VMware Virtual SAN Layer 2 and Layer 3 Network Topologies

Deployments

TECHNICAL WHITE PAPER

Table of Contents

Introduction

VMware Virtual SAN is a distributed object storage platform that depends on IP Network connectivity to provide access to storage resources and storage management infrastructure services. Virtual SAN requires that all of the participating hosts can communicate over an IP network and are members of the same vSphere Cluster.

The locally attached storage devices from all of the hosts are pooled and presented as a single datastore to all members of the cluster once they have established IP connectivity and can communicate on the same Ethernet Layer 2 domain.

Virtual SAN clusters can also be formed with hosts that are connected to different Layer 3 network segments. The network Layer 3 segments must first be configured with IP Multicast in order to make all segments reachable by all the members of the cluster.

Although the Virtual SAN network traffic and Virtual Machine traffic can coexist on the same networks, this paper will not cover the configuration semantics and tuning of Virtual Machine network traffic.

The focus of this paper is based on the physical network and vSphere related technologies that are required to deploy Virtual SAN across Layer 2 and Layer 3 topologies.

This paper will help virtualization, network, and storage implementation engineers, administrators, and architects interested in deploying Virtual SAN on Layer 2 and across Layer 3 network topologies.

Network and vSphere Technologies

This section provides an overview and description of the different physical network and vSphere technologies that are required for deployments of Virtual SAN across Layer 2 and Layer 3 IP network topologies.

Networking Related Technologies

IP Multicast

IP Multicast is an IP Network communication mechanism used to efficiently send communications to many recipients. The communication can be in the form of one source to many recipients (one-to-many) or many sources to many recipients (many-to-many).

The recipients may be located in the same Layer 3 segment or distributed across multiple Layer 3 segments. In the case where the recipients are in the same Layer 3 segment, the recipients will also share the same Ethernet Layer 2 domain.

An IP Multicast address is called a Multicast Group (MG). IP Multicast relies on communication protocols used by hosts, clients, and network devices to participate in multicast-based communications.

Communication protocols such as Internet Group Management Protocol (IGMP) and Protocol Independent Multicast (PIM) are integral components and dependencies for the use IP multicast communications.

IP Multicast is a fundamental requirement of Virtual SAN. Virtual SAN depends on IP multicast communication for the process of joining and leaving cluster groups as well as other intra-cluster communication services. IP multicast must be enabled and configured in the IP Network segments that will carry the Virtual SAN traffic service.

Internet Group Management Protocol (IGMP)

IGMP is a communication protocol used to dynamically add receivers to IP Multicast group memberships. The IGMP operations are restricted within individual Layer 2 domains. IGMP allows receivers to send requests to the Multicast Groups they would like to join.

Becoming a member of Multicast Groups allows the routers to know to forward traffic that is destined for the

Multicast Groups on the Layer 3 segment where the receiver is connected. This allows the switch to keep a table of the individual receivers that need a copy of the Multicast Group traffic.

The participating hosts in a Virtual SAN cluster will negotiate for IGMP version 3. If the network does not support IGMP version 3, the hosts will fall back to IGMP version 2. VMware recommends that the same version of IGMP be used in all Layer 3 segments.

Protocol-Independent Multicast (PIM)

Protocol-Independent Multicast (PIM) is a family of Layer 3 multicast routing protocols that provide different communication techniques for IP Multicast traffic to reach receivers that are in different Layer 3 segments from the Multicast Groups sources. There are different versions of PIM, each of which is best suited for different IP Multicast topologies. The main four versions of PIM are these:

- PIM Dense Mode (PIM-DM) – Dense Mode works by building a unidirectional shortest-path tree from each Multicast Groups source to the Multicast Groups receivers, by flooding multicast traffic over the entire Layer 3 Network and then pruning back branches of the tree where no receivers are present. Dense Mode is straightforward to implement and it is best suited for small Multicast deployments of one-to-many.
- PIM Sparse Mode (PIM-SM) – Sparse Mode avoids the flooding issues of Dense Mode by assigning a root entity for the unidirectional Multicast Groups shortest-path tree called a rendezvous point (RP). The rendezvous point is selected in a per Multicast Group basis.

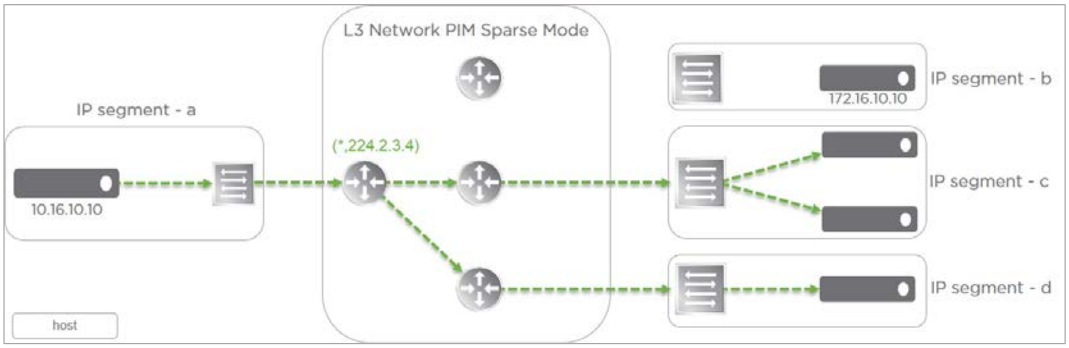


Figure 1: Layer 3 Network PIM Sparse Mode Communication Flow

Sparse Mode scales fairly well for larger Layer 3 Networks and is best suited for one-to-many Multicast topologies. If the network only supports IGMP version 2, VMware recommends the use of PIM-SM for Virtual SAN deployments over Layer 3.

- Bidirectional PIM (Bi-PIM) – Bidirectional PIM assumes that there are many MGs that have many sources and many receivers (many-to-many). Whereas Sparse Mode can manage many-to-many Multicast topologies, Bidirectional PIM does it by reducing the load on the Multicast routers as compared to Sparse Mode.

  Bidirectional PIM does not build a shortest-path tree, so MG data paths may have longer end-to-end delays than Sparse Mode, however Bidirectional PIM allows for a Multicast Group traffic to flow both ways over the same data path.
- PIM Source-Specific Multicast (PIM-SSM) – Source Specific Multicast is similar to Sparse Mode but it carries information about the IP of the source. Receivers join Multicast Groups based on the source of the Multicast Groups.

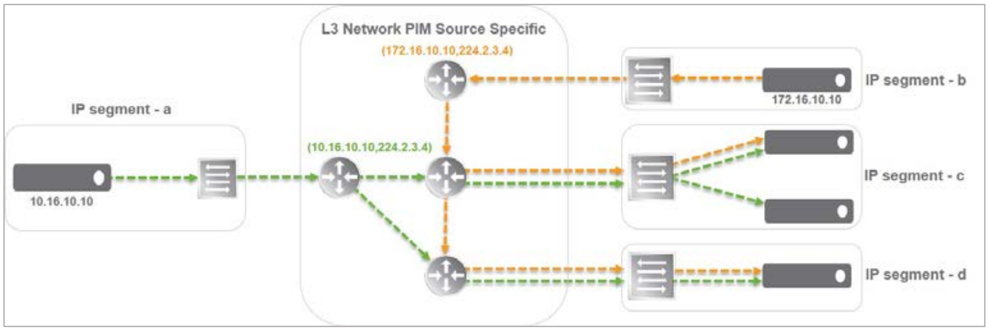


Figure 2: Layer 3 Network PIM Source Specific Mode Communication Flow

With Source Specific Multicast, shortest-path trees are built and are rooted in just one source, offering a more secure and scalable model for a limited amount of applications (mostly broadcasting of content).

