

1.5.5

Data Recovery Service

Date published: 2026-09-24

Date modified:

CLOUDERA

Legal Notice

© Cloudera Inc. 2026. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Data Recovery Service (DRS) overview.....	4
Understanding the Data Recovery and Restore Lifecycle.....	4
Data Recovery Service (DRS) prerequisites.....	5
About the Backup service.....	7
About the Restore service.....	10
Creating a backup.....	12
Triggering a Restore.....	15
Key Data Recovery Scenarios.....	16

Data Recovery Service (DRS) overview

The new DRS is a holistic backup and recovery solution encompassing the core Cloudera Control Plane and integrated Data Services (such as Cloudera Data Engineering, Cloudera Data Warehouse, and Cloudera Data Catalog).

The Cloudera Control Plane and supported Data Services are available in a unified experience.



Important: This new implementation is only available starting Cloudera Data Services on premises Service Pack 4 and higher versions. Applicable only for the cluster users of Cloudera Embedded Container Service. OpenShift Container Platform users should refer to the old DRS feature which is available in Cloudera Data Services on premises SP3 CHF3 and below releases.

Some of the new capabilities that DRS now offers includes:

- Unified framework consolidating all Data Services into a single interface
- Flexible policy management includes granular control over participating entities and scheduled and ad-hoc backups
- Ability to restore the backup to a secondary Cloudera Embedded Container Service cluster
- Support for incremental backups after an initial, full backup
- Configurable backup repositories (S3 compatible storage)
- Guided restore workflow with the ability to have configuration overrides

Core Architecture Overview and System Roles

The key operational roles and architecture components includes the following:

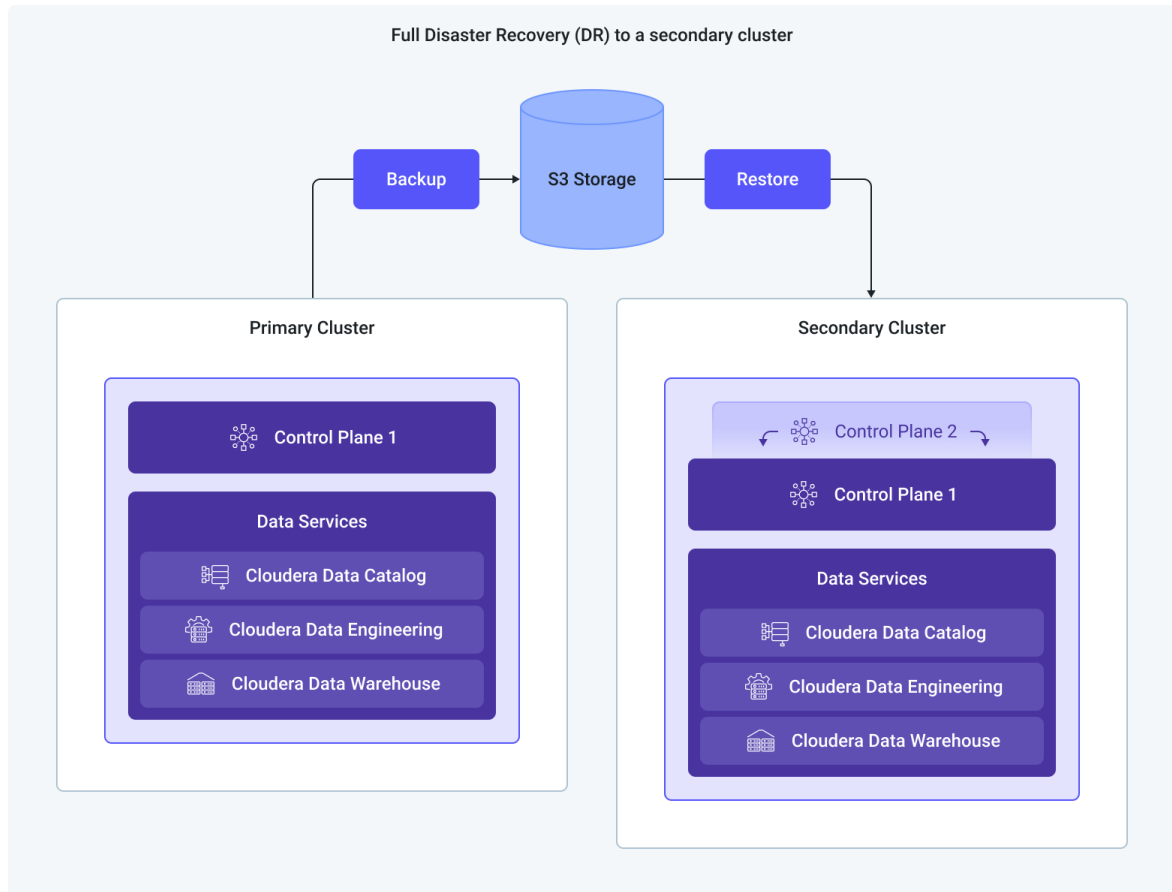
- Primary Cluster: Initiates backups and hosts active Cloudera Control Plane and Data Services (Cloudera Data Engineering, Cloudera Data Warehouse)
- Secondary Cluster: Clean target cluster receiving the restore payload. Must maintain Version Parity and adequate compute/storage capacity
- Central S3 Storage: Externalized S3-compatible storage containing both backup payloads and backup metadata manifests acting as the definitive source of truth
- Orchestration Layer: Cloudera wrapper that abstracts underlying execution engine, handles pre/post hooks, and coordinates with Data Services
- Execution Engine: Core underlying tools executing Kubernetes API backups, CSI volume snapshots, and incremental data deduplication/pushes

The security and login credentials transitions during the data recovery process. All user accounts, group permissions, and identity provider (IdP) configurations are saved directly within the application's environment. After restoring this environment on the secondary cluster, the backup copies your exact security setup automatically. Users can log in just like before, with zero manual re-configuration needed.

