Case_Study_1: IOT Data Processing

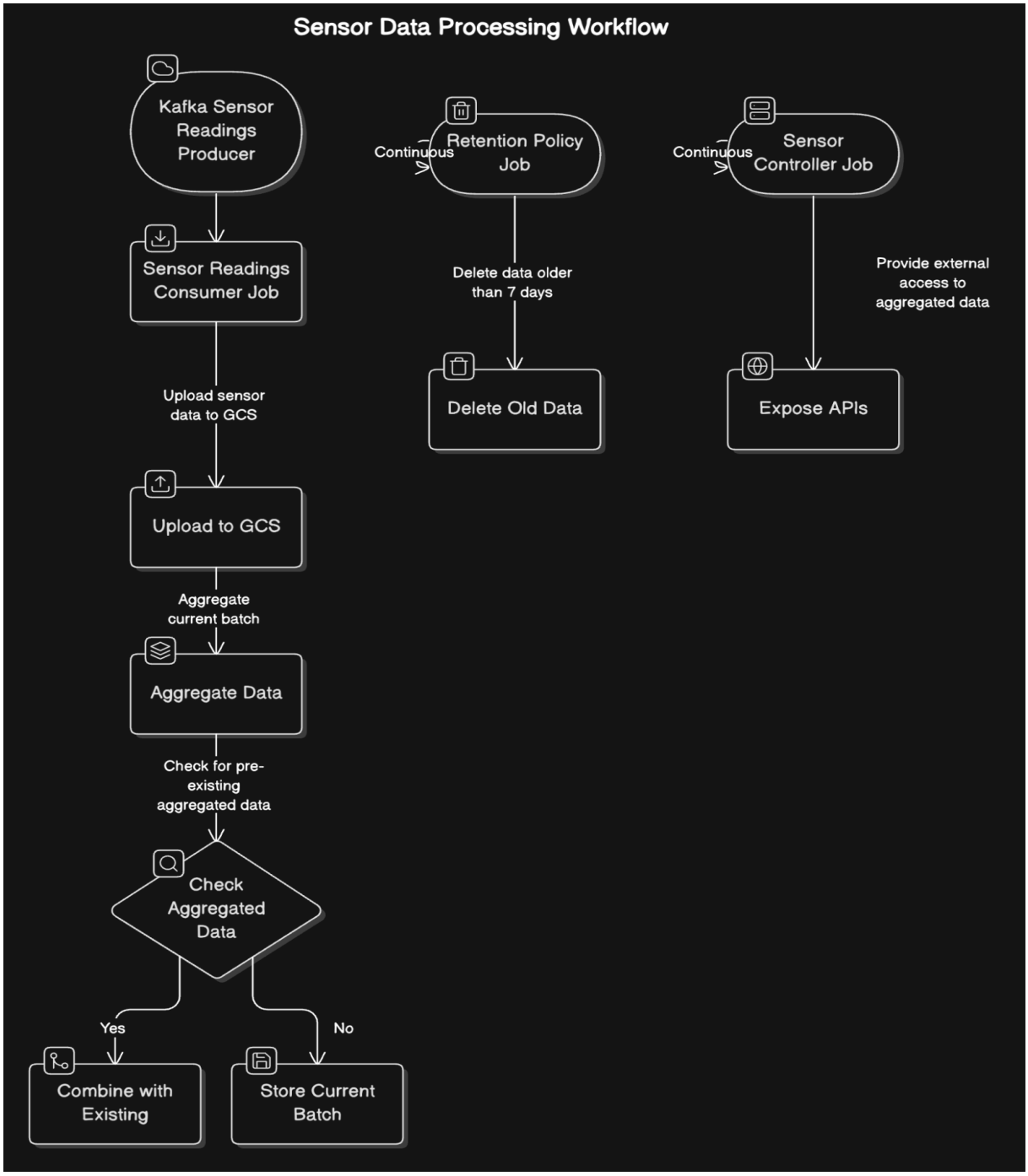
Bhargav Jangam

Detailed Code Workflow

The workflow consists of the following steps:

1. **Kafka Sensor Readings Producer**: This component produces sensor readings as JSON messages.
2. **Sensor Readings Consumer Job**: This job handles the following tasks:
    - Uploading the sensor data into a Google Cloud Storage (GCS) bucket under the path `raw/sensor-data` in Protobuf format, partitioned by the timestamp.
    - Aggregating the current batch of data.
    - Checking for any pre-existing aggregated data in the `aggregated/protobuf` folder. If previous aggregated data exists, it combines it with the current batch's aggregated data. If no previous data exists, it stores the current batch's aggregated data.
3. **Retention Policy Job**: This job runs continuously to delete folders in the `raw/sensor-data` path that are older than 7 days, ensuring data retention compliance.
4. **Sensor Controller Job**: This job runs continuously and exposes APIs related to the aggregated data, allowing external access to the processed information.

