Data Engineering Test Overview

Hello, this is a short test to help us understand how your skills and abilities would fit in at NuData. We conduct tests for all technical candidates, to help provide a fair opportunity for you to demonstrate your abilities.

Note:

- You will have **24 hours** from the receipt of the test to send us your completed work.
- Please feel free to use google, stackoverflow, or any other references you wish.
- Please read the full test description before starting.

Scenario

The Analytics Team has described for you a series of SQL queries that they regularly run against our data store to pull aggregate information. They use these queries to create daily reports and/or perform ad-hoc investigations.

Your task is to automate this process by creating a system which automatically extracts the information on a regular basis, and dumps that information to a local csv file.

The queries themselves are described below, as is the format of the csv's expected output.

Feel free to use whatever tooling/language you feel is most appropriate to the task.

Target Queries

1 How many players are in the table?
2. How many teams are there?
3. Who are the top 3 players who scored the most goals in order and how many goals?
4. How many teams had more than 250 goals from all their team players combined?

Output Format Specification

The csv is expected to be pipe-delimited with a header.

A raw extract might look like so:

```
Report DateTime|Total Players|Total Teams|Top Three Players Stats|Teams GT 250
2018-06-22 5:00|500|60|'[{"Name": "foobar", "Goals": 72}, {"Name": "bizbaz", "Goals": 66}, {"Name": "blipblorp", "Goals": 2}]'|3
2018-06-23 5:00|375|40|'[{"Name": "foobar", "Goals": 80}, {"Name": "bizbaz", "Goals": 50}, {"Name": "blipblorp", "Goals": 2}]'|2
```

As a table, it would like so:

Report DateTime	Total Players	Total Teams	Top Three Players Stats	Teams GT 250
18-06-22 5:00	500	60	[{"Name": "foobar", "Goals": 72}, {"Name": "bizbaz", "Goals": 66}, {"Name": "blipblorp", "Goals": 2}]	3
18-06-23 5:00	375	40	[{"Name": "foobar", "Goals": 80}, {"Name": "bizbaz", "Goals": 50}, {"Name": "blipblorp", "Goals": 2}]	2

Each run appends a new row of data to the csv.

“Report DateTime” is the Date and Time that the report was run.

Scheduling

The Analytics team is expecting this data every morning, so please include a cron file (or some scheduling system) to have your script/application execute every morning at 5am, Monday through Friday.

Datastore

You will be connecting to the attached SQLite DB, “de_test_db.sqlite”. The table you’ll be working with is called sqltest_table.

Bonus Requirements!

As an optional bonus, feel free to include one or all of the following requirements when writing your script/application:

- Assume your application could potentially run concurrently. Take whatever precautions you feel are necessary to deal with this scenario.

- Set up a rudimentary monitoring framework which emails a list of stakeholders whenever your script/application blows up, and/or unexpected conditions are encountered.

Tools and Resources:

http://www.w3schools.com/sql/sql_select.asp
http://www.sqlitetutorial.net/sqlite-cheat-sheet/
https://www.tutorialspoint.com/sqlite/sqlite_installation.htm

Finally:

· Please make sure to submit all files, source code, etc relating to your project. It's more important that we see *how* you work rather than just whether your code produces the expected output.

Good luck!