vSAN Stretched Clusters for Azure VMware Solution

VMware General

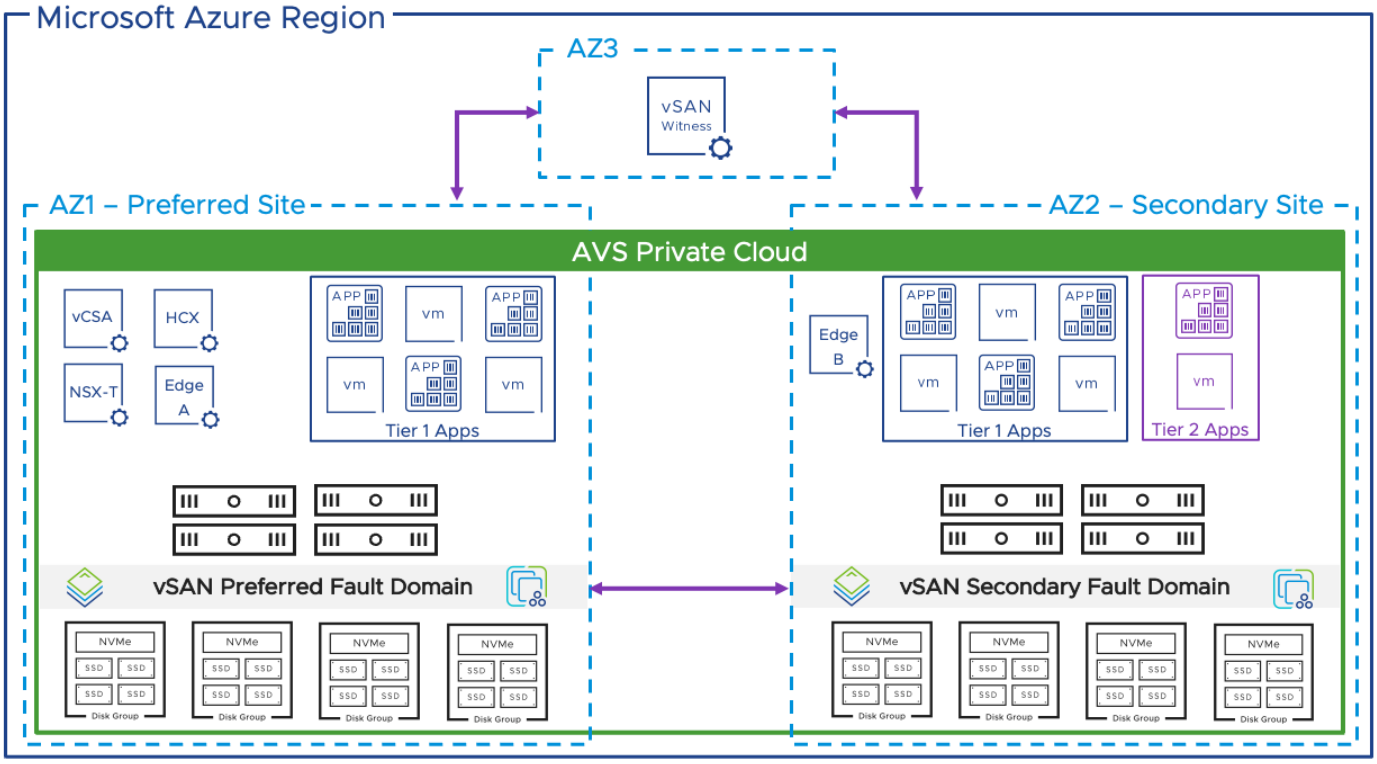
Table of contents

vSAN Stretched Clusters for Azure VMware Solution

Overview

The purpose of this document is to provide an overview and technical details specifically for vSAN Stretched Clusters for Azure VMware Solution (AVS). More generic details about vSAN on AVS covering local storage policies, Run commands, monitoring, responsibilities, and more can be found in our vSAN technical guide.

