

## Search

Date published: 2019-11-14

Date modified: 2025-06-11

# CLOUDERA

# Legal Notice

© Cloudera Inc. 2025. 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

<b>Search for assets.....</b>	<b>4</b>
Filters.....	4
Accessing Data Lakes.....	6
Download search results as CSV files.....	6
Integrating Cloudera Data Catalog with AWS Glue Data Catalog.....	7
Setting up AWS Glue Catalog with Cloudera Data Catalog.....	8
Working with AWS Glue.....	12
Accessing AWS Glue.....	12
Prerequisites for accessing Hue tables and databases.....	16
Accessing Hue assets.....	16
Accessing Hue tables and databases from the Search page.....	18
Accessing Hue tables and databases from the Asset Details page.....	19
Searching for assets using Atlas glossaries.....	20
Using terms in Cloudera Data Catalog.....	21
Mapping glossary terms.....	22
Searching for assets using glossary terms.....	24
Additional search options for asset types.....	25
Accessing tables based on Ranger policies.....	27
Creating classifications for selected assets.....	28
 <b>Viewing Data Asset details.....</b>	 <b>30</b>
Supported Data Asset fields for entities.....	30
Navigation in Asset Details.....	32
View Data Asset Access Audit Logs.....	34
Navigation support for hive entities within Lineage.....	35
Adding assets to one or more datasets.....	36
Viewing Atlas entity audits.....	38
Viewing Ranger access audits.....	40
Viewing Ranger policies.....	42

## Search for assets

On the Cloudera Data Catalog **Search** page, select a data lake and enter a search string in the search box to view all the assets with details that contain the search string.

When you enter the search terms **Search**, you are looking up names, types, descriptions, and other metadata collected by Cloudera Data Catalog. The search index includes metadata (not data) about your environment and cluster data assets and operations. You can make the search more powerful by associating your own information (business metadata) to the stored assets.

**Note:**

For the selected data lake, click the Atlas and Ranger links to navigate to the respective base cluster services in a new browser tab.

**Related Information**

[Understanding datasets](#)

## Filters

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

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 further refine your search results using filters as follows:

**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.

- 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

---

<sup>1</sup> Iceberg assets are discoverable in VM-based environments but they can be profiled only in Compute Cluster enabled environments.

- Iceberg Table<sup>1</sup>
- Impala Column Lineage
- Impala Process
- Impala Process Execution
- Kafka topic
- ML Model Build
- ML Model Deployment
- ML Project
- RDBMS Column
- RDBMS DB
- 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. For example, selecting the Hive Table type will enable the Column Tag filter.

### 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.



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

### Time Range

You can filter your assets by the **Created On** date (if provided by Atlas) after selecting an 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 terms set up.

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

In the resulting list of your matching assets, you can click a row and see the following:

- **Qualified name**
- **Database**
- **Classification**
- **Terms**


Clicking the Name of the entity will open its **Asset Details**.

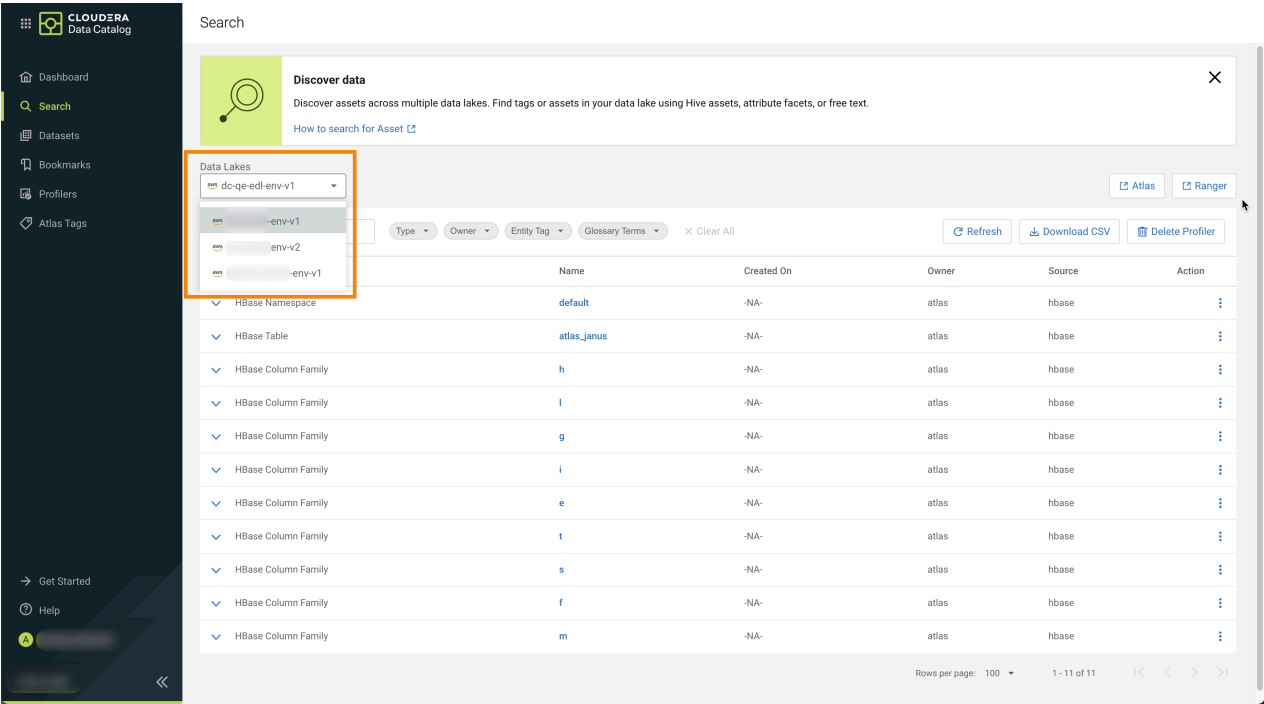
Accessing Data Lakes

In the **Search** page, the accessible data lakes are displayed in a drop-down.

Users have access to the lakes based on the permissions that are granted. You can choose the available lake by selecting the appropriate radio button.

For example, in the following diagram, the logged in user has access to all the listed data lakes.

 **Note:** You can search the assets of one data lake at a time.



Download search results as CSV files

You can download the search result for the current query with the selected data lake. The feature allows you to download up to 10000 rows for the current search query.

The CSV file format does not conform to any specific order or continuation in the downloaded results. For example, a user can download 10000 assets and later downloads the results for the same query again, then the downloaded CSV files may not contain the search results in the same order as it was downloaded previously.

Click **Download CSV** to start your download:

## 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 ExternalDataDiscoveryService (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:

- 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.

Role ARN: [arn:aws:iam::123456789012:role/Role for IDBroker](#)

Role description: Role for IDBroker | [Edit](#)

Instance Profile ARNs: [arn:aws:iam::123456789012:instance-profile/Instance Profile for IDBroker](#)

Path: /

Creation time: 2020-06-11 17:11 UTC+0530

Last activity: 2020-08-31 16:50 UTC+0530 (Today)

Maximum session duration: 1 hour [Edit](#)

**Permissions** | Trust relationships | Tags (1) | Access Advisor | Revoke sessions

▼ Permissions policies (2 policies applied)

[Attach policies](#) [Add inline policy](#)

Policy name	Policy type
aws-cdp-idbroker-assume-role-policy	Managed policy

[Policy summary](#) | [JSON](#) | [Edit policy](#) | [Simulate policy](#)

```

1 {
2   "Version": "2012-10-17",
3   "Statement": [
4     {
5       "Sid": "VisualEditor0",
6       "Effect": "Allow",
7       "Action": [
8         "sts:AssumeRole",
9         "sts:DecodeAuthorizationMessage",
10        "glue:*"
11      ],
12       "Resource": "*"
13     }
14   ]
15 }
```



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

1 2

A policy defines the AWS permissions that you can assign to a user, group, or role. You can create and edit a policy in the visual editor and using JSON. [Learn more](#)

**Visual editor** | **JSON** | [Import managed policy](#)

[Expand all](#) | [Collapse all](#)

▶ STS (2 actions)	<a href="#">Clone</a>   <a href="#">Remove</a>
▶ Glue (All actions)	<a href="#">Clone</a>   <a href="#">Remove</a>

[Add additional permissions](#)

1 2

A policy defines the AWS permissions that you can assign to a user, group, or role. You can create and edit a policy in the visual editor and using JSON. [Learn more](#)

Visual editor
JSON
Import managed policy

Expand all Collapse all

STS (2 actions)
Clone Remove

Glue (All actions)
Clone Remove

Select a service
Clone Remove

Service
Select a service below
Enter service manually

close

Glue ?

Actions
Choose a service before defining actions

Resources
Choose actions before applying resources

Request conditions
Choose actions before specifying conditions

Add additional permissions

Glue (All actions) 7 warnings
Clone Remove

Service
Glue

Actions
Specify the actions allowed in Glue ?
Switch to deny permissions ⓘ

close

Manual actions (add actions)

☒ All Glue actions (glue:\*)

Access level

☒ List (8 selected)

☒ Read (49 selected)

☒ Tagging (2 selected)

☒ Write (66 selected)

Resources
Specify catalog resource ARN for the BatchGetPartition and 39 more actions. ⓘ  
Specify connection resource ARN for the GetConnection and 5 more actions. ⓘ  
Specify database resource ARN for the BatchGetPartition and 28 more actions. ⓘ  
Specify mlTransform resource ARN for the UseMLTransforms and 10 more actions. ⓘ

Glue (All actions)
Clone Remove

Service
Glue

Actions
Manual actions

Resources
☐ Specific  
☒ All resources

Request conditions
Specify request conditions (optional)

Add additional permissions

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.

10

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::<xxxxxxxxxxxx>: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::<xxxxxxxxxxxx>:role/add-role \
--ranger-audit-role arn:aws:iam::<xxxxxxxxxxxx>: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
xxxx>: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. 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.

Data Catalog / Search

glue:aws-glue-demo

Search

Filters

TYPE

GLUE EXTERNAL TABLE

OBJECTSTORE

Type	Name	Location	Created On	Owner	Source
GLUE EXTERNAL TABLE	catalog_page_metadata	/catalog	Thu Mar 26 06:58:12 UTC 2020	am:aws sts:071390511469-assumed-...	AWS Glue
GLUE EXTERNAL TABLE	customer_address	/customer	Thu Mar 26 06:46:45 UTC 2020	NA	AWS Glue
GLUE EXTERNAL TABLE	demographics_metadata	/customer	Fri Mar 20 04:16:03 UTC 2020	am:aws sts:071390511469-assumed-...	AWS Glue

Register

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

Data Catalog / Asset Details

Name: catalog\_page\_metadata Type: GLUE EXTERNAL TABLE Data Lake: aws-glue-demo

Overview Schema

Asset Properties

Owner: am:aws sts:071390511469-assumed-role/AssumeAdmin/amit.kumar

Qualified Name: catalog.catalog\_page\_metadata

Created On: Thu Mar 26 2020 12:28:12 GMT+0530 (India Standard Time)

Table Type: EXTERNAL\_TABLE

Database: catalog

Retention: 0

Underlying File Type: csv

Register

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.

Data Catalog / Search

aws

glue

Search

Filters

TYPE

GLUE EXTERNAL TABLE

OBJECTSTORE

Type	Name	Location	Created On	Owner	Source
GLUE EXTERNAL TABLE	catalog_page_metadata	/catalog	Thu Mar 26 06:58:12 UTC 2020	am:aws sts:071390511469-assumed-...	AWS Glue
GLUE EXTERNAL TABLE	customer_address	/customer	Thu Mar 26 06:46:45 UTC 2020	NA	AWS Glue
GLUE EXTERNAL TABLE	demographics_metadata	/customer	Fri Mar 20 04:16:03 UTC 2020	am:aws sts:071390511469-assumed-...	AWS Glue

Register

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

The screenshot shows the 'Data Catalog / Asset Details' page for the asset 'customer\_address'. The asset is a 'GLUE EXTERNAL TABLE' located in the 'aws-glue-demo' database. A green box highlights the text 'customer\_address is registered to CDP'. A yellow box highlights the 'Registered' status. The 'Asset Properties' section shows: Owner: NA, Qualified Name: customer.customer\_address, Created On: Thu Mar 26 2020 12:16:45 GMT+0530 (India Standard Time), Table Type: EXTERNAL\_TABLE, Database: customer, Retention: 0, and Underlying File Type: csv.

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

The screenshot shows the 'Data Catalog / Search' page. A blue box highlights the 'Imported' status icon. The search results table lists various Hive Table assets. The 'customer\_address' asset is highlighted with a red box, showing its 'Imported' status. The table columns are: Type, Name, Location, Created On, Owner, and Source.

Type	Name	Location	Created On	Owner	Source
Hive Table	catalog_page_metadata	/glue_catalog	Tue Sep 22 2020	dprouffier	hive
Hive Table	compactions	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	scheduled_executions	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	scheduled_queries	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	cdh_version	/sys	Tue Sep 22 2020	hive	hive
Hive Table	partition_stats_view	/sys	Tue Sep 22 2020	hive	hive
Hive Table	table_stats_view	/sys	Tue Sep 22 2020	hive	hive
Hive Table	version	/sys	Tue Sep 22 2020	hive	hive
Hive Table	columns	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	locks	/sys	Tue Sep 22 2020	hive	hive
Hive Table	compactions	/sys	Tue Sep 22 2020	hive	hive
Hive Table	transactions	/sys	Tue Sep 22 2020	hive	hive
Hive Table	tables	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	views	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	column_privileges	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	table_privileges	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	schemata	/information_schema	Tue Sep 22 2020	hive	hive
Hive Table	customer_address	/glue_customer	Tue Sep 22 2020	dprouffier	hive
Hive Table	demographics_metadata	/glue_customer	Tue Sep 22 2020	dprouffier	hive



**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 screenshot shows the 'Data Catalog / Asset Details' page for a Glue asset named 'demographics\_metadata'. The asset is a HIVE TABLE located in the 'aws-' Data Lake, with Dataset ID 0. It was imported from AWS\_GLUE. The page includes tabs for Overview, Schema, Policy, and Audit. The Overview tab is active, showing 9 columns. The Asset Properties section lists the owner as 'dpprofiler', qualified name as 'glue\_customer.demographics\_metadata/cm', and creation time as 'Tue Sep 22 2020 14:46:03 GMT+0530 (India Standard T...)'. The Table Type is EXTERNAL\_T... and the Database is glue\_customer. The DB Catalog is cm. The Profilers section shows 2 profilers: Hive Column Profiler (Last run: - Status: NA, Next Schedule Run: Today at 5:30 PM) and Cluster Sensitivity Profiler (Generating Events, Profiling in progress (3 / 4) ...). The Managed Classifications section shows 0 classifications. The Lineage section shows a diagram with 'sample\_customer\_d...' and 'demographics\_meta...' connected by a green arrow.

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.

The screenshot shows the 'Data Catalog / Search' page. The search bar contains 'glueaws-glue-demo'. The search results are displayed in a table with columns: Type, Name, Location, Created On, Owner, and Source. The table lists three assets: 'catalog\_page\_metadata', 'customer\_address', and 'demographics\_metadata'. All three assets are of type 'GLUE EXTERNAL TABLE' and are marked as 'registered'. The 'demographics\_metadata' asset is highlighted in blue.

Type	Name	Location	Created On	Owner	Source
GLUE EXTERNAL TABLE	catalog_page_metadata	/catalog	Thu Mar 26 06:58:12 UTC 2020	arn:aws:sts:071390511469:assum...	AWS Glue (registered)
GLUE EXTERNAL TABLE	customer_address	/customer	Thu Mar 26 06:46:45 UTC 2020	NA	AWS Glue (registered)
GLUE EXTERNAL TABLE	demographics_metadata	/customer	Fri Mar 20 04:16:03 UTC 2020	arn:aws:sts:071390511469:assum...	AWS Glue

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.

Policy Type **Access** ⓘ Add Validity Period

Policy ID **63**

Policy Name \*  **Enabled** **Normal**

Policy Label

S3 Bucket \*

Path \*  **Recursive**

Description

Audit Logging **Yes**

**Allow Conditions:** hide

Select Role	Select Group	Select User	Permissions	Delegate Admin
<input type="text" value="Select Roles"/>	<input type="text" value="x_c_ranger_admins_5059c750"/>	<input type="text" value="x_rangerraz"/> <input type="text" value="x_dpprofiler"/>	<b>Read</b> <b>Write</b> <input type="text" value=""/>	<input checked="" type="checkbox"/> <input type="text" value=""/>



**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.

## Prerequisites for accessing Hue tables and databases

Apache Hue is a part of the Cloudera Data Warehouse service. Before you can access Apache 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, Apache Hue links for the selected assets are displayed in Cloudera Data Catalog.

## Accessing Hue assets

Cloudera Data Catalog helps you to explore assets within your Data Lake. Apache 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 Hue instance from the Search page or the Asset Details.



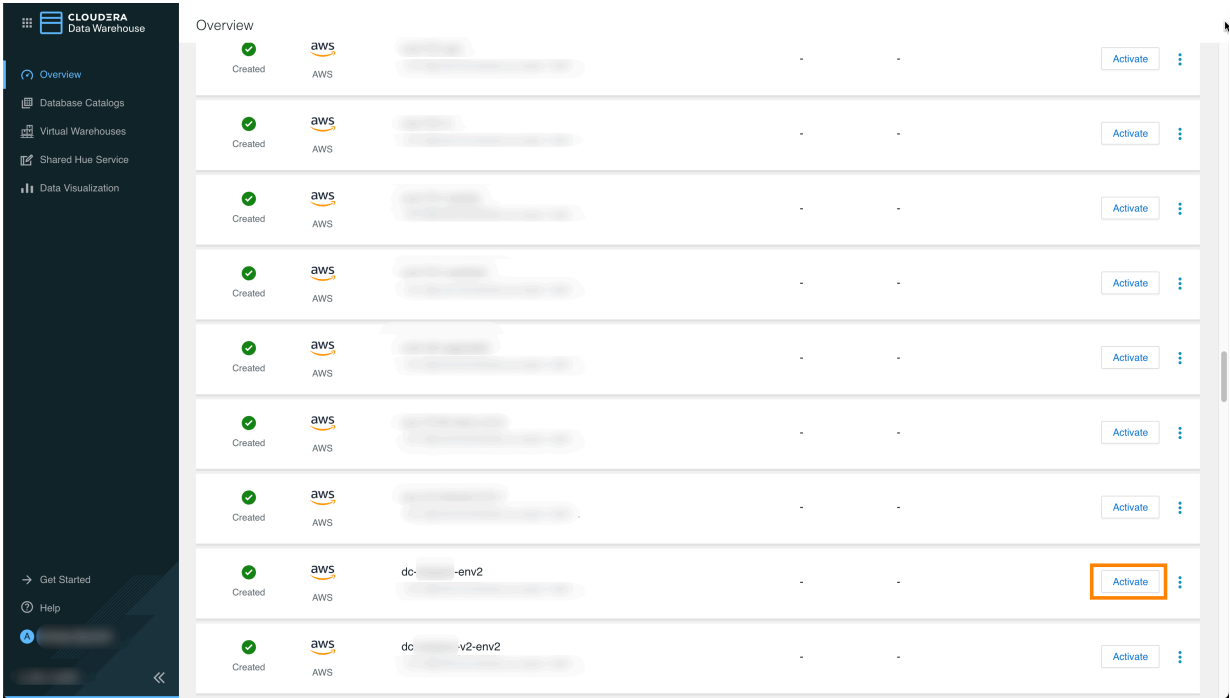
Before you begin

You must fulfill all conditions listed in [Prerequisites for accessing Hue tables and databases](#) on page 16 for viewing the external link button.

You must activate Cloudera Data Warehouse on your selected environment.

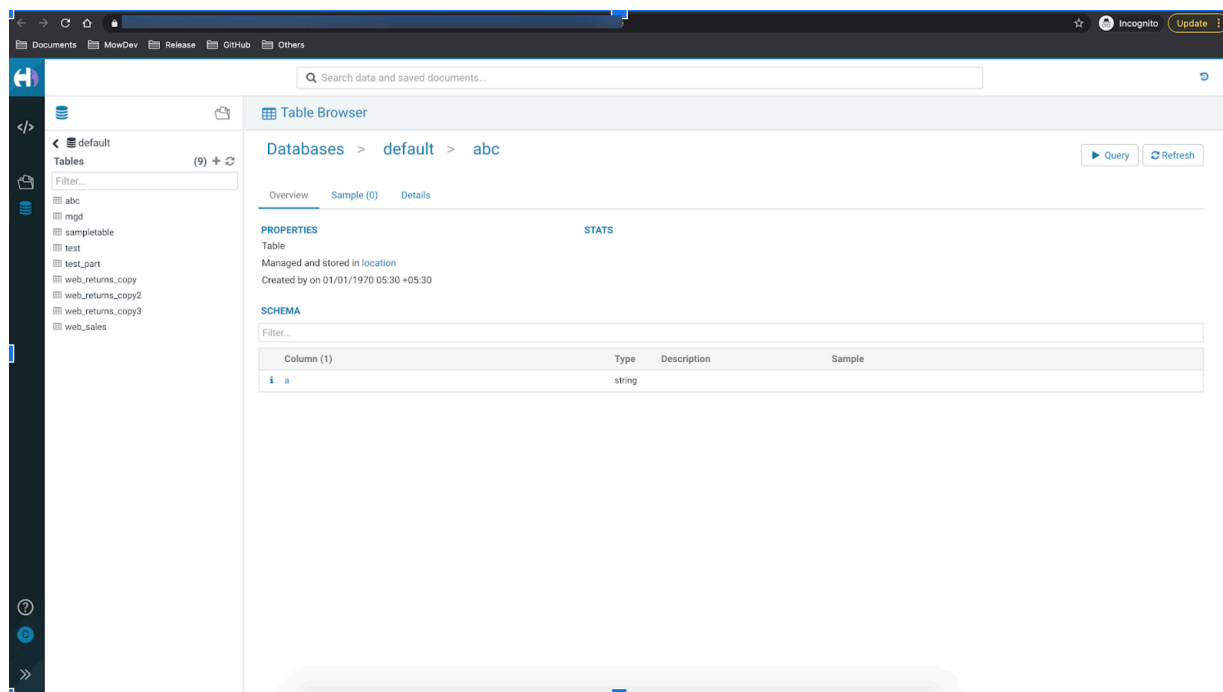
Procedure

- 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.



- 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 Hue assets on these virtual warehouses using the links provided by Cloudera Data Catalog.

## Accessing Hue tables and databases from the Search page

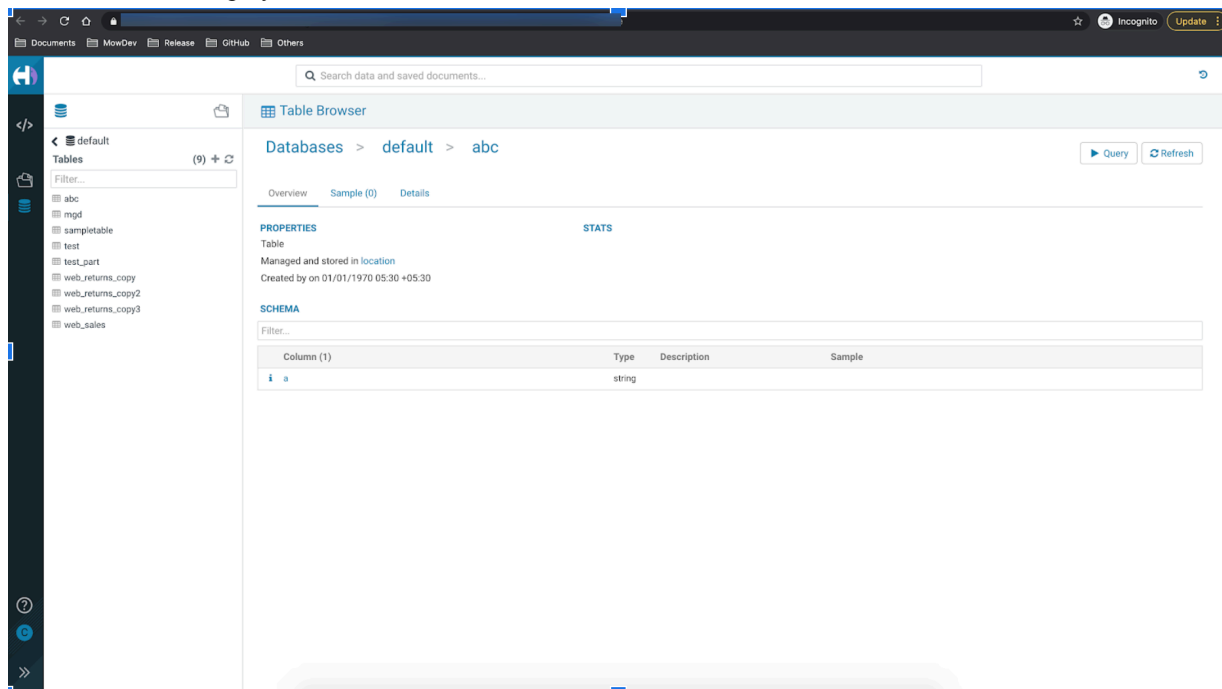
When you log into Cloudera Data Catalog, you can access web links to the Apache Hue instances from the Search page.

### Procedure

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

A Link to Experiences pop-up appears. It contains a list of Hue instances within the virtual warehouses that are associated with this particular table.

- Click the link to open Hue web interface.  
The Hive table is displayed on the Hue Table Browser.



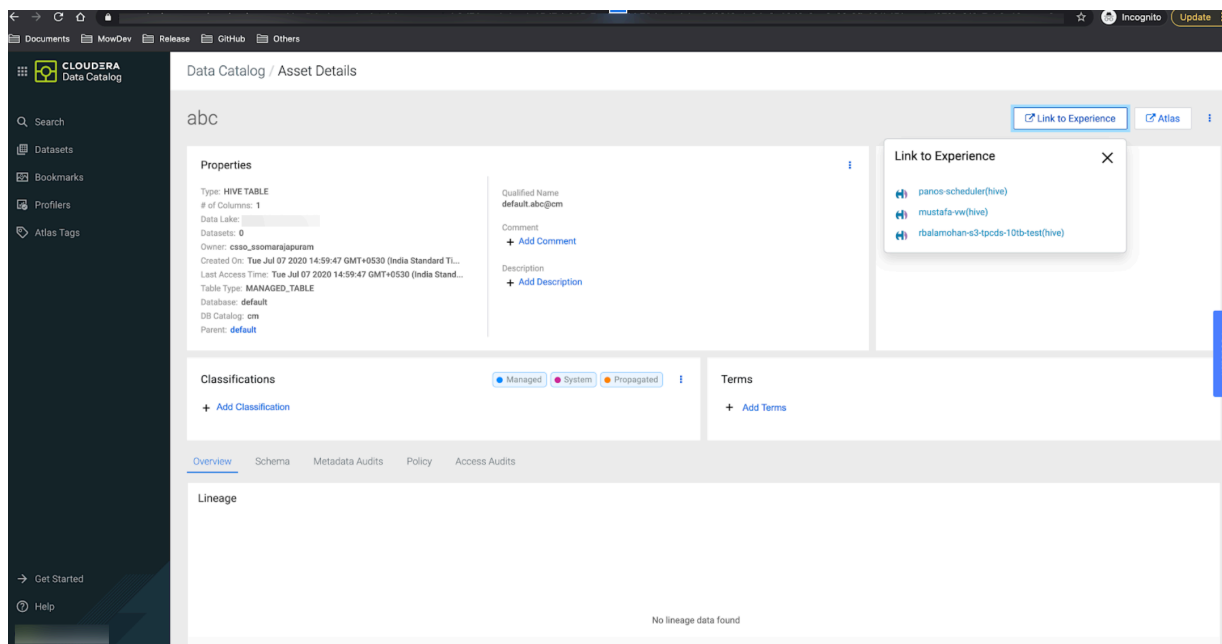
## Accessing Hue tables and databases from the Asset Details page

When you log into Cloudera Data Catalog, you can access web links to the Apache Hue instances from the Asset Details page.

