WHITE PAPER – JANUARY 2020

# VMware vSphere Cluster Resiliency and High Availability

## Amazon RDS on VMware

vmware®

# Table of contents

### VMWARE VSPHERE VERSIONS SUPPORTED FOR RDS ON VMWARE

- *6.5 GA, U1, U2, U3*
- *6.7 GA, U1, U2*
- *Certified on NSX (not required)*
- *Certified on vSAN (not required)*
- *vCloud Foundation*

### RDS ON VMWARE HARDWARE REQUIREMENTS

- *Recommended a vSphere cluster dedicated to RDS on VMware*
- *Hardware must be supported by VMware, according to VMware's hardware compatibility list.*
- *For Production, Cluster size should meet redundancy, with minimum 3 ESXi hosts, >=1 Gbps Ethernet connectivity between ESXi hosts in same cluster*

## Introduction

Amazon RDS on VMware is a service that makes it easy for customers to set up, operate, and scale databases in VMware-based software-defined data centers and hybrid environments. Amazon RDS on VMware automates database provisioning, operating system and database patching, backup, point-in-time restore, storage and compute scaling, instance health monitoring, and failover.

This solution can also be used to enable low-cost, high-availability hybrid deployments, database disaster recovery to AWS, read replica bursting to Amazon RDS, and long-term database archival in Amazon Simple Storage Service (Amazon S3).

Amazon RDS on VMware can take advantage of durability and resiliency features that vSphere 6.5 onwards provide transparently to the solution. The service has been certified to leverage most of the resiliency, durability and high availability features available on vSphere 6.5 and 6.7.

We don't recommend deploying Amazon RDS on VMware atop of older vSphere editions.

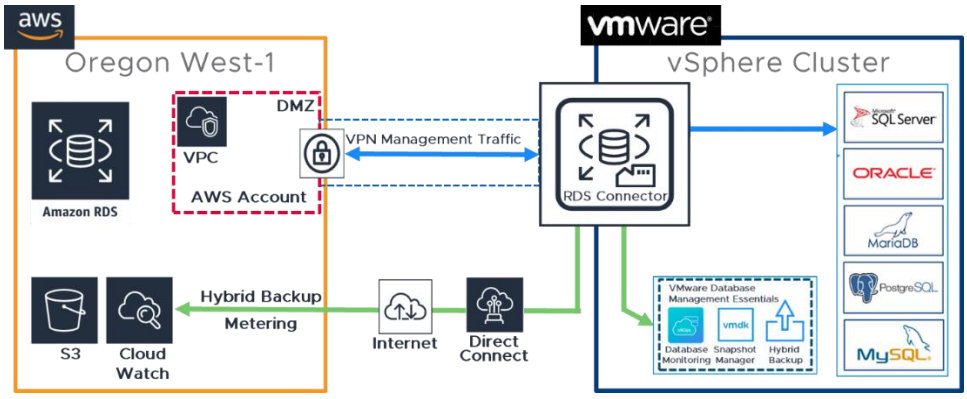
Amazon RDS on VMware supports Microsoft SQL Server, Oracle, PostgreSQL, MySQL, and MariaDB databases.

## Concepts

Each Amazon RDS on VMware Customer Availability Zone (AZ) is mapped to a single vSphere Cluster. A vSphere Cluster can only be mapped to a single Amazon RDS on VMware Customer Availability Zone.

A vSphere Cluster must be onboarded to become a Customer Availability Zone. Once onboarded, Amazon RDS on VMware deploys a few control plane proxy VMs into the corresponding vSphere Cluster.

In addition, such vSphere Cluster is also the target where the managed Amazon RDS on VMware Database instances will be deployed.



*Downtime Cost, whether planned or unplanned, brings considerable costs. However, solutions that 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, you can increase the baseline level of availability provided for all applications and provide higher levels of availability more easily and cost effectively*

# Downtime

## Planned

Planned downtime typically accounts for over 80% of private clouds 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.

