CDP Private Cloud Data Services 1.5.0

Upgrading CDP Private Cloud Data Services on the OpenShift Container Platform

Date published: 2020-12-16
Date modified: 2023-1-24

CLOUDERA

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

Update from 1.4.0-H1 or 1.4.1 to 1.5.0 (OCP).....4
Preparing for CDP Private Cloud Data Services update for CDE.....9
Updating CDP Private Cloud Data Services.....9
Completing post OCP update tasks.....14
Recovering a corrupted CDE Embedded database.....14
Migrating CDE Service Endpoint.....16

Update from 1.4.0-H1 or 1.4.1 to 1.5.0 (OCP)

You can update your existing CDP Private Cloud Data Services 1.4.0-H1 or 1.4.1 to 1.5.0 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.

Before you begin

Review the Software Support Matrix for OpenShift.

Ensure that you have the following before you update:

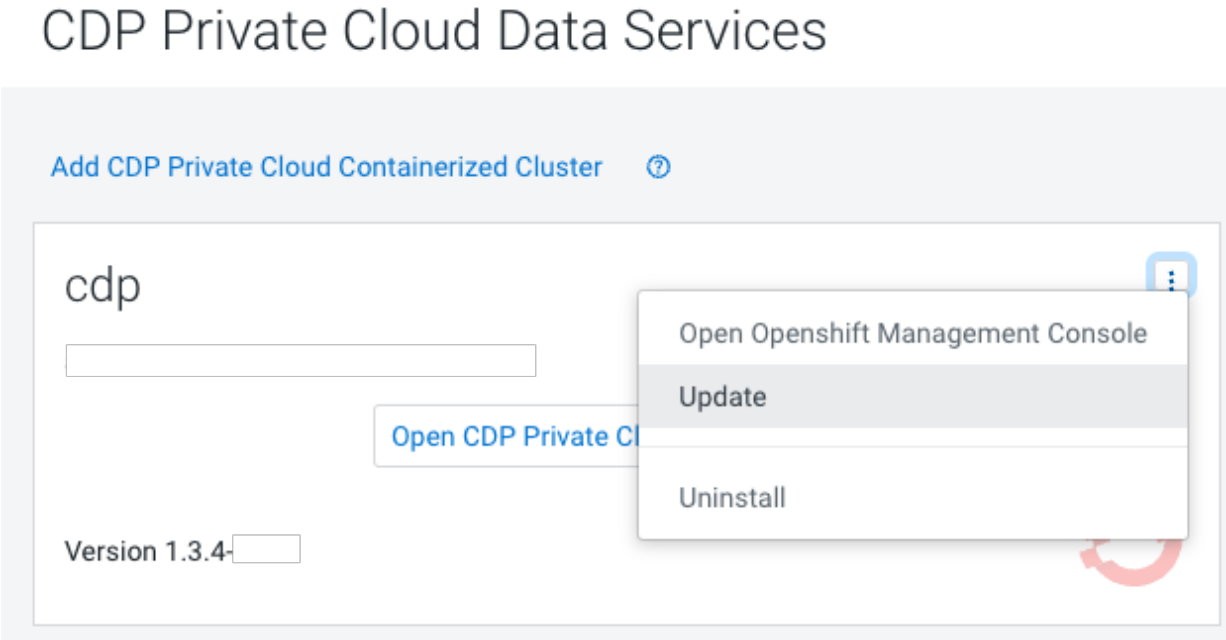
- Ensure that this kubeconfig has permissions to create Kubernetes namespaces.
- Back up all the external databases used by CDP Private Cloud Data Services.
- One or more environments registered in CDP Private Cloud Data Services.
- One of the registered environment has one or more Cloudera Data Warehouse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

Important:

If you are updating from CDP Private Cloud Data Services 1.4.0-H1 or 1.4.1 and you would like to use Cloudera Runtime 7.1.7 SP2, you must first upgrade the Cloudera Manager version to the Cloudera Manager 7.9.5-h1 hotfix release before proceeding with the CDP Private Cloud Data Services update.

