**SETTING UP AIRFLOW ON EC2:**

1. Creating EC-2 instance with the following requirements:
   1. t3.medium
   2. ubuntu - platform
   3. pem - key pair
   4. EMR master - Security group
2. Login putty with ubuntu username
3. Update all the packages:
   1. ***sudo apt update***
   2. ***sudo apt install -y python3-pip***
   3. ***sudo apt install -y build-essential libssl-dev libffi-dev python3-dev***
4. Creating virtual environment



1. Activate the environment
   1. source my\_env/bin/activate
   2. 
2. Installing Apache Airflow (Within Environment)
   1. pip install apache-airflow
   2. airflow db init
   3. airflow webserver (After creating DAG to run the website in local host)
3. Create DAG folder within airflow
   1. Creating python file (print Hello World)

Graphical user interface, text, application

Description automatically generated

* 1. Creating DAG DummyGraphical user interface, text, application

     Description automatically generated

from datetime import datetime

from airflow import DAG

from airflow.operators.python\_operator import PythonOperator

def get\_config():

pass

def read\_write\_data():

pass

def pre\_validation():

pass

def livy\_transformation():

pass

def post\_validation():

pass

dag = DAG('Dummy\_DAG', description='Simple tutorial DAG',

schedule\_interval='0 12 \* \* \*',

start\_date=datetime(2017, 3, 20), catchup=False)

get\_config\_file = PythonOperator(task\_id='get\_config\_file', python\_callable=get\_config, dag=dag)

read\_write\_data\_s3 = PythonOperator(task\_id='read\_write\_data\_s3', python\_callable=read\_write\_data, dag=dag)

prevalidation = PythonOperator(task\_id='prevalidation', python\_callable=pre\_validation, dag=dag)

1. Make sure to do tunnelling to enable localhost (Port 8080)
2. Run Webserver and Run Scheduler via duplicate session
   1. 
   2. 
3. **Creating Username and passwords:**

airflow users  create --role Admin --username admin --email admin --firstname admin --lastname admin --password admin