# VMware® Virtual SAN™ Hardware Guidance

**vm**ware®

**Table of Contents**

# Introduction

VMware® Virtual SAN™ is the industry's first scale-out, hypervisor-converged storage solution based on the industry-leading VMware vSphere® solution. Virtual SAN is a software-defined storage solution that enables great flexibility and vendor choice in hardware platform.

This document provides guidance regarding hardware decisions—based on creating Virtual SAN solutions using *VMware Compatibility Guide*–certified hardware—when designing and deploying Virtual SAN. These decisions include the selection of server form factor, SSD, HDD, storage controller, and networking components.

This document does not supersede the official hardware compatibility information found in the *VMware Compatibility Guide*, which is the single source for up-to-date Virtual SAN hardware-compatibility information and must be used for a list of officially supported hardware.

When designing a Virtual SAN cluster from a sum of *VMware Compatibility Guide*–certified vendor components, this guide should be used in combination with the *VMware Virtual SAN Design and Sizing Guide*; the Virtual SAN sizing tool, http://virtualsansizing.vmware.com; and other official Virtual SAN documentation from VMware Technical Marketing and VMware Technical Publications.

# Server Form Factors

The client server model has evolved over the years, with a prevalent data center server form factor evolving, from standalone and rackmount servers to multinode single-chassis server form factors that enable a converged infrastructure. Each of these form factors can be used to build a Virtual SAN solution, and each has different characteristics that affect the final hypervisor-converged solution design.

## Rackmount

Rackmount servers are an ideal form factor for Virtual SAN node, due to the number of available drive slots. Multiple drive slots enable servers to scale up performance via multiple flash devices and to scale up capacity via additional magnetic HDDs.

## Blade

Blade servers are supported, but they are not an optimal form factor for Virtual SAN 5.5 for two reasons. The first one regards the limited number of local drive slots in the typical blade server, usually two for half-height and half-width blades and four for full-height and full-width blades. The second is due to the fact that Virtual SAN 5.5 does not support disk enclosures external to servers. Blade servers can be used within Virtual SAN as converged compute and storage nodes for use cases that require limited capacity.

Another approach provides compute-only nodes via blade servers and compute and storage nodes via rackmount servers within a Virtual SAN cluster. Although this is a supported configuration, strict design and sizing considerations must be adhered to if considering this approach. Considerations for using this approach include properly sizing a cluster for the number of objects and corresponding components within the cluster. Compute-only nodes count against a Virtual SAN per-host 3,000-component limit. Consult the *VMware Virtual SAN Design and Sizing Guide* and the VMware Virtual SAN sizing tool for additional background and guidance on designing Virtual SAN clusters that include compute-only nodes.

### Converged Rackmount

New converged rackmount server form factors deliver a middle ground between the scalability of standard rackmount servers and the rack density of blade servers. Converged rackmount servers provide multiple hardware nodes in a 2U or 4U enclosure. Depending on the specific configuration, they can be an ideal hardware platform for a Virtual SAN solution.

As a starting point, for more information on suitable Virtual SAN server form factors, consult the *VMware Compatibility Guide* for supported servers.

# Server Boot Device

Because Virtual SAN takes advantage of local server storage slots for flash devices and HDD, the selection of a boot device is a critical decision point for your vSphere host.

vSphere 5.5 Update 1 with Virtual SAN 5.5 can support USB, SD cards, or HDD as the host boot device. Booting from SAN LUNs is also supported. SATADOM as a boot device is not supported.

Virtual SAN trace log devices are stored within either a scratch partition on the local boot HDD or a scratch partition on an available VMware vSphere VMFS formatted HDD, VMFS formatted LUN, or ramdisk. The following are ways that the selected boot device impacts how Virtual SAN trace logs are stored:

• For hosts with 512GB or less memory, SD cards and USB are supported as a boot device. In this scenario, the core dump partition is used for Virtual SAN traces. The core dump partition is configured on ramdisk in a compressed format. If a core dump occurs, 2.2GB of the USB is utilized. Net dump can be used in conjunction with this option.

• The minimum supported SD card or USB boot device size is 4GB; 8GB or larger is recommended.

• Virtual SAN trace logs present in ramdisk are not persisted during reboot or host failure. Consider this factor when selecting a boot device.

• For hosts configured with more than 512GB memory, either a local boot HDD or LUN or a scratch partition available on a VMFS formatted HDD or LUN is required for the Virtual SAN trace scratch partition.

*NOTE: The Virtual SAN trace file ramdisk size should not be increased over the default 300MB.*

# Flash Devices

Flash devices are used as a nonvolatile write buffer and read cache for Virtual SAN I/O. The performance class of a flash device has a significant impact on the overall performance of a Virtual SAN cluster. Flash devices using either a SAS, SATA, or PCIe interface are supported.

### Flash Performance

Flash devices in the *VMware Compatibility Guide* for Virtual SAN are categorized into classes, based on write performance. All flash devices are not created equal, and the class of the flash device can greatly affect the performance of a Virtual SAN cluster. The *VMware Compatibility Guide* specifies the following designated flash device classes:

• Class A: 2,500–5,000 writes per second
• Class B: 5,000–10,000 writes per second
• Class C: 10,000–20,000 writes per second
• Class D: 20,000–30,000 writes per second
• Class E: 30,000+ writes per second

For optimal performance, VMware recommends utilizing a flash device class that meets workload performance requirements.

A general guideline is to configure 10 percent of projected used HDD capacity before the *failures to tolerate* policy is applied. For example, if there are 100 virtual machines with 50GB VMDK, of which 50 percent is used, 250GB of flash is needed.

Specific flash device sizing is based on a particular use case. Consult the *VMware Virtual SAN Design and Sizing Guide* and the Virtual SAN sizing tool, http://virtualsansizing.vmware.com, for assistance in sizing a Virtual SAN solution for a specific use case.

## Flash Device Reliability

Flash device write metrics are the primary criteria used by flash device vendors to gauge reliability. VMware has validated that all flash devices within the *VMware Compatibility Guide* provide enterprise-level reliability by meeting the following minimum endurance metrics:

**Endurance requirements for SAS and SATA flash devices**

• The drive must support at least 10 drive writes per day (DWPD).

• *And* the drive must support random write endurance up to 3.5PB on 8KB transfer size per NAND module *or* up to 2.5PB on 4KB transfer size per NAND module.

**Endurance requirements for PCIe flash devices**

• The drive must support at least 10 drive writes per day (DWPD).

• *Or* the drive must support random write endurance up to 3.5PB on 8KB transfer size per NAND module *or* up to 2.5PB on 4KB transfer size per NAND module.

# Magnetic Hard Disk Drives

Magnetic hard disk drives (HDDs) store data within a Virtual SAN cluster. Data is stored on an HDD when destaged from the flash acceleration layer. HDDs are also used for read cache misses. The recommended practice for Virtual SAN workload sizing is to size an active working set to be contained within the flash device read cache, to limit the amount of read cache misses. When this guideline is adhered to, HDD speed mainly impacts write destaging of cold blocks. SAS, NL-SAS, and SATA drives are supported.

## HDD Performance

As a general rule, VMware recommends 7200RPM drives for capacity and 10000RPM drives for performance. Although *VMware Compatibility Guide*–supported 15000RPM drives can be used, 10000RPM drives typically meet the best balance of performance, capacity, and cost.

If optimal performance is a design goal, VMware recommends utilizing SAS or NL-SAS drives in preference to SATA. Even at identical HDD rotational speeds of 7200RPM for NL-SAS and SATA drives, Virtual SAN clusters optimize performance utilizing the NL-SAS interface, rather than SATA, due to the advanced command queuing that is available only through SAS and NL-SAS interfaces. VMware internal testing has observed as much as a 15 percent performance increase when using NL-SAS drives as compared to SATA drives of the same size and rotational speed.

