VMware vSAN 6.6 Technical Overview

July 13, 2017

Table of Contents

1. Introduction

.

1.1 Enterprise-Class Storage for HCI

VMware vSAN™ is enterprise-class storage for hyper-converged infrastructure (HCI). Native to the VMware vSphere® hypervisor, vSAN delivers flash-optimized, secure storage. It utilizes commodity x86 server components to lower costs up to 50% versus traditional server and storage array architectures.

Seamless integration with vSphere and the VMware ecosystem makes it the ideal storage platform for business-critical applications, disaster recovery sites, remote office and branch office (ROBO) implementations, test and development environments, management clusters, security zones, and virtual desktop infrastructure (VDI). Today, customers of all industries and sizes trust vSAN to run their most important applications.

All-flash configurations provide the highest levels of performance with very low latencies for demanding business-critical applications. Space efficiency features such as inline deduplication and compression minimize capacity consumption, which reduces capital expenditures. Per-virtual machine (VM) storage policy-based management lowers operational expenditures by enabling administrators to manage performance, availability, and capacity consumption with ease and precision. This means no more LUN management.

Many deployment options are available for vSAN. These options range from single, 2-node clusters for small implementations to multiple clusters each with as many as 64 nodes—all centrally managed by vCenter Server. Stretched clusters can easily be configured to enable cross-site protection with no downtime for disaster avoidance and rapid, automated recovery from entire site failure.

vSAN 6.6, the sixth generation of vSAN, is designed to help customers modernize their infrastructure by addressing three key IT needs: higher security, lower costs, and faster performance. For example, vSAN 6.6 further lowers total cost of ownership by providing more resilient, economical stretched clusters that are easy to deploy and maintain.

The industry’s first native HCI encryption solution and a highly available control plane is delivered in vSAN 6.6 to help customers evolve without risk without sacrificing flash storage efficiencies. Operational costs are reduced with 1-click firmware and driver updates, as well as, proactive cloud health checks for real-time support.

vSAN has been enhanced with up to 50% greater flash performance enabling customers to scale to tomorrow’s IT demands. vSAN storage services are integrated with the Photon Platform with full API management to support container technologies and take advantage of DevOps efficiency.

2. Architecture

.

2.1 Servers with Local Storage

A wide variety of deployment and configuration options make vSAN a flexible and highly scalable HCI storage solution. A single vSAN cluster consists of any number of physical server hosts from two to 64. Organizations can start with what is needed today and implement a “just-in-time” provisioning model for additional compute and storage capacity. Additional hosts can easily be added to an existing cluster in a matter of minutes. This method of purchasing, expanding, and refreshing an IT infrastructure is more efficient and less costly than provisioning large, monolithic “blocks” of capacity every few years.

Each host contains flash devices (all flash configuration) or a combination of magnetic disks and flash devices (hybrid configuration) that contribute cache and capacity to the vSAN distributed datastore. Each host has one to five disk groups. Each disk group contains one cache device and one to seven capacity devices.

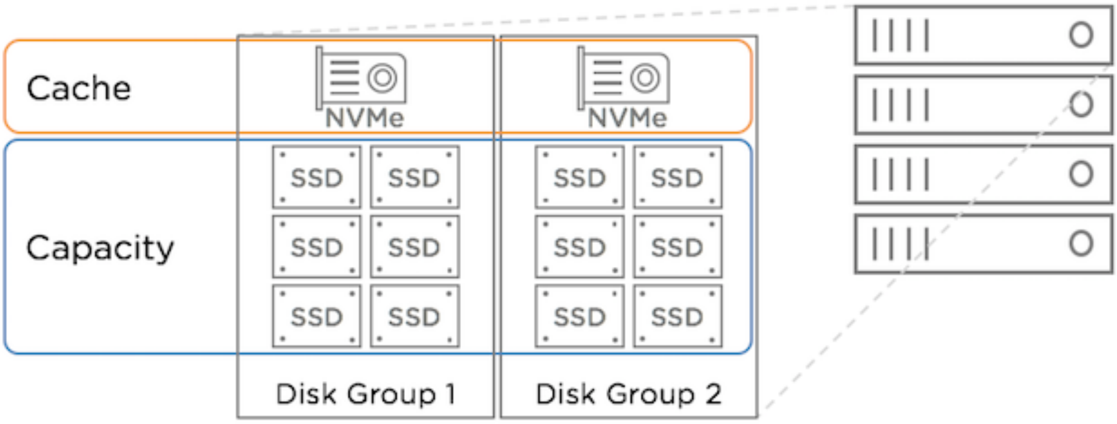


Figure 1. vSAN architecture

In all flash configurations, the flash devices in the cache tier are used for buffering writes. There is no need for read cache as performance from the capacity flash devices is more than sufficient. Two grades of flash devices are commonly used in an all flash vSAN configuration: Lower capacity, higher endurance devices for the cache layer and more cost effective, higher capacity, lower endurance devices for the capacity layer. Writes are performed at the cache layer and then de-staged to the capacity layer, as needed. This helps maintain performance while extending the usable life of the lower endurance flash devices in the capacity layer.

In hybrid configurations, one flash device and one or more magnetic drives are configured as a disk group. A disk group can have up to seven drives for capacity. One or more disk groups are used in a vSphere host depending on the number of flash devices and magnetic drives contained in the host. Flash devices serve as read cache and write buffer for the vSAN datastore while magnetic drives make up the capacity of the datastore. vSAN will uses 70% of the flash capacity as read cache and 30% as write cache.

2.2 Cluster Types

Standard Cluster

A standard vSAN cluster consists of a minimum of three physical nodes and can be scaled to 64 nodes. All the hosts in a standard cluster are commonly located at a single location and are well-connected on the same Layer-2 network. 10Gb network connections are required for all-flash configurations and highly recommended for hybrid configurations.

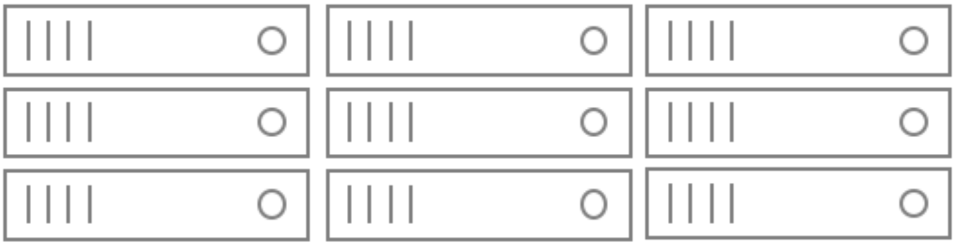


Figure 2. vSAN Standard Cluster

2-Node Cluster

As the name implies, a 2-node cluster consists of two physical nodes in the same location. These hosts are usually connected to the same network switch or connected directly using network crossover cables. Use of crossover cables eliminate the need to procure and manage an expensive network switch for a 2-node cluster, which lowers costs—especially in scenarios such as remote office deployments.

A third host called a “witness” is required for a 2-node configuration to avoid “split-brain” issues if network connectivity is lost between the two physical nodes. The witness is a virtual appliance provided by VMware that runs ESXi. We will discuss the witness virtual appliance in more detail shortly.

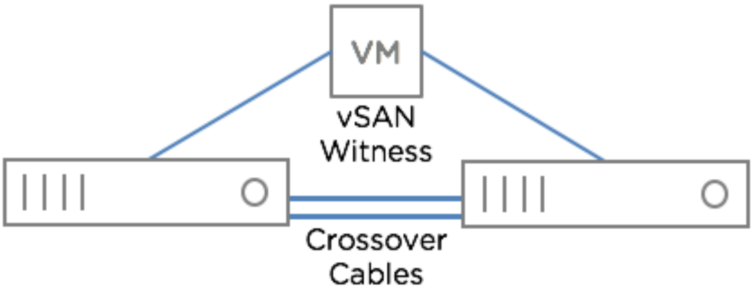


Figure 3. 2-Node cluster

Stretched Cluster

A vSAN stretched cluster provides resiliency against the loss of an entire site. The hosts in a stretched cluster are distributed evenly across two sites. The two sites are well-connected from a network perspective with a round trip time (RTT) latency of no more than five milliseconds. A witness virtual appliance is placed at a third site to avoid “split-brain” issues if connectivity is lost between the two stretched cluster sites.

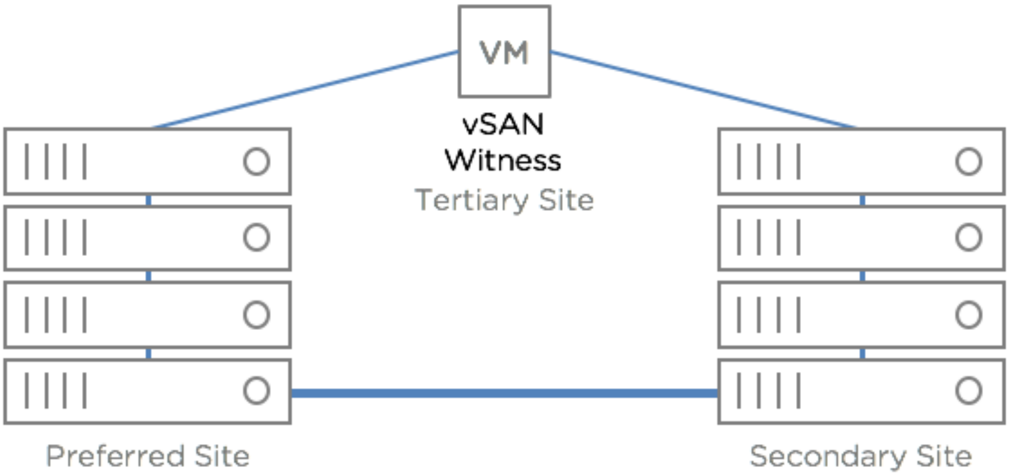


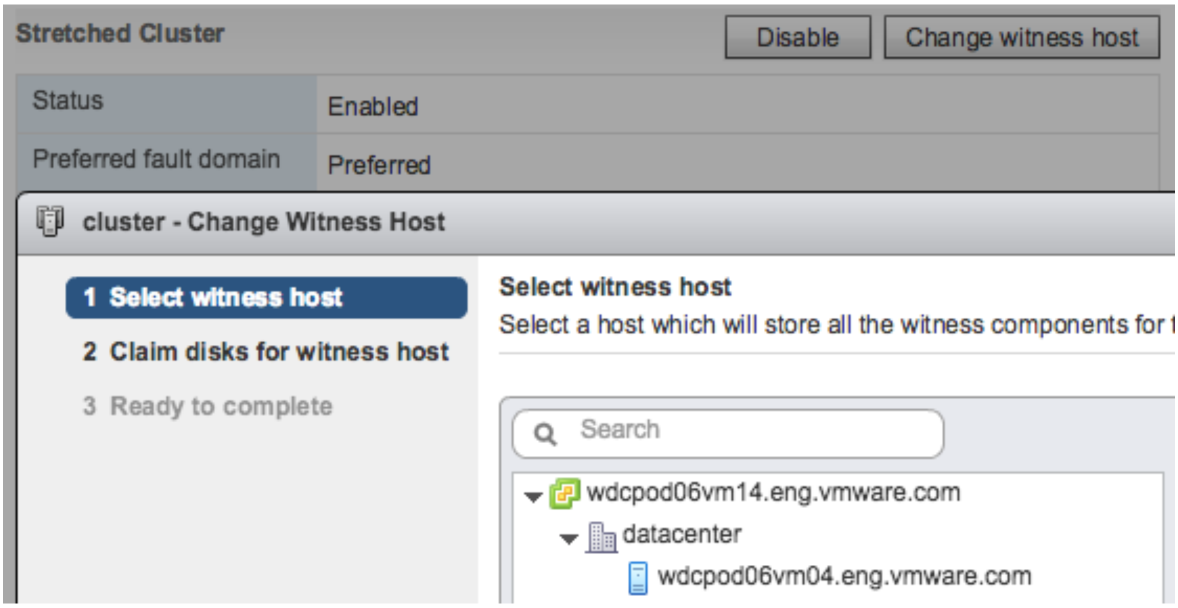
Figure 4. vSAN Stretched Cluster

vSAN Witness Virtual Appliance

2-node and stretched cluster vSAN deployments require the use of a witness virtual appliance that runs ESXi. The virtual appliance is easily deployed from a pre-configured OVA provided by VMware. The witness stores metadata commonly called “witness components” for vSAN objects. Virtual machine data such as virtual disks is not stored on the witness. The primary purpose of the witness is to serve as a “tie-breaker” in cases where the physical nodes in the configuration are disconnected.

Witness virtual appliances for remote office implementations commonly reside at a primary data center. Witness virtual appliances for stretched clusters must be located at a tertiary site that is independent of the Preferred and Secondary stretched cluster sites.

vSAN makes it easy to replace the witness virtual appliance. This capability minimizes the amount of time a stretched cluster must run without a witness in cases where the witness goes offline permanently. An example of this might be a catastrophic failure at the site where the witness host is running. The vSAN UI has a Change Witness Host button, which opens a simple, three-step wizard to change the witness host.



Status: Enabled
Preferred fault domain: Preferred

cluster - Change Witness Host

1 Select witness host
2 Claim disks for witness host
3 Ready to complete

Select witness host
Select a host which will store all the witness components for t

Figure 5. Change Witness Host

2.3 Hardware Deployment Options

A vSAN ReadyNode™ is an x86 server, available from all the leading server vendors, which is pre-configured, tested, and certified for vSAN. vSAN ReadyNodes provide an open, flexible approach when considering deployment methods. Organizations can continue to use their server vendor(s) of choice. Each ReadyNode is optimally configured for vSAN with the required amount of CPU, memory, network, I/O controllers and storage devices.

Turn-key appliances such as Dell EMC VxRail™ provide a fully integrated VMware hyper-converged infrastructure solution for a variety of applications and workloads. Simple deployment enables customers to be up and running in as little as 15 minutes. Dell EMC VxRack System SDDC™ powered by VMware Cloud Foundation™ is a rack-scale system with fully integrated hardware and software including vSAN. These solutions provide an agile and flexible infrastructure foundation that IT organizations can use as part of their transformation into an IT-as-a-Service operating model.

Custom configurations using jointly validated components from all the major OEM vendors is also an option. The vSAN Hardware Quick Reference Guide provides some sample server configurations as directional guidance and all components should be validated using the VMware Compatibility Guide for vSAN.

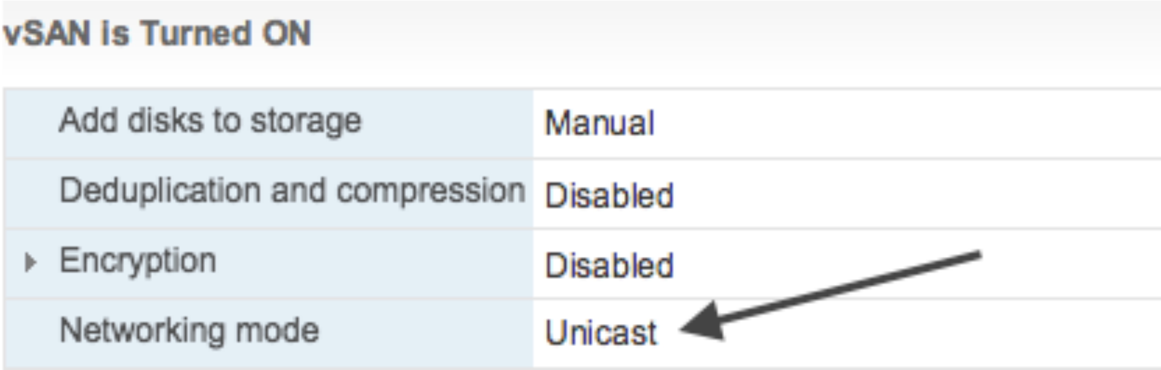
512e Drive Support

Disk drives have been using a native 512-byte sector size. Due to increasing demands for larger capacities, the storage industry introduced new formats that use 4KB physical sectors. These are commonly referred to as "4K native" drives or simply "4Kn" drives. Some 4Kn devices include firmware that emulates 512 byte (logical) sectors while the underlying (physical) sectors are 4K. These devices are referred to as "512B emulation" or "512e" drives.

vSphere 6.5 along with vSAN 6.5 and newer support the use of 512e drives. The latest information regarding support for these new drive types can be found in this VMware Knowledge Base Article: Support statement for 512e and 4K Native drives for VMware vSphere and vSAN (2091600)

2.4 Networking

All flash vSAN configurations require 10Gb network connectivity. 1Gb connections are supported for hybrid configurations although 10Gb is recommended. vSAN 6.6 simplifies design and deployment by removing the need for multicast network traffic (required for versions of vSAN prior to 6.6). When upgrading from a previous version of vSAN to vSAN 6.6, multicast is required until all hosts in the cluster are running version 6.6. vSAN automatically changes to unicast once the upgrade is complete.



vSAN Is Turned ON

Add disks to storage	Manual
Deduplication and compression	Disabled
Encryption	Disabled
Networking mode	Unicast

Figure 6. Unicast Networking Mode in vSAN 6.6

The network infrastructure for a vSAN environment should be designed with the same levels of redundancy as any other storage fabric. This helps ensure desired performance and availability requirements are met for the workloads running on vSAN. More information on vSAN network requirements and configuration recommendations can be found in the VMware vSAN Network Design guide.

Witness Networking

WAN connectivity between the witness and its 2-node or stretched cluster configuration is common. The throughput requirement for the connection between a witness virtual appliance and the rest of the cluster is small to help minimize network connection costs.

Witness network traffic can be separated from the network that connects the physical hosts in a 2-node cluster. This configuration option reduces complexity by eliminating the need to create and maintain static routes on the physical hosts. Security is also improved as the data network (between physical hosts) is completely separated from the WAN for witness traffic.