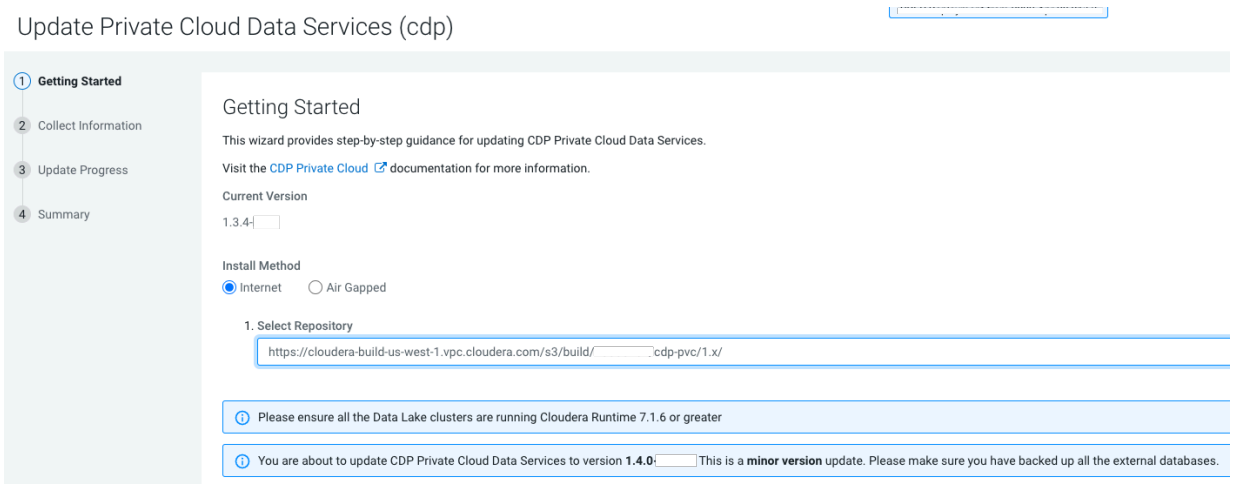
Procedure

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

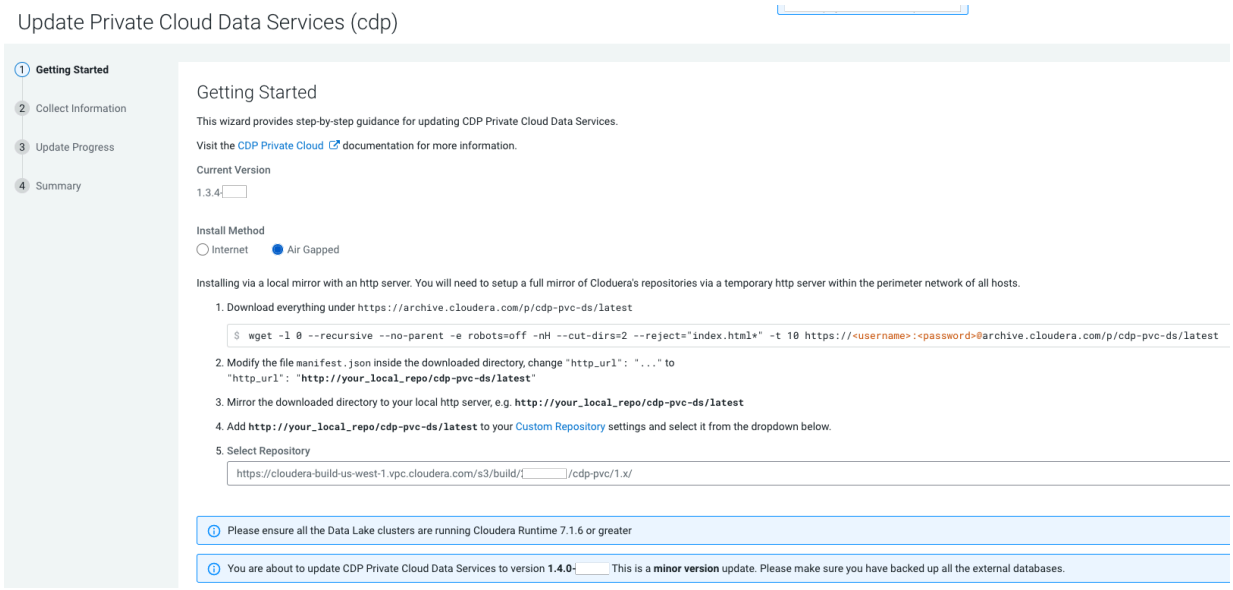


2. On the Getting Started page, you can select the Install method - Air Gapped or Internet and proceed.

Internet install method

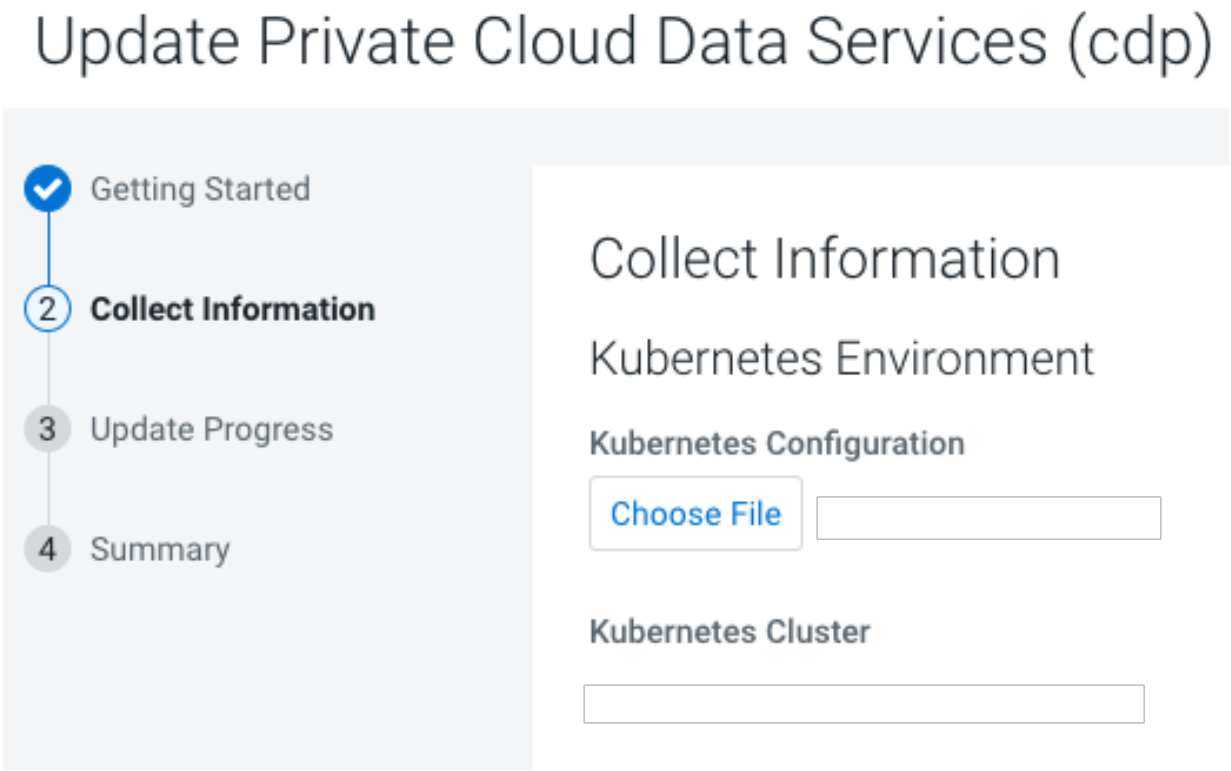


Air Gapped install method

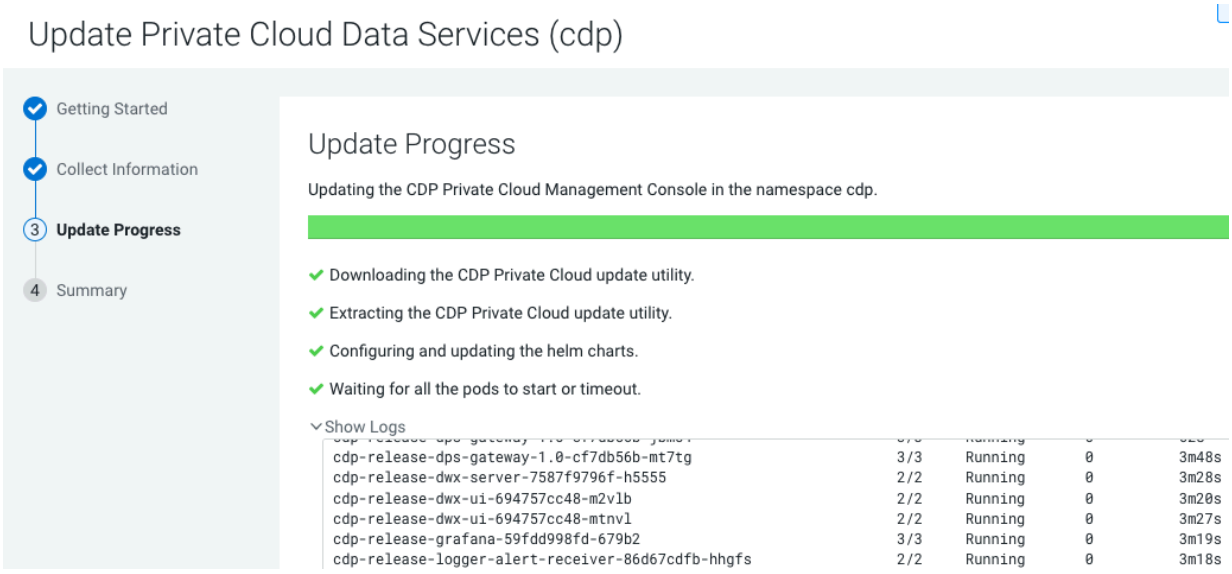


Click Continue.

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.