A stretched cluster is configured by deploying an AVS private cloud across two availability zones (AZs) within the same region and placing a vSAN witness in a third AZ in that same region. The vSAN witness continuously monitors all hosts within the cluster to serve as a quorum in case of split-brain scenarios, and is fully managed by Microsoft. The hosts within the private cloud are evenly deployed across regions and operate as a single entity, and storage policy-based synchronous replication ensures data is replicated across AZs delivering a recovery point objective (RPO) of zero. This means that if one AZ goes down due to a natural disaster or some other unexpected event, the other AZ can continue to operate and provide uninterrupted access to critical workloads.

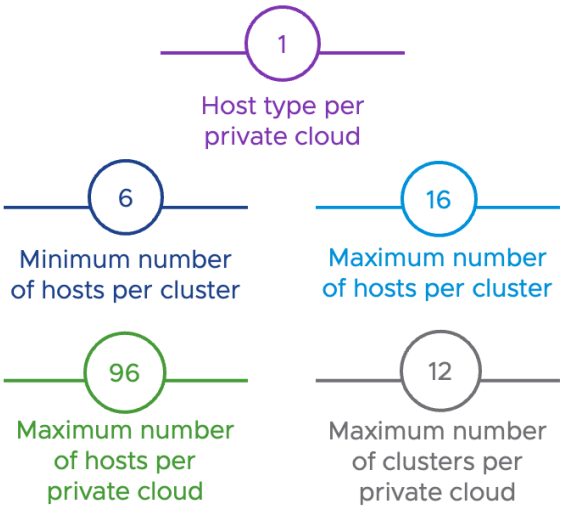


It should be noted that while stretched clusters offer additional protection and resiliency, they don't address all failure scenarios such as regional failures or certain inter-AZ failures. Designing application workloads for redundancy and having a proper DR plan in place is always highly recommended.

vSAN Clusters

Hosts and Capacity

AVS offers three different host types; in addition to having different CPU and RAM specs, they also have different storage footprints. vSAN datastore capacity is increased by adding additional hosts to the cluster. Currently, stretched clusters are only available for use with AV36 or AV36P depending on the region they're deployed in. Hosts must be deployed evenly across AZs within the region, with a minimum of 6 (3 per AZ) and a maximum of 16 (8 per AZ).



