Montando o ambiente Docker

1 - Clonar o projeto Git:

```
git clone https://github.com/cordon-thiago/airflow-spark
```

2 - Fazer um build da imagem:

```
cd airflow-spark/docker/docker-airflow
```

```
docker build --rm --force-rm -t docker-airflow-spark:1.10.7_3.1.2 .
```

3 – Verificar as imagens:

```
docker images
```

```
docker@engdados:~$ docker images
REPOSITORY                  TAG                 IMAGE ID            CREATED             SIZE
docker-airflow-spark        1.10.7_3.1.2        6a9db1c079ad        7 days ago          3.11GB
postgres                    9.6                 027ccf656dc1        10 months ago       200MB
bitnami/spark               3.1.2               eb5315fa7240        14 months ago       1.29GB
jupyter/pyspark-notebook    spark-3.1.2         b63f32deb3ba        14 months ago       3.22GB
docker@engdados:~$
```

4 – Acessar o banco postgres:

Host: localhost

User: airflow

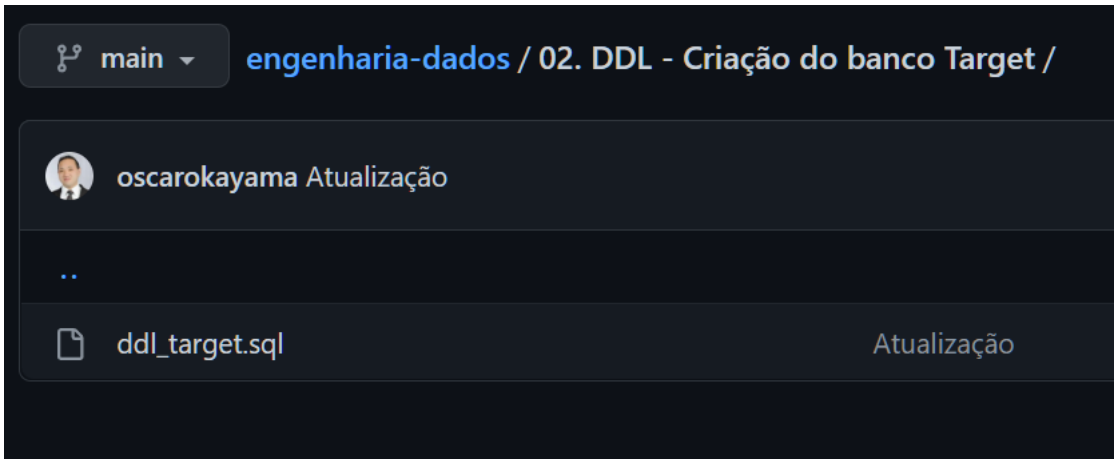
Password: airflow

```
CREATE DATABASE projeto OWNER airflow;
```

Criar o schema target no banco projeto:

```
CREATE SCHEMA target AUTHORIZATION airflow;
```

5 – Rodar o script ddl_target.sql:



Rodando o projeto de carga:

Acessar o Jupyter:

http://localhost:8888

Para pegar o token, executar o comando:

```
docker logs docker-jupyter-spark-1
```

```
[I 2022-12-27 19:54:37.112 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 19:54:37.121 NotebookApp] Serving notebooks from local directory: /home/jovyan
[I 19:54:37.124 NotebookApp] Jupyter Notebook 6.4.4 is running at:
[I 19:54:37.129 NotebookApp] http://129cc8dc9de6:8888/?token=0c14ab99331eba29089a67a141c5a9e40f5ec32aa434a07f
[I 19:54:37.129 NotebookApp]  or http://127.0.0.1:8888/?token=0c14ab99331eba29089a67a141c5a9e40f5ec32aa434a07f
[I 19:54:37.129 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation
[C 19:54:37.140 NotebookApp]

    To access the notebook, open this file in a browser:
        file:///home/jovyan/.local/share/jupyter/runtime/nbserver-6-open.html
    Or copy and paste one of these URLs:
        http://129cc8dc9de6:8888/?token=0c14ab99331eba29089a67a141c5a9e40f5ec32aa434a07f
     or http://127.0.0.1:8888/?token=0c14ab99331eba29089a67a141c5a9e40f5ec32aa434a07f
[I 20:18:06.286 NotebookApp] 302 GET / (192.168.99.1) 2.180000ms
/opt/conda/lib/python3.9/json/encoder.py:257: UserWarning: date_default is deprecated since jupyter_client 7.0.0. Us
  return _iterencode(o, 0)
Exception in callback <TaskWakeupMethWrapper object at 0x7fbafdfd15b0>(<Future finis...cdc"\r\n\r\n'>)
```

Acessar a url:

`http://127.0.0.1:8888/?token=0c14ab99331eba29089a67a141c5a9e40f5ec32aa434a07f`

Abrir o notebook disponibilizado no Github: