

# 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

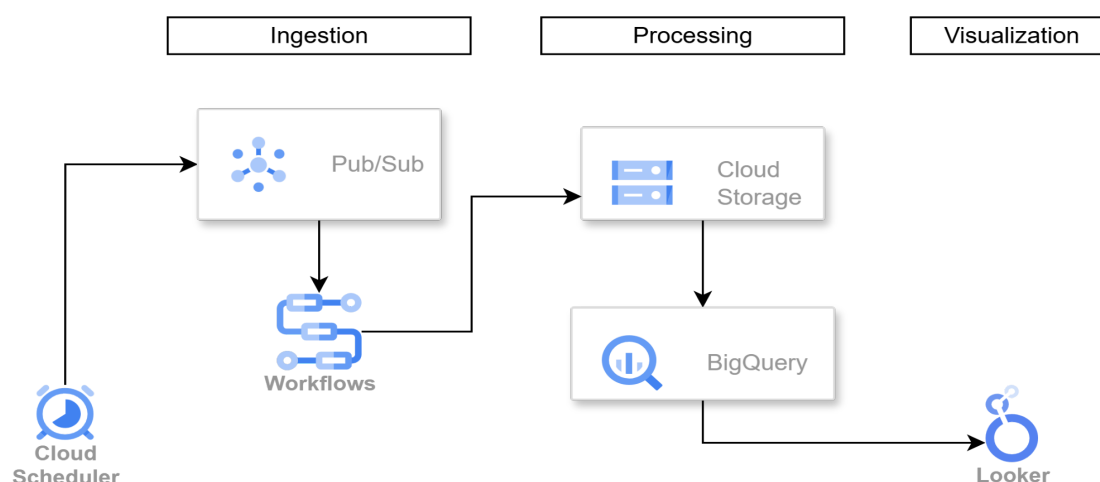


Image created through [draw.io](https://draw.io) software

# Cloud Scheduler

Google Cloud

My First Project

Search (/) for resources, docs, products and more

Search

Cloud Scheduler / Jobs

Jobs

CREATE JOB

REFRESH

FORCE RUN

EDIT

COPY

PAUSE

RESUME

DELETE

LEARN

SCHEDULER JOBS

APP ENGINE CRON JOBS

Filter

Filter jobs

Name

↑

Status of last execution

Region

State

Description

Frequency

Target

Last run

Next run

Last updated

Actions

run-carbon-workflow

Success

europe-west1

Enabled

0 9 \* \* 1  
(Asia/Calcutta)

Topic : projects/diesel-freehold-461615-j3/topics/trigger-emissions-load

7 Jul 2025,  
09:00:00

14 Jul 2025,  
09:00:00

3 Jun 2025,  
00:04:29

# Pub/Sub

Pub/Sub / Topics / Topic: trigger-emissions-load

Pub/Sub

Topics

Subscriptions

Snapshots

Schemas

Pub/Sub Lite

Lite reservations

Lite topics

Lite Subscriptions

Release notes

trigger-emissions-load

EDIT

+ TRIGGER CLOUD RUN FUNCTION

IMPORT

DELETE

LEARN

HIDE INFO PANEL

Export options have moved to the Create subscription drop-down menu under the Subscriptions tab below.

GOT IT

Topic name

projects/diesel-freehold-461615-j3/topics/trigger-emissions-load

SUBSCRIPTIONS

SNAPSHOTS

METRICS

DETAILS

MESSAGES

Only subscriptions attached to this topic are displayed. A subscription captures the stream of messages published to a given topic. You can also stream messages to BigQuery or Cloud Storage by creating a subscription from a Cloud Dataflow job. [Learn more](#)

CREATE SUBSCRIPTION

EXPORT

Filter

Filter subscriptions

Subscription ID

Subscription name

Project

eventarc-europe-west1-on-pubsub-emissions-sub-324

projects/diesel-freehold-461615-...

diesel-...

trigger-emissions-load

PERMISSIONS

LABELS

STORAGE POLICY

Edit or delete roles below, or select 'Add principal' to grant new access.

ADD PRINCIPAL

Show inherited roles in table

Display roles inherited from the parent resources in the table below

Filter

Enter property name or value

Role/Principal

↑

Inheritance

Artifact Registry Service Agent (1)

Cloud Build Service Account (1)

Cloud Build Service Agent (1)

Cloud Functions Service Agent (1)

Cloud Scheduler Service Agent (1)

Container Registry Service Agent (1)

Editor (3)

# Workflow

Google Cloud

My First Project

Search (/) for resources, docs, products and more

Search

Workflows / Workflows

Workflows

+ CREATE

LEARN

Filter

Filter workflows

Workflow name

Location

Latest revision

Created

Last updated

Labels

Tags

Actions

carbon\_emissions\_loader

europe-west1

000007-637

02/06/2025, 23:45

14/06/2025, 19:49

None

None

```
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

The screenshot shows the Google Cloud Storage interface for a bucket named `ss_carbon_fp-bucket`. The bucket is located in the `eu` region, uses the `Standard` storage class, and has `Not public` access with `Soft delete` protection. The `Objects` tab is active, showing a folder browser view with the bucket's contents. A table lists the objects in the bucket:

Name	Size	Type	Created	Storage
<code>rides.csv</code>	49.9 KB	text/csv	14 Jun 2025, 19:22:41	Standard

## Big Query

The screenshot shows the Google Cloud BigQuery interface. The `ride_emissions` table is selected, and the `Schema` tab is active. A message indicates that the table is partitioned. The schema table is displayed below:

Field name	Type	Mode	Key	Collation	Default value	Policy tags	Description
<code>ride_id</code>	STRING	NULLABLE	-	-	-	-	-
<code>vehicle_type</code>	STRING	NULLABLE	-	-	-	-	-
<code>distance_km</code>	FLOAT	NULLABLE	-	-	-	-	-
<code>city</code>	STRING	NULLABLE	-	-	-	-	-
<code>timestamp</code>	TIMESTAMP	NULLABLE	-	-	-	-	-

Google Cloud My First Project Search (/) for resources, docs, products and more Search

Explorer + Add data

Search BigQuery resources

Show starred only

diesel-freehold-461615-j3

Repositories

Queries

Notebooks

Data canvases

Data preparations

Pipelines

External connections

Carbon\_Tracker\_EU\_ACCESS

Ride\_emissions\_co2\_V

Carbon\_tracker\_EU

ride\_emissions

Ride\_emissions\_co2\_V Query Open in Share Copy Delete Refres

Schema Details Table explorer Preview Insights Lineage Data profile Data Quality

Storage info

Logical views are virtual and provide a reusable reference to a set of data, but don't physically store any data. [Learn more](#)

Query

```

1 Select *,
2 distance_km* CASE WHEN lower(vehicle_type)='petrol_car' THEN 192
3                 WHEN lower(vehicle_type)='diesel_car' THEN 171
4                 WHEN lower(vehicle_type)='electric_bike' THEN 21
5                 WHEN lower(vehicle_type)='bus' THEN 105
6                 else 0
7                 end as co2_grams
8 from `diesel-freehold-461615-j3.Carbon_tracker_EU.ride_emissions`

```

Edit query

## Looker Studio

