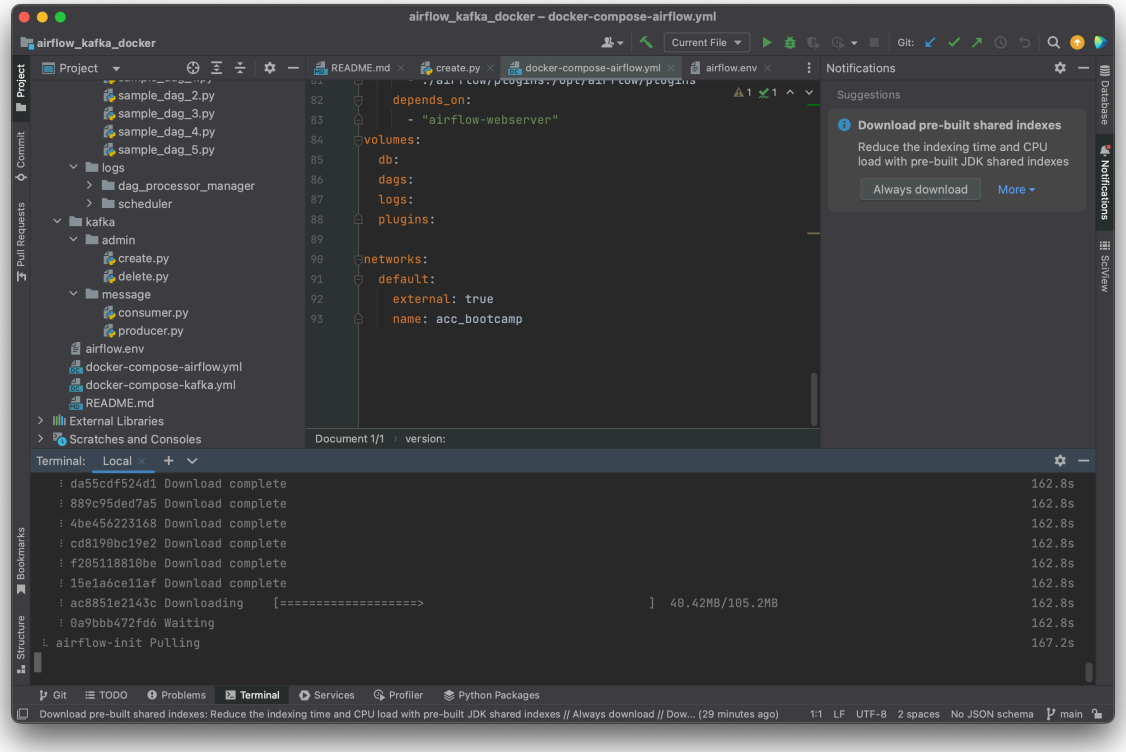
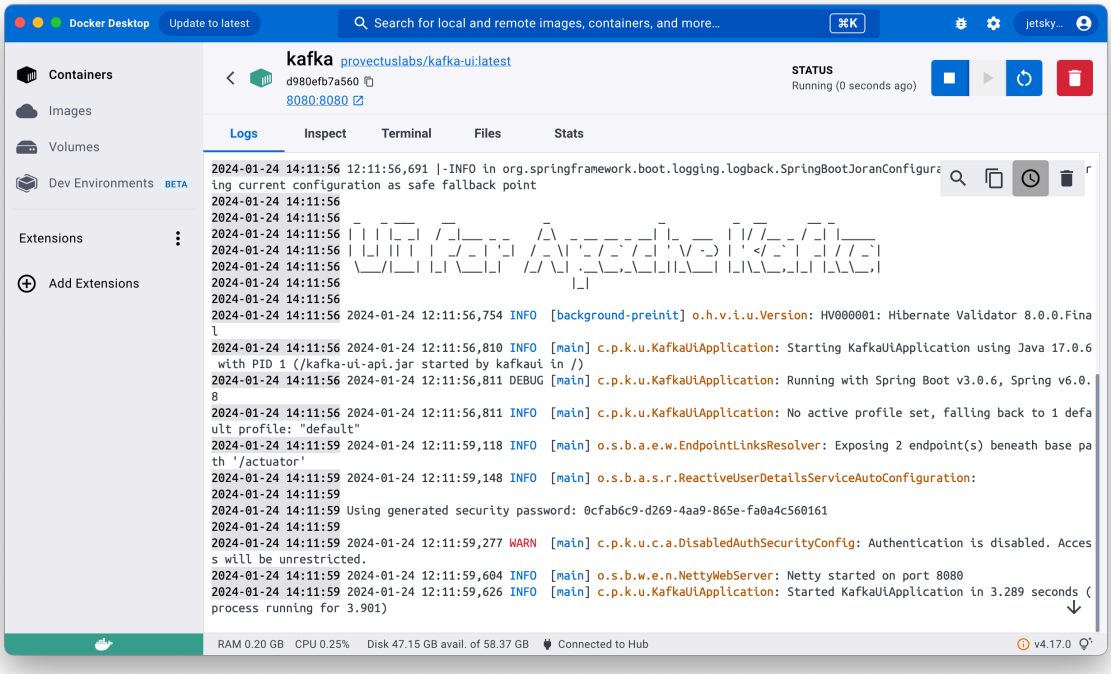
Day 13 – Airflow – Accenture bootcamp

