**Janarish Saju C**

AI/ML Engineer

Apache Airflow

November 20, 2022

# About

Apache Airflow is an open-source platform for developing, scheduling, and monitoring batch-oriented workflows. 

# Key Terms

* ETL pipelines that extract data from multiple

sources

* DAG Representation
* Workflow Management
* Job Scheduling

## 

## 

# Installation Steps

***Install Ubuntu*** - If using Docker (optional)

docker run -dit -p 8000:80 --name airflow ubuntu

***Get into docker container terminal*** -If using Docker (optional)

docker exec -it "container\_id" /bin/bash

***Install necessary programs***

apt-get update

apt-get install sudo

Sudo apt-get install vim

Sudo apt-get install vi

Sudo apt-get install nano

***Install Python3***

sudo apt-get install python3

sudo apt-get install python3-pip

***Set Airflow Home Environment***

cd home - Get into ubuntu home

mkdir airflow - create directory “airflow”

cd airflow - Get into airflow home

mkdir dags - create directory “dags” (here we are storing DAG’s python file and text data)

export AIRFLOW\_HOME=`pwd’ - Set airflow home environment

***Install Airflow***

pip install apache-airflow

***Check Airflow Installed***

airflow version

***Database Initialization***

airflow db init

***Create Username & Password***

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

***Copy files from Host to docker Container*** -If using Docker (optional)

exit - exit console

*Move my\_simple\_dag.py from host to container*

docker cp d:\airflow\my\_simple\_dag.py airflow:/home/airflow/dags

*Move greet.txt from host to container*

docker cp d:\airflow\greet.txt airflow:/home/airflow/dags

***Return to docker container terminal*** -If using Docker (optional)

docker exec -it "container\_id" /bin/bash

***Verify airflow.cfg has correct path***

cd home/airflow

vim airflow.cfg

dags\_folder = /home/airflow/dags - Just verify

***Error - If timezone error occurred***

apt-get install -y tzdata - try only if error occurred

# Running Steps

***Run Airflow***

airflow standalone

airflow webserver

airflow scheduler

***Check Airflow UI***

<http://localhost:8080/home>

Username: admin

Password: admin

Running successfully...