Cloudera Public Cloud

CDP Public Cloud Upgrade Advisor

Date published: 2019-08-22
Date modified: 2025-08-18

https://docs.cloudera.com/

Legal Notice

Contents

Getting started with upgrades for Cloudera on cloud.....5
- FAQ for Cloudera on cloud upgrades.....5
- Supported Cloudera Runtime versions for Cloudera on cloud.....9

Preparing for an upgrade.....9

Upgrading to Cloudera Runtime 7.3.1.....14
- Identify cluster version details.....15
- Supported Cloudera Runtime 7.3.1 upgrade paths.....18
- Review the prerequisites.....20
- Upgrade Data Lakes and all Cloudera Data Hub clusters to the latest 7.3.1 version.....20
  - Upgrade database to PostgreSQL 14.....21
  - Changing the default Java version to 8.....21
  - Upgrade Data Lake Cloudera Runtime version to 7.3.1 and OS to RHEL 8.10.....22
  - Upgrade Cloudera Runtime version of Cloudera Data Hub clusters to 7.3.1.....23
- Upgrade Data Lakes and all Cloudera Data Hub clusters to a specific 7.3.1 patch version with CDP CLI.....24
  - Changing the default Java version to 8.....24
  - Upgrading a Data Lake manually via CLI.....25
  - Upgrading a Cloudera Data Hub cluster with the CDP CLI.....28
- Upgrade to a version that enables 7.3.1 upgrade.....30
  - Resize to Enterprise Data Lake.....30
  - Changing the default Java version to 8.....31
  - Upgrade Data Lake Cloudera Runtime version to 7.3.1 and OS to RHEL 8.10.....32
  - Upgrade Cloudera Runtime version of Cloudera Data Hub clusters to 7.3.1.....32
  - Upgrade OS version of Cloudera Data Hub clusters to RHEL 8.10.....33

Upgrading to Cloudera Runtime 7.2.18.....33
- Identify cluster version details.....34
- Identify your upgrade path.....37
- Review the prerequisites.....39
- High-level upgrade steps.....39
  - Step 1: Upgrade Data Lake and all Cloudera Data Hub clusters to the latest service pack and OS.....39
  - Step 2: Upgrade FreeIPA to RHEL 8.....41
  - Step 3: Upgrade Data Lake and Cloudera Data Hub clusters to the latest Cloudera Runtime 7.2.17 and the latest RHEL 8 OS image.....41
  - Step 4: Resize Medium Duty Data Lake to Enterprise Data Lake.....43
  - Step 5: Upgrade Data Lake and Cloudera Data Hub clusters from 7.2.17 to 7.2.18.....44

Upgrading from CentOS to RHEL.....45
- Prerequisites for upgrading from CentOS to RHEL.....45
- Upgrade clusters to RHEL.....46

Upgrading from Medium Duty to Enterprise Data Lake.....47

Upgrading from Spark 2 to Spark 3..... 48

Troubleshooting upgrade operations..... 48

Data Lake upgrade - out of memory error..... 48

Data Lake upgrade - Kafka consumer not available yet..... 49

Getting started with upgrades for Cloudera on cloud

Compute resources deployed via cloud services are generally considered transient in nature. With the separation of compute and storage, compute clusters in Cloudera can be stopped or decommissioned, while the data used or created by workloads running on these clusters generally remains accessible on persistent cloud storage.

There are some exceptions to the above, most importantly SDX metadata and cloud service-specific metadata. SDX metadata is stored in the Data Lake, while metadata specific to a cloud service may be stored in databases, local configurations, or even locally attached disks. This local storage can also be persistent (block storage volumes) or transient (ephemeral disks).

In cloud services where compute is elastic and state is transient, we need safeguards to protect all data that is not persistent, especially when changes are performed on the services themselves. In Cloudera on cloud, there can be several changes, most notably Runtime service pack upgrades, Runtime minor/major software version upgrades, and OS upgrades.

In general, there are two main approaches to upgrading cloud services:

1. Backup, re-create and restore
2. In-place upgrade

These two approaches have similarities: a prior backup should be performed, service endpoints should remain stable after the operation, and they should result in the same outcome. Cloudera on cloud supports both approaches. While the first approach may be convenient for simple data workloads, complex data applications and their custom data pipelines spanning multiple clusters may require an in-place upgrade path.

In this guide we will describe the high-level steps of performing in-place upgrade of Data Lake and Cloudera Data Hub clusters. For steps required for the backup and restore approach, refer to the respective documentation on backing up Data Lakes and Cloudera Data Hub and performing metadata restore (automated for Data Lake clusters only).

FAQ for Cloudera on cloud upgrades

During the preparation for an upgrade, it is recommended to carefully review the questions and answers below.

General questions related to upgrading Cloudera

What is the length of the available maintenance window?

Currently, Data Lake backup and restore requires a maintenance window, where no metadata changes occur. Furthermore, Cloudera Data Hub clusters need to be stopped during an upgrade.

The Cloudera environment on cloud does not need to be upgraded in one go: you may opt to upgrade the Data Lake and all attached Data Hubs together, or start with the Data Lake upgrade only and perform individual Data Hub upgrades consecutively. Whether or not a Cloudera Data Hub upgrade is required depends on the Cloudera Runtime version of the Cloudera Data Hub cluster:

- If your Cloudera Data Hub clusters are on Cloudera Runtime version 7.2.15 or earlier, they must run the same major/minor version of Cloudera Runtime as the Data Lake. In this scenario, after a Data Lake upgrade you are required to upgrade any Cloudera Data Hub clusters that are 7.2.15 or earlier to the same version as the Data Lake.
- If your Cloudera Data Hub clusters are on Cloudera Runtime version 7.2.16 or later, they are compatible with newer versions of the Data Lake (7.2.17 or later). You can independently upgrade your Cloudera Data Hub clusters at a later time if you choose to, though it is not required.

What type of upgrade is required?

Currently, there are three types of upgrades available to Data Lake and Cloudera Data Hub clusters: service pack upgrades; minor/major version upgrades; and OS upgrades. Service pack and minor/major version upgrades install a newer version of Cloudera Manager and/or Cloudera Runtime. OS upgrades for Data Lakes and Cloudera Data Hub clusters are complementary and will bring the image of the cluster hosts to a newer version. If you plan to also perform an OS upgrade, plan the maintenance window accordingly.

Are ephemeral disks used for user or workload-related persistent data?

Major/minor version upgrades as well as service pack upgrades will bring Cloudera Manager and Cloudera Runtime to the selected version without impacting the underlying VM. However, OS upgrades will recreate the underlying VM with a fresh image, which results in the loss of any data stored on ephemeral disks.

If you are currently storing user or workload-related data on volumes using ephemeral disks, please reach out to Cloudera support while planning for the upgrade.

What Data Hub cluster templates are in use? Are you using custom templates?

Check whether in-place upgrade is supported for your built-in or custom Cloudera Data Hub template. Depending on the type and version of the Cloudera Data Hub cluster, additional backup steps, manual configuration changes or post-upgrade steps may be required. Check specific steps for upgrading the OS if you use Cloudera Flow Management. Cloudera Operational Database clusters have a different upgrade process.

What is the size of the SDX / Data Lake metadata?

SDX metadata includes the Hive Metastore database, Ranger audit log index, as well as Atlas metadata. If you are planning to perform a Data Lake backup before an upgrade (which is recommended), prepare your maintenance window accordingly. Cloudera supports skipping the backup of certain metadata to reduce the time required for backup and restore operations.

Are you using Data Services?

If you have deployed Cloudera Data Engineering, Cloudera Data Warehouse, Cloudera DataFlow, or Cloudera AI in your environment, be sure to check the *Preparing for an upgrade* topic to verify compatibility between the data service version and the Data Lake version or desired features/Cloudera Runtime services.

Questions related to upgrading to Cloudera Runtime 7.2.18

Will upgrading to Cloudera Runtime 7.2.18 and changing from CentOS to RHEL 8 cost Cloudera customers money?

Upgrading the Cloudera Runtime version to 7.2.18 in Cloudera Public Cloud, and changing the operating system from CentOS 7 to RHEL 8 will not incur additional costs, either to your organization or to the Cloud Service Provider. This update is designed for a frictionless transition and continued support, without financial impact.

What are the key features of Cloudera Runtime 7.2.18?

- RHEL 8 is set as the default operating system, in anticipation of the CentOS 7 sunset on June 30, 2024, to ensure a modern and fully supported infrastructure.
- Rolling upgrades allow for seamless updates of services without operational interruptions.
- Iceberg support is now fully integrated with Atlas, enriching data management capabilities with comprehensive data lineage support.
- The transition from Medium Duty to Enterprise Data Lakes enhances performance and scalability, aligning with advanced workload requirements.
- Amazon S3 Express One Zone support provides a fast and cost-effective data storage option.

What are Rolling Upgrades, and how do they affect my operations?

Rolling Upgrades allow you to upgrade without causing any interruption to ongoing operations. This means customers can continue using their services while the upgrade process is underway.

While Rolling Updates are not available for all services, key services including Data Lakes, Cloudera Operational Database, and Cloudera Streams Messaging Data Hub clusters now support this feature, significantly enhancing operational efficiency and minimizing disruption during upgrades.

Why is the upgrade to Cloudera Runtime 7.2.18 recommended?

Upgrading to Cloudera Runtime 7.2.18 is recommended for several reasons:

- Cloudera Runtime 7.2.18 transitions to RHEL for modern, fully supported Linux infrastructure.
- Cloudera Runtime 7.2.18 introduces rolling upgrades for key services.
- Cloudera Runtime 7.2.18 integrates Iceberg with Atlas for enhanced data management and data lineage.
- Cloudera Runtime 7.2.18 supports Amazon S3 Express One Zone for cost-effective, high-speed storage.

How does the transition to RHEL 8 benefit Cloudera customers?

The transition to RHEL 8 benefits Cloudera customers by ensuring a modern, fully supported Linux infrastructure that meets the latest industry standards. It enhances security, streamlines InfoSec approval processes, and offers greater automation capabilities. This transition supports advanced functionality such as Generative AI and real-time streaming, providing faster time-to-value with no additional costs.

Can you explain the Iceberg support with Atlas integration?

The integration of Iceberg with Atlas in the 7.2.18 update completes Cloudera's Iceberg integration story by providing comprehensive data lineage support. With Atlas, users gain visibility into the lineage of their Iceberg data. This enhancement enriches the platform's ability to manage and understand data across its lifecycle, facilitating better data governance and compliance.

What happens if a customer does NOT upgrade to RHEL 8 by June 30, 2024?

If you do not upgrade to RHEL 8 by the time it is deprecated, Cloudera will still accept support cases. We do ask customer account teams to file for an extension with CentOS. However, Cloudera will not publish OS patches or CVE fixes for CentOS-based images after June 2024.

How do I prepare my environment for upgrading to Cloudera Runtime 7.2.18?

To prepare your environment for upgrading to 7.2.18, check out Upgrading to Cloudera Runtime 7.2.18 for more information. Here you will find instructions on how to identify cluster versions, identify your upgrade path, and more.

How do I prepare my environment for upgrading to RHEL 8?

To prepare your environment for upgrading to RHEL 8, follow the guidance provided in Upgrading from CentOS to RHEL.

Is there a recommended upgrade path for users on various Cloudera Runtime versions?

Yes, and Cloudera offers documentation that guides you through the upgrade process. Check out Upgrading to Cloudera Runtime 7.2.18 for more information.

Can I roll back to a previous Cloudera Runtime version after upgrading to 7.2.18?

No, rolling back to a previous version after upgrading to 7.2.18 is not supported due to compatibility risks. Should you run into any errors or issues, Cloudera support is here to help.

Questions related to upgrading to Cloudera Runtime 7.3.1

What are the key features of Cloudera Runtime 7.3.1?

- Unified Runtime for Cloudera on cloud and Cloudera on premises.
- RHEL 8.10 is set as the default operating system, while RHEL 8.8 continues to be supported.
- Python 3.9 is set as the new default version.

- OpenJDK 17 becomes the new default Java runtime version, new clusters will be launched with this version.

  OpenJDK 11 is no longer supported, and OpenJDK 11 clusters cannot be upgraded to Cloudera Runtime 7.3.1.

  After upgrading clusters from Runtime 7.2.x to 7.3.1, the JDK version remains the same, that is, JDK8.

What features will no longer be available after the upgrade to Cloudera Runtime 7.3.1?

- Customers need to rebase their applications to Spark 3 as Spark 2 is no longer supported and clusters with Spark 2 will not be allowed to upgrade toCloudera Runtime 7.3.1.
- Zeppelin service is turned off for Cloudera Runtime 7.3.1 clusters and gets removed automatically during the upgrade.
- Cloudera Data Engineering 1.23 and earlier versions are not supported on Cloudera Runtime 7.3.1

Why is the upgrade to Cloudera Runtime 7.3.1 recommended?

Upgrading to Cloudera Runtime 7.3.1 is recommended because the Cloudera Runtime 7.3.1 unified Runtime provides seamless portability of workloads across cloud and on-prem environments without rewrite.

How do I prepare my environment for upgrading to Cloudera Runtime 7.3.1?

To prepare your environment for upgrading to version 7.3.1, check out Upgrading to Cloudera Runtime 7.3.1 for more information. Here you will find instructions on how to identify cluster versions, identify your upgrade path, and more.

Is there a recommended upgrade path for users on various Cloudera Runtime versions?

Yes, and Cloudera offers documentation that guides you through the upgrade process. Check out Upgrading to Cloudera Runtime 7.3.1 for more information.

Can I roll back to a previous Cloudera Runtime version after upgrading to 7.3.1?

No, rolling back to a previous version after upgrading to 7.3.1 is not supported due to compatibility risks. Should you run into any errors or issues, Cloudera support is here to help.

How do I prepare my Cloudera Data Hub cluster for Spark 2 to Spark 3 migration?

If you have a Cloudera Data Engineering Data Hub cluster, during the Cloudera Runtime upgrade process to version 7.3.1, you need to remove Spark 2 while creating a new cluster with Spark 3 and migrate your data. Depending on your environment and Spark applications, besides the cluster upgrade tasks, you need to perform Spark application migration tasks and sidecar migration tasks for the Cloudera Data Hub cluster in multiple steps. For details, see Upgrading Apache Spark.

What happens to the Zeppelin service?

Zeppelin service is turned off for Cloudera Runtime 7.3.1 clusters and gets removed automatically during the upgrade. Therefore, you need to back up your data from the Zeppelin service before starting the Cloudera Data Hub upgrade to Runtime 7.3.1.

How do I back up data from the Zeppelin service?

You can back up Zeppelin data, to ensure all necessary components are preserved.

1. **Interpreter Configuration**: Copy the interpreter.json file located in the zeppelin_home/conf folder. This file contains the interpreter settings crucial for Zeppelin operations.
2. **Notebooks**: Back up the entire notebooks folder from the configured repository. This folder includes all Zeppelin notebooks.
3. **Cloudera Manager Configuration**: Access the Cloudera Manager UI, locate the Zeppelin configurations and copy any custom configurations set for the site.xml and env.sh files. Ensure any safety valve variables added in Cloudera Manager are also documented as part of the backup.

Related Information
Preparing for an upgrade

Supported Cloudera Runtime versions for Cloudera on cloud

Before upgrading, review the list of supported Cloudera Runtime versions and their Release Notes to learn more about the new features, known issues, fixed issues and behavioral changes.

Cloudera Runtime version	End of Support (EoS) date	Release Notes
7.2.17.0 including service packs	June 2025	Cloudera Runtime 7.2.17 Release Notes
7.2.18.0 including service packs	September 2025	Cloudera Runtime 7.2.18 Release Notes
7.3.1.0 including service packs	December 2026	Cloudera Runtime 7.3.1 Release Notes

Related Information
Cloudera Support Lifecycle Policy

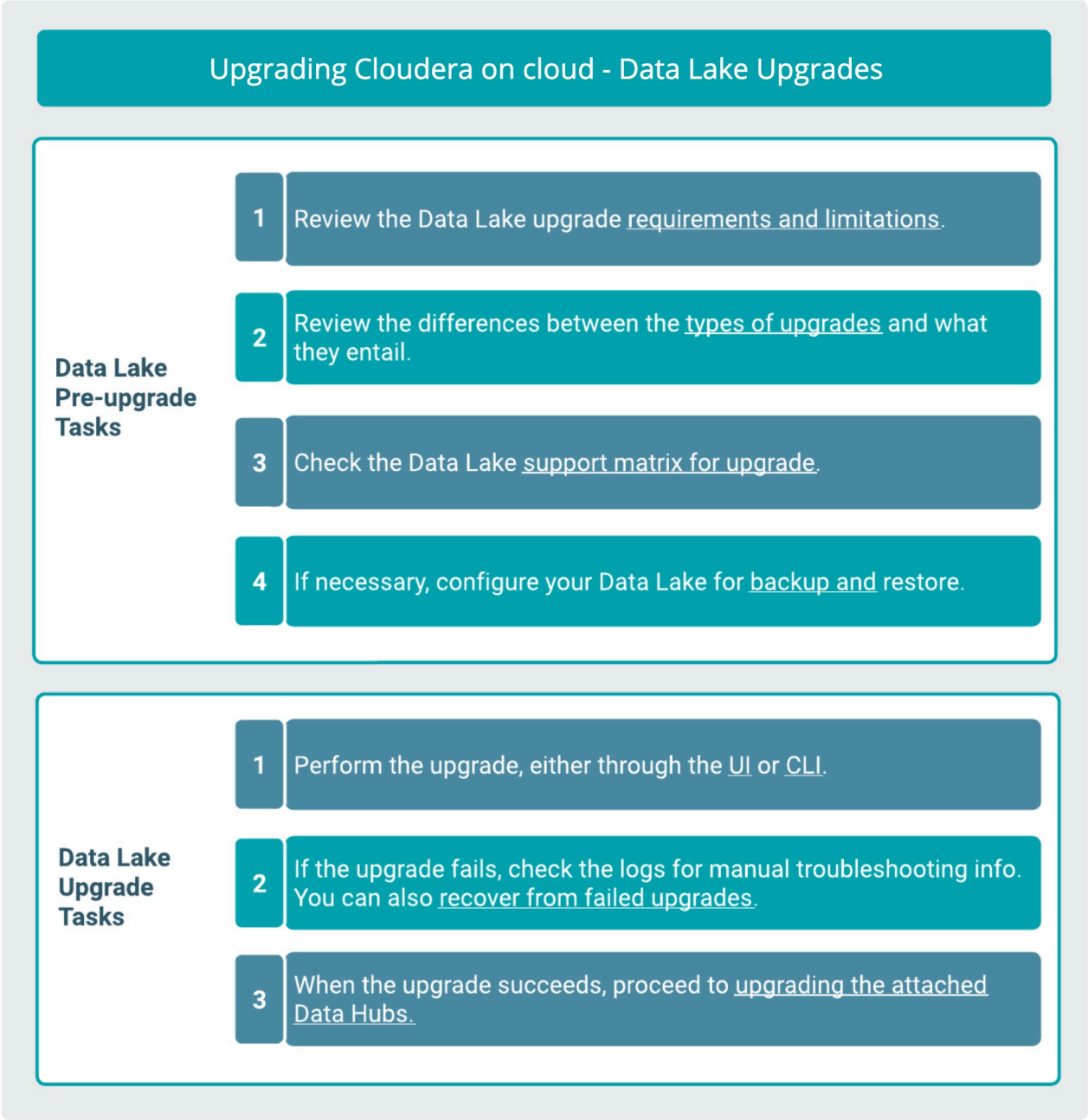
Preparing for an upgrade

Upgrading Cloudera on cloud consists of two major steps: upgrading the Data Lake within an environment and then potentially upgrading the attached Data Hubs. If applicable, you can upgrade any data services after the Data Lake upgrade.

