

Replication Manager Reference

Date published: 2019-11-07

Date modified: 2024-07-08

CLOUDERA

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

CDP CLI for Cloudera Replication Manager.....	4
CDP CLI options for Replication Manager.....	4
Adding cloud credentials in Replication Manager using CDP CLI.....	8
Creating HDFS replication policy using CDP CLI.....	10
HDFS replication policy definition JSON file.....	11
Managing HDFS replication policies using CDP CLI.....	14
Creating Hive replication policy using CDP CLI.....	16
Hive replication policy definition JSON file.....	17
Managing Hive replication policies using CDP CLI.....	22
Creating HBase replication policy using CDP CLI.....	24
HBase replication policy definition JSON file.....	26
Managing HBase replication policies using CDP CLI.....	29

CDP CLI for Cloudera Replication Manager

You can use CDP CLI commands to create and manage HDFS, Hive, and HBase replication policies in Cloudera Replication Manager. You can also register the ABFS and AWS cloud credentials to use in Replication Manager. The CDP CLI commands for Replication Manager are under the "replicationmanager" CDP CLI option.

Prerequisites

To use CDP CLI commands for Replication Manager, ensure that the following are available:

1. CDP CLI client.

For information about installing a CDP CLI client, see [Installing CDP CLI client](#).

2. Access to CDP CLI.

Choose one of the following methods to log into CDP CLI:

- Interactive method. This login method grants a 12-hour access key to the CLI. For more information, see [Logging into CLI/SDK](#).
- Traditional method. In this method, you generate access credentials and configure the `~/.cdp/credentials` file with the key pair. This login method allows you to withdraw the access permission by removing the access credentials from the `~/.cdp/credentials` file. For more information, see [Generating an API access key](#) and [Configuring CDP client with the API access key](#).

Access CLI help

CDP CLI includes help that can be accessed by using the `cdp help` command. For information about a certain CDP CLI, use the `cdp [***MODULE-NAME***] [***COMMAND-NAME***] help` command.

You can also find all of the CDP CLI commands in the [CDP CLI Reference documentation](#).

Related Information

[Introduction to Replication Manager](#)

[Support matrix for Cloudera Replication Manager](#)

[Using HDFS replication policy](#)

[Using Hive replication policy](#)





CDP CLI options for Replication Manager

You can use the CDP CLI commands to create, activate, and delete HDFS, Hive, and HBase replication policies.

CDP CLI options


You can use the following CDP CLI options to perform tasks in Replication Manager:

CDP CLI option	Description
<code>abort-policy*</code>	Aborts all running instances of the specified replication policy. Provide the CRN of the cluster where the replication policy is created and stored.
<code>activate-hbase-policy</code>	Resumes the specified suspended HBase replication policy. If the target Cloudera Manager API version is higher than 50, the command resumes all the HBase replication policies between the same source and destination cluster. However, if the target Cloudera Manager API version is lower than 45, the command resumes only the specified HBase replication policy. This command is not available for target Cloudera Manager API versions between 45 and 50.

CDP CLI option	Description
activate-policy	Resumes the specified suspended HDFS or Hive replication policy. Provide source CRN for HDFS replication policies and target CRN for Hive replication policies.
activate-snapshot-policy*	Resumes a suspended snapshot policy run.
collect-diagnostic-bundle	Triggers the diagnostic bundle collection for the specified replication policy in the target Cloudera Manager. Use the commandId in the command output to download the diagnostic bundle.
continue-hbase-setup	Continues the first-time setup on the target cluster.  Note: If an HBase replication policy has a classic cluster source, the HBase service on the source is not restarted automatically. You must restart the source manually and then run the continue-hbase-setup command. This command is not required if the source cluster is not a classic cluster.
create-abfs-credential	Adds a ABFS cloud credentials to use in Replication Manager.  Important: Before you register the ABFS cloud credentials in Replication Manager, ensure that you update cloud credentials in the source cluster Cloudera Manager UI. For more information, see Registering ABFS Cloud account in Replication Manager .
create-aws-credential	Adds AWS cloud credentials to use in Replication Manager.  Important: Before you add AWS credentials, see Registering Amazon S3 cloud account .
create-gcs-credential*	Adds GCS cloud credentials to use in Replication Manager.
create-hbase-policy	Creates an HBase replication policy with the given name on the specified cluster.
create-policy	Creates an HDFS or Hive replication policy based on the provided parameters. Provide source CRN for HDFS replication policies and target CRN for Hive replication policies.  Important: Only a non-machine user can run this CDP CLI command. Otherwise, an HTTP 500 error appears.
create-snapshot-policy*	Creates a HDFS or HBase snapshot policy depending on the details you provide in the snapshot definition.
delete-credential	Deletes the registered credentials from Replication Manager.
delete-hbase-policy	Deletes the specified HBase replication policy permanently. Enter --force if a normal delete fails. For example, when the source cluster is unreachable. Otherwise, enter --no-force.
delete-policy	Deletes the specified HDFS or Hive replication policy. Provide source CRN for HDFS replication policies and target CRN for Hive replication policies.
delete-snapshot-policy*	Deletes the specified snapshot policy.
download-diagnostic-bundle	Generates a bundleFile, a binary string in base64 encoded format, as a ZIP file. Run a script to save the response as a file to your machine. For example, cat response.json jq -r '.bundleFile' base64 -D > bundle.zip. The file is generated only if the bundleStatus in the get-command-status command shows <i>DOWNLOADABLE WITH CLI</i> .
get-cluster-config	Retrieves the configuration of a specified cluster.

CDP CLI option	Description
get-command-status	Returns the current status of the collect-diagnostic-bundle command. You can also view the status of any Cloudera Manager command using the relevant input command ID.
get-credentials	Returns the cloud credentials that are available on the specified cluster. If you provide the credential name and credential ID, the ID is considered by the CDP CLI command.
get-hbase-time-series*	<p>Returns the time series data for an HBase replication peer depending on the provided parameters.</p> <p>You can enter one of the following HBase replication peer metric:</p> <ul style="list-style-type: none"> age_of_last_shipped_operation log_edits_filtered_rate log_edits_read_rate log_queue_size log_read_in_bytes_rate shipped_batches_rate shipped_ops_rate shipped_size_in_kb_rate age_of_last_shipped_operation_25th_percentile age_of_last_shipped_operation_75th_percentile age_of_last_shipped_operation_90th_percentile age_of_last_shipped_operation_98th_percentile age_of_last_shipped_operation_99th_percentile age_of_last_shipped_operation_999th_percentile age_of_last_shipped_operation_max age_of_last_shipped_operation_mean age_of_last_shipped_operation_median age_of_last_shipped_operation_min age_of_last_shipped_operation_rate shipped_hfiles_rate size_of_hfile_refs_queue <p>You can enter one of the following aggregate rollup level for time series data:</p> <ul style="list-style-type: none"> RAW TEN_MINUTELY HOURLY SIX_HOURLY DAILY WEEKLY
get-snapshot-policy*	Retrieves the details about the specified snapshot policy.
list-all-credentials	Provides a detailed list of cloud credentials across all clusters that are available for Replication Manager.
list-cluster-service-statuses	Provides the current status of the services on all the clusters that are available for Replication Manager.
list-clusters	Provides a detailed list of all the clusters that are available for Replication Manager.
list-paired-hbase-clusters	Lists the paired clusters to use for HBase replication policies. The first-time setup is complete for paired clusters.
list-policies	Lists the available replication policies across all the clusters. Enter the source cluster CRN to view the HDFS replication policies and target cluster CRN to view all the Hive and HBase replication policies.

