Cloudera Data Services on premises 1.5.5

Upgrading Cloudera Data Services on premises on the OpenShift Container Platform

Date published: 2023-12-16
Date modified: 2025-11-08

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

Upgrading Cloudera Data Services on premises using OpenShift Container Platform

You can upgrade Cloudera Data Services on premises 1.5.4 or 1.5.5 versions to 1.5.5 SP1 version without uninstalling existing installation. 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 Cloudera Data Engineering virtual cluster. Hence, you must back up Cloudera Data Engineering jobs in the Cloudera Data Engineering virtual cluster, and then delete the Cloudera Data Engineering service and Cloudera Data Engineering virtual cluster. Restore it after the upgrade. For more information about backup and restore Cloudera Data Engineering jobs, see Backing up and restoring Cloudera Data Engineering jobs.

Before you begin

- Review the Software Support Matrix for OpenShift.
- Make a backup of the OpenShift routes before upgrading to 1.5.5 SP1.

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.

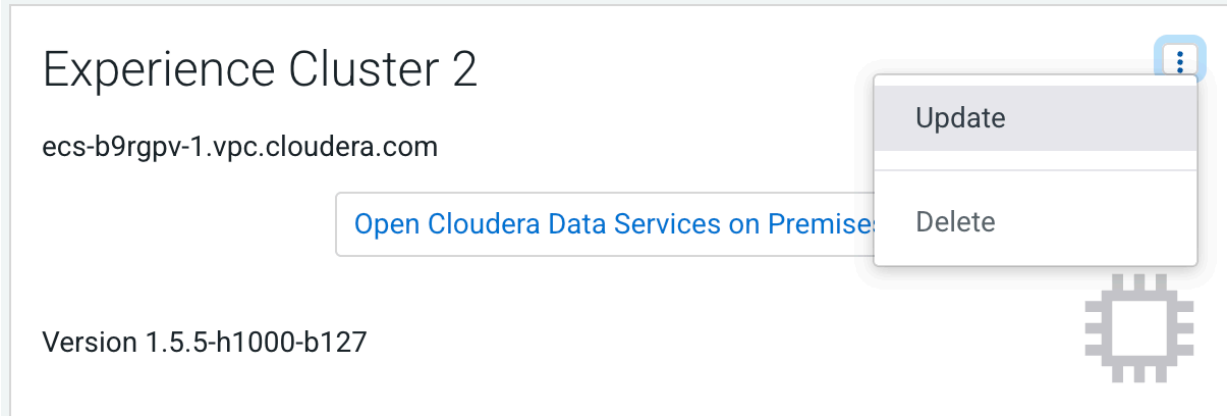
Ensure that you have the following before you upgrade:

- 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 Cloudera Data Services on premises.
- One or more environments registered in Cloudera Data Services on premises.
- One of the registered environment has one or more Cloudera Data Warehouse or Cloudera AI experience workspaces.
- Access to the Cloudera on premises repositories (archive.cloudera.com)
- Administrator access to OCP and Privileged access to your external Vault

Note: Cloudera Data Services on premisesß require the tmpfs file systems (example: '/tmp') mounts to be without the 'noexec' flag.

