA Checklist

The vSphere HA checklist contains requirements that you must be aware of before creating and using a vSphere HA cluster.

Review this list before you set up a vSphere HA cluster. For more information, follow the appropriate cross reference.

- All hosts must be licensed for vSphere HA.
- A cluster must contain at least two hosts.
- All hosts must be configured with static IP addresses. If you are using DHCP, you must ensure that the address for each host persists across reboots.
- All hosts must have at least one management network in common. The best practice is to have at least two management networks in common. You should use the VMkernel network with the **Management traffic** checkbox enabled. The networks must be accessible to each other and vCenter Server and the hosts must be accessible to each other on the management networks. See"Best Practices for Networking," on page 40.
- To ensure that any virtual machine can run on any host in the cluster, all hosts must have access to the same virtual machine networks and datastores. Similarly, virtual machines must be located on shared, not local, storage otherwise they cannot be failed over in the case of a host failure.

  NOTE vSphere HA uses datastore heartbeating to distinguish between partitioned, isolated, and failed hosts. So if some datastores are more reliable in your environment, configure vSphere HA to give preference to them.

- For VM Monitoring to work, VMware tools must be installed. See "VM and Application Monitoring," on page 18.
- vSphere HA supports both IPv4 and IPv6. See "Other vSphere HA Interoperability Issues," on page 32 for considerations when using IPv6.
- For VM Component Protection to work, hosts must have the All Paths Down (APD) Timeout feature enabled.
- To use VM Component Protection, clusters must contain ESXi 6.0 hosts or later.
- Only vSphere HA clusters that contain ESXi 6.0 or later hosts can be used to enable VMCP. Clusters that contain hosts from an earlier release cannot enable VMCP, and such hosts cannot be added to a VMCP-enabled cluster.
- If your cluster uses Virtual Volume datastores, when vSphere HA is enabled a configuration Virtual Volume is created on each datastore by vCenter Server. In these containers, vSphere HA stores the files it uses to protect virtual machines. vSphere HA does not function correctly if you delete these containers. Only one container is created per Virtual Volume datastore.

Create a vSphere HA Cluster

To enable your cluster for vSphere HA, you must first create an empty cluster. After you plan the resources and networking architecture of your cluster, use the vSphere Web Client to add hosts to the cluster and specify the cluster's vSphere HA settings.

A vSphere HA-enabled cluster is a prerequisite for Fault Tolerance.

Prerequisites

- Verify that all virtual machines and their configuration files reside on shared storage.
- Verify that the hosts are configured to access the shared storage so that you can power on the virtual machines by using different hosts in the cluster.
- Verify that hosts are configured to have access to the virtual machine network.
- Verify that you are using redundant management network connections for vSphere HA. For information about setting up network redundancy, see "Best Practices for Networking," on page 40.
- Verify that you have configured hosts with at least two datastores to provide redundancy for vSphere HA datastore heartbeating.
- Connect vSphere Web Client to vCenter Server using an account with cluster administrator permissions.

Procedure

1. In the vSphere Web Client, browse to the data center where you want the cluster to reside and click **Create a Cluster**.
2. Complete the New Cluster wizard.

   Do not turn on vSphere HA (or DRS).
3. Click **OK** to close the wizard and create an empty cluster.
4. Based on your plan for the resources and networking architecture of the cluster, use the vSphere Web Client to add hosts to the cluster.
5. Browse to the cluster and enable vSphere HA.
   a. Click the **Manage** tab and click **Settings**.
   b. Select **vSphere HA** and click **Edit**.
   c. Select **Turn ON vSphere HA**.
6. Select **Host Monitoring**

   Enabling Host Monitoring allows hosts in the cluster to exchange network heartbeats and allows vSphere HA to take action when it detects failures. Host Monitoring is required for the vSphere Fault Tolerance recovery process to work properly.
7. Choose a setting for **Virtual Machine Monitoring**.

   Select **VM Monitoring Only** to restart individual virtual machines if their heartbeats are not received within a set time. You can also select **VM and Application Monitoring** to enable application monitoring.
8. Click **OK**.

You have a vSphere HA cluster, populated with hosts.

What to do next

Configure the vSphere HA settings as appropriate for your cluster.

- Failure conditions and VM response
- Admission Control
- Datastore for Heartbeating
- Advanced Options

See "Configuring vSphere HA Cluster Settings," on page 35.

Configuring vSphere HA Cluster Settings

When you create a vSphere HA cluster or configure an existing cluster, you must configure settings that determine how the feature works.

In the vSphere Web Client, you can configure following the vSphere HA settings:

Failure conditions and VM response	Provide settings here for VM restart priority, Host isolation response, VM monitoring sensitivity, and VM Component Protection.
Admission Control	Enable or disable admission control for the vSphere HA cluster and choose a policy for how it is enforced.
Datastore for Heartbeating	Specify preferences for the datastores that vSphere HA uses for datastore heartbeating.
Advanced Options	Customize vSphere HA behavior by setting advanced options.

NOTE You can check the status of vSphere HA configuration tasks on each of the hosts in the Tasks console of the vSphere Web Client.

Configure Virtual Machine Responses

The Failure conditions and VM response page allows you to choose settings that determine how vSphere HA responds to host failures and isolations. These settings include the VM restart priority, host isolation response, settings for VM Component Protection, and VM monitoring sensitivity.

Virtual Machine Response page is editable only if you enabled vSphere HA.

Procedure

1 In the vSphere Web Client, browse to the vSphere HA cluster.

2 Click the **Manage** tab and click **Settings**.

3 Under Settings, select **vSphere HA** and click **Edit**.

4 Expand **Failure Conditions and VM Response** to display the configuration options.

Option	Description
VM restart priority	The restart priority determines the order in which virtual machines are restarted when the host fails. Higher priority virtual machines are started first. This priority applies only on a per-host basis. If multiple hosts fail, all virtual machines are migrated from the first host in order of priority, then all virtual machines from the second host in order of priority, and so on.
Response for Host Isolation	The host isolation response determines what happens when a host in a vSphere HA cluster loses its console network connection, but continues running.

Option	Description
Response for Datastore with Permanent Device Loss (PDL)	This setting determines what VMCP does in the case of a PDL failure. You can choose to have it **Issue Events** or **Power off and restart VMs**.
Response for Datastore with All Paths Down (APD)	This setting determines what VMCP does in the case of an APD failure. You can choose to have it **Issue Events** or **Power off and restart VMs** conservatively or aggressively.
Delay for VM failover for APD	This setting is the number of minutes that VMCP waits before taking action.
Response for APD recovery after APD timeout	You can choose whether or not VMCP resets a VM in this situation.
VM monitoring sensitivity	Set this by by moving the slider between **Low** and **High**. You can also select **Custom** to provide custom settings.

5 Click **OK**.

Your Virtual Machine Response settings take effect.

Configure Admission Control

After you create a cluster, admission control allows you to specify whether virtual machines can be started if they violate availability constraints. The cluster reserves resources to allow failover for all running virtual machines on the specified number of hosts.

The Admission Control page appears only if you enabled vSphere HA.

