

Create a Secure Data Lake on Cloud Storage: Challenge Lab

Task 1

Create a Cloud Storage bucket

1. Sign into the project as **User 1**
2. Create a regional Cloud Storage bucket using the following bucket name:[PROJECT-ID]-bucket and replace the PROJECT_ID in the bucket name with the project ID provided at the left side of the lab instructions.
3. Use the same bucket you have created in the above step to attach as an asset to the zone

Objective: As **User 2**, create a dataset and a table from a GCS Avro file.

1. **Go to the BigQuery Console:** <https://console.cloud.google.com/bigquery>
2. **Switch to the correct project** (if needed).
3. In the **Explorer pane**, click on the project name > click **Create Dataset**.

Dataset Configuration:

- **Dataset ID:** `Raw_data`
 - **Data location:** Choose **Regional** (e.g., us-central1)
 - Leave all other options as default.
 - Click **Create Dataset**
4. After creating the dataset, click the **Raw_data** dataset > **Create Table**

Table Configuration:

- **Create table from:** Google Cloud Storage
- **Select file from GCS bucket:** `spls/gsp1145/users.avro`

- **File format:** AVRO
- **Table name:** public-data
- Leave the rest as default.
- Click **Create Table**

Task 2

Create a lake in Dataplex and add a zone to your lake

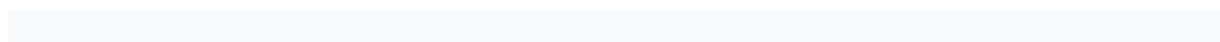
1. Sign into the project as **User 2**.
2. Create a lake in Dataplex using the below information:

Property	Value
Display Name	Customer-Lake
ID	Leave the default value.
Region	Region from the Lab Details panel which is located at the left side of the lab instructions

3. Add a zone to your lake:**Customer-Lake**. Use the information below:

Display Name	Value
Display Name	Public-Zone
ID	Leave the default value.
Type	Raw zone
Data locations	Regional
Discovery settings	Enable metadata discovery

4. For the **Labels** set key_1 as: domain_type and value_1 as: source_data



1. Go to **Dataplex** in the console: <https://console.cloud.google.com/dataplex>
2. Click on the **Lake** named `public-lake`
3. Under **Zones**, click **Add Zone**

Zone Configuration:

- **Display Name:** `temperature raw data`
- **ID:** Leave default
- **Type:** Raw zone
- **Data location:** Select **Regional**
- Leave everything else as default and click **Create**

Task 3

Create an entry group

1. Navigate to the Entry groups.
2. Use the below configuration for the entry group creation.

Property	Value
Entry Group Name	Custom entry group
Location	Region from the Lab Details panel which is located at the left side of the lab instructions

This will create the entry group.

1. In **Dataplex**, stay within the `public-lake`
2. Navigate to the **temperature raw data** zone
3. Click **Add Asset**

Asset Configuration:

- **Type:** BigQuery dataset
- **Display Name:** `Customer Details Dataset`
- **ID:** Leave default
- **Dataset:** Select the dataset named `customer_reference_data`

Discovery settings:

- Set to **Inherit from zone**

Click **Create**

Task 4

Create a tag template

To begin tagging data, an asset must be added to the zone . A tag template is a reusable structure that can be used to swiftly create new tags. You are required to structure the tags by topic using tag templates. Therefore, in this task you must create a tag template and attach a tag template to an asset and related fields as detailed in the following table.

Tag template name	Tag template ID	Location	Fields	Type
Customer Data Tag Template	customer_data_tag_template	Use the default region	Data Owner	String
			PII Data	Enumerated
				<ul style="list-style-type: none">• Value 1: Yes• Value 2: No

Therefore, attach a tag to the **Storage bucket** Data Catalog entry under the `CLOUD STORAGE` source system. Use the Customer Data Tag Template template to tag this entry, and provide the values for the tag fields provided in the following table.

Tag field	Value
Data Owner	Enter your name here
PII Data	Yes

Search for assets using a Tag Template: **Customer Data Tag Template**

1. Go to **Data Catalog**: <https://console.cloud.google.com/datacatalog>
2. Click **Tag Templates** > **Create Tag Template**

Template Configuration:

- **Template Display Name:** Protected Data Template
- **Template ID:** Leave default
- **Location:** Use default region
- **Visibility:** Public

3. Click **Add field**:

Field Configuration:

- **Field Display Name:** Protected Data Flag
- **Field ID:** Leave default
- **Type:** Enumerated
- Add Values:
 - Value 1: YES
 - Value 2: NO
- Click **Create**