TECHNICAL WHITE PAPER:
November 2025

Running Microsoft SQL Server Failover Cluster Instance on VMware vSAN with VMware Cloud Foundation 9

Table of contents

Executive Summary

Business Case

Modern enterprises prioritize high performance, availability, and cost efficiency in their Database Management Systems (DBMS). Over the years, Clustering Databases have become the mainstream choice over standalone databases in production environments. Clustering improves the availability of Microsoft SQL Server (further referenced as SQL Server) instances by providing a failover mechanism to a new node in a cluster in the case of physical or operating system failures. The most used model is the active-passive nodes creating an embedded base of clustering systems that run on the VMware Cloud Foundation™ platform powered by VMware vSAN™.

VMware vSAN has achieved widespread adoption as a storage solution for business-critical applications, including SQL Server with Very Large Database (VLDB) use cases over 50TB and beyond. It delivers a highly scalable, available, reliable, and high-performance storage infrastructure utilizing cost-effective hardware, specifically direct-attached disks in VMware ESXi™ hosts. vSAN's policy-based storage management paradigm streamlines and automates complex management workflows, thereby simplifying configuration and clustering compared to traditional enterprise storage systems.

Furthermore, vSAN Stretched Clusters extend the vSAN cluster from a single data site to two sites for a higher level of availability and inter-site load balancing. Stretched clusters can be used to manage planned maintenance and avoid disaster scenarios, because maintenance or loss of one site does not affect the overall operation of the cluster. vSAN Stretched Clusters further provide a solution for clustered applications like SQL Server to use shared disk across sites. This allows data center administrators to run workloads using legacy clustering technologies on vSAN across two data centers which can fully leverage compute resources on the data centers while having the capability to sustain one site failure.

vSAN supports running Windows SQL Server Failover Clusters Instances (FCI) natively since version 6.7 Update 3. This advancement empowers data center administrators to deploy workloads utilizing traditional clustering technologies on vSAN, supporting shared target storage locations when the storage target is exposed through vSAN's native capabilities for SQL Server Failover Cluster Instances.

vSAN Native Support for Windows Server Failover Clusters (WSFC)

With VMware Cloud Foundation 9, vSAN provides native support for virtualized Windows Server Failover Clusters (WSFC). It supports SCSI-3 Persistent Reservations (SCSI3PR) on a virtual disk level required by WSFC to arbitrate access to a shared disk between nodes. Support of SCSI-3 PRs enables configuration of WSFC with a disk resource shared between VMs natively on vSAN datastores.

The validated/tested limits for running WSFC on vSAN is as listed below:

- 16 maximum number of WSFC nodes per ESXi hosts
- 50 maximum number of WSFC clusters per vSAN cluster
- 45 maximum number of shared disks per WSFC node
- 256 maximum number of shared disks per vSAN cluster
- 58 maximum number of shared disks per ESXi host
- 6 maximum number of nodes in a WSFC cluster

Note the above limits are recommended maximum deployment scale of shared disks for WSFC running on vSAN. If your deployment scale is beyond the recommended maximum limits, you can reach out to your Broadcom representatives for support.

Solution Overview

This reference architecture validates the solution of a Microsoft SQL Server Failover Cluster Instance using shared disks backed by vSAN for the following two scenarios.

- Deploy SQL Server Failover Cluster Instance on a standard vSAN cluster
- Deploy SQL Server Failover Cluster instance on a vSAN stretched cluster

We also showcase the SQL Server Failover Cluster Instance role failover for the above two scenarios.

Solution Architecture

Figure 1 illustrates the architecture for deploying SQL Server FCI on a standard vSAN cluster. This configuration utilizes SCSI-3 Persistent Reservations to arbitrate shared access to clustered disk resources through shared VMDK disks. vSAN offers flexible storage policies that can be tailored to the specific storage requirements of various database workloads. For vSAN ESA, the recommended erasure coding configuration is RAID-5/6, which provides enhanced storage efficiency while maintaining performance levels comparable to RAID-1 mirroring.

