Technical Report

NetApp E-Series storage with Veeam Backup & Replication

Reference architecture and best practices

Mitch Blackburn and Alonso DeVega, NetApp; Pete Ybarra, Veeam; Matthias Beller, Advanced UniByte
September 2024 | TR-4948

In partnership with

Abstract

This document outlines the reference architecture and best practices when using NetApp® E-Series storage in a Veeam Backup & Replication environment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Executive summary

The challenge

With data growing at astounding rates, IT managers depend more and more on reliable data backup and recovery. High-growth businesses require a complete data protection solution that is reliable, flexible, and easy to use. Virtualizing an environment provides increased levels of data availability but meeting aggressive recovery point objectives (RPOs) and recovery time objectives (RTOs) becomes increasingly difficult.

Traditional backup tools were not created for virtualized environments. That fact makes it hard for many organizations to take full advantage of their virtualized environment, and many IT managers struggle with:

- Unreliable backups
- Recovery that takes too long
- High costs that are associated with managing backup data and secondary storage
- An inability to provide reliable and true backups for compliance purposes
- Lost productivity because of management complexity
- The need to scale backup operations for growth
- Using immutability to protect against malware

The solution

To meet these challenges, Veeam and NetApp collaborated to offer high-performance storage with reliable data protection that is designed for virtualized environments.

Veeam and NetApp help you modernize your data protection strategy with a solution that is designed to manage large data volumes and to handle the increasing performance and availability demands of a 21st-century infrastructure.

Veeam Backup & Replication unifies backup and replication in a single solution, increasing the value of backup and reinventing data protection for VMware vSphere and Microsoft Hyper-V virtual environments. The Veeam agentless design provides multiple backup options to meet your needs. Features such as source-side deduplication and compression, change block tracking, parallel processing, and automatic load balancing provide fast and efficient backups.

The Veeam Linux Hardened Repository protects your backup data and prevents data loss due to malware activity or unplanned actions.

NetApp E-Series storage provides simple and reliable SAN storage that integrates seamlessly with most application environments. Its modular design helps decrease operating expenses while offering many options for connectivity, capacity, and performance that easily scale to meet the demands of a growing backup environment.

Together, Veeam and NetApp create an optimal staging area for backups, reducing backup ingest bottlenecks and providing faster backups through parallel processing.

In addition, Veeam Backup & Replication provides:

- Granular recovery of virtual machines (VMs) and files, including Microsoft Exchange and SharePoint application items
- The ability to automatically verify every backup, VM, and replica every time
- Self-service recovery of VMs and guest files without direct network connection to the VM, user permissions, or the need to deploy costly agents
- Instant VM recovery to recover a failed VM

- A choice to back up and recover what you need, where you need it, and when you need it, whether it is on site, on tape, or in the cloud

Veeam and NetApp offer the right solution for performance, flexibility, and reliability, providing an impressive modern disaster recovery solution for your vSphere or Hyper-V environment.

This document is a reference architecture for enabling a collaborative backup and recovery solution on NetApp E-Series with Veeam Backup & Replication data protection software.

Introduction

Veeam and NetApp jointly developed this reference architecture to guide successful Backup & Replication deployments with E-Series storage and to enable data and application availability.

NetApp E-Series and Veeam Backup & Replication combine to offer a data protection and availability solution through this tested reference architecture from industry leaders NetApp and Veeam. This solution is optimized for virtual environments, providing disk-to-disk backup and recovery on high-capacity, flexible, performance-oriented NetApp E-Series storage arrays. This solution provides you with superior data management for virtual environments and high availability while also making your data highly available.

NetApp E-Series arrays provide a high-performing backup repository to store Veeam-created backups. With this capability, the recovery technologies that are enabled through Veeam can satisfy stringent RTOs. Recovery technologies such as instant VM recovery, SureBackup, and on-demand sandbox can leverage backup repositories that are capable of high I/O to achieve their full potential.

These technologies enable you to restore from your backups faster and enable capabilities such as automated recovery verification. The technologies can also leverage backup data as an ad hoc testing environment. This capability changes the way that users have used backups in the past because more benefits are associated with having backups. No longer do backups sit idle, waiting for an emergency restore; you can apply your backups for many creative uses.

Features include:

- Fast recovery of a failed VM
- Near-continuous data protection with built-in replication
- Fast, agentless item recovery and e-discovery for Microsoft Exchange, SharePoint, and Active Directory, along with transaction-level recovery of SQL Server databases
- Automatic recoverability testing of every backup and every replica, every time
- Off-site backups made up to 50 times faster than the speed of standard file copy with built-in WAN acceleration
- Fast and secure cloud backups with Veeam Cloud Connect
- Immutable backups with Veeam Linux Hardened Repository
- Deduplication and compression to minimize storage consumption
- Off-site recovery with one-click site failover and support for facilitated data center migrations with zero data loss

About NetApp

NetApp creates innovative products, including storage systems and software that help customers around the world store, manage, protect, and retain one of their most precious corporate assets: their data. We are recognized throughout the industry for continually pushing the limits of today’s technology so that our customers don’t have to choose between saving money and acquiring the capabilities that they need to be successful.

We find ways to enable our customers to do things that they couldn't do before and at a speed that they never thought possible. We partner with industry leaders to create efficient and cost-effective solutions that are optimized for customers' IT needs and to deliver to and support these customers worldwide. Leading organizations worldwide count on NetApp for software, systems, and services to manage and store their data. Customers value our teamwork, expertise, and passion for helping them succeed now and into the future NetApp.com.

About Veeam

Veeam is the home of Radical Resilience. Our mission is to help every company in the world not just bounce back from an outage or data loss, but to bounce forward. We call this radical resilience, and we're obsessed with creating innovative new ways to help our customers achieve it. Veeam.com.

About Advanced UniByte

Advanced UniByte is one of the leading system integrators for datacenter infrastructure, storage solutions, as well as cloud and managed services. Advanced UniByte was founded in 1994, has currently 250 employees and is based in Metzingen, Germany. Our portfolio covers storage, backup, virtualization, compute, network and security products as well as cloud and managed services. Together with these solutions we successfully deploy FlexPod solutions and have been awarded as FlexPod Partner of the year 2023. Key factors in our ability to lead are the deep expertise in the products we use and having our own lab for testing solutions. Advanced UniByte celebrates 20 years of NetApp partnership in 2024, additionally one of the largest NetApp LSC Partners in Germany.

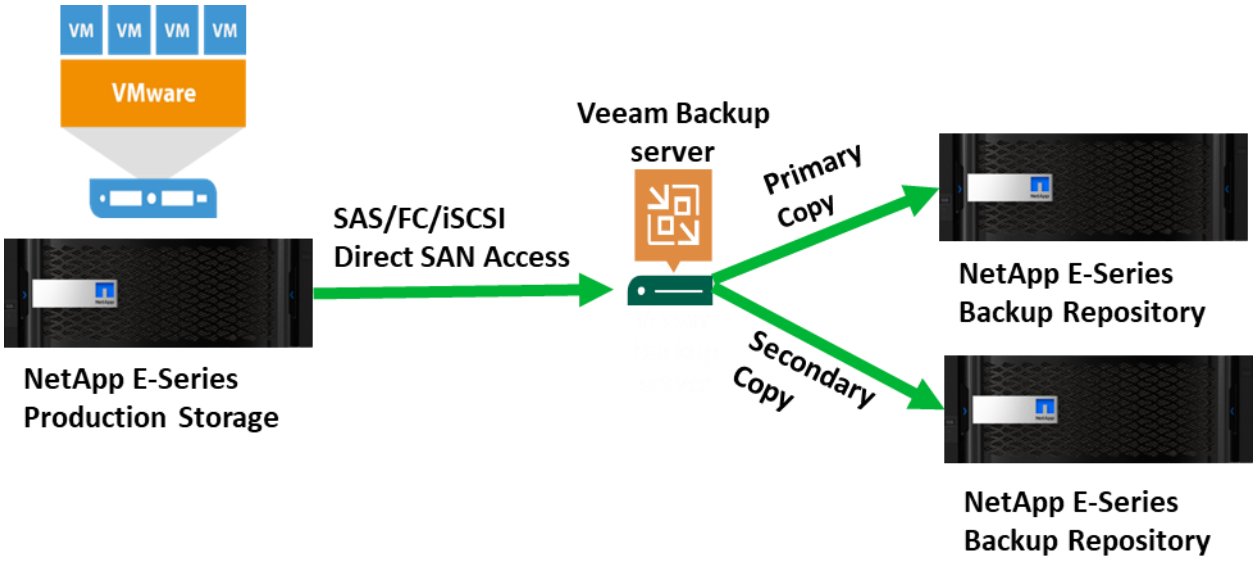
Reference architecture overview

This section details reference architectures that range from those of small environments that protect a few terabytes of data to those in enterprise-size environments with petabytes of data under management.

NetApp E-Series as Veeam Backup & Replication repositories for backup and archiving

Figure 1 gives us a graphical representation of a NetApp E-Series and Veeam setup.

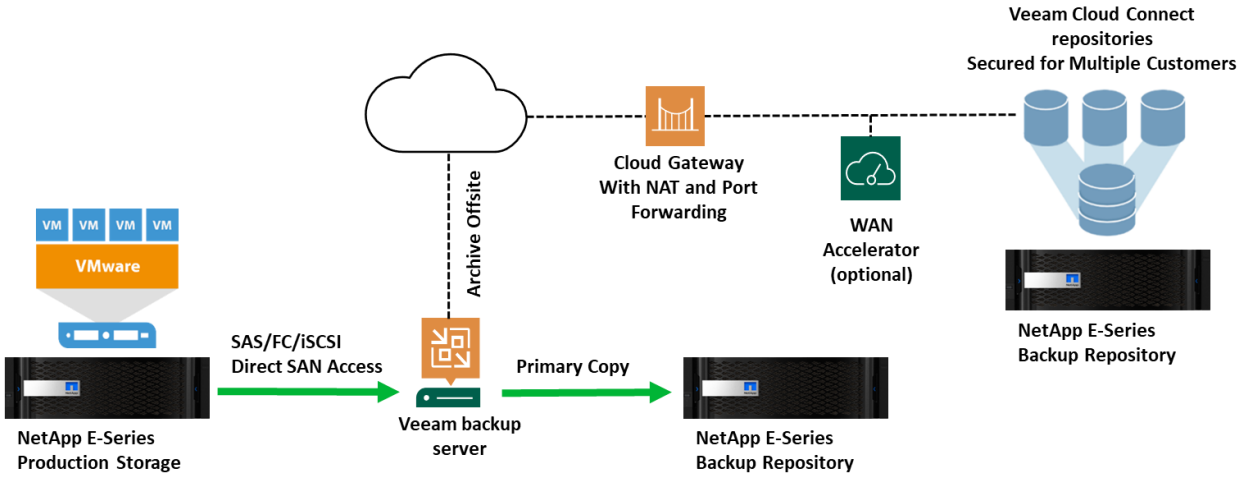
Figure 1) NetApp E-Series as Veeam Backup & Replication repositories for backup and archiving.



NetApp E-Series as a Veeam Backup & Replication backup repository and an off-site cloud repository

Figure 2 shows NetApp E-Series as a Veeam Backup & Replication backup repository and an off-site cloud repository.

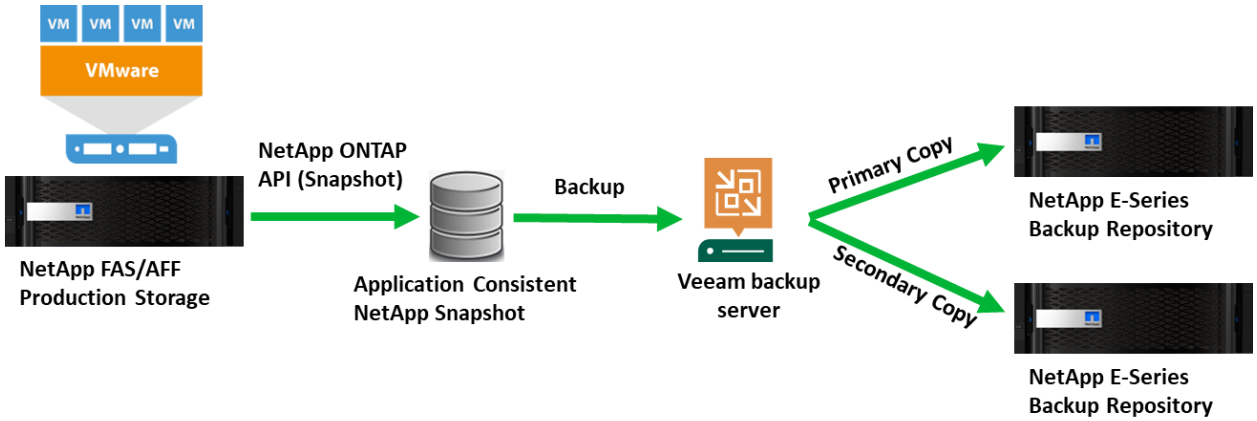
Figure 2) NetApp E-Series as a Veeam Backup & Replication backup repository and an off-site cloud repository.



NetApp E-Series as a Veeam Backup & Replication backup repository for NetApp FAS production storage

Figure 3 illustrates Veeam integration with a NetApp FAS series production storage array, with newly created backups going to an E-Series array for storage. To provide disaster recovery, backups can also be sent off the premises to another backup repository (another E-Series system). Veeam provides a backup copy job for such scenarios; this job can be leveraged for backups off the premises or for long-term archiving by using Veeam's built-in grandfather-father-son (GFS)–type retention.

Figure 3) NetApp E-Series as a Veeam Backup & Replication backup repository and off-premises backup repository for ONTAP production storage.



Leveraging Veeam's backup copy job architecture is important for achieving that last level of protection. The off-site copy provides safeguards for an entire data center–level disaster. Veeam also provides an optional WAN acceleration component that can help reduce bandwidth utilization. This component can

play a huge role in environments that have active-active sites or that have low available bandwidth to start with. The forever-incremental nature of the backup copy job enables only incremental change data to be transmitted off site after the initial copy. Pre-seeding options are available for the initial transfer of data for environments that need it.

Deployment scenarios

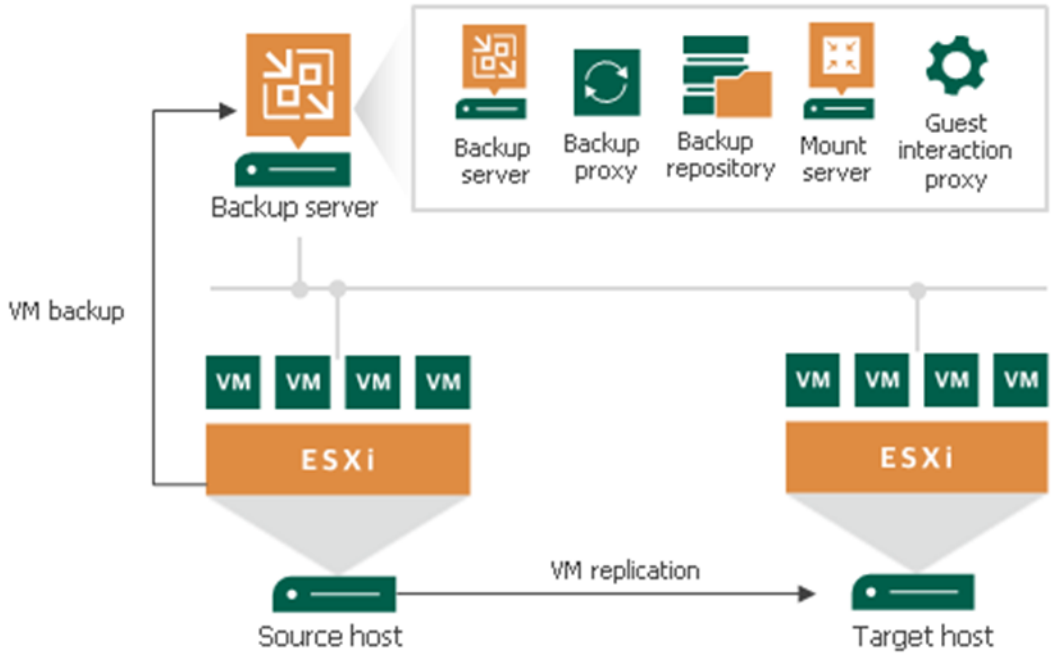
Simple deployment

In a simple deployment scenario, one instance of Veeam Backup & Replication is installed on a physical or virtual Windows-based machine. This installation is referred to as a Veeam backup server.

Simple deployment (see Figure 4) implies that the Veeam backup server fills three major roles:

- It functions as a management point, coordinates all jobs, controls job scheduling, and performs other administrative activities.
- It acts as the default backup proxy for handling job processing and for transferring backup traffic. All services that are necessary for the backup proxy functionality are installed on the Veeam backup server locally.
- It is used as the default backup repository. During installation, Veeam Backup & Replication checks volumes of the machine on which you install the product and identifies a volume with the greatest amount of free disk space. On this volume, Veeam Backup & Replication creates the backup folder that is used as the default backup repository.

Figure 4) Simple deployment.



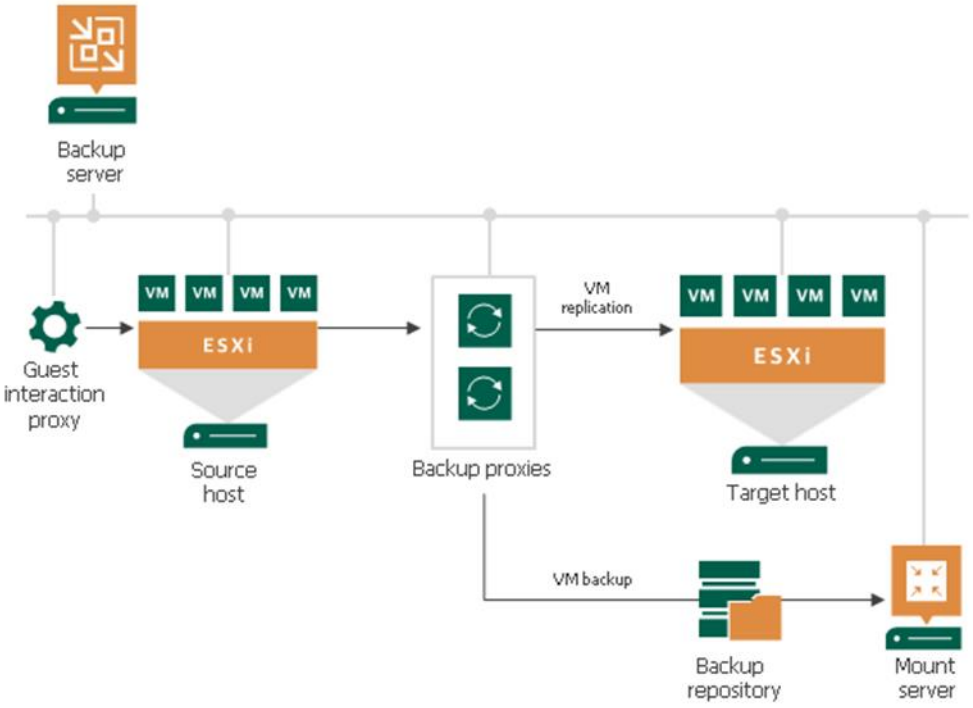
If you plan to back up and replicate only a few VMs or evaluate Veeam Backup & Replication, this configuration is enough to get you started. Veeam Backup & Replication is ready for use right out of the box; as soon as it is installed, you can start using the solution to perform backup and replication operations. To balance the load of backing up and replicating your VMs, you can schedule jobs at different times.

