

Sidecar migration from HDP to CDP

Date published: 2022-02-14

Date modified:



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

Sidecar migration from HDP to CDP.....	4
Process Overview.....	4
Pre-migration steps.....	5
Planning.....	5
Preparing your hosts.....	5
Replicate configurations.....	5
Set up TLS.....	6
Set up Kerberos.....	6
KMS ACL export.....	6
Migrating Services.....	6
Atlas Sidecar migration from HDP to CDP.....	6
Ranger Sidecar migration from HDP to CDP.....	6
Migrating Ranger Policies from HDP.....	6
HDFS Sidecar Migration to CDP.....	7
YARN Capacity Scheduler Sidecar migration from HDP to CDP.....	7
HBase Sidecar migration from HDP to CDP.....	8
Solr Sidecar migration from HDP to CDP.....	8
Oozie Sidecar migration to CDP.....	8
Accumulo Sidecar migration to OpDB.....	9
Post-migration steps.....	10
HDF Sidecar migration from HDF to CDP.....	10
Test the replicated data.....	10

Sidecar migration from HDP to CDP

Sidecar migrations allow you to migrate your HDP deployment to CDP on new hardware,

“Side-car” migration allows you to migrate an existing HDP deployment to a new cluster deployed on new (or different) hardware running the latest version of CDP Private Cloud Base. Migrations completed using this process have several advantages, including minimizing downtime during the migration process. You also will have the ability to test your workloads at production scale on the new cluster while keeping your production system running until you determine that the new cluster is ready.

Cloudera provides a variety of tools and documentation to help you move your data, service and cluster configurations, security configurations and policies to the new cluster. The exact procedures to follow may depend on the mix of services deployed on your cluster. This document can help you to learn how to plan and sequence your upgrade. You may need to engage Cloudera Professional Services to plan and execute some parts of the upgrade that could require special procedures not covered in this document.

This document will also provide you with links to useful information about what you need to know about changes in Cloudera Runtime components and how those might affect your workloads. Cloudera recommends that you review those documents closely to avoid unexpected changes in behavior that may delay your migration.

This method of upgrading your cluster has several advantages:

- Service-level agreements for individual workloads are easier to meet because the legacy versions and the CDP versions can run in parallel without impacting each other.
- Single, well-contained tenants can move one at a time without requiring a single coordinated event across all tenants.
- Rollback requires coordination only at the workload or tenant level rather than at the whole cluster level.
- Enables moving directly to CDP from any HDP 2 or HDP 3 version.
- Sidecar migration provides the ability to test your workloads at production scale on the new cluster without impacting your live production environment.

You can also use a “rolling side-car migration”, where you decommission some of the existing cluster hosts, install CDP on the decommissioned hosts, and then migrate workloads one at a time. As workloads are migrated, you again decommission more hosts and repeat the process.

For an overview of other upgrade and migration options, see [The Four Upgrade and Migration Paths to CDP from Legacy Distributions](#).

Process Overview

A side-car migration consists of the following phases, which are discussed in detail in other topics.

1. Pre-migration and planning

In this phase, you analyze your workloads and datasets to determine the type and quantity of hosts required for the migration. You also need to consider changes and upgrades to various components used in the source cluster. Some components are no longer available in CDP and you will need to understand how to migrate to services that provide similar functionality.

After determining the specifications for the destination cluster, you can begin to provision your hosts with a supported operating system and install the CDP Private Cloud Base software, which includes Cloudera Manager and Cloudera Runtime.

2. Migration

In this phase, you migrate your data and workloads to the destination cluster and begin testing for functionality and performance. After migrating your data, you can configure the replication process to use “snapshots” so that you can keep the data on the destination cluster synchronized with the source cluster as you complete your workload migration and testing.

If the cluster is using HDFS Transparent Data Encryption, you may also need to migrate KMS ACLs to the new CDP Cluster.