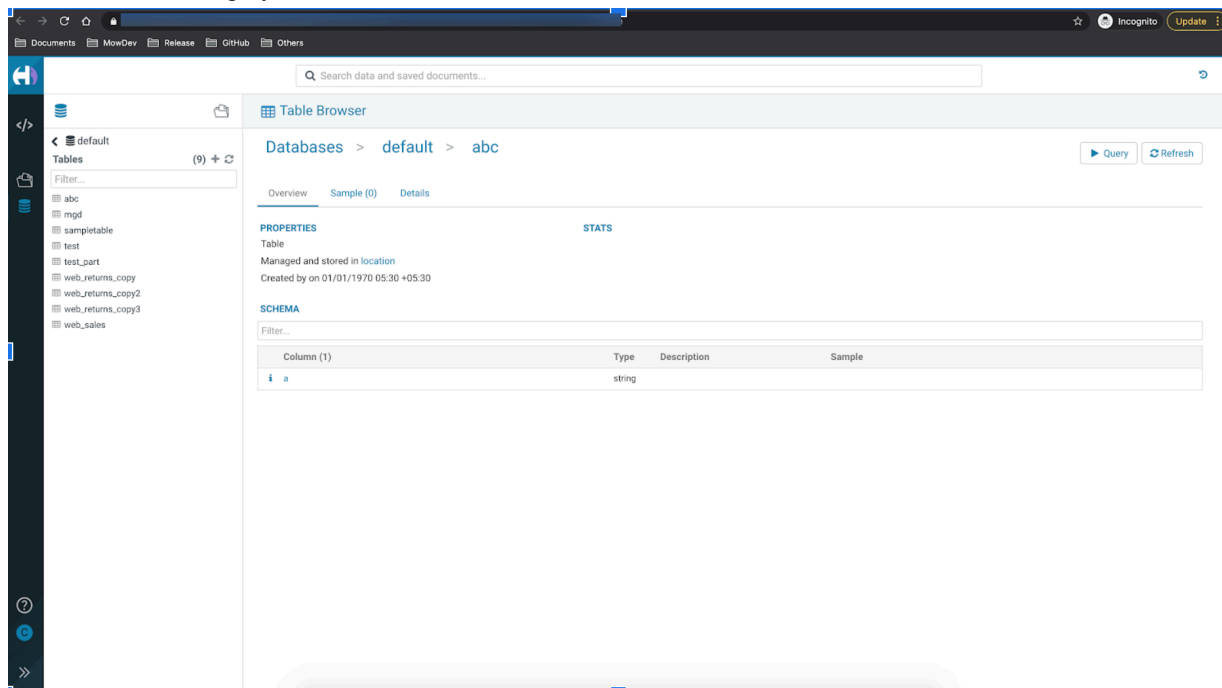
### Procedure

- Log into Cloudera Data Catalog.
- Select the Data Lake associated with your environment.
- Select the asset type under Filters, for example Hive Table.

- 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:



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



## 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.

The screenshot shows the Cloudera Data Catalog Search page. On the left is a sidebar with navigation links: Dashboard, Search, Datasets, Bookmarks, Profilers, and Atlas Tags. The main content area is titled 'Search' and includes a search bar and a 'Glossary Terms' filter. The filter is highlighted with an orange box and shows a dropdown menu with 'Search Glossary Terms' and 'First Atlas Glossary term@Test Glossary' options. Below the filter is a table of assets with columns: Type, Name, Owner, Source, and Action. The table lists various HBase and HDFS assets, including namespaces, tables, and column families.

Type	Name	Owner	Source	Action
▼ HBase Namespace	default	atlas	hbase	⋮
▼ HBase Table	atlas_janus	atlas	hbase	⋮
▼ HBase Column Family	m	atlas	hbase	⋮
▼ HBase Column Family	l	atlas	hbase	⋮
▼ HBase Column Family	g	atlas	hbase	⋮
▼ HBase Column Family	i	atlas	hbase	⋮
▼ HBase Column Family	h	atlas	hbase	⋮
▼ HBase Column Family	f	atlas	hbase	⋮
▼ HBase Column Family	s	atlas	hbase	⋮
▼ HBase Column Family	t	atlas	hbase	⋮
▼ HBase Column Family	e	atlas	hbase	⋮
▼ HDFS Path	Test-hdfs-from-Atlas-by-	-NA-	hdfs	⋮

## 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.

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.



**Note:** When you work with terms in Cloudera Data Catalog and map them to your assets, you can search for the same datasets in Atlas by using the corresponding terms.

The screenshot shows the 'Asset Details' page for an asset named 'default'. The left sidebar contains navigation links: Dashboard, Search, Datasets, Bookmarks, Profilers, and Atlas Tags. The main content area is divided into several sections:

- Properties:** Type: HBASE NAMESPACE, Data Lake: dc, Owner: atlas, Created On: -NA-, Update Time: -NA-, Created By: atlas, Updated By: cseo, Status: ACTIVE. Qualified Name: default@cm. Description: default.
- Classifications | 2:** Managed, System, Propagated. Two tags are visible: 'test\_tag\_created\_in\_dc\_atlas...' and 'Test\_tag\_created\_in...'.
- Terms:** A button labeled '+ Add Terms' is highlighted with an orange box.
- Content:** A table with columns: Type, Name, Location, Created On, Owner, Source. The table contains one row: hbase\_table, atlas\_janus, /default, -NA-, atlas, hbase.

## Mapping glossary terms

Cloudera Data Catalog contains the glossary terms that are created in Apache Atlas.

You can search for those terms in Cloudera Data Catalog and map specific terms with assets. 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.

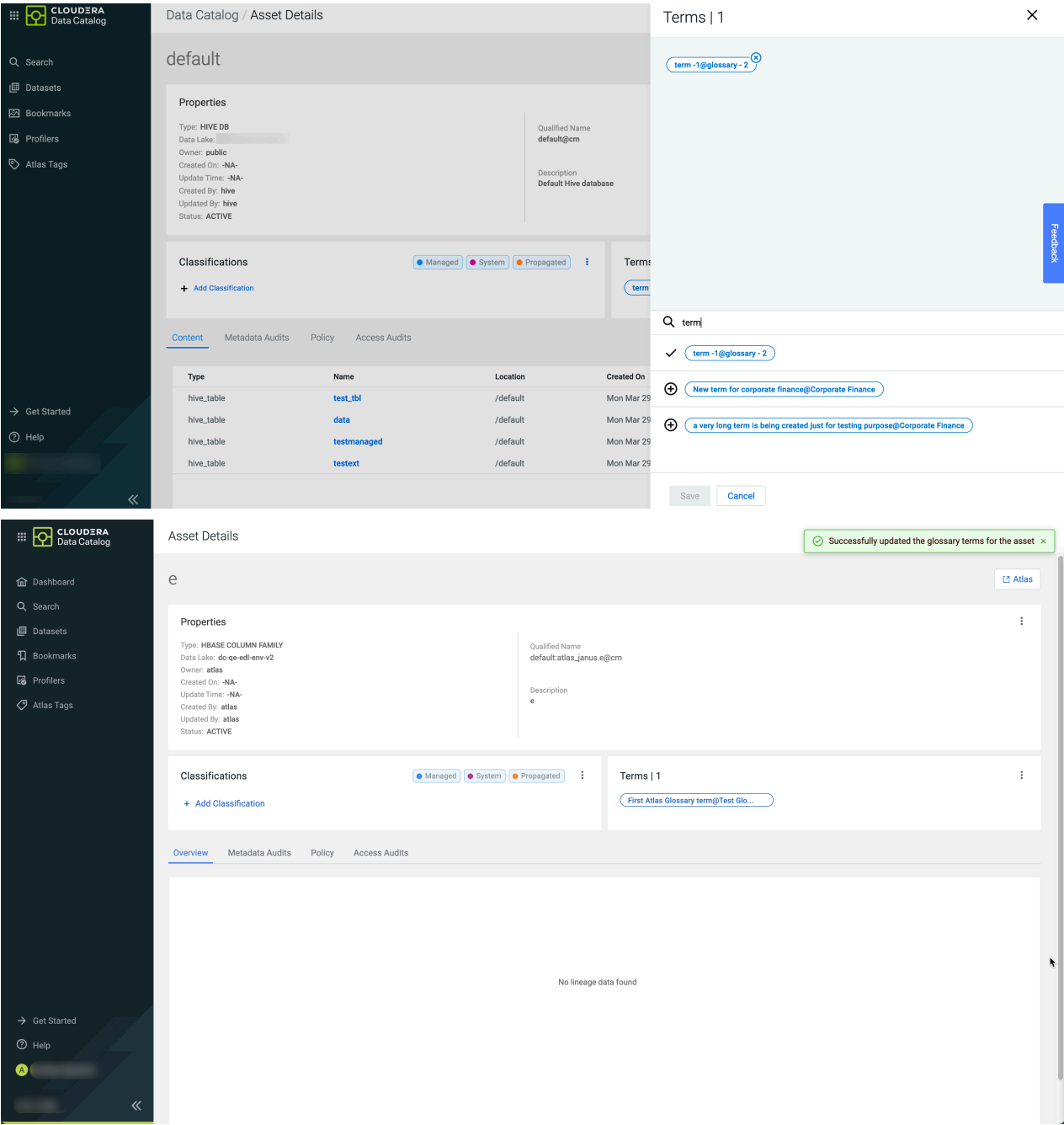
When you map a specific term for your dataset, the term is displayed in the following format:

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

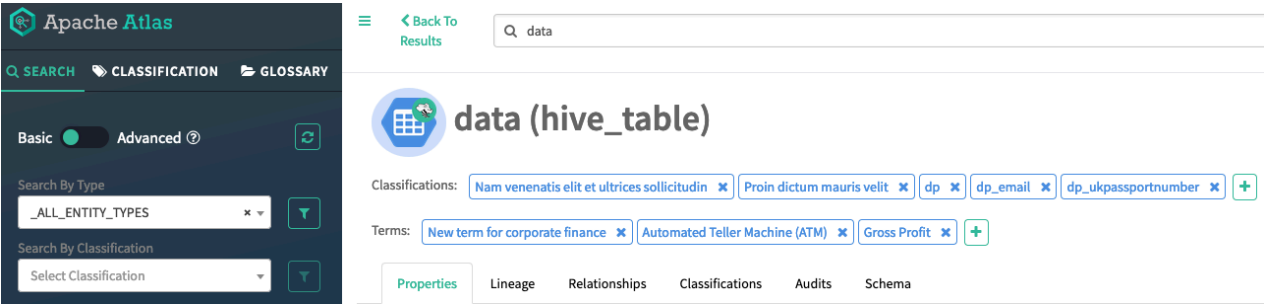
The screenshot shows the 'Asset Details' page for an asset named 'data'. The left sidebar is the same as the previous screenshot. The main content area is divided into several sections:

- Properties:** Type: HIVE TABLE, # of Columns: 12, Data Lake: dc, Datasets: 0, Owner: dpprofiler, Created On: Mon Mar 29 2021 14:48:32 GMT+0530 (India Standard Time), Last Access Time: Mon Mar 29 2021 14:48:32 GMT+0530 (India Standard Time), Table Type: EXTERNAL\_TABLE, Database: default, DB Catalog: cm, Parent: default. Qualified Name: default.data@cm. Comment: + Add Comment. Description: + Add Description.
- Classifications | 5:** Managed, System, Propagated. Three tags are visible: 'dp', 'Nam venenatis elit et ultrices sollicitudin', and 'Proin dictum mauris velit'.
- Terms | 3:** A box highlights three terms: 'New term for corporate finance', 'Automated Teller Machine (ATM)', and 'Gross Profit'.
- Profilers | 2:** Cluster Sensitivity Profiler (Last run: a day ago, Status: SUCCESS, Next Schedule Run: Tomorrow at 3:50 PM) and Hive Column Profiler (Last run: a day ago, Status: SUCCESS, Next Schedule Run: Today at 11:30 PM).
- Lineage:** A diagram shows the data flow from '/user/dpprofiler/...' to 'data' (highlighted with a red circle) to 'test\_tbl'.