CDP CLI option	Description
list-policy-jobs*	<p>Returns the list of jobs triggered by the replication policy. This command is a paginated operation, therefore multiple API calls might be required to retrieve the entire dataset of results. You can use one of the following methods to retrieve the results:</p> <ul style="list-style-type: none"> Use <code>--no-paginate</code> argument to disable pagination. Use <code>--max-items [***TOTAL NUMBER OF ITEMS TO RETURN***]</code> <code>--starting-token [***TOKEN TO START PAGINATING***]</code> <code>--page-size [***PAGE SIZE***]</code> to return a specific number of jobs at a time.
list-snapshot-policies*	Returns a list of all the snapshot policies across all available clusters.
list-snapshot-policy-jobs*	Retrieves the snapshot history of the specified snapshot policy.
repair-hbase-policy*	Runs the last failed command for a failed HBase replication policy.
rerun-policy*	Runs the specified HDFS or Hive replication policy.
restore-snapshot*	Restores the specified snapshot.
retry-failed-hbase-first-time-setup*	<p>Runs the first time setup configuration between the clusters in the specified HBase replication policy if the first time setup failed for the replication policy.</p> <p>You can use the following arguments as required:</p> <ul style="list-style-type: none"> <pre>--machine-user [***ENTER USERNAME AND PASSWORD OF MACHINE USER***]</pre> <p>Syntax: user=[***STRING***], password=[***STRING***], createUser=[***ENTER 'true' TO CREATE A NEW MACHINE USER, OR 'false' TO RETURN AN ERROR IF THE USER DOES NOT EXIST***]</p> <pre>--source-restart-type [***ENTER RESTART OR ROLLING_RESTART TO SPECIFY THE RESTART TYPE***]</pre> <pre>--target-restart-type [***ENTER RESTART OR ROLLING_RESTART TO SPECIFY THE RESTART TYPE***]</pre>
retry-failed-hbase-snapshots*	Reruns only the failed initial snapshots in the replication policy if the policy failed to replicate existing data in some or all the tables
suspend-hbase-policy	Pauses an active HBase replication policy.
suspend-policy	Stops all the replication tasks defined for the HDFS or Hive replication policy. Provide source CRN for HDFS replication policies and target CRN for Hive replication policies.
suspend-snapshot-policy*	Suspends the specified snapshot policy.
update-abfs-credential*	Updates the ABFS cloud credentials.
update-aws-credential*	Updates the AWS cloud credentials for Replication Manager.
update-gcs-credential*	Updates the GCS cloud credentials for Replication Manager.
update-hbase-policy	Modifies an existing HBase replication policy. You can change the name and description of the replication policy. You can also delete one or more tables in the replication policy.

CDP CLI option	Description
update-policy*	<p>Modifies the HDFS or Hive replication policy.</p> <p>To modify the replication policy, make appropriate changes in the JSON file, save the file, and use it in the update-policy command. Use the rerun-policy command to run the policy.</p> <p> Note: Some elements such as cloud credential, source or target cluster, or source dataset cannot be modified. Create another replication policy to use the new values.</p>
update-snapshot-policy*	Modifies an existing snapshot policy depending on the arguments you choose.
verify-hbase-cluster-pair	<p>Verifies whether the specified pair of clusters are paired, not paired, or wrongly paired.</p> <p>The wrongly paired status indicates that one or both of the specified clusters are paired to other clusters.</p>
<p>*The option is a technical preview feature and is not ready for production deployment. The components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the use of preview components, which should be used by customers at their own risk. For more information, contact your Cloudera account team.</p>	

CDP CLI options to create a replication policy



Note: Only a non-machine user can run the replicationmanager create-policy CDP CLI command to create a replication policy.

The following parameters are available in the replicationmanager create-policy CDP CLI option:

Parameter	Description
--cluster-crn	<p>Enter the cluster CRN. To determine the cluster CRN, run the list-clusters command.</p> <p>Provide the following CRN when you create a replication policy:</p> <ul style="list-style-type: none"> • Source cluster CRN for HDFS replication policy. • Target cluster CRN for Hive replication policy. • Target cluster CRN for HBase replication policy.
--policy-name	Enter a unique name for the replication policy.
--policy-definition	Enter the policy definition in the \$(cat [***POLICY DEFINITION FILE NAME**]) format, and then enter the cluster CRN and policy name as command arguments.
--cli-input-json	Enter the policy definition JSON file path using the cat command to read the data from the file to create and run the replication policy.
--generate-cli-skeleton	Shows a policy definition template in JSON format. You can copy the output of this command to a file, add the required parameters, and save it as a JSON file. You can use the filename while creating a replication policy.

Related Information

[Technical preview CDP CLI options for Replication Manager](#)

Adding cloud credentials in Replication Manager using CDP CLI

To replicate data to a storage cloud account, you must register the cloud credentials, so that the Replication Manager can access your cloud account. The supported cloud storage accounts are Amazon S3 and Azure Blob Filesystem

(ABFS). You can add, update, or delete AWS or ABFS cloud credentials to use in Replication Manager using CDP CLI.

About this task

Perform one of the following steps, as necessary, to manage cloud credentials in Replication Manager using CDP CLI:

Procedure

- Add ABFS credentials:

```
replicationmanager create-abfs-credential --name [***CREDENTIAL NAME***]
--clusters [***CLUSTER CRNs SEPARATED BY SPACE***] --type [***ACCESSKEY
OR CLIENTKEY***] --access-key [***ABFS ACCESS KEY***] --storage-accou
nt-name [***ABFS STORAGE ACCOUNT NAME***] --client-id [***CLIENT ID
OF ACTIVE DIRECTORY SERVICE PRINCIPAL ACCOUNT***] --client-secret-
key [***CLIENT KEY OF ACTIVE DIRECTORY SERVICE PRINCIPAL ACCOUNT***] --t
enant-id [***TENANT ID OF ACTIVE DIRECTORY SERVICE PRINCIPAL ACCOUNT***]
```



Note:

- Provide the access key and storage account name if you choose ACCESSKEY type.
- Provide the client ID, client secret key, and tenant ID if you choose CLIENTKEY type.

Before you add ABFS credentials, see [Registering ABFS Cloud account in Replication Manager](#).

