Feature Brief: Stretched Clusters

VMware General

Table of contents

Feature Brief: Stretched Clusters

Introduction

Stretched Clusters for VMware Cloud on AWS is designed to protect against an AWS availability zone failure. Applications can span multiple AWS availability zones within a VMware Cloud on AWS cluster. If an application instance fails the second instance in another availability zone can take over. Let’s take a closer look at the Stretched Clusters feature.

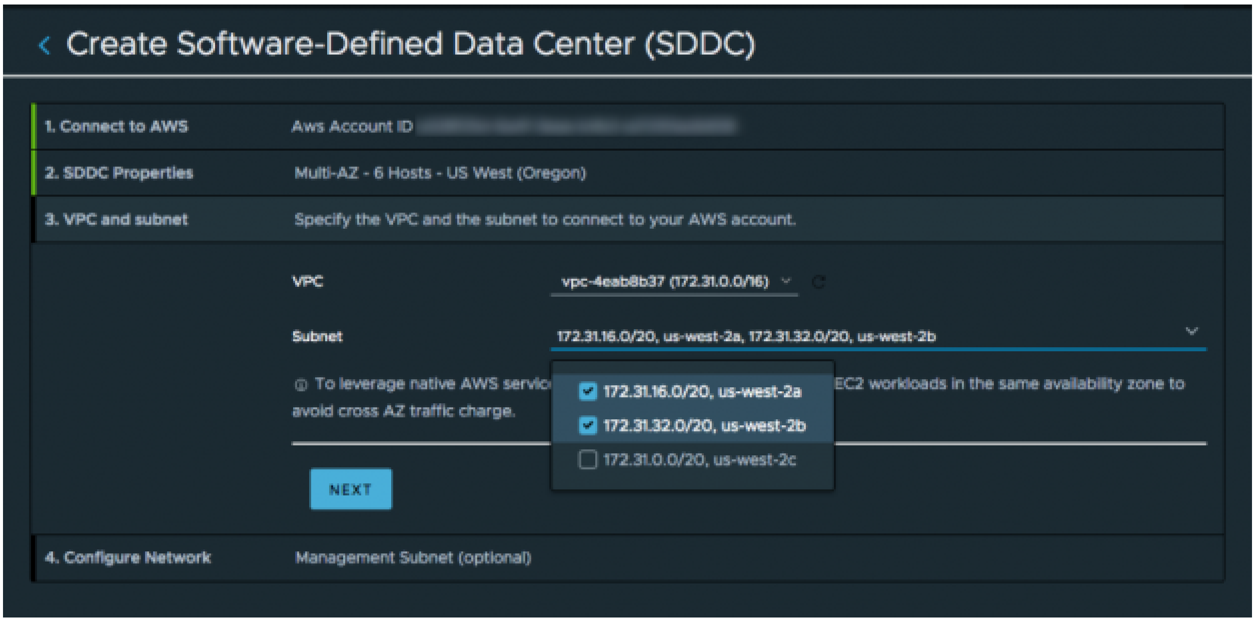
AWS Regions and Availability Zones

Amazon's global infrastructure is broken up into Regions. Each Region supports the services for a given geography. Within each Region, Amazon builds isolated and redundant islands of infrastructure called Availability Zones (AZ). When VMware deploys a vSphere Cluster as part of the VMware Cloud on AWS managed service; all hosts for a given cluster are placed into a single AZ. As a result, VMware Cloud on AWS clusters are vulnerable to the incredibly rare but real threat of an AZ failure. To negate this consideration, Amazon recommends deploying a given service across multiple Availability Zones; utilizing network failover to mitigate any failures.

Architecture

When selected, a vSAN Stretched Cluster is created across three AZ's, creating a vSphere Cluster that can survive the loss of entire Availability Zone. This requires an AWS VPC with two subnets, one subnet per availability zone. The subnets determine an ESXi host placement between the two availability zones. To protect against split-brain scenarios and help measure site health a managed vSAN Witness is also created in the third AZ. The Witness has been engineered to run on an EC2 M5.XL AMI to reduce the cost to the customer.

During the Create SDDC request, the cloud admin selects which VPC subnet should be linked to the tenant workload logical network.   In doing so they are implicitly also designating which AZ the vSphere hosts are deployed in.



