Cloudera on Premises Data Services 1.5.4

Upgrading Cloudera Data Services on premises on the Cloudera Embedded Container Service

Date published: 2023-12-16
Date modified: 2024-12-20

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

Upgrading

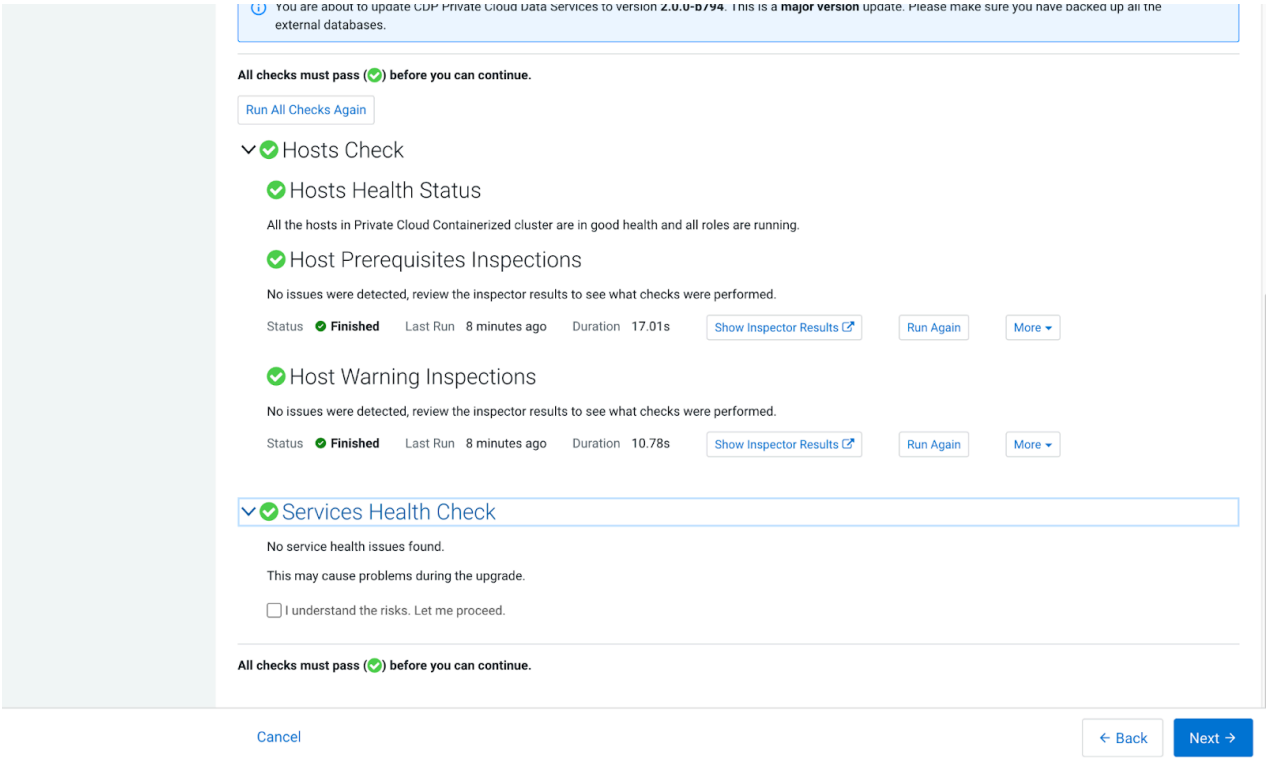
Pre-upgrade checklist

There is a list of pre-upgrade checks that will run after the upgrade version has been chosen. This checklist verifies if your cluster is ready for upgrade.

About this task

Note: Pre-upgrade checklist is added in the Upgrade wizard of the Cloudera Manager UI in the Cloudera Data Services on premises 1.5.4 SP2 release.

In the Cloudera Manager UI, under Getting Started for Upgrade Step in the Upgrade wizard, select the repository URL for upgrade. A pre-upgrade checklist will appear that verifies that the hosts and services in your cluster are ready for the upgrade.