Amazon RDS on VMware can take advantage of specific vSphere features to have its workloads (Control Plane and native instances) dynamically moved to different physical servers without downtime or service interruption. Server maintenance can be performed without requiring application and service downtime, helping on:

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

To achieve transparent migration of workloads to other members of the cluster, Amazon RDS on VMware can leverage **vSphere vMotion**. Administrators can perform faster and completely transparent maintenance operations, moving Amazon RDS on VMware workloads to different physical servers without being forced to schedule inconvenient maintenance windows.

## Unplanned

While vSphere ESXi host provides a robust platform for running applications, customers must also protect themselves from unplanned downtime caused from hardware or application failures. Amazon RDS on VMware can take advantage from such vSphere capabilities to prevent unplanned downtime.

These vSphere capabilities are part of the virtual infrastructure and are transparent to the vSphere certified guest operating systems and applications running in virtual machines.

These features can be configured and utilized by all the Amazon RDS on VMware virtual machines on a physical system, reducing the cost and complexity of providing higher availability.

# Key vSphere built-in availability capabilities

## Shared Storage

Amazon RDS on VMware requires to be deployed on a vSphere Cluster, and all the vSphere ESXi hosts part of that cluster must be connected to the same shared storage to eliminate single points of failure. All the Amazon RDS on VMware instances (Local Control Plane or native VMware database instances) should store its virtual machine files on shared storage.

The use of SAN/vSAN mirroring, along with the Amazon RDS on VMware native data replication features, can be used to keep updated data of each of the database instances locally ready for a quick recovery or cloning on demand.

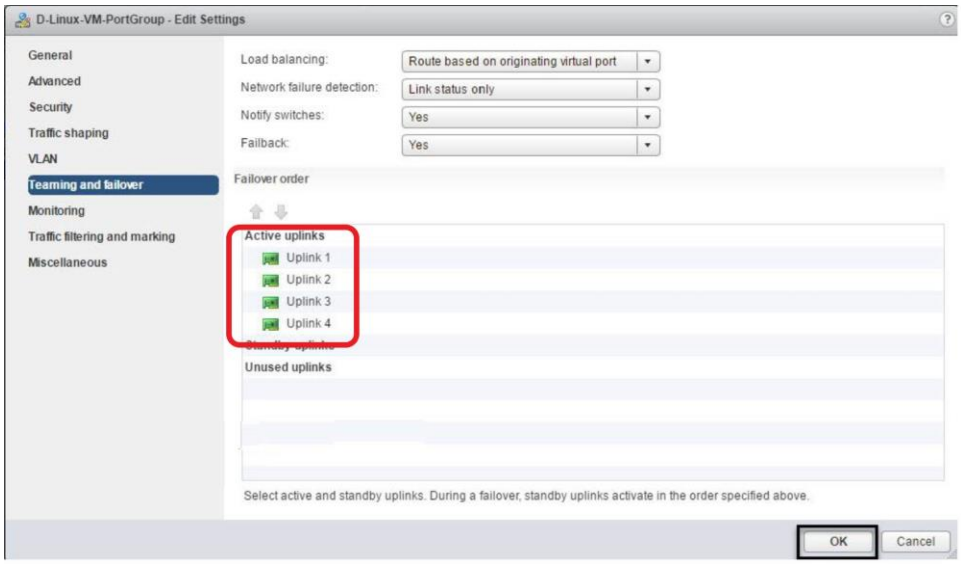
**Note** *On this version of Amazon RDS on VMware, resiliency and durability of the data needs to be achieved by leveraging Storage Mirroring, although very soon, Amazon RDS on VMware will provide a durability solution through Hybrid Backup up to S3, being able to be deployed on cheaper storage locally*

---

**Note** *It is recommended for Amazon RDS on VMware to include NIC teaming on the virtual switch associated to the Custom Control Network across separate physical switches.*

## Network interface teaming

Provide tolerance of individual network card failures. This provides also a path to Virtual Networking, which is specially designed to prevent network routing failures. You can use NIC teaming to increase the network bandwidth available in a network path.




