vSAN 2-Node Cluster on VxRail Planning Guide

November 2020

Abstract

This guide provides information for the planning of a VMware vSAN 2-node cluster infrastructure on a VxRail platform. This guide focuses on the VxRail implementation of the vSAN 2-node cluster, including minimum requirements and recommendations.

D&LL Technologies

Contents

Revision history

Date	Description
June 2019	Removed Remote or Branch Office (ROBO) licensing requirement
March 2020	Updated to support switch changes made in VxRail v4.7.410
May 2020	Updated to support VxRail v7.0
July 2020	Updated to add D-series to the supported hardware section
November 2020	Updated NIC requirements

Overview

VMware vSAN 2-node cluster is a configuration that is implemented in environments where a minimal configuration is a key requirement.

VxRail v4.7.100 was the first release to support the vSAN 2-node cluster with direct-connect configuration. Starting with VxRail v4.7.410, vSAN 2-node cluster with switch configuration is also supported.

Note: A minimum of four ports are required for both configurations.

This guide provides information for the planning of a vSAN 2-node cluster infrastructure on a VxRail platform. This guide focuses on the VxRail implementation of the vSAN 2-node cluster, including minimum requirements and recommendations.

For detailed information about VMware vSAN 2-node cluster architecture and concepts, see the VMware vSAN 2-Node Guide.

Introduction

A VMware vSAN 2-node cluster on VxRail consists of a cluster with two VxRail nodes, and a Witness host deployed as a virtual appliance. The VxRail cluster is deployed and managed by VxRail Manager and VMware vCenter Server.

A vSAN 2-node configuration is very similar to a stretched-cluster configuration. If there is a failure, the Witness host is the component that provides quorum for the two data nodes. As in a stretched-cluster configuration, the requirement for one Witness per cluster still applies.

Unlike a stretched cluster, the vCenter Server and the Witness host are typically located in a main data center, as shown in Figure 1. The two vSAN data nodes are typically deployed in a remote location. Even though the Witness host can be deployed at the same site as the data nodes. The most common deployment for multiple 2-node clusters is to have multiple Witnesses hosted in the same management cluster as the vCenter Server. This deployment optimizes the infrastructure cost by sharing the vSphere licenses and the management hosts.

This design is facilitated by the low bandwidth that is required for the communication between data nodes and the Witness.

A vSAN 2-node configuration maintains the same high availability characteristics as a regular cluster. Each physical node is configured as a vSAN fault domain. This means that the virtual machines can have one copy of data on each fault domain. If a node or a device fails, the virtual machine remains accessible through the alternate replica and Witness components.

When the failed node is restored, the Distributed Resource Scheduler (DRS) automatically rebalances the virtual machines between the two nodes. DRS is not required but highly recommended. It requires a vSphere Enterprise edition license or higher.

Requirements, Recommendations, and Restrictions

VxRail hardware

The vSAN 2-node configuration is supported using the VxRail D560/F, E560, E560F/N, E665/F/N, P570/F, V570/F, and S570 appliances. The systems can be configured with the following Network Daughter Cards:

Starting with 7.0.100, 25 GbE is supported.

VxRail earlier than 7.0.010:

- 4 x 10 GbE
- 2 x 1 GbE + 2 x 10 GbE

VxRail 7.0.010:

- 4 x 10 GbE
- 2 x 1 GbE + 2 x 10 GbE
- Customer-Supplied VDS (still requires four ports).

VxRail 7.0.100:

- 4 x 10 GbE
- 2 x 1 GbE + 2 x 10 GbE
- 4 x 25 GbE (2 x 25 GbE on NDC + 2 x 25 GbE on PCIe)
- Customer-Supplied VDS (still requires four ports).

VxRail software version

VxRail v4.7.100 or later supports the vSAN 2-node cluster with direct-connect configuration.

VxRail v4.7.410 or later supports the vSAN 2-node cluster with switch configuration.

