



# What's New in VMware® Virtual SAN™

TECHNICAL WHITE PAPER  
V1.0/FEBRUARY 2014 UPDATE

**Table of Contents**

- 1. Introduction ..... 4
  - 1.1 Software-Defined Datacenter ..... 4
  - 1.2 Software-Defined Storage ..... 4
  - 1.3 VMware Virtual SAN ..... 4
- 2. Requirements ..... 5
  - 2.1 vSphere Requirements ..... 5
    - 2.1.1 vCenter Server ..... 5
    - 2.1.2 vSphere ..... 5
  - 2.2 Storage Requirements ..... 5
    - 2.2.1 Disk Controllers ..... 5
    - 2.2.2 Hard Disk Drives ..... 5
    - 2.2.3 Flash-Based Devices ..... 5
  - 2.3 Network Requirements ..... 5
    - 2.3.1 Network Interface Cards ..... 5
    - 2.3.2 Supported Virtual Switch Types ..... 6
    - 2.3.3 VMkernel Network ..... 6
- 3. Install and Configure ..... 7
  - 3.1 Creating a Virtual SAN Cluster ..... 7
    - 3.1.1 Manually Add Disks to Disk Groups ..... 7
    - 3.1.2 Automatic Creation of Disk Groups ..... 7
    - 3.1.3 Virtual SAN Cluster Creation Example ..... 7
  - 3.2 The Virtual SAN Shared Datastore ..... 8
    - 3.2.1 Virtual SAN Datastore Properties ..... 8
  - 3.3 Defining Virtual Machine Requirements ..... 8
  - 4.1 Distributed RAID ..... 9
  - 4.2 Witnesses and Replicas ..... 9
  - 4.3 Flash-Based Devices in Virtual SAN ..... 9
    - 4.3.1 Read Cache ..... 9
    - 4.3.2 Write Cache (Write Buffer) ..... 9
- 5. Storage Policy Based Management ..... 10
  - 5.1 Virtual SAN Capabilities ..... 11
    - 5.1.1 Number of Failures to Tolerate ..... 11
    - 5.1.2 Number of Disk Stripes per Object ..... 11
    - 5.1.3 Flash Read Cache Reservation ..... 12

- 5.1.4 Object Space Reservation ..... 12
- 5.1.5 Force Provisioning ..... 12
- 5.2 Witness Example .....13
- 5.3 Virtual Machine Storage Policies.....13
  - 5.3.1 Enabling Virtual Machine Storage Policies .....14
  - 5.3.2 Creating Virtual Machine Storage Policies.....14
  - 5.3.3 Assigning a Virtual Machine Storage Policy During  
Virtual Machine Provisioning.....15
  - 5.3.4 Virtual Machine Objects.....15
  - 5.3.5 Matching Resource .....15
  - 5.3.6 Compliance .....16
- Conclusion ..... 16
- Acknowledgments..... 16
- About the Author .....17

# 1. Introduction

## 1.1 Software-Defined Datacenter

The annual VMware® user conference, VMworld®, introduced the vision of VMware for the software-defined data center (SDDC) in 2012. The SDDC is the VMware cloud architecture in which all pillars of the data center—including compute, storage, networks, and associated services—are virtualized. In this white paper, we look at one aspect of the VMware SDDC, the storage pillar. We specifically discuss how a new product, VMware Virtual SAN™, fits into this vision.

## 1.2 Software-Defined Storage

The VMware software-defined storage strategy focuses on a set of VMware initiatives regarding local storage, shared storage, and storage and data services. Software-defined storage is designed to provide storage services and service-level agreement (SLA) automation through a software layer on the hosts that integrates with and abstracts the underlying hardware. With software-defined storage, virtual machine storage requirements can be dynamically instantiated. There is no need to repurpose LUNs or volumes. Virtual machine workloads might change over time, and the underlying storage can be adapted to the workload at any time.

A key factor for software-defined storage is Storage Policy Based Management (SPBM), which is featured in the VMware vSphere® 5.5 release and can be considered the next generation of VMware vSphere Storage Profile features. vSphere Storage Profile was introduced with vSphere 5.0. In vSphere 5.5, an enhanced feature, virtual machine storage policies, was introduced.

SPBM is a critical component for VMware in implementing software-defined storage. Using SPBM and VMware vSphere APIs, the underlying storage technology provides vSphere administrators with an abstracted pool of storage space for virtual machine provisioning. The technology's various capabilities relate to performance, availability, and storage services such as replication. A vSphere administrator can then create a virtual machine storage policy using a subset of the capabilities required by the application running in the virtual machine.

At the time of deployment, the vSphere administrator selects the virtual machine storage policy appropriate for the needs of that virtual machine. SPBM pushes the requirements down to the storage layer. Datastores that provide the capabilities included in the virtual machine storage policy are made available for selection. So, based on storage policy requirements, the virtual machine is always instantiated on the appropriate underlying storage. If the virtual machine's workload changes over time, a new policy with updated requirements that reflect the new workload is applied.

## 1.3 VMware Virtual SAN

Virtual SAN is a new software-defined storage solution that is fully integrated with vSphere. Virtual SAN aggregates locally attached disks in a vSphere cluster to create a storage solution that rapidly can be provisioned from VMware vCenter™ during virtual machine provisioning operations. It is an example of a hypervisor-converged platform—that is, a solution in which storage and compute for virtual machines are combined into a single device, with storage's being provided within the hypervisor itself as opposed to via a storage virtual machine running alongside other virtual machines.

Virtual SAN is an object-based storage system designed to provide virtual machine-centric storage services and capabilities through a SPBM platform. SPBM and virtual machine storage policies are solutions designed to simplify virtual machine storage placement decisions for vSphere administrators.

Virtual SAN is fully integrated with core vSphere enterprise features such as VMware vSphere High Availability (vSphere HA), VMware vSphere Distributed Resource Scheduler™ (vSphere DRS), and VMware vSphere vMotion®. Its goal is to provide both high availability and scale-out storage functionality. It also can be considered in the context of quality of service (QoS) because virtual machine storage policies can be created to define the levels of performance and availability required on a per-virtual machine basis.

