Hewlett Packard Enterprise

Implementing vSAN on HPE ProLiant servers with VMware vSphere

Contents

Technical white paper

Introduction

VMware® vSAN™ provides a software-defined storage solution that supports the storage needs of virtual machines in a vSphere cluster and provides support for key vSphere capabilities such as VMware vSphere® vMotion®, VMware vSphere® High Availability (vSphere HA), and VMware vSphere® Distributed Resource Scheduler™ (DRS). vSAN aggregates the local server storage of multiple HPE ProLiant servers within a vSphere cluster into a shared storage pool that is available to all hosts within that vSphere cluster. Hard disk drives (HDDs) and flash-based storage, such as solid-state drives (SSDs) and PCIe accelerators, from servers in a vSAN-enabled vSphere cluster contribute to the performance and capacity of the vSAN solution.

Note

For the remainder of this paper, flash-based storage will be generalized and referred to as SSDs.

This paper focuses on the requirements for a successful vSAN implementation on ProLiant servers with vSphere 5.5 and 6.0. When planning a vSAN implementation, take into account the configuration and support restrictions discussed in this paper. Additionally, this paper assumes that you are familiar with vSphere and ProLiant servers. For additional information on these products, see hpe.com/info/proliant and h22168.www2.hpe.com/us/en/partners/vmware/.

For additional documentation, see VMware®'s Technical Papers portal: vmware.com/vmtn/resources/.

vSAN overview

VMware vSAN, as its name implies, provides an alternative to traditional shared storage solutions, such as iSCSI, Fibre Channel, and Fibre Channel over Ethernet (FCoE), by utilizing the local storage available to each ProLiant server in the vSphere cluster. In vSAN 5.5, the local storage on each ProLiant server must be comprised of a combination of SSD and HDD. This is known as a hybrid configuration. The SSD drive(s) on each server provide read and write cache to accelerate IOPS and the HDDs provide persistent storage. With vSAN 6.0, there is the additional option of an all-flash configuration, which swaps out HDDs for SSDs. With vSAN, the servers now provide both the compute and shared storage capabilities required in a vSphere cluster.

Data access

When configuring vSAN, vSphere administrators define disk groups in the vSphere Web Client. Up to five disk groups can be configured on an ESXi host in a vSAN cluster with each disk group being comprised of one caching tier SSD and up to seven-capacity tier HDD/SSD. Each disk group is presented as a vSAN shared datastore, which are utilized to store virtual machine objects such as virtual disks, configuration files, and snapshots.

The data on a vSAN datastore is accessed in one of two methods. If the requested data is local to the host, then the I/O is serviced directly from the local disks. However, if the data is remote, then the data is read from or written to the remote host over the Ethernet network.

In a hybrid configuration, the SSDs in the caching tier form a caching layer with 70 percent dedicated for read cache and 30 percent for write buffer. In an all-flash configuration, the caching tier only serves as a write buffer while the capacity tier is used for read cache and storing data. The read cache keeps a list of commonly accessed disk blocks in order to reduce I/O read latency. If the requested block is not in cache, it is retrieved directly from the capacity layer. Important to note is that a host can access the read cache of any of the hosts in the same cluster. vSAN maintains a directory of cached blocks that can be referenced to find the requested block and then retrieved over the Virtual SAN network.

Write operations go into the write buffer first and later de-staged to the disks in the capacity layer. When a write is initiated, it is sent to the local cache on the host and additionally to the write cache on a remote host. In the event of a failure, this ensures a copy of the data is available. Once those writes land on both hosts, the write operation is acknowledged back to the application. The data then is de-staged to the capacity layer at regular intervals.

High availability

vSAN high availability is addressed by replicating the data to one or more additional hosts within the vSAN-enabled cluster. In addition to acting as read/write cache, the SSDs in the caching layer also support the vSAN replication activities by acting as the initial write copy target for the replicated data.

vSAN provides the ability for an administrator to define the number of complete mirror copies of data, referred to as replicas, within the vSAN cluster. The administrator defines the desired number of replicas on a per-VM basis, which provides the flexibility to adjust the desired level of protection based on the criticality of the VM data.

Scale-out storage

As new ProLiant servers are added into a vSAN cluster, the potential capacity of existing shared datastores is increased. Up to 32 nodes can be added into the cluster in vSAN 5.5, which matches the number of hosts supported in a vSphere cluster. vSAN 6.0 supports up to 64 nodes.

In addition to extending existing shared datastores, scale out is possible by creating new datastores through the addition of new hosts into the cluster, by adding additional disks into existing members of the cluster, or by performing a combination of those two options. Depending on the vSphere administrator preference, the configuration and addition of the new storage can be performed automatically by VMware® vCenter Server™ or manually be the administrator.

Limitations

As discussed previously, vSAN supports shared datastores and key vSphere capabilities such as VMware vSphere vMotion, vSphere High Availability (vSphere HA), and vSphere Distributed Resource Scheduler. However, there are some limitations that should be taken into account when determining the appropriate storage solution for your environment. The following list describes key limitations to take into account:

- vSAN does not support Distributed Power Management (DPM).

