Bước 1: Tạo project scrapy để crawl dữ liệu

Ảnh có chứa văn bản, máy tính, phần mềm, màn hình

Mô tả được tạo tự động

Bước 2: Thiết lậy middleware để lưu dữ liệu lên Mongo(Chỉ lưu những url các job chưa có trong database)

Ảnh có chứa văn bản, đồ điện tử, máy tính, màn hình

Mô tả được tạo tự động

Bước 3: Download docker compose airflow từ trang chủ về với câu lệnh

Lệnh: curl 'https://airflow.apache.org/docs/apache-airflow/2.8.4/docker-compose.yaml' -o docker-compose.yaml

* Kết quả:

Ảnh có chứa văn bản, màn hình, máy tính, phần mềm

Mô tả được tạo tự động

Bước 4: Tạo các folder dags, logs, plugins, config ngang cấp với docker-compose.yaml với mkdir hoặc bằng tay

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phần mềm đa phương tiện

Mô tả được tạo tự động

Bước 5: Mở docker desktop và chạy lệnh:

docker compose up airflow-init

Ảnh có chứa văn bản, máy tính, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

Bước 6: Chạy lệnh docker compose up

docker compose up

Ảnh có chứa văn bản, màn hình, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

Bước 7: Kiểm tra các dịch vụ đã chạy chưa với lệnh: docker ps, nếu thấy airflow webservices up nghĩa là đã chạy thành công server ui ảo localhost:8080

Bước 8: Vào web mở <http://localhost:8080> (tài khoản: airflow, mật khẩu airflow)

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Biểu tượng máy tính

Mô tả được tạo tự động

Bước 9: Kiểm tra flow của dag file mình đã tạo bằng graph

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Biểu tượng máy tính

Mô tả được tạo tự động

Bước 9: chọn vào 1 task và nhấn run trigger

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Biểu tượng máy tính

Mô tả được tạo tự động