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.



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.

Overview **Schema** Metadata Audits Policy Access Audits

Search Column Edit

	Chart Type	Column Name	Type	Unique Values *	Null Values	Max	Min	Mean	Comment	Classifications	Terms
▼		age	int	21	18	49	1	23.66		Nam venenatis elit et. + 1	
▼		cabin	string	9	0					Nam venenatis elit et. + 1	Accounting Rate of ... + 1
▼		embarked	string	3	0					dp.ukpassportnumbe + 2	Compound Annual G... + 1
▼		fare	float	35	0	262.38		23.78		dp.ukpassportnumbe + 1	New term for corpor... + 5
▼		name	string	54	0					dp.ukpassportnumbe + 6	New term for corpor... + 5
▼		parch	int	3	0	2		0.42		dp.ukpassportnumbe + 1	New term for corpor... + 6
▼		passengerid	int	50	0	53	1	27		dp.ukpassportnumbe + 2	New term for corpor... + 2
▼		pclass	int	3	0	3	1	2.42		dp.ukpassportnumber	New term for corpor... + 5
▼		sex	string	2	0					dp.ukpassportnumber	a very long term is b... + 6
▼		sibsp	int	4	0	8		0.43			
▼		survived	int	2	0	1		0.72			
▼		ticket	string	48	0						

Rows per page: 20 1 - 12 of 12

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

Cloudera Data Catalog

Search

Datasets

Bookmarks

Profilers

Atlas Tags

Get Started

Help

Data Catalog / Asset Details

Overview **Schema** Metadata Audits Policy Access Audits

Search Column Edit

	Chart Type	Column Name	Type	Unique Values *	Null Values	Max	Min	Mean	Comment	Classifications	Terms
▼		age	int	21	18	49	1	23.66		Nam venenatis elit et. + 1	a very long term is b... + 7
▼		cabin	string	9	0					Nam venenatis elit et. + 1	a very long term is b... + 7
▼		embarked	string	3	0					dp.ukpassportnumbe + 2	a very long term is b... + 7
▼		fare	float	35	0	262.38		23.78		dp.ukpassportnumbe + 1	a very long term is b... + 7
▼		name	string	54	0					dp.ukpassportnumbe + 6	a very long term is b... + 7
▼		parch	int	3	0	2		0.42		dp.ukpassportnumbe + 1	a very long term is b... + 7
▼		passengerid	int	50	0	53	1	27		dp.ukpassportnumbe + 2	a very long term is b... + 7
▼		pclass	int	3	0	3	1	2.42		dp.ukpassportnumber	a very long term is b... + 7
▼		sex	string	2	0					dp.ukpassportnumber	a very long term is b... + 7
▼		sibsp	int	4	0	8		0.43			a very long term is b... + 7
▼		survived	int	2	0	1		0.72			a very long term is b... + 7
▼		ticket	string	48	0						a very long term is b... + 7

Rows per page: 20 1 - 12 of 12

## 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.



## 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.

Data Catalog / Asset Details

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

Data Catalog / Asset Details

ww\_customers

Properties

Type: HIVE TABLE

# of Columns: 40

Data Lake:

Datasets: 0

Owner: hive

Created On: Tue Mar 09 2021 10:48:45 GMT+0530 (India Stand...

Last Access Time: Tue Mar 09 2021 10:48:45 GMT+0530 (Indi...

Table Type: EXTERNAL\_TABLE

Database: hortoniabank

DB Catalog:

Parent: hortoniabank

Qualified Name

hortoniabank.ww\_customers@cm

Comment

passport\_number

Description

visa\_number

Cancel

Save

Profilers | 2

Cluster Sensitivity Profiler

Last run: 9 hours ago | Status: SUCCESS

Next Schedule Run: Thursday at 11:50 AM

Run

Hive Column Profiler

Last run: 8 hours ago | Status: SUCCESS

Next Schedule Run: Tomorrow at 5:30 PM

Run

Click Save to save your changes.

Data Catalog / Asset Details

Asset details were updated successfully.

ww\_customers

Properties

Type: HIVE TABLE

# of Columns: 40

Data Lake:

Datasets: 0

Owner: hive

Created On: Tue Mar 09 2021 10:48:45 GMT+0530 (India Stand...

Last Access Time: Tue Mar 09 2021 10:48:45 GMT+0530 (Indi...

Table Type: EXTERNAL\_TABLE

Database: hortoniabank

DB Catalog:

Parent: hortoniabank

Qualified Name

hortoniabank.ww\_customers@cm

Comment

passport\_number

Description

visa\_number

Profilers | 2

Cluster Sensitivity Profiler

Last run: 9 hours ago | Status: SUCCESS

Next Schedule Run: Thursday at 11:50 AM



Run

Hive Column Profiler

Last run: 8 hours ago | Status: SUCCESS

Next Schedule Run: Tomorrow at 5:30 PM

Run

 **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:

ww\_customers (hive\_table)

Classifications: +

Terms: +

Properties

Lineage

Relationships

Classifications

Audits

Schema

Technical properties

columns (40)

comment passport\_number

createTime 03/09/2021 10:48:45 AM (IST)

