CDP Public Cloud / Data Catalog

Search

Date published: 2019-11-14
Date modified: 2026-03-31



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2026. 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

Search filters....4

Searching for assets using Atlas glossaries....7
Using terms in Cloudera Data Catalog....7
Searching for assets using glossary terms....10

Viewing Data Asset details....10
Supported Data Asset fields for entities....12
Navigation in Asset Details....13
Creating a classification from Asset Details....15
Creating a classification from Search page....18
Additional search options for asset types....18
Mapping glossary terms....21
Navigation support for hive entities within Lineage....25
Adding assets to one or more datasets....26
Viewing Atlas entity audits....29
Viewing Ranger access audits....31
Viewing Ranger policies....32

Accessing tables based on Ranger policies....33

Prerequisites for accessing Cloudera Data Explorer (Hue) tables and databases....34
Accessing Cloudera Data Explorer (Hue) assets....34
Accessing Cloudera Data Explorer (Hue) tables and databases from the Search page....36
Accessing Cloudera Data Explorer (Hue) tables and databases from the Asset Details page....36

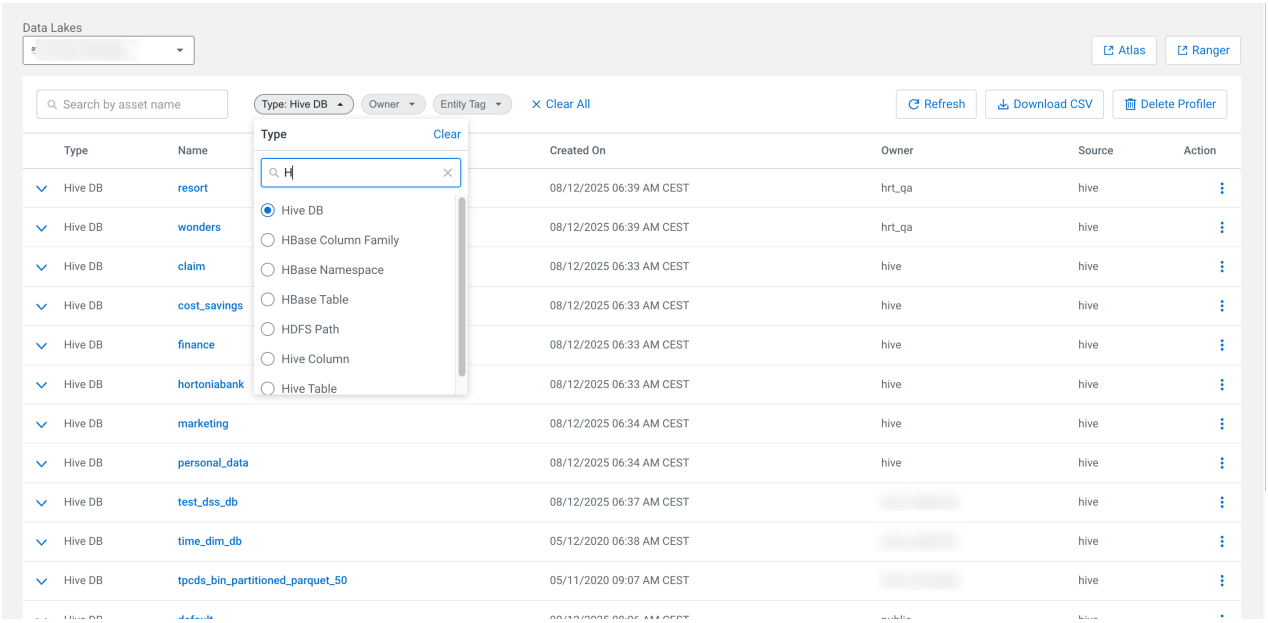
Integrating Cloudera Data Catalog with AWS Glue Data Catalog....38
Setting up AWS Glue Catalog with Cloudera Data Catalog....38
Working with AWS Glue....42
Accessing AWS Glue....42

Search filters

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

Cloudera recommends selecting your asset Type first because certain asset types enable more filters specific to that type.

Figure 1: Type filter



Note: The Type filter can have only one asset type selected at a time, but you can limit the number of displayed asset types by starting to type in the type name to help finding the type to selected.

For example, table assets usually have different column filters.

Figure 2: HBase table time range filter

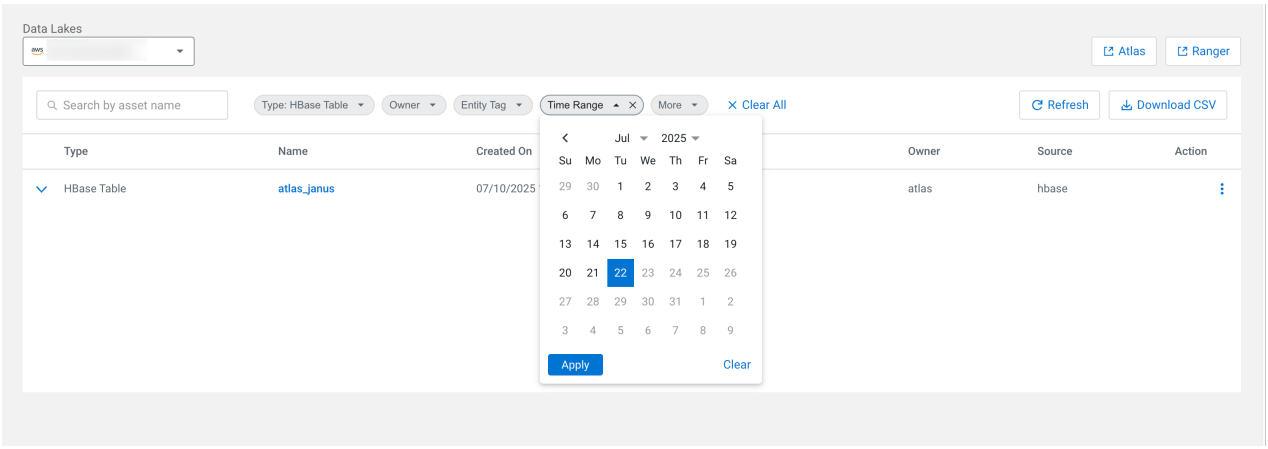
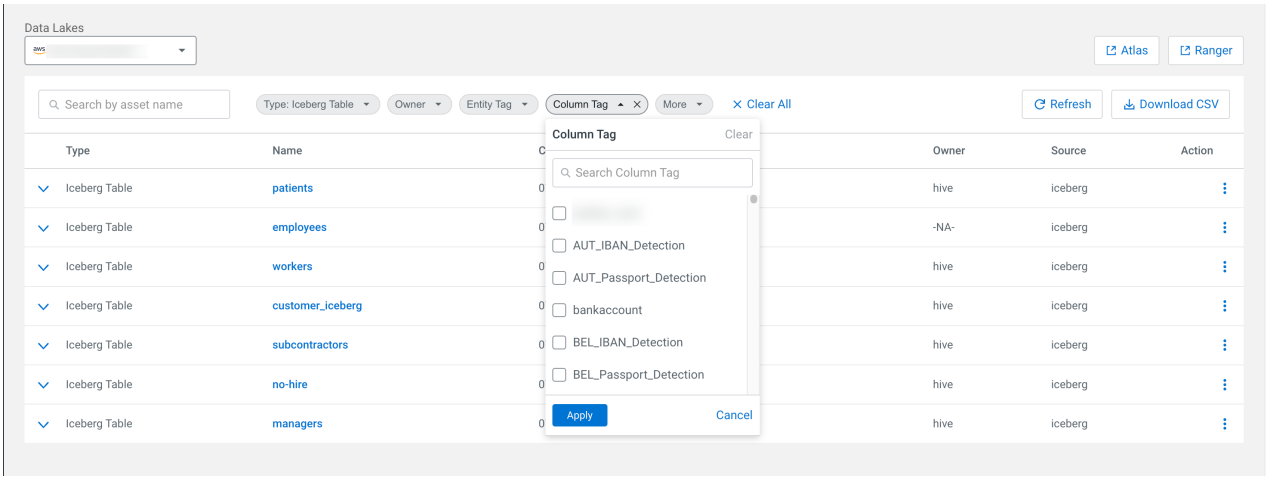


Figure 3: Column Tag filters



Note: Some filters display fewer child filters, as irrelevant ones are hidden when no data is recorded or synchronized from Atlas.

You can use the following filters:

Type

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

- Azure BLOB
- Azure Container
- Azure Directory
- AWS S3 Bucket
- AWS S3 Object
- AWS S3 Pseudo Dir
- AWS S3 V2 Bucket
- AWS S3 V2 Directory
- AWS S3 V2 Object
- HBase Column Family
- HBase Namespace
- HBase Table
- HDFS path
- Hive Column
- Hive DB
- Hive Table
- Iceberg Column
- Iceberg Table [1]
- Impala Column Lineage
- Impala Process
- Impala Process Execution
- Kafka topic
- ML Model Build
- ML Model Deployment
- ML Project
- RDBMS Column
- RDBMS DB

[1] Iceberg assets are discoverable in VM-based environments but they can be profiled only in Compute Cluster-enabled environments.

- RDBMS Foreign key
- RDBMS Index
- RDBMS Table
- Spark Application
- Spark Column
- Spark Column Lineage
- Spark DB
- Spark ML Directory
- Spark ML Model
- Spark ML Pipeline
- Spark Process
- Spark Process Execution
- Spark Table

Note: After selecting an entity type, further filters related to that type will be available under the More filter option. For example, selecting the Hive Table type will enable the Column Tag filter.

Owner

Select the owner to further refine the results.

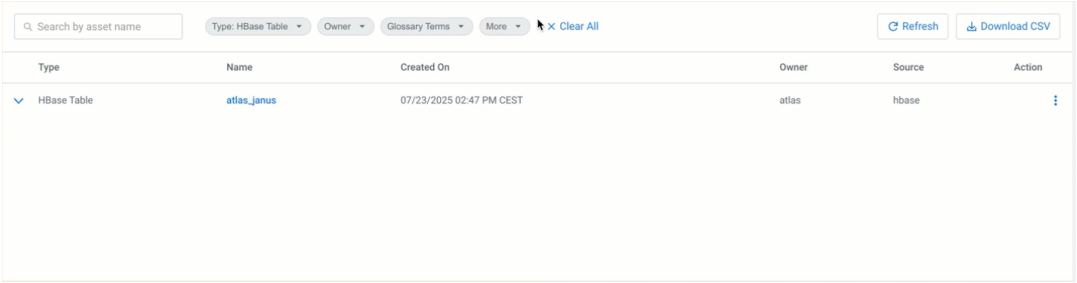
Entity Tag

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.

Note: When searching for Hive or Iceberg columns based on Owner, Entity Tag or Glossary Terms, select the asset Type first (Hive Column, Iceberg Column).

Time Range

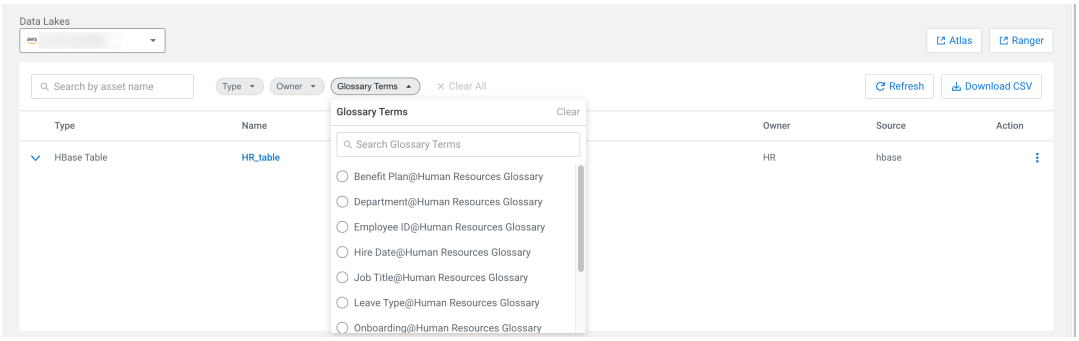
You can filter your assets by the **Created On** date, if provided by Atlas, after selecting and asset Type. Use the calendar widget to select a range and click Apply.



Glossary Terms

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 glossary 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.

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.

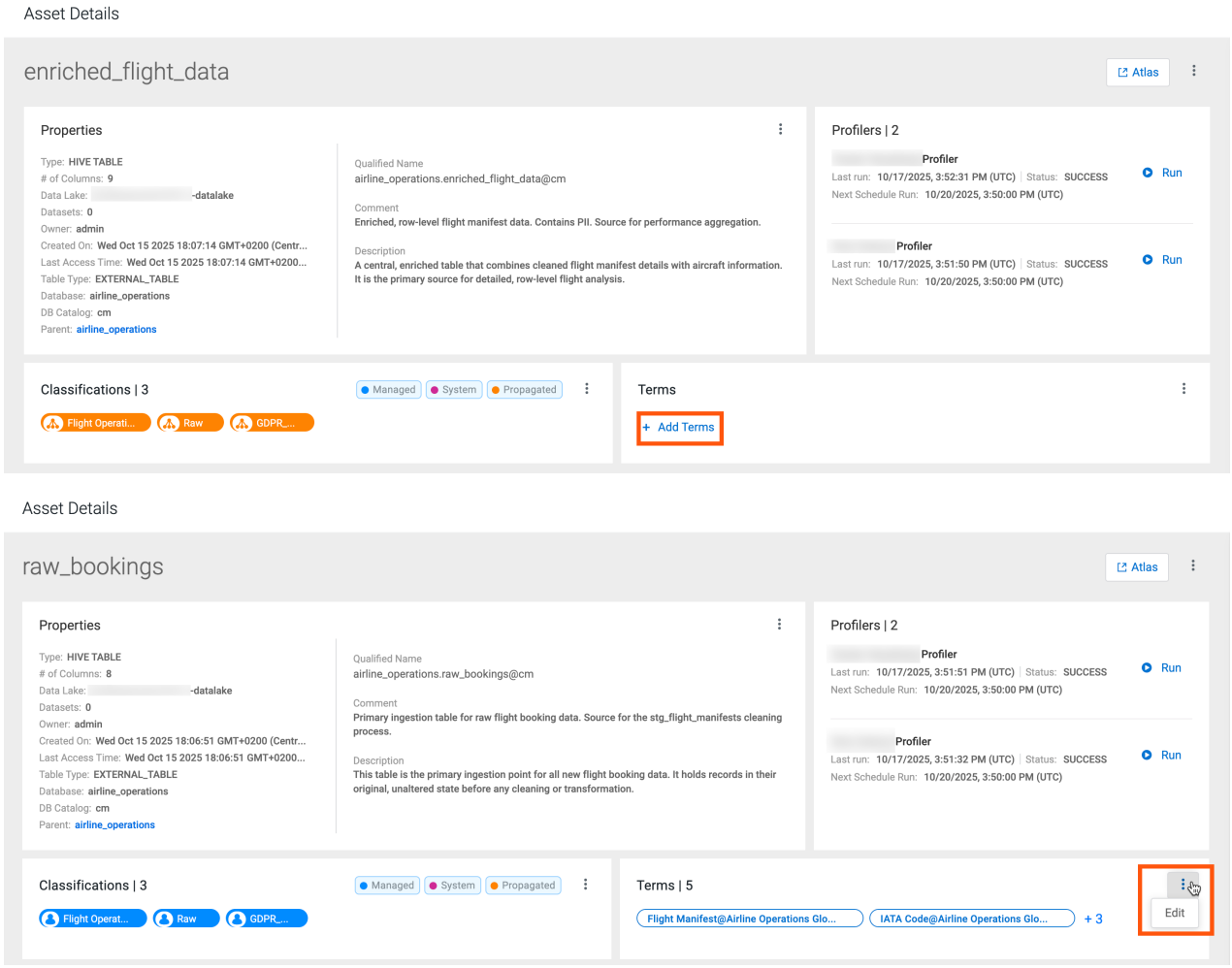
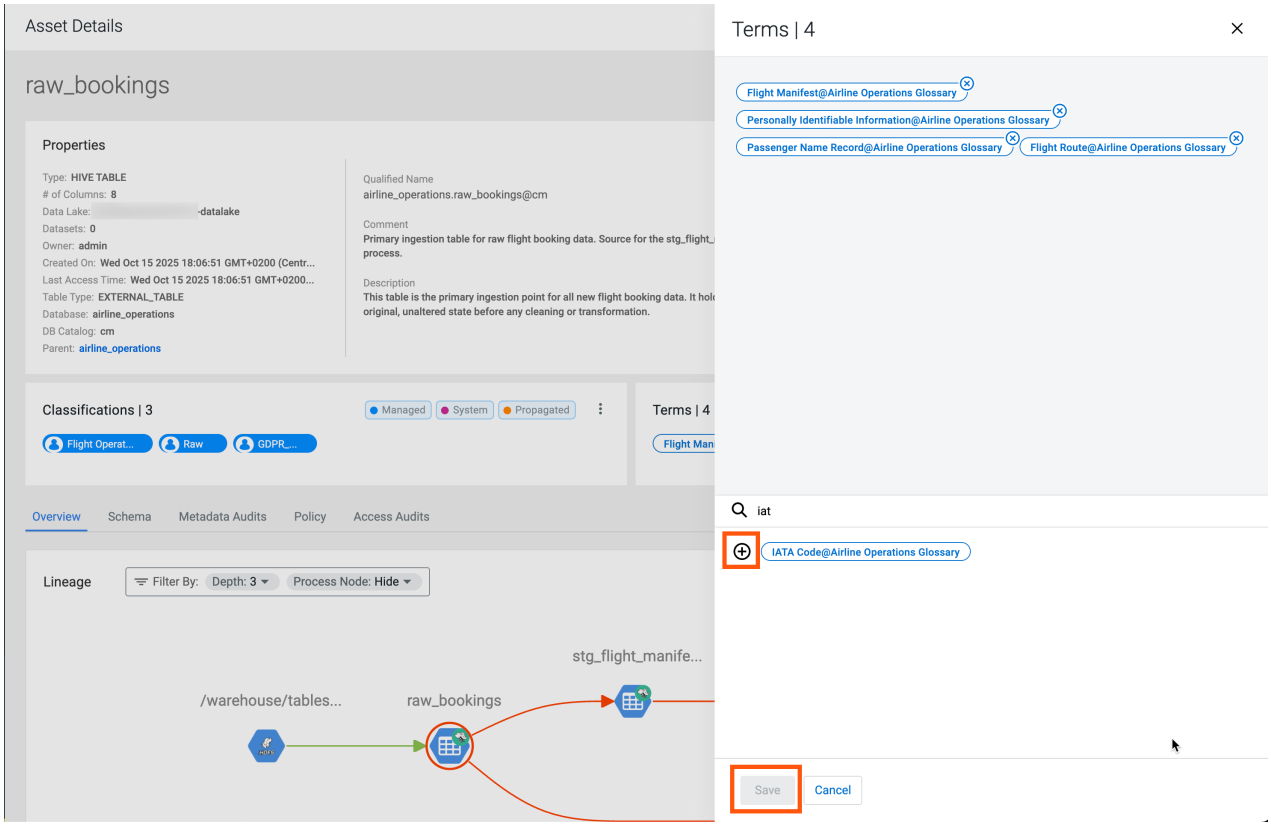
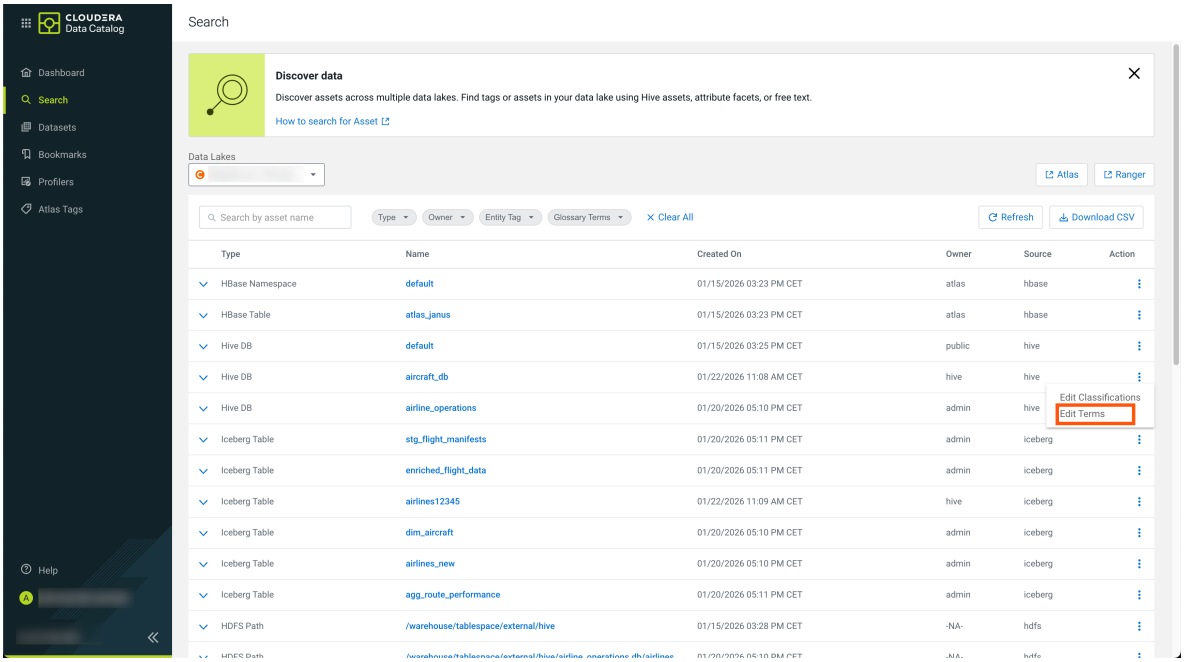


Figure 4: Editing glossary terms of an asset



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



Related Information

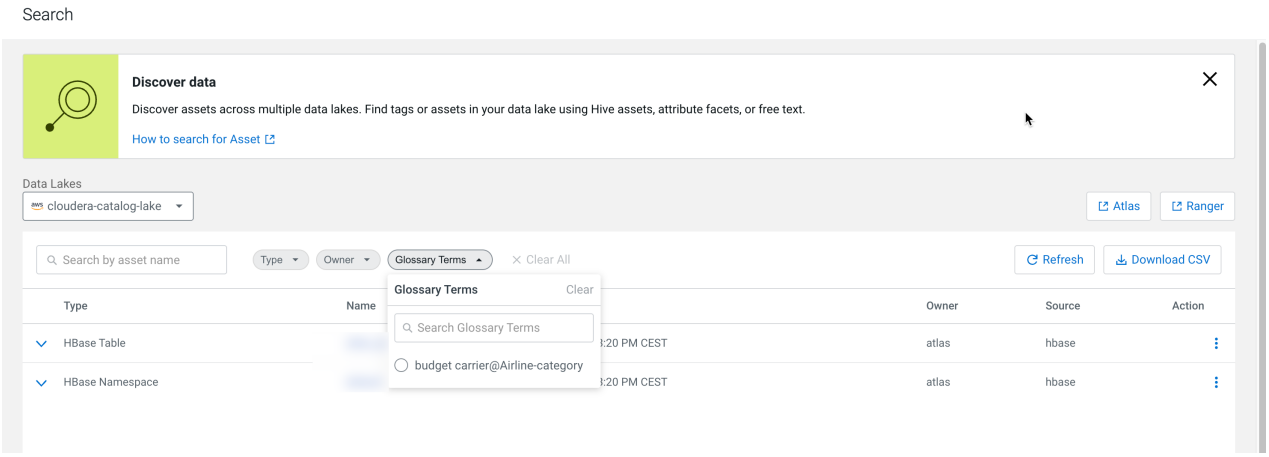
Glossaries overview
Creating glossaries
Creating terms
Searching for assets using Atlas glossaries
Search filters

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. Select your Data Lake.
3. Click the Glossary Terms drop-down and selected the term to be searched.



