Troubleshooting Cloudera Data Catalog

Date published: 2019-11-14
Date modified: 2025-10-17

CLOUDERA

Legal Notice

Contents

Missing authorization for viewing assets

Cloudera Data Catalog users must have appropriate authorization set up in Apache Ranger to view certain asset types.

Lack of permission to view a dataset

If you do not have access to a dataset, a similar error message to the following can be seen:
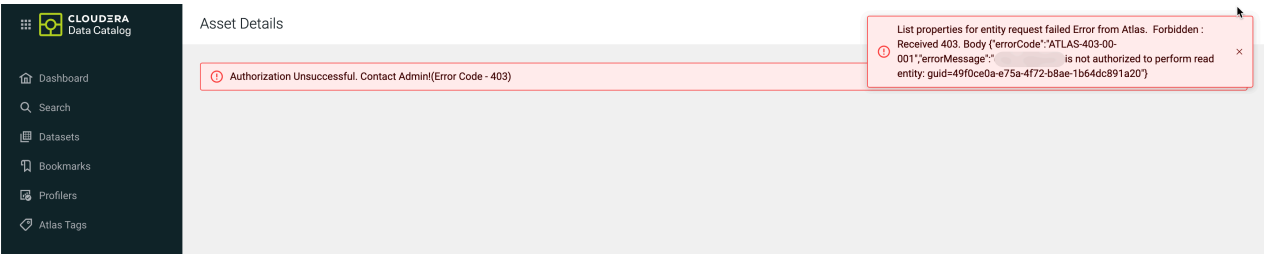
The previous error message is the result of a Hive policy setting similar to the following:
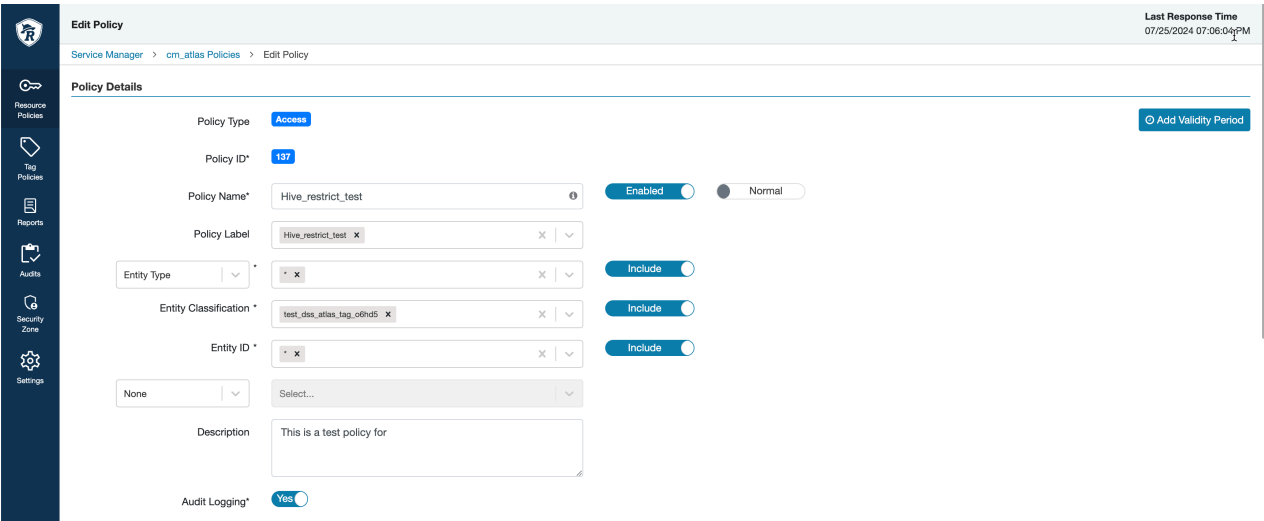
The following image displays the **Deny Conditions** set for the specific user.
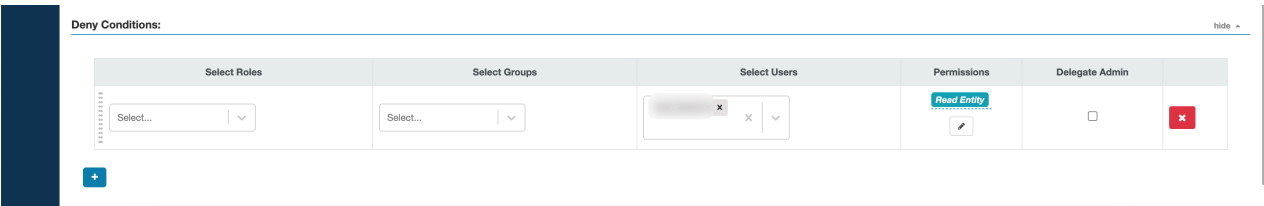
Hive Ranger policy

You must set up Hive Ranger resource-based policies as per your work requirements for Hive assets in Cloudera Data Catalog.

The following diagram provides a sample Hive Ranger policy with all permissions.

Atlas Ranger policy

Additionally, you must set up Ranger policies for Atlas to work with asset search and tag flow management.

Note: When a subscriber is provided with user access either as EnvironmentAdmin or EnvironmentUser, by default, the user is synchronized with the default Atlas policies in Ranger. This implies that the same user has access to all the datasets in the environment.

The following diagram provides a sample Atlas Ranger policy.

Reducing resource consumption with restricted users

When you plan to restrict data access, note the following:

- Audit summarization for the asset evolves from the Ranger Audit Profiler and Metrics service.
- Various Hive Column Statistical metrics for columns of the asset evolves from Atlas as part of the profile_data of a column.

To ensure that the data related to audit summary and Hive Column Statistics are not visible to the subscribers, you must make sure to turn off the Ranger Audit Profiler/Activity Profiler and the Hive Column Profiler/Statistics Collector Profiler respectively.

Related Information
Providing role access
Ranger Policies Overview
Resource-based Services and Policies
Configuring resource-based policies

FAQs

This page provides answers to frequently asked questions about problems that you might encounter while you use Cloudera Data Catalog.

Issue while accessing Cloudera Data Catalog

You must note about specific permissions that are needed to access Cloudera Data Catalog.

Users have to add themselves to at least one Data Lake as an Environmental   Admin or Environmental User. This allows you to search for assets on the Data Lake, add tags, and business glossaries to these assets.

In order to use Profilers (Cloudera Data Hub clusters specific to Cloudera Data Catalog or Kubernetes based profiling jobs in Compute Cluster based environments) related features, you must have a PowerUser role. This runs Spark jobs and provides some valuable insights / suggestions.

Associating tags to Hive assets

Associating tags to Hive assets fails on 7.2.16 cluster.

Note: This issue only applies for VM-based environments.

The issue can be resolved by adding the following Spark configuration:

spark.sensitive.profiler.tableSamplePercentage=50 to profiler scheduler conf by adding the configuration under the key - profiler_scheduler_spark_conf.

The value passed in the configuration must be an integer value ranging from 1 to 100.

This value indicates the percentage of rows to be picked up as a sample for the profiler job.