vSAN Operations Guide

First Published On: 07-04-2016
Last Updated On: 10-12-2018

Table of Contents

1. vSAN Basics
    1.1. vSAN Basics
2. vSAN Cluster Operations
    2.1. Creating a vSAN Cluster
    2.2. Disabling a vSAN Cluster
    2.3. Powering Down a vSAN Cluster
    2.4. Adding Hosts (Scaling Out)
    2.5. Removing a Host
    2.6. Compute Only Hosts
    2.7. Migrating Hybrid to All-Flash vSAN
    2.8. Configuring Fault Domains
    2.9. Enabling Deduplication and Compression
3. Network Operations
    3.1. Network Operations
    3.2. Creating a vSwitch
    3.3. Creating a vDS
    3.4. Creating a vSAN VMkernel Port Group
    3.5. Creating a NIC Team/Failover Order/LACP
    3.6. Setting up a VLAN on a network switch
    3.7. Shared NIC/Dedicated NIC?
    3.8. Enabling Multicast for vSAN on a Network Switch
    3.9. Creating a Static Route for vSAN Networking
    3.10. Configuring NIOC for vSAN – Bandwidth Allocation
    3.11. Configuring VLANs
    3.12. Configuring Multicast
    3.13. Configuring Jumbo Frames
    3.14. Migrating from vSS to vDS
4. Disk Operations
    4.1. Disk Operations
    4.2. Creating a Disk Group (Hybrid/All-Flash)
    4.3. Removing a Disk Group
    4.4. Removing a Cache Disk (Failure) from a Disk Group
    4.5. Adding a Capacity Tier Device to a Disk Group
    4.6. Mark a Disk as Local/Remote
    4.7. Mark a Disk as Flash or HDD
    4.8. Removing a Capacity Disk
    4.9. Balance the Disk Usage
    4.10. Removing a Partition From a Disk
    4.11. Blinking a Disk LED
5. Datastore Operations
    5.1. Datastore Operations
    5.2. Browsing vSAN Datastore Contents
    5.3. Uploading files to vSAN Datastore
    5.4. Maintaining Sufficient Slack (Free) Space
6. VM Storage Policies Operations
    6.1. VM Storage Policies Operations
    6.2. Creating a Policy
    6.3. Editing a Policy
    6.4. Deleting a Policy
    6.5. Applying a Policy
    6.6. Changing a Policy On-the-Fly (What Happens)
    6.7. Bulk Assign Storage Policies to Multiple VMs
    6.8. Checking Compliance Status
    6.9. Backing up Policies
    6.10. Restoring Policies
    6.11. Storage Policy recommendations for VMs in stretched clusters
    6.12. Storage Policy Naming Considerations
7. Maintenance Mode Operations
    7.1. Enter Maintenance Mode

7.2. Set Default Maintenance Mode Operation
8. Host Operations
   8.1. Patching and Updates of Hosts
   8.2. Configuring Log Locations
   8.3. Improving Visibility of Host Restarts
9. vCenter Operations
   9.1. vCenter Operations
   9.2. Updating vCenter in a vSAN Cluster
   9.3. Certificates
   9.4. Moving a vSAN Cluster
   9.5. Replacing a vCenter Server for existing vSAN hosts
10. Compression and Deduplication Operations
   10.1. Compression and Deduplication
   10.2. Enabling Dedup/Compression on a New Cluster
   10.3. Enabling Dedup/Compression on an Existing Cluster
   10.4. Disabling Dedupe/Compression
   10.5. Monitoring Progress of Enabling/Disabling
   10.6. Allow Reduced Redundancy
   10.7. Adding a capacity Tier Disk
   10.8. Removing a Cache Disk
   10.9. Removing a Capacity Disk From a Disk Group
   10.10. Failure Considerations for Cache Disk
   10.11. Failure Considerations for Capacity Disks
11. Checksum Operations
   11.1. Checksum Operations
   11.2. Defining a VM Storage Policy for Checksum
   11.3. Applying Policy with a VM Storage Policy
   11.4. Manually Disabling Checksum on a VM or Object
   11.5. Enabling Checksum on a VM or Object
12. Performance Service Operations
   12.1. Performance Service Operations
   12.2. Enable Performance Service
   12.3. Disable Performance Service
   12.4. Change policy on Performance Service
13. Stretched Cluster Operations
   13.1. Stretched Cluster Operations
   13.2. Deploying a Witness Appliance
   13.3. Configuring a Stretched Cluster
   13.4. Replacing a Witness Appliance
   13.5. DRS Settings
   13.6. HA Settings
   13.7. Affinity Rules
   13.8. Decommissioning a Stretched Cluster
14. Upgrading vSAN
   14.1. Upgrading vSAN
15. Monitoring vSAN
   15.1. Monitoring vSAN Cluster Heath
   15.2. Monitoring vSAN Datastore Capacity
   15.3. Monitoring Disk Capacity
   15.4. Monitoring Dedupe/Compression
   15.5. Monitoring Checksum
   15.6. Monitoring vSAN with the Performance Service
   15.7. Monitoring Resync Activity
   15.8. Configure Alarms/Traps/Emails
16. vRealize Operations Manager
   16.1. vRealize Operations Manager
   16.2. Deploy vRealize Operations Manager
   16.3. Configure vROps to Monitor vSphere
   16.4. Install the Management Pack for Storage Devices
   16.5. Configure the MPSD Adapter Instance
   16.6. Integrating vRealize Log Insight with vSAN

16.7.Integration vRLI with vROps for vSAN

1. vSAN Basics

Before we will describe all of the different operational procedures around vSAN 6.x we would like to ensure that everyone has a basic understand of vSAN first.

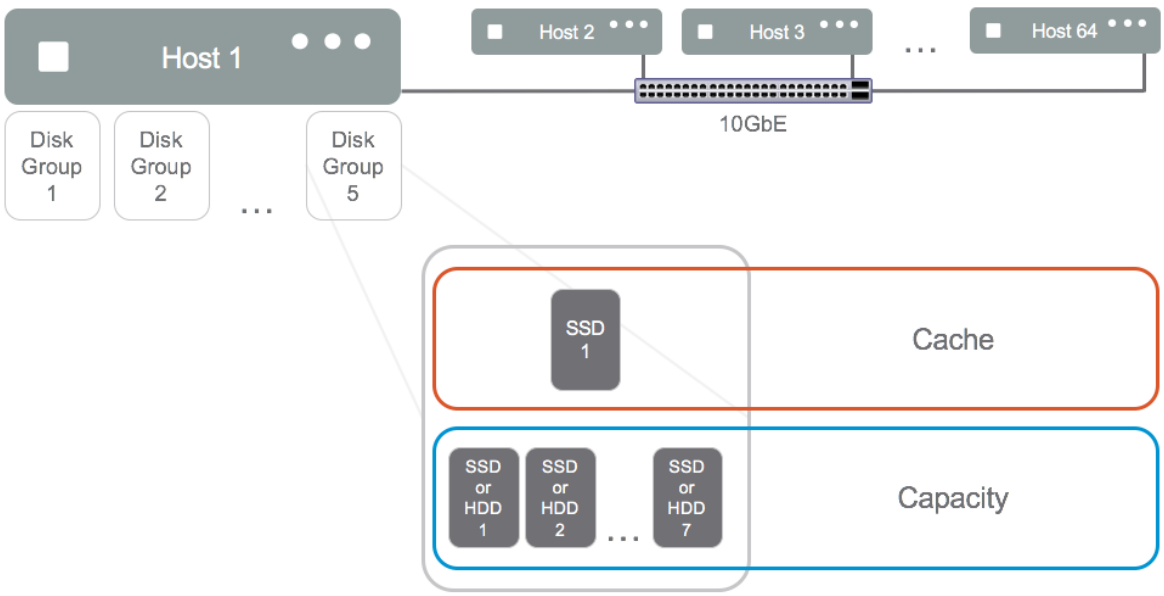
1.1 vSAN Basics

Before we describe operational procedures for vSAN, we would first like to ensure that everyone has a basic understanding of vSAN. If you are already familiar with terms like clusters, disk groups, replicas, objects, and components, you can probably skip this section. However, it is a good refresher even if you are familiar with vSAN.

vSAN clusters contain two or more physical hosts that contain either a combination of magnetic disks and flash devices (hybrid configuration) or all flash devices (all-flash configuration). These devices are used exclusively by vSAN to construct the cache and capacity tiers of a vSAN datastore. There is one vSAN datastore per cluster and it is accessible by all of the vSphere hosts in the vSAN cluster.

While the minimum number of physical hosts is two, a more common starting point is three or four physical hosts. Two hosts often referred to a 2-node configuration are more common in use cases requiring a very small amount of computing and storage resources such as a remote or branch office. The maximum number of hosts in a vSAN cluster is 64.

Each host in a vSAN cluster usually has one or more vSAN disk groups. A disk group consists of exactly one cache device, which must be a flash device and one to seven capacity devices which can be magnetic drives or flash devices. A disk group can have up to seven capacity devices. The diagram below provides a high-level look at vSAN architecture.

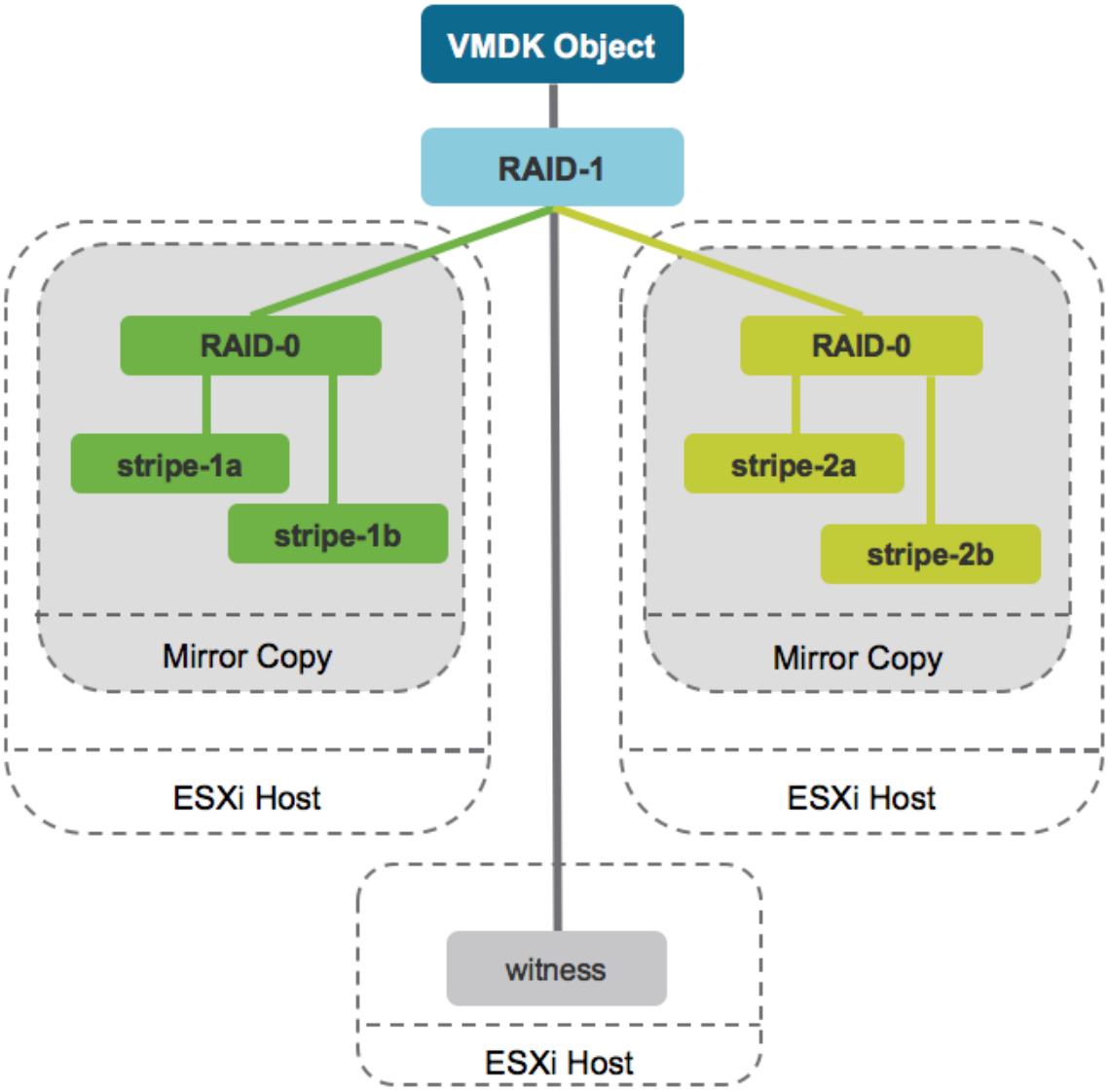


All-flash configurations are the most common type of clusters deployed. In an all-flash configuration, the flash devices in the cache tier are used for write buffering only (no read cache). Read performance directly from capacity flash devices is more than sufficient. Two different grades of flash devices are commonly used in an all-flash vSAN configuration: Lower capacity, higher endurance devices for the cache tier and more cost-effective, higher capacity, lower endurance devices for the capacity layer. Writes are performed at the cache tier and then de-staged to the capacity tier as needed. This helps extend the usable life of the lower endurance flash devices in the capacity tier.

In a hybrid configuration, one flash device is required for each disk group, which is part of the cache tier. One or more magnetic drives are part of each disk group and these

drives make up the capacity tier. A disk group can have up to seven capacity devices. A hybrid vSAN configuration uses 30% of the cache tier for write buffering and 70% of the cache tier for read caching.

vSAN is an object-based storage platform. This means that VMs (and their virtual disks) are stored as objects on a vSAN datastore. A storage policy is assigned to each object. The components that make up each object are distributed across disks in various hosts in the vSAN cluster based on the rules defined in the storage policy. Below is an example of a 400GB virtual disk (VMDK) that has the Default vSAN Storage Policy assigned. This policy contains the rules Primary Level of Failures to Tolerate = 1 and Fault Tolerance Method = RAID-1 mirroring. Mirrored copies or "replicas" of the data are placed on separate hosts so that the loss of one host can be tolerated. Since the maximum component size is 255GB, the object on each host is split into two components or "stripes" that are 200GB in size. A witness component is placed on a third host to serve as a tie-breaker and preserve data integrity if a "split-brain" scenario occurs.



More details on the concepts above are available in other documents here on StorageHub and in the vSAN documentation. We could certainly spend more time on those and other topics here, but the focus of this document is operations and management, which is up next.

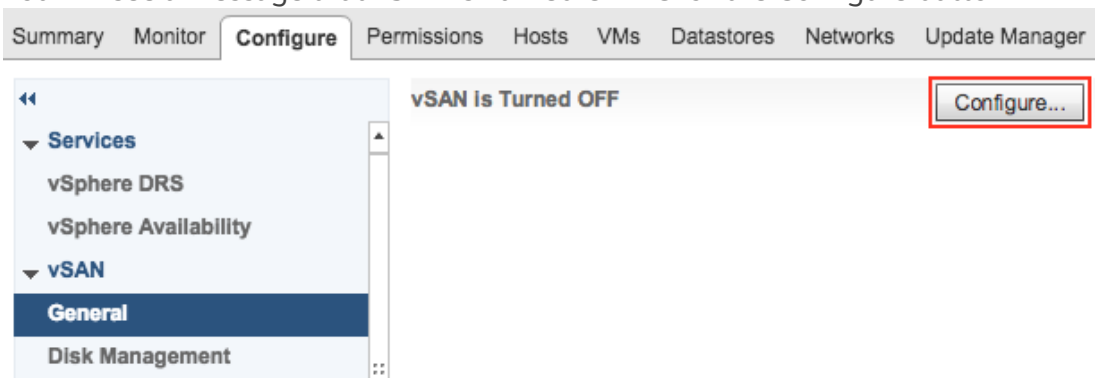
2. vSAN Cluster Operations

In this section we will describe the various operational procedures which occur on a vSAN cluster level.

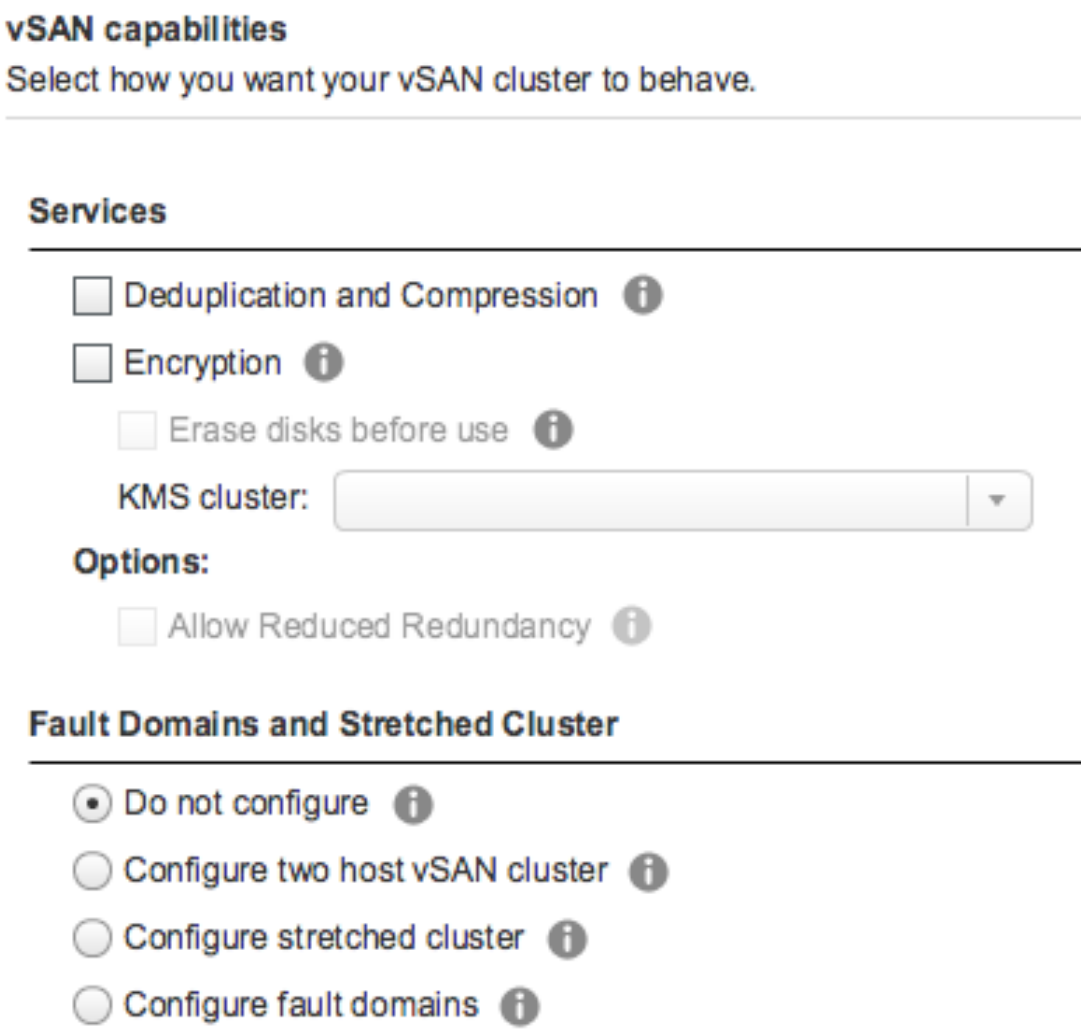
2.1 Creating a vSAN Cluster

The creation of a vSAN cluster is simple. In the scenario below we already have hosts in the cluster (HA/DRS), and we are going to turn on vSAN. To enable vSAN on an existing cluster, follow the click-through demo, vSAN 6.5 - Turning on vSAN or the procedure below:

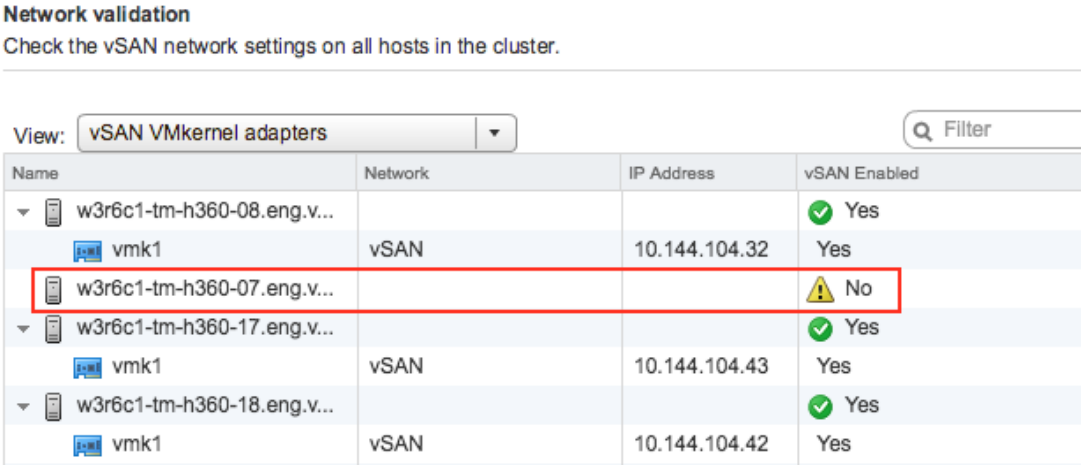
1. Open the vSphere Web Client.
2. Click Hosts and Clusters.
3. Select the cluster on which you want to enable vSAN.
4. Click the Configure tab and select General under vSAN.
5. You will see a message that vSAN is Turned OFF. Click the Configure button.



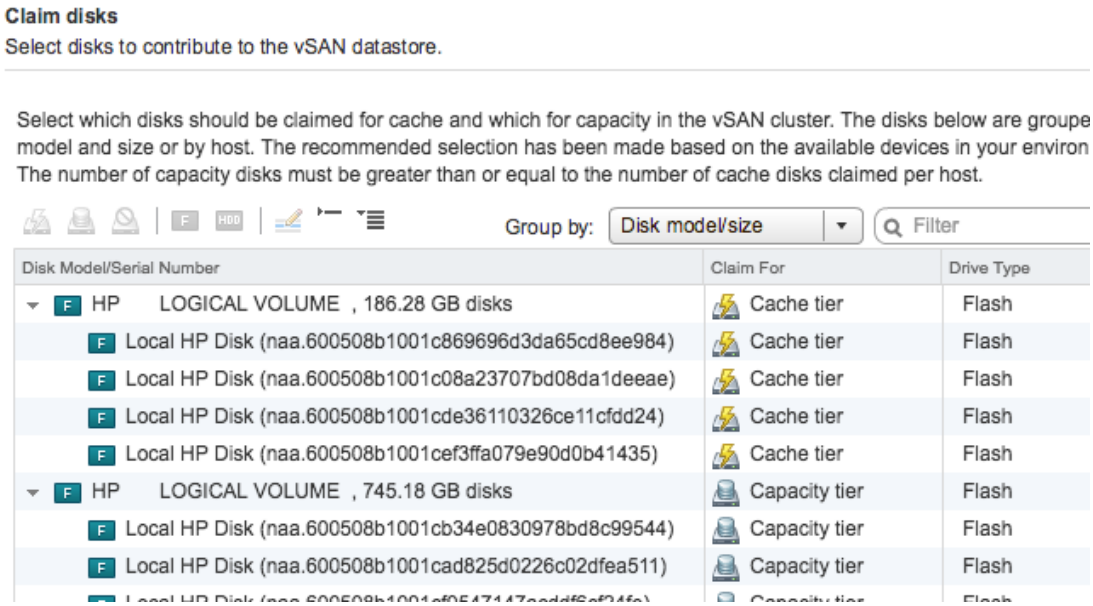
6. Select the vSAN capabilities you want to enable such as Deduplication and Compression. **Recommendation** : Enable services such as Encryption, Deduplication, and Compression before placing virtual machines on the vSAN datastore. While these services can be enabled at a later time, this process requires a rolling reformat of every disk group in the cluster. This operation can take a considerable amount of time and it reduces the usable capacity of the vSAN datastore until the operation is complete.
7. Optionally, you can enable a 2-node configuration, a stretched cluster configuration, or fault domains by selecting one of these options in the vSAN Capabilities window.



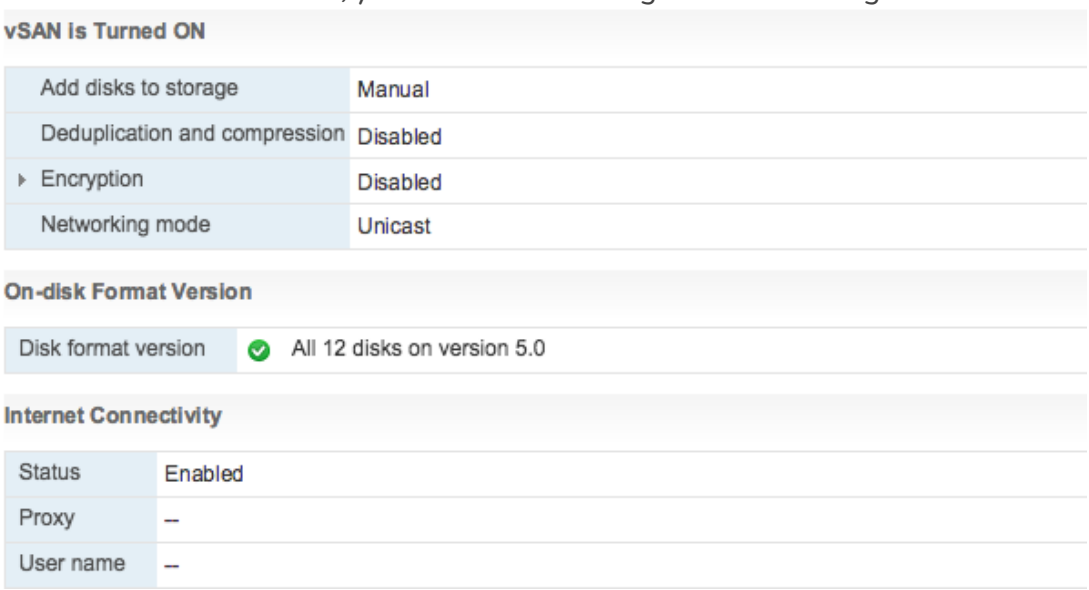
8. The Network Validation step in this process simply verifies that each host has a VMkernel adapter with the vSAN service enabled. The figure below shows a host that does not have the vSAN service enabled on a VMkernel adapter. This issue should be resolved before proceeding.



9. Select the storage devices in each host that will be used for the cache and capacity tiers.



10. Verify the configuration and click Finish. The process of turning on vSAN can take some time. When finished, you will see something similar to the figure below.



2.2 Disabling a vSAN Cluster

When you disable a vSAN cluster, all the virtual machines on the shard vSAN datastore become inaccessible. If you need to use this VMs while vSAN is disable, migrate them to another datastore.

To disable vSAN on an existing cluster:

1. Open the vSphere Web Client.
2. (Optional) Migrate all VMs off the cluster.
3. Select **Hosts and Clusters**.
4. Select the *cluster* on which you want to disable vSAN.
5. Click the **Configure** tab.
6. Under vSAN, click **General**.

7. Click **Edit** at the top where it says "vSAN is turned On".

8. Uncheck "Turn on vSAN".
9. Click **OK**.
10. Read the warning, "If you turn off vSAN, virtual machines on the vSAN datastore become inaccessible", and click **OK** when understanding the impact of turning off vSAN.

Click to see topic media

For more details see KB 2058322.

2.3 Powering Down a vSAN Cluster

The following steps describe how to power down a vSAN Cluster.

1. Power down all Virtual Machines that are running on the vSAN Cluster except vCenter Server.
2. Verify that no vSAN components are currently resyncing.
    - Using vSphere Web Client, navigate to the vSAN Cluster.
    - Select the **Monitor** tab and click **vSAN**.

    - Select **Resyncing Components** to determine if any resync operations are in progress. If any are, wait until they are completed to proceed.
3. If the vCenter Server is running on the vSAN cluster
    - Migrate the vCenter Server to the first host.
    - Shutdown the vCenter Server. The vSphere Web Client will no longer be accessible.
4. Place all ESXi hosts into Maintenance Mode. You must perform this operation through one of the CLI methods that supports setting the vSAN mode when entering Maintenance Mode. You can either do this by logging directly into the ESXi Shell and running ESXCLI locally or you can invoke this operation on a remote system using ESXCLI.

    - `esxcli system maintenanceMode set -e true -m noAction`

5. Once the host enters maintenance mode, shutdown all ESXi hosts using either the vSphere C# Client, ESXi Shell, SSH or the Host Client.

For more information see KB 2142676.

2.4 Adding Hosts (Scaling Out)

vSAN allows you to both scale up (add resources to existing hosts) and scale out (add hosts). Follow the procedure in vSphere 6.5 Add Hosts to the vSAN Cluster section, or follow the click-through demo vSAN 6.5 Scale Out by Adding a Host.

2.5 Removing a Host

Using vSAN Health, validate you have a fully functioning vSAN cluster before removing hosts. Also, verify there will be a sufficient number of hosts and sufficient storage capacity after a host is removed. The following steps are recommended:

1. Click **Hosts and Clusters**.
2. Select the *host* that needs to be removed, right-click and select **Maintenance Mode>Enter Maintenance Mode**.
3. Select the *Full Data Migration* option. This ensures that VMs remain compliant with their assigned storage policy after removing the host.
4. Click **OK**.
5. When the host enters Maintenance Mode, right-click the *host* and select **Move To**.
6. Select the *new location* and click **OK**.

The host is now removed from the cluster. If you want to add it back to the vSAN cluster at some point it is recommended that you remove all partitions first from the disks.

Click to see topic media

2.6 Compute Only Hosts

It is also possible to add hosts which are not contributing storage capacity to the vSAN Datastore to the cluster. The procedure is similar to the Adding Hosts. You will **NOT** add the diskgroups to the datastore in this instance.

NOTE: If you configured vSAN to automatically claim all empty disks, you will need to first switch to manual.

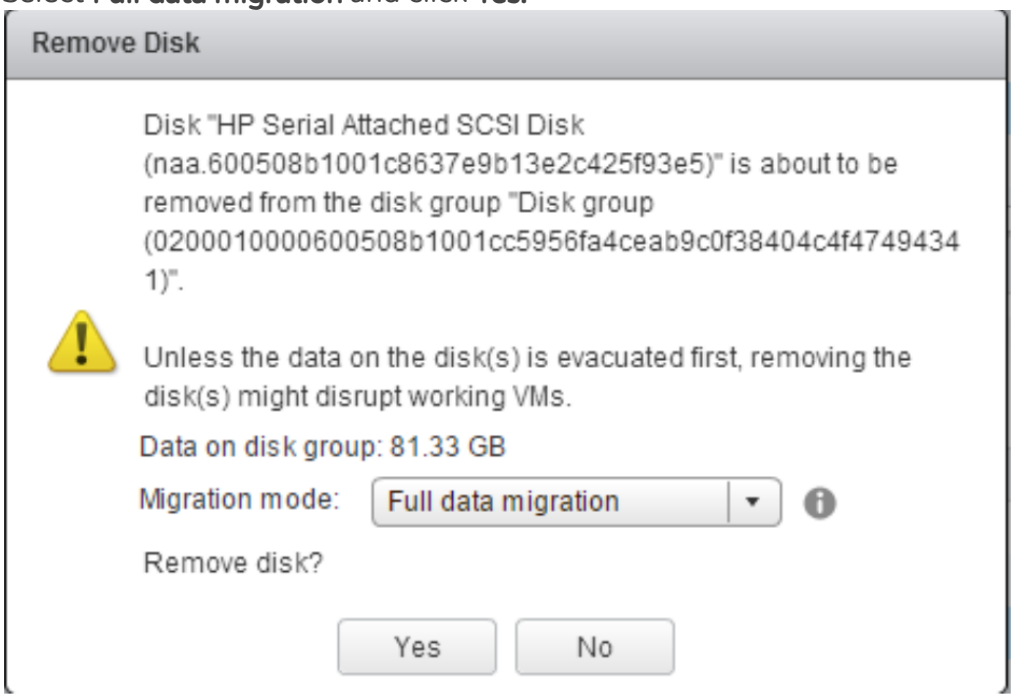
NOTE: It is not recommended to create clusters where a few hosts are providing capacity and others are simply consuming. The reason for this is that from an availability and performance standpoint a broader distributed datastore is more beneficial. The impact of failures are high when only few hosts contribute storage. See Considerations for Compute-Only Hosts.

2.7 Migrating Hybrid to All-Flash vSAN

The following steps describe the procedure of how to migrate from a hybrid vSAN cluster to an all-flash vSAN cluster. **NOTE**: In order to be able to run all-flash at a minimum the "Advanced" license is required at the time of writing.

1. Open the vSphere Web Client.
2. Remove the hybrid disk group:
    - Click the **Hosts and Clusters** tab.
    - Select the *cluster* you want to migrate to all-flash vSAN.
    - Click the **Configure** tab.

- Under vSAN, click **Disk Management**.
- Select the *Disk Group* to remove and click the **Remove Disk Group icon**.
- Select **Full data migration** and click **Yes.**



3. Remove the physical HDDs from the host.
4. Add the flash devices to the host. Ensure there are no partitions on the flash devices.
5. Create the all-flash disk group on each host. Follow the Create a Disk Group procedure.

Repeat above steps for each host in the cluster.

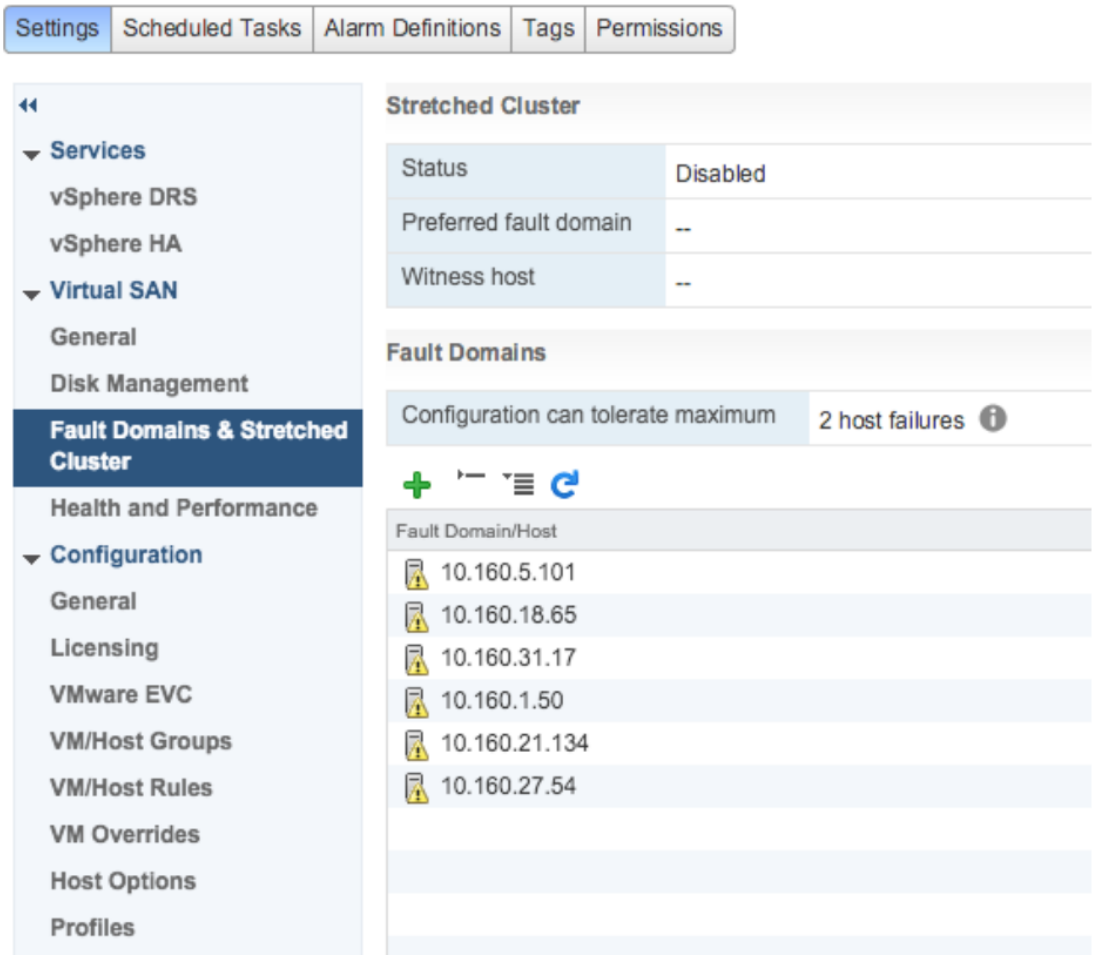
Click to see topic media

2.8 Configuring Fault Domains

To configure fault domains on an existing vSAN cluster:

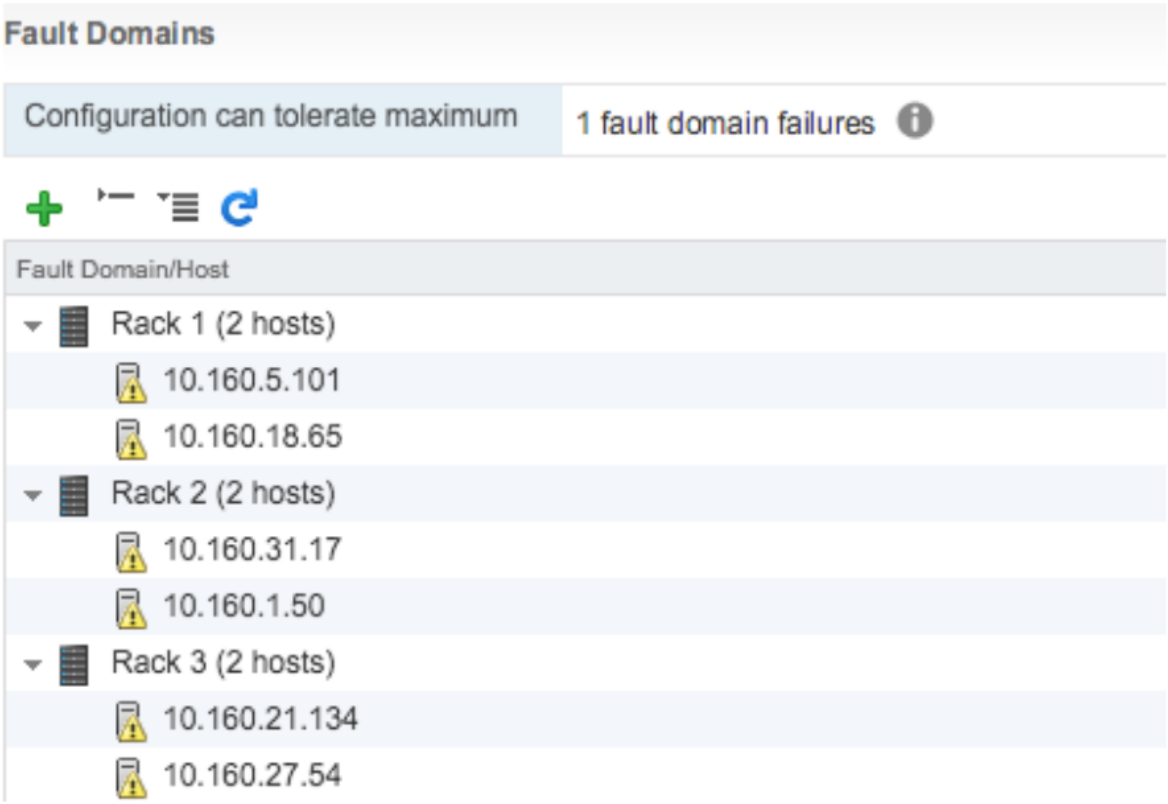
- Open the vSphere Web Client.
- Select **Hosts and Clusters**.
- Select the *cluster* on which you want to configure fault domains for vSAN.
- Click the **Configure** tab.
- Under vSAN, click **Fault Domains & Stretched Cluster**.

- Click the Create a new fault domain icon (+).



- Enter a *name* for the fault domain.
- Select one or more *hosts* for this fault domain.
- Click **OK**.

Go through the above procedure for each Fault Domain you need to create. The outcome should look something similar to the below.



Click to see topic media

2.9 Enabling Deduplication and Compression

Deduplication and compression can be enabled during the creation of a cluster, but it can also be enabled after the creation of a cluster. **NOTE**: this process may take several hours, depending on the size of the datastore.

1. Open the vSphere Web Client.
2. Select the **Hosts and Clusters**.
3. Select the *cluster* you want to enable deduplication and compression on.
4. Click the **Configure** tab.
5. Under vSAN, select **General**.
6. Set the disk claiming mode to **Manual**.
7. Click **Edit** at the top where it says "vSAN is turned On".
8. Select **Enabled** on the Deduplication and compression dropdown
9. Click **OK**.

Now the vSAN Datastore will be reconfigured, this may take several hours, depending on the size of the datastore. vSAN must convert each disk group one at a time. vSAN evacuates data from a disk group, removes the disk group, and recreates it with the new format.

You can monitor the progress on the **Tasks and Events** tab.

Click to see topic media

3. Network Operations

In this section of the vSAN Operations Guide, command network operations pertaining to vSAN are examined.

3.1 Network Operations

In this section of the vSAN Operations Guide, command network operations pertaining to vSAN are examined. Many of these operation are not vSAN specific, as many of them are standard vSphere operations. Where applicable, a link is provided to the appropriate vSphere administration guide.

3.2 Creating a vSwitch

The procedure to create a vSphere standard switch is available in the Create a vSphere Standard Switch. No changes are needed for vSAN.

3.3 Creating a vDS

vSAN licenses entitle a customer to a vSphere Distributed Switch (vDS) irrespective of their vSphere license. The procedure to create a vDS is available in the vSphere 6.5 Networking section. No changes are needed for vSAN.

3.4 Creating a vSAN VMkernel Port Group

The procedure to create a VMkernel port group on a standard vSwitch is available in the vSphere 6.5 Networking section. You must select **vSAN** in the *enable services* section.

The procedure to create a VMkernel port group on a distributed vSwitch is available in the vSphere 6.5 Networking section. You must select **vSAN** in the *enable services* section.

3.5 Creating a NIC Team/Failover Order/LACP

vSAN network traffic has not been designed to load balance across multiple network interfaces when these interfaces are teamed together. While some load balancing may occur when using LACP, NIC teaming can be best thought of as providing a way of making the vSAN traffic network “highly available”. Should one adapter fail, the other adapter will take over the communication. It should not be considered in terms of improved performance (although some improvement in performance may be observed).

The procedure for adding physical NICs, and teaming them, on a vSphere standard switch is available in the vSphere 6.5 Networking section.

If you wish to use LACP on the vSAN network, there is a section describing how in the vSphere 6.5 Networking section.

3.6 Setting up a VLAN on a network switch

Setting up a VLAN on a network switch

to be worked on

3.7 Shared NIC/Dedicated NIC?

For small, hybrid vSAN environments, such as 2-node remote office/branch office (ROBO) deployments and 3-node clusters, 1GbE NICs are supported. However these

NICs must be dedicated to the vSAN network. With larger hybrid environments, 10GbE NICs are recommended. For all-flash environments, 10GbE NICs are required. 10GbE NICs do not have to be completely dedicated to the vSAN network. These NICs may be shared with other traffic types.

3.8 Enabling Multicast for vSAN on a Network Switch

Cisco

(Default is IGMP snooping on).

```
switch# configure terminal
switch(config)# vlan 500
switch(config-vlan)# no ip igmp snooping
switch(config-vlan)# do write memory
```

Brocade ICX

(Default is IGMP snooping off)

```
Switch# configure
Switch(config)# VLAN 500
Switch(config-vlan-500)# multicast disable-igmp-snoop
Switch(config-vlan-500)# do write memory
```

HP ProCurve

(Default is IGMP snooping on).

```
switch#    **configure terminal
switch(config)# VLAN 500 ip IGMP
switch(config)# no VLAN 500 ip IGMP querier
switch(config)# write memory
```

3.9 Creating a Static Route for vSAN Networking

There are some vSAN use cases where static routes may be required. This is because in the current release of vSphere, there can only be a single default gateway, so all routed traffic will try to reach its destination by this gateway by default.

Examples where routed traffic is needed are 2-node (ROBO) deployments where the witness is on a different network, and stretched cluster, where both the data sites and the witness host are on different sites.

There is no way to create a static route via the vSphere Web Client. These must be created via the command line (ESXCLI). Here is an example of such a command.

```
esxcli network ip route ipv4 add –n <remote-network> -g <gateway-to-use>
```

<remote-network> refers to the remote network that this host wishes to have a path to

<gateway-to-use> refers to the interface to use when traffic is sent to the remote network

3.10 Configuring NIOC for vSAN – Bandwidth Allocation

As mentioned earlier, customers who purchase vSAN are automatically entitled to DVS, distributed switches, irrespective of the version of vSphere that they use. Included with the DVS is a Quality of Service (QoS) feature called Network I/O Control (NIOC). This allows administrators to set bandwidth limits on certain traffic types.

This is useful when customers are using a 10GbE NIC for vSAN traffic other traffic types, especially vMotion. vMotion is notorious for consuming as much bandwidth as possible to complete a migration as possible. This may impact vSAN traffic if there is a lot of network activity when the vMotion is initiated. To avoid vMotion traffic impacting vSAN traffic on a shared NIC, NIOC can be used to set a bandwidth allocation. For example, you may like to set the bandwidth allocation for vMotion to 4Gb/s of the 10Gb/s available.

Steps on how to set NIOC bandwidth allocation on different network traffic types can be found in the vSphere 6.5 Networking section.

3.11 Configuring VLANs

Customers can use VLANs to isolate network traffic. This also applies to vSAN traffic. Details on how to use VLANs for network isolation is in vSphere 6.5 Networking section.

3.12 Configuring Multicast

Multicast is a requirement for the vSAN network. There are no server side configuration steps necessary to implement multicast. The configuration steps are all done on the physical switch. VMware recommends the use of IGMP (Internet Group Management Protocol) so that multicast frames are only sent to members of the same group. The group refers to the set of physical switch ports to which the uplinks carrying the vSAN traffic are connected. This avoids sending these multicast frames to every port. Details on how to configure IGMP and multicast are covered earlier in this operations guide.

3.13 Configuring Jumbo Frames

vSAN supports MTU sizes greater than 1500, more commonly referred to as jumbo frames. If jumbo frames are used across your network infrastructure, it can lead to reduced CPU cycles on the ESXi hosts for managing network traffic. Jumbo frames need to be configured in a number of different places, such as the physical switch ports and the virtual switch.

To change the MTU size of a vSS the procedure is in the vSphere 6.5 Networking section.

To change the MTU size of a vDS, the procedure is in the vSphere 6.5 Networking section.

3.14 Migrating from vSS to vDS

Before we begin, this procedure is rather complicated, and can easily go wrong. The only real reason why one would want to migrate from vSS (standard vSwitches) to a vDS (Distributed vSwitch) is to make use of the Network I/O Control feature that is only available with vDS. This will then allow you to place bandwidth allocation QoS (Quality of Service) on the various traffic types such as vSAN traffic.

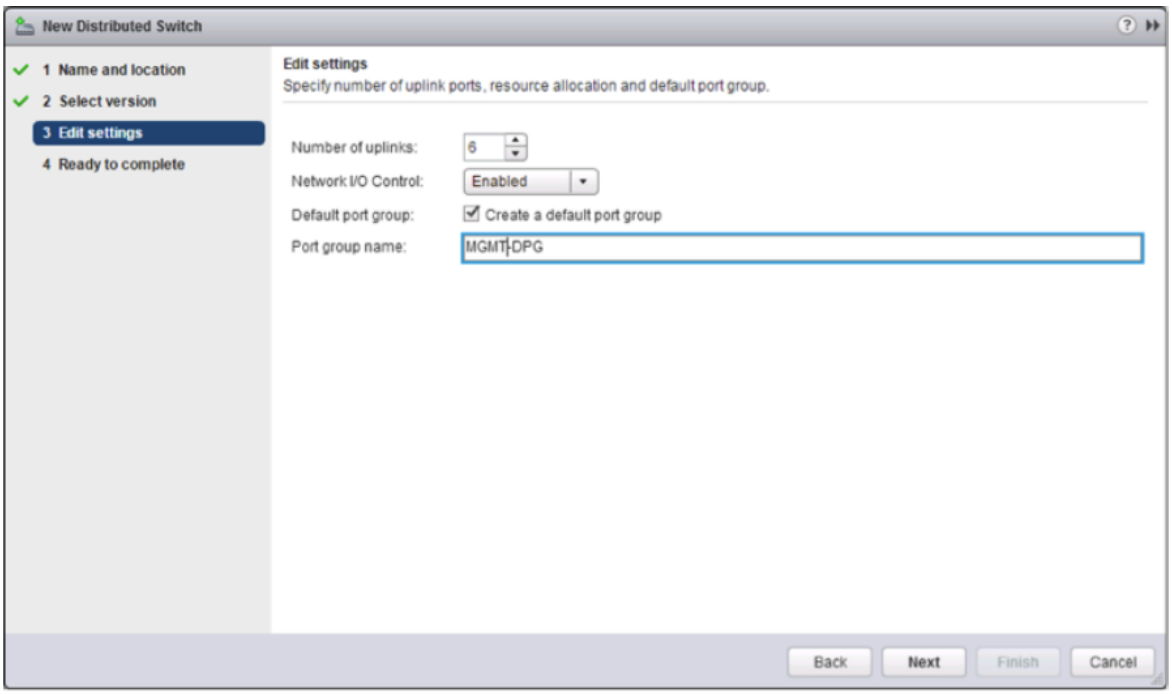
NOTE: Please ensure that you have console access to the ESXi hosts during this procedure. If everything goes well, you will not need it. However, should something go wrong, you may need to access the console of the ESXi hosts.

Create vSphere Distributed Switch

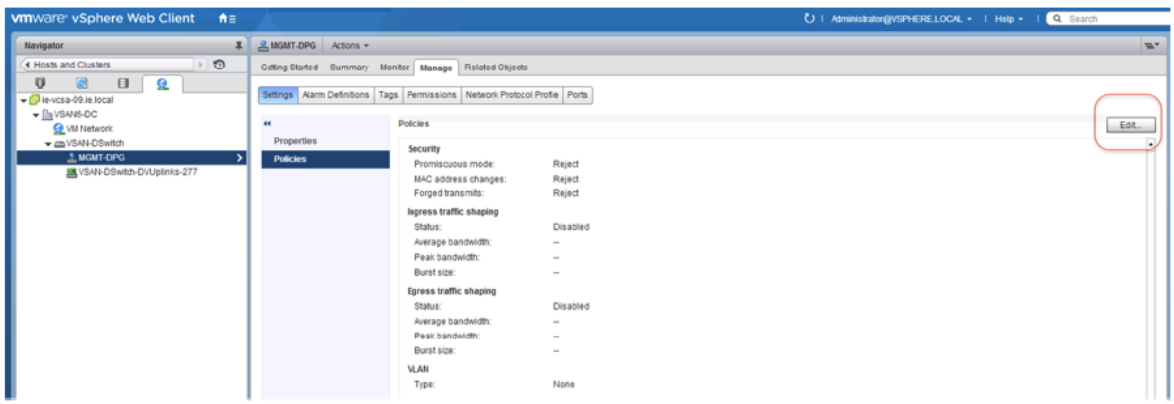
To begin with, create the distributed switch. This is a relatively straight forward procedure which can be found in the Create a vSphere Distributed Switch

When you create a new distributed switch you will be prompted first to provide a name for the new distributed switch. Next, select the version of the vDS, for example, 6.5.0. At this point, we get to add the settings. First, you will need to determine how many uplinks you are currently using for networking. Let's assume for example that we are using six; one for management, one for vMotion,

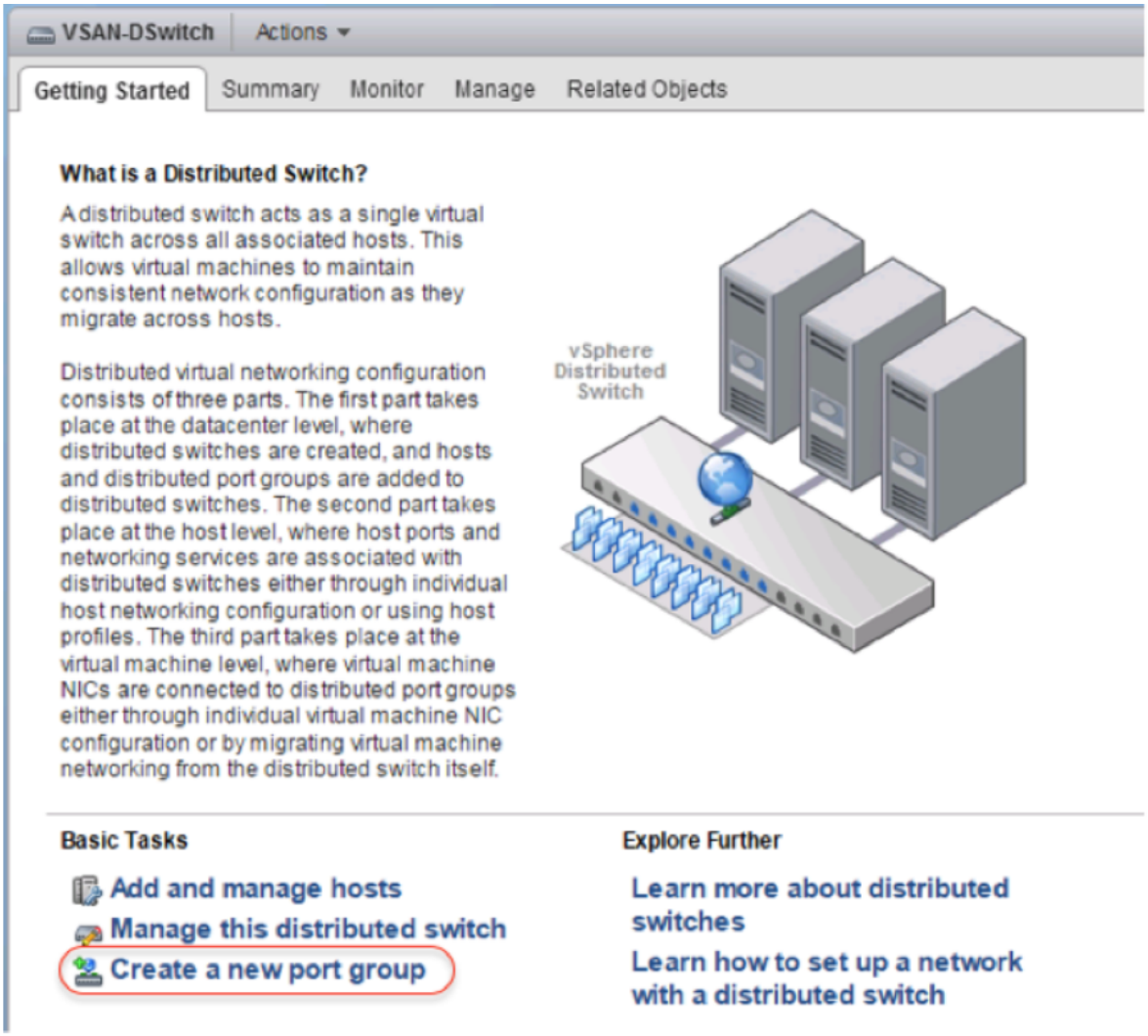
one for virtual machines and three for vSAN. Therefore, when we are prompted for the number of uplinks, we select “6”. This may differ in your environment but you can always edit it later on. Another point to note here is that a default portgroup can be created. You can certainly create a port group at this point, such as a portgroup for the management network shown below, but there will be additional port groups that need to be created shortly. At this point, the distributed switch can be completed.



