

INSTALLATION DU AIRFLOW

Nous l'avons installé à partir de docker :

- commandes d'installation et démarrage des services

```
export AIRFLOW_HOME=~/.airflow
# install from pypi using pip
pip install apache-airflow
# initialize the database
airflow initdb
# start the web server, default port is 8080
airflow webserver -p 8080
# start the scheduler
airflow scheduler
```

- création du user

```
(base) brice@wambdevps:~$ airflow users create --username admin --firstname admin --lastname admin --role Admin --email admin@domain.com
[2022-11-05 05:53:36,366] {manager.py:824} WARNING - No user yet created, use flask fab command to do it.
Password:
Repeat for confirmation:
[2022-11-05 05:55:25,757] {manager.py:212} INFO - Added user admin
User "admin" created with role "Admin"
(base) brice@wambdevps:~$
```

- l'interface Web Apache-Airflow à l'aide de l'URL (<http://localhost:8080/home#>) et ainsi qu'observation d'une liste de DAG

localhost:8080/home#

Airflow DAGs Datasets Security Browse Admin Docs 05:09 UTC AA

Do not use SQLite as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. [Click here](#) for more information.

Do not use SequentialExecutor in production. [Click here](#) for more information.

DAGs

All 40 Active 0 Paused 40

Filter DAGs by tag Search DAGs

Auto-refresh

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks
dataset_consumes_1 consumes dataset-scheduled	airflow	0 of 1 datasets updated				
dataset_consumes_1_and_2 consumes dataset-scheduled	airflow	0 of 2 datasets updated				
dataset_consumes_1_never_scheduled consumes dataset-scheduled	airflow	0 of 2 datasets updated				
dataset_consumes_unknown_never_scheduled dataset-scheduled	airflow	0 of 2 datasets updated				
dataset_produces_1						

- Création de notre premier DAG

```
dag_1.py 3, U X
home > brice > docker-airflow > dags > dag_1.py > ...
1 #step 1
2
3 from airflow import DAG
4 from datetime import timedelta, timedelta
5 from airflow.operators.dummy_operator import DummyOperator
6
7 #step 2
8
9 default_args = {
10     'owner': 'airflow',
11     'depends_on_past': False,
12     'started_date': datetime(2022, 5, 11),
13     'retries': 0
14 }
15
16 #step 3
17
18 dag=DAG(dag_id='DAG-1',default_args=default_args,catchup=False,scheduler_interval='@once')
19
20 #step 4
21
22 start = DummyOperator(task_id='start',dag=dag)
23 end= DummyOperator(task_id='end',dag=dag)
24
25 #step 5
26
27 start >> end
```