**Uber Data Analytics | Modern Data Engineering GCP Project**

**A diagram of a company

Description automatically generated**

This diagram illustrates a data processing workflow using Google Cloud services, Mage, and Looker. Here's a breakdown of each component and its role:

1. **Raw Data (Cloud Storage)**:
   * This represents where the raw data is stored, such as unprocessed files, logs, or datasets.
   * It is hosted on Google Cloud Storage, which is a scalable, secure, and durable storage solution.
2. **ETL Process (Mage on Compute Engine)**:
   * The ETL (Extract, Transform, Load) process is managed by **Mage**, an open-source tool for building data pipelines.
   * Mage runs on a **Compute Engine Virtual Machine (VM)**, a Google Cloud service that provides scalable and customizable virtual machines.
   * **ETL steps**:
     + **Extract**: Mage retrieves raw data from Cloud Storage.
     + **Transform**: It processes and cleanses the data, making it ready for analysis.
     + **Load**: The transformed data is loaded into BigQuery for analytics.
3. **Analytics (BigQuery)**:
   * **BigQuery** is a Google Cloud data warehouse that allows users to run fast, SQL-like queries on large datasets.
   * It is used here to store and analyze the processed data produced by the ETL process.
4. **Visualization and Reporting (Looker)**:
   * **Looker** is a business intelligence and data visualization tool that connects to BigQuery.
   * Users can create dashboards, reports, and visualizations to derive insights and make data-driven decisions.

**Workflow Summary**:

* Raw data in Google Cloud Storage → Processed by Mage ETL on Compute Engine → Loaded into BigQuery for analytics → Visualized in Looker for business intelligence.