vSAN Operations Guide

Recommendations for vSAN in VMware Cloud Foundation 9.1

May 5, 2026

vmware by Broadcom

Table of Contents

Introduction 5

Scope of Topics 5

Section 1: Cluster 5

Pre-Flight Checks Prior to Introducing Clusters into Production 5

Maintenance Work for L2/L3 Switching in Production 5

Configuring Fault Domains 7

Migrate to a Different vSAN Cluster Type 8

Section 2: Network 9

Using NIOC for vSAN 9

Using Large MTU Sizes with vSAN 10

Create and Manage Broadcast Domains for Multiple vSAN Clusters 11

Change IP Addresses of Hosts in a vSAN Cluster 12

Migrate vSAN Traffic to Another VMkernel Port 12

Introducing RDMA into a vSAN Environment 14

Section 3: Storage Devices 16

Adding and Removing Storage Devices in vSAN ESA 16

Secure Erase of Data on a Decommissioned vSAN Storage Device 16

Section 4: vSAN Datastore 18

Capacity Management Guidance 18

Automatic Rebalancing in a vSAN Cluster 18

Managing Orphaned Objects in a Datastore 19

Section 5: Storage Policies 20

Using Workload Limiting Polies (IOPS Limits) 20

Using Object Space Reservation (OSR) 20

Section 6: Host and EMM Operations 20

What EMM Option to Choose for Host Maintenance 20

Restarting a Host 22

Cluster Shutdown and Power-Up 22

Section 7: Guest VM Operations 24

Configuring TRIM/UNMAP in vSAN 24

Tuning of Workloads after Migration to vSAN 25

Section 8: Data Services 28

Compression (ESA) 28

Global Deduplication (ESA) 28
Data-at-Rest Encryption: Enabling and Disabling 28
Data-in-Transit Encryption: Enabling in-flight Encryption on a vSAN Cluster 29
Data-at-Rest Encryption: Using vSAN and vSphere Encryption Together 29
Data-at-Rest Encryption: Performing a Shallow Rekey 30
Data-at-Rest Encryption: Performing a Deep Rekey 30
iSCSI: Identification and Management of iSCSI Objects in a vSAN Cluster 30
File Services: Introducing it into an Existing Environment 32
Section 9: Stretched Clusters 33
Section 10: 2-Node Clusters 33
Section 11: vCenter Server Maintenance and Event Handling 34
Upgrade Strategies for vCenter Server Powering One or More vSAN Clusters 34
Replacing vCenter Server on an Existing vSAN Cluster 34
Protecting vSphere Distributed Switches Powering vSAN 34
Section 12: Upgrade Operations 36
Upgrading and Patching vSAN Hosts 36
Tips for using vLCM in an Existing Environment 36
Upgrade Considerations for Different vSAN Topology Types 36
Multi-Cluster Upgrading Strategies 37
Upgrading Large vSAN Clusters 38
Upgrading Firmware and Drivers for NICs and Storage Controllers 38
Section 13: Customizing Utility Storage Space 39
Resize Custom Namespace Objects 39
Section 14: Monitoring vSAN Health 40
Remediating vSAN Health Alerts 40
Checking Object Status and Health during a Failure 40
Monitoring and Management of vSAN Object Components 41
Viewing vSAN Cluster Partitions in the Health Service UI 41
Monitoring and Management of Isolated vSAN Environments 42
Section 15: Monitoring vSAN Performance 44
Navigating Across the Different Levels of Performance Metrics 44
Troubleshooting vSAN Performance 46
Monitoring Resynchronization Activity 47
Network Monitoring of vSAN Powered Cluster 51

Summary 52
Additional Resources 52
About the Author 52

Introduction

VMware vSAN provides enterprise-class storage that is robust, flexible, powerful, and easy to use. While vSAN-powered clusters share many similarities to vSphere clusters in a three-tiered architecture, the unique abilities and architecture of vSAN means that some operational practices and recommendations may be different than that of traditional environments.

This document provides practical guidance in the day-to-day operations of vSAN-powered clusters. It augments the step-by-step instructions found in Broadcom TechDocs, knowledge articles, and other detailed guidance found in vSAN blogs and other vSAN content on the VMware Resource Center. **This operations guide is not intended to be "how-to" documentation**. It offers general guidance and recommendations applicable to a large majority of environments. Requirements unique to a specific environment may dictate slightly different operational practices, thus the reason for no single "best practice." New topics may be added periodically. Please check to ensure the latest copy is used.

The guidance provided in this document reflects recommendations in accordance with the latest version of vSAN at the time of this writing. New features in vSAN will often impact operational recommendations. When guidance differs based on recent changes introduced to vSAN, it will be noted. The guidance will not retain an ongoing history of practices for previous versions of vSAN.

Scope of Topics

The information provided in this document will assume the use of vSAN 9.1, and/or VMware Cloud Foundation (VCF) 9.1. VCF deployments may have additional requirements and support limitations that fall outside of the scope of this document. The release of vSAN in VCF 9.1 marks the 6th release of vSAN ESA, and as a result, this document will primarily focus on space efficiency features in vSAN ESA. For space efficiency features related to the vSAN Original Storage Architecture (OSA), see earlier versions of this document.

Section 1: Cluster

Pre-Flight Checks Prior to Introducing Clusters into Production

Introducing a new vSAN cluster into production is a simple process. Features such as vSAN health checks help provide guidance to ensure proper configuration, while VM migrations to a production cluster can be transparent to the consumers of those VMs. Supplementing the introduction of a new vSAN cluster into production with additional steps to ensure that, once the system is powering production workloads, you get the expected outcomes.

Preparation

Preparation helps reduce potential issues when VMs rely on the services provided by the cluster. It also helps establish a troubleshooting baseline. The following may be helpful in a cluster deployment workflow:

- **What is the intention of this cluster? And what data services reflect those intentions?** Do VMs need to be encrypted on this cluster? Do you plan on using vSAN Global Deduplication. Generally, cluster-wide data services are best enabled or disabled at the time the cluster is provisioned. This pertains primarily toward the OSA.
- **Has host count in cluster size been sufficiently considered?** Perhaps you planned on introducing a new 24-node cluster to the environment. You may want to evaluate whether a single cluster or multiple clusters are the correct fit, especially as they relate to network design, and operational objectives. While this can be changed later, evaluating at initial deployment is most efficient.

Recommendation: Always run a synthetic test (HCIBench) prior to introducing the system into production. This can verify that the cluster behaves as expected and can be used for future comparisons should an issue arise, such as network card firmware hampering performance. See step 1 in the "Troubleshooting vSAN Performance" document for more information.

Maintenance Work for L2/L3 Switching in Production

Redundant configuration

VMware vSAN recommends configuring redundant switches paired with vDS NIC teaming and failover so that the loss of one switch or path does not permanently cause a switch outage. For more information, see the post: "vSAN Networking - Teaming for Redundancy."

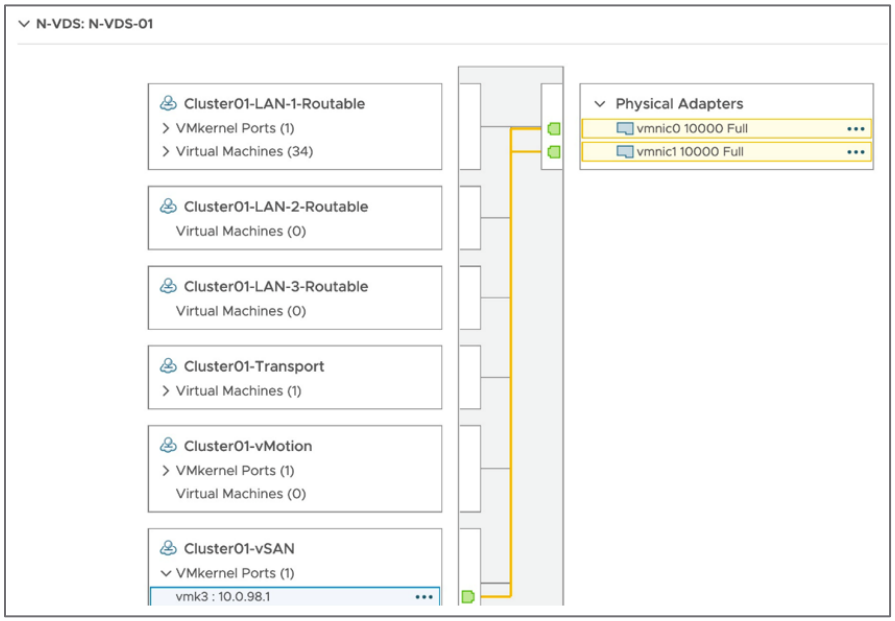


Figure. Virtual Switch and port group configuration

Health Findings

Prior to performing maintenance, review the vSAN networking health findings Health findings tied to connectivity, latency, or cluster partitions can help identify situations where one of the two paths is not configured correctly, or is experiencing a health issue.

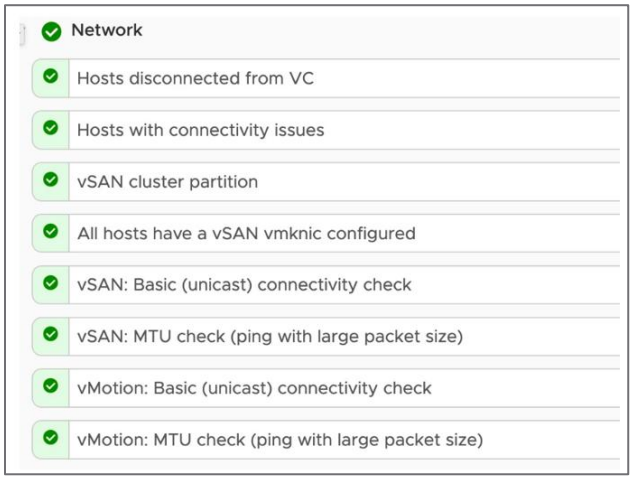


Figure. Network-related health checks in the vSAN health UI in vCenter

Understanding the nature of the maintenance can also help you understand what health alarms to expect. Basic switch patching can sometimes be performed non-disruptively. Switch upgrades that can be performed as an in-service software upgrade (ISSU) may not be noticeable, while physically replacing a switch may lead to several connectivity alarms. Discuss the options with your networking vendor.

Testing failure impacts

It is a good idea to simulate a path failure on a single host (disable a single port) before taking a full switch offline. If VMs on that host become unresponsive, or if HA is triggered, this may imply an issue with pathing that should be resolved prior to switch removal or reboot.

Controlled maintenance

If fault domains are used with multiple racks of hosts using different switches, consider limiting maintenance to a single fault domain and verify its health before continuing on. For stretched clusters, limit maintenance to one side at a time to reduce potential impacts.

Summary

In a vSAN environment, configuration of virtual switches, and the respective uplinks used follows practices commonly recommended in traditional three-tier architectures. With the added responsibility of serving as the storage fabric, ensuring that the proper configuration is in place will help the abilities of vSAN to perform as expected.

Configuring Fault Domains

Each host in a vSAN cluster is an implicit fault domain by default. vSAN distributes data across fault domains (hosts) to provide resilience against drive and host failure. This is sufficient to provide the right combination of resilience and flexibility for data placement in a cluster in most environments. There are use cases that call for fault domain definitions spanning across multiple hosts. Examples include protection against server rack failure, such as rack power supplies and top-of-rack networking switches.

vSAN includes an optional ability to configure explicit fault domains that include multiple hosts. vSAN distributes data across these fault domains to provide resilience against larger domain failure—an entire server rack, for example.

Preparation

vSAN requires a minimum of three fault domains. At least one additional fault domain is recommended to ease data resynchronization in the event of unplanned downtime, or planned downtime such as host maintenance and upgrades. The diagram below shows a vSAN cluster with 24 hosts. These hosts are evenly distributed across six server racks.

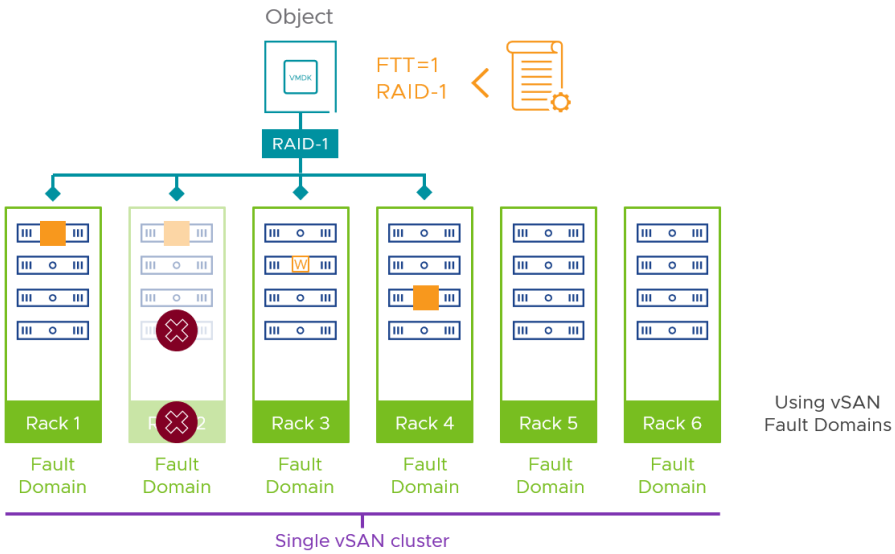


Figure. A conceptual illustration of a vSAN cluster using 24 hosts and 6 explicit fault domains

With the example above, you would configure six fault domains—one for each rack—to help maintain access to data in the event of an entire server rack failure. This process takes only a few minutes using the vSphere Client. The "Design and Operation Considerations When Using vSAN Fault Domains" post offers practical guidance for some of the most commonly asked questions when designing for vSAN fault domains. The post "Using Fault Domains in vSAN ESA" describes how the feature is different when using vSAN ESA.

Recommendation: Prior to deploying a vSAN cluster using fault domains, ensure that rack-level redundancy is a requirement of the organization. Fault domains can increase the considerations in design and management, thus determining the actual requirement up front can result in a design that reflects the actual needs of the organization.

vSAN is also capable of delivering multi-level replication or "nested fault domains." This is fully supported with vSAN stretched cluster architectures and 2-Node clusters. Nested fault domains provide an additional level of resilience at the expense of higher-capacity consumption.

Migrate to a Different vSAN Cluster Type

In some cases, an administrator may want to migrate VMs from one vSAN cluster to another. In most cases, performing a simple vMotion and storage vMotion will achieve the desired result. Some data services such as vSAN File Services and vSAN iSCSI target services may make this task more complex, but in most cases, the transition should be straightforward. See the "VMware vSAN Migration Guide" for more information.

Migrating from vSAN OSA to ESA

As customers refresh to the latest hardware and are anxious to migrate to vSAN ESA, questions arise about how best to migrate their environment to clusters running the ESA. For a detailed explanation on options available, see the post "Migrating to the Express Storage Architecture in vSAN 8"

Section 2: Network

Using NIOC for vSAN

vSphere Network I/O Control (NIOC) has mechanisms to reserve bandwidth for system traffic based on the capacity of the physical adapters on a host. It enables fine-grained resource control at the VM network adapter level, similar to the model used for allocating CPU and memory resources. NIOC is only supported on the VMware Distributed Switch (vDS) and is enabled per switch.

Planning

It is recommended to **not enable limits**. Limits artificially restrict vSAN traffic even when bandwidth is available. Reservations should also be avoided because reservations do not yield free bandwidth back for non-VMkernel port uses. On a 10Gbps interface uplink, a 9Gbps vSAN reservation would result in only 1Gbps of traffic available for VMs even when vSAN is not passing traffic. Limits also do not work well given that the ESA can utilize much more bandwidth if the guest VMs demand it, thus limits would not be very adaptable to these conditions. Reservations can also undermine desired outcomes, and are generally not recommended.

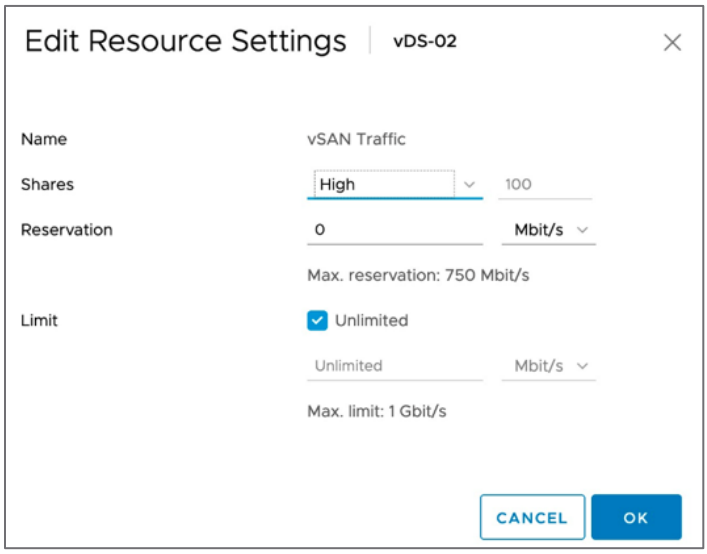
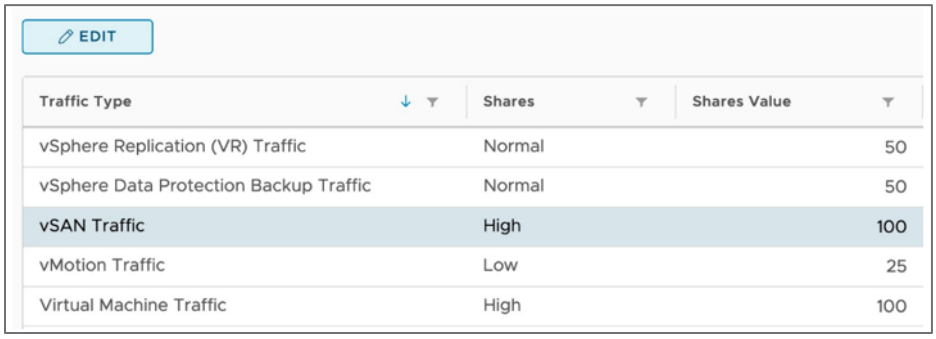


Figure. Setting shares in NIOC to balance network resources under contention

Under times of contention, such as a failed NIC in a vSAN host, **shares are the recommended way to prioritize traffic for VMware vSAN**. Raise the vSAN shares to "High."



Traffic Type	Shares	Shares Value
vSphere Replication (VR) Traffic	Normal	50
vSphere Data Protection Backup Traffic	Normal	50
vSAN Traffic	High	100
vMotion Traffic	Low	25
Virtual Machine Traffic	High	100

Figure. An example of a configuration of shares for a vSAN-powered cluster (OSA)

Other network quality of service (QoS) options

It is worth noting that NIOC only provides shaping services on the host's physical interfaces. It does not provide prioritization in switch-to-switch links and does not have awareness of contention caused by over saturated leaf/spine uplinks, or data center–to–data center links for stretched clustering. Discuss these options with your networking teams, and switch vendors for optimal configuration guidance.

Using Large MTU Sizes with vSAN

Jumbo frames are Ethernet frames larger than 1,500 bytes of payload. The most common jumbo configuration is a payload size of 9,000, although modern switches can often go up to 9,216 bytes.

Planning

Consult with your switch vendor and identify if jumbo frames are supported and what maximum transmission units (MTUs) are available. If multiple switch vendors are involved in the configuration, be aware they measure payload overhead in different ways in their configuration. Also identify if a larger MTU is needed to handle encapsulation such as VxLAN. Identify all configuration points that must be changed to support jumbo frames end to end. If Witness Traffic Separation is in use, be aware that an MTU of 1,500 may be required for the connection to the witness.

Implementation

Start the changes with the physical switch and distributed switch. To avoid dropped packets, make the change last to the VMkernel port adapters used for vSAN.

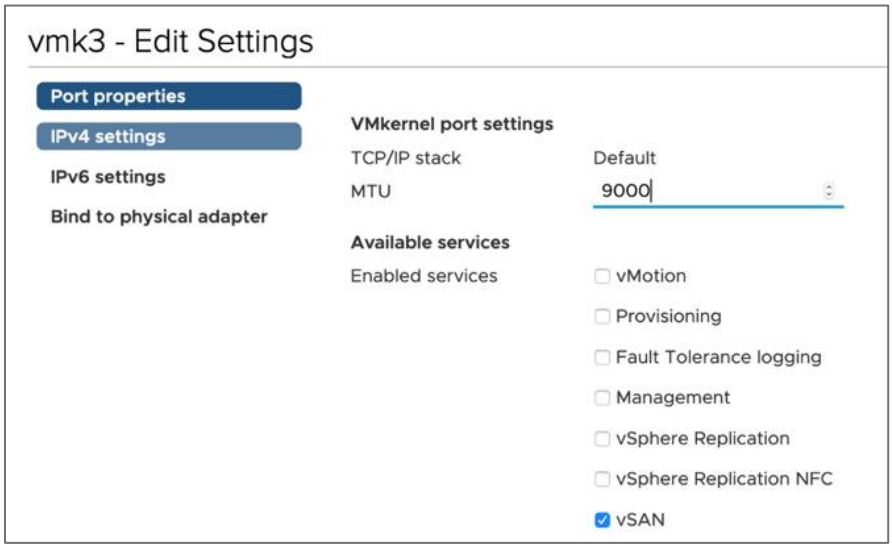


Figure. Changing the MTU size of virtual distributed switch (VDS)

Validation

The final step is to verify connectivity. To assist with this, vSAN: MTU check (ping with large packet size) will perform a ping test with large packet sizes from each host to all other hosts to verify connectivity end to end.

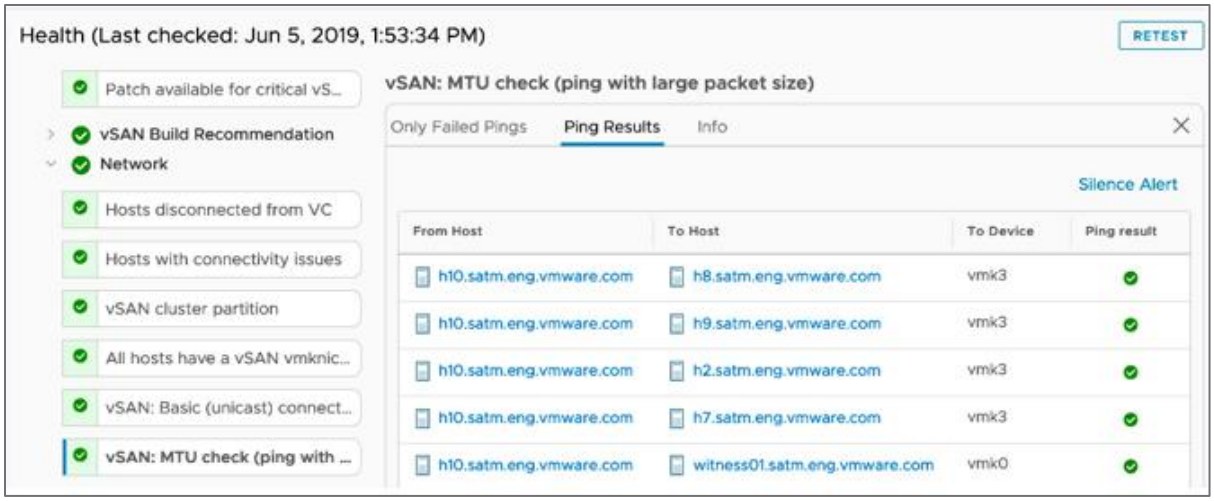


Figure. Verifying connectivity using the vSAN MTU check health check.

Create and Manage Broadcast Domains for Multiple vSAN Clusters

It is recommended, when possible, to **dedicate unique broadcast domains** (or collections of routed broadcast domains for Layer 3 designs) for vSAN. Benefits to unique broadcast domains include:

- **Fault isolation.** Spanning tree, configuration mistakes, entering duplicate IP address, and other failures can cause a broadcast domain to fail, or failures to propagate across a broadcast domain.
- **Security.** While vSAN hosts have automatic firewall rules created to reduce attack surface, data over the vSAN network is not encrypted unless by higher-level solutions (VM encryption, for example). To reduce the attack surface, restrict the broadcast domain to only contain VMkernel ports dedicated to the vSAN cluster. Dedicating isolated broadcast domains per cluster helps ensure security barriers between clusters.

Planning

There are several ways to isolate broadcast domains. The most basic is physically dedicated and isolated interfaces and switching. The most commonly chosen is to tag VLANs onto the port groups used by the vSAN VMkernel ports. Prior to this, configure the switches between the hosts to carry this VLAN for these ports. Other encapsulation methods for carrying VLANs between routed segments (ECMP fabrics, VxLAN) are supported.

Implementation

The first step is to configure the VLAN on the port group. This can also be set up when the VDS and port groups are created using the Cluster Quickstart.

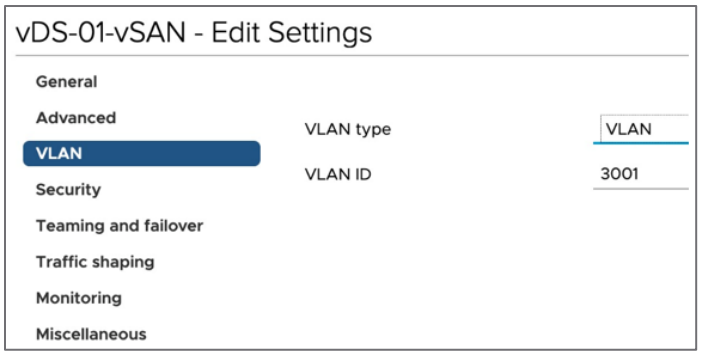


Figure: Configuring a port group to use a new VLAN

Validation

A number of built-in health checks can help identify if a configuration problem exists, preventing the hosts from connecting. To ensure proper functionality, all vSAN hosts must be able to communicate. If they cannot, a vSAN cluster splits into multiple partitions (i.e., subgroups of hosts that can communicate but not to other subgroups). When that happens, vSAN objects might become unavailable until the network misconfiguration is resolved. To help troubleshoot host isolation, the vSAN network health checks can detect these partitions and ping failures between hosts.

Recommendation: VLAN design and management does require some levels of discipline and structure. Discuss with your network team the importance of having discrete VLANs for your vSAN clusters up front, so that it lays the groundwork for future requests.

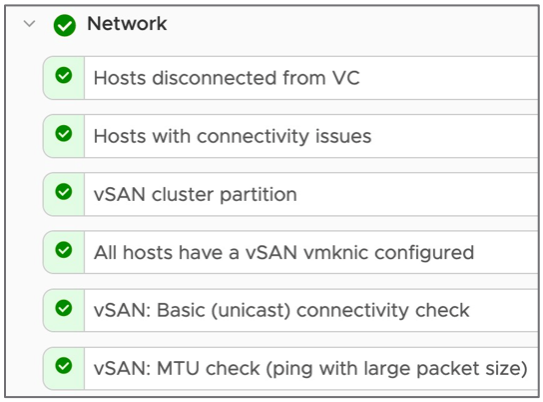


Figure. Validating that changes pass network health findings

Change IP Addresses of Hosts in a vSAN Cluster

vSAN requires networking between all hosts in the cluster for VMs to access storage and maintain the availability of storage. Operationally migrating IP addresses of storage networks need extensive care to prevent loss of connectivity to storage or loss of quorum to objects.

Planning

Identify if you do this as an online process or as a disruptive offline process (powering off all VMs). If disruptive, make sure to power off all VMs following the cluster shutdown guidance.

Implementation

If new VMkernel ports are used prior to removing old ones, a number of techniques can be used to validate networking and test hosts before removing the original VMkernel ports.

- Use vmkping to source pings between the new VMkernel ports.
- Put hosts into maintenance mode, or evacuate VMs before removing the original vSAN VMkernel port.
- Check the vSAN object health alarms to confirm that the cluster is at full health once the original VMkernel port has
- been removed.
- Once the host has left maintenance mode, vSphere vMotion, a test VM to the host and confirm that no health alarms are alerting before continuing to the next host.

Validation

Before restoring the host to service, confirm that networking and object health is returning normal health.

Migrate vSAN Traffic to Another VMkernel Port

There are cases where the vSAN network needs to be migrated from to a different segment. For example, the implementation of a new network infrastructure or the migration of vSAN standard cluster (non-routed network) to a vSAN stretched cluster (routed network). Recommendations and guidance on this procedure is given below.

Prerequisites

Check Skyline Health for vSAN to verify there are no issues. This is recommended before performing any planned maintenance operations on a vSAN cluster. Any issues discovered should be resolved before proceeding with the planned maintenance.

Set up the new network configuration on your vSAN hosts. This procedure will vary based on your environment. Ensure that the new vSAN network subnet does not overlap with the existing one. vSphere will not allow the vSAN service to run simultaneously on two VMkernel ports on the same subnet. Attempting to do this using esxcli will produce an error like the one shown below.

```
esxcli vsan network ip add -i vmk2

Failed to add vmk2 to CMMDS: Unable to complete Sysinfo operation. Please see the VMkernel log file for more details.

Vob Stack: [vob.vsan.net.update.failed.badparam]: Failed to ADD vmknic vmk2 with vSAN because a parameter is incorrect.
```

Note that you might see warnings in vSAN Health as you add new VMkernel adapters with the vSAN service--specifically, the "vSAN: Basic (unicast) connectivity check" and "vSAN: MTU check (ping with large packet size)" health checks, as shown below. This is expected if the vSAN service on one host is not able to communicate with other hosts in the vSAN cluster. These warnings should be resolved after the new VMkernel adapters for vSAN have been added and configured correctly on all hosts in the cluster. Use the "Retest" button in vSAN Skyline Health to refresh the health checks status.

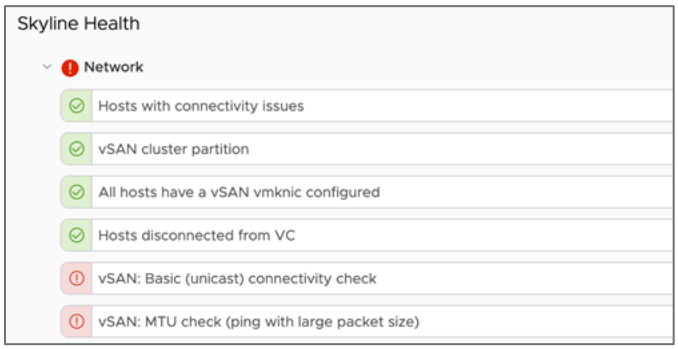


Figure. vSAN Skyline Health warnings

Use vmkping to verify the VMkernel adapter for the new vSAN network can ping the same VMkernel adapters on other hosts. This KB article provides guidance on using vmkping to test connectivity: https://knowledge.broadcom.com/external/article?legacyId=1003728

1. Shut down all running virtual machines that are using the vSAN datastore. This will minimize traffic between vSAN nodes and ensure all changes are committed to the virtual disks before the migration occurs.
2. After configuring the new vSAN network on every host in the vSAN cluster, verify the vSAN service is running on both VMkernel adapters. This can be seen in the UI by checking the Port Properties for both VMkernel adapters in the UI or by running esxcli vsan network list. You should see an output similar to the text below.

```
[root@host01:~] esxcli vsan network list Interface

VmkNic Name: vmk1

...

Traffic Type: vsan
```

```
Interface

VmkNic Name: vmk2

...

Traffic Type: vsan
```

1. Click the "Retest" button in vSAN Skyline Health to verify there are no warnings while the vSAN service is enabled on both VMkernel adapters on every host. If there are warnings, it is most likely because one of more hosts do not have the vSAN service enabled on both VMkernel adapters. Troubleshoot the issue and use the "Retest" option in vSAN Skyline Health until all issues are resolved.
2. Disable the vSAN service on the old VMkernel adapters.
3. Click the "Retest" button in vSAN Skyline Health to verify there are no warnings.
4. Power on the virtual machines.

Recommendation: While it is possible to perform this migration when VMs on the vSAN datastore are powered on, it is NOT recommended and should only be considered in scenarios where shutting down the workloads running on vSAN is not possible.

Introducing RDMA into a vSAN Environment

Running vSAN over RDMA introduces all new levels of capabilities in efficiency and performance. **For clusters running the vSAN ESA, RDMA may provide even more beneficial**, as often times the bottleneck in an ESA environment is the network, and not the storage stack in a server.

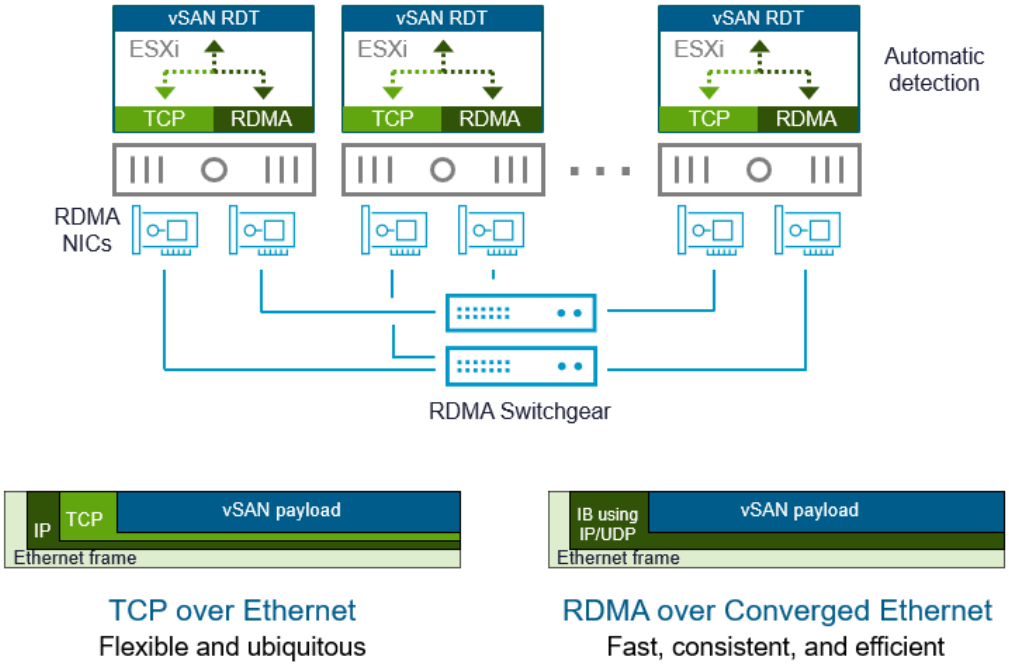


Figure. vSAN over RDMA

Recommendation: Ensure that a host added to a vSAN cluster running RDMA is in fact, fully compatible with RDMA. Adding a single host that is not compatible with RDMA will make the cluster fail back to using TCP over ethernet.

Introducing RDMA into an environment requires the use of certified hardware (RDMA NIC adapters and switches). See the BCG for vSAN RDMA Network Adapters for more information. For more information, see the post: "vSAN Networking – Is

RDMA Right for You?" vSAN clusters using RDMA may be subject to additional limitations of supported features or functionality, including, but not limited to:

- vSAN cluster sizes are limited to 32 hosts
- vSAN cluster must not be running the vSAN iSCSI services
- vSAN cluster must not be running in a stretched cluster configuration
- vSAN cluster must be using RDMA over Layer 2. RDMA over Layer 3 is not supported.
- vSAN cluster running vSAN over RDMA is not supported with disaggregated environments including vSAN storage clusters, or vSAN HCI with datastore sharing.
- vSAN cluster must not be using a teaming policy based on IP Hash or any active/active connection where sessions are balanced across two or more uplinks.

Section 3: Storage Devices

Both vSAN OSA and vSAN ESA have the ability to easily add and remove storage devices that contribute to the vSAN cluster. The management of these devices varies slightly between architectures, as noted below.

Adding and Removing Storage Devices in vSAN ESA

The claiming of devices in vSAN ESA is relatively flexible and easy. ESA removes the construct of a disk group found in OSA. Instead, each host shows claimed devices in a "storage pool." There is only one storage pool per host, and it simply reflects the collection of storage devices claimed for use by vSAN. Claiming can be performed manually, automatically through "managed disk claim" or prescriptively through "prescriptive disk claim" capability introduced in vSAN 8 U2. For more information, see the post: "vSAN Prescriptive Disk Claim for the ESA in vSAN 8 U2."

With each storage device in vSAN ESA being its own boundary of failure and management, accommodation for these two conditions is much more efficient than in vSAN OSA. For more information, see the post: "The Impact of a Storage Device Failure in vSAN ESA versus OSA"

Recommendation: When removing any storage device permanently, use the "full data migration option." This ensures that objects remain compliant with the respective storage policies. Use LED indicators to identify the appropriate devices that needs to be removed from the server.

Secure Erase of Data on a Decommissioned vSAN Storage Device

Just as with many storage systems, discrete storage devices decommissioned from a storage system typically need an additional step to meet the National Institute of Standards and Technology (NIST) to ensure that all data previously stored on a device can no longer be accessed. This involves a step often referred to as "secure erase" or "secure wipe." The goal of a secure wipe is to prevent data spillage, which could occur if a system or device was repurposed to a less sensitive environment. It also plays a critical role in a declassification procedure, which may involve the formal demotion of the hardware to a less secure environment. The method discussed here achieves a properly and securely erased device for both of those purposes.

vSAN provides a secure wipe process. It can be achieved through API or PowerCLI, with the latter being a much more convenient option for administrators. It should be the final step in the decommissioning process if the requirements dictate this level of security. To ensure the protection of data occurring as a result of an inadvertent command, the wipe option will only be supported if the "Evacuate" all data" was chosen at the time of removing the disk from the disk group. More information can be found in the "vSAN Encryption Services" document.

Recommendation: Be patient. The secure wipe procedure may take some time. Claiming the device in vSAN must wait for the secure wipe process to complete.

PowerCLI command Syntax

The PowerCLI commands for wiping a disk will include:

```
Wipe-Disk – Given a list of disks, issues a wipe disk.  Syntax:  Wipe-Disk –Disk <Disk[]> -
RunAsync

Query-Wipe-Status – Given a list of disks, returns a lists of wipe disk status.  Syntax:
Query-Wipe-Status <Disk[]>

Abort-Wipe-Disk – Given a list of disks, cancel the sanitization of them and return the
status.  Syntax:  Abort-Wipe-Disk <Disk[]>
```

The following image shows an example of these secure wipe commands.

```
1. login server

PS C:\windows\system32> Connect-VIServer '              ' -Username 'administrator@vsphere.local' -Pass
                

2. Get host

PS C:\windows\system32> $h = Get-VMHost

3. Query wipe status

PS C:\windows\system32> $s = Get-VsanWipeDiskStatus -VMHost $h[0] -CanonicalName ("mpx.vmhba0:C0:T3:L0
```

Figure. Example of the secure wipe commands

The disk wipe activity log will capture:

- Date/time wipe initiated
- Date/time wipe completed
- Status of job.
- Relevant information as to which host and cluster the activity occurred
- Status of success or failure

Section 4: vSAN Datastore

Capacity Management Guidance

Capacity Management becomes substantially different in vSAN for VCF 9.1. vSAN Auto-RAID paired with the new "Effective Capacity" capability in vSAN makes vSAN capacity management behave very similar to traditional storage. This removes many of the capacity concepts that were unique to vSAN, such as rendering storage capacity in terms of raw capacity.

Overview

The new "Effective Capacity" view in vSAN for VCF 9.1 provides clear and easy-to-interpret capacity information. Unlike previous versions, "Total usable capacity" is just as the name implies: actual capacity for your workloads. All overheads and free space requirements by vSAN are already accounted for. This gives a clean and clear understanding of what the cluster can provide. For more information, see the post: "Simplifying Storage with the New Effective Capacity View in VMware vSAN for VCF 9.1."

The "Space Efficiency" will provide clear indicators of how much capacity has been reclaimed through space efficiency techniques. For more information, see the "vSAN Space Efficiency Technologies" document.

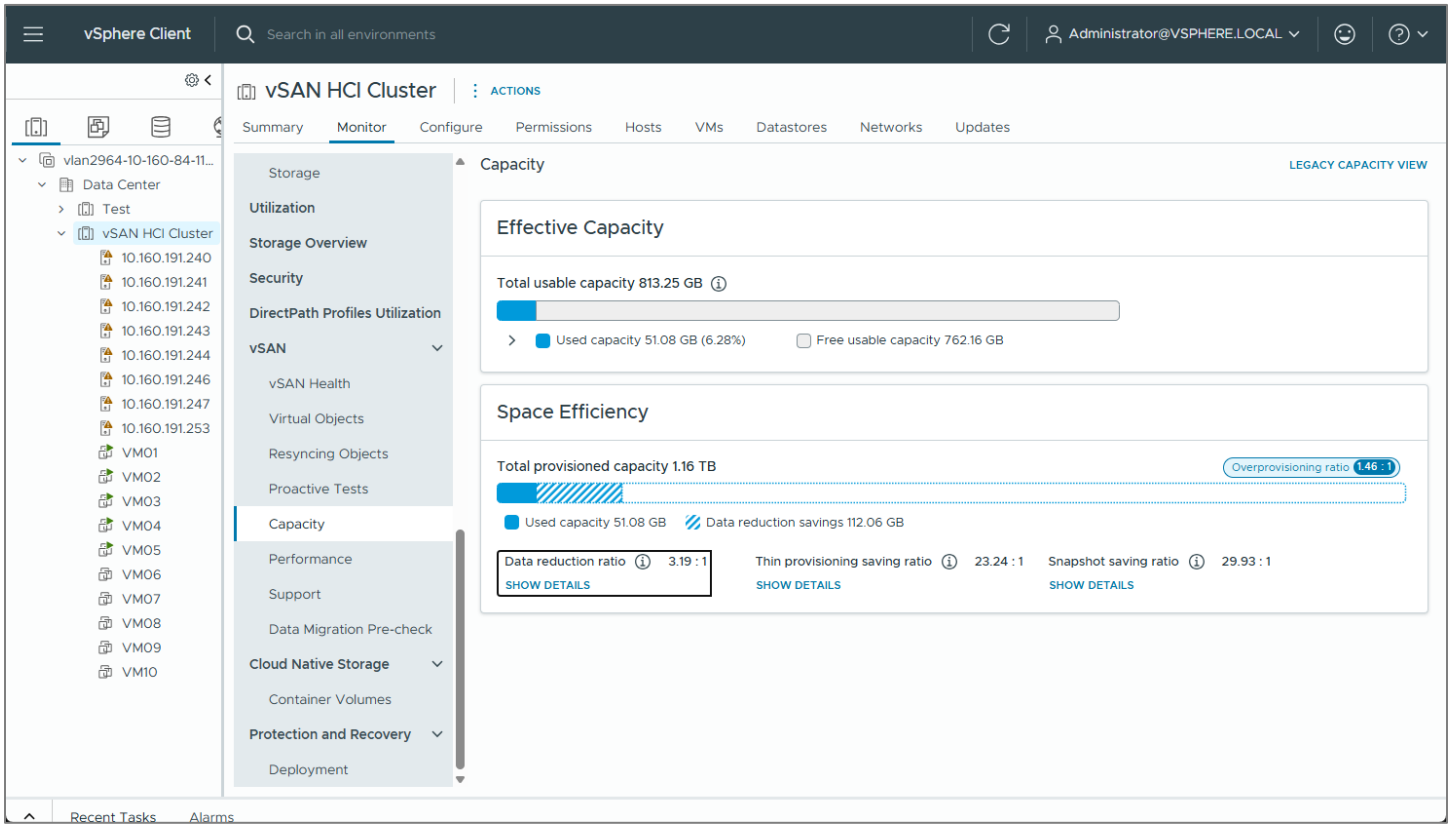


Figure. New "Effective Capacity" view in vSAN for VCF 9.1

Automatic Rebalancing in a vSAN Cluster

vSAN provides an ability to automatically rebalance data across a vSAN cluster to optimize the use of resources across the cluster. For more extensive information on this capability, see the post: "Should Automatic Rebalancing be Enabled in a vSAN Cluster?"

In most cases, it is recommended to enabled the cluster-based toggle. This will help distribute the data for an optimal level of resource utilization across the cluster.

Managing Orphaned Objects in a Datastore

vSAN is an object-based datastore. The objects typically represent entities such as virtual machines, performance history database, iSCSI objects, Persistent volumes, and vSphere Replication Data. An object may inadvertently lose its association with a valid entity and become orphaned. Objects in this state are termed as orphaned or unassociated objects. While orphaned objects do not critically impact the environment, they contribute to unaccounted capacity and skew reporting.

Common causes for orphaned objects include but not limited to:

- Objects that were created manually instead of using vCenter or an ESXi host
- Improper deletion of a virtual machine such as deleting files through a command-line interface(CLI)
- Using vSAN datastore to store non-standard entities such as ISO images
- Manage files directly through vSAN datastore browser
- Residual objects caused by incorrect snapshot consolidation or removal by 3rd party utilities

Identification and Validation

Unassociated objects can be ascertained through command-line utilities such as Ruby vSphere Console(RVC) and Go-based vSphere CLI(GOVC). RVC is embedded as part of the vCenter Server Appliance(vCSA). GOVC is a single static binary that is available in GitHub and can be installed across different OS platforms.

Here are the steps to identify the specific objects,

GOVC

```
Command Syntax: govc datastore.vsan.dom.ls -ds <datastorename> -l -o

Sample Command: govc datastore.vsan.dom.ls -ds vsanDatastore -l -o

<Command does not return an output if no unassociated objects are found>
```

Additional Reference for this task can be found at KB 70726

Recommendation: Contact Technical Support to help validate and delete unassociated objects. Incorrect detection and deletion of unassociated objects may lead to loss of data.

Section 5: Storage Policies

Historically, storage policies in vSAN provided the ability to prescribe outcomes that addressed both desired resilience levels, and other VM specific settings. In vSAN for VCF 9.1, Auto-RAID determines and manages the optimal resilience settings for all data residing on the datastore.

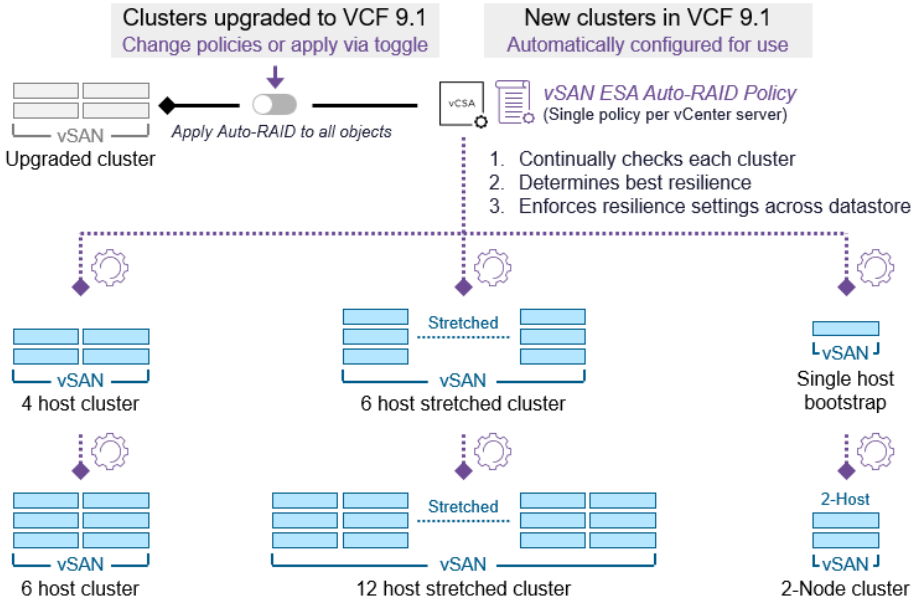


Figure. vSAN Auto-RAID in vSAN for VCF 9.1

Since vSAN Auto-RAID will only create one storage policy per vCenter Server, yet manage multiple different vSAN cluster types, this will have a profound improvement on management at scale. For more information, see the post: "Auto-RAID in VMware vSAN for VCF 9.1 - Comprehensive System-Managed Data Resilience."

Using Workload Limiting Polies (IOPS Limits)

IOPS limits is one of the ways that administrators can reduce the potential impact of one VM consuming an abnormally large amount of resources. While it sounds like an ideal solution for those situations, it does come with tradeoffs. To learn more about IOPS limits in vSAN, and when and when not to use them, see the post: "Performance Metrics when using IOPS Limits with vSAN – What you Need to Know."

Using Object Space Reservation (OSR)

The object space reservation (OSR) storage policy rule serves as a capacity management capability for VMs where you want to ensure they will have all of the available capacity that has been assigned. Setting the OSR on a VM does not impact the performance of a VM one way or the other.

Section 6: Host and EMM Operations

What EMM Option to Choose for Host Maintenance

All hosts in a vSAN cluster contribute to a single shared vSAN datastore for that specific cluster. If a host goes offline due to any planned or unplanned process, the overall storage capacity for the cluster is reduced. From the perspective of storage capacity, placing the host in maintenance mode is equivalent to its being offline. During the decommissioning period, the storage devices of the host in maintenance mode won't be part of the vSAN cluster capacity.

Maintenance mode is mainly used when performing upgrades, patching, and hardware maintenance such as replacing a drive, adding or replacing memory, or updating firmware. For network maintenance that has a significant level of disruption in connectivity to the vSAN cluster and other parts of the infrastructure, a cluster shutdown procedure may be most

appropriate. Rebooting a host is another reason to use maintenance mode. For even a simple host restart, it is recommended to place the host in maintenance mode.

Placing a given host in maintenance mode impacts the overall storage capacity of the vSAN cluster. Here are some pre-requisites that should be considered before placing a host in decommission mode:

- It is always better to decommission one host at a time.
- Maintain sufficient free space for operations such as VM snapshots, component rebuilds, and maintenance mode.
- Verify the vSAN health condition of each host.
- View information about the number of objects that are currently being synchronized in the cluster, the estimated time to finish the resynchronization, the time remaining for the storage objects to fully comply with the assigned storage policy, and so on.

A pre-check simulation is performed on the data that resides on the host so that vSAN can communicate to the user the type of impact the EMM will have, all without moving any data. If the pre-check results show that a host can be seamlessly placed in maintenance mode, decide on the type of data migration. Take into account the storage policies that have been applied within the cluster. Some migration options might result in a reduced level of availability for some objects. Let's look at the three potential options for data migration:

Full data migration—Evacuate all components to other hosts in the cluster.
This option maintains compliance with the FTT number but requires more time as all data is migrated from the host going into maintenance mode. It usually takes longer for a host to enter maintenance mode with Full data migration versus Ensure accessibility. Though this option assures the absolute availability of the objects within the cluster, it causes a heavy load of data transfer. This might cause additional latency if the environment is already busy. When it is recommended to use Full data migration:

- If maintenance is going to take longer than the rebuild timer value.
- If the host is going to be permanently decommissioned.
- If you want to maintain the FTT method during the maintenance.

Ensure accessibility (default option)—Instructs vSAN to migrate just enough data to ensure every object is accessible after the host goes into maintenance mode. vSAN searches only for data with RAID-0 and move/regenerate them on a host different than the one entering in maintenance mode. All the other objects with RAID-1 and higher, should already have at least one copy residing on different host within the cluster. Once the host comes back to operational, the data components left on the host in maintenance mode update with changes that have been applied on the components from the hosts that have been available. Keep in mind the level of availability might be reduced for objects that have components on the host in maintenance mode.

This maintenance mode is intended to be used for software upgrades or node reboots. Ensure accessibility gives the opportunity to avoid needless Full data migration, since the host will be back to operational in a short time frame. It is the most versatile of all EMM options

No data migration—No data is migrated when this option is selected.
A host will typically enter maintenance mode quickly with this option, but there is a risk if any of the objects have a storage policy assigned with FTT=0. When it is recommended to use No data migration:

- This option should be applied while some network changes are to be applied. In that specific case, all the nodes from the cluster should be placed in maintenance mode, selecting the "No data migration" option.
- This option is best for short amounts of planned downtime where all objects are assigned a policy with FTT=1 or higher, or where downtime of objects with FTT=0 is acceptable.

Our recommendation is to always build a cluster with the minimum number of hosts n + 1. This configuration allows vSAN to self-heal in the event of a host failure or a host entering in maintenance mode.

There is no need to keep a host in maintenance mode in perpetuity to achieve an N+1 or hot spare objective. vSAN's distributed architecture already achieves this. To ensure a cluster is properly sized for N+1 or greater failures, use the vSAN Sizer, and follow the recommendations in the vSAN Design Guide.

Restarting a Host

For typical host restarts with ESXi, most administrators get a feel for roughly how long a host takes to restart, and simply wait for the host to reappear as "connected" in vCenter. This may be one of the many reasons why out-of-band host management isn't configured, available, or a part of operational practices. However, hosts in a vSAN cluster can take longer to reboot than non-vSAN hosts because they have additional actions to perform during the host reboot process. Many of these additional tasks simply ensure the safety and integrity of data. Incorporating out-of-band console visibility into your operational practices can play an important role for administering a vSAN environment.

Recommendation: Use out-of-band management to view vSphere DCUI during host restarts.

Cluster Shutdown and Power-Up

Occasionally a graceful shutdown of a vSAN cluster may need to occur. Whether it be for server relocation, or for a sustained power outage where backup power cannot sustain the cluster indefinitely. Since vSAN is a distributed storage system, care must be taken to ensure that the cluster is shut down properly. The guidance offered here will be dependent on the version of vSAN used.

The recommendations below assume that guest VMs in the cluster are shut down gracefully before beginning this process. The order that guest VMs are powered down is dependent on the applications and requirements of a given customer environment and is ultimately the responsibility of the administrator.

vSAN provides a guided workflow built right into vCenter Server makes a cluster power down and power up process easy, predictable, and repeatable. This feature is a management task of the cluster. It is available in the vCenter Server UI when highlighting a given vSAN cluster, and selecting **vSAN > Shutdown Cluster**.

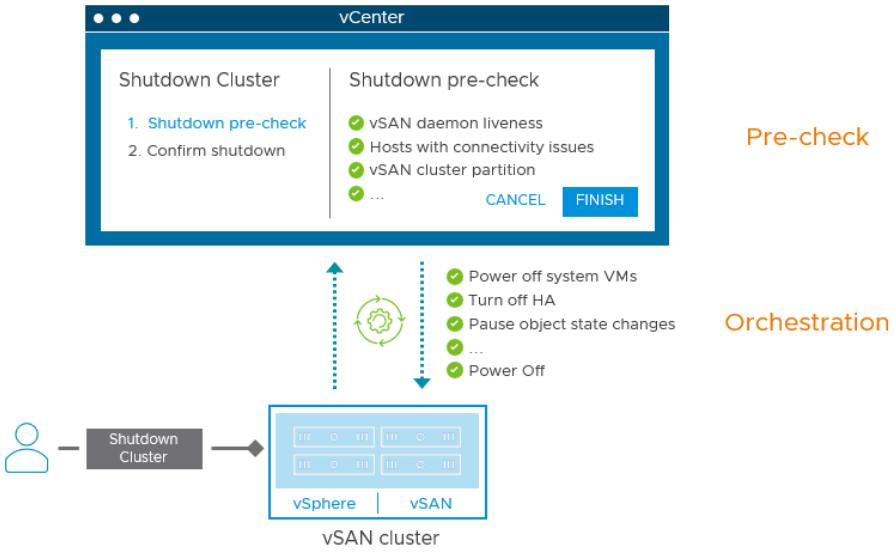


Figure. The logic of the new cluster shutdown workflow.

Note that **in vSAN 8, the Shutdown Cluster orchestration was enhanced to provide improved robustness** under a variety of conditions, and in vSAN 8 U1, the Shutdown Cluster workflow can be executed using PowerCLI.

The workflow accommodates vSAN clusters that are powering the vCenter Server. The process elects an orchestration host that assists in this cluster shutdown and the startup process once the vCenter Server VM is powered off. The selection of

the orchestration host is arbitrary, but if the cluster powers a vCenter Server, it will typically elect the host that the vCenter Server VM is associated with.

Powering down the cluster will be orchestrated by this new built-in workflow. A high-level overview of the steps includes:

- Pre-validation health checks (e.g. is HA disabled, and all VMs powered off, etc.). The workflow will be halted if it does not pass.
- Hosts set a new flag so vSAN's object manager will pause all change control processes.
- If vCenter server resides in the same cluster, vCenter will shut down and management of subsequent tasks will be delegated to the orchestration host.
- All hosts enter maintenance mode using "no action" option to prevent unnecessary data migration.
- Hosts are shut down.

Powering up the cluster will also be orchestrated by the new built-in workflow. A high-level overview of the steps includes:

- Administrator powers on ALL hosts in the cluster (using OOB management like IPMI, iDRAC, ILO, etc.)
- The orchestration host will set the flag in vSAN's object manager back to its original state to accept CCPs.
- If vCenter Server is within the cluster that was shut down, it will be automatically powered on.
- vCenter Server will perform a health check to verify the power state and alert of any issues.
- The administrator can power on VM's.

The workflow also supports stretched cluster and 2-node topologies, but will not power down the witness host appliance, as this is an entity that resides outside of the cluster, and may also be responsible for duties with other clusters. The feature will also be available when the ESXi host lockdown mode is enabled on the hosts in the cluster.

Some system may be automatically managed during the shutdown process, while others may not. Examples of other system-related VMs that will need to be managed manually include:

- File Services. No prechecks or automation workflows are included at this time.
- VKS pod VMs. These must be manually shut down.
- NSX management VMs. These must be manually shut down.

Recommendation: Regardless of the version of vSAN used, become familiar with the shutdown cluster process by testing it in a lab environment. This will help ensure that your operational procedures are well understood for these scenarios.

A vSAN cluster **assumes that network connectivity is fully available while the cluster performs the shutdown.** If the network is powered down first, or becomes unavailable, then the cluster will remain partitioned. This is a protective mechanism to preserve the integrity of the data (preventing split brain, partial commits, etc.). For more information on failure handling scenarios, see: "vSAN Availability Technologies."

Powering up a vSAN cluster

A commonly overlooked step in the powering up of a vSAN cluster is to **ensure all hosts in the cluster are powered on and fully initialized prior to powering on guest VMs**. This is different than a vSphere cluster using a traditional three-tier architecture where a host that was powered on and initialized would not necessarily need to wait for other hosts to be powered on before VMs could be started. Since vSAN provides the storage resources in a distributed manner, A VM hosted on one host may have its data reside on other hosts, thus the need to ensure that all hosts are ready prior to powering on guest VMs.

Section 7: Guest VM Operations

Configuring TRIM/UNMAP in vSAN

vSAN supports thin provisioning, which lets you use just as much storage capacity as currently needed in the beginning and then add the required amount of storage space at a later time. Using the vSAN thin provisioning feature, you can create virtual disks in a thin format. For a thin virtual disk, ESXi commits only as much storage space as the disk needs for its initial operations. To use vSAN thin provisioning, set the SPBM policy for Object Space Reservation (OSR) to its default of 0.

One challenge to thin provisioning is that VMDKs, once grown, will not shrink when files within the guest OS are deleted. This problem is amplified by the fact that many file systems always direct new writes into free space. A steady set of writes to the same block of a single small file eventually use significantly more space at the VMDK level. Previous solutions to this required manual intervention and migration with Storage vMotion to external storage, or powering off a VM. To solve this problem, automated TRIM/UNMAP space reclamation was created for vSAN 6.7U1.

Additional information can be found on the "UNMAP/TRIM space reclamation on vSAN" section of the vSAN Space Efficiency Technologies document. The post: "The Importance of Space Reclamation for Data Usage Reporting in vSAN" will also be of use in better understanding TRIM/UNMAP functionality.

Planning

If implementing this change on a cluster with existing VMs, identify the steps to clean previously non-reclaimed space. In Linux, this can include scheduling file system (FS).Trim to run by timer, or in Windows, running the disk optimization tools or the Optimize-Volume PowerShell command. Identify any operating systems in use that may not natively support TRIM/UNMAP.

UNMAP commands do not process through the mirror driver. This means that snapshot consolidation will not commit reclamation to the base disk, and commands will not process when a VM is being migrated with VMware vSphere Storage vMotion. To compensate for this, run asynchronous reclamation after the snapshot or migration to reclaim these unused blocks. This may commonly be seen if using VADP-based backup tools that open a snapshot and coordinate log truncation prior to closing the snapshot. One method to clean up before a snapshot is to use the pre-freeze script.

Identify any VMs that you wish to not reclaim space with. For these VMs you can use a VMX flag disk.scsiUnmapAllowed set to False.

Implementation

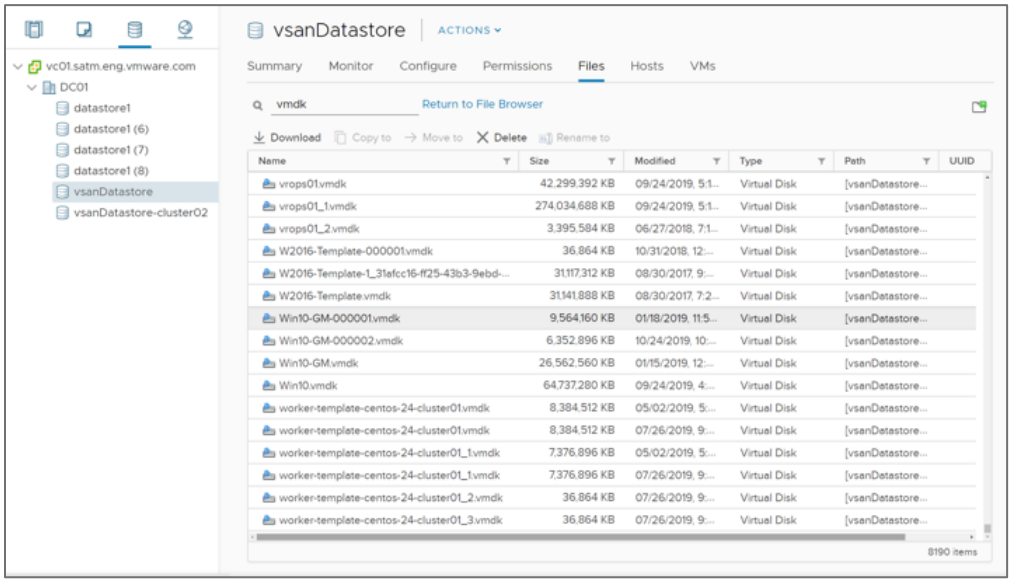


Figure. Viewing the size of a virtual disk within the vSAN Datastore view of Center

Validation

After making the change, reboot a VM and manually trigger space reclaim. Monitor the backend UNMAP throughput and total free capacity in the cluster increasing.

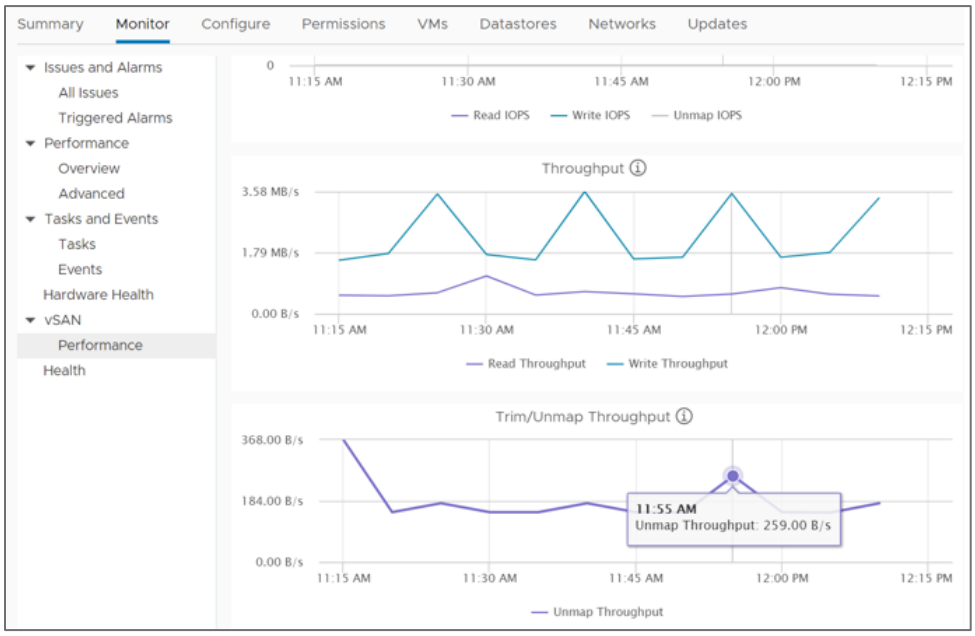


Figure. Viewing TRIM/UNMAP throughput on the host-level vSAN performance metrics

Tuning of Workloads after Migration to vSAN

In production environments, it is not uncommon to tune VMs to improve the efficiency or performance of the guest OS or applications running in the VM. Tuning generally comes in two forms:

- **VM tuning**—Achieved by adjusting the VM's virtual hardware settings.
- **OS/application tuning**—Achieved by adjusting OS or application-specific settings inside the guest VM.

The following provides details on the tuning options available, and general recommendations in how and when to make adjustments.

VM Tuning

VM tuning is common in traditional three-tier architectures as well as vSAN. Ensuring sufficient but properly sized virtual resources of compute, memory, and storage has always been important. Additionally, vSAN provides the ability to tune storage performance and availability settings per VM or VMDK using storage policies. VM tuning that is non-vSAN-specific includes, but is not limited to:

- Virtual CPU
- Amount of virtual memory
- Virtual disks
- Type and number of virtual SCSI controllers
- Type and number of virtual NICs

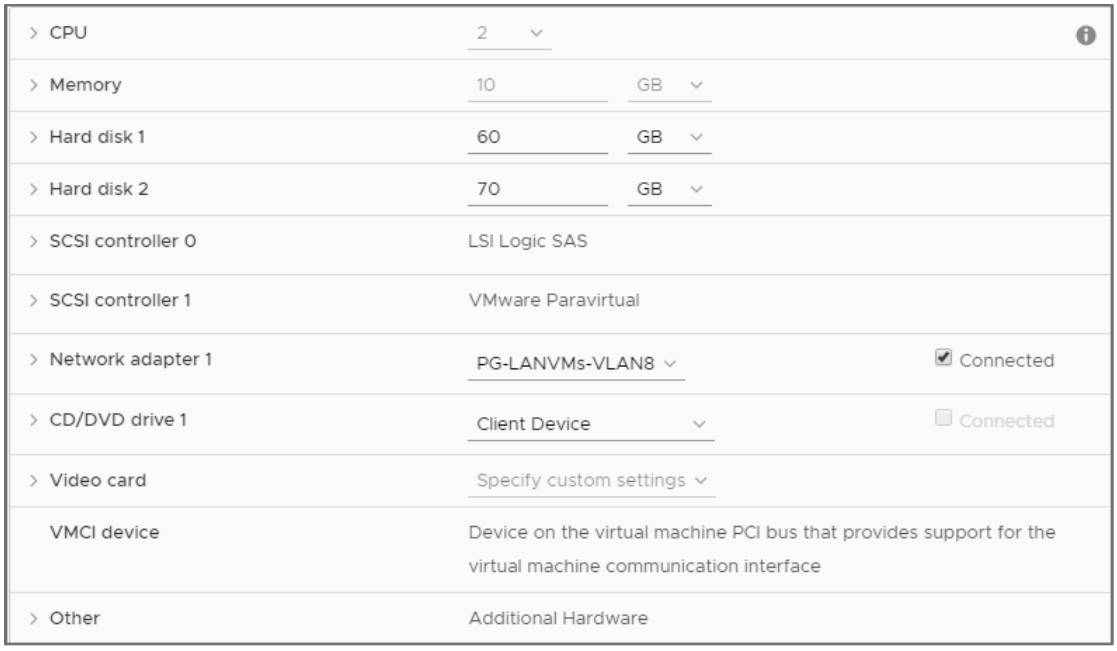


Figure. Virtual hardware settings of a VM

Determining the optimal allocation of resources involves monitoring the VM's performance metrics in vCenter Server, or augmenting this practice with other tools such as VMware Aria Operations to determine if there are any identified optimizations for VMs.

Recommendation: For VMs using more than one VMDK, use multiple virtual SCSI adapters. This provides improved parallelism and can achieve better performance. It also allows one to easily use the much more efficient and better performing Paravirtual SCSI controllers on these additional VMDKs assigned to a VM. See the "Troubleshooting vSAN Performance" document for more information.

OS/application tuning

OS/application tuning is generally performed to help the OS or application optimize its behavior to the existing environment and applications. Often you may find this tuning in deployment guides by an application manufacturer, or in a reference architecture. Note: Sometimes, if the recommendations come from a manufacturer, they may not take a virtualized OS or application into account and may have wildly optimistic recommendations.

For high performing applications such as SQL server, **ensure the guest VM volumes use a proper disk/partition alignment.** Some applications such as SQL demand a highly efficient storage system to ensure that serialized, transactional updates can be delivered in a fast and efficient manner. Sometimes, due to how a guest OS volume or partition is created, I/O requests will be unaligned, causing unnecessary Read, Modify, Write (RMW) events, increasing I/O activity unnecessarily, and impacting performance . See the post "Enhancing Microsoft SQL Server Performance on vSAN (and VMC on AWS) with SQL Server Trace Flag 1800" for information on how to determine if there is I/O unalignment of your SQL Server VM, and how to correct it. In some circumstances, it can have a dramatic impact on performance. While the link above showcases the issue and benefit on Microsoft SQL Server running on Windows Server, it can can occur with other applications.

Recommendation: Avoid over-ambitious OS/application tuning unless explicitly defined by a software manufacturer, or as outlined in a specific reference architecture. Making OS and application adjustments in a non-prescriptive way may add unnecessary complexity and result in undesirable results. If there are optimizations in the OS and application, make the adjustments one at a time and with care. Once the optimizations are made, document their settings for future reference.

VM tuning, as well as OS/application tuning can sometimes stem from identified bottlenecks. The "Troubleshooting vSAN Performance" document on core.vmware.com provides details on how to isolate the largest contributors to an identified performance issue, and the recommended approach for remediation.

Section 8: Data Services

Compression (ESA)

For vSAN in VCF 9.1, data compression enhancements were made to deliver improved compression rates, with minimal CPU usage. As a result, data compression is a cluster-wide feature that is always-on, and cannot be disabled in the UI. For more information, see the post: "More Capacity with VMware vSAN Compression and Global Deduplication in VCF 9.1."

Global Deduplication (ESA)

vSAN in VCF 9.1 provides the general availability of vSAN Global Deduplication. This is a cluster-based service enabled as a toggle within the configuration of a cluster. When turning deduplication off, it only pauses deduplication. It does not revert the data to an unduplicated state. Once you have enabled deduplication, leave it on.

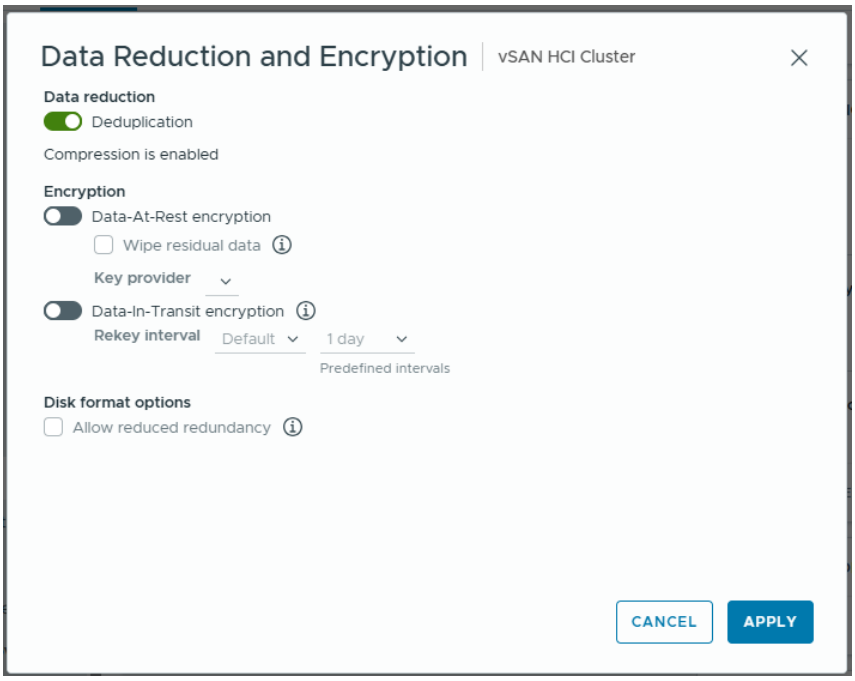


Figure. Enabling Global Deduplication in vSAN for VCF 9.1.

Data-at-Rest Encryption: Enabling and Disabling

Enabling or disabling data at rest encryption on a vSAN cluster is relatively easy. When vSAN Data-at-Rest Encryption is enabled or disabled, it performs a rolling reformat that copies data to available capacity on another device or host, and then encrypts or decrypts each device to prepare it for the encryption setting preferred (enabled or disabled)..

Enabling or disabling encryption introduces a rolling reformat of the storage devices that are claimed by vSAN. This rolling reformat does not place the host into maintenance mode and simply reformats one device at a time. **While the hosts are not placed into maintenance mode during the enabling or disabling of vSAN Data-at-Rest encryption, this rolling reformat may generate a substantial amount of data movement across the network during the transition.**

Data-at-Rest encryption is a cluster-based toggle. Enabling or disabling the capability changes the data format. We recommend making this decision on its use prior to the deployment of a cluster running vSAN. Otherwise, the effort that it takes to roll through disk format process can be time consuming, and can impact performance. Once enabled, encryption services can impact the amount of processing required for storing data.

For more information on Data-At-Rest Encryption, see the vSAN Encryption Services document.

Data-in-Transit Encryption: Enabling in-flight Encryption on a vSAN Cluster

Data-in-Transit Encryption allows for vSAN traffic to be securely transmitted from host to host in a vSAN cluster. Data-in-transit encryption provides a complete over the wire encryption solution that address host/member authentication, data integrity/confidentiality, and embedded key management. It can be used on its own, or in conjunction with vSAN Data-at-Rest Encryption to provide an end-to-end encryption solution. Both capabilities use the same vSphere based FIPS 140-3 validated cryptographic modules.

Just like vSAN Data-at-Rest Encryption, Data-in-Transit Encryption is enabled and disabled at the cluster level. Unlike Data-at-Rest Encryption, **it does not use an external key management server** (KMS) which can make it extremely simple to operationalize. If a cluster uses both encryption features, the features will be independently responsible for its key management. Data-at-Rest Encryption will use an external KMS, while Data-in-Transit will manage its own host keys.

Health and Management of Data-in-Transit Encryption

The vSAN Health Services will periodically check the configuration state of the hosts that comprise the vSAN cluster. The Health Service will be the first place to check if there are difficulties with enabling Data-in-Transit Encryption.

Host BIOS settings and AES-NI Offloading

The vSphere cryptographic modules used for both methods of encryption can take advantage of AES-NI offloading to minimize the CPU consumption of the hosts. Modern CPUs are much more efficient with this offloading than older CPUs, so the impact of this offloading will depend on the generation of CPUs in the hosts

Recommendation: Prior to deployment, check the BIOS to ensure that AES offloading is enabled.

The Potential Impact on Performance

Data-in-transit encryption is an additional data service that, as one might expect, demands additional resources. Performance considerations and expectations should be adjusted when considering these types of security features. The degree of impact will be dependent on the workloads and hardware specifications of the environment. **Data-in-Transit does have the potential to impact guest VM latencies, since all over-the-wire communication to synchronously replicate the data must be encrypted and decrypted in flight.**

Data-at-Rest Encryption: Using vSAN and vSphere Encryption Together

These two features provide encryption at different points in the stack, and have different pros and cons for using each. **We highly recommend using vSAN-based encryption capabilities as opposed to VM encryption**. Using vSAN-based encryption will ensure that space efficiency capabilities like global deduplication are maintained.

Skyline Health for vSAN reports when a VM has an encryption policy (for VM encryption) and also resides on an encrypted vSAN cluster. This alert is only cautionary.

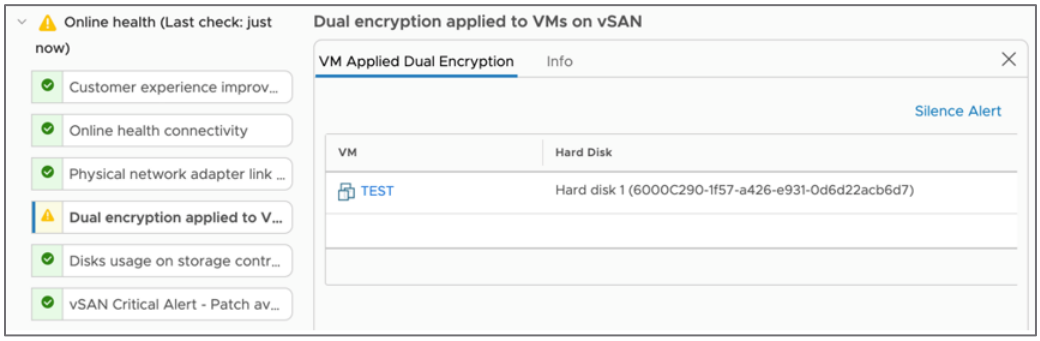


Figure. The vSAN health check reporting the use of multiple encryption types used together

Data-at-Rest Encryption: Performing a Shallow Rekey

Key rotation is a strategy often used to prevent long-term use of the same encryption keys. When encryption keys are not rotated on a defined interval, it can be difficult to determine their trustworthiness. If the encryption keys have not been changed (or rotated), the data could possibly be decrypted and recovered from the suspected failed storage device. For more information, see the "vSAN Encryption Services" document.

Shallow rekey operations are extremely light weight. All Disk Encryption Keys (DEK)s remain the same but are re-wrapped with a new KEK. Shallow rekey operations are very fast due to little to no data movement, and are the most common of the two key rotation methods.

Recommendation: Implement a KEK rotation strategy that aligns with organizational security and compliance requirements.

Data-at-Rest Encryption: Performing a Deep Rekey

While vSAN ESA uses a cluster-wide Disk Encryption Key (DEK), the object data has object-specific keys that are used. This ensures that object data in VM uses different keys than any other VM. Note however that a rekey operation will perform a rekey across all data on the cluster, and does not give a choice for rekeying discrete objects.

Deep rekey operations are a more resource intensive endeavor, and careful consideration should be taken when performing the action. For more information, see the post: "Key Rotation Options for vSAN ESA in vMware Cloud Foundation 5.1 and vSAN 8 U2."

iSCSI: Identification and Management of iSCSI Objects in a vSAN Cluster

vSAN iSCSI objects can appear different than other vSAN objects in vSAN reports—usually reporting as "unassociated" because they aren't mounted directly into a VM as a VMDK—but rather via a VM's guest OS iSCSI initiator, into which vSAN has no visibility. If you use the vCenter RVC to query vSAN for certain operations, be aware that iSCSI LUN objects as well as vSAN performance service objects (both of which are not directly mounted into VMs) will be listed as "unassociated"—this does NOT mean they are unused or safe to be deleted. For more information, see the "vSAN iSCSI Target Usage Guide."

So, how can you tell if objects are in use by the performance service or iSCSI? After logging in to the vCenter server, iSCSI objects or performance management objects could be listed and shown as unassociated when querying with RVC command vsan.obj_status_report.

These objects are not associated with a VM, but they may be valid vSAN iSCSI objects on the vSAN datastore and should not be deleted. If the intention is to delete some other unassociated objects and save space, **please contact the VMware GS team for assistance.** The following shows how to identify unassociated objects as vSAN iSCSI objects and verify from the vSphere web client.

Login to vCenter via ssh and launch the RVC, then navigate to the cluster:

```
root@sc-rdops-vm03-dhcp-93-66 [ ~ ]# rvc administrator@vsphere.local@localhost

Welcome to RVC. Try the 'help' command.

0 /

1 localhost/

> cd localhost/
```

Run the vSAN object status report in RVC:

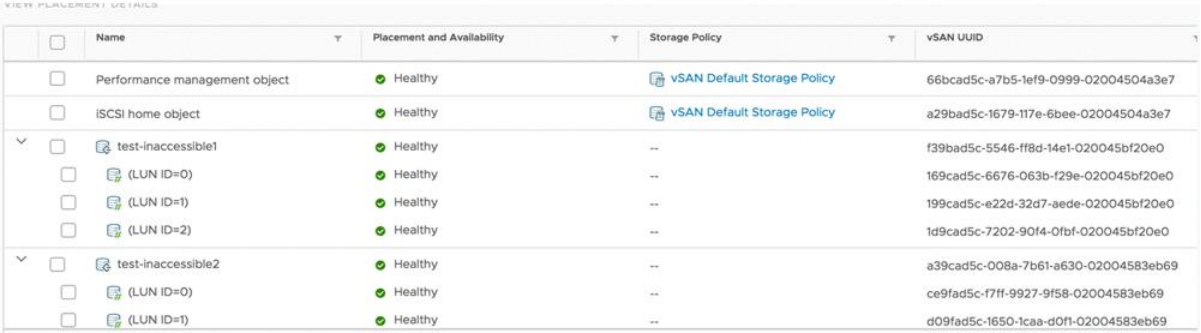
```
/localhost> vsan.obj_status_report -t /localhost/VSAN-DC/computers/VSAN-Cluster/…

Histogram of component health for non-orphaned objects

+-------------------------------------+------------------------------+
```

```
| Num Healthy Comps / Total Num Comps | Num objects with such status |
+-------------------------------------+------------------------------+
| 3/3 (OK) | 10 |
+-------------------------------------+------------------------------+
Total non-orphans: 10
Histogram of component health for possibly orphaned objects
+-------------------------------------+------------------------------+
| Num Healthy Comps / Total Num Comps | Num objects with such status |
+-------------------------------------+------------------------------+
+-------------------------------------+------------------------------+
Total orphans: 0
Total v9 objects: 10
+----------------------------------------+---------+---------------------------+
| VM/Object | objects | num healthy / total comps |
+----------------------------------------+---------+---------------------------+
| Unassociated objects | | |
| a29bad5c-1679-117e-6bee-02004504a3e7 | | 3/3 |
| ce9fad5c-f7ff-9927-9f58-02004583eb69 | | 3/3 |
| a39cad5c-008a-7b61-a630-02004583eb69 | | 3/3 |
| d49fad5c-bace-8ba3-9c7a-02004583eb69 | | 3/3 |
| d09fad5c-1650-1caa-d0f1-02004583eb69 | | 3/3 |
| 66bcad5c-a7b5-1ef9-0999-02004504a3e7 | | 3/3 |
| 169cad5c-6676-063b-f29e-020045bf20e0 | | 3/3 |
| f39bad5c-5546-ff8d-14e1-020045bf20e0 | | 3/3 |
| 199cad5c-e22d-32d7-aede-020045bf20e0 | | 3/3 |
| 1d9cad5c-7202-90f4-0fbf-020045bf20e0 | | 3/3 |
+----------------------------------------+---------+---------------------------+
```

Cross-reference the “Unassociated objects” list UUIDs with the vSAN iSCSI objects, as well as the “iSCSI home object” and the “performance management object” in the vSphere web client under vSAN Cluster → Monitor → vSAN → Virtual Objects and compare the UUIDs under the “vSAN UUID” column with those in the “Unassociated objects” report from RVC. If UUIDs appear in both lists, they are NOT safe to remove.



Name	Placement and Availability	Storage Policy	vSAN UUID
Performance management object	Healthy	vSAN Default Storage Policy	66bcad5c-a7b5-1ef9-0999-02004504a3e7
iSCSI home object	Healthy	vSAN Default Storage Policy	a29bad5c-1679-117e-6bee-02004504a3e7
test-inaccessible1	Healthy	--	f39bad5c-5546-ff8d-14e1-020045bf20e0
(LUN ID=0)	Healthy	--	169cad5c-6676-063b-f29e-020045bf20e0
(LUN ID=1)	Healthy	--	199cad5c-e22d-32d7-aede-020045bf20e0
(LUN ID=2)	Healthy	--	1d9cad5c-7202-90f4-0fbf-020045bf20e0
test-inaccessible2	Healthy	--	a39cad5c-008a-7b61-a630-02004583eb69
(LUN ID=0)	Healthy	--	ce9fad5c-f7ff-9927-9f58-02004583eb69
(LUN ID=1)	Healthy	--	d09fad5c-1650-1caa-d0f1-02004583eb69

Figure. Enumerated objects, related storage policies, and vSAN UUIDs

Again, if in any doubt, **please contact the VMware GS team for assistance**.

File Services: Introducing it into an Existing Environment

vSAN File Services allows for vSAN administrators to easily deliver file services on a per cluster basis using any vSAN cluster. Providing both NFS and SMB file services in a manner that is native to the hypervisor allows for a level of flexibility and ease of administration that is otherwise difficult or costly to achieve with stand-alone solutions. For more information on vSAN File Services, see the "vSAN File Services" document.

Enabling vSAN File Services in an environment introduces several operational considerations. vSAN File Services can be unique in that it may require additional considerations with the infrastructure that may or may not be related to the hypervisor. Some of those considerations include:

- Supported topology types
- Authentication options through Active Directory for SMB and Kerberos for NFS
- Supported protocol versions and how to connect clients to the shares

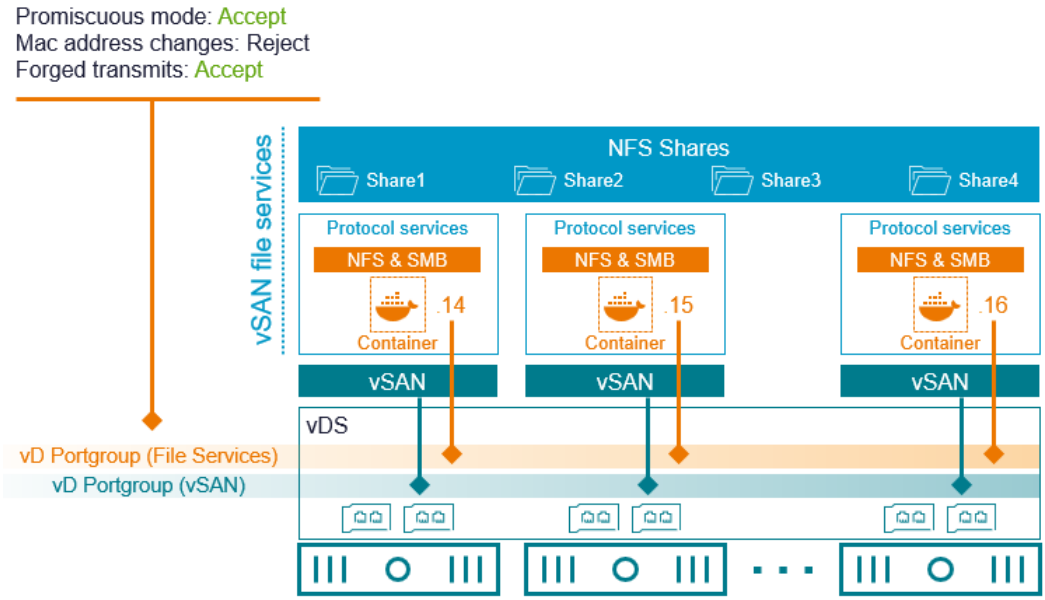
Note that in the VMware documentation and in the product UI, the term "share" may be used interchangeably with both SMB and NFS. The term "share" is used to simplify the language when discussing multiple protocols. Windows-based SMB have historically referenced them as "shares" while Unix and Linux based systems typically refer to them as an "NFS export" that the NFS client will mount.

Recommendations on Introducing vSAN File Services into your environment

Since vSAN File Services is a relatively new feature, successfully introducing it into an environment can be achieved with preparation and familiarity.

- **Run the latest edition of vSAN**. File Services have improved substantially over recent versions, filling in some of the initial gaps in scaling, performance, and capabilities compared to early editions.
- **Understand the limits.** vSAN File Services does not allow ESXi hosts to connect directly to File Services via NFS for the purpose of presenting storage for VMs. A share served by vSAN File Services can only be used for SMB, or NFS: Not both concurrently. The vSAN File Services FAQ and VMware Docs outline the common limits that you should be aware of prior to deployment.
- **Become familiar with the prerequisites required for the setup of vSAN file services.** Enabling and configuring vSAN File Services will require additional IP addresses for the respective protocol services containers (up to a maximum of 64 for 64 host clusters) with forward and reverse DNS records.
- **Decide on the approach used for port group used by vSAN File Services.** The port group that is used for vSAN File Services will automatically enable promiscuous mode and forged transmits if those settings are not enabled already. If NSX-based networks are being used, ensure that similar settings are configured for the provided network entity from the NSX admin console, and all the hosts and File Services nodes are connected to the desired NSX-T network. An administrator may decide to put the IP addresses of protocol service container created by vSAN File

Services on their own dedicated port group, or use another existing port group that previously did not have these settings enabled. Internal requirements (e.g. security) or other constraints may dictate the decision. Both approaches are supported.



- **Build out a test cluster to become familiar with the deployment process and configuration settings**. This will allow for easy experimentation to become familiar with the feature and the configuration. It can also serve as a way to test out the upgrade process, as well as review future editions of vSAN File Services.
- **Use the test cluster to ensure the proper configuration of Active Directory.** Configuration of Active Directory and Kerberos settings for vSAN File Services will be highly dependent on your organization's Active Directory Configuration. The deployment wizard also has guidance with this, including the requirements of a dedicated OU in Active Directory for use by vSAN File Services.
- **Set quotas.** vSAN file services can provide as much capacity for file shares as provided by the cluster. vSAN provides share warning thresholds as well as a hard quota to protect against the consumption of storage capacity beyond what is intended.
- **Become familiar with creating shares and their associated connection strings.** A connection string is the string of text an NFS or SMB client will use to establish a connection to the share. This connection string will be different for SMB, NFS v3, and NFS v4.1. Learn where to find these strings, and how to connect the clients.
- **Learn how to monitor.** vSAN provides the ability to monitor the activities of vSAN File Services. The share can be selected in the vSAN Performance Service to look at the demand on the share over a period of time. The Skyline Health checks also continuously check for various aspects of the cluster related to vSAN File Service health. The vCenter Server UI even allows you to see which objects make up the given file share. This can be found in the "Virtual Objects" view followed by clicking the "File Shares" icon to filter the object listing.

Recommendation: Do not use vSAN File Services as a location for important host logging, core dumps or scratch locations of the hosts that comprise the same cluster providing the file services. This could create a circular dependency and prevent the logging and temporary data from being available during an unexpected condition that requires further diagnostics.

Section 9: Stretched Clusters

See the vSAN Stretched Cluster Guide for more information on how to best operationalize vSAN when deployed as a stretched cluster.

Section 10: 2-Node Clusters

See the vSAN 2-Node Cluster Guide for more information on how to best operationalize vSAN when deployed as a 2-node cluster.

Section 11: vCenter Server Maintenance and Event Handling

Upgrade Strategies for vCenter Server Powering One or More vSAN Clusters

It is common for vCenter to host multiple vSAN clusters, these could be at different ESXi versions to each other, and as such, different vSAN versions. This is a fully supported configuration, but it is a good idea to ensure that your vCenter and ESXi versions are compatible with one-another.

Recommendation: When in doubt, simply run the very latest version of vCenter Server. The hosts that the vCenter Server manages do not need to match. This also sets up the environment well for upgrades, as vCenter Server is typically the first aspect to be upgraded.

Replacing vCenter Server on an Existing vSAN Cluster

There may be instances in which you need to replace the vCenter server that hosts some vSAN clusters. While vCenter acts as the interaction point for vSAN and is used to set it up and manage it, it is not the only source of truth and is not required for steady-state operations of the cluster. If you replace the vCenter server, your workloads will continue to run without it in place.

Replacing the vCenter server associated with a vSAN cluster can be done, but is not without its challenges or requisite planning. For more detailed information, see the blog post: "Replacing a vCenter Server for Existing vSAN Hosts."

Scenario: An all-flash vSAN OSA cluster on 7 U2 needs migrated to a new vCenter server with deduplication and compression enabled.

- Ensure the target vCenter server has the same vSphere version as the ESXi hosts, or higher (same is preferable).
- Create a new cluster on the new vCenter server with the same settings as the source cluster (vSAN enabled, DD&C, encryption, HA, DRS) and ensure the Disk Claim Mode is set to Manual.
- If you are using a Distributed Switch on the source vCenter server, export the vDS and import it into the new vCenter server, ensure "Preserve original distributed switch port group identifiers" is NOT checked upon import.
- Recreate all SPBM policies on the target vCenter server to match the source vCenter server.
- Disconnect all hosts from the source vCenter server.
- Remove hosts from the source vCenter server inventory.
- Add hosts into the new vCenter server.
- Drag the hosts into the new cluster.
- Verify hosts and VMs are contactable.
- Run esxcfg-advcfg -s 0 /VSAN/IgnoreClusterMemberListUpdates on all hosts.
- Configure hosts to use the imported vDS one by one, ensuring connectivity is maintained.
- Reconfigure a VM with the same policy as source—ensuring no resynchronization when the VM is reconfigured.
- For each SPBM policy, reconfigure one of each VM as a test to ensure no resynchronization is performed.
- Once verified, reconfigure all VMs in batches with their respective SPBM policies.

Recommendation: If you are not completely comfortable with the above procedure and doing this in a live environment, please open a ticket with GS and have them guide you through the procedure.

Protecting vSphere Distributed Switches Powering vSAN

Virtual switches and the physical uplinks that are associated with them are the basis for connectivity in a vSAN powered cluster. Connectivity between hosts is essential for vSAN clusters since the network is the primary storage fabric, as opposed to three-tier architectures that may have a dedicated storage fabric.

VMware recommends the use of vSphere Distributed Switches (VDS) for vSAN. Not only do they provide additional capabilities to the hosts, they also provide a level of consistency, as the definition of the vSwitch and the associated port groups are applied to all hosts in the cluster. Since a vDS is a management construct of vCenter, it is recommended to ensure these are protected properly, in the event of unknown configuration changes, or if vCenter server is being recreated and introduced to an exist ing vSAN cluster.

Procedures

The specific procedures for exporting, importing, and restoring VDS configurations can be found at "Backing Up and Restoring a vSphere Distributed Switch Configuration." The process for each respective task is quite simple, but it is advised to become familiar with the process, and perhaps experiment with a simple export and restore in a lab environment to become more familiar with the task. This will help minimize potential confusion for when it is needed most. Inspecting the data.xml file included in the zip file of the backup can also provide a simple way to review the settings of the vDS.

Recommendation: The VDS export option provides the ability to export just the vDS, or the vDS and all port groups. You may find it helpful to perform the export twice, using both options. This will allow for maximum flexibility in any potential restoration activities.

vDS can apply to more than one cluster. In any type of scenarios in which the VDS is restored from a backup, it will be important for the administrator to understand what clusters and respective hosts it may impact. Understanding this clearly will help minimize the potential impact of unintended consequences, and may also influence the naming/taxonomy of the vDS used by an organization, as the number of clusters managed by vCenter continues to grow.

Section 12: Upgrade Operations

Upgrading and Patching vSAN Hosts

Upgrading and patching vSAN hosts is very similar to the process for vSphere hosts using traditional storage. The unique role that vSAN plays means there are additional considerations to be included in operational practices to ensure predictable results.

vSAN is a cluster-based storage solution. Each ESXi host in the vSAN cluster participates in a manner that provides a cohesive, single storage endpoint for the VMs running in the cluster. Since it is built directly into the hypervisor, ESXi depends heavily on the expected interaction between hosts to provide a unified storage system. This can be dependent on consistency of the following:

- The version of ESXi installed on the host.
- Firmware versions of key components, such as storage controllers, NICs, and BIOS.
- Driver versions (VMware-provided "inbox" or vendor-provided "async") for the respective devices.

Inconsistencies between any or all of these may change the expected behavior between hosts in a cluster. Therefore, **avoid mixing vSAN/ESXi versions in the same cluster for general operation.** Limit inconsistency of versions on hosts to the cluster upgrade process, where updates are applied one host at a time until complete. The Knowledge Base contains additional recommendations on vSAN upgrade best practices.

Recommendation: Subscribe to the VCG Notification service in order to stay informed of changes with compatibility and support of your specified hardware and the associated firmware, drivers, and versions of vSAN.

Tips for using vLCM in an Existing Environment

The recommendations listed below will improve the transition to vLCM in your environment.

- **Download and install the latest vendor plugin(s).** This is the first step in establishing the intelligence between vLCM, the respective vendor repositories, VMware, and the hosts in your cluster. Depending on the server vendor, the configuration and operation of the plugin may vary slightly. VMware built this flexibility with the vendors in mind to best accommodate their repositories and capabilities.
- **Ensure you have the server OEM's vLCM plugin/HSM downloaded and installed.** This will allow you to experience, and experiment with more complete lifecycle management of the host.
- **Ensure your operational run-books for server updates are updated to reflect vLCM.** Make sure old procedures such as manually updating VIBs does not undermine the lifecycle management of the cluster. Since a vLCM enabled cluster manages this differently, applying those same practices may incorrectly introduce "drift" (move away from the desired state) into a cluster.

The benefits vLCM brings show up most when scale, consistency, and frequency of updates are top-of-mind. Clusters are managed by a single desired-state image which helps ensure consistency and reduces the obstacles typically associated with the lifecycle management process.

Upgrade Considerations for Different vSAN Topology Types

Lifecycle management using vLCM will be very similar regardless of the type of vSAN deployment, or topology.

- **2-Node topologies.** In 2-node clusters, for all clusters running vSAN 7 U1 or later, the witness host should be updated prior to upgrading the hosts in the cluster that use the witness host. In later versions of vSAN, vLCM will be able to update the virtual witness host appliance.
- **Stretched cluster topologies.** In stretched clusters, for all clusters running vSAN 7 U1 or later, the witness host should be updated prior to upgrading the hosts in the cluster that use the witness host. In later versions of vSAN, vLCM will be able to update the virtual witness host appliance.

- **vSAN storage clusters and/or vSAN HCI with datastore sharing**. In a disaggregated topology, the server cluster will generally maintain an independent lifecycle management from the client clusters that mount the datastore. While it is not required, it may be best to factor in the dependency or relationship to these clusters during upgrades. However, since vLCM will always update a vSAN cluster one host at a time per cluster, the impact should be minimal.

Multi-Cluster Upgrading Strategies

While VMware continues to introduce to vSAN all new levels of performance, capabilities, robustness, and ease of use, the respective vSAN clusters must be updated to benefit from these improvements. While the upgrading process continues to be streamlined, environments running multiple vSAN clusters can benefit from specific practices that will deliver a more efficient upgrade experience.

vCenter server compatibility

In a multi-cluster environment, vCenter server must be running the version equal to, or greater than the version to be installed on any of the hosts for the clusters it manages. Ensuring that vCenter server is always running the very latest edition will guarantee compatibility among all potential host versions running in a multi-cluster arrangement and introduce enhancements to vCenter that independent from the clusters it is managing.

Phasing in new versions of vSAN

As noted in the "Upgrading and Patching vSAN Hosts" section, vSAN is a cluster-based solution. Therefore, upgrades should be approached per cluster, not per host. With multi-cluster environments, IT teams can phase in a new version of vSAN per cluster to meet any of their own vetting, documentation, and change control practices. Similar to common practices in application maintenance, upgrades can be phased in on less critical clusters for testing and validation prior to rolling out the upgrade into more critical clusters.

Parallel upgrades

While vSAN limits the upgrade process to one host at a time within a vSAN cluster, cluster upgrades can be performed concurrently if desired. vLCM supports up to 64 concurrent cluster update activities. This can speed up host updates across larger data centers. Whether to update one cluster or multiple clusters at a time is at your discretion based on understanding tradeoffs and your procedural limitations.

Updating more hosts simultaneously should be factored into the vSAN cluster sizing strategy. More clusters with fewer hosts allows for more parallel remediation than fewer clusters with more hosts. For example, an environment with 280 hosts could cut remediation time in half if the design was 20 clusters of 14 hosts each, as opposed to 10 clusters of 28 hosts each.

Since a vSAN cluster is its own discrete storage system, administrators may find greater agility in operations and troubleshooting. "vSAN Cluster Design—Large Clusters Versus Small Clusters" discusses the decision process of host counts and cluster sizing in great detail.

Larger environments with multiple vSAN clusters may have different generations of hardware. Since drivers and firmware can cause issues during an update process, concurrent cluster upgrades may introduce operational challenges to those managing and troubleshooting updates. Depending on the age and type of hardware, a new version of vSAN could be deployed as a pilot effort to a few clusters individually, then could be introduced to a larger number of clusters simultaneously. Determine what level of simultaneous updates is considered acceptable for your own organization.

Recommendation: Focus on efficient delivery of services during cluster updates, as opposed to speed of update. vSAN restricts parallel host remediation. A well-designed and -operating cluster will seamlessly roll through updating all hosts in the cluster without interfering with expected service levels.

Why are vSAN clusters restricted to updating one host at a time? Limiting to a single host per cluster helps reduce the complexity of subtracting not only compute resources but storage capacity and performance. Factoring in available capacity in addition to compute resources is unique to an HCI architecture. Total available host count can also become important for some data placement policies such as FTT=2 using RAID-6 erasure coding. Limiting the update process to one host at a time per cluster also helps avoid this complexity, while reducing the potential need for data movement due to resynchronization.

Upgrading Large vSAN Clusters

Standard vSAN clusters can range from 3 to 64 hosts. Since vSAN provides storage services per cluster, a large cluster is treated in the same way as a small cluster: as a single unit of services and management. Maintenance should occur per cluster and is sometimes referred to as a "maintenance domain."

Upgrading vSAN clusters with a larger quantity of hosts is no different than upgrading vSAN clusters with a smaller quantity of hosts. In addition, those described in "Upgrading and Patching vSAN Hosts," there are a few additional host upgrade considerations to be mindful of during these update procedures.

vLCM is limited to updating one host at a time in a vSAN cluster. The length of time for the cluster to complete an update is proportional to the number of hosts in a cluster. To upgrade more than one host at a time, reduce the size of the maintenance domain by creating more clusters comprising fewer hosts. This smaller maintenance domain will allow for more hosts (one per cluster) to perform parallel upgrades.

Designing an environment that has a modest maintenance domain is one of the most effective ways to improve operations and maintenance of a vSAN-powered environment. For more information on this approach, see the topic "Multi-Cluster Upgrading Strategies."

While no more than one host per vSAN cluster can be upgraded at a time, there are some steps that can be taken to potentially improve the upgrade speed.

- Use hosts that support the Quick Boot feature. This can help host restart times. Since hosts in a vSAN cluster are updated one after the other, reducing host restart times can significantly improve the completion time of the larger clusters.
- If a large cluster has relatively few resources used, an administrator may be able to place multiple hosts into maintenance mode safely without running short of storage and capacity resources. Updates will still occur one host at a time, but this may save some time placing the respective hosts into maintenance mode. This would only be possible in large clusters that are underused, and actual time savings may be negligible.

Recommendation: Focus on efficient delivery of services during cluster updates, as opposed to speed of update. vSAN restricts parallel host remediation of hosts. A well-designed and -operating cluster will seamlessly roll through updating all hosts in the cluster without interfering with expected service levels.

Larger vSAN clusters may better absorb reduced resources as a host enters maintenance mode for the update process. Proportionally, each host contributes a smaller percentage of resources to a cluster. Large clusters may also see slightly less data movement than much smaller clusters to comply with the "Ensure accessibility" data migration option when a host is entered into maintenance mode. For more information on the tradeoffs between larger and smaller vSAN clusters, see "vSAN Cluster Design—Large Clusters Versus Small Cluster"

Upgrading Firmware and Drivers for NICs and Storage Controllers

Outdated or mismatched firmware and drivers for NICs and Storage Controllers can impact VM and or vSAN I/O handling. While VUM handles updates of firmware and drivers for a limited set of devices, firmware and driver updates remain a largely manual process. Whether installed directly on an ESXi server from the command line or deployed using vLCM, ensure the correct firmware and drivers are installed, remain current to the version recommended, and are a part of the cluster lifecycle management process.

vLCM strives to simplify the coordination of firmware and driver updates for select hardware. It has a framework to coordinate the fetching of this software from the respective vendors for the purposed of building a single desired state image for the hosts.

Recommendation: Subscribe to the VCG Notification service in order to stay informed of changes with compatibility and support of your specified hardware and the associated firmware, drivers, and versions of vSAN.

Section 13: Customizing Utility Storage Space

Resize Custom Namespace Objects

The ability to resize custom namespace objects such as ISO directories and content libraries was introduced in vSAN 8 U1, supporting both ESA and OSA. This capability is not available in the UI, but is available via API, and through PowerCLI using multiple cmdlets. The example below demonstrates the ability to view and resize the desired namespace object.

After connecting to VI server,

```
PS C:\Users\User1> $services = Get-view 'ServiceInstance'

PS C:\Users\User1> $datastoreMgr = Get-view $services.Content.DatastoreNamespaceManager

PS C:\Users\User1> $datacenter = get-datacenter

PS C:\Users\User1>
$datastoreMgr.DeleteDirectory($datacenter.ExtensionData.MoRef,"/vmfs/volumes/vsan:526916282a8
ec9e1-95c4972ba093a2ec/6771b663-4b89-170b-f416-0200368ec988")

PS C:\Users\User1> $datastore = get-datastore

PS C:\Users\User1>
$datastoreMgr.CreateDirectory($datastore.ExtensionData.MoRef,"CodyTest2",$null,16777216)

/vmfs/volumes/vsan:526916282a8ec9e1-95c4972ba093a2ec/fa77b663-8748-3a1e-e9a6-0200368ec988

PS C:\Users\User1>
$datastoreMgr.QueryDirectoryInfo($datacenter.ExtensionData.MoRef,"/vmfs/volumes/vsan:52691628
2a8ec9e1-95c4972ba093a2ec/fa77b663-8748-3a1e-e9a6-0200368ec988")

Capacity Used

-------- ----

16777216 2223

PS C:\Users\User1> $datastoreMgr.IncreaseDirectorySize($datacenter.ExtensionData.MoRef,
"/vmfs/volumes/vsan:526916282a8ec9e1-95c4972ba093a2ec/fa77b663-8748-3a1e-e9a6-0200368ec988",
33554432)
```

Section 14: Monitoring vSAN Health

Remediating vSAN Health Alerts

The vSAN Skyline health UI provides an end-to-end approach to monitoring and managing the environment. Health finding alerts are indicative of an unmet condition or deviation from expected behavior.

The alerts can typically stem out of:

- Configuration inconsistency
- Exceeding software/hardware limits
- Hardware incompatibility
- Failure conditions

The ideal methodology to resolve a Skyline health alert is to correct the underlying situation. An administrator can choose to suppress the alert in certain situations.

vSAN 8 U1 introduced a new way to understand the severity and priority of triggered health checks in a vSAN cluster. Through a sophisticated set of relationships and weighting, Skyline health for vSAN can help you understand how severe the triggered health checks are, and what triggered health check is most important to remediate. For more information, see the post: "Skyline Health Scoring, Diagnostics, and Remediation in vSAN 8 U1."

Checking Object Status and Health during a Failure

An object is a fundamental unit in vSAN around which availability and performance are defined. This is done by abstracting the storage services and features of vSAN and applying them at an object level through SPBM. For a primer on vSAN objects and components, see the post: "vSAN Objects and Components Revisited."

At a high level, an object's compliance with the assigned storage policy is enough to validate its health. In certain scenarios, it may be necessary to inspect the specific state of the object, such as in a failure.

In the event of a failure, ensure all objects are in a healthy state or recovering to a healthy state. vSAN object health check provides a cluster-wide overview of the object's health and its respective states. This health check can be accessed by clicking on the vSAN cluster and viewing the Monitor tab. The data section comprises information specific to the object health check.

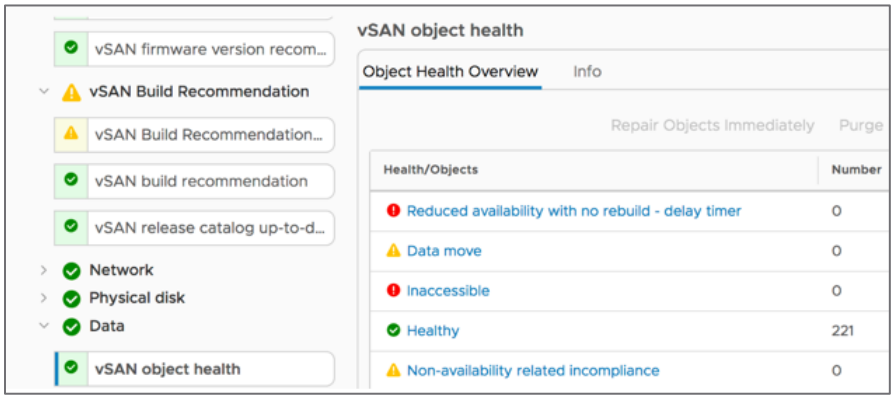


Figure. Viewing object health with the vSAN health checks.

On failure detection, vSAN natively initiates corrective action to restore a healthy state. This, in turn, reinstates the object's compliance with the assigned policy. The health check helps quickly assess the impact and validates that restoration is in progress. In certain cases, based on the nature of failure and the estimated restoration time, an administrator may choose to override or expedite the restoration.

Monitoring and Management of vSAN Object Components

VMware vSAN uses a data placement approach that is most analogous to an object store. VMs that live on vSAN storage are comprised of several storage objects - which can be thought of as a unit of data. VMDKs, VM home namespace, VM swap areas, snapshot delta disks, durability data, and snapshot memory maps are all examples of storage objects in vSAN. Object data is placed across hosts in the cluster in a manner that ensures data resilience. Resilience, space efficiency, security, and other settings related to a vSAN object are easily managed by the Administrator through the use of storage policies. For a primer on vSAN objects and components, see the post: "vSAN Objects and Components Revisited."

A vSAN object is comprised of one or more "components." Depending on the object size, applied storage policy, and other environmental conditions, an object may consist of more than one component. This sharded data is simply an implementation detail of vSAN and not a manageable entity.

Note that the **vSAN ESA has a maximum component limit of 27,000 per host.** It also has a maximum component limit of 3,000 data components per disk, and 3,000 metadata components per disk. It is best practice to ensure you have at least 6 storage devices in a host to make sure that you are not subject to the per-storage device limit.

A component limit per host exists primarily to keep host resource consumption to reasonable levels. A distributed scale-out storage system like vSAN must create and manage data about the data. This metadata is what allows for the seamless scalability and adaptability of a vSAN cluster. As data is sharded into more components, additional resources may be consumed to manage the data. The component limit helps keep memory and CPU requirements of vSAN to reasonable levels while still maintaining resources for guest VM consumption.

The component and VM limits add another dimension to capacity management considerations that will impact both the design of a vSAN cluster, as well as the operation of a vSAN cluster. See the section topic "Estimating Approximate 'Free/Usable Space in vSAN Cluster" for more details.

Recommendations to Mitigate

There can be some circumstances where component counts in a vSAN cluster are approaching their limit. In those relatively rare cases, one can alleviate this issue by adding another host to the vSAN cluster. Adding another host can quickly relieve the pressure of clusters approaching a per-host maximum. Let's use a 7-host vSAN cluster with a theoretical component limit of 63,000 as an example. A component count of 57,000 would trigger a red health alert error, as it exceeds 90% utilization. Adding a host to the cluster would increase the theoretical component limit to 72,000. The component count of 57,000 would not even trigger a health alert warning, as it falls below the 80% threshold. Once the host was added, the Automatic Rebalancing in a vSAN cluster would take steps to evenly distribute the data across the hosts.

Viewing vSAN Cluster Partitions in the Health Service UI

vSAN employs a highly resilient and distributed architecture. The network plays an important role in accommodating this distributed architecture.

Each host in the vSAN cluster is configured with a VMkernel port tagged with vSAN traffic and should be able to communicate with other hosts in the cluster. If one or more hosts are isolated, or not reachable over the network, the objects in the cluster may become inaccessible. To restore, resolve the underlying network issue.

There are multiple network-related validations embedded as part of the health service to detect and notify when there is an anomaly. These alerts ought to be treated with the highest priority, specifically the vSAN cluster partition. Health service UI can provide key diagnostic information to help ascertain the cause.

Recommendation: Focus on discovering the root cause of the cluster partition issue. A triggered cluster partition health check is often a symptomatic triggered alert as a result of some other issue that was the cause. If the cause is from an issue captured by another Skyline health check, this will show up in the new health check correlation feature as the "Primary issue." of the cluster partition.

Accessing the health service UI

The vSAN Skyline Health service UI provides a snapshot of the health of the vSAN cluster and highlights areas needing attention. Each health check validates whether a certain condition is met. It also provides guidance on remediation when there is a deviation from expected behavior. The UI can be accessed by clicking on the vSAN cluster and viewing the Monitor tab. The specific "vSAN cluster partition" health check is a good starting point to determine the cluster state. A partition ID represents the cluster as a single unit. In an ideal state, all hosts reflect the same partition ID. Multiple subgroups within the cluster indicate a network partition requiring further investigation. At a micro level, this plausibly translates to an object not having access to all of its components.

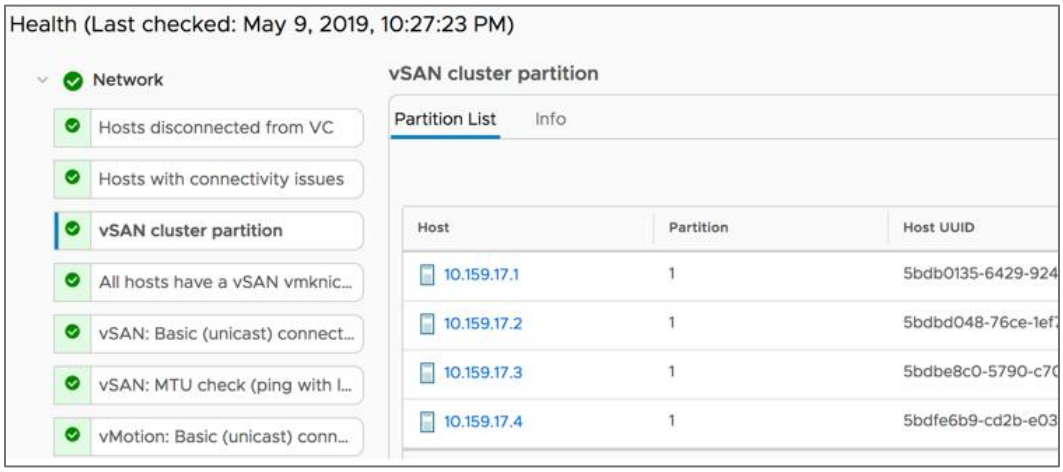


Figure. Identifying unhealthy network partitions in a vSAN cluster

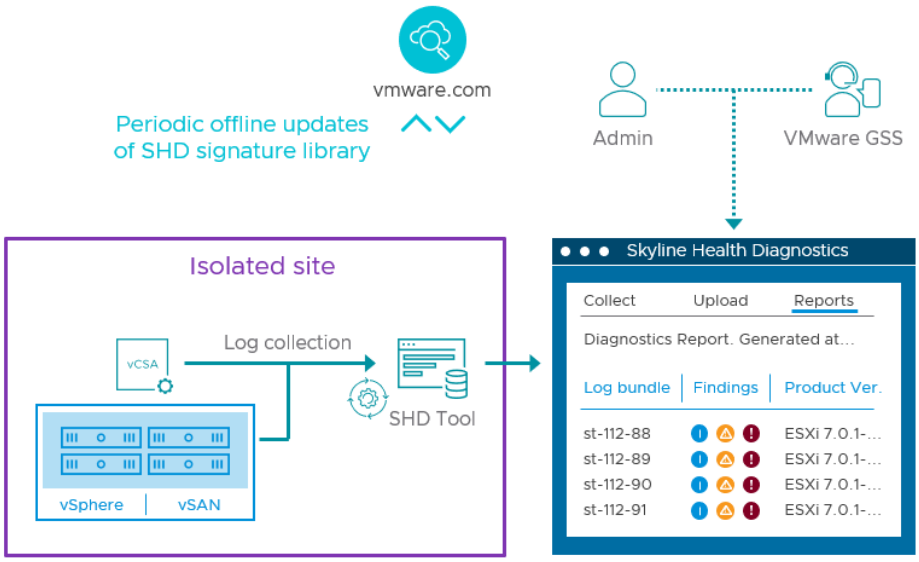
The network section in the health service UI has a plethora of network tests that cover some basic yet critical diagnostics, such as ping, MTU Check, Unicast connectivity, and host connectivity with vCenter. Each health check can systematically confirm or eliminate a layer in the network as the cause.

Recommendation: As with any network troubleshooting, a layered methodology is strongly recommended (top-down or bottom-up).

Monitoring and Management of Isolated vSAN Environments

Some environments require full isolation from any management access of a vSAN cluster to the Internet. While this is quite easy to do, it can pose additional operational challenges in asynchronous health check updates, troubleshooting and incident support with VMware Global Support.

vSAN provides a Skyline Health Diagnostics Tool (SHD). The Skyline Health Diagnostics tool is a self-service tool that brings some of the benefits of Skyline health directly to an isolated environment. The tool is run by an administrator at a frequency they desire. It will scan critical log bundles to detect issues, and give notifications and recommendations to important issues and their related KB articles. The goal for our customers is a faster time to resolution for issues, and for isolated environments, this is the tool to help with that.


vmware.com
Periodic offline updates
of SHD signature library
Admin
VMware GSS
Isolated site
vCSA
Log collection
SHD Tool
vSphere
vSAN
Skyline Health Diagnostics
Collect
Upload
Reports
Diagnostics Report. Generated at...
Log bundle
Findings
Product Ver.
st-112-88
st-112-89
st-112-90
st-112-91
ESXi 7.0.1-...
ESXi 7.0.1-...
ESXi 7.0.1-...
ESXi 7.0.1-...

Section 15: Monitoring vSAN Performance

Navigating Across the Different Levels of Performance Metrics

The vSAN performance service provides storage-centric visibility to a vSAN cluster. It is responsible for collecting vSAN performance metrics and presents them in vCenter. A user can set the selectable time window from 1 to 24 hours, and the data presented uses a 5-minute sampling rate. The data may be retained for up to 90 days, although the actual time retained may be shorter based on environmental conditions.

vSAN 8 U1 introduced **high resolution performance metrics**. This allows the ability for the administrator to monitor critical performance metrics using 30 second intervals, which will be much more representative of the actual workload than the longer, 5 minute intervals. For more information, see the post: "High Resolution Performance Monitoring in vSAN 8 U1." This capability is available in the ESA and the OSA.

Levels of navigation

The vSAN performance service presents metrics at multiple locations in the stack. vSAN-related data can be viewed at the VM level, the host level, the disk and disk group level, and the cluster level. Some metrics such as IOPS, throughput, and latency are common at all locations in the stack, while more specific metrics may only exist at a specific location, such as a host. The performance metrics can be viewed at each location simply by highlighting the entity (VM, host, or cluster) and clicking on

Monitor → vSAN → Performance.

Figure. Collects and renders performance data at multiple levels

The metrics are typically broken up into a series of categories, or tabs at each level. Below is a summary of the tabs that can be found at each level.

- VM Level
  - **VM:** This tab presents metrics for the frontend VM traffic (I/Os to and from the VM) for all VMs on the selected host.
  - **Virtual disk:** This presents metrics for the VM, broken down by the individual VMDK and especially helpful for VMs with multiple VMDKs.
- Host Level
  - **VM:** This tab presents metrics for the frontend VM traffic (I/Os to and from the VM) for all VMs on the selected host.

- **Backend:** This tab presents metrics for all backend traffic, as a result of replica traffic and resynchronization data.
- **Disks:** This tab presents performance metrics for the selected disk group, or the individual devices that compose the disk group(s) on a host.
- **Physical adapters:** This tab presents metrics for the physical uplink for the selected host.
- **Host network:** This tab presents metrics for the specific or aggregate VMkernel ports used on a host.
- **iSCSI:** This tab presents metrics for objects containing data served up by the vSAN iSCSI service.

- Cluster Level
  - **VM:** This tab presents metrics for the frontend VM traffic (I/Os to and from the VM) for all VMs living on the selected host.
  - **Backend:** This tab presents metrics for all backend traffic as a result of replica traffic and resynchronization data.
  - **iSCSI:** This tab presents metrics for objects containing data served up by the vSAN iSCSI service.

vSAN also includes a VM consolidated performance view. This solves some of the difficulties when attempting to compare performance metrics of more than one VM, side by side, and can be extremely helpful in doing comparisons and correlations. vSAN also has a Top Contributors view at the cluster level. This will help administrators quickly see VMs and disk groups that contribute to the most demand on resources provided by the vSAN cluster, and pair nicely with the VM Consolidated performance view.

Typically, the cluster level is an aggregate of a limited set of metrics, and the VM level is a subset of metrics that pertain to only the selected VM. The host level is the location at which there will be the most metrics, especially as it pertains to the troubleshooting process. A visual mapping of each category can be found in the image below.

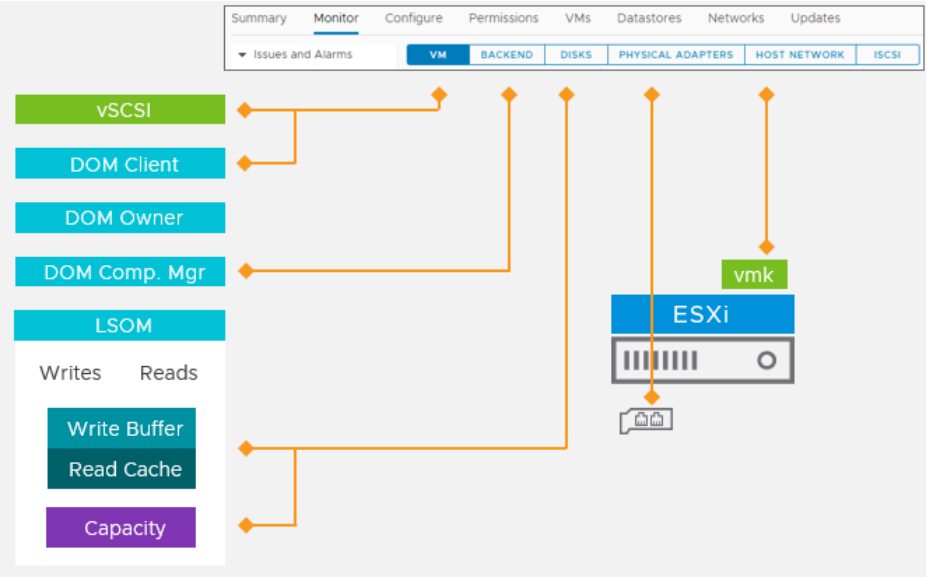


Figure. Provides vSAN-specific metrics and other vSphere-/ESXi-related metrics

Note that the performance service can only aggregate performance data up to the cluster level. It will not be able to provide aggregate statistics from multiple vSAN clusters. Aria Operations can achieve that result. Which are most important? They all relate to each other in some form or another. The conditions of the environment and the root cause of a performance issue will dictate which metrics are more significant than another. For more general information on troubleshooting vSAN, see the topic "Troubleshooting vSAN Performance" in this document. For a more detailed understanding of troubleshooting performance as well as definitions to specific metrics found in the vSAN performance service, see "Troubleshooting vSAN Performance"

The information provided by the vSAN performance service (rendered in the vCenter Server UI) is the preferred starting point for most performance data collection and analysis scenarios. Depending on the circumstances, there may be a need for additional tooling that exposes different types of data, such as the vSAN's IOInsight or vSAN's VM I/O Trip Analizer. This aims to help administrators identify the primary points of contention (bottlenecks) more easily. A list of common tools used for performance diagnostics are listed in Appendix B of the Troubleshooting vSAN Performance document.

Troubleshooting vSAN Performance

Troubleshooting performance issues is a common challenge for many administrators, regardless of the underlying infrastructure and topology. A distributed storage platform like vSAN also introduces other elements that can influence performance, and the practices for troubleshooting should accommodate those. Use the metrics in the vSAN performance service to isolate the sources of the performance issue.

While originally developed prior to the debut of the ESA, the framework described is generally applicable to ESA clusters as well.

The performance troubleshooting workflow

The basic framework for troubleshooting performance in a vSAN environment is outlined in the image below. Each of the five steps is critical to identifying the root cause properly and mitigating it systematically.

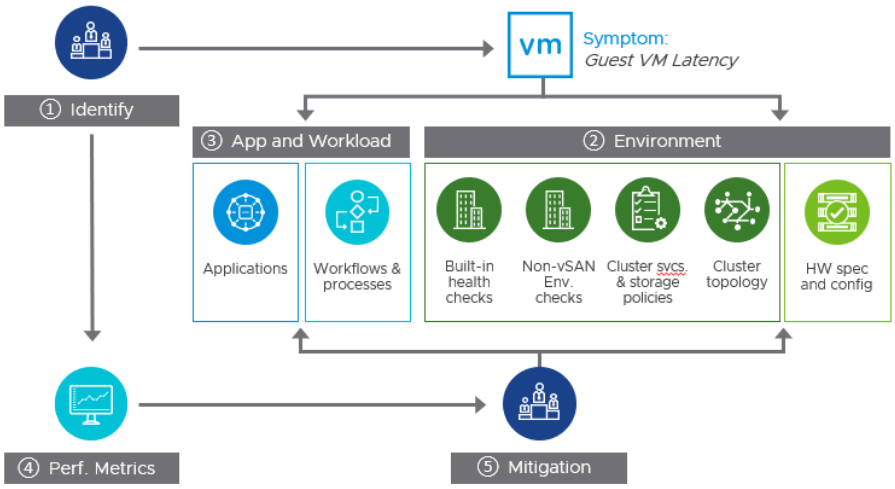


Figure. The troubleshooting framework

"Troubleshooting vSAN Performance" contains a more complete understanding of the performance troubleshooting process.

The order of review for metrics

Once steps 1–3 have been completed, begin using the performance metrics. The order in which the metrics are viewed can help decipher what level of contention may be occurring. The figure below shows the order in which to better understand and isolate the issue; it is the same order used in "Appendix C: Troubleshooting Example" in "Troubleshooting vSAN Performance."

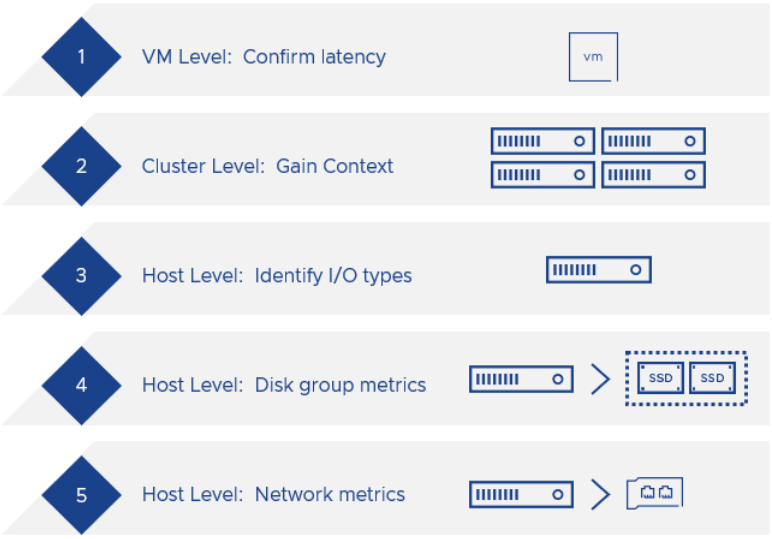


Figure. Viewing order of performance metrics

Here is a bit more context to each step:

- View metrics at the VM level to confirm unusually high storage related latency. This must be verified that there is in fact storage latency as seen by the guest VM.
- View metrics at the cluster level to provide context and look for other anomalies. This helps identify potential "noise" coming from somewhere else in the cluster.
- View metrics on the host to isolate the type of storage I/O associated with the latency.
- View metrics on the host, looking at the disk group level to determine type and source of latency.
- View metrics on the host, looking at the host network and VMkernel metrics to determine if the issue is network related.

Steps 3–5 assume that one has identified the hosts where the VM's objects reside. Host-level metrics should look at only the hosts where the objects reside for the particular VM in question. For further information on the different levels of performance metrics in vSAN, see the topic "Navigating Across the Different Levels of Performance Metrics."

Recommendation: Be diligent and deliberate when changing your environment to improve performance. Changing multiple settings at once, overlooking a simple configuration issue, or not measuring the changes in performance can often make the situation worse, and more complex to resolve.

Monitoring Resynchronization Activity

Resynchronizations are a common activity that occur in a vSAN environment. They are simply the process of replicating the data across the vSAN cluster so it adheres to the conditions of the assigned storage policy that determines levels of resilience, space efficiency, and performance. Resynchronizations occur automatically and are the result of policy changes to an object, host or disk group evacuations, rebalancing of data across a cluster, and object repairs should vSAN detect a failure condition.

Methods of visibility

Resynchronization visibility occurs in multiple ways: through vCenter, Aria Operations, and PowerCLI. The best method depends on what you attempt to view, and familiarity with the tools available.

Viewing resynchronizations in vCenter

Resynchronization activity can be found in vCenter in two different ways:

- At the cluster level as an enumerated list of objects currently being resynchronized

- At the host level as time-based resynchronization metrics for IOPS, throughput, and latency

Find the list of objects resynchronizing in the cluster by highlighting the cluster and clicking on Monitor → vSAN → Resyncing Objects.

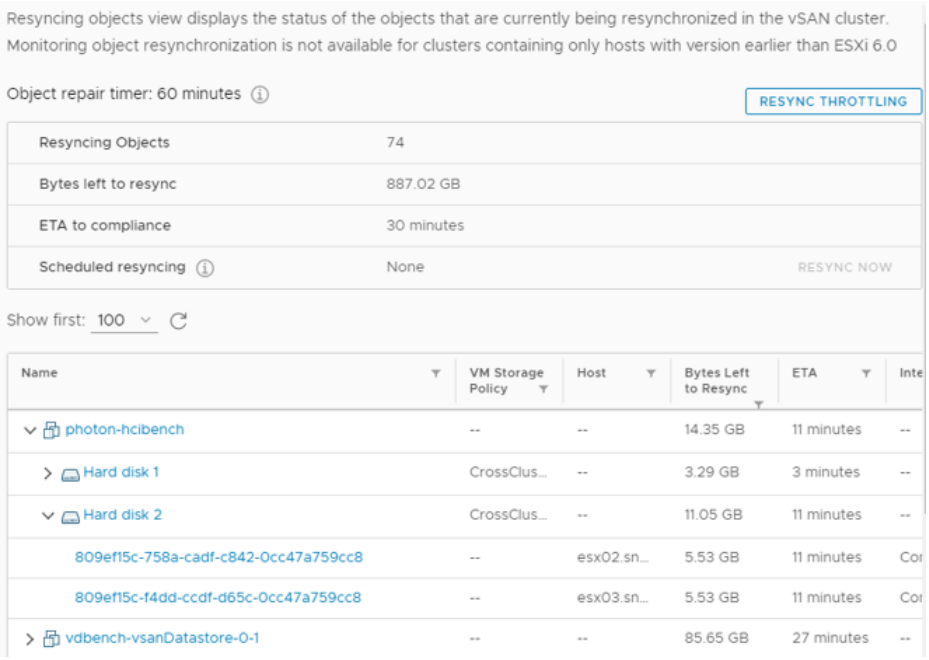


Figure. Viewing the status of resynchronization activity at the cluster level

Find time-based resynchronization metrics by highlighting the desired host and clicking on Monitor → vSAN → Performance → Backend.

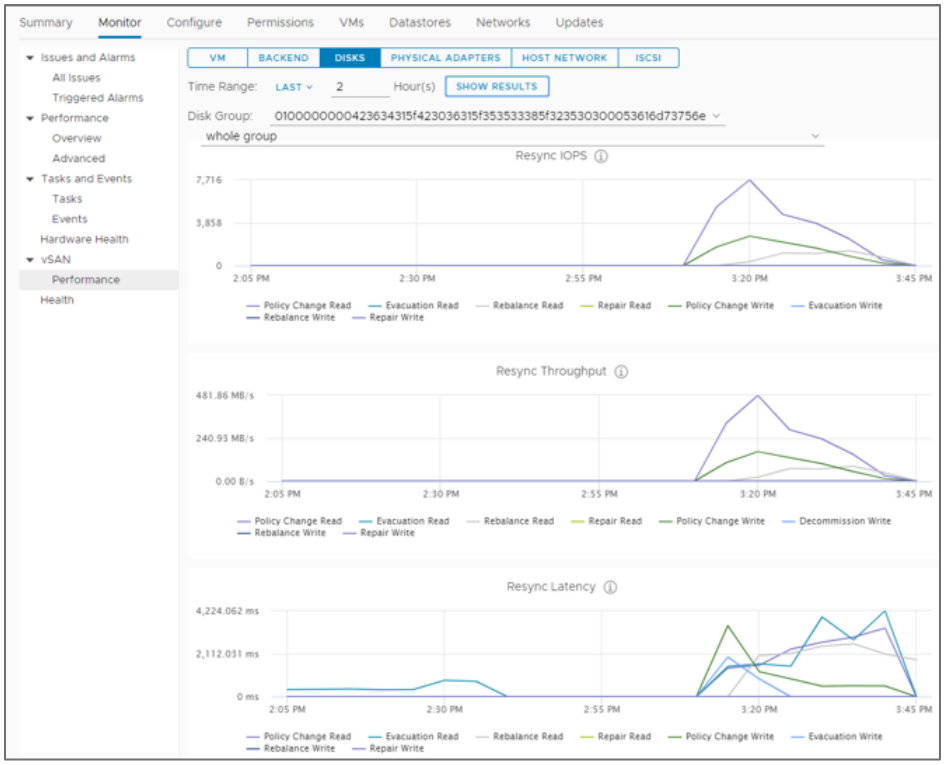


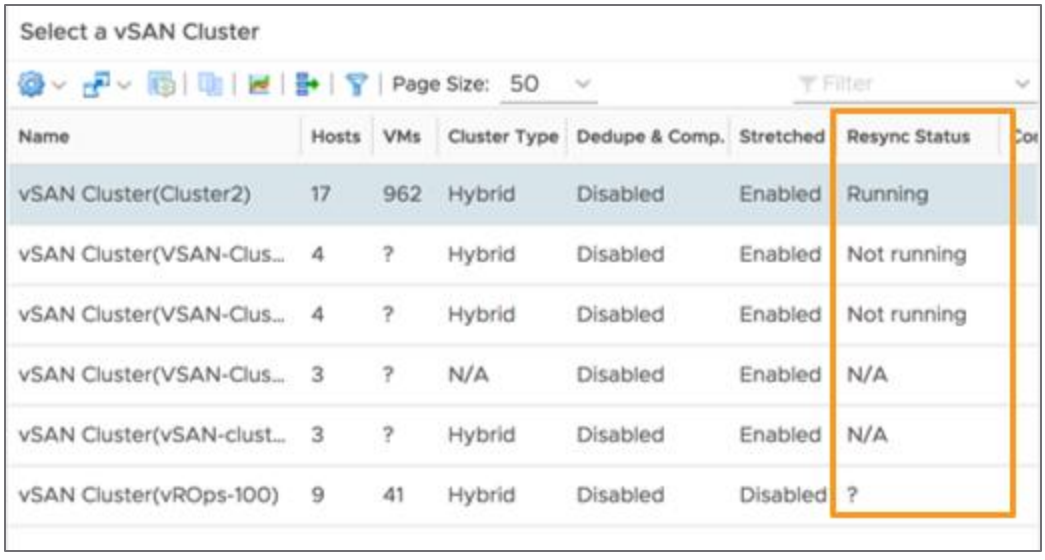
Figure. A breakdown of resynchronization types found in the host-level view of the vSAN performance metrics

Recommendation: Discrete I/O types can be "unticked" in these time-based graphs. This can provide additional clarity when deciphering the type of I/O activity occurring at a host level.

Viewing resynchronizations in VCF Operations

VCF Operations has visibility for resynchronizations in a vSAN cluster. It can be used to augment the information found in vCenter, as the resynchronization intelligence found in Aria Operations is not readily available within the vSAN performance metrics found in Center.

VCF Operations can provide an easy-to-read resynchronization status indicator for all vSAN clusters managed by the vCenter server.



Name	Hosts	VMs	Cluster Type	Dedupe & Comp.	Stretched	Resync Status
vSAN Cluster(Cluster2)	17	962	Hybrid	Disabled	Enabled	Running
vSAN Cluster(VSAN-Clus...	4	?	Hybrid	Disabled	Enabled	Not running
vSAN Cluster(VSAN-Clus...	4	?	Hybrid	Disabled	Enabled	Not running
vSAN Cluster(VSAN-Clus...	3	?	N/A	Disabled	Enabled	N/A
vSAN Cluster(vSAN-clust...	3	?	Hybrid	Disabled	Enabled	N/A
vSAN Cluster(vROps-100)	9	41	Hybrid	Disabled	Disabled	?

Figure. Resynchronization status of multiple vSAN clusters

VCF Operations provides burn down rates for resynchronization activity over time. Measuring a burn down rate helps provide the context in a way that can be difficult to understand using simple resynchronization throughput statistics. A burn down graph for resynchronization activity provides an understanding of the extent of data queued for resynchronization, how far along the process is, and a trajectory toward completion. Most importantly, it measures this at the cluster level, eliminating the need to gather this data per host to determine the activity across the entire cluster.

VCF Operations renders resynchronization activity in one of two ways:

- Total objects left to resynchronize
- Total bytes left to resynchronize

A good example of this is illustrated in a simple dashboard, where several VMs had their storage policy changed from using RAID-1 mirroring to RAID-5 erasure coding.

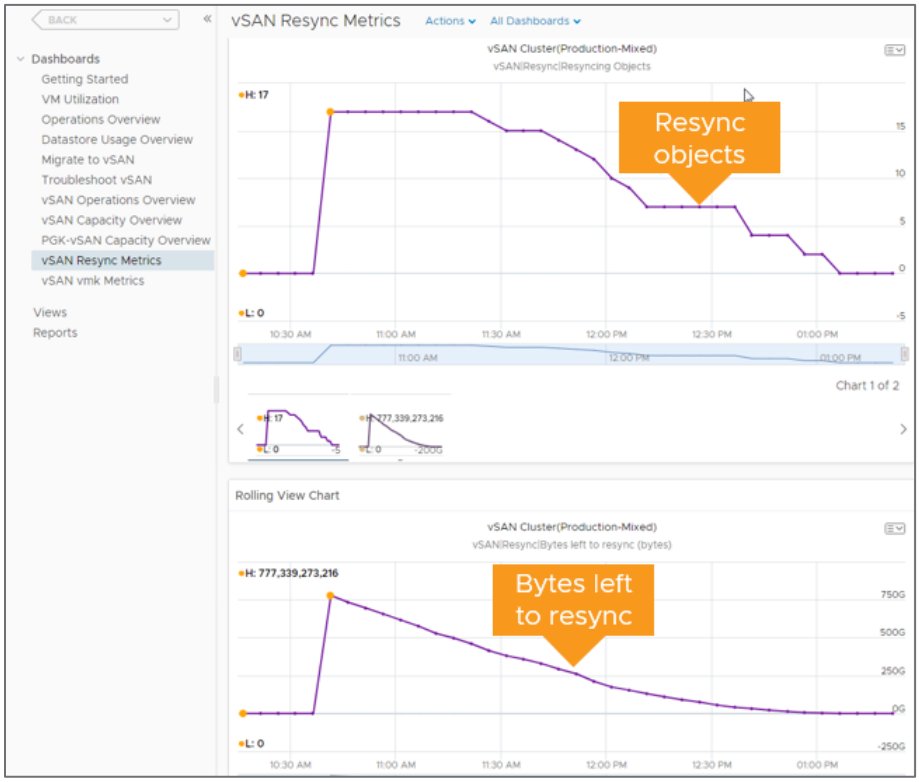


Figure. Resynchronization burn down rates for objects, and bytes remaining

When paired, the “objects remaining” and “bytes left” can help us understand the correlation between the number of objects to be resynchronized, and the rate at which the data is being synchronized. Observing rates of completion using these burn down graphs helps better understand how Adaptive Resync in vSAN dynamically manages resynchronization rates during periods of contention with VM traffic. These charts are easily combined with VM latency graphs to see how vSAN helps prioritize different types of traffic under these periods of contention.

Burn down graphs can provide insight when comparing resynchronization activities at other times, or in other clusters. For example, the figure below shows burn down activity over a larger time window. We can see that the amount of activity was very different during the periods that resynchronizations occurred.

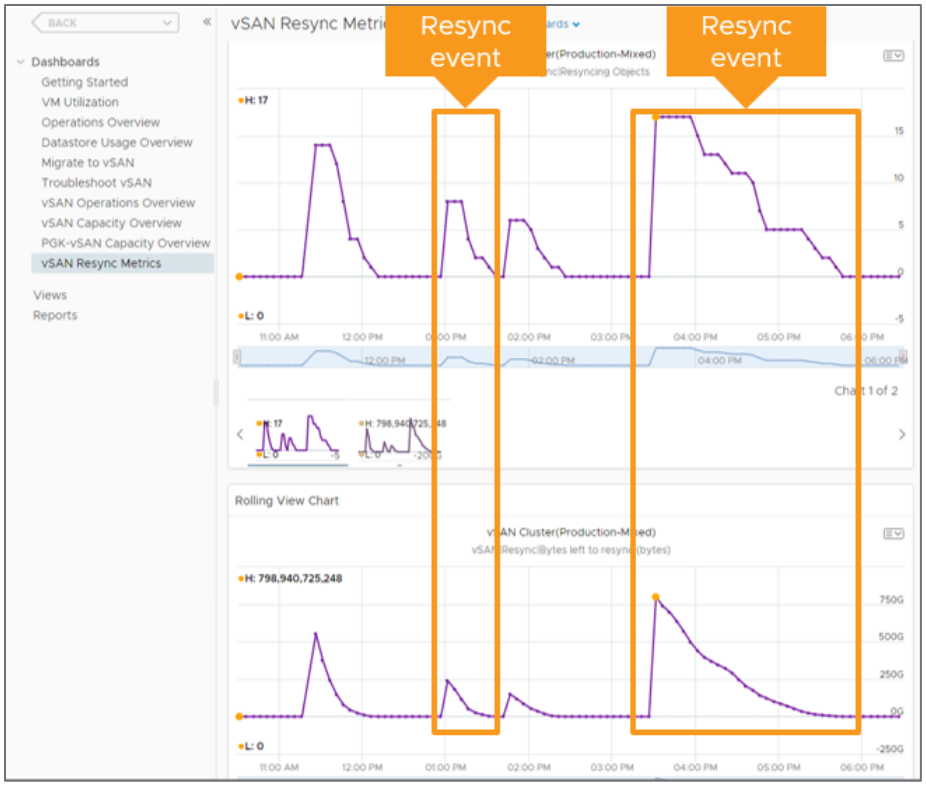


Figure. Comparing resynchronization activity—viewing burn down rates across a larger time window

The two events highlighted represent a different quantity of VMs that had their policies changed. This is the reason for the overall difference in the amount of data synchronized.

Viewing resynchronizations in PowerCLI

Resynchronization information can be gathered at the cluster level using the following PowerCLI command:

```
Get-VsanResyncingComponent -Cluster (Get-Cluster -Name "Clustername")
```

Additional information will be shown with the following:

```
Get-VsanResyncingComponent -Cluster (Get-Cluster -Name "Clustername") |fl
```

See the "PowerCLI Cookbook for vSAN" for more PowerCLI commands and how to expose resynchronization data.

Network Monitoring of vSAN Powered Cluster

Understanding the health and performance of a network is an important part of ensuring a hyper converged platform like vSAN is running at its very best. A distributed storage system like vSAN depends heavily on the network that connects the hosts, as it is the hosts in the cluster that makes up the storage system. Network interruptions can generate packet loss in ways that even at relatively low levels, can degrade the effective throughput of communication required for transmission of storage.

vSAN has several network related metrics to monitor. In addition to the new and existing network metrics listed above, there are additional metrics found in the "Performance for Support" section of the vCenter Server UI. These are metrics typically intended for GSS cases, but still visible to the administrator.

Recommendation: Use some tools for visibility into the operation of the network switches. Network switch configurations and performance are outside of the management domain of vSphere, but should be monitored with equal levels of importance.

Mitigating Network Connectivity Issues in a vSAN cluster

As with any type of distributed storage system, VMware vSAN is highly dependent on the network to provide reliable and consistent communication between hosts in a cluster. When network communication suffers, impacts may not only be seen in the expected performance of the VMs in the cluster but also with vSAN's automated mechanisms that ensure data remains available and resilient in a timely manner.

In this type of topology, issues unrelated to vSAN can lead to the potential of a systemic issue across the cluster because of vSAN's dependence on the network. Examples include improper firmware or drivers for the network cards used on the hosts throughout the cluster, or perhaps configuration changes in the switchgear that are not ideal. Leading indicators of such issues include:

- **Much higher storage latency than previously experienced.** This would generally be viewed at the cluster level, by highlighting the cluster, clicking Monitor > vSAN > Performance and observing the latency.
- **Noticeably high levels of network packet loss.** Degradations in storage performance may be related to increased levels of packet loss occurring on the network used by vSAN. Recent editions of vSAN have enhanced levels of network monitoring and can be viewed by highlighting a host, clicking on vSAN > Performance > Physical Adapters, and looking at the relevant packet loss and drop rates.

Remediation of such issues may require care to minimize potential disruption and expedite the correction. When the above conditions are observed, **VMware recommends holding off on any corrective actions such as host restarts and reaching out to VMware Global Support Services (GS) for further assistance.**

Summary

With the proper guidance, operational tasks and other monitoring activities related to vSAN can be easily incorporated into existing data centers. This operational guidance, paired with Broadcom's official vSAN documentation is a great way to deploy and operate vSAN efficiently.

Additional Resources

The following are a collection of useful links that relate to bandwidth sizing for vSAN stretched clusters.

Performance Recommendations for vSAN ESA. This is a collection of recommendations to help achieve the highest levels of performance in a vSAN ESA cluster. Many of these same recommendations apply to vSAN storage clusters.

Design and Sizing for vSAN ESA clusters. This post offers some nice guidance on using the vSAN Sizer for the ESA that summarizes some key points that can be found in the VMware vSAN Design Guide.

vSAN Network Design Guide. This network design guide applies to environments running vSAN 8 and later.

vSAN technical blogs. Stay up to date on the most recently published technical information about vSAN. These posts are created by the vSAN Technical Marketing team.

VMware Resource Center. The location for design guides, operations guides and other technical white papers on vSAN. These assets are created by the vSAN Technical Marketing and Product Enablement teams.

Official vSAN documentation. The location for all "how to" documentation on vSAN.

About the Author

Pete Koehler is a Product Marketing Engineer in the VCF division at Broadcom. With a primary focus on vSAN, Pete covers topics such as design and sizing, operations, performance, troubleshooting, and integration with other products and platforms.