As alluded to earlier, lets now configure and create the additional port groups. So far, a single default port group has been created for the management network. There was little in the way of configuration that could be done at that time. It is now important to edit this port group to make sure it has all the characteristics of the management port group on the vSS, such as VLAN and NIC teaming and failover settings. Select the distributed port group, and click on the Edit button if it is necessary to change the VLAN and to tag the distributed port group accordingly.



Once the management distributed port group taken care of, you will also need to create distributed port groups for vMotion, virtual machine networking and of course vSAN networking. In the “Getting Started” tab of the distributed switch, there is a basic task link called “Create a new port group”.

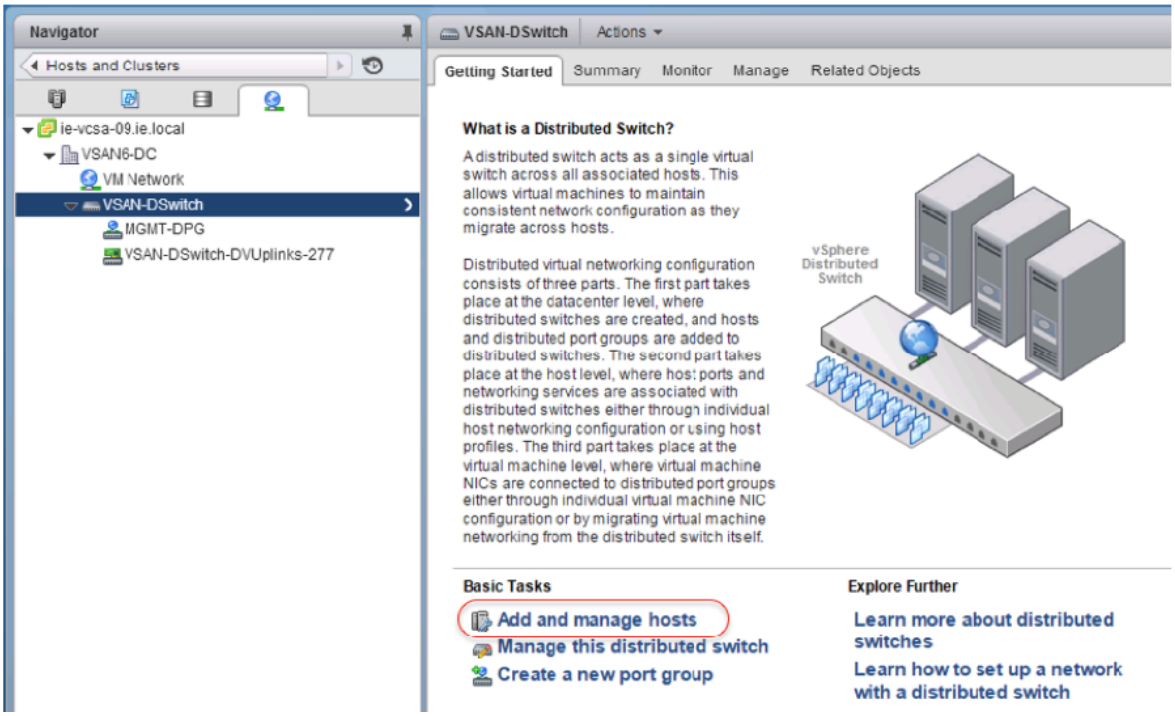


We shall now create a port group for the vMotion network. Again, you will need to provide a name for the new distributed port group, configure distributed port group settings, such as VLAN, then click Finish to complete creating the new distributed port group. Once all the distributed port groups are created on the distributed switch, the uplinks, VMkernel networking and virtual machine networking can be migrated to the distributed switch and associated distributed port groups.

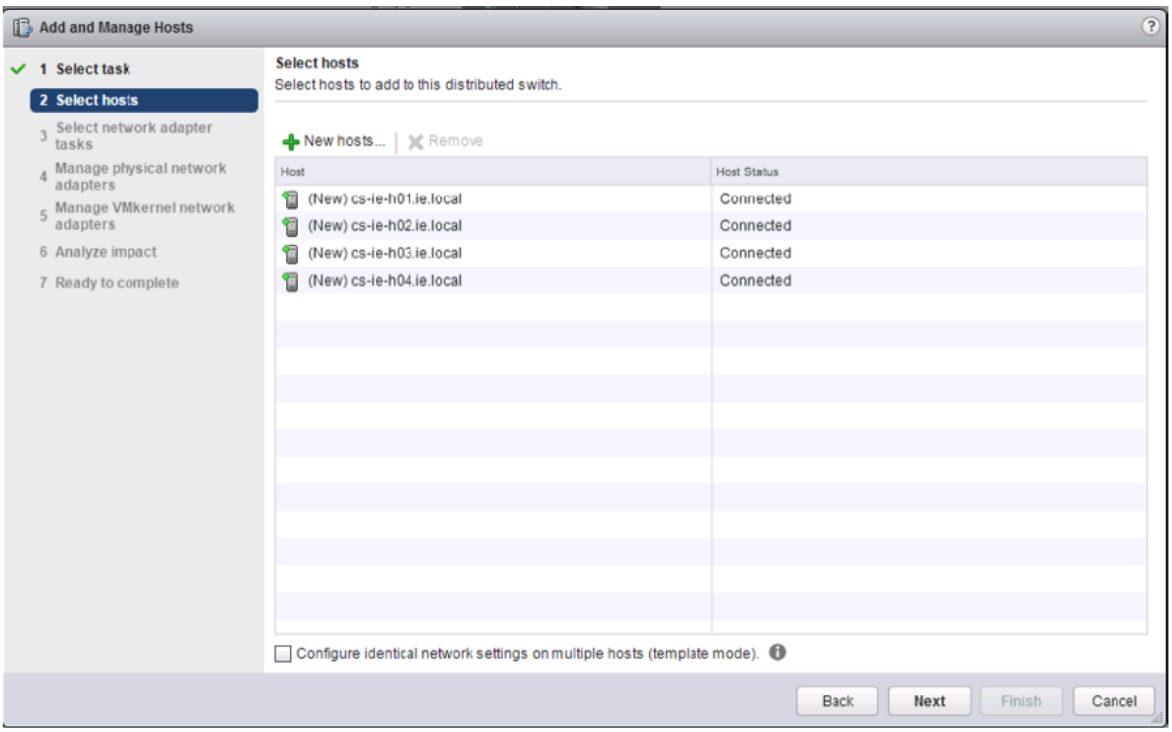
Warning: While the migration wizard allows many uplinks and many networks to be migrated concurrently, we recommend migrating the uplinks and networks step-by-step to proceed smoothly and with caution. For that reason, this is the approach we use here.

Migrate Management Network

To begin, let's migrate just the management network (vmk0) and its associated uplink, which in this case is vmnic0 from VSS to DVS. To begin, select "Add and manage hosts" from the basic tasks in the Getting started tab of the DVS.



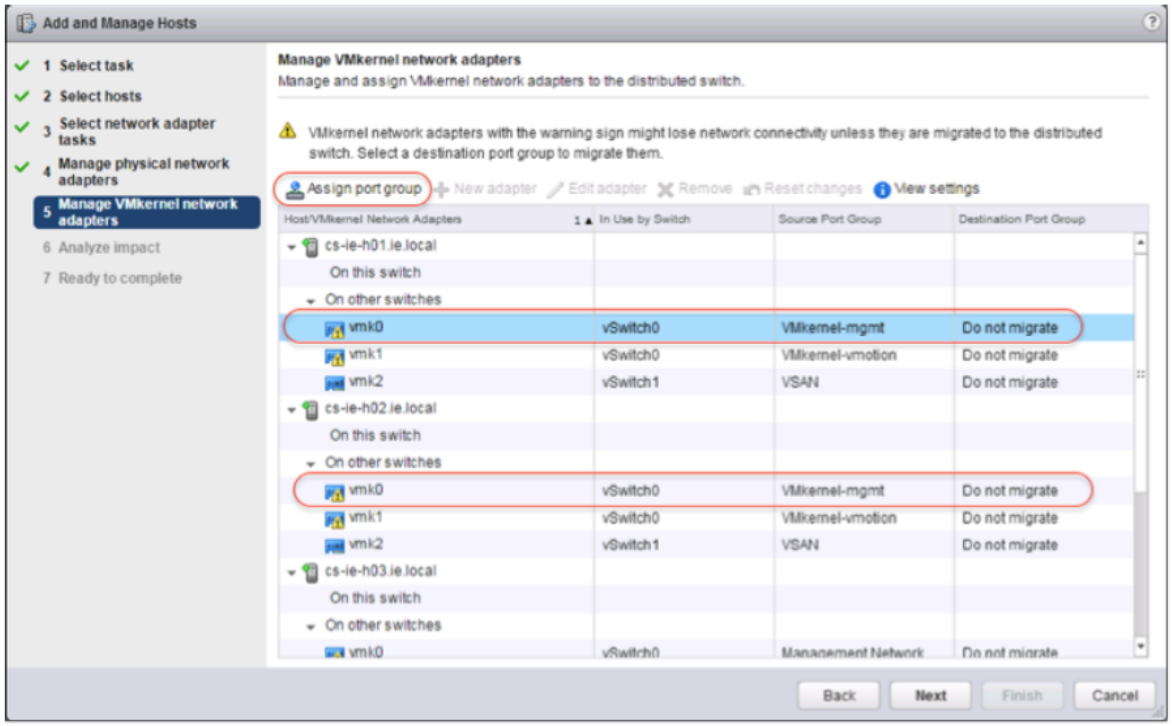
The first step is to add hosts to the DVS. Click on the green + and add all four hosts from the cluster.



The next step is to manage both the physical adapters and VMkernel adapters. To repeat, what we wish to do here is migrate both uplinks and VMkernel adapters to the DVS. Select physical adapters and VMkernel adapters, then select an appropriate uplink on the DVS for the physical adapter, for example Uplink1. Now the uplink (uplink1) to physical adapter vmnicX.

With the physical adapter selected and an uplink chosen, the next step is to migrate the management network from the VSS to the VDS. Leave the other VMkernel adapters for the moment and just migrate the management network VMkernel adapter.

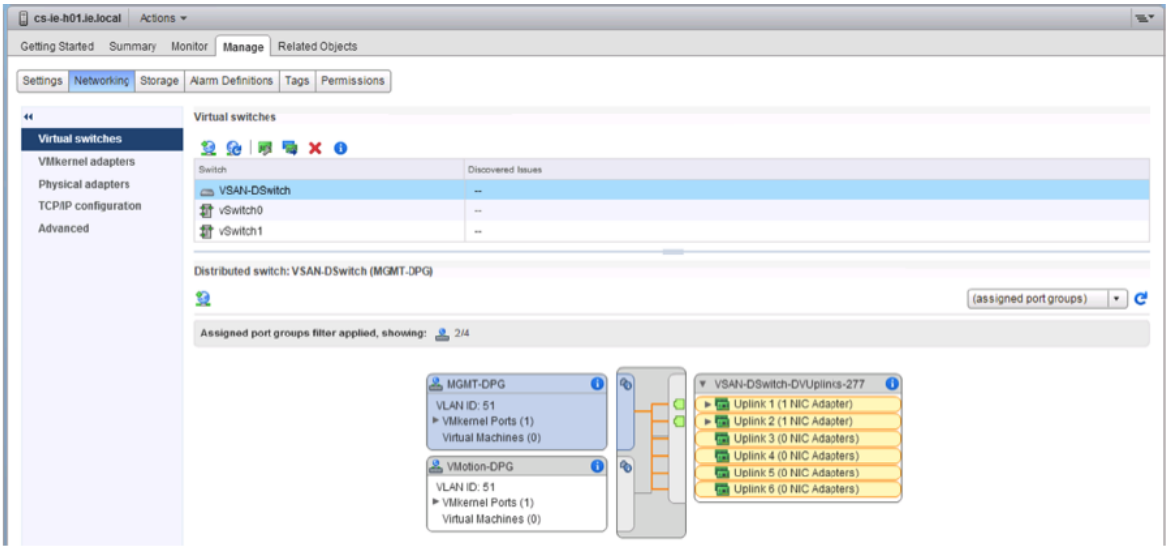
Select the management vmkernel, and then click on the "Assign port group". The port group assigned should be the newly created distributed port group created for the management network earlier. Remember to do this for each host.



Click through the analyze impact screen since it only checks iSCSI and is not relevant to vSAN. At the finish screen, you can examine the changes. It will display the number of hosts that are being added, the number of uplinks (vmnicX from each host) and number of VMkernel adapters (vmkX from each host). When the networking configuration of each host is now examined, you should observe the new DVS, with one uplink (vmnicX) and the vmkX management port on each host. You will now need to repeat this for the other networks.

Migrate vMotion Network

Migrating the vMotion network takes the exact same steps as the management network. Before you begin, ensure that the distributed port group for the vMotion network has all the same attributes as the port group on the standard (VSS) switch. Then it is just a matter of migrating the uplink used for vMotion (in this case vmnic1) along with the VMkernel adapter (vmk1). As mentioned already, this takes the same steps as the management network.

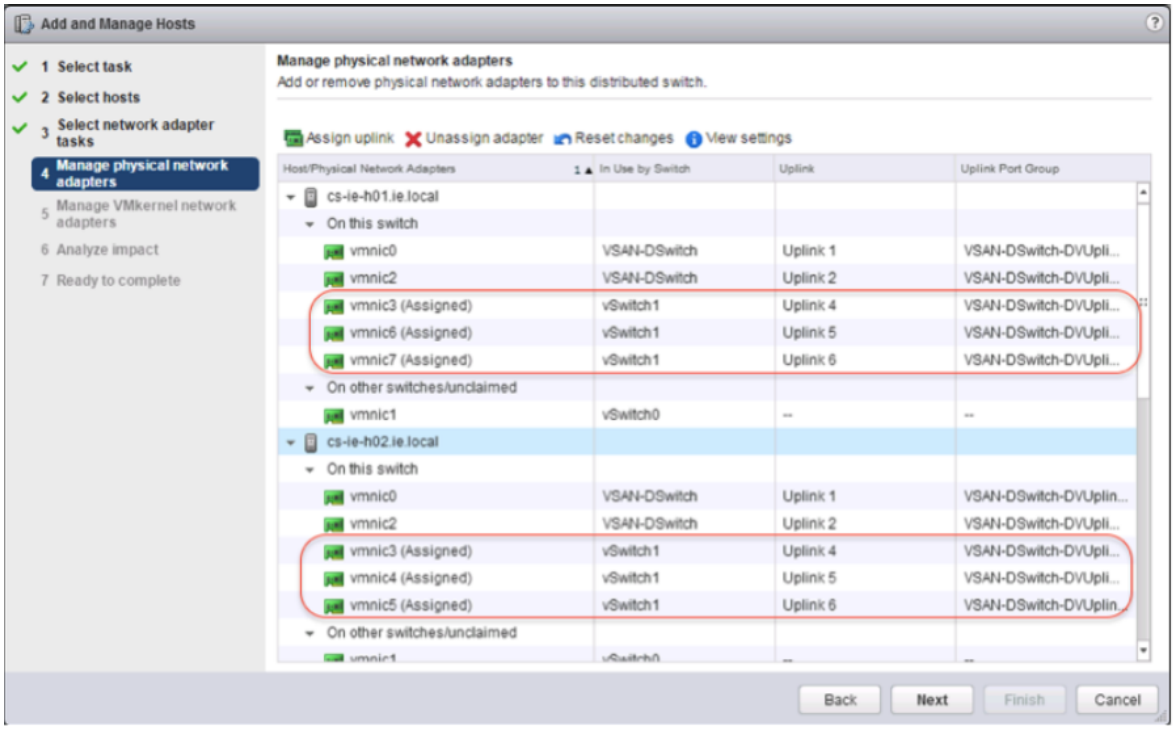


Migrate vSAN Network

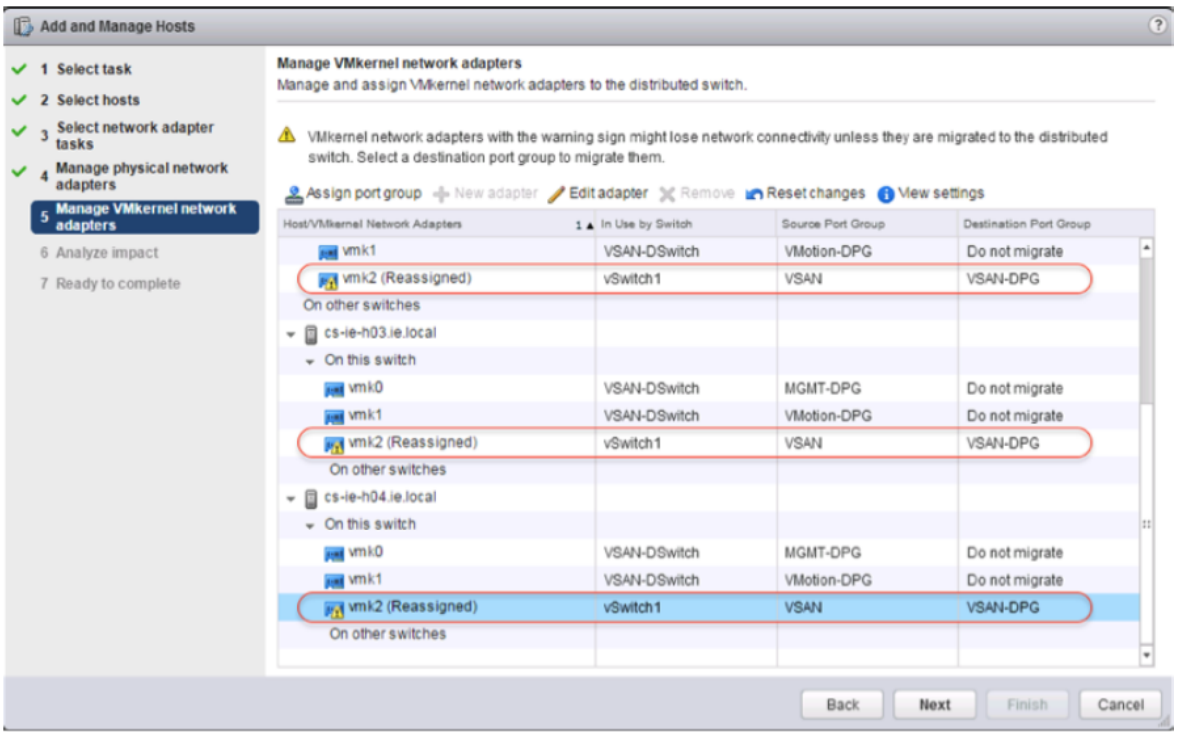
If you are using a single uplink for the vSAN network, then the process becomes the same as before.

However, if you are using more than one uplink, then there are additional steps to be taken. If the vSAN network is using a feature such as Link Aggregation (LACP), or it is on a different VLAN to the other VMkernel networks, then you will need to place some of the uplinks into an unused state for certain VMkernel adapters.

For example, in a scenario where the VMkernel adapter vmk2 and uplinks vmnic3, 4 and 5 are used for vSAN (which are in turn in a LACP configuration), all other vmnics (0, 1 and 2) must be placed in an unused state for vmk2. Similarly, for the management adapter and vMotion adapter, the vSAN uplinks/vmnics should be placed in an unused state. It is advisable to have uniform uplink configurations across all hosts to make thing easier. This may not always be the case, as in the example below, where the hosts are using different vmnics for the vSAN network.



Modifying the settings of the distributed port group and changing the path policy/failover appropriately do this. In the manage physical network adapter, the steps are similar as before except that now you are doing this for multiple adapters. As before, the vSAN VMkernel adapter should be assigned to the distributed port group for vSAN.



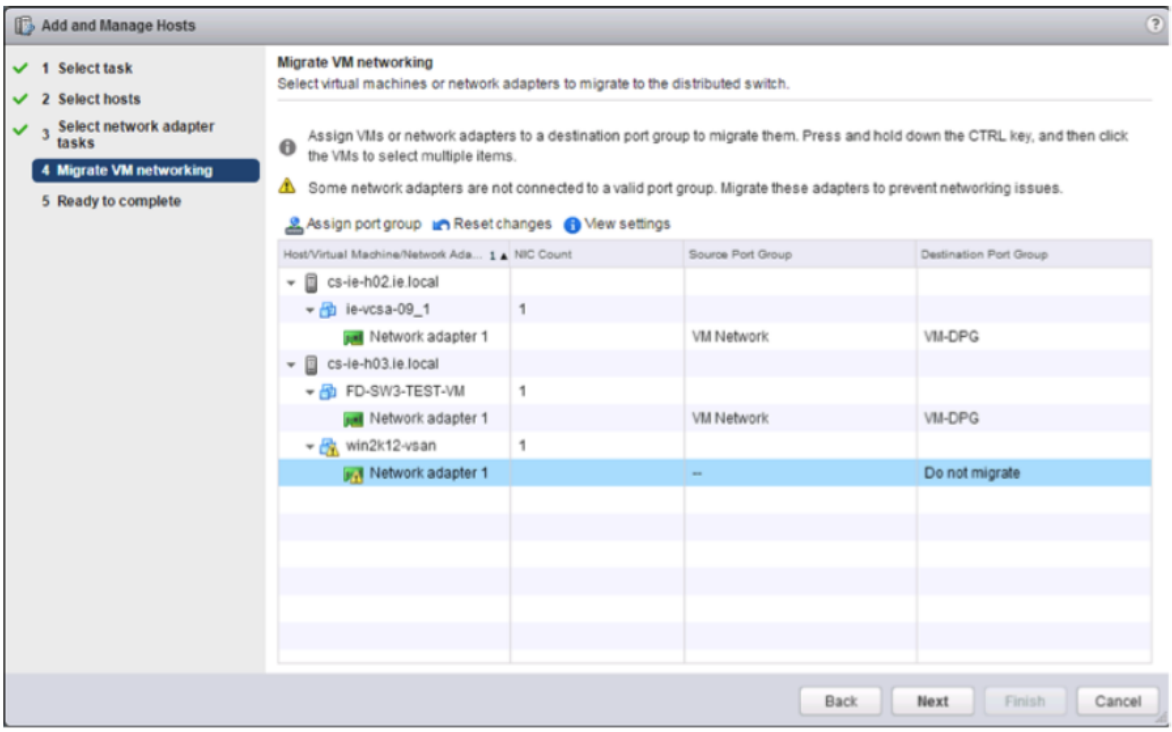
Note: If you are only now migrating the uplinks for the vSAN network, you may not be able to change the distributed port group settings until after the migration. During this time, vSAN may have

communication issues. After the migration, move to the distributed port group settings and make any policy changes and mark any uplinks that should be unused. vSAN networking should then return to normal when this task is completed. Use the Health Check plugin to verify that everything is functional once the migration is completed.

That completed the VMkernel adapter migrations. The final step is to move the VM networking.

Migrate VM Network

This is the final step of migrating the network from a standard vSwitch (VSS) to a distributed switch (DVS). Once again, we use the "Add and manage hosts", the same link used for migrating the VMkernel adapters. The task is to manage host networking. Select all the hosts in the cluster, as all hosts will have their virtual machine networking migrated to the distributed switch.



You may or may not need to move any uplinks depending on the configurtion. However, if the VM networking on your hosts uses a different uplink, then this of course would also need to be migrated from the VSS. Select the VMs that you wish to have migrated from a virtual machine network on the VSS to the new virtual machine distributed portgroup on the DVS. Click on the "Assign port group" option like we have done many times before, and select the distributed port group for virtual machine traffic.

Review the final screen. Note that in this procedure we are only moving to VMs. Note that any templates using the original VSS virtual machine network will need to be converted to virtual machines, edited and the new distributed port group for virtual machines will need to be selected as the network. This step cannot be achieved through the migration wizard.

Clean up

The VSS should no longer have any uplinks of port groups and can be safely removed. This completes the migration from a standard vSwitch (vSS) to a Distributed Switch (vDS).

4. Disk Operations

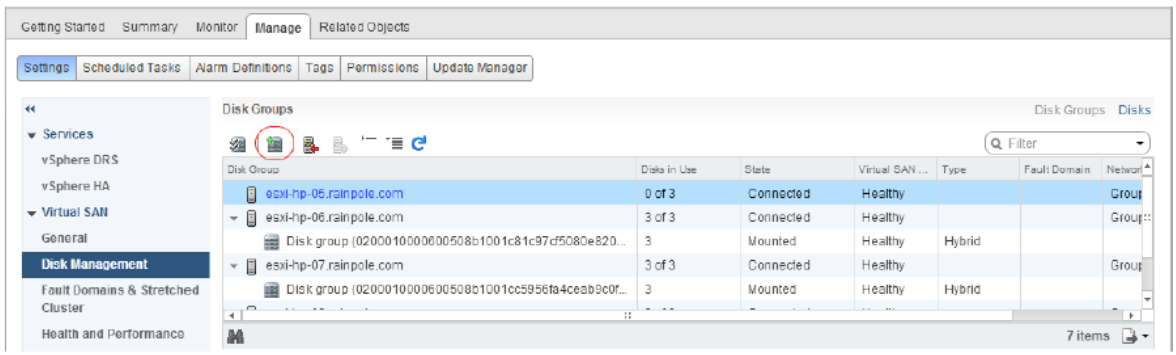
In this section of the vSAN Operations Guide, operations related to the disk subsection are discussed. This covers both cache and capacity tier devices, as well as hybrid and all-flash vSAN

4.1 Disk Operations

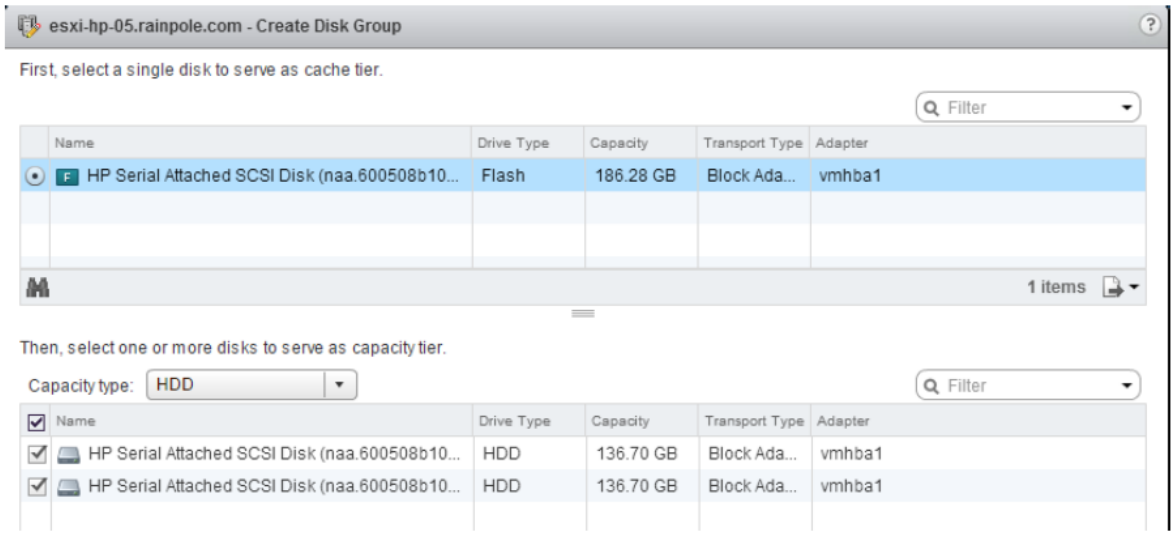
In this section of the vSAN Operations Guide, operations related to the disk subsection are discussed. This covers both cache and capacity tier devices, as well as hybrid and all-flash vSAN configurations.

4.2 Creating a Disk Group (Hybrid/All-Flash)

If the vSAN cluster is in manual mode, it will not automatically claim cache and capacity devices to build disk groups. An administrator may also want to have full control over which devices are used to make up a particular disk group. Therefore there is an option to manually create disk groups. To create a disk group, navigate to the cluster object in the inventory, select the Manage tab, then Disk groups. Next select the host on which the disk group is to be created, and the click on the disk group icon with the green plus sign to start selecting the devices for the disk group.



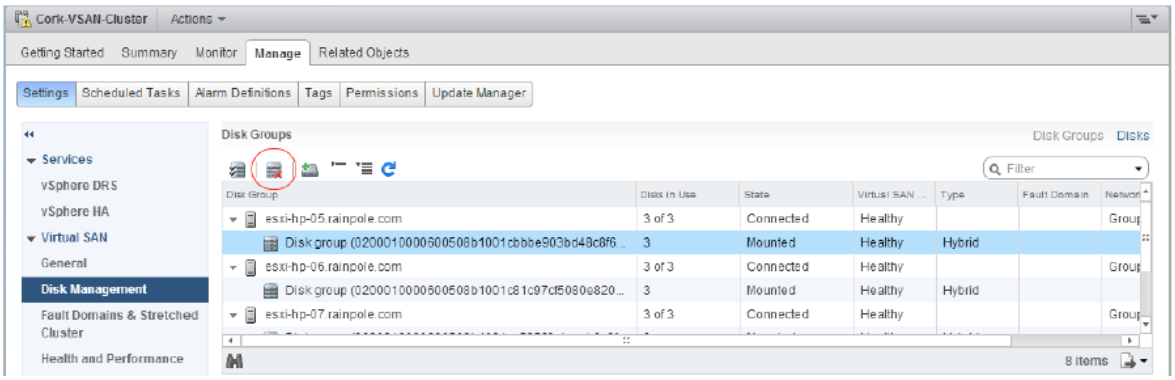
Now on hybrid configurations, you will be prompted for a cache device and one or more capacity devices. The cache device is a flash device and the capacity device is a HDD, which makes it easy to differentiate them. Here is such an example.



With all-flash, this is a little more complicated, where the devices are all flash devices. Therefore administrators need to pick an appropriate flash device for the cache tier and an appropriate flash device for the capacity tier.

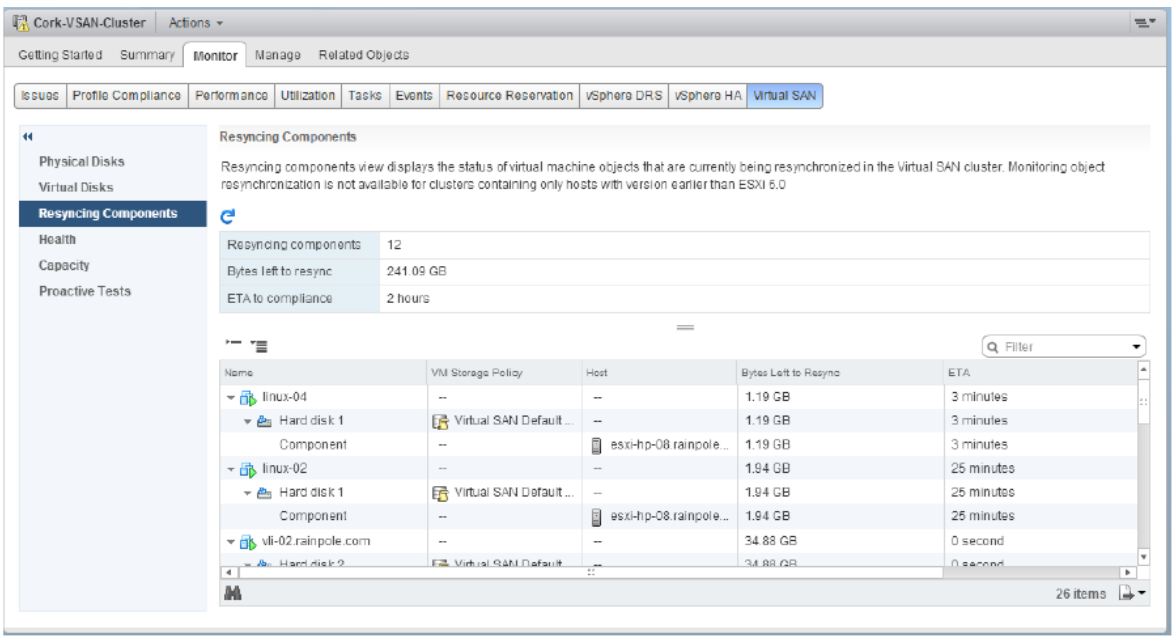
4.3 Removing a Disk Group

To remove a disk group, select the cluster object in the inventory, navigate to Manage and then Disk Groups. Select the disk group that you wish to remove. There will be an icon that represents a disk group with a red X, as shown below. Click that to begin the process of removing the disk group.



Once this option is selected, the administrator is then prompted as to whether or not they wish to evacuate all of the data from the disk group. VMware recommends that you respond 'yes' to this request as this will mean that your virtual machines remain protected even when the disk group is removed. Of course, this is only possible if there are enough resources (hosts and storage capacity) in the cluster.

Administrators can monitor the progress of a disk group evacuation by monitoring the resyncing components activity, as shown below.



4.4 Removing a Cache Disk (Failure) from a Disk Group

Removing a cache tier device is identical to removing the whole disk group. A disk group cannot exist without a cache tier device, and this process is effectively the same as removing the whole of a disk group as discussed previously.

4.5 Adding a Capacity Tier Device to a Disk Group

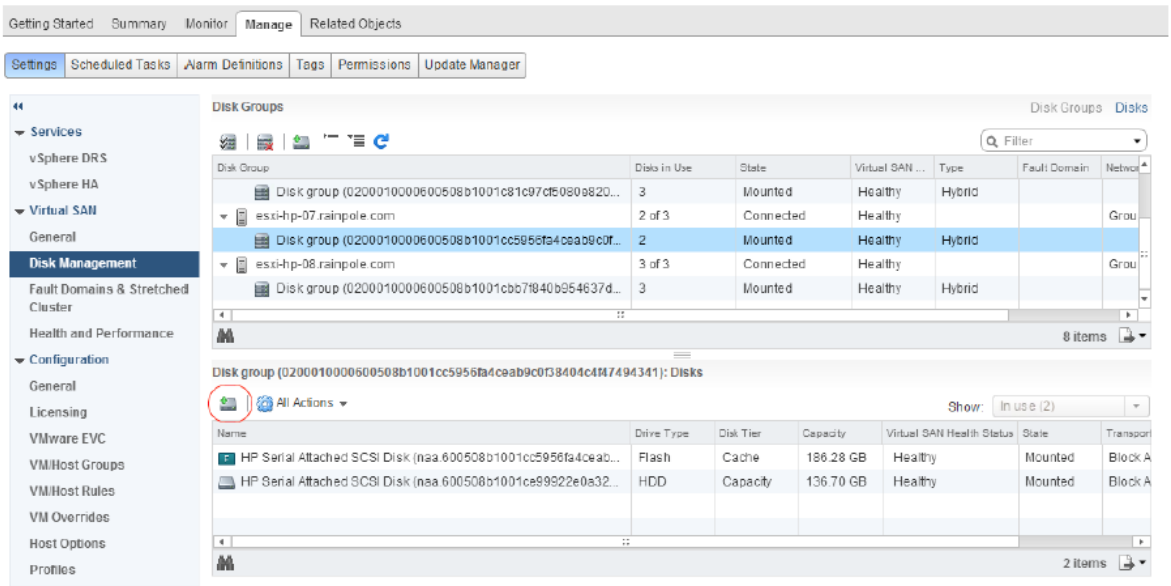
This step is only necessary if the cluster is in **manual** mode. If vSAN is configured in automatic mode, it will automatically claim any local, empty storage devices presented to the ESXi host.

Prerequisites:

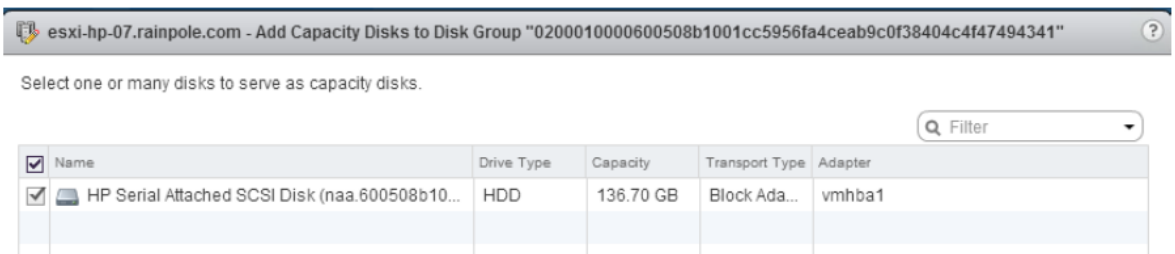
- vSAN cluster disk claiming is manual.
- The new disk must be the same as existing devices, such as SSD or magnetic disks.
- The new disk can NOT contain any partitions. See [vSphere 6.5 Remove Partition From Devices section].

Procedure:

1. Open vSphere Web Client.
2. Click **Hosts and Clusters**.
3. Select the *cluster* you are adding to, click the **Configure** tab.
4. Under vSAN, click **Disk Management**.
5. Find the host that contains the disk, and the select the appropriate *disk group.*
6. Click the Add a disk group icon



The UI will display a list of eligible storage devices that may be added to the disk group. Check the box on any of the devices that you wish to have added to the disk group, as shown below.

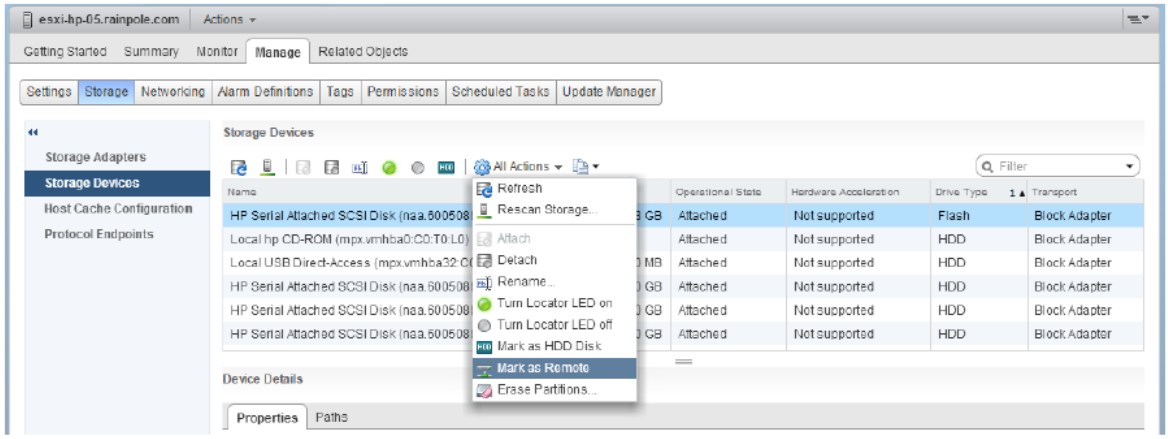


When the disk has been successfully added to the disk group, the vSAN datastore's capacity should grow appropriately.

4.6 Mark a Disk as Local/Remote

Some storage controllers allow their devices to be shared by more than one host. In cases like these, the ESXi host is not aware if the device is dedicated to this host, or if it is being accessed by another host. For this reason, ESXi marks any devices behind this controller as remote.

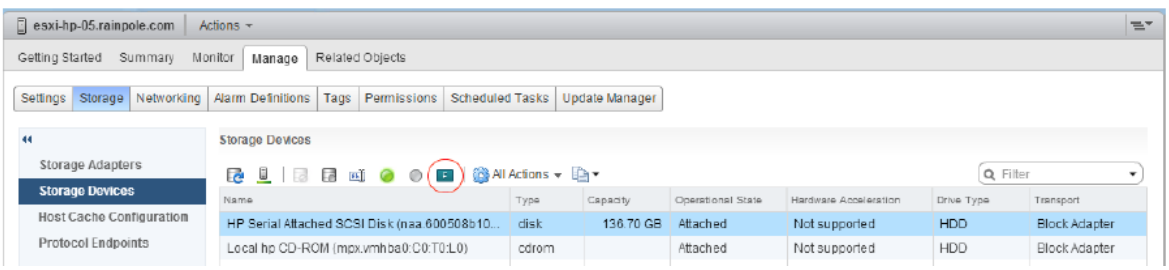
vSAN requires devices to be local, and will not automatically claim devices that are not local. Therefore administrators may need to mark a device that shows up as non local as local for vSAN to automatically claim it. This is a single step in the UI. Select the host in the inventory, then Manage > Storage > Storage Devices. Select the device in question, then right-click and select match the device as local. Similarly, the same procedure can be followed should there ever be a need to make the device as remote.



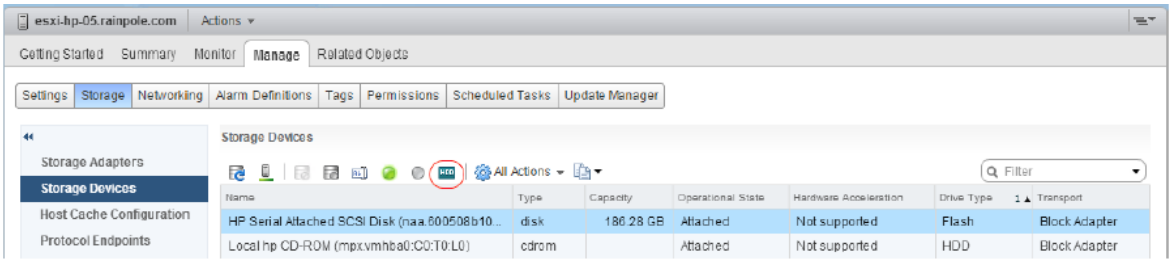
4.7 Mark a Disk as Flash or HDD

There may be occasions, especially when a storage controller cannot do pass-thru (which implies that each physical devices needs to be encapsulated in a RAID-0 volume) that the physical characteristics of a device are not made visible to the ESXi host. One of these characteristics is the type of device. In other words, is it a flash device such (SSD) or a spinning disk disk (HDD)? When flash devices are surfaced up as HDDs, then vSAN cannot consume them for the cache tier. Therefore these devices must be tagged as SSDs/flash via the UI.

Here is where to tag a flash device, that has been detected as an HDD, as flash. Select the host in the inventory, then Manage > Storage > Storage Devices.



Similarly, in the case of all-flash clusters in vSAN 6.x, vSAN needed to be informed that flash devices are actually HDD devices so that they can be consumed for the capacity tier. This is no longer the case with later versions of vSAN, as there is a more intuitive way to claim flash devices for both the cache and capacity tiers, but this is where a flash device can be tagged as a HDD so that it could be consumed for the capacity tier in earlier versions of vSAN.

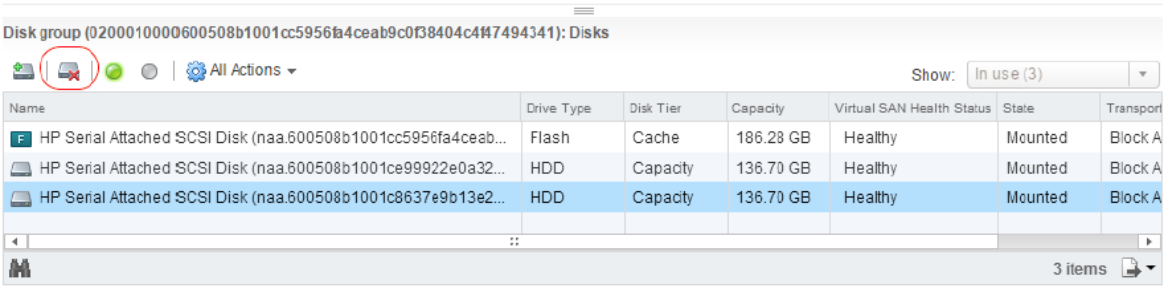


4.8 Removing a Capacity Disk

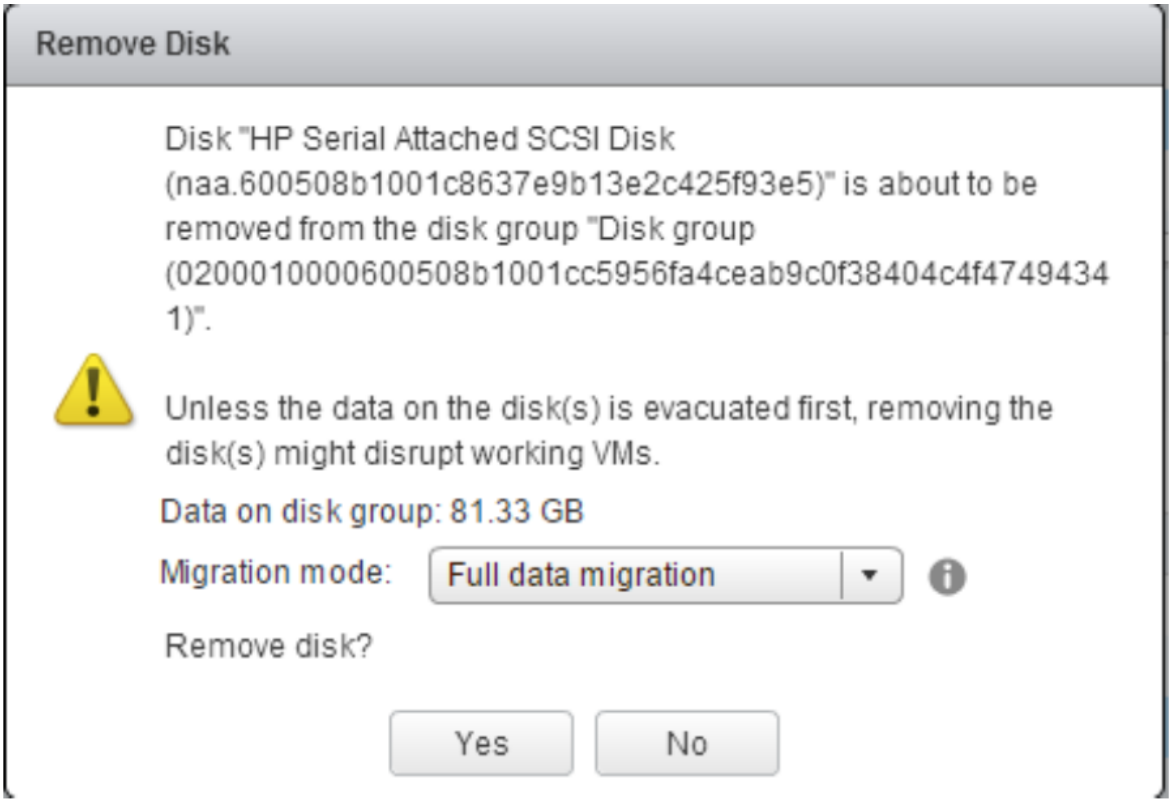
In this section, the procedure to remove a disk from a disk group is discussed. This procedure can be as a result of different activities, such as replacing a failed disk or replacing a capacity device for a larger one.

In order to remove a disk from a disk group, navigate to the vSAN cluster object in the vCenter inventory, then select Manage, followed by Disk Management. Find the host that contains the disk, and the select the appropriate disk group.

If the cluster is in manual mode, which it needs to be for this operation to succeed, a red X will be available when the physical disk is selected in the disk group. This is visible in the figure below.



When this icon is clicked to remove a disk from a disk group, you will be prompted to evacuate the existing components that are on the disk. In order to maintain full protection for the virtual machines, it is always recommended that you do a full data evacuation. This means that even after the disk has been removed, none of the objects are at risk of a failure elsewhere in the cluster impacting availability or accessibility.



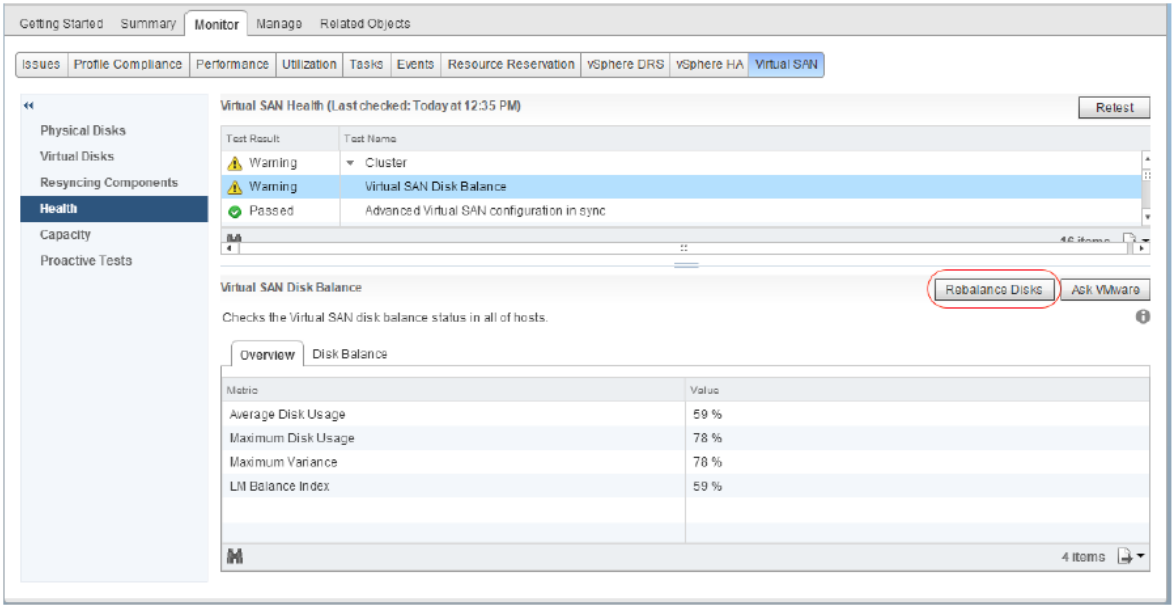
Leave the migration mode at "Full data migration" and click ok.

Caution: If there are not enough resources in the cluster to do a full data migration (e.g not enough nodes, not enough space), you need to be aware that your virtual machines are at risk while you replace this disk, and rebuild the components that were on the original disk.

4.9 Balance the Disk Usage

As capacity device are evacuated and removed from the vSAN Cluster, and new capacity devices are added, you may find that the vSAN cluster becomes unbalanced from a capacity usage perspective. This unbalanced is reported as part of the vSAN health checks, and is easily rectified via the vSAN health check.