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



Important:

During the "Upgrade Control Plane" step of the CDP upgrade process, the grafana pod can get stuck in the terminating sate. 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-datasou
rces]
```

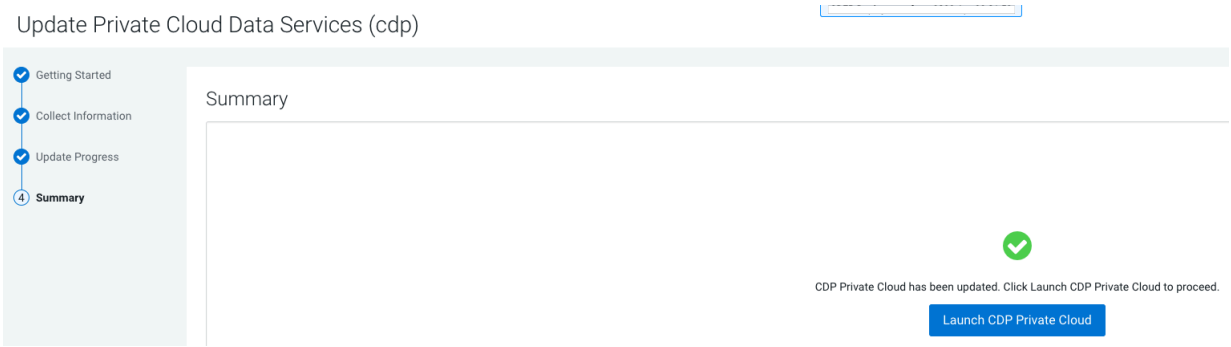
If you search for the terminating pod id in the kubelet log on the host, the following error message can be found:

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

If there is a granfa 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.



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 cluters. 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.

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> --r
   eplicas 0
   ```

   Kubernetes CLI

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

Updating CDP Private Cloud Data Services

You can update your existing CDP Private Cloud Data Services 1.4.0-H1 or 1.4.1 to 1.5.0 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.