### Number of Spindles

It is a common storage practice to enhance performance by increasing the number of spindles. Although the flash acceleration layer is the most impactful factor for performance in a Virtual SAN cluster, the number of spindles plays a role in optimizing performance for workloads that utilize working sets larger than the available flash acceleration layer within a Virtual SAN cluster.

In a Virtual SAN solution, scaling up performance within a single node is most effective through the addition of disk groups, which provides more flash device acceleration as well as HDD spindles. See the Virtual SAN sizing tool to assist in configuring your Virtual SAN solution.

### HDD Reliability

VMware requires that any magnetic HDD within the *VMware Compatibility Guide* meet the following minimum endurance metrics:

• The drive must have a minimum useful life of five years.

• The drive must meet an annualized failure rate (AFR) of no more than 0.73 percent *or* mean time between failures (MTBF) of 1,200,000 hours.

# Storage Controllers

Virtual SAN supports storage controllers in two modes: pass-through and RAID 0. One of the major considerations when choosing a storage controller for Virtual SAN is whether it supports pass-through mode, RAID 0 mode, or both.

### Storage Controller Pass-Through Mode

Storage controller HBAs and RAID adapters can support a mode of operation commonly known as pass-through mode, where the vSphere hypervisor is given direct access to the underlying drives. For storage controller HBAs, this is also known as initiator target mode (IT mode); for RAID controllers that support pass-through, this is known as JBOD mode. Regardless of the nomenclature, the end result gives Virtual SAN complete control of the local SSDs and HDDs attached to the storage controller.

### Storage Controller RAID 0 Mode

For storage controller RAID adapters that do not support pass-through mode, Virtual SAN fully supports RAID 0 mode. RAID 0 mode is implemented by creating a single-drive RAID 0 set via the storage controller software, utilizing all SSDs and HDDs within the Virtual SAN cluster. The single-drive RAID 0 sets are then presented to Virtual SAN. For Virtual SAN to differentiate between the SSD and HDD RAID 0 sets, the single-drive SSD RAID 0 sets must be tagged within the vSphere esxcli. See vSphere 5.5 Update 1 documentation and VMware Knowledge Base article 2013188, http://kb.vmware.com/kb/2013188, for more information on this procedure.

When RAID 0 mode is utilized, the following occurs:

• Virtual SAN does not manage hot-plug capabilities of drives.

• Hot-plug is managed by the storage controller firmware.

### Storage Controller Feature Types

When selecting a storage controller for Virtual SAN, four storage controller feature types are listed within the *VMware Compatibility Guide*. The following are the controller feature names listed in the guide, along with their corresponding support for Virtual SAN in pass-through or RAID 0 mode:

• Virtual SAN SAS – IT mode controller, supports pass-through

• Virtual SAN SATA – IT mode controller, supports pass-through

• Virtual SAN RAID 0 – RAID controller, supports RAID 0 mode

• Virtual SAN Pass-Through – RAID controller, supports pass-through (JBOD mode)

### Pass-Through Versus RAID 0 Mode

Pass-through and RAID 0 performance for the same storage controller should be similar in most instances within a Virtual SAN solution. The supported storage controller queue depth is the most important factor in determining storage controller performance in a Virtual SAN solution. When utilizing RAID 0 mode, the storage controller cache should be disabled—this is configurable on some but not all storage controllers—to ensure that it does not conflict with the cache of the SSD drives controlled by Virtual SAN. When the storage controller cache cannot be completely disabled in a RAID 0 configuration, configure it for 100 percent read cache, effectively disabling the write cache.

The main consideration when utilizing RAID 0 mode for storage controllers within Virtual SAN is the impact on the operational model. To manage the addition and removal of drives, RAID 0 mode requires interaction with the storage controller firmware through hardware vendor–specific software tools.