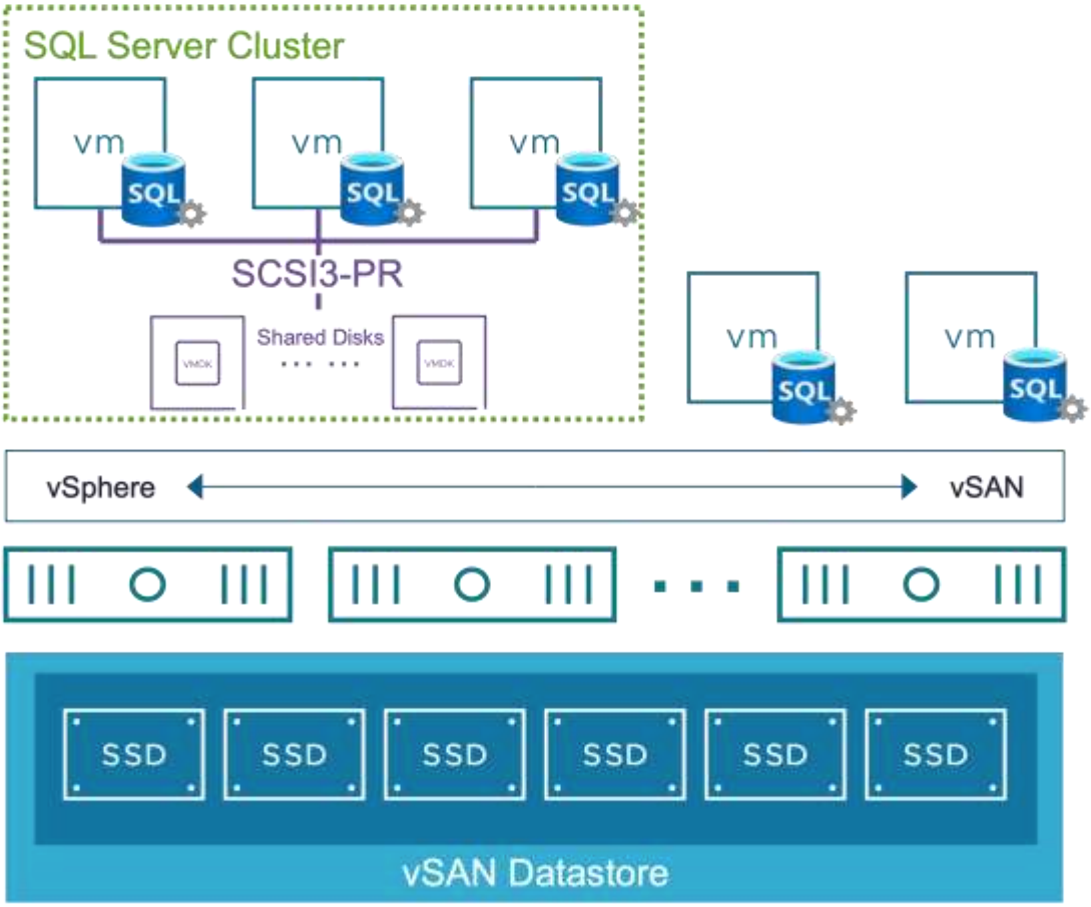
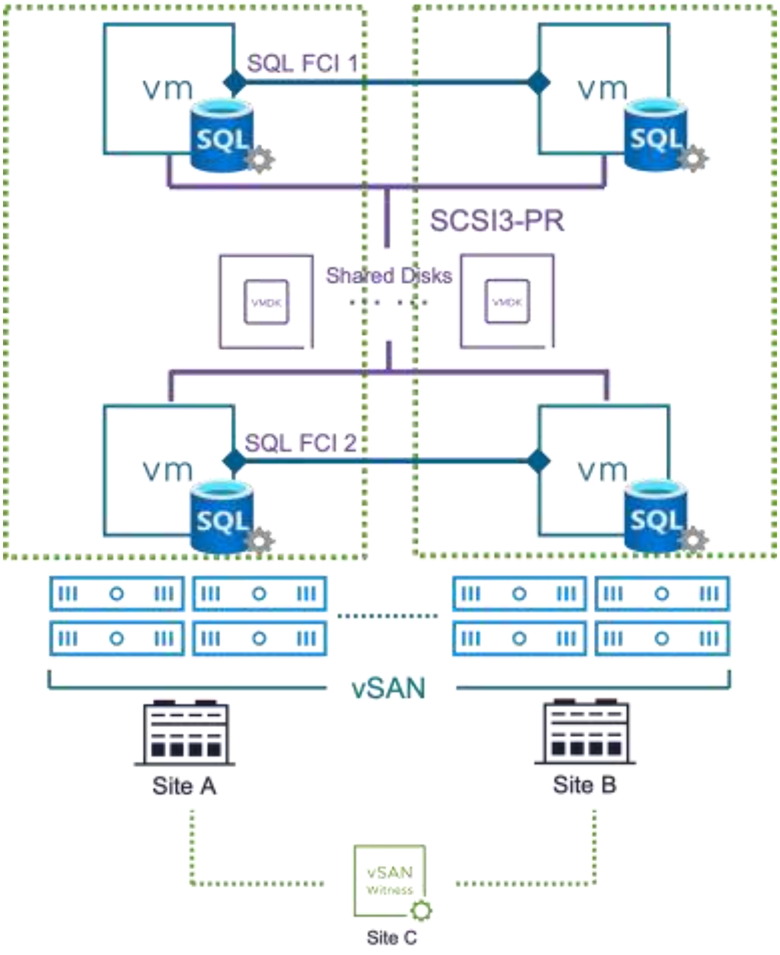


Figure 2 depicts the solution architecture for deploying SQL Server FCI on a vSAN stretched cluster. In this setup, two SQL Server FCI clusters are configured, with one active node strategically placed on Site A and the other active node on Site B. A vSAN Witness appliance is deployed on Site C, facilitating Layer 3 routing to bridge vSAN traffic between Site A and Site B. This distributed architecture ensures resilience and optimal performance across geographically separated data centers.



vSAN stretched clusters

Shared Disks Configuration on vSAN

To enable shared disk support on vSAN, the virtual machines within the SQL Server Failover Cluster Instance must adhere to specific requirements to ensure proper functionality and data integrity. These requirements are critical for the successful implementation of shared storage in a vSAN environment:

- **SCSI Bus Sharing Configuration:** Shared disks must be accessed through a SCSI controller configured with **SCSI Bus Sharing** set to **Physical**. This setting is essential to enable the use of SCSI-3 Persistent Reservations (SCSI3PR), which are required by Windows Server Failover Clusters (WSFC) to arbitrate access to a shared disk among multiple nodes.
- **Disk Mode Setting:** To prevent unsupported snapshot operations on the shared disks, the **Disk Mode** for all disks within the cluster must be set to **Independent** – **Persistent**. This configuration ensures that changes to the shared disks are immediately and permanently written, preventing potential data inconsistencies or loss that could arise from snapshot-related issues.

