



Availability of vCenter Server

Table of contents

Availability of vCenter Server	3
vCenter Availability Options	4
vSphere HA	4
vCenter HA	4
vSphere Fault Tolerance	5
Comparison	6
High Availability Topologies for vCenter	8
vSphere HA	8
vSphere HA: vSAN Stretched Cluster	8
vCenter HA	9
vCenter HA: vSAN Stretched Cluster	11
vSphere Fault Tolerance	11
Summary	13
Useful Links	13

Availability of vCenter Server

When considering the availability of vCenter Server, there are a few options to provide that availability including vSphere HA, vCenter HA, vSphere Fault Tolerance and even vSAN. In this article we're going to explore all these availability features and how they apply to vCenter Server.

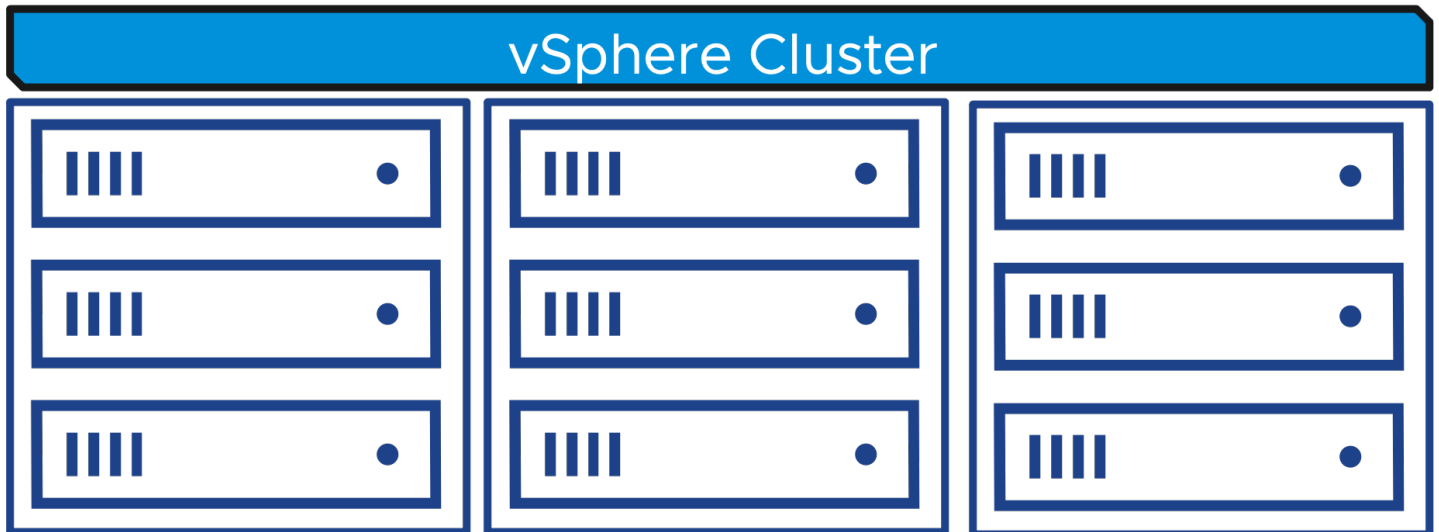
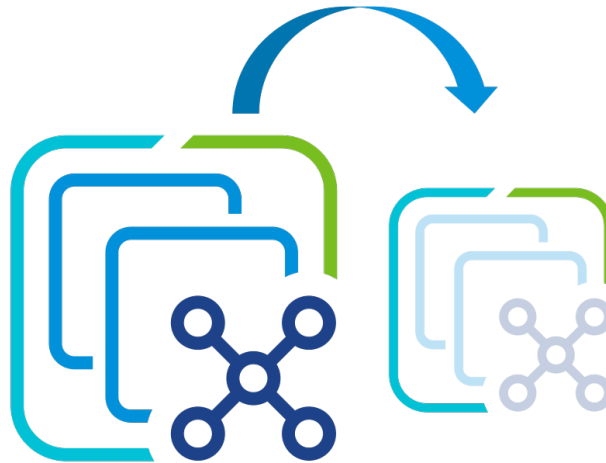
vCenter Availability Options

Let's look at three vSphere features that can provide availability to vCenter Server.

