

Dataplex: Qwik Start - Console | Google Cloud Skills Boost

Qwiklabs : 9-11 minutes

GSP1143



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 Google Cloud console.

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**.

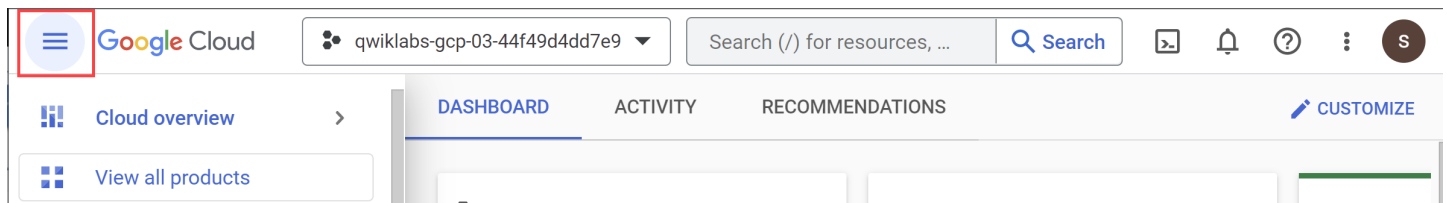
3. If necessary, copy the **Username** from the **Lab Details** panel and paste it into the **Sign in** dialog. Click **Next**.
4. Copy the **Password** from the **Lab Details** panel and paste it into the **Welcome** dialog. Click **Next**.

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.



Enable the Cloud Dataplex API

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**.

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 create a lake to start building a data mesh.

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	sensors
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.

Click *Check my progress* to verify the objective. Create a data lake **sensors**

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 tables.

In this task, you create a raw zone for working with files in a Cloud Storage bucket.

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	temperature raw data
ID	Leave the default value.
Type	Raw 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**.

Click *Check my progress* to verify the objective. Add a zone **temperature raw data** to the lake sensors

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 attach a Cloud Storage bucket that you create in the Google Cloud console.

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	Storage bucket
Display Name	measurements

ID Leave the default value.

Leave the other default values.

5. For **Bucket**, click **Browse**.

You can attach an existing Cloud Storage bucket or create a new one without leaving Dataplex. In the next steps, you create a new Cloud Storage bucket and attach it to the zone.

6. Click **Create new bucket** ().

7. Provide your project ID as the bucket name (), and then click **Continue**.

8. For **Location type**, select **Region**, and then select .

Leave other default values.

9. Click **Create**.

If prompted Public access will be prevented, click **Confirm**.

10. Click **Select** to select the bucket you just created, and then click **Continue**.

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

12. Click **Submit**.

Click *Check my progress* to verify the objective. Create a cloud storage bucket and attach an asset **measurements** to the zone **temperature raw data**

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 detach the asset from the zone, then delete the zone, and last, delete the lake.

Detach an asset

1. On the left menu, click on **Manage** tab, and then click on the name of your lake.
2. On the **Zones** tab, click on the name of your zone.
3. On the **Assets** tab, enable the checkbox to the left of the asset name.
4. Click **Delete assets**.
5. Click **Delete** to confirm.

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

Delete a zone

1. On the left menu, click on **Manage** tab, and then click on the name of your lake.
2. On the **Zones** tab, enable the checkbox to the left of the zone name.
3. Click **Delete zone**.
4. Click **Delete** to confirm.

Delete the lake

1. On the left menu, click on **Manage** tab, and then click on the name of your lake.
2. At the top of the page, click **Delete**.
3. Confirm deletion by typing **delete** into the text box.
4. Click **Delete lake** to confirm.

Click *Check my progress* to verify the objective. Delete lake **sensors** and nested resources

Congratulations!

You used the Google Cloud console 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 May 3, 2023

Lab Last Tested May 3, 2023

Copyright 2023 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.