File report\_data.xlsx bao gồm dữ liệu về các đơn hàng và khách hàng tại thị trường Mỹ của 1 trang thương mại điện tử.

Yêu cầu:

* Import dữ liệu từ file excel report\_data.xlsx vào SQL server thành 2 bảng tên là “Order” và “Customer”
* Xử lý dữ liệu (bằng SQL hoặc Python) nếu cần thiết
* Kết nối Power BI desktop vào SQL server, tạo 1 report thể hiện tổng quan tình hình hoạt động của trang thương mại điện tử này (từ cả khía cạnh order lẫn customer)
* Trình bày những nhận định của bạn về tập khách hàng và hoạt động của trang thương mại điện tử này vào 1 file văn bản (ví dụ Microsoft Word)
* Giả sử report power bi cần được update dữ liệu hàng ngày. Hãy tự động hóa việc chuẩn bị dữ liệu cho report power bi của bạn theo logic sau:

+ Mỗi ngày bạn đều nhận được 1 file excel report\_data.xlsx với cấu trúc không thay đổi, và chỉ thay đổi về dữ liệu (dữ liệu được cập nhật mới từng ngày)

+ Bạn chọn 1 trong 2 cách sau để cập nhật dữ liệu trong SQL server:

Cách 1:

* + - import file dữ liệu nhận được vào SQL server thành 2 bảng Order\_temp và Customer\_temp.
    - Viết code SQL lọc ra dữ liệu của những order mới xuất hiện hôm nay (có trong bảng Order\_temp mà không có trong bảng Order) và cập nhật thêm vào bảng Order.
    - Lưu dữ liệu từ bảng Customer vào bảng Customer\_hist (là bảng để lưu lịch sử thay đổi dữ liệu Customer, tuy nhiên để cho đơn giản thì ta chỉ lưu lại lần gần nhất, chính là dữ liệu hiện có trong bảng Customer). Sau đó update lại toàn bộ bảng Customer theo bảng Customer\_temp (xóa toàn bộ dữ liệu bảng Customer đi và đưa dữ liệu từ Customer\_temp vào)
    - Sử dụng SQL server agent hoặc Microsoft Windows Task Scheduler để tự động hóa việc cập nhật dữ liệu này

Cách 2:

* + - Import file dữ liệu nhận được vào Python thành 2 bảng (object) Order\_temp và Customer\_temp.
    - Import dữ liệu 2 bảng Order và Customer từ SQL server vào Python và đặt tên các object trong Python giống như vậy.
    - Xử lý tương tự logic ở cách 1 để cập nhật dữ liệu vào các object Order, Customer, Customer\_hist trong Python.
    - Load dữ liệu từ các object Order, Customer, Customer\_hist trong Python ghi vào các bảng với tên tương ứng trong SQL server (xóa hết dữ liệu từ các bảng này đi và load toàn bộ dữ liệu mới từ Python vào)
    - Sử dụng Microsoft Windows Task Scheduler để tự động hóa việc cập nhật dữ liệu này
* Sau khi tự động hóa xong, hãy quay video màn hình máy tính (1 video ngắn khoảng 5, 10p) trình bày :

+ code cập nhật dữ liệu của bạn và cách bạn tự động hóa nó. Thử thay đổi dữ liệu trong file excel và chạy job để thể hiện code tự động đã hoạt động đúng và trả ra kết quả như mong muốn.

+ show báo cáo Power BI và giới thiệu qua về các phần chính trong báo cáo (không cần trình bày insights, chỉ cần nêu có những phần nào, mỗi phần/chart có tác dụng gì)