Advanced deployment

In large-scale virtual environments with numerous jobs, the load on the Veeam backup server is heavy. In this case, NetApp recommends using the advanced deployment scenario (see Figure 5), which moves the backup workload to dedicated backup proxies and backup repositories.

The essence of the advanced deployment is that the backup proxy takes off part of Veeam backup server activities (namely, it collects and processes data and moves backup traffic from the source to the target). In addition, the Veeam backup server no longer acts as a storage location. The backup proxy transports VM data to a dedicated backup repository that keeps backup files, VM copies, metadata, and so on. The Veeam backup server in this scenario functions as a manager for deploying and maintaining backup proxies and repositories.

Figure 5) Advanced deployment.



To deploy a backup proxy or a backup repository, add a server to Veeam Backup & Replication and assign a proxy or a repository role to it, as applicable. Veeam Backup & Replication automatically installs lightweight components and services on these servers. A backup proxy does not require separate configuration server database; all settings are stored centrally, within the SQL or PostgreSQL database that the Veeam backup server uses.

With the advanced deployment scenario, you can easily meet your current and future data protection requirements. You can expand your backup infrastructure horizontally in a matter of minutes to match the amount of data that you want to process and the available network throughput. Instead of increasing the number of backup servers or constantly tuning job scheduling, you can install multiple backup proxies and repositories and distribute the backup workload among them. The installation process is fully automated, which simplifies deploying and maintaining the backup infrastructure in your virtual environment.

In virtual environments with several proxies, Veeam Backup & Replication dynamically distributes backup traffic among those proxies. You can explicitly map a job to a specific proxy, or you can let Veeam Backup & Replication choose the most suitable proxy. If you opt for the latter, Veeam Backup &

Replication checks the settings of available proxies and selects the most appropriate one for the job. The proxy server to be used should have access to the source and target hosts as well as to the backup repository to which files are written.

The advanced deployment scenario can be a good choice for backing up and replicating off site. You can deploy a backup proxy in the production site and another one in the disaster recovery (DR) site, closer to the backup repository. When a job is performed, backup proxies on both sides establish a stable connection, so this architecture also allows efficient data transport over a slow network connection or WAN.

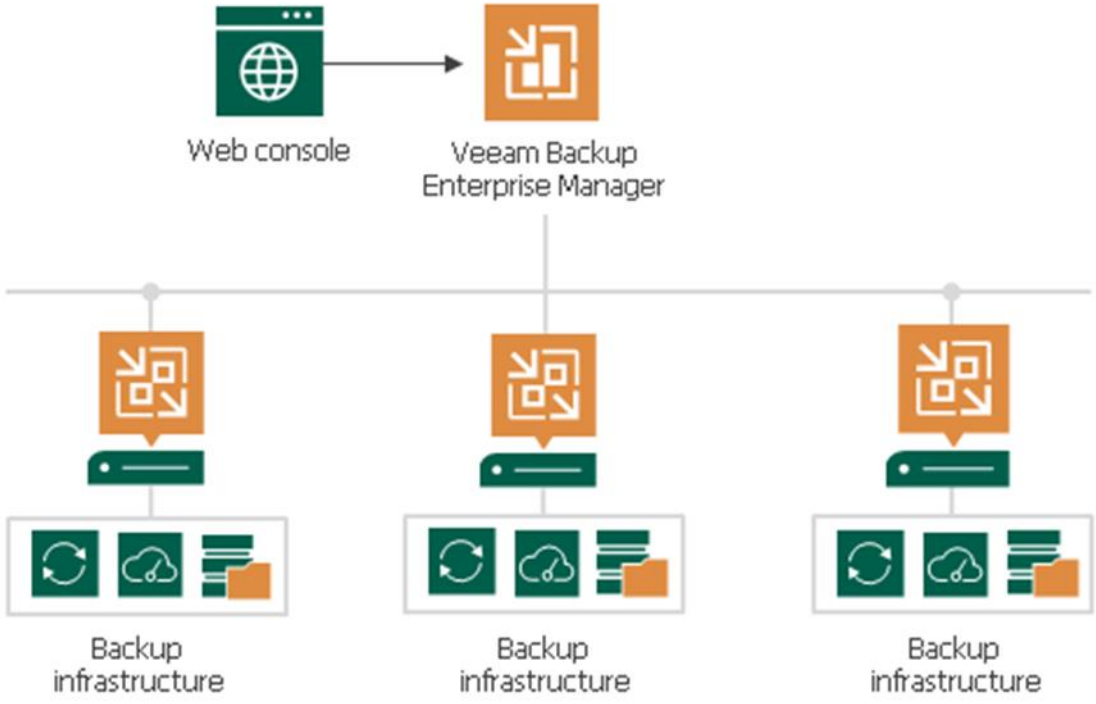
To regulate backup load, you can specify the maximum number of concurrent tasks per proxy and set up throttling rules to limit proxy bandwidth. The maximum number of concurrent tasks can also be specified for a backup repository in addition to the value of the combined data rate for it.

Another advantage of the advanced deployment scenario is that it contributes to high availability. Jobs can migrate between proxies if one of them becomes overloaded or unavailable.

Distributed deployment

NetApp recommends the distributed deployment scenario (see Figure 6) for large geographically dispersed virtual environments with multiple Veeam backup servers that are installed across different sites. These backup servers are federated under Veeam Backup Enterprise Manager, an optional component that provides centralized management and reporting for these servers through a web interface.

Figure 6) Distributed deployment.



Veeam Backup Enterprise Manager collects data from Veeam backup servers and enables you to run backup and replication jobs across the entire backup infrastructure through a single interface. You can also edit those jobs and clone jobs by using a single job as a template. Enterprise Manager also provides

reporting data for various areas (for example, all jobs that were performed within the past 24 hours or 7 days, all VMs that were engaged in these jobs, and so on).

By using indexing data that is consolidated on one server, Veeam Backup Enterprise Manager provides advanced capabilities to search for VM guest OS files in VM backups that are created on all Veeam backup servers. You can search even if the files are stored in repositories on different sites, and you can recover them in a single click. Searching for VM guest OS files is enabled through Veeam Backup Enterprise Manager; to streamline the search process, you can optionally deploy a Veeam Backup Search server in your backup infrastructure.

With flexible delegation options and security roles, IT administrators can delegate the necessary file restore or VM restore rights to authorized personnel in your organization. For example, they can allow database administrators to restore Oracle or SQL Server VMs.

If you use Veeam Backup Enterprise Manager in your backup infrastructure, you do not need to install licenses on every Veeam backup server that you deploy. Instead, you can install one license on the Veeam Backup Enterprise Manager server, and it is applied to all servers across your backup infrastructure. This approach simplifies tracking license usage and license updates across multiple Veeam backup servers.

In addition, VMware administrators can benefit from the Veeam plug-in for vSphere Web Client, which can be installed by using Veeam Backup Enterprise Manager. Administrators can analyze cumulative information about used and available storage space; view statistics on processed VMs; and review success, warning, and failure counts for all jobs. Administrators can also easily identify unprotected VMs and perform capacity planning for repositories, all directly from vSphere.

NetApp E-Series arrays

Our E-Series arrays are the go-to system for so many because of the simplicity and reliability they deliver. From mid-sized businesses driving data-intensive applications like analytics, video surveillance, and disk-based backup, to small enterprises and remote offices needing mixed-workload performance for their dedicated apps.

As can be seen on the Veeam Alliance Technical Programs website, both NetApp E-Series E2800 and E5700 are Veeam Ready and are verified backup storage that supports all Veeam backup and restore features.

E2800 and E5700 hybrid arrays

The NetApp E-Series E2800 and E5700 are industry-leading storage systems that deliver high input/output operations per second (IOPS) and bandwidth with consistently low latency to support the demanding performance and capacity needs of backup and recovery applications such as Veeam.

The E2800 and E5700 provide the following benefits:

- Support for wide-ranging workloads and performance requirements
- Fully redundant I/O paths, advanced protection features, and proactive support monitoring and services for high levels of availability, integrity, and security
- Increased IOPS performance by up to 20% compared to the previous high-performance generation of E-Series products
- A level of performance, density, and economics that leads the industry
- Interface protocol flexibility to support FC host and iSCSI host workloads simultaneously

As shown in Table 1, E2800 is available in three shelf options and E5700 is available in two shelf options, which support both hard-disk drives (HDDs) and solid-state drives (SSDs) to meet a wide range of performance and backup application requirements.

Table 1) E2800 and E5700 controller shelf and drive shelf models.

Controller Shelf Model	Drive Shelf Model	Number of Drives	Type of Drives
E2860/E5760	DE460C	60	2.5” and 3.5” SAS drives (HDDs and SSDs)
E2824/E5724	DE224C	24	2.5” SAS drives (HDDs and SSDs)
E2812	DE212C	12	2.5” and 3.5” SAS drives (HDDs and SSDs)

Each controller shelf includes two controllers, with each controller providing two ethernet management ports for out-of-band management. The system also supports in-band management access and has two 12Gbps (x4 lanes) wide-port SAS drive expansion ports for redundant drive expansion paths. The controllers might also include two built-in host ports, which can be configured as either 16Gb FC or 10Gb iSCSI. The following host interface cards (HICs) can be installed in each controller:

Note: Both controllers in an array must be identically configured.

- 4-port 12Gb SAS Wide Port (SAS-3 connector)
- 4-port 32Gb FC
- 4-port 25Gb iSCSI
- 2-port 100Gb InfiniBand (IB) (E5700 controllers only)

Note: E2800 can also be configured having only one controller.

SANtricity OS

The NetApp E-Series controllers and SANtricity OS use the on-box, browser-based management interface, SANtricity System Manager.

E-Series storage systems are shipped preloaded with SANtricity OS, which includes SANtricity System Manager. To discover multiple storage systems running SANtricity OS from a central view, download SANtricity Unified Manager (which includes the Web Services Proxy) from the NetApp Support site. Then load it on a management server that has IP access to the storage systems.

For further information about the SANtricity Unified Manager and the SANtricity System Manager, see the E-Series and SANtricity documentation resources page.

SANtricity security features

The new-generation E-Series arrays running the latest SANtricity OS are Common Criteria certified (NDcPP v2 certification). SANtricity security features include Lightweight Directory Access Protocol (LDAP), role-based access control (RBAC), and Secure Sockets Layer (SSL) certificates. For complete details and workflow examples, see the following resources:

- TR-4712: NetApp SANtricity Management Security Feature Details and Configuration Guide,
- TR-4855: Security Hardening Guide for NetApp SANtricity
- TR-4813: Managing Certificates for NetApp E-Series Storage Systems

SANtricity drive security technology provides comprehensive security for data at rest without sacrificing system performance or ease of use and supports both internal and external key management.

For more information about disk encryption, see TR-4474: SANtricity drive security.

SANtricity reliability features

Table 2 provides a list of SANtricity reliability features and a brief explanation of each with references to additional information.

Table 2) SANtricity features for long-term reliability.

Reliability features with SANtricity
Proactive drive monitor and data evacuator. Nonresponsive drives are automatically power-cycled to see if the fault condition can be cleared. If the condition cannot be cleared, the drive is flagged as failed. For predictive failure events, the evacuator feature starts to remove data from the affected drive to move the data before the drive fails. If the drive fails, the rebuild resumes where the evacuator was disrupted, reducing the rebuild time.
Automatic drive fault detection, failover, and rebuild. You can perform these tasks by using global hot spare drives for standard RAID and spare pool capacity for Dynamic Disk Pools (DDP).
SSD wear-life tracking and reporting. This metric is found in the hardware tab's Drive Settings dialog box. It indicates the wear life of SSDs and replaces two SSD wear-life metrics (average erase count and spare blocks remaining) that were in previous versions of SANtricity. The metric is percent endurance used; to access it, select a drive from the hardware view and then select Settings.
Online drive firmware upgrade. This feature upgrades one drive at a time and tracks writes to the affected drives during the upgrade window; it should be used only during low write I/O periods. Parallel drive firmware upgrades are supported offline to upgrade multiple drives more quickly during a maintenance window.
Automatic load balancing. This feature provides automated I/O workload balancing and confirms that incoming I/O traffic from hosts is dynamically managed and balanced across both controllers. The workload of each controller is continually monitored and analyzed in the background. When I/O on one controller significantly exceeds the I/O on the other controller for a prolonged, predictable period, SANtricity can change volume ownership from the busy controller to the less busy controller. The feature does not react to short-term changes in I/O patterns. However, when a change of ownership is needed, SANtricity interacts with the affected host multipath driver to initiate an implicit path failover. Most current server operating systems and associated multipath drivers support implicit failover. For more information, search for "What is automatic load balancing?" in the System Manager online help.
Embedded SNMP agent. For the EF600 controller, SNMP is supported natively. The embedded SNMP agent complies with the SNMP V2C standard and RFC 1213 (MIB-II). For more information, search for "manage SNMP alerts" in the System Manager online help.
Automatic alerts. This feature sends email alerts to notify data center support staff about events on the storage array.
Event Monitor and system log. The SANtricity Storage Manager Event Monitor automatically records events that occur on the storage array. Syslog enables a second level of activity tracking that allows you to connect events with associated changes recorded in the system log.
AutoSupport. E-Series products have supported AutoSupport for several releases.
Ability to enable or disable AutoSupport maintenance window. AutoSupport includes an option for enabling or suppressing automatic ticket creation on error events. Under normal operation mode, the storage array uses AutoSupport to open a support case if there is an issue. To enable or disable the AutoSupport maintenance window, select Support > Access Management > AutoSupport.

SANtricity data protection features

E-Series has a reputation for reliability and availability. Many of the data protection features in these systems can be beneficial in a backup environment.

Background media scan and data assurance (T10 PI)

Media scan is a background process that is performed by the controllers to provide error detection on the drive media. The main purpose of the feature is to detect and repair media errors on disk drives that are

infrequently read by user applications and where data loss might occur if other drives in the volume group fail. A secondary purpose is to detect redundancy errors such as data/parity mismatches. A background media scan can find media errors before they disrupt normal drive reads and writes.

The data assurance feature provides controller-to-drive data integrity protection through the SCSI direct-access block device protection information model. This model protects user data by appending protection information to each block of user data. The protection model is sometimes referred to as data integrity field protection or T10 PI. This model makes sure that an I/O has completed without any bad blocks written to or read from disk. It protects against displacement errors, data corruption resulting from hardware or software errors, bit flips, and silent drive errors, such as when the drive delivers the wrong data on a read request or writes to the wrong location.

You need both data assurance and media scan. They complement each other to protect your data.

Unreadable sector management

This feature provides a controller-based mechanism for handling unreadable sectors detected both during normal I/O operation of the controller and during long-lived operations such as reconstructions. The feature is transparent to the user and requires no special configuration.

Dynamic Disk Pools

With Dynamic Disk Pools (DDP) technology, NetApp SANtricity OS and management software allows you to create pools in addition to traditional volume groups (generally referred to as RAID groups). A pool can range in size from a minimum of 11 drives to as large as all the drives in a storage system, which is up to 420 SAS drives in the NetApp EF600 system. Pools can consist of either HDDs or SSDs. In addition, pools and volume groups can coexist in the same system.

For more information about DDP, see TR-4652: SANtricity OS Dynamic Disk Pools.

Veeam Backup & Replication overview

Veeam Backup & Replication is a data protection and disaster recovery solution for VMware vSphere and Microsoft Hyper-V virtual environments of any size or complexity. By combining all the necessary functions in one intuitive interface, Veeam Backup & Replication solves the most critical problems of virtualized infrastructure management. The solution also protects mission critical VMs from both hardware and software failures.

Solution architecture

Veeam Backup & Replication is composed of the following three elements:

- Backup server
- Backup proxy
- Backup repository

Veeam 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 Veeam backup server performs all types of administrative activities. This server:

- Coordinates backup, replication, recovery verification, and restore tasks
- Controls job scheduling and resource allocation

- Is used to set up and manage backup infrastructure components and to specify global settings for the backup infrastructure

In addition to its primary functions, a newly deployed Veeam backup server also acts as a default backup proxy and the backup repository (it manages data handling and data storing tasks).

The Veeam backup server uses the following services and components:

- **Veeam Backup Service** is a Windows service that coordinates all the operations that Veeam Backup & Replication performs, such as backup, replication, recovery verification and restore tasks. The Veeam Backup Service runs under the local system account or an account that has the local administrator permissions on the backup server.
- **Veeam Broker Service** interacts with the virtual infrastructure to collect and cache the virtual infrastructure topology. It jobs and tasks query information about the virtual infrastructure topology from the Broker Service, which accelerates job and task performance.
- **Veeam Guest Catalog Service** manages guest OS file system indexing for VMs and replicates system index data files to enable search through guest OS files. Index data is stored in the Veeam Backup Catalog, which is a folder on the backup server. The Veeam Guest Catalog Service running on the backup server works in conjunction with search components that are installed on Veeam Backup Enterprise Manager and (optionally) a dedicated Microsoft Search Server.
- **Veeam Mount Service** mounts backups and replicas for file-level access, browsing the VM guest file system and restoring VM guest OS files and application items to the original location.
- **Veeam backup proxy services**: In addition to dedicated services, the backup server runs a set of data mover services. For details, see the following “Backup Proxy” section.
- **Veeam Backup & Replication Configuration database** stores data about the backup infrastructure, jobs, sessions, and so on. The database instance can be on an SQL Server or PostgreSQL Server (PostgreSQL is the default configuration database when a clean install is performed) that is installed either locally (on the same machine where the backup server is running) or remotely.
- **Veeam Backup & Replication console** provides the application UI and allows user access to the application’s functionality.
- **Veeam Backup PowerShell snap-in** is an extension for Microsoft Windows PowerShell 2.0. Veeam Backup PowerShell adds a set of cmdlets that enable you to perform backup, replication, and recovery tasks through the PowerShell CLI or run custom scripts to fully automate Veeam Backup & Replication operation.

