**Airflow Documentation**

**Installing Apache Airflow**

1. Optional: create a python virtual environment, so all libraries will be installed here.
2. Install Airflow:

pip install "apache-airflow==2.2.3" --constraint "https://raw.githubusercontent.com/apache/airflow/constraints-2.2.3/constraints-no-providers-3.9.txt"

1. Make an Airflow db:

airflow db init 🡪 there will be a folder named airflow at $HOME

cd ~/airflow

1. Create user credential:

airflow users create \

--username admin \

--password admin \

--firstname <FirstName> \

--lastname <LastName> \

--role Admin \

--email <YourEmail>

1. Start Airflow web server

airflow webserver -D

Airflow interface bisa dibuka di 0.0.0.0:8080 atau 127.0.0.1:8080

masukkan user ‘admin’ dan password ‘admin’

1. Run the scheduler:

airflow scheduler -D

**Make an Airflow Direct Acyclic Graph (DAG) data pipeline**

The data pipeline will get the current datetime from the Terminal, process it, and save it to a CSV file. Pretty simple, but you’ll learn how Airflow’s Bash and Python operators work, and also how to communicate between tasks using Xcoms, and how to schedule your data pipelines.

1. Create a new dags folder:

mkdir dags

1. Create a new dag python file:

vi my\_dag.py

1. Isi my\_dag.py dengan pipeline

Buat beberapa task yg akan dijalankan pada pipeline

xcom akan menangkap output dari task sebelumnya

Variable akan dibuat pada webserver GUI dari Airflow

1. Jalankan my\_dag.py:

airflow tasks test <dag\_name> <task\_name> <date\_in\_the\_past>

e.g.: airflow tasks test first\_airflow\_dag get\_datetime 2022-2-1

<date\_in\_the\_past> 🡪 the date time input at which the pipeline start running its process

1. Membuat variable di Airflow web server:

Klik admin 🡪 Variable

Key = my\_dag\_csv\_path

Val = home/mukhlishga/airflow/datetime.csv

Variable ini akan ditangkap oleh my\_dag.py

1. Pastikan task pada my\_dag.py disusun berurutan:

task\_get\_datetime >> task\_process\_datetime >> task\_save\_datetime

1. Run dag di Airflow web GUI, lihat hasil pada datetimes.csv