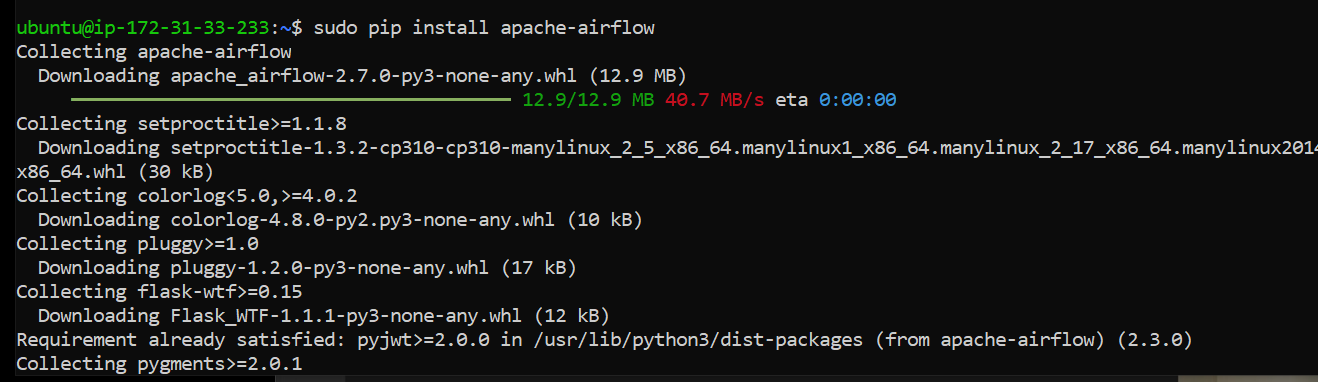
Step 1: Connect to any API and fetch the data

Step 2: Convert that data to json format

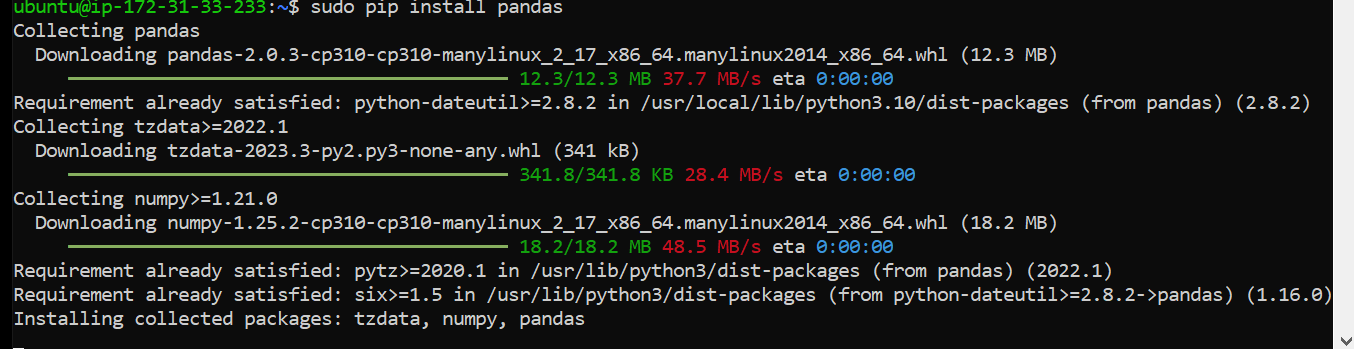
Step 3: Launch ec2 instance with ubuntu as OS