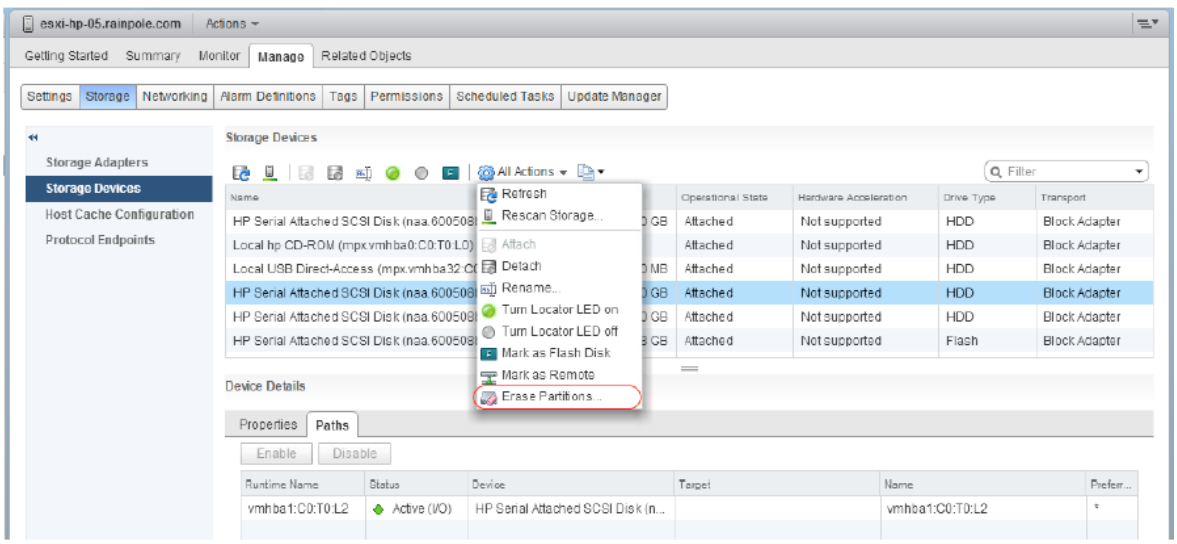
Navigate to the vSAN Cluster > Monitor > vSAN > Health view, and under the Cluster checks, there is a check called vSAN Disk Balance. If the maximum variance is above a particular threshold, administrators have the option to rebalance the disk usage by click on the "Rebalance disks" button highlighted below. This will move components from the over-utilized disks to the under-utilized ones.



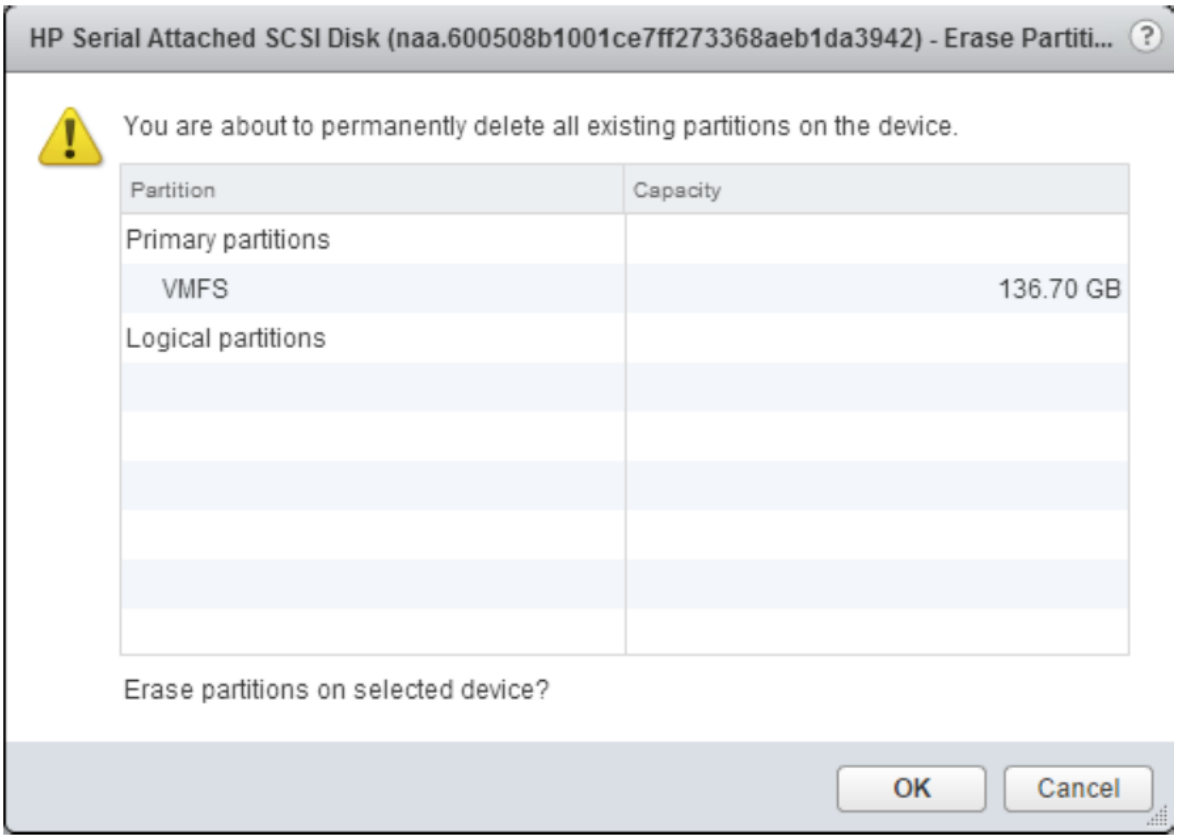
VMware recommends that this operation is done during non-production hours as it may introduce some additional overhead in the cluster. If at any time the rebalance operation is impacting the cluster in any way, the administrators can choose to stop the rebalance operation at any time, and resume it again at some point in the future.

4.10 Removing a Partition From a Disk

vSAN can only claim local, empty disks. If a disk was previously used for other storage, such as a VMFS volume, it cannot be automatically consumed by vSAN. First, the disk will need to have the existing partition information removed. This can be done via the UI. Select the host, then Manage, Storage and then Storage Devices. In the "All Actions" dropdown menu, there is an option to erase partitions from a disk device.



In this example, there is an existing VMFS partition displayed, so administrators can be sure that they are erasing the partitions from the correct disk.

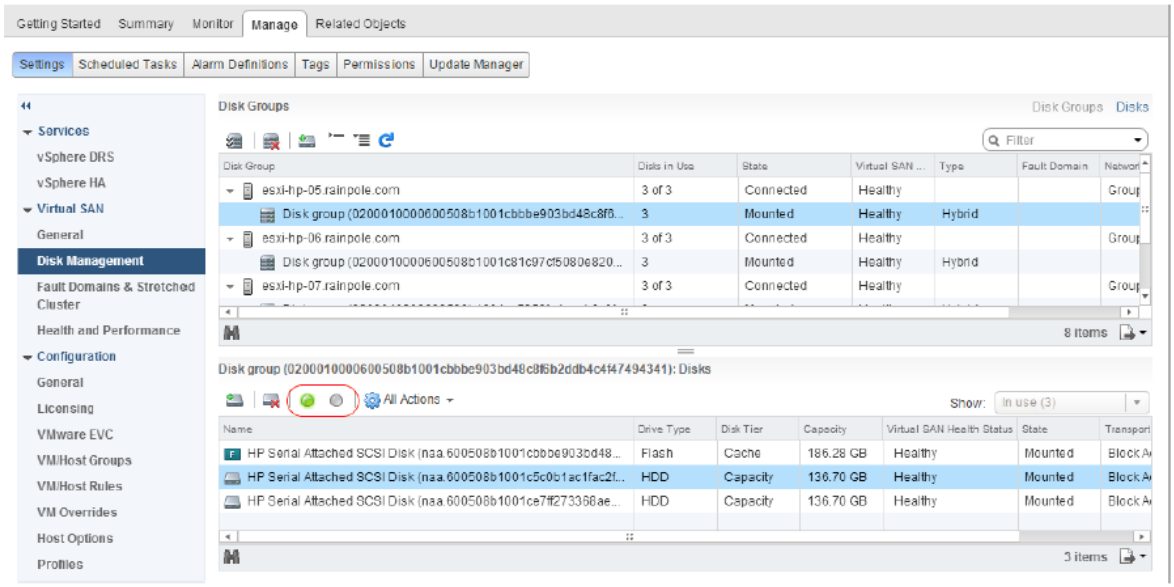


Once the existing partition information is removed and the device is empty, it can be claimed by vSAN (as long as it is a local device).

4.11 Blinking a Disk LED

This feature is design to assist administrators in identifying disks in very large vSAN farms. The utility, which can be driven through the vSphere UI or CLI, blinks the LED on the front of a disk drive for easy recognition. Note that there may be a requirement to have special VIBs installed on the ESXi host for this functionality to work. In many cases (e.g. DELL, HP), the special OEM builds of ESXi from these partners already includes any necessary VIBs required to make this functionality work. Otherwise the appropriate VIBs may need to be downloaded from the server vendor and installed before this feature can be used.

The icons to blink the LEDs on and off are found in the disk view. Select the cluster in the inventory, then the Manage tab, then Disk groups and then select the disk that you wish to blink the LEDs on. there are two icons for this task; one turns on the LED blinking and the other turns it off again. These are shown in the figure below.



5. Datastore Operations

Before undertaking any actions through the Datastore Browser, please be noted that it is not recommended to delete VMs through the browser.

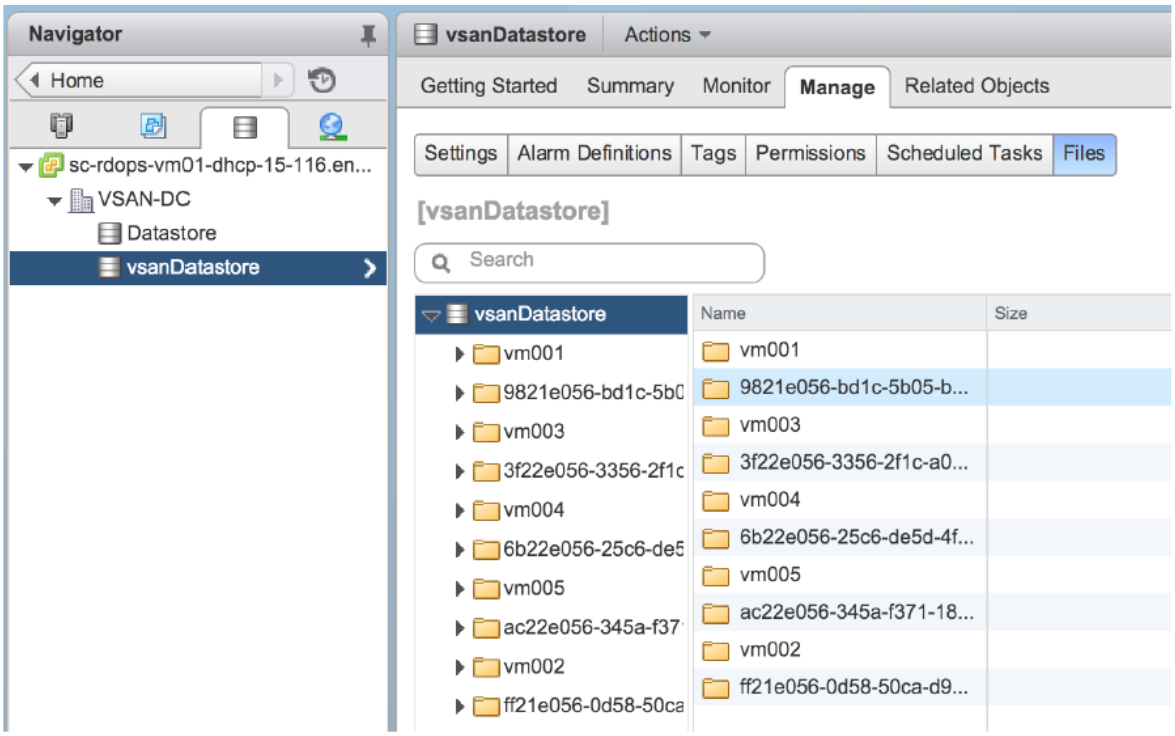
5.1 Datastore Operations

Before undertaking any actions through the Datastore Browser, please be noted that it is not recommended to delete VMs through the browser. In order to delete VMs, please use the "remove/delete" option in the vSphere Web Client. Same applies to programmatically deleting VMs.

5.2 Browsing vSAN Datastore Contents

Browsing a vSAN datastore is no different than any other datastore in your environment.

1. Open the vSphere Web Client.
2. Click the **Storage** tab.
3. Right-click the *VSAN Datastore* and click **Browse Files**.



5.3 Uploading files to vSAN Datastore

Files can be uploaded to the vSAN, however it should be pointed out that files should not be stored in the root folder. Before uploading a directory should be created where the files will be stored.

1. Open the vSphere Web Client.
2. Click the **Storage** tab.
3. Right-click the *VSAN Datastore* and click **Browse Datastore**.
4. Click the **Create Folder** icon .
5. Provide a *name* and click **OK**.
6. Select the *new folder* by clicking it

7. Click the **Upload to a datastore** icon.

8. Browse to the *file* and click **Open**

5.4 Maintaining Sufficient Slack (Free) Space

vSAN "slack space" is simply free space that is set aside for operations such as host maintenance mode data evacuation, component rebuilds, rebalancing operations, and VM snapshots. Activities such as rebuilds and rebalancing can temporarily consume additional raw capacity. Host maintenance mode reduces the total amount of raw capacity a cluster has while a host is in maintenance mode. This is because the local drives on a host that is in maintenance mode do not contribute to vSAN datastore capacity until the host exits maintenance mode.

Recommendation: Maintain 25-30% slack space when designing and running a vSAN cluster.

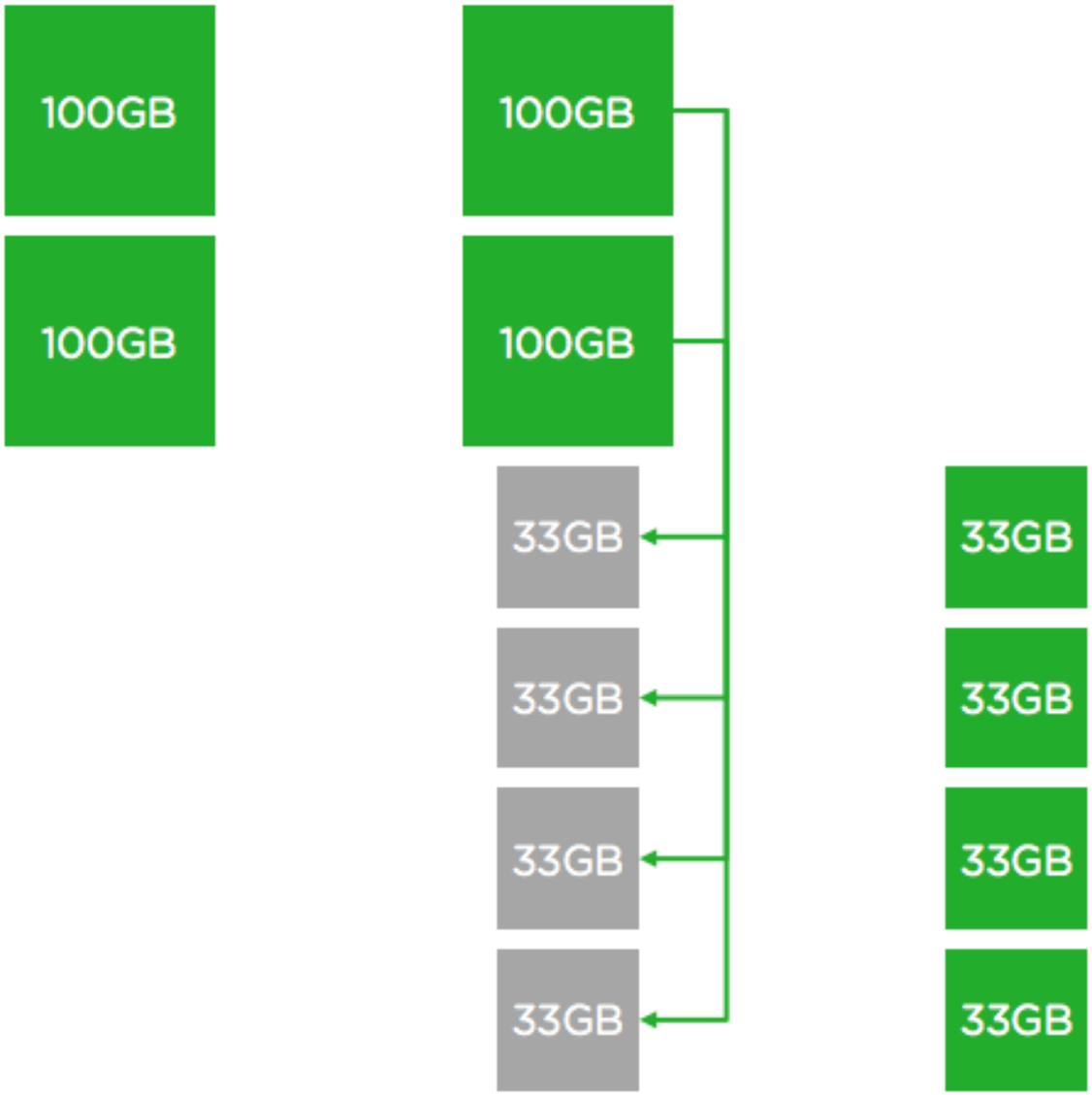
For example, a vSAN datastore with 20TB of raw capacity should always have 5-6TB of free space available for use as slack space. This recommendation is not exclusive to vSAN. Most other HCI storage solutions follow similar recommendations to allow for fluctuations in capacity utilization.

When one or more storage devices that contribute capacity to a vSAN datastore are more than 80% utilized, vSAN will automatically initiate a reactive rebalance of the data across vSAN storage devices in an attempt to bring utilization below 80%. This rebalancing activity naturally generates additional IO in the vSAN cluster. Maintaining 25-30% slack space minimizes the need for rebalancing operations while accommodating temporary fluctuations in utilization due to the various activities mentioned above.

To help gain a better understanding, here is one example of a temporary increase in capacity utilization due to a storage policy change:

A 100GB virtual disk is assigned a storage policy that includes the rules Primary Level of Failures to Tolerate = 1 and Failure Tolerance Method = RAID-1 mirroring. vSAN creates two full mirrors or "replicas" of the virtual disk and places them on separate hosts. Each replica consists of one component. There is also a witness component created, but we will not factor that in as witness components are very small – typically, around 2MB. The two replicas for the 100GB virtual disk objects consume up to 200GB of raw capacity (objects on a vSAN datastore are "thin provisioned" by default). We will assume deduplication and compression are not enabled to keep this example simple.

A new storage policy is created. Primary Level of Failures to Tolerate = 1 and Failure Tolerance Method = RAID-5/6 erasure coding. The new policy is assigned to that same 100GB virtual disk. vSAN begins copying the existing mirrored components to a new set of components distributed in a RAID-5 erasure coding configuration. Data integrity and availability are maintained as the mirrored components continue to serve reads and writes while the new RAID-5 component set is built. This process naturally consumes additional raw capacity as the new components are built. Once the new components are completely built, IO is transferred to the new components and the old mirrored components are deleted. The new RAID-5 component set consumes up to 133GB of raw capacity. This means all of the components for this object could consume as much as 333GB of raw capacity just before the resync is complete and the old RAID-1 mirrored components are deleted. After the RAID-1 components are deleted, the capacity that was consumed by these components is automatically freed up for use with other operations.



As you can imagine, performing this storage policy change on multiple VMs concurrently could cause a considerable amount of additional raw capacity to be consumed. Likewise, if a storage policy that is assigned to many VMs is modified, more capacity will likely be needed temporarily to make the necessary changes to components that make up these VMs.

6. VM Storage Policies Operations

With the initial release of vSAN, there were five Virtual Machine storage policies that could be chosen.

6.1 VM Storage Policies Operations

With the initial release of vSAN, there were five Virtual Machine storage policies that could be chosen. These were:

- Number of failures to tolerate
- Number of disk objects to stripe
- Force Provisioning
- Flash Read Cache Reservation (%)
- Object Space Reservation (%)

In vSAN 6.2, an extra three policies are introduced:

- Failure Tolerance Method - also known as erasure coding or RAID-5/RAID-6
- Software Checksum
- IOPS limit per object.

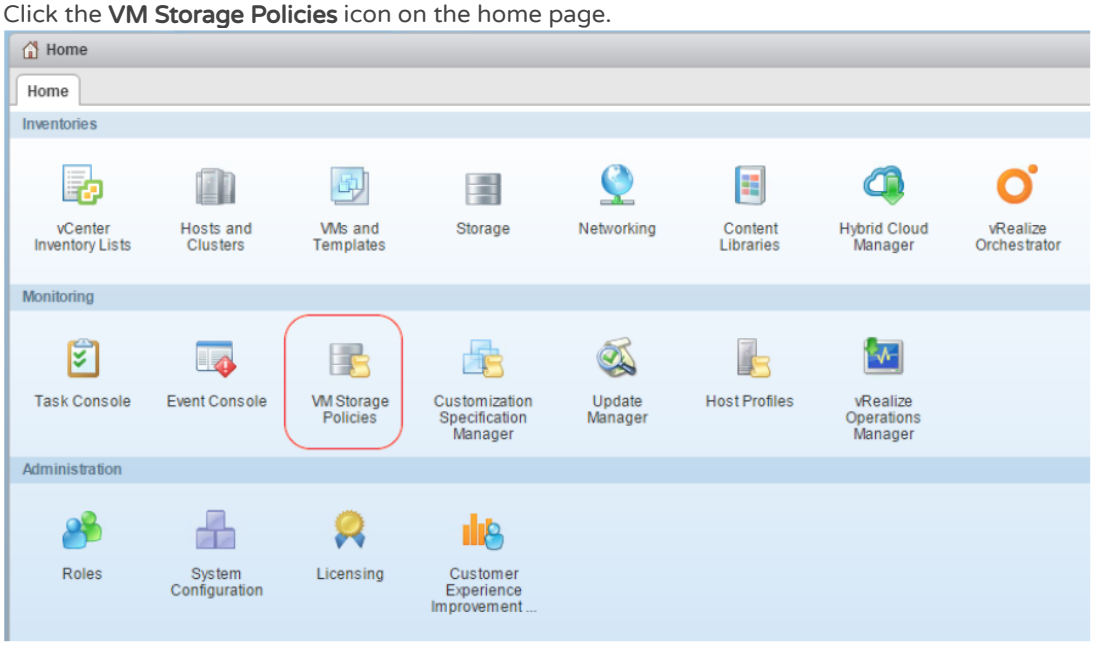
The operations guide will not describe each of these capabilities in detail. This information can be found in the vSAN Administraton Guide.

6.2 Creating a Policy

vSphere has two default storage policies available: one for vSAN and one for VVols (virtualized volumes). To create a new VM Storage Policy, follow the click-through demo, vSAN 6.5 - Create and Assign a Storage Policy, follow the vSphere 6.5 Creating and Managing VM Storage Policies or the following procedure:

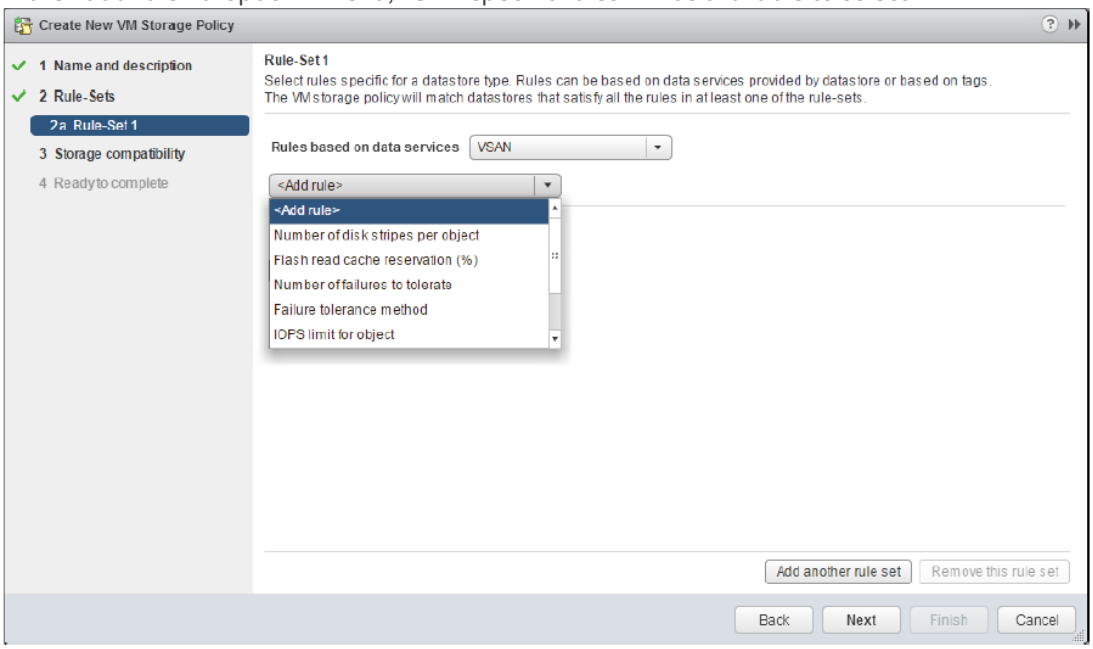
Define your VM storage policy prior to starting this process.

1. Open the vSphere Web Client.
2. Click the **VM Storage Policies** icon on the home page.

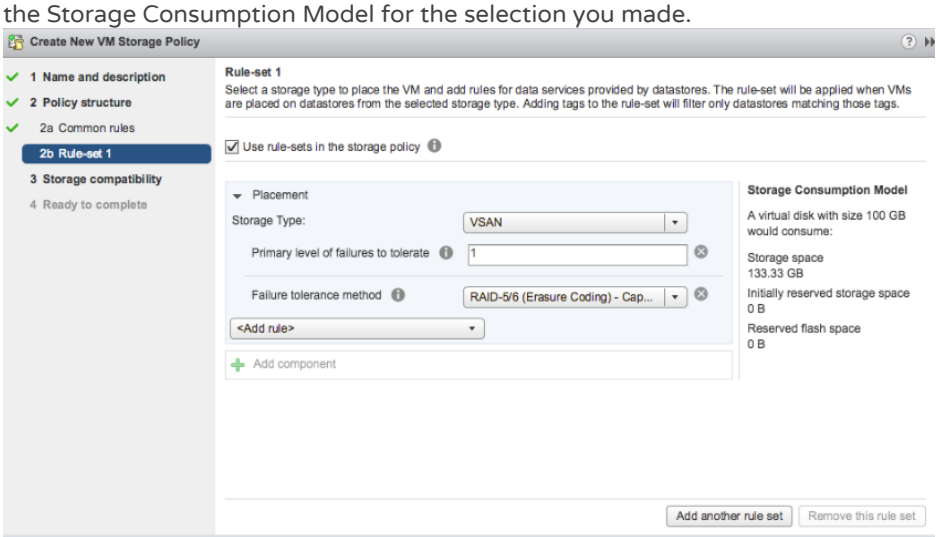


3. Click the **Create VM Storage Policy**. This will start the create VM storage policy wizard.
4. Select the *vCenter Server* for the policy.
5. Provide a *name* and *description* for the policy. Click **Next**.

6. You will be presented with a description of rule-sets, and how multiple rule-sets can be defined for a single policy if necessary. In this example, we will keep this simple and only create a single rule-set in our policy. Click **Next**.
7. If you are not using common rules, click **Next**.
8. Select the checkbox for **Use rule-sets in the storage policy.**
9. Select **VSAN** from the *storage type* dropdown.
10. In the <add rule> dropdown menu, vSAN specific rules will be available to select.



   - For each rule:
     - Select the rule from the dropdown, such Number of failures to tolerate.
     - The next screen will show the default value, which can be modified. It will show the Storage Consumption Model for the selection you made.



     - Select the <Add rule> dropdown until all of your rules are defined.
     - Click **Next**.
11. The storage compatibility screen is displayed. This should tell you whether the policy you chose is compatible with the vSAN configuration. For example, if I tried to create a RAID-5/RAID-6 configuration on a hybrid array, or if I tried to set Number of failures to tolerate to a higher value and there were not enough hosts in the cluster (to tolerate n failures with RAID-1 mirroring, there needs to be 2n + 1 host in the cluster, then the vSAN datastore would not show up as compatible. On this screen, an incompatibility reason is also provided (e.g. cluster is not all-flash, or there are not enough hosts/fault domains in the cluster). The storage compatibility

screen should always be examined to make sure that the policy you are creating can be satisfied by the cluster. Click **Next**.

12. The "Ready to complete" screen lets you review the policy before creating it. If you are done, click **Finish**.

The new policy is now in the list of available policies and may be chosen a VM provisioning time, or indeed applied to already existing VMs.

6.3 Editing a Policy

To edit a VM Storage policy, follow the vSphere 6.5 Creating and Managing VM Storage Policies.

You can edit the name or the set of capabilities. You can also check the compatibility with the vSAN datastore. If the storage policy has already been applied to a virtual machine(s) you will be able to reapply the changed policy immediately or later.

6.4 Deleting a Policy

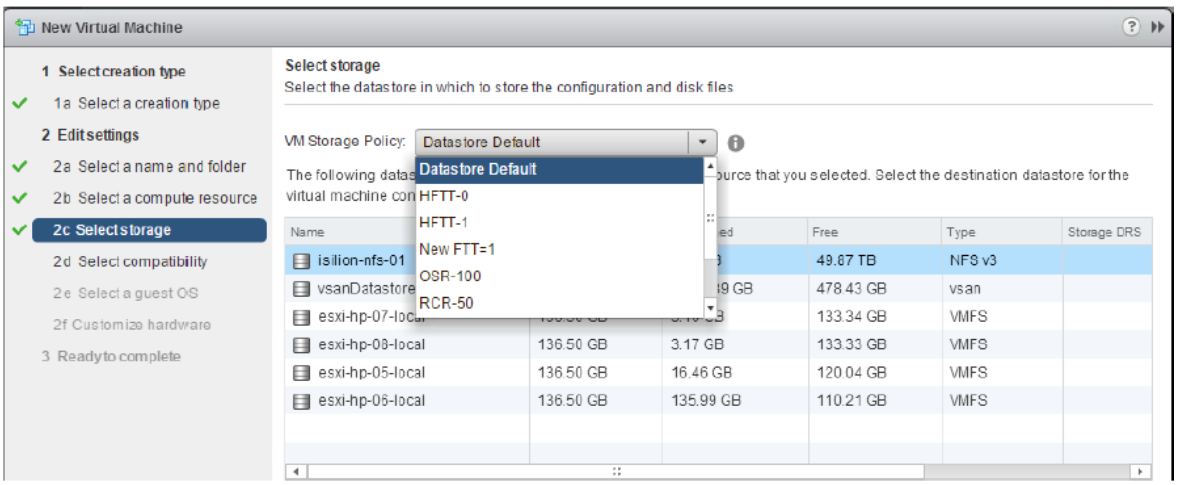
Deleting a policy is also very straight-forward. Follow the procedure in the vSphere 6.5 Creating and Managing VM Storage Policies.

NOTE: If a storage policy is in use by a VM, it cannot be deleted. A new policy needs to be associated with that VM. This procedure will be covered shortly.

6.5 Applying a Policy

To apply a storage policy to a new virtual machine or change it on-the-fly, follow the procedures in the vSphere 6.5 Storage Policies and Virtual Machines.

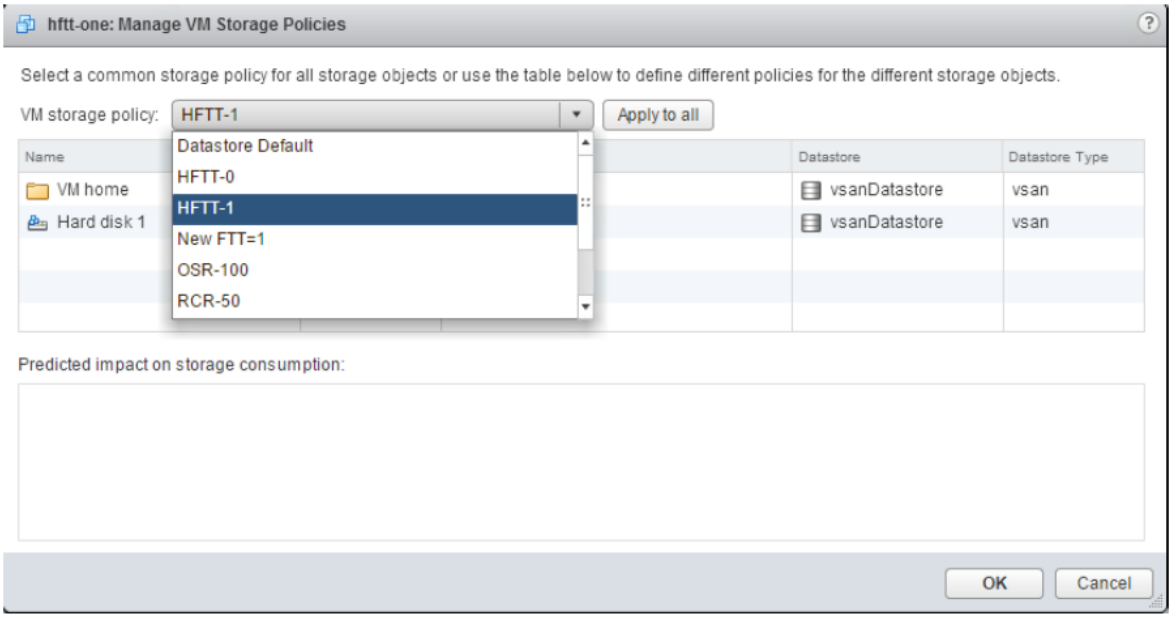
Policies can be chosen when a virtual machine is first deployed, but it may also be changed when the VM is already running. Here is an example where the policy is chosen when the VM is being deployed. In the next section, how to change the policy on-the-fly is discussed.



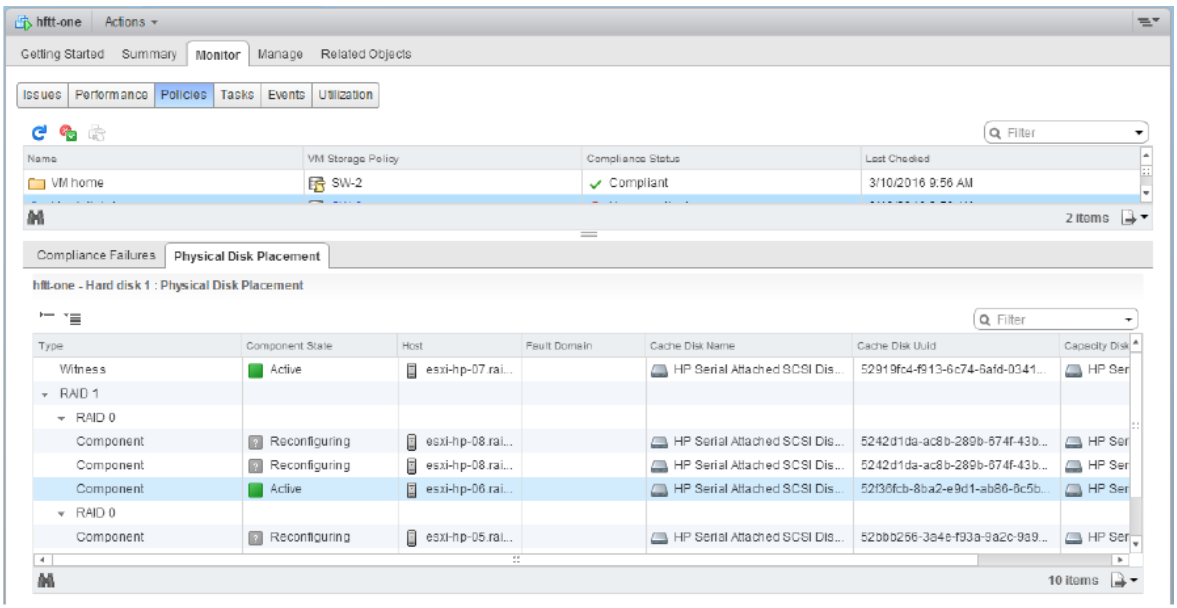
6.6 Changing a Policy On-the-Fly (What Happens)

To change a storage policy on-the-fly, follow the procedures in the vSphere 6.5 Storage Policies and Virtual Machines.

NOTE: If you want this policy to apply to both the VM Home Namespace object and the VMDK objects, the "**Apply to all**" button should be clicked. Otherwise, the new policy change will only apply to the VM home object.

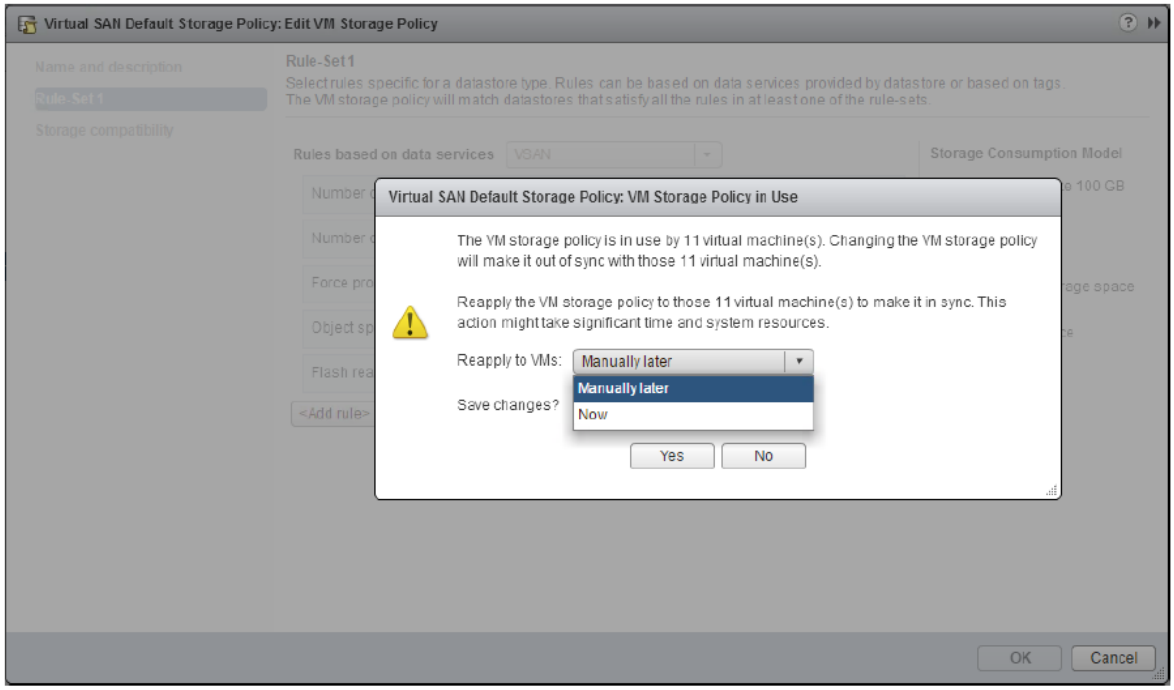


In many cases, this operation will result in a build of new objects to match the requirements of the new policy. For example, if you wish to increase the stripe width, or you wish to reserve some space on the vSAN datastore for the object. In other cases, such as reducing the number of failures to tolerate value for a RAID-1 object, vSAN simply needs to remove one replica, so there is no need to build new objects. To observe this activity, once again select the VM, then the Monitor tab, then Policies, followed by the Physical Disk Placement tab. This will show any objects that are reconfigured as a result of a policy change.



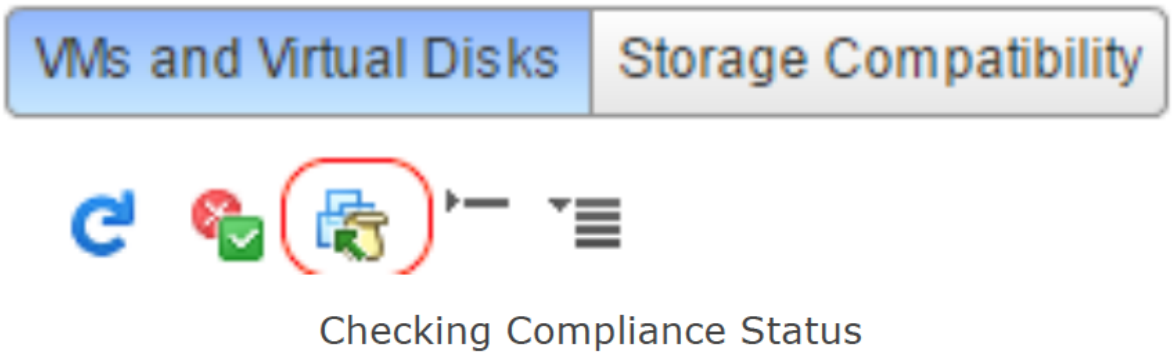
6.7 Bulk Assign Storage Policies to Multiple VMs

There might be an occasion where you would like to change the policy associated with multiple virtual machines at the same time. The assumption here is that the VMs in question are all sharing the same common policy. The first step is to make the appropraite changes in the policy that they VMs are sharing. When the policy is changed, SPBM (Storage Policy Based Management) knows how many VMs are using the policy, and prompts the administrator to apply the new policy to the VMs either now or later.



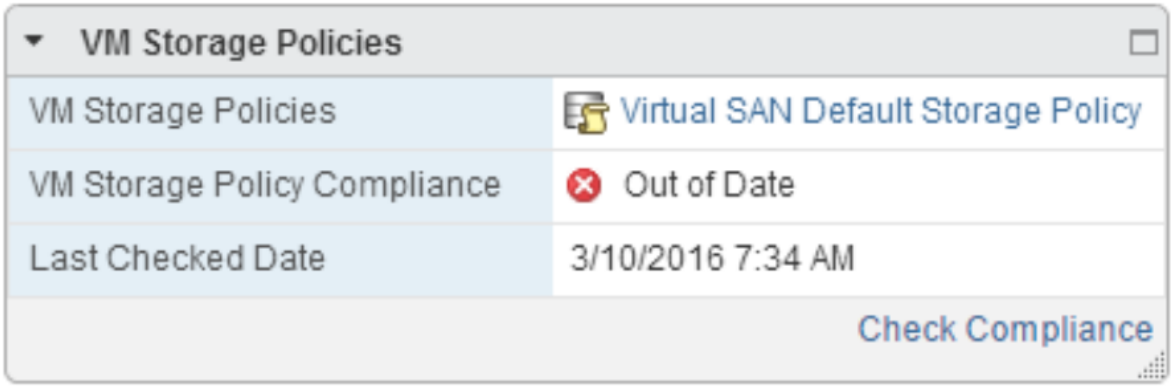
In this example, 11 VMs are using the default policy. If we change this policy, and reapply it to the VMs now, multiple new components could be rebuilt and resynced to the current objects depending on the change. Alternatively, if the administrator decides to do it later, the compliance will be shown as "Out Of date".

At a later point, you can bring the objects to compliance by navigating to VM Storage Policies, selecting the policy in question, then the Monitor tab. In the VM and Virtual Disk view, select all the VM, and then click on the icon (shown below) to reapply the policy to all out of date entities.

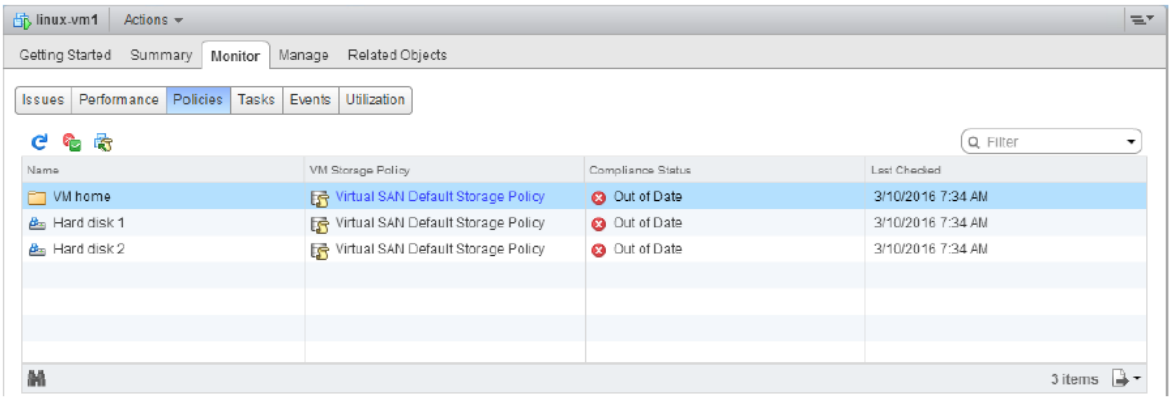


6.8 Checking Compliance Status

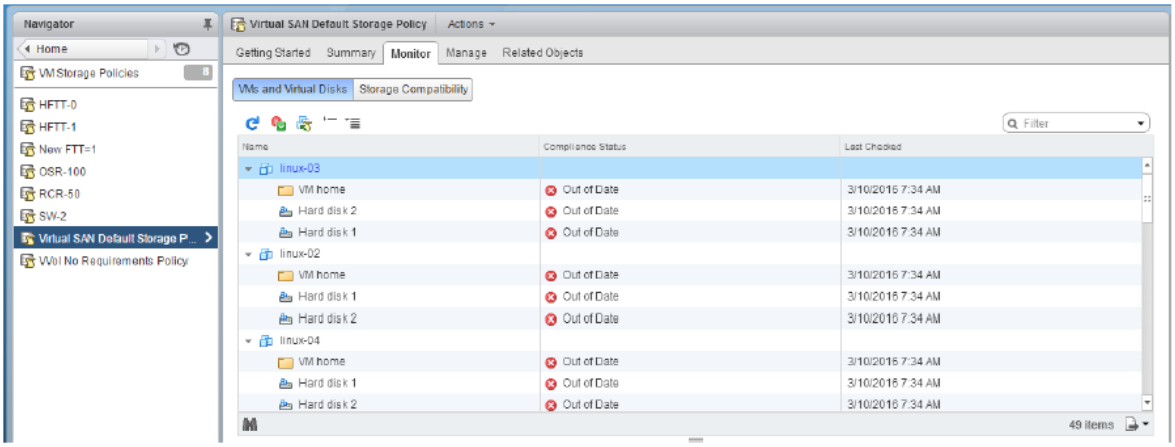
Compliance status an be checked in a number of places. Individual VMs compliance can be checked via the Summary tab of the VM.



The individual components of a VM can be examined by selecting the VM, then Monitor and Policy view.



To look at the compliance of all VMs using a particular policy, revert to the VM Storage Policies section, select the policy in question and then Monitor. The VMs and Virtual Disks shows all VMs that are using the policy and their respective Compliance Status.



6.9 Backing up Policies

There is no way to specifically backup a VM Storage Policy outside of backing up vCenter Server. However, it should be noted that even if the vCenter server where the policies were created are lost, it has no impact on the alreayd running VMs. They continue to use the policy attributes assigned to them

by SPBM, and these VMs' policies can be interogated even when vCenter no longer exists through Ruby vSphere Console (RVC) commands.

```
**vsan.vm_object_info -h**
usage: vm_object_info [opts] vms...
Fetch VSAN object information about a VM
vms: Path to a VirtualMachine
-c, --cluster=                    Cluster on which to fetch the object info
-p, --perspective-from-host=      Host to query object info from
-i, --include-detailed-usage      Include detailed usage info
-h, --help                        Show this message
```

Some sample information returned is as follows, where capabilities like number of failures to tolerate, stripe width, etc, are clearly visible:

```
Disk backing: [vsanDatastore] 8b559d56-d63b-2296-405f-a0369f56ddc0/linux-vm1.vmdk
DOM Object: 90559d56-143b-d6ac-cb00-a0369f56ddc0 (v3, owner: esxi-
hp-08.rainpole.com, policy: forceProvisioning = 0, hostFailuresToTolerate = 1,
spbmProfileId = aa6d5a82-1c88-45da-85d3-3d74b91a5bad, proportionalCapacity = 0,
spbmProfileGenerationNumber = 1, cacheReservation = 0, **stripeWidth = 1)
```

6.10 Restoring Policies

It is also possible to recover VM Storage Policies in the event of a complete vCenter Server failure. If a new vCenter must be deployed, the existing VMs can be queried, and their respective policies can be rebuilt. This is once again achievable via the Ruby vSphere Console (RVC).

```
**vsan.recover_spbm -h**
usage: recover_spbm [opts] cluster_or_host
SPBM Recovery
cluster_or_host: Path to a ClusterComputeResource or HostSystem
-d, --dry-run    Don't take any automated actions
-f, --force      Answer all question with 'yes'
-h, --help       Show this message
```

6.11 Storage Policy recommendations for VMs in stretched clusters

Use separate SPBM policies for VMs in stretched clusters

vSAN stretched clusters are an easy, fast, and flexible way to deliver cluster level redundancy across sites using a capability built right into vSphere. Since it is enabled at the cluster level, a mix of stretched clusters and non-stretched clusters can easily co-exist and all be managed by the same vCenter server.

This flexibility can lead to operational decisions in the management of SPBM policies: The rules that govern the performance and protection requirements for your VMs. vSAN stretched clusters have a few policy rules that adopt a slightly different behavior when running in stretched cluster environments. Therefore, **creating and using separate, purpose-built storage policies specifically for VMs in stretched clusters is recommended for single, and multi-cluster environments.** Let's go into more detail about this recommendation.

Two storage policy rules will be the focus of this post as we look at differences between non-stretched vSAN clusters and stretched vSAN clusters.

- " **Failure Tolerance Method** " (FTM): Defines the actual data placement, or parity method used to tolerate a failure. The FTM can be set to "RAID-1 (Mirroring)" or "RAID-5/6 (Erasure Coding)."

- " **Failures to Tolerate** " (FTT): Defines the number of failures an object can tolerate while still being accessible. Valid preset FTT values for RAID-1 object mirroring would be from 0 – 3, while RAID-5/6 supports an FTT of 1 – 2.

In any type of vSAN environment, the FTM description of "RAID-5/6 (Erasure Coding)" refers to two RAID levels. The actual RAID scheme used under this policy setting are dictated by the associated FTT level assigned in the policy. A policy setting of FTT=1 will mean that assigned VMs will use RAID-5, while an FTT=2 means that assigned VMs will use RAID-6. For the purposes of clarity, the SPBM policy names used in this post are to help explain their settings. In practice, they can be named whatever suites an environment best.

Adjusted definition for vSAN 6.6 and newer

vSAN 6.6 introduced the ability to assign an additional, secondary layer of protection when spanning a mirrored copy of data across the two sites of stretched cluster. The ability for secondary "local protection" was introduced, but needed to fit within the existing policy structure. In vSAN 6.6, there are two types of a level of failure to tolerate. There is now a "Primary Level of Failures to Tolerate" (PFTT), and a "Secondary Levels of Failure to Tolerate" (SFTT). The location they apply at in the topology depends on whether or not the stretched cluster feature is enabled on a specific vSAN cluster.

With a non-stretched cluster in vSAN 6.6, the Failures to Tolerate, or FTT is now called "Primary Level of Failures to Tolerate" or PFTT. This defines the number of failures to tolerate within a cluster at a single, local site. Just as described earlier, the PFTT in this case could have a setting 0 – 3 depending on the circumstances and FTM chosen. The options available for an FTM of a non-stretched cluster are RAID-1, and RAID-5/6. Figure 1 shows how the FTM and the PFTT are represented in a non-stretched cluster.

