DELL Technologies

Technical White Paper

Dell EMC PowerProtect Data Manager Protecting OpenShift Workloads

Abstract

This white paper describes the integration of Red Hat OpenShift with Dell EMC™ PowerProtect Data Manager and how OpenShift Kubernetes workloads can be protected.

September 2021

H18715.1

Revisions

Date	Description
May 2021	Initial release
September 2021	Document revised for PowerProtect Data Manager 19.9 release

Acknowledgments

Author: Charu

Table of contents

Executive summary

As global organizations embark on their digital transformation, container technologies are widely being adopted based on their ease of use, portability, cost savings and independence between applications and infrastructure. Containers have their own file system, CPU, memory, and process space which allows organizations to build and run scalable applications in modern, dynamic environments such as public, private, and hybrid clouds. Containers provide a simple way for development and operations teams to increase productivity consistently, through agile application creation, continuous development, environmental consistency across development, application-centric management, efficient resource allocation and resource isolation.

Kubernetes is an open-source container management platform that unifies a cluster of machines into a single pool of compute resources. OpenShift is a Platform-as-a-Service (PaaS) which is built on top of Kubernetes. It helps to develop, deploy, and manage container-based applications. OpenShift provides a self-service platform to create, modify and deploy applications on demand, thus enabling faster development and release life cycles. As business-critical applications move to the OpenShift platform, organizations need to protect these applications and application data. But, protecting an OpenShift environment is not as simple as applying a traditional backup and DR solution to this container space.

As an innovative leader in protecting Kubernetes, Dell Technologies has evolved, innovated, and integrated with OpenShift to address the needs of the new container infrastructure. Dell Technologies offers proper protection for OpenShift, including the unique complexities that come with it. Dell EMC PowerProtect Data Manager offers a centralized platform to protect OpenShift workloads. It ensures high availability and consistent and reliable backup and restore of workloads.

1 Introduction

Containers are transforming modern IT infrastructure. Containers provide an environment to run the applications independent of infrastructure and operating system. Kubernetes is a container orchestrator for managing containerized workloads and services, that facilitates both declarative configuration and automation. OpenShift is an open-source container platform based on Kubernetes, which automates the development to deployment workflow. OpenShift provides the capability to deploy the application using predefined image builders or using the Docker images. With currently distributed container deployment, it is important to protect the workloads. Yet, protection for cloud native workloads is a major challenge in container adoption.

PowerProtect Data Manager protects OpenShift Kubernetes workloads ensuring data is easy to backup and restore, remains available, consistent, and durable in a Kubernetes workload or disaster recovery situation. PowerProtect Data Manager provides a centralized management UI where protection policies can be defined to manage the clusters, namespaces, and other OpenShift components.

1.1 OpenShift components

1.1.1 Namespaces

Namespaces provide the scope (or context) for names. More specifically, namespaces provide the scope for named resources that describes the application and how it should be deployed. Namespaces are also a way to divide cluster resources between multiple uses.

1.1.2 Projects

A project is a concept added in OpenShift which manages the access to the namespace. Projects in OpenShift therefore provide the walls between namespaces, ensuring that users, or applications, can only see and access what they have the permission for.

1.1.3 Web Console

OpenShift provides a web-based user interface (UI) to visualize, browse, and manage the contents of the project. The web console provides an easier to use environment based on application templates.

1.1.4 Pods

Pods are the most basic unit in OpenShift. They consist of one or more containers guaranteed to be running on the same host. The containers within a pod share a unique IP address. Each pod is sized for the workload and has explicit resource reservations for that workload.

1.1.5 Persistent Volumes (PV) and Persistent Volume Claims (PVC)

A Persistent Volume is storage defined for the cluster, 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 plugins like Volumes but have a life cycle independent of any individual Pod that uses the PV. It captures the details of the implementation of NFS, iSCSI, or a cloud-provider-specific storage system. A Persistent Volume Claim (PVC) is a request for storage by a user. Similar to how Pods consume node resources, PVCs consume PV resources.

1.2 PowerProtect Data Manager components

1.2.1 Cloud Native Data Manager

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

1.2.2 PowerProtect Controller

PowerProtect Controller is the component which gets installed on the Kubernetes cluster when the cluster gets discovered by PowerProtect Data Manager. The backup and restore controllers manage BackupJob Custom Resource (CR) and RestoreJob CR definitions and are responsible for the backup and restore of Persistent Volumes.

1.2.3 Velero

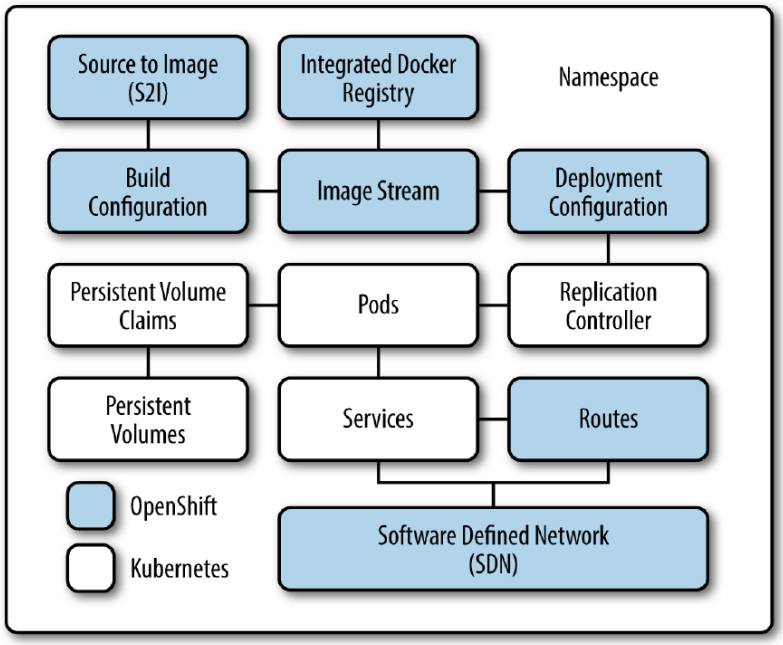
Velero is an open-source tool which is integrated with PowerProtect Data Manager. It is built-in and does not require separate installation or management. The Velero component is pushed into a Kubernetes cluster by the PowerProtect controller pod, using the Velero deployment object. It is responsible for backing up and restoring metadata.

1.2.4 Containerized Proxy (cProxy)

The stateless cProxy gets installed on the Kubernetes cluster when the backup and restore process initiates and is deleted once those processes are 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 Persistent Volumes from target storage and making the data available for attaching to Pods. Also, it acts as an agent plug-in orchestrator for application aware backups.

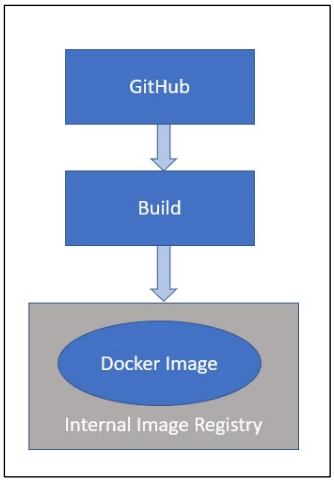
2 Architecture

OpenShift is a PaaS platform and adds several additional components on top of the standard Kubernetes meta data components including Build, BuildConfig, ImageStream, ImageStreamTag, ImageTag, DeploymentConfig. These additional components support Source to Image and Image to deployment workflow that takes an existing source code repository and converts it to associated container or Docker images. When restoring OpenShift namespaces, these components need to be protected. Red Hat released a special **OpenShift plug-in** that allows these components to be protected.



2.1 Build Configuration

The build configuration contains a description of how to build source code and a base image into a new image, which is the primary method for delivering changes to the application. The OpenShift platform converts source code to container images (Docker) that are stored either in the internal image registry or an external repository such as Docker Hub. The output of the build process is an image, which is stored in an integrated Docker registry ready for distribution out to nodes when the application is deployed.



Below is an example of a BuildConfig object definition, which results in a new build every time a container image tag or the source code changes.

```
kind: "BuildConfig"
apiVersion: "v1"
metadata:
  name: "ruby-sample-build" (1)
spec:
  runPolicy: "Serial" (2)
  triggers: (3)
    -
      type: "GitHub"
      github:
        secret: "secret101"
    - type: "Generic"
      generic:
        secret: "secret101"
    -
      type: "ImageChange"
  source: (4)
    git:
      uri: "https://github.com/openshift/ruby-hello-world"
  strategy: (5)
    sourceStrategy:
      from:
        kind: "ImageStreamTag"
        name: "ruby-20-centos7:latest"
  output: (6)
    to:
      kind: "ImageStreamTag"
      name: "origin-ruby-sample:latest"
  postCommit: (7)
      script: "bundle exec rake test"
```

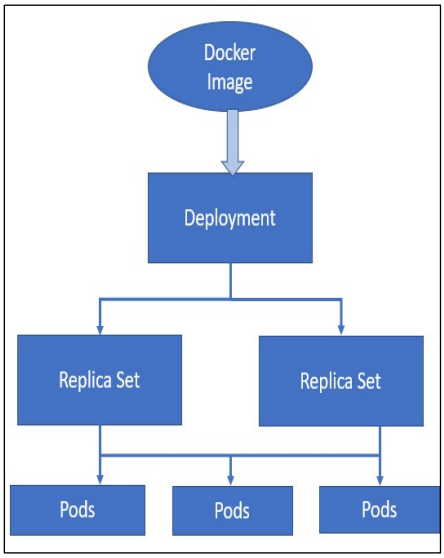
1. This specification creates a new BuildConfig named “ruby-sample-build”.
2. The runPolicy field decides if the builds created from this build configuration can be run simultaneously. The default value is Serial, which means that new builds run sequentially, not simultaneously.
3. A list of triggers can be specified, which causes a new build to be created.
4. The source section describes the source of the build which can be either git, Dockerfile or binary to accept binary payloads.
5. The strategy section defines the build strategy used to run the build.
6. After the container image is successfully built, it is pushed into the repository described in the output section.
7. The postCommit section describes an optional build hook.

2.2 Image Stream

The image stream is how OpenShift tracks the image and its versions. An image stream and its associated tags provide an abstraction for referencing container images from within the OpenShift Container Platform.

2.3 Deployment Configuration

The deployment configuration defines the template for a pod and manages deploying new images or configuration changes. The result of a deployment is the replication controller, which then manages the pods and keeps them running.



