



Hedvig as backup target for Veeam

Solution Whitepaper

Version 1.0

April 2018

Table of contents

Executive overview	3
Introduction	3
Solution components	4
Hedvig	4
Hedvig Virtual Disk (vDisk)	4
Hedvig Storage Service	4
Hedvig Storage Proxy.....	5
Veeam	5
Backup server.....	5
Backup proxy.....	6
Backup repository.....	6
Integrating Hedvig with Veeam	7
Create Hedvig virtual disk	7
NFS.....	7
Block.....	7
Create Veeam backup repository	11
Linux server.....	11
Microsoft windows server	13
Hedvig specific settings for Veeam	15
Backup repository settings.....	15
Backup proxy settings.....	15
Backup Job settings	15
Hedvig deployment.....	16
Hyperscale.....	16
Hyperconverged	16
Cloud support	16
Cloud replica using Veeam's Backup Copy.....	16
Storage native cloud replica.....	17
Cloud native backup	17
Summary and conclusion	17
Additional resources	17

Executive overview

Digital transformation has exposed the legacy roots of backup and has magnified backup's stodgy pace compared to the more agile technologies that cloud and distributed computing systems provide to enterprises today. All modern applications are being designed with high-availability and cloud compatibility in mind, and most storage solutions are struggling to keep up.

The Hedvig Distributed Storage Platform provides a seamless way to integrate with your backup applications and modernize your backup infrastructure. Hedvig significantly reduces your backup storage costs with the use of commodity hardware and advanced storage reduction algorithms, enables a pay-as-you-grow model, and provides hassle-free public cloud support. Hedvig removes gravity from your backup data that is generally associated with traditional storage solutions and enables data movement across racks, data centers, and public clouds.

Introduction

This document outlines how Hedvig can be used as a backup target for Veeam Backup & Replication™. Veeam® delivers availability for all workloads — virtual, physical and cloud including support for VMware vSphere and Microsoft Hyper-V — from a single management console. It provides fast, flexible, and reliable recovery of your applications and data, and brings backup and replication together into a single software solution.

With the combined solution of Veeam and Hedvig, enterprises can achieve:

- Lower costs using commodity hardware and advanced storage reduction algorithms
- Ease of consumption with intuitive interfaces and instantaneous storage provisioning
- On-demand, pay-as-you-grow backup storage
- Hybrid cloud and multi-cloud deployments
- Integration with current workflows using the Hedvig REST API

Solution components

Hedvig

The Hedvig Distributed Storage Platform is a software-defined storage solution that lets you tailor an elastic storage system built with low-cost commodity server hardware to support any application, hypervisor, container, or cloud. Hedvig is built with cloud-like provisioning and simplicity for infrastructure-as-a-service (IaaS) and plugs seamlessly into cloud orchestration solutions to deliver automated, on-demand storage provisioning at any scale.

Hedvig Virtual Disk (vDisk)

The virtual disk is the fundamental abstraction unit of the Hedvig Distributed Storage Platform. Organizations can spin up any number of concurrent vDisks — each thinly provisioned and instantly available. A series of user-configurable attributes can be set when provisioning a vDisk (Figure 1). These attributes are explained in detail later in this document.