- Based on the cluster size, disk group, and disk count restrictions, there is a physical limit to the capacity support of a vSAN configuration.

- vSAN and VMware NSX® can co-exist on the same vSphere infrastructure but NSX does not support vSAN network traffic.

- vSAN requires a minimum of three vSphere hosts in each cluster.

- vSAN storage is only available to the hosts within the vSAN cluster.

- Up to 100 VMs per host in vSAN 5.5. Up to 200 VMs per host in vSAN 6.0 (See VMware's vSphere® Configuration Maximums white paper).

- Blade systems with direct attached storage systems are supported in vSAN 6.0. This support does not exist in vSAN 5.5. See the VCG for certified systems.

vSAN enablement process

Figure 1 provides an overview of the process for enabling a successful vSAN cluster. These steps provide a high-level overview and of planning and implementation process. It is highly recommended to review the VMware technical white paper, "What's New in VMware vSAN."

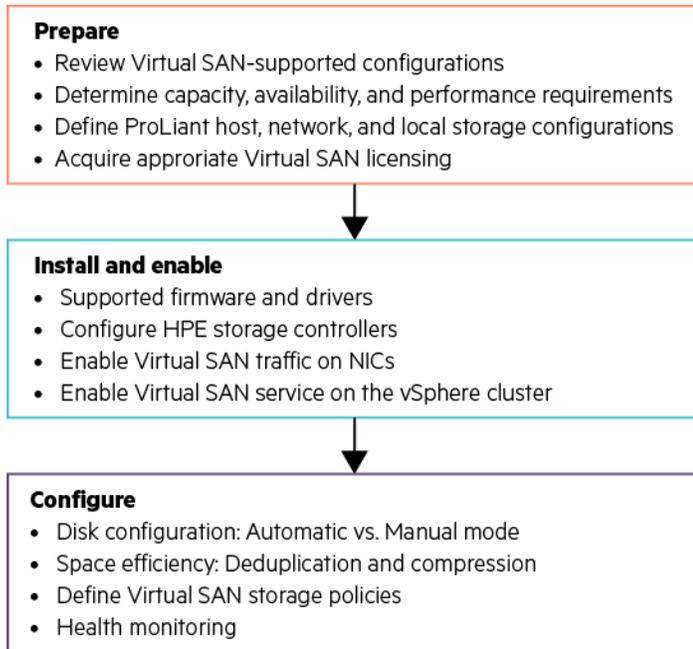**Prepare**
- Review Virtual SAN-supported configurations
- Determine capacity, availability, and performance requirements
- Define ProLiant host, network, and local storage configurations
- Acquire approriate Virtual SAN licensing

Install and enable
- Supported firmware and drivers
- Configure HPE storage controllers
- Enable Virtual SAN traffic on NICs
- Enable Virtual SAN service on the vSphere cluster

Configure
- Disk configuration: Automatic vs. Manual mode
- Space efficiency: Deduplication and compression
- Define Virtual SAN storage policies
- Health monitoring

Figure 1. vSAN enablement process overview

Prepare

Determining the hardware components and physical configuration for a vSAN implementation is not a trivial task. A large number of factors must be taken into account. This section provides an overview of the planning steps required to implement a supported vSAN configuration, highlights unique factors to take into account for HPE solutions, and provides references to additional resources.

The following steps define a rough outline and recommended steps to implement a successful vSAN configuration. It will become clear as the administrator performs the steps that the factors and requirements discussed in each step have multiple interdependencies that influence each other. In most cases, it will be an iterative process to finalize the configuration.

1. **Review possible options for building a supported configuration**: To implement a fully supported vSAN configuration, a vSAN Ready Node configuration can be implemented or a custom configuration (Build Your Own based on Certified Components) can be defined if the following requirements are met:

   – ProLiant server is listed on the VMware Compatibility Guide for Systems/Servers with support for the targeted vSphere release.

   – ProLiant server supports the targeted storage controller and drives. This can be determined by working with an HPE sales representative or by reviewing each product's QuickSpecs available on hpe.com.

   – Storage controllers, HDDs, and SSDs are listed on the VMware Compatibility Guide for vSAN. The VMware Compatibility Guide (VCG) has specific hardware listed along with their supported firmware and drivers. VMware is emphasizing the importance of following the VCG to avoid issues stemming from unsupported configurations.

   vSAN-certified solutions

   HPE has certified several vSAN Ready Node configurations. HPE Ready Nodes are certified combinations of ProLiant servers, RAID/Storage controllers, NICs, and drives specific for vSAN usage. These configurations are categorized into Ready Node Profiles based

on VMware's vSAN Hardware Quick Reference Guide. This guide provides a reference for selecting the right Ready Node server for your environment.

To build your own server configuration, HPE server components have been certified and listed on the VMware Compatibility Guide. See the "Build Your Own based on Certified Components" section of the VCG.

Refer to the VMware Compatibility Guide for the latest list of vSAN certified offerings.

