Data Engineering technical challenge @Flink

1. Implement a **Python** module that reads the sample data file provided and performs the following actions (keep in mind that the system could be larger ~10GB):
    a. Validates the schema of the file -> it should check that the values have the expected type (keep in mind that the schema may change in the future)
    b. Inserts the data into a Postgres/MySQL DB

2. **Monitoring** of the pipeline:
    a. Integrate (or implement it independently) a mechanism that checks if all the data in the file has been properly stored
    b. Can you think of any different aspects of the pipeline that would be worth monitoring?

3. **CI/CD**:
    a. Describe from a high level the different steps you would implement to transition the pipeline from a development environment into a production one

4. *(Optional:)* Bonus points if everything is dockerised. From a high level, how would you deploy it in Kubernetes?

Delivery of the challenge: please, create a repository in **GitHub**, and share the link with us.

Note: It is more important to provide explanations and high-level approach to the problem than submitting a fully complete and running solution. E.g. if a function would take you too long to develop, then instead just document what the function should do for you.

Have fun!