## 2. Requirements

### 2.1 vSphere Requirements

#### 2.1.1 vCenter Server

Virtual SAN requires VMware vCenter Server™ 5.5 Update 1. Both the Microsoft Windows version of vCenter Server and the VMware vCenter Server Appliance™ can manage Virtual SAN. Virtual SAN is configurable and monitored from only VMware vSphere Web Client.

#### 2.1.2 vSphere

Virtual SAN requires three or more vSphere hosts to form a supported cluster in which each host contributes local storage. The minimum, three-host, configuration enables the cluster to meet the lowest availability requirement of tolerating at least one host, disk, or network failure. The vSphere hosts require vSphere version 5.5 or later.

### 2.2 Storage Requirements

#### 2.2.1 Disk Controllers

Each vSphere host that contributes storage to the Virtual SAN cluster requires a disk controller. This can be a SAS or SATA host bus adapter (HBA) or a RAID controller. However, the RAID controller must function in one of two modes:

- Pass-through mode
- RAID 0 mode

Pass-through mode, commonly referred to as JBOD or HBA mode, is the preferred configuration for Virtual SAN because it enables Virtual SAN to manage the RAID configuration settings for storage policy attributes based on availability and performance requirements that are defined on a virtual machine.

For a list of the latest Virtual SAN certified hardware and supported controllers, check the *VMware Compatibility Guide* for the latest information: <http://www.vmware.com/resources/compatibility/search.php>

#### 2.2.2 Hard Disk Drives

Each vSphere host must have at least one SAS, near-line SAS (NL-SAS), or SATA magnetic hard-disk drive (HDD) to participate in the Virtual SAN cluster. HDDs account for the storage capacity of the Virtual SAN shared datastore. Additional magnetic disks increase the overall capacity and can also improve virtual machine performance, because the virtual machine storage objects might be striped across multiple spindles. This topic is covered in greater detail in the “Virtual Machine Storage Policies” section of this paper.

#### 2.2.3 Flash-Based Devices

Each vSphere host must have at least one flash-based device—SAS, SATA, or PCI Express SSD—to participate in the Virtual SAN cluster. Flash-based devices provide both a write buffer and a read cache. The larger the flash-based device capacity per host, the larger the number of I/Os that can be cached and the greater the performance results that can be achieved.

*NOTE: Flash-based devices do not contribute to the overall size of the distributed Virtual SAN shared datastore. They count only toward the capacity of the Virtual SAN caching tier.*

### 2.3 Network Requirements

#### 2.3.1 Network Interface Cards

Each vSphere host must have at least one network adapter. It must be 1Gb Ethernet or 10Gb Ethernet capable, but VMware recommends 10Gb. For redundancy, a team of network adapters can be configured on a per-host basis. VMware considers this to be a best practice but not necessary in building a fully functional Virtual SAN cluster.

### 2.3.2 Supported Virtual Switch Types

Virtual SAN is supported on both the VMware vSphere Distributed Switch™ (VDS) and the vSphere standard switch (VSS). No other virtual switch types are supported in the initial release.

### 2.3.3 VMkernel Network

On each vSphere host, a VMkernel port for Virtual SAN communication must be created. A new VMkernel virtual adapter type has been added to vSphere 5.5 for Virtual SAN. The VMkernel port is labeled **Virtual SAN traffic**.

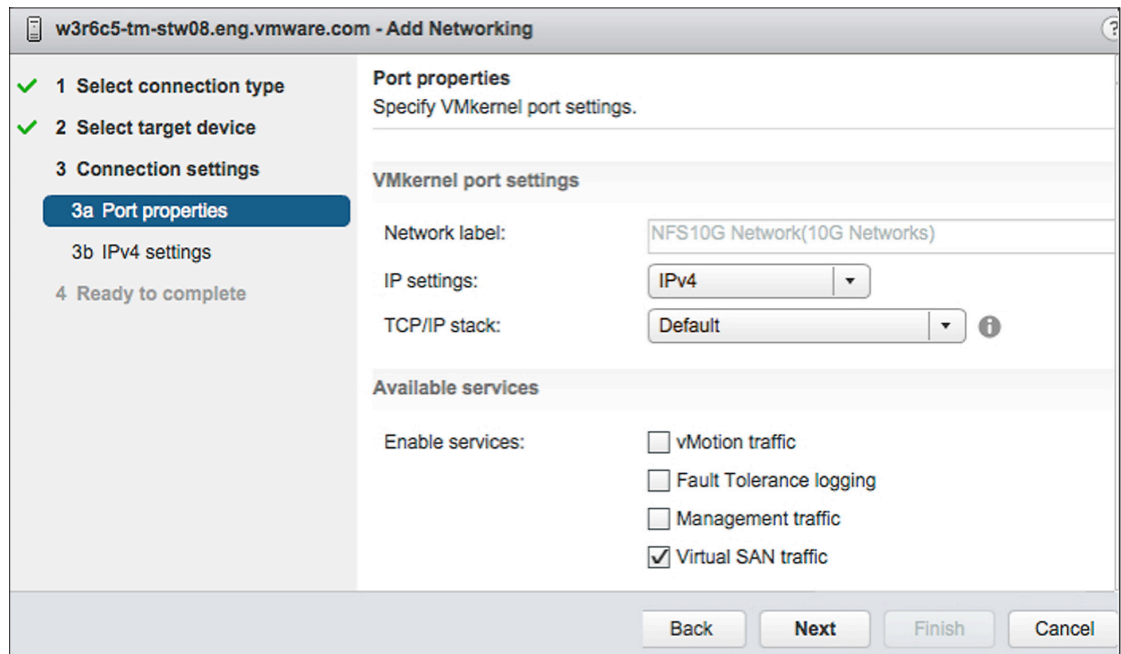


Figure 1. Virtual SAN VMkernel Adapter Type

This new interface is used for host intracluster communications as well as for read and write operations whenever a vSphere host in the cluster is the owner of a particular virtual machine but the actual data blocks making up that virtual machine's objects are located on a remote host in the cluster. In this case, I/O must traverse the network configured between the hosts in the cluster. If this interface is created on a VDS, the VMware vSphere Network I/O Control feature can be used to set shares or reservations for the Virtual SAN traffic.

## 3. Install and Configure

### 3.1 Creating a Virtual SAN Cluster

The process and procedures for creating a Virtual SAN cluster are identical to those that vSphere administrators use to set up other vSphere enterprise cluster features such as vSphere DRS and vSphere HA. When a cluster object is created in the vCenter Server inventory, users can choose either to enable the Virtual SAN cluster feature and then add the hosts to the cluster or to add the hosts first and then enable Virtual SAN.

When the Virtual SAN feature is enabled, the option of how to add disks to storage is displayed, asking the vSphere administrator to choose the manual or automatic mode. This gives the vSphere administrator the choice of how to claim disks in the cluster.

- Automatic: Enable Virtual SAN to discover all of the local disks on the hosts and automatically add the disks to the Virtual SAN shared datastore.
- Manual: Manually select the disks to add to the Virtual SAN shared datastore.

#### 3.1.1 Manually Add Disks to Disk Groups

When the **Manual** add-disk-to-storage option is selected, the Virtual SAN cluster is still formed, but the Virtual SAN shared datastore is initially 0 bytes in size. The vSphere administrator must manually select the disk to create disk groups on a per-host basis and add at least one magnetic disk and one flash-based device to each disk group. Each group can contain only one flash-based device and a maximum of seven magnetic disks.

Each vSphere host in the cluster can be configured with as many as five disk groups, each containing one flash-based device and one to seven magnetic disks. After each disk group has been created on a per-host basis, the size of the Virtual SAN shared datastore grows according to the number of disks per disk group, the number of disk groups, and the size of the magnetic disks added.

*NOTE: The flash-based devices function as read caches and write buffers and are not included in the capacity of the Virtual SAN shared datastore.*

#### 3.1.2 Automatic Creation of Disk Groups

When the **Automatic** add-disk-to-storage option is selected, the system automatically discovers all local magnetic disks and flash-based devices on each host and builds the disk groups on every host in the cluster. All hosts with valid storage have one or multiple disk groups containing their local magnetic disks and flash-based devices.

Finally, after this has been completed, the Virtual SAN shared datastore is created; its size reflects the capacity of all magnetic disk groups across all hosts in the cluster, except for some allotted to metadata overhead.

*NOTE: For metadata overhead, we calculate about 1GB per drive as a general guideline.*

vSphere hosts that are part of the cluster and do not have valid storage or locally attached storage can still access the Virtual SAN shared datastore. This is a very advantageous feature of Virtual SAN, because a cluster now can be scaled for compute requirements in addition to storage requirements.

#### 3.1.3 Virtual SAN Cluster Creation Example

After the vSphere host network and storage configuration requirements have been met, the creation of the Virtual SAN cluster can be started. As previously mentioned, the configuration is identical to that of a vSphere HA or vSphere DRS cluster, and it is user interface (UI) driven.

To create a Virtual SAN cluster, create a cluster object in the vCenter inventory, add the vSphere hosts to the cluster object that you intend to include in the cluster, and enable Virtual SAN on the cluster.

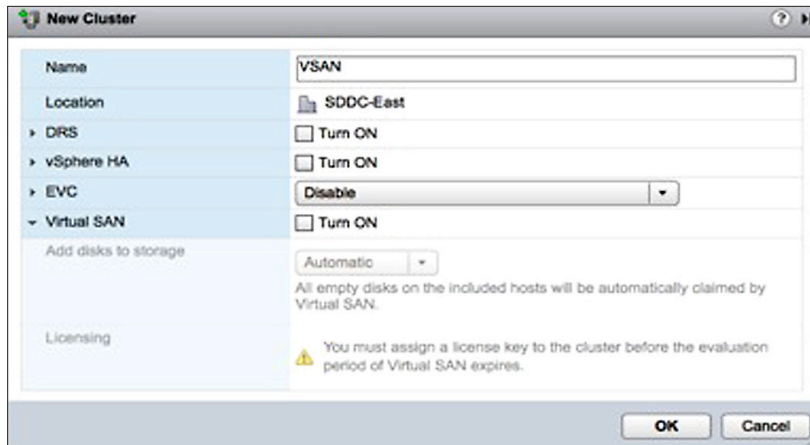


Figure 2. Create a New Virtual SAN Cluster

## 3.2 The Virtual SAN Shared Datastore

### 3.2.1 Virtual SAN Datastore Properties

The size and capacity of the Virtual SAN shared datastore are dictated by the number of magnetic disks per disk group in a vSphere host and by the number of vSphere hosts in the cluster. For example, if a cluster is composed of eight vSphere hosts, where each host contains one disk group composed of seven magnetic disks of 2TB in size each, the total raw capacity of the Virtual SAN shared datastore is 111.9TB after subtracting the metadata overhead capacity.

- Formula: Eight (8) vSphere hosts x one (1) disk group x seven (7) magnetic disks x 2TB = 112TB of raw capacity
- 112TB raw capacity – 56GB metadata overhead = 111.9TB of usable raw capacity

After the Virtual SAN shared datastore has been formed, a number of datastore capabilities are surfaced up to vCenter Server. These capabilities are based on storage capacity, performance, and availability requirements and are discussed in greater detail in the “Storage Policy Based Management” section of this paper. The essential point is that they can be used to create a policy that defines the storage requirements of a virtual machine.

