

# Case\_Study\_1: IOT Data Processing

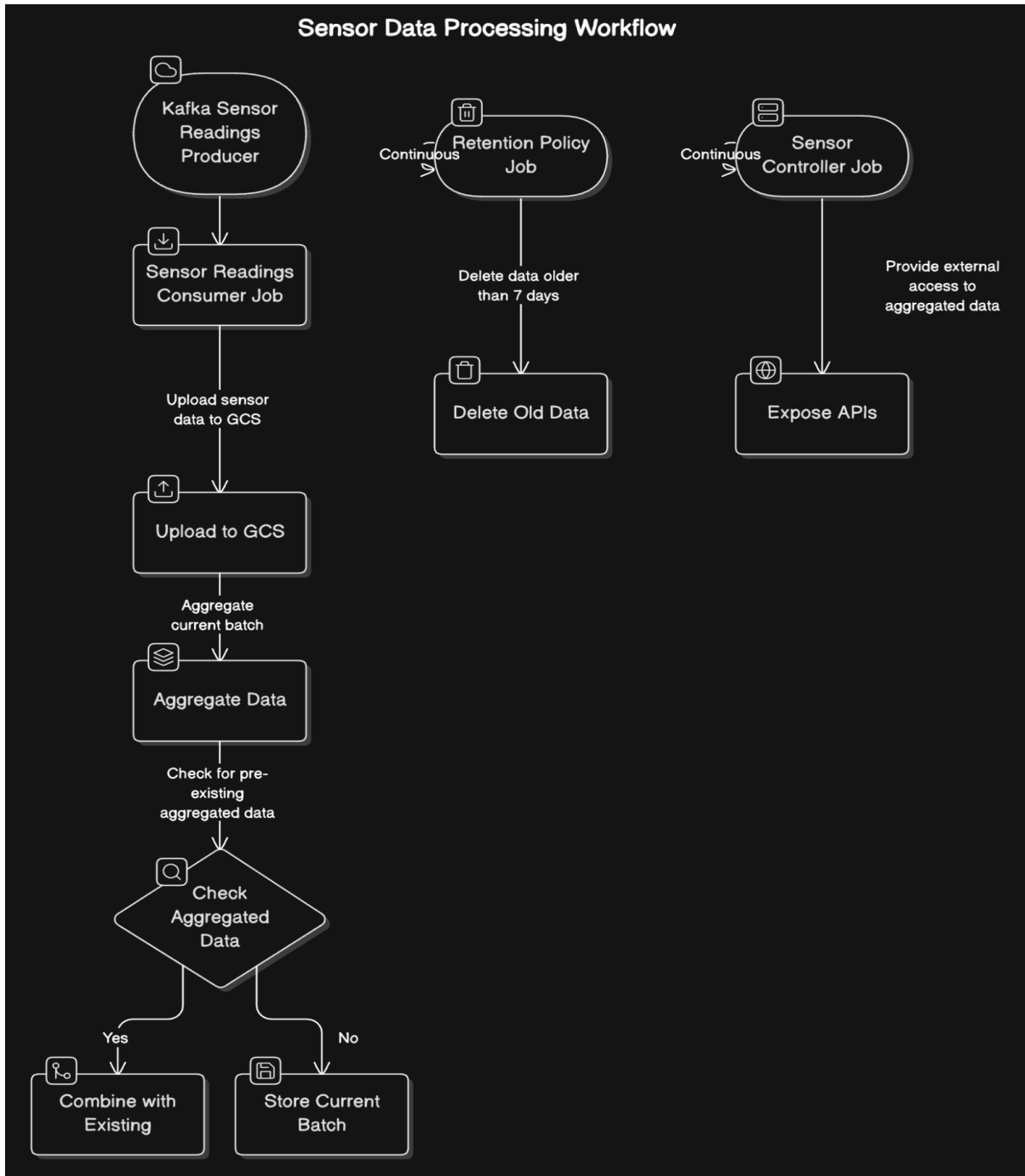
Bhargav Jangam

## 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.
bhargavjangan@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":1734247994029,"temperature":-2.0162125,"humidity":41.20615}
{"sensorId":10,"timestamp":1734247994149,"temperature":52.593872,"humidity":45.53692}
{"sensorId":54,"timestamp":1734247994269,"temperature":-8.243908,"humidity":94.725426}
{"sensorId":75,"timestamp":1734247994389,"temperature":-8.754467,"humidity":90.95502}
{"sensorId":30,"timestamp":1734247994509,"temperature":125.334946,"humidity":31.077433}
{"sensorId":74,"timestamp":1734247994629,"temperature":60.930824,"humidity":32.009792}
{"sensorId":55,"timestamp":1734247994749,"temperature":45.437256,"humidity":46.98705}
{"sensorId":47,"timestamp":1734247994869,"temperature":-32.51921,"humidity":61.223526}
{"sensorId":93,"timestamp":1734247994989,"temperature":-6.595146,"humidity":73.37569}
{"sensorId":62,"timestamp":1734247995109,"temperature":105.17395,"humidity":56.7289}
{"sensorId":20,"timestamp":1734247995229,"temperature":97.76599,"humidity":69.836266}
{"sensorId":47,"timestamp":1734247995349,"temperature":-11.792339,"humidity":55.22512}
{"sensorId":76,"timestamp":1734247995469,"temperature":44.995407,"humidity":89.01199}
{"sensorId":50,"timestamp":1734247995589,"temperature":55.539978,"humidity":2.4913132}
{"sensorId":60,"timestamp":1734247995709,"temperature":138.19638,"humidity":75.98818}
{"sensorId":83,"timestamp":1734247995829,"temperature":-8.153854,"humidity":83.104866}
