NetApp

Technical Report

NetApp Scale-out Data Protection
Solution Deployment

Jeannine Walter, NetApp
Sarbjit Singh, Commvault
October 2020 | TR-4864-DEPLOY

In partnership with

COMMVAULT

Abstract

This paper introduces NetApp® Scale-out Data Protection (SDP), a combined NetApp and Commvault® solution that offers expanded rapid recovery support for critical data across a data fabric powered by NetApp technology. This document provides configuration, implementation, and sizing guidelines for the NetApp SDP solution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Solution Overview

With many organizations struggling to meet the availability and rapid recovery goals that their business demands, NetApp and Commvault have partnered to create an appliancelike verified architecture specifically for this purpose. Running on NetApp® HCI and powered by Commvault® software, the primary recovery medium is fast all-flash storage, with a cloud tier based on NetApp StorageGRID® software for long-term retention. This clever design meets the most demanding recovery needs and keeps costs under control, while exceeding modern ease-of-use requirements for scalability and day-to-day management.

About NetApp

In a world full of generalists, NetApp is a specialist. We're focused on one thing, helping your business get the most out of your data. NetApp brings the enterprise-grade data services you rely on into the cloud, and the simple flexibility of cloud into the data center. Our industry-leading solutions work across diverse customer environments and the world's biggest public clouds.

As a cloud-led, data-centric software company, only NetApp can help build your unique data fabric, simplify and connect your cloud, and securely deliver the right data, services, and applications to the right people—anytime, anywhere.

About Commvault

Commvault believes in data readiness. Commvault helps organizations of all sizes intelligently manage data through solutions that store, protect, manage, and use their most critical asset—their data. Commvault software, solutions, and services are available from the company and through a global ecosystem of trusted partners. Commvault employs more than 2,300 highly skilled individuals across markets worldwide, is publicly traded on NASDAQ (CVLT), and is headquartered in Tinton Falls, New Jersey, in the United States.

Solution Technology

Scale-out Data Protection (SDP) is a verified architecture from NetApp. The components include NetApp HCI, NetApp AFF, and NetApp StorageGRID, combined with Commvault Complete™ Backup & Recovery (Commvault Complete).

The Commvault Complete CommServe® and MediaAgents are installed and running on NetApp HCI. AFF becomes the primary landing zone for the backup data, also referred to as the performance ContentStore. And NetApp StorageGRID holds long-term retention data.

Figure 1 shows the technical components of the solution.

Figure 1) Solution overview.
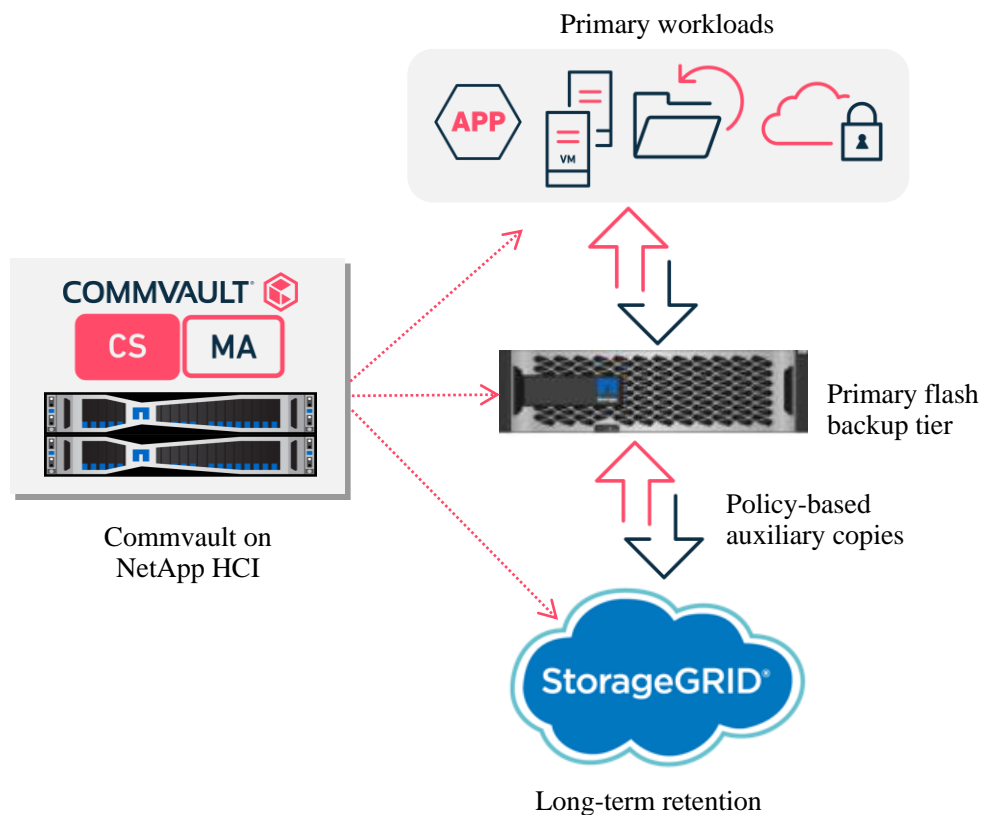
Use Case Summary

This solution applies to the following use cases:

- Snapshot and replication orchestration and management
- Rapid recovery of critical applications
- Protection of recovery workloads and virtual machines (VMs)
- Ransomware protection
- Long-term backup retention for governance
- Disaster recovery (DR)
- Multiple cloud use cases, including data and VM migration and cloud DR

Scale-out Data Protection

SDP is a highly scalable solution. Each piece of technology can be sized and scaled independently, which can make sizing and deploying the solution challenging. This document details each step of the process and provides overall guidelines and recommendations.

Before beginning deployment, each piece of the solution should be sized correctly. Some assumptions were made for sizing of workloads. Those assumptions are described in section 9, Solution Sizing.

