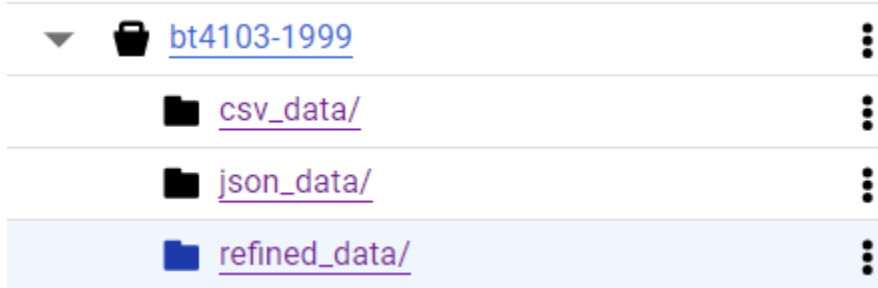


Pipeline Instructions

Google Cloud Storage

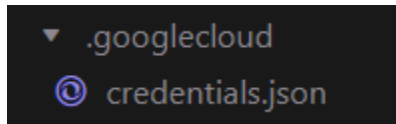
1. Create a new Google Bucket and name it as you like.
2. Within that Google Bucket, create 3 folders called "csv_data", "json_data", "refined_data"



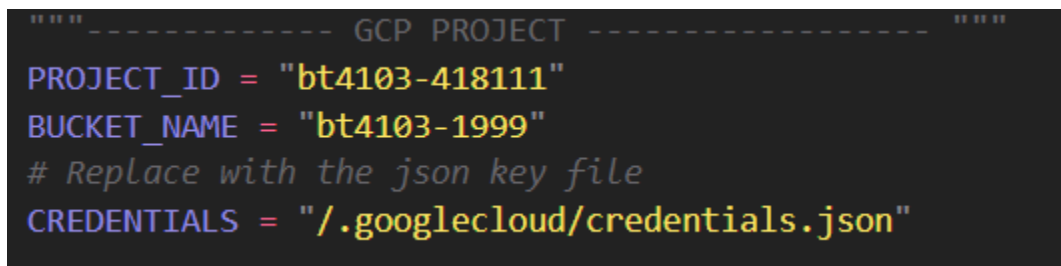
3. Upload your json files (GA) data into the folder 'json_data'
4. Upload your SQL data into 'csv_data'
5. Go to IAM & Admin > Service Accounts to get your Google credentials file. Download it and rename as 'credentials'.

Docker Airflow

1. Drop your credentials.json file into the /.googlecloud directory.



2. Go to the constants.py file and change the GCP project details (PROJECT_ID and BUCKET_NAME).



3. Open up a terminal and run the following commands.
 - cd into airflow_proj: `cd airflow_proj`
 - Create customised docker image: `docker build . -f Dockerfile --pull --tag bt4301-airflow:0.0.1`
 - Run airflow-init container: `docker compose up airflow-init`
 - Run all containers: `docker compose up`
4. Open a chrome browser and type '<http://localhost:8080/>'
5. Airflow login details are: username: airflow, password: airflow
6. Trigger the sicpama_dag