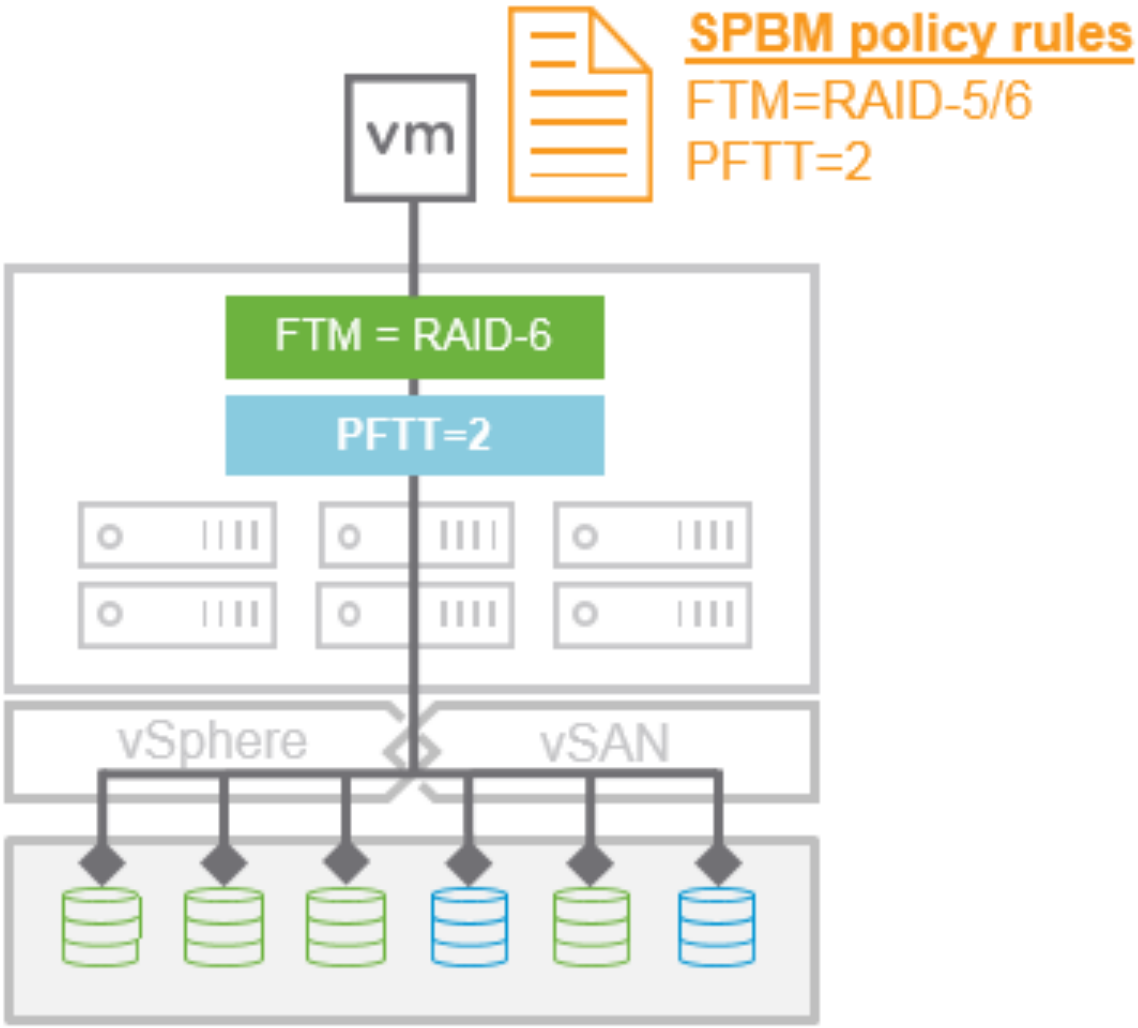


Figure 1. How assigned FTM and PFTT policy rules look in a non-stretched cluster

With a stretched cluster , the PFTT definition is different. The PFTT defines the number of failures to tolerate across the two sites, with a valid setting of 0, or 1. A setting of 0 (and paired with an "affinity" rule) is a way of setting site affinity, and would mean that it would not be protected across sites. This is a useful setting for VMs that may already have availability mechanisms at the application layer, or do not need cross-site availability.

A "Secondary Level of Failures to Tolerate" or SFTT, defines the number of failures it can tolerate within each local site. Valid preset SFTT values of for RAID-1 object mirroring would be from 0-3, while RAID-5/6 supports an FTT of 1-2. The FTM chosen in the policy setting remains as the way to define the data placement method (mirroring, versus erasure coding) used to tolerate a failure. Figure 2 shows how the PFTT and SFTT are represented in a stretched cluster.

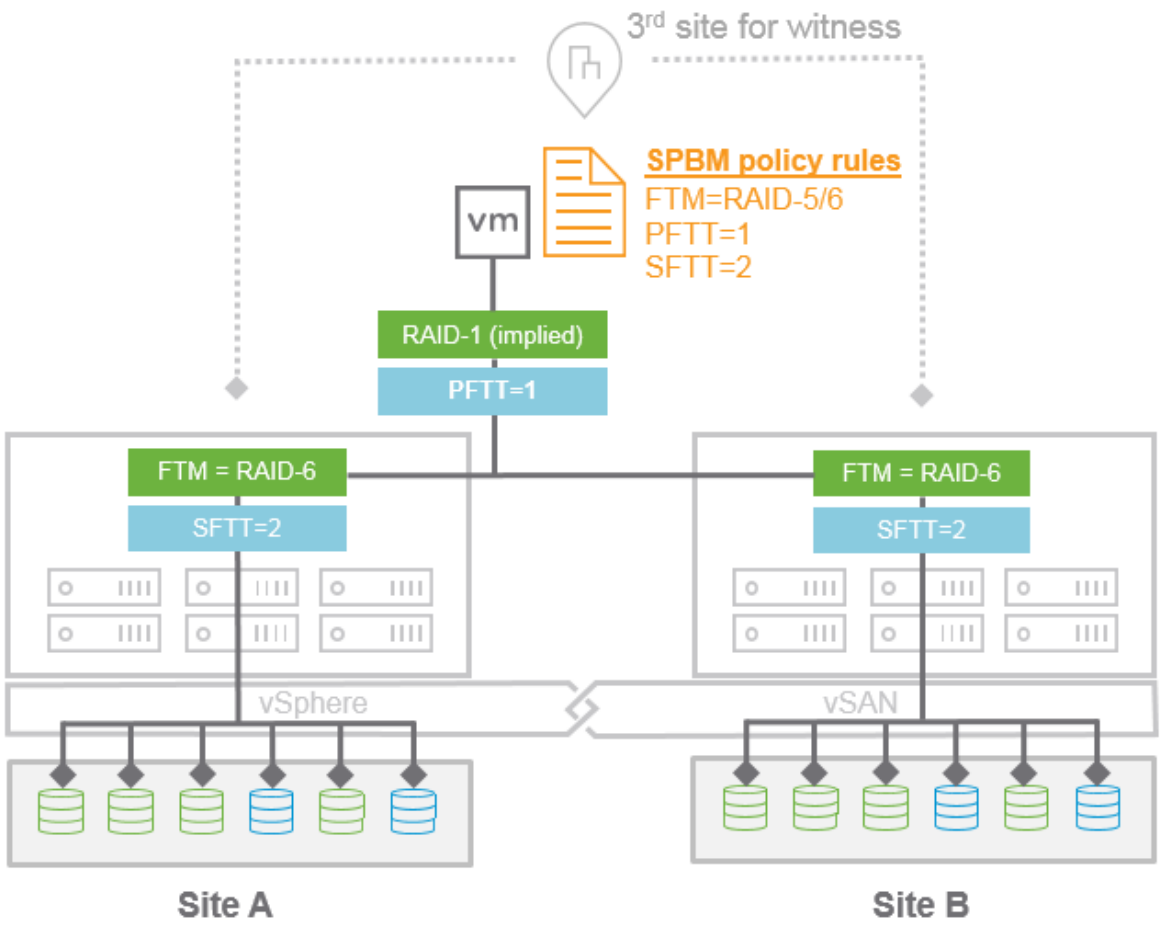


Figure 2. How assigned FTM, PFTT and SFTT policy rules look in a stretched cluster

In a stretched cluster environment, the implied FTM across sites is always a RAID-1 mirror. The FTM setting in the policy definition can be set to either "RAID-1 (Mirroring)" or "RAID-5/6 (Erasure Coding)." **In a stretched cluster, the FTM rule determines the RAID scheme used for the secondary local protection (SFTT), when it is defined.**

Behaviors when enabling or disabling stretched clusters without explicitly defined SPBM policies

How does vSAN handle VMs with non-stretched cluster specific storage policies when transitioning from a non-stretched cluster to a stretched cluster? Take a look at Figure 3.

VM	Assigned Policy	Behavior of VM while in non-stretched cluster	Behavior of VM after enabling stretch cluster
SVM-01	ForStdCluster-FTM-R1-PFTT1	RAID-1 with PFTT=1 applied to local site as expected	Protected across sites. No secondary (SFTT) local site protection applied (none defined). May show error if PFTT value was greater than 1 prior to enabling stretched clusters.
SVM-02	ForStdCluster-FTM-R5-PFTT1	RAID-5 with PFTT=1 applied to local site as expected	No secondary (SFTT) local site protection applied. UI will warn that virtual objects are non-compliant. A "Reduced Availability" status will show in vSAN Health and VM virtual object status.

Figure 3. Behavior of non-stretched cluster policies when transitioning to a stretched cluster

Once stretched clustering is enabled and configured in a vSAN cluster, the VMs do not impart any secondary, local protection logic for data placement. **The PFTT that was designated at the local site prior to enabling a stretched cluster is now set across sites.** As a result, the policy may not be able to provide compliance depending on the original policy setting.

Let's look at the behavior in the other direction, where we have VMs with stretched cluster specific storage policies, and transition from a stretched to a non-stretched cluster. Figure 4 shows this behavior.

VM	Assigned Policy	Behavior of VM while in stretched cluster	Behavior of VM after disabling stretch cluster
SVM-01	ForStretchedCluster-FTM-R1-PFTT1/SFTT1	RAID-1 mirror (PFTT) across sites with SFTT=1 applied to local site as expected	May show compliance, but will have additional artifacts and data placement arrangement from stretched cluster, such as additional witness components
SVM-02	ForStretchtedCluster-FTM-R5-PFTT1/SFTT1	RAID-1 with mirror (PFTT) across sites with RAID-5 protection scheme from SFTT=1 applied to local site as expected	May show compliance, but will have additional artifacts and data placement arrangement from stretched cluster, such as additional witness components

Figure 4. Behavior of stretched cluster policies when transitioning to a non-stretched cluster

When disabling a stretched cluster with policies built for stretched clusters, you may find some artifacts from the data placement and arrangement of the objects. vSAN is smart enough to clean this up, and will do so when you apply policies that do not have stretched cluster specific rules in them.

Now let's look at what happens when we attempt to change a VM's policy to a stretched cluster specific storage policy, even though the cluster does not have stretched clustering configured. Figure 5 details this behavior.

VM	Assigned Policy	Attempting to change VM to stretched cluster policy when stretched cluster is not enabled
SVM-01	ForStdCluster-FTM-R1-PFTT1	Will not apply storage policy. Will provide the following alert: "SFTT failures to tolerate should be 0 when stretched cluster is disabled."
SVM-02	ForStdCluster-FTM-R5-PFTT1	Will not apply storage policy. Will provide the following alert: "SFTT failures to tolerate should be 0 when stretched cluster is disabled."

Figure 5. Attempting to apply stretched cluster specific policies when cluster is not stretched

In this case, we are attempting to change a VM's policy to one specifically designed for a stretched cluster (perhaps the policy was built when stretched clustering had been enabled at in a previous scenario), while the cluster is not currently configured for stretched clusters. vSAN will not allow this. Furthermore, a policy that has these rules included in a policy will not be visible in the UI when a stretched cluster has been disabled. In vSAN 6.6 and 6.6.1, the ability to set or configure the SFTT rule or any other rule specific (e.g. "affinity" rule) to stretched clusters in a policy cannot be performed until stretched clustering is enabled.

SPBM policy recommendations for stretched clusters

The easiest way to accommodate a mix of stretched, and non-stretched vSAN clusters is to have **separate policies for stretched clusters** . You could have policies that are exclusive to that specific vSAN stretched cluster, or build stretched cluster specific policies that could be applied to multiple stretched clusters. Based on the topology, a blend of both strategies might be most fitting for your environment. Perhaps cluster specific policies for larger purpose-built clusters, along with a single set of policies for all smaller branch offices. Additional policies can easily be created by cloning existing SPBM policies, modifying accordingly, then assigning to the appropriate VMs. Having multiple policies for VMs in stretched and non-stretched clusters is also good for a single cluster environment where you need to tear down and recreate the stretched cluster.

Adjusting existing policies will always impact all VMs that are using the adjusted policy, whether they live in a stretched cluster, or non-stretched cluster. Adjustments in this scenario could introduce unnecessary resynchronization traffic when an administrator is trying to remediate an unexpected policy condition. This is another reason why dedicated SPBM policies for VMs running in stretched clusters are recommended.

Summary

vSAN stretched clusters use SPBM to provide extraordinary levels of flexibility and granularity for any vSAN environment, and is one of the staples behind vSAN's ease of use. Using separate polices for VMs in stretched clusters is a simple operational practice that can help virtualization administrators become more comfortable with introducing and managing one or more stretched clusters in a vSAN powered environment.

6.12 Storage Policy Naming Considerations

Practical guidance for storage policy naming

The use of naming conventions in the data center can play an important part in operational efficiency. Naming conventions are most often used for devices such as hosts and switches, along with other entities such as VMs, virtual switches, and Microsoft Group Policy Objects. Choosing a consistent approach to naming can reduce time for change requests, and minimize operational mistakes as an environment grows.

Determining the right approach for naming is often an exercise in tradeoffs. Overly simplistic standards can result in ad-hoc naming that lacks consistency. Overly complex standards often fail because governance becomes too difficult. The ideal naming convention is one that addresses the needs of the organization, is descriptive and flexible, all while maintaining simplicity.

How naming conventions can help with storage policies

Storage Policy Based Management (SPBM) can also benefit from naming conventions. When used with vSAN or VVols, the SPBM framework takes what used to be all-or-nothing settings defined on external storage, and allows the administrator to assign storage policies to individual VMs. The administrator can easily specify performance, availability, and space efficiency settings on a per VM, or even per VMDK basis to accommodate application requirements. Policies can be named in whatever way that suites an organization best.

Depending on the need, an environment may require just a few storage policies, or dozens. Before deciding on an approach that works best for your organization, let's review a few characteristics of storage policies with SPBM.

- A maximum of 1024 SPBM policies can exist per vCenter server.
- A storage policy is stored and managed per vCenter server, but can be applied to VMs in one or more clusters.
- A storage policy can define one or many rules around performance, availability, space efficiency, etc.
- Storage policies are not additive. Only one policy (that contains one or more policy rules) can be applied per object.
- A storage policy can be applied to a group of VMs, a single VM, or even a single VMDK within a VM.
- A storage policy name can consist of up to 80 characters.
- A storage policy name is not the true identifier. Storage policies use a unique identifier for system management.

With a high level of flexibility, users are often faced with the decision of how best to name policies, and apply them to their environments.

Balancing descriptiveness and simplicity

Policy names are most effective when they include two descriptors: intention, and scope.

- **Intention** refers to what the policy aims to achieve. Perhaps the goal of the policy is to apply high performing mirroring using a Failure Tolerance Method (FTM) of RAID-1, with an increased level of protection by using a Level of Failures to Tolerate (FTT) of 2.
- **Scope** refers to where the policy will be applied. Maybe a policy is for a farm of servers hosting the company ERP solution. or perhaps it is for just the respective VMDKs holding databases in a specific vSAN cluster.

To improve readability of policy names, you may wish to associate specific terms to a policy to indicate their settings. For instance, "BasicProtect" may be associated with all policies using an FTT of 1, while "EnhancedProtect" might be associated with all policies using an FTT of 2. This type of approach can also be used for Performance ("EnhancedPerf" associated with RAID-1 and a stripe width of 4) along with capacity consumption ("SpaceEfficient" associated with RAID-5/6). The naming flexibility lets you decided on how much technical detail you wish to expose in the name to administrators, application owners, or automation teams interacting with policies.

Determining the realistic needs of the organization is an important step in finding the best storage policy naming conventions for an environment. A few questions to ask yourself might include:

- What is the size of the environment?
- Are there multiple clusters? If so, how many?
- Are there stretched clusters?
- Is there a preference to indicate actual performance/protection settings within names, or adopt a gold/silver/bronze approach?
- Is there a need for application specific storage policies?
- Is there a need for VMDK specific storage policies?
- What type of delimiter will work best? Spaces? Hyphens? Periods? What will work best in conjunction with scripting?
- Are there specific departments, or business units that need representation in a storage policy name?

Who is the intended audience? Virtualization Administrators? Application Owners? Automation teams? This can have an impact on the level of detail you wish to provide in a policy name.

The answer to these questions will help determine how you might want to name storage policies, and the level of sophistication to a naming convention used.

Examples

The following examples show how different approaches can be used to simplify the management of storage policies. These are only examples. **Storage policy names can be as simple or descriptive as you wish them to be.** The flexibility of storage policies allows you to decide how best to apply them for your conditions. The intention with these examples are to inspire a way for you to name storage policies that suites your organization best.

Example 1: The example below shows a storage policy that could be used for a collection of general purpose workloads. One can optionally provide additional setting indicators of the policy for readability. Specifying a cluster name is optional, but might be helpful for environments that have multiple clusters, or stretched cluster environments that will be best served by dedicated policies.

Naming structure:

[ClusterName]-[Workloads]-[IntentionofPolicy]-[OptionalSettingIndicators]

Functional examples:

Cluster01-ManagementVMs-BasicProtectionEnhancedPerf

MultiCluster-ProductionVMs-BasicProtectionEnhancedPerf

MultiCluster-ProductionVMs

Example 2: The example below shows a storage policy that could be used for application specific workloads. They could be assigned to one or more VMs that provide the services of the application. This type of approach would be most appropriate for environments with targeted performance and protection SLAs that could not be met by a general policy as shown in Example 1.

Naming structure:

[ApplicationName]-[IntentionofPolicy]-[OptionalSettingIndicators]

Functional examples:

App-SharePointWebFarm-BasicProtectionSpaceEfficient

App-SharePointSQLBackEnd-EnhancedProtectionEnhancedPerf

App-VDIPoolX-BasicProtectionEnhancedPerf

App-VDIPoolX

Example 3: The example below shows a storage policy to define performance and protection settings for a department, business unit, or predefined service tier.

Naming structure:

[Department]-[IntentionofPolicy]-[OptionalSettingIndicators]

Functional examples:

Group-DevelopomentCodeCompling-BasicProtect-UltraHighPerf

Group-CustomerX-BasicProtectionSpaceEfficient-ValueTier

Group-CustomerX

Example 4: The example below shows a storage policy that could be used for performance characteristics of a targeted VMDK type, in a single VM, or multiple VMs. In this case, the policy rules will limit the number of IOPS for all VMDKs using this policy so that they do not interfere with the performance of other VMs (the noisy neighbor phenomenon).

Naming structure:

[ApplicationorWorkloadName(s)]-[IntentionofPolicy]-[OptionalSettingIndicators]

Functional examples:

App-LogAnalyticsCollectorVMDKs-1500IOPSCap

App-LogIndexingVMDKs-1000IOPSCap

App-LogIndexing

In Figure 1, we see how the examples above look when viewed in vCenter. The examples represent policies that are applied across large collections of VMs, specific departments with an organization, as well as targeted applications. Naming standards make them easier to understand and manage.

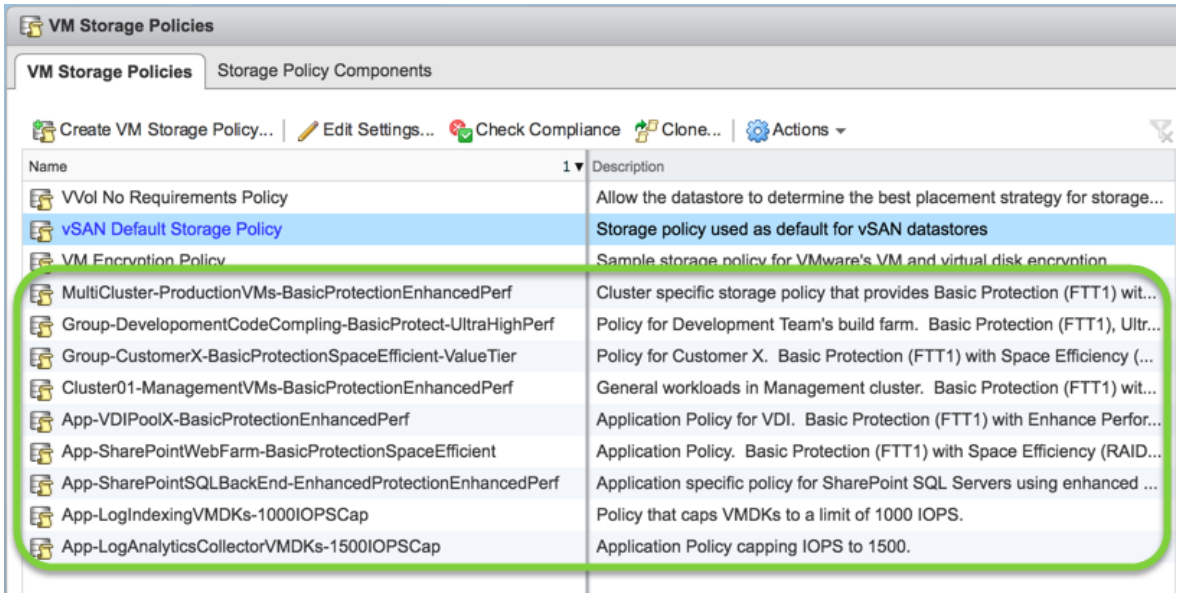


Figure 1. Example policies added in vCenter

More policies could easily be created that address requirements such as site affinity, or workloads with application redundancy. You may also find the " vSAN Storage Policy Example Creation " script created by Jared Lutgen on code.vmware.com helpful.

Other tips

Here are a few other operational tips related to storage policies.

- **Do not attempt perfection of a naming standard.** Try what works, and adjust as necessary. There is no right or wrong way to name policies.
- **Avoid using and changing the "vSAN Default Storage Policy."** Create and clone storage policies as needed.
- **Use the "Description" field in a storage policy.** This is a great way to define what the intention and scope of a policy is for, and assist in self-documentation efforts.
- **Find the right balance of descriptiveness in a policy name.** Short names are easy to consume, but may not be descriptive enough. Naming conventions that are too cryptic often do a good job of keeping the name length to a minimum, but do little to help the user. Be mindful that a policy name is limited to 80 characters.
- **Move VMs to another policy instead of changing existing policy rules.** Changing existing policy rules can introduce a large amount of resynchronization traffic, and immediately applies to all VMs using that policy. Some settings like a change in the FTM, or stripe width can generate a lot of data movement. By moving VMs to a new policy, resynchronization traffic will be limited to those VMs that are assigned a new policy.
- **Unsure of what you need? Start with generic storage policies, then add application specific policies as needed.** This is a way to achieve a mix of policies that can address groups of VMs with undetermined storage requirements, while adding custom policies intended for a single application, or the collection of VMs that make up an application.

Conclusion

As data centers continue to move toward a more elastic, software defined model, control of that infrastructure will be through policy engines like SPBM for vSAN and VVols, as well as what is found in

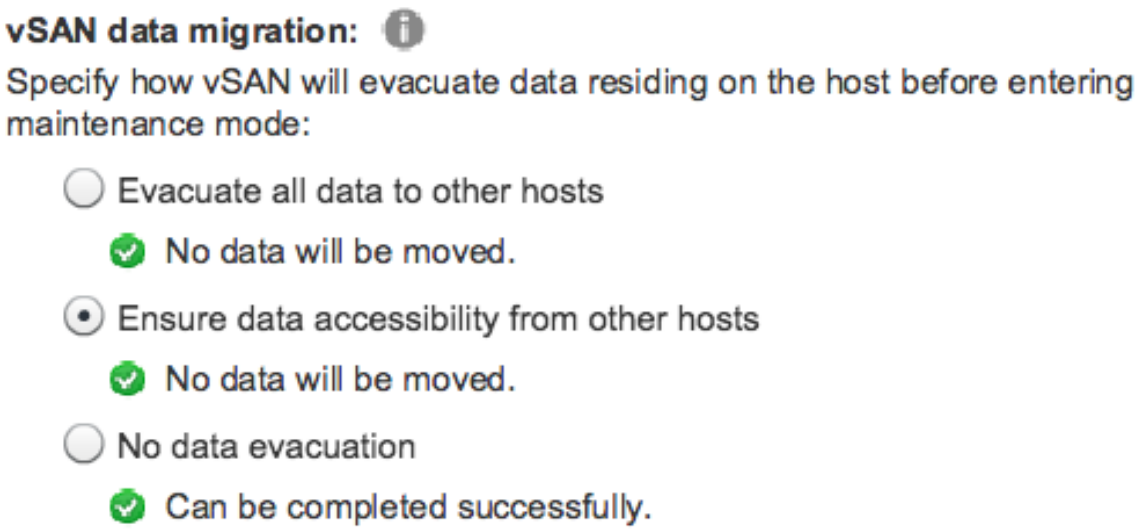
solutions like NSX. An administrator has tremendous flexibility in determining what policies are applied, where they are applied, and how they are named. Having an approach to naming conventions for policies that drive the infrastructure will allow you to make changes to your environment with confidence.

7. Maintenance Mode Operations

When maintenance needs to be performed on an ESXi host it is recommended to put the host in maintenance mode.

7.1 Enter Maintenance Mode

Let's start with a brief explanation of vSAN data migration options for maintenance mode. This diagram shows the options in the vSphere Web Client.



Evacuate all data to other hosts

This option moves all of the vSAN components from the host entering maintenance mode to other hosts in the vSAN cluster. This option is commonly used when a host will be offline for an extended period of time or permanently decommissioned.

Ensure data accessibility from other hosts

vSAN will verify whether an object remains accessible even though one or more components will be absent due to the host entering maintenance mode. If the object will remain accessible, vSAN will not migrate the component(s). If the object would become inaccessible, vSAN will migrate the necessary number of components to other hosts ensuring that the object remains accessible. This option is the default and it is commonly used when the host will be offline for just a short amount of time, e.g., a host reboot. It minimizes the amount of data that is migrated while ensuring all objects remain accessible. However, the level of failure tolerance will likely be reduced for some objects until the host exits maintenance mode.

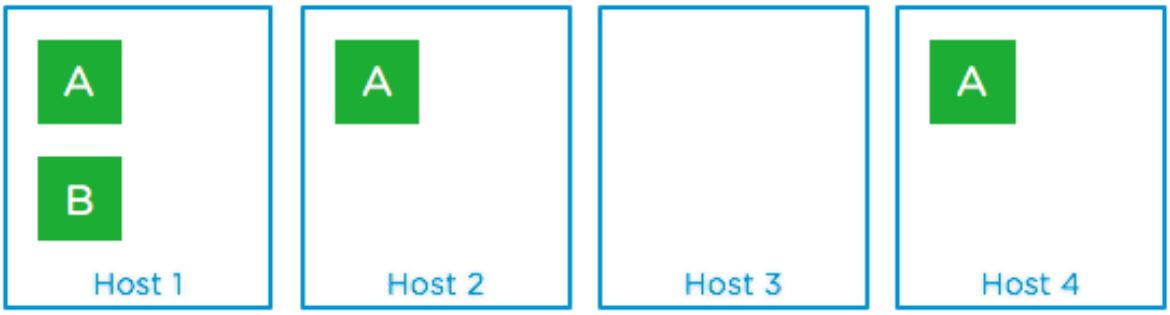
No data evacuation

Data is not migrated from the host as it enters maintenance mode. This option can also be used when the host will be offline for a short period of time. All objects will remain accessible as long as they have a storage policy assigned where the Primary Level of Failures to Tolerate is set to one or higher.

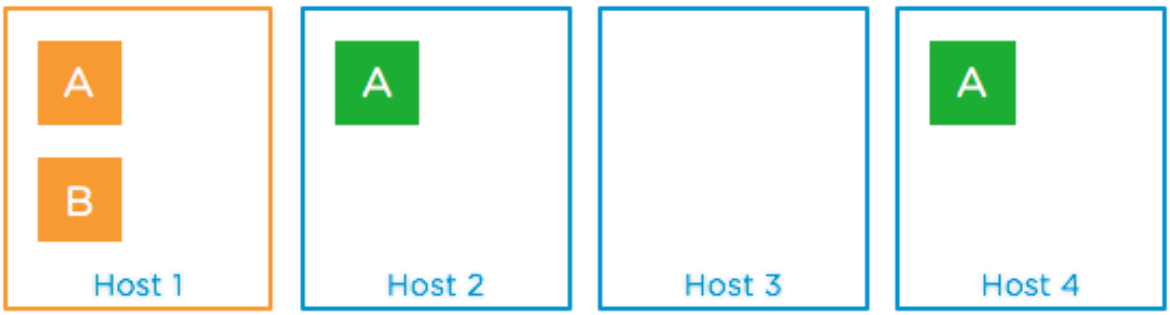
Perform the following steps to put a host into maintenance mode:

1. Click Hosts and Clusters.
2. In the Navigator column, right-click the host you wish to put into maintenance mode.
3. Select Maintenance Mode > Enter Maintenance Mode.
4. Click the desired vSAN data migration option.
5. Review the "what-if" information.
6. Click OK.

If an object has a storage policy where the Primary Level of Failures to Tolerate (PFTT) is set to zero, using the No Data Evacuation option might cause the object to become inaccessible. The diagrams below illustrate why. Object A has PFTT = 1 (mirroring). Object B has PFTT = 0.

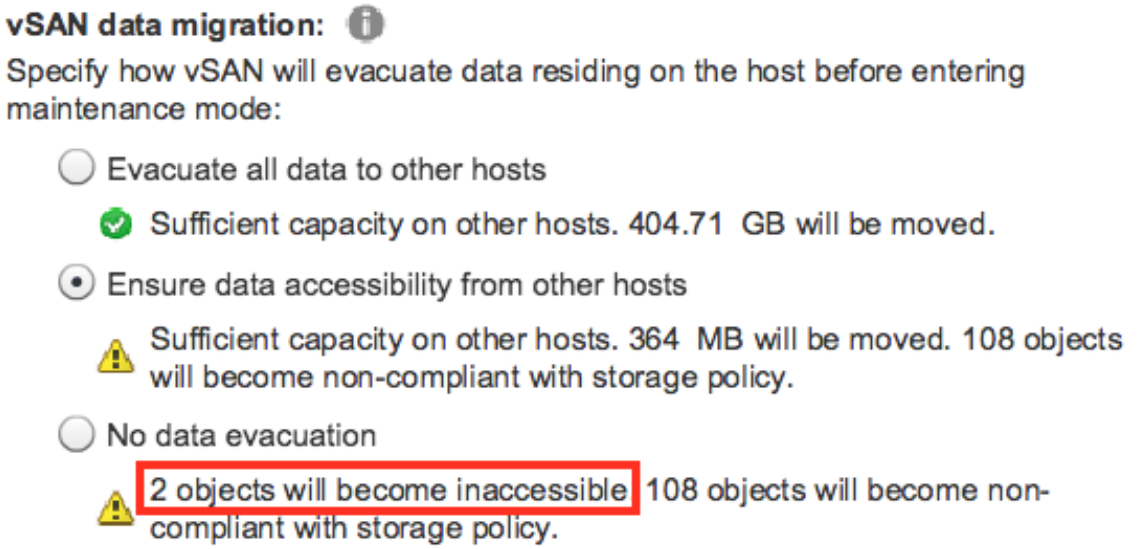


If we put Host 1 into maintenance mode using the No Data Evacuation option, Object A remains accessible. Object B becomes inaccessible as the only copy of that data is on Host 1.



You might be wondering if data loss with Object B occurred when the host entered maintenance mode. The answer is no – the data is still on the host, but it will not be accessible until the host exits maintenance mode.

Recommendation: Pay close attention to the "what-if" information displayed in the Maintenance Mode UI. It will tell you if objects will become inaccessible as a result of putting the host into maintenance mode. An example is shown below.



To avoid objects becoming inaccessible, select the Ensure Data Accessibility From Other Hosts option. vSAN will migrate the components required to keep the objects accessible to other hosts. You can also assign a storage policy to the objects where PFTT is set to one or higher. Just be sure all objects are in compliance with their storage policies before continuing with the No Data Evacuation option.

Ensure Accessibility

vSAN ensures that all virtual machines on this host will remain accessible if the host is shut down or removed from the cluster. Only partial data migration is needed. This is the default option.

Full Data Migration

vSAN migrates all data that resides on this host. This option results in the largest amount of data transfer and consumes the most time and resources. It also ensures that all virtual machines are still compliant with their selected policy.

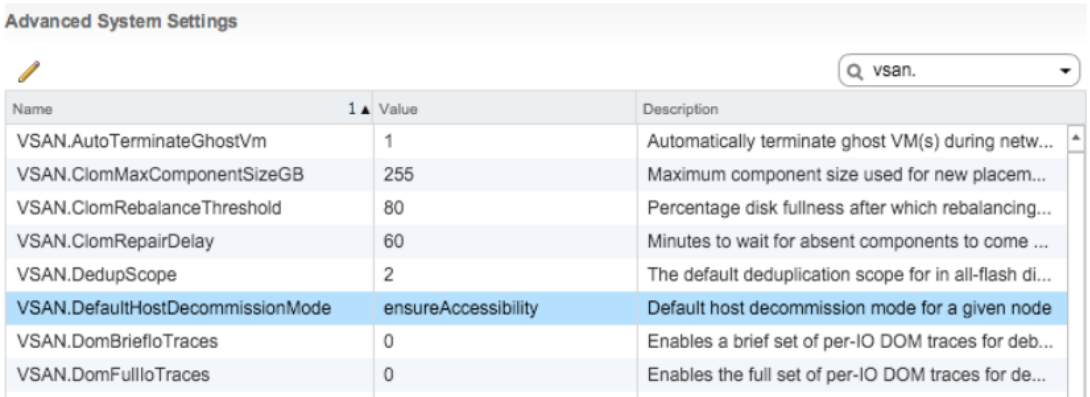
No Data Migration

vSAN will not migrate any data from this host. Some virtual machines might become inaccessible if the host is shut down or removed from the cluster. Do not use this unless there is no other option. There is a risk of data loss using this option.

7.2 Set Default Maintenance Mode Operation

To change the default Maintenance Mode Operation on a vSAN cluster the following steps should be taken.

1. Open the vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the *cluster* on which you want change the default maintenance mode operation.
4. Select the first host in the cluster.
5. Click the **Manage** tab.
6. Click **Settings**.
7. Click **Advanced System Settings**
8. Filter on "vsan."
9. Select VSAN.DefaultHostDecommissionMode entry



10. Change entry to either of the following three options, where the first option is the default:
    - ensureAccessibility
    - evacuateAllData
    - noAction

NOTE: When "noAction" is selected it could result in data loss as even when VMs only have a single copy of their data stored on the vSAN datastore the data will not be migrated.

8. Host Operations

In this section, the most common host operations related to vSAN are discussed.

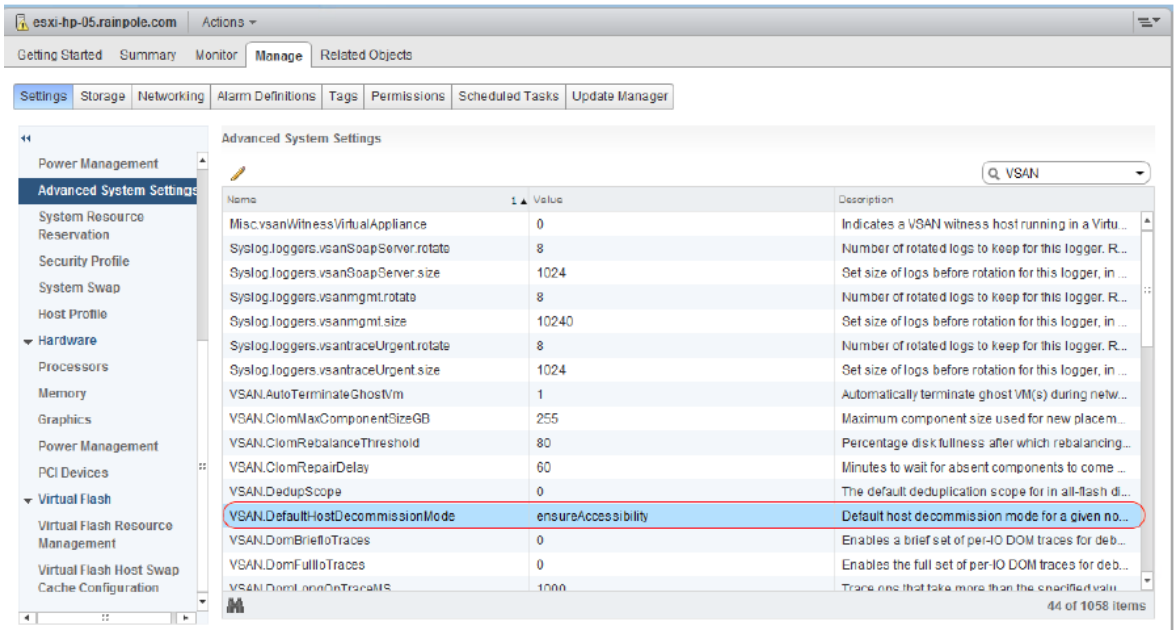
8.1 Patching and Updates of Hosts

Patching and Updates of hosts (e.g. VUM) in a vSAN cluster

Patching and updating ESXi hosts in a vSAN cluster requires some additional consideration. In order for the virtual machines to remain fully available and have no risk, each host in the cluster would need to be placed into maintenance mode, have its data evacuated, upgraded/updated applied, rebooted, and on a successful reboot, the host is then taken out of maintenance mode and can rejoin the cluster. This then has to repeated for all the ESXi hosts in the cluster.

One should note that the default maintenance mode/decommission mode used by VUM (vSphere Update Manager) when dealing with vSAN hosts was "Ensure Accessibility". This meant that a host could be placed in maintenance mode even when virtual machines only have one copy of the data available in a RAID-1 mirrored configuration. This therefore meant that there was some risk to virtual machine availability, should a failure occur while this host was in maintenance mode.

In vSAN 6.1 and later, there is a new advanced option which allows administrators to set the maintenance mode/decommission mode. The option is called VSAN.DefaultHostCommissionMode, and can be found in the Advanced System Settings on each host. By changing this to "evacuateAllData", VUM will now ensure that each host is fully evacuated when updating an ESXi host that is a member of a vSAN Cluster.



The complete set of options for this advanced parameter are shown in the table below:

VSAN Decommission Mode Value	Description
ensureAccessibility	vSAN data reconfiguration should be performed to ensure storage object accessibility
evacuateAllData	vSAN data evacuation should be performed such that all storage object data is removed from the host
noAction	No special action should take place regarding vSAN data

Note that the advanced option needs to be set identically on all hosts on the cluster.

The advanced setting can also be set at the command line. To configure the default vSAN maintenance mode option using ESXCLI, run the following command:

```
esxcli system settings advanced set -o /VSAN/DefaultHostDecommissionMode -s
<DECOMISSION_MODE>
```

8.2 Configuring Log Locations

This is a conversation that comes up regularly again. The main consideration related to what sort of device is used for booting the ESXi host that is participating in vSAN. Is it booting from a regular HDD, from an SD/USB device or from a SATADOM.

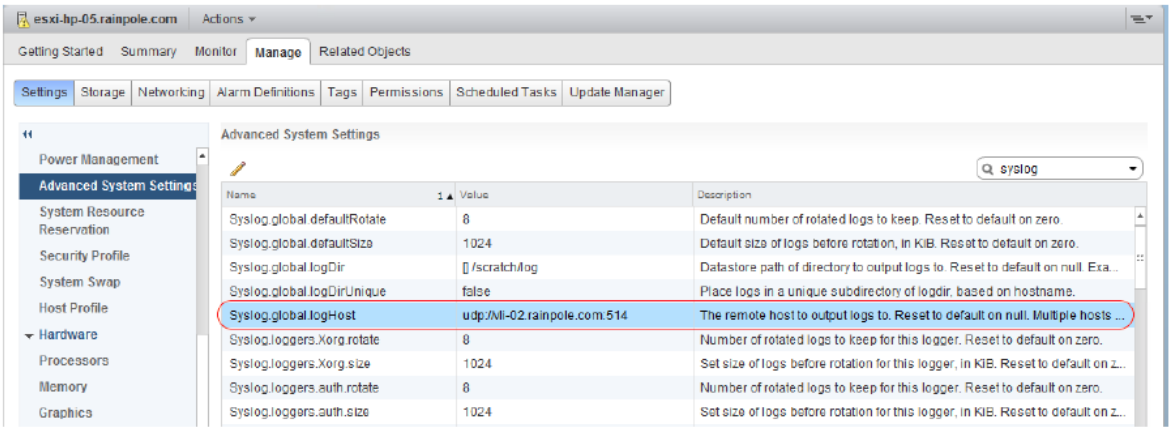
If the ESXi host is booting from an HDD or a SATDOM (which looks like a HDD), then there are different partition layouts and different considerations for logging and tracing when compared to an ESXi host that is booting from a USB/SD device.

In a nutshell, when ESXi is booting from SD/USB, then RAMdisks are used for both vSAN traces and log files. This is to prevent burnout of the SD/USB device, which historically did not have high endurance. When the host is shutdown (gracefully or in a uncontrolled manner), the contents in the RAMdisks are stored on the USB/SD device. However, due to the finite space on the USB/SD device, it is not always possible to capture all the logs and log file contents.

There are a number of advanced options and configuration steps which can help, both for hosts that boot from UDB/SD, and from HDD/SATADOM. These are covered here in detail.

Configuring syslog

Many customers use a dedicated syslog server, such a vRealize Log Insight, for capturing and storing all of the logs from their ESXi hosts. ESXi hosts that participate in a vSAN cluster are no different and can redirect their logs to a remote host. This is done via an advanced setting called *Syslog.global.logHost* and should be done on each host in the cluster:



Note that it is not supported at this time to send syslog output to a vSAN datastore. Always worth noting in the above screen shot is *Syslog.global.logDir.* This is pointing to the scratch location on this host which is booted from a USB device, so scratch, in this case, is a RAM disk. In this example, logs are being sent to both scratch partition and syslog host. Regardless of the additional syslog configuration specified using the options above, logs continue to be placed in the default locations on the ESXi host.

Further information regarding configuring syslog, see vSphere 6.5 Configure Syslog on ESXi Hosts. For additional information see KB 2003322.

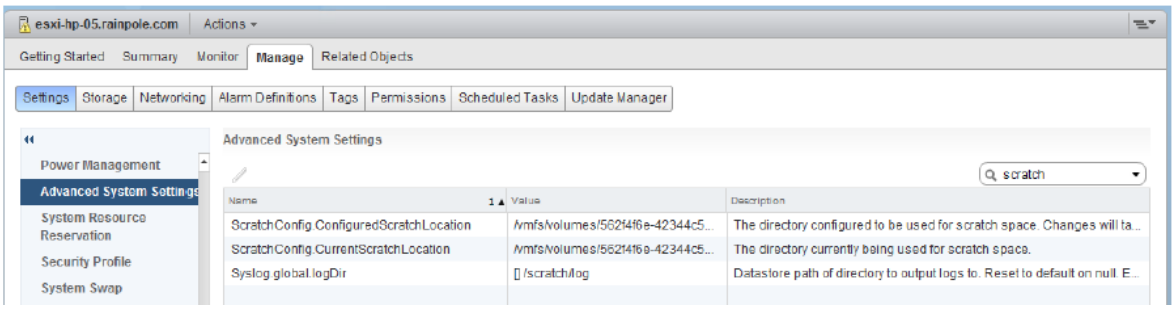
Configuring netdumper

Rather than dumping cores on local storage of the ESXi host, vSphere provides a mechanism called netdumper to transfer core files to a location outside of the ESXi host. It is a post crash feature that sends the core dump "unreliably" over a UDP connection. Unfortunately, this tool does have some limitations, as one transmission failure will result in a failed core dump collection, and thus there will be no core dump for root cause analysis.

Details on how to configure the netdumper can be found in the Managing Core Dumps section of the vSphere 6.5 Command Line Reference Guide.

Configuring scratch

Once again, if booting from an SD/USB device, the "scratch" location where temporary files and log files are stored, is a RAMdisk. It might be desireable to redirect the scratch to a persistent storage device rather than a RAMdisk. Once again, this would need to be done individually on a host by host basis. Once the advanced setting ScratchConfig.ConfiguredScratchLocation has been updated, the ESXi host would need to be rebooted for the change to take effect. After the reboot, both ScratchConfig.ConfiguredScratchLocation and ScratchConfig.CurrentScratchLocation should match. The advanced option for scratch location is shown in the screenshot below.



Note that it is not supported at this time to place scratch on a vSAN datastore.

Configuring vSAN traces

This is another significant considerations for logging, and is once more dependent on whether the ESXi host is booted from USD/SD or HDD/SATADOM. Let's start with ESXi hosts that are booting from either USB sticks, or SD cards. I'm grouping these together since the considerations are more or less the same from a vSAN trace perspective. As outlined earlier, when an ESXi host that is booting from one of these devices is also running vSAN, vSAN traces are written to a RAM disk. Since the RAM disk is non-persistent, these logs are written to persistent storage either during host shutdown or on system crash (PANIC). This means that the vSAN traces, which are typically quite write intensive, do not burn out the boot media. This method of first writing the traces to RAM disk and later moving them to persistent store is handled automatically by the ESXi host and there is no user action required. This is the only support method of handling vSAN traces when booting an ESXi from either a USB stick or an SD card. You cannot write vSAN traces directly to SD or USB boot devices at this time.

This is not such a concern when booting ESXi from HDD or SATADOMs. SATADOMs, short for Serial ATA Disk on Modules, are basically flash memory modules designed to be inserted into the SATA connector of a server. In vSAN 6.0 and later, vSAN supports ESXi hosts booting from SATADOM, as long as they met specific requirements. On ESXi hosts that boot from SATADOM, the vSAN traces are written directly to the SATADOM. In other words, there is no RAM disk involved. This is why specification requirements for SATADOM are documented in the vSAN Administration Guide, and the requirement is for an SLC (single level cell) device. The SLCs have higher endurance and quality when compared to other flash devices. The reason for this is once again to prevent any sort of burn-out occurring on the boot device when trace files are being written to it.

With the release of vSAN 6.2, it is now possible to send urgent vSAN traces to syslog. In fact this feature is now on by default. It is also possible to redirect vSAN traces to a persistent storage such as an NFS. This can only done via the ESXCLI however; there is no advanced option to redirect vSAN Traces. Here is an example of using the "get" parameter to display the current settings. The parameter "set" can be used to change any of this.

```
[root@esxi-hp-05:~] esxcli vsan trace get
VSAN Traces Directory: /scratch/vsantraces
Number Of Files To Rotate: 8
Maximum Trace File Size: 180 MB
Log Urgent Traces To Syslog: true
```

8.3 Improving Visibility of Host Restarts

Use Out-of-Band Management to View vSphere DCUI During Host Restarts

Visibility to vSphere hosts through out-of-band management has always been a convenient way to access the vSphere Direct Console User Interface (DCUI). Traditionally, remote management using Intelligent Platform Management Interface (IPMI) or some other method, provided the pre-boot visibility and control necessary to update firmware on the host, and see the host state if it was not accessible using traditional, in-band methods. 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.

Yet, the DCUI access can play an important role for administering a vSAN environment, which is why **incorporating out-of-band console visibility into your operational practices is recommended**. Let's look a bit more as to why this practice makes sense.

A host in a vSAN based cluster has additional actions to perform during the host reboot process. Many of these additional tasks during a host reboot simply ensure the safety and integrity of data. Looking at the DCUI during a host restart will reveal a few vSAN related activities. The most prominent message, and perhaps the one that may take the most time is "*vSAN: Initializing SSD… Please wait…*" similar to what is shown in Figure 1.

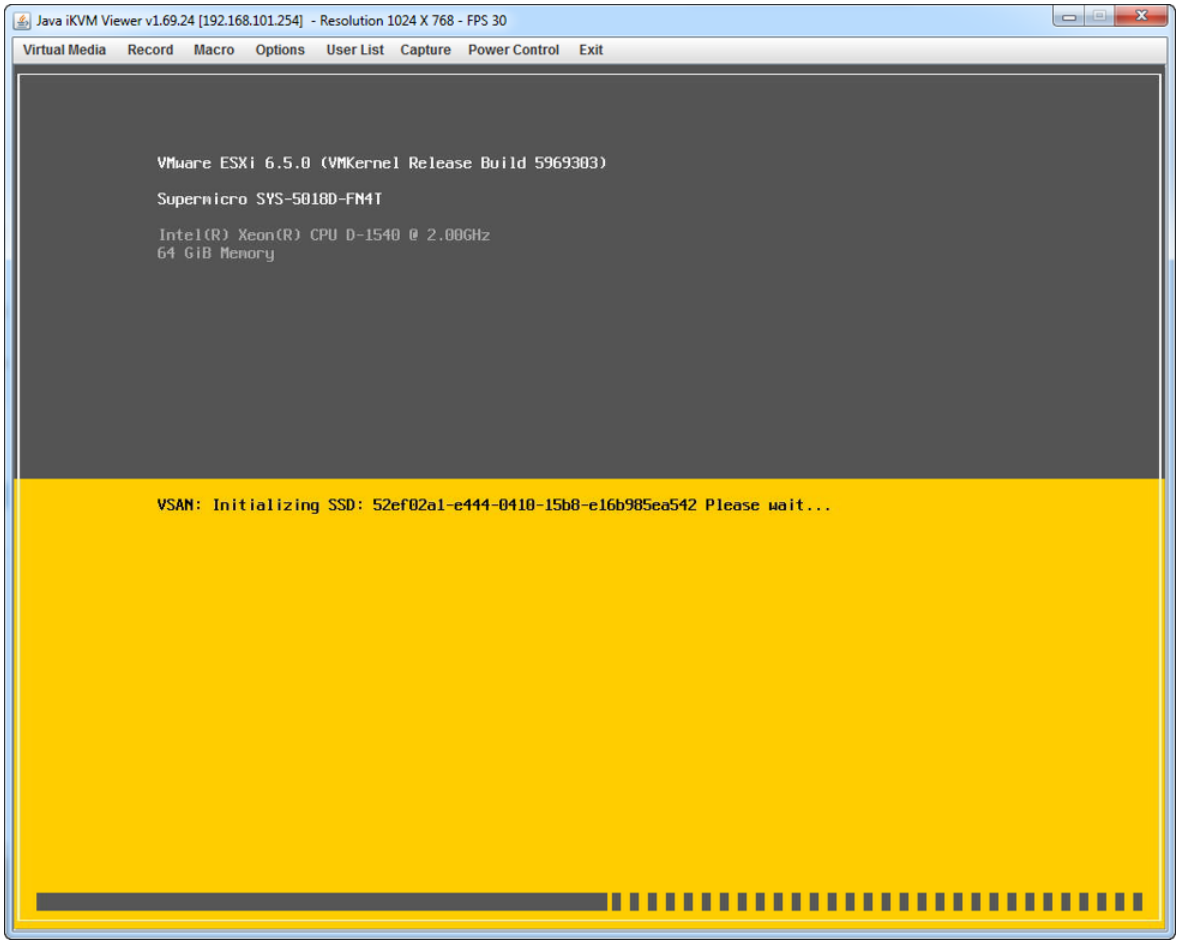


Figure 1. DCUI showing the "Initializing SSD" status.

During this step, vSAN is processing data, and digesting the log entries in the buffer to generate all required metadata tables. More detail on a variety of vSAN initialization activities can be exposed by hitting ALT + F11 or ALT + F12 in the DCUI, as shown in Figure 2.

```
2017-09-18T18:05:20.056Z cpu6:65920)VSAN: Initializing SSD: 52ef02a1-e444-0410-15b8-e16b985ea542 Please wait...
2017-09-18T18:05:20.057Z cpu12:66981)PLOG: PLOGNotifyDisks:4495: MD 0 with UUID 522e3049-9c56-d398-cab6-af9342573d45 with state
2017-09-18T18:05:20.057Z cpu12:66981)VSANServer: VSANServer_InstantiateServer:2063: Instantiated VSANServer 0x430504d0d198
2017-09-18T18:05:20.059Z cpu13:66631)Created VSAN Slab RcSsdParentsSlab_0x430aa47fbba0 (objSize=208 align=64 minObj=2500 maxObj=
2017-09-18T18:05:20.060Z cpu13:66631)Created VSAN Slab RcSsdIoSlab_0x430aa47fbba0 (objSize=65536 align=64 minObj=64 maxObj=25000
2017-09-18T18:05:20.060Z cpu13:66631)Created VSAN Slab RcSsdMdBElemSlab_0x430aa47fbba0 (objSize=32 align=64 minObj=4 maxObj=4096
2017-09-18T18:05:20.060Z cpu13:66631)Created VSAN Slab RCInvBmapSlab_0x430aa47fbba0 (objSize=56 align=64 minObj=1 maxObj=1 overh
```

Figure 2. Detailed log entries during "Initializing SSD" state.

The specific activities that relate to the duration of the "Initializing SSD" activity are the Physical Log, and Object Manager entries. You might see entries such as:

SSDLOGLogEnumProgress:948: Estimated time for recovering 712459 log blks is 95221 ms
PLOG_Recover:970: Doing plog recovery on SSD
PLOGRecDisp:988: PLOG recovery complete

Entries like the examples shown above are a normal part of this "Initializing SSD" step, and show that vSAN is making progress in the processing of this data. Since each vSAN host contributes to the overall storage footprint available to the VMs, the processing and reconciliation of data during a restart is expected behavior of a host in a vSAN environment.