Procedure

1 In the vSphere Web Client, browse to the vSphere HA cluster.

2 Click the **Manage** tab and click **Settings**.

3 Under Settings, select **vSphere HA** and click **Edit**.

4 Expand **Admission Control** to display the configuration options.

5 Select an admission control policy to apply to the cluster.

Option	Description
Define failover capacity by static number of hosts	Select the maximum number of host failures that you can recover from or to guarantee failover for. Also, you must select a slot size policy.
Define failover capacity by reserving a percentage of the cluster resources	Specify a percentage of the cluster's CPU and Memory resources to reserve as spare capacity to support failovers.
Use dedicated failover hosts	Select hosts to use for failover actions. Failovers can still occur to other hosts in the cluster if a default failover host does not have enough resources.
Do not reserve failover capacity	This option allows virtual machine power-ons that violate availability constraints.

6 Click **OK**.

Admission control is enabled and the policy that you chose takes effect.

Configure Datastore for Heartbeating

vSphere HA uses datastore heartbeating to distinguish between hosts that have failed and hosts that reside on a network partition. Datastore heartbeating allows vSphere HA to monitor hosts when a management network partition occurs and to continue to respond to failures that occur.

You can specify the datastores that you want to be used for datastore heartbeating.

Procedure

1 In the vSphere Web Client, browse to the vSphere HA cluster.

2 Click the **Manage** tab and click **Settings**.

3 Under Settings, select **vSphere HA** and click **Edit**.

4 Expand **Datastore for Heartbeating** to display the configuration options for datastore heartbeating.

5 To instruct vSphere HA about how to select the datastores and how to treat your preferences, choose from the following options:

Table 2-3.

Datastore Heartbeating Options
Automatically select datastores accessible from the host
Use datastores only from the specified list
Use datastores from the specified list and complement automatically if needed

6 In the **Available heartbeat datastores** pane, select the datastores that you want to use for heartbeating.

The datastores listed are those shared by more than one host in the vSphere HA cluster. When a datastore is selected, the lower pane displays all the hosts in the vSphere HA cluster that can access it.

7 Click **OK**.

Set Advanced Options

To customize vSphere HA behavior, set advanced vSphere HA options.

Prerequisites

Verify that you have cluster administrator privileges.

NOTE Because these options affect the functioning of vSphere HA, change them with caution.

Procedure

1 In the vSphere Web Client, browse to the vSphere HA cluster.

2 Click the **Manage** tab and click **Settings**.

3 Under Settings, select **vSphere HA** and click **Edit**.

4 Expand **Advanced Options**.

5 Click **Add** and type the name of the advanced option in the text box.

You can set the value of the option in the text box in the Value column.

6 Repeat step 5 for each new option that you want to add and click **OK**.

The cluster uses the options that you added or modified.

What to do next

Once you have set an advanced vSphere HA option, it persists until you do one the following:

- Using the vSphere Web Client, reset its value to the default value.
- Manually edit or delete the option from the fdm.cfg file on all hosts in the cluster.

vSphere HA Advanced Options

You can set advanced options that affect the behavior of your vSphere HA cluster.

Table 2-4. vSphere HA Advanced Options

Option	Description
`das.isolationaddress[...]`	Sets the address to ping to determine if a host is isolated from the network. This address is pinged only when heartbeats are not received from any other host in the cluster. If not specified, the default gateway of the management network is used. This default gateway has to be a reliable address that is available, so that the host can determine if it is isolated from the network. You can specify multiple isolation addresses (up to 10) for the cluster: `das.isolationaddressX`, where X = 0-9. Typically you should specify one per management network. Specifying too many addresses makes isolation detection take too long.
`das.usedefaultisolationaddress`	By default, vSphere HA uses the default gateway of the console network as an isolation address. This option specifies whether or not this default is used (true|false).
`das.isolationshutdowntimeout`	The period of time the system waits for a virtual machine to shut down before powering it off. This only applies if the host's isolation response is Shut down VM. Default value is 300 seconds.
`das.slotmeminmb`	Defines the maximum bound on the memory slot size. If this option is used, the slot size is the smaller of this value or the maximum memory reservation plus memory overhead of any powered-on virtual machine in the cluster.
`das.slotcpuinmhz`	Defines the maximum bound on the CPU slot size. If this option is used, the slot size is the smaller of this value or the maximum CPU reservation of any powered-on virtual machine in the cluster.
`das.vmmemoryminmb`	Defines the default memory resource value assigned to a virtual machine if its memory reservation is not specified or zero. This is used for the Host Failures Cluster Tolerates admission control policy. If no value is specified, the default is 0 MB.
`das.vmcpuminmhz`	Defines the default CPU resource value assigned to a virtual machine if its CPU reservation is not specified or zero. This is used for the Host Failures Cluster Tolerates admission control policy. If no value is specified, the default is 32MHz.
`das.iostatsinterval`	Changes the default I/O stats interval for VM Monitoring sensitivity. The default is 120 (seconds). Can be set to any value greater than, or equal to 0. Setting to 0 disables the check. **NOTE** Values of less than 50 are not recommended since smaller values can result in vSphere HA unexpectedly resetting a virtual machine.
`das.ignoreinsufficienthbdatastore`	Disables configuration issues created if the host does not have sufficient heartbeat datastores for vSphere HA. Default value is false.
`das.heartbeatdsperhost`	Changes the number of heartbeat datastores required. Valid values can range from 2-5 and the default is 2.

Table 2-4. vSphere HA Advanced Options (Continued)

Option	Description
fdm.isolationpolicydelaysec	The number of seconds system waits before executing the isolation policy once it is determined that a host is isolated. The minimum value is 30. If set to a value less than 30, the delay will be 30 seconds.
das.respectvmvmantiaffinityrules	Determines if vSphere HA enforces VM-VM anti-affinity rules. Default value is "false", whereby the rules are not enforced. Can also be set to "true" and rules are enforced (even if vSphere DRS is not enabled). In this case, vSphere HA does not fail over a virtual machine if doing so violates a rule, but it issues an event reporting there are insufficient resources to perform the failover. See *vSphere Resource Management* for more information on anti-affinity rules.
das.maxresets	The maximum number of reset attempts made by VMCP. If a reset operation on a virtual machine affected by an APD situation fails, VMCP retries the reset this many times before giving up
das.maxterminates	The maximum number of retries made by VMCP for virtual machine termination.
das.terminateretryintervalsec	If VMCP fails to terminate a virtual machine, this is the number of seconds the system waits before it retries a terminate attempt
das.config.fdm.reportfailoverfailevent	When set to 1, enables generation of a detailed per-VM event when an attempt by vSphere HA to restart a virtual machine is unsuccessful. Default value is 0. In versions earlier than vSphere 6.0, this event is generated by default.
vpxd.das.completemetadataupdateintervalsec	The period of time (seconds) after a VM-Host affinity rule is set during which vSphere HA can restart a VM in a DRS-disabled cluster, overriding the rule. Default value is 300 seconds.
das.config.fdm.memreservationmb	By default vSphere HA agents run with a configured memory limit of 250 MB. A host might not allow this reservation if it runs out of reservable capacity. You can use this advanced option to lower the memory limit to avoid this issue. Only integers greater than 100, which is the minimum value, can be specified. Conversely, to prevent problems during master agent elections in a large cluster (containing 6,000 to 8,000 VMs) you should raise this limit to 325 MB. **Note** Once this limit is changed, for all hosts in the cluster you must run the Reconfigure HA task. Also, when a new host is added to the cluster or an existing host is rebooted, this task should be performed on those hosts in order to update this memory setting.

