

CDP Private Cloud Data Services 1.5.3

Upgrading CDP Private Cloud Data Services on the OpenShift Container Platform

Date published: 2023-12-16

Date modified: 2024-03-23

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with a stylized 'E' that has a horizontal bar extending to the right.

<https://docs.cloudera.com/>

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

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

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

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

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

Contents

Upgrade from 1.5.1 or 1.5.2 to 1.5.3 (OCP).....	4
Pre-upgrade - Preparing for CDP Private Cloud Data Services update for CDE.....	9
Pre-upgrade - Upgrading CDE Service for Endpoint Stability.....	9
Prerequisites for upgrading CDE Service with endpoint stability.....	9
Backing up CDE service using the docker image.....	11
OCP upgrade steps for CDP Private Cloud Data Services 1.5.3.....	12
Upgrading CDP Private Cloud Data Services.....	13
Completing post OCP update tasks.....	18
Post-upgrade - Ozone Gateway validation.....	18
Recovering a corrupted CDE Embedded database.....	20
Post-upgrade - Restoring CDE Service for Endpoint Stability.....	21
Restoring a CDE service.....	21
Rolling back the CDE Service Endpoint migration.....	23
Limitations.....	24

Upgrade from 1.5.1 or 1.5.2 to 1.5.3 (OCP)

You can upgrade your existing CDP Private Cloud Data Services 1.5.1 or 1.5.2 to 1.5.3 without performing an uninstall. After the update is complete, you may need to upgrade the underlying OpenShift Container Platform. See the [Software Support Matrix for OpenShift](#) for more information about supported OCP versions.

About this task

If you are upgrading the OCP version to **4.10.x or higher**, while the CDE service is enabled, it fails to launch the Jobs page in the old CDE virtual cluster. Hence, you must back up CDE jobs in the CDE virtual cluster, and then delete the CDE service and CDE virtual cluster. Restore it after the upgrade. For more information about backup and restore CDE jobs, see [Backing up and restoring CDE jobs](#).



Note: CML Customers on 1.5.0, 1.5.1, or 1.5.1 CHF should directly upgrade to 1.5.2-h6 and skip all the intermediate releases like 1.5.1, 1.5.1 CHF, and 1.5.2. There were some issues discovered with ML upgrade paths to 1.5.1 and 1.5.2 which have been fixed with 1.5.2-h6. For more details, please refer to the [KB](#) article.

Before you begin

- Review the [Software Support Matrix for OpenShift](#).
- Make a backup of the OpenShift routes before upgrading to 1.5.3.



Important:

Ensure that the OpenShift namespace name is 29 characters or less. Do not proceed with the upgrade if the namespace name is 30 or more characters in length.

Ensure that you have the following before you upgrade:

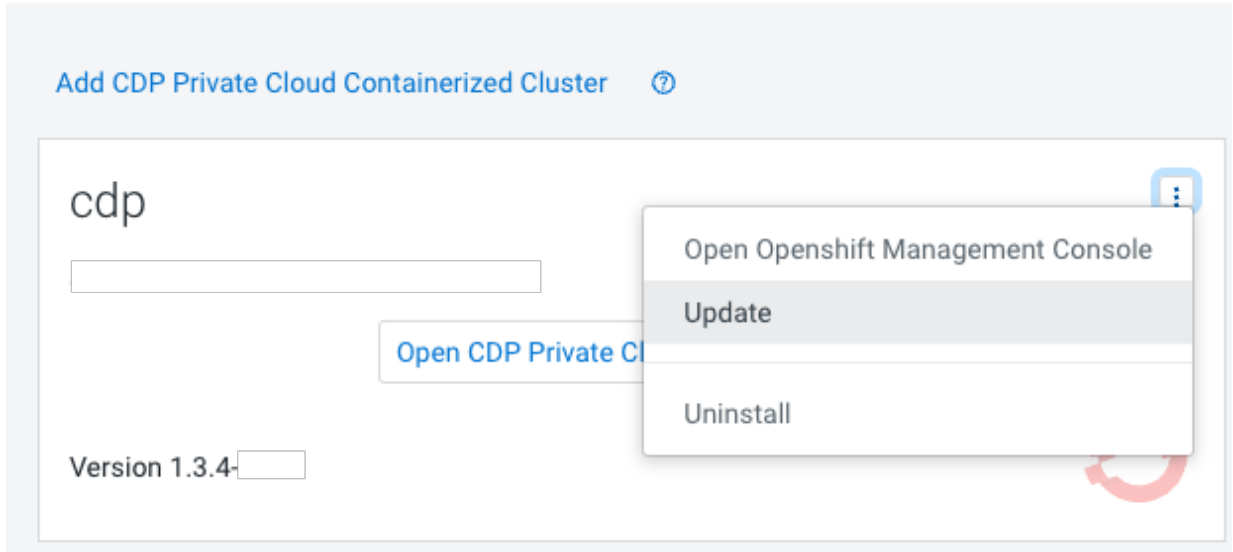
- A kubeconfig file for the OCP cluster
- Ensure that the kubeconfig file has permissions to create Kubernetes namespaces.
- Back up all the external databases used by CDP Private Cloud Data Services.
- One or more environments registered in CDP Private Cloud Data Services.
- One of the registered environment has one or more Cloudera Data Warehouse (CDW) or Cloudera Machine Learning (CML) experience workspaces.
- Access to the Cloudera Private Cloud repositories (archive.cloudera.com)
- Administrator access to OCP and Privileged access to your external Vault

Procedure

1.

In Cloudera Manager, navigate to CDP Private Cloud Data Services and click . Click Update.

CDP Private Cloud Data Services



2. On the Getting Started page, you can select the Install method - Air Gapped or Internet and proceed.

Internet install method

Update Private Cloud Data Services (cdp)

1 Getting Started

2 Collect Information

3 Update Progress

4 Summary

Getting Started

This wizard provides step-by-step guidance for updating CDP Private Cloud Data Services.