{"sensorId":57,"timestamp":1734247995949,"temperature":64.438736,"humidity":87.46969}
```

## GCS uploaded Proof for Raw Sensor Data

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

OBSERVABILITY

INVENTORY REPORTS

OPERATIONS

Folder browser

▼

📁

bhargav-assignments

:

▶

📁

Day16And17Task/

:

▶

📁

Day18And19Task/

:

▶

📁

Day18AndDay19Task/

:

▼

📁

final\_project/

:

▼

📁

case\_study\_1/

:

▶

📁

aggregated/

:

▼

📁

raw/

:

▼

📁

sensor-data/

:

▼

📁

2024/

:

▼

📁

12/

:

▼

📁

15/

:

▶

📁

13/

:

Buckets > bhargav-assignments > final\_project > case\_study\_1 > raw > sensor-data > 2024 > 12 > 15 > 13

CREATE FOLDER

UPLOAD

TRANSFER DATA

OTHER SERVICES

Filter by name prefix only

Filter

Filter objects and folders

Show

Live objects only

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	
<input type="checkbox"/>	<a href="#">_SUCCESS</a>	0 B	application/octet-stream	Dec 15, 2024, 1:08:01 PM	Standard	<div><div>📄</div><div>⋮</div></div>
<input type="checkbox"/>	<a href="#">part-00000-489ba93c-ef12-4cd8-b7</a>	6 KB	application/octet-stream	Dec 15, 2024, 1:07:51 PM	Standard	<div><div>📄</div><div>⋮</div></div>
<input type="checkbox"/>	<a href="#">part-00001-489ba93c-ef12-4cd8-b7</a>	3.7 KB	application/octet-stream	Dec 15, 2024, 1:07:54 PM	Standard	<div><div>📄</div><div>⋮</div></div>
<input type="checkbox"/>	<a href="#">part-00002-489ba93c-ef12-4cd8-b7</a>	15.3 KB	application/octet-stream	Dec 15, 2024, 1:07:56 PM	Standard	<div><div>📄</div><div>⋮</div></div>

## Current Batch Data

