huddle

ETL Developer

Take-home project

Overview

This test is designed to give you an opportunity to demonstrate your knowledge of dealing with extraction, transformation and loading of the data.

Scenario

Our Quantitative Analysts are working on an NBA quantitative model. They realized they were missing parts of the team and player statistics for the season 2017/2018 and 2018/2019. Help them in filling the database with data they need.

Objective

1. **Download**
   From the website, https://www.nba.com/ find and download the required data for the missing seasons:
   a. Team statistics (per mode: game): these are the aggregate statistics for all the teams, averaged per game

b. Player statistics (per mode: game): these are the aggregate statistics for each player, averaged per game

⚠️ Make sure to download all the data for all the required seasons.

2. **Store**
   Define an appropriate schema for the just-downloaded data and persist them in a suitable relational database (preferably PostgreSQL).

   ⚠️ You are free to choose how to create the tables and to insert the data in the DB, however the usage of enterprise-grade best practices will be recognized and rewarded.

3. **Query**
   Provide a query which returns a players table with the following requirements:
   a. Each row represents a player
   b. 5 columns:
      i. Player full name
      ii. His team's name
      iii. Average points scored by the player
      iv. How much better (or worse) he is with respect to his team average
      v. How much better (or worse) he is with respect to the whole average
   c. Display only the best 3 players per team
   d. Sort by point average, descending order

   ⚠️ Query complexity and joins performance will be evaluated

Technology requirements

- It is expected for the assignment solution to be runnable on our side:
  - The preferred and recommended way is to deliver a containerized solution, executable on any machine containing Docker Engine. Please include the Dockerfile used to build the container image as part of the project.
  - Alternatively, thoroughly describe how to reproduce the environment used to run the application. This option is discouraged and should be avoided if possible.
  - Non-runnable solutions will be disregarded

- Feel free to use any additional tools (workflow managers, libraries, etc.) you see fit for this project
- The whole technology stack should be open source

Hand-in requirements

The code and any additional files, as well as the downloaded data, must be placed in a compressed archive and sent to `gabriele@huddle.tech`