Here is the pre-upgrade checklist:

Checklist	Description
Hosts check	This verifies the host health status, runs the host prerequisite inspections, and host warning inspections. • Host Health Status- This check verifies that there are no hosts in bad health or concerning health. It also checks for any stopped roles on the hosts. • Host Prerequisites Inspections - These are host inspections that must pass in order for you to proceed

Checklist	Description
	to upgrade. Currently the prerequisite inspection includes: • EcsHostDnsInspection - Checks to make sure that there are less than three nameserver entries in the /etc/resolv.conf file, and checks the connections to the Cloudera Manager cluster and the CDP console. It also checks to see if vault.localhost.localdomain's ping can be resolved. If not, it is likely that the host /etc/nswitch.conf file is misconfigured. If this inspection fails: • Check the /etc/resolv.conf and /etc/nswitch.conf files and ensure that /etc/resolv.conf does not contain three or more nameservers, and that /etc/nswitch.conf must contain myhostname under the hosts field. • Check to see the connections are resolved correctly. If the connection to the CDP console fails, check to see if your DNS wildcard is configured properly. • Host Warning Inspections - These are host inspections that are used to detect potential factors that can cause issues during an upgrade. Currently the warning inspections include: • SecuritySoftwareInspection - Checks to make sure that ther are no security software processes running on the hosts in the cluster.
Services Health Check	This verifies that there are no services in bad or concerning health.

Note: Fix any pre-check issues and run the check again to continue the upgrade.

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.

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 the Cloudera Data Engineering service.

About this task

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. Login into the ECS Server host using SSH and create an external kubeconfig file. The following command assumes that your home directory, that is, ~/ is the working directory.

```
sed -e 's/certificate-authority-data/#&/' -e "s/server: .*/server: https
:\/\/`hostname`:6443/" -e '/server/a \ \ \ \ insecure-skip-tls-verify: t
rue' /etc/rancher/rke2/rke2.yaml > ~/kubeconfig && cat ~/kubeconfig
```

   This command creates a file named kubeconfig in the working directory which is the external kubeconfig file.
2. Copy the CDP Credentials file named credentials of the DEAdmin user into the ECS Server host's working directory as follows:
   a) In the Cloudera Data Platform (CDP) console, click the Management Console tile.
   b) Click User Management and select the user.
   c) Click Generate Access Key Download credentials file .
   d) Copy the CDP Credentials file into the ECS Server host with the name credentials.
   e) Verify if the credentials are present in the ECS Server host:

```
ls -l credentials
```

3. Set the environment variables in the ECS Server host by running the following command:

```
export PATH=$PATH:/opt/cloudera/parcels/ECS/installer/install/bin/linux/:/
opt/cloudera/parcels/ECS/docker export KUBECONFIG=~/kubeconfig
```

4. Download the dex-upgrade-utils docker image tarball. The file naming convention is dex-upgrade-utils-*[***VERSION-NUMBER***]-[***BUILD-NUMBER***]*.tar.gz.
5. Load the downloaded docker 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
```

   Sample output:

   368243204766.dkr.ecr.us-west-2.amazonaws.com/cdp-private/cloudera/dex/dex-upgrade-utils:1.20.1-b48

   Important: The version of the utility must be same as the version of the CDE control plane that you are upgrading to.
6. Create the required folders on the ECS Server host and copy the credentials and kubeconfig secret files.

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

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

7. Place the CDP credentials file of the *DEADMIN* user and *ADMINISTRATOR* kubeconfig file in the $BASE_WORK_DIR/secrets directory.

```
cp ~/credentials secrets/
cp ~/kubeconfig secrets/
```

8. 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:ECS
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=ECS
user_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 CDE 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.

9. 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 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 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 CDE cluster ID.

Related Information

Upgrading CDP on the Cloudera Embedded Container Service
Managing cluster resources using Quota Management

Backing up Cloudera Data Engineering service using the docker image

You must run the docker image to take a 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 one active Cloudera Data Engineering service at a time.

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 that you copy the logs of the commands that are 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.

Procedure

1. Set the following environment variables in the ECS Server host terminal:

```
export PATH=$PATH:/opt/cloudera/parcels/ECS/installer/install/bin/linux/:/
opt/cloudera/parcels/ECS/docker
export KUBECONFIG=~/kubeconfig
export BASE_WORK_DIR=/opt/backup-restore
export BACKUP_OUTPUT_DIR=/home/dex/backup
```

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