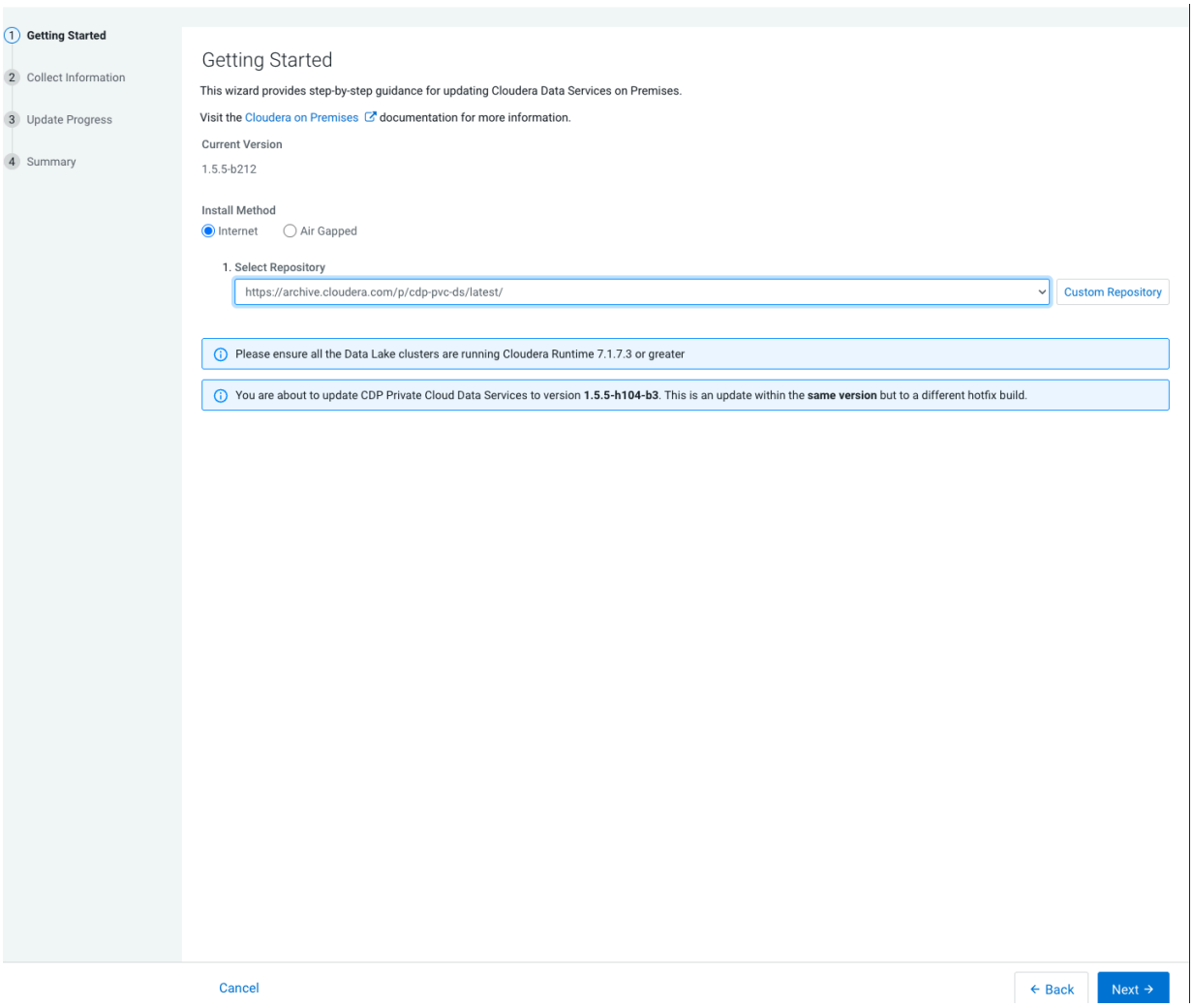
Procedure

1. In Cloudera Manager, navigate to Cloudera Data Services on premises and click . Click Update.

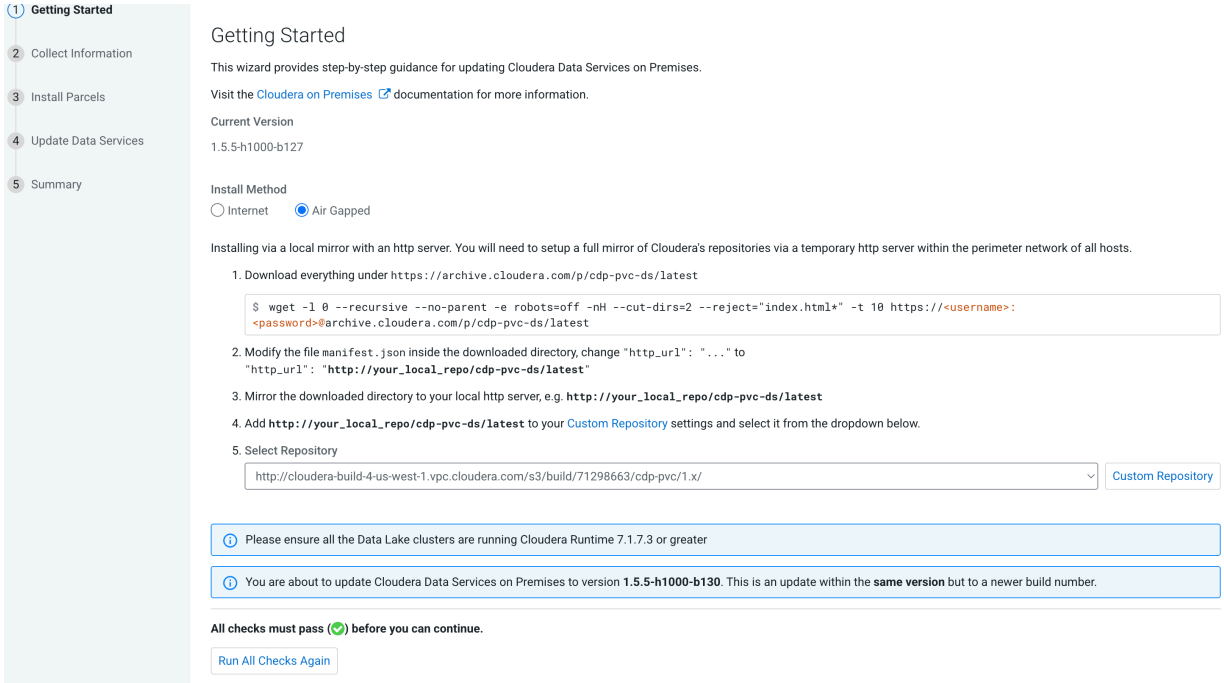


2. On the Getting Started page, you can select the Install method - Air Gapped or Internet and proceed.

Internet install method

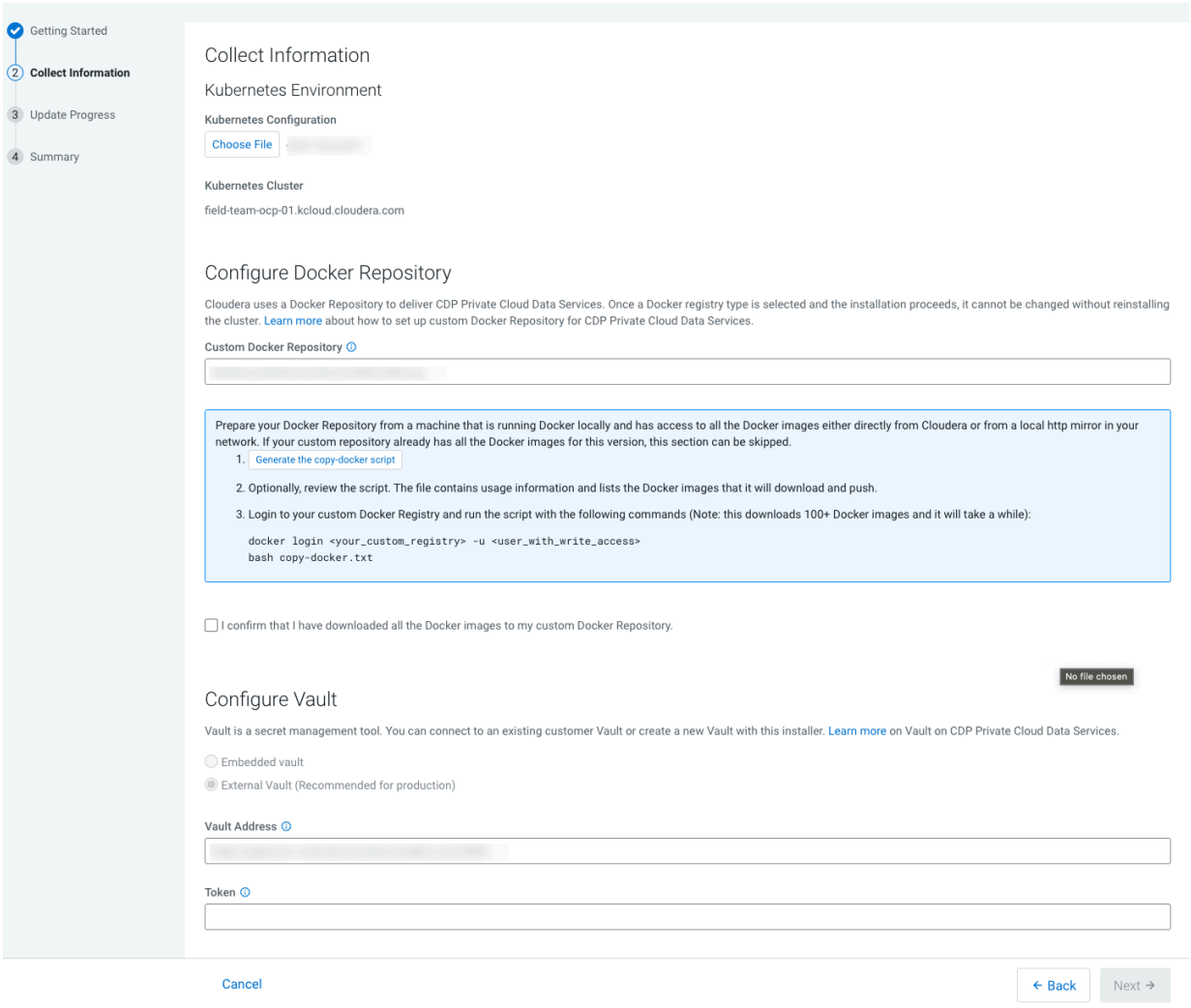


Air Gapped install method



Click Continue.

