Build a Data Mesh with Dataplex: Challenge Lab

experiment Lab schedule 1 hour 30 minutes universal_currency_alt No cost

show_chart Introductory

GSP514

Overview

In a challenge lab you're given a scenario and a set of tasks. Instead of following step-by-step instructions, you will use the skills learned from the labs in the course to figure out how to complete the tasks on your own! An automated scoring system (shown on this page) will provide feedback on whether you have completed your tasks correctly.

When you take a challenge lab, you will not be taught new Google Cloud concepts. You are expected to extend your learned skills, like changing default values and reading and researching error messages to fix your own mistakes.

To score 100% you must successfully complete all tasks within the time period!

This lab is recommended for students who have enrolled in the Build a Data Mesh with Dataplex skill badge. Are you ready for the challenge?

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.

Challenge scenario

You are just starting your junior data engineer role. So far you have been helping teams create and manage Dataplex assets.

You are expected to have the skills and knowledge for these tasks.

Your challenge

You are asked to help a newly formed development team with building a new data mesh using Dataplex. Specifically, you need to create a Dataplex lake with multiple zones and assets. You also need to tag assets in the new lake and assess them for data quality; you receive the following request to complete the following tasks:

- Create a Dataplex lake with two zones and two assets.
- Create and apply a tag template to tag an entire zone as protected data.
- Assign a Dataplex IAM role to another user.
- Create and upload a data quality specification file to Cloud Storage.
- Define and run a data quality job in Dataplex.

Some standards you should follow:

- Ensure that any needed APIs (such as Dataplex, Data Catalog, and Dataproc) are successfully enabled.
- Create all resources in the `us-west1` region, unless otherwise directed.

Each task is described in detail below, good luck!

Task 1. Create a Dataplex lake with two zones and two assets

The Cloud Storage bucket and BigQuery dataset for step 2 have been pre-created in this lab.

1. Create a Dataplex lake named **Sales Lake** with two regional zones:

- Raw zone named **Raw Customer Zone**
- Curated zone named **Curated Customer Zone**

2. Attach one pre-created asset to each zone:

- To the raw zone, attach the Cloud Storage bucket named `qwiklabs-gcp-02-fbb8015ae5b2-customer-online-sessions` as a new asset named **Customer Engagements**.

- To the curated zone, attach the BigQuery dataset named `qwiklabs-gcp-02-fbb8015ae5b2.customer_orders` as a new asset named **Customer Orders**.

Click **Check my progress** to verify the objective.

Create a Dataplex lake with two zones and two assets

Check my progress

Assessment Completed!

Task 2. Create and apply a tag template to a zone

1. Create a public tag template named **Protected Customer Data Template** with two enumerated fields:

- First field named **Raw Data Flag** with two values: `Yes` and `No`.
- Second field named **Protected Contact Information Flag** with two values: `Yes` and `No`

2. Use this template to tag the **Raw Customer Zone** using a value of `Yes` for both flags.

Click **Check my progress** to verify the objective.

Create and apply a tag template to a zone

Check my progress

Please create the tag template using the provided name and configuration.

Task 3. Assign a Dataplex IAM role to another user

- Using the principle of least privilege, assign the appropriate Dataplex IAM role to User 2 (`student-00-d577a2150fce@qwiklabs.net`) that allows them to upload new Cloud Storage files to the Dataplex asset named **Customer Engagements**.

Click **Check my progress** to verify the objective.

Assign a Dataplex IAM role to another user

Check my progress

Assessment Completed!

Task 4. Create and upload a data quality specification file to Cloud Storage

The Cloud Storage bucket for step 2 has been pre-created in this lab.

1. Create a data quality specification file named **dq-customer-orders.yaml** with the following specifications:

- **NOT NULL** rule applied to the **user_id** column of the **customer_orders.ordered_items** table
- **NOT NULL** rule applied to the **order_id** column of the **customer_orders.ordered_items** table

2. Upload the file to the Cloud Storage bucket named `qwiklabs-gcp-02-fbb8015ae5b2`**-dq-config**.

Click **Check my progress** to verify the objective.

Create and upload a data quality specification file

Check my progress

Assessment Completed!

Task 5. Define and run a data quality job in Dataplex

The BigQuery dataset for step 1 has been pre-created in this lab.

1. Define a data quality job using the **dq-customer-orders.yaml** file with the following specifications:

Property	Value
Data Quality Job Name	**Customer Orders Data Quality Job**
BigQuery destination table for the results	`qwiklabs-gcp-02-fbb8015ae5b2`**.orders_dq_dataset.results**
User service account	**Compute Engine default service account**

2. Run the data quality job immediately.

It can take several minutes for the job to run. You may need to refresh the page to see that the job has run successfully.

Click **Check my progress** to verify the objective.

Define and run a data quality job in Dataplex

Check my progress

Assessment Completed!

Congratulations!

You built a Data Mesh by creating a Dataplex lake with multiple zones and assets, tagging assets in the new lake, and assessing them for data quality.

Google Cloud

Build a Data Mesh with Dataplex

Smart Analytics

SKILL BADGE · INTRODUCTORY

Earn your next skill badge

This self-paced lab is part of the Build a Data Mesh with Dataplex skill badge. Completing this skill badge quest earns you the badge above, to recognize your achievement. Share your badge on your resume and social platforms, and announce your accomplishment using #GoogleCloudBadge.

This skill badge quest is part of Google Cloud's Data Engineer learning path. If you have already completed the other skill badge quests in this learning path, search the catalog for other skill badge quests in which you can enroll.

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 5, 2024

Lab Last Tested February 29, 2024