Cloudera Data Engineering 1.20.

Upgrading Cloudera Data Engineering

Date published: 2023-10-10
Date modified:

CLOUDƎRA

Legal Notice

Contents

Upgrading Cloudera Data Engineering

Cloudera Data Engineering (CDE) supports in-place upgrades on both AWS and Azure. You can upgrade your service from CDE 1.19.2 and above for AWS and from CDE 1.19.4 and above for Azure to the latest CDE release. The upgrades can be triggered by an Admin from the CDE user interface.

About this task

When you upgrade to the latest version in Cloudera Data Engineering, the upgrade process will prepare your service for the upgrade with a click of a button. Your information will also be backed up for you in this process. Learn how to complete the upgrade process here.

> **Important:**
> - In-place upgrades GA do not work with AWS and Azure private clusters which are in tech preview. Instead of using an in-place upgrade for AWS and Azure private clusters, you must upgrade to the latest CDE version by cloning your CDE service and copying over jobs using the jobs backup & restore function.
> - Virtual clusters for this service are unavailable during the upgrade process and the upgrade will put your service into maintenance. During this time, scheduled jobs are paused and any active jobs are killed. The upgrade can take some time so be advised and alert your team of the downtime.

The following are not included in the backup:

- Python-venv resources
- Airflow custom operators & libraries
- Logs
- Job run history

Before you begin

Ensure that the option catchup is not enabled for any user's Airflow jobs. You must manually disable the Airflow DAG catchup options if they are enabled before the backup starts.

Procedure

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile.

   The CDE **Home** page displays.
2. Click Administration in the left navigation menu, select a Service, and click Upgrade.
3. Click the Maintenance tab and click Upgrade Service.
4. Click Start Preparation. CDE prepares the service for you. If the service can not be prepared, you can click Retry or Cancel Upgrade. Canceling the upgrade returns your service to the original state. Optionally, you can click Download Logs to view the Diagnostics page where you can generate a diagnostic bundle. If the preparation is successful, proceed to the next step. If the preparation and cancellation fails, you must contact Cloudera support.
5. Click Start Backup. CDE automatically creates a backup for you. Optionally, you can click Download Logs to view the Diagnostics page where you can generate a diagnostic bundle.

   > **Note:** If the backup fails, you have the option to click Retry Upgrade or Contact Support. If the upgrade fails a second time, you have the option to contact support or run a side-by-side upgrade. If the backup is successful, you can proceed with the upgrade.

6. When the backup is complete, you can click Backup Library to view the backup that was created.
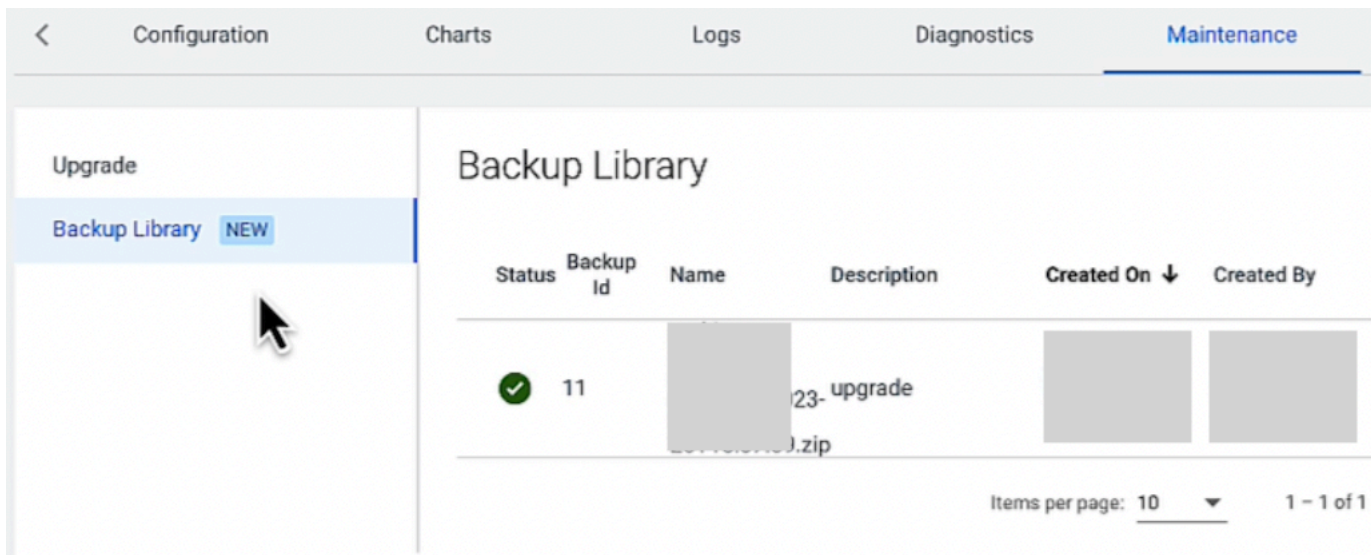
7. Click Start Upgrade.

8. Once the upgrade is complete, click Resume Service to enable all paused jobs and sessions.

The service restarts.

Results

Once the upgrade is complete, and you have resumed service, the service will operate on the upgraded version of CDE. If you wish to cancel the upgrade, your service will resume to the previous state which will enable all paused jobs and resume all sessions.

Related Information

Side-by-side upgrade for CDE

Handling upgrade failures for Cloudera Data Engineering

If your upgrade of Cloudera Data Engineering (CDE) fails, you have the option to clones the service with the latest version of CDE. Learn how to handle an upgrade failure.

About this task

During the Upgrading Cloudera Data Engineering process, linked below, a backup is created as part of the upgrade preparation process. This procedure will use that backup to be restored in a new cluster. The list of service backups is available in the Backup Library. To locate the Backup Library, in CDE, click Administration in the left navigation menu, click Service Details, click the Maintenance tab, and click Backup Library. Also, the list of all available backups can be obtained with the cdp de list-backups CDP CLI command. The list of backups of services that were associated with a specific CDP environment can be obtained with the "cdp de list-backups --filter   "environment(eq)<cdp-env-name>" command, where <cdp-env-name> is the name of CDP environment. The list of service backups are available in the Backup Library in the CDE UI by going to Administration > Service Details > Maintenance > Backup Library.

The CDE backup includes the following:

• CDE Service configurations
• Virtual cluster names
• Virtual cluster configurations
• Virtual cluster file-based resources
• Spark job definitions

- Airflow job definitions
- Spark Python-env resources

The following are not yet included in the backup:

- Non file-based resources, for example, Python-venv resources and custom runtimes
- Airflow custom operators & libraries
- Logs
- Job run history
- End points

Before you begin

⚠️ **Important:** Ensure that the option catchup is not enabled for any user's Airflow jobs. You must manually disable the Airflow DAG catchup options if they are enabled before the backup starts. This procedure will change the Virtual Cluster end-points. By default, the restored service will have the name and ID of the original backed-up service. This means that the original service, which failed to upgrade during Upgrading Cloudera Data Engineering linked below, must be deleted first. Otherwise, an attempt to restore from the backup will fail due to the name and ID conflicts.

Steps:

1. You must re-name and assign a new ID by using the following options:

```
--service-id and --service-name
```

📝 **Note:** Assigning a new ID results in new fully qualified domain names (FQDNs) for the service and all its Virtual Clusters, meaning no endpoint stability.

Procedure

1. Restore the service from the backup. Note that backup-id is the ID of the backup that you are restoring from, and environment-crn is the Customer Resource Number (CRN) of the Cloudera Data Platform (CDP) environment with which a restored CDE service will be associated. Currently, the CDE service can be restored only to the same CDP environment where a backed-up service was associated:

```
cdp de restore-service --backup-id <backup-id> --environment-crn
    <environment-crn>
```

For example:

```
cdp de restore-service --backup-id 2 --environment-crn crn:cdp:environme
nts:us-west-1:9d74eee4-1cad-45d7-b645-7ccf9edbb73d:environment:c67b9089-
2d3b-4579-861d-c0df12a105b1
```

2. To obtain a list of backups, use the following command:

```
cdp de list-backups
```

a) You can describe a particular backup with the following command:cdp de describe-backup --backup-id   < backup-id>

For example:

```
$ cdp de describe-backup --backup-id 2 --profile priv
{
    "backup": {
        "id": 2,
        "serviceID": "cluster-cf6h74lq",
        "serviceName": "dex-priv-default-azure-env-1689008683873",
```

```
        "environmentName": "dex-priv-default-azure-env",
        "environmentCrn": "crn:cdp:environments:us-west-1:9d74eee4-1ca
d-45d7-b645-7ccf9edbb73d:environment:c67b9089-2d3b-4579-861d-c0df12a105b
1",
        "creator": "crn:altus:iam:us-west-1:9d74eee4-1cad-45d7-b645-7ccf
9edbb73d:user:0f9a97a7-23a7-43bd-bc71-ecdb2aa34ed5",
        "cloudPlatform": "AZURE",
        "status": "completed",
        "created": "2023-07-17T18:02:58.385455Z"
    }
}
```

Related Information

Upgrading CDE