## Storage multipathing

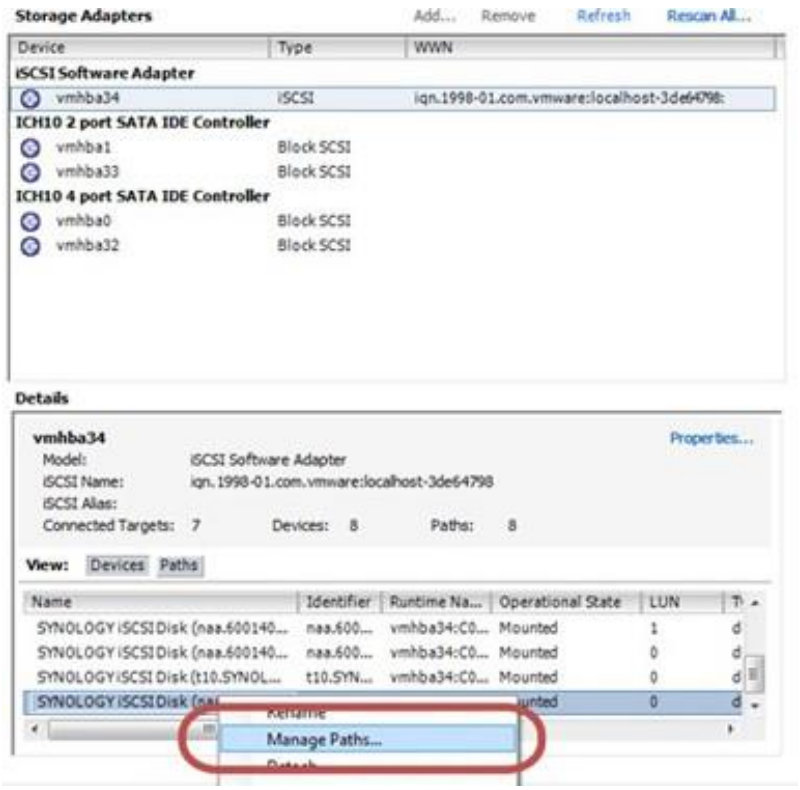
vSphere can tolerate storage path failures. To maintain a constant connection between a host and its storage, ESXi supports multipathing. Multipathing is a technique that lets you use more than one physical path that transfers data between the host and an external storage device.

ESXi provides an extensible multipathing module called the Native Multipathing Plug-In (NMP). VMware NMP supports all storage arrays listed on the VMware storage HCL and provides a default path selection algorithm based on the array type

**Note** It is recommended than Amazon RDS on VMware runs on a datastore cluster storage that is listed on the HCL

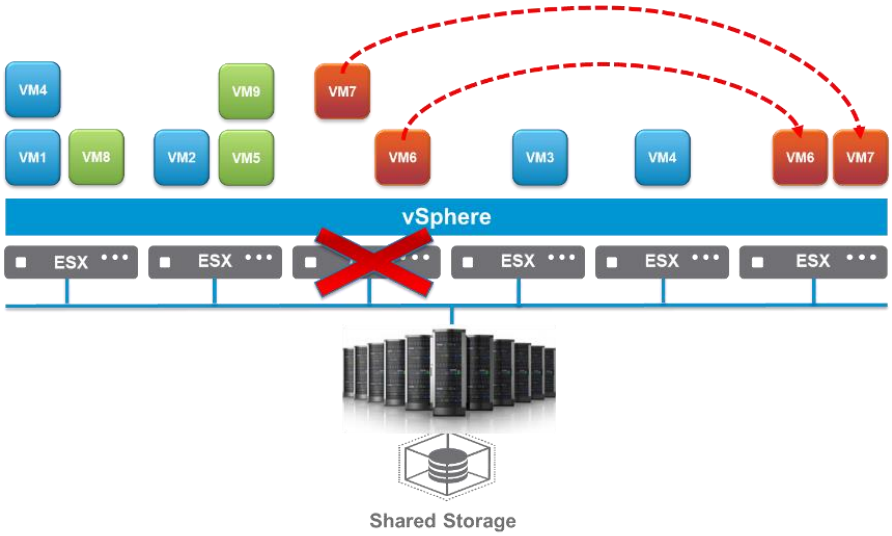
In case of a failure of any element in the SAN network, such as an adapter, switch, or cable, ESXi can switch to another physical path, which does not use the failed component. This process of path switching to avoid failed components is known as path failover.