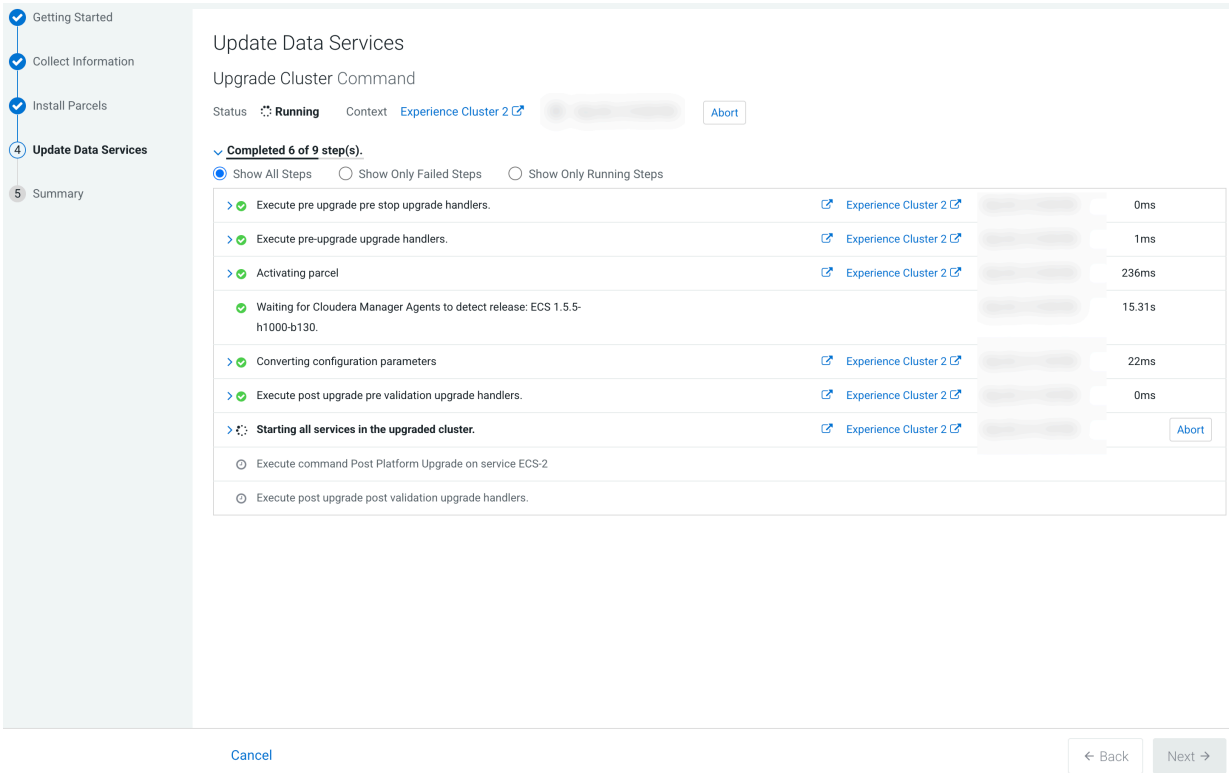
3. On the Collect Information page, click Next.



For the kubeconfig upload step (collect information) : After we upload the KUBECONFIG file:

- If you are using non default repository, confirm all the images are uploaded.
- If you are using External Vault, provide the details with respect to Vault Address and Token information with respect to the External Vault.

4. On the Update Data Services page, you can see the progress of your update. Click Next.



Important:

During the "Upgrade Cloudera Control Plane" step of the Clouderadat platform upgrade process, the grafana pod can get stuck in the terminating sate. This usually means that all other Cloudera 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]
```

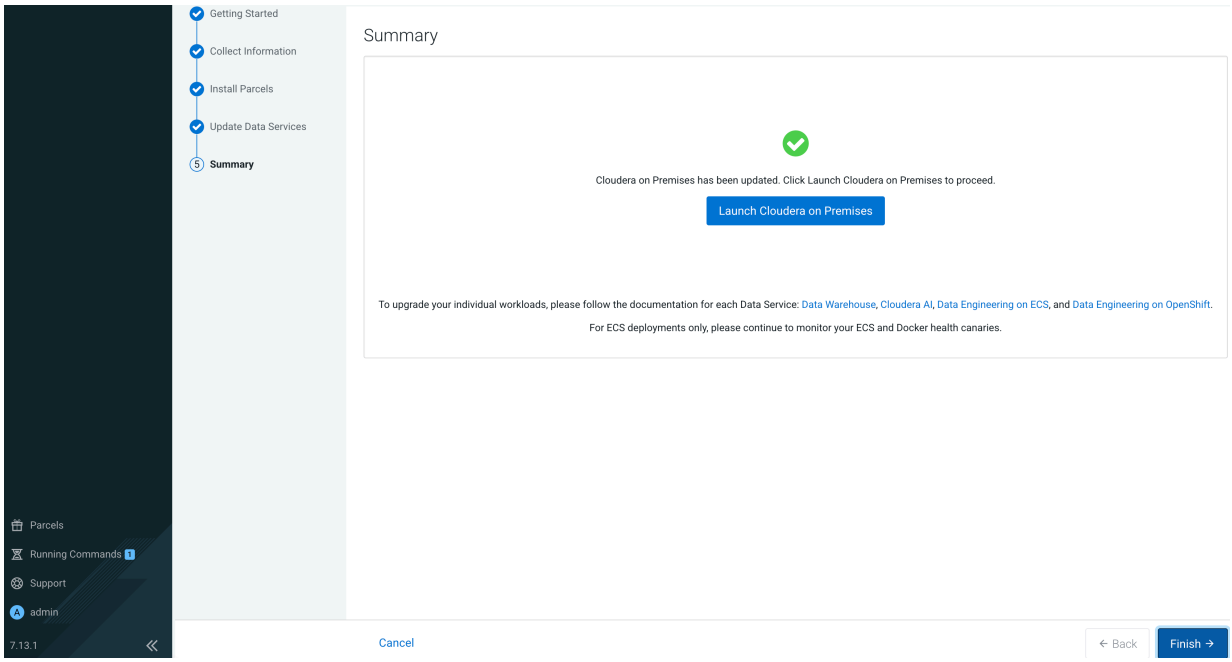
If you search for the terminating pod id in the kubelet log on the host, the following error message can be found:

```
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 Cloudera Embedded Container Service Server host:

```
<GRAFANA-POD-ID> --force --grace-period=0
```

5. After the update is complete, the Summary page appears. You can now Launch Cloudera on premises from here.



   Or you can navigate to the Cloudera Data Services on premises page and click Open Cloudera on premises.

   Cloudera Data Services on premises opens up in a new window.

6. After the update is complete, delete the old Cloudera Data Engineering 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.

Pre-upgrade - Preparing for Cloudera Data Services on premises update for Cloudera Data Engineering

To upgrade the Cloudera Data Engineering service, no jobs must be running or scheduled in any virtual cluster under that Cloudera Data Engineering service.