Backup proxy

When Veeam Backup & Replication is first installed, the Veeam backup server coordinates all job activities and handles data traffic. So, when you run a backup, replication, VM copy, or VM migration job or perform restore operations, VM data is moved from the source to the target through the Veeam backup server. This scenario is acceptable for virtual environments in which few backup jobs are performed. In large-scale environments, however, the workload on the Veeam backup server is significant.

To take the workload off the Veeam backup server, Veeam Backup & Replication uses backup proxies. A backup proxy is an architecture component that sits between the data source and the target and is used to process jobs and to deliver backup traffic. In particular, the backup proxy tasks include retrieving VM data from the production storage. The tasks also include compressing the data and sending it 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). As the data handling task is assigned to the backup proxy, the Veeam backup server becomes the point of control for dispatching jobs to proxy servers.

The role of a backup proxy can be assigned to a dedicated Windows or Linux Server (physical or virtual) in your environment. You can deploy backup proxies both in the primary site and in remote sites. To optimize the performance of several concurrent jobs, you can use several backup proxies. In this case, Veeam Backup & Replication distributes the backup workload between available backup proxies.

By using backup proxies, you can easily scale your backup infrastructure up and down based on your demands. Backup proxies run lightweight services that take a few seconds to deploy. Deployment is fully automated: Veeam Backup & Replication installs the necessary components on a Windows-based server when you add it to the product console. As soon as you assign the role of a backup proxy to the added server, Veeam Backup & Replication starts the required services on it.

The primary role of the backup proxy is to provide an optimal route for backup traffic and to enable efficient data transfer. Therefore, when deploying a backup proxy, you must analyze the connection between the backup proxy and the storage with which it is working. Depending on the type of connection, the backup proxy can be configured in one of the following ways (starting with the most efficient):

- A machine used as a backup proxy should have direct access to the production storage on which the VMs reside or, in case of restores, the storage to which VM data is written. In this way, the backup proxy retrieves data directly from the production storage which will result in optimal performance. Depending on the protocol used on production storage to connect the VMware Datastores, different modes are available: direct SAN access (FC or iSCSI) for SAN storage systems and direct NFS access if production storage uses a NFS storage solution like NetApp ONTAP. It is recommended to use a physical server as backup proxy for better performance, if using direct SAN access with Fibre Channel (FC) the server must be physical, because of the physical FC connection. With FC connections Veeam is able to bypass the LAN for a LAN-free backup solution.
- If using VMware HotAdd feature, the backup proxy must be a VM with access to VM disks on the source datastore. Please be aware that this mode is recommended for HCI, VSAN, local disks or if direct storage access is not possible. It is not recommended for NFS source storage and may need additional configuration, for more details see Veeam KB articles KB1054: Appliance Mode (Hotadd) Requirements and Troubleshooting and KB1681: VM Loses Connection During Snapshot Removal.
- If neither of the preceding scenarios is possible, you have alternatives. You can assign the role of the backup proxy to a machine on the network that is closer to the source or closer to the target storage with which the proxy works. In this case, VM data is transported over the LAN by using the Network Block Device (NBD) protocol.

Depending on the type of backup proxy and your backup architecture, the backup proxy can use different transport modes. If the source storage system is added to the Backup & replication console, the backup proxy can also use the Backup from Storage Snapshots mode. You can select the transport mode or let Veeam Backup & Replication automatically choose it.

The backup proxy uses the following services and components:

- **Veeam Installer Service** is an auxiliary service that is installed and started on any Windows/Linux server after it is added to the list of managed servers in the Veeam Backup & Replication console. This service analyzes the system and installs and upgrades necessary components and services depending on the role that is selected for the server.
- **Veeam Transport** is responsible for deploying and coordinating executable modules that act as data movers and that perform the main job activities on behalf of Veeam Backup & Replication. These activities include communicating with VMware Tools, copying VM files, performing data deduplication and compression, and so on.

Backup repository

A backup repository is a location that Veeam Backup & Replication jobs use to store backup files, copies of VMs, and metadata for replicated VMs. Technically, a backup repository is a folder on the backup storage. By assigning different repositories to jobs and by limiting the number of parallel jobs for each one, you can balance the load across your backup infrastructure.

Veeam Backup & Replication requirements

System requirements

Please view the Veeam Backup & Replication 12 User Guide, subsection Supported Platforms and Applications, for server requirements.

Note: Nutanix AHV is not supported.

Capacity sizing

This section describes the procedure and parameters to be considered while estimating the amount of disk space required.

When estimating the amount of required disk space, you should know the following:

- Total size of VMs being backed up
- Frequency of backups
- Retention period for backups
- Whether jobs use forward or reverse incremental

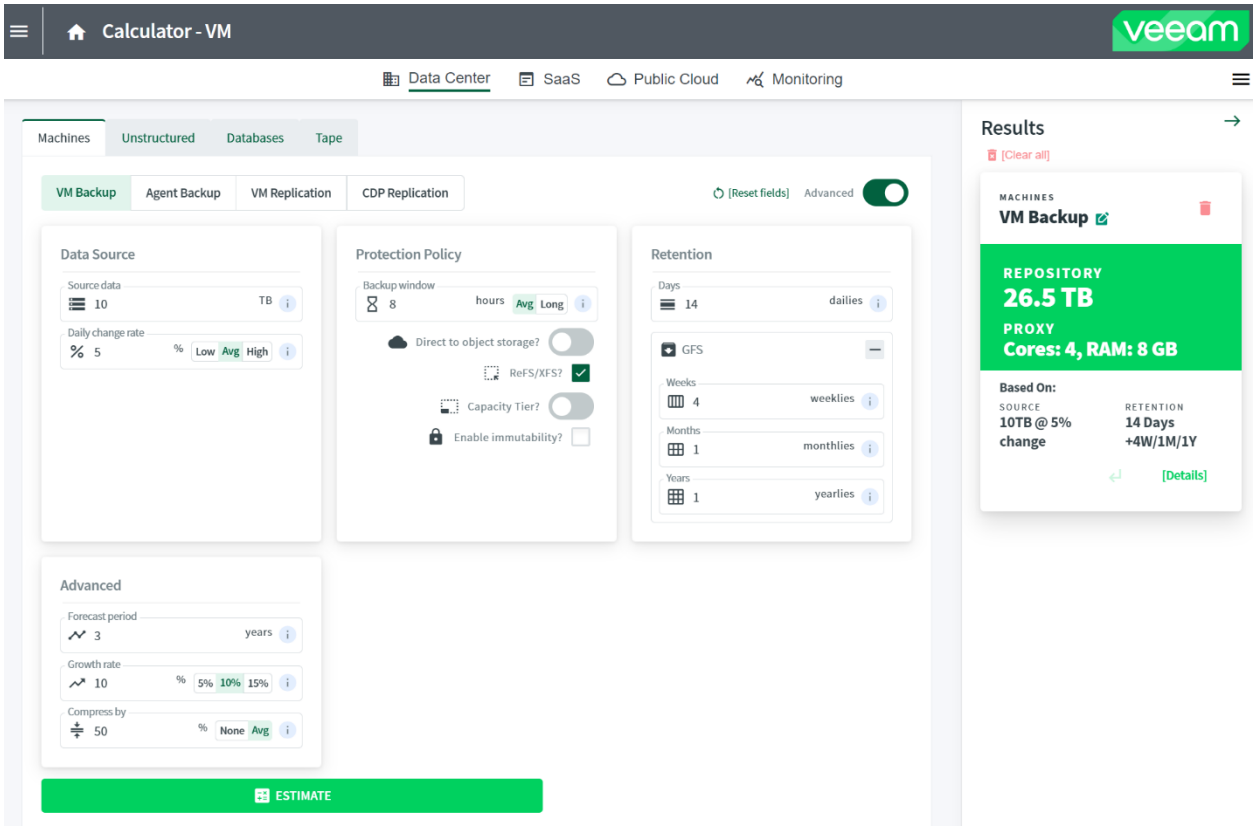
Also, when testing is not possible beforehand, you should make assumptions about compression and deduplication ratios, change rates, and other factors. The following figures are typical for most deployments; however, it is important to understand the specific environment to figure out possible exceptions:

- Data reduction because of compression and deduplication is usually 2:1 or more; it's common to see 3:1 or better, but you should always be conservative when estimating required space.
- Typical daily change rate is between 2% and 5% in a midsize or enterprise environment. This rate can greatly vary among servers; some servers show much higher values. If possible, run monitoring tools such as Veeam ONE to have a better understanding of the real change rate values.
- Include additional space for one-off full backups.
- Include additional space for backup chain transformation (forward forever incremental, reverse incremental): at least the size of a full backup multiplied by 1.25x.

Veeam backup capacity calculator

A capacity planning tool that can be used for estimation is available at https://www.veeam.com/calculators. Figure 7 shows an overview of the backup capacity calculator by Veeam.

Figure 7) Veeam Backup Capacity Calculator.



Configure E-Series Storage Based on Veeam Calculator

We will be using the backup capacity of 26.5TB from the Veeam Capacity Calculator values listed in **Error! Reference source not found.** to configure our E-Series storage.

The usable capacity required is 26500GB. We can configure our array using the E2812, which is a 2U shelf containing 12 drives. For this example, we will use ten 4TB NL-SAS drives.

Now that we have our drives, we can create a RAID6 (8+2) volume group which will provide a usable capacity of 29.07TB. From the RAID6 volume group, volumes (LUNs) can be created and assigned to the Veeam backup server to use as backup repositories.

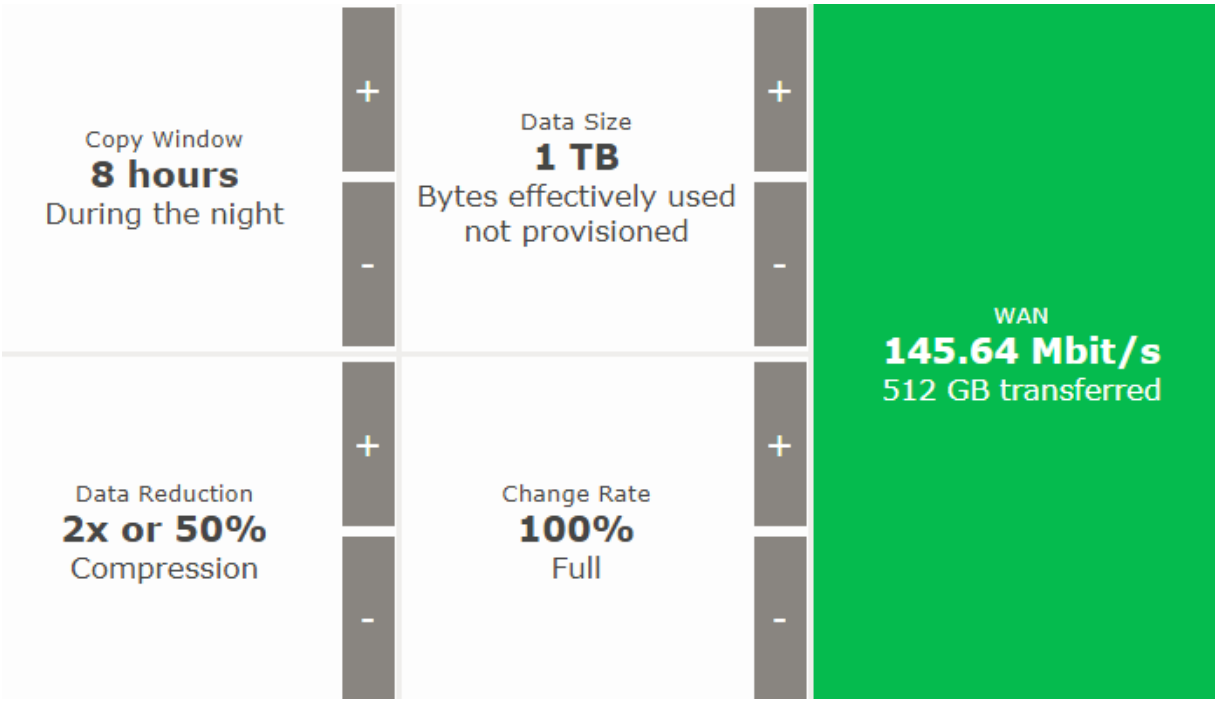
Network planning and sizing

Veeam has an unofficial tool that can assist in network planning. It can be accessed at http://rps.dewin.me/bandwidth/. Figure 8 gives an example of the tool. Overall, it has four parameters that need input, based on which it gives an approximate bandwidth required.

- Copy Window: This parameter is divided into four categories; select the input based on the expected copy window:
  - Short Slot: your copy window is below 6 hours.
  - During the night: Your copy window is below 8 to 9 hours.
  - During business hours: Your copy window is below 24 hours.
  - Over multiple days: Your copy window can span multiple days.
- Data Size: Make sure you input the bytes effectively used, not provisioned.
- Data Reduction: This parameter is divided into four categories:

  - Reduction Disabled: No compression or data reduction is carried out.
  - Compression: From 40% to 80% data is transferred after compression and reduction.
  - Extreme Compression: From 33% to 36% data is transferred after compression and reduction.
  - Veeam WAN Acceleration: WAN acceleration is used to reduce the data transferred below 30%.
- Change Rate: This parameter should be selected based on the data change rate expected. Average is ~10%. Databases have a change rate ~20% to 50%.

Figure 8) Network planning tool.



Veeam Backup & Replication design considerations

Direct SAN access mode

Veeam proxy servers directly connect to the storage fabric because it is the fastest way to perform backups with Backup & Replication. An often-discussed point about direct SAN access for the Veeam proxy is the possibility that the Veeam proxy can mount VMware Virtual Machine File System (VMFS) LUNs and write a Windows signature onto the LUN. Doing so results in an inaccessible VMFS, which requires VMware support to fix the issue. Veeam Knowledge Base (KB) article 1446 provides instructions on how to configure SAN access for use with Veeam Backup & Replication.

Although Microsoft Windows 2003 had automount enabled by default, the SAN policy in the Windows Server 2008 R2 Enterprise and Datacenter editions is set to Offline Shared by default. However, even if an administrator changed the policy to Online All, Veeam reverts the policy back to Offline Shared. The result is that Windows does not mount and re-signature VMFS LUNs unless an administrator changes this setting again or manually mounts VMFS LUNs while ignoring Windows warnings. NetApp recommends the direct SAN access transport mode for VMs whose disks are on shared VMFS SAN LUNs that are connected to ESXi hosts through FC or iSCSI.

In the direct SAN access transport mode, Veeam Backup & Replication leverages VMware VADP to transport VM data directly from and to FC and iSCSI storage over the SAN. VM data travels over the

SAN, bypassing ESXi hosts and the LAN. The direct SAN access transport method provides the fastest data transfer speed and produces no load on the production network.

You can use the direct SAN access transport mode for all operations in which the backup proxy is engaged, including:

- Backup
- Replication
- Full VM restore
- VM disk restore
- Replica failback
- Quick migration

Requirements for the direct SAN access transport mode

To use the direct SAN access transport mode, you must meet the following requirements:

- The backup proxy that uses the direct SAN access transport mode must have direct access to the production storage through a hardware or software host bus adapter (HBA). If a direct SAN connection is not configured or is not available when a job or a task starts, the job or the task fails.
- For restore operations, the backup proxy must have write access to LUNs where the VM disks are located.

Limitations for the direct SAN access transport mode

The direct SAN access transport mode has the following limitations:

- The direct SAN access transport mode is not supported for VMs that reside on a virtual SAN (VSAN). You can use the Virtual Appliance or Network transport mode to process such VMs. For details about VSAN restrictions, see the VDDK 5.5 Release Notes.
- If at least one VM disk is on a VVol, you cannot use the direct SAN access mode.
- You can use the direct SAN access transport mode only for the initial run of the replication job. For subsequent replication job runs, Veeam Backup & Replication uses the Virtual Appliance or Network transport mode.
- You can use the direct SAN access transport mode to restore only thick VM disks.
- Because of VMware limitations, you cannot use the direct SAN access transport mode for incremental restore. You must either disable changed block tracking (CBT) for VM virtual disks for the duration of the restore process or select another transport mode for incremental restore.

For VMware vSphere 5.5 and later, integrated development environment (IDE) and SATA disks can be processed in the direct SAN access transport mode.

