Google Sheets Data to GCS & BigQuery - Automated
- A Case Study

Problem Statement

Currently, the client stores the data in Google Sheets and creates reports using Looker. Because of this, the business is not able to save historical data due to continuous changes in Google Sheets.

To overcome this situation the client wants to analyze the data by performing various transformations in BigQuery and save the data in GCS Data lakes. This complete process needs to be automated using the Orchestration tool.

Methodology

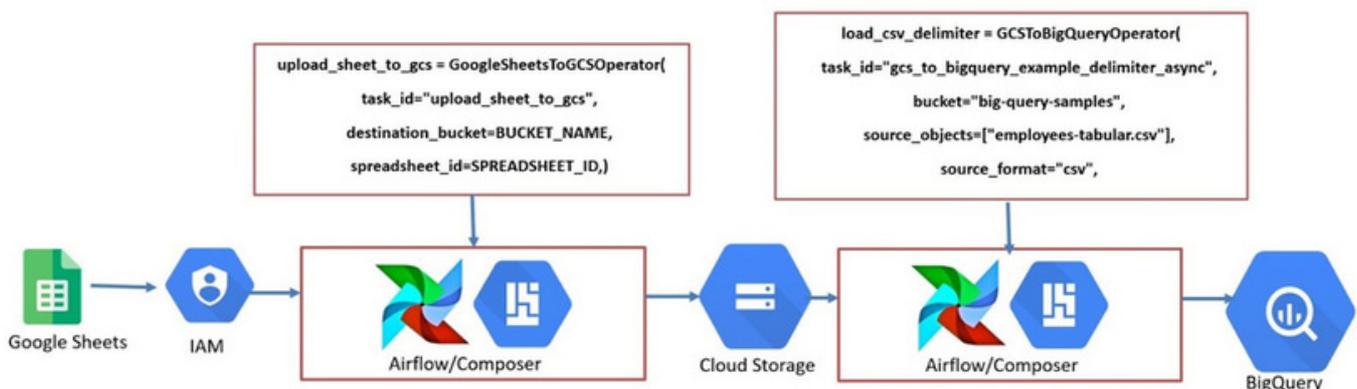The data is extracted from Google Sheets and ingested into BigQuery using Airflow Composer

- Creating Data lakes in GCS to store the Google Sheets Data
- Querying the Google Sheets Data in BigQuery for Analysis
- Automation of Data pipeline process from Google Sheets to BigQuery

⚙ Technology

1. Google Cloud Platform Components / Tools
   a. Google Cloud Storage
   b. Cloud Composer
   c. BigQuery
2. Google Sheets

Project Architecture

Google Sheets Data to GCS & BigQuery - Automated
- A Case Study

Our Approach

- Created a Service Account with a Viewer role under the project.
- Assigned the Service account to Google Sheets by using the share option as to provide access.
- The data extraction process from Google Sheets to Google Cloud Storage is done by using Airflow Composer.
- Used Google Transfer Operator which is GoogleSheetsToGCSOperator in the Airflow DAG.
- Created Dataset, and Table in the BigQuery using Airflow operators and loaded the data to BQ using Bash operator.
- Specify the Source Sheet ID, and Bucket Name in the Airflow Dag.

  Ex: upload_sheet_to_gcs = GoogleSheetsToGCSOperator(
  task_id="upload_sheet_to_gcs",
  destination_bucket=BUCKET_NAME,
  spreadsheet_id=SPREADSHEET_ID,)

- Check the BigQuery dataset for loaded data