Dell EMC VxRail™ 7.0 vSAN Stretched Cluster Planning Guide

Abstract

This planning guide provides best practices and requirements for using stretched clusters with VxRail appliances.

May 2020

The information in this publication is provided "as is." Dell Technologies makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software that is described in this publication requires an applicable software license.

Copyright © 2017-2020 Dell/EMC or its subsidiaries. All Rights Reserved. Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners. [5/5/2020] [Planning Guide] [H18306]

Contents

1 Executive summary

A stretched cluster is a deployment model in which two or more virtualization host servers are part of the same logical cluster but are located in separate geographical locations. The vSAN stretched cluster feature enables synchronous replication of data between sites. This feature allows for an entire site failure to be tolerated. It extends the concept of fault domains to data center awareness domains.

vCenter Server is the centralized platform for managing a VMware environment. It is the primary point of management for both server virtualization and vSAN. It is also the enabling technology for advanced capabilities such as vMotion, Distributed Resource Scheduler (DRS), and HA. vCenter scales to enterprise levels where a single vCenter can support up to 2,500 hosts (VxRail nodes) and 30,000 virtual machines. vCenter supports a logical hierarchy of data centers, clusters, and hosts, which allow resources to be segregated by use cases or lines of business. It also allows resources to be moved as needed dynamically. These operations all done from a single interface.

1.1 Intended Use and Audience

This guide is intended for customers, Dell EMC and business partners, and implementation professionals. It is designed to help you understand the requirements for stretched cluster support with the Dell EMC VxRail™ Appliance. Services from Dell EMC or an authorized VxRail Services Partner are required for implementation of stretched clusters.

This document is not intended to replace the implementation guide or to bypass the service implementation required for stretched clusters. An attempt to set up stretch clusters on your own will invalidate support.

Note: For versions earlier than VxRail 7.0.000, refer to the *Dell EMC VxRail vSAN Stretched Cluster Planning Guide* located at https://www.dellemc.com/en-us/collaterals/unauth/white-papers/products/converged-infrastructure/h15275-vxrail-planning-guide-virtual-san-stretched-cluster.pdf

2 Overview

This planning guide provides best practices and requirements for using stretched cluster with a VxRail Appliance. This guide assumes that the reader is familiar with the *vSAN Stretched Cluster Guide*. This guide is for use with a VxRail Appliance only.

The vSAN stretched cluster feature creates a stretched cluster between two geographically separate sites, synchronously replication data between sites. This feature allows for an entire site failure to be tolerated. It extends the concept of fault domains to data center awareness domains.

The following is a list of the terms that are used for vSAN stretched clusters:

- *Preferred or Primary site* – one of the two data sites that is configured as a vSAN fault domain.
- *Secondary site* – one of the two data sites that is configured as a vSAN fault domain.
- *Witness host* – a dedicated ESXi host or vSAN witness appliance can be used as the witness host. The witness components are stored on the witness host and provide a quorum to prevent a split-brain scenario if the network is lost between the data sites. This is the third fault domain.

The vSAN storage policies that impact stretched cluster are:

- *Dual site mirroring (stretched cluster)* – enables protection across sites.
- *None – keep data on Preferred (stretched cluster)* – Keep data on primary site only, no cross-site protection.
- *None – keep data on Non-preferred (stretched cluster)* – Keep data on secondary site only, no cross-site protection.
- *Failures to Tolerate* – defines how many disk or node failures can be tolerated for each site. For stretched cluster, it's '2n+1' (n is the number to tolerate). For erasure coding, it's 4 or 6 (1 or 2 failures respectively). All-Flash is required for erasure coding (RAID5 or 6).

2.1 Distributed Resource Scheduler

For vSAN stretched cluster functionality on VxRail, vSphere Distributed Resource Scheduler (DRS) is required. DRS provides initial placement assistance, and automatically migrates virtual machines to the corrected site in accordance with the Host and VM affinity rules. It can also help relocate virtual machines to their correct site when a site recovers after a failure.

2.2 Fault domains

Fault domains (FD) provide the core functionality of vSAN stretched cluster. The supported number of fault domains in a vSAN stretched cluster is three. The first Fault Domain can be referred to as Preferred data site. The second Fault Domain can be referred to as Secondary data site, and the third Fault Domain is the witness host site. Keep utilization per data site below 50% to ensure proper availability, if either the Preferred or Secondary site goes offline.

2.3 VxRail cluster nodes

vSAN stretched clusters are deployed across two sites in an active/active configuration. An identical number of ESXi hosts are recommended. Unbalanced configuration is supported it with the use of Site Affinity Rule.

2.3.1 VxRail cluster deployment options

You must plan the VxRail stretched cluster deployment before installation. Depending on the number of nodes in the VxRail cluster, you can:

- Deploy up to 16 nodes on initial deployment or
- Initially deploy the minimum number of nodes per site and then scale out additional nodes either at installation or during the VxRail stretched cluster life cycle.