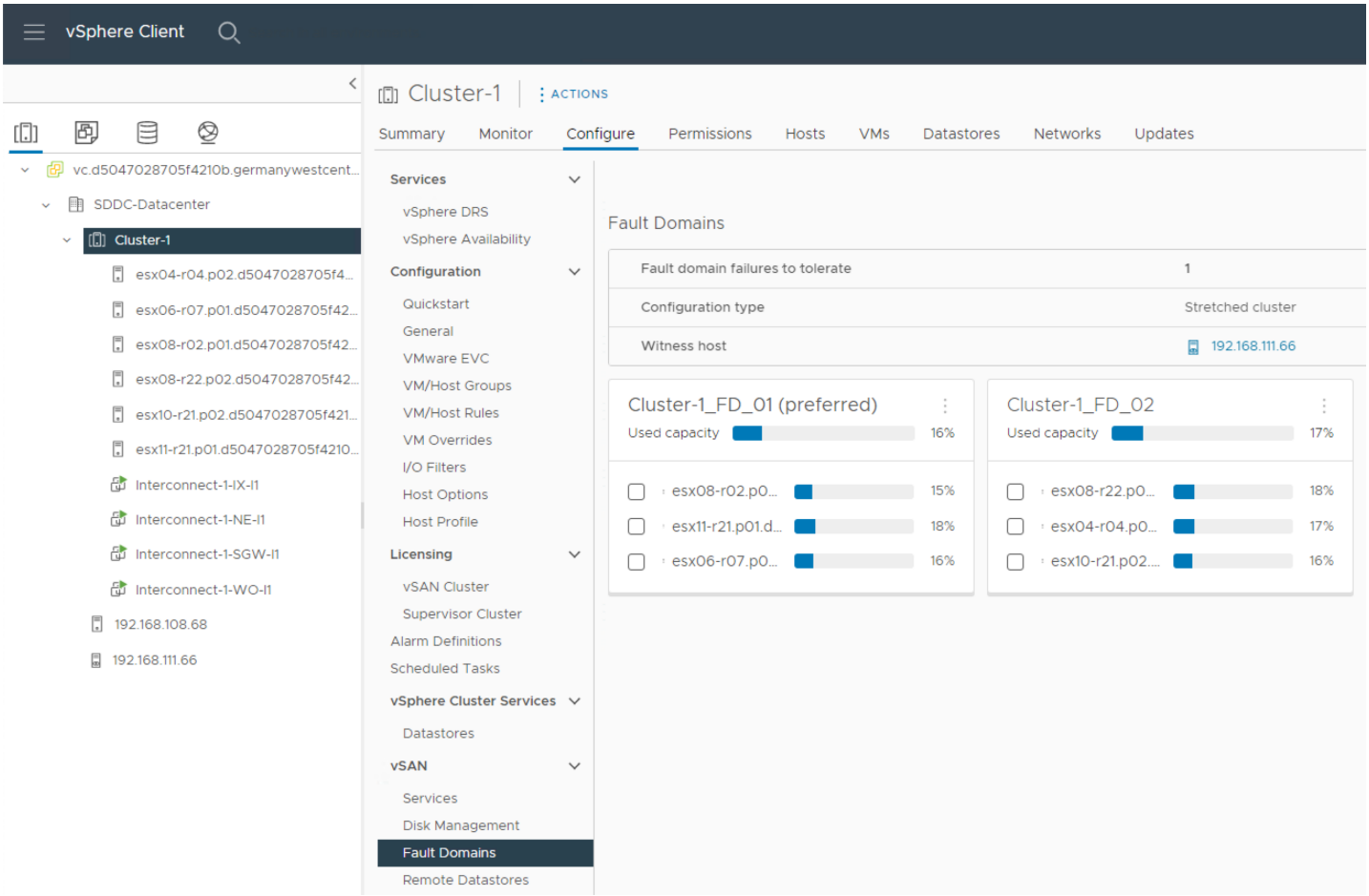
		AV36		AV36P	
Hosts	CPU	RAM	Storage	RAM	Storage
6	216	3.4 TB	92 TB	4.5 TB	115 TB
8	288	4.5 TB	123 TB	6 TB	153 TB
10	360	5.6 TB	154 TB	7.5 TB	192 TB
12	432	6.75 TB	185 TB	9 TB	230 TB
14	504	7.9 TB	215 TB	10.5 TB	268 TB
16	576	9 TB	246 TB	12 TB	307 TB

- AV36 Availability: UK South, West Europe, Germany West Central
- AV36P Availability: Australia East

Note: All vSAN storage numbers are listed as total raw capacity. Regional availability for new host types limited during initial rollout, check with Microsoft for timing in your region.

The total amount of storage per host is shown as raw capacity, for example, a standard 3 host AV36 cluster provides 46.2 TB of raw capacity, a stretched cluster would technically have 92.4 TB of raw capacity. However, the total usable capacity will vary based on a number of variables such as RAID, Primary Failures to Tolerate (PFTT, or site disaster tolerance), and Secondary Failures to Tolerate (SFTT, or local FTT), compression, deduplication, thin vs thick provisioning, and slack space. With these variables in mind, it's likely you'll have far less available capacity than anticipated, and it's strongly recommended to ensure appropriate cluster sizing based on capacity needs.

All hosts are separated evenly into two fault domains labeled **Cluster-*n*_FD_01 and 02**. Each fault domain is associated with a specific AZ. You can also see which hosts reside in which AZ both in the vSphere Client, and in the Azure Portal by viewing or editing the cluster.



Stretched Cluster Limitations

- A new, separate, Azure subscription must be used when deploying a stretched cluster.
- Standard and stretched clusters cannot co-exist in the same private cloud.
- The type of cluster – standard or stretched – must be chosen at the time of deployment; a cluster cannot be converted after it has been deployed.
- Clusters must contain the same host type.
- Some GA features of standard clusters may not be available yet for stretched clusters.
- Scale in and scale out of a stretched cluster must be in pairs.

vSAN Configuration

Microsoft manages the vSAN cluster configuration, and customers do not have the ability to modify the configuration.