2. **Size and define the characteristics of the vSphere cluster**: Initially, the traditional planning exercises required to define the compute, memory, and networking requirements of a vSphere cluster should be completed. At the end of this step, a rough number of ESXi hosts, their CPU and memory configuration, and the physical and virtual networking should be defined. In addition, the virtual machines/application storage capacity and storage performance requirements along with their I/O profile should be well understood. The configuration defined in this step should be viewed as a starting point, which will likely be altered when the vSAN requirements and needs are taken into account. See VMware's TCO and Sizing Calculator vsantco.vmware.com/vsan/si/siev.

   – A minimum of three hosts are required for vSAN, although in a ROBO configuration you can have two hosts with an external witness. This minimum configuration only allows for a single host failure. HPE recommends having four or more hosts for additional availability options (failures, maintenance, etc.).

   – The minimum memory requirement for a host is 6 GB.

   – A dedicated 1 Gb or 10 Gb network for vSAN traffic is required with a recommendation for 10 Gb networks in most environments. For all-flash configurations, the 10 Gb network is required due to the increase in network traffic.

   – Load balancing is not supported for vSAN network traffic. However, it is recommended to include more than one adapter and to use NIC teaming to provide redundancy. This means configuring physical NICs in an active/standby configuration.

   – vSAN requires multicast traffic be enabled on the Layer 2 network between ESXi hosts in a cluster. Layer 3 multicast is not required.

   – USB and SD devices are supported as ESXi boot devices in vSAN.

3. **Size and define the characteristics of the vSAN configuration**: Capacity, performance, and availability requirements for storage, VMs, and networking are the primary factors to take into account in this step. The VMware white paper, "VMware vSAN Design and Sizing Guide" provides a detailed analysis of each of the factors and provides formulas to assist in the planning process. See vmware.com/files/pdf/products/vsan/vsan_design_and_sizing_guide.pdf.

Information to take into account during the planning exercises in this step:

   – If ESXi will be installed on a local disk, that disk cannot be utilized for a vSAN disk group. Any RAID level supported by the storage controller can be used on the install disk(s).

   – It is not required for all ESXi hosts in the cluster to contribute storage to the vSAN cluster; however, a minimum of three hosts must contribute storage.

   – Achieving optimal performance depends on the I/O load. HPE recommends to ensure that the storage controllers have a queue depth of at least 256.

   – vSAN supports storage controllers in two modes: either pass-through mode or RAID 0 mode. Pass-through mode is also commonly referred to as HBA mode.

   – vSAN supports multiple HPE controllers on a single host. This allows for disk configurations beyond the limit of a single controller.

   – HPE SAS expanders are supported with VSAN with maximum of 16 drives.

   – HPE recommends to have the cache size of the vSAN datastore be at least 10 percent of the anticipated consumed capacity. This ratio should be calculated before the number of failures to tolerate setting is considered. This applies to both hybrid and all-flash configurations.

   – Up to five disk groups can be configured on an ESXi host in a vSAN cluster with each disk group being comprised of one caching tier SSD and up to seven-capacity tier HDD/SSD.

   – With the limitation of one caching tier SSD per disk group, HPE recommends to have multiple disk groups to improve performance and to protect against failure.

   – SSDs are categorized by their endurance and performance classes. HPE offers a variety of VSAN certified SSDs (SAS, SATA, NVMe, and PCIe). See VMware's VCG here.

At the end of this step, the questions below should be answered and the initial design adjusted accordingly:

- How many shared datastores are required and what are the minimum raw capacity requirements for each of those?

- What is the desired caching-to-capacity disk ratio?

- How many VMs can be supported by the cluster and per host?

- What is the expected host CPU and memory overhead for vSAN processing?

- What are the networking requirements for vSAN?

- What is the minimum number of ESXi hosts required?

4. **Define the hardware configuration**: In this step, the ESXi host compute, memory, and storage requirements should be mapped to hardware. As mentioned in step 1, the VMware Compatibility Guide should be referenced to determine the supported hardware with vSAN. Keep in mind that depending on the HPE ProLiant server and drive tray configuration, the number of physical disks that can be supported by each system will vary. See VMware's Hardware Guidance white paper vmware.com/files/pdf/products/vsan/vmware-tmd-virtual-san-hardware-guidance.pdf and VMware's Hardware Quick Reference Guide partnerweb.vmware.com/programs/vsan/virtual%20san%20hardware%20quick%20start%20guide.pdf.

   HPE supports vSAN configurations with both HPE Smart Array controllers and Smart Host Bus Adapters (HBAs). In either case, when a PCIe standup controller is utilized, it should be placed in a PCIe Gen3 slot for optimal performance.

5. **Procure appropriate vSAN licensing**: vSAN licenses have a per CPU capacity. When a license is applied to a cluster, the total number of CPUs in the participating hosts count towards that license capacity. See VMware's licensing guide for additional information: vmware.com/files/pdf/products/vsan/vmware-vsan-62-licensing-guide.pdf.

Install and enable

Firmware and drivers: To ensure you are running a supported vSAN configuration, update the drivers and firmware on the storage controllers, SSDs and/or HDDs to match the supported versions listed on VMware's VCG.

