Take-home Test for the position of

Software Engineers (Junior Level)

**Question 1:**

Consider the situation that we need to process all files under a certain directory, and insert records stored in each file into a database. The file hierarchy is as follows:

* There is no nested folder inside the target directory
* All files are in CSV format

Example:

|- target-directory/

|- file-1.csv

|- file-2.csv

|- …

|- file-N.csv

Also, all CSV files have the following specs:

* The first row of file contains header names
* There are three columns in every CSV file
  + timestamp: in ISO8601 date-time format with time zone
  + sensorName: name of sensor
  + value: numeric value of the sensor reading
* Each file has large amounts of data (> 1GB), but you can assume a file is always able to be fitted in memory (e.g., < 8 GB)

Each record represents a reading from an IoT sensor, note that each sensor can only emit only 1 record at a time. Please design and implement:

1. A relational database schema with proper configuration (e.g., indexes, constraints, etc.)
2. A program that allows users to configure (a) the target directory and (b) target database

With the following requirements:

1. The program should be able to process files in parallel
2. The number of files in a directory could be large (e.g., 1000 files)
3. Records are not guaranteed to be in valid format. For example,
   1. Timestamp is not a ISO8601 string
   2. Value contains non-numeric string (e.g., true, “ABC”, etc.)
4. Records can also be duplicated in the same/different file(s)

Please also provide the setup script/test cases/mock data if appropriate.

You can choose any tech stack of your choice. The application should be able to be executed in the docker environment. Automated installation, e.g., shell script, is preferred when suitable. Please note there is no hard performance goal to meet, but the program design, maintainability, and extendibility will be considered.

Please also prepare a short demonstration for the interview. And a git repository (avoid words like “ATAL,” “interview” in the repository name) containing your work. We are interested to know more about the task details in the upcoming interview.

**<End>**