

# Rolling Data Lake Upgrades (Preview)

Date published: 2023-07-11

Date modified: 2023-07-11

# Legal Notice

© Cloudera Inc. 2023. All rights reserved.

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

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

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

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

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

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

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

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

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

# Contents

<b>Legal Notice</b>	<b>2</b>
<b>Contents</b>	<b>3</b>
Rolling Data Lake Upgrades	4
Rolling Upgrade Limitations	7

## Rolling Data Lake Upgrades

With the release of Cloudera Runtime 7.2.17, you can now upgrade your Data Lake in a rolling manner. The 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.

Data Lake rolling upgrades are a technical preview feature under entitlement and are currently available only through the [beta CDP CLI](#). Similarly to the current [Data Lake upgrade logic](#), a Data Lake rolling upgrade first upgrades the Runtime version and then the OS.

The new beta CDP CLI option to enable rolling upgrades is: `rolling-upgrade-enabled`

Use the `rolling-upgrade-enabled` option in the `cdp datalake upgrade-datalake` CLI command. For example:

```
cdp datalake upgrade-datalake --datalake-name test-dl --runtime
7.2.17 --rolling-upgrade-enabled
```

The `cdp datalake upgrade-datalake` 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]
  [--rolling-upgrade-enabled]
  [--cli-input-json <value>]
  [--generate-cli-skeleton]
```

Option	Description
<code>--datalake-name</code> (string)	Required. The name or CRN of the Data Lake to upgrade.

<code>--image-id (string)</code>	The ID of an image to upgrade to.
<code>--runtime (string)</code>	The Runtime version to upgrade to. In a maintenance upgrade this parameter is the current Runtime version. 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. If you specify an invalid Runtime version, you'll receive an error message that the version is not supported for upgrade.
<code>--lock-components   --no-lock-components (boolean)</code>	Use <code>--lock</code> components to perform an OS upgrade only.
<code>--dry-run   --no-dry-run (boolean)</code>	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.
<code>--show-available-images   --no-show-available-images (boolean)</code>	Returns the list of images that are eligible to upgrade to.
<code>--show-available-image-per-runtime   --no-show-available-image-per-runtime (boolean)</code>	Returns the latest image that is eligible to upgrade to, for each Runtime version with at least one available upgrade candidate.
<code>--skip-backup   --no-skip-backup</code>	If provided, will skip the backup flow for the upgrade process.
<code>--skip-ranger-hms-metadata   --no-skip-ranger-hms-metadata</code>	Skips the backup of the databases backing HMS/Ranger services. Redundant if <code>--skip-backup</code> is included. If this option is not provided, the HMS/Ranger services are backed up by default.
<code>--skip-atlas-metadata   --no-skip-atlas-metadata</code>	Skips the backup of the Atlas metadata. Redundant if <code>--skip-backup</code> is included. If this option is not provided, the Atlas metadata is backed up by default.
<code>--skip-ranger-audits   --no-skip-ranger-audits</code>	Skips the backup of the Ranger audits. Redundant if <code>--skip-backup</code> is included. If this option is not provided, Ranger audits are backed up by default.

<p><code>--skip-backup-validation</code>    <code>--no-skip-backup-validation</code></p>	<p>Skips the validation steps that run prior to the backup. Redundant if <code>--skip-backup</code> is included. If this option is not provided, the validations are performed by default.</p>
<p><code>--rolling-upgrade-enabled</code></p>	<p>If provided, enables the rolling upgrade.</p>
<p><code>--cli-input-json</code> (string)</p>	<p>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.</p>
<p><code>--generate-cli-skeleton</code> (boolean)</p>	<p>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>.</p>

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.

**Note:** With backward compatibility between Data Lakes and Data Hubs also being introduced in Cloudera Runtime 7.2.17, it is no longer required that you perform Data Hub upgrades in lock-step with the Data Lake upgrade. You can upgrade your Data Hubs at a later time if you choose to. This requires that your Data Hubs are on Cloudera Runtime version 7.2.16 or newer.

**Note:** Rolling upgrade is only supported for highly available Data Lakes (medium duty and enterprise).

## Rolling Upgrade Limitations

- 1) 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.

These services and their endpoints will continue to function normally after the Data Lake rolling upgrade finishes.

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.

- 2) Certain workloads may experience downtime during the Data Lake rolling upgrade operations:
  - a) Any workloads configured to use a single HMS endpoint (Hive Warehouse Connector configurations).
  - b) Hue File Browser may be unavailable for a short period during a rolling upgrade in a RAZ-enabled environment.
  - c) 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.
  - d) During the rolling upgrade, Hive and HBase grant and revoke commands will not function.
  - e) 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.
- 3) Rolling upgrades for a medium duty Data Lake or 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.
- 4) Certain operations (create, upgrade, and resume) for Data Hubs and Data Services are not recommended during a Data Lake rolling upgrade.