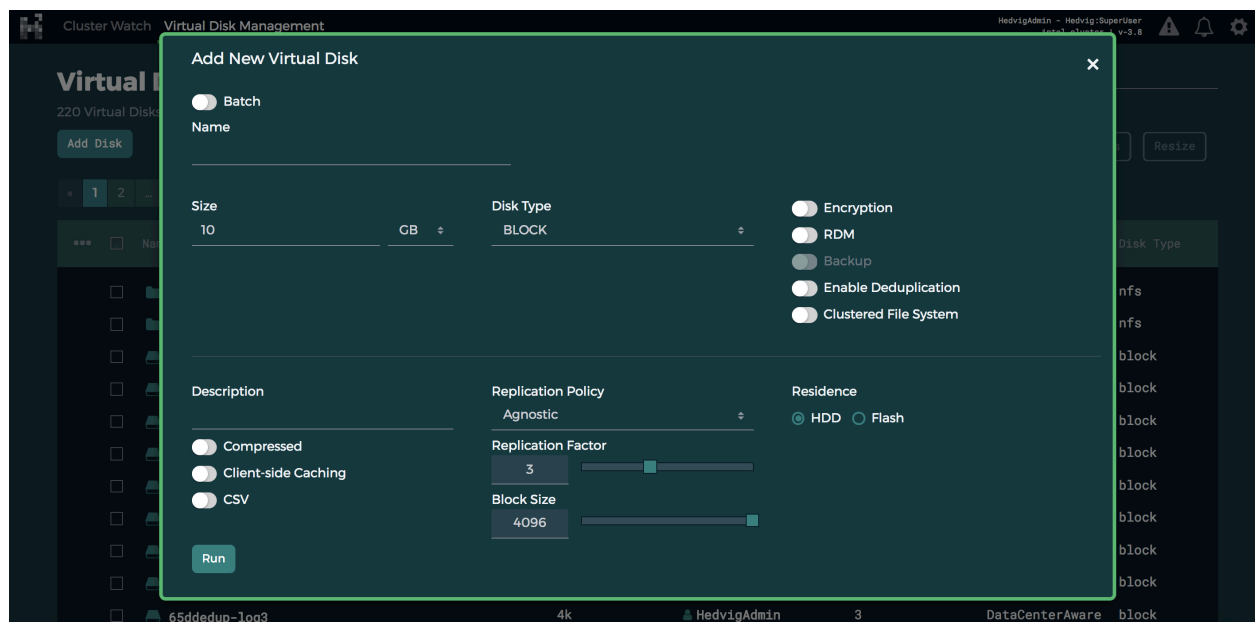


Figure 1

Hedvig Storage Service

The Hedvig Storage Service is Hedvig's patented distributed systems engine. It installs on commodity x86 servers or cloud instances and transforms existing server and storage assets —

including SSD/flash media and hard disk — into a full-featured elastic storage cluster. The software deploys to an on-premises infrastructure or to hosted or public clouds to create a single storage cluster that is implicitly hybrid.

Hedvig Storage Proxy

The Hedvig Storage Proxy (HSP) 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 HSP provides intelligent access to the hyperscale storage nodes, directing I/O requests to the relevant backend storage nodes based on latency response times. The HSP runs in the user space and can be managed by any virtualization management and orchestration tool. It acts as a protocol (iSCSI/NFS/S3/Swift) converter, load balances I/O requests to the backend storage nodes, and provides edge caching using local flash (SSD/NVMe) devices to optimize storage performance directly at the application hosts. It also caches data fingerprints for dedupe-enabled vDisks and eliminates the transmission of duplicate data over network links.

Veeam

Veeam Backup & Replication is a backup solution developed for VMware vSphere and Microsoft Hyper-V virtual environments. Veeam Backup & Replication provides a set of features for performing data protection and disaster recovery tasks. Here are the key components of the Veeam backup infrastructure:

Backup server

The Veeam backup server is a Windows-based physical or virtual machine on which Veeam Backup & Replication is installed. It is the core component in the backup infrastructure that fills the role of the “configuration and control center.” The backup server performs the following types of administrative activities:

- Coordinating backup, replication, recovery verification and restore tasks
- Controlling job scheduling and resource allocation

- Setting up and managing backup infrastructure components, as well as specifying global settings for the backup infrastructure

Backup proxy

A Veeam backup proxy is an architecture component that sits between the backup server and other components of the backup infrastructure. While the backup server administers tasks, the backup proxy processes jobs and delivers backup traffic. Basic backup proxy tasks include the following:

- Retrieving VM data from production storage
- Compressing VM data
- Deduplicating VM data
- Encrypting VM data
- Sending VM data to the backup repository (for example, if you run a backup job) or to another backup proxy (for example, if you run a replication job)

Backup repository

A Veeam backup repository is a storage location where Veeam keeps backup files, VM copies, and metadata for replicated VMs. To configure a backup repository, you can use the following servers or storage systems:

- Microsoft Windows server
- Linux server
- CIFS (SMB) share
- Backup repositories with rotated drives

Veeam information sourced from 'User Guide for Veeam VMware vSphere':

<https://helpcenter.veeam.com/docs/backup/vsphere/overview.html?ver=95>

Configuring Hedvig with Veeam

After a Hedvig Storage Cluster is deployed, two steps are involved in consuming a Hedvig vDisk as a Veeam backup repository. First, you need to create a Hedvig vDisk as per your backup requirements, using the Hedvig interface. Second, you need to set up a Veeam backup repository and point it to the Hedvig vDisk.

Create Hedvig virtual disk

You can select from 2 disk types when creating a Hedvig vDisk.

