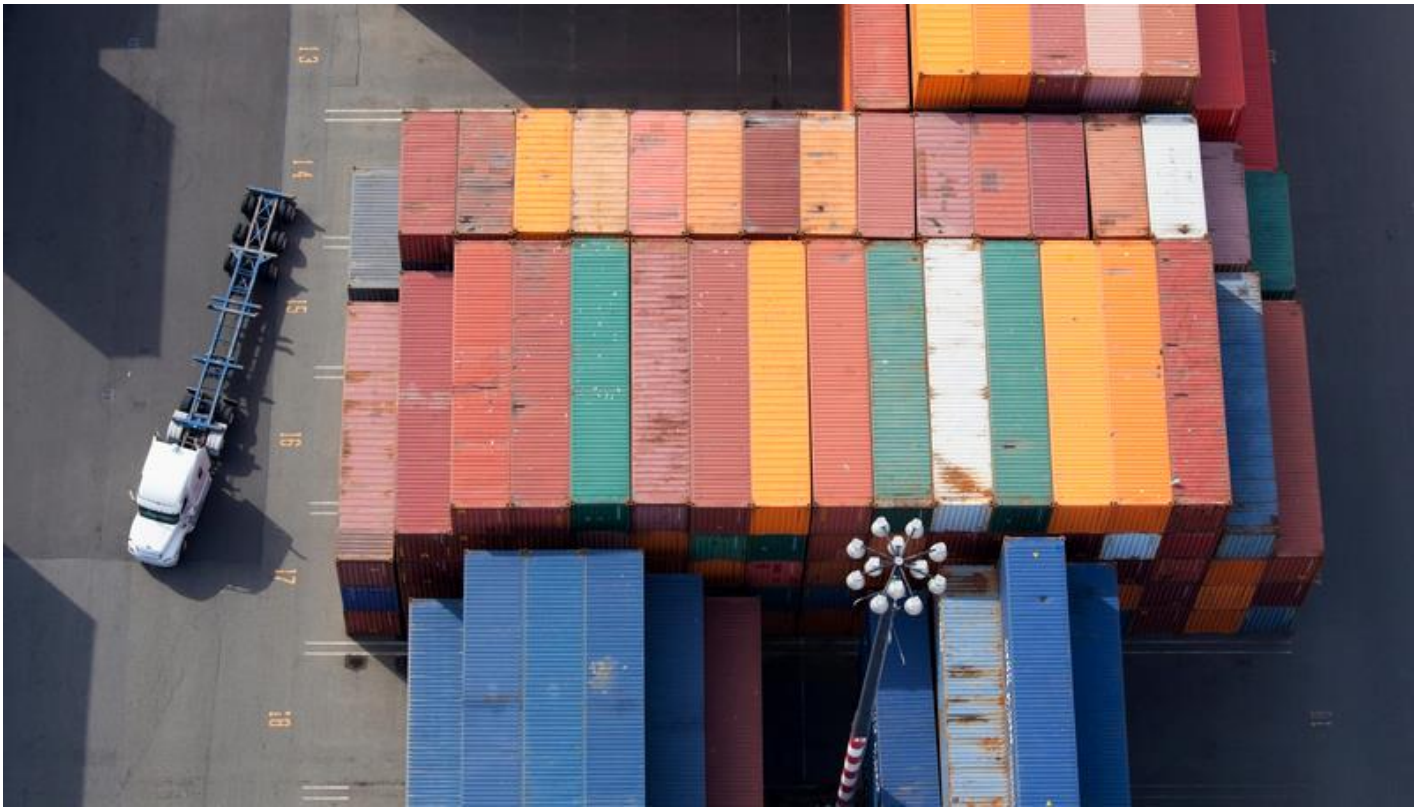
Hewlett Packard Enterprise | COMMVAULT

Technical white paper

Data Protection for Kubernetes using Commvault Backup & Recovery, HPE Apollo Servers, and HPE CSI Driver for Kubernetes



Contents

Executive summary

Container technology adoption is increasing at a fast pace. The global application container market size is expected to reach USD $12 billion by 2028, rising at a market growth of 33.1% CAGR during the forecast period, according to a report by ReportLinker.[1]

The rapid container technology adoption growth is due to the ease that containerized applications can be ported and deployed across different environments. Containers package the applications with everything they need to run with and isolate them within the deployment environment. This enables containerized applications to easily run on different environments such as local desktops, virtual and physical servers, development, testing and production environments, and private or public clouds.

Another factor driving growth of containers is their increasing prevalence for deploying artificial intelligence (AI) and analytics applications. Containers are becoming the standard way to build and deploy machine learning (ML) models, create real-time analytics pipelines, and run batch analytics and extract, transform, and load (ETL) jobs. Their portability across different environments makes containers the perfect vehicle to manage the full lifecycle of AI/ML models and most analytics applications.

With the massive adoption of containers for analytics and AI/ML, data protection requirements are inevitable. These requirements include recovering containerized applications from failures and disasters, replicating environments for migrating a test and development environment to production (or replicating environments from production to staging before an upgrade), and being able to move container clusters more easily.

To protect containerized applications, there are key requirements that must be met. These include seamless operations and policies across on-premises and clouds, operational simplicity for container deployment and data management policies that can span across multiple environments, and backup/restore at the application level—not at virtual machine (VM)/server level.