VMware vCenter Server

The vSAN 2-node cluster must be connected to an external vCenter Server at the time of its deployment. VxRail-supplied vCenter is not supported.

- VMware vCenter Server version requirements:
  - Version 6.7u1 or later is required for 2-node cluster with direct-connect configuration.

- Version 6.7u3 or later is required for switch configurations.

- The vCenter Server must be deployed before the deployment of the 2-node cluster.

- vCenter Server <u>cannot</u> be deployed on the 2-node cluster.

Witness virtual appliance

VMware supports both physical ESXi hosts and a virtual appliance as a vSAN Witness host. VxRail v4.7.100 supports using the vSAN Witness virtual appliance as a vSAN Witness host only. The Witness virtual appliance does not consume extra vSphere licenses and does not require a dedicated physical host.

Note: The Witness ESXi OVA host deploys a Virtual Standard witness Switch (vSS). See the VMware documentation for more details.

Software version

- vSAN Witness appliance version 6.7u1 or later is required.

- Witness appliance must be at the same vSAN version as the ESXi hosts.

- The vSphere license is included and hard-coded in the Witness virtual appliance.

Installation

- The Witness appliance must be installed, configured, and added to vCenter inventory before the vSAN 2-node cluster on VxRail deployment.

- The Witness appliance must have connectivity to both vSAN nodes.

- The Witness vSAN traffic must be on a different vLAN than the data nodes witness traffic.

- The Witness appliance must be managed by the same vCenter Server that is managing the 2-node cluster.

    In VxRail earlier than v7.0.100, a Witness appliance can only be connected to one vSAN 2-node cluster. In v7.0.100 and later, a single witness can support up to 64 2-node clusters.

- The Witness can run in the same physical site as the vSAN data nodes. However, it cannot be placed in the 2-node cluster to which it provides quorum.

- The general recommendation is to place the vSAN Witness host in a different data center, such as a main data center or a cloud provider.

- It is possible to deploy the Witness appliance on another 2-node cluster, but it is not recommended. A VMware RPQ is required for this solution design.

- Direct-connect configurations require Ports 1 and 2 to be connected to the switches. Ports 3 and 4 from Node 1 are directly connected to Ports 3 and 4 of Node 2 respectively.

- Switched configurations require all four ports to be connected to the switches.

Sizing

- There are three typical sizes for a witness appliance that can be selected during deployment: Tiny, normal, and large. Each option has different requirements for compute, memory, and storage. Select the appropriate size from the deployment drop-down menu.

- The general recommendation is to use the normal size. However, 2-node clusters with up to 25 VMs are good candidates for the "Tiny" option because they are less likely to reach or exceed 750 components.

    - Each storage object is deployed on vSAN as a RAID tree and each leaf of the tree is said to be a component. For instance, when we deploy a VMDK with a RAID 1 mirror, we will have a replica component in one host and another replica component in another host. The number of stripes that are used has an effect. For example, if using two stripes we will have two replica components in each host.

Physical network In releases earlier than 4.7.410, the two vSAN data nodes must be directly connected using SFP+ or Cat6 cables. Starting with VxRail 4.7.410, the two vSAN data nodes can be connected through the switch.

The following physical layout is enforced:

- Either a 1/10/25 GbE switch is supported. 1 GbE switch is only supported in a directly connected configuration.

- Ports 1 and 2 of the VxRail Appliances are connected to a switch and used for the management and witness traffic. If the ports are connected to a 1 GbE switch, port speed auto-negotiates down to 1 Gb.

- If direct-connect is desired, Ports 3 and 4 from Node 1 are directly connected to Ports 3 and 4 of Node 2 respectively and are used for vSAN and vMotion traffic.

