

vSphere 7 Microsoft Windows Server Failover Cluster with Clustered VMDKs

VMware Cloud Workloads Technical Topic Series

Introduction

VMware vSphere® 7.0 provides support for SCSI-3 Persistent Reservations (SCSI3-PRs) on a virtual disk (VMDK) level. Support for SCSI-3 PRs enables configurations of a Windows Server Failover Cluster (WSFC) across VMware ESXi™ hosts, with clustered VMDKs as disk resources, in alignment with the requirements and supportability outlined in this document.

Technology Overview

WSFC uses SCSI-3 PRs commands to coordinate access to a clustered disk resource. These commands (SCSI Persistent Reserve Out and Persistent Reserve IN) are emulated at the VSCSI layer on a datastore, assuming support for this feature exists within the datastore. A datastore configured to host clustered VMDKs will be referred to as a clusteredVMDK datastore further in this document.

Requirements

VMware ESXi, VMware vCenter®, VMware vSphere VMFS

- Require all hosts connected to a clusteredVMDK datastore to be on ESXi version 7.0 or higher, managed by the same vCenter instance, version 7.0 and higher.
- All ESXi hosts involved in hosting nodes of a WSFC must be managed by the same vCenter instance. A cross vCenter WSFC is not supported (i.e., when ESXi hosts, hosting VM nodes of a WSFC are managed by different vCenter instances)
- Requires VMFS version 6

Virtual Machine (VM)

- Clustered VMDKs must be Eager Zeroed Thick (EZT).
- Clustered VMDKs must be attached to a virtual SCSI controller with bus sharing set to physical. VM Boot disk (and all VM non-shared disks) should be attached to a separate virtual SCSI controller with bus sharing set to none. Mixing clustered and non-shared disks on a single virtual SCSI controller is not supported.
- Multi-writer flag must NOT be used.

vSphere Cluster

- Require DRS Affinity Rules to separate VMs, nodes of the cluster, to different ESXi hosts (must rule enforcement). For guidance on the creation of rules, click [here](#).

Datastore

- Require the datastore capability Clustered VMDK support to be enabled or the flag should be set after the datastore is created..

(To unset the flag, no VMs should be using clustered disks on the specific datastore, or VMs using the clustered disk should be powered off [if the attribute is removed, a VM will not power on]).

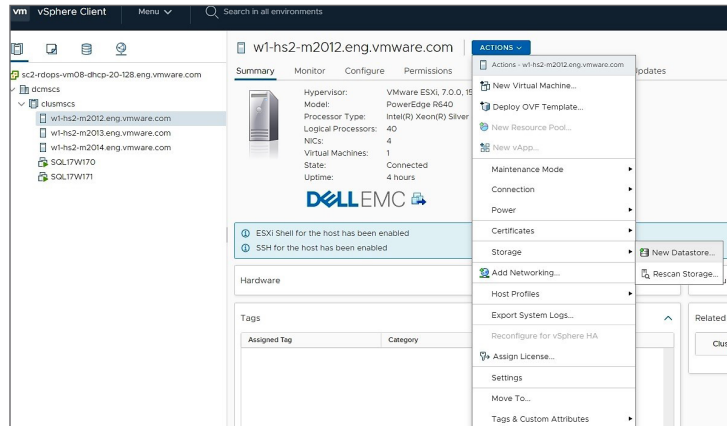
Storage Array/LUN

- Support SCSI-3 reservation type WEAR (i.e., write exclusive all registrant).
- Supported storage system access protocol – Fibre Channel (FC) only.
- The physical disk on which VMDKs are stored should support ATS commands.
- Require 512/512e sector size disks.