These storage capabilities enable the vSphere administrator to create virtual machine storage policies that specify storage service requirements that must be satisfied by the storage system during virtual machine provisioning operations. This simplifies the virtual machine provisioning operations process by empowering the vSphere administrator to easily select the correct storage for virtual machines.

### 3.3 Defining Virtual Machine Requirements

When the Virtual SAN cluster is created, the shared Virtual SAN datastore—which has a set of capabilities that are surfaced up to vCenter—is also created.

When a vSphere administrator begins to design a virtual machine, that design is influenced by the application it will be hosting. This application might potentially have many sets of requirements, including storage requirements.

The vSphere administrator uses a virtual machine storage policy to specify and contain the application's storage requirements in the form of storage capabilities that will be attached to the virtual machine hosting the application; the specific storage requirements will be based on capabilities surfaced by the storage system. In effect, the storage system provides the capabilities, and virtual machines consume them via requirements placed in the virtual machine storage policy.



## 4.1 Distributed RAID

In additional storage environments, redundant array of independent disks (RAID) refers to disk redundancy inside the storage chassis to withstand the failure of one or more disk drives.

Virtual SAN uses the concept of distributed RAID, by which a vSphere cluster can contend with the failure of a vSphere host, or of a component within a host—for example, magnetic disks, flash-based devices, and network interfaces—and continue to provide complete functionality for all virtual machines. Availability is defined on a per-virtual machine basis through the use of virtual machine storage policies.

vSphere administrators can specify the number of host component failures that a virtual machine can tolerate within the Virtual SAN cluster. If a vSphere administrator sets zero as the number of failures to tolerate in the virtual machine storage policy, one host or disk failure can impact the availability of the virtual machine.

Using virtual machine storage policies along with Virtual SAN distributed RAID architecture, virtual machines and copies of their contents are distributed across multiple vSphere hosts in the cluster. In this case, it is not necessary to migrate data from a failed node to a surviving host in the cluster in the event of a failure.

## 4.2 Witnesses and Replicas

Replicas are copies of the virtual machine storage objects that are instantiated when an availability capability is specified for the virtual machine. The availability capability dictates how many replicas are created. It enables virtual machines to continue running with a full complement of objects when there is a host, network, or disk failure in the cluster.

Witnesses are part of every storage object. They are components that contain metadata but not data. They act as tiebreakers when availability determinations are made in the Virtual SAN cluster. A witness consumes about 2MB of space for metadata on the Virtual SAN shared datastore.

*NOTE: For an object to be accessible in Virtual SAN during a failure, more than 50 percent of its components must be accessible.*

## 4.3 Flash-Based Devices in Virtual SAN

Flash-based devices serve two purposes in Virtual SAN. They are used to build the flash tier in the form of a read cache and a write buffer, which dramatically improves the performance of virtual machines. In some respects, Virtual SAN can be compared to a number of “hybrid” storage solutions on the market that also use a combination of flash-based devices and magnetic disk storage to boost the performance of the I/O and that have the ability to scale out based on low-cost magnetic disk storage.

### 4.3.1 Read Cache

The read cache keeps a cache of commonly accessed disk blocks. This reduces the I/O read latency in the event of a cache hit. The actual block that is read by the application running in the virtual machine might not be on the same vSphere host on which the virtual machine is running.

To handle this behavior, Virtual SAN distributes a directory of cached blocks between the vSphere hosts in the cluster. This enables a vSphere host to determine whether a remote host has data cached that is not in a local cache. If that is the case, the vSphere host can retrieve cached blocks from a remote host in the cluster over the Virtual SAN network. If the block is not in the cache on any Virtual SAN host, it is retrieved directly from the magnetic disks.

### 4.3.2 Write Cache (Write Buffer)

The write cache performs as a nonvolatile write buffer. The fact that Virtual SAN can use flash-based storage devices for writes also reduces the latency for write operations.

Because all the write operations go to flash storage, Virtual SAN ensures that there is a copy of the data elsewhere in the cluster. All virtual machines deployed onto Virtual SAN inherit the default availability policy settings, ensuring that at least one additional copy of the virtual machine data is available. This includes the write cache contents.

After writes have been initiated by the application running inside of the guest operating system (OS), they are sent in parallel to both the local write cache on the owning host and the write cache on the remote hosts. The write must be committed to the flash storage on both hosts before it is acknowledged.

This means that in the event of a host failure, a copy of the data exists on another flash device in the Virtual SAN cluster and no data loss will occur. The virtual machine accesses the replicated copy of the data on another host in the cluster via the Virtual SAN network.

## 5. Storage Policy Based Management

Storage Policy Based Management (SPBM) plays a major role in the policies and automation for the VMware software-defined storage strategy. Using virtual machine storage policies, administrators can specify a set of required storage capabilities for a virtual machine, or more specifically a set of requirements for the application running in the virtual machine.

This set of required storage capabilities is pushed down to the storage layer, which then checks where the storage objects for that virtual machine can be instantiated to match this set of requirements. For instance, are there enough disks in the cluster to comply with the stripe widths policy if this virtual machine requires it? Or are there enough hosts in the cluster to provide the number of failures to tolerate? If the Virtual SAN datastore is compatible with the capabilities placed in the virtual machine storage policies, it is said to be a matching resource and is highlighted as such in the provisioning wizard.