2.4 Witness host

A vSAN Witness Appliance, or a physical host, can be used for the Witness function. The vSAN Witness Appliance includes licensing, while a physical host must be licensed accordingly.

The Witness OVA can be deployed on an ESXi 6.5 or higher, but the ESXi host CPU must meet the requirements of vSphere you are installing.

Deploying the witness for a normal configuration requires 10GB of cache device. The physical host does not require flash nor SSD device. The capacity device must be 350GB or greater. Traditional spinning drives are enough because the witness will mark those devices as needed.

Each vSAN stretched-cluster configuration requires a Witness host, a witness cannot be shared among the clusters. The Witness must reside on a third site that has independent paths to each data site. While the Witness host must be part of the same vCenter as the hosts in the data sites, it must not be on the same cluster as the data site hosts. The Witness ESXi OVA is deployed using a virtual standard switch (vSS). Network Address Translation (NAT) is not supported.

NOTE: The Witness host OVA file includes a license, it does not consume a vSphere license. However, a physical host requires a vSphere license.

2.5 VxRail cluster requirements

This section describes the requirements to implement vSAN stretched clusters in a VxRail Cluster running VxRail v7.0.

- The VxRail cluster must be deployed across two physical sites in an active/active configuration.
- The VxRail cluster must be running VxRail version 7.0 or later.
- Failure Tolerance Method of RAID5 or 6 must be all-flash.
- The minimum supported configuration is 1+1+1 (2 nodes +1 witness). See the *vSAN 2-Node Cluster on VxRail Planning Guide* for more detailed information.
- The maximum supported configuration is 15+15+1 (30 nodes+1 witness).
- A witness host must be installed on a separate site as part of the installation engagement. The witness must be running the same version as the vSphere. See Table 1 for version compatibility.

VxRail Version	Witness Host OVA Version
VxRail 7.0.x	OVA Version 7.0

Table 1 **VxRail/Witness Host OVA Compatibility Chart**

2.6 vCenter Server requirements

With VxRail 7.0, either a VxRail vCenter Server or a customer-supplied vCenter Server can be used for stretched clusters. See the *VxRail vCenter Server Planning Guide* for limitations of using VxRail vCenter Server.

Customer-supplied vCenter Server Appliance is the recommended choice.

2.7 Customer-supplied vCenter Server requirements

The following are the customer-supplied vCenter Server requirements:

- The customer must provide the vSphere Enterprise Plus license.
- The customer-supplied vCenter Server version must be in the VxRail and external vCenter interoperability matrix.
- Check the VxRail Release Notes for to determine the proper version numbers.
    - VxRail 7.x and vSphere 7.x, version details can be found *in VxRail Appliance Software 7.x Release Notes*.

To join the customer-supplied vCenter Server, you must:

- Know the customer-supplied vCenter Server FQDN.
- Know the Customer Existing Single Sign-on domain (SSO) (For example vsphere.local).
- Create or select a data center on the customer-supplied vCenter Server for the VxRail Cluster to join.
- Specify the name of the cluster that will be created by VxRail in the selected data center when the cluster is built. It will also be part of the distributed switch name. This name must be unique and not used anywhere in the data center on the customer-supplied vCenter Server.
- Verify that the customer DNS server can resolve all VxRail ESXi hostnames before deployment.
- Create or reuse a VxRail management user and password for this VxRail cluster on the customer-supplied vCenter Server. The user must be:
    - Created with no permissions
    - Created with no roles assigned to it
- (Optional) If the administrator account cannot be used, create a VxRail admin user and password for VxRail on the Customer-Supplied vCenter Server.

3 Networking and latency

3.1 Layer 2 and Layer 3 support

Both Layer 2 and Layer 3 support vSAN connectivity between data nodes. Layer 2 does not require a static route, but Layer 3 does.

Using Layer 3 is recommended for connectivity between the data sites and the witness. Witness Traffic Separation (WTS) can be used with Stretched Cluster configurations under the following conditions:

- If a VMkernel interface other than the Management VMkernel interface is tagged with "witness" traffic, static routes are required to communicate with the vSAN Witness Host VMkernel interface tagged for vSAN Traffic.
- If the Management VMkernel interface is tagged with "witness" traffic, static routes are not required if the host can already communicate with the vSAN Witness Host VMkernel interface using the default gateway.

MTU can be different. For instance, MTU of 9000 between data nodes, MTU of 1500 between data nodes and the witness.

3.2 Witness traffic separation

vCenter Server 6.7u1 supports Witness Traffic Separation (WTS) for VxRail stretched-cluster deployments. This feature allows an alternate VMkernel interface to be designated to carry traffic that is destined for the Witness rather than the vSAN tagged VMkernel interface. This feature supports more flexible network configurations by allowing separate networks for node-to-node and node-to-witness traffic. From a routing perspective, this feature allows two independent subnets and routes to be advertised from each Data Node site to the Witness site.

