CDP Private Cloud Data Services / Data Catalog 1.5.5

Search

Date published: 2023-12-16
Date modified: 2025-11-10

https://docs.cloudera.com/

Legal Notice

Contents

Search filters

Use filters to refine the overview of all your available assets.

Note: You must have access to at least one data lake to search and filter your results. By default, a data lake is already selected for you if you have access to it.

You can use the following filters:

Owner

From all the owner names that appear, you can select the owner to further refine the results and display those search results with the selected owner.

Type

Select an entity type to view all the assets stored in that type of database.

Additional Type filters can be added by clicking + Add New Value.

Note: For information purposes, the Database filter is displayed as Namespace in case of HBase tables.

By default, only the following options are visible:

- Hive Table
- HBase Table

Entity Tag

Use entity tags to refine your search results. You can add business metadata as entity tags in Atlas as classifications, or in the **Atlas Tags** menu. Use these tags to refine your search results and view the details of the required data asset.

Additional entity tags can be added by clicking + Add New Value.

Column Tag

You can search for the following type of table assets by tags that have been applied on their children entities, that is, columns or column families using the column tags filter:

- Hive
- HBase

Created Within

You can choose to refine your search results of assets within the data lake to view the data assets created within the last 7 days, 15 days, or 30 days. You can also add custom values such as 5 days or 10 days to view specific information.

Created Before

Depending on the time when the assets were created, you can choose to refine the search results and view data assets created before 1 day, 7 days, or 15 days. You can add custom values to view data assets created before the days of your preference such as 8 days or 12 days.

Note: The Created Within and Created Before filters are only visible when Atlas provides the created time for the assets.

Glossary

You can filter assets based on business glossary terms. You can search for any asset without any entity type restrictions.

Note: This filter appears only if Atlas has terms set up.

Click Cancel for any filter to clear the selection or Clear All to reset all your filters.

Searching for assets using Atlas glossaries

Use Apache Atlas glossaries to define a common set of search terms that data users across your organization use to describe their data.

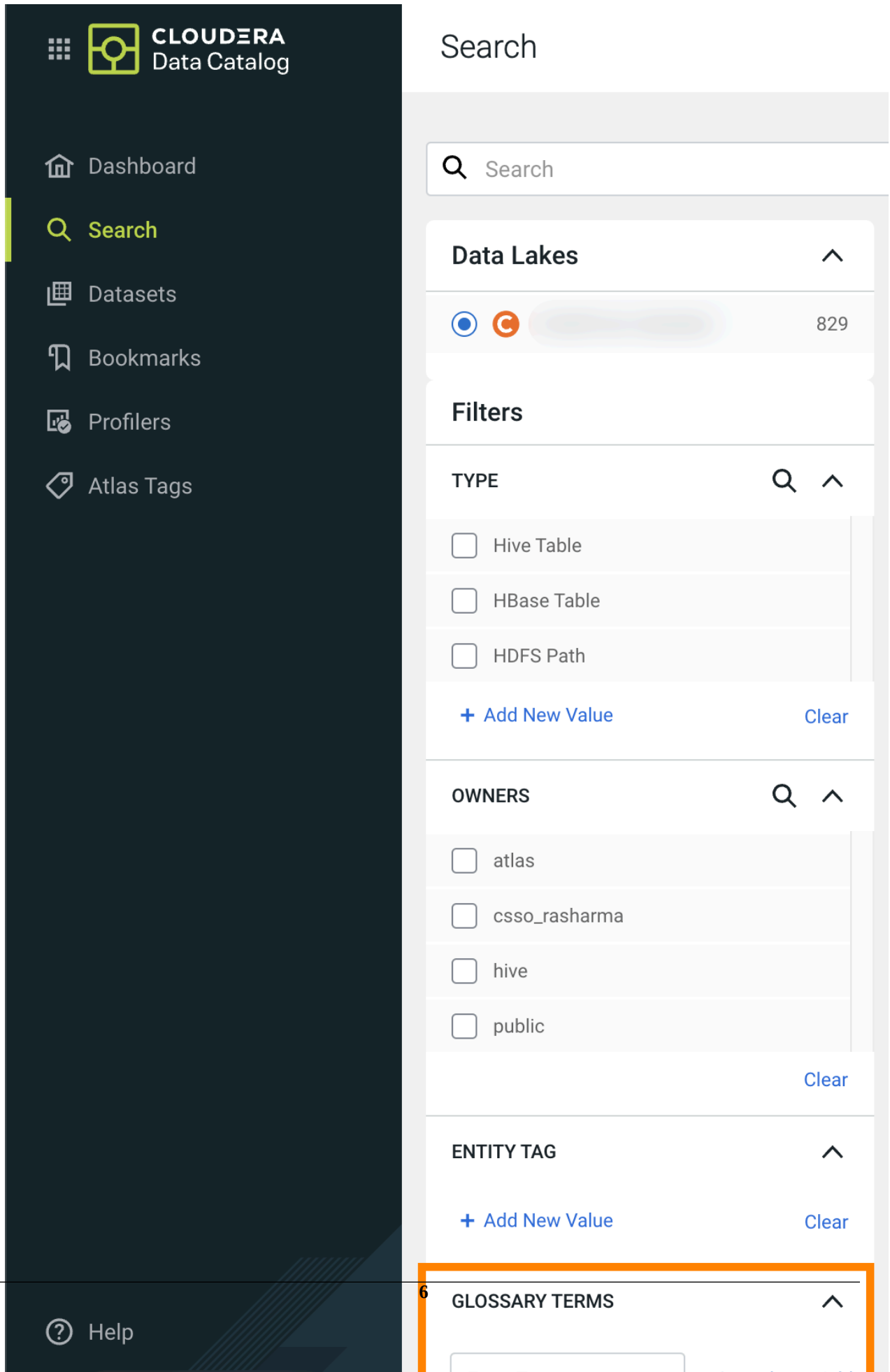
Data can describe a wide variety of content: lists of names or text or columns full of numbers. You can use algorithms to describe data as having a specific pattern, of being within a range or having wide variation, but what's missing from these descriptions is what does the data mean in a given business context and what is it used for? Is this column of integers the count of pallets that entered a warehouse on a given day or number of visitors for each room in a conference center?

The glossary is a way to organize the context information that your business uses to make sense of your data beyond what can be figured out just by looking at the content. The glossary holds the terms you've agreed upon across your organization so business users can use familiar terms to find what they are looking for.

Glossaries enable you to define a hierarchical set of business terms that represents your business domain.

Glossary terms can be thought of as of a flat (but searchable) list of business terms organized by glossaries. Unlike classifications, terms are not propagated through lineage relationships: the context of the term is what's important, so propagation may or may not make sense.

You can search for the datasets using the Glossary Terms filter available on the **Search** page.


CLOUDERA
Data Catalog
Dashboard
Search
Datasets
Bookmarks
Profilers
Atlas Tags
Help
Search
Search
Data Lakes
829
Filters
TYPE
Hive Table
HBase Table
HDFS Path
+ Add New Value
Clear
OWNERS
atlas
csso_rasharma
hive
public
Clear
ENTITY TAG
+ Add New Value
Clear
GLOSSARY TERMS

Using terms in Cloudera Data Catalog

You can use the Asset Details page to add or modify Apache Atlas glossary terms for your selected assets. You can search these glossary terms by using the search filters.

Use Atlas to define rich glossary vocabularies using the natural terminology (technical terms and/or business terms) of your industry. You can also create semantic relationships between your terms. Then, in Cloudera Data Catalog, use the **Terms** widget in the **Asset Details** page to map assets to glossary terms.

You can use terms in Cloudera Data Catalog to search for entities, filter them by glossary term(s), and also search for entities associated with them in Atlas.

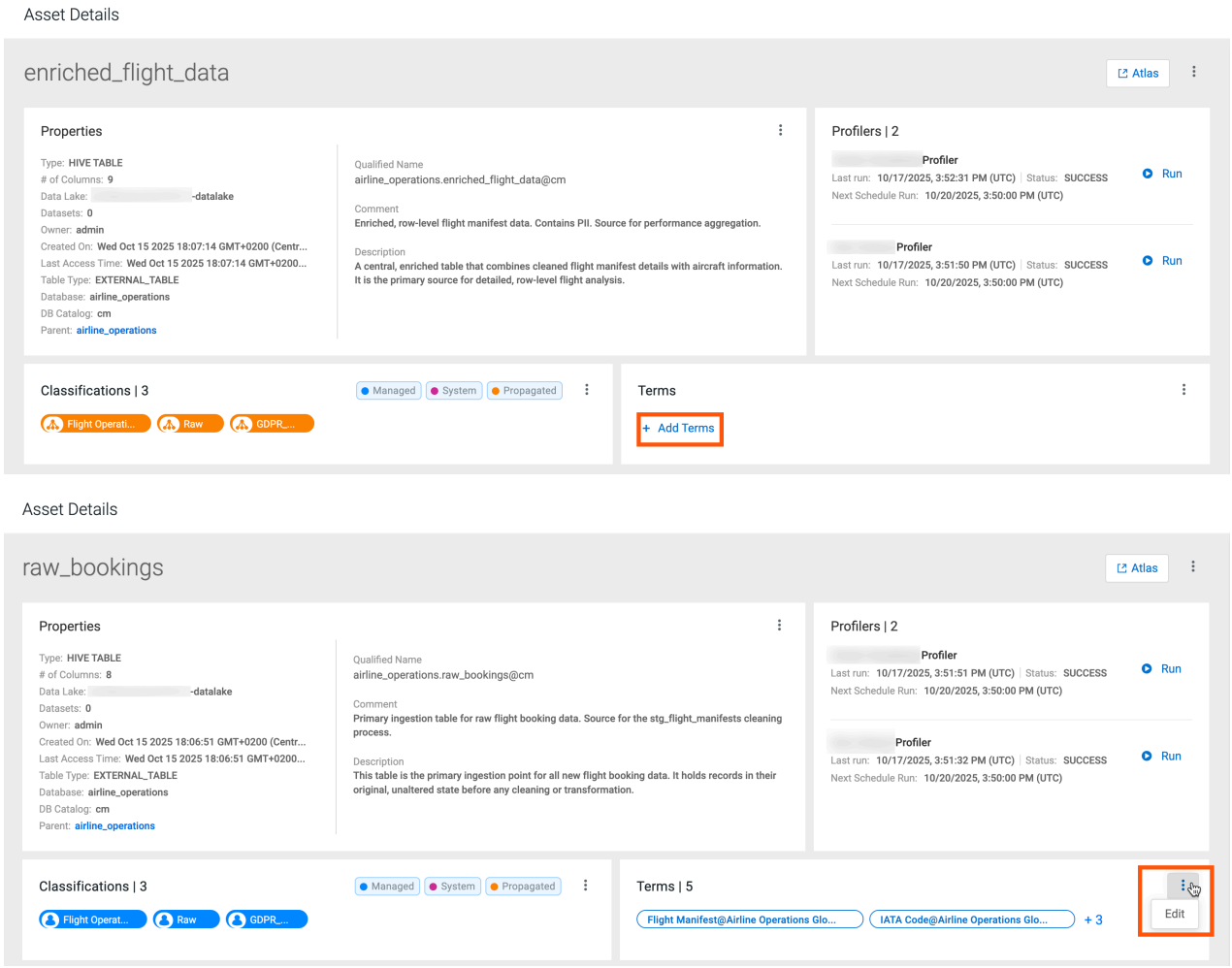
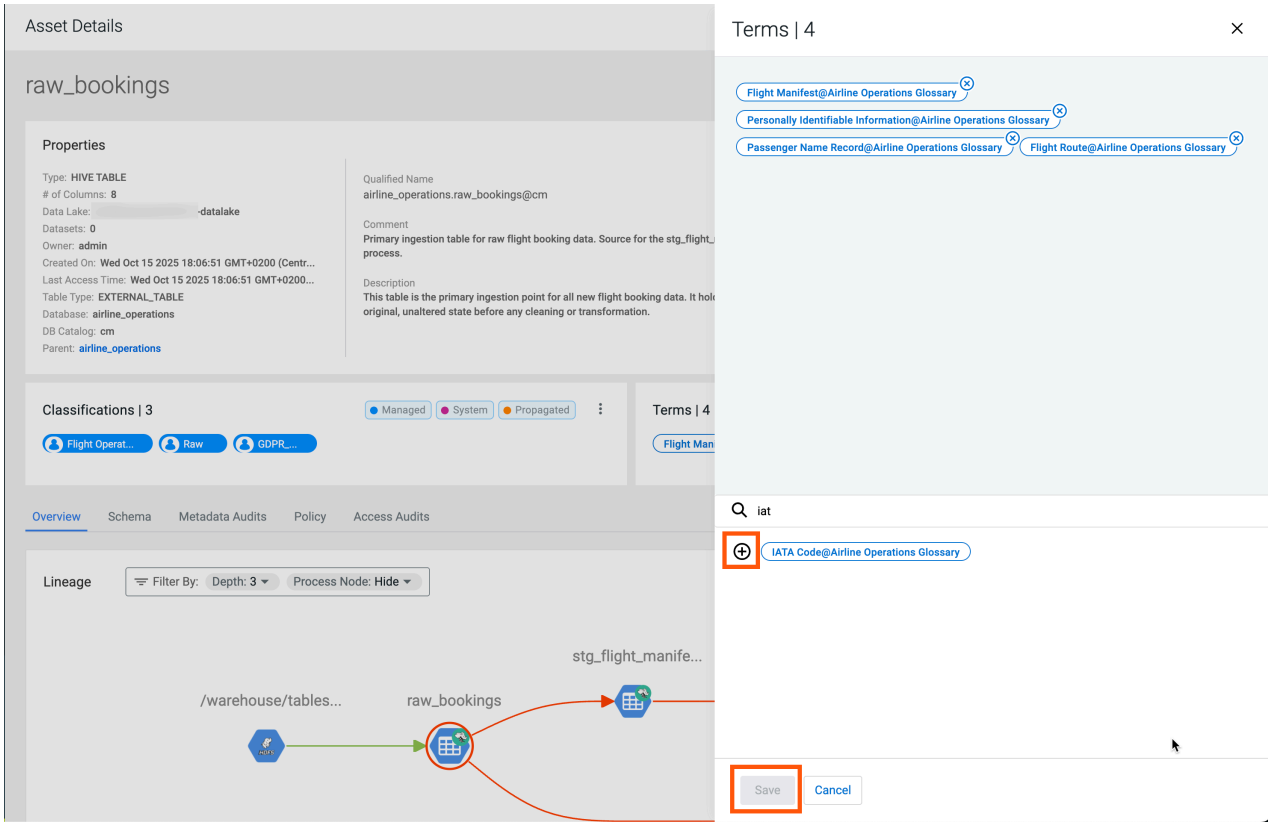
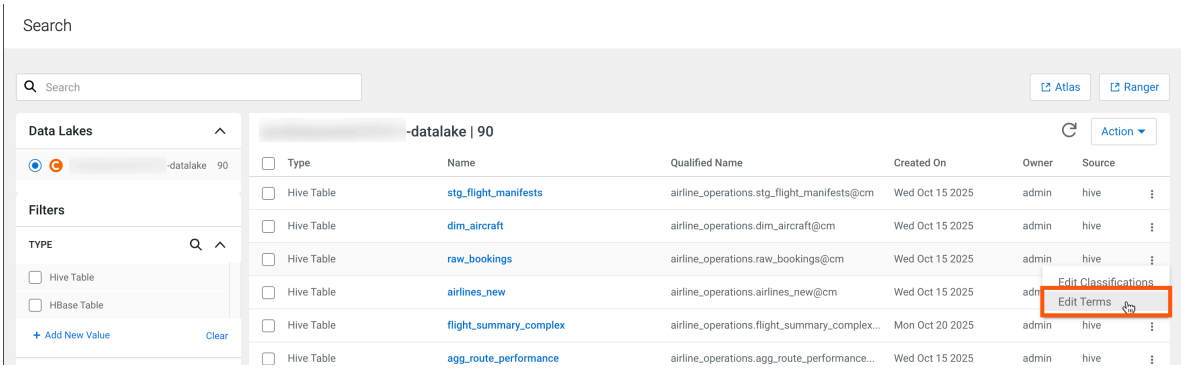


Figure 1: Editing glossary terms of an asset



Note: You can also access the **Terms** pane from **Search**.