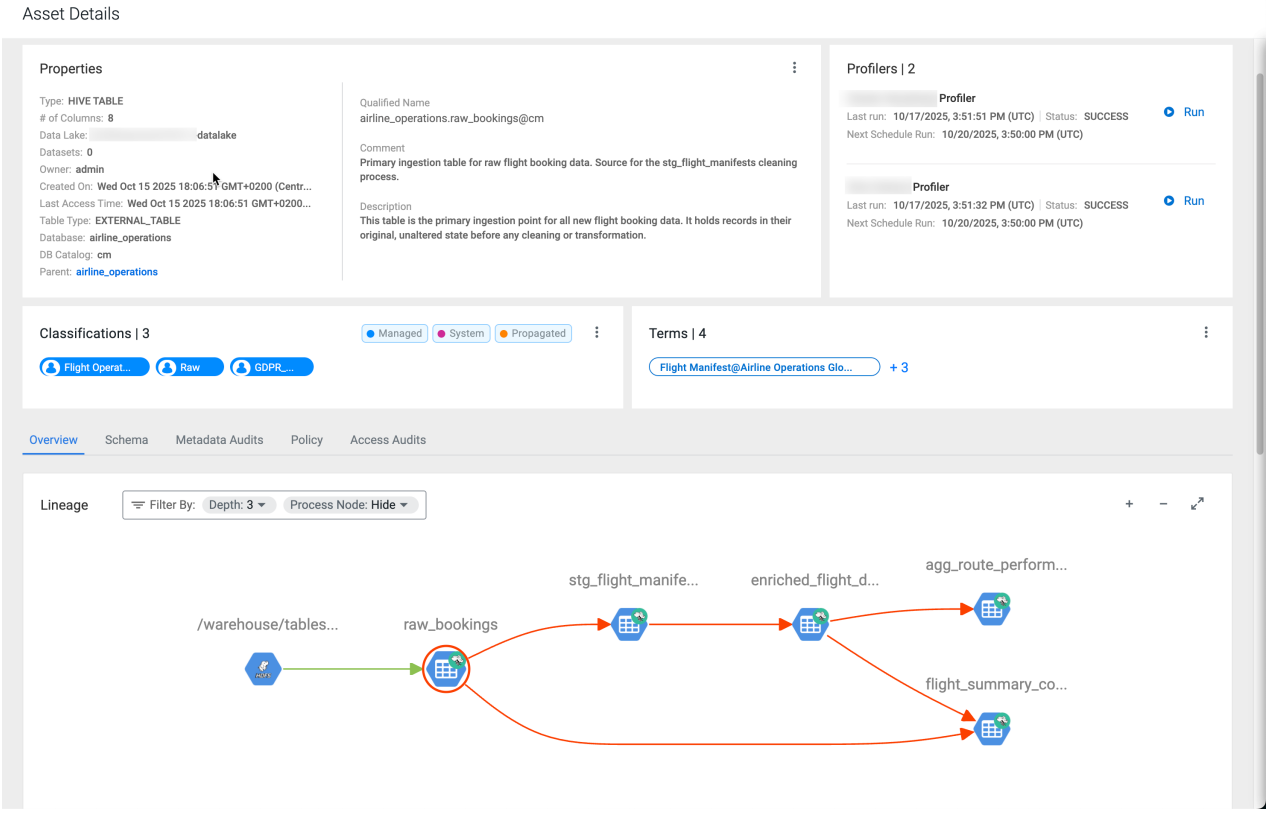
Related Information

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
aws_S3_bucket	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
aws_S3_Object	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
aws_S3_pseudo_dir	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes
aws_s3_v2_object	Yes	Yes	Not Supported	Not Supported	Not Supported	Yes	Yes
aws_s3_v2_directory	Yes	Yes	Not Supported	Not Supported	Not Supported	Yes	Yes
aws_s3_v2_bucket	Yes	Yes	Not Supported	Not Supported	Not Supported	Yes	Yes
adls_gen2_directory	Yes	Yes	Not Supported	Not Supported	Not Supported	Yes	Yes
adls_gen2_blob	Yes	Yes	Not Supported	Not Supported	Not Supported	Yes	Yes
adls_gen2_container	Yes	Yes	Not Supported	Not Supported	Not Supported	Yes	Yes
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
Hbase Column Family	Yes	Yes	Not Supported	Not Supported	Yes	Yes	Yes
Iceberg Table[2]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Iceberg Column	Yes	Yes	Yes	Not supported	Yes	Yes	Yes
impala_process	Yes	Yes	Not Supported	Not Supported	Not Supported	Not Supported	Yes

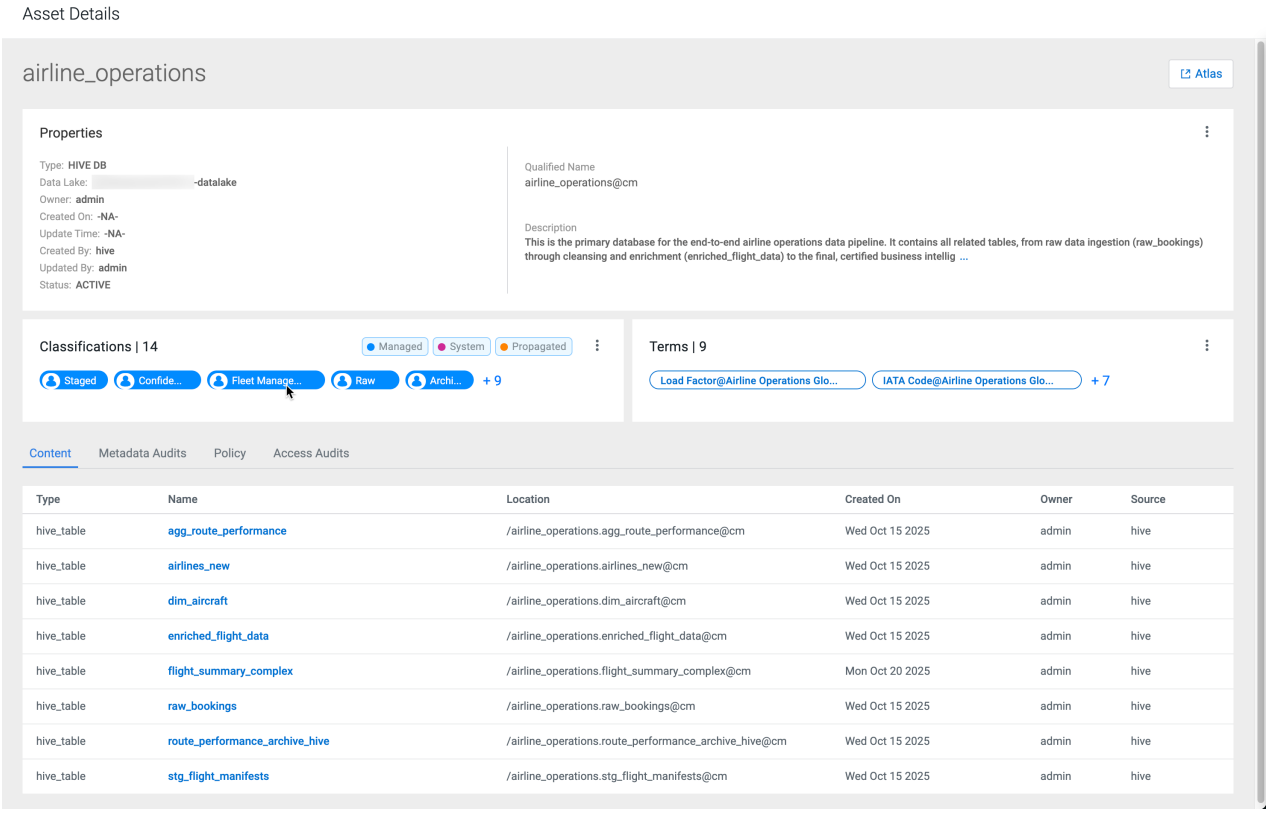
[2] Iceberg assets are discoverable in VM-based environments but they can be profiled only in Compute Cluster enabled environments.

Asset Type	Lineage	Tagging	Access Metrics	Schema	Policy	Metadata and Access Audits	Atlas Punch out
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

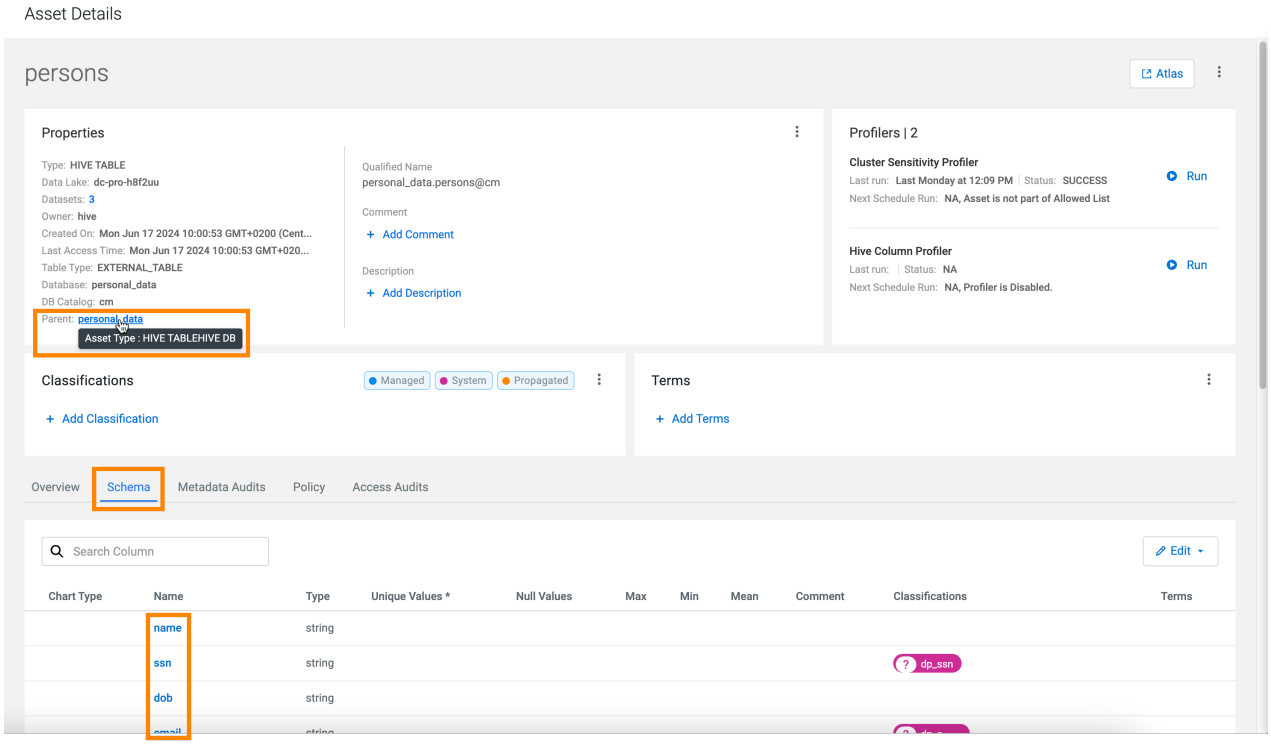
A generic Assets Details page is available for container data types like buckets and databases.

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 or Iceberg 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, Iceberg Table
Hive Table, Iceberg Table[3]	Hive DB	Hive column, Iceberg column
HBase Namespace	-	HBase Table
ML Project	-	ML Model Build
ML Model Build	ML Project	ML Model Deployment
AWS S3 Bucket	-	AWS S3 Pseudo Dir
AWS S3 Pseudo Dir	AWS S3 Bucket	AWS S3 Object
RDBMS Instance	-	RDBMS DB
RDBMS DB	RDBMS Instance	RDBMS Table

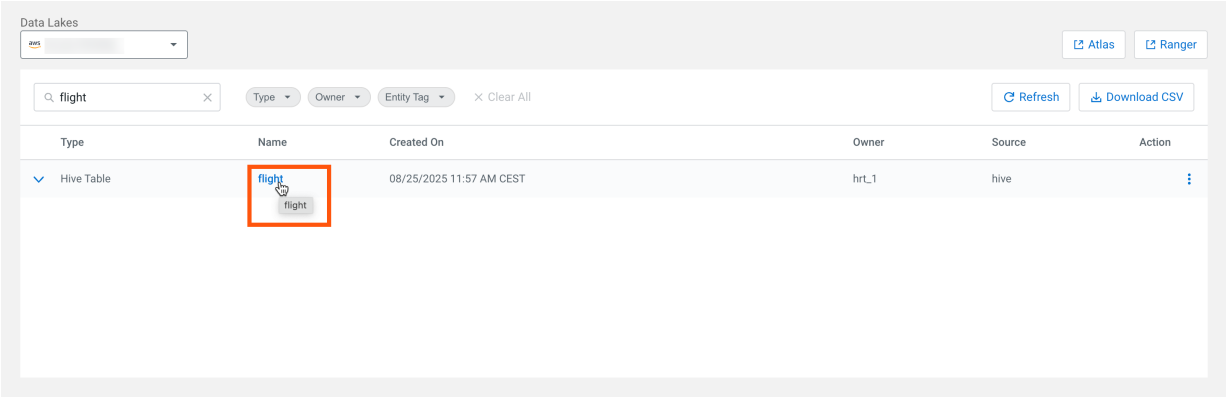
Creating a classification from Asset Details

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

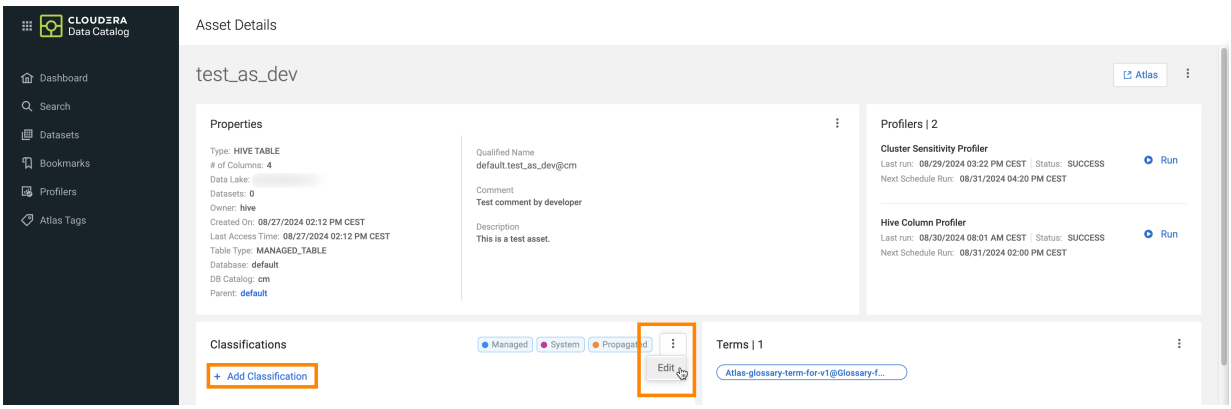
1. Go to the **Asset Details** page of an asset.

[3] Iceberg assets are discoverable in VM-based environments but they can be profiled only in Compute Cluster enabled environments.



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

Figure 5: Adding Atlas classifications



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

Figure 6: Creating a new Atlas classification from Cloudera Data Catalog

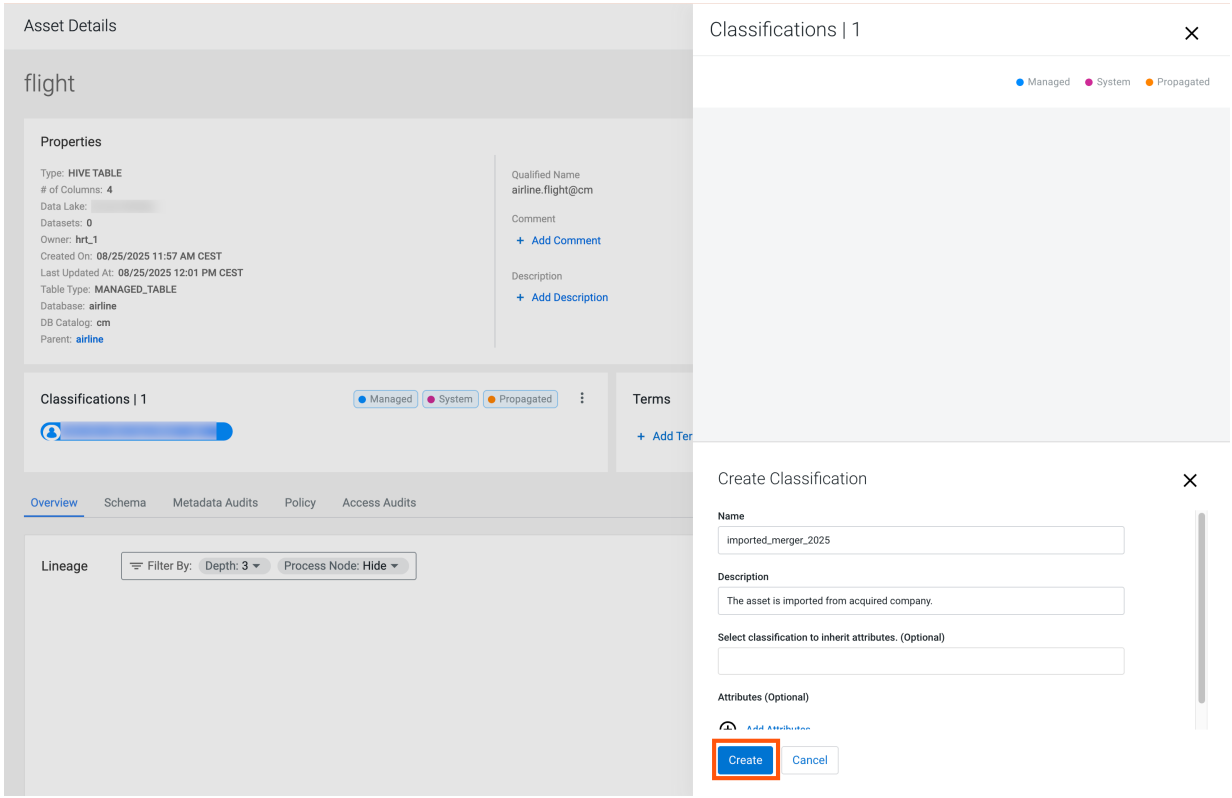
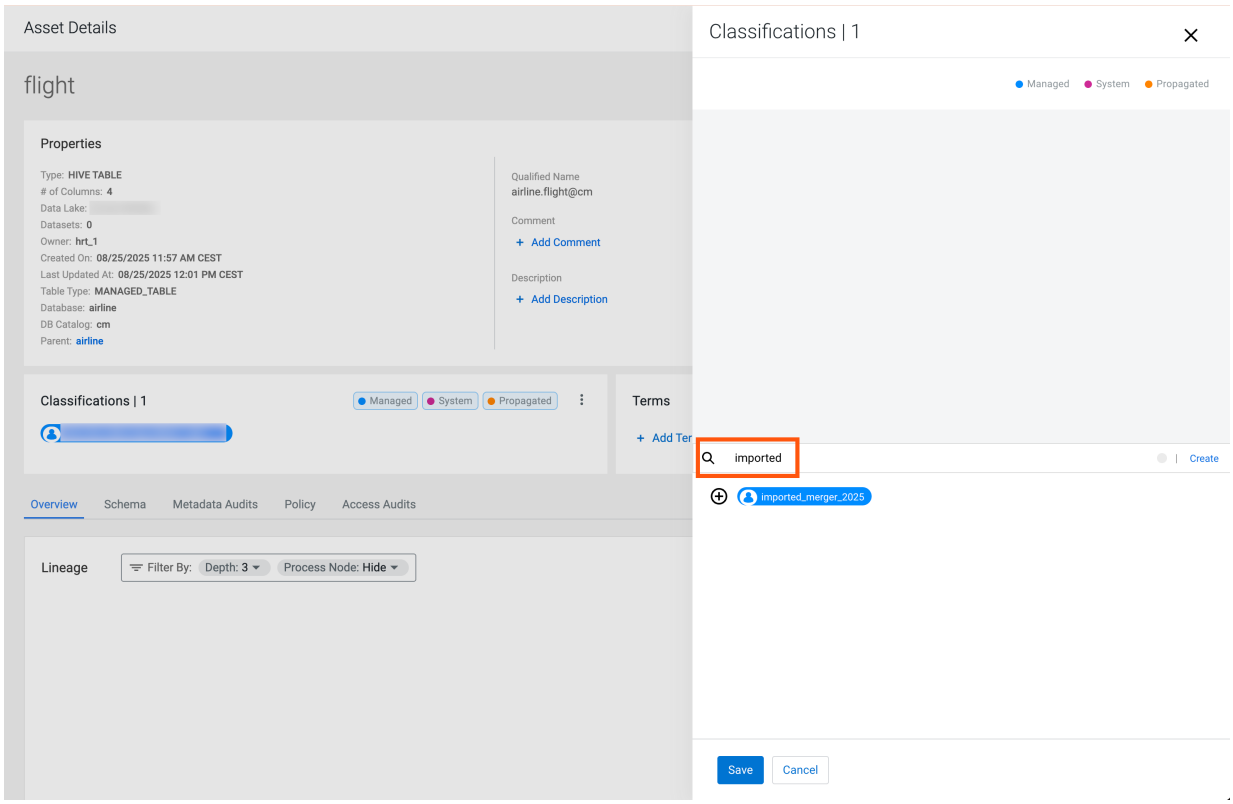


