$sudo apt install python3-pip

$pip3 install apache-airflow

$sudo apt-get install postgresql postgresql-contrib

$sudo -u postgres psql

CREATE DATABASE airflow;

CREATE USER airflow WITH ENCRYPTED PASSWORD 'airflow';

GRANT ALL PRIVILEGES ON DATABASE airflow to airflow;

\du ->To check if the user was created.

$pip3 install apache-airflow[postgres]

$pip3 install psycopg2

$sudo nano /etc/postgresql/12/main/pg\_hba.conf

Comment IPV4 address line in this file

host all all 0.0.0.0/0 trust

$sudo service postgresql restart

$sudo nano /etc/postgresql/12/main/postgresql.conf

listen\_addresses=’\*’

Open /home/hadoop/airlfow/airflow.cfg

sql\_alchemy\_conn = postgresql+psycopg2://airflow:airflow@localhost/airflow

Create a directory called dags in /home/hadoop/airflow

$airflow initdb

$airflow webserver -p 8080

$airflow scheduler

airflow users create \

--username admin \

--firstname test \

--lastname test \

--role Admin \

--email admin@example.org

Here is a couple of options you can use for your schedule\_interval. You can choose to use some preset argument or cron-like argument:

|  |  |  |
| --- | --- | --- |
| **preset** | **meaning** | **cron** |
| None | Don’t schedule, use for exclusively “externally  triggered” DAGs |  |
| @once | Schedule once and only once |  |
| @hourly | Run once an hour at the beginning of the hour | 0 \* \* \* \* |
| @daily | Run once a day at midnight | 0 0 \* \* \* |
| @weekly | Run once a week at midnight on Sunday morning | 0 0 \* \* 0 |
| @monthly | Run once a month at midnight of the first day  of the month | 0 0 1 \* \* |
| @yearly | Run once a year at midnight of January 1 | 0 0 1 1 \* |

**Example usage**:

* Daily schedule:
  + schedule\_interval='@daily'
  + schedule\_interval='0 0 \* \* \*'