FreeIPA upgrades

You should periodically upgrade the environment (FreeIPA cluster) to ensure that you are running the latest OS-level security patches, but this is not required at the same time as upgrading the Data Lake and Cloudera Data Hub clusters. While FreeIPA upgrade can be performed without a downtime, Cloudera recommends to perform this separately from other operations, preferably before upgrading the Data Lake and Cloudera Data Hub clusters. See upgrade the FreeIPA cluster.

Data Lake upgrade workflow



Pre-upgrade tasks

1. Review the Data Lake upgrade requirements and limitations.
2. Carefully review the differences between the types of upgrades and what they entail.
3. Check the Data Lake support matrix for upgrade to verify which Runtime versions you can upgrade to and from.
4. If you have not configured your Data Lake for backup and restore, you will need to do so. The backup and restore process is integrated into the upgrade flow automatically, but successful upgrade requires that the correct IAM policies exist on the DATALAKE_ADMIN_ROLE and RANGER_AUDIT_ROLE (for backup), and the DATALAKE_ADMIN_ROLE, RANGER_AUDIT_ROLE, and LOG_ROLE (for restore).

   If your roles are not configured correctly, Cloudera will not be able to write the backup to the BACKUP_LOCATION_BASE path of your cloud storage.

5. If you are performing the backup manually (as opposed to the integrated backup available during the upgrade process), you can launch the Data Lake backup from the UI or CLI. When using the CLI, you can specify to skip certain backup actions (skip HMS, Atlas metadata or Ranger audit log index backup). You can monitor the backup process using the CLI.
6. From the Data Lake UI, run the Validate and Prepare option to check for any configuration issues and begin the Cloudera Runtime parcel download and distribution. Using the validate and prepare option does not require downtime and makes the maintenance window for an upgrade shorter. Validate and prepare also does not make any changes to your cluster and can be run independently of the upgrade itself. Although you can begin the upgrade without first running the validate and prepare option, using it will make the process smoother and the downtime shorter. (The parcels that are downloaded and distributed by the validate and prepare option are specific to the Runtime version that you have selected, so if you use validate and prepare and then decide to upgrade to a different Runtime version instead, you will need to re-run validate and prepare. Be aware that if you use validate and prepare for multiple major/minor Runtime versions, the parcels for different versions are not cleaned up and may saturate the disk. These parcels are cleaned up only once the upgrade is complete.)

Data Lake upgrade tasks

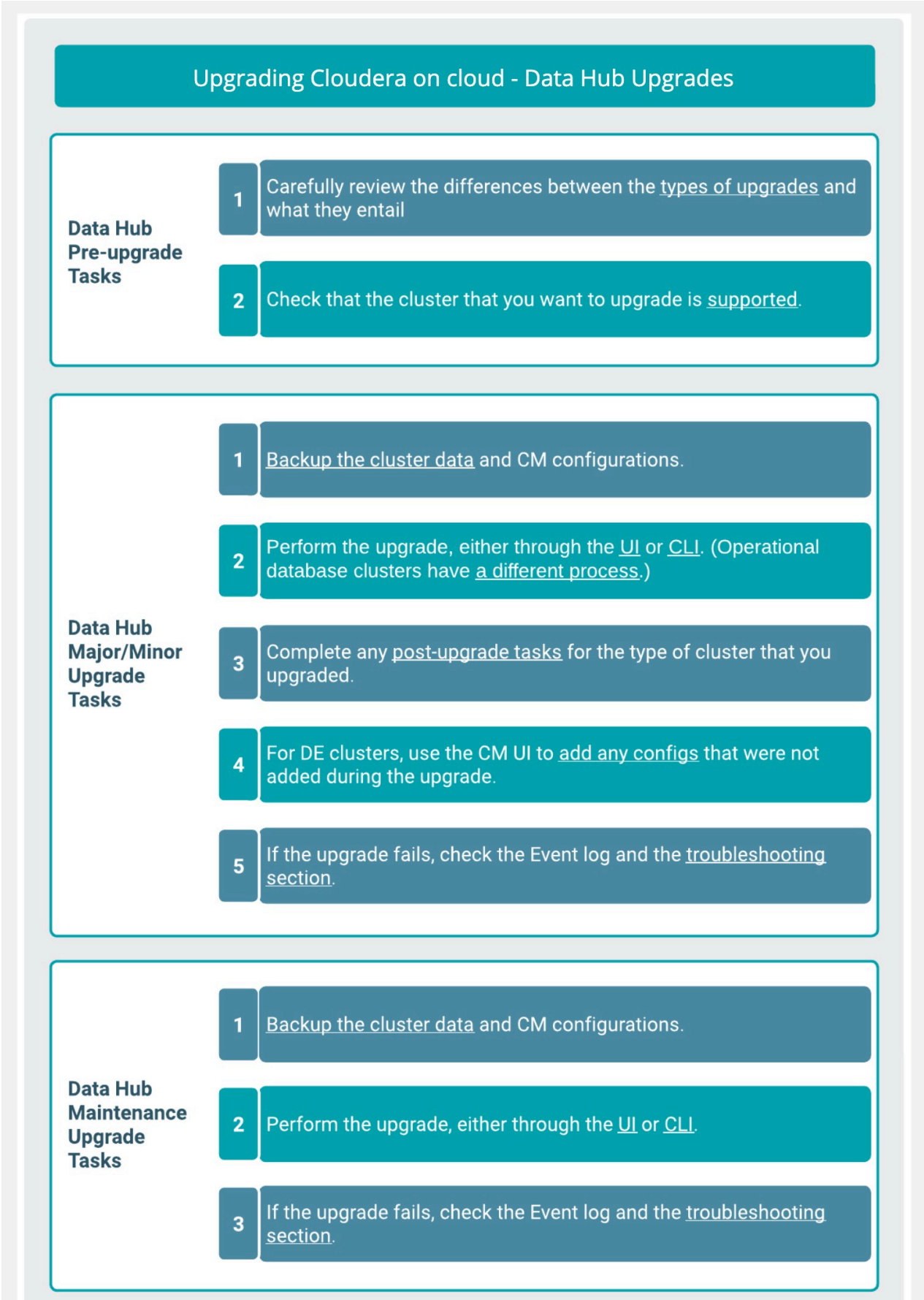
1. Perform the upgrade, either through the UI or CLI. The type of upgrade that you perform will depend on whether a major/minor or service pack version of Runtime is available. A new OS image may also be available to upgrade to.
2. If the upgrade fails, check the logs for manual troubleshooting info. You can also recover from failed upgrades.
3. When the upgrade succeeds, proceed to upgrading the attached Cloudera Data Hub clusters if required.

Data Services upgrades

Any data services in use should be upgraded after the Data Lake upgrade. Some data services have their own compatibility matrix with different versions of the Data Lake. Other data services contain features that may not be compatible with every Data Lake (Cloudera Runtime) version. Refer to the data services documentation for information on Data Lake compatibility and upgrading these services.

- Cloudera AI and Data Lake Compatibility & Cloudera AI service with Data Lake upgrades
- Cloudera Data Engineering and Data Lake Compatibility
- Cloudera DataFlow feature support matrix
- Cloudera Data Warehouse version mapping

Cloudera Data Hub upgrade workflow



Pre-upgrade tasks

1. Carefully review the differences between the types of upgrades and what they entail.
2. Check the Cloudera Data Hub upgrade support matrix to verify that the Cloudera Runtime versions you want to upgrade to and from are supported for your clusters.
3. From the Cloudera Data Hub UI, run the Validate and Prepare option to check for any configuration issues and begin the Cloudera Runtime parcel download and distribution. Using the validate and prepare option does not require downtime and makes the maintenance window for an upgrade shorter. Validate and prepare also does not make any changes to your cluster and can be run independently of the upgrade itself. Although you can begin the upgrade without first running the validate and prepare option, using it will make the process smoother and the downtime shorter. (The parcels that are downloaded and distributed by the validate and prepare option are specific to the Runtime version that you have selected, so if you use validate and prepare and then decide to upgrade to a different Runtime version instead, you will need to re-run validate and prepare. Be aware that if you use validate and prepare for multiple major/minor Runtime versions, the parcels for different versions are not cleaned up and may saturate the disk. These parcels are cleaned up only once the upgrade is complete.)

Major/minor Runtime version upgrade tasks

1. Backup the cluster data and Cloudera Manager configurations.
2. Perform the upgrade, either through the UI or CLI. (Cloudera Operational Database clusters have a different process.)
3. Complete any post-upgrade tasks for the type of cluster that you upgraded.
4. For Cloudera Data Engineering clusters, use the Cloudera Manager UI to add any configs that were not added during the upgrade.
5. If the upgrade fails, check the Event log and the troubleshooting section.
6. Complete any post-upgrade tasks for the type of cluster that you upgraded.
7. For Cloudera Data Engineering clusters, use the Cloudera Manager UI to add any configs that were not added during the upgrade.
8. If the upgrade fails, check the Event log and the troubleshooting section.

Service pack upgrade tasks

1. Backup the cluster data and Cloudera Manager configurations.
2. Perform the upgrade, either through the UI or CLI.
3. If the upgrade fails, check the Event log and the troubleshooting section.

OS upgrade tasks

1. Review the Before you begin section to verify that there is no data belonging to NiFi or NiFi Registry on the root disk of the VM. Note that during an OS upgrade, any data on the root volume (parcels, service logs, custom software) will be lost.
2. Unlike the Data Lake upgrade, OS upgrades are not integrated in the larger upgrade flow and must be performed separately, either through the UI or CLI.

Rolling upgrades

With the release of Cloudera Runtime 7.2.18, rolling upgrades for certain Data Lakes and Cloudera Data Hub clusters are now available.

Rolling upgrades allow you to upgrade the Data Lake Cloudera Runtime and OS, or the Cloudera Data Hub Runtime and OS, without stopping the cluster and its services. This means that you can upgrade a Data Lake without stopping the attached Cloudera Data Hub clusterss and Data Services, or upgrade a Cloudera Data Hub without workload downtime.

Rolling upgrades for the Data Lake are limited to certain Data Lake Cloudera Runtime versions and shapes. Rolling upgrades for Cloudera Data Hub clusters are limited to certain Cloudera Runtime versions and cluster templates.

Upgrading to Cloudera Runtime 7.3.1

Cloudera Runtime 7.3.1 introduces significant new features and improvements. As a result, if you are planning to upgrade the Cloudera Runtime version in your existing Data Lake or Cloudera Data Hub clusters to 7.3.1, you might be required to perform this in multiple steps.

Cloudera Runtime 7.3.1.500 Service Pack 3 (SP3)

Some of the important Cloudera Runtime 7.3.1.500 SP3 features include:

- FreeIPA, Data Lake and RDS on ARM-based architecture
- Cloudera Data Hub clusters on ARM-based architecture:
  - Data Discovery and exploration - Spark 3
  - Data Mart
  - Real-time Data Mart - Spark 3
  - Streams Messaging

Cloudera Runtime 7.3.1.400 Service Pack 2 (SP2)

On June 27, Cloudera Runtime 7.3.1.400 Service Pack 2 (SP2) was released that includes a set of fixed issues and improvements. For more information, see the Cloudera Runtime Release Notes.

Cloudera Runtime 7.3.1.200 Service Pack 1 (SP1)

Some of the important Cloudera Runtime 7.3.1.200 SP1 features include:

- General Availability (GA) of Cloudera Operational Database on ARM-based architecture
- Technical Preview of Cloudera Data Engineering Data Hub clusters on ARM-based architecture

Cumulative Hotfix

On June 2, Cloudera Runtime 7.3.1.300 Service Pack 1 (SP1) Cumulative Hotfix 1 (CHF1) was released that includes a set of fixed issues and improvements. For more information, see the Cloudera Runtime Release Notes.

Cloudera Runtime 7.3.1

Some of the important Cloudera Runtime 7.3.1 features include:

- Unified Runtime for Cloudera on cloud and Cloudera on premises.
- RHEL 8.10 is set as the default operating system.
- Python 3.9 is set as the new default version.
- OpenJDK 17 becomes the new default Java runtime version, new clusters will be launched with this version.
- OpenJDK 11 is no longer supported. If you are using Cloudera Runtime and with JDK 11, you need to update the JDK version to 8 to be able to upgrade Cloudera Runtime to 7.3.1. For more information, see the Changing the default Java version to 8 documentation.
- After upgrading clusters from Cloudera Runtime 7.2.x to 7.3.1, the JDK version remains the same, that is, JDK8.
- Spark 2 applications need to be rebased to Spark 3 as Spark 2 is no longer supported and clusters with Spark 2 will not be allowed to upgrade to Cloudera Runtime 7.3.1.
- Zeppelin service is turned off for Cloudera Runtime 7.3.1 clusters and gets removed automatically during the upgrade.
- Cloudera Data Engineering 1.23 and earlier versions are not supported on Cloudera Runtime 7.3.1
- Spark in Cloudera AI is not certified to work with the Data Lake 7.3.1 version.

Cumulative Hotfix

On March 18, Cloudera Runtime 7.3.1.100 Cumulative Hotfix 1 (CHF1) was released that includes a set of fixed issues and improvements. For more information, see the Cloudera Runtime Release Notes.

What to do next?

The following pages provide step-by-step guidance on identifying the upgrade path for your source cluster version and upgrading to Cloudera Runtime 7.3.1.

Identify cluster version details

Identify your Data Lake and Cloudera Data Hub cluster's Cloudera Runtime version, Data Lake, Cloudera Data Hub, and FreeIPA image version and operating system (OS) type, and Data Lake shape.

Your upgrade path to versions 7.2.18 or 7.3.1 may depend on the current Cloudera Runtime version, image version, and OS of your cluster. Cloudera recommends that you identify your version information as follows:

1. Identify cluster's Cloudera Runtime version on page 34
2. Identify cluster's image version, image creation date, OS type on page 35
3. Identify Data Lake shape on page 37

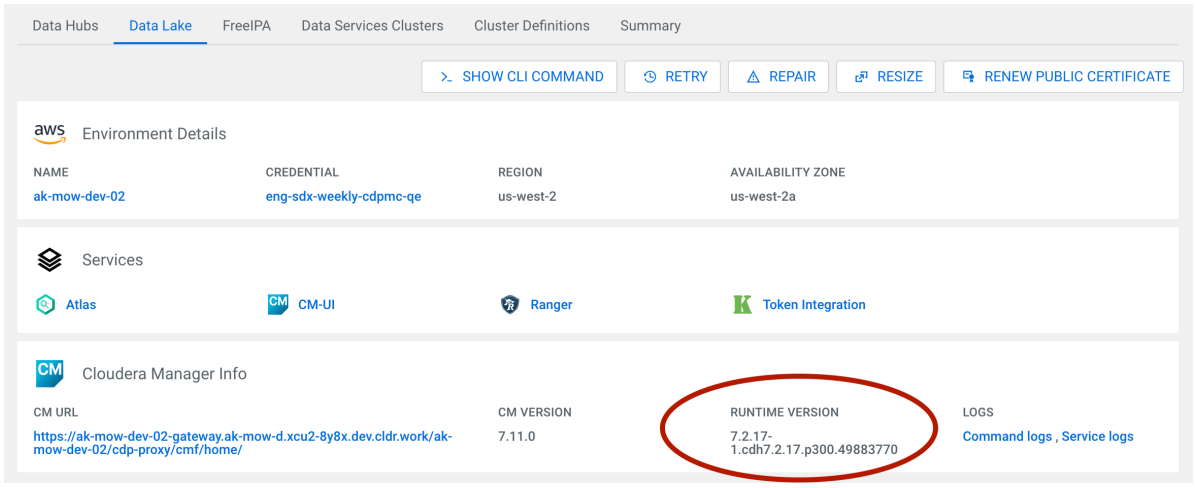
Identify cluster's Cloudera Runtime version

Follow these instructions to identify your Data Lake and Cloudera Data Hub's Cloudera Runtime version. You should follow these steps separately for each cluster.

Steps

For Cloudera UI

1. In the Cloudera Management Console, navigate to the details page of the Data Lake or Cloudera Data Hub cluster that you would like to upgrade.
2. In the Cloudera Manager Info section find the field called "Runtime version":



For CDP CLI

Get the Cloudera Runtime version of your Data Lake from the CDP CLI by using the following command:

```
cdp datalake describe-datalake --datalake-name <DL-NAME>
```

Get the Cloudera Runtime version of your Cloudera Data Hub from the CDP CLI by using the following command:

```
cdp datahub describe-cluster --cluster-name <DH-NAME>
```

The version information is displayed in the following format:

7.2.17-1.cdh7.2.17.p200.46967063

The highlighted digits show the current Cloudera Runtime version and the number after p indicates the current Service Pack version. In the above example, the cluster is using Cloudera Runtime version 7.2.17.200.