### Storage Controller Performance

Choices regarding storage controllers can have a significant impact on optimizing performance and capacity sizing within a Virtual SAN cluster. The following storage controller characteristics impact performance:

• Queue depth supported by the storage controller

• Storage controller interface speed

• Number of drives supported per storage controller

The most important performance factor regarding storage controllers in a Virtual SAN solution is the supported queue depth. VMware recommends storage controllers with a queue depth of greater than 256 for optimal Virtual SAN performance.

For optimal performance of storage controllers in RAID 0 mode, disable the write cache, disable read-ahead, and enable direct I/Os.

# Network

Virtual SAN utilizes a dedicated VMkernel port type and proprietary transport protocol for internode traffic. Virtual SAN supports both 1Gb Ethernet (GbE) and 10GbE for the transport network. If using 1GbE, it must be dedicated to Virtual SAN. 10GbE network adapters can be shared with other traffic types, utilizing VMware vSphere Network I/O Control to guarantee quality of service (QoS) for the multiple traffic types over the same uplink. Because vSphere Network I/O Control requires the use of a VMware vSphere Distributed Switch™ (VDS), all editions of Virtual SAN include VDS capability. For optimal performance, VMware recommends the use of a dedicated storage network.

Hosts cannot be cross-connected to create a Virtual SAN cluster. A network switch is required because VMware ESXi™ hosts cannot pass VMkernel traffic. Virtual SAN requires Ethernet for network traffic; InfiniBand is not supported.

### Multicast

Layer 2 multicast is required for VMkernel ports utilized by Virtual SAN. IGMP snooping configured with an IGMP snooping querier can be used to limit the physical switch ports participating in a multicast group to only Virtual SAN VMkernel port uplinks.

When multiple Virtual SAN clusters reside on the same layer 2 network, change the default multicast address within the additional Virtual SAN clusters to prevent multiple clusters from receiving all multicast streams. Consult VMware Knowledge Base article 2075451, http://kb.vmware.com/kb/2075451, for details on this procedure.

### Jumbo Frames

Jumbo frames are supported. For greenfield installations, they are recommended. For brownfield installations, the operational cost of enabling jumbo frames in an existing infrastructure should be weighed. VMware testing shows minimal performance throughput benefit when utilizing jumbo frames for the Virtual SAN transport network. Jumbo frames optimize CPU utilization in high-performance Virtual SAN clusters.

## Conclusion

VMware Virtual SAN is a groundbreaking storage solution that enables unprecedented hardware configuration flexibility through building an individual solution based on preferred server vendor components. The guidance provided in this document enables users to make the best choice regarding their particular storage needs for their software-defined datacenter based on VMware vSphere. When selecting hardware components for a Virtual SAN solution, users should always utilize the *VMware Compatibility Guide* as the definitive resource tool.

# Acknowledgments

I would like to thank Kiran Madnani of VMware Product Management, Rawlinson Rivera of VMware Storage and Availability Technical Marketing, Charu Chaubal of VMware Storage and Availability Technical Marketing, Duncan Epping of VMware R&D, Cormac Hogan of VMware Integration Engineering, and Simon Todd of VMware Global Support Services for reviewing this paper.

# About the Author

Wade Holmes, VCDX #15, CISSP, CCSK, is a senior technical marketing architect at VMware, currently focusing on software-defined storage reference architectures. Wade has more than 17 years of industry experience in the design and implementation of complex computing environments of all scopes and sizes. He has presented at many industry conferences and is a co-author of *VMware vCloud Architecture Toolkit*. Wade holds a bachelor's degree in information technology and a master's degree in information assurance.

• Follow Wade's blogs:
  • http://blogs.vmware.com/vsphere/storage
  • http://www.vwade.com
• Follow Wade on Twitter:
  • @wholmes

**vm**ware®