

PHÂN TÍCH TÌNH HÌNH BÁN HÀNG CỦA SIÊU THỊ

MỤC LỤC

MỤC LỤC.....	1
1. Thiết kế kho dữ liệu, NDS, DDS	2
1.1. Kho dữ liệu	2
1.2. Thiết kế NDS	3
1.3. Thiết kế DDS	13
2. Quy trình ETL.....	22
2.1. Nạp dữ liệu vào Source	22
2.2. Từ Source vào Stage.....	24
2.3. Từ Stage vào NDS.....	25
2.4. Từ NDS vào DDS.....	30
3. OLAP	39
3.1. Thiết kế OLAP Cube	39
3.2. Tình hình mua hàng của khách hàng theo từng chi nhánh, từng loại sản phẩm, theo thời gian, hình thức thanh toán.....	40
4. Report.....	43
4.1. Thống kê số thành viên mua hàng theo ngày, tháng, năm	43
4.2. Thống kê doanh thu của khách hàng (member, normal) theo ngày, tháng, năm và theo chi nhánh	47
4.3. Thống kê số lượng khách thanh toán theo cash/debit/... ở từng chi nhánh theo từng tháng trong năm	53
4.4. Thống kê số lượng rating của khách hàng (member, normal) theo từng loại sản phẩm (ProductLine).....	56
4.5. Thống kê số lượng sản phẩm bán được theo từng thời điểm (time)	57
4.6. Thống kê số lượng khách hàng nữ đã mua theo từng loại sản phẩm.....	61
5. Tạo job chạy tự động ETL	63

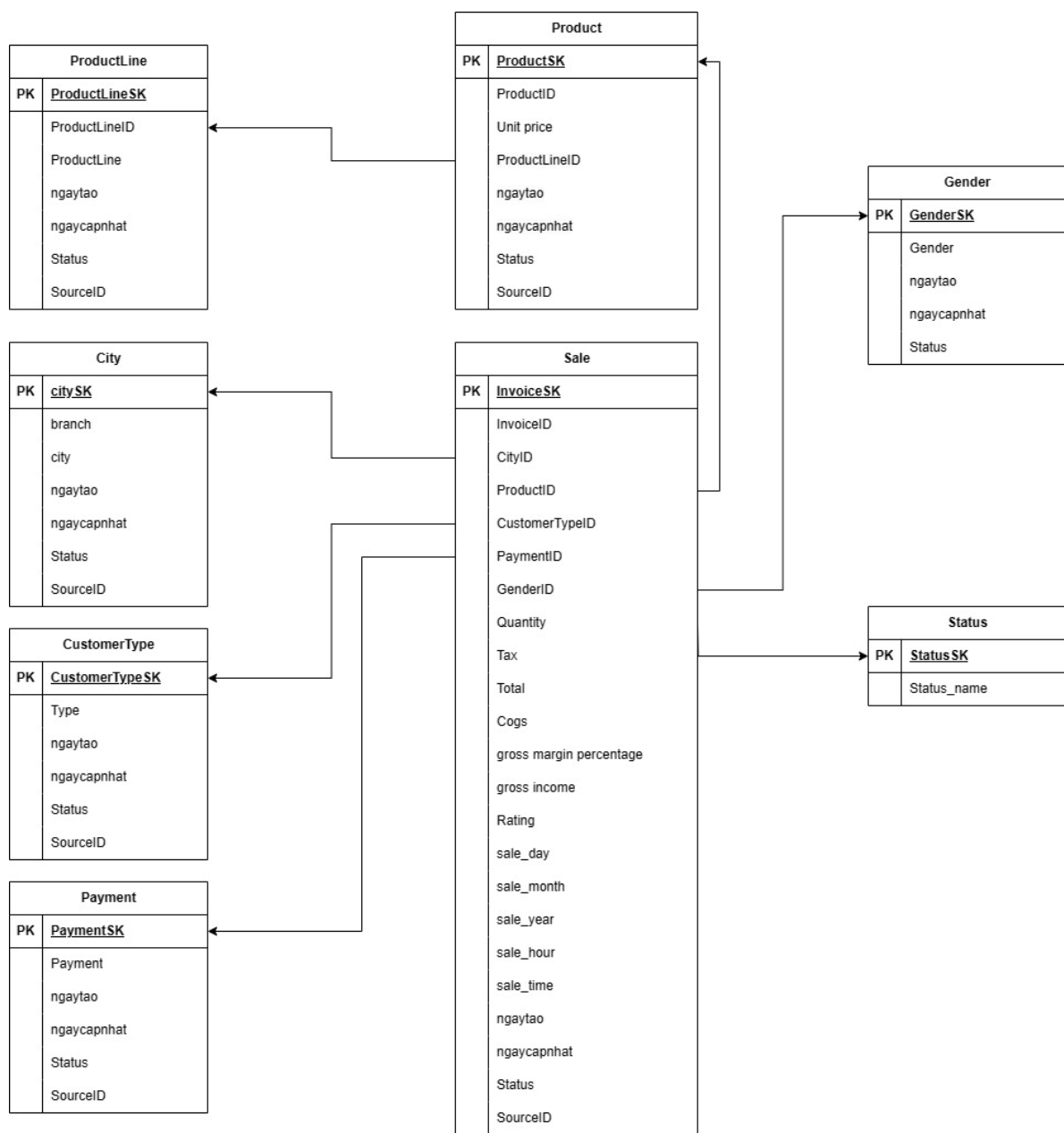
1. Thiết kế kho dữ liệu, NDS, DDS

1.1. Kho dữ liệu

Kho dữ liệu bao gồm:

- 1 database Source: để lưu tất cả dữ liệu từ các nguồn khác nhau phục vụ cho việc incremental extract vào Stage
- 1 database Stage: trung gian giữa Source và NDS, chỉ lấy những dữ liệu mới được thêm vào hoặc mới được cập nhật để thêm vào NDS.
- 1 database NDS: để lưu dữ liệu trong các bảng đã được chuẩn hoá ở dạng chuẩn 3 trở lên.
- 1 database DDS: để phục vụ tạo OLAP Cube.
- 1 database Metadata: để lưu thông tin thời gian thực hiện ETL thành công gần nhất vào Stage và vào DDS.

1.2. Thiết kế NDS



Bảng ProductLine

STT	Thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả
1	ProductLineSK	integer		Khóa tự tăng của bảng ProductLine
2	ProductLineID	nvarchar	Từ bảng product_line_Stage	Mã loại sản phẩm trong source system
3	ngaytao	datetime		Thời gian mà record được tạo
4	ngaycapnhat	datetime		Thời gian mà record được cập nhật
5	status	integer		Trạng thái của dòng record
6	sourceID	integer	Từ bảng product_line_Stage	Nguồn chứa record

Bảng Product

STT	Thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả
1	ProductSK	integer		Khóa tự tăng của bảng Product
2	ProductID	nvarchar	Từ bảng product_Stage	Mã sản phẩm trong source system
3	UnitPrice	decimal (8,2)	Từ bảng product_Stage	Giá cả của từng sản phẩm
4	ProductLineID	integer	Từ bảng product_line_Stage	Loại sản phẩm của dòng record
5	ngaytao	datetime		Thời gian mà record được tạo

6	ngaycapnhat	datetime		Thời gian mà record được cập nhật
7	status	integer		Trạng thái của dòng record
8	sourceID	integer	Từ bảng product_Stage	Nguồn chứa record

Bảng CustomerType

STT	Thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả
1	CustomerTypeSK	integer		Khóa tự tăng của bảng CustomerType
2	customerType	nvarchar	Từ bảng supermarket_sale_Stage	Loại khách hàng
3	ngaytao	datetime		Thời gian mà record được tạo
4	ngaycapnhat	datetime		Thời gian mà record được cập nhật
5	status	integer		Trạng thái của dòng record
6	sourceID	integer	Từ bảng supermarket_sale_Stage	Nguồn chứa record

Bảng City

STT	Thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả
1	citySK	integer		Khóa tự tăng của bảng City
2	branch	nvarchar	Từ bảng city_Stage	Mã thành phố
3	city	nvarchar	Từ bảng city_Stage	Tên thành phố
4	ngaytao	datetime		Thời gian mà record được tạo
5	ngaycapnhat	datetime		Thời gian mà record được cập nhật
6	status	integer		Trạng thái của dòng record
7	SourceID	integer	Từ bảng city_Stage	Nguồn chứa record

Bảng Payment

STT	Thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả
1	PaymentSK	integer		Khóa tự tăng của bảng Payment
2	Payment	nvarchar	Từ bảng supermarket_sale_Stage	Phương thức thanh toán
3	ngaytao	datetime		Thời gian mà record được tạo
4	ngaycapnhat	datetime		Thời gian mà record được cập nhật
5	Status	integer		Trạng thái của dòng record
6	SourceID	integer	Từ bảng supermarket_sale_Stage	Nguồn chứa record

Bảng Gender

STT	Thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả
1	GenderSK	integer		Khóa tự tăng của bảng Gender
2	Gender	nvarchar	Từ bảng supermarket_sale_Stage	Giới tính
3	ngaytao	datetime		Thời gian mà record được tạo
4	ngaycapnhat	datetime		Thời gian mà record được cập nhật
5	Status	integer		Trạng thái của dòng record
6	SourceID	integer	Từ bảng supermarket_sale_Stage	Nguồn chứa record

Bảng Status

STT	Thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả
1	StatusSK	integer		Khóa tự tăng của bảng Status
2	Status_name	nvarchar	Từ bảng supermarket_sale_Stage	Trạng thái

Bảng Sale

STT	Thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả
1	InvoiceSK	integer		Khóa tự tăng của bảng InvoiceSK
2	InvoiceID	nvarchar	Từ bảng supermarket_sale_Stage	Mã hóa đơn
3	CityID	integer	Từ bảng city_Stage	Khóa ngoại tới bảng City thể hiện thành phố
4	ProductID	integer	Từ bảng product_Stage	Khóa ngoại tới bảng Product thể hiện sản phẩm
5	CustomerTypeID	integer	Từ bảng supermarket_sale_Stage	Khóa ngoại tới bảng CustomerType thể hiện loại khách hàng
6	PaymentID	integer	Từ bảng supermarket_sale_Stage	Khóa ngoại tới bảng Payment thể hiện phương thức thanh toán
7	GenderID	integer	Từ bảng supermarket_sale_Stage	Khóa ngoại tới bảng Gender thể hiện giới tính khách hàng mua
8	Quantity	integer	Từ bảng supermarket_sale_Stage	Số lượng sản phẩm khách hàng mua
9	Tax	decimal (8,2)	Từ bảng supermarket_sale_Stage	Phí thuế cho khách hàng mua hàng (5%)

10	Total	decimal (8,2)	Từ bảng supermarke t_sale_Stag e	Tổng giá bao gồm thuế
11	Cogs	decimal (8,2)	Từ bảng supermarke t_sale_Stag e	Giá vốn hàng bán
12	gross margin percentage	decimal (5,2)	Từ bảng supermarke t_sale_Stag e	Tỷ lệ lợi nhuận gộp
13	gross income	decimal (8,2)	Từ bảng supermarke t_sale_Stag e	Tổng thu nhập
14	rating	decimal (4,1)	Từ bảng supermarke t_sale_Stag e	Xếp hạng của khách hàng (Trên thang điểm từ 1 đến 10)
15	Sale_day	integer	Từ bảng supermarke t_sale_Stag e	Ngày mua của record hóa đơn
16	Sale_mont h	integer	Từ bảng supermarke t_sale_Stag e	Tháng mua của record hóa đơn
17	Sale_year	integer	Từ bảng supermarke t_sale_Stag e	Năm mua của record hóa đơn

18	Sale_hour	integer	Từ bảng supermarket_sale_Stage	Giờ mua của record hoá đơn
19	Sale_time		Từ bảng supermarket_sale_Stage	Thời gian mua của record hoá đơn
20	ngaytao	datetime		Thời gian mà record được tạo
21	ngaycapnh at	datetime		Thời gian mà record được cập nhật
22	Status	integer		Trạng thái của dòng record
23	SourceID	integer	Từ bảng supermarket_sale_Stage	Nguồn chứa record

ProductLine (ProductLineID, ProductLine, ngaytao, ngaycapnh
at, Status, SourceID)

PHỤ THUỘC HÀM $F = \{ \text{ProductLineID} \Rightarrow \text{ProductLine}, \text{ngaytao}, \text{ngaycapnh}, \text{at}, \text{Status}, \text{SourceID} \}$

DẠNG CHUẨN: BCK

Product (ProductID, ProductLineID, UnitPrice, ngaytao, ngaycapnh
at, Status, SourceID)

PHỤ THUỘC HÀM $F = \{ \text{ProductID} \Rightarrow \text{UnitPrice}, \text{ProductLineID}, \text{ngaytao}, \text{ngaycapnh}, \text{at}, \text{Status}, \text{SourceID} \}$

DẠNG CHUẨN: BCK

Gender (GenderSK, Gender, ngaytao, ngaycapnh
at, Status)

PHỤ THUỘC HÀM $F = \{GenderSK \Rightarrow Gender, ngaytao, ngaycapnhat, Status\}$

DẠNG CHUẨN: BCK

City (CitySK, branch, city, ngaytao, ngaycapnhat, Status, SourceID)

PHỤ THUỘC HÀM $F = \{CitySK \Rightarrow branch, city, ngaytao, ngaycapnhat, Status, SourceID\}$

DẠNG CHUẨN: BCK

CustomerType (CustomerTypeSK, Type, ngaytao, ngaycapnhat, Status, SourceID)

PHỤ THUỘC HÀM $F = \{CustomerTypeSK \Rightarrow Type, ngaytao, ngaycapnhat, Status, SourceID\}$

DẠNG CHUẨN: BCK

Payment (PaymentSK, Payment, ngaytao, ngaycapnhat, Status, SourceID)

PHỤ THUỘC HÀM $F = \{PaymentSK \Rightarrow Payment, ngaytao, ngaycapnhat, Status, SourceID\}$

DẠNG CHUẨN: BCK

Status (StatusSK, Status_name)

PHỤ THUỘC HÀM $F = \{StatusSK \Rightarrow Status_name\}$

DẠNG CHUẨN: BCK

Sale (InvoiceID, CityID, ProductID, CustomerTypeID, PaymentID, GenderID, Quantity, Tax, Total, Cogs, gross_margin_percentage, gross_income, Rating, sale_day, sale_month, sale_year, sale_hour, sale_time, ngaytao, ngaycapnhat, Status, SourceID)

PHỤ THUỘC HÀM $F = \{ \text{InvoiceID} \Rightarrow \text{CityID}, \text{ProductID}, \text{CustomerTypeID}, \text{PaymentID}, \text{GenderID}, \text{Quantity}, \text{Tax}, \text{Total}, \text{Cogs}, \text{gross_margin_percentage}, \text{gross_income}, \text{Rating}, \text{sale_day}, \text{sale_month}, \text{sale_year}, \text{sale_hour}, \text{sale_time}, \text{ngaytao}, \text{ngaycapnhat}, \text{Status}, \text{SourceID} \}$

DẠNG CHUẨN: BCK

1.3. Thiết kế DDS

Phân tích yêu cầu:

- Sự kiện: khi khách hàng mua 1 sản phẩm
- Bối cảnh sự kiện:
 - Ai: khách hàng
 - Khi nào: ngày mua hàng
 - Cái gì: sản phẩm
 - Ở đâu: tại cửa hàng, thành phố
- Đo lường: doanh thu, rating, số lượng

Phân tích bảng Fact:

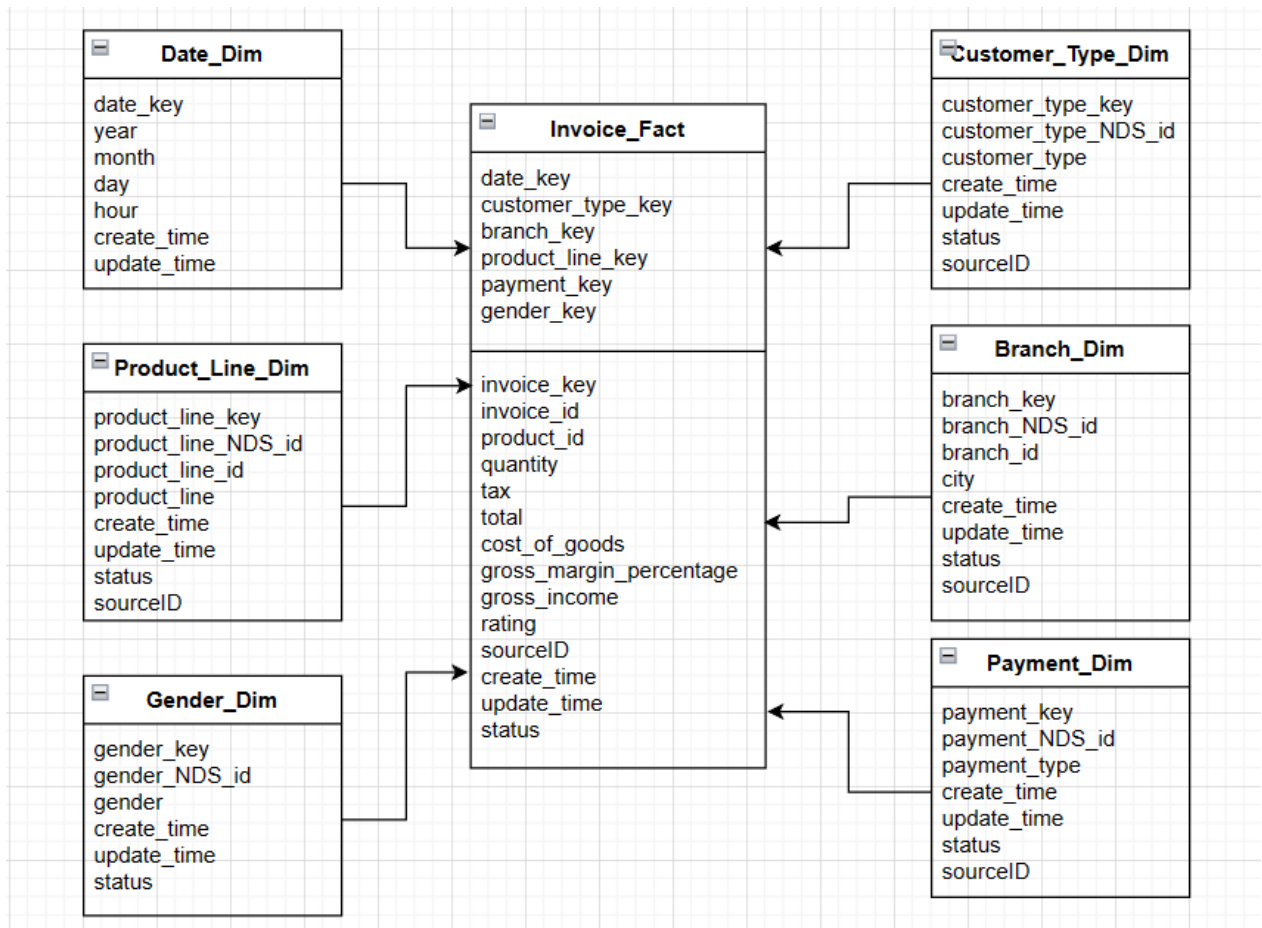
- Các giá trị có sẵn từ nguồn: gross_income, rating, quantity
- Các giá trị phải tính toán:
- Cấp chi tiết dữ liệu:
 - Đơn vị nhỏ nhất xảy ra sự kiện: một dòng trong fact table tương ứng một lần mua sản phẩm của một khách hàng tại một chi nhánh

Phân tích thiết kế chiều:

- Các chiều liên quan đến sự kiện phân tích:
 - Date dimension
 - Customer type dimension
 - Branch dimension
 - Payment dimension
 - Product line dimension
 - Gender dimension
- Branch dimension:

- Trong tương lai có thể thay đổi
- SCD loại 2: lưu giá trị lịch sử theo dòng
- Các chiều còn lại:
 - Không yêu cầu lưu lịch sử
 - SCD loại 1: lưu đề giá trị

Thiết kế DDS



Bảng dim_gender

Tên thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả	Transformation
gender_key	integer		Khóa tự tăng của bảng dim_gender	
gender_NDS_id	integer	Từ bảng gender_NDS	Khóa tự tăng lấy từ bảng gender_NDS	
gender	nvarchar	Từ bảng gender_NDS	Giới tính khách hàng	
create_time	datetime		Thời gian tạo trong DDS	
update_time	datetime		Thời gian cập nhật trong DDS	
status_type	integer	Từ bảng gender_NDS	Trạng thái của dòng	
sourceID	integer	Từ bảng gender_NDS	Mã của nguồn trong source system	

Bảng dim_product_line

Tên thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả	Transformation
product_line_key	integer		Khóa tự tăng của bảng dim_product_line	
product_line_NDS_id	integer	Từ bảng product_line_NDS	Khóa tự tăng lấy từ bảng	

			product_line_NDS	
product_line_id	nvarchar	Từ bảng product_line_NDS	Mã phân loại nhóm hàng	
product_line	nvarchar	Từ bảng product_line_NDS	Tên phân loại nhóm hàng	
create_time	datetime		Thời gian tạo trong DDS	
update_time	datetime		Thời gian cập nhật trong DDS	
status_type	integer	Từ bảng product_line_NDS	Trạng thái của dòng	
sourceID	integer	Từ bảng product_line_NDS	Mã của nguồn trong source system	

