

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN KHOA HỌC THÔNG
THÔNG TIN



BÁO CÁO ĐỒ ÁN MÔN HỌC

KHO DỮ LIỆU VÀ OLAP

ĐỀ TÀI:

**PHÂN TÍCH DỮ LIỆU CHUỖI CUNG ỨNG CỦA
CÔNG TY DATACO**

GVHD: ThS. Nguyễn Thị Kim Phụng

Nhóm: Nhóm 20

1. Lê Quang Nhật 20521705

2. Trần Anh Thy 21521517

TP. HỒ CHÍ MINH, NĂM 2023

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

....., ngày tháng năm 2023

Người nhận xét

(Ký tên và ghi rõ họ tên)

MỤC LỤC

BẢNG PHÂN CÔNG CÔNG VIỆC	7
TỔNG QUAN ĐỀ TÀI	9
CHƯƠNG 1. GIỚI THIỆU KHO DỮ LIỆU	10
1.1. Mô tả dataset.....	10
1.2. Lọc dữ liệu.....	11
1.3. Danh sách thuộc tính được phân tích	18
1.4. Mô tả chi tiết thuộc tính.....	22
1.1.1. Category	22
1.1.1.1. 1.4.2. Type	24
1.1.1.2. 1.4.3. Shipping Mode	24
1.1.1.3. 1.4.4. Delivery Status.....	25
1.5. 1.5. Sơ đồ dữ liệu xây dựng.....	25
1.5.1. Lược đồ hình sao (Star schema).....	25
1.5.2. Lược đồ hình bông tuyết (Snowflake schema)	26
1.6. 1.6. Xây dựng kho dữ liệu	27
1.6.1. Sơ đồ hình sao minh họa	27
1.6.2. Mô tả chi tiết các bảng dữ liệu	27
1.1.1.4. 1.6.2.1 Bảng Fact.....	27
1.1.1.5. 1.6.2.2. Dim_Customers	28
1.1.1.6. 1.6.2.3. Dim_Categories	28
1.1.1.7. 1.6.2.4. Dim_Products	28
1.1.1.8. 1.6.2.5. Dim_Locations	29
1.1.1.9. 1.6.2.6. Dim_Times.....	29
1.1.1.10. 1.6.2.7. Dim_Order_Statuses	29
1.1.1.11. 1.6.2.8. Dim_Type.....	29
1.1.1.12. 1.6.2.9. Dim_Shipping_Modes	30
1.1.1.13. 1.6.2.10. Dim_Delivery_Statuses.....	30
CHƯƠNG 2. QUÁ TRÌNH XÂY DỰNG KHO DỮ LIỆU (SSIS)	30
2.1. Chuẩn bị các công cụ.....	30

2.2.	Chuẩn bị cơ sở dữ liệu.....	33
2.3.	Tạo mới project SSIS	37
2.4.	Kiểm tra dữ liệu các bảng.....	98
	CHƯƠNG 3. PHÂN TÍCH DỮ LIỆU	103
3.1.	Cài đặt công cụ	103
3.1.1.	Cài đặt Microsoft Analysis Services Projects extension	103
3.1.2.	Cài đặt Analysis Services	106
3.2.	Tạo mới Project SSAS.....	114
3.3.	Xác định dữ liệu nguồn (Data Sources)	115
3.4.	Xác định Data Source Views.....	120
3.5.	Xây dựng khối (Cubes) và xác định các độ đo (Measures).....	123
3.6.	Xác định các chiều (Dimensions).....	128
3.6.1.	Cấu hình deployment. Cấu hình và tạo thuộc tính phân cấp cho “Dim Products.dim”	133
3.6.2.	Cấu hình và tạo thuộc tính phân cấp cho “Dim Locations.dim”	146
3.6.3.	Cấu hình và tạo thuộc tính phân cấp cho “Dim Times.dim”	162
3.6.4.	Cấu hình cho “Dim Statuses.dim”	179
3.6.5.	Cấu hình cho “Dim Customers.dim”.....	181
3.6.6.	Cấu hình cho “Dim Shipping Modes.dim”	183
3.6.7.	Cấu hình cho “Dim Delivery Statuses.dim”.....	185
3.7.	Độ đo, KPI, Name Set	187
3.7.1.	Đổi tên và thuộc tính các độ đo ban đầu.	187
3.7.2.	Tạo các độ đo mới	189
3.7.3.	Tạo KPI	192
3.7.3.1.	Tạo KPI về sự tăng trưởng lợi nhuận. Giá trị là trung bình lợi nhuận trên toàn thế giới.....	192
3.7.4.	Tạo nameset.....	197
3.7.4.1.	Top 5 sản phẩm được mua nhiều nhất	197
3.7.4.2.	Danh sách các nước có doanh thu âm	199
3.7.4.3.	Top 5 quốc gia có lợi nhuận cao nhất không phải Mỹ.....	201
3.7.4.4.	Danh sách các khách hàng đem lại lợi nhuận > 700 từ Cash hoặc > 1200 bằng Debit.....	206

3.7.4.5.	Danh sách Top 10 khách hàng đem lại lợi nhuận nhiều nhất....	208
3.7.4.6.	Danh sách Top 1 quốc gia đem lại nhiều lợi nhuận nhất	210
3.7.4.7.	Danh sách top 10 quốc gia đem lại ít lợi nhuận nhất (không âm) 212	
3.7.5.	Tạo Calculated Member	215
	Tạo Calculated “Late Risk”	215
3.8.	15 câu truy vấn	218
3.8.1.	Các truy vấn về Việt Nam	218
	Câu truy vấn 1: Thống kê số lượng item được bán ra và tổng lợi nhuận theo tháng và bắn tại Việt Nam.....	218
	Câu truy vấn 2: Thống kê doanh thu, số lượng item bán ra, số ngày giao hàng trung bình trên kế hoạch và thực tế tại Việt Nam, Thái Lan và Trung Quốc trong năm 2017	220
	Câu truy vấn 3: Thống kê KPI doanh thu mua hàng tại Việt Nam trong hàng tháng trong năm có vượt qua mức trung bình chung, sử dụng KPI Growth .	222
	Câu truy vấn 4: Thống kê KPI tăng trưởng doanh thu mua hàng tại Việt Nam trong năm.	225
	Câu truy vấn 5: Thống kê top 5 sản phẩm (cùng với tên loại sản phẩm) được mua nhiều nhất tại Việt Nam	227
3.8.2.	Các truy vấn về sales.....	229
	Câu truy vấn 6: Liệt kê danh sách các nước có lợi nhuận, sắp xếp từ nhỏ đến lớn nhất	229
	Câu truy vấn 7: Tổng doanh thu và lợi nhuận của châu Âu, sắp xếp theo tháng năm.....	231
	Câu truy vấn 8: Liệt kê danh sách các nước không phải Mỹ đem lại nhiều lợi nhuận nhất.....	233
	Câu truy vấn 9: Danh sách 5 nước có doanh thu thấp nhất (không âm).....	235
	Câu truy vấn 10: Danh sách các sản phẩm luôn nằm trong top 10 doanh thu trong các năm 2015-16-17	236
	Câu truy vấn 11: Tình trạng giao hàng tại châu Á – Thái Bình Dương.....	238
	Câu truy vấn 12: Thống kê số lượng item trong đơn hàng được tạo trong các thứ trong tuần.....	240

Câu truy vấn 13: Thống kê tình trạng giao hàng của 4 loại vận chuyển tại Trung Quốc	242
3.8.3. Các câu truy vấn về khách hàng.....	245
Câu truy vấn 14: Danh sách khách hàng đem lại lợi nhuận từ Cash > 700 hoặc từ Debit trên 1200.....	245
Câu truy vấn 15: Top 10 khách hàng đem lại lợi nhuận nhiều nhất từ quốc gia đem lại lợi nhuận nhiều nhất.....	247
Chương 4: SSRS.....	248
4.1. Cài đặt các công cụ cần thiết	248
4.2. Cài đặt Report Server Project.....	248
4.3. Tạo và cấu hình cho SSRS Project	249
4.4. Tạo và cấu hình cho Power BI	253
4.5. Các Reports	278
4.5.1. Report Builder.....	278
4.5.1.1. Vietnam China Benefit	278
4.5.1.2. Revenue of the markets	279
4.5.1.3. Top sale each shipping mode through year	279
4.5.2. Power BI.....	280
4.5.2.1. Vietnam China Benefit	280
4.5.2.2. Revenue of the markets	280
4.5.2.3. Top sale each shipping mode through year	281
CHƯƠNG 5: QUÁ TRÌNH KHAI THÁC DỮ LIỆU (DATA MINING)	281
5.1. Import các thư viện cần thiết	281
5.2. Đọc file dataset	282
5.3. Kiểm tra giá trị null	282
5.4. Thay đổi dữ liệu và chọn ra những cột cần thiết	282
5.5. Chia data test và train theo tỉ lệ 7:3	283
5.6. Khởi chạy và đánh giá thuật toán	283
DANH MỤC TÀI LIỆU THAM KHẢO	286

BẢNG PHÂN CÔNG CÔNG VIỆC

BẢNG PHÂN CÔNG CÔNG VIỆC		
	Trần Anh Thy	Lê Quang Nhật
Nguồn dữ liệu, mô tả chi tiết dữ liệu	Tìm nguồn dataset đồ án.	Tìm nguồn data bài tập lab A1.1 A1.2
Xử lý dữ liệu	<ul style="list-style-type: none"> - Hoàn thiện preprocess, - Preprocess chia dims fact, phân bảng, lưu vào excel. 	<ul style="list-style-type: none"> - Đọc dữ liệu - Preprocess data lần 2. - Quay video preprocess.
Mô tả dữ liệu, đồ án	<ul style="list-style-type: none"> - Mô tả kiểu dữ liệu SSIS 	<ul style="list-style-type: none"> - Mô tả dữ liệu dataset gốc - Mô tả dự án, giới thiệu dự án
SSIS	<ul style="list-style-type: none"> - Viết Script SQL database - SSIS process - Quay Videos SSIS 	<ul style="list-style-type: none"> - Vẽ Schema - Script SQL database
SSAS	<ul style="list-style-type: none"> - SSAS process, Power BI, Cube, Mdx. - Truy vấn trên chủ đề Vietnam, Sales, Customers - Quay video SSAS 	<ul style="list-style-type: none"> - Excel - Truy vấn trên chủ đề Logistics
SSRS	Thực hiện chính	-
Data Mining	-	Thực hiện chính
Làm Slide	50%	50%
Viết báo cáo	50%	50%
Format và chỉnh lỗi chính tả trong báo cáo	50%	50%

Mức độ hoàn thành công việc	100%	100%
------------------------------------	------	------

TỔNG QUAN ĐỀ TÀI

Trong thế giới kinh doanh hiện đại, quản lý chuỗi cung ứng không chỉ là một phần quan trọng mà còn là yếu tố quyết định sự thành công của một doanh nghiệp. Để hiểu rõ hơn về tầm quan trọng của việc này, chúng tôi đã tiến hành phân tích một bộ dữ liệu về chuỗi cung ứng được sử dụng bởi công ty DataCo Global.

Bộ dữ liệu này không chỉ cung cấp thông tin về các hoạt động quan trọng trong chuỗi cung ứng như cung cấp, sản xuất, bán hàng, phân phối thương mại mà còn cho phép liên kết dữ liệu có cấu trúc với dữ liệu không có cấu trúc để tạo ra kiến thức mới.

Trong môi trường kinh doanh ngày nay, nhanh chóng và dựa trên dữ liệu, quản lý chuỗi cung ứng đã trở thành một yếu tố quan trọng cho sự thành công của một doanh nghiệp. Báo cáo này sẽ đi sâu vào phân tích dữ liệu chuỗi cung ứng, nơi mà phân tích các bộ dữ liệu toàn diện cung cấp những thông tin quý giá giúp tối ưu hóa hoạt động hiệu quả và đưa ra quyết định chiến lược, chủ yếu là về hiệu suất giao hàng.

Mục tiêu của chúng tôi là phát triển một bảng điều khiển hiệu suất giao hàng toàn diện để tạo điều kiện cho quyết định dựa trên dữ liệu và phát triển chiến lược kinh doanh thông minh. Chúng tôi sẽ bao gồm biến thể thời gian giao hàng để phát hiện bất kỳ sai lệch nào so với thời gian giao hàng dự kiến và đánh giá hiệu quả của quá trình giao hàng. Ngoài ra, chúng tôi cũng sẽ tính toán tỷ lệ giao hàng đúng hẹn để cho phép người dùng theo dõi và tối ưu hóa hiệu suất vận chuyển dựa trên các chỉ số hiệu suất chính (KPIs). Cuối cùng, chúng tôi sẽ cung cấp thông tin về phân phối khu vực thông qua bảng điều khiển, giúp người dùng ra quyết định có thông tin về các chuyến đi của đội vận chuyển, hợp tác với các nhà cung cấp bên thứ ba và cơ hội mở rộng tiềm năng.