See HPE's white paper on deploying updates on HPE ProLiant servers here: h20195.www2.hpe.com/v2/GetDocument.aspx?docname=4aa4-7994enw.

Storage controllers: Before enabling vSAN, the storage controllers must be configured in either pass-through or RAID 0 mode. VMware recommends running in pass-through mode when possible so that the hypervisor is given direct access to the physical disks. VSAN can then setup a distributed configuration across all hosts in a cluster. In RAID 0 mode, single disk RAID 0 sets will need to be configured for each disk using HPE Smart Storage Administrator (see user guide here). In this case, VSAN will not manage hot-plug operations (instead relying on the storage controller's software to manage that), and local SSDs will need to be tagged as SSD before they can be claimed in vSAN. HPE SSA can be used to switch between modes and configure the RAID 0 sets. HPE Smart Storage Administrator CLI is also supported and can be used:

- To switch between pass-through and RAID 0 modes on HPE storage controllers:

  # esxcli hpssacli cmd -q "controller slot=<slotnumber> modify hbamode=<on/off> forced"

- To create a single disk RAID 0 set:

  # esxcli hpssacli cmd -q "controller slot=<slotnumber> create type=ld drives=<driveaddress> raid=0"

For more information about HPSSACLI, see the HPE VMware Utilities User Guide: h20564.www2.hpe.com/hpsc/doc/public/display?docId=c05063271.

Additionally in RAID 0 mode, it is recommended to configure the storage controller cache for 100 percent read cache. This effectively disables write caching to minimize conflicts with vSAN's caching layer. In RAID 0 mode, there are two supported vSAN Smart Array configurations that can be configured using HPSSACLI commands:

1. Lower q-depth writes—generally Smart Array default:

   a. # esxcli hpssacli cmd "controller slot=<slotnumber> modify drivewritecache=disable"

   b. # esxcli hpssacli cmd "controller slot=<slotnumber> ld all modify caching=enable"

2. High q-depth writes:

a. # esxcli hpssacli cmd "controller slot=<slotnumber> modify drivewritecache=enable"

b. # esxcli hpssacli cmd "controller slot=<slotnumber>ld all modify caching=disable"

Finally, VMware recommends disabling the HPE SSD Smart Path feature on HPE Smart Array controllers. See <u>VMware KB 2092190</u>.

Networking: The VSAN service must be enabled on a VMkernel port for each host. HPE recommends dedicating a VMkernel to vSAN traffic and separating it from the management traffic. Edit the VMkernel port and check the box for vSAN Traffic.
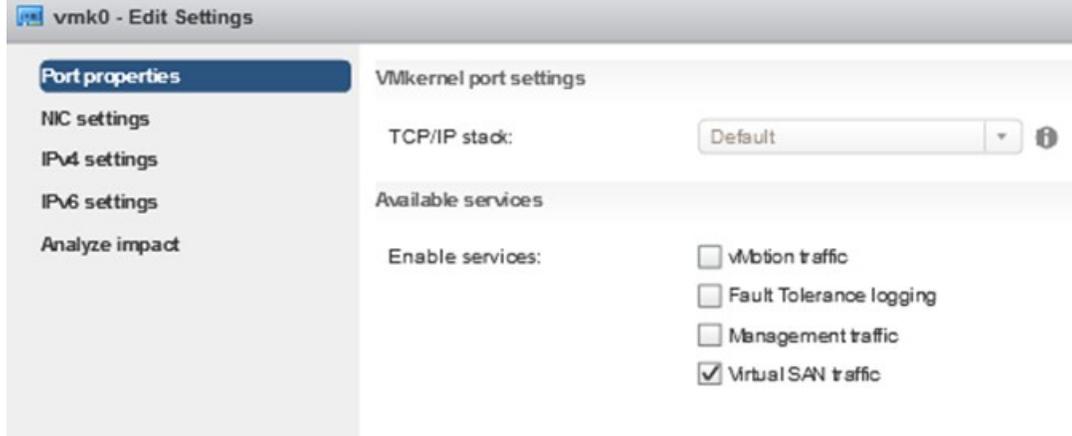
Figure 2. VMkernel port settings

If vMotion, Storage vMotion, or HA is desired on the cluster, an additional VMkernel port is needed. vSAN traffic requires a 1 Gb or 10 Gb network, and standard switches are supported but vSphere Distributed Switches are recommended.

Enable VSAN on cluster: Now create the cluster and add the desired hosts. To enable vSAN, select the cluster, go to "Manage," "Settings," and "General." Click the "Configure" button and complete the wizard to enable VSAN.