Flow Chart



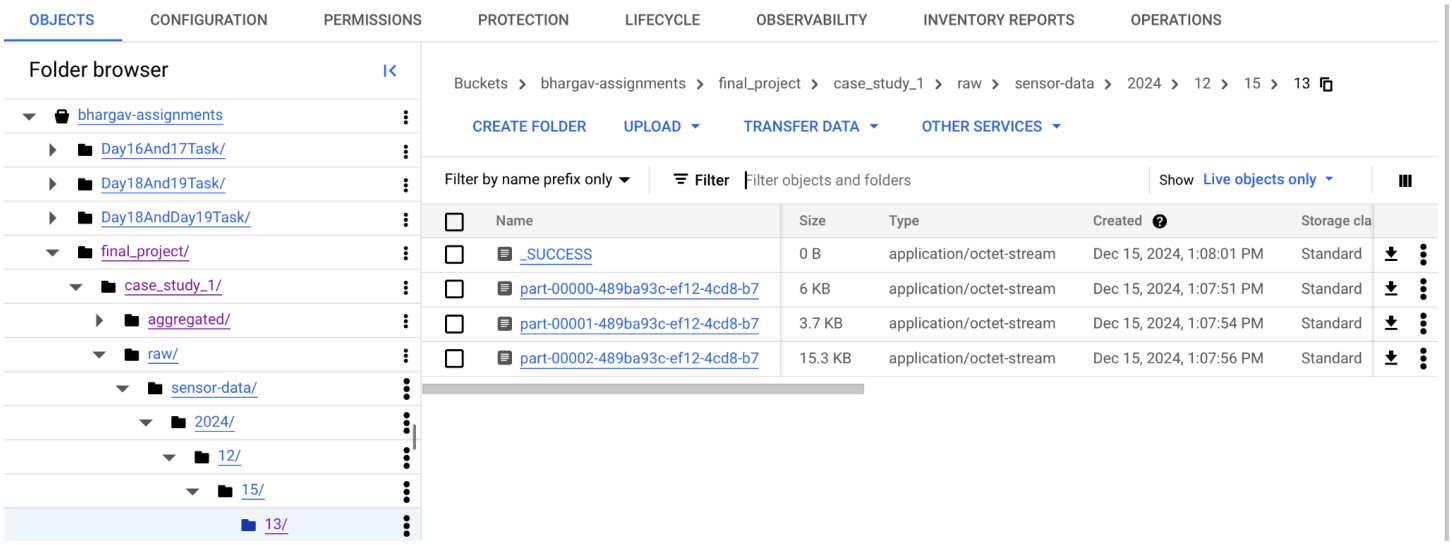
Below are the proof of execution and the resulting images showcasing the outcomes of the process.

