CDP Public Cloud

CDP Public Cloud Upgrade Advisor

Date published: 2022-07-11
Date modified: 2022-07-11

CLOUDERA

Legal Notice

Contents

CDP Public Cloud upgrade advisor

Compute resources deployed via cloud services are generally considered transient in nature. With the separation of compute and storage, compute clusters in CDP 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 CDP Public 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. CDP Public 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 Data Hub clusters. For steps required for the backup and restore approach, refer to the respective documentation on backing up Data Lakes and Data Hubs and performing metadata restore (automated for Data Lake clusters only).

Upgrade Checklist FAQ

During the preparation for an upgrade, Cloudera recommends carefully reviewing the questions and answers below.

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, Data Hubs need to be stopped during an upgrade.

> The CDP Public Cloud environment 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 Data Hub upgrade is required depends on the Runtime version of the Data Hub:

> - If your Data Hubs are on Runtime version 7.2.15 or earlier, they must run the same major/minor version of Runtime as the Data Lake. In this scenario, after a Data Lake upgrade you are required to upgrade any Data Hubs that are 7.2.15 or earlier to the same version as the Data Lake.
> - If your Data Hubs are on 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 Data Hubs 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 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 Data Hubs 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 data hub template. Depending on the type and version of the Data Hub, additional backup steps, manual configuration changes or post-upgrade steps may be required. Check specific steps for upgrading the OS if you use Flow Management. 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. CDP 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, Data Warehouse, Data Flow, or Machine Learning 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/ Runtime services.

Related Information

Preparing for an upgrade

Preparing for an upgrade

Upgrading CDP Public 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 Data Hubs. 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 Data Hubs. See upgrade the FreeIPA cluster.

Data Lake upgrade workflow
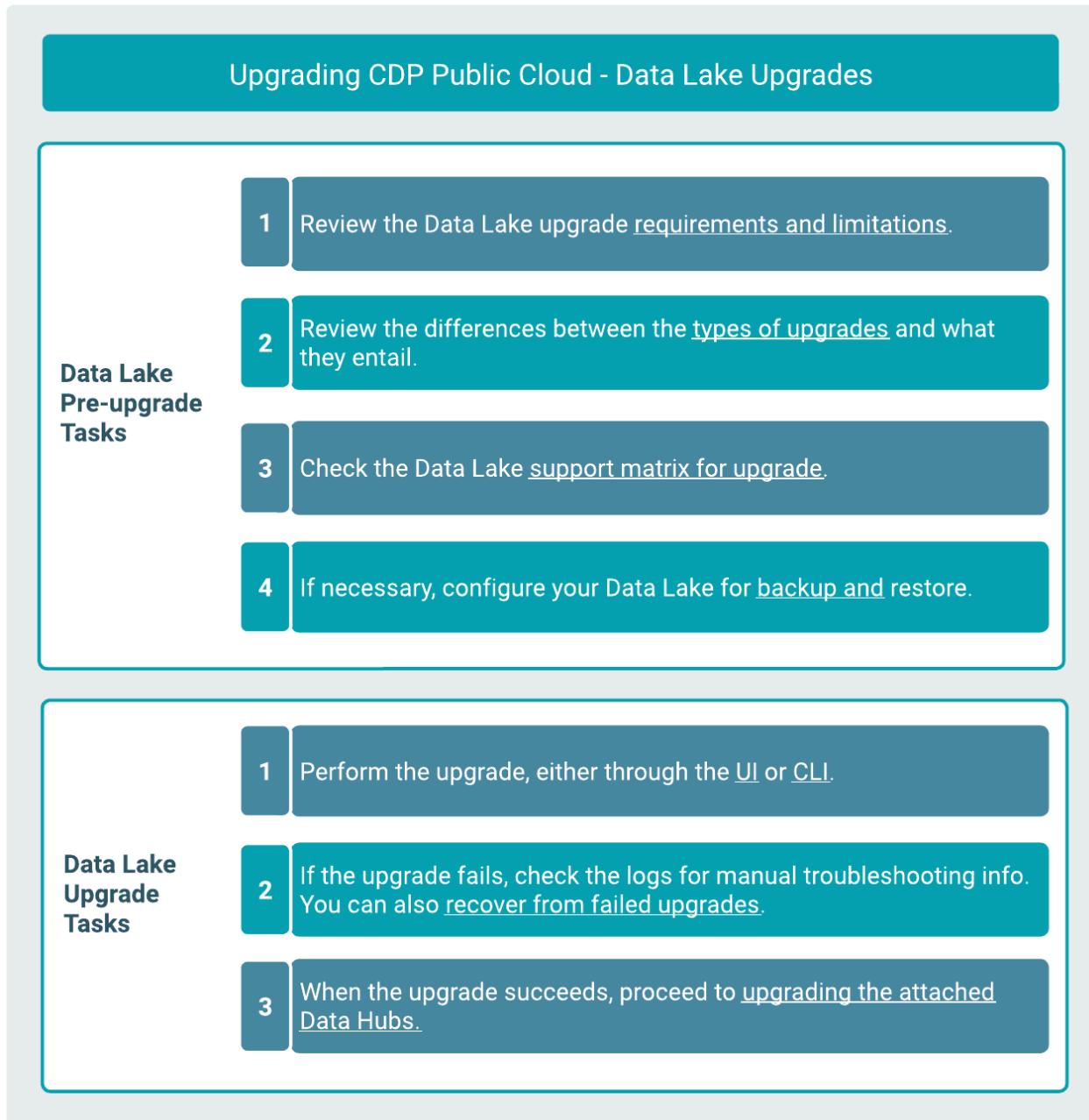
Pre-upgrade tasks

