Cloudera Runtime 7.1.9

Migrating from Navigator to Atlas using CDH 6 with Cloudera Manager 7

Date published: 2023-08-31
Date modified:

CLOUDERA

Legal Notice

Contents

About migrating from Cloudera Navigator to Atlas using Cloudera Manager 7

Navigator role instances under the **Cloudera Management Service** are no longer available while using **Cloudera Runtime 7.1.9** release. You must perform the migration process before you upgrade to Cloudera Manager 7.11.3 Latest cumulative hotfix version.

Note: This change does not impact Navigator Encrypt and Navigator KTS functionality.

Attention: Migrating from Navigator to Atlas is **NOT** required if:

- Cloudera Navigator is not installed on the Cloudera Enterprise 6 cluster that you are planning to upgrade to Cloudera Runtime 7.1.9 version.
- Cloudera Navigator is installed on your CDH 6 cluster and you do not require the pre-upgrade lineage information from Navigator.

Caution: Currently, you cannot perform the Nav2Atlas transition and the pre-upgrade lineage data will not be available after the CDH 6 to CDP 7.1.9 CHF upgrade process is completed. You can only export the Cloudera Navigator lineage using the cnav.sh tool and later proceed with the remaining steps of the upgrade process.

The result of this change is that, during the Nav2Atlas migration's first step, the Navigator lineage export process must be carried out before upgrading Cloudera Manager 7.11.3 Latest cumulative hotfix.

Previous versions of Cloudera Manager 7 contained the cnav.sh script and included dependencies which are responsible to perform the Nav2Atlas export step.

Note the following regarding the Cloudera Manager 7.11.3 Latest cumulative hotfix / Cloudera Runtime 7.1.9 release:

- CDH 5 version is **NOT** supported in CDP 7.1.9 release.
- CDH 6 version is supported in CDP 7.1.9 release without Cloudera Navigator.
- (**Optional**) If Navigator service is available on your CDH cluster, before the CDH 6 cluster upgrade process, you must plan to export the lineage information to a zip file using the cnav tool and later remove Navigator. When you upgrade to Cloudera Manager 7.11.3 Latest cumulative hotfix / Cloudera Runtime 7.1.9 version, Navigator service is not available.

Supported CDH upgrade paths

Regular upgrade for CDH is supported using a couple of scenarios.

- **Cloudera Manager 7** + **CDH 6** -> **CDP 7.1.9** / **Cloudera Manager 7.11.3 Latest cumulative hotfix**
- **Cloudera Manager 6** + **CDH 6** -> **CDP 7.1.9** / **Cloudera Manager 7.11.3 Latest cumulative hotfix**

In the earlier upgrade processes, Cloudera Manager from version 6 was upgraded to 7 and later the Navigator metadata export was executed using the cnav tool. With the release of Cloudera Manager 7.11.3 Latest cumulative hotfix, cnav export process takes precedence over the Cloudera Manager upgrade operation.

Migrating Cloudera Navigator to Atlas while upgrading Cloudera Manager 7 with CDH 6

Learn about migrating Navigator to Atlas while upgrading your CDH 6 clusters with Cloudera Manager 7 to Cloudera Runtime 7.1.9 version. You must perform the migration process before you upgrade to Cloudera Manager 7.11.3 Latest cumulative hotfix.

Remember: Migrating from Navigator to Atlas is **NOT** required if:

- Cloudera Navigator is not installed on the Cloudera Enterprise 6 cluster that you are planning to upgrade to Cloudera Runtime 7.1.9 version.
- Cloudera Navigator is installed on your CDH 6 cluster and you do not require the pre-upgrade lineage information from Navigator.

As a pre-upgrade step, you must first complete the transition of Navigator data using the customized scripts

Transitioning Navigator data using customized scripts

You can run the customized scripts generated by the Cloudera Manager wizard to configure the Apache Atlas installation and start the Navigator-to-Atlas data migration process when you step into the CDP upgrade wizard. You can also run the migration scripts independently from the CDP upgrade.

The transition has three phases: extraction, transformation, and import. If you have not already, you must estimate the time and resource requirements for the migration steps.

Related Information

Estimating the time and resources needed for transition

Estimating the time and resources needed for transition

While the cluster is starting up, you can plan for and start the transition process.

- Inspect Navigator installation to determine the number of Navigator entities that will be transitioned.
- Estimate the time and disk space required for each phase of the transition.

  The following transition rates are approximate and depend on the resources available on the Atlas host and other unknown factors. Note that the number of entities actually imported may be considerably less that the number of entities extracted. The transition process discards HDFS entities that are not referenced by processes that are transitioned (Hive, Impala, Spark).

Transition Phase	Transition Rate	Disk Space	Output File Size	Trial Data Points
Extraction	4 minutes / 1 million entities	100 MB / 1 million entities, less as volumes increase	65 MB / 1 million entities	10 million entities takes about 30 minutes; 256 million takes about 18 hours.
Transformation	1.5 minutes / 1 million entities	100 to 150 MB / 1 million entities, higher end of range with larger volumes	150 MB / 1 million entities	10 million entities takes about 20 minutes; 256 million takes about 6 hours.
Import	35 minutes / 1 million migrated entities	N/A	N/A	10 million entities takes about 4 hours; 256 million takes about 6 days.