NFS

We recommend using Hedvig's NFS disk type for Veeam. Create a thinly provisioned Hedvig vDisk of required size with 'NFS' as disk type, and select the 'Backup' option (Figure 2). The 'Backup' option is needed to use Hedvig as a backup target in order to utilize the following optimizations:

- **Increased write throughput:** Hedvig utilizes Veeam's property of a dedicated backup proxy for a given backup job and makes sure there is no overhead of synchronizing concurrent writes to the same backup file(s) across multiple backup proxies. Also, Hedvig leverages the WORM (Write Once Read Many) property of backup data and makes sure data is instantly available to restore (if needed) across multiple backup proxies.
- **Efficient restore speed:** Hedvig provides an aggressive pre-fetch mechanism for VM restores. Any flash disk can be provided to the Hedvig Storage Proxy installed in the same hypervisor as the Veeam backup proxy. This will make sure that Hedvig utilizes the power of the distributed cluster, and large chunks of data will be restored at local SSD speed.

Block

You can also create a block vDisk for a backup repository (if needed) by selecting 'block' as disk type during Hedvig's vDisk creation. A Block vDisk can be used to provide a directly attached volume to your repository server.

During vDisk creation, you also have multiple optional parameters to choose from, based on your backup requirements:

- a) **Deduplication:** Backup data predominantly provides high deduplication ratios, and generally it's a good idea to select this option. This will not only make sure that all VM backups are deduplicated in any given repository, it will also deduplicate backup data across all of Hedvig's dedupe-enabled backup repositories. Also, Hedvig recommends disabling Veeam's "inline data deduplication" in order to take full advantage of Hedvig's "global deduplication."

Another option that you will be asked to select, along with deduplication, is known as 'Retention Policy' (Figure 2). You can select any value from 'TWO_WEEKS' to 'SIX_MONTHS,' and Hedvig will optimize the garbage collection of expired data that is being referenced by multiple VM backups.

Compression is always enabled by default when the deduplication option is selected. Compression has minimal overhead in Hedvig due to the execution of compression algorithms in each of the individual storage nodes before persisting to disk.