- If switched configuration is desired, all four ports must be connected to the switches.
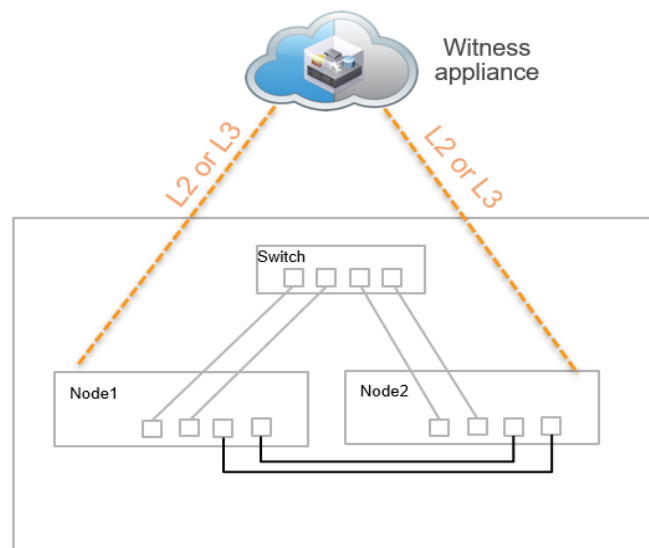
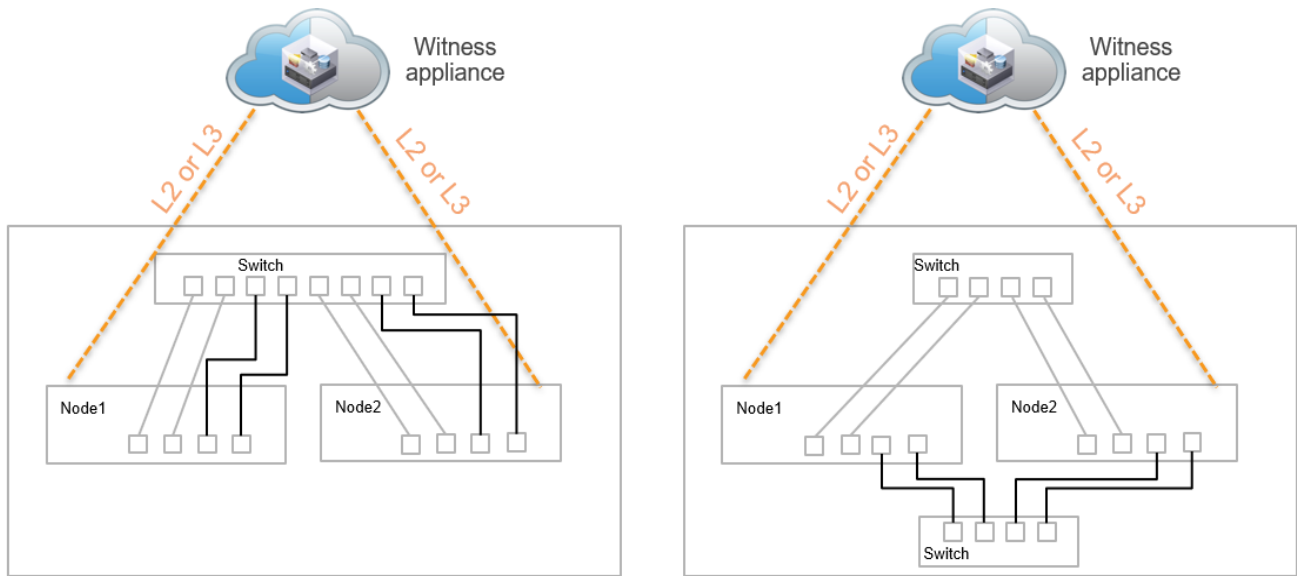**Figure 1.   4 x 10 GbE Direct-connect port configuration**

Figure 2.   4 x 10 GbE ToR switch port configuration single switch (left) dual switch (right)

Port requirements

The following table lists the services that are needed. The incoming and outgoing firewall ports for these services should be opened. See https://ports.vmware.com/home/vSphere for the latest list of firewall ports.