Here are the steps for sizing:

1. Calculate workload sizing with the Commvault sizing tool.

2. Follow the NetApp HCI medium, large, extra-large recommendations listed in section 9, Solution Sizing.

3. Calculate the NetApp AFF A220 tier by using the NetApp Fusion sizing tool.

4. Calculate the StorageGRID 6060 tier by using the Solution Designer spreadsheet.
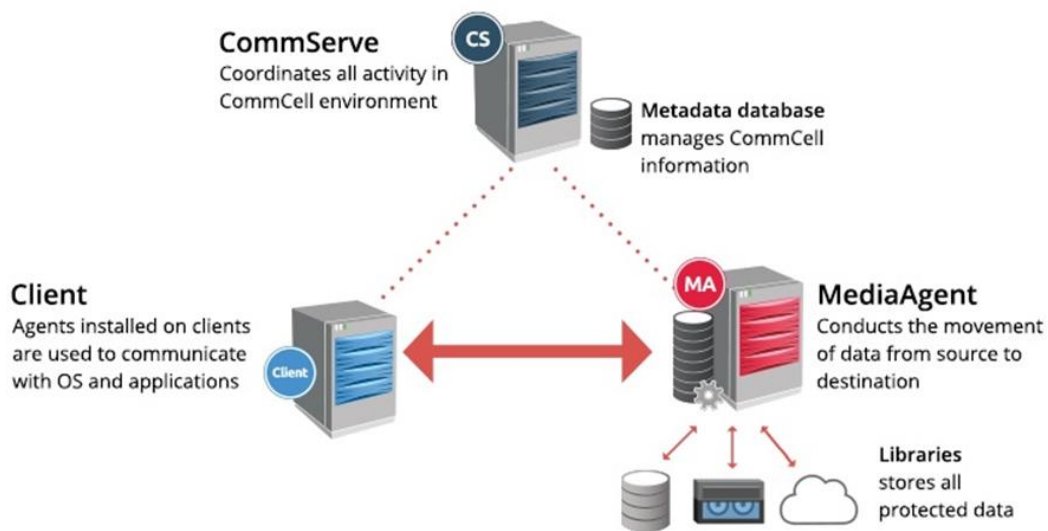
After the sizing has been determined, the solution can be installed and deployed. The steps and recommendations for each component are detailed in the following sections.

Commvault Deployment and Configuration

Commvault Architecture

Figure 2 shows the major components of a Commvault Complete Backup & Recovery system.

Figure 2) Commvault architecture.

CommServe Server

The CommServe host is a physical or virtual system in an environment that acts as the central management engine for the CommCell architecture. A Commvault Complete deployment allows the creation of only a single CommServe server in a protected environment. A CommServe server hosts an internal MySQL database that contains all necessary information for the protection of the converged infrastructure, including all configuration, security, and operational history for the environment.

MediaAgent

The MediaAgent manages the data storage libraries and oversees all data transmission operations in the protected environment. Like the CommServe server, it can be deployed in the converged infrastructure as a physical or virtual host. However, unlike the CommServe server, you can deploy multiple MediaAgent servers to further scale operations and provide high-performance data movement.

CommCell Console

The CommCell Console is a simple UI for the centralized management of all operations in the protected environment. It enables administrators to monitor and control all active jobs, and it enables users to review all activities and event logs.

Command Center

WebServer and WebConsole are installed by default along with the CommServe server. They provide front-end and back-end functionality for management of the protected environment. These components allow end users to remotely manage their data through an HTML5-based web interface known as the Command Center. Through this interface, users can create reports, modify backup jobs and schedules, and manage the virtual machines deployed in the converged infrastructure.

Clients and Agents

Clients are the targets of backup operations in the CommCell environment, whether they are physical or virtual entities deployed in the converged infrastructure. Installing agents on each client provides advanced backup functionality. Some agents provide backup and recovery for simple file systems, whereas other agents enable application-aware backups to create consistent restore points. Another class of agents enables database backup and recovery by quiescing database operations to create a stable state before a new backup copy is taken.

ContentStore

A hardware-agnostic virtual repository. The back-end repository for all Commvault-managed information. ContentStore consolidates protection and archiving of data to eliminate inefficient data silos that waste resources and infrastructure. An intelligent index provides global awareness for your data so you can quickly find what you need, when you need it. Commvault software has a single, deduplicated index across the ContentStore, making it the only enterprise-class solution like it in the market.

Disk Library

A storage resource with an associated mount path that is used in the Commvault backup environment to store index information backups.

Core Setup

Before installing the Commvault software for the NetApp Scale-out Data Protection solution, you should review the following system requirements. You should also make sure that the virtual machines needed for running Commvault CommServe and MediaAgents are available per the required CPU, memory, disk, and IOPS specifications.

Before beginning, review the following guidelines:

- System Requirements for the CommServe server
- System Requirements for the MediaAgent

Other packages that are installed along with the Commvault software might require a license. To review the list of packages, see Additional Packages Installed with the CommServe Server.

For full information about Commvault licensing, see License Administration.

Install the Commvault Package

Use the Commvault package to set up a CommCell environment for protecting databases, file systems, and cloud applications.

Before You Begin

Download the Commvault Installer for the current mainstream version from [Commvault Maintenance Advantage](#) (example v11 SP20).

Verify that the computer where you plan to install the package satisfies the minimum requirements specified in [System Requirements](#).

Procedure

To install the Commvault package, follow these steps.

1. Start the Commvault installer, select where to extract the installation files, and then click Extract.
2. On the Language Selection page, select a language and then click Next.
3. On the Welcome page, select the I Agree check box and then click Next.
4. On the Choose the Installation Type page, click Install Packages on This Computer and then click Next.
5. On the Install Option page, click Add Packages and then click Next.
6. On the Select Packages page, select the following packages and then click Next:
   - CommServe
   - Active Directory
   - Command Center
   - File System Agent
   - MediaAgent
   - Virtual Server Agent
   - CommCell Console
   - Web Server
   - Workflow Engine

   Note: The installer autoselects other required packages.
7. On the Select Hotfix Packages page, select the hotfix package to install and then click Next.
8. The installation begins.
9. On the Client Computer Information page, the computer name and host name are entered automatically. Click Next.
10. Complete the steps in the installation wizard.
11. On the Completion page, make a note of the Command Center URL and then click Finish.
12. When the setup is complete, provide the Command Center URL to backup administrators, database administrators, and other users who might need to access the console.