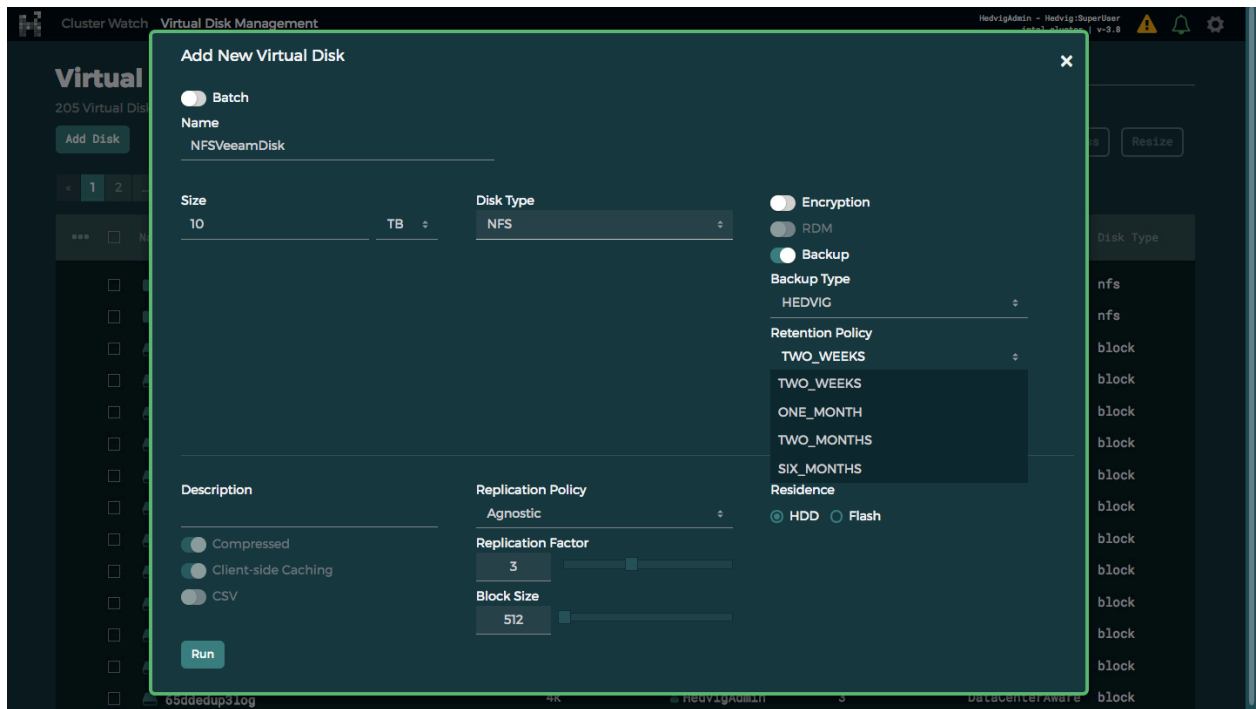


Figure 2

- b) **Compression:** This option can be used even if you don't want to select the deduplication feature. You might want to explicitly select this option if the dataset is not dedupe-friendly, for instance if you are trying to back up data that is already encrypted or exists in binary format, or if your source application always generates unique data, such as timestamped system log data.
- c) **Client-side caching:** This option allows you to increase restore performance by aggressively pre-fetching data and enabling local SSD speed for the majority of read operations.
- d) **Replication factor:** You can choose to generate 1-6 copies of backup data depending on your data protection requirements. Data is striped at a granularity such that if you choose to generate 'n' copies of data, data will be spread to more than just 'n' storage nodes, so that you can utilize the full potential of the distributed cluster for your backup and restore performance. Hedvig will make sure that copies of your data reside in different storage

nodes and that applications can withstand failures, such as bad disk, storage node outage, and network interruptions.

- e) **Replication policy:** This option gives you control over the residence of all copies of your backup data selected using ‘replication factor.’ There are 3 options to choose from (Figure 3):
1. **Agnostic:** Hedvig dynamically figures out the best set of storage nodes for your replicated data, based on cluster capacity distribution.
 2. **Data Center Aware:** The Hedvig Storage Cluster can span multiple data centers, and you can select the data center(s) on which your backup data copies should reside. A data center can be a group of physical x86 storage servers, virtual storage nodes, or cloud (AWS/Azure/GCP) instances of storage nodes. This enables hybrid-cloud and multi-cloud installations for your backup applications.
 3. **Rack Aware:** If you have multiple racks in your data centers, you can select this option make sure copies of data reside in different storage nodes across multiple racks.