Configure
Disk configuration: Within the same wizard, two modes will be available to configure the disk-claiming mode:
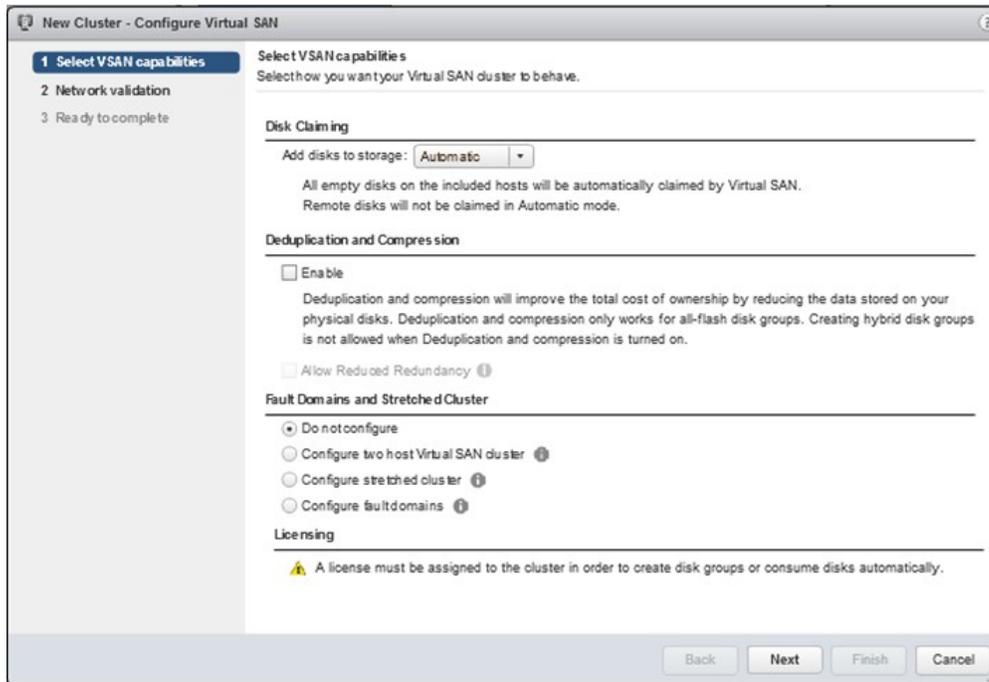
Figure 3. New cluster disk configuration

- **Automatic mode**: It will automatically discover and claim local empty SSDs and/or HDDs on each host in the cluster and build disks groups on each host. Once this is completed, an aggregate datastore is created. The size of the datastore will be the aggregate of the disks in the capacity layer in the VSAN. If additional storage is added to a host or a host is added to the cluster, the new disks will be claimed by VSAN, disk groups will be created, and the VSAN datastore will grow accordingly. VSAN will not automatically claim "remote" disks. These can be manually added to disk groups. If the controller is in RAID 0 mode, disks will not be automatically claimed.

Note
In an all-flash configuration, SSDs intended for the capacity layer will need to be tagged as such. There is currently no GUI option. CLI must be used.

- **Manual mode**: With this mode, disk groups will have to be manually created by selecting individual disks to use in each disk group. This can be done through the "Configure vSAN" wizard or through the cluster's disk management.
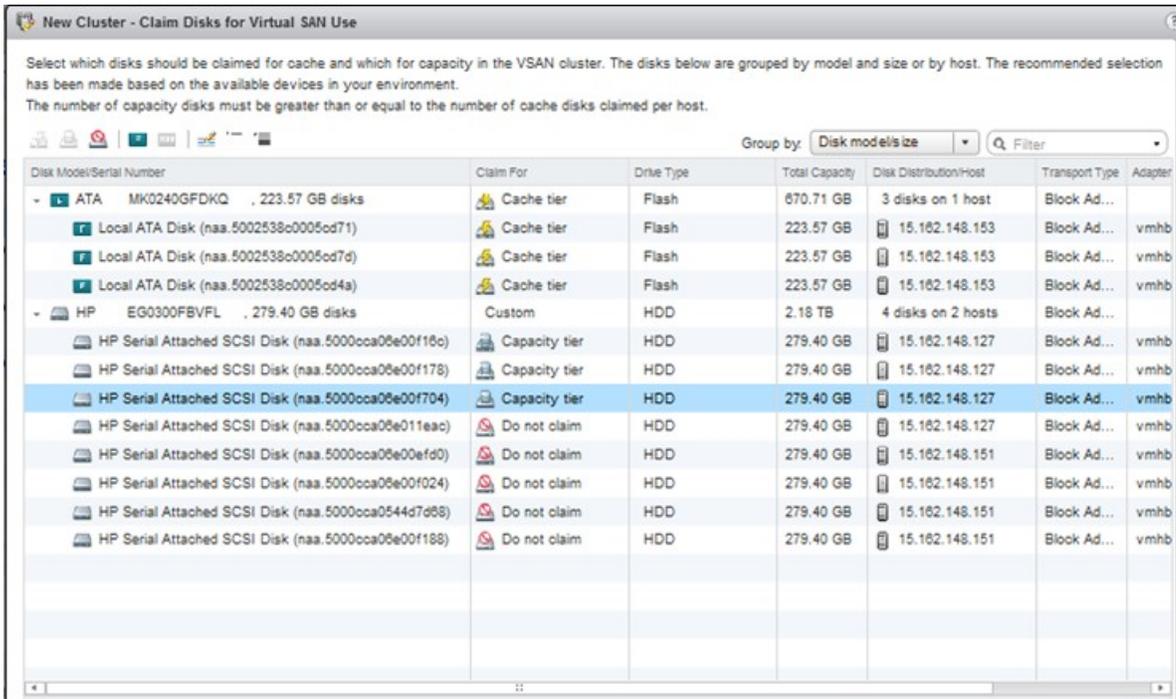
Figure 4. Claiming disks

If the storage controllers are in RAID 0 mode, there's an additional step required to specify the Drive Type. In the Claim Disks window, there are buttons in the toolbar to tag drives as SSD or HDD.