```
Current Batch Data
+-----+-----+-----+-----+
|sensorId|    timestamp|temperature| humidity|
+-----+-----+-----+-----+
|      12|1734248059789|   75.684525| 69.08004|
|      97|1734248059908|   48.3256|36.445694|
|     111|1734248060029|   49.931396| 41.39127|
|     161|1734248060149|  136.62585| 82.22545|
|     641|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

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

OBSERVABILITY

INVENTORY REPORTS

OPERATIONS

Folder browser

▼

📁

bhargav-assignments

⋮

▶

📁

Day16And17Task/

⋮

▶

📁

Day18And19Task/

⋮

▶

📁

Day18AndDay19Task/

⋮

▼

📁

final\_project/

⋮

▼

📁

case\_study\_1/

⋮

▼

📁

aggregated/

⋮

📁

json/

⋮

▼

📁

protobuf/

⋮

▼

📁

2024/

⋮

▼

📁

12/

⋮

▼

📁

15/

⋮

▶

📁

13/

⋮

Buckets

>

bhargav-assignments

>

final\_project

>

case\_study\_1

>

aggregated

>

protobuf

>

2024

>

12

>

15

>

13

📁

CREATE FOLDER

UPLOAD

TRANSFER DATA

OTHER SERVICES

Filter by name prefix only

Filter

Filter objects and folders

Show

Live objects only

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	
<input type="checkbox"/>	📁 <a href="#">_temporary/</a>	—	Folder	—	—	⋮
<input type="checkbox"/>	📄 <a href="#">part-00000-05969e75-bb70-46e3-9</a>	305 B	application/octet-stream	Dec 15, 2024, 1:36:39 PM	Standard	📄 ⋮
<input type="checkbox"/>	📄 <a href="#">part-00002-05969e75-bb70-46e3-9</a>	671 B	application/octet-stream	Dec 15, 2024, 1:36:45 PM	Standard	📄 ⋮
<input type="checkbox"/>	📄 <a href="#">part-00003-05969e75-bb70-46e3-9</a>	671 B	application/octet-stream	Dec 15, 2024, 1:36:51 PM	Standard	📄 ⋮
<input type="checkbox"/>	📄 <a href="#">part-00004-05969e75-bb70-46e3-9</a>	671 B	application/octet-stream	Dec 15, 2024, 1:37:04 PM	Standard	📄 ⋮
<input type="checkbox"/>	📄 <a href="#">part-00005-05969e75-bb70-46e3-9</a>	670 B	application/octet-stream	Dec 15, 2024, 1:37:15 PM	Standard	📄 ⋮
<input type="checkbox"/>	📄 <a href="#">part-00007-05969e75-bb70-46e3-9</a>	671 B	application/octet-stream	Dec 15, 2024, 1:37:22 PM	Standard	📄 ⋮
<input type="checkbox"/>	📄 <a href="#">part-00011-05969e75-bb70-46e3-9</a>	671 B	application/octet-stream	Dec 15, 2024, 1:37:28 PM	Standard	📄 ⋮
<input type="checkbox"/>	📄 <a href="#">part-00013-05969e75-bb70-46e3-9</a>	671 B	application/octet-stream	Dec 15, 2024, 1:37:32 PM	Standard	📄 ⋮

## GCS upload proof for aggregated Json format

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

OBSERVABILITY

INVENTORY REPORTS

OPERATIONS

Folder browser

▼

📁

bhargav-assignments

▶

📁

Day16And17Task/

▶

📁

Day18And19Task/

▶

📁

Day18AndDay19Task/

▼

📁

final\_project/

▼

📁

case\_study\_1/

▼

📁

aggregated/

▼

📁

json/

▼

📁

2024/

▼

📁

12/

▼

📁

15/

▼

📁

13/

▶

📁

\_temporary/

Buckets > bhargav-assignments > final\_project > case\_study\_1 > aggregated > json > 2024 > 12 > 15 > 13