For VMware vSphere 5.1 and earlier, you should consider the following:

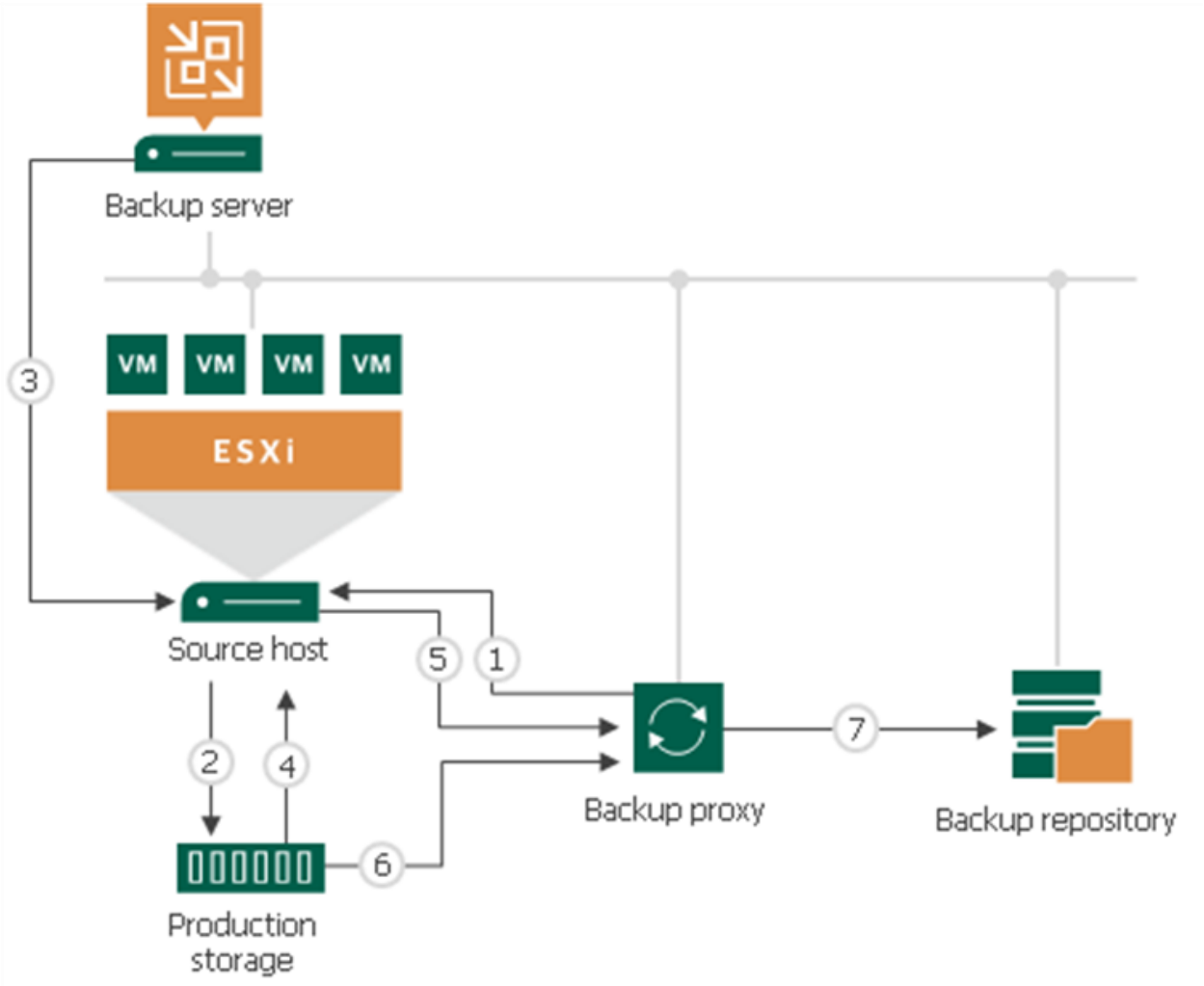
- IDE disks can be backed up in the direct SAN access transport mode. However, restore of IDE disks in the direct SAN access transport mode is not supported.
- If a VM disk fails to be processed in the direct SAN access transport mode, Veeam Backup & Replication does not fail over to the network mode.
- If a VM has some disks that cannot be processed in the direct SAN access transport mode, Veeam Backup & Replication uses the network mode for VM disk processing.

Data Backup in direct SAN access mode

To retrieve VM data blocks from a SAN LUN during backup, the backup proxy uses metadata about the layout of VM disks on the SAN; see Figure 9. Data backup in the direct SAN access transport mode includes the following steps:

1. The backup proxy sends a request to the ESXi host to locate the necessary VM on the datastore.
2. The ESXi host locates the VM.
3. Veeam Backup & Replication triggers VMware vSphere to create a VM snapshot.
4. The ESXi host retrieves metadata about the layout of VM disks on the storage (physical addresses of data blocks).
5. The ESXi host sends metadata to the backup proxy.
6. The backup proxy uses metadata to copy VM data blocks directly from the source storage over the SAN.
7. The backup proxy processes copied data blocks and sends them to the target.

Figure 9) Data backup in direct SAN access mode.

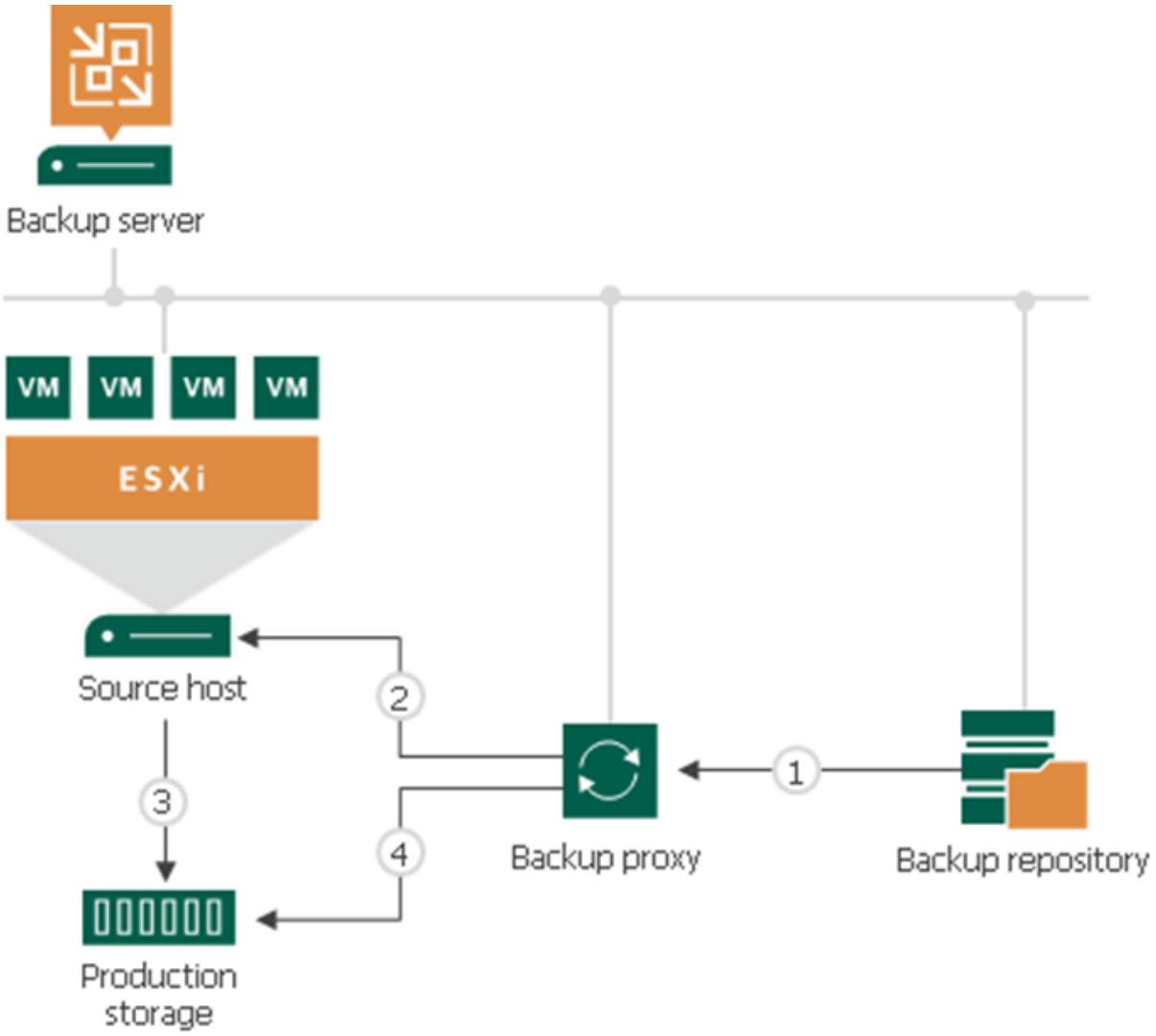


Data Restore in Direct SAN access mode

Data restore in the Direct SAN access transport mode includes the following steps (see Figure 10):

1. The backup proxy retrieves data blocks from the backup repository or a datastore in the target site.
2. The backup proxy sends a request to the ESXi host on the source site to restore data to a necessary datastore.
3. The ESXi host on the source site allocates space on the datastore.
4. Data blocks that are obtained from the backup proxy are written to the datastore.

Figure 10) Data restore in direct SAN access mode.



You can use the direct SAN access transport mode to restore VMs only with thick disks. Before VM data is restored, the ESXi host must allocate space for the restored VM disk on the datastore:

- When thick disks are restored, the ESXi host allocates space on the disk before writing VM data.
- When thin disks are restored, the ESXi host attempts to allocate space dynamically as requests for data block restores are received.

As a result, restoring thin disks involves extra allocation overhead when compared with restoring thick disks, which results in decreased performance.

To restore VMs with thin disks, you can use the Virtual Appliance mode or the network mode. If you plan to process a VM that has both thin and thick disks, select the direct SAN access transport mode and choose to fail over to the network mode if the SAN becomes inaccessible. In that case, Veeam Backup & Replication uses the direct SAN access transport mode to restore thick disks and uses the Network transport mode to restore thin disks. Alternatively, you can restore all VM disks as thick.

Off-host backup (Hyper-V)

In off-host backup mode, backup processing is shifted from the source Hyper-V host to a dedicated machine: an off-host backup proxy. The off-host backup proxy acts as a data mover. The Veeam transport service that runs on it retrieves VM data from the source datastore, processes it, and transfers it to the destination. This type of backup does not impose load on the Hyper-V host; while resource-intensive backup operations are performed on the off-host backup proxy, production hosts remain unaffected.

To perform off-host backup, Veeam Backup & Replication uses transportable shadow copies. Transportable shadow copy technology lets you create a snapshot of a data volume on one server and import, or mount, it onto another server within the same subsystem (SAN) for backup and other purposes. The transport process is accomplished in a few minutes, regardless of the amount of data. The process is performed at the SAN storage layer, so it does not affect host CPU usage or network performance.

To perform off-host backup, you must meet the following requirements:

1. You must configure an off-host backup proxy. You can assign the role of an off-host backup proxy only to a Windows Server machine with the Hyper-V role enabled.

The version of the Hyper-V host and off-host backup proxy must be the same. For example, you use a Microsoft Windows 2022 machine with the Hyper-V role enabled as a Hyper-V host. In that case, you should deploy the off-host backup proxy on a Microsoft Windows 2022 machine with the Hyper-V role enabled.

2. In the properties of a backup or replication job, you must select the off-host backup method. If necessary, you can point the job to a specific proxy.
3. The source Hyper-V host and the off-host backup proxy must be connected through a SAN configuration to the shared storage.
4. To create and manage volume shadow copies on the shared storage, you must install and properly configure a Volume Shadow Copy Service (VSS) hardware provider that supports transportable shadow copies on the off-host proxy and the Hyper-V host. Typically, when configuring a VSS hardware provider, you must specify a server that controls the LUN and disk array credentials to provide access to the array.

The VSS hardware provider is usually distributed as part of the client components that are supplied by the storage vendor. Any VSS hardware provider that is certified by Microsoft is supported. Some storage vendors might require additional software and licensing to work with transportable shadow copies.

5. If you back up VMs whose disks reside on Cluster Shared Volumes (CSV) with data deduplication enabled, make sure that you:
   a. Use a Microsoft Windows Server machine as an off-host backup proxy.
   b. Enable data deduplication on this off-host backup proxy.

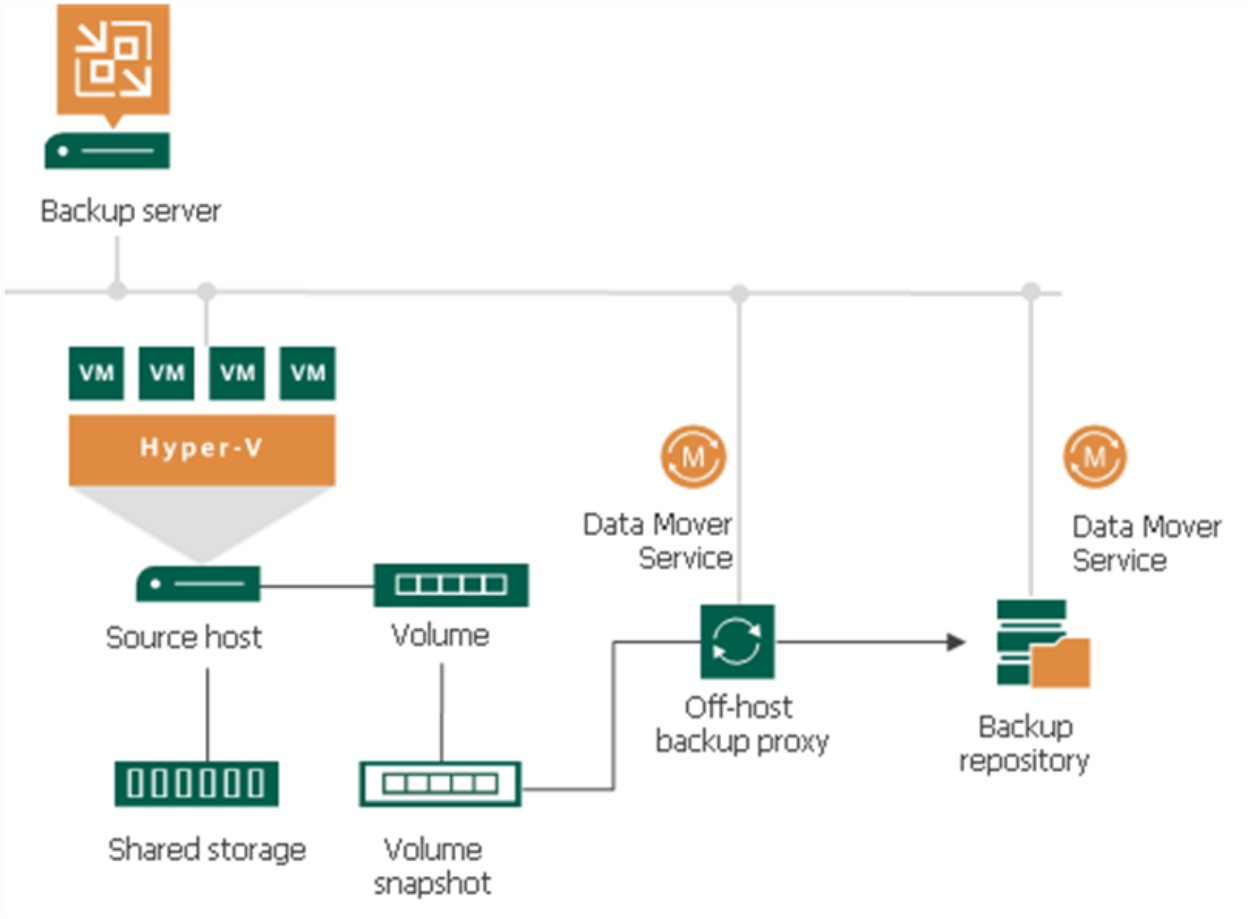
Otherwise, off-host backup fails.

The off-host backup process (see Figure 11) includes the following steps:

1. Veeam Backup & Replication triggers a snapshot of the necessary volume on the production Hyper-V host.

2. The created snapshot is split from the production Hyper-V server and is mounted to the off-host backup proxy.
3. The Veeam transport service that runs on a backup proxy uses the mounted volume snapshot to retrieve VM data. The VM data is processed on the proxy server and is copied to the destination.
4. When the backup process is complete, the snapshot is dismounted from the off-host backup proxy and is deleted from the SAN.

Figure 11) Process of off-host backup mode.



Note: If you plan to perform off-host backup for a Hyper-V cluster with CSV, deploy an off-host backup proxy on a host that is not part of a Hyper-V cluster.

When a volume snapshot is created, this snapshot has the same LUN signature as the original volume. The Microsoft Cluster Service does not support LUNs with duplicate signatures and partition layout. For this reason, volume snapshots must be transported to an off-host backup proxy outside the cluster. If the off-host backup proxy is deployed on a node of a Hyper-V cluster, a duplicate LUN signature is generated, and the cluster fails during backup or replication.

Off-Host backup proxy

By default, when you perform backup, replication, or VM copy jobs in the Hyper-V environment, VM data is processed directly on the source Hyper-V host where the VMs reside. The data is then moved to the target, bypassing the Veeam backup server.

VM data processing can produce unwanted overhead on the production Hyper-V host and can affect the performance of the VMs that run on this host. To take data processing off the production Hyper-V host, use the off-host backup mode.

The off-host mode shifts the backup and replication load to a dedicated machine: an off-host backup proxy. The off-host backup proxy functions as a data mover that retrieves VM data from the source datastore, processes it, and transfers it to the destination.

The machine that performs the role of an off-host backup proxy must meet the following requirements:

- The role can be assigned only to a Windows Server machine with the Hyper-V role enabled.
- The off-host backup proxy must have access to the shared storage that hosts the VMs to be backed up, replicated, or copied.
- To create and manage volume shadow copies on the shared storage, you must install a VSS hardware provider that supports transportable shadow copies on the off-host proxy and the Hyper-V host. The VSS hardware provider is usually distributed as part of the client components that are supplied by the storage vendor.

When you assign the role of an off-host backup proxy to the selected machine, Veeam Backup & Replication automatically installs on it the lightweight components and services that are required for backup proxy functioning. Unlike the Veeam backup server, backup proxies do not require a dedicated SQL Server database; all the settings are stored centrally, within the SQL Server database that the Veeam backup server uses.

To enable a Hyper-V host or a Windows machine to act as an off-host backup proxy, Veeam Backup & Replication installs the following services on it:

- Veeam Installer Service is an auxiliary service that is installed and started on any Windows (or Hyper-V) server after it is added to the list of managed servers in the Veeam Backup & Replication console. This service analyzes the system and installs and upgrades necessary components and services.
- Veeam Transport is responsible for deploying and coordinating executable modules that act as data movers and that perform the main job activities on behalf of Veeam Backup & Replication. These activities include performing data deduplication, compression, and so on.
- Veeam Hyper-V Integration Service is responsible for communicating with the VSS framework during backup, replication, and other jobs and for performing recovery tasks. The service also deploys a driver that handles changed block tracking for Hyper-V.

Veeam Backup & Replication components

Table 3 describes each component in Veeam Backup & Replication.

Table 3) Veeam Backup & Replication components and their description.

Component	Description
Veeam 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 Veeam backup server performs all types of administrative activities: • Coordinates backup, replication, recovery verification, and restore tasks • Controls job scheduling and resource allocation • Is used to set up and manage backup infrastructure components and to specify global settings for the backup infrastructure

Component	Description
	• In addition to its primary functions, a newly deployed Veeam backup server also performs the roles of the default backup proxy and the backup repository (it manages data handling and data storing tasks).
Veeam Backup Service	This Windows service coordinates all the operations that are performed by Veeam Backup & Replication, such as backup, replication, recovery verification, and restore tasks. The Veeam Backup Service runs under the local system account or the account that has the local administrator permissions on the Veeam backup server.
Veeam Backup Shell	The Veeam Backup Shell provides the application UI and enables user access to the application's functionality.
Veeam Backup Catalog Service	This Windows service manages guest OS file system indexing for VMs and replicates system index data files to enable search through guest OS files. Index data is stored in the Veeam Backup Catalog, a folder on the Veeam backup server. The Veeam Backup Catalog Service running on the Veeam backup server works with search components that are installed on Veeam Backup Enterprise Manager and (optionally) a dedicated Microsoft Search Server.
Veeam Backup SQL Server database	This database is used by Veeam Backup Service, Veeam Backup Shell, and Veeam Backup Catalog Service to store data about the backup infrastructure, jobs, sessions, and so on. The database instance can be on a SQL Server installed either locally (on the same machine where the Veeam backup server runs) or remotely.
Veeam Backup PowerShell snap-in	This snap-in is an extension for Microsoft Windows PowerShell 2.0. Veeam Backup PowerShell adds a set of cmdlets to enable you to perform backup, replication, and recovery tasks through the PowerShell CLI or run custom scripts to fully automate Veeam Backup & Replication operation.
Backup proxy services	In addition to dedicated services, the Veeam backup server runs a set of Data Mover Services.
Backup proxy	A backup proxy is an architecture component that sits between the data source and the target and is used to process jobs and deliver backup traffic. The backup proxy retrieves VM data from the production storage. It also compresses the retrieved data and sends it 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). As the data handling task is assigned to the backup proxy, the Veeam backup server becomes the point of control for dispatching jobs to proxy servers.
Backup repository	A backup repository is a location that Veeam Backup & Replication jobs use to store backup files, copies of VMs, and metadata for replicated VMs. Technically, a backup repository is a folder on the backup storage. By assigning different repositories to jobs and by limiting the number of parallel jobs for each one, you can balance the load across your backup infrastructure.