Before you begin

Review the Software Support Matrix for OpenShift.

Ensure that you have the following before you update:

- Ensure that the kubeconfig 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.

- A Kubeconfig file for the OCP cluster.
- Access to the Cloudera Private Cloud repositories (archive.cloudera.com).
- Administrator access to OCP and Privileged access to your external Vault.

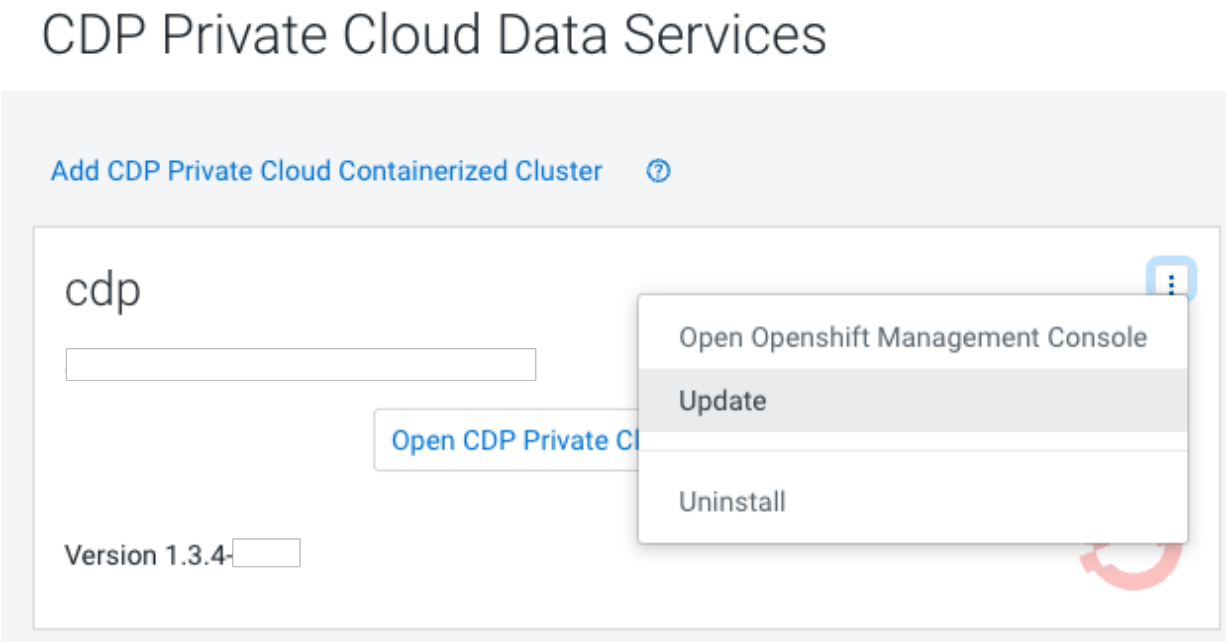
Important:

If you are updating from CDP Private Cloud Data Services 1.4.0-H1 or 1.4.1 and you would like to use Cloudera Runtime 7.1.7 SP2, you must first upgrade the Cloudera Manager version to the Cloudera Manager 7.9.5-h1 hotfix release before proceeding with the CDP Private Cloud Data Services update.