Note

Hosts that are later added to the cluster will not have their local storage added to any disk group. This applies to disks added to any existing hosts as well. Disk groups will have to be manually created or modified to add the new disks.

Once completed, previously unclaimed disks now show as being claimed by vSAN:
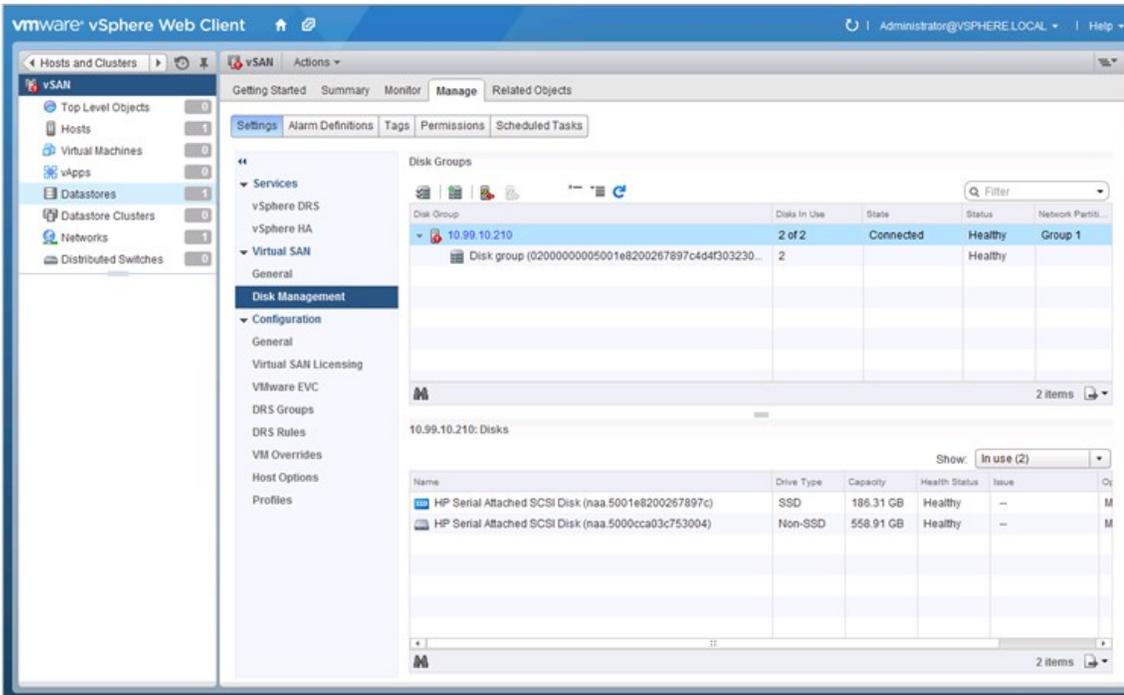
Figure 5. VSAN disk management

Space efficiency: Deduplication and compression are now available with VSAN 6.2. These two space efficiency techniques are only available for all-flash configurations. They are enabled together at the cluster level and cannot be enabled individually.
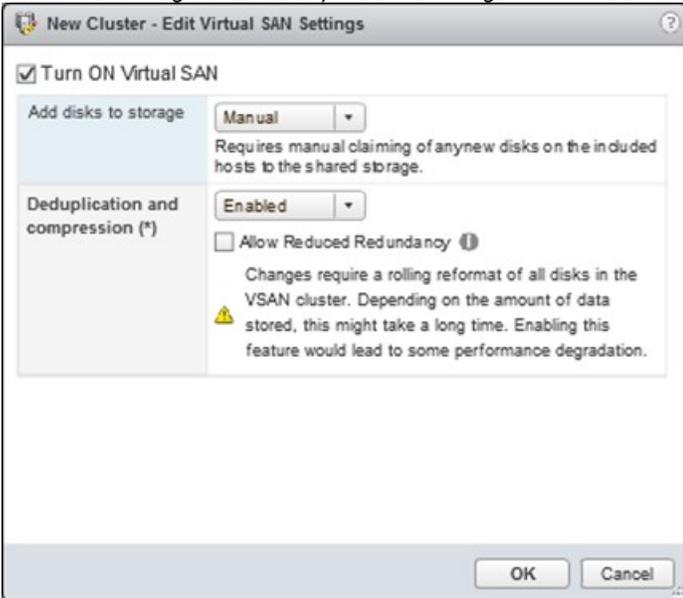
Figure 6. Space efficiency options

Deduplication occurs when data is de-staged from the caching tier to the capacity tier, and is performed within each disk group. Redundant blocks across multiple disk groups will not be deduplicated. Compression is then applied after the eduplication has occurred but before the data is written to the capacity tier. For more information, see VMware's VSAN 6.2 Space Efficiency white paper here.

