**BỘ MÔN HỆ THỐNG THÔNG TIN – KHOA CÔNG NGHỆ THÔNG TIN**

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH, ĐẠI HỌC QUỐC GIA TP HCM**

GV phụ trách: ThS. Hồ Thị Hoàng Vy

ThS.Tiết Gia Hồng

Nhóm sinh viên thực hiện: 2023.CQ.BI.06

Đồ án/bài tập môn học  - BUSINESS INTELLIGENCE

HỌC KỲ II – NĂM HỌC 2023-2024

**môn HỆ THỐNG THÔNG TIN PHỤC VỤ TRÍ Tuệ kinh doanh**

**MỤC LỤC**

[**MỤC LỤC** 1](#_Toc155201466)

[**BẢNG THÔNG TIN CHI TIẾT NHÓM** 2](#_Toc155201467)

[**BẢNG PHÂN CÔNG & ĐÁNH GIÁ HOÀN THÀNH CÔNG VIỆC** 2](#_Toc155201468)

[**YÊU CẦU ĐỒ ÁN- BÀI TẬP** 4](#_Toc155201469)

[**A.** **Yêu cầu của Đồ án/Bài tập** 4](#_Toc155201470)

[**B.** **Kết quả** 4](#_Toc155201471)

[**1.** **Mô tả ý nghĩa các thuộc tính trong tập dữ liệu** 4](#_Toc155201472)

[**a.** **Sheet ‘supermarket\_sales’** 4](#_Toc155201473)

[**b.** **Sheet ‘product’** 5](#_Toc155201474)

[**c.** **Sheet ‘ProductLine’** 5](#_Toc155201475)

[**d.** **Sheet ‘city’** 6](#_Toc155201476)

[**TÀI LIỆU THAM KHẢO** 7](#_Toc155201477)

**BẢNG THÔNG TIN CHI TIẾT NHÓM**

|  |  |  |
| --- | --- | --- |
| **Mã nhóm:** | **2023.CQ.BI.06** | |
| **Tên nhóm:** | **Nhóm 6** | |
| **Số lượng:** | **4** | |
| **MSSV** | **Họ tên** | **Email** |
| 20120429 | Nguyễn Quốc Anh | 20120429@student.hcmus.edu.vn |
| 20120431 | Tô Trần Sơn Bá | 20120431@student.hcmus.edu.vn |
| 20120439 | Hoàng Văn Cầu | 20120439@student.hcmus.edu.vn |
| 20120446 | Nguyễn Đình Cường | 20120446@student.hcmus.edu.vn |

**BẢNG PHÂN CÔNG & ĐÁNH GIÁ HOÀN THÀNH CÔNG VIỆC**

|  |  |  |  |
| --- | --- | --- | --- |
| **Công việc thực hiện** | **Người thực hiện** | **Mức độ hoàn thành** | **Đánh giá của nhóm** |
|  |  | 100% | 4/4 |
|  |  | 100% | 4/4 |
|  |  | 100% | 4/4 |
|  |  | 100% | 4/4 |

# **YÊU CẦU ĐỒ ÁN- BÀI TẬP**

|  |  |
| --- | --- |
| **Loại bài tập** | **Lý thuyết 🗹 Thực hành 🗹 Đồ án Bài tập** |
| **Ngày bắt đầu** | **4/1/2024** |
| **Ngày kết thúc** | **4/1/2024** |

# **Yêu cầu của Đồ án/Bài tập**

Bộ dữ liệu của đồ án mô tả tình hình bán hàng tại Suppermarket\_sales ở 3 chi nhánh ở 3 thành phố Yangon, Mandalay, Naypyitaw trong 3 tháng đầu năm 2019 yêu cầu thực hiện:

1. Mô tả ý nghĩa các thuộc tính trong tập dữ liệu
2. Thiết kế & cài đặt CSDL (NDS, DDS)
3. Cài đặt ETL nạp dữ liệu
4. Report
5. MDX, OLAP
6. Mining