Execution Proof and Resulting Images

Kafka - Sensor Readings Produced messages

```
Created topic sensor-readings.
bhargavjangam@apple-ka-MacBook-Pro config % kafka-console-consumer.sh --bootstrap-server localhost:9092 -topic sensor-readings --group consumer-group-2  --from-beginning
{"sensorId":77,"timestamp":1734247992109,"temperature":138.23384,"humidity":97.050644}
{"sensorId":64,"timestamp":1734247992228,"temperature":77.046844,"humidity":50.77436}
{"sensorId":15,"timestamp":1734247992349,"temperature":58.22834,"humidity":61.393143}
{"sensorId":10,"timestamp":1734247992467,"temperature":62.755905,"humidity":19.628292}
{"sensorId":12,"timestamp":1734247992589,"temperature":46.531174,"humidity":20.904022}
{"sensorId":99,"timestamp":1734247992709,"temperature":-12.388466,"humidity":68.10223}
{"sensorId":52,"timestamp":1734247992829,"temperature":28.304817,"humidity":62.89461}
{"sensorId":49,"timestamp":1734247992949,"temperature":44.60669,"humidity":73.36029}
{"sensorId":1,"timestamp":1734247993069,"temperature":101.1277,"humidity":68.22141}
{"sensorId":45,"timestamp":1734247993189,"temperature":77.69631,"humidity":90.87563}
{"sensorId":12,"timestamp":1734247993309,"temperature":37.35888,"humidity":26.069016}
{"sensorId":22,"timestamp":1734247993430,"temperature":-19.744896,"humidity":67.54745}
{"sensorId":86,"timestamp":1734247993549,"temperature":-20.581604,"humidity":31.041777}
{"sensorId":72,"timestamp":1734247993669,"temperature":-20.452166,"humidity":8.7509575}
{"sensorId":66,"timestamp":1734247993789,"temperature":73.27059,"humidity":44.32997}
{"sensorId":34,"timestamp":1734247993909,"temperature":-22.265076,"humidity":95.37437}
{"sensorId":60,"timestamp":1734247997388,"temperature":-2.0162125,"humidity":41.20615}
{"sensorId":10,"timestamp":1734247997507,"temperature":52.593872,"humidity":45.53692}
{"sensorId":54,"timestamp":1734247997630,"temperature":-8.243908,"humidity":94.725426}
{"sensorId":75,"timestamp":1734247997750,"temperature":-8.754467,"humidity":90.95502}
{"sensorId":30,"timestamp":1734247997869,"temperature":125.334946,"humidity":31.077433}
{"sensorId":74,"timestamp":1734247997989,"temperature":60.930824,"humidity":32.009792}
{"sensorId":55,"timestamp":1734247998108,"temperature":45.437256,"humidity":46.98705}
{"sensorId":47,"timestamp":1734247998228,"temperature":-32.51921,"humidity":61.223526}
{"sensorId":93,"timestamp":1734247998350,"temperature":-6.595146,"humidity":73.37569}
{"sensorId":62,"timestamp":1734247998469,"temperature":105.17395,"humidity":56.7289}
{"sensorId":20,"timestamp":1734247998590,"temperature":97.76599,"humidity":69.836266}
{"sensorId":47,"timestamp":1734247998709,"temperature":-11.792339,"humidity":55.22512}
{"sensorId":76,"timestamp":1734247998829,"temperature":44.995407,"humidity":89.01199}
{"sensorId":50,"timestamp":1734247998948,"temperature":55.539978,"humidity":2.4913132}
{"sensorId":60,"timestamp":1734247999069,"temperature":138.19638,"humidity":75.98818}
{"sensorId":83,"timestamp":1734247999188,"temperature":-8.153854,"humidity":83.104866}
{"sensorId":57,"timestamp":1734247999309,"temperature":64.438736,"humidity":87.46969}
```

GCS uploaded Proof for Raw Sensor Data



Current Batch Data

```
Current Batch Data
+--------+-------------+-----------+---------+
|sensorId|    timestamp|temperature| humidity|
+--------+-------------+-----------+---------+
|      12|1734248059789|  75.684525| 69.08004|
|      97|1734248059908|    48.3256|36.445694|
|      11|1734248060029|  49.931396| 41.39127|
|      16|1734248060149|  136.62585| 82.22545|
|      64|1734248060268|   64.32188|11.170721|
+--------+-------------+-----------+---------+
only showing top 5 rows
```

