Shivani Sinha, IUBH

Development and Reflection Phase

During this phase, the core batch-processing system planned in Phase 1 was fully implemented using modular, serverless components in the Google Cloud Platform. My goal was to build a cost-effective, reliable, and scalable solution within GCP's limitations. The system ingests time-stamped ride data from Cloud Storage, processes it using BigQuery SQL transformations, and delivers analytical insights via a Looker Studio dashboard.

To simulate real-world data, I generated a synthetic dataset with 1,000 rides, including vehicle types and distances. The reference of this data was taken from taxi rides data from kaggle but as we needed just a few columns for creating the final result, I implemented an inspired data file instead of using the actual data . The ingestion microservice was built using **Cloud Scheduler**, **Pub/Sub**, **Cloud Workflows** and **BigQuery Tables** replacing the originally intended Cloud Function due to UI limitations. The transformation logic to compute `co2_grams` was implemented as a reusable **BigQuery view**, ensuring easy updates and clean separation of concerns.

Looker Studio was used to visualize total emissions, emissions by city and vehicle type, and monthly emission trends. Sorting issues in the dashboard were addressed by creating a helper field (`MonthSort`) and formatting months using `FORMAT_DATE`, ensuring correct chronological order. The entire pipeline was tested repeatedly using new data uploads and manual triggers.

While Docker was not used in practice (due to the GCP-native design), each service was treated as an independent microservice. The working solution aligns closely with the goals of reliability, maintainability, and reproducibility.