- vSphere HA
- vCenter HA
- vSphere Fault Tolerance

vSphere HA

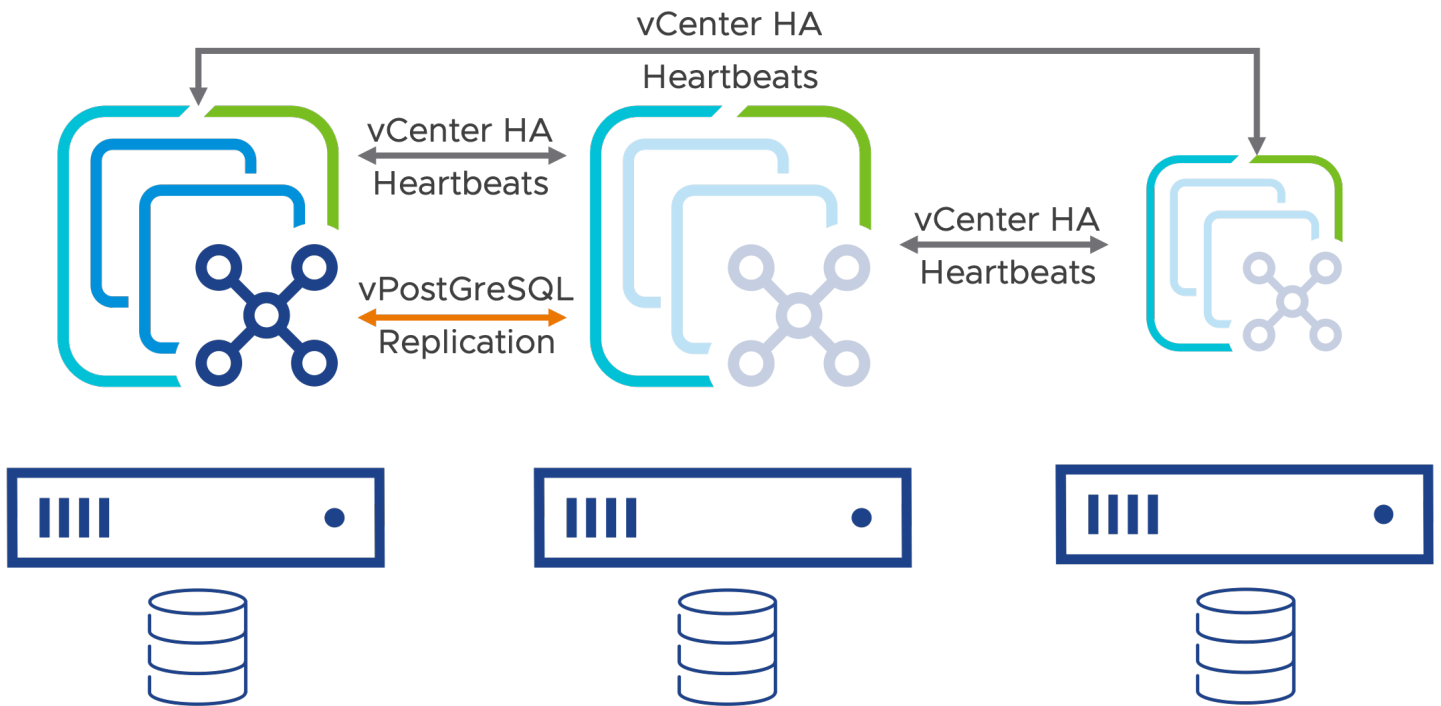
vSphere HA provides a simple, out-of-the-box, high availability solution for all virtual machines in a vSphere cluster, including vCenter Server. vSphere HA restarts the vCenter Server, typically on a new host. There is downtime while the VM and services restart.



vSphere HA feature such as VM and Application monitoring supports VMware Tools heartbeat monitoring for vCenter Server and VM Component Protection (VMCP) helps protect against datastore accessibility issues and can restart vCenter Server on another host if the current host has issues accessing the datastore.

vCenter HA

vCenter HA requires a comparatively more complex setup procedure to activate as well as a second dedicated vCenter HA network to service vPostgreSQL, and file-based, replication and the vCenter HA heartbeat traffic. The vCenter HA passive and witness nodes are full VM clones in terms of compute and storage utilization, however the witness node is reconfigured to reduce the compute footprint to 1 vCPU and 1GB memory. The storage utilization is not reduced. This means, using vCenter HA requires 3x the storage space and more than 2x the compute resources.