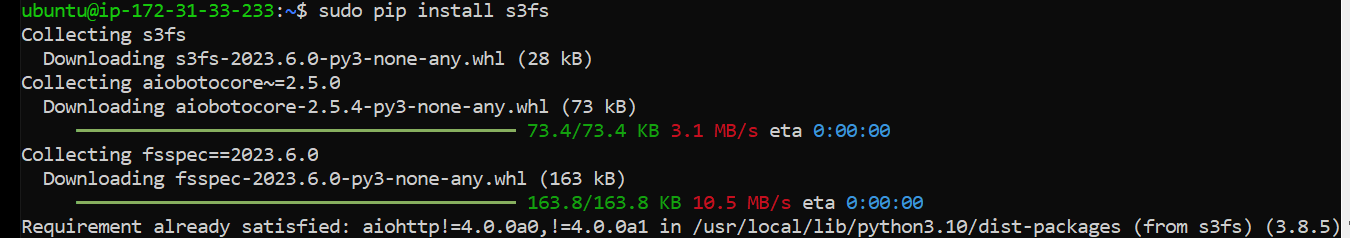
Step 4:install python and airflow



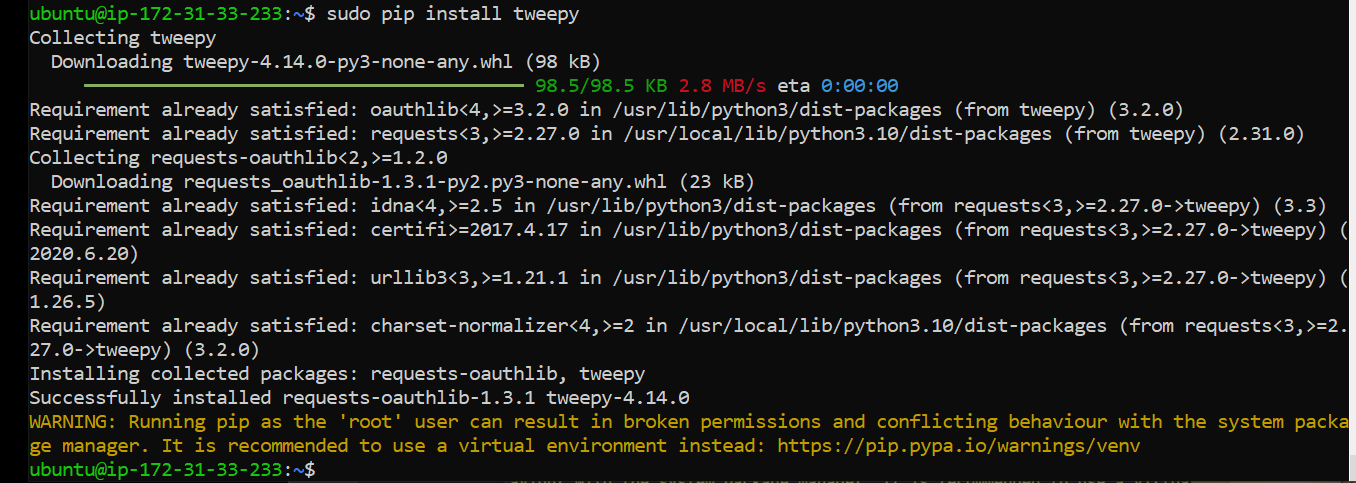
Step 5: install pandas



Step 6: install s3fs for amazon bucketing

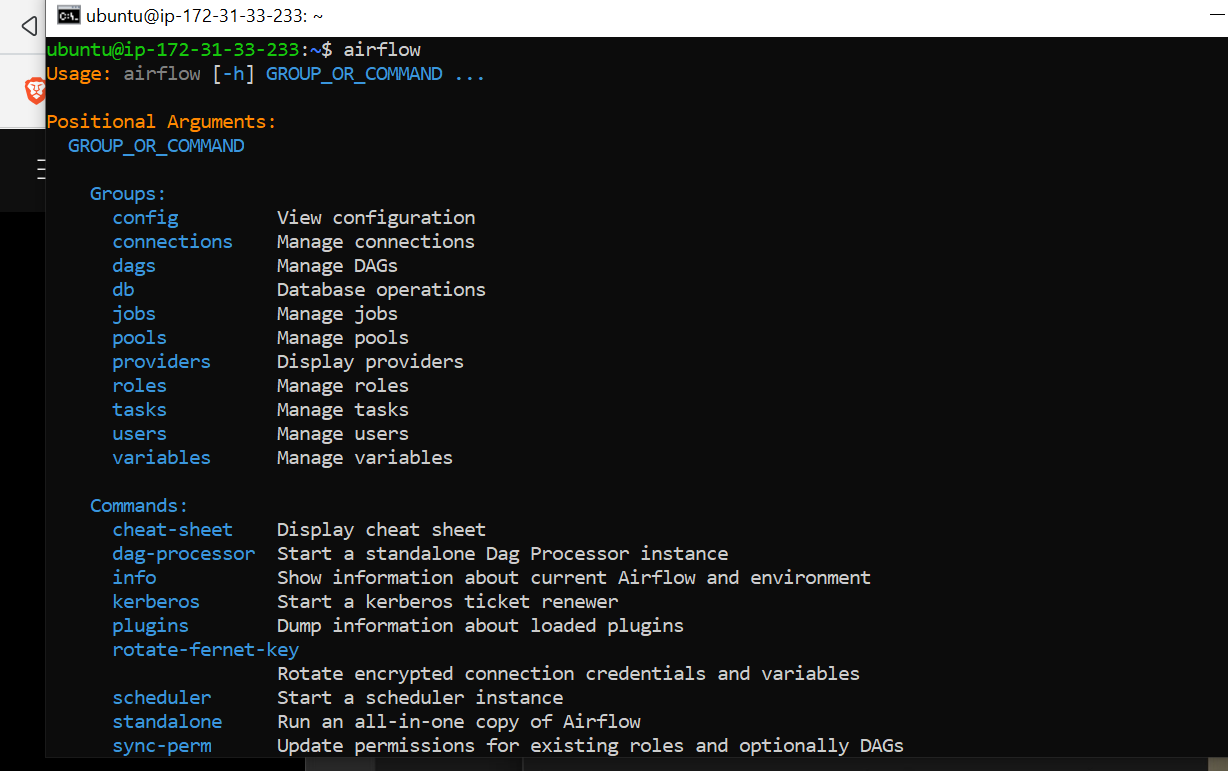


Step 7:

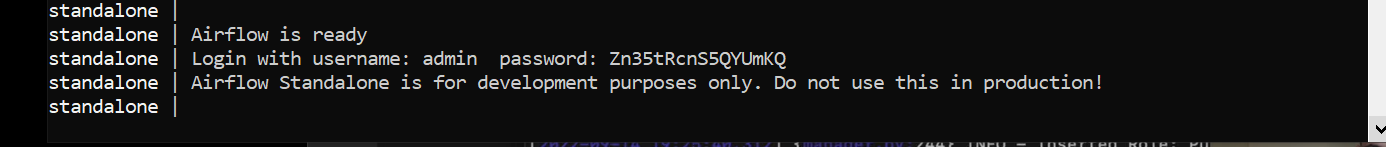


Step 8:

See the system



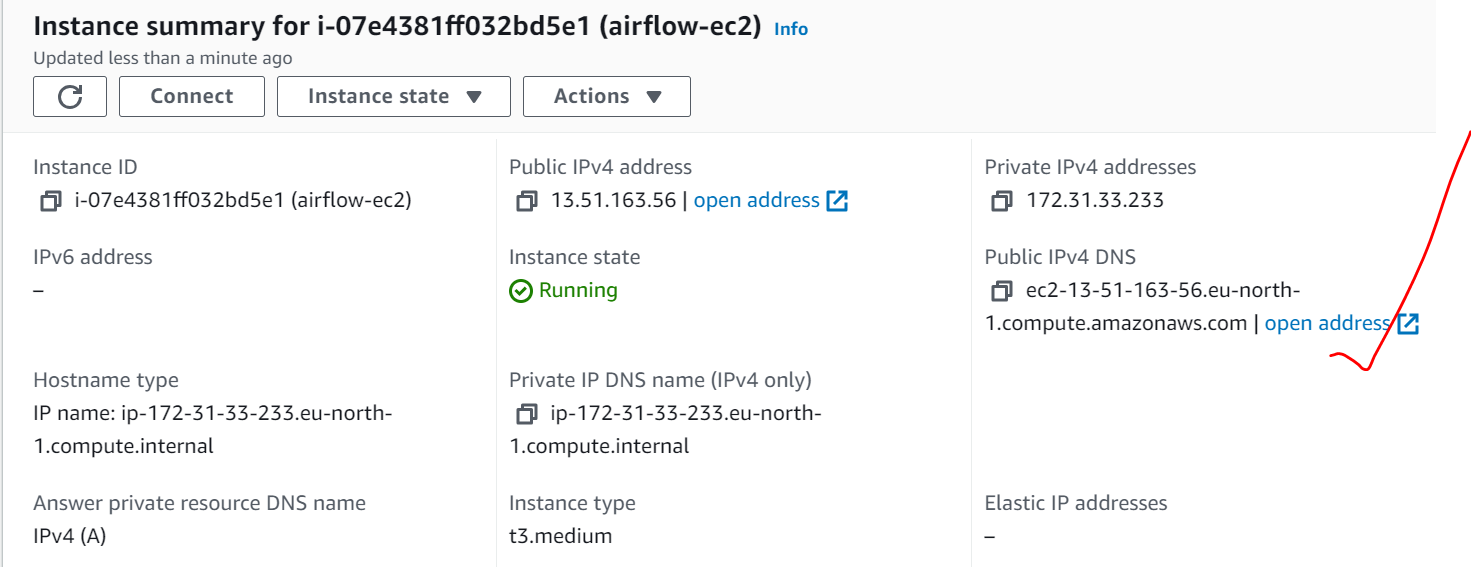
Run the Airflow standalone



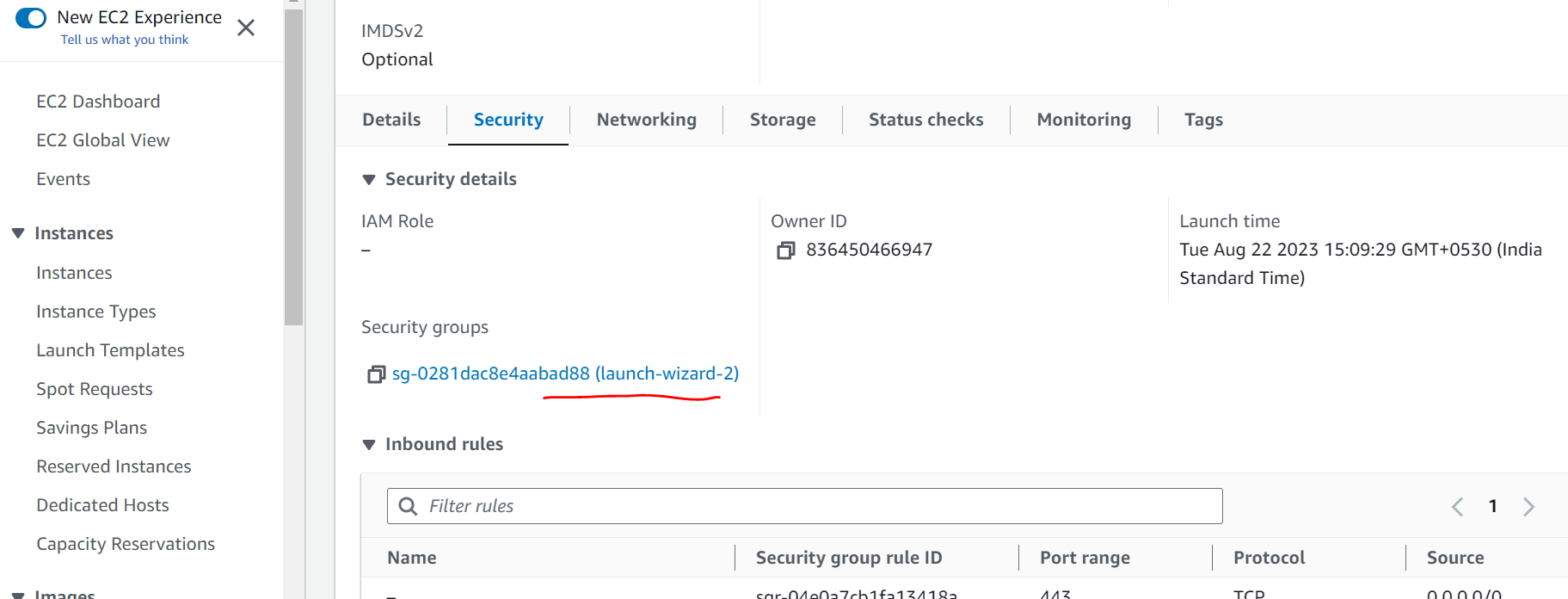
Login with username: admin

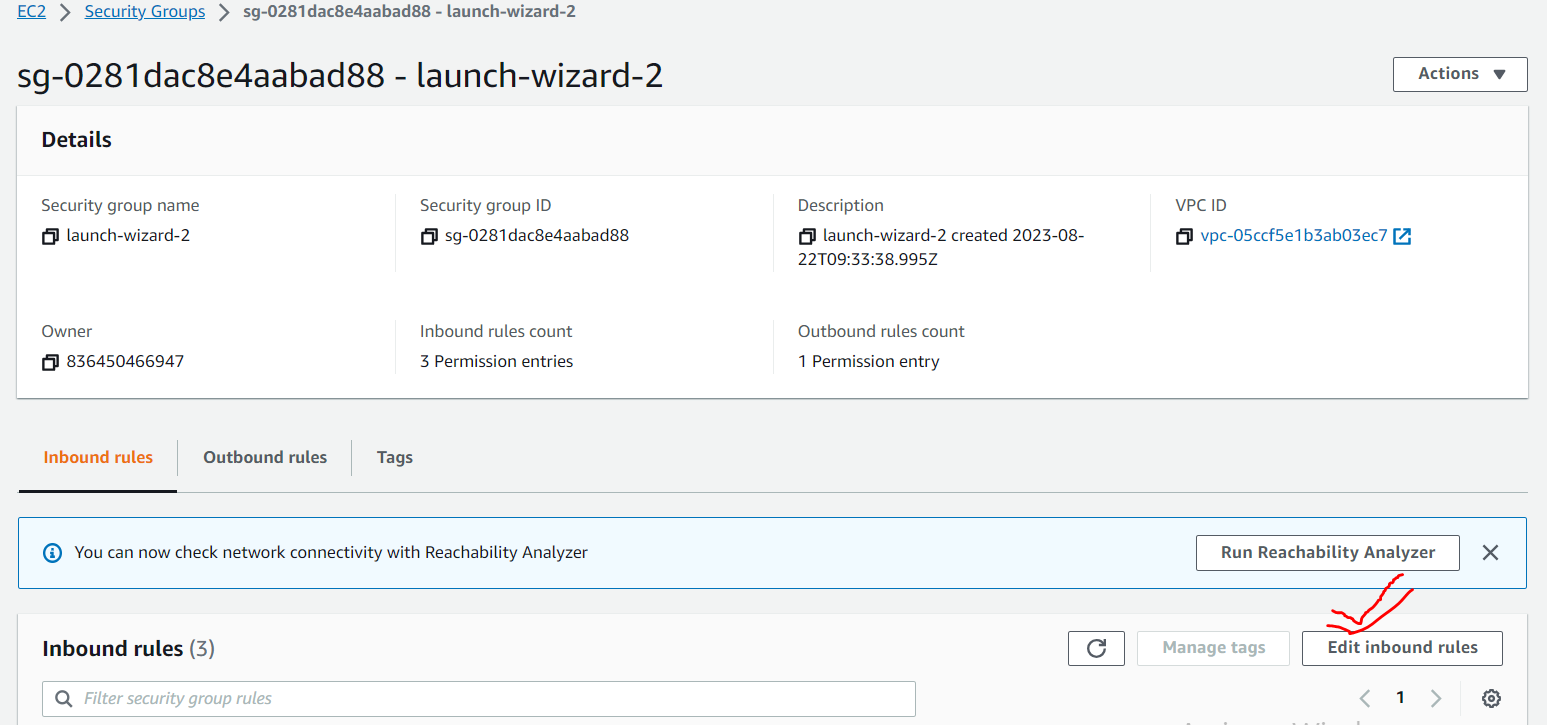
password: Zn35tRcnS5QYUmKQ

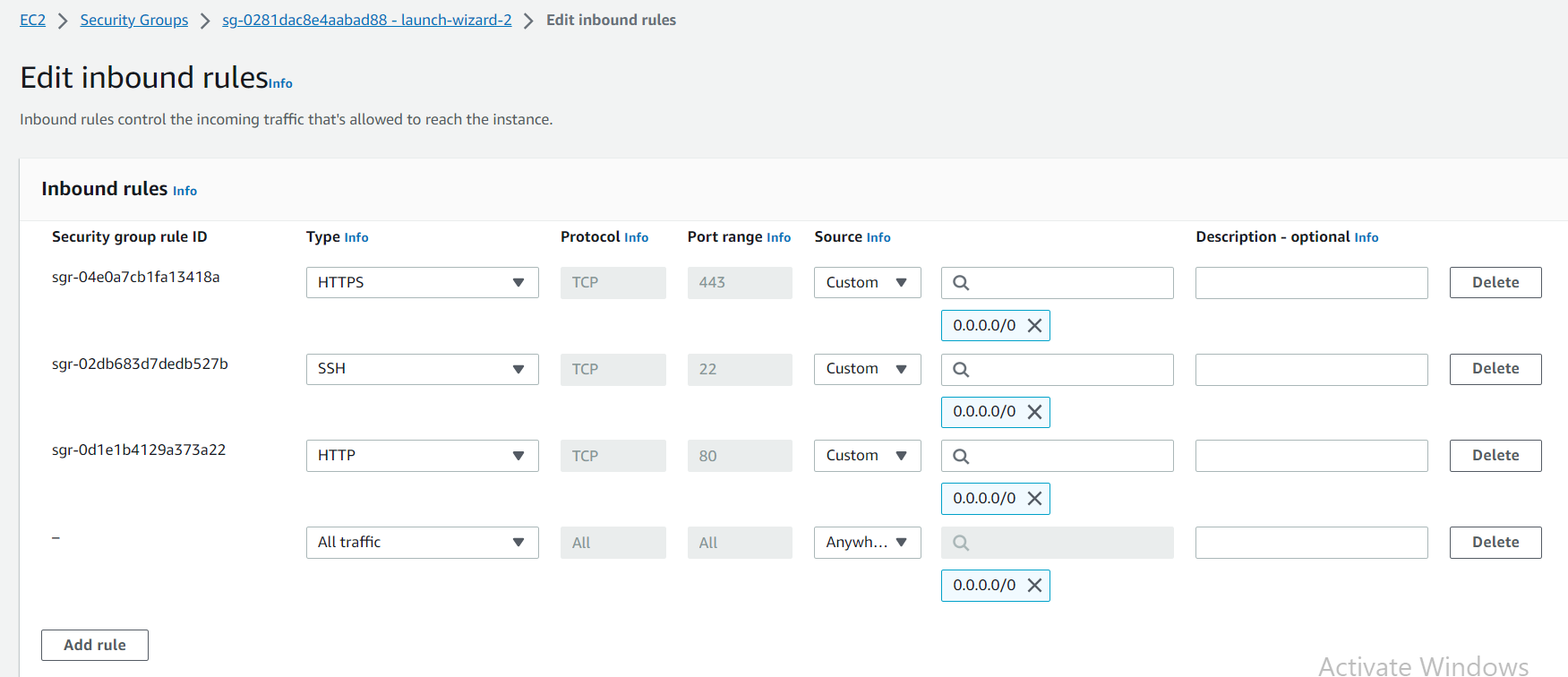
Step 9: Open the instance inside the chrome