```
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-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 which is being backed up and -o is the backup output directory path 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/s
cripts/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
prepare-for-upgrade -s cluster-c2dhkp22 -o $BACKUP_OUTPUT_DIR
```

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 Data Platform (CDP) 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.default.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.default, then do not change the resource pool size.
- If the Cloudera Data Engineering service uses a custom resource pool (for example, root.default.primary.secondary), the resource pool size of the last level (that is, secondary level in the example) must be doubled using the 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.

Related Information

Managing cluster resources using Quota Management

Prerequisites for upgrading Cloudera Data Engineering Service with endpoint stability

Cloudera Data Services on premises supported in-place upgrade paths

Supported upgrade paths for Cloudera Data Services on premises.

The tables below detail the supported upgrade paths for upgrades to Cloudera Data Services on premises. Before upgrading, ensure that you select a compatible version of Cloudera Manager. See Cloudera Manager support for CDH, Cloudera Runtime and CDP Private Cloud Data Services.

For CDP Private Cloud Data Services

For supported versions of Operating system, Database, JDK versions, see https://supportmatrix.cloudera.com/.

Table 1: Upgrade paths for Cloudera Data Services on premises 1.5.4

Source version			Target version	
Cloudera Manager	**Cloudera Base on premises**	**Cloudera Data Services on premises**	**Cloudera Manager**	**Cloudera Base on premises**
Greenfield deployment of Cloudera Data Services on premises 1.5.4				
NA	NA	NA	7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6
NA	NA	NA	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher
NA	NA	NA	7.11.3 Cumulative hotfix 6	7.1.7 SP3
Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.3				
7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 6	NA	7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6
7.7.3 Latest cumulative hotfix	7.1.8 Cumulative hotfix 22 or higher	NA	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher
7.7.1 Latest cumulative hotfix	7.1.8 Cumulative hotfix 22 or higher	NA	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher
7.11.3 Cumulative hotfix 5	7.1.7 SP3	NA	7.11.3 Cumulative hotfix 6	7.1.7 SP3
Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premi				
7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 6	1.5.3	7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6
7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 22 or higher	1.5.3	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher
7.11.3 Cumulative hotfix 4	7.1.7 SP3	1.5.3	7.11.3 Cumulative hotfix 6	7.1.7 SP3
7.11.3 Cumulative hotfix 1	7.1.8 Cumulative hotfix 22 or higher	1.5.2	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher

Table 2: Upgrade paths for Cloudera Data Services on premises 1.5.3

Source version			Target version	
Cloudera Manager	**Cloudera Base on premises**	**Cloudera Data Services on premises**	**Cloudera Manager**	**Cloudera Base on premises**
Greenfield deployment of Cloudera Data Services on premises 1.5.3				
NA	NA	NA	7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 3
NA	NA	NA	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher
NA	NA	NA	7.11.3 Cumulative hotfix 4	7.1.7 SP2
Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.3				
7.11.3 Cumulative hotfix 1	7.1.9 Cumulative hotfix 3	NA	7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 3
7.7.3 Latest cumulative hotfix	7.1.8 Cumulative hotfix 19 or higher	NA	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher
7.7.1 Latest cumulative hotfix	7.1.8 Cumulative hotfix 19 or higher	NA	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
7.6.7 Latest cumulative hotfix	7.1.7 SP2	NA	7.11.3 Cumulative hotfix 4	7.1.7 SP2
Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premi				
7.11.3 Cumulative hotfix 1	7.1.9 Cumulative hotfix 3	1.5.2	7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 3
7.11.3 Cumulative hotfix 1	7.1.8 Cumulative hotfix 19 or higher	1.5.2	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher
7.11.3 Cumulative hotfix 1	7.1.7 SP2	1.5.2	7.11.3 Cumulative hotfix 4	7.1.7 SP2
7.10.1	7.1.8 Cumulative hotfix 19 or higher	1.5.1	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher
7.10.1	7.1.7 SP2	1.5.1	7.11.3 Cumulative hotfix 4	7.1.7 SP2

Table 3: Upgrade paths for Cloudera Data Services on premises 1.5.2

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
Greenfield deployment of Cloudera Data Services on premises 1.5.2				
NA	NA	NA	7.11.3 Cumulative hotfix 1	7.1.9
NA	NA	NA	7.11.3 Cumulative hotfix 1	7.1.9 Cumulative hotfix 1
NA	NA	NA	7.11.3 Cumulative hotfix 1	7.1.8 CHF11 or higher
NA	NA	NA	7.11.3 Cumulative hotfix 1	7.1.7 SP2
Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.2				
7.11.3	7.1.9	NA	7.11.3 Cumulative hotfix 1	7.1.9
7.11.3	7.1.9 Cumulative hotfix 1	NA	7.11.3 Cumulative hotfix 1	7.1.9 Cumulative hotfix 1
7.7.3 Latest cumulative hotfix	7.1.8 Cumulative hotfix 11 or higher	NA	7.11.3 Cumulative hotfix 1	7.1.8 Cumulative hotfix 11 or higher
7.7.1 Latest cumulative hotfix	7.1.8 Cumulative hotfix 11 or higher	NA	7.11.3 Cumulative hotfix 1	7.1.8 Cumulative hotfix 11 or higher
7.6.7 Latest cumulative hotfix	7.1.7 SP2	NA	7.11.3 Cumulative hotfix 1	7.1.7 SP2
Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premi				
7.10.1	7.1.7 SP2	1.5.1	7.11.3 Cumulative hotfix 1	7.1.7 SP2
7.10.1	7.1.8 Cumulative hotfix 11 or higher	1.5.1	7.11.3 Cumulative hotfix 1	7.1.8 Cumulative hotfix 11 or higher
7.9.5	7.1.8 Cumulative hotfix 11 or higher	1.5.0	7.11.3 Cumulative hotfix 1	7.1.8 Cumulative hotfix 11 or higher
7.9.5	7.1.8	1.5.0	7.11.3 Cumulative hotfix 1	7.1.8
7.9.5	7.1.7 SP2	1.5.0	7.11.3 Cumulative hotfix 1	7.1.7 SP2

Important: Upgrading from Cloudera Manager 7.7.3 version to Cloudera Manager 7.10.1 is currently not supported. Note that, Cloudera Manager 7.7.3 version is supported on Python 3 and Cloudera Manager 7.10.1 supports only Python 2. Cloudera Data Services on premises 1.5.1 support for users using Cloudera Manager 7.7.3 is intended to be made available in the future with a new version of Cloudera Manager that has support for both Python 3 and 2 versions respectively.

Table 4: Upgrade paths for Cloudera Data Services on premises 1.5.1

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
Greenfield deployment of Cloudera Data Services on premises 1.5.1				
NA	NA	NA	7.10.1	7.1.7 SP2
NA	NA	NA	7.10.1	7.1.8 CHF4
Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.1				
7.6.7	7.1.7 SP2	NA	7.10.1	7.1.8 CHF4 [1]
7.7.1	7.1.8	NA	7.10.1	7.1.8 CHF4 [2]
7.6.7	7.1.7 SP2	NA	7.10.1	7.1.8
7.7.1	7.1.8	NA	7.10.1	7.1.8
Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premi				
7.8.1	7.1.7 SP1	1.4.1	7.10.1	7.1.7 SP1
7.8.1	7.1.8	1.4.1	7.10.1	7.1.8
7.9.5	7.1.7 SP2	1.5.0	7.10.1	7.1.7 SP2
7.9.5	7.1.7 SP1	1.5.0	7.10.1	7.1.7 SP1
7.9.5	7.1.8	1.5.0	7.10.1	7.1.8

Important: Upgrading from Cloudera Manager 7.7.3 version to Cloudera Manager 7.9.5 is currently not supported. Note that, Cloudera Manager 7.7.3 version is supported on Python 3 and Cloudera Manager 7.9.5 supports only Python 2. Cloudera Data Services on premises 1.5.0 support for users using Cloudera Manager 7.7.3 is intended to be made available in the future with a new version of Cloudera Manager that has support for both Python 3 and 2 versions respectively.

Table 5: Upgrade paths for Cloudera Data Services on premises 1.5.0

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
Greenfield deployment of Cloudera Data Services on premises 1.5.0				
NA	NA	NA	7.9.5	7.1.7 SP2
NA	NA	NA	7.9.5	7.1.7 SP1
NA	NA	NA	7.9.5	7.1.8
Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.0				
7.6.7	7.1.7 SP2	NA	7.9.5	7.1.7 SP2
7.6.1	7.1.7 SP1	NA	7.9.5	7.1.7 SP1
7.7.1	7.1.8	NA	7.9.5	7.1.8
Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premi				
7.8.1	7.1.7 SP1	1.4.1	7.9.5	7.1.7 SP1
7.8.1	7.1.8	1.4.1	7.9.5	7.1.8

[1] Upgrade from 7.1.7 SP2 to 7.1.8 CHF4 is not a mandatory upgrade. In 7.1.8 CHF4, you can install Ozone as a parcel.
[2] Upgrade from 7.1.8 to 7.1.8 CHF4 is not a mandatory upgrade. In 7.1.8 CHF4, you can install Ozone as a parcel.

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
7.6.5	7.1.7 SP1	1.4.0-H1	7.9.5	7.1.7 SP1
7.6.5	7.1.7	1.4.0-H1	7.9.5	7.1.7

For information on the pre-requisites on the upgrade, see Prerequisites for upgrading Cloudera Data Engineering service with endpoint stability on page 5 and Backing up Cloudera Data Engineering service using the docker image on page 8

For information on the upgrade, see Upgrade from 1.5.2 or 1.5.3 to 1.5.4 on Cloudera Embedded Container Service on page 13.

For information on the post-upgrade, see Restoring a Cloudera Data Engineering service on page 22, Rolling back the Cloudera Data Engineering service endpoint migration on page 24, and Limitations of Cloudera Data Engineering service endpoint migration on page 25

.

Upgrading Cloudera Manager

You must use Cloudera Manager version 7.11.3 CHF 6 to install or upgrade to Cloudera Data Services on premises 1.5.4.

If you already have a Cloudera Base on premises 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 6 before proceeding with the Cloudera Data Services on premises update.

Note: You must use Cloudera Manager version 7.11.3 CHF 11 to install or upgrade to Cloudera Data Services on premises 1.5.4 SP1.

Note: You must use minimum version of Cloudera Manager 7.13.1 CHF1 to install or upgrade to Cloudera Data Services on premises 1.5.4 SP2.

Related Information

Upgrading Cloudera Manager