Name: Pothumulla Kankanamge Mewan Madhusha

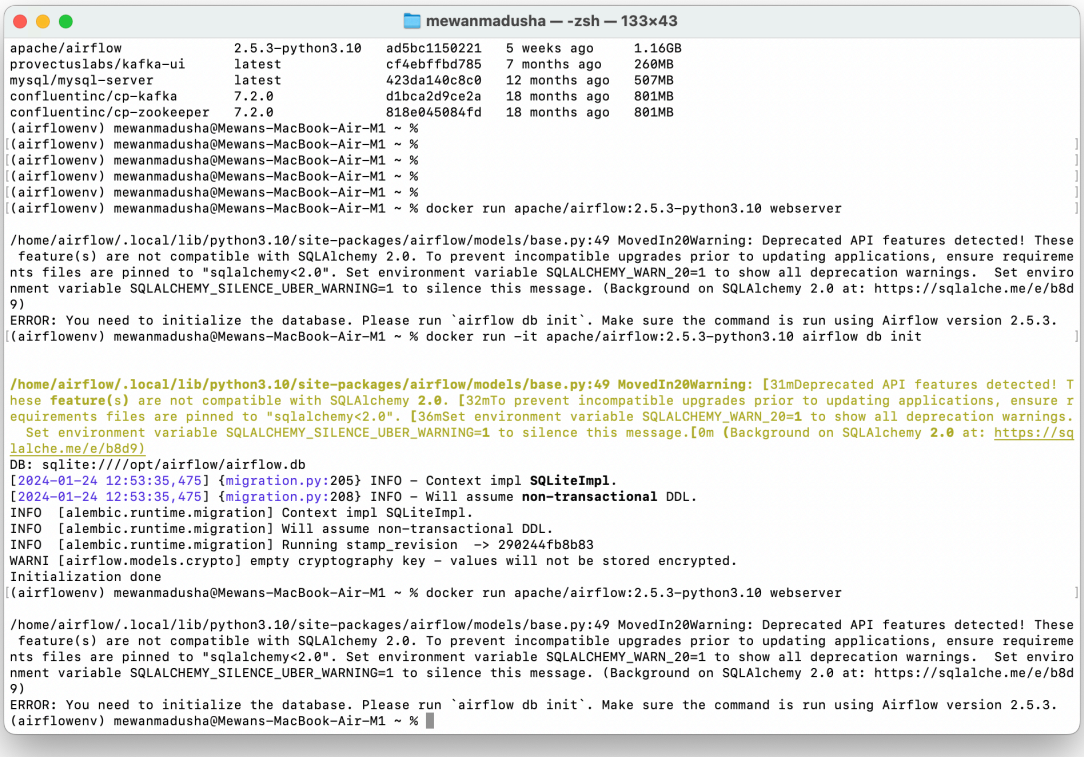
Clone and compose given docker repository

```
docker-compose -f docker-compose-airflow.yml -f docker-compose-kafka.yml up
```





```
docker run apache/airflow:2.5.3-python3.10 webserver
```