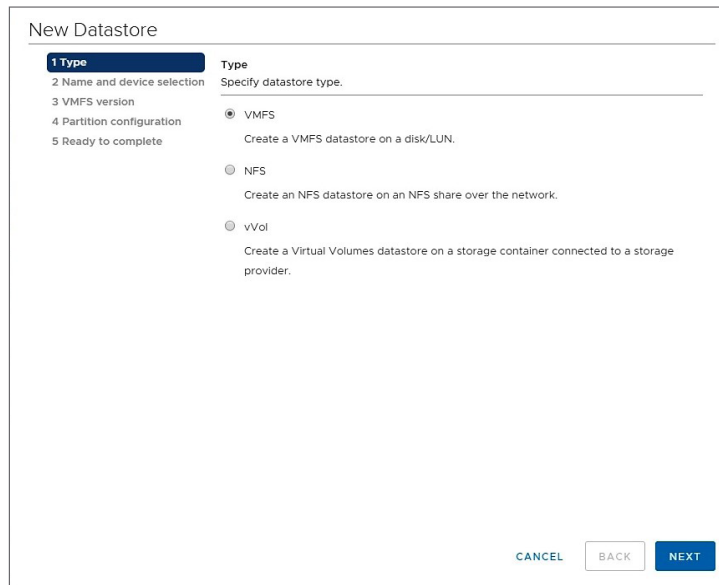
(Consult the VMware Compatibility Matrix for the list of supported arrays.)

Configuration Flow

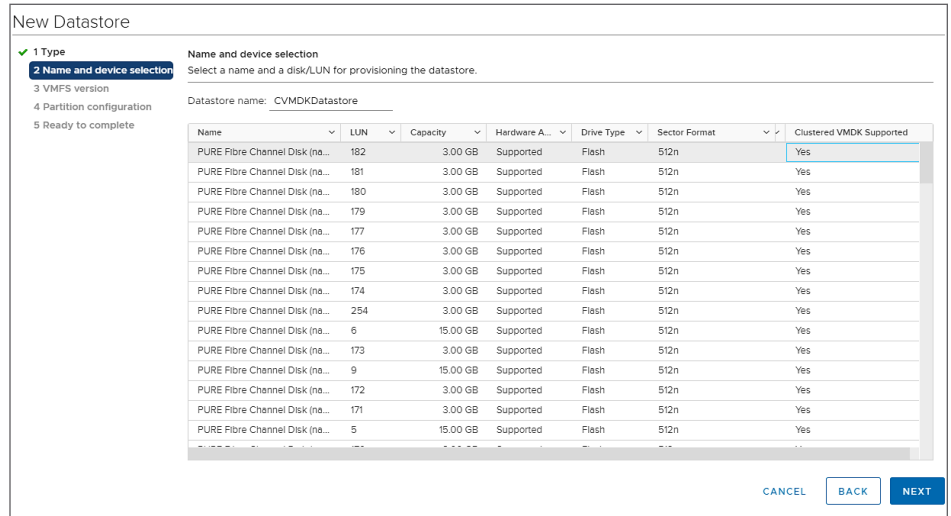
- Check that all requirements are met.
 - Enable Clustered VMDK Support on a datastore from the vSphere Client UI.
1. Navigate to any inventory object that is a valid parent object of a datastore, such as a host, a cluster, or a data center, and right-click to select **Storage** and then **New Datastore**.



2. On the **New Datastore** page, choose datastore type **VMFS**.



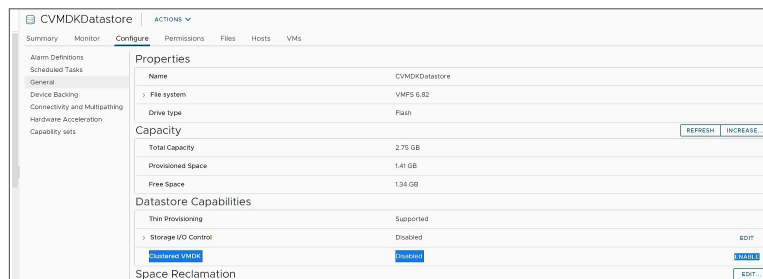
3. On the **New Datastore** page, choose a device with the attribute **Clustered VMDK Supported** set to **Yes**.