CHƯƠNG 1. GIỚI THIỆU KHO DỮ LIỆU

1.1. Mô tả dataset

Tên bộ dữ liệu: DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS.

The screenshot shows the dataset page for "DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS" on Mendeley Data. At the top, there's a navigation bar with a user profile icon, the title, a '241' badge, a 'New Notebook' button, a download link ('Download (27 MB)'), and other options. Below the title, there's a thumbnail image of boxes in a warehouse. The main content area has a section titled "About Dataset". It includes a "Content" section describing a dataset of supply chains used for analysis, mentioning machine learning algorithms and R software. It also lists important activities like provisioning, production, sales, and commercial distribution, noting the correlation of structured and unstructured data. A "Usability" rating of 7.65 is shown. The "License" is CC0: Public Domain. The "Expected update frequency" is listed as "Never". Under "Tags", "Data Visualization" and "Exploratory Data Analysis" are selected. The "Acknowledgements & Source" section lists contributors: Fabian Constante, Instituto Politecnico de Leiria Escola Superior de Tecnologia e Gestao, and Fernando Silva, António Pereira. A link to the dataset on Mendeley is provided at the bottom: <https://data.mendeley.com/datasets/8gx2fg2k6/5>.

Đây là tập dữ liệu về chuỗi cung ứng được công ty DataCo sử dụng để phục vụ cho việc phân tích dữ liệu lớn. Bộ dữ liệu chứa thông tin về các hoạt động chuỗi cung ứng như khách hàng, đơn đặt, phân phối, vận chuyển, sản phẩm từ 2015 đến 2018.

Cập nhật gần nhất: 13/03/2019.

Tên tác giả: Fabian Constante, Fernando Silva, António Pereira

Bộ dữ liệu gồm 3 tệp: tệp chính gồm 180520 dòng dữ liệu và 53 cột thuộc tính, một tệp chứa thông tin mô tả các cột dữ liệu, và một tệp lưu thông tin log truy cập.

Sau khi lọc dữ liệu ta được 50 thuộc tính để sử dụng cho việc phân tích đề tài.

Nguồn: [DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS - Mendeley Data](#)

1.2. Lọc dữ liệu

Sử dụng python để lọc dữ liệu

Bước 1: Nhập các thư viện cần thiết và đọc file dataset

```
import pandas as pd
# writer = pd.ExcelWriter('dataframes.xlsx', engine='xlsxwriter')

DataCoSupplyChainDataset = pd.read_csv(".\DataCoSupplyChainDataset\DataCoSupplyChainDataset.csv", encoding='latin1')
DescriptionDataCoSupplyChain = pd.read_csv(".\DataCoSupplyChainDataset\DescriptionDataCoSupplyChain.csv")
```

DescriptionDataCoSupplyChain		
[4]	FIELDS	DESCRIPTION
0	Type	: Type of transaction made
1	Days for shipping (real)	: Actual shipping days of the purchased product
2	Days for shipment (scheduled)	: Days of scheduled delivery of the purchased...
3	Benefit per order	: Earnings per order placed
4	Sales per customer	: Total sales per customer made per customer
5	Delivery Status	: Delivery status of orders: Advance shipping...
6	Late_delivery_risk	: Categorical variable that indicates if send...
7	Category Id	: Product category code
8	Category Name	: Description of the product category
9	Customer City	: City where the customer made the purchase
10	Customer Country	: Country where the customer made the purchase
11	Customer Email	: Customer's email
12	Customer Fname	: Customer name
13	Customer Id	: Customer ID
14	Customer Lname	: Customer lastname
15	Customer Password	: Masked customer key
16	Customer Segment	: Types of Customers: Consumer , Corporate , ...
17	Customer State	: State to which the store where the purchase...
18	Customer Street	: Street to which the store where the purchase...
19	Customer Zipcode	: Customer Zipcode
20	Department Id	: Department code of store
21	Department Name	: Department name of store
22	Latitude	: Latitude corresponding to location of store

DataCoSupplyChainDataset.head()														
	Type	Days for shipping (real)	Days for shipment (scheduled)	Benefit per order	Sales per customer	Delivery Status	Late_delivery_risk	Category Id	Category Name	Customer City	...	Order Zipcode	Product Card Id	Product Category Id
0	DEBIT	3	4	91.250000	314.640015	Advance shipping	0	73	Sporting Goods	Caguas	...	NaN	1360	73
1	TRANSFER	5	4	-249.089996	311.359985	Late delivery	1	73	Sporting Goods	Caguas	...	NaN	1360	73
2	CASH	4	4	-247.779999	309.720001	Shipping on time	0	73	Sporting Goods	San Jose	...	NaN	1360	73
3	DEBIT	3	4	22.860001	304.809998	Advance shipping	0	73	Sporting Goods	Los Angeles	...	NaN	1360	73
4	PAYMENT	2	4	134.210007	298.250000	Advance shipping	0	73	Sporting Goods	Caguas	...	NaN	1360	73

5 rows x 53 columns

Bước 2: Lọc các dòng dữ liệu trùng nhau và loại bỏ các cột có quá nhiều giá trị NaN.

```
print("origin shape: " + str(DataCoSupplyChainDataset.shape) + " -- Shape after Drop duplicate: " + str(DataCoSupplyChainDataset.drop_duplicates().shape))

origin height: (180519, 53) -- Drop duplicate: (180519, 53)
```

DataCoSupplyChainDataset.drop(columns=['Product Description', 'Order Zipcode', 'Customer Zipcode'], axis=1, inplace=True) DataCoSupplyChainDataset.dropna()															
	Type	Days for shipping (real)	Days for shipment (scheduled)	Benefit per order	Sales per customer	Delivery Status	Late_delivery_risk	Category Id	Category Name	Customer City	...	Order State	Order Status	Product Card Id	Product Category Id
0	DEBIT	3	4	91.250000	314.640015	Advance shipping	0	73	Sporting Goods	Caguas	...	Java Occidental	COMPLETE	1360	73
1	TRANSFER	5	4	-249.089996	311.359985	Late delivery	1	73	Sporting Goods	Caguas	...	Rajastán	PENDING	1360	73
2	CASH	4	4	-247.779999	309.720001	Shipping on time	0	73	Sporting Goods	San Jose	...	Rajastán	CLOSED	1360	73
3	DEBIT	3	4	22.860001	304.809998	Advance shipping	0	73	Sporting Goods	Los Angeles	...	Queensland	COMPLETE	1360	73
4	PAYMENT	2	4	134.210007	298.250000	Advance shipping	0	73	Sporting Goods	Caguas	...	Queensland	PENDING_PAYMENT	1360	73
...	
180514	CASH	4	4	40.000000	399.980011	Shipping on time	0	45	Fishing	Brooklyn	...	Shanghái	CLOSED	1004	45
180515	DEBIT	3	2	-613.770019	395.980011	Late delivery	1	45	Fishing	Bakersfield	...	Osaka	COMPLETE	1004	45
180516	TRANSFER	5	4	141.110001	391.980011	Late delivery	1	45	Fishing	Bristol	...	Australia del Sur	PENDING	1004	45

SỬ DỤNG PYTHON TÁCH FILE THÀNH CÁC DIM VÀ FACT.

Bước 3: Dim LOCATION - Đánh index Location_Id theo nhóm Market, OrderCity, OrderCountry, OrderRegion, OrderState.

```
# Make a copy of the DataFrame
TempData = DataCoSupplyChainDataset.copy()

# Reset index after dropping duplicates in TempData
TempData.reset_index(drop=True, inplace=True)

# Add an ID for each unique combination of location columns in DataCoSupplyChainDataset
DataCoSupplyChainDataset['Location Id'] = TempData.groupby(['Market', 'Order City', 'Order Country', 'Order Region', 'Order State']).ngroup() + 1
DataCoSupplyChainDataset['Location Id']

[✓ 0.1s]

0      2330
1      2351
2      2351
3      3022
4      3022
...
180514    2918
180515    2535
180516    2229
180517    2229
180518    2777

Name: Location Id, Length: 180519, dtype: int64
```

```
# List of location-related columns
location_columns = ['Location Id', 'Market', 'Order City', 'Order Country', 'Order Region', 'Order State']

# Create a new DataFrame containing only location-related columns and drop duplicates
dim_locations = DataCoSupplyChainDataset[location_columns].drop_duplicates()

# Reset index after dropping duplicates
dim_locations.reset_index(drop=True, inplace=True)

# Sort dim_locations DataFrame based on 'Location Id'
dim_locations_sorted = dim_locations.sort_values(by='Location Id')

# Reset index after sorting
dim_locations_sorted.reset_index(drop=True, inplace=True)
```

Bước 4: Dim TIMES – Đánh index Time_Id cho từng nhóm Dates mà đã bỏ đi các giá trị trùng lặp.

```
DataCoSupplyChainDataset['shipping date (DateOrders)']

0      2/3/2018 22:56
1      1/18/2018 12:27
2      1/17/2018 12:06
3      1/16/2018 11:45
4      1/15/2018 11:24
...
180514    1/20/2016 3:40
180515    1/19/2016 1:34
180516    1/20/2016 21:00
180517    1/18/2016 20:18
180518    1/19/2016 18:54

Name: shipping date (DateOrders), Length: 180519, dtype: object

# Make a copy of the DataFrame
TempData = DataCoSupplyChainDataset.copy()

# Drop duplicates from TempData
TempData['shipping date (DateOrders)'].drop_duplicates()

# Reset index after dropping duplicates in TempData
TempData.reset_index(drop=True, inplace=True)

# Add an ID for each unique combination of location columns in DataCoSupplyChainDataset
DataCoSupplyChainDataset['Time Id'] = TempData.groupby(['shipping date (DateOrders)']).ngroup() + 1
```

```

# List of time-related columns
time_columns = ['Time Id', 'shipping date (DateOrders)']

# Create a new DataFrame containing only time-related columns and drop duplicates
dim_times = DataCoSupplyChainDataset[time_columns].drop_duplicates()

# Reset index after dropping duplicates
dim_times.reset_index(drop=True, inplace=True)

# Sort dim_times DataFrame based on 'Time Id'
dim_times_sorted = dim_times.sort_values(by='Time Id')

# Reset index after sorting
dim_times_sorted.reset_index(drop=True, inplace=True)

```

Bước 5: Tương tự với Dim STATUS

```

# Make a copy of the DataFrame
TempData = DataCoSupplyChainDataset.copy()

# Drop duplicates from TempData
TempData['Order Status'].drop_duplicates()

# Reset index after dropping duplicates in TempData
TempData.reset_index(drop=True, inplace=True)

# Add an ID for each unique combination of location columns in DataCoSupplyChainDataset
DataCoSupplyChainDataset['Order Status Id'] = TempData.groupby(['Order Status']).ngroup() + 1

```

Bước 6: Dim TYPE

```

# Make a copy of the DataFrame
TempData = DataCoSupplyChainDataset.copy()

# Drop duplicates from TempData
TempData['Type'].drop_duplicates()

# Reset index after dropping duplicates in TempData
TempData.reset_index(drop=True, inplace=True)

# Add an ID for each unique combination of location columns in DataCoSupplyChainDataset
DataCoSupplyChainDataset['Type Id'] = TempData.groupby(['Type']).ngroup() + 1

```

```

# List of time-related columns
type_columns = ['Type Id', 'Type']

# Create a new DataFrame containing only time-related columns and drop duplicates
types = DataCoSupplyChainDataset[type_columns].drop_duplicates()

# Reset index after dropping duplicates
types.reset_index(drop=True, inplace=True)

# Sort order_statuses DataFrame based on 'Type Id'
dim_types_sorted = types.sort_values(by='Type Id')

# Reset index after sorting
dim_types_sorted.reset_index(drop=True, inplace=True)