Bảng dim_date

Tên thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả	Transformation
date_key	integer		Khóa tự tăng của bảng dim_date	
d_year	integer	Từ bảng supermarket_sale_NDS	Năm mua hàng của hóa đơn	
d_month	integer	Từ bảng supermarket_sale_NDS	Tháng mua hàng	

			của hóa đơn	
d_day	integer	Từ bảng supermarket_sale_NDS	Ngày mua hàng của hóa đơn	
d_hour	integer	Từ bảng supermarket_sale_NDS	Giờ mua hàng của hóa đơn	
create_time	datetime		Thời gian tạo trong DDS	
update_time	datetime		Thời gian cập nhật trong DDS	

Bảng dim_payment

Tên thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả	Transformation
payment_key	integer		Khóa tự tăng của bảng dim_payment	
payment_NDS_id	integer	Từ bảng payment_NDS	Khóa tự tăng lấy từ bảng payment_NDS	
payment_type	nvarchar	Từ bảng payment_NDS	Phương thức thanh toán khi mua hàng	
create_time	datetime		Thời gian tạo trong DDS	

update_time	datetime		Thời gian cập nhật trong DDS	
status_type	integer	Từ bảng payment_NDS	Trạng thái của dòng	
sourceID	integer	Từ bảng payment_NDS	Mã của nguồn trong source system	

Bảng dim_branch

Tên thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả	Transformation
branch_key	integer		Khóa tự tăng của bảng dim_branch	
branch_NDS_id	integer	Từ bảng city_NDS	Khóa tự tăng lấy từ bảng city_NDS	
branch_id	nvarchar	Từ bảng city_NDS	Chi nhánh cửa hàng	
city	nvarchar	Từ bảng city_NDS	Thành phố của chi nhánh	
create_time	datetime		Thời gian tạo trong DDS	
update_time	datetime		Thời gian cập nhật trong DDS	
status_type	integer	Từ bảng city_NDS	Trạng thái của dòng	
sourceID	integer	Từ bảng city_NDS	Mã của nguồn trong source system	

Bảng **dim_customer_type**

Tên thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả	Transformation
customer_type_key	integer		Khóa tự tăng của bảng dim_customer_type	
customer_NDS_id	integer	Từ bảng customer_type_NDS	Khóa tự tăng lấy từ bảng customer_type_NDS	
customer_type	nvarchar	Từ bảng customer_type_NDS	Loại khách hàng	
create_time	datetime		Thời gian tạo trong DDS	
update_time	datetime		Thời gian cập nhật trong DDS	
status_type	integer	Từ bảng customer_type_NDS	Trạng thái của dòng	
sourceID	integer	Từ bảng customer_type_NDS	Mã của nguồn trong source system	

Bảng **fact_invoice**

Tên thuộc tính	Kiểu dữ liệu	Nguồn	Mô tả	Transformation
date_key	integer	Từ bảng dim_date	Khóa tự tăng lấy từ bảng dim_date	

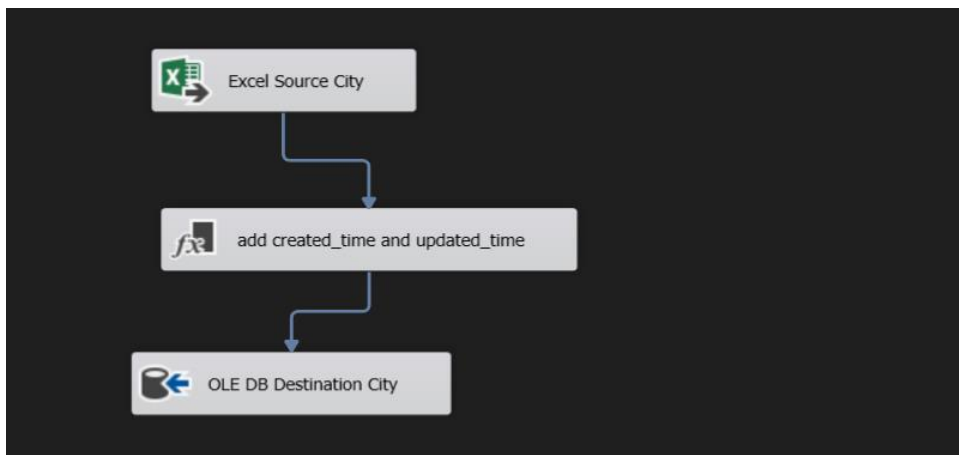
product_line_key	integer	Từ bảng dim_product_line	Khóa tự tăng lấy từ bảng dim_product_line	
gender_key	integer	Từ bảng dim_gender	Khóa tự tăng lấy từ bảng dim_gender	
payment_key	integer	Từ bảng dim_payment	Khóa tự tăng lấy từ bảng dim_payment	
branch_key	integer	Từ bảng dim_branch	Khóa tự tăng lấy từ bảng dim_branch	
customer_type_key	integer	Từ bảng dim_customer_type	Khóa tự tăng lấy từ bảng dim_customer_type	
invoice_key	integer		Khóa tự tăng của bảng fact_invoice	
invoice_NDS_key	integer	Từ bảng supermarket_sale_NDS	Khóa tự tăng lấy từ bảng supermarket_sale_NDS	
invoice_id	nvarchar	Từ bảng supermarket_sale_NDS	Mã hóa đơn	
product_id	integer	Từ bảng supermarket_sale_NDS	Mã sản phẩm	
quantity	integer	Từ bảng supermarket_sale_NDS	Số lượng sản phẩm khách hàng mua	

tax	decimal (8,2)	Từ bảng supermarket_sale_NDS	Phí thuế cho khách hàng mua hàng (5%)	
total	decimal (8,2)	Từ bảng supermarket_sale_NDS	Tổng giá bao gồm thuế	
cogs	decimal (8,2)	Từ bảng supermarket_sale_NDS	Giá vốn hàng bán	
grossMarginPercentage	decimal (5,2)	Từ bảng supermarket_sale_NDS	Tỷ lệ lợi nhuận gộp	
grossIncome	decimal (8,2)	Từ bảng supermarket_sale_NDS	Tổng thu nhập	
rating	decimal (4,1)	Từ bảng supermarket_sale_NDS	Xếp hạng phân tầng của khách hàng về trải nghiệm mua sắm tổng thể của họ (Trên thang điểm từ 1 đến 10)	
create_time	datetime		Thời gian tạo trong DDS	
update_time	datetime		Thời gian cập nhật trong DDS	
status_type	integer	Từ bảng supermarket_sale_NDS	Trạng thái của dòng	
sourceID	integer	Từ bảng supermarket_sale_NDS	Mã của nguồn trong source system	

2. Quy trình ETL

2.1. Nạp dữ liệu vào Source

2.1.1. Bảng City

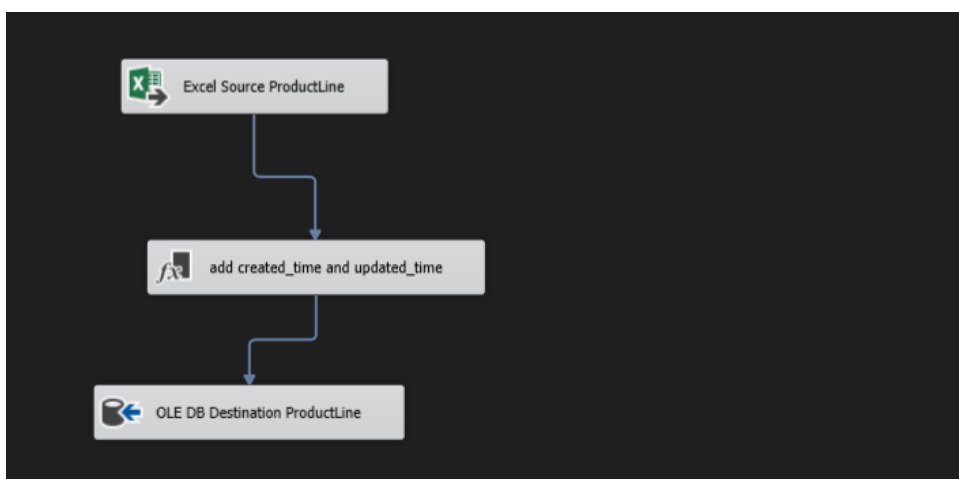


Bước 1: Lấy dữ liệu từ sheet city trong file excel supermarket_sales.

Bước 2: Thêm 2 trường dữ liệu là thời gian tạo và thời gian cập nhật.

Bước 3: Mapping trường dữ liệu từ file excel vào bảng city trong source.

2.1.2. Bảng Product Line

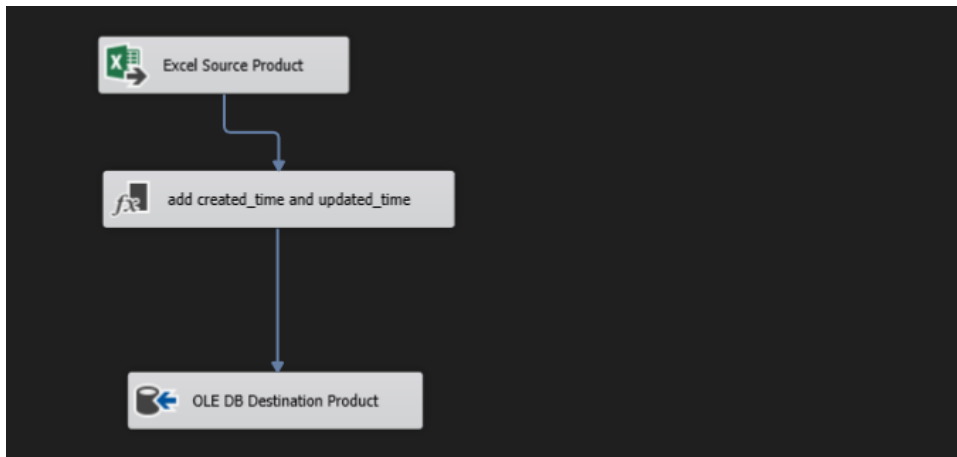


Bước 1: Lấy dữ liệu từ sheet product_line trong file excel supermarket_sales.

Bước 2: Thêm 2 trường dữ liệu là thời gian tạo và thời gian cập nhật.

Bước 3: Mapping trường dữ liệu từ file excel vào bảng product_line trong source.

2.1.3. Bảng Product

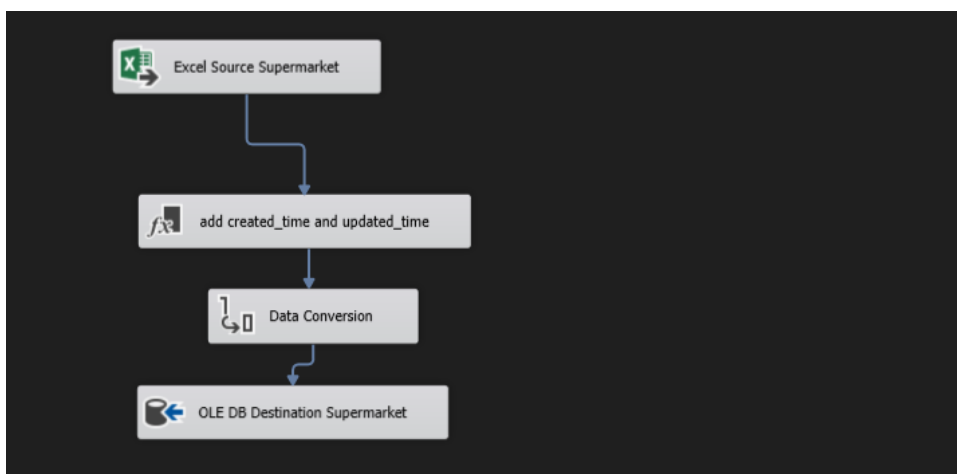


Bước 1: Lấy dữ liệu từ sheet product trong file excel supermarket_sales.

Bước 2: Thêm 2 trường dữ liệu là thời gian tạo và thời gian cập nhật.

Bước 3: Mapping trường dữ liệu từ file excel vào bảng product trong source.

2.1.4. Bảng Supermarket_Sale



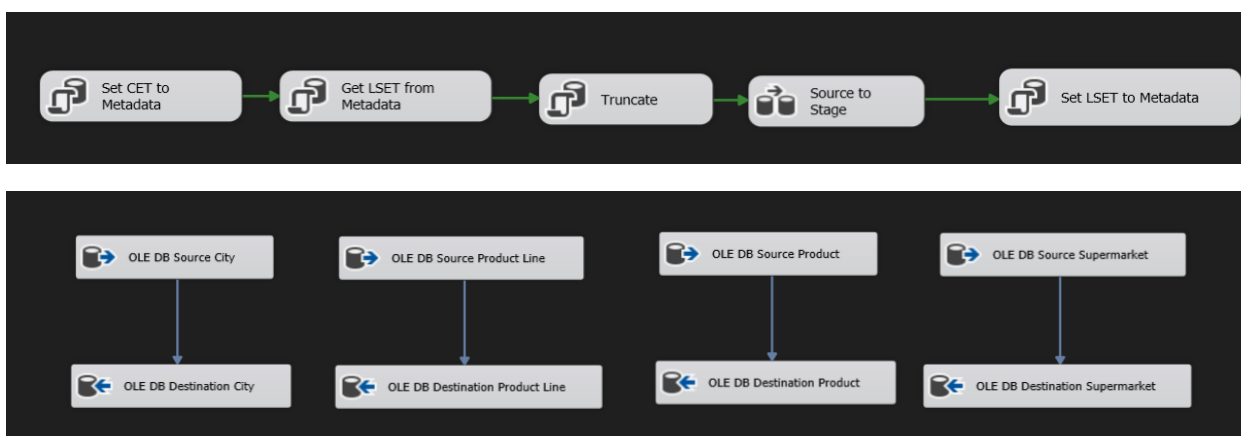
Bước 1: Lấy dữ liệu từ sheet supermarket_sales trong file excel supermarket_sales.

Bước 2: Thêm 2 trường dữ liệu là thời gian tạo và thời gian cập nhật.

Bước 3: Chuyển đổi kiểu dữ liệu của 2 trường Date, Time.

Bước 4: Mapping trường dữ liệu từ file excel vào bảng supermarket_sales trong source.

2.2. Từ Source vào Stage



Quy trình đổ dữ liệu từ Source sang Stage:

Bước 1: Cập nhật thông tin CET trong Metadata bằng thời gian khởi động Package.

Bước 2: Lấy thông tin LSET, CET được lưu trong Metadata.

Bước 3: Truncate các bảng trong Stage: TRUNCATE TABLE <Stage Table>.

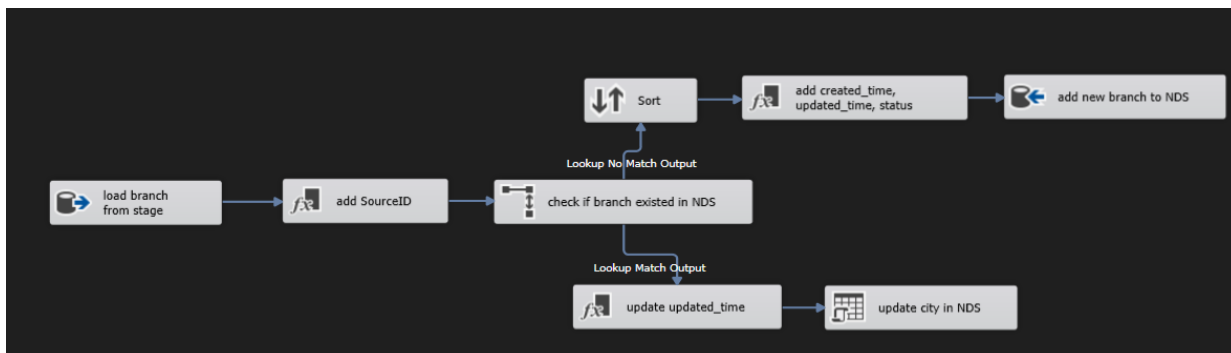
Bước 4: Rút trích dữ liệu từ bảng Source sang bảng Stage: SELECT * FROM <Source Table> where (create_time >= LSET and create_time < CET) or (update_time >= LSET and update_date < CET).

Bước 5: Cập nhật lại LSET trong Metadata.

2.3. Từ Stage vào NDS

2.3.1. Bảng City NDS

Quy trình đồ dữ liệu từ Stage sang NDS của bảng city_NDS:



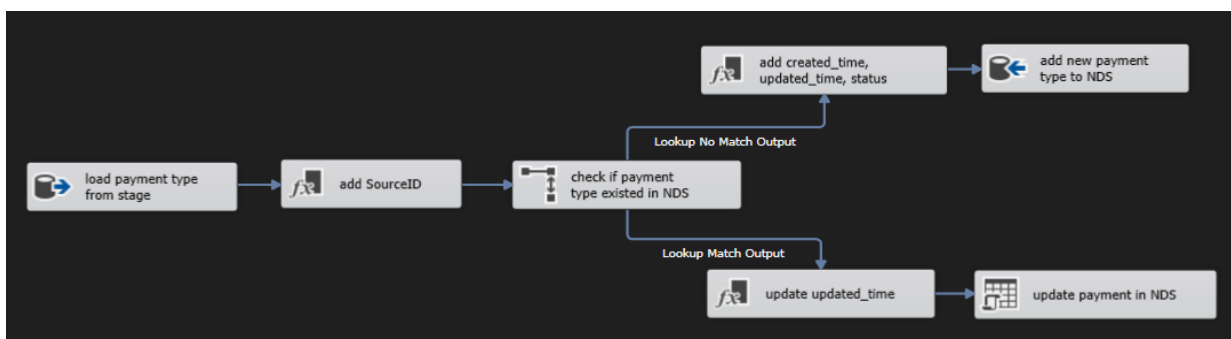
Bước 1: Đồ dữ liệu city_Stage vào để xử lý.

Bước 2: Thêm cột SourceID.

Bước 3: Kiểm tra xem branch đã từng tồn tại trong NDS chưa, nếu có thì cập nhật lại ngày cập nhật bằng thời gian hiện tại. Nếu chưa thì thêm các trường created_time, updated_time, status và thêm vào bảng city_NDS.

2.3.2. Bảng Payment NDS

Quy trình đồ dữ liệu từ Stage sang NDS của bảng payment_NDS:



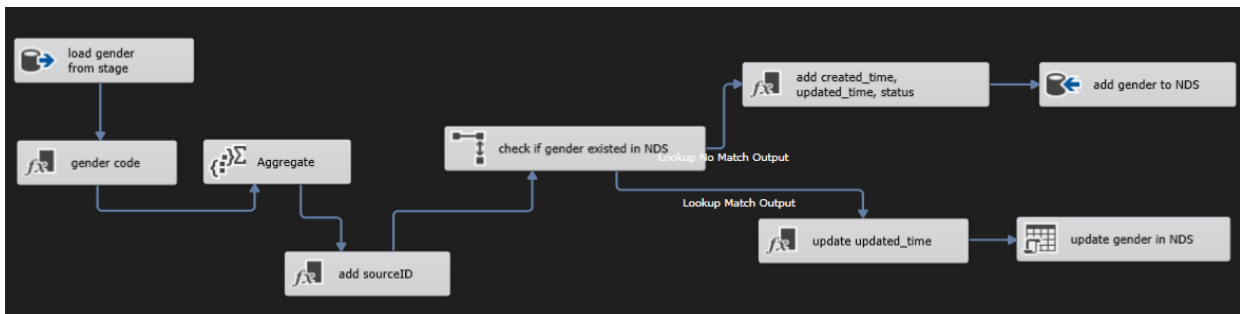
Bước 1: Đồ dữ liệu payment_Stage vào để xử lý.

Bước 2: Thêm cột SourceID.

Bước 3: Kiểm tra xem payment đã từng tồn tại trong NDS chưa, nếu có thì cập nhật lại ngày cập nhật bằng thời gian hiện tại. Nếu chưa thì thêm các trường created_time, updated_time, status và thêm vào bảng payment_NDS.

2.3.3. Bảng Gender NDS

Quy trình đổ dữ liệu từ Stage sang NDS của bảng gender_NDS:



Bước 1: Đổ dữ liệu gender_Stage vào để xử lý.

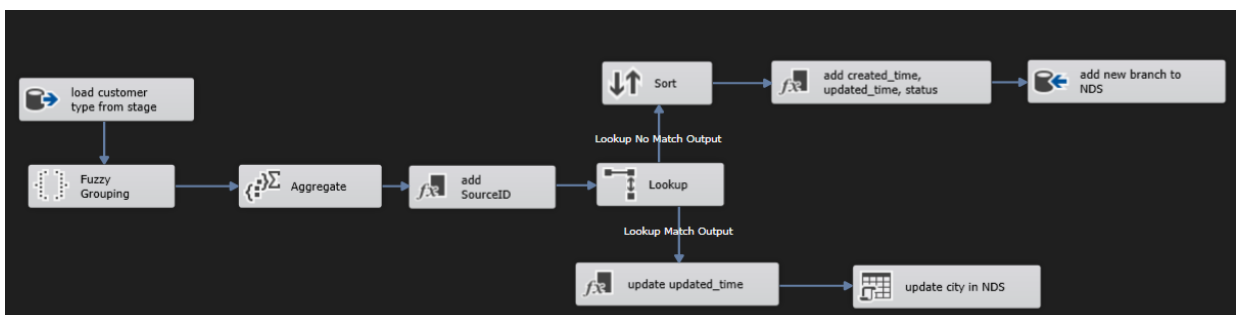
Bước 2: Xử lý dữ liệu giới tính thành ‘female’ và ‘male’ do nguồn dữ liệu lưu ở dạng ‘Female’, ‘F’, ‘Male’, ‘M’.