2.5 Storage Controller Virtual Appliance Disadvantages

Storage in a hyper-converged infrastructure (HCI) requires compute resources that have been traditionally offloaded to dedicated storage arrays. Nearly all other HCI solutions require the deployment of storage virtual appliances to some or all hosts in the cluster. These appliances provide storage services to each host. Storage virtual appliances typically require dedicated CPU and/or memory to avoid resource contention with other virtual machines.

Running a storage virtual appliance on every host in the cluster reduces the overall amount of compute resources available to run regular virtual machine workloads. Consolidation ratios are lower and total cost of ownership rises when these storage virtual appliances are present and competing for the same resources as regular virtual machine workloads.

Storage virtual appliances can also introduce additional latency, which negatively affects performance. This is due to the number of steps required to handle and replicate write operations as shown in the figure below.

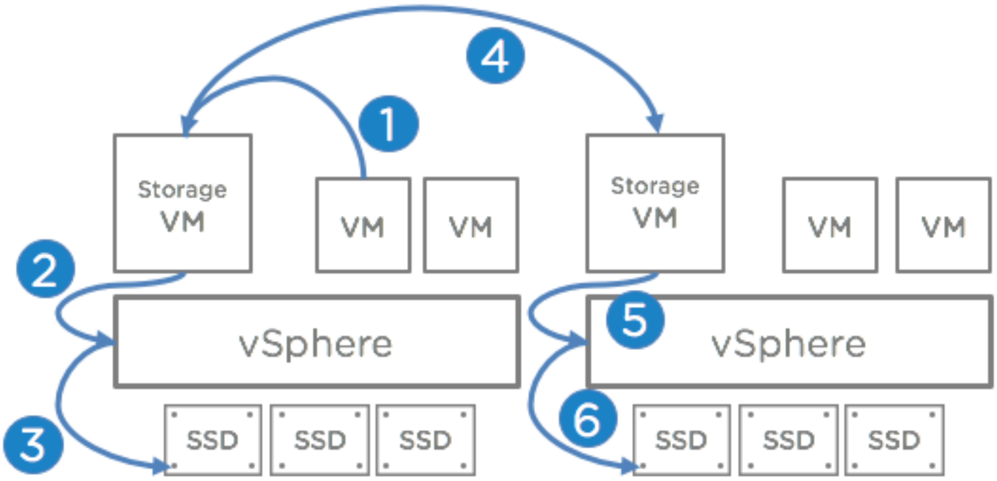


Figure 7. Storage controller virtual appliances

2.6 vSAN is Native in the vSphere Hypervisor

vSAN does not require the deployment of storage virtual appliances or the installation of a vSphere Installation Bundle (VIB) on every host in the cluster. vSAN is native in the vSphere hypervisor and typically consumes less than 10% of the compute resources on each host. vSAN does not compete with other virtual machines for resources and the I/O path is shorter.

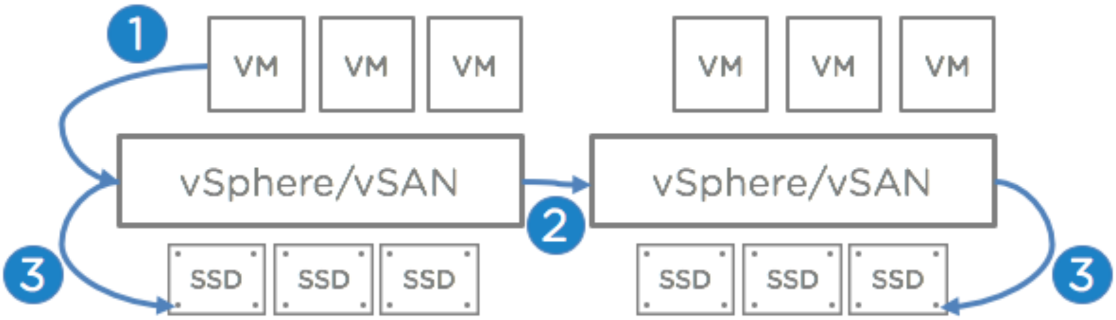


Figure 8. vSAN is native in the vSphere hypervisor

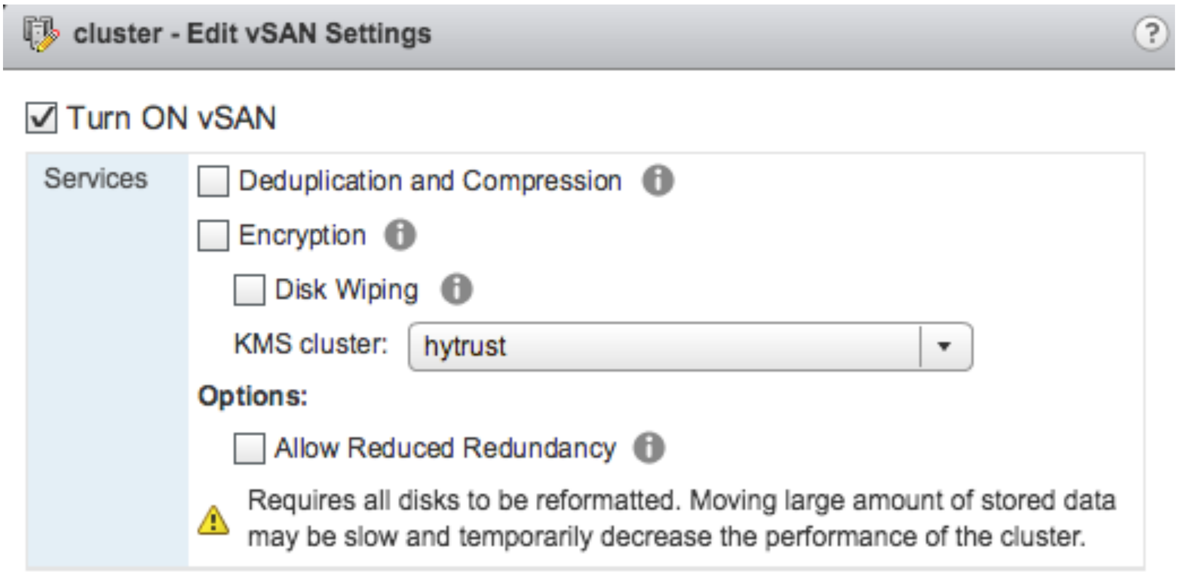
A shorter I/O path and the absence of resource-intensive storage virtual appliances enables vSAN to provide excellent performance with minimal overhead. Higher virtual machine consolidation ratios translate into lower total costs of ownership.

3. Enabling vSAN

.

3.1 Turn on vSAN

vSAN is enabled with just a few mouse clicks. There is no requirement to install additional software and/or deploy virtual storage appliances to every host in the cluster. Simply click the Enable vSAN checkbox to start the process. Deduplication and compression, as well as, encryption can also be enabled when turning on vSAN.



cluster - Edit vSAN Settings

☑ Turn ON vSAN

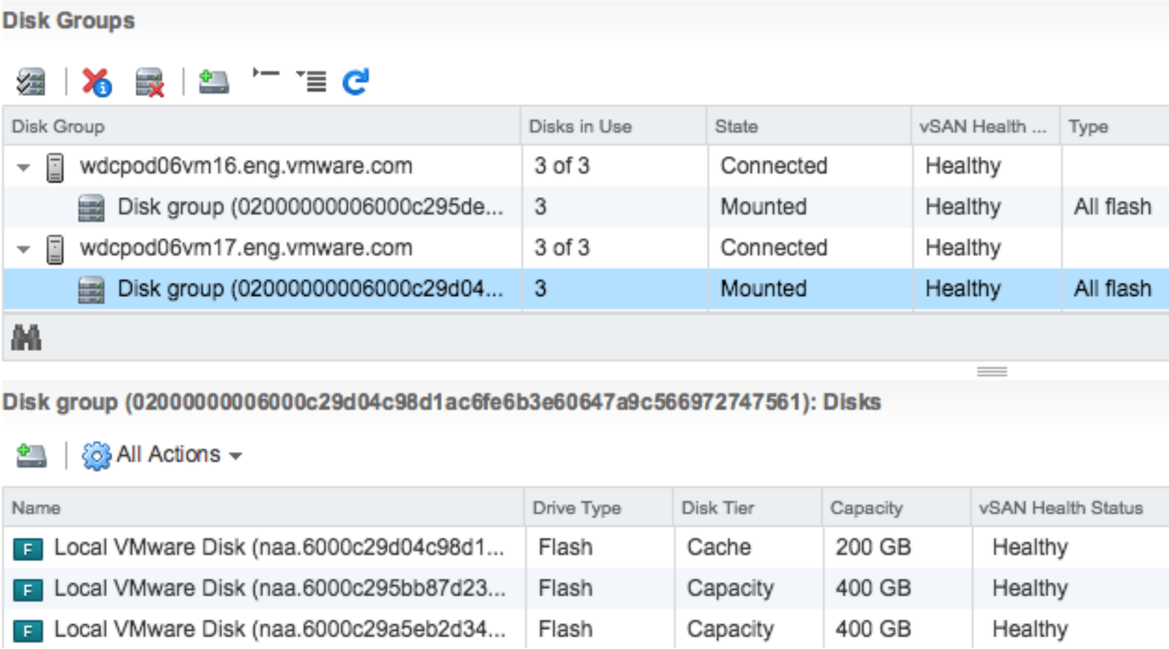
Services
- □ Deduplication and Compression
- □ Encryption
  - □ Disk Wiping
  - KMS cluster: hytrust

Options:
- □ Allow Reduced Redundancy

⚠ Requires all disks to be reformatted. Moving large amount of stored data may be slow and temporarily decrease the performance of the cluster.

Figure 9. Enable vSAN

The next step is claiming local storage devices in each host for the vSAN cache and capacity tiers. One or more disk groups are created in each host. Each disk group contains one cache device and one or more capacity devices. vSAN pools these local storage devices together to create a shared datastore.



Disk Groups

Disk Group	Disks in Use	State	vSAN Health ...	Type
wdcpod06vm16.eng.vmware.com	3 of 3	Connected	Healthy	
Disk group (02000000006000c295de...	3	Mounted	Healthy	All flash
wdcpod06vm17.eng.vmware.com	3 of 3	Connected	Healthy	
Disk group (02000000006000c29d04...	3	Mounted	Healthy	All flash

Disk group (02000000006000c29d04c98d1ac6fe6b3e60647a9c566972747561): Disks

All Actions

Name	Drive Type	Disk Tier	Capacity	vSAN Health Status
Local VMware Disk (naa.6000c29d04c98d1...	Flash	Cache	200 GB	Healthy
Local VMware Disk (naa.6000c295bb87d23...	Flash	Capacity	400 GB	Healthy
Local VMware Disk (naa.6000c29a5eb2d34...	Flash	Capacity	400 GB	Healthy

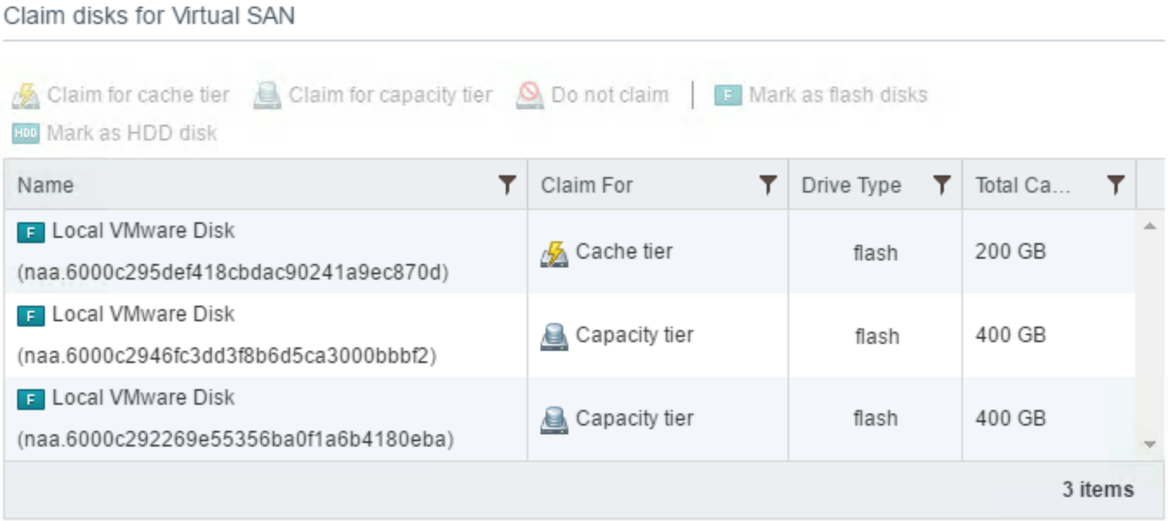
Figure 10. vSAN disk group

After disk groups have been configured, the vSAN datastore is available for use. In most cases, the time required to enable vSAN and configure disk groups is measured in minutes.

3.2 Easy Install

Deployment of a vSAN cluster is easier than ever before. The vCenter Server Appliance (VCSA) installation wizard enables administrators to install vSphere on a single host, configure a vSAN datastore, and deploy a VCSA to this datastore. Additional vSphere hosts are then added to finish the build out of the vSAN cluster. This eliminates the need to provision additional disks to run the VCSA prior to enabling vSAN, which simplifies and streamlines the deployment process. This enhancement is especially useful when deploying a new cluster where there is no existing infrastructure to host the VCSA.

The following figure shows a step in the VCSA deployment wizard where the vSAN datastore is initially configured. In this case, three flash devices will be utilized in the host—one for write caching and the other two for capacity.



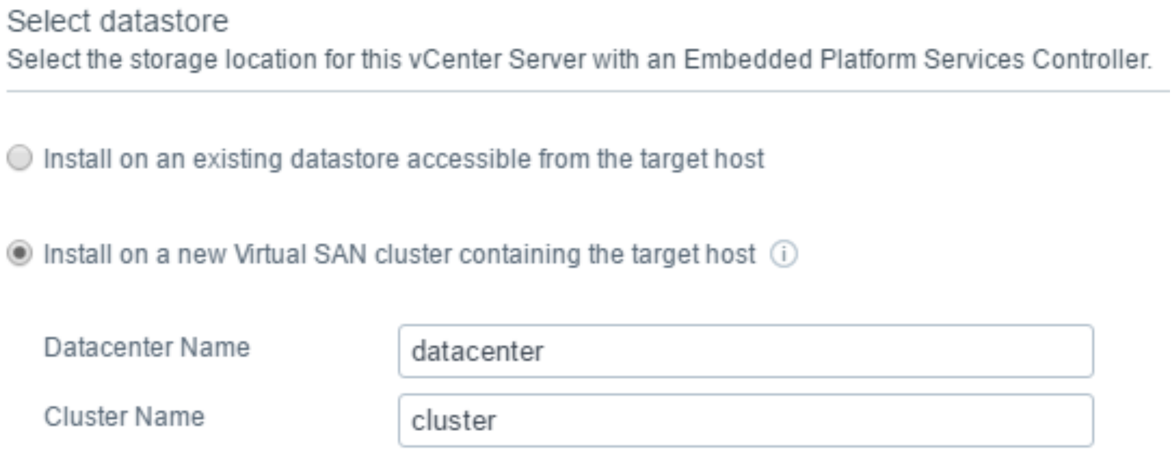
Claim disks for Virtual SAN

Name	Claim For	Drive Type	Total Ca...
F Local VMware Disk (naa.6000c295def418cbdac90241a9ec870d)	Cache tier	flash	200 GB
F Local VMware Disk (naa.6000c2946fc3dd3f8b6d5ca3000bbbf2)	Capacity tier	flash	400 GB
F Local VMware Disk (naa.6000c292269e55356ba0f1a6b4180eba)	Capacity tier	flash	400 GB

3 items

Figure 11. Claiming disks for the vSAN datastore

The next figure shows the selection of the vSAN datastore where the VCSA will be deployed. A datacenter name and cluster name are also specified in this step.



Select datastore

Select the storage location for this vCenter Server with an Embedded Platform Services Controller.

○ Install on an existing datastore accessible from the target host

● Install on a new Virtual SAN cluster containing the target host

Datacenter Name: datacenter

Cluster Name: cluster

Figure 12. Selecting the new vSAN datastore for the VCSA

With just a few more clicks, a vSAN datastore is configured on the host and the VCSA is deployed on this new datastore. The Easy Install functionality enables simple, rapid deployment of a vSAN cluster managed by vCenter Server.

4. Security

.

4.1 Native Encryption

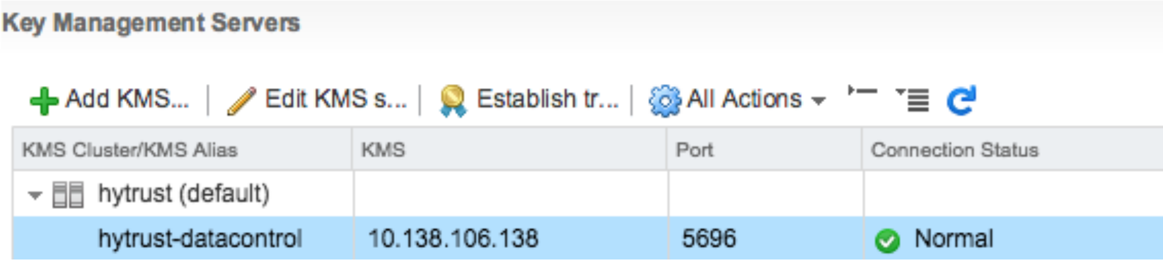
Data-at-rest encryption is an option for vSAN datastores to further improve security and provide compliance with increasingly stringent regulatory requirements. vSAN datastore encryption uses an AES 256 cipher. vSAN encryption eliminates the extra cost, limitations, and complexity associated with purchasing and maintaining self-encrypting drives.

vSAN datastore encryption is enabled and configured at the datastore level. In other words, every object on the vSAN datastore is encrypted when this feature is enabled. Data is encrypted when it is written to persistent media in the cache and capacity tiers of a vSAN datastore.

Encryption occurs just above the device driver layer of the vSphere storage stack, which means it is compatible with all vSAN features such as deduplication and compression, RAID-5/6 erasure coding, stretched cluster configurations. All vSphere features including vSphere vMotion, vSphere Distributed Resource Scheduler (DRS), vSphere High Availability (HA), and vSphere Replication are supported.

A Key Management Server (KMS) is required to enable and use vSAN encryption. Multiple KMS vendors are compatible including HyTrust, Gemalto (SafeNet), Thales e-Security, CloudLink, and Vormetric. These solutions are commonly deployed in clusters of hardware appliances or virtual appliances for redundancy and high availability.

Initial configuration is done in the vCenter Server UI of the vSphere Web Client. The KMS cluster is added to vCenter Server and a trust relationship is established. The process for doing this varies depending on the KMS vendor, but it is usually quite simple.

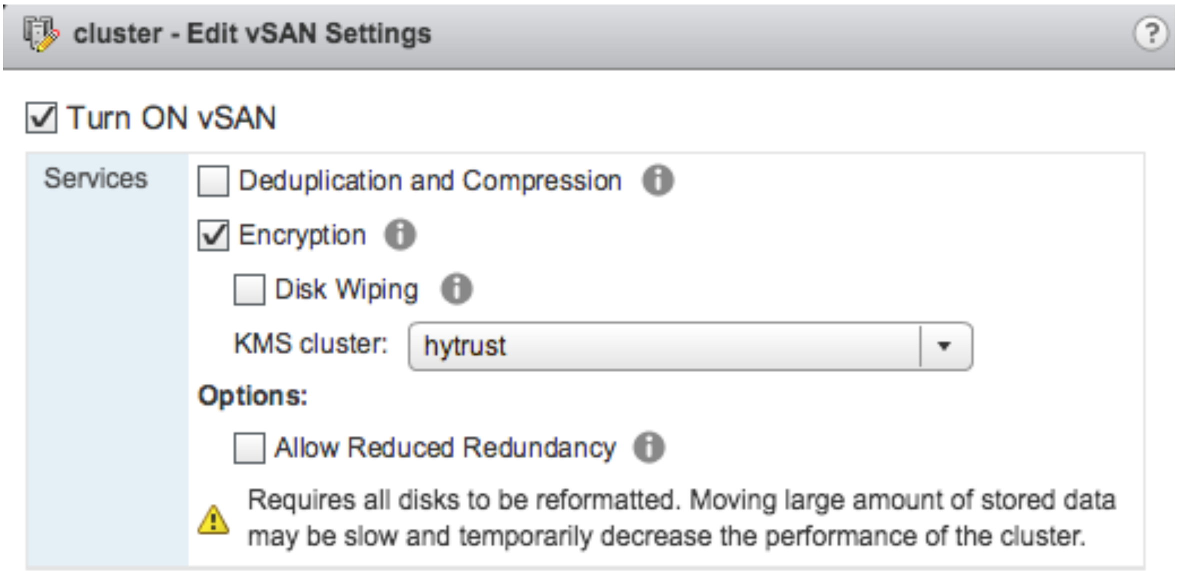


KMS Cluster/KMS Alias	KMS	Port	Connection Status
hytrust (default)			
hytrust-datacontrol	10.138.106.138	5696	✓ Normal

Figure 13. KMS configured for use with vCenter Server

Turning on encryption is a simple matter of clicking a checkbox. Encryption can be enabled when vSAN is turned on or after, with or without virtual machines residing on the datastore.

Note: A rolling reformat is required when encryption is enabled. This can take a considerable amount of time—especially if large amounts of existing data must be migrated as the rolling reformat takes place.



cluster - Edit vSAN Settings

☑ Turn ON vSAN

Services

□ Deduplication and Compression

☑ Encryption

□ Disk Wiping

KMS cluster: hytrust

Options:

□ Allow Reduced Redundancy

Requires all disks to be reformatted. Moving large amount of stored data may be slow and temporarily decrease the performance of the cluster.

Figure 14. Enabling vSAN Encryption

Encryption keys are transferred to vSAN hosts using the Key Management Interoperability Protocol (KMIP). Industry standards and compliance with regulations often require the generation of new keys on a regular basis. This reduces the risk of a key being exposed or compromised by brute force. Generating new keys is performed in the vSAN UI with just a few clicks.

Encryption can be disabled for a cluster. Like enabling encryption, a rolling disk format change is required. Disabling encryption can take a significant amount of time.

4.2 Compliance

vSAN is native to the vSphere hypervisor and, because of that tight integration, shares the robust security and compliance benefits realized by the vSphere platform. 2-factor authentication methods, such as RSA SecurID and Common Access Card (CAC), are supported by vCenter Server, vSphere and vSAN.

vSAN is part of the core vSphere STIG. These native features help organizations achieve industry certifications and comply with regulatory requirements. Visit the Security portal on vmware.com for more details on VMware’s well-established programs and practices to identify and remediate security vulnerabilities.

5. Operational Efficiency

.

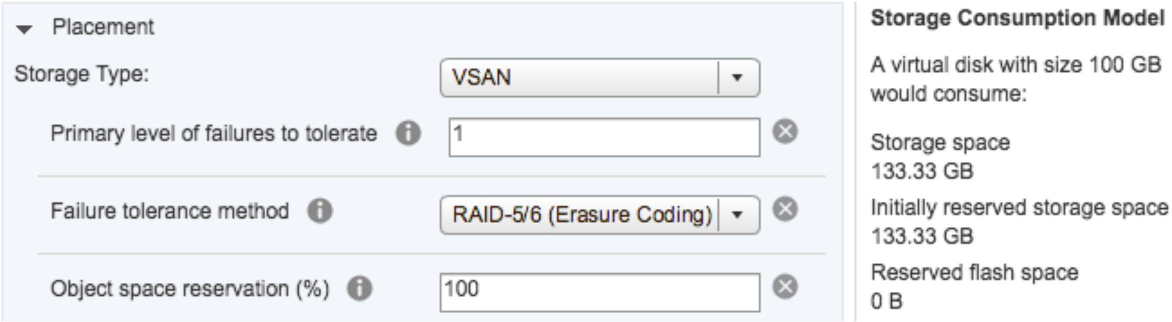
5.1 Storage Policy-Based Management

Traditional storage solutions commonly use LUNs or volumes. A LUN or a volume is configured with a specific disk configuration such as RAID to provide a specific level of performance and availability. The challenge with this model is each LUN or volume is confined to providing only one level of service regardless of the workloads that it contains. This leads to provisioning numerous LUNs or volumes to provide the right levels of storage services for various workload requirements. Maintaining many LUNs or volumes increases complexity. Deployment and management of workloads and storage in traditional storage environments is often a manual process that is time consuming and error prone.

Storage Policy-Based Management (SPBM) from VMware enables precise control of storage services. Like other storage solutions, vSAN provides services such as availability levels, capacity consumption, and stripe widths for performance. A storage policy contains one or more rules that define service levels.

Storage policies are created and managed using the vSphere Web Client. Policies can be assigned to virtual machines and individual objects such as a virtual disk. Storage policies are easily changed or reassigned if application requirements change. These modifications are performed with no downtime and without the need to migrate virtual machines from one datastore to another. SPBM makes it possible to assign and modify service levels with precision on a per-virtual machine basis.

The following figure shows storage policy rules. This policy contains three rules. The first rule, “Primary level of failures to tolerate," defines how many failures an object can tolerate before it becomes unavailable. The second rule indicates the failure tolerance method that will be used. This policy uses RAID-5/6 erasure coding to minimize capacity consumption. The third rule, “Object space reservation," enables administrators to reserve capacity for objects.



Placement

Storage Type: VSAN

Primary level of failures to tolerate: 1

Failure tolerance method: RAID-5/6 (Erasure Coding)

Object space reservation (%): 100

Storage Consumption Model

A virtual disk with size 100 GB would consume:

Storage space
133.33 GB

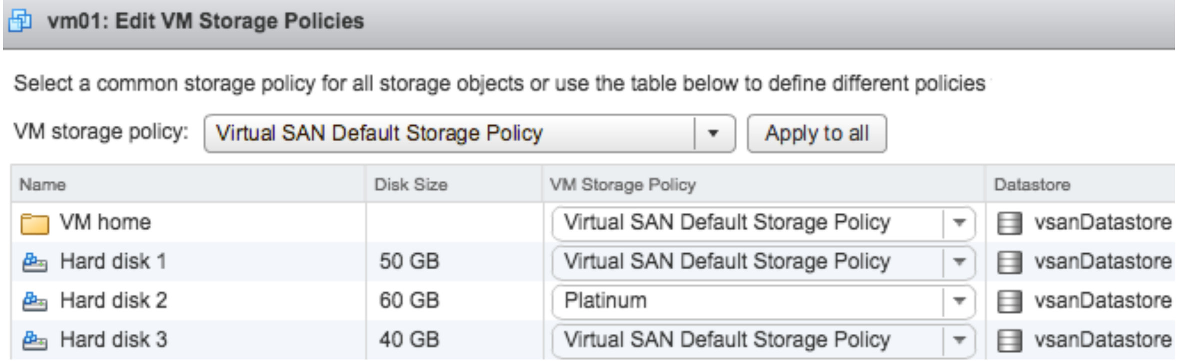
Initially reserved storage space
133.33 GB

Reserved flash space
0 B

Figure 15. vSAN storage policy

Note: By default, vSAN objects are thin-provisioned with an implicit object space reservation of 0%. Specifying an object space reservation of 100% is like thick-provisioning on traditional vSphere datastores.

As mentioned previously, storage policies can be applied to all objects that make up a virtual machine and to individual objects such as a virtual disk. The figure below shows a virtual machine with three virtual disks. One of the virtual disks has the “Platinum” storage policy assigned while the rest of the objects have the default storage policy assigned.



vm01: Edit VM Storage Policies

Select a common storage policy for all storage objects or use the table below to define different policies

VM storage policy: Virtual SAN Default Storage Policy

Name	Disk Size	VM Storage Policy	Datastore
VM home		Virtual SAN Default Storage Policy	vsanDatastore
Hard disk 1	50 GB	Virtual SAN Default Storage Policy	vsanDatastore
Hard disk 2	60 GB	Platinum	vsanDatastore
Hard disk 3	40 GB	Virtual SAN Default Storage Policy	vsanDatastore

Figure 16. Assigning storage policies

5.2 Automation

VMware PowerCLI is one of the most widely adopted extensions to the PowerShell framework. VMware PowerCLI includes a very comprehensive set of functions that abstract the vSphere API down to simple and powerful cmdlets including many for vSAN. This makes it easy to automate several actions from enabling vSAN to deployment and configuration of a vSAN stretched cluster. Here are a few simple examples of what can be accomplished with vSAN and PowerCLI:

Assigning a Storage Policy to Multiple VMs

Sparse Virtual Swap Files

vSAN features an extensive management API and multiple software development kits (SDKs) to provide IT organizations options for rapid provisioning and automation. Administrators and developers can orchestrate all aspects of installation, configuration, lifecycle management, monitoring, and troubleshooting in vSAN environments. Recent updates include commandlets for performance monitoring, cluster upgrades, and vSAN iSCSI operations. A Host-level API can query cluster-level information. S.M.A.R.T. device data can also be obtained through the vSAN API.

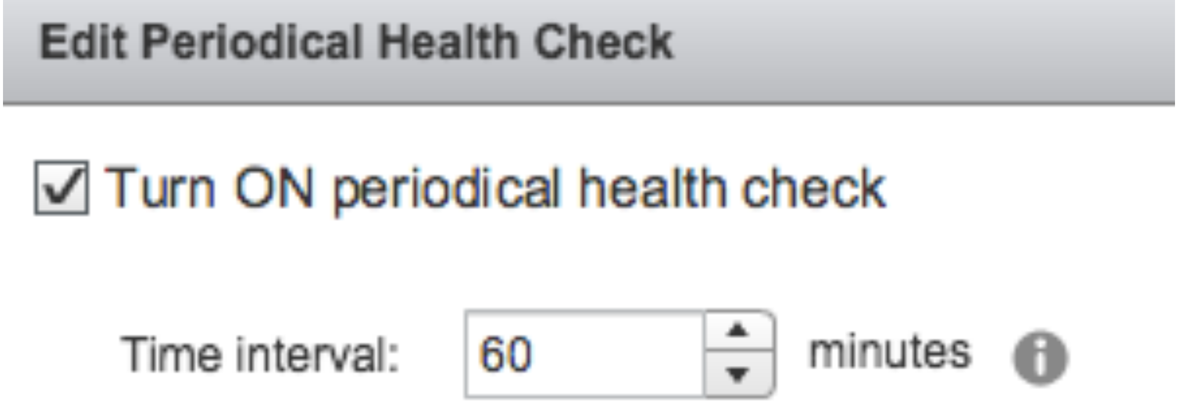
SDKs are available for several programming languages including .NET, Perl, and Python. The SDKs are available for download from VMware {code} and include libraries, documentation, and code samples.

vSphere administrators and DevOps shops can utilize these SDKs and PowerCLI cmdlets to lower costs by enforcing standards, streamlining operations, and enabling automation for vSphere and vSAN environments.

5.3 Health

vSAN features a comprehensive health service appropriately called vSAN Health that actively tests and monitors many items such as hardware compatibility, verification of storage device controllers, controller queue depth, and environmental checks for all-flash and hybrid vSAN configurations. vSAN Health examines network connectivity and throughput, disk and cluster health, and capacity consumption. Proactive monitoring and alerting in vSAN Health helps ensure the environment is optimally configured and functioning properly for the highest levels of performance and availability.

vSAN Health is enabled by default and configured to check the health of the vSAN environment every 60 minutes. The 60-minute time interval is the recommended setting, but this setting can be changed.



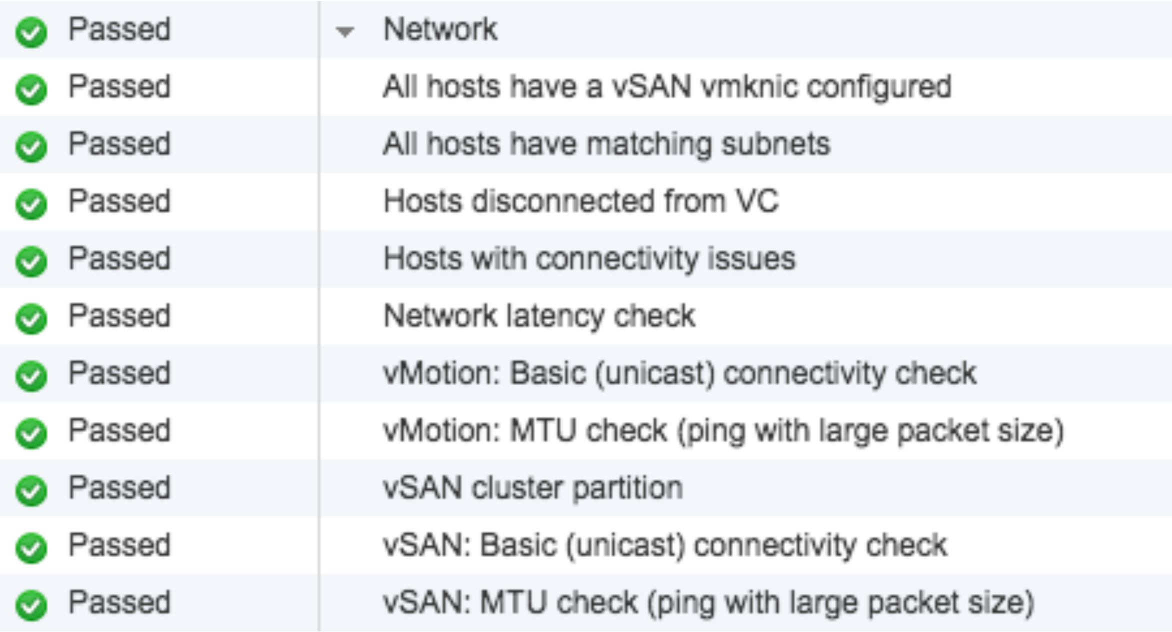
Edit Periodical Health Check

☑ Turn ON periodical health check

Time interval: 60 minutes

Figure 17. Periodical health check time interval

vSAN Health is thorough in the number of tests it performs. As an example, there are 10 tests just in the Network section of the vSAN Health UI.



✓ Passed	▾ Network
✓ Passed	All hosts have a vSAN vmknic configured
✓ Passed	All hosts have matching subnets
✓ Passed	Hosts disconnected from VC
✓ Passed	Hosts with connectivity issues
✓ Passed	Network latency check
✓ Passed	vMotion: Basic (unicast) connectivity check
✓ Passed	vMotion: MTU check (ping with large packet size)
✓ Passed	vSAN cluster partition
✓ Passed	vSAN: Basic (unicast) connectivity check
✓ Passed	vSAN: MTU check (ping with large packet size)

Figure 18. Network health

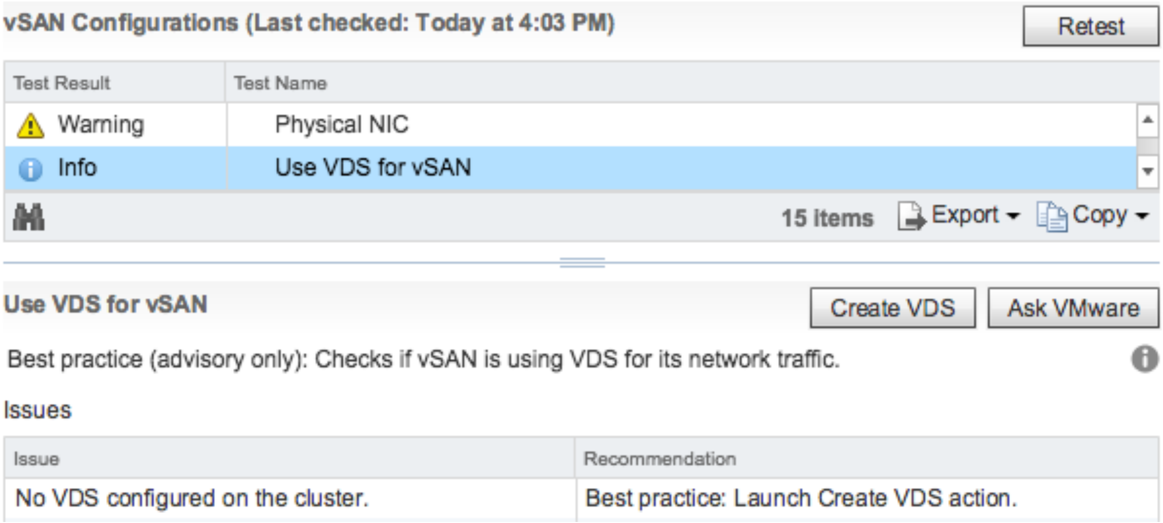
If an issue is detected, a warning is immediately visible in the vSAN UI. Clicking the warning provides more details about the issue. In addition to providing details about the warning, vSAN Health also has an Ask VMware button, which brings up the relevant VMware Knowledge Base article. This simplifies and streamlines remediation efforts.

5.4 vSAN Configuration Assist

An important aspect of a healthy vSAN environment is ensuring correct configurations, device firmware, and device drivers. vSAN 6.6 includes vSAN Configuration Assist to check hardware compatibility, burn-in testing, network configuration, vSAN configuration, and adherence to VMware cluster recommendations.

As example of a recommendation or “best practice” is each vSAN vmknic should be backed by two physical network interface cards (NICs). Configuration Assist will check the configuration of the vSAN vmknic and recommend two NICs if only one is configured.

Another example is the recommendation to use a virtual distributed switch (VDS) for vSAN. In the figure below, the recommendation to use a VDS is visible and a “Create VDS” button is provided to start the process.



vSAN Configurations (Last checked: Today at 4:03 PM)

Test Result	Test Name
⚠ Warning	Physical NIC
ⓘ Info	Use VDS for vSAN

Use VDS for vSAN

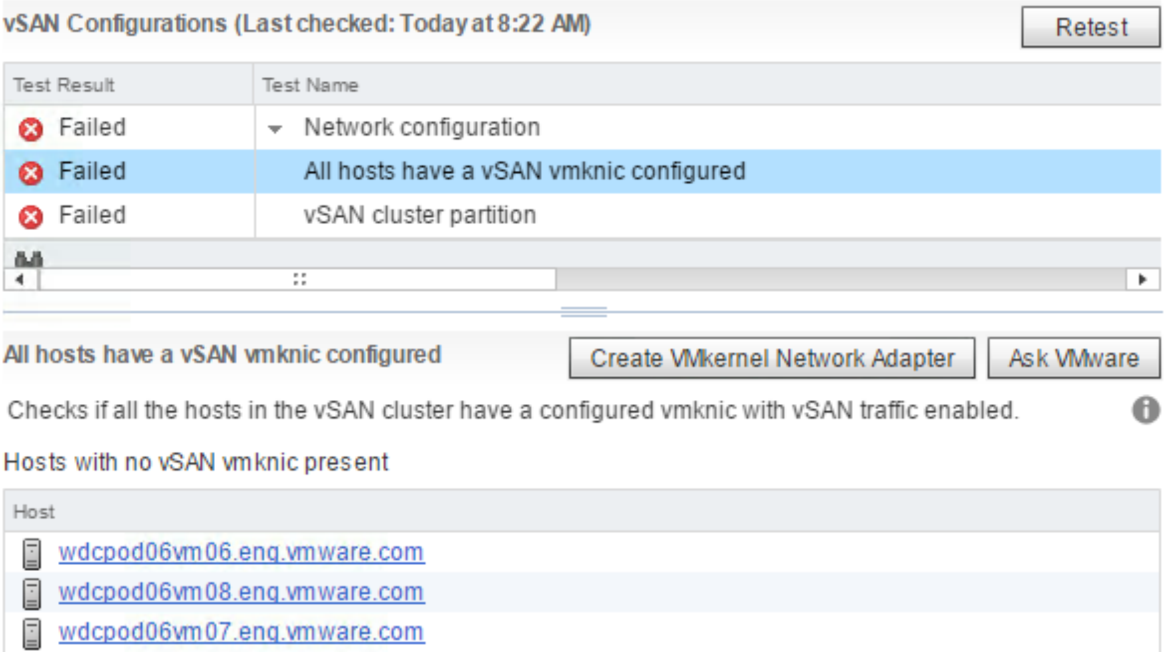
Best practice (advisory only): Checks if vSAN is using VDS for its network traffic.

Issues

Issue	Recommendation
No VDS configured on the cluster.	Best practice: Launch Create VDS action.

Figure 19. vSAN configuration checks

vSAN Configuration Assist also simplifies the creation of a VMkernel network adapter for vSAN on each host in a vSAN cluster. The following figure shows a 3-node cluster where the hosts do not have a vSAN vmknic configured. Configuration Assist identifies the issue configuration issue and includes the “Create VMkernel Network Adapter” button. Clicking this button initiates the process of configuring the vSAN vmknics on a virtual distributed switch. It eliminates a manual process and ensures consistency across the cluster.



vSAN Configurations (Last checked: Today at 8:22 AM)

Test Result	Test Name
☒ Failed	▾ Network configuration
☒ Failed	All hosts have a vSAN vmknic configured
☒ Failed	vSAN cluster partition

All hosts have a vSAN vmknic configured

Checks if all the hosts in the vSAN cluster have a configured vmknic with vSAN traffic enabled.

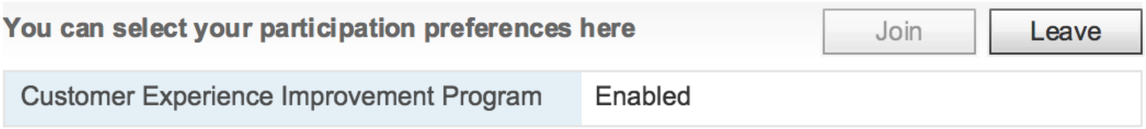
Hosts with no vSAN vmknic present

Host
wdcpod06vm06.eng.vmware.com
wdcpod06vm08.eng.vmware.com
wdcpod06vm07.eng.vmware.com

Figure 20. vSAN Configuration Assist Create VMkernel Adapter

5.5 Proactive Cloud Health Checks

Participating in the Customer Experience Improvement Program (CEIP) enables VMware to provide higher levels of proactive and reactive customer assistance. Benefits of participating in this program include streamlined troubleshooting, real-time notifications and recommendations for your environment, diagnostics, and the potential to remedy issues before they become problems.

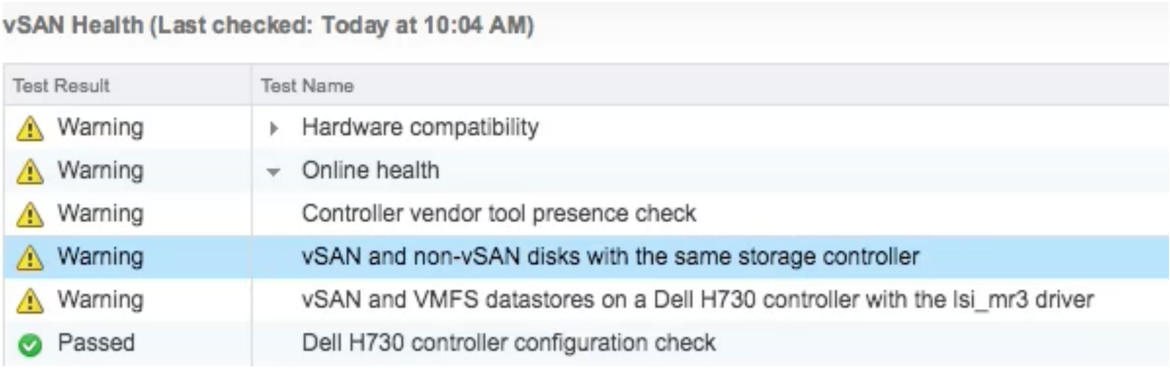


You can select your participation preferences here

Customer Experience Improvement Program	Enabled

Figure 21. Customer Experience Improvement Program enabled

vSAN Health is cloud-connected. New health checks appear as new VMware Knowledge Base (KB) articles are created and published. An “Ask VMware” button is supplied in the user interface, which guides administrators to the relevant VMware knowledge base article. This benefit is delivered without the need to upgrade vSphere and vSAN. This enhancement consistently provides administrators with latest checks and remedies for optimal reliability and performance. An example is shown below where VMware has detected vSAN and non-vSAN disks attached to the same storage controller.



vSAN Health (Last checked: Today at 10:04 AM)

Test Result	Test Name
⚠ Warning	▸ Hardware compatibility
⚠ Warning	▾ Online health
⚠ Warning	Controller vendor tool presence check
⚠ Warning	vSAN and non-vSAN disks with the same storage controller
⚠ Warning	vSAN and VMFS datastores on a Dell H730 controller with the lsi_mr3 driver
✓ Passed	Dell H730 controller configuration check

Figure 22. Online health recommendation

An “Ask VMware” button opens the Best practices when using vSAN and non-vSAN disks with the same storage controller (2129050) KB article. New health checks appear as new KB articles are created and published. This benefit is delivered without the need to upgrade vSphere and vSAN. This feature also provides data to VMware, which is correlated and analyzed to help identify opportunities for product improvement. VMware uses this insight to provide customers effective recommendations on design, configuration, and management of vSAN clusters.

Note: CEIP provides VMware with information that enables VMware to improve its products and services, to fix problems, and to advise you on how best to deploy and use our products. As part of the CEIP, VMware collects technical information about your organization’s use of VMware products and services on a regular basis in association with your organization’s VMware license key(s). This information does not personally identify any individual. For additional information regarding the CEIP, please see the Trust & Assurance Center.

5.6 vRealize Management Pack for vSAN

VMware vRealize Operations streamlines and automates IT operations management. Intelligent operations management from applications to infrastructure across physical, virtual, and cloud environments can be achieved with vRealize Operations. vRealize Operations Management Packs extend the capabilities of vRealize Operations by including prebuilt dashboards that focus on design, management, and operations for a variety of solutions and use cases.

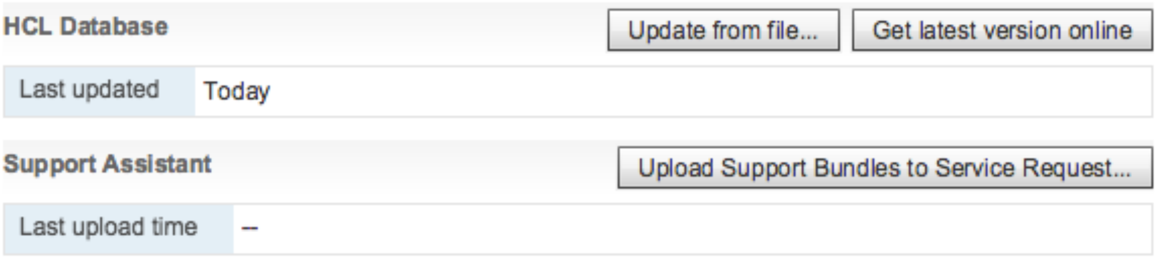
The vRealize Management Pack for vSAN provides additional visibility into and analytics for vSAN environments. An incredible number of metrics are exposed to assist with monitoring and issue remediation. vRealize Operations makes it easier to correlate data from multiple sources to speed troubleshooting and root cause analysis.

5.7 Hardware Compatibility

vSphere and vSAN support a wide variety of hardware configurations. The list of hardware components and corresponding drivers that are supported with vSAN can be found in the VMware Compatibility Guide. It is very important to use only hardware, firmware, and drivers found in this guide to ensure stability and performance.

The list of certified hardware, firmware, and drivers is contained in a hardware compatibility list (HCL) database. vSAN makes it easy to update this information. If the environment has Internet connectivity, updates are obtained directly from VMware. Otherwise, HCL database updates can be downloaded for offline use.

If an issue does arise that requires the assistance of VMware Global Support Services, it is easy to upload support bundles to help expedite the troubleshooting process. Clicking the Upload Support Bundles to Service Request button enables you to enter an existing support request (SR) number and easily upload the necessary logs.


HCL Database

Last updated: Today

Support Assistant

Last upload time: --

Update from file...
Get latest version online
Upload Support Bundles to Service Request...

Figure 23. vSAN HCL database and Support Assistant

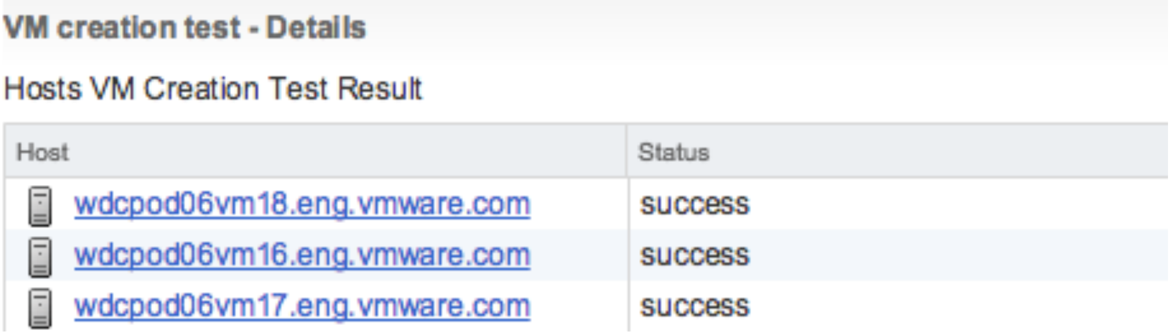
5.8 Hardware Lifecycle Management

vSAN streamlines common hardware maintenance tasks by automating the hardware update process. Outdated firmware and driver versions are identified in the UI and the option to download and install the latest supported software is provided. Software upgrades are initiated by a single click and orchestrated across the entire cluster to ease hardware lifecycle management. This feature eliminates manual update tasks and helps ensure all hosts in the cluster are at the same firmware and driver levels for best performance and stability.

Note: A subset of the OEMs and server models that run vSAN are currently supported. Some do not yet support automated remediation. The vSAN Health UI shows items that need attention, but remediation is a manual process in these cases.

5.9 Proactive Tests

vSAN Proactive Tests enable administrators to verify vSAN configuration, stability, and performance to minimize risk and confirm that the datastore is ready for production use. As an example, the VM creation test creates and deletes a small virtual machine on each host confirming basic functionality. The following figure shows the results of a VM creation test.



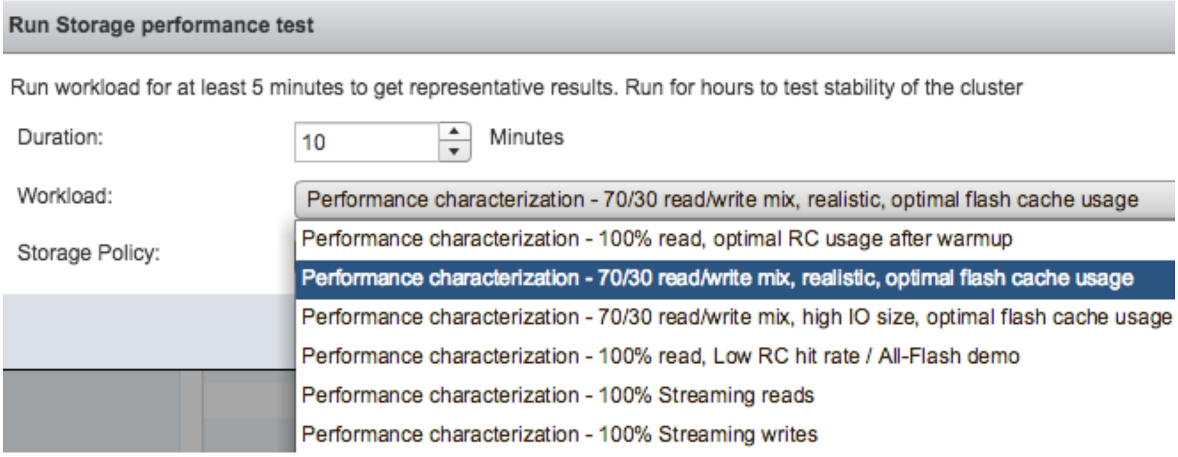
VM creation test - Details

Hosts VM Creation Test Result

Host	Status
wdcpod06vm18.eng.vmware.com	success
wdcpod06vm16.eng.vmware.com	success
wdcpod06vm17.eng.vmware.com	success

Figure 24. VM creation test results

The storage performance test is used to check the stability of the vSAN cluster under heavy I/O load. There are several workload profiles that can be selected for this test as shown below. Keep in mind the storage performance test can affect other workloads and tasks. This test is intended to run before production virtual machine workloads are provisioned on vSAN.



Run Storage performance test

Run workload for at least 5 minutes to get representative results. Run for hours to test stability of the cluster

Duration: 10 Minutes

Workload: Performance characterization - 70/30 read/write mix, realistic, optimal flash cache usage

Storage Policy:

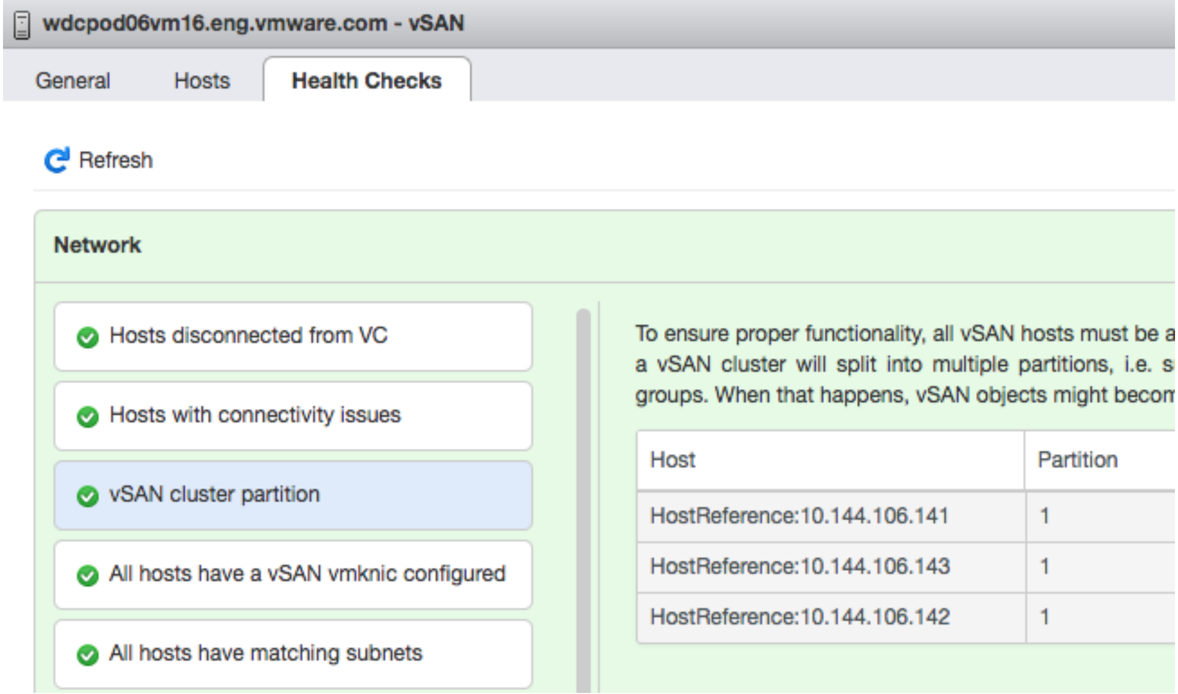
- Performance characterization - 100% read, optimal RC usage after warmup
- Performance characterization - 70/30 read/write mix, realistic, optimal flash cache usage
- Performance characterization - 70/30 read/write mix, high IO size, optimal flash cache usage
- Performance characterization - 100% read, Low RC hit rate / All-Flash demo
- Performance characterization - 100% Streaming reads
- Performance characterization - 100% Streaming writes

Figure 25. Storage performance test workload profiles

5.10 Highly Available Control Plane for Health Checks

Previous versions of vSAN required vCenter Server and the vSphere Web Client server (part of vCenter Server) to be online to check the health of the cluster. vSAN 6.6 includes the ability to perform vSAN health checks using the VMware Host Client in the rare event vCenter Server is offline.

Hosts in a vSAN cluster cooperate in a distributed fashion to check the health of the entire cluster. Any host in the cluster can be used to view vSAN Health. This provides redundancy for the vSAN Health data to help ensure administrators always have this information available. The figure below shows the Network – vSAN Cluster Partition health.



wdcpod06vm16.eng.vmware.com - vSAN

Network

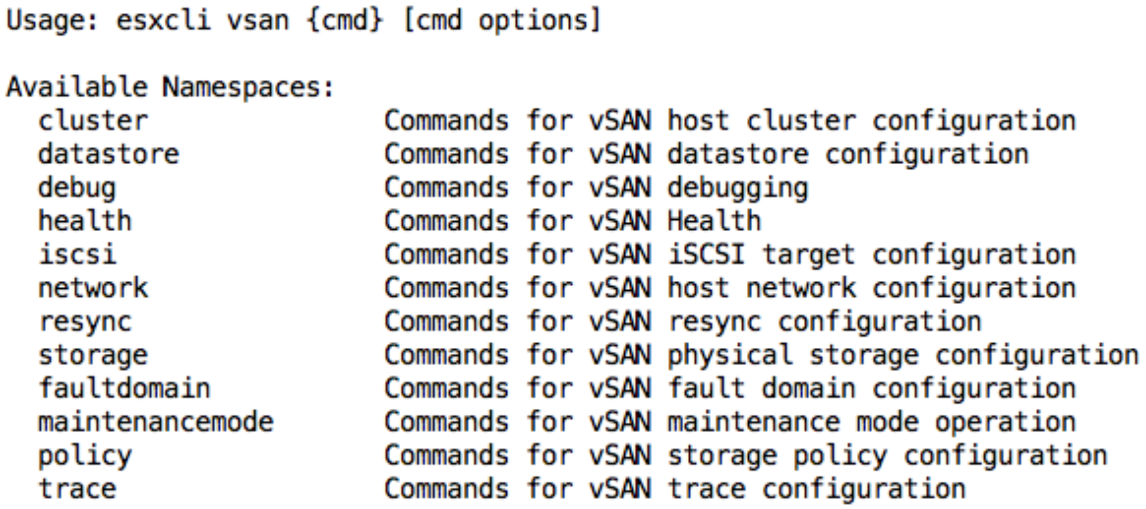
- ✓ Hosts disconnected from VC
- ✓ Hosts with connectivity issues
- ✓ vSAN cluster partition
- ✓ All hosts have a vSAN vmknic configured
- ✓ All hosts have matching subnets

To ensure proper functionality, all vSAN hosts must be a
a vSAN cluster will split into multiple partitions, i.e. s
groups. When that happens, vSAN objects might becom

Host	Partition
HostReference:10.144.106.141	1
HostReference:10.144.106.143	1
HostReference:10.144.106.142	1

Figure 26. vSAN Health in the VMware Host Client

Command line functionality is keeping pace with information available in the vSphere Web Client and VMware Host Client graphical user interfaces. Administrators can use esxcli vsan commands to check vSAN health, perform debugging, and manage configurations for items such as fault domains, storage policies, and iSCSI targets.



```
Usage: esxcli vsan {cmd} [cmd options]

Available Namespaces:
  cluster                Commands for vSAN host cluster configuration
  datastore              Commands for vSAN datastore configuration
  debug                  Commands for vSAN debugging
  health                 Commands for vSAN Health
  iscsi                  Commands for vSAN iSCSI target configuration
  network                Commands for vSAN host network configuration
  resync                 Commands for vSAN resync configuration
  storage                Commands for vSAN physical storage configuration
  faultdomain            Commands for vSAN fault domain configuration
  maintenancemode        Commands for vSAN maintenance mode operation
  policy                 Commands for vSAN storage policy configuration
  trace                  Commands for vSAN trace configuration
```

Figure 27. vsan esxcli commands

6. Availability

.

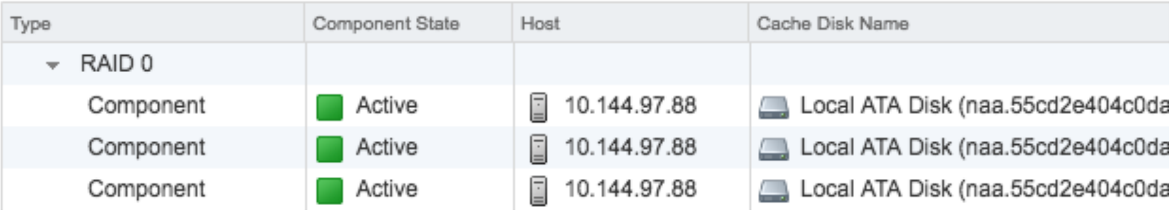
6.1 Objects and Component Placement

vSAN is an object datastore with a mostly flat hierarchy of objects and containers (folders). Items that make up a virtual machine are represented by objects. These are the most prevalent object types you will find on a vSAN datastore:

- VM Home, which contains virtual machine configuration files and logs, e.g., VMX file
- Virtual machine swap
- Virtual disk (VMDK)
- Delta disk (snapshot)
- Performance database

There are a few other objects that are commonly found on a vSAN datastore such as the vSAN performance service database, memory snapshot deltas, and VMDKs that belong to iSCSI targets.

Each object consists of one or more components. The number of components that make up an object depends primarily on a couple things: The size of the objects and the storage policy assigned to the object. The maximum size of a component is 255GB. If an object is larger than 255GB, it is split up into multiple components. The image below shows a 600GB virtual disk split up into three components.

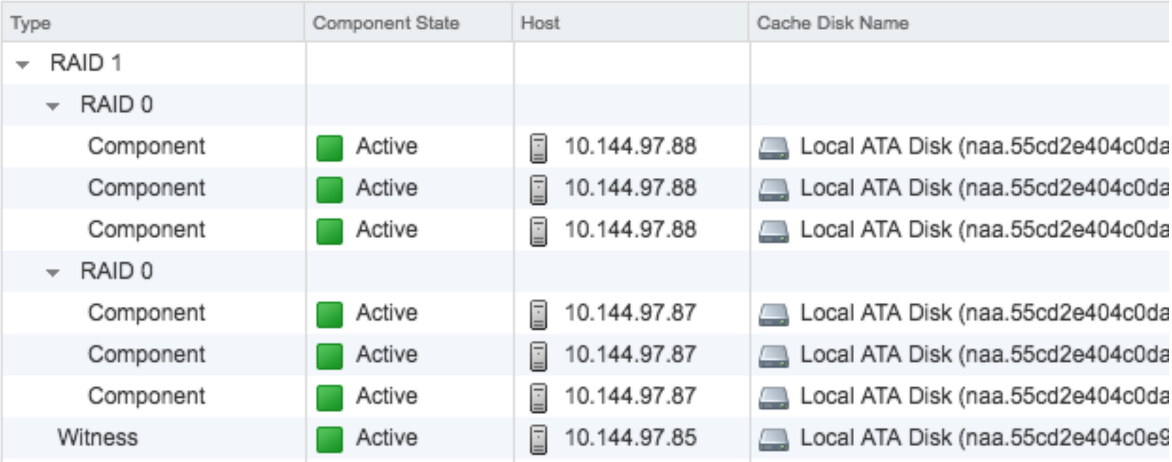


Type	Component State	Host	Cache Disk Name
RAID 0			
Component	🟩 Active	10.144.97.88	Local ATA Disk (naa.55cd2e404c0da
Component	🟩 Active	10.144.97.88	Local ATA Disk (naa.55cd2e404c0da
Component	🟩 Active	10.144.97.88	Local ATA Disk (naa.55cd2e404c0da

Figure 28. 600GB virtual disk object split up into three components

vSAN will break down a large component into smaller components in certain cases to help balance capacity consumption across disks, optimize rebuild and resynchronize activities, and improve overall efficiency in the environment.

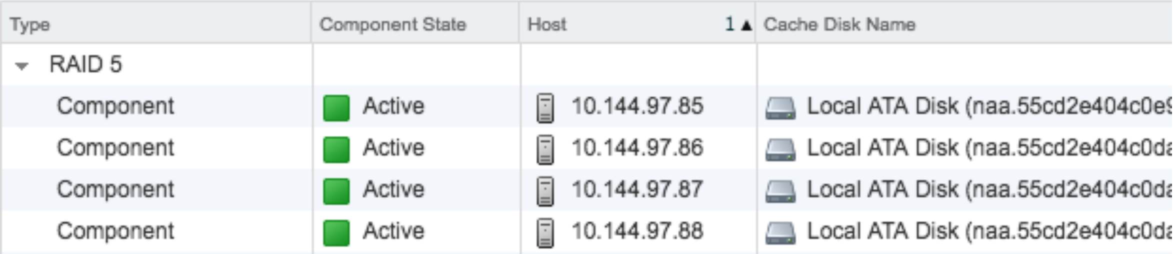
In most cases, a VM will have a storage policy assigned that contains availability rules such as Number of Failures to Tolerate and Failure Tolerance Method. These rules affect the number of components that make up an object. As an example, let's take that same 600GB virtual disk and apply the vSAN Default Storage Policy, which uses the RAID-1 mirroring failure tolerance method and has the number of failures to tolerate set to one. The 600GB object with three components will be mirrored on another host. This provides two full copies of the data distributed across two hosts so that the loss of a disk or an entire host can be tolerated. The figure below shows the six components (three on each host). A seventh component, a witness component, is created by vSAN to "break the tie" and achieve quorum in the event of a network partition between the hosts. The witness object is placed on a third host.



Type	Component State	Host	Cache Disk Name
RAID 1			
RAID 0			
Component	🟩 Active	10.144.97.88	Local ATA Disk (naa.55cd2e404c0da
Component	🟩 Active	10.144.97.88	Local ATA Disk (naa.55cd2e404c0da
Component	🟩 Active	10.144.97.88	Local ATA Disk (naa.55cd2e404c0da
RAID 0			
Component	🟩 Active	10.144.97.87	Local ATA Disk (naa.55cd2e404c0da
Component	🟩 Active	10.144.97.87	Local ATA Disk (naa.55cd2e404c0da
Component	🟩 Active	10.144.97.87	Local ATA Disk (naa.55cd2e404c0da
Witness	🟩 Active	10.144.97.85	Local ATA Disk (naa.55cd2e404c0e9

Figure 29. 600GB virtual disk object protected by RAID-1 mirroring

In this last example of component placement, we apply storage policy with RAID-5 erasure coding (Failures to Tolerate = 1) to an object. The object consists of four components—three data components and a parity component. These components are distributed across four hosts in the cluster. If disk or host containing any one of these components is offline, the data is still accessible. If one of these components is permanently lost, vSAN can rebuild the lost data or parity component from the other three surviving components.



Type	Component State	Host 1 ▲	Cache Disk Name
▾ RAID 5			
Component	🟩 Active	10.144.97.85	Local ATA Disk (naa.55cd2e404c0e
Component	🟩 Active	10.144.97.86	Local ATA Disk (naa.55cd2e404c0d
Component	🟩 Active	10.144.97.87	Local ATA Disk (naa.55cd2e404c0d
Component	🟩 Active	10.144.97.88	Local ATA Disk (naa.55cd2e404c0d

Figure 30. Virtual disk object protected by RAID-5 erasure coding

vSAN requires a minimum number of hosts depending on the failure tolerance method and number of failures to tolerate (FTT) configuration. For example, a minimum of three hosts are needed for FTT=1 with RAID-1 mirroring. A minimum of four hosts are required for FTT=1 with RAID-5 erasure coding. More details on cluster sizing minimums and recommendations can be found in the vSAN Design and Sizing Guide.

6.2 Fault Domains

"Fault domain" is a term that comes up often in availability discussions. In IT, a fault domain usually refers to a group of servers, storage, and/or networking components that would be impacted collectively by an outage. A common example of this is a server rack. If a top-of-rack switch or the power distribution unit for a server rack would fail, it would take all the servers in that rack offline even though the server hardware is functioning properly. That server rack is considered a fault domain.

Each host in a vSAN cluster is an implicit fault domain. vSAN automatically distributes components of a vSAN object across fault domains in a cluster based on the Number of Failures to Tolerate rule in the assigned storage policy. The following diagram shows a simple example of component distribution across hosts (fault domains). The two larger components are mirrored copies of the object and the smaller component represents the witness component.

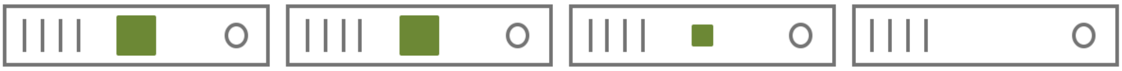


Figure 31. vSAN components distributed across hosts

vSAN Rack Awareness

The failure of a disk or entire host can be tolerated in the previous example scenario. However, this does not protect against the failure of larger fault domains such as an entire server rack. Consider out next example, which is a 12-node vSAN cluster. It is possible that multiple components that make up an object could reside in the same server rack. If there is a rack failure, the object would be offline.

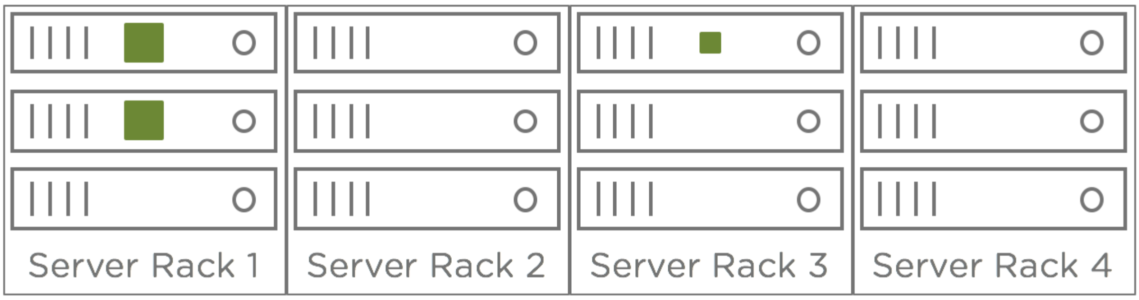


Figure 32. Multiple vSAN components in a server rack

To mitigate this risk, place the servers in a vSAN cluster across server racks and configure a fault domain for each rack in the vSAN UI. This instructs vSAN to distribute components across server racks to eliminate the risk of a rack failure taking multiple objects offline. This feature is commonly referred to as "Rack Awareness". The diagram below shows component placement when three servers in each rack are configured as separate vSAN fault domains.

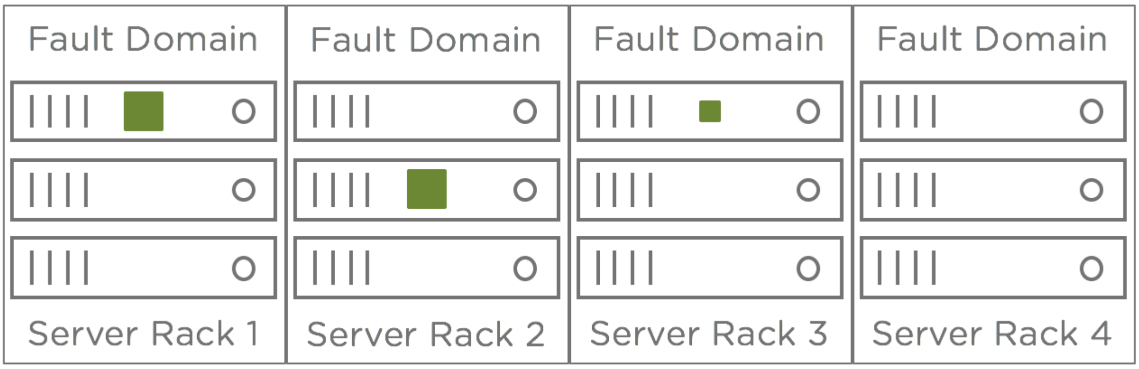


Figure 33. vSAN components distributed across fault domains

6.3 Rebuild and Resynchronize

vSAN achieves high availability and extreme performance through the distribution of data across multiple hosts in a cluster. Data is transmitted between hosts using the vSAN network. There are cases where a significant amount of data must be copied across the vSAN network. One example is when you change the fault tolerance method in a storage policy from RAID-1 mirroring to RAID-5 erasure coding. vSAN copies or “resynchronizes” the mirrored components to a new set of striped components.

Another example is repair operations such as when a vSAN components are offline due to a host hardware issue. These components are marked “absent” and colored orange in the vSAN user interface. vSAN waits 60 minutes by default before starting the repair operation. vSAN has this delay as many issues are transient. In other words, vSAN expects absent components to be back online in a reasonable amount of time and we want to avoid copy large quantities of data unless it is truly necessary. An example is a host being temporarily offline due to an unplanned reboot.

vSAN will begin the repair process for absent components after 60 minutes to restore redundancy. For example, an object such as a virtual disk (VMDK file) protected by a RAID-1 mirroring storage policy will create a second mirror copy from the healthy copy. This process can take a considerable amount of time depending on how much data must be copied. The rebuild process continues even if the absent copy comes back online in versions of vSAN prior to 6.6.

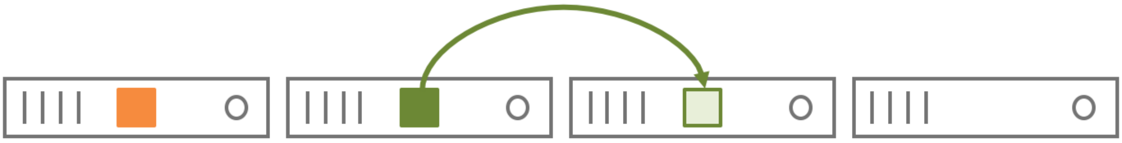


Figure 34. vSAN Component Rebuild

The object repair process is improved in vSAN 6.6. If absent components come back online while vSAN is rebuilding another copy, vSAN will determine whether it is more efficient to continue building an entirely new copy or update the existing copy that came back online. vSAN will restore redundancy using the most efficient method and cancel the other action. This enhancement in vSAN 6.6 rebuilds improves the speed and efficiency of object repair operations to reduce risk and minimize resource usage.

In cases where there are not enough resources online to comply with all storage policies, vSAN 6.6 will repair as many objects as possible. This helps ensure the highest possible levels of redundancy in environments affected by unplanned downtime. When additional resources come back online, vSAN will continue the repair process to comply with storage policies.

There are a few other operations that can temporarily increase vSAN “backend” traffic flow. Rebalancing of disk utilization is one of these operations. When a disk has less than 20% free space, vSAN will automatically attempt to balance capacity utilization by moving data from that disk to other disks in the vSAN cluster. Achieving a well-balanced cluster from a disk capacity standpoint can be more challenging if there are many large components. vSAN 6.6 improves efficiency by splitting large components into smaller components to achieve a better balance.

Excessive amounts of vSAN backend resynchronization traffic might affect cluster performance. Resynchronization operations in previous versions of vSAN are automated and controlled entirely by vSAN. In other words, administrators are unable to adjust the rate at which resynchronization operations are performed.

vSAN 6.6 introduces the option to adjust the throughput of resynchronization operations. If cluster performance is being severely impacted by excessive resynchronizing activity, you can minimize the impact by reducing the throughput allowed for resynchronizing.

Note: It is highly recommended to use the default setting (throttling disabled). Throttling resynchronization traffic will increase the amount of time needed to rebuild and/or resynchronization components, which increases the risk of downtime.

6.4 Stretched Clusters

vSAN stretched clusters provide resiliency against the loss of an entire site. vSAN is integrated tightly with vSphere HA. If a site goes offline unexpectedly, vSphere HA will automatically restart the virtual machines affected by the outage at the other site with no data loss. The virtual machines will begin the restart process in a matter of seconds, which minimizes downtime.

vSAN stretched clusters are also beneficial in planned downtime and disaster avoidance situations. Virtual machines at one site can be migrated to the other site with VMware vMotion. Issues such as an impending storm or rising flood waters typically provide at least some time to prepare before disaster strikes. Virtual machines can easily be migrated out of harm’s way in a vSAN stretched cluster environment.

Deployment

vSAN features the capability to create a stretched cluster across two sites. The two sites could be separate rooms at opposite ends of the same building, two buildings on the same campus, two campuses in separate cities, and so on. There are many possibilities.

The limitations of what is possible centers on network bandwidth and round trip time (RTT) latency. Nearly all stretched cluster solutions need a RTT latency of 5 ms or less. Writes to both sites must be committed before the writes are acknowledged. RTT latencies higher than 5 ms introduce performance issues. vSAN is no exception. A 10Gbps network connection with 5 ms RTT latency or less is required between the preferred and secondary sites of a vSAN stretched cluster.

Up to 15 hosts per site is currently supported. In addition to the hosts at each site, a "witness" must be deployed to a third site. The witness is a VM appliance running ESXi that is configured specifically for use with a vSAN stretched cluster. Its purpose is to enable the cluster to achieve quorum when one of the two main data sites is offline. The witness also acts as “tie-breaker” in scenarios where a network partition occurs between the two data sites. This is sometimes referred to as a “split-brain” scenario. The witness does not store virtual machine data such as virtual disks. Only metadata such as witness components are stored on the witness.

Up to 200 ms RTT latency is supported between the witness site and data sites. The bandwidth requirements between the witness site and data sites varies and depends primarily on the number of vSAN objects stored at each site. A minimum bandwidth of 100Mbps is required and the general rule of thumb is at least 2Mbps of available bandwidth for every 1000 vSAN objects. The vSAN Stretched Cluster Bandwidth Sizing guide provides more details on networking requirements.

Fault Domains

A vSAN stretched cluster consists of exactly three primary fault domains. The physical hosts in the primary or “preferred” location make up one fault domain; the physical hosts in the secondary location are the second fault domain; the witness is the third fault domain placed at a tertiary location.

When a virtual machine is deployed to a vSAN stretched cluster, a RAID-1 mirroring policy with FTT=1 is applied to the virtual machine by default. One copy of the virtual machine is located at the preferred site. Another copy is placed at the secondary site. Metadata (vSAN witness objects) is placed on the witness virtual appliance at a third site. If any one of the sites go offline, there are enough surviving components to achieve quorum so the virtual machine is still accessible.

Starting with vSAN 6.6, it is possible to configure a secondary level of failures to tolerate. This feature enables resiliency within a site, as well as, across sites. For example, RAID-5 erasure coding protects objects within the same site while RAID-1 mirroring protects these same objects across sites.

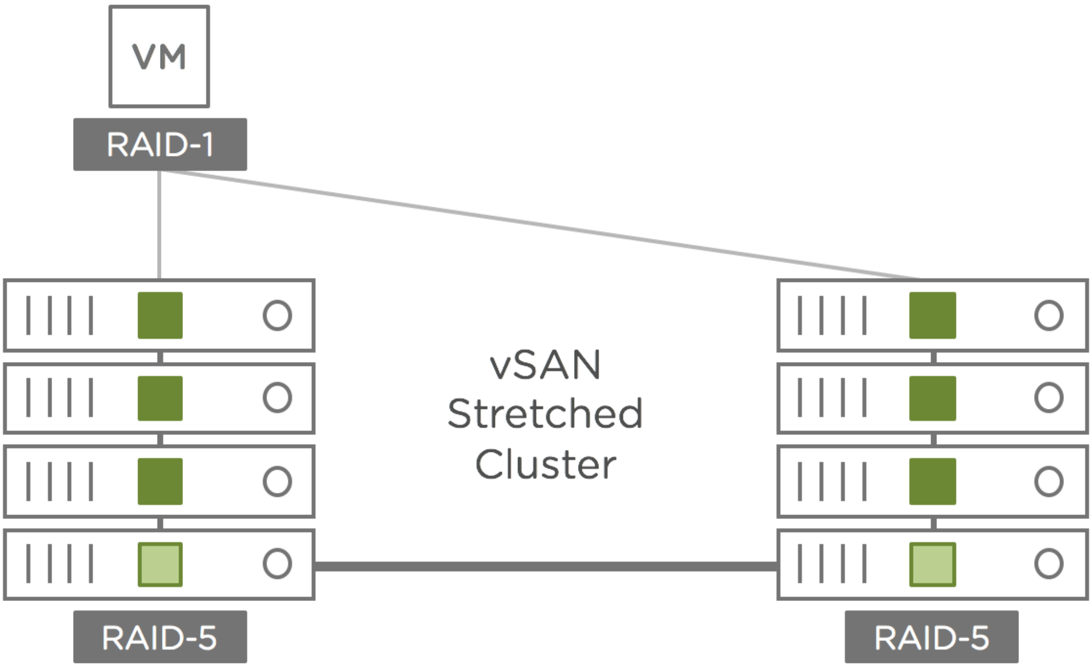
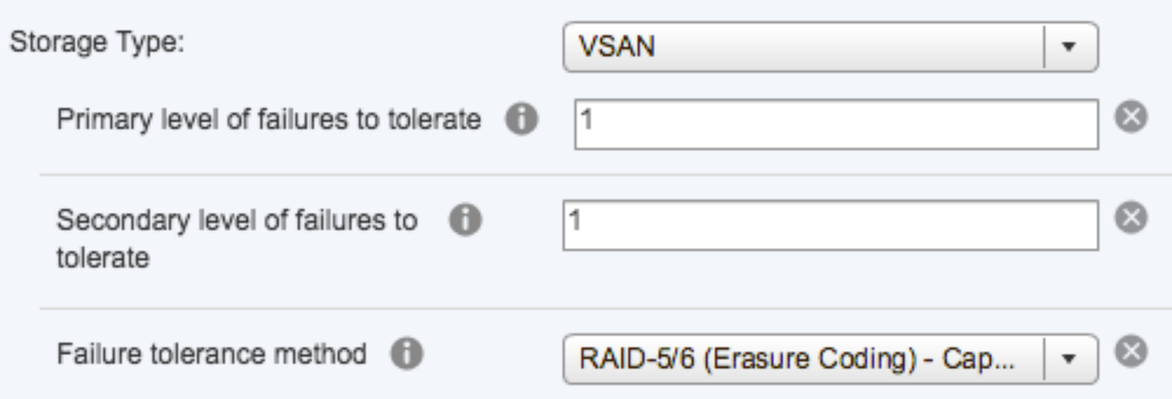


Figure 35. Stretched cluster with RAID-5 erasure coding local failure protection

Local failure protection within a vSAN stretched cluster further improves the resiliency of the cluster to minimize unplanned downtime. This feature also reduces or eliminates cross-site traffic in cases where components need to be resynchronized or rebuilt. vSAN lowers the total cost of ownership of a stretched cluster solution as there is no need to purchase additional hardware or software to achieve this level of resiliency.

This is configured and managed through a storage policy in the vSphere Web Client. The figure below shows rules in a storage policy that is part of an all-flash stretched cluster configuration. The primary level of failures to tolerate is set to 1, which instructs vSAN to mirror data across the two main sites of the stretched cluster. The secondary level of failures to tolerate specifies how data is protected within the site. In the example storage policy below, RAID-5 erasure coding is used, which can tolerate the loss of a host within the site.



Storage Type: VSAN

Primary level of failures to tolerate: 1

Secondary level of failures to tolerate: 1

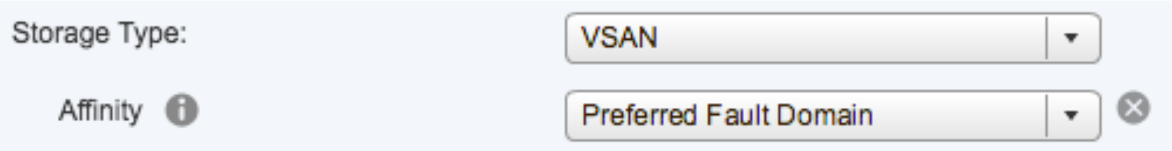
Failure tolerance method: RAID-5/6 (Erasure Coding) - Cap...

Figure 36. Stretched Cluster with Local Failure Protection Storage Policy Rule

Site Affinity

vSAN 6.6 improves the flexibility of storage policy-based management for stretched clusters by introducing the Affinity rule. You can specify a single site to locate virtual machine objects in cases where cross-site redundancy is not necessary. Common examples include applications that have built-in replication or redundancy such as Microsoft Active Directory and Oracle Real Application Clusters (RAC). This capability reduces costs by minimizing the storage and network resources used by these workloads.

Affinity is easy to configure and manage using storage policy-based management. A storage policy is created and the Affinity rule is added to specify the site where a virtual machine's objects will be stored.



Storage Type: VSAN

Affinity: Preferred Fault Domain

Figure 37. Stretched Cluster Site Affinity Storage Policy Rule

6.5 Degraded Device Handling

VMware continues to improve how vSAN handles hardware issues such as a storage device that is showing symptoms of impending failure. In some cases, storage devices issues are easily detected through errors reported by the device, e.g., SCSI sense codes. In other cases, issues are not so obvious. To proactively discover these types of issues, vSAN will track performance characteristics of a device over time. A significant discrepancy in performance is a good indicator of a potential problem with a device. vSAN uses multiple samples of data to help avoid “false positives” where the issue is transient in nature.

When failure of a device is anticipated, vSAN evaluates the data on the device. If there are replicas of the data on other devices in the cluster, vSAN will mark these components as “absent”. “Absent” components are not rebuilt immediately as it is possible the cause of the issue is temporary. vSAN waits for 60 minutes by default before starting the rebuild process. This does not affect the availability of a virtual machine as the data is still accessible using one or more other replicas in the cluster. If the only replica of data is located on a suspect device, vSAN will immediately start the evacuation of this data to other healthy storage devices.

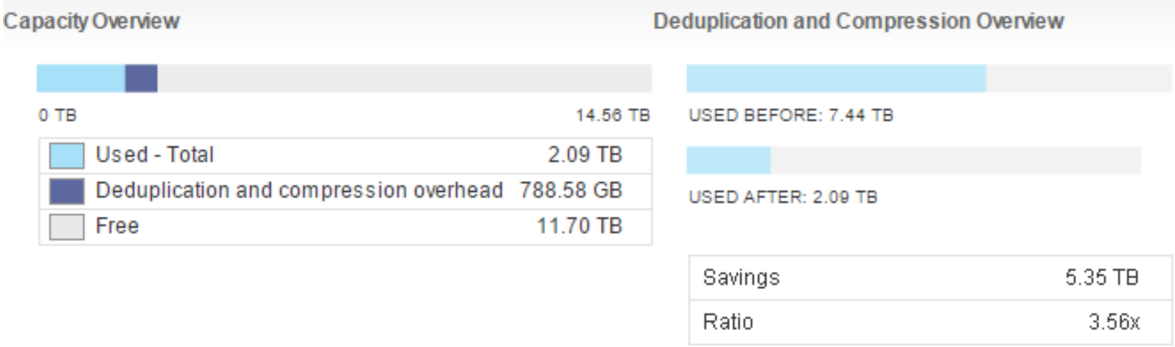
Intelligent, predictive failure handling drives down the cost of operations by minimizing the risk of downtime and data loss.

7. Capacity Reporting

.

7.1 Capacity Overview and Breakdown

Capacity overviews make it easy for administrators to see used and free space at a glance. Deduplication and compression information is displayed. Information is also available showing how much capacity various object types are consuming.



Capacity Overview

🟦	Used - Total	2.09 TB
🟦	Deduplication and compression overhead	788.58 GB
⬜	Free	11.70 TB

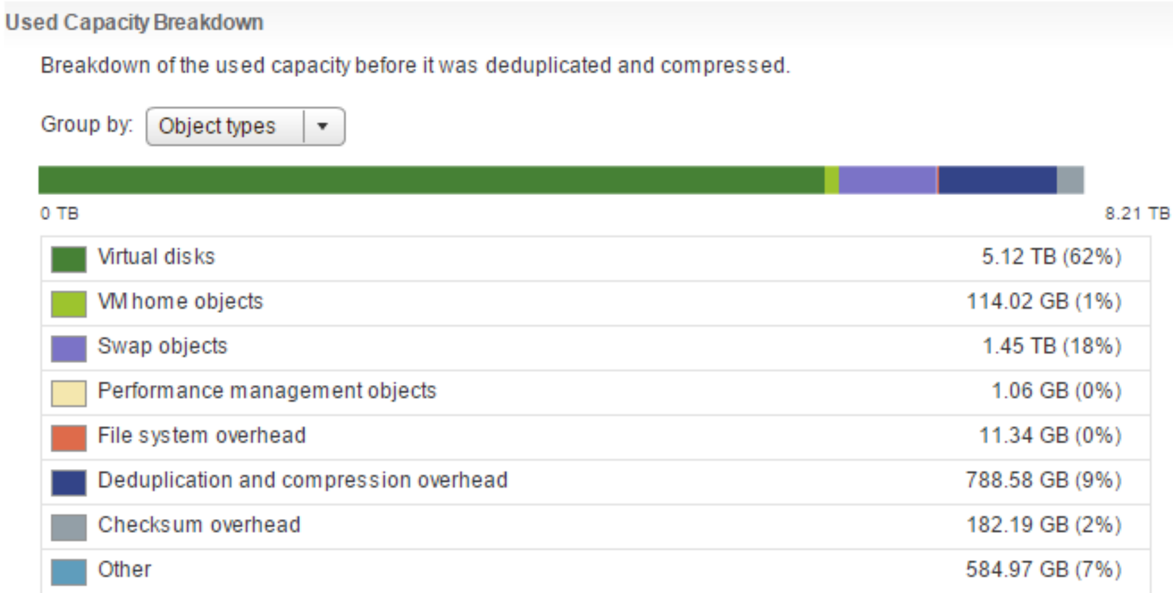
Deduplication and Compression Overview

USED BEFORE: 7.44 TB

USED AFTER: 2.09 TB

Savings	5.35 TB
Ratio	3.56x

Figure 38. Capacity overview charts



Used Capacity Breakdown

Breakdown of the used capacity before it was deduplicated and compressed.

Group by: Object types

🟩	Virtual disks	5.12 TB (62%)
🟩	VM home objects	114.02 GB (1%)
🟪	Swap objects	1.45 TB (18%)
🟨	Performance management objects	1.06 GB (0%)
🟧	File system overhead	11.34 GB (0%)
🟦	Deduplication and compression overhead	788.58 GB (9%)
⬜	Checksum overhead	182.19 GB (2%)
🟦	Other	584.97 GB (7%)

Figure 39. Used capacity breakdown

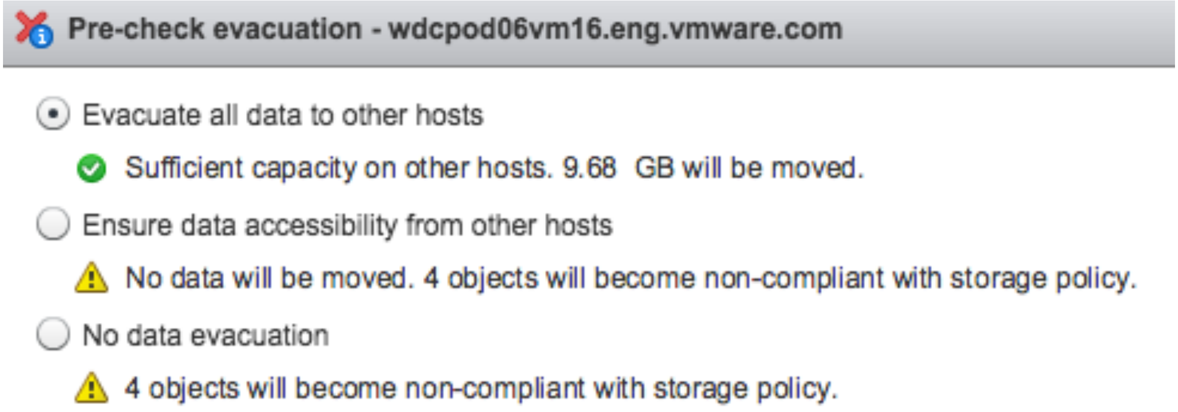
Note: Percentages are of used capacity, not of total capacity.

This following list provides more details on the object types in the Used Capacity Breakdown chart.

- Virtual disks: Virtual disk consumption before deduplication and compression
- VM home objects: VM home object consumption before deduplication and compression
- Swap objects: Capacity utilized by virtual machine swap files
- Performance management objects: When the vSAN performance service is enabled, this is the amount of capacity used to store the performance data
- File system overhead: Capacity required by the vSAN file system metadata
- Deduplication and compression overhead: deduplication and compression metadata, such as hash, translation, and allocation maps
- Checksum overhead: Capacity used to store checksum information
- Other: Virtual machine templates, unregistered virtual machines, ISO files, and so on that are consuming vSAN capacity

7.2 Host Evacuation

A maintenance mode pre-check is included in vSAN 6.6 to help ensure adequate capacity remains after a host is evacuated. This function is also used when removing a disk or disk group. Conducting this pre-check prior to evacuating a host provides a better understanding of how the cluster will be impacted from a capacity standpoint. Changes in storage policy compliance are also indicated. This is one more example of how vSAN reduces complexity, minimizes risk, and lowers operational overhead.



Pre-check evacuation - wdcpod06vm16.eng.vmware.com

● Evacuate all data to other hosts

✓ Sufficient capacity on other hosts. 9.68 GB will be moved.

○ Ensure data accessibility from other hosts

⚠ No data will be moved. 4 objects will become non-compliant with storage policy.

○ No data evacuation

⚠ 4 objects will become non-compliant with storage policy.

Figure 40. Pre-Check Evacuation

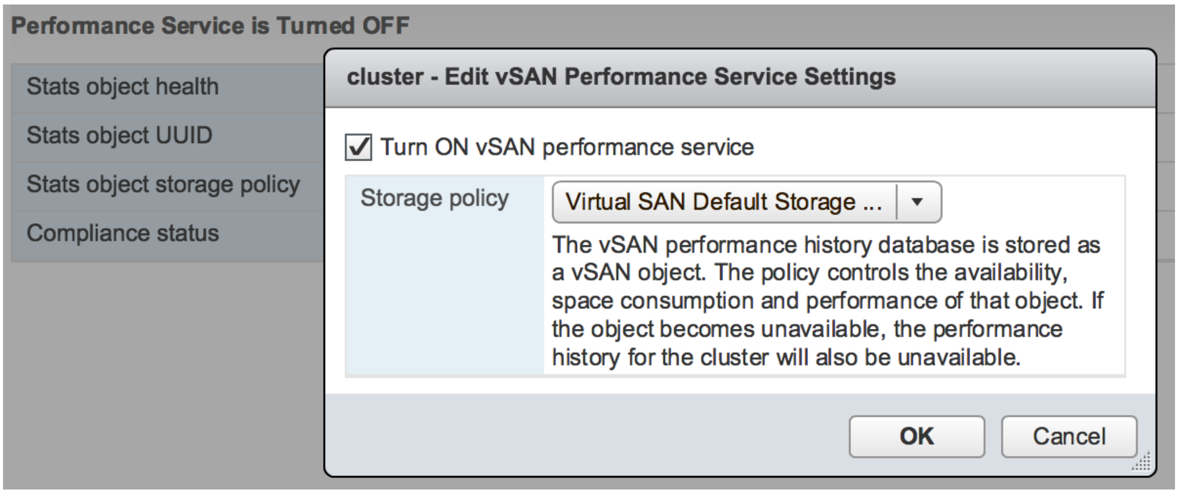
8. Performance Reporting

.

8.1 Turning on the Performance Service

A healthy vSAN environment is one that is performing well. vSAN includes many graphs that provide performance information at the cluster, host, network adapter, virtual machine, and virtual disk levels. There are many data points that can be viewed such as IOPS, throughput, latency, packet loss rate, write buffer free percentage, cache de-stage rate, and congestion. Time range can be modified to show information from the last 1-24 hours or a custom date and time range. It is also possible to save performance data for later viewing.

The performance service is enabled at the cluster level. The performance service database is stored as a vSAN object independent of vCenter Server. A storage policy is assigned to the object to control space consumption and availability of that object. If it becomes unavailable, performance history for the cluster cannot be viewed until access to the object is restored.



Performance Service is Turned OFF

- Stats object health
- Stats object UUID
- Stats object storage policy
- Compliance status

cluster - Edit vSAN Performance Service Settings

☑ Turn ON vSAN performance service

Storage policy: Virtual SAN Default Storage ...

The vSAN performance history database is stored as a vSAN object. The policy controls the availability, space consumption and performance of that object. If the object becomes unavailable, the performance history for the cluster will also be unavailable.

Figure 41. Turn on vSAN performance service

Note: The performance service is turned off by default.

8.2 Metrics

Cluster Metrics

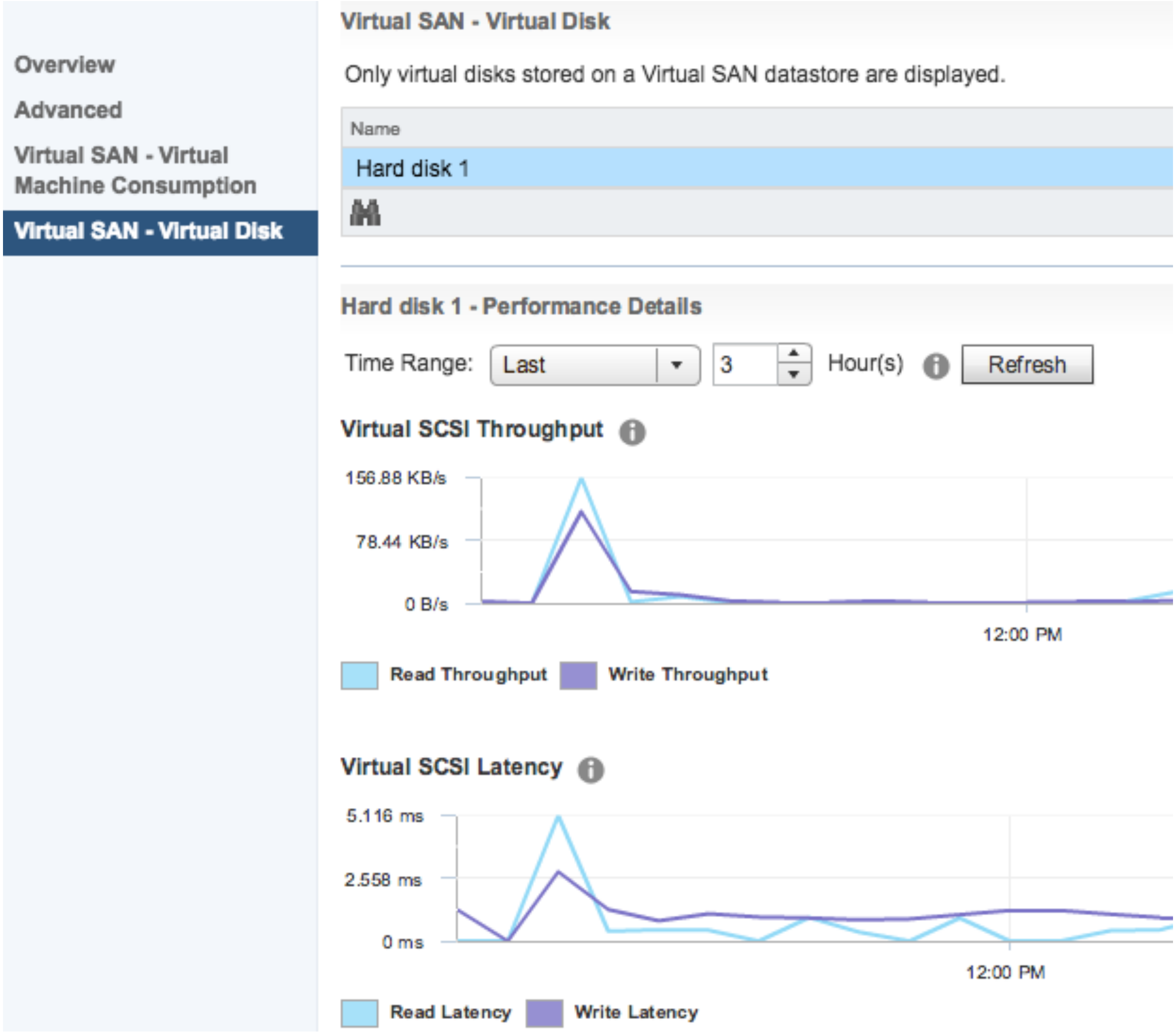
These metrics provide visibility to the entire vSAN cluster. Graphs show IOPs, throughput, latency, congestion, and outstanding I/O. “vSAN – Virtual Machine Consumption” graphs show metrics generated by virtual machines across the cluster. In addition to normal virtual machines reads and writes, “vSAN – Backend” consumption adds traffic such as metadata updates, component rebuilds, and data migrations.

Host Metrics

In addition to virtual machine and backend metrics, disk group, individual disk, physical network adapter, and VMkernel adapter performance information is provided at the host level. Seeing metrics for individual disks eases the process of troubleshooting issues such as failed storage devices. Throughput, packets per second, and packet loss rate statistics for network interfaces help identify potential networking issues.

Virtual Machine Metrics

Virtual machine metrics include IOPS, throughput, and latency. It is also possible to view information at the virtual disk level. The figure below shows virtual disk-level Virtual SCSI throughput and latencies for reads and writes.



Overview
Advanced
Virtual SAN - Virtual Machine Consumption
Virtual SAN - Virtual Disk

Virtual SAN - Virtual Disk

Only virtual disks stored on a Virtual SAN datastore are displayed.

Name
Hard disk 1

Hard disk 1 - Performance Details

Time Range: Last 3 Hour(s)

Figure 42. Virtual disk performance metrics

9. Space Efficiency

.

9.1 Reducing Total Cost of Ownership

Space efficiency features such as deduplication, compression, and erasure coding reduce the total cost of ownership (TCO) of storage. Even though flash capacity is currently more expensive than magnetic disk capacity, using space efficiency features makes the cost-per-usable-gigabyte of flash devices the same as or lower than magnetic drives. Add in the benefits of higher flash performance and it is easy to see why all-flash configurations are more popular.

9.2 Deduplication and Compression

Enabling deduplication and compression can reduce the amount of physical storage consumed by as much as 7x. Environments with redundant data such as homogenous operating systems typically benefit the most. Likewise, compression offers more favorable results with data that compresses well like text, bitmap, and program files. Data that is already compressed such as certain graphics formats and video files, as well as files that are encrypted, will yield little or no reduction in storage consumption from compression. Deduplication and compression results will vary based on the types of data stored in an all flash vSAN environment.

Deduplication and compression is a single cluster-wide setting that is disabled by default and can be enabled using a simple drop-down menu.

Note: A rolling format of all disks in the vSAN cluster is required when deduplication and compression are enabled on an existing cluster. This can take a considerable amount of time. However, this process does not incur virtual machine downtime.

Deduplication and compression are implemented after writes are acknowledged in the vSAN cache tier to minimize impact to performance. The deduplication algorithm utilizes a 4K fixed block size and is performed within each disk group. In other words, redundant copies of a block within the same disk group are reduced to one copy, but redundant blocks across multiple disk groups are not deduplicated.

“Cold” data in the cache tier that is ready to be de-staged is moved to memory were it is deduplicated and compressed and then it is written to the capacity tier.

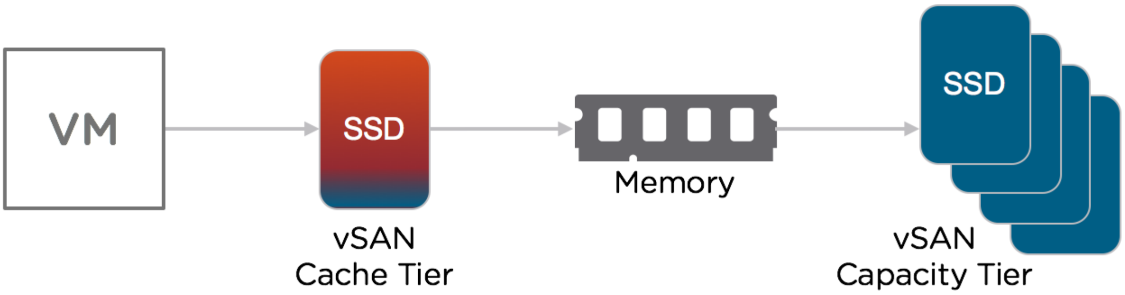


Figure 43. De-staging with deduplication and compression enabled

The compression algorithm is applied after deduplication has occurred just before the data is written to the capacity tier. Considering the additional compute resource and allocation map overhead of compression, vSAN will only store compressed data if a 4K block can be reduced to 2K or less. Otherwise, the block is written uncompressed to avoid the use of additional resources.

The processes of deduplication and compression on any storage platform incur overhead and potentially impact performance in terms of latency and maximum IOPS. vSAN is no exception.

However, considering deduplication and compression are only supported in all flash vSAN configurations, these effects are predictable in most use cases. The extreme performance and low latency of flash devices easily outweigh the additional resource requirements of deduplication and compression. The space efficiency generated by deduplication and compression lowers the cost-per-usable-GB of all flash configurations.

9.3 RAID-5/6 Erasure Coding

RAID-5/6 erasure coding is a space efficiency feature optimized for all flash configurations. Erasure coding provides the same levels of redundancy as mirroring, but with a reduced capacity requirement. In general, erasure coding is a method of taking data, breaking it into multiple pieces and spreading it across multiple devices, while adding parity data so it may be recreated in the event one of the pieces is corrupted or lost.

Unlike deduplication and compression, which offer variable levels of space efficiency, erasure coding guarantees capacity reduction over a mirroring data protection method at the same failure tolerance level. As an example, let's consider a 100GB virtual disk. Surviving one disk or host failure requires 2 copies of data at 2x the capacity, i.e., 200GB. If RAID-5 erasure coding is used to protect the object, the 100GB virtual disk will consume 133GB of raw capacity—a 33% reduction in consumed capacity versus RAID-1 mirroring.

RAID-5 erasure coding requires a minimum of four hosts. Let's look at a simple example of a 100GB virtual disk. When a policy containing a RAID-5 erasure coding rule is assigned to this object, three data components and one parity component are created. To survive the loss of a disk or host (FTT=1), these components are distributed across four hosts in the cluster.

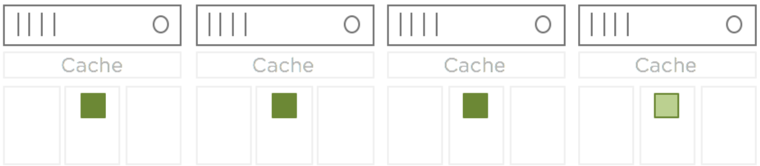


Figure 44. RAID-5 erasure coding

RAID-6 erasure coding requires a minimum of six hosts. Using our previous example of a 100GB virtual disk, the RAID-6 erasure coding rule creates four data components and two parity components. This configuration can survive the loss of two disks or hosts simultaneously (FTT=2)

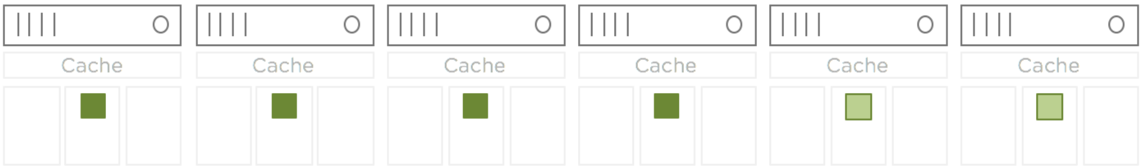


Figure 45. RAID-6 erasure coding

While erasure coding provides significant capacity savings over mirroring, understand that erasure coding requires additional processing overhead. This is common with any storage platform. Erasure coding is only supported in all flash vSAN configurations. Therefore, performance impact is negligible in most use cases due to the inherent performance of flash devices.

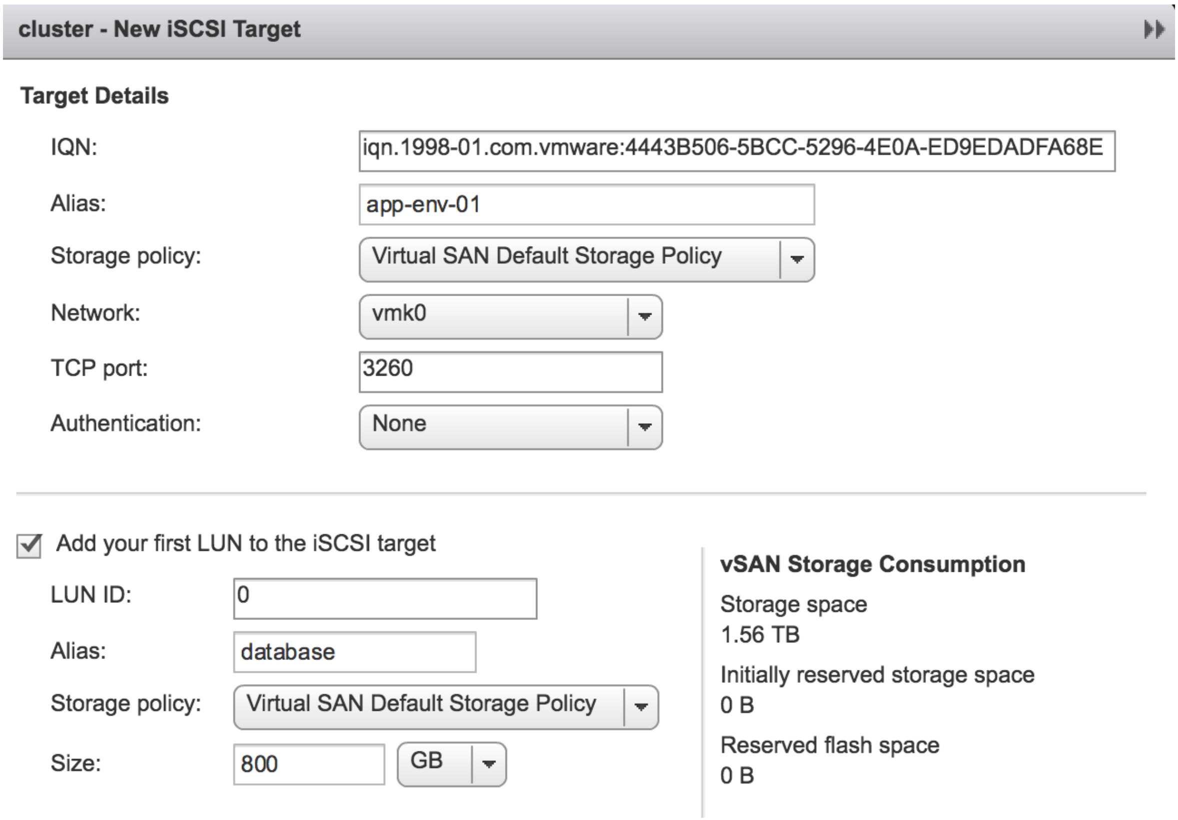
10. vSAN for Physical Workloads

.

10.1 iSCSI Target Service

Block storage can be provided to physical workloads using the iSCSI protocol. The vSAN iSCSI target service provides flexibility and potentially avoids expenditure on purpose-built, external storage arrays. In addition to capital cost savings, the simplicity of vSAN lowers operational costs.

The vSAN iSCSI target service is enabled with just a few mouse clicks. CHAP and Mutual CHAP authentication is supported. vSAN objects that serve as iSCSI targets are managed with storage policies just like virtual machine objects. After the iSCSI target service is enabled, iSCSI targets and LUNs can be created. The figure below shows vSAN iSCSI target configuration.



cluster - New iSCSI Target

Target Details

IQN: iqn.1998-01.com.vmware:4443B506-5BCC-5296-4E0A-ED9EDADFA68E

Alias: app-env-01

Storage policy: Virtual SAN Default Storage Policy

Network: vmk0

TCP port: 3260

Authentication: None

☑ Add your first LUN to the iSCSI target

LUN ID: 0

Alias: database

Storage policy: Virtual SAN Default Storage Policy

Size: 800 GB

vSAN Storage Consumption

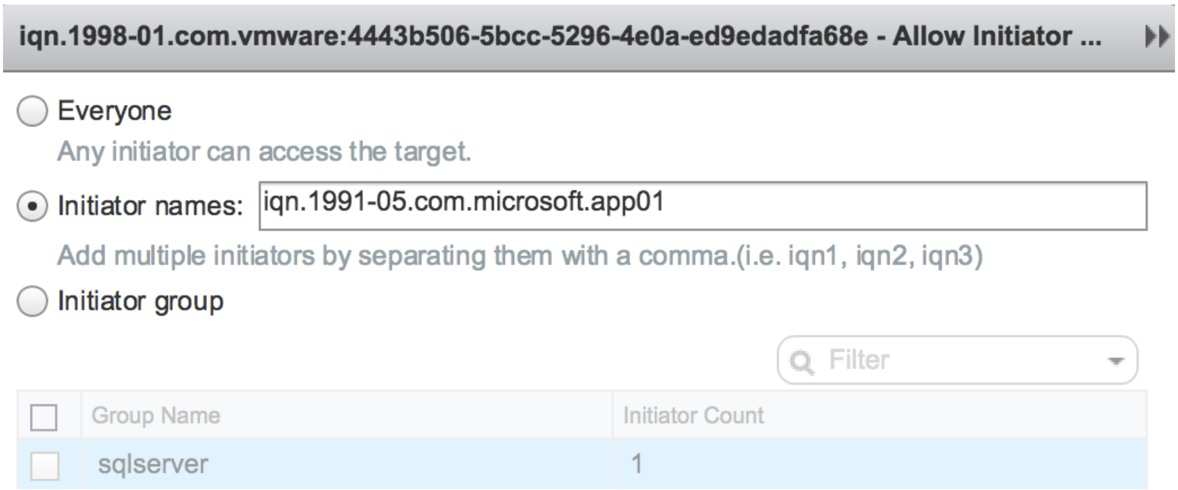
Storage space
1.56 TB

Initially reserved storage space
0 B

Reserved flash space
0 B

Figure 46. vSAN iSCSI target configuration

The last step is adding initiator names or an initiator group, which controls access to the target. It is possible to add individual names or create groups for ease of management.



iqn.1998-01.com.vmware:4443b506-5bcc-5296-4e0a-ed9edadfa68e - Allow Initiator ...

○ Everyone

Any initiator can access the target.

● Initiator names: iqn.1991-05.com.microsoft.app01

Add multiple initiators by separating them with a comma.(i.e. iqn1, iqn2, iqn3)

○ Initiator group

Filter

□	Group Name	Initiator Count
□	sqlserver	1

Figure 47 Adding iSCSI initiator names

In nearly all cases, it is best to run workloads in virtual machines to take full advantage of vSAN's simplicity, performance, and reliability. However, for those use cases that truly need block storage, it is now possible to utilize vSAN iSCSI Target Service.

10.2 Certified File Services Solutions

Certified solutions for file services are available through the VMware Ready for vSAN™ program. Customers can deploy these solutions with confidence to extend HCI environments with proven, industry-leading solutions. Using these solutions with vSAN provides benefits such as simplified setup and management, documented recommendations, and robust support. The diagram below shows a virtual storage appliance providing standard file services to physical workloads.

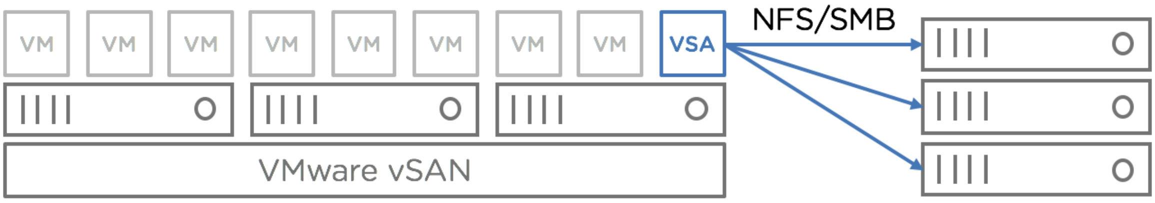


Figure 48. Physical workloads using vSAN

11. Quality of Service

.

11.1 Limiting IOPS

vSAN can limit the number of IOPS a virtual machine or virtual disk generates. There are situations where it is advantageous to limit the IOPS of one or more virtual machines. The term noisy neighbor is often used to describe when a workload monopolizes available IO or other resources, which negatively impact other workloads or tenants in the same environment.

An example of a possible noisy neighbor scenario is month-end reporting. Management requests delivery of these reports on the second day of each month so the reports are generated on the first day of each month. The virtual machines that run the reporting application and database are dormant most of the time. Running the reports take just a few hours, but this generates very high levels of storage I/O. The performance of other workloads in the environment are sometimes impacted while the reports are running. To remedy this issue, an administrator creates a storage policy with an IOPS limit rule and assigns the policy to the virtual machines running the reporting application and database.

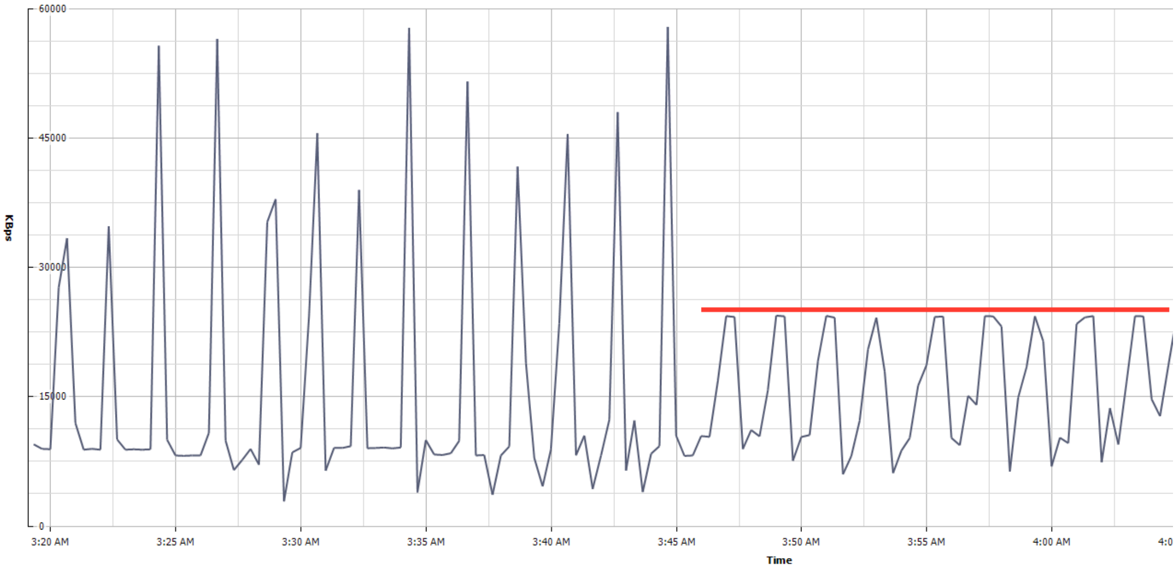


Figure 49. Graph showing IOPS limit

The IOPS limit eliminates possible performance impact to the other virtual machines. The reports take longer, but they are still finished in plenty of time for delivery the next day.

Keep in mind storage policies can be dynamically created, modified, and assigned to virtual machines. If an IOPS limit is proving to be too restrictive, simply modify the existing policy or create a new policy with a different IOPS limit and assign it to the virtual machines. The new or updated policy will take effect just moments after the change is made.

12. Cloud Native Application Storage

.

12.1 New Application Technologies

New application architecture and development methods have emerged that are designed to run in today's mobile-cloud era. Container technologies such as Docker and Kubernetes are a couple of the many solutions that have emerged as options for deploying and orchestrating these applications. VMware is embracing these new application types with a number products and solutions. Here are a few examples:

- Photon OS – a minimal Linux container host optimized to run on VMware platforms
- Photon Controller – a distributed, multi-tenant ESXi host controller for containers
- Lightwave – an enterprise-grade identity and access management services
- Admiral – a highly scalable and very lightweight container management platform

Cloud native applications naturally require persistent storage just the same as traditional applications. Deploying vSAN for Photon Platform enables the use of a vSAN cluster in cloud native application environments managed by Photon Controller.

VMware vSphere Integrated Containers™ provides enterprise container infrastructure to help IT Ops run both traditional and containerized applications side-by-side on a common platform. Supporting containers in a virtualized environments provides a number of benefits: IT teams get the security, isolation and management of VMs, while developers enjoy the speed and agility of containers—all within the familiar vSphere platform. Availability and performance features in vSphere and vSAN can be utilized by vSphere Integrated Containers just the same as traditional storage environments.

VMware vSphere Docker Volume Service enables vSphere users to create and manage Docker container data volumes on vSphere storage technologies such as VMFS, NFS, and vSAN. This driver makes it very simple to use containers with vSphere storage and provides the following key benefits:

- DevOps-friendly API for provisioning and policy configuration.
- Seamless movement of containers between vSphere hosts without moving data.
- Single platform to manage—run virtual machines and containers side-by-side on the same vSphere infrastructure.

vSAN along with the solutions above provides an ideal storage platform for developing, deploying, and managing cloud native applications.

13. Summary

.

13.1 HCI Powered by vSAN

Hyper-Converged Infrastructure, or HCI, consolidates traditional IT infrastructure silos onto industry-standard servers. The physical infrastructure is virtualized to help evolve data centers without risk, reduce total cost of ownership (TCO), and scale to tomorrow with timely support for new hardware, applications, and cloud strategies.

HCI solutions powered by VMware consist of a single, integrated platform for storage, compute and networking that build on the foundation of VMware vSphere, the market-leading hypervisor, and VMware vSAN, the software-defined enterprise storage solution natively integrated with vSphere. vCenter Server provides a familiar unified, extensible management solution.

Seamless integration with vSphere and the VMware ecosystem makes it the ideal storage platform for business-critical applications, disaster recovery sites, remote office and branch office (ROBO) implementation, test and development environments, management clusters, security zones, and virtual desktop infrastructure (VDI). Today, customers of all industries and sizes trust vSAN to run their most important applications.

VMware provides the broadest choice of consumption options for HCI:

- VMware Cloud Foundation™, a unified platform that brings together VMware's vSphere, vSAN and VMware NSX™ into an integrated stack for private and public clouds
- Dell EMC VxRail™ and VxRack SDDC™, turn-key HCI solutions tailored for ease of use, rapid deployment, and scale
- vSAN ReadyNodes™ with 200+ systems from all major server vendors catering to flexibility of deployment needs and vendor preferences

vSAN 6.6 focuses on enabling customers to modernize their infrastructure by enhancing three key areas of today's IT need: higher security, lower cost, and faster performance.

The industry's first native encryption solution for HCI is delivered through vSphere and vSAN. A highly available control plane is built in to help organizations minimize risk without sacrificing flash storage efficiencies.

vSAN lowers TCO by providing highly available, powerful, and economical stretched clusters that are 60% less than leading legacy storage solutions. Operational costs are also reduced with new, intelligent operations that introduce 1-click hardware updates for predictable hardware experience and pro-active cloud health checks for custom, real-time support.

vSAN is designed to scale to tomorrow's IT needs by optimizing flash performance for traditional and next-generation workloads. This enables organizations to realize the benefits of vSAN for all workloads.