Complete the Core Setup Wizard

When you log on to the Command Center for the first time, you can use the wizard to finish the core setup. The wizard helps you to set up a disk storage pool and to modify the server backup plan according to your requirements.

Before You Begin

Note the location of the NetApp CIFS share to be used as a disk library.

Procedure

To complete the Core Setup wizard, follow these steps:

1. Go to the Command Center URL that your administrator provided. The URL has the following format: http://*webhost*/adminconsole.

2. Enter your user name and password.

3. Click Login.

4. From the navigation pane, go to Guided Setup. The initial application setup page opens.

5. Click Let's Get Started.

6. The Core Setup wizard opens.

7. On the Add Storage Pool tab of the wizard, add storage pool settings and then click Save.

    For example, add a storage pool path (a disk library) and a deduplication database (DDB) partition path.

8. On the Create Server Backup Plan tab, modify the plan according to your requirements and then click Save.

Configuring Solutions in the Commvault Command Center

After you complete the core setup, the Guided Setup page opens on the Command Center navigation pane.

On the Guided Setup page, you can configure backup and recovery for file servers/NAS, virtual machines, databases, and cloud apps.

Before You Begin Complete the Core Setup.

Procedure

To configure solutions in the Commvault Command Center, follow these steps.

1. From the navigation pane, go to Guided Setup. The Welcome page opens.

    Use the quick links in the right section of the page to perform the following tasks:

    Configuring an Email Server

    Customizing the Theme

    Creating a Server Backup Plan

    Creating a Company

    Downloading Software

2. Select what you want to set up:

    On the Protect tab, click the tile for the application that you want to back up and recover

    Click the More button to expand the list and open the links below to configure data protection for your specific workload:

    File Servers

    Virtualization

    Laptop

    Databases

    Office 365

    Exchange

[Archiving](#)

[Cloud apps](#)

[Oracle EBS](#)

[SharePoint](#)

[Active Directory](#)

[Big data apps](#)

[G Suite](#)

[Salesforce](#)

3. You are redirected to the page for the solution you configured. For example, when you configure the Virtualization workload, you are redirected to the Protect > Virtualization page for further configuration and management.

Additional Configuration

To add the NetApp AFF system as a disk target that uses CIFS/NFS/ISCSI, go to [Configuring a Disk Storage.](#) Also see section 5, NetApp AFF Deployment and Configuration.

To add a NetApp StorageGRID bucket as a cloud storage target for Commvault, go to [Configuring Cloud Storage](#). Also see section 7, NetApp StorageGRID Deployment and Configuration.

To start protecting your data, create one or more [server backup plans](#) according to your RPO and RTO needs. Review [System Requirements for the NetApp Storage Array](#) for information about protecting primary workloads on NetApp technology.

Tuning Parameters

To enhance the performance of data movement operations, the number of data readers was increased on the CommServe server. Data readers determine the number of parallel read operations while the data is being backed up. Configuring multiple data readers per subclient on disk arrays can improve the backup performance of clients. The number of data readers was increased to 25 during the validation testing.

Sizing Recommendations for CommServe and MediaAgents

For the NetApp Scale-out Data Protection solution, Commvault components were deployed on NetApp HCI. The CommServe server runs on a Windows VM, while the MediaAgents can be on Windows or Linux. This configuration included one CommServe server and four MediaAgent VMs. In this solution the sizing recommendations follow the requirements for a large or extra-large CommServe server and large or extra-large MediaAgents (see section 9, Solution Sizing). The MediaAgent VMs connect to the NetApp AFF storage over CIFS and to StorageGRID over S3 protocol.

Before installing the Commvault software, review the following requirements:

- [System Requirements for the CommServe Server](#)
- [System Requirements for the MediaAgent](#)

Commvault Workload Sizing

The next step in the process is to determine the size of the workload that will be stored on the AFF tier. To correctly size the AFF tier, NetApp data protection specialists can use the Commvault [Total Backup Storage Calculator](#) tool to estimate the backup infrastructure requirements. The tool requires Commvault Partner Portal access. [Sign up for access,](#) if necessary.

The following tasks can be used to perform discovery for sizing the data protection solution:

- Identify the system or application and database workloads and corresponding front-end capacity (in terabytes [TB]) to be protected.
- Identify the VM or file workload and similar front-end capacity (in TB) to be protected.
- Identify short-term and long-term retention requirements.
- Identify the daily % change rate for the datasets or workloads identified.
- Identify projected data growth over the next 12, 24, and 36 months.
- Define the RTO and RPO for data protection or recovery according to business needs.

When this information is available, the backup infrastructure sizing can be done, resulting in a breakdown of required storage capacities for the AFF tier and the StorageGRID tier. The flash and object storage tiers can be sized independently. The flash tier is typically used for short-term retention and for workloads that require faster recovery, typically tier 0 or tier 1 workloads. Tier 2 and tier 3 workloads can be directly backed up to the StorageGRID tier for short-term and long-term retention.

For sizing assistance specific to your environment, contact your NetApp or Commvault representative.

Commvault Data Tiering

Commvault software uses the auxiliary copy operation to tier the data between the AFF and StorageGRID tiers in the NetApp SDP solution. The copy operation is deduplication aware and is performed at the Commvault MediaAgent level. The same process is also used for replicating backup data across different data centers efficiently across MediaAgents, regardless of their operating system type. For more information, refer to Performance Tuning for the Auxiliary Copy Operation.

Commvault Licensing

Other packages that are installed along with the Commvault software might require a license. To review the list of packages, see Additional Packages Installed with the CommServe Server.

Client licensing is based on front-end terabytes (FET), or it can be purchased as instances.