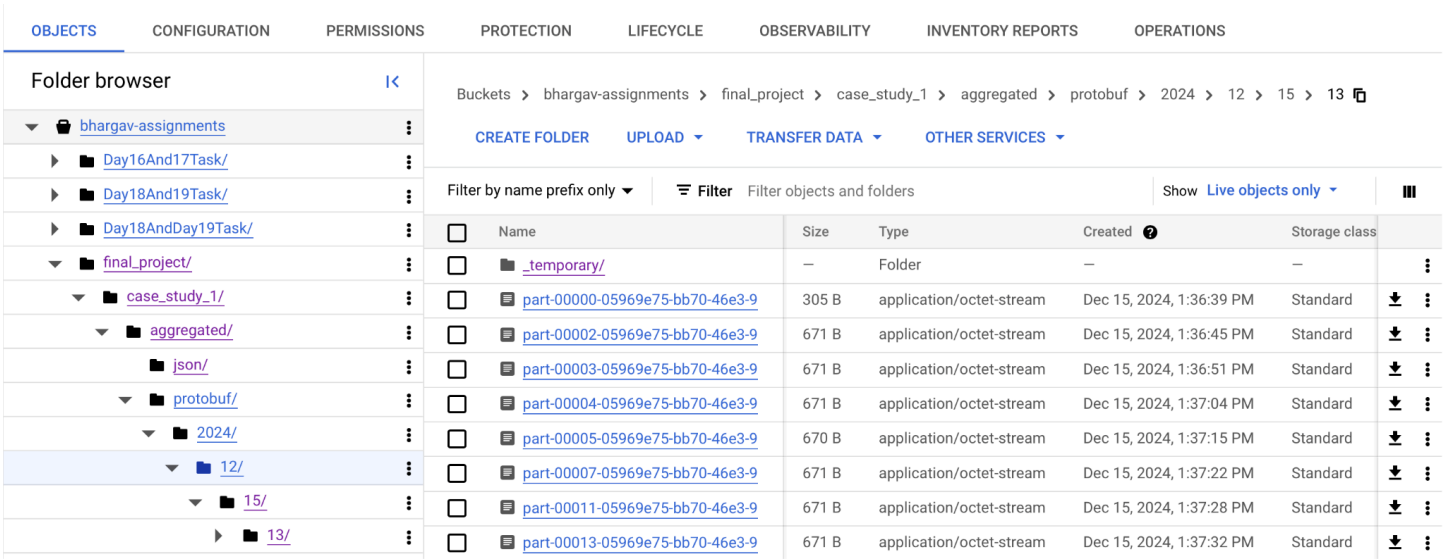
Current Batch Aggregated Data

```
Current Batch Aggregated Data
+--------+-----------------+--------------+-----------------+-----------------+--------------+--------------+-----------+
|sensorId|averageTemperature|averageHumidity|minimumTemperature|maximumTemperature|minimumHumidity|maximumHumidity|noOfRecords|
+--------+-----------------+--------------+-----------------+-----------------+--------------+--------------+-----------+
|      31|        46.962433|      53.987217|        -49.560177|         148.95181|     3.2515109|       99.51542|         85|
|      85|        54.726604|        48.5518|         -43.64445|         149.59799|     1.0321617|      97.400085|        103|
|      65|         42.58269|      46.316166|        -47.757866|         148.44275|      2.135229|       94.33407|         97|
|      53|        55.427742|      47.603973|        -48.458908|         149.87958|    0.20794272|       99.81757|         98|
|      78|        51.478294|      46.518616|        -49.199722|         149.36684|     0.5998552|       99.80707|        101|
+--------+-----------------+--------------+-----------------+-----------------+--------------+--------------+-----------+
only showing top 5 rows
```