Create Software-Defined Data Center (SDDC)

1. Connect to AWS — Aws Account ID [redacted]

2. SDDC Properties — Multi-AZ - 6 Hosts - US West (Oregon)

3. VPC and subnet — Specify the VPC and the subnet to connect to your AWS account.

VPC: vpc-4eab8b37 (172.31.0.0/16)

Subnet: 172.31.16.0/20, us-west-2a, 172.31.32.0/20, us-west-2b

- ☑ 172.31.16.0/20, us-west-2a
- ☑ 172.31.32.0/20, us-west-2b
- □ 172.31.0.0/20, us-west-2c

ⓘ To leverage native AWS servic EC2 workloads in the same availability zone to avoid cross AZ traffic charge.

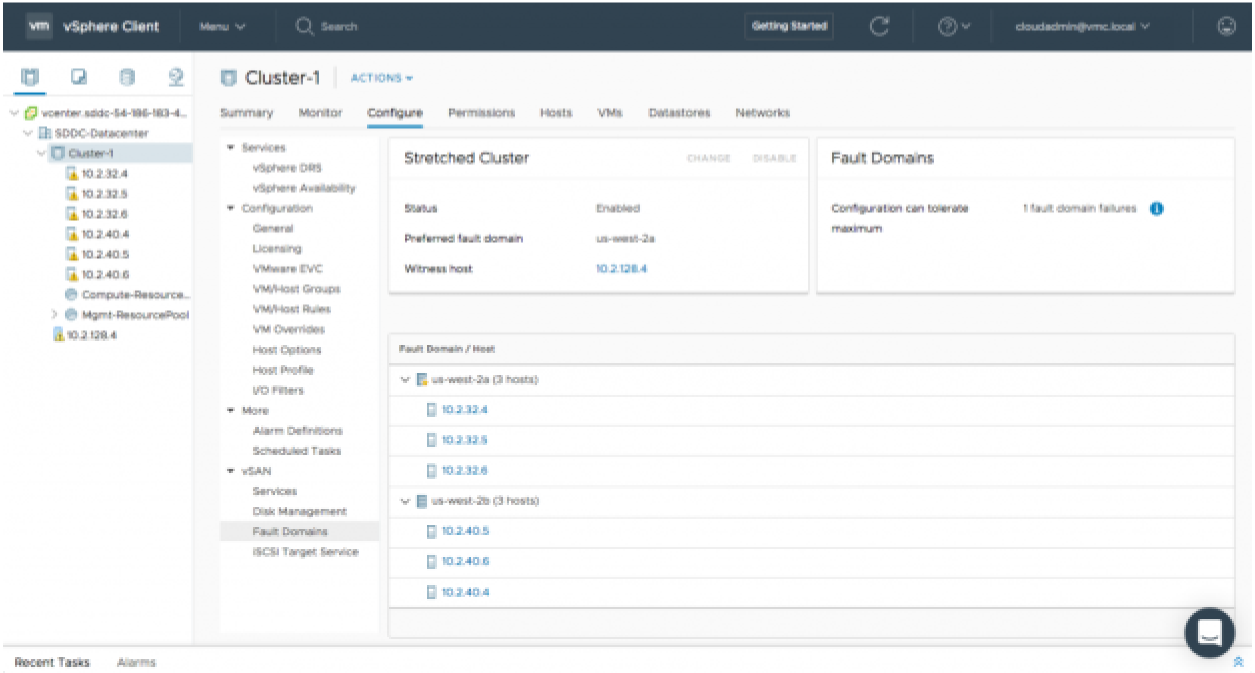
4. Configure Network — Management Subnet (optional)

Finally, vSAN Fault Domains are configured to inform vSphere and vCenter which Hosts reside in which Availability Zones. Each Fault domain is named after the AZ it resides within to increase clarity. Logical networks are also extended using NSX to support workload mobility across the two AWS availability zones.

The following demo shows how to deploy a stretched cluster and also explains the failover process:

Managing a Stretched Cluster

Upon initially logging into your newly provisioned Multi-AZ SDDC vCenter instance you will discover what appears to be a standard vSphere Cluster. Taking a closer look, Fault Domain configuration reveals a fuller story. Here we can see how the nodes have been evenly split between the two AZs specified in the SDDC creation wizard. We can also see the vSAN Witness node deployed as a stand-alone host (which, as mention above, is actually an EC2 instance rather than an actual physical host). Furthermore, we can see that the Cluster can survive the loss of any single Fault Domain without impacting running workload.