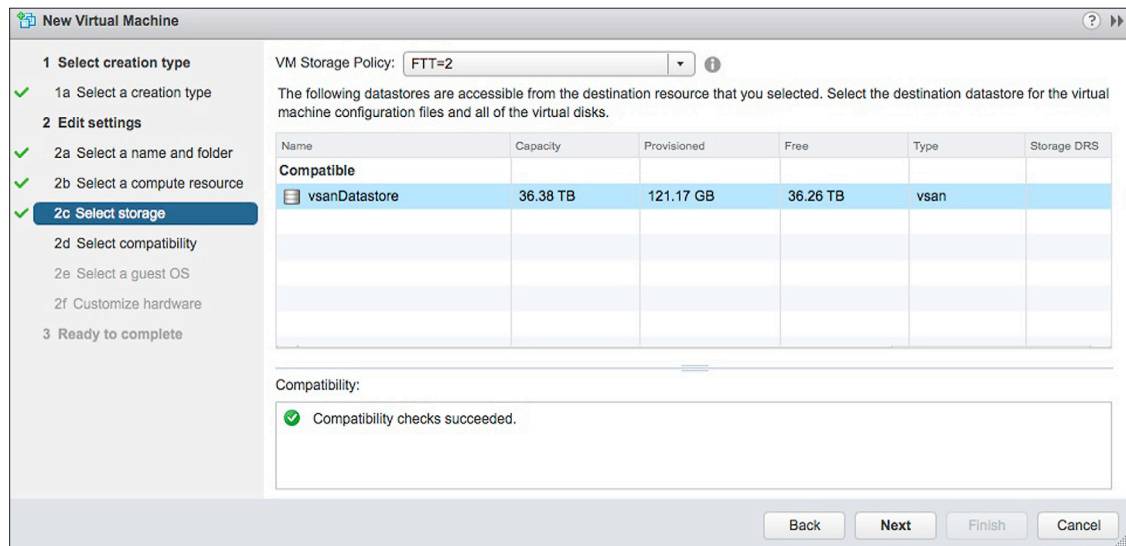


Figure 3. Virtual Machine Storage Policy Compatibility

Subsequently, when a virtual machine is deployed, if the requirements in the virtual machine storage policy attached to the virtual machine can't be satisfied by the Virtual SAN shared datastore, the datastore is said to be compliant from a storage perspective in its own summary window. If the Virtual SAN datastore is overcommitted or cannot meet the capability requirements, it might still be shown as a matching resource in the deployment wizard, but the provisioning task might fail.

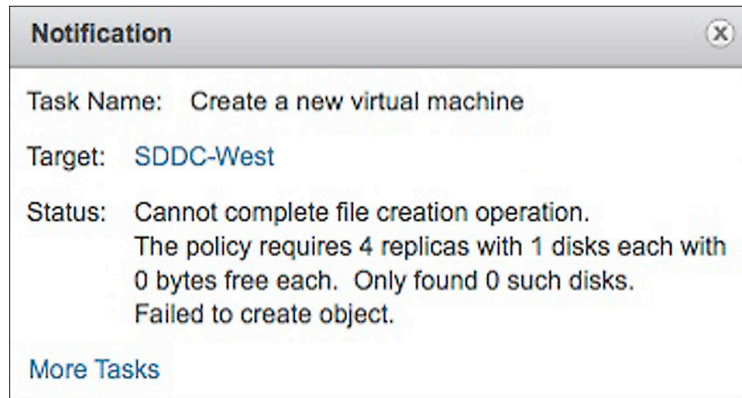


Figure 4. Failed Notification Event

Based on the required storage capabilities defined in the virtual machine storage policies, SPBM provides an automated, policy-driven mechanism for selecting appropriate datastores for virtual machines.

## 5.1 Virtual SAN Capabilities

This section examines the storage capabilities that can be selected and saved in virtual machine storage policies. These capabilities, which are surfaced by the Virtual SAN shared datastore when the cluster is successfully configured, highlight the availability, performance, and sizing requirements that can be demanded from the storage on a per-virtual machine basis.

### 5.1.1 Number of Failures to Tolerate

This property requires the storage object to tolerate a defined number of concurrent host, network, or disk failures in the cluster and still ensure availability of the objects.

If this property is populated, it specifies that a virtual machine configuration must withstand at least  $\text{NumberOfFailuresToTolerate} + 1$  replicas and might also contain an additional number of witness objects to ensure that the objects' data is available—to maintain quorum in split-brain cases—even in the presence of as many as  $\text{NumberOfFailuresToTolerate}$  concurrent host failures. Therefore, to tolerate  $n$  failures, at least  $(n + 1)$  copies of the object must exist and at least  $(2n + 1)$  hosts are required.

*NOTE: Any disk failure on a single host is treated as a "failure" for this metric. Therefore, the object cannot persist if there is a disk failure on host A and another disk failure on host B when  $\text{NumberOfFailuresToTolerate}$  is set to 1. Virtual SAN implements a default availability policy in all virtual machines. The default policy is equivalent to  $\text{NumberOfFailuresToTolerate}$  equals 1.*

### 5.1.2 Number of Disk Stripes per Object

This capability defines the number of physical disks across which each replica of a storage object is distributed. A value higher than 1 might result in better performance if read caching is not effective, but it will also result in a greater use of system resources.

To understand the impact of disk stripes, we examine this first in the context of write operations and then in the context of read operations. Because all writes go to the flash device write buffer, the value of an increased disk stripes number might not improve write performance. This is because there is no guarantee that the new stripe will use a different flash-based device. The new stripe might be in the same disk group and therefore use the same flash-based device. The only instance in which an increased disk stripes number might add value is where there are many writes to destage from flash-based devices to magnetic disks.

From a reading perspective, an increased disk stripes number helps when a user is experiencing many cache misses. Using the example of a virtual machine that consumes 2,000 read operations per second and is experiencing a cache-hit rate of 90 percent, there are 200 read operations per second that must be serviced from the magnetic disks. In this case, one magnetic disk might not be able to service those read operations, so an increase in the disk stripes number would help.

In general, the default disk stripes number of 1 should meet the requirements of most if not all virtual machine workloads. The disk stripes requirement should be changed only when a few high-performance virtual machines are running.

### 5.1.3 Flash Read Cache Reservation

This is the amount of flash capacity reserved on the flash-based devices as read cache for the storage objects. It is specified as a percentage of the logical size of the virtual machine disk storage object. This is expressed as a percentage value (%) with as many as four decimal places.