```

Bước 7: Dim SHIPPING_MODE

```
# Make a copy of the DataFrame
TempData = DataCoSupplyChainDataset.copy()

# Drop duplicates from TempData
TempData['Shipping Mode'].drop_duplicates()

# Reset index after dropping duplicates in TempData
TempData.reset_index(drop=True, inplace=True)

# Add an ID for each unique combination of location columns in DataCoSupplyChainDataset
DataCoSupplyChainDataset['Shipping Mode Id'] = TempData.groupby(['Shipping Mode']).ngroup() + 1
```

```
# List of time-related columns
shipping_mode_columns = ['Shipping Mode Id', 'Shipping Mode']

# Create a new DataFrame containing only time-related columns and drop duplicates
shipping_modes = DataCoSupplyChainDataset[shipping_mode_columns].drop_duplicates()

# Reset index after dropping duplicates
shipping_modes.reset_index(drop=True, inplace=True)

# Sort order_statuses DataFrame based on 'Type Id'
dim_shipping_modes_sorted = shipping_modes.sort_values(by='Shipping Mode Id')

# Reset index after sorting
dim_shipping_modes_sorted.reset_index(drop=True, inplace=True)
```

Bước 8: Dim DELIVERY_STATUS

```
# Make a copy of the DataFrame
TempData = DataCoSupplyChainDataset.copy()

# Drop duplicates from TempData
TempData['Delivery Status'].drop_duplicates()

# Reset index after dropping duplicates in TempData
TempData.reset_index(drop=True, inplace=True)

# Add an ID for each unique combination of location columns in DataCoSupplyChainDataset
DataCoSupplyChainDataset['Delivery Status Id'] = TempData.groupby(['Delivery Status']).ngroup() + 1
```

```
# List of time-related columns
delivery_status_columns = ['Delivery Status Id', 'Delivery Status']

# Create a new DataFrame containing only time-related columns and drop duplicates
delivery_statuses = DataCoSupplyChainDataset[delivery_status_columns].drop_duplicates()

# Reset index after dropping duplicates
delivery_statuses.reset_index(drop=True, inplace=True)

# Sort order_statuses DataFrame based on 'Delivery Status'
dim_delivery_statuses_sorted = delivery_statuses.sort_values(by='Delivery Status Id')

# Reset index after sorting
dim_delivery_statuses_sorted.reset_index(drop=True, inplace=True)
```

Bước 9: Dim CUSTOMER

```
# List of customer-related columns
customer_columns = ['Customer Id', 'Customer Fname', 'Customer Lname']

# Create a new DataFrame containing only customer-related columns and drop duplicates
dim_customers = DataCoSupplyChainDataset[customer_columns].drop_duplicates()

# Reset index after dropping duplicates
dim_customers.reset_index(drop=True, inplace=True)

# Update the DataFrame (if needed)

# Sort dim_customers DataFrame based on 'Customer Id'
dim_customers_sorted = dim_customers.sort_values(by='Customer Id')

# Reset index after sorting
dim_customers_sorted.reset_index(drop=True, inplace=True)
```

Bước 11: Dim CATEGORY

```
# List of category-related columns
category_columns = ['Category Id', 'Category Name']

# Create a new DataFrame containing only category-related columns and drop duplicates
dim_categories = DataCoSupplyChainDataset[category_columns].drop_duplicates()

# Reset index after dropping duplicates
dim_categories.reset_index(drop=True, inplace=True)

# Sort dim_categories DataFrame based on 'Category Id'
dim_categories_sorted = dim_categories.sort_values(by='Category Id')

# Reset index after sorting
dim_categories_sorted.reset_index(drop=True, inplace=True)
```

Bước 12: Dim PRODUCT

```
# List of product-related columns
product_columns = ['Product Card Id', 'Product Category Id', 'Product Name', 'Product Price']

# Create a new DataFrame containing only product-related columns and drop duplicates
dim_products = DataCoSupplyChainDataset[product_columns].drop_duplicates()

# Reset index after dropping duplicates
dim_products.reset_index(drop=True, inplace=True)

# Sort dim_products DataFrame based on 'Product Card Id'
dim_products_sorted = dim_products.sort_values(by='Product Card Id')

# Reset index after sorting
dim_products_sorted.reset_index(drop=True, inplace=True)
```

Bước 13: Dim ORDER

```
# List of time-related columns
order_status_columns = ['Order_Status_Id', 'Order_Status']

# Create a new DataFrame containing only time-related columns and drop duplicates
order_statuses = DataCoSupplyChainDataset[order_status_columns].drop_duplicates()

# Reset index after dropping duplicates
order_statuses.reset_index(drop=True, inplace=True)

# Sort order_statuses DataFrame based on 'Order_Status_Id'
dim_order_statuses_sorted = order_statuses.sort_values(by='Order_Status_Id')

# Reset index after sorting
dim_order_statuses_sorted.reset_index(drop=True, inplace=True)
```

Bước 14: Dim ORDER_STATUS

```
# List of time-related columns
order_status_columns = ['Order_Status_Id', 'Order_Status']

# Create a new DataFrame containing only time-related columns and drop duplicates
order_statuses = DataCoSupplyChainDataset[order_status_columns].drop_duplicates()

# Reset index after dropping duplicates
order_statuses.reset_index(drop=True, inplace=True)

# Sort order_statuses DataFrame based on 'Order_Status_Id'
dim_order_statuses_sorted = order_statuses.sort_values(by='Order_Status_Id')

# Reset index after sorting
dim_order_statuses_sorted.reset_index(drop=True, inplace=True)
```

Bước 15: FACT TABLE

```
fact_order_item_columns = ['Order_Item_Id',
                           'Location_Id', 'Time_Id',
                           'Product_Card_Id', 'Customer_Id',
                           'Order_Status_Id', 'Type_Id', 'Shipping_Mode_Id', 'Delivery_Status_Id',
                           'Days_for_shipping_(real)', 'Days_for_shipment_(scheduled)',
                           'Order_Item_Quantity', 'Order_Item_Discount_Rate',
                           'Order_Item_Profit_Ratio']

fact_order_items = DataCoSupplyChainDataset[fact_order_item_columns]
```

Bước 16: Lưu các dim và Fact table chuẩn bị ra file Excel cho các bước xử lý sau này

```
# Specify the full path along with the file name
file_path = 'Preprocessed_DataCoSupplyChainDataset.xlsx'

# Write each DataFrame to a separate worksheet
with pd.ExcelWriter(file_path, engine='xlsxwriter') as writer:
    ... dim_customers_sorted.to_excel(writer, sheet_name='dim_customers', index=False)
    ... dim_categories_sorted.to_excel(writer, sheet_name='dim_categories', index=False)
    ... dim_products_sorted.to_excel(writer, sheet_name='dim_products', index=False)
    ... dim_locations_sorted.to_excel(writer, sheet_name='dim_locations', index=False)
    ... dim_times_sorted.to_excel(writer, sheet_name='dim_times', index=False)
    ... dim_order_statuses_sorted.to_excel(writer, sheet_name='dim_order_statuses', index=False)
    ... dim_types_sorted.to_excel(writer, sheet_name='dim_types', index=False)
    ... dim_shipping_modes_sorted.to_excel(writer, sheet_name='dim_shipping_modes', index=False)
    ... dim_delivery_statuses_sorted.to_excel(writer, sheet_name='dim_delivery_statuses', index=False)
    ... fact_order_items.to_excel(writer, sheet_name='fact_order_items', index=False)

print(f"DataFrames saved to '{file_path}' successfully.")
```

DataFrames saved to 'Preprocessed_DataCoSupplyChainDataset.xlsx' successfully.

1.3. Danh sách thuộc tính được phân tích

STT	Tên thuộc tính	Tên đầy đủ	Ý nghĩa	KDL
1	Type	Loại	Loại giao dịch đã thực hiện	varchar
2	Days for shipping (real)	Ngày giao hàng (thực tế)	Số ngày giao hàng thực tế của sản phẩm được mua	number
3	Days for shipment (scheduled)	Ngày giao hàng (dự kiến)	Số ngày dự kiến giao hàng của sản phẩm được mua	number
4	Benefit per order	Lợi nhuận trên mỗi đơn hàng	Lợi nhuận trên mỗi đơn hàng được đặt	number
5	Sales per customer	Doanh số trên mỗi khách hàng	Tổng doanh số trên mỗi khách hàng đã thực hiện	number

7	Category Name	Tên danh mục	Mô tả về danh mục sản phẩm	varchar
8	Customer City	Thành phố của khách hàng	Thành phố nơi khách hàng thực hiện mua hàng	varchar
9	Customer Country	Quốc gia của khách hàng	Quốc gia nơi khách hàng thực hiện mua hàng	varchar
10	Customer Email	Email của khách hàng	Email của khách hàng	varchar
11	Customer Fname	Họ của khách hàng	Tên của khách hàng	varchar
12	Customer Id	ID của khách hàng	ID của khách hàng	number
13	Customer Lname	Tên của khách hàng	Họ của khách hàng	varchar
14	Customer Password	Mật khẩu của khách hàng	Khóa khách hàng được che giấu	varchar
15	Customer Segment	Phân khúc của khách hàng	Các loại khách hàng: Người tiêu dùng, Doanh nghiệp, Văn phòng tại nhà	varchar
16	Customer State	Tỉnh/thành phố của khách hàng	Tỉnh/thành phố mà cửa hàng đăng ký mua hàng thuộc về	varchar
17	Customer Street	Đường phố của khách hàng	Đường phố mà cửa hàng đăng ký mua hàng thuộc về	varchar

18	Market	Thị trường	Thị trường mà đơn hàng được giao: Châu Phi, Châu Âu, LATAM, Thái Bình Dương, USCA	varchar
18	Order City	Thành phố đích của đơn hàng	Thành phố đích của đơn hàng	varchar
19	Order Country	Quốc gia đích của đơn hàng	Quốc gia đích của đơn hàng	varchar
20	Order Customer Id	ID đơn hàng của khách hàng	Mã đơn hàng của khách hàng	number
21	order date (DateOrders)	Ngày đặt hàng	Ngày mà đơn hàng được đặt	date
22	Order Id	ID đơn hàng	Mã đơn hàng	number
23	Order Item Cardprod Id	ID sản phẩm trong đơn hàng	Mã sản phẩm được tạo ra thông qua đầu đọc RFID	number
24	Order Item Discount	Giảm giá mặt hàng đơn hàng	Giá trị giảm giá mặt hàng đơn hàng	number
25	Order Item Discount Rate	Tỷ lệ giảm giá mặt hàng	Phần trăm giảm giá mặt hàng đơn hàng	number
26	Order Item Id	ID mặt hàng trong đơn hàng	Mã mặt hàng của đơn hàng	number
27	Order Item Product Price	Giá sản phẩm mặt hàng đơn hàng	Giá của sản phẩm không có giảm giá	number
28	Order Item Profit Ratio	Tỷ lệ lợi nhuận mặt hàng đơn hàng	Tỷ lệ lợi nhuận mặt hàng đơn hàng	number
29	Order Item Quantity	Số lượng sản phẩm trong đơn hàng	Số lượng sản phẩm trên mỗi đơn hàng	number
30	Sales	Doanh số	Giá trị trong doanh số	number
31	Order Item Total	Tổng số mặt hàng trong đơn hàng	Tổng số tiền trên mỗi đơn hàng	number
32	Order Profit Per Order	Lợi nhuận trên mỗi đơn hàng	Lợi nhuận trên mỗi đơn hàng	number

33	Order Region	Khu vực đơn hàng	Khu vực trên thế giới mà đơn hàng được giao: Đông Nam Á, Nam Á, Đại Dương, Đông Á, Tây Á, Tây Hoa Kỳ, Trung tâm Hoa Kỳ, Tây Phi, Trung Phi, Bắc Phi, Tây Âu, Bắc Âu, Caribbean, Nam Mỹ, Đông Phi, Nam Âu, Đông Hoa Kỳ, Canada, Nam Phi, Trung Á, Châu Âu, Trung Mỹ, Đông Âu, Nam của Hoa Kỳ	varchar
34	Order State	Trạng thái đơn hàng	Trạng thái của đơn hàng: HOÀN CHỈNH, ĐANG CHỜ, ĐÓNG, ĐANG CHỜ THANH TOÁN, ĐÃ HỦY, ĐANG XỬ LÝ, NỮA, ĐANG GẶP NGUY CƠ LÙA ĐẢO, ĐANG ĐƯỢC GIỮ, XEM XÉT THANH TOÁN	varchar
35	Product Card Id	ID sản phẩm	Mã sản phẩm	number
36	Product Category Id	ID danh mục sản phẩm	Mã danh mục sản phẩm	number
37	Product Description	Mô tả sản phẩm	Mô tả về sản phẩm	varchar
38	Product Image	Hình ảnh sản phẩm	Liên kết để xem và mua sản phẩm	varchar
39	Product Name	Tên sản phẩm	Tên sản phẩm	varchar
40	Product Price	Giá sản phẩm	Giá sản phẩm	number
41	Product Status	Trạng thái sản phẩm	Trạng thái của hàng tồn kho: Nếu là 1 không có sẵn, 0 sản phẩm có sẵn	varchar
42	Customer Zipcode	Mã Zip của khách hàng	Mã Zip của khách hàng	varchar

43	Latitude	Vĩ độ	Vĩ độ tương ứng với vị trí của cửa hàng	number
44	Longitude	Kinh độ	Kinh độ tương ứng với vị trí của cửa hàng	number
45	Shipping date (DateOrders)	Ngày và giờ giao hàng	Ngày và giờ chính xác của việc giao hàng	datetime
46	Shipping Mode	Chế độ giao hàng	Các chế độ giao hàng được trình bày: Lớp tiêu chuẩn, Lớp đầu tiên, Lớp thứ hai, Cùng ngày	varchar

1.4. Mô tả chi tiết thuộc tính

1.1.1. Category

- Dictionary thuộc tính: Category.
- Thuộc tính Category chứa các danh mục ngành hàng.

Category		
STT	Tên	Ý nghĩa
1	Soccer	Bóng đá
2	Baseball & Softball	Bóng chày & Bóng mềm
3	Basketball	Bóng rổ
4	Lacrosse	Lacrosse
5	Tennis & Racquet	Tennis & Vợt
6	Hockey	Khúc côn cầu
7	Cardio Equipment	Thiết bị Cardio
8	Strength Training	Tập luyện sức mạnh
9	Fitness Accessories	Phụ kiện tập luyện sức khỏe
10	Boxing & MMA	Quyền Anh & MMA
11	Electronics	Điện tử
12	As Seen on TV!	Nhìn thấy trên TV!
13	Cleats	Giày đinh
14	Men's Footwear	Giày nam

15	Women's Apparel	Quần áo nữ
16	Girls' Apparel	Quần áo cho bé gái
17	Shop By Sport	Mua theo môn thể thao
18	Men's Golf Clubs	Gậy golf nam
19	Women's Golf Clubs	Gậy golf nữ
20	Golf Apparel	Quần áo golf
21	Golf Shoes	Giày golf
22	Golf Bags & Carts	Túi & Xe đẩy golf
23	Golf Gloves	Găng tay golf
24	Golf Balls	Bóng golf
25	Electronics	Điện tử
26	Kids' Golf Clubs	Gậy golf cho trẻ em
27	Accessories	Phụ kiện
28	Trade-In	Đổi trả
29	Camping & Hiking	Cắm trại & Leo núi
30	Hunting & Shooting	Săn bắn
31	Fishing	Câu cá
32	Indoor/Outdoor Games	Trò chơi trong nhà/ngoài trời
33	Water Sports	Thể thao dưới nước
34	Books	Sách
35	Baby	Em bé
36	CDs	Đĩa CD
37	CDs	Đĩa CD
38	Cameras	Máy ảnh
39	Children's Clothing	Quần áo trẻ em
40	Computers	Máy tính
41	Consumer Electronics	Điện tử tiêu dùng

42	Crafts	Thủ công mỹ nghệ
43	DVDs	Đĩa DVD
44	Garden	Vườn
45	Health and Beauty	Sức khỏe và Làm đẹp
46	Men's Clothing	Quần áo nam
47	Music	Âm nhạc
48	Pet Supplies	Đồ dùng cho thú cưng
49	Sporting Goods	Hàng thể thao
50	Toys	Đồ chơi
51	Video Games	Trò chơi điện tử
52	Women's Clothing	Quần áo nữ

1.1.1.1. 1.4.2. Type

- Dictionary thuộc tính: Type.
- Thuộc tính Type chứa các loại hình thanh toán mặt hàng.

Type		
STT	Tên	Ý nghĩa
1	CASH	Tiền mặt
2	DEBIT	Nợ
3	PAYMENT	Thanh toán
4	TRANSFER	Chuyển khoản

1.1.1.2. 1.4.3. Shipping Mode

- Dictionary thuộc tính: Shipping Mode.
- Thuộc tính Shipping Mode cho biết loại hình của việc giao hàng.

Shipping Mode		
STT	Tên	Ý nghĩa
1	First Class	Loại ưu tiên I

2	Same Day	Trong ngày
3	Second Class	Loại ưu tiên II
4	Standard Class	Tiêu chuẩn

1.1.1.3.**1.4.4. Delivery Status**

- Dictionary thuộc tính: Delivery Status.
- Thuộc tính Delivery Status cho biết trạng thái giao hàng.

Delivery Status		
STT	Tên	Ý nghĩa
1	Advance shipping	Vận chuyển nâng cao
2	Late delivery	Vận chuyển chậm
3	Shipping canceled	Hủy vận chuyển
4	Shipping on time	Vận chuyển đúng giờ

1.5. 1.5. Sơ đồ dữ liệu xây dựng**1.5.1. Lược đồ hình sao (Star schema)**

Lược đồ sao là loại lược đồ Kho dữ liệu đơn giản nhất. Nó được gọi là lược đồ sao vì cấu trúc của nó giống như một ngôi sao. Trong lược đồ hình sao, tâm của ngôi sao có thể có một bảng sự kiện (fact) và số lượng bảng chiều (dimension) được liên kết. Nó còn được gọi là Star Join Schema và được tối ưu hóa để truy vấn các tập dữ liệu lớn.

Ưu điểm:

- Các truy vấn đơn giản hơn - logic nối lược đồ sao thường đơn giản hơn logic nối được yêu cầu để truy xuất dữ liệu từ lược đồ giao dịch được chuẩn hóa cao.
- Tăng hiệu suất truy vấn - các lược đồ sao có thể cung cấp các cải tiến hiệu suất cho các ứng dụng báo cáo chỉ đọc khi so sánh với các lược đồ được chuẩn hóa cao.
- Tổng hợp nhanh - các truy vấn đơn giản hơn đối với lược đồ sao có thể dẫn đến hiệu suất được cải thiện cho các hoạt động tổng hợp.

Nhược điểm:

- Tính toàn vẹn dữ liệu không được thực thi tốt vì nó không ở trạng thái chuẩn hóa cao. [1]
- Tốn nhiều dung lượng lưu trữ do dư thừa dữ liệu.
- Lược đồ hình sao không linh hoạt về mặt nhu cầu phân tích như mô hình dữ liệu chuẩn hóa.

1.5.2. Lược đồ hình bông tuyết (Snowflake schema)

Lược đồ bông tuyết là một phần mở rộng của Lược đồ hình sao và nó thêm các bảng bổ sung. Lược đồ bông tuyết được coi là một trường hợp đặc biệt của lược đồ hình sao. Nó được gọi là bông tuyết vì sơ đồ của nó giống như một Bông tuyết.

Ưu điểm:

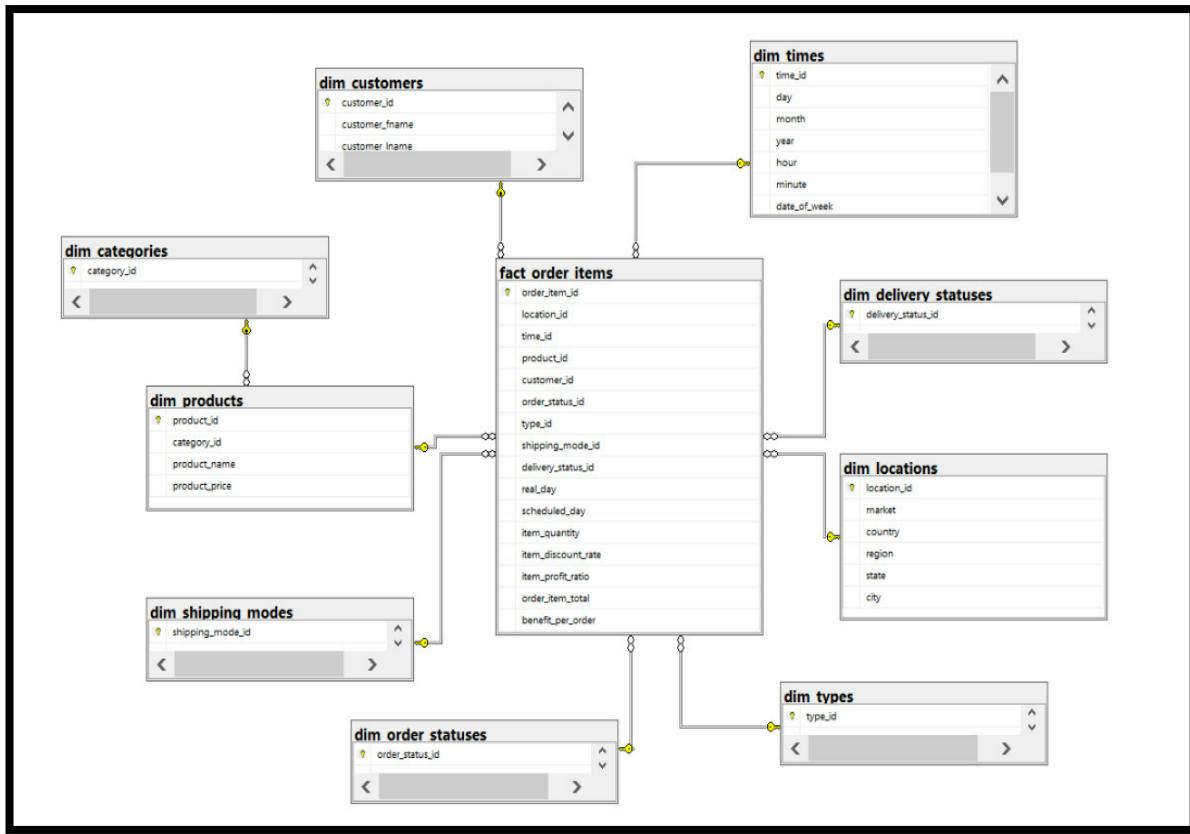
- Sử dụng cấu trúc dữ liệu chuẩn hóa.
- Đơn giản hóa các thuộc tính dẫn đến sự tiết kiệm không gian lưu trữ dữ liệu.
- Một số chiều được phân cấp để thể hiện rõ ràng định chuẩn của bảng chiều.
- Một số công cụ mô hình hóa cơ sở dữ liệu đa chiều (OLAP) được tối ưu hóa cho các lược đồ bông tuyết.

Nhược điểm:

- Mức chuẩn hóa thuộc tính yêu cầu nhiều bảng hơn và các phép nối phức tạp hơn, làm chậm hiệu suất truy vấn so với lược đồ hình sao.

1.6.1.6. Xây dựng kho dữ liệu

1.6.1. Sơ đồ hình sao minh họa



1.6.2. Mô tả chi tiết các bảng dữ liệu

1.1.1.4.

1.6.2.1 Bảng Fact

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Order Item Id	int	<u>PK</u>	Id của bảng Fact
2	Location Id	int	FK	Id dim Location
3	Time Id	int	FK	Id dim Time
4	Product Card Id	int	FK	Id dim Product Card
5	Customer Id	int	FK	Id dim Customer
6	Order Status Id	int	FK	Id dim Order Status
7	Type Id	int	FK	Id dim Type
8	Shipping Mode Id	int	FK	Id dim Shipping Mode
9	Delivery Status Id	int	FK	Id dim Delivery Status

10	Days for shipping (real)	int		Số ngày giao hàng thực tế
11	Days for shipment (scheduled)	int		Ngày giao hàng dự kiến
12	Order Item Quantity	int		Số lượng hàng
13	Order Item Discount Rate	int		Tỉ lệ giảm giá
14	Order Item Profit Ratio	int		Tỉ lệ lợi nhuận

1.1.1.5.**1.6.2.2. Dim_Customers**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Customer Id	Int	PK	Id
2	Customer Fname	Varchar		Tên của khách hàng
3	Customer Lname	Varchar		Họ của khách hàng

1.1.1.6.**1.6.2.3. Dim_Categories**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Category Id	Int	PK	Id
2	Category Name	Varchar		Tên thể loại

1.1.1.7.**1.6.2.4. Dim_Products**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Product Card Id	Int	PK	Id
2	Product Category Id	Int	FK	Khóa ngoại bảng Category
3	Product Name	Varchar		Tên sản phẩm
4	Product Price	Float		Giá sản phẩm

1.1.1.8.**1.6.2.5. Dim_Locations**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Location Id	Int	PK	Id
2	Market	Varchar		Thị trường tiêu thụ
3	Order City	Varchar		Tên thành phố của người mua hàng
4	Order Country	Varchar		Quốc gia của người đặt hàng
5	Order Region	Varchar		Vùng của người đặt hàng
6	Order State	Varchar		Bang của người đặt hàng

1.1.1.9.**1.6.2.6. Dim_Times**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Time Id	Int	PK	Id
2	Date (DateOrders)	Timestamp		Ngày giờ đặt hàng

1.1.1.10.**1.6.2.7. Dim_Order_Statuses**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Order Status Id	Int	PK	Id
2	Order Status	Enum		Tên loại trạng thái đặt hàng

1.1.1.11.**1.6.2.8. Dim_Type**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Type Id	Int	PK	Id
2	Type	Enum		Tên loại thanh toán

1.1.1.12.**1.6.2.9. Dim_Shipping_Modes**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Shipping Mode Id	Int	PK	Id
2	Shipping Mode	Enum		Tên loại hình giao hàng

1.1.1.13.**1.6.2.10. Dim_Delivery_Statuses**

STT	Tên thuộc tính	KDL	Ràng buộc	Ý nghĩa
1	Delivery Status Id	Int	PK	Id
2	Delivery Status	Enum		Tên loại hình vận chuyển

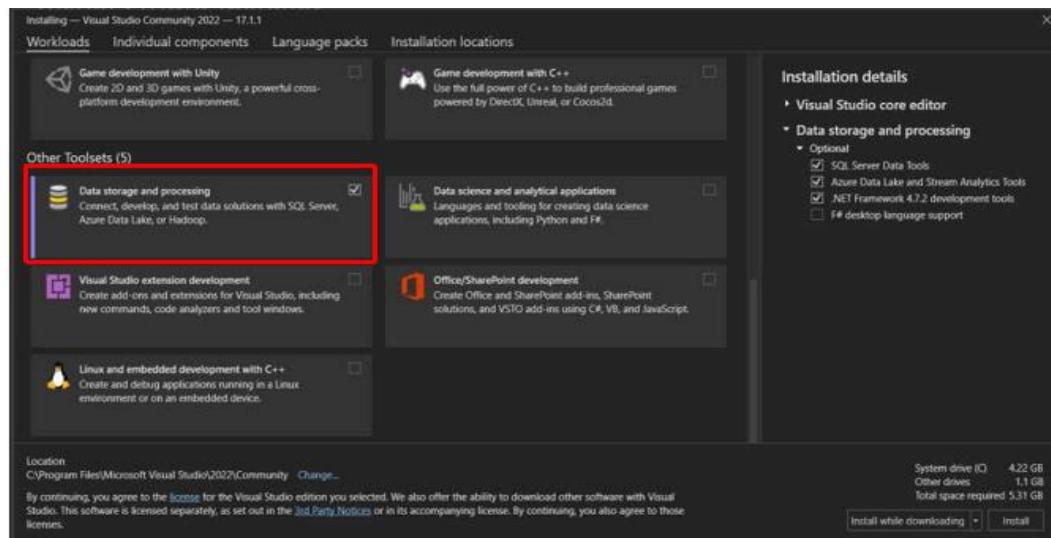
CHƯƠNG 2. QUÁ TRÌNH XÂY DỰNG KHO DỮ LIỆU (SSIS)

2.1. Chuẩn bị các công cụ

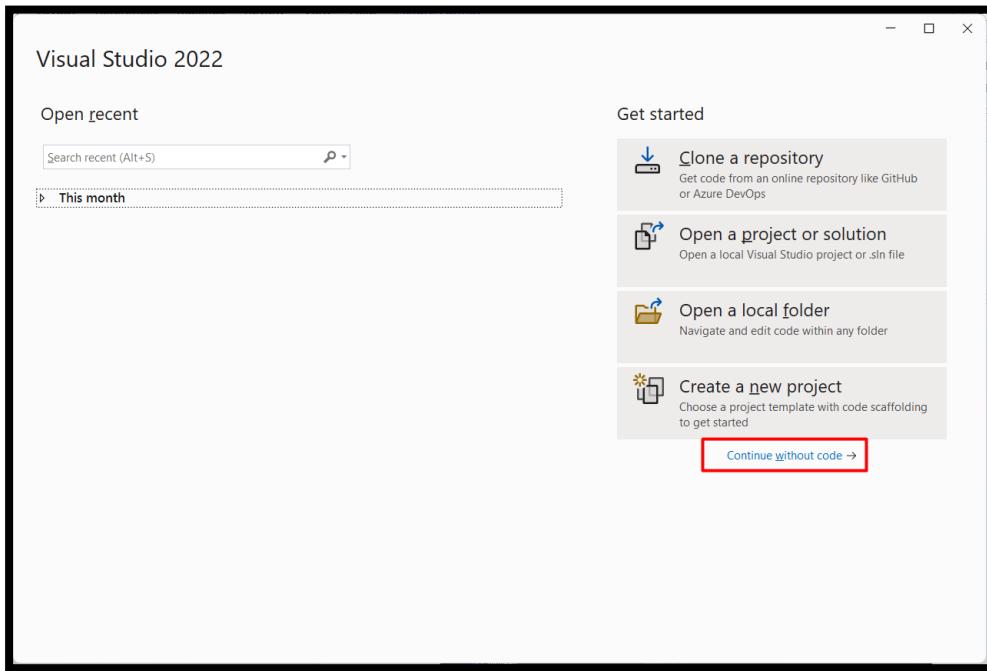
Để thực hiện được quá trình SSIS ta cần chuẩn bị và cài đặt các công cụ sau:

- Visual Studio Community 2022
- SQL Server Integration Services Project

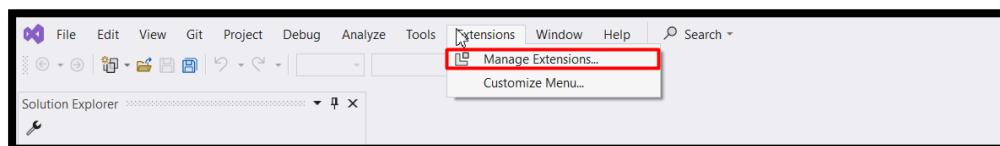
Bước 1: Tải [Visual Studio Community 2022](#) về máy. Trong lúc cài đặt, chọn mục “Data storage and processing” để cài đặt SQL Server Data Tools. Sau đó chọn Install để tiếp tục hành cài đặt



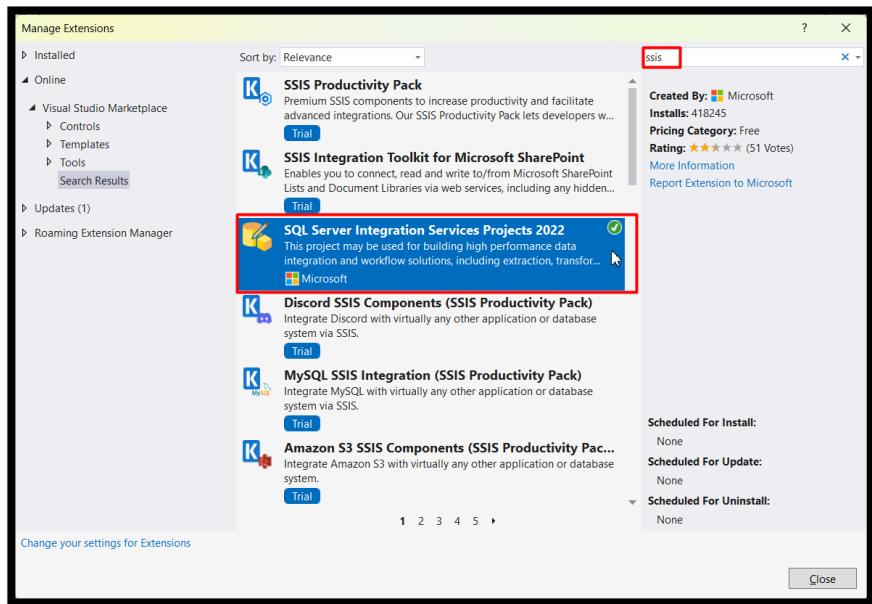
Bước 2: Mở Visual Studio 2019 và chọn “Continue without code”



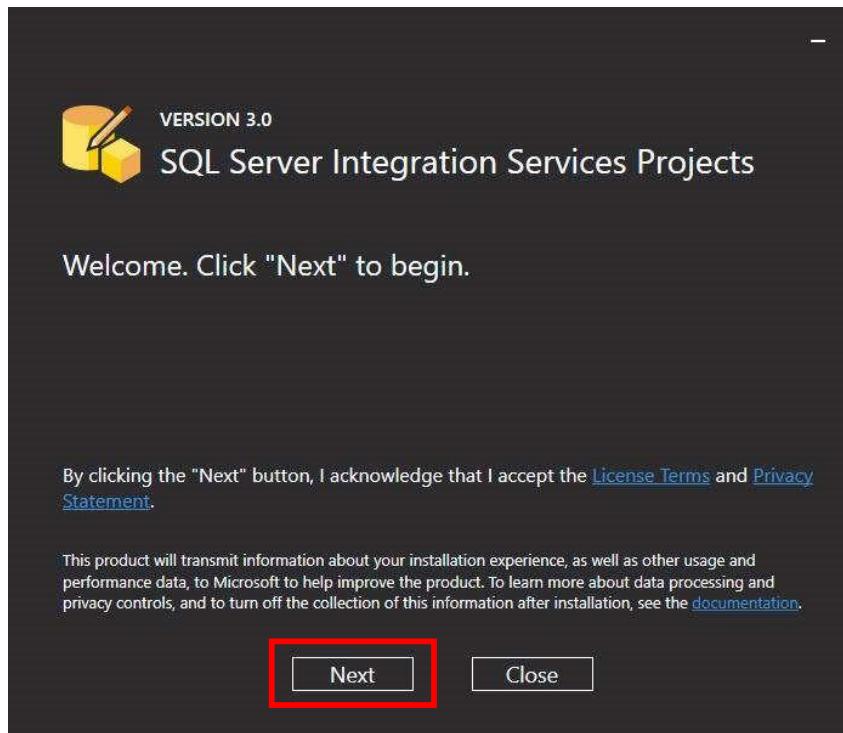
Bước 3: Trong giao diện chính, click chọn “Extensions” > “Manage Extensions”



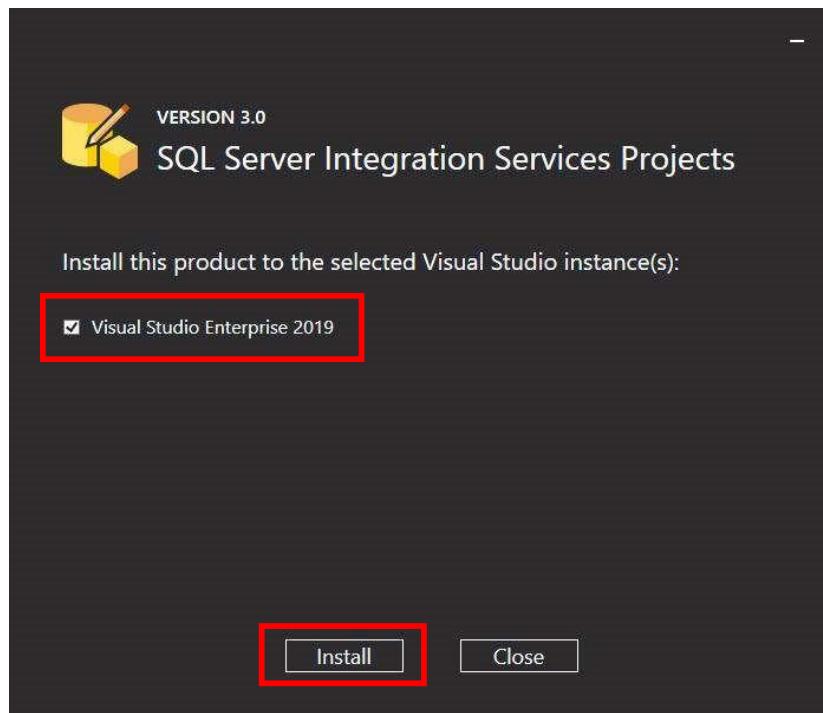
Bước 4: Tìm kiếm với từ khóa “ssis” và tải về công cụ SQL Server Integration Services Projects.



