QSAN Official Document

VMware Clustered VMDK

Application Note

July 2022

ANNOUNCEMENT

Copyright

QSAN believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

Trademarks

- QSAN, the QSAN logo, QSAN.com, XCubeFAS, XCubeSAN, XCubeNXT, XCubeNAS, XCubeDAS, XEVO, SANOS, and QSM are trademarks or registered trademarks of QSAN Technology, Inc.
- Microsoft, Windows, Windows Server, and Hyper-V are trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries.
- Linux is a trademark of Linus Torvalds in the United States and/or other countries.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Mac and OS X are trademarks of Apple Inc., registered in the U.S. and other countries.
- Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
- VMware, ESXi, and vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other countries.
- Citrix and Xen are registered trademarks or trademarks of Citrix Systems, Inc. in the United States and/or other countries.
- Other trademarks and trade names used in this document to refer to either the entities claiming the marks and names or their products are the property of their respective owners.

TABLE OF CONTENTS

FIGURES

TABLES

NOTICES

Information contained in this document has been reviewed for accuracy. But it could include typographical errors or technical inaccuracies. Changes are made to the document periodically. These changes will be incorporated in new editions of the publication. QSAN may make improvements or changes in the products. All features, functionality, and product specifications are subject to change without prior notice or obligation. All statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

PREFACE

Executive Summary

This document provides technical guidance for setting up WSFC (Windows Server Failover Clustering) in a VMware environment. A comparison of different deployment methods is presented. And point out the advantages of the new feature cluster VMDK. It also provides installation tips to help users not get tripped up.

Audience

This document is intended for QSAN customers and partners who are familiar with VMware technology and Windows failover clustering. It assumes the reader is familiar with QSAN products and has general IT experience, including knowledge as a system or network administrator. If there is any question, please refer to the user manuals of products, or contact QSAN support for further assistance.

Technical Support

Do you have any questions or need help trouble-shooting a problem? Please contact QSAN Support, we will reply to you as soon as possible.

- Via the Web: https://www.qsan.com/technical_support
- Via Telephone: +886-2-77206355
- (Service hours: 09:30 - 18:00, Monday - Friday, UTC+8)
- Via Skype Chat, Skype ID: qsan.support
- (Service hours: 09:30 - 02:00, Monday - Friday, UTC+8, Summer time: 09:30 - 01:00)
- Via Email: support@qsan.com

Information, Tip, and Caution

This document uses the following symbols to draw attention to important safety and operational information.

INFORMATION

INFORMATION provides useful knowledge, definition, or terminology for reference.

TIP

TIP provides helpful suggestions for performing tasks more effectively.

CAUTION

CAUTION indicates that failure to take a specified action could result in damage to the system.

1. OVERVIEW

This chapter provides an overview of WSFC (Windows Server Failover Clustering) and a brief introduction to its setup in a VMware environment. VMware's new feature cluster VMDK is released in ESXi 7. Finally, a comparison is made between the traditional architecture and the new features.

1.1. Windows Server Failover Clustering

A failover cluster is a group of independent computers that work together to improve the availability and scalability of the cluster role. To reduce system downtime and ensure high availability for Windows, you can cluster servers (called nodes) so that if one node fails, one or more other nodes automatically take over processing. This is also referred to as Windows clustering.

