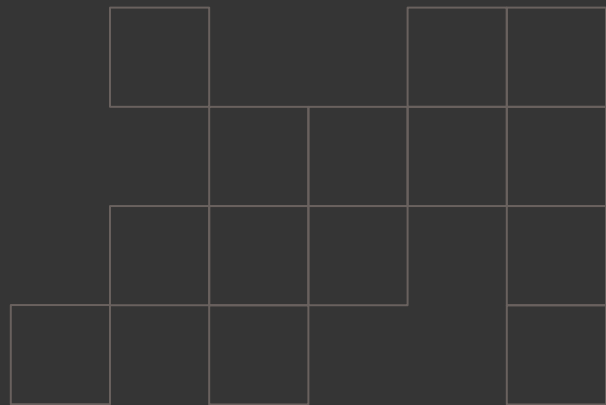


neuefische

# Weekly Update: Ingestion

Chirayu Upadhyay



# Contents

1.

Ingestion diagram

2.

Screenshots for ingestion layer

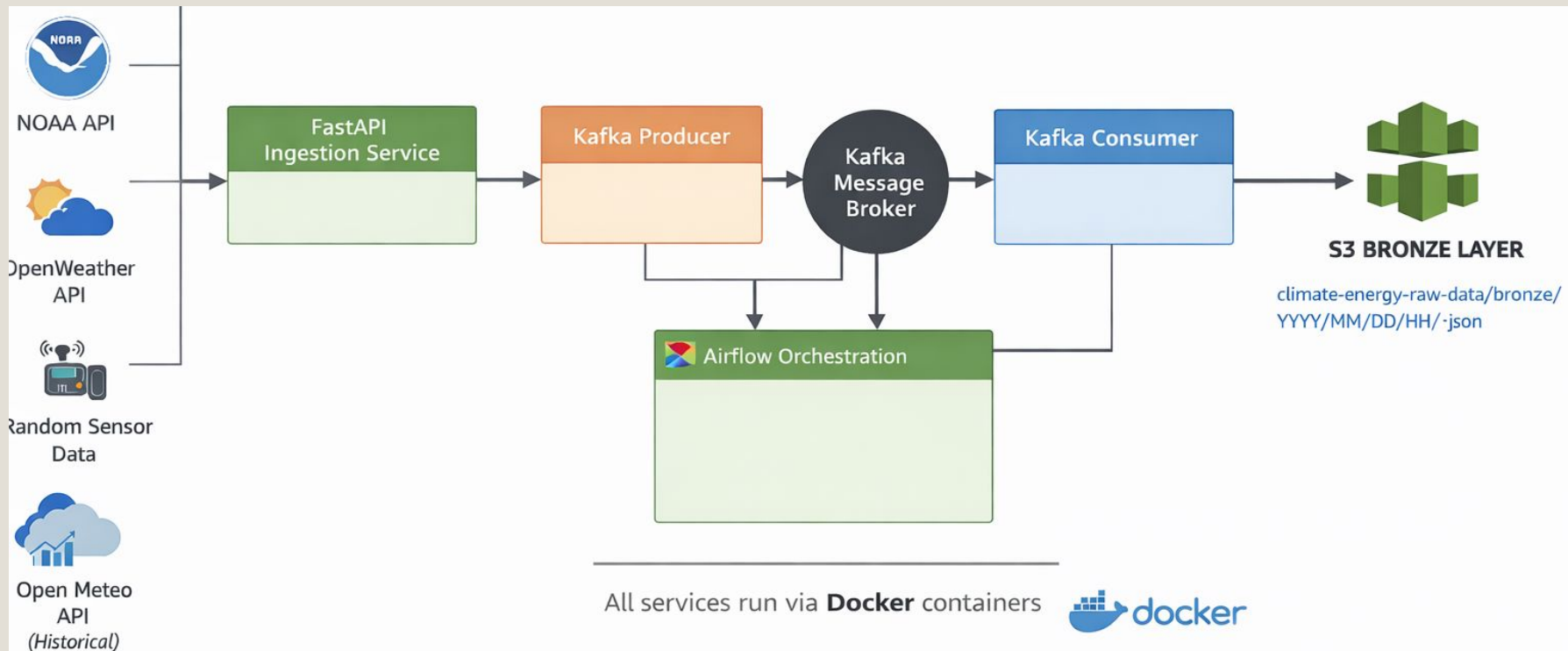
3.

API data snippets in json

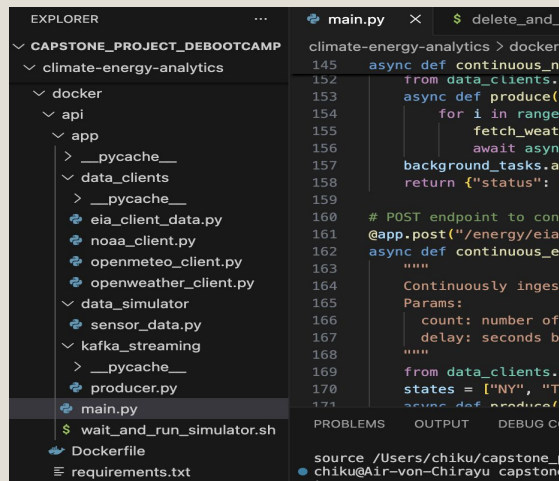
4.