Note If you change the value of any of the following advanced options, you must disable and then re-enable vSphere HA before your changes take effect.

- das.isolationaddress[...]
- das.usedefaultisolationaddress
- das.isolationshutdowntimeout

Customize an Individual Virtual Machine

Each virtual machine in a vSphere HA cluster is assigned the cluster default settings for VM Restart Priority, Host Isolation Response, VM Component Protection, and VM Monitoring. You can specify specific behavior for each virtual machine by changing these defaults. If the virtual machine leaves the cluster, these settings are lost.

Procedure

1 In the vSphere Web Client, browse to the vSphere HA cluster.

2 Click the **Manage** tab and click **Settings**.

3 Under Settings, select **VM Overrides** and click **Add**.

4 Use the **+** button to select virtual machines to which to apply the overrides.

5 Click **OK**.

6 (Optional) You can change other settings, such as the **Automation level**, **VM restart priority**, **Host isolation response**, VMCP settings,**VM Monitoring**, or **VM monitoring sensitivity** settings.

NOTE You can view the cluster defaults for these settings by first expanding **Relevant Cluster Settings** and then expanding **vSphere HA**.

7 Click **OK**.

The virtual machine's behavior now differs from the cluster defaults for each setting that you changed.

Best Practices for vSphere HA Clusters

To ensure optimal vSphere HA cluster performance, you should follow certain best practices. This section highlights some of the key best practices for a vSphere HA cluster.

You can also refer to the *vSphere High Availability Deployment Best Practices* publication for further discussion.

Best Practices for Networking

Observe the following best practices for the configuration of host NICs and network topology for vSphere HA. Best Practices include recommendations for your ESXi hosts, and for cabling, switches, routers, and firewalls.

Network Configuration and Maintenance

The following network maintenance suggestions can help you avoid the accidental detection of failed hosts and network isolation because of dropped vSphere HA heartbeats.

- When making changes to the networks that your clustered ESXi hosts are on, suspend the Host Monitoring feature. Changing your network hardware or networking settings can interrupt the heartbeats that vSphere HA uses to detect host failures, and this might result in unwanted attempts to fail over virtual machines.
- When you change the networking configuration on the ESXi hosts themselves, for example, adding port groups, or removing vSwitches, suspend Host Monitoring. After you have made the networking configuration changes, you must reconfigure vSphere HA on all hosts in the cluster, which causes the network information to be reinspected. Then re-enable Host Monitoring.

NOTE Because networking is a vital component of vSphere HA, if network maintenance needs to be performed inform the vSphere HA administrator.

Networks Used for vSphere HA Communications

To identify which network operations might disrupt the functioning of vSphere HA, you should know which management networks are being used for heart beating and other vSphere HA communications.

- On legacy ESX hosts in the cluster, vSphere HA communications travel over all networks that are designated as service console networks. VMkernel networks are not used by these hosts for vSphere HA communications. To contain vSphere HA traffic to a subset of the ESX console networks, use the `allowedNetworks` advanced option.
- On ESXi hosts in the cluster, vSphere HA communications, by default, travel over VMkernel networks. With an ESXi host, if you wish to use a network other than the one vCenter Server uses to communicate with the host for vSphere HA, you must explicitly enable the **Management traffic** checkbox.

To keep vSphere HA agent traffic on the networks you have specified, configure hosts so vmkNICs used by vSphere HA do not share subnets with vmkNICs used for other purposes. vSphere HA agents send packets using any pNIC that is associated with a given subnet if there is also at least one vmkNIC configured for vSphere HA management traffic. Consequently, to ensure network flow separation, the vmkNICs used by vSphere HA and by other features must be on different subnets.

Network Isolation Addresses

A network isolation address is an IP address that is pinged to determine whether a host is isolated from the network. This address is pinged only when a host has stopped receiving heartbeats from all other hosts in the cluster. If a host can ping its network isolation address, the host is not network isolated, and the other hosts in the cluster have either failed or are network partitioned. However, if the host cannot ping its isolation address, it is likely that the host has become isolated from the network and no failover action is taken.

By default, the network isolation address is the default gateway for the host. Only one default gateway is specified, regardless of how many management networks have been defined. You should use the `das.isolationaddress[...]` advanced option to add isolation addresses for additional networks. See "vSphere HA Advanced Options," on page 38.

Network Path Redundancy

Network path redundancy between cluster nodes is important for vSphere HA reliability. A single management network ends up being a single point of failure and can result in failovers although only the network has failed. If you have only one management network, any failure between the host and the cluster can cause an unnecessary (or false) failover activity if heartbeat datastore connectivity is not retained during the networking failure. Possible failures include NIC failures, network cable failures, network cable removal, and switch resets. Consider these possible sources of failure between hosts and try to minimize them, typically by providing network redundancy.

The first way you can implement network redundancy is at the NIC level with NIC teaming. Using a team of two NICs connected to separate physical switches improves the reliability of a management network. Because servers connected 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 management network, configure the vNICs in vSwitch configuration for Active or Standby configuration. The recommended parameter settings for the vNICs are:

- Default load balancing = route based on originating port ID
- Failback = No

After you have added a NIC to a host in your vSphere HA cluster, you must reconfigure vSphere HA on that host.

In most implementations, NIC teaming provides sufficient heartbeat redundancy, but as an alternative you can create a second management network connection attached to a separate virtual switch. Redundant management networking allows the reliable detection of failures and prevents isolation or partition conditions from occurring, because heartbeats can be sent over multiple networks. The original management network connection is used for network and management purposes. When the second management network connection is created, vSphere HA sends heartbeats over both management network connections. If one path fails, vSphere HA still sends and receives heartbeats over the other path.

NOTE Configure the fewest possible number of hardware segments between the servers in a cluster. The goal being to limit single points of failure. Additionally, routes with too many hops can cause networking packet delays for heartbeats, and increase the possible points of failure.

Using IPv6 Network Configurations

Only one IPv6 address should be assigned to a given network interface used by your vSphere HA cluster. Assigning multiple IP addresses increases the number of heartbeat messages sent by the cluster's master host with no corresponding benefit.

Best Practices for Interoperability

Observe the following best practices for allowing proper interoperability between vSphere HA and other features.

vSphere HA and Storage vMotion Interoperability in a Mixed Cluster

In clusters where ESXi 5.x hosts and ESX/ESXi 4.1 or prior hosts are present and where Storage vMotion is used extensively or Storage DRS is enabled, do not deploy vSphere HA. vSphere HA might respond to a host failure by restarting a virtual machine on a host with an ESXi version different from the one on which the virtual machine was running before the failure. A problem can occur if, at the time of failure, the virtual machine was involved in a Storage vMotion action on an ESXi 5.x host, and vSphere HA restarts the virtual machine on a host with a version prior to ESXi 5.0. While the virtual machine might power on, any subsequent attempts at snapshot operations could corrupt the vdisk state and leave the virtual machine unusable.

Using Auto Deploy with vSphere HA

You can use vSphere HA and Auto Deploy together to improve the availability of your virtual machines. Auto Deploy provisions hosts when they power up and you can also configure it to install the vSphere HA agent on such hosts during the boot process. See the Auto Deploy documentation included in vSphere Installation and Setup for details.