Table 1.   Service ports on VxRail Appliance

Services	Port	Protocol	To/From
vSAN Clustering Service	12345, 23451	UDP	vSAN Hosts
vSAN Transport	2233	TCP	vSAN Hosts
vSAN VASA Vendor Provider	8080	TCP	vSAN Hosts and vCenter Server
vSAN Unicast Agent to the Witness Host	12321	UDP	vSAN Hosts and Witness Appliance

**Witness and Management network topology**

VMware recommends that the vSAN communications between vSAN nodes and the vSAN Witness host be:

- Layer 2 (same subnet) for configurations with the Witness host in the same location

- Layer 3 (routed) for configurations with the Witness host in an alternate location such as at the main data center

    - A static route is required.

The maximum supported roundtrip time (RTT) between the vSAN 2-node cluster and the Witness is 500 milliseconds (250 milliseconds each way).

In the VxRail implementation of the vSAN 2-node cluster, a VMkernel interface is designated to carry traffic destined for the Witness host.
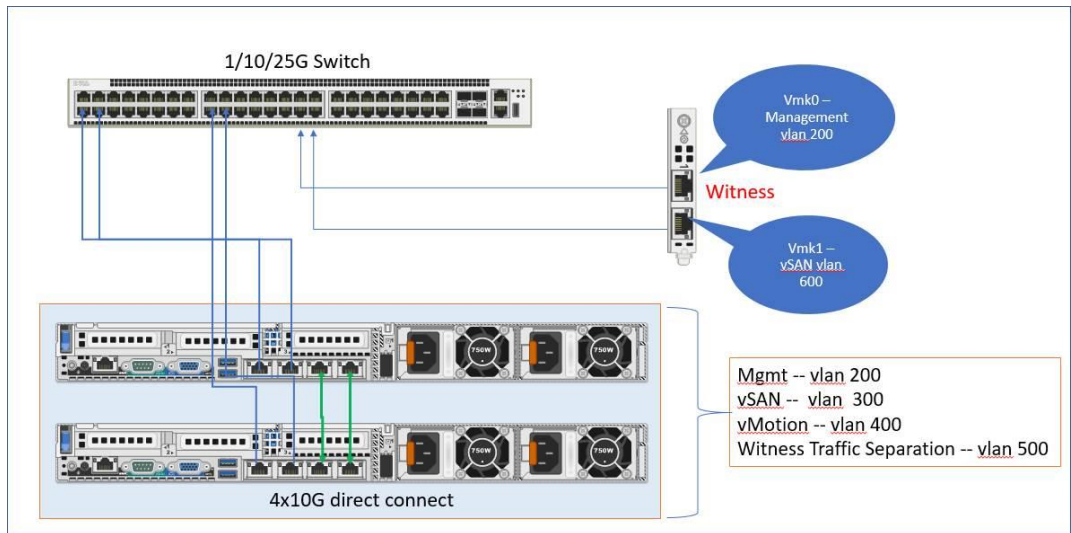
Figure 3.   4x10G Direct-connect port configuration

Each vSAN host's vmk5 VMkernel interface is tagged with "witness" traffic. When using Layer 3, each vSAN host must have a static route configured for vmk5 and be able to properly access the vmk1 on the vSAN Witness host, which is tagged with "vSAN" traffic.

Likewise, the vmk1 interface on the Witness host must have a static route configured to properly communicate with vmk5 on each vSAN host.

Network layout   The following chart illustrates the network layout that is used by VxRail in the configuration of a vSAN 2-node cluster. Although VMware supports using an alternative vLAN other than vSAN vLAN for WTS, VxRail requires the WTS to have its own vLAN. This layout is specific to the VxRail vSAN 2-node cluster. The configuration of the management cluster is slightly different as described in the *VxRail Network Planning Guide.*

