Name: Meghana G
USN: 1RVU23CSE264

Ex No: 6 Date: 15/10/25	Apache Airflow

Objective:

Apache Airflow is an open-source platform created by Airbnb (now part of the Apache Software Foundation) for programmatically authoring, scheduling, and monitoring workflows. It is widely used in data engineering and ETL (Extract, Transform, Load) pipelines

Outcomes:

1. Connecting to Airflow using Docker compose.
2. Installing Apache Airflow.
3. Running the DAGs(postgres) in Airflow.

Materials:

- ENV flies (`.env`)
- Csv file(users.csv)
- Yaml source file (`docker-compose`) and dockerfile.
- Python source file like dagselt_pipeline_postgres

Lab Procedure:

Stage 1: Install Apache Airflow for windows PC

1. Open the command prompt and navigate to your lab folder (cd "c:\Users\< >\lab6)
2. Install the docker and docker compose by commands :

sudo apt-get update

sudo apt-get install docker.io docker-compose –y

3. This builds a fresh image using your `Dockerfile`.

docker build -t airflowsqlserver -f Dockerfile --no-cache .

docker-compose up

Name: Meghana G
USN: 1RVU23CSE264

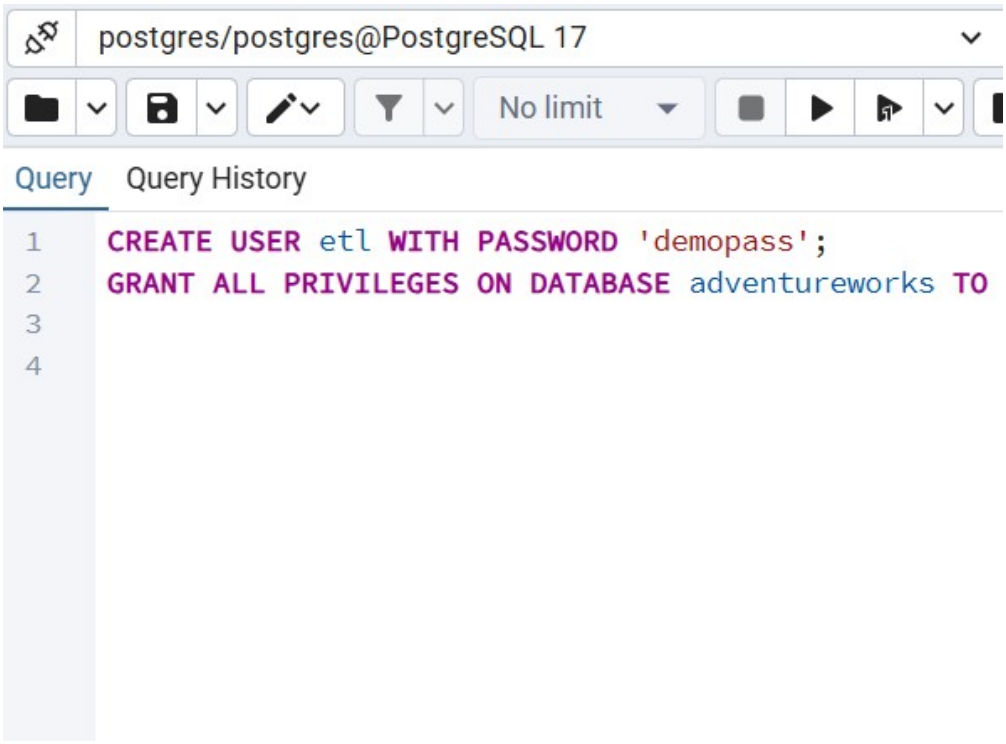
docker ps

- You should see containers for Airflow, Postgres, etc.
- Open a browser → Go to `http://localhost:8080` (usually Airflow UI).

4. Add this schema to pgadmin4 then save and run

   CREATE SCHEMA IF NOT EXISTS etl_staging;

   GRANT ALL PRIVILEGES ON SCHEMA etl_staging TO etl;