db hortoniabank

dcProfiledData { samplePercent: "100.0", rowCount: 50000, }

description visa\_number

User-defined properties

Labels

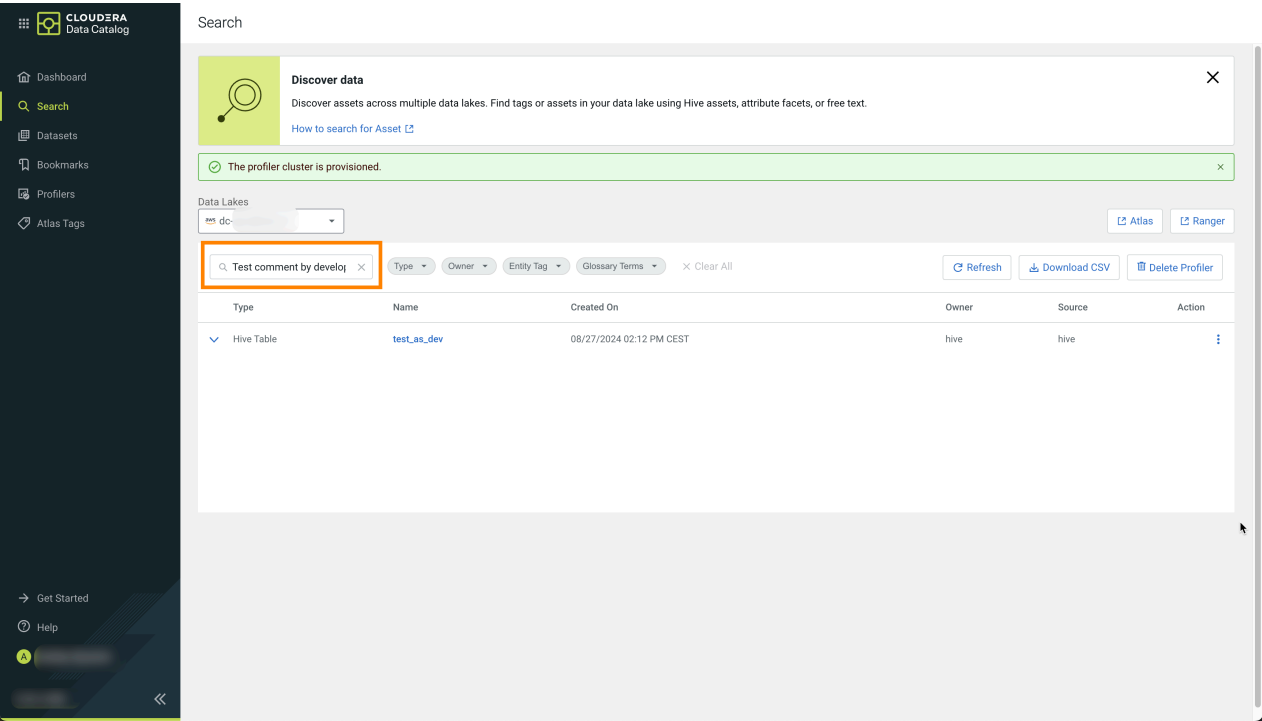
Business Metadata

Switch to Beta UI

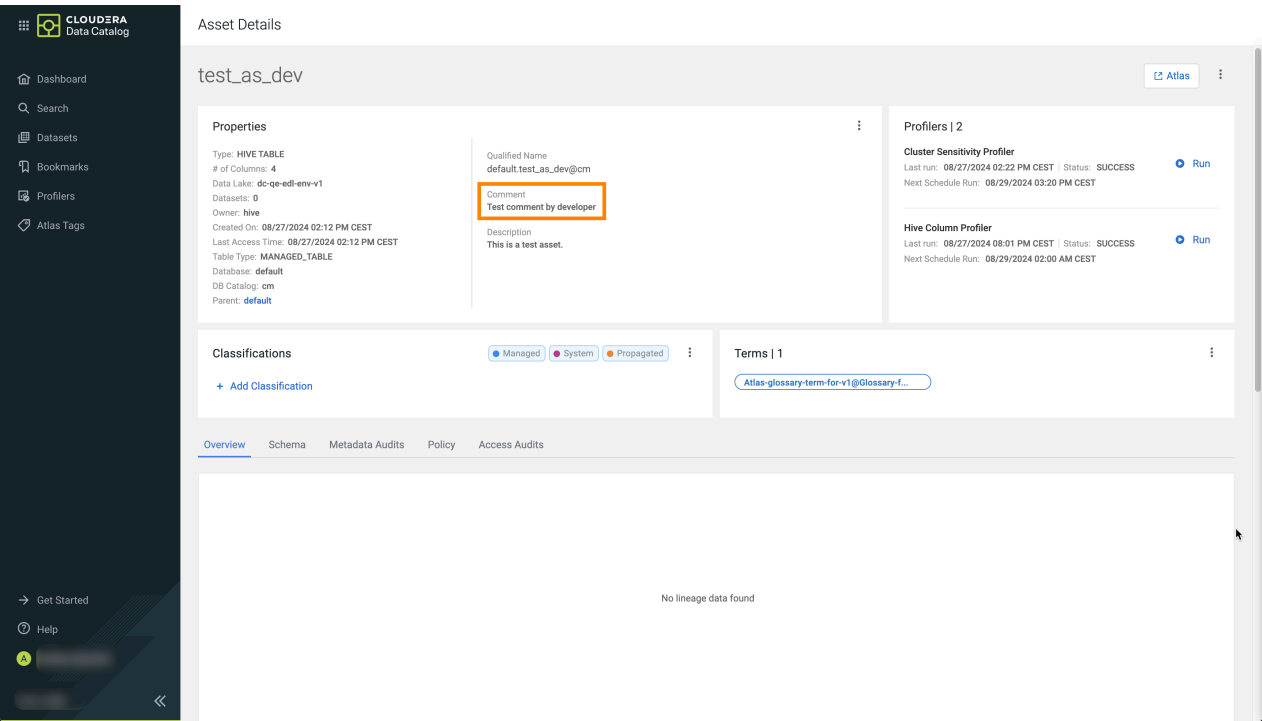
26

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.

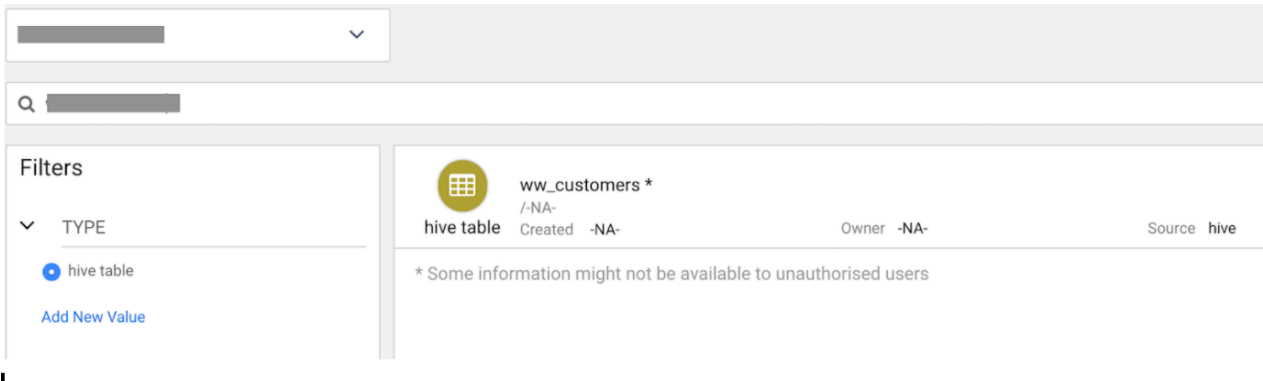


Accessing tables based on Ranger policies

When a table (in blue color link) is clicked, 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.

As seen in the following diagram, if users are not able to view the table details, a message appears next to the same table "Some information might not be available to unauthorised users".



In the next example diagram, tables that have the permissions to view are displayed with a blue color link. The ones that do not have read permissions are visible in grey.


<div>▼</div> <div>CREATED BEFORE</div> <div>Clear</div> <div><div><div><div></div><div>Last 1 day</div></div><div><div></div><div>Last 7 days</div></div><div><div></div><div>Last 15 days</div></div></div><div>Add New Value</div></div>	<div><div><div><div></div><div>Hive Table</div></div><div><div>/sys</div><div>Created</div><div>Tue Apr 07 2020</div><div>Owner</div><div>hive</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>scheduled_queries</div></div><div><div>/information_schema</div><div>Created</div><div>Tue Apr 07 2020</div><div>Owner</div><div>hive</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>schemata</div></div><div><div>/information_schema</div><div>Created</div><div>Tue Apr 07 2020</div><div>Owner</div><div>hive</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>table_stats_view</div></div><div><div>/sys</div><div>Created</div><div>Tue Apr 07 2020</div><div>Owner</div><div>hive</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>scheduled_executions</div></div><div><div>/information_schema</div><div>Created</div><div>Tue Apr 07 2020</div><div>Owner</div><div>hive</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>andromeda</div></div><div><div>/-</div><div>Created</div><div>-</div><div>Owner</div><div>-</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>milky</div></div><div><div>/-</div><div>Created</div><div>-</div><div>Owner</div><div>-</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>bear</div></div><div><div>/-</div><div>Created</div><div>-</div><div>Owner</div><div>-</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>n170</div></div><div><div>/-</div><div>Created</div><div>-</div><div>Owner</div><div>-</div><div>Source</div><div>hive</div></div></div></div> <div><div><div><div></div><div>umajor5</div></div><div><div>/-</div><div>Created</div><div>-</div><div>Owner</div><div>-</div><div>Source</div><div>hive</div></div></div></div>
--	---

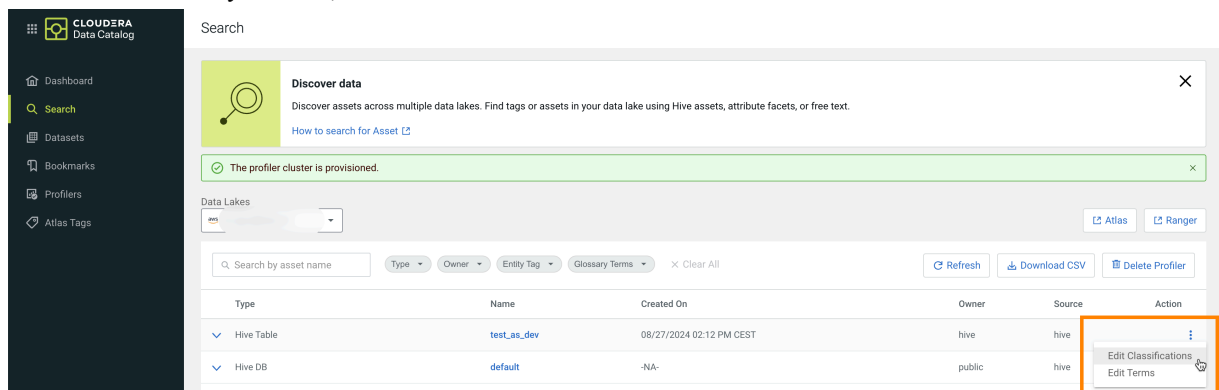
Creating classifications for selected assets

You can create classifications in multiple pages. These classifications can be associated with an asset. Then, you can use these classifications to filter your assets both in Cloudera Data Catalog and Apache Atlas.

Creating a classification from the Search page


- 1. Navigate to the Search page.

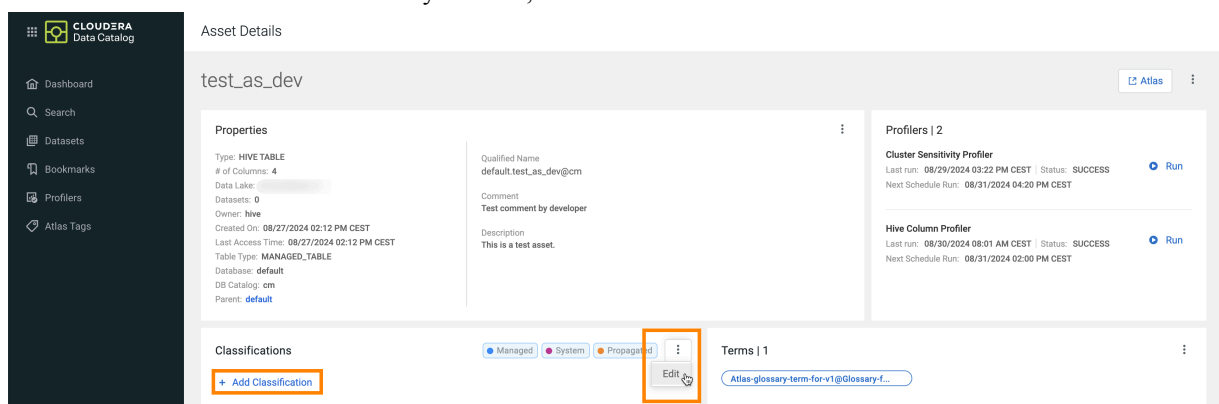
- Click the  icon by an asset, then select Edit Classifications.



