

Globant Data Engineer Challenge

Candidate: Luciano Martins dos Santos.

Development Read me

Date:10/04/2025

1- Objects:

- BigQueryAPI.py
- test_BigQueryAPI.py
- BigQuerySQL.txt
- Dockerfile
- requirements.txt
- hired_employees.csv
- jobs.csv
- departments.csv

2- Descriptions:

- **BigQueryAPI.py:** This script loads the files departments.csv, jobs.csv, and hired_employees.csv into pandas dataframes and makes some transformations. Then loads the dataframes into Google Big Query tables

The csv files are hosted in my GitHub public folder. The link is pointed in the dataframe creation method, so it can run anywhere.

Warning: I needed to create an IAM credential token as a credentials variable to connect to Big Query. It was not recommended to upload it to a public folder, so ask the evaluation team to create a credentials file in the Big Query Console path: IAM and administrator >> Service Account, download the JSON file, and place the file path in here:

```
# Credentials:
credentials = service_account.Credentials.from_service_account_file(
    'C:\\Users\\lucia\\Documents\\GBQ.json',
    scopes=['https://www.googleapis.com/auth/bigquery']
)
```

- **test_BigQueryAPI.py:** This script is the BigQueryAPI.py test case.

```
(base) PS C:\Users\lucia\Projetos> pytest test_BigQueryAPI.py
===== test session starts =====
platform win32 -- Python 3.12.7, pytest-7.4.4, pluggy-1.0.0
rootdir: C:\Users\lucia\Projetos
plugins: anyio-4.2.0
collected 4 items

test_BigQueryAPI.py ....

===== warnings summary =====
..\.anaconda3\Lib\site-packages\jupyter_client\connect.py:22
C:\Users\lucia\anaconda3\Lib\site-packages\jupyter_client\connect.py:22: DeprecationWarning: Jupyter is migrating its paths to use standard platformdirs
given by the platformdirs library. To remove this warning and
see the appropriate new directories, set the environment variable
`JUPYTER_PLATFORM_DIRS=1` and then run `jupyter --paths`.
The use of platformdirs will be the default in `jupyter_core` v6
  from jupyter_core.paths import jupyter_data_dir, jupyter_runtime_dir, secure_write
-- Docs: https://docs.pytest.org/en/stable/how-to/capture-warnings.html
===== 4 passed, 1 warning in 5.36s =====
(base) PS C:\Users\lucia\Projetos>
```

- **BigQuerySQL.txt:** This is an SQL script to be run in BigQuery. There are created two analytical views called vw_hired_employees_quarters and vw_hired_employees_count.
- **Dockerfile:** This file contains the script to build and run a Docker container. It runs both files, first test_BigQueryAPI.py and then BigQueryAPI.py

It runs the requirements.txt file as well.

- **requirements.txt:**
there are the dependencies to be used in the Dockerfile:
google-cloud-bigquery, pandas and pytest

After the deployment, please run the analytical views and in your BigQuery account:

department ▼	job ▼	Q1 ▼	Q2 ▼	Q3 ▼	Q4 ▼
Accounting	Account Representative IV	1	0	0	0
Accounting	Actuary	0	1	0	0
Accounting	Analyst Programmer	0	0	1	0
Accounting	Budget/Accounting Analyst III	0	1	0	0
Accounting	Cost Accountant	0	1	0	0
Accounting	Database Administrator III	0	0	0	1
Accounting	Desktop Support Technician	0	0	1	0
Accounting	Food Chemist	1	0	0	0

id ▼	department ▼	hired_count ▼	mean ▼
8	Support	221	139.17
5	Engineering	208	139.17
6	Human Resources	204	139.17
7	Services	204	139.17
4	Business Development	187	139.17
3	Research and Development	151	139.17
9	Marketing	143	139.17