Configuration Steps

You may follow the configuration steps described in the Knowledge Base article - Configuring a shared disk resource for Windows Server Failover Cluster (WSFC) and migrating SQL Server Failover Cluster Instance (FCI) from SAN (RDMs) to vSAN

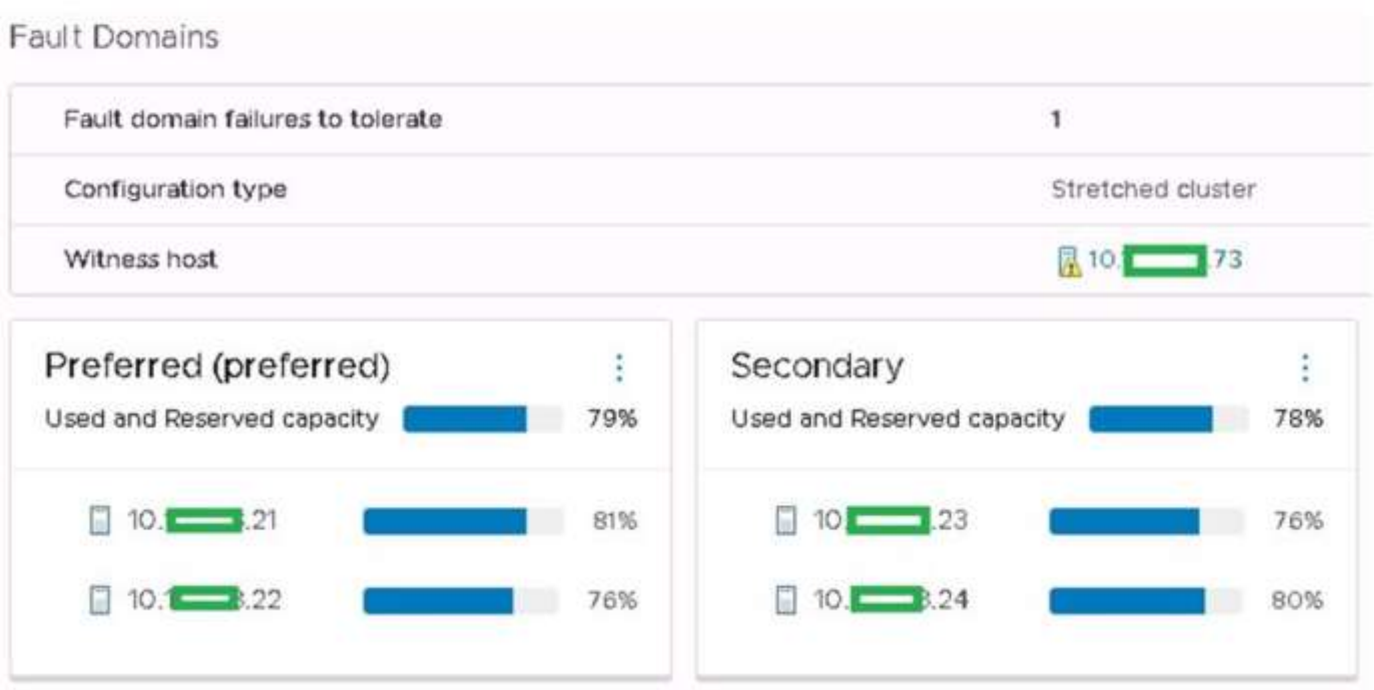
vSAN Storage Policy for shared disks

We recommend using the following storage policy for shared disks on vSAN that consumed by SQL Server FCI

- For vSAN OSA, use RAID-5 for data file disks and RAID-1 for log file disks
- For vSAN ESA, use RAID-5 or RAID-6 according to the desired failure tolerance level (FTT) of the database workloads.
- Thin provisioning is the default settings for shared disks on vSAN. You may override the settings by creating a dedicated storage policy with the desired Object Space Reservation (OSR) for the shared disks.

Node placement for vSAN stretched cluster

In the sample configuration of shared disks on vSAN stretched cluster, we configured two fault domains with two nodes in each domain. The vSAN witness host resided on a management cluster in a different site. See the figure below for the fault domain configuration for vSAN stretched cluster.



Fault Domains

Fault domain failures to tolerate	1
Configuration type	Stretched cluster
Witness host	10.[redacted]73

Preferred (preferred)

Used and Reserved capacity 79%

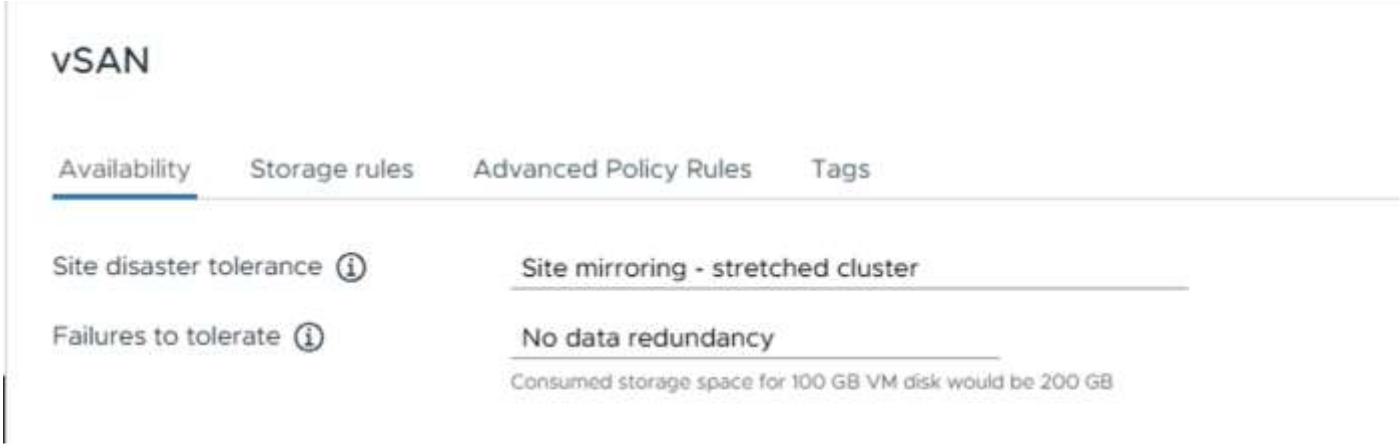
10.[redacted].21	81%
10.[redacted].22	76%

Secondary

Used and Reserved capacity 78%

10[redacted].23	76%
10.[redacted].24	80%

As for WSFC cluster 1, the active node was placed on site A, and the standby node was placed on site B, and WSFC cluster 2 nodes were placed with the reversed order. We set the “Site disaster tolerance” rule of the VM Storage Policy to “Site mirroring (stretched cluster)” for VM home, OS, DB/Log and TempDB VMDK, and the quorum VMDK. See the figure below for the cross-site RAID-1 setting in the VM Storage policy. Note for failures to tolerate within each site, the test configuration of two nodes per site only allowed data protection across sites. Data protection within sites requires three or more hosts to satisfy local site protection policies and you may use RAID-1 mirroring or RAID 5/6 depending on your workload business SLA requirements.



vSAN

Site disaster tolerance ⓘ: Site mirroring - stretched cluster

Failures to tolerate ⓘ: No data redundancy

Consumed storage space for 100 GB VM disk would be 200 GB

Application Role Failover

For the standard vSAN cluster, we verified the application role failover for SQL Server FCI cluster using Benchmark Factory REST API to detect the job status and check the running status of the workload job. If the job was stopped, we restarted the job immediately.

Our test results demonstrated the following:

- The SQL Server application role failover duration was about 20 seconds to bring online both instances hosting two-sized databases with different numbers of shared disks.
- Using REST API can detect and trigger the job restart in 8 seconds, but the test application needs some time to prepare and restart the test. Shown in the table below, the test client needed 48 seconds and 77 seconds to restart the application. This duration was longer than the instance failover time. That means if the failover duration (20 seconds) can be shorter than the restart duration of the application (48 seconds and 77 seconds), the failover will not cause the connection issue. Developers can use the failover duration as reference for their application timeout setting or try-catch-retry logic.

Table 4. Application Role Failover Duration

Database size and number of shared disks	Failover Cluster brings online the resources	BMF restarts job duration
300GB (5 shared disks)	~ 20 seconds	~ 48 seconds
600GB (7 shared disks)		~ 77 seconds

For vSAN stretched clusters, application role failover for SQL FCI can happen from one site to another if the active node is on site A and non-active node is on site B. The duration of a SQL Server instance failover from active node to standby node(s) in a WSFC only requires a few seconds. We validated the following scenarios and demonstrated the advantages of running SQL Server FCI on vSAN stretched clusters.

- **Scenario 1 – Application Role Failover:** we manually moved the SQL Server FCI from one node to another node
- **Scenario 2 – Host shutdown (non-primary node of WSFC):** we shut down the host which hosted the non-primary node of the WSFC to emulate the unplanned host failure. Before this failure validation, we initiated the workload on one database.
- **Scenario 3 – Site shutdown (primary node of WSFC):** we shut down all the hosts on the site where the primary node of WSFC ran to emulate the unplanned site failure. This scenario will cause VMs restarting on hosts in a different site.

Here’s the validation results:

- Application role failover as expected by moving FCI active nodes from one node to another.
- Failure of the non-primary node did not cause workload interruption, and vSphere HA restarted the impacted VM on the remaining hosts.
- Failure of the primary node (caused by site failure) restarted the VM on the other site and no cluster service down was monitored when disk witness is configured as the quorum disk for WSFC on vSAN stretched cluster.

Recommendations for running SQL Server FCI on vSAN stretched cluster

- Less than four milliseconds inter-site (round trip) latency is recommended for tier-1 SQL Server databases running on vSAN stretched clusters.
- Enable DRS VM/Hosts rule and create rules to separate the VMs of one WSFC on different ESXi hosts. And enable VM/Hosts rule to separate the VMs of different WSFC nodes on different ESXi hosts for performance consideration.