Plan for next sprint : Transformation

# 1. Ingestion layer Diagram



## 2. Screenshots of Ingestion layer: Code



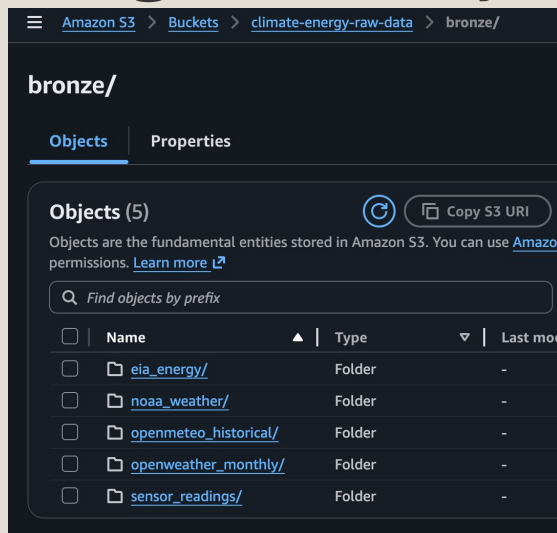
The screenshot shows a code editor with a file explorer on the left. The file explorer lists the following structure:

- EXPLOSER
  - ...
    - CAPSTONE\_PROJECT\_DEBOOTCAMP
      - climate-energy-analytics
        - docker
          - api
            - app
              - \_\_pycache\_\_
              - data\_clients
              - \_\_pycache\_\_
              - eia\_client\_data.py
              - noaa\_client.py
              - openmeteo\_client.py
              - openweather\_client.py
            - data\_simulator
              - sensor\_data.py
            - kafka\_streaming
              - \_\_pycache\_\_
              - producer.py
            - main.py
            - wait\_and\_run\_simulator.sh
            - Dockerfile
            - requirements.txt

The main.py file is open, showing Python code for a FastAPI application. It includes endpoints for energy data, weather data, and a Kafka streaming endpoint. The code uses async/await for asynchronous processing and includes a Dockerfile and requirements.txt file.

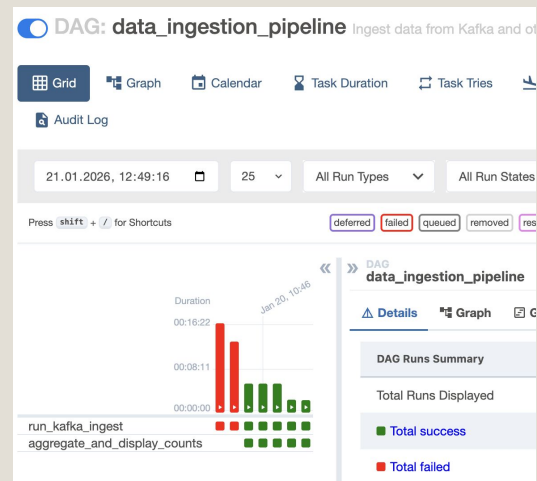
Project directory:

Shows the each client\_api files with Dockerfile



AWS S3 Bucket directory

Shows all the kafka topic folders are created inside the /bronze layer and data saved with the date, month, day, time partitioned folders



Airflow Dag

Shows the airflow dag for the ingestion and it depicts the failure and success runs

### 3. Api Data Snippets

```
"{
  \"source\": \"eia\",
  \"topic\": \"eia_energy\",
  \"timestamp\": \"2026-01-15T16:14:36.149\",
  \"data\": {
    \"series_id\": \"EIA-TEST\",
    \"period\": \"2026-01\",
    \"state\": \"NY\",
    \"sector\": \"residential\",
    \"price\": 0.12
  }
}
```

eia\_energy api data

```
=
{
  \"source\": \"noaa\",
  \"topic\": \"noaa_weather\",
  \"timestamp\": \"2026-01-18T14:17:25.18\",
  \"data\": {
    \"date\": \"2024-12-30\",
    \"datatype\": \"TMIN\",
    \"station_id\": \"GHCND:USW00094728\",
    \"value\": 75.0
  }
}
```

NOAA weather api data

```
=
{
  \"source\": \"sensor\",
  \"topic\": \"sensor_readings\",
  \"timestamp\": \"2026-01-20T12:41:24.044601\",
  \"data\": {
    \"meter_id\": \"SENS_4555\",
    \"power_kw\": 7.01,
    \"voltage\": 230,
    \"location\": \"TX\",
    \"timestamp\": 1768912884,
    \"sensor_id\": \"SENSOR_051\",
    \"temperature_c\": 20.47,
    \"humidity_pct\": 68.84
  }
}
```