Related Information

Glossaries overview
Creating glossaries
Creating terms
Searching for assets using Atlas glossaries
Search filters

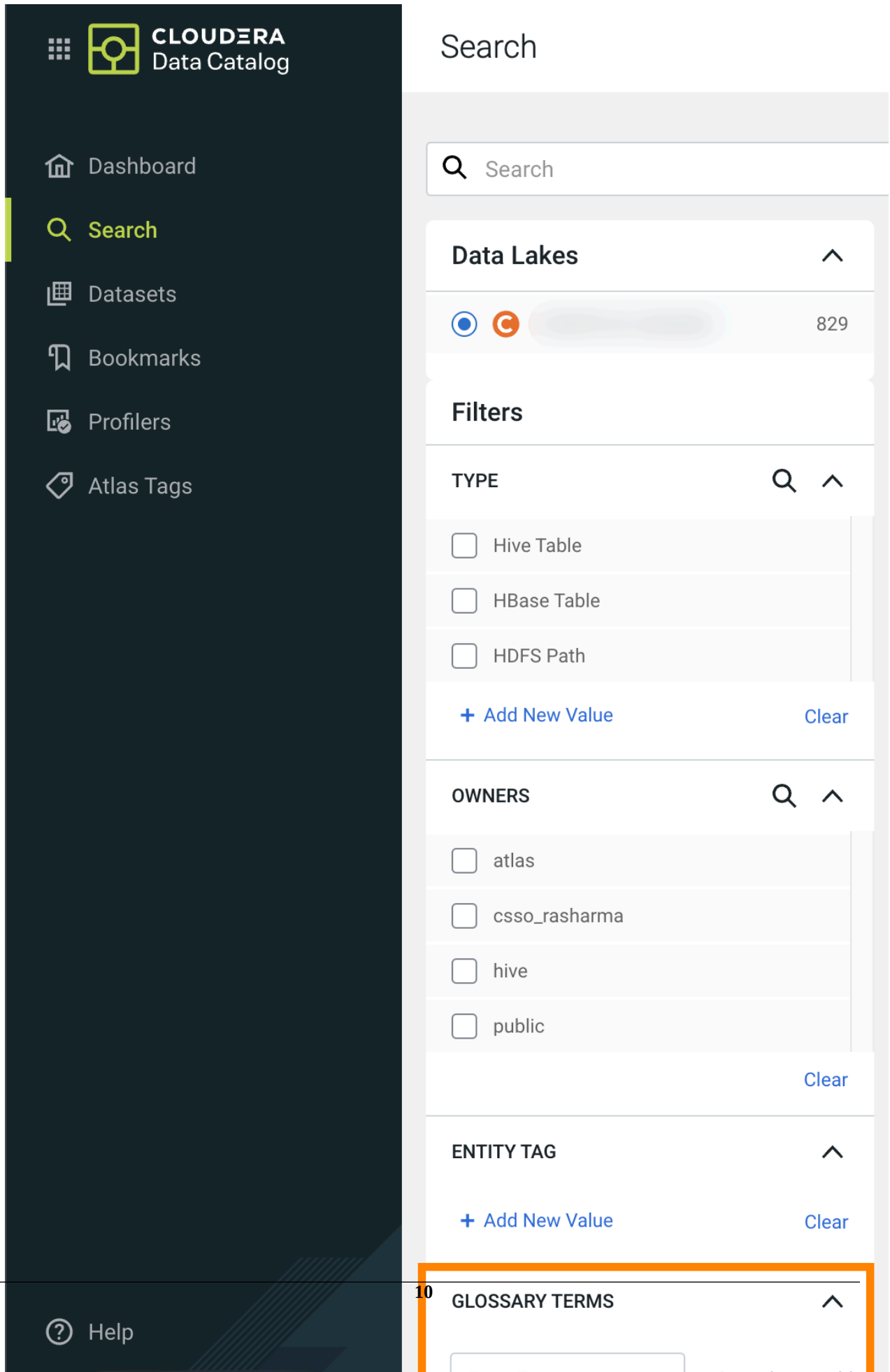
Searching for assets using glossary terms

You can search for the datasets using the Glossary Terms filter available on the Search page.

Note: The option for searching based on Glossary terms appears only if there are terms available in Apache Atlas.

1. Go to **Search**.

2. Scroll down to **GLOSSARY TERMS**.
3. Click + Add Term, and select the term to be used for your search.


CLOUDERA
Data Catalog
Search
Dashboard
Search
Datasets
Bookmarks
Profilers
Atlas Tags
Help
Search
Data Lakes
829
Filters
TYPE
Hive Table
HBase Table
HDFS Path
+ Add New Value
Clear
OWNERS
atlas
csso_rasharma
hive
public
Clear
ENTITY TAG
+ Add New Value
Clear
GLOSSARY TERMS

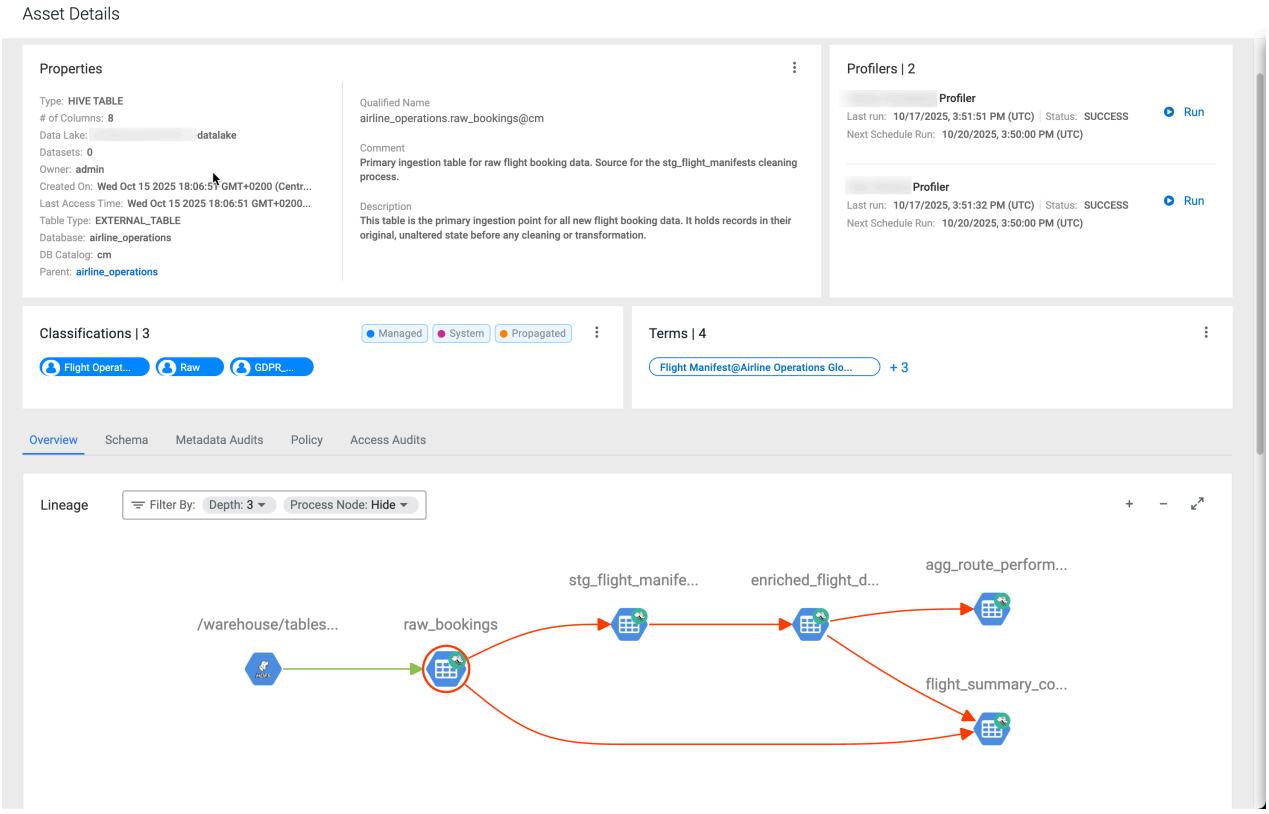
Related Information

Glossaries overview
Creating glossaries
Creating terms
Searching for assets using Atlas glossaries
Search filters

Viewing Data Asset details

Use the Asset Details menu to learn about the specifics of your asset including its relation to other asset, metadata and access updates related to it or Apache Ranger policies affecting it.

To access the **Asset Details** menu, click an asset in the **Search** menu. This brings you to the **Overview** tab.



- **Properties**:
  - **Type**
  - **Number of columns**
  - **Data Lake**
  - **Owner**
  - **Created On** date
  - **Modified Time**
  - **Namespace GUID** (HBase entity specific)
  - **Last Access Time**
  - **Table Type**
  - **DB Catalog**
  - **Parent**
  - **Qualified Name**
  - **Comment**

    Note: This field is available for Hive tables and columns and Iceberg tables and columns.

  - **Description**
- **Profilers**: The list of available on-demand profilers and details of the last run.
- **Classifications** and **Terms**: The list of assigned metadata.
- **Overview**: Shows the chain of custody for the data from relevant metadata repositories such as Apache Atlas. Lineage overview shows both upstream paths (lineage) into and downstream paths (impact) out of a given asset.
- **Schema**: Displays the schema of the data asset for structured data (such as Hive tables) from the relevant metadata repositories (such as Atlas).
- **Metadata Audits**: Shows the list of metadata changes.
- **Policy**: Shows security (authorization) policies defined on assets such as those present in Apache Ranger. It includes both resource (physical asset based) as well as classification based policies. This helps you understand how data access is secured and protected: what users can see what data (or metadata) under what conditions (security policies, data protection, and anonymization).
- **Access Audits**: Shows the most recent access audits from Apache Ranger.

Supported Data Asset fields for entities

The Asset Details menu displays all the Apache Atlas metadata associated with a particular data asset.

Supported fields

The following matrix captures the supported fields for different asset types:

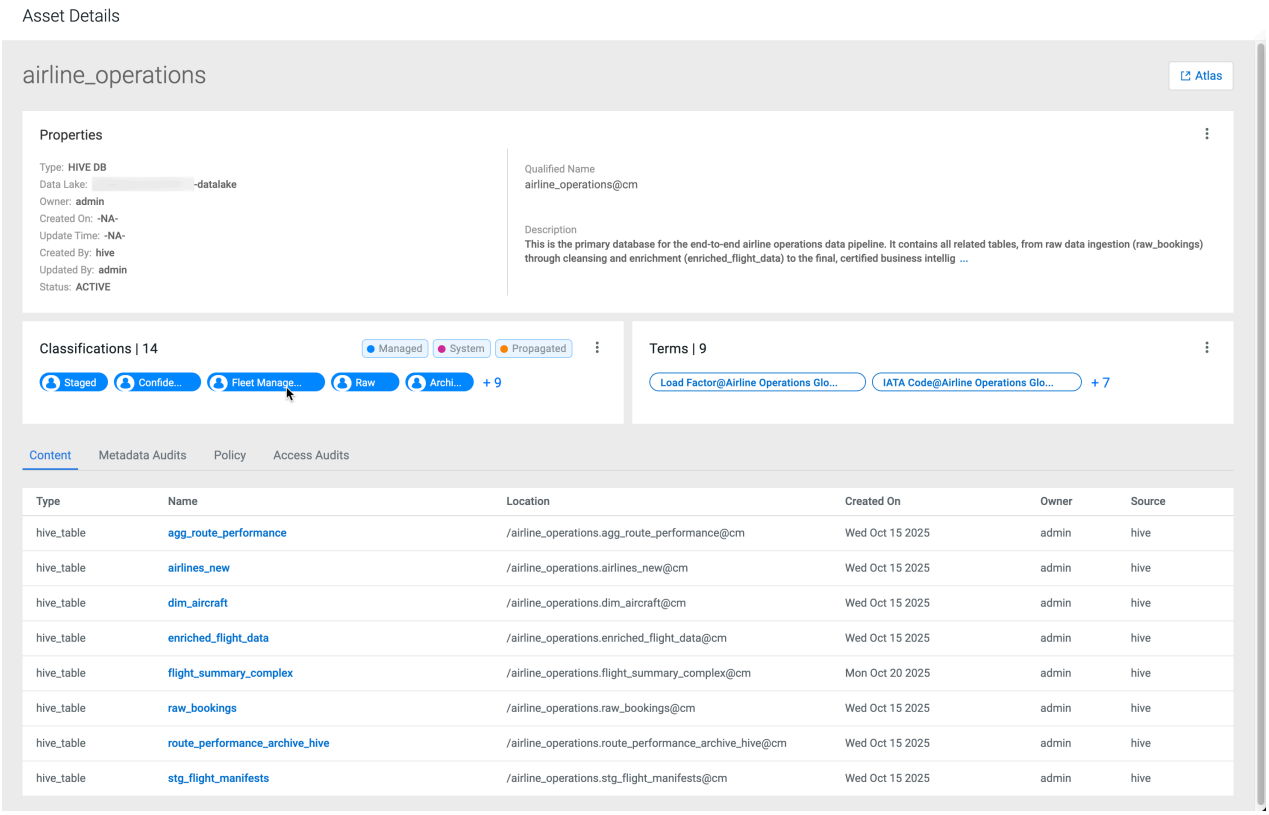
Asset Type	Lineage	Tagging	Access Metrics	Schema	Policy	Metadata and Access Audits	Atlas Punch out
Hive DB	Not Supported	Yes	Not Supported	Not Supported	Yes	Yes	Yes
Hive Table	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hive Column	Yes	Yes	Not Supported	Not Supported	Yes	Yes	Yes
Hbase Namespace	Yes	Yes	Not Supported	Not Supported	Yes	Yes	Yes
Hbase Table	Yes	Yes	Not Supported	Yes	Yes	Yes	Yes

Asset Type	Lineage	Tagging	Access Metrics	Schema	Policy	Metadata and Access Audits	Atlas Punch out
Hbase Column Family	Yes	Yes	Not Supported	Not Supported	Yes	Yes	Yes
impala_process	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
impala_column_lineage	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
impala_process_execution	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
ML_Project	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
ML_Model_Build	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
ML_Model_Deploy	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
rdbms_db	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
rdbms_column	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
rdbms_foreign_key	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
rdbms_index	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
rdbms_instance	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
rdbms_table	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_process	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_application	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_column	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_column_lineage	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_db	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_ml_directory	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_ml_model	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_ml_pipeline	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_process_execution	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
spark_table	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes

Navigation in Asset Details

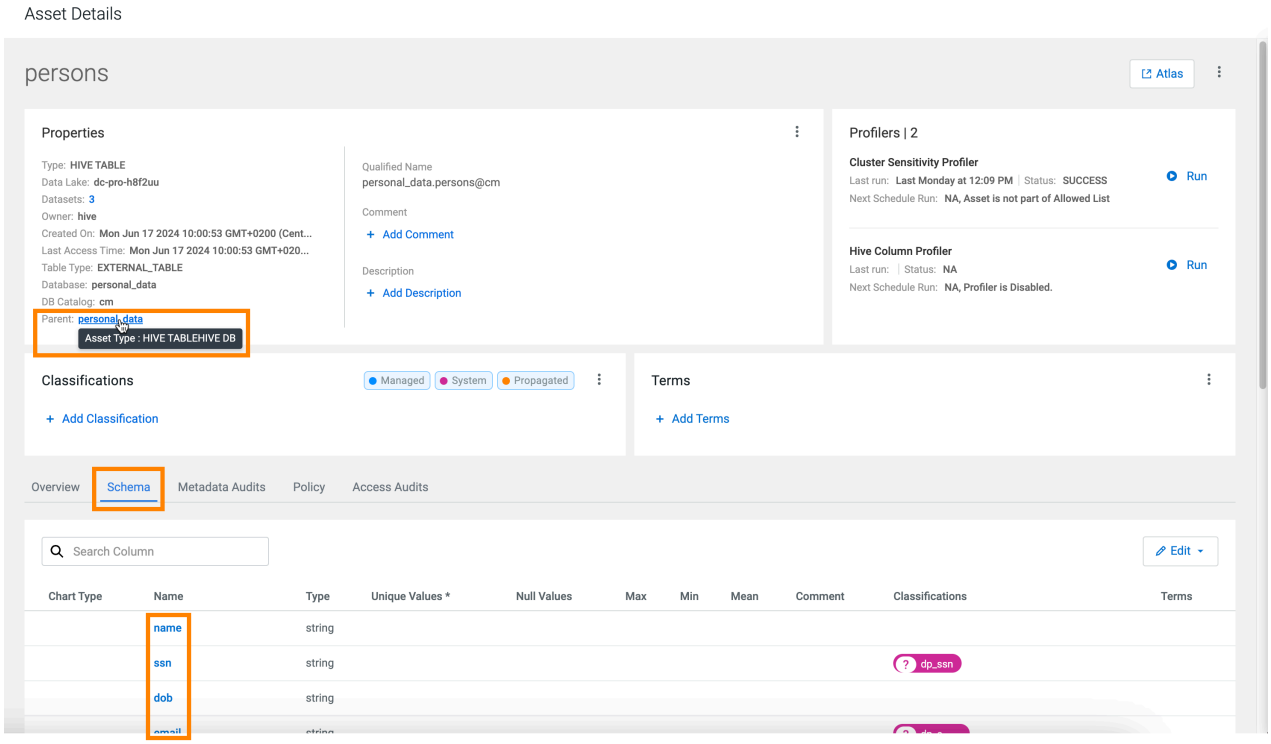
A generic Assets Details page is available for container data types like buckets and databases.

The **Content** tab (similar to the **Schema** tab) lists all the contents of the selected entity. Clicking on any element available in the **Content** tab navigates you to the **Asset Details** page.



For example, for a database entity having a list of tables, clicking on any listed table navigates to the **Asset Details** page of the same table. This page helps you understand the parent-child relationship between related assets. The **Contents** tab displays entities that are contained within assets of container entity types. The entities in the table of **Contents** tab are clickable, which will allow you to navigate to the **Asset Details** page of these contained assets.

In these contained assets, the Parent row in the **Properties** panel allows you to go one level up in the hierarchy to the container element. In certain elements, such as Hive tables, the **Schema** tab contains further elements whose name can be clicked to open individual contained elements.



The following table lists the entity types, their parent, and contents.

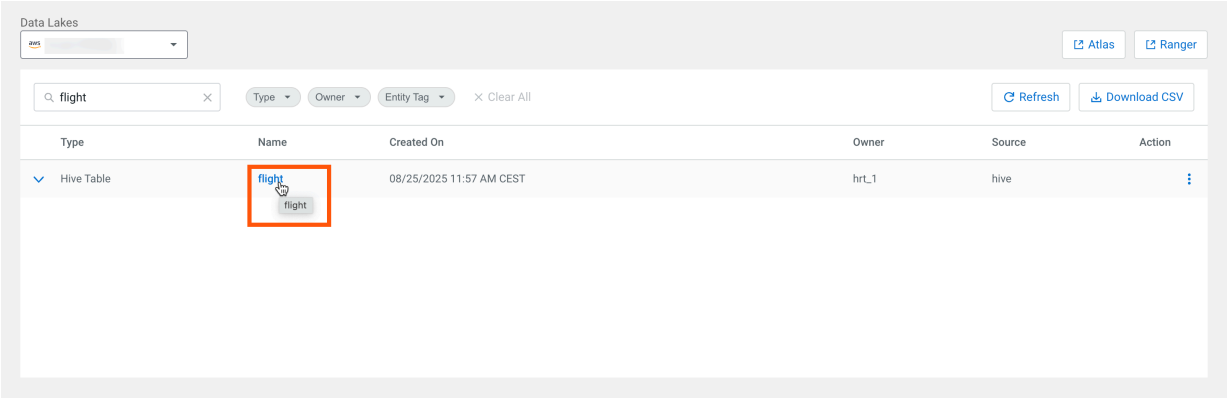
Type	Parent	Content
Hive DB	-	Hive Table
Hive Table	Hive DB	Hive column
HBase Namespace	-	HBase Table
ML Project	-	ML Model Build
ML Model Build	ML Project	ML Model Deployment
RDBMS Instance	-	RDBMS DB
RDBMS DB	RDBMS Instance	RDBMS Table

Creating a classification from Asset Details

You can create and add Atlas classifications in in multiple places to label data assets, which simplifies data discovery and helps in consistently applying governance policies like security and access control in Apache Ranger.

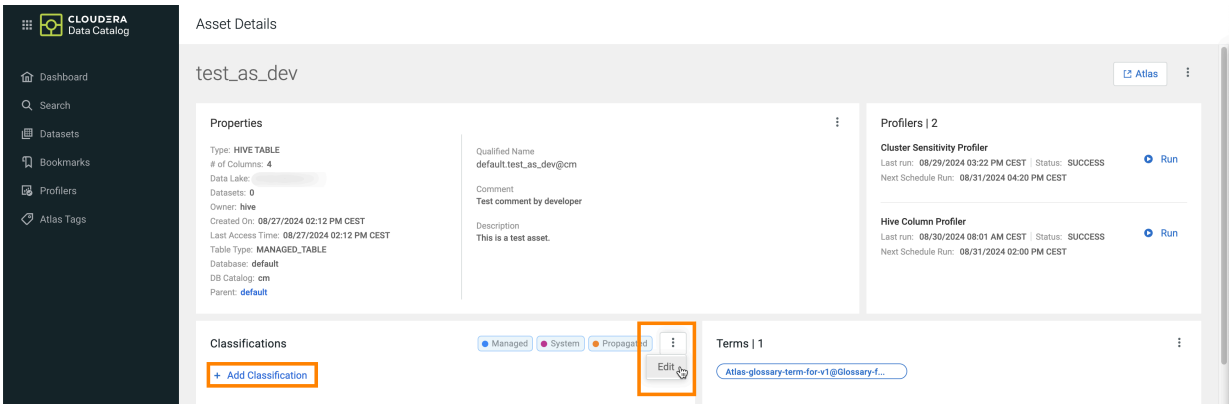
Procedure

1. Go to the **Asset Details** page of an asset.



2. Click Add Classification or ⋮ icon by an asset, then select Edit.

Figure 2: Adding Atlas classifications



3. Search for a previously created classification or create a new one.

Figure 3: Creating a new Atlas classification from Cloudera Data Catalog

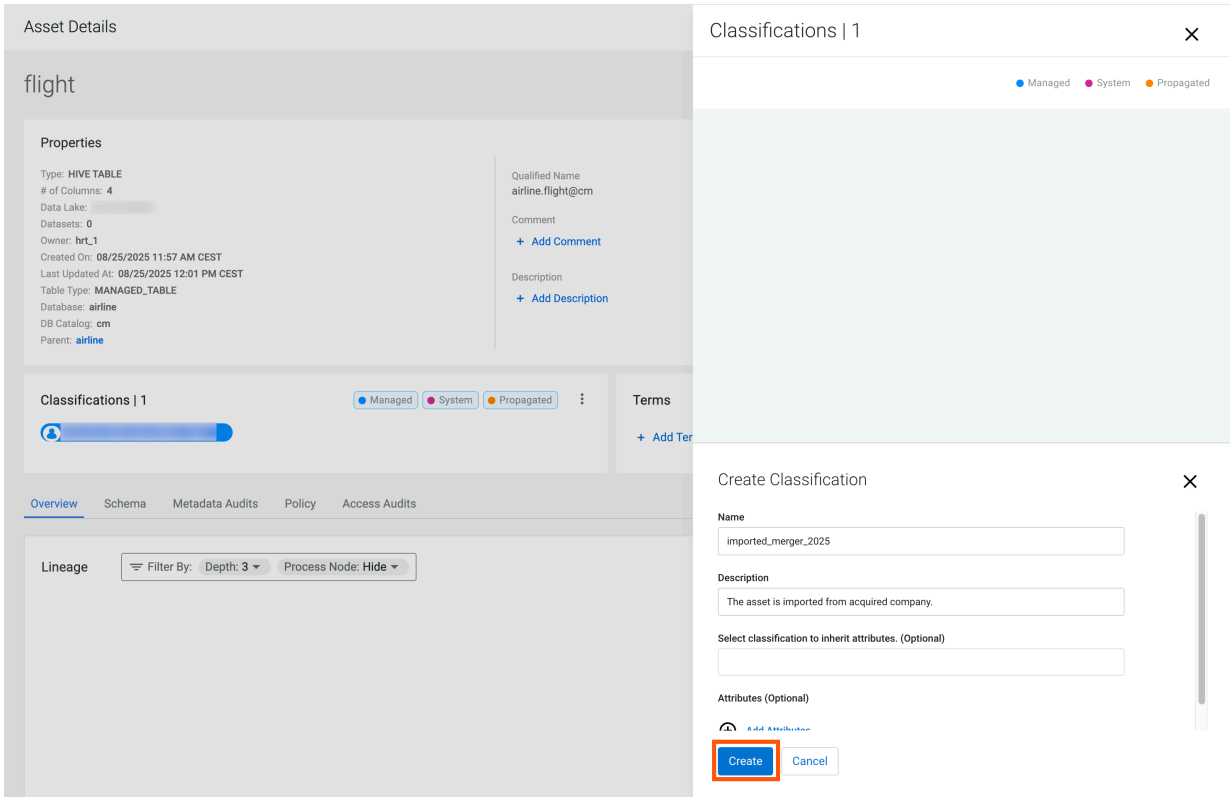
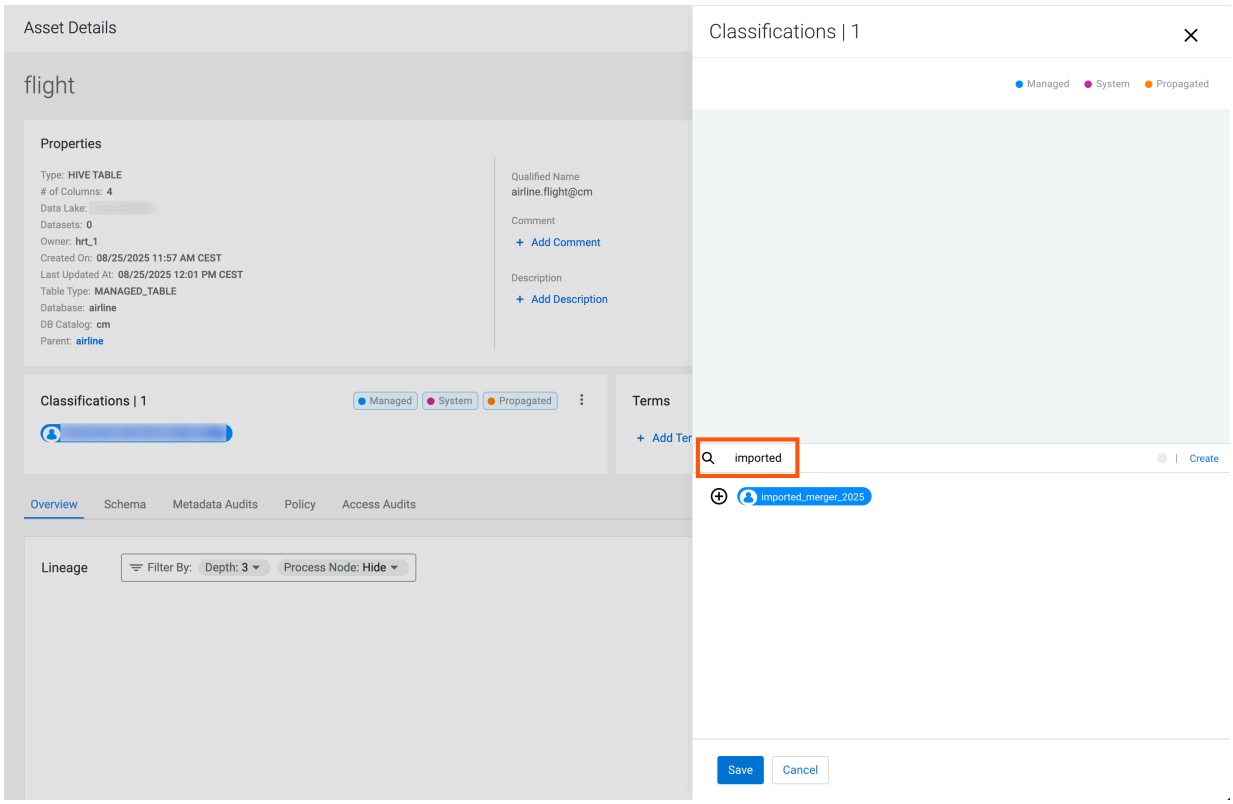


Figure 4: Searching for an Atlas classification



4. Click Save to finalize your changes.

Related Information

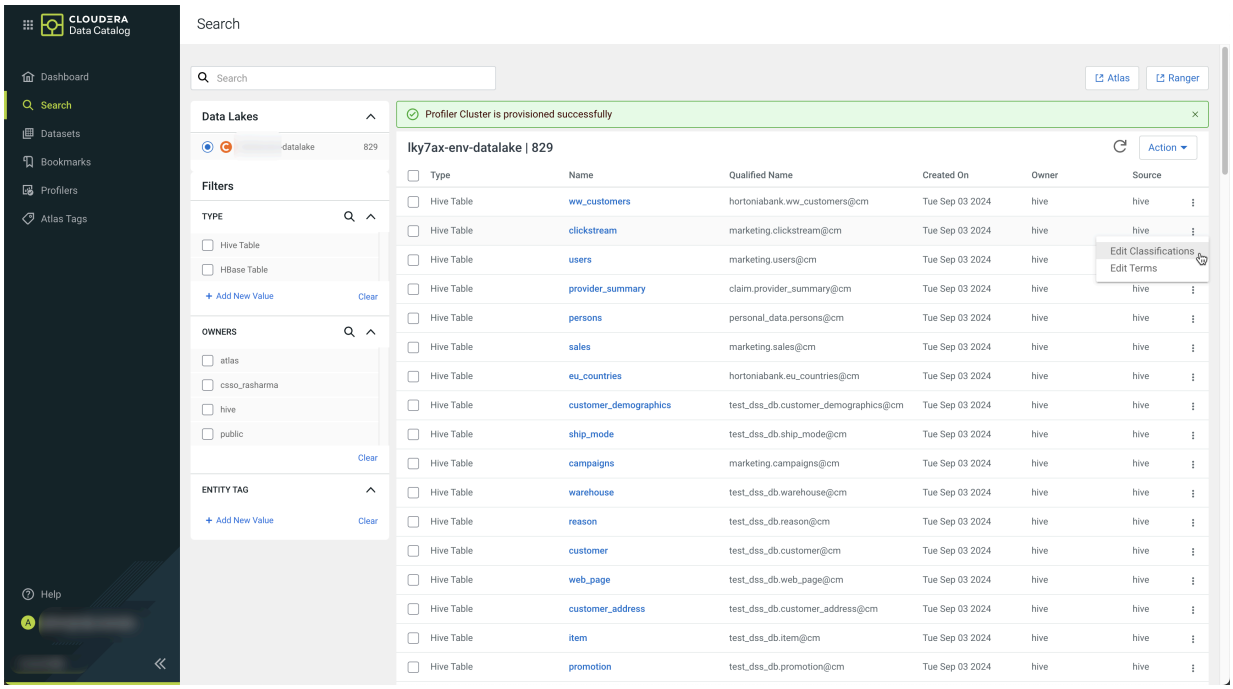
Apache Atlas metadata attributes
Working with Atlas classifications and labels
Creating classifications
Tag-based Services and Policies
Atlas classifications drive Ranger policies
Atlas tag management
Creating a classification from Search page

Creating a classification from Search page

You can create and add Atlas classifications in in multiple places to label data assets, which simplifies data discovery and helps in consistently applying governance policies like security and access control in Apache Ranger.

Procedure

1. Go to the **Search** page.
2. Click the ⋮ icon by an asset, then select Edit Classifications.



3. Search for a previously created classification or create a new one.
4. Click Save to finalize your changes.

Related Information

Apache Atlas metadata attributes
Working with Atlas classifications and labels
Creating classifications
Tag-based Services and Policies
Atlas classifications drive Ranger policies
Atlas tag management
Creating a classification from Asset Details

Additional search options for asset types

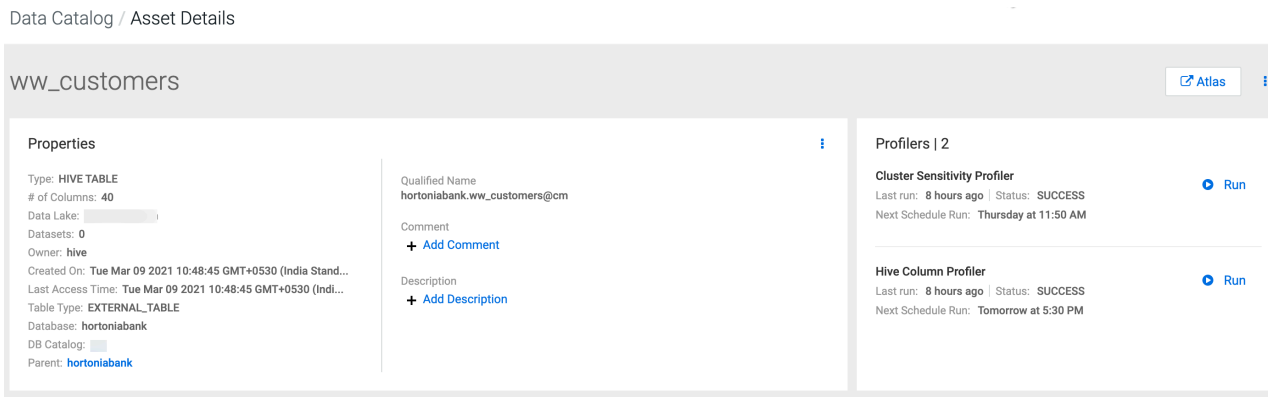
Using Cloudera Data Catalog, you can add or edit asset description values to search for data assets across both Cloudera Data Catalog and Apache Atlas services by using the asset content. These values can be searched.

Adding comments and descriptions to assets

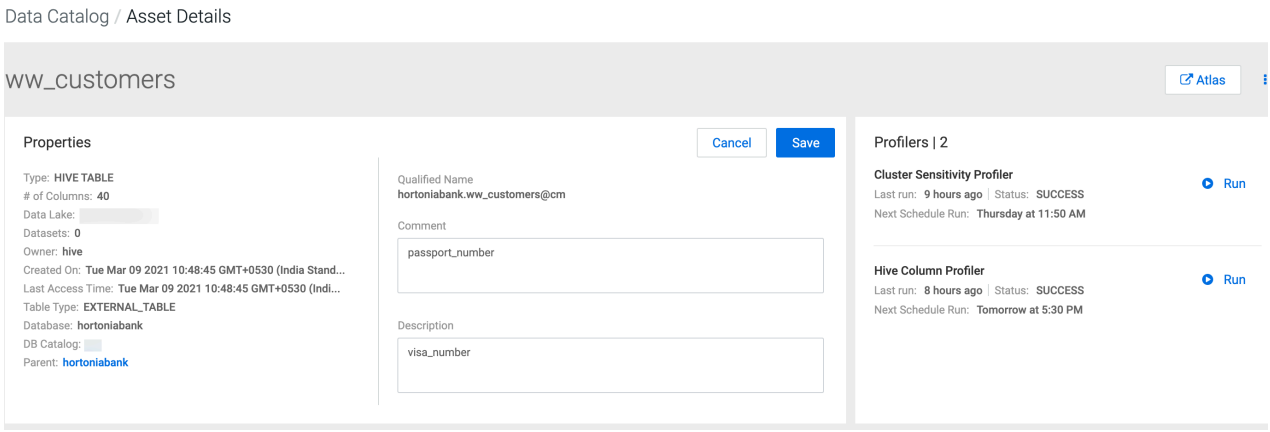
In the **Asset Details** page for each asset type that you select, you can add or edit **comment** or **description** fields. Including these values for the selected asset helps you to identify your chosen asset.

Using the same set of values (comment or description), you can also search for the asset types in Atlas.

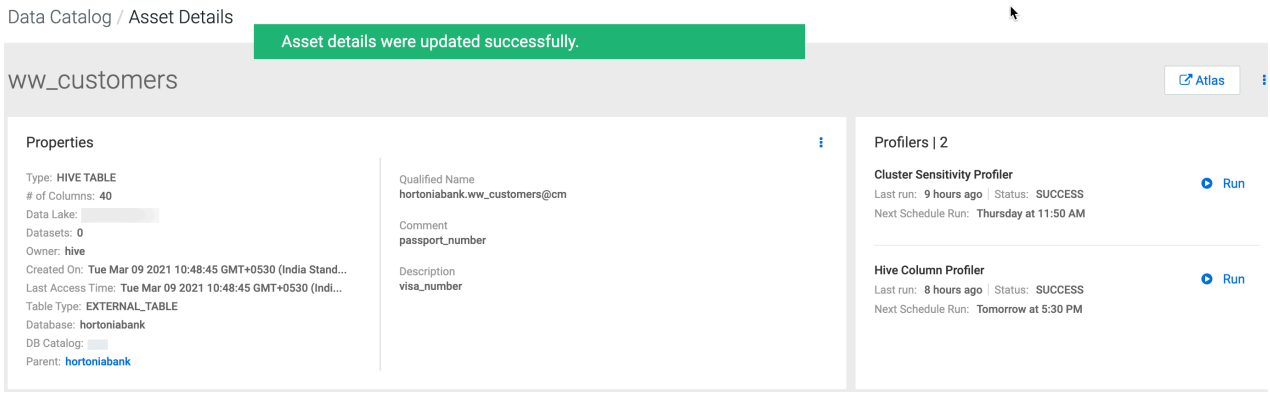
Note: The comment and description options are supported only for Hive table and Hive Column assets. For other asset types, only the description option is supported.



Click + Add Comment or + Add Description fields to include the respective values.

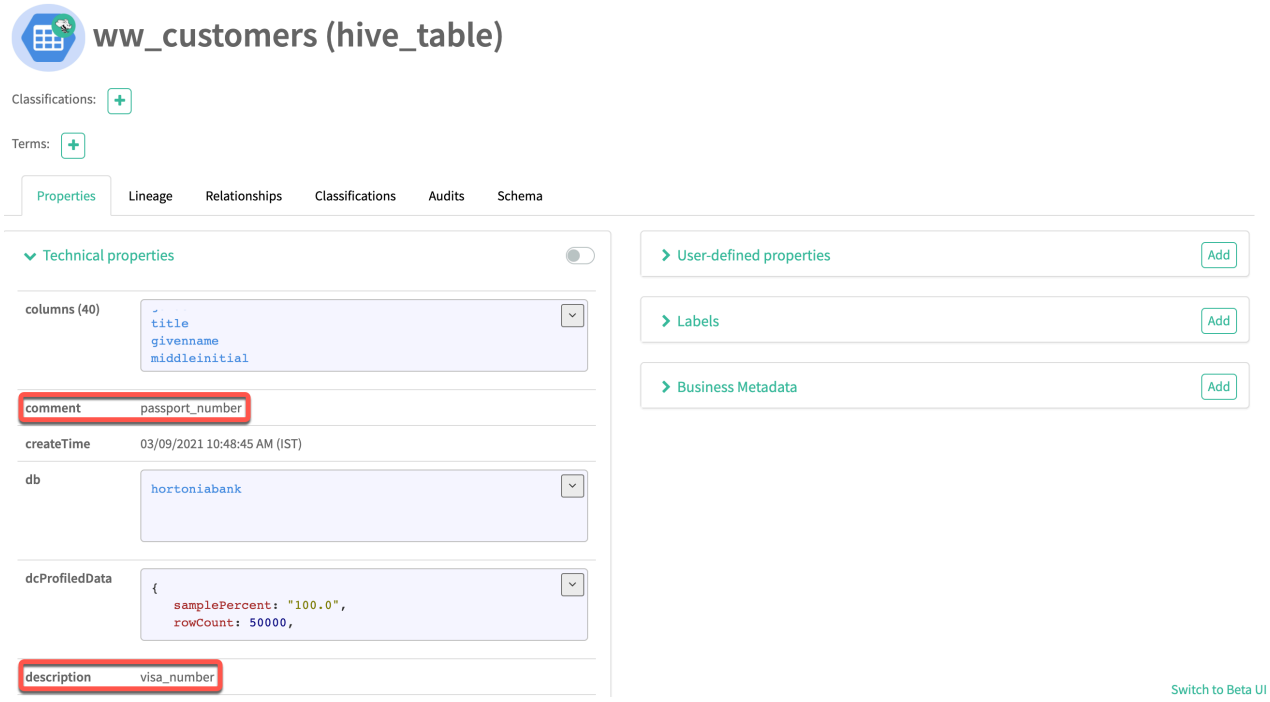


Click Save to save your changes.



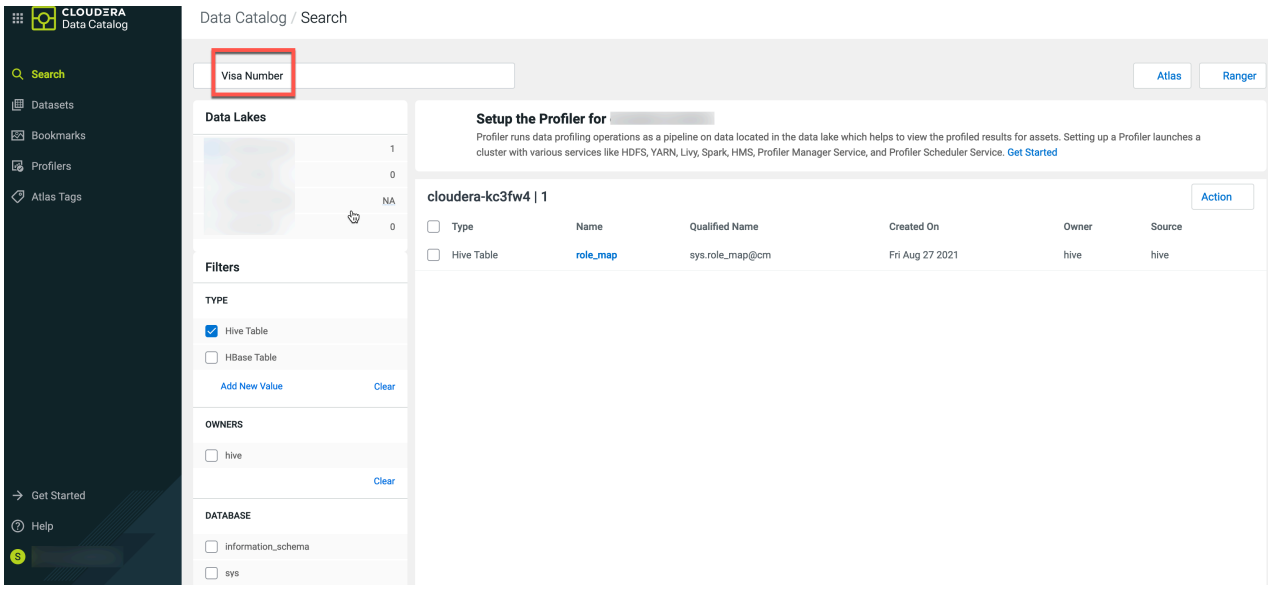
Note: You can also edit the already saved valued by clicking the icon.

Clicking on the Atlas button will navigate to the corresponding Atlas asset page as shown:

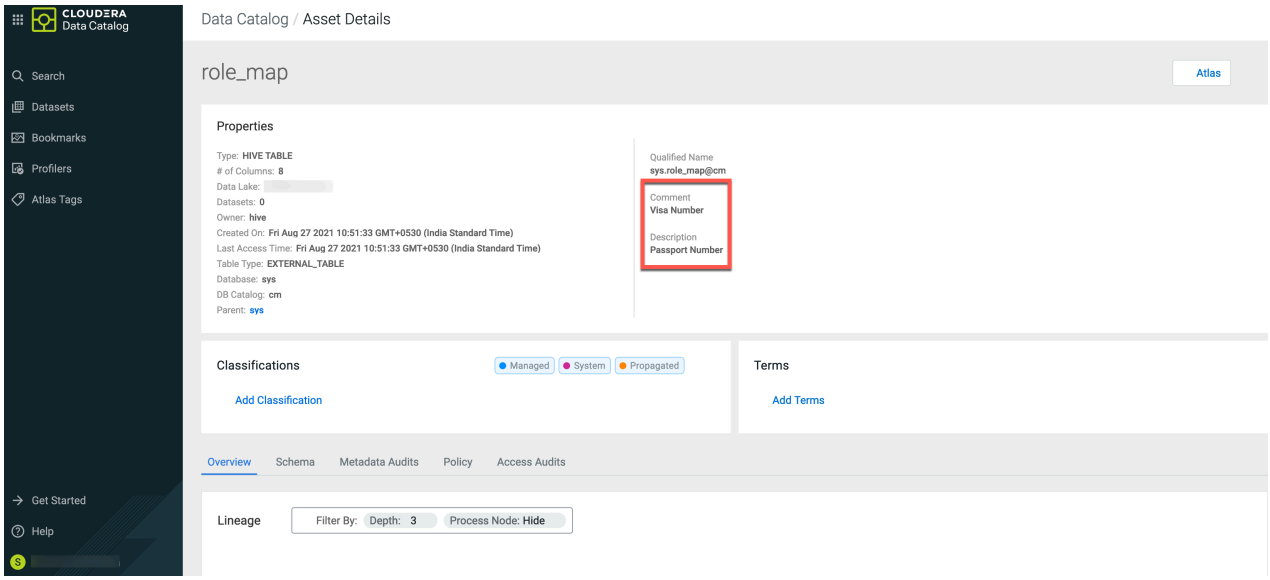


Searching comments and descriptions

The values of the **Comment** or **Description** fields can be searched in the **Search** menu. The result page displays the assets where you added your comments and descriptions without the use of filters.



Clicking on the asset type displays the comment and description values.



Mapping glossary terms

Cloudera Data Catalog synchronizes the glossary terms that are created in Apache Atlas. These can be seen in Search and Asset Details.

You can search for Atlas glossary terms in **Search** and map specific terms with assets in **Asset Details**. You can also search for terms to delete them from the selected asset. The selected asset displays the total number of terms associated or mapped accordingly.