Understanding the Data Recovery and Restore Lifecycle

Review the End to End Disaster Recovery and Restore Lifecycle and its components that make up for the data recovery and restore phases.

This flow details how the secondary target cluster discovers backups, cleans resources, restores data, and applies cluster-specific re-configurations.



Execution workflow

1. **Discovery and Selection:** The administrator reads available backup entities directly from externalized S3 metadata.
2. **Pre-Restore operations:** Local secondary cluster configurations (Control Plane URLs, FQDNs, and Ingress rules) are captured into local memory/secrets.
3. **Restoration engine:**
 - Target cluster namespaces defined in the backup are cleaned.
 - K8s manifests, secrets, and Persistent Volumes are pulled from S3 and restored as-is.
4. **Post-Restore and Cluster Re-hydration:**
 - **Cloudera Control Plane Post-Restore:** Updates restored manifests with secondary cluster FQDN/URLs. The restored Vault secret is used to unseal the Vault instance.
 - **Environment and Cloudera Manager Updates:** If the Cloudera Base on premises cluster details/credentials are changed, then the environments (or environment objects) are updated.
 - **Data Services Post-Restore:** Post restore refresh tasks from participating data services are then invoked (For example, certificate regeneration, internal URL updates).

Data Recovery Service (DRS) prerequisites

Before you proceed further, you must review and note the required prerequisites.

- The cluster configurations for both the primary and secondary clusters must be the same.

- You must link your S3 storage to the primary and secondary clusters. Navigate to Cloudera Management Console # Administration # Backup and Restore.



Important: The configured S3 location must be available and accessible at all times from both primary and secondary clusters.

On the Backup and Restore Settings page, click Add New Location to add the S3 storage details for the location name, S3 Endpoint URL, External Bucket Name, Region Code, Access Key, Secret Key, upload the CA Certificate, and you have the option to enable the default location to the new backup settings. Click Test Connection to verify the connectivity between both the primary and secondary clusters. And later save the changes.

Location Settings

* Location Name

drs-velero

* S3 Endpoint Url

https://ccycloud-1.drs-cm1.root.comops.site:9879

* External Bucket Name

velero-drs-test-nk

Region Code

us-west-1

Default: us-west-1

* Access Key

.....

* Secret Key

.....

CA Certificate ⓘ

☐ File Upload

☒ Direct Input

Enter text

Default Location

☐ Enable default

This new location will be set as default for new backup settings.

Test Connection

Save Location

Cancel

- Secondary cluster configurations must be the same as the primary cluster. For example:
 - Container UID
 - Network CIDR range
 - IP range or subnet
- If the Node tainting feature was enabled in the primary cluster, then the same/similar configurations must be enabled in the secondary cluster as well.
- Environment mapping consistency: When restoring data / contents from the primary cluster to secondary cluster, the environments are automatically rewired as part of the restore process, with an option to override the default

destination, if needed. You must map all environments to the new secondary base clusters **when base clusters change** (when environments in the secondary cluster are configured to use a different base cluster, possibly a secondary base cluster)). Inconsistent environment mapping could cause restore operations to fail.

- Storage requirements: Cloudera recommends that you retain at least 30–40% of each node's Longhorn capacity as a free storage. Backups temporarily provision a CSI snapshot and a full-size, replicated Persistent Volume Claim (PVC) per volume, which are released after the data is copied over to the external S3 location.
- Any settings or configurations changes performed outside of the Cloudera Management Console to enable SAML needs to be reviewed after a secondary restore. References of 'primary control plane' in such external configurations should be changed to 'secondary control plane' after secondary cluster restore is successful. Failure to perform these steps can cause auth/login issues.
- Users must revert to the old external DRS settings that was supported using Longhorn's [BackupTarget](#) feature. These settings can negatively impact the new DRS solution by consuming higher storage/resources/network and might cause backup process to take a long time to complete or fail.

