Lab Summary: Dataplex and Google Workspace Tools

Overview:

Today, I explored **Google Cloud Dataplex**, an intelligent data fabric that allows centralized **data discovery, management, monitoring, and governance** across various data sources such as data lakes, warehouses, and marts. I learned how Dataplex supports **data mesh architecture**, enabling decentralized data ownership without requiring data movement or duplication.

What I Did

1. Enabled the Dataplex API

- Activated the **Dataplex API** in Google Cloud.
- Set up **environment variables** for `PROJECT_ID` and `REGION` in Cloud Shell.

2. Created and Managed Lakes and Zones

- Created a **lake** named `ecommerce` representing a business domain.
- Added a **curated zone** (`orders-curated-zone`) for analytics-ready data.

3. Attached BigQuery Assets

- Created a **BigQuery dataset** (`orders`).
- Attached it as an **asset** within the curated zone using the Dataplex CLI.
- Enabled **metadata discovery** for automated integration.

4. Deleted Assets, Zones, and Lakes

- Practiced **detaching assets**, deleting **zones**, and finally removing the **lake** to understand full lifecycle management in Dataplex.

Key Learnings

- Understood the **hierarchical structure** of Dataplex: *Lake → Zone → Asset*.
- Learned how **metadata discovery** simplifies governance and data cataloging.
- Practiced using **gcloud commands** for Dataplex operations directly from Cloud Shell.
- Reinforced the importance of using **Dataplex Universal Catalog** for centralized metadata management.

Additional Tools Explored

Along with Dataplex, I also familiarized myself with essential **Google Workspace tools** including:

- **Google Drive** – for file storage and sharing
- **Google Calendar** – for scheduling and time management
- **Gmail** – for communication

- **Google Meet** – for virtual meetings
- **Google Sheets** – for collaborative data handling