E-Series as storage for Veeam Hardened Repository

To protect your backup data and prevent data loss because of malware activity or unplanned actions, you can add a hardened repository based on a Linux server. The hardened repository supports the following features:

- Immutability: when you add a hardened repository you specify the days where the backup file is immutable. During this time the backup file cannot be modified or deleted.
- Single-use credentials: These are the credentials Veeam uses only once to deploy the services like Veeam Data Mover or transport service when you add the Linux server to the backup infrastructure for the first time. These credentials are not stored in the configuration database. Even if the Veeam Backup & Replication server is compromised, the attacker cannot get the credentials and connect to the hardened repository.

A limitation is that the hardened repository, per best practice policies, must only be used as a repository. For this reason, you cannot assign other roles to the Linux server. It is possible to use the hardened repository server as a VMware backup proxy (Network mode only), but it is not recommended. For complete documentation please see Hardened Repository.

Below is an abbreviated list of requirements for a Linux Hardened Repository:

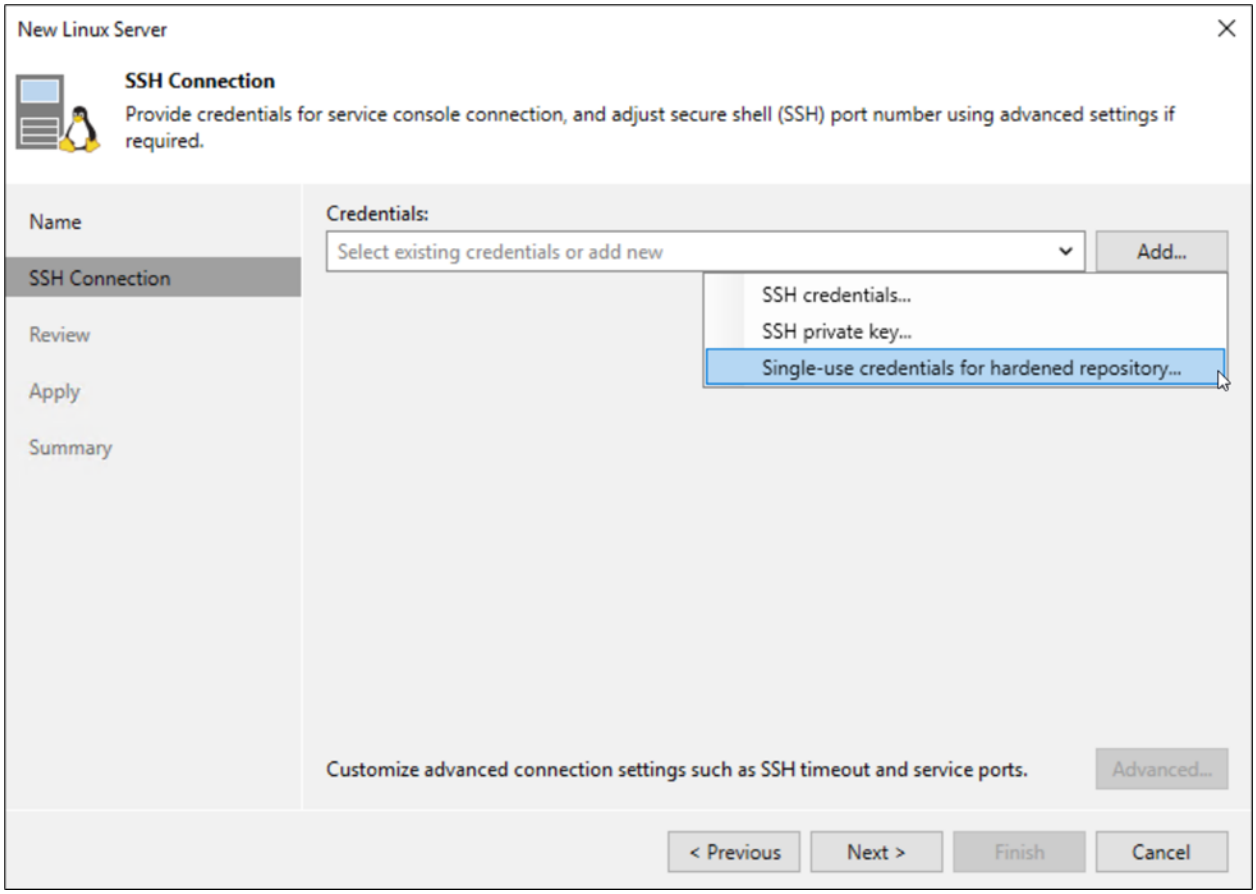
- Use supported Linux distribution.
- Requires block storage (no NFS, Shares, etc.) because the functions for immutable backups are based on filesystem level features.
- Use XFS with Extended Attributes to also leverage fast block cloning.
- While iSCSI as Data Protocol is technically possible – it is not recommended in case of performance and security. The recommended way to connect your E-Series to the Repository Server is to direct attach via SAS or Fibre Channel. At Scale also zoning via FC switch is possible.
- Do not use a VM, plan for a dedicated hardware server with an HBA.
- Don’t forget OS hardening, for example using DISA STIG profiles.
- As a Best Practice disable SSH and remove sudo permissions after the installation, but keep in mind that you have to do extra work to log into the system.

For a full list of system requirements, please see system requirements for Hardened Repository.

Below is not a step-by-step guide on how to configure the repository, but the steps highlight some key features of the concept. A step-by-step guide can be found here.

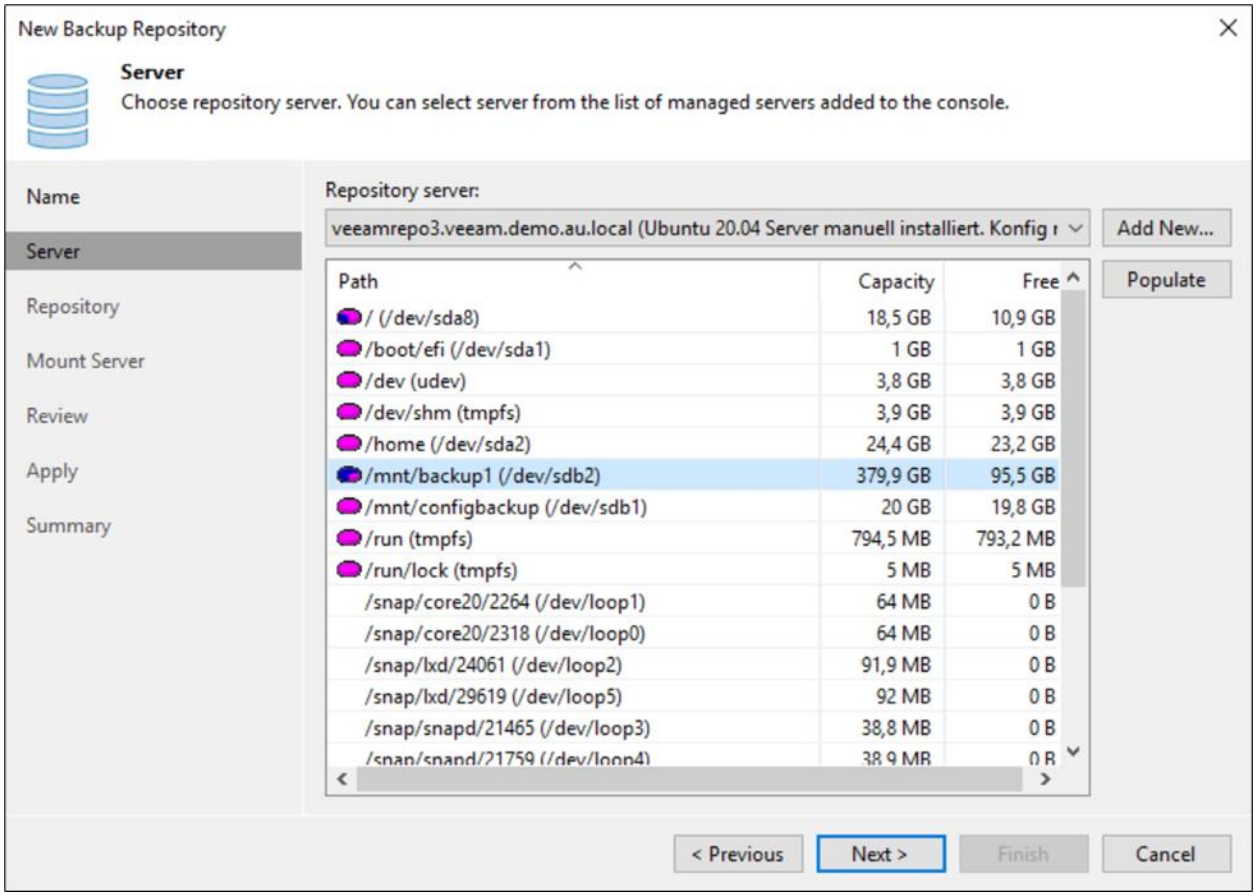
When you add the Linux host for the first time Veeam will prompt for credentials. Use the single-use credentials here, as these credentials will not be stored in the configuration database, see Figure 12.

Figure 12) Login credentials selection for hardened repository.



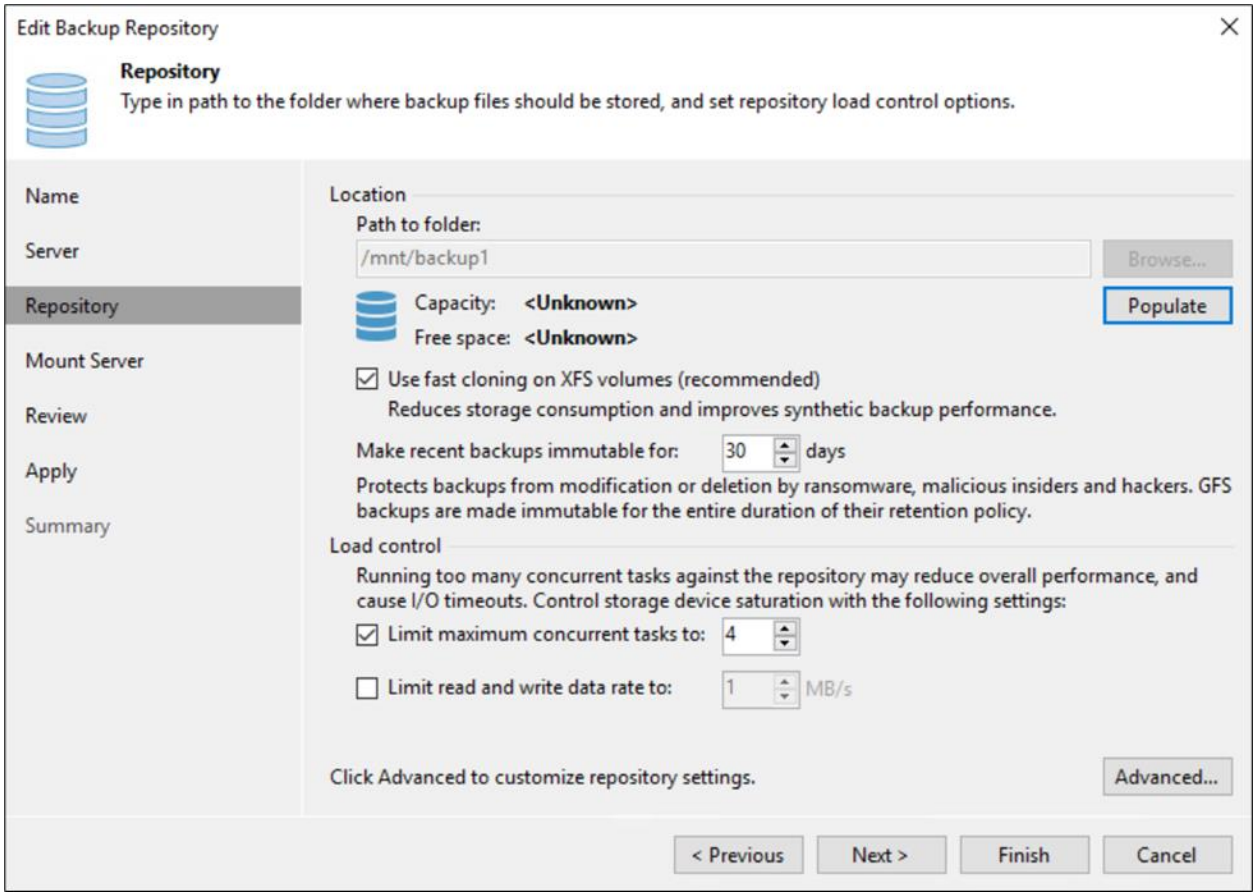
After adding the Linux host you can deploy the repository to the server. In this example the storage was mounted under “/mnt/backup1”, see Figure 13. Ensure the LUN has been mounted to the Linux server and that it has been formatted with XFS.

Figure 13) Mounting repository to Linux server.



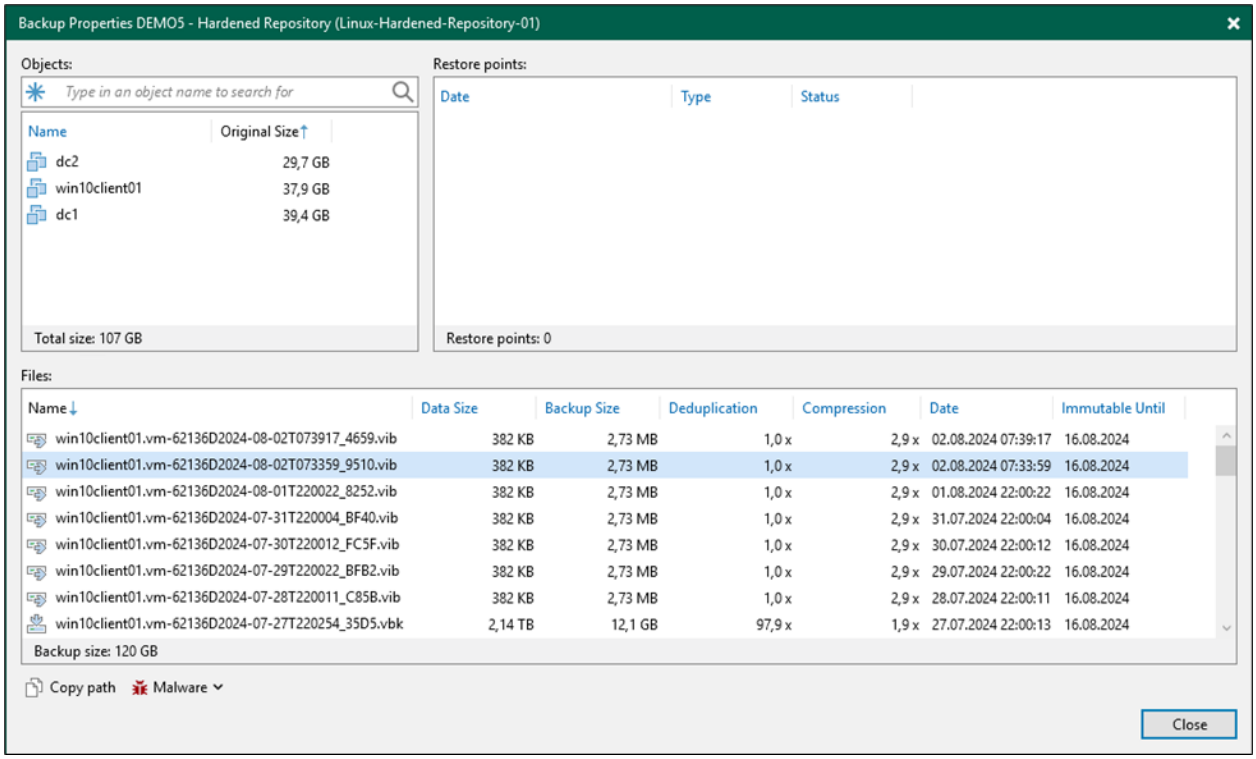
The Immutability duration is defined on the repository level. There is no easy way to delete backups in case the repository is running out of space so it may be best to start with less days and modify it afterwards to a higher value. See Figure 14.

Figure 14) Selecting immutability duration.



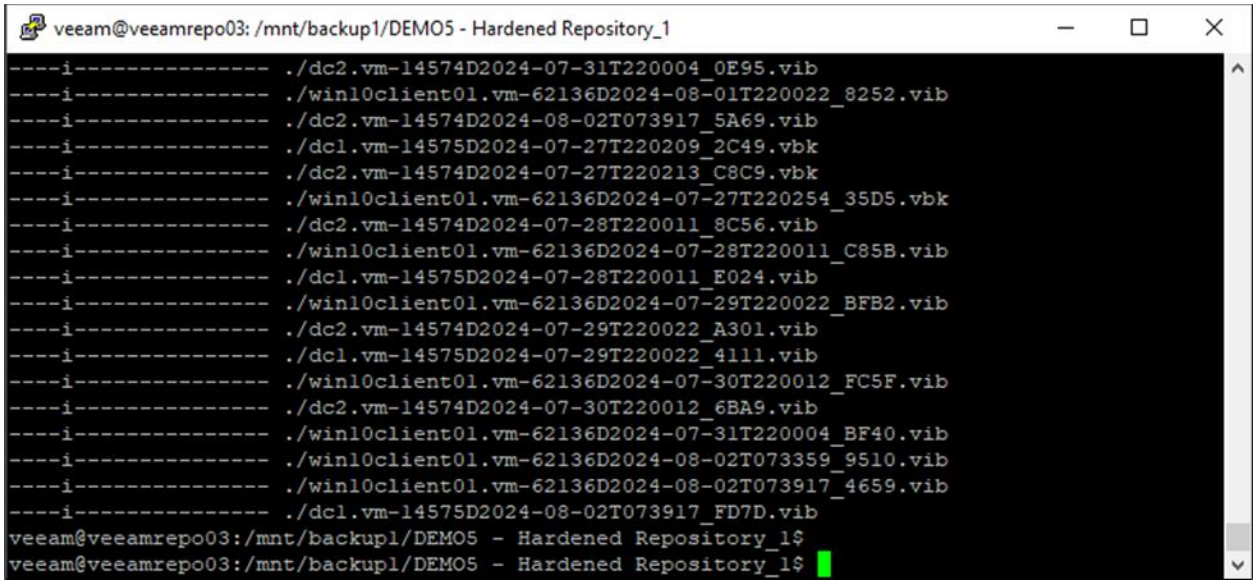
Once the repository setup has been completed the “Immutable Until” date is filled. It’s expected that the date here is longer than the configured immutability, for details see Block Generation. See Figure 15.