During a vCenter HA failover event, the vCenter Server services must start on the new active node. The failover is not instantaneous and there will be some down-time until the services have completely started. vCenter HA heartbeat monitoring is also available via the `vmware-vcba` service to monitor the liveliness of the three vCenter HA nodes from the perspective of the vCenter HA cluster.

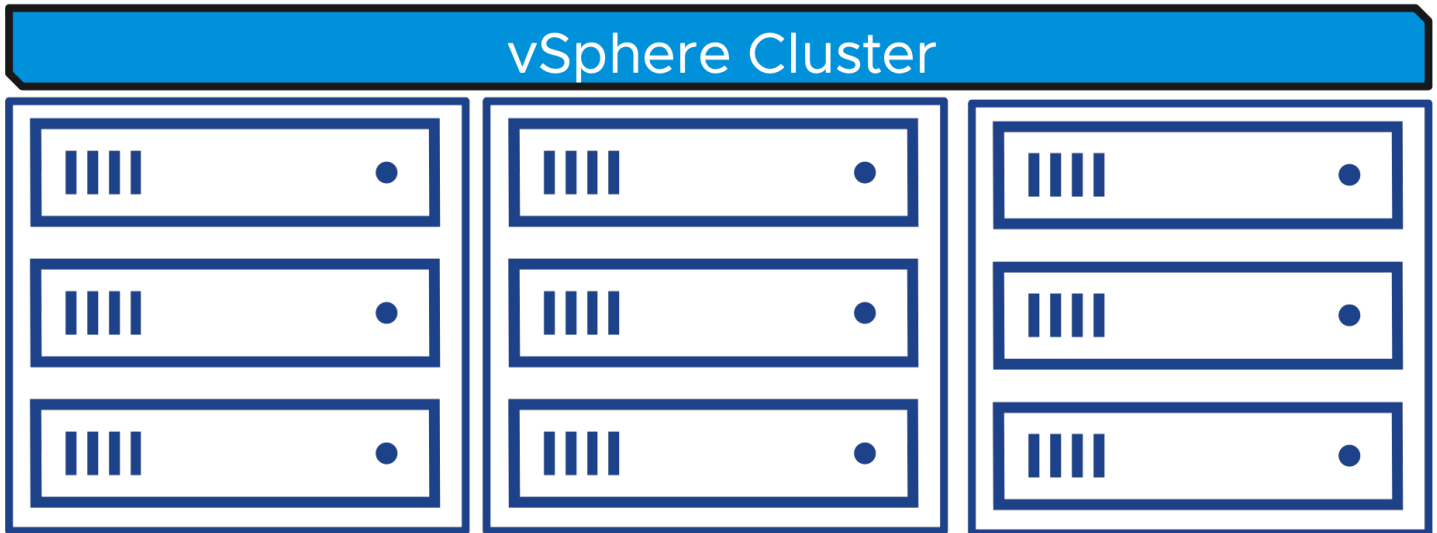
There are several [management and maintenance limitations](#) imposed when using vCenter HA. Such as, patching and updating vCenter Server, SSO domain repointing, certificate management, IP/FQDN changes and using snapshots, are not supported when using vCenter HA. To perform these actions you must deactivate vCenter HA completely, perform the desired maintenance, and reactivate vCenter HA from scratch. File-based backups of a vCenter HA environment are supported. Only the current active node is backed up. Upon restoration, vCenter HA must be reactivated from scratch.

One last consideration when using vCenter HA is that SSH must be enabled when using vCenter HA. This is required to facilitate the use of `rsync` replication between the active and passive nodes.

vSphere Fault Tolerance

vSphere Fault Tolerance is relatively easy to activate for a virtual machine, including vCenter Server, once you have satisfied the specific requirements. vSphere FT requires a 10-Gbit network between ESXi hosts in the cluster, a dedicated 10-Gbit network exclusively for FT is recommended. vSphere FT supports up to 8 vCPUs on a single VM, which means vCenter Server instances of size Large or greater cannot be protected using vSphere FT.

During a failover event, there is no down time. The secondary VM immediately becomes active.



Several vSphere features are not supported on vSphere FT enabled VMs. In the case of a vCenter Server, this would mean snapshots and storage vMotion are not supported and the vCenter Server cannot reside on a virtual volumes datastore.