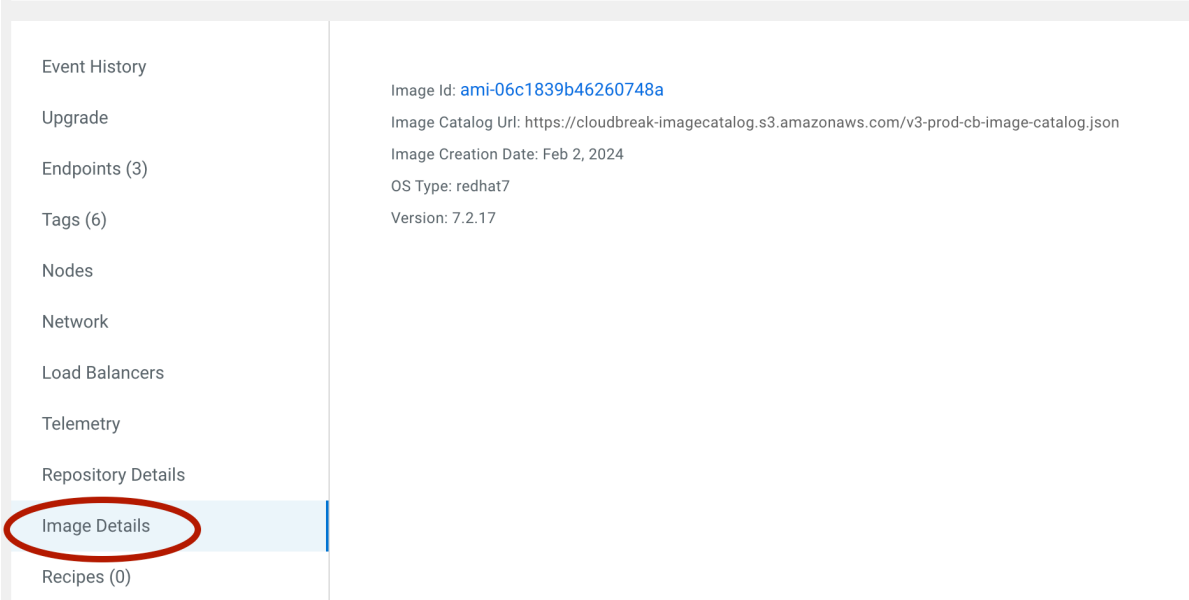
Identify cluster's image version, image creation date, OS type

Follow these instructions to identify Data Lake and Cloudera Data Hub image version and OS type. You should follow these steps separately for each cluster.

Steps

For Cloudera UI

1. In the Cloudera Management Console, navigate to the details page of the Data Lake or Cloudera Data Hub cluster that you would like to upgrade.
2. Scroll down to access the left navigation pane.
3. From the left navigation pane, select Image Details:



4. You will see information similar to the following:

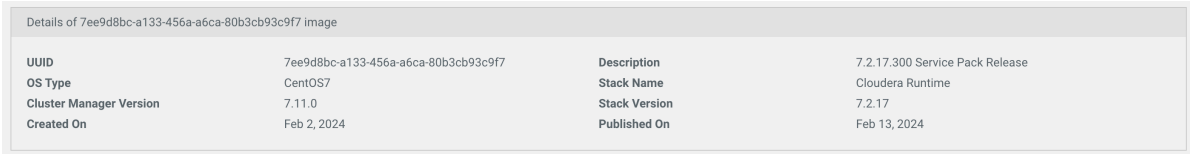
```
Image Id: ami-0e95ecddb7fc75ea8
Image Catalog Url: https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-
prod-cb-image-catalog.json
Image Creation Date: Jan 9, 2024
OS Type:redhat7
Version: 7.2.17
```

Under OS Type:

- If you see "redhat7", your cluster is using a CentOS 7 OS image. This is a known issue.

- If you see “redhat8”, your cluster is using a RHEL 8 OS image.

Clicking on image ID will redirect you to the image catalog and provide you with further information, including the actual OS Type:



For CDP CLI

Get Data Lake’s image version from the Cloudera CLI by using the following command:

```
cdp datalake upgrade-datalake --show-latest-available-image-per-runtime --
datalake-name <DL-NAME>
```

Get Cloudera Data Hub cluster’s image version from the CDP CLI by using the following command:

```
cdp datahub upgrade-cluster --show-latest-available-image-per-runtime --
cluster-name <DH-NAME>
```

The output of this command looks like this, providing you with the OS information:

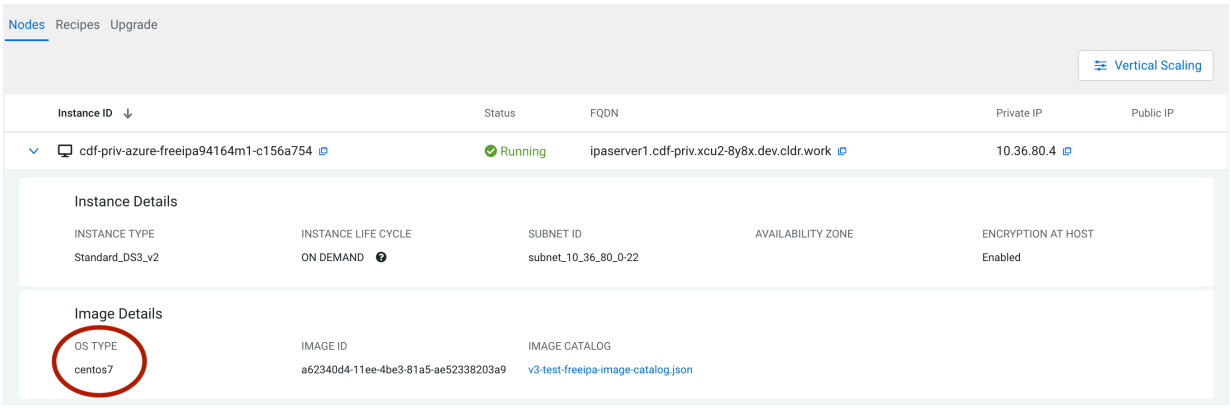
```
{
 "targetImage": {
 "catalog": "https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-prod-fr
eeipa-image-catalog.json",
 "id": "a5b2580e-7487-4ee6-a80f-d991fa00cd13",
 "os": "redhat8",
 "imageName": "ami-0935d849de6a25e45",
 "date": "2024-01-09"
 },
 "originalImage": {
 "catalog": "https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-prod-f
reeipa-image-catalog.json",
 "id": "e1f92688-3773-47dd-89ff-561933540108",
 "os": "centos7",
 "imageName": "ami-0db1b7000c8a27fd4",
 "date": "2023-11-22"
 },
 "operationId": "3a50f30f-33f3-4f81-b9bb-30c119023ff0"
}
```

Follow these instructions to identify FreeIPA OS type.

Steps

1. In the Cloudera Management Console, navigate to the environment details page,.
2. Click on the FreeIPA tab.
3. In the Nodes tab (which is open by default), expand one section corresponding to an instance ID.

4. Under Image Details, you can find the OS Type:

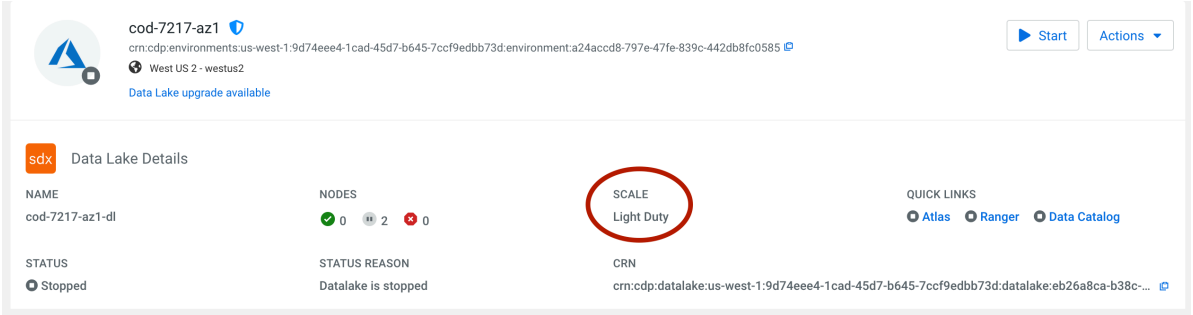


Identify Data Lake shape

Follow these instructions to identify Data Lake's shape.

For Cloudera UI

1. In the Cloudera Management Console, navigate to the details page of the Data Lake that you would like to upgrade:



2. Find the Scale field. There are three possible values: "Light Duty", "Medium Duty HA" and "Enterprise Data Lake".

For CDP CLI

Get the Data Lake shape from the CDP CLI by using the following command:

```
cdp datalake describe-datalake –datalake-name <DL-NAME>
```

Find the "scale" field in the output of this command. There are three possible values: "Light Duty", "Medium Duty HA" and "Enterprise Data Lake".

Supported Cloudera Runtime 7.3.1 upgrade paths

Learn more about the supported upgrade paths to Cloudera Runtime 7.3.1 for Data Lakes and Cloudera Data Hub clusters.

When upgrading to Cloudera Runtime 7.3.1 using the Cloudera Management Console, you will be upgraded to the latest available Service Pack (SP) or Cumulative Hotfix (CHF) after selecting 7.3.1 as the target Cloudera Runtime version. If you are on a source Cloudera Runtime version that does not support a direct upgrade path to Cloudera Runtime 7.3.1, you need to upgrade to a version that enables the direct upgrade.

If you do not want to upgrade to the latest available Cloudera Runtime version, but to a previously released version such as 7.3.1.0 or 7.3.1.100, you can use CDP CLI and manually provide the target Cloudera Runtime version with

the image ID. In this case, ensure that there is a supported upgrade path to Cloudera Runtime 7.3.1 for your source version.

You can use the following table to check if there is a direct upgrade path available to Cloudera Runtime 7.3.1 for your source cluster version:

Source version	Target version: 7.3.1.0	Target version: 7.3.1.10x	Target version: 7.3.1.200+
7.2.16.x			
7.2.17.0			
7.2.17.100			
7.2.17.200	##	##	##
7.2.17.300	##	##	##
7.2.17.400	##	##	##
7.2.17.500	##	##	##
7.2.17.600			##
7.2.17.700			##
7.2.17.800			##
7.2.17.900			##
7.2.17.1000			##
7.2.17.1100			##
7.2.17.1200			##
7.2.18.0	##	##	##
7.2.18.100	##	##	##
7.2.18.200	##	##	##
7.2.18.300			##
7.2.18.400			##
7.2.18.500			##
7.2.18.600			##
7.2.18.700			##
7.2.18.800			##
7.2.18.900			##
7.2.18.1000			##
7.2.18.1101			##

For more information about the latest available Cloudera Runtime version, see the Cloudera Runtime Release Notes.

What to do next?

- If you want to upgrade to the latest available 7.3.1 version and your current version supports a direct upgrade, follow the steps described in Upgrade to the latest 7.3.1 patch version.
- If you want to upgrade to the latest available 7.3.1 version and your current version does not support a direct upgrade, follow the steps described in Upgrade to a version that enables 7.3.1 upgrade.
- If you want to upgrade to a previously released 7.3.1 version, such as 7.3.1.0 or 7.3.1.10x, and your current version supports a direct upgrade, follow the steps described in Upgrade to a specific 7.3.1 patch version.

Review the prerequisites

Before you start upgrading your clusters, ensure that you meet the prerequisites for each upgrade that is required.

Refer to the following documentation for detailed prerequisites:

- Review Data Lake Upgrade: Before you begin and Cloudera Data Hub Upgrade: Limitations and Prerequisites.
- If you are upgrading a Data Lake, use the Backup and Restore functionality. This will allow you to restore the SDX metadata if required.
- Certain Data Hubs might require certain pre-upgrade and post-upgrade steps. Review the related documentation before upgrading Data Hubs.
- If you are upgrading from CentOS to RHEL, review the Prerequisites for upgrading from CentOS to RHEL.
- If your upgrade path involves a Data Lake resize, refer to Review Data Lake resizing: Prerequisites.

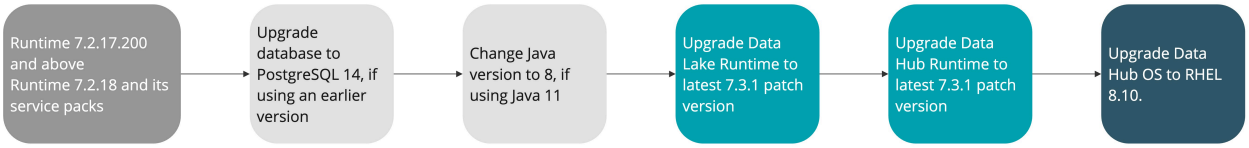
Upgrade Data Lakes and all Cloudera Data Hub clusters to the latest 7.3.1 version

When upgrading to Cloudera Runtime 7.3.1 using the Cloudera Management Console, you will be upgraded to the latest available Service Pack (SP) or Cumulative Hotfix (CHF) after selecting 7.3.1 as the target Cloudera Runtime version.

Upgrading from 7.3.1.x

If you are using Cloudera Runtime 7.3.1 or any of its patch versions, you can directly upgrade to the latest Cloudera Runtime Service Pack (SP) or Cumulative Hotfix (CHF) version using the steps for Performing a Cloudera Data Hub service pack upgrade.

Upgrading from 7.2.17.200 and higher versions or 7.2.18.x



If you are using 7.2.17.200 and higher versions or 7.2.18 and its service packs, you can directly upgrade to the latest Cloudera Runtime, 7.3.1.300 SP1 CHF1, using the following steps:

1. Upgrade database to PostgreSQL 14 - Only required if you are currently using a lower version than PostgreSQL 14
2. Changing the default Java version to 8 - Only required if the current Java version of your Data Lake and Cloudera Data Hub cluster is 11
3. Upgrade Data Lake Cloudera Runtime to 7.3.1 and OS to RHEL 8.10
4. Upgrade Cloudera Runtime version of Cloudera Data Hub clusters to 7.3.1

Important:

- Each of these operation has its own prerequisites and postrequisites so in addition to the actual upgrade steps, the prerequisites and postrequisites need to be performed as well.
- You will need to upgrade all Data Lakes and Cloudera Data Hub clusters separately, first the Data Lakes and then each Cloudera Data Hub cluster.
- For Data Lakes, the Runtime and OS upgrade is performed in one step. For Cloudera Data Hub clusters, the Cloudera Runtime needs to be updated first, and then the OS in a separate step.
- You can perform each upgrade operation either from the Upgrade UI of the respective cluster or using the CDP CLI.
- The upgrade user interface in the Cloudera Management Console shows you the next Data Lake and Cloudera Data Hub upgrade step available, so you can refer to it as a guideline.

Upgrade database to PostgreSQL 14

Before being able to upgrade to 7.3.1, you need to upgrade Data Lakes and all databases to a to PostgreSQL 14.

For information on upgrading Data Lake and databases to PostgreSQL 14, see Upgrading Data Lake/ database.

Note:

If you are using an Azure environment, upgrading the PostgreSQL database used by the Data Lake and clusters also upgrades Azure Single Server to Azure Flexible Server as Single Server is not supported with PostgreSQL 14.

For more information, see Upgrading Azure Single Server to Flexible Server. Make sure to review both the database upgrade and the Azure Single to Flexible Server upgrade documentation.

Changing the default Java version to 8

As JDK 11 is no longer supported in Cloudera Runtime 7.3.1, you need to change the default Java version to 8 to be able to upgrade to Cloudera Runtime 7.3.1.

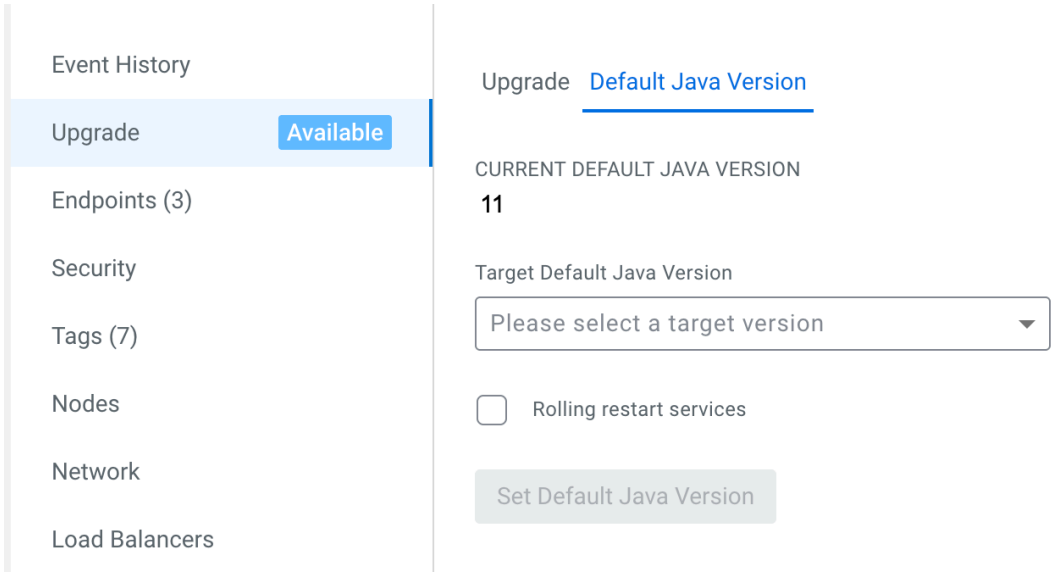
Important: After changing the default Java version, Cloudera Manager and the Cloudera Runtime services will restart.

For Cloudera UI

You can change the default Java version from 11 to 8 for your Data Lake and Cloudera Data Hub cluster on Cloudera Management Console.

1. Navigate to the **Upgrade** tab of your Data Lake and Cloudera Data Hub cluster.
2. Select the Default Java Version tab.
3. Choose Java 8 as the **Target Default Java Version**.
4. Check the Rolling restart services box to enable rolling restart for the Data Lake and Cloudera Data Hub services.

5. Click Set Default Java Version.



For CDP CLI

Use the following command to change the default Java version for the Data Lake:

```
cdp datalake set-default-java-version --cluster-crn [***CRN***] --java-version 8
```

Use the following command to change the default Java version for the Cloudera Data Hub cluster:

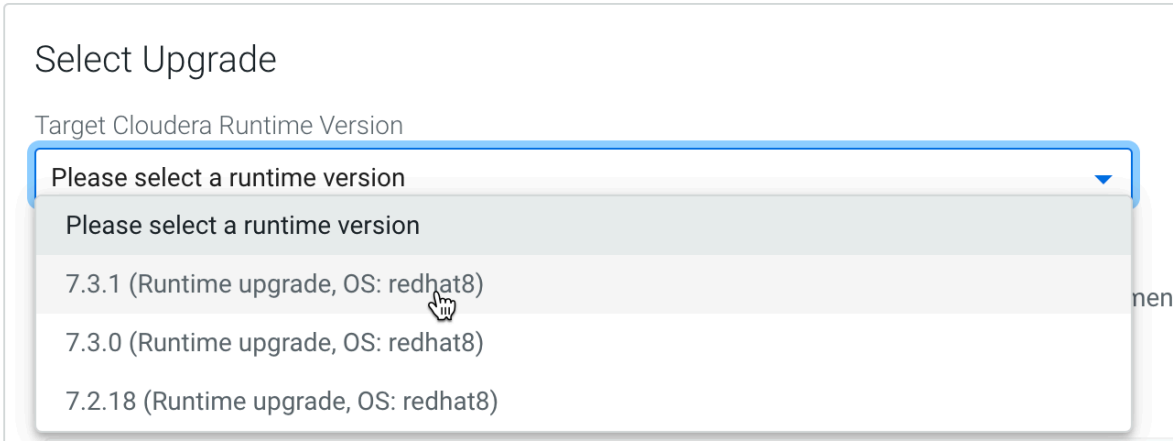
```
cdp datahub set-default-java-version --cluster-crn [***CRN***] --java-version 8
```

Upgrade Data Lake Cloudera Runtime version to 7.3.1 and OS to RHEL 8.10

From the supported Cloudera Runtime 7.2.17 or 7.2.18 versions, you need to upgrade Data Lake Cloudera Runtime version to 7.3.1 and OS to RHEL 8.10 in one step prior to upgrading Cloudera Data Hub clusters.