- Add AWS credentials:

```
replicationmanager create-aws-credential --name [***CREDENTIAL NAME***]
--clusters [***CLUSTER CRNs SEPARATED BY SPACE***] --type [***IAM OR
ACCESSKEY***] --access-key [***AWS ACCESS KEY***] --secret-key [***AWS
SECRET KEY***]
```



Important: Before you add AWS credentials, see [Registering Amazon S3 cloud account](#).

- Update ABFS credentials:

```
replicationmanager update-abfs-credential --name [***CREDENTIAL NAME***]
--type [***ACCESSKEY OR CLIENTKEY***] --access-key [***ABFS ACCESS
KEY***] --storage-account-name [***ABFS STORAGE ACCOUNT NAME***] --clien
t-id [***ABFS CLIENT ID***] --client-secret-key [***ABFS CLIENT SECRET
KEY***] --tenant-id [***ABFS TENANT ID***]
```

- Update AWS credentials:

```
replicationmanager update-aws-credential --name [***CREDENTIAL NAME***] --
type [***IAM OR ACCESSKEY***] --access-key [***ABFS ACCESS KEY***] --secre
t-key [***AWS SECRET KEY***]
```



Note:

- No parameters are required for the IAM type.
- Provide the access key and secret key if you choose the ACCESSKEY type.

- View all the available registered credentials in a cluster:

```
replicationmanager get-credentials --cluster-crn [***CLUSTER CRN***]
```

- Delete a registered credential from Replication Manager:

```
replicationmanager delete-credential --name [***CREDENTIAL NAME***]
```

Creating HDFS replication policy using CDP CLI

You can use CDP CLI to create an HDFS replication policy. Only a non-machine user can run the "replicationmanager create-policy" CDP CLI command to create a replication policy.

Procedure

1. Log into Replication Manager CDP CLI setup:

```
cdp --profile [***PROFILE NAME***] replicationmanager
```

2. List the clusters to verify whether the required clusters are available:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-clusters
```

3. Verify whether the required services are running on the source cluster:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-cluster-service
-statuses
```

4. Ensure that the cloud credentials are available:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-all-credentials
```

5. Perform the following steps to create a policy definition JSON file:

- a) Open a policy definition JSON file, or copy the output of the following command to a JSON file to generate a policy definition JSON file.

```
cdp --profile [***PROFILE NAME***] replicationmanager create-policy --
generate-cli-skeleton
```

```
cdp --profile hdfs1 replicationmanager create-policy --generate-cli-
skeleton > rm_hdfs1.json
```

- b) Enter the required parameters. Remove the key-value pairs that are not required in the policy definition JSON file for the specific policy. For example, you can remove the hiveArguments key-value pairs when you create an HDFS replication policy.
- c) Save the file.

6. Create the HDFS replication policy.

```
cdp --profile [***PROFILE NAME***] replicationmanager create-policy --cli-
input-json [***POLICY DEFINITION JSON FILE PATH USING CAT***]
```

The cat command reads the data from the policy definition JSON file.

```
cdp --profile local-dev replicationmanager create-policy --cli-input-json "$(cat temp/rm_hdfs1.json)"
```

Results

Replication Manager creates the replication policy and initiates data replication.

What to do next

Verify whether the replication policy is running as expected on the Cloudera Replication Manager Replication Policies page, or run the following command:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-policies --cluste
r-crn [***CRN OF THE CLUSTER WHERE THE REPLICATION POLICIES ARE STORED***]
```

Related Information

[Creating HDFS replication policy using Replication Manager](#)


HDFS replication policy definition JSON file

The policy definition JSON file contains all the parameters required to create an HDFS replication policy. When you edit the file to define an HDFS replication policy, remove the parameters that are not required for the replication policy.

Parameters in HDFS replication policy definition JSON file

The following table lists the parameters in the policy definition JSON file that are required for an HDFS replication policy:

Parameter	Description
clusterCrn	Provide the source cluster CRN for the HDFS replication policy. Replication Manager saves the replication policy in the specified cluster CRN.
policyName	Provide a unique name for the replication policy.
name	Provide the unique name for the policy.
type	Provide FS to create an HDFS replication policy.
path	Provide the HDFS file path in the source cluster.
mapReduceService	Provide the MapReduce or YARN service for the replication policy to use.
logPath	Provide an alternate path for the logs, if required.
replicationStrategy	<p>Provide one of the following options to determine whether the file replication tasks must be distributed among the mappers statically or dynamically:</p> <ul style="list-style-type: none"> • STATIC - Static replication distributes file replication tasks among the mappers up front to achieve an uniform distribution based on the file sizes. • DYNAMIC - Dynamic replication distributes the file replication tasks in small sets to the mappers, and as each mapper completes its tasks, it dynamically acquires and processes the next unallocated set of tasks. <p>Default is DYNAMIC.</p>
skipChecksumChecks	<p>Provide true to skip checksum checks.</p> <p>Default is true.</p> <p>Checksums are used to perform the following tasks:</p> <ul style="list-style-type: none"> • To skip replication of files that have already been copied. When set to true, the replication job skips copying a file if the file lengths and modification times are identical between the source and destination clusters. Otherwise, the job copies the file from the source to the destination. • To redundantly verify the integrity of data. However, checksums are not required to guarantee accurate transfers between clusters. HDFS data transfers are protected by checksums during transfer and storage hardware also uses checksums to ensure that data is accurately stored. These two mechanisms work together to validate the integrity of the copied data.

Parameter	Description
skipListingChecksumChecks	<p>Provide true to skip checksum check while comparing two files to determine whether they are the same or not. Otherwise, the file size and last modified time are used to determine if files are the same or not. Skipping the check improves performance during the mapper phase.</p> <p> Note: If you set skipChecksumChecks to false, the skipListingChecksumChecks is also set to false by default.</p>
abortOnError	<p>Provide true to stop the policy job when an error occurs. This ensures that the files copied up to that point remain on the destination, but no additional files are copied.</p> <p>Default is false.</p>
abortOnSnapshotDiffFailures	<p>Provide true to stop the replication job if a snapshot diff fails during replication.</p>
preserve	<p>Provide true to preserve the block size, replication count, permissions (including ACLs), and extended attributes (XAttrs) as they exist on the source file system.</p> <ul style="list-style-type: none"> blockSize replicationCount permissions extendedAttributes <p>Provide false to use the settings as configured on the destination file system.</p> <p>By default, the source system settings are preserved.</p> <p>In an HDFS replication policy:</p> <ul style="list-style-type: none"> when the permissions parameter is set to true and both the source and destination clusters support ACLs, replication preserves ACLs. Otherwise, ACLs are not replicated. when extendedAttributes is set to true and both the source and destination clusters support extended attributes, the replication process preserves them. when you select one or more of the Preserve options, and you are replicating <i>to</i> S3 or ADLS, the values of all of these options are saved in metadata files on S3 or ADLS. When you replicate <i>from</i> S3 or ADLS to HDFS, you can set the options you want to preserve.
deletePolicy	<p>Provide one of the following options:</p> <ul style="list-style-type: none"> KEEP_DELETED_FILES - Retains the destination files even when they no longer exist at the source. DELETE_TO_TRASH - Moves files to the trash folder if the HDFS trash is enabled. (Not supported when replicating to S3 or ADLS.) DELETE_PERMANENTLY - Uses the least amount of space; use with caution. <p>Default is KEEP_DELETED_FILES.</p>
alerts	<p>Configure the following parameters as required:</p> <ul style="list-style-type: none"> onFailure - Provide true to generate alerts when the replication job fails. onStart - Provide true to generate alerts when the replication job starts. onSuccess - Provide true to generate alerts when the replication job completes successfully. onAbort - Provide true to generate alerts when the replication job is aborted.
exclusionFilters	<p>Provide one or more directory paths to exclude from replication.</p>