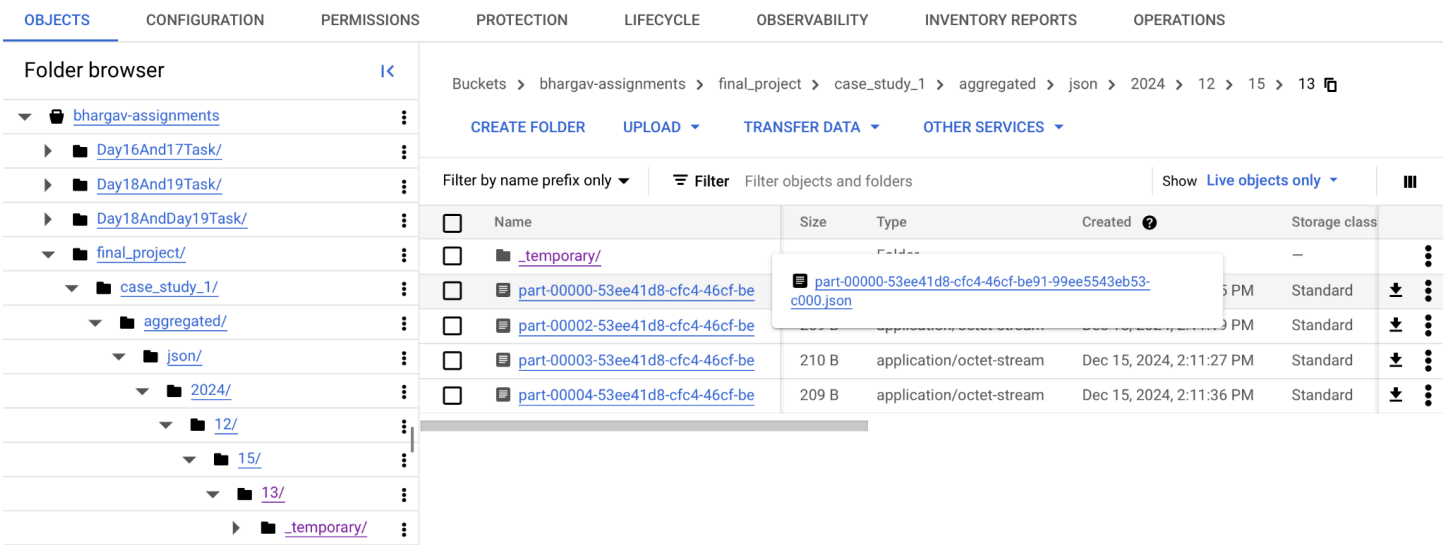
New Aggregated Data

```
New Aggregated Data
+--------+-----------------+--------------+-----------------+-----------------+--------------+--------------+-----------+
|sensorId|averageTemperature|averageHumidity|minimumTemperature|maximumTemperature|minimumHumidity|maximumHumidity|noOfRecords|
+--------+-----------------+--------------+-----------------+-----------------+--------------+--------------+-----------+
|       1|        50.898552|       53.89796|        -49.882008|          148.9321|     1.4179289|      99.301315|        144|
|       2|         48.99007|      55.650764|        -49.765682|         147.50931|     1.0541439|       99.40332|        114|
|       3|        55.502296|       51.44175|         -49.82182|         148.44707|     1.1851907|       99.25376|        162|
|       4|         49.54507|      50.719257|        -49.553013|         148.68245|      3.654939|       99.40032|        153|
|       5|        51.615704|      49.743484|         -47.71577|         148.81497|     2.0555258|       99.62398|        146|
+--------+-----------------+--------------+-----------------+-----------------+--------------+--------------+-----------+
only showing top 5 rows
```