Figure 15) Immutable Until date listed.



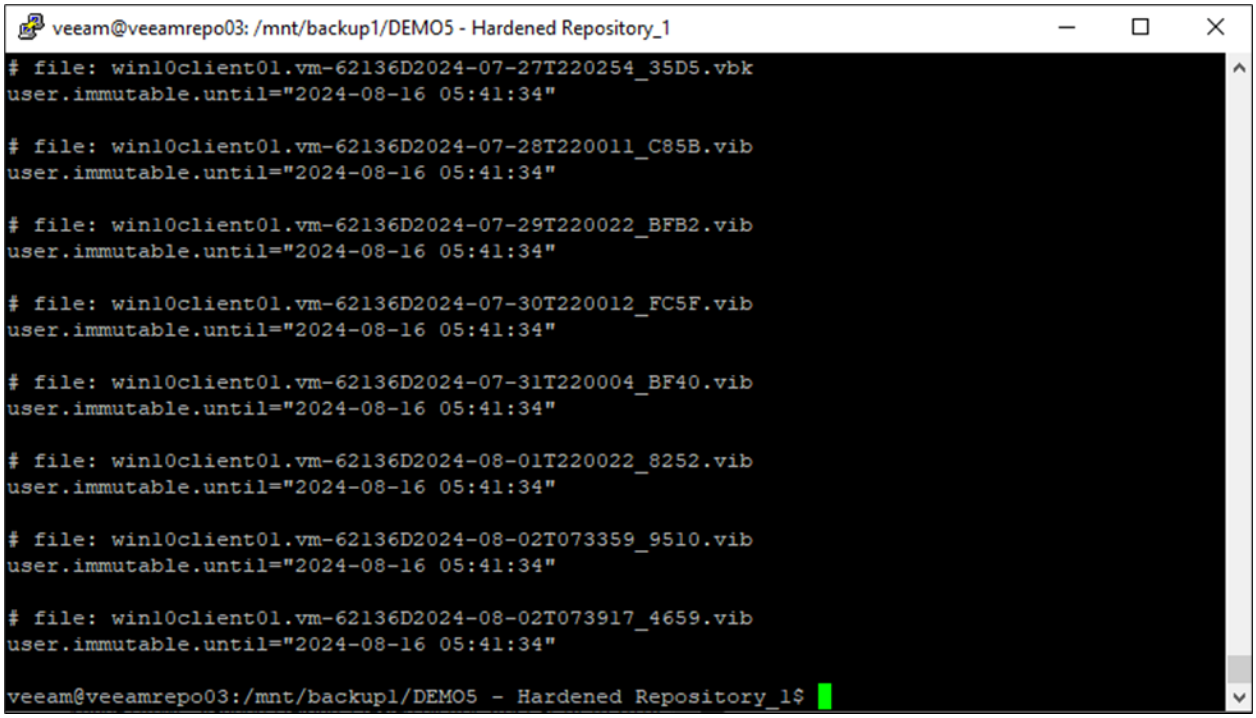
To verify immutability on the XFS filesystem, the “lsattr” command can be run on the command line interface (CLI). The immutability flag, “i” should be shown for the appropriate files, see Figure 16.

Figure 16) Immutability flag set.



Use command “getfattr * -n user.immutable.until” in Linux CLI to verify the same for the same immutability end time that is reported in the Veeam console, see Figure 17.

Figure 17) Linux immutability end time.



NetApp E-Series volume configuration guidelines

For optimal performance, NetApp recommends that you follow these guidelines:

- Use RAID 6 (8+2) volume groups.
- Create multiple volume groups, having an even number when possible so that you can achieve balance between owning controllers. Make sure of drawer loss protection when applicable to reduce the risk of data unavailability and/or data loss.
- Create a single standard (not thin) volume per volume group.
- When creating the volumes, select a 512KB segment size to match the 512KB transfer size that Veeam Backup & Replication presents as a sequential write to the system after sequencing I/O.
- Run multiple backup jobs to each repository.

Note: The preceding setup can be configured without any hot spares too, to achieve better performance in case of drive failure. A caveat is that data loss might occur if more than two drives fail at once.

For a large configuration (configuration with 60 drives or more), NetApp recommends that you follow these guidelines:

- Use dynamic disk pools (DDP) technology to maximize ease of use and for fast rebuild times.
- Create multiple dynamic disk pools (15 or 30 disk pools), having an even number when possible so that you can achieve balance between owning controllers. Make sure of drawer loss protection when applicable to reduce the risk of data unavailability and/or data loss.

- Create a single standard (not thin) volume per DDP.
- When creating the volumes with DDP, the default segment size is 128KB. No additional selection is required.
- Run multiple backup jobs to each repository.

Note: RAID 10 can also be considered as an option for volume configuration, but it is not recommended due to a single parity and capacity penalty. Also, write throughput performance of RAID 6 and DDP is better in comparison to RAID 10.

Note: NetApp recommends avoiding thin volumes while setting up NetApp E-Series with Veeam.

NetApp E-Series host configuration guidelines

Host connectivity

NetApp E-Series arrays provide multiple options for connectivity.

Host connectivity for E2800 and E5700

The E2800 controller has the following base hardware features:

- Dual Ethernet ports for management-related activities
- Dual 12Gb SAS drive expansion ports to attach expansion drive shelves

Note: Some older E2800 and E5700 controllers also have either two optical FC/iSCSI or two RJ-45 iSCSI baseboard ports for host connection.

Note: For host connectivity a host interface card (HIC) is required for the E2800 and E5700 controllers.

For all host connectivity options please reference the Introduction to NetApp E-Series E2800 arrays and the Introduction to NetApp E-Series E5700 arrays.

Your optimal design is based on your existing environment and how the data flows from primary storage through the Veeam server to its destination on the E-Series array. One design approach is to leverage 12Gb SAS for connectivity from the Veeam server to the E-Series array. This approach provides the target backup repository (E-Series) to Veeam as a direct connection. Having a direct connection to the E-Series array prevents having to write the backup file across an existing network.

This configuration might not be ideal for larger environments in which Veeam's distributed architecture is implemented and in which multiple proxy servers process backup data. Having a dedicated network for backup targets might make more sense in those cases. Either way, the options are there for any environment, and Veeam provides a bottleneck detector to help optimize the backup data flow as you progress through your implementation.

NetApp E-Series storage host mapping configuration for direct SAN access

On the E-Series system that hosts the virtual environment's storage, verify that the Veeam backup server is associated with the same host group as the host of the virtual environment. See the example shown for NetApp SANtricity in Figure 18.

Host_Cluster_Veeam is the host cluster that has the ESXi Server (Host_ESXi_Server) and the Veeam Server (Host_Veeam_Server). As seen, all 16 volumes are shared between both the host members.

Figure 18) Backup server mapping.

Create ▾ | Assign Volumes | Unassign Volumes | View/Edit Settings | Delete

Name	Type	Associated Objects	Total Assigned Volumes	Reported Capacity (GiB)	Host Type	Edit
− Default Cluster	Default Cluster	1 Host Cluster(s) 2 Host(s)	16	23600.00	Windows	
− Host_Cluster_Veeam	Cluster	Default Cluster 2 Host(s)	16	23600.00	Windows	
Host_ESXi_Server	Host Member	Host_Cluster_Veeam	16	23600.00	Windows	
Host_Veeam_Server	Host Member	Host_Cluster_Veeam	16	23600.00	Windows	

Total rows: 4

As a result of hosting both the virtual environment and the backup repository on the E-Series arrays, Veeam Backup & Replication has resource location awareness in the environment. The resulting data transfers take place over the SAN.

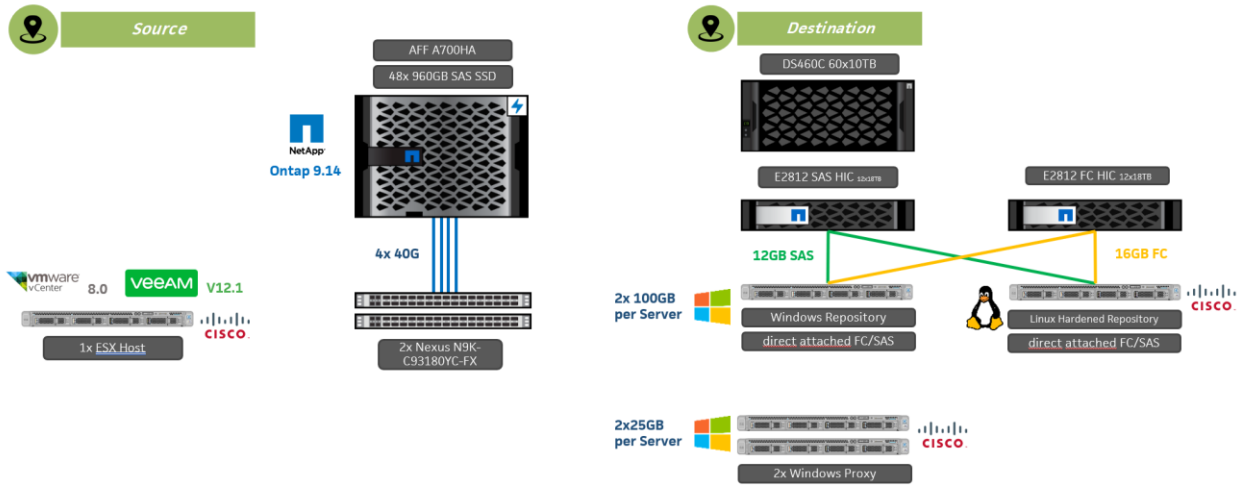
Performance with Veeam and NetApp E-Series

The performance testing conducted by Advanced UniByte with NetApp E2800 arrays and Veeam proved to be a robust solution that meets the backup and restore needs of modern data centers. The testing objective was to determine the best practices in addition to those previously highlighted using different host protocols, different drive quantities and different volume (LUN) configurations.

Test environment and setup

Figure 19 provides a graphical display of the test environment setup.

Figure 19) Test environment diagram.



The test environment consisted of two parts, the source and the destination configurations. The source configuration consisted of the following:

- A700 running ONTAP 9.14.

- 1x Cisco C220 M6 with 2x Intel Gold 6362 processors, 384GB RAM and VIC 1467 10/25GB iSCSI host adapters.
- Windows Server 2022 with ReFS installed on Cisco servers.
- VMware ESXi 8.0 OS installed on Cisco server with 4 iSCSI datastores, 4 virtual machines (VM) per datastore, each VM using 50GB.
- Veeam 12.1 (Build 12.1.1.56) virtual install with 2x Cisco servers running Windows Server 2022 used as Veeam proxies.
- Veeam compression was disabled (only for benchmarking to get maximum throughput on the repository).

The destination configuration consisted of the following:

- 1x E2812 running SANtricity 11.80GA with 12Gb, 4-port SAS HICs and 12x 18TB NL-SAS drives.
- 1x E2812 running SANtricity 11.80GA with 16Gb, 4-port FC HICs and 12x 18TB NL-SAS drives.
- 1x DE460C with 60x 10TB NL-SAS drives used as expansion when needed.
- Dynamic Load Balancing was disabled on both E2812s.
- 2x Cisco C220 M6 with 2x Intel Gold 6362 processors and 384GB RAM, one server running Windows Server 2022 with ReFS and one running Ubuntu 20.04 with XFS for Linux Hardened Repositories.

Note: See Veeam Backup & Replication 12 – Hardened Repository for more information on Linux Hardened Repositories.

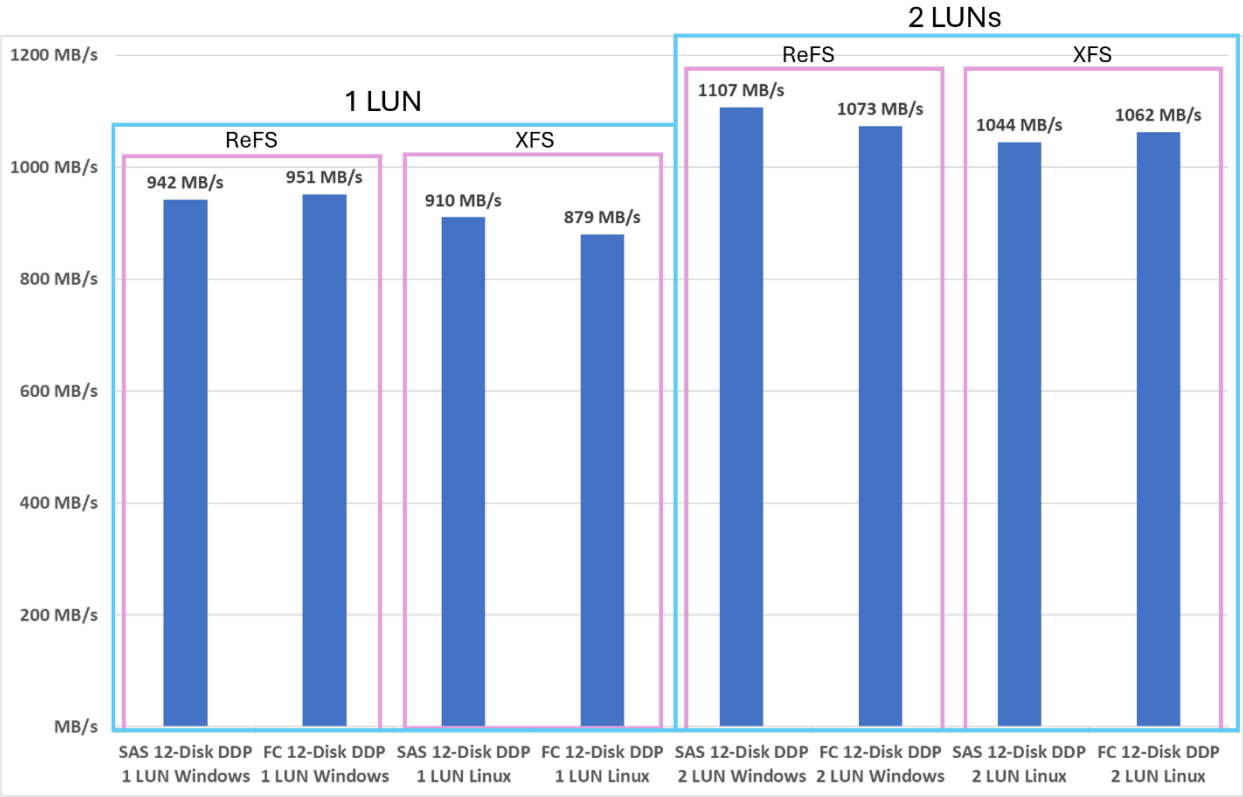
LUNs/volumes were equally distributed across the controllers, multiple LUNs were used as Scale-Out repository in Veeam and LUNs were mounted to the operating systems as one LUN equaled one device.

File system comparison

Figure 20 below compares throughput when using Windows (ReFS) and Linux (XFS) file systems on the E2812 volumes.

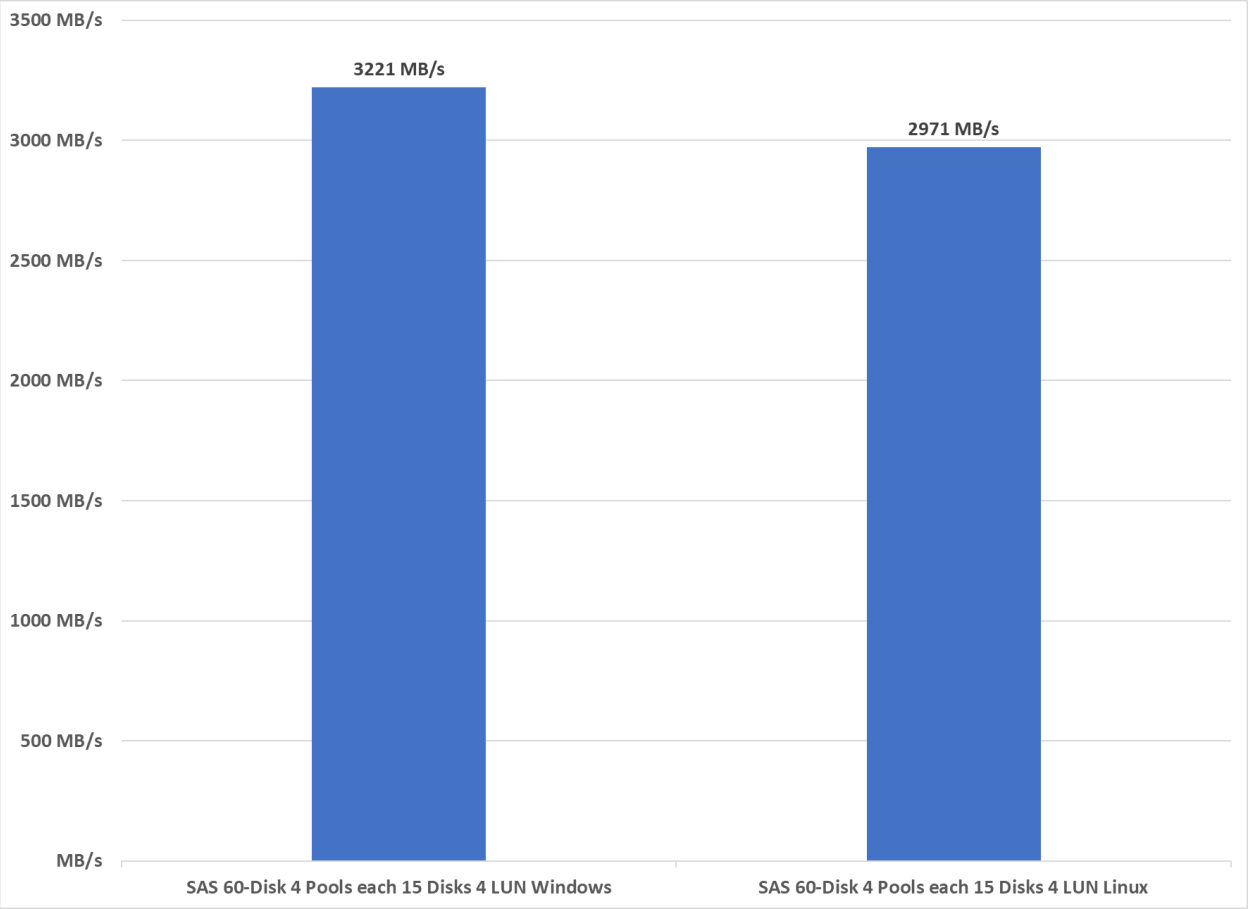
From the tests, Linux was a bit slower than Windows with only 12 disks in the pool.