All Cloudera Data Hub clusters in your environment should be stopped and Data Services should not be running workloads for the duration of the operation.

- **1.** Once Runtime 7.3.1 (Runtime upgrade, OS: redhat8) is available to select from the Upgrade UI, select it and run Validate and Prepare.



Important: When selecting a main version, such as 7.3.1, for Cloudera Runtime version, you are upgraded to latest available Cumulative Hotfix (CHF) of the latest Service Pack (SP).

2. Once Validate and Prepare completes, return to this page, select 7.3.1 (Runtime upgrade) again and run Upgrade.

Tip: When upgrading a Data Lake, Cloudera recommends allocating a 60-minute maintenance window for this upgrade step.

For more information, see Upgrading a Data Lake.

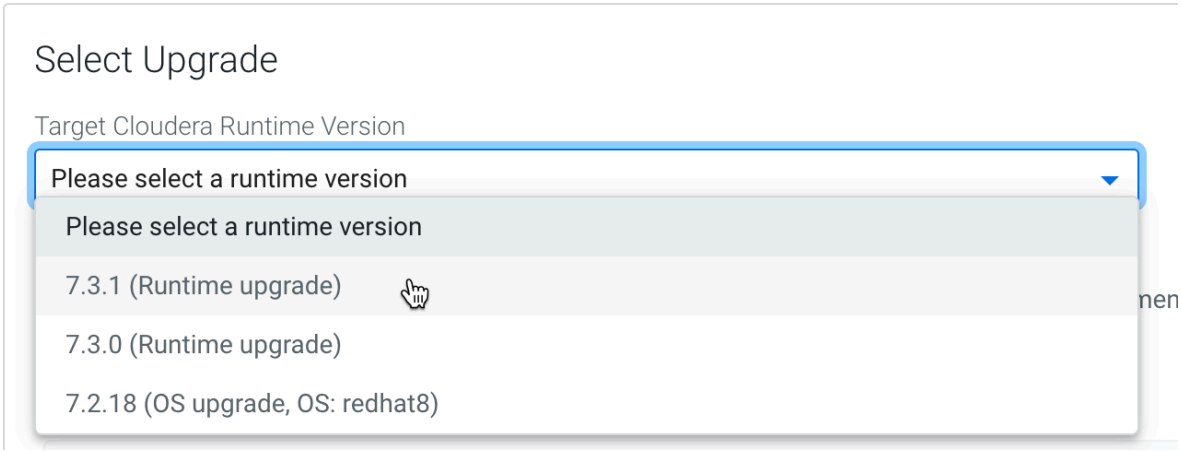
Upgrade Cloudera Runtime version of Cloudera Data Hub clusters to 7.3.1

Once the Data Lake Cloudera Runtime version has been upgraded to 7.3.1 and simultaneously, the OS has been upgraded to RHEL 8.10, the Cloudera Runtime version of each Cloudera Data Hub cluster needs to be upgraded to 7.3.1.

Note:

If you have a Cloudera Data Engineering Data Hub cluster, during the Cloudera Runtime upgrade process to version 7.3.1, you need to remove Spark 2 while creating a new cluster with Spark 3 and migrate your data. Depending on your environment and Spark applications, besides the cluster upgrade tasks, you need to perform Spark application migration tasks and sidecar migration tasks for the Cloudera Data Hub cluster in multiple steps. For more information, see Upgrading Apache Spark.

- **1.** Once Runtime 7.3.1 (Runtime upgrade) is available to select from the Upgrade UI, select it and run Validate and Prepare.



Important: When selecting a main version, such as 7.3.1, for Cloudera Runtime version, you are upgraded to latest available Cumulative Hotfix (CHF) of the latest Service Pack (SP).

2. Once Validate and Prepare completes, return to this page, select 7.3.1 (Runtime upgrade) again and run Upgrade.

Tip: When upgrading a Cloudera Data Hub cluster,allocate 90-120 minutes (as you will be performing this upgrade in two steps).

For more information, see Upgrading Cloudera Data Hub clusters.

Upgrade Data Lakes and all Cloudera Data Hub clusters to a specific 7.3.1 patch version with CDP CLI

In case you want to upgrade to a previously released 7.3.1 version, such as 7.3.1.0 or 7.3.1.10x, you can upgrade your Data Lake and Cloudera Data Hub clusters using CDP CLI.

Changing the default Java version to 8

As JDK 11 is no longer supported in Cloudera Runtime 7.3.1, you need to change the default Java version to 8 to be able to upgrade to Cloudera Runtime 7.3.1.

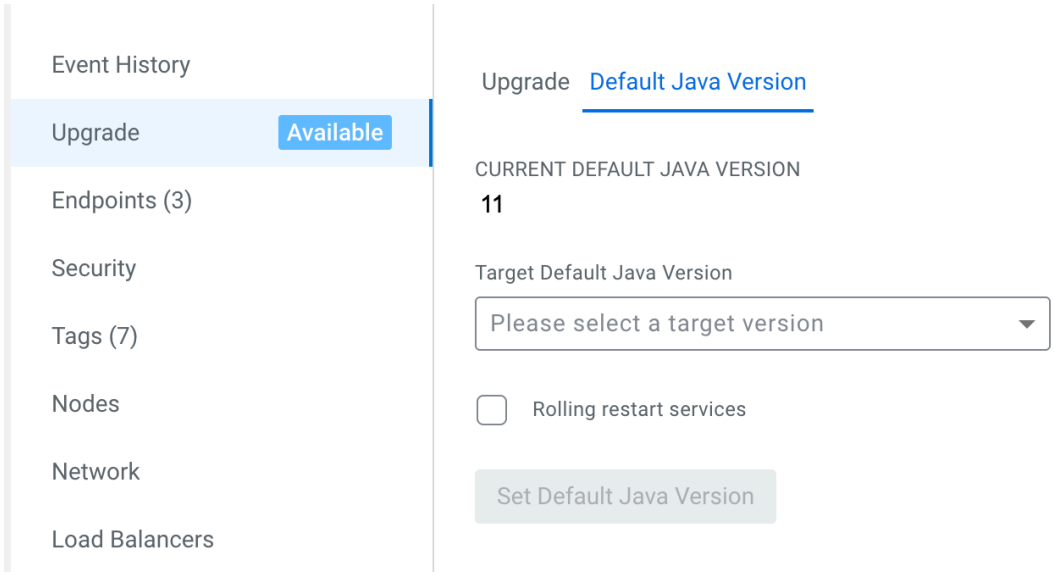
Important: After changing the default Java version, Cloudera Manager and the Cloudera Runtime services will restart.

For Cloudera UI

You can change the default Java version from 11 to 8 for your Data Lake and Cloudera Data Hub cluster on Cloudera Management Console.

1. Navigate to the **Upgrade** tab of your Data Lake and Cloudera Data Hub cluster.
2. Select the Default Java Version tab.
3. Choose Java 8 as the **Target Default Java Version**.
4. Check the Rolling restart services box to enable rolling restart for the Data Lake and Cloudera Data Hub services.

5. Click Set Default Java Version.



For CDP CLI

Use the following command to change the default Java version for the Data Lake:

```
cdp datalake set-default-java-version --cluster-crn [***CRN***] --java-
version 8
```

Use the following command to change the default Java version for the Cloudera Data Hub cluster:

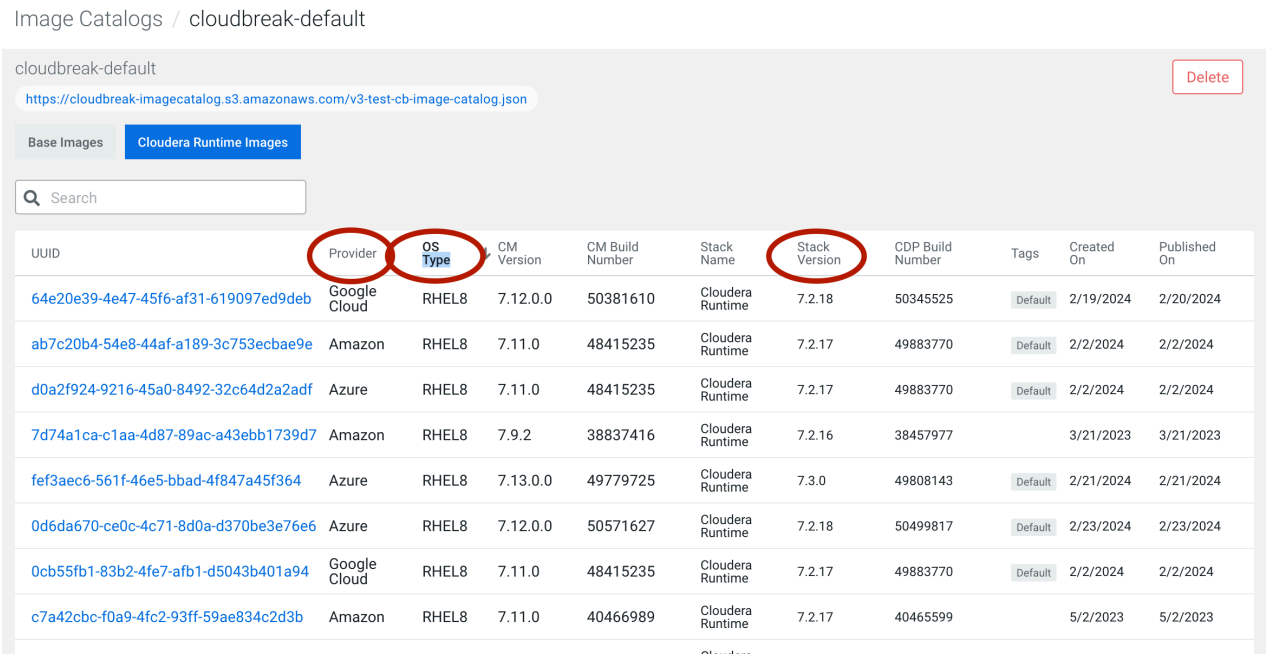
```
cdp datahub set-default-java-version --cluster-crn [***CRN***] --java-vers
ion 8
```

Upgrading a Data Lake manually via CLI

You can initiate a Data Lake upgrade with the CDP CLI. Using the same CLI command, you can also search for and validate available images to upgrade to, and generate JSON templates for specific upgrade scenarios.

Obtain image ID

If your Data Lake upgrade includes upgrading from CentOS to RHEL 8, prior to attempting an upgrade you need to obtain an ID of a target RHEL 8 image. You can obtain it from the image catalog by finding an image with your target Runtime version which has an OS Type of RHEL8.



UUID	Provider	OS Type	CM Version	CM Build Number	Stack Name	Stack Version	CDP Build Number	Tags	Created On	Published On
64e20e39-4e47-45f6-af31-619097ed9deb	Google Cloud	RHEL8	7.12.0.0	50381610	Cloudera Runtime	7.2.18	50345525	Default	2/19/2024	2/20/2024
ab7c20b4-54e8-44af-a189-3c753ecbae9e	Amazon	RHEL8	7.11.0	48415235	Cloudera Runtime	7.2.17	49883770	Default	2/2/2024	2/2/2024
d0a2f924-9216-45a0-8492-32c64d2a2adf	Azure	RHEL8	7.11.0	48415235	Cloudera Runtime	7.2.17	49883770	Default	2/2/2024	2/2/2024
7d74a1ca-c1aa-4d87-89ac-a43ebb1739d7	Amazon	RHEL8	7.9.2	38837416	Cloudera Runtime	7.2.16	38457977		3/21/2023	3/21/2023
fef3aec6-561f-46e5-bbad-4f847a45f364	Azure	RHEL8	7.13.0.0	49779725	Cloudera Runtime	7.3.0	49808143	Default	2/21/2024	2/21/2024
0d6da670-ce0c-4c71-8d0a-d370be3e76e6	Azure	RHEL8	7.12.0.0	50571627	Cloudera Runtime	7.2.18	50499817	Default	2/23/2024	2/23/2024
0cb55fb1-83b2-4fe7-afb1-d5043b401a94	Google Cloud	RHEL8	7.11.0	48415235	Cloudera Runtime	7.2.17	49883770	Default	2/2/2024	2/2/2024
c7a42cbc-f0a9-4fc2-93ff-59ae834c2d3b	Amazon	RHEL8	7.11.0	40466989	Cloudera Runtime	7.2.17	40465599		5/2/2023	5/2/2023

Once you have identified the ID, you can provide it in the upgrade CLI command by using the --image-id flag.

Upgrade steps

1. Run the cdp datalake upgrade-datalake command. In order to use this command for upgrading from CentOS to RHEL, ensure to provide an image ID of a RHEL 8 image.

   The command has the following options:

```
cdp datalake upgrade-datalake
          --datalake-name <value>
          [--image-id <value>]
          [--runtime <value>]
          [--lock-components | --no-lock-components]
          [--dry-run | --no-dry-run]
          [--show-available-images | --no-show-available-images]
          [--show-available-image-per-runtime | --no-show-available-image-
per-runtime]
          [--skip-backup | --no-skip-backup]
          [--skip-ranger-hms-metadata | --no-skip-ranger-hms-metadata]
          [--skip-atlas-metadata | --no-skip-atlas-metadata]
          [--skip-ranger-audits | --no-skip-ranger-audits]
          [--skip-backup-validation | --no-skip-backup-validation]
          [--cli-input-json <value>]
          [--generate-cli-skeleton]
```

Option	Description
--datalake-name (string)	Required. The name or CRN of the Data Lake to upgrade.
--image-id (string)	The ID of an image to upgrade to. If upgrading from CentOS to RHEL, make sure to provide an image ID of a target RHEL image.
--runtime (string)	The Cloudera Runtime clusters version to upgrade to. In a maintenance upgrade this parameter is the current Cloudera Runtime version. When you specify the Cloudera Runtime version, the upgrade uses the latest image ID of the given Cloudera Runtime version from the same image catalog used for Data Lake creation. If you specify an invalid Runtime version, you'll receive an error message that the version is not supported for upgrade.

Option	Description
--lock-components | --no-lock-components (boolean)	Use --lock components to perform an OS upgrade only.
--dry-run | --no-dry-run (boolean)	Checks the eligibility of an image to upgrade. Can be used in conjunction with any other parameter, returning the available image (with respect to image Id, Runtime or lock-components set) without performing any actions.
--show-available-images | --no-show-available-images (boolean)	Returns the list of images that are eligible to upgrade to.
--show-available-image-per-runtime | --no-show-available-image-per-runtime (boolean)	Returns the latest image that is eligible to upgrade to, for each Cloudera Runtime version with at least one available upgrade candidate.
--skip-backup | --no-skip-backup	If provided, will skip the backup flow for the upgrade process.
--skip-ranger-hms-metadata | --no-skip-ranger-hms-metadata	Skips the backup of the databases backing HMS/Ranger services. Redundant if –skip-backup is included. If this option is not provided, the HMS/Ranger services are backed up by default.
--skip-atlas-metadata | --no-skip-atlas-metadata	Skips the backup of the Atlas metadata. Redundant if –skip-backup is included. If this option is not provided, the Atlas metadata is backed up by default.
--skip-ranger-audits | --no-skip-ranger-audits	Skips the backup of the Ranger audits. Redundant if –skip-backup is included. If this option is not provided, Ranger audits are backed up by default.
--skip-backup-validation | --no-skip-backup-validation	Skips the validation steps that run prior to the backup. Redundant if –skip-backup is included. If this option is not provided, the validations are performed by default.
--cli-input-json (string)	Performs service operation based on the JSON string provided. The JSON string follows the format provided by --generate-cli-skeleton. If other arguments are provided on the command line, the CLI values will override the JSON-provided values.
--generate-cli-skeleton (boolean)	Prints a sample input JSON to standard output. Note the specified operation is not run if this argument is specified. The sample input can be used as an argument for --cli-input-json.

When you run the cdp datalake upgrade-datalake command to initiate an upgrade, you have one of three options:

a. Specify one of either --image-id, --runtime, or --lockComponents, which makes an explicit choice of the exact image, Runtime (latest OS), or latest OS (same Runtime) for upgrade.
b. Specify both --image-id and --lockComponents, which specifies an image and ensures the image represents an OS only upgrade.
c. Specify none of the --image-id, --runtime, or --lockComponents parameters, which initiates a Runtime/CM upgrade to the latest compatible version and OS image.