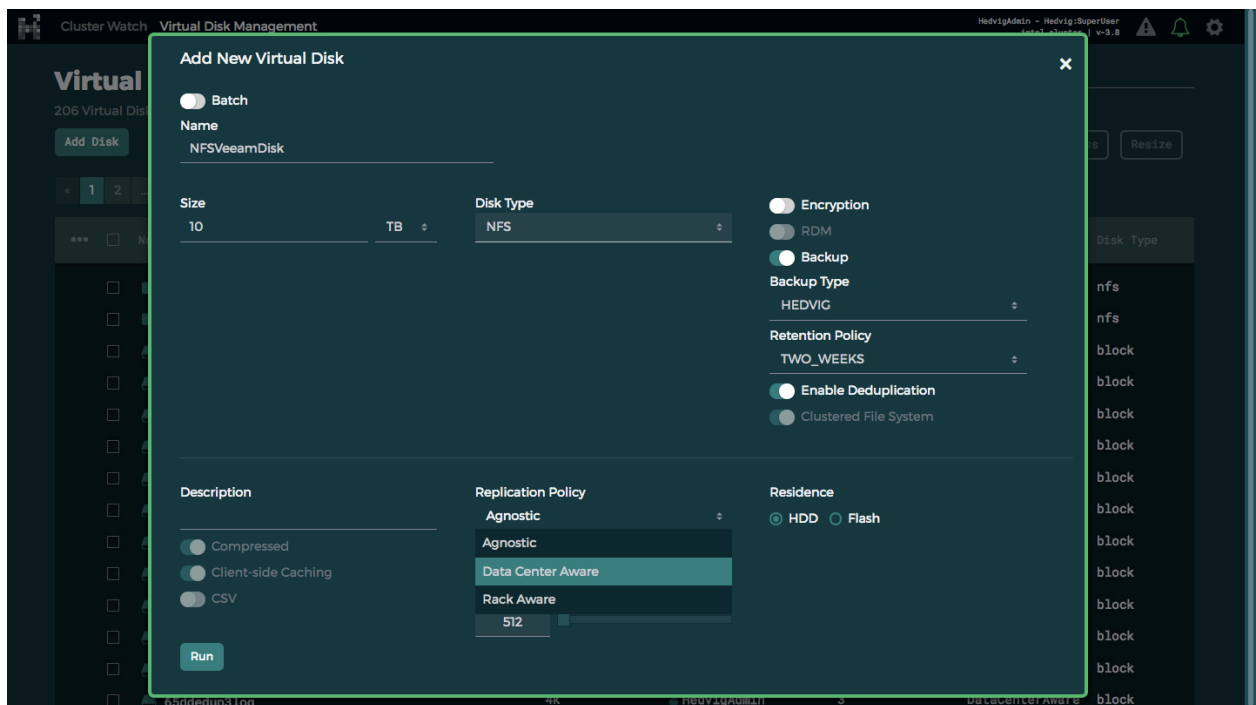


Figure 3

- f) **Residence:** If you have a mixed cluster with HDDs as well as SSDs available to the storage nodes, you can choose to pin certain volumes' data to SSD and make sure critical backup jobs are prioritized over others.
- g) **Encryption:** You can enable end-to-end security for your backup jobs by selecting this option. Hedvig makes sure data is encrypted in-flight, in-use, and at-rest, and encrypted data is decrypted only when requested by the backup application.

Create Veeam backup repository

Hedvig supports NFS as well as block vDisk types for creating a Veeam backup repository. Veeam currently does not support adding a locally mounted NFS volume as a backup repository. There are two Veeam backup repository types that are supported with Hedvig:

Linux server

1. NFS (recommended): Create Hedvig vDisk with NFS as disk type and mount it to a Linux server (Figure 4):

```
mount -t nfs -o nolock,hard,intr,nfsvers=3
-o rw <HedvigProxyServer>:/exports/HedvigNFSDisk <Path to local directory>
```

```
[root@poc-lix1 ~]# mount -t nfs -o nolock,hard,intr,nfsvers=3 -o rw blacks3cvm.external.hedviginc.com:/exports/NFSVeeamDisk /mnt/NFSVeeamDisk/
[root@poc-lix1 ~]# df -kh
Filesystem      Size  Used Avail Use% Mounted on
/dev/mapper/vg_bonesclnt4-lv_root
26G  7.9G  17G  32% /
tmpfs            1.9G   0  1.9G   0% /dev/shm
/dev/sda1        477M   63M  389M  14% /boot
/dev/sdb         16G   33M   16G   1% /hedvig1
blacks3cvm.external.hedviginc.com:/exports/NFSVeeamDisk
10T  8.0M  10T   1% /mnt/NFSVeeamDisk
[root@poc-lix1 ~]# hostname
poc-lix1.external.hedviginc.com
[root@poc-lix1 ~]#
```

Figure 4

Create a new repository server and provide Linux host credentials. After you click 'populate,' you will be able to see the mounted Hedvig volume (Figure 5) and consume it as an externally mounted NFS volume.

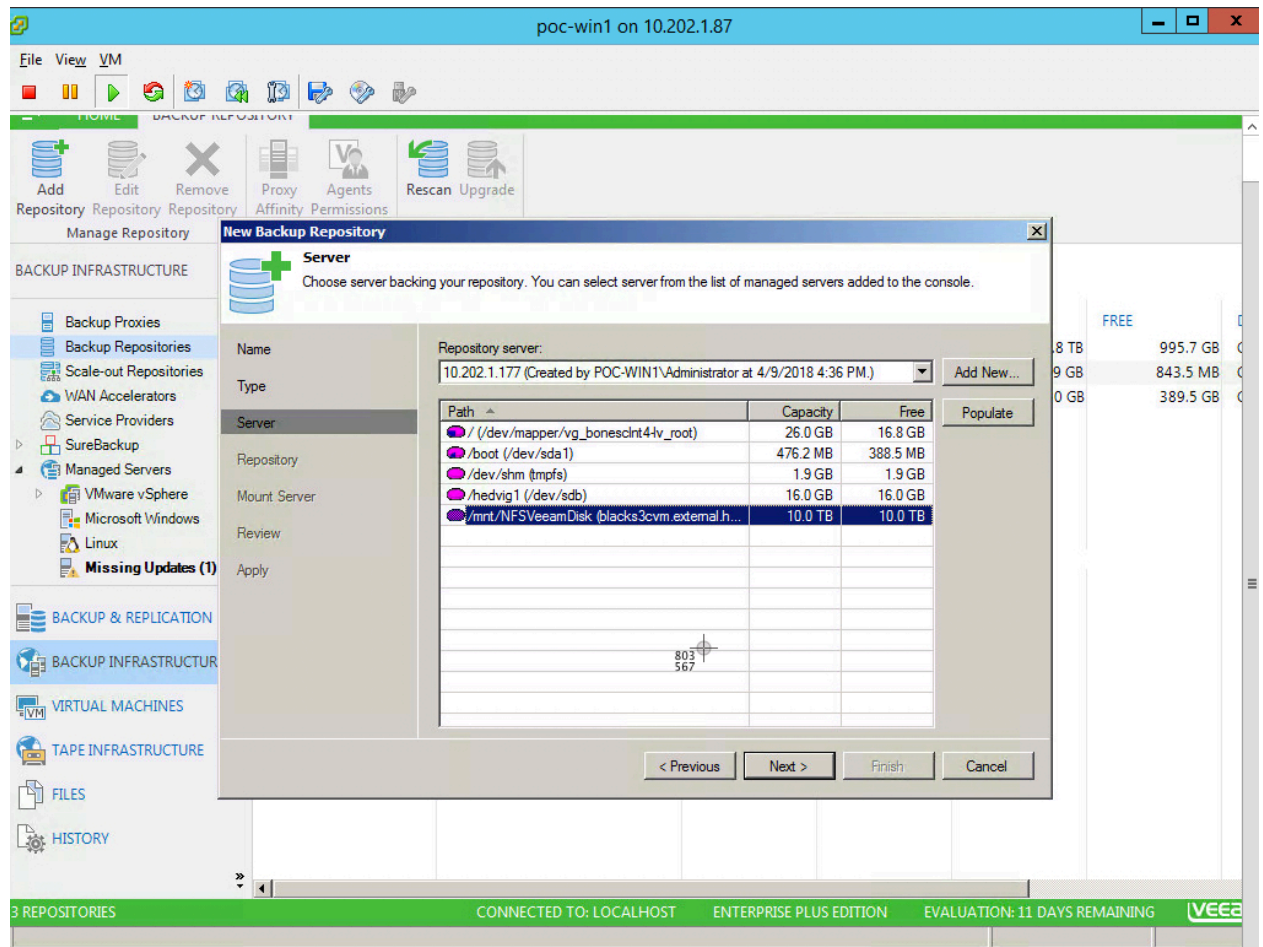


Figure 5

- Linux: Create a Hedvig vDisk with 'block' type and provide access to any Linux host. Create a filesystem on this disk, mount it to a local directory, and use this remote Linux volume as a backup repository.

Microsoft windows server

1. NFS: If the Veeam repository server is set up as a virtual machine, the Hedvig NFS vDisk can be mounted as an NFS datastore at the hypervisor layer (Figure 6). A new disk can be carved out of this datastore and provided to the Veeam repository server, which in turn can create a filesystem and present it as a directly attached volume to the backup repository (Figure 7).

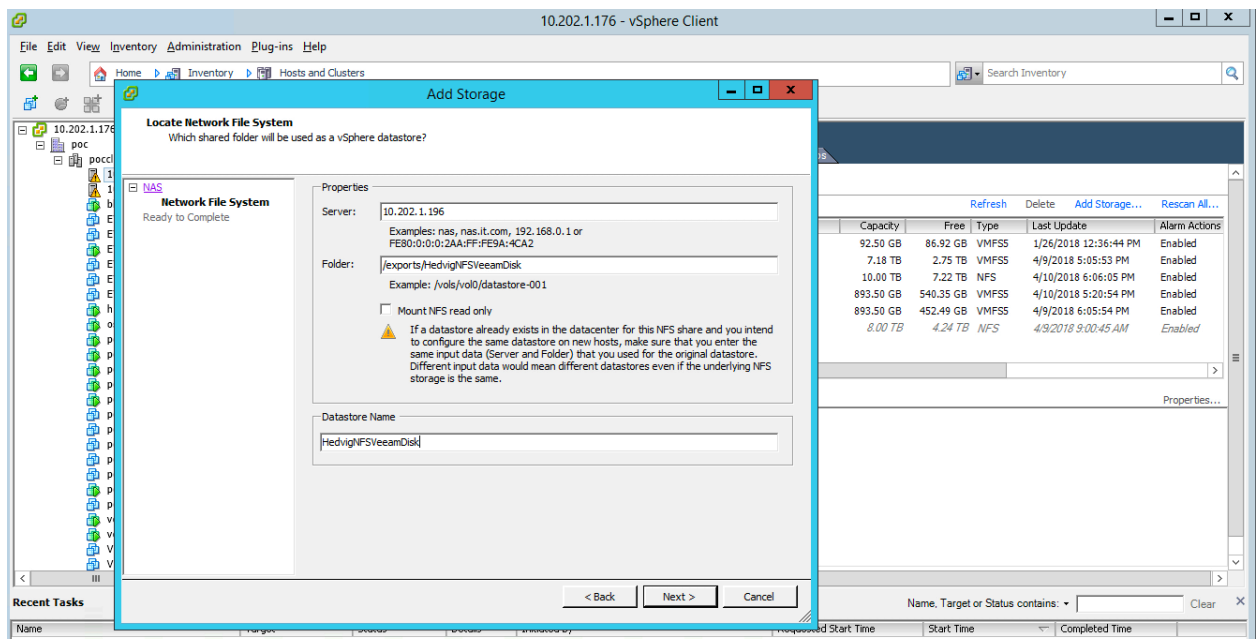


Figure 6

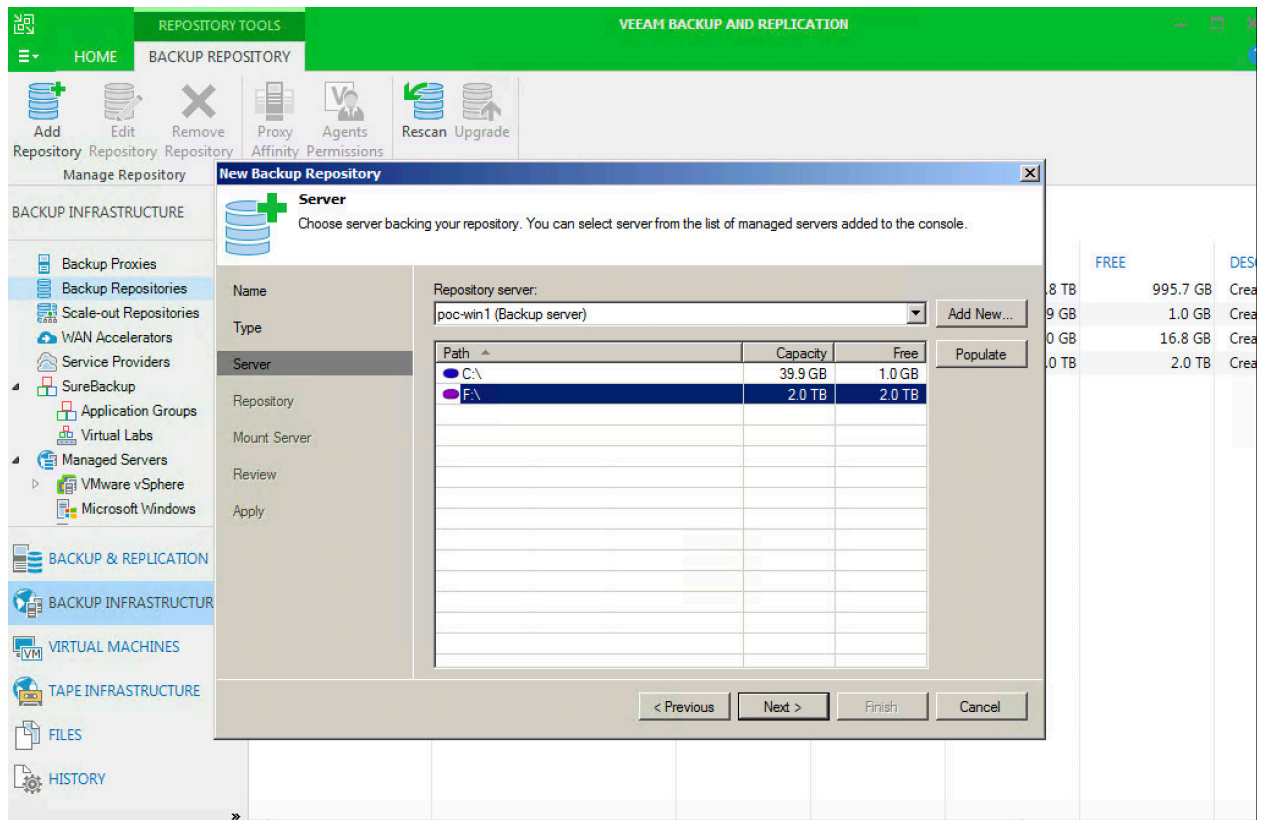


Figure 7

2. Block: Create a Hedvig virtual disk with 'block' type, and provide your Veeam repository server access to this vDisk. You can discover this vDisk in your repository server using Windows "iSCSI Initiator," create a filesystem, and then present it as a directly attached volume. This will work for physical as well as virtual backup repository servers.

