

Teste Técnico - Staff Data Reliability Engineer

Abaixo os erros encontrados e a solução para funcionamento do ambiente:

1 - Erro na inicialização devido a falta da variável AIRFLOW_UID, baseado na documentação, foi necessário criar um arquivo .env através do comando abaixo.

O valor da variável foi definida com UID do usuário, no caso utilizei o usuário ROOT, portanto o UID é 0.

Executado comando: **echo -e "AIRFLOW_UID=\$(id -u)" > .env**

2 - Foi necessário criar os diretórios logs, plugins e config para mapeá-los no container.

Estas configurações já estavam feitas no arquivo docker-compose.

```
AIRFLOW__API__AUTH__BACKENDS: 'airflow'
volumes:
  - ./dag:/opt/airflow/dags
  - ./logs:/opt/airflow/logs
  - ./plugins:/opt/airflow/plugins
USER: "5000"
```

Executado comando: **mkdir -p ./logs ./plugins ./config** o diretório **dags** existe no repositório git, portanto não foi necessário criá-lo.

```
root@ip-172-31-24-134:~/dre-3-test# ls -l
total 44
-rw-rw-r-- 1 root root 11357 Sep 17 21:56 LICENSE
-rw-rw-r-- 1 root root 10 Sep 17 21:56 README.md
drwxrwxr-x 2 root root 4096 Sep 17 22:20 config
drwxrwxr-x 3 root root 4096 Sep 19 22:34 dag
drwxrwxr-x 2 root root 4096 Sep 19 22:25 dags
-rw-rw-r-- 1 root root 4341 Sep 17 22:23 docker-compose.yml
drwxrwxr-x 5 root root 4096 Sep 19 22:48 logs
drwxrwxr-x 2 root root 4096 Sep 17 22:06 plugins
```

3 - Configuração do Postgres errada, sendo necessário adicionar na inicialização a criação da ROLE (usuário) chamado airflow.

- Criado script para criação da Role (config/init.sql)
 - **#!/bin/bash**
set -e
psql -v ON_ERROR_STOP=1 --username "\$POSTGRES_USER"
--dbname "\$POSTGRES_DB" <<-EOSQL
CREATE USER airflow;
CREATE DATABASE airflow;
GRANT ALL PRIVILEGES ON DATABASE airflow TO airflow;
EOSQL
- Configurado no serviço postgres no arquivo docker-compose.yml a inicialização do init.sql, para que a Role seja criada na inicialização do container.

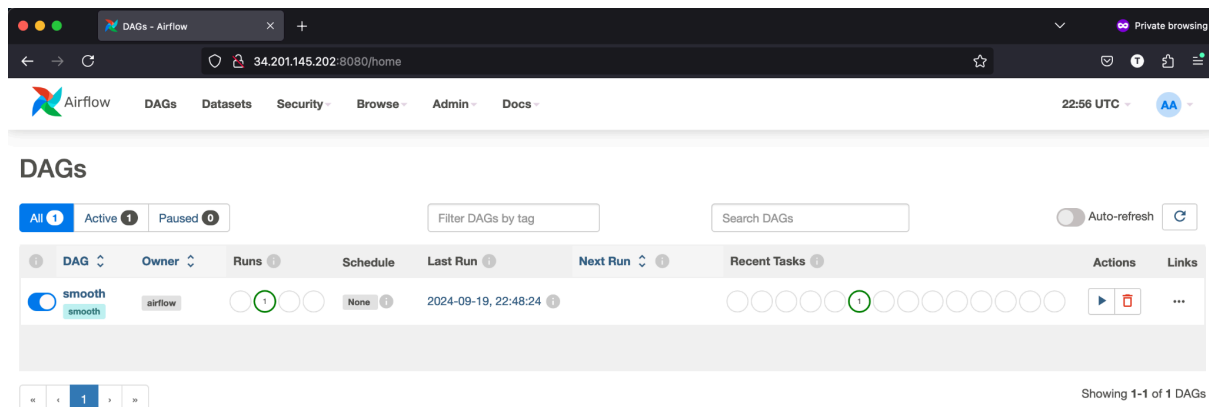
```
postgres:
  image: postgres:13
  ports:
    - 5432:5432
  environment:
    POSTGRES_USER: airflow
    POSTGRES_PASSWORD: airflow
    POSTGRES_DB: airflow
  volumes:
    - postgres-db-volume:/var/lib/postgresql/data
    - ./config/init.sql:/docker-entrypoint-initdb.d/init-database.sh
  healthcheck:
    test: ["CMD", "pg_isready", "-U", "airflow"]
    interval: 5s
    retries: 5
    restart: always
```

4 - Havia um erro no código python na declaração da função smooth, quando declarado a função smooth é necessário adicionar : ao final, conforme abaixo.

correção - linha 12 - def smooth ():

Conclusão:

Após os passos acima, o ambiente ficou funcional, podendo ser acessado via WEB e foi possível executar a DAG smooth.py.



Dificuldades encontradas:

Não tenho muito conhecimento em Airflow, portanto tive que ler a documentação para iniciar a configuração do ambiente.

Tive dificuldade em executar em meu computador, por falta de recursos, quando o ambiente era inicializado retornava muitos erros, sendo o erro mais latente o abaixo:

ERROR - Triggerer's async thread was blocked for 0.80 seconds, likely by a badly-written trigger. Set PYTHONASYNCIODEBUG=1 to get more information on overrunning coroutines.

Fiz buscas em fóruns técnicos e identifiquei que o problema era falta de recursos, portanto subi o ambiente na AWS em uma máquina em uma EC2 t3.xlarge (4 CPU com 16 GB de RAM), após essa alteração o ambiente subiu sem erros e consegui executar a DAG.