In addition to path failover, multipathing provides load balancing. Load balancing is the process of distributing I/O loads across multiple physical paths. Load balancing reduces or removes potential bottlenecks.




## VMware vSphere vMotion

Enables the live migration of running virtual machines from one physical server to another with zero downtime. vSphere vMotion is a key enabling technology for creating the dynamic, available, and self-optimizing data center.



HOW VMWARE VSPHERE HA WORKS?

VMware HA provides high availability for virtual machines by pooling them 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.

### VMware vSphere High Availability

vSphere HA provides easy-to-use, cost-effective, high availability for applications running in virtual machines. In the event of physical server failure, affected virtual machines are automatically restarted on other production servers with spare capacity.

### VMware vSphere App HA

VMware vSphere App HA monitors the status of application services running in the guest operating system (OS) and performs remediation if the service is unavailable.

### VMware vSphere Fault Tolerance

vSphere FT provides continuous availability for applications in the event of server failures by creating a live shadow instance of a virtual machine that is in virtual lockstep with the primary instance. By enabling instantaneous failover between the two instances in the event of hardware failure, vSphere FT eliminates even the smallest chance of disruption.

## VMware vSphere High Availability in detail

### Primary and Secondary Hosts in a VMware HA Cluster

When you add a host to a VMware HA cluster, an agent is uploaded to the host and configured to communicate with other agents in the cluster. The first five hosts added to the cluster are designated as primary hosts, and all subsequent hosts are designated as secondary hosts.

The primary hosts maintain and replicate all cluster state and are used to initiate failover actions. If a primary host is removed from the cluster, VMware HA promotes another (secondary) host to primary status.

If a primary host is going to be offline for an extended period, you should remove it from the cluster, so that it can be replaced by a secondary host.

Any host that joins the cluster must communicate with an existing primary host to complete its configuration (except when you are adding the first host to the cluster). At least one primary host must be functional for VMware HA to operate correctly. If all primary hosts are unavailable (not responding), no hosts can be successfully configured for VMware HA.

You should consider this limit of five primary hosts per cluster when planning the scale of your cluster. Also, if your cluster is implemented in a blade server environment, do not place more than four primary hosts in a single blade chassis. If all five of the primary hosts are in the same chassis and that chassis fails, your cluster loses VMware HA protection.

One of the primary hosts is also designated as the active primary host and its responsibilities include:

**Note** *When a host fails, VMware HA does not fail over any virtual machines to a host that is in maintenance mode.*

**Note** *If you ensure that the network infrastructure is sufficiently redundant and that at least one network path is always available, host network isolation should be a rare occurrence*

- Deciding where to restart virtual machines.
- Keeping track of failed restart attempts.
- Determining when it is appropriate to keep trying to restart a virtual machine.

If the active primary host fails, another primary host replaces it.

## Failure Detection and Host Network Isolation

Agents communicate with each other and monitor the liveness of the hosts in the cluster. This communication is done through the exchange of heartbeats, by default, every second. If a 15-second period elapses without the receipt of heartbeats from a host, and the host cannot be pinged, it is declared as failed.

In the event of a host failure, the virtual machines running on that host are failed over, that is, restarted on alternate hosts.

Host network isolation occurs when a host is still running, but it can no longer communicate with other hosts in the cluster. With default settings, if a host stops receiving heartbeats from all other hosts in the cluster for more than 12 seconds, it attempts to ping its isolation addresses. If this also fails, the host declares itself as isolated from the network. An isolation address is pinged only when heartbeats are not received from any other host in the cluster.

When the isolated host's network connection is not restored for 15 seconds or longer, the other hosts in the cluster treat the isolated host as failed and attempt to fail over its virtual machines. However, when an isolated host retains access to the shared storage it also retains the disk lock on virtual machine files. To avoid potential data corruption, VMFS disk locking prevents simultaneous write operations to the virtual machine disk files and attempts to fail over the isolated host's virtual machines fail. By default, the isolated host shuts down its virtual machines, but you can change the host isolation response to Leave powered On or Power off.

## Using VMware HA and DRS Together

Using VMware HA with Distributed Resource Scheduler (DRS) combines automatic failover with load balancing. This combination can result in faster rebalancing of virtual machines after VMware HA has moved virtual machines to different hosts.