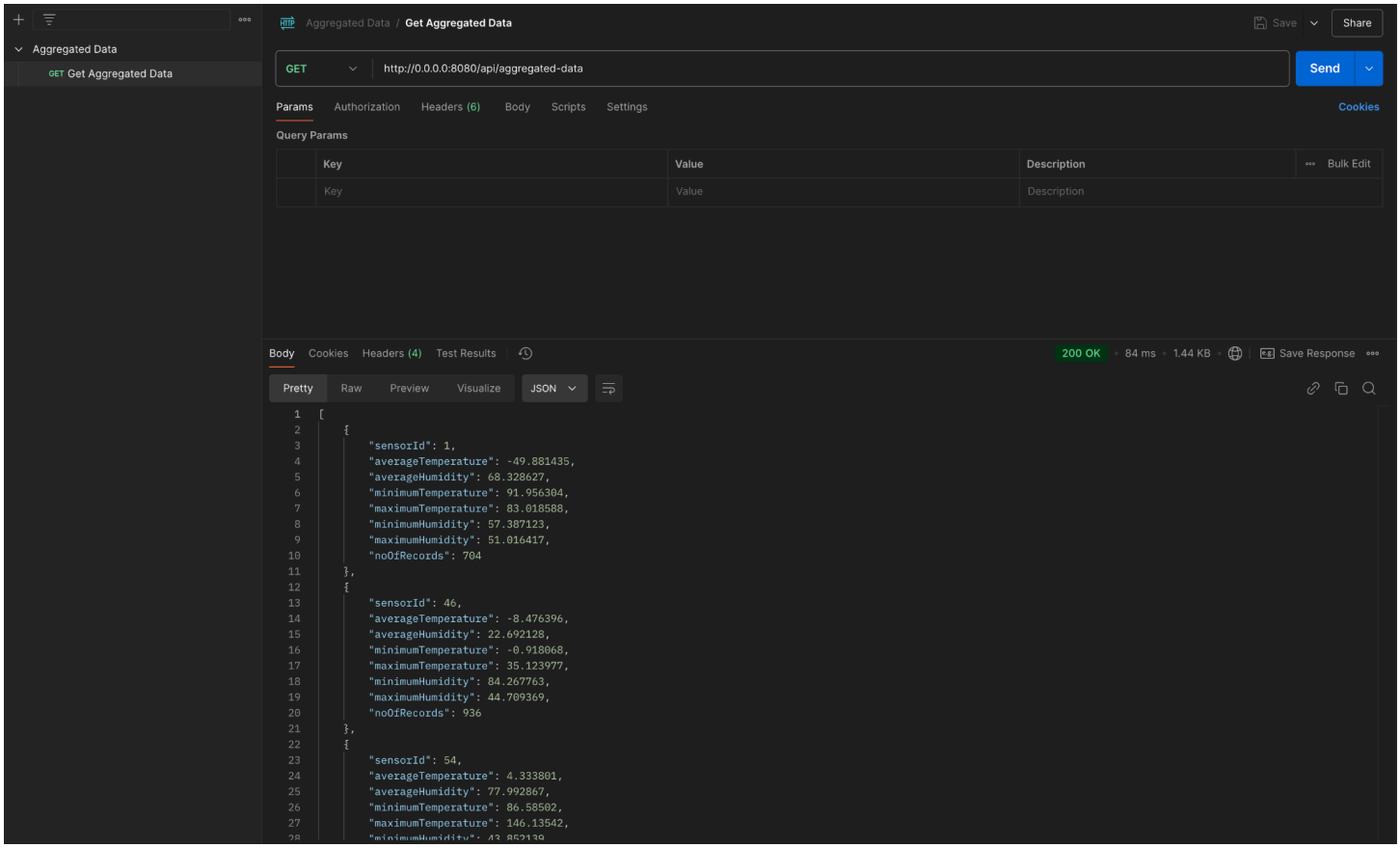
GCS upload proof for aggregated protobuf format



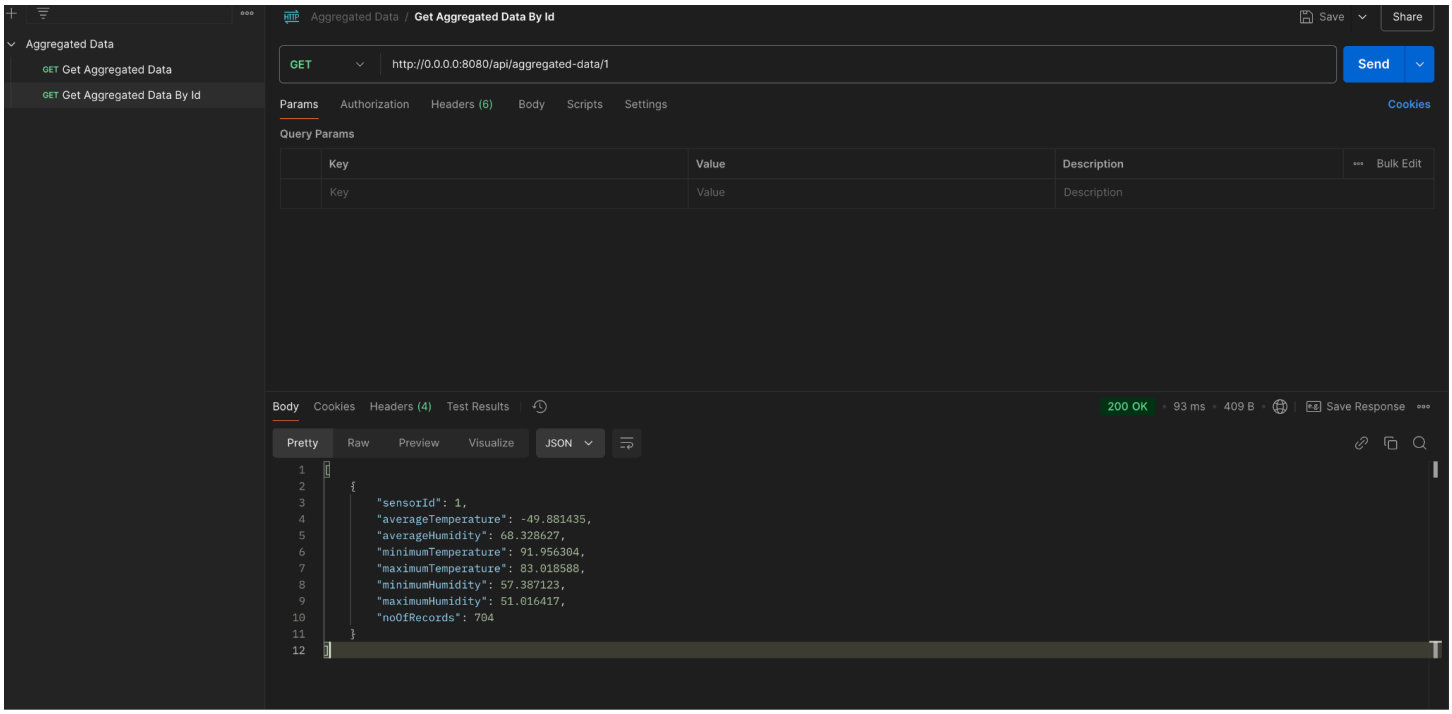
GCS upload proof for aggregated Json format