Outside of upgrade, you can use the following options:

- --show-available-images/--no-show-available-images
- --show-available-images-per-runtime/--no-show-available-images-per-run time
- --dry-run

Examples of valid inputs:

```
cdp datalake upgrade-datalake --datalake-name my-datalake --dry-run
```

```
cdp datalake upgrade-datalake --datalake-name my-datalake --image-id d1c
520b1-987d-461f-7860-918f43994c04
cdp datalake upgrade-datalake --datalake-name my-datalake --image-id d1c52
0b1-987d-461f-7860-918f43994c04 --dry-run
cdp datalake upgrade-datalake --datalake-name my-datalake --runtime 7.2.11
cdp datalake upgrade-datalake --datalake-name my-datalake --runtime 7.2
.11 --dry-run
cdp datalake upgrade-datalake --datalake-name my-datalake--lock-components
cdp datalake upgrade-datalake --datalake-name my-datalake --show-availa
ble-image-per-runtime
cdp datalake upgrade-datalake --datalake-name my-datalake- --show-availabl
e-images
```

Examples of incorrect inputs:

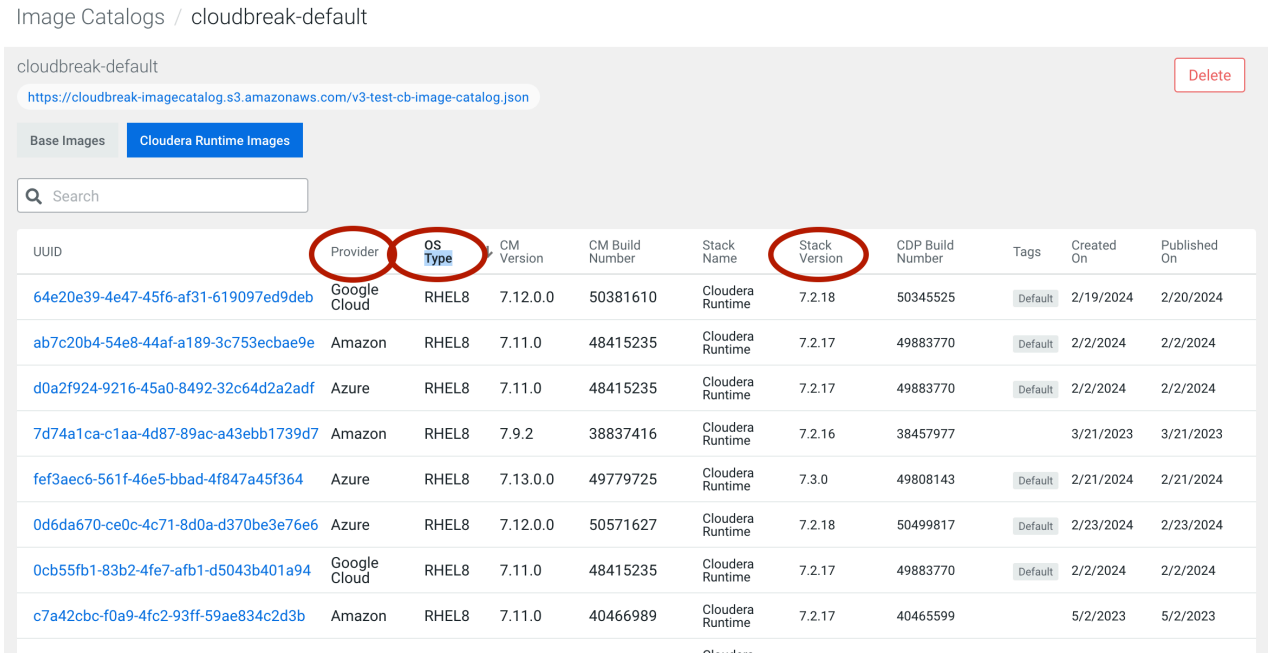
```
cdp datalake upgrade-datalake --datalake-name my-datalake --image-id 7.2
.11
cdp datalake upgrade-datalake --datalake-name my-datalake--runtime d1c52
0b1-987d-461f-7860-918f43994c04
cdp datalake upgrade-datalake --datalake-name my-datalake--lock-components
 --imageid imageid --runtime runtime
cdp datalake upgrade-datalake --datalake-name my-datalake --show-availabl
e-image-per-runtime --show-available-images
cdp datalake upgrade-datalake --datalake-name my-datalake --show-availab
le-image-per-runtime --dry-run
cdp datalake upgrade-datalake --datalake-name my-datalake --show-availabl
e-images --dry-run
```

Upgrading a Cloudera Data Hub cluster with the CDP CLI

You can initiate a Cloudera Data Hub cluster upgrade (either OS, Runtime, or both) with the CDP CLI. Using the same CLI command, you can also search for and validate available images to upgrade to, and generate JSON templates for specific upgrade scenarios.

Obtain image ID

If your Data Lake upgrade includes upgrading from CentOS to RHEL 8, prior to attempting an upgrade you need to obtain an ID of a target RHEL 8 image. You can obtain it from the image catalog by finding an image with your target Runtime version which has an OS Type of RHEL8.



Once you have identified the ID, you can provide it in the upgrade CLI command by using the --image-id flag.

Upgrade steps

The cdp datahub upgrade-cluster command includes the following options:

```
cdp datahub upgrade-cluster

          --cluster-name <value>
          [--image-id <value>]
          [--runtime <value>]
          [--lock-components | --no-lock-components]
          [--dry-run | --no-dry-run]
          [--show-available-images | --no-show-available-images]
          [--show-available-image-per-runtime | --no-show-available-image-
per-runtime]
          [--cli-input-json <value>]
          [--generate-cli-skeleton]
```

In order to use this command for upgrading from CentOS to RHEL, ensure to provide an image ID of a RHEL 8 image.

Important: The --runtime option does not upgrade the OS. Upgrading the OS is a separate process that requires specifying the --lock-components option.

Table 1:

Option	Description
--cluster-name (string)	Required. The name or CRN of the Cloudera Data Hub cluster to upgrade.
--image-id (string)	The ID of an image to upgrade to. If upgrading from CentOS to RHEL, ensure to provide a RHEL 8 image ID.
–runtime (string)	The Runtime version to upgrade to. When you specify the Cloudera Runtime version, the upgrade uses the latest image ID of the given Cloudera Runtime version from the same image catalog used for Cloudera Data Hub cluster creation.
--lock-components | --no-lock-components (boolean)	Use --lock components to perform an OS upgrade only.
--dry-run | --no-dry-run (boolean)	Checks the eligibility of an image to upgrade. Can be used in conjunction with any other parameter, returning the available image (with respect to image Id, Runtime or lock-components set) without performing any actions.
--show-available-images | --no-show-available-images (boolean)	Returns the list of images that are eligible to upgrade to.
--show-latest-available-image-per-runtime | --no-show-latest-available-image-per-runtime (boolean)	Returns the latest image that is eligible to upgrade to, for each Runtime version with at least one available upgrade candidate.

When you run the cdp datahub upgrade-cluster command to initiate an upgrade, you have one of three options:

1. Specify one of either --image-id, --runtime, or --lockComponents, which makes an explicit choice of the exact image, Cloudera Runtime (latest OS), or latest OS (same Runtime) for upgrade.
2. Specify both --image-id and --lockComponents,which specifies an image and ensures the image represents an OS only upgrade.
3. Specify none of the --image-id, --runtime, or --lockComponents parameters, which initiates a Cloudera Runtime/Cloudera Manager upgrade to the latest compatible version and OS image.

Outside of upgrade, you can use the following options:

```
--show-available-images/--no-show-available-images
--show-available-images-per-runtime/--no-show-available-images-per-runtime
```

```
--dry-run
```

Upgrade Data Lakes and all Cloudera Data Hub clusters to a version from which it is supported to upgrade to Cloudera Runtime 7.3.1

Before being able to upgrade to Cloudera Runtime 7.3.1, you need to upgrade Data Lakes and all Cloudera Data Hub clusters to a Cloudera Runtime 7.2.17 or 7.2.18 version from which it is supported to upgrade to Cloudera Runtime 7.3.1.

Important: This step is only required if your current Cloudera Runtime version is 7.2.17.100 or earlier.

As part of this upgrade step, you need to upgrade Data Lakes and Cloudera Data Hub clusters to the latest 7.2.17 service pack version from which upgrade to Cloudera Runtime 7.3.1 is supported.

Upgrading Cloudera Runtime 7.2.16 or earlier version to Cloudera Runtime 7.2.17

If the Cloudera Runtime version of your Data Lakes and Cloudera Data Hub clusters is 7.2.16 or earlier, perform steps 1 through 4 in Identify your upgrade path of the Upgrading to Cloudera Runtime 7.2.18 documentation. It is not required to upgrade to Cloudera Runtime 7.2.18, there is a direct upgrade path from Cloudera Runtime 7.2.17.200 and its service pack to Cloudera Runtime 7.3.1.

To be able to upgrade to Cloudera Runtime 7.3.1, you must perform the following steps:

1. Upgrade Data Lake and Cloudera Data Hub clusters to the latest service pack of Cloudera Runtime and OS current version.
2. Upgrade FreeIPA to RHEL 8 if using CentOS.
3. Upgrade Data Lake and Cloudera Data Hub clusters Runtime to 7.2.17.
4. Upgrade Data Lake and Cloudera Data Hub clusters OS to RHEL 8.

Upgrading Cloudera Runtime 7.2.17.0 or 7.2.17.100 to Cloudera Runtime 7.2.17 latest service pack

If the Cloudera Runtime version of your Data Lakes and Cloudera Data Hub clusters is 7.2.17.0 or 7.2.17.100, a direct upgrade to Cloudera Runtime 7.3.1 is not supported. To upgrade to a Cloudera Runtime 7.2.17 service pack from which the direct upgrade to Cloudera Runtime 7.3.1 is supported, perform the following steps:

1. Upgrade the Cloudera Runtime of your Data Lakes and Cloudera Data Hub clusters to the latest Cloudera Runtime 7.2.17.200 service pack.

   For more information, Performing a Cloudera Data Hub service pack upgrade.
2. If you are using CentOS, upgrade to RHEL 8.

   For more information, see Upgrading from CentOS to RHEL.

Resize to Enterprise Data Lake if using Medium Duty Data Lake

If you are upgrading from Cloudera Runtime 7.2.17.200-7.2.17.500 to Cloudera Runtime 7.3.1, you must resize your Data Lake from Medium Duty to Enterprise Data Lake as Medium Duty Data Lake is not supported starting with Cloudera Runtime 7.2.18.

Important: This step applies only to Data Lake clusters that have been upgraded to 7.2.17 and have a Medium Duty shape. If your cluster is already using RHEL 8 and has the Enterprise Data Lake shape then you are not required to perform this step. You also do not need to perform this if you are using a Light Duty Data Lake.

For information on why the Data Lake resize is needed, see Upgrading from Medium Duty to Enterprise Data Lake.

For information on how to perform the resize, see Data Lake resizing.

Prerequisites

Prior to resizing the Data Lake, ensure that the following are in place:

- Review the notes, limitations, and prerequisites listed in Data Lake resizing.

During this operation, the metadata maintained in your current Data Lake is automatically backed up, a new Enterprise Data Lake is created within the environment, and the metadata is automatically restored to this new cluster. As mentioned above, any custom cluster configuration that you previously made will need to be reapplied after the resize completes.

The maintenance window required for this operation depends on the size of your SDX metadata. When you open the Resize cluster UI, it will show you the estimated duration of the operation.

Changing the default Java version to 8

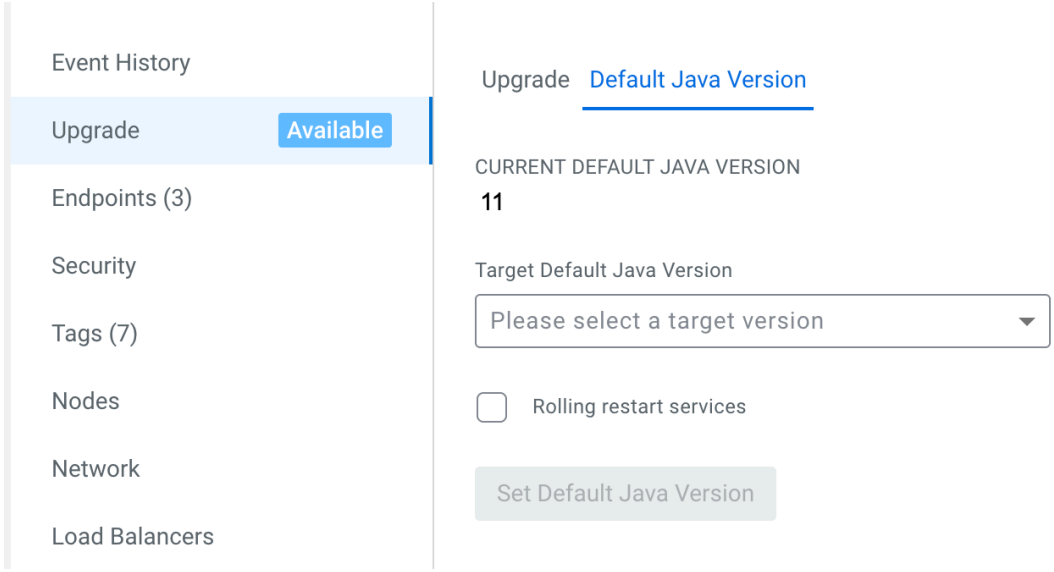
As JDK 11 is no longer supported in Cloudera Runtime 7.3.1, you need to change the default Java version to 8 to be able to upgrade to Cloudera Runtime 7.3.1.

Important: After changing the default Java version, Cloudera Manager and the Cloudera Runtime services will restart.

For Cloudera UI

You can change the default Java version from 11 to 8 for your Data Lake and Cloudera Data Hub cluster on Cloudera Management Console.

1. Navigate to the **Upgrade** tab of your Data Lake and Cloudera Data Hub cluster.
2. Select the Default Java Version tab.
3. Choose Java 8 as the **Target Default Java Version**.
4. Check the Rolling restart services box to enable rolling restart for the Data Lake and Cloudera Data Hub services.
5. Click Set Default Java Version.



For CDP CLI

Use the following command to change the default Java version for the Data Lake:

```
cdp datalake set-default-java-version --cluster-crn [***CRN***] --java-version 8
```

Use the following command to change the default Java version for the Cloudera Data Hub cluster:

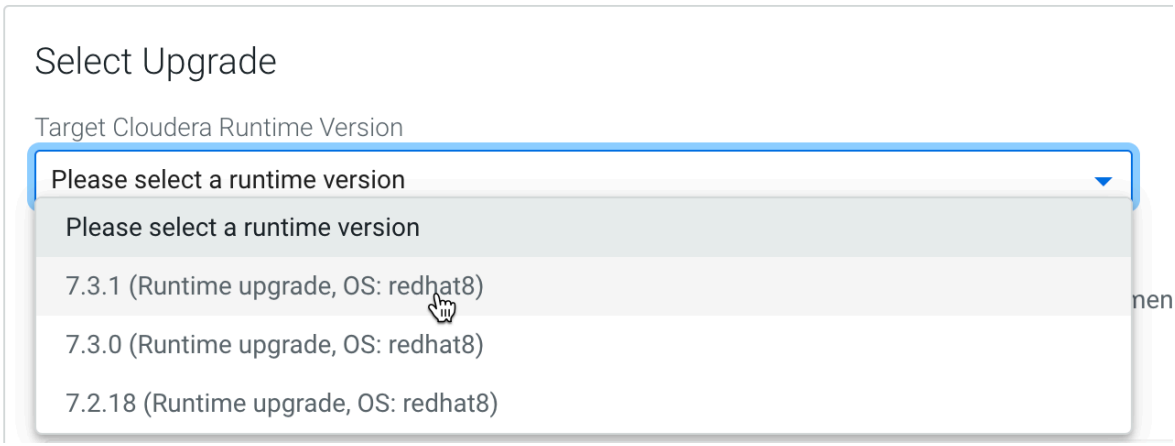
```
cdp datahub set-default-java-version --cluster-crn [***CRN***] --java-vers
ion 8
```

Upgrade Data Lake Cloudera Runtime version to 7.3.1 and OS to RHEL 8.10

From the supported Cloudera Runtime 7.2.17 or 7.2.18 versions, you need to upgrade Data Lake Cloudera Runtime version to 7.3.1 and OS to RHEL 8.10 in one step prior to upgrading Cloudera Data Hub clusters.

All Cloudera Data Hub clusters in your environment should be stopped and Data Services should not be running workloads for the duration of the operation.

- **1.** Once Runtime 7.3.1 (Runtime upgrade, OS: redhat8) is available to select from the Upgrade UI, select it and run Validate and Prepare.



Important: When selecting a main version, such as 7.3.1, for Cloudera Runtime version, you are upgraded to latest available Cumulative Hotfix (CHF) of the latest Service Pack (SP).

- **2.** Once Validate and Prepare completes, return to this page, select 7.3.1 (Runtime upgrade) again and run Upgrade.

Tip: When upgrading a Data Lake, Cloudera recommends allocating a 60-minute maintenance window for this upgrade step.

For more information, see Upgrading a Data Lake.

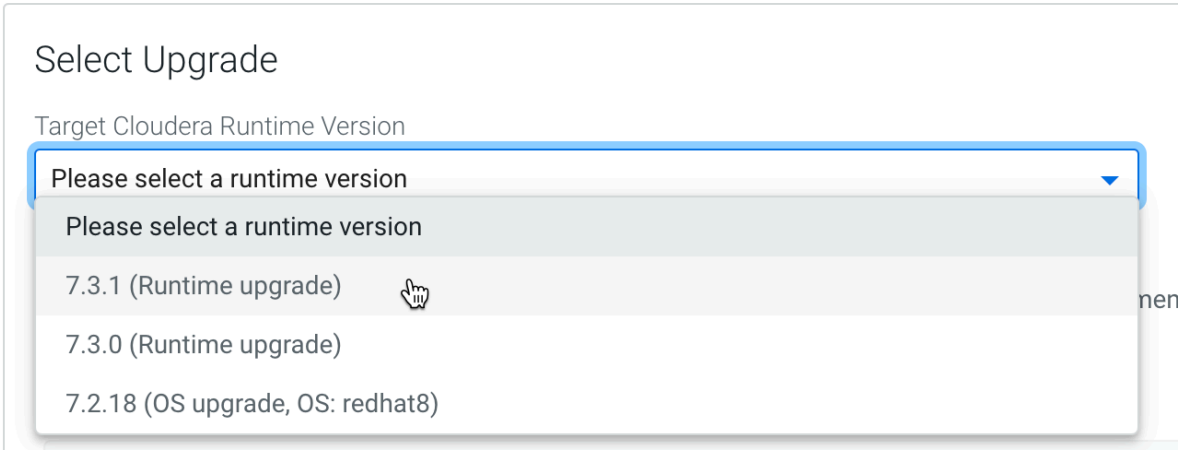
Upgrade Cloudera Runtime version of Cloudera Data Hub clusters to 7.3.1

Once the Data Lake Cloudera Runtime version has been upgraded to 7.3.1 and simultaneously, the OS has been upgraded to RHEL 8.10, the Cloudera Runtime version of each Cloudera Data Hub cluster needs to be upgraded to 7.3.1.

Note:

If you have a Cloudera Data Engineering Data Hub cluster, during the Cloudera Runtime upgrade process to version 7.3.1, you need to remove Spark 2 while creating a new cluster with Spark 3 and migrate your data. Depending on your environment and Spark applications, besides the cluster upgrade tasks, you need to perform Spark application migration tasks and sidecar migration tasks for the Cloudera Data Hub cluster in multiple steps. For more information, see Upgrading Apache Spark.

- **1.** Once Runtime 7.3.1 (Runtime upgrade) is available to select from the Upgrade UI, select it and run Validate and Prepare.



Important: When selecting a main version, such as 7.3.1, for Cloudera Runtime version, you are upgraded to latest available Cumulative Hotfix (CHF) of the latest Service Pack (SP).

2. Once Validate and Prepare completes, return to this page, select 7.3.1 (Runtime upgrade) again and run Upgrade.

Tip: When upgrading a Cloudera Data Hub cluster,allocate 90-120 minutes (as you will be performing this upgrade in two steps).

For more information, see Upgrading Cloudera Data Hub clusters.

Upgrade OS version of Cloudera Data Hub clusters to RHEL 8.10

During a Data Lake upgrade, the Data Lake OS is automatically upgraded to RHEL 8.10. In case of Cloudera Data Hub clusters, you need to upgrade the OS to RHEL 8.10 manually.