Hewlett Packard Enterprise offers a data protection solution for a Kubernetes environment deployed on HPE Apollo systems with Commvault® Backup & Recovery, hereinafter referred to as "Commvault software" in this paper. The solution leverages the flexibility and high-density storage optimization for big data on HPE Apollo systems by deploying Commvault software to protect a containerized application—to meet the needs of demanding AI/ML and Deep Learning (DL) workloads—with the right compute, flexible I/O, and storage options.

Target audience: Presales consultants, solution architects, storage operators, data center managers, enterprise architects, and deployment and implementation engineers. A working knowledge of Kubernetes and Commvault software is recommended.

Document purpose: This technical paper describes a solution that highlights how a containerized application in a Kubernetes environment running on HPE Apollo servers can be protected with Commvault software.

This technical white paper describes solution testing performed by Hewlett Packard Enterprise in October 2022.

Introduction

The combination of HPE Apollo systems and Commvault software provides end-to-end enterprise-grade container protection, safeguarding containers wherever they live (on-premises or the cloud or hybrid) and restoring them wherever they are needed. Through the HPE Complete program, Hewlett Packard Enterprise provides one-stop shopping where customers can purchase validated turnkey backup and recovery solutions, reducing risk and improving recovery readiness while protecting their data. Commvault integrates with a multitude of HPE storage and servers, including HPE Alletra, HPE Primera, HPE 3PAR StoreServ, HPE Nimble Storage, HPE StoreOnce, HPE Apollo servers, and HPE ProLiant DL servers, as well as HPE Alletra dHCI, HPE Nimble dHCI, and HPE SimpliVity.

One of the key challenges for protecting Kubernetes environments is managing the dynamic deployment, because the containers are not bundled with physical servers or virtual machines. Containerized applications can run on different environments such as local desktops, virtual and physical servers, development, testing, and production environments, and private or public clouds. Enterprises not only need to quickly deploy containerized AI and analytics applications but protect and restore these applications. Supporting containers with data backup and mobility at scale is a critical need for AI/ML environments.

This technical white paper provides an overview of the requirements for backing up a containerized application in a Kubernetes environment using Commvault software, then restoring that application to a new namespace for test/development.

[1] Global News Wire - ReportLinker

Solution components

Hewlett Packard Enterprise validated protecting a Kubernetes cluster with Commvault software in a lab environment. This section provides details on each major component incorporated into the solution.

Figure 1 illustrates the lab environment's configuration.

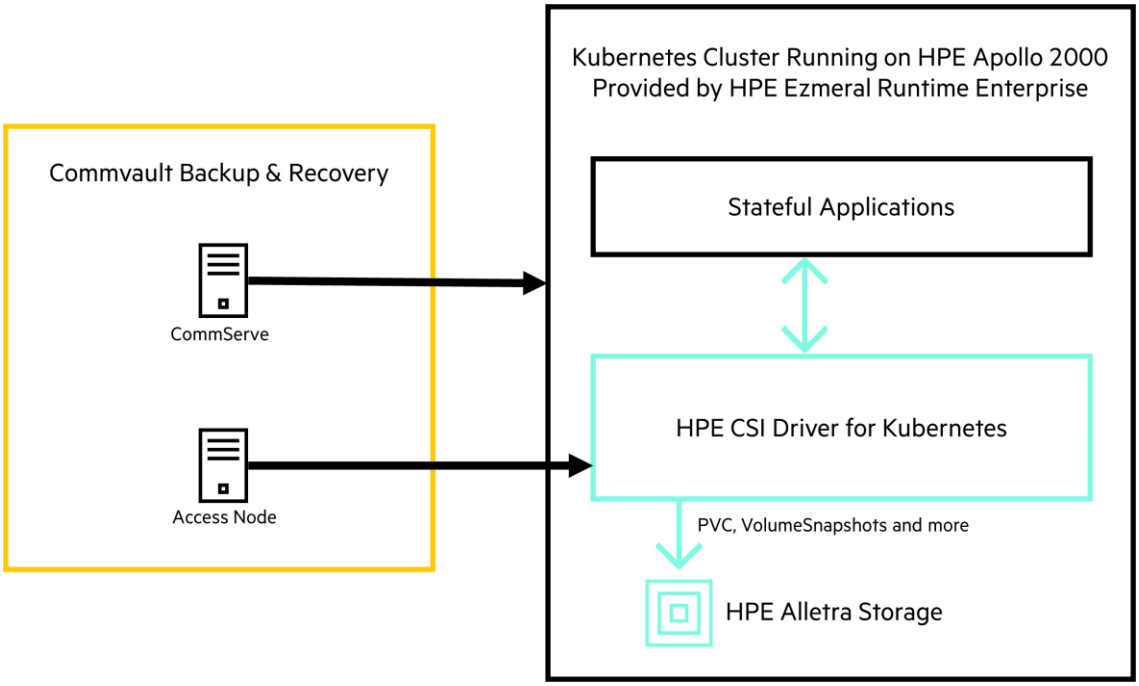


Figure 1. Solution overview

HPE Apollo 2000 servers

New infrastructure must support a diversity of processor technologies and data-intensive workloads in the architecture to enable the converged use of analytics, AI, and high performance compute (HPC) to unlock the potential of your data and accelerate innovation. HPE Apollo systems deliver the flexibility to tailor the system to the precise needs of demanding HPC workloads with the right compute, flexible I/O, and storage options.