When you map a specific term for your dataset, the term is displayed in the following format:

```
<termname>@glossaryname>
```

Figure 5: Glossary term in Search

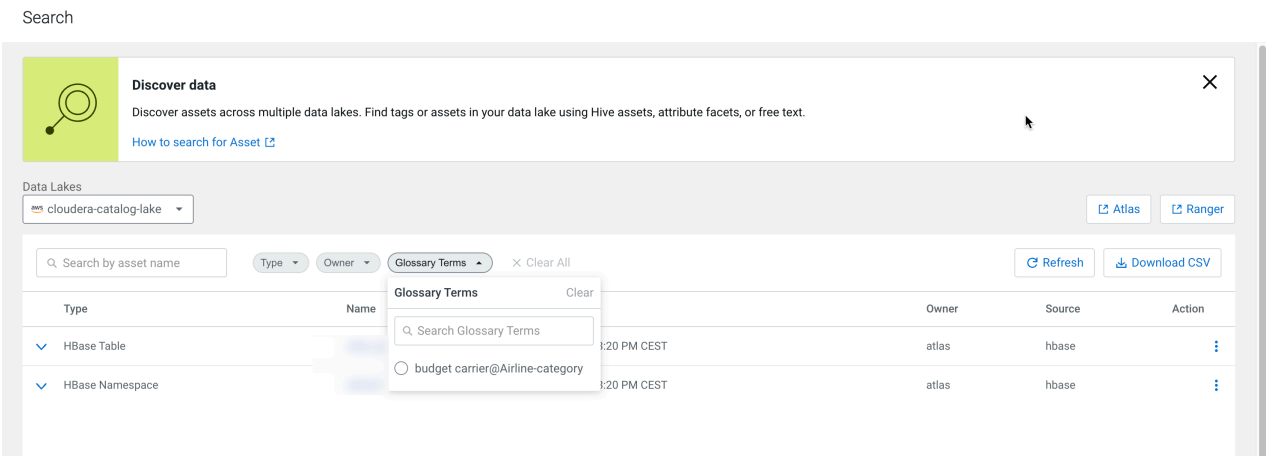
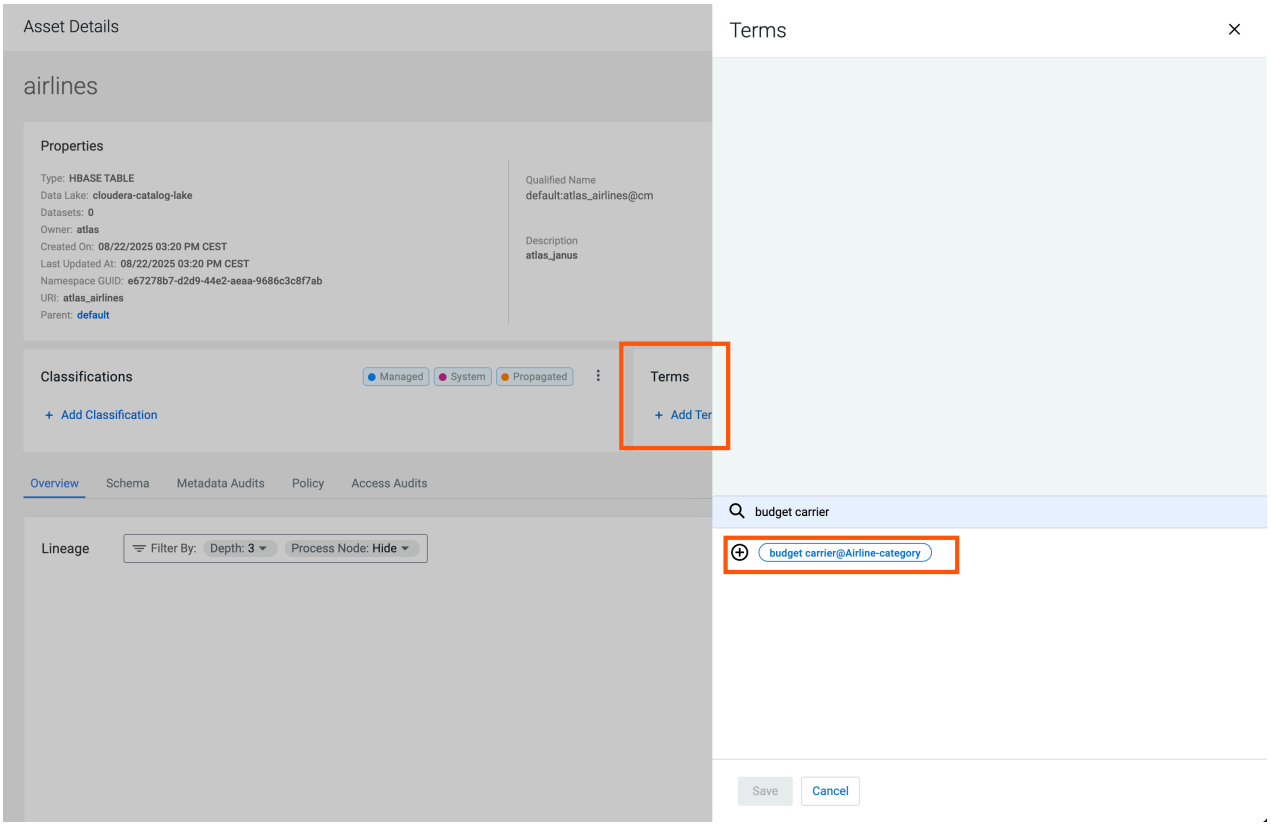


Figure 6: Glossary term in Asset Details



You can use the icon in the **Terms** widget on the **Asset Details** page to add new terms for your assets. Click Save to save the changes.

Figure 7: Mapping terms in Asset Details

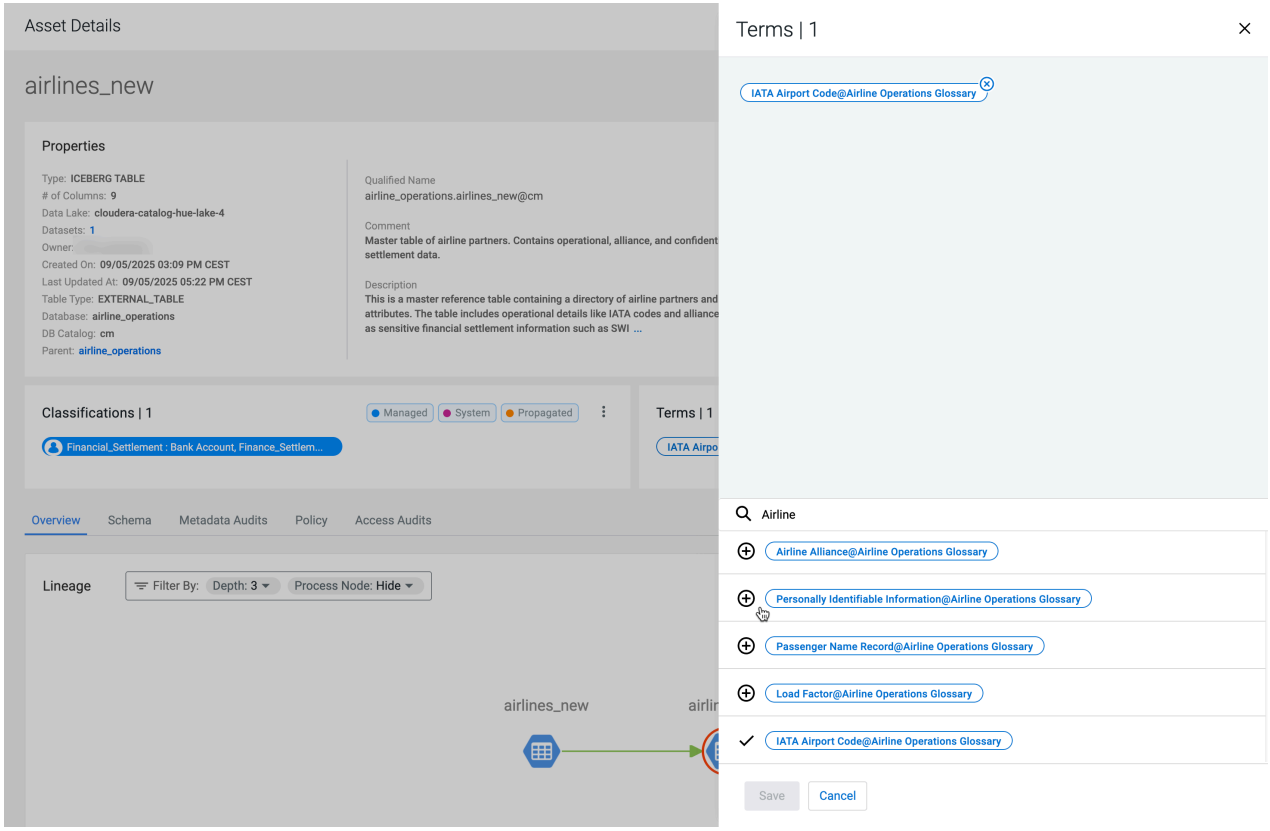
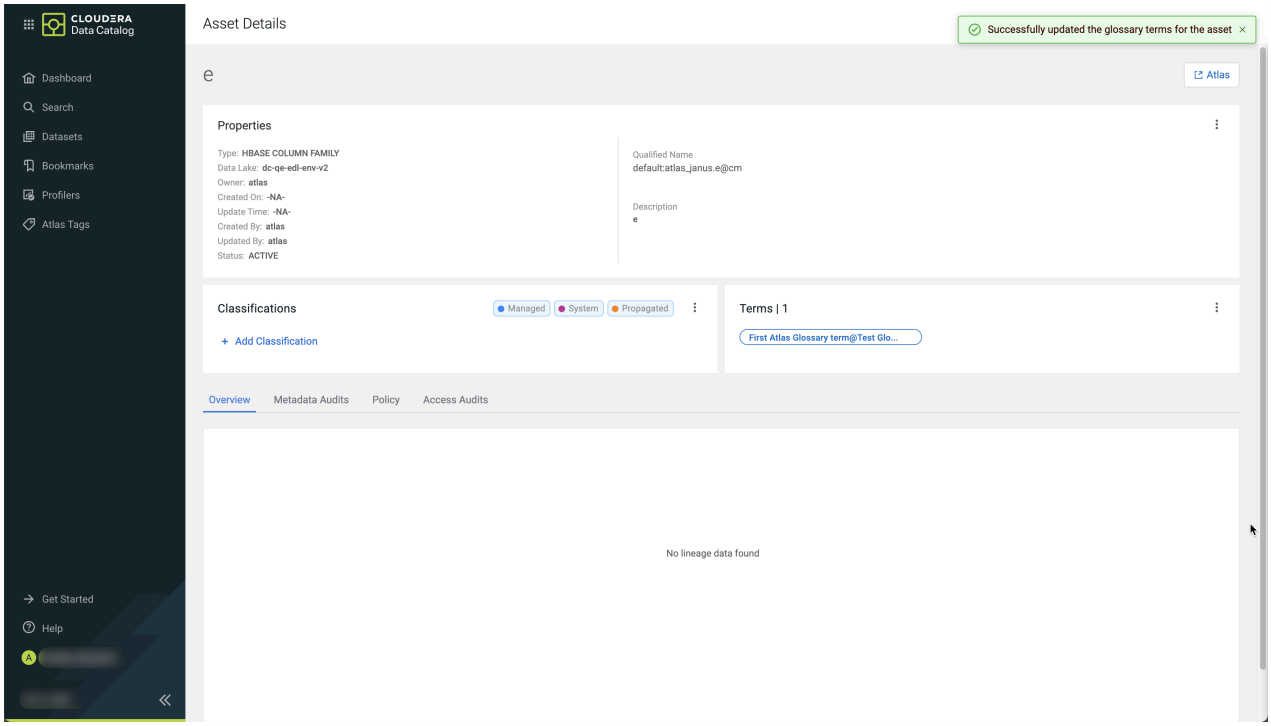
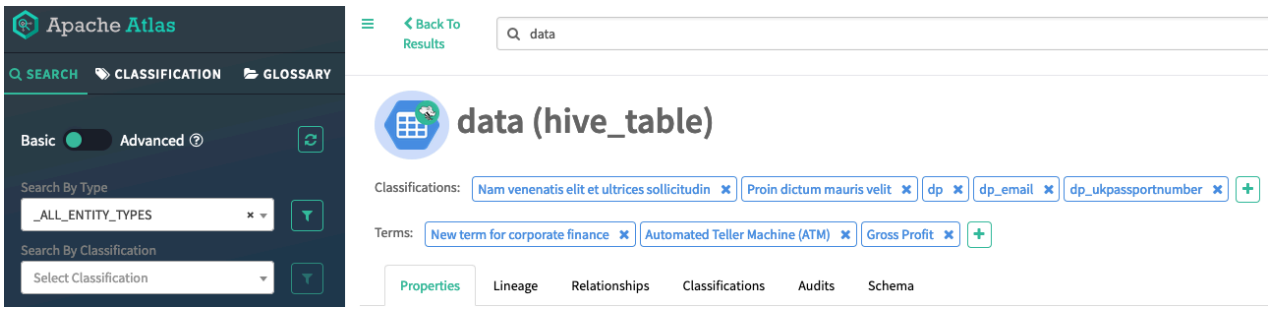


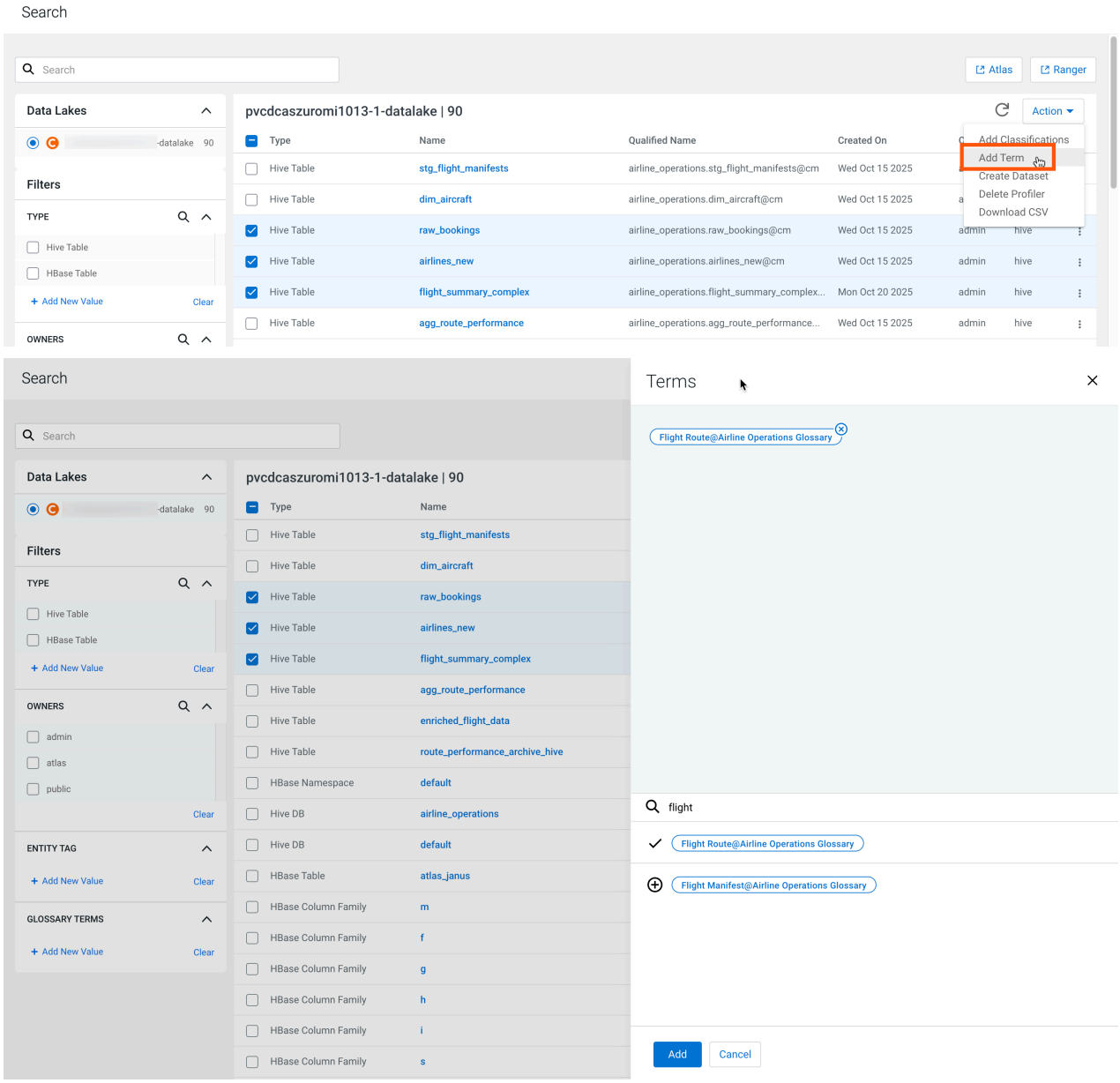
Figure 8: Mapped glossary term



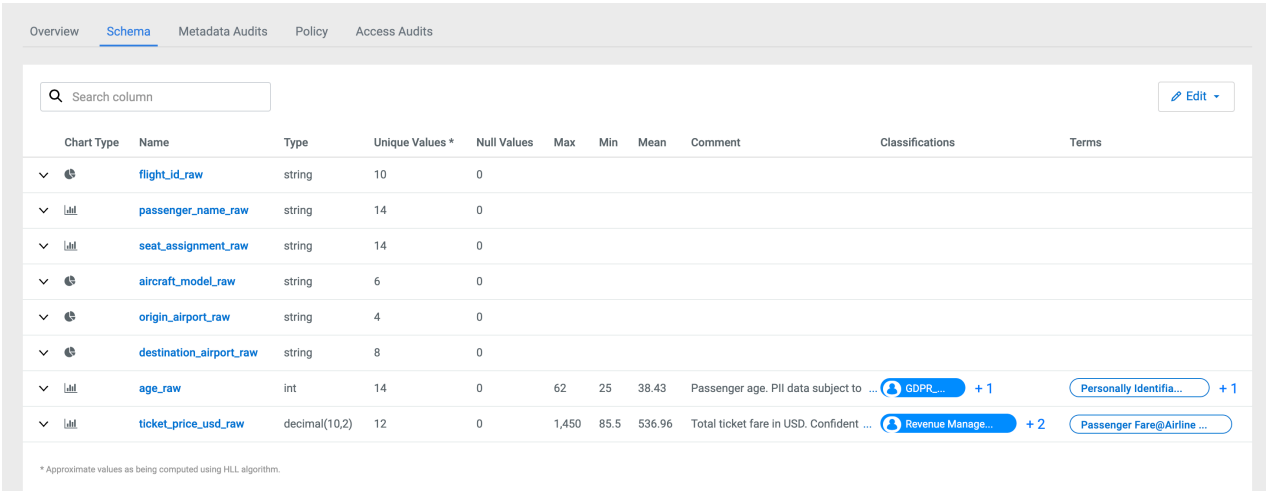
You can search for the same asset in the corresponding Atlas environment as shown in the example image.



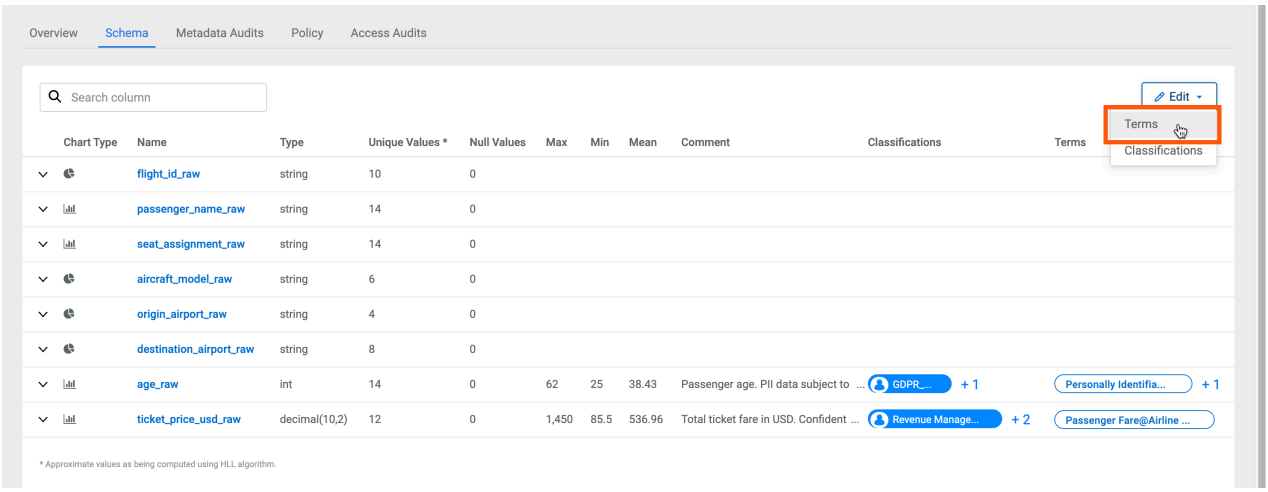
Additionally, you can also associate terms to your datasets by selecting one or more assets on the **Search** page. You can associate terms with multiple datasets at a time.



When you select a Hive table asset and navigate to the **Asset Details** page, under the **Schema** tab, you can view the list of terms associated with the asset.



You can add or update the terms for the associated datasets by clicking the Edit button.



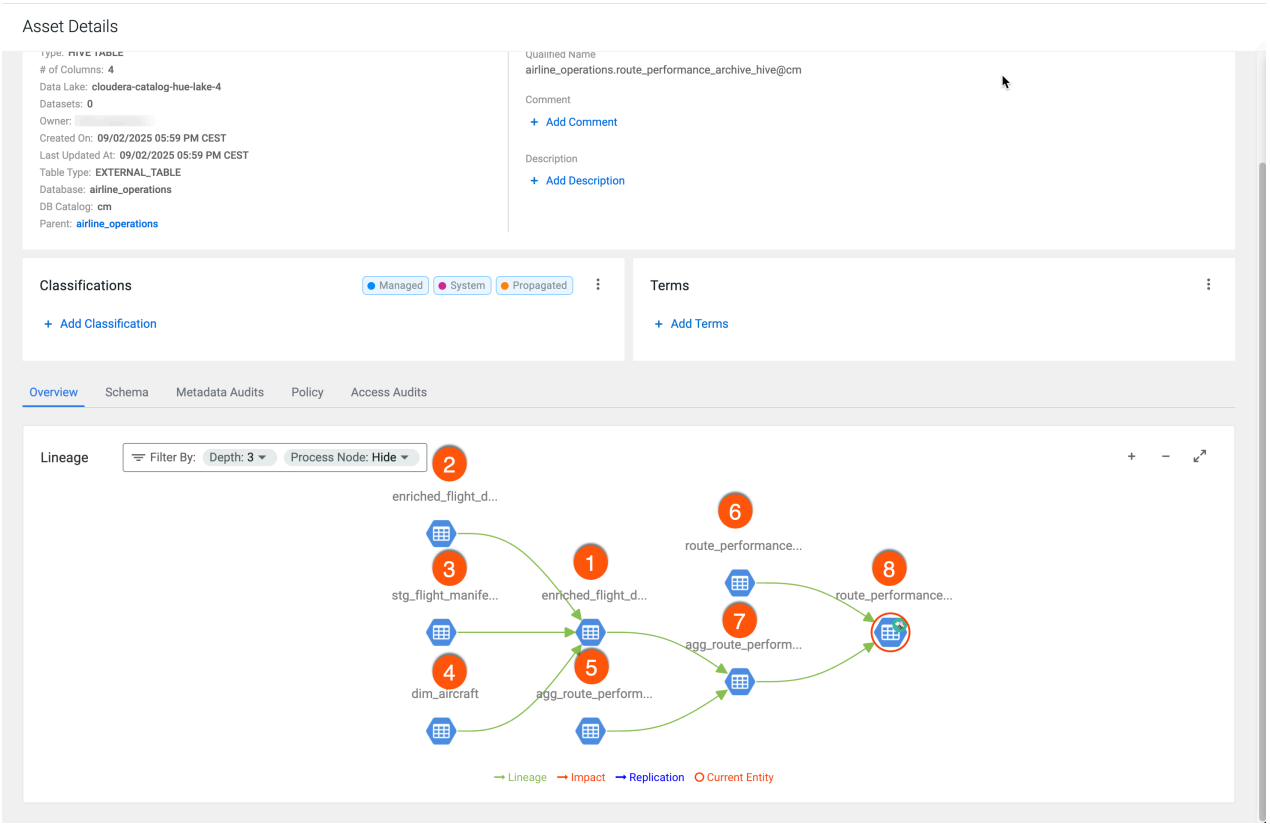
Related Information

Glossaries overview
Creating glossaries
Creating terms
Searching for assets using Atlas glossaries
Search filters

Navigation support for hive entities within Lineage

The Lineage in the Overview tab shows the chain of custody for the data from relevant metadata repositories such as Apache Atlas. Lineage overview shows both upstream paths (lineage) into and downstream paths (impact) out of a given asset.

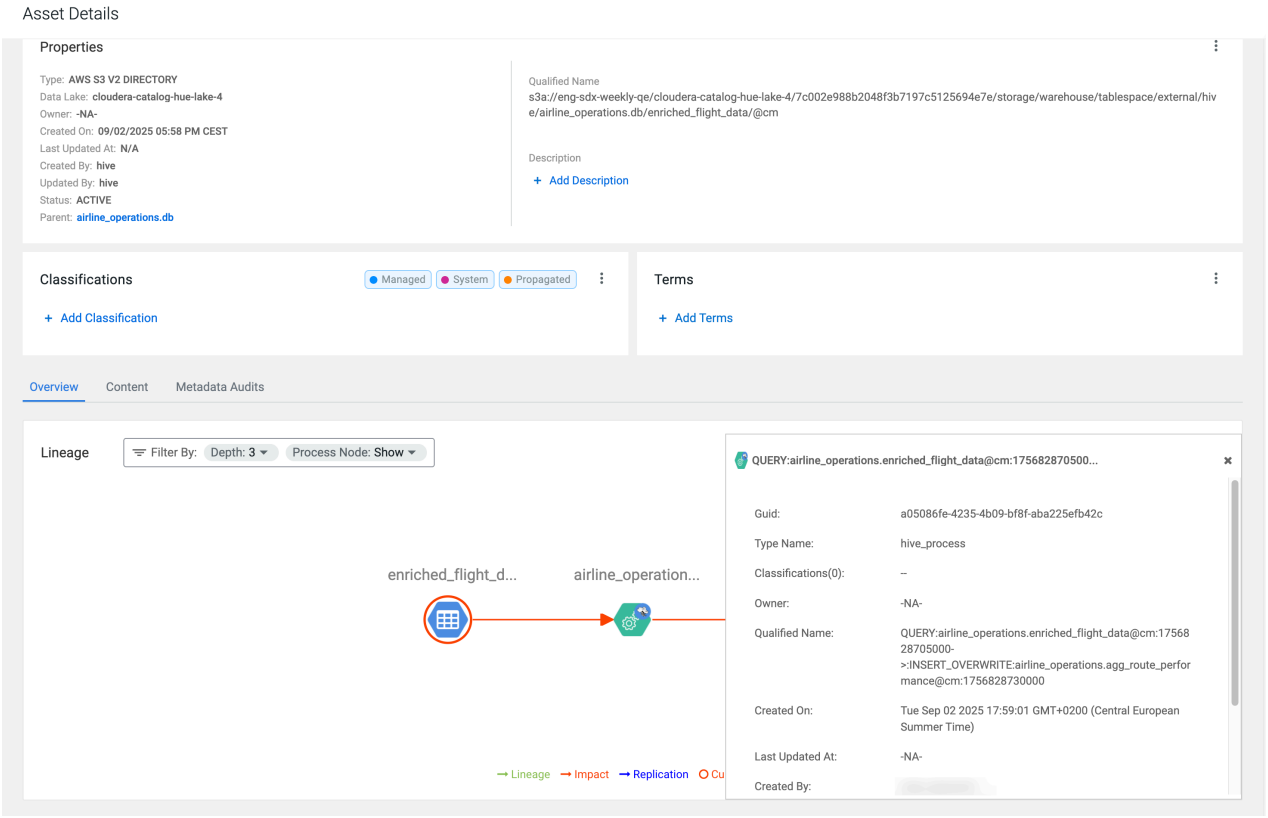
When you click a Hive entity within a lineage, the **Asset Details** page of the selected Hive entity is displayed. In the **Overview** tab, you can change the detail level of the lineage with the Depth drop-down and Show Process:



- The enriched_flight_data table (1) is created by joining the following source tables:
  - airline_operations.stg_flight_manifests@cm (3): This is the main source of transactional data.
  - airline_operations.dim_aircraft@cm (4): This is a dimension or lookup table. It provides supplemental information.
- s3a://eng-sdx-weekly-qe/cloudera-catalog-hue-lake-4/7c002e988b2048f3b7197c5125694e7e/storage/warehouse/tablespace/external/hive/airline_operations.db/enriched_flight_data/@cm (5) is the physical path for storing airline_operations.stg_flight_manifests@cm and airline_operations.dim_aircraft@cm.
- airline_operations.agg_route_performance@cm (7) is table created from aggregated some items from enriched_flight_data with an additional query.
- s3a://eng-sdx-weekly-qe/cloudera-catalog-hue-lake-4/7c002e988b2048f3b7197c5125694e7e/storage/warehouse/tablespace/external/hive/airline_operations.db/agg_route_performance/@cm (6) is the physical path for storing airline_operations.agg_route_performance@cm.
- airline_operations.route_performance_archive_hive@cm (8) is a direct copy of the aggregation table, intended for historical record-keeping in this example.

Alternatively, if you do not want to navigate away from the current page and want to view the information with respect to any entity, hover on the entity and click the information icon to view the details.

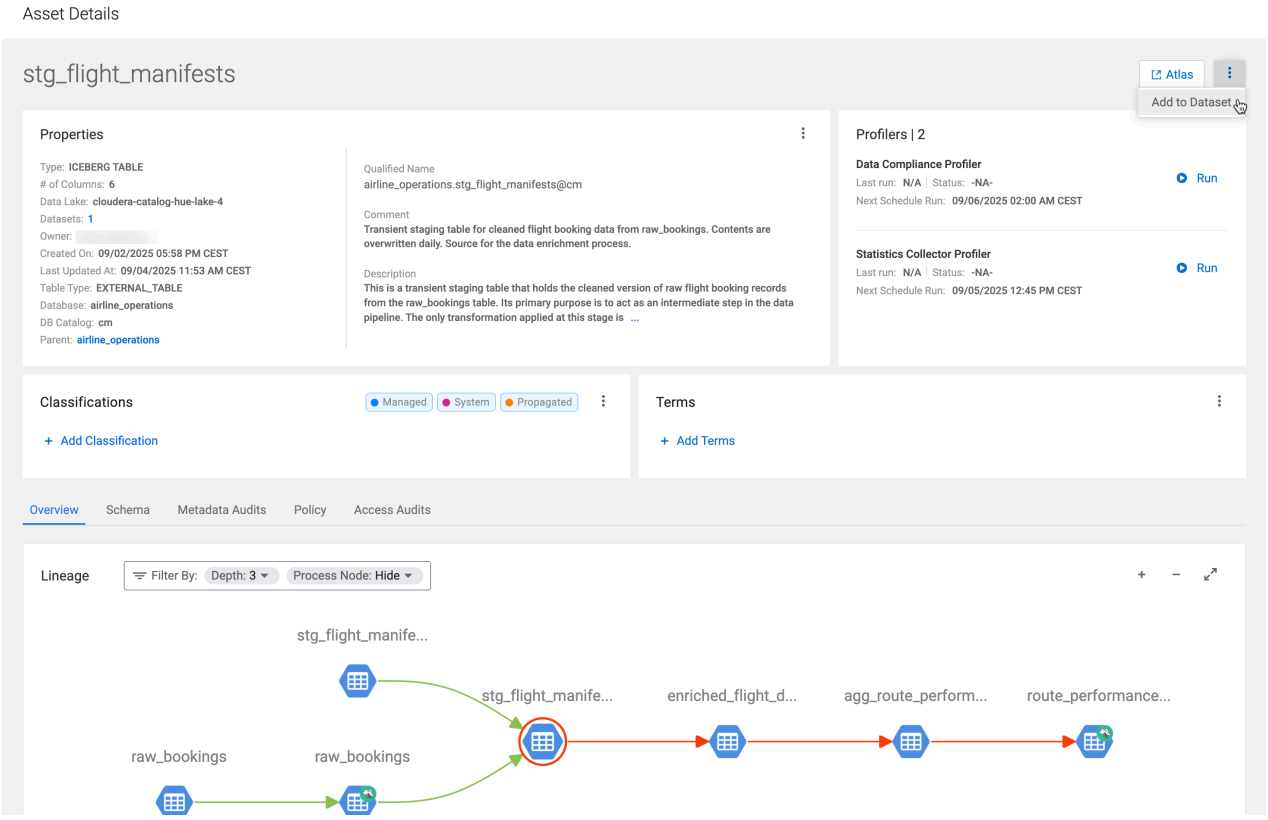
The screenshot depicts the slider information for the clicked entity:



This field shows the Hive operation writing data from airline_operations.enriched_flight_data@cm to the airline_operations.agg_route_performance@cm table.

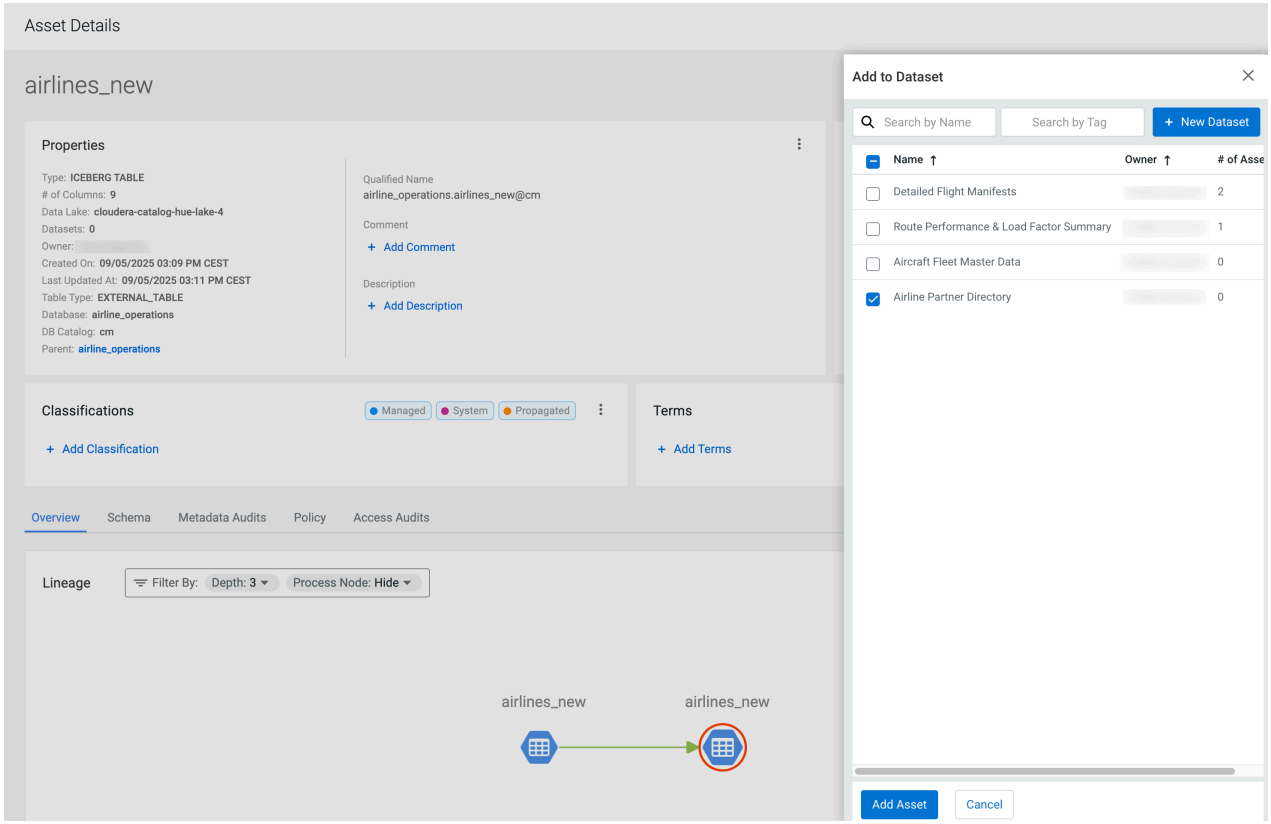
Adding assets to one or more datasets

On the Asset Details screen, users are provided with an option to add the asset to a dataset for easier management and searching.

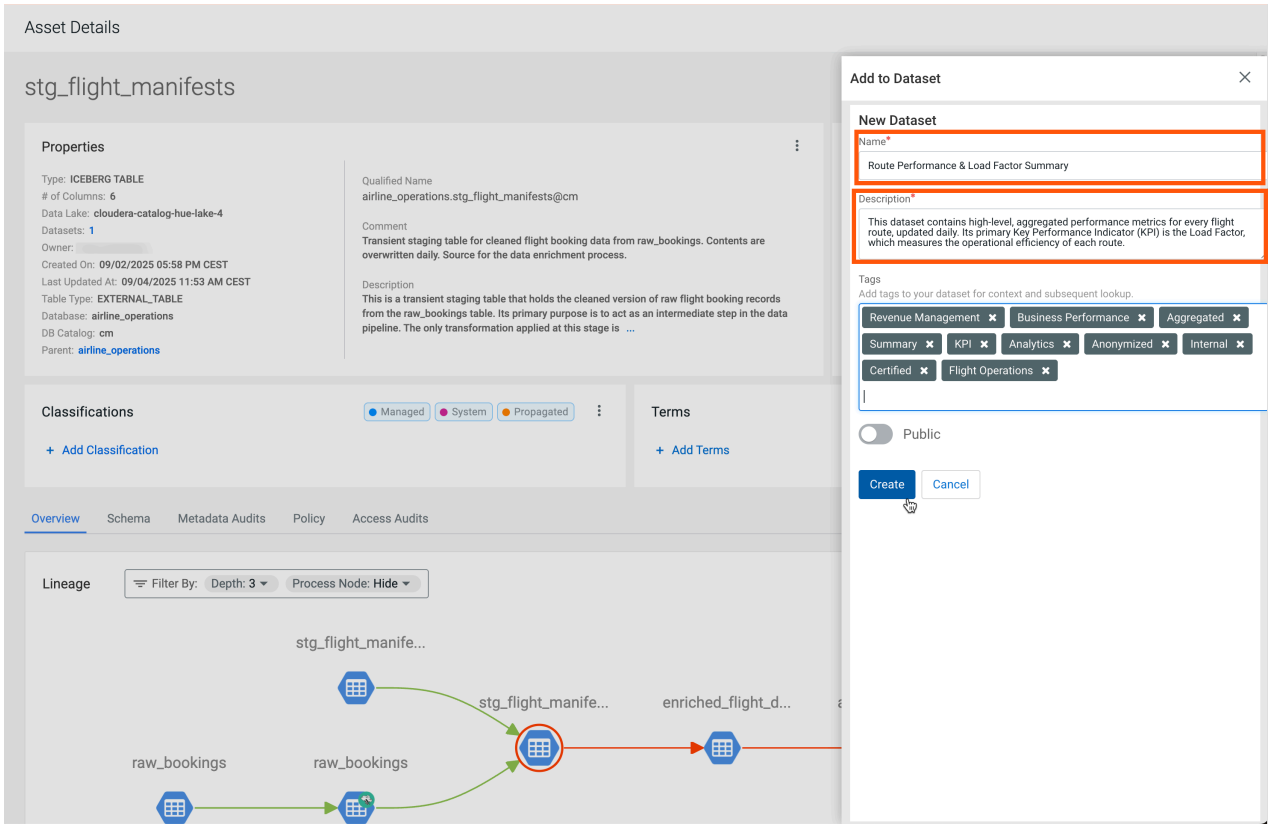


The **Add to Dataset** pane provides an option to add the asset into one or more existing datasets or even create a new one.

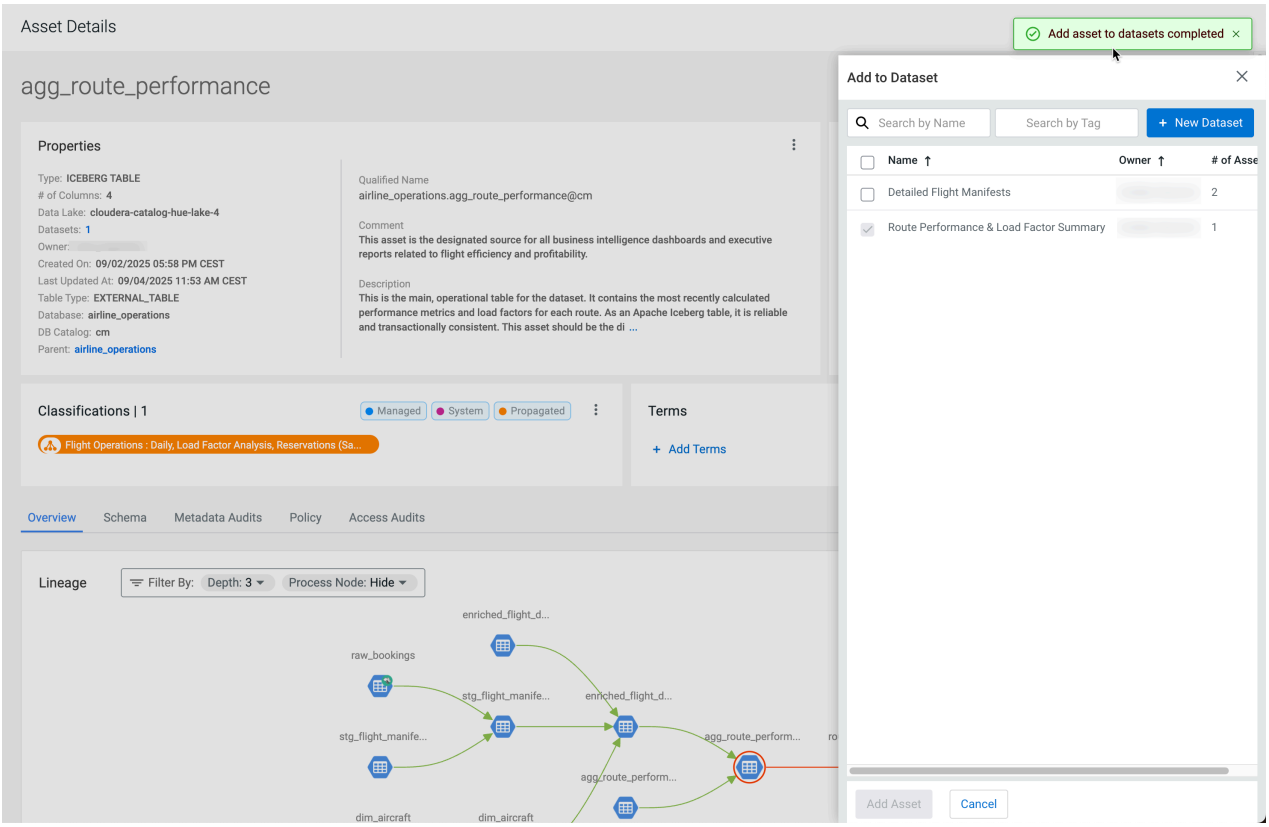
Datasets that already contain the asset are disabled and marked as checked. Datasets which are currently in edit state are disabled and marked with the characters a *.



Users can search for an existing dataset by name or by tags applied on the dataset. Users can select one or more datasets from the list and then click on the Add Asset button which adds the asset to these dataset(s).



There are instances, where there are no datasets present or the user just wants to create a new dataset to add the asset. In that case, the user can click on the New Dataset button which opens up a new dataset form. Once the user fills in the form and clicks on the Create button, a new dataset with the given properties is created and the asset is added to it automatically. This is reflected in the datasets list where the newly added dataset is highlighted.

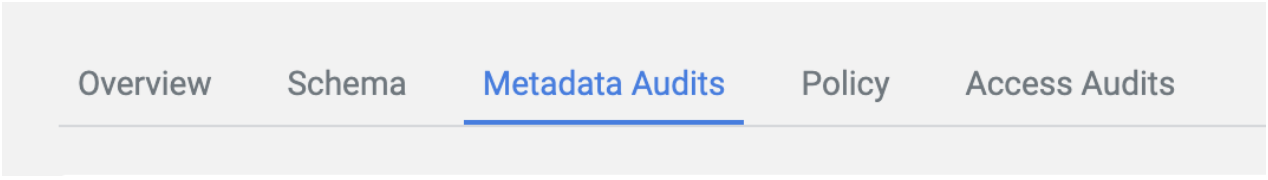


Viewing Atlas entity audits

Atlas audits help Data Stewards to identify and track the entity changes or modifications that are performed over a period of time.

Information about the Apache Atlas entity audit events are displayed for each entity in the **Asset Details**. Using this information, Data Stewards can distinguish between entity audits and data audits that originate from Ranger.

On the **Asset Details** page, the **Metadata Audits** tab displays information related to the selected entity type and about the events that occurred based on the user activities.



Clicking on **Metadata Audits**, tab, you can view manage information about:

- The user who made the changes to the specific entity
  - Users also include the profilers
- The time when the entity was changed
- The kind of change that was made to the entity
  - For example, adding comments, descriptions and the entity itself being created

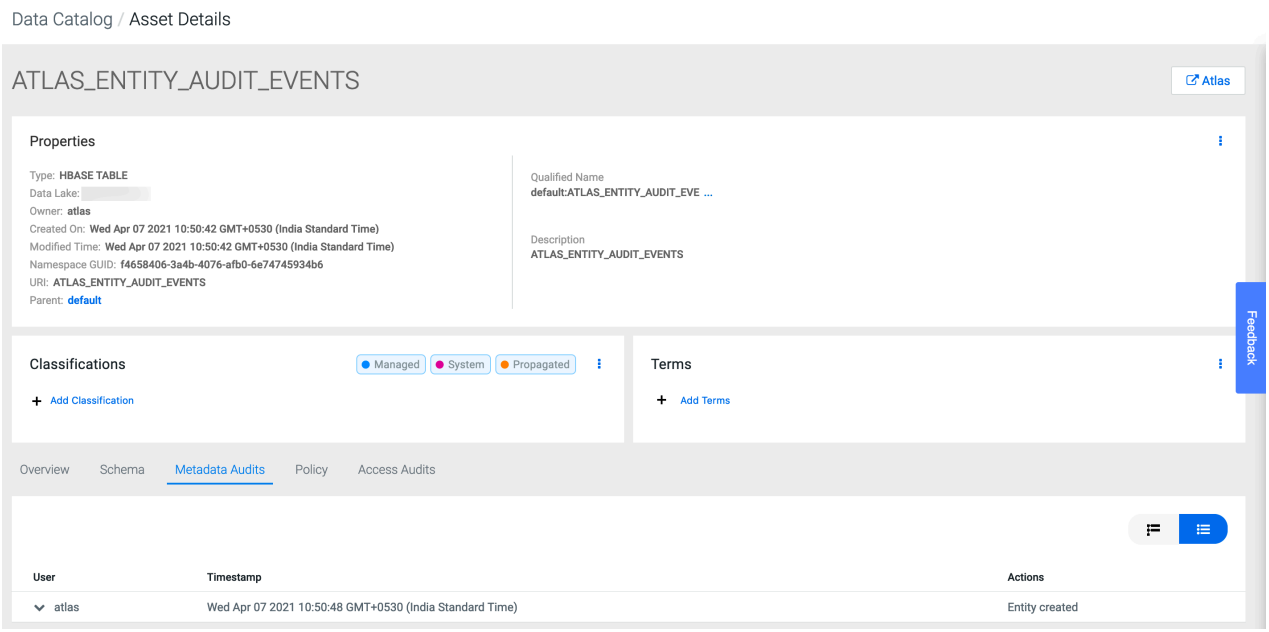
- Any other relevant changes pertaining to the audit entries

The changes that can be identified for:

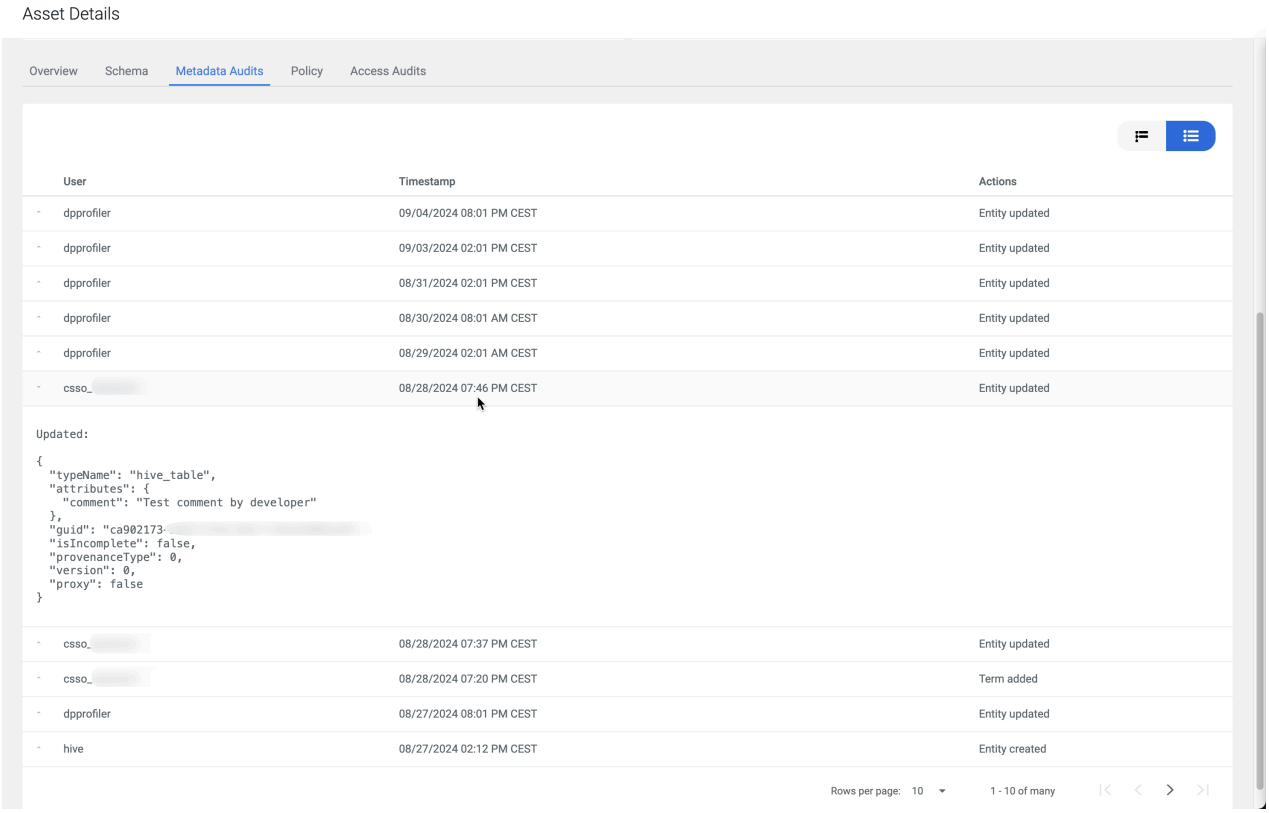
- Created entities and related updates
- Tagged entities
- Labeled entities
- Export and Import operations

For example, the following image displays the Atlas audit entity creation event that is recorded by each Atlas entity that is displayed in the **Asset Details** page.

