Cloudera Runtime 7.1.0

Migrating data from Navigator to Atlas

Date published: 2020-05-30
Date modified:

CLOUDⅬRA

Legal Notice

Contents

Migrating Navigator content to Atlas

As part of upgrading a CDH cluster to CDP, you can choose to migrate much of your Navigator metadata to Atlas.

Cloudera Runtime 7 includes Apache Atlas to collect technical metadata from cluster services. Atlas replaces Cloudera Navigator Data Management for these clusters. Cloudera has incorporated many features from Navigator into Apache Atlas to make sure that the rich metadata collected in Navigator can be represented in Atlas. Atlas provides scalable and robust infrastructure that supports metadata searches and lineage across enterprise production clusters.

> **Note:** Governance through Apache Atlas is just one element of a secure production cluster: Cloudera supports Atlas when it runs on a cluster where Kerberos is enabled to authenticate users. When upgrading from to Cloudera Runtime 7.1.1 and running Apache Atlas, the new cluster must have Kerberos enabled.

You may choose not to migrate Navigator content to Atlas at all: this document describes how to think about archiving your Navigator audits and metadata.

Whether you choose to migrate Navigator contents to Atlas or not, this document describes how to use Atlas to accomplish the tasks you are accustomed to performing in Navigator.

What's migrated?

Business metadata is migrated into Atlas, including:

- Tags
- Custom properties (definitions and entity assignments)
- Managed metadata properties (definitions and entity assignments)
- Original and updated entity names and descriptions

Technical metadata from the following sources are migrated into Atlas:

- Hive
- Impala
- Spark
- Referenced HDFS / S3

What's NOT migrated?

- Audits. In CDP, Ranger collects audit information for successful and failed access to objects under its control. This audit system is focused and powerful, but it's enough different from how Navigator collected audits that migration isn't appropriate. This document includes information on how to transition your auditing to Ranger and how to archive your existing Navigator audit information.
- Entity Metadata. The following metadata entities in Navigator are not migrated to Atlas:

    - Unreferenced S3 and HDFS entities. Files in HDFS and S3 that are not included in lineage from Hive, Spark, or Impala entities are not migrated.
    - Metadata for Sqoop, Pig, Map-Reduce v1 and v2, Oozie, and YARN.
- Policies. Navigator policies are not migrated to Atlas.
- Configuration settings. Configuration properties you've set in Cloudera Manager that determine Navigator behavior are not migrated to the new environment. If you have properties that may apply in Atlas, such as authentication credentials, you'll need to reset them in the new environment.

Will Navigator still run in Cloudera Manager?

After upgrading Cloudera Manager to CDP, Navigator continues to collect metadata and audit information from CDH cluster services. There are no changes to Navigator functionality; all Navigator data is retained in the Cloudera Manager upgrade.

After upgrading a CDH cluster, services that previously sent metadata and audit information to Navigator, such as Hive, Impala, Spark, and HBase, are configured to pass metadata to Atlas. Navigator audit collection for those services is disabled. You can still access audits and metadata through Navigator; however, Navigator will not collect new information from cluster services. When you decide you have exhausted the value of the Navigator audits and after you've converted Navigator metadata to Atlas content, you can disable Navigator servers.

Related Information

High-level migration process

Assumptions and prerequisites

Installing Atlas in the Cloudera Manager upgrade wizard

Migrating Navigator data using customized scripts

Mapping Navigator metadata to Atlas

Transitioning Navigator audits

What's new in Atlas for Navigator Users?

Atlas technical metadata migration reference