Bước 3: Thêm cột SourceID.

Bước 4: Kiểm tra xem gender đã từng tồn tại trong NDS chưa, nếu có thì cập nhật lại ngày cập nhật bằng thời gian hiện tại. Nếu chưa thì thêm các trường created_time, updated_time, status và thêm vào bảng gender_NDS.

2.3.4. Bảng CustomerType NDS

Quy trình đổ dữ liệu từ Stage sang NDS của bảng customer_type_NDS:



Bước 1: Đổ dữ liệu customer_type_Stage vào để xử lý.

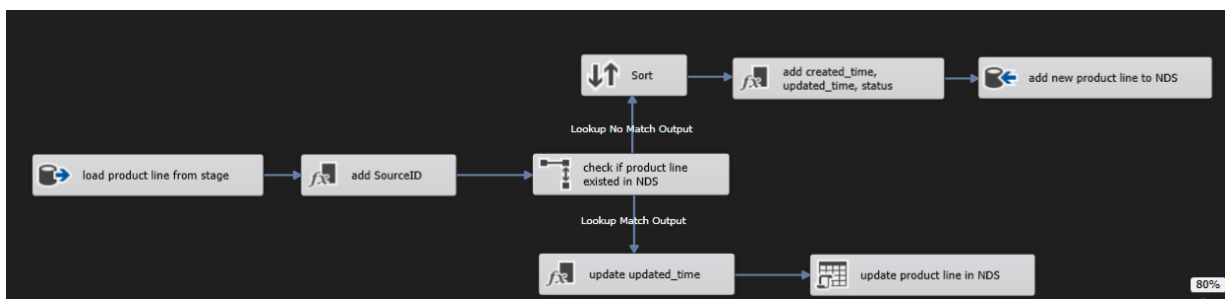
Bước 2: Nhóm dữ liệu lại.

Bước 3: Thêm cột SourceID trong data pipeline dữ liệu.

Bước 4: Kiểm tra xem customer_type đã từng tồn tại trong NDS chưa, nếu có thì cập nhật lại ngày cập nhật bằng thời gian hiện tại. Nếu chưa thì thêm các trường created_time, updated_time, status và thêm vào bảng customer_type_NDS.

2.3.5. Bảng ProductLine NDS

Quy trình đổ dữ liệu từ Stage sang NDS của bảng product_line_NDS:



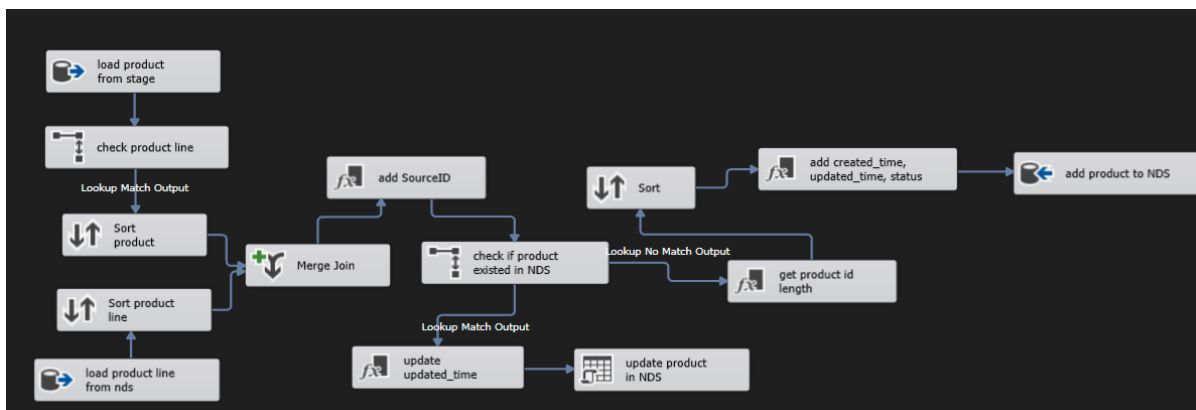
Bước 1: Đổ dữ liệu product_line_Stage vào để xử lý.

Bước 2: Thêm cột SourceID trong data pipeline dữ liệu.

Bước 3: Kiểm tra xem product_line đã từng tồn tại trong NDS chưa, nếu có thì cập nhật lại ngày cập nhật bằng thời gian hiện tại. Nếu chưa thì thêm các trường created_time, updated_time, status và thêm vào bảng product_line_NDS.

2.3.6. Bảng Product NDS

Quy trình đổ dữ liệu từ Stage sang NDS của bảng product_NDS:



Bước 1: Đổ dữ liệu product_Stage và product_line_NDS vào để xử lý.

Bước 2: Kiểm tra product line đã có trong product chưa.

Bước 3: Sắp xếp các dòng dữ liệu từ product_Stage và product_line_NDS theo product line.

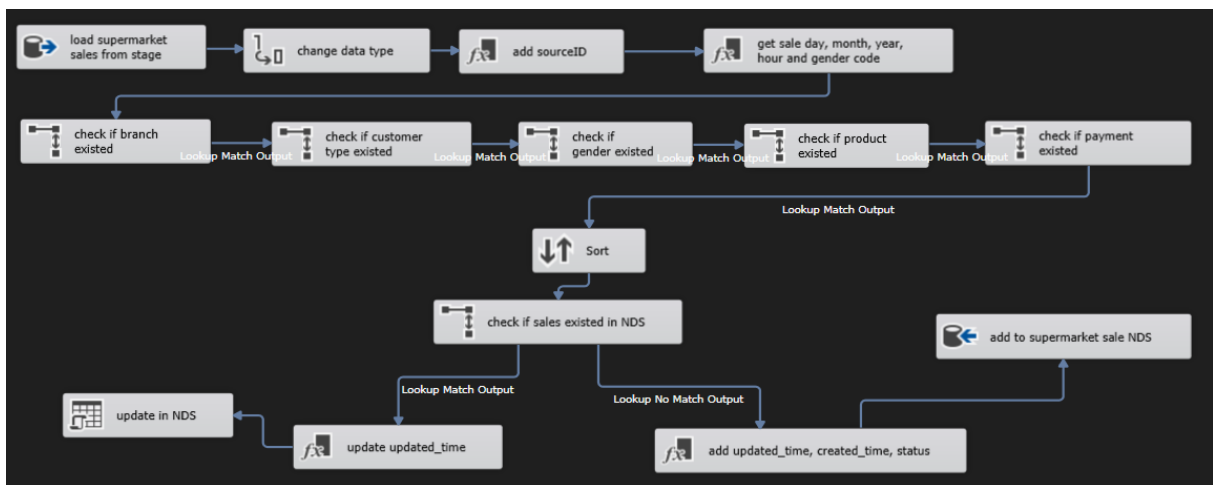
Bước 3: Merge Join 2 bảng với nhau để lấy id của product line và thêm cột SourceID.

Bước 4:

- Kiểm tra xem product đã từng tồn tại trong NDS chưa, nếu có thì cập nhật lại ngày cập nhật của dòng bằng thời gian hiện tại.
- Nếu chưa thì thêm các trường created_time, updated_time, status và thêm vào bảng product_NDS.

2.3.7. Bảng Sale NDS

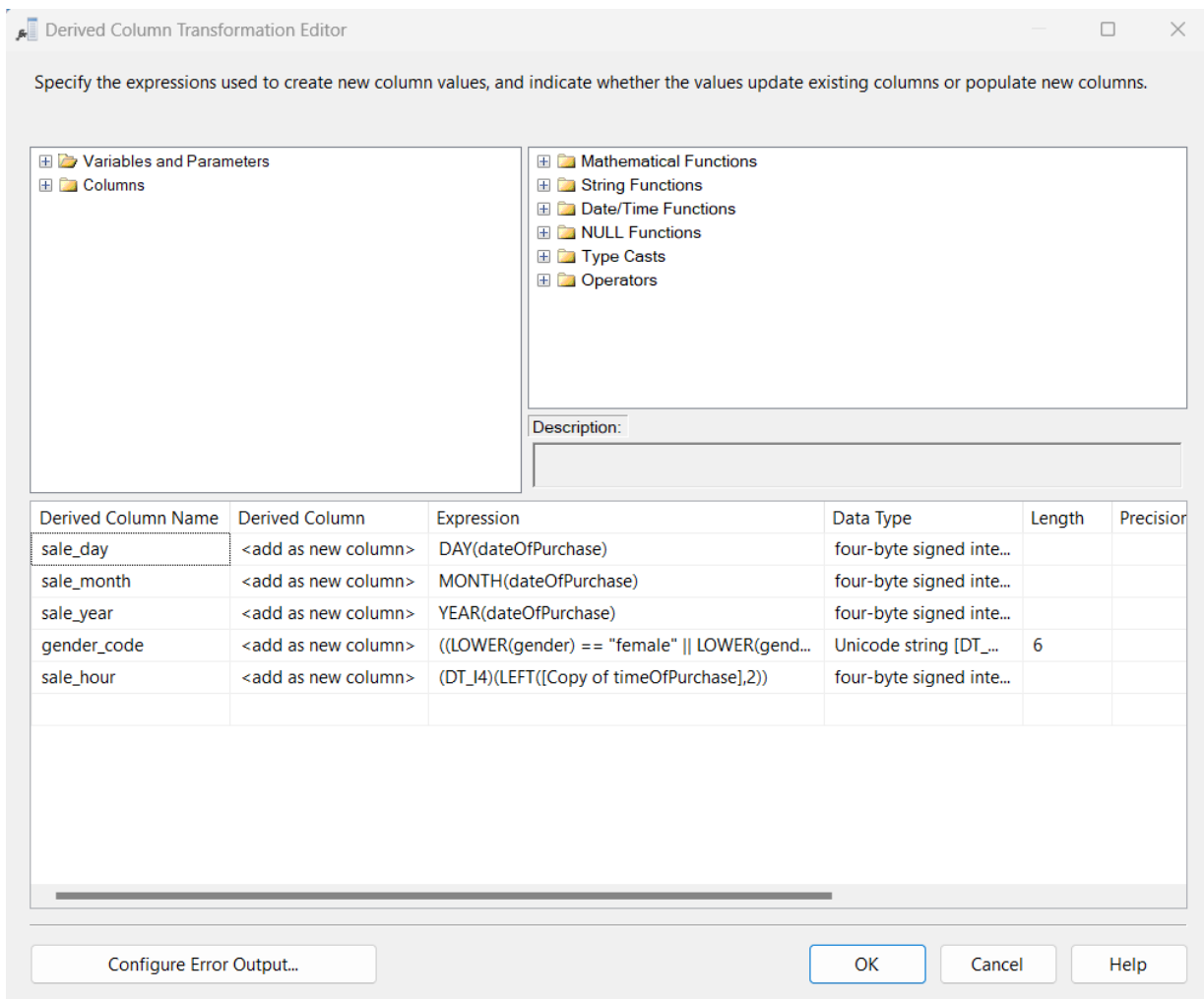
Quy trình đổ dữ liệu từ Stage sang NDS của bảng supermarket_sale_NDS:



Bước 1: Đổ dữ liệu supermarket_sales_Stage vào để xử lý.

Bước 2: Chỉnh sửa kiểu của các trường dữ liệu từ dữ liệu nguồn và thêm cột SourceID.

Bước 3: Sau khi có được dữ liệu, ta đổ dữ liệu vào các trường dữ liệu sale_day, sale_month, sale_year, gender_code, sale_hour.



Bước 4: Kiểm tra tiếp branch, customer_type, gender, product, payment đã tồn tại trong NDS.

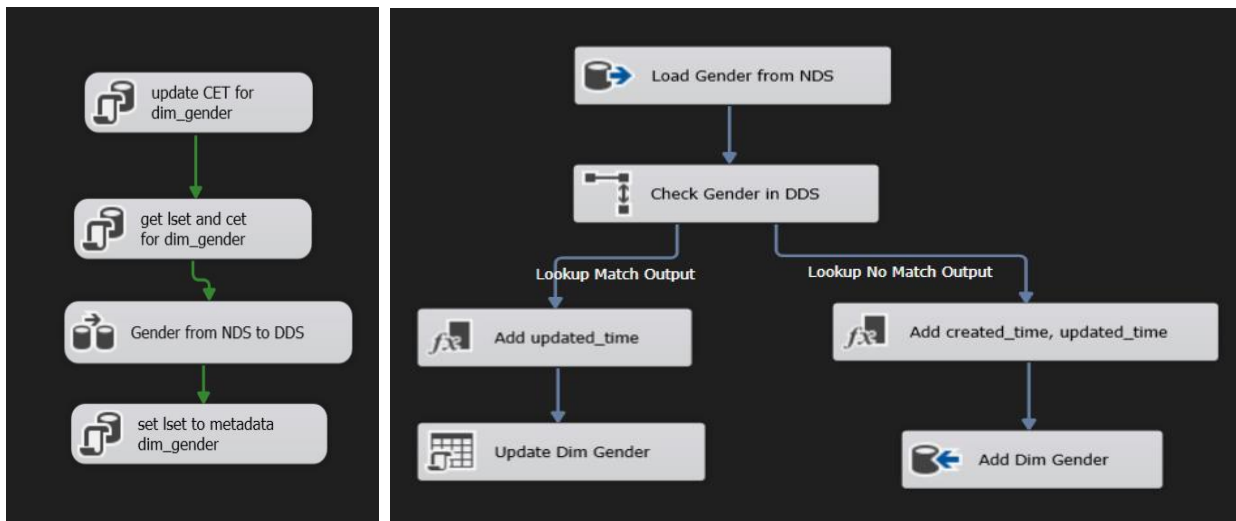
Bước 5:

- Kiểm tra xem sales đã từng tồn tại trong NDS chưa, nếu có thì cập nhật lại ngày cập nhật của dòng bằng thời gian hiện tại.
- Nếu chưa thì thêm các trường created_time, updated_time, status và thêm vào bảng sales_NDS.

2.4. Từ NDS vào DDS

2.4.1. Bảng Gender Dim

Quy trình đồ dữ liệu từ NDS sang DDS của bảng dim_gender:



Bước 1: Cập nhật thông tin CET cho bằng thời gian khởi động Package.

Bước 2: Lấy thông tin LSET, CET được lưu trong Metadata.

Bước 3: Tiến hành đồ dữ liệu từ NDS vào DDS cho bảng dim_gender:

- Đầu tiên lấy dữ liệu mới cập nhật hoặc mới được thêm vào từ bảng gender_NDS
select *

from gender_NDS

where (ngaytao >= lset and ngaytao <= cet) or (ngaycapnhat >= lset and ngaycapnhat <= cet)

- Sau đó tiến hành kiểm tra xem từng dòng Gender đã từng tồn tại trong bảng dim_gender hay không.

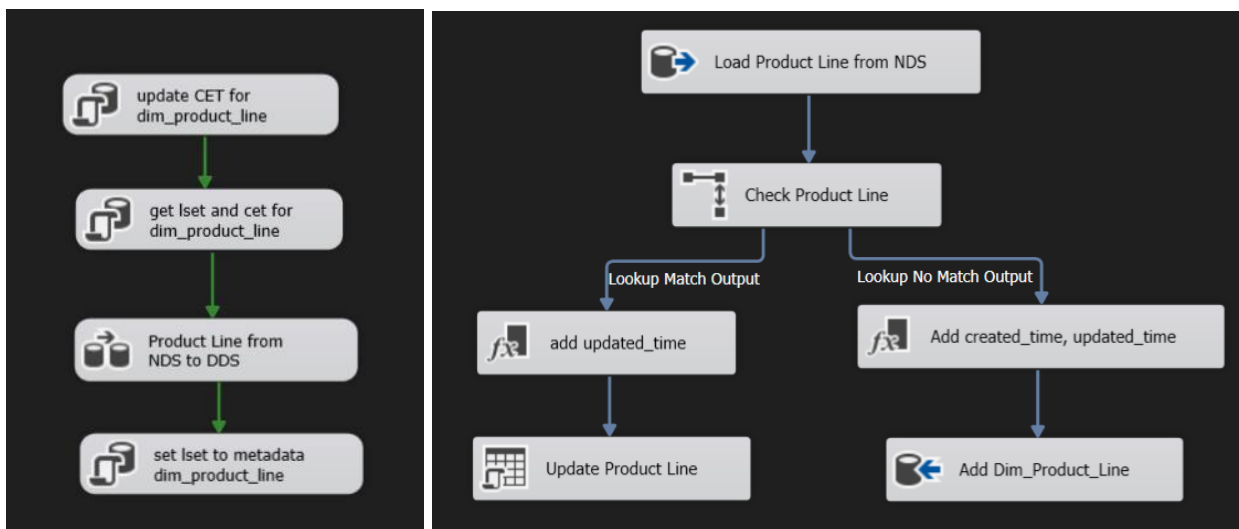
- Nếu dòng đã tồn tại thì sửa lại thời gian cập nhật cho dòng.

- Nếu dòng chưa tồn tại thì tiến hành thêm dòng dữ liệu vào dim_gender kèm thông tin thời gian tạo và thời gian cập nhật.

Bước 4: Đặt lại LSET = CET trong metadata.

2.4.2. Bảng Product Line Dim

Quy trình đồ dữ liệu từ NDS sang DDS của bảng dim_product_line:



Bước 1: Cập nhật thông tin CET cho bảng thời gian khởi động Package.

Bước 2: Lấy thông tin LSET, CET được lưu trong Metadata.

Bước 3: Tiến hành đổ dữ liệu từ NDS vào DDS cho bảng dim_product_line:

- Đầu tiên lấy dữ liệu mới cập nhật hoặc mới được thêm vào từ bảng product_line_NDS

select *

from product_line_NDS

where (ngaytao >= lset and ngaytao <= cet) or (ngaycapnhat >= lset and ngaycapnhat <= cet)

- Sau đó tiến hành kiểm tra xem từng dòng Product Line đã từng tồn tại trong bảng dim_product-line hay không.

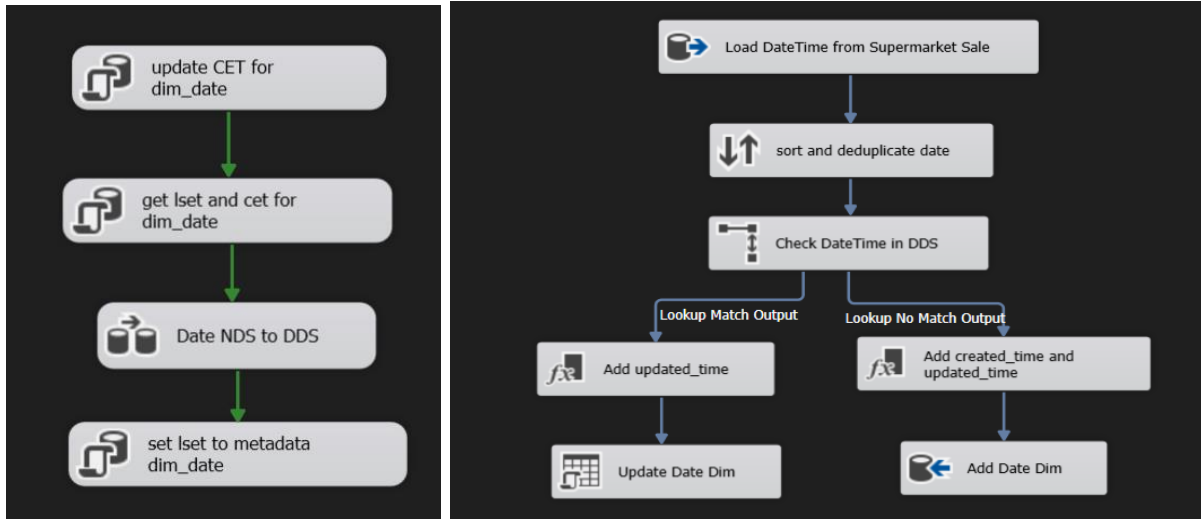
- Nếu dòng đã tồn tại thì sửa lại thời gian cập nhật cho dòng.

- Nếu dòng chưa tồn tại thì tiến hành thêm dòng dữ liệu vào dim_product_line kèm thông tin thời gian tạo và thời gian cập nhật.

Bước 4: Đặt lại LSET = CET trong metadata.

2.4.3. Bảng Date Dim

Quy trình đổ dữ liệu từ NDS sang DDS của bảng dim_date:



Bước 1: Cập nhật thông tin CET cho bảng thời gian khởi động Package.

Bước 2: Lấy thông tin LSET, CET được lưu trong Metadata.