3. Post-migration

Some components require additional steps that must be performed after the data and workload migration are completed. Review the sections for your components in this publication to learn more about these requirements. When these steps are complete and you have completed your testing, you can move your new cluster(s) into production and de-commission the workloads running on the source cluster.

Pre-migration steps

There are steps to take before you begin a Sidecar migration.

Planning

To plan your migration, you need to consider a variety of items.

- When you provision the new CDP deployment, you should consider the type and capacity of the hosts you deploy in your clusters. Ensure that hardware is sufficient in speed and capacity to handle your current and planned workloads.
- Depending on the software running on the source cluster, not all services deployed in HDP clusters have exact matches in CDP. Consider how your services will map to updated versions of those services or to new services designed to replace the legacy software. Configurations and performance tuning may need to be adjusted.
- Plan for any updates that may be required in your application code. Newer versions may have updated or deprecated APIs and replacement services could require you to modify or re-architect your applications to use the correct APIs.

Preparing your hosts

Procedure

1. Install a supported operating system on all hosts. To learn about supported operating systems, see the [Cloudera Support Matrix](#).
2. Install Cloudera Manager and Cloudera Runtime on the hosts. See the [CDP Private Cloud Base Installation Guide](#)

Replicate configurations

When you create your new CDP Private Cloud Base cluster(s) you may need to copy some of your custom/tuning configuration from the source cluster to the new cluster. These can include properties like thresholds, memory limits, ports numbers, directory locations, log file locations, etc. These configurations exist for services, clusters, and hosts. Note that many CDP services have been updated from earlier versions and there may not be a one-to-one correspondence of configuration parameters from the old service to the new service. You will need to ensure that you

understand the new set of configuration options in CDP and that your configurations are valid and appropriate for your workloads and environment in which the new cluster is deployed..

Use Ambari to view configurations for each service as a guide for configuring the services in the new cluster, using Cloudera Manager.

Set up TLS

See:

- [Configuring TLS Encryption for Cloudera Manager Using Auto-TLS](#)
- [Manually Configuring TLS Encryption for Cloudera Manager](#)

Set up Kerberos

See [Configuring Authentication in Cloudera Manager](#)

Important: If you enable Apache Ranger on the cluster, you must configure Kerberos to enforce strong authentication of users. Running Apache Ranger on a non-Kerberized CDP environment is not supported.

KMS ACL export

Coming soon.

Migrating Services

The steps required in this phase of a Sidecar migration can vary depending on the components deployed in the source cluster and planned for the destination cluster. In some cases, the order in which you complete these migrations should be considered due to dependencies among some components.

Atlas Sidecar migration from HDP to CDP

You can manually migrate Atlas from HDP to CDP.

You must manually migrate tags and metadata configurations to the destination cluster. Apache Atlas will start to collect lineage information in the destination cluster. Lineage for the source cluster remains only in the Atlas instance deployed in the source cluster.

Cloudera Professional Services can assist you with planning your Atlas migration.

Ranger Sidecar migration from HDP to CDP

You can migrate Ranger policies from HDP to CDP.

Migrating Ranger Policies from HDP

Procedure

1. Open the Ranger UI on the source cluster
2. Select Access Manager Resource-based.

3. Click Export.
The JSON file is downloaded.
4. Choose Access ManagerTag-based
5. Click Export.
The JSON file is downloaded.
6. Copy the JSON files to the Ranger server host on the destination cluster.
7. Click Import.
8. Navigate to each exported JSON file and click Import.

HDFS Sidecar Migration to CDP

You can migrate data stored in HDFS to CDP.

To migrate data stored in HDFS from HDP to CDP, use the distcp command-line tool. See "[HDFS Replication](#)". See:

- (HDP 2 to CDP) [Using DistCp to Copy Files](#).
- (HDP 3 to CDP) [Using DistCp to Copy Files](#).