*NOTE: A fine, granular unit size is required so administrators can express sub-1 percent units. The following example is of a 1TB disk: If we limited the read cache reservation to 1-percent increments, it would result in cache reservations in increments of 10GB, which in most cases is far too large for a single virtual machine.*

*It is not required that a reservation be set for virtual machines to get flash read cache resources. VMware recommends not using reservations on a per-virtual machine basis. It is preferable that Virtual SAN distributed resource scheduler provide the read flash cache capacity based on demand.*

If cache space is not reserved, the Virtual SAN scheduler manages fair cache allocation. Cache reserved for one virtual machine is not available to other virtual machines. Unreserved cache is shared fairly between all objects.

Even if read performance is acceptable as is, cache can be added to prevent more reads from going to the magnetic disks and to decrease cache misses. This enables more writes to the magnetic disks.

Therefore, write performance can be improved indirectly by increasing flash read cache reservation. However, setting a reservation should be considered only if flushing data from flash devices to magnetic disk storage creates a bottleneck.

### 5.1.4 Object Space Reservation

This capability defines the percentage of the logical size of the storage object that should be reserved on magnetic disks during initialization. By default, provisioning on the Virtual SAN shared datastore is “thin.” The object space reservation is the amount of space to reserve on the Virtual SAN shared datastore, specified as a percentage of the virtual machine disk.

The value is the minimum amount of capacity that must be reserved for that particular disk. If object space reservation is set to 100 percent, all of the storage capacity of the virtual machine is offered up front.

*NOTE: If a virtual machine is provisioned and the “thick” disk format—either lazy zeroed or eager zeroed—is selected, this setting overrides the object space reservation setting in the virtual machine storage policy.*

### 5.1.5 Force Provisioning

If this option is enabled, the object will be provisioned even if the capabilities specified in the virtual machine storage policy cannot be satisfied with the resources available in the cluster at the time. Virtual SAN attempts to bring the object into compliance if and when resources become available.

If this parameter is set to a nonzero value, the object will be provisioned even if the policy specified in virtual machine storage policies is not satisfied by the datastore. However, if there is not enough space in the cluster to satisfy the reservation requirement of at least one replica, the provisioning will fail even if force provisioning is turned on. This option is disabled by default.

## 5.2 Witness Example

The following example of a witness involves the deployment of a virtual machine that has a disk stripes number of 1 and a NumberOfFailuresToTolerate number of 1. In this case, two replica copies of the virtual machine are created. Effectively, this is a RAID-1 virtual machine with two replicas.

However, with two replicas, there is no way to differentiate between a network partition and a host failure. Therefore, a third entity, the “witness,” is added to the configuration. For an object on Virtual SAN to be available, the following two conditions must be met:

- At least one replica must be intact for data access.
- More than 50 percent of all components must be available.

In the provided example, the object is accessible only when there is access to one replica copy and a witness or to two replica copies. That way, one part of the cluster, at most, can ever access an object in the event of a network partition.

## 5.3 Virtual Machine Storage Policies

Virtual machine storage policies in Virtual SAN work in a similar manner to vSphere Storage Profiles, introduced in vSphere 5.0, because a user builds a policy containing their virtual machine provisioning requirements. There is a major difference in the way virtual machine storage policies work in Virtual SAN as compared to how previous versions of vSphere Storage Profiles worked.

With the original version of vSphere Storage Profiles, the capabilities in the policy were used to select an appropriate datastore when provisioning the virtual machine. The new virtual machine storage policies not only select the appropriate datastore but also communicate to the underlying storage layer that there are certain availability and performance requirements for a specified virtual machine.

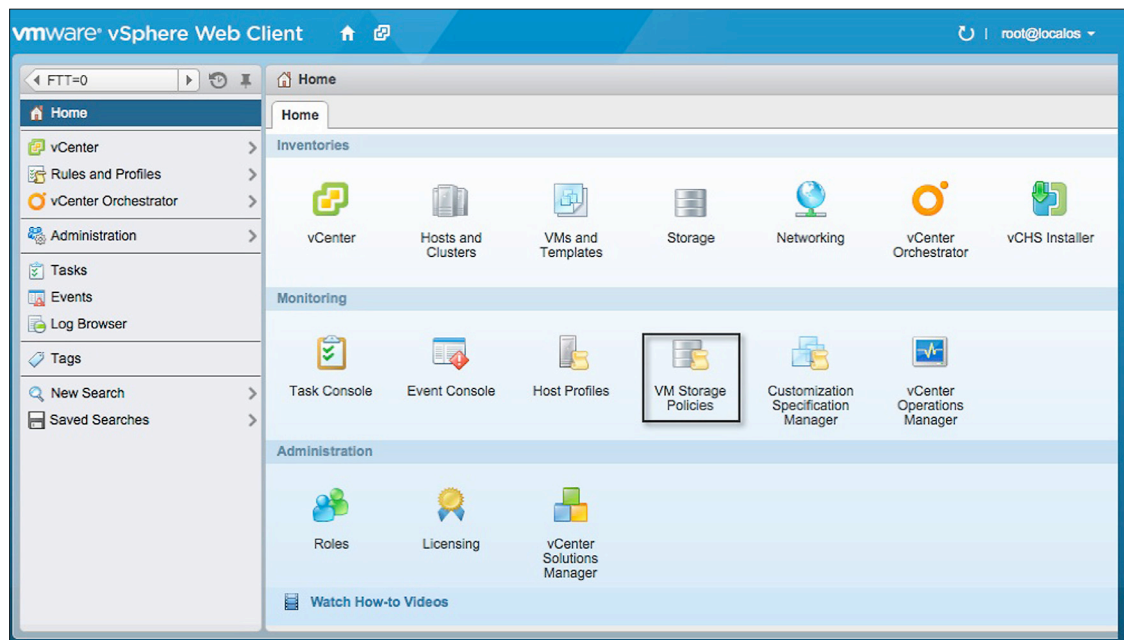


Figure 5. Virtual Machine Storage Policies

Therefore, the Virtual SAN datastore might be the destination datastore when that virtual machine is provisioned with a virtual machine storage policy, but other policy settings stipulate a requirement for the number of replicas of the virtual machine files for availability and might also contain a stripe-width requirement for performance.