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, CDP 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 Data Hubs 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 (Runtime) version. Refer to the data services documentation for information on Data Lake compatibility and upgrading these services.

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

Data Hub upgrade workflow

Upgrading CDP Public Cloud - Data Hub Upgrades

Data Hub Pre-upgrade Tasks

1. Carefully review the differences between the types of upgrades and what they entail

2. Check that the cluster that you want to upgrade is supported.

Data Hub Major/Minor Upgrade Tasks

1. Backup the cluster data and CM configurations.

2. Perform the upgrade, either through the UI or CLI. (Operational database clusters have a different process.)

3. Complete any post-upgrade tasks for the type of cluster that you upgraded.

4. For DE clusters, use the CM 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.

Data Hub Maintenance Upgrade Tasks

1. Backup the cluster data and CM configurations.

2. Perform the upgrade, either through the UI or CLI.

3. If the upgrade fails, check the Event log and the troubleshooting section.

Pre-upgrade tasks

1. Carefully review the differences between the types of upgrades and what they entail.
2. Check the Data Hub upgrade support matrix to verify that the Runtime versions you want to upgrade to and from are supported for your clusters.
3. From the 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 CM configurations.
2. Perform the upgrade, either through the UI or CLI. (Operational database clusters have a different process.)
3. Complete any post-upgrade tasks for the type of cluster that you upgraded.
4. For DE clusters, use the CM 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 DE clusters, use the CM 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 CM 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.

Upgrading to Runtime 7.2.18

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

Some of the important 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, Runtime 7.2.18 supports RHEL 8-based images only. New deployments of Data Lakes and Data Hubs with Runtime 7.2.18 and upgrades to 7.2.18 are only possible with RHEL 8. Data Lake and Data Hub clusters running Runtime 7.2.17 support both CentOS 7 and RHEL 8. Earlier 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 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 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 Data Hub's Runtime version, Data Lake, Data Hub, and FreeIPA image version and operating system (OS) type, and Data Lake shape.