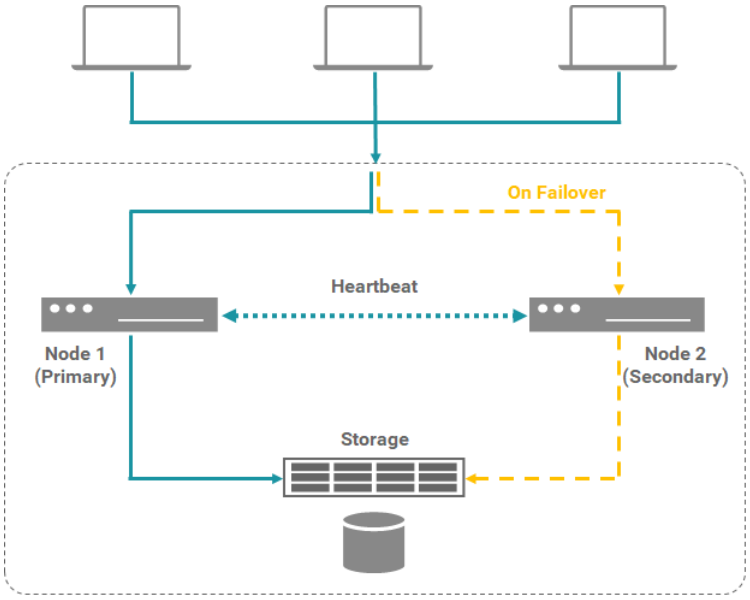

Figure 1-1 Windows Server Failover Clustering Architecture

Cluster servers are connected by physical cables and software. If one or more cluster nodes fail, other nodes begin to provide service (a process known as failover). Additionally, cluster roles are actively monitored to verify that they are functioning properly. If they don't work, they will be restarted or moved to another node.

Failover clustering has many practical applications, including highly available or continuously available file share storage for applications such as Microsoft SQL Server and Hyper-V virtual machines.

1.2. How Windows Clustering Works

Cluster software is required to monitor the health of the primary node and initiate recovery operations when problems are detected. High availability clustering also need a way to ensure that in the event of a failure, the secondary node is accessing the newest data in storage. In most cases, this is achieved by connecting all nodes of the cluster to the same shared storage.

Failover clustering also provides the CSV (Cluster Shared Volume) feature, which provides a consistent distributed namespace that cluster roles can use to access shared storage from all nodes. With failover clustering, users can minimize service interruptions.

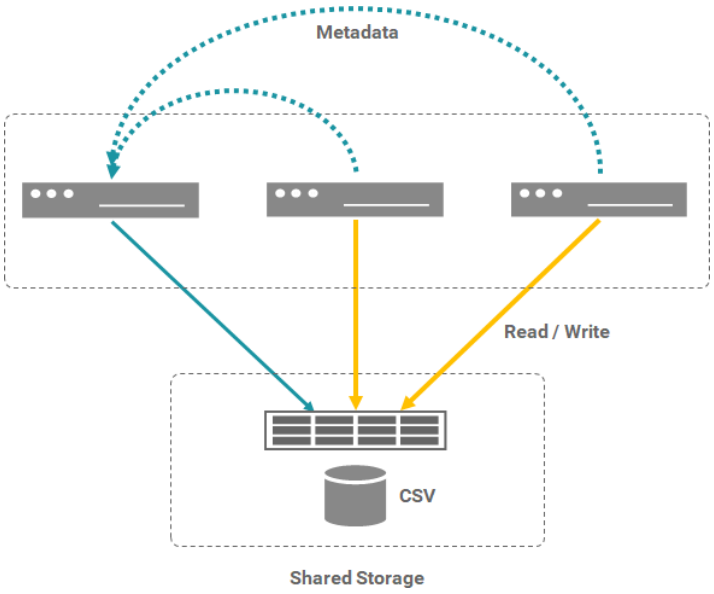

Figure 1-2 Cluster Shared Volume

With CSV feature, all cluster nodes can access the CSV at the same time. Server-side metadata synchronization avoids I/O interruptions. It is recommended that cluster nodes should be geographically separated to protect applications from site-area disasters.

1.3. VMware RDM

WSFC deployments can be virtualized. VMware vSphere supports Windows clustering using WSFC across virtual machines. Clustering virtual machines can reduce the hardware costs of traditional high-availability Windows clusters.

RDM (Raw Device Mapping) is VMware virtualization technology that allows a VM (Virtual Machine) to directly access LUN (Logical Unit Number). It is a special mapping file in a VMFS volume that manages the metadata of its mapped devices. The mapping file is provided to the management software as a normal disk file and can be used for file system operations. For virtual machines, the storage virtualization layer presents mapped devices as virtual SCSI devices.