Enabled by default:

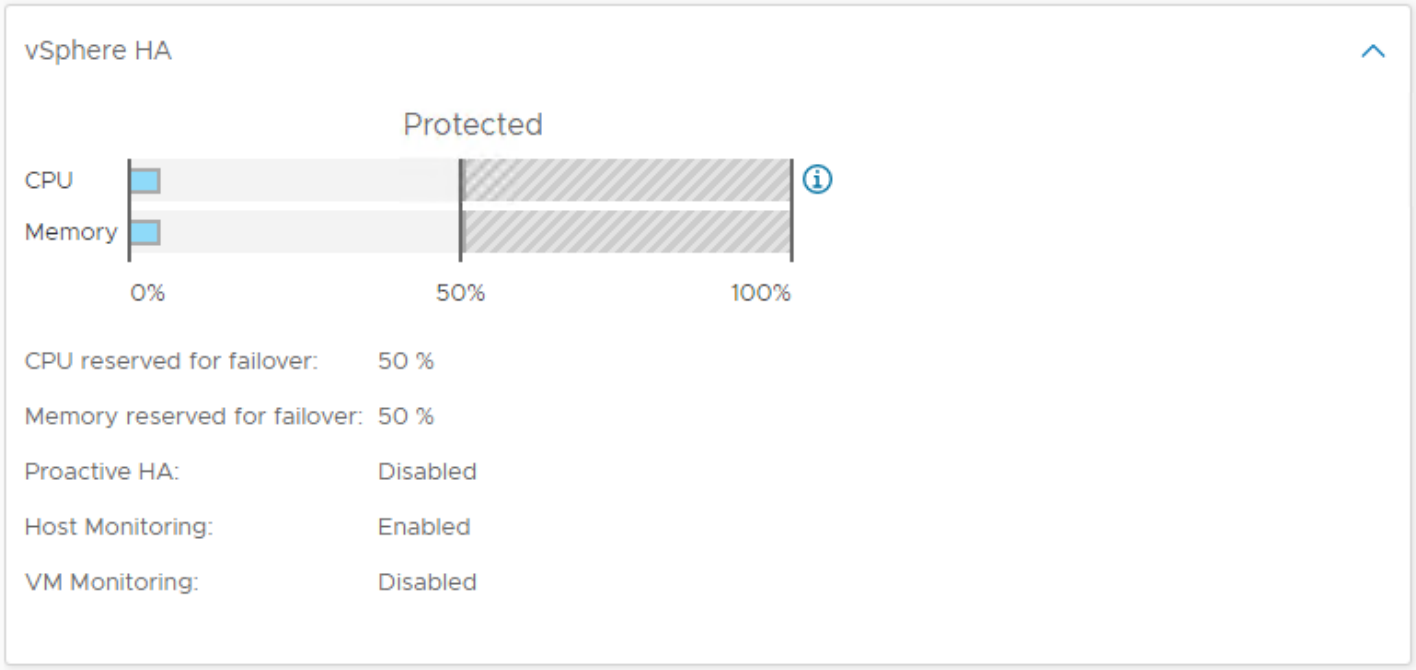
- Space efficiency services (deduplication and compression)
- Data-at-rest encryption
  - Provider managed, encryption keys are created at the time of deployment and stored in Azure Key Vault.
  - If a host if removed from a cluster, the data on the local devices is invalidated immediately.
  - Customers have the option to manage their own keys after deployment.

Disabled by default:

- Operations reserve
- Host rebuild reserve
- vSAN iSCSI Target Service
- File Service