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

### Creating a classification from Asset Details

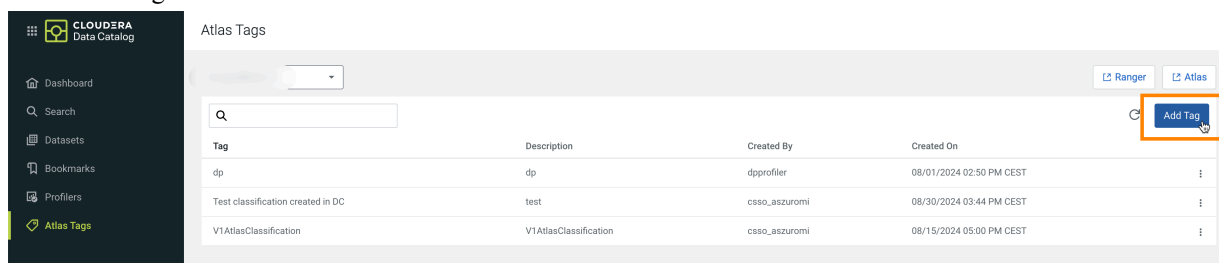
- Navigate to the **Asset Details** page of an asset.
- Click Add Classification or  icon by an asset, then select Edit.



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

### Creating a classification in Atlas Tags

- Navigate to **Atlas Tags**.
- Click Add Tag.



- Fill in the details and Save your changes.



**Note:** Your classification still needs to be added to an asset in the **Search** or **Asset Details** menu.



**Note:** Classifications are synchronized between Apache Atlas and Cloudera Data Catalog.

## 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.
- **Classification** 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 <sup>2</sup>	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
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

<sup>2</sup> 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
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 **Contents** 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.



Data Catalog / Asset Details

default

Atlas

Properties

Type: HBASE NAMESPACE  
Data Lake:   
Owner: atlas  
Created On: -NA-  
Update Time: -NA-  
Created By: atlas  
Updated By:   
Status: ACTIVE

Qualified Name  
default@cm

Description  
default

Classifications | 1

Managed

System

Propagated

rachit\_new\_inherit

Terms | 1

new\_term1@new\_glossary

Content

Metadata Audits

Policy

Access Audits

Type	Name	Location	Created On	Owner	Source
hbase_table	ATLAS_ENTITY_AUDIT_EVENTS	/default	Wed Apr 07 2021	atlas	hbase

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.

Asset Details

persons

Atlas

Properties

Type: HIVE TABLE  
Data Lake: dc-pro-h8f2uu  
Datasets: 3  
Owner: hive  
Created On: Mon Jun 17 2024 10:00:53 GMT+0200 (Cent...  
Last Access Time: Mon Jun 17 2024 10:00:53 GMT+020...  
Table Type: EXTERNAL\_TABLE  
Database: personal\_data  
DB Catalog: cm  
Parent: 

personal\_data  
Asset Type: HIVE TABLEHIVE DB

Qualified Name  
personal\_data.persons@cm

Comment  
+ Add Comment

Description  
+ Add Description

Classifications

Managed

System

Propagated

+ Add Classification

Profilers | 2

Cluster Sensitivity Profiler  
Last run: Last Monday at 12:09 PM | Status: SUCCESS  
Next Schedule Run: NA, Asset is not part of Allowed List  
+ Run

Hive Column Profiler  
Last run: | Status: NA  
Next Schedule Run: NA, Profiler is Disabled.  
+ Run

Terms

+ Add Terms

Overview

Schema

Metadata Audits

Policy

Access Audits

Q Search Column

Edit

Chart Type	Name	Type	Unique Values *	Null Values	Max	Min	Mean	Comment	Classifications	Terms
	name	string								
	ssn	string							dp_ssn	
	dob	string								
	email	string								

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

Type	Parent	Content
Hive DB	-	Hive Table, Iceberg Table

33

Type	Parent	Content
Hive Table, Iceberg Table <sup>3</sup>	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

## View Data Asset Access Audit Logs

The Asset Details Access Audit page displays all the Apache Ranger audit events associated with a particular data asset. This helps you to view who has accessed what data from a forensic audit or compliance perspective, and to visualize access patterns and identify anomalies.

### Procedure

1. From the **Search** menu, select a data asset.  
The **Asset Details** opens.
2. Click the **Access Audit** tab.  
The Audit table shows the most recent raw audit event data by type of access and access outcome (authorized/unauthorized).
3. You can filter the audit results by **Access Type** or **Result**.  
The available options are the following for Hive Table:



**Note:** Different asset types have different filter options.

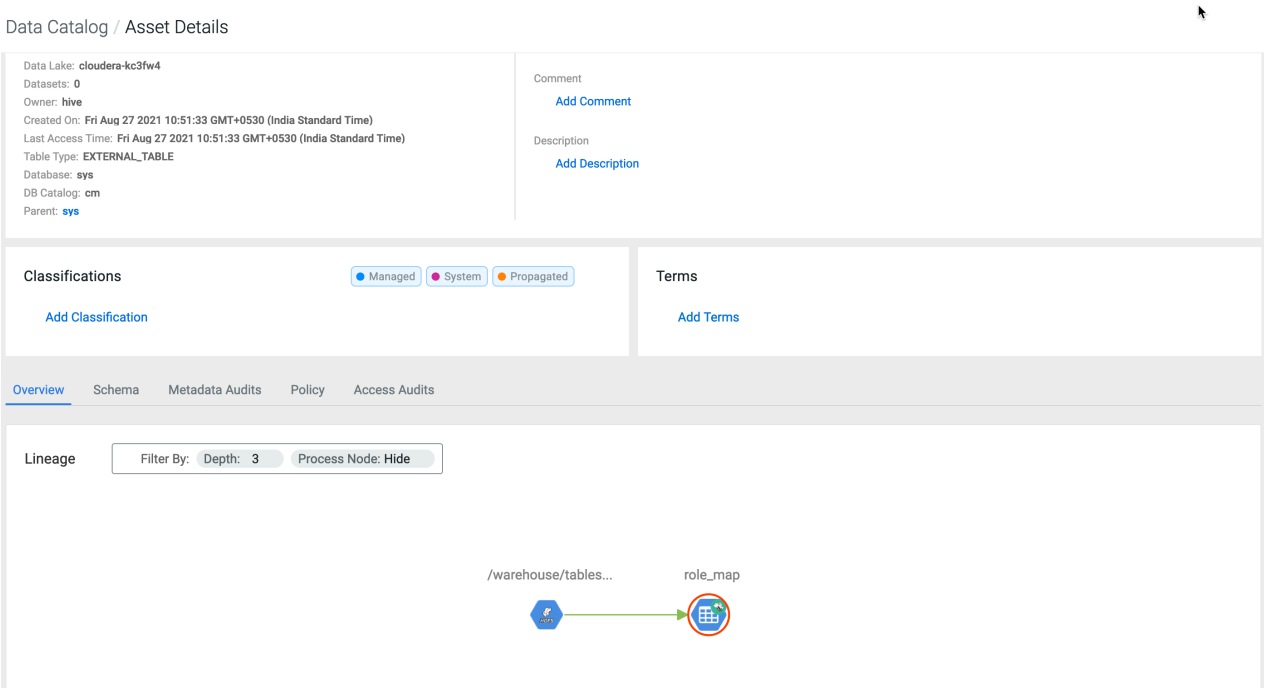
- **Access Type**
  - SELECT
  - UPDATE
  - CREATE
  - DROP
  - ALTER
  - INDEX
  - READ
  - WRITE
- **Result**
  - ALL
  - ALLOWED
  - DENIED


<sup>3</sup> Iceberg assets are discoverable in VM-based environments but they can be profiled only in Compute Cluster enabled environments.

## 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:



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:

## Data Catalog / Asset Details

Table Type: EXTERNAL\_TABLE  
Database: sys  
DB Catalog: cm  
Parent: sys

[Add Description](#)

**Classifications**

[Add Classification](#)

Managed System Propagated

**Terms**

[Add Terms](#)

**Overview** Schema Metadata Audits Policy Access Audits

**Lineage**

Filter By: Depth: 9 Process Node: Show

→ Lineage → Impact → Replication ○

**partition\_params**

Guid: [redacted]

Type Name: hive\_table

Classifications(0): --

Owner: hive

Qualified Name: sys.partition\_params@cm

Created On: Fri Aug 27 2021 10:51:31 GMT+0530 (India Standard Time)

Last Access Time: Fri Aug 27 2021 10:51:31 GMT+0530 (India Standard Time)

Table Type: EXTERNAL\_TABLE

Database: sys

DB Catalog: cm

## 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.

## Data Catalog / Asset Details

**campaigns**

[Atlas](#) [Feedback](#)

[Add to Dataset](#)

**Properties**

Type: HIVE TABLE  
# of Columns: 5  
Data Lake: [redacted]  
Datasets: 0  
Owner: hive  
Created On: Wed Apr 07 2021 15:27:03 GMT+0530 (India Stand...  
Last Access Time: Wed Apr 07 2021 15:27:03 GMT+0530 (Indi...  
Table Type: EXTERNAL\_TABLE  
Database: marketing  
DB Catalog: cm  
Parent: marketing

Qualified Name  
marketing.campaigns@cm

Comment  
This table contains marketing campaigns information.

Description  
[+ Add Description](#)

**Profilers | 2**

**Cluster Sensitivity Profiler**  
Last run: 2 hours ago | Status: SUCCESS  
Next Schedule Run: NA, Profiler is Disabled. [Run](#)

**Hive Column Profiler**  
Last run: 2 hours ago | Status: SUCCESS  
Next Schedule Run: NA, Profiler is Disabled. [Run](#)

**Classifications**

[+ Add Classification](#)

Managed System Propagated

**Terms | 1**

[new\\_term1@new\\_glossary](#)

**Overview** Schema Metadata Audits Policy Access Audits

**Lineage**

[T](#) [Q](#) [Q](#) [v](#)

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 \*.

The screenshot shows the 'campaigns' asset details page. The 'Add to Dataset' modal is open, displaying a table of existing datasets. The modal has search bars for 'Name' and 'Tag', and a '+ New Dataset' button. The table lists datasets with checkboxes for selection.