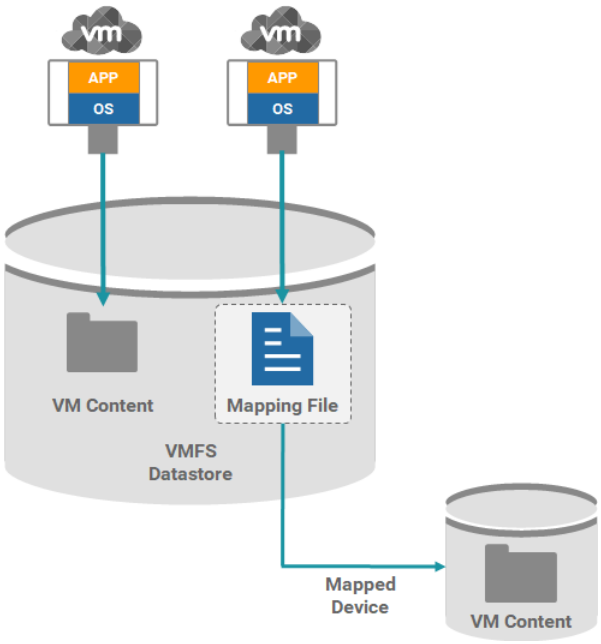

Figure 1-3 Raw Device Mapping

INFORMATION

Regarding RDM, see the following documentation for more details.

- About Raw Device Mapping
- Differences between Virtual and Physical RDM

1.4. Clustered VMDK

In ESXi 7.0, when the VMs hosting the cluster nodes are on different ESXi hosts, clustered VMDKs (Virtual Machine Disk Format) on VMFS (Virtual Machine File System) datastores are supported. VMware has added support for SCSI-3 PR (Persistent Reservations) at the virtual disk level. You can now deploy WSFC using a clustered (shared) VMDK.

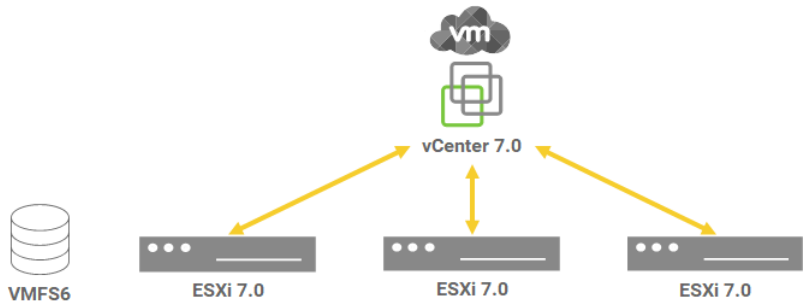

Figure 1-4 Enabling Clustered VMDK

Clustered VMDK support can be enabled when creating a new VMFS6 datastore, or enabled on an existing VMFS6 datastore. Before enabling clustered VMDK support, ensure that all hosts connected to the datastore are using ESXi 7.0 or later and managed by vCenter 7.0 or later. All hosts connected to the datastore must be managed by the same vCenter with the cluster VMDK flag disabled or enabled on the datastore. With the cluster VMDK flag enabled or disabled, the host can be managed by any vCenter with version 7.0 or higher.

1.5. RDM vs. Clustered VMDK

Both RDM and clustered VMDK can help you to setup WSFC, but what we'll highlight here is how RDM compares to clustered VMDK in VMware.

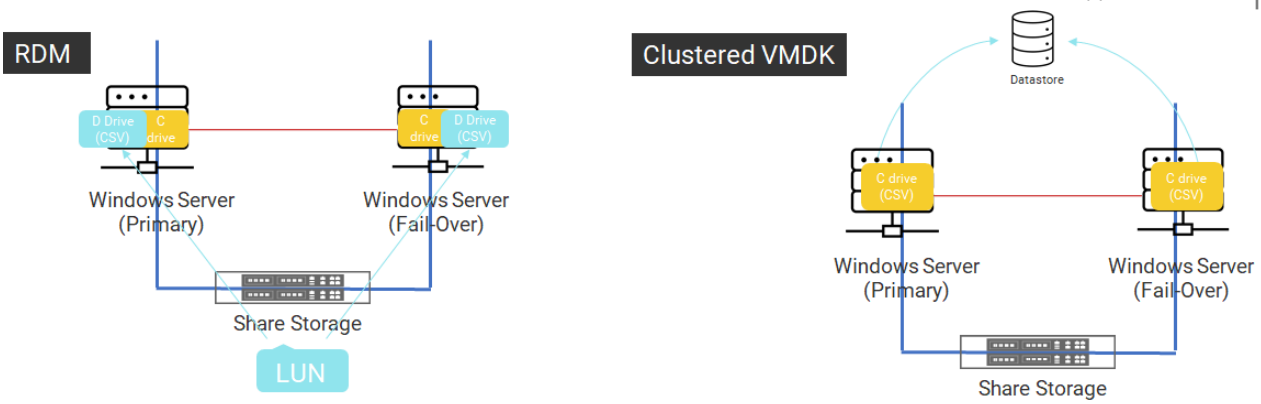

Figure 1-5 RDM vs. Clustered VMDK

Storage in a clustered environment should have a locking mechanism to prevent writes to the same block. It essentially uses this command to lock the volume so only active nodes are allowed to write to it. But since VMFS has its own locking mechanism, these SCSI commands are intercepted and dropped by traditional virtual disks. Therefore, RDM disks need to be used as mapping devices for physical LUNs.

Clustered VMDK allows SCSI-3 PR commands to be issued to virtual disks, which means you will no longer need a dedicated physical LUN to start a Windows failover cluster.

Table 1-1 RDM vs. Clustered VMDK

	RDM	CLUSTERED VMDK
Compare Items	• Local C drive for OS only • Need to map new LUN for D drive • Set D drive into CSV[1] for data saving	• Local C drive can be set as CSV[1] • Extra capacity share the same datastore
Conclusion	• Extra LUN mapping • Extra drive for data	• No extra LUN mapping • One drive only • Reduce setting configuration process

[1] CSV (Cluster Shared Volume)

In summary, supporting clustered VMDK simplified the process of VM application environment when setting up WSFC. You can now migrate and delete those RDMs created in your environment to handle failover clusters, allowing these Windows VMs to access VMware's unified and simplified virtual disk management.

2. INSTALLATION TIPS

This chapter provides tips for installing WSFC on VMware using a clustered VMDK. We emphasize tips rather than complete installation steps, as you can find some detailed documentations on the Internet.

2.1. Prerequisites for Clustered VMDK

Clustered VMDKs come with a bunch of limitations In VMware documents. Some of which are related to arrays.

- Only supported with arrays using FC (Fibre Channel) for connectivity.
- The array must support ATS SCSI commands.
- The array must support SCSI-3 PR (Persistent Reservations), specifically WEAR (Write Exclusive - All Registrants).

INFORMATION

Clustered VMDK is supported in SANOS firmware 2.0.4 and above.

There are some prerequisites to VMware and WSFC for using clustered VMDKs.

- CIB (Cluster in a Box) configuration is not supported.
- The datastore must be formatted with VMFS 6 (VMFS 5 is not supported).
- VMDK must be Eager Zero Thick (no thin provisioned VMDKs).
- If you have DRS configured in your environment, you must create an anti-affinity rule so that the VMs can run on separate hosts.
- vCenter server 7.0 and higher.
- Snapshots, cloning, and storage vMotion are not supported (no backup of nodes is possible, because backup software uses snapshots).

- Fault tolerance, hot change to the VMS virtual hardware, and hot expansion of clustered disks are not.
- vMotion is supported, but only for hosts that meet the same requirements.

2.2. WSFC Topology

The WSFC environment on VMware is shown below.

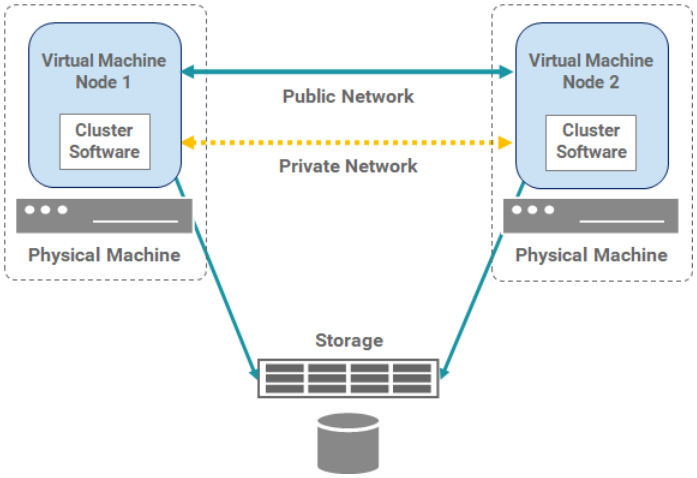

Figure 2-1 Virtual Machines Clustered Across Hosts

Clustering of virtual machines across physical ESXi hosts protects against software and hardware failures on physical ESXi hosts by placing cluster nodes on separate ESXi hosts. This configuration requires shared storage for cluster disk resources. Note that the clustered VMDK supports CAB (Cluster Across Boxes) instead of CIB (Cluster in a Box).

The above figure shows a CAB setup.

- Two virtual machines on two different ESXi hosts running WSFC.
- Virtual machines share private and public network connections for private heartbeats.
- Each virtual machine is connected to shared storage.

2.3. Setup Clustered VMDK Tips

We provide some tips for setting up WSFC via a clustered VMDK, not full steps, as there is already detailed documentation on the installation.

- When you create a datastore in vSphere 7, there is a new column in the Create Datastore wizard called **Clustered VMDK Supported** that will tell you if the array device is.
- Click the **Enable** button during datastore creation.