Procedure

1. Pause all Airflow jobs and scheduled Spark jobs.
2. Kill all the the running jobs in the Cloudera Data Engineering virtual clusters under all Cloudera Data Engineering services or wait for them to complete.

Related Information

Enabling, disabling, and pausing scheduled jobs

Pre-upgrade - Upgrading Cloudera Data Engineering service with endpoint stability

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

The Cloudera Data Engineering 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.

Important:

- After upgrading Cloudera Data Engineering from version 1.5.4 SP2 or earlier to 1.5.5, the Cloudera Data Engineering Services and Virtual Clusters that were created in the earlier version does not work in Cloudera Data Engineering 1.5.5. Cloudera recommends you not to use the old Cloudera Data Engineering Services and Virtual Clusters that were created before upgrade. Instead, create new Services and Virtual Clusters in Cloudera Data Engineering 1.5.5 and use them.
- After upgrading Cloudera Data Engineering, the upgraded Virtual Cluster retains the same base OS images that were used in the source Virtual Cluster before upgrade to ensure maximum compatibility particularly for jobs that depend on particular python and scala versions such as Spark jobs. For example, if an old Virtual Cluster is Redhat insecure based, then the new restored Virtual Cluster will also be Redhat insecure based only and if the old Virtual Cluster is security hardened based, then the new restored Virtual Cluster will also be security hardened based only. No automation path is supported from Redhat insecure to security hardened or vice versa.

Prerequisites for upgrading Cloudera Data Engineering Service with endpoint stability

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

Before you begin

Ensure that the host meets the following conditions:

- Docker must be running in the host.
- OCP Kubernetes APIs must be reachable from this host
- The Cloudera Control Plane must be reachable from this host.

Important: If Cloudera Data Engineering is installed and you plan to enable SAML authentication, ensure that the Cloudera Data Engineering service is upgraded using LDAP first. Once the upgrade is successful, you can proceed to enable SAML authentication.

Procedure

1. Download the dex-upgrade-utils docker image tarball. The file naming convention is dex-upgrade-utils-*[***VERSION-NUMBER***]-[***BUILD-NUMBER***]*.tar.gz.

   Important: To download the dex-upgrade-utils docker image tarball, see the following locations:

   - Cloudera Data Services on premises 1.5.5
   - Cloudera Data Services on premises 1.5.5 SP1

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-b48.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 Cloudera Data Engineering 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 as follows:
   a) Create the cde-upgrade-util.properties file and save it in the $BASE_WORK_DIR directory.
   b) Update the following information in the cde-upgrade-util.properties file:

```
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
format. 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 re
sume the jobs that were paused during the backup phase. The jobs will be
 resumed after you restore the CDE service.>
platform_type:OCP
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 i
s 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.>
```

   For example: The following options are the minimum recommended options that you must include in the cde-upgrade-util.properties file:

```
cdp_k8s_namespace=cdp
cdp_endpoint=https://console-cdp.apps.host-1.ecs-pvc1.kcloud.cloudera.
com
credential_file_path=/home/dex/.cdp/credentials
de_admin_user=cdpuser1
de_admin_password=VGVzdDEyMw==
tls_insecure=true
auto_unpause_jobs=true
platform_type=OCP
```

```
use_stored_user=false
```

Important:

- The cdp_k8s_namespace, cdp_endpoint, de_admin_user, and de_admin_password values must be updated based on your cluster.
- The de_admin_password password is the base64 encoded password of the de_admin_user. You can use echo -n <pwd> | base64 to encode it.
- You must always set the value of the credential_file_path property as /home/dex/.cdp/credentials and must 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 Cloudera Data Engineering service that is being restored. But, if you change the environment during restore, it leads 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 Cloudera Data Engineering service that is being migrated. This information is required if you are using a Cloudera 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 console, click the Data Engineering tile. The Cloudera Data Engineering 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 Cloudera Data Engineering service whose endpoint you want to migrate.
   d. Make a note of the Cloudera Data Engineering cluster ID.

Backing up Cloudera Data Engineering service using the docker image

You must run the docker image to take backup of a Cloudera Data Engineering service. It takes backup of all the active virtual clusters in that Cloudera Data Engineering service. You can take backup of only an active Cloudera Data Engineering 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 upgrading Cloudera Data Engineering Service with endpoint stability* 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 Cloudera Data Engineering 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 helps during debugging or 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/scr
ipts/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-CLUSTER-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 Cloudera Data Engineering service ID and -o is the backup output directory path in the container. The backup output directory value must always be $BACKUP_OUTPUT_DIR and must 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-b48
 prepare-for-upgrade -s cluster-c2dhkp22 -o $BACKUP_OUTPUT_DIR
```

To identify the *CDE-CLUSTER-ID*, do the following steps:

a. In the Cloudera console, click the Data Engineering tile. The Cloudera Data Engineering home page displays.
b. On the left navigation menu, click Administration.
c. In the Services column, click  for the Service for which you want to identify the Cluster ID.
d. Note the CLUSTER ID displayed on the page.

Results

You have now taken the Cloudera Data Engineering 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 Cloudera Data Engineering service.

What to do next

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

Important:

- During the restore operation, both old and the new Cloudera Data Engineering services use the same resources allocated to the existing Cloudera Data Engineering service. Hence, you must double the resource pool size using the Quota Management option. For example, if root.<env_pool>.cde.sales is the pool that is used for the old or existing Cloudera Data Engineering 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. Consider the following conditions and plan whether to modify the resource pool size or not:
  - If the Cloudera Data Engineering service uses the default resource pool, that is root.<env_pool>.cde, then do not change the resource pool size.
  - If the Cloudera Data Engineering service uses a custom resource pool (for example, root.<env_pool>.cde.primary.secondary), the resource pool size of the last level (that is., secondary level in the example) must be doubled using Quota Management option. The additional capacity required after doubling the last level's pool size is allocated from the levels above it, starting from the higher levels and progressing downward. In this example, when you double the secondary level (last level), the extra resource pool capacity required is initially added to the primary level pool. Then the newly added resource pool capacity is added to the secondary level pool, resulting in an overall doubling of the resource pool size of the last level.
  - The resource capacity at the Cloudera Data Engineering service and the Virtual Cluster level must not be changed. Modifying the pool size at the resource pool level is sufficient.
- In Cloudera Data Engineering 1.5.5 SP1 and higher versions, along with doubling the resource pool size, you must change the memory units of the pools to the corresponding binary memory unit, such as Ki, Mi, and Gi. For example, if the memory unit is G, change it to Gi.

Related Information

Upgrading CDP on OpenShift
Prerequisites for upgrading CLoudera Data Engineering Service with endpoint stability
Managing cluster resources using Quota Management

Scaling down Cloudera Data Engineering embedded database

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

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. Scale down Cloudera Data Engineering embedded database

   Access the OpenShift cluster with OpenShift CLI or Kubernetes CLI, and scale down the Cloudera Data Engineering 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
   ```

OCP upgrade steps for Cloudera Data Services on premises 1.5.5

Updating Cloudera Data Services on premises 1.5.3 or 1.5.4 to 1.5.5 on OCP requires you to upgrade earlier OCP versions to a version supported by Cloudera Data Services on premises 1.5.4. This process involves incremental OCP version upgrades in conjunction with Cloudera Data Services on premises updates and validations. Intermediate OCP versions are validated with the existing workload to work with older OCP while upgrade is being done to a newer OCP. However, it is recommended to stick to a support OCP version for a given data services release as provided on Cloudera Support Matrix .

