Hedvig

Hedvig Virtualized Storage for Microsoft Hyper-V Clusters
User Guide

Table of Contents

Overview of Virtualized Storage

Microsoft's Hyper-V technology has gained significant traction in many virtualized computing environments — from small or medium businesses to large enterprise companies.

By building its virtualization platform directly into Windows Server, Microsoft has fostered Hyper-V adoption and provided an easy way to create and manage a virtualized computing environment.

One of the advantages of moving to a hypervisor-based compute environment is the ability to form a highly available — or *failover* — cluster that supports application resiliency. A prerequisite for this type of cluster is that shared storage be accessible to all hosts in the cluster. Storage platforms that deliver seamless high availability and built-in DR across multiple sites complement a Microsoft Hyper-V Failover Cluster.

Hedvig's scale-out distributed storage platform, with real-time multi-site replication, provides protection of data and delivers maximum availability in the event of server or data center outages. Hedvig's hyperscale architecture is ideal for Microsoft Hyper-V virtualization deployments, providing highly elastic and flexible storage that can be scaled independently of compute.

See also https://www.commvault.com/software-defined-storage.
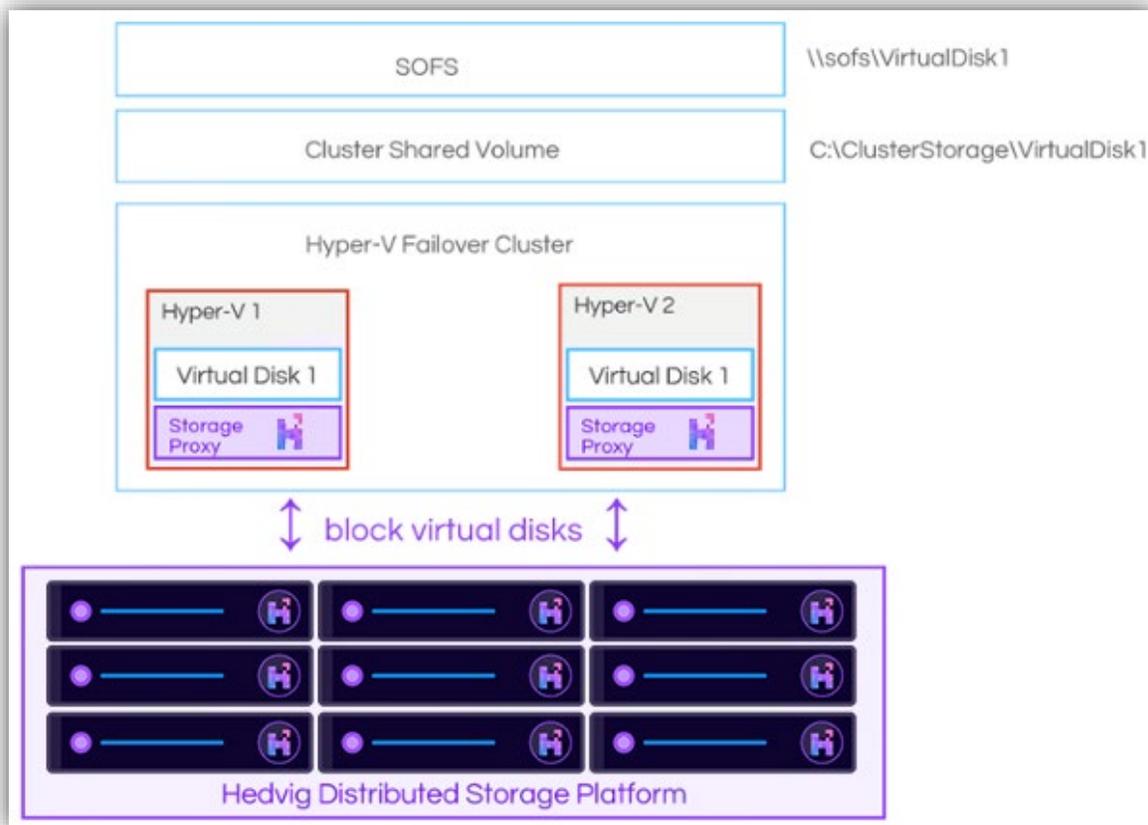
Microsoft Hyper-V reference architecture

Figure 1: Hedvig + Hyper-V reference architecture

Microsoft Windows Server 2012 ships with a multitude of server roles and technologies. Many Windows Servers running Hyper-V roles can be pooled together to form a highly available Hyper-V cluster by leveraging the Windows *Failover Clustering* feature. The combination of virtualization and Failover Clustering provides continuous access to applications and services, thereby eliminating any single point of failure.

Figure 1: Hedvig + Hyper-V reference architecture illustrates the Hedvig Distributed Storage Platform operating with a Hyper-V Failover Cluster. Hedvig's scale-out storage software installs on standard x86 servers (also known as *Hedvig Cluster Storage Nodes*) and can span a storage cluster across multiple data centers and/or clouds that network together to form a storage resource pool. The compute tier (Hyper-V hosts) accesses the storage resource pool through a Hedvig software client called the *Hedvig Storage Proxy.*

The storage proxy is a component that resides completely in the compute tier. In this reference architecture, each Hyper-V host has its own storage proxy running as a VM. The storage proxy masquerades as an iSCSI target, providing storage to the corresponding Hyper-V host in the form of one or more block devices. The storage proxy is completely stateless and runs as an HA pair.

Cluster Shared Volumes (CSV)

Cluster Shared Volumes (CSV) is a Windows Failover Clustering feature that provides a layer of abstraction between Hyper-V applications and underlying storage, thereby allowing multiple nodes in the Failover Cluster simultaneous access to the same disk (LUN) provisioned as an NTFS volume.

CSV enable quick failover between nodes without having to unmount and remount a volume.

CSV simplify the management of many disks by using a consistent file namespace (for example, `C:\ClusterStorage\VolumeX`) across each of the storage cluster nodes in the Failover Cluster.

CSV ownership is managed by Failover Cluster nodes through *SCSI-3 Persistent Reservations* (SCSI-3 PR).

Since each storage cluster node accesses storage resource through its own local storage proxy, which masquerades as an iSCSI target, SCSI-3 PR disks are tracked at the storage *cluster* level, as opposed to at the storage *proxy* level.

CSV ownership in Windows Server 2012 R2 is automatically distributed and rebalanced across the Failover Cluster nodes.

Hyper-V Cluster Setup with Hedvig

Prerequisites for Hyper-V Clusters with Hedvig

- You must have already set up:

  - Hyper-V Failover Cluster

  - Hedvig Cluster

  - Hedvig Storage Proxies running as HA pairs (one pair per Hyper-V host)

- Run the CSV configured with Redirected Access.

- If the Failover Cluster has an odd number of nodes in the cluster, there is no need to configure a quorum disk. It is preferred that the quorum disk be on a distinct share. This is because the quorum disk needs to be seen by all Hyper-V nodes at all times. If the quorum disk resides on Hedvig, inactive iSCSI connections to the quorum disk can result in inaccessibility to the CSV, potentially resulting in a file system-level inconsistency within the `.vhdx` files.

- Hyper-V CSV are very sensitive to iSCSI timeouts. Do not leave this value at the default — set it to 140 seconds.

For a Hyper-V Failover Cluster quick start:

- Set up a Failover Cluster on Windows 2012/2016 using normal best practices. For reference, see:

  https://thesolving.com/virtualization/how-to-set-up-a-hyper-v-failover-cluster/

- A quorum should ideally not be used for an odd number of nodes. Also, for an even number of nodes, it is not recommended that you use a quorum disk on Hedvig.

- Hyper-V networks should be 10G+ or higher.

- On all Hyper-V nodes, set the following registry property (which requires that the MPIO feature be installed on the Hyper-V nodes). A reboot will be required for this.

  ```
  HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\mpio\Parameters\PDORemovePeriod = 130
  ```

- A storage proxy pair is deployed on each Hyper-V node, using the storage proxy `.vhdx` file imported into Hyper-V. The storage proxy must be configured with 4vcpu and 8g of memory at a minimum, and a secondary SSD cache disk of at least 300GB.

- After the Failover Cluster is up, add disks configured with the CSV property and client-side caching, and expose them to the Hyper-V nodes. For CSV disks, addlun and addaccess must be done to all storage proxy pairs in the Hyper-V cluster

- Scan the disks using the Microsoft iSCSI initiator, and bring them online on one Hyper-V node first, and format the disk. Keep the disks offline on the other nodes during this time.

- In the Failover Cluster manager, add the disks to available storage. The Failover Cluster will manage disk ownership from now on.

- After the disks are added to available storage, right-click on the disk, and convert it to CSV (cluster shared volumes).

- After CSV are available, convert them to Redirected Access by following these instructions:

  In the Windows Server Failover Cluster Manager, expand **Storage**, select **Disks**, highlight/select your cluster disk in the upper pane, then highlight/select the CSV volume in the lower pane, right-click the CSV volume and select **Turn On Redirected Access.**

  For reference, see:

  https://serverfault.com/questions/613404/turn-on-redirected-access-mode-for-server-2012r2-cluster-shared-volume

Setting up Hyper-V Clusters with Hedvig

1. Create a virtual disk of type **BLOCK** with the **CSV** (cluster shared volumes) option enabled.
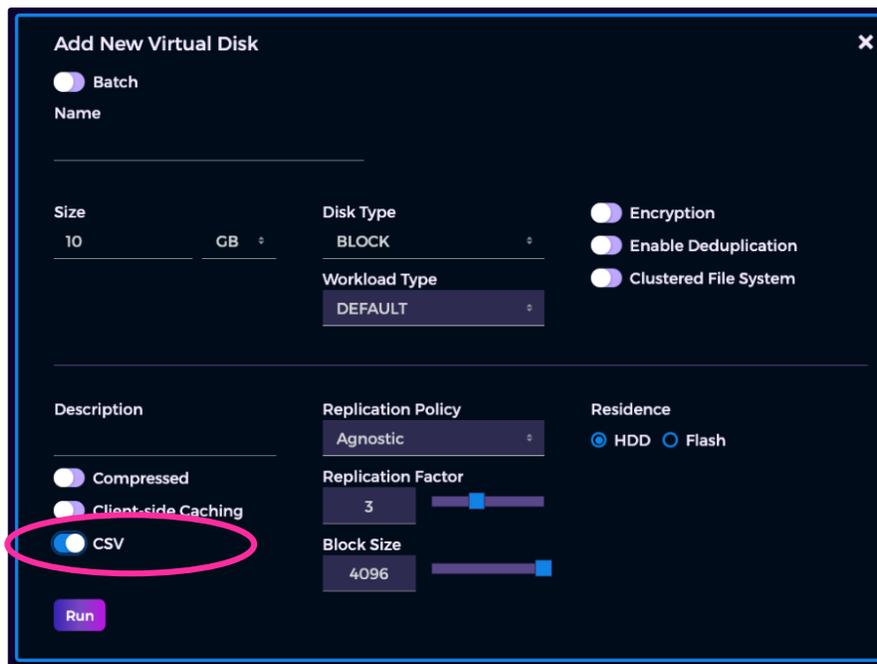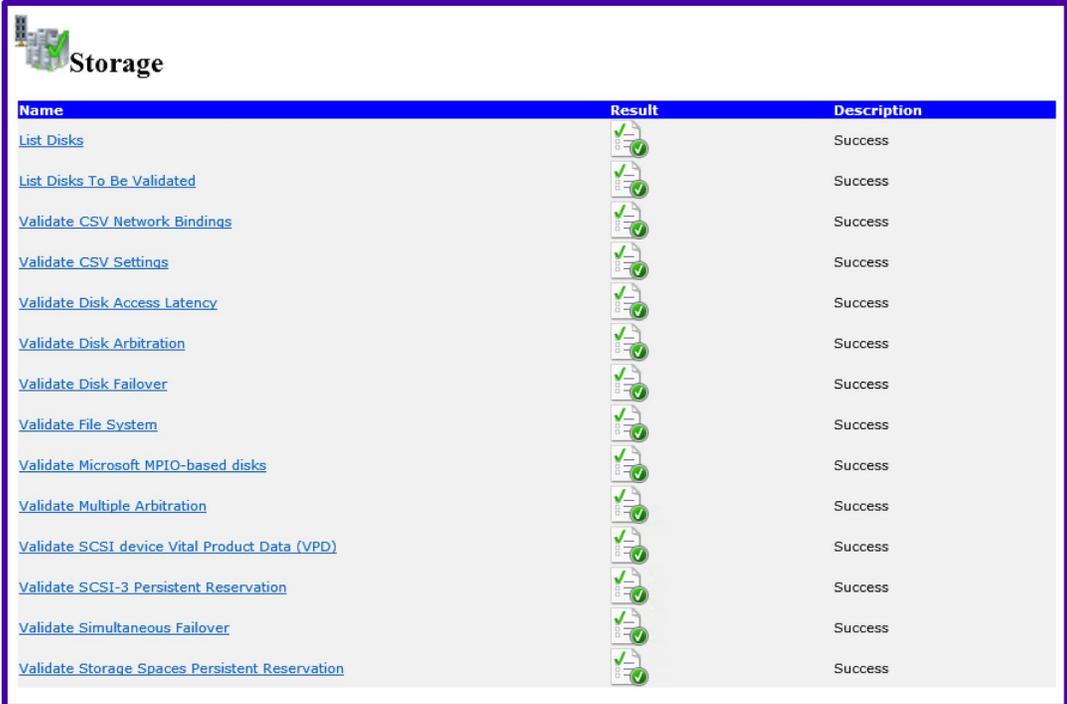

Figure 2: Creating a block virtual disk with the CSV option

2. Add this virtual disk as a LUN on all Hedvig Storage Proxies running on Hyper-V hosts. This is necessary because each Hyper-V host will access the same virtual disk using its local storage proxy.

3. Update the ACL information on the storage proxies so that each Hyper-V host can consume the provisioned virtual disk as an iSCSI LUN only from its local storage proxy, which masquerades as an iSCSI target.

4. Using the Microsoft iSCSI initiator on each Hyper-V cluster node, login to the iSCSI target to connect to the provisioned virtual disk. After the disks are connected, bring these disks online through the Disk Management console.

5. Format the virtual disk with NTFS on any one of the Hyper-V cluster nodes. You do not need to mount them or assign a drive letter after formatting.

6. Run the Failover Cluster Validation tests. This is necessary because before you can assign the NTFS volume to clustered roles, it must pass the Failover Cluster Validation tests to verify that the underlying storage array supports Failover Cluster operations.

Figure 3: Successful Failover Cluster Validation Tests

7. After all validation tests have passed, add the NTFS volume to the Failover Cluster, where it will appear as a clustered volume.

8. Add this clustered volume to a CSV to enable a seamless and quick failover between the Hyper-V cluster nodes. When you add a clustered volume to a CSV, you are asked to choose a filesystem path where it will be mounted on each Hyper-V cluster node. The default path for a CSV is:

```
C:\ClusterStorage\<VolumeName>
```

Glossary

This glossary contains definitions of terms used in this document.

Table 3: Glossary of terms

term	definition
ACL	An *access control list* is a list of permissions attached to an object.
CSV	*Cluster Shared Volumes* is a feature of *Failover Clustering* (or *HA* clustering) first introduced in Windows Server 2008 R2 for use with the Hyper-V role. A CSV is a shared disk containing an NTFS or ReFS (ReFS: Windows Server 2012 R2 only) volume that is made accessible for read and write operations by all nodes within a Windows Server *Failover Cluster.*
DR	*Disaster recovery* involves a set of policies and procedures to enable the recovery or continuation of vital technology infrastructure and systems following a natural or human-induced disaster.
failover	*Failover* is the ability of applications to migrate from one server to another if a server fails, so that these applications can continue to operate without interruption.
Failover Cluster	A *Failover Cluster* (also known as an *HA cluster*) is a group of computers that support server applications that can be reliably utilized with a minimum amount of down-time.
HA	*High availability* is the ability of a system to continue to function after the failure of one or more of its servers.
Hedvig Storage Cluster	A *Hedvig Storage Cluster* is an elastic cluster, formed by using any type of commodity server(s).
Hedvig Storage Cluster Node	A *Hedvig Storage Cluster Node* is an individual commodity server running *Hedvig Storage Service* software.
Hedvig Storage Proxy	A *Hedvig Storage Proxy* is a lightweight software component that deploys at the application tier as a virtual machine or Docker container, or on bare metal, to provide storage access to any physical host or virtual machine in the application tier. The storage proxy presents block, file, and object (Amazon S3) storage access to app hosts, accelerates read performance with flash caching, drives efficiency with deduplication, and secures data with encryption. *This may also be referred to as an HSP, controller, CVM, target, or tgt.*

term	definition
hyperscale	In a *hyperscale* architecture, compute and storage scale independently of each other.
hypervisor	A *hypervisor* or *VMM (virtual machine monitor or virtual machine manager)* is a piece of computer software, firmware, or hardware that creates and runs VMs. A computer on which a hypervisor runs one or more VMs is called a *host* machine, and each VM is called a *guest* machine. The *Red Hat Virtual Machine Manager* is also known as *virt-manager.*
Hyper-V	*Microsoft Hyper-V,* formerly known as *Windows Server Virtualization*, is a native hypervisor that can create VMs on x86-64 systems running Windows.
iSCSI	*Internet Small Computer System Interface* is an IP-based storage networking standard for linking data storage facilities.
LUN	A *logical unit number* is a number that identifies a logical unit, which is a device addressed by the SCSI protocol or SAN protocols, which encapsulate SCSI, such as Fibre Channel or iSCSI.
NTFS	*New technology file system* is a proprietary file system developed by Microsoft.
RDBMS	A *relational database management system* is a database management system that is based on the relational model as invented by E. F. Codd, of IBM's San Jose Research Laboratory.
RDSMS	A *relational data stream management system* is a distributed, in-memory DSMS (data stream management system) that is designed to use standards-compliant SQL queries to process unstructured and structured data streams in real-time.
ReFS	*Resilient file system* is a Microsoft proprietary file system introduced in Windows Server 2012 with the intent of becoming the next generation file system after NTFS.
SAN	A *storage area network* is a network that provides access to consolidated, block-level data storage.
SCSI	*Small Computer System Interface* is a set of standards for physically connecting and transferring data between computers and peripheral devices.

term	definition
SCSI reservations	*SCSI reservations* are used to control access to a shared SCSI device such as a disk or tape drive. An initiator sets a reservation on a LUN to prevent another initiator from making changes to the LUN. This is similar to the file-locking concept. SCSI reservations are always set by a host initiator. Ideally, the same initiator would perform a SCSI release on the affected LUN. See also *SCSI-3 PR.*
SCSI-3 PR	*SCSI-3 Persistent Reservations* add the ability for a SCSI reservation to persist even if the bus is reset for error recovery. See also *SCSI reservations.*
SMB	*Server Message Block* operates as an application-layer network protocol, primarily used for providing shared access to printers, etc.
SOFS	A *Scale-Out File Server* is a Windows Failover Clustering feature that provides continuously available scale-out file shares for file-based server application storage.
SQL	*Structured query language* is a special-purpose programming language designed for managing data held in an RDBMS or for stream processing in an RDSMS.
target	See *Hedvig Storage Proxy.*
VHD, VHDX	Hyper-V in Windows Server 2012 introduced a new version of the *virtual hard disk* (VHD) format called *VHDX,* which has a much larger storage capacity than the older VHD format.
virtual disk	A *virtual disk* is an abstracted logical disk volume presented to a computer or application for read/write use.
VM	A *virtual machine* is a software computer that, like a physical computer, runs an operating system and applications.
VMM	See *hypervisor*.

Software-defined AES-256, FIPS compliant encryption of data in flight and at rest