CDP Private Cloud Data Services 1.5.3

Upgrading CDP Private Cloud Data Services on the Embedded Container Service

Date published: 2023-12-16
Date modified: 2024-03-23

CLOUDERA

https://docs.cloudera.com/

Legal Notice

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

Upgrading

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*

Related Information

Upgrading CDP on the Embedded Container Service

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

Managing cluster resources using Quota Management

Prerequisites for migrating CDE endpoints

Upgrading Cloudera Manager

You must use Cloudera Manager version 7.11.3 CHF 4 to install or upgrade to CDP Private Cloud Data Services 1.5.3.

If you already have a CDP Private Cloud Base cluster set up using an earlier version of Cloudera Manager, you must first upgrade the Cloudera Manager version to Cloudera Manager 7.11.3 CHF 4 before proceeding with the CDP Private Cloud Data Services update.

Related Information

Upgrading Cloudera Manager

Upgrade from 1.5.1 or 1.5.2 to 1.5.3 (ECS)

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.

Before you begin

- Review the Software Support Matrix for ECS.
- The Docker registry that is configured with the cluster must remain the same during the upgrade process. If CDP Private Cloud Data Services 1.5.1 or 1.5.2 was installed using the public Docker registry, CDP Private Cloud Data Services 1.5.3 should also use the public Docker registry, and not be configured to use the embedded Docker registry. If you would like to use a different configuration for the Docker registry, you must perform a new installation of CDP Private Cloud Data Services.

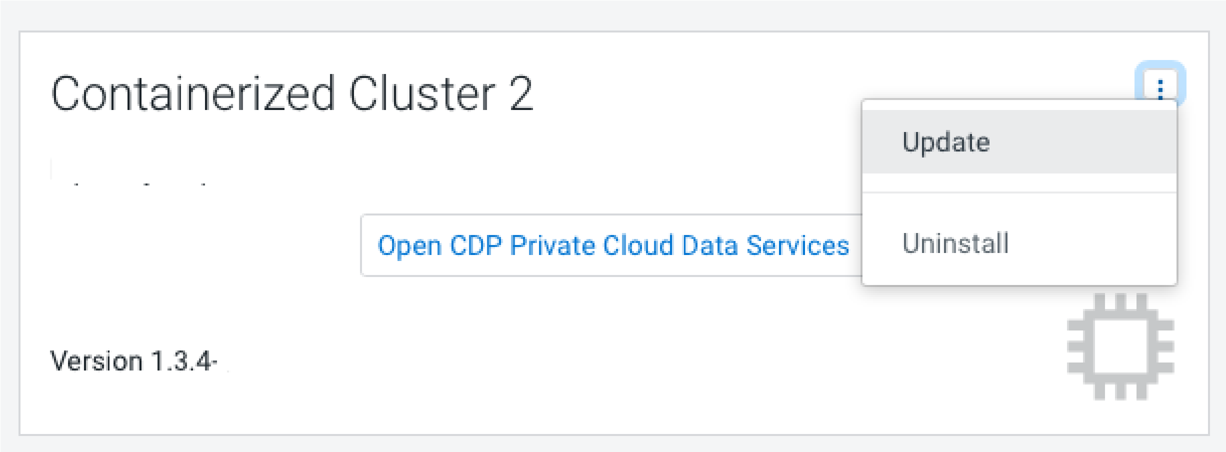
About this task

Note: ECS services will be unavailable to users for a period of time during this upgrade procedure. However, you should not stop the ECS cluster prior to upgrade. Upgrade requires the ECS cluster to be running and in a healthy state.

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.