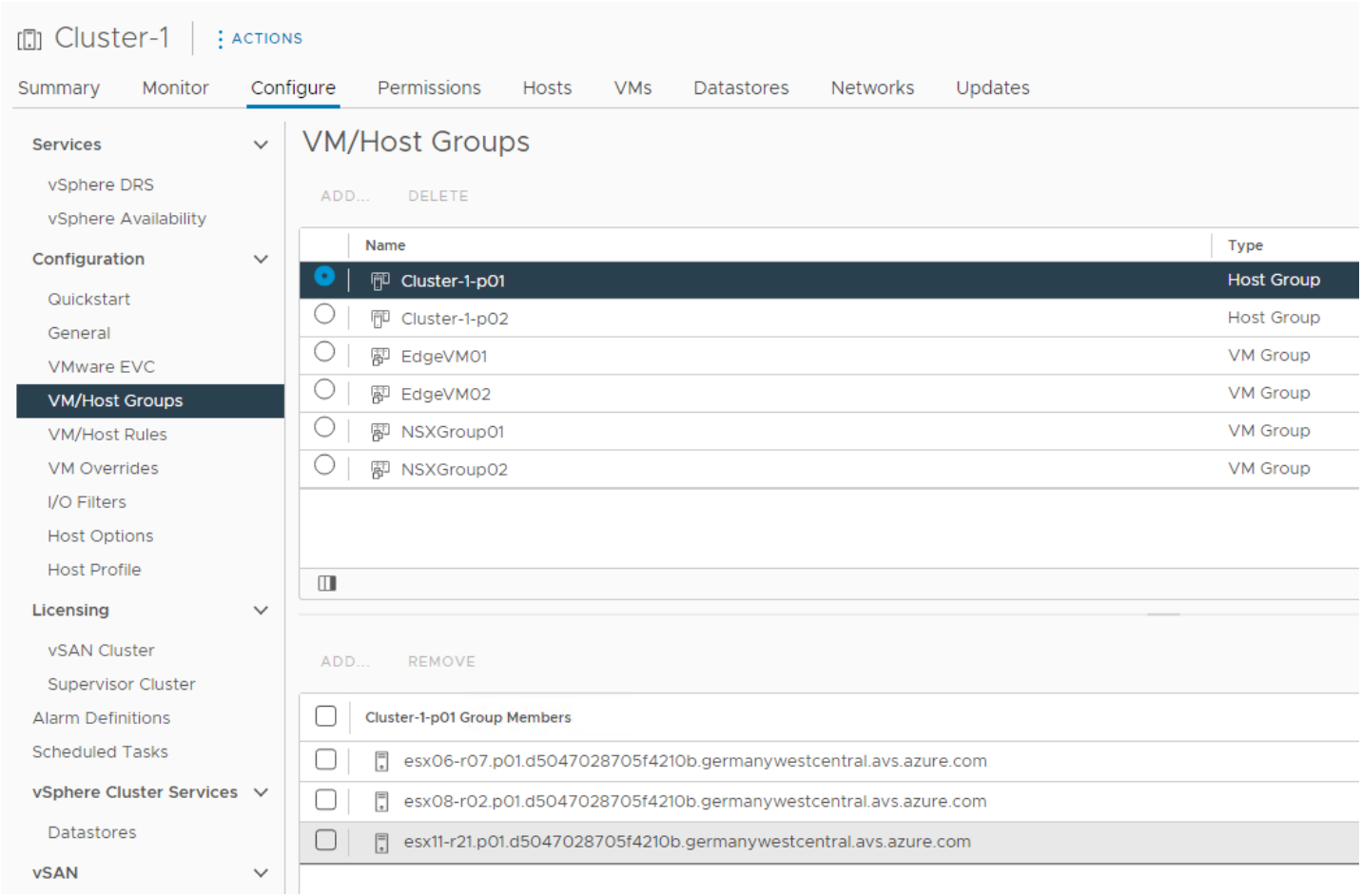
vSphere HA

vSphere HA is configured to restart VMs during host failure or isolation. Admission control failover capacity is configured by reserving 50% of the CPU and memory capacity of the entire stretched cluster to ensure that VMs can fail between AZs and have adequate resources. All management VMs (vCenter, NSX Managers, etc.) are given the highest restart priority in the event of a failure.



vSphere DRS

vSphere DRS is configured as fully automated, with a migration threshold set to 3 by default. Special DRS Host Groups are created for each fault domain, and an NSX Edge VM is placed in each group.


