

# Data Lake Upgrade (Preview)

Date published: 2021-06-03

Date modified: 2021-07-07

## Legal Notice

© Cloudera Inc. 2021. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 ("ASLv2"), the Affero General Public License version 3 (AGPLv3), or other license terms.

Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER'S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## Contents

<b>Legal Notice</b>	<b>2</b>
<b>Contents</b>	<b>3</b>
Data Lake Upgrade Overview	4
Runtime and CM Major/Minor Version Upgrades	5
Maintenance/'Hotfix' Upgrades	5
OS Upgrades	6
Upgrade Paths	6
Before You Begin	6
Upgrading the Data Lake	7
Upgrading the Data Lake with the CDP CLI	8
Examples of valid inputs:	10
Examples of incorrect inputs:	11

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## Data Lake Upgrade Overview

When new versions of Cloudera Runtime/Cloudera Manager are available for the Data Lake service, you can initiate a Data Lake upgrade. You may also have the option to upgrade to a new OS image. Use either the CDP CLI or the Management Console to initiate an upgrade.

There are several potential types of Data Lake upgrades:

- Runtime and Cloudera Manager version upgrades, which are available when a new Runtime version is released.
- Maintenance or ‘hotfix’ upgrades, which do not change the major or minor Runtime or CM release, but upgrade the Data Lake to the latest CM and/or Runtime hotfix versions available in the given version. These upgrades are made available as needed and can deliver targeted bug fixes for Cloudera Runtime, Cloudera Manager, or both.
- OS upgrades, which do not change any CM or Runtime builds, but update the underlying cloud image.

The type of upgrade that is available for a Data Lake is explicit on the **Upgrade** tab of the Data Lake details page. Any available upgrades are visible in the *Target Cloudera Runtime Version* drop-down menu. For example:

Event History Endpoints (6) Tags (6) Hardware Network Telemetry Repository Details Image Details Recipes (0) Cloud Storage Attached clusters (0) Database [Upgrade](#)

Upgrade Data Lake ↻

Target Cloudera Runtime Version

7.2.9

Version details

	Patch Level	CM Version	CM Build number	CDP Version	CDP Build number
Current	2021-05-03	7.4.0	11637958	7.2.8	12570397
Target	2021-05-03	7.4.1	12308673	7.2.9	12752730

[Upgrade](#)

*A version upgrade, where the major/minor versions of CM and Runtime are upgraded together.*

Event History Endpoints (6) Tags (6) Hardware Network Telemetry Repository Details Image Details Recipes (0) Cloud Storage Attached clusters (0) Database [Upgrade](#)

Upgrade Data Lake ↻

Target Cloudera Runtime Version

7.2.9

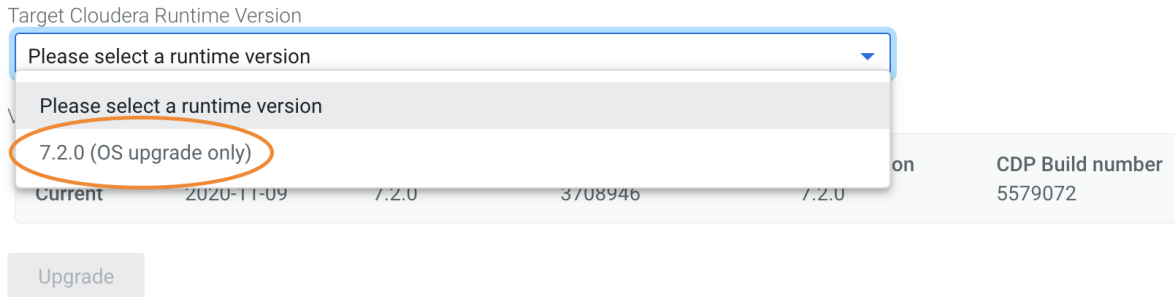
Version details

	Patch Level	CM Version	CM Build number	CDP Version	CDP Build number
Current	2021-04-27	7.4.1	12308673	7.2.9	12682656
Target	2021-05-03	7.4.1	12308673	7.2.9	12752730

[Upgrade](#)

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided ‘as is’ without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

*A maintenance/'hotfix' upgrade, where no version upgrade is available, but a new Cloudera Runtime and/or Cloudera Manager hotfix is available for upgrade.*



*An OS upgrade is designated by “(OS Upgrade Only)” when you select the drop-down menu.*

There are important differences between the nature of version, hotfix, and OS upgrades. Read the details below carefully to understand which type of upgrade is appropriate for your situation.

### *Major/Minor Version Upgrades*

Version upgrades are available as new versions of Cloudera Runtime and Cloudera Manager are released. Version upgrades combine a Runtime and CM upgrade into one operation. For example, this upgrade might involve upgrading from Runtime 7.2.8 to 7.2.9, and Cloudera Manager from 7.4.0 to 7.4.1.

Data Lake version upgrades require you to stop all of the Data Hubs in the environment before performing the Data Lake upgrade. The upgrade process stops all of the Data Lake services, downloads and distributes the new Cloudera Runtime, restarts the services, and deletes the old Cloudera Runtime. This process includes launching entirely new instance(s) from new cloud image(s). Once the Data Lake upgrade is complete, you can then either upgrade your Data Hub clusters to the same version as the Data Lake (if your Data Hub cluster type is supported for upgrade), or delete and recreate the clusters with the new Data Lake version.

**Note:** Data Hub clusters in the Environment must be running a compatible version of Cloudera Runtime.

### *Maintenance/'Hotfix' Upgrades*

The 'hotfix' upgrade process checks to see if a new Cloudera Manager or Cloudera Runtime build is available, and then upgrades the Data Lake to the newest build by default. If desired, specific older hotfixes can also be chosen using the CDP CLI. The hotfix upgrade process follows the same steps as the version upgrade process.

Hotfix upgrades do not upgrade to a new **version** of Cloudera Manager or Runtime; they only upgrade to the latest hotfix.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## OS Upgrades

OS upgrades may be available sporadically as new images are created. OS upgrades are typically released to address security vulnerabilities or other issues on the host system. The OS upgrade process includes launching entirely new instances with the new OS image.

## Upgrade Paths

The following Data Lake upgrade paths are supported:

### AWS

Current Runtime Version	Supported Upgrade Paths
7.1.0	7.2.2, 7.2.6, 7.2.7, 7.2.8, 7.2.9, 7.2.10
7.2.0	7.2.2, 7.2.6, 7.2.7, 7.2.8, 7.2.9, 7.2.10
7.2.1	7.2.2, 7.2.6, 7.2.7, 7.2.8, 7.2.9, 7.2.10
7.2.2	7.2.6, 7.2.7, 7.2.8, 7.2.9, 7.2.10
7.2.6	7.2.7, 7.2.8, 7.2.9, 7.2.10
7.2.7	7.2.8, 7.2.9
7.2.8	7.2.8, 7.2.9
7.2.9	7.2.9, 7.2.10
7.2.10	7.2.10

### Azure

Current Runtime Version	Supported Upgrade Paths
7.1.0	7.2.9, 7.2.10
7.2.0	7.2.2, 7.2.6, 7.2.7, 7.2.9, 7.2.10
7.2.2	7.2.6, 7.2.7, 7.2.8, 7.2.9, 7.2.10
7.2.8	7.2.9
7.2.9	7.2.9, 7.2.10
7.2.10	7.2.10

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## GCP

Current Runtime Version	Supported Upgrade Paths
7.2.9	7.2.10
7.2.10	7.2.10

## Before You Begin

- The Data Lake must be running and in a healthy state.
- Plan the upgrade during a time of low activity. You should stop any Experiences (Data Engineering, CML, CDW, Operational Database, DataFlow, or Data Visualization instances) that are running, and these should run the latest available releases that are available to the service.
- Create a backup of the Data Lake. For instructions, see [Backup and restore for the Data Lake](#). If the upgrade fails for any reason, you can restore the Data Lake from the backup that you took.
- Plan for at least two hours of downtime while the upgrade completes.
- The upgrade requires 27 GB space on the CM server node and 20 GB on every other instance. If space is insufficient on your Data Lake, upgrade will not be permitted.

Note the following limitations for the Data Lake upgrade:

- Data Lake upgrade does not include the upgrade of the FreeIPA software or the operating system on the instance(s) running FreeIPA. No upgrades are supported for FreeIPA at this time.
- Data Lake resizing during an upgrade is not supported at this time.
- Zero-downtime upgrade of Data Lake services is not yet supported.

## Upgrading the Data Lake

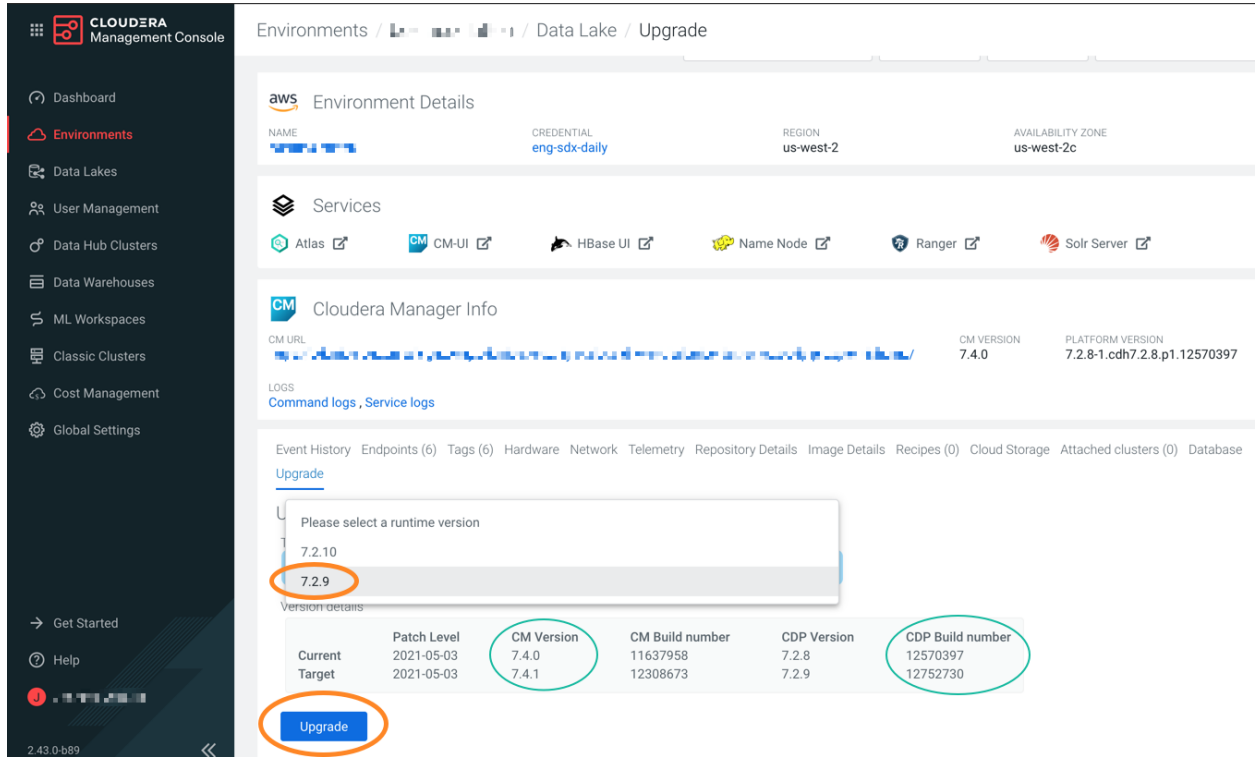
If a new CM or CDP build is available for the Data Lake, you can initiate an upgrade from the Management Console.

- 1) Stop all Data Hubs attached to the environment. In order to upgrade the Data Lake, all attached Data Hub clusters must be stopped before you begin.
- 2) From the Management Console, click **Data Lakes > Environment Name**, scroll to the bottom of the Data Lake details page, and click the **Upgrade** tab.
- 3) Click the **Target Cludera Runtime Version** drop-down menu to see any available upgrades for a given Runtime version.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cludera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

If a new build is available for the selected version, the UI displays the current and target versions and build numbers. If only an OS upgrade is available, the UI displays “(OS upgrade only).”

- 4) Select the desired version for upgrade and verify the version or builds that you are upgrading to.



- 5) Click **Upgrade** to initiate the upgrade.
- 6) Click the **Event History** tab to monitor the upgrade process and verify that it completes successfully.

**What to do next:** Upgrade (or if necessary, delete and re-create) your Data Hub clusters and restart any Experiences. Data Hub clusters must run the same Runtime version as the Data Lake.

## Upgrading the Data Lake with the CDP 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.

The `cdp datalake upgrade-datalake` command has the following options:

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*



## CLUDERA TECHNICAL PREVIEW DOCUMENTATION

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]
[--cli-input-json <value>]
[--generate-cli-skeleton]

```

Option	Synopsis
--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.
--runtime (string)	The Runtime version to upgrade to. When you specify the Runtime version, the upgrade uses the latest image ID of the given Runtime version from the same image catalog used for Data Lake 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-available-image-per-runtime   --no-show-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.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

<code>--cli-input-json</code> (string)	Performs service operation based on the JSON string provided. The JSON string follows the format provided by <code>--generate-cli-skeleton</code> . If other arguments are provided on the command line, the CLI values will override the JSON-provided values.
<code>--generate-cli-skeleton</code> (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 <code>--cli-input-json</code> .

When you run the `cdp datalake upgrade-datalake` 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, 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 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-runtime
--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
d1c520b1-987d-461f-7860-918f43994c04
cdp datalake upgrade-datalake --datalake-name my-datalake --image-id
d1c520b1-987d-461f-7860-918f43994c04 --dry-run
cdp datalake upgrade-datalake --datalake-name my-datalake --runtime 7.1.1
cdp datalake upgrade-datalake --datalake-name my-datalake --runtime 7.2.0
    
```

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## CLOUDERA TECHNICAL PREVIEW DOCUMENTATION

```
cdp datalake upgrade-datalake --datalake-name my-datalake --runtime 7.2.0
--dry-run
cdp datalake upgrade-datalake --datalake-name my-datalake --runtime 7.3.0
cdp datalake upgrade-datalake --datalake-name my-datalake--lock-components
cdp datalake upgrade-datalake --datalake-name my-datalake
--show-available-image-per-runtime
cdp datalake upgrade-datalake --datalake-name my-datalake-
--show-available-images
```

### Examples of incorrect inputs:

```
cdp datalake upgrade-datalake --datalake-name my-datalake --image-id 7.1.1
cdp datalake upgrade-datalake --datalake-name my-datalake--runtime
d1c520b1-987d-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-available-image-per-runtime --show-available-images
cdp datalake upgrade-datalake --datalake-name my-datalake
--show-available-image-per-runtime --dry-run
cdp datalake upgrade-datalake --datalake-name my-datalake
--show-available-images --dry-run
```

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*