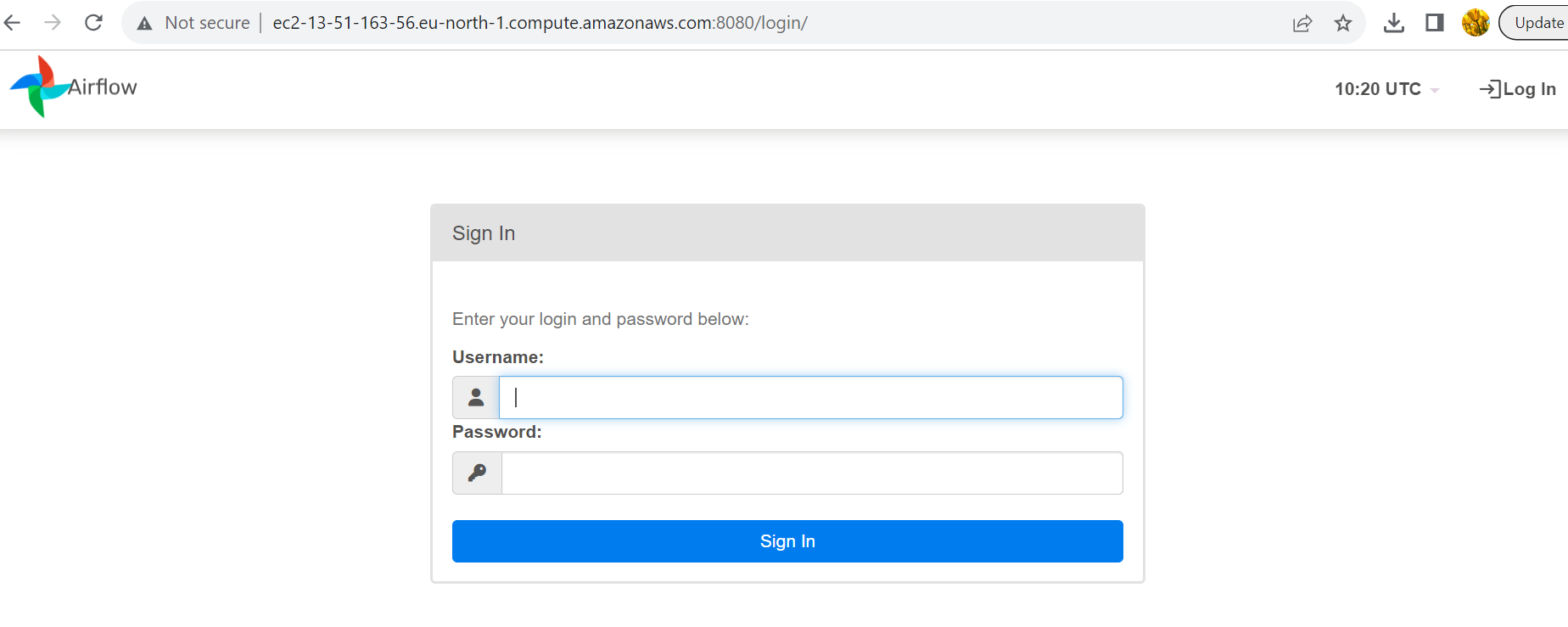
Step 10: Change the configuration

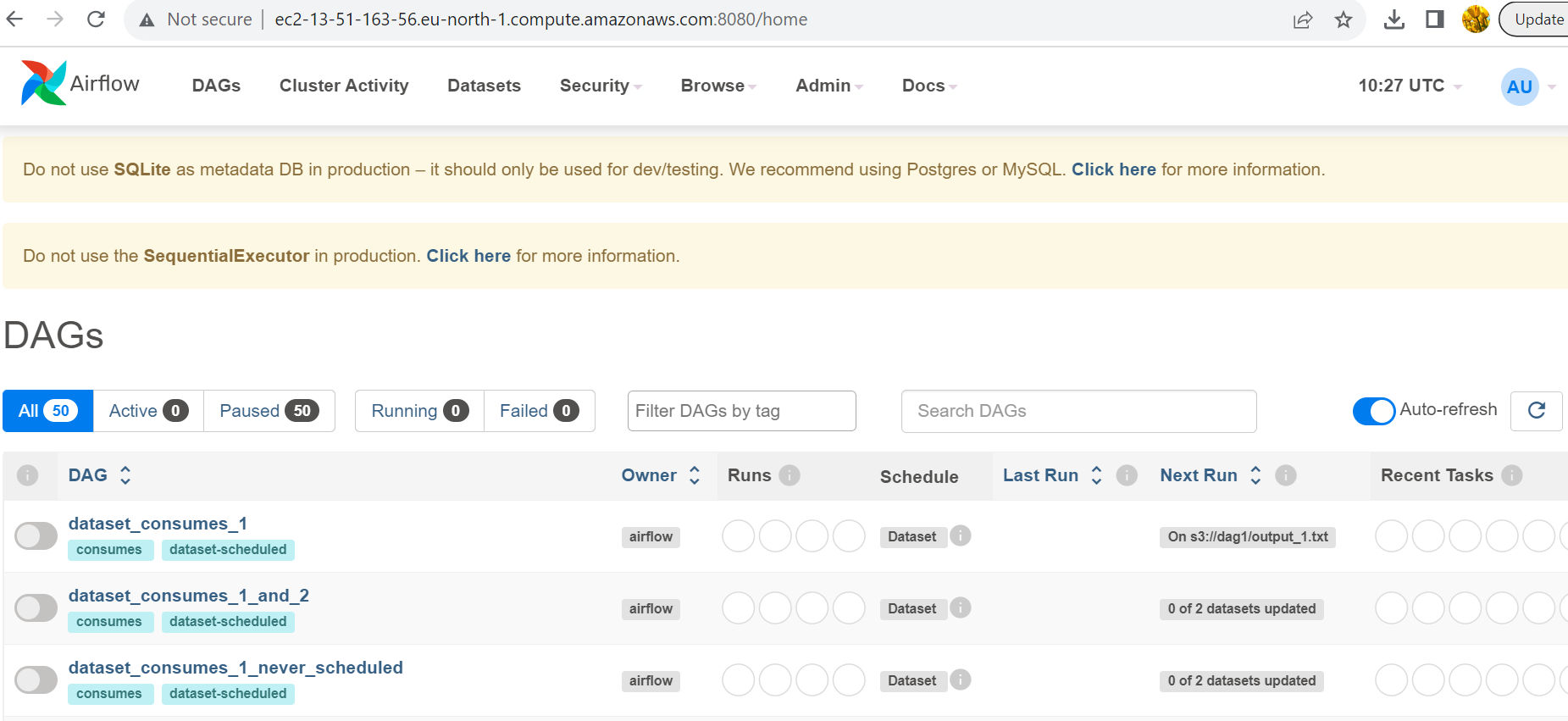






**Now run the airflow:**





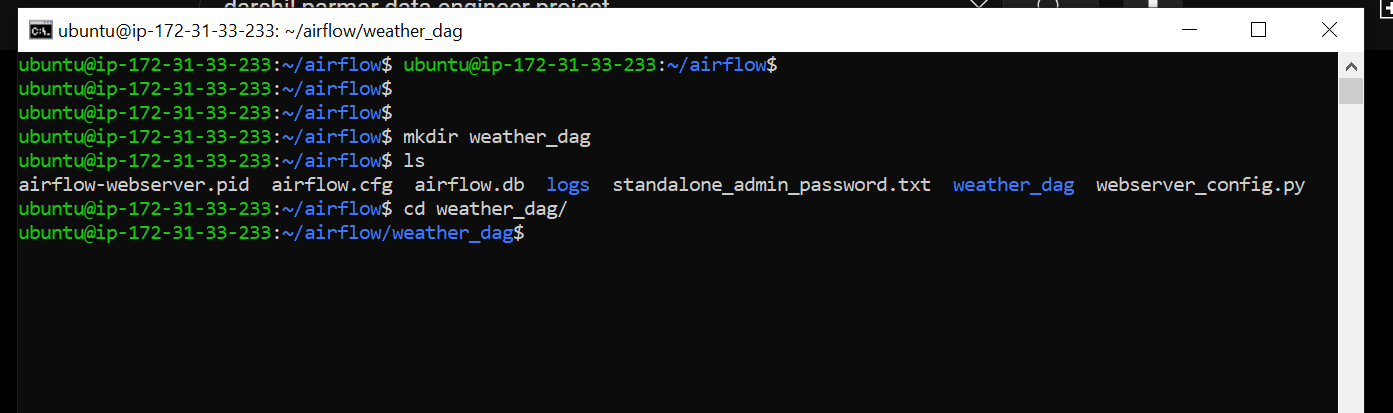
Step: create two py files inside vscode

Weather\_etl.py

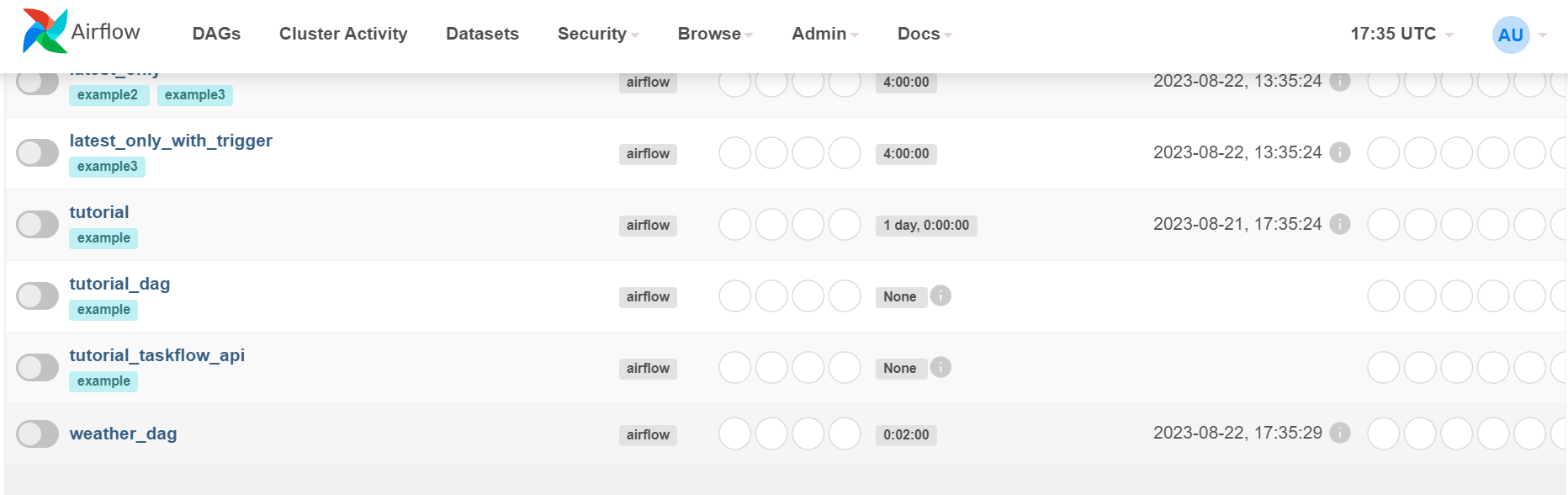
Weather\_dag.py

Now got to ec2 and change the config file path

Step: next make a folder and save both files