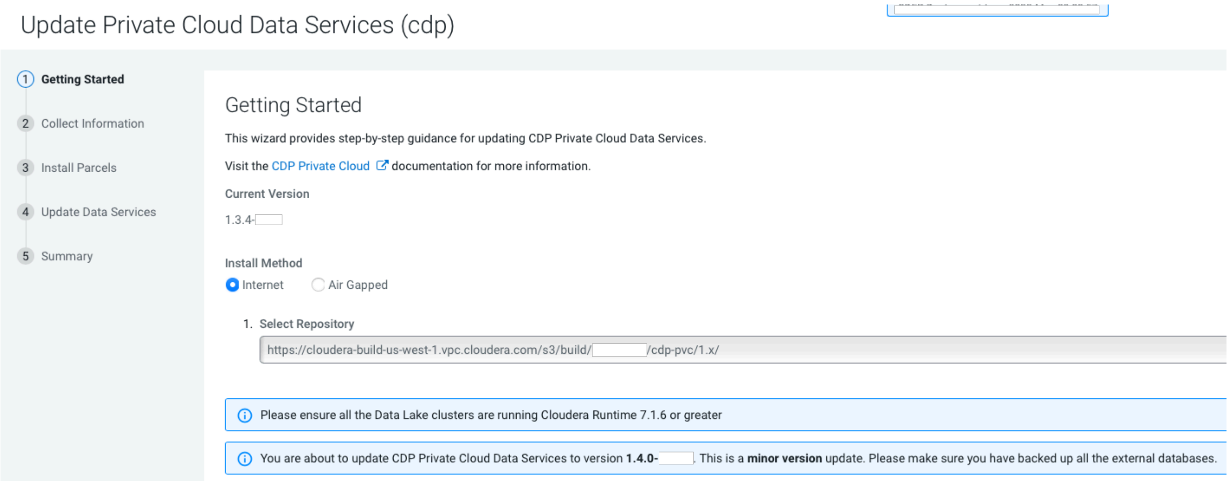
Procedure

1. In Cloudera Manager, navigate to CDP Private Cloud Data Services and click the ⋮ icon, then click Update.



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

   Internet install method



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

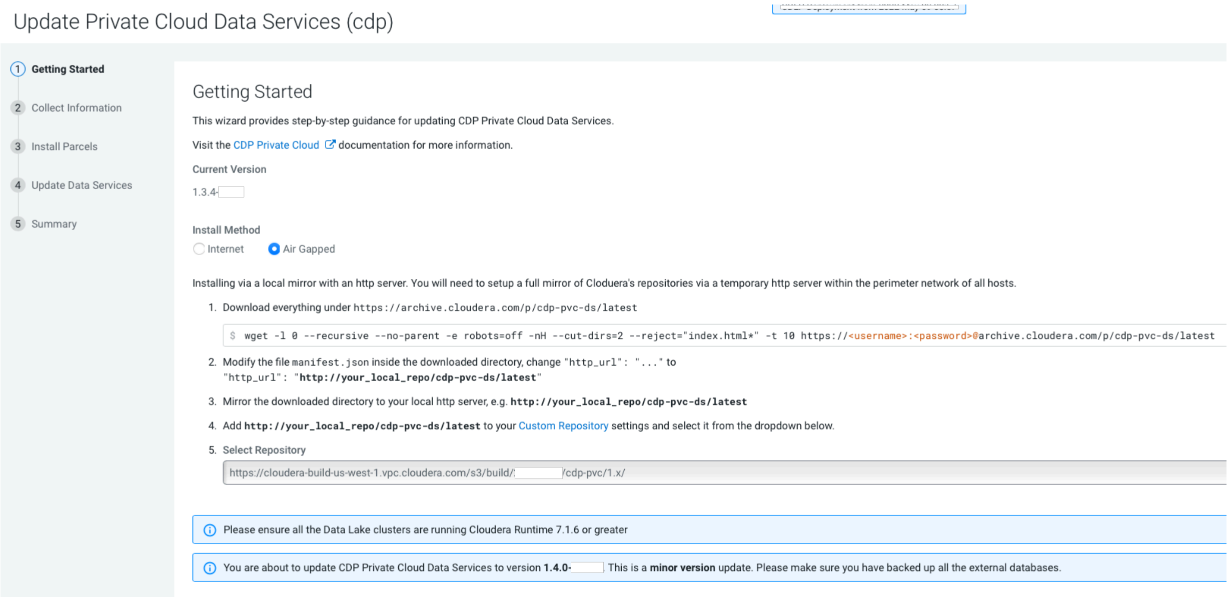
1. Select Repository

   https://cloudera-build-us-west-1.vpc.cloudera.com/s3/build/ /cdp-pvc/1.x/

ⓘ 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



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 Cloduera's repositories via a temporary http server within the perimeter network of all hosts.