The HPE Apollo 2000 system provides a density-optimized compute platform that doubles the density of traditional 1U servers, providing up to four 2P servers in a 2U form factor. The HPE ProLiant XL170r Gen10 server is an ideal node for compute-intensive and in-memory analytics in real-time, interactive, and batch workloads.

HPE Ezmeral Runtime Enterprise

HPE Ezmeral Runtime Enterprise is a secure, enterprise-grade platform to build and deploy cloud-native and non-cloud-native (i.e., legacy) applications at scale across data centers, multiple clouds, and at the edge for a wide range of use cases. It provides all the tools enterprise customers need to build, modernize, deploy, monitor, and manage a wide range of AI and analytics workloads to unleash their data's full potential and accelerate their data-driven digital transformation.

In addition, HPE Ezmeral Runtime Enterprise provides more than just a container orchestration solution. HPE Ezmeral Runtime Enterprise's differentiated value-adds include an integrated data fabric to connect and manage data, enterprise-grade security, GitOps-based policy management and drift management, public cloud cluster imports (unified control plane makes it easy to import external Kubernetes clusters, including from cloud vendors such as Amazon EKS, Google™ GKE™, and Azure AKS), and multi-tenancy.

The Kubernetes cluster is provided by HPE Ezmeral Runtime Enterprise.

HPE CSI Driver for Kubernetes

The HPE CSI Driver for Kubernetes allows you to use a Container Storage Provider (CSP) to perform data management operations on storage resources. Figure 2 illustrates the CSI driver architecture. The architecture of the Container Storage Interface (CSI) driver allows block storage vendors to implement a CSP that follows the REST API specification.

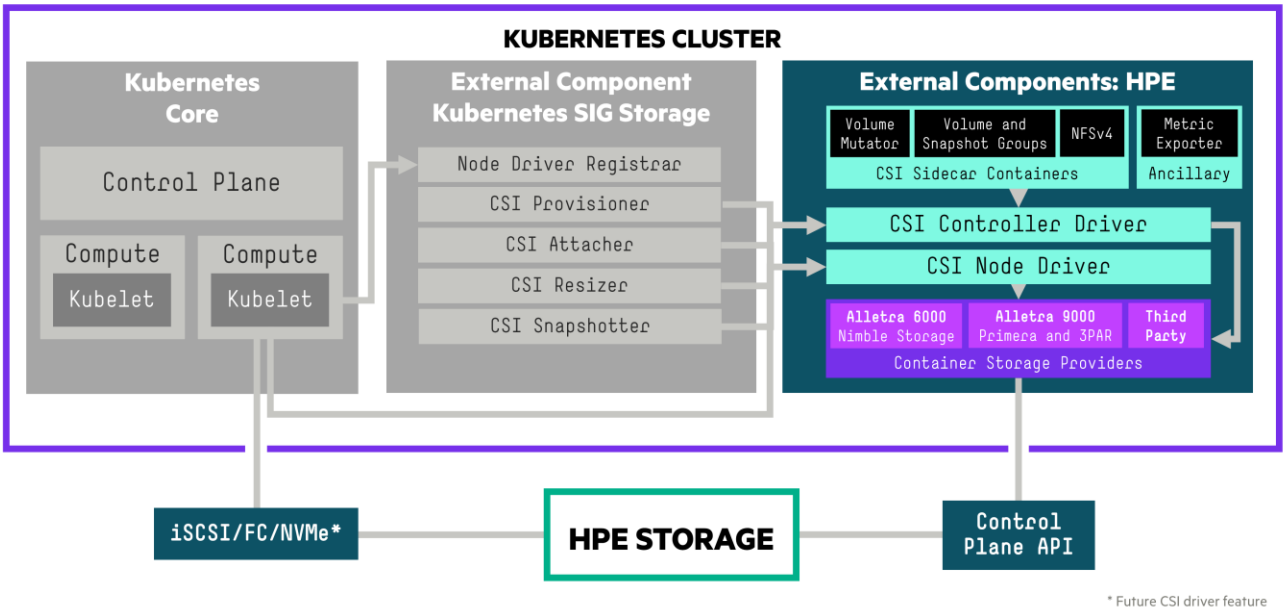


Figure 2. Container Storage Interface with HPE CSI Driver for Kubernetes

The CSI driver architecture provides a complete separation of concerns between upstream Kubernetes core, SIG Storage (CSI owners), CSI driver author (HPE), and the back-end CSP developer.

The HPE CSI Driver for Kubernetes exposes array-level volume snapshots via the Kubernetes `VolumeSnapshotClass`. When configured, Commvault will use `VolumeSnapshots` as an efficient mechanism for persistent storage backup.

Version 2.2.0 of the driver was used in this environment.

Commvault software

Commvault Backup & Recovery is a world-class data protection platform that enables customers to utilize, protect, and move data across on-premises and the cloud. Hewlett Packard Enterprise recommends Commvault Platform Release 2022E (11.28.14) or newer when backing up, restoring, and migrating applications and data for Cloud Native Computing Foundation (CNCF) certified Kubernetes distributions.