- Use quorum disk witness as the cluster service quorum setting and vSAN stretched cluster can ensure the witness disk accessibility for WSFC in a site failure without tearing down the FCI cluster service.

Conclusion

vSAN natively supports the deployment of SQL Server Failover Cluster Instances by enabling shared VMDKs through SCSI-3 Persistent Reservations. This approach simplifies configuration by eliminating the need for complex LUN settings typically associated with third-party storage, thereby reducing the management overhead of pRDMs from legacy storage. vSAN offers unparalleled flexibility for deploying Windows Failover Clustering-related applications, providing robust support for both standard and stretched cluster configurations without compromise.

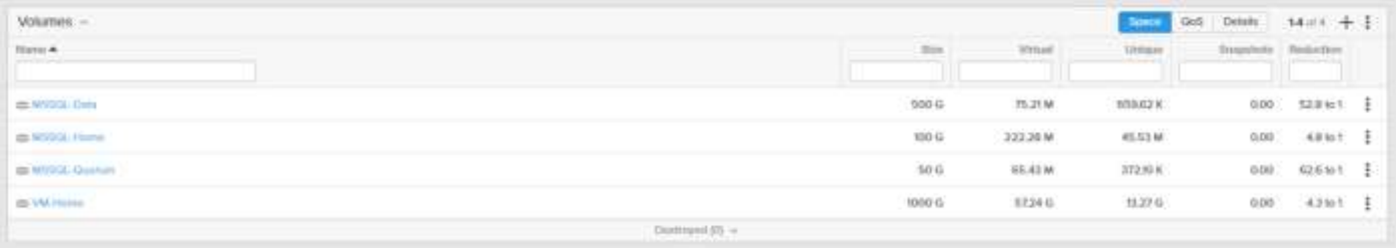
Migration of shared disks from SAN to vSAN

Migration Prerequisites

To simulate a shared disk use case for pRDM, we created four LUNs on a 3rd party storage with the capacity and usage described in Table 5. The VM Home LUN is mounted as a VMFS datastore for VM Operating System disks. The other 3 LUNs are used as pRDM disks for the purpose of SQL Server FCI home directory, user database and WSFC Witness respectively.

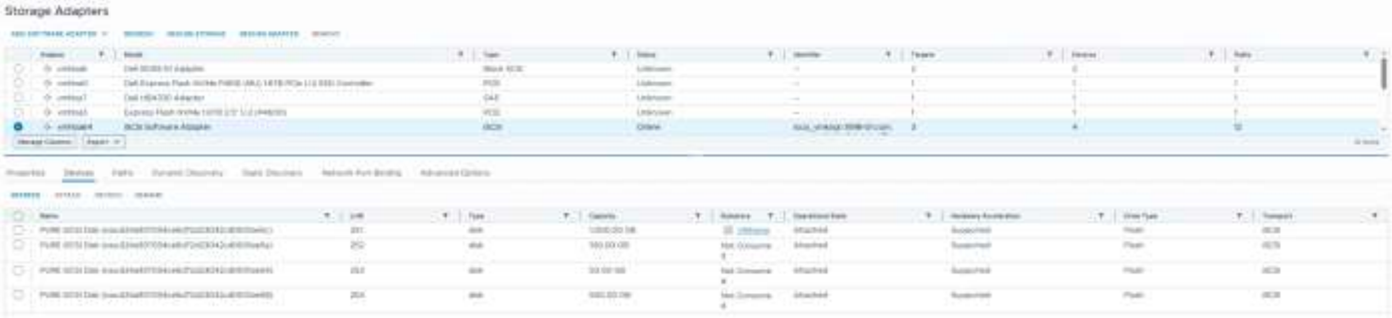
Name	Size (GB)	Purpose
VM Home LUN	1000	Shared datastore for VM Operating System disk
SQL Home Directory LUN	100	SQL Server system database disk
SQL User Database LUN	500	SQL Server user database disk for data and log files
WSFC Witness LUN	50	WSFC quorum disk

Figure 6 shows the created LUNs from 3rd party storage for the preparation of the migration from SAN to vSAN.