Procedure

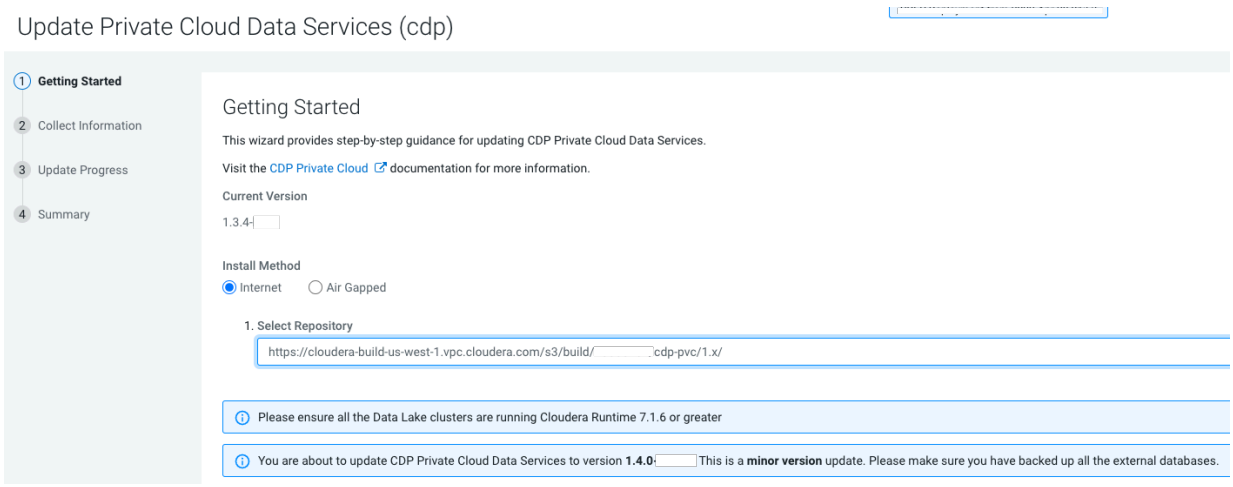
1.