Visit the [CDP Private Cloud](#) documentation for more information.

Current Version
1.3.4

Install Method
 Internet Air Gapped

1. Select Repository

Please ensure all the Data Lake clusters are running Cloudera Runtime 7.1.6 or greater

You are about to update CDP Private Cloud Data Services to version 1.4.0. This is a **minor version update. Please make sure you have backed up all the external databases.**

Air Gapped install method

Update Private Cloud Data Services (cdp)

1 Getting Started

2 Collect Information

3 Update Progress

4 Summary

Getting Started

This wizard provides step-by-step guidance for updating CDP Private Cloud Data Services.

Visit the [CDP Private Cloud](#) documentation for more information.

Current Version
1.3.4

Install Method
 Internet Air Gapped

Installing via a local mirror with an http server. You will need to setup a full mirror of Cloudera's repositories via a temporary http server within the perimeter network of all hosts.

- Download everything under `https://archive.cloudera.com/p/cdp-pvc-ds/latest`

```
$ wget -l 0 --recursive --no-parent -e robots=off -nH --cut-dirs=2 --reject="index.html*" -t 10 https://<username>:<password>@archive.cloudera.com/p/cdp-pvc-ds/latest
```
- Modify the file `manifest.json` inside the downloaded directory, change `"http_url": "..."` to `"http_url": "http://your_local_repo/cdp-pvc-ds/latest"`
- Mirror the downloaded directory to your local http server, e.g. `http://your_local_repo/cdp-pvc-ds/latest`
- Add `http://your_local_repo/cdp-pvc-ds/latest` to your `Custom Repository` settings and select it from the dropdown below.
- Select Repository

Please ensure all the Data Lake clusters are running Cloudera Runtime 7.1.6 or greater

You are about to update CDP Private Cloud Data Services to version 1.4.0. This is a **minor version update. Please make sure you have backed up all the external databases.**

Click Continue.

3. On the Collect Information page, upload a Kubernetes configuration (kubeconfig) file from your existing environment. You can obtain this file from your OpenShift Container Platform administrator. Click Continue.

Update Private Cloud Data Services (cdp)

Getting Started

2 Collect Information

3 Update Progress

4 Summary

Collect Information

Kubernetes Environment

Kubernetes Configuration

[Choose File](#)

Kubernetes Cluster

4. On the Update Progress page, you can see the progress of your update. Click Continue.

Update Private Cloud Data Services (cdp)

- ✓ Getting Started
- ✓ Collect Information
- 3 Update Progress
- 4 Summary

Update Progress

Updating the CDP Private Cloud Management Console in the namespace cdp.

- ✓ Downloading the CDP Private Cloud update utility.
- ✓ Extracting the CDP Private Cloud update utility.
- ✓ Configuring and updating the helm charts.
- ✓ Waiting for all the pods to start or timeout.

▼ Show Logs

cdp-release-dps-gateway-1.0-cf7db56b-mt7tg	3/3	Running	0	3m48s
cdp-release-dwx-server-7587f9796f-h5555	2/2	Running	0	3m28s
cdp-release-dwx-ui-694757cc48-m2v1b	2/2	Running	0	3m20s
cdp-release-dwx-ui-694757cc48-mtnv1	2/2	Running	0	3m27s
cdp-release-grafana-59fdd998fd-679b2	3/3	Running	0	3m19s
cdp-release-logger-alert-receiver-86d67cdfb-hhgfs	2/2	Running	0	3m18s



Important:

During the "Upgrade Control Plane" step of the CDP upgrade process, the grafana pod can get stuck in the terminating state. This usually means that all other Control Plane pods are in the running state, but for Grafana, there is one pod that is in running state and there is one pod that is stuck in terminating state. The terminating pod has the following message:

```
containers with incomplete status: [multilog-init grafana-sc-datasources]
```

If you search for the terminating pod id in the kubelet log on the host, the following error message can be found:

```
E0531 2209 kuberuntime_sandbox.go:70] CreatePodSandbox for pod "<pod id>" failed: rpc error: code = Unknown desc = error reading container (probably exited) json message: EOF
```

If there is a grafana pod stuck in terminating state, run the following command on the ECS Server host:

```
<grafana-pod-id> --force --grace-period=0
```

5. After the update is complete, the Summary page appears. You can now Launch CDP Private Cloud from here.

Update Private Cloud Data Services (cdp)

- ✓ Getting Started
- ✓ Collect Information
- ✓ Update Progress
- 4 Summary

Summary

CDP Private Cloud has been updated. Click Launch CDP Private Cloud to proceed.

Launch CDP Private Cloud

Or you can navigate to the CDP Private Cloud Data Services page and click Open CDP Private Cloud Data Services.

CDP Private Cloud Data Services opens up in a new window.

6. After the update is complete, delete the old CDE service and the underlying virtual clusters. You may also need to upgrade the underlying OpenShift Container Platform. See the [Software Support Matrix for OpenShift](#) for more information about supported OCP versions.

Pre-upgrade - Preparing for CDP Private Cloud Data Services update for CDE

Upgrading the OpenShift Container Platform (OCP) version while CDE service is enabled, can cause database corruption in the embedded MySQL database used for CDE. Follow the below steps before starting the OCP version upgrade.