Traffic	Requirements	Members	NIOC Shares	UPLINK1 VMNIC0	UPLINK2 VMNIC1	UPLINK3 VMNIC2	UPLINK4 VMNIC3
Management Network	Same VLAN as vCenter Server network	ESXi vmk2,	40	Standby	Active	Unused	Unused
vCenter Server Network	Same VLAN as management network	vCenter, VxM VNIC0				Unused	Unused
VxRail Management	VLAN 3939, can't change the physical port.	ESXi vmk0, VxM VNIC1(New)		Standby	Active	Unused	Unused
Witness Traffic		ESXi vmk5		Active	Standby		
vSAN		ESXi vmk3	100	Unused	Unused	Active	Standby
vMotion		ESXi vmk4	50	Unused	Unused	Standby	Active
Virtual Machines		Virtual Machines	60	Active	Standby	Unused	Unused

Figure 4.   Network layout of a VxRail vSAN 2-node cluster

vSAN 2-Node Cluster on VxRail Planning Guide   **9**

Capacity planning considerations

In this section, we offer general recommendations for storage, CPU, memory, and link bandwidth sizing.

Storage capacity

- A minimum of 25% to 30% of spare storage capacity is required for a 2-node cluster.

- Note that in a 2-node cluster, the protection method will be RAID 1. If a node fails, the surviving node will continue to operate with a single object's component.

CPU and memory capacity

- When defining CPU and memory capacity, consider the minimum capacity required to satisfy the VM requirements while in a failed state.

- The general recommendation is to size a cluster to operate below 50% of the max CPU required, taking in consideration the projected growth in consumption.
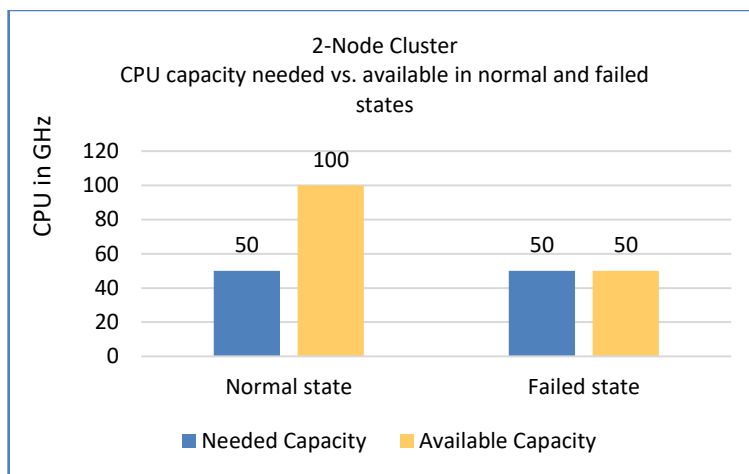

Figure 5. CPU capacity planning

Network Bandwidth

Our measurements indicate that a regular T1 link can satisfy the network bandwidth requirements for the communications between Data Nodes <> vCenter Server and Data Nodes <> Witness Appliances. However, for the purpose of adapting the solution to different service level requirements, it is important to understand in more details the requirements for:

- Normal cluster operations

- Witness contingencies

- Services, such as maintenance, lifecycle management, and troubleshooting
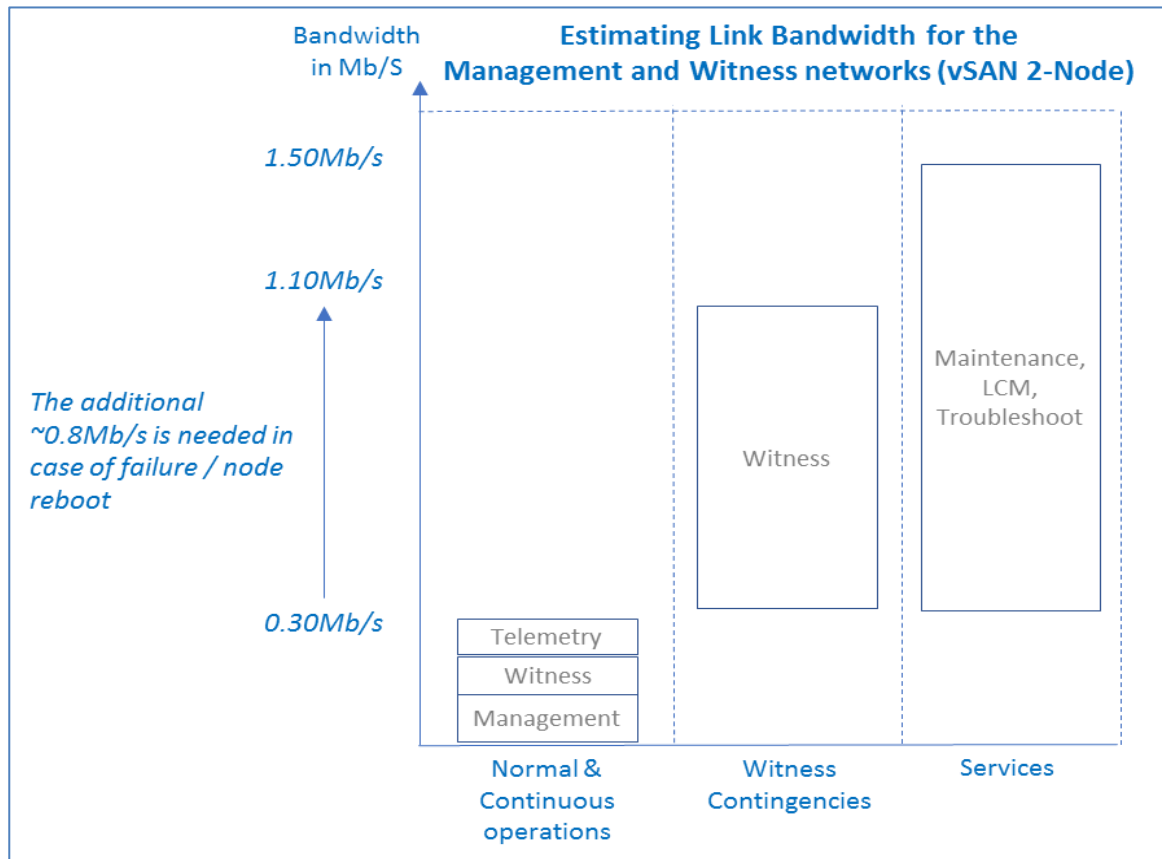
Figure 6.   Network bandwidth planning considerations

Normal cluster operations

- Normal cluster operations include the traffic between data nodes, vCenter Server, and the Witness appliance.

- During normal operations, the bulk of the traffic is between data nodes and vCenter Server. This traffic is affected primarily by number of VMs and number of components but, is typically very light load.

- Our measurements of a cluster with 25 VMs and near 1,000 components indicated a bandwidth consumption lower than 0.3 Mbps.

Witness contingencies

- The Witness appliance does not maintain any data, only metadata component.

- The Witness traffic can be influenced by the I/O workload running in the cluster, but in general, this is very small traffic while the cluster is in a normal state.

- In the event the primary node fails or is partitioned, the following occurs:

  - vSAN powers off the VMs in the failed host.

  - The secondary node is elected as the HA primary. The Witness host sends updates to the new primary, followed by the acknowledgment from the primary that the ownership is updated.

  - 1,138 bytes is required for each component update.

- When the update is completed, quorum is formed between the secondary host and the Witness host, allowing the VMs to have access to their data and be powered on.

- The failover procedure requires enough bandwidth to allow for the ownership of components to change within a short interval of time.

- Our recommendation for a 2-node cluster with up to 25 VMs is that at least 0.8 Mbps be available to ensure a successful failover operation.

Maintenance, lifecycle management and troubleshooting

- The amount of bandwidth reserved for maintenance, lifecycle management and troubleshooting are determined primarily by the desired transfer times for large files.

- The log files that are used in troubleshooting are compressed and typically can be transferred in a reasonable time.