Stretched Cluster

Status: Enabled
Preferred fault domain: us-west-2a
Witness host: 10.2.128.4

Fault Domains

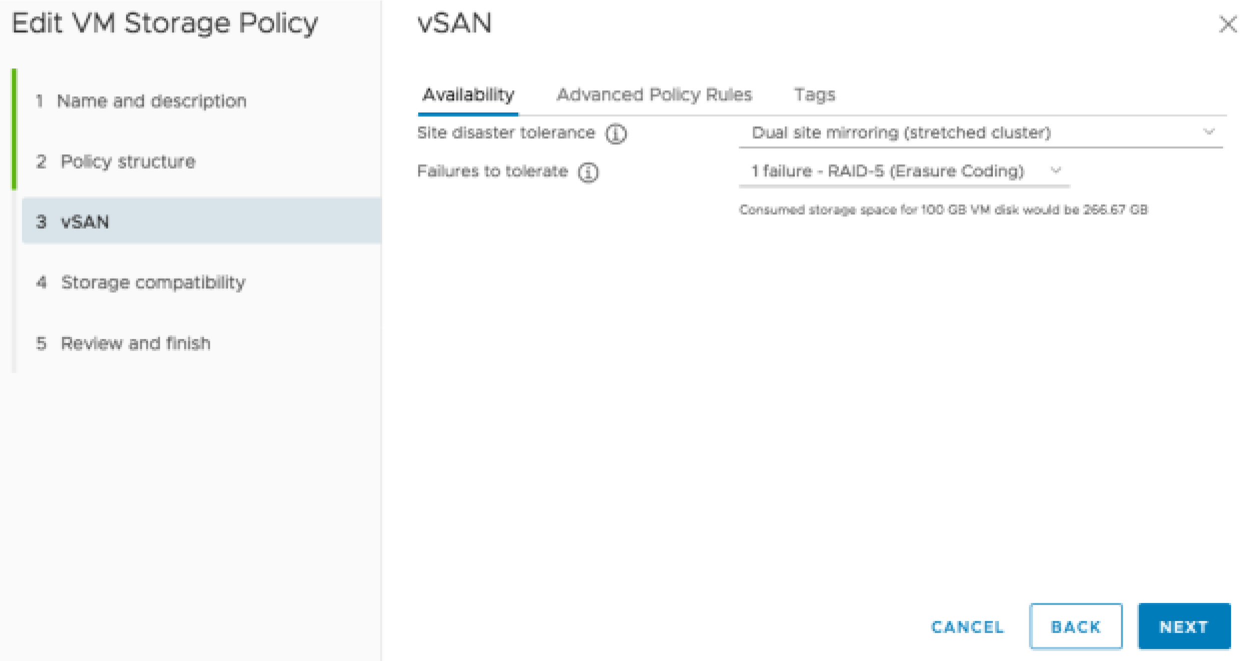
Configuration can tolerate maximum: 1 fault domain failures

Fault Domain / Host
us-west-2a (3 hosts)
10.2.32.4
10.2.32.5
10.2.32.6
us-west-2b (3 hosts)
10.2.40.5
10.2.40.6
10.2.40.4

However, there isn't anything to "manage" on the VMC Cluster.   VMC on AWS abstracts away all the complexity traditionally associated with running a multi-site active/active vSphere Cluster. Perhaps more importantly, VMware operations are responsible for the monitoring and maintenance of the Cluster.   The only thing that customers need do to make VMs capable of surviving the loss of an entire Availability Zone is to configure the VM Storage Policy.

Policy-Based Site and data management

The vSAN policy panel's Availability tab combines both local and site data management into a single view.