Bước 3: Tiến hành đổ dữ liệu từ NDS vào DDS cho bảng dim_date:

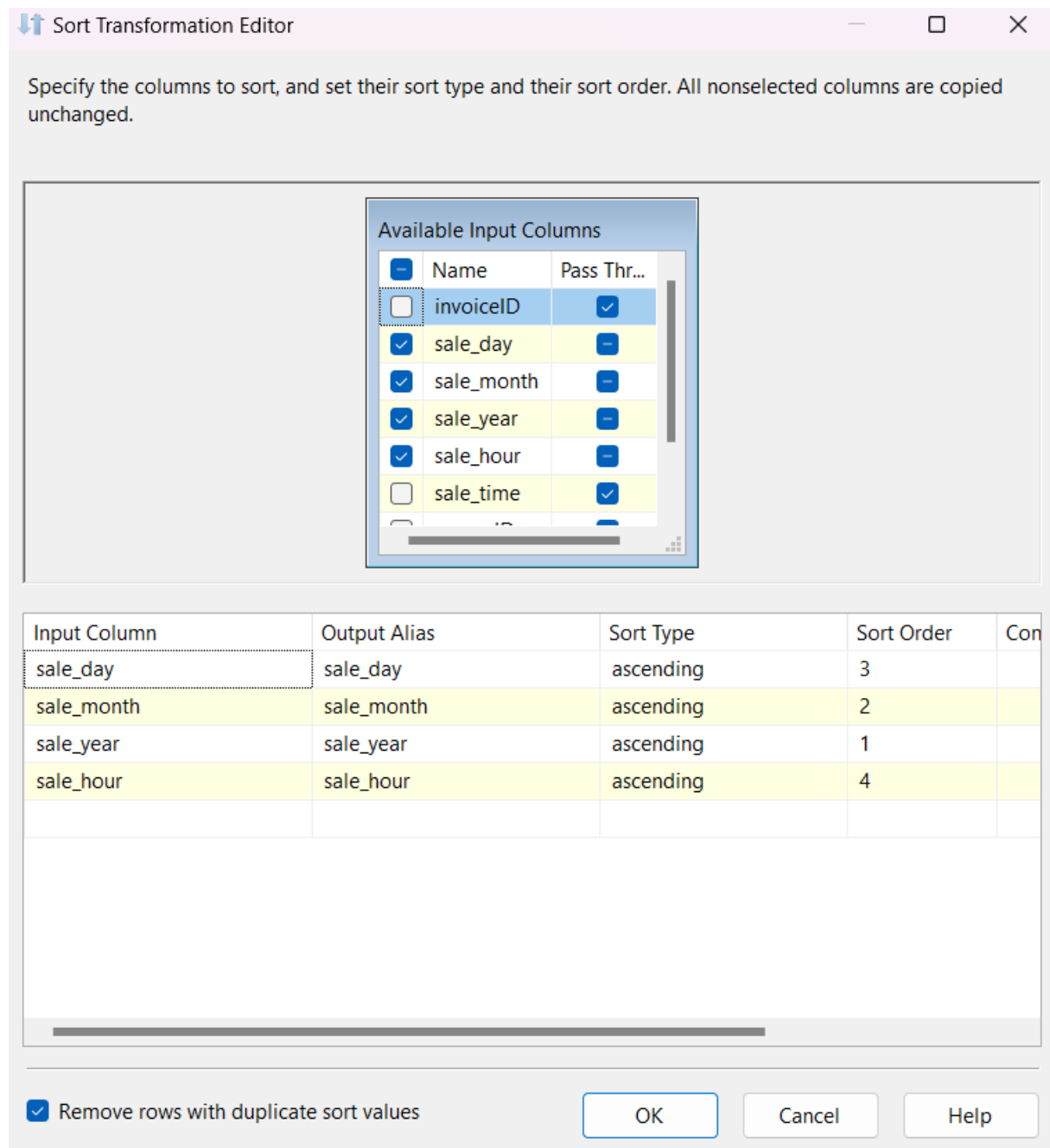
- Đầu tiên lấy dữ liệu mới cập nhật hoặc mới được thêm vào từ bảng supermarket_sale_NDS

```
select *
```

```
from supermarket_sale_NDS
```

```
where (ngaytao >= lset and ngaytao <= cet) or (ngaycapnhat >= lset and ngaycapnhat <= cet)
```

- Sau khi lấy được dữ liệu từ NDS, ta sẽ tiến hành sắp xếp lại các dòng dữ liệu theo thứ tự ưu tiên là theo năm, tháng, ngày, giờ và xóa các dòng dữ liệu bị trùng lặp.

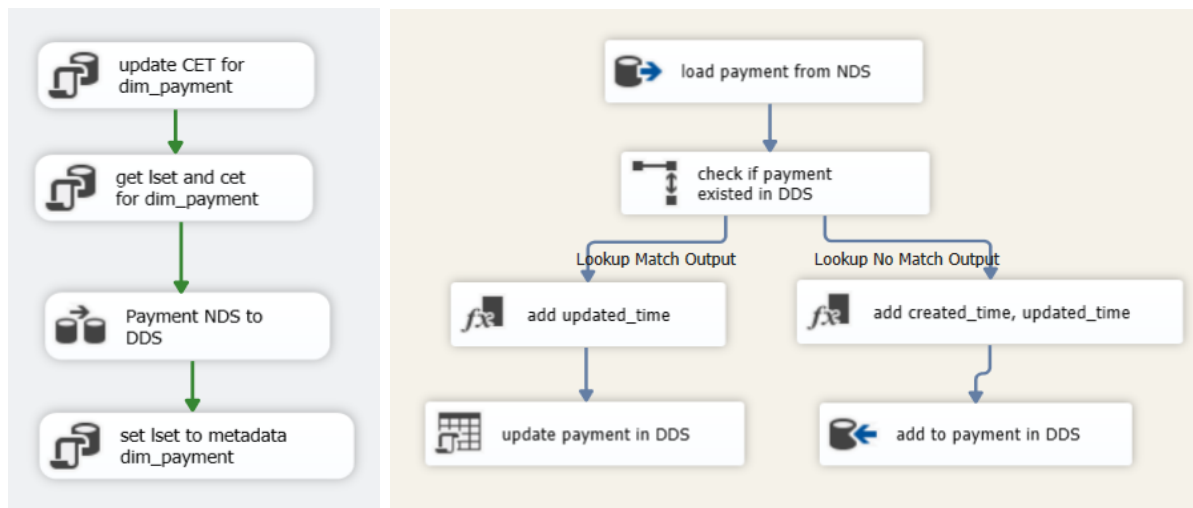


- Tiếp theo ta sẽ tiến hành kiểm tra xem từng dòng dữ liệu về ngày tháng năm giờ đã từng tồn tại trong bảng dim_date hay không.
- Nếu dòng đã tồn tại thì sửa lại thời gian cập nhật cho dòng.
- Nếu dòng chưa tồn tại thì tiến hành thêm dòng dữ liệu vào dim_date kèm thông tin thời gian tạo và thời gian cập nhật.

Bước 4: Đặt lại LSET = CET trong metadata.

2.4.4. Bảng Payment Dim

Quy trình đồ dữ liệu từ NDS sang DDS của bảng dim_payment:



Bước 1: Cập nhật thông tin CET cho bảng thời gian khởi động Package.

Bước 2: Lấy thông tin LSET, CET được lưu trong Metadata.

Bước 3: Tiến hành đồ dữ liệu từ NDS vào DDS cho bảng dim_date:

- Đầu tiên lấy dữ liệu mới cập nhật hoặc mới được thêm vào từ bảng payment_NDS
select *

from payment_NDS

where (ngaytao >= lset and ngaytao <= cet) or (ngaycapnhat >= lset and ngaycapnhat <= cet)

- Tiến hành kiểm tra xem từng dòng dữ liệu về payment đã từng tồn tại trong bảng dim_payment hay không.

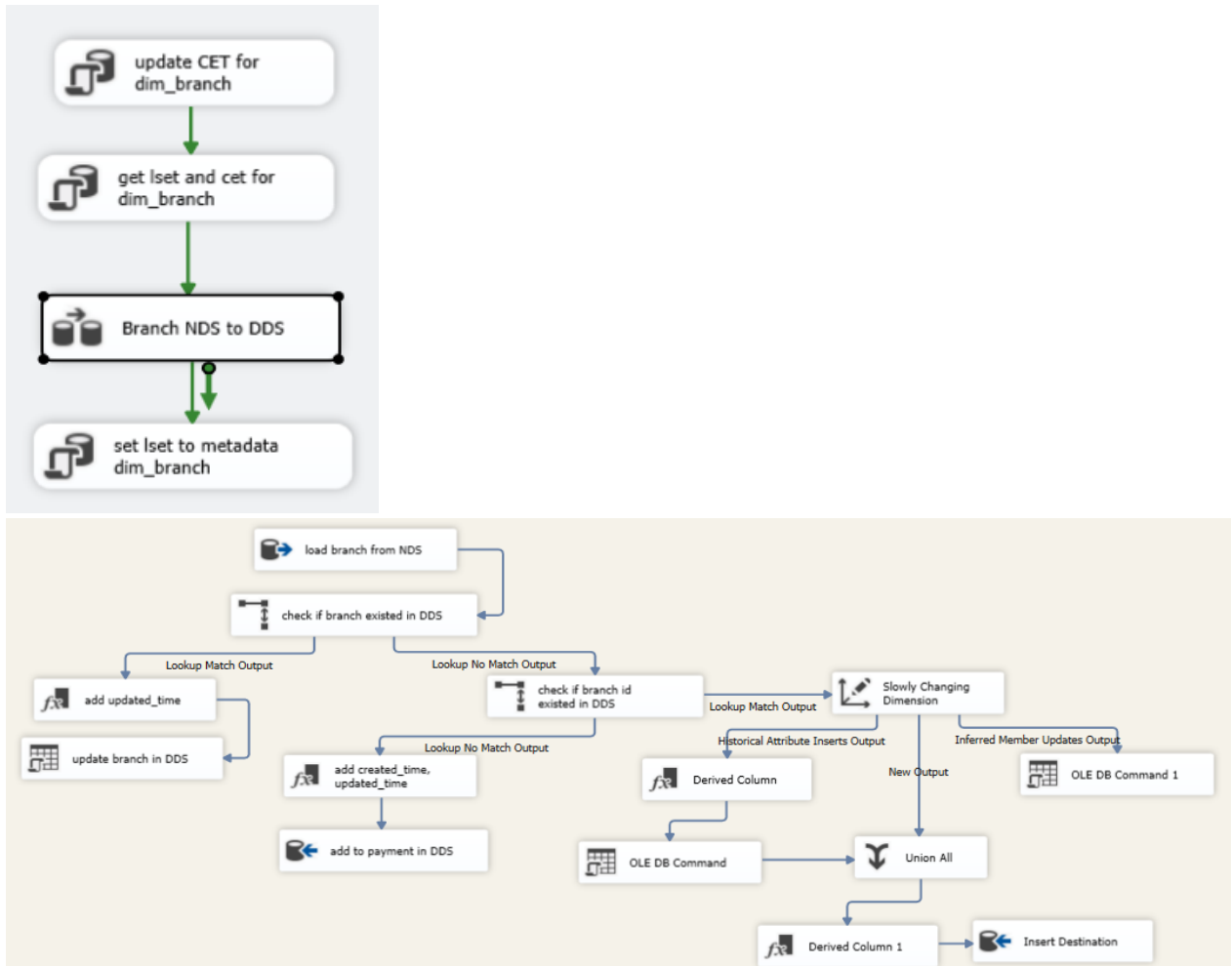
- Nếu dòng đã tồn tại thì sửa lại thời gian cập nhật cho dòng.

- Nếu dòng chưa tồn tại thì tiến hành thêm dòng dữ liệu vào dim_payment kèm thông tin thời gian tạo và thời gian cập nhật.

Bước 4: Đặt lại LSET = CET trong metadata.

2.4.5. Bảng Branch Dim

Quy trình đổ dữ liệu từ NDS sang DDS của bảng dim_branch:



Bước 1: Cập nhật thông tin CET cho bảng thời gian khởi động Package.

Bước 2: Lấy thông tin LSET, CET được lưu trong Metadata.

Bước 3: Tiến hành đổ dữ liệu từ NDS vào DDS cho bảng dim_branch:

- Đầu tiên lấy dữ liệu mới cập nhật hoặc mới được thêm vào từ bảng city_NDS

```
select *  
from city_NDS
```

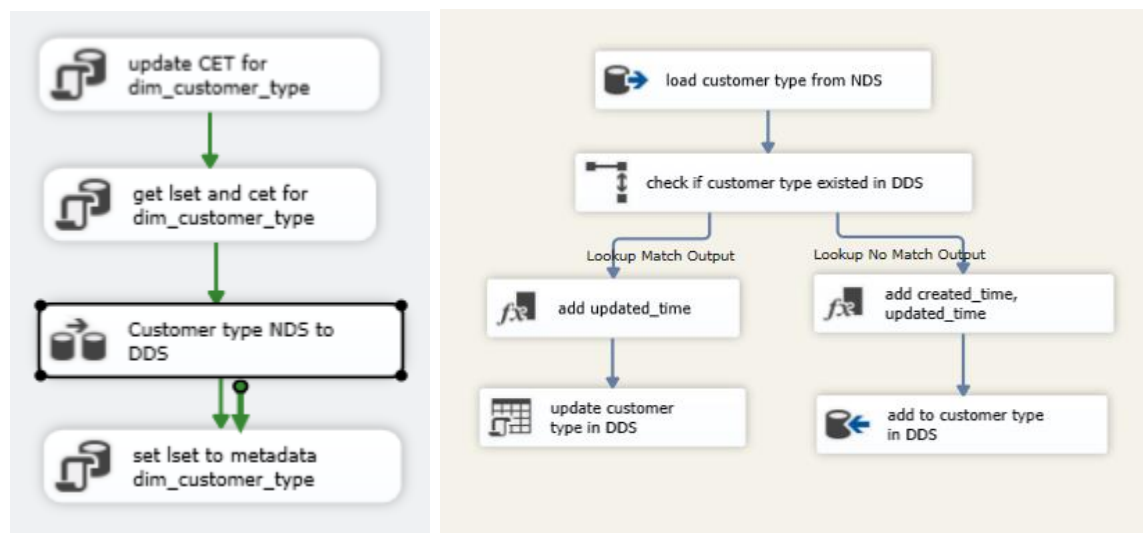
where (ngaytao >= lset and ngaytao <= cet) or (ngaycapnhat >= lset and ngaycapnhat <= cet)

- Tiến hành kiểm tra xem từng dòng dữ liệu về branch đã từng tồn tại trong bảng dim_branch hay không.
- Nếu dòng đã tồn tại thì sửa lại thời gian cập nhật cho dòng.
- Nếu dòng chưa tồn tại thì kiểm tra xem branch id đã tồn tại trong dim_branch chưa.
- Nếu chưa thì tiến hành thêm dòng dữ liệu vào dim_payment kèm thông tin thời gian tạo và thời gian cập nhật.
- Nếu đã tồn tại branch id thì cập nhật thêm dòng mới với branch id và city mới, branch id ở dòng cũ trạng thái sẽ trở thành 2.

Bước 4: Đặt lại LSET = CET trong metadata.

2.4.6. Bảng Customer Type Dim

Quy trình đồ dữ liệu từ NDS sang DDS của bảng dim_customer_type:



Bước 1: Cập nhật thông tin CET cho bảng thời gian khởi động Package.

Bước 2: Lấy thông tin LSET, CET được lưu trong Metadata.

Bước 3: Tiến hành đồ dữ liệu từ NDS vào DDS cho bảng dim_customer_type:

- Đầu tiên lấy dữ liệu mới cập nhật hoặc mới được thêm vào từ bảng customer_type_NDS

select *

from customer_type_NDS

where (ngaytao >= lset and ngaytao <= cet) or (ngaycapnhat >= lset and ngaycapnhat <= cet)

- Tiến hành kiểm tra xem từng dòng dữ liệu về customer type đã từng tồn tại trong bảng dim_payment hay không.

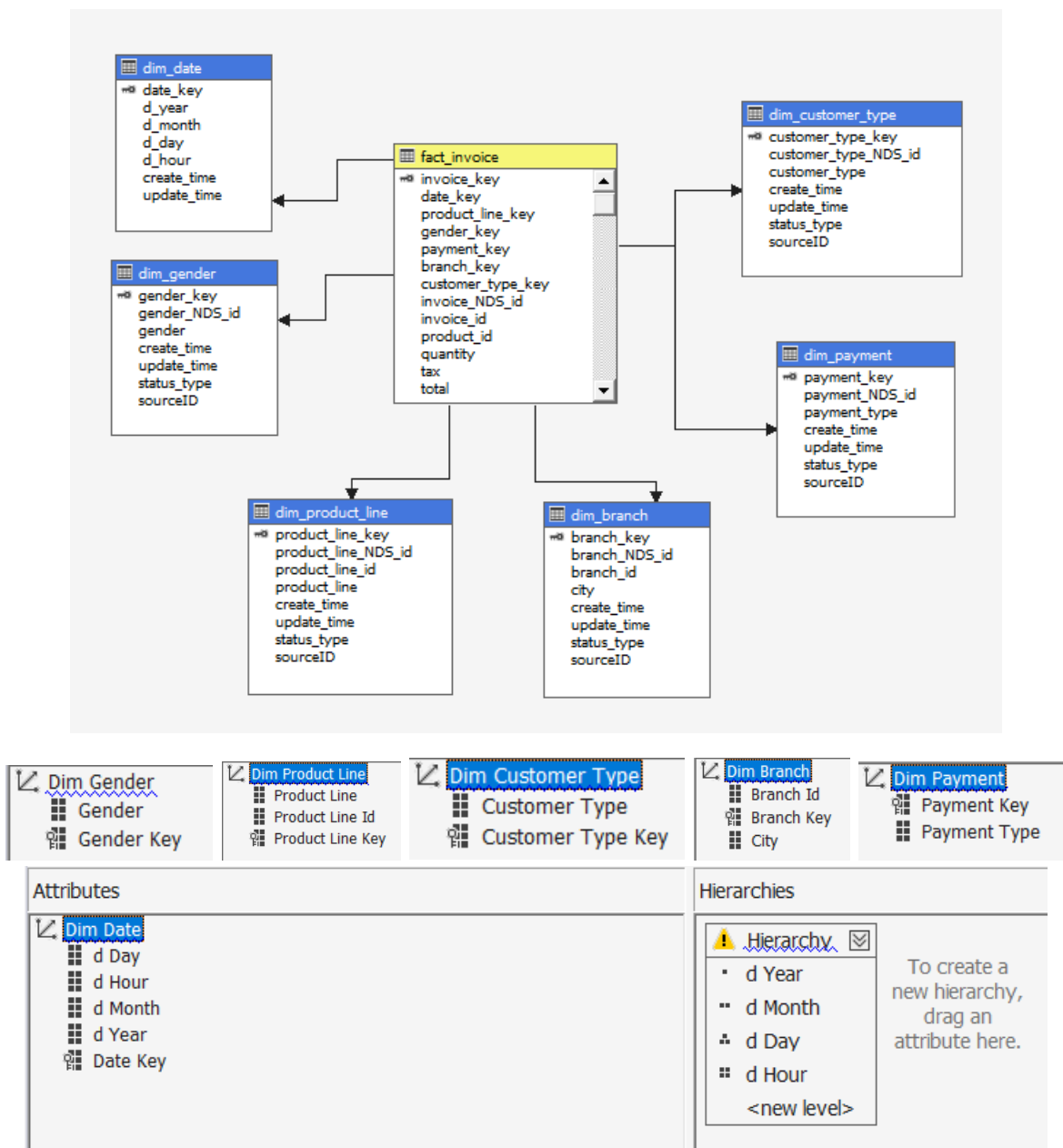
- Nếu dòng đã tồn tại thì bỏ qua, tiến hành sửa lại thời gian cập nhật cho dòng.

- Nếu dòng chưa tồn tại thì tiến hành thêm dòng dữ liệu vào dim_customer_type kèm thông tin thời gian tạo và thời gian cập nhật.

Bước 4: Đặt lại LSET = CET trong metadata.

3. OLAP

3.1. Thiết kế OLAP Cube



3.2. Tình hình mua hàng của khách hàng theo từng chi nhánh, từng loại sản phẩm, theo thời gian, hình thức thanh toán

Theo từng chi nhánh

```
--1. Tình hình mua hàng của khách hàng theo từng chi nhánh
select [Dim Branch].[City].[City] on rows,
{[Measures].[Fact Invoice Count], [Measures].[Total], [Measures].[Gross Income]} on columns
from Supermarket_sales
```

	Fact Invoice Count	Total	Gross Income
Mandalay	338	107360.39	5110.93
Naypyitaw	329	111019.09	5285.36
Yangon	349	109687.05	5221.8

- Tại chi nhánh Yangon có nhiều hoá đơn mua hàng nhất nhưng thu về doanh thu và lợi nhuận thấp nhất, ngược lại với chi nhánh Naypyitaw.

Theo từng loại sản phẩm

```
--2. Tình hình mua hàng của khách hàng theo từng loại sản phẩm
select [Dim Product Line].[Product Line].[Product Line] on rows,
{[Measures].[Fact Invoice Count], [Measures].[Quantity], [Measures].[Total], [Measures].[Gross Income]} on columns
from Supermarket_sales
```

	Fact Invoice Count	Quantity	Total	Gross Income
Electronic accessories	168	918	53711.72	2556.99
Fashion accessories	179	962	57506.11	2737.7
Food and beverages	172	959	61721.9	2938.48
Health and beauty	153	826	45850.42	2182.71
Home and lifestyle	170	922	51743.91	2463.35
Sports and travel	174	1005	57532.47	2738.86

- 3 mặt hàng thu về nhiều lượt mua hàng nhất: Fashion accessories, Sports and travel và Food and beverages.
- 3 mặt hàng doanh thu và lợi nhuận thu về cao nhất: Food and beverages, Fashion accessories, Sports and travel.
- Mặt hàng thu về doanh thu và lợi nhuận thấp nhất là Health and beauty.