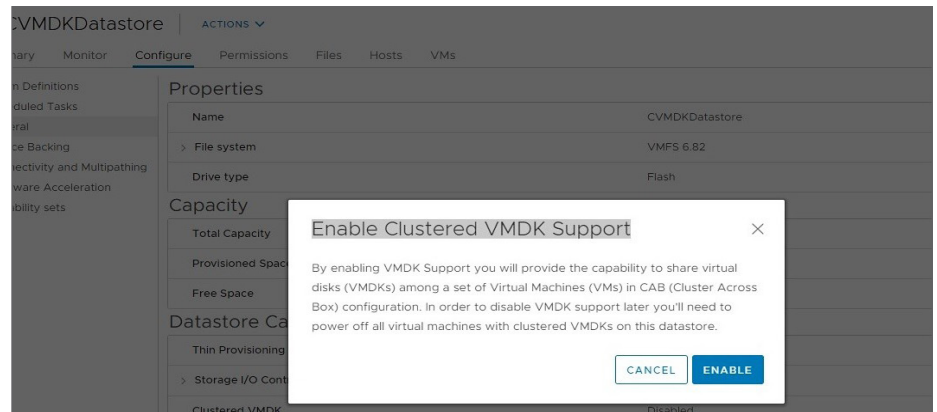
4. Create a **VMFS 6** version datastore.



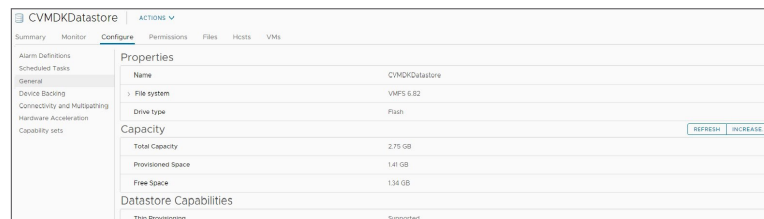
5. Review the disk layout and specify the partition configuration details. Click on **FINISH** to create the clusteredVMDK datastore. Once the datastore is created, go to the **Configure Datastore** tab.



- Under **General**, go to the **Datastore Capabilities** section. Click **Enable** for **Clustered VMDK**.



- Once enabled, use the **Disable** button to disable, if required, a clustered VMDK. Disable will not work if there are any VMs using clustered VMDKs on the datastore. All VMs with shared disks on the datastore must be powered off or migrated out of the datastore.



- Create a first VM (the first node of a WSFC):
 - Follow the wizard to create a new VM. Assign required resources (CPU and Memory) and select the ESXi cluster to host the VM. Finish the wizard.
 - Add a new virtual SCSI controller (VMware Paravirtual/LSI SAS [VMware Paravirtual is recommended]).
 - Change the SCSI bus-sharing configuration option for the newly added controller to **physical**.
 - Add a new VMDK (E2T) on a prepared clusteredVMDK datastore.
 - Assign the new added VMDK to the new vSCSI controller.
- Create a second VM (the second node of a cluster):
 - Follow the wizard to create a new VM. Assign required resources (CPU and Memory) and select the ESXi cluster to host the VM. Finish the wizard.
 - Add a new virtual SCSI controller (VMware Paravirtual /LSI SAS [VMware Paravirtual is recommended]).
 - Change the SCSI bus-sharing configuration option for the newly added controller to **physical**.
 - Point to an existing VMDK from the first node of the WSFC (use **add an existing disk** configuration option). Assign the VMDK to the new vSCSI controller created previously.
 - Use the same SCSI IDs (Virtual device Node) while assigning disks to all additional nodes

Supportability

vSphere Version

- Datastores with a mix of ESXi and older hosts are not supported. Ensure that no older hosts are zoned to the physical LUN hosting the datastore or upgrade them to vSphere 7.0.