Procedure

1. Stop running jobs and pause scheduled workloads
 - a) Kill all the running Spark jobs in the CDE virtual clusters under all CDE services or wait for them to complete.
 - b) Pause all Airflow jobs and scheduled Spark jobs.
2. Identifying the CDE Namespace
 - a) Navigate to the Cloudera Data Engineering Overview page by clicking the Data Engineering tile in the Cloudera Data Platform (CDP) management console.
 - b) In the CDE Services column, click Service Details for the CDE service.
 - c) Note the Cluster ID shown in the page. For example, if the Cluster ID is *cluster-abcd1234*, then the CDE Namespace is *dex-base-abcd1234*.
 - d) Use this CDE Namespace (in the above example, it is *dex-base-abcd1234*) in the following instructions to run kubernetes commands.
3. Scale down CDE embedded database

Access the OpenShift cluster with OpenShift CLI or Kubernetes CLI, and scale down the CDE embedded database statefulset to 0 with the following command:

OpenShift CLI

```
oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 0
```

Kubernetes CLI

```
kubectl scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 0
```

Pre-upgrade - Upgrading CDE Service for Endpoint Stability

You can seamlessly upgrade an old Cloudera Data Engineering (CDE) service to a new version with endpoint stability. This enables you to access the CDE service of the new version with the previous endpoint. Thus, you can use the existing endpoints without changing configurations at the application level.

The CDE service endpoint migration process lets you migrate your resources, jobs, job run history, spark jobs' logs and event logs from your old cluster to the new cluster.

Prerequisites for upgrading CDE Service with endpoint stability

You must first download the docker image and create the `cde-upgrade-util.properties` file to back up CDE services.

Procedure

1. Download the [dex-upgrade-utils](#) docker image tarball. The file naming convention is `dex-upgrade-utils-<version-number>-<build-number>.tar.gz`.
2. Load the downloaded image into the host machine docker runtime:

```
docker load < dex-upgrade-utils-<version-number>-<build-number>.tar.gz
```

Example:

```
docker load < dex-upgrade-utils-1.20.1-b20.tar.gz
```

3. Create a directory in the host machine and export that path as `BASE_WORK_DIR`.

```
mkdir <host_machine_path>
export BASE_WORK_DIR=<host_machine_path>
```

Example:

```
mkdir /opt/backup-restore
export BASE_WORK_DIR=/opt/backup-restore
```

4. Create backup and secrets directories in the `BASE_WORK_DIR` directory and update the access permissions. The secrets directory stores the kubeconfig and CDP DE Admin credentials files. The backup directory will store the backup file which will be generated when you backup the CDE Service.

```
cd $BASE_WORK_DIR
mkdir backup secrets
chmod 775 backup
```

5. Place the CDP credentials file of the *DEAdmin* user and *administrator* kubeconfig file in the `$BASE_WORK_DIR/secrets` directory.
6. Create the `cde-upgrade-util.properties` file with the information described below and save it in the `$BASE_WORK_DIR` directory.

```
cdp_k8s_namespace:<CDP control plane k8s namespace>
cdp_endpoint:<CDP control plane endpoint>
credential_file_path:<Path to the DEAdmin user CDP credentials file>
de_admin_user:<DEAdmin user-id>
de_admin_password:<DEAdmin user's password must be in base64 encoded form
at. Use the echo -n <password> | base64 command to encode the password.>
tls_insecure:<Keep it true if you are using a self-signed certificate>
auto_unpause_jobs:<Specify it as "true" if you want to automatically res
ume the jobs that were paused during the backup phase. The jobs will be
resumed after you restore the CDE service.>
platform_type:<Platform type, can be ECS (Embedded Container Service) or
OCP (OpenShift Container Platform)>
use_stored_user:<(optional) Boolean property which can be true or false.
Use this property in conjunction with do-as described below.>
do_as:<(optional) if the value of use_stored_user is set to true, this v
alue is used as a fallback when the stored user is not valid. Otherwise,
this is directly used as job owner. If the use_stored_user parameter is
set to false and no value is supplied in the do_as parameter, then no
validation will be performed for the job's username and it will be resto
red as it is.>
```

Example:

```
$ cat $BASE_WORK_DIR/cde-upgrade-util.properties
cdp_k8s_namespace=cdp-test
```

```

cdp_endpoint=https://console-cdp-test.apps.cluster.example.com
credential_file_path=/home/dex/.cdp/credentials
de_admin_user=user
de_admin_password=cGFzc3dvcmQ=
tls_insecure=true
auto_unpause_jobs=true
platform_type=ECS
use_stored_user=true
do_as=cdp-test-user

```



Note: You must always set the value of the `credential_file_path` property as `/home/dex/.cdp/credentials` and should not be changed.



Warning: You can specify the `cdp_env_override:<environment-name>` optional property in the `cde-upgrade-util.properties` file, if you want to change the environment of the CDE service that is being restored. But, if you change the environment during restore, it will lead to loss of old spark jobs' logs and event logs that were there in old virtual clusters.

7. Make a note of the details of the CDE service that is being migrated. This information is required if you are using a CDP database that is external and is not accessible from the container which is running the `cde-upgrade` endpoint stability commands. Identify the cluster endpoint:
 - a. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
 - b. Click Administration in the left navigation menu. The Administration page displays.
 - c. In the Services column on the left, click the Cluster Details icon corresponding to the CDE service whose endpoint you want to migrate.
 - d. Make a note of the CDE cluster ID.
8. During the restore operation, both old and the new CDE services use the same resources allocated to existing CDE service. Hence, you must double the resource pool size using the Quota Management option. For example, if `root.default.sales` is the pool that is used for the old or existing CDE service, you must double the CPU and memory resources of this pool. Also, make sure that you have sufficient hardware when doubling the resource pool size.

What to do next

You must now expand the resource pool, and then upgrade your Cloudera Data Platform (CDP) Platform before you restore the CDE service. For information about configuring resource pool and capacity, see *Managing cluster resources using Quota Management*

Backing up CDE service using the docker image

You must run the docker image to take backup of a Cloudera Data Engineering (CDE) service. It takes backup of all the active virtual clusters in that CDE service. You can take backup of only an active CDE service.