3.3 Supported geographical distances

For vSAN stretched clusters, support is based on network latency and bandwidth requirements, rather than distance. The key requirement is the actual latency numbers between sites.

3.4 Data site to data site network latency

Latency or Round-Trip Time (RTT) between sites hosting virtual machine objects should not be greater than 5 milliseconds (<= 2.5 milliseconds one-way).

3.5 Data site to data site bandwidth

Bandwidth between sites hosting virtual machine objects are workload-dependent. For most workloads, VMware recommends a minimum of 10 Gbps or greater bandwidth between sites.

3.6 Data site to witness network latency

In most vSAN stretched-cluster configurations, latency or RTT between sites hosting VM objects and the witness nodes should not be greater than 200 milliseconds (100 milliseconds one-way).

The latency to the witness is dependent on the number of objects in the cluster. VMware recommends that on vSAN stretched-cluster configurations up to 10+10+1, a latency of less than or equal to 200 milliseconds is acceptable. If possible, a latency of less than or equal to 100 milliseconds is preferred. For configurations that are greater than 10+10+1, VMware recommends a latency of less than or equal to 100 milliseconds is required.

3.7 Data site to witness network bandwidth

Bandwidth between sites hosting VM objects and the witness nodes are dependent on the number of objects residing on vSAN. Size the data site to witness bandwidth appropriately for both availability and growth. A standard rule of thumb is 2 Mbps for every 1000 objects on vSAN.

3.8 Intersite MTU consistency

Unless Witness Traffic Separation is used, you must maintain a consistent MTU (maximum transmission unit) size between data nodes and the witness in a stretched-cluster configuration. Ensuring that each VMkernel interface designated for vSAN traffic is set to the same MTU size prevents traffic fragmentation. The vSAN Health Check looks for a uniform MTU size across the vSAN data network, and reports on any inconsistencies.

3.9 Connectivity

- Management network: Connectivity to all three sites
- VM network: Connectivity between the data sites (the witness will not run virtual machines that are deployed on the vSAN cluster).
- vMotion network: Connectivity between the data sites (virtual machines are never migrated from a data host to the witness host).
- vSAN network: Connectivity to all three sites

Appendix A: VxRail stretched cluster setup checklist

Required Reading	✓ Read the *VMware vSAN Stretched Cluster Guide.* ✓ Read the *VxRail vCenter Server Planning Guide.*
VxRail Version	✓ No mixed clusters are supported (that is, VxRail 4.7 and 7.0 in the same cluster).
vSphere License	✓ vSphere Enterprise Plus license is required. ✓ You cannot reuse the VxRail vCenter Server license on any other deployments.
Number of Nodes	✓ The minimum supported configuration is 1+1+1 (2 nodes+1 witness). ✓ The maximum supported configuration is 15+15+1 (30 nodes+1 witness).
customer-supplied vCenter Server (Recommended choice)	✓ The customer-supplied vCenter Server version must be in the *VxRail and external vCenter interoperability matrix.*
Fault Domains	✓ Must have 3 Fault Domains (preferred, secondary, and witness host).
Network Topology	✓ vSAN traffic between the data sites can be Layer 2 or Layer 3. ✓ vSAN traffic between the witness host and the data sites should be Layer 3. ✓ Witness Traffic Separation allowed users to use an alternative VMkernel interface for traffic to the Witness other than the vSAN interface.
Data Site to Data Site Network Latency	✓ Latency or RTT between data sites should not be greater than 5 milliseconds. (<2.5 milliseconds one-way)
Data Site to Data Site Bandwidth	✓ A minimum of 10 Gbps is required.
Data Site to Witness Network Latency	✓ For configurations up to 10+10+1, latency or RTT less than or equal to 200 milliseconds is acceptable. ✓ For configuration greater than 10+10+1, latency or RTT less than or equal to 100 milliseconds is required.
Data Site to Witness Network Bandwidth	✓ The rule of thumb is 2 Mbps for every 1000 objects on vSAN.
Intersite MTU consistency	✓ Required to be consistent between data sites. ✓ MTU to the Witness can be different.
Network Ports	✓ Review *Appendix B* for required port connectivity.

Appendix B: VxRail stretched cluster open port requirements

The following table lists the open port requirements for a VxRail stretched cluster. See https://ports.vmware.com/home/vSphere for the latest list of port connectivity requirements.

Description	Connectivity To and From	L4 Protocol	Port
vSAN Clustering Service	vSAN Hosts	UDP	12345, 23451
vSAN Transport	vSAN Hosts	TCP	2233
vSAN VASA Vendor Provider	vSAN Hosts and vCenter	TCP	8080
vSAN Unicast Agent (to Witness Host)	vSAN Hosts and vSAN Witness Appliance	UDP	12321