Theo thời gian

```
--3. Tình hình mua hàng của khách hàng theo thời gian
select [Dim Date].[d Year].[d Year] on rows,
{[Measures].[Fact Invoice Count], [Measures].[Quantity], [Measures].[Total], [Measures].[Gross Income]} on columns
from Supermarket_sales
```

	Fact Invoice Count	Quantity	Total	Gross Income
2019	1000	5510	322962.47	15375.09
2020	16	82	5104.06	243

```
select [Dim Date].[d Month].[d Month] on rows,
{[Measures].[Fact Invoice Count], [Measures].[Quantity], [Measures].[Total], [Measures].[Gross Income]} on columns
from Supermarket_sales
```

	Fact Invoice Count	Quantity	Total	Gross Income
1	355	1980	117510.42	5594.27
2	309	1690	99335.9	4729.05
3	352	1922	111220.21	5294.77

- Tháng 1 ở các chi nhánh bán được nhiều nhất, mang lại doanh thu và lợi nhuận cao nhất trong 3 tháng.
- Tháng 2 ở các chi nhánh bán được nhiều nhất, mang lại doanh thu và lợi nhuận cao nhất trong 3 tháng.

```
select [Dim Date].[d Day].[d Day] on rows,
{[Measures].[Fact Invoice Count], [Measures].[Quantity], [Measures].[Total], [Measures].[Gross Income]} on columns
from Supermarket_sales
```

	Fact Invoice Count	Quantity	Total	Gross Income
1	29	172	10281.4	489.47
10	32	176	9865.05	469.62
11	29	165	10209.53	486.07
12	32	188	12107.67	576.44
13	28	138	5448.91	259.34
14	39	218	13635.18	649.14
15	46	268	16377.45	779.72
16	27	157	9947.19	473.55
17	31	171	10599.92	504.63
18	23	105	5569.21	265.08
19	41	239	14883.06	708.55
2	41	229	12925.75	615.33
20	35	199	11819.96	562.75
21	21	88	5835.45	277.77
22	29	153	7767.86	369.78
23	36	202	12428.66	591.69
24	33	200	11601.85	552.35
25	43	193	12232.13	582.34
26	39	166	8828.03	420.23
27	38	229	13398.01	637.85
28	30	169	9326	443.97
29	20	121	7539.73	358.96
3	36	214	12399.1	590.31
30	20	106	7045.24	335.41
31	14	82	5232.44	249.11
4	29	157	7957.48	378.79
5	42	215	13075.64	622.47
6	33	188	9612.08	457.58
7	40	238	12703.5	604.74
8	42	226	13538.77	644.51
9	38	230	13874.28	660.54

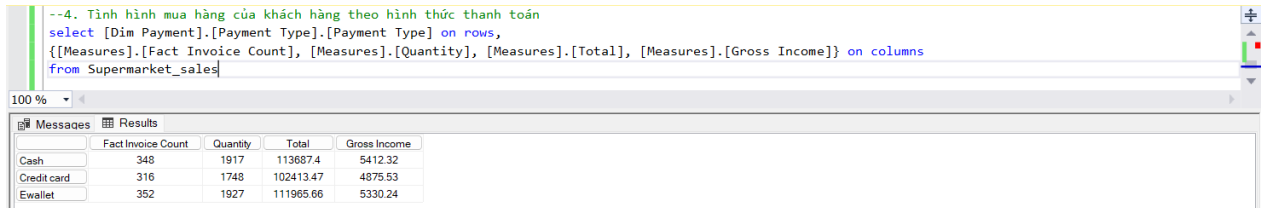
- Doanh thu và lợi nhuận cao nhất rơi vào các ngày đầu tháng và thấp nhất vào các ngày cuối tháng

```
select [Dim Date].[d Hour].[d Hour] on rows,
{[Measures].[Fact Invoice Count], [Measures].[Quantity], [Measures].[Total], [Measures].[Gross Income]} on columns
from Supermarket_sales
```

	Fact Invoice Count	Quantity	Total	Gross Income
10	104	539	32402.07	1542.5
11	93	530	31516.71	1500.41
12	90	503	26135.23	1244.18
13	103	585	34722.8	1653.06
14	83	495	30828.04	1467.66
15	104	535	31298.88	1489.99
16	79	436	26269.92	1250.63
17	76	422	24831.21	1182.12
18	94	478	26202.19	1247.38
19	115	667	40890.31	1946.73
20	75	402	22969.17	1093.43

-
- Số lượng hoá đơn, doanh thu và lợi nhuận cao nhất vào thời điểm 19 giờ và thấp nhất vào thời điểm 12 giờ.

Theo hình thức thanh toán



```
--4. Tình hình mua hàng của khách hàng theo hình thức thanh toán
select [Dim Payment].[Payment Type].[Payment Type] on rows,
{[Measures].[Fact Invoice Count], [Measures].[Quantity], [Measures].[Total], [Measures].[Gross Income]} on columns
from Supermarket_sales
```

	Fact Invoice Count	Quantity	Total	Gross Income
Cash	348	1917	113687.4	5412.32
Credit card	316	1748	102413.47	4875.53
Ewallet	352	1927	111965.66	5330.24

- Các hoá đơn được thanh toán nhiều nhất bằng Ewallet, ít nhất bằng Credit card.
- Doanh thu và lợi nhuận thu về nhiều nhất từ các khách hàng thanh toán bằng Cash.

4. Report

4.1. Thống kê số thành viên mua hàng theo ngày, tháng, năm

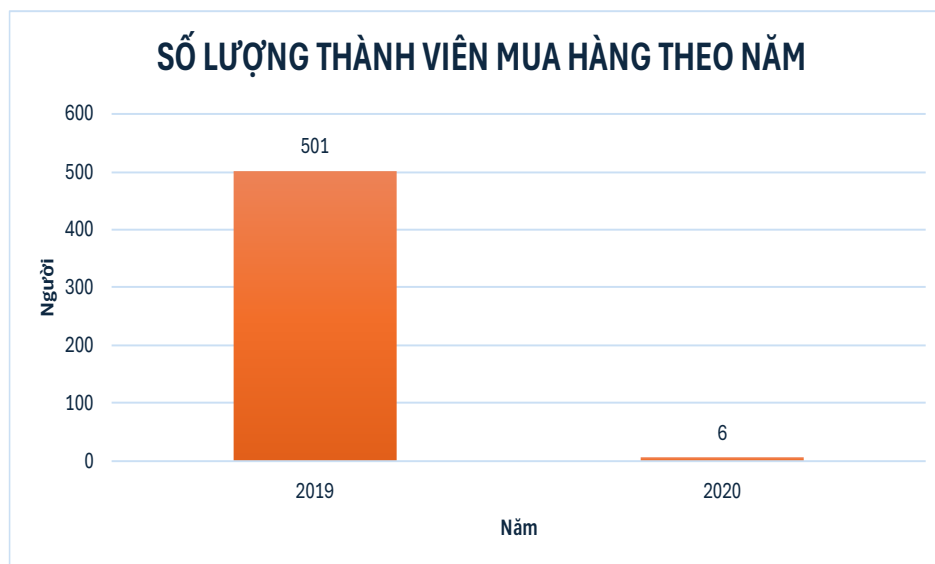
Thống kê số lượng thành viên mua hàng theo năm.

MDX Query và kết quả:

```
-- 1.1 Thống kê số lượng thành viên mua hàng theo năm
select [Measures].[Fact Invoice Count] on columns,
[Dim Date].[d Year].[d Year] on rows
from [Supermarket_sales]
where [Dim Customer Type].[Customer Type].&[Member];
```

Fact Invoice Count	
2019	501
2020	6

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Ở bộ dữ liệu này, số lượng thành viên mua hàng của năm 2019 lớn hơn rất nhiều so với năm 2020.
- Cụ thể số lượng thành viên mua hàng ở năm 2019 là 501 thành viên, còn trong năm 2020 số lượng thành viên mua hàng chỉ có 6 thành viên.

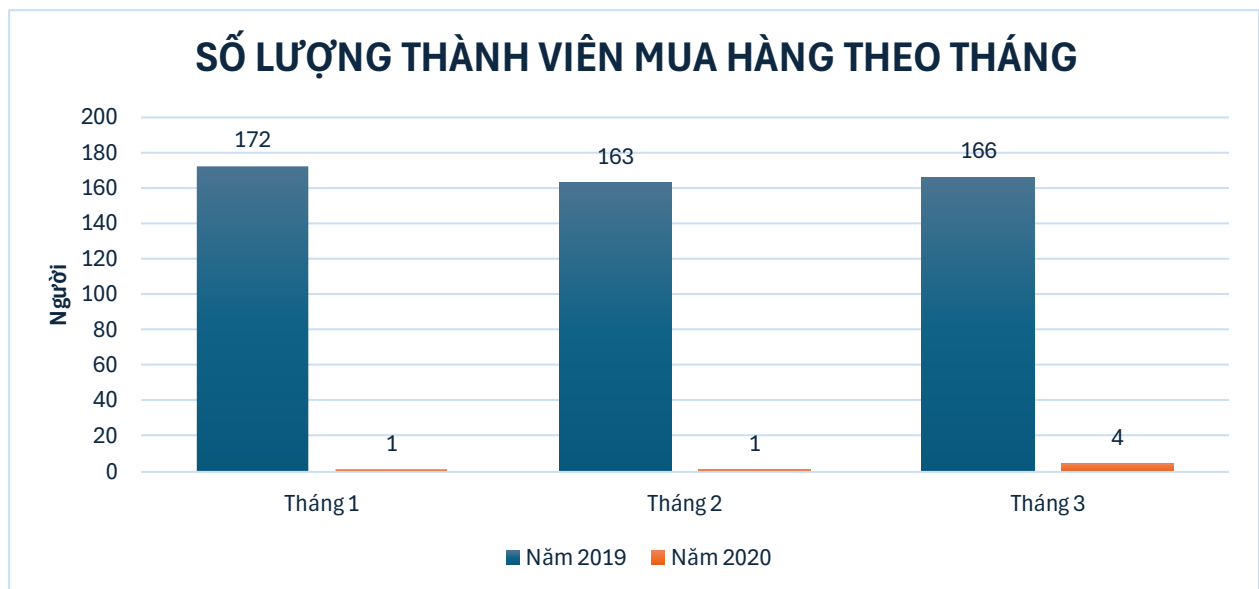
Thống kê số lượng thành viên mua hàng theo tháng

MDX Query và kết quả:

```
-- 1.2 Thống kê số lượng thành viên mua hàng theo tháng
WITH
MEMBER [Measures].[2019] AS SUM ([Dim Date].[d Year].[2019] , [Measures].[Fact Invoice Count])
MEMBER [Measures].[2020] AS SUM ([Dim Date].[d Year].[2020] , [Measures].[Fact Invoice Count])
select
{ [Measures].[2019], [Measures].[2020] } on columns,
[Dim Date].[d Month].[d Month] on rows
from [Supermarket_sales]
where [Dim Customer Type].[Customer Type].[Member];
```

	2019	2020
1	172	1
2	163	1
3	166	4

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Số lượng thành viên mua hàng dường như không thay đổi đáng kể qua các tháng trong năm.

Thống kê số lượng thành viên mua hàng theo ngày của năm 2019

MDX Query và kết quả:

-- 1.3 Thống kê số lượng thành viên mua hàng theo ngày của năm 2019

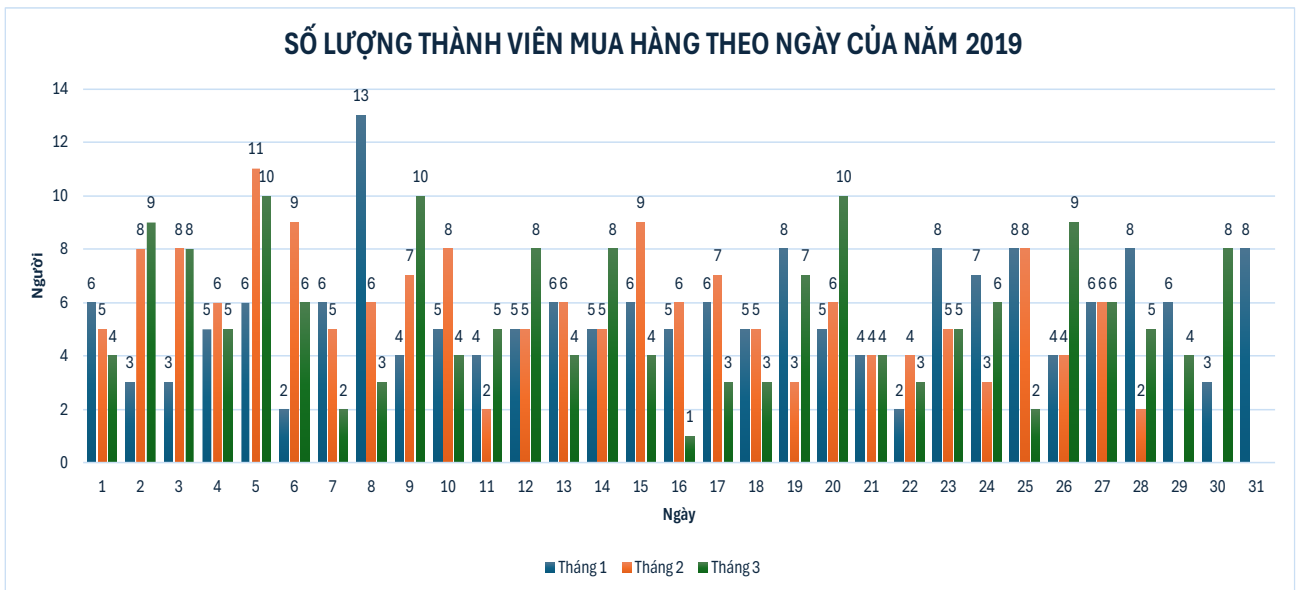
```

WITH
MEMBER [Measures].[1] AS SUM ([Dim Date].[d Month].&[1] , [Measures].[Fact Invoice Count])
MEMBER [Measures].[2] AS SUM ([Dim Date].[d Month].&[2] , [Measures].[Fact Invoice Count])
MEMBER [Measures].[3] AS SUM ([Dim Date].[d Month].&[3] , [Measures].[Fact Invoice Count])
select
{ [Measures].[1], [Measures].[2], [Measures].[3] } on columns,
NON EMPTY CROSSJOIN ([Dim Date].[d Year].&[2019], [Dim Date].[d Day].[d Day] ) on rows
from [Supermarket_sales]
where [Dim Customer Type].[Customer Type].&[Member];

```

		1	2	3
2019	1	6	5	4
2019	10	5	8	4
2019	11	4	2	5
2019	12	5	5	8
2019	13	6	6	4
2019	14	5	5	8
2019	15	6	9	4
2019	16	5	6	1
2019	17	6	7	3
2019	18	5	5	3
2019	19	8	3	7
2019	2	3	8	9
2019	20	5	6	10
2019	21	4	4	4
2019	22	2	4	3

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Số lượng thành viên mua hàng thay đổi qua từng ngày trong năm 2019.

- Trong đó, ngày 8 của tháng 1 năm 2019 là ngày có số lượng thành viên mua hàng nhiều nhất với 13 thành viên. Còn ngày 16 tháng 3 năm 2019 là ngày có số lượng thành viên mua hàng ít nhất chỉ có 1 thành viên.

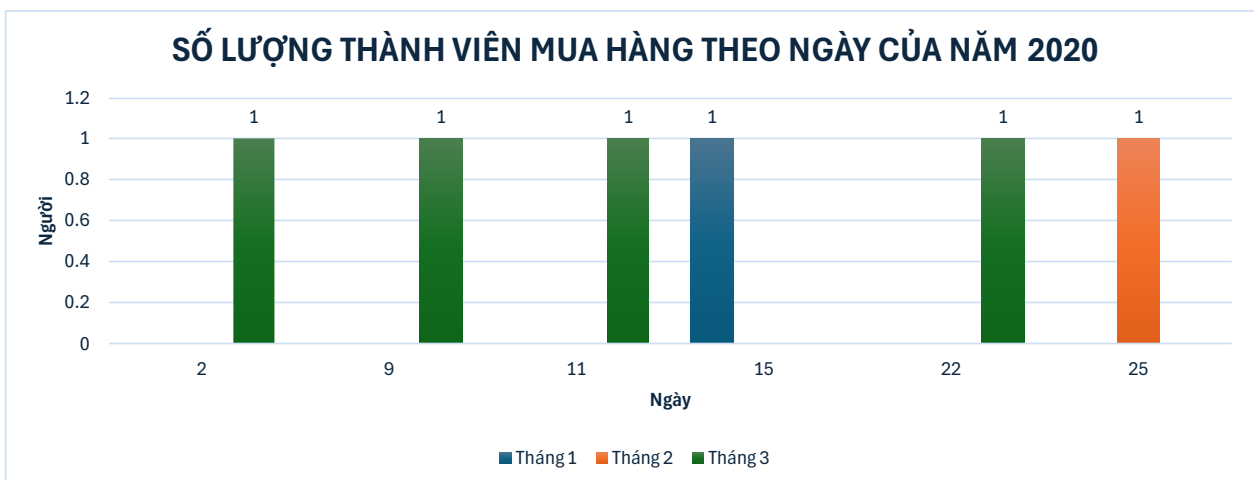
Thống kê số lượng thành viên mua hàng theo ngày của năm 2020

MDX Query và kết quả:

```
-- 1.4 Thống kê số lượng thành viên mua hàng theo ngày của năm 2020
WITH
MEMBER [Measures].[1] AS SUM ([Dim Date].[d Month].&[1] , [Measures].[Fact Invoice Count])
MEMBER [Measures].[2] AS SUM ([Dim Date].[d Month].&[2] , [Measures].[Fact Invoice Count])
MEMBER [Measures].[3] AS SUM ([Dim Date].[d Month].&[3] , [Measures].[Fact Invoice Count])
select
{ [Measures].[1], [Measures].[2], [Measures].[3] } on columns,
NON EMPTY CROSSJOIN ([Dim Date].[d Year].&[2020], [Dim Date].[d Day].[d Day] ) on rows
from [Supermarket_sales]
where [Dim Customer Type].[Customer Type].&[Member];
```

		1	2	3
2020	11	(null)	(null)	1
2020	15	1	(null)	(null)
2020	2	(null)	(null)	1
2020	22	(null)	(null)	1
2020	25	(null)	1	(null)
2020	9	(null)	(null)	1

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Số lượng thành viên mua hàng trong năm 2020 không thay đổi qua các ngày.