Before you begin

You must download the `dex-upgrade-utils` docker image and create the `cde-upgrade-util.properties` file before backing up jobs as described in the *Prerequisites for migrating CDE endpoints* section.



Warning: You must make sure to allocate sufficient downtime before you proceed further. If you start the backup procedure, you cannot create, edit, or run jobs in the existing CDE service and its associated virtual clusters until the backup is complete. The virtual clusters will be in the read-only mode after you backup the service and until you restore it.



Important: It is recommended to copy the logs of the commands run from the terminal and save them on your machine. This will be helpful in debugging or when raising a support ticket. You can also increase the terminal buffer size and save the terminal logs of each command for reference.

Procedure

Run the dex-upgrade-utils docker image on the host machine:

```
$ export BACKUP_OUTPUT_DIR=/home/dex/backup

$ docker run \
-v <kubeconfig_file_path>:/home/dex/.kube/config:ro \
-v <cdp_credential_file_path>:/home/dex/.cdp/credentials:ro \
-v <cde-upgrade-util.properties_file_path>:/opt/cde-backup-restore/scripts/
backup-restore/cde-upgrade-util.properties:ro \
-v <local_backup_directory>:$BACKUP_OUTPUT_DIR \
-e KUBECONFIG=/home/dex/.kube/config \
<docker_image_name>:<docker_image_version> prepare-for-upgrade -s <cde-clust
er-id> -o $BACKUP_OUTPUT_DIR
```



Important: All the paths to the right side of colon (:) in volume mounts, that is, paths inside the container are fixed paths and must not be changed. Here -s is the CDE service ID and -o is the backup output filepath in the container. The backup output directory value must always be \$BACKUP_OUTPUT_DIR and should not be changed.

Example:

```
$ docker run \
-v $BASE_WORK_DIR/secrets/kubeconfig:/home/dex/.kube/config:ro \
-v $BASE_WORK_DIR/secrets/credentials:/home/dex/.cdp/credentials:ro \
-v $BASE_WORK_DIR/cde-upgrade-util.properties:/opt/cde-backup-restore/scr
ipts/backup-restore/cde-upgrade-util.properties:ro \
-v $BASE_WORK_DIR/backup:$BACKUP_OUTPUT_DIR \
-e KUBECONFIG=/home/dex/.kube/config \
docker-private.infra.cloudera.com/cloudera/dex/dex-upgrade-utils:1.20.1 pr
epare-for-upgrade -s cluster-c2dhkp22 -o $BACKUP_OUTPUT_DIR
```

Results

You have now taken the Cloudera Data Engineering (CDE) service backup as a ZIP file. You can make a note of the Zip file name from the logs to use it while restoring the CDE service.

What to do next

You must now expand the resource pool, and then upgrade your Cloudera Data Platform (CDP) Platform before you restore the CDE service. For information about configuring resource pool and capacity, see *Managing cluster resources using Quota Management*.

Related Information

[Upgrading CDP on OpenShift](#)

[Prerequisites for migrating CDE endpoints](#)

[Managing cluster resources using Quota Management](#)

OCP upgrade steps for CDP Private Cloud Data Services 1.5.3

Updating CDP Private Cloud Data Services 1.5.1 or 1.5.2 to 1.5.3 on OCP requires you to upgrade earlier OCP versions to a version supported by CDP Private Cloud Data Services 1.5.3. This process involves incremental OCP version upgrades in conjunction with CDP Private Cloud Data Services updates and validations.

Before you begin

**Important:**

Ensure that the OpenShift namespace name is 29 characters or less. Do not proceed with the upgrade if the namespace name is 30 or more characters in length.

OCP upgrade steps for CDP Private Cloud Data Services 1.5.1 to 1.5.3

Starting point: CDP Private Cloud Data Services 1.5.1 on OCP 4.11.

1. Update CDP Private Cloud Data Services from 1.5.1 to 1.5.3.
2. Perform post-upgrade validations of existing resources and workloads.
3. Upgrade OCP from 4.11 to 4.12.
4. Upgrade new and existing resources as required.
5. Perform validations on all data services to verify that the new and preexisting data and workloads continue to function properly.

OCP upgrade steps for CDP Private Cloud Data Services 1.5.2 to 1.5.3

Starting point: CDP Private Cloud Data Services 1.5.2 on OCP 4.12.

1. Update CDP Private Cloud Data Services from 1.5.2 to 1.5.3.
2. Upgrade new and existing resources as required.
3. Perform validations on all data services to verify that the new and preexisting data and workloads continue to function properly.

Upgrading CDP Private Cloud Data Services

You can upgrade your existing CDP Private Cloud Data Services 1.5.1 or 1.5.2 to 1.5.3 without performing an uninstall. After the upgrade is complete, you may need to upgrade the underlying OpenShift Container Platform. See the [Software Support Matrix for OpenShift](#) for more information about supported OCP versions.

Before you begin

- Review the [Software Support Matrix for OpenShift](#).
- Make a backup of the OpenShift routes before upgrading to 1.5.2.

**Important:**

Ensure that the OpenShift namespace name is 29 characters or less. Do not proceed with the upgrade if the namespace name is 30 or more characters in length.