This means that hosts in a vSAN cluster can take longer to reboot than non-vSAN hosts. The message may appear only momentarily on the DCUI screen, or it may take several minutes per disk group to complete this step and proceed with the remainder of the host reboot.

This task may fail if there is an underlying issue with SSD, or perhaps when one is using a storage controller not on the HCL. In rare circumstances, long periods of time can also indicate possible health issues with some metadata associated with components that make up an object in vSAN. Metadata health can be easily viewed in the UI using the vSAN Health check service.

The time this initialization actually takes depends on a number of factors. One of the primary variables is the amount of data, or blocks, that are in the write buffer at the time of the host restart. **During this "initializing SSD" period, further reboots of hosts in this state should be avoided.** Having out-of-band access to the host DCUI is one of the best ways to provide proper visibility, and avoid unnecessary, additional host restarts during these moments where it is performing tasks, but not available to the cluster.

DCUI accessibility via remote management should also be incorporated into defined maintenance workflows such as host restarts. It doesn't mean that an administrator needs to watch the DCUI every time they restart a host. The objective would be to 1.) adjust expectations on what typical restart times are for a vSAN host during the reboot process. 2.) Instill a good operational practice that in the event that the status of a host is uncertain during a normal host restart, how should the administrator proceed as a next step.

While some operational practices for vSAN are noticeably different than a traditional infrastructure, many practices, like this one, are a simple reminder of best practices for almost any environment. Overly anxious, hard resets of hosts have never been ideal, but unfortunately these types of practices become a customary troubleshooting step for many organizations. By understanding how vSAN can potentially change the boot time of a host, this helps bring to light the importance of adhering to proper operational procedures.

9. vCenter Operations

Although vSAN is fully integrated in the vSphere Web Client, there is no direct dependency on the availability of vCenter Server itself when it comes to how vSAN functions.

9.1 vCenter Operations

Although vSAN is fully integrated in the vSphere Web Client, there is no direct dependency on the availability of vCenter Server itself when it comes to how vSAN functions. A vCenter Server can even be fully replaced with a new vCenter Server instance if desired and vSAN will keep functioning. There are some considerations around policy management, but those are covered in the policy section.

9.2 Updating vCenter in a vSAN Cluster

When it comes to updating vCenter Server there are no considerations for vSAN. We highly recommend making a backup of vCenter Server before upgrading, and / or exporting your VM Storage Policies and when applicable your Distributed Switch configurations.

9.3 Certificates

When the default vCenter Server certificates are replaced, the Health Check may be unavailable. To ensure the health check functions, see KB 2133384. It includes details on the exact problem and the steps to solve it.

9.4 Moving a vSAN Cluster

In some cases instead of upgrading a vCenter Server instance it may be desired to deploy a new vCenter Server instance. How do you do this when using vSAN? The steps are straight forward:

1. In the Web Client for the new vCenter Server instance, create a new HA / DRS / vSAN cluster (Creating a vSAN Cluster).
2. Add the first *host* from your old cluster to the new cluster.
3. Wait until it is configured, an error will pop up that says " *Misconfiguration detected* ". This is expected as you only have 1 host in your cluster.
4. Now add the rest of the hosts one by one to the new cluster.
5. After completing the full migration, the *Misconfiguration Detected* error should be gone.

Note that the policies will need to be exported and imported, details around this can be found in the VM Storage Policies section

For more information on introducing a new vCenter server to an existing vSAN cluster, see Replacing a vCenter server for existing vSAN hosts in this operations guide.

9.5 Replacing a vCenter Server for existing vSAN hosts

Replacing a vCenter server for existing vSAN hosts

Since VMware vCenter is used as a common control and management plane for a vSphere cluster, questions may arise when determining how a vSAN cluster reacts when a vCenter server must be rebuilt from a new installation, or restored from a backup. Accounting for unplanned events is always a top-of-mind concern for data center administrators. The scenario shown below helps describe how vSAN, and vCenter behaves when adding hosts previously participating in a vSAN cluster to a new, pristine vCenter server.

While vCenter plays an important role in the interactive management of a vSphere cluster, vSAN is sufficiently decoupled from vCenter to ensure continued operation if vCenter is offline, or rebuilt from a new installation. vCenter has never been responsible for any data path, or object management activities with vSAN.

vSAN 6.6 transitioned from the use of multicast to unicast for all host membership activities, and under the new architecture, maintains this in vCenter, as well as distributing the membership information across the hosts in the cluster. The vCenter authority health check is a new health check introduced to vSAN 6.6.1 to verify host membership consistency between vCenter and the hosts in the cluster. This health check will check for, and remediate any consistency issues that are seen with host membership and settings. These health checks can be especially useful in scenarios that include adding hosts to a vSAN cluster during periods in which vCenter was offline, restoring an older vCenter server from backup, and creating a new installation of a vCenter Server for an existing vSAN cluster. John Nicholson describes this nicely in the post vCenter Recoverability Improvements .

There are additional cluster health checks that can aid in the effort of rebuilding a vCenter server for an existing vSAN cluster. The vSAN health service can check for the consistency of cluster-wide settings such as deduplication and compression, encryption, and fault domains. Cluster-wide settings like these may be easily overlooked by an administrator during the process of building a new vCenter server, creating a new data center and cluster, and adding vSAN hosts previously associated with a vCenter server no longer available.

Scenario and remediation options

One or more cluster health check failures may surface when adding vSAN hosts managed by a vCenter server no longer available, into a cluster on a new vCenter server. As shown in Figure 1, the “vCenter state is authoritative” and the “vSAN cluster configuration consistency” health checks both failed. In this scenario, not only did the health check recognize that this is a new vCenter server not previously managing the hosts, it also identified that one or more cluster settings are inconsistent with the settings residing on the hosts. Looking at the message further, it states the issue is that deduplication and compression is enabled on the hosts, but not on the cluster.

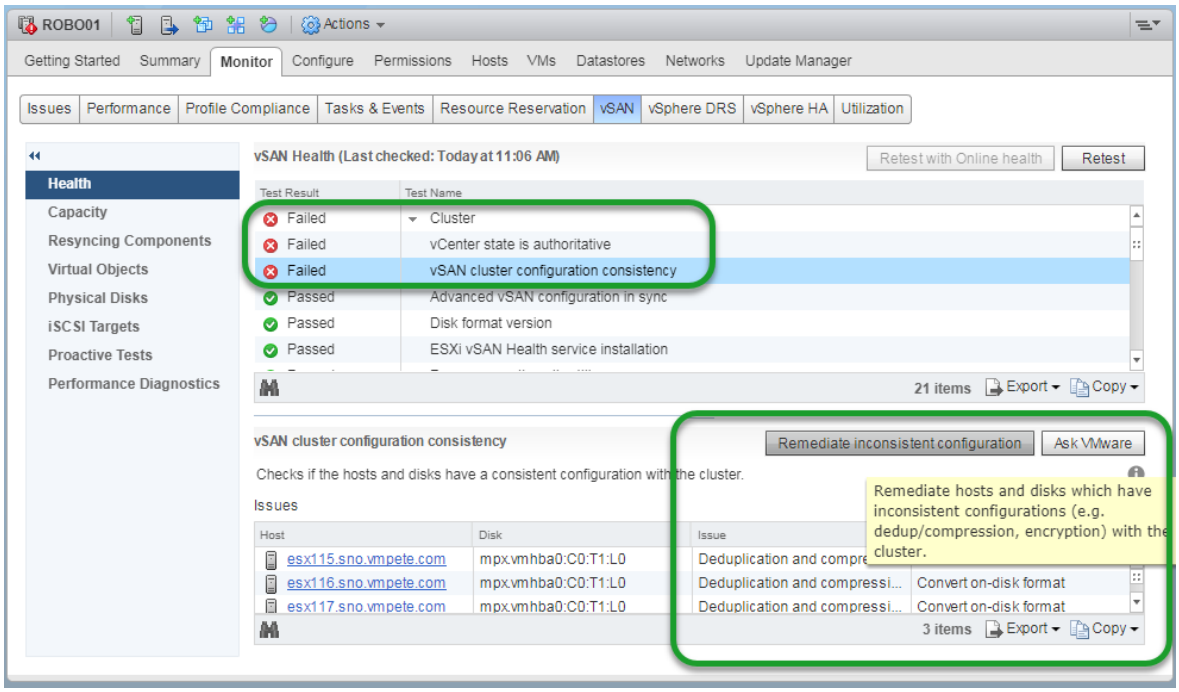


Figure 1. vSAN Cluster configuration consistency

The remediation of the health check errors shown above will depend on the specific health checks that failed, the configuration of the cluster, and the steps taken by the administrator.

In this example, the options for remediation shown below describe the effective result of hosts that were in a vSAN cluster with deduplication and compression **enabled** , where the original vCenter server

is no longer available. The hosts were added to a newly created vCenter server, where the cluster-wide setting of deduplication and compression is **not enabled** .

Option #1. Ticking the deduplication and compression checkbox in vCenter. This will enable deduplication and compression as seen by vCenter, and will provide consistency between vCenter and hosts. Since the hosts were already running deduplication and compression, this action involves a small metadata update, and does not introduce any rolling disk group evacuations common with enabling or disabling deduplication and compression. This will also eliminate the vCenter Authority health check alert, as vCenter will update the generation ID, and be identified as the source of truth after the change.

Option #2. Temporarily removing the recently added hosts to the new vCenter server, ticking the cluster-wide, deduplication and compression checkbox in vCenter, then re-adding the hosts. This will effectively eliminate the previously generated "vSAN cluster configuration consistency" failure, and leave only the "vCenter state is authoritative" health check failure. In this scenario, remediation of the "vCenter state is authoritative" health check would be a very light weight effort, as there were no other inconsistencies with cluster-wide settings.

Option #3. Clicking on "Remediate inconsistent configuration" in the health check UI. The current vSAN cluster-wide settings as defined in vCenter will be pushed down to all hosts participating in the vSAN cluster. In this case, this will kick off a rolling upgrade across all vSAN hosts to reflect setting of deduplication and compression NOT enabled. Any enabling or disabling of deduplication and compression on an active cluster can be a resource intensive operation, and is discouraged.

Option #4. Clicking on "Update ESXi Configuration" in the health check UI. This is similar to option #3, where vSAN cluster-wide settings of this new vCenter server will be pushed down to all hosts participating in the vSAN cluster. There is a warning of the impact of this change, as shown in Figure 2. Depending on the settings, this could be a lightweight metadata change, such as updating the generation ID, or in this scenario, a resource intensive operation, as it would push a new deduplication and compression setting to each host in the cluster.

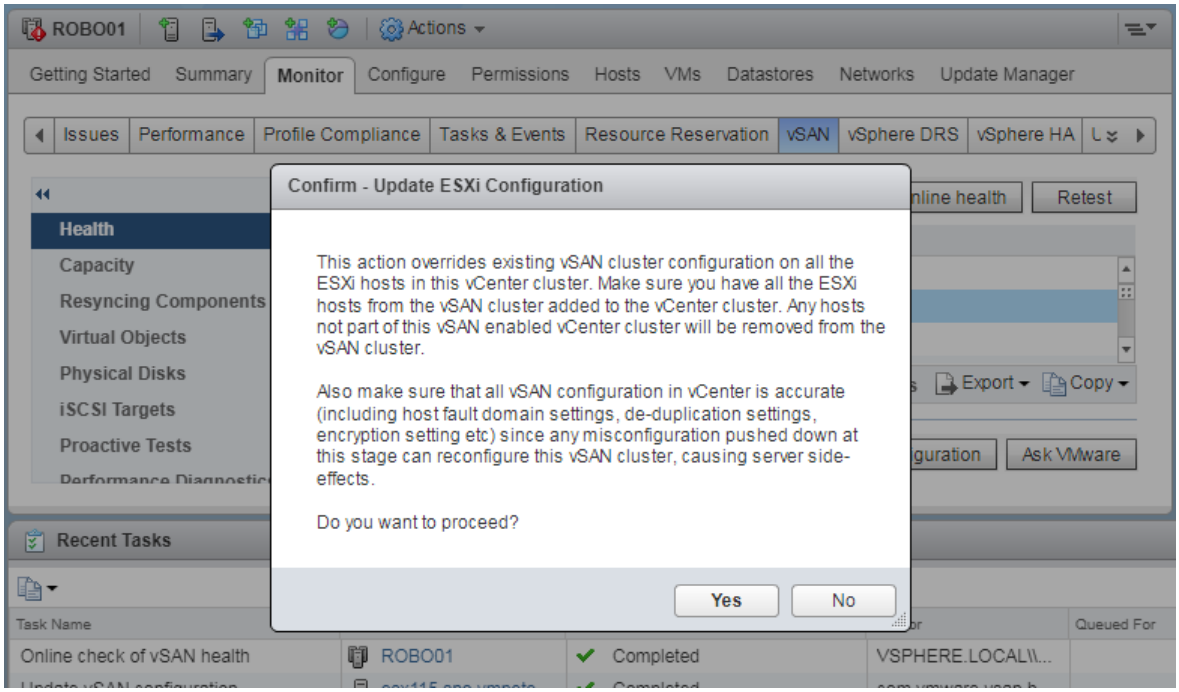


Figure 2. The confirmation dialog box for "Update ESXi Configuration"

In situations where the cluster wide services are consistent, then the only cluster health check alert may be the "vCenter state is authoritative" alert. This can be a light weight fix made by vCenter updating the generation ID on the hosts to reestablish consistency.

For clusters using vSAN encryption, additional steps may be necessary when replacing vCenter when vSAN encryption is enabled. Dave Morera describes his experiences with his post Replacing vCenter with vSAN encryption enabled, and is just an example of some additional factors to consider. Using an isolated lab to test the procedure specific to your environment is highly recommended.

Additional recommendations

Introducing a new vCenter server to an existing vSAN cluster can be made easier by adopting the following practices:

- **Run the latest version of vCenter.** After the initial installation of a new vCenter server, always run the "Check Updates" in the vCenter Appliance Management Interface, as shown in Figure 3. This will ensure that vCenter is always running the latest version, and is compatible with the version of ESXi running on the hosts.

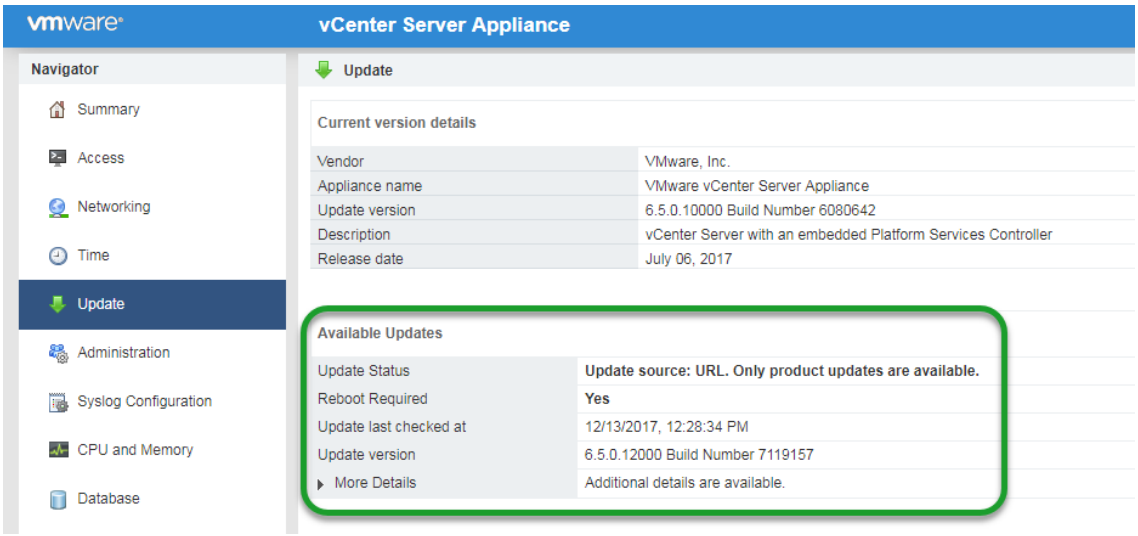


Figure 3. Updating vCenter using the vCenter Appliance Management Interface

- **Add licensing.** Add the vSphere host, vCenter, and vSAN licenses to vCenter prior to adding the hosts to streamline the process of adding the hosts and enabling services.
- **Set and verify cluster-wide settings.** Ensure that as many cluster-wide settings are configured the same on the new vCenter Server as they were on the old vCenter server. This would include, but is not limited to Data Center and cluster object names, HA and DRS settings, as well as all vSAN cluster-wide settings.
- **Verify your protection strategies for vCenter.** Ensure that application and system level protection of your vCenter server are made per organizational requirements. This might include guest-level backups, and exporting of configuration settings such as SPBM policies , vSphere distributed switches (VDS) and other items. This can make efforts in restoring much easier.

Conclusion

Maintaining availability of a vCenter server is a desired goal for any data center powered by vSphere and vSAN. In situations where the recovery of a vCenter server is not possible, the architecture of vSAN paired with the continued improvement of the integrated health checks for vSAN in vCenter allow for an easy, predictable experience of introducing a new vCenter server to an existing vSAN cluster. For more information on this topic, see Recovering a vCenter Server and vCenter Recovery Example with vSAN on StorageHub .

10. Compression and Deduplication Operations

Deduplication and compression on a vSAN cluster can be used as a space efficiency technique to eliminate duplicate data and reduce the amount of space needed to store data.

10.1 Compression and Deduplication

Deduplication and compression on a vSAN cluster can be used as a space efficiency technique to eliminate duplicate data and reduce the amount of space needed to store data. Starting in vSAN 6.2, deduplication occurs when data is de-staged from the cache tier to the capacity tier of an all-flash vSAN datastore. Compression is applied after deduplication has occurred and before the data is written to the capacity tier

Deduplication and compression is a vSAN cluster-wide setting. Some important notes on deduplication and compression:

- Only available on *all-flash*
- On-disk format version 3.0 or later is required
- Capacity overhead is approximately 5% of total raw capacity

Deduplication and compression are enabled as a unit. It is not possible to enable deduplication or compression individually. Deduplication and compression can be used with:

- Two Node vSAN Cluster (ROBO)
- Stretched vSAN Cluster configuration
- vSAN Cluster with Fault Domains

For new clusters, deduplication and compression can be enabled during cluster creation phase. For existing clusters, deduplication and compression can be enabled turning by selecting Deduplication and compression as a property of an existing cluster. As a consequence a rolling reformat of every disk group on every host in the vSAN cluster is required, which can take a considerable amount of time to evacuate data. This process does not incur virtual machine downtime.

10.2 Enabling Dedup/Compression on a New Cluster

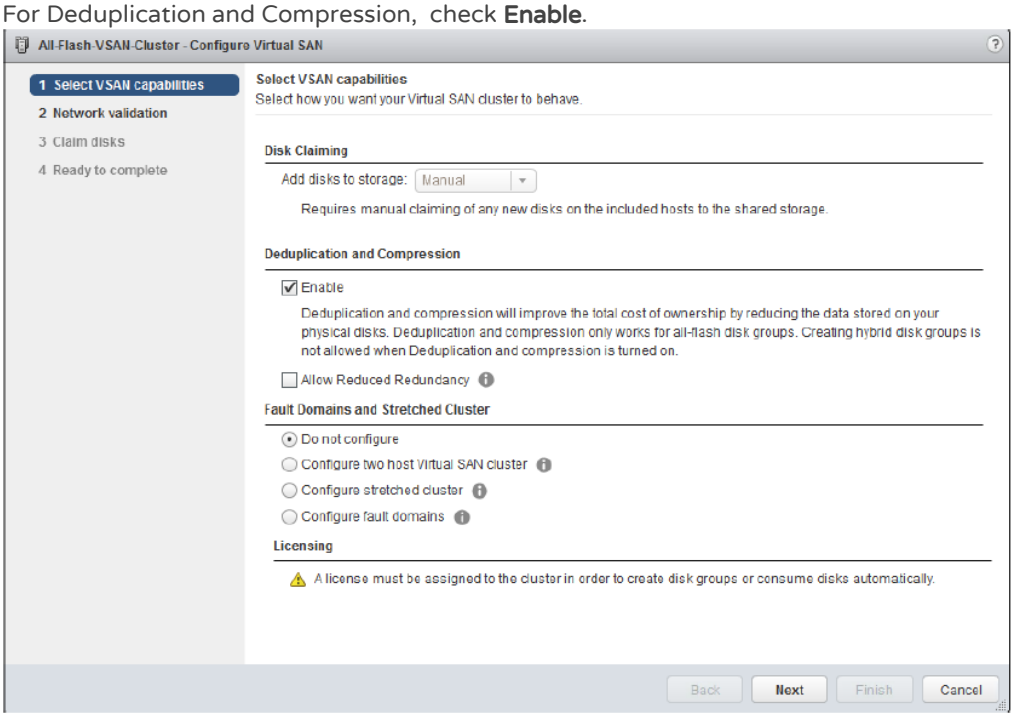
Prerequisites :

- vSAN 6.2 or later, which includes vCenter 6.0 U2 and ESXi 6.0 U2
- Flash Devices for both Cache and Capacity devices. At least 1 SSD for cache tier and 1 SSD for capacity per host or node
- A valid license to enable deduplication and compression on a cluster
- At least 3 hosts or nodes contributing storage
- vSAN Networking is configured properly.
- An existing vSAN cluster created on Virtual Center

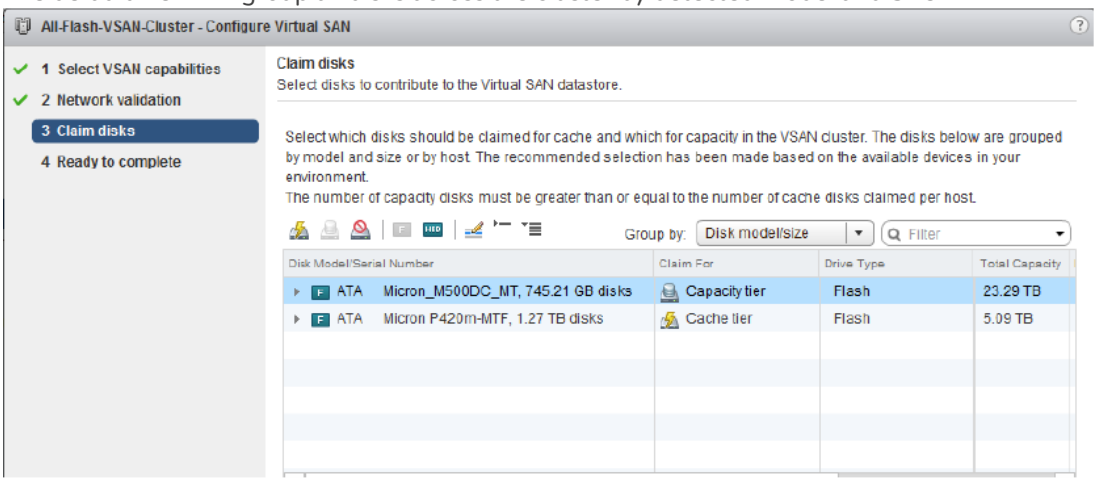
Procedure:

1. Open vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the *cluster* you want to enable Dedup/Compression on.
4. Click the **Configure** tab.
5. In the vSAN is turned on pane, select **General** and click **Configure vSAN** button.
6. Configure deduplication and compression on the cluster:
    - For "Add disks to storage", if necessary, change to **Manual**.

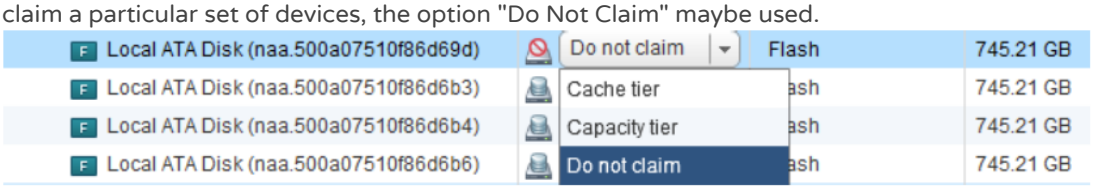
- For Deduplication and Compression, check **Enable**.



7. The wizard will recommend a configuration based on the available devices across the cluster. The default view will group all disks across the cluster by detected Model and Size .

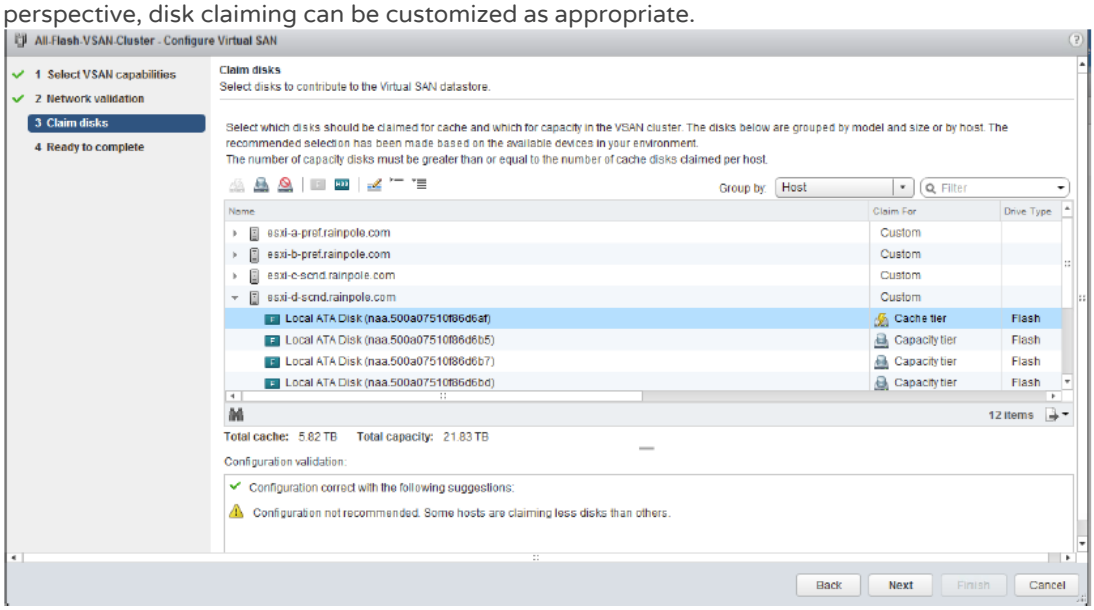


8. Individual disks can be selected for either cache or capacity, if an administrator does not wish to claim a particular set of devices, the option "Do Not Claim" maybe used.



9. A possibly more useful view can be used to display available devices by changing "Group by:" to "Host". This will allow the administrator to select cache and capacity devices from a host

perspective, disk claiming can be customized as appropriate.



10. Once the cluster is formed and disks are formatted , Disk Claiming can be changed to Automatic

Effects

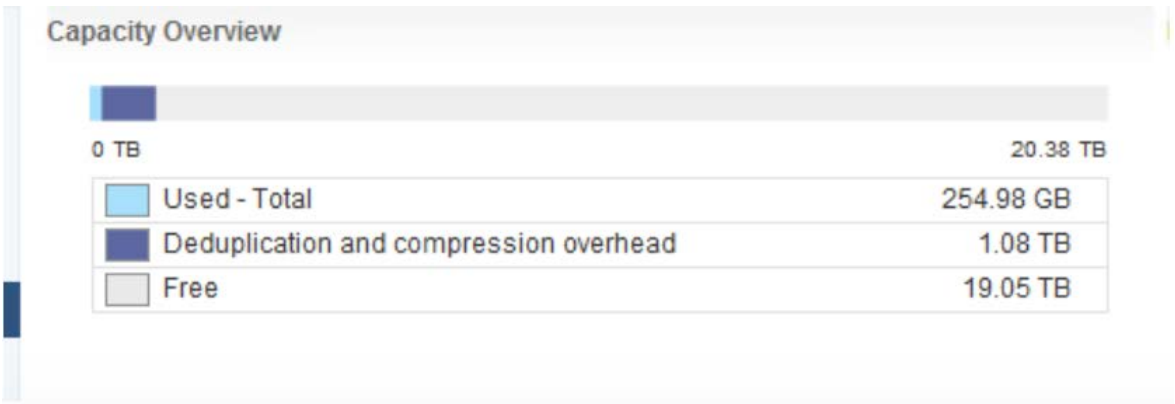
The following Advanced Options will be changed on each hosts participating in a vSAN enabled Cluster: */VSAN/DedupScope* - will be set to the value "2". You can inspect this through the console or SSH as follows:

```
esxcli system settings advanced list -o /VSAN/DedupScope
```

Enabling a cluster will format each disk group on each host with on disk format of V3. Once all configuration tasks have completed, the vSAN cluster will be formed with Deduplication and Compression enabled.

Deduplication and Compression Overheads can be displayed from the vSphere Client:

1. Open vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the *cluster*, click the **Monitor** tab.
4. Click on **Capacity**.



The raw capacity of the above example shows 20.38 TB, with vSAN Deduplication and Compression enabled. The overhead shown above equates to an approximate 5% overhead. In other words:

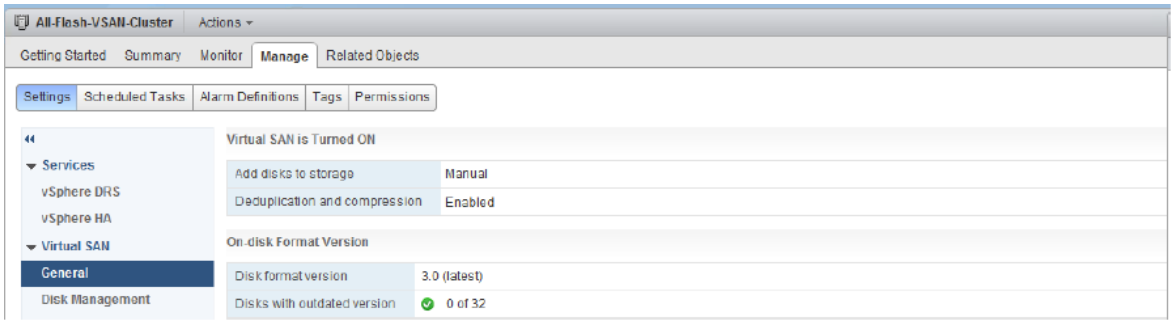
Deduplication and Compression Overhead = ((dedup and compression overhead / raw capacity) * 100)), Working example from above screenshot:((1.08TB/20.38TB)*100) = 5.3%

Effects

The following Advanced Options will be changed on each hosts participating in a vSAN enabled Cluster:
`/VSAN/DedupScope` - will be set to the value "2"

You can inspect this through the console or SSH as follows:
`esxcli system settings advanced list -o /VSAN/DedupScope`

Enabling a cluster will format each disk group on each host with on disk format of V3 Once all configuration tasks have completed, vSAN cluster will be formed with Deduplication and compression enabled.



Deduplication and Compression Overheads can be displayed from the vSphere Client:

1. Select vSAN Cluster
2. Click on the Monitor tab
3. Click on Capacity

The raw capacity of the above example shows 20.38 TB, with vSAN Deduplication and Compression enabled. The overhead shown above equates to an approximate 5% overhead.

In other words:
Deduplicaton and Compression Overhead = ((dedup and compression overhead / raw capacity) * 100))

Worked Example from above screenshot:
((1.08TB/20.38TB)*100) = 5.3%

10.3 Enabling Dedup/Compression on an Existing Cluster

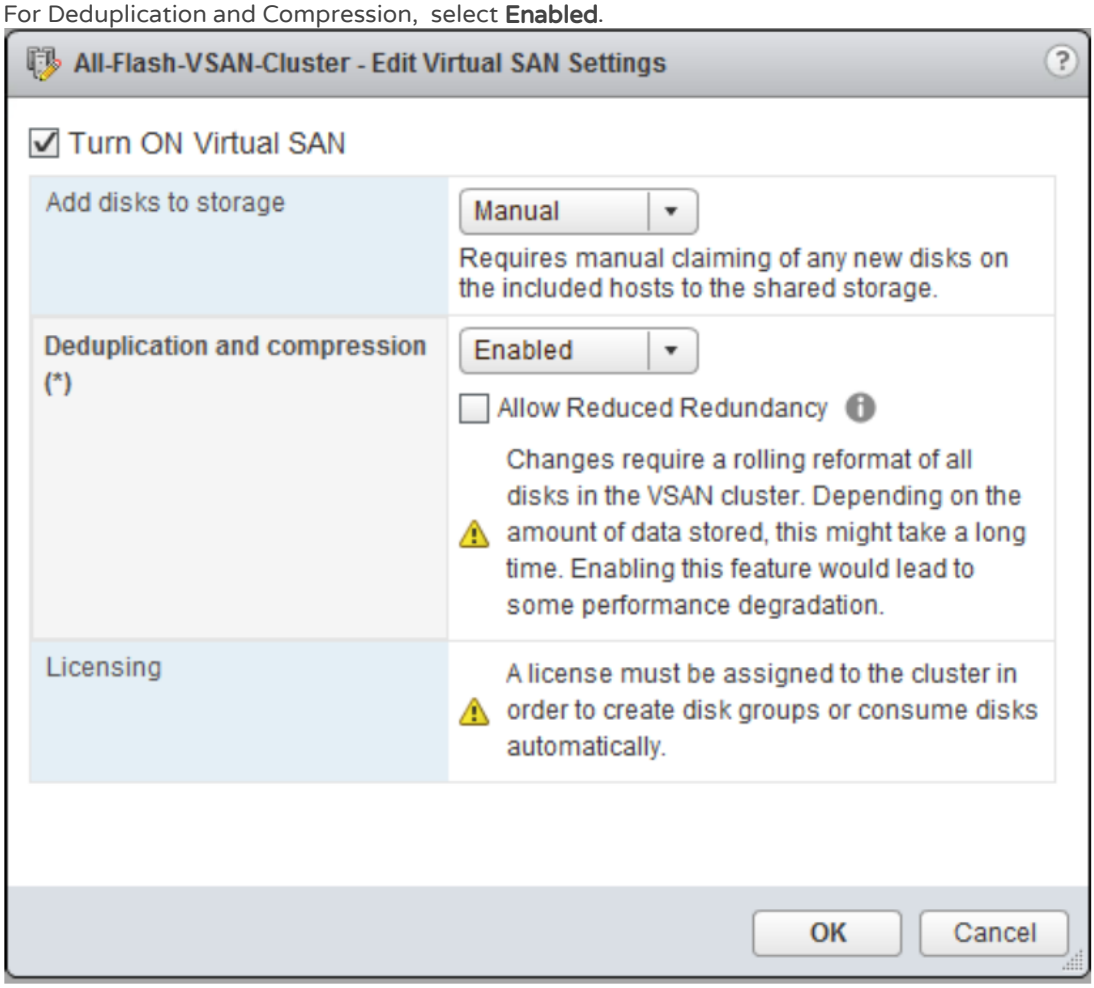
To enable Deduplication and Compression on an existing vSAN cluster follow the procedure below:

Prerequisites:

- Disk claiming must be set to manual.
- All hosts must be connected to vCenter.
- Performing a vSAN health check is **highly recommended**.

Procedure:

1. Open vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the *cluster* you want to enable Dedup/Compression on.
4. Click the **Configure** tab.
5. In the vSAN is turned on pane, click **Edit**.
6. For "Add disks to storage", if necessary, change to **Manual**.
7. For Deduplication and Compression, select **Enabled**.



8. Click **OK** to apply the configuration change

This process will trigger the following effects to a cluster

vSAN will select a host at random and will perform the following steps:

1. Data evacuation from the affected disk group(s) to another host and diskgroup
2. Remove the disk group(s) from the host
3. Update advanced parameter "/VSAN/DedupScope" on each host to the value of 2
4. Re-Add the disk group(s) with Dedup and compression enabled. (existing disk group configurations will be retained).

Steps 1-4 will be repeated as necessary on each host in a given vSAN Cluster

Note:- Performing above procedure will not trigger a virtual machine migration and works independently of DRS. Depending on how much data to move from one disk group, and the amount of hosts in a cluster, this operation may take a long time as it is a network and disk intensive operation

10.4 Disabling Dedupe/Compression

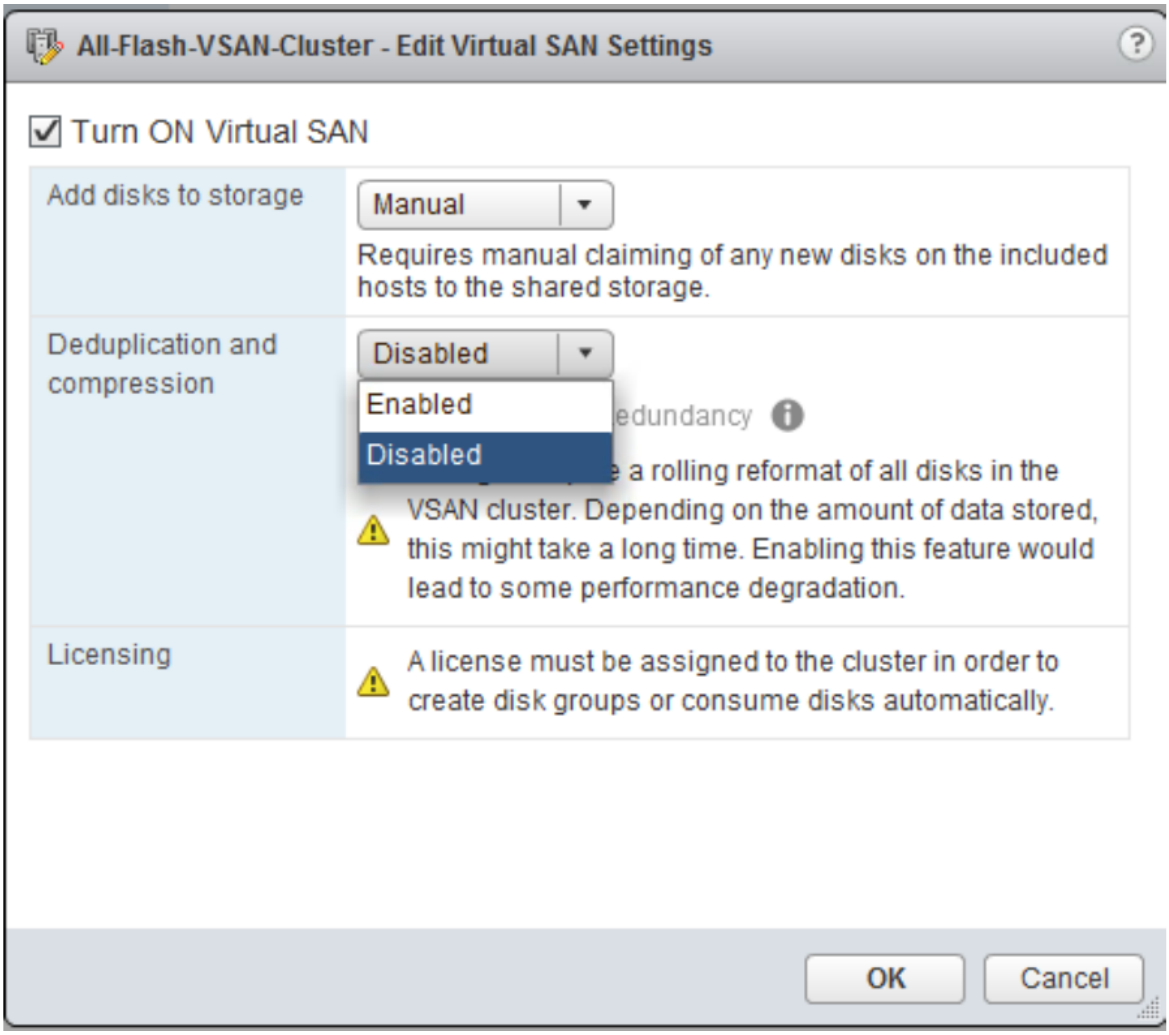
To disable Deduplication and Compression on an existing vSAN cluster follow the procedure below:

Prerequisites:

- Disk claiming must be set to M**anual**.
- All hosts must be connected to vCenter.
- Performing a vSAN health check is **highly recommended**.

Procedure:

1. Open vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the *cluster* you want to enable Dedup/Compression on.
4. Click the **Configure** tab.
5. In the vSAN is turned on pane, click **Edit**.
6. For "Add disks to storage", if necessary, change to **Manual**.
7. For Deduplication and Compression, select **Disabled**.
8. Click **OK** to apply the configuration change.



This process will trigger the following effects to a cluster

Disabling Dedup/Compression requires an on-disk format conversion. This will trigger the following effects to a cluster. vSAN will select a host and will perform:

vSAN will select a host at random and will perform the following steps:

1. Data evacuation from the affected disk group(s) to another host and diskgroup
2. Remove the disk group(s) from the host
3. Update advanced parameter "/VSAN/DedupScope" on each host to the value of 0
4. Re-Add the disk group(s) without Dedup and compression enabled. (existing disk group configurations will be retained).

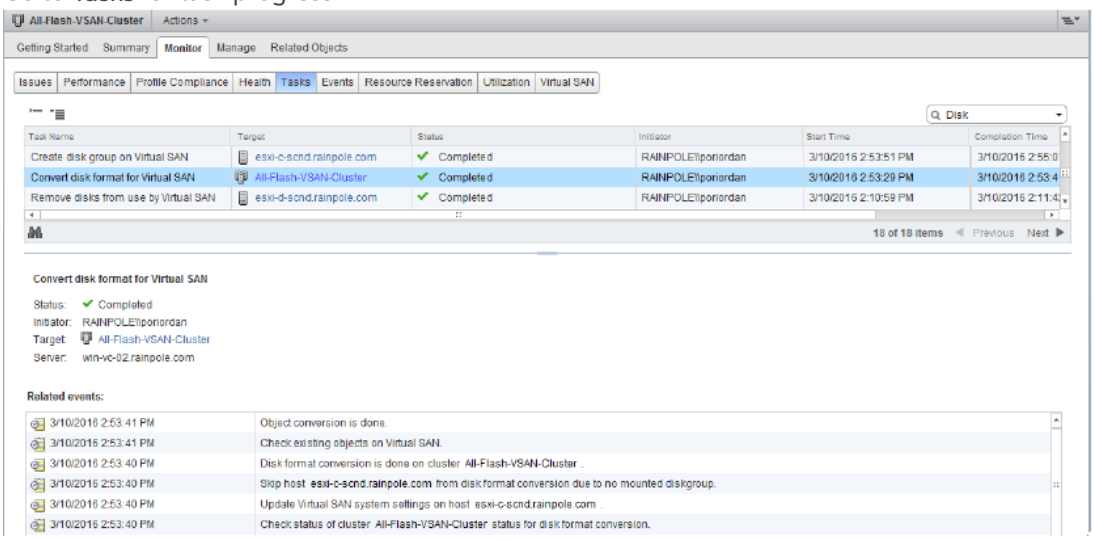
Steps 1-4 will be repeated as necessary on each host in a given vSAN Cluster

Note: Performing above procedure will not trigger a virtual machine migration and works independently of DRS. Depending on how much data there is to move from one disk group, and the number of hosts in a cluster, this operation may take a long time as it is a network and disk intensive operation

Monitoring Progress of Enabling or Disabling Dedup/Compression

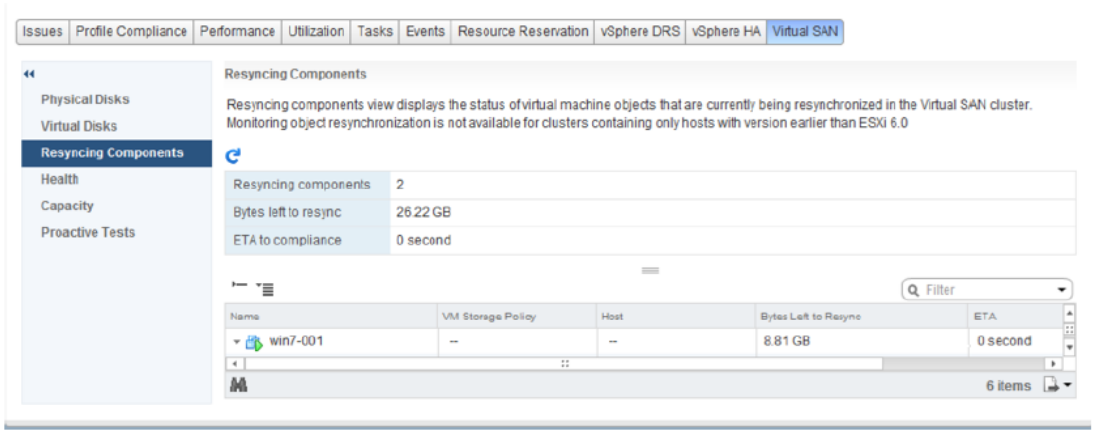
Since a rolling reformat of every disk group on every host in the vSAN cluster is required, and the task can take a considerable amount of time to complete it may be necessary to track the progress of the operation(s). Progress can be monitored from the vSphere Client from Tasks and Events:

1. Select the *Cluster.*
2. Click on the **Monitor** tab.
3. Go to **Tasks** for task progress.



Resyncing components can also be Monitored from vSphere Web Client:

1. Select the *Cluster.*
2. Click on the **Monitor** tab.
3. Select the **Virtual SAN** tab.
4. Select **Resyncing Components.**



Overall Disk conversion process logs can be found on your vCenter Server instance at:

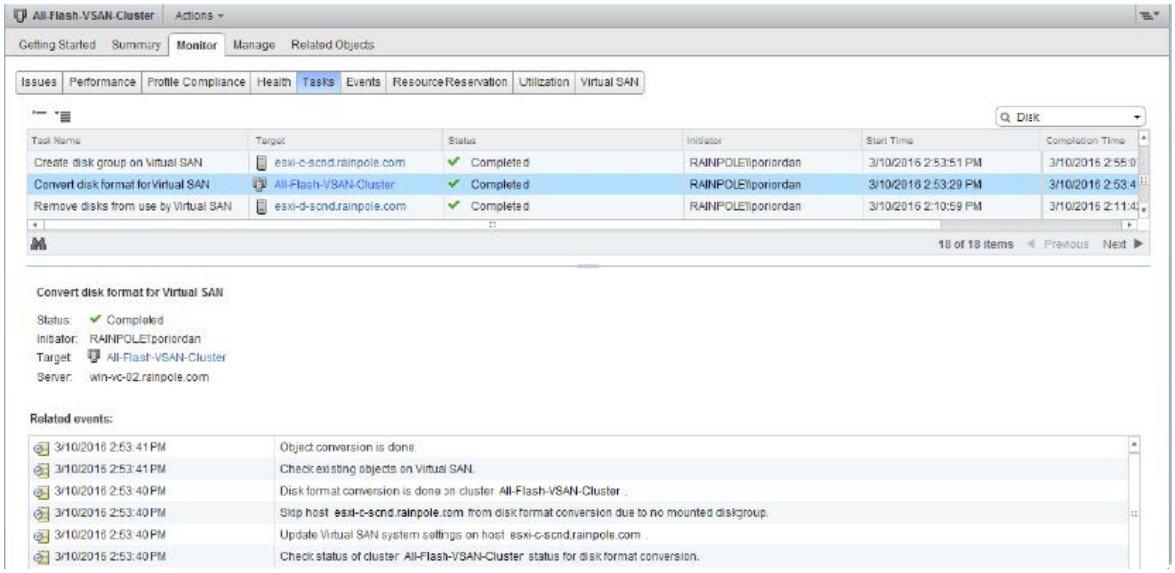
Windows:
%ProgramData%\VMware\vCenterServer\logs\vsan-health\vmware-vsan-health-service.log

VCSA:
/var/log/vmware/vsan-health/vmware-vsan-health-service.log

10.5 Monitoring Progress of Enabling/Disabling

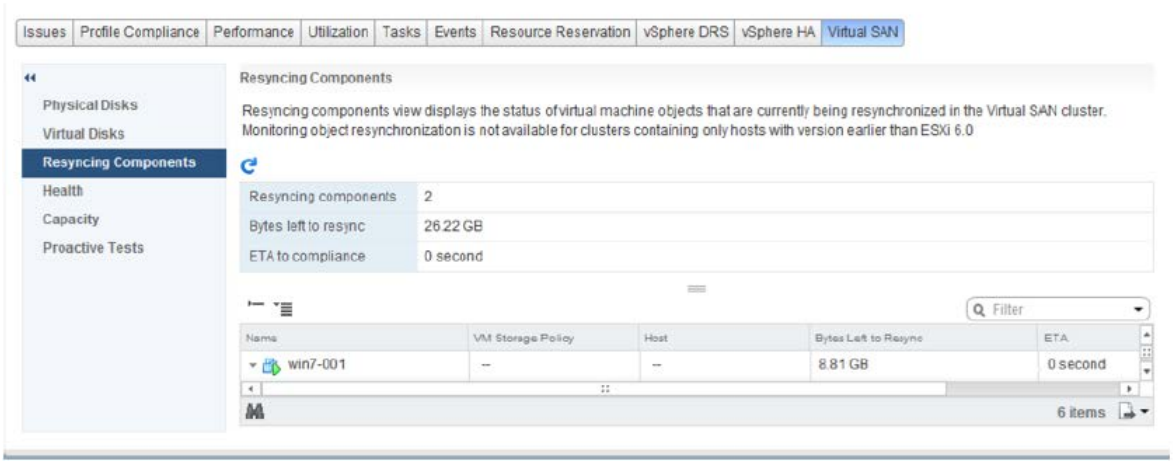
Since a rolling reformat of every disk group on every host in the vSAN cluster is required, and the task can take a considerable amount of time to complete it maybe necessary to track progress of the operation(s). Progress can be monitored from the vSphere Client from Tasks and Events:

1. Open vSphere Web Client.
2. Click **Hosts and Clusters**.
3. Select the cluster, click on the **Monitor**>**Task & Events** tab.



Resyncing components can also be monitored from vSphere Web Client:

1. Open vSphere Web Client.
2. Click **Hosts and Clusters**.
3. Select the cluster, click on the **Monitor** tab.
4. Select **vSAN>Resyncing Components**.



Overall Disk conversion process logs can be found on your vCenter Server instance at:

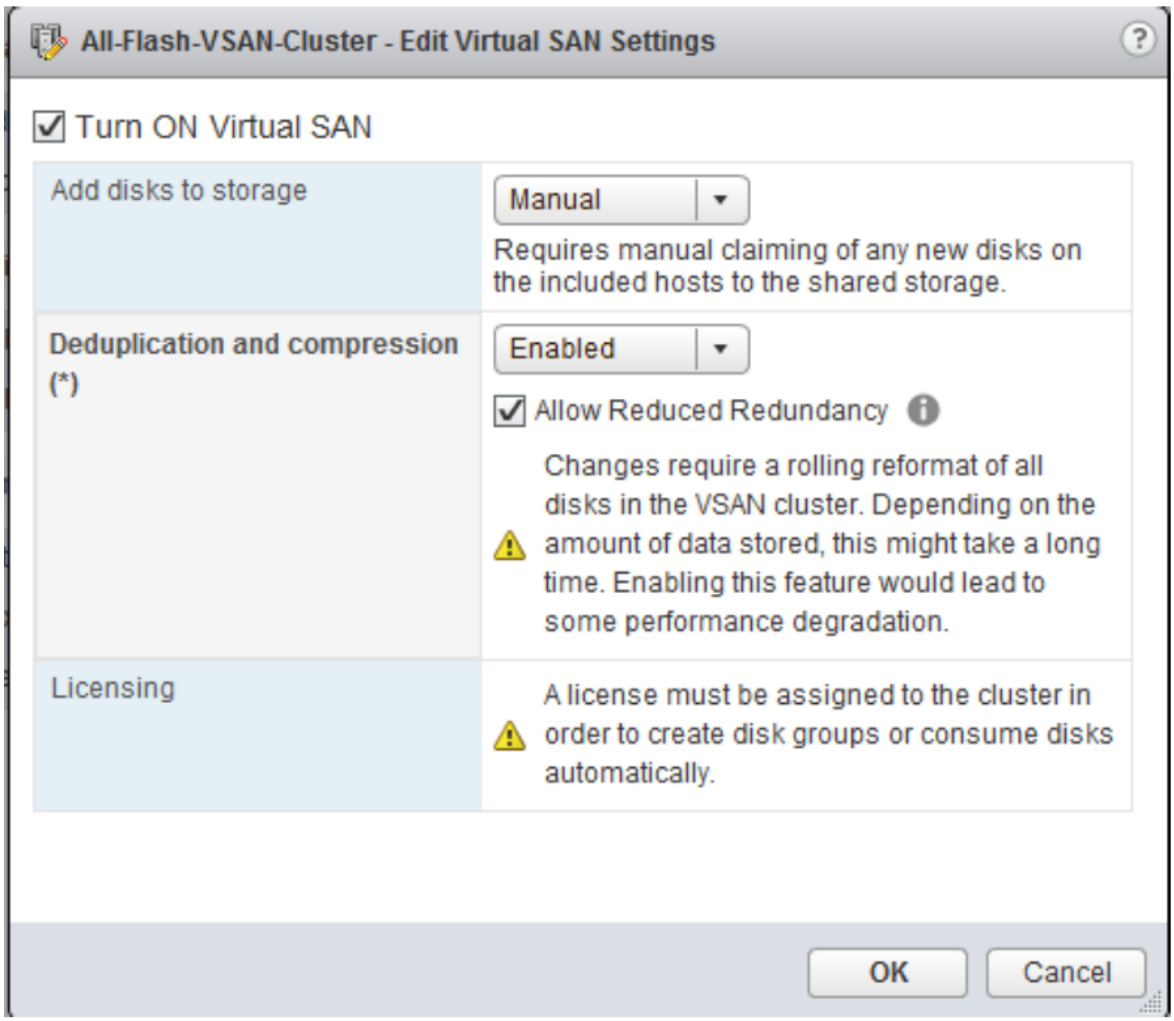
vCenter Server Windows:
%ProgramData%\VMware\vCenterServer\logs\vsan-health\vmware-vsan-health-service.log
vCetner Server Appliance:
/var/log/vmware/vsan-health/vmware-vsan-health-service.log

10.6 Allow Reduced Redundancy

Because a disk format conversion requires a data evacuation from the disk group, data availability must be maintained when performing a disk evacuation. Depending on the fault Tolerance Method used this poses a problem in the following use cases:

- Three mode clusters as data, objects cannot be evacuated to another host
- Erasure coding (raid-5 or raid-6) is used for Fault Tolerance Method
  - For example Raid-5 (host failures to tolerate = 1) will require a minimum of 4 fault domains (or hosts)
  - For example Raid-6 erasure coding ((host failures to tolerate = 1) will require a minimum of 6 fault domains (or hosts)
- Insufficient capacity in the cluster to evacuate a hosts disk group(s)

For these reasons the option "Allow Reduced Redundancy" is exposed when Enabling or Disabling Dedup/Compression.