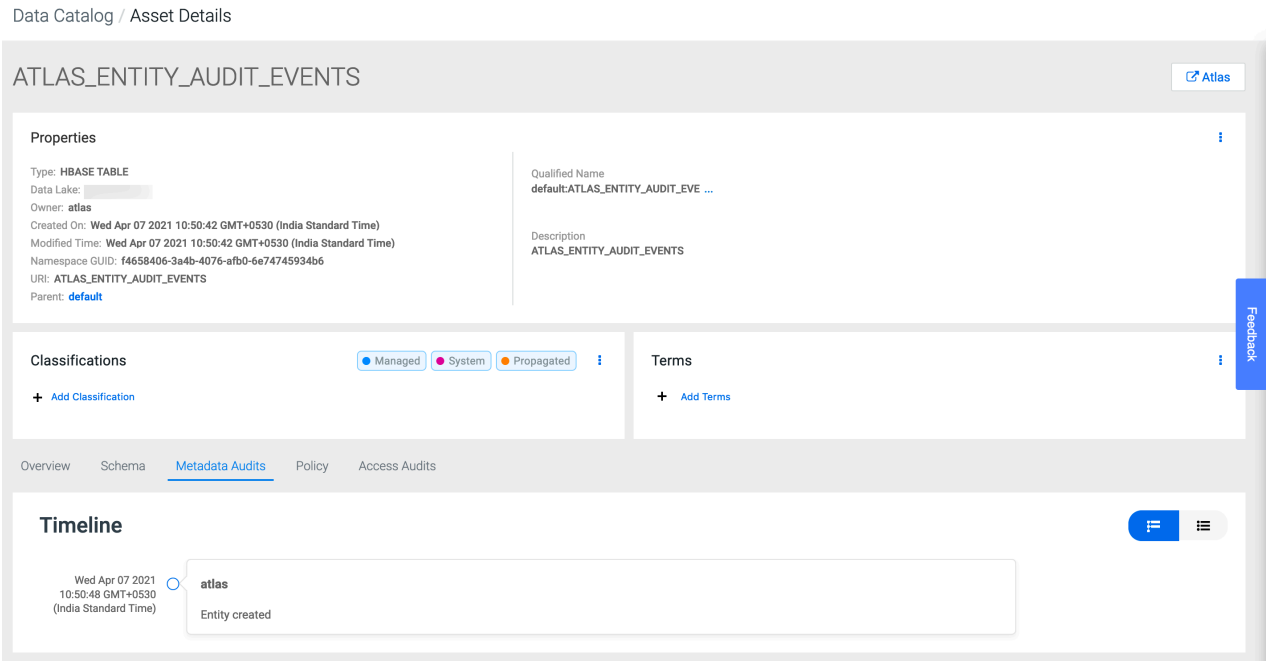
a



Clicking on any line item displays the change in JSON format, which is directly derived from Atlas. For example, adding an Atlas glossary term:



Use the toggle icon (on the top-right corner) for viewing Atlas audits in different formats. By default, you can view the audits in tabular format. When you toggle the view icon, you can view the **Timeline** format.



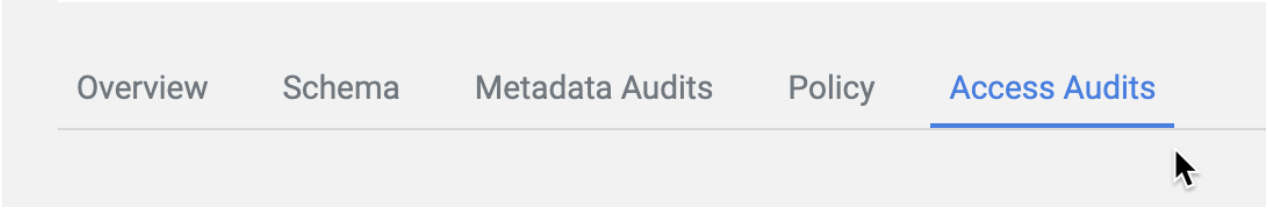
Clicking on a user in the **Timeline** format displays the JSON data.

Viewing Ranger access audits

Apache Ranger audits help Data Stewards to identify and track the policies affecting the Apache Atlas entities over a period of time. More specifically, the accesses and the relevant policies can be also identified.

Information about the Apache Ranger policy audit events are displayed for each entity in the **Asset Details**. Using this information, Data Stewards can distinguish between policies originating from Ranger to have an overview of all users accessing (or being prevented from accessing) an entity. They can also use this to troubleshoot access issues.

On the **Asset Details** page, the **Access Audits** tab displays information related to the selected entity type and about the events that occurred based on the user activities trying to access the entity.



Clicking on **Access Audits** tab, you can view manage information about:

- The relevant Ranger policy ID
- The time when the entity access was attempted
- The user who accessed the specific entity (or was being barred from accessing it)

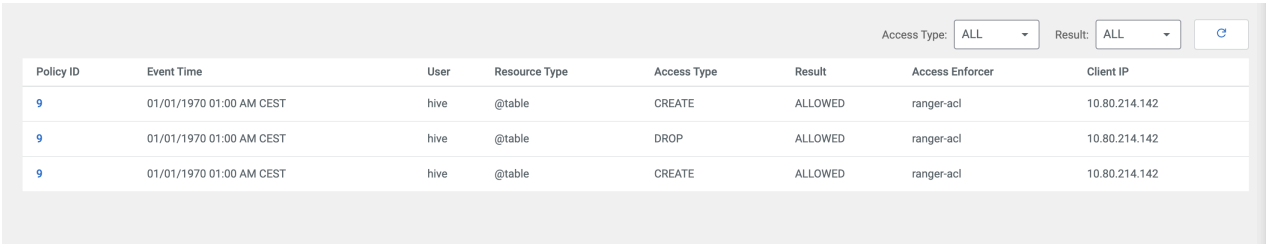
  Note: Users also include the profilers.

- The resource type accessed
- The access type
  - For example, by Hive entities the following filter options are available
    - SELECT
    - UPDATE
    - CREATE
    - DROP
    - ALTER
    - INDEX
    - READ
    - WRITE
- The access attempt result
  - ALL
  - ALLOWED
  - DENIED
- The client IP

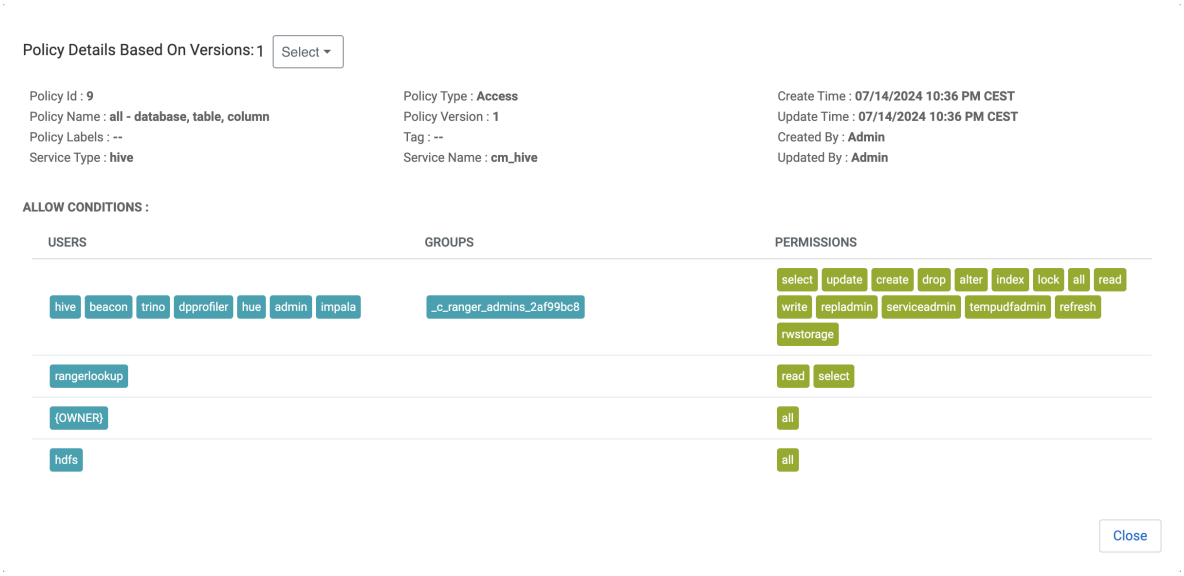
The accesses can be identified for:

- Created entities and related updates
- Tagged entities
- Labeled entities

For example, the following image displays the entity creation event recorded for a Hive table:



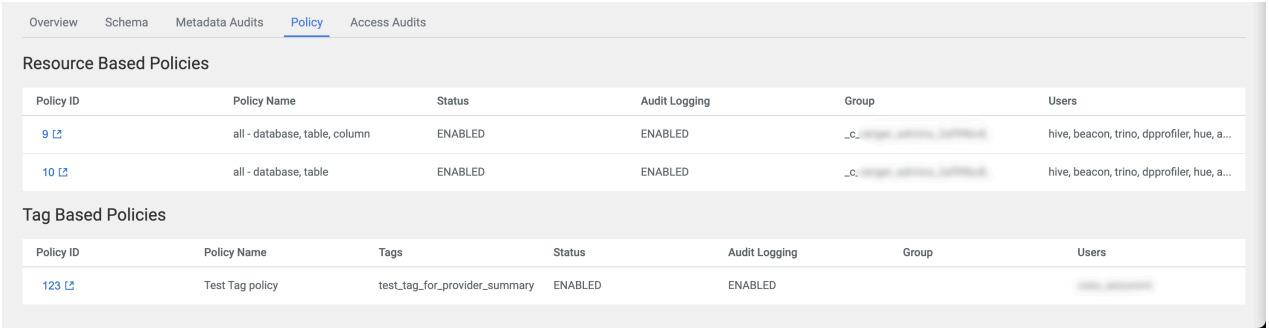
Note: If you click an event row, you can see an overview of the users set for that policy including the set permissions.



Viewing Ranger policies

Apache Ranger audits help Data Stewards to identify and track the policies affecting the Apache Atlas entities.

In Asset Details Policy , you can have an overview of the Ranger policies currently affecting the selected entity.

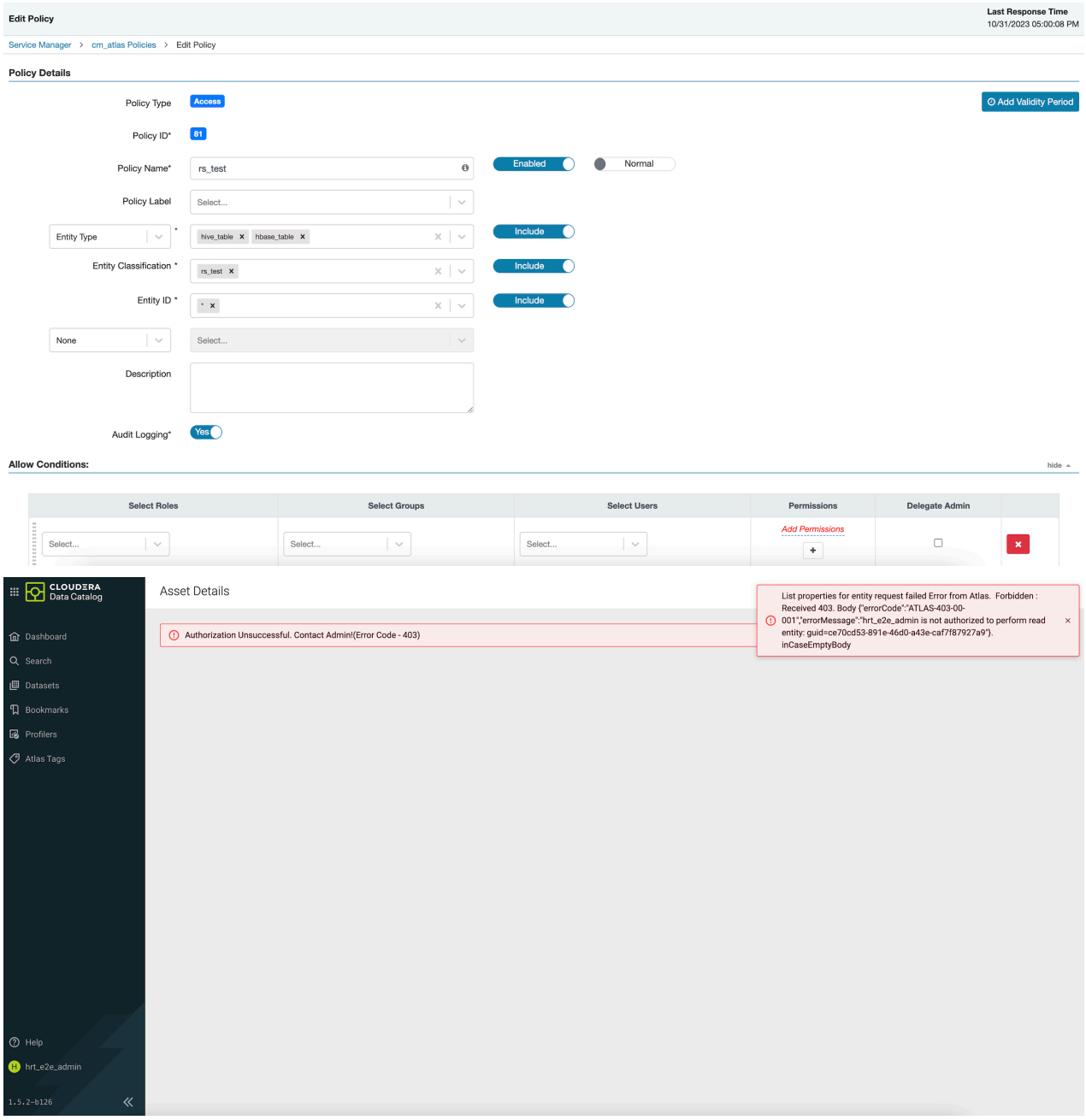


By clicking the Policy ID, you can directly open the relevant policy in Ranger for editing.

Accessing tables based on Ranger policies

When a table (in blue color link) is clicked in Search results, the Asset Details view page is displayed.

If a user is not authorized to click or view table details, it implies that the user permissions have not been set up in the Apache Ranger.



Related Information

Providing role access

Ranger Policies Overview

Resource-based Services and Policies

Configuring resource-based policies

Missing authorization for viewing assets