CDP Private Cloud Data Services 1.5.0

CDP Private Cloud Data Services Data Recovery

Date published: 2020-12-16
Date modified: 2023-01-25

https://docs.cloudera.com/

Legal Notice

Contents

Data Recovery Service overview

The Data Recovery Service (DRS) is a microservice in CDP Private Cloud that allows you to back up and restore Kubernetes namespaces and resources on both Embedded Container Service (ECS) and OpenShift Container Platform (OCP). Cloudera recommends that you create a backup of your Kubernetes namespace using DRS before a maintenance activity or in general, as a best practice.

Note: In the CDP Private Cloud 1.5.0 release, you can backup and restore data of the Control Plane and Cloudera Data Warehouse (CDW) data service using CDP CLI.

Using DRS, you can take a backup of:

- the Kubernetes resources associated with the cdp namespace and the embedded vault namespaces of the Control Plane. The resources include deployment-related information, stateful sets, secrets, and configmaps.
- the data used by the stateful pods, such as the data in the embedded database and Kubernetes persistent volume claim.

By default, the data recovery service is located in the [***CDP_INSTALLATION_NAMESPACE***]-drs namespace. For example, if the CDP Private Cloud Data Services installation is located in the cdp namespace, the data recovery service namespace is automatically named cdp-drs. If you have multiple CDP Private Cloud Data Services installations (as in OCP), the data recovery service is named accordingly.

You must have the “PowerUser” role to run the DRS commands using the CDP CLI.

The Data Recovery Service requires CSI snapshots to back up and restore Kubernetes namespaces and resources. The CSI snapshots are enabled on ECS by default. You might require an additional license to enable CSI snapshots in Red Hat ODF storage on OCP.

Sequence of backup and restore events when using DRS

Learn about the high-level steps that are performed when you create and restore a backup using Data Recovery Service (DRS).

Backup event

When you create a backup, DRS:

1. initiates the backup event of the Control Plane,
2. assigns an ID called “backupCrn” to the backup event,

   You can specify the backupCrn in the `describe-backup` CDP CLI command to track the progress of the backup event and to identify whether the event completed successfully. You can also use the `get-logs` CDP CLI command to retrieve detailed information about the event.
3. archives the information to a ZIP file,
4. saves the ZIP file on the same cluster, and
5. takes persistent volume claim snapshots in the OpenShift Container Platform (OCP) cluster and persistent volume claims clones in the Embedded Container Service (ECS) cluster.

The backup event does not have any downtime impact and you can backup the Control Plane while it is running.

Restore event

When you start the restore event, DRS:

1. initiates the restore event based on the specified backupCrn.

2. assigns an ID called “restoreCrn” to the restore event,

   You an specify the restoreCrn in the `describe-restore` CDP CLI command to track the progress of the restore event and to identify whether the event completed successfully. You can also use the `get-logs` CDP CLI commnd to retrieve detailed information about the event.
3. deletes the existing resources and data,

   During this stage of the restore event, the ECS restore vault is sealed and the POD is down which might appear as a failure in the control plane environment. After the restore event is complete, the vault and POD are auto-recovered and restored. Depending on the number of resources and data, this step might take a maximum of 10 minutes to complete.
4. restores the resources and data from the specified backupCrn.

   The restore event has a downtime impact because the pods and data are recreated.

   Important: Ensure that you do not delete the *[***CDP_INSTALLATION_NAMESPACE***]*-drs namespace while the restore event is in progress.

Consider the following points before you initiate the restore event:

- When you initiate the `restore-backup` event, the CDP User Management System (UMS) is up and running, therefore, the restore event initiates without any issues. During the restore event, the UMS goes down and comes up eventually. However, if the UMS is corrupted, contact Cloudera Support for further assistance.
- When the restore event crosses the time set in the `POD_CREATION_TIMEOUT` environment property of the cdp-release-thunderhead-drsprovider deployment in the drs namespace, a timeout error appears. By default, the property is set to 900 seconds. In this scenario, you must manually verify whether the pods are up or not.