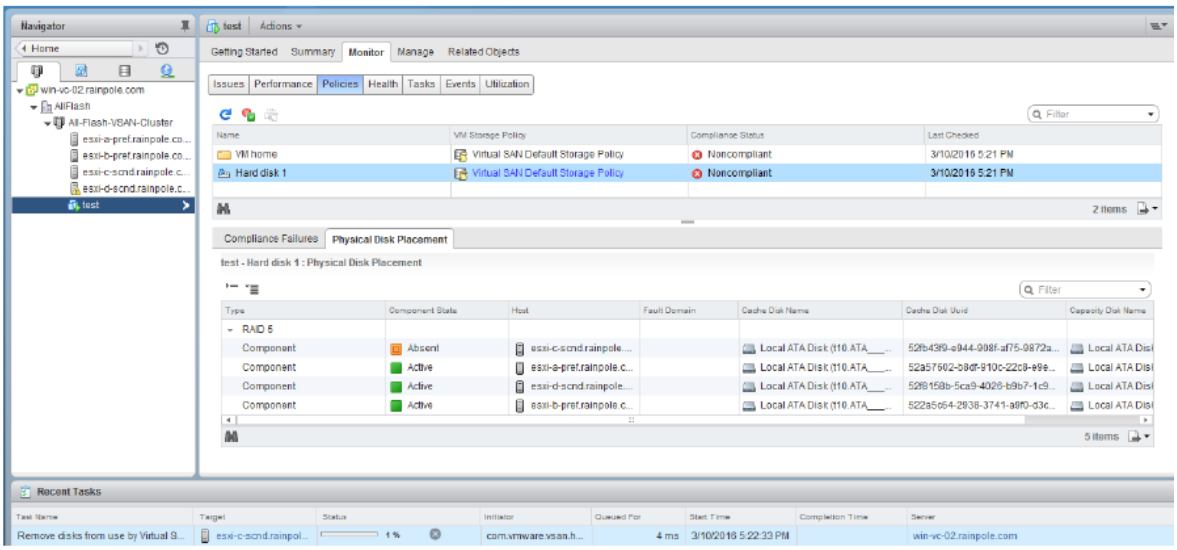
With this option set:

1. vSAN **removes** the redundant copy of components from the objects marking them as absent.
2. Removes the affected disk group(s) from a host.
3. Recreates the disk group with the new on-disk format,
4. Resynchs the component(s) before moving onto the next disk group.

If an object has a host failures to tolerate = 0 the object will be "moved" to a different Disk Group, assuming there is a adequate capacity to successfully perform this task.

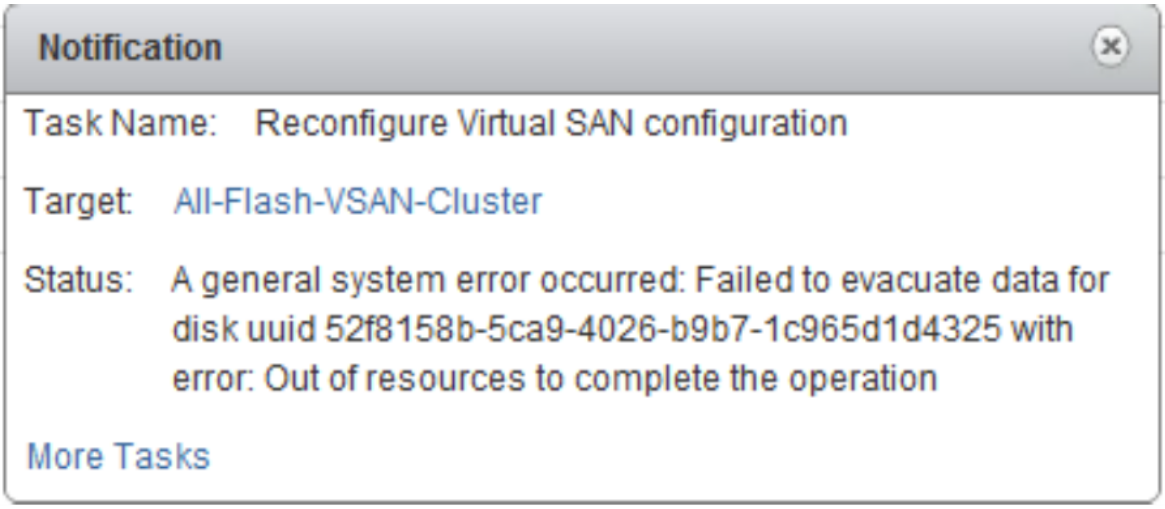
There is a large amount of operational risk associated with selecting this option for example in a 3 node cluster, setting Allow Reduced Redundancy, will remove a component of an object to remove and add a disk group. This means the object cannot tolerate another failure if another host or disk group goes unavailable. It is highly recommended by VMware that more nodes or capacity should be added to a cluster to ensure there are adequate resources to perform a disk format conversion.

Below figure shows a vSAN Disk object , protected using Erasure Coding (Raid-5) for Fault Tolerance Method. As you can see when Allow Reduced Redundancy is selected during disk conversion (disk remove and re-add) the Virtual Disk object from a policy perspective, is non complaint, as one of the components is absent. If another host or disk group becomes unavailable while the disk conversion is in progress then the Virtual Machine and its data will become unavailable.



If an Administrator attempts to enable or disable deduplication and compression, but does not have enough resources to complete data evacuation, the conversion task will fail. This is sample of the event failure:

```
"A general system error occurred: Failed to evacuate data for disk uuid xxxxxxx-
xxxx-xxxx-xxxx-xxxxxxxxxxx with error: Out of resources to complete the
operation"
```



10.7 Adding a capacity Tier Disk

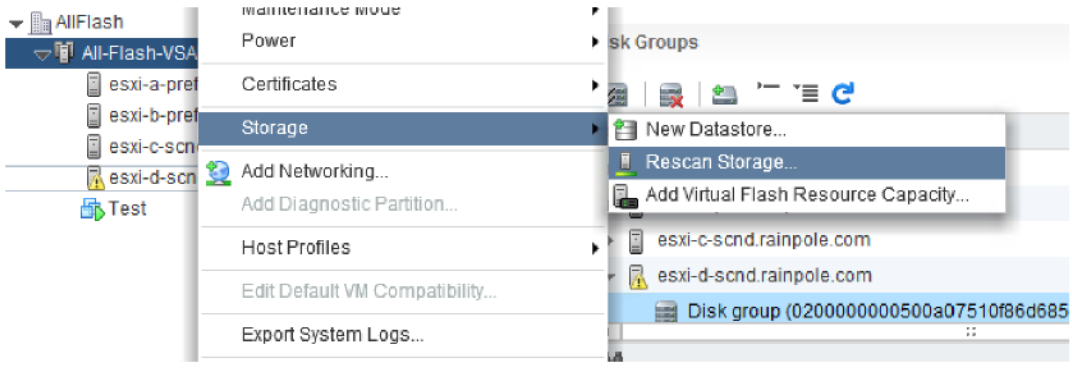
It is possible to add a capacity tier disk to a dedupe/compression enabled disk group. However, dedupe data and metadata hash tables are spread out across all the capacity disks in a disk group, it is not operationally efficient to do this. For more efficient deduplication and compression, instead of adding capacity disks to an existing disk group, consider creating a new disk group to increase cluster storage capacity or performing a full data migration, removing and recreating a disk group with additional capacity .

From a procedural perspective it is the same as the no-dedupe use case:

Manual Disk Claiming Mode

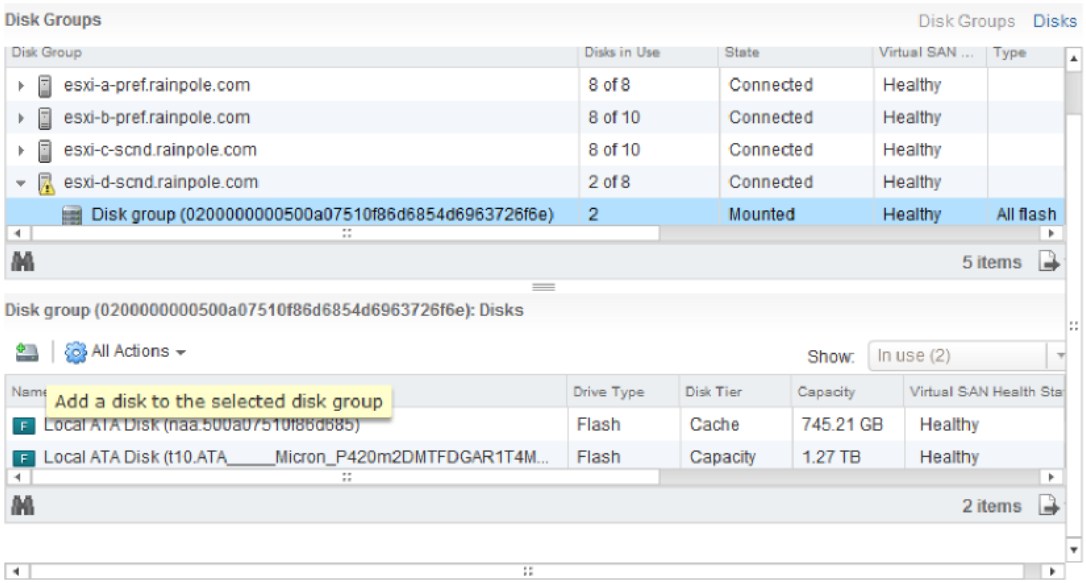
If required, physically insert disk into specific host

1. Perform a rescan of Storage to ensure ESXi host recognizes new disks



2. Select "Scan for new Storage Devices" This rescans all host bus adapters for new storage devices. Ensure device is visible and operational from a ESXi host perspective and is recognized as a SSD. Since Dedupe/Compression is enabled, All Flash is a requirement.

3. From vSphere Web Client, Select V**irtual SAN Enabled Cluster > Manage > vSAN > Disk Management**
4. Select host and disk group to add new flash capacity



5. Add desired disk(s) and click **OK**.

This will add the claim rule "enable_capacity_flash" on the flash disk and add it to an existing disk group by formatting it with the vSAN file system format.

Automatic Disk Claiming Mode

Newly discovered flash devices will **not** be automatically added to a vSAN Disk Group as they will not have the tag "enable_capacity_flash" attribute

If an administrator wishes they may manually tag a new flash disk as using esxcli
`esxcli vsan storage tag add -d <naa.xxxxxxxxxxxxx> -t capacityFlash`

Disks will then be automatically added to new or existing disk groups.

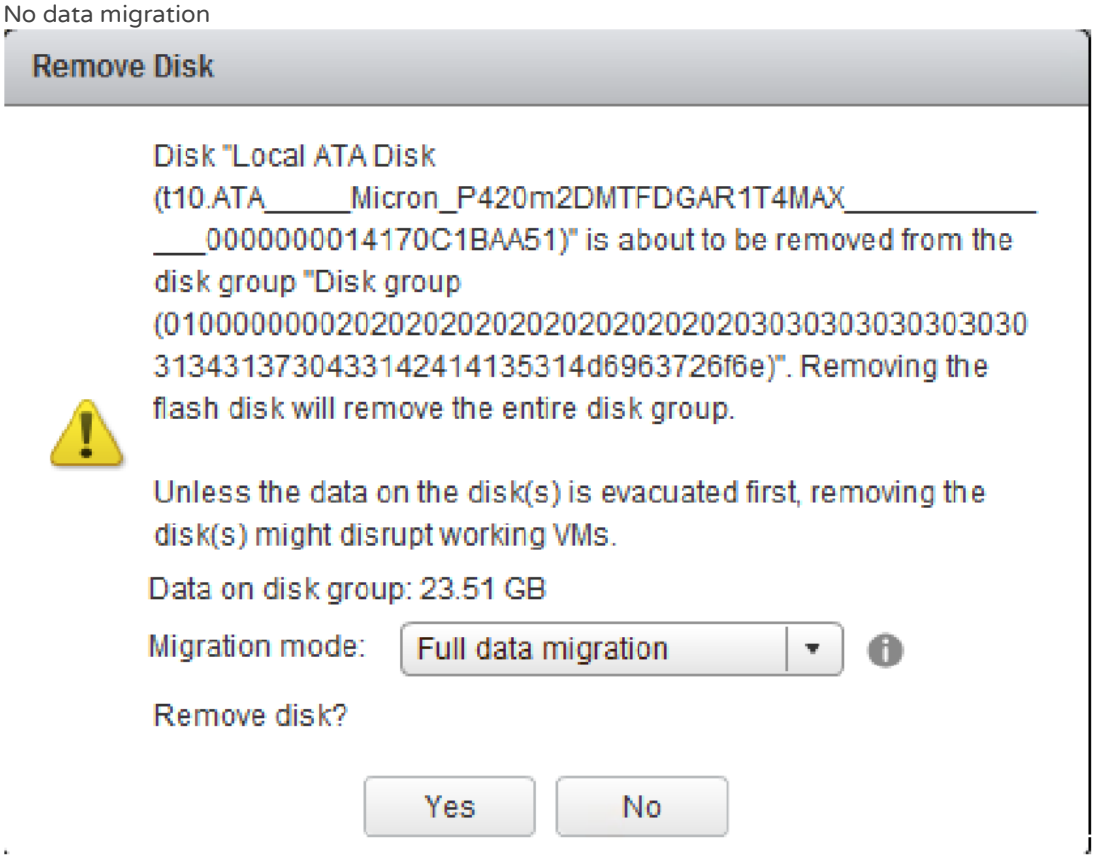
10.8 Removing a Cache Disk

Removing a disk assigned as the caching or cache tier will result in removal of the entire disk group. If the user requests to remove a cache tier disk from a Dedup and Compression enabled diskgroup a data evacuation will task will be triggered.

Depending if the disk group is utilized or not, and the protection mechanism implemented on the objects, the options are:

- Full Data Migration (or evacuation)
- Ensure accessibility

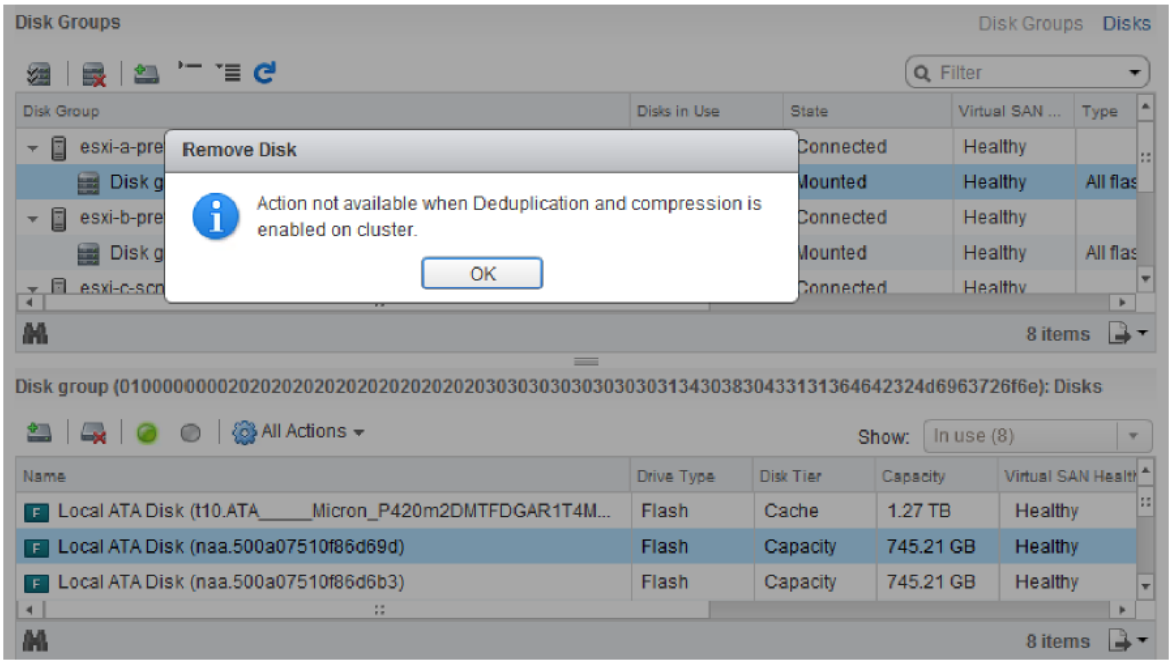
- No data migration



This is similar to the options used for Maintenance Mode, see vSphere 6.5 Working with Maintenance Mode for a detailed description.

10.9 Removing a Capacity Disk From a Disk Group

This operation is **not** possible when deduplication and compression are enabled because deduplication is implemented at a disk group level,. Dedup data and metadata (hashes) are stored in a stripe across all disks in a disk group. The hash tables are spread out across all the capacity disks in a disk group. As a result it is not possible to remove capacity disks from a disk group after the space savings features are enabled. If an administrator attempts to remove a disk from a disk group, the action is not available. the vSphere web client will inform the user "Action not available when Deduplication and compression is enabled on cluster"



10.10 Failure Considerations for Cache Disk

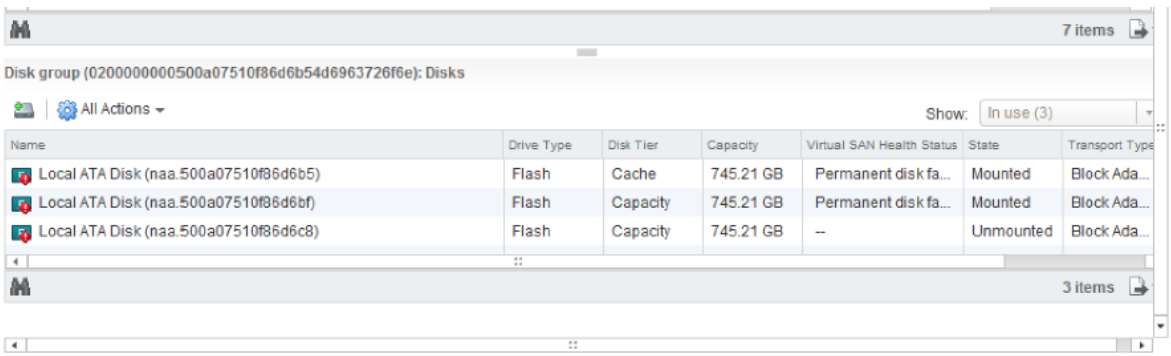
If a cache disk suffers a failure in a dedup and compression enabled disk group, the entire disk group goes offline, and all components on that disk group are marked as "Degraded". This behaviour is no different from a non dedup/compression enabled disk group. Ensuring adequate capacity to allow for a disk group failure is highly recommended.

10.11 Failure Considerations for Capacity Disks

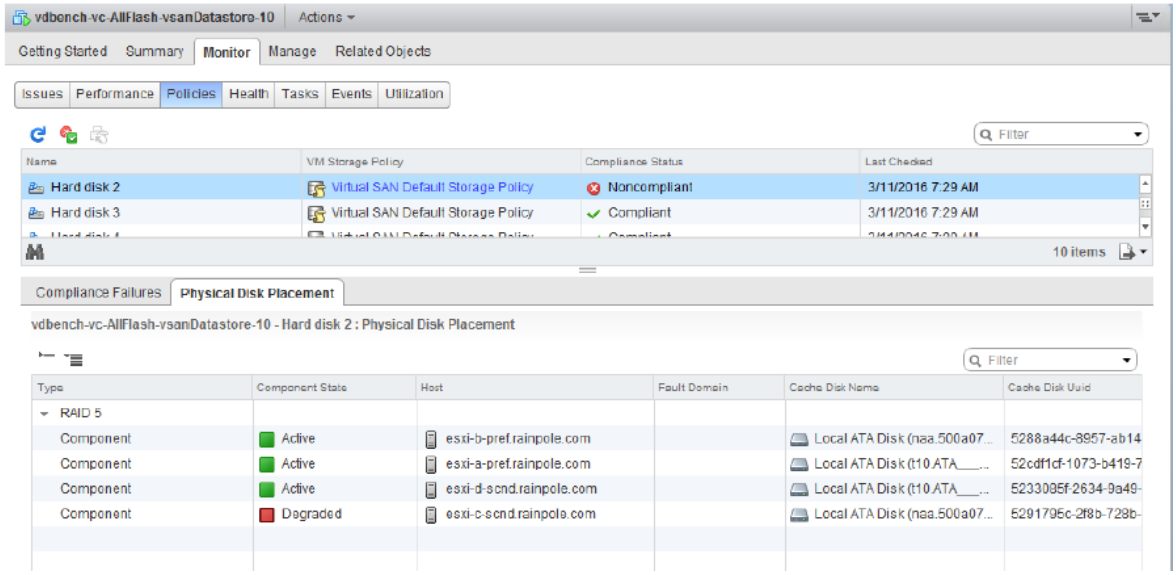
If a capacity disk suffers a failure in a dedup and compression enabled disk group, the entire disk group goes offline, and all components on that disk group are marked as "Degraded" this is different from a non dedup/compression enabled disk group.

The reason the entire disk group goes offline is dedupe data and metadata (hashes) are stored in a stripe across all disks in a disk group. so a failure of a capacity tier disk will render the data on the disk group defunct.

Figure below describes a failed capacity disk and the effect on a disk group with dedup and compression enabled. Capacity disk `naa.500a07510f86d6c8` had an error, however all disks in disk group went offline as dedup/compression was enabled.

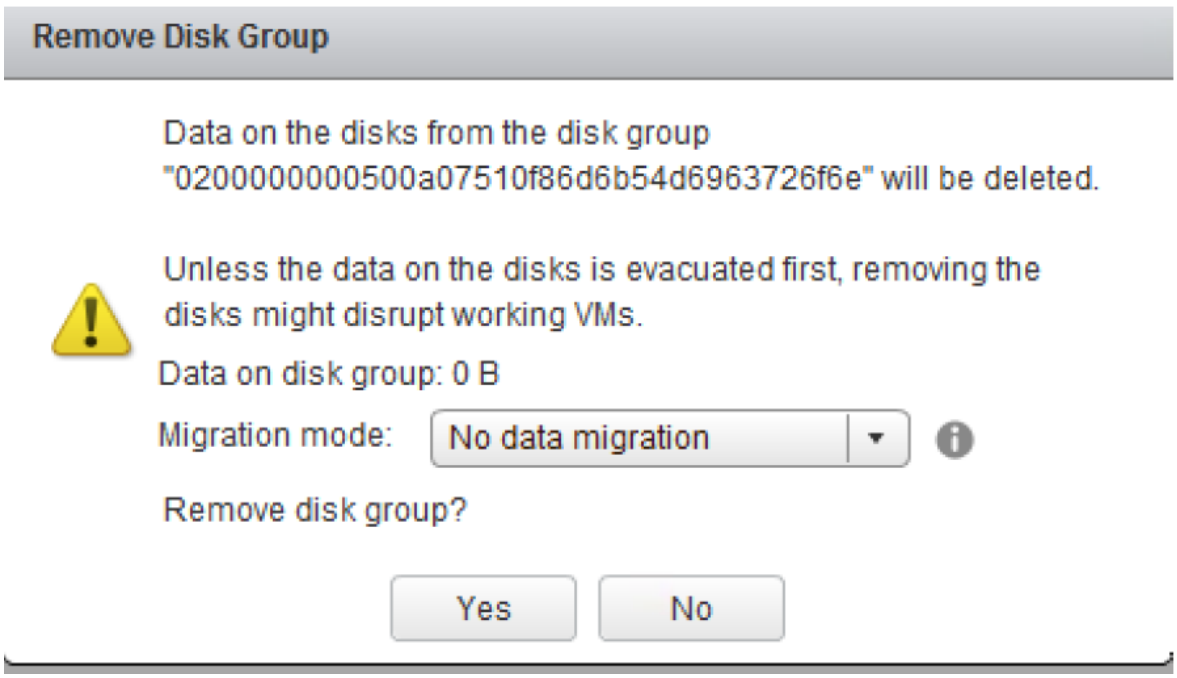


From a Virtual Machine object perspective, any affected components will go into "Degraded" immediately, and the compliance status will be "Non compliant".



To resolve the issue, the failing component must be identified and replaced as necessary.

If the failed disk group is required to be removed, then the option with the option "No Data Migration" should be selected as the disk group is unhealthy. If an entire disk group is "unhealthy" you cannot evacuate data in EvacuateAllData mode.



A new disk group can then be created.

11. Checksum Operations

End-to-End Software Checksum helps customers avoid data integrity issues that may arise due to problems on the underlying storage media.

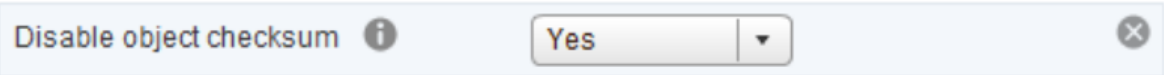
11.1 Checksum Operations

The vSAN capability for checksum is called '*Disable object checksum*' It may be disabled or enabled when creating or modifying a VM Storage Policy. By default is always enabled without the need for an explicit rule added to a given policy. It may be enabled or disabled on per virtual machine/object basis.

11.2 Defining a VM Storage Policy for Checksum

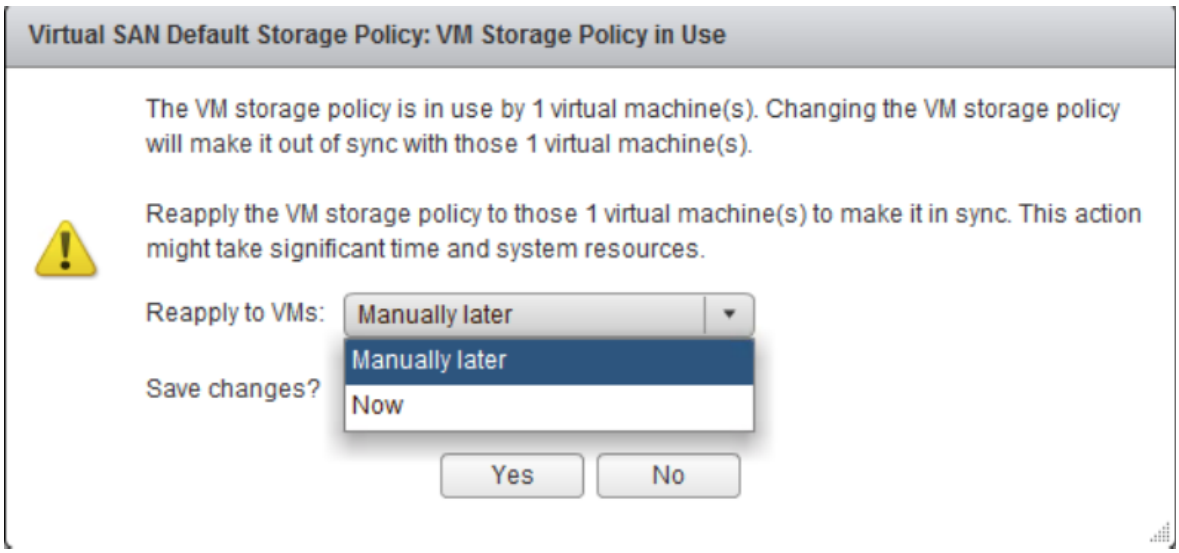
Software checksum can be explicitly disabled via VM Storage Policy Procedure:

1. From the vSphere Web Client home, click P**olicies and Profiles > VM Storage Policies**.
2. Click the **VM Storage Policies** tab.
3. Select a *storage policy*, and click **Edit a VM storage policy**.
4. From Rule-Set 1 screen, click**Add rule**.
5. From the dropdown list select *Disable object checksum*
6. Select **Yes**.
7. Click **OK**.



11.3 Applying Policy with a VM Storage Policy

Changing an existing policy will prompt and Admin to either apply the new ruleset immediately or apply it manaully to a VM or object later For example modification of an existing "in use" VM Storage Policy with the a rule-set will prompt an Administrator to Reapply Policy "Manually Later" or "Now"

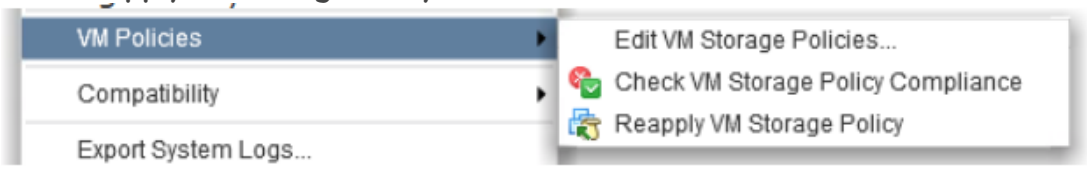


If an Admin chooses "Now" this will the new rule-set will be applied to all the Virtual Machines or objects that use a VM Storage Policy. This may trigger several reconfigure task on a Virtual Center system. The "VM Storage Policy in Use" dialogue will inform the admin how many virtual machines would be affected.

11.4 Manually Disabling Checksum on a VM or Object

Procedure - Per VM:

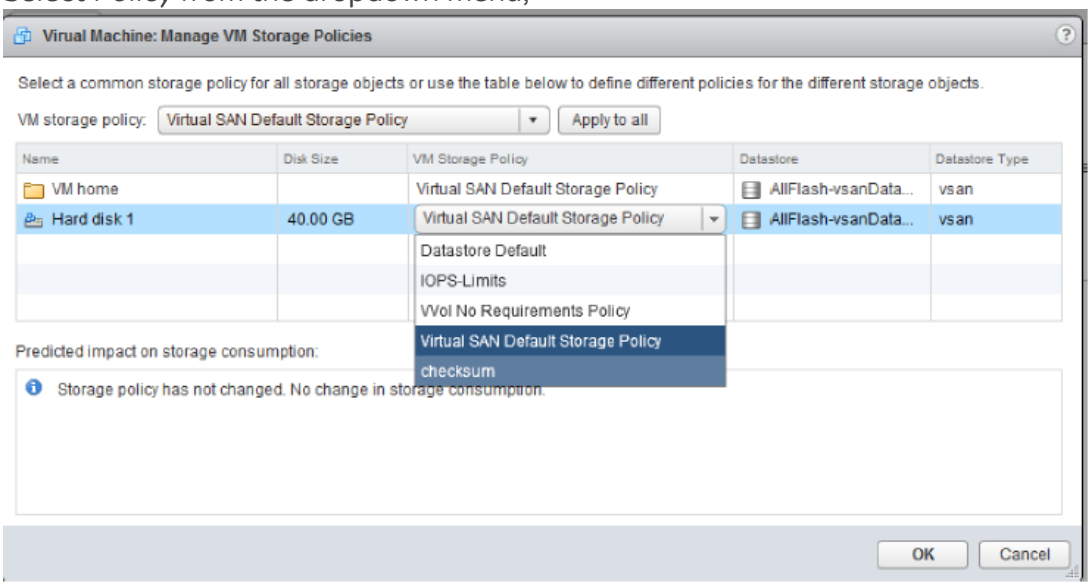
1. Open the vSphere Client.
2. Select a *Virtual Machine*, right-click and select **VM Policies**.
3. Click **Reapply VM Storage Policy**.



4. Click **Yes** to confirm.

Procedure - Per Object:

1. Open the vSphere Client.
2. Select an *object*, right-click and select **VM Policies>Edit VM Storage Policies**.
3. Select the desired *object* , e.g Home Namespace or Hard Disk.
4. Select *Policy* from the dropdown menu,



5. Click **OK** to finish.

11.5 Enabling Checksum on a VM or Object

Checksum is disabled by default, unless explicitly added as a rule in VM Storage Policy rule-set. If software checksum is intentionally disabled, it is simply a matter of removing the "Disable software checksum" rule from a VM Storage Polices rule-set.

Procedure:

1. Open the vSphere Web Client.
2. Click **Policies and Profiles > VM Storage Policies**.
3. Select a *Virtual Machine Policy*, right-click and select **Edit a VM storage policy**. You can either:
    - Modify the existing rule *Disable object checksum:* the value "No"

    - Simply remove the *Disable object checksum* rule from the rule set
4. Click **OK** to save the VM Storage Profile and decide to Apply now or manually later to a VM or object.

An Administrator can either

- Modify the existing rule "Disable object checksum: the value "No"
- Simply remove the "Disable object checksum" rule from the rule set

12. Performance Service Operations

Performance Service is a new feature introduced in vSAN 6.2. It allows for end-to-end monitoring of a virtual machine's performance, all the way down to physical disk level.

12.1 Performance Service Operations

Performance Service is a new feature introduced in vSAN 6.2. It allows for end-to-end monitoring of a virtual machine's performance, all the way down to physical disk level. It also provides two unique views of performance, both the front-end VM view, and the back-end vSAN view. This is easily explained if we take the example of a virtual machine with a RAID-1, mirrored VMDK object. If the VM generates 500 write IOPS, then at the back-end there will be 1,000 IOPS generated, 500 writes to each replica.

12.2 Enable Performance Service

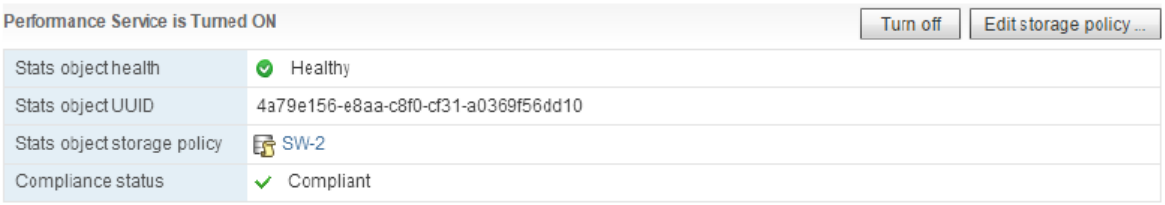
By default the performance service is disabled. To enable the performance service see the vSphere 6.5 Monitoring vSAN Performance for details.

Once you click on the edit button, you will be prompted to pick a policy from the list of existing VM Storage Policies. This will provide a degree of resilience to the stats database object, which is stored on the vSAN datastore. It means that the performance service can continue to function even if there is failure on the cluster. By default, the vSAN Default policy is chosen.

That completes the steps for enabling the performance service. Chart displaying performance metrics should now be visible in the various performance view for cluster, hosts and virtual machines. Note that these are normalized over a 5 minute period, so you will have to wait for at least 5 minutes for any meaningful performance data to appear.

12.3 Disable Performance Service

To turn off the performance service, navigate to the performance service as mentioned in the previous operation, and click **Turn off** as shown in the screenshot below. This disabled the performance service.



12.4 Change policy on Performance Service

To change the storage policy associated with the performance service, click on the edit storage policy button as shown above, and change the storage policy to the new policy from the dropdown list of policies. Click **OK**. The stats object will now be configured to take into account the new policy settings.

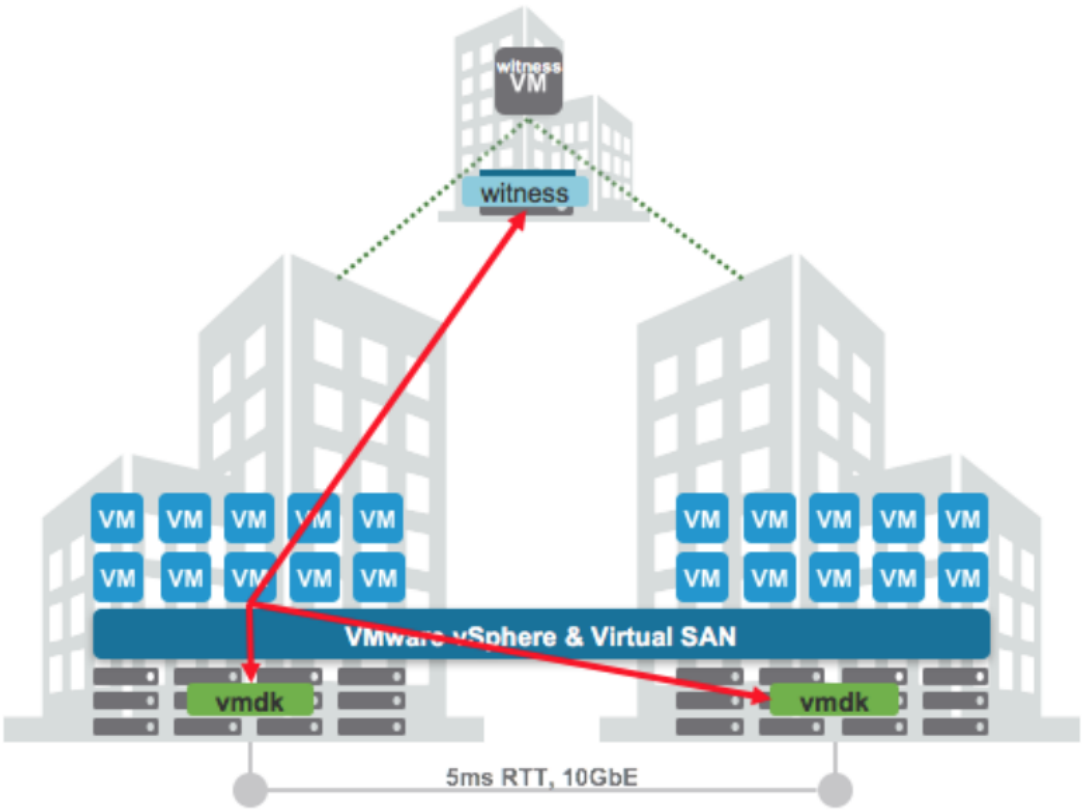
13. Stretched Cluster Operations

With vSAN you have the ability to stretch a cluster across distance.

13.1 Stretched Cluster Operations

With vSAN you have the ability to stretch a cluster across distance. At the time of writing the maximum distance is specified in Round Trip Time (RTT) latency, which for vSAN is 5ms at most between the sites hosting data. Before we dive in to some of the operational aspects we want to point out that we will not be going in to any significant level of depth in terms of architecture and design. There is a great white paper on this topic which can be found here.

The vSAN Stretched Cluster solution is based on Fault Domains. Instead of creating a fault domain per rack, now complete sites or data centers are considered to be a fault domain. The following diagram illustrates this situation. Note that there are two data centers/sites where data is hosted, and there is a requirement for a witness as well in a third location.

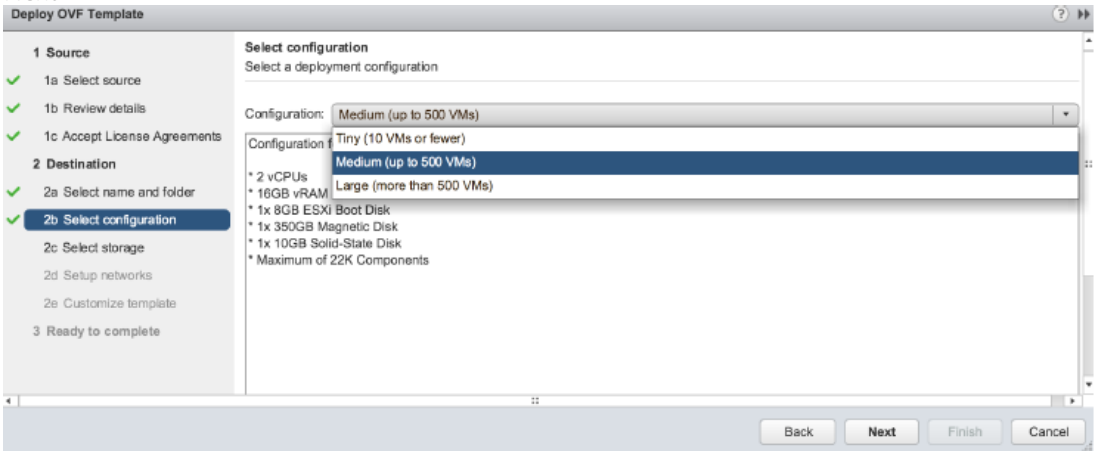


13.2 Deploying a Witness Appliance

The first step, of course, is downloading the Witness Appliance. It can be found here on the Download VMware vSAN Witness Appliance 6.5 page under "VMware vSAN Tools, Plug-ins and Appliances".

1. Download VMware vSAN Witness Appliance 6.5 and save it to your local drive.
2. Open the vSphere Web Client.
3. Click the **Hosts and Clusters** tab.
4. Right-click the *cluster* or *host* on which you want to deploy the witness appliance and click **Deploy OVF Template**.
5. Select the *.ova* file you downloaded and click **Next**.
6. Review the details and click **Next**.
7. Accept the License Agreement and click **Next**.
8. Enter the *name* for the Witness Appliance and select the *folder* or data center where it lands and click **Next**.

9. Depending on the size of your environment select the Witness Appliance configuration, click **Next**.



10. Select the VM Storage Policy (when applicable) and the Datastore the Witness Appliance needs to be stored on and click **Next**.
11. Select a network for the management network. This gets associated with both network interfaces (management and vSAN) at deployment, so later on the vSAN network configuration will need updating. Click **Next**.
12. Give a root password for the witness ESXi host and click **Next**.
13. Review the selected configuration and characteristics and click **Finish**.

When deployment has finished there are a couple of steps that will need to be taken before the Witness Appliance can be used to finalize the configuration:

1. Right-click the *Witness Appliance* and click **Edit Settings**.
2. The Network for the second Network Adapter needs to be changed. It is currently set to the network selected during the provisioning, the vSAN network segment needs to be selected.
3. Select the vSAN Network segment and click **OK**.

Now the Appliance can be powered on. After power on the Management VMkernel interface of the Witness Appliance should be changed unless having DHCP available. The steps to do this are:

1. Open the *Witness Appliance VM Console*
2. Press **F2** and go to the *Network Adapters* view
3. On Network Adapters ensure there is at least 1 vmnic selected for transport.
4. Navigate to the IPv4 Configuration section. This will be using DHCP by default. Select the static option as shown below and add the appropriate IP address, subnet mask and default gateway for this witness ESXi's management network.
5. The next step is to configure DNS. A primary DNS server should be added and an optional alternate DNS server can also be added. The FQDN, fully qualified domain name, of the host should also be added at this point.

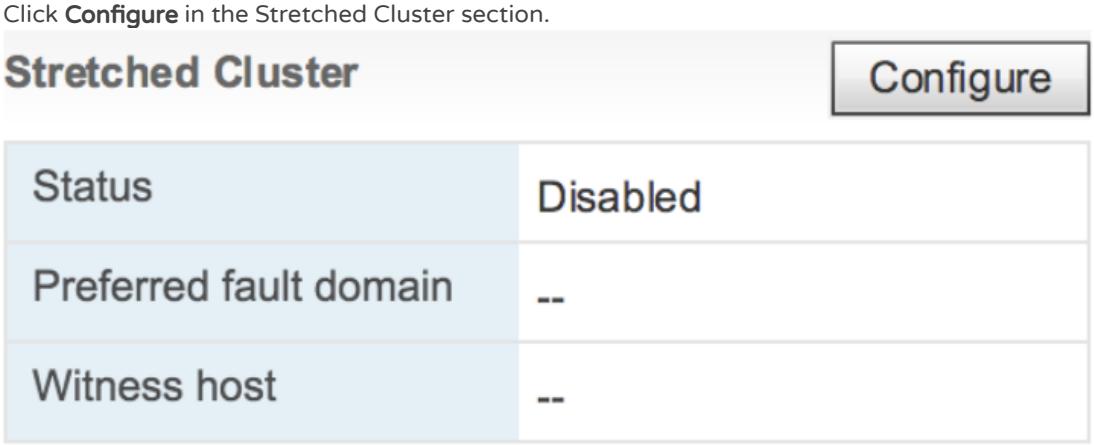
Next, the Witness Appliance can be added to vCenter Server as a regular vSphere host. Note that it is also needed to configure the vSAN VMkernel Interface, this can, however, be done through the Web Client just like with a normal vSphere host. If you are not familiar with how to do this, see KB 2058368.

13.3 Configuring a Stretched Cluster

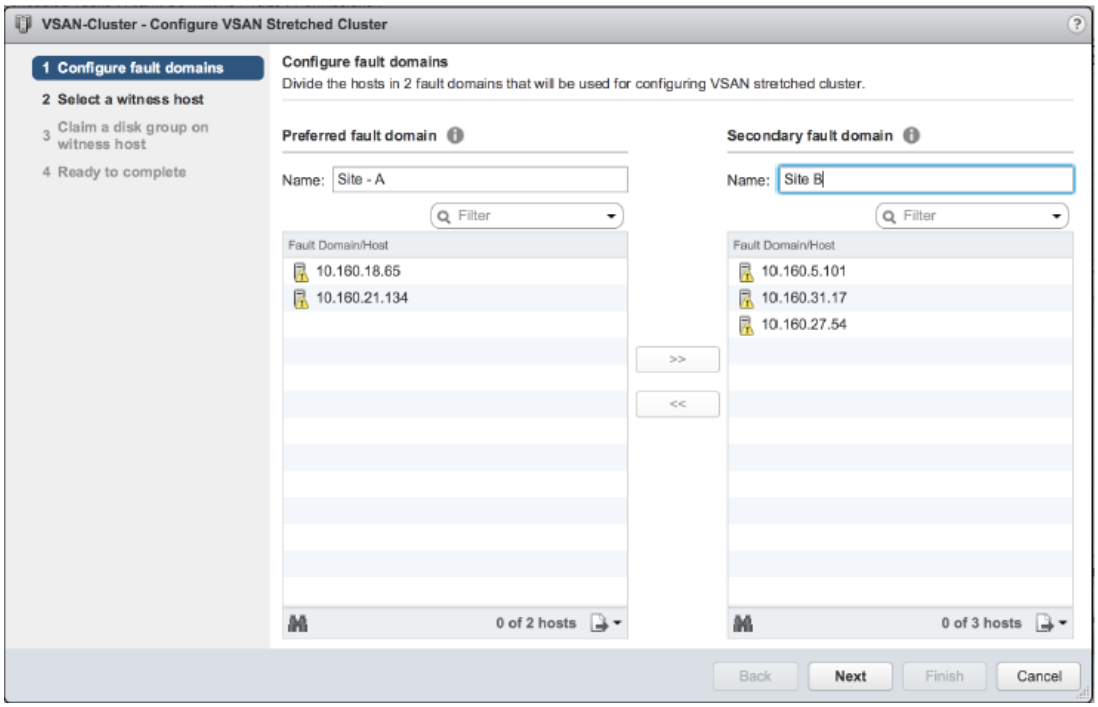
Configuring a stretched cluster from a vSAN point of view is very easy. It takes a couple of minutes and the steps are described below:

1. Open the vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the *vSAN cluster*.
4. Click **Manage>Fault Domains & Stretched Cluster**.

5. Click **Configure** in the Stretched Cluster section.



6. Provide a *name* for both "data" sites/fault domains.
7. Select the *hosts* for each of the sites and click **Next**.



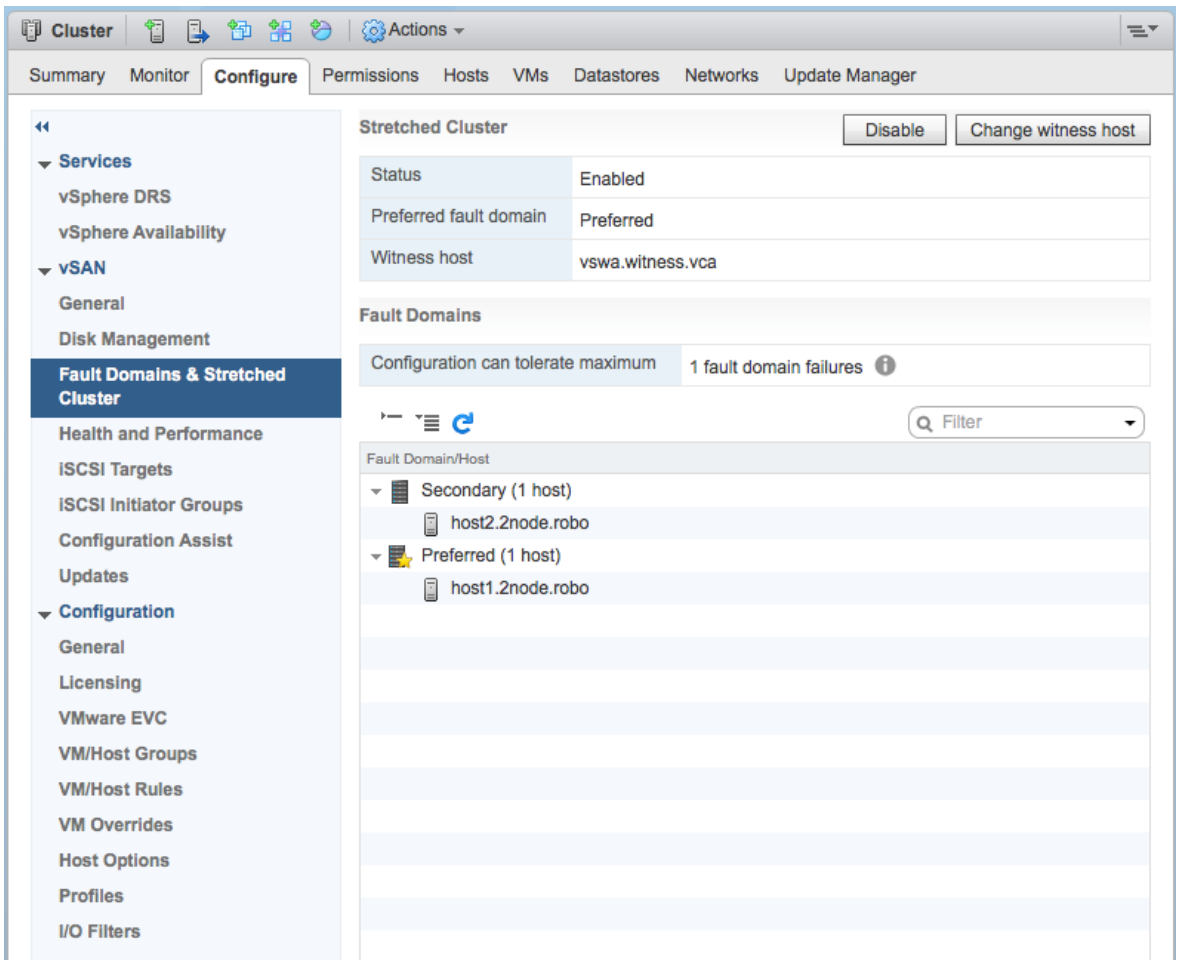
8. Select the *Witness host,* this can be the Appliance or a host you have installed and click **Next**.
9. Select the *Caching and Capacity devices* for the Witness. This is used to store the witness on. Click **Next**.
10. Review the configuration and when correct click **Finish**.