Edit VM Storage Policy

1 Name and description
2 Policy structure
3 vSAN
4 Storage compatibility
5 Review and finish

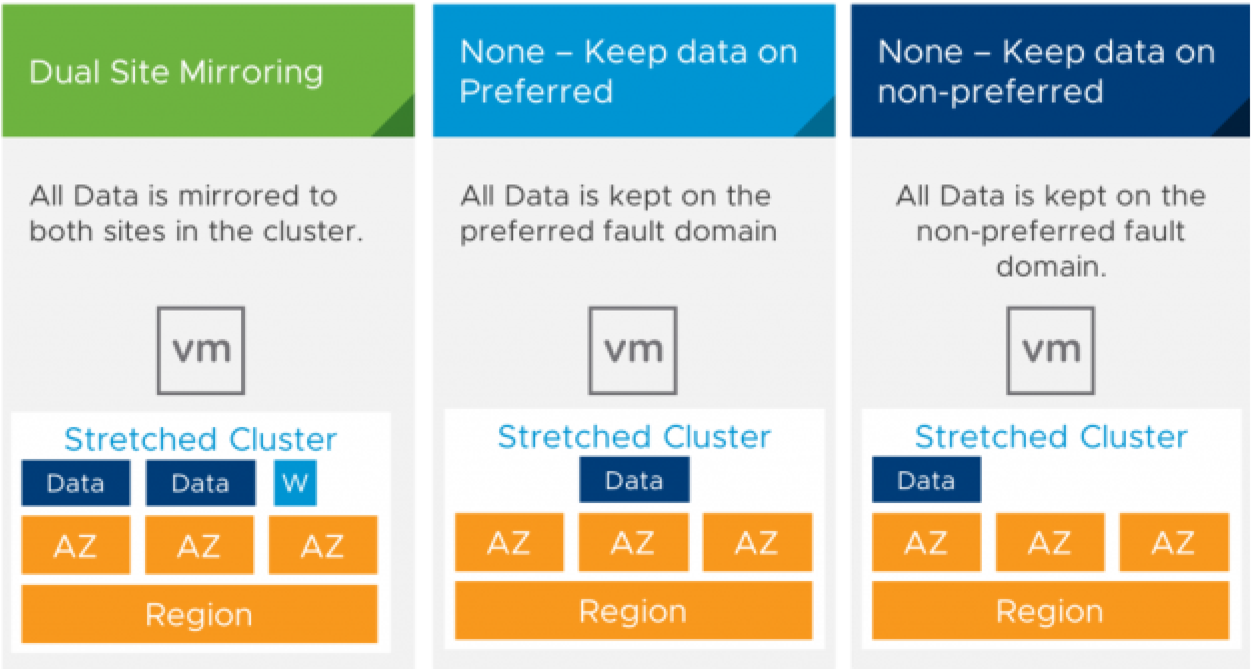
vSAN

Site disaster tolerance: Dual site mirroring (stretched cluster)

Failures to tolerate: 1 failure - RAID-5 (Erasure Coding)

Consumed storage space for 100 GB VM disk would be 266.67 GB

Diving into the Site Disaster Tolerance settings, we discover that we can control where our data is stored. There are three options for managing data placement.



Dual Site Mirroring

All Data is mirrored to both sites in the cluster.

