

Cloudera Replication Manager

CDP Private Cloud Data Services Replication Manager Overview

Date published: 2022-11-18

Date modified: 2023-06-09

CLouDERA

<https://docs.cloudera.com/>

Legal Notice

© Cloudera Inc. 2024. 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

Introduction to Replication Manager.....	4
Replication Manager service web interface.....	4
Overview page.....	5
Clusters panel.....	5
Policies panel.....	5
Jobs panel.....	5
Issues & Updates panel.....	6
Clusters page.....	6
Replication Policies page.....	7

Introduction to Replication Manager

Replication Manager is a service in CDP Private Cloud Data Services. You can use this service to copy and migrate HDFS data, Hive external tables, and Ozone data between CDP Private Cloud Base 7.1.8 or higher clusters using Cloudera Manager version 7.7.3 or higher.

The following support matrix table lists the lowest supported Cloudera Manager version, Cloudera Runtime version, clusters, and supported services for Replication Manager:

Source cluster	Lowest supported source Cloudera Manager version	Lowest supported source Cloudera Runtime version	Destination cluster	Supported services on Replication Manager
CDP Private Cloud Base	7.7.3	7.1.8	CDP Private Cloud Base	HDFS, external Hive tables, Ozone data

You can access the Replication Manager service on the CDP Private Cloud Data Services web interface. To replicate data between clusters, add the source and destination clusters on the [Management Console Clusters](#) page, and then create HDFS and Hive replication policies in Replication Manager. The **Replication Policies** page shows the progress and status of replication policy jobs.

The Replication Manager supports the following functionalities:

HDFS replication policies

These policies replicate files or directories in HDFS.

Some use cases to use these replication policies are to:

- replicate required data to another cluster to run load-intensive workflows on it which optimizes the primary cluster performance.
- deploy a complete backup-restore solution for your enterprise.

Hive replication policies

These policies replicate Hive metadata and Hive external tables.

Some use cases where you can use these replication policies are to:

- backup legacy data for future use or archive cold data.
- replicate or move data to cloud clusters to run analytics.

Ozone replication policies

These policies replicate Ozone data.

Some use cases where you can use these replication policies are to:

- replicate or migrate the required Ozone data to another cluster to run load-intensive workloads.
- back up data.

You can monitor the replication policies and its progress on the **Overview** page and the other pages of Replication Manager. You can also monitor the statistics related to each replication policy on the [CDP Private Cloud Data Services Management Console Dashboard Monitoring Dashboard](#) page.

Replication Manager service web interface

You can access the Replication Manager service by logging into “CDP Private Cloud Data Services”. The Replication Manager web interface provides insights into issues and updates related to replication policies and replication policy jobs, shows the geolocation of each cluster, and enables you to create and monitor replication policies.

After you click Replication Manager on the **CDP Private Cloud Data Services** web interface, the **Overview** page of the Replication Manager appears.

The Replication Manager has the following pages:

- Overview
- Clusters
- Replication Policies

Overview page

When you click Replication Manager on the CDP Private Cloud Data Services web interface, the Overview page appears. The page provides a snapshot of the Replication Manager service. It provides insights into issues and updates related to various entities and resources through dashboards such as Clusters, Policies, Jobs, and Issues & Updates.

The following panels appear on the **Overview** page:

- Clusters
- Policies
- Jobs
- Issues & Updates

Click Create Policy to create a replication policy.

Clusters panel

The Clusters panel on the Overview page is updated after you add a CDP Private Cloud Base 7.1.8 or higher cluster on the Clusters page in the Management Console.

The **Clusters** panel shows the following cluster status:

- Number of clusters that are currently not running as expected or in **Error** state.
- **Active** clusters that are currently available to run the replication jobs.
- Clusters having less than 10% disk capacity appear as **Warning**. Click the number to open a table to track the cluster name and the exact disk capacity remaining.
- **Total** number of clusters that are in use.

To investigate the issues associated with clusters that have an error or warning status, use Cloudera Manager or Management Console.

Policies panel

The Policies panel on the Overview page tracks the number of replication policies that are in use and their status.

The **Policies** panel shows the following status for the replication policies:

- **Active** replication policies in **Submitted** or **Running** state.
- Replication policies that are **Suspended** by the administrator.
- **Unhealthy** replication policies are associated with a cluster designated as **Error** on the **Clusters** panel. Click the number to find out the policy names, the names of its associated source and destination clusters, and the services that have stopped on the source cluster or destination cluster.
- **Total** number of running policies.

Jobs panel

The Jobs panel on the Overview page tracks the total number of running and failed replication jobs and their status in Replication Manager.

The **Jobs** panel shows the following details:

- **In Progress** jobs.
- **Failed Last** replication policies are the policies for which the last job failed to complete.
- **Failed in Last 10** replication policies are the policies for which at least one of the last ten jobs failed.
- **Total** number of available jobs in Replication Manager.