Figure 20) Throughput comparison between Windows (ReFS) and Linux (XFS) with 12-disks.



Another set of comparison tests between Windows (ReFS) and Linux (XFS) were run with four 15-disk pools, one volume per pool, using the SAS array only. Figure 21 shows the results for these tests.

Figure 21) Throughput comparison between ReFS and XFS with 60-disks.



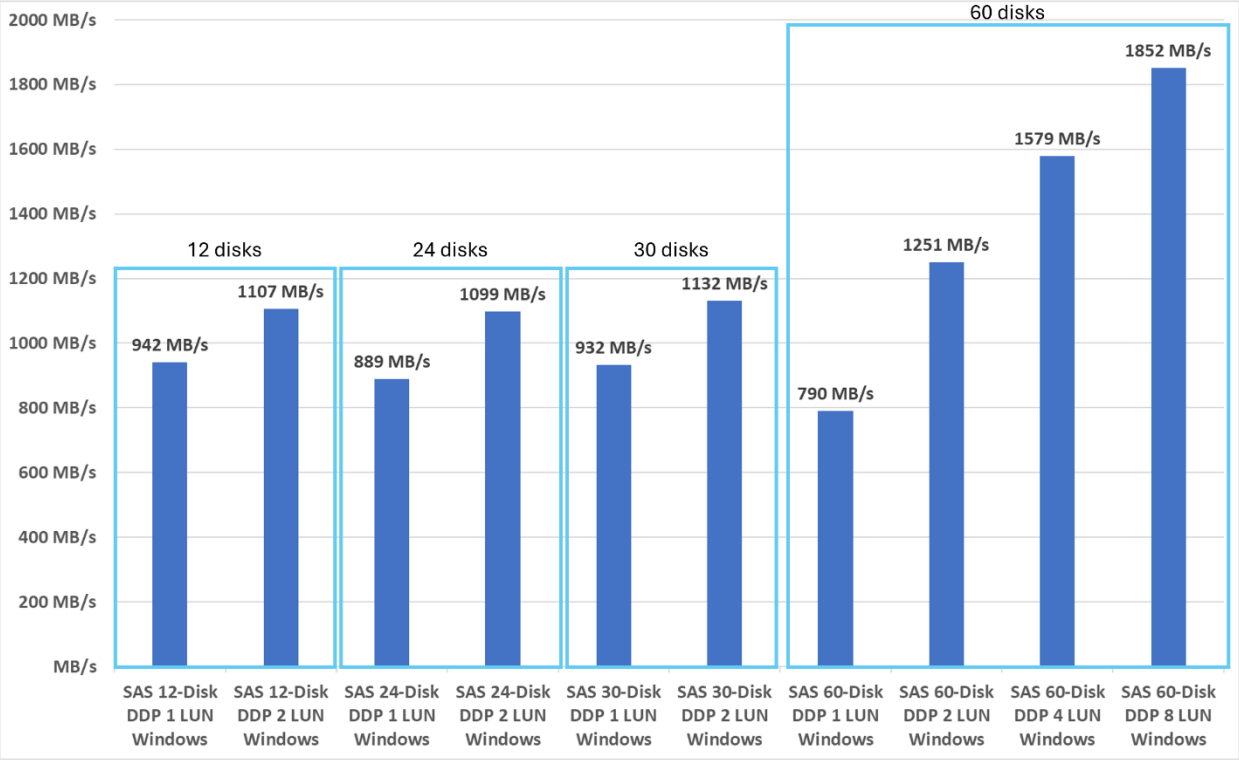
Again, the results above show that XFS is slower than ReFS when using pools made up with small number of disks, in this case 4x 15 disks in a dynamic disk pool (DDP).

Scaling throughput

Disk scaling

Figure 22 below shows throughput after increasing the number of disks per pool. The tests were run on the SAS array using dynamic disk pools with Windows host. There is no speed improvement by adding more disks while only using one LUN. As a result of this, with 60 disks or more, either use multiple LUNs or, even better, smaller RAID/DDP sets with multiple LUNs.

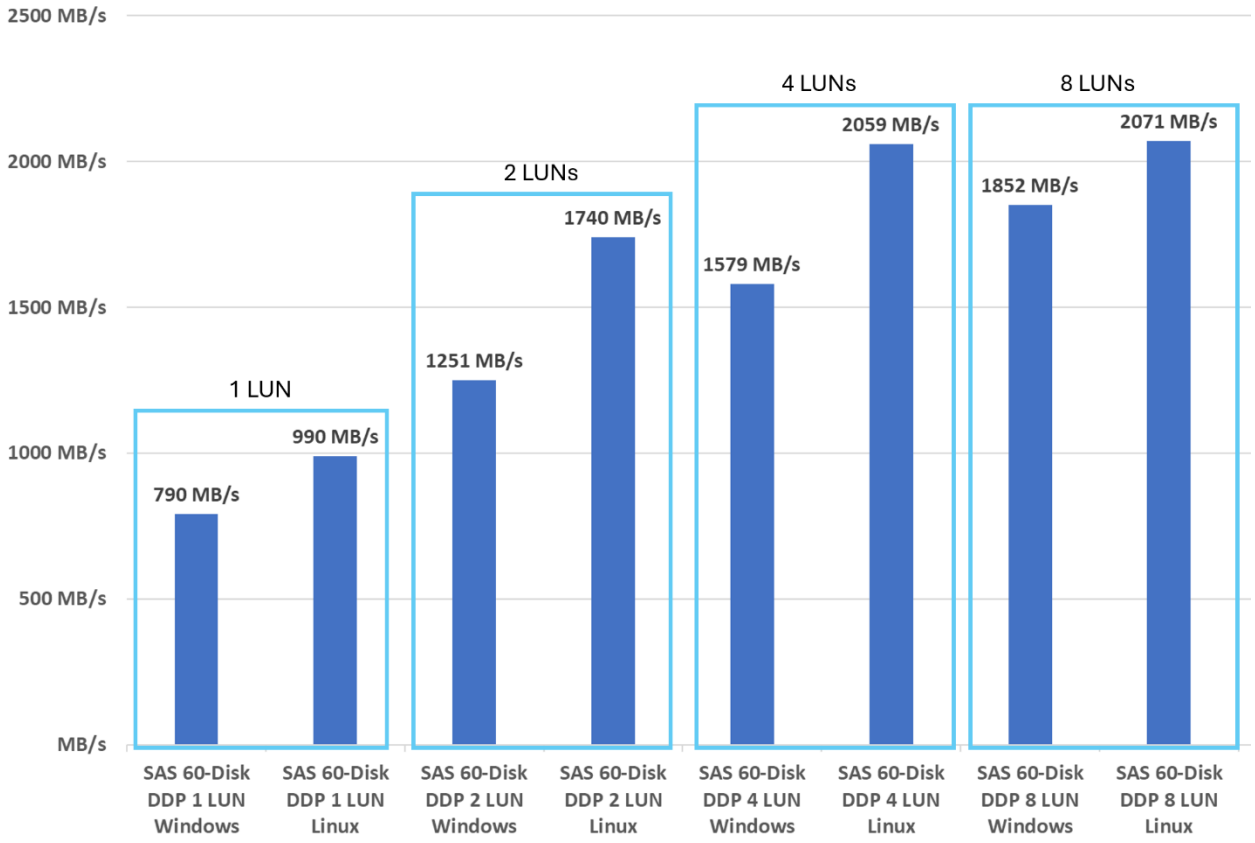
Figure 22) Throughput when increasing disks in a pool.



LUN/volume scaling

Figure 23 shows throughput with 60-disk pool but increasing the number of LUNs/volumes assigned to either Windows or Linux hosts. Again, the tests were run using the SAS array.

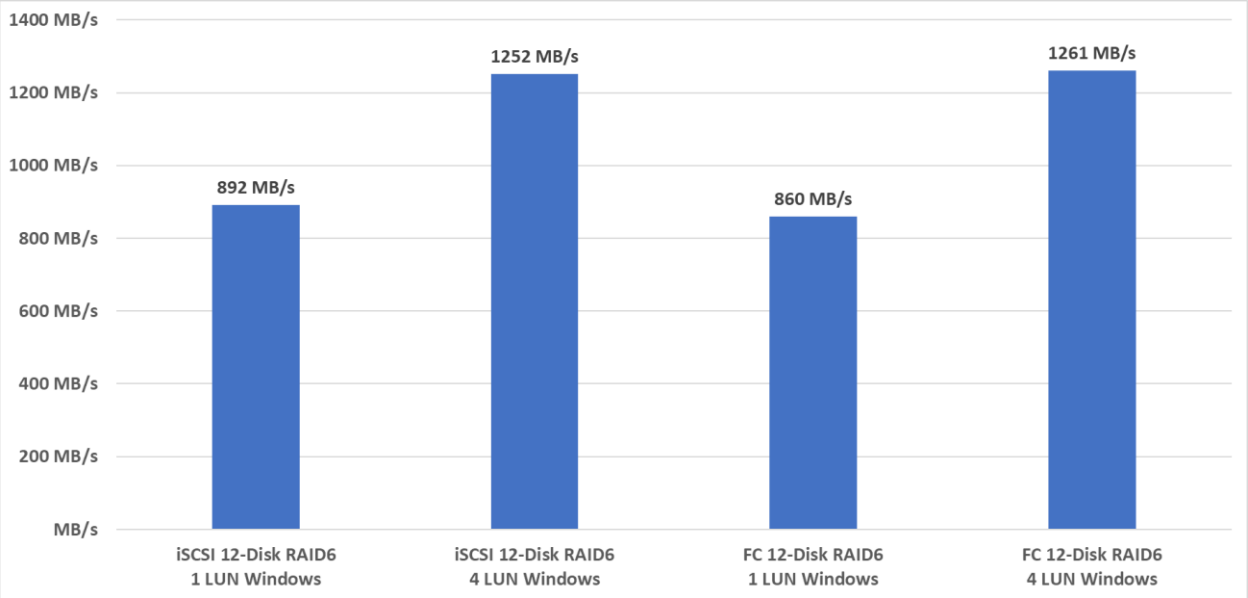
Figure 23) Throughput when increasing LUNs in a pool.



With a 60-disk DDP, XFS on Linux is faster than ReFS on Windows, which is opposite to what was observed when using 12-disk DDP, see Figure 20. From the above results, it shows XFS benefits from more disks within a RAID set.

Although most tests were performed using disk pools, there were comparison tests run using RAID 6 volume groups. The tests were run with 12-disk volume groups using Windows host to see the difference between 1 and 4 LUNs as well as throughput difference between 16Gb FC and 10Gb iSCSI. Jumbo frames were not enabled for the iSCSI tests. Results from this test can be seen below in Figure 24.

Figure 24) RAID6 FC versus iSCSI.



The results above show there isn't much difference between 16Gb FC and 10Gb iSCSI but there is improvement in throughput when going from one volume to four volumes, because both controllers are being utilized and the I/O load is spread across more drives.

Simulated tests

Advanced UniByte also performed simulated workload testing using **diskspd**, see Veeam KB article KB2014: How to Simulate Veeam Backup & Replication Disk I/O. The tests were run against the E2812 with SAS HIC.

The "tasks" referenced in Figure 26 and Figure 28 were separate command prompt windows each running **diskspd**. If more **diskspd** windows were open than there were LUNs available, that means multiple **diskspd** instances were run against a single LUN.

Simulated active full backup

Figure 25 shows the command used with **diskspd** to simulate Veeam active full backup load.

Figure 25) Diskspd simulated full command.

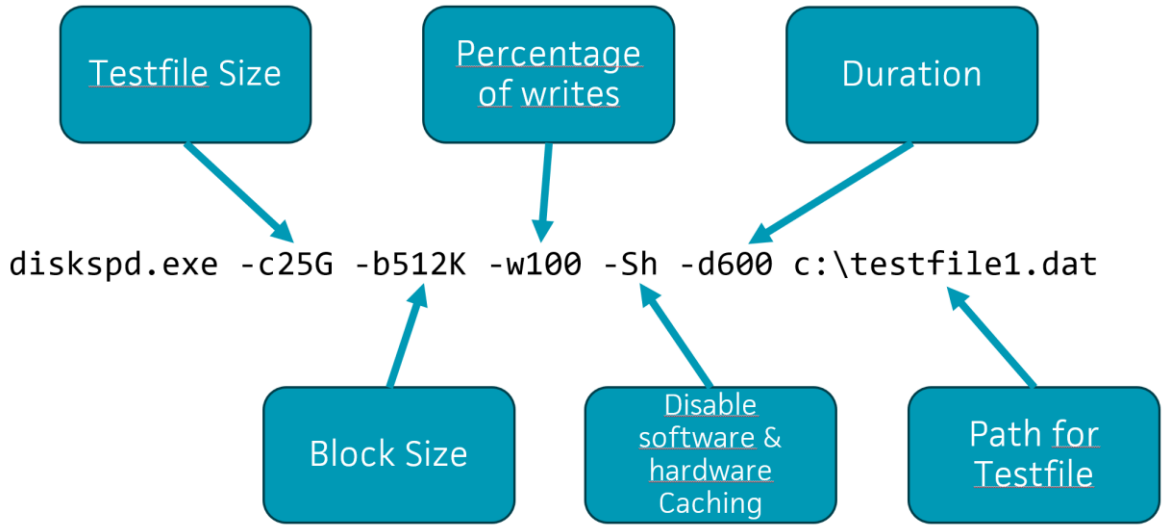
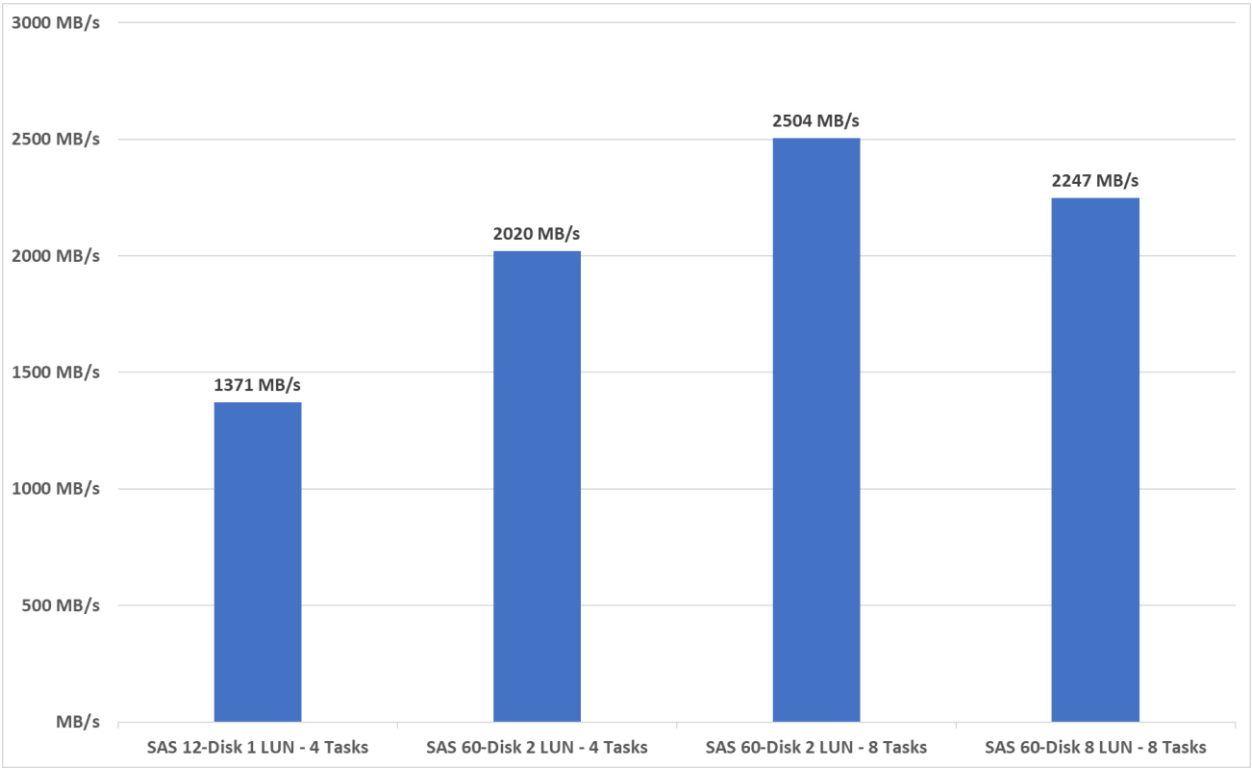


Figure 26 shows the results of the simulated active full backup tests. The 2247 MB/s for the 60-disk, 8 LUN test below is close to the 60-disk, 8 LUNs test that achieved 2071 MB/s, see Figure 23.

Figure 26) Diskspd simulated full results.



Simulated active restore

Figure 27 shows the command used with **diskspd** to simulate Veeam restore I/O load.

Figure 27) Diskspd simulated restore command.