Stage 2: Access to the Airflow

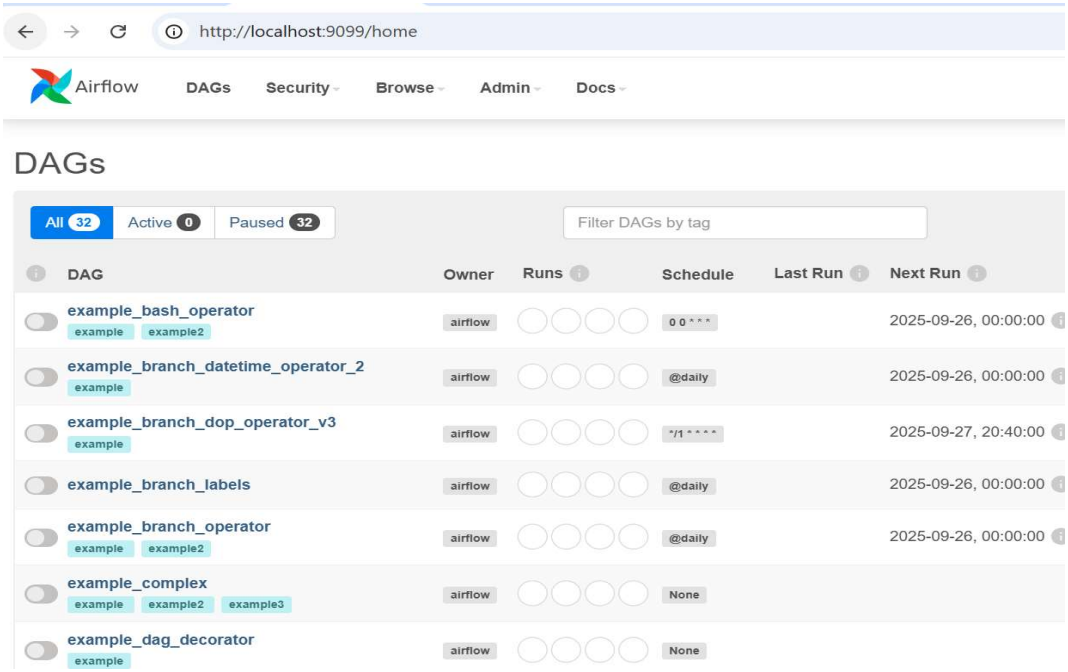
1. Open your browser:
   http://localhost:8080
2. Default Airflow login (unless changed in `.env`):

   username: airflow

   password: airflow

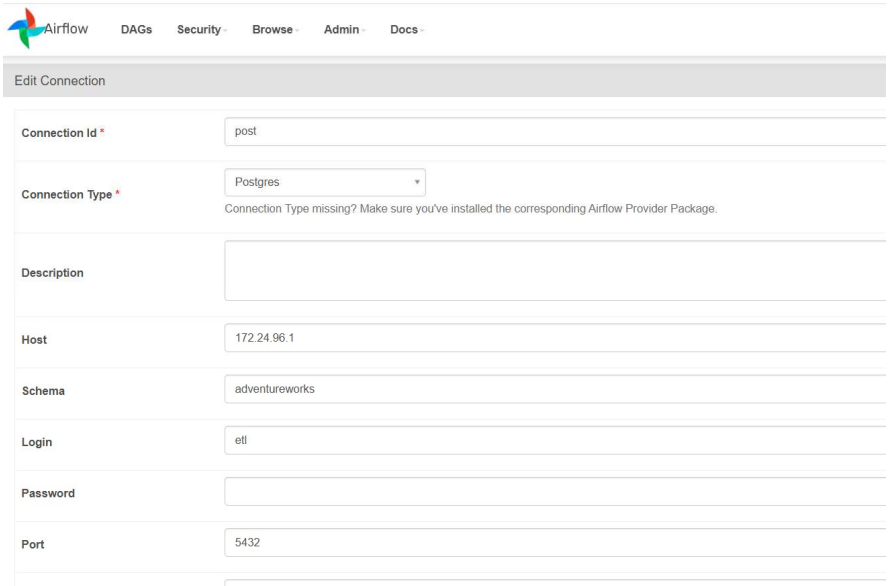
3. Verify the DAGs

Name: Meghana G
USN: 1RVU23CSE264



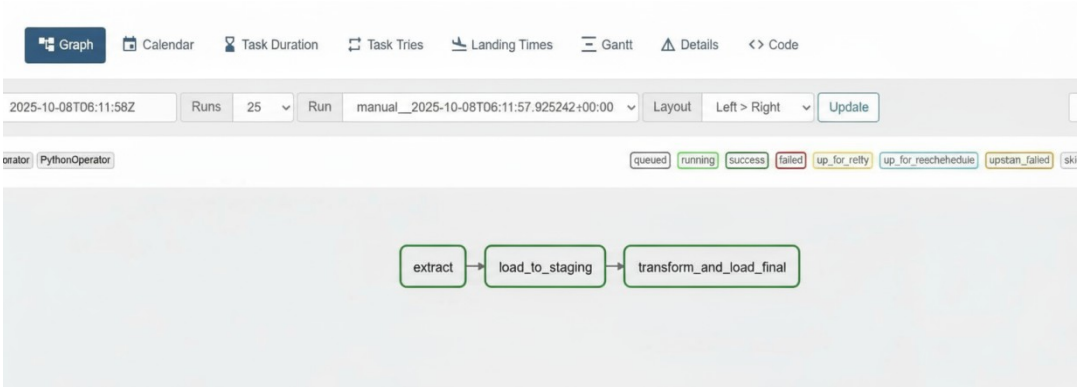
- In the Airflow UI, go to the **DAGs** tab.
- You should see your pipeline DAG from `dagselt_pipeline_postgres.py`.
- Turn the toggle **ON** and trigger it.

4. Go to Edit connection and give host as your pc ip address , login and password save it



5. Run the Dag

Name: Meghana G
USN: 1RVU23CSE264



It shows all green which means all stages run successfully.

Github link : https://github.com/meghana1653/Data-Engineering/blob/main/Lab-6.zip