Architecture Diagram

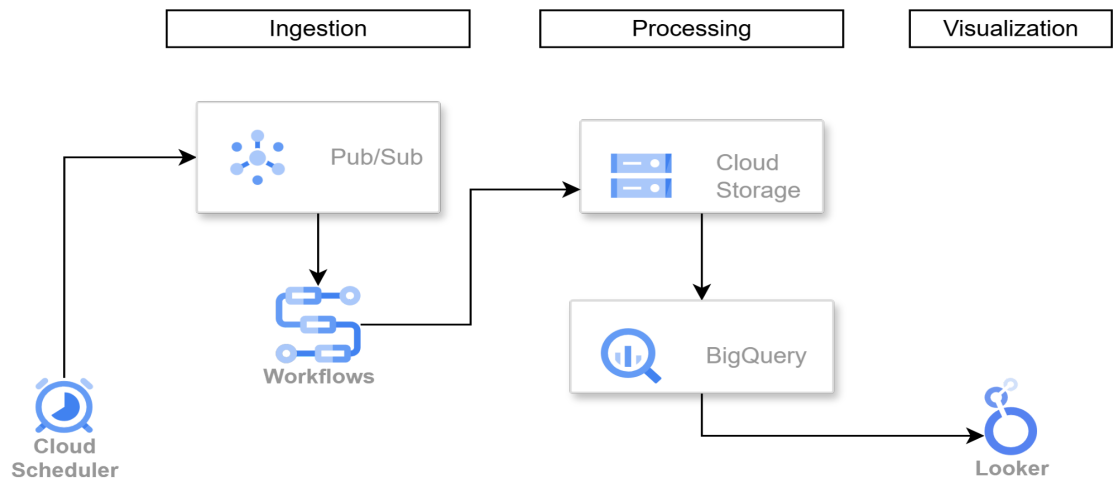
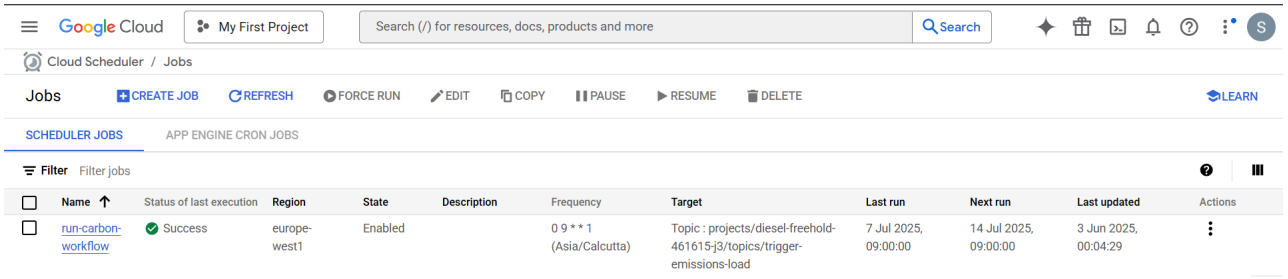
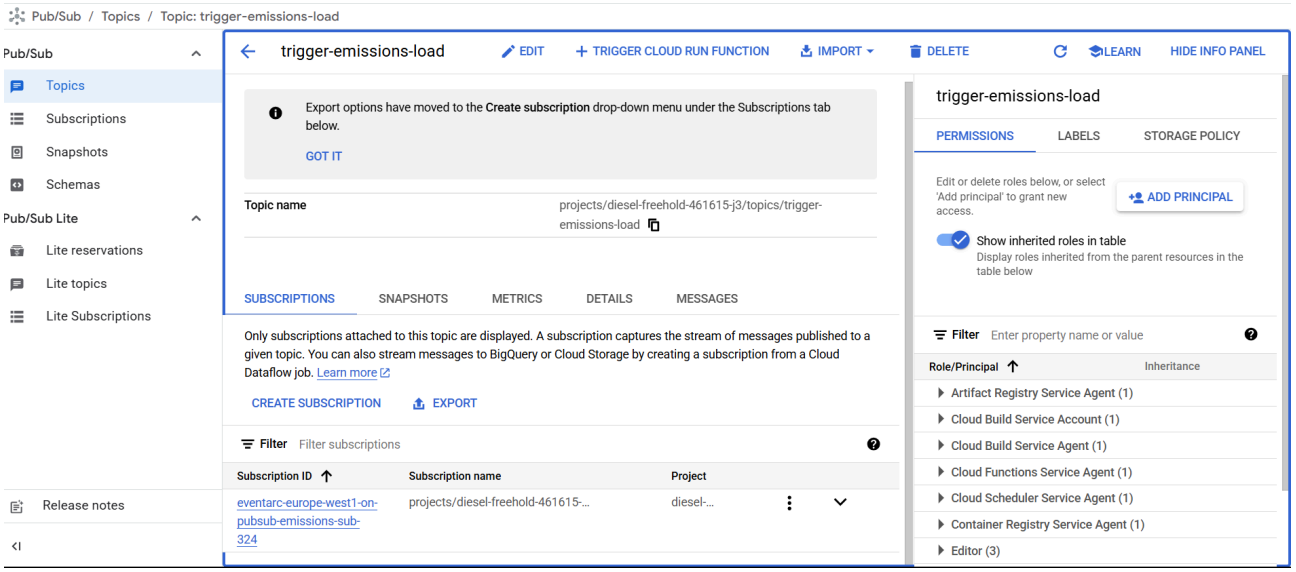