Comparison

This table shows a simple comparison of the three availability features we discussed. An important aspect to note, is that none of these features are disaster recovery solutions.

vSphere HA	vCenter HA	Fault Tolerance
	Management limitations	Management limitations

This table highlights some of the management and maintenance limitations of each solution as it relates to a vCenter Server. For more details, see the Management limitations links above.

Feature	vSphere HA	vCenter HA	Fault Tolerance
SSH Enabled			
Snapshots			
Hostname/IP Change			
SSO Domain Repointing			
Certificate Management			
Storage vMotion			

* These actions have not been tested, by VMware, on a vCenter Server protected by Fault Tolerance.

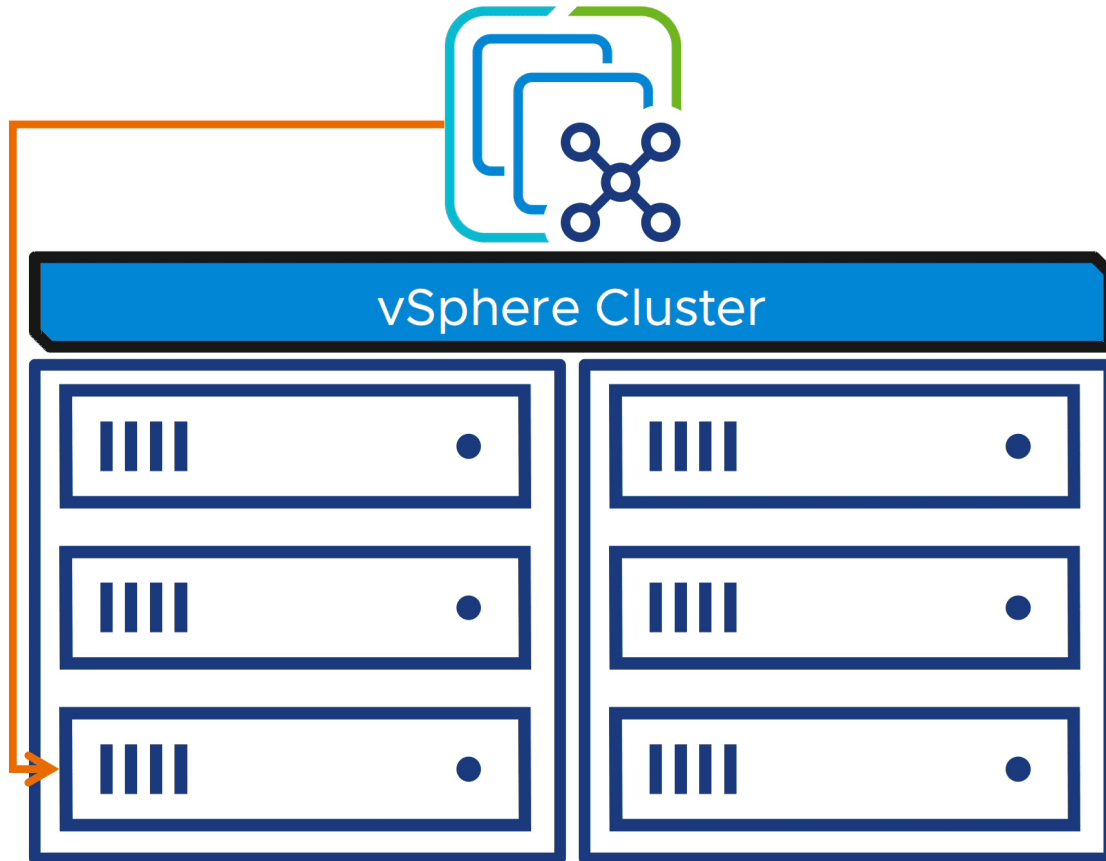
You can weigh up the complexity, requirements and limitations of each option as well as the level of availability each solution provides to make your decision.