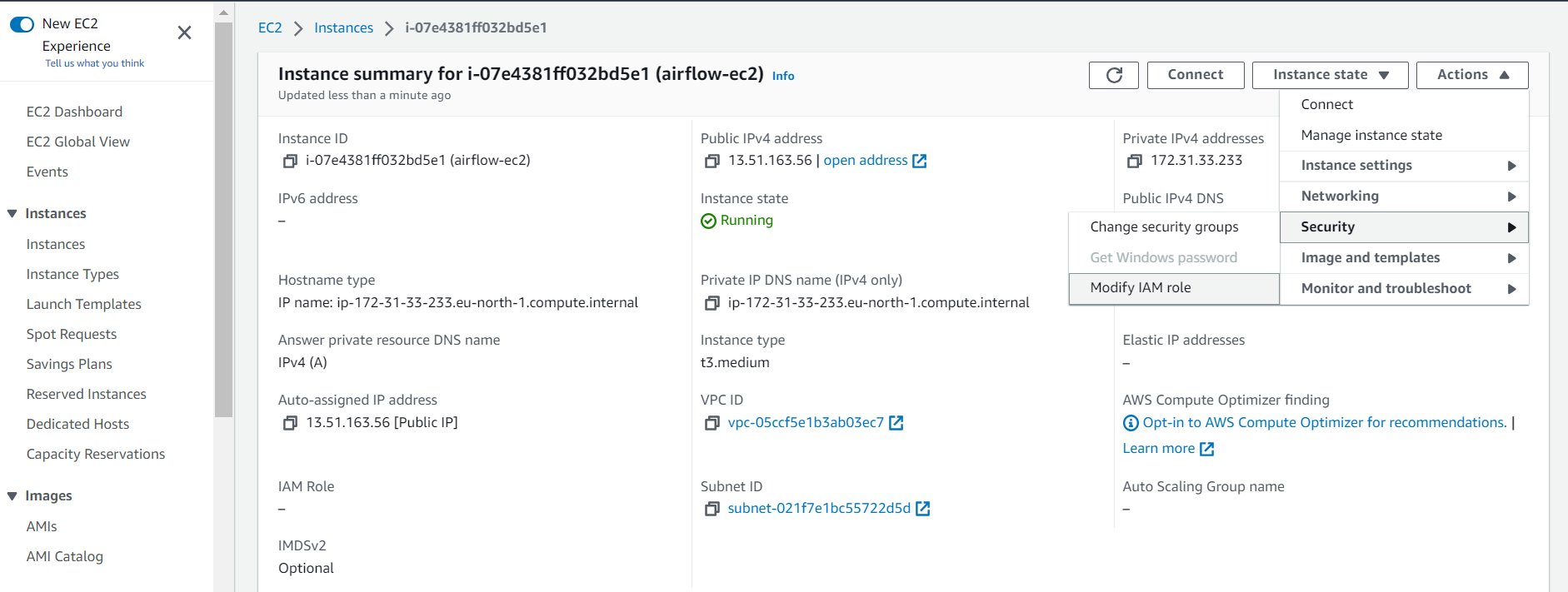


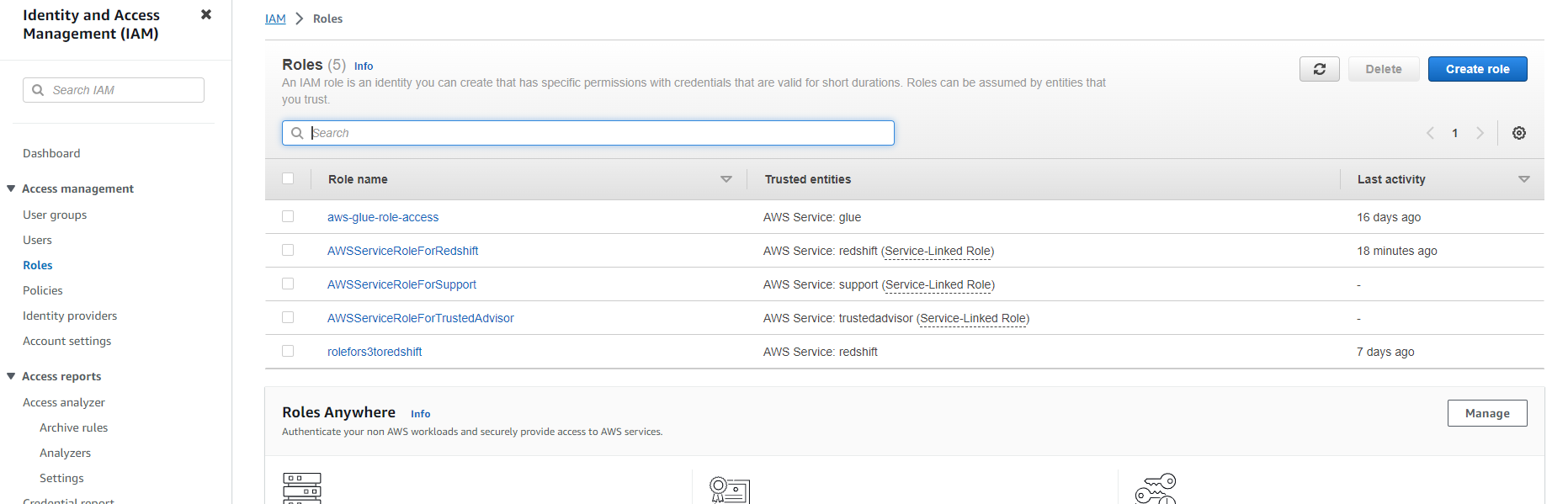
Step: Run the airflow standalone

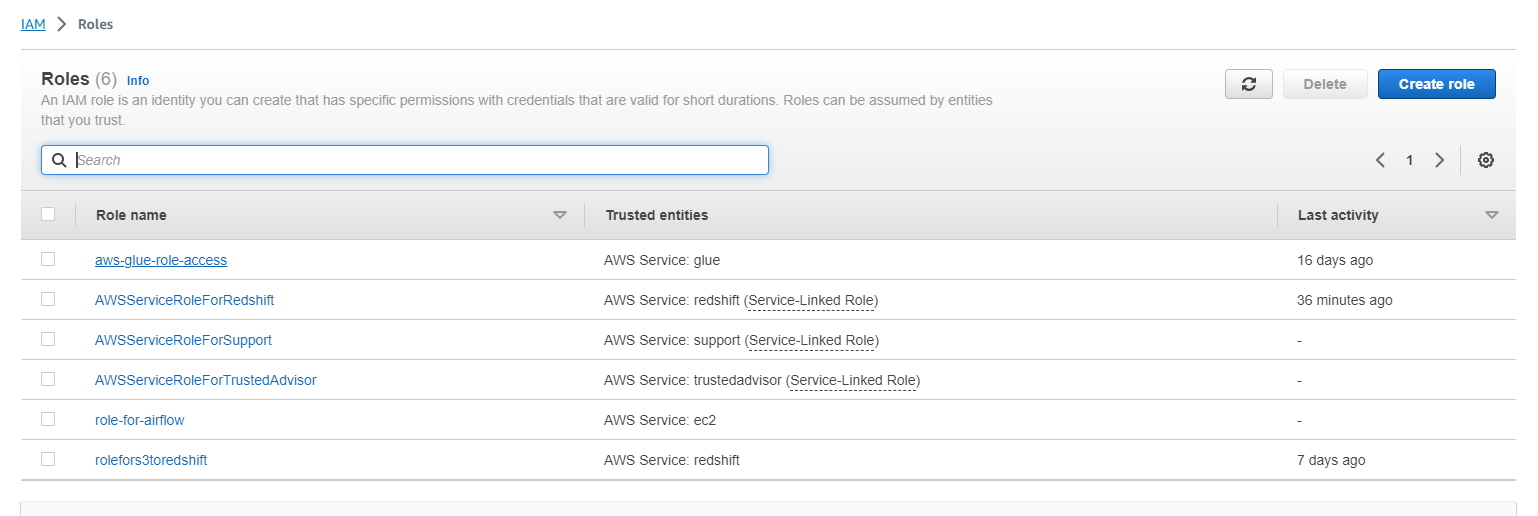


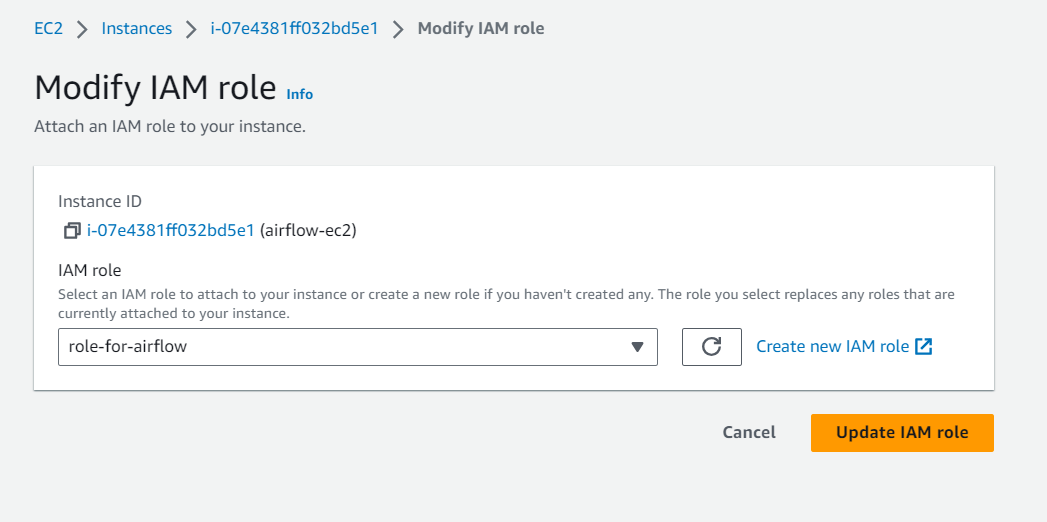
Step :