For full information about licensing, see License Administration.

Additional Information

For CommServe server sizing, see Hardware Specifications for the CommServe Server.

For MediaAgent sizing, see the Deduplication Building Block Guide.

NetApp HCI Deployment and Configuration

Empower your organization to move faster while reducing costs with NetApp HCI. Scale compute and storage resources independently so you never pay for more than you use. And deploy in minutes with a turnkey cloud infrastructure that eliminates the complex management of traditional three-tier architectures. Integration into the data fabric delivered by NetApp means that you can unleash the full potential of your applications, with the data services they require, across any cloud.

For this solution, Commvault is deployed on a NetApp HCI system. When deployment is complete, the NetApp HCI system provides compute and storage for the Commvault CommServe server and MediaAgents.

Note: This solution is intended to be a closed system. Nothing other than the Commvault CommServe server and MediaAgents is intended to run on the NetApp HCI system. No other workloads should be run on this system.

NetApp HCI provides both storage and compute resources, combining them to build a VMware vSphere environment backed by the capabilities of NetApp Element® software.

After successful deployment of NetApp HCI, you manage the system from the VMware vSphere Web Client. The NetApp Element Plug-in for vCenter Server (also referred to as the vCenter Plug-in, or VCP) is used to manage storage resources, including datastores, volumes, quality of service, storage cluster components, and data protection. Compute nodes appear as ESXi hosts, which can be managed in vSphere.

Deployment Using the NetApp Deployment Engine

NetApp HCI simplifies day 0 deployment with the NetApp Deployment Engine (NDE), completing more than 400 deployment and configuration tasks in a fraction of the time it would take to deploy manually.

The NetApp HCI 1.8 Deployment Guide describes the steps for configuring and deploying the NetApp HCI compute and storage nodes.

The NDE UI is the software wizard that is used to install NetApp HCI.

NetApp HCI uses a storage node management network IPv4 address for initial access to the NDE. As a best practice, connect from the first storage node.

To deploy NetApp HCI, you need to access the NetApp Deployment Engine on one of the NetApp H-Series storage nodes via the IPv4 address assigned to the Bond1G interface. The Bond1G interface is the logical interface that combines ports A and B for storage nodes. This storage node becomes the controlling storage node for the deployment process. Depending on your environment, you need to either configure the IPv4 address or retrieve it from one of the storage nodes.

The NDE completely automates the installation and configuration of vSphere components and prepares the system according to established best practices.

The next steps include configuring internal and external networks and port groups, configuring vSphere, and creating VMDKs for the CommServe server and MediaAgents.

Complete the prerequisite steps and deployment steps described in the NetApp HCI 1.8 Deployment Guide.

Sizing and Performance Recommendations

For the NetApp Scale-out Data Protection solution, Commvault components (CommServe server and MediaAgent) were deployed as Windows VMs on the NetApp HCI system. This deployment included one CommServe server and four MediaAgent VMs. In this solution, the sizing recommendations follow the requirements for a large or extra-large CommServe server and large or extra-large MediaAgents.

Before installing the Commvault software, review the following requirements:

- System Requirements for the CommServe Server
- System Requirements for the MediaAgent

When the details and requirements for the CommServe server and MediaAgents have been determined, refer to the NetApp Fusion sizing tool for the correct system to meet the performance level required.

Section 9, Solution Sizing, contains "t-shirt" sizing recommendations for NetApp HCI systems to simplify sizing this solution.

For full information about the Fusion sizing tool, see Welcome to Fusion.

Note: If you choose to perform your own NetApp HCI sizing and not to use the recommendations in Section 9, note the following points about Fusion sizing calculations:

- No oversubscription on vCPUs can be calculated.

- No storage efficiencies can be used. The best deduplication ratio you can expect is 1.2:1. The NetApp HCI storage is running only Commvault and the Commvault deduplication database. 85% of the data will be unique.

Note: Be sure to include the RAM and vCPU requirements for the NetApp HCI management node and the vCenter server in your sizing calculations for NetApp HCI; see Table 1 . Additionally, if you choose not to include the SG1000 StorageGRID load-balancing appliance, be sure to include the management and gateway node requirements in the NetApp HCI calculations as well; see table 2.

Table 1) Management node and vCenter requirements.

Management Node Requirements	vCenter Requirements
6 vCPUs	4 vCPUs
12GB RAM	16GB RAM
400GB disk space	300GB disk space

Table 2) Gateway and admin node requirements for StorageGRID.

Gateway Node Requirements	Admin Node Requirements
8 vCPUs	8 vCPUs
24GB RAM	24GB RAM
200GB disk space	400GB disk space

Additional Information

For NetApp HCI specifications, see the NetApp HCI datasheet.

For additional information about NetApp HCI systems, see the NetApp HCI Documentation Center.

NetApp AFF Deployment and Configuration

NetApp AFF is a robust scale-out platform built for virtualized environments, combining low-latency performance with best-in-class data management, built-in efficiencies, integrated data protection, multiprotocol support, and nondisruptive operations.

Validation testing was completed using AFF, which provides the speed and flexible deployment options required for this solution.

AFF-based storage can be presented to Commvault via CIFS/NFS or ISCSI protocol. NAS connectivity offers the additional advantage of sharing the Commvault disk library across MediaAgent VMs for high availability and no single point of failure. For full information about configuring disk libraries, see Disk Libraries—Best Practices.

The following steps are required for providing a backup target that uses either NFS, CIFS, or ISCSI:

1. Complete Systems Installation and Setup Instructions on the NetApp AFF array.
2. Complete cluster administration to provision clusters, networking, and physical and logical storage.
3. Complete either NFS, CIFS, or iSCSI configuration to be used to connect to MediaAgents:
   - Provisioning for NAS protocols
   - Provisioning for SAN protocols

Sizing and Performance Recommendations

To properly size the all-flash tier, the Commvault sizing tool must be used first. After the type of data, the retention, the change rate, and the growth rate are entered into the Commvault sizing tool, the tool produces a sizing estimate for the primary performance ContentStore on the AFF. See section 3.9, Commvault Workload Sizing.