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

Identify cluster's Runtime version

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

Steps

For CDP UI

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

For CDP CLI

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

```
cdp datahub describe-cluster --datahub-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 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 Data Hub image version and OS type. You should follow these steps separately for each cluster.

Steps

For CDP UI

1. In the Management Console, navigate to the details page of the Data Lake or Data Hub that you would like to upgrade.
2. Scroll down to access the left navigation pane.
3. From the left navigation pane, select Image Details:
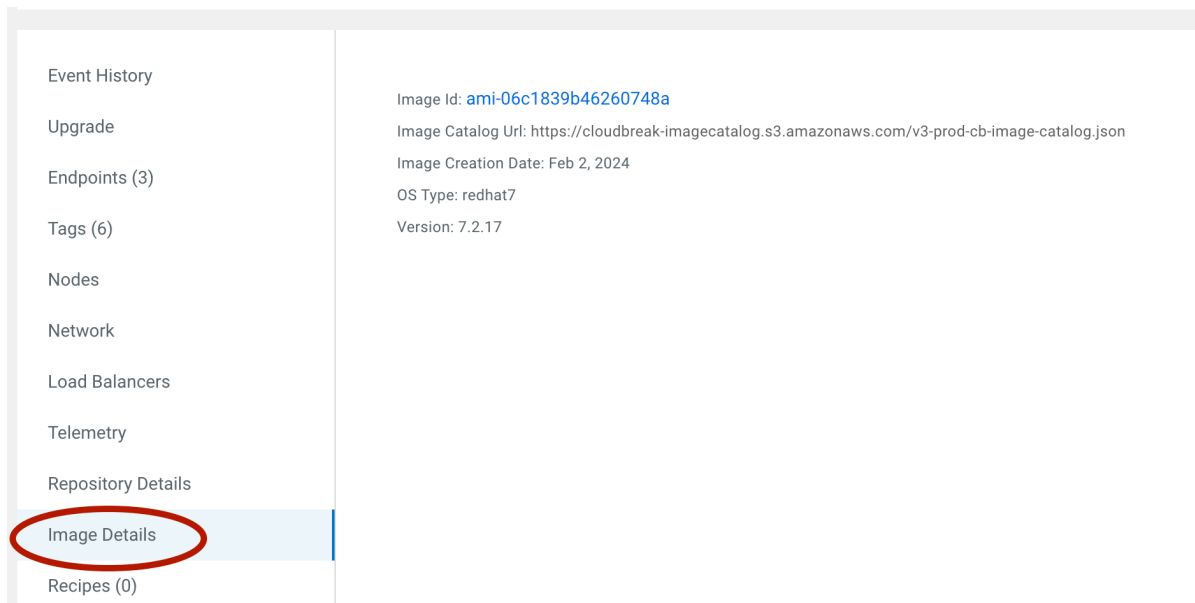
4. You will see information similar to the following:

```
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:

Details of 7ee9d8bc-a133-456a-a6ca-80b3cb93c9f7 image			
UUID	7ee9d8bc-a133-456a-a6ca-80b3cb93c9f7	Description	7.2.17.300 Service Pack Release
OS Type	CentOS7	Stack Name	Cloudera Runtime
Cluster Manager Version	7.11.0	Stack Version	7.2.17
Created On	Feb 2, 2024	Published On	Feb 13, 2024

For CDP CLI

Get Data Lake's image version from the CDP CLI by using the following command:

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

Get Data Hubs'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 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 CDP UI

1. In the 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 Data Hubs 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 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 Data Hubs a service pack upgrade and an OS upgrade in Step 1. If you started with Runtime 7.2.16, after performing the operations in Step 1 you would have your Data Lake and

all Data Hubs upgraded to the latest service pack of Runtime 7.2.16 and your cluster OS would be patched with up-to-date Python 3 binaries.

Step 1

If using Runtime 7.2.16 or earlier AND your cluster is affected by TSB-664, upgrade **Data Lake** and all **Data Hubs** to the latest service pack of their Runtime version and OS.

Step 2

If using CentOS, upgrade **FreeIPA** to RHEL 8.

Step 3

Upgrade **Data Lake** and **Data Hubs** to the latest Runtime 7.2.17.

If using CentOS, also upgrade **Data Lake** and **Data Hubs** to RHEL 8.

Step 4

If using Medium Duty **Data Lake**, resize to Enterprise Data Lake.

Step 5

Upgrade **Data Lake** and **Data Hubs** from Runtime 7.2.17 to 7.2.18.

Operations:
- DL upgrade
- DH service pack upgrade
- DH OS upgrade

Result:
- DL and DH use the latest SP and OS of your current Runtime version.
- Upgraded Python 3 binaries

Operations:
- FreeIPA upgrade
- DL upgrade
- DH Runtime version upgrade
- DH OS upgrade

Result:
- DL and DH are upgraded to the latest 7.2.17 version.
- FreeIPA, DL and DH use RHEL 8.

Operations:
- DL resize

Result:
- DL is EDL.

Operations:
- DL upgrade
- DH Runtime version upgrade

Result:
- DL and DH use the latest 7.2.18.

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

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

For example, if your clusters are using 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 Data Hubs 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 Management Console shows you the next Data Lake and Data Hub upgrade step available, so you can refer to it as a guideline.

USEFUL LINKS:

- Data Lake upgrade
- Data Hub upgrade
- Rolling upgrades

- 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 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 Data Hubs 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 Management Console shows you the next Data Lake and Data Hub upgrade step available, so you can refer to it as a guideline.

1. Upgrade Data Lake and all Data Hubs to the latest service pack and OS
2. Upgrade FreeIPA to RHEL 8
3. Upgrade Data Lake and Data Hubs to the latest 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 Data Hubs from 7.2.17 to 7.2.18

Step 1: Upgrade Data Lake and all Data Hubs to the latest service pack and OS

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

> **Important:**
> This step applies only to Data Lake and Data Hub clusters running 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 Data Hub separately. The examples and screenshots below assume that your cluster is using Runtime 7.2.16. For older Runtimes, just replace 7.2.16 with your actual Runtime version.

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

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

   a. Select the upgrade target that matches your current Runtime version. If you are upgrading a Data Lake, you will see 7.2.16 (Runtime upgrade, OS; centos7); for Data Hubs you will see Select 7.2.16 (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:

   ## Select Upgrade

   Target Cloudera Runtime Version

Please select a runtime version ▼
Please select a runtime version
7.2.16 (Runtime upgrade, OS: centos7)

   For Data Lakes, this will perform the OS image upgrade, so step 2 below is required for Data Hubs only.

2. Upgrade the OS image on all Data Hubs.

   Additionally, you are required to upgrade the OS image on all Data Hubs:

   a. Open the Data Hub Upgrade UI again and select the upgrade target that matches your current 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.

   ## Select Upgrade

   Target Cloudera Runtime Version

Please select a runtime version ▼
Please select a runtime version
7.2.16 (OS upgrade, OS: centos7)

   Tip:

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

   When upgrading a 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 Data Hub service pack upgrade (UI)
- Performing a Data Hub OS upgrade (UI)
- Upgrading a Data Hub with CLI

Step 2: Upgrade FreeIPA to RHEL 8

Before your Data Lake or 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 Data Hubs to the latest 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 Data Hub clusters using a CentOS 7 image. If your clusters are already using 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 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 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 Runtime upgrade:

a. If you are upgrading a Data Lake, select 7.2.17 (Runtime upgrade, OS: redhat8). For upgrading a Data Hub, select 7.2.17 (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

Select Upgrade

Target Cloudera Runtime Version

Please select a runtime version	▼
Please select a runtime version	
7.2.17 (Runtime upgrade, OS: redhat8)	

Data Hub UI

Select Upgrade

Target Cloudera Runtime Version

Please select a runtime version	▼
Please select a runtime version	
7.2.17 (Runtime upgrade)	

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

3. Upgrade all Data Hubs 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.

Select Upgrade

Target Cloudera Runtime Version

Please select a runtime version	▼
Please select a runtime version	
7.2.17 (OS upgrade, OS: centos7)	
7.2.17 (OS upgrade, OS: redhat8)	

> **Tip:** When upgrading a Data Lake, all Data Hubs in your environment should be stopped and CDP 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 Data Hub OS upgrade
- Upgrading a 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 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.
- 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.
- 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.
- 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 Data Hub through the Cloudera Manager UI.
- With Cloudera Data Flow 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 Data Hubs from 7.2.17 to 7.2.18

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

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

1. Once Runtime 7.2.18 (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 (Runtime upgrade, OS: redhat8) again and run Upgrade.