About the Backup service

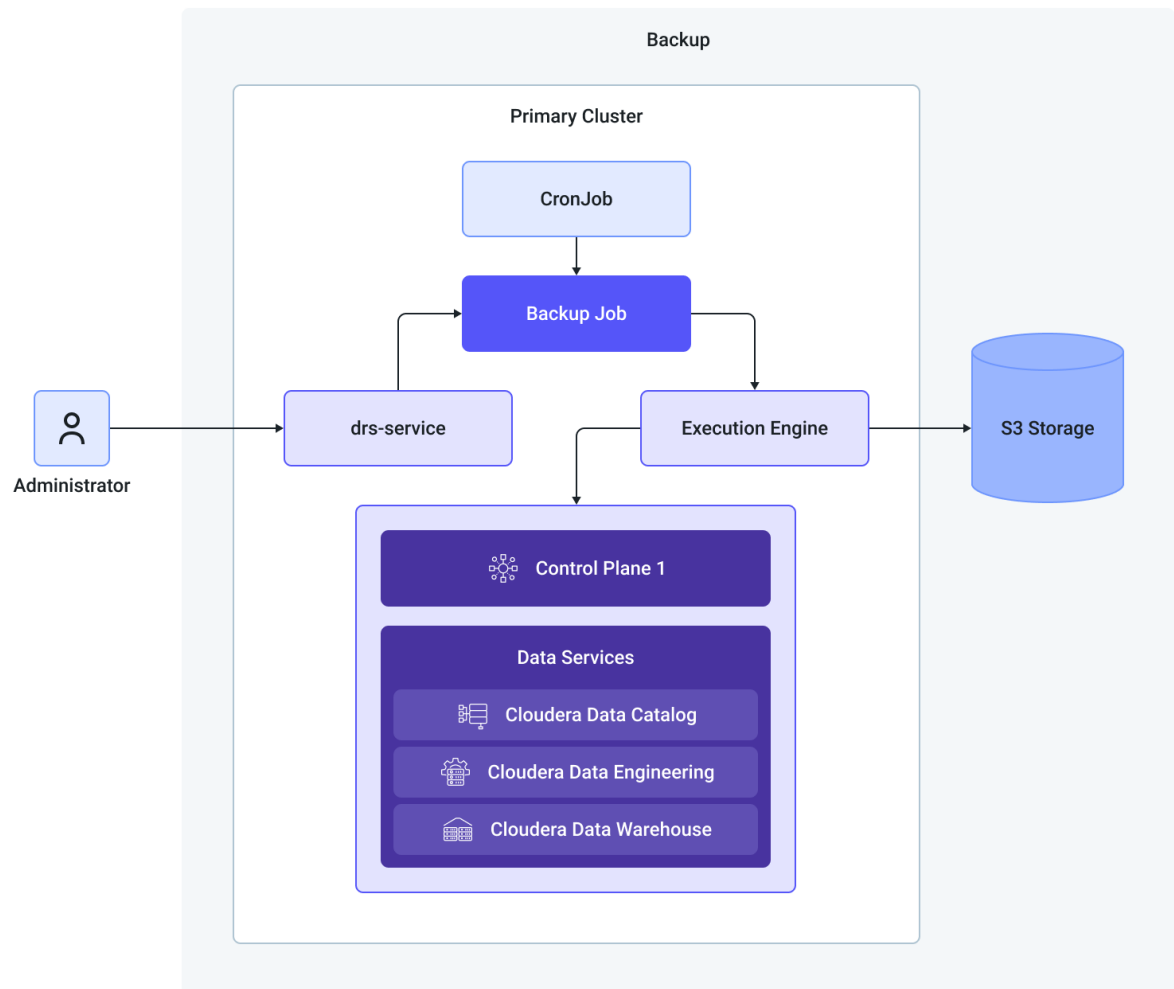
The Backup operation of the Data Recovery Service performs a backup of the entire Cloudera Data Services on premises cluster.

Backup operation is initiated from the primary cluster. Backing up the data includes the K8s resources and Persistent Volumes (PV) that are required to restore services as-is on the secondary cluster. Cloudera Control Plane is a required entity for every backup process, irrespective of which Data Services that are involved as part of backing up the data is concerned. Additionally, backup and restore support for only the Cloudera Control Plane entity is available.

When Data Services are deployed, they register their backup entities, which is a logical unit of recovery and its associated namespaces. The registration process will be completed during the respective service creation lifecycle.

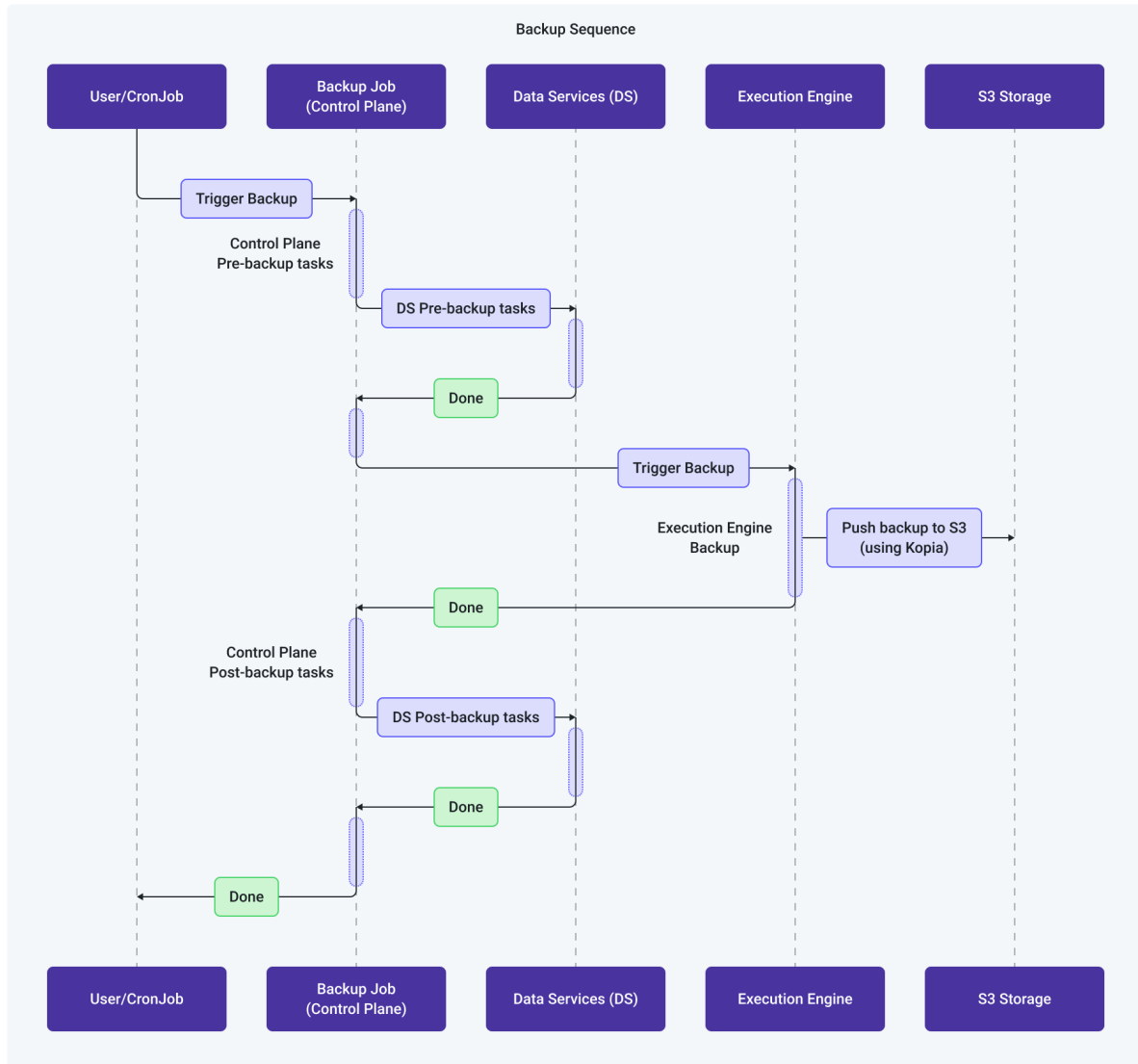
In Cloudera Data Engineering, a backup entity consists of a combination of one 'CDE service' + one or more 'CDE Virtual Clusters'. This expands to a 'dex-base-***' namespace + one 'dex-app-***' namespace for each virtual cluster.