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 Cloudera Data Services on premises 1.5.3 to 1.5.4

Starting point: Cloudera Data Services on premises 1.5.3 on OCP 4.12.

Cloudera Private Cloud Data Services

1. Update Cloudera Data Services on premises from 1.5.3 to 1.5.4.
2. Perform post-upgrade validations of existing resources and workloads.

   Important: If you are using Cloudera Data Services on premises, then post-upgrade validations of existing resources and workloads are not supported until you upgrade OCP version to 4.14 and upgrade Cloudera Data Services on premises workloads version to 1.5.4. For more information about upgrading Cloudera Data Services on premises workloads version to 1.5.4, see *Post-upgrade - Restoring Cloudera Data Services on premises service for endpoint stability*.
3. Upgrade OCP from 4.12 to 4.14.
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 Cloudera Private Cloud Data Services 1.5.4 to 1.5.5 SP1

Starting point: Cloudera Private Cloud Data Services 1.5.4 CHF1 or 1.5.4 CHF3 OCP 4.14 and 1.5.4 SP1or 1.5.4 SP2 on OCP 4.18.

1. Update Cloudera Private Cloud Data Services from 1.5.4 to 1.5.5.

2. Perform post-upgrade validations of existing resources and workloads.

   Important: If you are using Cloudera Data Services on premises, then post-upgrade validations of existing resources and workloads are not supported until you upgrade OCP version to 4.17 and upgrade Cloudera Data Engineering workloads version to 1.5.5. For more information about upgrading Cloudera Data Engineering workloads version to 1.5.5, see *Post-upgrade - Restoring Cloudera Data Engineering service for endpoint stability*.
3. Upgrade OCP from 4.14 or 4.16 to 4.18.
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.

Related Information

Post-upgrade - Restoring Cloudera Data Engineering service for endpoint stability

Upgrading Cloudera Data Services on premises using OpenShift Container Platform

You can upgrade Cloudera Data Services on premises 1.5.4 or 1.5.5 versions to 1.5.5 SP1 version without uninstalling the existing installation. 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 1.5.5 SP1.
- To upgrade Cloudera Data Engineering, make sure that you have performed Pre-upgrade - Upgrading Cloudera Data Engineering service with endpoint stability and taken a backup of the Cloudera Data Engineering service.

Important:

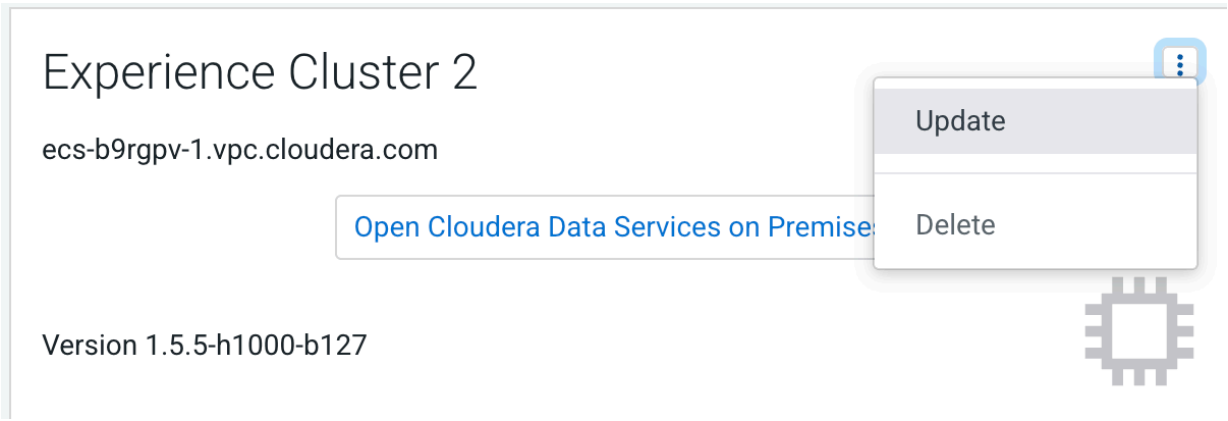
Ensure that the OpenShift namespace name is 29 characters or less. Do not proceed with the upgrade if the namespace name is 30 or more characters in length.

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 Cloudera Data Services on premises.
- One or more environments registered in Cloudera Data Services on premises.
- One of the registered environment has one or more Cloudera Data Warehouse or Cloudera AI experience workspaces.
- Access to the Cloudera on premises repositories (archive.cloudera.com).
- Administrator access to OCP and Privileged access to your external Vault.