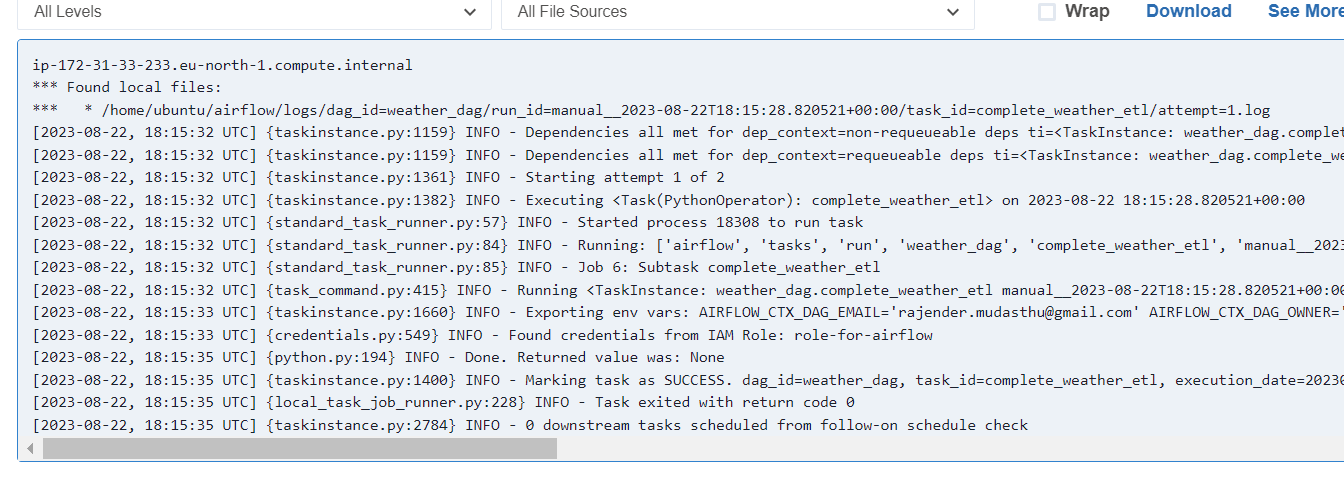
Change the IAM role access the s3 bucket

