Challenge #1

You are a data engineer at Globant and you are about to start an important project. This project is big data migration to a new database system. You need to create a PoC to solve the next requirements:

1. Move historic data from files in CSV format to the new database.
2. Create a Rest API service to receive new data. This service must have:
    2.1. Each new transaction must fit the data dictionary rules.
    2.2. Be able to insert batch transactions (1 up to 1000 rows) with one request.
    2.3. Receive the data for each table in the same service.
    2.4. Keep in mind the data rules for each table.
3. Create a feature to backup for each table and save it in the file system in AVRO format.
4. Create a feature to restore a certain table with its backup.

You need to publish your code in GitHub. It will be taken into account if frequent updates are made to the repository that allow analyzing the development process.

Clarifications

- You decide the origin where the CSV files are located.
- You decide the destination database type, but it must be a SQL database.
- The CSV file is comma separated.
- "Feature" must be interpreted as "Rest API, Stored Procedure, Database functionality, Cron job, or any other way to accomplish the requirements".

Not mandatory, but taken into account:

- Create a markdown file for the Readme.md
- Security considerations for your API service
- Use the Git workflow to create versions
- Create a Dockerfile to deploy the package
- Use cloud tools instead of local tools

You can use Python, Java, Go or Scala to solve it!

Data Rules

- Transactions that don't accomplish the rules must not be inserted but they must be logged.
- All the fields are required.

Csv files structures

hired_employees.csv:

id	INTEGER	Id of the employee
name	STRING	Name and surname of the employee
datetime	STRING	Hire datetime in ISO format
department_id	INTEGER	Id of the department which the employee was hired for
job_id	INTEGER	Id of the job which the employee was hired for

4535,Marcelo Gonzalez,2021-07-27T16:02:08Z,1,2
4572,Lidia Mendez,2021-07-27T19:04:09Z,1,2
Link to the file

departments.csv

id	INTEGER	Id of the department
department	STRING	Name of the department

1, Supply Chain
2, Maintenance
3, Staff
Link to the file

jobs.csv:

id	INTEGER	Id of the job
job	STRING	Name of the job

1, Recruiter
2, Manager
3, Analyst
Link to the file

Challenge #2

You need to explore the data that was inserted in the first challenge. The stakeholders ask for some specific metrics they need. You should create an end-point for each requirement.

Requirements

- Number of employees hired for each job and department in 2021 divided by quarter. The table must be ordered alphabetically by department and job.

  Output example:

department	job	Q1	Q2	Q3	Q4
Staff	Recruiter	3	0	7	11
Staff	Manager	2	1	0	2
Supply Chain	Manager	0	1	3	0

- List of ids, name and number of employees hired of each department that hired more employees than the mean of employees hired in 2021 for all the departments, ordered by the number of employees hired (descending).

  Output example:

id	department	hired
Staff	Staff	45
Staff	Supply Chain	12

Not mandatory, but taken into account:

- Create a visual report for each requirement using your favorite tool