To determine the proper AFF series system and the amount of capacity required after the Commvault sizing, NetApp has provided the Fusion sizing tool specifically for AFF. Fusion provides system sizing estimates and creates a simple report to facilitate ordering.

Best Practices

For best performance, the number of data readers may need to be increased from the default of 5; 25 data readers were used during validation testing.

The layout and choice of server virtual machines recommends setting SVMs up per application.

When creating aggregates using SSDs, there is no performance gain over 5 SSDs. Therefore, you can create aggregates as needed as long as the root aggregate is separate.

SnapLock Considerations

NetApp SnapLock® compliance software is supported with Commvault. When configuring a disk library for use with SnapLock, Commvault refers to this as read-only files on storage systems. The read-only lock mechanism on the destination systems saves the archive and backup files as read-only files. The expiration date for the read-only lock is set to match the data retention time (backup or archiver retention days, whichever is higher) established in the storage policy copies. These files cannot be modified or deleted by any user or application until the specified retention date. Once the retention expires, the system deletes the files as a part of data aging.

For security best practices and hardening the AFF system, review the hardening guide in [NetApp Product Security](#).

To use SnapLock, there must be a separate aggregate dedicated to SnapLock. Every volume needs a protection policy. The best practice is to enable write once, read many (WORM) on secondary and tertiary copies that are not accessed frequently. This practice helps to minimize over usage, the storage of extra data that can't be deleted until the SnapLock retention date is met.

Deduplication

Deduplication is integrated into Commvault Complete, which reduces backup times while saving storage and network resources by identifying and eliminating duplicate blocks of data during backups.

Commvault recommends not using other hardware or software deduplication applications with Commvault deduplication. Using a second software or hardware deduplication application on top of Commvault deduplication can incur unnecessary performance and resource utilization on systems and offers no further benefit to the already deduplicated data.

Additional Information

For more information about NetApp AFF, see [AFF A-Series All Flash Arrays.](#)

For the latest AFF installation guides, see [the AFF Installation and Setup Guides](#).

For the latest Fusion sizing tool, see [Welcome to Fusion](#).

NetApp EF-Series Deployment and Configuration

NetApp EF-Series storage systems provide extreme throughput performance with new, faster host interfaces and can scale to up to 1.8PB of raw capacity to support fast, large-capacity applications. With its versatility—including multiple host interface choices, multiple RAID choices, and drive choices from entry-level capacity to enterprise capacity—the EF-Series systems are modern, ready-to-work, all-flash storage systems. The EF570 system delivers industry-leading performance, an excellent interface, and configuration flexibility.

All testing and validation for this solution were completed using the NetApp AFF series systems. When performance must be weighed against cost, and when ONTAP® tools (SnapLock, SnapMirror®, SnapVault®, and so on) are not required, the all-flash EF-Series can be substituted for the AFF tier.

There are some trade-offs when substituting EF-Series systems for AFF. EF systems do not support SnapLock, so it is not possible to create an immutable backup tier. You can still use the capabilities of Commvault's artificial intelligence and machine learning to detect anomalies, but you will not be able to create a completely immutable backup target or NetApp Snapshot™ copy.

EF-Series systems support only ISCSI and SAN protocols. When configuring an EF system as a disk library, that disk library must be attached to a single MediaAgent, unlike an NFS mounted disk library that can be shared among several MediaAgents. In case of a MediaAgent failure, a shared disk library can be accessed via a different MediaAgent. In case of a failure, an attached disk library would have to be manually attached to another MediaAgent.

Sizing and Performance Recommendations

To properly size the EF-Series tier, the Commvault sizing tool must be used first. After the type of data, the retention, the change rate, and the growth rate are entered into the Commvault sizing tool, the tool produces a sizing estimate for the primary performance ContentStore on the EF-Series system. See section 3.9, Commvault Workload Sizing.

To determine the correct EF-Series system and the amount of capacity required after the Commvault sizing, NetApp provides Fusion, a sizing tool specifically for EF-Series systems. Fusion provides system sizing estimates and creates a simple report to facilitate ordering.

Once the sizing has been determined, the next step is to configure the EF-Series system to work with the Commvault MediaAgents. The EF-Series system is part of the high-performance ContentStore, configured in Commvault as a disk library connected to a MediaAgent.

To complete the disk library creation, refer to the [Commvault Documentation Disk Library,](#)

Best Practices

For best performance, the number of data readers may need to be increased from the default of 5; 25 data readers were used during validation testing.

NetApp recommends the use of NetApp Dynamic Disk Pools rather than RAID configurations. NetApp Dynamic Disk Pool technology enables dynamic rebalancing of drive count changes to easily expand storage capacity as backup needs grow, while providing added data protection with faster rebuild times if a drive fails.

For large and extra-large configurations, consider the following best practices:

- Use NetApp Dynamic Disk Pools to maximize ease of use and rebuild times.
- Create LUNs (not thin) from the volume group of 9TB if possible, regardless of the spindle size installed in the system
- For a Dynamic Disk Pool, the default segment size (file system typical) is 128KB.

Additional Information

For more information about NetApp AFF, see [EF-Series Flash Storage Arrays.](#)

For the latest EF-Series installation guide, see [E-Series and SANtricity 11 Documentation Center](#).

For the latest Fusion sizing tool, see [Welcome to Fusion](#).

For full information about Commvault integration with E-Series systems, see TR-4320, [NetApp E-Series and Commvault Data Platform V11](#).

NetApp StorageGRID Deployment and Configuration

NetApp StorageGRID is a software-defined, object-based storage solution that supports industry-standard object APIs, including the Amazon Simple Storage Service (S3) API and the OpenStack Swift API.

StorageGRID provides secure, durable storage for unstructured data at scale. Integrated, metadata-driven lifecycle management policies optimize where your data lives throughout its life. Content is placed in the right location, at the right time, and on the right storage tier to reduce cost.

The StorageGRID system uses four types of grid nodes:

- Admin nodes provide management services such as system configuration, monitoring, and logging. Each grid must have one primary admin node and might have any number of nonprimary admin nodes for redundancy. Admin nodes can also be used to load balance S3 and Swift client traffic.

- Storage nodes manage and store object data and metadata. Each StorageGRID system must have at least three storage nodes. If you have multiple sites, each site in your StorageGRID system must also have three storage nodes.

- Gateway nodes (optional) provide a load-balancing interface that client applications can use to connect to StorageGRID. A load balancer seamlessly directs clients to an optimal storage node at an optimal site, so that the failure of nodes or even an entire site is transparent. You can use a combination of gateway nodes and admin nodes for load balancing, or you can implement a third-party HTTP load balancer.

- Archive nodes (optional) provide an interface through which object data can be archived to tape.

Sizing and Performance Recommendations

In order to perform StorageGRID sizing, you must understand the workload:

- Usable capacity

- Average object size

- Performance requirements

- ILM policy applied

To find the amount of usable capacity, review the customer's requirement for the size of the backup workload tiered to StorageGRID.

Average object size is an input parameter that helps with sizing for performance in a StorageGRID environment. The average object sizes used for a Commvault workload depend on the type of backup. Table 3 lists the object sizes based on the type of backup.

Table 3) Average object size used in Commvault backups to StorageGRID.

Backup Type	Average Object Size
Auxiliary copy to StorageGRID	32MB
Direct backup to StorageGRID (deduplication enabled)	8MB
Direct backup to StorageGRID (deduplication disabled)	32MB

In addition, understanding the customer's performance requirements for full backups and incremental backups can help determine sizing for the number of StorageGRID storage nodes.

The StorageGRID information lifecycle management policy data protection methods determine the capacity overhead required for storing Commvault backups, which has an impact on sizing a grid for capacity.

StorageGRID ILM replication is one of two mechanisms that StorageGRID uses to store object data. When StorageGRID matches objects to an ILM rule that is configured to create replicated copies, the system creates exact copies of object data and stores the copies on storage nodes.

Erasure coding is the second method that StorageGRID uses to store object data. When StorageGRID matches objects to an ILM rule that is configured to create erasure-coded copies, it slices object data into data fragments, computes additional parity fragments, and stores each fragment on a different storage node. When an object is accessed, it is reassembled using the stored fragments. If data or a parity fragment becomes corrupt or lost, the erasure-coding algorithm can recreate that fragment by using a subset of the remaining data and parity fragments.

For example:

- Storing two whole replica copies of backup objects results in 2x storage overhead.
- Storing 2+1 erasure-coded copy of backup objects results in 1.5x storage overhead.

StorageGRID is typically deployed in two or more sites with data protection policies that replicate data to protect against node- and site-level failures. When data is committed to StorageGRID, it is protected by making multiple copies of the data or using erasure coding to protect in a single site or across many sites. StorageGRID can also use capacity in other cloud storage systems such as AWS, S3 compatible, and Azure Blob.

Table 4) StorageGRID ILM policy.

Rule	Qualifiers	Ingest Behavior
Erasure code 2+1	Object larger than 200KB	Balanced
Two copy (default rule)	None	Dual commit

The ILM 2-copy policy is the default rule. The policy applied for testing was 2+1 erasure coding to any object 200KB or larger. The default rule was applied for objects smaller than 200KB. This is in accordance with StorageGRID best practices.

When determining the correct sizing, be sure to include the admin node and optional gateway node requirements in your sizing. If you don't want to run the admin and gateway nodes as VMs on the NetApp HCI system, you can add the SG1000 appliance as a physical admin node and load balancer. In the large and extra-large systems, we configured two SG1000s as part of the solution. In the medium systems, the admin and gateway nodes for StorageGRID were run as VMs on the NetApp HCI system.

The validated architecture was configured with tier 2 and higher data backed up directly to the StorageGRID 6060. Commvault tuned the object size to be optimal for performance for puts, gets, and deletes with StorageGRID. This makes StorageGRID a viable backup target for daily and long-term retention workloads. Using StorageGRID for tier 2 and higher workloads can greatly reduce required capacity for the all-flash tier.

To determine proper sizing for the StorageGRID tier, you must first complete the Commvault sizing to determine the estimated capacity required for the StorageGRID tier, taking into account daily and long-term workloads. For full information, see section 3.9, Commvault Workload Sizing.

Note: Although StorageGRID can back up and restore workloads, it *cannot* perform any Commvault Live functions, including LiveMount, LiveRestore, and LiveBrowse. If your workload requires those features, do not stage the workload directly to StorageGRID first. Those workloads should go directly to the all-flash tier.

Table 5) StorageGRID admin node and gateway node requirements.

Node Type	Quantity	vCPU	RAM	Storage
Gateway node	1	8	24GB	100GB LUN for OS
Admin node	1	8	24GB	• 100GB LUN for operating system • 200GB LUN for admin node tables • 200GB LUN for admin node audit log

StorageGRID Solution Designer Is a tool for sizing StorageGRID. This tool is not a part of Fusion; it is a separate spreadsheet-style sizing tool. After you select the StorageGRID model that best fits your workloads, use the spreadsheet to help determine the correct sizing.

As shown in Figure 3, the first step in performing StorageGRID sizing is to determine the correct appliance model. In the validated architecture, the SG6060 appliance is recommended. In smaller environments, the SG5760 appliance can be used.

To perform StorageGRID sizing, follow these steps.

1. Select SG6060 as the Appliance Model.
2. Select the Usable Capacity Required.
3. Select the Average Object Size:
   a. Average object size for workloads direct to StorageGRID = 8MB.
   b. Average object size for workloads auxiliary copied to StorageGRID = 32MB.
4. Select the Small Object Ingest Rate (Figure 4):
   a. Hover over the field to show ingest rate by appliance model.
   b. The SG6060 = 2180 Obj/s.
5. Select the Large Object Ingest Throughput (Figure 5):
   a. Hover over the field to show ingest throughput by appliance model.
   b. The SG6060 = 770MB/s.
6. Select the Data Protection policies (Figure 5):
   a. Select the Hardware-Level Data Protection.
   b. Select the ILM rule.

   Note: The average object size and the data protection policies have a great impact on the number of nodes required.

Figure 3) StorageGRID Solution Sizer.

Figure 4) Small object ingest rate.

Figure 5) Large object ingest rate.
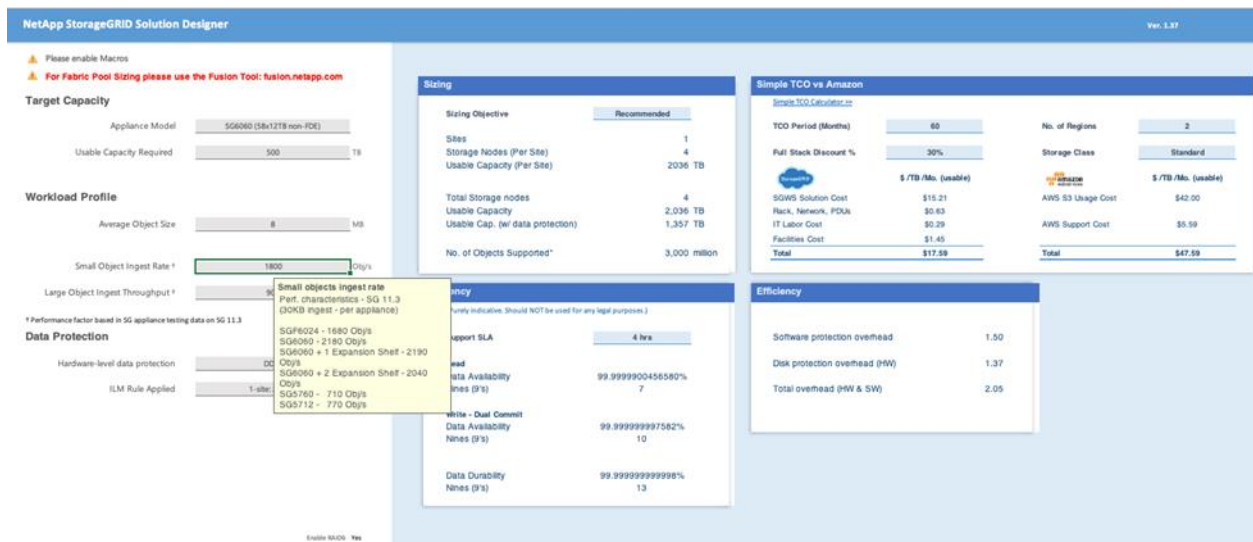
Once the sizing has been determined, the next step is to configure StorageGRID to work with the Commvault MediaAgents. StorageGRID is part of the capacity ContentStore, which is configured in Commvault as a cloud library connected to a MediaAgent.

For the steps to complete creating the disk library, refer to the [Commvault Documentation Cloud Library](#).

For the NetApp StorageGRID bucket configuration, the recommended consistency level is "Available."

Additional Information

For full information about deploying StorageGRID, see the [StorageGRID Documentation Center](#).

Networking

To support the high performance that this system can attain, redundant Mellanox switches should be included in this deployment. The SN2700 offers simple deployment, easy maintenance, and the lowest power consumption in the market. The SN2700 supports speeds of 1, 10, 25, 40, 50, 56, and 100GbE with 32 ports and is hot swappable. The SN2700 is the high-performance switch of choice for this solution, providing enhanced scalability and future-proof unprecedented performance.

Additional Information

For full deployment details for the Mellanox switches, see [TR-4836 NetApp HCI with Mellanox SN2100 and SN2700 Network Switch, Best Practice Cabling Guide.](#)

For detailed information about the SN2700, see [SN2700 Open Ethernet Switch Product Brief.](#)

Solution Sizing

This solution is complex, and it may be difficult to size without guidelines. To simplify the sizing process, here are "t-shirt" sizing recommendations for medium, large, and extra-large systems. To understand how the following configurations were determined, review the previous sections in this document.

Here are the assumptions made in the sizing. These sizing numbers are for a single site. Replication and offsite copies require additional hardware.

The sizing provided here is a base guideline. The numbers vary depending on retention period and daily rate of change in your environment.

- 30% of workloads, specifically tier 0 and tier 1, go to AFF.

- 70% of tier 2+ goes directly to NetApp StorageGRID. (SG6060 with flash makes this viable.)

- The total dataset (front end in TB) is considered equal to the amount of usable data to be protected.

- To calculate the total backup disk target, the following rates of change are assumed:

  - 0-300TB = 1% daily rate of change

  - 300-800TB = 0.5% daily rate of change

  - 800TB and higher = 0.25% daily rate of change

- Yearly data growth is not factored into this calculation.

- StorageGRID is also used to support long-term retention with monthly and yearly backups. See Table 6 for retention period assumptions.

- Short-term retention is targeted at the all-flash tier, and long-term retention (monthly and yearly full backups) is targeted at StorageGRID.

Table 6) Backup retention assumptions.

Backup	Retention Period	Target Disk
Daily Incrementals	14 days	AFF
Weekly	2 weeks	AFF
Monthly	12 months	StorageGRID
Yearly	5 years	StorageGRID

The following steps were performed; the results are detailed below.

1. **Step 1:** Calculate workload sizing with the Commvault sizing tool.
2. **Step 2:** Use NetApp HCI medium, large, and extra-large recommendations.
3. **Step 3:** Calculate NetApp AFF A220 tier by using Fusion.
4. **Step 4:** Calculate the StorageGRID 6060 tier by using the Solution Designer spreadsheet.

Medium Configuration

The medium configuration is designed to protect approximately 300 front-end terabytes (FETs) or approximately 1,000 VMs.