WSFC Related

- Supports up to five (5) nodes per WSFC, equal to the maximum for physical RDMs.
- Maximum number of 128 clustered VMDKs per ESXi host.
- Supports up to three (3) WSFC clusters (i.e., multi-cluster) running on the same ESXi host.
- Mixing of clustered VMDKs and other types of clustered disks (e.g., pRDMs, vVol) in a single VM is not supported.
- Placing all VMs, nodes of a WSFC on the same ESXi host (i.e. Cluster-in-a-Box (CiB) is not supported).
- VMs, nodes of a WSFC, must be placed on different ESXi hosts (i.e. Cluster Across Boxes (CAB)). The placement must be enforced with DRS MUST Anti-Affinity Rules.
- Change/increase in the GOS the WSFC Parameters.
 - (get-cluster -name <cluster-name>).QuorumArbitrationTimeMax = 60
 - (get-cluster -name <cluster-name>).SameSubnetThreshold = 10
 - (get-cluster -name <cluster-name>).CrossSubnetThreshold = 20
 - (get-cluster -name <cluster-name>).RouteHistoryLength = 40

Windows OS Qualified

- Windows Server 2012, 2012R2, 2016, 2019.
(*Microsoft SQL Server versions 2016, 2017 were used to validate the configuration.*)

vSphere Features

- Follow the wizard to create a new VM. Assign required resources (CPU and Memory) and select the ESXi cluster to host the VM. Finish the wizard.
- Live VMware vMotion® is supported when the destination host meets all the requirement listed above.
- Moving more than one node of a WSFC to the same ESXi host is not allowed. vMotion will fail.
- Snapshot and on-line VMware vSphere Storage vMotion are not supported (similar to that of RDM-based clustering).
- Offline Storage vMotion require all VMs, nodes of WSFC, be powered off. After performing offline Storage vMotion of all VMs, nodes of WSFC, ensure to re-point all clustered VMDKs to the new storage location.
- Cloning of a VM with clusteredVMDK is not supported.
- vSphere Fault Tolerance (FT) is not supported.
- Hot changes to the VM hardware, including hot expansion of the clustered disk, is not supported.
- Diagnostic and other ESXi partitions are not supported on the clusteredVMDK datastore.
- A clusteredVMDK datastore cannot be used as an ESXi installation disk.
- A clustered VMDK cannot be used to host a boot disk for GOS

VMFS

- Spanned VMFS datastores are not supported. To expand a clusteredVMDK datastore, expand the backend LUN to increase the datastore capacity.
- VMFS versions 5.0 and older are not supported.
- A physical LUN with a clustered VMDK datastore cannot have any additional VMFS datastores.

Storage

- Only datastores accessible via Fibre Channel (FC) are currently supported.
- Non-SCSI back-end or non-SCSI virtual front-end (e.g., NVMe, vNVMe) are not supported.
- SCSI-2 Reservations will not be supported on clustered disks/datastores.

Windows OS Qualified

- Windows Server 2012, 2012R2, 2016, 2019.
(*SQL Server 2016, 2017 were used to validate the configuration.*)

Notes

- Performance impact – SCSI-3 WEAR type reservation is not expected to introduce any perceivable performance or latency impact.

Best Practices

1. Ensure that all VMs hosting nodes of a WSFC are migrated off or powered off properly before removing them from a clusteredVMDK datastore so that resources (e.g., heartbeat [HB] slots), are freed. In the event of a VM crash or an APD on the clusteredVMDK datastore during power-off, always power on and power off before removing the VM from a cluster.
2. It's not recommended to use a clusteredVMDK datastore for storing non-clustered VMDKs. However, keeping both clustered and non-clustered disks on a clustered VMDK datastore is supported. VMs using non-clustered disks on a clustered VMDK datastore will continue to work normally and all operations(i.e. snapshot, clone etc.), are supported.
3. Do not keep clustered VMDKs for different WSFC clusters on the same clustered VMDK datastore. Use different clusteredVMDK datastores for different clusters.
4. Consider setting vHardware (virtual compatibility) to vSphere 7.0 while using the clustered VMDK feature.

¹To ensure consistent and clear terminology, all references to a virtual disk (VMDK) accessed simultaneously by two or more VMs for the purpose of creating an HA solution are referred to as a clustered VMDK.



VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 vmware.com Copyright © 2020 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies. Item No: VMW_0120_2134_VMDK_Whitepaper_1.5_MMT 3/20