

Migrating data from Navigator to Atlas

Date published: 2020-05-30

Date modified: 2022-03-30



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

Transitioning Navigator content to Atlas.....	4
--	----------

Transitioning Navigator content to Atlas

During the transition from CDH to CDP, you can transition the metadata from Navigator to Apache Atlas, for a scalable and robust infrastructure that supports enhanced metadata searchability by collecting of metadata directly from your cluster.

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 transition Navigator content to Atlas at all: this document describes how to think about archiving your Navigator audits and metadata.

Whether you choose to transition 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 transitioned?

Business metadata is transitioned 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 transitioned into Atlas:

- Hive
- Impala
- Spark
- Referenced HDFS / S3

What's NOT transitioned?

- 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 transition 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 transitioned 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 transitioned.
 - Metadata for Sqoop, Pig, Map-Reduce v1 and v2, Oozie, and YARN.
- Policies. Navigator policies are not transitioned to Atlas.
- Configuration settings. Configuration properties you've set in Cloudera Manager that determine Navigator behavior are not transitioned 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](#)

[Transitioning 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](#)