Sensor Readings data

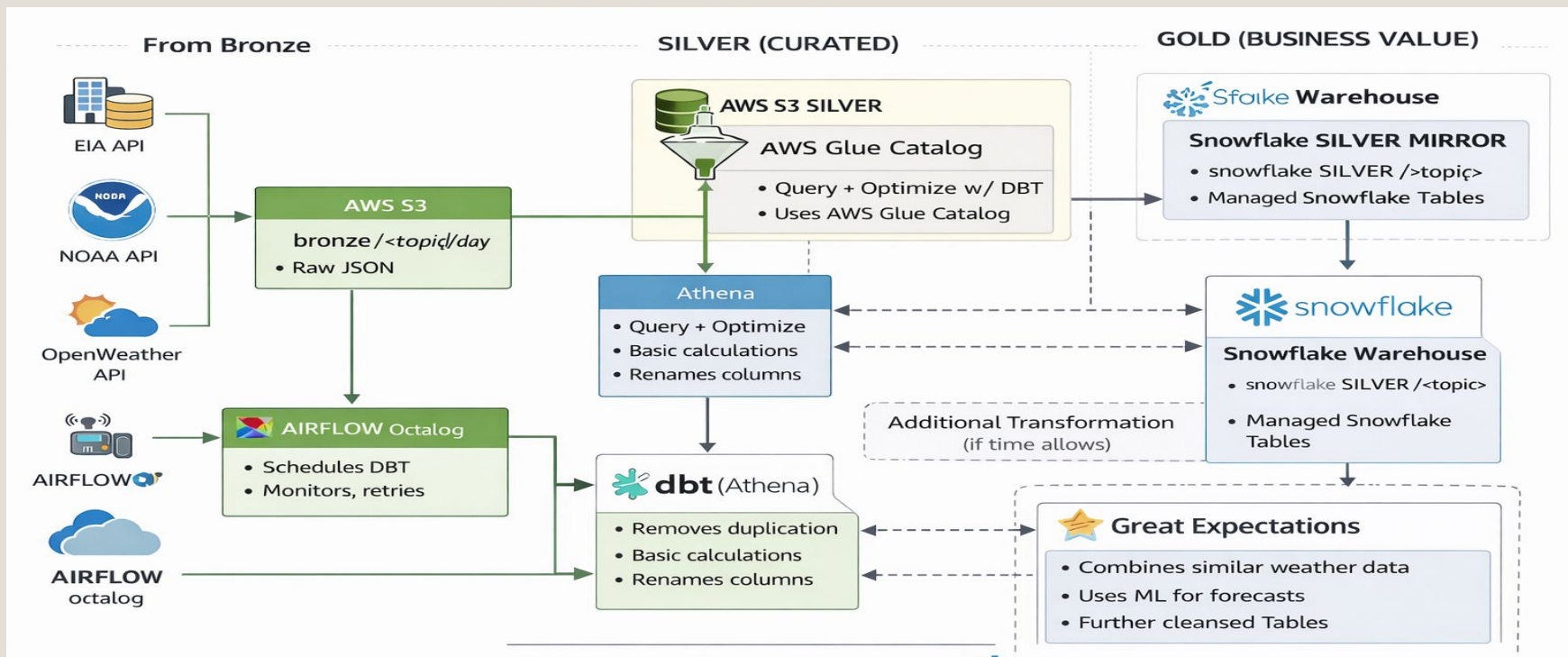
```
"
{
  \"source\": \"openweather\",
  \"topic\": \"openweather_monthly\",
  \"timestamp\": \"2026-01-17T14:28:10.304989\",
  \"data\": {
    \"city\": \"new york\",
    \"dt\": 1768499999,
    \"temperature_c\": 22.5,
    \"feels_like_c\": 21.0,
    \"humidity_pct\": 60,
    \"wind_speed_mps\": 3.2,
    \"visibility_m\": 10000,
    \"weather_main\": \"Clear\"
  }
}
```

OpenWeather Api data

```
=
{
  \"source\": \"openmeteo\",
  \"topic\": \"openmeteo_historical\",
  \"timestamp\": \"2026-01-17T16:59:06.98300\",
  \"data\": {
    \"dt\": \"2025-08-25\",
    \"temperature_c\": 34.1,
    \"temperature_min_c\": 19.9,
    \"precipitation_mm\": 0.0,
    \"windspeed_max_mps\": 12.9,
    \"city\": \"los angeles\"
  }
}
```

Open Meteo weather api data

## 4. Plan for next sprint: Transformation



Note: the dotted line connection will be optional (except GE that will i perform) but the ML thing in gold layer would be optional to perform



Thank you