Related Information

Migrating Cloudera Navigator to Atlas while upgrading Cloudera Manager 7 with CDH 6

Run the extraction for CDH 6 with Cloudera Manager 7

Extraction must be executed for all Navigator Metadata Server clusters prior to the Cloudera Manager upgrade.

Procedure

1. Determine the required information for the extraction:

   • Planned host for the Atlas service
   • JAVA_HOME location
   • Navigator Metadata Server cluster names
   • Navigator Metadata Server admin users
   • Keystore and truststore paths and passwords

```
$ ssh <planned Atlas host>
$ export JAVA_HOME=…
$ /opt/cloudera/cm-agent/service/navigator/cnav.sh -n http://<NAVIGATOR_
METADATA_HOST>:7187 -u<NAVIGATOR_USERNAME> -p<NAVIGATOR_PASSWORD> -c '<C
LUSTER_NAME>' -o '/tmp/nav2atlas/<cluster_name>_navigator_data.zip'
```

2. Open a terminal window or command prompt from where you can access the cluster.

3. Using the provided command, SSH into the Atlas host.

4. Run the export command pointing to the location of the JDK, to make sure the JAVA_HOME variable is set.

5. Customize the extraction command to include the Navigator admin user and password.

6. Run the extraction command.
   When the extraction is complete, you can view a status message in the command output.

   If Navigator is configured with TLS/SSL, the cnav script needs the following credential information:

   • Truststore path
   • Truststore password
   • Keystore path
   • Keystore password

   To make these parameters available, run the following commands before running the cnav script:

   export KEYSTORE_PATH=<keystore-path>;

   export   KEYSTORE_PASSWORD=<keystore-password>;

   export TRUSTSTORE_PATH=<truststore-path>;

   export   TRUSTSTORE_PASSWORD=<truststore-password>

   For example, the command sequence might look similar to the following (line breaks are introduced for readability):

   export   KEYSTORE_PATH=/opt/cloudera/CMCA/trust-store/acme_truststore.jks;

   export

   KEYSTORE_PASSWORD=Q01FAeH53dn1HLY74D68Kk1yMAQVGtOI_cLznArccid48DDzS0VXY-DWnz
   pz0Ug10BvikGMoovYaZT2EEEdBGgLPiDCRKHyzFExE3OITRGazjKPtZxAaXOUzgKMMmQQgJKw-5J
   W9I6WgLGBhCpKfBa7vP3z6PFtm6XfYB-o3R6qmcdzZLwslDIQ18mowuFV1ouQIzA;

```
export
                             TRUSTSTORE_PATH=/var/lib/cloudera-scm-agent/a
gent-cert/cm-auto-host_keystore.jks;
```

   export TRUSTSTORE_PASSWORD=123420978alngdfdfjliaiu;

```
/opt/cloudera/cm-agent/service/navigator/cnav.sh -n
                         https://acme-finance-1.acme-finance:7187 -u admin
 -p adminpass -c "Cluster 2" -o
```

```
/tmp/nav2atlas/nav2altas_nav_export.zip
```

Note: The -c flag parameter in the nav2atlas script represents the Navigation Metadata Server (NMS) cluster name and NOT the Cloudera Manager cluster name. You must retrieve the value of -c from the Navigator UI to use it while running the nav2atlas script.

Attention: Nav2Atlas data migration can fail with **FileNotFoundException** error. The cnav tool changes the file name to lowercase if the zip file has uppercase. To overcome this behavior, you must:

- Change the file name to lower case.
- Configure atlas.nav2atlas.use.spawner=false.
- Rerun cnav.sh tool but with the -r   ON option supplied and the -o option pointing to the existing cnav output zip file.

Important: In a CDH multi-cluster environment, multiple clusters are managed by one Navigator instance. In a Cloudera Private Cloud Base deployment, each cluster has its own Atlas servers manage the respective cluster metadata. As of Cloudera Runtime 7.1.8 release, in a CDH multi-cluster environment, one of the CDH clusters was upgraded to CDP (where Atlas collected the service metadata) and the other CDH cluster originally communicated with Navigator.

Starting with **CDP Runtime 7.1.9** release, Navigator service is not available. Before upgrading to Cloudera Manager, the Navigator data (including lineage information) must be extracted from all the CDH clusters before removing the Navigator role instances.

Removing Navigator role instances

You must stop and remove the Navigator role instances in your Cloudera Manager instance.

Procedure

1. Navigate to Cloudera Manager -> Cloudera Management Service -> Instances.
2. Select Navigator Audit Server and Navigator Metadata Server.
3. 3. Under the "Actions for selected" scroll-down menu, select Stop.

4. Select Delete.

What to do next

Once the Navigator role instances are successfully removed from your Cloudera Manager instance, upgrade to Cloudera Manager 7.11.3 / CDP 7.1.9 version. For more information, See CDH 6 to CDP 7.1.9 upgrade documentation.

Post upgrade operations for your CDH 6 clusters

Once the CDH 6 with Cloudera Manager 7 upgrade to Cloudera Runtime 7.1.9 version process is completed, you must perform some of the post-upgrade operations.

Run the transformation for CDH 6 with Cloudera Manager 7

You can run the transformation as soon as the CDP upgrade process is completed. To customize and run the transformation command. You can customize and run the transformation command.

Procedure

1. Go back to the editor where you saved the transformation commands, from copy the transformation command text from the step "Click Add Atlas Service."

```
$ ssh finance-3.finance.acme-core.site
$ export JAVA_HOME=…
$/opt/cloudera/parcels/CDH-7.1.9-1.cdh7.1.9.p0.xxxxxxxx/lib/atlas/tools/
nav2atlas.sh -f tmp/nav2atlas/cluster_1_navigator_data.zip' -o '/tmp/nav
2atlas/cluster_1_atlas_data.zip' clusterName 'Cluster 1'
```

2. If needed, open a terminal window or command prompt where you have access to the cluster.

3. If needed, SSH into the Atlas host.

4. If needed, make sure the JAVA_HOME variable is set; if it isn't, run the export command pointing to the location of the JDK.

5. Run the transformation command.

When the transformation is complete, you'll see a status message in the command output.

Run the import for CDH 6 with Cloudera Manager 7

When Atlas starts up, it checks the output file location for a completed file from the transformation phase. When Atlas finds the completed file, it begins importing entities automatically. To view the progress of the import, follow these steps.

Procedure

1. Open the Atlas UI.

You can open Atlas from the Atlas service page in Cloudera Manager.
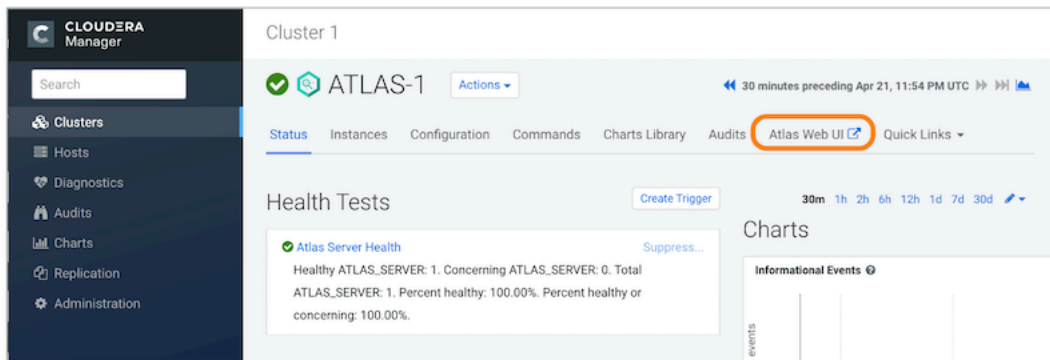
2. Review transition progress in the Statistics page.

The normal Atlas UI does not appear in migration mode; instead you shall view the Statistics page, which shows a real-time report of the number of entities, classifications, and other metadata that have been created in Atlas.

Moving Atlas out of migration mode

After installation, Atlas runs in migration mode and you must move Atlas out from the migration mode.

Before you begin

- Atlas does not collect metadata from services running on the cluster. The metadata remains in Kafka topics and is collected later

- Atlas starts importing metadata when it finds a final transformation file in the location you specified in output file location from the step Click Add Atlas Service.

  > ⚠️ **Attention:** This is the location where Atlas will look for the content to import. Make sure it matches the location you plan to use for the output of the transformation command.

  1. Open Cloudera Manager to the Atlas service.
  2. Go to the Configuration tab.
  3. Filter the list of properties by typing "Safety" in the filter field.
  4. Remove the migration-specific entries from the Atlas Server Advanced Configuration Snippet (Safety Valve) for conf/atlas-application.properties

     You must remove the following properties:

     - atlas.migration.data.filename
     - atlas.migration.mode.batch.size
     - atlas.migration.mode.workers
  5. Reset the Atlas Max Heapsize property back to the default value.

Click to set the default value

Atlas Max Heapsize

atlas_max_heap_size

Atlas Server Default Group ↩

16   GiB ⬍
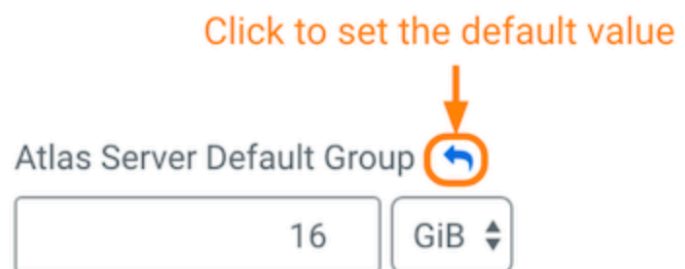
  6. Click Save Changes.
  7. Restart Atlas. Choose Action.
  8. Restart.

Related Information

Run the import for CDH 6 with Cloudera Manager 7