13.4 Replacing a Witness Appliance

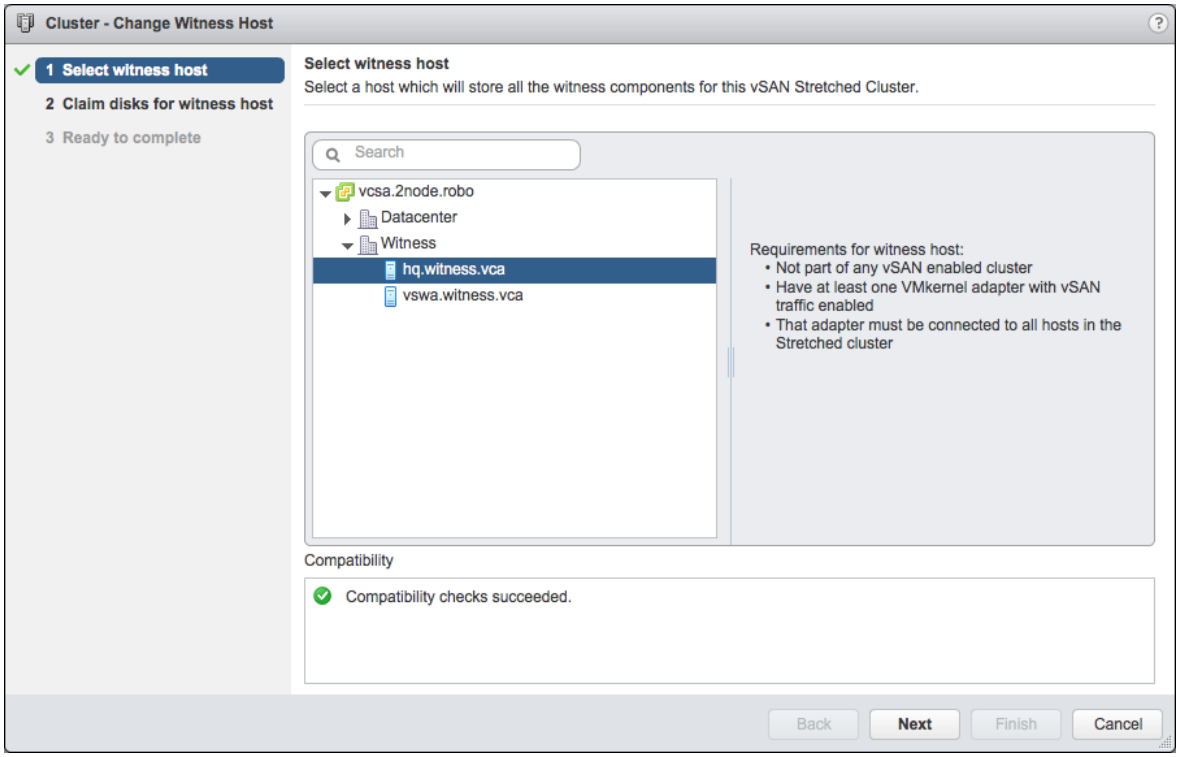
vSAN 6.1-6.5
In vSAN versions previous up to 6.5, it is not possible to replace a Witness. Administrators can introduce a new Witness by disabling Stretched Clustering and re-enabling the Stretched Cluster with the same sites using the new Witness.
How to do this is described in Configuring a stretched cluster and Deploying a Witness Appliance. Please refer to those sections.

vSAN 6.6
In vSAN 6.6, the capability of replacing the Witness has been added to the vSAN UI.

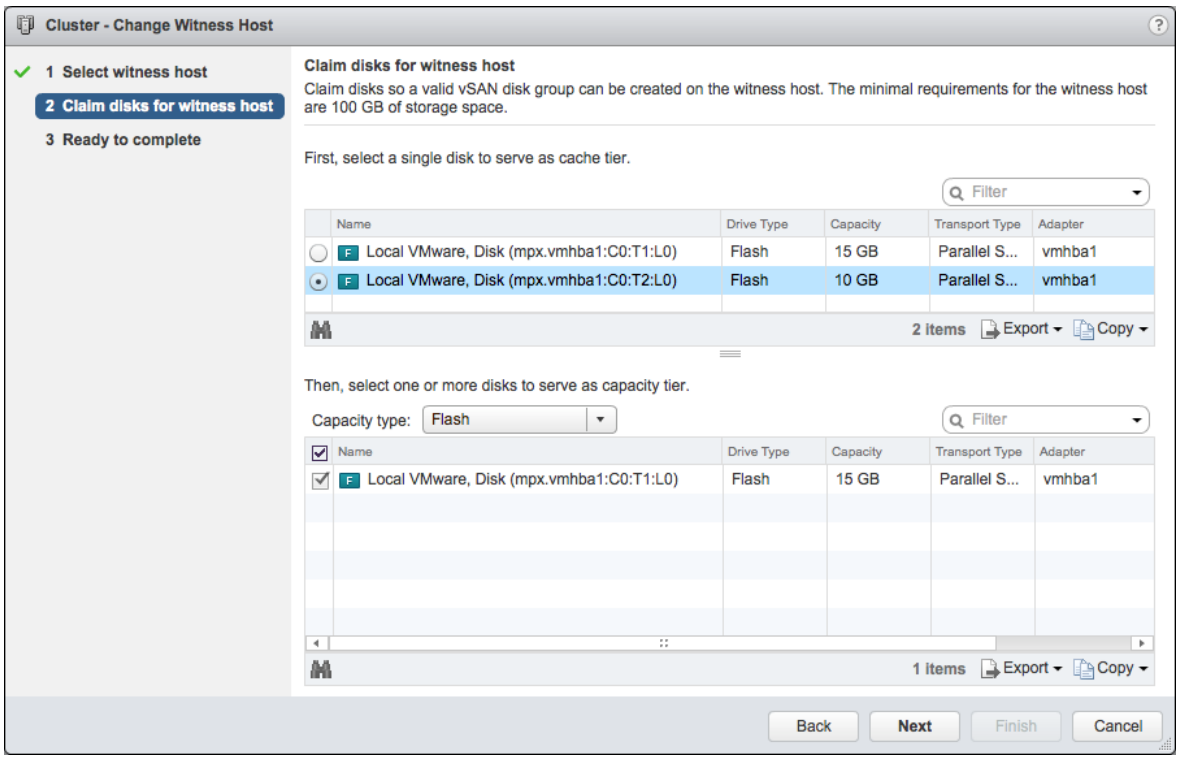
To change the Witness, select the **Change witness host** button



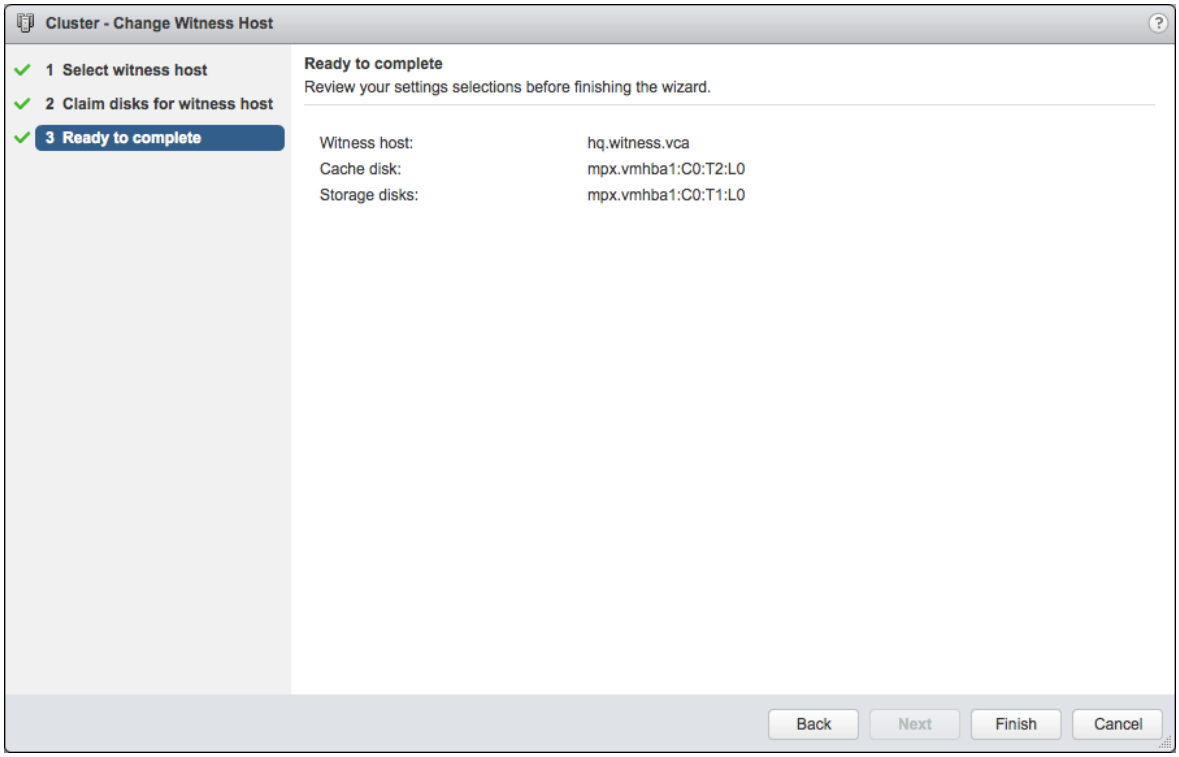
Select the new Witness and insure compatibility checks succeed.



If using the vSAN Witness Appliance, select the 10GB disk for the cache tier, and the 15GB disk for the capacity tier.



Complete the Witness replacement.



13.5 DRS Settings

VMware recommends enabling DRS in an entirely automated fashion in a stretched cluster and using Affinity Rules to control the placement of Virtual Machines. How to enable and configure DRS is described in the vSphere 6.5 Edit Cluster Settings.

13.6 HA Settings

VMware recommends enabling HA to allow for fully automated restarts of workloads in a stretched cluster configuration. How to enable HA is described in the vSphere 6.5 Create a vSphere HA Cluster. There are several additional recommended settings for a stretched cluster as described in the Stretched Cluster Guide in particular and we will describe how to set these here:

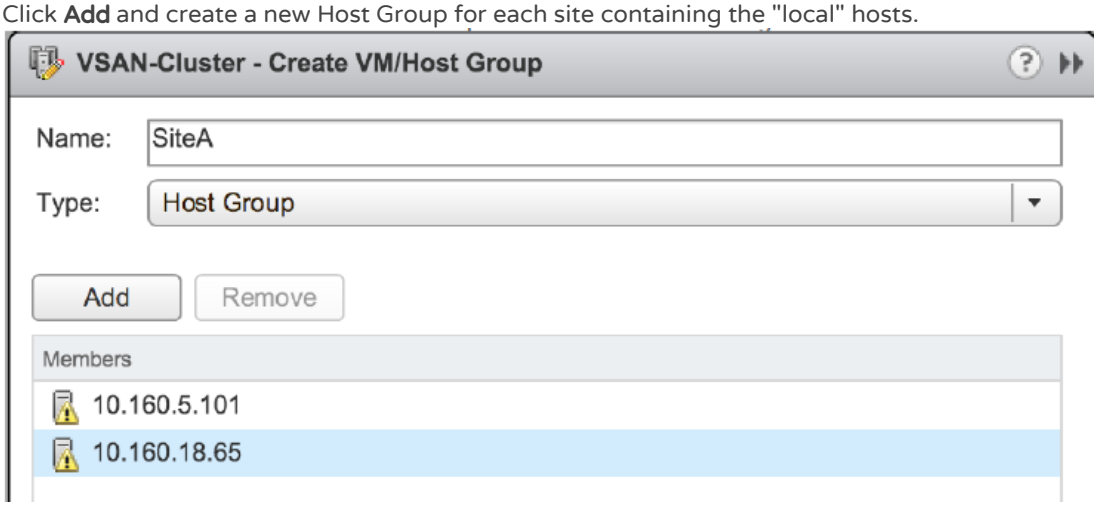
1. Open the vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the *vSAN cluster*.
4. Click the **Manage** tab and go to **vSphere HA**.
5. Click **Edit** and **enable** HA if not yet enabled and click **OK**.
6. When HA is correctly configured, we will now set some for the recommended advanced settings by clicking **Edit** again.

7. Expand the **Failure conditions** and **VM response** section and set **Response for Host Isolation** to "*Power Off and Restart VMs*".
8. Expand the **Admission Control** option and set it to *Define a fail-over capacity by reserving a percentage of the cluster resources.*
    - Set the resource to 50% for both memory and CPU as that is the only way to guarantee a restart after a full site failure.
9. Expand the **Advanced Options** section.
10. Add the following advanced options, underneath each we will explain what it is used for.
    - `das.isolationaddress0`
    - `das.isolationaddress1`
    - Isolation Address is used when a host has been isolated. We need to set one per site, which needs to be site local so that even in the case of a site isolation and a potential host isolation we have a local address to verify isolation against.
    - `das.useDefaultIsolationAddress=false`
    - This disables the use of the default gateway for isolation purposes. In most stretched cluster environments it is preferred to use a reliable site local address.
11. Click **OK**.

13.7 Affinity Rules

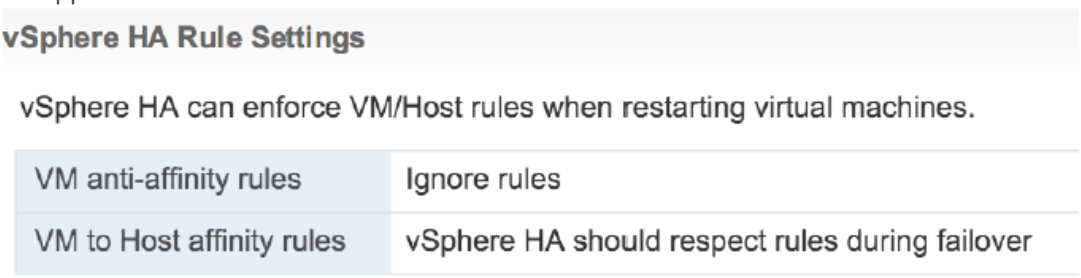
VM/Host Affinity rules can be created to ensure that VMs always reside on a certain side of the stretched cluster. This can be useful for example for Active Directory, where multiple AD hosts will be located at both sites so that in the case of a full site failure the service is still available. This can be configured as follows:

1. Open the vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the *vSAN cluster*.
4. Click the **Manage** tab and go to **VM/Host Groups**.
5. Click **Add** and create a new Host Group for each site containing the "local" hosts.



6. Click **Add** and create a new VM Group for each site containing the VMs that need to reside on that site.
7. Click **VM/Host Rules**.
8. Click **Add**.
9. Provide a *name* for the rule and select *Virtual Machines to Hosts* from the type drop down.
10. Make sure to select "Should run on hosts in group" as "must rules" to prevent HA from restarting VMs when a full site failure occurs.

11. Click **OK**.
12. Now the rule has been created, click **Edit** on the "vSphere HA Rule Settings".
13. Select "vSphere HA should respect VM to Host affinity rules during failover" and click **OK.** This will ensure that when a single host fails, VMs will be restarted on one of the host specified in the applicable rule.



13.8 Decommissioning a Stretched Cluster

It is possible to decommission a stretched cluster configuration. The following steps should be taken to do so:

1. Open the vSphere Web Client.
2. Click the **Hosts and Clusters** tab.
3. Select the v*SAN cluster.*
4. Click the **Manage** tab and go to the **Fault Domains & Stretched Cluster** section.
5. Click **Disable** and confirm the decommissioning by clicking **Yes**.
    - This will remove the witness host, but will leave 2 fault domains in tact
6. Remove the two remaining fault domains by selecting the Fault Domain in the Fault Domain view and click the red **X**.
7. Confirm the removal of the fault domain by clicking **Yes**, and repeat this for the second Fault Domain.

In order to ensure full availability for you virtual machines it is highly recommended to immidiately repair your objects. As the Witness Appliance has been removed all witness components are missing for your workloads. You can recreate these instantly as follows:

1. Click on the **Monitor** tab and click on **vSAN**.
2. Click on **Health** and check the “vSAN object health” under vSAN Object Health.
    - Most likely it will be “red” as the “witness components” have gone missing. vSAN will repair this automatically by default in 60 minutes.
3. Click **repair object immediately,** now witness components will be recreated and the vSAN cluster will be healthy again.
4. Click **retest** after a couple of minutes

14. Upgrading vSAN

There is a specific order that needs to be applied for a successful upgrade.

14.1 Upgrading vSAN

Upgrading vSAN is a two phase operation:

- vCenter Server upgrade and vSphere host upgrade
- vSAN Object and Disk format upgrade

There is a specific order that needs to be applied for a successful upgrade. The general order of events would be to perform vCenter upgrade first, (which may include VMware Update Manager), followed by vSphere host upgrade. vSAN Object conversion and disk format conversion or DFC, which may be considered to be the longest operation. This guide will focus primarily on vSAN object and disk format conversion

The guidance would be to verify the following:

1. Backup the Virtual Machines hosted on a vSAN cluster.
2. Verify that you have enough capacity to tolerate a failure and data evacuation prior to performing a disk format conversion.
3. Verify all hosts are healthy and not in maintenance mode.
4. Verify all software, hardware, drivers, firmware, and storage I/O controllers are on the vSAN HCL.

Upgrading vCenter server in a vSAN Cluster

regardless if vCenter Server is hosted on a vSAN enabled cluster, the upgrade process is agnostic to this fact and has no dependence on vSAN General guidelines would be:

- Read the vSphere Release notes for known issues. for example VMware vCenter Server 6.0 Update 2 Release Notes
- Ensure your system meets the minimum hardware and software requirements. Check the following:
  - VMware Product Interoperability requirements before the upgrade. VMware Product Interoperability Matrixes
  - VMware vCenter Server supported host operating systems. Supported host operating systems for VMware vCenter Server installation
  - vSphere vCenter Upgrade Requirements

Upgrading ESXi hosts in a vSAN Cluster

The main caveat for ESXi hosts is that all components are on VMware vSAN Hardware Compatibility guide. Strict adherence is required to ensure a successful upgrade. Ensure the hardware you plan on using are supported by vSAN 6.2 and later, and are listed on the VMware vSAN Compatibility Guide. It is of extreme importance that all the software and hardware components are supported, but specifically:

- Storage I/O controllers
  - Drivers and firmware verified to be on vSAN HCL
- Disks and SSDs
  - Minimum supported firmware is verified on the vSAN HCL.

Read the vSphere Release notes for known issues. For example: VMware ESXi 6.0 Update 2 Release Notes

Upgrading the on-disk format of vSAN 5.5 on-disk format from V1 to V3.

There are two major parts during the on-disk format upgrade, software prerequisites must include:

1. vCenter 6.0 U2
2. ESXi 6.0 U2
3. strict adherence to vSAN HCL driver and firmware guidance

Part I

10% to 15% is fixing object aliment in preparation for for on-disk format v3 features. There are two sections to this. The first part is realigning objects and their components to have a 1mb address space. This is specific to on-disk format V1. The second section is realigning vsansparse objects to be 4k aligned and upgrading objects from v2 to v2.5. Depending on how many objects there are, this can take considerable time

Part II

The second and final part is the actual on-disk format process itself. This includes three parts, on per disk group basis.

- Data evacuation of a disk group.
- Removal of a disk group.
- ReAdd of a diskgroup.

This process will be repeated on a per diskgroup basis and can take considerable time. As outlined above, part I and part II are part of the on-disk Format Upgrade.

Upgrading the on-disk format of vSAN 6.0 / 6.0 U1 on-disk format from V2 to V3.

There are two major parts during the on-disk format upgrade. Software Prerequisites must include:

1. vCenter 6.0 U2
2. ESXi 6.0 U2
3. strict adherence to vSAN HCL driver and firmware guidance

Part I

10% to 15% is fixing object aliment in preparation for for ondisk format v3 features. There are two sections to this. The first part is realigning objects and their components to have a 1MB address space. This is specific to on-disk format V1. The second section is realigning vsansparse objects to be 4KB aligned and upgrading objects from v2 to v2.5. Depending on how many objects there are, this can take considerable time

Part II

The second and final part is the actual ondisk format process itself. This includes three parts, on per disk group basis:

1. Data evacuation of a disk group.
2. Removal of a disk group.
3. Re-add of a diskgroup.

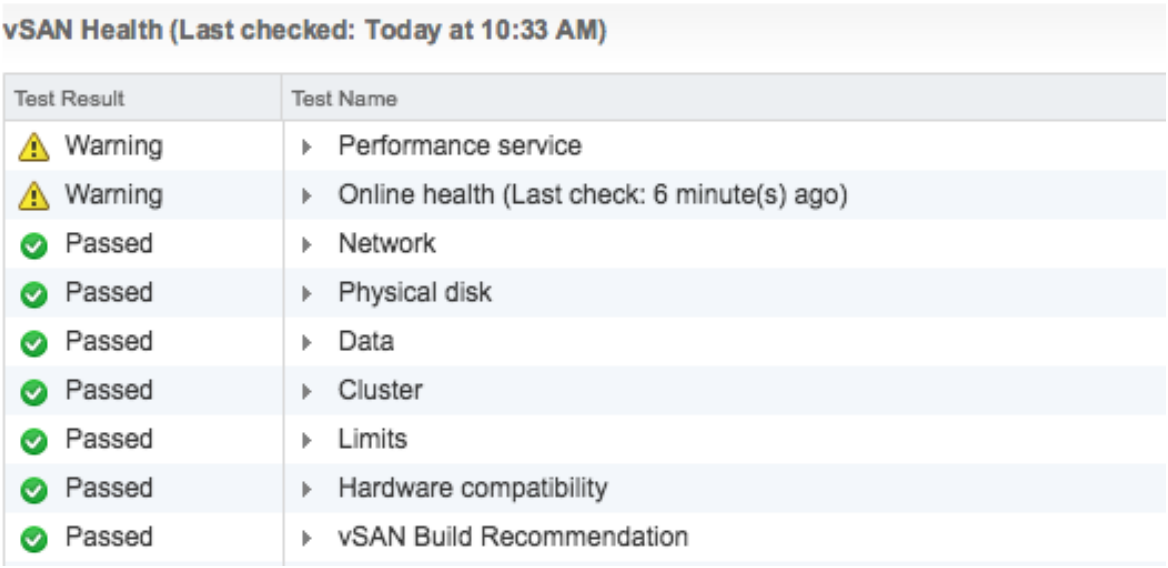
This process will be repeated on a per diskgroup basis and can take considerable time. As outlined above, Part I and Part II are part of on-disk Format Upgrade.

15. Monitoring vSAN

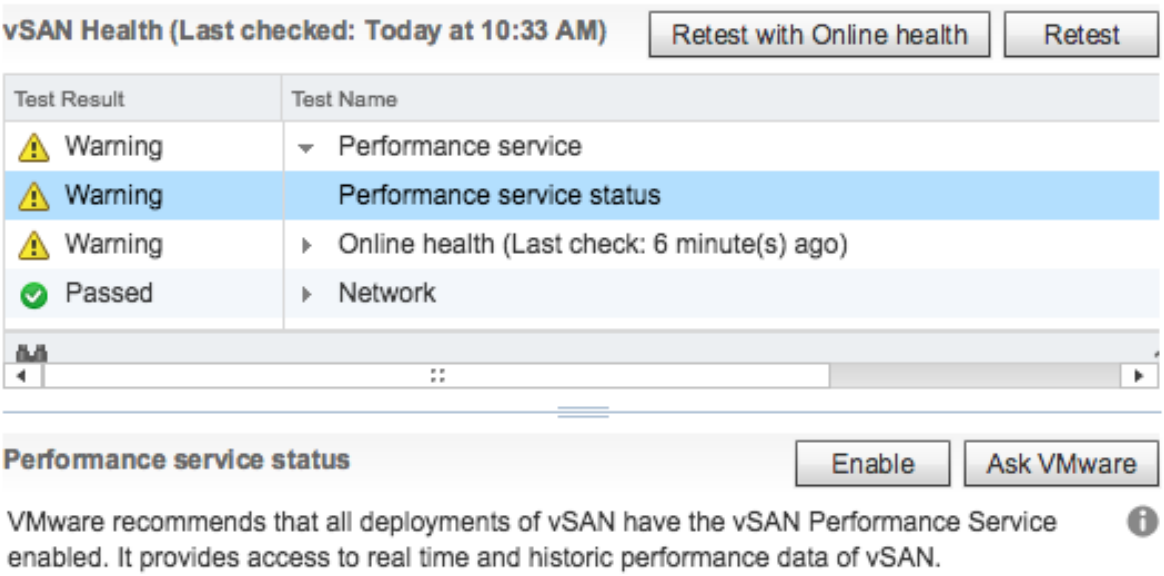
There are various placed that vSAN can be monitored.

15.1 Monitoring vSAN Cluster Heath

There are a few ways to monitor the health of a vSAN cluster. The primary method is using vSAN Health in the vSphere Web Client shown in the following figure.



As you can see, a number of items are monitored in this central location. If there are issues, these are surfaced to the top of the list for visibility. Expanding a vSAN Health item provides more details. In many cases, an Ask VMware button is provided, which is a direct link to the relevant VMware Knowledge Base article. Some items also include a button to resolve the issue. An example of these buttons is shown below.

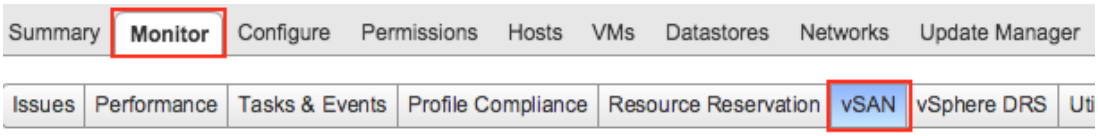


In the case above, clicking the Enable button provides the option to turn on vSAN Performance Service without the need to navigate elsewhere in the vSphere Web Client. these features can help reduce troubleshooting efforts and resolution times.

To access vSAN Health, perform these steps:

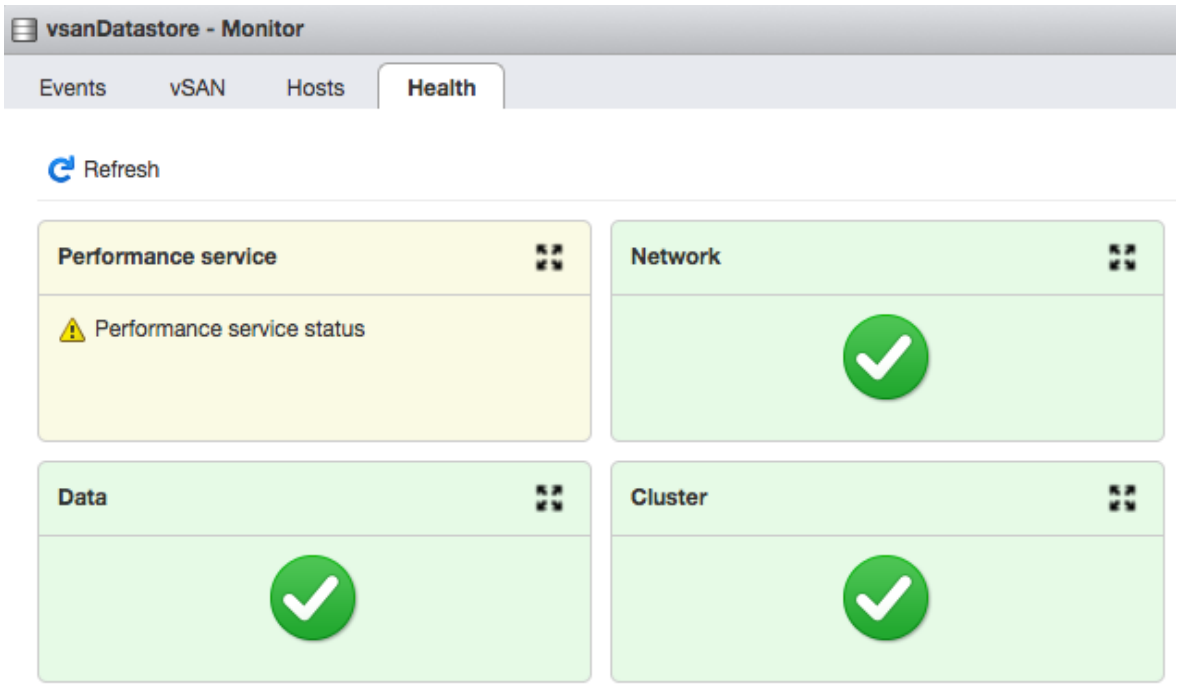
1. Log in to the vSphere Web Client.

2. Click Hosts and Clusters.
3. Click a cluster in the Navigator column.
4. Click the Monitor tab.
5. Click vSAN



6. Click Health.

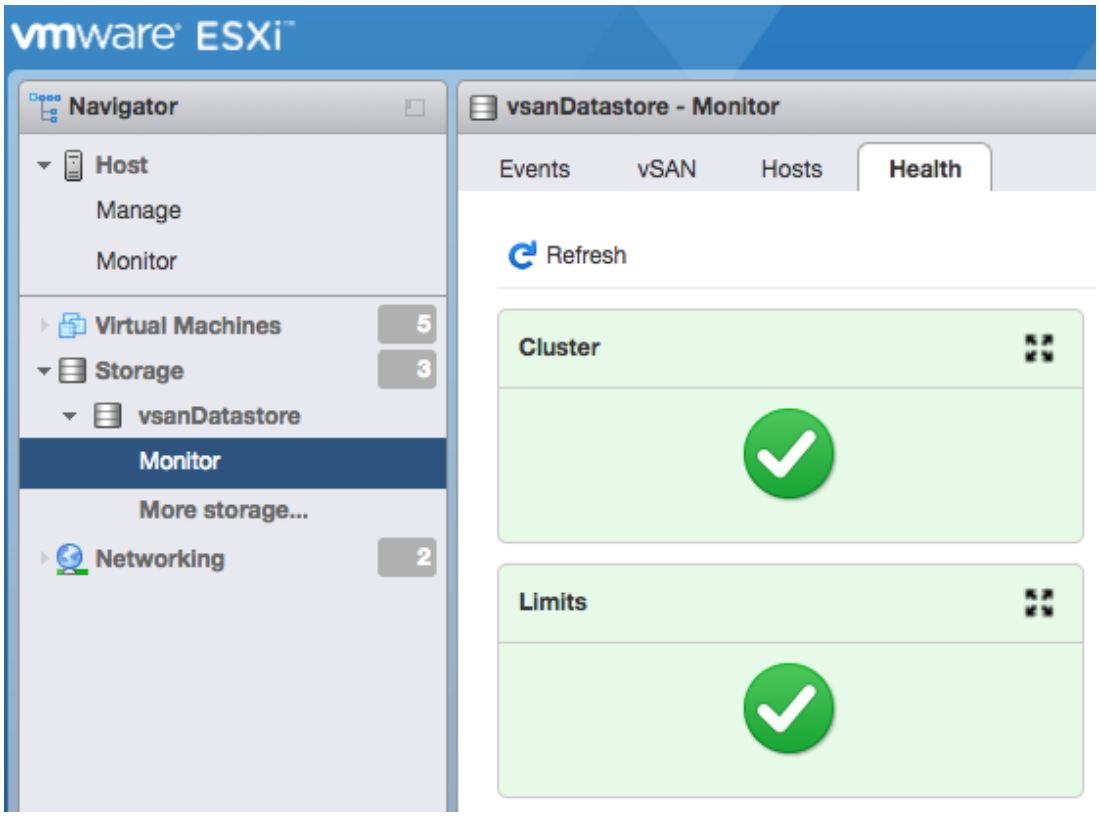
A variation of vSAN Health is available in the vSphere Host Client. While this option does not provide as many details or options for remediation, it does give administrators a way to monitor the health of a vSAN cluster when vCenter Server is offline. The figure below shows an example of vSAN Health in the vSphere Host Client. It shows there is an issue with the vSAN Performance service. In this case, it was not turned on.



View vSAN Health in the vSphere Host Client by following these steps:

1. Connect a web browser directly to a vSphere host, e.g., https://vsphere-host-name (or IP address).
2. Enter the local credentials for the host, e.g. Username: root and the corresponding password.
3. Click Storage in the Navigator column.
4. Click the vSAN datastore in the list of datastores. This will expand more items under Storage in the Navigator column.
5. Click Monitor in the Navigator column.

6. Click the Health tab.



In addition to graphical user interface (GUI) for monitoring vSAN health, a command line interface (CLI) can be used. Here are a few examples of esxcli commands and output:

```
esxcli vsan health cluster list
```

```
Health Test Name                                     Status
---------------------------------------------------  ----------
Overall health                                       green (OK)
Cluster                                              green
  ESXi vSAN Health service installation              green
  vSAN Health Service up-to-date                     green
  Advanced vSAN configuration in sync                green
  vSAN CLOMD liveness                                green
  vSAN Disk Balance                                  green
  Resync operations throttling                       green
  Software version compatibility                     green
  Disk format version                                green
Network                                              green
  Hosts disconnected from VC                         green
  Hosts with connectivity issues                     green
```

```
esxcli vsan health cluster get -t "vSAN cluster partition"
```

```
Partition list
Host                 Partition      Host UUID
---------------------------------------------------------------------
10.144.97.86         1              58c8898d-f132-5cdc-7772-002590c61436
10.144.97.85         1              58c95966-c4ad-73e5-c9bf-002590c61478
10.144.97.87         1              58c975b0-2ec2-54f5-972a-002590c61472
10.144.97.88         1              58c97d3b-3f5b-98cf-f583-002590c61434
```

PowerCLI is another good solution as demonstrated in this article: https://www.altaro.com/vmware/how-to-generate-vsan-html-report-powercli/

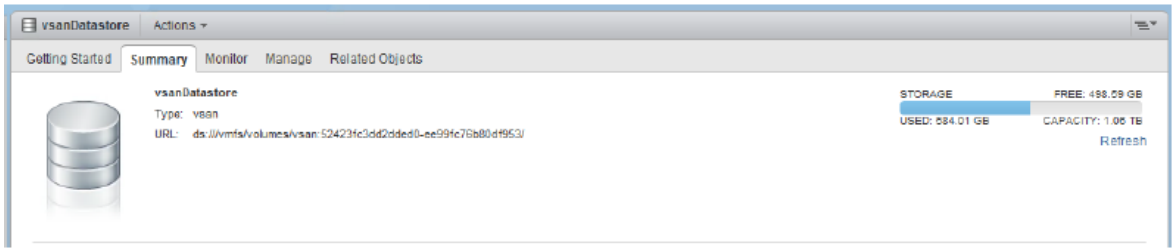
There is a Python health check script on the vCenter Server Virtual Appliance (VCSA):

```
python /usr/lib/vmware-vpx/vsan-health/vsan-vc-health-status.py
```

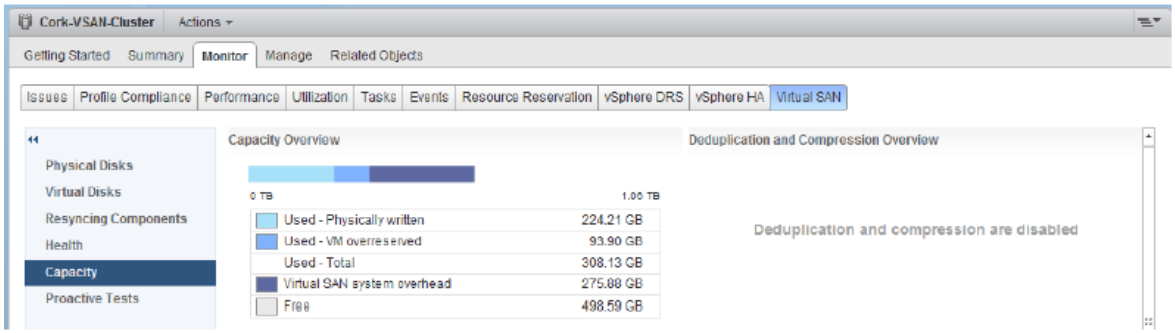
Additional options for monitoring vSAN Health include vSAN Management SDKs , the Ruby vSphere Console (RVC), and vSAN Observer.

15.2 Monitoring vSAN Datastore Capacity

The capacity of the vSAN datastore can be monitored from a number of locations. First, one can select the datastore view, and view the summary tab for the vSAN datastore. This will show you the capacity, used and free space.

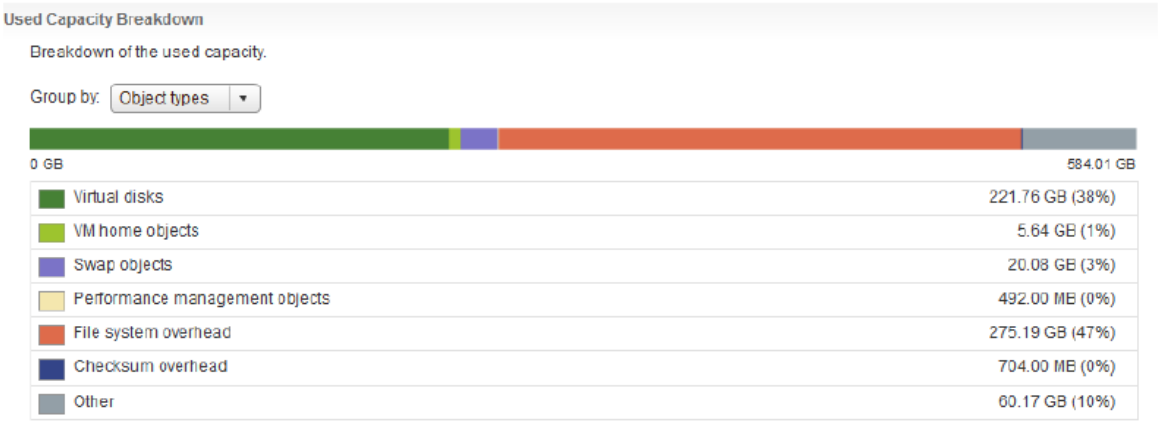


In vSAN 6.2, because of the new data services introduced, there is also a view of the vSAN datastore in the Cluster > Monitor > vSAN > Capacity view. This gives a more granular break down of what is consuming space on the vSAN datastore, including overheads.



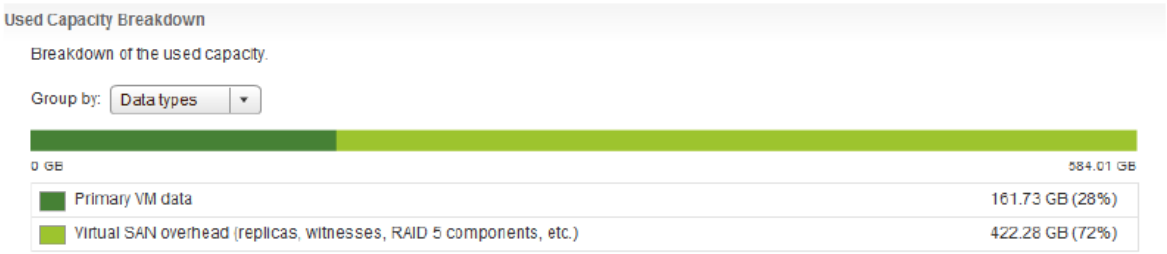
In this example, we can also see the vSAN system overhead. What we can also see in the "Used - VM overreserved" metric how much space have been provisioned for virtual machines, and not yet consumed. This is a hybrid system do deduplication and compression are not enabled.

On the same view, there is also a way to break down the capacity usage into space consumed by object types and space consumed by data types.



These are all the different object types one might find on the vSAN datastore. We have VMDKs, VM Home namespaces, and swap objects for virtual machines. We also have performance management objects when the performance service is enabled. There are also the overheads associated with on-disk format file system, and checksum overhead. Other refers to objects such as templates and ISO images, and anything else that doesn't fit into a category above.

To see a different view where capacity is monitored grouped by data type, the following is the breakdown of data types one may see:

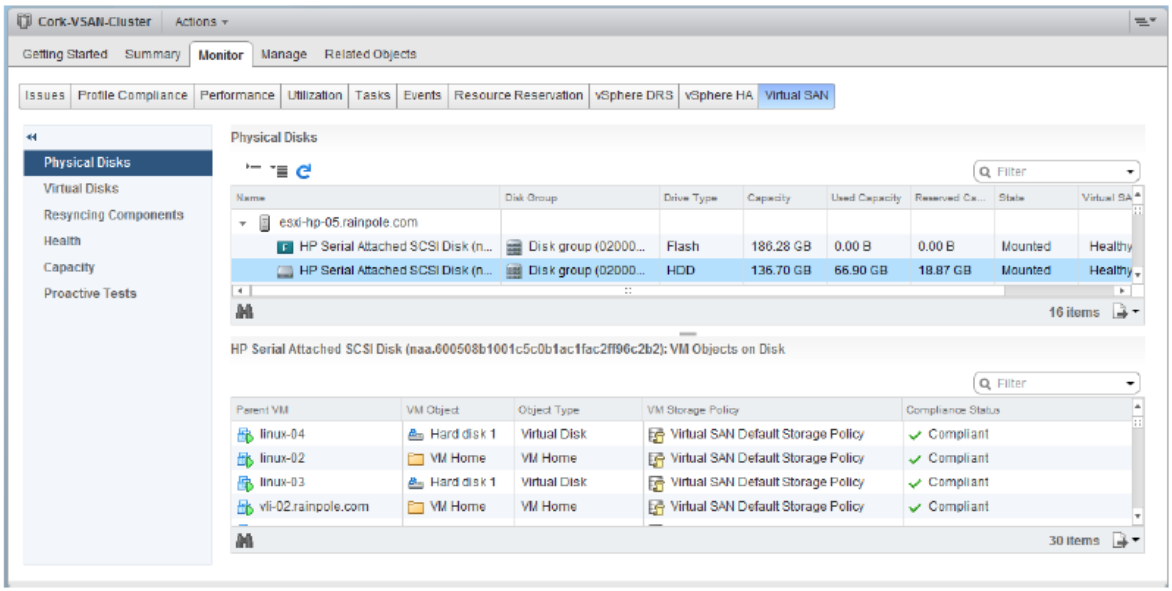


In this view, we can see how much data is taken up for VM data, and then, depending on the policy, we can see any capacity consumed to create replica copies of the data, witness components or RAID-5/ RAID-6 parity components.

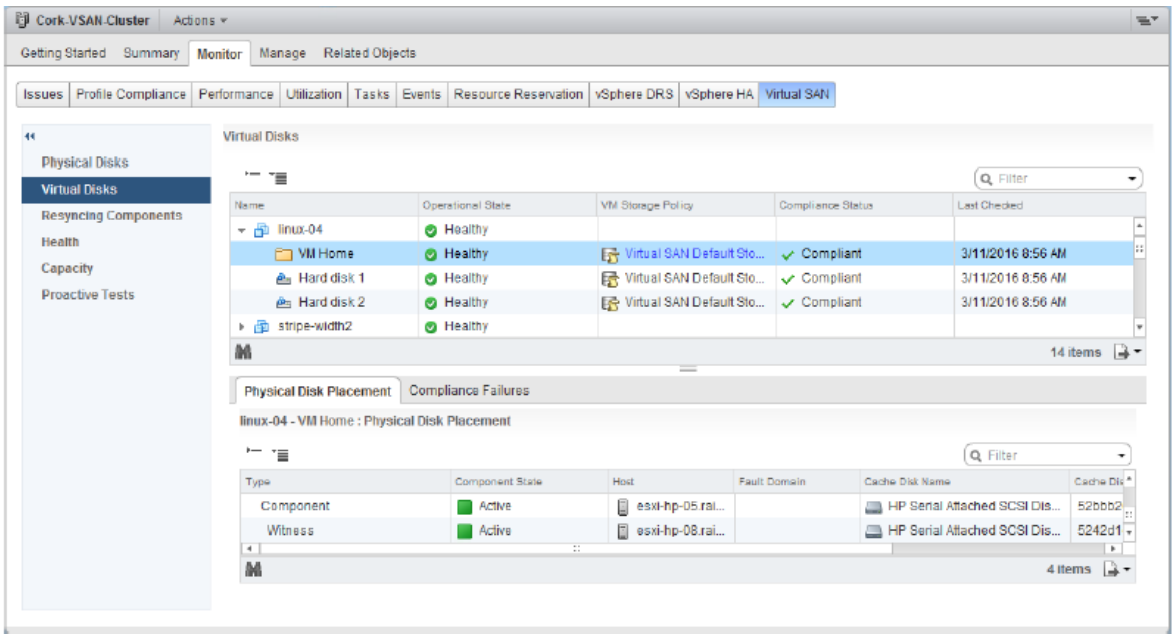
15.3 Monitoring Disk Capacity

There are a number of places where disk capacity and usage can be observed. The Cluster > Manage > Settings > vSAN Disk Management is one such place, where the individual devices that go to make up the cache tier and capacity tier of each disk group are shown.

Another place where the physical disks can be viewed in the Cluster > Monitor > vSAN > Physical Disks view. If any of the capacity tier devices are selected, you can see both the capacity and the consumed size, as well as a list of the components residing on the device.



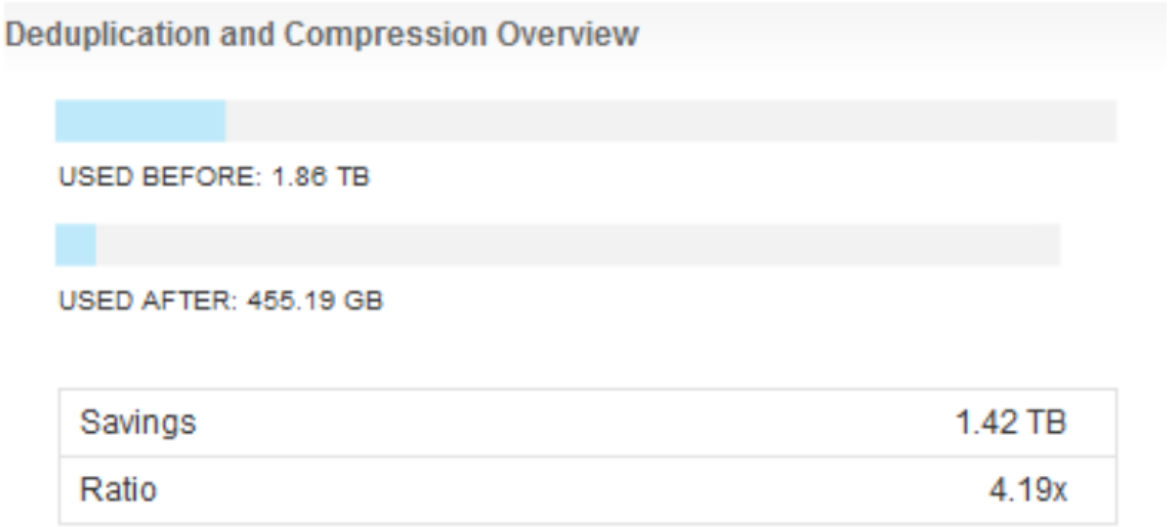
In the same screen, there is a Virtual Disks View. This displays a list of virtual machines deployed on the vSAN datastore, and if a virtual machine is selected and expanded, a list of objects that make up the virtual machine is displayed. By selecting one of the objects, the components and their makeup (RAID-0, RAID-1, RAID-5, RAID-6) is displayed and the location of each component (disk and host) is presented.



15.4 Monitoring Dedupe/Compression

Note that this feature is only available on all-flash vSAN configurations running version 6.2 and later. In a previous section, we saw how this view looked on hybrid vSANs, where Deduplication and Compression are disabled. When deduplication and compression are enabled, this view displays how much capacity has been saved by deduplication/compression (Savings). This can also be used to determine how much space would be required to re-inflate the deduped and compressed data if these

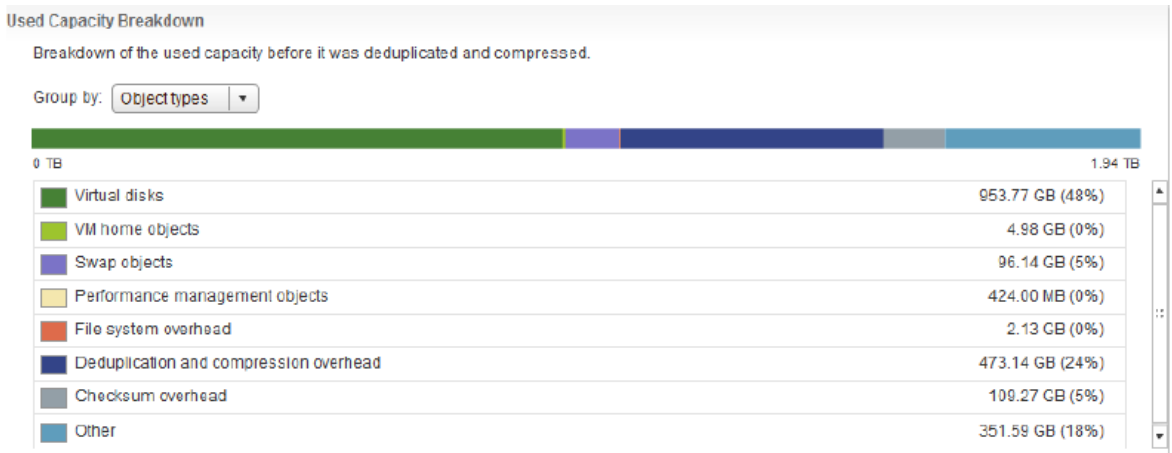
features are once again disabled. Also shown is the dedupe/compression ratio currently achieved on the system.



The overhead of deduplication and compression can also be seen in the Used Capacity Breakdown, Group by Object types.

15.5 Monitoring Checksum

Checksum overhead can be seen in the Capacity view. In the Used Capacity Breakdown, Group by Object types, checksum overhead is displayed.

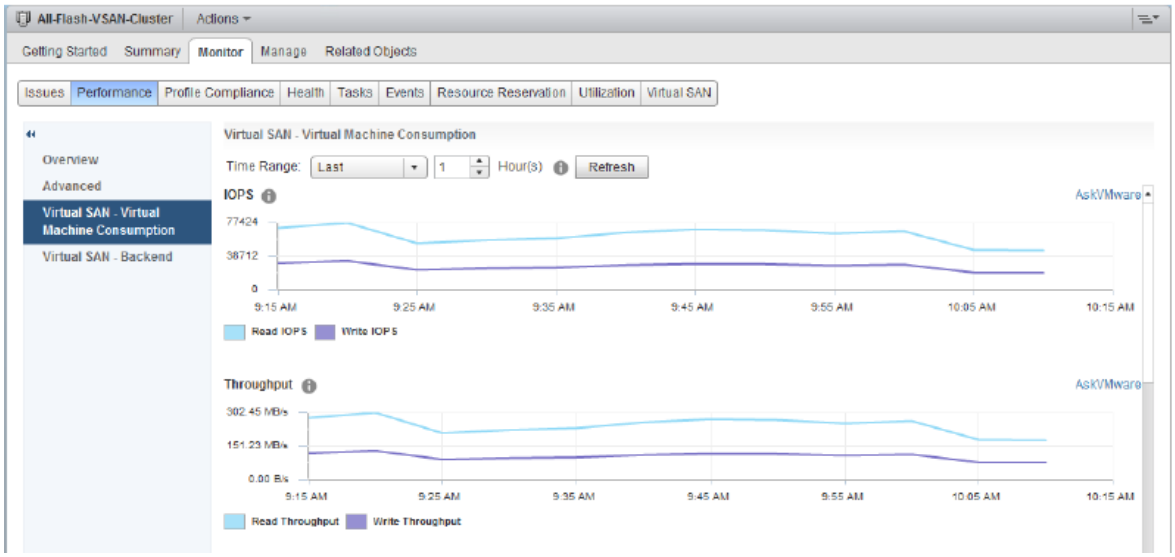


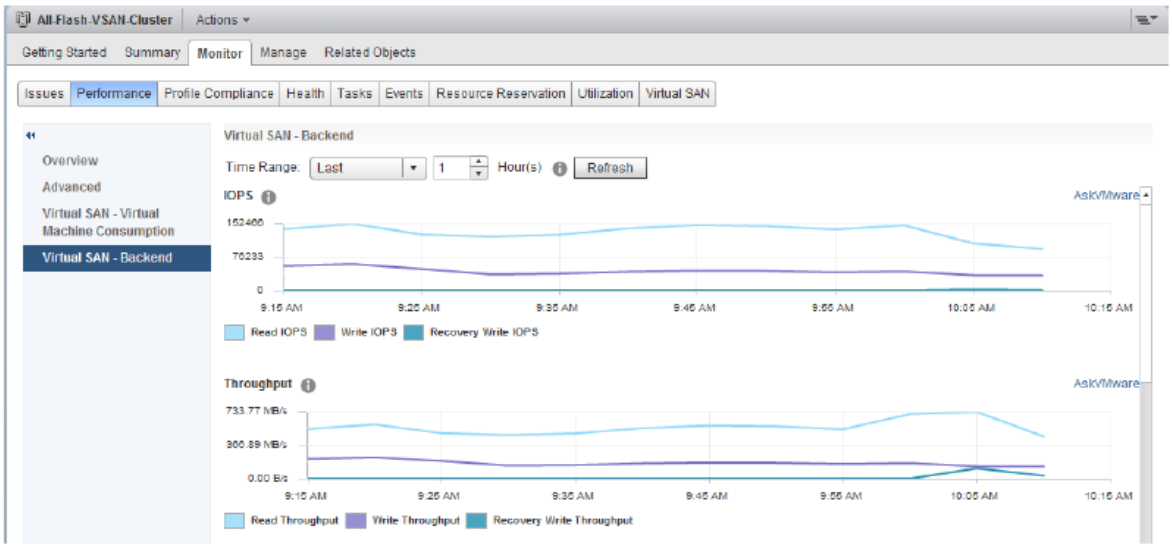
15.6 Monitoring vSAN with the Performance Service

vSAN 6.2 introduced a new performance service. This allows administrators to view performance end-to-end on a vSAN. The Performance Services gives visibility at the cluster, host, disk group, disk and VM perspective. There is also visibility into front-end virtual machine performance, and back-end vSAN performance. For example, if a VM is generating 500 write IOPS, and the VMDK is in a RAID-1 mirrored configuration, then there will be 1,000 IOPS at the back-end, 500 to each replica.

