Installation of Airflow

1. Create a python virtual environment.

sudo apt install python3-virtualenv

virtualenv -p /usr/bin/python3 ~/Envs/venv

1. Activate the virtual environment.

cd ~/Envs/venv/bin

source ./activate

1. Execute the following to install airflow and its dependencies. (Postgres is used here)

AIRFLOW\_VERSION=2.1.0

PYTHON\_VERSION=3.8

CONSTRAINT\_URL=<https://raw.githubusercontent.com/apache/airflow/constraints-${AIRFLOW_VERSION}/constraints-${PYTHON_VERSION}.txt>

curl -L –proxy http://www.proxy.ms.com:8080/$CONSTRAINT\_URL | tee airflow-constraints.txt

pip install "apache-airflow[async,postgres,google]==${AIRFLOW\_VERSION}" --constraint "${CONSTRAINT\_URL}"

1. Database configuration and initialization

Open airflow.cfg and change the sql\_alchemy\_conn as follows

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

setup the db and privileges

CREATE DATABASE airflow;

CREATE USER airflow WITH PASSWORD 'secret';

GRANT ALL PRIVILEGES ON DATABASE airflow TO airflow;

Initialize the DB

airflow db init

airflow users create \

--username admin \

--role Admin

5. Start server and scheduler

airflow webserver –port 8080

airflow schedular