# **Kết quả**

1. **Mô tả ý nghĩa các thuộc tính trong tập dữ liệu**
   1. **Sheet ‘supermarket\_sales’**

|  |  |
| --- | --- |
| **Thuộc tính** | **Ý nghĩa** |
| **Invoice ID** | Mã số định danh hóa hóa đơn bán hàng. |
| **Branch** | Chi nhánh cửa hàng (tương tự như trong sheet 'city' - dùng để liên kết với các thuộc tính trong sheet 'city') |
| **Customer type** | Loại khách hàng (có thể là retail hoặc wholesale). |
| **Gender** | Giới tính của khách hàng |
| **ProductID** | Mã số định danh sản phẩm (tương tự như trong sheet ‘product’ - dùng để liên kết với các thuộc tính trong sheet ‘product’) |
| **Quantity** | Số lượng sản phẩm được bán |
| **Tax 5%** | Thuế 5% áp dụng cho sản phẩm. |
| **Total** | Tổng giá trị của hóa đơn sau khi tính thuế. |
| **Date** | Ngày mua hàng. |
| **Time** | Thời gian mua hàng. |
| **Payment** | Phương thức thanh toán. |
| **cogs** | Cost of Goods Sold - Chi phí của hàng bán được, tức giá vốn của sản phẩm |
| **gross margin percentage** | Tỷ lệ lợi nhuận gộp (gross margin) tính bằng công thức: (Total - cogs) / Total \* 100 |
| **gross income** | Lợi nhuận gộp, tức là tổng giá trị hóa đơn trước thuế. |
| **Rating** | Điểm đánh giá từ khách hàng |

* 1. **Sheet ‘product’**

|  |  |
| --- | --- |
| **Thuộc tính** | **Ý nghĩa** |
| **ProductID** | Mã số định danh sản phẩm (tương tự như trong sheet 'supermarket\_sales' - dùng để liên kết với các thuộc tính trong sheet 'supermarket\_sales'). |
| **Unit price** | Giá của một đơn vị sản phẩm |

* 1. **Sheet ‘ProductLine’**

|  |  |
| --- | --- |
| **Thuộc tính** | **Ý nghĩa** |
| **Product line** | Loại sản phẩm hoặc dòng sản phẩm. |
| **ProductLineID** | Mã số định danh của dòng sản phẩm |

* 1. **Sheet ‘city’**

|  |  |
| --- | --- |
| **Thuộc tính** | **Ý nghĩa** |
| **Branch** | Chi nhánh cửa hàng (tương tự như trong sheet 'supermarket\_sales' – dùng để liên kết với các thuộc tính trong sheet 'supermarket\_sales') |
| **City** | Thành phố tương ứng với chi nhánh. |

1. **Tiền xử lí file ‘supermarket\_sales.xls’**
   1. **Phân tích**

Theo yêu cầu của đồ án, nhóm phải thực hiện trên file excel ‘supermarket\_sales.xls’ tuy nhiên qua quá trình quan sát nhóm có một vài nhận định sau:  
**\* Vấn đề thứ nhất:** Cột “Gender” mang dữ liệu không đồng đều, cụ thể: trong các dòng của cột “Gender” chứa cả bốn kiểu giá trị: “Male - Female”, “M - F” nhưng thực tế chỉ cần chứa hai kiểu giá trị: “Male - Female” hoặc “ M - F”

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

**Vì vậy:** Nhóm thống nhất sẽ convert các giá trị về thành hai giá trị thống nhất là “Male - Female”

**\* Vấn đề thứ hai:** Cột “Date” mang dữ liệu không đồng đều, cụ thể: trong các dòng của cột “Date” chứa hai kiểu dữ liệu về ngày tháng năm: “1/5/2019” hoặc “6-Mar-19” nhưng thực tế chỉ cần mang một trong hai kiểu giá trị: “1/5/2019” hoặc “6-Mar-19”