4.2. Thống kê doanh thu của khách hàng (member, normal) theo ngày, tháng, năm và theo chi nhánh

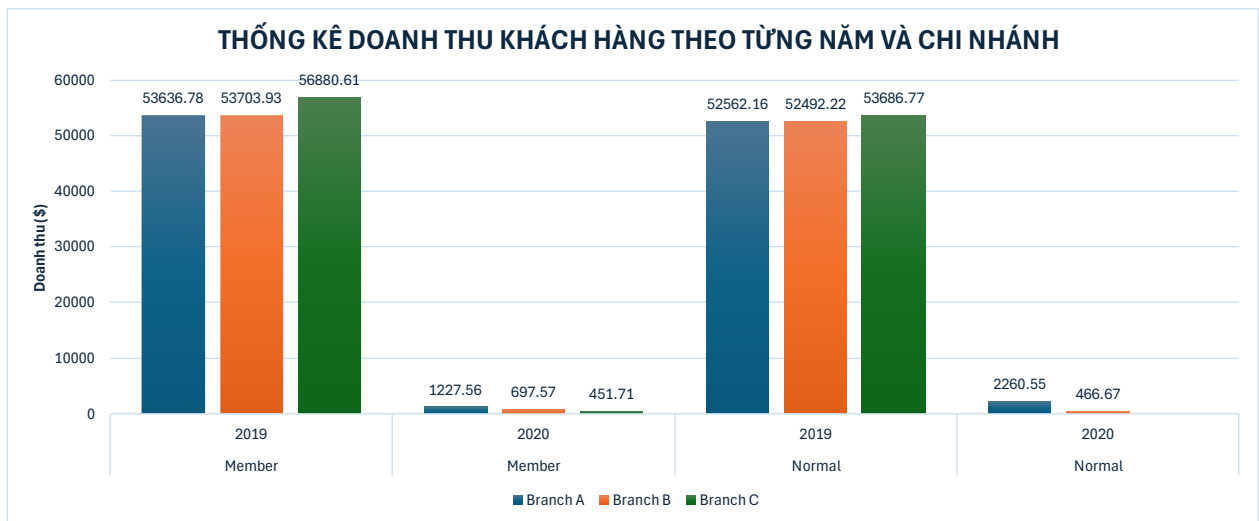
Thống kê doanh thu của khách hàng theo năm và theo chi nhánh

MDX Query và kết quả:

```
-- 2.1 Thống kê doanh thu của khách hàng theo năm và theo chi nhánh
WITH
MEMBER [Measures].A AS SUM ([Dim Branch].[Branch Id].&[A] , [Measures].[Total])
MEMBER [Measures].B AS SUM ([Dim Branch].[Branch Id].&[B] , [Measures].[Total])
MEMBER [Measures].C AS SUM ([Dim Branch].[Branch Id].&[C] , [Measures].[Total])
select
{ [Measures].A, [Measures].B, [Measures].C } on columns,
NON EMPTY CROSSJOIN ([Dim Customer Type].[Customer Type].[Customer Type] , [Dim Date].[d Year].[d Year]) on rows
from [Supermarket_sales];
```

		A	B	C
Member	2019	53636.78	53703.93	56880.61
Member	2020	1227.56	697.57	451.71
Normal	2019	52562.16	52492.22	53686.77
Normal	2020	2260.55	466.67	(null)

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Doanh thu giữa khách hàng thành viên và khách hàng thông thường không có sự chênh lệch đáng kể giữa các chi nhánh với nhau.
- Trong đó, doanh thu khách hàng thành viên của chi nhánh C trong năm 2019 là cao nhất,

Thống kê doanh thu của khách hàng theo tháng và theo chi nhánh của năm 2019

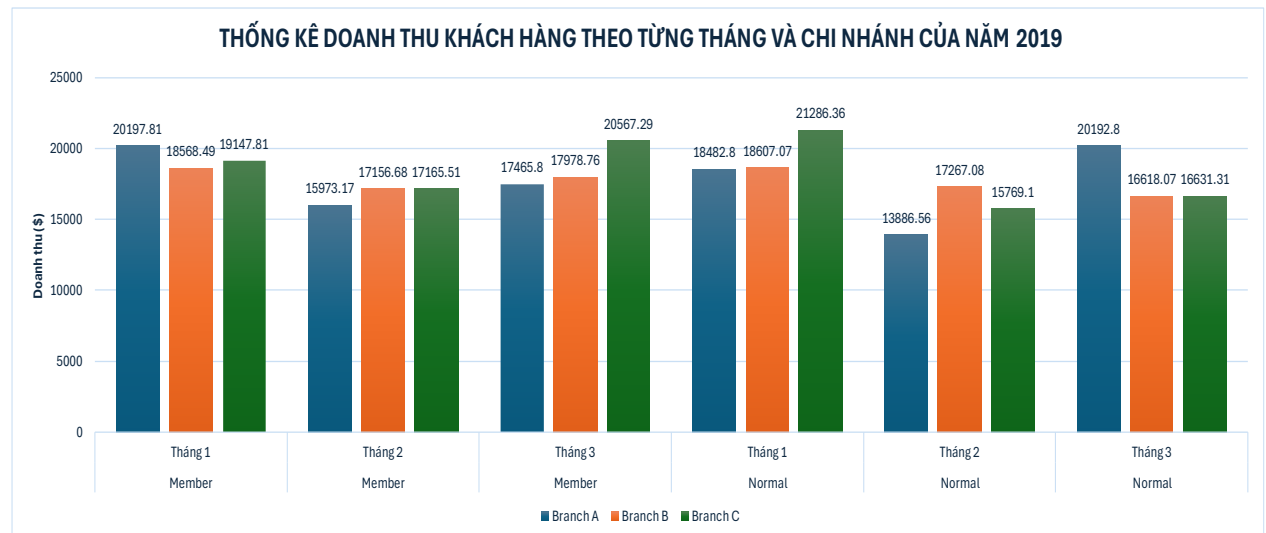
MDX Query và kết quả:

```
-- 2.2 Thống kê doanh thu của khách hàng theo tháng và theo chi nhánh của năm 2019
WITH
MEMBER [Measures].A AS SUM ([Dim Branch].[Branch Id].&[A] , [Measures].[Total])
MEMBER [Measures].B AS SUM ([Dim Branch].[Branch Id].&[B] , [Measures].[Total])
MEMBER [Measures].C AS SUM ([Dim Branch].[Branch Id].&[C] , [Measures].[Total])

select
{ [Measures].A, [Measures].B, [Measures].C } on columns,
NON EMPTY CROSSJOIN ([Dim Customer Type].[Customer Type].[Customer Type] , [Dim Date].[d Month].[d Month] ) on rows
from [Supermarket_sales]
where [Dim Date].[d Year].&[2019];
```

		A	B	C
Member	1	20197.81	18568.49	19147.81
Member	2	15973.17	17156.68	17165.51
Member	3	17465.8	17978.76	20567.29
Normal	1	18482.8	18607.07	21286.36
Normal	2	13886.56	17267.08	15769.1
Normal	3	20192.8	16618.07	16631.31

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Doanh thu đến từ khách hàng thông thường của chi nhánh C trong tháng 1 năm 2019 là cao nhất.
- Doanh thu thấp nhất cũng khách hàng thông thường của chi nhánh A trong tháng 2 năm.

Thống kê doanh thu của khách hàng theo tháng và theo chi nhánh của năm 2020

MDX Query và kết quả:

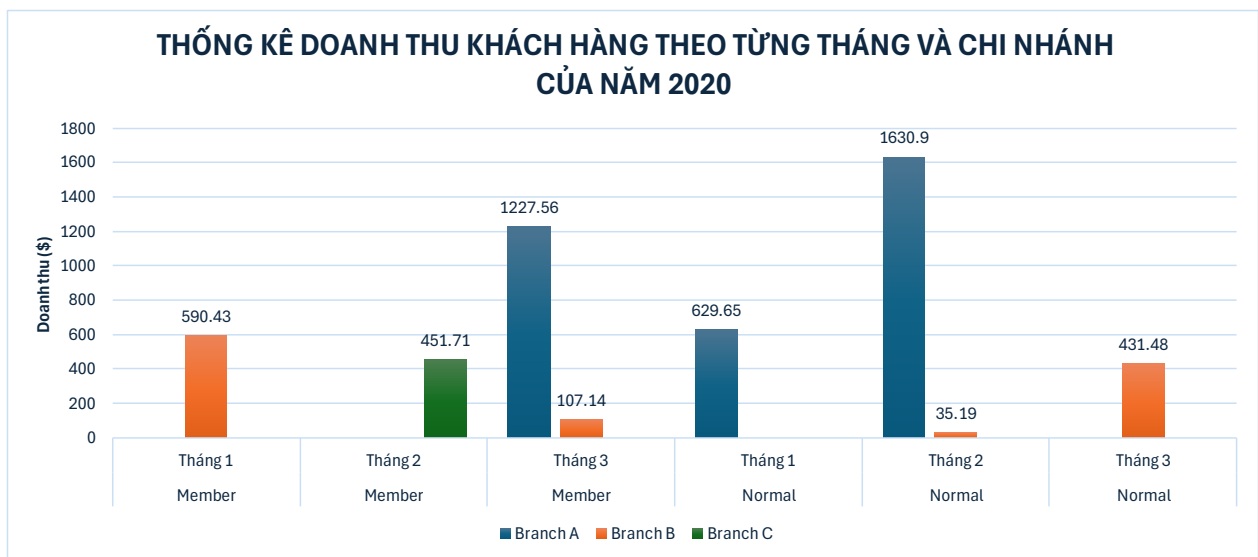
```
-- 2.3 Thống kê doanh thu của khách hàng theo tháng và theo chi nhánh của năm 2020
```

```
WITH
MEMBER [Measures].A AS SUM ([Dim Branch].[Branch Id].&[A] , [Measures].[Total])
MEMBER [Measures].B AS SUM ([Dim Branch].[Branch Id].&[B] , [Measures].[Total])
MEMBER [Measures].C AS SUM ([Dim Branch].[Branch Id].&[C] , [Measures].[Total])

select
{ [Measures].A, [Measures].B, [Measures].C } on columns,
NON EMPTY CROSSJOIN ([Dim Customer Type].[Customer Type].[Customer Type] , [Dim Date].[d Month].[d Month] ) on rows
from [Supermarket_sales]
where [Dim Date].[d Year].&[2020];
```

		A	B	C
Member	1	(null)	590.43	(null)
Member	2	(null)	(null)	451.71
Member	3	1227.56	107.14	(null)
Normal	1	629.65	(null)	(null)
Normal	2	1630.9	35.19	(null)
Normal	3	(null)	431.48	(null)

Trực quan hóa kết quả bằng biểu đồ Excel:

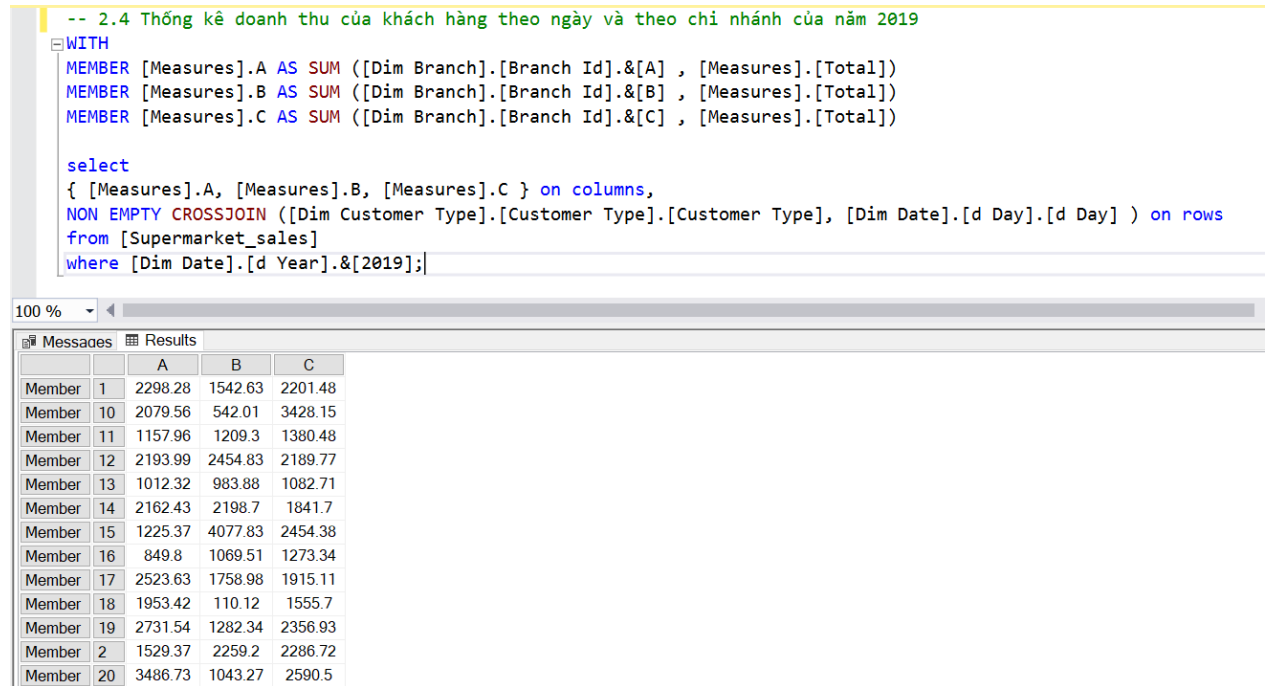


Nhận xét:

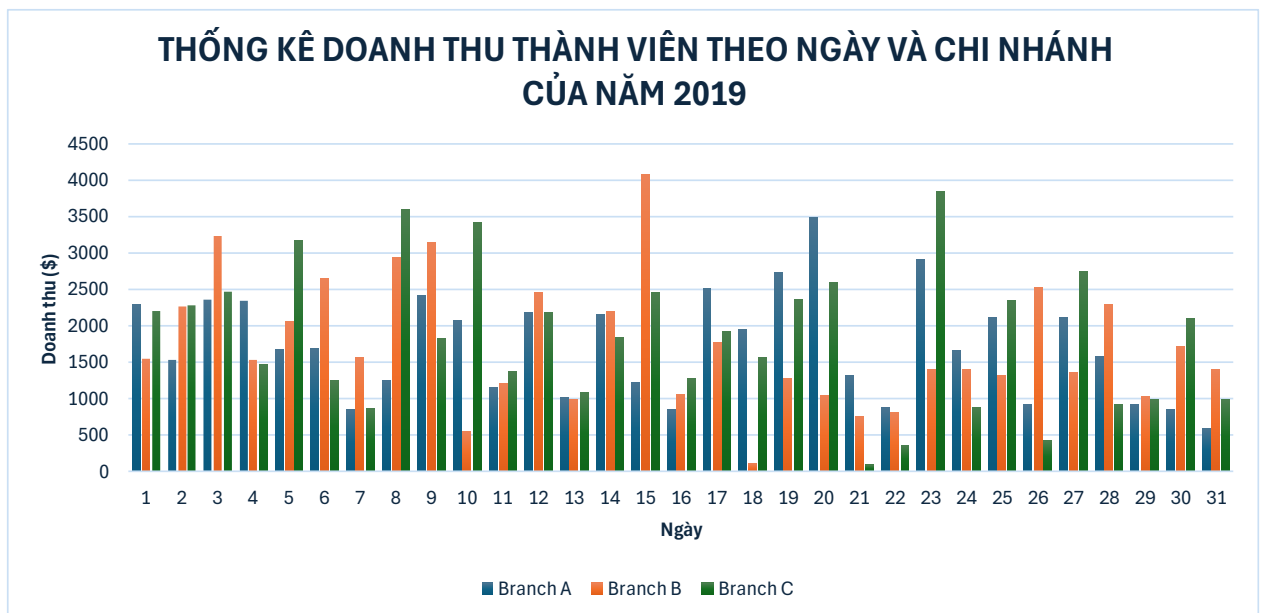
- Doanh thu đến từ khách hàng thông thường của chi nhánh A vào tháng 2 năm là cao nhất.
- Doanh thu thấp nhất cũng đến từ khách hàng thông thường tháng 2 của chi nhánh B.

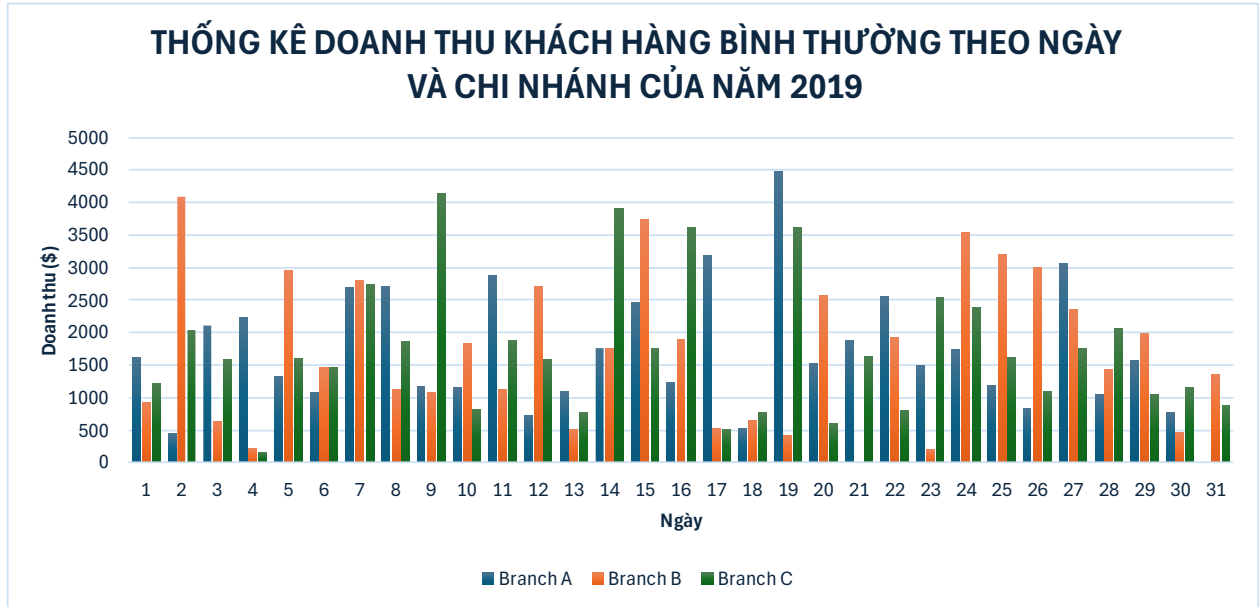
Thống kê doanh thu của khách hàng theo ngày và theo chi nhánh của năm 2019

MDX Query và kết quả:



Trực quan hóa kết quả bằng biểu đồ Excel:





Nhận xét:

- Doanh thu khách hàng của các chi nhánh thay đổi qua từng ngày.
- Trong đó, doanh thu thành viên cao nhất đến từ ngày 15 của chi nhánh B, doanh thu thành viên thấp nhất đến từ ngày 21 của chi nhánh C.
- Tương tự, doanh thu khách hàng thông thường cao nhất đến từ ngày 19 của chi nhánh A và doanh thu khách hàng thông thường thấp nhất đến từ ngày 4 của chi nhánh C.

Thống kê doanh thu khách hàng theo ngày và theo chi nhánh của năm 2020

MDX Query và kết quả:

-- 2.5 Thống kê doanh thu của khách hàng theo ngày và theo chi nhánh của năm 2020

```

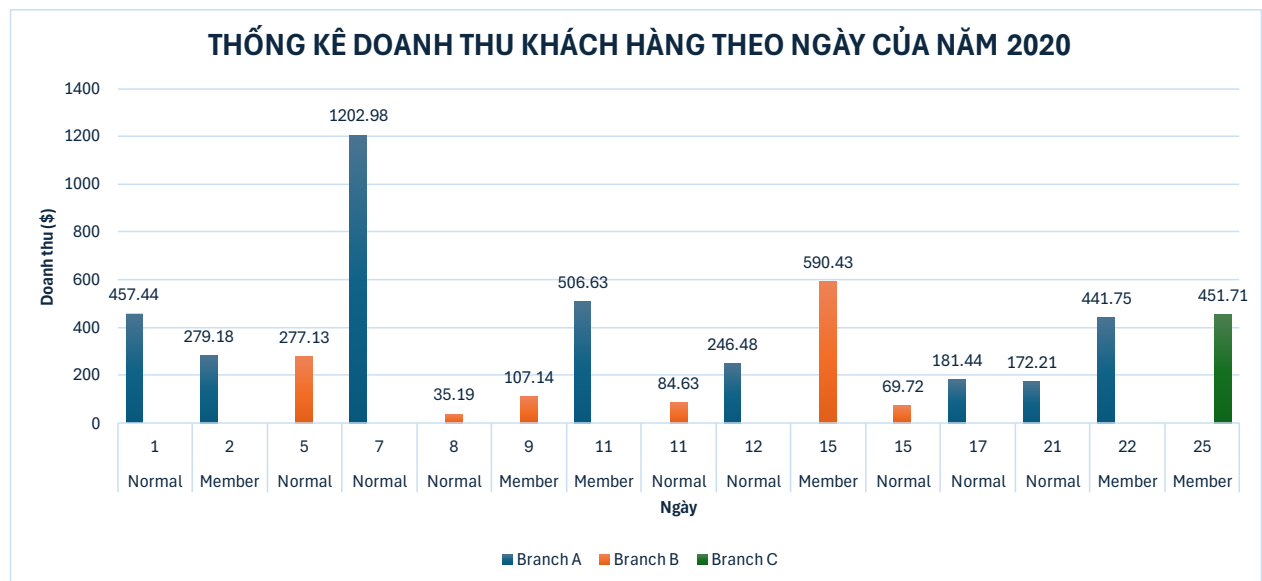
WITH
MEMBER [Measures].A AS SUM ([Dim Branch].[Branch Id].&[A] , [Measures].[Total])
MEMBER [Measures].B AS SUM ([Dim Branch].[Branch Id].&[B] , [Measures].[Total])
MEMBER [Measures].C AS SUM ([Dim Branch].[Branch Id].&[C] , [Measures].[Total])

select
{ [Measures].A, [Measures].B, [Measures].C } on columns,
NON EMPTY CROSSJOIN ([Dim Customer Type].[Customer Type].[Customer Type], [Dim Date].[d Day].[d Day] ) on rows
from [Supermarket_sales]
where [Dim Date].[d Year].&[2020];

```

		A	B	C
Member	11	506.63	(null)	(null)
Member	15	(null)	590.43	(null)
Member	2	279.18	(null)	(null)
Member	22	441.75	(null)	(null)
Member	25	(null)	(null)	451.71
Member	9	(null)	107.14	(null)
Normal	1	457.44	(null)	(null)
Normal	11	(null)	84.63	(null)
Normal	12	246.48	(null)	(null)
Normal	15	(null)	69.72	(null)
Normal	17	181.44	(null)	(null)
Normal	21	172.21	(null)	(null)
Normal	5	(null)	277.13	(null)
Normal	7	1202.98	(null)	(null)

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Doanh thu đến từ khách hàng thông thường vào ngày 7 của chi nhánh A là cao nhất. Doanh thu thấp nhất cũng đến từ khách hàng thông thường vào ngày 8 của chi nhánh B.