When VMware 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. VMware HA uses the virtual machine's CPU and memory reservation to determine if a host has enough spare capacity to accommodate the virtual machine.

In a cluster using DRS and VMware 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

RECOMMENDATION

Amazon RDS on VMware should be deployed on a cluster with DRS and HA enabled to achieve better durability and resiliency

the event of a failure. You must manually migrate the virtual machines off of the hosts using vMotion.

In some scenarios, VMware 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 cannot be used by virtual machines for failover.

In such cases, VMware HA will 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. VMware HA does not perform a failover if doing so would violate such a rule.

## vCenter Server High Availability Options

Amazon RDS on VMware requires vCenter to be deployed highly available and on a resilient manner.

VMware vSphere Fault Tolerance (FT) can be utilized to provide continuous availability for vCenter Server by having identical vCenter Server virtual machines running on separate hosts. A transparent failover occurs if the host running the Primary vCenter Server virtual machine fails, in which case the Secondary vCenter Server virtual machine is immediately activated to replace the failed virtual machine.

A new secondary virtual machine is started, and Fault Tolerance redundancy is reestablished automatically. Because support of FT for up to 4 virtual CPUs (vCPU) is available only in vSphere 6.x Enterprise Plus edition, FT can only be used to protect vCenter Server for the tiny and small deployment type, 2 vCPU and 4 vCPU, respectively.

VMware Service Lifecycle Manager  is another feature available on supported vCenter versions for Amazon RDS on VMware. It monitors and protects vCenter Server providing better availability by periodically checking the vCenter Server processes (PID Watchdog) or the vCenter Server API (API Watchdog).

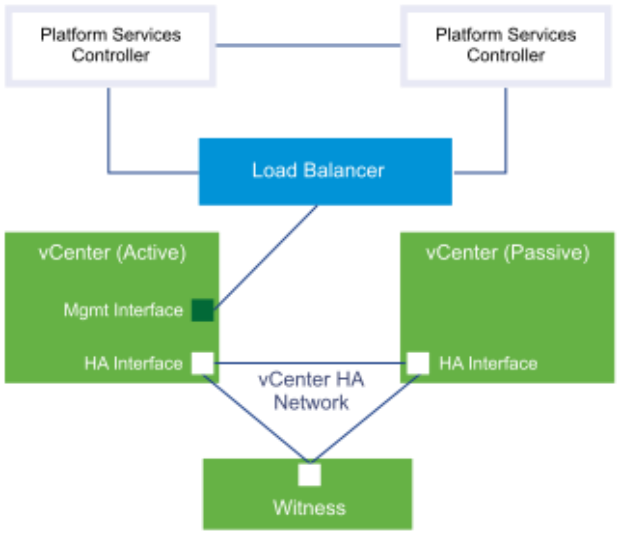
If the VMware Service Lifecycle Manager service detects that APIs are not running or responding, it will attempt to restart the service two times; on the third attempt, depending on vCenter configuration, it will reboot the vCenter Server's host OS.

VMware vSphere vCenter HA provides robust general-purpose protection against hardware and operating system failures for vCenter Server instances running on a virtual machine. Once configured, vSphere HA monitors hosts and virtual machines, and takes the user-configured action with or without vCenter Server availability.

The active-passive architecture of the solution can also help reduce downtime significantly when the vCenter Server Appliance is patched. vCenter HA is only available for the vCenter Server Appliance.

## vCenter HA with an Embedded Platform Services Controller

When you use vCenter HA with an embedded Platform Services Controller, the environment setup is as follows, vCenter HA with an Embedded Platform Services Controller




This setup is as follows:

- The user provisions the vCenter Server Appliance with an embedded Platform Services Controller.
- Cloning of the vCenter Server Appliance to a Passive and a Witness node occurs.
- In a Basic configuration, the configuration creates and configures the clones.
- In an Advanced configuration, the user creates and configures the clones.
- As part of the clone process, Platform Services Controller and all its services are cloned as well.
- When configuration is complete, vCenter HA performs replication to ensure that the Passive node is synchronized with the Active node. The Active node to Passive node replication includes Platform Services Controller data.
- When configuration is complete, the vCenter Server Appliance is protected by vCenter HA. In case of failover, Platform Services Controller and all its services are available on the Passive node.