The backup metadata is the sole basis which implies that it is the definitive evidence for initiating the recovery process. By using the backedup metadata, the secondary cluster inspects the contents of a backup and executes restore operations with full context of the original cluster's state.



The Backup Execution Lifecycle

This flow illustrates the sequential orchestration from trigger to external S3 persistence.



Understanding the backup operations

The end-user manages the end-to-end flow of the backup processes which comprises fine grain details of the data, storage entities, and governance capabilities.

Based on the backup entities registered by data services, the end user will be able to pick and choose the candidates for each configured backup. The default part of all the backup operations is managed by the Cloudera Control Plane entity. Users can choose the required entities that need to be restored from the backup.

S3-compatible storage is the one and the only single, centralized repository for all backup and recovery data.

Why this policy matters for disaster recovery:

- **Complete Independence:** If your primary cluster goes down completely, the secondary cluster doesn't need to talk to the failed system or rely on its internal database.
- **Instant Discovery:** The secondary cluster can read the metadata straight from S3 to see all available restore points.
- **Seamless Failover:** You can reconstruct your environment on the secondary cluster without any dependency on the primary site.

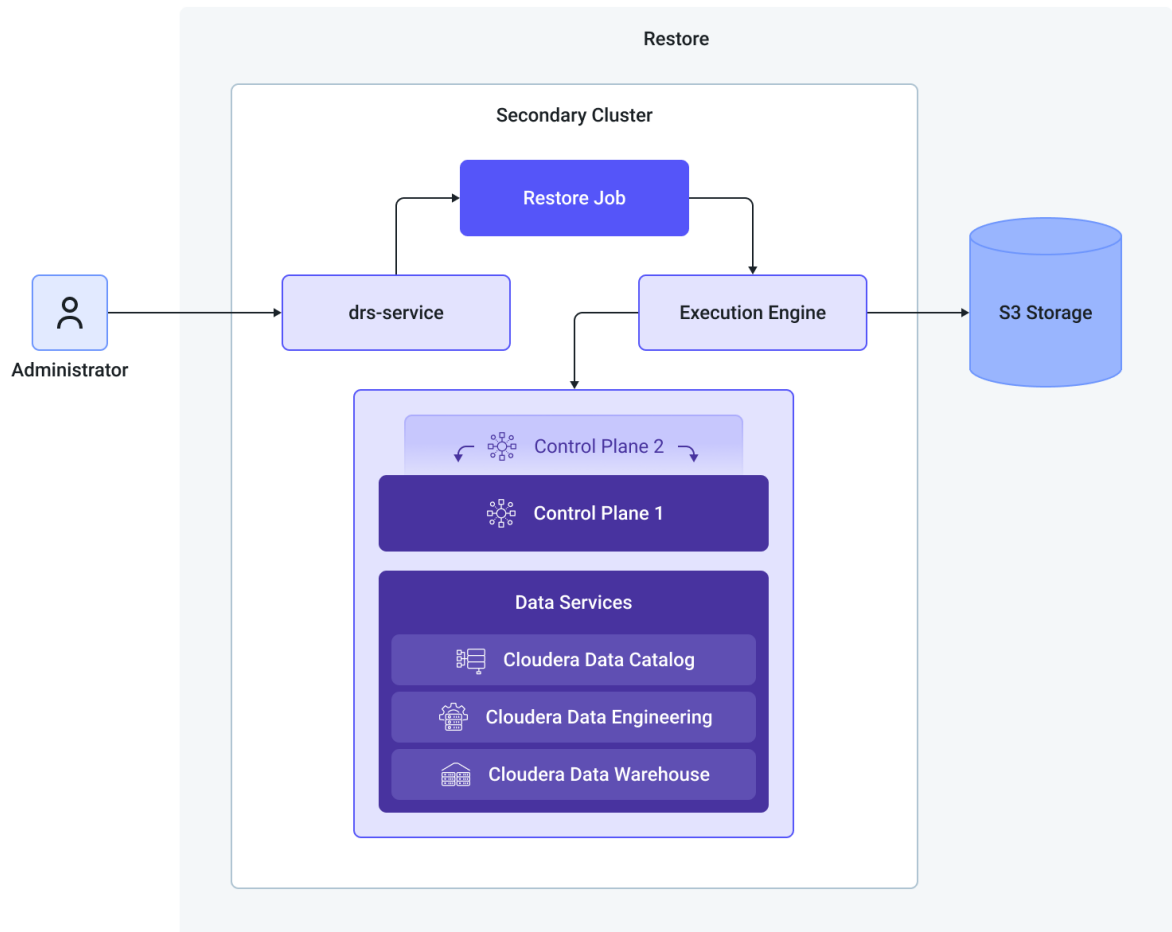
About the Restore service

The Disaster Recovery process enables an end-to-end restore of the Data Services on-premises cluster from an S3 compatible external storage to a target secondary cluster.