Here are a few additional things to keep in mind:

1. The Veeam repository server host could be the same as, or different from, the Veeam backup server host.
2. You are free to choose physical or virtual Veeam backup server(s) and repository server(s).
3. There are no additional Hedvig resource requirements for Veeam components.
4. Hedvig volumes are highly available, and no additional intervention is needed with a reboot of Veeam components.

5. If you need to delete a Hedvig vDisk, you should first remove all references to that vDisk from Veeam backup repositories. Then, you can use the Hedvig interface to physically delete a vDisk.

Hedvig specific settings for Veeam

Backup repository settings

1. **Use per-VM backup files:** Use this option if you are utilizing Hedvig's Linux-mounted NFS volume. This will enable increased parallelism for concurrent read/write operations of multiple files.
2. **Limit maximum concurrent tasks to:** Maximum value can be set to 64.

Backup proxy settings

Max concurrent tasks: Maximum value can be set to 64.

Backup Job settings

1. **Enable inline data deduplication:** Set this option to false, in order to maximize Hedvig's deduplication potential. Hedvig provides inline global deduplication and significantly reduces the amount of data transferred from the repository server to backend storage, and also deduplicates data across multiple backup proxies and repository servers.
2. **Compression level:** Select 'Optimal' if Hedvig's deduplication is disabled. Otherwise, select 'Dedupe-friendly'.
3. **Storage optimization:** Select 'local target' since Hedvig takes care of optimizing data transferred from the repository server to backend storage.

All other Veeam options can be selected as per Veeam recommendations.

Hedvig deployment

The Hedvig Distributed Storage Platform components can be configured to support two types of deployments: hyperscale and hyperconverged. Hedvig provides the flexibility to leverage both in the same Hedvig Storage Cluster:

Hyperscale – Hyperscale deployments scale storage resources independently from application compute resources. With hyperscale, storage capacity and performance scale out horizontally by adding commodity servers running the Hedvig Storage Service. Application hosts consuming storage resources scale separately with the Hedvig Storage Proxy, allowing for the most efficient usage of storage and compute resources.

Hyperconverged – Hyperconverged deployments scale compute and storage in lockstep, with workloads and applications residing on the same physical nodes as data. In this configuration, the Hedvig Storage Proxy and the Hedvig Storage Service software are packaged and deployed as VMs on a compute host with a hypervisor installed. Hedvig provides plug-ins for hypervisor and virtualization tools, such as VMware vCenter, to provide a single management interface for a hyperconverged solution.

Cloud support

A Hedvig Storage Cluster can be deployed to integrate with public cloud providers in multiple ways.

Cloud replica using Veeam's Backup Copy

Create two Hedvig NFS vDisks and assign 'DataCenterAware' replication policies to both of them. Select on-prem data center(s) as replica(s) of the first NFS vDisk and cloud data center(s) as replica(s) of the second NFS vDisk. Assign the first NFS vDisk as a backup repository for your primary backup job and the second NFS vDisk for a secondary backup copy job.

Storage native cloud replica

You can also delegate the responsibility of cloud replication to Hedvig. Create a Hedvig NFS vDisk, select the 'DataCenterAware' replication policy, and assign on-prem data center(s) as well as public cloud instance(s) for NFS volume replication. You can utilize this NFS volume as a backup repository to seamlessly achieve hybrid-cloud installation natively from the storage layer.

Cloud native backup

Hedvig can be installed purely as a cloud installation. Hedvig provides multi-cloud capabilities, such as cross-region resilience and cross-cloud fault tolerance across multiple-cloud providers. You can select multiple-cloud providers, such as AWS, Azure, and GCP, for your NFS volume's replication policy, as part of the same Hedvig Storage Cluster installation.

Summary and conclusion

As enterprises increasingly adopt digital transformation — and also a cloud-first or a cloud-native approach — they are recognizing that backup is not just backup anymore. It is not the “passive” data that traditional backup data once was. Nowadays, your backup data is also active data, so read and write performance become equally important, along with global and secure access of the data.

The Hedvig Distributed Storage Platform together with Veeam Backup & Replication not only provide high-availability, security, and broad application support, but also make your applications future-proof with the flexibility of selecting any hardware vendor and built-in public cloud mobility.

Additional resources

Hedvig architecture overview:

<https://www.hedvig.io/technical-overview-whitepaper>

SDS for backup and recovery:

https://www.hedviginc.com/hubfs/Website_Resources/Hedvig-backup-and-recovery-solution-brief.pdf?t=1515174582999