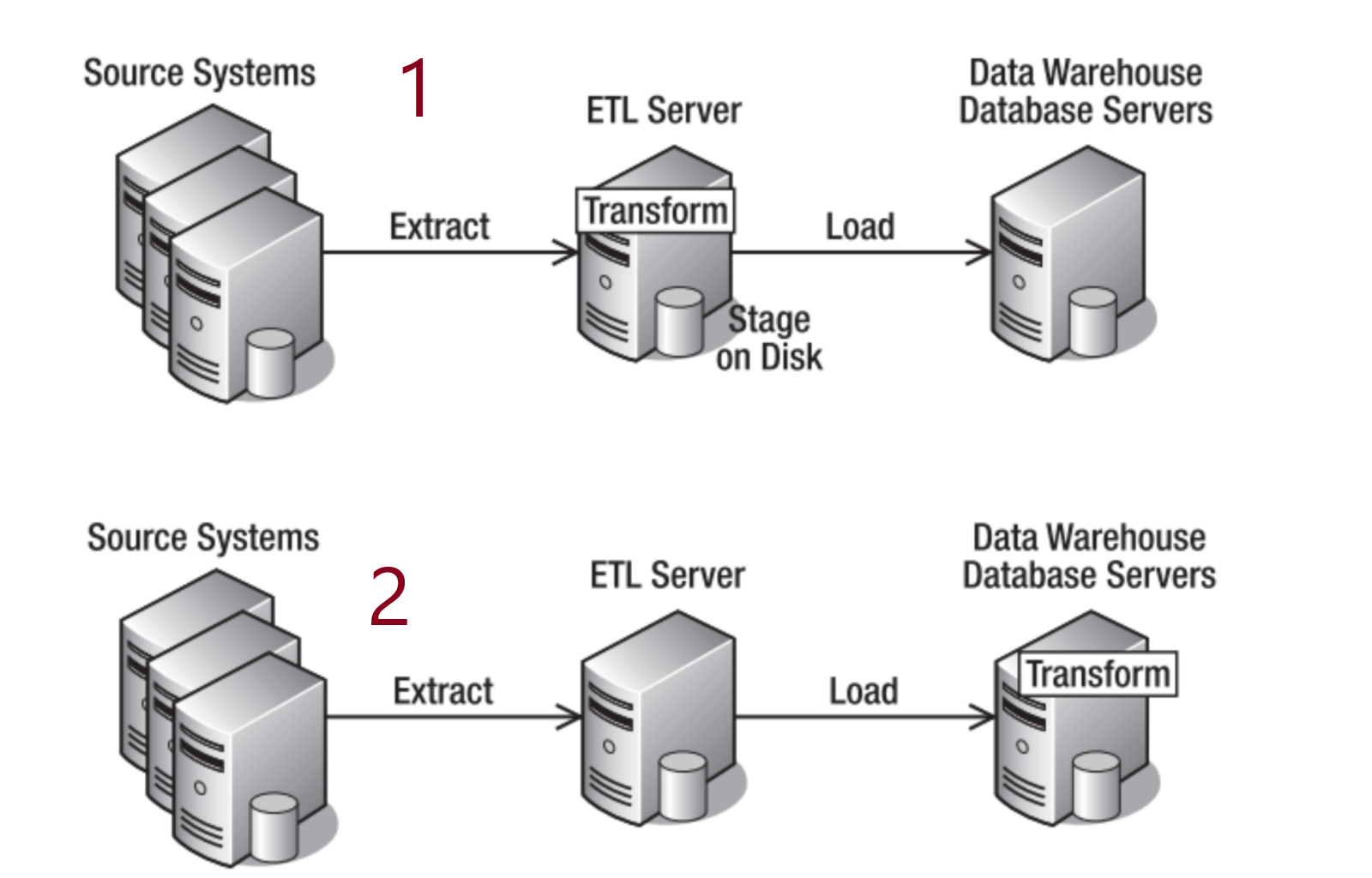
**Vì vậy:** Nhóm thống nhất sẽ convert kiểu dữ liệu “6-Mar-19” về thành kiểu dữ liệu thống thống nhất là “1/5/2019”**Ảnh có chứa văn bản, Phông chữ, ảnh chụp màn hình, số

