www.cloudskillsboost.google /games/4111/labs/26485

Dataplex: Qwik Start - Command Line | Google Cloud Skills Boost

Qwiklabs：11-13 minutes

GSP1144

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. Specifically, you can use Dataplex to build a data mesh architecture, which is an organizational and technical approach that decentralizes data ownership among domain data owners.

Dataplex manages data in a way that doesn't require data movement or duplication. As you add new data assets, Dataplex harvests the metadata for both structured and unstructured data, and automatically registers all metadata in a secure, unified metastore. Data and metadata can then be assessed via Google Cloud services such as Data Catalog and BigQuery.

In this lab, you learn how to start building your own data mesh by creating and removing lakes, zones, and assets in Dataplex using the command line.

What you'll do

- Enable the Dataplex API

- Create a lake

- Add a zone to your lake

- Attach and detach assets

- Delete zones and lakes

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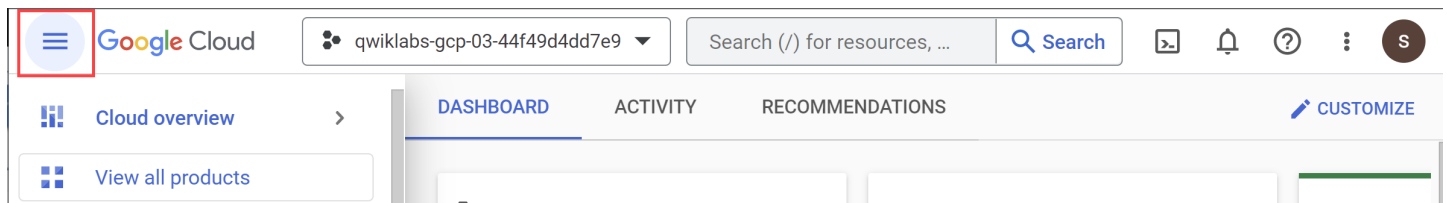
   Important: You must use the credentials from the left panel. Do not use your Google Cloud Skills Boost credentials. **Note:** Using your own Google Cloud account for this lab may incur extra charges.

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.

Google Cloud   qwiklabs-gcp-03-44f49d4dd7e9 ▼   Search (/) for resources, ...   🔍 Search

||| **Cloud overview**   >   DASHBOARD   ACTIVITY   RECOMMENDATIONS   ✏ CUSTOMIZE

▦ View all products

Activate Cloud Shell

Cloud Shell is a virtual machine that is loaded with development tools. It offers a persistent 5GB home directory and runs on the Google Cloud. Cloud Shell provides command-line access to your Google Cloud resources.

1. Click **Activate Cloud Shell** 🖥 at the top of the Google Cloud console.

When you are connected, you are already authenticated, and the project is set to your **PROJECT_ID**. The output contains a line that declares the **PROJECT_ID** for this session:

Your Cloud Platform project in this session is set to YOUR_PROJECT_ID

`gcloud` is the command-line tool for Google Cloud. It comes pre-installed on Cloud Shell and supports tab-completion.

2. (Optional) You can list the active account name with this command:

gcloud auth list

3. Click **Authorize**.

4. Your output should now look like this:

Output:

ACTIVE: * ACCOUNT: student-01-xxxxxxxxxxxx@qwiklabs.net To set the active account, run: $ gcloud config set account `ACCOUNT`

5. (Optional) You can list the project ID with this command:

gcloud config list project

Output:

[core] project = <project_ID>

Example output:

[core] project = qwiklabs-gcp-44776a13dea667a6 **Note:** For full documentation of `gcloud`, in Google Cloud, refer to the gcloud CLI overview guide.

Enable the Dataplex API and set variables

1. In Cloud Shell, run the following command to enable the Dataplex API.

gcloud services enable \ dataplex.googleapis.com

2. Run the following command to create a variable for project ID:

```
export PROJECT_ID=$(gcloud config get-value project)
```

3. Run the following command to create a variable for region:

```
export REGION={{{project_0.default_region | ""}}} gcloud config set compute/region $REGION
```

Task 1. Create a lake

In Dataplex, a lake is the highest organizational domain that represents a specific data area or business unit. For example, you can create a lake for each department or data domain in your organization, so that you can organize and provide data for specific user groups.

In this task, you use the command line to create a lake to start building a data mesh.

- In Cloud Shell, run the following command to create a new lake called **Ecommerce**:

```
gcloud dataplex lakes create ecommerce \ --location=$REGION \ --display-name="Ecommerce" \ --description="Ecommerce Domain"
```

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

If you receive an error of Status code: 403. Permission 'dataplex.lakes.create' denied, wait a few minutes and try again. The permissions may need more time to be granted after enabling the Dataplex API in the previous Setup section.

The final output should look similar to the following:

Created [ecommerce] Lake created in [projects/$PROJECT_ID/locations/$REGION]. Create a Dataplex lake

Task 2. Add a zone to your lake

After you create a lake, you can add zones to the lake. Zones are subdomains within a lake that you can use to categorize data further. For example, you can categorize data by stage, usage, or restrictions.

There are two types of zones:

- Raw zones contain data in raw formats (such as files in Cloud Storage buckets) and are not subject to strict type-checking.
- Curated zones contain data that is cleaned, formatted, and ready for analytics such as BigQuery datasets.

In this task, you use the command line to create a curated zone for working with BigQuery datasets.

- In Cloud Shell, run the following command to create a new curated zone called **Orders Curated Zone** with metadata discovery enabled:

```
gcloud dataplex zones create orders-curated-zone \ --location=$REGION \ --lake=ecommerce \ --display-name="Orders Curated Zone" \ --resource-location-type=SINGLE_REGION \ --type=CURATED \ --
```

discovery-enabled \ --discovery-schedule="0 * * * *"

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

The final output should look similar to the following:

Created [orders-curated-zone] Zone created in
[projects/$PROJECT_ID/locations/$REGION/lakes/ecommerce]. Add a zone to your lake

Task 3. Attach an asset to a zone

Data stored in Cloud Storage buckets or BigQuery datasets can be attached as assets to zones within a Dataplex lake.

In this task, you use the command line to create a BigQuery dataset and then attach the dataset to the previously created zone.

Create a BigQuery dataset

- In Cloud Shell, run the following command to create a new BigQuery dataset called **orders**:

bq mk --location=$REGION --dataset orders

While this dataset does not contain any tables or data, you can attach it to the zone now, and newly created tables and loaded data will automatically be integrated into the zone.

The final output should look similar to the following:

Dataset '$PROJECT_ID:orders' successfully created.

Attach the BigQuery dataset to the zone

- In Cloud Shell, run the following command to attach the BigQuery dataset to the zone as an asset called **Orders Curated Data** with metadata discovery enabled:

gcloud dataplex assets create orders-curated-dataset \ --location=$REGION \ --lake=ecommerce \ --zone=orders-curated-zone \ --display-name="Orders Curated Dataset" \ --resource-type=BIGQUERY_DATASET \ --resource-name=projects/$PROJECT_ID/datasets/orders \ --discovery-enabled

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

The final output should look similar to the following:

Created [orders-curated-dataset] Asset created in
[projects/$PROJECT_ID/locations/$REGION/lakes/ecommerce/zones/orders-curated-zone]. Attach an asset to a zone

Task 4. Delete assets, zones, and lakes

To delete a lake, you must first detach assets and then delete the zones.

In this task, you use the command line to detach the asset from the zone, then delete the zone, and last, delete the lake.

Detach an asset

- In Cloud Shell, run the following command to detach the BigQuery dataset from the zone:

gcloud dataplex assets delete orders-curated-dataset --location=$REGION --zone=orders-curated-zone --lake=ecommerce

If prompted to confirm, enter Y.

This action does delete the underlying data in the BigQuery dataset. It simply removes the BigQuery dataset from being accessible or discoverable using the lake in Dataplex.

The final output should look similar to the following:

Deleted asset [orders-curated-dataset].

Delete a zone

- In Cloud Shell, run the following command to delete the zone:

gcloud dataplex zones delete orders-curated-zone --location=$REGION --lake=ecommerce

If prompted to confirm, enter Y.

The final output should look similar to the following:

Deleted zone [orders-curated-zone].

Delete the lake

- In Cloud Shell, run the following command to delete the lake:

gcloud dataplex lakes delete ecommerce --location=$REGION

If prompted to confirm, enter Y.

The final output should look similar to the following:

Deleted lake [ecommerce]. Delete assets, zone and dataplex lake

Congratulations!

You used the command line to create and delete lakes, zones, and assets in Dataplex.

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 April 26, 2023

Lab Last Tested April 26, 2023