Upgrade from 1.5.2 or 1.5.3 to 1.5.4 on Cloudera Embedded Container Service

You can upgrade your existing Cloudera Data Services on premises version 1.5.2 or 1.5.3 to 1.5.4 without performing uninstalling the previous version.

Before you begin

Note: If you are on Cloudera Data Services on premises 1.5.2 or 1.5.2 hotfixes, you MUST upgrade to Cloudera Data Services on premises 1.5.4 first, and then to 1.5.4 CHF1 or 1.5.4 SP1.

- 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 Cloudera Data Services on premises 1.5.2 or 1.5.3 was installed using the public Docker registry, Cloudera Data Services on premises 1.5.4 should also use the public Docker registry, and not be configured to use the embedded Docker registry. To use a different configuration for the Docker registry, you must perform a new installation of Cloudera Data Services on premises.

About this task

Note: Cloudera Embedded Container Service 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 Cloudera Embedded Container Service cluster to be running and in a healthy state.

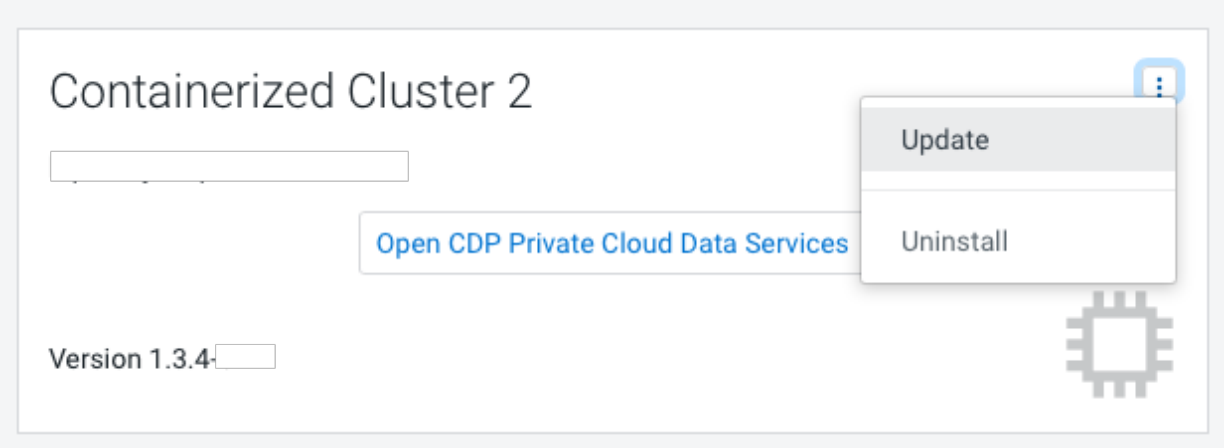
Important:

RHEL 7.x support on Cloudera Embedded Container Service has been dropped in Cloudera Data Services on premises 1.5.4 and higher versions. If you are running RHEL 7.x, you must upgrade to a higher version before upgrading.

Procedure

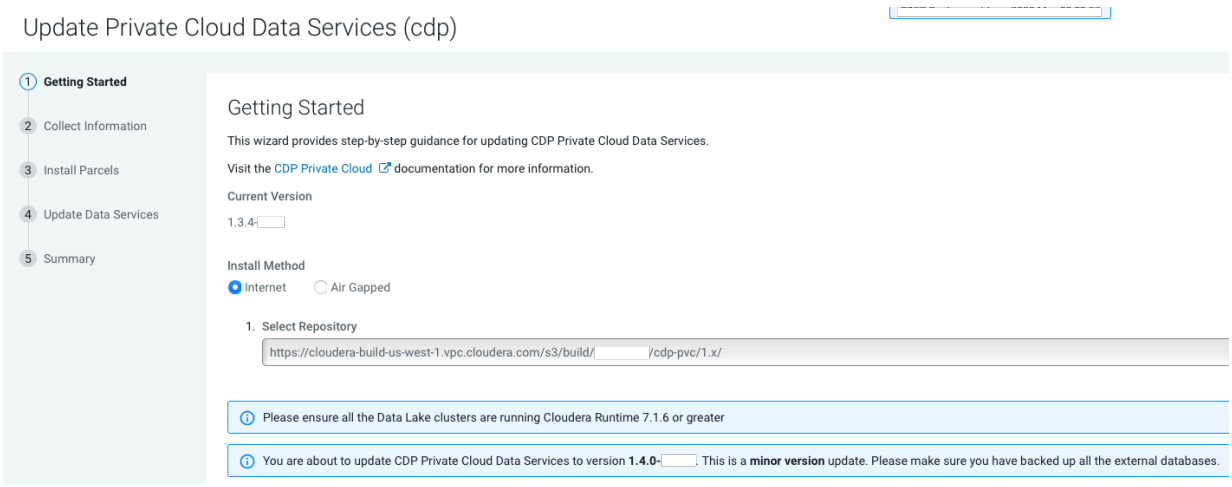
1.

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

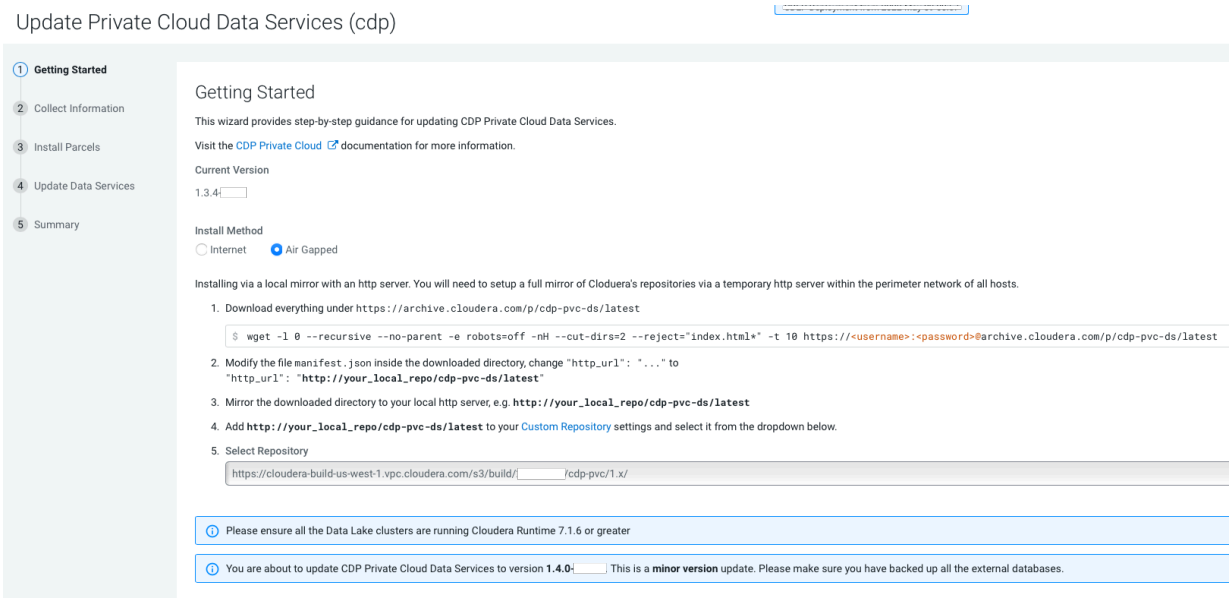


2. On the Getting Started page, you can select the Install method - Air Gapped or Internet and proceed.

   Internet install method

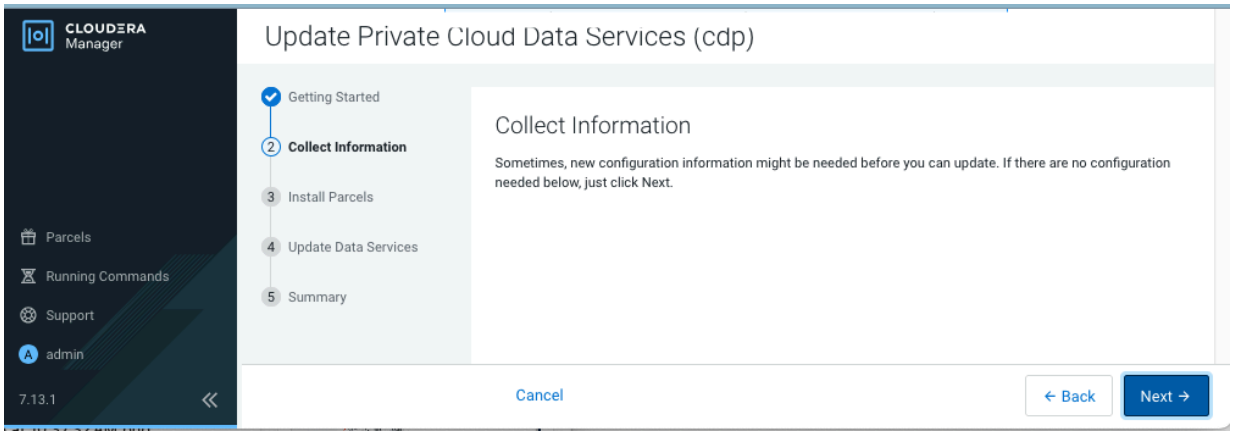


   Air Gapped install method

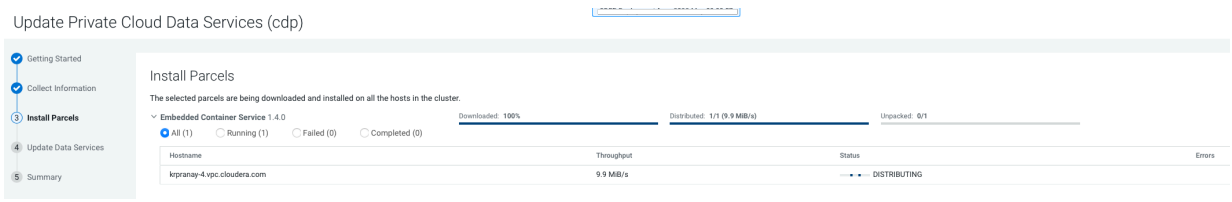


   Click Continue.

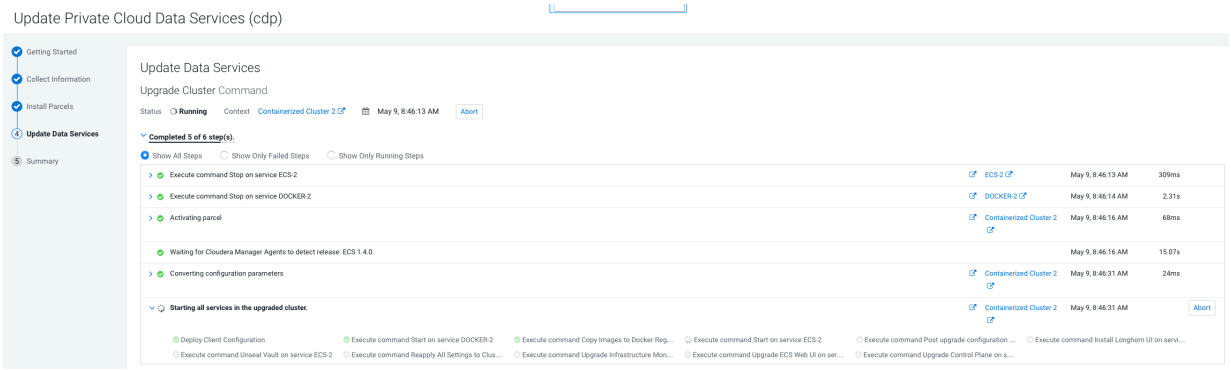
3. On the Collect Information page, click Next.



4. On the Install Parcels page, click Continue.



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



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 looks 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:

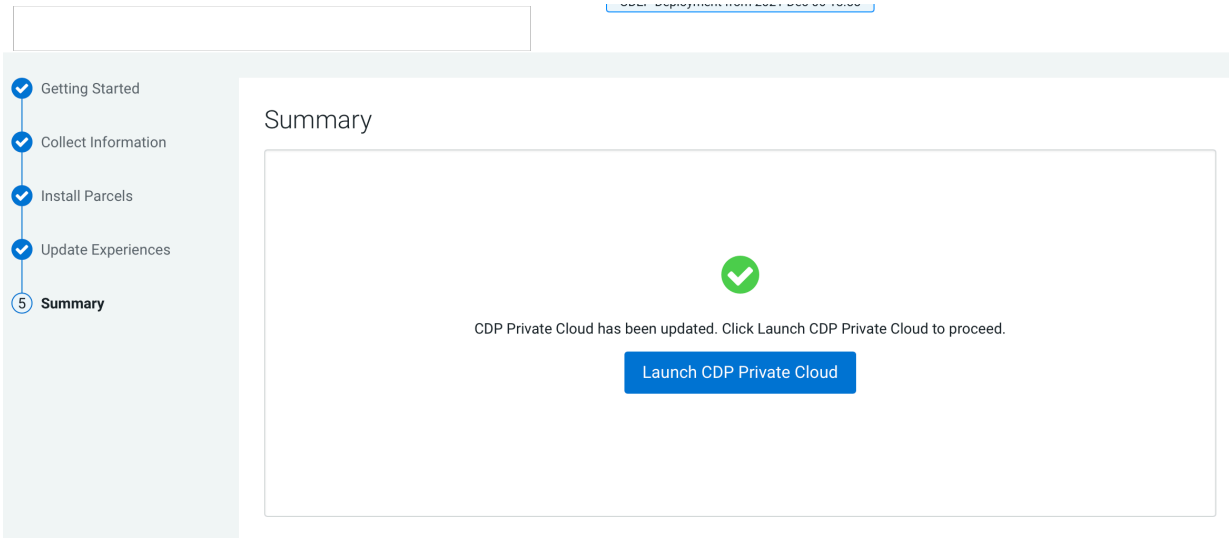
  ```
  NAME                      STATUS     ROLES
  AGE     VERSION
  ecs-abc-1.vpc.myco.com   Ready      control-plane,etcd,master
  3h47m   v1.21.11+rke2r1
  ecs-abc-2.vpc.myco.com   NotReady   <none>
  3h45m   v1.21.8+rke2r2
  ecs-abc-3.vpc.myco.com   NotReady   <none>
   3h45m   v1.21.8+rke2r2
  ecs-abc-4.vpc.myco.com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

  If you any nodes display a status of NotReady, click the Abort option 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 option 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.



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 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:

```
export PATH=$PATH:/opt/cloudera/parcels/ECS/installer/install/bin/linux/
:/opt/cloudera/parcels/ECS/docker
export KUBECONFIG=~/kubeconfig

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. If there are any pods in the "Pending" state, then you can try to reschedule the pods in the "Pending state" by restarting the yunikorn-scheduler. Run the following commands to restart yunikorn-scheduler:

```
kubectl get pods -n yunikorn

kubectl get deploy -n yunikorn

kubectl scale --replicas=0 -n yunikorn deployment/yunikorn-scheduler

kubectl get deploy -n yunikorn

kubectl scale --replicas=1 -n yunikorn deployment/yunikorn-scheduler

kubectl get deploy -n yunikorn
```

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

The Longhorn dashboard opens.

5. 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.
6. In the LongHorn UI, go to the Volume tab and check if any of the volumes are in the "Detached" state. If any are in the "Detached" state, then restart the associated pods or reattach them to the host manually.

- 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.

- If a new release-dwx-server pod is unable to start because of an existing release-dwx-server pod failing to start:
  - Delete the pod manually by executing the following command:

```
kubectl delete -n cdp pod cdp-release-dwx-server-<pod_id>
```

  - Resume the upgrade wizard if it had timed out.

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 Cloudera Embedded Container Service 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 Cloudera Private Cloud Data Services to 1.5.2 and higher, 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 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 1.5.2 and 1.5.3 to 1.5.4 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

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 on premises 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 will be helpful in debugging or when 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 the 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. Set the following environment variables in case you have exited from the ECS Server host:

```
export PATH=$PATH:/opt/cloudera/parcels/ECS/installer/install/bin/linux/:/
opt/cloudera/parcels/ECS/docker
export KUBECONFIG=~/kubeconfig
export BASE_WORK_DIR=/opt/backup-restore
export BACKUP_OUTPUT_DIR=/home/dex/backup
```

3. Run the dex-upgrade-utils docker image on the ECS Server host to restore the service.

```
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-b
48 restore-service -s cluster-c2dhkp22 -f $BACKUP_OUTPUT_DIR/cluster-c2d
hkp22-2023-03-10T06_00_05.zip
```

4. If you are using a Cloudera 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 have 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.

```
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 from 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 = '2025-05-05 06:16:37.786199' where id =
 'cluster-c2dhkp22'
```

5. 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.

6. 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. Do not delete the service if you want to rollback to the old service.

Related Information

Upgrading CDP on the Cloudera Embedded Container Service

Ozone Gateway validation

Prerequisites for upgrading Cloudera Data Engineering Service with endpoint stability

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 in debugging or when 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 CDE service must be in the Failed or Installed state before you perform the rollback command.

Procedure

1. Set the following environment variables in case you have exited from the ECS Server host:

```
export PATH=$PATH:/opt/cloudera/parcels/ECS/installer/install/bin/linux/:/
opt/cloudera/parcels/ECS/docker
export KUBECONFIG=~/kubeconfig
export BASE_WORK_DIR=/opt/backup-restore
export BACKUP_OUTPUT_DIR=/home/dex/backup
```

2. 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-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/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-b
48 rollback-restore-service -s cluster-92c2fkgb  -f $BACKUP_OUTPUT_DIR/c
luster-c2dhkp22-2023-03-10T06_00_05.zip
```

Limitations of Cloudera Data Engineering service endpoint migration

This page lists 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, 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.