If the networks are configured with IGMP version 3, then Source Specific Multicast requires the receivers to support IGMP version 3.

vSphere Related Technologies

vSphere Virtual Switch

VMware Virtual SAN supports the use of both the vSphere Standard Switch and vSphere Distributed Switch. However, VMware recommends the use of the vSphere Distributed Switch to take advantage of its centralized management capabilities as well as advanced network features.

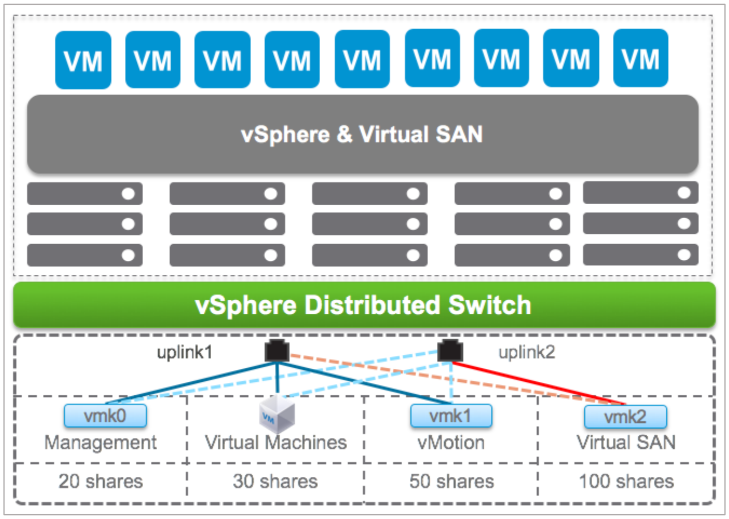


Figure 3: VMware Distributed Switched - QoS with Network I/O Control

The Virtual SAN network configuration can be implemented with vSphere standard or distributed switches. In either case, the networking configuration requirements and behavior remain relatively the same.

vSphere Distributed switches provide several advantages around management, advanced network features, and scalability capabilities that are all conducive the benefits and values of VMware Virtual SAN.

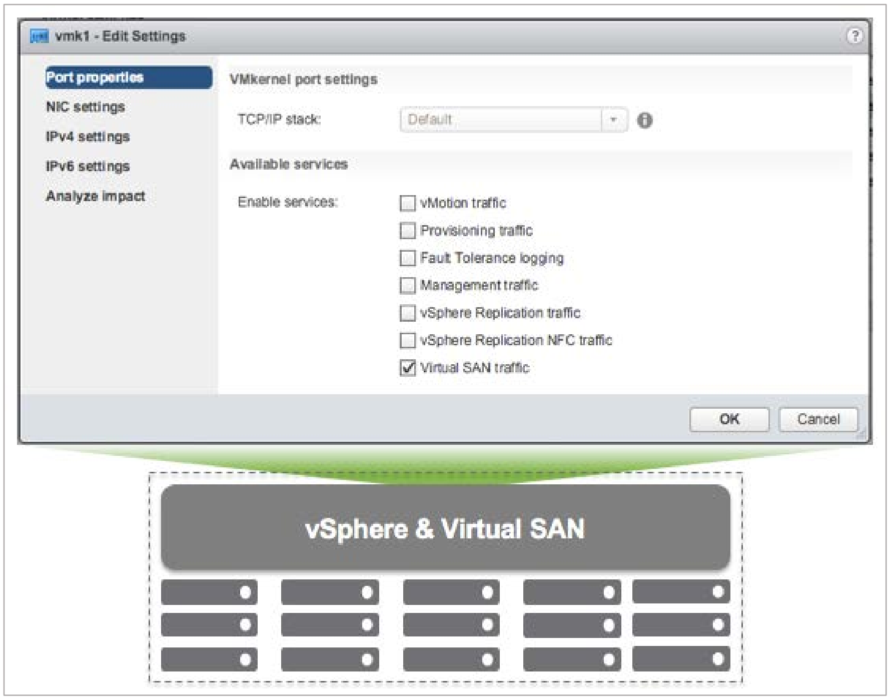
vSphere Distributed Switches facilitate large scale deployments with the support of up to 500 hosts per switch. They also provide access to advanced network features such as Network I/O Control and IP Multicast Filtering.

For scenarios where different network traffic services share physical network adapters, VMware recommends the use of Network I/O Control as mechanism for bandwidth allocation control for traffic management optimization (QoS).

Note: While the use of the vSphere Distributed Switch and the Network I/O Control feature are typically part of the vSphere Enterprise Plus licensing SKU, their use is also exclusively included as part of the VMware Virtual SAN license agreement.

VMkernel Network Interface

The VMkernel networking layer provides network connectivity to hosts and also handles the standard system traffic of multiple vSphere network services such as vSphere vMotion, IP storage, Fault Tolerance, Virtual SAN, and others.



VMkernel port settings

TCP/IP stack: Default

Available services

Enable services:
□ vMotion traffic
□ Provisioning traffic
□ Fault Tolerance logging
□ Management traffic
□ vSphere Replication traffic
□ vSphere Replication NFC traffic
☑ Virtual SAN traffic

Figure 4: Creating a VMkernel network Interface associated with Virtual SAN Traffic Service

Any host that is going to participate as a member of a Virtual SAN cluster must have the Virtual SAN traffic service associated with a VMkernel network interface.

The Virtual SAN traffic service will automatically assign the default multicast address settings to each host which will then make them eligible to send frames to a default Multicast Group, and Multicast Group Agent.

- Virtual SAN Default Multicast Group address 224.1.2.3
- Virtual SAN Default Multicast Group Agent address 224.2.3.4

The physical uplinks used by the Virtual SAN network interfaces should be connected to physical switches that are configured with IGMP and IGMP Snooping version 2 or version 3 on a common network segment that will carry the Virtual SAN network traffic.

When deploying on a Layer 2 network, one of the switches on that network segment (VLAN) should be configured as the IGMP Querier.

Alternatively, when the deployment is being performed across Layer 3 network segments, a Layer 3 capable device (router or switch) with a connection and access to the same Layer 3 network segments can be configured as the IGMP Querier.

At this point, the hosts will establish their method of communication by joining the Virtual SAN default Multicast Group addresses, 224.1.2.3 and default Multicast Group Agent addresses 224.2.3.4.

In order to avoid unnecessary IP multicast floods within the Layer 2 segments, VMware recommends configuring IGMP snooping with an IGMP Querier in order to control the number of physical ports on the switches that will receive IP multicast frames.

For optimal network communication and efficiency, Virtual SAN multicast frames should be exclusively forwarded to the ports that are associated with the uplinks of the VMkernel network interfaces that are configured to carry the Virtual SAN traffic.

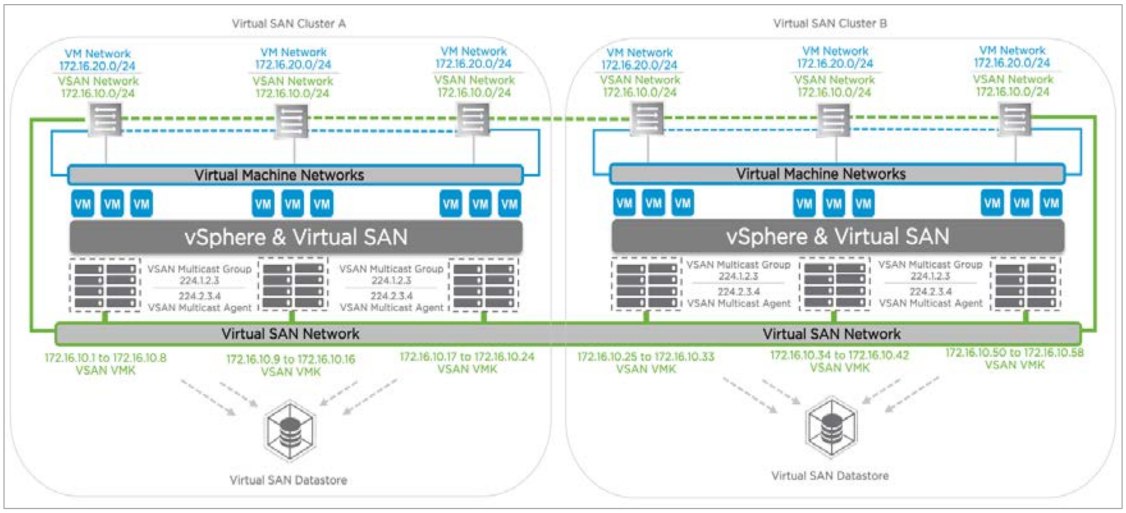


Figure 5: Multiple Virtual SAN Clusters

In scenarios with multiple Virtual SAN clusters, VMware recommends changing the default Multicast Group address and the default Multicast Group Agent address when the different clusters will share the same Layer 2 network segment.

This will prevent the clusters from receiving unnecessary multicast frames from one another.

In scenarios where members of a cluster have been deployed across different network segments (Layer 3), VMware recommends changing the default Multicast Group address and default Multicast Group Agent address.

VMware recommends the use of the Multicast Address range of 239.0.0.0/8 when changing the default addresses. Also, consult with members of the network team in order to identify the adequate Multicast Group addresses to use in order to comply with any potential Multicast Addressing policies that may exist.

For detailed instruction on how to change the default multicast address for Virtual SAN, please refer to the VMware Knowledge Base article 2075451.

http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2075451

TCP/IP Stacks

vSphere 6.0 introduced a new TCP/IP Stack architecture where multiple TPC/IP stacks can be utilized to manage different VMkernel network interfaces and their associated traffic.

As a result, the new architecture provides the ability to configure traffic services such vMotion, Management, Fault Tolerance, etc. on completely isolated TCP/IP stacks with the ability to use multiple default gateways.

For network traffic isolation and security requirements, VMware recommends deploying the different traffic services onto different network segments in an order to prevent the different traffic services from traversing through the same default gateway.

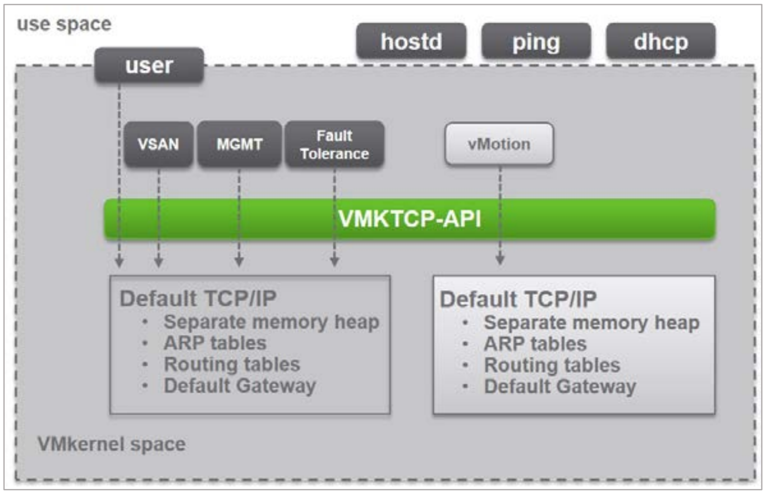


Figure 6: vSphere Multiple TCP/IP Stack Architecture

In order to configure the traffic services onto separate TCP/IP stacks, each traffic service type needs to be deployed onto their own network segments.

The network segments will be accessed through a physical network adapter with VLAN segmentation and individually mapped to dissimilar VMkernel network interfaces with the respective traffic services (Virtual SAN, vMotion, Management, etc.) enabled.

Built-in TCP/IP stacks available in vSphere:

- **Default TCP/IP Stack** – multi-purpose stack that can be used to manage any of the host related traffic services. Shares a single default gateway between all configured network services.
- **vMotion TCP/IP Stack –** utilized to isolate vMotion traffic onto its own stack. The use of this stack completely removes or disable vMotion traffic from the default TCP/IP stack.
- **Provisioning TCP/IP Stack –** utilized to isolate some virtual machine related operations such as cold migrations, cloning, snapshot, NFC related traffic.

It is assumed that environments with isolated network requirements for the vSphere traffic services will not be able to use the same default gateway to direct traffic.

The use of the different TCP/IP stacks facilitates the management for traffic isolation with the ability to use different default gateways.

Currently, vSphere 6.0 does not include a dedicated TCP/IP stack for the Virtual SAN traffic service nor the supportability for the creation of custom Virtual SAN TCP/IP stack.

To ensure Virtual SAN traffic in Layer 3 network topologies leaves over the Virtual SAN VMkernel network interface, add the Virtual SAN VMkernel network interface to the Default TCP/IP Stack and define static routes for all of the Virtual SAN cluster members.

Static Routes

The use of static routes is required by traffic services for which vSphere does not provide a non-Default TCP/IP stack.

In the VMware recommended deployment scenario where the Management and Virtual SAN traffic services are configured to use different Layer 3 network segments, they will share the Default TCP/IP Stack but be configured in different Layer 2 domains.

The default route for the Default TCP/IP Stack should remain with the Management VMkernel network interface. Static routes will be added for the Virtual SAN traffic to egress of the Virtual SAN VMkernel network interface.

It is only necessary to configure a single static route per host for each remote Virtual SAN Layer 3 segment or a single summary static route if the Virtual SAN Layer 3 segment addressing plan allows it.

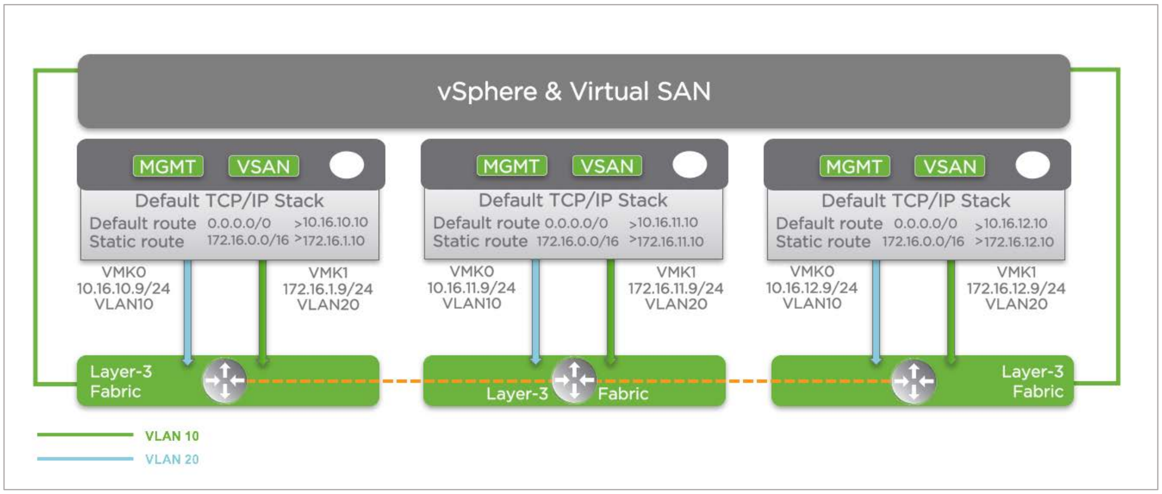


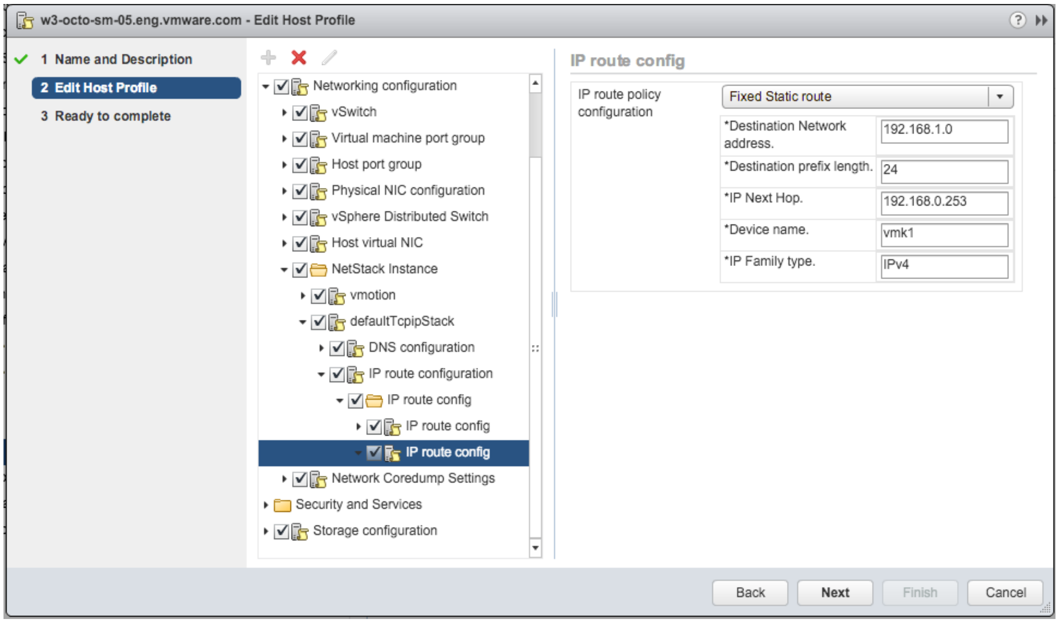
Figure 7: Static Route Logical Diagram

Hosts Profiles

Consider the use of Host Profiles as a management option to deal with the operating management functions of the communications paths that are established with the use of static routes.

Host Profiles provide an automated and centrally managed mechanism for host configuration and compliance. The use of Host Profiles reduces configuration risks, and can improve efficiency by reducing reliance on repetitive, manual tasks.

Host Profiles provide the ability to capture the configuration of a pre-configured host, and store the configuration as a managed object and use the catalog of parameters contained within to configure networking, storage, security and other host-level parameters.



IP route config

IP route policy configuration	Fixed Static route
*Destination Network address.	192.168.1.0
*Destination prefix length.	24
*IP Next Hop.	192.168.0.253
*Device name.	vmk1
*IP Family type.	IPv4

Figure 8: Host Profiles to Storing Static Routes

Static routes are stored within the Hosts Profiles as part of their catalog parameters. Host Profiles can be applied to either individual hosts or a cluster; applying a Host Profile to a cluster will affect all hosts in the cluster and result in a consistent configuration across all hosts in that cluster.

Host Profiles can also be used to validate the system configuration by checking compliance for any host or cluster against an associated standardized Host Profile.

Supported Network Topologies

This section covers the different supported network topologies and the impact they introduce to the overall deployment and management of Virtual SAN in different network scenarios.

Layer 2 Network Topologies

Layer 2 network topologies are defined as networking architectures that are composed of devices that operate at the Data Link layer (Layer 2) of the OSI model.

This network topology is responsible for forwarding packets through intermediate Layer 2 devices such as hosts, bridge, or switches.

It is required that all of the hosts participating in a Virtual SAN cluster are able to establish communication through the VMkernel interface connected to a common Layer 2 network segment.

The Layer 2 network topology offers the least complex implementation and management of the IP Multicast requirements for Virtual SAN while constraining the radius of the cluster.

All cluster members will send IGMP join requests over the VMkernel network interfaces that are used for the Virtual SAN traffic service.

By default, the hosts will negotiate their communication for IGMP version 3 and failback to IGMP version 2 whenever the physical network device does not support IGMP version 3.

For maximum Layer 2 traffic efficiency, VMware recommends the use and configuration of IGMP Snooping in all the switches configured in the Layer 2 network segment where Virtual SAN is present.

IGMP Snooping allows physical network devices to forward Multicast frames to only the interfaces where IGMP Join requests are being observed.

Layer 2 Physical Network Configuration

This section covers the physical network configuration procedures to enable IP Multicast for Virtual SAN. The configuration is focused on IGMP snooping and IGMP snooping Querier.

We will assume all members of the cluster are in the same Layer 2 network segment, represented by VLAN 10. In this scenario the role of IGMP Querier will be performed by a physical switches and not a router.

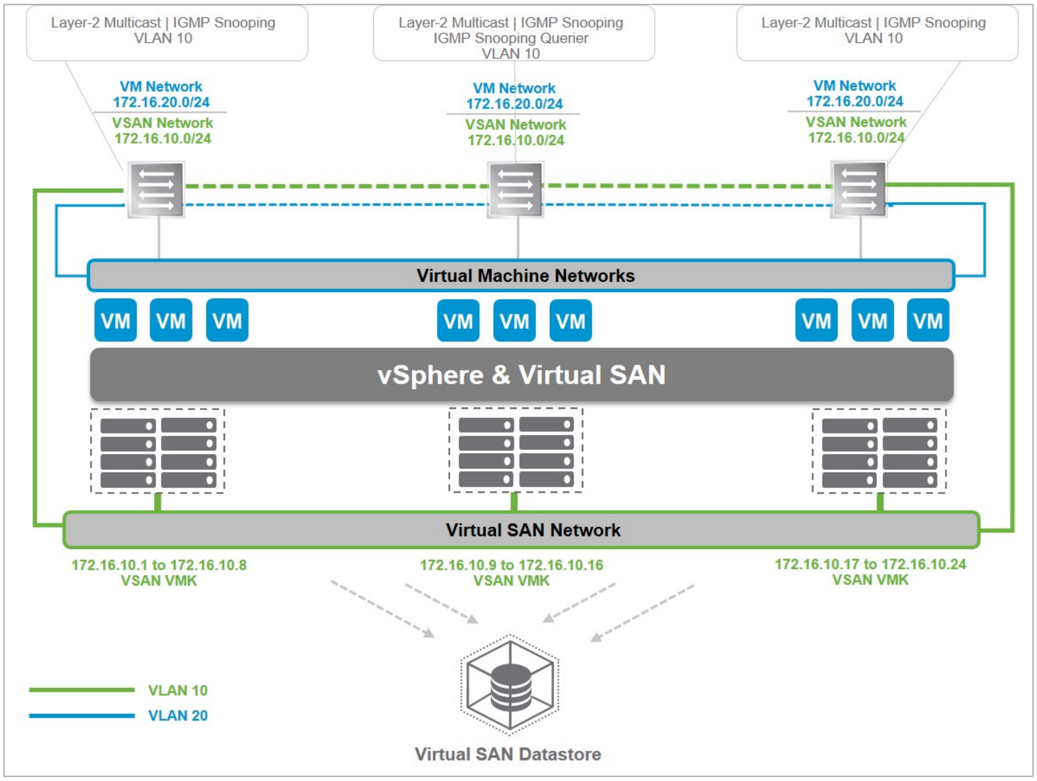


Figure 9: Virtual SAN Layer 2 Network Topology

For this scenario we will provide two different configuration examples that will be based on two different vendor platforms. The first example is based on the configuration of Cisco Nexus switch devices operating with the Cisco Nexus platform OS with IGMP version 3.

The second example is based on the configuration of Brocade VDX switch devices with IGMP version 2. Currently, Brocade VDX switch devices do not support IGMP version 3 and therefore the configuration will be based on IGMP version 2.

The configuration procedures for IP Multicast varies between different vendors and their respective network devices. Consult the network device vendor documentation for in-depth details and specific advanced procedures that go beyond the scope of this document.

Cisco Hardware Devices

The following a sample configuration of IGMP version 3 (enabled by default per VLAN) in Nexus 6000 running NX-OS 7.0(3):

Cisco Switch 1

```
configure terminal
ip igmp snooping
interface vlan 10
ip igmp snooping
```

Cisco Switch 2

```
configure terminal
ip igmp snooping
interface vlan 10
ip igmp snooping
ip igmp snooping querier 172.16.10.253
```

Cisco Switch 3

```
configure terminal
ip igmp snooping
interface vlan 10
ip igmp snooping
```

Brocade Hardware Devices

The following is a sample configuration of IGMP version 2 in VDX 6740s running NOS 7.0.0:

Brocade Switch 1

```
configure terminal
ip igmp snooping enable
interface vlan 10
ip igmp snooping enable
```

Brocade Switch 2

```
configure terminal
ip igmp snooping enable
interface vlan 10
ip igmp snooping enable
ip igmp snooping querier enable
```

Brocade Switch 3

```
configure terminal
ip igmp snooping enable
interface vlan 10
ip igmp snooping enable
```

Layer 3 Network Topologies

Layer 3 network topologies are defined as networking architectures that are composed of devices that are capable of operating at the network layer (Layer 3) of the OSI model.

This network topology is responsible for routing packets through intermediate Layer 3 capable devices such as routers and Layer 3 capable switches.

All Virtual SAN cluster members are required to join the cluster's Multicast Group by sending IGMP Join requests over the VMkernel network interfaces that are being used for the Virtual SAN traffic service.

Whenever hosts are deployed across different Layer 3 network segments, the result is a routed network topology.

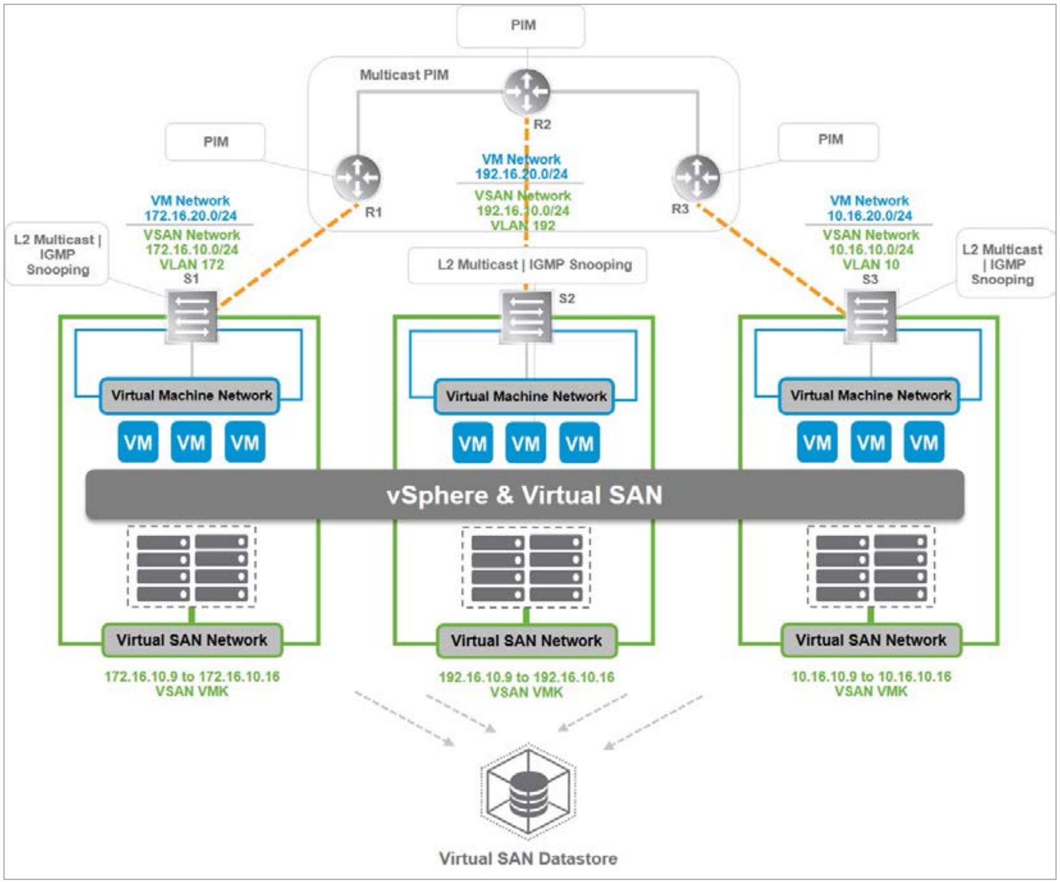


Figure 10: Virtual SAN Over a Layer 3 Network Topology

However, since there is a need for those requests to be sent by each Layer 3 segment Default Gateway, the IGMP Querier has to be the Default Gateway itself.

The Default Gateway will use the Multicast Group memberships from the IGMP Joins to update the PIM protocol running.

In Layer 3 Network topologies, VMware recommends the use and configuration of IGMP Snooping in all the switches configured in the Layer 2 domains where hosts participating in the Virtual SAN cluster will be present.

Layer 3 Physical Network Configuration

This section covers the configuration procedures for IGMP snooping, IGMP Querier and PIM. We will assume that there are three Layer 2 domains, each with its own Layer 3 segment. The Layer 2 domains will be represented by VLANs 10, 172 and 192, as shown in the figure below.

Two configuration examples are provided: one based on the Cisco Nexus platform (with IGMP version 3 and Source Specific Multicast) and the Brocade VDX (with IGMP version 2 and Sparse Mode).

Configuration procedures are typically different based on hardware vendor's implementation. Consult the hardware vendor documentation for in-depth and specific procedures that are beyond the scope of this document.

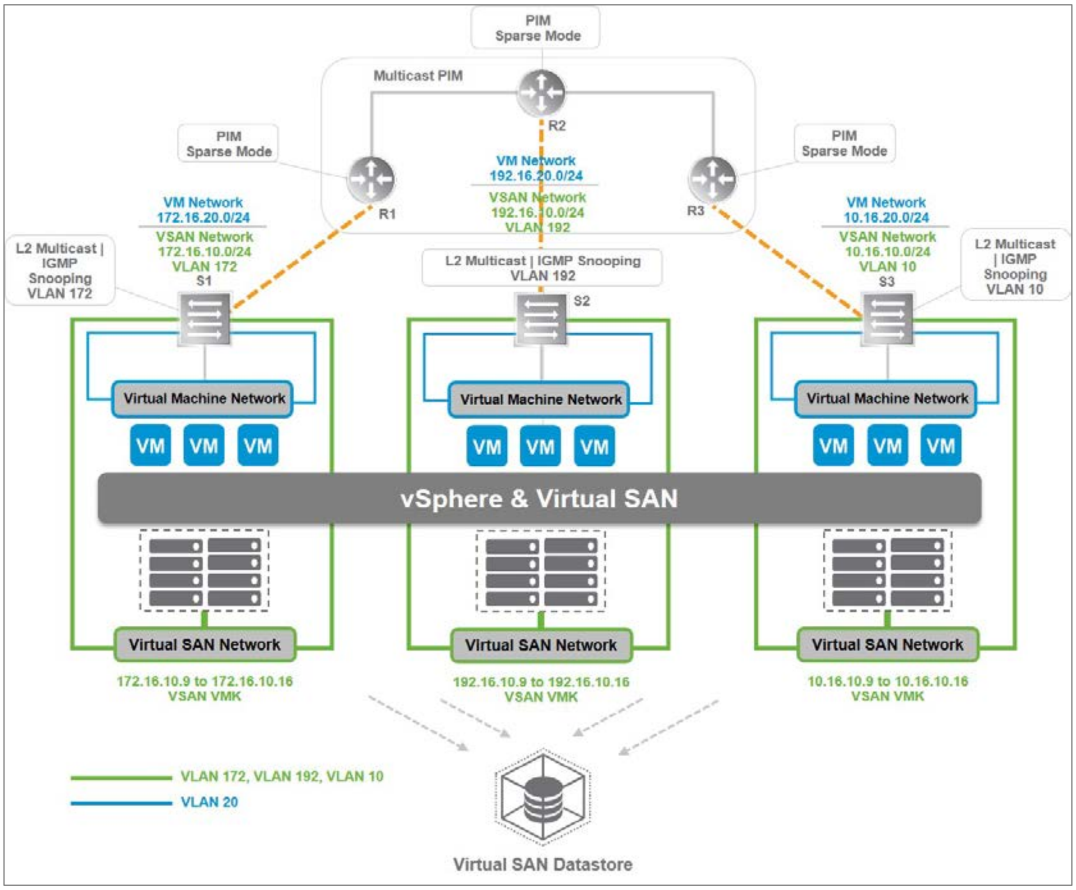


Figure 11: Layer 3 Network Logical Design

NETWORKS	SUBNETS	VLAN	ROUTER	LO0	MGM	AGM	RP
VSAN1	172.16.10.0/24	172	R1	1.1.1.1	224.1.2.3	224.2.3.4	2.2.2.2
VSAN2	192.16.10.0/24	192	R2	2.2.2.2	224.1.2.3	224.2.3.4	2.2.2.2
VSAN3	10.16.10.0/24	10	R3	3.3.3.3	224.1.2.3	224.2.3.4	2.2.2.2

Table 1: Network Information Configuration Table

MGM - Master Group Multicast
AGM - Agent Group Multicast

Lo0 - Loopback Interface 0
RP - Rendezvous Point

Cisco Hardware Devices

The following a sample configuration of IGMP version 3 (enabled by default per VLAN) and Sparse Mode in Nexus 6000 running NX-OS 7.0(3)

Cisco Switch 1 (S1)

```
configure terminal
ip igmp snooping
vlan configuration 172
ip igmp snooping
```

Cisco Switch 2 (S2)

```
configure terminal
ip igmp snooping
vlan configuration 192
ip igmp snooping
```

Cisco Switch 3 (S3)

```
configure terminal
ip igmp snooping
vlan configuration 10
ip igmp snooping
```

Cisco Router 1 (R1)

```
configure terminal
feature pim
ip pim rp-address 2.2.2.2 group-list 224.1.2.3/32
ip pim rp-address 2.2.2.2 group-list 224.2.3.4/32
interface vlan 201
description Network Uplink
ip address 20.1.1.1/30
ip pim sparse-mode

interface vlan 172
ip address 172.16.10.253/24
ip router ospf 9 area 0.0.0.0
ip igmp snooping
ip igmp snooping querier 172.16.10.253

interface Loopback 1
ip address 1.1.1.1/32
ip router ospf 9 area 0.0.0.0
```

Cisco Router 2 (R2)

```
configure terminal
feature pim
ip pim rp-address 2.2.2.2 group-list 224.1.2.3/32
ip pim rp-address 2.2.2.2 group-list 224.2.3.4/32
interface vlan 202

description Network Uplink
ip address 20.1.2.1/30
ip pim sparse-mode
interface vlan 192
ip address 192.16.10.253/24
ip router ospf 9 area 0.0.0.0
ip igmp snooping
ip igmp snooping querier 192.16.10.253

interface Loopback 2
ip address 2.2.2.2/32
ip router ospf 9 area 0.0.0.0
```

Cisco Router 3 (R3)

```
configure terminal
feature pim
ip pim rp-address 2.2.2.2 group-list 224.1.2.3/32
ip pim rp-address 2.2.2.2 group-list 224.2.3.4/32

interface vlan 203
description Network Uplink
ip address 20.1.3.1/30
ip pim sparse-mode

interface vlan 10
ip address 10.16.10.253/24
ip router ospf 9 area 0.0.0.0
ip igmp snooping
ip igmp snooping querier 10.16.10.253

interface Loopback 3
ip address 3.3.3.3/32
ip router ospf 9 area 0.0.0.0
```

Brocade Hardware Devices

The following a sample configuration of IGMP version 2 and Sparse Mode in VDX 6740s running NOS 7.0.0.

Brocade Switch 1 (S1)

```
configure terminal
ip igmp snooping enable
interface vlan 172
ip igmp snooping enable
```

Brocade Switch 2 (S2)

```
configure terminal
ip igmp snooping enable
interface vlan 192
ip igmp snooping enable
```

Brocade Switch 3 (S3)

```
configure terminal
ip igmp snooping enable
interface vlan 10
ip igmp snooping enable
```

Brocade Router 1 (R1)

```
configure terminal
interface vlan 201
interface vlan 172
ip igmp snooping enable
ip igmp snooping querier enable

rbridge-id 101
router pim
rp-address 2.2.2.2

router ospf
area 0.0.0.0

interface loopback 1
ip address 1.1.1.1/32
ip ospf area 0.0.0.0
no shutdown

interface ve 201
description Network Uplink
ip address 20.1.1.1/30
ip ospf area 0.0.0.0
ip pim-sparse
no shutdown

interface ve 172
ip address 172.16.10.1/24
ip ospf area 0.0.0.0
no shutdown
```

Brocade Router 2 (R2)

```
configure terminal
interface vlan 202
interface vlan 192
ip igmp snooping enable
ip igmp snooping querier enable

rbridge-id 102
router pim
rp-address 2.2.2.2

router ospf
area 0.0.0.0

interface loopback 2
ip address 2.2.2.2/32
ip ospf area 0.0.0.0
no shutdown

interface ve 202
description Network Uplink
ip address 20.1.2.1/30
ip ospf area 0.0.0.0
ip pim-sparse
no shutdown

interface ve 192
ip address 192.16.10.1/24
ip ospf area 0.0.0.0
no shutdown
```

Brocade Router 3 (R3)

```
configure terminal
interface vlan 203
interface vlan 10
ip igmp snooping enable
ip igmp snooping querier enable

rbridge-id 103
router pim
rp-address 2.2.2.2

router ospf
area 0.0.0.0

interface loopback 3
ip address 3.3.3.3/32
ip ospf area 0.0.0.0
no shutdown

interface ve 203
description Network Uplink
ip address 20.1.3.1/30
ip ospf area 0.0.0.0
ip pim-sparse
no shutdown

interface ve 10
ip address 10.16.10.1/24
ip ospf area 0.0.0.0
no shutdown
```

Virtual Network Configuration

This section details the configuration procedures for the virtual network components and features such as vSphere Distributed Switch, vSphere Distributed Port Groups, VMkernel Network Interfaces, Virtual SAN Traffic service, and hosts static routes.

Creating vSphere Distributed Switch

Create a vSphere distributed switch on a data center to manage the networking configuration of multiple hosts at a time from a central place.

- From the vSphere Web Client, navigate to a data center.
- In the navigator, right-click the data center and select Distributed Switch > New Distributed Switch.
- In Name and Location, type a name for the new distributed switch and click Next.
- Select version, select the compatible with ESXi 6.0 and later and click Next
- In Edit Settings configure the distributed switch settings according to environment requirements. Click next, then Finish.

Creating vSphere Distributed Port Groups

Add a distributed port group to a vSphere Distributed Switch to create a distributed switch network to associate with VMkernel adapters.

- From the vSphere Web Client, navigate to the distributed switch.
- Right-click the distributed switch and select Distributed port group > New distributed port group.
- In the Select name and location section, type the name of the new distributed port group, VSAN1, and click Next.
- In the Configure settings section, configure VLAN (172), and Failover Order. Set one uplink to active, and the other to standby, then keep the default settings beyond that and click Next, and then Finish.

Creating VMkernel Network Interface for Virtual SAN

Create a VMkernel adapter on a host that is associated with a distributed switch to provide network connectivity to the host and to handle the traffic for Virtual SAN. Dedicate a single distributed port group per VMkernel adapter. For better isolation, you should configure one VMkernel adapter with one traffic type.

- From the vSphere Web Client, navigate to the host
- Under Manage, select Networking and then select VMkernel adapters.
- Click Add host networking.
- On the Select connection type page, select VMkernel Network Adapter and click Next.
- From the Select an existing network option, select a distributed port group and click Next.
- On the Port properties page, configure the settings for the VMkernel adapter based on the network information listed on table 2. Enable the Virtual SAN traffic service, then click Next, then Finish.

Host Configuration Information

NETWORKS	HOSTS	VSAN VMK IP	SUBNETS	VLAN
VSAN1	octo.vsan.a.01	172.16.10.9/24	172.16.10.0/24	172
VSAN1	octo.vsan.a.02	172.16.10.10/24	172.16.10.0/24	172
VSAN1	octo.vsan.a.03	172.16.10.11/24	172.16.10.0/24	172
VSAN1	octo.vsan.a.04	172.16.10.12/24	172.16.10.0/24	172
VSAN2	octo.vsan.b.01	192.16.10.9/24	192.16.10.0/24	192
VSAN2	octo.vsan.b.02	192.16.10.10/24	192.16.10.0/24	192
VSAN2	octo.vsan.b.03	192.16.10.11/24	192.16.10.0/24	192
VSAN2	octo.vsan.b.04	192.16.10.12/24	192.16.10.0/24	192
VSAN3	octo.vsan.c.01	10.16.10.9/24	10.16.10.0/24	10
VSAN3	octo.vsan.c.02	10.16.10.10/24	10.16.10.0/24	10
VSAN3	octo.vsan.c.03	10.16.10.11/24	10.16.10.0/24	10
VSAN3	octo.vsan.c.04	10.16.10.12/24	10.16.10.0/24	10

Table 2: Host Network Information Configuration Table

Adding Host Static Routes

Static routes are used to instruct the Default TCP/IP Stack to use a different default gateway to direct the Virtual SAN traffic through the necessary paths to reach the remote Virtual SAN networks.

Static routes are required by all the hosts between all the different individual Virtual SAN networks.

NETWORKS	SUBNETS	GATEWAYS	VLANS	ROUTERS
VSAN1	172.16.10.0/24	172.16.10.253	172	R1
VSAN2	192.16.10.0/24	192.16.10.253	192	R2
VSAN3	10.16.10.0/24	10.16.10.253	10	R3

Table 3: Virtual SAN Network Addresses

- Static Routes for hosts on VSAN 1 Network:

```
esxcli network ip route ipv4 add –g 172.16.10.253 -n 192.168.10.0/24
esxcli network ip route ipv4 add –g 172.16.10.253 -n 10.16.10.0/24
```

- Static Routes for hosts on VSAN 2 Network:

```
esxcli network ip route ipv4 add –g 192.168.10.253 -n 172.16.10.0/24
esxcli network ip route ipv4 add –g 192.168.10.253 -n 10.16.10.0/24
```

- Static Routes for hosts on VSAN 3 Network:

```
esxcli network ip route ipv4 add –g 10.16.10.253 -n 172.16.10.0/24
esxcli network ip route ipv4 add –g 10.16.10.253 -n 10.16.10.0/24
```

After adding the static routes, the Virtual SAN traffic connectivity should be available across all networks.

Use the vmkping command test and confirm communication between the different networks by pinging the different default gateway from all three networks.

- Test connectivity to remote hosts from VSAN 1 Network:

```
vmkping –I vmk3 192.168.10.253
vmkping –I vmk3 10.16.10.253
```

- Test connectivity to remote hosts from VSAN 2 Network:

```
vmkping –I vmk3 172.16.10.253
vmkping –I vmk3 10.16.10.253
```

- Test connectivity to remote hosts from VSAN 2 Network:

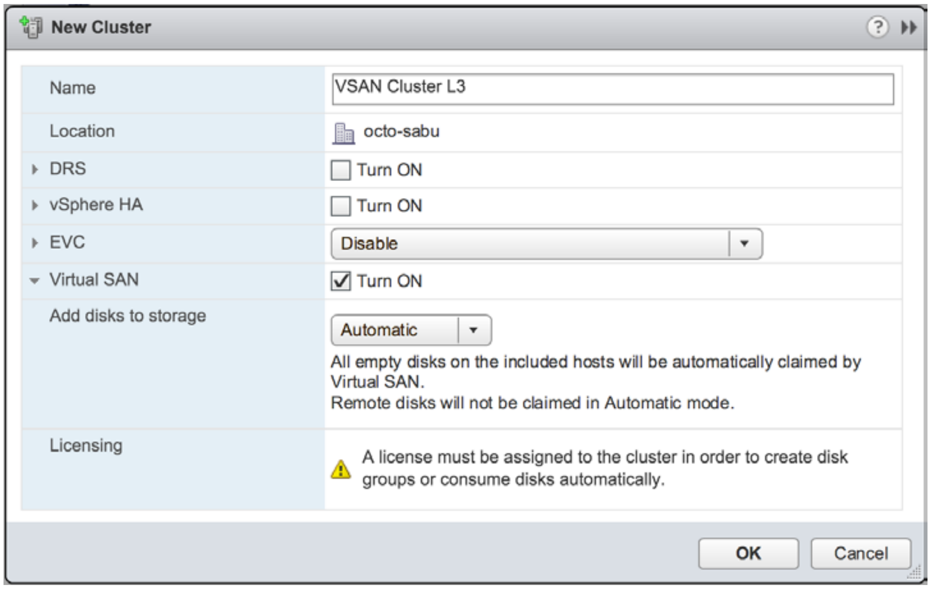
  ```
  vmkping –I vmk3 192.168.10.253
  vmkping –I vmk3 172.16.10.253
  ```

Note: Use vmkping to validate the connectivity across all hosts in all three networks after the VMkernel network interfaces have been created on each host.

Enable and Configure Virtual SAN

Once all the necessary physical and virtual networking configurations have been successfully implemented, it is time to enable Virtual SAN. Virtual SAN can be enabled during or after a vSphere Cluster is created.

- From the vSphere Web Client, navigate to a data center.
- In the navigator, right-click the data center and select right click > New cluster.
- Click **Turn On** Virtual SAN



Name	VSAN Cluster L3
Location	octo-sabu
DRS	□ Turn ON
vSphere HA	□ Turn ON
EVC	Disable
Virtual SAN	☑ Turn ON
Add disks to storage	Automatic All empty disks on the included hosts will be automatically claimed by Virtual SAN. Remote disks will not be claimed in Automatic mode.
Licensing	⚠ A license must be assigned to the cluster in order to create disk groups or consume disks automatically.