Parameter	Description
frequencyInSec	Auto-populated after the policy runs successfully. Shows the time duration between two replication jobs in seconds.
targetDataset	Auto-populated after the policy runs successfully. Shows the target location where the replicated files are available on the target cluster.
cloudCredentials	Provide the cloud credentials.
sourceCluster	Shows the source cluster name.
targetCluster	Shows the target cluster name in the dataCProvideName\$cluster name format. For example, "DC-US\$My Destination 17".
startTime	Shows the start time of the replication job in the YYYY-MM-DDTHH:MM:SSZ format.
endTime	Shows the end time of the replication job in the YYYY-MM-DDTHH:MM:SSZ format.
distcpMaxMaps	Provide the maximum map slots to limit the number of map slots per mapper. Default is 20.
distcpMapBandwidth	Provide the maximum bandwidth to limit the bandwidth per mapper. Default is 100 MB.
queueName	Provide a YARN queue name, if necessary. Default queue name is Default.
tdeSameKey	Provide true if the source and destination are encrypted with the same TDE key.
description	Provide a description for the policy.
enableSnapshotBasedReplication	Provide true to enable snapshot-based replication.
cloudEncryptionAlgorithm	Provide the cloud encryption algorithm.
cloudEncryptionKey	Provide the cloud encryption key.
plugins	Provide the plugins to deploy on all the nodes in the cluster if you have multiple repositories configured in your environment.
cmPolicySubmitUser	Provide the following options: <ul style="list-style-type: none"> • userName - Provide the user name that you are using to run the policy. • sourceUser - Provide the source cluster username, if any.

Sample HDFS replication policy definition JSON file

The following snippet shows the contents of the HDFS replication policy definition JSON file.

```
{
  "name": "string",
  "type": "FS"|"HIVE",
  "sourceDataset": {
    "hdfsArguments": {
      "path": "string",
      "mapReduceService": "string",
      "logPath": "string",
      "replicationStrategy": "DYNAMIC"|"STATIC",
      "errorHandling": {
        "skipChecksumChecks": true|false,
        "skipListingChecksumChecks": true|false,
        "abortOnError": true|false,
        "abortOnSnapshotDiffFailures": true|false
      }
    }
  }
}
```

```

    "preserve": {
      "blockSize": true|false,
      "replicationCount": true|false
      "permissions": true|false,
      "extendedAttributes": true|false
    },
    "deletePolicy": "KEEP_DELETED_FILES" | "DELETE_TO_TRASH" | "DELETE_PERMANENTLY",
  },
  "alerts": {
    "onFailure": true|false,
    "onStart": true|false,
    "onSuccess": true|false,
    "onAbort": true|false
  },
  "exclusionFilters": ["string", ...]
},
"hiveArguments": {
  "databasesAndTables": [
    {
      "database": "string",
      "tablesIncludeRegex": "string",
      "tablesExcludeRegex": "string",
    }
  ],
  ...
},
"sentryPermissions": "INCLUDE" | "EXCLUDE",
"skipUrlPermissions": true|false,
"numThreads": integer
}
},
"frequencyInSec": integer,
"targetDataset": "string",
"cloudCredentials": "string",
"sourceCluster": "string",
"targetCluster": "string",
"startTime": "string",
"endTime": "string",
"distcpMaxMaps": integer,
"distcpMapBandwidth": integer,
"queueName": "string",
"tdeSameKey": true|false,
"description": "string",
"enableSnapshotBasedReplication": true|false
"cloudEncryptionAlgorithm": "string",
"cloudEncryptionKey": "string",
"plugins": ["string", ...],
"hiveExternalTableBaseDirectory": "string",
"cmPolicySubmitUser": {
  "userName": "string",
  "sourceUser": "string"
}
}
}

```



Important: When you edit the file, ensure that you remove the key-value pairs that are not required for the policy. For example, you can remove queueName if you do not want to configure it for the replication policy.

Managing HDFS replication policies using CDP CLI

You can use CDP CLI to perform various actions on a replication policy.

About this task

You can perform the following actions to manage an HDFS replication policy:

Procedure

- Suspend a running policy job:

```
cdp --profile [***PROFILE NAME***] replicationmanager suspend-policy --cluster-crn [***SOURCE CLUSTER CRN***] --policy-name [***POLICY NAME***]
```

- Activate a suspended policy job:

```
cdp --profile [***PROFILE NAME***] replicationmanager activate-policy --cluster-crn [***SOURCE CLUSTER CRN***] --policy-name [***POLICY NAME***]
```

- Download the diagnostic bundle for the specified replication policy:

- a) Initiate the diagnostic bundle collection operation for the specified replication policy on the target Cloudera Manager.

```
cdp --profile [***PROFILE NAME***] replicationmanager collect-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --policy-name [***POLICY NAME***]
```

The following sample snippet shows the command output:

```
{
  "commandId": 58747,
  "name": "Replication Diagnostics Collection",
  "active": true,
  "startTime": "2022-11-07T12:27:25.872Z"
}
```

- b) View the diagnostic bundle collection status. Optionally, you can use the following command to get the current status of any Cloudera Manager command.

```
cdp --profile [***PROFILE NAME***] replicationmanager get-command-status --cluster-crn [***TARGET CLUSTER CRN***] --policy-name [***POLICY NAME***]
```

The command returns the following responses:

- The diagnostic bundle collection is **IN PROGRESS** on the Cloudera Manager server.
- The diagnostic bundle collection is complete, and is **DOWNLOADABLE WITH URL** using the URL specified in the `resultDataUrl` field in the command output.
- The diagnostic bundle collection is complete, and is **DOWNLOADABLE WITH CLI** using the `download-diagnostic-bundle` CDP CLI operation in Step 3.
- The diagnostic bundle collection has **FAILED** on the Cloudera Manager server.



Tip: The `commandId` in the output is used in Step 3 to download the bundle.