The S3 backup repository contains all Kubernetes resources and Persistent Volume (PV) data required to restore services to their original state.

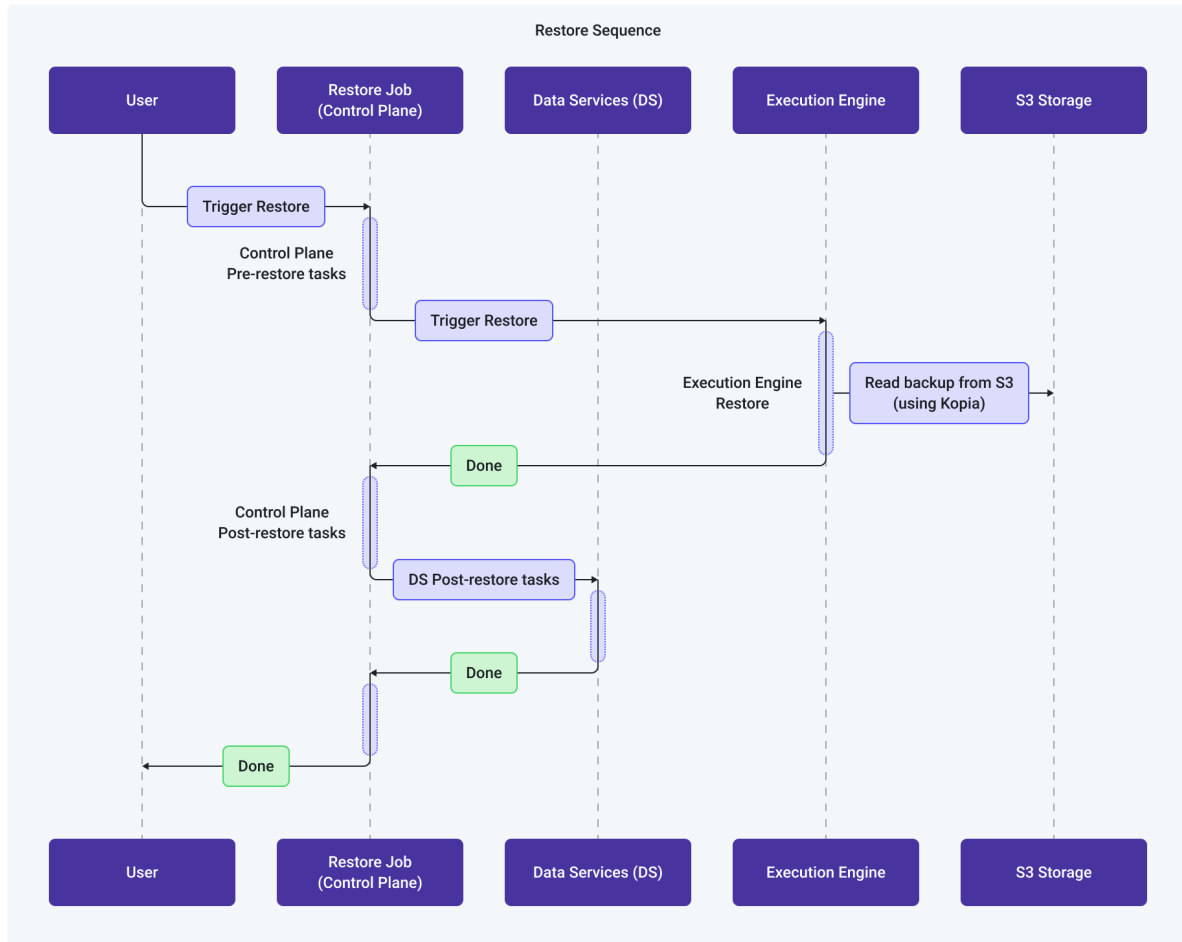


Caution: The Data Recovery restore process is **destructive**. Initiating a restore will completely purge/clean all pre-existing namespaces in the target cluster that match those in the S3 backup object. Once initiated, the original state of the target cluster cannot be recovered.



The Restore Execution Lifecycle

The restore process includes retrieving the backed up data, mapping the entities, backing up attributes, clearing up target cluster namespaces, and post restore operations.



Understanding the restore operations

You must review the following information before initiating the restore process.

1. The **Restore** Operation is initiated directly on the target secondary cluster.
2. Based on the target cluster's configured storage, the operator queries the S3 storage bucket and displays available backup candidates.
3. The end-user selects a specific backup candidate.
4. Using the backup metadata stored alongside the data in S3, the user can select specific entities to restore or opt for a full cluster restore.
5. You can select the scope which is a subset of the backed up data.



Important: Cloudera Data Warehouse service is currently not supported for the secondary cluster restore when pointing to a secondary Base cluster.

Backup Pre-Restore local state

To ensure cluster accessibility and continuity after the restore, the Cloudera Control Plane automatically backs up key attributes unique to the target secondary cluster before destructive actions occur.