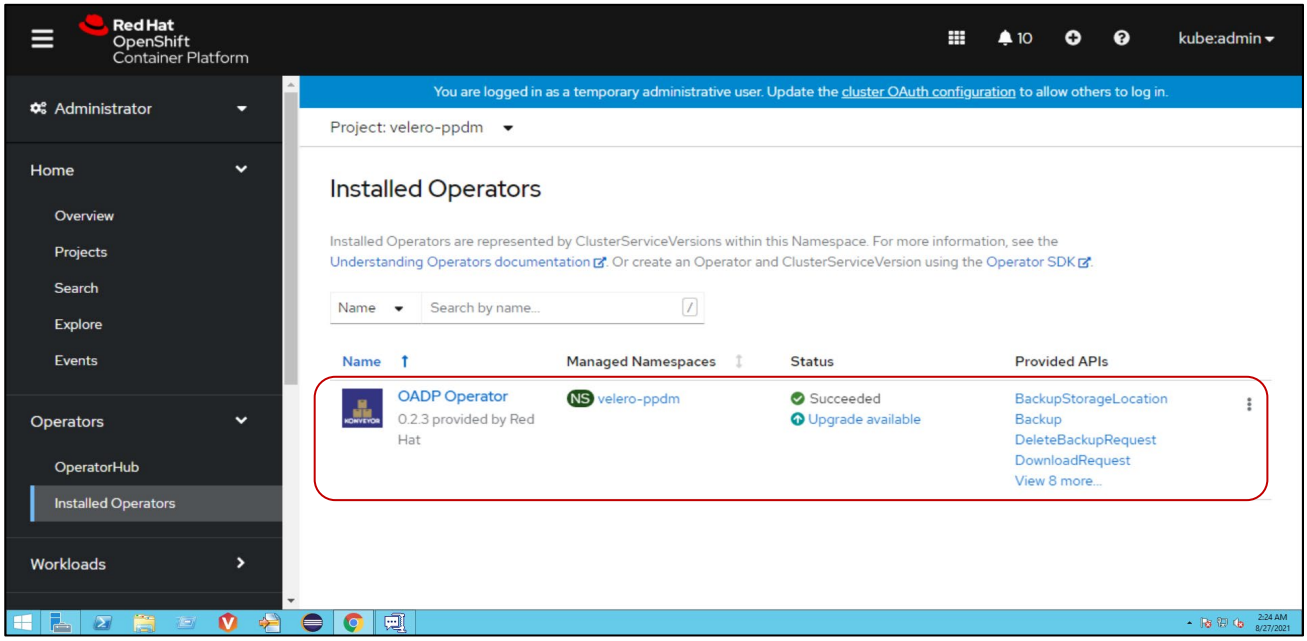
The following example describes the parameters of deploymentConfig resource.

```
kind: "DeploymentConfig"
apiVersion: "v1"
metadata:
  name: "frontend"
spec:
  template: (1)
    metadata:
      labels:
        name: "frontend"
    spec:
      containers:
        - name: "helloworld"
          image: "openshift/origin-ruby-sample"
          ports:
            - containerPort: 8080
              protocol: "TCP"
  replicas: 5 (2)
  triggers:
    - type: "ConfigChange" (3)
    - type: "ImageChange" (4)
      imageChangeParams:
        automatic: true
        containerNames:
          - "helloworld"
        from:
          kind: "ImageStreamTag"
          name: "origin-ruby-sample:latest"
  strategy: (5)
    type: "Rolling"
  paused: false (6)
  revisionHistoryLimit: 2 (7)
  minReadySeconds: 0 (8)
```

1. The replication controller template named “frontend” describes a simple Ruby application.
2. The number of replicas defined is 5 by default.
3. A configuration change trigger causes a new deployment to be created any time the replication controller template changes.
4. An image change trigger causes a new deployment to be created each time a new version of the origin-ruby-sample:latest image repository is available.
5. The Rolling strategy is the default and may be omitted.

3 OpenShift data protection

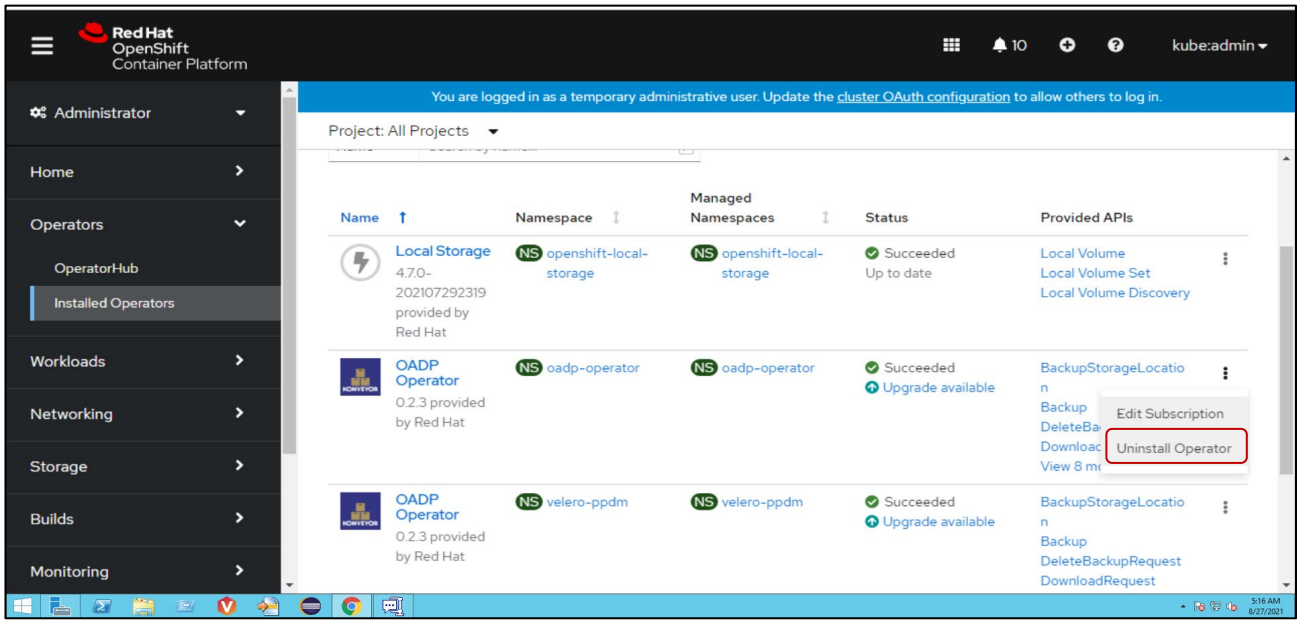
PowerProtect Data Manager controller automatically installs the OpenShift Application Data Protection (OADP) operator in velero-ppdm namespace/project when there is an OpenShift cluster detected during the discovery process and the operator further deploys Velero and the required plugins.



During backups and restores, the OpenShift plug-in will be leveraged to back up the associated OpenShift components. This process is transparent to the user in terms of policy creation and during restores.

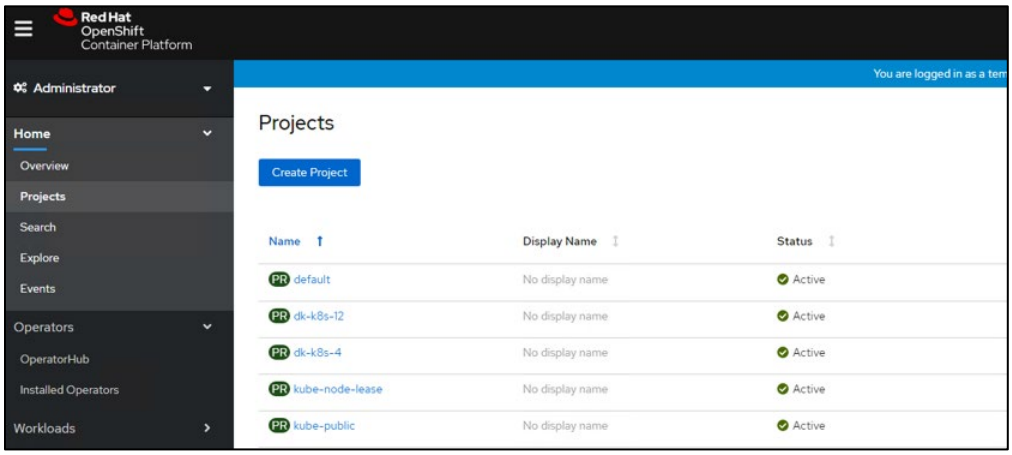
OpenShift is supported for both Container Storage Interface (CSI) snapshots and VMware Cloud Native Storage variants.

Note: Remove any OADP operator installed in a different namespace other than velero-ppdm to avoid any potential custom resource definitions (CRDs) conflict.

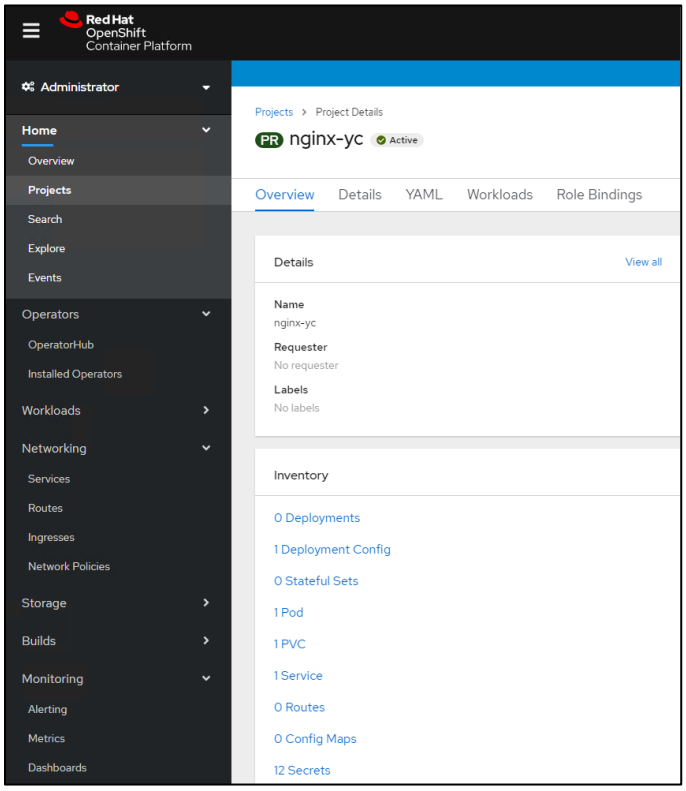


3.1 Assets to be protected

The namespaces that are available to be protected can be seen in the OpenShift UI under Projects.



Each namespace can be expanded, and its inventory can be explored.



This white paper looks at protecting the components of namespaces including pods, deploymentconfig, and imagestream.

```
[core@irv-13-2 ~]$ oc get all,pvc -n nginx-yc
NAME                 READY   STATUS      RESTARTS   AGE
pod/nginx-1-deploy   0/1     Completed   0          45d
pod/nginx-1-kcnzj    1/1     Running     0          45d

NAME                            DESIRED   CURRENT   READY   AGE
replicationcontroller/nginx-1   1         1         1       45d

NAME            TYPE        CLUSTER-IP     EXTERNAL-IP   PORT(S)   AGE
service/nginx   ClusterIP   172.30.70.21   <none>        80/TCP    45d

NAME                                     REVISION   DESIRED   CURRENT   TRIGGERED BY
deploymentconfig.apps.openshift.io/nginx   1          1         1         config,image(nginx:latest)

NAME                                  IMAGE REPOSITORY                                                                TAGS     UPDATED
imagestream.image.openshift.io/nginx   default-route-openshift-image-registry.apps.ocpk8s.ocpdellemc.com/nginx-yc/nginx   latest   6 weeks ago

NAME                                    STATUS   VOLUME                                     CAPACITY   ACCESS MODES   STORAGECLASS   AGE
persistentvolumeclaim/nginx-pvc-claim   Bound    pvc-0e1cbcfc-937b-426b-9c5b-63d1abd3c7a2   2Gi        RWO            mongodb-sc     45d
[core@irv-13-2 ~]$
```

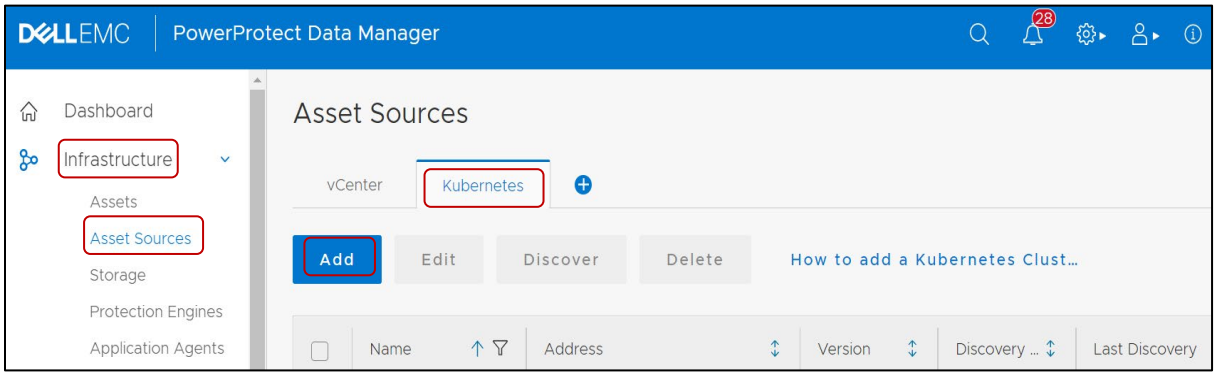
4 Configure PowerProtect Data Manager to protect OpenShift Kubernetes workloads

This section describes the process for registering OpenShift Kubernetes clusters with PowerProtect Data Manager and how these clusters are protected.