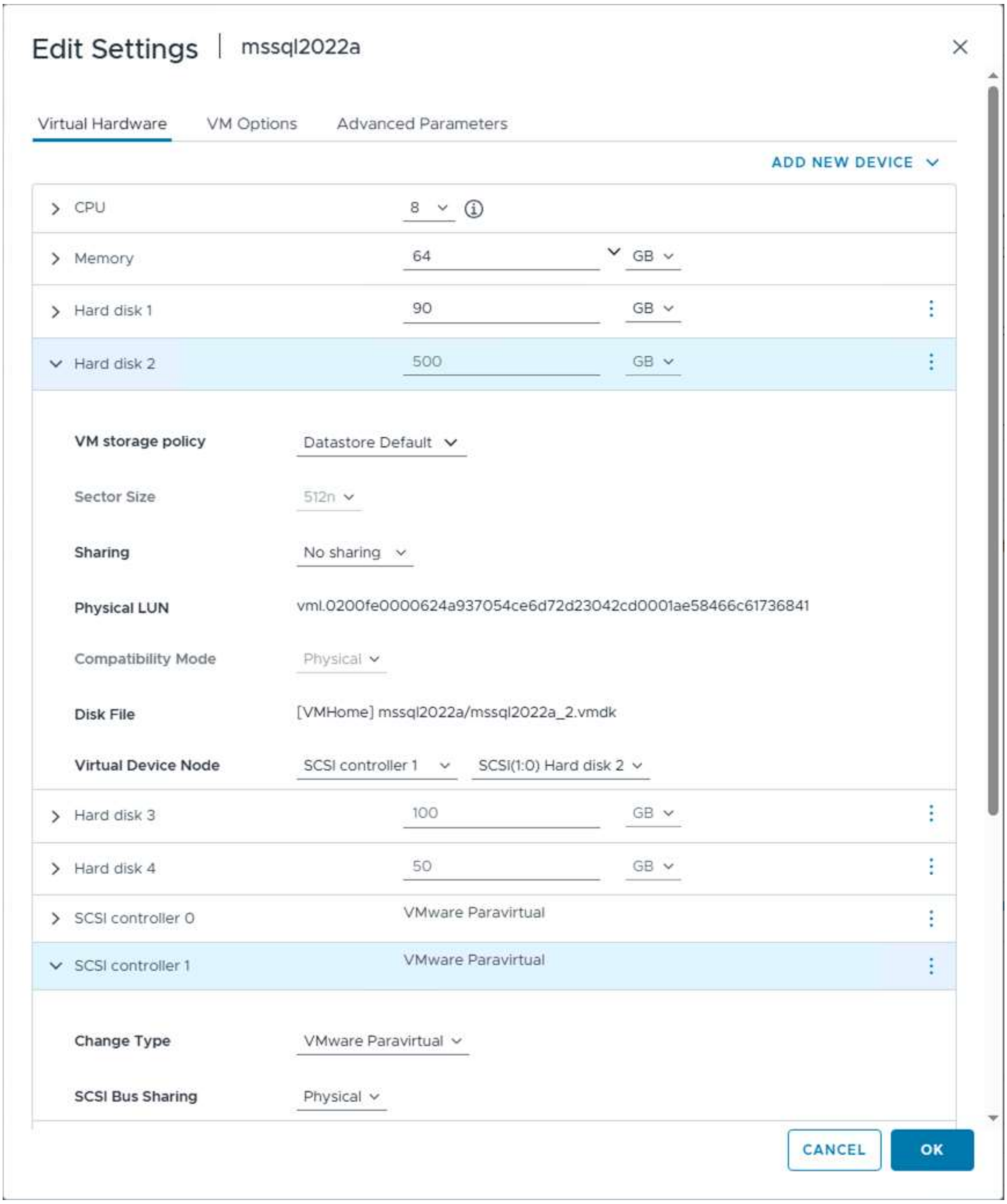


Name	Size	Virtual	Unique	Snapshots	Reduction
MSSQL-Data	500 G	75.21 M	959.62 K	0.00	52.8 to 1
MSSQL-Home	100 G	322.26 M	45.53 M	0.00	4.8 to 1
MSSQL-Quorum	50 G	65.43 M	372.10 K	0.00	62.6 to 1
VM-Home	1000 G	57.24 G	13.27 G	0.00	4.3 to 1

We created iSCSI interface on the 3rd party storage and connected to the interface from the ESXi hosts. As shown in Figure 7, the disks can be accessed by each ESXi host through the iSCSI software initiator.



The disk exposed as RDM in physical compatibility mode or pRDM to SQL Server, virtual machine will have fixed capacity (greyed-out) and use the physical mode of the SCSI Controller. Note the vmdk file associated with the pRDM disk is only a mapping file and we will migrate the pRDM to shared VMDK on vSAN using the storage migration wizard.



CPU	8	
Memory	64	GB
Hard disk 1	90	GB
Hard disk 2	500	GB
VM storage policy	Datastore Default	
Sector Size	512n	
Sharing	No sharing	
Physical LUN	vml.0200fe0000624a937054ce6d72d23042cd0001ae58466c61736841	
Compatibility Mode	Physical	
Disk File	[VMHome] mssql2022a/mssql2022a_2.vmdk	
Virtual Device Node	SCSI controller 1	SCSI(1:0) Hard disk 2
Hard disk 3	100	GB
Hard disk 4	50	GB
SCSI controller 0	VMware Paravirtual	
SCSI controller 1	VMware Paravirtual	
Change Type	VMware Paravirtual	
SCSI Bus Sharing	Physical	