- Tracked Attributes: Target Control Plane URL, local ingress settings, and cluster-specific endpoints.

Namespace Purging and Data Restore



Caution: During the data restore operations, the Cloudera Management Console gets into an inoperable state.

1. The operator cleans/purges target cluster namespaces that match those in the incoming backup object.
2. Kubernetes resource manifests and PV data are transferred from S3 and applied as-is to the target cluster.

Post-Restore Reconciliation

Because the restored backup metadata inherently contains configuration details from the primary cluster (such as primary URLs), a reconciliation step runs immediately after the restore operation:

The Cloudera Control Plane retrieves the local secondary cluster attributes saved during the pre-restore step.

Deployment manifests, Ingress resources, and service configurations are dynamically updated to reflect the secondary cluster's parameters (for example, updating Ingress URLs from the primary cluster endpoint to the secondary cluster endpoint).

Cluster services return to a fully operational state.

This is necessary because the backup file you are restoring contains the web address of the original (primary) cluster. Once the restore operation is completed, the system makes use of the saved settings to update the web addresses across the applications and connections, making sure everything points to the new cluster and is fully operational.

Creating a backup

You can backup your existing data on the primary cluster to the S3 storage.

About this task

Procedure to perform the backup operation to secure the data.

Before you begin

As a prerequisite, you must have linked the S3 storage to the primary cluster. The cluster configurations for both the primary and secondary clusters must be the same.

About this task

As a prerequisite, you must link the S3 storage to the primary cluster. The cluster configurations for both the primary and secondary clusters must be the same.

Procedure

1. Navigate to Cloudera Management Console# Backup and Restore # Select the Backups tab.

Backup and Restore

Backups Restores Backup Policies

Search

Create Backup

Status	Name	Storage Location	Started	Last Updated
Completed	[AUTO] [Hourly] Full backup	In Cluster	09/16/2026 10:05 PM IST	09/16/2026 10:05 PM IST
Completed	[AUTO] [Hourly] Full backup	In Cluster	09/16/2026 8:05 PM IST	09/16/2026 8:06 PM IST
Completed	[AUTO] [Weekly] Full backup	In Cluster	09/16/2026 3:35 PM IST	09/16/2026 3:35 PM IST
Completed	[AUTO] [Daily] Full backup	In Cluster	09/16/2026 3:05 PM IST	09/16/2026 3:05 PM IST

2. Click Create Backup.
3. Under the Create Backup pane, enter the following information:

- Provide a name for the backup.
- Under the Storage Destination pane, select the Custom Storage Location or In Cluster.

Selecting Custom Storage Location enables additional flowpath where a Backup Scope option is displayed. The Backup Scope pane defines the boundaries for the data that is being backed up.



Note: Enter the storage location where you want to capture the data. This is a mandatory field before you proceed further.

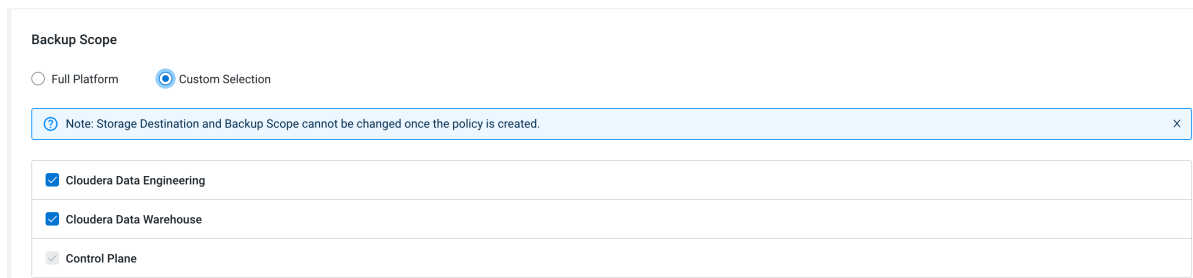
4. Select Full Platform or Custom Selection option.

- Full Platform backup picks up the data from the entire cluster.



Caution: Once the Backup policy is created, the Storage Location and Backup Scope cannot be changed or modified. Optionally, you can discard the changes by clicking Cancel.

- Custom selection specifically backs up from one or more entities from the following:
 - Control Plane (default)
 - Control Plane + Supported Data Services



Backup Scope

☐ Full Platform ☒ Custom Selection

Note: Storage Destination and Backup Scope cannot be changed once the policy is created.

☒ Cloudera Data Engineering

☒ Cloudera Data Warehouse

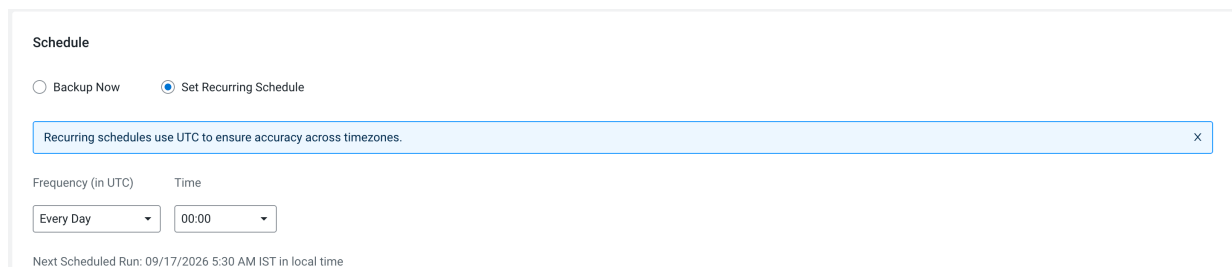
☐ Control Plane

- Under the Schedule pane, select the Frequency (in UTC) for the backup and enter the valid values. The next scheduled backup schedule is displayed alongside.