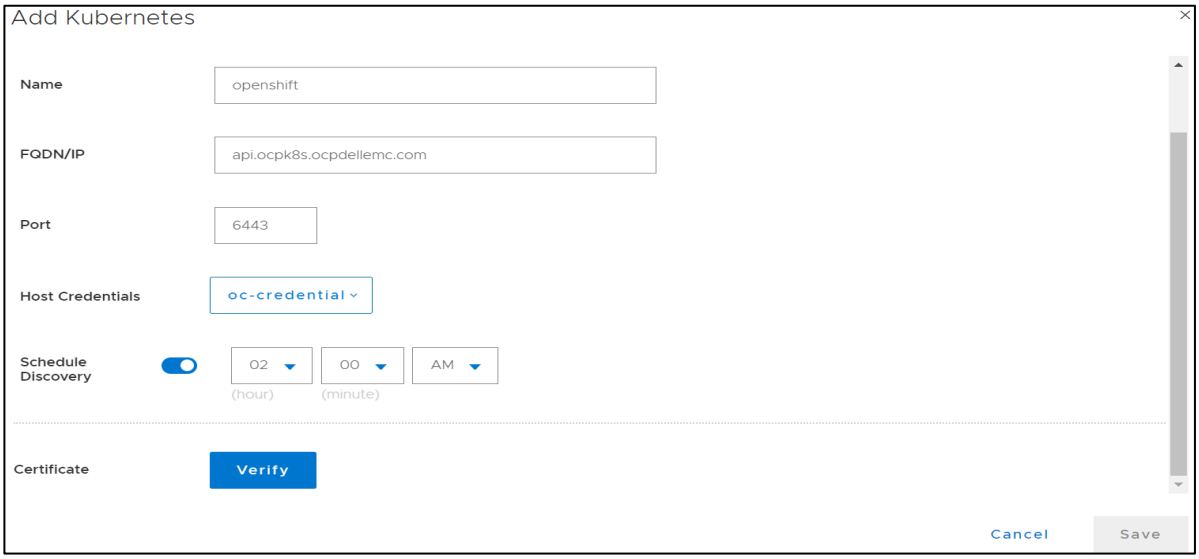
4.1 Asset Discovery

To discover the OpenShift Kubernetes cluster and respective namespaces:

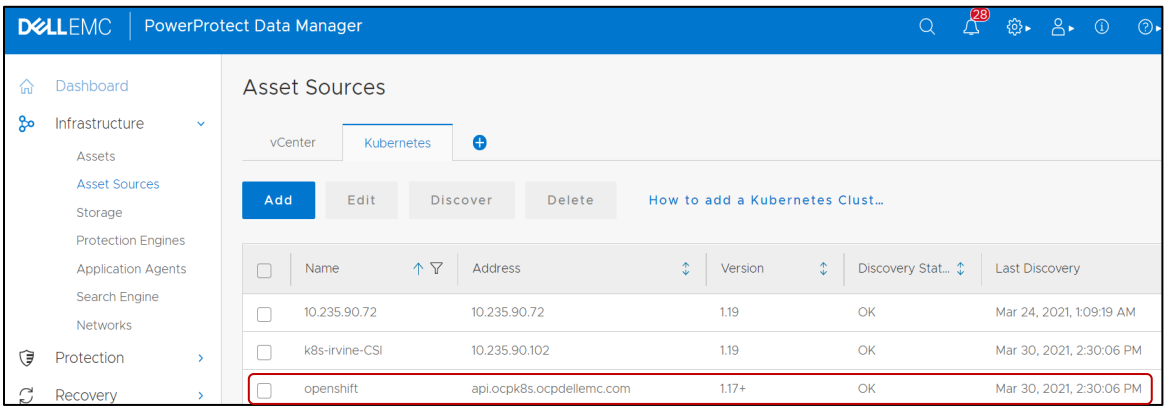
1. Log in to PowerProtect Data Manager UI with administrator credentials.
2. On the left pane of the PowerProtect Data Manager UI, click **Infrastructure**.
3. Select **Asset Sources**, and at the top select **Kubernetes.**
4. Click **Add.**



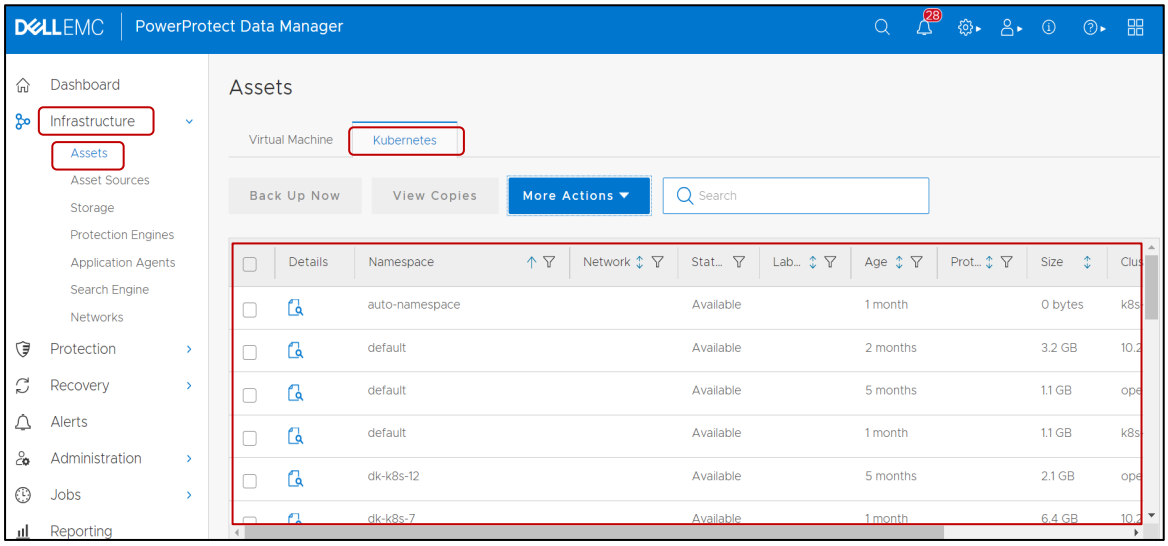
- **Name**: Specify the name
- **FQDN/IP**: Specify the IP address or fully qualified domain name
- **Port**: 6443 (can be changed as per the configuration)
- **Host Credentials**: Select the specified host credential for the cluster
- **Scheduled Discovery**: This is optional and toggle if needed to specify automated discovery at given time schedule
- Click **VERIFY** to authenticate the credentials
- Once verified click **Save**



The OpenShift Kubernetes cluster is now available under **Asset Sources.**

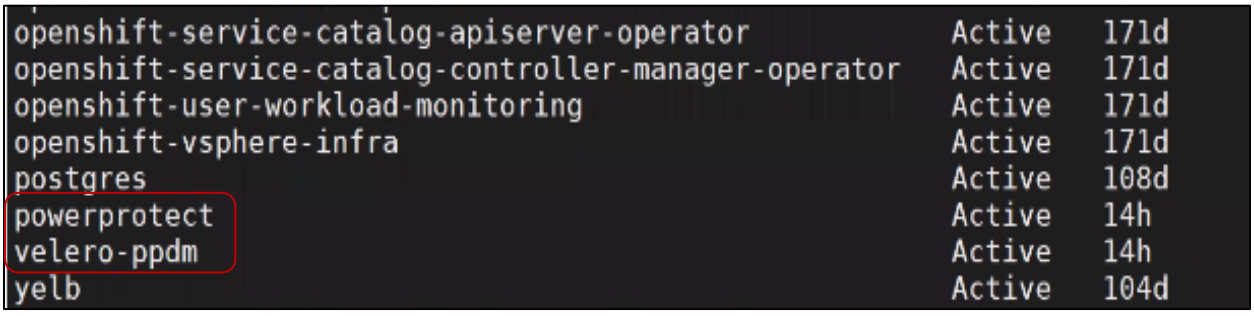


5. Click **Assets** > **Kubernetes** to view the namespaces.



Note:

1. The **powerprotect** and **velero-ppdm** namespaces are created automatically once the cluster is integrated to PowerProtect Data Manager.



2. During the discovery process, when OpenShift cluster is detected in PowerProtect Data Manager, OADP and Velero pod gets automatically installed in the velero-ppdm namespace. These pods will deploy the required plugins such as OpenShift, vSphere and DDR plugins.

```
[core@irv-12-181 ~]$ oc get pods -n velero-ppdm
NAME                                READY   STATUS    RESTARTS   AGE
backup-driver-6fdcb48666-p6x66      1/1     Running   0          5d15h
oadp-operator-74d9f55bbc-k92ct      1/1     Running   0          5d15h
velero-5d74cd5bf8-txds5             1/1     Running   0          5d15h
```

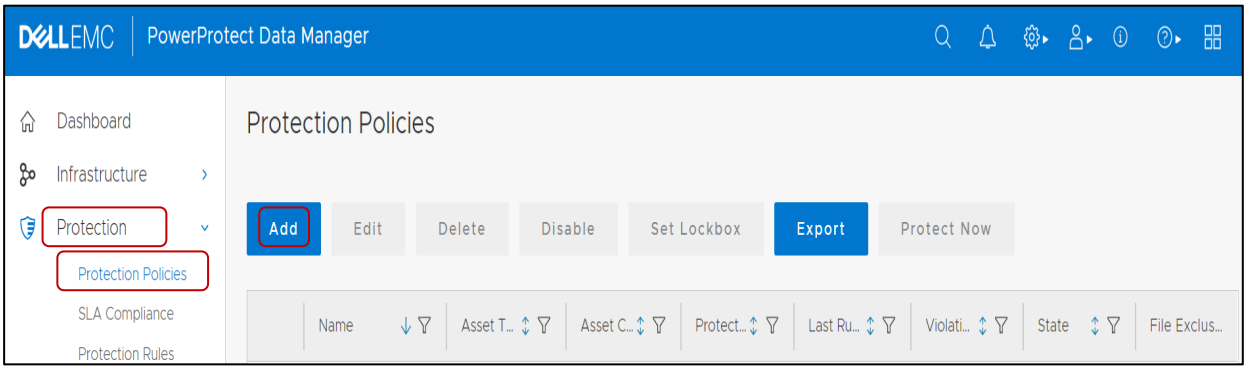
4.2 Backup Configuration

4.2.1 Create a protection policy

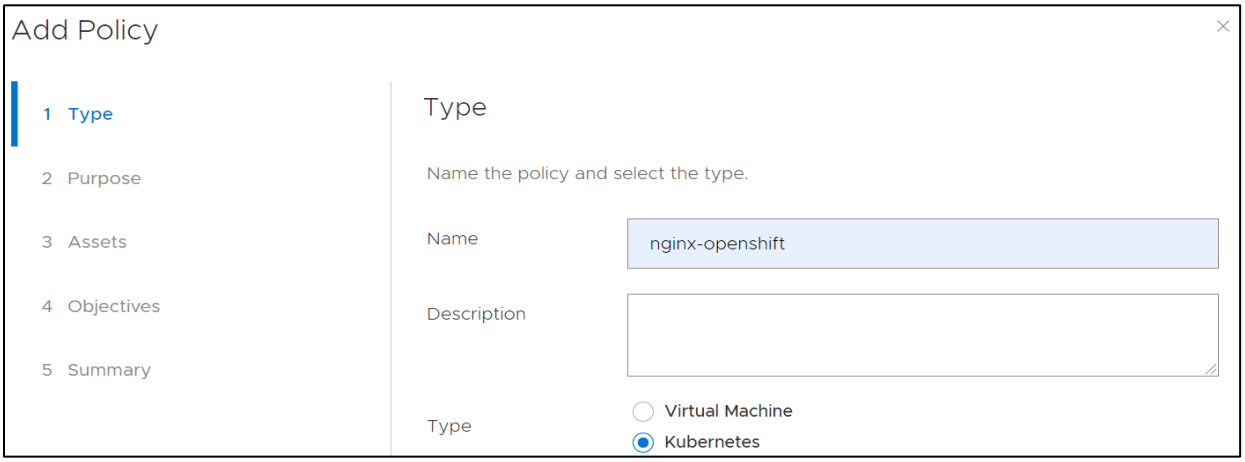
Backup can be scheduled as well as manually initiated. PowerProtect Data Manager UI enables users to create a protection policy to select the namespace that is to be protected and run the backup.