YARN Capacity Scheduler Sidecar migration from HDP to CDP

To migrate YARN from a HDP cluster you must migrate the Capacity Scheduler to CDP. The migration includes converting Capacity Scheduler placement rules to the new JSON format.

Procedure

To migrate the Capacity scheduler from HDP to CDP:

1. Get and copy the capacity-scheduler.xml configuration file from Ambari.
2. Log in to Cloudera Manager in the destination cluster and ensure that the YARN Queue Manager is not started and that the YARN service is installed and running.
3. Go to the YARN service.
4. Click the Configurations tab.
5. Search for the following property: Capacity Scheduler Configuration Advanced Configuration Snippet (Safety Valve)
6. Copy and paste the contents of the capacity-scheduler.xml configuration file you copied from Ambari into the text box.
7. Restart YARN. (Click Actions > Restart)
8. Open the YARN Queue Manager UI (Click Clusters > [Cluster Name] > YARN Queue Manager UI)
9. Queue Manager reads the configuration from the safety-valve and if it detects that the placement rules are in the format used by the source cluster, it automatically converts it into the new JSON-based placement rules format (this format was introduced in CDP Private Cloud Base 7.1.6).
10. Verify that Queue Manager correctly converted the configuration. You can test the conversions by going to the Placement Rules tab and changing the value of a property temporarily. If you see an "out of sync" error message, then the configurations are not in sync.

Results

From now on, Queue Manager can be used for configuration.



Note:

The `capacity-scheduler.xml` Advanced Configuration Snippet works differently than the usual Advanced Configuration Snippet. Normally, Advanced Configuration Snippets are used as a powerful tool to edit configuration. In the case of the YARN Queue Manager UI, Queue Manager continuously writes into the Advanced Configuration snippet as the user changes the configuration via the YARN Queue Manager UI. Therefore the configuration set in the Advanced Configuration Snippet cannot differ from the Queue Manager configuration. If you experience out-of-sync errors, restart the YARN and Queue Manager services, which will override the Advanced Configuration Snippet with the latest settings applied in Queue Manager.

HBase Sidecar migration from HDP to CDP

You can use HBase replication and snapshots to migrate HBase tables to the destination cluster.

1. Prepare for the migration. Learn about changes in HBase and what actions you need to take before migrating your data. See [Preparing for data migration](#).
2. Use the HBase shell on the source cluster to take a snapshot of all tables you want to migrate to the destination cluster and then export the snapshots to the destination. See: [Using Snapshots \(HDP 2\)](#) or [Using Snapshots in HBase \(HDP 3\)](#)
3. Export the snapshots to the destination cluster. See: [Using Snapshots \(HDP 2\)](#) [Using Snapshots in HBase \(HDP 3\)](#)
4. Import the snapshot on the destination cluster by running the following command in the HBase shell:

```
restore_snapshot <snapshot name>
```

5. If you are migrating from HDP 2, you may need to update your client application code. See [Known Incompatibilities when Upgrading from CDH and HDP](#).
6. Review the following topics for details for additional incompatibilities with HDP 2:
 - [Remove PREFIX_TREE Data Block Encoding](#)
 - [Validate HFiles](#)
 - [Check co-processor classes](#)
7. Verify and validate the migrated data on the destination cluster. See [Verifying and validating if your data is migrated](#)
8. Test your applications or queries.
9. When you have verified and tested your application code and are ready to move the applications to production, repeat the above steps to synchronize the HBase data with the source cluster.

You can use the Hash Table and Sync Table tool to synchronize table data. This tool can help improve performance when synchronizing large tables. See [Use HashTable and SyncTable Tool](#).

Solr Sidecar migration from HDP to CDP

To migrate Solr from HDP to CDP, you copy the data to the destination cluster and then re-index the data.

About this task

Indexes used by Solr HDP are not compatible with CDP. To migrate Solr to CDP:

Procedure

1. Copy all of the data you are indexing from the source to the destination cluster. See other sections in this guide for information about copying data.
2. After copying, you will need to re-index your collections. See [Reindexing Solr collections](#).