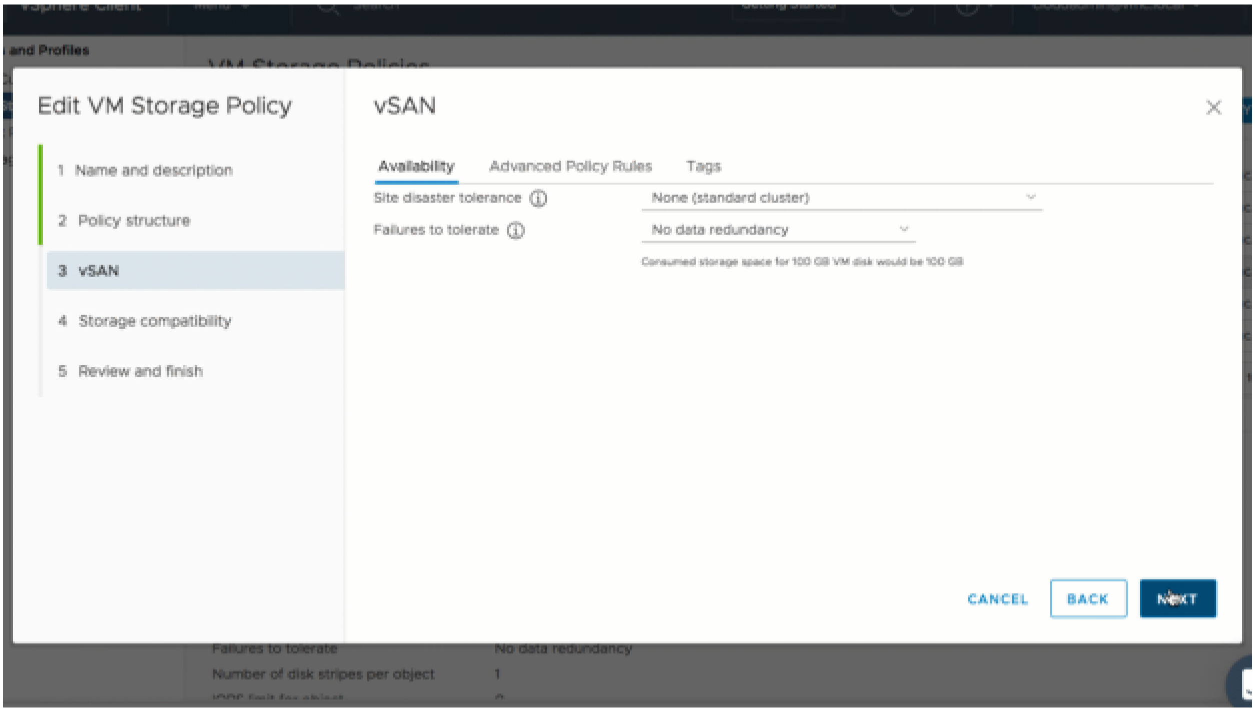
None – Keep data on Preferred

All Data is kept on the preferred fault domain

None – Keep data on non-preferred

All Data is kept on the non-preferred fault domain.

Similarly, the cloud admin can specify how vSAN should store the data within each fault domain via the Failures to Tolerate setting. Failure to Tolerate is a simple concatenation of the previous Failure Toleration Methodology (FTM) and Secondary Failures to Tolerate (SFTT) settings.   The new interface removes all guesswork from policy authoring.



Edit VM Storage Policy

1 Name and description
2 Policy structure
3 vSAN
4 Storage compatibility
5 Review and finish

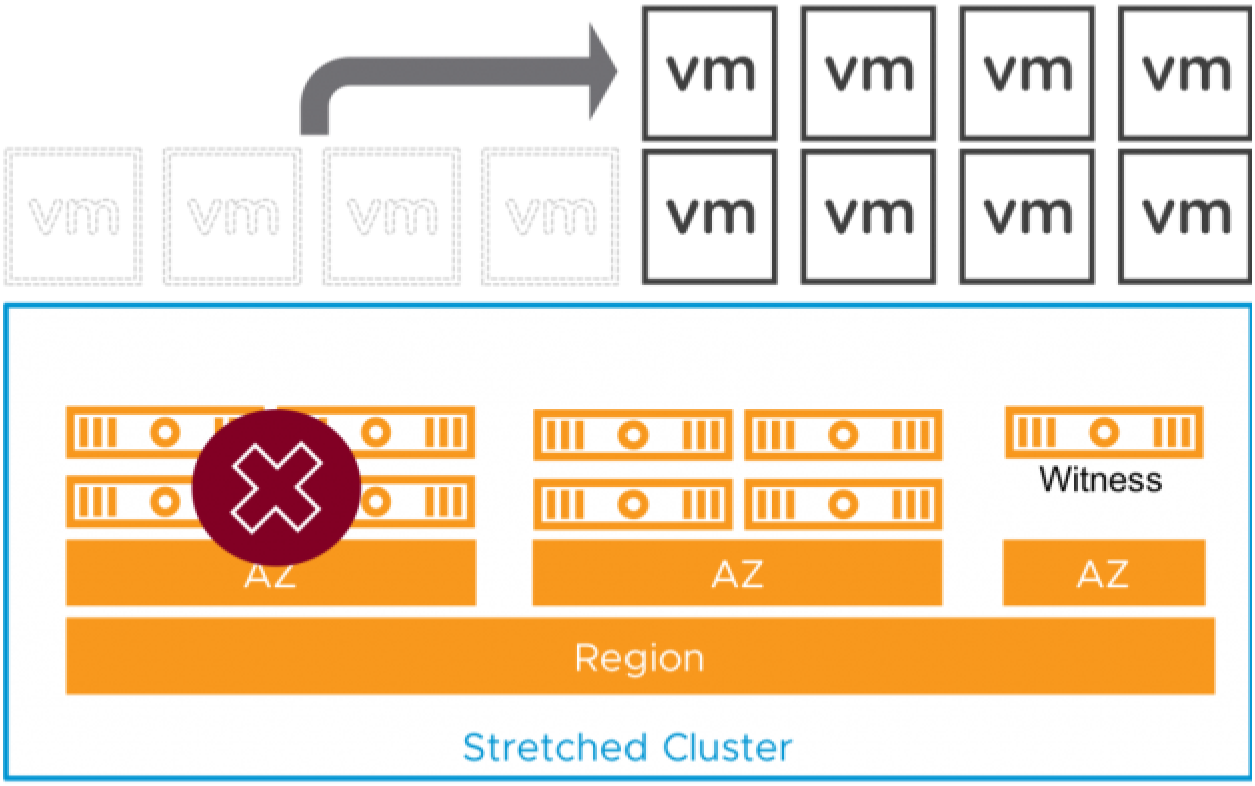
vSAN

Site disaster tolerance: None (standard cluster)

