|  |  |  |
| --- | --- | --- |
| TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP THÀNH PHỐ HỒ CHÍ MINH  **KHOA CNTT** | **ĐỀ THI GIỮA KỲ**  **Môn thi :** Kỹ Thuật Dữ Liệu  **Lớp/Lớp học phần:** DHKHDL18A  **Ngày thi:** 28/03/2024  Thời gian làm bài: 60 phút  *(Không kể thời gian phát đề)* |  |
|  |  |  |

*Họ và tên thí sinh : Nguyễn Thị Viên; MSSV: 22003025*

**Đề thi:**

1. Sử dụng thư viện *faker* để phát sinh dữ liệu giả gồm 10 records các trường thông tin *(‘id’, ‘name’, age’, ‘city’, ‘salary’)* lưu vào file *employees.csv*

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Viết file ‘*docker-compose.yaml*’ để tạo một container với Airflow và PostgresSQL (hoặc MongoDB)

A computer screen with many colorful lines

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A computer screen with many colorful lines and text

Description automatically generated

Viết flie *demo.py* đọc file dữ liệu *employees.csv* và thực hiện pipeline như sau:

* Chọn ngẫu nhiên một dòng (records)
* In ra dòng chữ: “Chuẩn bị thêm dữ liệu vào hệ thống”
* Thực hiện insert vào Database.
  + Nếu thành công, thông báo ra dòng chữ “Chèn dữ liệu thành công”,
  + Ngược lại nếu trùng id thì thông báo “Đã tồn tại bản ghi trong hệ thống” và thử lại sau 3 phút; số lần thử tối đa là 3 lần.
* In ra dòng chữ: “Hoàn thành việc chèn dữ liệu”.

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer screen

Description automatically generated**

**Import file lên airflow**

**A screenshot of a computer

Description automatically generated**

**Hiển thị ngày giờ chạy: (này là em sáng em quên chụp nên gờ em chạy chụp lại)**

**A screenshot of a computer

Description automatically generated**

**Chèn vào mongobd**

**A screenshot of a computer

Description automatically generated**

**Nộp bài:** Tạo thư mục có định dạng: stt\_hoten\_GK.zip bên trong chứa các tập tin như sau:

* docker-compse.yaml
* dockerfile
* demo.py
* Chụp hình screenshot và dán vào file word này, điền thông tin cá nhân

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

*Lưu ý: - Đề thi (được) sử dụng tài liệu.*