CREATE FOLDER

UPLOAD

TRANSFER DATA

OTHER SERVICES

Filter by name prefix only

Filter

Filter objects and folders

Show Live objects only

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	
<input type="checkbox"/>	📁 _temporary/				—	
<input type="checkbox"/>	📄 part-00000-53ee41d8-cfc4-46cf-be					
<input type="checkbox"/>	📄 part-00002-53ee41d8-cfc4-46cf-be					
<input type="checkbox"/>	📄 part-00003-53ee41d8-cfc4-46cf-be	210 B	application/octet-stream	Dec 15, 2024, 2:11:27 PM	Standard	<div> <div>📄</div> <div>part-00000-53ee41d8-cfc4-46cf-be91-99ee5543eb53-c000.json</div> </div>
<input type="checkbox"/>	📄 part-00004-53ee41d8-cfc4-46cf-be	209 B	application/octet-stream	Dec 15, 2024, 2:11:36 PM	Standard	

# API Response for aggregated Data

+

☰

...

Aggregated Data

Get Aggregated Data

GET

http://0.0.0.0:8080/api/aggregated-data

Send

Params

Authorization

Headers (6)

Body

Scripts

Settings

Cookies

Query Params

	Key	Value	Description	...	Bulk Edit
	Key	Value	Description		

Body

Cookies

Headers (4)

Test Results

200 OK

84 ms

1.44 KB

Save Response

...

Pretty

Raw

Preview

Visualize

JSON

...

```
1  [
2    {
3      "sensorId": 1,
4      "averageTemperature": -49.881435,
5      "averageHumidity": 68.328627,
6      "minimumTemperature": 91.956304,
7      "maximumTemperature": 83.018588,
8      "minimumHumidity": 57.387123,
9      "maximumHumidity": 51.016417,
10     "noOfRecords": 704
11   },
12   {
13     "sensorId": 46,
14     "averageTemperature": -8.476396,
15     "averageHumidity": 22.692128,
16     "minimumTemperature": -0.918068,
17     "maximumTemperature": 35.123977,
18     "minimumHumidity": 84.267763,
19     "maximumHumidity": 44.709369,
20     "noOfRecords": 936
21   },
22   {
23     "sensorId": 54,
24     "averageTemperature": 4.333801,
25     "averageHumidity": 77.992867,
26     "minimumTemperature": 86.58502,
27     "maximumTemperature": 146.13542,
28     "minimumHumidity": 40.852130
```

# API Response for aggregated Data by Id

+

☰

...

Aggregated Data

Get Aggregated Data

Get Aggregated Data By Id

GET

http://0.0.0.0:8080/api/aggregated-data/1

Send

Params

Authorization

Headers (6)

Body

Scripts

Settings

Cookies

Query Params

	Key	Value	Description	...	Bulk Edit
	Key	Value	Description		

Body

Cookies

Headers (4)

Test Results

200 OK

93 ms

409 B

Save Response

...

Pretty

Raw

Preview

Visualize

JSON

...

```
1  {
2    "sensorId": 1,
3    "averageTemperature": -49.881435,
4    "averageHumidity": 68.328627,
5    "minimumTemperature": 91.956304,
6    "maximumTemperature": 83.018588,
7    "minimumHumidity": 57.387123,
8    "maximumHumidity": 51.016417,
9    "noOfRecords": 704
10  }
```

# Aggregation Test

Project

case\_study\_1 ~/Documents/TrainingProjects/Scala-Project/scala

bsp

idea

project [case\_study\_1-build] sources root

src

main

resources

spark-gcs-key.json

scala

controller

SensorController

KafkaConsumer

SensorDataConsumer

KafkaProducer

SensorReadingsProducer

AggregatedDataTest.scala

```
7 class AggregatedDataTest extends AnyFunSuite {
17   val sensorData = Seq(
23     .toDF("sensorId", "timestamp", "temperature", "humidity")
24   )
25   // Perform transformation
26   val result = getBatchAggregatedDF(sensorData)
27
28   // Expected DataFrame
29   val expectedData = Seq(
30     (1, 48.226665f, 46.27f, -32.56f, 120.46f, 23.49f, 80.76f, 3L),
31     (2, 39.34f, 27.16f, 0.8f, 78.68f, 0.8f, 54.32f, 2L)
32   ).toDF("sensorId", "averageTemperature", "averageHumidity", "minimumTemperature", "maximumTemperature", "minimumHumidity", "maximumHumidity")
33
34   // Assertion to check if the transformation result matches expected result
35   assert(result.collect().mkString("Array(", ",", ")") == expectedData.collect().mkString("Array(", ",", ")"), "The results do not match")
36 }
```

Run

AggregatedDataTest.generateBatchAggregateDF should...

