

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 phân: DHKHD18A

Ngày thi: 28/03/2024

Thời gian làm bài: 60 phút

(Không kể thời gian phát đề)

Họ và tên thí sinh; MSSV:

Đề thi:

- 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*
- Viết file '*docker-compose.yaml*' để tạo một container với Airflow và PostgreSQL (hoặc MongoDB)
- Viết file *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".

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.