**MongoDB**

***# Cập nhật danh sách gói***

sudo apt update

***# Nâng cấp các gói hiện có***

sudo apt upgrade

***# Thêm khóa GPG của MongoDB***

wget -qO - https://www.mongodb.org/static/pgp/server-6.0.asc | sudo apt-key add -

***# Tạo file danh sách nguồn cho MongoDB***

echo "deb [ arch=amd64,arm64 ] https://repo.mongodb.org/apt/ubuntu jammy/mongodb-org/6.0 multiverse" | sudo tee /etc/apt/sources.list.d/mongodb-org-6.0.list

***# Cập nhật lại danh sách gói sau khi thêm nguồn MongoDB***

sudo apt update

***# Cài đặt MongoDB***

sudo apt install -y mongodb-org

***# Khởi động dịch vụ MongoDB***

sudo systemctl start mongod

***# Đảm bảo dịch vụ MongoDB tự động khởi động khi hệ thống khởi động***

sudo systemctl enable mongod

***# Truy cập vào MongoDB shell***

mongosh

***# Chuyển sang cơ sở dữ liệu myDatabase***

use myDatabase

***# Thoát khỏi MongoDB shell***

exit

sudo service mongod start

sudo service mongod status

sudo service mongod stop

**MySQL**

***# Thiết lập cơ sở dữ liệu backend cho Airflow***

***# Cập nhật danh sách gói***

sudo apt-get update

***# Cài đặt các gói cần thiết***

sudo apt-get install build-essential libssl-dev libffi-dev python3-dev pkg-config

***# Cài đặt các thư viện MySQL***

sudo apt-get install default-libmysqlclient-dev libmariadb-dev

***# Cập nhật lại danh sách gói***

sudo apt-get update

***# Cài đặt MySQL server***

sudo apt-get install mysql-server

***# Khởi động dịch vụ MySQL***

sudo service mysql start

***# Truy cập vào MySQL với quyền root***

sudo mysql -u root -p

***# Tạo cơ sở dữ liệu mới cho Airflow***

CREATE DATABASE airflow\_db;

***# Tạo người dùng mới cho Airflow và cấp quyền***

CREATE USER 'airflow\_user'@'localhost' IDENTIFIED BY 'your\_password';

GRANT ALL PRIVILEGES ON airflow\_db.\* TO 'airflow\_user'@'localhost';

***# Áp dụng các thay đổi quyền***

FLUSH PRIVILEGES;

**Airflow**

***# Cập nhật hệ thống***

sudo apt update

***# Cài đặt các gói cần thiết***

sudo apt install -y python3-pip python3-dev libmysqlclient-dev libssl-dev libffi-dev

***# Cài đặt và tạo môi trường ảo***

sudo apt install python3-venv

python3 -m venv airflow\_env

source airflow\_env/bin/activate

***# Thiết lập biến môi trường cho Airflow***

export AIRFLOW\_HOME=~/airflow

***# Cài đặt Apache Airflow, csdl backend, celery cho task song song***

pip install apache-airflow

pip install 'apache-airflow[mysql]'

pip install 'apache-airflow[celery]'

***# Cài đặt một broker cho Celery: Celery yêu cầu một message broker như Redis hoặc RabbitMQ để phân phối các task.***

sudo apt-get install redis-server

sudo service redis-server start

***# Khởi tạo cơ sở dữ liệu cho Airflow***

airflow db init

***# Tạo người dùng quản trị cho Airflow***

airflow users create \

--username admin \

--firstname Admin \

--lastname User \

--role Admin \

--email admin@example.com \

--password admin

***# Cấu hình song song trên airflow***

***# Mở file airflow.cfg trong Folder airflow***

***# Tìm phần [database], chỉnh sửa để trỏ đến MySQL:***

[database]

sql\_alchemy\_conn = mysql+mysqldb://airflow\_user:your\_password@localhost/airflow\_db

***# Tìm phần [core] trong tệp airflow.cfg, sửa để sử dụng CeleryExecutor:***

[core]

executor = CeleryExecutor

***# Tìm phần [celery] trong tệp airflow.cfg, sửa hoặc thêm nếu chưa tồn tại:***

***# Thường nằm dưới phần [logging]***

[celery]

broker\_url = redis://localhost:6379/0

result\_backend = db+mysql://airflow\_user:your\_password@localhost/airflow\_db

***# Số lượng worker mà Celery có thể xử lý đồng thời***

worker\_concurrency = 16

***# Kết nối timeout khi worker không phản hồi (số giây)***

task\_acks\_late = True

worker\_prefetch\_multiplier = 1

***# Khởi động máy chủ web Airflow***

airflow webserver --port 8080

*# Truy cập giao diện web của Airflow tại địa chỉ http://localhost:8080*

***# Mở một tab mới để chạy scheduler***

***# Thiết lập môi trường airflow (luôn luôn phải thiết lập mỗi lần chạy mới)***

source ~/airflow\_env/bin/activate

airflow scheduler

***# Mở một tab mới để chạy worker***

source ~/airflow\_env/bin/activate

airflow celery worker

***# Mở một tab mới để ném code vào và chạy***

***# Tạo thư mục chứa code***

mkdir ~/airflow/dags

cp sample.py ~/airflow/dags

***# Dừng web Server, scheduler, worker***

Ctr + C

**Extend**

***# Fix RAM WSL***

*###*

*Windows Subsystem for Linux (WSL) sử dụng chung tài nguyên RAM và CPU với hệ điều hành Windows*

*Để kiểm tra và thay đổi cấu hình, bằng cách tạo hoặc chỉnh sửa tệp .wslconfig trong thư mục người dùng (ví dụ: C:\Users\<Tên người dùng>\.wslconfig)*

*###*

[wsl2]

memory= 6GB # Giới hạn RAM

processors= 4 # Giới hạn số lượng CPU

***# Chạy lệnh để khởi động lại WSL***

wsl –shutdown

### Thực hiện lệnh: free -h để kiểm tra

***# Install library for python***

***# Trỏ đến môi trường ảo airflow***

source ~/airflow\_env/bin/activate

***# Tạo file install\_me.sh***

touch install\_me.sh

***# Mở file install\_me.sh và thêm nội dung sau***

#!/bin/bash

sudo apt update

sudo apt install -y python3-pip

pip3 install apache-airflow requests beautifulsoup4 pymongo

***# Lưu file và cấp quyền thực thi***

chmod +x install\_me.sh

***# Chạy script để cài đặt các thư viện***

./install\_me.sh