The following sample snippet shows the command output when the `bundleStatus` is **DOWNLOADABLE WITH CLI**:

```
{
  "commandId": 58741,
  "success": true,
  "active": false,
  "name": "Replication Diagnostics Collection",
}
```

```
"resultDataUrl": "http://[***CM HOST***]:[***CM PORT***]/cmf/command/58741/download",
"resultMessage": "Replication diagnostics collection succeeded.",
"bundleStatus": "DOWNLOADABLE WITH CLI",
"bundleStatusMessage": "The bundle can be downloaded with the download-diagnostic-bundle operation."
}
```

- c) Run the following command only if the bundleStatus shows **DOWNLOADABLE WITH CLI** in the Step 2 command output. The command output appears as a binary string in base64 encoded format on the screen.

```
cdp --profile [***PROFILE NAME***] replicationmanager download-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --command-id [***COMMAND ID***]
```

You can use any method to parse the response. Alternatively, you can also use one of the following methods to parse the response:

Method

Save the diagnostic bundle in the specified file in JSON format and download to your machine

Command

```
cdp --profile [***PROFILE NAME***] replicationmanager download-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --command-id [***COMMAND ID***] > [***<FILE>.JSON***]
```

Save the diagnostic bundle to the specified ZIP file and download to your machine.

```
cdp --profile [***PROFILE NAME***] replicationmanager download-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --command-id [***COMMAND ID***] > [***<FILENAME>.JSON***] | jq -r '.bundleFile' | base64 -D > [***<FILENAME>.ZIP***]
```

Save the diagnostic bundle as a ZIP file and extract it to the specified location on your machine automatically.

```
cdp --profile [***PROFILE NAME***] replicationmanager download-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --command-id [***COMMAND ID***] > [***<FILENAME>.JSON***] | jq -r '.bundleFile' | base64 -D > | bsdtar -xvf -f [***LOCATION***]
```

- Delete the replication policy:

```
cdp --profile [***PROFILE NAME***] replicationmanager delete-policy --cluster-crn [***SOURCE CLUSTER CRN***] --policy-name [***POLICY NAME***]
```

Creating Hive replication policy using CDP CLI

You can use CDP CLI to create an Hive replication policy. Only a non-machine user can run the "replicationmanager create-policy" CDP CLI command to create a replication policy.

Procedure

1. Log into Replication Manager CDP CLI setup:

```
cdp --profile [***PROFILE NAME***] replicationmanager
```

2. List the clusters to verify whether the required clusters are available:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-clusters
```

3. Verify whether the required services are running on the source cluster:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-cluster-service
-statuses
```

4. Ensure that the cloud credentials are available:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-all-credentials
```

5. Perform the following steps to create a policy definition JSON file.

- a) Open a policy definition JSON file, or copy the output of the command to a JSON file to generate a policy definition JSON file.

```
cdp --profile [***PROFILE NAME***] replicationmanager create-policy --
generate-cli-skeleton
```

```
cdp --profile hive1 replicationmanager create-policy --generate-cli-
skeleton > rm_hive1.json
```

- b) Enter the required parameters. Remove the key-value pairs that are not required in the policy definition JSON file for the specific policy.
- c) Save the file.

6. Create the Hive replication policy.

```
cdp --profile [***PROFILE NAME***] replicationmanager create-policy --cli-
input-json [***POLICY DEFINITION JSON FILE PATH USING CAT***]
```

The cat command reads the data from the policy definition JSON file.

```
cdp --profile local-dev replicationmanager create-policy --cli-input-json "$(cat temp/rm_hive1.json)"
```

Results

Replication Manager creates the replication policy and initiates data replication.

What to do next

Verify whether the replication policy is running as expected on the Cloudera Replication Manager Replication Policies page, or run the following command:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-policies --cluste
r-crn [***CRN OF THE CLUSTER WHERE THE REPLICATION POLICIES ARE STORED***]
```

Related Information


[Creating Hive replication policy using Replication Manager](#)

Hive replication policy definition JSON file

The policy definition JSON file contains all the parameters required to create a Hive replication policy. When you edit the file to define a Hive replication policy, remove the parameters that are not required for the replication policy.

Parameters in Hive replication policy definition JSON file

The following table lists the parameters in the policy definition JSON file that are required for a Hive replication policy:

Parameter	Description
name	Provide the unique name for the policy.
type	Provide HIVE to create a Hive replication policy.
mapReduceService	Provide the MapReduce or YARN service for the replication policy to use.
logPath	Provide an alternate path for the logs, if required.
replicationStrategy	<p>Provide one of the following options to determine whether the file replication tasks must be distributed among the mappers statically or dynamically:</p> <ul style="list-style-type: none"> • STATIC - Static replication distributes file replication tasks among the mappers up front to achieve a uniform distribution based on the file sizes. • DYNAMIC - Dynamic replication distributes the file replication tasks in small sets to the mappers, and as each mapper completes its tasks, it dynamically acquires and processes the next unallocated set of tasks. <p>Default is DYNAMIC.</p>
skipChecksumChecks	<p>Provide true to skip checksum checks.</p> <p>Default is true.</p> <p>Checksums are used to perform the following tasks:</p> <ul style="list-style-type: none"> • To skip replication of files that have already been copied. When set to true, the replication job skips copying a file if the file lengths and modification times are identical between the source and destination clusters. Otherwise, the job copies the file from the source to the destination. • To redundantly verify the integrity of data. However, checksums are not required to guarantee accurate transfers between clusters. HDFS data transfers are protected by checksums during transfer and storage hardware also uses checksums to ensure that data is accurately stored. These two mechanisms work together to validate the integrity of the copied data.
skipListingChecksumChecks	<p>Provide true to skip checksum check while comparing two files to determine whether they are the same or not. Otherwise, the file size and last modified time are used to determine if files are the same or not. Skipping the check improves performance during the mapper phase.</p> <p> Note: If you set skipChecksumChecks to false, the skipListingChecksumChecks is also set to false by default.</p>
abortOnError	<p>Provide true to stop the policy job when an error occurs. This ensures that the files copied up to that point remain on the destination, but no additional files are copied.</p> <p>Default is false.</p>
abortOnSnapshotDiffFailures	Provide true to stop the replication job if a snapshot diff fails during replication.