<input type="checkbox"/>	Name	Owner	# of Assets
<input checked="" type="checkbox"/>	Credit card	Srinivas Sudhindra	1
<input type="checkbox"/>	Town1234	Srinivas Sudhindra	1
<input checked="" type="checkbox"/>	City	Srinivas Sudhindra	1
<input type="checkbox"/>	Visa	Srinivas Sudhindra	1
<input checked="" type="checkbox"/>	SSN1	Srinivas Sudhindra	1
<input type="checkbox"/>	Passport123	Srinivas Sudhindra	1

Buttons at the bottom of the modal: Add Asset, Cancel.

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).

The screenshot shows the 'warehouse' asset details page. The 'New Dataset' modal is open, allowing the user to create a new dataset. The modal includes fields for Name, Description, and Tags, along with a 'Public' toggle and 'Create' and 'Cancel' buttons.

Fields in the 'New Dataset' modal:

- Name:
- Description:
- Tags:
- Public: ☒

Buttons at the bottom of the modal: Create, Cancel.

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.

Data Catalog / Asset Details

Add asset to datasets completed

warehouse

**Properties**

Type: HIVE TABLE  
 # of Columns: 14  
 Data Lake:   
 Datasets: 6  
 Owner:   
 Created On: Mon May 11 2020 13:02:22 GMT+0530 (India Stan...  
 Last Access Time: Mon May 11 2020 13:02:22 GMT+0530 (Indi...  
 Table Type: EXTERNAL\_TABLE  
 Database:   
 DB Catalog: cm  
 Parent:

Qualified Name  
 Comment  
 + Add Comment  
 Description  
 + Add Description

**Classifications**  
 + Add Classification

Overview Schema Metadata Audits Policy Access Audits

Lineage

**Add to Dataset**

Search by Name Search by Tag + New Dataset

<input type="checkbox"/>	Name↑	Owner↑	# of Assets↑
<input checked="" type="checkbox"/>	Link test	Srinivas Sudhindra	1
<input checked="" type="checkbox"/>	Demotest	Srinivas Sudhindra	1
<input checked="" type="checkbox"/>	Social Security Number	Srinivas Sudhindra	1
<input checked="" type="checkbox"/>	Visa number	Srinivas Sudhindra	1
<input type="checkbox"/>	passport12	Srinivas Sudhindra	0
<input type="checkbox"/>	Demo1234	Srinivas Sudhindra	0
<input checked="" type="checkbox"/>	Test1357	Srinivas Sudhindra	2
<input type="checkbox"/>	Demo	Srinivas Sudhindra	1
<input checked="" type="checkbox"/>	Hive123	Srinivas Sudhindra	2

Feedback

## 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.

Overview Schema **Metadata Audits** Policy Access Audits

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.

ATLAS\_ENTITY\_AUDIT\_EVENTS

Atlas

Properties

Type: HBASE TABLE

Data Lake:

Owner: atlas

Created On: Wed Apr 07 2021 10:50:42 GMT+0530 (India Standard Time)

Modified Time: Wed Apr 07 2021 10:50:42 GMT+0530 (India Standard Time)

Namespace GUID: f4658406-3a4b-4076-afb0-6e74745934b6

URI: ATLAS\_ENTITY\_AUDIT\_EVENTS

Parent: default

Qualified Name

default:ATLAS\_ENTITY\_AUDIT\_EVE ...

Description

ATLAS\_ENTITY\_AUDIT\_EVENTS

Classifications

Managed

System

Propagated

+ Add Classification

Terms

+ Add Terms

Overview

Schema

Metadata Audits

Policy

Access Audits

Feedback

≡

≡

User	Timestamp	Actions
atlas	Wed Apr 07 2021 10:50:48 GMT+0530 (India Standard Time)	Entity created

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

ATLAS\_ENTITY\_AUDIT\_EVENTS

Atlas

Properties

Type: HBASE TABLE

Data Lake:

Owner: atlas

Created On: Wed Apr 07 2021 10:50:42 GMT+0530 (India Standard Time)

Modified Time: Wed Apr 07 2021 10:50:42 GMT+0530 (India Standard Time)

Namespace GUID: f4658406-3a4b-4076-afb0-6e74745934b6

URI: ATLAS\_ENTITY\_AUDIT\_EVENTS

Parent: default

Qualified Name

default:ATLAS\_ENTITY\_AUDIT\_EVE ...

Description

ATLAS\_ENTITY\_AUDIT\_EVENTS

Classifications

Managed

System

Propagated

+ Add Classification

Terms

+ Add Terms

Overview

Schema

Metadata Audits

Policy

Access Audits

Feedback

≡

≡

User	Timestamp	Actions
atlas	Wed Apr 07 2021 10:50:48 GMT+0530 (India Standard Time)	Entity created

Created:

```
{
  "typeName": "hbase_table",
  "attributes": {
    "owner": "atlas",
    "isNormalizationEnabled": false,
  }
}
```

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.

Data Catalog / Asset Details

ATLAS\_ENTITY\_AUDIT\_EVENTS

Properties

Type: HBASE TABLE

Data Lake:

Owner: atlas

Created On: Wed Apr 07 2021 10:50:42 GMT+0530 (India Standard Time)

Modified Time: Wed Apr 07 2021 10:50:42 GMT+0530 (India Standard Time)

Namespace GUID: f4658406-3a4b-4076-afb0-6e74745934b6

URI: ATLAS\_ENTITY\_AUDIT\_EVENTS

Parent: default

Qualified Name

default:ATLAS\_ENTITY\_AUDIT\_EVE ...

Description

ATLAS\_ENTITY\_AUDIT\_EVENTS

Classifications

Managed

System

Propagated

+ Add Classification

Terms

+ Add Terms

Overview

Schema

Metadata Audits

Policy

Access Audits

Timeline

Wed Apr 07 2021 10:50:48 GMT+0530 (India Standard Time)

atlas

Entity created

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

Data Catalog / Asset Details

ATLAS\_ENTITY\_AUDIT\_EVENTS

Properties

Type: HBASE TABLE

Data Lake:

Owner: atlas

Created On: Wed Apr 07 2021 10:50:42 GMT+0530 (India Standard Time)

Modified Time: Wed Apr 07 2021 10:50:42 GMT+0530 (India Standard Time)

Namespace GUID: f4658406-3a4b-4076-afb0-6e74745934b6

URI: ATLAS\_ENTITY\_AUDIT\_EVENTS

Parent: default

Qualified Name

default:ATLAS\_ENTITY\_AUDIT\_EVE ...

Description

ATLAS\_ENTITY\_AUDIT\_EVENTS

Classifications

Managed

System

Propagated

+ Add Classification

Terms

+ Add Terms

Overview

Schema

Metadata Audits

Policy

Access Audits

Timeline

Wed Apr 07 2021 10:50:48 GMT+0530 (India Standard Time)

atlas

Entity created

Details

atlas

Snapshot

Created:

{  
 "typeName": "hbase\_table",  
 "attributes": {  
 "owner": "atlas",  
 ...  
 }  
}

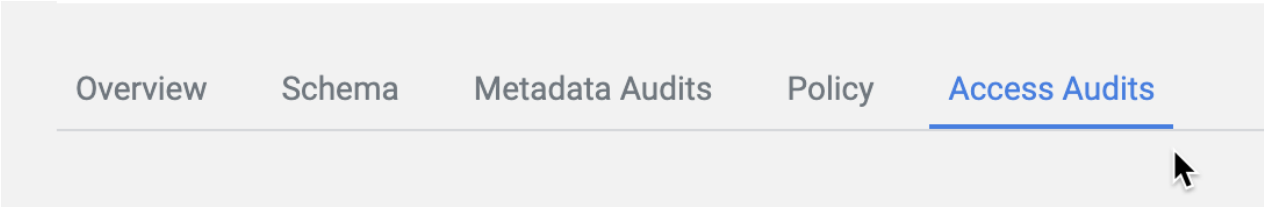
Viewing Ranger access audits

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
  - 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:

Access Type: ALL Result: ALL							
Policy ID	Event Time	User	Resource Type	Access Type	Result	Access Enforcer	Client IP
9	01/01/1970 01:00 AM CEST	hive	@table	CREATE	ALLOWED	ranger-acl	10.80.214.142
9	01/01/1970 01:00 AM CEST	hive	@table	DROP	ALLOWED	ranger-acl	10.80.214.142
9	01/01/1970 01:00 AM CEST	hive	@table	CREATE	ALLOWED	ranger-acl	10.80.214.142



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

Policy Details Based On Versions: 1

Select

Policy Id : 9

Policy Name : all - database, table, column

Policy Labels : --

Service Type : hive

Policy Type : Access

Policy Version : 1

Tag : --

Service Name : cm\_hive

Create Time : 07/14/2024 10:36 PM CEST

Update Time : 07/14/2024 10:36 PM CEST

Created By : Admin

Updated By : Admin

ALLOW CONDITIONS :

USERS

hive

beacon

trino

dpProfiler

hue

admin

impala

rangerlookup

{OWNER}

hdfs

GROUPS

\_c\_ranger\_admins\_2af99bc8

PERMISSIONS

select

update

create

drop

alter

index

lock

all

read

write

repladmin

serviceadmin

tempudfadmin

refresh

rwstorage

read

select

all

all

Close

## 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.

Overview	Schema	Metadata Audits	Policy	Access Audits		
Resource Based Policies						
Policy ID	Policy Name	Status	Audit Logging	Group	Users	
9	all - database, table, column	ENABLED	ENABLED	_c_ranger_admins_2af99bc8	hive, beacon, trino, dpProfiler, hue, a...	
10	all - database, table	ENABLED	ENABLED	_c_ranger_admins_2af99bc8	hive, beacon, trino, dpProfiler, hue, a...	
Tag Based Policies						
Policy ID	Policy Name	Tags	Status	Audit Logging	Group	Users
123	Test Tag policy	test_tag_for_provider_summary	ENABLED	ENABLED		

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

42