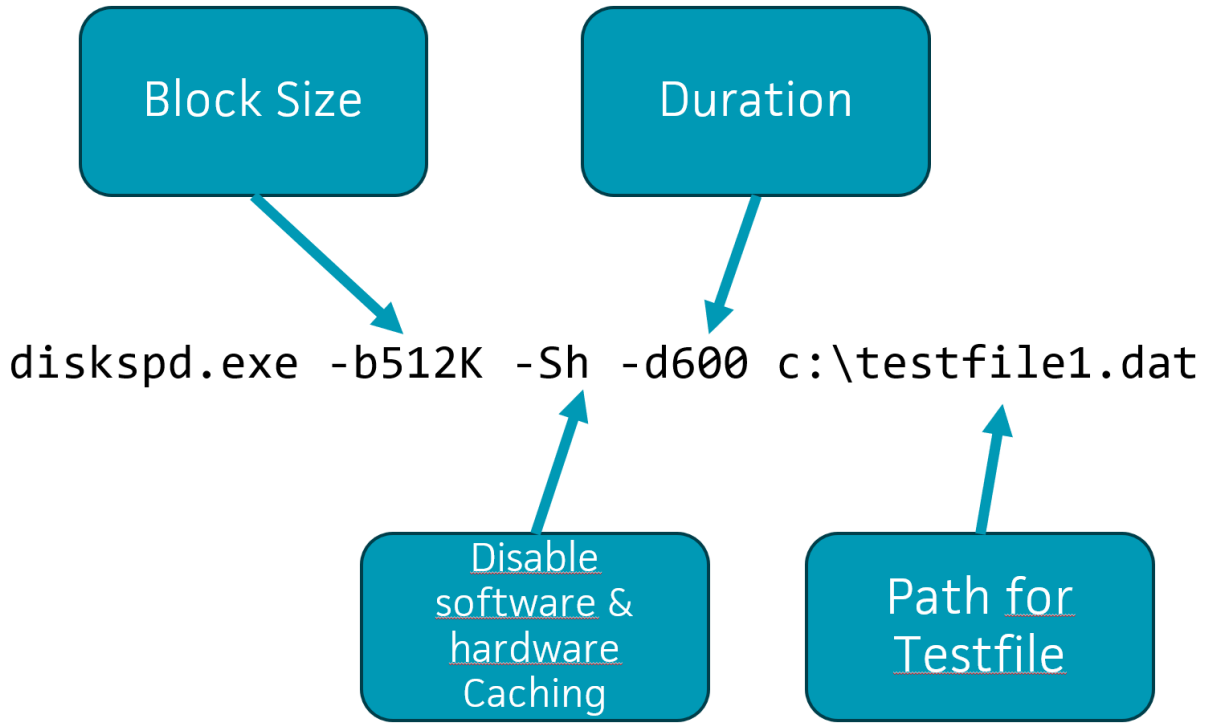
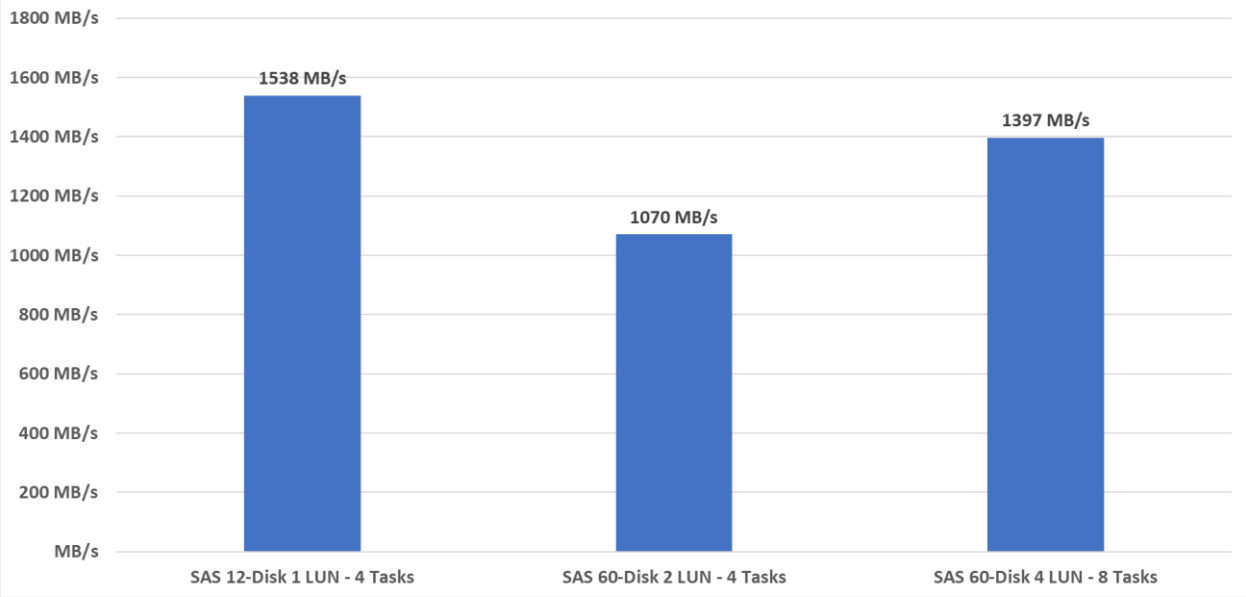


Figure 28 shows the results of the simulated active restore tests.

Figure 28) Diskspd simulated restore results.



Key Take-aways

Several key take-aways were found during the Advanced UniByte testing:

- An obvious one was poor performance while volumes were initializing, which can take many hours to days to complete, depending on the volume size and disk type used.
- Poor performance without cache, which was discovered because one controller had it’s backup battery unit fail.
- A volume’s capacity can be completely used without affecting performance.
- The workload labels given to a volume during creation have no impact on the volume’s performance or usage, it is just a tag.
- Veeam cannot benefit from Full Stripe Write Acceleration (FSWA) because the block size in Veeam operations is not consistent.
- NIC Teaming in Windows can reduce performance.
- Network load balancing with Link Aggregation Control Protocol (LACP), multiple ports, etc. does not increase performance and can hinder performance in some cases.
- For Proxy or Repository Servers choose the highest bandwidth available.

Performance best practices

Based on the performance tests carried out, a list of best practices was created. These help setup and tune the system to achieve excellent backup/restore performance.

Set task limits

When you start a data protection or disaster recovery job, Veeam Backup & Replication analyzes the list of VMs added to the job and creates a separate task for every disk of every VM to be processed. Veeam Backup & Replication then defines what backup infrastructure components must be used for the job, checks what backup infrastructure components are currently available, and assigns necessary components to process the created job tasks.

If you use the parallel data processing mode and/or schedule several jobs to run in parallel, backup infrastructure components typically process several tasks at the same time. You can limit the number of tasks that backup infrastructure components must process concurrently. Task limitations helps you balance the workload across the backup infrastructure and avoid performance bottlenecks.

Task limits set for backup infrastructure components influence the job performance. For example, you add a VM with four disks to a job and assign a backup proxy that can process a maximum of two tasks concurrently for the job. In this case, Veeam Backup & Replication creates four proxy tasks (one task per each VM disk) and starts processing two tasks in parallel. The other two tasks are pending.

Backup infrastructure components

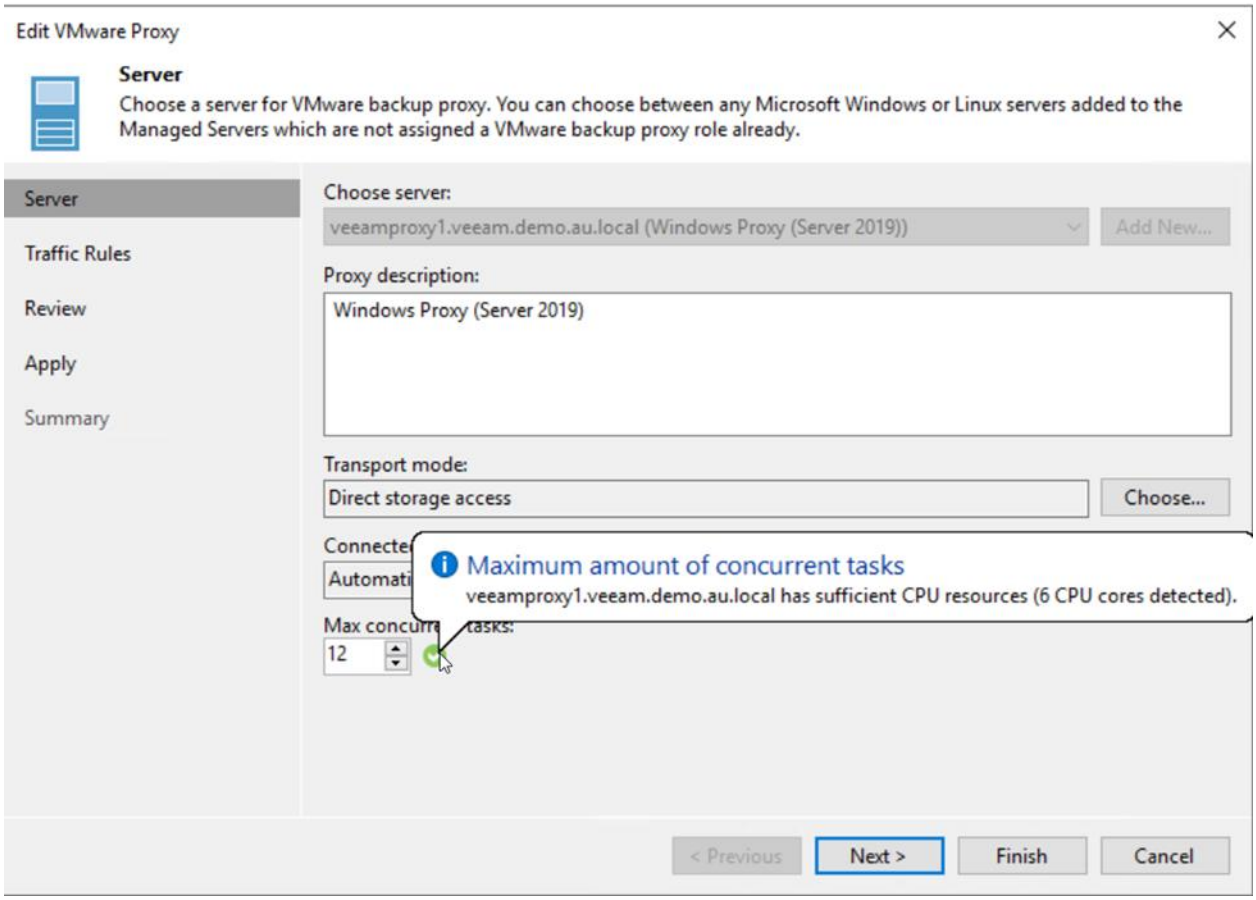
Veeam Backup & Replication lets you limit the number of concurrent tasks for the following backup infrastructure components: backup proxies and backup repositories.

Backup proxies

To limit the number of concurrent tasks on a backup proxy, you must define the maximum concurrent tasks setting for the backup proxy; see Figure 29.

The maximum number of concurrent tasks depends on the number of CPU cores available on the backup proxy. It is strongly recommended that you define task limitation settings using the following rule: one physical CPU core = two proxy tasks. For example, if a backup proxy has four CPU cores, it is recommended that you limit the number of concurrent tasks for this backup proxy to a maximum of eight. This rule only applies fi the server is only used as a proxy, otherwise the resources must be shared between the roles, such as proxy and repository.

Figure 29) Restricting concurrent tasks per backup proxy.



Backup repositories

To limit the number of concurrent tasks on a backup repository, you must enable the Limit maximum concurrent tasks to <N> option on the backup repository and define the necessary task limit; see Figure 30.

The maximum number of concurrent tasks depends on the number of CPU cores available on the backup repository. It is strongly recommended that you define task limitation settings using the following rule: one task = one CPU core if the server is only sued as a repository server, otherwise CPU cores must be shared between roles.

It is recommended to configure 4GB RAM per core. In the case of shared folder backup repositories, the same number of resources is required for gateway servers.

Synthetic operations performed on the backup repository (such as synthetic full backup, backup files merge and transform) are also regarded as tasks. The number of tasks performed during these operations depends on the type of backup chains stored on the backup repository:

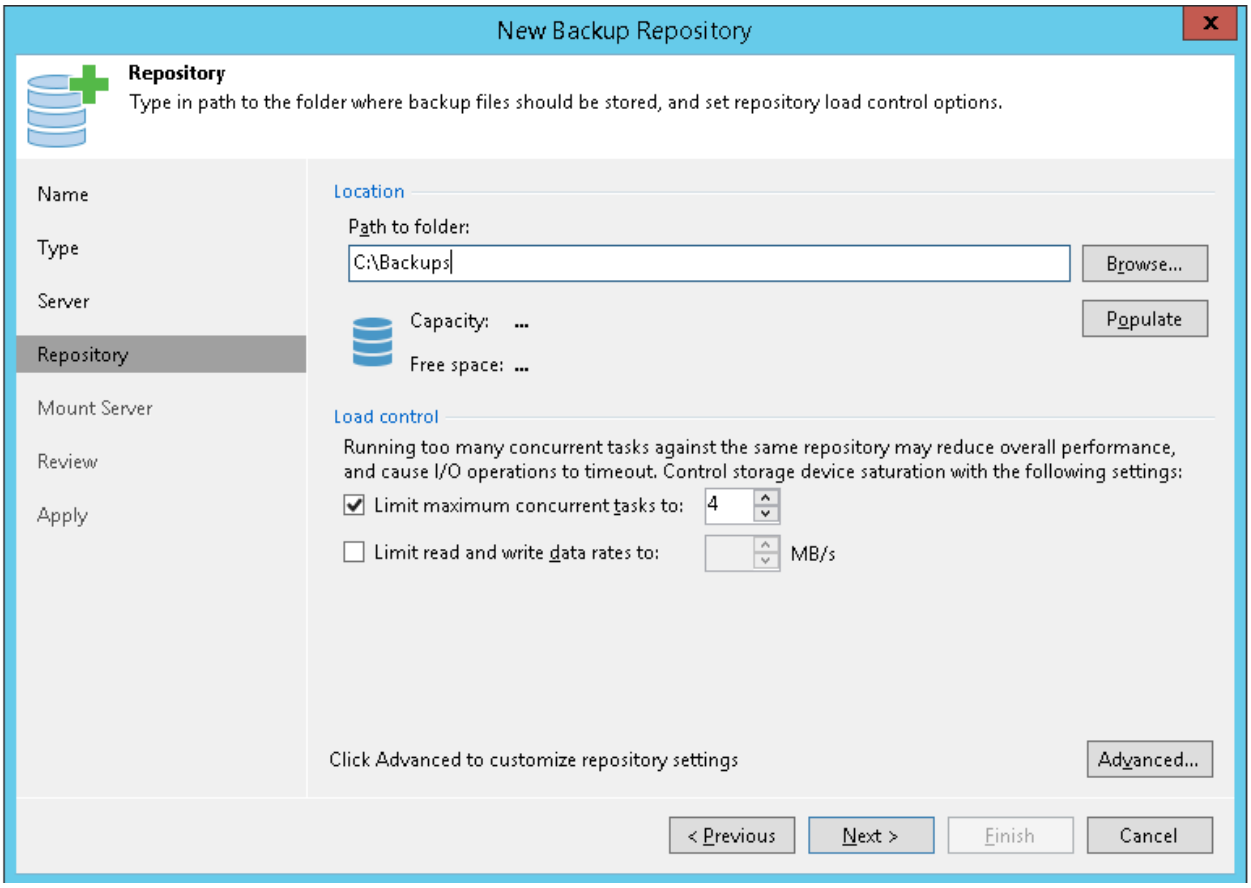
- For regular backup chains, Veeam Backup & Replication creates one task per job.
- For per-VM backup chains, Veeam Backup & Replication creates one task per every VM chain (that is, every VM added to the job).

If you use backup repositories for backup copy jobs, you must also consider tasks for read operations.

Note: When you limit the number of tasks for the backup repository, bear in mind the storage throughput. If the storage system is not able to keep up with the number of tasks that you have

assigned, it will be the limiting factor. It is recommended that you test components and resources of the backup infrastructure to define the workload that they can handle.

Figure 30) Restricting concurrent tasks per backup repository.



Adjust the number of active snapshots per datastore to be in line with the number of concurrent tasks per backup repository. By default, it is restricted to four as a “protection method” to avoid filling up a datastore. If left at default value, your number of tasks per backup repository will be restricted to four.

Note: The default 4 active snapshot per datastore value can be modified by creating a registry DWORD value in ‘HKEY_LOCAL_MACHINE\SOFTWARE\Veeam\Veeam Backup and Replication\’ called MaxSnapshotsPerDatastore and use the appropriate hex or decimal value. Any number can be selected but ensure you have enough capacity on your backup repository.

Use the recommended guidelines with regards to Direct SAN access mode, this is the recommended transfer mode.

NetApp SANtricity storage plug-in for vCenter

The vSphere Client is a single management interface that you can use to manage the VMware infrastructure and all your day-to-day storage needs. The following functions are available in the NetApp SANtricity Storage Plugin for vCenter:

- View and manage discovered storage arrays in the network.
- Perform batch operations on groups of multiple storage arrays.
- Perform upgrades on the software operating system.

- Import settings from one storage array to another.
- Configure volumes, SSD cache, hosts, host clusters, pools, and volume groups.
- Launch the System Manager interface for additional management tasks on an array.

Note: The plugin is not a direct replacement for the System Manager software. System Manager is still required for performing certain storage administration tasks on a single array.

The plugin requires a VMware vCenter Server Appliance deployed in the VMware environment and an application host to install and run the plugin web server.

You can download the plugin from the NetApp Support site, NetApp Support Site - Downloads - All Downloads.

You can find installation and configuration documentation in the SANtricity Storage Plugin for vCenter.

Summary

Veeam Backup & Replication offers powerful cutting-edge capabilities in the data protection industry, but without a proper repository, backup windows and recoveries can be negatively affected. Veeam provides more recovery options and faster restoration capabilities, but to realize these benefits, the storage system must have the necessary performance profile. With technologies such as instant VM recovery, you can run an application directly from your backup file, but how is that application going to perform? NetApp E-Series arrays offer the performance that you need when recovering one or more applications and give you confidence that the data that you backed up is protected and is available when you need it.

Where to find additional information

To learn more about the information that is described in this document, review the following documents and/or websites:

- E-Series and SANtricity 11 Documentation Center
  https://docs.netapp.com/us-en/e-series-family/index.html
- Veeam Backup & Replication technical documentation
  http://www.veeam.com/documentation-guides-datasheets.html
- Instant Recovery to VMware vSphere
  https://helpcenter.veeam.com/docs/backup/vsphere/instant_recovery.html?ver=120
- NetApp Knowledge Base (https://kb.netapp.com/) articles regarding Veeam and E-Series:
  - VMware snapshots stun and unstun times are to long
  - Why does a virtual machine freeze during a VMware snapshot

Version history

Version	Date	Document version history
Version 2.0	September 2024	Updated with Advanced UniByte testing results and added Linux Hardened Repository section.
Version 1.0	December 2022	Initial release.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

TR-4948-0924

NetApp