Since provided docker file airflow version is not working, tried with official Apache/airflow image

https://hub.docker.com/r/apache/airflow

This method also didn't work.

Next install airflow standalone with pip install.

```
AIRFLOW_VERSION=2.8.1

# Extract the version of Python you have installed. If you're currently using a Python version that is not supported by Airflow, you may want to set this manually.
# See above for supported versions.
PYTHON_VERSION="$(python --version | cut -d " " -f 2 | cut -d "." -f 1-2)"

CONSTRAINT_URL="https://raw.githubusercontent.com/apache/airflow/constraints-${AIRFLOW_VERSION}/constraints-${PYTHON_VERSION}.txt"
```

```
# For example this would install 2.8.1 with python 3.8:
https://raw.githubusercontent.com/apache/airflow/constraints-
2.8.1/constraints-3.8.txt

pip install "apache-airflow==${AIRFLOW_VERSION}" --constraint
"${CONSTRAINT_URL}"
```

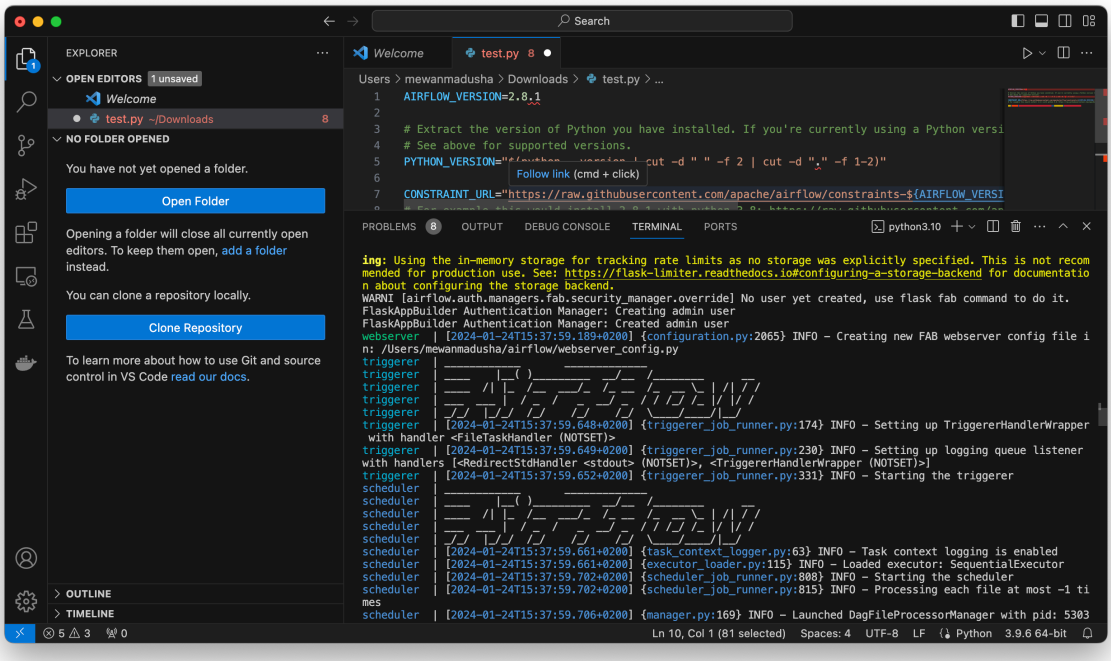
Start webserver.

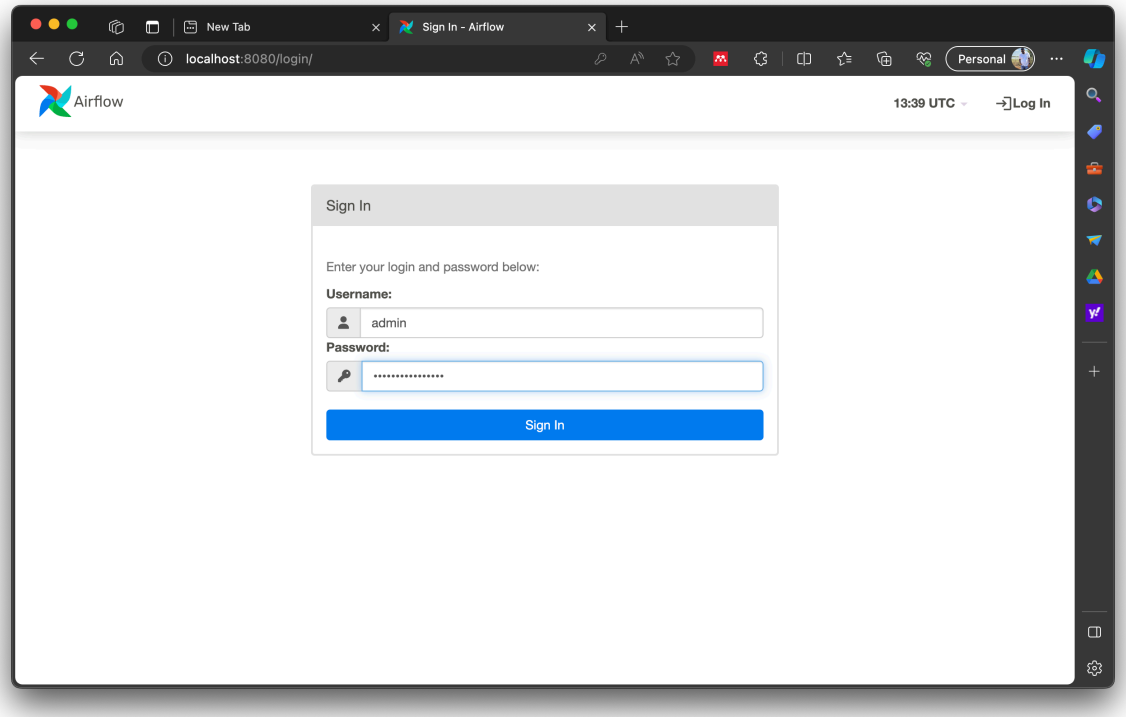
Run all at once.

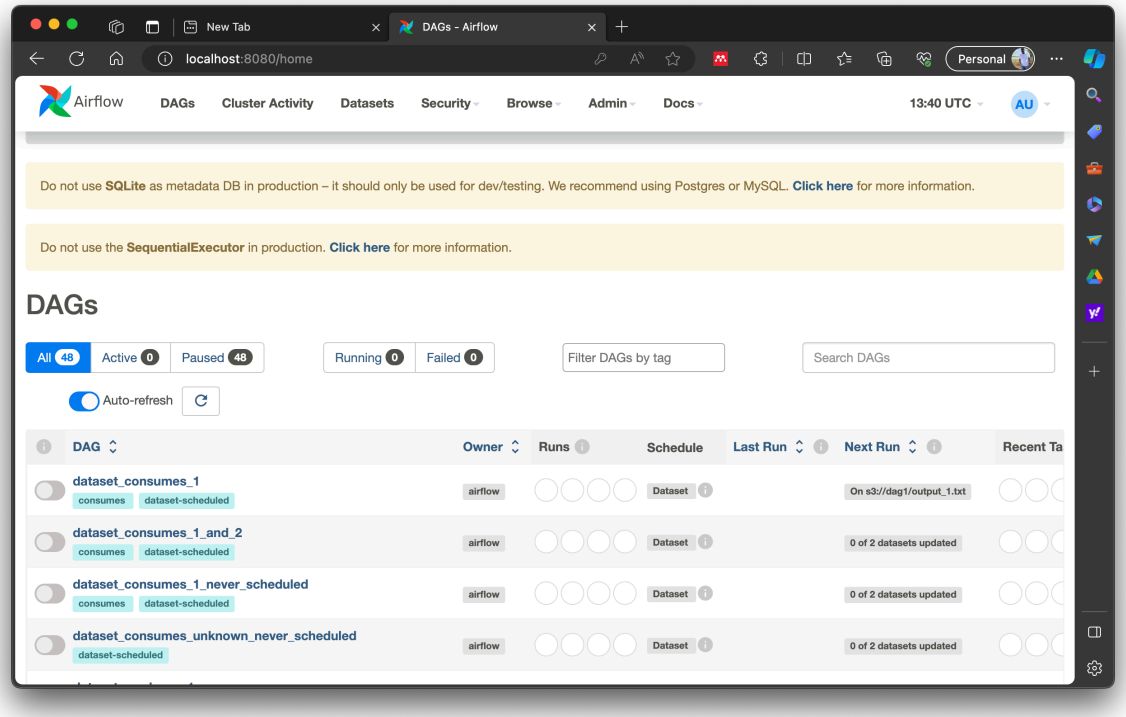
```
airflow standalone
```

Or run each functionality separately.

```
airflow db migrate

airflow users create \
    --username admin \
    --firstname Peter \
    --lastname Parker \
    --role Admin \
    --email spiderman@superhero.org

airflow webserver --port 8080

airflow scheduler
```



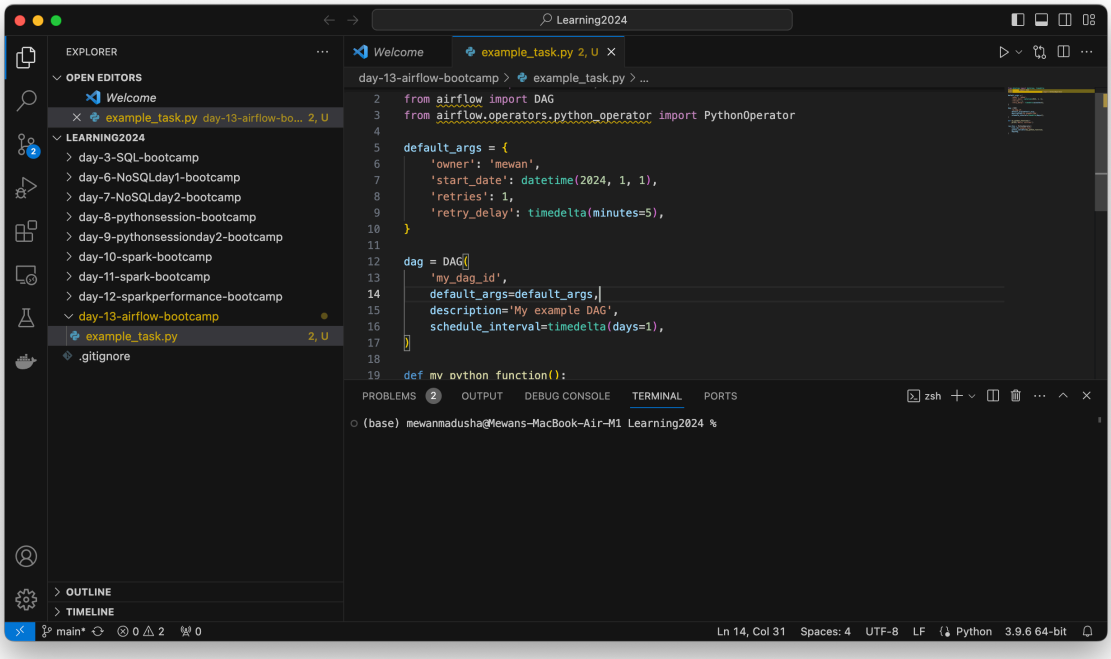

New Tab
Sign In - Airflow
localhost:8080/login/
Personal
Airflow
13:39 UTC
Log In
Sign In
Enter your login and password below:
Username:
admin
Password:
Sign In


New Tab
DAGs - Airflow
localhost:8080/home
Personal
Airflow
DAGs
Cluster Activity
Datasets
Security
Browse
Admin
Docs
13:40 UTC
AU
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 the SequentialExecutor in production. Click here for more information.
DAGs
All 48
Active 0
Paused 48
Running 0
Failed 0
Filter DAGs by tag
Search DAGs
Auto-refresh
DAG
Owner
Runs
Schedule
Last Run
Next Run
Recent Ta
dataset_consumes_1
consumes
dataset-scheduled
airflow
Dataset
On s3://dag1/output_1.txt
dataset_consumes_1_and_2
consumes
dataset-scheduled
airflow
Dataset
0 of 2 datasets updated
dataset_consumes_1_never_scheduled
consumes
dataset-scheduled
airflow
Dataset
0 of 2 datasets updated
dataset_consumes_unknown_never_scheduled
dataset-scheduled
airflow
Dataset
0 of 2 datasets updated

Create python file (you can do it into dag file as well but its good to process external py file with its path!)

Create dag as attached file

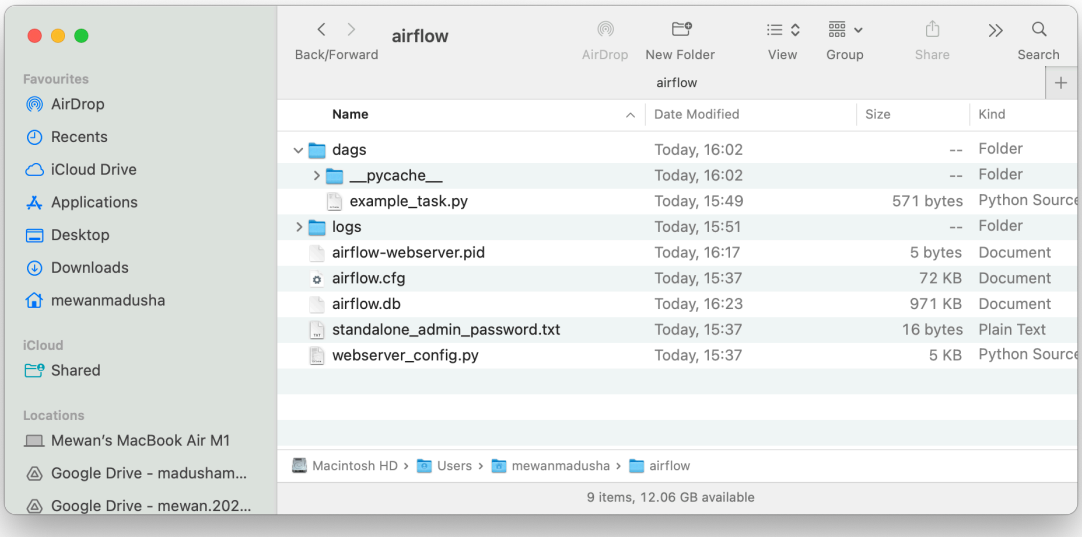
Create external python file.