After performing Cloudera Runtime version upgrade of the Cloudera Data Hub cluster, open the Upgrade UI again and the 7.3.1 (OS upgrade, OS: redhat8)option will now be available in the drop-down menu.

1. Select 7.3.1 (OS upgrade, OS: redhat8) and run Validate and Prepare.
2. Once Validate and Prepare completes, return to this page, select 7.3.1 (OS upgrade, OS: redhat8) again and run Upgrade.

Upgrading to Cloudera Runtime 7.2.18

Cloudera Runtime 7.2.18 introduces significant new features and improvements. As a result, if you are planning to upgrade the Cloudera Runtime version in your existing Data Lake or Cloudera Data Hub clusters to 7.2.18, you might be required to perform this in multiple steps.

Some of the important Cloudera Runtime 7.2.18 features include:

- RHEL 8 images as CentOS reaches End of Life (EOL)

  As of June 30, 2024, CentOS reaches End of Life (EOL), and consequently, Cloudera Runtime 7.2.18 supports RHEL 8-based images only. New deployments of Data Lakes and Cloudera Data Hub clusters with Cloudera Runtime 7.2.18 and upgrades to 7.2.18 are only possible with RHEL 8. Data Lake and Cloudera Data Hub clusters running Cloudera Runtime 7.2.17 support both CentOS 7 and RHEL 8. Earlier Cloudera Runtime versions support CentOS 7 only. Cloudera will not publish any updates or fixes for CentOS-based images after June 2024.

  For Azure environments, Cloudera only provides RHEL 8 images via Azure Marketplace.

- Discontinuation of Medium Duty Data Lake

  Starting with Cloudera Runtime 7.2.18, Medium Duty Data Lake is discontinued as it does not support rolling upgrades and horizontal scaling, which have been introduced with the Enterprise Data Lake . While earlier Cloudera Runtime versions support Medium Duty Data Lake (7.2.17 supports Enterprise, Medium Duty, Light Duty, and older versions support Medium Duty and Light Duty), 7.2.18 only supports Enterprise and Light Duty Data Lakes.

Use the following documentation to identify your Data Lake cluster's version details and then based on that determine your upgrade path:

Identify cluster version details

Identify your Data Lake and Cloudera Data Hub cluster's Cloudera Runtime version, Data Lake, Cloudera Data Hub, and FreeIPA image version and operating system (OS) type, and Data Lake shape.

Your upgrade path to versions 7.2.18 or 7.3.1 may depend on the current Cloudera Runtime version, image version, and OS of your cluster. Cloudera recommends that you identify your version information as follows:

1. Identify cluster's Cloudera Runtime version on page 34
2. Identify cluster's image version, image creation date, OS type on page 35
3. Identify Data Lake shape on page 37

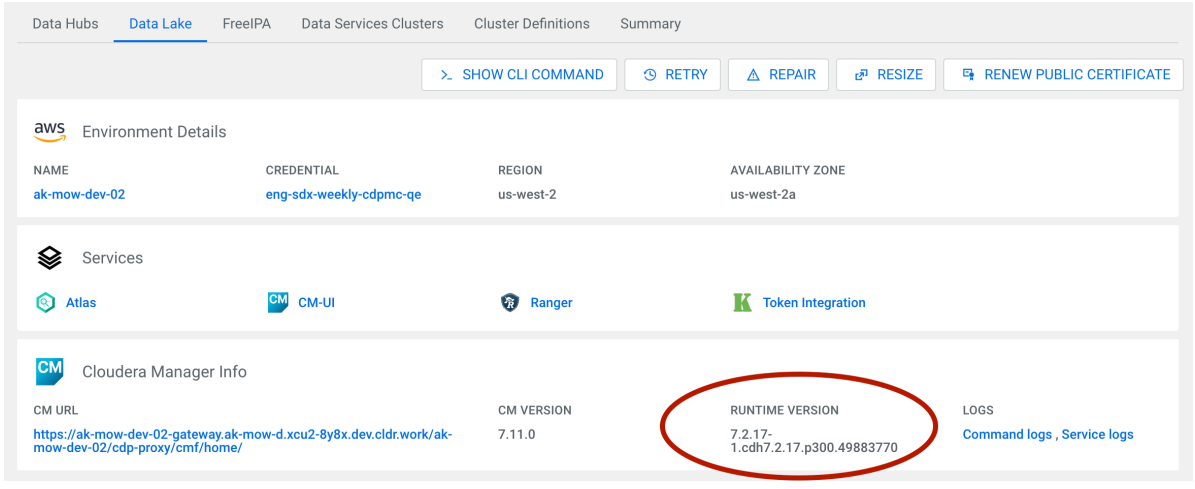
Identify cluster's Cloudera Runtime version

Follow these instructions to identify your Data Lake and Cloudera Data Hub's Cloudera Runtime version. You should follow these steps separately for each cluster.

Steps

For Cloudera UI

1. In the Cloudera Management Console, navigate to the details page of the Data Lake or Cloudera Data Hub cluster that you would like to upgrade.
2. In the Cloudera Manager Info section find the field called "Runtime version":



For CDP CLI

Get the Cloudera Runtime version of your Data Lake from the CDP CLI by using the following command:

```
cdp datalake describe-datalake --datalake-name <DL-NAME>
```

Get the Cloudera Runtime version of your Cloudera Data Hub from the CDP CLI by using the following command:

```
cdp datahub describe-cluster --cluster-name <DH-NAME>
```

The version information is displayed in the following format:

7.2.17-1.cdh7.2.17.p200.46967063

The highlighted digits show the current Cloudera Runtime version and the number after p indicates the current Service Pack version. In the above example, the cluster is using Cloudera Runtime version 7.2.17.200.

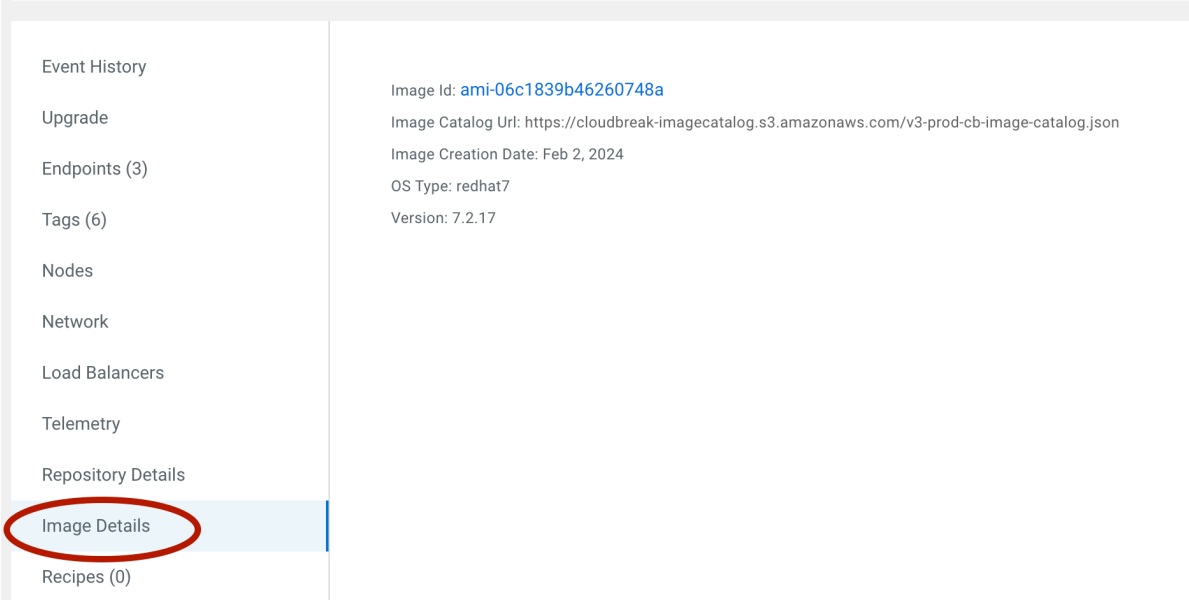
Identify cluster's image version, image creation date, OS type

Follow these instructions to identify Data Lake and Cloudera Data Hub image version and OS type. You should follow these steps separately for each cluster.

Steps

For Cloudera UI

1. In the Cloudera Management Console, navigate to the details page of the Data Lake or Cloudera Data Hub cluster that you would like to upgrade.
2. Scroll down to access the left navigation pane.
3. From the left navigation pane, select Image Details:



4. You will see information similar to the following:

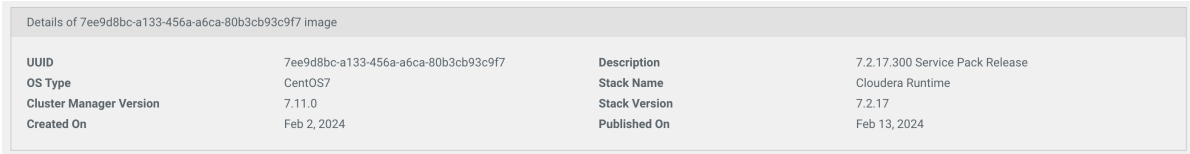
```
Image Id: ami-0e95ecddb7fc75ea8
Image Catalog Url: https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-
prod-cb-image-catalog.json
Image Creation Date: Jan 9, 2024
OS Type:redhat7
Version: 7.2.17
```

Under OS Type:

- If you see "redhat7", your cluster is using a CentOS 7 OS image. This is a known issue.

- If you see “redhat8”, your cluster is using a RHEL 8 OS image.

Clicking on image ID will redirect you to the image catalog and provide you with further information, including the actual OS Type:



For CDP CLI

Get Data Lake’s image version from the Cloudera CLI by using the following command:

```
cdp datalake upgrade-datalake --show-latest-available-image-per-runtime --
datalake-name <DL-NAME>
```

Get Cloudera Data Hub cluster’s image version from the CDP CLI by using the following command:

```
cdp datahub upgrade-cluster --show-latest-available-image-per-runtime --
cluster-name <DH-NAME>
```

The output of this command looks like this, providing you with the OS information:

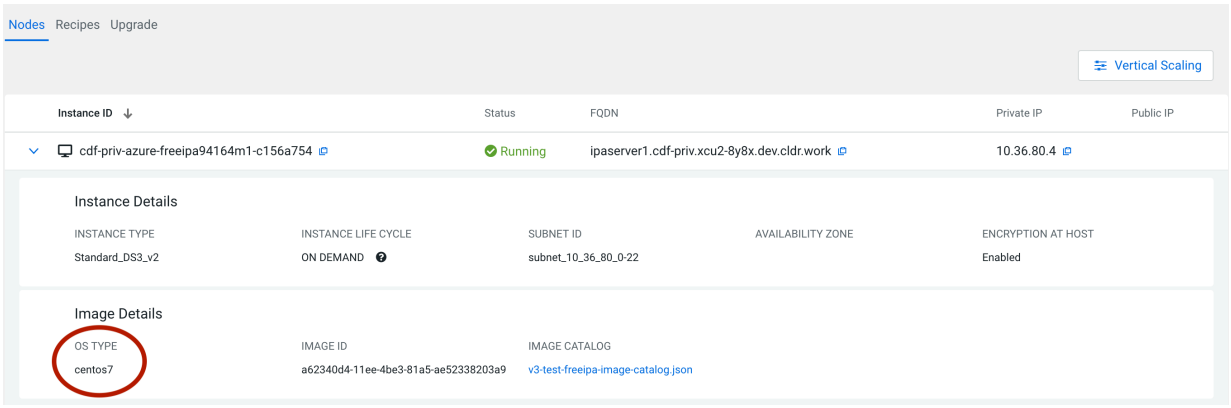
```
{
 "targetImage": {
 "catalog": "https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-prod-fr
eeipa-image-catalog.json",
 "id": "a5b2580e-7487-4ee6-a80f-d991fa00cd13",
 "os": "redhat8",
 "imageName": "ami-0935d849de6a25e45",
 "date": "2024-01-09"
 },
 "originalImage": {
 "catalog": "https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-prod-f
reeipa-image-catalog.json",
 "id": "e1f92688-3773-47dd-89ff-561933540108",
 "os": "centos7",
 "imageName": "ami-0db1b7000c8a27fd4",
 "date": "2023-11-22"
 },
 "operationId": "3a50f30f-33f3-4f81-b9bb-30c119023ff0"
}
```

Follow these instructions to identify FreeIPA OS type.

Steps

1. In the Cloudera Management Console, navigate to the environment details page,.
2. Click on the FreeIPA tab.
3. In the Nodes tab (which is open by default), expand one section corresponding to an instance ID.

4. Under Image Details, you can find the OS Type:

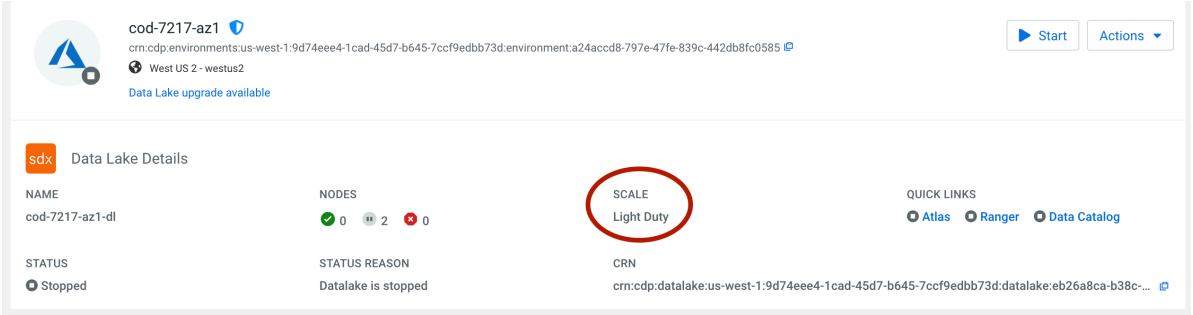


Identify Data Lake shape

Follow these instructions to identify Data Lake's shape.

For Cloudera UI

1. In the Cloudera Management Console, navigate to the details page of the Data Lake that you would like to upgrade:



2. Find the Scale field. There are three possible values: "Light Duty", "Medium Duty HA" and "Enterprise Data Lake".

For CDP CLI

Get the Data Lake shape from the CDP CLI by using the following command:

```
cdp datalake describe-datalake –datalake-name <DL-NAME>
```

Find the "scale" field in the output of this command. There are three possible values: "Light Duty", "Medium Duty HA" and "Enterprise Data Lake".

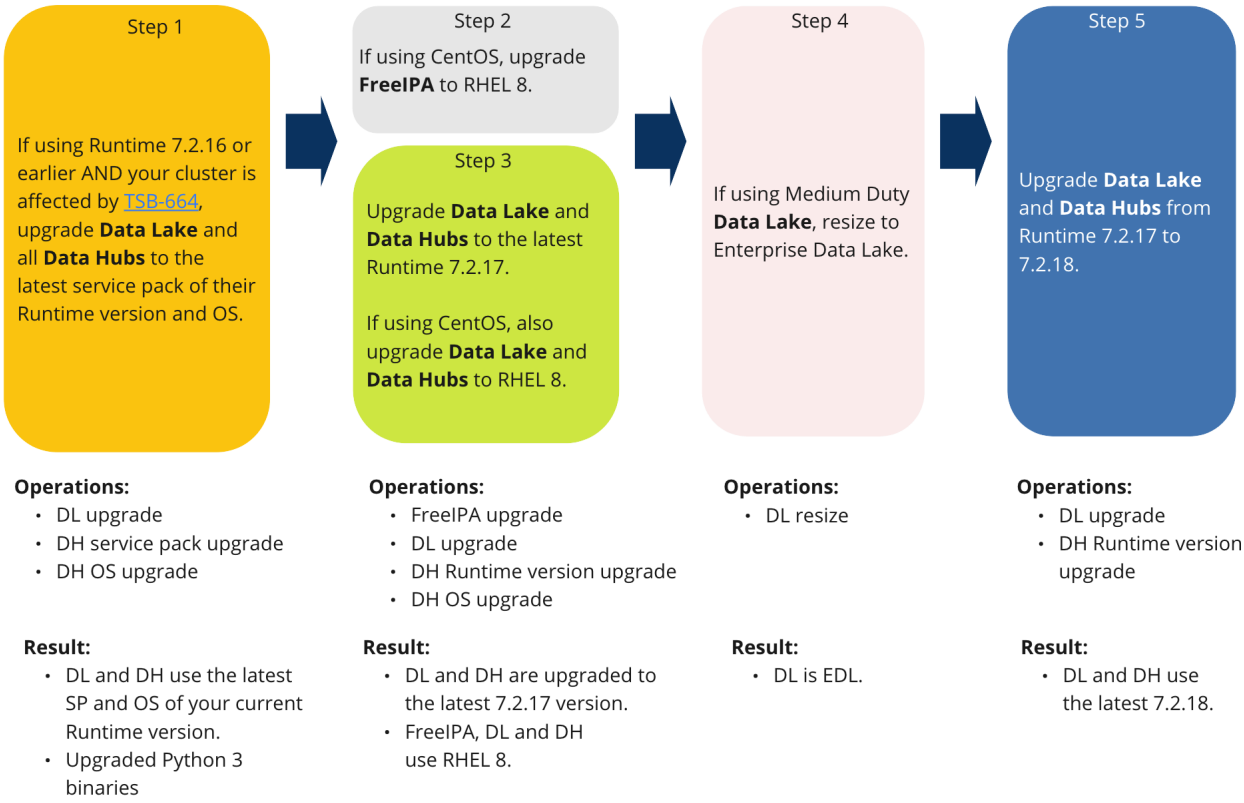
Identify your upgrade path

Refer to the following diagram for a high-level overview of the upgrade steps. Your starting point will vary, depending on the Runtime version and image version that your environment, Data Lake, and Cloudera Data Hub clusters are using.

The high-level upgrade steps are presented horizontally, from step 1 to step 5 and under each step you can see:

- Operations: A list of the actual operations that are performed as part of this step
- Result: The resulting state after these operations have been performed

For example, if your clusters are using Cloudera Runtime 7.2.16 and if Step 1 is applicable to your clusters, you would be performing a Data Lake upgrade and for each of your Cloudera Data Hub clusters a service pack upgrade and an OS upgrade in Step 1. If you started with Cloudera Runtime 7.2.16, after performing the operations in Step 1 you would have your Data Lake and all Cloudera Data Hub clusters upgraded to the latest service pack of Cloudera Runtime 7.2.16 and your cluster OS would be patched with up-to-date Python 3 binaries.