1. 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
```

2. Modify the file `manifest.json` inside the downloaded directory, change `"http_url": "..."` to `"http_url": "`**`http://your_local_repo/cdp-pvc-ds/latest`**`"`
3. Mirror the downloaded directory to your local http server, e.g. **`http://your_local_repo/cdp-pvc-ds/latest`**
4. Add **`http://your_local_repo/cdp-pvc-ds/latest`** to your Custom Repository settings and select it from the dropdown below.
5. Select Repository

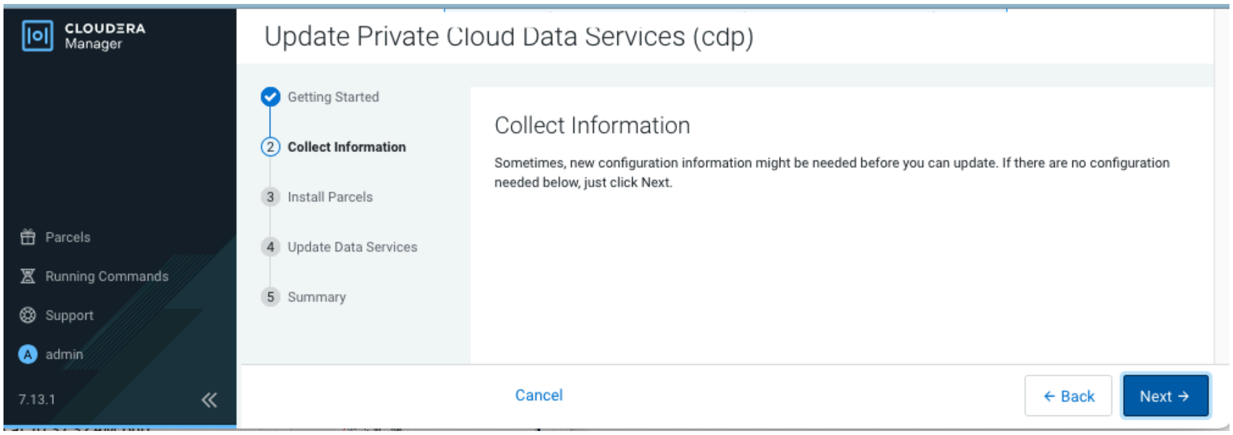
   https://cloudera-build-us-west-1.vpc.cloudera.com/s3/build/ /cdp-pvc/1.x/

ⓘ 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.

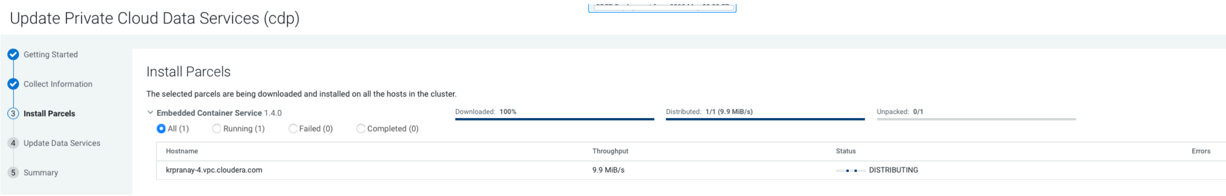
3. On the Collect Information page, click Next.



Collect Information

Sometimes, new configuration information might be needed before you can update. If there are no configuration needed below, just click Next.

4. On the Install Parcels page, click Continue.



Update Private Cloud Data Services (cdp)

Install Parcels

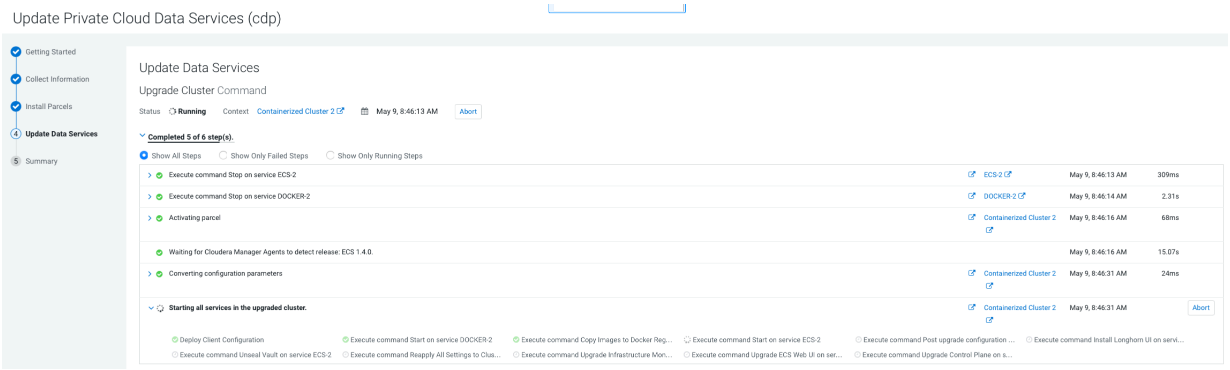
The selected parcels are being downloaded and installed on all the hosts in the cluster.

Embedded Container Service 1.4.0 Downloaded: 100% Distributed: 1/1 (9.9 MiB/s) Unpacked: 0/1

● All (1) ○ Running (1) ○ Failed (0) ○ Completed (0)

Hostname	Throughput	Status	Errors
krpranay-4.vpc.cloudera.com	9.9 MiB/s	DISTRIBUTING	

5. On the Update Progress page, you can see the progress of your upgrade. Click Continue after the upgrade is complete .



Update Private Cloud Data Services (cdp)