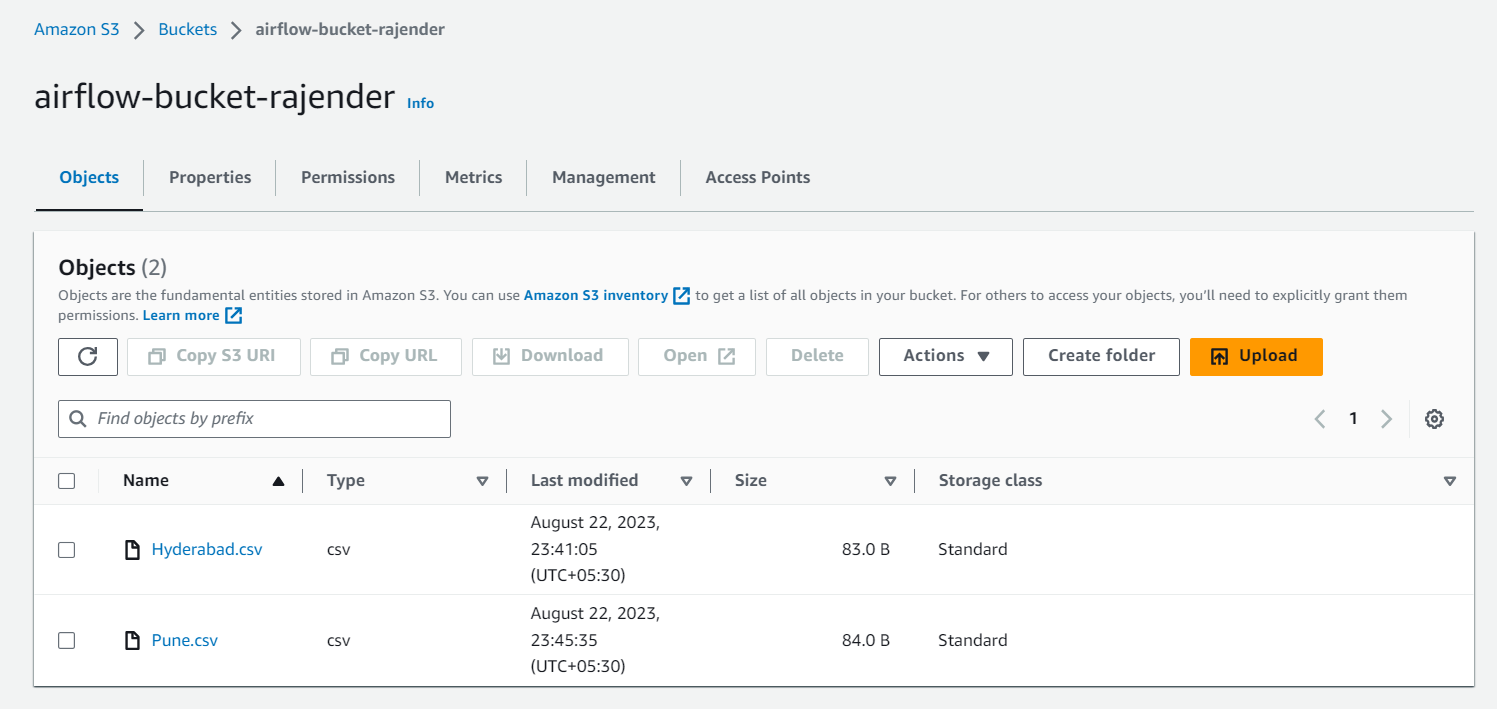












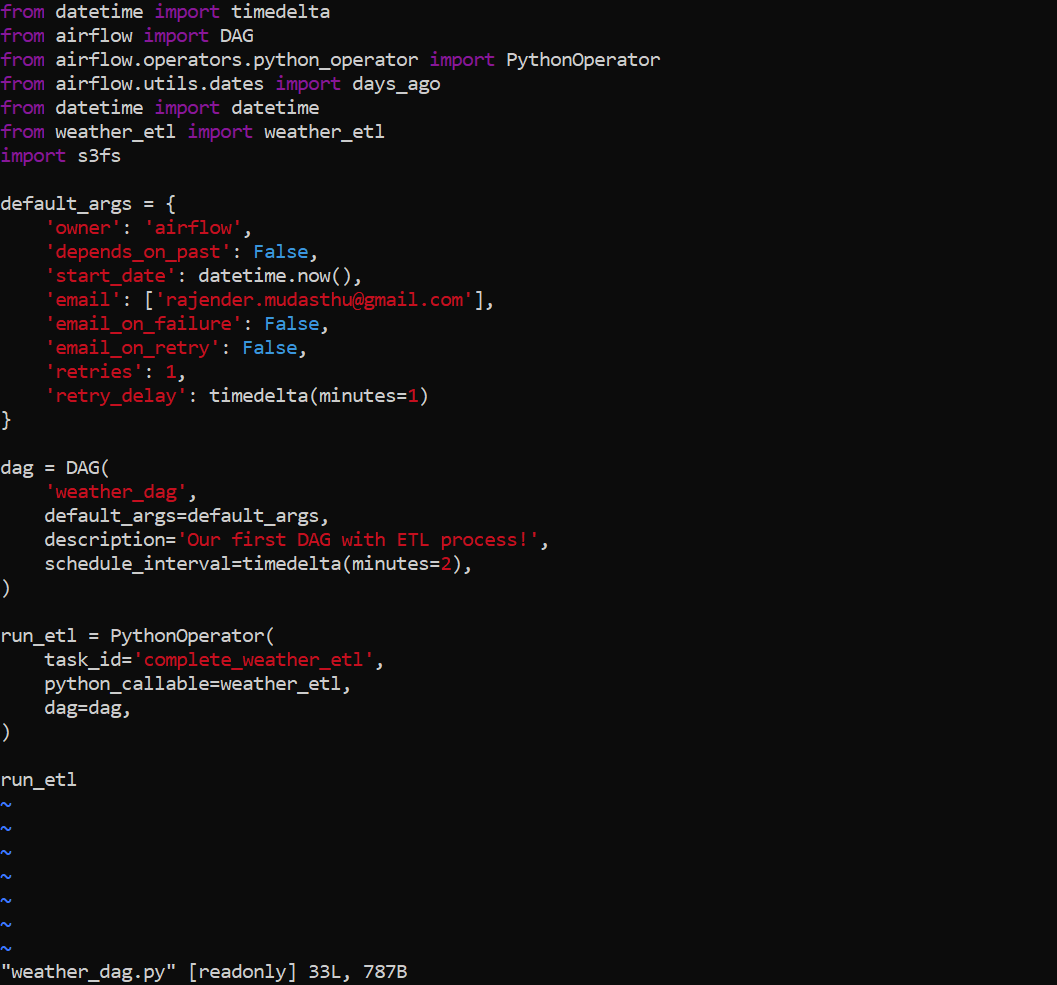
Now test for getting the data according to each hour:

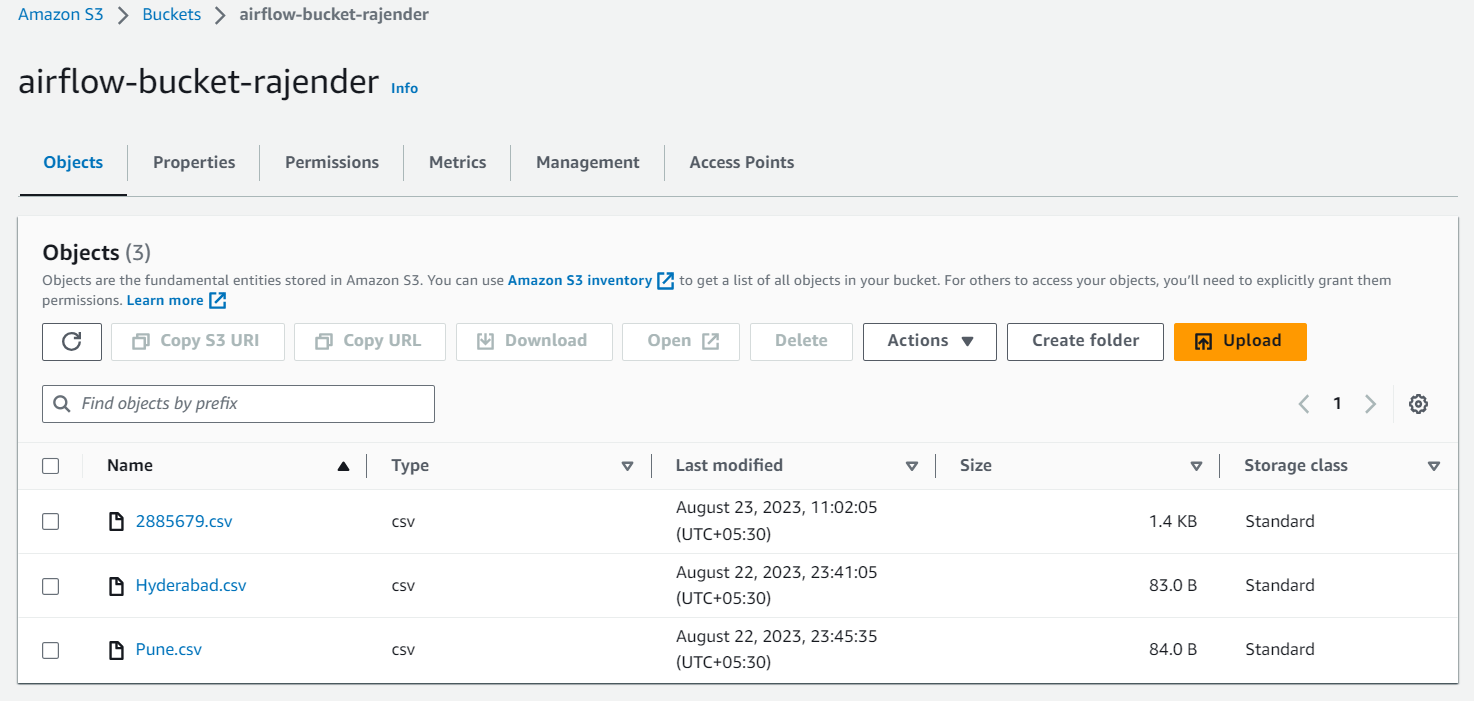
Go and take academic subscription

Then login to the weather api and get the data as shown below

Create a dataframe and save it on the S3 stage







import requests import pandas as pd import json from datetime import datetime import s3fs def weather\_etl(id=2885679): api\_key='70de1cd6a4b552df328b9d37c66b7892' data=requests.get(f"http://history.openweathermap.org/data/2.5/history/city?id=2885679&type=hour&appid=70de1cd6a4b552df328b9d37c66b7892") weather = data.json() weather\_data = [] for entry in weather["list"]: weather\_data.append({ "city\_id": weather["city\_id"], "dt": entry["dt"], "temp": entry["main"]["temp"], "feels\_like": entry["main"]["feels\_like"], "pressure": entry["main"]["pressure"], "humidity": entry["main"]["humidity"], "temp\_min": entry["main"]["temp\_min"], "temp\_max": entry["main"]["temp\_max"] }) # Create a pandas DataFrame df = pd.DataFrame(weather\_data) # Create a DataFrame from the weather data return df.to\_csv('s3://airflow-bucket-rajender/{}.csv'.format(id))

**Example of location:**

{

"id": 1269843,

"city": {

"id": {

"$numberLong": "1269843"

},

"name": "Hyderabad",

"findname": "HYDERABAD",

"country": "IN",

"coord": {

"lon": 78.474442,

"lat": 17.37528

},

"zoom": {

"$numberLong": "6"

}

}