Parameter	Description
preserve	<p>Provide true to preserve the block size, replication count, permissions (including ACLs), and extended attributes (XAttrs) as they exist on the source file system.</p> <ul style="list-style-type: none"> blockSize replicationCount permissions extendedAttributes <p>Provide false to use the settings as configured on the destination file system.</p> <p>By default, the source system settings are preserved.</p>
deletePolicy	<p>Provide one of the following options:</p> <ul style="list-style-type: none"> KEEP_DELETED_FILES - Retains the destination files even when they no longer exist at the source. DELETE_TO_TRASH - Moves files to the trash folder if the HDFS trash is enabled. (Not supported when replicating to S3 or ADLS.) DELETE_PERMANENTLY - Uses the least amount of space; use with caution. <p>Default is KEEP_DELETED_FILES.</p>
alert	<p>Configure the following parameters as required:</p> <ul style="list-style-type: none"> onFailure - Provide true to generate alerts when the replication job fails. onStart - Provide true to generate alerts when the replication job starts. onSuccess - Provide true to generate alerts when the replication job completes successfully. onAbort - Provide true to generate alerts when the replication job is aborted.
exclusionFilters	Provide one or more directory paths to exclude from replication.
databasesAndTables	<p>Configure the parameter as required:</p> <ul style="list-style-type: none"> database - Provide one or more database names to include from replication. tablesIncludeRegex - Provide one or more regular expression-based paths to include in replication. <p>For example, if you Provide</p> <pre>table1 table2 table3</pre> <p>, Replication Manager includes the specified tables for replication. If you Provide</p> <pre>DB :db_name Table : (?!table1 table2 table3) . +</pre> <p>, Replication Manager includes all the tables in the 'db_name' database and excludes 'table1', 'table2', and 'table3' from replication.</p> <p>tablesExcludeRegex is a legacy option. You can provide one or more regular expression-based paths of tables to exclude in replication.</p>
sentryPermissions	Provide INCLUDE to import both Hive object and URL permissions.
skipUrlPermissions	Provide true to import only the Hive object permissions.
numThreads	Provide the number of threads to use during replication.

Parameter	Description
frequencyInSec	Auto-populated after the policy runs successfully. Shows the time duration between two replication jobs in seconds.
targetDataset	Auto-populated after the policy runs successfully. Shows the target location where the replicated files are available on the target cluster.
cloudCredential	Provide the cloud credentials.
sourceCluster	Shows the source cluster name.
targetCluster	Shows the target cluster name in the dataCProvideName\$clustername format. For example, "DC-US\$My Destination 17".
startTime	Shows the start time of the replication job in the YYYY-MM-DDTHH:MM:SSZ format.
endTime	Shows the end time of the replication job in the YYYY-MM-DDTHH:MM:SSZ format.
distcpMaxMaps	Provide the maximum map slots to limit the number of map slots per mapper. Default is 20.
distcpMapBandwidth	Provide the maximum bandwidth to limit the bandwidth per mapper. Default is 100 MB.
queueName	Provide a YARN queue name, if necessary. Default queue name is Default.
tdeSameKey	Provide true if the source and destination are encrypted with the same TDE key.
description	Provide a description for the policy.
enableSnapshotBasedReplication	Provide true to enable snapshot-based replication.
cloudEncryptionAlgorithm	Provide the cloud encryption algorithm.
cloudEncryptionKey	Provide the cloud encryption key.
plugins	Provide the plugins to deploy on all the nodes in the cluster if you have multiple repositories configured in your environment.
hiveExternalTableBaseDirectory	Provide the Hive external table base directory path.
cmPolicySubmitUser	Provide the following options: <ul style="list-style-type: none"> • userName - Provide the user name that you are using to run the policy. • sourceUser - Provide the source cluster username, if any.

Sample Hive replication policy definition JSON file

The following snippet shows the contents of the Hive replication policy definition JSON file. While editing the file, ensure that you remove the key-value pairs that are not required for the Hive replication policy.

```
{
  "name": "string",
  "type": "HIVE",
  "sourceDataset": {
    "hdfsArguments": {
      "path": "string",
      "mapReduceService": "string",
      "logPath": "string",
      "replicationStrategy": "DYNAMIC" | "STATIC",
    },
    "errorHandling": {
      "skipChecksumChecks": true|false,
      "skipListingChecksumChecks": true|false,
    }
  }
}
```

```

    "abortOnError": true|false,
    "abortOnSnapshotDiffFailures": true|false
  },
  "preserve": {
    "blockSize": true|false,
    "replicationCount": true|false
    "permissions": true|false,
    "extendedAttributes": true|false
  },
  "deletePolicy": "KEEP_DELETED_FILES" | "DELETE_TO_TRASH" | "DELETE_PERMANENTLY",
  "alert": {
    "onFailure": true|false,
    "onStart": true|false,
    "onSuccess": true|false,
    "onAbort": true|false
  },
  "exclusionFilters": ["string", ...]
},
"hiveArguments": {
  "databasesAndTables": [
    {
      "database": "string",
      "tablesIncludeRegex": "string",
      "tablesExcludeRegex": "string",
    }
    ...
  ],
  "sentryPermissions": "INCLUDE" | "EXCLUDE",
  "skipUrlPermissions": true|false,
  "numThreads": integer
}
},
"frequencyInSec": integer,
"targetDataset": "string",
"cloudCredential": "string",
"sourceCluster": "string",
"targetCluster": "string",
"startTime": "string",
"endTime": "string",
"distcpMaxMaps": integer,
"distcpMapBandwidth": integer,
"queueName": "string",
"tdeSameKey": true|false,
"description": "string",
"enableSnapshotBasedReplication": true|false
"cloudEncryptionAlgorithm": "string",
"cloudEncryptionKey": "string",
"plugins": ["string", ...],
"hiveExternalTableBaseDirectory": "string",
"cmPolicySubmitUser": {
  "userName": "string",
  "sourceUser": "string"
}
}
}

```



Important: When you edit the file, ensure that you remove the key-value pairs that are not required for the policy.

Managing Hive replication policies using CDP CLI

You can use CDP CLI to perform various actions on a replication policy. You can suspend a running Hive replication policy job and then activate it. You can also delete a replication policy.

About this task

You can perform the following actions to manage a Hive replication policy:

Procedure

- Suspend a running policy job:

```
cdp --profile [***PROFILE NAME***] replicationmanager suspend-policy --cluster-crn [***SOURCE CLUSTER CRN***] --policy-name [***PROFILE NAME***]
```

- Activate a suspended policy job:

```
cdp --profile [***PROFILE NAME***] replicationmanager activate-policy --cluster-crn [***TARGET CLUSTER CRN***] --policy-name [***PROFILE NAME***]
```

- Download the diagnostic bundle for the specified replication policy:

- a) Initiate the diagnostic bundle collection operation for the specified replication policy on the target Cloudera Manager.

```
cdp --profile [***PROFILE NAME***] replicationmanager collect-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --policy-name [***POLICY NAME***]
```

The following sample snippet shows the command output:

```
{
  "commandId": 58747,
  "name": "Replication Diagnostics Collection",
  "active": true,
  "startTime": "2022-11-07T12:27:25.872Z"
}
```

- b) View the diagnostic bundle collection status. Optionally, you can use the following command to get the current status of any Cloudera Manager command.

```
cdp --profile [***PROFILE NAME***] replicationmanager get-command-status --cluster-crn [***TARGET CLUSTER CRN***] --policy-name [***POLICY NAME***]
```

The command returns the following responses:

- The diagnostic bundle collection is **IN PROGRESS** on the Cloudera Manager server.
- The diagnostic bundle collection is complete, and is **DOWNLOADABLE WITH URL** using the URL specified in the resultDataUrl field in the command output.
- The diagnostic bundle collection is complete, and is **DOWNLOADABLE WITH CLI** using the download-diagnostic-bundle CDP CLI operation in Step 3.
- The diagnostic bundle collection has **FAILED** on the Cloudera Manager server.