> ⚠️ **Attention:**
>
> Enterprise Data Lakes created with 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 Data Hubs and workloads on any data services that are running in the environment. To determine if your cluster is eligible, refer to Rolling upgrades.
>
> Cloudera recommends that you perform the rolling upgrade during a maintenance window, when no critical workloads are running on your attached Data Hub clusters or data services.
>
> To perform a rolling upgrade of the Data Lake, select "Rolling Upgrade":

Select Upgrade

Target Cloudera Runtime Version

7.2.18 (Runtime upgrade, OS: redhat8)   ▼

☐ Perform 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 Data Hub major/minor version upgrade (UI)
- Upgrading a Data Hub with CLI
- Rolling upgrades

Upgrading from CentOS to RHEL

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

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 Data Hubs via the Data Hub OS upgrade web UI or CLI.

> ⚠️ **Attention:** Your Data Lake or Data Hub must be using 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 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 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 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 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 CDP web UI. See CDP images hosted in Azure Marketplace.

RHEL 8 images are only available through Azure Marketplace. In order for CDP 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 CDP web UI or Azure CLI . If you do not accept the terms and conditions, CDP will not be able to access the images hosted in Azure Marketplace and consequently environment creation will fail. For more information, see CDP images hosted in Azure Marketplace.

Upgrade clusters to RHEL

To upgrade your environment, Data Lake, and Data Hubs to RHEL, follow the usual operating system (OS) upgrade documentation for FreeIPA, Data Lake, and 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:

Upgrade FreeIPA ⟳

Target Image

FreeIPA (Latest, OS: redhat8)	▼

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

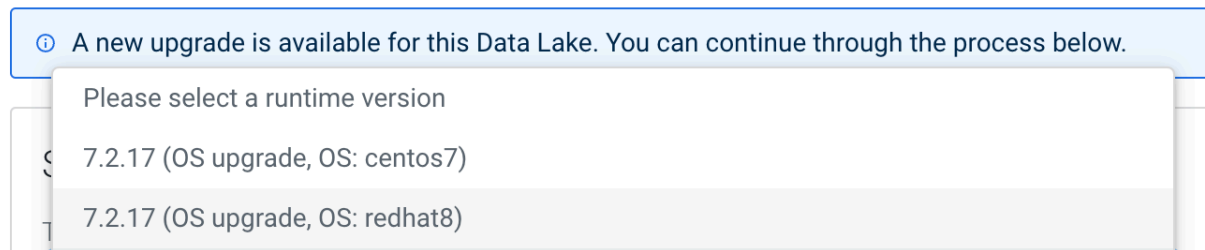For detailed CDP UI and CLI instructions, see Upgrade FreeIPA.

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:

Upgrade Data Lake ⟳

Current Data Lake Version: 7.2.17

ⓘ A new upgrade is available for this Data Lake. You can continue through the process below.

Please select a runtime version

7.2.17 (OS upgrade, OS: centos7)

7.2.17 (OS upgrade, OS: redhat8)

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 Data Hubs to RHEL

When upgrading via the web interface, you navigate to your 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.