High Availability Topologies for vCenter

In this section we will illustrate a few, not all possible, vSphere topologies that can contribute to the availability of vCenter Server.

vSphere HA

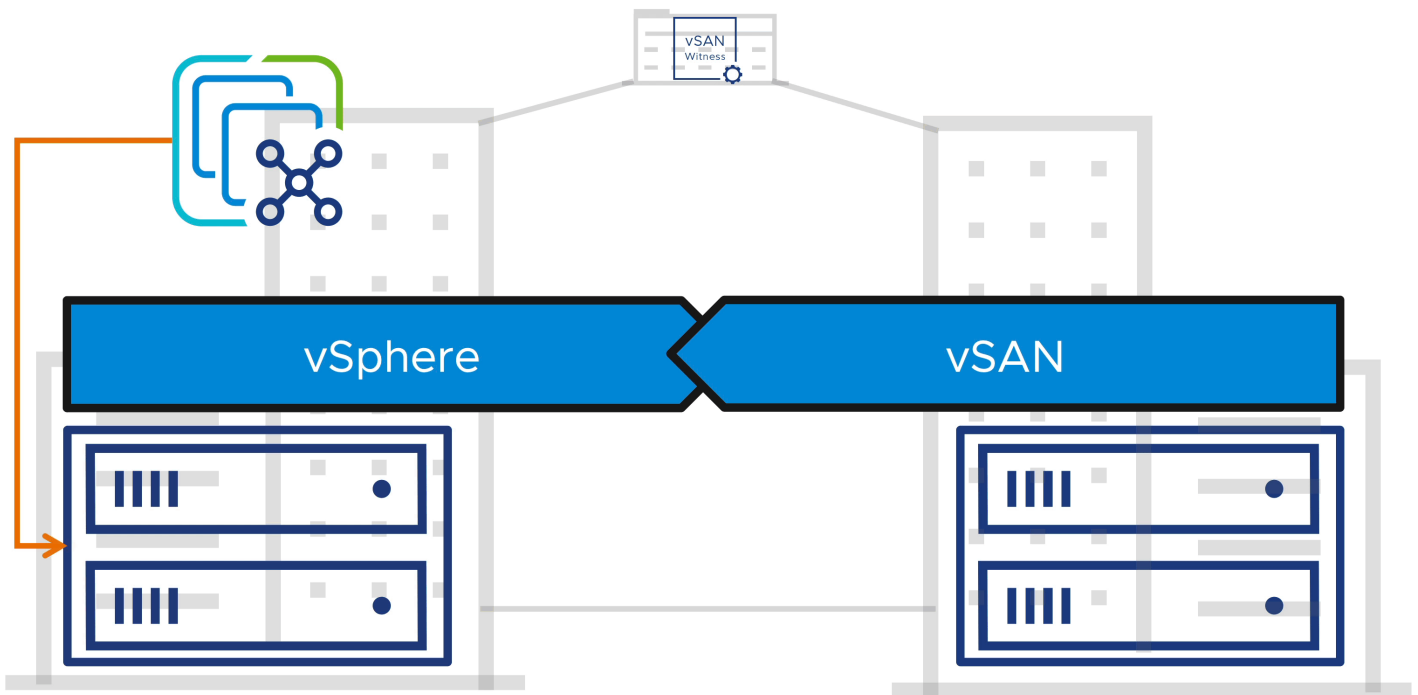
Assuming you have already designed your infrastructure for failure, meaning, your vSphere cluster is made up of ESXi hosts from multiple racks, redundant power supplies, redundant top-of-rack switches, etc. vSphere HA will restart the vCenter Server VM on a new available host if the current host or even the entire rack encounters a failure. vSphere HA operates within the boundaries of a cluster. vSphere HA cannot failover VMs between clusters.



Be mindful of any [vSphere HA admission control](#) settings that may impact the clusters ability to satisfy a large failure.

vSphere HA: vSAN Stretched Cluster

You can combine vSphere HA protection of a vCenter Server with a vSAN stretched cluster to provide additional physical availability. The vSAN cluster is comprised of hosts that span a physical location. vSphere HA can restart the vCenter Server on hosts in the opposite physical location.