Update Data Services

Upgrade Cluster Command

Status Running Context Containerized Cluster 2 May 9, 8:46:13 AM Abort

Completed 5 of 6 step(s).

● Show All Steps ○ Show Only Failed Steps ○ Show Only Running Steps

✓ Execute command Stop on service ECS-2	ECS-2	May 9, 8:46:13 AM	309ms
✓ Execute command Stop on service DOCKER-2	DOCKER-2	May 9, 8:46:14 AM	2.31s
✓ Activating parcel	Containerized Cluster 2	May 9, 8:46:16 AM	68ms
✓ Waiting for Cloudera Manager Agents to detect release: ECS 1.4.0.		May 9, 8:46:16 AM	15.07s
✓ Converting configuration parameters	Containerized Cluster 2	May 9, 8:46:31 AM	24ms
Starting all services in the upgraded cluster.	Containerized Cluster 2	May 9, 8:46:31 AM	Abort

✓ Deploy Client Configuration
✓ Execute command Start on service DOCKER-2
✓ Execute command Copy Images to Docker Reg...
Execute command Start on service ECS-2
Execute command Post upgrade configuration ...
Execute command Install Longhorn UI on servi...
Execute command Unseal Vault on service ECS-2
Execute command Reapply All Settings to Clus...
Execute command Upgrade Infrastructure Mon...
Execute command Upgrade ECS Web UI on ser...
Execute command Upgrade Control Plane on s...

Note: The upgrade might occasionally fail with error messages or conditions such as the following:

- Error message: During the following step: Execute command Install Tolerations Webhook on service ECS-3 the Upgrade progress page mentions a failure waiting for kube-proxy to come up.

  Workaround:

  a. Log in using ssh to one of the ECS Server nodes and run the following command:

  ```
  /var/lib/rancher/rke2/bin/kubectl get nodes
  ```

  The output will look similar to the following:

  ```
  NAME                       STATUS     ROLES
  AGE     VERSION
  ecs-abc-1.vpc.myco.com    Ready      control-plane,etcd,master
  4h50m   v1.21.8+rke2r2
  ecs-abc-2.vpc.myco.com    NotReady   <none>
   4h48m   v1.20.8+rke2r1
  ecs-abc-3.vpc.myco.com    Ready      <none>
   4h48m   v1.21.8+rke2r2
  ecs-abc-4.vpc.myco.com    NotReady   <none>
   4h48m   v1.20.8+rke2r1
  ecs-abc-5.vpc.myco.com    NotReady   <none>
  4h48m   v1.20.8+rke2r1
  ```

  If any of the version numbers in the last column are lower than the expected version, reboot those nodes. (For example, v1.20.8 in the output above.)

  b. In the Command Output window, in the step that failed, click Resume.

- Upgrade hangs on the Execute command Post upgrade configuration on service ECS step for more than an hour.

  Workaround:

  a. Log in to one of the ECS server nodes and run the following command:

  ```
  kubectl get nodes
  ```

  The output looks similar to the following:

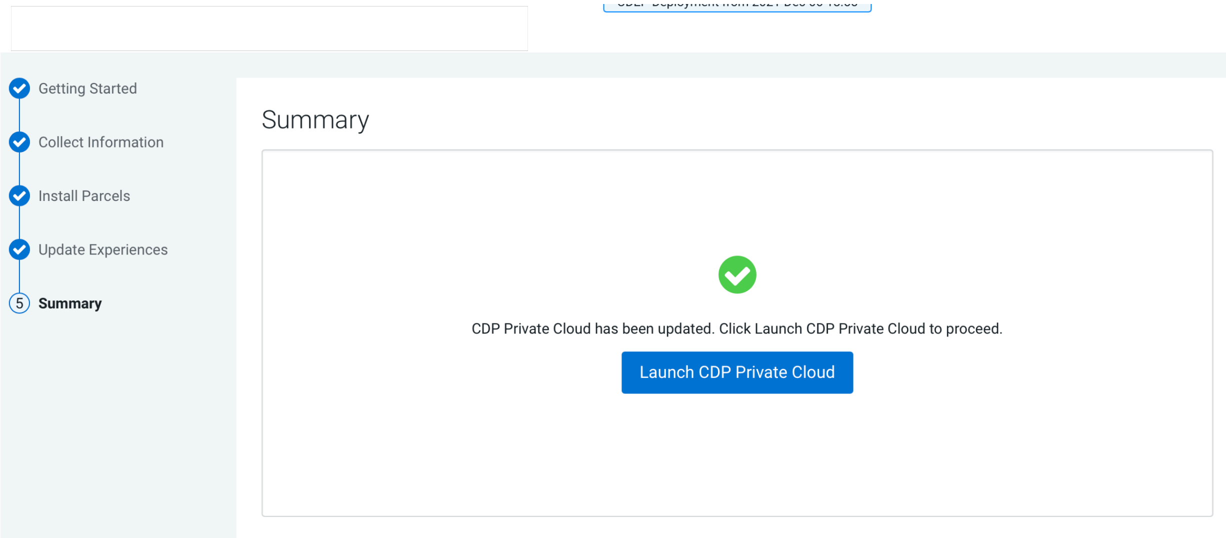
  ```
  NAME                       STATUS     ROLES
  AGE     VERSION
  ecs-abc-1.vpc.myco.com    Ready      control-plane,etcd,master
  3h47m   v1.21.11+rke2r1
  ecs-abc-2.vpc.myco.com    NotReady   <none>
  3h45m   v1.21.8+rke2r2
  ecs-abc-3.vpc.myco.com    NotReady   <none>
   3h45m   v1.21.8+rke2r2
  ecs-abc-4.vpc.myco.com    NotReady   <none>
   3h45m   v1.21.8+rke2r2
  ```

  If you any nodes display a status of NotReady, click the Abort button in the command output window.

  b. Reboot all nodes showing NotReady.

  c. Check the node status again as shown above. After all the nodes show Ready, click the Resume button in the command output window to continue with the upgrade.

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



Summary

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

If you see a Longhorn Health Test message about a degraded Longhorn volume, wait for the cluster repair to complete.

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.

- If the upgrade stalls, do the following:

1. Check the status of all pods by running the following command on the ECS server node:

```
kubectl get pods --all-namespaces
```

2. If there are any pods stuck in “Terminating” state, then force terminate the pod using the following command:

```
kubectl delete pods <NAME OF THE POD> -n <NAMESPACE> --grace-period=0 —f
orce
```

If the upgrade still does not resume, continue with the remaining steps.

3. In the Cloudera Manager Admin Console, go to the ECS service and click Web UIStorage UI.

The Longhorn dashboard opens.

4. Check the "in Progress" section of the dashboard to see whether there are any volumes stuck in the attaching/detaching state in. If a volume is that state, reboot its host.

- You may see the following error message during the Upgrade Cluster > Reapplying all settings > kubectl-patch :

```
kubectl rollout status deployment/rke2-ingress-nginx-controller -n kube-
system --timeout=5m
error: timed out waiting for the condition
```

If you see this error, do the following:

1. Check whether all the Kubernetes nodes are ready for scheduling. Run the following command from the ECS Server node:

```
kubectl get nodes
```

You will see output similar to the following:

```
NAME STATUS ROLES AGE VERSION
```

```
<node1> Ready,SchedulingDisabled control-plane,etcd,master 103m v1.21.
11+rke2r1
<node2> Ready <none> 101m v1.21.11+rke2r1
<node3> Ready <none> 101m v1.21.11+rke2r1
<node4> Ready <none> 101m v1.21.11+rke2r1
```

2. Run the following command from the ECS Server node for the node showing a status of SchedulingDisabled:

   ```
   kubectl uncordon <node1>
   ```

   You must add the `NODENAME` to the end of the command.

   You will see output similar to the following:

   ```
   <node1>node/<node1> uncordoned
   ```

3. Scale down and scale up the rke2-ingress-nginx-controller pod by running the following command on the ECS Server node:

   ```
   kubectl delete pod  rke2-ingress-nginx-controller-<pod number> -n kube-s
   ystem
   ```

4. Resume the upgrade.

What to do next

- After upgrading, 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"
  ```

- If you specified a custom certificate, select the ECS cluster in Cloudera Manager, then select Actions > Update Ingress Controller. This command copies the cert.pem and key.pem files from the Cloudera Manager server host to the ECS Management Console host.
- After upgrading, you can enable the unified time zone feature to synchronize the ECS cluster time zone with the Cloudera Manager Base time zone. When upgrading from earlier versions of CDP Private Cloud Data Services to 1.5.2+, unified time zone is disabled by default to avoid affecting timestamp-sensitive logic. For more information, see ECS unified time zone.

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_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 CDE CLI, see Using the Cloudera Data Engineering command line interface

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/script
s/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/scri
pts/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-c2dhk
p22-2023-03-10T06_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-92c2f
   kgb'
   ```

   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 the Embedded Container Service

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.