Upgrade Data Hub ⟳

Current Data Hub Version: 7.2.17
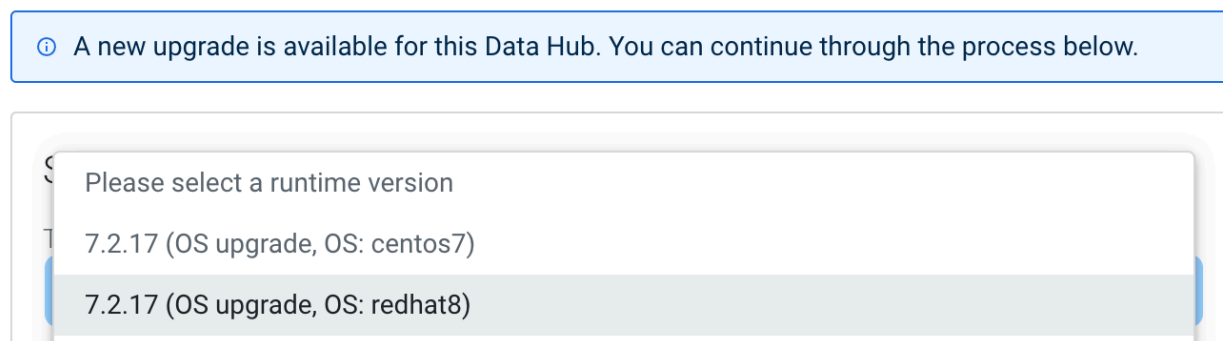
ⓘ A new upgrade is available for this Data Hub. You can continue through the process below.

Please select a runtime version

7.2.17 (OS upgrade, OS: centos7)

7.2.17 (OS upgrade, OS: redhat8)

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 Data Hub OS upgrade. For CLI instructions, see Upgrading a 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 (EDL). The resizing of Medium Duty Data Lakes to EDL should be performed on Runtime 7.2.17 and is a prerequisite for upgrading to 7.2.18.

Starting with Runtime 7.2.18, Medium Duty Data Lake is discontinued and is replaced by Enterprise Data Lake. While earlier 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.

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

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

Rolling upgrades

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

Rolling upgrades allow you to upgrade the Data Lake Runtime and OS, or the 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 Data Hubs and Data Services, or upgrade a Data Hub without workload downtime.

Rolling upgrades for the Data Lake are limited to certain Data Lake Runtime versions and shapes. Rolling upgrades for Data Hubs are limited to certain Runtime versions and cluster templates. See the following topics for available upgrade paths and limitations.

Data Lake rolling upgrades

The Data Lake rolling upgrade allows you to upgrade the Data Lake Runtime and OS without stopping attached Data Hubs or Data Services. This allows workloads to continue running during the Data Lake upgrade operation.

Similarly to the classic Data Lake upgrade logic, a Data Lake rolling upgrade first upgrades the Runtime version and then the OS.

To run a Data Lake rolling upgrade, the following requirements must be met:

- The Data Lake must be Runtime version 7.2.17.300+ to perform a rolling upgrade directly to 7.2.18+.
- The Data Lake OS must be RHEL 8. Rolling upgrades are not generally available from Data Lakes on CentOS.
- The Data Lake must be an Enterprise Data Lake (EDL). Rolling upgrades to 7.2.18 are not generally available from medium duty or light duty Data Lakes.

If your Data Lake does not meet these requirements, you will first have to do a traditional Runtime upgrade, upgrade your OS from CentOS to RHEL, or resize your Data Lake to EDL before a rolling upgrade can be performed.

Current Runtime version	Current OS	Current Data Lake Shape	Rolling upgrade support?
7.2.17.300+	RHEL 8	EDL	Yes, directly to 7.2.18+

In some circumstances, a rolling upgrade may not be supported for a Data Lake cluster, but can be enabled through entitlement. Some cluster services might become unavailable during this type of upgrade, and running workloads could be impacted. The Data Lake upgrade UI displays information about whether a rolling upgrade is available, unavailable, or may be available under entitlement. For instructions on performing a Data Lake upgrade, including rolling upgrades, see Upgrading a Data Lake. For information about obtaining an entitlement for rolling upgrade, contact Cloudera Customer Support.