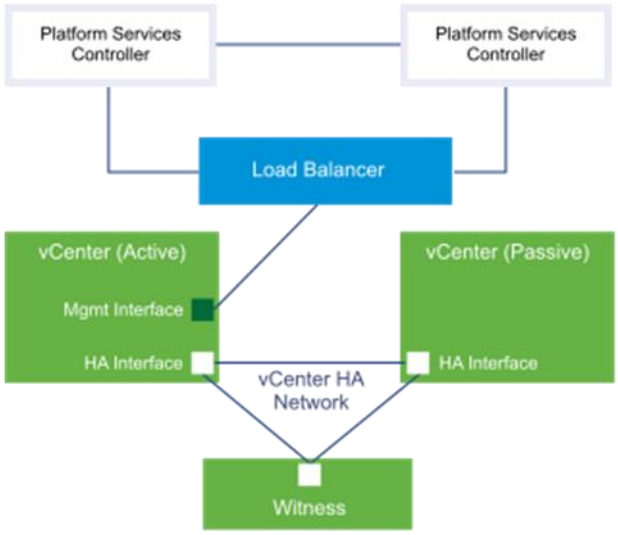
RECOMMENDATION

Amazon RDS on VMware should be deployed on a vSphere cluster where vCenter is deployed in High Availability.

## vCenter HA with an External Platform Services Controller

When you use vCenter HA with an external Platform Services Controller, you must set up an external load balancer to protect the Platform Services Controller. If one Platform Services Controller becomes unavailable, the load balancer directs the vCenter Server Appliance to a different Platform Services Controller.

The environment setup is as follows.




vCenter HA with External Platform Services Controller:

- The user sets up at least two external Platform Services Controller instances. These instances replicate vCenter Single Sign-On information and other Platform Services Controller information, for example, licensing.
- During provisioning of the vCenter Server Appliance, the user selects an external Platform Services Controller.
- The user sets up the vCenter Server Appliance to point to a load balancer that provides high availability for Platform Services Controller.
- The user or the Basic configuration clones the first vCenter Server Appliance to create the Passive node and Witness node.
- As part of the clone process, the information about the external Platform Services Controller and the load balancer is cloned as well.
- When configuration is complete, the vCenter Server Appliance is protected by vCenter HA.
- If the Platform Services Controller instance becomes unavailable, the load balancer redirects requests for authentication or other services to the second Platform Services Controller instance.

RECOMMENDATION

Amazon RDS on VMware would have better performance when storage SAN runs on VMFS compatible block model or under vSAN

## vCenter Server High Availability Options

Storage virtualization refers to a logical abstraction of physical storage resources and capacities from virtual machines and their applications.

Amazon RDS on VMware supports multiple types of shared storages:

| Storage Type | Datastore | Block-level | HA/DRS |
|---|---|---|---|
| Fiber Channel | VMFS | Yes | Yes |
| Fiber Channel Ethernet | VMFS | Yes | Yes |
| iSCSI | VMFS | Yes | Yes |
| NAS over NFS | NFS | No | No |
| vSAN | vSAN | No | Yes |

The datastores that you deploy on block storage devices use the native vSphere Virtual Machine File System (VMFS) format. It is a special high-performance file system format that is optimized for storing virtual machines.

### 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. 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. You can implement network redundancy at the NIC level with NIC teaming, or at the management network level. In most implementations, NIC teaming provides enough redundancy, but you can use or add management network redundancy if required. Redundant management networking allows the reliable detection of failures and prevents isolation conditions from occurring, because heartbeats can be sent over multiple networks. 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.

### Network Redundancy Using 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.

## Network Redundancy Using a Secondary Network

As an alternative to NIC teaming for providing redundancy for heartbeats, you can create a secondary management network connection, which is attached to a separate virtual switch. The primary management network connection is used for network and management purposes. When the secondary management network connection is created, vSphere HA sends heartbeats over both the primary and secondary management network connections. If one path fails, vSphere HA can still send and receive heartbeats over the other path.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com.