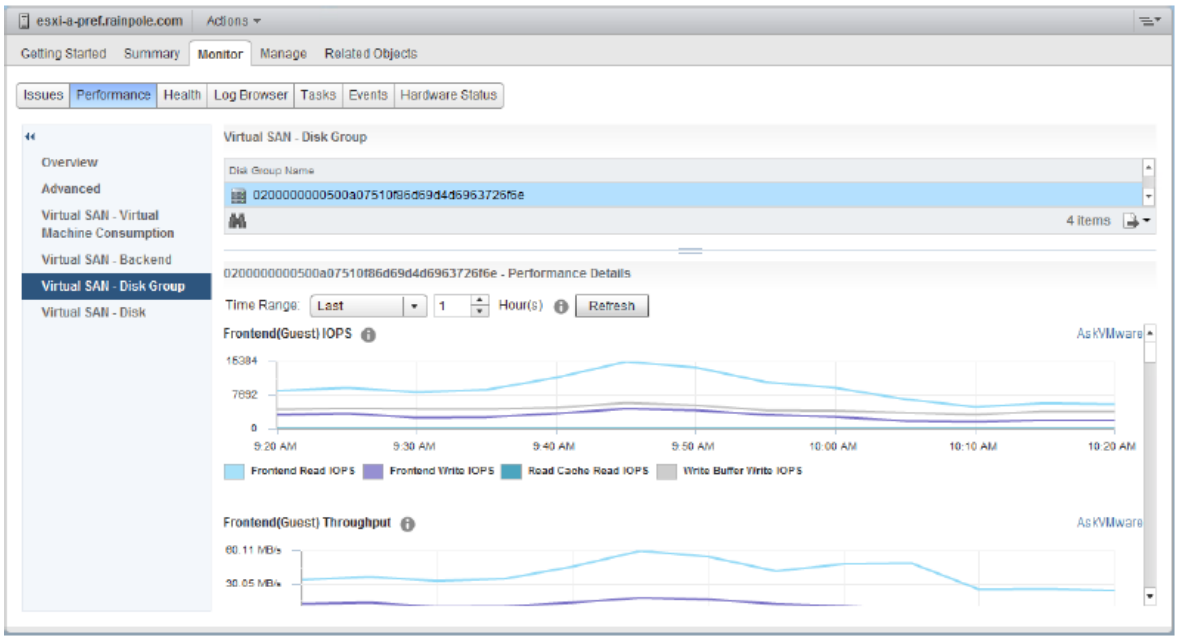
To look at a specific object's performance (cluster, host, VM), select the object in the vCenter inventory, Monitor tab, then Performance. There are two views: vSAN - Virtual Machine Consumption

and vSAN - Backend. There are a number of metrics available, such as IOPS, Throughput, Latency, etc. Below are two screenshots taken at the cluster level, from a front-end (VM) and back-end (vSAN) perspective.

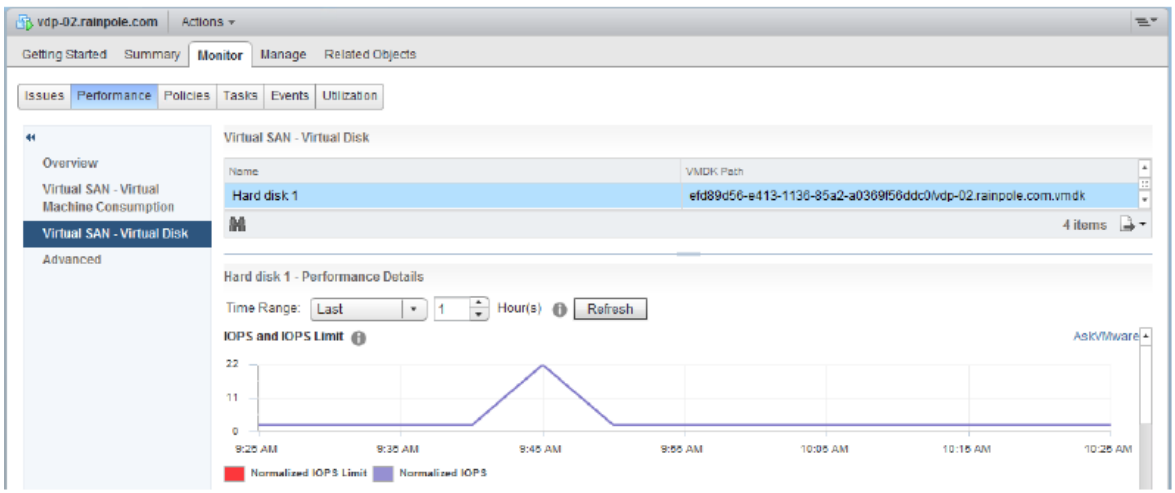




Similar views are available at the ESXi host level. Also included at the ESXi host view are performance views into disk groups and disk devices. The can be found by selecting the host object, then Monitor, Performance and then vSAN - Disk Group as shown below.

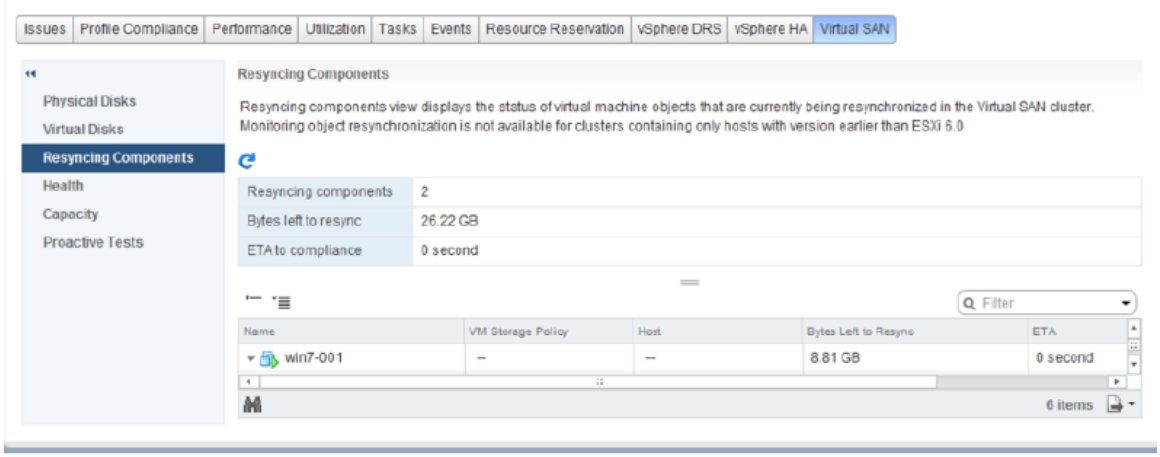


The final set of performance views relate to virtual machines. To view the performance of a virtual machine running on vSAN , select the VM, then Monitor, Performance and the appropriate view. Below is the Virtual Disk view.



15.7 Monitoring Resync Activity

Resync activity can be triggered for any number of reasons. It may be a host being placed into maintenance mode, and the administrator selects to evacuate all of the data from the host, or even ensure accessibility mode, and there are VMs with number of failures to tolerate set to 0. It could also be due to a change in the policy associated with a VM or an object, when a new object needs to be created to meet the new policy requirements. This new set of components then needs to be synchronized with the original components before those original components can be discarded. Of course, another scenario where there is resync activity due to a failure in the cluster. To monitor resync activity, select the vSAN cluster, then Monitor, vSAN and then Resyncing Components.



15.8 Configure Alarms/Traps/Emails

Configuring alarms, emails and SNMP traps on vSAN events are identical to how an administrator would do it for generic vCenter events. The procedure to do this is documented in the vSphere Administration Guide.

- To create alarms, refer to vSphere Create or Edit Alarms.
- To create SNMP traps as an alarm action when an alarm is raised, refer to vSphere Send SNMP Traps as an Alarm.
- To send an email as an alarm action when an alarm is raised, refer to vSphere Send Email as an Alarm Action.

16. vRealize Operations Manager

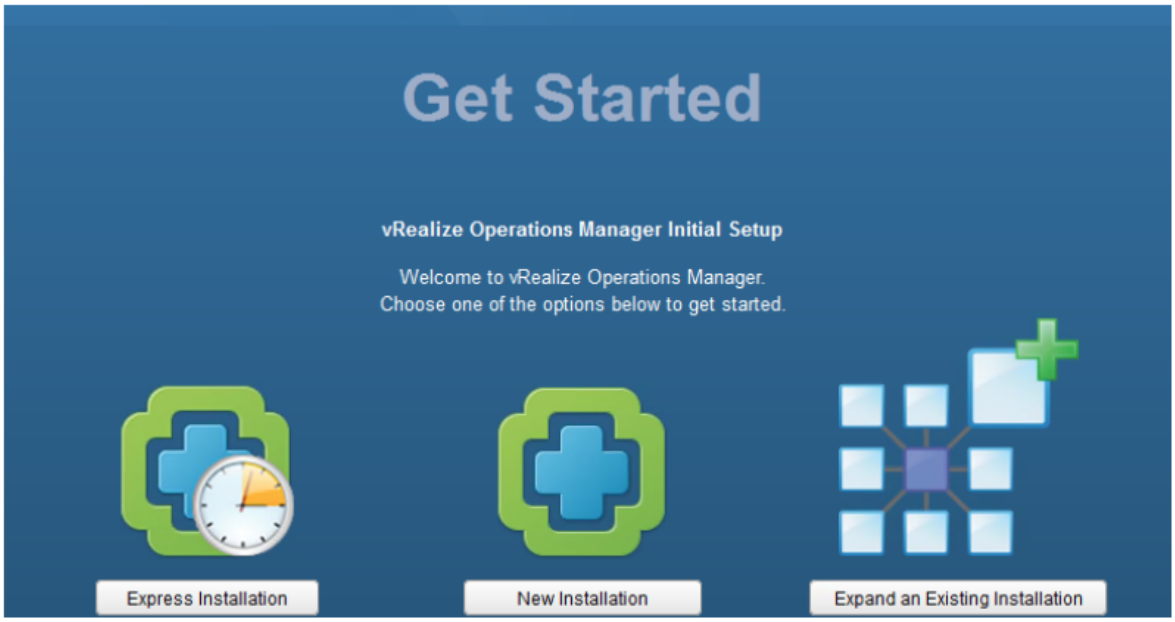
In this section, we will show how vSAN integrates with vRealize Operation Manager (VROps).

16.1 vRealize Operations Manager

In this section, we will show how vSAN integrates with vRealize Operation Manager (VROps). We will also show the steps to deploy and configure the Management Pack for Storage Devices (MPSD) which include a number of dashboards for reviewing vSAN Performance.

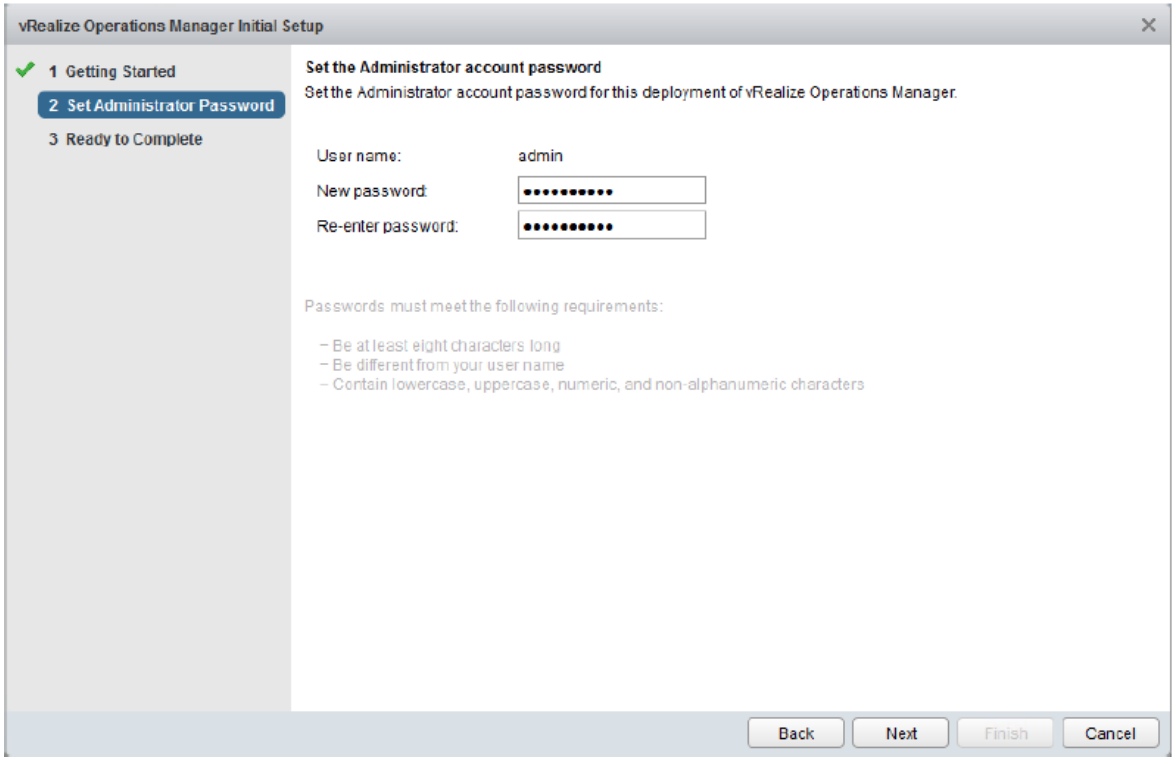
16.2 Deploy vRealize Operations Manager

In this example, the OVA associated with vROPs is deployed. Once successfully deployed and powered on, the administrator is presented with a number of options, including an express install, a new installation or an expansion to an existing installation.



In this example, an express installation is shown. For information on "New Installation" or to "Expand an Existing Installation", please refer to the official vROPs documentation found on https://www.vmware.com/support/pubs/vrealize-operations-manager-pubs.html.

In the express installation, the only item of note that an administrator needs to add is the password for the admin login:

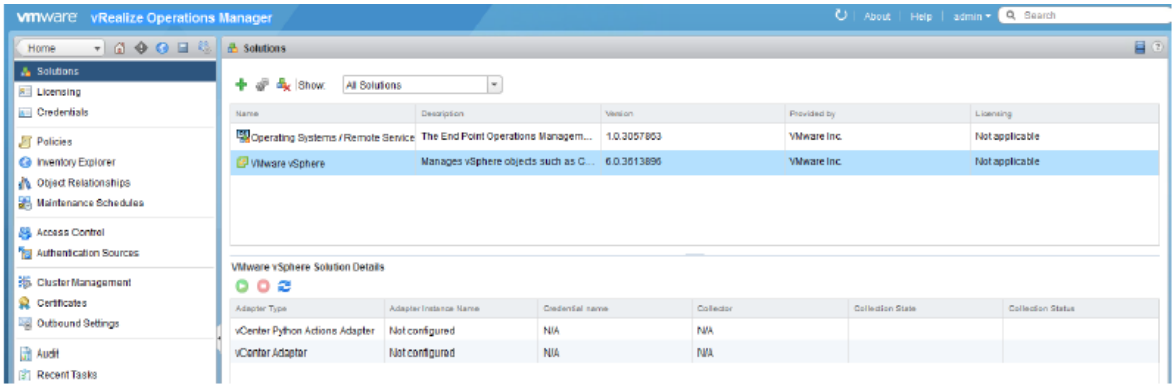


When the install is completed, the vROps smarts are installed and once completed; the administrator is prompted with a login prompt to log on to vROps and complete the configuration:

The next steps are to setup vROps to begin monitoring the vSAN/vSphere cluster. After that, the MPSD can be added and configured for vSAN specific dashboards.

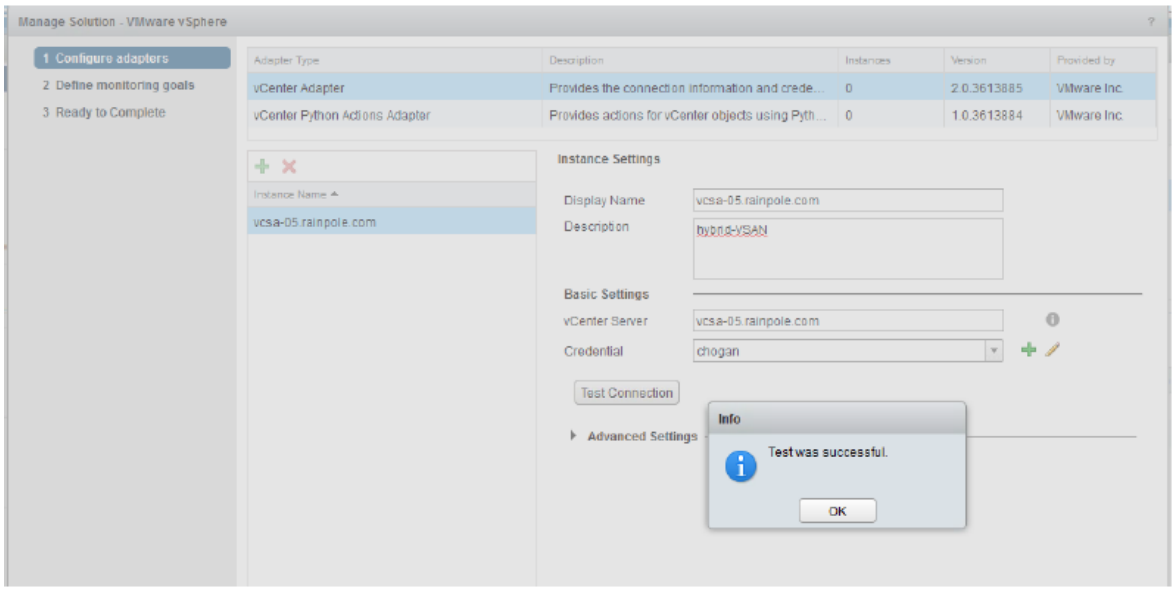
16.3 Configure vROps to Monitor vSphere

When the admin user logs into vROps, you are dropped to the management view initially. Here you will see the VMware vSphere solution. However, it will not be collecting any information until it is given an environment to monitor:

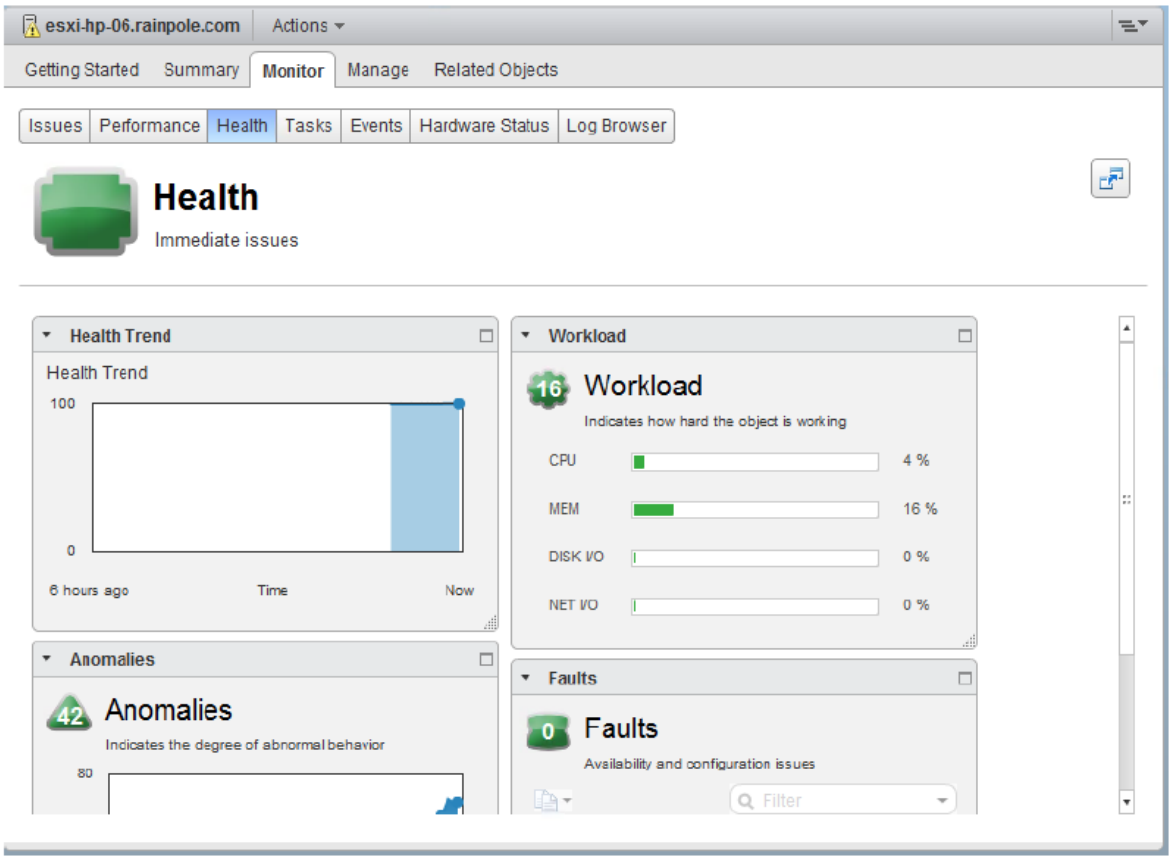


To begin the task of having this vROps instance monitor a vSphere environment (in particular our vSAN environment), simply click on the VMware vSphere solution, and then click on the configure icon (looks like a wheel/cog).

You will then be prompted to provide the details of the vCenter server managing the vSphere/vSAN environment, along with appropriate credentials. There is also an option to test the connection to ensure that the vCenter server and credentials are all functioning as expected. As you can see below the test was successful.



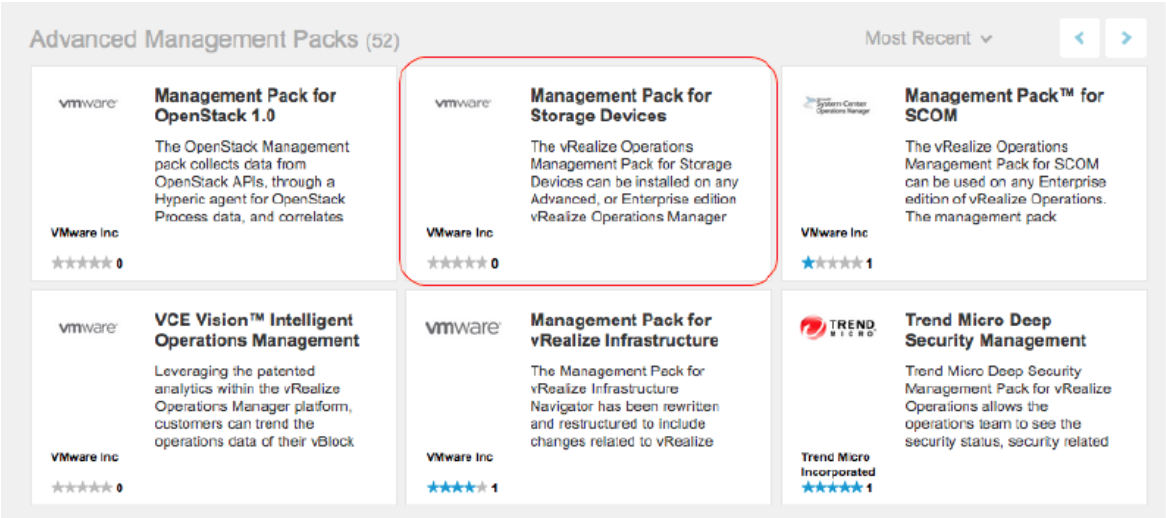
You should now save these settings. vROps now begins collecting information from vSphere, and after some minutes some useful metrics should begin to appear in the dashboards. On the vSphere web client, in each of the objects such as cluster, host, virtual machines, you should now see vROps counters related to health begin to appear:



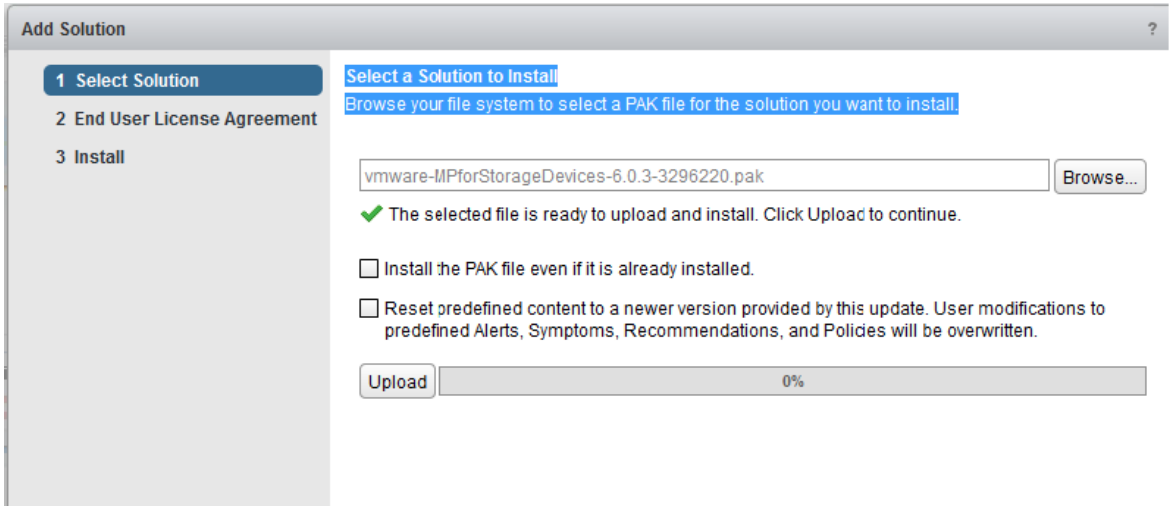
16.4 Install the Management Pack for Storage Devices

Now that vROps is monitoring the vSphere cluster, a management pack that looks at storage devices (including vSAN) can be installed and configured. This will allow an administrator to monitor some more specific vSAN metrics.

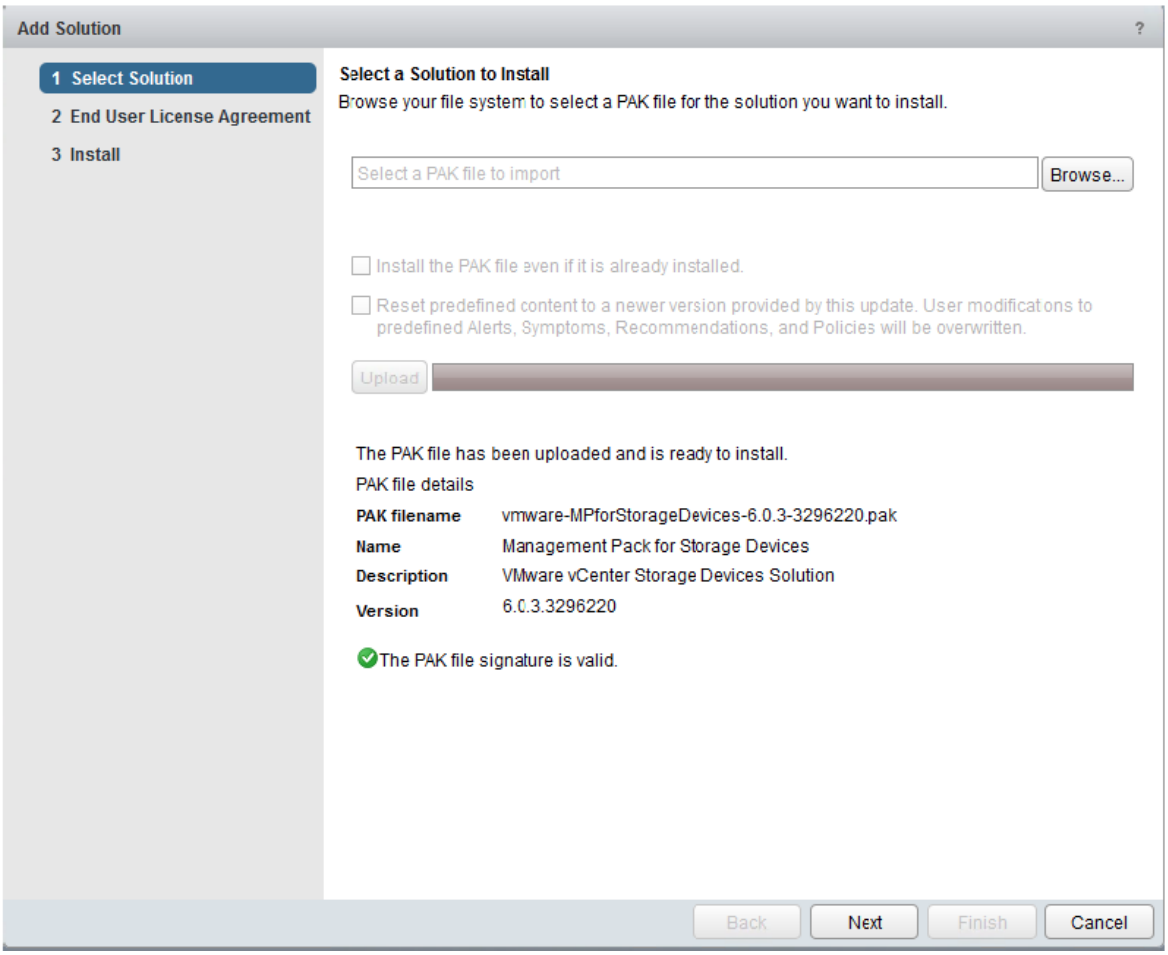
The management pack can be found on the VMware Solution Exchange. In the vSphere Operations section, Advanced Management Packs, you will find the MPSD:



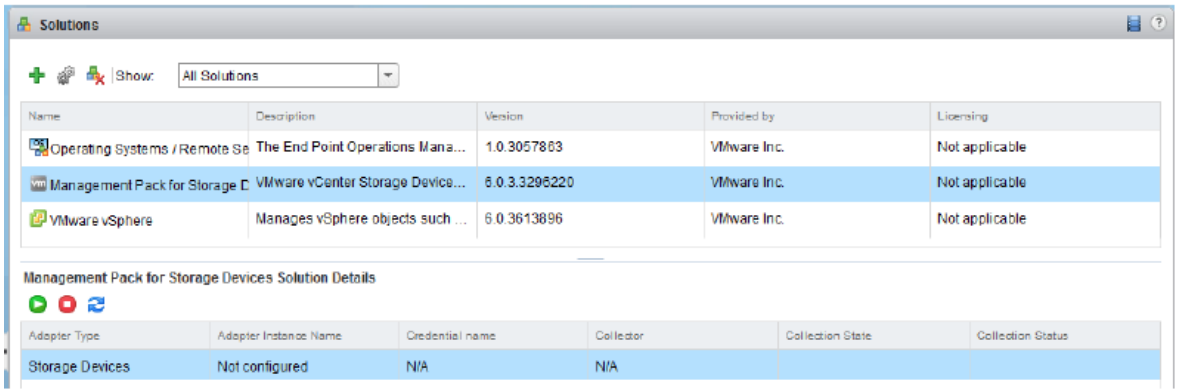
From the solutions view in vRops, click on the green + symbol to install a new solution. Use the browse button to select the MPSD '.pak' file:



Next, click on the "Upload" button. When the PAK file is fully uploaded, you should see details similar to those below, and a message to indicate that the PAK file signature is valid.



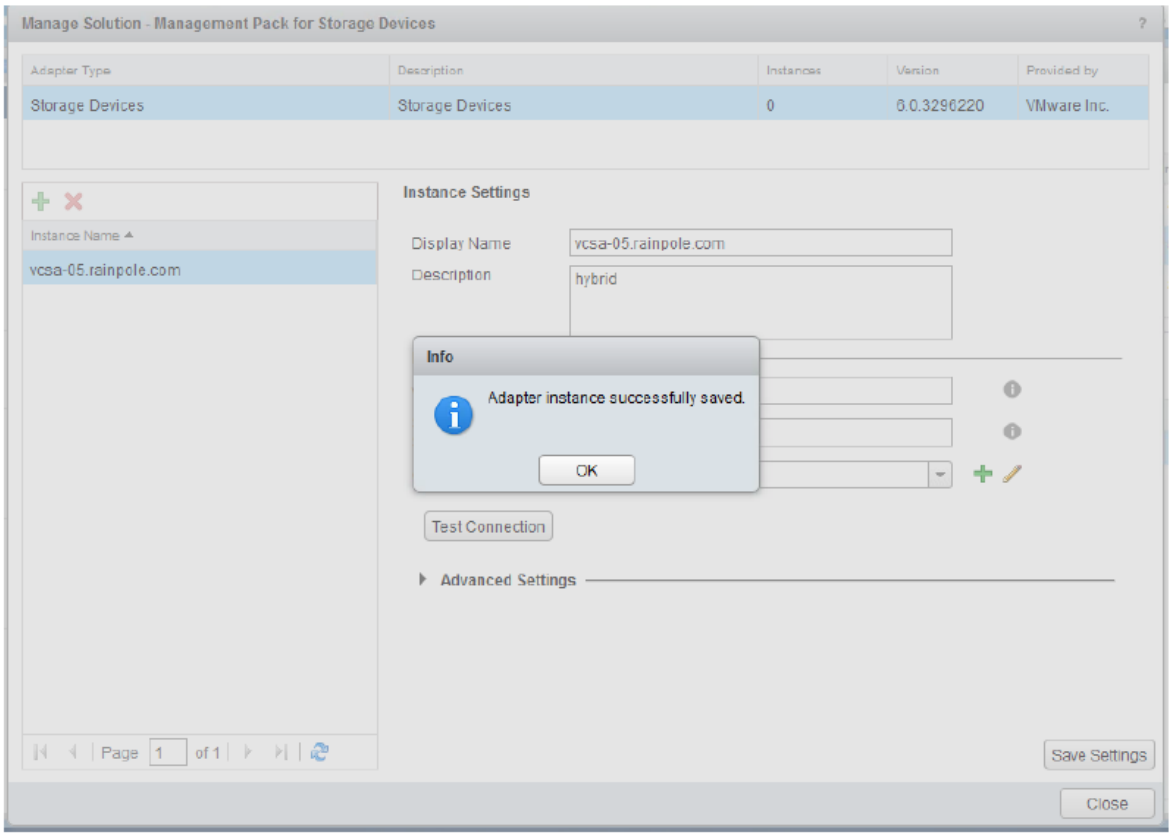
The only remain steps are to accept the EULA and complete the installation. When the install is complete, the new solution should be visible in the list of solutions:



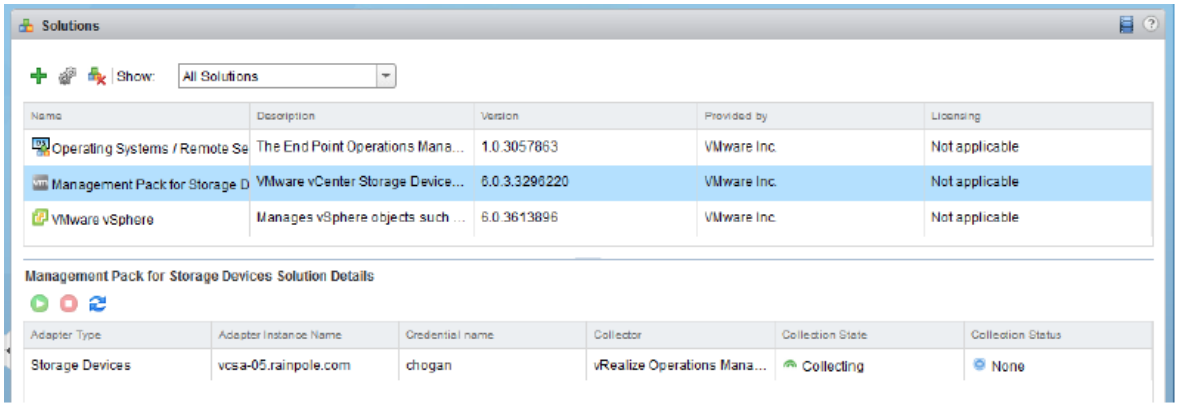
However, now that the Storage Devices Adapter Instance is still "Not configured". This is the next step.

16.5 Configure the MPSD Adapter Instance

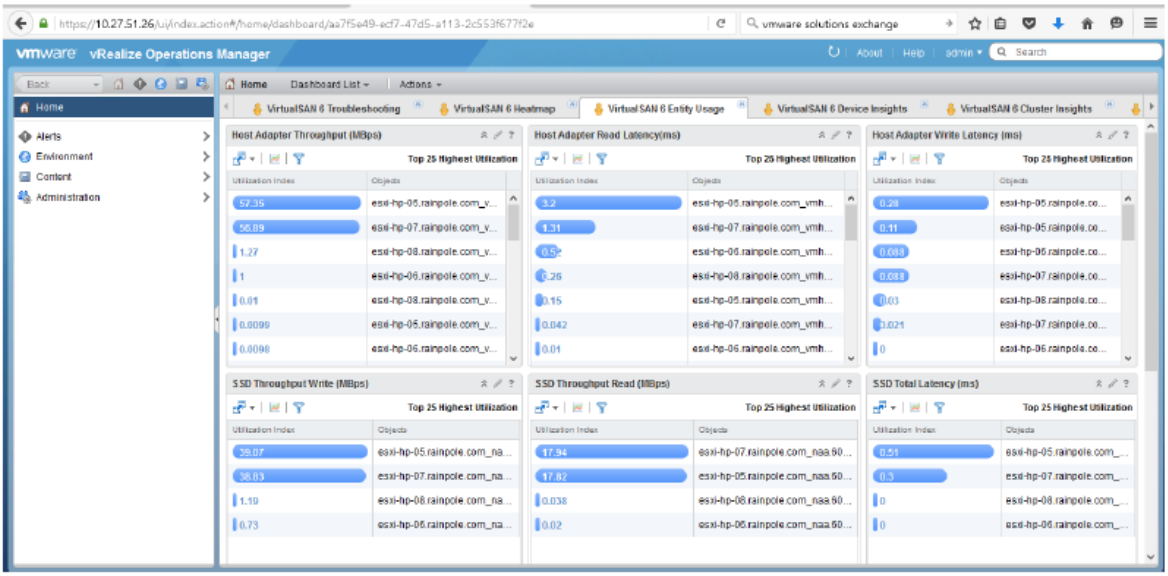
The configuration is very much the same as that carried out for the VMware vSphere solution done earlier. Select the solution, and then click on the configure option at the top of the screen (represented by the wheel/cogs icon). Fill in the vCenter server details and credentials as before, and once more test to make sure that they are functioning as expected:



When the settings are once again saved, the adapter instance should now be populated, and the collection state should changed to "Collecting" as shown below:



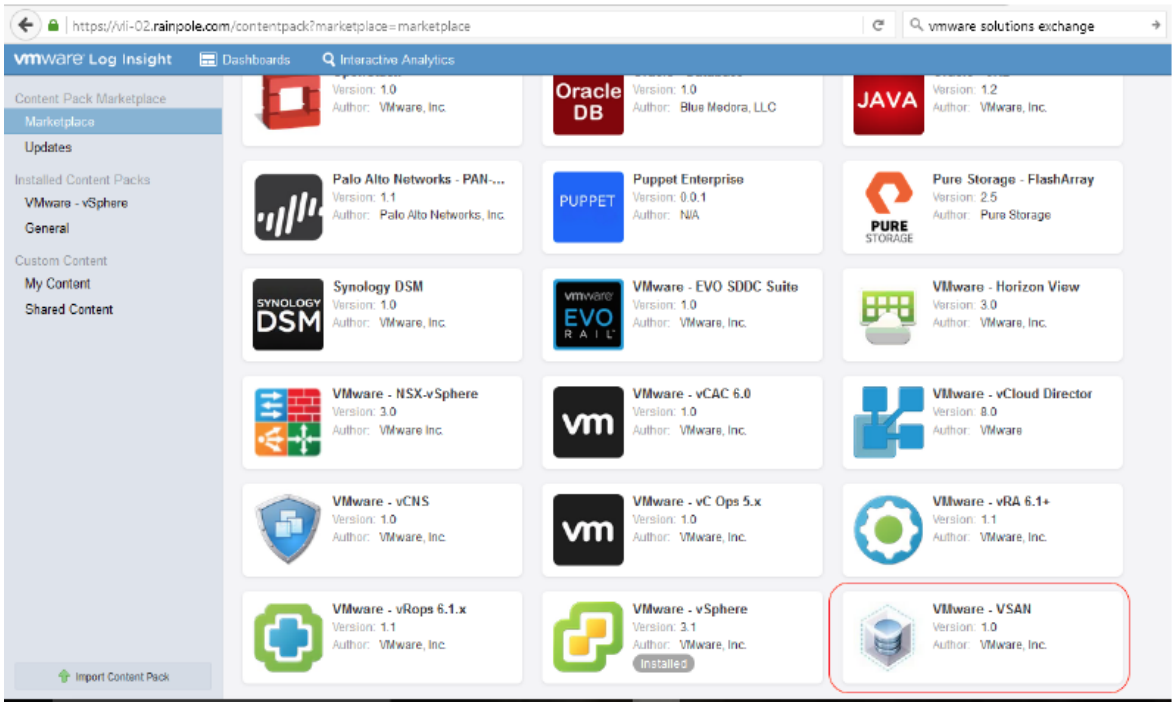
A new set of default vSAN dashboards are now presented to the admin. After a few minutes, the dashboards should begin to populate with vSAN specific information. Here are the dashboards along with some metrics from the Entity Usage dashboard.



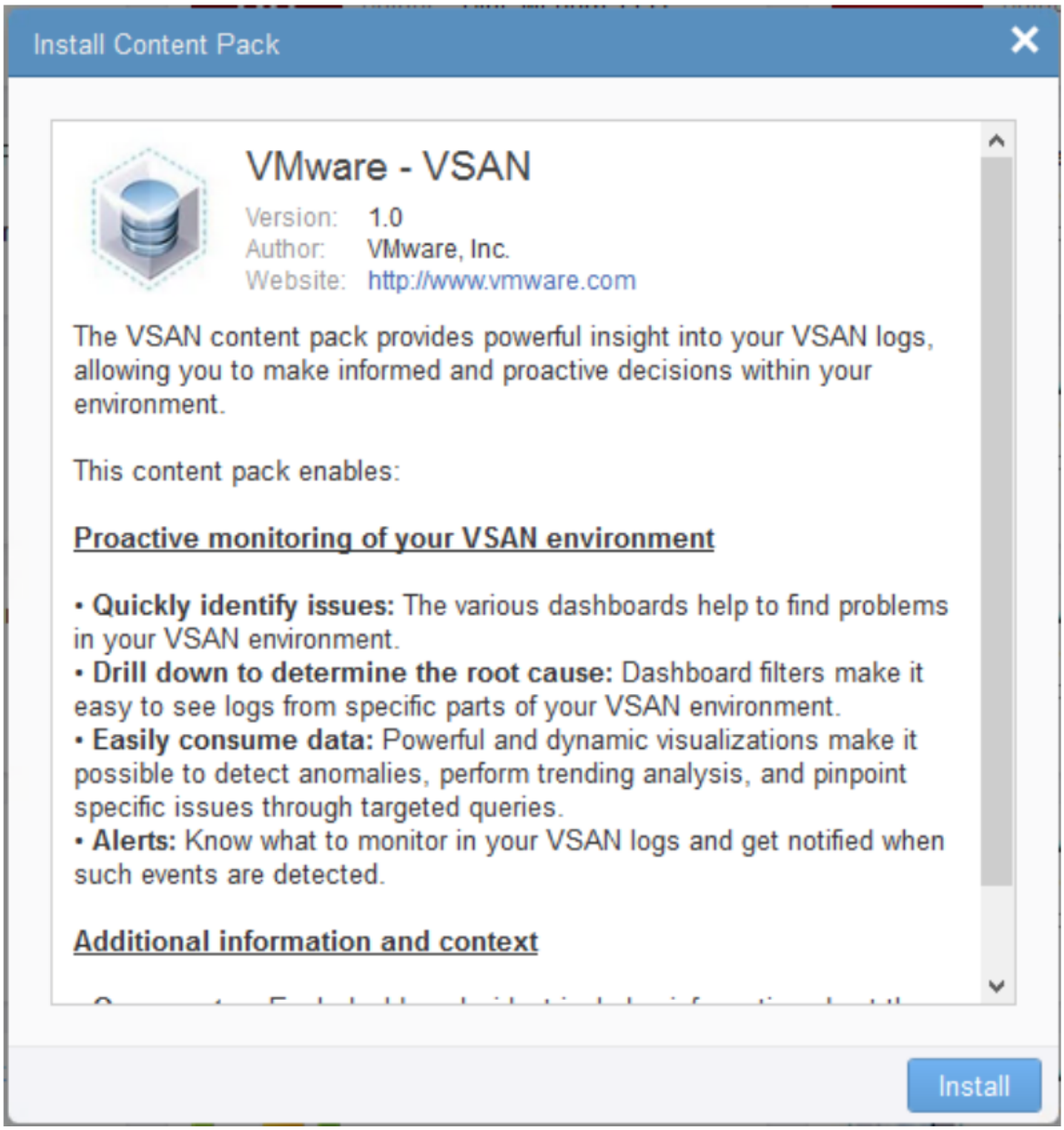
16.6 Integrating vRealize Log Insight with vSAN

Deploying vRealize Log Insight (vRLI) is not covered in this Operations Guide. Refer to the vRealize Log Insight documentation for this procedure.

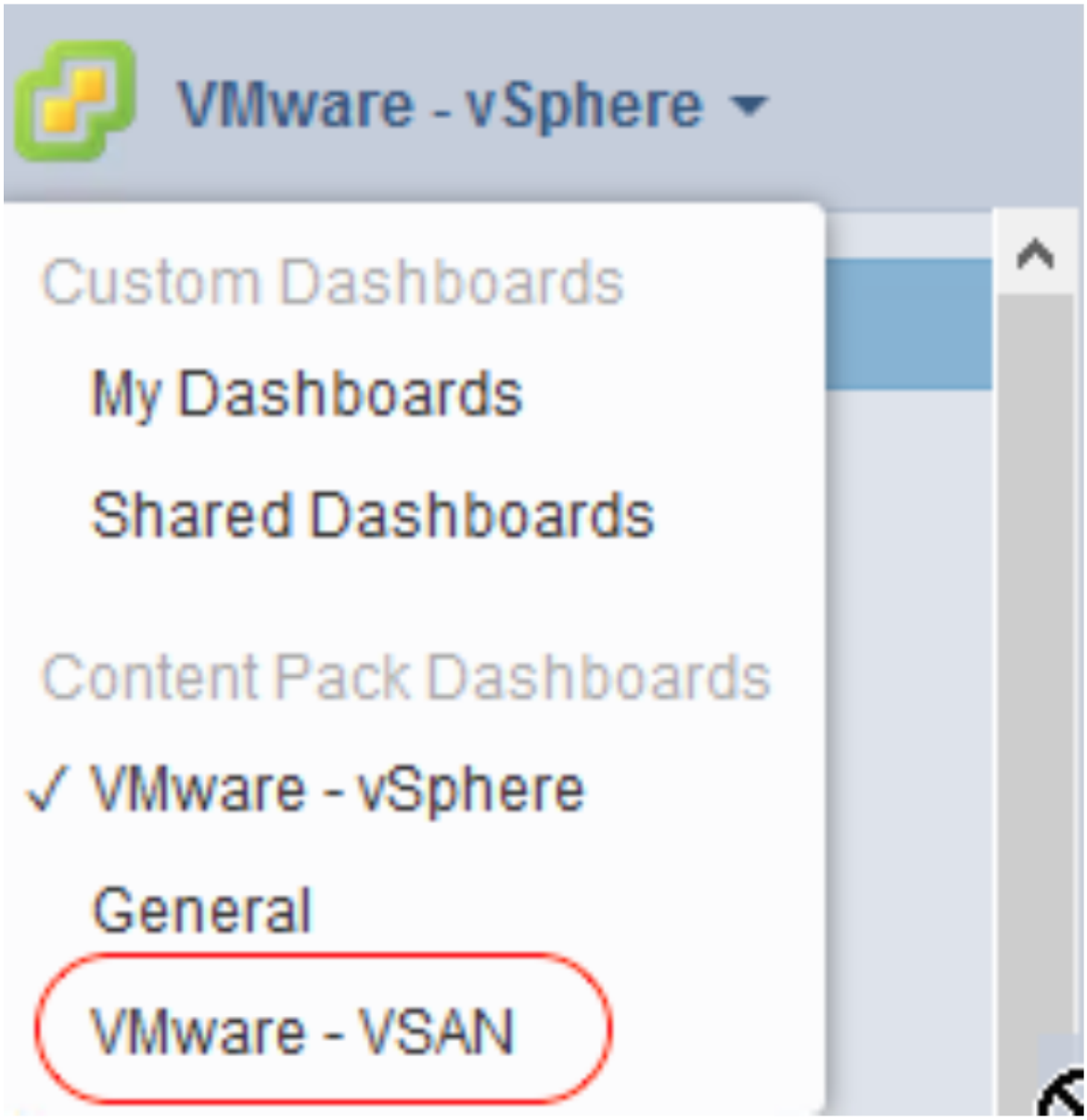
In this section, we will show you how to install a special content pack for vSAN into vRLI. This content pack is available in the solutions exchange and can even be accessed from within vRLI, as shown below:



Simply click on the content pack, review the pop-up screen and click install. When the install completes (in a matter of seconds), then details about the content pack and what it offers you for monitoring vSAN is also displayed.



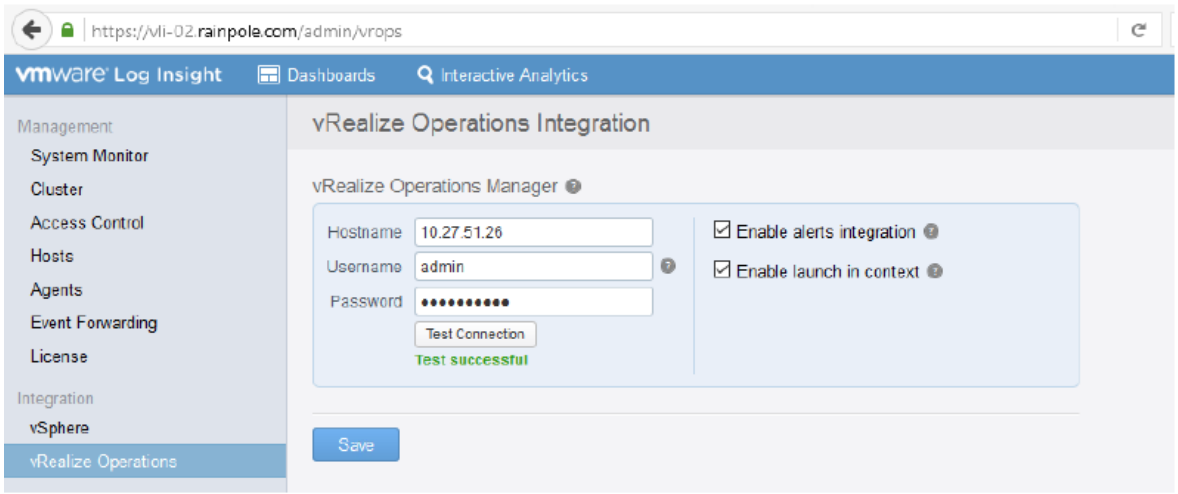
Now there are an additional set of dashboards related to vSAN. Simply select the new vSAN dashboards from the list of available dashboards in the top left-hand corner of the vRLI window, as shown below.



This provides you with a complete set of dashboards for monitoring vSAN events through logs and vRealize Log Insight.

16.7 Integration vRLI with vROps for vSAN

You can configure Log Insight to send alert notifications to vCenter Operations Manager. Integration is very simple. In vRLI, select the Administration section. There you will find a section called integration, which should already show vRLI integrated with vSphere. To integrate vRLI with vROPS, simply provide the appropriate vROps credentials, test the connection, then select Save.



If everything configures successfully, you should observe the following: