

VMware Storage



Table of contents

vSAN Proof of Concept: vSAN Stretched Cluster & Two-Node Overview & Testing
vSAN Stretched Cluster and Two node 4
Design and Overview
Witness Configuration 4
Witness Traffic Separation \ldots 6
Shared Witness
Stretched Cluster Setup
Stretched Cluster Setup: Manual Configuration 11
Stretched Cluster Hosts
Stretched Cluster Network Topology 13
Example Stretched Cluster Network Configurations 13
vSAN Stretched Cluster Lab Setup
Preferred / Secondary Site Details
vSphere HA Settings
Response to Host Isolation
Admission Control 20
Advanced Settings
Host/VM Groups and Site Affinity 22
DRS Settings
vSAN Stretched Cluster Local Failure Protection25
vSAN Stretched Cluster Site Affinity
vSAN Stretched Cluster Preferred Site Override 26
vSAN Stretched Cluster and Two Node Failure Scenarios 28
Scenario #1: Network Failure between Data Site and Witness
Trigger the Event
Cluster Behavior on Failure
Conclusion
Repair the Failure
Scenario #2: Network Failure from both Data sites to Witness
Trigger the Event



Cluster Behavior on Failure
Conclusion
Repair the Failure
Scenario #3: Secondary Site Failure
Trigger the Event
Cluster Behavior on Failure
Conclusion
Repair the Failure
Scenario #4: Primary Site Failure
Trigger the Event
Cluster Behavior on Failure
Conclusion 40
Repair the Failure 40
Scenario #5: Network Failure between Data Sites but Witness Still Accessible
Trigger the Event
Cluster Behavior on Failure 42
Conclusion 43
Repair the Failure



vSAN Stretched Cluster and Two node

Design and Overview

Good working knowledge of how vSAN Stretched Cluster is designed and architected is assumed. Readers unfamiliar with the basics of vSAN Stretched Cluster are urged to review the relevant documentation before proceeding. Details on how to configure a vSAN stretched cluster are found in the <u>vSAN Stretched Cluster Guide</u>. Both ESA and OSA architectures are supported with stretched cluster.

Witness Configuration

The 'witness' is usually deployed as an appliance, in a site (fault domain) separate to the two data sites. Although the witness appliance requires access to the vSAN network, only metadata stored on the appliance.

The Witness Appliance can be downloaded from <u>My VMware</u>. The Witness Appliance is deployed as an OVF template, as per the example below:



Note that while two VMkernel adapters are deployed by default on the Witness Appliance, another valid configuration is to tag both



vSAN and Management traffic on a single VMkernel adapter for simpler deployments.

After deployment of the OVF, add the Witness Appliance to the vCenter inventory as you would a physical host. Alternatively, customers can use a physical ESXi host for the witness.

The Witness Appliance host must not be added to a cluster, placement should be at the Datacenter level.

Enter the password as per the OVF properties:

Add Host	Connection settings			×
1 Name and location				
2 Connection settings	User name:	root		
2. Heat summary	Password:	******		
3 Host summary				
4 Assign license				
5 Lockdown mode				
6 VM location				
7 Ready to complete			CANCEL BACK	NEXT

Assign the embedded witness appliance license:

Add Host	Assign license Assign an existing license to this host			×
1 Name and location		•	Usage	Capacity
2 Connection settings	> Virtual SAN Witness for Embedded OEMs	,	• 1 CPUs (up t	1 CPUs (up to
3 Host summary	Normico Blue with Add on for Kubaratae		O CDUs (up to	3 items
4 Assign license	Assignment Validation for License 1			
5 Lockdown mode	O The license assignment is valid.			
6 VM location				
7 Ready to complete			CANCEL	CKNEXT

Ensure the location is set to the top-level Datacenter:

Add Host	VM location ×
1 Name and location	Select a location for this nost's virtual machines Datacenter
2 Connection settings	
3 Host summary	
4 Assign license	
5 Lockdown mode	
6 VM location	
7 Ready to complete	CANCEL BACK NEXT

Once added to vCenter, the Witness Appliance will appear as a host with special properties – again note that the host is deployed at the DC level (note also the OVA "OSA-Witness" deployed below):

\equiv vSphere Client $$ Q Search in			or@VSPHERE.LOCAL ~	♀ ?
Control Con	Image: WC-dxb-8.vsanpe.vmware.com : Accords and the second se	tacenters Hosts & Cluster	rs VMs Datastores	
 VSAN Cluster (single site) 10.159.21.10 10.159.21.11 10.159.21.2 10.159.21.9 OSA-Witness 10.156.144.153 	Name ↑ State □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □	Status	Cluster Versi 8.0.0))
Recent Tasks Alarms	LI EXPORT V		items per page	35 🗸 1 item

Witness Traffic Separation

When a vSAN Witness appliance is deployed, a separately tagged VMkernel interface may be used for witness traffic transit instead of extending the vSAN network to the witness host. This feature allows for a more flexible network configuration by allowing for separate networks for node-to-node vs. node-to-witness communication. Note that this capability can only be enabled from the command line.

Witness Traffic Separation provides the ability to directly connect vSAN data nodes in a 2-node configuration. Traffic destined for the Witness host can be tagged on an alternative physical interface separate from the directly connected network interfaces carrying vSAN traffic. Direct Connect eliminates the need for a dedicated switch at remote offices/branch offices (where the additional cost of the switch could be cost-prohibitive to the solution).





For the example illustrated above, to enable Witness Traffic on vmk1, execute the following on both hosts:

esxcli vsan network ip add -i vmkl -T=witness

Any VMkernel port not used for vSAN traffic can be used for Witness traffic.

In a more simplistic configuration (as shown below), the Management VMkernel interface (vmk0) could be tagged for Witness traffic. The VMkernel port tagged for Witness traffic needs to have IP connectivity to the vSAN traffic tagged interface on the vSAN Witness Appliance.





The minimum requirement for this setup is a single vnic uplink and vmkernel port carrying both management and vSAN traffic.

Shared Witness

A 'shared' witness can be used for several vSAN two-node setups (i.e. across multiple two-node clusters), reducing the number of separate witness appliances needed. Note that shared witness is not supported for stretched cluster.





For more information on vSAN Witness VM sizes refer to the official documentation: https://docs.vmware.com/en/VMware-vSphere/8.0/vsan-planning/GUID-05C1737A-5FBA-4AEE-BDB8-3BF5DE569E0A.html

Stretched Cluster Setup

To start the configuration wizard for either vSAN stretched cluster or two-node, navigate to [Cluster] > Configure > vSAN > Services and pick the appropriate option:





Follow the screens to configure the disks, as per a standard vSAN cluster. Then chose the vSAN witness host:

Configure cluster	Select witness host	×
1 Distributed switches	Select a host which will store all the witness components for this vSAN Stretched Cluster. Requirements for witness host:	
2 Advanced options	Not part of any vSAN enabled clusterHave at least one VMkernel adapter with vSAN traffic enabled	
3 Claim disks	That adapter must be connected to all hosts in the Stretched cluster	
4 Configure fault domains	Q Search	
5 Select witness host	 	
6 Claim disks for witness host	> 🔀 New Cluster	
7 Ready to complete		
	O Compatibility checks succeeded.	
	CANCEL BACK NE	кт

For an OSA cluster, select the disks for the witness host:



Configure cluster	Claim disks for witness host
1 Distributed switches	Select disks on the witness host to be used for storing witness components. First, select a single disk to serve as cache tier.
2 Storage traffic	Name \(\tau\) Drive Type \(\tau\) Capacity \(\tau\) Transport Type \(\tau\) Adapter
	C E Local VMware D.,. Flash 10.00 GB
3 Advanced options	C E Local VMware D., Flash 10.00 GB
4 Claim disks	C E Local VMware D.,, Flash 30.00 GB 6
5 Configure fault domains	Then, select one or more disks to serve as capacity tier. Capacity type: Flash
6 Select witness host	Name v Drive Type v Capacity v Transport Type v Adapter
	Local VMware D.,, Flash 10.00 GB
7 Claim disks for witness host	Local VMware D., Flash 10.00 GB
8 Ready to complete	Local VMware D.,, Flash 30.00 GB 6
	CANCEL BACK NEXT

Stretched Cluster Setup: Manual Configuration

Instead of using the configuration wizard above, to manually configure the stretched cluster (after vSAN has been enabled) navigate to **[vSAN cluster] > Configure > vSAN > Fault Domains** and click on '**Configure Stretched Cluster**':

📃 vSphere Client 🤇	χ Search in all environments	C S Admini	istrator@VSPHERE.LOCAL ~ 🙄 🕐 ~
> []] vSAN-8 Cluster Summary Monitor	Configure Permissions Hosts VMs Dat	tastores Networks Updates	
Host Profile	Fault Domains		
Licensing V	Fault domain failures to tolerate	1	
Trust Authority	Configuration type	Single site	CONFIGURE STRETCHED CLUSTER
Alarm Definitions Scheduled Tasks	To manage the fault domains, use the Actions me	nu or drag and drop the host.	
vSphere Cluster Ser ∨ Datastores	+		
vsan v	Click the plus icon or drag and drop here to		
Services	create new fault domain.		
Disk Management Fault Domains Remote Datastores	Standalone Hosts (4) (j)		
Recent Tasks Alarn	าร		

Configure the hosts into the appropriate domain:



Configure Stretched Cluster	Configure fault domains		×
1 Configure fault domains	Divide the hosts in 2 fault domains that	will be used for configuring vSAN stretched cluster.	
2 Select witness host	Preferred	Secondary	
3 Claim disks for witness host	10.159.21.1010.159.21.12	>>> [] 10.159.21.9	
4 Review			
		CANCEL	ΞХТ

Select the Witness host:

Configure Stretched Cluster	Select witness host X		
1 Configure fault domains	Select a host to store all witness components for this vSAN Stretched Cluster. A host must meet certain requirements to be used as a witness host. (
2 Select witness host	Q Search		
3 Claim disks for witness host4 Review	 Bhattd-vcsa-7u2.vsanpe.vmware.com Datacenter 10.156.144.173 		
	Compatibility checks succeeded.		

For an OSA witness, the first time you configure vSAN with a witness host you will claim the disks used by the witness, when adding additional clusters this step is skipped.



Configure Stretched Cluster 1 Configure fault domains	ster Claim disks for witness host Select disks on the witness host to be used for storing witness components. Select a single disk for the cache tier.			
2 Select witness host	Name T	Drive Type 🔻 Capacity 🔻	Transport Type 🔻 Adapter 🔻	
3 Claim disks for witness host	💿 ╞ Local VMware Disk (mpx.vmh	Flash 10.00 GB		
	Select one or more disks for the capacity ti Capacity type: HDD	er. 	1 item	
	Name	HDD 15.00 GB	Transport Type Y Adapter Y	
			1 item CANCEL BACK NEXT	

This is covered in greater detail in the vSAN Stretched Cluster Guide.

Stretched Cluster Hosts

Our example has four ESXi hosts in a cluster, two ESXi hosts on data site A (the "preferred" site), and two hosts on data site B (the "secondary" site). There is one disk group per host. The witness host/appliance is deployed in a 3^{rd} remote data center. This configuration is referred to as 2+2+1.

VMs are deployed on both the "Preferred" and "Secondary" sites of the vSAN Stretched Cluster. VMs are running/active on both sites.

Stretched Cluster Network Topology

As per the vSAN Stretched Cluster Guide, several different network topologies are supported for vSAN Stretched Cluster. Considerations include layer 2 (same subnet) or layer 3 (routed) connectivity between the three sites with or without Witness Traffic Separation (WTS) depending on the requirements.

Some of the common designs listed below. Options 1 and 2 are configurations without WTS. The only difference between them is whether the network is stretched (Layer 2) or routed (Layer 3) for vSAN data traffic. Option 3 uses Witness Traffic Separation. For simplicity, all options use L2 for VM traffic. During testing, you may choose to test one or another, or both options if you wish.

For more information on network design best practices for the stretched cluster, refer to the vSAN Stretched Cluster Guide.

Example Stretched Cluster Network Configurations

As per the vSAN Stretched Cluster Guide, several different network topologies are supported for vSAN Stretched Cluster. The options below provide some of the different for stretched cluster network configuration.

Option 1:

- L3 for witness traffic, without Witness Traffic Separation
- L2 for vSAN data traffic between data sites
- L2 for VM traffic





Option 2:

- L3 for witness traffic, without Witness Traffic Separation
- L3 for vSAN data traffic between data sites
- L2 for VM traffic





Option 3:

- L3 for witness traffic with Witness Traffic Separation
- L2 for vSAN data traffic between data sites
- L2 for VM traffic







vSAN Stretched Cluster Lab Setup

As per the vSAN Stretched Cluster Guide, several different network topologies are supported for vSAN Stretched Cluster.

The network topology deployed in this lab environment for our test case is layer 2 between the vSAN data sites and L3 between data sites and witness. ESXi hosts and vCenter are in the same L2 subnet for this setup. The VM network should be a stretched L2 between both data sites as the unique IP used by the VM can remain unchanged in a failure scenario.

There are four ESXi hosts in this cluster, two ESXi hosts on data site A (the "preferred" site) and two hosts on data site B (the "secondary" site). There is one disk group per host. The witness host/appliance is deployed in a 3^{rd} , remote data center. The configuration is referred to as 2+2+1.

VMs are deployed on both the "Preferred" and "Secondary" sites of the vSAN Stretched Cluster. VMs are running/active on both sites.

Below is a diagram detailing the environment used for the Stretched Cluster testing.



- This configuration uses a stretched (layer 2) network for vSAN traffic between sites
- A routed network is used between data each data site and the witness site
- For both data sites, the vSAN network is on VLAN 3921. The gateway is 172.21.0.1
- The Witness Appliance: vmk0 has been tagged for management and vSAN and is on the native vLAN (vLAN 0). The gateway is 10.156.159.253

The VM network is stretched (layer 2) between the data sites on VLAN 106 (since no VMs run on the Witness Appliance, there is no need to extend this network to the third site)

Preferred / Secondary Site Details

In vSAN Stretched Clusters, the 'preferred' site simply means the site that the witness will 'bind' to in the event of an inter-site link failure between the data sites. This will also be the site where all VMs will run when there is an inter-site link failure.

For our failure scenarios, we create two DVS port groups and add the appropriate vmkernel port to each, ready to test the failover behavior:



vSphere Client Q Search in all environments	C	© .
DSwitch : ACTIONS		
Summary Monitor Configure Permissions Ports Hosts VMs Networks		
Distributed Port Groups Uplink Port Groups		
		Y vsan
Name ↑ VLAN ID NSX Port Group ID	VNI	Port Binding
ULAN access: 3921		Static binding (elastic)
ULAN access: 3921		Static binding (elastic)
EXPORT ~		ltems per page 35 \lor 2 items
Recent Tasks Alarms		

On each of the hosts, vSAN traffic is tagged on vmk1, with the default gateway set on the vSAN network

IU.133.21.25 : *	CHONS					
Summary Monitor Cor	nfigure	Permissions VMs Dat	astores Network	s Updates		
Storage	\sim	VMkernel adapters				
Storage Adapters						
Storage Devices		ADD NETWORKING REFRE	:5H			
Host Cache Configuration		Device	т			
Protocol Endpoints		: >> 🖭 vmk0		VMkernel network a	adapter: vmk1	
I/O Filters		: 🔍 🖭 vmk1		All Properties IP Settir	ngs Policies	
Networking	~	: >> == vmk2		_		
Virtual switches				✓ Port properties		
VMkernel adapters				Network label	vSAN-Preferred-vlan3921	
Physical adapters				TCP/IP stack	Default	
TCP/IP configuration				Enabled services	VSAN	
Virtual Machines	~			> NIC settings		
VM Startup/Shutdown						
Agent VM Settings				✓ IPv4 settings		
Default VM Compatibility				DHCP	Disabled	
Swap File Location				IPv4 address	172.21.4.5 (static)	
System	~			Subnet mask	255.255.240.0	
Licensing				Default gateway	172.21.0.1	
Host Profile				DNS server addresses	10.159.18.10	
Time Configuration				> IPv6 settings		
Authentication Services		3 items				

The routing table shows the gateways set:



\equiv vSphere Client Q) Search			Ca	Administrator@VSPHER		© ~
Summary Monitor	ACTION	e Permissions VMs Datastores Network:	s Updates				
Networking Virtual switches VMkernel adapters Physical adapters	~	TCP/IP CONTIGUIATION	TCP/IP Stack	: Default	IPv6 Routing Table	Advanced	X
Virtual Machines VM Startup/Shutdown Agent VM Settings Default VM Compatibility	~	: >> mirror : >> Provisioning : >> vMotion	Network Address 10.159.16.0 172.21.0.0 Default	Subnet Mask 255.255.240.0 255.255.240.0 0.0.0.0	Gateway Local Subnet Local Subnet 10.159.31.253	Device vmk0 vmk1 vmk0	
Swap File Location System Licensing List Brafile	~	1-5/5 < 1 >					3 items

The witness appliance has one vmkernel adapter, with both management and vSAN tagged. The default gateway is on the management network.

vSphere Client Q Searc	h in all environments C & Administrator@VSPHERELOCAL ~ 😳	?
■ 10.156.144.153 : AC	lions	
Summary Monitor Configur	e Permissions VMs Resource Pools Datastores Networks Updates	
Networking V	VMkernel adapters	
Virtual switches	ADD NETWORKING REFRESH	
VMkernel adapters		
Physical adapters	Device T VMkernel network adapter: vmk0	\sim
TCP/IP configuration		×
Virtual Machines V	All Properties IP Settings Policies	
VM Startup/Shutdown		
Agent VM Settings	 Port properties 	
Default VM Compatibility	Network label Management Network	
Swap File Location	VLAN ID None (0)	
System 🗸	TCP/IP stack Default	
Licensing	Enabled services Management	
Host Profile	VSAN	
Time Configuration	> NIC settings	
Authentication Services		
Certificate	✓ IPv4 settings	
Power Management	DHCP Enabled	
Advanced System Settings	IPv4 address 10.156.144.153 (DHCP)	
System Resource Reservation	Subnet mask 255,255,240.0	
Firewall	Default gateway 10.156.159.253	
Services	DNS server addresses 10.156.128.10	
Security Profile	10.196.128.11	
System Swap	1item > IPv6 settings	
Recent Tasks Alarms		

Note: The Witness Appliance is a nested ESXi host and requires the same treatment as a standard ESXi host (i.e, for lifecycle management). Keep all ESXi hosts in a vSAN cluster at the same update level, including the Witness appliance.

vSphere HA Settings

vSphere HA plays a critical part in Stretched Cluster. HA is required to restart virtual machines on other hosts and even the other site depending on the different failures that may occur in the cluster. The following section covers the recommended settings for vSphere HA in a Stretched Cluster environment.

Response to Host Isolation

The recommendation is to "Power off and restart VMs" on isolation, as shown below. In cases where the virtual machine can no longer access the majority of its object components, it may not be possible to shut down the guest OS running in the virtual machine. Therefore, the "Power off and restart VMs" option is recommended.



onere HA			
lures and responses	Admission Control	Heartbeat Datastores	Advanced Options
supported: host, host	isolation, VM compone	ent protection (datastore with	PDL and APD), VM and application.
> Host Failure Response	5	Restart VMs ~	
Response for Host Iso	lation	Disabled ✓ Power off and restart VMs Shut down and restart VM	s 1s
Datastore with PDL		Power off and restart VMs	<u>~</u>
Datastore with APD		Power off and restart VMs -	- Conservative restart policy \smallsetminus
VM Monitoring		Disabled	<u> </u>
the Monitoring			

Admission Control

If a full site fails, the desire is to have all virtual machines run on the remaining site. To allow a single data site to run all virtual machines if the other data site fails, the recommendation is to set Admission Control to 50% for CPU and Memory as shown below.



Edit Cluster Settings	OSA-Stretched		×
potential host failures will increase the	ne availability constraints and capacity reserved.		
Host failures cluster tolerates	1 Ç Maximum is one less than number of hosts in cluster.		
Define host failover capacity by	Cluster resource Percentage		
	Override calculated failover capacity.		
	Reserved failover CPU capacity:	50	
	Reserved failover Memory capacity:	50	≎ % Memo
	Reserve Persistent Memory failover capacity ((i)	
		[_ .	
		CANCEL	ок

Advanced Settings

The default isolation address uses the default gateway of the management network. This will not be useful in a vSAN Stretched Cluster when the vSAN network is broken. Therefore, the default isolation response address should be turned off. This is done via the advanced setting *das.usedefaultisolationaddress* to *false*.

To deal with failures occurring on the vSAN network, VMware recommends setting at least one isolation address which is reachable from each of the data sites. In this example, we use the gateway IP address of the witness host network. Use the advanced setting *das.isolationaddress0* to set the isolation address for the IP gateway address to reach the witness host.

Finally, we set *das.ignoreRedundantNetWarning* to silence the warnings for redundant networks.



Edit Cluster Settings OSA-Stretched >					
vSphe	ere HA 🚺				
Failur	es and responses Admission Control	Heartbeat Datastores	Advanced Options		
Option	Value		ADD		
	Ontion	Value			
:	das.ignoreRedundantNetWarning	true			
÷	das.isolationaddress0	10.156.1	44.153		
÷	das.usedefaultisolationaddress	false			
				3 items	
				CANCEL	

Note: Since vSAN 6.5 there is no need for VM anti-affinity rules or VM to host affinity rules in HA

Host/VM Groups and Site Affinity

The next step is to configure the Host/VM groups and define which VMs should run on each site. VMs will be restarted on the remote site only if there is a catastrophic failure (or a significant resource shortage).

First, we create host groups, by navigating to [Cluster] > Configure > Configuration > VM/Host Groups.

\equiv vSphere Client	Q	Search in all environments	? ~
> (OSA-Stretch	ed	ACTIONS	
Summary Monitor	Cor	figure Permissions Hosts VMs Datastores Networks Updates	
Services	~	VM/Host Groups	
vSphere DRS vSphere Availability		ADD DELETE T	
Configuration	\sim	Name Type	
Quickstart		\bigtriangledown	
General			
Key Provider			
VMware EVC		No items found	
VM/Host Groups			0 items
VM/Host Rules			
VM Overrides			
I/O Filters			
Host Options			
Host Profile			
Licensing	\sim	No VM/Host group selected	
vSAN Cluster			
Trust Authority			
Alarm Definitions			
Scheduled Tasks			
Recent Tasks Al	larms		

Create groups for the primary and secondary site for both the hosts and VMs. In the example below, the Host Groups are named Preferred and Secondary:



\equiv vSphere Client Q	Search in all environments C & Administrator@VSPHERE.LOCAL ~ 🕞	?~
> 🗊 OSA-Stretched	ACTIONS	
Summary Monitor C	onfigure Permissions Hosts VMs Datastores Networks Updates	
Configuration V	VM/Host Groups	
Quickstart	ADD DELETE T	
General	Name Type	
Key Provider	Image: Contract of the second seco	
VMware EVC		
VM/Host Groups		
VM/Host Rules		
VM Overrides	Image: Secondary size Vm Group	4.14
Host Options		4 items
Host Profile		
Licensing V		
vSAN Cluster		
Trust Authority		
Alarm Definitions	No VM/Host group selected	
Scheduled Tasks		
vSphere Cluster Services ∨		
Datastores		
Desired State		
Recent Tasks Alarms	3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	

Next, we define the affinity rules that specify where the VMs should run. Navigate to [Cluster] > Configure > Configuration > VM/Host Rules

📃 vSphere Client	Q	Search in all environments				C	Administrator@VSPHER	e.local V	٢	?~
> (1) OSA-Stretche	ed _{Cor}	ACTIONS	VMs Datastores Networks	Updates						
Services vSphere DRS vSphere Availability	~	VM/Host Rules								
Configuration		Name	Туре	Enable	d		Conflicts	Defined By		
Comguration	Ť	💿 📋 Primary VMs	Run VMs on Hosts	Yes			0	User		
Quickstart										
General										
VMware EVC										
VM/Host Groups										
VM/Host Rules										
VM Overrides										1 item
I/O Filters										
Host Options		VM/Host Rule Details								
Host Profile		Virtual Machines that are members of	the VM Group should run on hosts th	at are men	bers of the Ho	ost Group.				
Licensing	~	ADD REMOVE			ADD RE	EMOVE				
VSAN Cluster		VMs Preferred Site Group Members		\uparrow	Hosts Pr	referred Site	Group Members			\uparrow
Alarm Definitions		Star VM 1 Primary			□ E 10	.159.21.25				
Scheduled Tasks		M 2 Primary			□ □ □ 10.	159.21.26				
vSphere Cluster Services	s∨					159 21 27				
Datastores						150 01 00				
Desired State					U [10.	.159.21.28				
Desired State	~			4 items						4 items
Recent Tasks Ala	arms									

Here we add a *should* "VM to Host" rule that binds the primary site VMs to the preferred site:



Create VM/Ho	ost Rule OSA-Stret	tched ×
Name	Primary VMs	Enable rule.
Туре	Virtual Machines to H	Hosts ~
Site should run on host VM Group: VMs Preferred Site	group Hosts Preferred Site.	~
Should run on hosts in	group	~
Host Group:		
Hosts Preferred Site		~
		CANCEL

Note that "should" rules (as opposed to "must" rules) mean that every attempt will be made to adhere to the affinity rules. However, if this is not possible (due lack of resources), the other site will be used for hosting the virtual machine.

Also, note that the vSphere HA rule setting is set to "should". This means that if there is a catastrophic failure on the site to which the VM has an affinity, HA will restart the virtual machine on the other site. If this was a "must" rule, HA would not start the VM on the other site.

The same settings are necessary on both the primary VM/Host group and the secondary VM/Host group.

DRS Settings

In this example, DRS is set to 'partially automated', which is the recommended setting for a stretched cluster. This is to avoid failback of VMs occurring while rebuild activity is still taking place. More on this later.

vSphere DRS is Turned ON

> DRS Automation	Partially Automated
> Additional Options	Expand for policies
> Power Management	Off
> Advanced Options	None



vSAN Stretched Cluster Local Failure Protection

We build on resiliency by including local failure protection, which provides storage redundancy within each site and across sites. Local failure protection is achieved by implementing local RAID-1 mirroring or RAID-5/6 erasure coding within each site. This means that we can protect the objects against failures within a site. For example, if there is a host failure on site 1, vSAN can selfheal within site 1 without having to go to site 2 if properly configured.



Local Failure Protection is configured and managed through storage policies. The figure below shows rules in a storage policy that is part of an all-flash stretched cluster configuration. The "Site disaster tolerance" is set to **Dual site mirroring (stretched cluster)**, which instructs vSAN to mirror data across the two main sites of the stretched cluster. The "Failures to tolerate" specifies how data is protected within the site. In the example storage policy below, **1 failure - RAID-5 (Erasure Coding)** is used, which can tolerate the loss of a host within the site.

Create VM Storage Policy	vSAN	×
1 Name and description	Availability Storage rules	Advanced Policy Rules Tags
2 Policy structure	Site disaster tolerance (j)	Site mirroring - stretched cluster
3 vSAN	Failures to tolerate (j)	1 failure - RAID-5 (Erasure Coding) Consumed storage space for 100 GB VM disk would be
4 Storage compatibility		266.67 GB
5 Review and finish		
		CANCEL BACK NEXT

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



to achieve this level of resiliency.

vSAN Stretched Cluster Site Affinity

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

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



vSAN Stretched Cluster Preferred Site Override

Preferred and secondary sites are defined during cluster creation. If it is desired to switch the roles between the two data sites, you can navigate to **[vSAN cluster] > Configure > vSAN > Fault Domains**, select the ellipses to the right of the 'Secondary' site pane and click the 'Set as preffered' option.



\equiv vSphere Client $$ $$ $$ $$ $$ $$		C & Adminis	trator@VSPHERE.LOCAL ~
VSAN-OSA Clus Summary Monitor Co VM/Host Rules VM Overrides I/O Filters	ter2 : ACTIONS onfigure Permissions Hosts VMs Dat Fault Domains Fault domain failures to tolerate	astores Networks Updates	
Host Options Host Profile	Witness host	■ 10.156.144.144	CHANGE WITNESS HOST
vSAN Cluster Trust Authority Alarm Definitions Scheduled Tasks vSphere Cluster Services ~ Datastores	Preferred (preferred) :: Used capacity 2% 10.159.16.116 2% 10.159.16.115 1%	Secondary Used capacity Secondary Set as preferred Edit Delete	<u>.</u>
vSAN ✓ Services Disk Management Fault Domains Remote Datastores			
Desired State Image Configuration Recent Tasks Alarms			



vSAN Stretched Cluster and Two Node Failure Scenarios

In this section, we will look at how to inject various network failures in a vSAN Stretched Cluster configuration. We will see how the failure manifests itself in the cluster, focusing on the vSAN health check and the alarms/events as reported in the vSphere client.

Note that network failover scenarios for stretched Cluster (with or without witness traffic separation) and 2-node (with or without direct connect) are the same because the Witness traffic is always connected via routed (L3) network.

Scenario #1: Network Failure between Data Site and Witness



Trigger the Event

To make the either the primary or secondary site lose access to the witness site, one can simply remove the gateway configured on the vSAN VMkernel adapter.

First choose which site to operate on (either the primary or secondary hosts). Open an SSH session to each host on the chosen site. Our first step is to confirm which VMkernel adapter is being used for vSAN:

```
[root@localhost:~] esxcli vsan network list | grep VmkNic
VmkNic Name: vmk1
```

Then get the details for this adapter, and make a note of the output:

[root	@localhost:~]	esxcli network	ip interface ipv	4 get -i vmkl		
Name	IPv4 Address	IPv4 Netmask	IPv4 Broadcast	Address Type	Gateway	DHCP DNS
vmk1	172.21.4.12	255.255.240.0	172.21.15.255	STATIC	172.21.0.1	false

We can then effectively remove the gateway set for this adapter by re-applying the same IP address and omitting the gateway

Cluster Behavior on Failure

In such a failure scenario (where the witness is isolated from one of the data sites) the implication is that there is no communication to both the master node AND the backup node. In stretched clusters, the master node and the backup node are placed on different fault domains (i.e. sites).

This is the case in this failure scenario. The witness becomes isolated, and the nodes on the preferred and secondary sites remain in the cluster. We can see how this is shown in vCenter below.

To begin with, the **Cluster Summary** view shows us a few errors:

Clicking on the 'view all issues' link shows the full extent of the issues:

📃 vSphere Client	Q Sea			C	2	Administrator@VSPHERE.	LOCA	L ~	? ~
> OSA-Stretche	ed Config	ACTIONS ure Permissions Hosts VMs Datastores Networks U	pdates						
Issues and Alarms	~	All Issues							
All Issues Triggered Alarms		PREVIOUS NEXT ACKNOWLEDGE RESET TO GREEN							
Deutemann		Issue	т	Туре	т	Trigger Time	т	Status	т
Performance	~	VSAN performance service alarm 'Stats DB object'		Triggered Alarm		11/22/2022, 04:14 PM		Alert	
Overview		vSAN data alarm 'vSAN object'		Triggered Alarm		11/22/2022, 04:14 PM		Alert	
Advanced		vSAN network alarm 'vSAN: MTU check (ping with large packet size)'		Triggered Alarm		11/22/2022, 04:14 PM		Alert	
Tasks and Events	~	VSAN network alarm 'vSAN: Basic (unicast) connectivity check'		Triggered Alarm		11/22/2022, 04:14 PM		Alert	
Tasks		VSAN network alarm 'vSAN cluster partition'		Triggered Alarm		11/22/2022, 04:14 PM		Alert	
Events									
vSphere DRS	~								
Recommendations									
Faults									
History									
VM DRS Score									
CPU Utilization									
Memory Utilization									
Network Utilization									5 items
 Recent Tasks Ala 	arms								

On navigating to the **[vSAN Cluster] > Monitor > vSAN > Health** view, there are a lot of checks showing errors, including failed pings to the witness appliance:

= vSphere Client Q Search in all environments C & Administrator@VSPHERELOCAL ~ © ?									
> OSA-Stretched	ACTIONS	oras Notworks Undatos							
CPU	Skyline Health	vSAN: MTU check (ping with	large packet size)		SILENCE ALERT				
Memory Persistent Memory	Last checked: 11/22/2022, 10:52:40 AM RETEST	A This health finding error might be	caused by vSAN: Basic (unicast) conr	nectivity check					
Storage	🚳 Overview	Only failed pings Info							
Storage Overview Security	∽ Network	From Host	To Host	To Device vmkO	Ping result				
vSphere Cluster Services V	vSAN cluster partition	. 10.159.21.29	. 10.156.144.153	vmk0	0				
Health vSAN V	vSAN: MTU check (ping with larg	. 10.159.21.31	10.156.144.153	vmkO	0				
Skyline Health Virtual Objects	+ 8 healthy findings	10.159.21.32	10.156.144.153	vmkO	0				
Resyncing Objects Proactive Tests	 ✓ Data ● vSAN object health 	10.156.144.153	. 10.159.21.29	vmk1	0				
Capacity Performance	+ 1 healthy findings	. 10.156.144.153	. 10.159.21.31	vmk1	0				
Performance Diagnostics	✓ Performance service	10.156.144.153	10.159.21.32	vmk1	0				
Support Data Migration Pre-check	Stats DB object								
Cloud Native Storage V Container Volumes	> Online health				8 items				
Recent Tasks Alarms									

One final place to examine is virtual machines. Navigate to **[vSAN cluster] > Monitor > vSAN > Virtual Objects > View Placement Details**. It should show the witness absent from the secondary site perspective. However, virtual machines should still be running and fully accessible.

\equiv vSphere Client $ ext{Q}$ Sea			C Administr	ator@VSPHERE.LOCAL	~ © @~
	VM 1 Primary	D 🖸 🛃 🔯 🕴 🗄 ACTIONS	tuerles Cranshetz	Ludates	
 Datacenter OSA-Stretched 10.159.21.25 10.159.21.26 	Issues and Alarms V All Issues Triggered Alarms	Physical disk placement Group components by host placement Virtual Object Components	tworks snapsnots	Opdates	
 10.159.21.27 10.159.21.28 10.159.21.29 	Performance V Overview Advanced	Type	Component State	Host	Fault Domain
 10.159.21.30 10.159.21.31 	Tasks and Events V	> RAID 5			
 10.159.21.32 WM 1 Primary VM 1 Secondary 	Events Utilization	✓ RAID 5 Component	🔮 Active	. 10.159.21.32	Secondary
VM 2 Primary	vSAN V	Component	Active	10.159.21.29	Secondary
Two 2 Secondary	Performance	Component	Active	10.159.21.28	Secondary
 WM 3 Secondary VM 4 Primary VM 4 Secondary 	i/O Trip Analyzer	Component Witness	 Active Absent 	 10.159.21.31 10.156.144.153 	Secondary
> []] vSAN-ESA > [] 10.156.144.153			27 vSAN c	omponents on 9 hosts <	< 1/2 > >
Recent Tasks Alarms					

Returning to the health check, selecting **Data** > **vSAN object health**, you can see the error 'Reduced availability with no rebuild - delay timer'

SA-Stretched	ACTIONS		
ummary Monitor Conf	figure Permissions Hosts VMs Data	stores Networks Updates	
Memory	Skyline Health	vSAN object health	SILENCE ALERT
Persistent Memory	Last checked: 11/22/2022, 10:52:40 AM RETEST	This health finding error might be caused by vSAN cluster partition	
Storage	View Health History	Ouenieus lofe	
Utilization	🙆 Overview		
Storage Overview		REPAIR OBJECTS IMMEDIATELY PURGE INACCESSIBLE VM SWAP OBJECT	TS
Security	✓ Network	Health/Objects	Object count
vspriere cluster services ~	() vSAN cluster partition	(Reduced availability with no rebuild - delay timer) VIEW DETAILS	5 31
Health	• vSAN: Basic (unicast) connectivity		
	• vSAN: MTU check (ping with large		
Skyline Health Virtual Objects	+ 8 healthy findings		
Resyncing Objects	+ o nearry maings		
Proactive Tests	∨ Data		
Capacity	vSAN object health		
Performance	+ 1 healthy findings		
Performance Diagnostics	✓ Performance service		
Support			1 item

Conclusion

The loss of the witness does not impact the running virtual machines on the secondary site. There is still a quorum of components available per object, available from the data sites. Since there is only a single witness host/site, and only three fault domains, there is no rebuilding/resyncing of objects.

Repair the Failure

Add the gateway IP address back to the VMkernel adapter on the affected hosts, for example:

esxcli network ip interface ipv4 set -i vmk1 -t static -g 172.21.0.1 -I 172.21.4.12 -N 255.255.240.0

Trigger the Event

Similar to the last test, we can remove the gateway configured on the vSAN VMkernel adapter.

Here we open an SSH session to every host (on both sites). Again, confirm which VMkernel adapter is being used for vSAN:

```
[root@localhost:~] esxcli vsan network list | grep VmkNic
    VmkNic Name: vmk1
```

Then get the details for this adapter, and make a note of the output:

[root	@localhost:~]	esxcli network	ip interface ipv	4 get -i vmkl		
Name	IPv4 Address	IPv4 Netmask	IPv4 Broadcast	Address Type	Gateway	DHCP DNS
vmk1	172.21.4.12	255.255.240.0	172.21.15.255	STATIC	172.21.0.1	false

We can then effectively remove the gateway set for this adapter by re-applying the same IP address and omitting the gateway address. For example:

esxcli network ip interface ipv4 set -i vmk1 -t static -I 172.21.4.12 -N 255.255.240.0

Optionally, we can achieve this using a quick script (remember to change the values of the VMkernel adaptor and netmask):


```
ip=$(esxcli network ip interface ipv4 get -i vmk1 | grep vmk1 | awk '{print $2}')
esxcli network ip interface ipv4 set -i vmk1 -t static -I $ip -N 255.255.240.0
```

For a more detailed script, see:

https://github.com/vmware-tanzu-experiments/vsphere-with-tanzu-proof-of-concept-samples/blob/main/VCF/vSAN-net-gw-operations.md

Cluster Behavior on Failure

The events observed are for the most part identical to those observed in the previous test.

\equiv vSphere Client $~Q~$ Search in all environments	C & Administrator@VSPHERE.LOCAL > (2) (?) >
VSphere Client Q Search in all environments VSphere Client Configure Permissions Hosts VMs D CPU Memory Persistent Memory Storage Utilization Storage Overview Security VSphere Cluster Services Health Virtual Objects Resyncing Objects Proactive Tests Capacity VSAN object health Virtual Objects Resyncing Objects Proactive Tests Capacity	C Administrator@VSPHERELOCAL V • <td< th=""></td<>
Capacity Performance Performance Diagnostics Recent Tasks Alarms	Stats DB object

And again, we have a vSAN network partition, with the data sites and witness appliance separated:

\equiv vSphere Client $ ext{Q}$ Sea			C	Administrator@VSPHERE.LOCAL ~	©
Summary Monitor Config	ACTIONS	ores Networks Updates			
Configuration Issues Datastores under APD or PDL Resource Allocation V CPU	Skyline Health Last checked: 11/22/2022, 527:35 PM RETEST View Health History	VSAN cluster partition	be caused by vSAN: Basic (ur	iicast) connectivity check	SILENCE ALERT
Memory Persistent Memory		Partition list Info	be caused by vSAN: MTU che	ck (ping with large packet size)	
Storage Utilization Storage Overview Security	vSAN: Basic (unicast) connectivity	Host	Partition	Host UUID 6377a85d-c870-5076-6bf0-00505699ba	a71
vSphere Cluster Services ∨ Health	 VSAN: MTU check (ping with large + 8 healthy findings 	. 10.159.21.26 . 10.159.21.25	2	63625066-492d-38ba-2182-a4bf012fb018 636251ee-5434-2a1c-69ce-a4bf012fa94c	3
Skyline Health Virtual Objects Resurcing Objects	V Data VSAN object health + 1 healthy findings	. 10.159.21.27 . 10.159.21.28	2	63736c86-f433-edb7-6f5d-a4bf0130462	d
Proactive Tests Capacity Performance	Verformance service Stats DB object	 10.159.21.29 10.159.21.30 	2	6374ece5-ff5f-d737-336e-a4bf012fab77 6374eda2-9d31-527a-c783-a4bf012fa9d8	
Performance Diagnostics Support Data Migration Pre-check	+ 4 healthy findings > Online health	 10.159.21.31 10.159.21.32 	2	6374ee11-93ed-76eb-b586-a4bf012fae97 6374ee4f-f7a8-c13c-819a-a4bf012fab95	·
Cloud Native Storage Container Volumes	> Physical disk > Cluster				9 items
Recent Tasks Alarms					

Conclusion

When the vSAN network fails between the witness site and both the data sites (as in the witness site fully losing its WAN access), it does not impact the running virtual machines. There is still a quorum of components available per object, available from the data sites. However, as explained previously, since there is only a single witness host/site, and only three fault domains, there is no rebuilding/resyncing of objects.

Repair the Failure

Just as before, add the gateway IP address back to the VMkernel adapter on all the hosts. For example:

esxcli network ip interface ipv4 set -i vmk1 -t static -g 172.21.0.1 -I 172.21.4.12 -N 255.255.240.0

Scenario #3: Secondary Site Failure

Trigger the Event

Power off the hosts in the secondary site (either physically or via the host management console).

Cluster Behavior on Failure

In the first instance, hosts will show communication failures, such as:

We can see the hosts are shown as 'not responding' and the 'unexpected number of fault domains' indicates that we have lost a site.

📃 vSphere Client	vSphere Client Q Search in all environments C 2 Administrator@VSPHERE.LOCAL ~ 😳							?~	
) 🕼 OSA-Stretch	ed	ACTIONS							
Summary Monitor	Config	ure Permissions Hosts VMs Datastores Networks	Update	lS					
Issues and Alarms	~	All Issues							
All Issues		PREVIOUS NEXT ACKNOWLEDGE RESET TO GREEN							
I riggered Alarms		Issue	т	Туре	т	Trigger Time	т	Status	Ŧ
Periorinance	Ť	VSAN performance service alarm 'Stats DB object'		Triggered Alarm		11/23/2022, 04:07 PM		() Alert	
Overview		VSAN stretched cluster alarm 'Unexpected number of fault domains'		Triggered Alarm		11/23/2022, 04:07 PM		() Alert	
Advanced		VSAN data alarm 'vSAN object'		Triggered Alarm		11/23/2022, 04:07 PM		() Alert	
Tasks and Events	~	VSAN network alarm 'Hosts disconnected from VC'		Triggered Alarm		11/23/2022, 04:07 PM		() Alert	
Tasks		10.159.21.32: vSphere HA host status		Triggered Alarm		11/23/2022, 04:06 PM		() Alert	
Events		10.159.21.31: vSphere HA host status		Triggered Alarm		11/23/2022, 04:06 PM		() Alert	
vSphere DRS	~	10.159.21.29: vSphere HA host status		Triggered Alarm		11/23/2022, 04:06 PM		() Alert	
Recommendations		10.159.21.28: vSphere HA host status		Triggered Alarm		11/23/2022, 04:06 PM		() Alert	
Faults		10.159.21.31: Host connection and power state		Triggered Alarm		11/23/2022, 04:05 PM		() Alert	
History		10.159.21.29: Host connection and power state		Triggered Alarm		11/23/2022, 04:05 PM		() Alert	
VM DRS Score		10.159.21.32: Host connection and power state		Triggered Alarm		11/23/2022, 04:05 PM		Alert	
CPU Utilization		10.159.21.28: Host connection and power state		Triggered Alarm		11/23/2022, 04:05 PM		Alert	
Memory Utilization									
Network Utilization									12 items
Recent Tasks Al	larms								

Clicking on 'view all issues' shows more information and indicates that we have a HA event:

		C Administrator@VSPHERE.LOCAL > (2) (?
 	OSA-Stretched : ACTIONS Summary Monitor Configure Permissions Issues and Alarms VSAN performance service alarm 'Stats DB object' VSAN stretched cluster alarm 'Unexpected number of fault c vSAN data alarm 'VSAN object' VIEW ALL ISSUES (12)	Hosts VMs Datastores Networks Updates Actions domains' Actions Actions
 In the second and the	Cluster Details Total Processors: 160 Total vMotion 5 Migrations: Prefere Pault Domains: Prefere	Image: Secondary Image: Secondary <td< th=""></td<>

Thus, all VMs from the secondary data site will be restarted via HA on the Preferred data site. As we only have one data site (fault domain) left, the storage policy will be non-compliant. If we look at the VMs on the cluster, we can see that this is the case. Below we have added the columns for 'host', 'uptime' and 'VM storage policy compliance' to illustrate this:

\equiv vSphere Client $ ext{Q}$ Search in all en					C 2 · © ? ·
> CSA-Stretched : ACTIONS					
Summary Monitor Configure Per	missions Hos	ts VMs	Datastores Netwo	orks Update	25
Virtual Machines VM Templates vA	pps				
					▼ second
Name	State	Status	Host	Uptime 个	VM Storage Policies Compliance
🗌 📱 🔠 VM 1 Secondary	Powered On	🗸 Normal	10.159.21.30	7 minutes	1 Noncompliant (Nov 23, 2022 4:14 PM)
🗌 📱 🔂 VM 2 Secondary	Powered On	V Normal	10.159.21.27	7 minutes	Interpretent (Nov 23, 2022 4:14 PM)
🗌 🗄 🔂 VM 5 Secondary	Powered On	✓ Normal	10.159.21.30	7 minutes	Interpretended in the image of the image
🗌 📱 🔂 VM 6 Secondary	Powered On	V Normal	10.159.21.27	7 minutes	① Noncompliant (Nov 23, 2022 4:14 PM)
🗌 📱 🔂 VM 4 Secondary	Powered On	V Normal	10.159.21.25	8 minutes	① Noncompliant (Nov 23, 2022 4:14 PM)
🗌 🕴 🔂 VM 7 Secondary	Powered On	V Normal	10.159.21.25	8 minutes	① Noncompliant (Nov 23, 2022 4:14 PM)
🗌 📱 健 VM 3 Secondary	Powered On	V Normal	10.159.21.26	8 minutes	🕕 Noncompliant (Nov 23, 2022 4:14 PM)
🔲 🗏 🔠 VM 8 Secondary	Powered On	✓ Normal	10.159.21.26	8 minutes	Noncompliant (Nov 23, 2022 4:14 PM)
EXPORT ~					ltems per page 35 \lor 8 items
Recent Tasks Alarms					

Verify on each host or via [vSAN cluster] -> VMs if all VMs were restarted on the preferred site.

Conclusion

When the secondary site fails, it does not impact the running VMs on the primary data site as quorum exists. VMs on the failed site will be restarted via HA on the available data site. Again, since there are only three fault domains, there is no rebuilding or resyncing of objects (and thus the storage policies are shown as non-compliant).

Repair the Failure

Power on the hosts on the secondary site. If DRS was set to 'fully automated' then VMs will be moved back to the secondary site. Otherwise, they will need to be moved back to the correct site.

Scenario #4: Primary Site Failure

Trigger the Event

Power off the hosts in the secondary site (either physically or via the host management console).

Cluster Behavior on Failure

As before, the first error is reported back from the remaining powered on hosts (i.e. on the secondary site)

Looking at the cluster issues, we see the host power state and subsequent HA event:

📃 vSphere Client	Q Search in all environments	C & Administrator@VSPHERE.LOCAL ~ 😧 ?
CSA-Stretche	d EACTIONS	
Summary Monitor	Configure Permissions Hosts VMs Datastores Networks Update	s
Issues and Alarms	 All Issues 	
All Issues Triggered Alarms	PREVIOUS NEXT ACKNOWLEDGE RESET TO GREEN	
Performance	V Issue T	Type Y Trigger Time Y Status Y
0	10.159.21.30: vSphere HA host status	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
Overview	10.159.21.27: vSphere HA host status	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
Advanced	10.159.21.26: vSphere HA host status	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
Tasks and Events	V 10.159.21.25: vSphere HA host status	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
Tasks	VSAN performance service alarm 'Stats DB object'	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
Events	VSAN stretched cluster alarm 'Preferred fault domain unset'	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
vSphere DRS	VSAN stretched cluster alarm 'Unexpected number of fault domains'	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
Recommendations	VSAN data alarm 'vSAN object'	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
Faults	VSAN network alarm 'Hosts disconnected from VC'	Triggered Alarm 11/24/2022, 03:25 PM ① Alert
History	10.159.21.30: Host connection and power state	Triggered Alarm 11/24/2022, 03:24 PM ① Alert
VM DRS Score	10.159.21.25: Host connection and power state	Triggered Alarm 11/24/2022, 03:24 PM ① Alert
CPU Utilization	10.159.21.27: Host connection and power state	Triggered Alarm 11/24/2022, 03:24 PM ① Alert
Memory Utilization	10.159.21.26: Host connection and power state	Triggered Alarm 11/24/2022, 03:24 PM ① Alert
Network Utilization		13 items
Recent Tasks Ala	rms	

And, as per the previous test, the VMs on the affected site are restarted to run on the active site:

\equiv vSphere Client $ ext{ Q}$ Search in all environments			C 🛆 Admi	nistrator@VSPHERE.LOCAL ~	?~
Control Con	Image: OSA-Stretched : ACTIONS Summary Monitor Configure Permission Virtual Machines VM Templates vApps	ons Hosts VMs D	atastores Networks	Updates	
 GSA-Stretched 10.159.21.25 (Not responding) 10.159.21.27 (Not responding) 10.159.21.27 (Not responding) 10.159.21.28 10.159.21.30 10.159.21.30 (Not responding) 10.159.21.31 10.159.21.31 10.159.21.32 VM 1 Primary VM 1 Secondary VM 2 Primary 	Name Stat II III VM 1 Primary Pori IIII IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	Status vered On < Normal	Host Uptim 10 1059.21.32 52 st 10.159.21.32 52 st 10.159.21.32 52 st 10 10.159.21.38 52 st 10 10.159.21.38 52 st 10 10.159.21.31 69 st 10 10.159.21.31 69 st 10 10.159.21.31 69 st	Y pri e + VM Storage Policies Compliance econds ① Noncompliant (Nov 24, 2022 3 econds ① Noncompliant (Nov 24, 2022 3	225 PM) 225 PM) 225 PM) 225 PM) 225 PM) 225 PM) 225 PM) 225 PM)
ⓓ VM 2 Secondary	EXPORT -			litems per page $35 \vee$	8 items

The storage policy is shown as non-compliant as we do not have enough fault domains to fulfil the policy requirements. Looking at the objects in the cluster, we can see the status as 'reduced availability with no rebuild' – as there is no other fault domain to rebuild to.

📃 vSphere Client	Q Sea	rch in all env	ironments			C	O Administrator@VSPH	IERE.LOCAL ~) ?~		
) 🕼 OSA-Stretche	ed	ACTIONS									
Summary Monitor	Config	ure Perm	nissions Hosts VMs	Datastores Netwo	orks Update	s					
Issues and Alarms	~	Virtual Of	ojects								
All Issues Triggered Alarms		A There are connectivity issues in this cluster. One or more hosts are unable to communicate with the vSAN datastore. Data below does not reflect the real state of the system.									
Performance	~	Browse all	virtual objects and check the	r state in real time and vi	iew their placer	ent across the physical infra	astructure. Get information	about each obiect sta	te and		
Overview		common re	mediation. About vSAN Obje	ect Health 🖸				,,,,,,,			
Advanced		Reduced a	vailability with no rebuild - delay tir	ner 66)							
Tasks and Events	~	VIEW PLACE	MENT DETAILS VIEW PERFOR	MANCE CLEAR FILTERS							
Tasks											
Events			Name	т Туре	т	Object State	Storage	Policy	т		
vSphere DRS	~	> 🗆	VM 1 Primary	VM		Reduced availability with no re	ebuild - delay t 📄 🔜 Str	etched Cluster R1 withl R	85 local		
Recommendations			🛱 VM 1 Secondary	VM		Reduced availability with no re	ebuild - delay t) 🖙 Str	etched Cluster R1 withl R	85 local		
Faults			E un en l								
History		> U	VM 2 Primary	VM		Reduced availability with no re	ebuild - delay t	etched Cluster R1 withi R	to local		
CRUUItilization		> 🗆	VM 2 Secondary	VM		Reduced availability with no re	ebuild - delay t 🛛 🔒 Str	etched Cluster R1 withl R	85 local		
Memory Utilization		> 🗆	🗇 VM 3 Primary	VM		Reduced availability with no re	ebuild - delay t 🔒 Str	etched Cluster R1 withI R	85 local		
Network Utilization		> 🗆	🗇 VM 3 Secondary	VM		Reduced availability with no re	ebuild - delay t 🖳 Str	etched Cluster R1 withI R	85 local		
vSphere HA	~		🗇 VM 4 Primary	VM		Reduced availability with no re	ebuild - delay t) 🗖 Str	etched Cluster R1 withI R	85 local		
Summary											
Heartbeat									19 items		
~ ~ ~ ~ ~											
Recent Tasks Al	arms										

Conclusion

As per the previous test, after losing a data site, HA restarts the VMs on the remaining, active site. Again, there is no rebuild of components as only one data site is available.

Repair the Failure

Power on the hosts on the primary site. If DRS was set to 'fully automated' then VMs will be moved back to the secondary site. Otherwise, they will need to be moved back to the correct site.

Scenario #5: Network Failure between Data Sites but Witness Still Accessible

Trigger the Event

Link failure between preferred and secondary data sites simulates a datacenter link failure while witness traffic remains up and running.

For this test, we can either physically disable the network link between the data sites or use the DVS traffic filter function.

To use the DVS traffic filter, navigate to the port group used for the preferred site, then **Configure > Traffic filtering and marking**, as shown below:

 ✓ VSAN-Preferred-vlan3921 : ACTIONS ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑ ☑	vSphere Client Q Search in all environ		C Administrator@VSPHERE.LOCAL ~ 🙄 ? ~
Image: Solution DVUplinks-30 No. Rule Name Action Traffic Direction Traffic Qualifiers Image: DSwitch-Management Network Alarm Definitions Image: DSwitch-VM Network-1003 Image: DSwitch-VM Network-1004 Image: DSwitch-VM Network-1004	Control Con	VSAN-Preferred-vlan3921 Summary Monitor Configure Permissions Settings Properties Policies Traffic filtering and marking Alarm Definitions No.	TIONS Ports Hosts VMs ng and marking is ENABLE AND REORDER DELETE Rule Name Action Traffic Direction Traffic Qualifiers Voltems found No items found

Here, we create filter rule for **each** host on the preferred site to **all** hosts on the secondary site, selecting the action as 'drop' with

'any' protocol between the IP addresses (thus, for four hosts, we need 16 rules in total):

Settings				
Name	25 <> 29 drop			
Action	Drop ~			
Traffic direction	Ingress/Egress V	(j)		
Traffic Qualifie	rs			
IP MAC S	ystem traffic			
Enable qualifier				
Protocol number	any ~			
Source and Des	tination Addresses			
Source IP address	is v	10.159.21.25		
Destination IP address	is ~	10.159.21.29		
			CANCEL	

We can then enable the newly created DVS filters:

Summary Monitor	red-v _{Confi}	rlan39)21 Per	MACTIONS missions Ports	Hosts VMs			
Settings Properties Policies	~	Traf	fic	filtering and r	narking is Turned C 	DFF	ENABL	E AND REORDER
Traffic filtering and ma	arking			No.	Rule Name	Action	Traffic Direction	Traffic Qualifiers
Alarm Definitions		0	~	1 IP, Protocol numbe	25 <> 29 drop r: any, Source: 10.159.21.25, D	Drop Destination: 10.159.21.29	Ingress/Egress	(q)
		0	~	2 IP, Protocol number	Network Traffic Rule 1 r: any, Source: 10.159.21.25, D	Drop vestination: 10.159.21.28	Ingress/Egress	qI
		0	~	3 IP, Protocol numbe	Network Traffic Rule 2 r: any, Source: 10.159.21.25, D	Drop vestination: 10.159.21.31	Ingress/Egress	q
		•	~	4	Network Traffic Rule 3	Drop	Ingress/Egress	(IP) 16.it

Cluster Behavior on Failure

As expected, we observe network errors and a cluster partition:

The VMs are restarted by HA on the preferred site:

\equiv vSphere Client Q Search	in all environments					ninistrator@VSPH	iere.local ~	© ?~
<	I OSA-Stretched	ACTIONS						
	Summary Monitor Configur	e Permissions	Hosts	VMs Datas	stores Networks	Updates		
VC-dxb-8.vsanpe.vmware.com	Virtual Machines VM Templat	tes vApps						
							T VM	
10.159.21.25								
10.159.21.26	Name ↑	State	Status	Fault Domain	Provisioned Space	Used Space	Host CPU	Host Mem
10.159.21.27	VM 1 Primary	Powered On	 Normal 	Preferred	244.53 GB	1.34 GB	0 Hz	77 MB
10.159.21.28	VM 1 Secondary	Powered On	 Normal 	Preferred	244.11 GB	844 MB	0 Hz	74 MB
10.159.21.29	VM 2 Primary	Powered On	 Normal 	Preferred	244.49 GB	1.3 GB	0 Hz	77 MB
10.159.21.30	VM 2 Secondary	Powered On	 Normal 	Preferred	244.06 GB	800 MB	0 Hz	74 MB
10.159.21.31	VM 3 Primary	Powered On	 Normal 	Preferred	244.53 GB	1.34 GB	0 Hz	77 MB
10.159.21.32	🗌 📱 🔡 VM 3 Secondary	Powered On	 Normal 	Preferred	244.09 GB	828 MB	0 Hz	74 MB
🔂 VM 1 Primary	🗌 📱 🔂 VM 4 Primary	Powered On	 Normal 	Preferred	244.47 GB	1.28 GB	0 Hz	77 MB
VM 1 Secondary	VM 4 Secondary	Powered On	V Normal	Preferred	244.09 GB	828 MB	0 Hz	74 MB
VM 2 Primary	🗌 🗄 🕼 VM 5 Primary	Powered On	 Normal 	Preferred	184.46 GB	544 MB	0 Hz	77 MB
VM 2 Secondary	🔲 🗄 🔁 VM 5 Secondary	Powered On	 Normal 	Preferred	184.34 GB	400 MB	0 Hz	74 MB
VM 3 Primary	🗌 🗄 🕼 VM 6 Primary	Powered On	 Normal 	Preferred	184.43 GB	516 MB	0 Hz	35 MB
VM 3 Secondary	🔲 🗄 🔡 VM 6 Secondary	Powered On	V Normal	Preferred	184.34 GB	392 MB	0 Hz	34 MB
VM 4 Primary	🔲 🗄 🔂 VM 7 Primary	Powered On	 Normal 	Preferred	184.44 GB	520 MB	0 Hz	35 MB
WM 5 Primany	🗌 🗄 🔡 VM 7 Secondary	Powered On	V Normal	Preferred	184.3 GB	356 MB	0 Hz	34 MB
WM 5 Primary	🔲 📱 🔂 VM 8 Primary	Powered On	V Normal	Preferred	184.41 GB	496 MB	0 Hz	35 MB
WM 6 Primary	🗌 📱 🔡 VM 8 Secondary	Powered On	 Normal 	Preferred	184.34 GB	396 MB	0 Hz	34 MB
WM 6 Secondary								
WM 7 Primary	EXPORT -						Items per page	35 🗸 16 items
Recent Lasks Alarms								

Conclusion

In the failure scenario, if the data link between data centers is disrupted, HA will start the VMs on the preferred site. There is still a quorum of components available per object, available from the data sites. As previously discussed, there is only one data site remaining, therefore there is no rebuilding/resyncing of objects.

Repair the Failure

Disable the DVS filter rules and rerun the health check tests. Verify that all tests pass successfully and that there are no remaining errors on the cluster, hosts or VMs.