Figure 12: Enable Virtual SAN

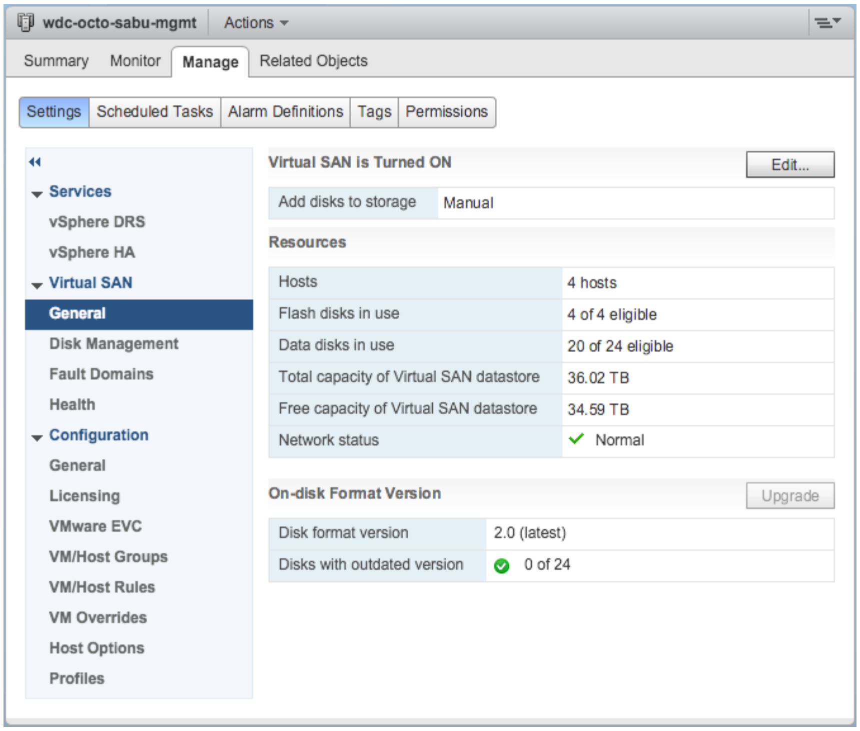
After enabling Virtual SAN, the Virtual SAN storage provider is automatically registered with the vCenter Server and the Virtual SAN datastore is created across the Layer 3 fabric.

Validating Virtual SAN Configuration and Health

Once Virtual SAN has been enabled, the cluster's communication and membership can be validated in multiple ways ranging from the vSphere Web Client to multiple command line interface tools available in vSphere.

The vSphere Web Client offers multiple locations in the UI that offer overall configuration status as well as the health and validation of the network configuration.

Overall Network Status – navigate to the cluster management view and general settings. If all the members of the cluster are successfully communicating via the assigned multicast group and address, the network status is displayed as normal.



Virtual SAN is Turned ON

Add disks to storage	Manual

Resources

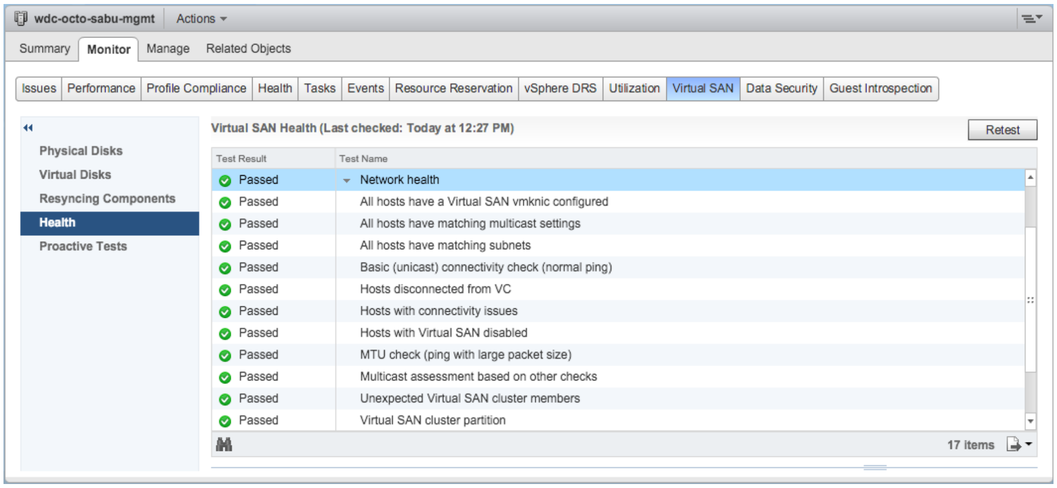
Hosts	4 hosts
Flash disks in use	4 of 4 eligible
Data disks in use	20 of 24 eligible
Total capacity of Virtual SAN datastore	36.02 TB
Free capacity of Virtual SAN datastore	34.59 TB
Network status	✓ Normal

On-disk Format Version

Disk format version	2.0 (latest)
Disks with outdated version	✓ 0 of 24

Figure 13: Virtual SAN Network Communication Status

Detailed Network Health and Multicast Assessment – navigate to the clusters monitoring view for Virtual SAN. Review the Network health section that contains several checkpoints for network health and configuration validation points.



Virtual SAN Health (Last checked: Today at 12:27 PM)

Test Result	Test Name
✓ Passed	▾ Network health
✓ Passed	All hosts have a Virtual SAN vmknic configured
✓ Passed	All hosts have matching multicast settings
✓ Passed	All hosts have matching subnets
✓ Passed	Basic (unicast) connectivity check (normal ping)
✓ Passed	Hosts disconnected from VC
✓ Passed	Hosts with connectivity issues
✓ Passed	Hosts with Virtual SAN disabled
✓ Passed	MTU check (ping with large packet size)
✓ Passed	Multicast assessment based on other checks
✓ Passed	Unexpected Virtual SAN cluster members
✓ Passed	Virtual SAN cluster partition

17 items

Figure 14: Virtual SAN Network Configuration Health

We recommend re-testing the network health and multicast assessment after making any future network changes by clicking the Retest button.

Regardless of the deployment model of choice, Virtual SAN supported hardware options are based on industry standard storage components.

Summary

VMware Virtual SAN is the next evolution in Storage Virtualization. Virtual SAN implementations leverage the already existing IP Network infrastructure to maximize return on investment while reducing OPEX.

From a deployment perspective, the Virtual SAN network stack is flexible and supported over Layer 2 and Layer 3 network topologies.

Virtual SAN implementations over Layer 2 network topologies present the least amount of network complexity to implement and simplest option to manage and maintain when compared to Layer 3 network topology deployments.

Either way, VMware Virtual SAN deployments can be performed on Layer 2 as well as Layer 3 networking topologies right out-of-the box.

Acknowledgments

I would like to thank: Christos Karamanolis, CTO of the Storage and Availability Business Unit at VMware; Duncan Epping, Chief Architect; Cormac Hogan, Senior Staff Engineer in the Office of the CTO of the Storage and Availability Business Unit; Madhu Nalamati, Staff Network Engineer; Ken Werneburg, Manager of the Storage and Availability Technical marketing team at VMware for reviewing this paper.

I would also like to extend a special thanks to Elver Sena Sosa, CCIE 7321 and VCDX 154 Network Virtualization for his time and contributing to this paper.

Author

Rawlinson Rivera is a Principal Architect in the Office of the CTO of the Storage and Availability Business Unit at VMware, Inc. He specializes in cloud enterprise architectures, Hyper-converged Infrastructures (HCI).

Primarily focused on Software-Defined Storage such as Virtual SAN, vSphere Virtual Volumes, as well as storage related solutions for OpenStack and Cloud-Native Applications. He serves as a trusted adviser to VMware's customers primarily in the US.

Rawlinson is among the few VMware Certified Design Experts (VCDX #86) in the world, and author of multiple books based on VMware and other technologies. He is the owner and main author of virtualization blog punchingclouds.com.

- Follow Rawlinson’s blogs:
  http://blogs.vmware.com/virtualblocks/
  http://www.punchingclouds.com/
- Follow Rawlinson on Twitter: @PunchingClouds