Upgrading Hosts in a Cluster Using Virtual SAN

If you are upgrading the ESXi hosts in your vSphere HA cluster to version 5.5 or higher, and you also plan to use Virtual SAN, follow this process.

1 Upgrade all of the hosts.

2 Disable vSphere HA.

3 Enable Virtual SAN.

4 Re-enable vSphere HA.

Best Practices for Admission Control

Observe the following best practices for the configuration and usage of admission control for vSphere HA.

The following recommendations are best practices for vSphere HA admission control.

- Select the Percentage of Cluster Resources Reserved admission control policy. This policy offers the most flexibility in terms of host and virtual machine sizing. When configuring this policy, choose a percentage for CPU and memory that reflects the number of host failures you want to support. For example, if you want vSphere HA to set aside resources for two host failures and have ten hosts of equal capacity in the cluster, then specify 20% (2/10).
- Ensure that you size all cluster hosts equally. For the Host Failures Cluster Tolerates policy, an unbalanced cluster results in excess capacity being reserved to handle failures because vSphere HA reserves capacity for the largest hosts. For the Percentage of Cluster Resources Policy, an unbalanced cluster requires that you specify larger percentages than would otherwise be necessary to reserve enough capacity for the anticipated number of host failures.
- If you plan to use the Host Failures Cluster Tolerates policy, try to keep virtual machine sizing requirements similar across all configured virtual machines. This policy uses slot sizes to calculate the amount of capacity needed to reserve for each virtual machine. The slot size is based on the largest reserved memory and CPU needed for any virtual machine. When you mix virtual machines of different CPU and memory requirements, the slot size calculation defaults to the largest possible, which limits consolidation.
- If you plan to use the Specify Failover Hosts policy, decide how many host failures to support and then specify this number of hosts as failover hosts. If the cluster is unbalanced, the designated failover hosts should be at least the same size as the non-failover hosts in your cluster. This ensures that there is adequate capacity in case of failure.

Best Practices for Cluster Monitoring

Observe the following best practices for monitoring the status and validity of your vSphere HA cluster.

Setting Alarms to Monitor Cluster Changes

When vSphere HA or Fault Tolerance take action to maintain availability, for example, a virtual machine failover, you can be notified about such changes. Configure alarms in vCenter Server to be triggered when these actions occur, and have alerts, such as emails, sent to a specified set of administrators.

Several default vSphere HA alarms are available.

- Insufficient failover resources (a cluster alarm)
- Cannot find master (a cluster alarm)
- Failover in progress (a cluster alarm)
- Host HA status (a host alarm)
- VM monitoring error (a virtual machine alarm)
- VM monitoring action (a virtual machine alarm)
- Failover failed (a virtual machine alarm)

NOTE The default alarms include the feature name, vSphere HA.

Monitoring Cluster Validity

A valid cluster is one in which the admission control policy has not been violated.

A cluster enabled for vSphere HA becomes invalid when the number of virtual machines powered on exceeds the failover requirements, that is, the current failover capacity is smaller than configured failover capacity. If admission control is disabled, clusters do not become invalid.

In the vSphere Web Client, select **vSphere HA** from the cluster's **Monitor** tab and then select **Configuration Issues**. A list of current vSphere HA issues appears.

DRS behavior is not affected if a cluster is red because of a vSphere HA issue.

Providing Fault Tolerance for Virtual Machines 3

You can utilize vSphere Fault Tolerance for your virtual machines to ensure business continuity with higher levels of availability and data protection than is offered by vSphere HA.

Fault Tolerance is built on the ESXi host platform, and it provides continuous availability by having identical virtual machines run on separate hosts.

To obtain the optimal results from Fault Tolerance you should be familiar with how it works, how to enable it for your cluster and virtual machines, and the best practices for its usage.

Fault Tolerance Protection for Virtual Machines (http://link.brightcove.com/services/player/bcpid2296383276001?bctid=ref:video_fault_tolerance_protection_vms)

This chapter includes the following topics:

How Fault Tolerance Works

You can use vSphere Fault Tolerance (FT) for most mission critical virtual machines. FT provides continuous availability for such a virtual machine by creating and maintaining another VM that is identical and continuously available to replace it in the event of a failover situation.

The protected virtual machine is called the Primary VM. The duplicate virtual machine, the Secondary VM, is created and runs on another host. The Secondary VM's execution is identical to that of the Primary VM and it can take over at any point without interruption, thereby providing fault tolerant protection.

The Primary and Secondary VMs continuously monitor the status of one another to ensure that Fault Tolerance is maintained. A transparent failover occurs if the host running the Primary VM fails, in which case the Secondary VM is immediately activated to replace the Primary VM. A new Secondary VM is started and Fault Tolerance redundancy is reestablished automatically. If the host running the Secondary VM fails, it is also immediately replaced. In either case, users experience no interruption in service and no loss of data.

A fault tolerant virtual machine and its secondary copy are not allowed to run on the same host. This restriction ensures that a host failure cannot result in the loss of both VMs.

NOTE You can also use VM-Host affinity rules to dictate which hosts designated virtual machines can run on. If you use these rules, be aware that for any Primary VM that is affected by such a rule, its associated Secondary VM is also affected by that rule. For more information about affinity rules, see the *vSphere Resource Management* documentation.

Fault Tolerance avoids "split-brain" situations, which can lead to two active copies of a virtual machine after recovery from a failure. Atomic file locking on shared storage is used to coordinate failover so that only one side continues running as the Primary VM and a new Secondary VM is respawned automatically.

vSphere Fault Tolerance can accommodate symmetric multiprocessor (SMP) virtual machines with up to four vCPUs. Earlier versions of vSphere used a different technology for Fault Tolerance (now known as legacy FT), with different requirements and characteristics (including a limitation of single vCPUs for legacy FT VMs). If compatibility with these earlier requirements is necessary, you can instead use legacy FT. However, this involves the setting of an advanced option for each VM. See "Legacy Fault Tolerance," on page 57 for more information.

Fault Tolerance Use Cases

Several typical situations can benefit from the use of vSphere Fault Tolerance.

Fault Tolerance provides a higher level of business continuity than vSphere HA. When a Secondary VM is called upon to replace its Primary VM counterpart, the Secondary VM immediately takes over the Primary VM's role with the entire state of the virtual machine preserved. Applications are already running, and data stored in memory does not need to be re-entered or reloaded. This differs from a failover provided by vSphere HA, which restarts the virtual machines affected by a failure.

This higher level of continuity and the added protection of state information and data informs the scenarios when you might want to deploy Fault Tolerance.

- Applications that need to be available at all times, especially those that have long-lasting client connections that users want to maintain during hardware failure.
- Custom applications that have no other way of doing clustering.
- Cases where high availability might be provided through custom clustering solutions, which are too complicated to configure and maintain.

Another key use case for protecting a virtual machine with Fault Tolerance can be described as On-Demand Fault Tolerance. In this case, a virtual machine is adequately protected with vSphere HA during normal operation. During certain critical periods, you might want to enhance the protection of the virtual machine. For example, you might be executing a quarter-end report which, if interrupted, might delay the availability of mission critical information. With vSphere Fault Tolerance, you can protect this virtual machine prior to running this report and then turn off or suspend Fault Tolerance after the report has been produced. You can use On-Demand Fault Tolerance to protect the virtual machine during a critical time period and return the resources to normal during non-critical operation.

