Data Catalog 1.5.2

Data Catalog Reference

Date published: 2023-11-02
Date modified:

CLOUDERA

Legal Notice

Contents

Sample Workflows

When you work with Data Catalog, you can use the workflows described in this section.

Tag Management and Search workflow

In Data Catalog, you can create a tag, assign the tag to an asset, and search for an asset using various filters.

About this task
You can create a tag and search for any asset using the tag. You can also apply a tag to a column. Later, you can search assets based on the Entity and Column filters.

Procedure

1. Data Catalog > Atlas Tags > Create a new tag
2. Enter the name, description for the tag.
3. Inserting classification to inherit attributes and adding new attributes are optional.
4. Click Save.
   The new tag is created.

   Next, you must assign the newly created tag to one of the existing assets in Data Catalog.

   For example, we can assign the tag to the Hive table.
5. Data Catalog > Search for a Hive asset > click on the asset to view the Asset Details page.
6. Under the Manage Classification panel, click + to enable the tag list view.
7. Select the tag that you newly created and assign the tag to the asset.
8. Click Save.

   Next, you can add the tag to a column in the selected asset.
9. Data Catalog > Search for a Hive asset > click on the asset to view the Asset Details page.
10. Select the Schema tab and click Edit Tags.
11. Click + and select the tag from the tag drop-down.
    The tag list contains System Tags and Managed Tags. Managed Tags are selected by default.
12. Select a tag and click Save.
    Once tags are added to the table and column, you can search for the corresponding assets using the tag name.

    To search for the asset using the Entity filter:
13. Data Catalog > Search > Enter the tag name in the Entity filter.
    The asset for which the tag was added is displayed in the search result.

    To search for the asset using the Column filter:
14. Data Catalog > Search > Enter the tag name in the Column filter.
    The asset for whose column the tag was added is displayed in the search result.

Auto Tagging workflow with Custom Cluster Sensitivity Profiler Rules

You can auto tag workflows while working with Custom Cluster Sensitivity Profiler.

About this task
Use the following information to create a custom tag and assign the same to the Custom Sensitivity Profiler.

Procedure

1. Data Catalog > Profilers > Select the Tag Rules tab.
2. Click + New to open the Custom Rule window.
3. Under the Resources pane, click + to open the Regular Expression Editor window.
4. Enter the name and input the regular expression value in such a way that it matches the test string.
   If your regular expression value is [a-z][a-z][a-z][a-z] and the test string is "baby", there is a match.
5. Click Save.
6. On the Custom Rule window, enter the name and description.
7. Enter the tags value and select the Column Expression Name from the drop-down.

   You must select the same regular expression you had created under the Resources pane.
8. Enter the tags value and select the Column Value Expression from the drop-down.

   You must select the same regular expression you had created under the Resources pane.
9. Click Save & Validate.
   The Data For Validation window appears.
10. Enter the sample values to validate if Column Expression Name and Column Value Expression entities match.
    Make sure that the correct data lake is selected to validate the entries.
11. Click Submit Validation.
    The status for the newly created regular expression validation is displayed on the Tags Rule tab. Once the validation is successful, you can deploy the rule.
12. Click Done.
    On the Rule Groups pane, verify if the rule is available under the Custom Deployed list. You can also suspend the tag by selecting the same from the list.

    Once the Cluster Sensitivity Profiler job or On-Demand Profiler picks up the Hive asset for profiling, the newly set-up custom tag must get applied on the Hive column, provided the asset has the column(s) which meet the custom rule criteria.