As presented in the diagram, you may be required to perform the following operations:

- Runtime service pack upgrades for your Data Lake and each Cloudera Data Hub cluster
- Major/Minor Runtime and OS upgrades for your Data Lake and each Cloudera Data Hub cluster
- CentOS to RHEL 8 upgrade for your FreeIPA, Data Lake, and each Cloudera Data Hub cluster
- Data Lake resize

For example, if your clusters are using Cloudera Runtime 7.2.16 or earlier and are impacted by Python 3.8 dependency described in TSB-664, you need to start at Step 1, but if this does not apply to you, you may start at a later step. Furthermore, depending on the operating system that your clusters are using, the CentOS to RHEL 8 upgrade steps (which are part of in Step 2) may not apply to you. and if you are not using Medium Duty Data Lake, you skip Enterprise Data Lake resize step (Step 4).

Important:

- Each of these operation has its own prerequisites and postrequisites so in addition to the actual upgrade steps, the prerequisites and postrequisites need to be performed as well.
- You will need to upgrade all Data Lakes and Cloudera Data Hub clusters separately, first the Data Lakes and then each Cloudera Data Hub cluster.
- You can perform each upgrade operation either from the Upgrade UI of the respective cluster or using the CDP CLI.
- The upgrade user interface in the Cloudera Management Console shows you the next Data Lake and Cloudera Data Hub upgrade step available, so you can refer to it as a guideline.

USEFUL LINKS:

- Data Lake upgrade

- Cloudera Data Hub upgrade
- Rolling upgrades - Data Lake
- Rolling upgrades - Cloudera Data Hub
- Data Lake resizing
- Upgrading from CentOS to RHEL

Review the prerequisites

Before you start upgrading your clusters, ensure that you meet the prerequisites for each upgrade that is required.

Refer to the following documentation for detailed prerequisites:

- Review Data Lake Upgrade: Before you begin and Cloudera Data Hub Upgrade: Limitations and Prerequisites.
- If you are upgrading a Data Lake, use the Backup and Restore functionality. This will allow you to restore the SDX metadata if required.
- Certain Data Hubs might require certain pre-upgrade and post-upgrade steps. Review the related documentation before upgrading Data Hubs.
- If you are upgrading from CentOS to RHEL, review the Prerequisites for upgrading from CentOS to RHEL.
- If your upgrade path involves a Data Lake resize, refer to Review Data Lake resizing: Prerequisites.

High-level upgrade steps

Refer to this section for high-level upgrade steps and links to more detailed upgrade documentation. This document is not meant to provide you with complete upgrade steps, but rather it outlines the general upgrade path and provides you with the links to existing upgrade documentation.

Important:

- Each of these operations has its own prerequisites and postrequisites so in addition to the actual upgrade steps, the prerequisites and postrequisites need to be performed as well.
- You will need to upgrade all Data Lakes and Cloudera Data Hub clusters separately, first the Data Lakes and then each Data Hub.
- You can perform each upgrade operation either from the Upgrade UI of the respective cluster or using the CDP CLI.
- The upgrade user interface in the Cloudera Management Console shows you the next Data Lake and Cloudera Data Hub upgrade step available, so you can refer to it as a guideline.

1. Upgrade Data Lake and all Cloudera Data Hub clusters to the latest service pack and OS
2. Upgrade FreeIPA to RHEL 8
3. Upgrade Data Lake and Cloudera Data Hub clusters to the latest Cloudera Runtime 7.2.17 and the latest RHEL 8 OS image
4. Resize Medium Duty Data Lake to Enterprise Data Lake
5. Upgrade Data Lake and Cloudera Data Hub clusters from 7.2.17 to 7.2.18

Step 1: Upgrade Data Lake and all Cloudera Data Hub clusters to the latest service pack and OS

Upgrade Data Lake and Cloudera Data Hub to the latest service pack of their Cloudera Runtime version.

Important:

This step applies only to Data Lake and Cloudera Data Hub clusters running Cloudera Runtime 7.2.16 or earlier AND using an OS image that has been published before March 2023 (that is, clusters impacted by Python 3.8 dependency described in TSB-664).

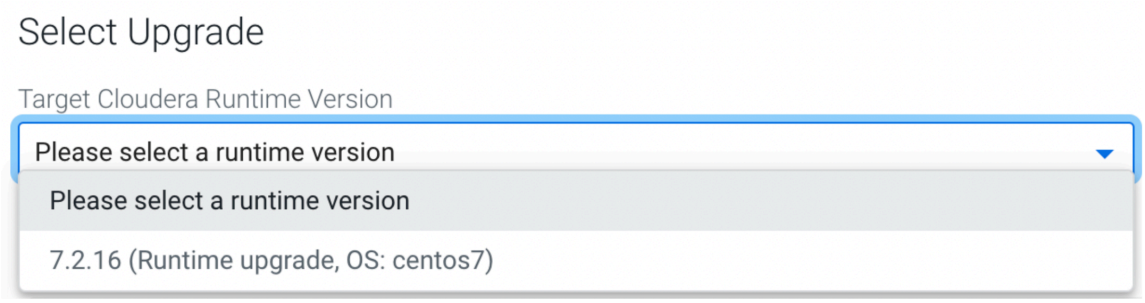
Another way to identify whether you are required to perform this step is opening the Upgrade UI and verifying that 7.2.17 is available from the drop-down as an upgrade target. If it is, you can skip to Step 2.

Follow these steps first for your Data Lake and then for each attached Cloudera Data Hub separately. The examples and screenshots below assume that your cluster is using Cloudera Runtime 7.2.16. For older Cloudera Runtime, just replace 7.2.16 with your actual Cloudera Runtime version.

1. Upgrade Data Lake and Cloudera Data Hub to the latest service pack of their Runtime version.

   To upgrade your Data Lake and all Cloudera Data Hub clusters to the latest service pack of their Runtime version, perform the following for your Data Lake and each Cloudera Data Hub from the Upgrade UI each cluster:

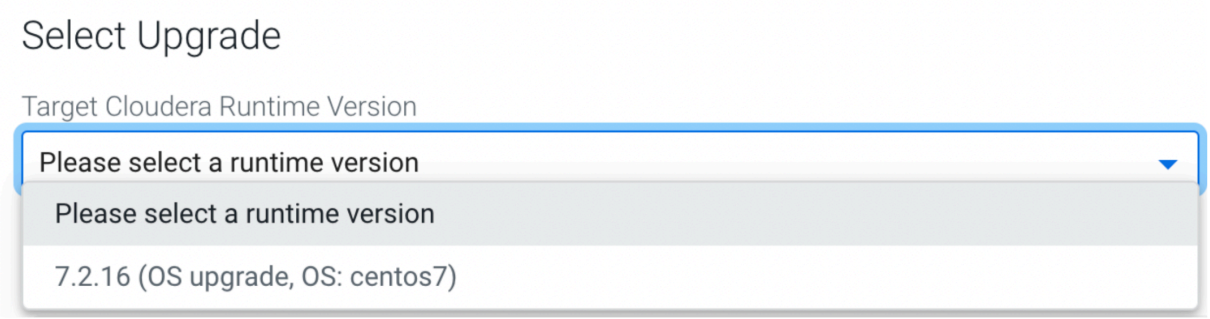
   a. Select the upgrade target that matches your current Cloudera Runtime version. If you are upgrading a Data Lake, you will see 7.2.16 (Cloudera Runtime upgrade, OS; centos7); for Cloudera Data Hub clusters you will see Select 7.2.16 (Cloudera Runtime upgrade). Select this line and then run Validate and Prepare.
   b. Once Validate and Prepare completes, return to this page, select the same line again and run Upgrade:



   For Data Lakes, this will perform the OS image upgrade, so step 2 below is required for Cloudera Data Hub clusters only.
2. Upgrade the OS image on all Cloudera Data Hub clusters.

   Additionally, you are required to upgrade the OS image on all Cloudera Data Hub clusters:

   a. Open the Cloudera Data Hub Upgrade UI again and select the upgrade target that matches your current Cloudera Runtime version and run Validate and Prepare. Select 7.2.16 (OS upgrade, OS; centos7).
   b. Once Validate and Prepare completes, return to this page, select the same lineagain and run Upgrade.



Tip:

When upgrading a Data Lake, we recommend allocating a 60 minutes maintenance window for this upgrade step. All Cloudera Data Hub clusters in your environment should be stopped and Data Services should not be running workloads for the duration of the operation.

When upgrading a Cloudera Data Hub, allocate 90-120 minutes (as you will be performing this upgrade in two steps).

USEFUL LINKS:

- Upgrading a Data Lake (UI)
- Upgrading a Data Lake (CLI)
- Performing a Cloudera Data Hub service pack upgrade (UI)
- Performing a Cloudera Data Hub OS upgrade (UI)

- Upgrading a Cloudera Data Hub with CLI

Step 2: Upgrade FreeIPA to RHEL 8

Before your Data Lake or Cloudera Data Hub clusters can be upgraded to RHEL 8, you first need to upgrade your environment (FreeIPA) cluster.

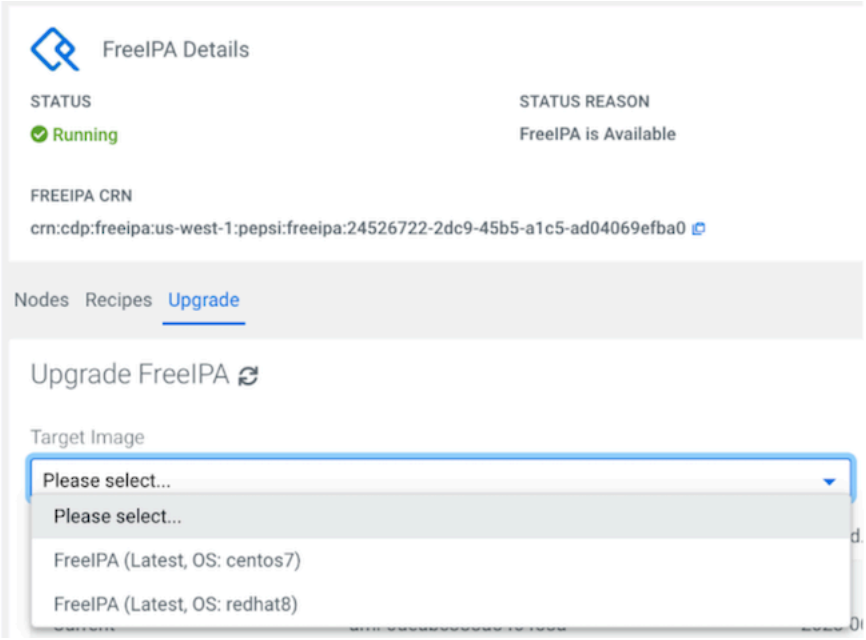
Important:

This step applies only to FreeIPA clusters using a CentOS 7 image.

Open the FreeIPA tab on your environment's details page and select Upgrade. From the drop-down menu select FreeIPA (Latest, OS: redhat8) and run the upgrade:

Note:

If in the drop-down selector described below you do not see the option FreeIPA (Latest, OS: centos7), then your cluster is already using RHEL 8 and you can skip step completely.



Tip:

We recommend allocating a 120-minute maintenance window for this upgrade step for a 3-node FreeIPA cluster (1-node and 2-node clusters will take less time).

USEFUL LINKS:

- Upgrading FreeIPA (UI and CLI)

Step 3: Upgrade Data Lake and Cloudera Data Hub clusters to the latest Cloudera Runtime 7.2.17 and the latest RHEL 8 OS image

Upgrade your clusters to the latest 7.2.17 and to RHEL 8.

Important:

This step applies only to Data Lake and Cloudera Data Hub clusters using a CentOS 7 image. If your clusters are already using Cloudera Runtime 7.2.17 and a RHEL 8 image, then you are not required to perform this step.

Perform the following steps for the Data Lake and then each Cloudera Data Hub:

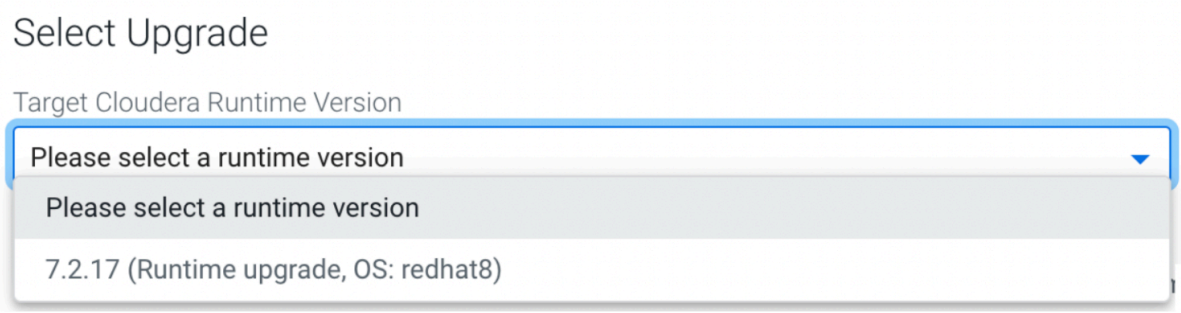
1. Understand RHEL 8 requirements.

   Before performing the OS version change from CentOS 7 to RHEL 8 make sure you have understood the requirements mentioned in Upgrading from CentOS to RHEL.
2. Upgrade to the latest 7.2.17 service pack if needed.

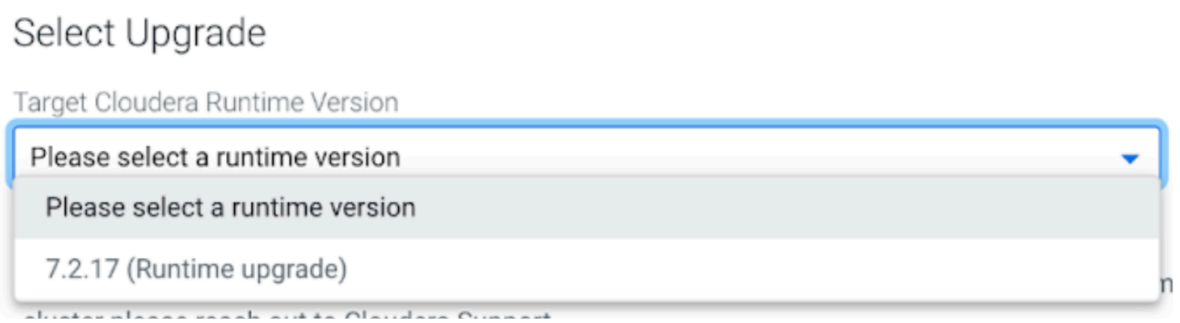
   To determine whether you need to perform a Cloudera Runtime and OS upgrade, or only an OS upgrade, check whether you see the target version in the drop-down labeled 7.2.17 (OS upgrade, OS: redhat8). If you do, you only need an OS upgrade. and you can skip to step 3 below. If you do not see it, perform the following steps for a Cloudera Runtime upgrade:

   a. If you are upgrading a Data Lake, select 7.2.17 (Cloudera Runtime upgrade, OS: redhat8). For upgrading a Cloudera Data Hub, select 7.2.17 (Cloudera Runtime upgrade). Next, run Validate and Prepare.
   b. Once Validate and Prepare completes, return to this page, select the same line again, and run Upgrade.

   Data Lake UI

   Cloudera Data Hub UI

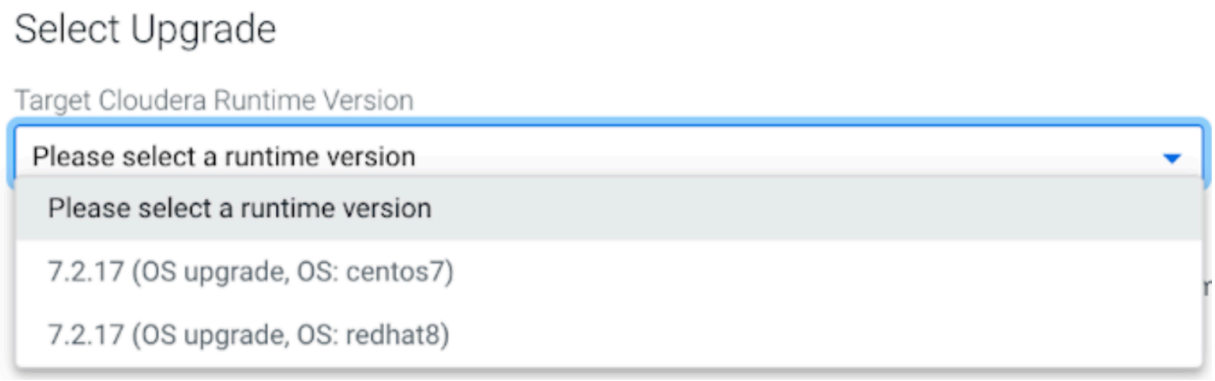
   

   For Data Lakes this will perform the RHEL 8 upgrade as well.

3. Upgrade all Cloudera Data Hub clusterss to RHEL.

   Open the Upgrade UI again and the 7.2.17 (OS upgrade, OS: redhat8) option will now be available in the drop-down menu.

   a. Select Select 7.2.17 (OS upgrade, OS: redhat8) and run Validate and Prepare.
   b. Once Validate and Prepare completes, return to this page, select Select 7.2.17 (OS upgrade, OS: redhat8) again and run Upgrade.



Tip: When upgrading a Data Lake, all Cloudera Data Hub clusters in your environment should be stopped and Cloudera data services should not be running workloads for the duration of the operation. We recommend allocating a 90-120 minute maintenance window for this upgrade step.

USEFUL LINKS:

- Upgrading a Data Lake (UI)
- Upgrading a Data Lake (CLI)
- Performing a Cloudera Data Hub OS upgrade
- Upgrading a Cloudera Data Hub with CLI

Step 4: Resize Medium Duty Data Lake to Enterprise Data Lake

If using a Medium Duty Data Lake, you should resize it to Enterprise Data Lake.

Important: This step applies only to Data Lake clusters that have been upgraded to 7.2.17 and have a Medium Duty shape. If your cluster is already using RHEL 8 and has the Enterprise Data Lake shape then you are not required to perform this step. You also do not need to perform this if you are using a Light Duty Data Lake.