Related Information

CDP CLI drscp

CLI reference for using DRS on Control Plane

You can use the Data Recovery Service (DRS) commands using the CDP CLI to backup and restore resources and data in the Control Plane of CDP Private Cloud Data Services.

The following table provides the CDP CLI options to backup and restore the Control Plane:

CDP CLI options	Description
create-backup	Creates a backup and archives it as a ZIP file on the same cluster. The item-name is optional for Control Plane data recovery service. You can provide a unique backup name, so that you can identify the backup easily during restore. When you run this command, the service initiates the backup process and returns the assigned ID or `backupCrn` for the backup.
delete-backup	Deletes the specified backup (`backupCrn`) permanently. **Important:** Alerts are not generated when you run this command.
describe-backup	Shows the progress of the current status of the specified `backupCrn` (backup event).
describe-restore	Shows the progress of the current status of the specified `restoreCrn` (restore event).

CDP CLI options	Description
get-logs	Returns logs about the specified backup, restore, or delete job and automatically creates a diagnostic bundle. You can download the bundle to your machine to analyze an issue or share it with Cloudera Support for further troubleshooting.
list-backup-entities	Lists the entities that you can backup, which includes the Control Plane namespace and its corresponding vault namespace (if embedded).
list-backups	Lists the successful backup jobs of `backupCrn`. You can filter the backup jobs using the NOT_STARTED; IN_PROGRESS; COMPLETED; PARTIALLY_FAILED; and FAILED job states.
list-restores	Lists the past restore events.
restore-backup	Restores the backup of the specified `backupCrn`. During the restore event, the existing Kubernetes resources and data are deleted and then recreated using the information in the backup. When you run the command, the service initiates the restore event and returns a `restoreCrn` value.

Related Information

CDP CLI drscp

Using DRS with CDW

You can back up and restore Kubernetes namespaces behind Cloudera Data Warehouse (CDW) entities (for example, Database Catalogs, Virtual Warehouses) on demand using the Data Recovery Service (DRS). CDW leverages DRS and provides CDP CLI endpoints which you can use to create and restore backups for CDW namespaces to back up CDW metadata and configurations such as Kubernetes objects, persistent volumes, autoscaling configuration, and so on.

The following limitations apply for CDW data service if you are on Embedded Container Service (ECS) or using an embedded database on Red Hat OpenShift Container Platform:

- The embedded database that CDW uses is part of the Control Plane. You cannot back up only CDW-related entities from the embedded database using the `dw create-backup` command. You must take a backup of the Control Plane service.
- You must restore the entire Control Plane configurations to restore configurations stored in the CDW database. This recreates the Control Plane namespace.

List of data recovery sub-commands for CDW

The following table lists the commands and CLI endpoints for backing up and restoring Kubernetes namespaces behind CDW entities:

DRS sub-commands for CDW	Description
create-backup	Creates an on-demand backup for the Data Warehouse including Kubernetes objects, persistent volumes, and so on. Backup requests are processed asynchronously and instantaneously.
delete-backup	Deletes an existing Data Warehouse backup. The call returns immediately. It returns a delete CRN, which is the deletion process identifier.
describe-backup	Returns the description of an existing Data Warehouse backup.
restore-backup	Restores the state of the Data Warehouse from an existing backup. It returns a restore CRN, which is the identifier of the restoration process.

DRS sub-commands for CDW	Description
describe-restore	Returns the description of the Data Warehouse restore operation.
list-backup-entities	Lists potential backup entities associated with the Data Warehouse.
list-backups	Lists backups associated with the Data Warehouse.
list-restores	Lists restores associated with the Data Warehouse.
get-logs	Returns the job logs corresponding to the specified CRN.

Related Information

CDP CLI reference of DRS commands for CDW