You can select the frequency based on a daily, weekly, and monthly basis.



Important: A frequency that has a combination of a monthly schedule and a selected day as 31 will skip for the specific month which does not have that date.



Schedule

☐ Backup Now ☒ Set Recurring Schedule

Recurring schedules use UTC to ensure accuracy across timezones.

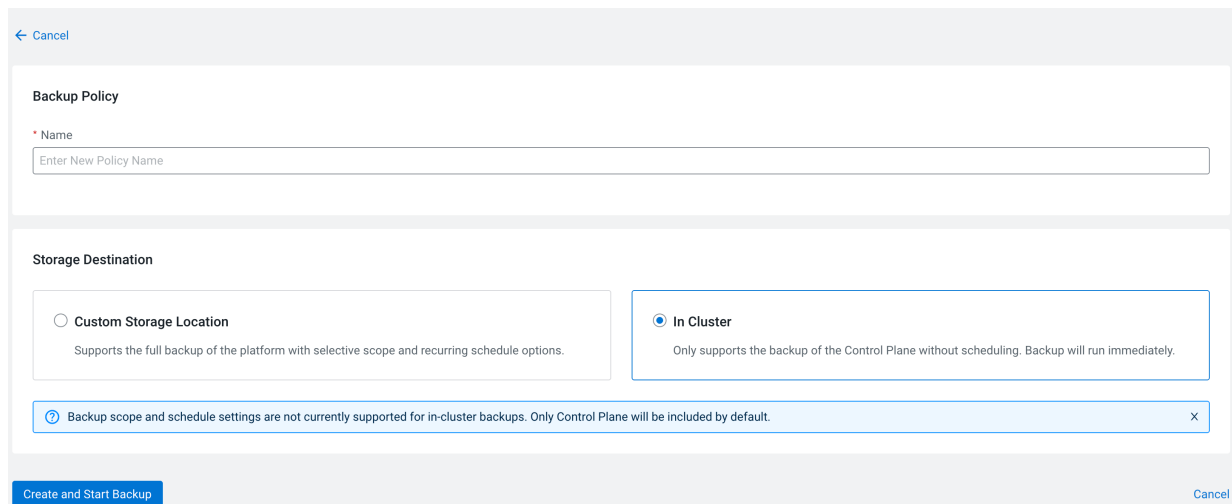
Frequency (in UTC) Time

Every Day 00:00

Next Scheduled Run: 09/17/2026 5:30 AM IST in local time

- Under the Storage Destination pane, select In Cluster to only backup the Cloudera Control Plane data. Backup process starts immediately.

Backup scope and schedule settings are not currently supported for in-cluster backups. Only Cloudera Control Plane will be included by default.



← Cancel

Backup Policy

* Name

Enter New Policy Name

Storage Destination

☐ Custom Storage Location
Supports the full backup of the platform with selective scope and recurring schedule options.

☒ In Cluster
Only supports the backup of the Control Plane without scheduling. Backup will run immediately.

Note: Backup scope and schedule settings are not currently supported for in-cluster backups. Only Control Plane will be included by default.

Create and Start Backup Cancel

- Click Create Backup Policy.

This clickable button appears when you select Custom Storage Location to fully backup the platform with selective scope and recurring schedule options.

- Click Create and Start Backup.

This clickable button appears when you select In Cluster to only backup the Cloudera Control Plane data.

The newly created backup policy is displayed under the Backup Policies tab in the Backup and Restore homepage. The Backup Policies page displays all the backups that are either in progress or scheduled for the next run. You can view the backup details, understand the scope of the backup, edit backup schedule or delete the backup policy.

Triggering a Restore

You can restore the backed up data on the S3 storage.

About this task

Procedure to enable the restore operations to secure the data.

Before you begin

As a prerequisite, you must have backed-up your data on the S3 storage and also you must have linked the S3 storage to the secondary cluster and later perform the restore operation. The cluster configurations for both the primary and secondary clusters must be the same.

Procedure

1. Navigate to Cloudera Management Console # Backup and Restore # Select the Backups tab.

The list of backed up data is displayed.

2.



Select the backed up file and click on the icon and select Restore Backup.



Caution: When restoring data / contents from the primary cluster to secondary cluster, the environments are automatically rewired as part of the restore process, with an option to override the default destination, if needed. You must map all environments to the new secondary Base clusters **when base clusters change** (when environments in the secondary cluster are configured to use a different Base cluster - likely a secondary Base cluster). Inconsistent environment mapping could cause restore operations to fail.

Backup and Restore



Backups				
Backups Restores Backup Policies				
Search Create Backup				
Status	Name	Storage Location	Started	Last Updated
Completed	velero-drs-1	drs-velero	09/17/2026 11:53 AM IST	09/17/2026 11:52 AM IST
Completed	[AUTO] [Hourly] Full backup	In Cluster	09/17/2026 11:05 AM IST	09/17/2026 11:06 AM IST
Completed	drs2	drs-velero	09/17/2026 10:03 AM IST	09/17/2026 10:01 AM IST
Completed	[AUTO] [Hourly] Full backup	In Cluster	09/17/2026 8:35 AM IST	09/17/2026 8:36 AM IST
Completed	[AUTO] [Weekly] Full backup	In Cluster	09/17/2026 12:05 AM IST	09/17/2026 12:05 AM IST
Completed	[AUTO] [Daily] Full backup	In Cluster	09/16/2026 11:35 PM IST	09/16/2026 11:35 PM IST
Completed	drs1	drs-velero	09/17/2026 10:00 AM IST	09/16/2026 2:38 PM IST
Completed	naina	drs-velero	09/17/2026 10:00 AM IST	09/10/2026 7:17 PM IST

- View Details
- View Scope
- View Log
- Restore Backup
- Delete Backup

3. On the Restore Scope pane, select Full Platform or Custom Selection option.

- Full Platform backup picks up the data from the entire cluster.
- Custom selection specifically backs up from one or more entities from the following:
 - Control Plane (default)
 - Control Plane + Supported Data Services

Restore Backup



← Cancel

Restore Scope

☒ Full Platform ☐ Custom Selection

Environments

The environments below will be restored automatically.