4.3. Thống kê số lượng khách thanh toán theo cash/debit/... ở từng chi nhánh theo từng tháng trong năm

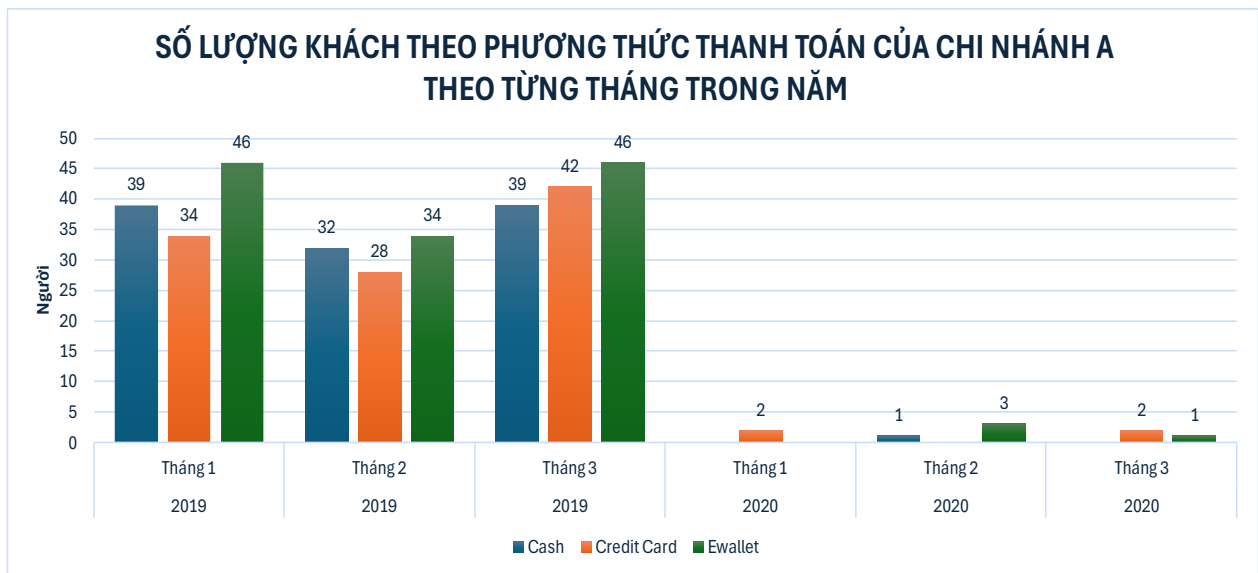
Thống kê số lượng khách theo phương thức thanh toán ở chi nhánh A theo từng tháng trong năm

MDX Query và kết quả:

```
-- 3.1 Thống kê số lượng khách theo phương thức thanh toán ở chi nhánh A theo từng tháng trong năm
WITH
MEMBER [Measures].Cash AS SUM ([Dim Payment].[Payment Type].[Cash] , [Measures].[Fact Invoice Count])
MEMBER [Measures].Credit_Card AS SUM ([Dim Payment].[Payment Type].[Credit card] , [Measures].[Fact Invoice Count])
MEMBER [Measures].Ewallet AS SUM ([Dim Payment].[Payment Type].[Ewallet] , [Measures].[Fact Invoice Count])
select
{ [Measures].Cash, [Measures].Credit_Card, [Measures].Ewallet } on columns,
NON EMPTY CROSSJOIN ([Dim Date].[d Year].[d Year],[Dim Date].[d Month].[d Month] ) on rows
from [Supermarket_sales]
where [Dim Branch].[Branch Id].[A];
```

		Cash	Credit_Card	Ewallet
2019	1	39	34	46
2019	2	32	28	34
2019	3	39	42	46
2020	1	(null)	2	(null)
2020	2	1	(null)	3
2020	3	(null)	2	1

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Trong năm 2019, số lượng khách hàng thanh toán bằng Ewallet là cao nhất qua các tháng, còn phương thức Credit Card là phương thức thanh toán ít được sử dụng nhất.
- Trong năm 2020, chỉ có phương thức Credit Card được sử dụng tại tháng 1, tại tháng 2 có 2 phương thức được sử dụng là Cash và Ewallet và cuối cùng tháng 3 chỉ có phương thức Credit Card và Ewallet được sử dụng.

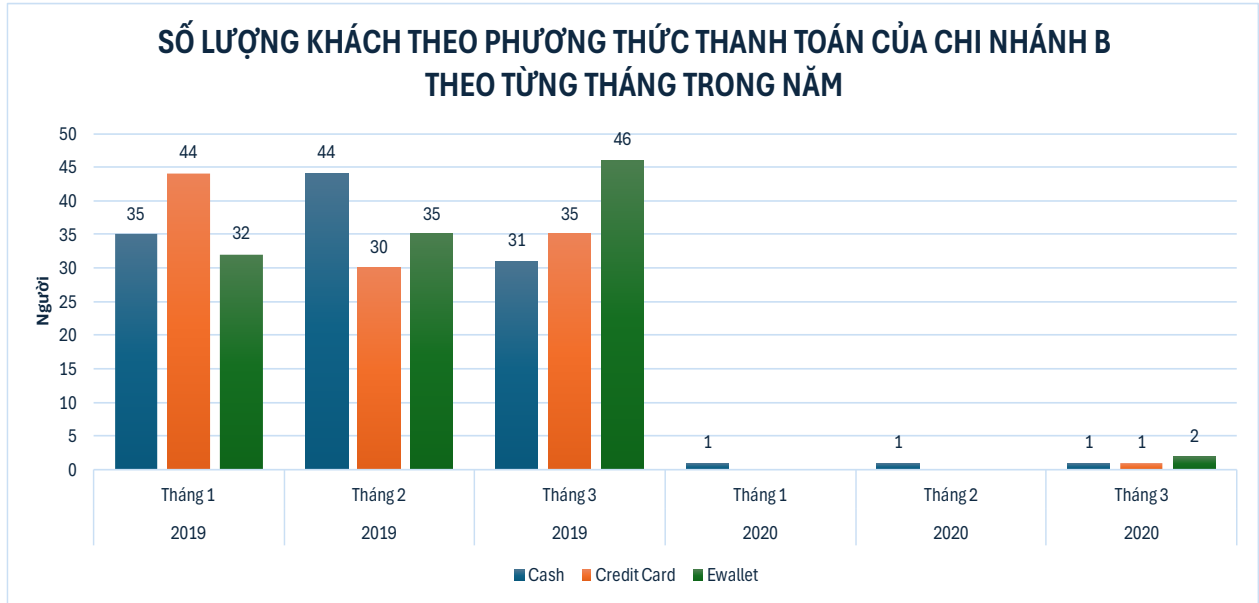
Thống kê số lượng khách theo phương thức thanh toán ở chi nhánh B theo từng tháng trong năm

MDX Query và kết quả:

```
-- 3.2 Thống kê số lượng khách theo phương thức thanh toán ở chi nhánh B theo từng tháng trong năm
WITH
MEMBER [Measures].Cash AS SUM ([Dim Payment].[Payment Type].[Cash] , [Measures].[Fact Invoice Count])
MEMBER [Measures].Credit_Card AS SUM ([Dim Payment].[Payment Type].[Credit card] , [Measures].[Fact Invoice Count])
MEMBER [Measures].Ewallet AS SUM ([Dim Payment].[Payment Type].[Ewallet] , [Measures].[Fact Invoice Count])
select
{ [Measures].Cash, [Measures].Credit_Card, [Measures].Ewallet } on columns,
NON EMPTY CROSSJOIN ([Dim Date].[d Year].[d Year],[Dim Date].[d Month].[d Month] ) on rows
from [Supermarket_sales]
where [Dim Branch].[Branch Id].&[B];
```

		Cash	Credit_Card	Ewallet
2019	1	35	44	32
2019	2	44	30	35
2019	3	31	35	46
2020	1	1	(null)	(null)
2020	2	1	(null)	(null)
2020	3	1	1	2

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Trong năm 2019, nhu cầu sử dụng các phương thức thanh toán thay đổi liên tục qua các tháng. Trong tháng 1, phương thức thanh toán được sử dụng nhiều nhất là Credit Card. Qua tháng 2, phương thức thanh toán được sử dụng nhiều nhất là Cash. Tháng thứ 3, phương thức thanh toán được sử dụng nhiều nhất lại là Ewallet.
- Trong năm 2020, ở tháng 1 và 2 thì chỉ có phương thức thanh toán Cash được sử dụng. Qua tháng 3 thì cả 3 phương thức thanh toán đều được sử dụng.

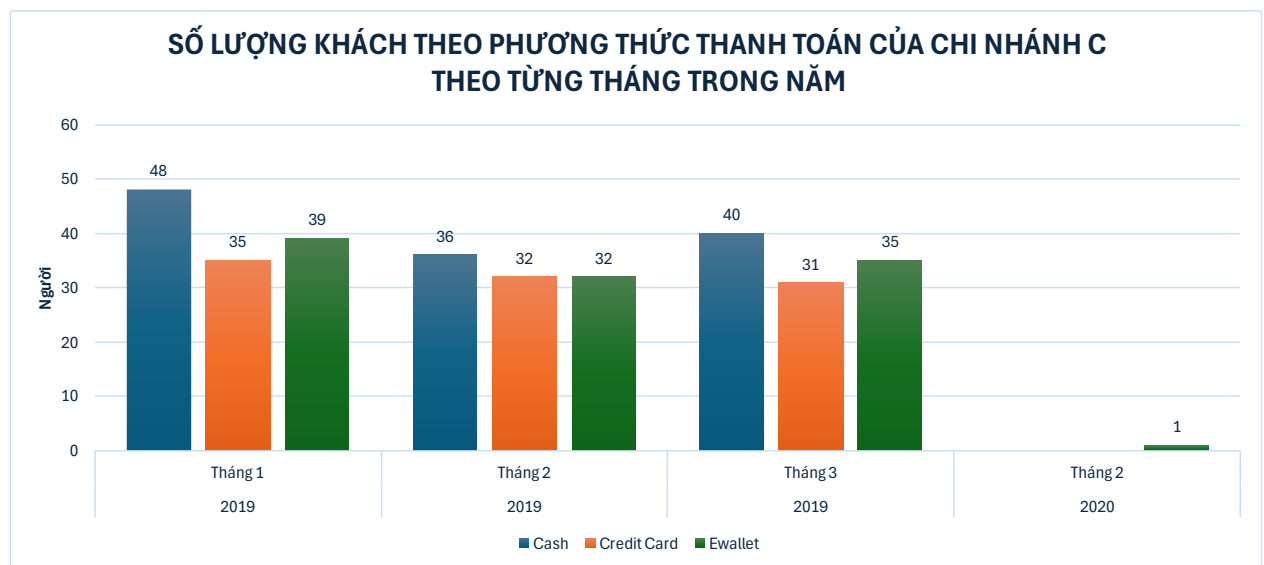
Thống kê số lượng khách theo phương thức thanh toán ở chi nhánh C theo từng tháng trong năm

MDX Query và kết quả:


```
-- 3.3 Thống kê số lượng khách theo phương thức thanh toán ở chi nhánh C theo từng tháng trong năm
WITH
MEMBER [Measures].Cash AS SUM ([Dim Payment].[Payment Type].[Cash] , [Measures].[Fact Invoice Count])
MEMBER [Measures].Credit_Card AS SUM ([Dim Payment].[Payment Type].[Credit card] , [Measures].[Fact Invoice Count])
MEMBER [Measures].Ewallet AS SUM ([Dim Payment].[Payment Type].[Ewallet] , [Measures].[Fact Invoice Count])
select
{ [Measures].Cash, [Measures].Credit_Card, [Measures].Ewallet } on columns,
NON EMPTY CROSSJOIN ([Dim Date].[d Year].[d Year],[Dim Date].[d Month].[d Month] ) on rows
from [Supermarket_sales]
where [Dim Branch].[Branch Id].&[C];
```

		Cash	Credit_Card	Ewallet
2019	1	48	35	39
2019	2	36	32	32
2019	3	40	31	35
2020	2	(null)	(null)	1

Trực quan hóa kết quả bằng biểu đồ Excel:

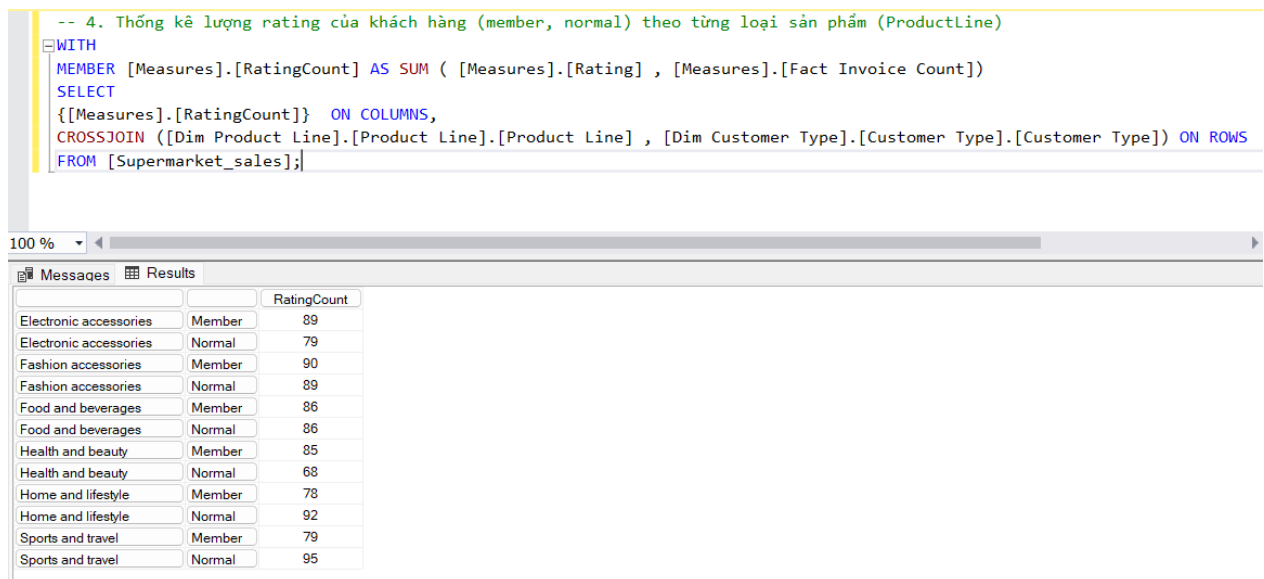


Nhận xét:

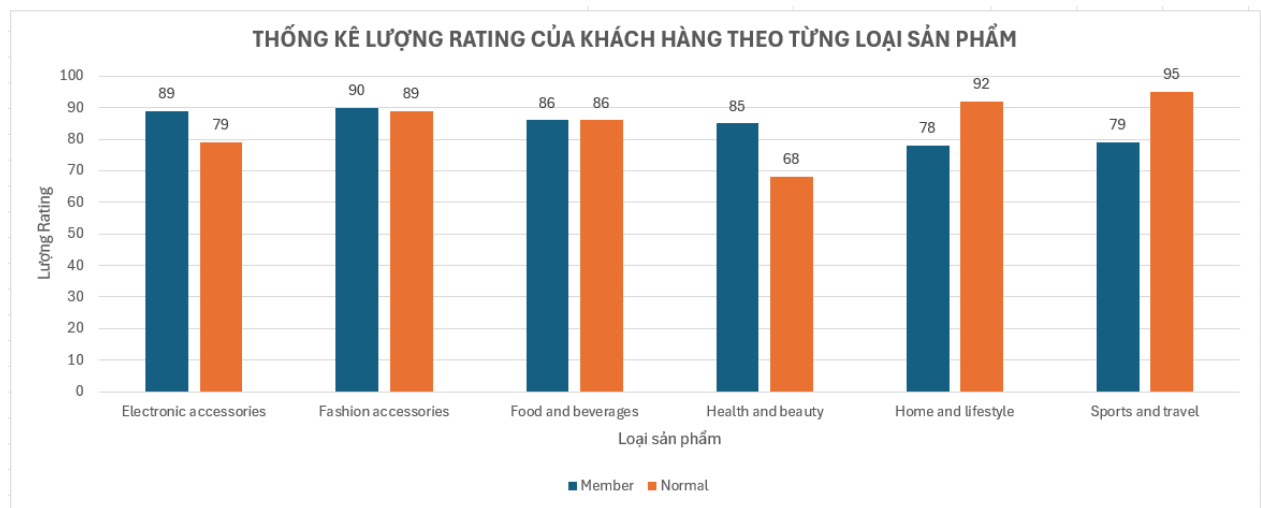
- Trong năm 2019, phương thức thanh toán Cash được sử dụng nhiều nhất qua các tháng, thấp nhất là phương thức Credit Card.
- Trong năm 2020, chỉ có duy nhất phương thức Ewallet được sử dụng vào tháng 2.

4.4. Thống kê số lượng rating của khách hàng (member, normal) theo từng loại sản phẩm (ProductLine)

MDX Query và kết quả:



Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Số lượng khách hàng mỗi loại khi đánh giá sản phẩm khá đồng đều, không quá chênh lệch, cho thấy sản phẩm được bày bán phù hợp với thị hiếu khách hàng, khách hàng rất quan tâm chất lượng sản phẩm.

4.5. Thống kê số lượng sản phẩm bán được theo từng thời điểm (time)

MDX Query và kết quả:

-- 5. Thống kê số lượng sản phẩm bán được theo từng thời điểm (time)

WITH

MEMBER [Measures].[SumProduct] AS SUM ([Measures].[Quantity])

SELECT [Measures].[SumProduct] ON COLUMNS,

{CROSSJOIN(

[Dim Date].[d Year].[d Year],

[Dim Date].[d Month].[d Month],

[Dim Date].[d Day].[d Day],[Dim Date].[d Hour].[d Hour]}}

ON ROWS

from [Supermarket_sales];

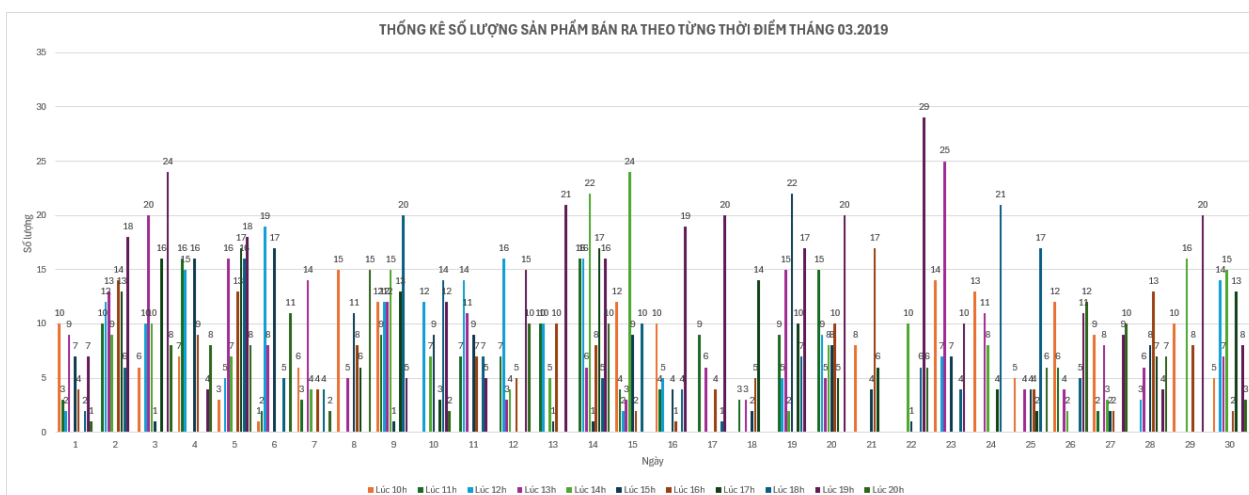
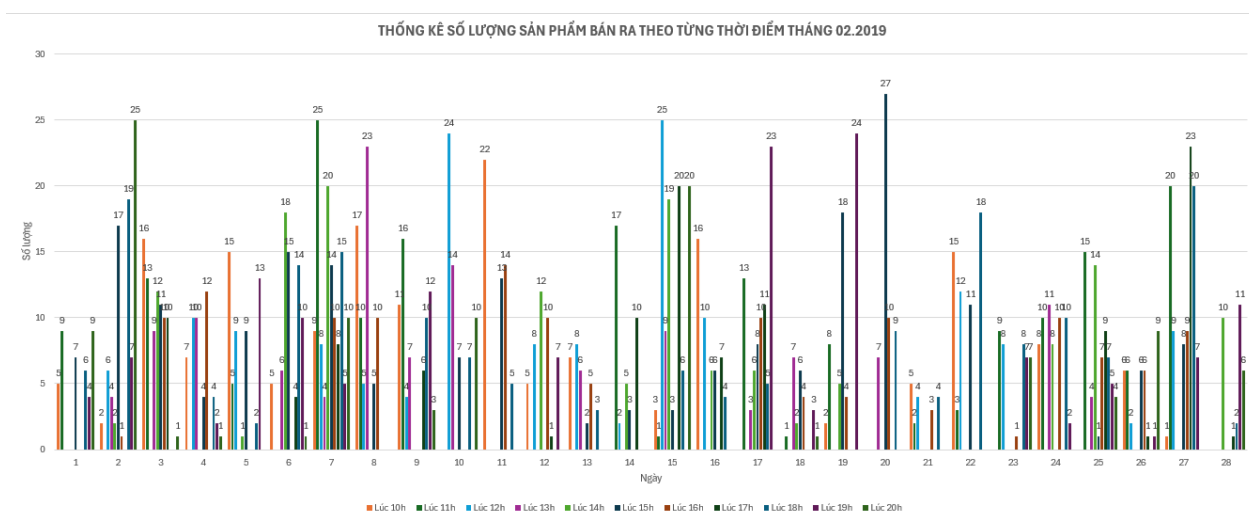
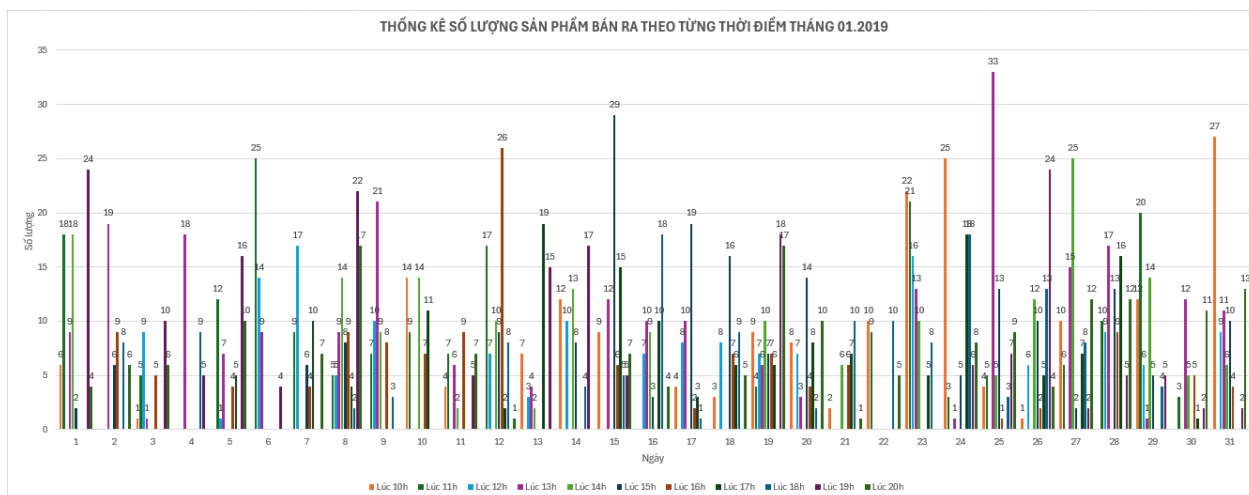
100 %