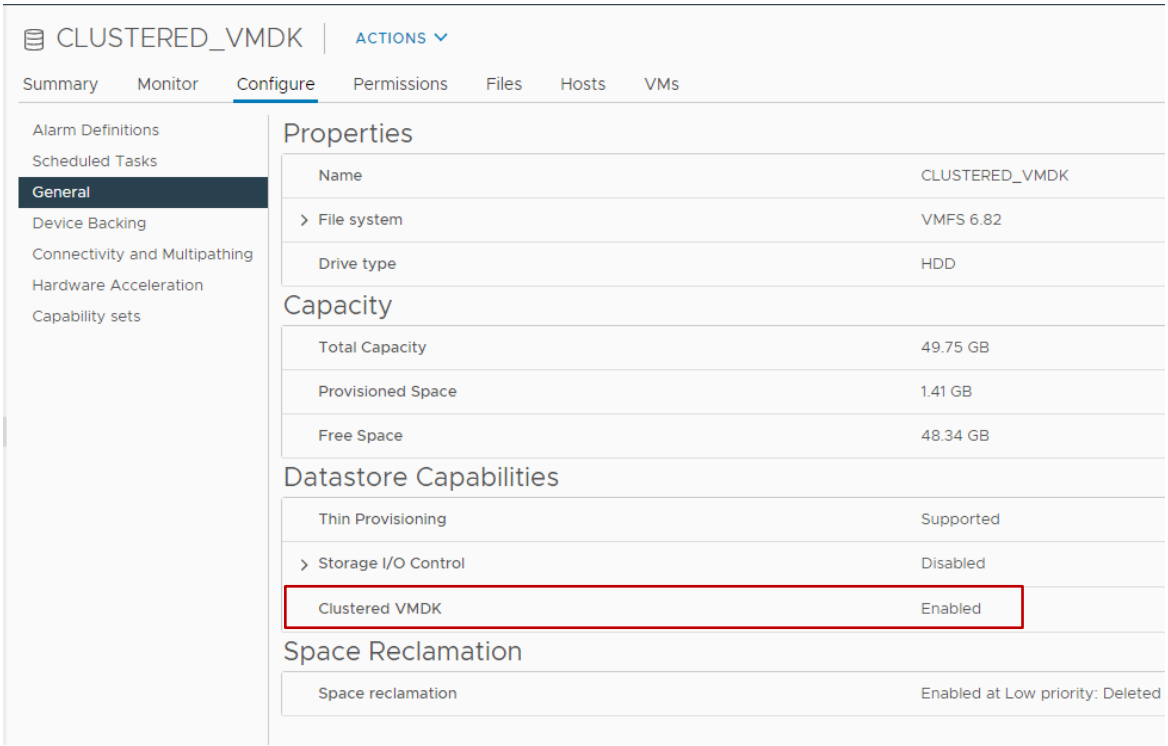

Figure 2-2 Enable Clustered VMDK

- When creating the 1st virtual machine, select the **Disk Provisioning** to **Thick Provision Eager Zeroed** and select the **Virtual Device Node** to **New SCSI controller**.
- On the 2nd virtual machine, click the **Add New Device** with **Existing Hard Disk**.
- Set the Windows Cluster Parameter **QuorumArbitrationTimeMax** to 60.

INFORMATION

About **QuorumArbitrationTimeMax**, see the following documentation for more details.

- QuorumArbitrationTimeMax

2.4. Not Supported for WSFC Setups

The following environments and functions are not supported for WSFC setups with vSphere 7. Some of which are related to arrays.

- Using VMDKs on NFS datastore as a shared disk resource for WSFC.
- Increasing the size of a shared disk.

The second point means that the datastore cannot be expanded online. Therefore users must plan enough capacity for the application.

3. PERFORMANCE RESULTS

This chapter provides performance results for building a WSFC using traditional RDMs and clustered VMDKs.

3.1. RDM vs. Clustered VMDK Performance Results

We verified the random IOPS and throughput through the IOmeter benchmark too. This is the environment and results.

Test Equipments and Configurations

- Storage
  - Model: XCubeSAN XS5224D
    Memory: 8GB per controller
    Firmware: SANOS 2.0.4
    Host Card: 2 x 4-port 16Gb FC (SFP+)
    SAS HDD: 3 x Seagate NL-SAS HDD 12.0 Gb/s 1TB
  - Pools: 2 x (3 x HDDs per Pool for RAID 5)
  - Volumes: 1 x 100GB in Pool 1 (Ctrl 1)
  - Volume Block Size: 512 Bytes
- Server
  - Model: 2 x ASUS RS700
  - 16Gb FC HBA: Marvell QLogic QLE2672
  - OS: VMware ESXi 7.0U2
- Software
  - VMware vSphere 7.0U2
  - VM OS: 2 x Windows Server 2016 with Failover Cluster feature
- IOmeter
  - Version: 1.1.0
  - Workers: 1
  - Queue Depth: 128

Test Results

The following are the performance results of random 4K and sequential 32K.

Table 3-1 Performance Results of RDM and Clustered VMDK

	RDM	CLUSTERED VMDK
Random 4K	Read IOPS: 117K Write IOPS: 26.7K	Read IOPS: 125K Write IOPS: 22.6K
Sequential 32K	Read Throughput: 838 MB/s Write Throughput: 749 MB/s	Read Throughput: 1,557 MB/s Write Throughput: 1,307 MB/s

The random 4K in the table doesn't make much difference, but in terms of throughput, using clustered VMDK is better than RDM.

4. CONCLUSION

In ESXi 7.0, the WSFC configuration supports clustered VMDK on VMFS datastore. It simplifies the process of VM application environment. QSAN storage follows in VMware's footsteps, supports clustered VMDKs, and is well tested. In terms of performance results, using clustered VMDK also has better throughput behavior. QSAN storage is an ideal solution for building Windows cluster architecture. Although there are some limitations for using VMDK, users can avoid them when planning.

5. APPENDIX

5.1. Apply To

- XEVO firmware 2.1.3 and above.
- SANOS firmware 2.0.4 and above.

5.2. Reference

Setup WSFC Documents

- Setup for Windows Server Failover Clustering

About VMware RDM

- About Raw Device Mapping
- Differences between Virtual and Physical RDM

About Clustered VMDK

- Setup Windows Server Failover Cluster with Clustered VMDKs on vSphere 7 with Hitachi VSP Series
- VMware Clustered VMDK, SCSI3-PR and WEAR