Image created through draw.io software

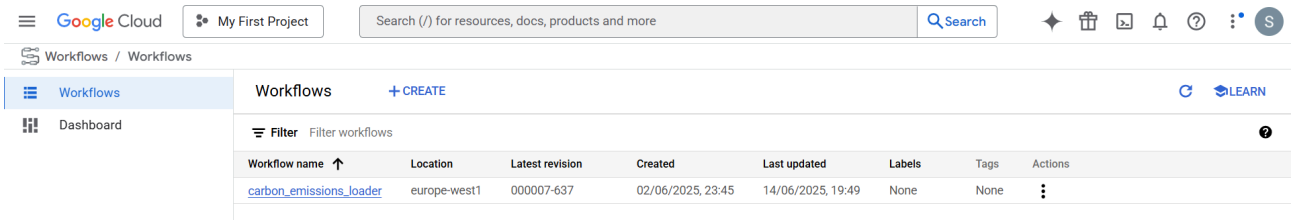
Cloud Scheduler



Pub/Sub



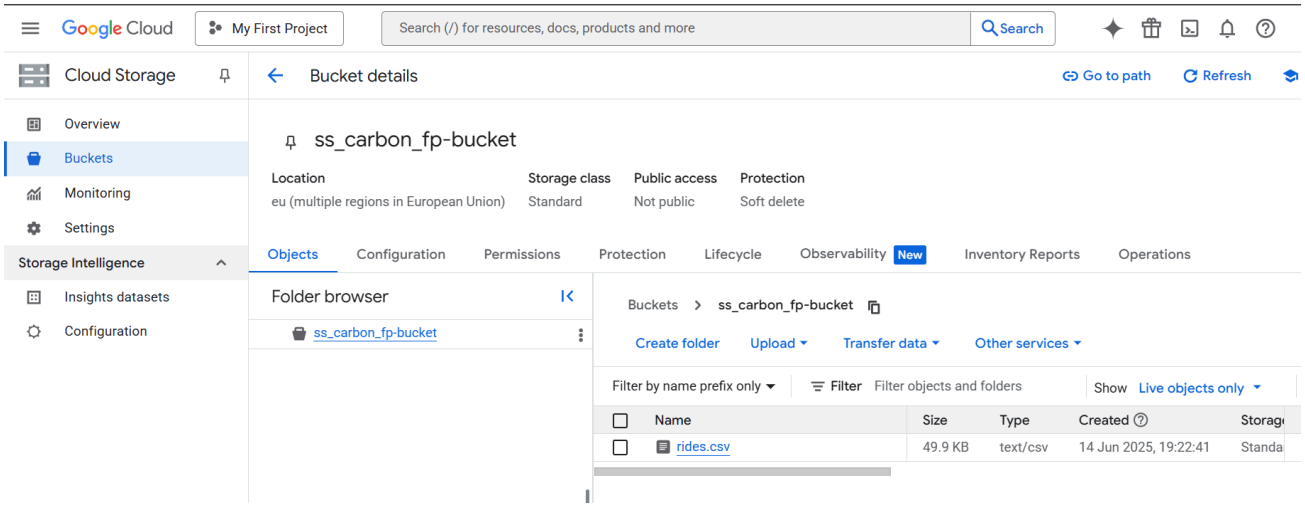
Workflow



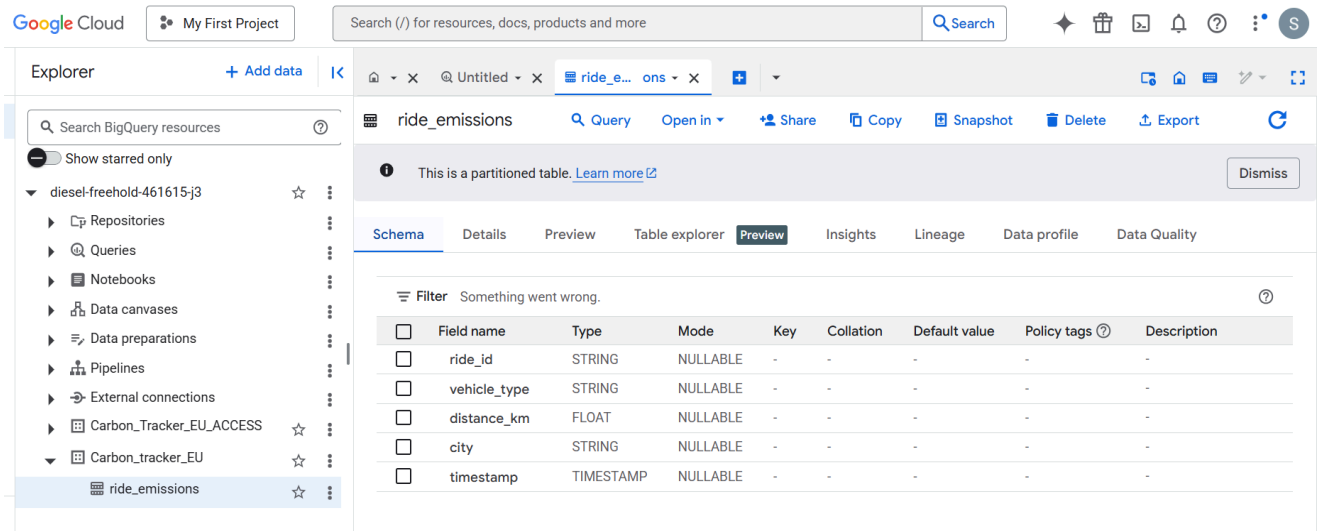

```
main:
  params: [event]
  steps:
    - init:
        assign:
          - bucket: "ss_carbon_fp-bucket"
          - file: "rides.csv"
          - dataset: "Carbon_tracker_EU"
          - table: "ride_emissions"
          - source_uri: "gs://ss_carbon_fp-bucket/rides.csv"
    - load_to_bigquery:
        call: googleapis.bigquery.v2.jobs.insert
        args:
          projectId: diesel-freehold-461615-j3
```