NetApp HCI

One large CommServe server and two large MediaAgents; this system requires:

- 36 total virtual CPU (vCPU) cores

- 18TB usable capacity on NetApp HCI storage calculated with 1.2:1 efficiency (running the Commvault deduplication database)

- 160GB RAM

NetApp HCI Recommendation

- 3 HCI compute nodes model: H615C-75031

- 5 HCI storage nodes model: H610S-1 with 12 x 960GB drives

EF-Series

High-performance ContentStore 1 EF570: 104TB estimated usable

EF-Series Recommendation

One EF570-Series system with 21x 7.6TB drives

StorageGRID

ContentStore 4 SG5760 (4TB drives): 472TB estimated usable

StorageGRID Recommendation

4 NetApp StorageGRID 5760s with 60x 4TB drives each

Commvault

Commvault Complete licensing: 100x 10 VM packs

Instance licenses in packs of 10 for less than 500 front-end terabytes (FET) are less expensive than FET licenses.

Table 7 lists the medium Commvault license configuration.

Table 7) Medium configuration.

Product	Quantity	Model	Drive Type
Compute node	3	H615C-75031	–
Storage node	5	H610S-1	12 x 960GB
NetApp EF-Series	1	EF570	21 x 7.6TB
StorageGRID	4	SG5760	60 x 4TB
Mellanox switching	2	SN2700	–

Note: The admin and gateway nodes (for load balancing) for StorageGRID for the medium configuration are run on the NetApp HCI system as VMs.

Large Configuration

The large configuration is designed to protect approximately 600 FETs or approximately 2,500 VMs.

NetApp HCI

One large CommServe server and four large MediaAgents; this system requires:

- 70 total virtual CPU (vCPU) cores
- 26TB usable capacity on NetApp HCI storage calculated with 1.2:1 efficiency (running the Commvault deduplication database)
- 348GB RAM

NetApp HCI Recommendation

- Three HCI compute nodes model: H615C-75031

- Six HCI storage nodes model: H610S-1 with 12 x 960GB drives

AFF

Performance ContentStore 1 AFF A220: 137TB estimated usable capacity

AFF Recommendation

One AFF A220 with 48x 3.84 TB drives

StorageGRID

NetApp StorageGRID ContentStore 4 SG6060 (8TB drives): 900TB estimated usable capacity

StorageGRID Recommendation

Four StorageGRID SG6060 with 58x 8TB drives

Commvault

Commvault Complete licensing: 600 FET

Table 8 lists the large Commvault license configuration.

Table 8) Large configuration.

Product	Quantity	Model	Drive Type
Compute node	3	H615C-75031	–
Storage node	6	H610S-1	12x 960GB
NetApp AFF A-Series	1	A220	48x 3.84TB
StorageGRID	4	SG6060	2x 800GB SSD/58x 8TB
Mellanox switching	2	SN2700	–
Load balancer	2	SG1000	–

The admin and gateway nodes for StorageGRID for the large and extra-large configurations are run on the SG1000 appliance. For the large and extra-large configurations, two SG1000s are recommended for load balancing and redundancy.

Extra Large

The extra-large configuration is designed to protect approximately 1200 FETs or approximately 5,000 VMs.

NetApp HCI

One extra-large CommServe server and four extra-large MediaAgents; this system requires:

- 90 total virtual CPU (vCPU) cores

- 26TB usable capacity on NetApp HCI storage calculated with 1.2:1 efficiency (running the Commvault deduplication database)

- 668GB RAM

NetApp HCI Recommendation
- Four NetApp HCI compute nodes model: H615C-75031
- Six NetApp HCI storage nodes model: H610S-1 with 12 x 960GB drives

AFF

Performance ContentStore 1 AFF A220: 276TB estimated usable capacity

AFF Recommendation

One AFF A220 with 48x 7.6TB drives

StorageGRID

NetApp StorageGRID® ContentStore 4 SG6060 (12TB drives): 1.3PB estimated usable capacity

StorageGRID Recommendation

Four StorageGRID SG6060 with 58 x 12TB drives

Commvault

Commvault Complete licensing: 1200 FET

Table 9 lists the large Commvault extra-license configuration

Table 9) Extra-large configuration.

Product	Quantity	Model	Drive Type
Compute node	4	H615C-75031	–
Storage node	6	H610S-1	12x 960GB
NetApp AFF A-Series	1	A220	48x 7.6TB
StorageGRID	4	SG6060	2x 800GB SSD/58x 12TB
Mellanox switching	2	SN2700	–
Load balancer	2	SG1000	–

The admin and gateway nodes for StorageGRID for the large and extra-large configurations are run on the SG1000 appliance. For the large and extra-large configurations, two SG1000s are recommended for load balancing and redundancy.

Conclusion

NetApp Scale-out Data Protection with Commvault provides fast recovery and infinite scalability to meet the ever-changing needs of enterprise customers. With Scale-out Data Protection, you can easily move, manage, protect, and orchestrate data across the data fabric. This solution enables you to transform your data protection strategy while providing a solid foundation to protect your data against threats like ransomware and natural disasters.

Where to Find Additional Information

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

- NetApp HCI Documentation Resources
  https://www.netapp.com/us/documentation/hci.aspx
- NetApp HCI Documentation Center
  https://docs.netapp.com/hci/index.jsp
- AFF and FAS System Documentation Center
  https://docs.netapp.com/platstor/index.jsp
- StorageGRID Documentation Resources
  https://www.netapp.com/us/documentation/storagegrid.aspx
- StorageGRID 11.3 Documentation Center
  https://docs.netapp.com/sgws-113/index.jsp
- NetApp Product Documentation
  https://docs.netapp.com
- Partner-facing TR
  NetApp Scale-out Data Protection with Commvault
- Field Portal TR
  NetApp Scale-out Data Protection with Commvault
- Solution brief
  NetApp Scale-out Data Protection

Version History

Version	Date	Document Version History
Version 1.0	October 2020	Jeannine Walter: Initial release.

Refer to the [Interoperability Matrix Tool (IMT)](https://mysupport.netapp.com) 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.

NetApp®