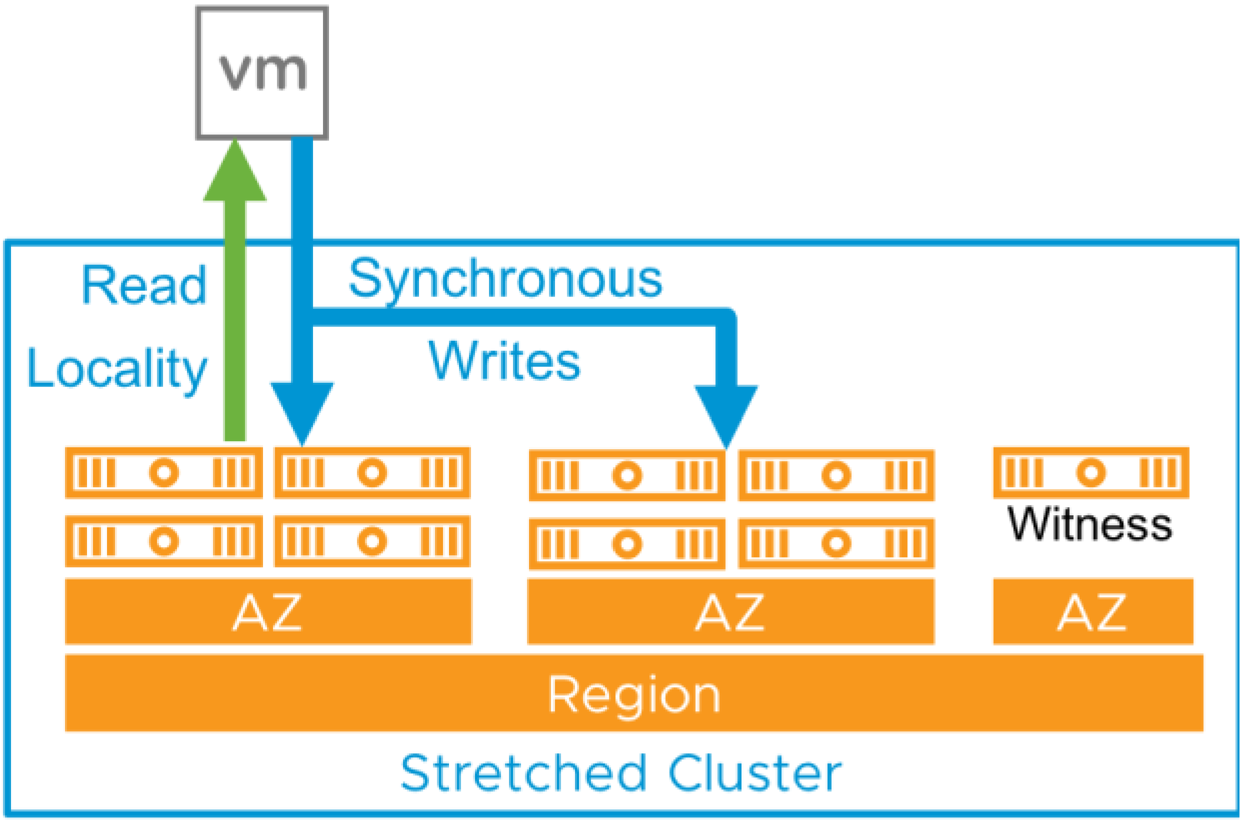
Failures to tolerate: No data redundancy

Consumed storage space for 100 GB VM disk would be 100 GB

Merely declare how you would like vSAN to protect any individual VM and or VMDK and the underlying storage system will make it so. By assigning a Site Disaster Tolerance setting of Dual Site Mirroring in a VM Storage Policy you can make any VM resilient to an AZ failure. In the event of a failure, vSphere HA will restart the affected VM on the surviving hosts.



This capability is the Cloud equivalent of an On-premises vSphere Metro Storage Cluster. The enabling technology here is vSAN, and its ability to synchronously commit any writes across two fault domains (AZ's).



However, Stretched Clusters for VMware Cloud on AWS allows customers to manage costs by enabling local and site-wide availability to be tuned to the needs of each individual item via VM Storage Policy assignment.