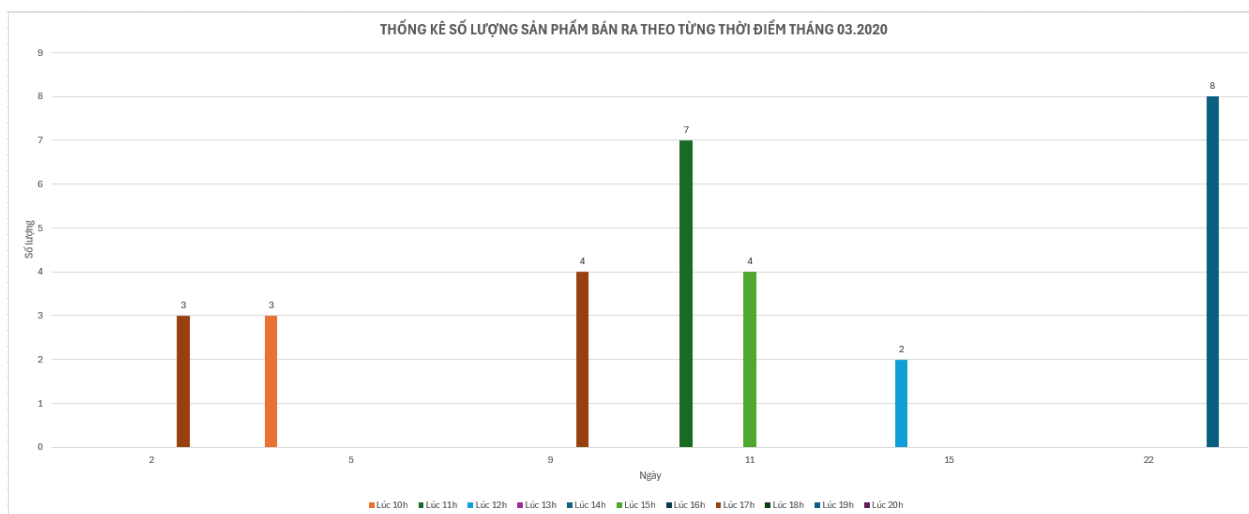
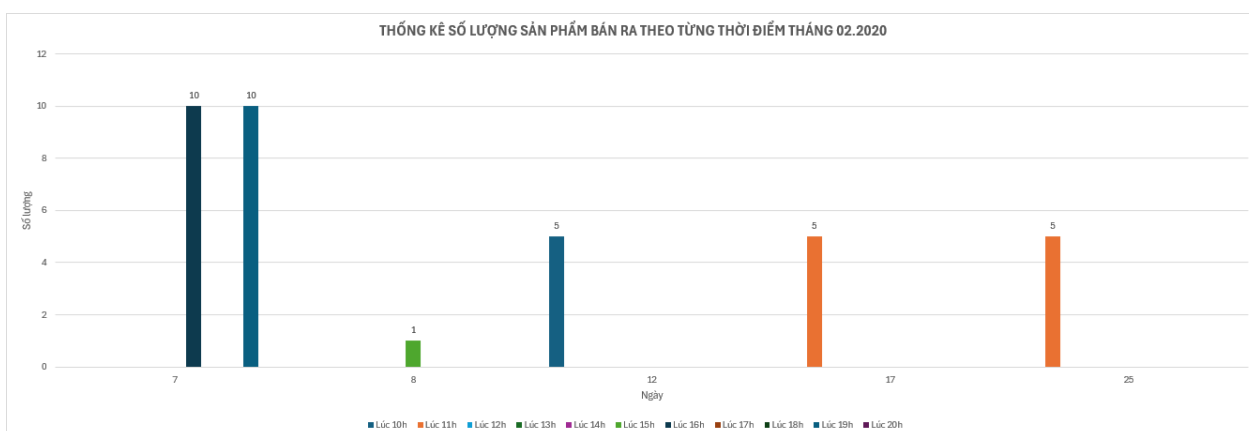
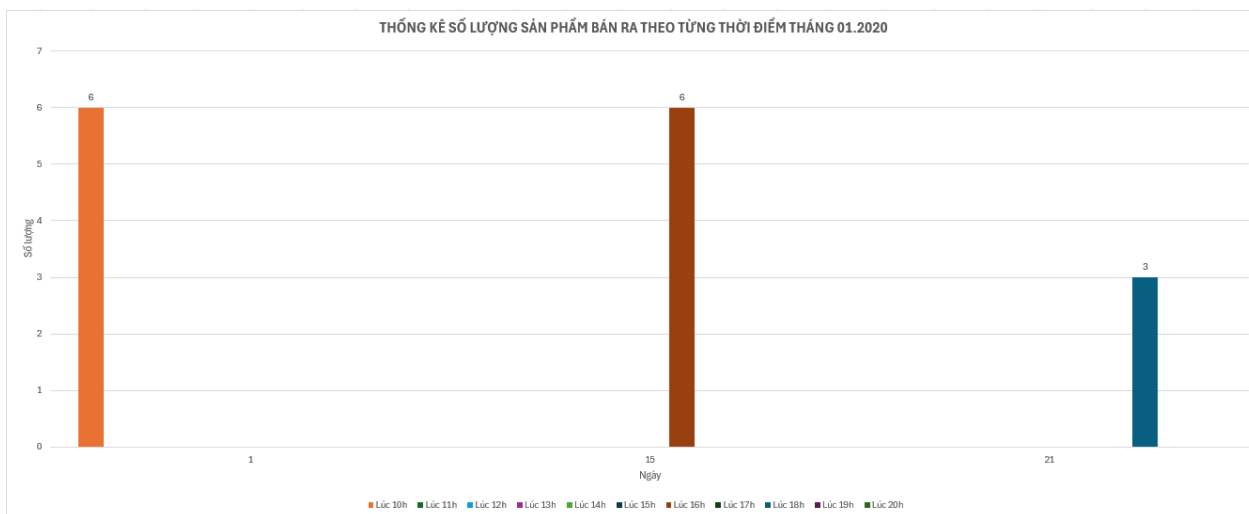
Messages

Results

				SumProduct
2019	1	1	10	6
2019	1	1	11	18
2019	1	1	13	9
2019	1	1	14	18
2019	1	1	15	2
2019	1	1	19	24
2019	1	1	20	4
2019	1	10	10	14
2019	1	10	11	9
2019	1	10	14	14
2019	1	10	16	7
2019	1	10	17	11
2019	1	11	10	4
2019	1	11	11	7
2019	1	11	13	6
2019	1	11	14	2
2019	1	11	16	9
2019	1	11	19	5
2019	1	11	20	7
2019	1	12	11	17
2019	1	12	12	7
2019	1	12	14	10
2019	1	12	15	9
2019	1	12	16	26

Trực quan hóa kết quả bằng biểu đồ Excel:





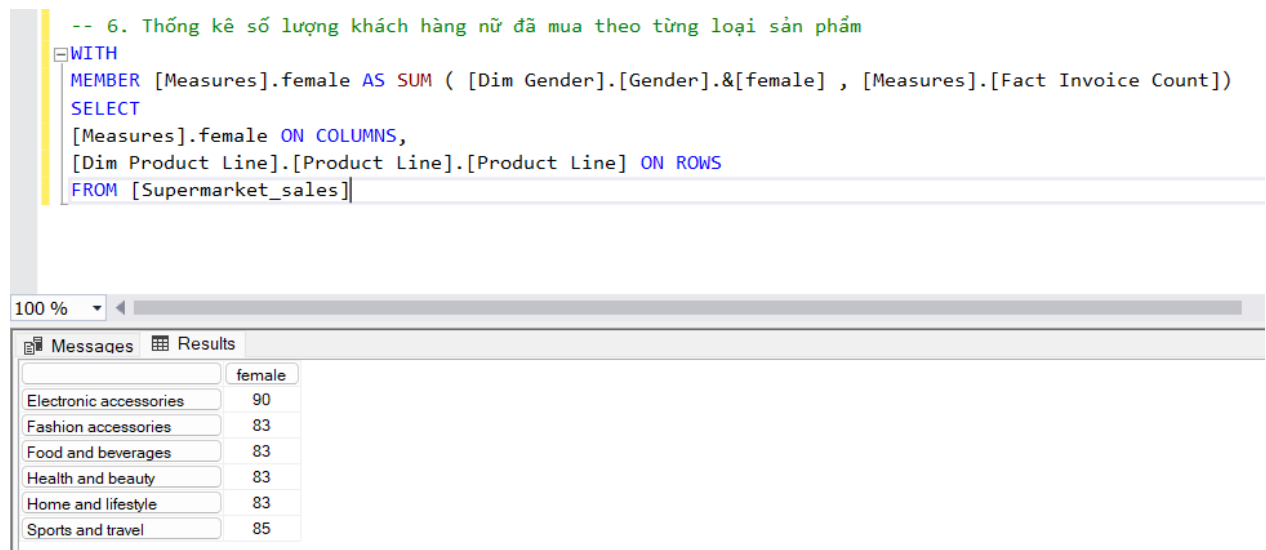
Nhận xét:

- Từ các biểu đồ trên, ta có thể nhận thấy lượng sản phẩm được bán cao nhất vào 3 tháng đầu năm 2019 vào các khung giờ 12h, 19h. Có thể giải thích cho điều này vì đây là 3 tháng đầu năm nên nhu cầu ăn mặc chỉ tiêu của người dân khá cao, và thường đi mua sắm vào các khung giờ nghỉ, giải lao của xã hội.
- Tuy nhiên, vào 3 tháng đầu năm 2020 thì số lượng sản phẩm bán ra cực ít, lẻ tẻ, có những ngày gần như không bán được sản phẩm nào. Có thể đã có vấn đề phát sinh ảnh hưởng tới kế hoạch buôn bán, doanh thu của doanh nghiệp như: phá sản, bê bối gây ảnh hưởng danh tiếng,...

4.6. Thống kê số lượng khách hàng nữ đã mua theo từng loại sản phẩm

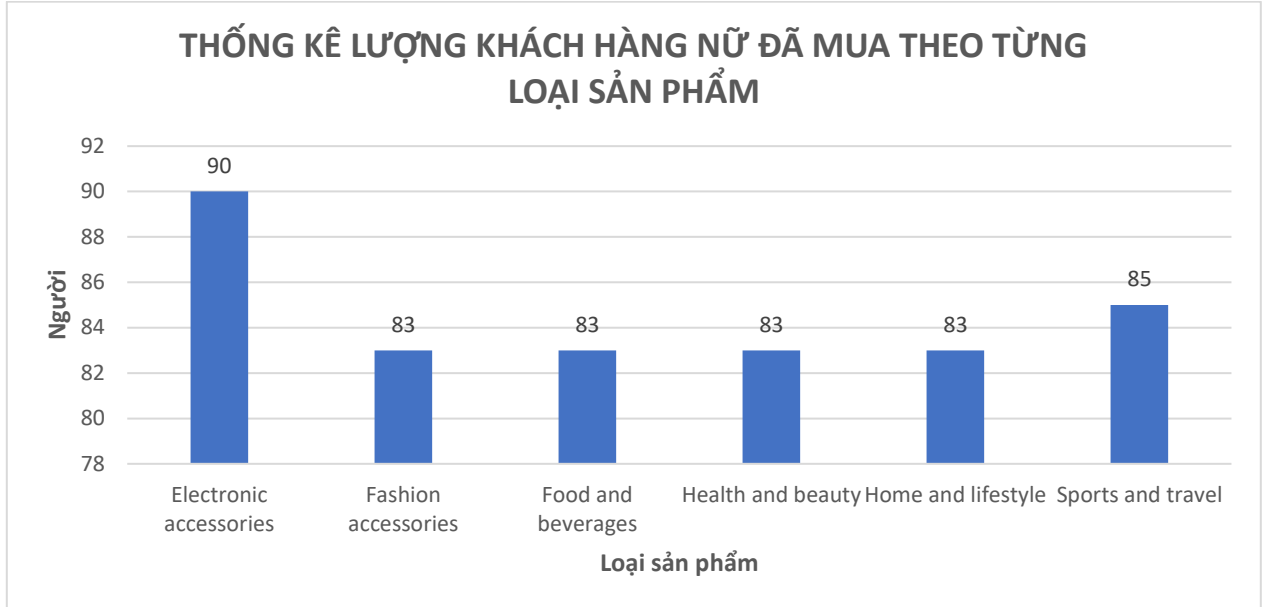
MDX Query và kết quả:

```
-- 6. Thống kê số lượng khách hàng nữ đã mua theo từng loại sản phẩm
WITH
MEMBER [Measures].female AS SUM ( [Dim Gender].[Gender].[female] , [Measures].[Fact Invoice Count])
SELECT
[Measures].female ON COLUMNS,
[Dim Product Line].[Product Line].[Product Line] ON ROWS
FROM [Supermarket_sales]
```



Product Line	Count
Electronic accessories	90
Fashion accessories	83
Food and beverages	83
Health and beauty	83
Home and lifestyle	83
Sports and travel	85

Trực quan hóa kết quả bằng biểu đồ Excel:



Nhận xét:

- Lượng khách hàng nữ mua loại sản phẩm “Phụ kiện điện tử” là cao nhất, cho thấy sự yêu thích, quan tâm của phái nữ đối với loại mặt hàng này.
- Với các loại sản phẩm còn lại, lượng khách hàng nữ đã mua sản phẩm của mỗi loại khá đồng đều, không quá chênh lệch, không thể hiện rõ độ phổ biến của mỗi loại.

5. Tạo job chạy tự động ETL

	Name [▲]	Enabled	Status	Last Run Outcome
[-]	Excel2Source	yes	Idle	Succeeded
[-]	NDS2DDS	yes	Idle	Succeeded
[-]	Source2Stage	yes	Idle	Succeeded
[-]	SSIS Server Maintenanc...	yes	Idle	Unknown
[-]	Stage2NDS	yes	Idle	Succeeded
[-]	syspolicy_purge_history	yes	Idle	Unknown

Bước 1: mở project SSIS trong Visual Studio, tiến hành deploy.

Lưu ý: cần tạo SSISDB trước ở trong SSMS.

Bước 2: kích hoạt SQL Server Agent.

Bước 3: ở job, chuột phải chọn New job.

Bước 4: đặt tên cho job ở phần Name.

Name:

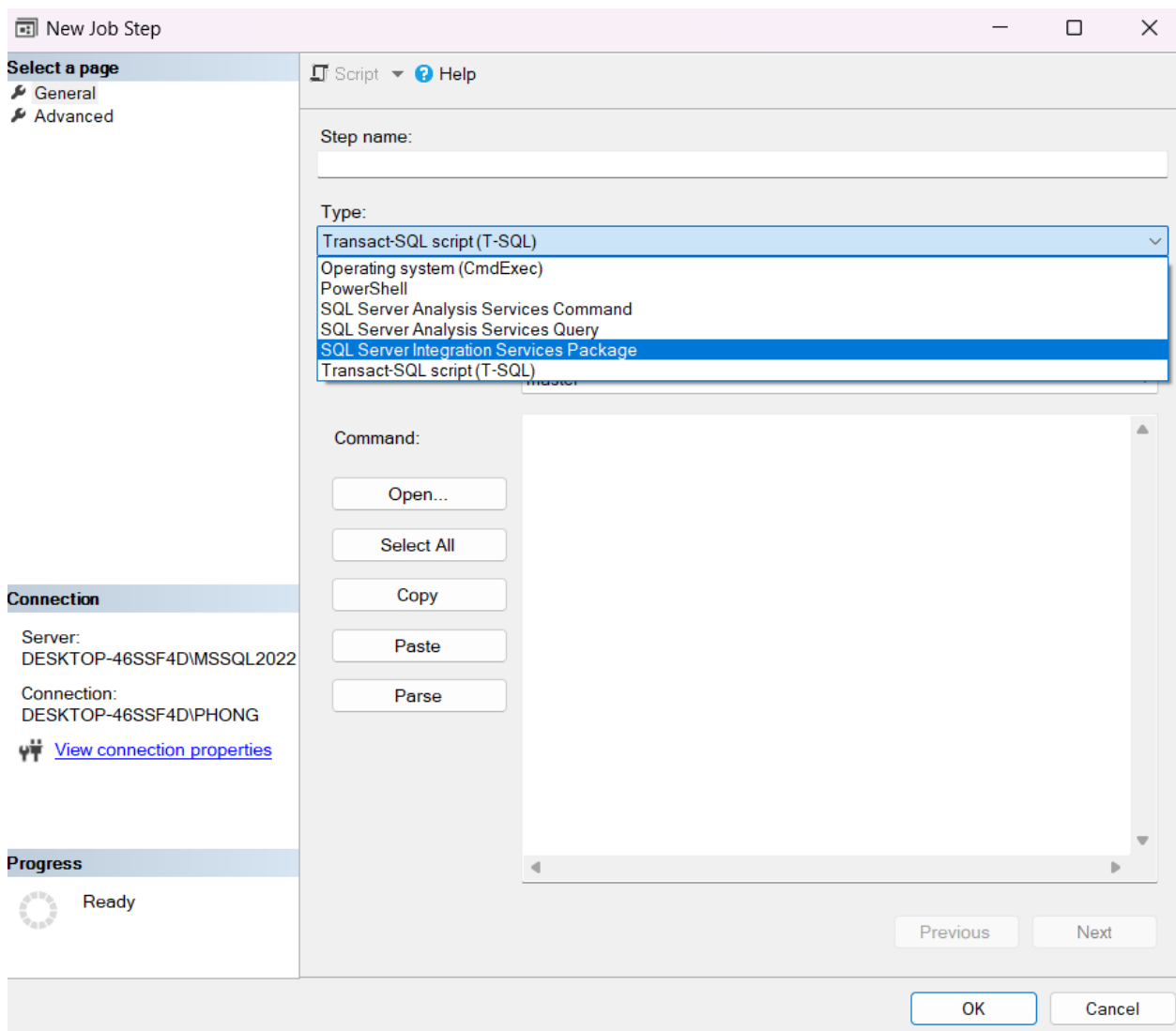
Owner: ...

Category: ...

Description:

☒ Enabled

Bước 5: chuyển qua phần Steps, nhấn vào New để tạo step mới.



Bước 6: ở phần type, chọn SQL Server Integration Services Package.

Bước 7: nhập tên server, chọn package cần lên lịch chạy tự động và chọn Ok.

New Job Step

Select a page

- General
- Advanced

Script Help

Step name:

Type:

SQL Server Integration Services Package

Run as:

SQL Server Agent Service Account

Package Configuration

Package source:

SSIS Catalog

Server:

Log on to the server

☒ Use Windows Authentication

☐ Use SQL Server Authentication

User name:

Password:

Package:

Previous Next

OK Cancel

Connection

Server:

DESKTOP-46SSF4D\MSSQL2022

Connection:

DESKTOP-46SSF4D\PHONG

[View connection properties](#)

Progress

Ready

Bước 8: chuyển qua Schedules, chọn New.

New Job Schedule

Name:

Schedule type: **Recurring** ☒ Enabled

One-time occurrence

Date: Time:

Frequency

Occurs: **Weekly**

Recurs every: week(s) on

☐ Monday ☐ Wednesday ☐ Friday ☐ Saturday
☐ Tuesday ☐ Thursday ☒ Sunday

Daily frequency

☒ Occurs once at:
☐ Occurs every: hour(s) Starting at:
Ending at:

Duration

Start date: ☐ End date:
☒ No end date:

Summary

Description:

OK Cancel Help

Bước 9: nhập tên cho Schedule, tùy chọn Occurs phù hợp, chọn thời gian cần thực hiện package tại Occurs once at, chọn ngày bắt đầu và nhấn Ok.

Lưu ý: ngày và thời gian được chọn cần phải trước thời gian hiện tại.

Bước 10: chờ đến đúng thời gian và job sẽ tự động chạy theo lịch đã lên sẵn.