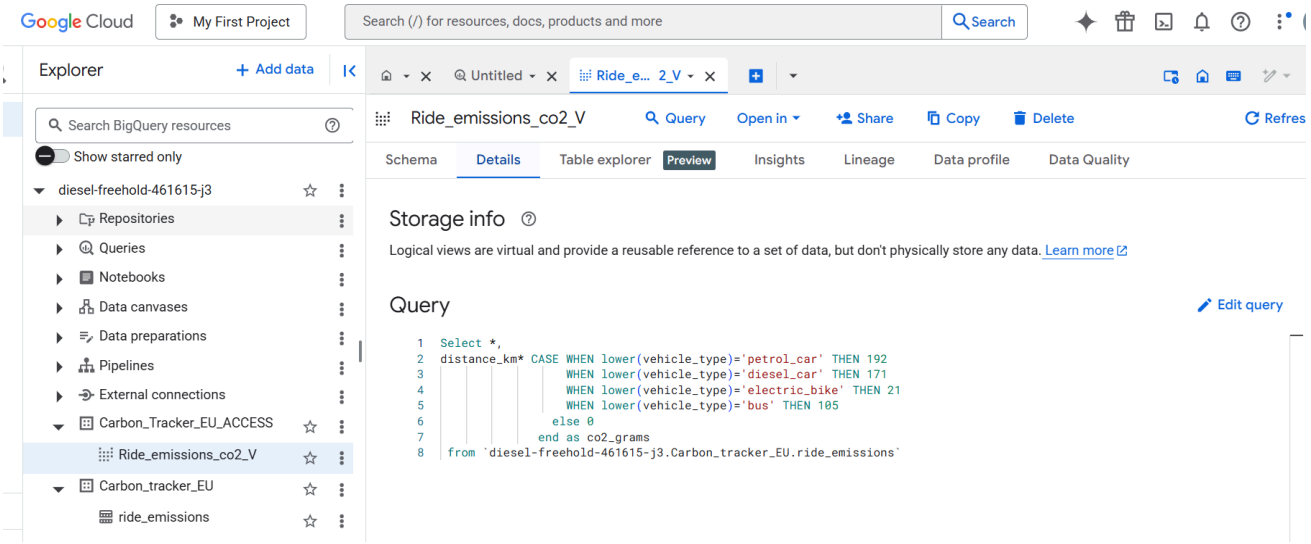
```
      body:
        configuration:
          load:
            destinationTable:
              projectId: diesel-freehold-461615-j3
              datasetId: ${dataset}
              tableId: ${table}
            sourceUris:
              - ${source_uri}
            skipLeadingRows: 1
            autodetect: true
            writeDisposition: WRITE_APPEND
      result: loadJob
```

Cloud Storage/ Bucket



Big Query





Looker Studio