Fault Tolerance Requirements, Limits, and Licensing

Before using vSphere Fault Tolerance (FT), consider the high-level requirements, limits, and licensing that apply to this feature.

Requirements

The following CPU and networking requirements apply to FT.

CPUs that are used in host machines for fault tolerant VMs must be compatible with vSphere vMotion or improved with Enhanced vMotion Compatibility. Also, CPUs that support Hardware MMU virtualization (Intel EPT or AMD RVI) are required. The following CPUs are supported.

- Intel Sandy Bridge or later. Avoton is not supported.
- AMD Bulldozer or later.

Use a 10-Gbit logging network for FT and verify that the network is low latency. A dedicated FT network is highly recommended.

Limits

In a cluster configured to use Fault Tolerance, two limits are enforced independently.

das.maxftvmsperhost	The maximum number of fault tolerant VMs allowed on a host in the cluster. Both Primary VMs and Secondary VMs count toward this limit. The default value is 4.
das.maxftvcpusperhost	The maximum number of vCPUs aggregated across all fault tolerant VMs on a host. vCPUs from both Primary VMs and Secondary VMs count toward this limit. The default value is 8.

Licensing

The number of vCPUs supported by a single fault tolerant VM is limited by the level of licensing that you have purchased for vSphere. Fault Tolerance is supported as follows:

- vSphere Standard and Enterprise. Allows up to 2 vCPUs
- vSphere Enterprise Plus. Allows up to 4 vCPUs

NOTE FT and legacy FT are not supported in vSphere Essentials and vSphere Essentials Plus.

Fault Tolerance Interoperability

vSphere Fault Tolerance faces some limitations concerning the vSphere features, devices, and other features it can interoperate with.

Before configuring vSphere Fault Tolerance, you should be aware of the features and products Fault Tolerance cannot interoperate with.

vSphere Features Not Supported with Fault Tolerance

When configuring your cluster, you should be aware that not all vSphere features can interoperate with Fault Tolerance.

The following vSphere features are not supported for fault tolerant virtual machines.

- Snapshots. Snapshots must be removed or committed before Fault Tolerance can be enabled on a virtual machine. In addition, it is not possible to take snapshots of virtual machines on which Fault Tolerance is enabled.

  NOTE Disk-only snapshots created for vStorage APIs - Data Protection (VADP) backups are supported with Fault Tolerance. However, legacy FT does not support VADP.

- Storage vMotion. You cannot invoke Storage vMotion for virtual machines with Fault Tolerance turned on. To migrate the storage, you should temporarily turn off Fault Tolerance, and perform the storage vMotion action. When this is complete, you can turn Fault Tolerance back on.

- Linked clones. You cannot use Fault Tolerance on a virtual machine that is a linked clone, nor can you create a linked clone from an FT-enabled virtual machine.
- VM Component Protection (VMCP). If your cluster has VMCP enabled, overrides are created for fault tolerant virtual machines that turn this feature off.
- Virtual Volume datastores.
- Storage-based policy management.
- I/O filters.

Features and Devices Incompatible with Fault Tolerance

Not all third party devices, features, or products can interoperate with Fault Tolerance.

For a virtual machine to be compatible with Fault Tolerance, the Virtual Machine must not use the following features or devices.

Table 3-1. Features and Devices Incompatible with Fault Tolerance and Corrective Actions

Incompatible Feature or Device	Corrective Action
Physical Raw Disk mapping (RDM).	With legacy FT you can reconfigure virtual machines with physical RDM-backed virtual devices to use virtual RDMs instead.
CD-ROM or floppy virtual devices backed by a physical or remote device.	Remove the CD-ROM or floppy virtual device or reconfigure the backing with an ISO installed on shared storage.
USB and sound devices.	Remove these devices from the virtual machine.
N_Port ID Virtualization (NPIV).	Disable the NPIV configuration of the virtual machine.
NIC passthrough.	This feature is not supported by Fault Tolerance so it must be turned off.
Hot-plugging devices.	The hot plug feature is automatically disabled for fault tolerant virtual machines. To hot plug devices (either adding or removing), you must momentarily turn off Fault Tolerance, perform the hot plug, and then turn on Fault Tolerance. **NOTE** When using Fault Tolerance, changing the settings of a virtual network card while a virtual machine is running is a hot-plug operation, since it requires "unplugging" the network card and then "plugging" it in again. For example, with a virtual network card for a running virtual machine, if you change the network that the virtual NIC is connected to, FT must be turned off first.
Serial or parallel ports	Remove these devices from the virtual machine.
Video devices that have 3D enabled.	Fault Tolerance does not support video devices that have 3D enabled.
Virtual EFI firmware	Ensure that the virtual machine is configured to use BIOS firmware before installing the guest operating system.
Virtual Machine Communication Interface (VMCI)	Not supported by Fault Tolerance.
2TB+ VMDK	Fault Tolerance is not supported with a 2TB+ VMDK.

Using Fault Tolerance with DRS

You can use vSphere Fault Tolerance with vSphere Distributed Resource Scheduler (DRS) only when the Enhanced vMotion Compatibility (EVC) feature is enabled. This process allows fault tolerant virtual machines to benefit from better initial placement.

When a cluster has EVC enabled, DRS makes the initial placement recommendations for fault tolerant virtual machines and allows you to assign a DRS automation level to Primary VMs (the Secondary VM always assumes the same setting as its associated Primary VM.)

When vSphere Fault Tolerance is used for virtual machines in a cluster that has EVC disabled, the fault tolerant virtual machines are given DRS automation levels of "disabled". In such a cluster, each Primary VM is powered on only on its registered host and its Secondary VM is automatically placed.

If you use affinity rules with a pair of fault tolerant virtual machines, a VM-VM affinity rule applies to the Primary VM only, while a VM-Host affinity rule applies to both the Primary VM and its Secondary VM. If a VM-VM affinity rule is set for a Primary VM, DRS attempts to correct any violations that occur after a failover (that is, after the Primary VM effectively moves to a new host).

Preparing Your Cluster and Hosts for Fault Tolerance

To enable vSphere Fault Tolerance for your cluster, you must meet the feature's prerequisites and you must perform certain configuration steps on your hosts. After those steps are accomplished and your cluster has been created, you can also check that your configuration complies with the requirements for enabling Fault Tolerance.

The tasks you should complete before attempting to enable Fault Tolerance for your cluster include the following:

- Ensure that your cluster, hosts, and virtual machines meet the requirements outlined in the Fault Tolerance checklist.
- Configure networking for each host.
- Create the vSphere HA cluster, add hosts, and check compliance.

After your cluster and hosts are prepared for Fault Tolerance, you are ready to turn on Fault Tolerance for your virtual machines. See "Turn On Fault Tolerance," on page 53.

Fault Tolerance Checklist

The following checklist contains cluster, host, and virtual machine requirements that you need to be aware of before using vSphere Fault Tolerance.

Review this list before setting up Fault Tolerance.

NOTE The failover of fault tolerant virtual machines is independent of vCenter Server, but you must use vCenter Server to set up your Fault Tolerance clusters.

