



Senior Data Engineer Case Study

Case Study Steps

You work as a **data engineer** and your mission is to build **batch and real-time data pipelines**.

Each explained part will be separate component inside your docker compose file. When your docker instances start running they should follow the below path of execution.

1. Streaming

- a. Use the ***measurements.parquet*** file and create a script that will read and write the data to your Kafka instance one by one.
- b. The file contains two different types of measurements (Temperature, Humidity). Take this into account when writing to your Kafka instance.

2. Consumer

- a. Read the data from your Kafka instance and write your data into any database instance you want. Please make sure to create it in the docker-compose.yaml file given within the project file.

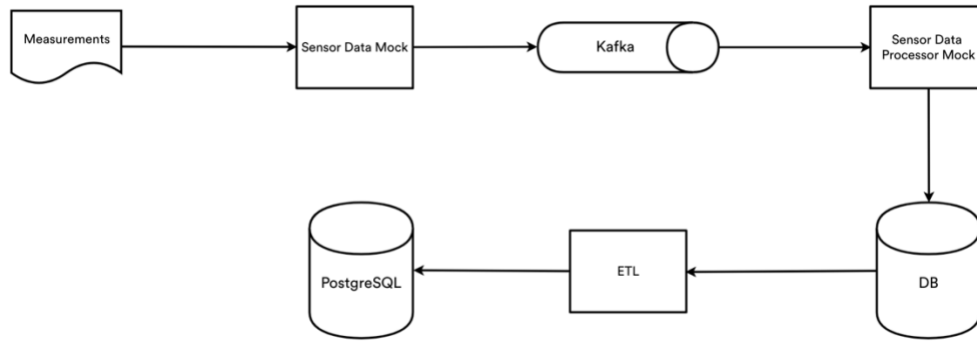
3. Batch Processing (ETL)

- a. Read the data from your created database instance and create the following reports.
 - i. Average summary of temperature data over a 10-minute interval.
 - ii. Average summary of humidity data over a 20-minute interval.
- b. Write your resulting data into a PostgreSQL instance.

After your docker instances have completed running we should be able to see your results inside your PostgreSQL instance within your docker compose.

Notes:

- You should **create the given architecture below**
 - a sensor data producer. (Sensor Data Mock)
 - a streaming app to read views from Kafka and integrate with API (Sensor Data Processor Mock)
 - a batch summary transformer (ETL)



Technical Details

Environment

Your solution should contain a docker compose file that when executed the data will appear inside the PostgreSQL instance. The Kafka instance is already provided within the docker-compose.yml file. It will start automatically.

Notes

- Preferred languages are Python and Scala. But you are free to choose other languages if you prefer. If so, please explain why you have chosen a different language.
- You are free to choose and install any other tools that you think might be a good fit.
- Please try to write unit tests if possible.
- Upon completion, please zip your files and share it with us.
- Please include your architectural drawing of the solution in a folder named solution-files.

You are given 7 days to complete the case study. If any questions occurred during the process, please get in touch with us.

We are eager to see your solution to the problem. There are no wrong answers here. So, keep it simple 😊