Dell PowerProtect Data Manager: Protecting Kubernetes Workloads

Abstract
Kubernetes is an open-source platform to manage and orchestrate containerized workloads and services. This document describes the architecture and integration of Kubernetes workloads with Dell PowerProtect Data Manager and details how they protect Kubernetes workloads.

May 2022
Revisions

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 2020</td>
<td>Initial release</td>
</tr>
<tr>
<td>January 2021</td>
<td>Dell PowerProtect Data Manager 19.6 updates</td>
</tr>
<tr>
<td></td>
<td>• PostgreSQL High Availability</td>
</tr>
<tr>
<td></td>
<td>• Cassandra Application Consistent Protection</td>
</tr>
<tr>
<td></td>
<td>• Restore Cluster Scoped Resources</td>
</tr>
<tr>
<td>February 2021</td>
<td>PowerProtect Data Manager 19.7 updates</td>
</tr>
<tr>
<td></td>
<td>• VMware Tanzu Kubernetes Cluster protection</td>
</tr>
<tr>
<td>May 2021</td>
<td>PowerProtect Data Manager 19.8 Updates</td>
</tr>
<tr>
<td></td>
<td>• Storage Class Mapping</td>
</tr>
<tr>
<td>May 2022</td>
<td>PowerProtect Data Manager 19.9 and 19.10 updates</td>
</tr>
<tr>
<td></td>
<td>• Structural custom resource definitions (CRDs)</td>
</tr>
<tr>
<td></td>
<td>• Quick Recovery to an alternate Kubernetes cluster</td>
</tr>
</tbody>
</table>

Acknowledgments

Author: Vinod Kumar Kumaresan

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

This document may contain certain words that are not consistent with Dell's current language guidelines. Dell plans to update the document over subsequent future releases to revise these words accordingly.

This document may contain language from third party content that is not under Dell's control and is not consistent with Dell's current guidelines for Dell's own content. When such third-party content is updated by the relevant third parties, this document will be revised accordingly.

Copyright © 2022 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners. [4/28/2022] [White Paper] [H18563.4]
# Table of contents

Revisions.........................................................................................................................2
Acknowledgments.............................................................................................................2
Table of contents .............................................................................................................3
Executive summary...........................................................................................................5
Audience ..........................................................................................................................5

1 Introduction.....................................................................................................................6
  1.1 Features of Kubernetes ..........................................................................................6
  1.2 PowerProtect Data Manager capabilities for Kubernetes ....................................6
  1.2.1 Efficient and flexible .......................................................................................6
  1.2.2 Built for Kubernetes .......................................................................................6
  1.3 Key components of PowerProtect Data Manager ................................................7
  1.3.1 Cloud Native Data Manager (CNDM) ..............................................................7
  1.3.2 PowerProtect controller ..................................................................................7
  1.3.3 VMware Velero ...............................................................................................7
  1.3.4 cProxy .............................................................................................................7
  1.4 Key components of Kubernetes ............................................................................7
  1.4.1 Cluster ..............................................................................................................7
  1.4.2 Node .................................................................................................................7
  1.4.3 Pods and containers .......................................................................................7
  1.4.4 Kubernetes API (kube-apiserver) ...................................................................8
  1.4.5 Persistent volume and persistent volume claim ..............................................8
  1.4.6 Container storage interface .............................................................................8
  1.4.7 Storage class ....................................................................................................8
  1.4.8 Namespaces.....................................................................................................9
  1.4.9 Custom resource .............................................................................................9

2 Deployment methods of Kubernetes ...........................................................................10
  2.1 Kubernetes on-premises .....................................................................................10
  2.1.1 Kubernetes running on virtual environment ................................................10
  2.1.2 On-premises Kubernetes on bare metal .........................................................10
  2.1.3 Using external CSI ........................................................................................11
  2.2 Kubernetes on Cloud ..........................................................................................11
  2.2.1 Kubernetes deployed on Infrastructure as a Service (IaaS) .............................11
  2.2.2 Kubernetes as a Service (KaaS) .....................................................................11

3 Reference architecture ...............................................................................................12
# Table of contents

3.1 Protection of Kubernetes clusters ................................................................................................................................................. 12
3.2 Structured custom resource definitions ............................................................................................................................................. 13
3.3 Protection of VMware Tanzu Kubernetes Grid ................................................................................................................................. 14

4 Configuring PowerProtect Data Manager for protecting Kubernetes workloads .......................................................................................... 15
4.1 Adding a Kubernetes cluster asset source ............................................................................................................................................... 15
4.1.1 Enable Kubernetes asset source .................................................................................................................................................. 15
4.1.2 Add a Kubernetes cluster ............................................................................................................................................................ 16
4.2 Add a protection policy for Kubernetes namespace protection ............................................................................................................. 18

5 Performing a backup of namespaces and PVCs on Kubernetes cluster ............................................................................................................. 20

6 Application-consistent database backups in Kubernetes .............................................................................................................................. 21
6.1.1 Supported database applications ................................................................................................................................................... 21
6.1.2 Kubernetes MySQL Application Consistent protection ...................................................................................................................... 21
6.1.3 Kubernetes PostgreSQL Application Consistent Protection ............................................................................................................... 23
6.1.1 Application template of PostgreSQL ........................................................................................................................................ 23
6.1.2 Kubernetes Cassandra application-consistent protection ................................................................................................................... 24

7 Restoring Kubernetes namespaces and PVCs ............................................................................................................................................ 25
7.1 Quick recovery to an alternate Kubernetes cluster ............................................................................................................................... 31
7.1.1 PowerProtect Data Manager quick recovery for Kubernetes cluster ............................................................................................ 31

A Technical support and resources ................................................................................................................................................................. 33
A.1 Related resources .................................................................................................................................................................................. 33
Executive summary

Traditionally, organizations used physical servers to run applications. There was no way to define resource boundaries for applications in a physical server, and this caused resource allocation issues. As a solution, virtualization was introduced, making it possible to run multiple virtual machines (VMs) on a single physical server’s CPU. It also enabled applications to be isolated within VMs and provided a level of security.

Modern infrastructure is being transformed by containers. Containers are like virtual machines but have relaxed isolation properties to share the operating system. The container has its own file system, CPU, memory, and process space. Agile application creation, continuous development, environmental consistency across development, application-centric management, efficient resource allocation, and resource isolation are the key benefits of containers. Kubernetes is an open-source container management platform that unifies a cluster of machines into a single pool of compute resources.

With a distributed container deployment, it is important to protect your workloads. Dell PowerProtect Data Manager protects the Kubernetes workloads and ensures high availability, and consistent and reliable backup and restore capabilities for Kubernetes workloads or for a disaster recovery (DR) situation. PowerProtect Data Manager offers centralized management, automation, multi-cloud options, and advanced integration to simplify managing workloads.

Audience

This white paper is intended for customers, partners, and other users who want to understand how PowerProtect Data Manager helps protect Kubernetes workloads.
1 Introduction
Kubernetes is an open-source container orchestrator for managing containerized workloads and services, enabling both declarative configuration and automation. It is portable, extensible, and scalable and has a large, rapidly growing ecosystem. Kubernetes services, support, and tools are widely available.

Dell PowerProtect Data Manager protects existing and newly discovered workloads. It allows IT operations and backup administrators to manage Kubernetes clusters and their protection through a single management UI and define protection policies for Kubernetes workloads from Kubernetes APIs. The policy-driven protection is defined by the protection policy mechanism. PowerProtect Data Manager discovers the namespaces, labels, and pods in the environment, and helps protect them by providing cluster credentials logging, monitoring, governance, and recovery.

1.1 Features of Kubernetes
Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications.

- Kubernetes automates Linux container operations and eliminates many of the manual processes involved in deploying and scaling containerized applications.
- Applications can be clustered together in group of hosts running Linux containers, and Kubernetes helps you easily and efficiently manage those clusters.
- Kubernetes is an ideal platform for hosting cloud-native applications that require rapid scaling.

1.2 PowerProtect Data Manager capabilities for Kubernetes
PowerProtect Data Manager provides enterprise-level protection for Kubernetes.

1.2.1 Efficient and flexible
- PowerProtect Data Manager provides a single platform for data protection. It manages different workloads such as VMs, applications, and the containers through one platform.
- Protection to deduplicated storage allows great total cost of ownership (TCO) with PowerProtect DD series.

1.2.2 Built for Kubernetes
PowerProtect Data Manager enables the following:

- Provides flexible protection for Kubernetes clusters using the Kubernetes APIs.
- Discovers, monitors, and protects Kubernetes resources—namespaces, labels, pods, and persistent volumes.
- Does not require installing a backup client container for each pod for the backup process.
- Provides protection to controllers per node to avoid cross-node traffic.
- Enables application consistency for MySQL and MongoDB databases.
- Can restore assets to another cluster that is connected to PowerProtect Data Manager.
- Provides protection for AWS-hosted Kubernetes clusters using PowerProtect Data Manager running on AWS and protected to PowerProtect Data Domain running on AWS.
1.3 Key components of PowerProtect Data Manager

1.3.1 Cloud Native Data Manager (CNDM)
The Cloud Native Data Manager (CNDM) is an integrated microservice component of PowerProtect Data Manager which communicates with the kube-apiserver of the cluster. This component is responsible for APIs for the backup and restore process.

1.3.2 PowerProtect controller
PowerProtect controller is the component which is installed on the Kubernetes cluster when the cluster is discovered by PowerProtect Data Manager. The backup and restore controllers manage BackupJob CR and RestoreJob CR definitions. The PowerProtect controller is responsible for the backup and restore of persistent volumes.

1.3.3 VMware Velero
VMware Velero is the open-source tool which is integrated with PowerProtect Data Manager. It is integrated and is not required to be installed separately. The Velero component is pushed into the Kubernetes cluster by the PowerProtect controller pod after the cluster is in the running state using the Velero deployment object. It is responsible for the backup and restore of metadata.

1.3.4 cProxy
The containerized proxy (cProxy) is a stateless containerized proxy. It is installed on the Kubernetes cluster when the backup and restore process is initiated, and is deleted after the process is completed. It is responsible for managing persistent volume snapshots (snap copies), mounting snapshots, and moving the data to the target storage. It is also responsible for restoring data into a persistent volume from target storage and making the data available for attaching to pods.

1.4 Key components of Kubernetes

1.4.1 Cluster
A Kubernetes cluster is a group of machines called nodes that run containerized applications and has a wanted state that defines which applications or workloads should be running. The cluster’s wanted state is defined with the Kubernetes API.

1.4.2 Node
A node is defined for a virtual or physical machine, depending on the cluster. Each node contains the services necessary to run pods and is managed by the control-plane components. There are two kinds of nodes: control-plane node and worker node.

1.4.3 Pods and containers
Pods operate at one level above the individual container. Multiple containers can be encapsulated within a pod. A Kubernetes pod is a group of containers that are deployed together on the same host. The pod is sometimes called as container when a single container is frequently deployed.
1.4.4 **Kubernetes API (kube-apiserver)**

The Kubernetes API server is the control-plane of the Kubernetes cluster that exposes the Kubernetes API. It serves as the foundation for the declarative configuration schema for the system. The `kubectl` command-line tool can be used to create, update, delete, and get API objects.

1.4.5 **Persistent volume and persistent volume claim**

A persistent volume (PV) is storage defined for the cluster that is provisioned by an administrator or dynamically provisioned using storage classes (SCs). It is a resource in the cluster similar to a node. PVs are volume plug-ins like volumes but have a life cycle independent of any individual pod that uses the PV. It captures the details of the implementation of the storage that is NFS, iSCSI, or a cloud-provider-specific storage system.

A persistent volume claim (PVC) is a request for storage by a user. It is like a pod, which consumes node resources. Similarly, PVCs consume PV resources. Pods can request specific levels of resources (CPU and memory).

1.4.6 **Container storage interface**

Container storage interface (CSI) defines a standard interface for container orchestration systems to expose arbitrary storage systems to respective container workloads. A CSI-compatible volume driver is deployed on a Kubernetes cluster so that users can use the CSI volume type to attach or mount the volumes exposed by the CSI driver.

1.4.7 **Storage class**

A storage class is described as the type of storage that is provisioned and allowed ranges for size and IOPS. When user creates a PVC, that specifies the storage class with size in GB and number of IOPS. A storage class is used to abstract the underlying storage platform.
1.4.8 Namespaces
A namespace is defined as Kubernetes object which partitions a single Kubernetes cluster into multiple virtual clusters. Namespaces are intended for the use in environment with many users spread across multiple teams or projects.

1.4.9 Custom resource
A resource in the Kubernetes environment is an endpoint for an API that stores a collection of API objects of a certain kind. A custom resource (CR) is an extension of the Kubernetes API that is not necessarily available in a default Kubernetes installation. It represents a customization of a particular Kubernetes installation.
2 Deployment methods of Kubernetes

Kubernetes is an open-source container orchestrator for automating deployment, scaling, and managing containerized workloads. There are few methods to deploy Kubernetes clusters and protected accordingly. Kubernetes can be deployed on-premises or on the cloud.

2.1 Kubernetes on-premises

Kubernetes is described as a cloud-native technology. However, the cloud-native concept includes the use of on-premises infrastructure.

2.1.1 Kubernetes running on virtual environment

The virtual environments are mostly VMware based. With this method of deployment, the persistent volumes are carved with VMware vSphere managed storage—VMware vSAN or VMware Virtual Volumes (vVols)—as a first-class disk. A first-class disk (FCD) refers to an improved virtual disk (IVD) which is a feature of VMware vSphere. The FCD is the independent disk which is not associated with VM. When Persistent Volumes for cloud native application is created, a virtual disk (VMDK) is attached to Kubernetes node. The disk backup of FCD is similar to that of other VMDKs.

The container orchestrates volume snapshot backup with which the volume snapshot is taken, mounted, and streamed block data to target storage. For application-level backups, pre- and post-hooks are used which quiesce the database, flush, and take snapshots of Persistent Volumes.

2.1.2 On-premises Kubernetes on bare metal

Bare-metal servers, such as Linux servers, run containers and Kubernetes nodes on physical servers. In this method of deployment, the Persistent Volumes are carved out from underlying storage that is SAN or NAS and connected through the storage vendor provided custom drivers or through standardized CSI drivers. The container orchestrates volume snapshot backup with which the volume snapshot is taken, mounted, and streamed block data to target storage. For application-level backups, the database dump backup is taken online and streamed data to target storage.
2.1.3 Using external CSI

Container storage interface (CSI) uses an open standard API that enables Kubernetes to expose arbitrary storage systems to containerized workloads. In this method of deployment, external CSI drivers are used to access infrastructure and external storage such as OpenStack.

2.2 Kubernetes on Cloud

There are few clouds which are supported for Kubernetes. Database backups are taken online and streamed data to cloud bucket. The container orchestrates cloud snapshots. Kubernetes deployment in the cloud that is Amazon Web Services (AWS), Microsoft Azure or Google Cloud have two variants.

2.2.1 Kubernetes deployed on Infrastructure as a Service (IaaS)

In this scenario, the Linux hosts running containers are deployed on IaaS plane (for example, AWS EC2 instance), which means the underlying infrastructure is provided and is similar to running on a virtualized environment. In such case, the user must deploy the Kubernetes clusters either upstream or through a distribution on the IaaS instance of the cloud provider.

2.2.2 Kubernetes as a Service (KaaS)

Various cloud providers offer Kubernetes as a service. In such case, the cloud provider manages the Kubernetes environment completely and offers it as a service to the user. Below is the list of cloud providers and the Kubernetes as a service:

- Kubernetes on Google Cloud: Google Kubernetes Engine (GKE)
- Kubernetes on Amazon Web Services (AWS): Elastic Kubernetes Service (EKS)
- Kubernetes on Microsoft: Azure Kubernetes Service (AKS)
- Kubernetes on Rancher: Rancher Kubernetes Engine (RKE)

Also, the Kubernetes environment can be a generic upstream instance that is an open-source version of Kubernetes, Distro (Kubernetes distribution) with Rancher and CoreOS. It automates the provisioning of the Kubernetes cluster using an installer script and OEM distribution that automates Kubernetes, which is Red Hat OpenShift, Anthos, or Pivotal Container Service (PKS).
3 Reference architecture

PowerProtect Data Manager protects Kubernetes workloads and ensures the data is consistent and highly available. PowerProtect Data Manager is a virtual appliance that is deployed on an ESXi host using OVA and is integrated with PowerProtect DD series appliances as protection target where backups are stored.

A Kubernetes cluster is group of a control-plane and worker nodes. The control-plane node manages the worker nodes, and it handles scheduling the pods across the nodes in the cluster. PowerProtect Data Manager integrates with Kubernetes cluster through Kubernetes APIs to perform the discovery. The CNDM is the component of PowerProtect Data Manager which communicates with the kube-apiserver of the cluster.

Once the discovery of cluster completes, the cluster is added as PowerProtect Data Manager asset source and associated namespaces as assets are available to be protected. During the process of the discovery, PowerProtect Data Manager creates below two namespaces in the cluster. The data is compressed and deduplicated at the source and sent to the target storage.

- **Velero-ppdm**: Contains a Velero pod to backup metadata and stage to the target storage if there is a BareMetal environment. It performs PVC and metadata backup if there is VMware Cloud Native Storage (CNS).
- **PowerProtect**: Contains a PowerProtect controller pod to drive Persistent Volume Claim snapshot and backup and push the backups to target storage using intermittently spawned cProxy pods.

3.1 Protection of Kubernetes clusters

PowerProtect Data Manager discovers the Kubernetes clusters using the IP address or FQDN. PowerProtect Data Manager uses the discovery service account and the token (kubeconfig file) to integrate with kube-apiserver. PowerProtect Data Manager protects two types of assets of Kubernetes cluster that are Namespaces and PersistentVolumeClaims (PVCs). The PowerProtect controller is responsible for the backup and restore of PVCs. The VMware Velero component is responsible for the backup and restore of metadata that is saved in the target storage.
A protection policy enables selecting a specific group of assets to be backed up. The protection policy is created using the PowerProtect Data Manager UI for the backup schedule and retention lock. When a protection policy is created, a new storage unit (SU) is created on the protection storage as part of protection policy configuration.

With Kubernetes workloads, a BackupStorageLocation containing the SU information is also created on the cluster. The PowerProtect controller running in the Kubernetes cluster creates a corresponding BackupStorageLocation in the Velero namespace whenever a BackupStorageLocation is created in the PowerProtect namespace.

### 3.2 Structured custom resource definitions

Starting with version 19.9, PowerProtect Data Manager uses structured customer resource definitions (CRDs) with Kubernetes to increase system security. PowerProtect Data Manager versions earlier than 19.9 used non-structural schemas that did not validate the consistent format of custom resources (CRs). When updating from older versions of PowerProtect Data Manager, the self-service recovery of the last backup copy taken before the update is not possible. To restore that backup, use the PowerProtect Data Manager user interface. There are important Kubernetes-specific factors to consider when updating PowerProtect Data Manager in a Kubernetes environment. For more information, see the [Dell PowerProtect Data Manager Kubernetes User Guide](#).

**Note:** For general information about updating PowerProtect Data Manager, see the [Dell PowerProtect Data Manager Kubernetes User Guide](#).
3.3 Protection of VMware Tanzu Kubernetes Grid

Data Manager 19.7 onwards introduces ability to protect Kubernetes workloads on VMware Tanzu Kubernetes Grid (TKG). VMware vSphere 7U1 re-architectures vSphere with native Kubernetes as its control plane. A TKG cluster is a Kubernetes cluster that runs inside the virtual machines on supervisor layer which allows to run Kubernetes with consistency. It is enabled through the TKG service for VMware vSphere and is upstream-compliant with open-source Kubernetes (guest cluster). The guest cluster is a consistent Kubernetes cluster that runs on VMs and consists of its control-plane VM, management plane VM, worker nodes, pods, and containers.

**Note:** For more information about Dell PowerProtect Data Manager protecting VMware Tanzu Kubernetes Clusters, see the document *PowerProtect Data Manager: Protecting VMware Tanzu Kubernetes Clusters*. 
4 Configuring PowerProtect Data Manager for protecting Kubernetes workloads

4.1 Adding a Kubernetes cluster asset source

Adding a Kubernetes cluster as an asset source in PowerProtect Data Manager enables you to protect namespaces and persistent volume claims (PVCs) within the cluster. You can use the asset sources window in the PowerProtect Data Manager UI to add a Kubernetes cluster asset source.

See the section “Prerequisites to Kubernetes cluster discovery” in the PowerProtect Data Manager Kubernetes User Guide before adding a Kubernetes cluster as an asset source with Data Manager.

4.1.1 Enable Kubernetes asset source

An asset source must be enabled in PowerProtect Data Manager before you can add and register the asset source for the protection of assets.

The Kubernetes asset source can be enabled from the PowerProtect Data Manager UI. Click Infrastructure > Asset Sources, and click + (plus) to view the New Asset Source tab. In the pane for the asset source that you want to add, click Enable Source. The Asset Sources window updates to display a tab for the new asset source.
## 4.1.2 Add a Kubernetes cluster

A Kubernetes cluster can be added as an asset source in PowerProtect Data Manager to protect the namespaces and PVCs within the cluster.

**Note:** Discovery of a Kubernetes cluster discovers namespaces that contain volumes from both container storage interface (CSI) and non-CSI based storage. However, backup and recovery are supported only from CSI-based storage. Also, only PVCs with the VolumeMode Filesystem are supported.

### Add Kubernetes

- **Name**: Cluster name
- **Address**: The fully qualified domain name (FQDN) or the IP address of the Kubernetes API server.
  **Note**: We recommend using the FQDN instead of the IP address.
- **Port**: Specify the port to use for communication when not using the default port, 443.
  **Note**: The use of any port other than 443 or 6443 requires you to open the port on PowerProtect Data Manager first to enable outgoing communication.

**Host Credentials**: The service account must have the following privileges:

- Get/Create/Update/List CustomResourceDefinitions
- Get/Create/Update ClusterRoleBinding for 'cluster-admin' role
Create/Update 'powerprotect' namespace
Get/List/Create/Update/Delete/List
Get/List/Create/Update/Delete all kinds of resources inside 'powerprotect' namespace
Get/List/Watch all namespaces in the cluster and PV, PVC, storageclass, deployments, and pods in these namespaces

Note: The admin-user service account in the kube-system namespace contains all these privileges. You can provide the token of this account, or an existing similar service account. Alternatively, create a service account that is bound to a cluster role that contains these privileges, and provide the token of this service account.

When adding the Kubernetes cluster as an asset source, a PowerProtect controller is installed on the cluster. This controller is also used to install Velero with the DD object-store plug-in and the vSphere plug-in.

For more details, see the section “Add a Kubernetes cluster” in the PowerProtect Data Manager Kubernetes User Guide.
The namespaces in the Kubernetes cluster appear in the Kubernetes tab of the Assets window.

4.2 Add a protection policy for Kubernetes namespace protection

Protection policies define a set of objectives that apply to specific periods of time. These objectives drive configuration, active protection, and copy-data-management operations that satisfy the business requirements for the specified data. Each plan type has its own set of user objectives. Users with the system admin role can create protection policies. A Kubernetes protection policy enables you to select namespaces in the Kubernetes cluster that you want to back up. Use the Data Manager UI to create a Kubernetes namespace protection policy.

Data Manager provides the following options when creating a Kubernetes cluster protection policy:

- **Crash Consistent**: Select this type for point-in-time backup of namespaces.
- **Exclusion**: Select this type if there are assets within the protection policy to exclude from data protection operations.
Also, the admin can select namespaces and associated PVCs statically or dynamically for inclusion or exclusion in protection policies, along with schedules, retention, and other protection operations.

From the Jobs window, the progress of the new Kubernetes cluster protection policy backup and associated tasks can be monitored.

For more information about creating a protection policy, see the section “Add a protection policy for Kubernetes namespace protection” in the PowerProtect Data Manager Kubernetes User Guide.
Performing a backup of namespaces and PVCs on Kubernetes cluster

Apart from the scheduled backup option, PowerProtect Data Manager supports the option to perform manual backups.

To perform a manual backup, go to Protection > Protection Policies. Select the Kubernetes protection policy and click Protect Now as shown below.

During the manual backup, select the backup type of either Full or Synthetic Full.

From the Jobs > Protection Jobs window, you can monitor the progress of the Kubernetes cluster backup.

For more information about performing backup and recovery of Kubernetes workloads using PowerProtect Data Manager, see the PowerProtect Data Manager Kubernetes User Guide.
6 Application-consistent database backups in Kubernetes

PowerProtect Data Manager supports agentless, application-consistent backups of database applications that reside in Kubernetes pods. Application-consistent backups occur when the database application is informed of a pending backup. The database completes all pending transactions and operations, while typically queuing new requests. This process places the database in a quiescent state of relative inactivity where the backup represents a true snapshot of the application. This backup now captures items that would have otherwise been stored only in memory. After the snapshot, the application resumes normal functionality. In most environments, the snapshot operation is instantaneous, so downtime is minimal.

These backups are agentless, in that the PowerProtect Data Manager can take a snapshot of containers without the need for software installation in the database application environment. Then, that snapshot is backed up using the normal procedures for the Kubernetes environment.

The PowerProtect Data Manager provides a standardized way to quiesce a supported database, back up the data from that database, and then return the database to operation. Application templates serve as a bridge between a specific database environment and the Kubernetes backup architecture for the PowerProtect Data Manager. Depending on the differences between database environments, each deployment may require a different configuration file.

6.1.1 Supported database applications

Supported applications include:

- **MySQL** in the following configurations:
  - Standalone deployment in one pod.
  - Cluster (primary/secondary) deployment with multiple StatefulSets or ReplicaSets. For example, through Helm.

- **MongoDB** without shards.

- **PostgreSQL** in the following configurations:
  - Standalone deployment in one pod.
  - Cluster (primary/secondary) deployment with multiple StatefulSets. For example, through Helm.

- **Cassandra** without shards.

Because data syncs from the primary pods to secondary pods, the PowerProtect Data Manager backs up secondary pods first.

6.1.2 Kubernetes MySQL Application Consistent protection

Data Manager has Application Consistent protection feature for MySQL and MongoDB databases. To utilize this feature, user creates a MySQL Application consistent backup using templates. PowerProtect Data Manager has in-built ‘ppdmctl.tar.gz’ file available which contains the templates for MySQL and MongoDB in YAML format. This file is pushed from PowerProtect Data Manager to Kubernetes cluster root directory to enable the application-consistent protection. The protection is provided only to the namespace which has asset defined in the protection policy. When the protection policy is applied with enabled MySQL template, the consistency of the backed-up data is application consistent and when the MySQL template is disabled, the consistency is crash consistent.
Kubernetes namespaces are already discovered as assets with PowerProtect Data Manager. There are the available namespaces for protection except powerprotect and velero-ppdm namespaces. To configure the application-consistent protection, particular namespace is specified. In this case, test-namespace is used to demonstrate.

1. Log in to the PowerProtect Data Manager CLI with administrator credentials.
2. Change the directory to cndm/misc using the below command.
   
   ```sh
   cd /usr/local/brs/lib/cndm/misc
   ```

3. Run the list command (ls) to view the content which is ppdmctl.tar.gz file.
   
   ```sh
   ls -ltrh
   ```

4. Run the below command to transfer file from PowerProtect Data Manager to Kubernetes cluster (K8s-master01 is for reference for Kubernetes cluster).
   
   ```sh
   scp ppdmctl.tar.gz tme@k8s-master01
   ```

5. Log in to Kubernetes Cluster, run the list command to list the content of the root directory, and confirm `ppdmctl.tar.gz` is available.
6. Run the tar command to untar the file. Once untar is completed, the ppdmctl directory is created.
   
   ```sh
   tar -xvf ppdmctl.tar.gz
   ```

7. Run the change directory command to enter the ppdmctl directory, and run ls to list the content.
   
   ```sh
   cd ppdmctl
   ls
   ```

8. Run the below command to enter the examples directory where all the MySQL and MongoDB templates are stored.
   
   ```sh
   cd examples
   ls
   ```

9. Copy the `mysqlapptemplatehelm.yaml` file to create a template copy from the example for MySQL with command (tmedemomysqlapptemplatestadalone.yaml is the existing template associated to the test-namespace).
   
   ```sh
   cp mysqlapptemplatehelm.yaml tmedemomysqlapptemplatestadalone.yaml
   ```

10. Run the below command to describe the MySQL stateful set.
    
    ```sh
    kubectl describe sts mysql -n test-namespace
    ```

11. Edit the tmedemomysqlapptemplatestadalone.yaml file to make the changes.

   - To change the name of primary pod as per the pod description, edit:
     ```yaml
     selectorExpression: "mysql" # (mysql is new name)
     ```
   - If there are no worker pods being used for MySQL, remove the `selectorTerms` section under `selectors` for worker.
   - To edit the MySQL password environment variable, preHook and postHook are edited. (Pod-based hooks execute hook code in a new pod derived from template in a deployment configuration, preHook is to quiesce and postHook is to be unquiesce.)
12. Run `cd..` to exit the examples directory to the ppdmctl directory.
13. To create a template for MySQL application-consistent protection from the edited (step11) template file in test-namespace, use the ppdmctl utility by running the following command.

```
./ppdmctl template create tmemysqltemplate -type=mysql -namespace=test-namespace -inputfile=examples/tmedemomysqlapptemplatestadalone.yaml
```

The application-consistent protection is configured for specific namespace (test-namespace in this demonstration). When the backup is run manually or using protection policy, the backup copy is saved as application consistent at the target storage.

14. To disable the MySQL template, run the below command.

```
./ppdmctl template disable tmemysqltemplate -namespace=test-namespace
```

### 6.1.3 Kubernetes PostgreSQL Application Consistent Protection

PowerProtect Data Manager supports both PostgreSQL and PostgreSQL HA (high availability). Both PostgreSQL and PostgreSQL HA configure a cluster with a primary-standby topology. The primary node has writing permission while replication is on the standby nodes which have read-only permissions. PowerProtect Data Manager has the integrated `ppdmctl.tar.gz` file which contains the templates for PostgreSQL in YAML format. This file is pushed to the primary node. Pgpool-II acts as proxy for PostgreSQL backend. It reduces connection overhead and used as a load balancer for PostgreSQL. Pgpool-II is responsible to spread the queries among nodes.

#### 6.1.1 Application template of PostgreSQL

**AppLabel:** “app=postgresql”  
**Type:** “PostgreSQL”  
**AppActions:**

Pod:
**PreHook:** 
```
"["/bin/bash","-c","PGPASSWORD=$POSTGRES_PASSWORD psql -U $POSTGRES_USER -c "select pg_start_backup('ppdm-backup',true);"""]"
```

**PostHook:** 
```
"["/bin/bash","-c","PGPASSWORD=$POSTGRES_PASSWORD psql -U $POSTGRES_USER -c "select pg_stop_backup();"""]"
```

---

**Application template of PostgreSQL HA:**

**AppLabel:** “app.kubernetes.io/name=postgresql-ha,app.kubernetes.io/component=postgresql”

# applabel to select
pod Type: “PostgreSQL”  
**AppActions:**

Pod:
**PreHook:** 
```
"["/bin/bash","-c","REPMGR_PRIMARY_HOST=`echo $REPMGR_PRIMARY_HOST | cut -f1 -d '.'`; if [ $HOSTNAME = $REPMGR_PRIMARY_HOST ]; then PGPASSWORD=$POSTGRES_PASSWORD psql -U $POSTGRES_USER -c "select pg_start_backup('ppdm-backup',false);"""]"
```

---

23 Dell PowerProtect Data Manager: Protecting Kubernetes Workloads | H18563.4
6.1.2 Kubernetes Cassandra application-consistent protection

Apache Cassandra is highly scalable, high-performance distributed database designed to handle large amounts of data with no single point of failure. It is a type of NoSQL database. Cassandra partitions data over storage nodes using consistent hashing algorithm. Each node stores multiple ranges of tokens and each range of token replicates to multiple nodes for fault-tolerance and high availability. PowerProtect Data Manager has Application Consistent protection feature for Cassandra database. PowerProtect Data Manager has the integrated `ppdmctl.tar.gz` file which contains the templates for Cassandra in YAML format.

**Application template of Cassandra:**

```
AppLabel: "app=cassandra"
Type: "Cassandra"
Enable: true
AppActions:
Pod:
PreHook: "\"/bin/bash\", \"-c\", \"nodetool flush\""]"
```

For more details, see the section “Application-Consistent Database Backups in Kubernetes” in the [PowerProtect Data Manager Kubernetes User Guide](#).
7 Restoring Kubernetes namespaces and PVCs

After protecting the Kubernetes cluster protection policy, restoring namespace and PVCs can be done from individual namespace backups. When a protection policy is successfully backed up, PowerProtect Data Manager displays details such as the name of the storage system containing the asset backup, location, the creation and expiry date, and the size.

To view backup copies available for restore, select Restore > Assets on the PowerProtect Data Manager UI. Select the asset and click Restore.

PowerProtect Data Manager provides options to recover the Kubernetes namespaces to the same or to an alternate cluster.

**Restore to Original Cluster**: Select this option to restore to a new namespace on the original cluster.

**Restore to an Alternate Cluster**: Select this option to restore to a new namespace on a different cluster, and then select the cluster from the list.

A restore to an alternate cluster can be useful when:

- Migrating namespaces from a cluster on-premises to a cluster in the cloud.
- Moving namespaces from a lower cluster version to a higher cluster version.
- Moving from one environment to another (for example, from a test environment to a production environment).

**Note**: When restoring to an alternate cluster, ensure that this Kubernetes cluster has been added and discovered in the PowerProtect Data Manager UI Asset Sources window.
**Restore Namespace and Select PVCs:** Select this option to restore namespace resources and selected persistent volume claims (PVCs). Optionally, you can also select **Include cluster scoped resources** to restore the cluster roles, cluster role bindings, and custom resource definitions (CRDs) that were backed up automatically as part of the Kubernetes protection policy. This option is only available for PowerProtect Data Manager 19.6 and later Kubernetes protection policy backups.

![Kubernetes Restore](image)

**Restore Cluster Scoped Resources:**

The resources that are scoped at a cluster level and not bound to any specific namespace are called cluster scoped resources (for example, cluster roles, cluster role bindings, and custom resource definitions (CRD)). When the CRD is created, Kubernetes API server creates a new RESTful resource path for the specific version created. The CRD can be either namespaced or cluster scoped as specified in scope field. This section examines how you can use Kubernetes backup copies and restore the cluster scoped resources such as service accounts, cluster roles, and cluster bindings. The Velero component performs the backup and restore of cluster scoped resources, including the backup of the namespace and the associated cluster scoped resources such as cluster roles, cluster role bindings. Custom resource definitions are included as a part of each namespace backup.

Restoring cluster resources is controlled by a check-box option in UI named **Include cluster scoped resources** and runs the restore process.

Below are the steps to verify the cluster role, cluster binding, and CRD.

1. Log in to Kubernetes cluster and run the below command to view cluster role.
   ```bash
   kubectl get clusterrole
   ```
2. Run the below command to view the clusterrolebinding.
   ```bash
   kubectl get clusterrolebinding
   ```
3. Run the below command to view the cluster resource definitions.
   ```bash
   kubectl get crd
   ```
4. Run the below command to view the service account associated with the namespace.
   ```bash
   kubectl get serviceaccount -n <namespace>
5. To delete all the cluster scoped resources associated with the namespace, run the below script on Kubernetes cluster.

   s./del-fcd.sh

After protecting the Kubernetes cluster protection policy, restoring namespace and PVCs can be done from individual namespace backups.

Use the following recovery options:

- **Restore to original namespace**: Restore to the original namespace on the original cluster.
- **Restore to new namespace**: Create a namespace and restore to this location on the original cluster or a different cluster.
- **Restore to existing namespace**: Restore to an existing namespace in the original cluster or a different cluster.
On the PVCs page, if the configuration of the namespace you want to restore is different from the configuration in the target namespace, perform the following:

- **Select Overwrite content of existing PVCs** to overwrite existing PVCs in the target location with the PVCs being restored if the PVCs have the same name.
- **Select Skip restore of existing PVCs** to restore selected PVCs without overwriting existing PVCs in the target location if they have the same name.

![Kubernetes Restore](image)

Optionally, if you choose to retire the storage class on the original cluster, perform the following:

- **Select Change storage class for PVCs to compatible storage class**. The PVCs that are part of the restore display.
- **Select the check box next to the PVCs for which you want to change the storage class on the target cluster**.

The storage class mapping feature with PowerProtect Data Manager 19.8 enables you to choose an alternate storage class for PVCs with a certain provisioner type while restoring persistent volumes. Storage class mapping enables restoring namespaces and PVCs from one cluster to another using different container storage. It is also useful when the migration of data from one storage class to another storage class and from on-premises to cloud or conversely.

**Note:** The storage class is not modified for existing PVCs being overwritten.
If Change storage class for PVCs to compatible storage class is selected as shown above, the Storage Class page appears with a list of supported storage classes on the target cluster.

1. Select the check box next to a PVC for which you want to change the storage class on the target cluster. Alternately, select multiple PVCs to change all selections to the same storage class.

   **Note:** When changing the PVC storage class on the target Kubernetes cluster, if you select more than one PVC at a time on this page, only the storage classes that apply to all selected PVCs are displayed. To view and select from all available storage classes, select one PVC at a time.

2. Click Target Storage Class to select from the available storage classes. The Select Storage Class dialog appears.

3. Select from one of the available storage classes, and click **Save** to save your changes and return to the Storage Class page.
From the **Summary** page, click **Restore** to initiate the restore job. An informational dialog box appears indicating that the restore has started.

![Summary page screenshot]

From the **Jobs > Protection Jobs** window, the restore progress can be monitored.

![Jobs Protection Jobs window screenshot]
7.1 Quick recovery to an alternate Kubernetes cluster

The quick recovery feature enables you to restore assets and data that you replicated to a destination system at a remote site. Quick recovery is supported for the protected assets of the following PowerProtect Data Manager asset sources:

- Virtual machines
- Kubernetes
- File system
- NAS

Quick recovery sends metadata from the source system to the destination system, following the same flow of backup copies. This metadata makes the replication destination aware of the copies and enables the recovery view. You can recover your workloads at the remote site before you restore the source PowerProtect Data Manager system.

7.1.1 PowerProtect Data Manager quick recovery for Kubernetes cluster

Starting with version 19.10, PowerProtect Data Manager supports quick recovery to an alternate Kubernetes cluster for Kubernetes workloads.

On a quick-recovery-enabled environment with a source and remote PowerProtect Data Manager configuration, Kubernetes assets that have protected and replicated copies created in the source PowerProtect Data Manager will be synced to the remote PowerProtect Data Manager.

The replicated copies will be available for restore to a Kubernetes cluster that is added to the remote PowerProtect Data Manager.
Remote PowerProtect Data Manager displaying the assets of the source PowerProtect Data Manager for restore:

![Image of PowerProtect Data Manager displaying assets]

Remote PowerProtect Data Manager displaying the replicated backup copies of the source PowerProtect Data Manager asset for restore:

![Image of PowerProtect Data Manager displaying replicated backup copies]
Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

The Data Protection Info Hub provides expertise that helps to ensure customer success with Dell data protection products.

A.1 Related resources

Dell PowerProtect Data Manager:

- PowerProtect Data Manager Kubernetes User Guide
- PowerProtect Data Manager Administration and User Guide
- PowerProtect Data Manager compatibility matrix