### 5.3.1 Enabling Virtual Machine Storage Policies

Virtual machine storage policies are automatically enabled when Virtual SAN is configured. To enable virtual machine storage policies manually, navigate to the VMware vSphere Client™ **Home** position and then choose **Rules and Profiles**. The virtual machine storage policies section is located here. Click **VM Storage Policies**; a number of icons will appear. One of these enables virtual machine storage policies functionality, which can be enabled on a per-host or a per-cluster basis.

### 5.3.2 Creating Virtual Machine Storage Policies

After virtual machine storage policies have been enabled, a vSphere administrator can click another icon in this window to create individual levels. As already mentioned, a number of capabilities related to availability and performance are surfaced by vSphere APIs. At this point, the administrator must decide which of these capabilities are required from availability and performance perspectives for applications running inside the virtual machines.

For example, how many component failures—hosts, network, and disk drive—does the administrator require this virtual machine to tolerate while continuing to function? Also, is the application running in this virtual machine considered demanding from an IOPS perspective? If so, a reservation of flash read cache might be required as a capability to meet performance needs. Other considerations might involve whether the virtual machine should be thinly or thickly provisioned.

In this example, a virtual machine storage policy with two capabilities—availability and performance—is being created. NumberOfFailuresToTolerate defines the availability, and disk stripes defines the performance.

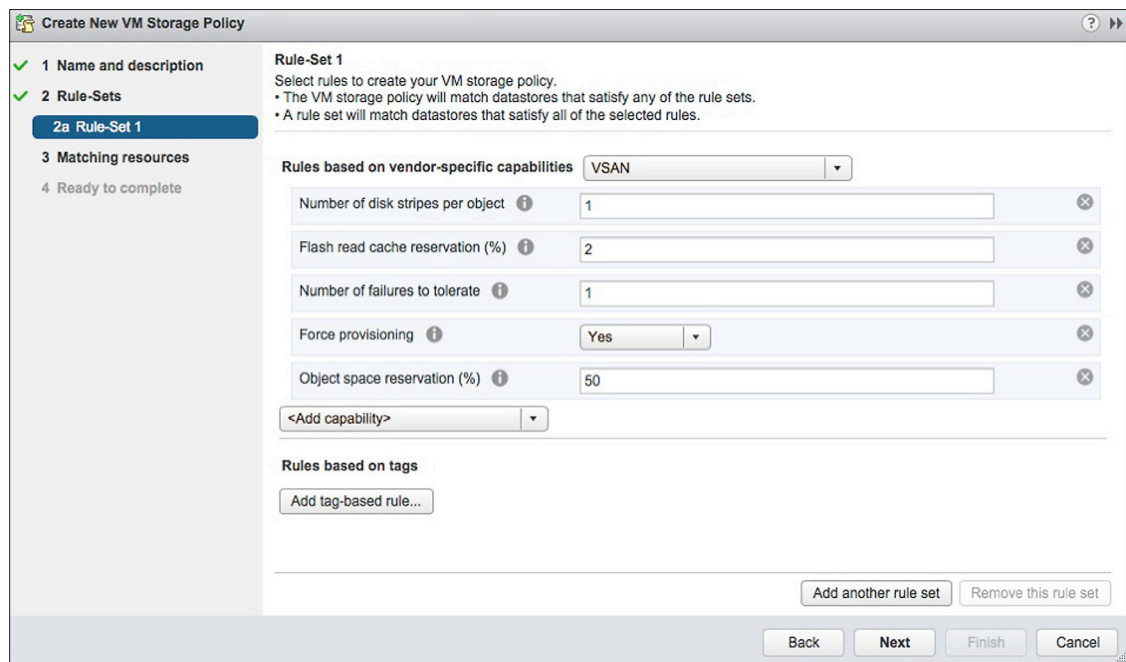


Figure 6. Create a New Virtual Machine Storage Policy

*NOTE: vSphere 5.5 also supports the use of tags for provisioning. Therefore, instead of using Virtual SAN datastore capabilities for the creation of a virtual machine storage policy, tag-based policies can be created. The use of tag-based policies is outside the scope of this white paper, but further information can be found in the vSphere storage documentation.*

### 5.3.3 Assigning a Virtual Machine Storage Policy During Virtual Machine Provisioning

The assignment of a virtual machine storage policy occurs during virtual machine provisioning. When the vSphere administrator must select a destination datastore, they select the appropriate level from the drop-down menu of available virtual machine storage policies. The datastores then are separated into compatible and incompatible datastores, enabling the vSphere administrator to choose the appropriate and correct placement of the virtual machine.

### 5.3.4 Virtual Machine Objects

Regarding the layout of objects on the Virtual SAN shared datastore, a virtual machine has various storage objects:

- VMDKs
- Virtual machine home
- Virtual machine swap
- Snapshot delta

The vSphere UI also enables administrators to interrogate the layout of a virtual machine object and to see where each component—stripes, replicas, and witnesses—of a storage object resides.