Data Lake rolling upgrade limitations and issues

The Data Lake rolling upgrade has the following limitations:

- Long running CDE and CML Spark jobs might loose connectivity to the Hive Catalog and fail during and after a Data Lake rolling upgrade. We recommend stopping these jobs prior to attempting the upgrade. The jobs will work again fine after re-submitting them.
- Cloudera recommends performing the upgrade outside of working hours, as user-facing UI/API endpoints may become unstable. Workloads running on Data Hubs and Data Services use different internal endpoints, so they are not affected. The impact of this is that you may not be able to view or edit Ranger permissions and the Ranger audit log, browse Atlas/Data Catalog, or make changes to Atlas at certain times during the rolling Data Lake upgrade (see Known Issues in Apache Atlas for more details on Atlas issues). If you are using custom-built applications that interact with the Data Lake using these endpoints, we recommend implementing retry logic in your clients to handle temporary unavailability of these endpoints. This is a best practice, irrespective of rolling upgrades.
- Atlas Authorization may return a "403-Access Denied" in response to Atlas REST API calls. After the rolling upgrade finishes and Ranger Admin is back up, these services and their endpoints will continue to function normally.
- The Ranger RAZ server becomes unreachable during and after the OS upgrade for some time, and an UnknownHostException may be seen in the RAZ client. During this period all authentication calls to the Ranger RAZ server are expected to fail.
- During OS upgrades, attempts to access Knox on the host being upgraded may produce occasional 403 HTTP responses. Wait and retry the failed requests.
- If Knox is HA and one of the Knox servers is down, then accessing the service through a Control Plane endpoint URL (i.e., through cloud load balancer) will take approximately 30 seconds to failover the request to the available Knox instance.

Disposition: / Status:
April 4, 2024 Commenting out the next bullet upon Balazs's request

- Solr supports rolling upgrades from release 7.2.18 and higher. Upgrading from a lower version means that all the Solr Server instances are shut down, parcels upgraded and activated, and then the Solr Servers are started again. This causes a service interruption of several minutes, the actual value depending on cluster size. Services like Atlas and Ranger that depend on Solr may face issues because of this service interruption.
- Certain workloads may experience downtime during the Data Lake rolling upgrade operations:

  - Any workloads configured to use a single HMS endpoint (Hive Warehouse Connector configurations).
  - Hue File Browser may be unavailable for a short period during a rolling upgrade in a RAZ-enabled environment.
  - Other clients in use in your workloads could be impacted. Cloudera recommends testing how your workloads function during Data Lake rolling upgrades, before you adopt this new feature.
  - During the rolling upgrade, Hive and HBase grant and revoke commands will not function.
  - Generally, if you have any workload that is using a single Data Lake service endpoint, it will likely experience a temporary outage. This may not necessarily result in a workload failure.

- Rolling upgrades for an enterprise Data Lake will take longer than a classic upgrade that requires downtime. This is because OS image upgrade will be performed sequentially, node by node.

- Certain operations (create, upgrade, and resume) for Data Hubs and Data Services are not recommended during a Data Lake rolling upgrade.

Data Hub rolling upgrades

The Data Hub rolling upgrade allows you to upgrade the Data Hub without stopping the Data Hub itself and incurring workload downtime.

Data Hub rolling upgrades are only available for the following cluster types:

- Cloudera Operational Database

Cloudera Operational Database clusters

For an Operational Database cluster to be eligible for rolling upgrade, the following requirements must be met:

- Currently, rolling OS upgrade is only supported on COD clusters with storage selected as HDFS or cloud without ephemeral storage.
- Rolling OS upgrade is not supported on Micro COD clusters.

Table 1:

Current Runtime version & template	Rolling upgrade support for Runtime?	Rolling OS upgrade support?
7.2.18	Yes, to 7.2.18+ (including service pack upgrades to 7.2.18.100+) Up to 20 nodes	Yes, for up to 20 nodes
7.2.17.x	Yes, to 7.2.18+	Yes

For instructions on performing COD rolling upgrades, see Performing a Cloudera Runtime upgrade (COD) and Performing a Cloudera operating system upgrade (COD).

Disposition: / Status:
Removing CSM patch upgrades as GA per QE (Sharmadha and Surya). See https://jira.cloudera.com/browse/DOCS-20159. The removed section is below.

Streams messaging clusters

For a Streams Messaging cluster to be eligible for rolling upgrade, the following requirements must be met:

- The cluster OS must be RHEL 8. Streams Messaging clusters on CentOS are not supported.
- The cluster must be Runtime version 7.2.18+. Rolling upgrades to 7.2.18 from certain Runtime 7.2.17 versions may be possible in certain scenarios. Contact Cloudera Customer Support for details.
- The cluster template must be Streams Messaging HA.