Commvault Backup & Recovery natively supports `kube-apiserver`, the core of the Kubernetes control plane. This integration allows for monitoring of the cluster and detection of new objects to protect. Snapshots and persistent volume creation are handled via native CSI integration. Figure 3 depicts an example backup lifecycle, wherein a temporary pod is deployed, and the previously mentioned snapshot volumes are mounted. This pod lives long enough to ensure the snapshot contents are backed up and then these resources are removed. The backup destination is external to the Kubernetes cluster and is defined by the Commvault Server Plan. There is no agent needed in the application containers nor in the Kubernetes nodes.

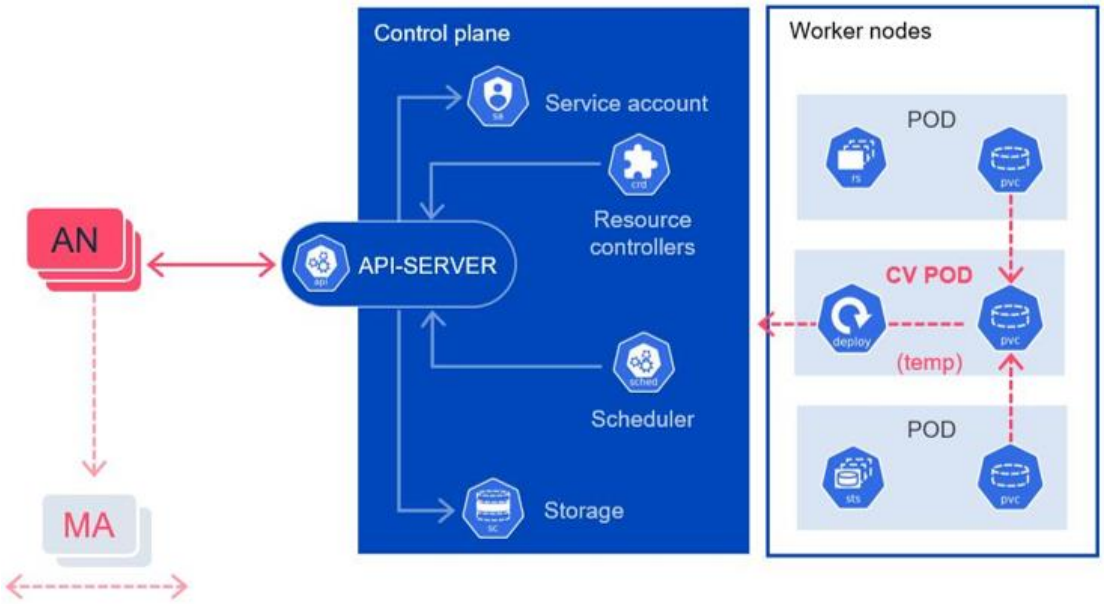


Figure 3. Commvault backup lifecycle of Kubernetes resource (Graphic courtesy of Commvault)

Commvault CommServe Server

The CommServe® Server is the command and control center of the CommCell® architecture. The CommServe Server was primarily used to communicate with a Commvault Access Node to initiate data protection, management, and recovery operations for the Kubernetes environment.

Commvault Access Node

The Commvault Access Node is responsible for coordinating protection operations, receiving data, and storing data and replicating for recovery events.

Testing overview

Installation of the HPE CSI Driver for Kubernetes

Review the HPE Storage Container Orchestrator Documentation to deploy and configure the HPE CSI Driver for the HPE primary storage in your environment. The following is an overview of the steps completed for this environment:

- Deploy using Helm to the "hpe-storage" `namespace`.
- Define `Secret` for the array, specifying IP, credentials, and service name.
- Define `StorageClass` to be used by Kubernetes application workloads. The `StorageClass` contains array-specific parameters that are applied during persistent volume creation.
- Enable CSI Snapshots in support for `VolumeSnapshots` by ensuring the appropriate CRDs are installed.
- Define a default `VolumeSnapshotClass`. Commvault will detect this default class and use it to take snapshots during backup operations.

With the Kubernetes environment prepared and the HPE CSI Driver configured, Commvault provides guidance to verify the Kubernetes cluster is ready for integration into Commvault.

Commvault installation and configuration guidance

The Commvault Platform Release 2022E Essential Kubernetes documentation is leveraged for installing and configuring Commvault protection for Kubernetes.

Review the Kubernetes Restrictions and Known Issues ("Expert" documentation, available with login) to understand the latest available protection schemes.

The following actions were taken to prepare and integrate the Kubernetes cluster into Commvault Command Center:

- The Commvault CommServe Server was installed on a host with a Windows Server operating system.
- A Linux® Access Node was added (RHEL) via the Command Center UI.
- Follow the Guided Setup documentation from Commvault. The following are a summary of steps to configure protection for the Kubernetes cluster:
  - Verify Kubernetes environment is compliant and pre-requisites are met.
  - Identify/Create a Kubernetes `Service Account`.
  - Obtain Kubernetes cluster API endpoint and `Service Account` token.
  - Specify Commvault Access Nodes and Server.
  - Define an initial namespace-level Application protection group.