Storage policies: VSAN relies on storage policies to define Virtual Machine storage requirements. Define the storage policies based on VM performance and availability requirements. When you enable vSAN on a cluster, a default vSAN policy is created and applied to the aggregate datastore. There are a number of settings that can be defined within a policy including the number of failures to tolerate, which allows a virtual machine to survive a number of hardware failures in a cluster.
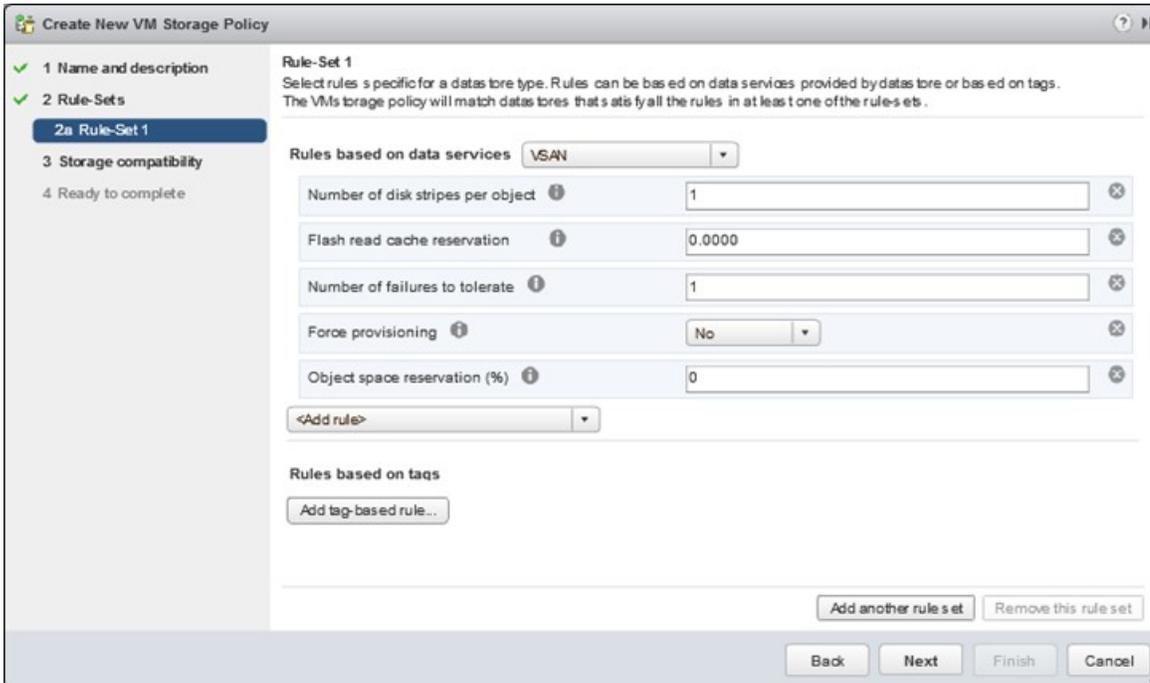

Figure 7. New VM storage policy

Once defined, policies can be applied when a VM is being created within the cluster. Alternatively, virtual machines can be assigned to the policy by editing their settings. If no policy is specified, the default policy will be applied with the following settings:

- Number of failures to tolerate = 1
- Number of disk stripes per object = 1
- Flash-read cache reservation = 0 percent
- Object space reservation = not used
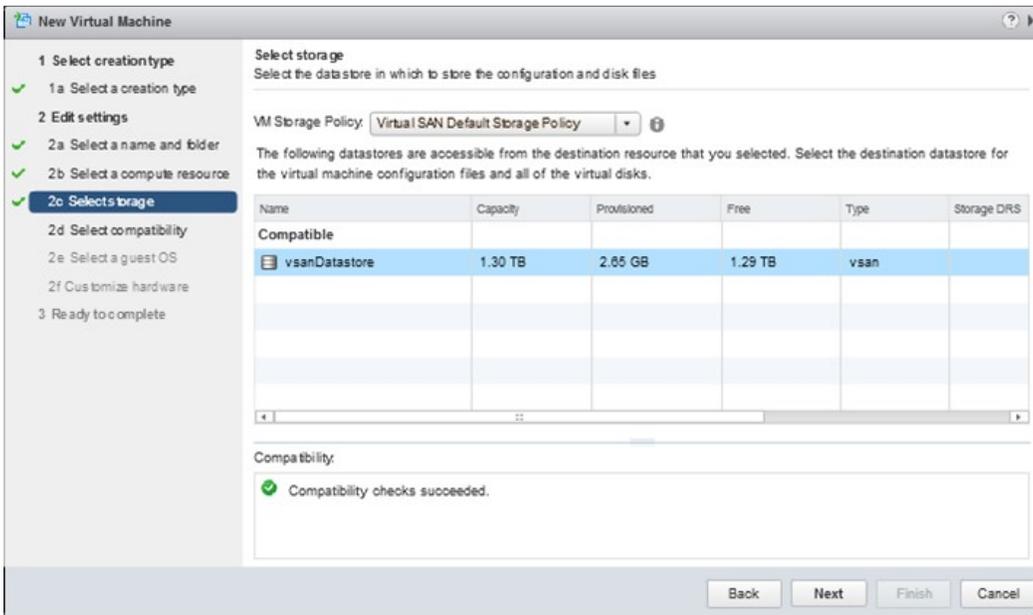- Force provisioning = disabled

