

Read ME interview env

Overview

Environment for performing the task for the position data engineer at Orange Egypt.

Pre-requisites

1. Docker
2. Git bash or any unix based terminal

Overview of the environment

Docker Environment includes the following, specified in the compose file

1. Airflow (UI on Port 8088), username: `airflow`, password: `airflow`
2. PySpark (Notebook on Port 8888, UI for jobs 4040), container named: `pyspark`
any code you create should be under the dir `spark_code` on your host machine, which is already created for you.
driver directory has postgres driver allowing spark to connect to postgres.

Tip: check the mounted volumes in the compose file for pyspark.

3. DBT. Container with dbt installed and the connector to postgres. container_name: `dbt`
Make sure to initialise and create any dbt related code under `dbt_project` dir on host or `/app` in the container.

Tip: check the mounted volumes in the compose file for dbt.

4. Postgres - Your data warehouse container. Separate from the postgres container for airflow. container_name: `postgres_warehouse`
Check the compose file for the credentials and database name.

How to run

Open the terminal and move to the directory.

1. `chmod u+x ./init.sh`
2. Pull and run the environment - `./init.sh`, total size of the environment is 5~6 gb, so first time running this command you will need internet connection.

3. To access notebook and use PySpark, run this command in the terminal `docker logs --tail 20 pyspark`, look for a line that looks something like this -

To access the server, open this file in a browser:

```
file:///home/jovyan/.local/share/jupyter/runtime/jpserver-7-open.html
```

Or copy and paste one of these URLs:

```
http://7a3899fb9869:8888/lab?token=a745a8cde0eafd1b6125e8395d1d02d0c836e677cd6c417a
```

```
http://127.0.0.1:8888/lab?token=a745a8cde0eafd1b6125e8395d1d02d0c836e677cd6c417a
```

Copy and paste the link with 127.0.0.1:8888?token=...

4. To close the environment , make sure you are in the terminal and under the same directory then execute this command - `docker compose down`