



VMware® Virtual SAN™ Stretched Cluster

Bandwidth Sizing Guidance

TECHNICAL WHITE PAPER



Table of Contents

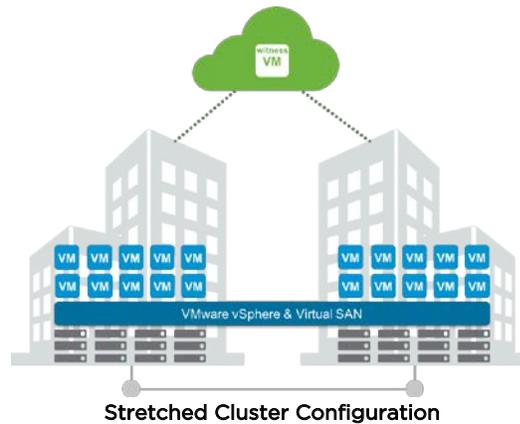
Overview	2
General Guidelines	2
Bandwidth Requirements Between Sites.....	3
Site to Site Examples	3
Bandwidth Requirements Between Witness & Data Sites	4
Witness to Site Examples.....	4
Witness Bandwidth for 2 Node Configurations.....	5
Next Steps.....	7
Additional Documentation.....	7
VMware Contact Information	7
About the Author	7

Overview

The purpose of this document is to explain how to size bandwidth requirements for Virtual SAN in Stretched Cluster configurations. This document only covers the Virtual SAN network bandwidth requirements.

In Stretched Cluster configurations, two data fault domains have one or more hosts, and the third fault domain contains a witness host or witness appliance. In this document each data fault domain will be referred to as a *site*.

Virtual SAN Stretched Cluster configurations can be spread across distances, provided bandwidth and latency requirements are met.



General Guidelines

The bandwidth requirement between the main sites is highly dependent on the workload to be run on Virtual SAN, amount of data, and handling of failure scenarios.

Under normal operating conditions, the basic bandwidth requirements are:

	Connectivity	Latency	Routing
Site to Site	Minimum of 10Gbps	<5ms latency RTT	*Layer 2 **Layer 3
Site to Witness	Minimum of 100Mbps	<200ms latency RTT	Layer 3
*Layer 2 with Multicast Recommended **Layer 3 with Multicast Supported			

Bandwidth Requirements Between Sites

Workloads are seldom all reads or writes, and normally include a general read to write ratio for each use case.

A good example of this would be a VDI workload. During peak utilization, VDI often behaves with a 70/30 write to read ratio. That is to say that 70% of the IO is due to write operations and 30% is due to read IO. As each solution has many factors, true ratios should be measured for each workload.

Using the general situation where a total IO profile requires 100,000 IOPS, of which 70% are write, and 30% are read, in a Stretched configuration, the write IO is what is sized against for inter-site bandwidth requirements.

With Stretched Clusters, read traffic is, by default, serviced by the site that the VM resides on. This concept is called Read Locality.

The required bandwidth between two data sites (B) is equal to Write bandwidth (Wb) * data multiplier (md) * resynchronization multiplier (mr):

$$B = Wb * md * mr$$

The data multiplier is comprised of overhead for Virtual SAN metadata traffic and miscellaneous related operations. VMware recommends a data multiplier of 1.4

The resynchronization multiplier is included to account for resynchronizing events. It is recommended to allocate bandwidth capacity on top of required bandwidth capacity for resynchronization events.

Making room for resynchronization traffic, an additional 25% is recommended.

Site to Site Examples

Workload 1

With an example workload of 10,000 writes per second to a workload on Virtual SAN, with a “typical” 4KB size write, that would require 40MB/s, or 320Mbps bandwidth.

$$B = 320 \text{ Mbps} * 1.4 * 1.25 = 560 \text{ Mbps.}$$

Including the Virtual SAN network requirements, the required bandwidth would be 560Mbps.

Workload 2

In another example, 30,000 writes per second, 4KB writes, would require 120MB/s, or 960Mbps bandwidth.

$$B = 960 \text{ Mbps} * 1.4 * 1.25 = 1680 \text{ Mbps or } \sim 1.7\text{Gbps}$$

The required bandwidth would be approximately 1.7Gbps.

Bandwidth Requirements Between Witness & Data Sites

Witness bandwidth isn't calculated in the same way as inter-site bandwidth requirements. Witnesses do not maintain VM data, but rather only component metadata.

It is important to remember that data is stored on Virtual SAN in the form of objects. Objects are comprised of 1 or more components of items such as:

- VM Home or namespace
- VM Swap object
- Virtual Disks
- Snapshots

Objects can be split into more than 1 component when the size is >255GB, and/or a Number of Stripes (stripe width) policy is applied. Additionally, the number of objects/components for a given Virtual Machine is multiplied when a Number of Failures to Tolerate (FTT) policy is applied for data protection and availability.

The required bandwidth between the Witness and each site is equal to $1138 B \times \text{Number of Components} / 5s$

$$1138 B \times \text{NumComp} / 5 \text{ seconds}$$

The 1138 B value comes from operations that occur when the Preferred Site goes offline, and the Secondary Site takes ownership of all of the components.

When the primary site goes offline, the secondary site becomes the master. The Witness sends updates to the new master, followed by the new master replying to the Witness as ownership is updated.

The 1138 B requirement for each component comes from a combination of a payload from the Witness to the backup agent, followed by metadata indicating that the Preferred Site has failed.

In the event of a Preferred Site failure, the link must be large enough to allow for the cluster ownership to change, as well ownership of all of the components within 5 seconds.