Ensure that you have the following before you upgrade:

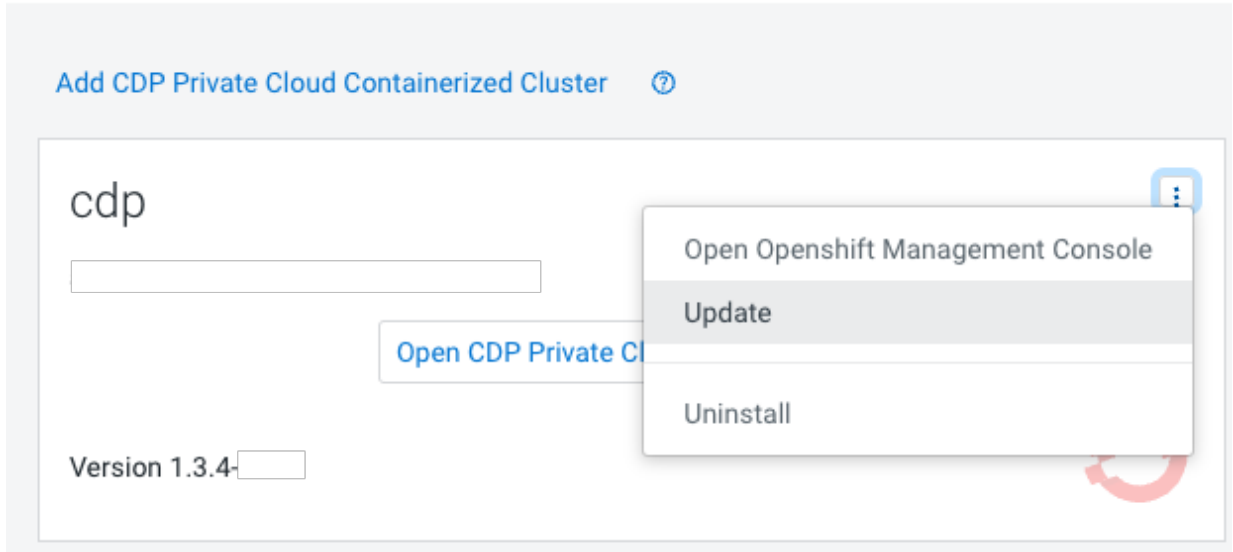
- A kubeconfig file for the OCP cluster.
- Ensure that the kubeconfig file has permissions to create Kubernetes namespaces.
- Back up all of the external databases used by CDP Private Cloud Data Services.
- One or more environments registered in CDP Private Cloud Data Services.
- One of the registered environment has one or more Cloudera Data Warehouse (CDW) or Cloudera Machine Learning (CML) experience workspaces.
- Access to the Cloudera Private Cloud repositories (archive.cloudera.com).
- Administrator access to OCP and Privileged access to your external Vault.

Procedure

1.

In Cloudera Manager, navigate to CDP Private Cloud Data Services and click . Click Update.

CDP Private Cloud Data Services



- On the Getting Started page, you can select the Install method - Air Gapped or Internet and proceed.

Internet install method

Update Private Cloud Data Services (cdp)

The screenshot shows the 'Getting Started' page for the Internet install method. The left sidebar has four steps: 1. Getting Started (selected), 2. Collect Information, 3. Update Progress, and 4. Summary. The main content area is titled 'Getting Started' and contains the following text:

This wizard provides step-by-step guidance for updating CDP Private Cloud Data Services. Visit the [CDP Private Cloud](#) documentation for more information.

Current Version: 1.3.4

Install Method: Internet Air Gapped

1. Select Repository:

Two warning messages are displayed at the bottom:

- Please ensure all the Data Lake clusters are running Cloudera Runtime 7.1.6 or greater
- You are about to update CDP Private Cloud Data Services to version 1.4.0. This is a **minor version** update. Please make sure you have backed up all the external databases.

Air Gapped install method

Update Private Cloud Data Services (cdp)

The screenshot shows the 'Getting Started' page for the Air Gapped install method. The left sidebar has four steps: 1. Getting Started (selected), 2. Collect Information, 3. Update Progress, and 4. Summary. The main content area is titled 'Getting Started' and contains the following text:

This wizard provides step-by-step guidance for updating CDP Private Cloud Data Services. Visit the [CDP Private Cloud](#) documentation for more information.

Current Version: 1.3.4

Install Method: Internet Air Gapped

Installing via a local mirror with an http server. You will need to setup a full mirror of Cloudera's repositories via a temporary http server within the perimeter network of all hosts.

- Download everything under `https://archive.cloudera.com/p/cdp-pvc-ds/latest`

```
$ wget -l 0 --recursive --no-parent -e robots=off -nH --cut-dirs=2 --reject="index.html*" -t 10 https://<username>:<password>@archive.cloudera.com/p/cdp-pvc-ds/latest
```
- Modify the file `manifest.json` inside the downloaded directory, change `"http_url": "..."` to `"http_url": "http://your_local_repo/cdp-pvc-ds/latest"`
- Mirror the downloaded directory to your local http server, e.g. `http://your_local_repo/cdp-pvc-ds/latest`
- Add `http://your_local_repo/cdp-pvc-ds/latest` to your `Custom Repository` settings and select it from the dropdown below.
- Select Repository:

Two warning messages are displayed at the bottom:

- Please ensure all the Data Lake clusters are running Cloudera Runtime 7.1.6 or greater
- You are about to update CDP Private Cloud Data Services to version 1.4.0. This is a **minor version** update. Please make sure you have backed up all the external databases.

Click Continue.

3. On the Collect Information page, upload a Kubernetes configuration (kubeconfig) file from your existing environment. You can obtain this file from your OpenShift Container Platform administrator. Click Continue.

Update Private Cloud Data Services (cdp)

Getting Started

2 Collect Information

3 Update Progress

4 Summary

Collect Information

Kubernetes Environment

Kubernetes Configuration

[Choose File](#)

Kubernetes Cluster

4. On the Update Progress page, you can see the progress of your upgrade. Click Continue.

Update Private Cloud Data Services (cdp)

- ✓ Getting Started
- ✓ Collect Information
- 3 Update Progress
- 4 Summary

Update Progress

Updating the CDP Private Cloud Management Console in the namespace cdp.