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

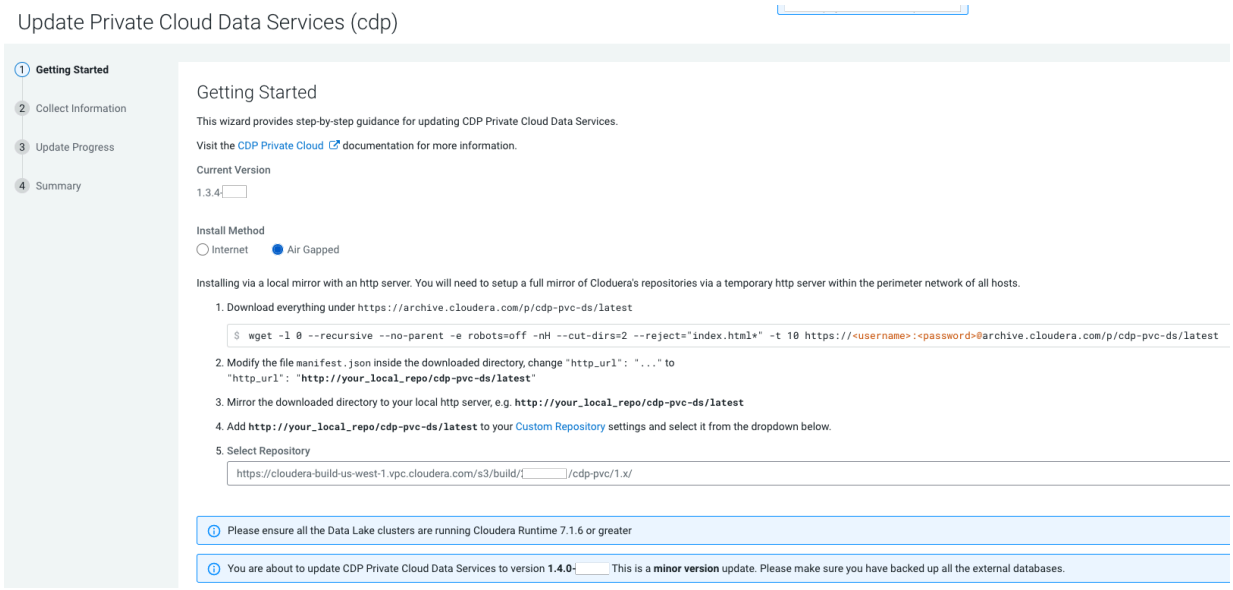


2. On the Getting Started page, you can select the Install method - Air Gapped or Internet and proceed.

Internet install method

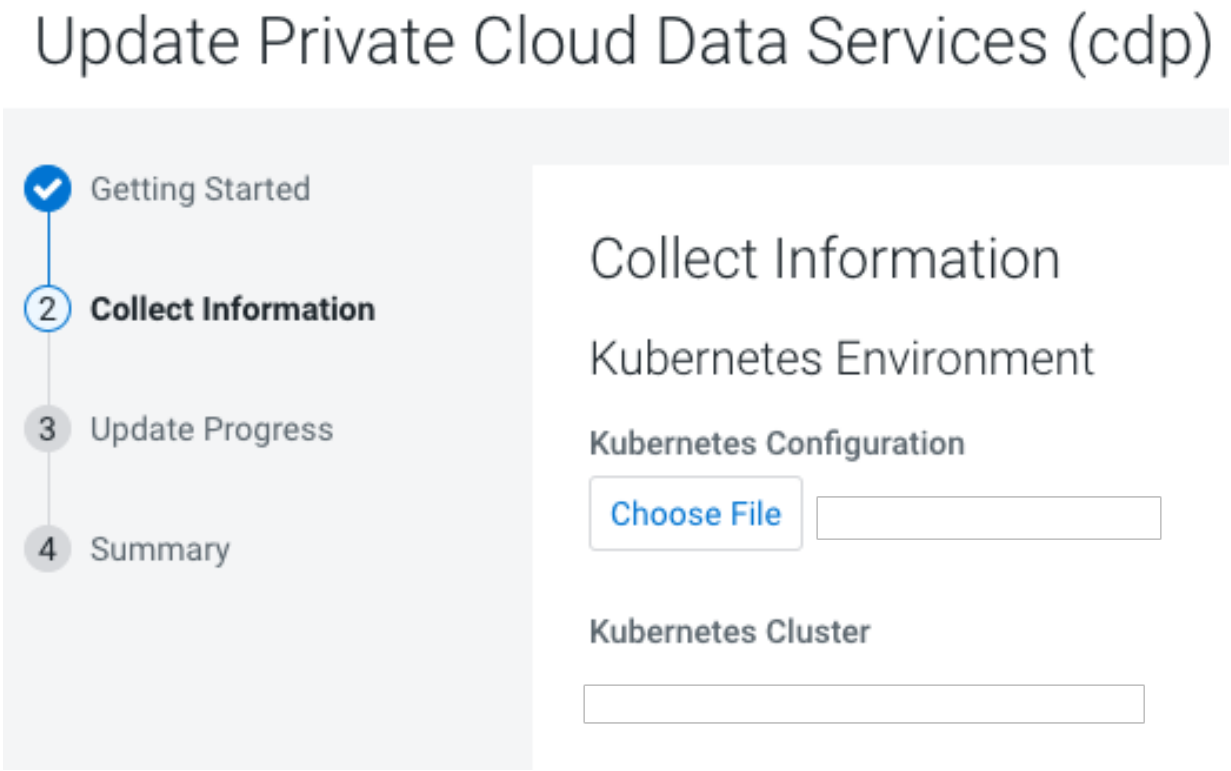


Air Gapped install method

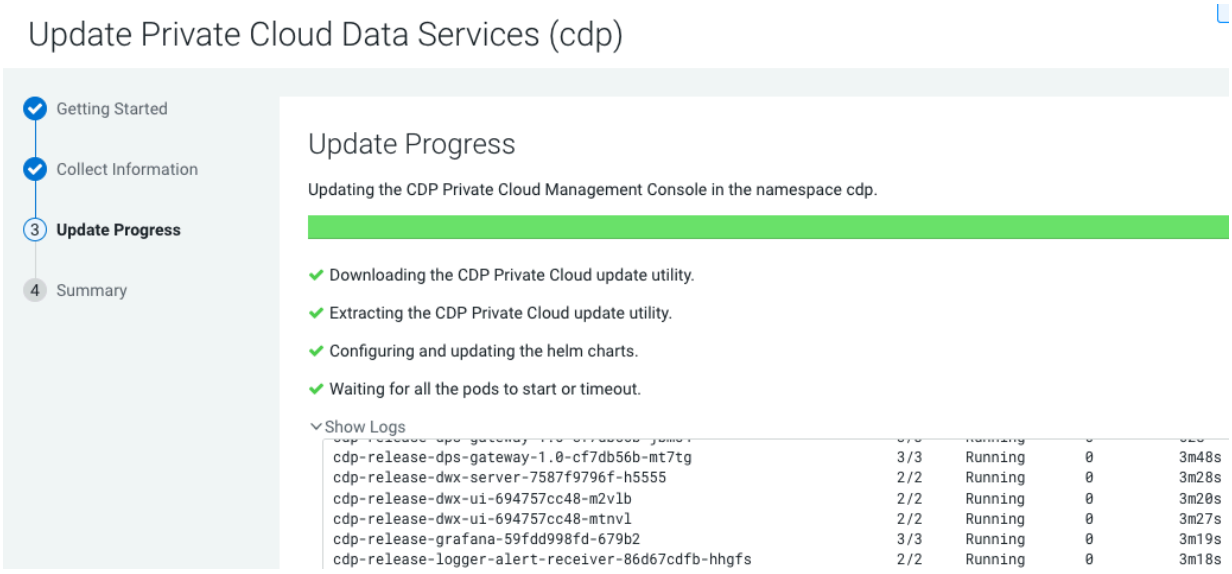


Click Continue.

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.



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



Important:

During the "Upgrade Control Plane" step of the CDP upgrade process, the grafana pod can get stuck in the terminating sate. 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-datasou
rces]
```

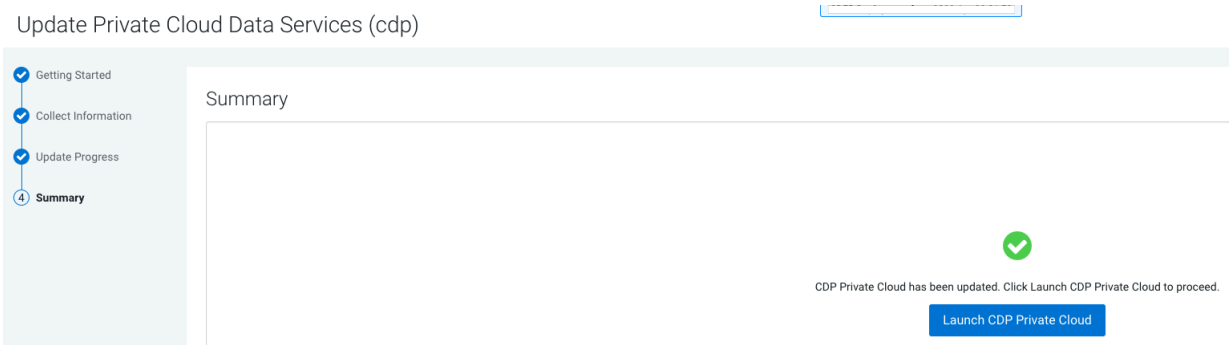
If you search for the terminating pod id in the kubelet log on the host, the following error message can be found:

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

If there is a granfa 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.



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.

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

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

   OpenShift CLI

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

   Kubernetes CLI

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

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> --r
   eplicas 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 ignor
   ed,
   # 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> --r
   eplicas 0

   oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --re
   plicas 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 Namesp
   ace>
   ```

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> --r
   eplicas 0

   oc scale statefulset/cdp-cde-embedded-db --namespace <CDE Namespace> --re
   plicas 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.

Migrating CDE Service Endpoint

You can now manually migrate an old CDE service endpoint to another new CDE service seamlessly. Migration provides endpoint stability and enables you to access the new CDE version with the previous endpoint. In upgrade scenarios, you can use the latest CDE version with the existing endpoints without changing configurations at the application level.

Contact your Cloudera Account team to help you in migrating your CDE Service endpoint manually.