Put into Dag folder

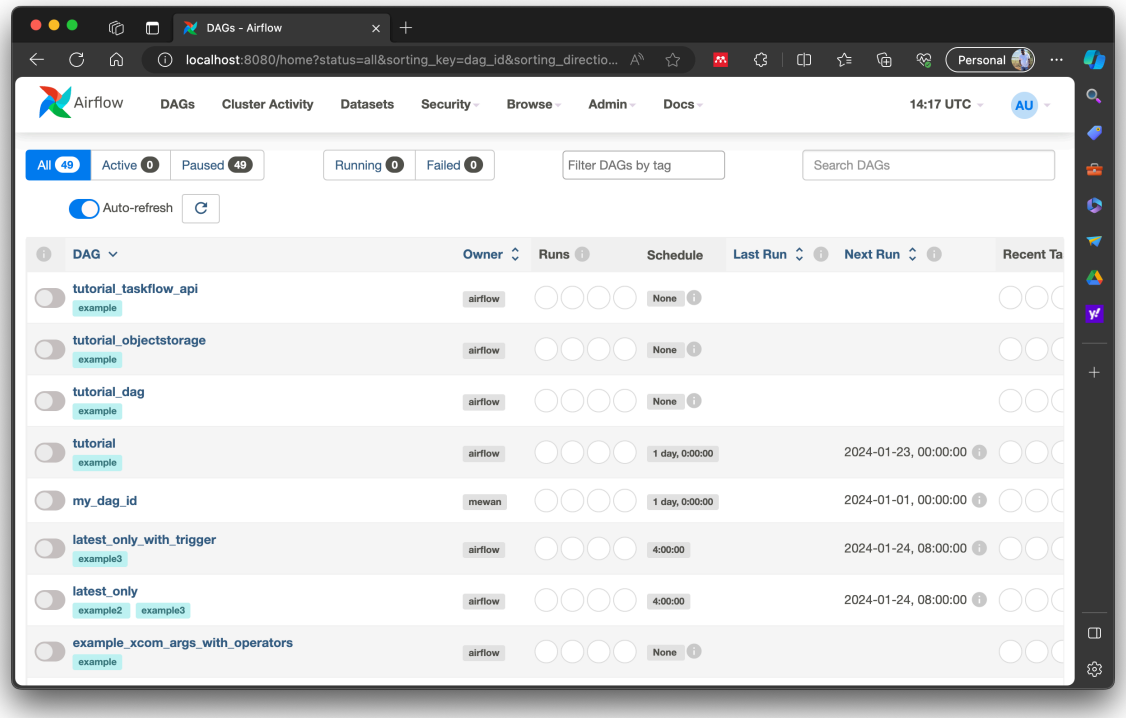
Past it into the dag folder

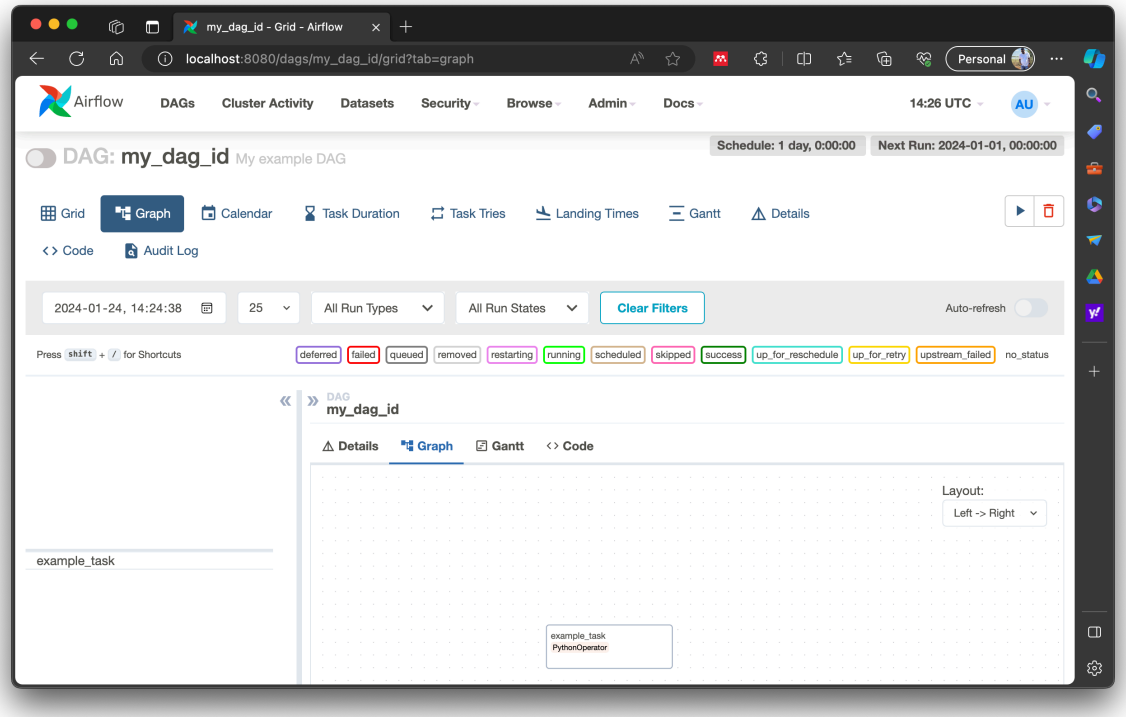
```
~/airflow/dags