Steps to be followed are:

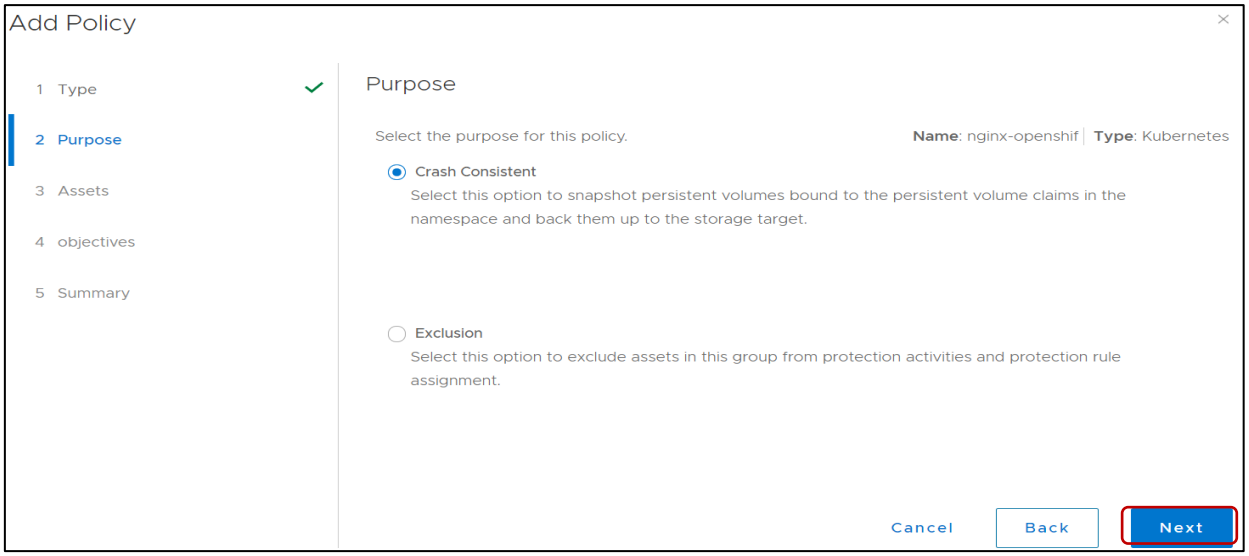
1. Select **Protection** > **Protection Policies**.
2. In the Protection Policies window, click **Add**. The Add Policy wizard appears.



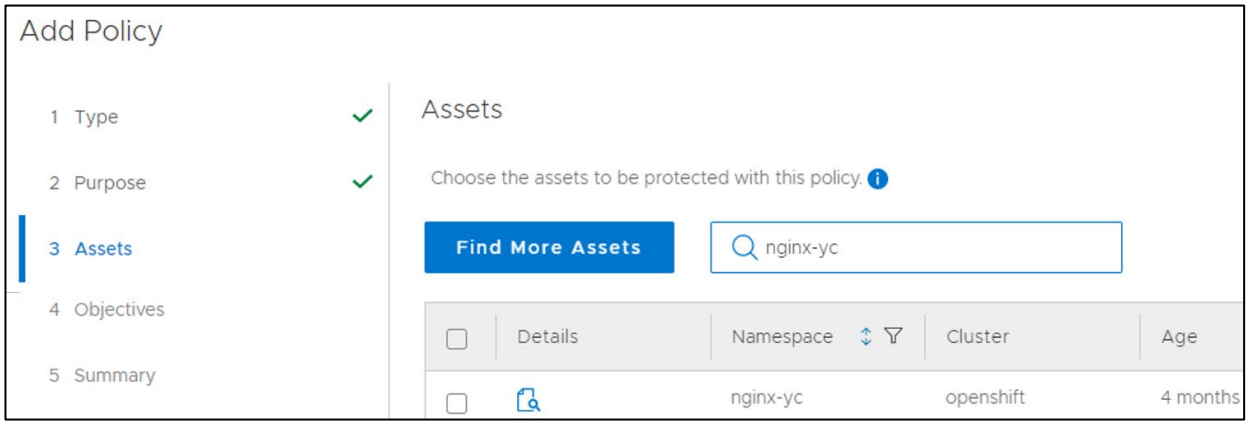
3. On the **Type** page, specify the following fields, and then click **Next.**
   - **Name**—a descriptive name for the protection policy.
   - **Description**—a description for the policy.
   - **Type**—For the policy type, select Kubernetes.



4. On the **Purpose** page, select from the following options to indicate the purpose of the new protection policy group, and then click **Next.**
   - **Crash Consistent**—Select this type for point-in-time backup of namespaces.
   - **Exclusion**—Select this type if there are assets within the protection policy that are to be excluded from data protection operations.

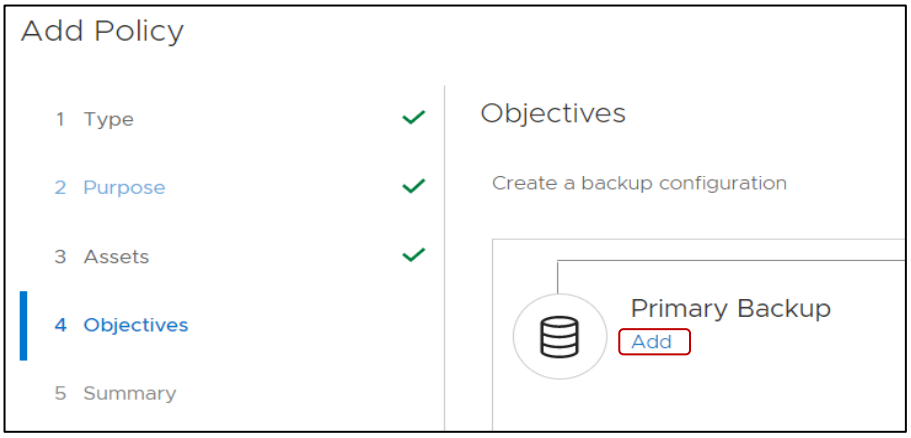


5. In the **Assets** page, select one or more unprotected namespaces that are to be backed up as part of this protection policy.

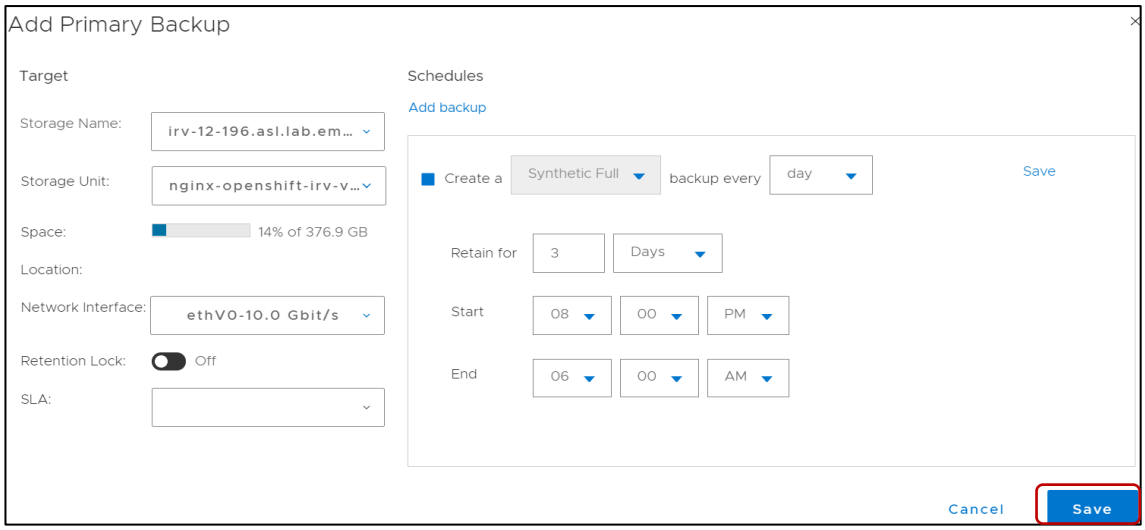


6. (Optional) For the selected namespaces, click the link in the PVCs Excluded column, if available, to clear any PVCs that are required to be excluded from the backup. By default, all PVCs are selected for inclusion.

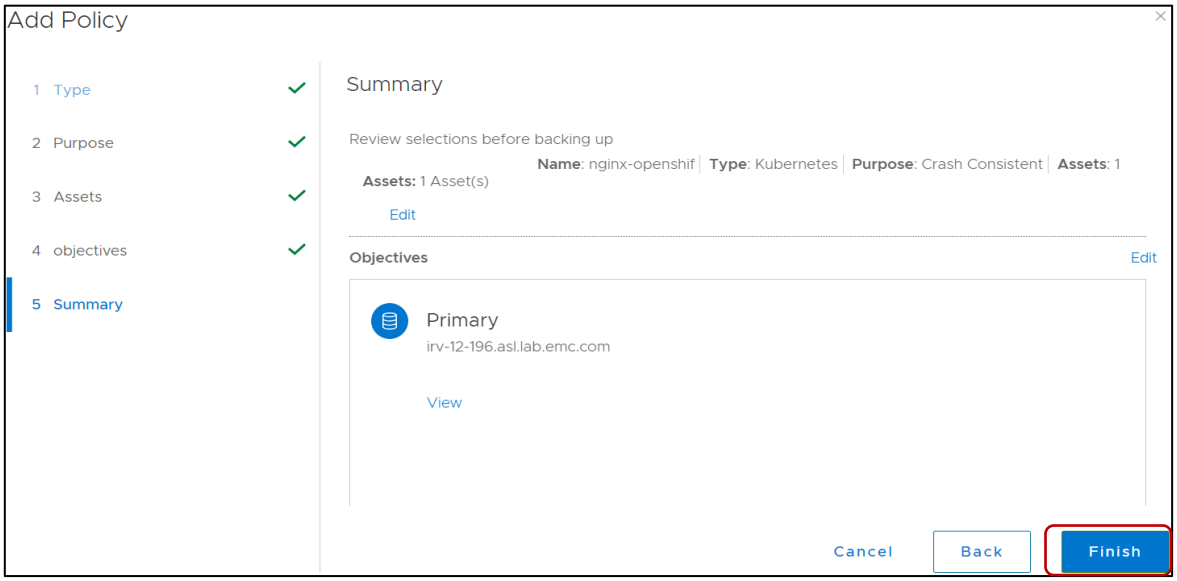
7. Click **Next**. In the **Objective** page, primary backup storage and schedule can be added.



8. Fill in the required details under **Add Primary Backup** section and click **Save**.
   - Under the **Target** section add the backup storage details.
   - Under the **Schedules** section add the backup schedule.
     - **Backup Every**: Specify how often to create a synthetic full backup.
     - **Retain for**: Specify the retention period for the backup.
     - **Start**: Specify the time of day to start initiating backups.
     - **End**: Specify the time of day to stop initiating backups.