Figure 4 shows the Add Cluster step from the Guided Setup.

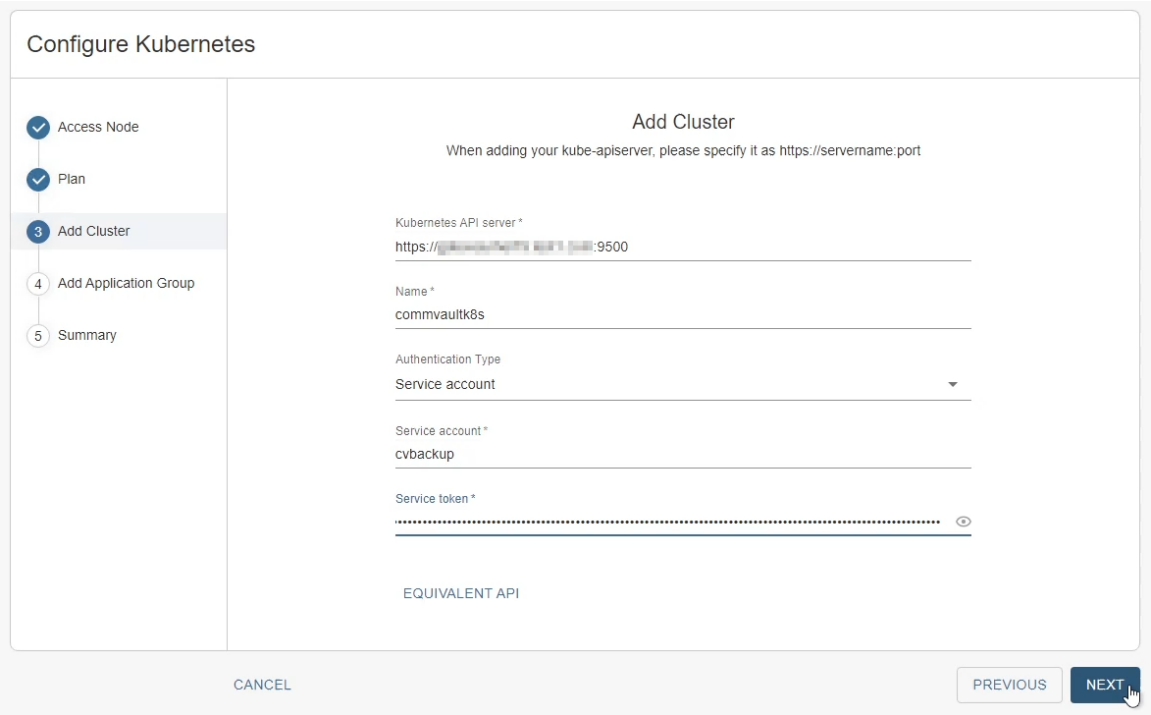


Figure 4. Adding a Kubernetes cluster: provide cluster API endpoint, SA and SA token, and relevant access node.

The newly added Kubernetes cluster can be viewed by selecting Protect → Kubernetes in the navigation pane of the Command Center.

Define protection for Kubernetes applications

With the Kubernetes cluster successfully added, there are several ways that Application group protection can be defined:

- Application-level: define protection for a specific `StatefulSet`.
- Namespace-level: automatically protect all applications under a namespace, including manifests for their `Pods`, `ConfigMaps`, `Secrets`, and other related API resources.
- Labels: allows for adding or excluding entities from a more encompassing protection specification.
- Full-cluster: automatically discovers and protects all namespaces, applications, `PersistentVolumeClaims`, and non-namespace (cluster-scoped) objects.

During the guided setup, a namespace-level application group "NS-commvault" was defined, as seen in Figure 5.

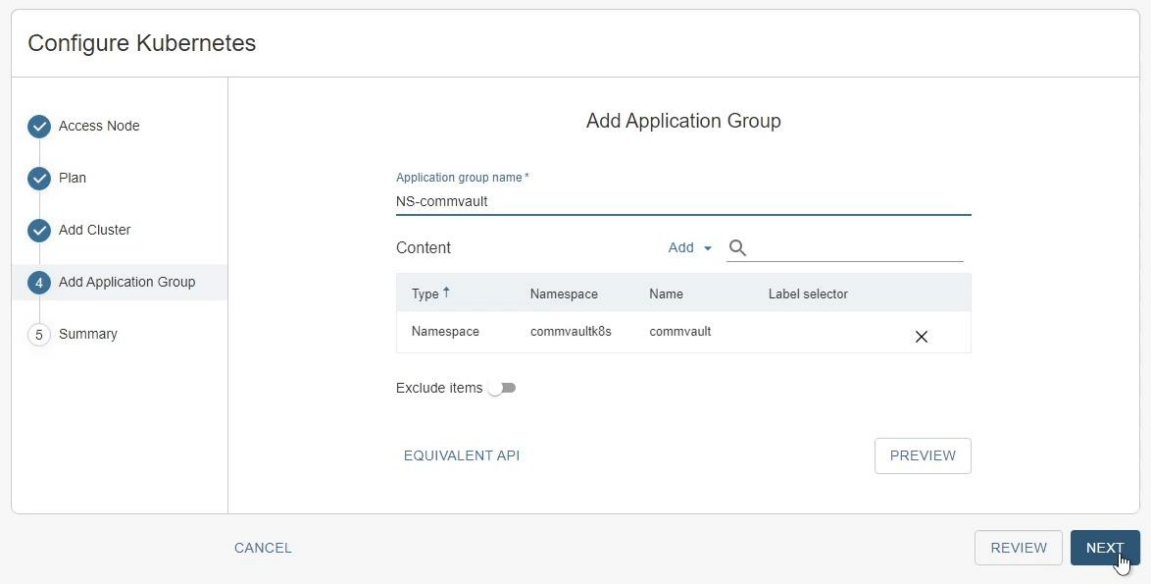


Figure 5. Defining namespace-level protection (application group) for the "commvault" namespace

This namespace will begin backup per the specified Server Plan. The recovery points for the namespace backup can be viewed in the Command Center UI under Protect → Kubernetes, then click the Application groups tab. From this list, "NS-commvault" was chosen. Figure 6 shows the "NS-commvault" application group and available recovery points.

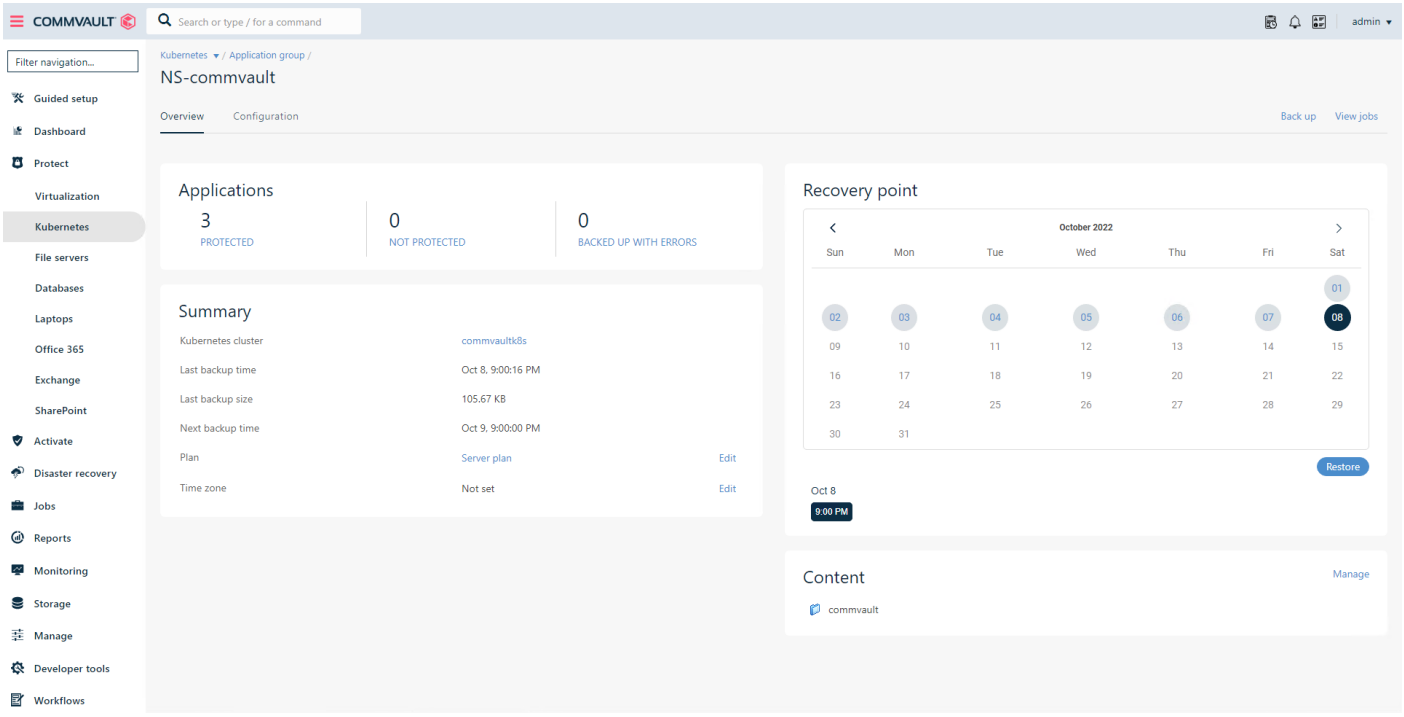


Figure 6. "NS-commvault" application group content and recovery points

Example DevOps use case

In addition to backup and recovery operations, another useful workflow is to restore a recovery point into a separate application for test and development. This can be done by sourcing an existing restore point from a production application and restoring it to a new namespace.

As a prerequisite, a separate namespace was created named "commvault-devtest." This will be the destination for the restore operation and is not generally included in backup protection.

From the Commvault Command Center, select Protect → Kubernetes, then click the Applications tab. From the list, the web application is chosen for this exercise. The view looks like that of Figure 6. Next, choose a recovery point, then click Restore.

Restore options are populated in the final step. As shown in Figure 7, Out of place is selected, because the application is being deployed to a different namespace from its parent. A new application name is given, "web-devtest-dml," as well as specification of the "commvault-devtest" namespace.

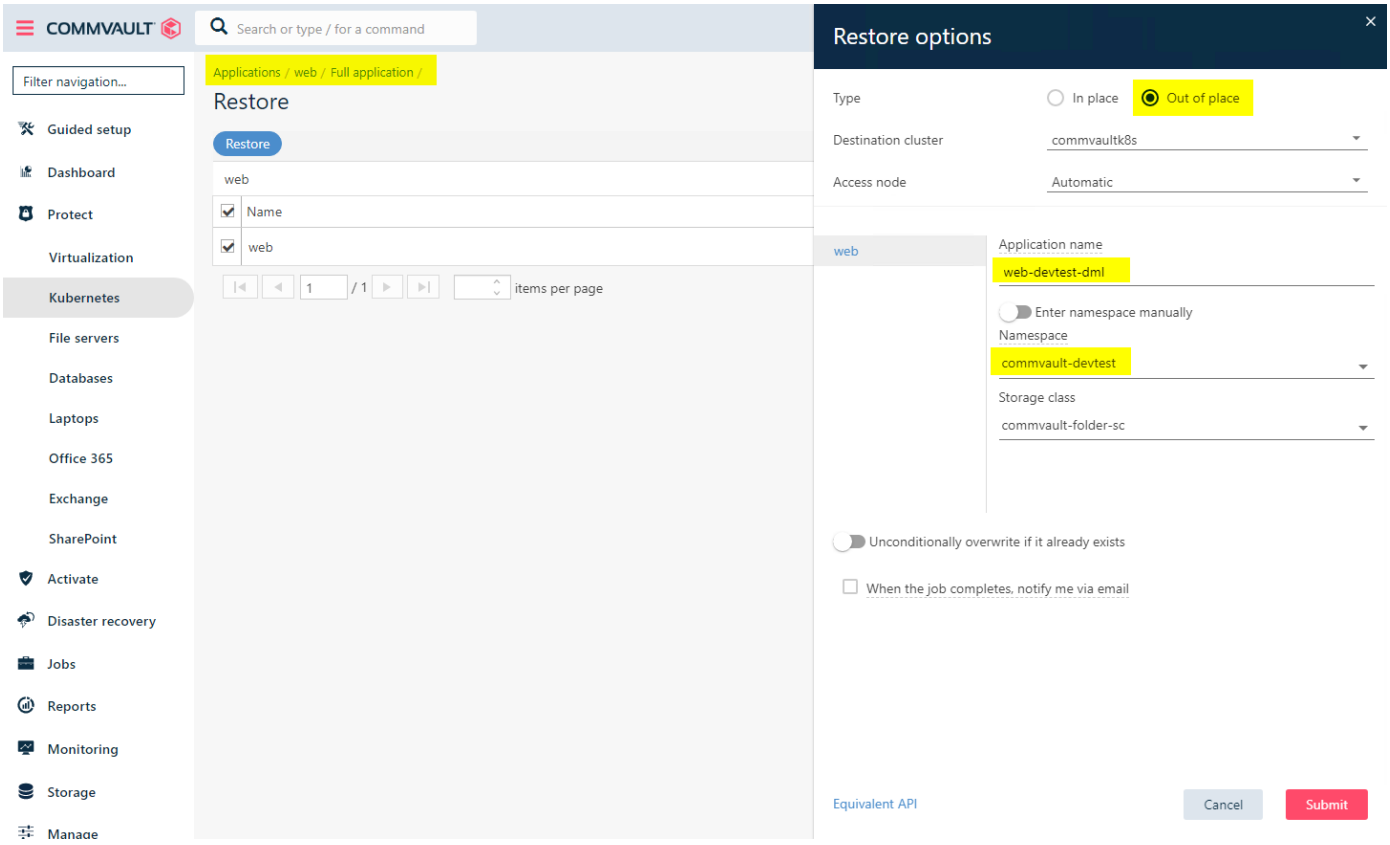


Figure 7. Create a dev/test environment with "Out of place" restore specified

After the restoration job is complete, the "web-devtest-dml" application can now be found in the "commvault-devtest" namespace. Figure 8 shows the Kubernetes Dashboard, as launched from the HPE Ezmeral Runtime Enterprise cluster operations view.

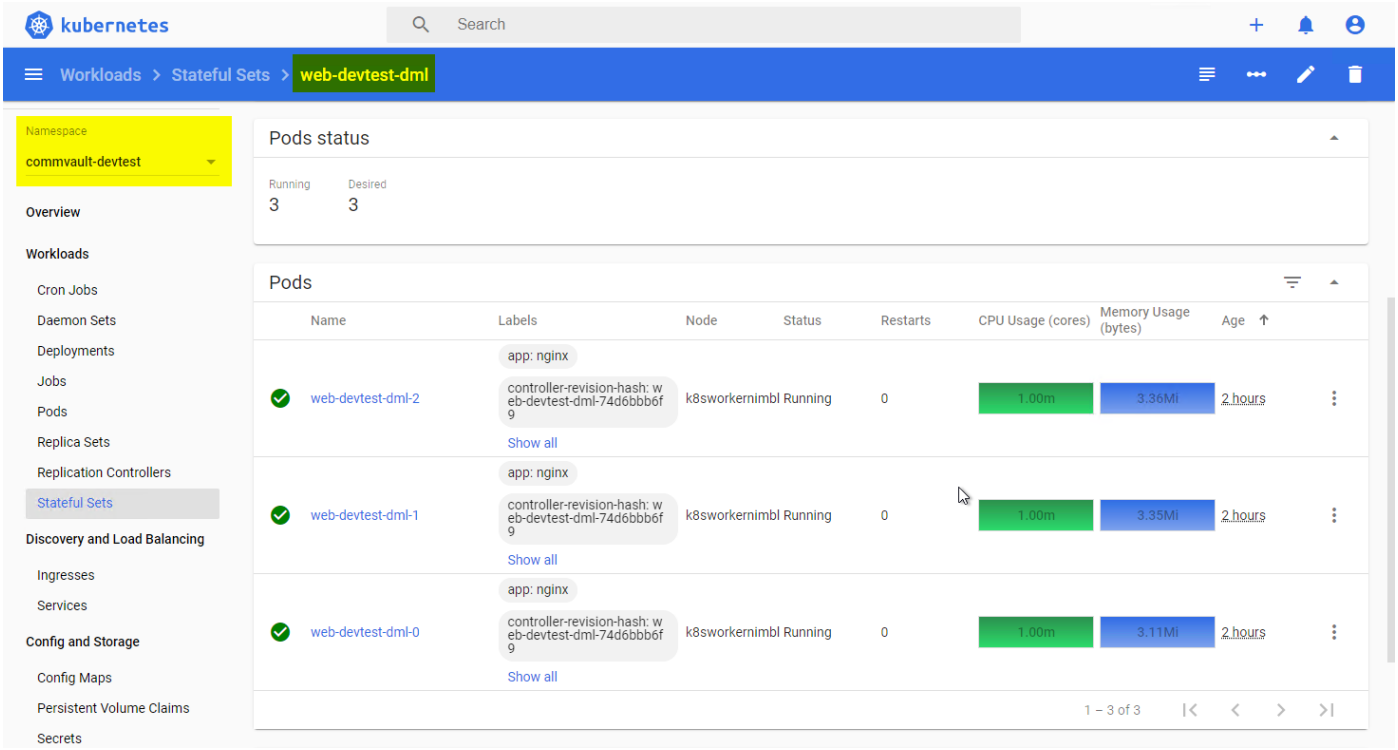


Figure 8. Kubernetes Dashboard UI, viewing the "commvault-devtest" namespace and newly cloned "web-devtest-dml" StatefulSet

Summary

This technical white paper has successfully demonstrated how to configure a Kubernetes cluster to utilize HPE Alletra persistent storage and to protect associated namespaces with Commvault. The steps provided offer guidance for the components and configuration required to add a Kubernetes cluster to the Commvault Command Center and then back up and restore a containerized application utilizing array-based snapshots. The specific use case tested was backing up a `StatefulSet` application, then restoring that application to an alternate `namespace`. The use case was chosen as it highlights the ease at which a parent/production environment can be spun up in a separate cluster or namespace for development purposes. Restoring a backup of the original application to a new `namespace` allows for concurrent workflows.

Due to the dynamic nature of Kubernetes environments, enterprises must be able to quickly deploy containerized AI and analytics applications typically in a variety of environments. That can include physical servers or virtual machines residing on-premises or in private or public cloud. Enterprises need a solution with the compute resources required for supporting containers, including data backup and mobility at scale.

Deploying a Kubernetes environment on HPE Apollo systems with HPE Ezmeral Runtime Enterprise provides flexibility to tailor the system to the precise needs of demanding HPC workloads with the right compute and flexible I/O, while the HPE CSI Driver for Kubernetes integrates HPE Alletra persistent storage options. Commvault Backup & Recovery software was utilized to protect a containerized application with the possibility to move it between clusters, which may be on-premises or in public cloud.

Hewlett Packard Enterprise and Commvault provide end-to-end enterprise-grade data protection, safeguarding data wherever it lives while also improving business continuity.

Resources, contacts, or additional links

HPE Apollo systems
hpe.com/apollo

HPE ProLiant DL servers
hpe.com/us/en/servers/proliant-dl-servers.html

HPE CSI Driver for Kubernetes documentation
scod.hpedev.io/welcome

HPE CSI Driver for Kubernetes and Commvault partner configuration
scod.hpedev.io/partners/commvault

Using Kubernetes CSI with HPE Ezmeral Container Platform
youtube.com/watch?v=1xEsUYm7G04

HPE Ezmeral Runtime Enterprise 5.4 documentation
docs.containerplatform.hpe.com/54

Commvault Backup & Recovery software
commvault.com/complete-data-protection/backup-and-recovery

Commvault Platform Release 2022E Kubernetes documentation:
documentation.commvault.com/2022e/essential/123634_protecting_kubernetes_with_commvault.html

Commvault Platform Release 2022E Expert documentation:
documentation.commvault.com/2022e/expert

HPE & Commvault Partnership
commvault.com/supported-technologies/hpe

HPE Complete Commvault software
buy.hpe.com/us/en/options/complete-storage-solution/commvault/p/1010634827

Learn more at
hpe.com/solutions

a00105578ENW, Rev. 2

Hewlett Packard Enterprise