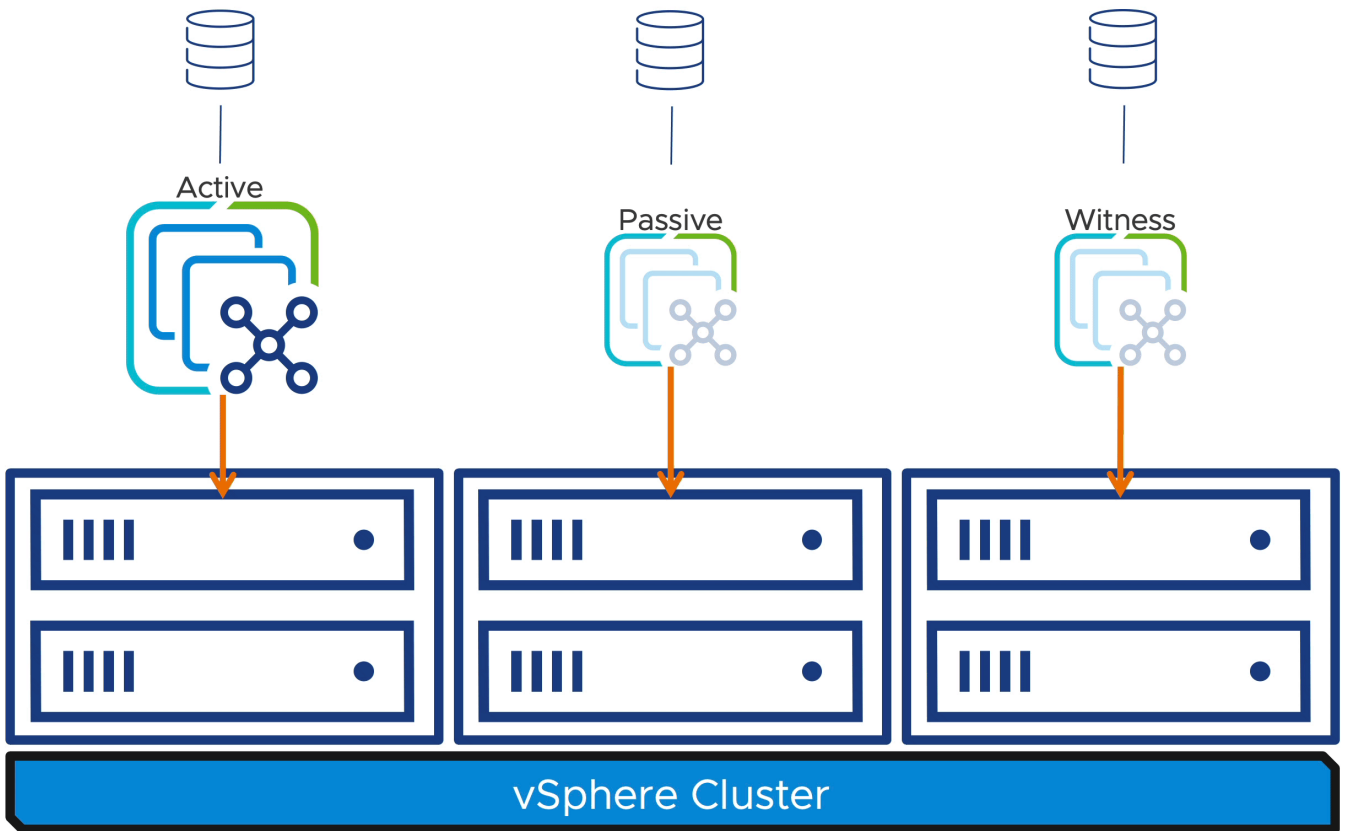


For a detailed look at vSAN stretched clusters, see the [vSAN stretched cluster guide](#).