Cluster Requirements for Fault Tolerance

You must meet the following cluster requirements before you use Fault Tolerance.

- Fault Tolerance logging and VMotion networking configured. See "Configure Networking for Host Machines," on page 50.
- vSphere HA cluster created and enabled. See "Creating and Configuring a vSphere HA Cluster," on page 32. vSphere HA must be enabled before you can power on fault tolerant virtual machines or add a host to a cluster that already supports fault tolerant virtual machines.

Host Requirements for Fault Tolerance

You must meet the following host requirements before you use Fault Tolerance.

- Hosts must use supported processors.
- Hosts must be licensed for Fault Tolerance.
- Hosts must be certified for Fault Tolerance. See http://www.vmware.com/resources/compatibility/search.php and select **Search by Fault Tolerant Compatible Sets** to determine if your hosts are certified.
- The configuration for each host must have Hardware Virtualization (HV) enabled in the BIOS.

NOTE VMware recommends that the hosts you use to support FT VMs have their BIOS power management settings turned to "Maximum performance" or "OS-managed performance".

To confirm the compatibility of the hosts in the cluster to support Fault Tolerance, you can also run profile compliance checks as described in "Create Cluster and Check Compliance," on page 51.

Virtual Machine Requirements for Fault Tolerance

You must meet the following virtual machine requirements before you use Fault Tolerance.

- No unsupported devices attached to the virtual machine. See "Fault Tolerance Interoperability," on page 47.
- Incompatible features must not be running with the fault tolerant virtual machines. See "Fault Tolerance Interoperability," on page 47.
- Virtual machine files must be stored on shared storage. Acceptable shared storage solutions include Fibre Channel, (hardware and software) iSCSI, NFS, and NAS.

Other Configuration Recommendations

You should also observe the following guidelines when configuring Fault Tolerance.

- If you are using NFS to access shared storage, use dedicated NAS hardware with at least a 1Gbit NIC to obtain the network performance required for Fault Tolerance to work properly.
- The memory reservation of a fault tolerant virtual machine is set to the VM's memory size when Fault Tolerance is turned on. Ensure that a resource pool containing fault tolerant VMs has memory resources above the memory size of the virtual machines. Without this excess in the resource pool, there might not be any memory available to use as overhead memory.
- Use a maximum of 16 virtual disks per fault tolerant virtual machine.
- To ensure redundancy and maximum Fault Tolerance protection, you should have a minimum of three hosts in the cluster. In a failover situation, this provides a host that can accommodate the new Secondary VM that is created.

Configure Networking for Host Machines

On each host that you want to add to a vSphere HA cluster, you must configure two different networking switches (vMotion and FT logging) so that the host can support vSphere Fault Tolerance.

To enable Fault Tolerance for a host, you must complete this procedure for each port group option (vMotion and FT logging) to ensure that sufficient bandwidth is available for Fault Tolerance logging. Select one option, finish this procedure, and repeat the procedure a second time, selecting the other port group option.

Prerequisites

Multiple gigabit Network Interface Cards (NICs) are required. For each host supporting Fault Tolerance, a minimum of two physical NICs is recommended. For example, you need one dedicated to Fault Tolerance logging and one dedicated to vMotion. Use three or more NICs to ensure availability.

NOTE The vMotion and FT logging NICs must be on different subnets. If you are using legacy FT, IPv6 is not supported on the FT logging NIC.

Procedure

1 In the vSphere Web Client, browse to the host.

2 Click the **Manage** tab and click **Networking**.

3 Click the **Add host networking** icon.

4 Select **VMkernel Network Adapter** on the Select Connection Type page and click **Next**.

5 Select **New standard switch** and click **Next**.

6 Assign free physical network adapters to the switch and click **Next**.

7 Provide a Network label and enable the services you want and click **Next**.

8 Provide an IP address and subnet mask and click **Finish** after reviewing your settings.

After you create both a vMotion and Fault Tolerance logging virtual switch, you can create other virtual switches, as needed. Add the host to the cluster and complete any steps needed to turn on Fault Tolerance.

What to do next

NOTE If you configure networking to support FT but subsequently suspend the Fault Tolerance logging port, pairs of fault tolerant virtual machines that are powered on remain powered on. If a failover situation occurs, when the Primary VM is replaced by its Secondary VM a new Secondary VM is not started, causing the new Primary VM to run in a Not Protected state.

Create Cluster and Check Compliance

vSphere Fault Tolerance is used in the context of a vSphere HA cluster. After you configure networking on each host, create the vSphere HA cluster and add the hosts to it. You can check to see whether the cluster is configured correctly and complies with the requirements for the enablement of Fault Tolerance.

Procedure

1 In the vSphere Web Client, browse to the cluster.

2 Click the **Monitor** tab and click **Profile Compliance**.

3 Click **Check Compliance Now** to run the compliance tests.

The results of the compliance test appear, and the compliance or noncompliance of each host is shown.

Using Fault Tolerance

After you have taken all of the required steps for enabling vSphere Fault Tolerance for your cluster, you can use the feature by turning it on for individual virtual machines.

Before Fault Tolerance can be turned on, validation checks are performed on a virtual machine.

After these checks are passed and you turn on vSphere Fault Tolerance for a virtual machine, new options are added to the Fault Tolerance section of its context menu. These include turning off or disabling Fault Tolerance, migrating the Secondary VM, testing failover, and testing restart of the Secondary VM.

Validation Checks for Turning On Fault Tolerance

If the option to turn on Fault Tolerance is available, this task still must be validated and can fail if certain requirements are not met.

Several validation checks are performed on a virtual machine before Fault Tolerance can be turned on.

- SSL certificate checking must be enabled in the vCenter Server settings.
- The host must be in a vSphere HA cluster or a mixed vSphere HA and DRS cluster.
- The host must have ESXi 6.x or greater installed (ESX/ESXi 4.x or greater for legacy FT).
- The virtual machine must not have snapshots.
- The virtual machine must not be a template.
- The virtual machine must not have vSphere HA disabled.
- The virtual machine must not have a video device with 3D enabled.

Checks for Powered-On Virtual Machines

Several additional validation checks are performed for powered-on virtual machines (or those that are in the process of being powered on).

- The BIOS of the hosts where the fault tolerant virtual machines reside must have Hardware Virtualization (HV) enabled.
- The host that supports the Primary VM must have a processor that supports Fault Tolerance.
- Your hardware should be certified as compatible with Fault Tolerance. To confirm that it is, use the VMware Compatibility Guide at http://www.vmware.com/resources/compatibility/search.php and select **Search by Fault Tolerant Compatible Sets**.
- The configuration of the virtual machine must be valid for use with Fault Tolerance (for example, it must not contain any unsupported devices).

Secondary VM Placement

When your effort to turn on Fault Tolerance for a virtual machine passes the validation checks, the Secondary VM is created. The placement and immediate status of the Secondary VM depends upon whether the Primary VM was powered-on or powered-off when you turned on Fault Tolerance.

If the Primary VM is powered on:

- The entire state of the Primary VM is copied and the Secondary VM is created, placed on a separate compatible host, and powered on if it passes admission control.
- The Fault Tolerance Status displayed for the virtual machine is **Protected**.

