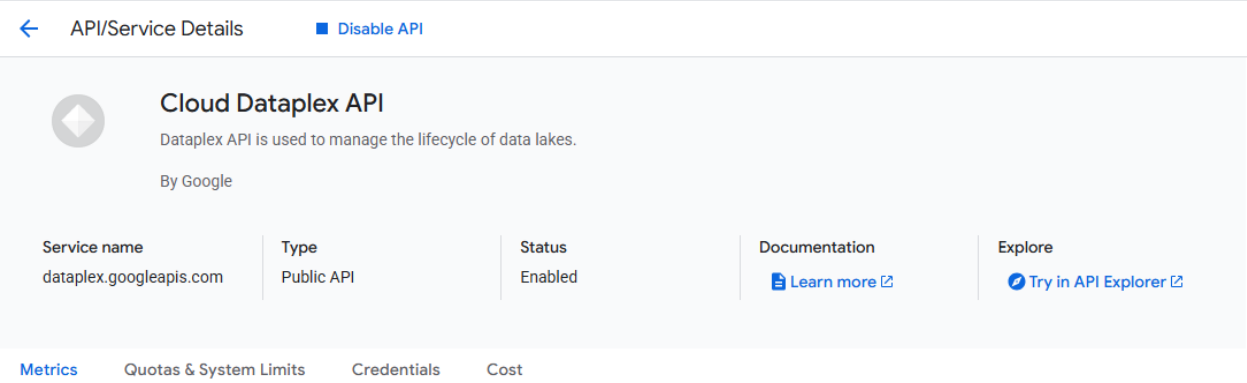
Dataplex: Qwik Start - Console

Objective

The goal of this lab is to familiarize users with **Dataplex Universal Catalog**, a tool that helps centralize the management, governance, and discovery of data across various data systems without requiring data movement. The lab simulates a data mesh setup where users create and manage lakes, zones, and assets via the Google Cloud Console.

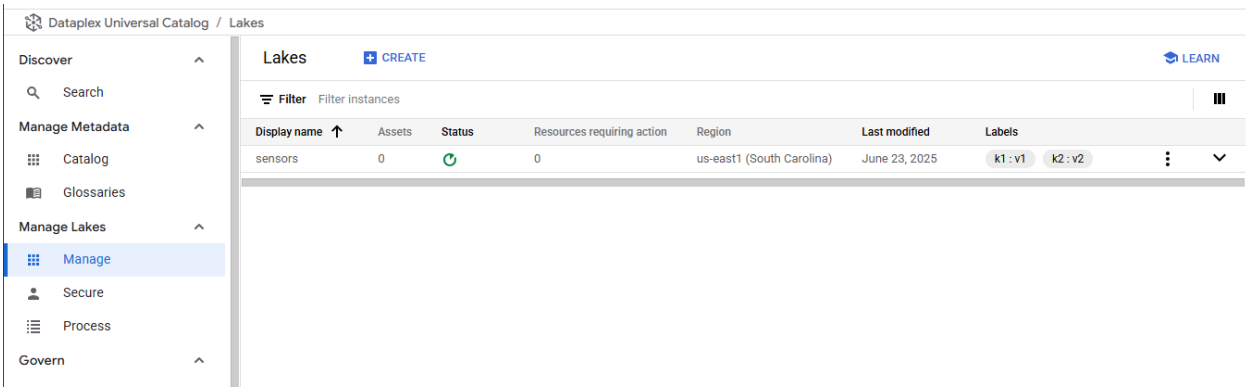
Task 1: Enable the Dataplex API

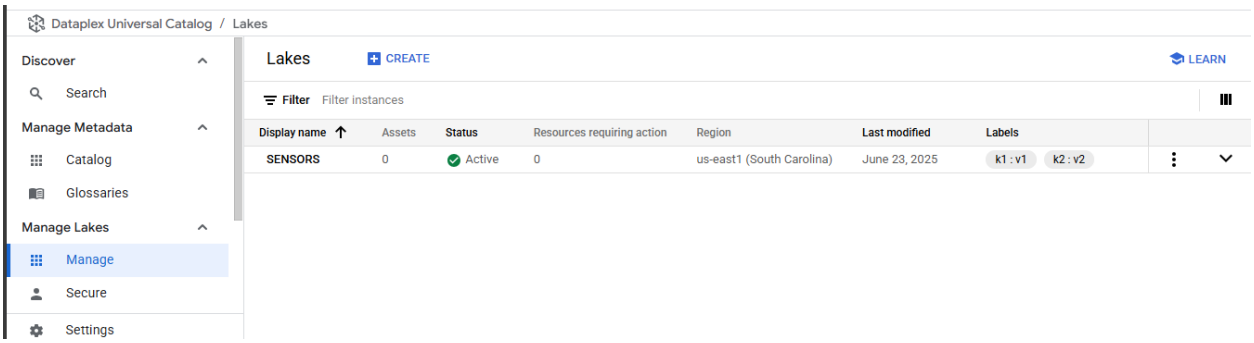
- Navigated to the **Dataplex API** in the Google Cloud Console.
- Enabled the API to access Dataplex services.



Task 2: Create a Lake

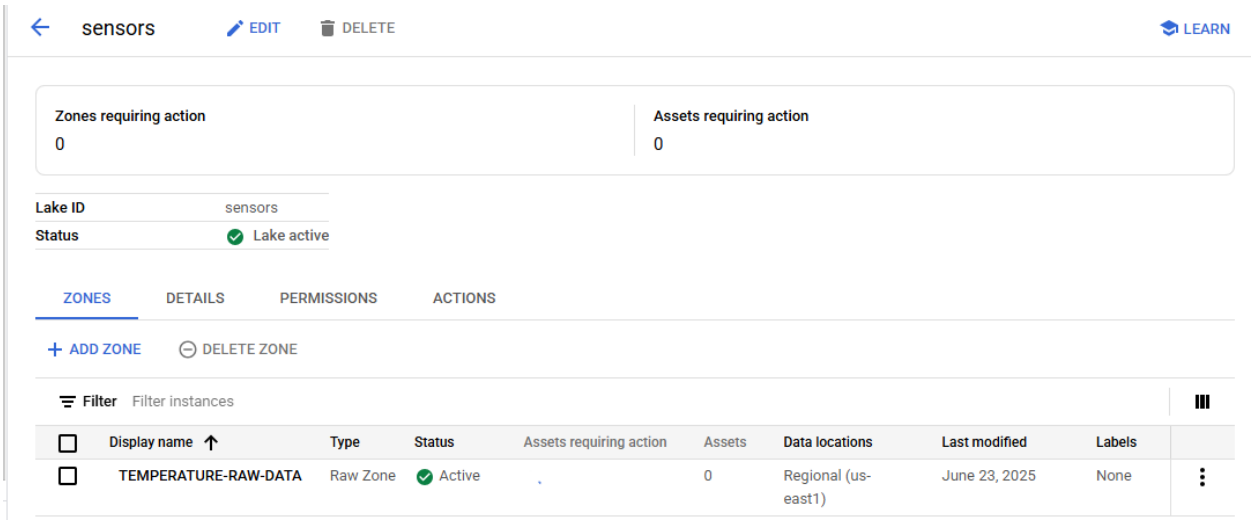
- Navigated to **Dataplex Universal Catalog** from the console.
- Created a **lake** named sensors with default ID and a selected region (e.g., us-central1).
- Verified lake creation after a short provisioning time.





Task 3: Add a Zone to the Lake

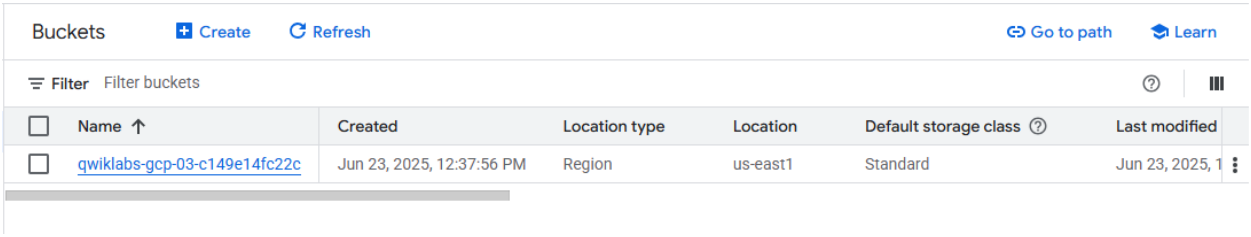
- Inside the sensors lake, created a **Raw Zone**:
  - **Name:** temperature raw data
  - **Type:** Raw zone
  - **Discovery settings:** Metadata discovery enabled by default
- Verified the zone was in an **Active** state before proceeding.



Task 4: Attach a Storage Asset to the Zone

- In the temperature raw data zone, created and attached a new **Cloud Storage bucket**:
  - **Asset Name:** measurements
  - **Bucket Name:** Used the project ID
  - **Location:** Regional
  - **Discovery Setting:** Inherited from the zone

- Confirmed asset was attached without moving or copying data.



Task 5: Delete Resources

- **Detached the asset** from the zone without deleting the underlying bucket data.
- **Deleted the zone** (temperature raw data) from the sensors lake.
- **Deleted the lake** after confirming removal of dependent components.

Learnings and Key Concepts

🔗 **Dataplex Universal Catalog**

- Provides a centralized catalog for **metadata** and governance without data duplication.
- Integrates with other GCP services like **BigQuery**, **Data Catalog**, and **Cloud Storage**.

🧩 **Data Mesh Architecture**

- Promotes **domain-based data ownership**.
- Enables scalable and decentralized data management.

📦 **Lakes, Zones, and Assets**

- **Lakes** group data domains (e.g., departmental data).
- **Zones** classify data stages (e.g., raw, curated).
- **Assets** are storage or datasets logically linked to zones

---

Conclusion

This lab provided hands-on exposure to Google Cloud's Dataplex Universal Catalog. By creating a lake, raw zone, and attaching a storage bucket as an asset, users simulated a basic data mesh architecture. The lab demonstrated how data discovery and governance

can be implemented efficiently without moving data, a key advantage for scalable enterprise data solutions.