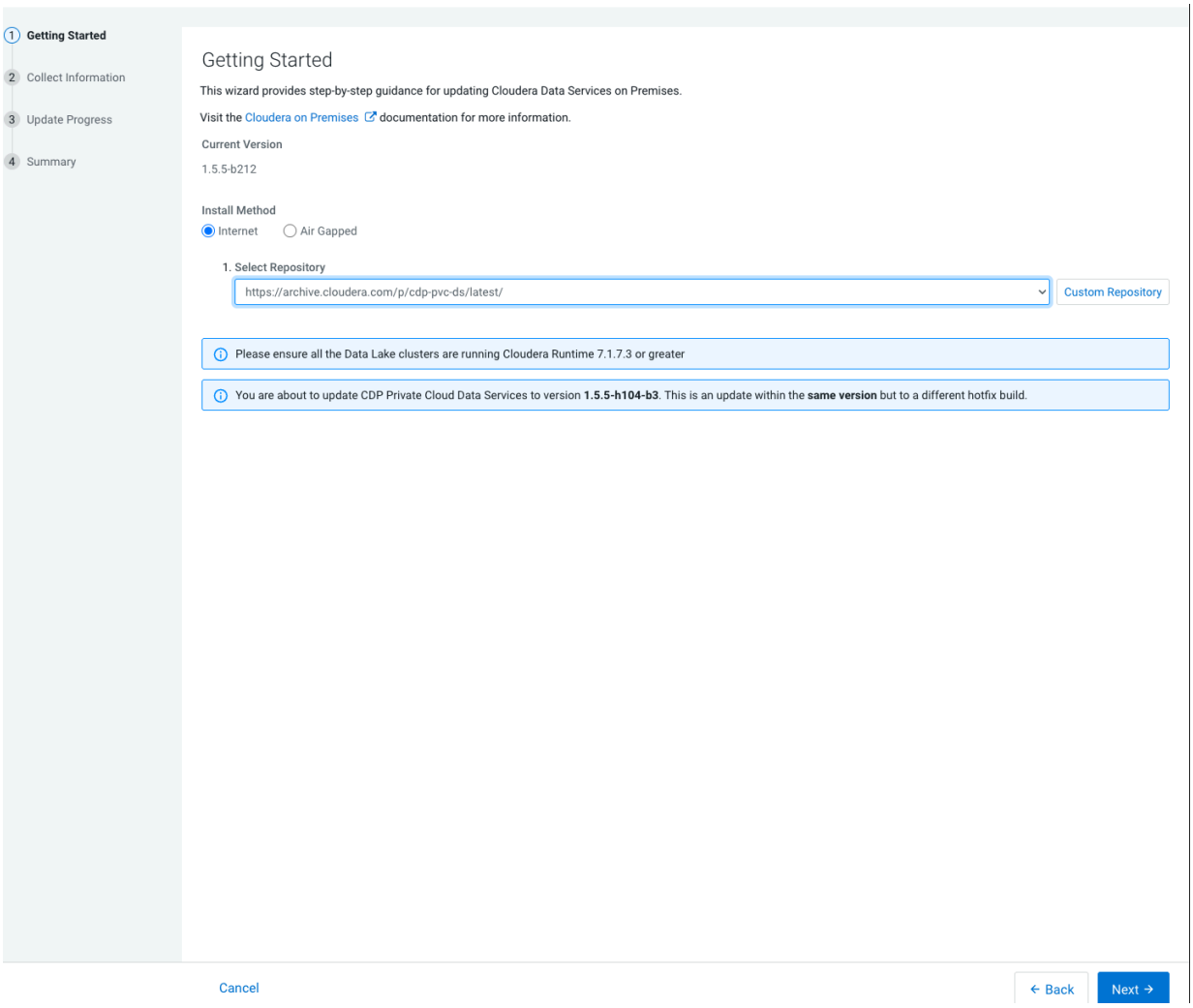
Procedure

1. In Cloudera Manager, navigate to Cloudera Data Services on premises and click . Click Update.

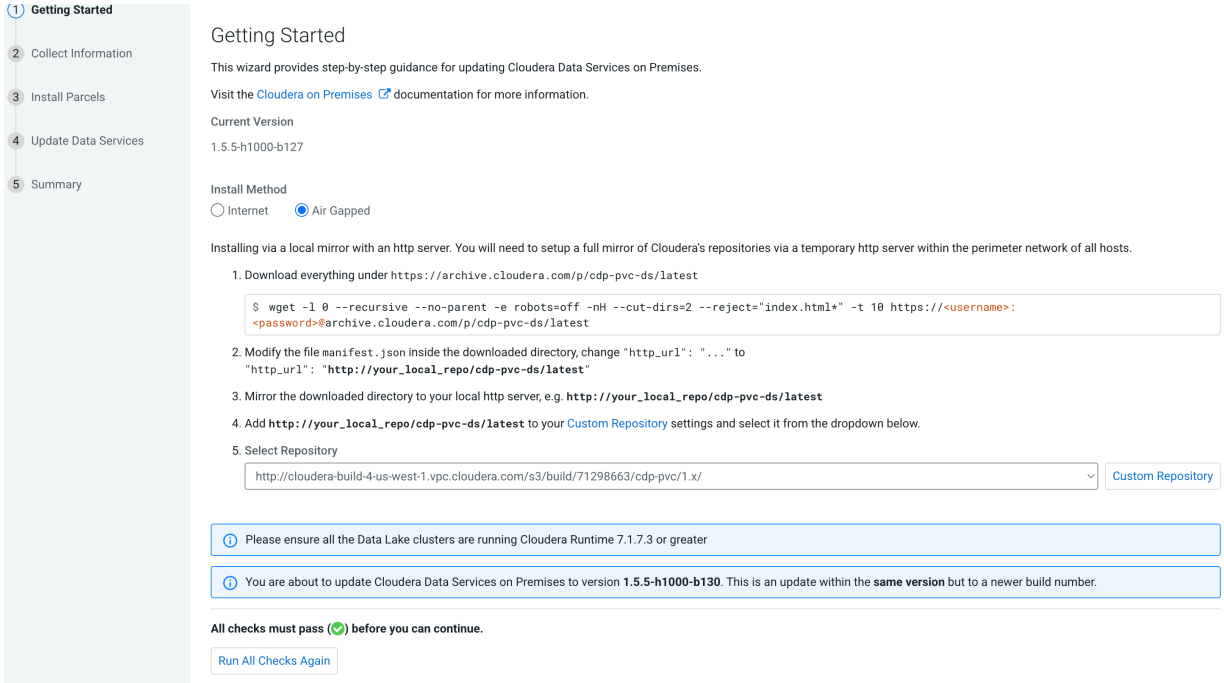


2. On the Getting Started page, you can select the Install method - Air Gapped or Internet and proceed.

Internet install method

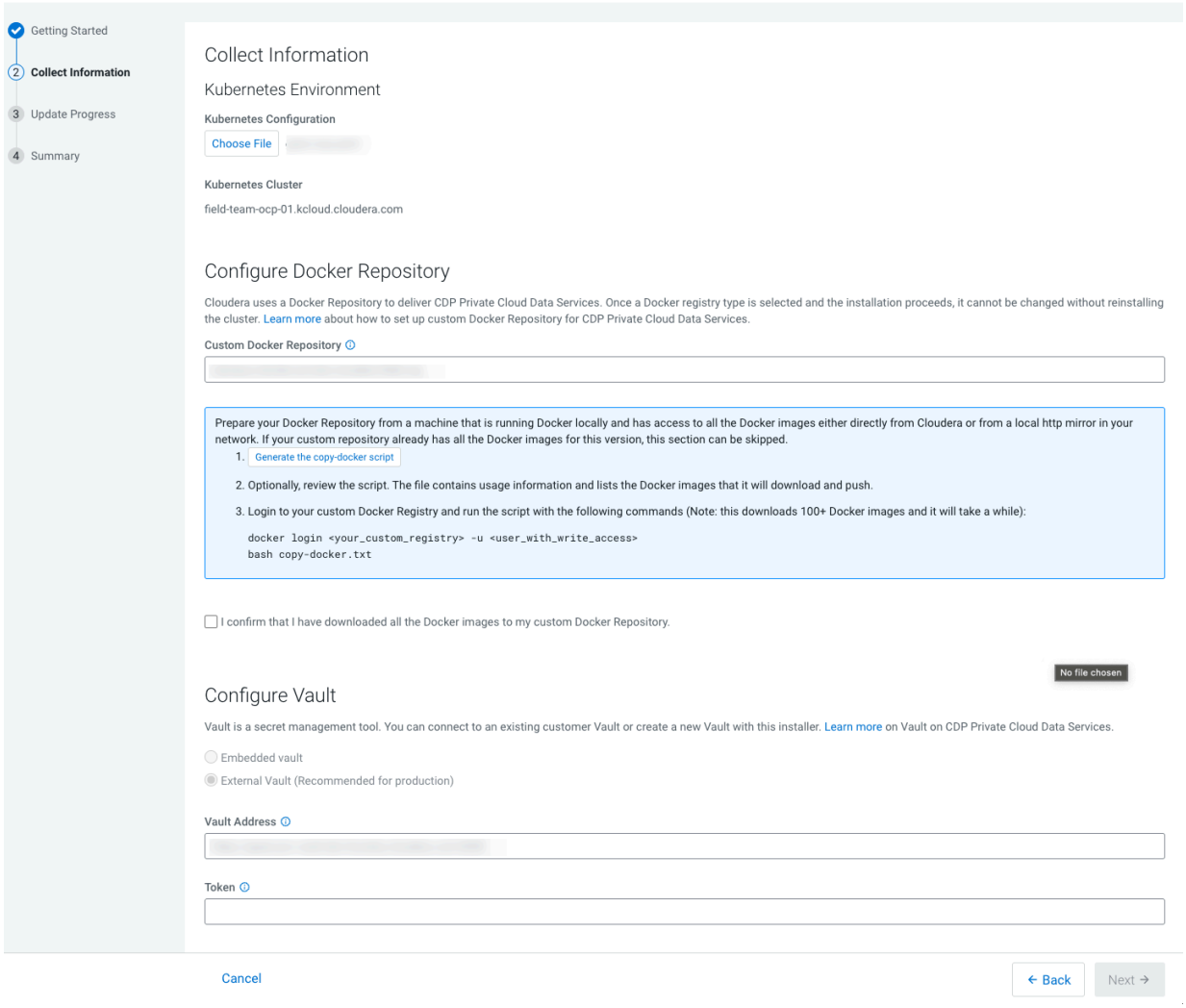


Air Gapped install method



Click Continue.

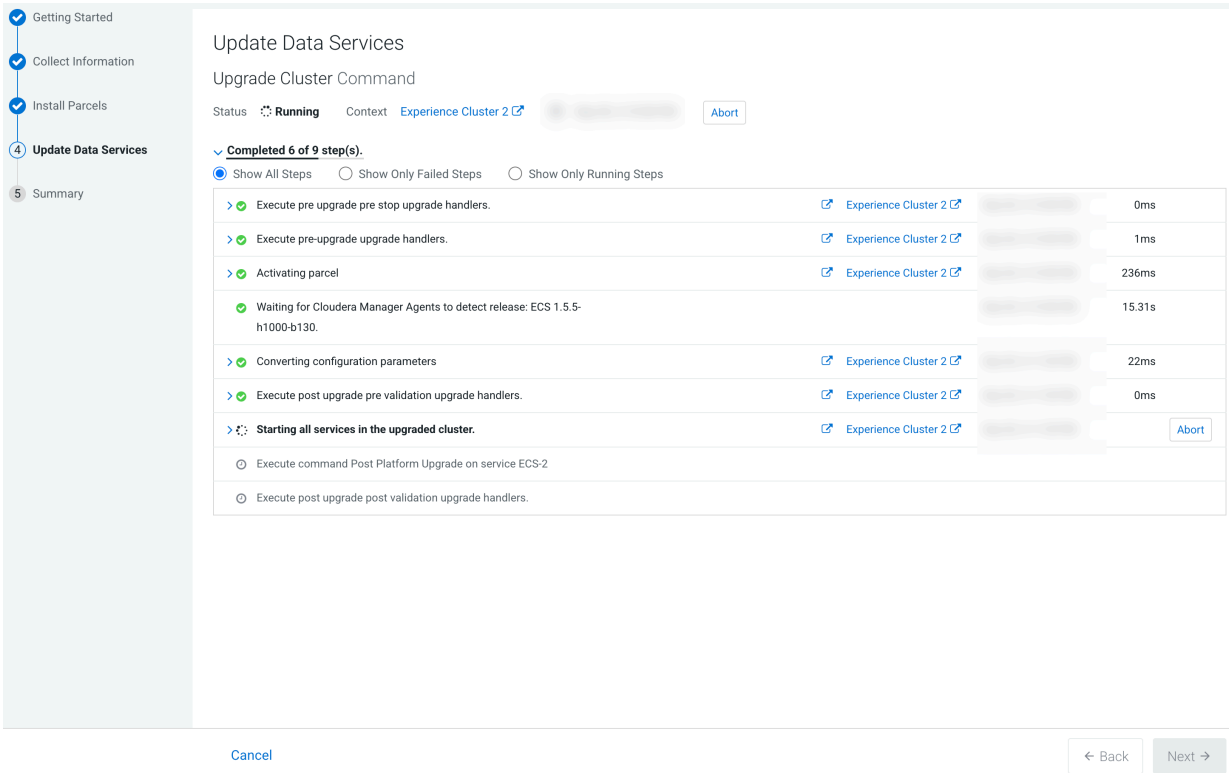
3. On the Collect Information page and click Next.



For the kubeconfig upload step (collect information) : After we upload the KUBECONFIG file:

- If you are using non default repository, confirm all the images are uploaded.
- If you are using External Vault, provide the details with respect to Vault Address and Token information with respect to the External Vault.

4. On the Update Data Services page, you can see the progress of your upgrade. Click Next.



Important:

During the "Upgrade Cloudera Control Plane" step of the Cloudera upgrade process, the grafana pod can get stuck in the terminating sate. This usually means that all other Cloudera 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]
```

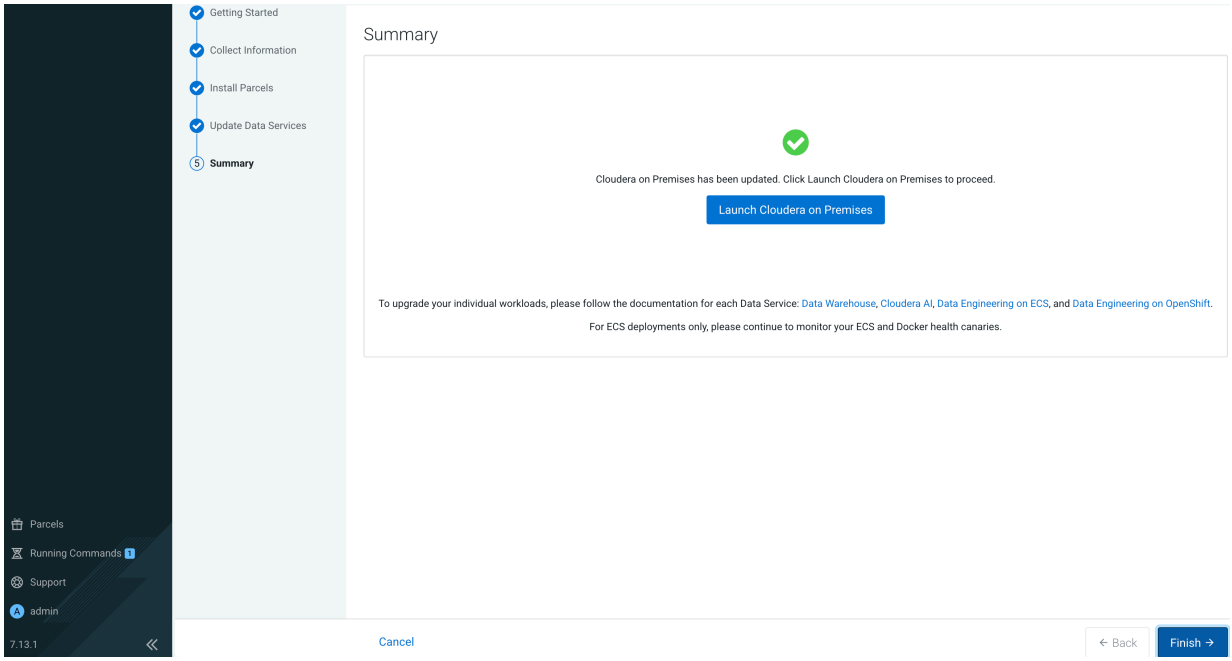
If you search for the terminating pod id in the kubelet log on the host, the following error message can be found:

```
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 Cloudera Embedded Container Service Server host:

```
<GRAFANA-POD-ID> --force --grace-period=0
```

5. After the upgrade is complete, the Summary page appears. You can now Launch Cloudera on premises from here.



   Or you can navigate to the Cloudera Data Services on premises page and click Open Cloudera on premises.

   Cloudera Data Services on premises 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 Cloudera Data Engineering, after you complete the OpenShift Container Platform (OCP) upgrade, ensure that the following step is performed before proceeding with *Post-upgrade - Restoring Cloudera Data Engineering Service for Endpoint Stability*.

Procedure

Scale back the Cloudera Data Engineering embedded database statefulset to 1.

OpenShift CLI

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

Kubernetes CLI

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

Related Information

Post-upgrade - Restoring Cloudera Data Engineering service for endpoint stability

Post upgrade - Ozone Gateway validation

If you are using Cloudera Data Engineering, after upgrading Cloudera Data Services on premises you must validate that the Ozone Gateway is working as expected. This procedure applies to both 1.5.3 and 1.5.4 to 1.5.5 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_enpoint> --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-endp
oint 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
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_enpoint> --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 Cloudera Data Engineering CLI, see Using the Cloudera Data Engineering command line interface

Recovering a corrupted Cloudera Data Engineering Embedded database

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

Procedure

1. Identifying the CDE Namespace
   a) Navigate to the Cloudera Data Engineering Overview page by clicking the Cloudera Data Engineering tile in the Cloudera Management Console .
   b) In the CDE Services column, click Service Details for the Cloudera Data Engineering 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 Cloudera Data Engineering 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 Cloudera Data Engineering 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 Cloudera Data Engineering 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 Cloudera Data Engineering Virtual Clusters to be accessible. This usually takes about 10 minutes.

Post-upgrade - Restoring Cloudera Data Engineering service for endpoint stability

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

Restoring a Cloudera Data Engineering service

You can restore the Cloudera Data Engineering 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 Cloudera Data Engineering service, expand the resource pool, and then upgrade your Cloudera to restore the Cloudera Data Engineering 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 helps you in debugging or raising a support ticket. You can also increase the terminal buffer size so that it does not throw away the logs 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 Cloudera Data Engineering 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***]
```

Where -s is the Cloudera Data Engineering service ID and -f is the backup output directory path 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/sc
ripts/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
estore-service -s cluster-c2dhkp22 -f $BACKUP_OUTPUT_DIR/cluster-c2dhkp2
2-2023-03-10T06_00_05.zip
```

3. If you are using a CDP database that is external and is not accessible from the container which is running the Cloudera Data Engineering 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-1-19-1' where id = 'cluster-c2dhkp22';
2023-05-17 13:02:29,551 [INFO]     update cluster set name = 'cde-base-
service' where id = 'cluster-92c2fkgb';
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 =
 '2025-05-05 06:16:37.786199' where id = 'cluster-c2dhkp22';
                -----------------------------------------------------
----------
```

   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 upgrading Cloudera Data Engineering Service with endpoint stability* section.

   a. Rename old Cloudera Data Engineering 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-1-19-1' where id = 'cluster-
c2dhkp22'
```

b. Rename the new Cloudera Data Engineering service to the old Cloudera Data Engineering service name.

```
update cluster set name = '[***OLD_CDE_SERVICE_NAME***]' where id =
 '[***NEW_CDE_CLUSTER_ID***]';
```

Example:

```
update cluster set name = 'cde-base-service' where id = 'cluster-92c2fkg
b'
```

c. Run the following command so that when the old Cloudera Data Engineering service is deleted or 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]***]' wh
ere id = '[***OLD_CDE_CLUSTER_ID***]';
```

Example:

```
update cluster set lastupdated = '2023-05-05 06:16:37.786199' where id =
 'cluster-c2dhkp22'
```

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 Cloudera Data Engineering UI, you can see the old Cloudera Data Engineering service is appended with a version number. For example, if the old Cloudera Data Engineering service name was cde-sales, after the restore, the old Cloudera Data Engineering service is something similar to cde-sales-1-19.1.

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

   Important: If you are upgrading from Cloudera Data Services on premises 1.5.4 or earlier, make sure that you update the Kerberos keytab file in Hadoop Authentication after upgrading to Cloudera Data Services on premises 1.5.5. For more information about how to update the keytab file, see Hadoop Authentication.

Related Information

Upgrading CDP on OpenShift
Upgrading CDP on OpenShift
Ozone Gateway validation

Rolling back the Cloudera Data Engineering service endpoint migration

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

About this task

Important: It is recommended to copy the logs of the commands run from the terminal and save them on your machine. This helps you during debugging or raising a support ticket. You can also increase the terminal buffer size so that it does not throw away the logs and save the terminal logs of each command for reference.

Before you begin

To rollback, the state of the Cloudera Data Engineering 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/scr
ipts/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-se
rvice -s [***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-b48
 rollback-restore-service -s cluster-c2dhkp25  -f $BACKUP_OUTPUT_DIR/clust
er-c2dhkp22-2023-03-10T06_00_05.zip
```

Limitations of Cloudera Data Engineering service endpoint migration

Learn about the limitations that you might run into while migrating your Cloudera Data Engineering 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, perform the following actions:
  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.
- Cloudera Data Engineering service does not support backup and restore of an Airflow Python environment resource after upgrade. You must create a new Airflow Python environment resource after restoring Cloudera Data Engineering service and then activate it manually.