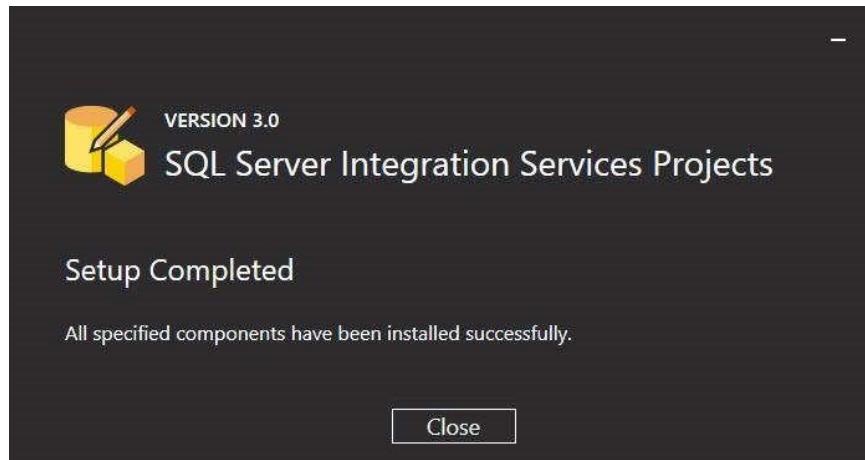
Bước 5: Mở file .exe vừa tải xuống. Chọn **Next** để tiếp tục



Bước 6: Tick chọn Visual Studio 2019 và chọn **Install**

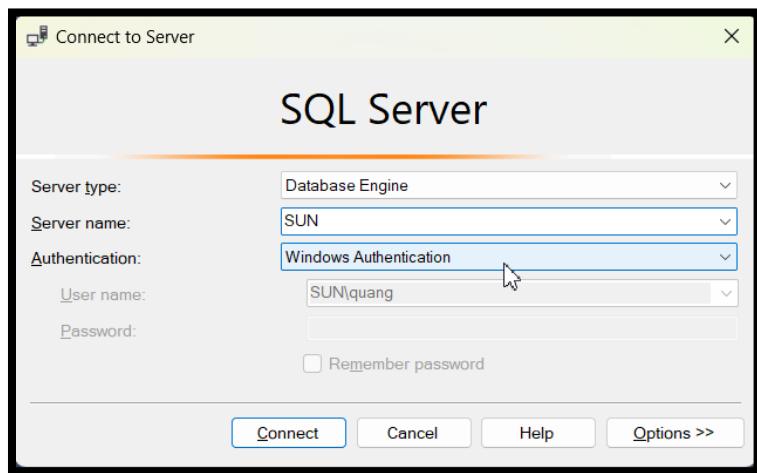


Sau khi cài đặt thành công sẽ nhận được thông báo:

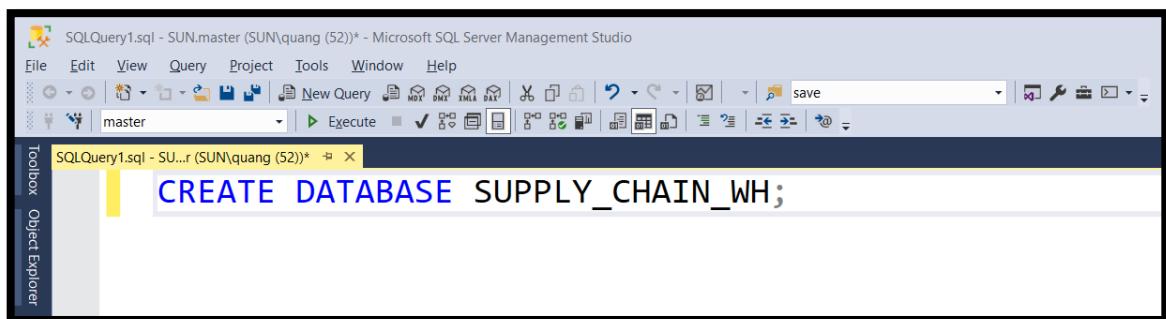


2.2. Chuẩn bị cơ sở dữ liệu

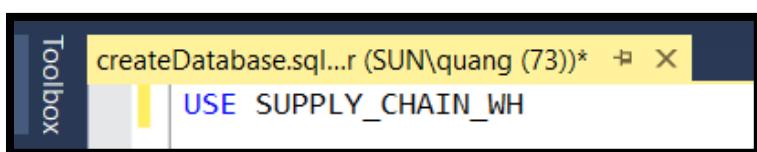
Bước 1: Mở SQL Server 2022 và kết nối với server bằng tài khoản user của windows (**Windows Authentication**).



Bước 2: Mở một console mới và khởi tạo một cơ sở dữ liệu có tên **SUPPLY_CHAIN_WH**, đây là nơi lưu các bảng Dim và bảng Fact cùng dữ liệu của các bảng đó.



Bước 3: Chọn cơ sở dữ liệu vừa tạo để tiến hành thao tác.



Bước 4: Xóa các database có sẵn nếu có tồn tại.

- *Bỏ các khóa ngoại đến các bảng khác nếu có.*

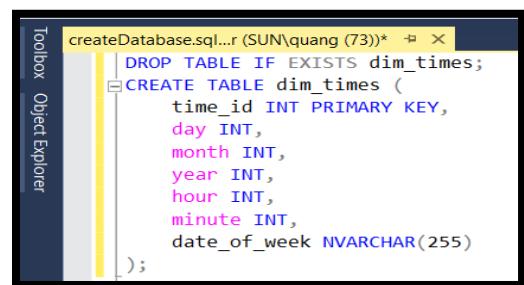
```
-- Drop constraints from fact_order_items table
ALTER TABLE fact_order_items DROP CONSTRAINT IF EXISTS FK_fact_order_items_location_id;
ALTER TABLE fact_order_items DROP CONSTRAINT IF EXISTS FK_fact_order_items_time_id;
ALTER TABLE fact_order_items DROP CONSTRAINT IF EXISTS FK_fact_order_items_product_id;
ALTER TABLE fact_order_items DROP CONSTRAINT IF EXISTS FK_fact_order_items_customer_id;
ALTER TABLE fact_order_items DROP CONSTRAINT IF EXISTS FK_fact_order_items_order_status_id;
ALTER TABLE fact_order_items DROP CONSTRAINT IF EXISTS FK_fact_order_items_shipping_mode_id;
ALTER TABLE fact_order_items DROP CONSTRAINT IF EXISTS FK_fact_order_items_type_id;
ALTER TABLE fact_order_items DROP CONSTRAINT IF EXISTS FK_fact_order_items_delivery_status_id;
```

- Xóa các bảng cần thiết.

```
-- Drop all tables
DROP TABLE IF EXISTS fact_order_items;
DROP TABLE IF EXISTS dim_delivery_statuses;
DROP TABLE IF EXISTS dim_shipping_modes;
DROP TABLE IF EXISTS dim_order_statuses;
DROP TABLE IF EXISTS dim_types;
DROP TABLE IF EXISTS dim_customers;
DROP TABLE IF EXISTS dim_products;
DROP TABLE IF EXISTS dim_locations;
DROP TABLE IF EXISTS dim_times;
```

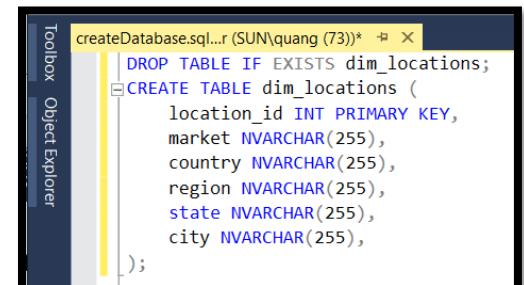
Bước 5: Lần lượt kiểm tra sự tồn tại và tạo các bảng dim.

- *dim_times*



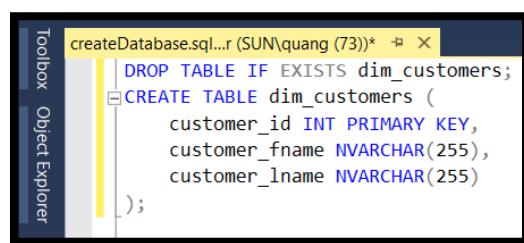
```
createDatabase.sql...r (SUN\quang (73))*
DROP TABLE IF EXISTS dim_times;
CREATE TABLE dim_times (
    time_id INT PRIMARY KEY,
    day INT,
    month INT,
    year INT,
    hour INT,
    minute INT,
    date_of_week NVARCHAR(255)
);
```

- *dim_locations*



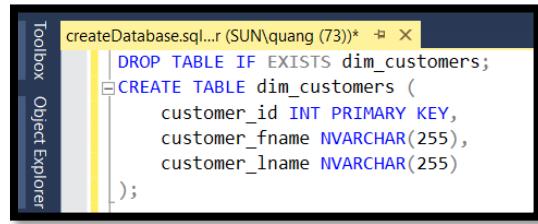
```
createDatabase.sql...r (SUN\quang (73))*
DROP TABLE IF EXISTS dim_locations;
CREATE TABLE dim_locations (
    location_id INT PRIMARY KEY,
    market NVARCHAR(255),
    country NVARCHAR(255),
    region NVARCHAR(255),
    state NVARCHAR(255),
    city NVARCHAR(255),
);
```

- *dim_customers*



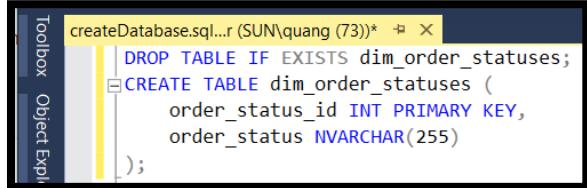
```
createDatabase.sql...r (SUN\quang (73))*
DROP TABLE IF EXISTS dim_customers;
CREATE TABLE dim_customers (
    customer_id INT PRIMARY KEY,
    customer_fname NVARCHAR(255),
    customer_lname NVARCHAR(255)
);
```

- *dim_types*



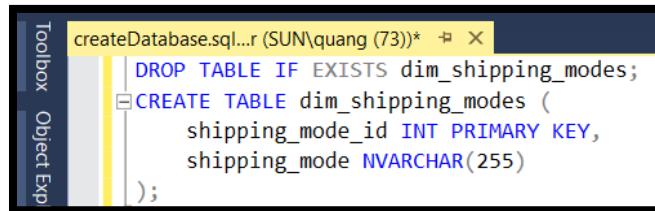
```
createDatabase.sql...r (SUN\quang (73))* ↵ X
DROP TABLE IF EXISTS dim_customers;
CREATE TABLE dim_customers (
    customer_id INT PRIMARY KEY,
    customer_fname NVARCHAR(255),
    customer_lname NVARCHAR(255)
);
```

- *dim_order_statuses*



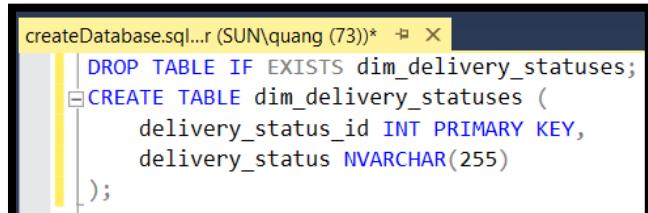
```
createDatabase.sql...r (SUN\quang (73))* ↵ X
DROP TABLE IF EXISTS dim_order_statuses;
CREATE TABLE dim_order_statuses (
    order_status_id INT PRIMARY KEY,
    order_status NVARCHAR(255)
);
```

- *dim_shipping_modes*



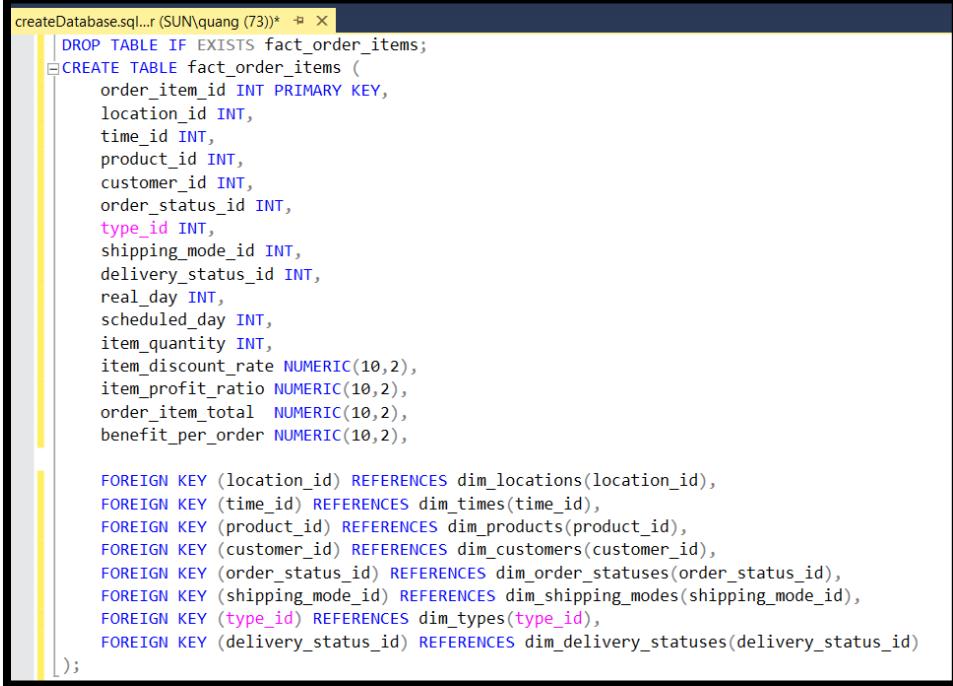
```
createDatabase.sql...r (SUN\quang (73))* ↵ X
DROP TABLE IF EXISTS dim_shipping_modes;
CREATE TABLE dim_shipping_modes (
    shipping_mode_id INT PRIMARY KEY,
    shipping_mode NVARCHAR(255)
);
```

- *dim_delivery_statuses*



```
createDatabase.sql...r (SUN\quang (73))* ↵ X
DROP TABLE IF EXISTS dim_delivery_statuses;
CREATE TABLE dim_delivery_statuses (
    delivery_status_id INT PRIMARY KEY,
    delivery_status NVARCHAR(255)
);
```

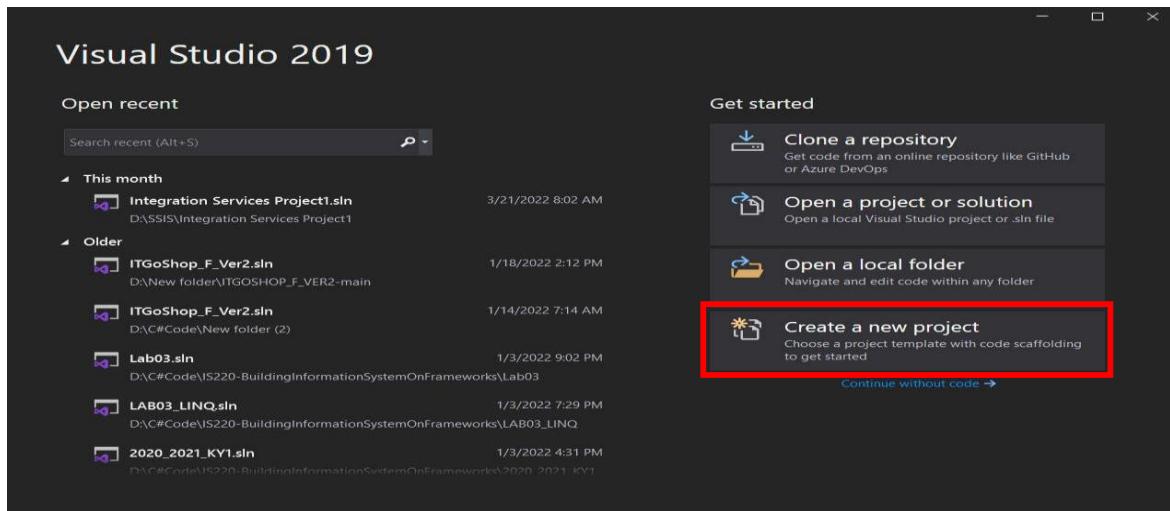
- *fact_order_items*



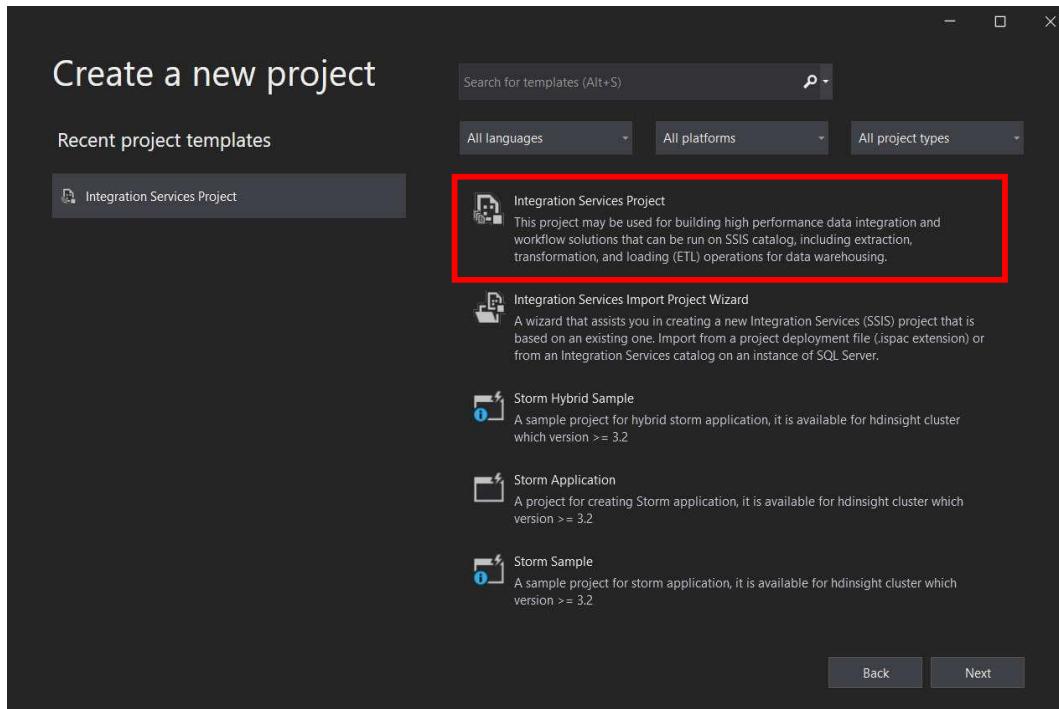
```
createDatabase.sql..r (SUN\quang (73))*
DROP TABLE IF EXISTS fact_order_items;
CREATE TABLE fact_order_items (
    order_item_id INT PRIMARY KEY,
    location_id INT,
    time_id INT,
    product_id INT,
    customer_id INT,
    order_status_id INT,
    type_id INT,
    shipping_mode_id INT,
    delivery_status_id INT,
    real_day INT,
    scheduled_day INT,
    item_quantity INT,
    item_discount_rate NUMERIC(10,2),
    item_profit_ratio NUMERIC(10,2),
    order_item_total NUMERIC(10,2),
    benefit_per_order NUMERIC(10,2),
    FOREIGN KEY (location_id) REFERENCES dim_locations(location_id),
    FOREIGN KEY (time_id) REFERENCES dim_times(time_id),
    FOREIGN KEY (product_id) REFERENCES dim_products(product_id),
    FOREIGN KEY (customer_id) REFERENCES dim_customers(customer_id),
    FOREIGN KEY (order_status_id) REFERENCES dim_order_statuses(order_status_id),
    FOREIGN KEY (shipping_mode_id) REFERENCES dim_shipping_modes(shipping_mode_id),
    FOREIGN KEY (type_id) REFERENCES dim_types(type_id),
    FOREIGN KEY (delivery_status_id) REFERENCES dim_delivery_statuses(delivery_status_id)
);
```

2.3. Tạo mới project SSIS

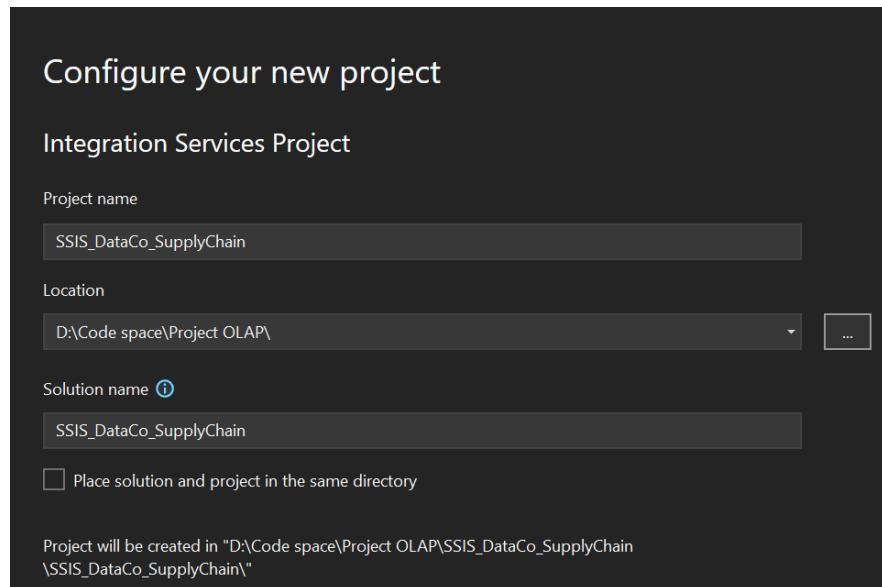
Bước 1: Mở Visual Studio 2022 và chọn “Create a new project”.



Bước 2: Chọn Integration Services Project và chọn Next



Bước 3: Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn **Create**.

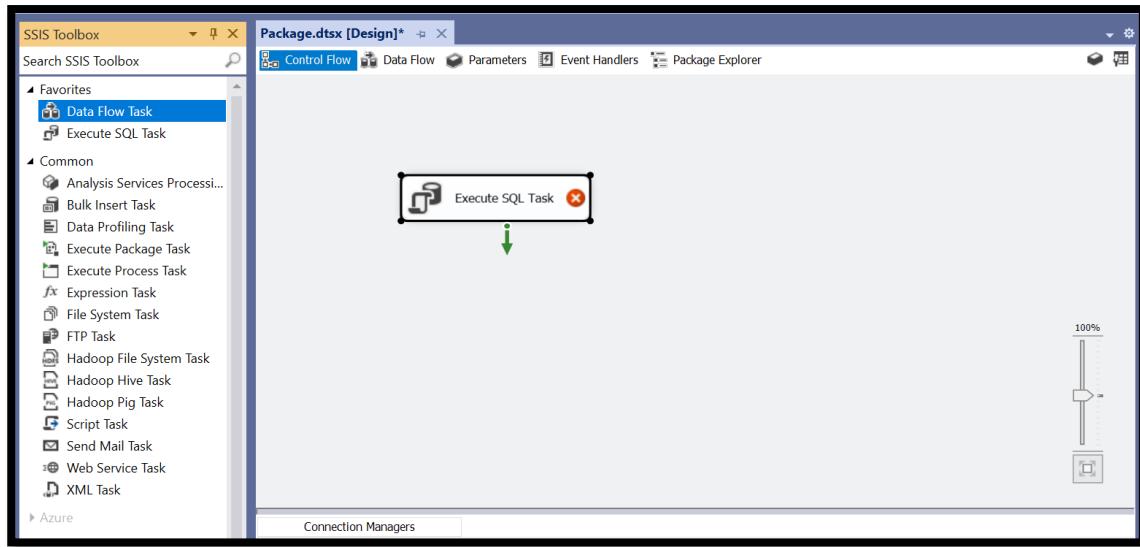


Tạo Control Flow cho quá trình SSIS

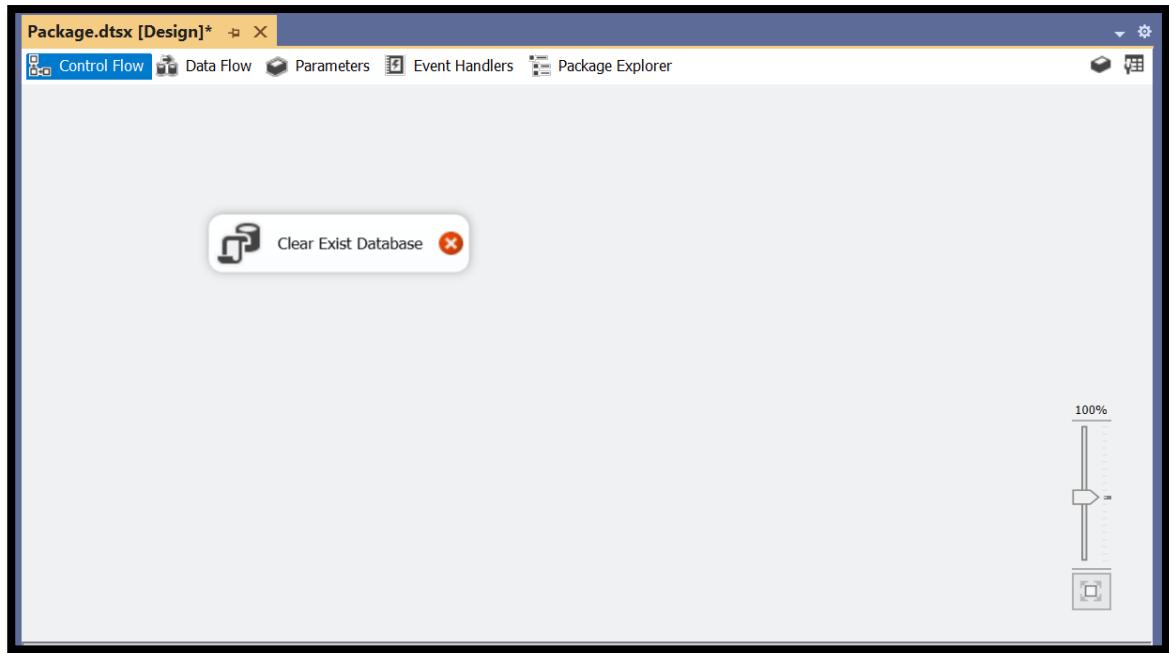
Ý tưởng của Control Flow sẽ đi từ Drop các database cũ nếu có -> tạo database mới -> đổ dữ liệu vào các bảng Dim nhằm tránh xung đột foreigne key sang đổ dữ liệu vào bảng fact.