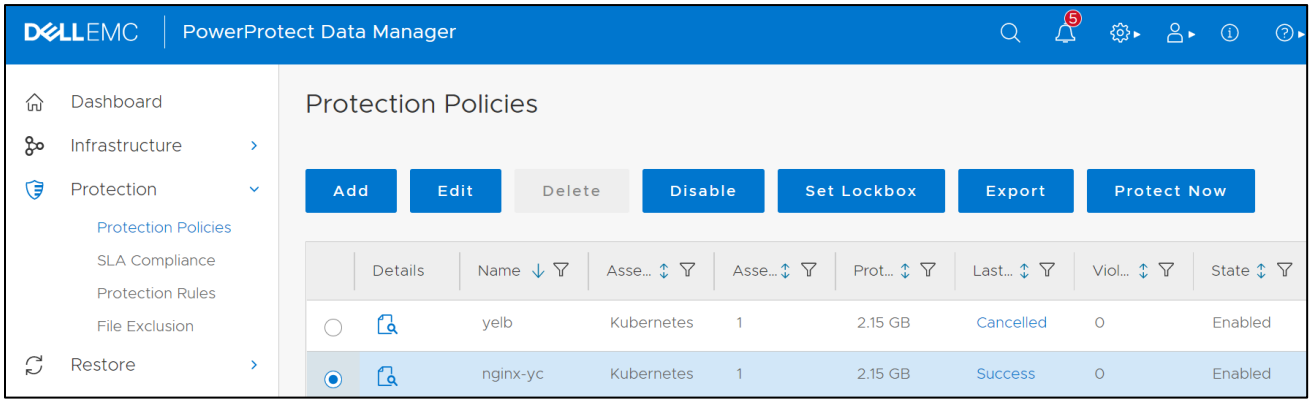
9. Verify the provided information is correct under **Summary** section. If it is correct, click **Finish**.



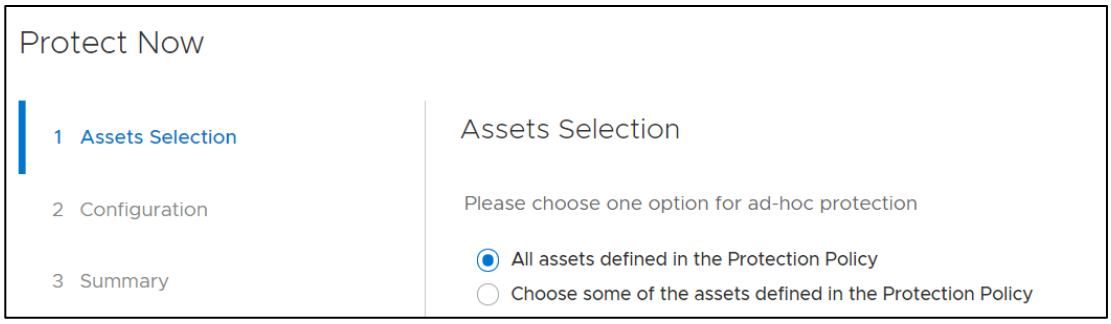
The protection policy is created successfully and triggers the backup at the scheduled time.

4.2.2 Configure the protection policy

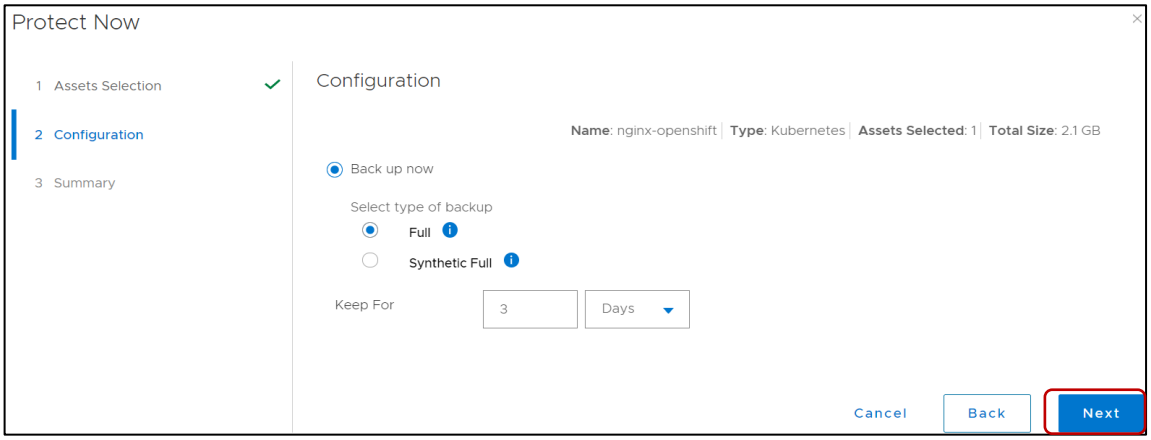
When the protection policy is created successfully, there are options to modify the existing policy that is Edit, Disable, Export and Protect Now.



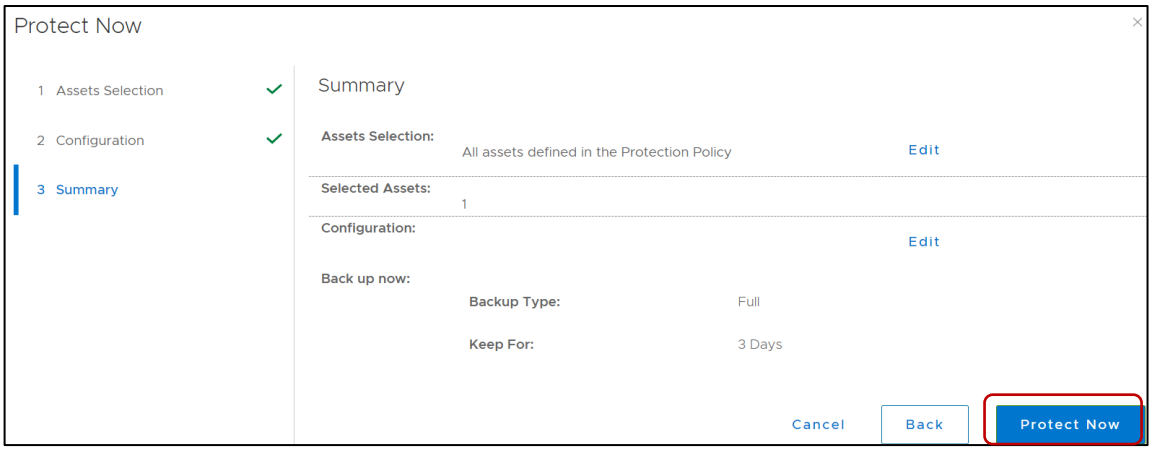
1. **Edit**: To edit the information or to change the schedule.
2. **Disable**: Backup Schedule is disabled with this option so backup would not occur.
3. **Export**: Downloadable file which contains the information about the asset protection.
4. **Protect Now**: This Option allows you to take a backup manually on an ad-hoc basis.
   - **Asset Selection**: It has two additional options to select the assets:
     - All assets defined in the protection policy
     - Choose some of the assets defined in the policy: This option allows you to select namespaces within the cluster



- **Configuration**:
  - This allows you to select type of backup. There are two options:
    a. **Full**: Backs up the namespace metadata and persistent volumes and creates a new full backup.
    b. **Synthetic full**: Backs up namespace metadata, changed blocks for persistent volumes on VMware first class disks, all data for other types of persistent volumes and creates a new full backup.
  - **Keep For**: Specify the retention period for the backup.
  - Click **Next**.

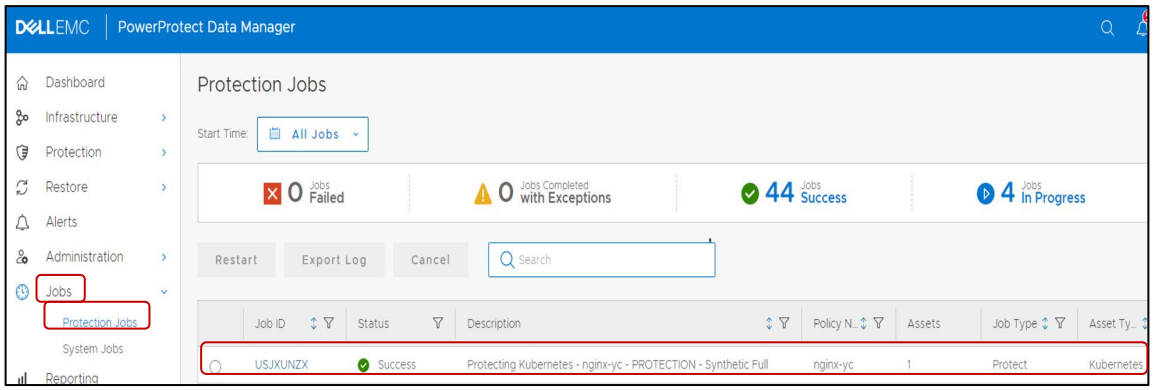


- **Summary**: Verify the information.



- Click **Protect now**.

- Monitor the backup job by navigating to **Protection Jobs** under **Jobs**.

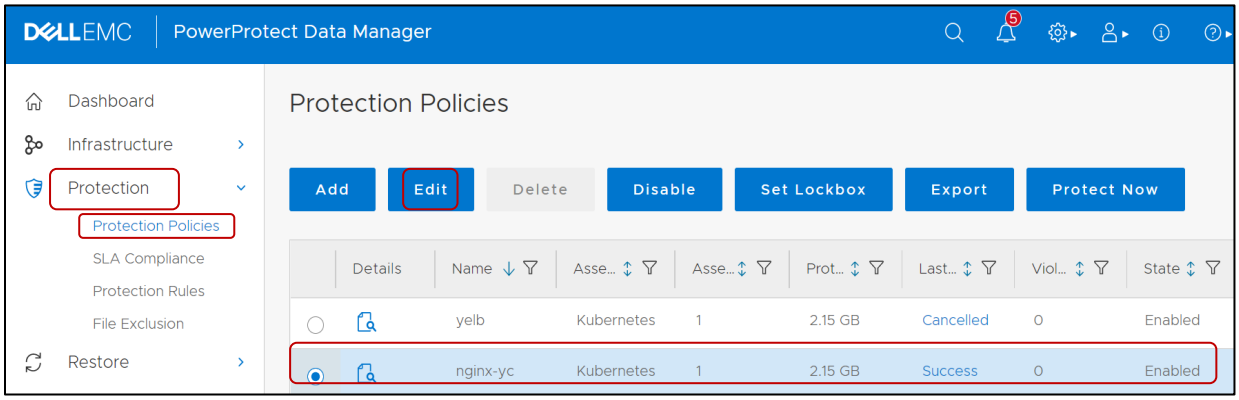


The backup job is completed successfully and details such as taskID, storage, and PVCs are available in backup job details.