Witness to Site Examples

Workload 1

With a VM being comprised of

- 3 objects
 - VM namespace
 - vmdk (under 255GB)
 - vmSwap
- Failure to Tolerate of 1 (FTT=1)
- Stripe Width of 1

Approximately 166 VMs with the above configuration would require the Witness to contain 996 components.

To successfully satisfy the Witness bandwidth requirements for a total of 1,000 components on Virtual SAN, the following calculation can be used:

Converting Bytes (B) to Bits (b), multiply by 8

$$B = 1138 B * 8 * 1,000 / 5s = 1,820,800 \text{ Bits per second} = 1.82 \text{ Mbps}$$

VMware recommends adding a 10% safety margin and round up.

$$B + 10\% = 1.82 \text{ Mbps} + 182 \text{ Kbps} = 2.00 \text{ Mbps}$$

With the 10% buffer included, a rule of thumb can be stated that for every 1,000 components, 2 Mbps is appropriate.

Workload 2

With a VM being comprised of

- 3 objects
 - VM namespace
 - vmdk (under 255GB)
 - vmSwap
- Failure to Tolerate of 1 (FTT=1)
- Stripe Width of 2

Approximately 1,500 VMs with the above configuration would require 18,000 components to be stored on the Witness.

To successfully satisfy the Witness bandwidth requirements for 18,000 components on Virtual SAN, the resulting calculation is:

$$B = 1138 B * 8 * 18,000 / 5s = 32,774,400 \text{ Bits per second} = 32.78 \text{ Mbps}$$
$$B + 10\% = 32.78 \text{ Mbps} + 3.28 \text{ Mbps} = 36.05 \text{ Mbps}$$

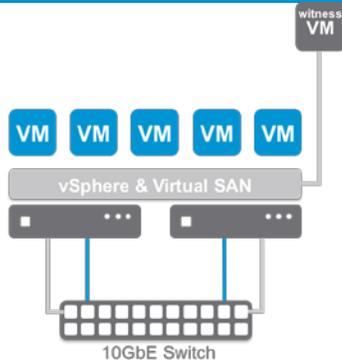
Using the general equation of 2Mbps for every 1,000 components, (NumComp/1000) X 2Mbps, it can be seen that 18,000 components does in fact require 36Mbps.

Witness Bandwidth for 2 Node Configurations

Remote Site Deployment

Virtual SAN introduced 2 Node support in version 6.1. This is a specialized use case of Stretched Clusters.

In cases where remote offices have a small complement of VMs, 2 Node configurations can be very cost effective.



2 Node Virtual SAN Configuration

Remote Site Example 1

Take the example of 25 VMs in a 2 Node configuration, each with a 1TB virtual disk protected at FTT=1 and a Stripe Width=1.

Each vmdk would be comprised of 8 components (vmdk and replica) and 2 components each for the VM namespace and swap file. The total number of components is 300 (12/VMx25VMs).

With 300 components, using the rule of thumb (300/1000 x 2Mbps), **600Kbps** of bandwidth is required.

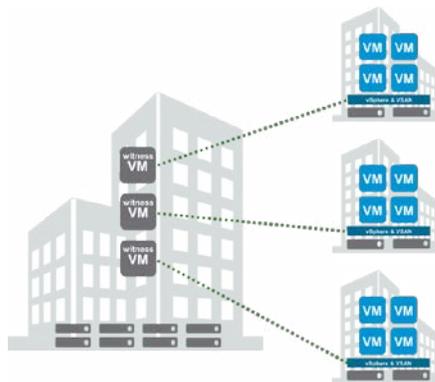
Remote Site Example 2

Take another example of 100 VMs on each host, of the same VM above, with 1TB virtual disk, FTT=1 & SW=1.

The total number of components would be 2,400. Using the rule of thumb (2,400/1000 x 2Mbps), **4.8Mbps** of bandwidth is required.

Multiple Remote Office Deployments

It is important to remember, when deploying 2 Node configurations to include enough bandwidth for each site.



Multiple 2 Node ROBO Configuration

The two examples would require a combined bandwidth of 5.4Mbps (600Kbps + 4.8Mbps)

Next Steps

Additional Documentation

For more information about VMware Virtual SAN, please visit the product pages at <http://www.vmware.com/products/virtual-san>

Below are some links to online documentation:

- [Product Documentation](#)
- [Design & Sizing Guide](#)
- [What's New](#)
- [Stretched Cluster Guide](#)
- [Stretched Cluster Performance and Best Practices](#)
- [Virtual SAN Community](#)
- [Support Knowledge base](#)

VMware vSphere and VMware vCenter Server Resources:

- [Product Overview](#)
- [Product Documentation](#)

VMware Contact Information

For additional information or to purchase VMware Virtual SAN, VMware's global network of solutions providers is ready to assist. If you would like to contact VMware directly, you can reach a sales representative at 1-877-4VMWARE (650-475-5000 outside North America) or email sales@vmware.com. When emailing, please include the state, country, and company name from which you are inquiring.

About the Author

This technical white paper was put together using content from various resources from Virtual SAN Engineering and Virtual SAN Product Management.

Jase McCarty is a Sr. Technical Marketing Architect, working in the Storage and Availability Business Unit at VMware, Inc. He specializes in Enterprise Architectures, Software-Defined Storage, and Hyper-Converged Infrastructures.

Jase is an author of multiple books based on VMware technologies, and is a repeat recipient of the VMware vExpert designation. He is the owner and main author of virtualization blog jsemccarty.com. He can be found on Twitter at [@jsemccarty](https://twitter.com/jsemccarty).