Cluster-1
ACTIONS
Summary
Monitor
Configure
Permissions
Hosts
VMs
Datastores
Networks
Updates
Services
vSphere DRS
vSphere Availability
Configuration
Quickstart
General
VMware EVC
VM/Host Groups
VM/Host Rules
VM Overrides
I/O Filters
Host Options
Host Profile
Licensing
vSAN Cluster
Supervisor Cluster
Alarm Definitions
Scheduled Tasks
vSphere Cluster Services
Datastores
vSAN
VM/Host Groups
ADD...
DELETE
Name
Type
Cluster-1-p01
Host Group
Cluster-1-p02
Host Group
EdgeVM01
VM Group
EdgeVM02
VM Group
NSXGroup01
VM Group
NSXGroup02
VM Group
ADD...
REMOVE
Cluster-1-p01 Group Members
esx06-r07.p01.d5047028705f4210b.germanywestcentral.avs.azure.com
esx08-r02.p01.d5047028705f4210b.germanywestcentral.avs.azure.com
esx11-r21.p01.d5047028705f4210b.germanywestcentral.avs.azure.com

Storage Policies

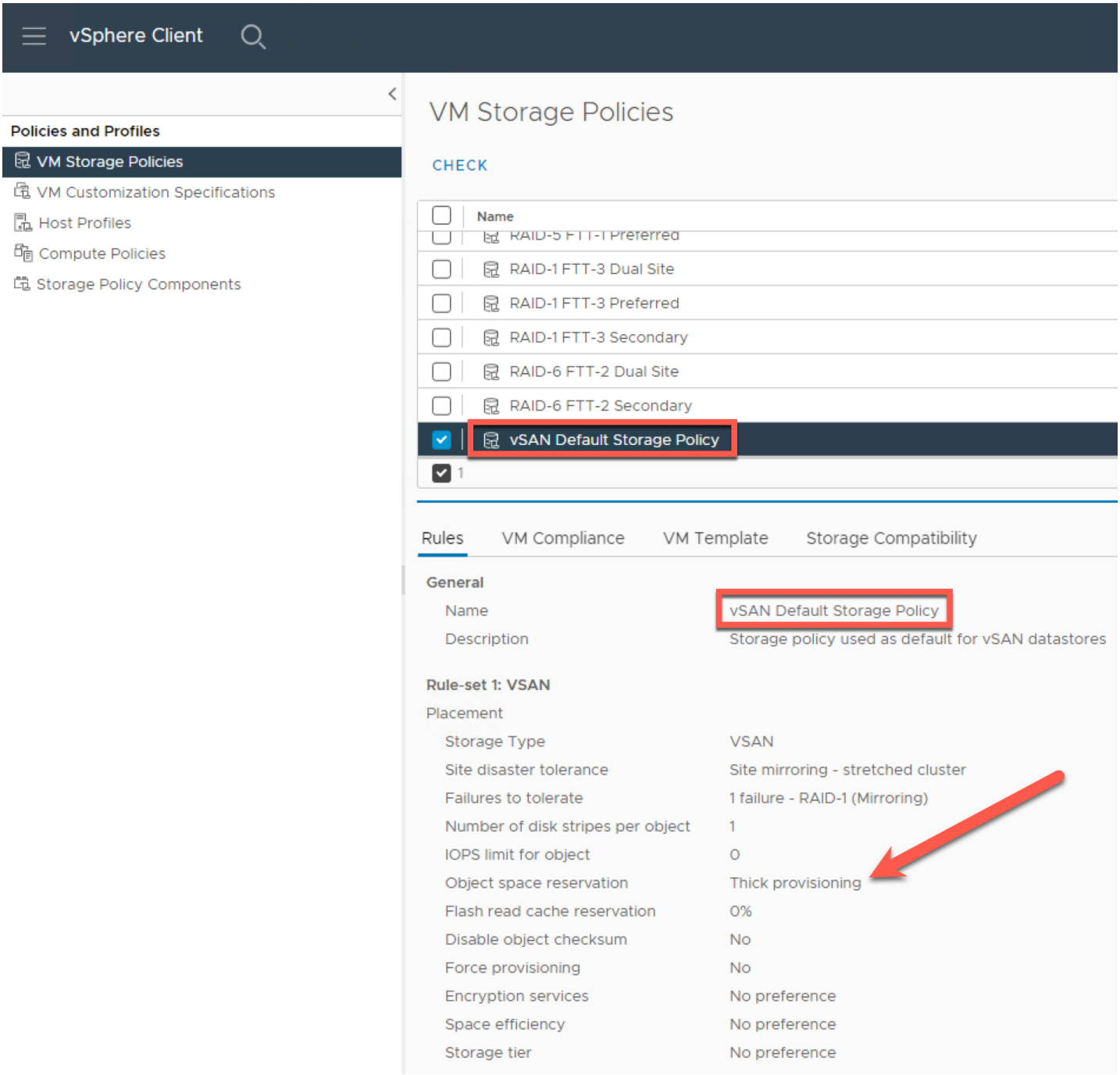
In addition to the standard cluster storage policies, 15 additional storage policies are available for Stretched Clusters. These policies are a combination of RAID-1,5, and 6 with PFTTs for Dual Site, Preferred, and Non-Preferred and various SFTTs.