Before you begin, note the following:

- Prior to attempting the resize, ensure that all Cloudera Data Hub clusters in your environment have been upgraded to Cloudera Runtime 7.2.17 and your data services have been upgraded to the latest version available.
- During the resize operation your cluster's Cloudera Manager and Cloudera Runtime configuration will be updated to the most recent recommendations (performance tuning, etc). You will be required to re-apply these custom configurations after you have performed the resize to Enterprise Data lake.
- There might be a requirement to vertically scale the Data Lake nodes to increase the available resources in case high resource utilization is observed after Data Lake resize to Enterprise and Operating System upgrade to RHEL from CentOS.
- The resize operation will perform an SDX backup and restore in the background. This involves writing and reading from cloud storage. If you have not yet used data lake backup, please make sure your environment's storage permissions are configured correctly for these operations.
- Ensure that all Data Hubs in your environment have been upgraded to Runtime 7.2.17 and your data services have been upgraded to the latest version available.
- The resizing operation requires a downtime and should be performed during a maintenance window. No metadata changes may occur during the resizing, as these changes will no longer be present once the resizing operation

completes (the previously backed up metadata is being restored). Suspend any operations that may result in any SDX metadata change during the resizing operation.
- Cloudera Data Hub clusters should be stopped before the resizing operation begins. For any cluster that cannot be stopped, stop all of the services on the Cloudera Data Hub through the Cloudera Manager UI.
- With Cloudera DataFlow 2.0 or lower, some flows must be re-created after a resizing operation.
- Review Data Lake resizing: Prerequisites.

Follow the steps for Data Lake resizing. During this operation, the metadata maintained in your current Data Lake is automatically backed up, a new Enterprise Data Lake is created within the environment, and the metadata is automatically restored to this new cluster. As mentioned above, any custom cluster configuration that you previously made will need to be reapplied after the resize completes.

The maintenance window required for this operation depends on the size of your SDX metadata. When you open the Resize cluster UI, it will show you the estimated duration of the operation.

USEFUL LINKS:

- Data Lake resizing prerequisites and limitations
- Resizing the Data Lake (UI)
- Resizing the Data Lake (CLI)

Step 5: Upgrade Data Lake and Cloudera Data Hub clusters from 7.2.17 to 7.2.18

Perform the following steps for your Data Lake and each Cloudera Data Hub:

If you are upgrading a Data Lake, all Cloudera Data Hub clusters in your environment should be stopped and Data Services should not be running workloads for the duration of the operation.

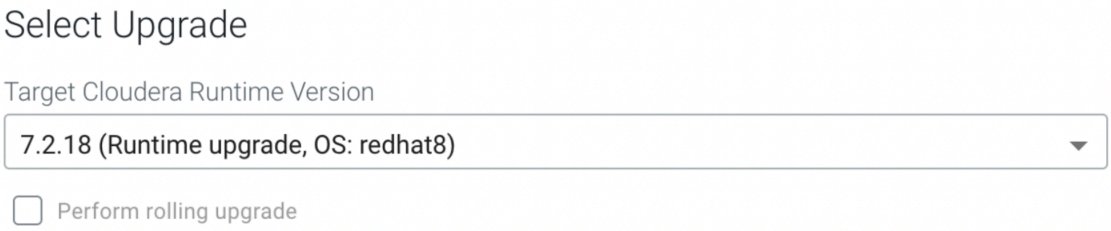
1. Once Cloudera Runtime 7.2.18 (Cloudera Runtime upgrade, OS: redhat8) is available to select from the Upgrade UI, select it and run Validate and Prepare.
2. Once Validate and Prepare completes, return to this page, select 7.2.18 (Cloudera Runtime upgrade, OS: redhat8) again and run Upgrade.

Attention:

Enterprise Data Lakes created with Cloudera Runtime 7.2.17.300 or newer and a RHEL 8 image can be directly upgraded in a rolling manner to 7.2.18. This allows the upgrade operation to be performed without stopping attached Cloudera Data Hub clusters and workloads on any data services that are running in the environment. To determine if your cluster is eligible, refer to Data Lake rolling upgrades and Cloudera Data Hub rolling upgrades documentations.

Cloudera recommends that you perform the rolling upgrade during a maintenance window, when no critical workloads are running on your attached Cloudera Data Hub clusters or data services.

To perform a rolling upgrade of the Data Lake, select “Rolling Upgrade”:



Tip:

We recommend allocating a 60 minutes maintenance window for this upgrade step

USEFUL LINKS:

- Upgrading a Data Lake (UI)
- Upgrading a Data Lake (CLI)
- Performing a Cloudera Data Hub major/minor version upgrade (UI)

- Upgrading a Cloudera Data Hub with CLI
- Rolling upgrades - Data Lake
- Rolling upgrades - Cloudera Data Hub

Upgrading from CentOS to RHEL

As part of FreeIPA, Data Lake, and Cloudera Data Hub upgrade, you have the option to upgrade the operating system (OS) on the virtual machines (VMs) from CentOS 7 to Red Hat Enterprise Linux 8 (RHEL 8).

As of June 30, 2024, CentOS reaches End of Life (EOL), and consequently, Cloudera Public Cloud 7.2.18 only supports RHEL 8-based images. New deployments of Data Lakes and Cloudera Data Hub clusters with Cloudera Runtime 7.2.18 and upgrades to 7.2.18 are only possible with RHEL 8. Data Lake and Cloudera Data Hub clusters running Cloudera Runtime 7.2.17 support both CentOS 7 and RHEL 8. Earlier Cloudera Runtime versions support CentOS 7 only. Cloudera will not publish any updates or fixes for CentOS-based images after June 2024.

You do not need to purchase a RHEL license for the upgrade. The license is covered by Cloudera.

Cloudera recommends that you perform the CentOS to RHEL in the following sequence:

1. Upgrade FreeIPA via the FreeIPA upgrade web UI or CLI.
2. Upgrade Data Lake via the Data Lake upgrade web UI or CLI.
3. Upgrade Cloudera Data Hub clusterss via the Cloudera Data Hub OS upgrade web UI or CLI.

Attention: Your Data Lake or Cloudera Data Hub must be using Cloudera Runtime 7.2.17 in order to access the CentOS to RHEL upgrade option.

Prerequisites for upgrading from CentOS to RHEL

Prior to attempting the CentOS to RHEL upgrade, ensure that you have taken care of the following prerequisites:

- Rewrite and test recipes if needed
- Check compatibility of third party software you might have installed on the nodes
- Opt in to use Azure Marketplace images (Azure only)

Rewrite and test recipes

If you are using FreeIPA, Data Lake, or Cloudera Data Hub recipes, you must ensure that they are compatible with RHEL. As some CentOS-specific commands may not work on RHEL, you may need to rewrite them. The steps are:

1. Rewrite your existing CentOS recipes so that they are compatible with RHEL.
2. Test the updated recipes on a newly created FreeIPA, Data Lake, or Cloudera Data Hub VM that uses RHEL operating system.
3. Detach your existing CentOS recipes and attach your updated RHEL recipes. You can find the steps here:
   a. Update FreeIPA recipes
   b. Update Data Lake recipes
   c. Update Cloudera Data Hub recipes

If after rewriting and testing your recipes are not working, contact Cloudera support for help.

Check compatibility of third party software

If you normally install additional software on your Data Lake, FreeIPA, or Cloudera Data Hub VMs (using recipes or manually), you may need to contact Cloudera support to make additional OS packages available. You should file a support ticket and provide details of the additional packages that you are installing (software name, version, and software URL).

Opt in to use Azure Marketplace images

Important:

To be able to upgrade Azure clusters to RHEL 8, you must first accept Azure Marketplace terms and conditions via Azure CLI or Cloudera web UI. See Cloudera images hosted in Azure Marketplace.

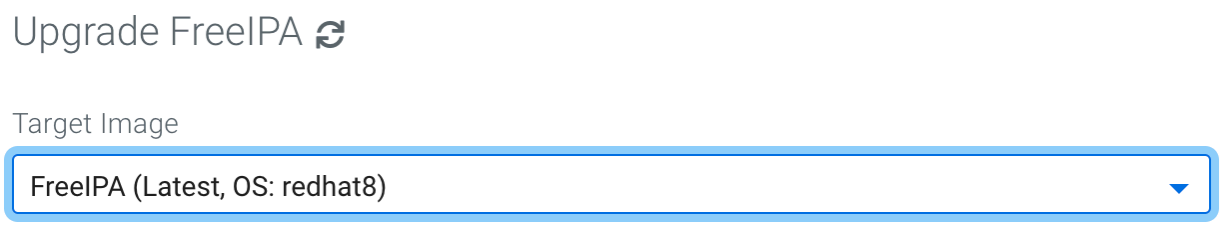
RHEL 8 images are only available through Azure Marketplace. In order for Cloudera to be able to load Cloudera-published virtual machine images in your subscription from the Azure Marketplace, you must first accept Azure Marketplace terms and conditions either via Cloudera web UI or Azure CLI . If you do not accept the terms and conditions, Cloudera will not be able to access the images hosted in Azure Marketplace and consequently environment creation will fail. For more information, see Cloudera images hosted in Azure Marketplace.

Upgrade clusters to RHEL

To upgrade your environment, Data Lake, and Cloudera Data Hub clusters to RHEL, follow the usual operating system (OS) upgrade documentation for FreeIPA, Data Lake, and Cloudera Data Hub.

Upgrade FreeIPA cluster to RHEL

When upgrading via the web interface, you navigate to the FreeIPA tab on your environment's details page and select Upgrade. Next, from the drop-down menu you select FreeIPA (Latest, OS: redhat8) and run the upgrade:



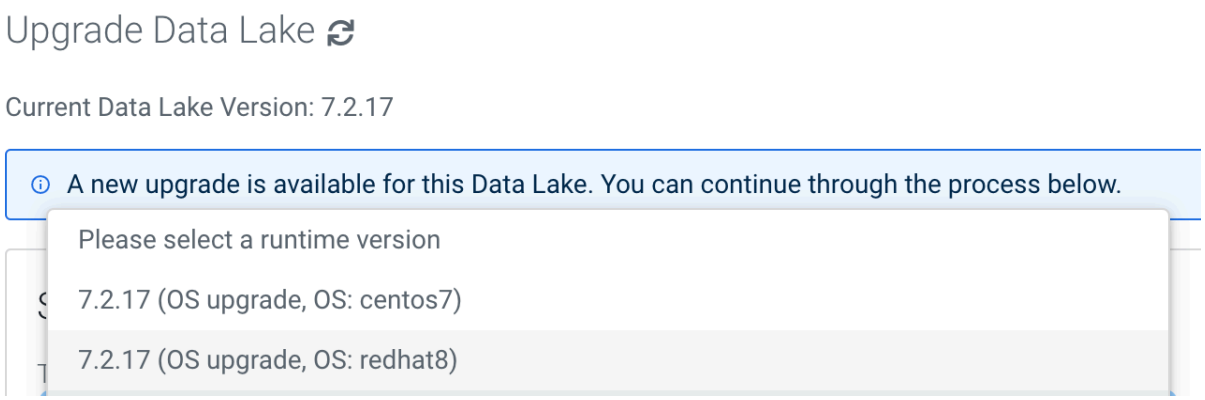
When upgrading via CDP CLI, you use the typical upgrade command and specify a RHEL image available for your cloud platform.

For detailed Cloudera UI and CLI instructions, see Upgrade FreeIPA.

Upgrade Data Lake to RHEL

When upgrading via the web interface, you navigate to your Data Lake's Upgrade page and perform the following steps:

1. Select 7.2.17 (OS upgrade, OS: redhat8) from the dropdown and run Validate and Prepare:



2. Once Validate and Prepare completes, you return to this page, select 7.2.17 (OS upgrade, OS: redhat8) again, and run Upgrade.

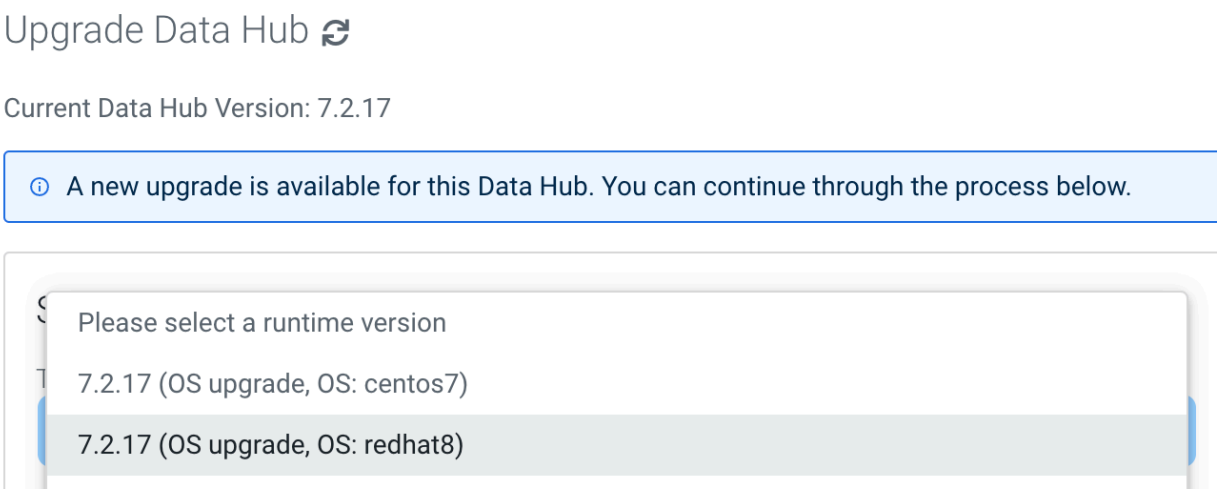
When upgrading via CDP CLI, you use the typical upgrade command and specify a RHEL image available for your cloud platform.

For full documentation including prerequisites and UI/CLI instructions, refer to Data Lake upgrade.

Upgrade Cloudera Data Hubs to RHEL

When upgrading via the web interface, you navigate to your Cloudera Data Hub's Upgrade page and perform the following steps:

1. Select 7.2.17 (OS upgrade, OS: redhat8) from the dropdown and run Validate and Prepare.



2. Once Validate and Prepare completes, return to this page, select 7.2.17 (OS upgrade, OS: redhat8) again, and run Upgrade.

When upgrading via CDP CLI, you use the typical upgrade command and specify a RHEL image available for your cloud platform.

For detailed instructions, see Performing a Cloudera Data Hub OS upgrade. For CLI instructions, see Upgrading a Cloudera Data Hub with the CDP CLI.

Upgrading from Medium Duty to Enterprise Data Lake

Refer to this content for an overview of the Data Lake resizing from Medium Duty to Enterprise Data Lake. The resizing of Medium Duty Data Lakes to Enterprise Data Lake should be performed on Cloudera Runtime 7.2.17 and is a prerequisite for upgrading to Cloudera Runtime 7.2.18 or 7.3.1.

Starting with Cloudera Runtime 7.2.18, Medium Duty Data Lake is discontinued and is replaced by Enterprise Data Lake. While earlier Cloudera Runtime versions support Medium Duty Data Lake (7.2.17 supports Enterprise, Medium Duty, Light Duty, and older versions support Medium Duty and Light Duty), 7.2.18 and later versions only support Enterprise and Light Duty Data Lakes.

Enterprise Data Lake enables Horizontal Scaling and Zero-Downtime Upgrades without increasing the infrastructure cost. For more information about Enterprise Data Lake, see Data Lake scale.

In order for existing Data Lakes to be upgraded to Cloudera Runtime 7.2.18 or later version, they must be using Enterprise or Light Duty Data Lake. If you are using Medium Duty Data Lake and would like to upgrade to Cloudera Runtime 7.2.18 or later version, you will need to upgrade to 7.2.17 first and then resize your Data Lake to Enterprise Data Lake using the existing Data Lake resize functionality. Afterwards, you will be able to proceed with the upgrade to Cloudera Runtime 7.2.18 or later version.

Note: After upgrading to EDL, Data Lakes with high resource utilization might require vertical scaling, which can increase cost. For more information, see Data Lake resizing

Upgrading from Spark 2 to Spark 3

If you have a Cloudera Data Engineering Data Hub cluster, during the Cloudera Runtime upgrade process to version 7.3.1, you need to remove Spark 2 while creating a new cluster with Spark 3 and migrate your data. Depending on your environment and Spark applications, besides the cluster upgrade tasks, you need to perform Spark application migration tasks and sidecar migration tasks for the Cloudera Data Hub cluster in multiple steps. For more information, see Upgrading Apache Spark.

Troubleshooting upgrade operations

This section provides some troubleshooting solutions for errors during upgrade operations.

Data Lake upgrade - out of memory error

Data Lake upgrade can fail due to not sufficient memory for Client Java Heap Size. The issue can be resolved by increasing the default Client Java Heap Size value in Cloudera Manager.

Condition

When upgrading the Data Lake from Cloudera Runtime 7.2.18 to one of its service pack versions or to 7.3.1, the upgrade fails with the following out of memory error during backup:

```
Upgrade not started, datalake backup failed. Failure message: Database: kini
t: Client 'hdfs/xxxxxxxxxx.CLOUDERA.SITE' not found in Kerberos database wh
ile getting initial credentials; kinit: Client 'hbase/xxxxxxxxxxxx.CLOUDERA
.SITE' not found in Kerberos database while getting initial credentials; kin
it: Client 'solr/uxxxxxxxxxxxx.CLOUDERA.SITE' not found in Kerberos databas
e while getting initial credentials; moveFromLocal: Failed with java.io.IOEx
ception while processing file/directory :[/xxxxxxx-xxx-xxxx-xxxx-xxxxxxxf_
database_backup/ranger_backup._COPYING_] in method:[java.util.concurrent.Exe
cutionException: java.lang.OutOfMemoryError: Direct buffer memory];
```

Cause

The default value of Client Java Heap Size in Bytes, 256 MB, is not the optimum value suitable for the upgrade operation.

Remedy

Procedure

1. Navigate to your environment in Cloudera Management Console, and click on the Data Lake tab on the environment details page.
2. Open Cloudera Manager by clicking on the CM URL.
3. Select Clusters core_settings in Cloudera Manager.
4. Click on the Configuration tab.
5. Search for Client Java Heap Size in Bytes in the search bar.

6. Increase the value from the default 256 MB to 1 GB (for example, 1073741824 bytes).
7. Click Save changes.
8. Restart the affected services as prompted by Cloudera Manager to apply the new configuration.
9. After the affected services are started again, retry the upgrade operation. Monitor the upgrade process to confirm the issue is resolved and the Data Lake upgrade finishes without errors.

Data Lake upgrade - Kafka consumer not available yet

After the Data Lake upgrade, the Atlas Hook does not function as the Kafka service is not available yet.

Condition

After upgrading the Data Lake to 7.3.1 or one of its patch versions, sometimes Atlas Hook does not function when Apache Atlas and Apache Kafka are started at the same time, thus Atlas is unable to connect to Kafka while Kafka is still being set up. When this happens, the following exception can be seen:

```
Exception in getKafkaConsumer ,WakeupException: null
```

The Kafka consumer creation should be retried if the Kafka service is unavailable during Atlas startup.

Cause

Atlas performs only three attempts to restart, but during this time, the Kafka service might not be available yet. This causes the Atlas Hook to not function and none of the messages from the Kafka topics are consumed.

Remedy

Procedure

1. Wait until the Kafka service is available after the upgrade.
2. Restart Atlas to trigger the reconnection to Kafka.