- The composite files that are used for software and firmware upgrades can be up to 4.0 GB and can take a long time to be transferred when using a T1 link. The bandwidth requirements should be evaluated in case you have specific maintenance window requirements.

  - As a reference, if using a T1 link, we expect that at least 1 Mb/s of bandwidth is available for the transfer of the composite file. We estimate that this transfer will take about nine hours.

Node upgrades   All components except for the witness nodes are upgraded using VxRail LCM. You must manually upgrade the witness nodes.

Licensing   The vSAN licensing editions that are listed in the following table can be used on a vSAN 2-node cluster.

	Standard	Advanced	Enterprise	Enterprise Plus
Storage Policy Based Mgmt.	✓	✓	✓	✓
Virtual Distributed Switch	✓	✓	✓	✓
Rack Awareness	✓	✓	✓	✓
Software Checksum	✓	✓	✓	✓
All-Flash Hardware	✓	✓	✓	✓
iSCSI Target Service	✓	✓	✓	✓
QoS – IOPS Limit	✓	✓	✓	✓
Cloud Native Storage	✓	✓	✓	✓
Deduplication and Compression		✓	✓	✓
RAID-5/6 Erasure Coding		✓	✓	✓
vRealize Operations with vCenter		✓	✓	✓
Data-at-Rest Encryption			✓	✓
Stretched Cluster with Local Failure Protection			✓	✓

Note: VxRail 2-node clusters do not support the following:

- Expansion to more than two nodes
- RAID-5/6 erasure coding

While vSAN ROBO licenses are supported for VxRail vSAN 2-node clusters, these licenses are not orderable through VxRail. Witness Appliance license is not required but the host where the Witness resides needs the appropriate vSphere license. For more information, see the VMware vSAN Licensing guide.

Deployment Types

Centralized management (Option 1)

In this scenario, the vCenter Servers and Witness virtual appliances are deployed at the same management cluster located at a main data center. One vCenter Server instance can manage multiple VxRail vSAN 2-node clusters but each VxRail vSAN 2-node cluster must have its own Witness.

Network bandwidth must meet the requirements stated earlier. Enhanced link mode is recommended.
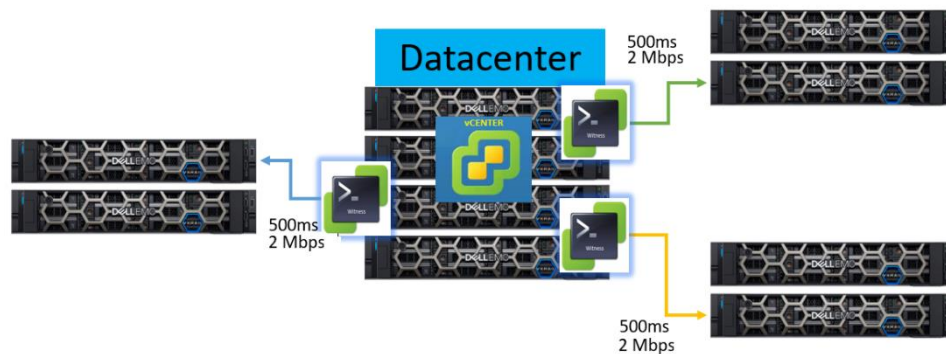
Figure 7.   Centralized management of vCenter Server and Witness appliances

**Centralized management, localized Witness (Option 2)**

In this deployment option, the vCenter server is located at the main data center, but the vSAN Witness appliance and the two VxRail nodes are at the same location. An additional ESXi host is required to host the vSAN Witness appliance. vSAN Witness appliance cannot be hosted in the VxRail 2-node cluster.
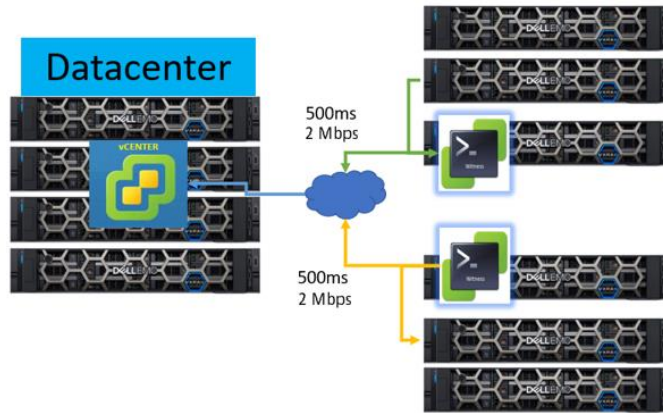
Figure 8.   Centralized vCenter with local Witness Appliances

Localized management and Witness (Option 3)

In this option, the three fault domains are at the same location; the vCenter Server, vSAN Witness appliance, and the VxRail nodes. An additional ESXi host is required to host vSAN Witness appliance and customer-supplied vCenter Server. vSAN Witness appliance, and customer-supplied vCenter cannot be hosted in the VxRail vSAN 2-node cluster.
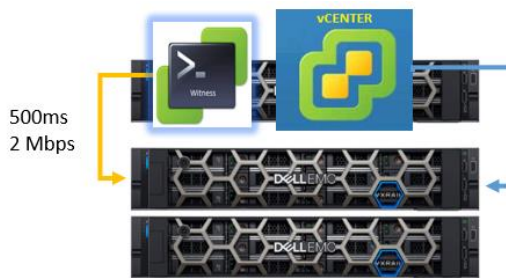
Figure 9.   Localized vCenter and Witness Appliance

Table 2.   Deployment options considerations

Option	Pros	Cons
Centralized Management and Witness	Single pane of glass for the management of multiple 2-node clusters Centralization of Witness appliances reduces licensing and hardware costs.	Network costs for vCenter and Witness communications
Centralized Management, localized Witness	Single pane of glass for the management of multiple 2-node clusters	Network costs for vCenter communications Software and hardware costs for deployment of witness appliances
Localized management and witness	Reduces network cost that is associated to normal operations and witness contingency	Software and hardware costs for deployment of multiple vCenter Servers and witness appliances Network Bandwidth that is still needed for maintenance and troubleshooting which is the larger bandwidth requirement

Conclusion

Starting with VxRail 4.7.100, VMware vSAN 2-node cluster direct connect is supported using VxRail E-Series platforms. Switch configuration is supported in v4.7.410 or later. A VMware vSAN 2-node cluster is a minimal configuration consisting of two vSAN data nodes and a Witness virtual appliance.

vSAN 2-node cluster can easily be deployed anywhere. Many vSAN 2-node clusters can be managed by a single vCenter instance. This minimal configuration continues to provide the same functional benefits of vSphere and vSAN. It enables an efficient centralized management with reduced hardware and software costs, while meeting the needs of environments with limited space, budget and/or IT personnel constraints.

References

- vSAN 2-Node Guide *(https://storagehub.vmware.com/t/vmware-vsan/vsan-2-node-guide/)*
- vSAN Stretched Cluster *(https://storagehub.vmware.com/t/vmware-vsan/vsan-stretched-cluster-guide/)*
- 2-Node vSAN – Witness Network Design Considerations *(https://cormachogan.com/2017/10/06/2-node-vsan-witness-network-design-considerations/)*
- vSAN Stretched Cluster Bandwidth Sizing *(https://storagehub.vmware.com/t/vmware-vsan/vsan-stretched-cluster-bandwidth-sizing/)*
- VxRail Network Guide *(https://www.dellemc.com/resources/en-us/asset/technical-guides-support-information/products/converged-infrastructure/h15300-vxrail-network-guide.pdf)*
- VxRail vCenter Server Planning Guide *(https://www.dellemc.com/resources/en-us/asset/technical-guides-support-information/products/converged-infrastructure/vxrail-vcenter-server-planning-guide.pdf)*