| **TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP**  **THÀNH PHỐ HỒ CHÍ MINH**    **KHOA CÔNG NGHỆ THÔNG TIN** |  |  |
| --- | --- | --- |
|  |  |  |

*Họ và tên thí sinh ………………….…………………; MSSV: ………………; Số Máy: ..........*

**Đề bài: Xây dựng data pipeline sử dụng airflow để xử lý và lưu trữ dữ liệu vào MongoDB và QdrantDB**

Giải nén thư mục ThiGiuaKy.zip. Cấu trúc thư mục ThiGiuaKy như sau:

ThiGiuaKy

airflow

.env

requirements.txt

Dockerfile

docker-compose.yaml

MyApp

app.py

Dockerfile

requirements.txt

data\_iuh\_new.json

pipeline.py

**Chuẩn bị:**

* **Đổi tên thư mục “airflow” thành “<hoten>\_<mssv>”**
* **Vào file .env đặt các biến theo yêu cầu**
* **Load image toàn bộ image được cung cấp vào docker bằng lệnh**

docker load --input <image\_name>

*Ví dụ: để load image airflow\_khdl16a1 ta thực hiện lệnh sau:*

docker load --input airflow\_khdl16a1.tar

Câu I. (2 điểm) Xây dựng Airflow với mongoDB và qdrantDB từ image *airflow\_khdl16a1*

*Hướng dẫn*: *hoàn thiện các file sau*:

* ThiGiuaKy/<hoten>\_<mssv>/Dockerfile
* ThiGiuaKy/<hoten>\_<mssv>/docker-compose.yaml

Câu II. (6.5 điểm) Build data pipeline and scheduler with airflow

Hoàn thiện file *pipeline.py* (**Thực hiện theo yêu cầu trong file**)

* Thêm DAG thành công và scheduler đúng yêu cầu vào airflow (1.5 điểm).
* Chạy thành công 2 **bash operator** (*start* và *end*) (1 điểm)
* Tạo collection qdrant thành công (1 điểm)
* Thêm dữ liệu vào mongoDB thành công (1 điểm)
* Thêm dữ liệu vào QdrantDB thành công (1 điểm)
* Thực hiện tìm kiếm trên QdrantDB (1 điểm)

Câu III. (1.5 điểm) Build app sử dụng QdrantDB kết hợp với pipeline cập nhật dữ liệu.

* Xây dựng app bằng cách hoàn thiện các file trong thư mục MyApp
* Thêm service có tên “**myapp**” vào trong file **docker-compose.yaml** và khởi chạy service “**myapp”**

*Lưu ý:*

* Đặt đúng tên theo yêu cầu được nêu ra trong các file hướng dẫn đi kèm
* Nộp bài: xóa file “data\_iuh\_new.json”, “airflow\_khdl16a1.tar” và “app\_khdl16a1.tar” nén toàn bộ thư mục ThiGiuaKy thành “ThiGiuaKy.zip” nộp lên LMS (https://lms.iuh.edu.vn/mod/assign/view.php?id=277882).

------------------------- Hết ------------------------------