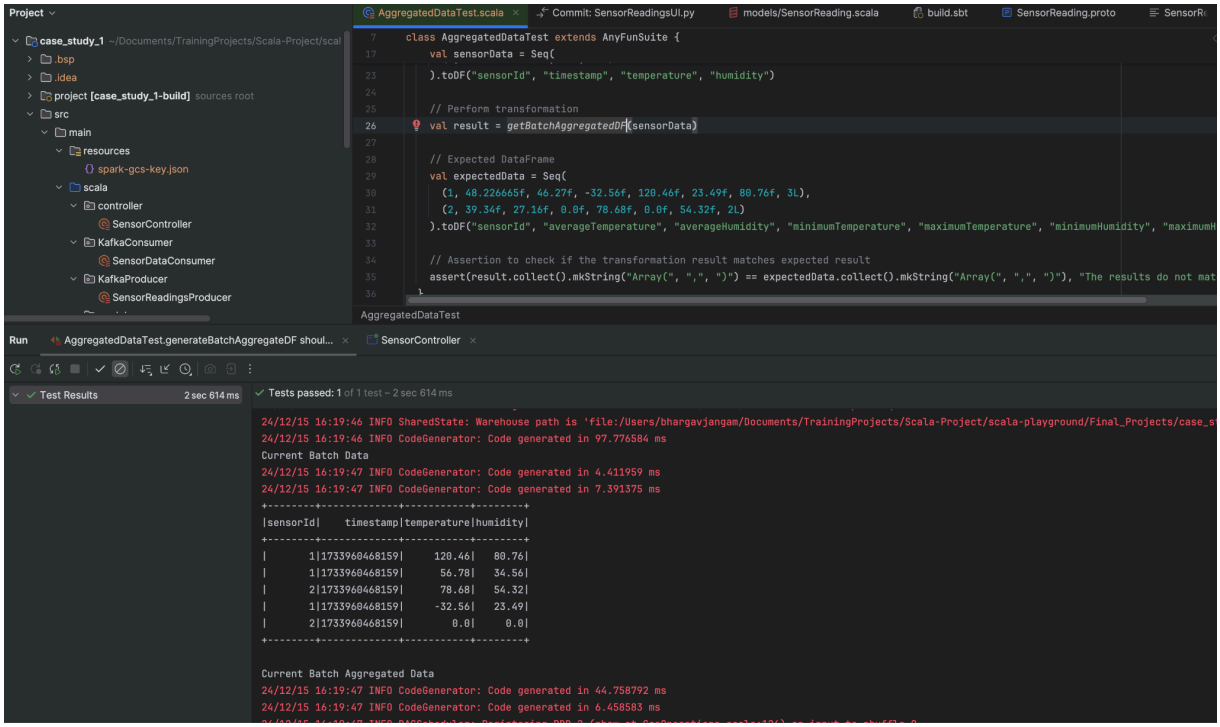
API Response for aggregated Data



API Response for aggregated Data by Id



Aggregation Test



Validation Test