Edit an environment to provide a custom Cloudera Manager destination if needed.

Environment Name	Cloudera Manager URL
host-2	https://ccycloud-1.drs-cm1.root.comops.site:7183

< 1 >

Selected Backup

Backup Name
velero-drs-1

Included Namespaces
cdp, vault-system, cdp-obs

Storage Location
drs-velero

Backup CRN
cm:altus.drs:us-west-1:altus.backup:c7c84d81-6848-4eb2-a6b2-ebff71a91242

Start Restore Cancel

CLUSTER Management Console

Restore Backup

← Cancel

Restore Scope

☐ Full Platform ☒ Custom Selection

☒ Data Catalog Profiler
☒ Cloudera Data Engineering
☒ Cloudera Data Warehouse
☒ Control Plane

Environments

The environments below will be restored automatically.

Edit an environment to provide a custom Cloudera Manager destination if needed.

Environment Name	Cloudera Manager URL
host-1	https://ecs-i2-base-1.vpc.cloudera.com:7183

< 1 >

Selected Backup

Backup Name
testttt

Included Namespaces
dcprofiler, dex-base-hhvhmt8, dex-app-6j9wz7f, warehouse-host-1, compute-test-hive, cdp, vault-system, cdp-obs

Storage Location
test-drs

Backup CRN
cm:altus.drs:us-west-1:altus.backup:7be745bd-c2cb-455d-bd82-1f65854d8ca2

Start Restore Cancel

4. Click Start Restore to commence restoring the selected entities.

Key Data Recovery Scenarios

The framework covers two primary data recovery models depending on the scope of the disaster.

- **Secondary Cluster Restore:** Used when an entire production (primary) cluster is destroyed, lost, or unreachable. Data is restored from external storage (S3) into a separate, dedicated Disaster Recovery (DR) secondary cluster.
- **Same-Cluster Restore:** Used for localized failures or data corruption (for example, accidental deletions) where the cluster remains healthy, but the system needs to revert to a previous healthy state.

Prerequisite Configuration

- Both the primary and secondary clusters must be connected to the exact same external S3 storage bucket destination.
- The secondary (DR) cluster operates in a clean state until a restore procedure is initiated.

Secondary Cluster Restore

In the event of an impending data loss situation in your primary cluster, you can restore the backed up data from the Primary cluster to a secondary cluster. This is referred to as Disaster Recovery.

Before you start this process, you must be aware about the existing functionality and take action accordingly:



Caution: Disaster Recovery Service supports backing up and restoring Cloudera Data Warehouse workloads, including Virtual Warehouses, Cloudera Data Visualization, and shared Data Explorer (Hue) instances. However, the target ECS cluster where you perform the restore must belong to the same base cluster environment as the source ECS cluster. Restoring Cloudera Data Warehouse workloads across different base clusters (or secondary base clusters) is not supported.

- As part of its creation and backing up the data to the external storage S3, the existing Primary cluster environment points to its own unique Cloudera Manager and Base versions.
- As part of the restore operations, you must re-point or update the Primary cluster based Cloudera Manager and Base cluster details with the Secondary cluster based Cloudera Manager and Base details to establish connectivity. The S3 bucket name must be the same on both the Primary and Secondary clusters.



Important: For example purposes, **CM1** and **CM2** are referred to as the supported Cloudera Manager instance 1 and 2 in primary and secondary clusters respectively.

1. **Backup generation:** The primary cluster continuously or periodically pushes control plane state, data, and service configurations to the designated external S3 bucket.
2. **Backup discovery:** The secondary cluster detects available backup snapshots from the shared S3 bucket in real time.
3. **Trigger restoration:** In a likely event of a primary cluster outage, initiate the Restore procedure directly from the secondary cluster.
4. **Cloudera Control Plane Rewiring (Re-pointing):**
 - During the restore configuration step, select the secondary cluster's environment settings.
 - This updates all references to point to the secondary Cloudera Manager (CM2) instead of the original primary Cloudera Manager (CM1).
5. **Cluster Environment Wipe:** The secondary cluster automatically clears its existing base environment to prepare for state replacement.
6. **State and Data Injection:** The Cloudera Control Plane state, data services, and rewired configuration pointers are pulled from S3 and applied to the secondary cluster, making it the new active primary.

Same-Cluster Restore

This is also referred to as In-Place restore.



Caution: Restore can only be executed to the same Cloudera Data Services on premises version that the backup was taken from. Backups from Cloudera Data Services on premises SP3 or older clusters cannot be used to take an Cloudera Data Services on premises SP4 cluster back to SP3 state (rollbacks are not supported using this DRS feature).

1. **Trigger Local Restore:** Navigate to the Backup and Restore page on the active primary cluster.
2. **Select Snapshot:** Choose the target historical backup snapshot stored in S3.
3. **Execute In-Place Restore:** Initiate Start Restore operation.



Note: Cloudera Manager re-pointing is skipped during this process because the underlying Cloudera Control Plane targets remain unchanged.