

CSC12107 – HTTT PV TRÍ TUỆ KINH DOANH

ĐỒ ÁN THỰC HÀNH

1. Thông tin chung

| | |
|----------------------------|--|
| Mã số bài tập: | 2023_DATH |
| Thời lượng dự kiến: | 6 tuần |
| Deadline nộp bài: | 15/12/2023 |
| Hình thức: | Bài tập nhóm |
| Hình thức nộp bài: | Nộp qua Moodle môn học |
| GV phụ trách: | Hồ Thị Hoàng Vy, Tiết Gia Hồng |
| Thông tin liên lạc với GV: | hthvy@fit.hcmus.edu.vn , tghong@fit.hcmus.edu.vn |

2. Chuẩn đầu ra cần đạt

Bài tập này nhằm mục tiêu đạt được các chuẩn đầu ra sau:

- G3.2: Giải thích vai trò và đặc trưng của kho dữ liệu, ý nghĩa của các bảng chiều và dữ kiện, các thuộc tính, phân cấp chiều trong lược đồ đa chiều
- G3.3: Thiết kế lược đồ chuẩn hoá, đa chiều (sao, bông tuyết) dựa vào dữ liệu hệ thống tác vụ và yêu cầu phân tích từ tình huống cho trước
- G4.2 Phân biệt được các ứng dụng khai phá dữ liệu vào kinh doanh thông minh. Giải thích đặc trưng và mục đích của khai phá dữ liệu, tầm quan trọng và một số kỹ thuật phổ biến
- G5.1 Triển khai quy trình ETL để rút trích dữ liệu từ nhiều nguồn, biến đổi, làm sạch dữ liệu, nạp dữ liệu vào KDL sử dụng công cụ SSIS
- G5.2 Xây dựng KDL đa chiều sử dụng công cụ SSAS và giải thích được lựa chọn phép toán OLAP phù hợp đối với 1 số yêu cầu phân tích.
- G5.3 Sử dụng một số công cụ biểu diễn dữ liệu (SSRS, powerBI, excel...) để biểu diễn kết quả phân tích, khai thác được (report, dashboard...)
- G5.4 Sử dụng SSAS và áp dụng các kỹ thuật mining tích hợp để thực hiện khai thác dữ liệu từ KDL xây dựng được.

3. Mô tả đề án

Bộ dữ liệu của đề án mô tả tình hình bán hàng tại Supermarket_sales ở 3 chi nhánh ở 3 thành phố Yangon, Mandalay, Naypyitaw trong 3 tháng đầu năm 2019:

Yêu cầu:

- *Mô tả dữ liệu:* Mô tả ý nghĩa các thuộc tính trong tập dữ liệu
 - **Link drive**

- Thiết kế kho dữ liệu (KDL), chuẩn hóa, tổng hợp, nạp dữ liệu các nguồn vào KDL, thiết kế và xây dựng cube.
 - Chuyển đổi dữ liệu ngày tháng sao cho có thể tạo được Date dimension với phân cấp chiều: năm > tháng > ngày > giờ
 - Xác định và thiết kế các phân cấp chiều khác để đáp ứng yêu cầu OLAP và report.
- Thiết kế dữ liệu
 - Thiết kế NDS
 - Thiết kế DDS
- Khai thác dữ liệu:
 - **Report:**
 1. Thống kê số thành viên mua hàng theo ngày, tháng, năm
 2. Thống kê doanh thu của khách hàng (member, normal) theo ngày, tháng, năm và theo chi nhánh.
 3. Thống kê số lượng khách thanh toán theo cash/ debit/... ở từng chi nhánh theo từng tháng trong năm
 4. Thống kê lượng rating của khách hàng (member, normal) theo từng loại sản phẩm (ProductLine)
 5. Thống kê số lượng sản phẩm bán được theo từng thời điểm (time / date)
 6. Thống kê số lượng khách hàng nữ đã mua theo từng loại sản phẩm
 - **OLAP:**

Tình hình mua hàng của khách hàng theo từng chi nhánh, từng loại sản phẩm, theo thời gian, hình thức thanh toán
 - **Mining**
 - Để dự đoán kết quả mua hàng của khách hàng có phải theo từng loại sản phẩm
 - SV có thể tự đề xuất 1 trường hợp
 - **Kết luận chung**

4. Các yêu cầu & quy định chi tiết cho bài nộp

- Có thể sử dụng word, excel, draw.io... để vẽ các lược đồ trên
- Bài làm ghi rõ mã nhóm, mssv, họ tên vào đầu bài làm, phân công và đánh giá mỗi thành viên và toàn nhóm.
- Đặt tên: **MaNhom_MSSV1_MSSV2**

5. Tài liệu tham khảo:

- Topic 01 – Hướng dẫn sử dụng SSIS
- Topic 02 – Hướng dẫn sử dụng SSAS - OLAP
- Topic 03 – Truy vấn OLAP với MDX
- Topic 04 – Hướng dẫn sử dụng SSAS – mining

6. Cách đánh giá

| STT | Nội dung | Điểm |
|-----|------------------------------------|------|
| 1 | Thiết kế & cài đặt CSDL (NDS, DDS) | 20% |
| 2 | Cài đặt ETL nạp dữ liệu | 40% |

| | | |
|---|-------------|-------------|
| 3 | Report | 10% |
| 4 | MDX, OLAP | 15% |
| 5 | Mining | 15% |
| | Tổng | 100% |

câu hỏi: Cho biết doanh thu của từng chi nhánh của các năm
cho biết lượng khách hàng thanh toán theo mỗi hình thức cash, debit... của mỗi chi nhánh /
của tất cả chi nhánh theo từng năm...

sv lập dashboard theo dõi doanh thu, top sp bán chạy, chi nhánh đạt doanh thu cao nhất, city
có đông lượt mua hàng nhất, theo dõi doanh thu của 1 chi nhánh bất kỳ

bài này nó khá gần đồ án lý thuyết, nên sv cũng ko khó làm
Enter

làm đồ án này tối đa 8đ.

sv tự chọn 1 dataset khác, mô tả các yêu cầu, tương tự, tổng hợp dl từ 2 nguồn trở lên: tối đa
10đ.