Default Storage Policy

The default policy for all of the SDDC management VMs—including the vCenter Server Appliance, NSX-T Manager, and NSX-T Edges—is labeled **Microsoft vSAN Management Storage Policy**.

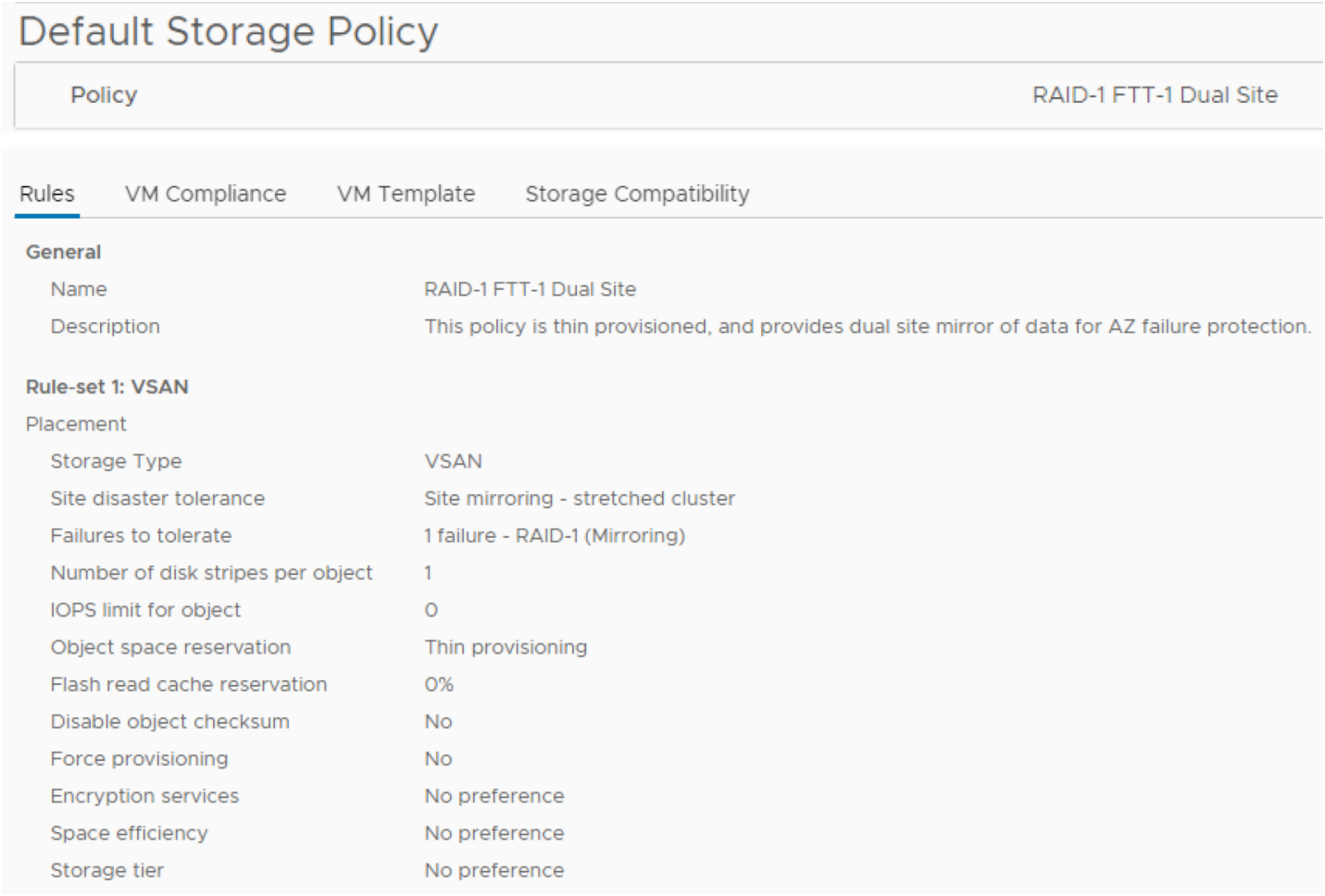
If you're familiar with vSAN, you know there is a **vSAN Default Storage Policy**. In AVS, this does exist, however it is **NOT** the default storage policy applied to the cluster. This policy exists for historical purposes only, and this is where confusion may set in. This is a vSAN Default Storage Policy … *policy*.