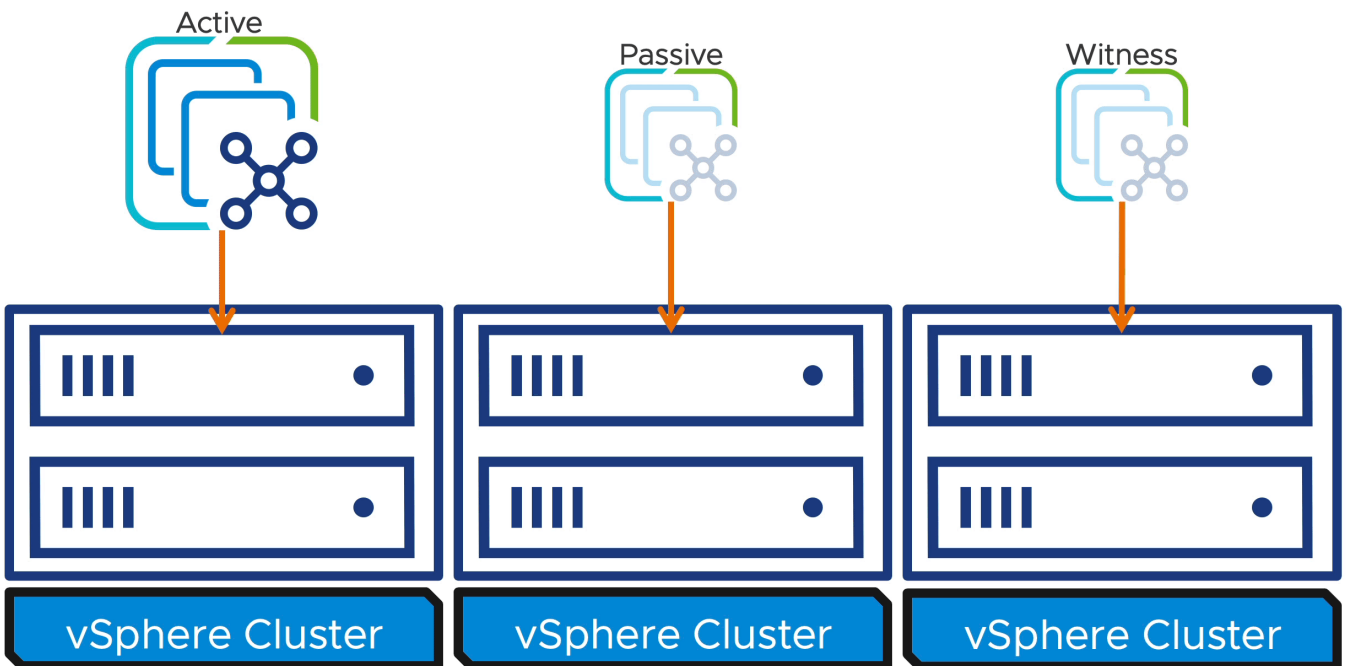
vCenter HA

vCenter HA is deployed across three ESXi hosts and three datastores (or single vSAN datastore). You can deploy vCenter HA within a single vSphere cluster, or you might choose to distribute the vCenter HA nodes across multiple vSphere clusters.

In the event that the current active node fails (be it the underlying host fails, or the node itself) then a failover event is triggered and the current passive node is promoted to the new active node. When the failed node recovers (either automatically, or through user intervention) it will re-join the vCenter HA cluster as the new passive node.



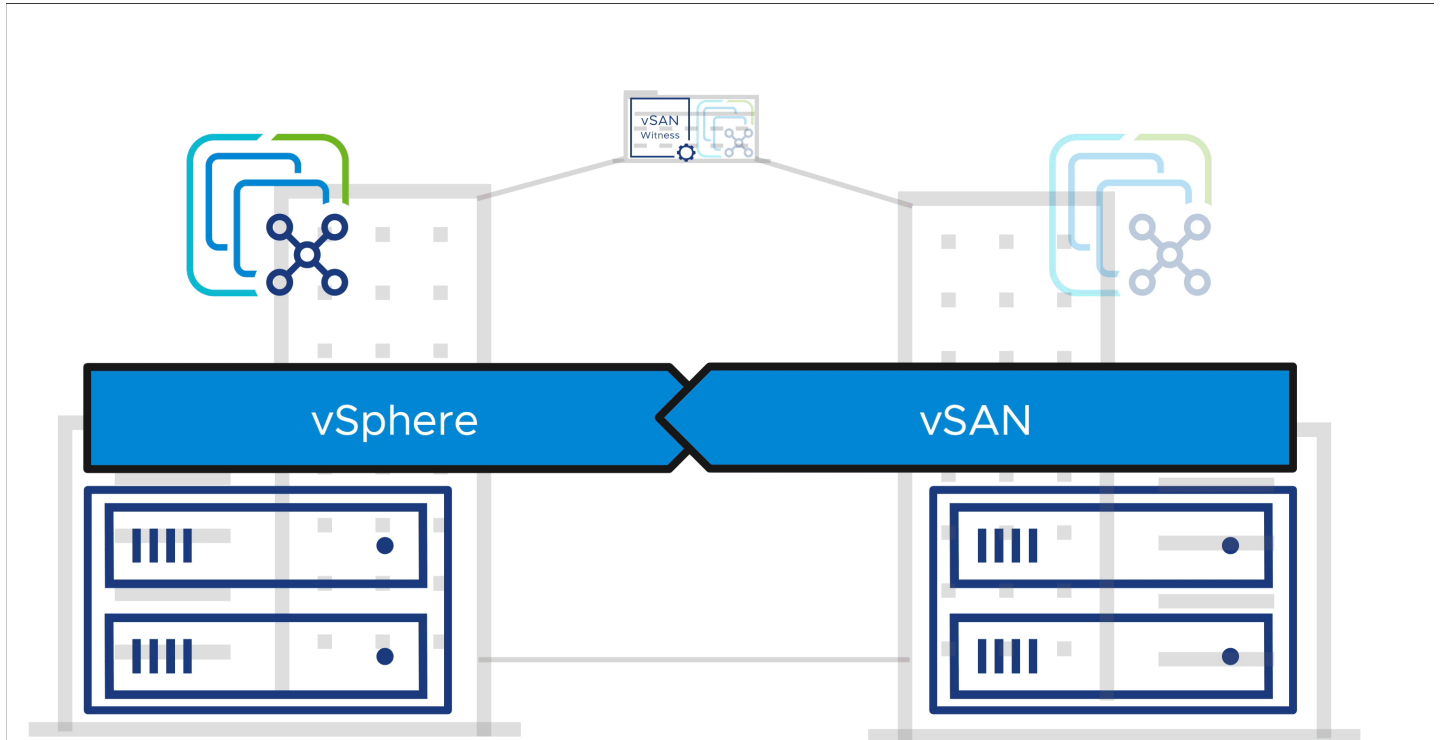
You could also deploy vCenter HA across three separate clusters. The level of availability is largely the same as a single cluster, if the cluster components, hosts, switches, racks, etc, are designed with availability in mind.