Figure 8. Assigning storage policy to VM

Health monitoring: Lastly, VMware introduced the Health Check plugin with VSAN 6.0. This feature provides administrators with the ability to monitor the health of a VSAN cluster. Once enabled, it runs a number of checks on the networking configuration, physical disks and the vSAN cluster configuration.
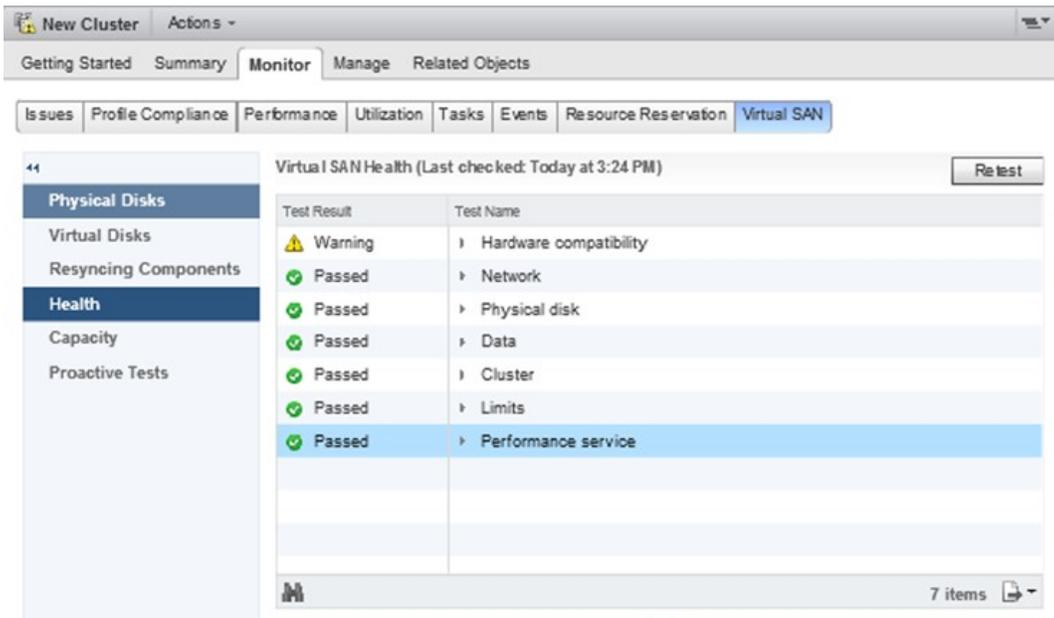

Figure 9. Health Check monitoring

Note
Installation requires a reboot on each host in the cluster.

One of the most useful aspects of this feature is the ability to ensure the hardware configuration is supported. The hardware can be checked against the VCG (either online or via uploading a copy of the VCG database). After the initial scan, the Health Check Plugin will continue monitoring the cluster on an ongoing basis and will highlight any issues found in the cluster. For additional information, see VMware's vSAN Health Check Plugin guide: vmware.com/files/pdf/products/vsan/vmw-gdl-vsan-health-check.pdf.

Additional recommendations

- For optimal performance, enable Jumbo Frames. The benefits will need to be weighed against the added complexity involved is deploying it throughout the switching infrastructure and NICs.

- Ensure SSD or HDD are empty and not pre-formatted before adding them to the cluster.

- For better performance, create multiple smaller disk groups instead of larger disk groups. The increase in cache drives could boost performance.

- While disparate hardware configurations are supported, it is recommended to use similar (if not identical) hardware configurations (number of disks, capacity, types of disk, etc.).

- Reference VMware vSAN Diagnostic and Troubleshooting Reference Manual for vSAN issues: vmware.com/files/pdf/products/vsan/vsan-troubleshooting-reference-manual.pdf.

Resources

HPE and VMware Alliance partner solutions
h22168.www2.hpe.com/us/en/partners/vmware/

HPE ProLiant servers
hpe.com/info/proliant

HPE Custom ESXi Images
hpe.com/info/esxidownload

HPE ProLiant server and option firmware and driver support recipe for VMware
vibsdepot.hpe.com

HPE Service Pack for ProLiant
hpe.com/info/spp

Deploying and updating VMware vSphere® on HPE ProLiant servers
h20195.www2.hpe.com/v2/getdocument.aspx?docname=4aa4-7994enw

VMware's technical papers portal
vmware.com/vmtn/resources/

VMware Compatibility Guide
vmware.com/resources/compatibility/search.php?devicecategory=vsan

Learn more at
hpe.com/info/proliant

Hewlett Packard
Enterprise