**Home Assignment for QA Automation Position**

This assignment will focus on database performance benchmarking.

You will implement a utility that benchmarks TPC-H queries over PostgreSQL database.

**Guidelines + Required prerequisites / Installations:**

Use Python 3.6 or higher

Install and use PostgreSQL

Scripts needs to be implemented in Python.

\*For the purpose of this task, we will only focus on two TPCH queries – tpch5 + tpch7.

TPCH data files can be downloaded from links below:

TPCH Schema: <https://github.com/dimitri/tpch-citus/blob/master/schema/tpch-schema.sql>

TPCH Data: <https://drive.google.com/file/d/1fczEwvvg0mOKxHn3hWkz-MGhs8WjIn5A/view?usp=sharing>

TPCH queries: <https://github.com/2ndQuadrant/pg-tpch/tree/master/queries>

**Timeframe for this assignment – up to 1 day.**

**Required Script Outputs + Functionalities:**

***./tpch\_util.py --create\_schema***

Will create the TPCH tables. \*In case the tables exist, it should be dropped and replaced.

***./tpch\_util.py --load\_data***

Will load the TPCH csv files to the corresponding tables.

***./tpch\_util.py --run-benchmark***

Will run the TPCH queries and output the query execution time

***./tpch\_util.py --run-benchmark --save-results***

When save results argument is provided, results should be saved to tpch\_results table in PostgreSQL database. Table definition should be as follows:

|  |  |
| --- | --- |
| **Field Name** | **Field Type** |
| run\_datetime | Timestamp |
| tpch\_query\_name | Text |
| benchmark\_result | Numeric |
|  |  |

***./tpch\_util.py --fetch-results***

Will fetch the results saved in results table and print it.