- ✓ Downloading the CDP Private Cloud update utility.
- ✓ Extracting the CDP Private Cloud update utility.
- ✓ Configuring and updating the helm charts.
- ✓ Waiting for all the pods to start or timeout.

▼ Show Logs

cdp-release-dps-gateway-1.0-cf7db56b-mt7tg	3/3	Running	0	3m48s
cdp-release-dwx-server-7587f9796f-h5555	2/2	Running	0	3m28s
cdp-release-dwx-ui-694757cc48-m2v1b	2/2	Running	0	3m20s
cdp-release-dwx-ui-694757cc48-mtnv1	2/2	Running	0	3m27s
cdp-release-grafana-59fdd998fd-679b2	3/3	Running	0	3m19s
cdp-release-logger-alert-receiver-86d67cdfb-hhgfs	2/2	Running	0	3m18s



Important:

During the "Upgrade Control Plane" step of the CDP upgrade process, the grafana pod can get stuck in the terminating state. This usually means that all other Control Plane pods are in the running state, but for grafana, there is one pod that is in running state and there is one pod that is stuck in terminating state. The terminating pod has the following message:

```
containers with incomplete status: [multilog-init grafana-sc-datasources]
```

If you search for the terminating pod id in the kubelet log on the host, the following error message can be found:

```
E0531 2209 kuberuntime_sandbox.go:70] CreatePodSandbox for pod "<pod id>" failed: rpc error: code = Unknown desc = error reading container (probably exited) json message: EOF
```

If there is a grafana pod stuck in terminating state, run the following command on the ECS Server host:

```
<grafana-pod-id> --force --grace-period=0
```

5. After the upgrade is complete, the Summary page appears. You can now Launch CDP Private Cloud from here.

Update Private Cloud Data Services (cdp)

- ✓ Getting Started
- ✓ Collect Information
- ✓ Update Progress
- 4 Summary

Summary

CDP Private Cloud has been updated. Click Launch CDP Private Cloud to proceed.

Launch CDP Private Cloud

Or you can navigate to the CDP Private Cloud Data Services page and click Open CDP Private Cloud Data Services.

CDP Private Cloud Data Services opens up in a new window.

What to do next

- After upgrade, the Cloudera Manager admin role may be missing the Host Administrators privilege in an upgraded cluster. The cluster administrator should run the following command to manually add this privilege to the role.

```
ipa role-add-privilege <cmadminrole> --privileges="Host Administrators"
```

Completing post OCP update tasks

If you are using CDE, after you complete the OpenShift Container Platform (OCP) upgrade, ensure that the steps done before OCP upgrade are reverted.

Procedure

- Unpause all Airflow jobs and scheduled Spark jobs that were paused before upgrade.
- Scale back the CDE embedded database statefulset to 1.

OpenShift CLI

```
oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 1
```

Kubernetes CLI

```
kubectl scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 1
```

Post-upgrade - Ozone Gateway validation

If you are using CDE, after upgrading CDP Private Cloud Data Services you must validate that the Ozone Gateway is working as expected. This procedure applies to both 1.5.1 CHF1 and 1.5.2 to 1.5.3 upgrades.

About this task

You can run the following commands to get the types of logs that are available with the job run.

Command 1

```
cde run logs --id <run_id> --show-types --vcluster-endpoint <job_api_url> --cdp-endpoint <cdp_control_plane_endpoint> --tls-insecure
```

For example,

```
cde run logs --id 8 --show-types --vcluster-endpoint https://76fsk4rz.cde-fm-ttv45d.apps.apps.shared-rke-dev-01.kcloud.cloudera.com/dex/api/v1 --cdp-endpoint https://console-cdp-keshaw.apps.shared-rke-dev-01.kcloud.cloudera.com --tls-insecure
```

Log:

TYPE	ENTITY	STREAM	ENTITY DEFAULT
driver/stderr	Driver	stderr	True
driver/stdout	Driver	stdout	False

TYPE	ENTITY	STREAM	ENTITY DEFAULT
executor_1/stderr	Executor 1	stderr	True
executor_2/stdout	Executor 2	stdout	False

Command 2

```
cde run logs --id <run_id> --type <log_type> --vcluster-endpoint <job_api_url>
--cdp-endpoint <cdp_control_plane_endpoint> --tls-insecure
```

For example,

```
cde run logs --id 8 --type driver/stderr --vcluster-endpoint https://76fsk4r
z.cde-fmttv45d.apps.apps.shared-rke-dev-01.kcloud.cloudera.com/dex/api/v1 --
cdp-endpoint https://console-cdp-keshaw.apps.shared-rke-dev-01.kcloud.cloude
ra.com --tls-insecure
```

Log:

```
Setting spark.hadoop.yarn.resourcemanager.principal to hive
23/05/22 09:27:28 INFO SparkContext: Running Spark version 3.2.3.1.20.71720
00.0-38
23/05/22 09:27:28 INFO ResourceUtils: =====
=====
23/05/22 09:27:28 INFO ResourceUtils: No custom resources configured for sp
ark.driver.
23/05/22 09:27:28 INFO ResourceUtils: =====
=====
23/05/22 09:27:28 INFO SparkContext: Submitted application: PythonPi
23/05/22 09:27:28 INFO ResourceProfile: Default ResourceProfile created, e
xecutor resources: Map(cores -> name: cores, amount: 1, script: , vendor: ,
memory -> name: memory, amount: 1024, script: , vendor: , offHeap -> name: o
ffHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cp
us, amount: 1.0)
23/05/22 09:27:29 INFO ResourceProfile: Limiting resource is cpus at 1 tasks
per executor
23/05/22 09:27:29 INFO ResourceProfileManager: Added ResourceProfile id: 0
23/05/22 09:27:29 INFO SecurityManager: Changing view acls to: sparkuser,c
dpuser1
23/05/22 09:27:29 INFO SecurityManager: Changing modify acls to: sparkuser,c
dpuser1
23/05/22 09:27:29 INFO SecurityManager: Changing view acls groups to:
23/05/22 09:27:29 INFO SecurityManager: Changing modify acls groups to:
23/05/22 09:27:29 INFO SecurityManager: SecurityManager: authentication en
abled; ui acls disabled; users with view permissions: Set(sparkuser, cdpuse
r1); groups with view permissions: Set(); users with modify permissions: Se
t(sparkuser, cdpuser1); groups with modify permissions: Set()
.....
.....
```

Results

- If you can see the driver pod logs, then Ozone Gateway is working as expected and you can go ahead with the upgrade.
- If the logs do not appear, then you can try restarting the Ozone Gateway and get Spark job's driver log to validate if Ozone gateway is healthy or not.
- If you do not get the Spark job driver log, then you must contact Cloudera Support.
- For more information about configuring CDE CLI, see [Using the Cloudera Data Engineering command line interface](#)

Recovering a corrupted CDE Embedded database

In case you did not stop the jobs and scale down CDE embedded databases but completed the upgrade of OpenShift Container Platform (OCP), there is a chance of the CDE embedded database getting corrupted which causes the virtual clusters to become inaccessible. Follow the below steps to recover the CDE embedded database.

Procedure

1. Identifying the CDE Namespace

- a) Navigate to the Cloudera Data Engineering Overview page by clicking the Data Engineering tile in the Cloudera Data Platform (CDP) management console.
- b) In the CDE Services column, click Service Details for the CDE service.
- c) Note the Cluster ID shown in the page. For example, if the Cluster ID is *cluster-abcd1234*, then the CDE Namespace is *dex-base-abcd1234*.
- d) Use this CDE Namespace (in the above example, it is *dex-base-abcd1234*) in the following instructions to run kubernetes commands.

2. Edit the dex-base-db-server-config configuration map and add the `innodb_force_recovery=4` configuration in the `[mysqld]` section.

OpenShift CLI

```
oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 1
```

Kubernetes CLI

```
kubectl scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 1
```

Example snippet:

```
# Please edit the object below. Lines beginning with a '#' will be ignored,
# and an empty file will abort the edit. If an error occurs while saving
# this file
# will be reopened with the relevant failures.
#
apiVersion: v1
data:
  my.cnf: |-
    [mysqld]
    port=3306
    default_authentication_plugin = mysql_native_password
    bind-address = 0.0.0.0
    innodb_force_recovery=4
```

3. Scale down and then back up the CDE embedded database statefulset to restart it.

OpenShift CLI

```
oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 0
```

```
oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 1
```

Kubernetes CLI

```
kubectl scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 0
```

```
kubectl scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 1
```

Wait for 10 minutes for the CDE embedded database to complete the recovery.

4. Edit the dex-base-db-server-config configuration map again by removing the previously added `innodb_force_recovery=4` configuration under the `[mysqld]` section.

OpenShift CLI

```
oc edit configmap/dex-base-db-server-config --namespace <CDE Namespace>
```

Kubernetes CLI

```
kubectl edit configmap/dex-base-db-server-config --namespace <CDE Namespace>
```

5. Scale down and then back up the CDE embedded database statefulset to restart it again.

OpenShift CLI

```
oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 0
```

```
oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 1
```

Kubernetes CLI

```
kubectl scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 0
```

```
kubectl scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --replicas 1
```

Wait for all the CDE Virtual Clusters to be accessible. This usually takes about 10 minutes.

Post-upgrade - Restoring CDE Service for Endpoint Stability

After you take backup of the CDE service and upgrade your CDP platform, you can restore the Cloudera Data Engineering (CDE) service with same endpoints.

Restoring a CDE service

You can restore the Cloudera Data Engineering (CDE) service with its jobs, resources, job run history, and job logs from a backed-up ZIP file.

Before you begin

You must back up the CDE service, expand the resource pool, and then upgrade your Cloudera Data Platform (CDP) Platform to restore the CDE service. Also, you must validate that the Ozone Gateway is working as expected by performing the steps listed in the *Post upgrade - Ozone Gateway validation* topic.



Important: It is recommended to copy the logs of the commands run from the terminal and save them on your machine. This will be helpful in debugging or when raising a support ticket. You can also increase the terminal buffer size and save the terminal logs of each command for reference.

Procedure

1. If you have exited from the previous terminal where the pre-upgrade commands were run for CDE Service being upgraded, then you have to export these variables before running any docker command.

```
export BASE_WORK_DIR=<host_machine_path>
export BACKUP_OUTPUT_DIR=/home/dex/backup
```

2. Run the dex-upgrade-utils docker image to restore the service on the same machine where you have completed the prerequisite steps.

```
$ docker run \
-v <kubeconfig_file_path>:/home/dex/.kube/config:ro \
-v <cdp_credential_file_path>:/home/dex/.cdp/credentials:ro \
-v <cde-upgrade-util.properties_file_path>:/opt/cde-backup-restore/scripts/backup-restore/cde-upgrade-util.properties:ro \
-v <local_backup_directory>:$BACKUP_OUTPUT_DIR \
-e KUBECONFIG=/home/dex/.kube/config \
<docker_image_name>:<docker_image_version> restore-service -s <cde-cluster-id> -f $BACKUP_OUTPUT_DIR/<backup-zip-file-name>
```

Here -s is the CDE service ID and -f is the backup output filepath in the container.

Example:

```
$ docker run \
-v $BASE_WORK_DIR/secrets/kubeconfig:/home/dex/.kube/config:ro \
-v $BASE_WORK_DIR/secrets/credentials:/home/dex/.cdp/credentials:ro \
-v $BASE_WORK_DIR/cde-upgrade-util.properties:/opt/cde-backup-restore/scripts/backup-restore/cde-upgrade-util.properties:ro \
-v $BASE_WORK_DIR/backup:$BACKUP_OUTPUT_DIR \
-e KUBECONFIG=/home/dex/.kube/config \
docker-private.infra.cloudera.com/cloudera/dex/dex-upgrade-utils:1.20.1
restore-service -s cluster-c2dhkp22 -f $BACKUP_OUTPUT_DIR/cluster-c2dhkp22-2023-03-10T06_00_05.zip
```

3. [Optional] If you are using a CDP database that is external and is not accessible from the container which is running the CDE upgrade command, then the following SQL statements are displayed in the logs.

Example:

```
2023-05-17 13:02:29,551 [INFO] CDP control plane database is external and
not accessible
2023-05-17 13:02:29,551 [INFO] Please rename the old & new cde service
name manually by executing below SQL statement
2023-05-17 13:02:29,551 [INFO]      update cluster set name = 'cde-base-
service-1' where id = 'cluster-92c2fkgb';
2023-05-17 13:02:29,551 [INFO]      update cluster set name = 'cde-
base-service-1-18-1' where id = 'cluster-rtc234g8';
2023-05-17 13:02:29,551 [INFO] Please update the lastupdated time of ol
d cde service in db to extend the expiry interval of db entry for suppor
ting CDE CLI after old CDE service cleanup
2023-05-17 13:02:29,551 [INFO]      update cluster set lastupdated =
'2023-05-05 06:16:37.786199' where id = 'cluster-rtc234g8';
```

You must execute the above SQL statements to complete the restore process.

If you have closed the terminal or do not have this information, run the following SQL statements and specify the cluster details. Use the cluster ID that you noted when performing the steps listed in the *Prerequisites for migrating CDE endpoints* section.

- a. Rename old CDE service to a different name.

```
update cluster set name = '<modified_service_name>' where id = '<old_cde_cluster_id>;'
```

Example:

```
update cluster set name = 'cde-base-service-1' where id = 'cluster-92c2fkgb'
```

- b. Rename the new CDE service to the old CDE service name.

```
update cluster set name = '<old_cde_service_name>' where id = '<new_cde_cluster_id>;'
```

Example:

```
update cluster set name = 'cde-base-service-1-18-1' where id = 'cluster-rtc234g8'
```

- c. Run the following command so that when the old CDE service is deleted/disabled then it is not cleared from the database for the next two years. The timestamp format must be the same and should be two years of the current time.

```
update cluster set lastupdated = 'YYYY-MM-DD hh:mm:ss[.nnn]' where id = '<old_cde_cluster_id>;'
```

Example:

```
update cluster set lastupdated = '2023-05-05 06:16:37.786199' where id = 'cluster-rtc234g8'
```

4. After the restore operation completes, validate that the jobs and resources are restored by running the `cde job list` and `cde resource list` CLI commands or check the virtual cluster job UI.

In the Administration page of the CDE UI, you can see the old CDE service is appended with a version number. For example, if the old CDE service name was `cde-sales`, after the restore, the old CDE service is something similar to `cde-sales-1-18.2`.

5. [Optional] You can now delete the old CDE service after validating that everything is working as expected. If you delete the old CDE service, then you can shrink the resource pool size back to its initial value which you expanded in the *Prerequisite* steps.

Related Information

[Upgrading CDP on OpenShift](#)

[Upgrading CDP on OpenShift](#)

[Ozone Gateway validation](#)

Rolling back the CDE Service Endpoint migration

You can use the rollback command to delete the new CDE service and restore the old CDE service in working condition.



Important: It is recommended to copy the logs of the commands run from the terminal and save them on your machine. This will be helpful in debugging or when raising a support ticket. You can also increase the terminal buffer size and save the terminal logs of each command for reference.

To rollback, the state of the CDE service must be in the Failed or Installed state before you restore it.

Procedure

Run the `rollback-restore-service` command.

```
docker run \
-v <kubeconfig_file_path>:/home/dex/.kube/config:ro \
-v <cdp_credential_file_path>:/home/dex/.cdp/credentials:ro \
-v <cde-upgrade-util.properties_file_path>:/opt/cde-backup-restore/scripts/
backup-restore/cde-upgrade-util.properties:ro \
-v <local_backup_directory>:$BACKUP_OUTPUT_DIR \
-e KUBECONFIG=/home/dex/.kube/config \
<docker_image_name>:<docker_image_version> rollback-restore-service -s <new-
new-service-id> -f <path-to-the-backup-file>
```

Example:

```
docker run \
-v $BASE_WORK_DIR/secrets/kubeconfig:/home/dex/.kube/config:ro \
-v $BASE_WORK_DIR/secrets/credentials:/home/dex/.cdp/credentials:ro \
-v $BASE_WORK_DIR/cde-upgrade-util.properties:/opt/cde-backup-restore/script
s/backup-restore/cde-upgrade-util.properties:ro \
-v $BASE_WORK_DIR/backup:$BACKUP_OUTPUT_DIR \
-e KUBECONFIG=/home/dex/.kube/config \
docker-private.infra.cloudera.com/cloudera/dex/dex-upgrade-utils:1.20.1 r
ollback-restore-service -s cluster-c2dhkp25 -f $BACKUP_OUTPUT_DIR/cluster-c
2dhkp22-2023-03-10T06_00_05.zip
```

Limitations

This page lists the current limitations that you might run into while migrating your CDE service endpoint.

- Airflow job logs of the old cluster will be lost after the Restore operation.
- The Spark UI tab for a completed job does not work on the first click. As a workaround, do the following:
 1. Click the Spark UI tab. Nothing is displayed.
 2. Click on some other tab. For example, the Logs tab.
 3. Click the Spark UI tab again. The Spark UI loads now.