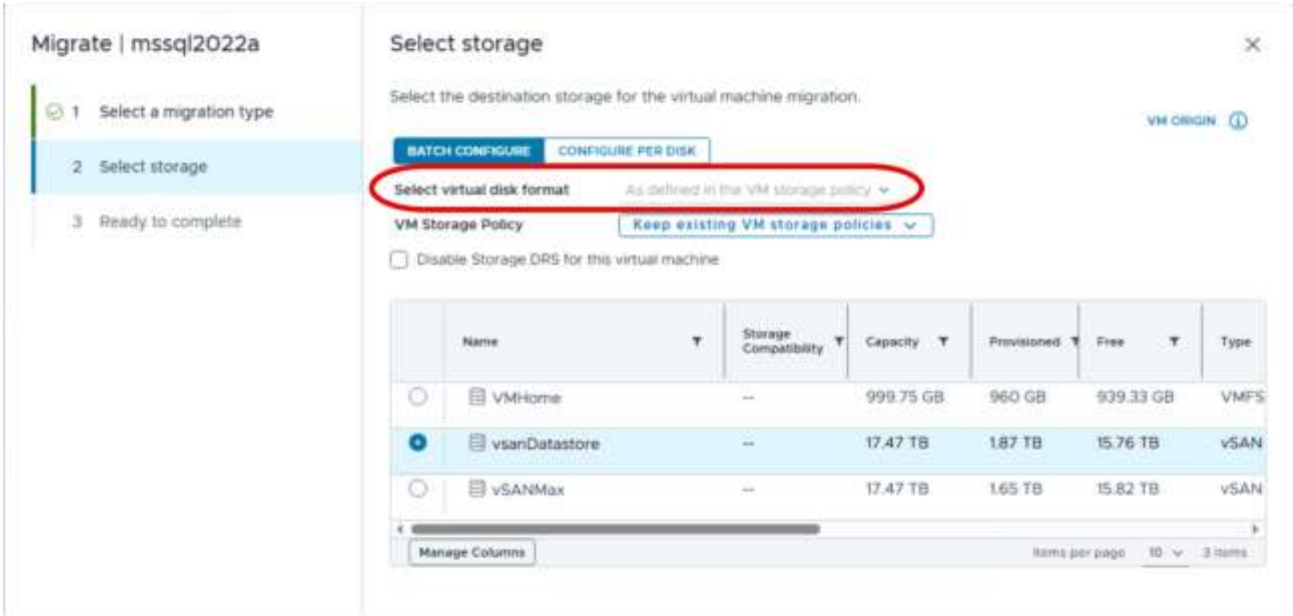
Migration Steps

Note that before migration, backup is highly recommended to avoid potential data loss. This migration operation is offline, and the duration is mainly for data copy. Make sure the offline time window is enough before moving forward.

To migrate the SQL Server FCI cluster using pRDMs as clustered disk resources to vSAN, follow the steps below:

1. Stop the SQL Server Cluster Role from the Windows Failover Cluster Manager.
2. Shut down all the VMs hosting nodes of Windows Failover Cluster gracefully, by clicking **Power** -> **shut down guest OS** or within the Guest OS
3. Migrate the first node of a cluster to vSAN by choosing **Change storage only** in the **Migrate** wizard. The migration process will convert pRDMs to VMDKs and apply the desired vSAN storage policies for clustered disks in the migration wizard.

Note to successfully migrate pRDM disk to VMDK, you must specify disk format to thin or thick provisioned as mentioned in Converting a Raw Device Mapping (RDM) into virtual disk (VMDK). However for vSAN the disk format is maintained in storage policy therefore the selection is greyed out in the Migrate wizard as shown in Figure 1.



Migrate | mssql2022a

1 Select a migration type

2 Select storage

3 Ready to complete

Select storage

Select the destination storage for the virtual machine migration.

Select virtual disk format: As defined in the VM storage policy

VM Storage Policy: Keep existing VM storage policies

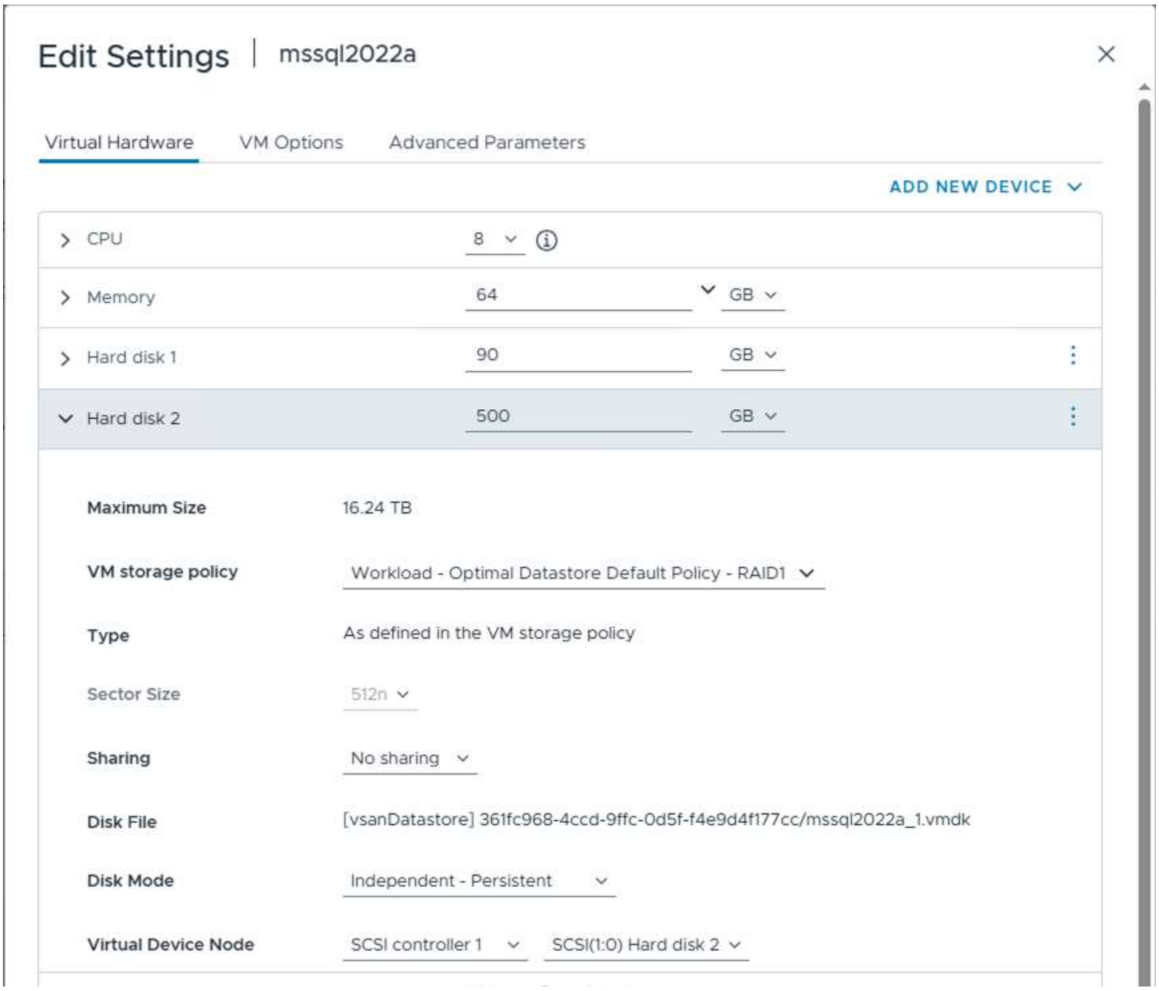
☐ Disable Storage DRS for this virtual machine