Upon closer inspection you'll notice that the Object space reservation is set to thick provisioning.



The actual default storage policy setting for stretched clusters is set to **RAID-1 FTT-1 Dual Site**, with Object space reservation

set to Thin provisioning.



The difference between the two is one is set to thick provisioning, and one is set to thin provisioning.

In a 6 host cluster, this policy enables RAID-1 mirroring and protects VMs not only against a single host failure, but a site failure as well. This policy does, however, require double the storage per virtual machine.

Configuring Storage Policies

As hosts are added to the cluster, it's recommended to change both the default storage policy and VM disk policies to ensure appropriate capacity utilization, availability, and performance for each virtual machine based on the number of hosts provisioned.

It's important to understand that storage policies are applied during initial VM deployment, or during specific VM operations – cloning or migrating. The default storage policy cannot be changed on a VM after it's deployed, but the storage policy can be changed per disk.

In most cases customers should be using RAID-5 or RAID-6 with FTT set to 1 or 2 depending on their cluster size. If you start off with a 6 host cluster, the policy will default to RAID-1 FTT-1. If you know that you'll expand the cluster in the near future, I recommend deploying with 8-12 hosts from the start.

The following are common storage policies are introduced when a stretched cluster is deployed:

Primary Failures to Tolerate (PFTT)

- Dual Site Mirroring
- None – Keep data on preferred fault domain
- None – Keep data on non-preferred fault domain

Secondary Failures to Tolerate (SFTT)

RAID	FTT	Minimum Hosts
RAID-1 (Default)	1	3 per AZ
RAID-5	1	4 per AZ
RAID-6	2	6 per AZ

Network Connectivity

The connectivity between AZs where the stretched clusters are deployed has bandwidth of at least 10 Gb and a maximum of 5ms round trip time (RTT).

If you're using an ExpressRoute circuit to connect your on-premises data center to your AVS private cloud, you should considering peering both AZs to this circuit with Global Reach to ensure connectivity isn't lost in the event of an AZ failure.

While standard charges will be incurred for things like hosts, ExpressRoute circuits, ExpressRoute gateways, etc, there are no charges for inter-AZ traffic or for the vSAN witness node.

Because this is a single AVS Private Cloud being deployed across AZs, your management network addresses are retained, and your NSX segments are extended, across AZs. Additionally, NSX-T edges are pinned to each AZ with BGP.

SLA

While vSAN Stretched Clusters are designed to provide a SLA of 99.99%, while standard clusters provide 99.9%, a number of conditions must be in place:

1. VMs must use a storage policy configured with a PFTT of Dual Site Mirroring and a minimum SFTT of 1.
2. Compliance with the **Additional Requirements** captured in the SLA Details of Azure VMware Solution.

Summary and Additional Resources

Additional Resources

For more information about vSAN, you can explore the following resources:

- vSAN for Azure VMware Solution
- VMware vSAN Stretched Cluster Guide
- Microsoft Documentation: Deploy vSAN Stretched Clusters
- VMware Documentation: Administering VMware vSAN
- VMware Documentation: vSAN Monitoring and Troubleshooting

Changelog

The following updates were made to this guide:

Date	Description of Changes
2023/06/13	

About the Author and Contributors

- Jeremiah Megie, Principal Cloud Solutions Architect, Cloud Services, VMware