SP1143 — Dataplex Universal Catalog

Overview

Dataplex Universal Catalog is an intelligent data fabric within Google Cloud that enables organizations to **centrally discover, manage, monitor, and govern data** across diverse data sources such as **data lakes, warehouses, and marts**. It provides a unified metadata management solution without requiring data duplication or relocation.

By using Dataplex Universal Catalog, organizations can implement a **data mesh architecture**, decentralizing data ownership among domain-specific teams while maintaining centralized governance and discoverability.

In this lab, the objective was to build foundational components of a data mesh by creating and managing **lakes**, **zones**, and **assets** within Dataplex Universal Catalog using the Google Cloud Console.

Learning Objectives

By the end of this lab, the following skills were demonstrated:

1. Enabling the **Dataplex API**.
2. Creating a **data lake** within the Universal Catalog.
3. Adding a **zone** (subdomain) to a lake.
4. Attaching a **Cloud Storage asset** to a zone.
5. Deleting assets, zones, and lakes to manage the full lifecycle of Dataplex resources.

Setup and Requirements

- **Platform:** Google Cloud Skills Boost
- **Service Used:** Dataplex Universal Catalog
- **Browser:** Google Chrome (Incognito Mode recommended)
- **Access Mode:** Temporary student account credentials
- **Estimated Duration:** Timed lab (non-pausable)

> Note: The lab was performed in a live Google Cloud environment, not a simulation. All operations were executed through the Google Cloud Console.

Tasks and Implementation

Task 1: Enable the Dataplex API

Objective:

Activate the Dataplex service to allow creation and management of lakes, zones, and assets.

Procedure:

1. In the Google Cloud Console, searched for **"Cloud Dataplex API"** in the top navigation bar.
2. Selected **Cloud Dataplex API** from the Marketplace results.
3. Clicked **Enable** to activate the API for the project.

Outcome:

The Dataplex API was successfully enabled, allowing access to Dataplex Universal Catalog features.

Task 2: Create a Lake

Objective:

Create a lake representing a high-level data domain.

Procedure:

1. Opened **Navigation menu → View all products → Analytics → Dataplex Universal Catalog**.
2. Selected **Manage lakes → Manage → +Create lake**.
3. Entered the following details:
   - **Display Name:** `sensors`

- **Region:** (selected an appropriate regional location)
- Other settings left at defaults.

4. Clicked **Create** and waited for the lake to become active.

Outcome:

A lake named **sensors** was successfully created, representing the top-level organizational domain for managing related data assets.

✅ *Objective verified: "Create a data lake sensors"*

Task 3: Add a Zone to the Lake

Objective:

Add a **Raw Zone** within the `sensors` lake to categorize unprocessed data.

Procedure:

1. On the **Manage** tab, selected the `sensors` lake.
2. Clicked **+Add zone**.
3. Entered the following information:
   - **Display Name:** `temperature raw data`
   - **Type:** Raw zone
   - **Data Location:** Regional
   - Discovery settings: Left defaults (metadata discovery enabled).
4. Clicked **Create**.

Outcome:

A **raw zone** named `temperature raw data` was successfully added to the lake `sensors`. This zone serves as the storage location for unprocessed or raw sensor data.

✅ *Objective verified: "Add a zone temperature raw data to the lake sensors"*

Task 4: Attach an Asset to a Zone

Objective:

Attach a Cloud Storage bucket as an asset to the `temperature raw data` zone.

Procedure:

1. Navigated to the **Zones** tab → selected the `temperature raw data` zone.
2. Opened the **Assets** tab → clicked **+Add Asset → +Add an Asset**.
3. Entered the following details:
   - **Type:** Storage bucket
   - **Display Name:** `measurements`
4. Created a new Cloud Storage bucket:
   - Used **Project ID** as the bucket name.
   - Set **Location type:** Regional.
   - Accepted default settings and created the bucket.
5. Attached the newly created bucket as an asset.
6. Chose **Inherit** for Discovery settings and finalized by clicking **Submit**.

Outcome:

The **Cloud Storage bucket** was successfully attached as an asset named `measurements` within the `temperature raw data` zone.

✅ *Objective verified: "Create a cloud storage bucket and attach an asset measurements to the zone temperature raw data"*

Task 5: Delete Assets, Zones, and Lakes

Objective:

Demonstrate cleanup and lifecycle management of Dataplex resources.

Procedure:

1. **Detach the asset:**

- Navigated to the zone and selected the asset checkbox.
- Chose **Delete assets** → Confirmed deletion.
- The action detached the bucket without deleting its contents.

2. **Delete the zone:**
   - Returned to the lake view → Selected the zone → Clicked **Delete zone** → Confirmed.
3. **Delete the lake:**
   - Selected the `sensors` lake → Clicked **Delete** → Typed `delete` to confirm.

Outcome:

All Dataplex resources (assets, zones, and lake) were deleted, completing the data lifecycle management workflow.

✅ *Objective verified: "Detach an asset, delete zone and lake"*

Key Learnings

- Dataplex Universal Catalog allows **metadata-driven data management** without physically moving data.
- **Lakes** serve as high-level organizational domains, while **zones** categorize data based on type or processing stage.
- **Assets** (e.g., Cloud Storage buckets or BigQuery datasets) are the actual data locations linked to zones.
- The **data mesh architecture** enables decentralized ownership while maintaining central visibility.
- Dataplex integrates seamlessly with services such as **BigQuery** for metadata discovery and analytics.

Conclusion

This lab provided hands-on experience with **Google Cloud Dataplex Universal Catalog**, demonstrating how to organize, manage, and govern data assets within a data mesh framework.

Through creating and managing **lakes, zones, and assets**, I gained a deeper understanding of how Dataplex facilitates **metadata-driven governance**, **centralized cataloging**, and **scalable data management** across multiple environments.

This foundational knowledge supports the implementation of **data mesh architectures** for large-scale, decentralized data ecosystems.