Click the number in the panel to apply a filter to the **Issues & Updates** table to view the required policies.

Issues & Updates panel

The Issues & Updates panel on the Overview page lists the replication policies that have running jobs with at least one job in Failed status in the most recent 10 jobs. If you do not see any policy, it indicates that the last 10 jobs of all the replication policies were successful.

The **Issues & Updates** panel has the following columns:

- Current **Job Status** of the job.

When the job status is running, the status circle icon and a progress bar appear. For jobs that are not running, the status circle icon appears with **Success**, **Failed**, or **Ignored** text. Hover over the **Failed** status and click **View Log** to view the job log.

- Replication **Policy** name. Click the policy name to view the policy details on the **Replication Policies** page.
- **Source** cluster associated with the replication policy.
- **Destination** cluster associated with the replication policy.
- **Service** shows **HDFS** for HDFS replication policies and **Hive** for Hive replication policies.
- **Policy History** of the last ten job statuses. The column also shows the status as colored dots which you can click to view the policy details on the **Replication Policies** page:
 - Green indicates that the job completed successfully.
 - Red indicates that the job did not complete.
 - Gray indicates that the job did not start because a previous instance of the policy is still in progress. Only one run of a job can be in progress at one time. If a job is consistently ignored, edit the replication policy to modify its frequency.

When you click the colored dots, the page appears with the filter preset to show the information about the specified policy.

- **Runtime** or time taken to complete the most recent job.
- Most recent job that **Started**.
- Most recent job that **Ended**.

The **Actions** menu shows the following options:

- **Abort Job** aborts a running job.
- **Re-run Job** starts another instance of the policy. This option is not available for running jobs.
- **Delete Policy** removes the replication policy permanently. The delete operation cannot be undone.
- **Suspend Policy** pauses a running replication policy job.
- **Activate Policy** resumes a suspended replication policy.

Clusters page

The Clusters page shows the geolocation of each cluster and helps you to easily identify the status of cluster services, using interactive color markers on the map. The page also shows the total number of clusters enabled for Replication Manager, the number of clusters that are active or in an error state, and the number of clusters for which a warning is issued.

The **Clusters** page shows the following information about each cluster:

- Cluster health **Status**
- **CDP Status** for clusters as **SYNCED** (cluster is registered and running), **SYNC_ERROR** (cluster is down), or **SYNC_IN_PROGRESS** (cluster registration is in progress).
- Source and destination cluster names
- Runtime versions
- Number of **Nodes** in the source cluster
- Number of replication **Policies** between the specified clusters

- **Location** of the cluster

Use the **Actions** menu to Add Policy (create a replication policy); Launch Cloudera Manager; or Sync the cluster configuration.

Click Add Cluster to register a CDP Private Cloud Base cluster 7.1.8 or higher cluster in **Management Console** to use in replication policies.

The map shows the geolocation of each cluster and helps you to easily identify the status of cluster services, using the following interactive markers on the map:

- **Red** indicates that at least one required service has stopped on the cluster.
- **Orange** indicates that all the required services are running on the cluster but the remaining disk capacity on the cluster is less than 10%.
- **Green** indicates that all the required services are running on the cluster and the remaining disk capacity is greater than 10%.

Hover over a marker on the map to view the data center associated with the cluster, the cluster name, and the number of Replication Manager policies that are associated with that cluster.

To investigate the issues associated with clusters that have an error or warning status, launch Cloudera Manager.

Replication Policies page

The Replication Policies page shows all the replication policies available in Replication Manager, their status, and a detailed view for each.

The **Replication Policies** page shows the number of policies that have the following status:

- **Error** or failed replication policies. Click the number to view the policy names, the names of the source and destination clusters, and the services that are stopped on the source or destination cluster.
- **Active** replication policies that are in **Submitted** or **Running** state.
- **Suspended** replication policies. Administrators can suspend a running policy.
- **Expired** replication policies.
- **Total** number of running policies.

You can also view the following policy details on the **Replication Policies** page:

- Current replication policy **Status**
- Policy **Type**
- Replication policy **Name**
- **Source** cluster name
- **Destination** cluster name
- **Jobs** that were run for the replication policy and the job status
- **Duration** to run the policy
- **Last Success** timestamp of the last successful run
- **Next Run** timestamp of the next scheduled run

Click **Actions** to perform more actions on a replication policy.

Expand a replication policy to view the following details about the policy:

- **Description** for the replication policy that you added while creating or editing the policy.
- **Source** cluster details
- **Destination** cluster details
- **Frequency** to run the policy which includes the next run timestamp and the last run timestamp.
- HDFS **Snapshot** shows whether it is enabled or not for the replication policy.
- **Queue** name for the policy otherwise it displays default.
- **Max Bandwidth** allocated for a map task

- **Max Maps** or maximum number of simultaneous copies per replication job