Mô tả được tạo tự động**. Đồng thời, tách cột “Date” thành các cột riêng biệt “Day”, “Month” và “Year” để phục vụ cho việc ETL trở nên dễ dàng hơn

* 1. **Thực thi**

1. **Tạo metadata**
2. **ETL từ Source sang Stage**
   1. **Kiến trúc**

Kiến trúc mà nhóm tiếp cận là kiến trúc số 2 vì dữ liệu sử dụng ở đây là thuộc loại dữ liệu E-commerce nên trong thực tế lượng dữ liệu sẽ lớn lên một cách nhanh chóng… Vì vậy thực hiện transform (xử lý) ở memory sẽ gây chậm cho hệ thống về sau này. Stage chỉ có nhiệm vụ nhận dữ liệu sau đó sẽ chuyển lên Data WareHouse mà không qua xử lí. Data WareHouse có nhiệm vụ nhận dữ liệu rồi sau đó transform



Ở đây đối tượng chủ động để lấy dữ liệu từ source system là ETL Server viết những thủ tục, chương trình để lấy dữ liệu từ source system một cách đều đặn. Source System đóng vai trò bị động. Hình ảnh dưới đây minh họa thể hiện rõ đối tượng lấy dữ liệu từ source system:Ảnh có chứa màu trắng, ảnh chụp màn hình, Phông chữ, văn bản

Mô tả được tạo tự động

* 1. **Cấu trúc của Stage**

Theo yêu cầu của đồ án, quá trình ETL từ Source vào Stage sẽ sử dụng 1 file excel source là file ‘supermarket\_sales.xls’ và đã được nhóm tinh chỉnh sang file ‘SupermarketSales\_Cleaned.xlsx’

Vì chỉ sử dụng một file source nên Stage sẽ được tạo tương ứng với 4 sheets ở trong file, cụ thể như sau:

**Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động**

* 1. **ETL**

❖ **Các bước thực hiện đổ từ Source vào Stage:**

**1.** Cập nhật giá trị CET của bảng tương ứng trong Metadata bằng với thời điểm thực thi (GetDate())

**UPDATE ETL\_METADATA\_STAGE**

**SET CET = getDate()**

**where TableName = 'STAGE'**

**2.** Lấy giá trị LSET, CET trong Metadata với bảng tương ứng

**SELECT LSET, CET**

**FROM ETL\_METADATA\_STAGE**

**WHERE TableName = 'CITY\_STAGE'**

**Với: Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động**

**3.** Truncate bảng Stage tương ứng hiện tại: **TRUNCATE TABLE <”Stage”>**

**4.** Rút trích dữ liệu từ file excel bằng “Excel Source” cụ thể vào trong bảng Stage tương ứng “ADO NET Destination”

**5.** Cập nhật LSET = CET trong Metadata với bảng Stage tương ứng

**UPDATE ETL\_METADATA\_STAGE**

**SET LSET = CET**

**WHERE TableName = 'CITY\_STAGE'**

**Flow khi thực hiện trên Visual Studio:  
Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, hàng

Mô tả được tạo tự động**

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, Phông chữ

Mô tả được tạo tự động

**Cụ thể trong Data Flow:  
Ảnh có chứa văn bản, ảnh chụp màn hình, hàng, Phông chữ

Mô tả được tạo tự động** **Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động** Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

### **Kết quả**

**Trước khi thực thi ETL:**

**Sau khi thực thi ETL:**

1. **Stage sang NDS**
2. **NDS sang DDS**
3. **Khai thác dữ liệu**
   1. **Report**
   2. **OLAP & MDX**
   3. **Mining**

# **TÀI LIỆU THAM KHẢO**