vCenter HA: vSAN Stretched Cluster

You can use vCenter HA in co-operation with a vSAN stretched cluster and a third location to host the vCenter HA witness.

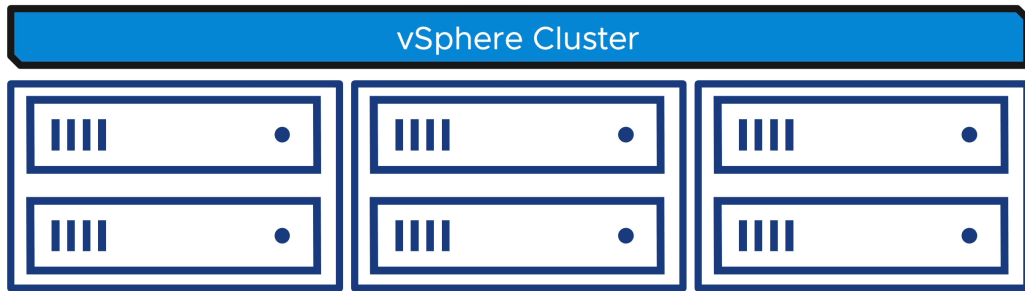
However, keep in mind that vCenter HA does not support the use of a WAN connection for the vCenter HA network and requires a minimum of 1 Gbps bandwidth with less than 10 milliseconds of latency.



vSphere Fault Tolerance

vSphere Fault Tolerance creates an identical duplicate of a virtual machine, and continuously replicates the state of the primary VM to the secondary VM in real-time. vSphere Fault Tolerance is bound to a single vSphere cluster. You cannot separate the primary and secondary VMs between different clusters.

Failover to the secondary is instantaneous and it becomes the new primary. Once that happens a new secondary VM is created and kept in real-time sync with the new primary.



Summary

In summary, vSphere provides various options to provide availability to virtual machines, including vCenter Server. There are pros and cons to each option and topology.

Ultimately you have options to design and architect the vSphere infrastructure for availability and can use as simple a solution as vSphere HA to maintain availability of vCenter Server. There is possibly less downtime with vCenter HA, but this comes at the cost of added complexity, management and maintenance limitations, which for your environment may outweigh the benefits of having a distributed vCenter HA cluster.

Similarly, while vSphere Fault Tolerance can provide a zero-downtime availability solution it also has some management limitations as well as a limit on the compute size of a VM it can protect.

At the end of the day, the decision is yours, and I hope this content helps you understand the options available.

Useful Links

- [vSphere HA: VM and Application Monitoring](#)
- [vSphere HA: VM Component Protection](#)
- [vSphere HA: Best Practices](#)
- [vCenter HA: FAQ](#)
- [vSphere Fault Tolerance: Architecture and Performance White Paper](#)
- [vSphere Fault Tolerance: vSphere Features not Supported](#)
- [vSphere Fault Tolerance: Features and Devices Incompatibility](#)
- [vSAN Stretched Cluster Guide](#)