Current Runtime version & template	Rolling upgrade support for Runtime 7.2.18+?	Rolling OS upgrade support?
7.2.18 Streams Messaging HA template (Clusters created on 7.2.18 only)	Yes (including service pack upgrades to 7.2.18.100+)	Yes, for up to 20 nodes. ⚠️ **Important:** Rolling OS upgrades replace nodes one-by-one and can take a significant amount of time, several hours or more, depending on the number of nodes.

In some circumstances, a rolling upgrade may not be supported for a Streams Messaging cluster, but can be enabled through entitlement. The Data Hub upgrade UI displays information about whether a rolling upgrade is available, unavailable, or may be available under entitlement. For instructions on performing a Data Hub upgrade, including

rolling upgrades, see Upgrading Data Hubs. For information about obtaining an entitlement for rolling upgrade, contact Cloudera Customer Support.

Data Hub rolling upgrade limitations and issues

Data Hub rolling upgrades have the following limitations:

Disposition: / Status:

Removing CSM patch upgrades as GA per QE (Sharmadha and Surya). See https://jira.cloudera.com/browse/DOCS-20159. The removed section is below.

Streams messaging clusters

- Rolling upgrades are not supported for Cruise Control or Streams Messaging Manager (SMM). When upgrading a Streams Messaging cluster, expect that both of these services will be temporarily unavailable during the upgrade. This, however, does not impact Kafka's ability to upgrade without downtime.
- Rolling upgrades for Schema Registry are only supported from Cloudera Runtime 7.2.18 or higher to higher versions. When upgrading a Streams Messaging cluster from a lower version, expect that clients connecting to the Schema Registry service might experience downtime. This, however, does not impact Kafka's ability to upgrade with zero downtime.

Cloudera Operational Database clusters

- HBase commands may fail during the rolling upgrade with the error "ServerNotRunningYetException: Server is not running yet." HBase retries DDL operations submitted while the master is initializing until the master is fully initialized to serve the request. However, a situation might arise where the default number of retries or intervals proves to be insufficient for an operation submitted by the client to complete.

  Implementing the following configuration adjustments in your client application can support the master getting initialized up to 10 minutes:

  ```
  <property>
      <name>hbase.client.pause</name>
      <value>300</value>
  </property>
  <property>
      <name>hbase.client.retries.number</name>
      <value>20</value>
  </property>
  ```

  If you have seen a longer or shorter master initialization period, you can modify these values accordingly. These retry settings apply to all types of calls to HBase service, encompassing GET, SCAN, MUTATE, and DDLs.
- In a rolling restart, if a COD cluster has less than 10 datanodes, existing writes can fail with an error indicating a new block cannot be allocated and all nodes are excluded. This is because the client has attempted to use all the datanodes in the cluster, and failed to write to each of them as they were restarted. This will only happen on small clusters of less than 10 datanodes, as larger clusters have more spare nodes to allow the write to continue.
- When performing a patch upgrade or other cluster upgrade, in some occasions there can been an error when the upgrade is nearly complete, but trying to re-start services/roles. The error is similar to: "Failed to start role hue-HUE_SERVER-8cc9321b2213cc5c6846c64e1fc6b1cb of service hue in cluster cod--xoaitnb0wnl1. This role requires the following additional parcels to be activated before it can start: [cdh]."

  This is due to an agent operation that sometimes is delayed and can interfere with the role start. When this happens, resume the failed upgrade from Cloudera Manager as a 'Full Admin' user.
- During the VM replacement as part of OS upgrade, every new node gets a new IP Address, and if the old IP address is cached somewhere, HDFS requests fail with UnknownHostException. It recovers after somet ime (10 mins max).
- During OS upgrades, attempts to access Knox on the host being upgraded may produce occasional 403 HTTP responses. Wait and retry the failed requests.

For more limitations of Cloudera Operational Database, see Rolling upgrade limitations (COD).