```

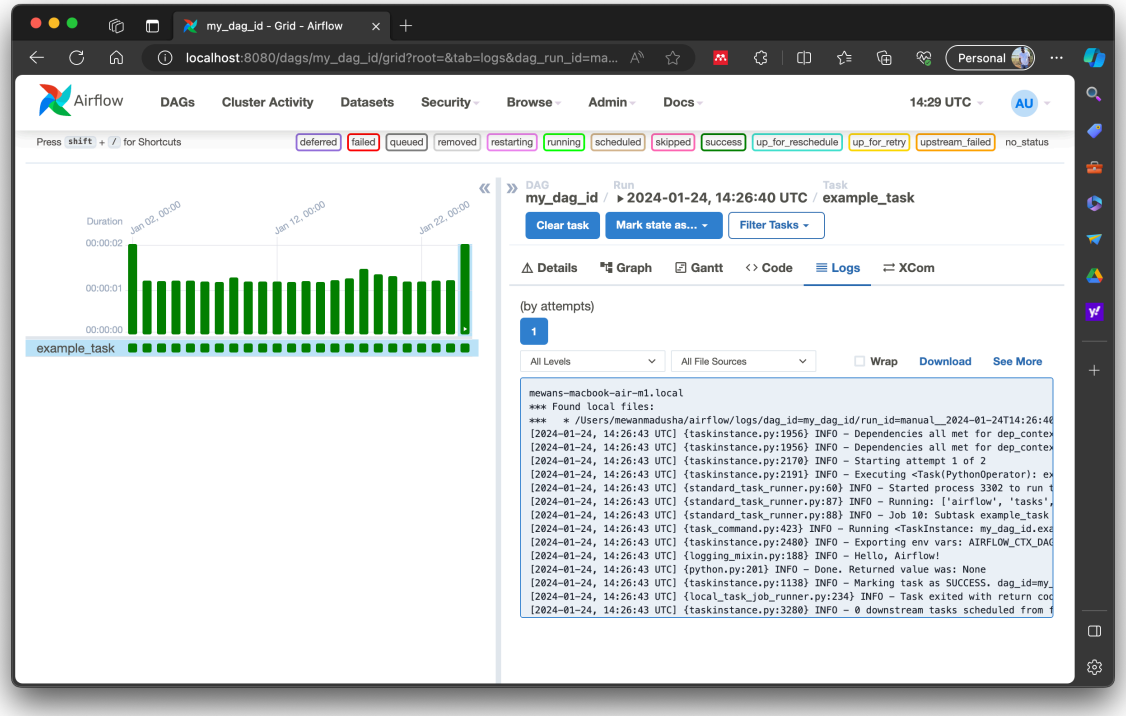


Couple of minutes later you can see on UI.

"my_dag_id" is external file







Print “Hello Airflow”