Figure 7: Searching for an Atlas classification



4. Click Save to finalize your changes.

Related Information

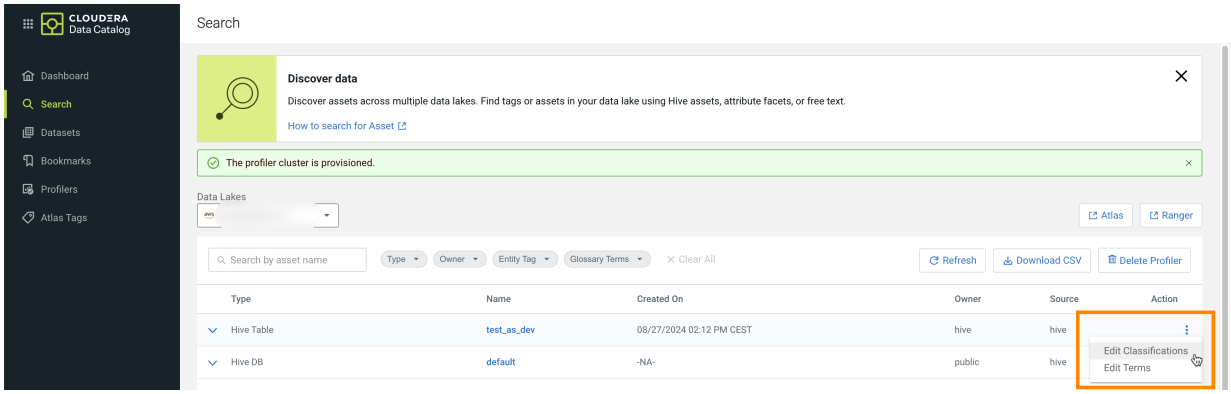
Apache Atlas metadata attributes
Working with Atlas classifications and labels
Creating classifications
Tag-based Services and Policies
Atlas classifications drive Ranger policies
Atlas tag management
Creating a classification from Search page

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

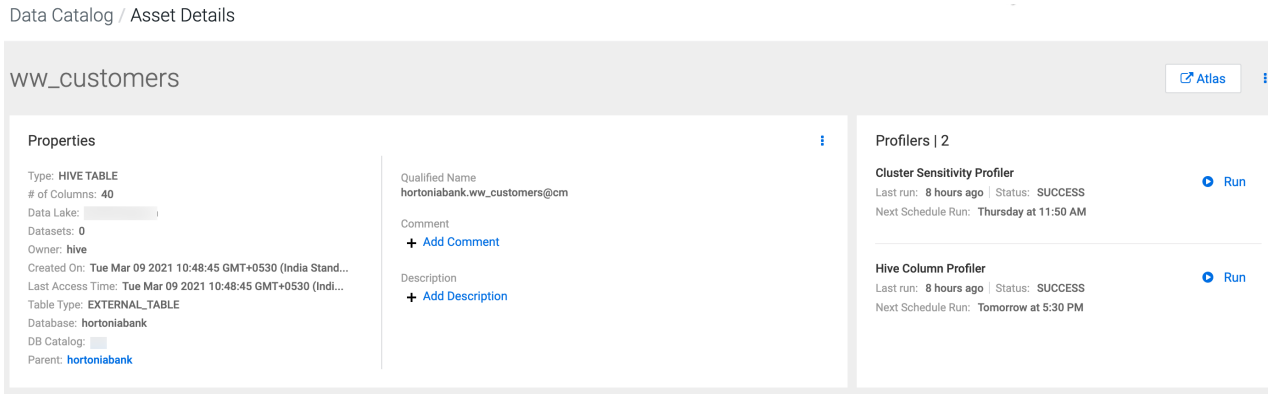
Using Cloudera Data Catalog, you can add or edit asset description values to search for data assets across both Cloudera Data Catalog and Apache Atlas services by using the asset content. These values can be searched.

Adding comments and descriptions to assets

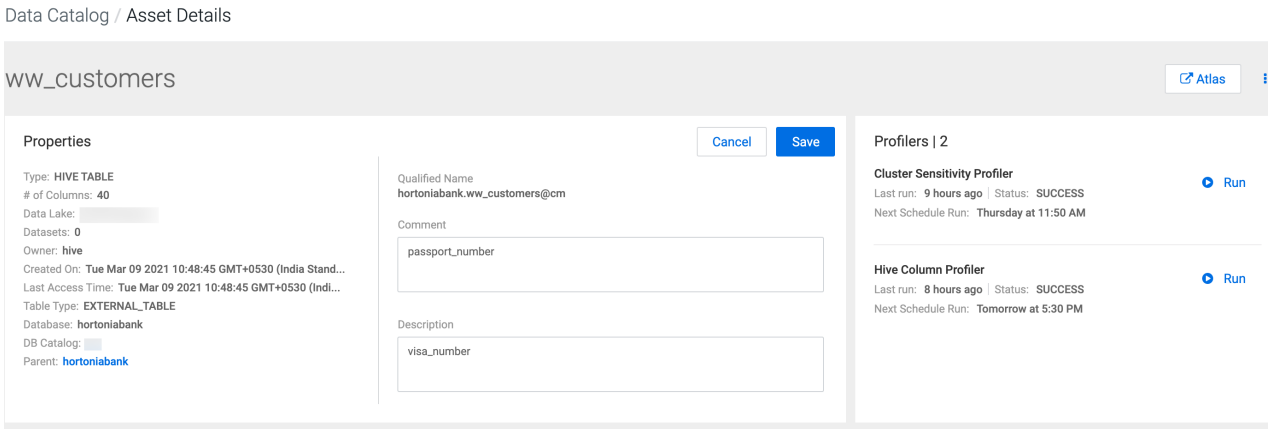
In the **Asset Details** page for each asset type that you select, you can add or edit **comment** or **description** fields. Including these values for the selected asset helps you to identify your chosen asset.

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

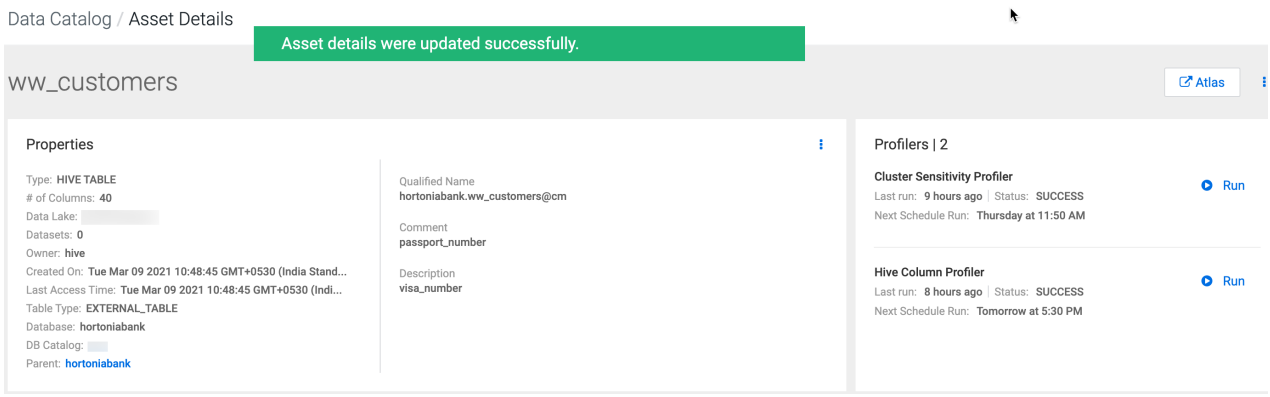
Note: The comment and description options are supported only for Hive table and Hive Column assets. For other asset types, only the description option is supported.



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

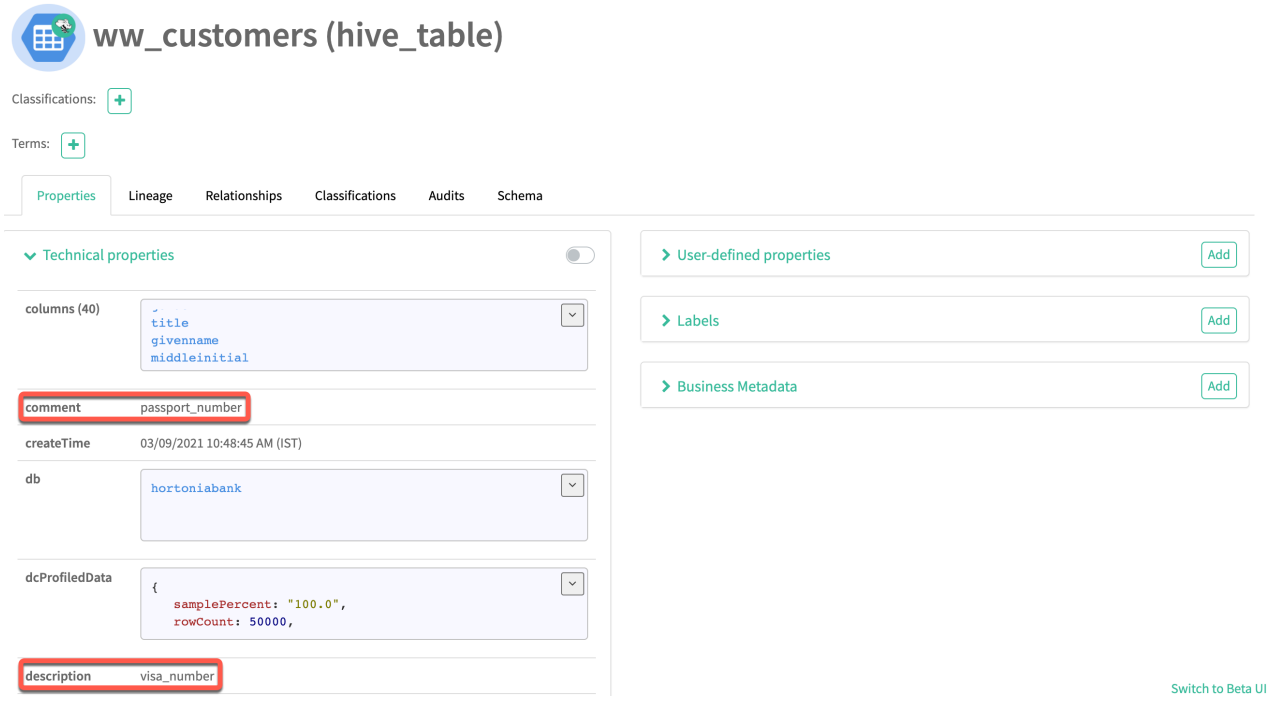


Click Save to save your changes.



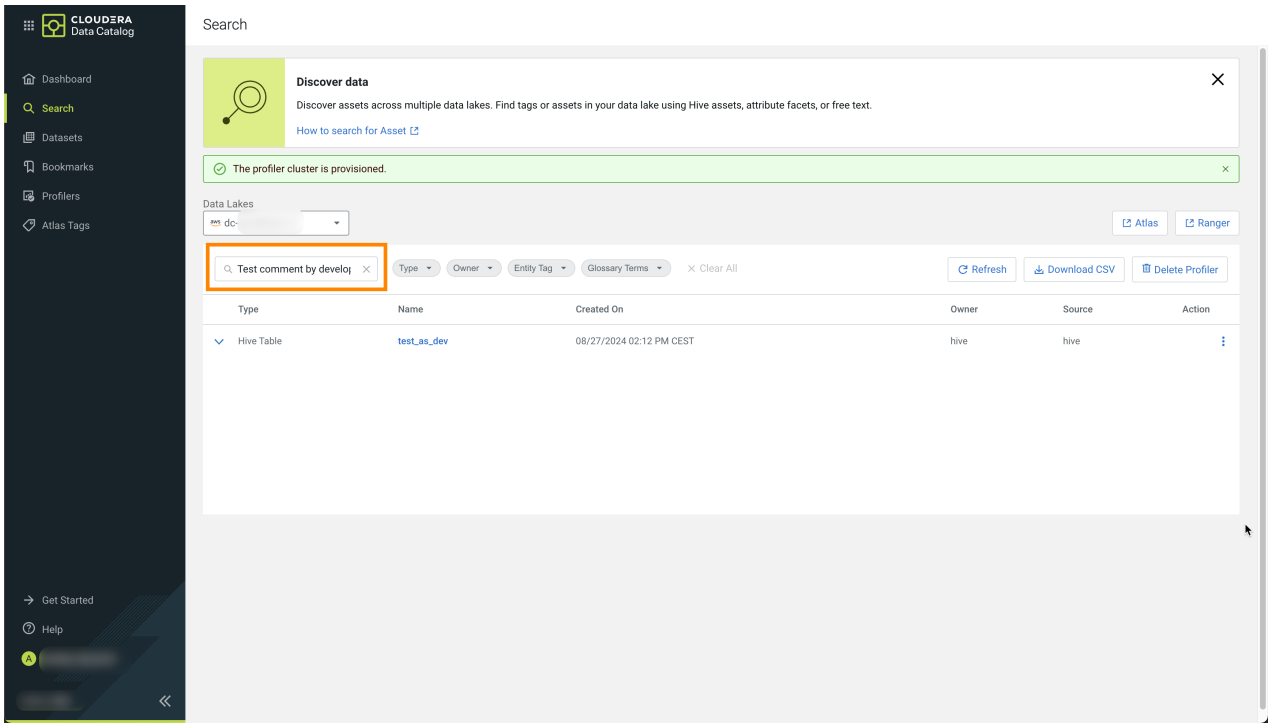
Note: You can also edit the already saved valued by clicking the icon.

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

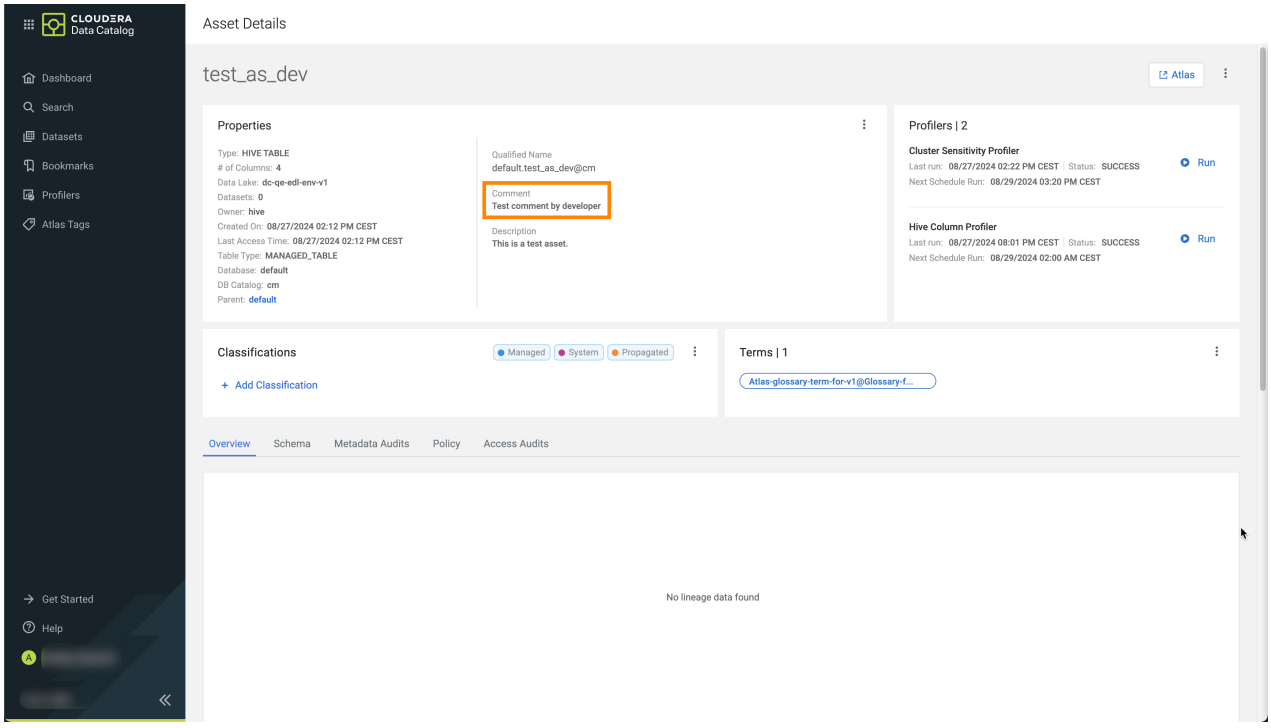


Searching comments and descriptions

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 8: Glossary term in Search

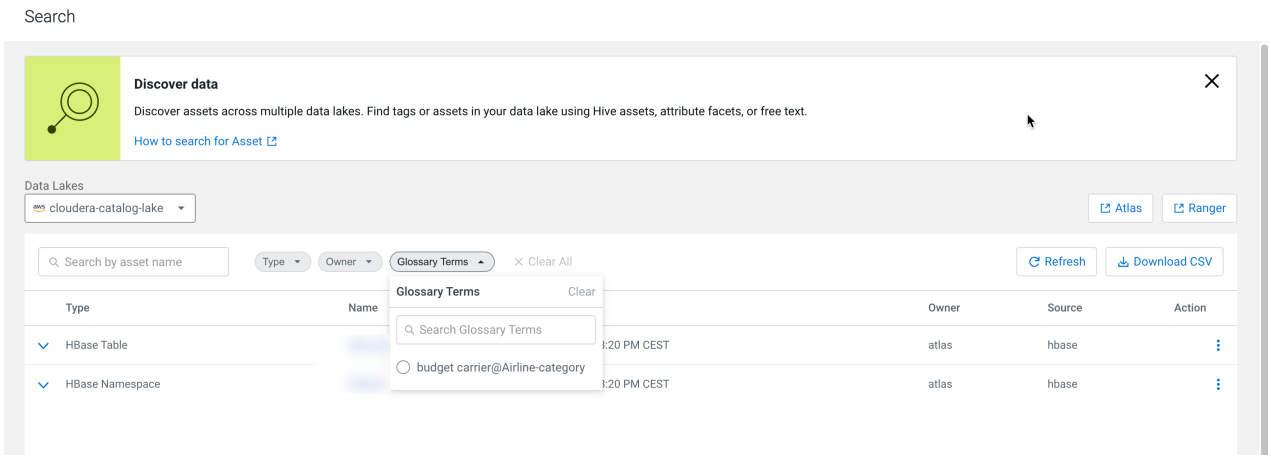
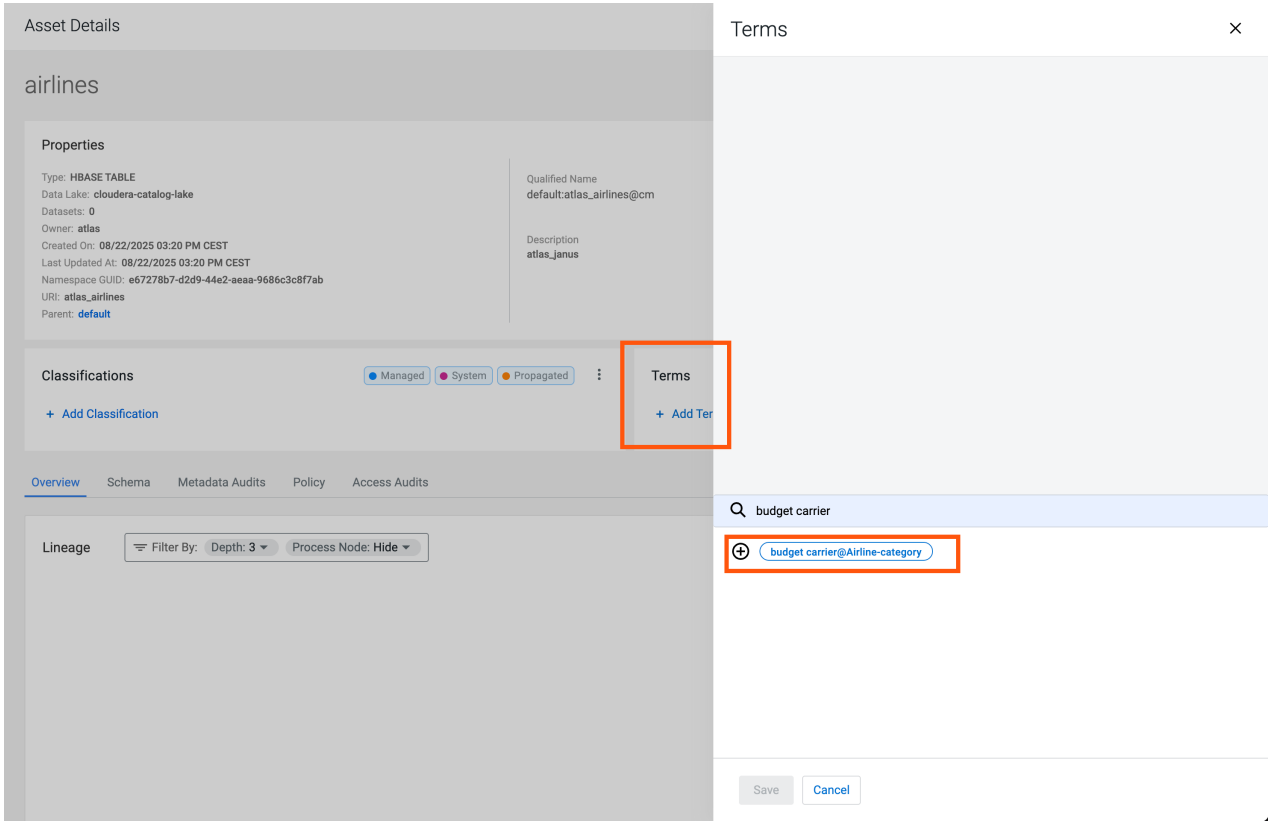