SensorController

Test Results

2 sec 614 ms

Tests passed: 1 of 1 test - 2 sec 614 ms

24/12/15 16:19:46 INFO SharedState: Warehouse path is 'file:/Users/bhargavjangam/Documents/TrainingProjects/Scala-Project/scala-playground/Final\_Projects/case\_study\_1/'
24/12/15 16:19:46 INFO CodeGenerator: Code generated in 97.776584 ms
Current Batch Data
24/12/15 16:19:47 INFO CodeGenerator: Code generated in 4.411959 ms
24/12/15 16:19:47 INFO CodeGenerator: Code generated in 7.391375 ms
+-----+
|sensorId| timestamp|temperature|humidity|
+-----+
1	1733960468159	120.46	80.76
1	1733960468159	54.78	34.54
2	1733960468159	78.68	54.32
1	1733960468159	-32.56	23.49
2	1733960468159	0.8	0.8
+-----+
Current Batch Aggregated Data
24/12/15 16:19:47 INFO CodeGenerator: Code generated in 44.758792 ms
24/12/15 16:19:47 INFO CodeGenerator: Code generated in 6.458583 ms
24/12/15 16:19:47 INFO CodeGenerator: Code generated in 4.411959 ms
24/12/15 16:19:47 INFO CodeGenerator: Code generated in 7.391375 ms

# Validation Test

case\_study\_1 ~/Documents/TrainingProjects/Scala-Project/scala

bsp

idea

project [case\_study\_1-build] sources root

src

main

resources

spark-gcs-key.json

scala

controller

SensorController

KafkaConsumer

SensorDataConsumer

KafkaProducer

SensorReadingsProducer

models

SensorReading

DataValidationTest.scala

```
8 class DataValidationTest extends AnyFunSuite {
17   val sensorData = Seq(
23     .toDF("sensorId", "timestamp", "temperature", "humidity")
24   )
25   test("validateSensorData") {
26     // Sample data with some invalid rows
27     val sensorData = Seq(
28       Row(1, 1733960468159L, 120.46f, 80.76f),
29       Row(1, 1733960468159L, -56.78f, 34.56f), // Invalid temperature
30       Row(2, 1733960468159L, 78.68f, 150.32f), // Invalid humidity
31       Row(3, null, 75.65f, 45.32f), // Invalid timestamp
32       Row(4, 1733960468159L, 200.00f, 40.00f), // Invalid temperature
33       Row(5, 1733960468159L, 23.45f, -5.00f) // Invalid humidity
34     )
35
36     // Create DataFrame using the schema
37     val sensorDataDF = spark.createDataFrame(spark.sparkContext.parallelize(sensorData), sensorDataSchema)
38
39     // Actual validation method
40   }
```

Run

DataValidationTest.validateSensorData

SensorController

Test Results

2 sec 75 ms

Tests passed: 1 of 1 test - 2 sec 75 ms

/Users/bhargavjangam/Library/Java/JavaVirtualMachines/corretto-11.0.25/Contents/Home/bin/java ...
Testing started at 4:42 pm ...
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
24/12/15 16:42:14 INFO SparkContext: Running Spark version 3.5.1
24/12/15 16:42:14 INFO SparkContext: OS info Mac OS X, 14.6, aarch64
24/12/15 16:42:14 INFO SparkContext: Java version 11.0.25
24/12/15 16:42:14 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
24/12/15 16:42:14 INFO ResourceUtils: =====
24/12/15 16:42:14 INFO ResourceUtils: No custom resources configured for spark.driver.
24/12/15 16:42:14 INFO ResourceUtils: =====
24/12/15 16:42:14 INFO SparkContext: Submitted application: Sensor Data Validation Test
24/12/15 16:42:14 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 1, script: , vendor: , memory -> )
24/12/15 16:42:14 INFO ResourceProfile: Limiting resource is cpu
24/12/15 16:42:14 INFO ResourceProfileManager: Added ResourceProfile id: 0
24/12/15 16:42:14 INFO SecurityManager: Changing view acls to: bhargavjangam
24/12/15 16:42:14 INFO SecurityManager: Changing modify acls to: bhargavjangam
24/12/15 16:42:14 INFO SecurityManager: Changing view acls groups to:
24/12/15 16:42:14 INFO SecurityManager: Changing modify acls groups to:

Sensor Metrics Viewer

Search by Sensor ID

Enter Sensor ID here

Search

Refresh Data

	averageHumidity	averageTemperature	maximumHumidity	maximumTemperature	minimumHumidity
0	53.945107	51.061460	99.954920	149.152130	48.555340
1	44.019104	43.699970	98.505720	149.915180	49.165474
2	45.997738	57.308464	99.817570	149.879580	48.458908
3	48.550945	49.871838	99.931595	147.816590	47.956690
4	50.906960	45.097115	97.654686	149.779920	49.871754
5	46.408936	49.638275	99.266780	147.020320	48.724556
6	46.320564	49.522880	99.623955	149.357790	47.800290
7	47.706680	54.111317	97.400085	149.597990	43.644450
8	47.455930	50.012287	99.807070	149.366840	49.199722
9	47.414455	43.261375	97.716820	148.388290	49.578728
10	52.796066	45.482204	98.776070	140.506820	49.949410
11	54.572510	53.773540	96.247670	147.440320	49.774887
12	53.810770	47.671890	99.515420	148.951810	49.560177
13	45.546080	44.550194	94.334070	148.442750	47.757866
14	49.340298	45.143800	99.460110	145.434160	47.849346
15	50.819630	48.898464	99.862220	149.547850	49.956383

✓ Fetched 16 records successfully!

Sensor Metrics Viewer

Search by Sensor ID

Enter Sensor ID here

Search

Refresh Data

id	maximumTemperature	minimumHumidity	minimumTemperature	noOfRecords	sensorId
920	149.152130	0.376070	-48.555340	170	26
720	149.915180	0.510061	-49.165474	165	27
570	149.879580	0.207943	-48.458908	156	53
595	147.816590	0.131553	-47.956690	168	81
686	149.779920	1.255494	-49.871754	160	76
780	147.020320	1.185304	-48.724556	141	44
955	149.357790	0.754827	-47.800290	160	93
085	149.597990	1.032162	-43.644450	161	85
070	149.366840	0.599855	-49.199722	160	78
820	148.388290	1.001364	-49.578728	156	28
070	140.506820	1.250297	-49.949410	162	12
670	147.440320	0.388360	-49.774887	161	91
420	148.951810	3.251511	-49.560177	138	31
070	148.442750	2.135229	-47.757866	155	65
110	145.434160	3.186470	-47.849346	150	34
220	149.547850	2.712959	-49.956383	184	22

✓ Fetched 16 records successfully!



UI (api/aggregated-data/{sensorId})

### Sensor Metrics Viewer

Search by Sensor ID

26

Search

Refresh Data

	averageHumidity	averageTemperature	maximumHumidity	maximumTemperature	minimumHumidity
0	53.945107	51.061460	99.954920	149.152130	

Found 1 records for Sensor ID 26.

### Sensor Metrics Viewer

Search by Sensor ID

26

Search

Refresh Data

id	maximumTemperature	minimumHumidity	minimumTemperature	noOfRecords	sensorId
1920	149.152130	0.376070	-48.555340	170	26

Found 1 records for Sensor ID 26.