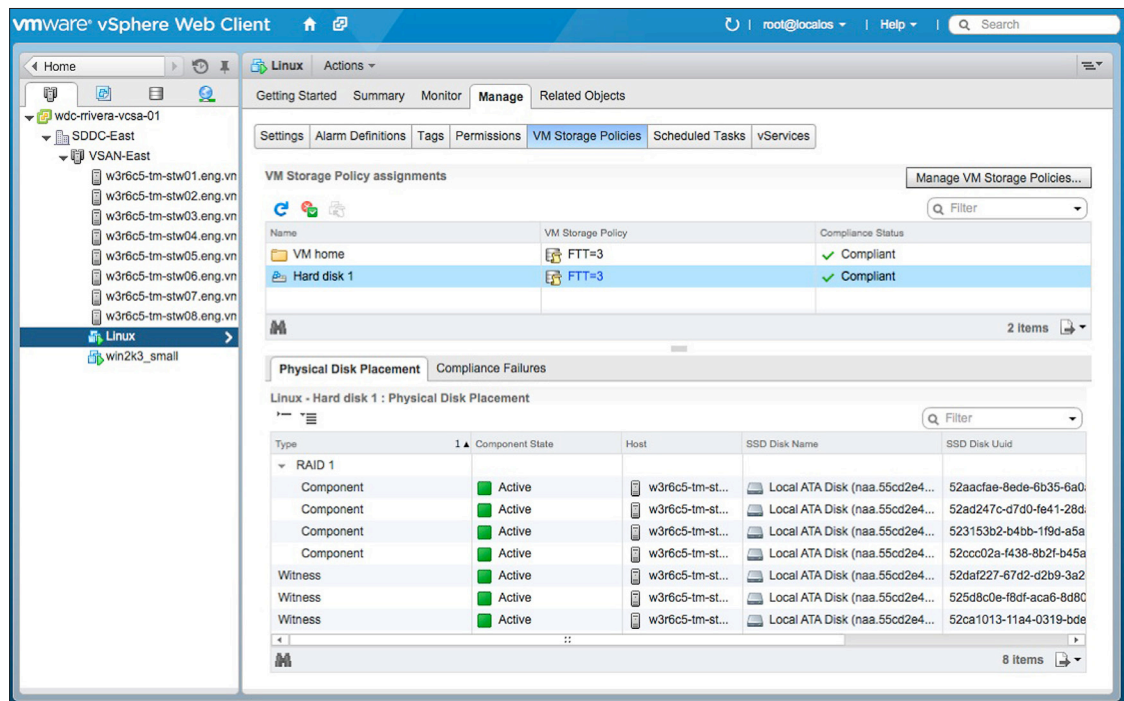


Figure 7. Storage Object Physical Mappings

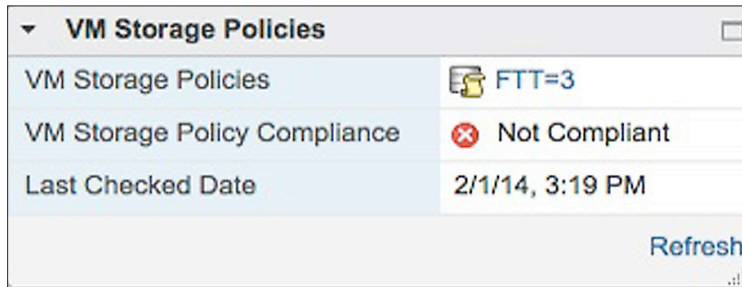
### 5.3.5 Matching Resource

When the virtual machine provisioning wizard reports that the Virtual SAN shared datastore is a matching resource for a virtual machine's storage policy, it guarantees that the requirements defined in the policy are compliant with the datastore.

This does not mean that the datastore will be able to meet those capabilities, and it's possible that the provisioning process will still fail. Compatibility does not guarantee that the datastore can actually meet the requirements. For example, if Virtual SAN capabilities are used, only Virtual SAN datastores will match, regardless of whether they have the resources to provision the virtual machine.

### 5.3.6 Compliance

The virtual machine **Summary** tab reveals the compliant state of a virtual machine. As long as Virtual SAN meets the capabilities defined in the virtual machine storage policy, the virtual machine is said to be **Compliant**. In the event of a component failure in the cluster, Virtual SAN might not be able to meet the policy requirements. In this case, the virtual machine is said to be **Not Compliant**. This **Compliant** or **Not Compliant** status is displayed in the virtual machine summary.



VM Storage Policies	
VM Storage Policies	FTT=3
VM Storage Policy Compliance	Not Compliant
Last Checked Date	2/1/14, 3:19 PM
Refresh	

Figure 8. Virtual Machine Summary - Not Compliant

If a failure persists for more than 60 minutes, Virtual SAN will rebuild the storage object and add new components to address the components of the storage objects that were impacted by the failure to bring the virtual machine back to a compliant state. This might involve the creation of new stripes, new replicas, or new witnesses. If the failure is addressed within this 60-minute window, no new components are created.

## Conclusion

VMware Virtual SAN is a hypervisor-converged storage platform that combines the compute and storage resources of vSphere hosts. It provides enterprise-class features and performance with a much simpler management experience for the user. It is a VMware-designed storage solution that makes software-defined storage a reality for VMware customers.

## Acknowledgments

I would like to thank Christian Dickmann and Christos Karamanolis of VMware R&D, whose deep knowledge and understanding of Virtual SAN was leveraged throughout this paper. I would also like to thank Cormac Hogan, senior architect in the Integration Engineering organization, for his contributions to this paper. Finally, I would like to thank Charu Chaubal, group manager of the Storage and Availability Technical Marketing team, for reviewing this paper.



## About the Author

Rawlinson Rivera is a senior architect in the Cloud Infrastructure Technical Marketing group at VMware. His focus is on storage virtualization, software-defined storage technologies, and the integration of VMware products and solutions with the OpenStack framework. Previously, he was an architect in the VMware Cloud Infrastructure and Management Professional Services organization, focused on vSphere and cloud enterprise architectures for VMware Fortune 100 and 500 customers.

Rawlinson is among the first VMware Certified Design Experts (VCDX#86) and is the author of multiple books based on VMware and other technologies.

Follow Rawlinson's blogs:

- <http://blogs.vmware.com/vsphere/storage>
- <http://www.punchingclouds.com>

Follow Rawlinson on Twitter:

- [@PunchingClouds](https://twitter.com/PunchingClouds)



**VMware, Inc.** 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 [www.vmware.com](http://www.vmware.com)

Copyright © 2014 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at <http://www.vmware.com/go/patents>. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies. Item No: VMW-TWP-VSAN-GA-edn-USLET-102

Docsource: OIC - 14VM004.10