If the Primary VM is powered off:

- The Secondary VM is immediately created and registered to a host in the cluster (it might be re-registered to a more appropriate host when it is powered on.)
- The Secondary VM is not powered on until after the Primary VM is powered on.
- The Fault Tolerance Status displayed for the virtual machine is **Not Protected, VM not Running**.
- When you attempt to power on the Primary VM after Fault Tolerance has been turned on, the additional validation checks listed above are performed.

  After these checks are passed, the Primary and Secondary VMs are powered on and placed on separate, compatible hosts. The virtual machine's Fault Tolerance Status is tagged as **Protected**.

Turn On Fault Tolerance

You can turn on vSphere Fault Tolerance through the vSphere Web Client.

When Fault Tolerance is turned on, vCenter Server resets the virtual machine's memory limit and sets the memory reservation to the memory size of the virtual machine. While Fault Tolerance remains turned on, you cannot change the memory reservation, size, limit, number of vCPUs, or shares. You also cannot add or remove disks for the VM. When Fault Tolerance is turned off, any parameters that were changed are not reverted to their original values.

Connect vSphere Web Client to vCenter Server using an account with cluster administrator permissions.

Prerequisites

The option to turn on Fault Tolerance is unavailable (dimmed) if any of these conditions apply:

- The virtual machine resides on a host that does not have a license for the feature.
- The virtual machine resides on a host that is in maintenance mode or standby mode.
- The virtual machine is disconnected or orphaned (its .vmx file cannot be accessed).
- The user does not have permission to turn the feature on.

Procedure

1. In the vSphere Web Client, browse to the virtual machine for which you want to turn on Fault Tolerance.
2. Right-click the virtual machine and select **Fault Tolerance > Turn On Fault Tolerance**.
3. Click **Yes**.
4. Select a datastore on which to place the Secondary VM configuration files. Then click **Next**.
5. Select a host on which to place the Secondary VM. Then click **Next**.
6. Review your selections and then click **Finish**.

The specified virtual machine is designated as a Primary VM, and a Secondary VM is established on another host. The Primary VM is now fault tolerant.

Turn Off Fault Tolerance

Turning off vSphere Fault Tolerance deletes the secondary virtual machine, its configuration, and all history.

Use the **Turn Off Fault Tolerance** option if you do not plan to reenable the feature. Otherwise, use the **Suspend Fault Tolerance** option.

NOTE If the Secondary VM resides on a host that is in maintenance mode, disconnected, or not responding, you cannot use the **Turn Off Fault Tolerance** option. In this case, you should suspend and resume Fault Tolerance instead.

Procedure

1. In the vSphere Web Client, browse to the virtual machine for which you want to turn off Fault Tolerance.
2. Right-click the virtual machine and select **Fault Tolerance > Turn Off Fault Tolerance**.
3. Click **Yes**.

Fault Tolerance is turned off for the selected virtual machine. The history and the secondary virtual machine for the selected virtual machine are deleted.

Suspend Fault Tolerance

Suspending vSphere Fault Tolerance for a virtual machine suspends its Fault Tolerance protection, but preserves the Secondary VM, its configuration, and all history. Use this option to resume Fault Tolerance protection in the future.

Procedure

1 In the vSphere Web Client, browse to the virtual machine for which you want to suspend Fault Tolerance.

2 Right-click the virtual machine and select **Fault Tolerance > Suspend Fault Tolerance**.

3 Click **Yes**.

Fault Tolerance is suspended for the selected virtual machine. Any history and the Secondary VM for the selected virtual machine are preserved and will be used if the feature is resumed.

What to do next

After you suspend Fault Tolerance, to resume the feature select **Resume Fault Tolerance**.

Migrate Secondary

After vSphere Fault Tolerance is turned on for a Primary VM, you can migrate its associated Secondary VM.

Procedure

1 In the vSphere Web Client, browse to the Primary VM for which you want to migrate its Secondary VM.

2 Right-click the virtual machine and select **Fault Tolerance > Migrate Secondary**.

3 Complete the options in the Migrate dialog box and confirm the changes that you made.

4 Click **Finish** to apply the changes.

The Secondary VM associated with the selected fault tolerant virtual machine is migrated to the specified host.

Test Failover

You can induce a failover situation for a selected Primary VM to test your Fault Tolerance protection.

This option is unavailable (dimmed) if the virtual machine is powered off.

Procedure

1 In the vSphere Web Client, browse to the Primary VM for which you want to test failover.

2 Right-click the virtual machine and select **Fault Tolerance > Test Failover**.

3 View details about the failover in the Task Console.

This task induces failure of the Primary VM to ensure that the Secondary VM replaces it. A new Secondary VM is also started placing the Primary VM back in a Protected state.

Test Restart Secondary

You can induce the failure of a Secondary VM to test the Fault Tolerance protection provided for a selected Primary VM.

This option is unavailable (dimmed) if the virtual machine is powered off.

Procedure

1 In the vSphere Web Client, browse to the Primary VM for which you want to conduct the test.

2 Right-click the virtual machine and select **Fault Tolerance > Test Restart Secondary**.

3 View details about the test in the Task Console.

This task results in the termination of the Secondary VM that provided Fault Tolerance protection for the selected Primary VM. A new Secondary VM is started, placing the Primary VM back in a Protected state.

Upgrade Hosts Used for Fault Tolerance

Use the following procedure to upgrade hosts used for Fault Tolerance.

Prerequisites

Verify that you have cluster administrator privileges.

Verify that you have sets of four or more ESXi hosts that are hosting fault tolerant virtual machines that are powered on. If the virtual machines are powered off, the Primary and Secondary VMs can be relocated to hosts with different builds.

NOTE This upgrade procedure is for a minimum four-node cluster. The same instructions can be followed for a smaller cluster, though the unprotected interval will be slightly longer.

Procedure

1 Using vMotion, migrate the fault tolerant virtual machines off of two hosts.

2 Upgrade the two evacuated hosts to the same ESXi build.

3 Suspend Fault Tolerance on the Primary VM.

4 Using vMotion, move the Primary VM for which Fault Tolerance has been suspended to one of the upgraded hosts.

5 Resume Fault Tolerance on the Primary VM that was moved.

6 Repeat Step 1 to Step 5 for as many fault tolerant virtual machine pairs as can be accommodated on the upgraded hosts.

7 Using vMotion, redistribute the fault tolerant virtual machines.

All ESXi hosts in a cluster are upgraded.

Best Practices for Fault Tolerance

To ensure optimal Fault Tolerance results, you should follow certain best practices.

The following recommendations for host and networking configuration can help improve the stability and performance of your cluster.

Host Configuration

Hosts running the Primary and Secondary VMs should operate at approximately the same processor frequencies, otherwise the Secondary VM might be restarted more frequently. Platform power management features that do not adjust based on workload (for example, power capping and enforced low frequency modes to save power) can cause processor frequencies to vary greatly. If Secondary VMs are being restarted on a regular basis, disable all power management modes on the hosts running fault tolerant virtual machines or ensure that all hosts are running in the same power management modes.

Host Networking Configuration

The following guidelines allow you to configure your host's networking to support Fault Tolerance with different combinations of traffic types (for example, NFS) and numbers of physical NICs.