Figure 9: Glossary term in Asset Details



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 10: Mapping terms in Asset Details

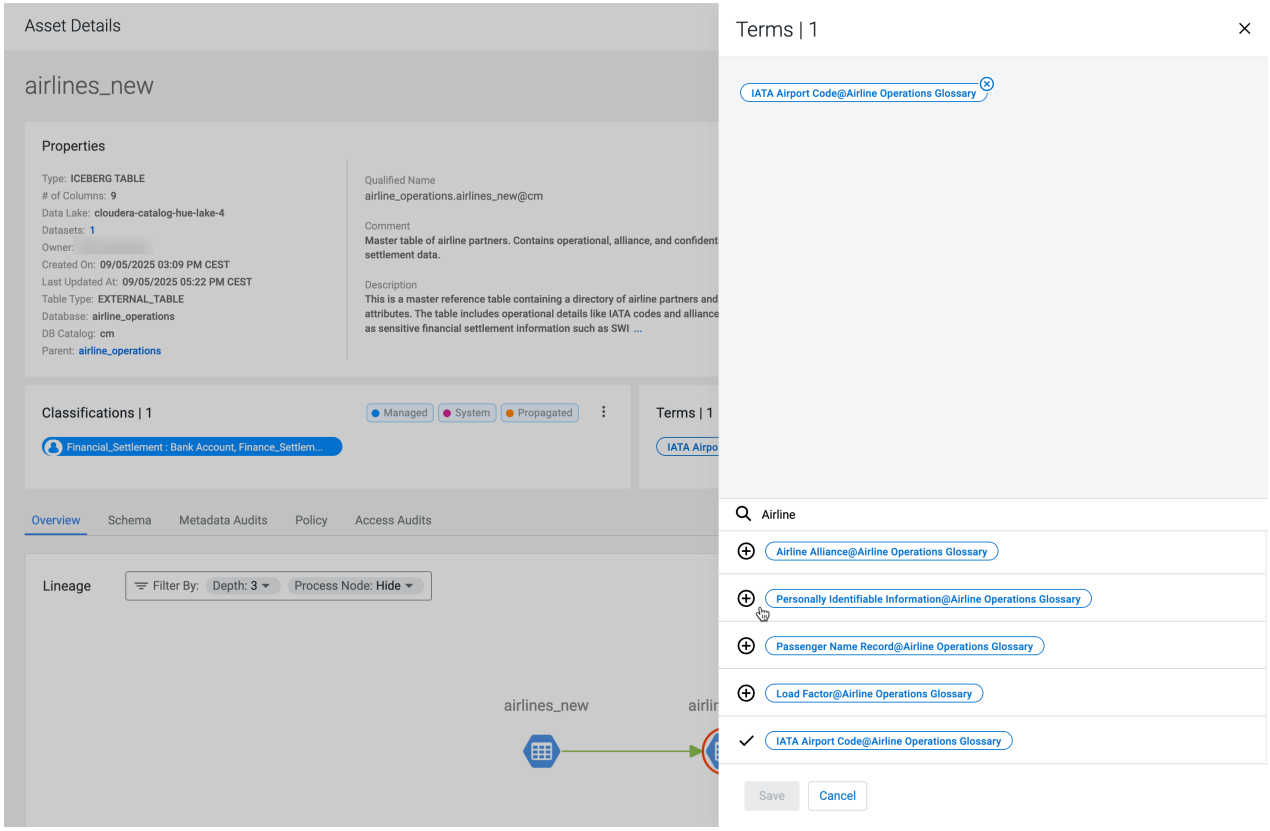
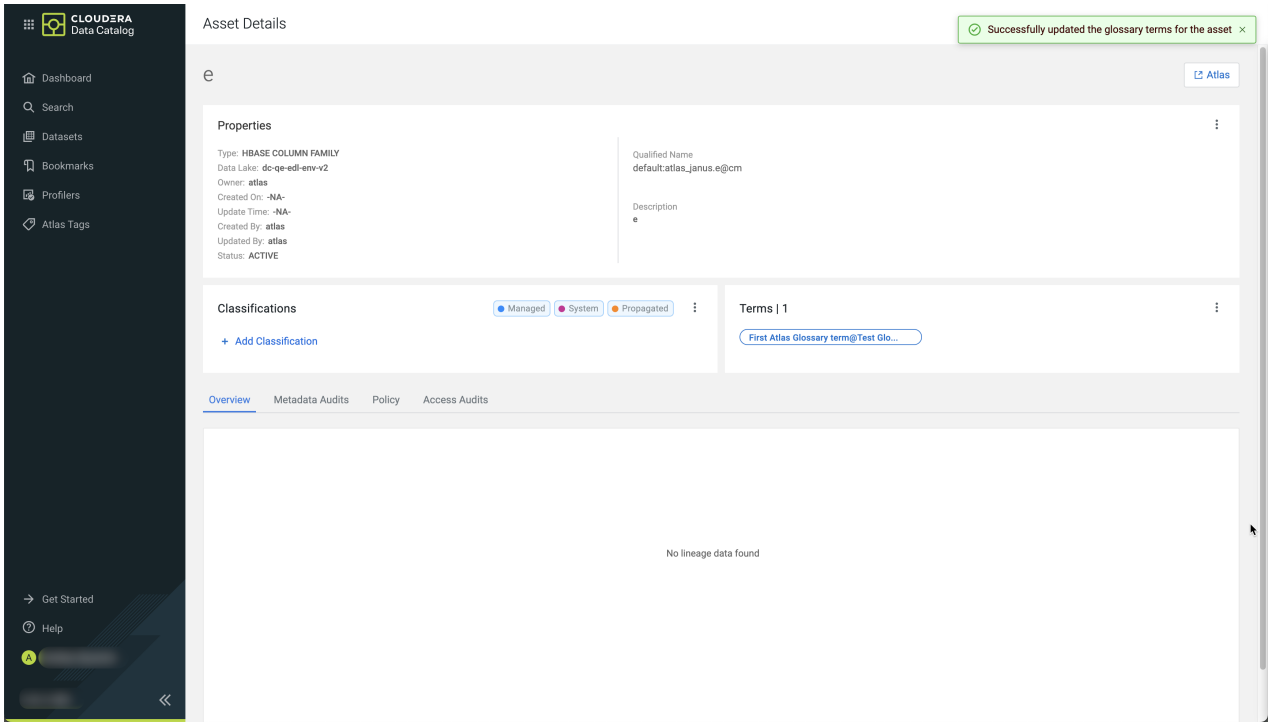
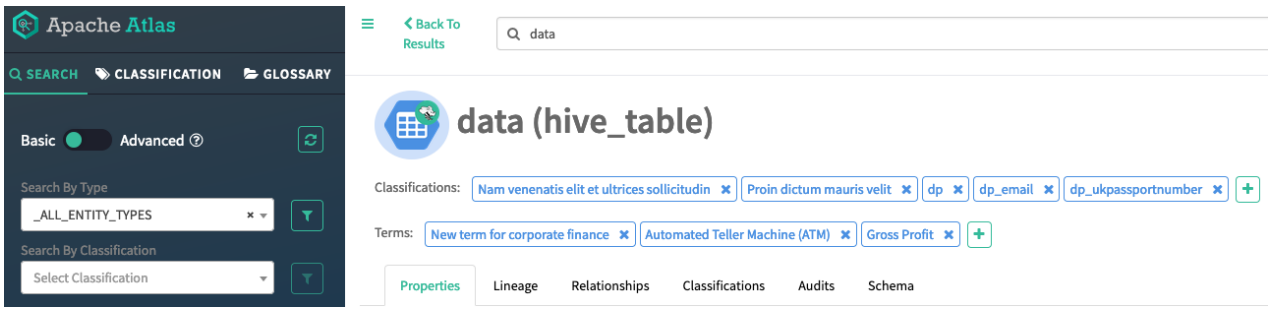


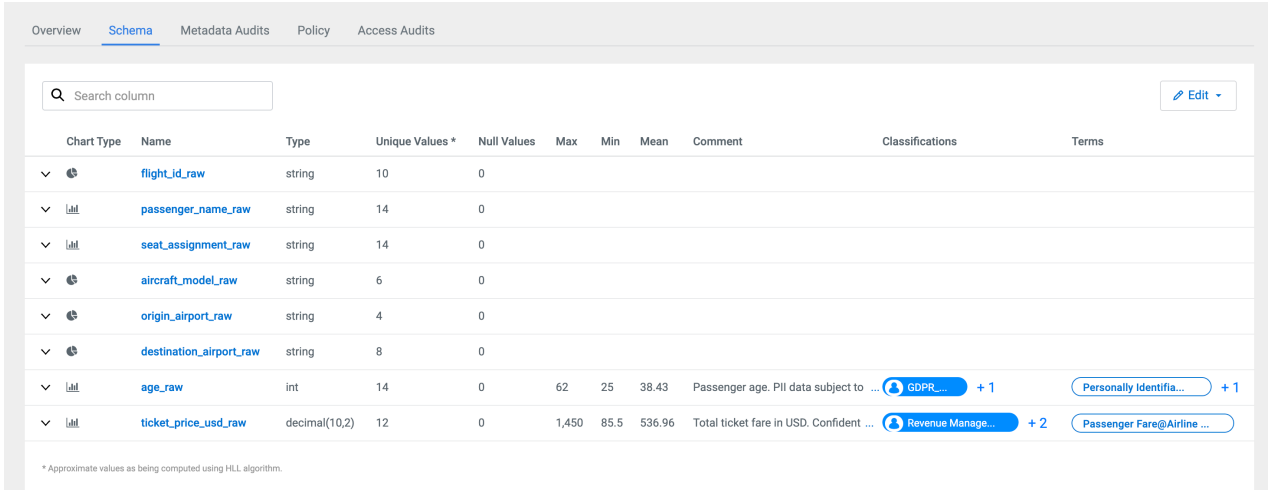
Figure 11: Mapped glossary term



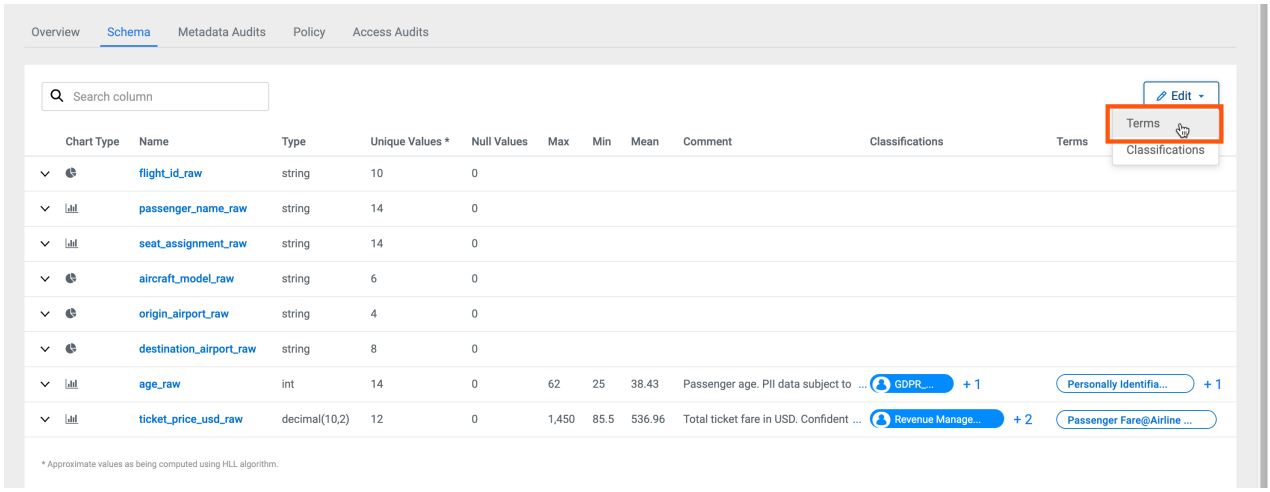
You can search for the same asset in the corresponding Atlas environment as shown in the example image.



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.



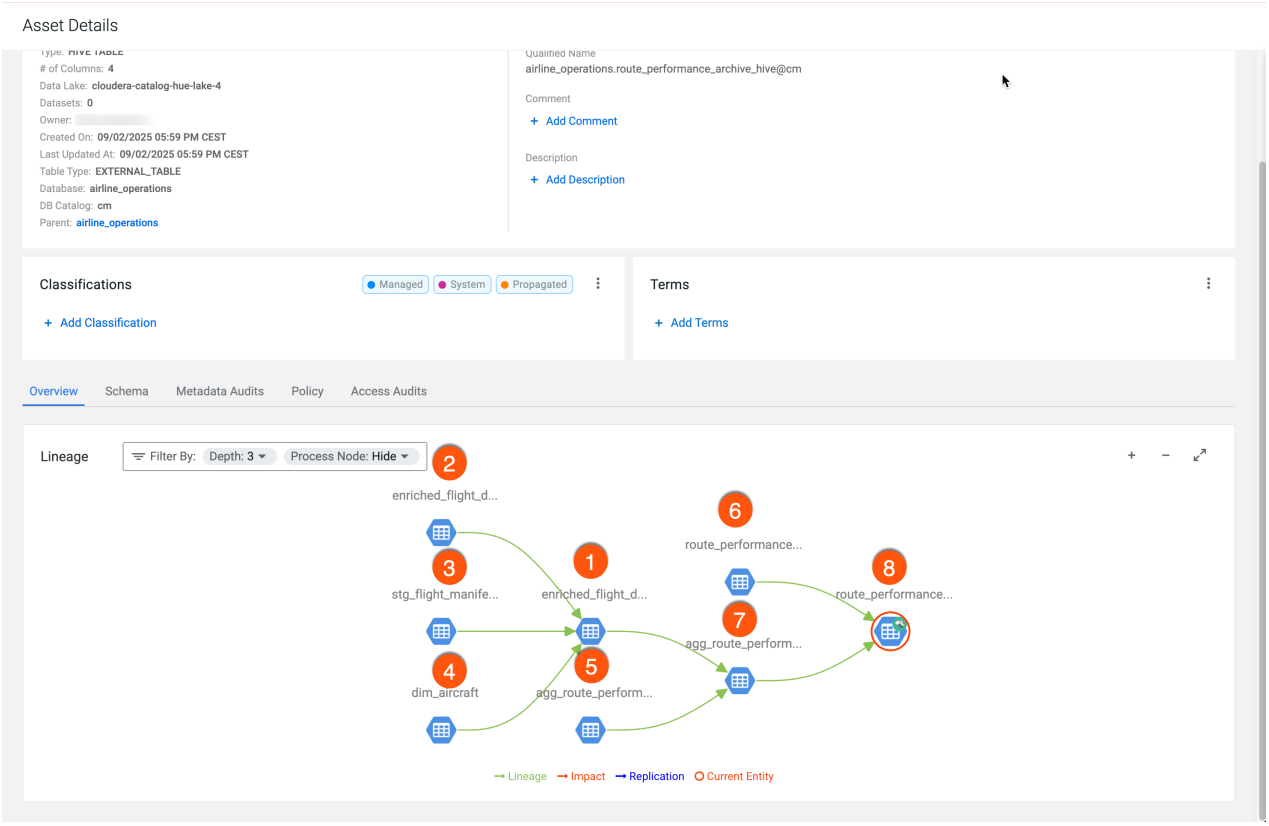
Related Information

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:



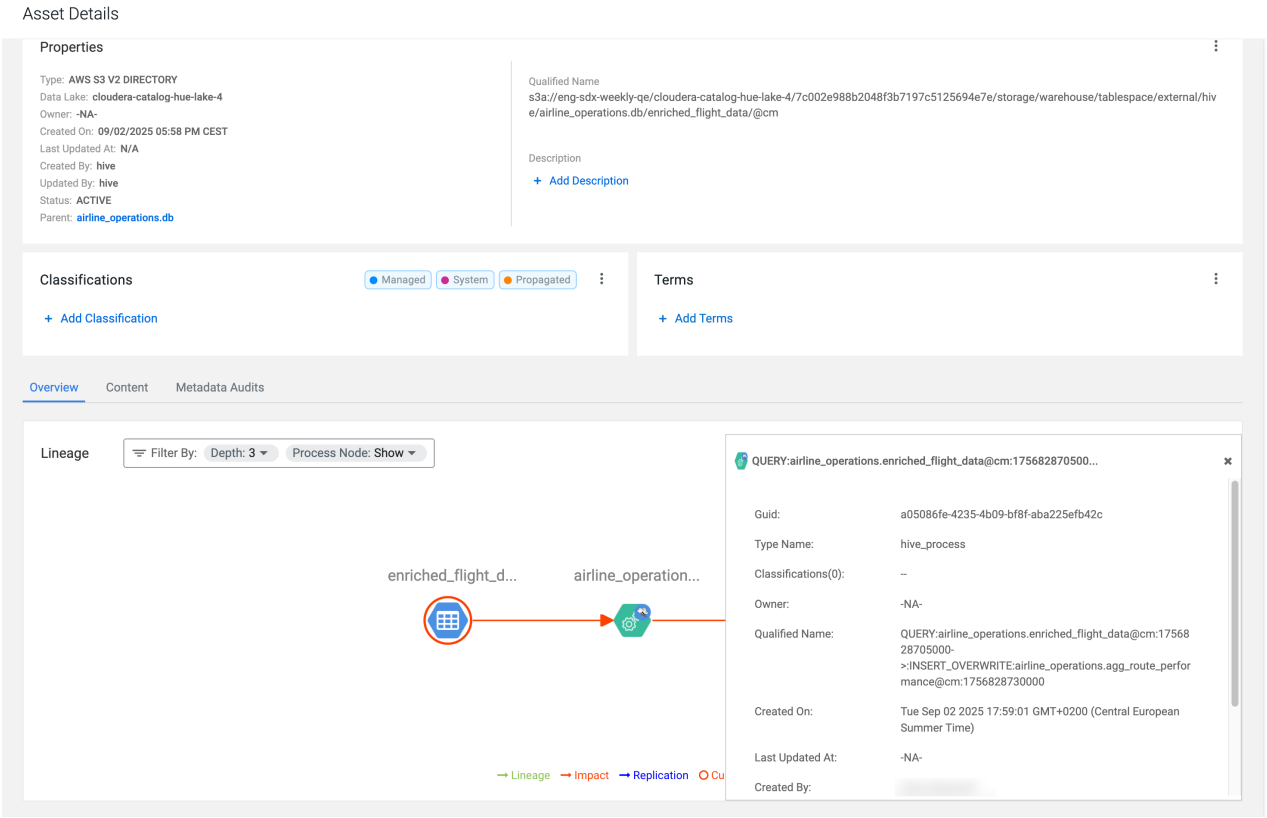
- The enriched_flight_data table (1) is created by joining the following source tables:
  - airline_operations.stg_flight_manifests@cm (3): This is the main source of transactional data.
  - airline_operations.dim_aircraft@cm (4): This is a dimension or lookup table. It provides supplemental information.
- ```
  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.

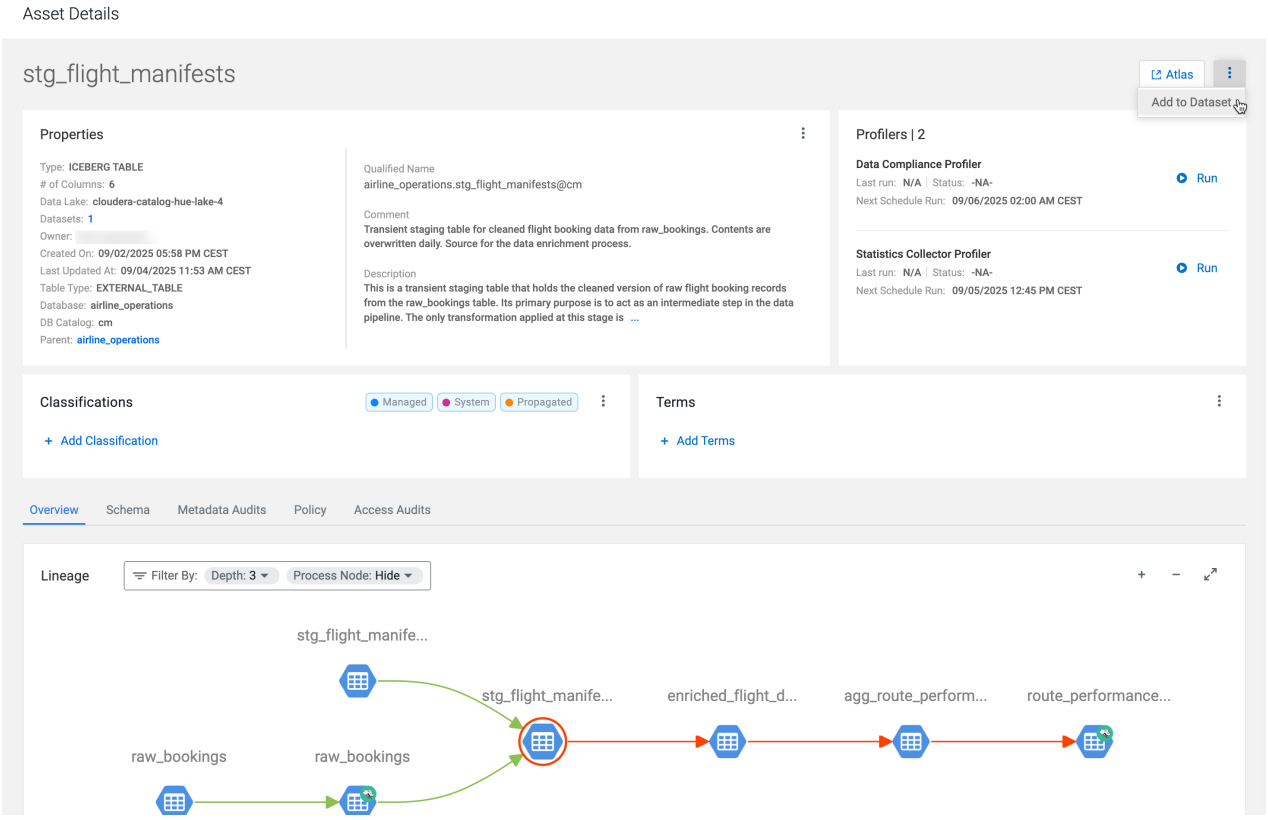
The screenshot depicts the slider information for the clicked entity:



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

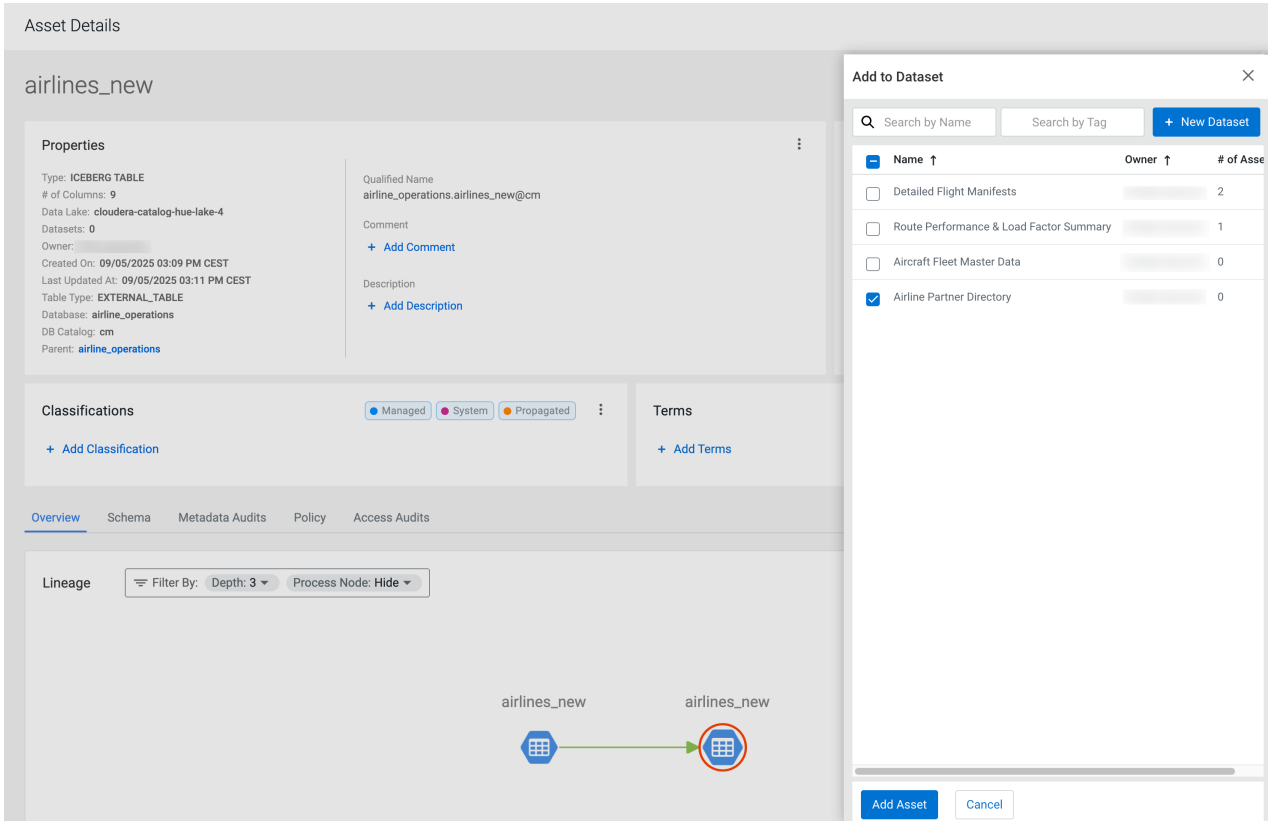
Adding assets to one or more datasets

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

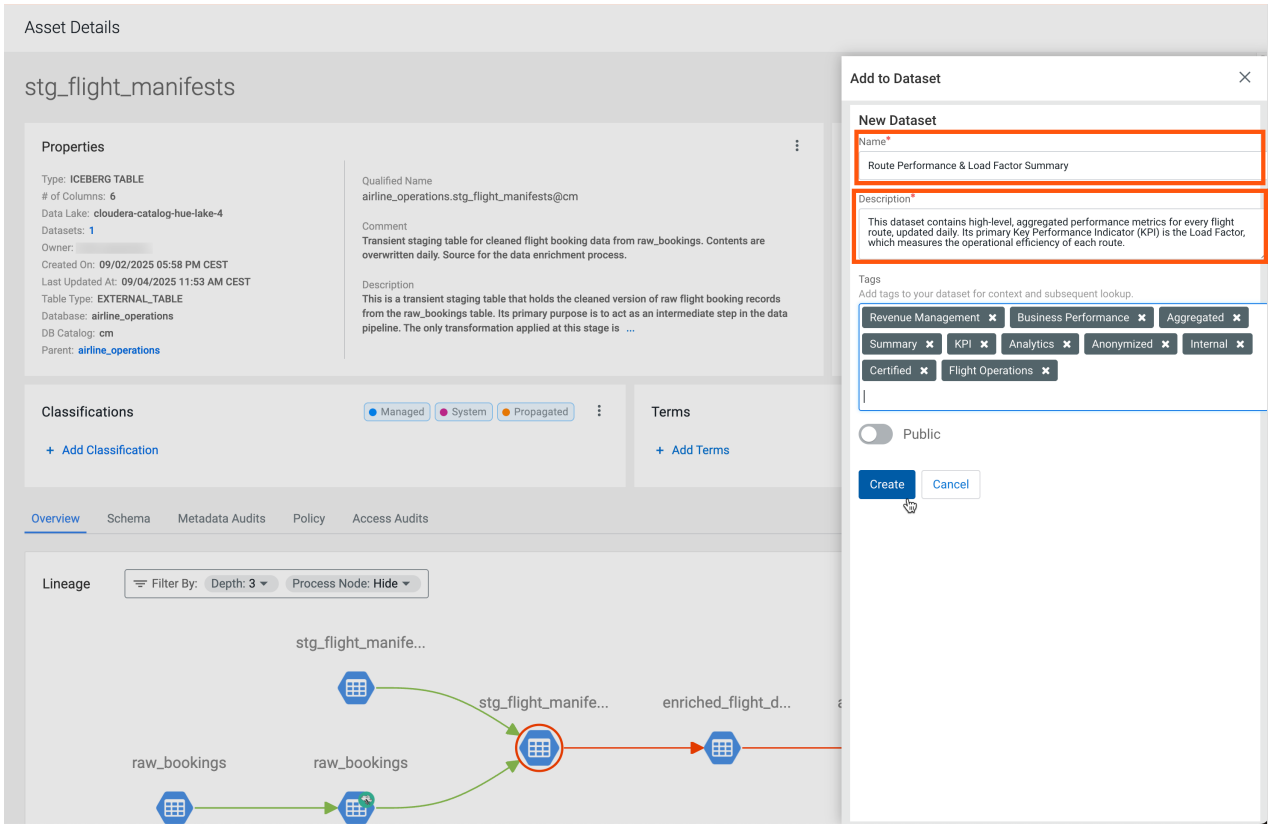


The **Add to Dataset** pane provides an option to add the asset into one or more existing datasets or even create a new one.

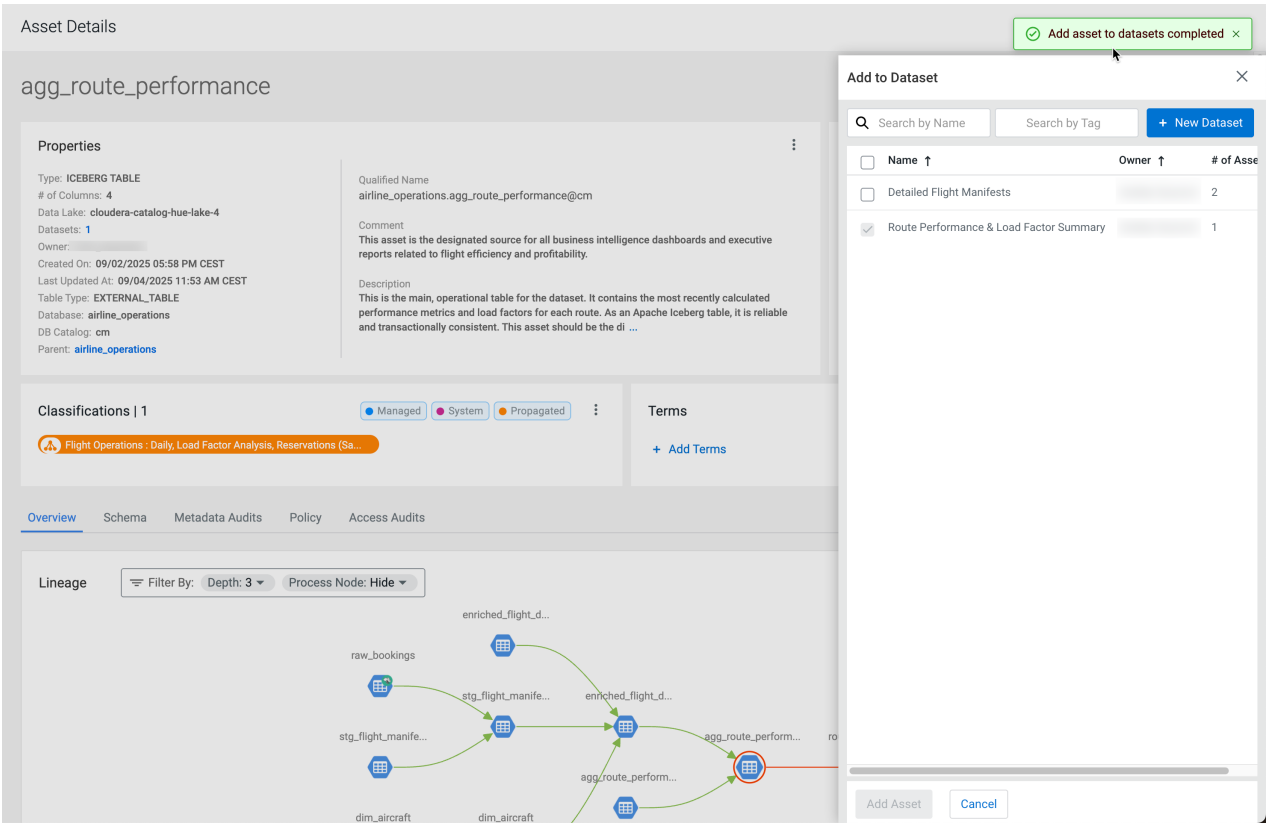
Datasets that already contain the asset are disabled and marked as checked. Datasets which are currently in edit state are disabled and marked with the characters a *.



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.

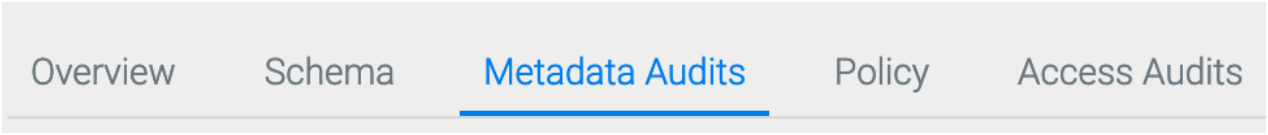


Viewing Atlas entity audits

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.



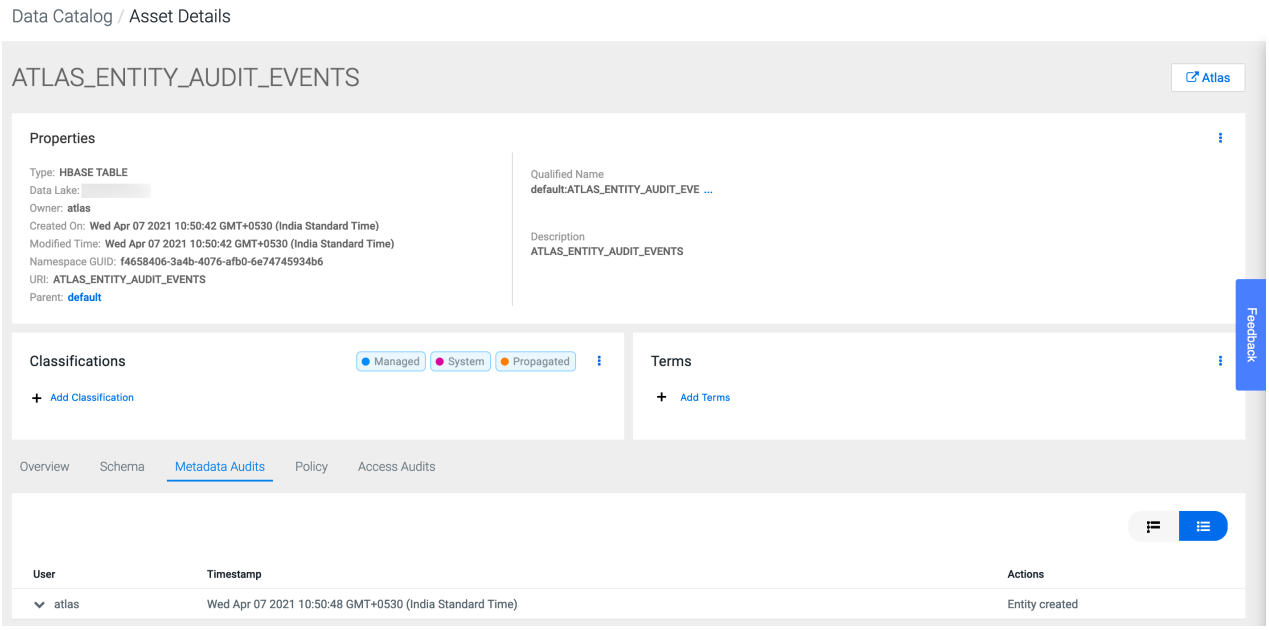
Clicking on **Metadata Audits**, tab, you can view manage information about:

- 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

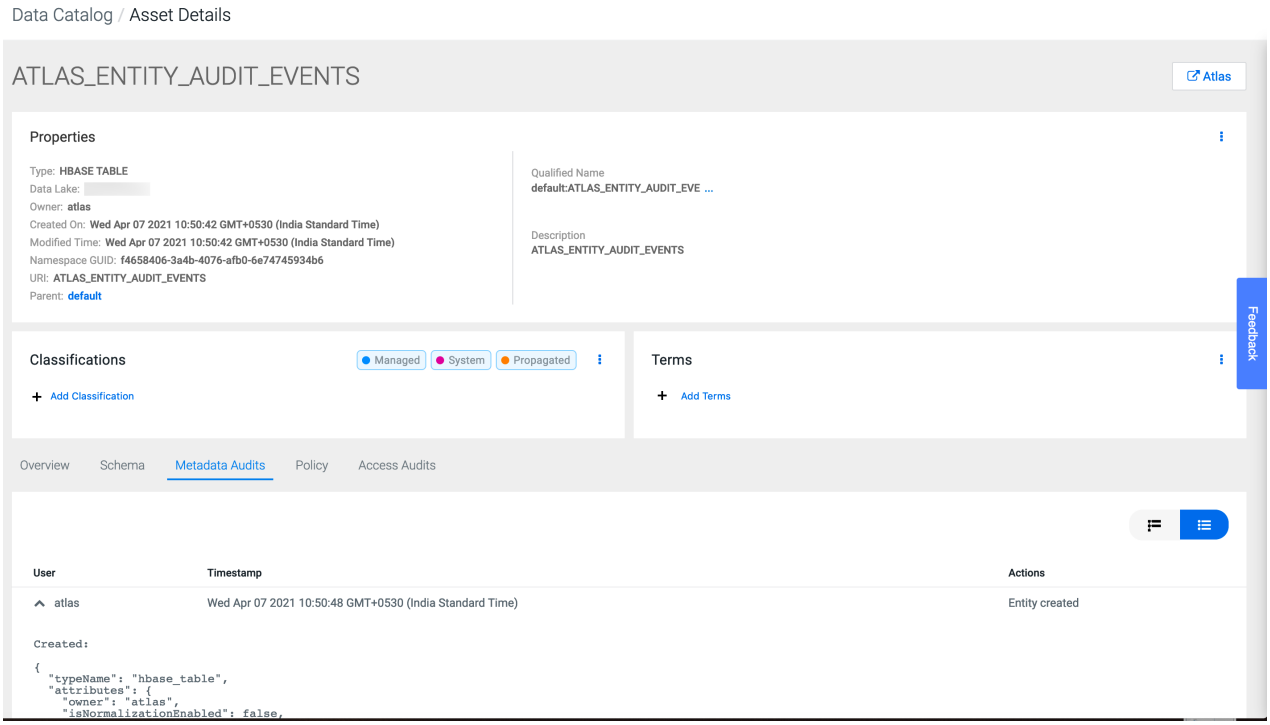
The changes that can be identified for:

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

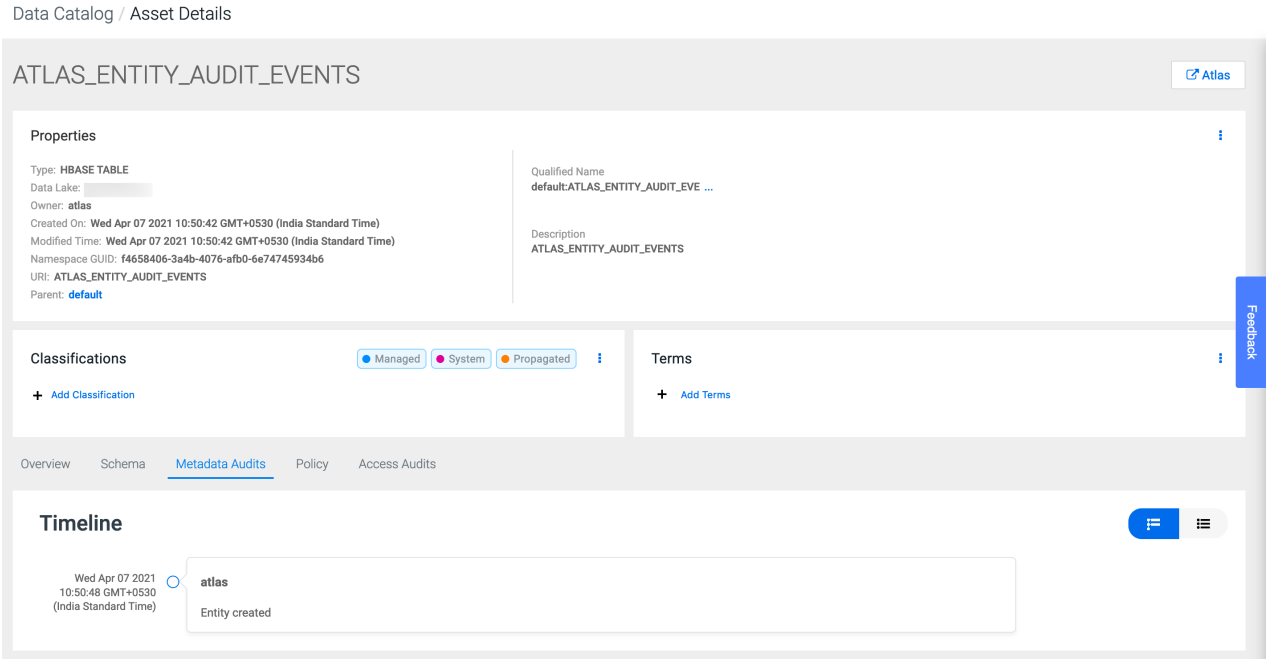
For example, the following image displays the Atlas audit entity creation event that is recorded by each Atlas entity that is displayed in the **Asset Details** page.



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.



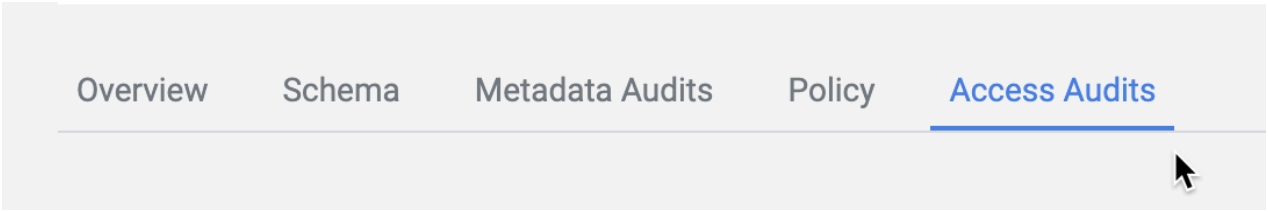
Clicking on a user in the **Timeline** format displays the JSON data.

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.

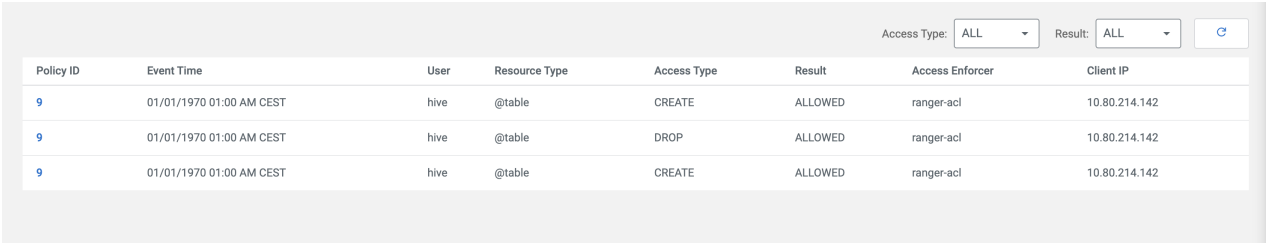
- The resource type accessed
- The access type
  - For example, by Hive entities the following filter options are available

    - SELECT
    - UPDATE
    - CREATE
    - DROP
    - ALTER
    - INDEX
    - READ
    - WRITE
- The access attempt result
    - ALL
    - ALLOWED
    - DENIED
- The client IP

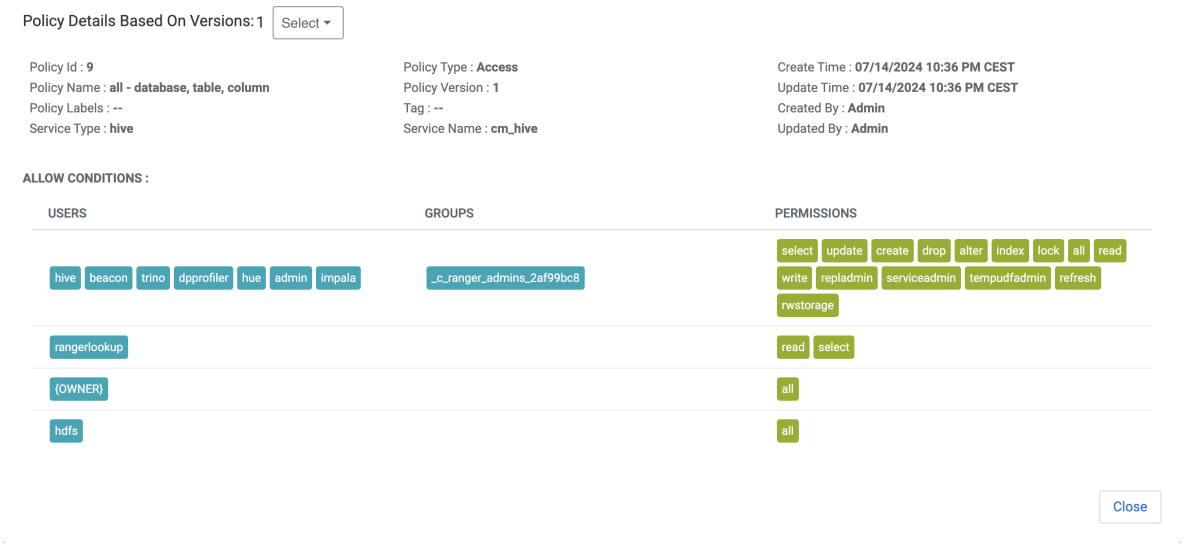
The accesses can be identified for:

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

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



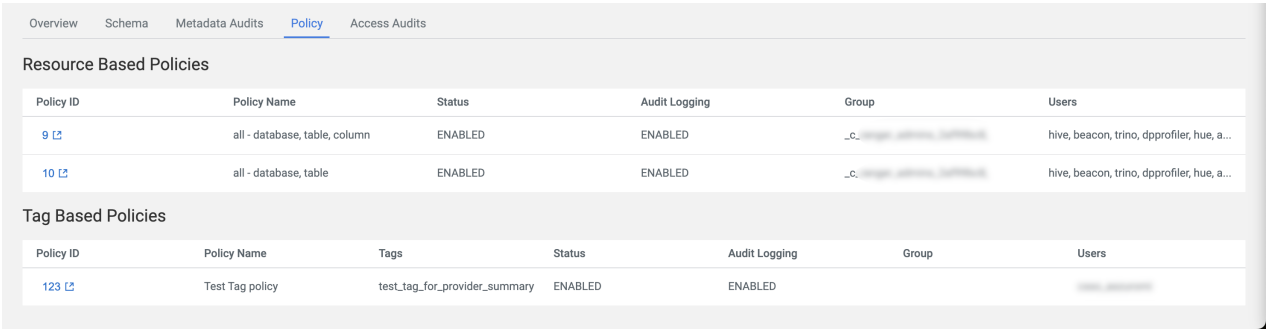
Note: If you click an event row, you can see an overview of the users set for that policy including the set permissions.



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.

If users are not able to view the table details, such as a column, an asterisks appears instead of the row in table.

Figure 12: Hive column read access denied

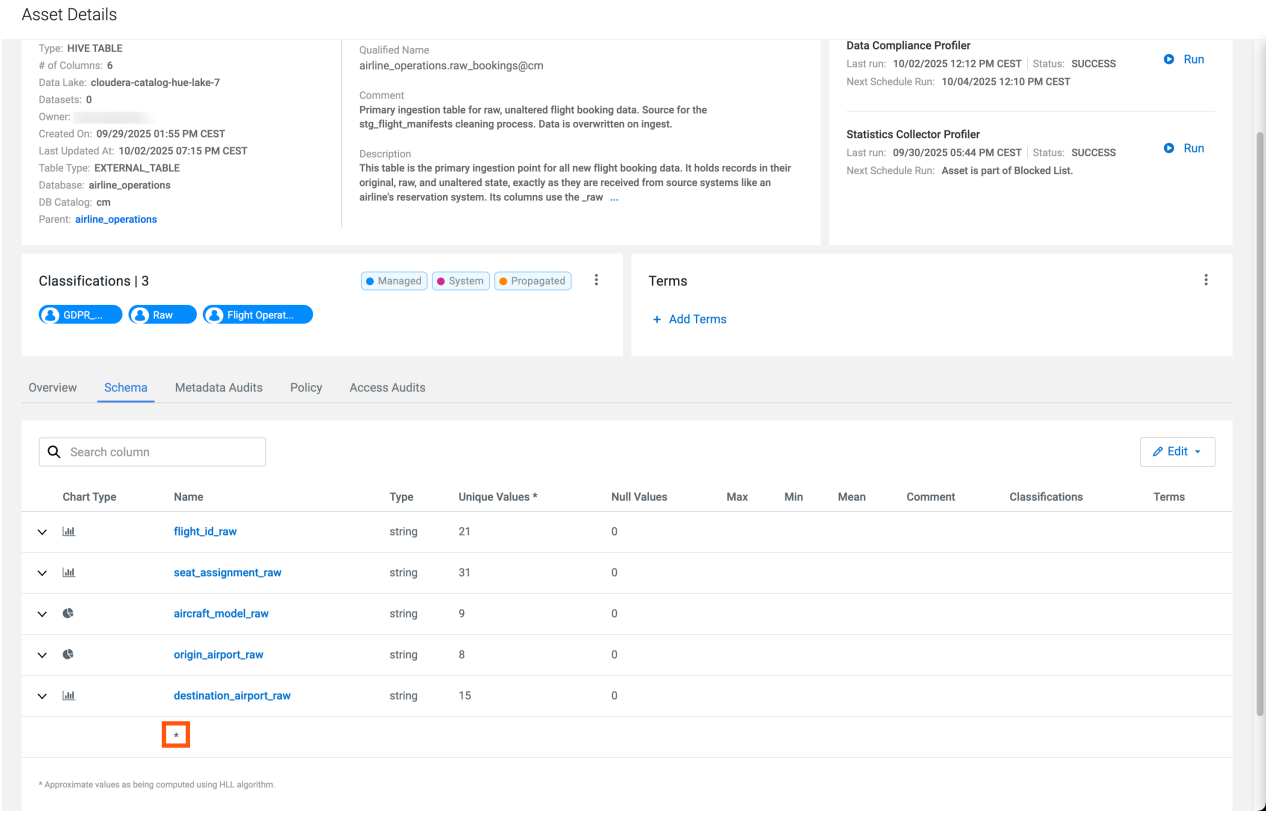
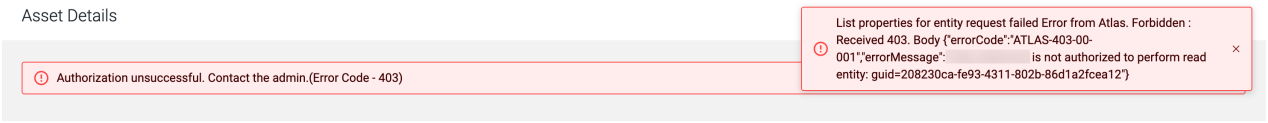


Figure 13: Hive table read access denied



Related Information

Providing role access
Ranger Policies Overview
Resource-based Services and Policies
Configuring resource-based policies
Missing authorization for viewing assets

Prerequisites for accessing Cloudera Data Explorer (Hue) tables and databases

Cloudera Data Explorer (Hue) is a part of the Cloudera Data Warehouse service. Before you can access Data Explorer (Hue) from Cloudera Data Catalog, you must set up the Cloudera Data Warehouse service.

Before you begin

You must have the DWAdmin role to perform these tasks. For more information, see Granting Cloudera users and groups access to CDW.

Procedure

1. Set up an AWS or Azure environment in Cloudera Data Warehouse that has been activated and registered in the Cloudera Management Console:

   Activating an AWS or Azure environment in Cloudera Data Warehouse enables using the existing Data Lakes that are present in the environment. This also creates a default Database Catalog that is used by a Virtual Warehouse.

   For more information, see Activating AWS environments or Activating Azure environments. Currently, Cloudera Data Warehouse only supports AWS and Azure Cloud environments
2. Set up a Hive or an Impala Virtual Warehouse created using Cloudera Data Warehouse. Depending on your data set type, create a Hive or an Impala Virtual Warehouse using Cloudera Data Warehouse.

   For more information, see Adding a new Virtual Warehouse.
3. Select the asset type.

   The asset type must be either a Hive Table or Hive DB.

   After these conditions are met, Data Explorer (Hue) links for the selected assets are displayed in Cloudera Data Catalog.

Accessing Cloudera Data Explorer (Hue) assets

Cloudera Data Catalog helps you to explore assets within your Data Lake. Cloudera Data Explorer (Hue) allows you to run SQL queries on Hive assets. The integration between Cloudera Data Catalog and Cloudera Data Warehouse service provides a direct link to the Data Explorer (Hue) instance from the Search page or the Asset Details.

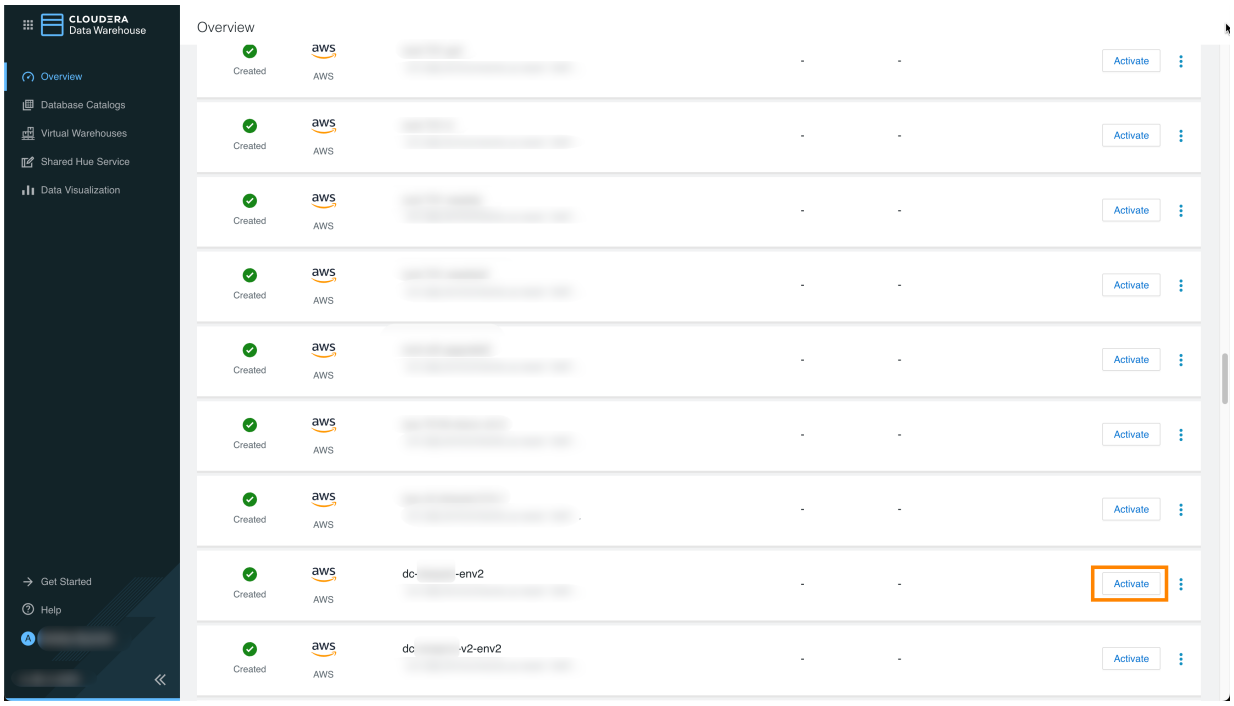
Before you begin

You must fulfill all conditions listed in Prerequisites for accessing Cloudera Data Explorer (Hue) tables and databases on page 34 for viewing the external link button.

You must activate Cloudera Data Warehouse on your selected environment.

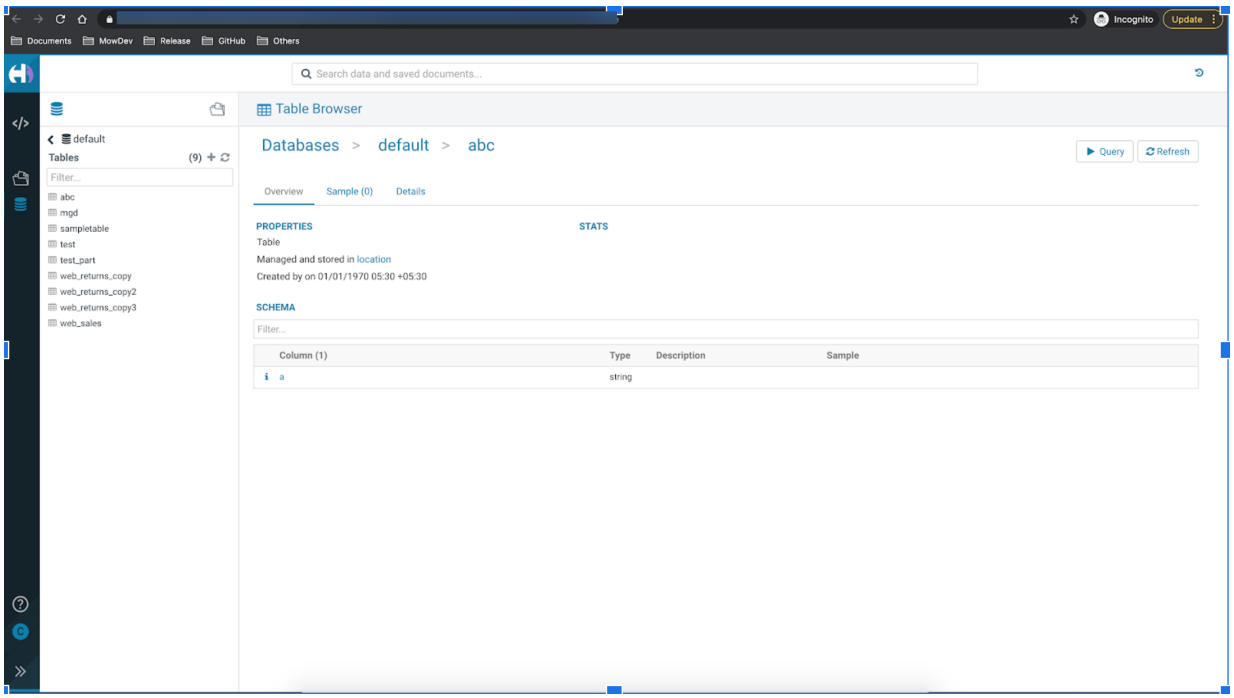
Procedure

1. Go to Cloudera Data Warehouse.
2. On the Overview page, search for your AWS or Azure environment.
3. Locate the enabled AWS / Azure environment and click Activate.
   This action also creates a default Database Catalogs.



4. Once the environment is activated and a DB Catalog is created and running, you can create a virtual warehouse as per the requirements.

   For more information about AWS and Azure Environments, see AWS environments overview and Azure environments overview.



You will now be able to navigate to the Data Explorer (Hue) assets on these virtual warehouses using the links provided by Cloudera Data Catalog.

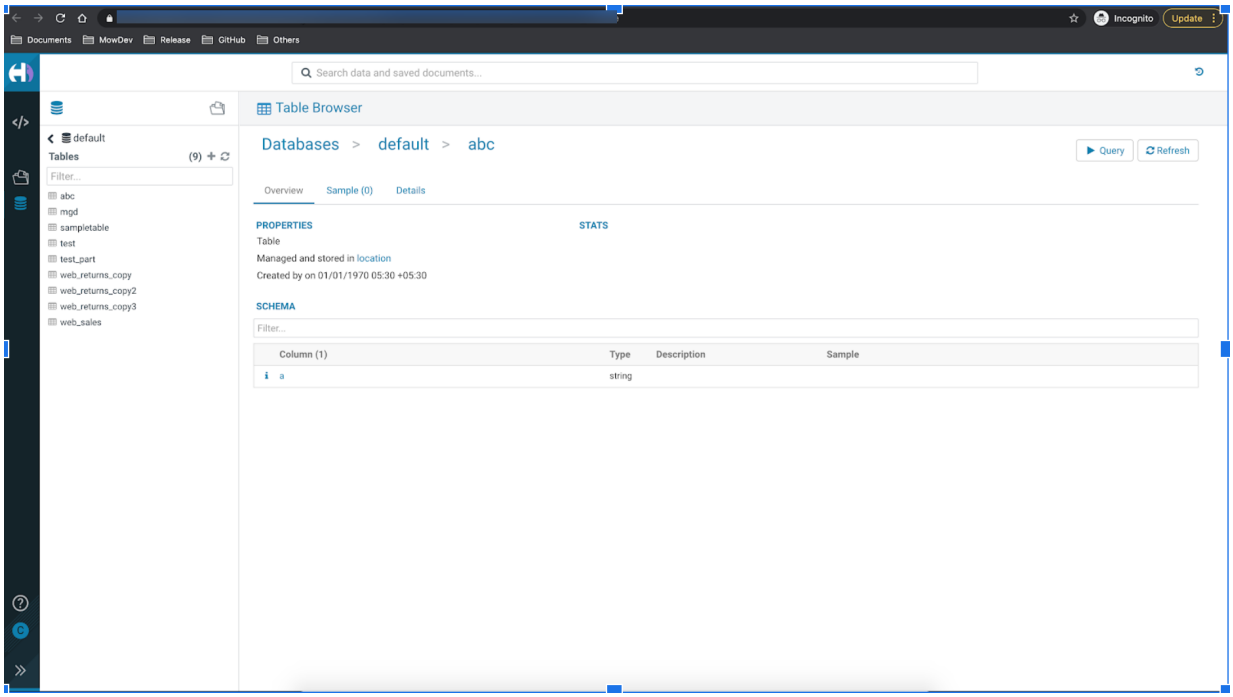
Accessing Cloudera Data Explorer (Hue) tables and databases from the Search page

When you log into Cloudera Data Catalog, you can access web links to the Data Explorer (Hue) instances from the Search page to run SQL queries on your assets.

Procedure

1. Log into Cloudera Data Catalog.
2. Select the Data Lake associated with your environment.
3. Select the asset type under Filters, for example Hive Table.
   Hive tables present within the Data Lake are displayed.
4. Click the icon at the end of the row for a table that you want to display in Data Explorer (Hue).

   A Link to Experiences pop-up appears. It contains a list of Data Explorer (Hue) instances within the virtual warehouses that are associated with this particular table.
5. Click the link to open the Data Explorer (Hue) web interface.
   The Hive table is displayed on the Data Explorer (Hue) Table Browser.



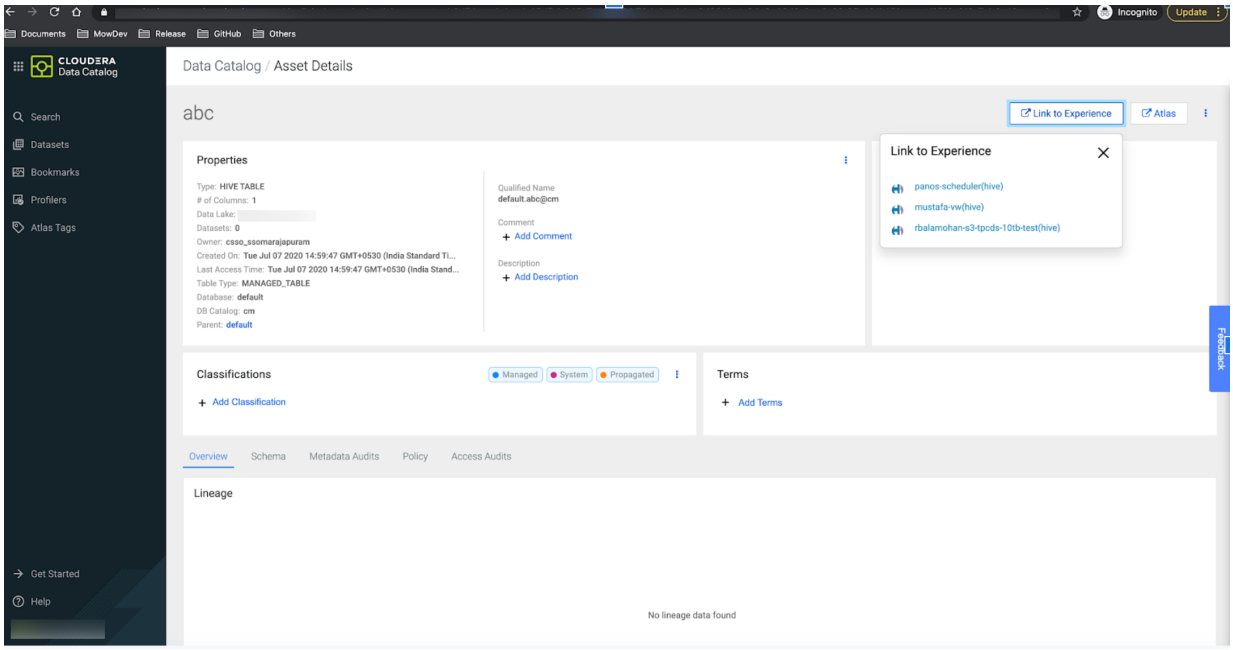
Accessing Cloudera Data Explorer (Hue) tables and databases from the Asset Details page

When you log into Cloudera Data Catalog, you can access web links to the Data Explorer (Hue) instances from the Asset Details page to run SQL queries on your assets.

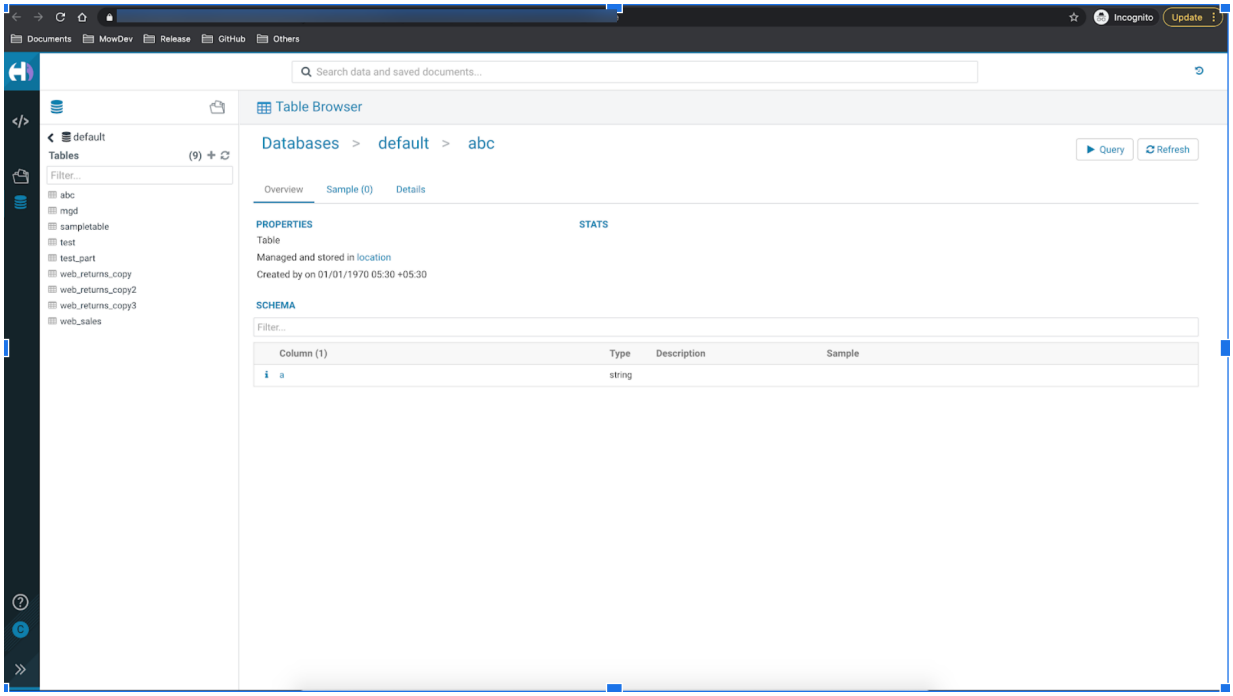
Procedure

1. Log into Cloudera Data Catalog.

2. Select the Data Lake associated with your environment.
3. Select the asset type under Filters, for example Hive Table.
4. Click on the asset name to open the **Asset Details** page. The **Asset Details** page contains the Link to Experience button as shown in the following image:



5. Click Link to Experience.
   The pop-up contains a list of Data Explorer (Hue) instances within the Virtual Warehouses that are associated with this particular table.
6. Click the link to open the Data Explorer (Hue) web interface.
   The Hive table is displayed on the Data Explorer (Hue) Table Browser.



Integrating Cloudera Data Catalog with AWS Glue Data Catalog

Integrating Cloudera Data Catalog with AWS Glue Catalog enables the users to browse and discover data as well as register data into Cloudera Shared Data Experience (through metadata translation or copy), so that it can be used with Cloudera Data Hub and other relevant experiences.

Note:

- This feature is only available in VM-based environments.
- This feature is available but is not ready for production deployment. Cloudera encourages you to explore this technical preview feature in non-production environments and provide feedback on your experiences through the Cloudera Community Forums or through your account managers and field team contacts.

While using AWS Glue in Cloudera Data Catalog, you will be able to experience a complete snapshot metadata view, along with other visible attributes that can power your data governance capabilities.

How integration works

Assuming that the Cloudera Shared Data Experience is running in the users' AWS account (that contains the same AWS account which has Glue Data Catalog and the data that has to be discovered), the credentials with the Exte rnalDataDiscoveryService (which is hosted in Cloudera Shared Data Experience) must be shared, so that these two entities can interact with each other. These credentials are used to launch Cloudera Shared Data Experience and other workload clusters on the users' AWS account.

Prerequisites:

- You must have full access to AWS Glue Catalog and also have access to the EMR cluster's Hive Metastore instance.
- You must set up the Cloudera platform.
- You must have access to your AWS IT Admin and CDP Admin user credentials, which is required to enable the Cloudera platform to access AWS/EMR managed data in Cloudera.

Note: AWS policies are managed by AWS IAM and the AWS roles are added to the Cloudera Management Console. Refer to Using Instance Profile in AWS and Using credentials in Cloudera Management Console. For more information about the AWS access, see AWS Environments.

Setting up AWS Glue Catalog with Cloudera Data Catalog

You must map your Cloudera Data Catalog instance with AWS Glue Catalog.

Procedure

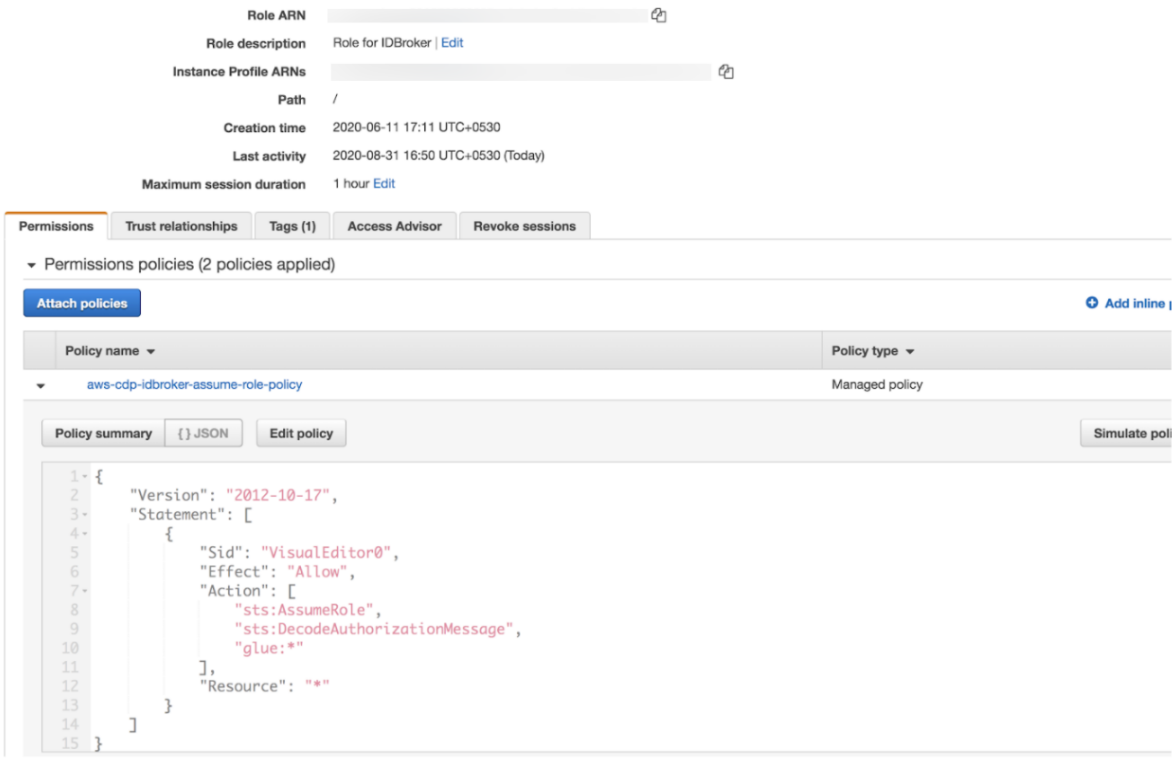
1. Enable the entitlement for your Cloudera Data Catalog instance by running the following command on your Cloudera environment. For example:

```
$ cdp coreadmin grant-entitlement --entitlement-name
DATA_CATALOG_ENABLE_AWS_GLUE --account-id {account_id}
```

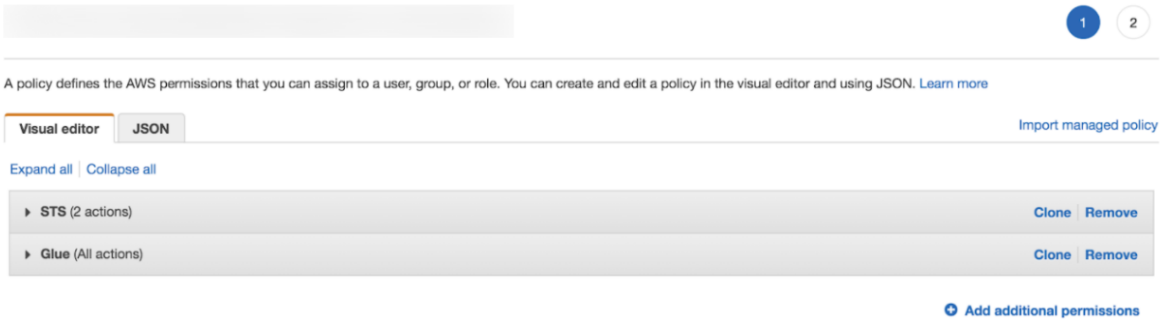
2. Add the relevant permissions in the corresponding AWS account:
   a) Include permission to access Glue Catalog service by editing the policy accordingly.

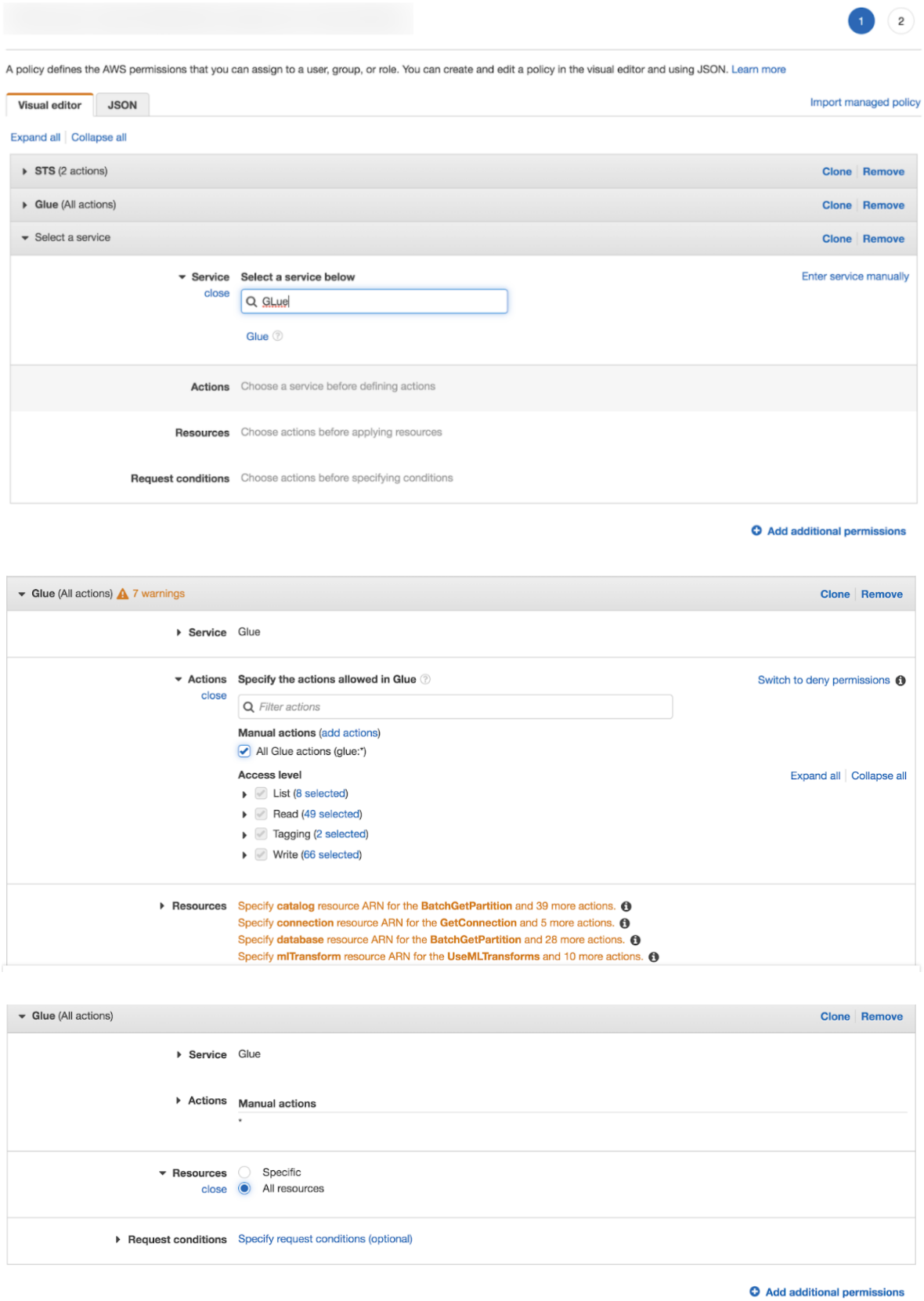
      Make a note of the Assumer Instance Profile role that you intend to use and include full access authorization for AWS Glue.

      Refer to the following images as a guidance to complete the set up.



Note: For Role ARN and Instance Profile ARNs, you must include the appropriate account number and role respectively.





b) Search for the role attached to the Instance Profile of the Cloudera environment. Use the Instance Profile that you have configured above with Glue related policy in your AWS Environment creation command.

Use the following examples to setup AWS environment and AWS Data Lake as part of the Glue setup:

```
cdp environments create-aws-environment --profile default --cli-input-js
on '
{"environmentName”:”ab-ds-cli-7321”,
 "credentialName”:”cd2d-1234”,
 "Region":"us-region-2”,
 "securityAccess":{-insert the value--"},
 "Authentication":{---insert the value---"},
 "logStorage":{"storageLocationBase":"s3a://demo-e2e-test-state-bucket/
ab-ds-cli-7321/logs","instanceProfile":"arn:aws:iam::<xxxxxxxxxxx>:insta
nce-profile/<role-name>"},
 "vpcId":"vpc-0123456”,
 "subnetIds":["subnet-04fe923b902aa5cf2","subnet-099c7a631f0ebed3c"],
 "s3GuardTableName":"dc-pro-cli-7210",
 "Description":"ab-ds-cli-7321",
"enableTunnel":false,
 "workloadAnalytics":false,
 "freeIpa":{"instanceCountByGroup":1},
 }'

cdp environments set-id-broker-mappings \
--environment-name "ab-ds-cli-7321" \
--profile default \
--set-empty-mappings \
--data-access-role arn:aws:iam::<xxxxxxxxxxx>:role/add-role \
--ranger-audit-role arn:aws:iam::<xxxxxxxxxxx>:role/add-role
```

Similarly, while setting up the Data Lake use the Instance Profile that you configured above with Glue related policy in your Data Lake creation command:

```
cdp datalake create-aws-datalake --profile default --runtime 7.2.12 --cl
i-input-json '
{"datalakeName":"ab-ds-cli-7321-sdx",
 "environmentName":"ab-ds-cli-7321",
 "cloudProviderConfiguration":{"instanceProfile":"arn:aws:iam::<xxxxxx
xxxxx>:instance-profile/<role-name>","storageBucketLocation":"s3a://demo
-e2e-test-state-bucket/ab-ds-cli-7321"},
 "scale":"LIGHT_DUTY",
 }'
```

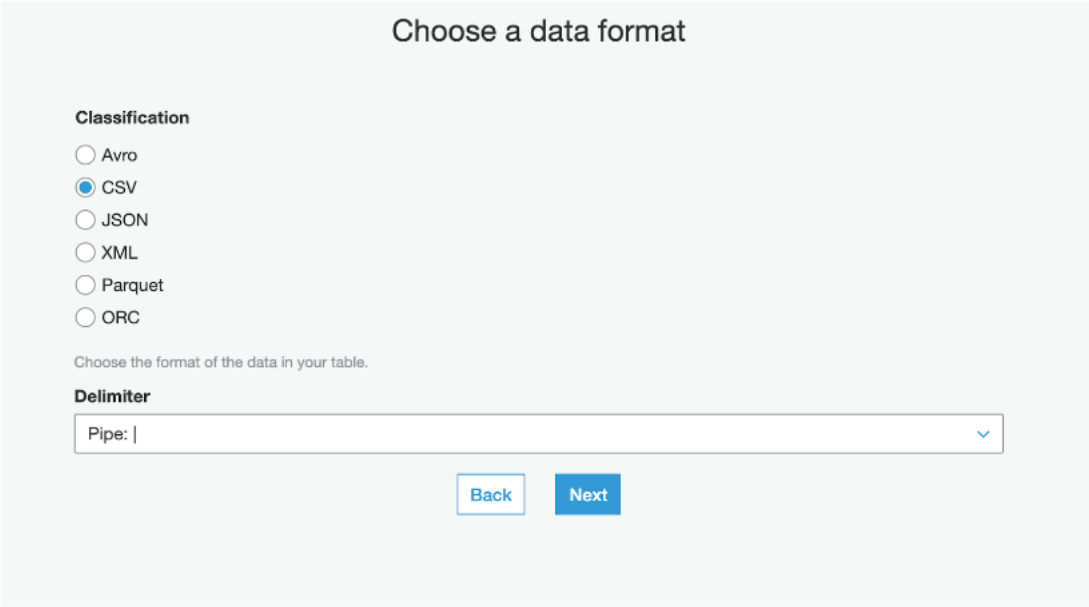
For more information, see Creating an AWS environment with a medium duty Data Lake using the CLI.

c) Navigate to the attached policy for the role.

d) When you manually create tables in AWS Glue Data Catalog, you must set the **fully qualified path** for the table location.

For example: s3://my-aws-server-node-1/something/something.amazonaws.com/dc-pro-721-storage/glue/

3. 3. You must set up the AWS Glue Data Catalog. For more information, see Populating the Glue Data Catalog. You must select only the CSV format which is currently supported for Cloudera Data Catalog and the delimiter which is used in the data.



4. While creating tables in AWS Glue Data Catalog manually, set the fully qualified path for location. For example: s3://my-aws-server-node-1/something/something/dc-pro-721-storage/glue/

What to do next

AWS Glue metadata must be registered with Cloudera Data Catalog.

Working with AWS Glue

The AWS Glue metadata must be registered with the Cloudera Data Catalog. The Glue contains the metadata that is synchronized with Cloudera Data Catalog. The Glue metadata is accessed using the data lake option in the Cloudera Data Catalog. After setting up AWS Glue with Cloudera Data Catalog and once the Glue synchronization with Cloudera Data Platform is complete, the AWS Glue data lake appears.

Note: For each AWS environment, there would be a separate listing under the data lake option. The Glue data lake name / identifier in Cloudera Data Catalog follows the format: <glue: Data Lake Name>. The Glue assets are of the type: GLUE EXTERNAL TABLE

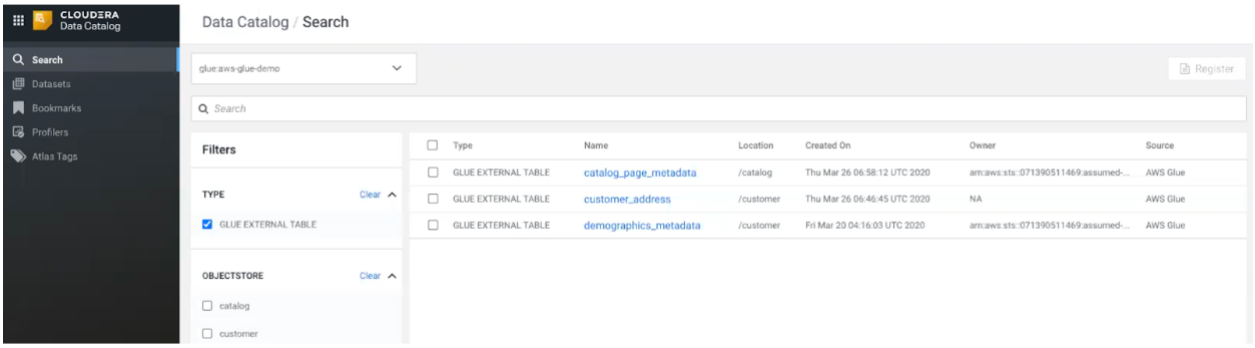
Accessing AWS Glue

To access the Glue metadata in Cloudera Data Catalog, you must note the following in your instance.

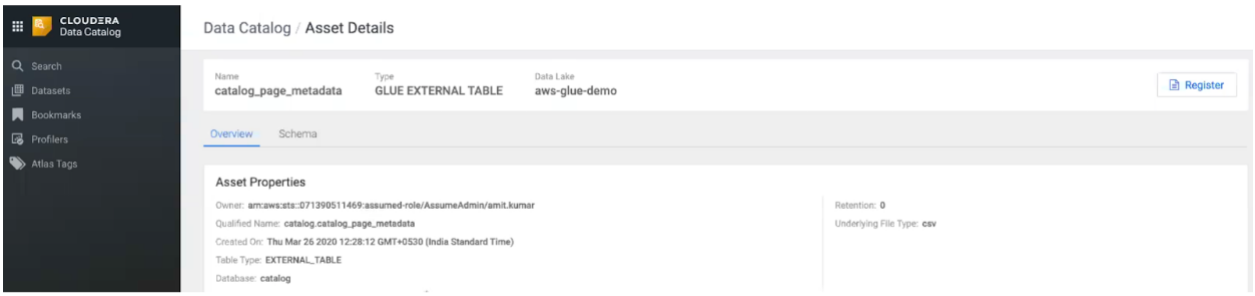
- List the Glue Metadata by selecting the Glue data lake.
- Select one or more Glue assets and register the same with Cloudera platform.
- Verify if the registered Glue assets are listed in the Cloudera Data Catalog owned data lake.
- Select the registered Glue asset and click to open the **Asset Details** page.

Listing the Glue assets

In Cloudera Data Catalog, when you select the AWS Glue data lake, you can view the list of Glue assets. These metadata assets are directly sourced from Glue.



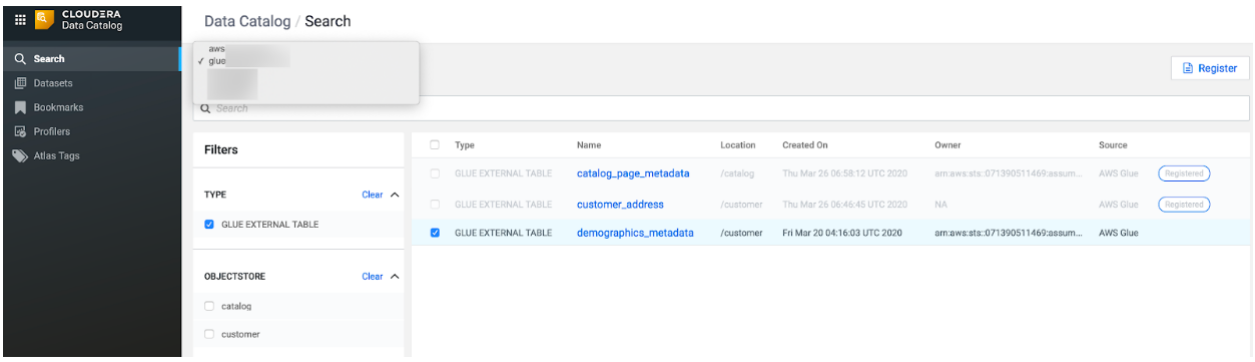
When you click on one of the assets, the **Asset Details** page is displayed.



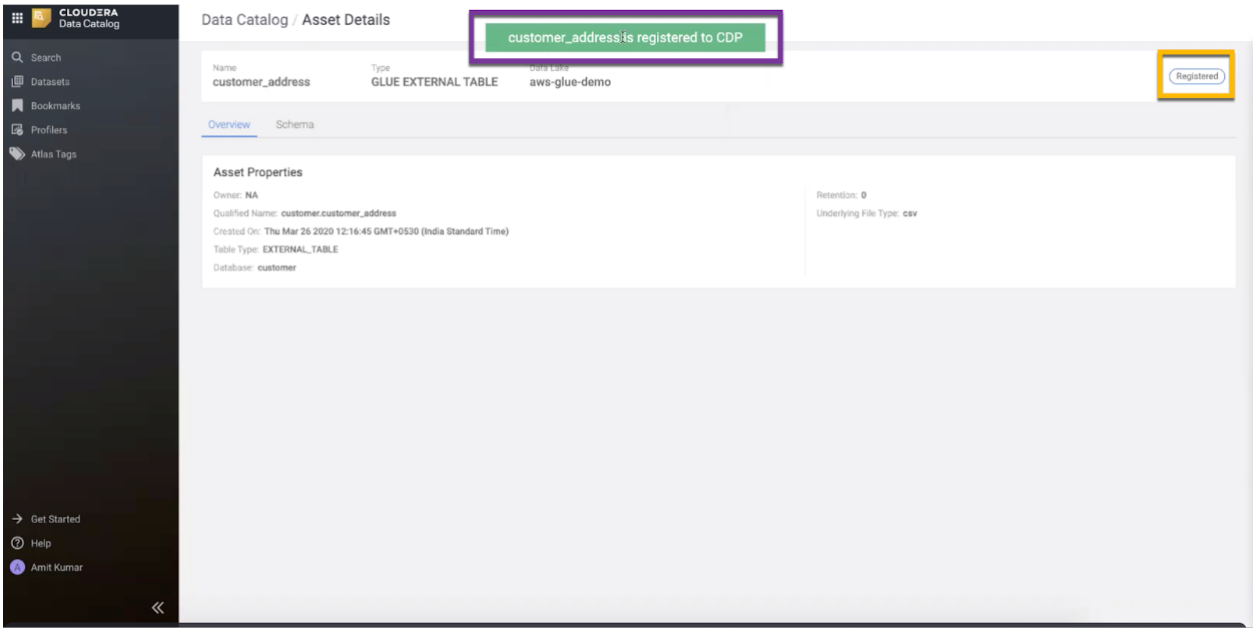
Next, on the main Cloudera Data Catalog page, you must select the Glue data lake and select one of the Glue assets and register the asset to the Cloudera platform. Click Register.

Optionally, you directly click on the Glue asset and register the asset on the **Asset Details** page.

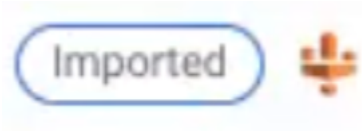
Note: You can select one or more Glue assets on the asset listing page to register them in Cloudera.

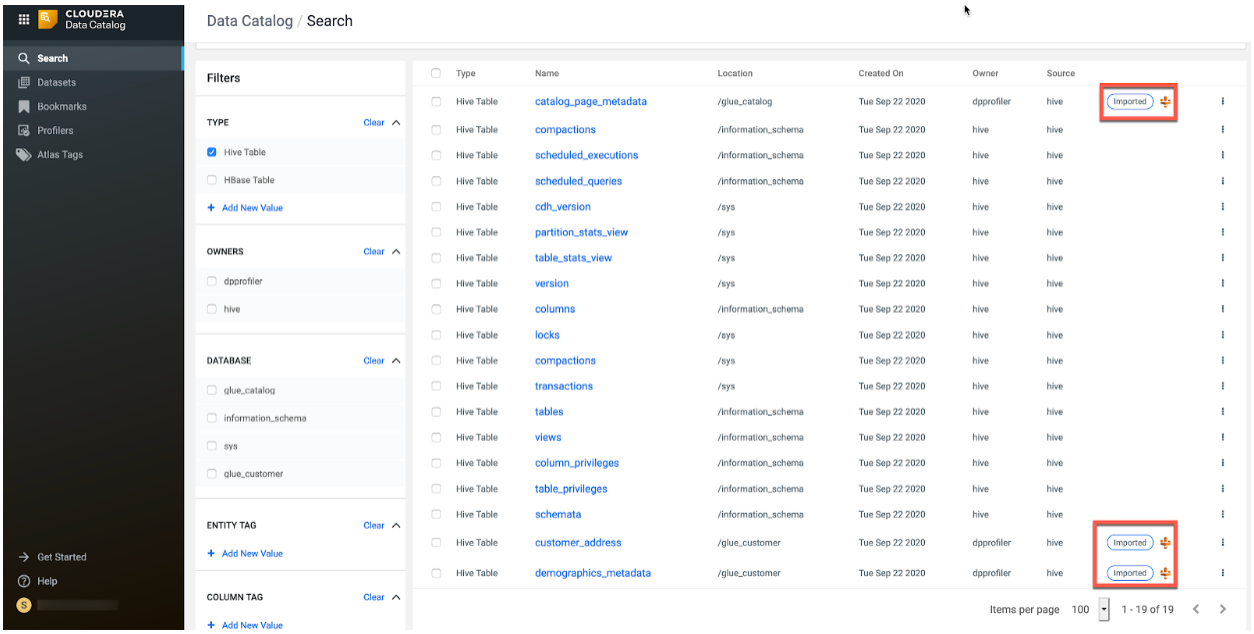


Once the Glue asset is registered, the asset is imported into Cloudera.



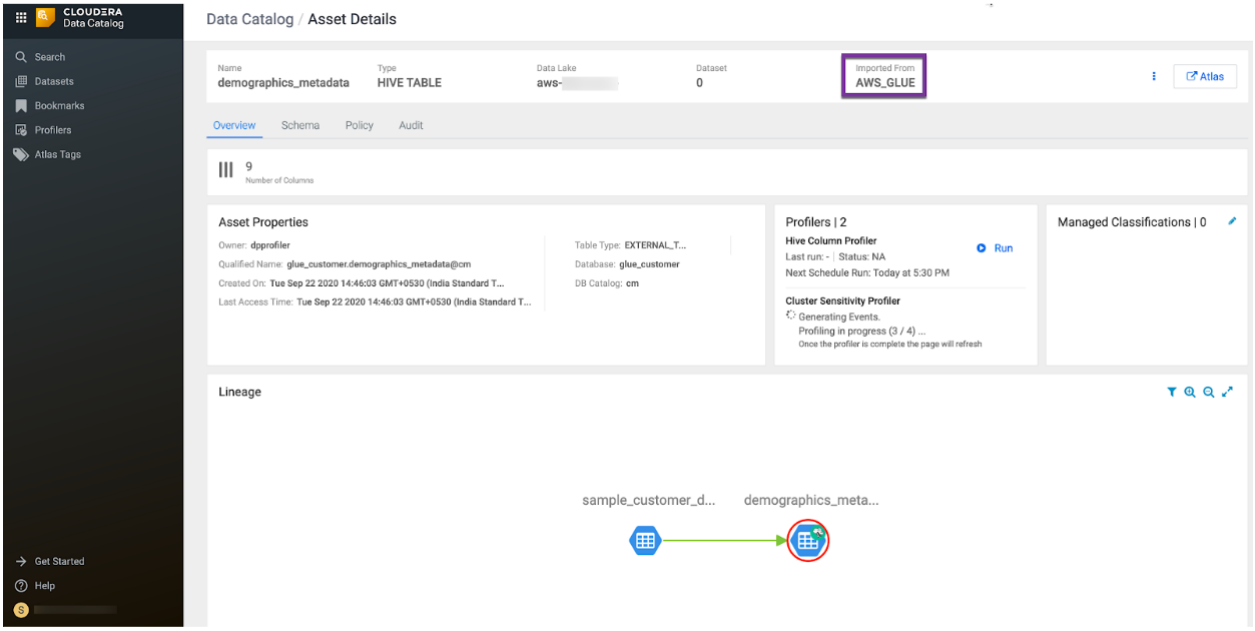
Next, navigate back to the Cloudera Data Catalog main page and select the Cloudera Data Catalog owned data lake and select the type as Hive Table. The search results lists all the Hive table assets and you can view the Glue registered asset(s) as well. The registered Glue asset can be identified using





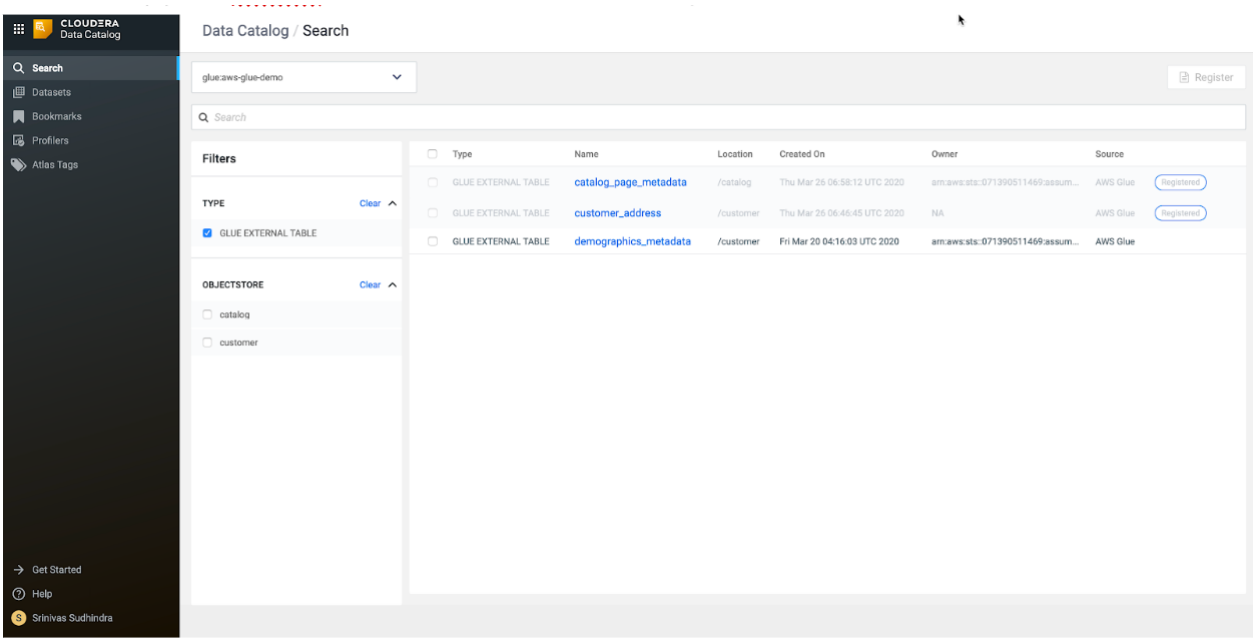
Note: The entries in Cloudera Data Catalog for Glue assets are created in the Hive Metastore.

Click on the Glue registered asset and you can view the **Asset Details** page for the selected Glue asset.



The **Asset Details** page for the Glue asset is populated by Atlas. While registering the Glue data, the data was written to the Hive Metastore and later Atlas synchronised the metadata.

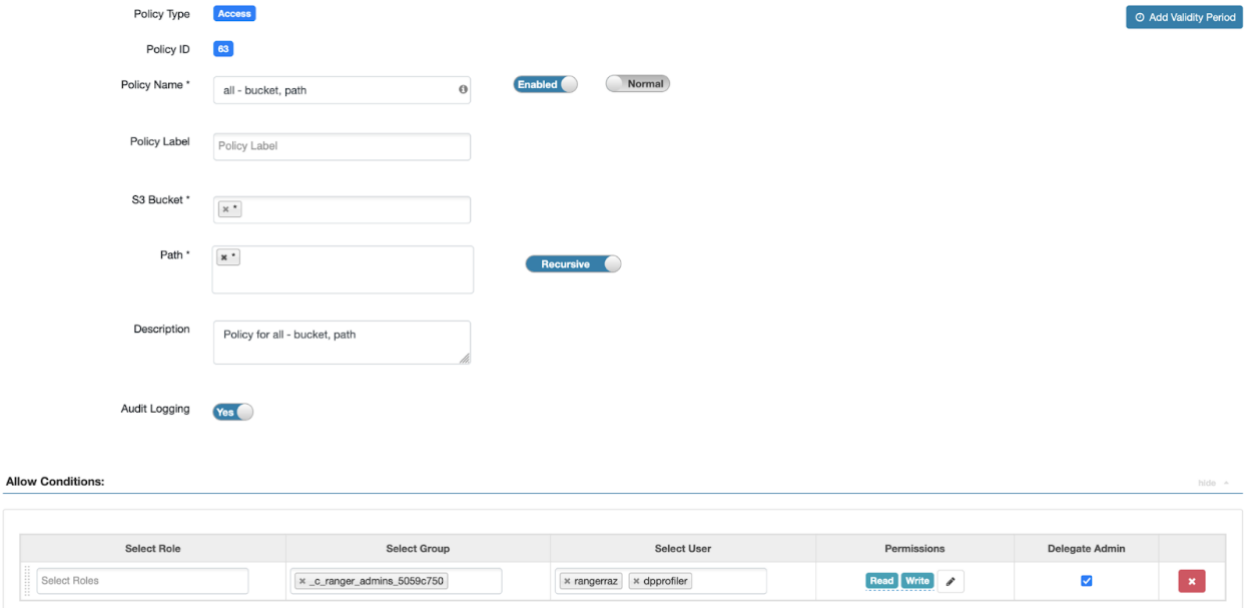
Go back to the main Cloudera Data Catalog page and select the Glue data lake. Note that the registered Glue asset(s) are greyed out or cannot be selected again.



You can still view the registered Glue assets (powered by Atlas) by clicking on the same and it navigates to the **Asset Details** page as seen above in the image.

Working with Ranger Authorization Service (RAZ) enabled AWS environment

For RAZ enabled AWS environment, you must employ the following permission settings to work with Cloudera Data Catalog - Glue integration.



Note: By default, the dpprofiler user is not included in the allowed users list. You must manually add the dpprofiler user in the allowed users list.