Tagging Dataplex Assets | Google Cloud Skills Boost

Qwiklabs : 9-11 minutes

GSP1145

Google Cloud Self-Paced Labs

Overview

Dataplex is an intelligent data fabric that enables organizations to centrally discover, manage, monitor, and govern their data across data lakes, data warehouses, and data marts to power analytics at scale.

Data Catalog is a fully managed, scalable metadata management service within Dataplex that you can use to tag data assets and search for assets to which you have access. Tags allow you to attach custom metadata fields to specific data assets for easy identification and retrieval (such as tagging certain assets as containing protected or sensitive data); you can also create reusable tag templates to rapidly assign the same tags to different data assets.

In this lab, you learn how to use Data Catalog within Dataplex to create a tag template, apply it to Dataplex assets, and then use the tag to search for assets.

What you'll do

- Enable the Dataplex and Data Catalog APIs

- Create a lake, zone, and asset in Dataplex

- Create a tag template

- Apply a tag template to Dataplex assets

- Search for assets using tags

Setup and requirements

Before you click the Start Lab button

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources will be made available to you.

This hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access Google Cloud for the duration of the lab.

To complete this lab, you need:

- Access to a standard internet browser (Chrome browser recommended).

Note: Use an Incognito or private browser window to run this lab. This prevents any conflicts between your personal account and the Student account, which may cause extra charges incurred to your personal account.

- Time to complete the lab---remember, once you start, you cannot pause a lab.

Note: If you already have your own personal Google Cloud account or project, do not use it for this lab to avoid extra charges to your account.

How to start your lab and sign in to the Google Cloud Console

1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is the **Lab Details** panel with the following:

    - The **Open Google Console** button
    - Time remaining
    - The temporary credentials that you must use for this lab
    - Other information, if needed, to step through this lab

2. Click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Sign in** page.

    Tip: Arrange the tabs in separate windows, side-by-side.

    Note: If you see the **Choose an account** dialog, click **Use Another Account**.

3. If necessary, copy the **Username** from the **Lab Details** panel and paste it into the **Sign in** dialog. Click **Next**.

4. Copy the **Password** from the **Lab Details** panel and paste it into the **Welcome** dialog. Click **Next**.

    Important: You must use the credentials from the left panel. Do not use your Google Cloud Skills Boost credentials. **Note:** Using your own Google Cloud account for this lab may incur extra charges.

5. Click through the subsequent pages:

    - Accept the terms and conditions.
    - Do not add recovery options or two-factor authentication (because this is a temporary account).
    - Do not sign up for free trials.

After a few moments, the Cloud Console opens in this tab.

Note: You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-left.
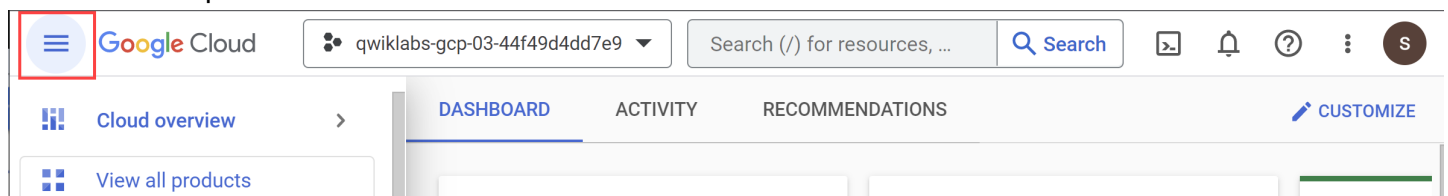
Enable Dataplex and Data Catalog APIs

1. In the Google Cloud Console, enter **Cloud Dataplex API** in the top search bar.

2. Click on the result for **Cloud Dataplex API** under Marketplace.

3. Click **Enable**.

4. Repeat steps 1-3 for **Google Cloud Data Catalog API**.

Task 1. Create a lake, zone, and asset

In this task, you create a new Dataplex lake to store customer order information, add a curated zone to the lake, and then attach a pre-created BigQuery dataset as a new asset in the zone.

Create a lake

1. In the Google Cloud Console, in the **Navigation menu** (≡), navigate to **Analytics** > **Dataplex**.

If prompted `Welcome to the new Dataplex experience`, click **Close**.

2. Under **Manage lakes**, click **Manage**.

3. Click **Create lake**.

4. Enter the required information to create a new lake:

Property	Value
Display Name	`Orders Lake`
ID	Leave the default value.
Region	

Leave the other default values.

5. Click **Create**.

It can take up to 3 minutes for the lake to be created.

Add a zone to the lake

1. On the **Manage** tab, click on the name of your lake.

2. Click **Add zone**.

3. Enter the required information to create a new zone:

Property	Value
Display Name	`Customer Curated Zone`
ID	Leave the default value.
Type	**Curated zone**

Data locations Regional

Leave the other default values.

For example, the option for **Enable metadata discovery** under **Discovery settings** is enabled by default and allows authorized users to discover the data in the zone.

    4. Click **Create**.

It can take up to 2 minutes for the zone to be created.

You can perform the next task once the status of the zone is **Active**.

Attach an asset to a zone

1. On the **Zones** tab, click on the name of your zone.

2. On the **Assets** tab, click **Add assets**.

3. Click **Add an asset**.

4. Enter the required information to attach a new asset:

Property	Value
Type	**BigQuery dataset**
Display Name	`Customer Details Dataset`
ID	Leave the default value.
Dataset	.customers

Leave the other default values.

5. Click **Done**.

6. Click **Continue**.

7. For **Discovery settings**, select **Inherit** to inherit the Discovery settings from the zone level, and then click **Continue**.

8. Click **Submit**.

Create a lake, zone, and asset in Dataplex

Task 2. Create a tag template

To start tagging data, you first need to create one or more tag templates. A tag template can be a public or private tag template. When you create a new tag template, the option to create a public tag template is the default and recommended option. Users who have the required view permissions for a data asset can view all the public tags associated with it. This supports simple search for discovery while also adhering to data access controls already implemented on the underlying data.

In this task, you create a public tag template to label BigQuery table columns with a protected status. With a public tag template, users who have access to the underlying BigQuery table columns will be able to see the tags applied to the columns.

1. On the left menu, under **Manage catalog**, click **Tag templates**.

2. Click **Create tag template**.

3. Enter the required information to define the tag template:

Property	Value
Template Display Name	Protected Data Template
Template ID	Leave the default value.
Location	
Visibility	**Public**

4. Click **Add field**, and enter the required information to add a new field to the template:

Property	Value
Field Display Name	Protected Data Flag
Field ID	Leave the default value.
Type	**Enumerated**

5. For **Enumerated values** > **Values 1**, enter YES.

6. Click **Add value**, and for **Values 2**, enter NO.

7. Click **Done**.

8. Click **Create**.

Create a tag template

Task 3. Apply a tag template to Dataplex assets

After you create a tag template, you can use it to attach tags to any number of desired data assets to which you have access.

In this task, you apply your previously created tag template to specific columns in the BigQuery table that you want to label with a protected data status.

1. On the left menu, under **Discover**, click **Search**.

2. For **Filters** > **Systems**, enable the checkbox for **Dataplex**.

3. Click on the **customer_details** table.

If you do not see the **customer_details** table, wait a few minutes and then refresh the page to allow the Dataplex asset list to be updated.

4. Click **Attach tags**.

5. For **Choose what to tag**, enable the checkboxes for the following columns:

   - zip
   - state
   - last_name
   - country
   - email
   - latitude
   - first_name
   - city
   - longitude

6. Click **OK**.

7. For **Choose the tag templates**, select **Protected data template**.

8. For **Protected data flag**, select **YES**.

9. Click **Save**.

Apply a tag template to Dataplex assets

Task 4. Search for assets using tags

After you have tagged data assets, you can search for them using Data Catalog within Dataplex.

In this task, you search for the assets that have been tagged using the Protected Data tag template.

1. On the left menu, under **Discover**, click **Search**.

2. For **Filters** > **Tags**, enable the checkbox for **Protected data template**.

3. Click on the **customer_details** table.

4. Click on the **Schema and column tags** tab to see the Protected Data tags on the specified columns.

Congratulations!

You used Data Catalog within Dataplex to create a tag template, applied it to Dataplex assets, and then used the tag to search for assets.

Google Cloud training and certification

...helps you make the most of Google Cloud technologies. Our classes include technical skills and best practices to help you get up to speed quickly and continue your learning journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. Certifications help you validate and prove your skill and expertise in Google Cloud technologies.

Manual Last Updated May 2, 2023

Lab Last Tested May 2, 2023