- Distribute each NIC team over two physical switches ensuring L2 domain continuity for each VLAN between the two physical switches.
- Use deterministic teaming policies to ensure particular traffic types have an affinity to a particular NIC (active/standby) or set of NICs (for example, originating virtual port-id).
- Where active/standby policies are used, pair traffic types to minimize impact in a failover situation where both traffic types will share a vmnic.
- Where active/standby policies are used, configure all the active adapters for a particular traffic type (for example, FT Logging) to the same physical switch. This minimizes the number of network hops and lessens the possibility of oversubscribing the switch to switch links.

NOTE FT logging traffic between Primary and Secondary VMs is unencrypted and contains guest network and storage I/O data, as well as the memory contents of the guest operating system. This traffic can include sensitive data such as passwords in plaintext. To avoid such data being divulged, ensure that this network is secured, especially to avoid 'man-in-the-middle' attacks. For example, you could use a private network for FT logging traffic.

Homogeneous Clusters

vSphere Fault Tolerance can function in clusters with nonuniform hosts, but it works best in clusters with compatible nodes. When constructing your cluster, all hosts should have the following configuration:

- Common access to datastores used by the virtual machines.
- The same virtual machine network configuration.
- The same BIOS settings (power management and hyperthreading) for all hosts.

Run **Check Compliance** to identify incompatibilities and to correct them.

Performance

To increase the bandwidth available for the logging traffic between Primary and Secondary VMs use a 10Gbit NIC, and enable the use of jumbo frames.

Store ISOs on Shared Storage for Continuous Access

Store ISOs that are accessed by virtual machines with Fault Tolerance enabled on shared storage that is accessible to both instances of the fault tolerant virtual machine. If you use this configuration, the CD-ROM in the virtual machine continues operating normally, even when a failover occurs.

For virtual machines with Fault Tolerance enabled, you might use ISO images that are accessible only to the Primary VM. In such a case, the Primary VM can access the ISO, but if a failover occurs, the CD-ROM reports errors as if there is no media. This situation might be acceptable if the CD-ROM is being used for a temporary, noncritical operation such as a patch.

Avoid Network Partitions

A network partition occurs when a vSphere HA cluster has a management network failure that isolates some of the hosts from vCenter Server and from one another. See "Network Partitions," on page 21. When a partition occurs, Fault Tolerance protection might be degraded.

In a partitioned vSphere HA cluster using Fault Tolerance, the Primary VM (or its Secondary VM) could end up in a partition managed by a master host that is not responsible for the virtual machine. When a failover is needed, a Secondary VM is restarted only if the Primary VM was in a partition managed by the master host responsible for it.

To ensure that your management network is less likely to have a failure that leads to a network partition, follow the recommendations in "Best Practices for Networking," on page 40.

Using Virtual SAN Datastores

vSphere Fault Tolerance can use Virtual SAN datastores, but you must observe the following restrictions:

- A mix of Virtual SAN and other types of datastores is not supported for both Primary VMs and Secondary VMs.
- Virtual SAN metro clusters are not supported with FT.

To increase performance and reliability when using FT with Virtual SAN, the following conditions are also recommended.

- Virtual SAN and FT should use separate networks.
- Keep Primary and Secondary VMs in separate Virtual SAN fault domains.

Legacy Fault Tolerance

By default, vSphere Fault Tolerance (FT) can accommodate symmetric multiprocessor (SMP) virtual machines with up to four vCPUs. If your virtual machine has only a single vCPU, however, you can use legacy FT instead for backward compatibility. Unless technically necessary, use of legacy FT is not recommended.

To use legacy Fault Tolerance, you must configure an advanced option for the virtual machine. After you complete this configuration, the legacy FT VM is different in some ways from other fault tolerant VMs.

Differences for VMs That Use Legacy FT

VMs that use FT and VMs that use legacy FT differ in several ways.

Table 3-2. Differences Between Legacy FT and FT

	Legacy FT	FT
Extended Page Tables/Rapid Virtualization Indexing (EPT/RVI)	Not supported	Required
IPv6	Not supported for legacy FT logging NICs.	Supported for FT-logging NICs.
DRS	Fully supported for initial placement, load balancing, and maintenance mode support.	Only power on placement of Secondary VM and maintenance mode are supported.

Table 3-2. Differences Between Legacy FT and FT (Continued)

	Legacy FT	FT
vStorage APIs - Data Protection backups	Not supported	Supported
Eager-zeroed thick .vmdk disk files	Required	Not required because FT supports all disk file types, including thick and thin
.vmdk redundancy	Only a single copy	Primary VMs and Secondary VMs always maintain independent copies, which can be placed on different datastores to increase redundancy.
NIC bandwidth	Dedicated 1-Gb NIC recommended	Dedicated 10-Gb NIC recommended
CPU and host compatibility	Requires identical CPU model and family and nearly identical versions of vSphere on hosts.	CPUs must be compatible with vSphere vMotion or EVC. Versions of vSphere on hosts must be compatible with vSphere vMotion.
Turn on FT on running VM	Not always supported. You might need to power off VM first.	Supported
Storage vMotion	Supported only on powered-off VMs. vCenter Server automatically turns off FT before performing a Storage vMotion action and then turns on FT again after the Storage vMotion action completes.	Not supported. User must turn off FT for the VM before performing the Storage vMotion action and then turn on FT again.
vlance networking drivers	Not supported	Supported

Additional Requirements for Legacy FT

In addition to the differences listed for legacy FT, it also has the following unique requirements.

- Your cluster must contain at least two FT-certified hosts that run the same Fault Tolerance version or host build number. The Fault Tolerance version number appears on a host's **Summary** tab in the vSphere Web Client.
- ESXi hosts must have access to the same virtual machine datastores and networks.
- Virtual machines must be stored in virtual RDM or virtual machine disk (VMDK) files that are thick provisioned. If a virtual machine is stored in a VMDK file that is thin provisioned and an attempt is made to use Fault Tolerance, a message indicates that the VMDK file must be converted. To perform the conversion, you must power off the virtual machine.
- Hosts must have processors from the FT-compatible processor group. Verify that the hosts' processors are compatible with one another.
- The host that supports the Secondary VM must have a processor that supports Fault Tolerance and is the same CPU family or model as the host that supports the Primary VM.
- When you upgrade hosts that contain fault tolerant VMs, verify that the Primary and Secondary VMs continue to run on hosts with the same FT version number or host build number (for hosts before ESX/ESXi 4.1).

  NOTE If you designated a VM to use legacy FT before you upgraded the hosts in the cluster, that VM continues to use legacy FT after the host upgrade.

Enable Legacy Fault Tolerance

To use legacy Fault Tolerance, you must configure an advanced option for the virtual machine.

Legacy FT can be used only with single vCPU virtual machines that are not already using FT. To enable legacy FT for each VM that is to use it, you must set the `vm.uselegacyft` advanced option to a value of **true**.

Procedure

1 In the vSphere Web Client, browse to the virtual machine.

2 Right-click the virtual machine and select **Edit Settings**.

3 Click the **VM Options** tab.

4 Open the **Advanced** section and next to **Configuration Parameters**, click **Edit Configuration**.

5 Click **Add Row** and enter `vm.uselegacyft` for Name and **true** for Value.

6 Click **OK**.

Legacy FT is now enabled for that virtual machine.

Index