Oozie Sidecar migration to CDP

To migrate Oozie to CDP, you must manually migrate workloads and configurations to the destination cluster.

The Oozie database on the source cluster is automatically configured to purge data that is older than 30 days. If you need to retain this data, Cloudera recommends that you back up the data and make it available for running simple SQL queries. See [Back Up the Databases](#).

Ensure that you have set up a new Oozie database on the destination cluster. (You may have already done this when you installed Cloudera Manager on the destination cluster.)

Most Oozie workloads and configurations will need to be migrated manually to the destination cluster. Although you can copy the XML and property files from the source cluster, these files will require extensive modifications before they can be used in the destination cluster. Cloudera Professional Services can assist you with these migrations.

In general, you will need to ensure that the following settings are correct for the destination cluster:

- Name-node settings for each workflow.
- Resource-manager settings for each workflow.
- Paths, if the locations are being changed
- JDBC URLs, including Hive must point to the new database,
- DFS paths may need to be replaced with S3, Azure, as appropriate
- New cluster-specific configurations for the Workflow (for example, Amazon S3 credentials.)
- Coordinator Jobs that have future sdf runs scheduled. Recreate these jobs to match what was included in HDP. For such Jobs, the Workflow and its configuration might be stored in Hue.
- Copy any custom sharelib jars from the source to the destination cluster. The jars are stored here: `/user/oozie/share/lib/lib_{TIMESTAMP}/{COMPONENT}`
- If you have custom application jars used by any of your workflows, please make sure you recompile them with the 3rd-party libraries used by the latest Cloudera runtime. For example on older versions of HDP, Oozie was using Guava 11, but in CDP 7 it is using Guava 28 or higher. If you compiled your application with Guava 11 you must also upgrade Guava.
- On the destination cluster, update the sharelib using the following command:

```
oozie admin -oozie {URL} -sharelibupdate
```

Accumulo Sidecar migration to OpDB

You can migrate the Accumulo service to Operational Database powered by Apache Accumulo (OpDB) on CDP

Before you begin

Follow these steps to migrate the Accumulo service's configuration and data to Operational Database powered by Apache Accumulo (OpDB).

Procedure

1. Download and install the Operational Database powered by Apache Accumulo CSD and install it on the destination cluster. See [Install Operational Database powered by Apache Accumulo CSD file](#).
2. Ensure that you have migrated the configurations for Apache Accumulo from the source cluster to the OpDB service on the destination cluster.
3. Migrate Accumulo data.
 - a) Export each table on the source cluster.
 - b) Collect the exported table data for transfer.
 - c) Move the collected data into the HDFS storage on the destination cluster.
 - d) Install and configure the Accumulo on CDP service on the destination cluster.
 - e) Import each table into the destination cluster.

For more information, see [Exporting Tables](#) in Apache Accumulo documentation.

Post-migration steps

After you configure your destination cluster and migrate your data and workloads, there may be additional steps required.

HDF Sidecar migration from HDF to CDP

You can migrate Schema Registry, Streams Messaging Manager and Kafka to CDP.

Procedure

1. Ensure that you have migrated Ranger Policies to the destination cluster. See [Migrating Ranger Policies from HDP](#) on page 6.
2. [Migrate Schema Registry](#).
Overview of the methods that can be used to migrate Schema Registry from HDF to CDP Private Cloud Base.
3. [Migrate Streams Messaging Manager](#)
Overview of the methods that can be used to migrate Streams Messaging Manager from HDF to CDP Private Cloud Base.
4. [Migrate Kafka Using Streams Replication Manager](#)
Overview of the methods that can be used to migrate Kafka from HDF to CDP Private Cloud Base using Streams Replication Manager.

Test the replicated data

After your data, configuration, and workloads are migrated to the destination cluster, you should thoroughly test your applications before moving the cluster to production.

Coming soon