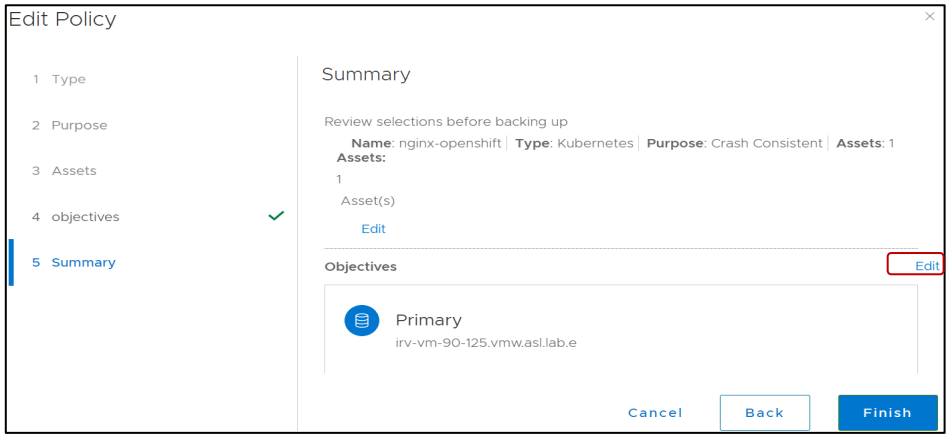
4.3 Replication Configuration

The replication is configured with an existing protection policy or a new policy can be created. Steps to configure replication on existing protection policy:

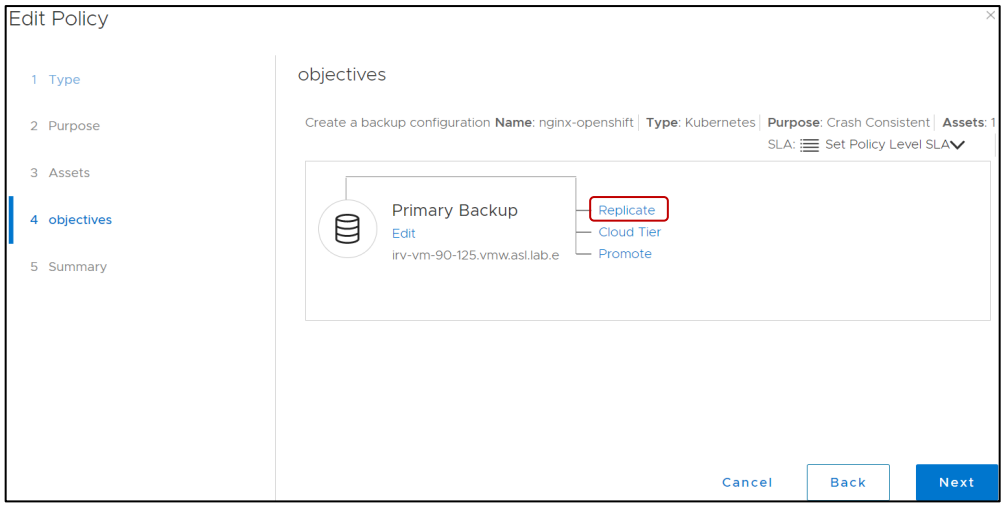
1. Login to **PowerProtect Data Manager UI** with administrator credentials.
2. On the left pane of the PowerProtect Data Manager UI, click **Protection** > **Protection Policies.**
3. Select the existing backup policy and click **Edit**.



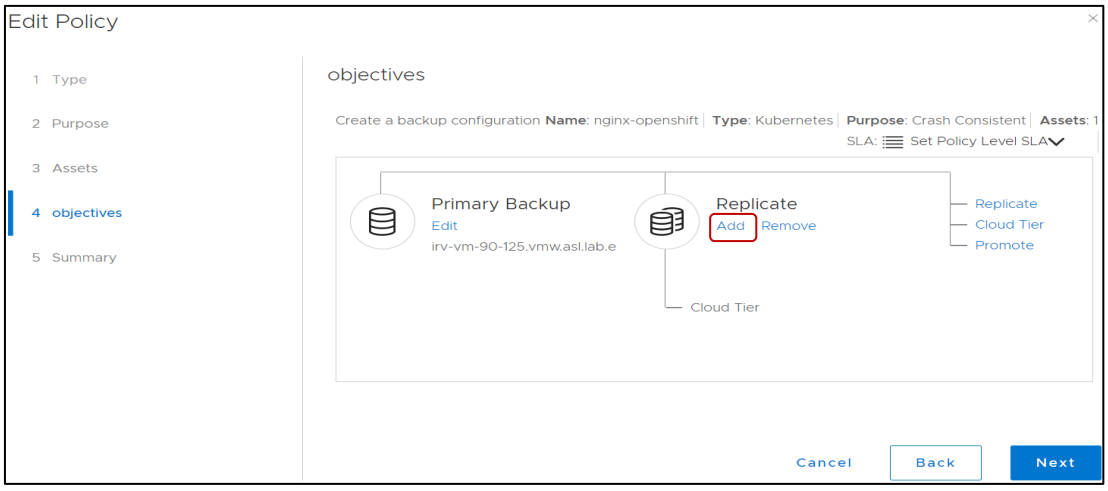
4. Click **Edit**.



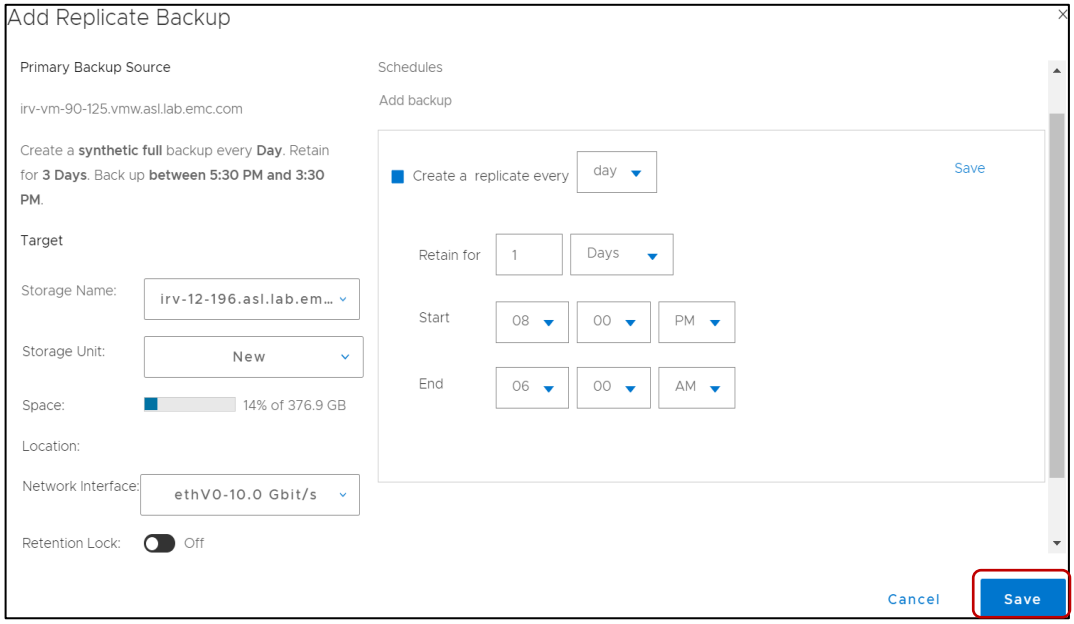
5. Click **Replicate.**



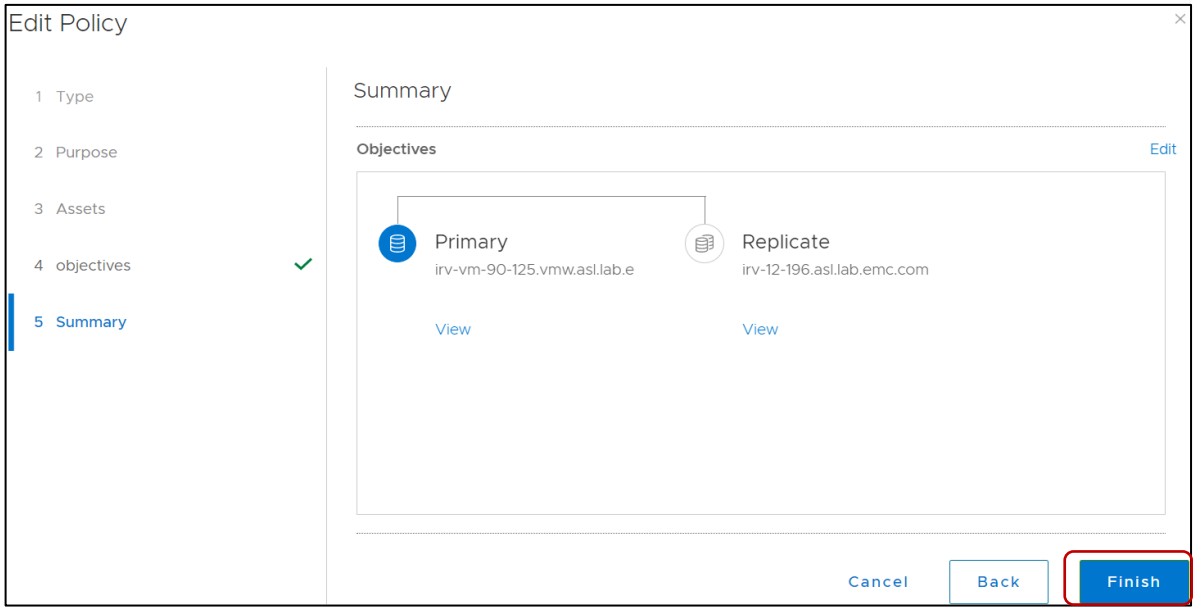
6. Click **Add** to add a replication schedule.



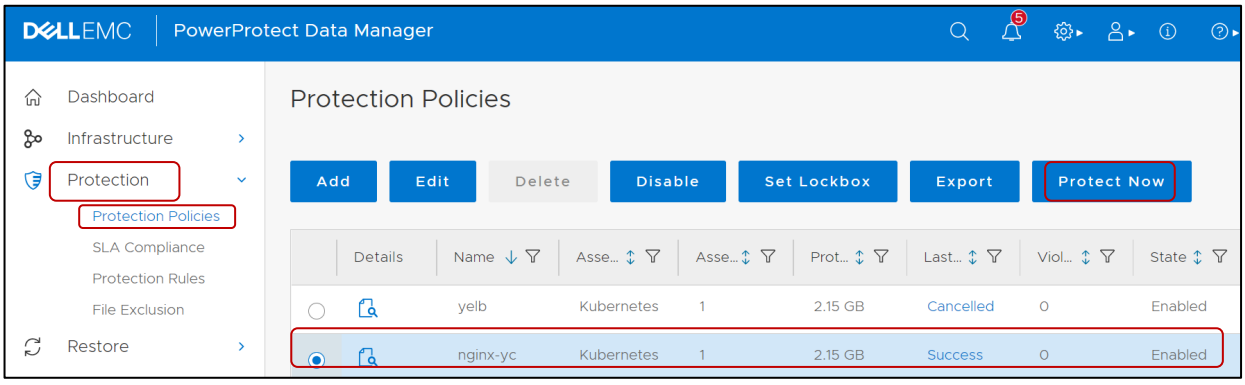
7. Fill in the required details under **Add Replicate Backup** section and click **Save**.
   - Under the **Target** section add the storage details.
   - Under the **Schedules** section add the replication schedule.
     - **Create a replica every**: Specify how often to create a replica.
     - **Retain for**: Specify the retention period for the replica.
     - **Start**: Specify the time of day to start initiating replication.
     - **End**: Specify the time of day to stop initiating replication.



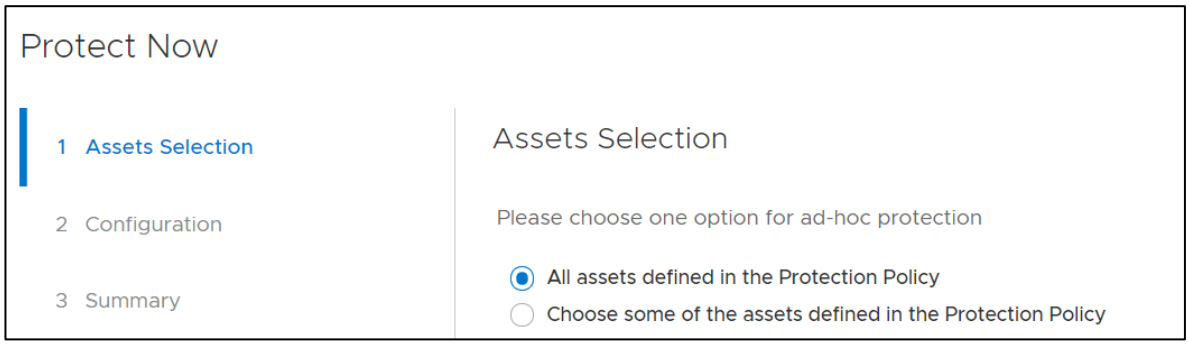
8. Review the summary and click **Finish**.



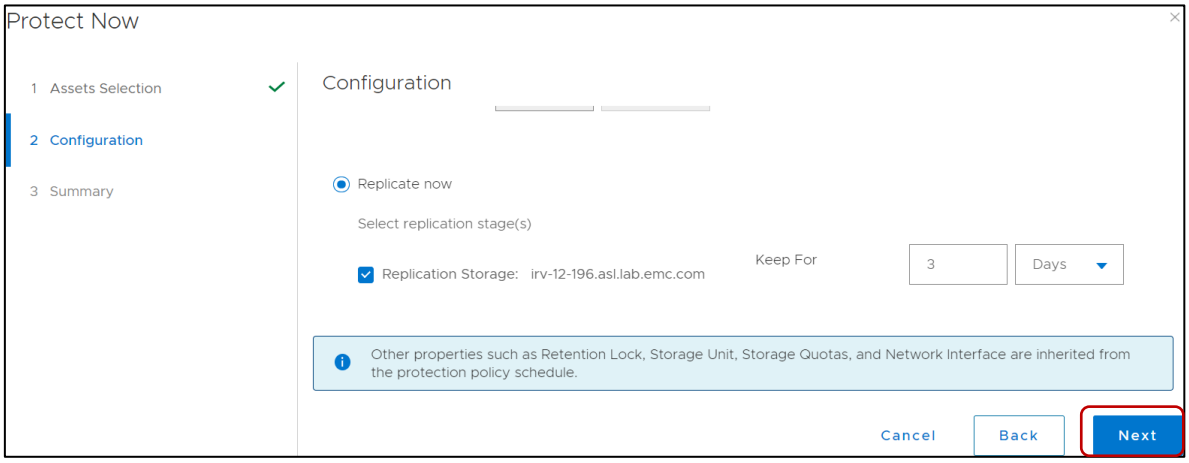
9. To run the replication now, select existing policies from **Protection** > **Protection Policies**.



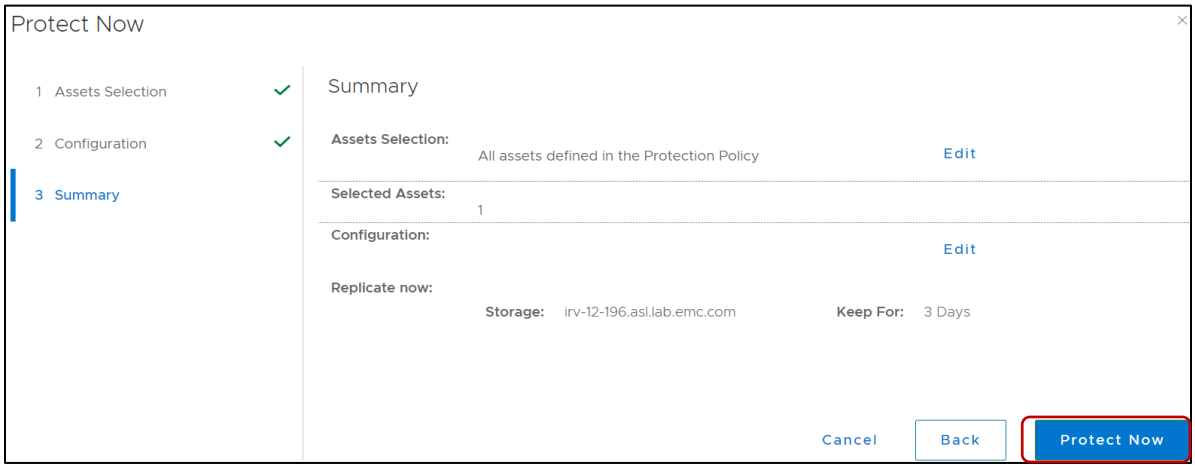
10. Click **Protect Now**.
    - **Asset Selection**: Choose one option for ad hoc protection.



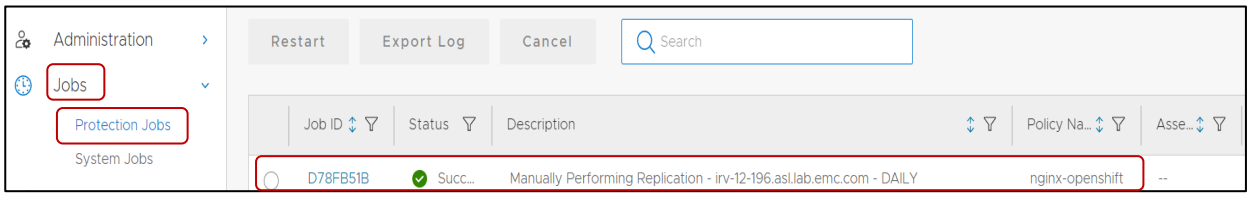
    - **Configuration**: Select **Replicate Now** option and check box to select replication storage.



    - **Summary**: Click **Protect Now** to start replication.



11. Verify the replication job is successfully completed under **Jobs** > **Protection Jobs**, click the details button for detailed results.



4.4 Restore Configuration

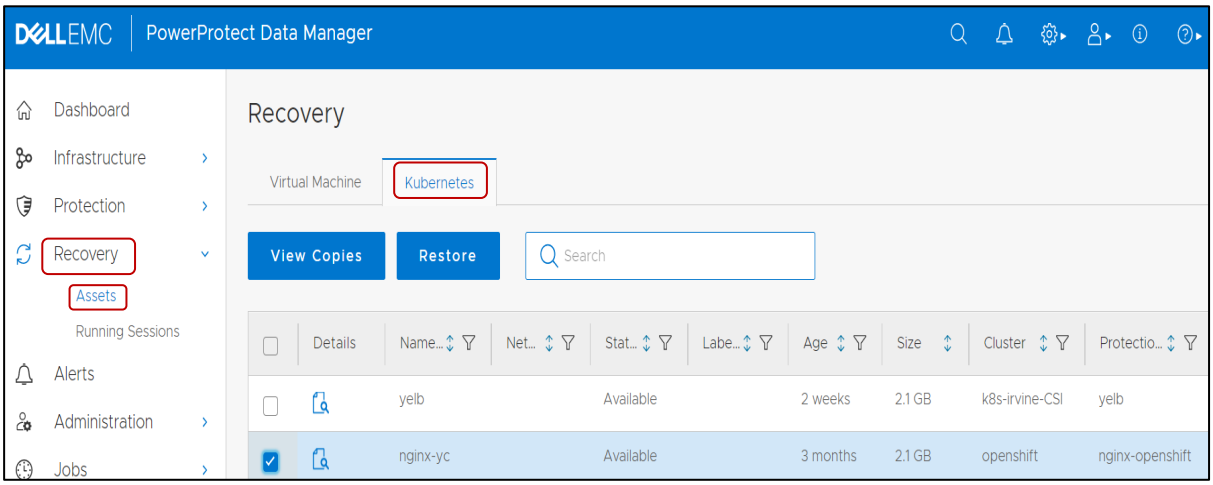
The recovery of assets is a manual process. With PowerProtect Data Manager, users can recover the Kubernetes namespace to the same cluster as well as alternate cluster.

1. Log in to PowerProtect Data Manager UI with admin credentials.

2. On the left pane of the PowerProtect Data Manager UI, click **Recovery.**

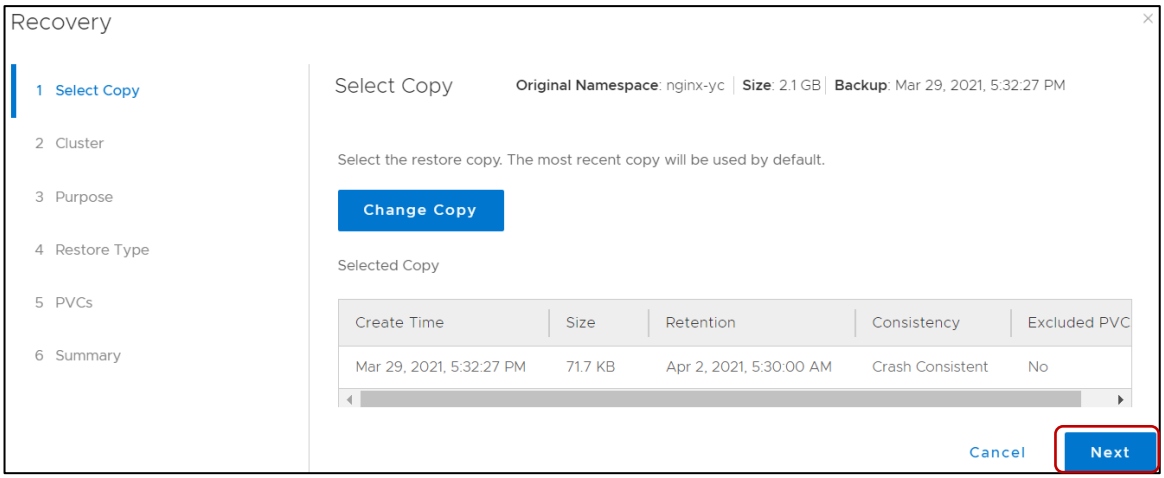
3. Click **Assets**.

4. Click **Kubernetes** on top and select the namespaces to be restored.

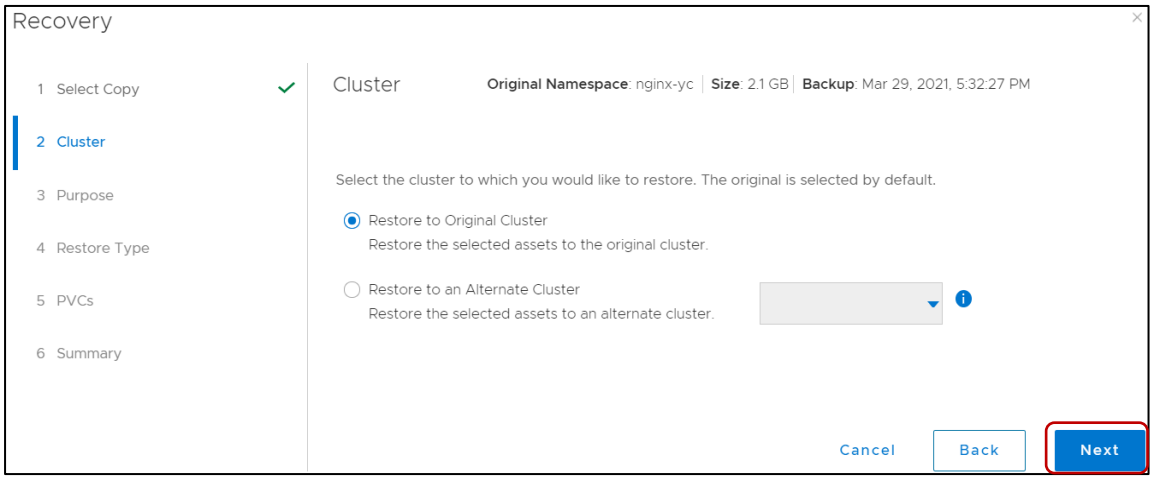


5. Click **Restore.**

- **Select Copy**:
  - Select the restore copy. The most recent copy will be used as default. To change from default copy, click **Change Copy**.
  - Click **Next**.

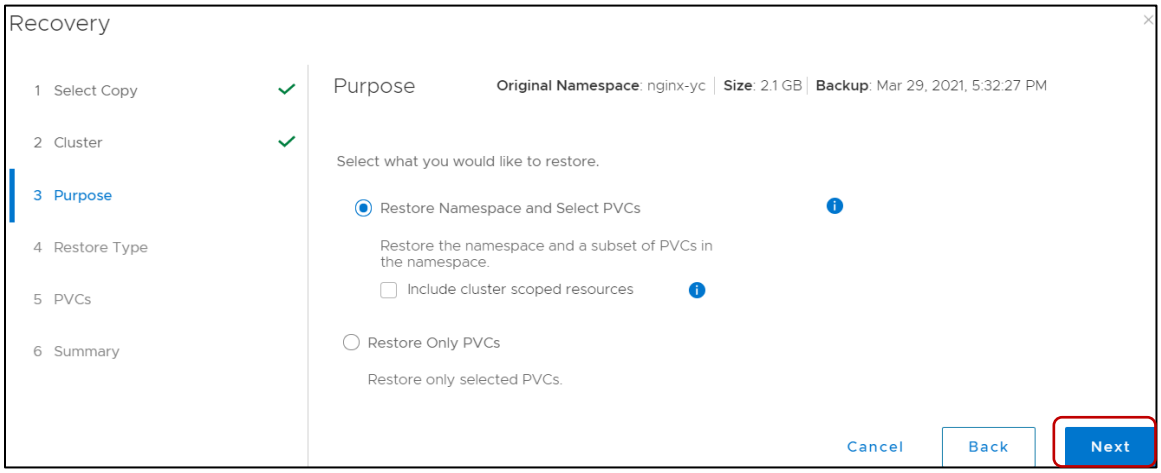


- **Cluster**: This provides the option to select the cluster on which assets to be restored.
  - **Restore to Original Cluster**: The Assets are restored to the source cluster from which the backup is taken.
  - **Restore to Alternate Cluster**: The Assets are restored on the alternate cluster. To use this option, the desired alternate target cluster must have previously been added as an asset source to PowerProtect Data Manager (as described in section 4.1).

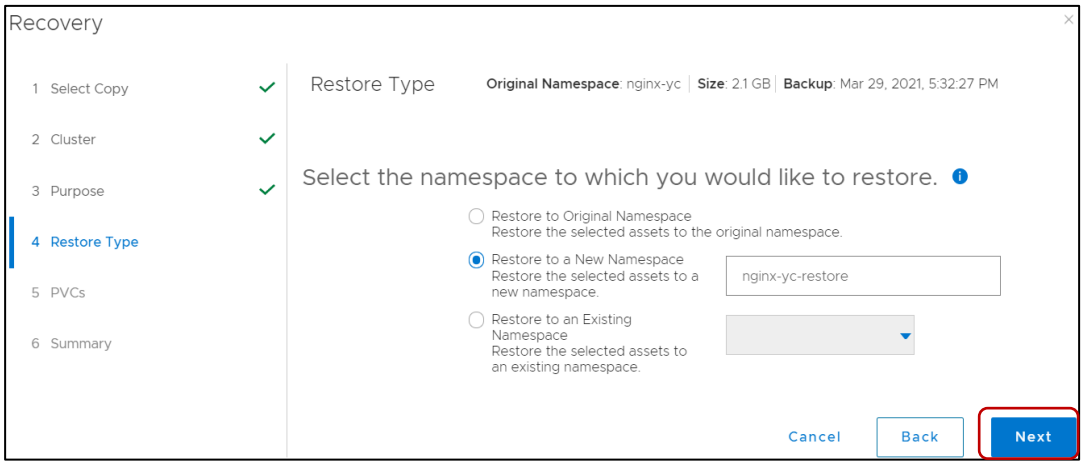


Note: Migration and restore to another cluster is not supported when using an integrated image registry, it is supported when using an external repository such as Docker Hub. For restore to another cluster, when using an integrated image repository, restore the individual PVCs instead of the entire namespace.

- **Purpose**: Select the option what is to be restored
  - **Restore Namespace and Select PVCs**: This option restores the namespace and a subset of PVCs in the namespace.
  - **Restore PVCs only**: This Option will restore only PVCs.

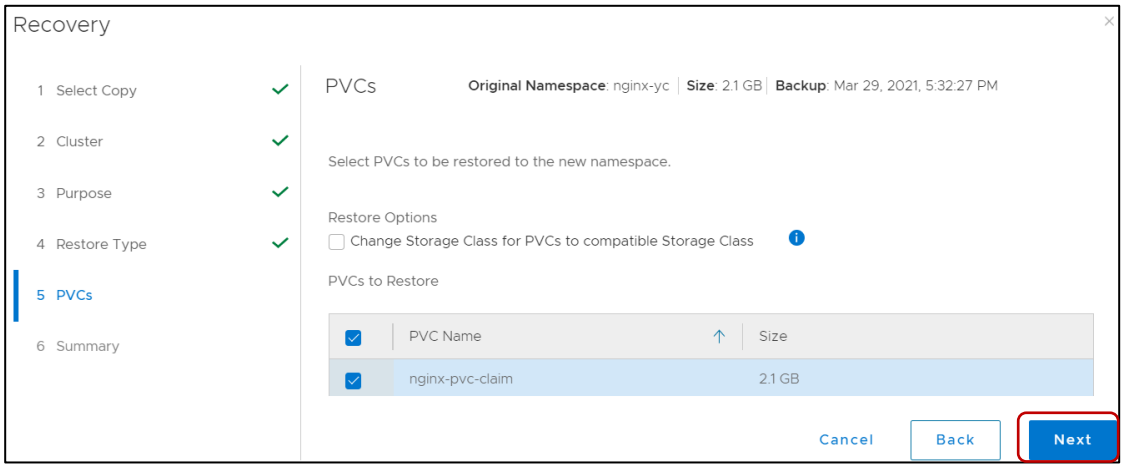


- **Restore Type**: Restore type has different options depending on the purpose of the restore.
  - If purpose is to Restore namespaces and PVCs, then options are
    a. **Restore to Original Namespace**,
    b. **Restore to New Namespace** and
    c. **Restore to an Existing Namespace**

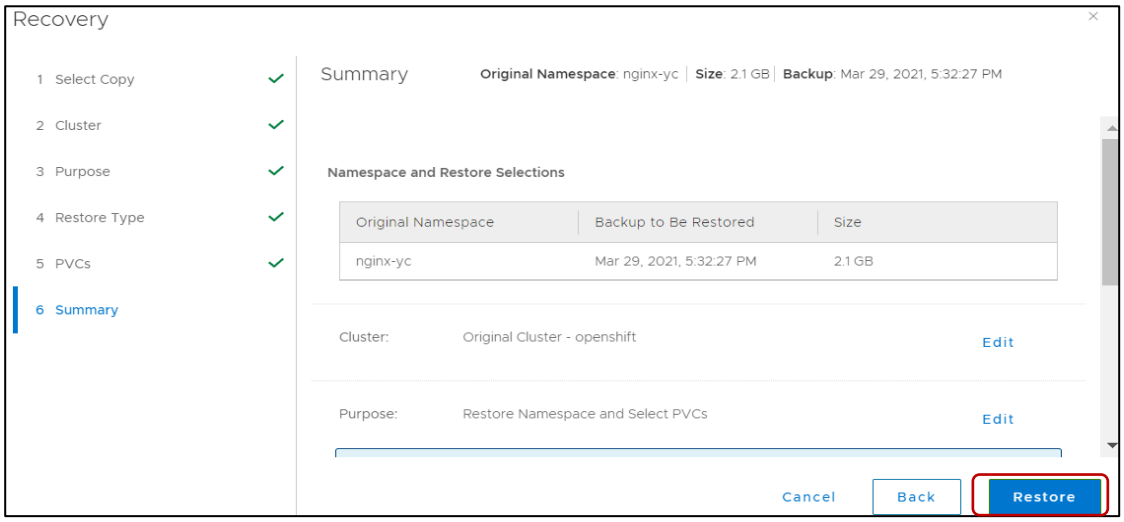


  - And if the purpose is to restore PVCs only, then the options are:
    a. Restore to Original Namespace and
    b. Restore to an Existing Namespace

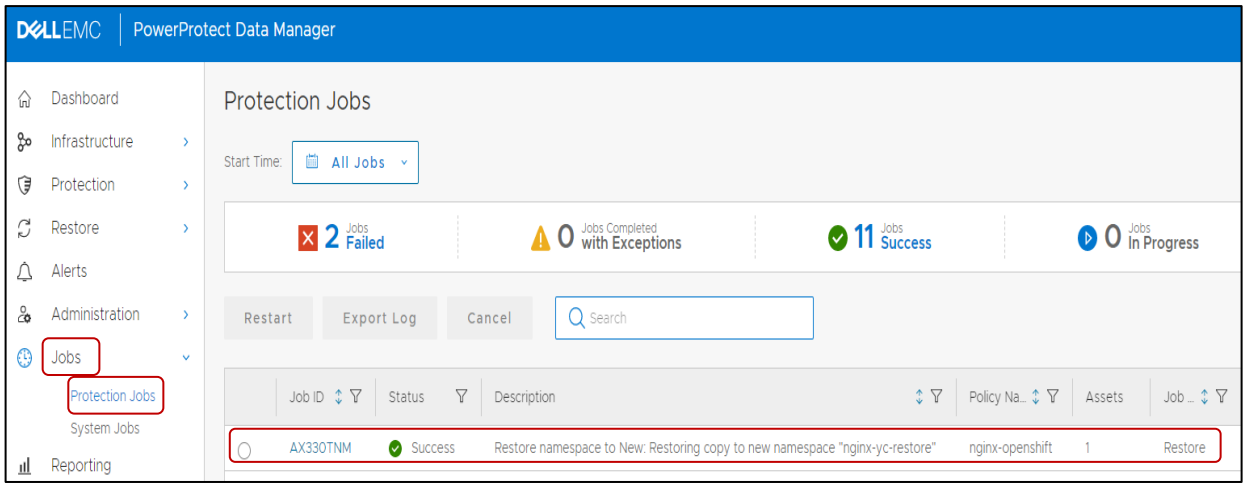
- **PVCs:** Select PVCs to be restored to the namespace.



- **Summary:** Verify the information and click **Restore**.



5. Progress of the recovery process can be seen under **Jobs** > **Protection Jobs**.



The components of the namespace that are restored can be seen by expanding the namespace which is selected for the restoration process in the OpenShift command line interface (CLI).

```
[core@irv-13-2 ~]$ oc get all,pvc -n nginx-yc-restore
NAME                READY   STATUS      RESTARTS   AGE
pod/nginx-1-deploy   0/1     Completed   0          45s
pod/nginx-1-dq2cr    1/1     Running     0          52s

NAME                            DESIRED   CURRENT   READY   AGE
replicationcontroller/nginx-1   1         1         1       45s

NAME            TYPE        CLUSTER-IP      EXTERNAL-IP   PORT(S)   AGE
service/nginx   ClusterIP   172.30.61.137   <none>        80/TCP    3m39s

NAME                                        REVISION   DESIRED   CURRENT   TRIGGERED BY
deploymentconfig.apps.openshift.io/nginx    1          1         1         config,image(nginx:latest)

NAME                                  IMAGE REPOSITORY
              TAGS     UPDATED
imagestream.image.openshift.io/nginx   default-route-openshift-image-registry.apps.ocpk8s.ocpdellemc.com/nginx-yc-r
estore/nginx   latest   46 seconds ago

NAME                                    STATUS   VOLUME                                     CAPACITY   ACCESS MODES
   STORAGECLASS   AGE
persistentvolumeclaim/nginx-pvc-claim   Bound    pvc-ee219f17-ccc3-45a1-9039-233a0351182a   2Gi        RWO
   mongodb-sc     3m38s
[core@irv-13-2 ~]$
```

5 Conclusion

This paper detailed how to discover OpenShift clusters, create protection policies, and walked through the backup and restore workflows with Dell EMC PowerProtect Data Manager. In summary, PowerProtect Data Manager provides the capability to protect OpenShift Kubernetes workloads, by ensuring that data is easy to back up and restore, always available, consistent, and durable in a Kubernetes workload or disaster recovery situation.

A Technical support and resources

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

Storage technical documents and videos provide expertise that helps to ensure customer success on Dell Technologies storage platforms.

A.1 Related resources

- OpenShift overview
- OpenShift for developers
- Understanding build configurations
- Understanding deployment configurations
- Understanding containers, images and image streams
- PowerProtect Data Manager Administrator and User guide
- Dell EMC PowerProtect Data Manager protecting VMware Tanzu Kubernetes Clusters
- Dell EMC PowerProtect Data Manager protecting Kubernetes Workloads