Tip: The commandId in the output is used in Step 3 to download the bundle.

The following sample snippet shows the command output when the bundleStatus is **DOWNLOADABLE WITH CLI**:

```
{
  "commandId": 58741,
  "success": true,
  "active": false,
  "name": "Replication Diagnostics Collection",
  "resultDataUrl": "http://[***CM HOST***]:[***CM PORT***]/cmf/command/58741/download",
  "resultMessage": "Replication diagnostics collection succeeded.",
  "bundleStatus": "DOWNLOADABLE WITH CLI",
  "bundleStatusMessage": "The bundle can be downloaded with the download-diagnostic-bundle operation."
}
```

- c) Run the following command only if the bundleStatus shows **DOWNLOADABLE WITH CLI** in the Step 2 command output. The command output appears as a binary string in base64 encoded format on the screen.

```
cdp --profile [***PROFILE NAME***] replicationmanager download-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --command-id [***COMMAND ID***]
```

You can use any method to parse the response. Alternatively, you can also use one of the following methods to parse the response:

Method

Save the diagnostic bundle in the specified file in JSON format and download to your machine

Command

```
cdp --profile [***PROFILE NAME***] replicationmanager download-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --command-id [***COMMAND ID***] > [***<FILE>.JSON***]
```

Save the diagnostic bundle to the specified ZIP file and download to your machine.

```
cdp --profile [***PROFILE NAME***] replicationmanager download-diagnostic-bundle --cluster-crn [***TARGET CLUSTER CRN***] --command-id [***COMMAND ID***] > [***<FILENAME>.JSON***] | j
```

Method

Save the diagnostic bundle as a ZIP file and extract it to the specified location on your machine automatically.

Command

```
q -r '.bundleFile' | base64 -D
> [***<FILENAME>.ZIP***]
```

```
cdp --profile [***PROFILE NAME***]
  replicationmanager download-
  diagnostic-bundle --cluster-crn
  [***TARGET CLUSTER CRN***] --
  command-id [***COMMAND ID***]
  > [***<FILENAME>.JSON***] | jq -
  r '.bundleFile' | base64 -D >| bs
  dtar -xvf -f [***LOCATION***]
```

- Delete the replication policy:

```
cdp --profile [***PROFILE NAME***] replicationmanager delete-policy --clus
ter-crn [***SOURCE CLUSTER CRN***] --policy-name [***POLICY NAME***]
```

Creating HBase replication policy using CDP CLI

You can use CDP CLI commands to create an HBase replication policy.

Before you begin

Ensure that the prerequisites are complete before you create an HBase replication policy. For more information about the prerequisites, see [Using HBase replication policies](#).

Procedure

1. Log into Replication Manager CDP CLI setup:

```
cdp --profile [***PROFILE NAME***] replicationmanager
```

2. List the clusters to verify whether the required clusters are available:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-clusters
```

3. Verify whether the required services are running on the required clusters:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-cluster-service
-statuses
```

4. Ensure that the cloud credentials are available:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-all-credentials
```


5. Verify whether the source cluster and target cluster are paired:

```
cdp --profile [***PROFILE NAME***] replicationmanager verify-hbase-cluster-pair
```

The command returns the following cluster pairing status:

pairingStatus	Description
PAIRED	<p>Indicates that the clusters are paired and Replication Manager attempts to initiate the first-time setup between the clusters. The following status appear during the pairing process:</p> <ul style="list-style-type: none"> ERROR - Indicates that the policy creation process cannot be continued unless the error is fixed and the pairing is reset. UNDER_SETUP - Indicates that you must wait for the process to complete before you continue to create the replication policy. RESETTABLE - Indicates that the pairing process completed successfully and that there are no existing replication policies using this pairing. You can continue to create the HBase replication policy. NOT_RESETTABLE - Indicates that the pairing process completed successfully and there are existing policies using this pairing. You can continue to create the HBase replication policy.
CAN_BE_FORCED	<p>Indicates that one or both of the specified clusters are part of other pairings that are in ERROR or RESETTABLE state.</p> <p>You can set up the pairing between the clusters by setting the force* parameter to true. This action removes the existing pairings.</p>
WRONG_PAIRING	<p>Indicates that one or both of the specified clusters are part of other pairings that are in NOT_RESETTABLE state.</p> <p>In this scenario, you cannot continue to create HBase replication policies.</p>
ON_HOLD	<p>Indicates that one or both of the specified clusters are part of other pairings that are in UNDER_SETUP state.</p> <p>In this scenario, you can retry the pairing process after some time.</p>
<p>*The option is a technical preview feature and is not ready for production deployment. The components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the use of preview components, which should be used by customers at their own risk. For more information, contact your Cloudera account team.</p>	



Tip: View the previously paired clusters:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-paired-hbase-clusters
```

6. Create a policy definition JSON file.

- Open a policy definition JSON file, or copy the output of the following command to a JSON file to generate a policy definition JSON file:

```
cdp --profile [***PROFILE NAME***] replicationmanager create-hbase-policy --generate-cli-skeleton
```

```
cdp --profile hbase1 replicationmanager create-hbase-policy --generate-cli-skeleton > rm_hbase1.json
```

- Enter the required parameters. Remove the key-value pairs that are not required in the policy definition JSON file for the specific policy.
- Save the file.

7. Create the HBase replication policy:

```
cdp --profile [***PROFILE NAME***] replicationmanager create-hbase-policy
--cli-input-json [***POLICY DEFINITION JSON***]
```

Optionally, you can use the cat command to read the data from the policy definition JSON file. For example, cdp --profile local-dev replicationmanager create-policy --cli-input-json “\$(cat temp/rm_hbase1.json)”



Important: If the source cluster is a classic cluster and the clusters have not been paired before, manual intervention is required for the initial setup.

Perform the following steps after you create the HBase replication policy to complete the first-time setup between the clusters:

a. View the policy status:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-policies | less
```

b. Restart the HBase service on the source Cloudera Manager when the policy status shows “WAITING _RESTART_ON_SRC” or WAITING_ON_SRC_RESTARTING_ON_DEST.

c. Ensure that the HBase service restart is complete.

d. Continue the first time setup process when the policy status shows WAITING_FOR_CONTINUE _SETUP_CALL:

```
cdp --profile [***PROFILE NAME***] replicationmanager continue
-hbase-setup --cluster-crn [***TARGET CLUSTER CRN***] --policy-
id [***POLICY ID***]
```

Results

Replication Manager creates the replication policy and initiates data replication.

What to do next

Verify whether the replication policy is running as expected on the Cloudera Replication Manager Replication Policies page, or run the following command:

```
cdp --profile [***PROFILE NAME***] replicationmanager list-policies --cluster-crn [***CRN OF THE CLUSTER WHERE THE REPLICATION POLICIES ARE STORED***]
```



HBase replication policy definition JSON file




The HBase replication policy definition JSON file contains all the parameters required to create an HBase replication policy.

Parameters in HBase replication policy definition JSON file

The following table lists the parameters in the policy definition JSON file that are required to create an HBase replication policy:

Parameter	Description	Required?
clusterCrn	Provide the target cluster CRN. Replication Manager saves the replication policy in the specified cluster CRN.	Required
policyName	Provide a unique name for the replication policy.	Required

Parameter	Description	Required?
policyDefinition	Provide the policy definition parameters as required.	Required
hbasePolicyArguments	Provide the HBase replication parameters.	Required
tables	<p>Provide the tables to be replicated where the key must be in the "namespace:tablename" format.</p> <p> Note: If the tables parameter has no value, Replication Manager replicates the entire database.</p>	Optional
credentialLocation*	<p>Provide SAFETY_VALVE if the credentials are available in the advanced configuration snippets (safety valves) on the source cluster, or provide EXTERNAL_ACCOUNT if the credentials are to be defined in the cloudCredential property.</p> <p>Default is EXTERNAL_ACCOUNT.</p>	Optional
cloudCredential	<p>Provide the cloud credentials to use to replicate the initial snapshot if the credentials are defined in an external account.</p> <p> Note: The cloudCredential parameter is mandatory if initialSnapshot=true and the cloud credentials are not available in the advanced configuration snippets.</p>	Optional
validateReplicationSetup*	Provide true to validate the replication setup after policy creation. Otherwise, provide false.	Optional
forceSetup*	<p>Provide true to force the first-time setup when one of the clusters is already paired with another cluster.</p> <p>Forced setup is only possible if there are no HBase replication policies between the pair of clusters or if the other cluster in the pair is currently unreachable.</p> <p>During the force setup, Replication Manager clears the existing pairing for the selected source or target cluster and initiates the first-time setup with the chosen new source or destination cluster.</p>	Optional
databaseArguments	<p>Provide one of the following values for the replicationStrategy parameter:</p> <ul style="list-style-type: none"> ALL_TABLES TABLES_WITH_REPLICATION_SCOPE_SET <p>For more information about replication scope, see Creating HBase replication policy.</p>	Optional
sourceCluster	Provide the name of the source cluster in the "dataCenterName\$cluster_name" format. For example, "DC-Europe\$My Source 42"	Required
targetCluster	Provide the name of the target cluster in the "dataCenterName\$cluster_name" format. For example, "DC-US\$My Destination 17"	Required
initialSnapshot	Provide true to replicate the existing data in the table. When you provide false, Replication Manager replicates only the data generated after the policy creation.	Required

Parameter	Description	Required?
description	Provide a brief description of the replication policy.	Optional
machineUser	<p>The credentials of the machine user that Replication Manager uses to run HBase replication policies.</p> <p>Provide the following parameters for the machine user:</p> <ul style="list-style-type: none"> • user • password • createUser - Provide true to create a new machine user. If you provide false, ensure that the username you provide exists in the Cloudera User Management System (UMS), otherwise an error message appears. <p>If you do not provide the machine user details, an HBase replication machine user is created automatically with an auto-generated username.</p>	Optional
queueName	<p>Provide a YARN queue name to use for the initial snapshot operation.</p> <p>Default is default.</p>	Optional
distcpMaxMaps	Provide the maximum map jobs to use for initial snapshot operation.	Optional
distcpMapBandwidth*	<p>Adjust the setting so that each map task is throttled to consume only the specified bandwidth.</p> <p>Default is 100 MB.</p>	Optional
sourceRestartType	<p>Provide RESTART or ROLLING_RESTART to restart the HBase service on the source cluster.</p> <p> Note: The sourceRestartType parameter is accepted only when the source cluster is not a classic cluster and the destination cluster has the restart-type support.</p> <p>Default is ROLLING_RESTART.</p> <p> Tip: During rolling restart, one node is restarted at a time and this continues until all the nodes in the cluster are restarted. This type of restart ensures that there is no disruption of service. During full restart, all the nodes are shut down at once and restarted simultaneously.</p>	Optional
targetRestartType	<p>Provide RESTART or ROLLING_RESTART to restart the HBase service on the destination cluster.</p> <p> Note: The targetRestartType parameter is accepted only if the cluster has restart-type support.</p> <p>Default is ROLLING_RESTART.</p>	Optional
<p>*The option is a technical preview feature and is not ready for production deployment. The components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the use of preview components, which should be used by customers at their own risk. For more information, contact your Cloudera account team.</p>		

Sample HBase replication policy definition JSON file

The following snippet shows the contents of an HBase replication policy definition JSON file:

```
{
  "clusterCrn": "string",
  "policyName": "string",
  "policyDefinition": {
    "hbasePolicyArguments": {
      "tables": [
        "string"
      ],
      "credentialLocation": "EXTERNAL_ACCOUNT"|"SAFETY_VALVE", (technical preview)
      "cloudCredential": "string",
      "validateReplicationSetup": true|false, (technical preview only)
      "forceSetup": true|false, (technical preview only)
      "databaseArguments": {
        "replicationStrategy": "ALL_TABLES"|"TABLES_WITH_REPLICATION_SCOPE_SET"
      }
    },
    "sourceCluster": "string",
    "targetCluster": "string",
    "initialSnapshot": true|false,
    "description": "string",
    "machineUser": {
      "user": "string",
      "password": "string",
      "createUser": true|false
    },
    "queueName": "string",
    "distcpMaxMaps": 0,
    "distcpMapBandwidth": 0, (technical preview only)
    "sourceRestartType": "RESTART"|"ROLLING_RESTART",
    "targetRestartType": "RESTART"|"ROLLING_RESTART"
  }
}
```



Important: When you edit the file, ensure that you remove the key-value pairs that are not required for the policy. For example, you can remove queueName if you do not want to configure it for the replication policy.

Managing HBase replication policies using CDP CLI

You can use CDP CLI commands to manage HBase replication policies.

Procedure

- Pause an active HBase replication policy:

```
cdp --profile [***PROFILE NAME***] replicationmanager suspend-hbase-policy
--cluster-crn [***TARGET CLUSTER CRN***] --policy-id [***POLICY ID***]
```

- Change the name and description, and delete one or more tables in the HBase replication policy:

```
cdp --profile [***PROFILE NAME***] replicationmanager update-hbase-policy
--cluster-crn [***TARGET CLUSTER CRN***] --policy-id [***POLICY ID***]
--update-hbase-policy-definition name=[***STRING***],description=[***STRING***],tables=[***STRING,STRING...***]
```

- Delete an HBase replication policy:

```
cdp --profile [***PROFILE NAME***] replicationmanager delete-hbase-policy  
--cluster-crn [***TARGET CLUSTER CRN***] --policy-id [***POLICY ID***] [--  
force] [--no-force]
```

Default is --no-force.

When the source cluster is not reachable, the error HBase peer not found appears. In this instance, you can append the --force option to delete the HBase replication policy.