	Name	Storage Compatibility	Capacity	Provisioned	Free	Type
○	VMHome	--	999.75 GB	960 GB	939.33 GB	VMFS
●	vsanDatastore	--	17.47 TB	1.87 TB	15.76 TB	vSAN
○	vSANMax	--	17.47 TB	1.65 TB	15.82 TB	vSAN

A workaround is to use PowerCLI to initiate the storage migration of pRDM disk to VMDK on vSAN by specifying **StorageFormat** parameter to **thin** or **thick** provisioning. You can check Appendix A for reference of sample code.

4. Power on the first node and validate that clustered disk resources are visible in the Windows Failover Cluster Manager and SQL Server Cluster Role can be started, and you may keep it online.
5. Detach pRDMs used to host clustered disk resources from all remaining nodes of the cluster, which are not migrated to vSAN yet.
6. (Optional) Migrate the remaining nodes to vSAN, if non-shared disks are planned to migrate to vSAN as well.
7. Attach disk resources back to remaining nodes of the cluster pointing to VMDKs from the first node stored on the vSAN datastore. Ensure that vSCSI controllers hosting disks are configured to use physical mode and share the VMDKs across virtual machines for the previous pointing disks on the cluster nodes.
8. Start up the virtual machines one by one and make sure the SQL Server Cluster Role is online on the first node, try to failover from the active node to the passive node to check if the other nodes can start SQL Server Cluster Role normally.

Figure 8 shows the disk was provisioned by vSAN instead of from pRDM and you can change the policy and size according to the requirement.



Edit Settings | mssql2022a

CPU	8	
Memory	64	GB
Hard disk 1	90	GB
Hard disk 2	500	GB

Maximum Size	16.24 TB
VM storage policy	Workload - Optimal Datastore Default Policy - RAID1
Type	As defined in the VM storage policy
Sector Size	512n
Sharing	No sharing
Disk File	[vsanDatastore] 361fc968-4ccd-9ffc-0d5f-f4e9d4f177cc/mssql2022a_1.vmdk
Disk Mode	Independent - Persistent
Virtual Device Node	SCSI controller 1 SCSI(1:0) Hard disk 2

After following the steps above, the shared virtual disks of SQL Server cluster are all on vSAN. Users can manage the virtual disks by using vSAN storage policy including expanding the disk and changing policy to follow the best practice to run SQL Server on vSAN to meet different business requirements.

Appendix A – Sample PowerCLI code to convert pRDM to VMDK on vSAN

```
# Variables
$vmName      = "mssql2022a"
$datastore   = "vsanDatastore"   # Target vSAN datastore
$diskNumbers = 2..4              # pRDM Disks to migrate (Hard disk 2, 3, and 4)

# Get VM object
$vm = Get-VM -Name $vmName

foreach ($diskNumber in $diskNumbers) {
    $hardDisk = Get-HardDisk -VM $vm | Where-Object { $_.Name -eq "Hard disk $diskNumber" }

    if ($hardDisk) {
        Write-Host "Migrating Hard disk $diskNumber of VM $vmName to $datastore as Thin provisioned..."
        Move-HardDisk -HardDisk $hardDisk -Datastore $datastore -StorageFormat Thin -Confirm:$false
    }
    else {
        Write-Warning "Hard disk $diskNumber not found on VM $vmName, skipping..."
    }
}

Write-Host "Migration completed for disks $($diskNumbers -join ', ') of VM $vmName."
```

Item No: vmw-bc-wp-tech-temp-a4-word-2024 1/24