Tạo task để dọn database cũ nếu có

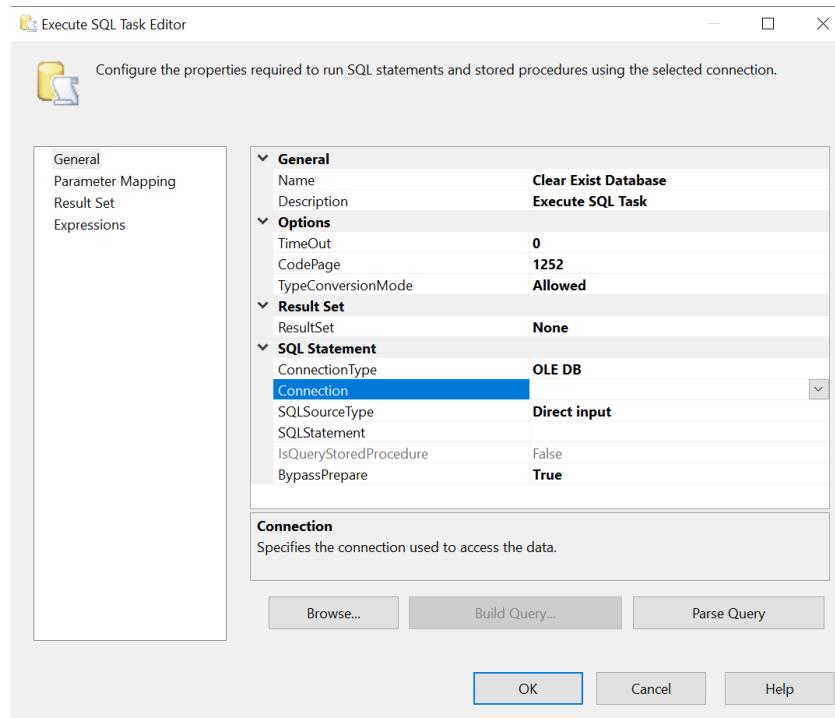
Bước 1: Tại SSIS toolbox kéo thả **Execute SQL Task** vào Control Flow panel.



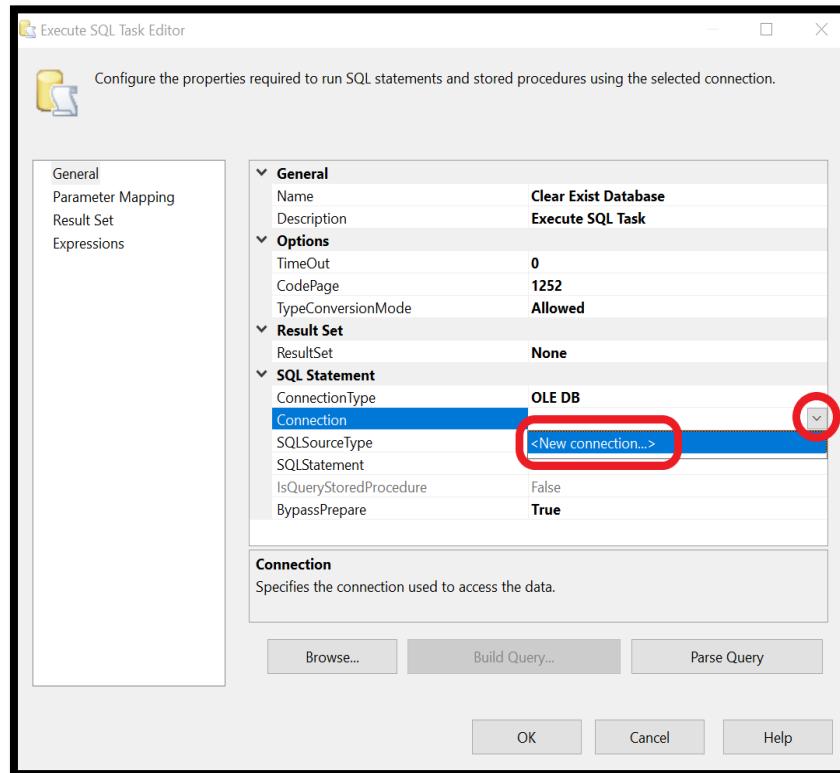
Bước 2: Click vào và chọn f2 để đổi tên task thành “Clear Exist Database”.



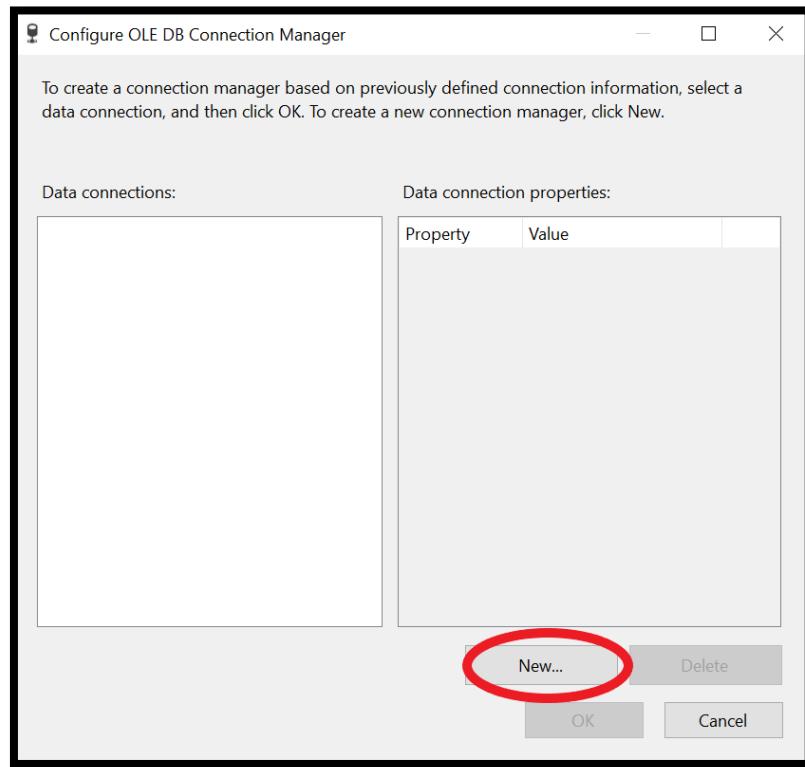
Bước 3: Double click vào “Clear Exist Database” để hiển thị dialog box



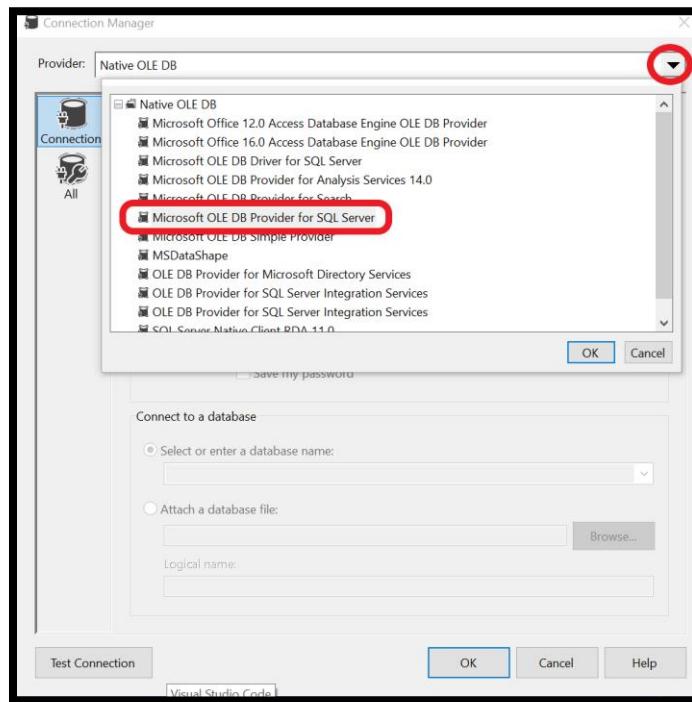
Bước 4: Nhấn vào nút mở rộng, chọn **New connection...** (Có thể chọn 1 kết nối khác có sẵn nếu có)



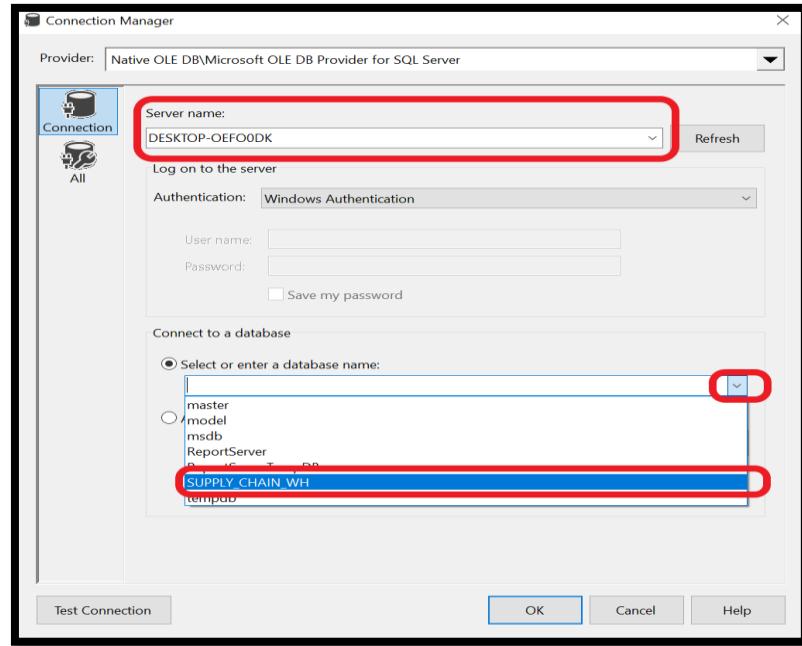
Bước 5: Chọn **New** (Có thể chọn 1 kết nối khác có sẵn nếu có).



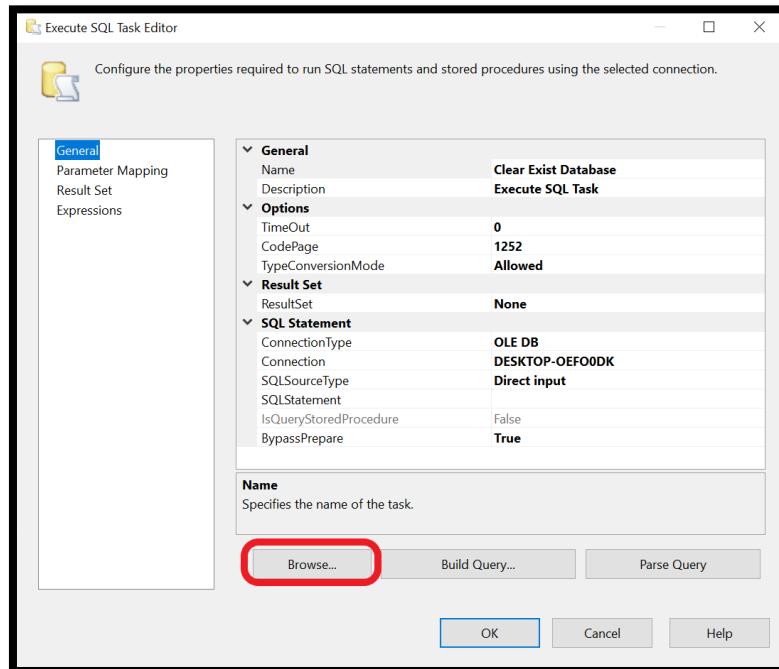
Bước 6: Tại Provider, mở rộng và chọn Microsoft OLE DB Provider for SQL Server, sau đó chọn OK



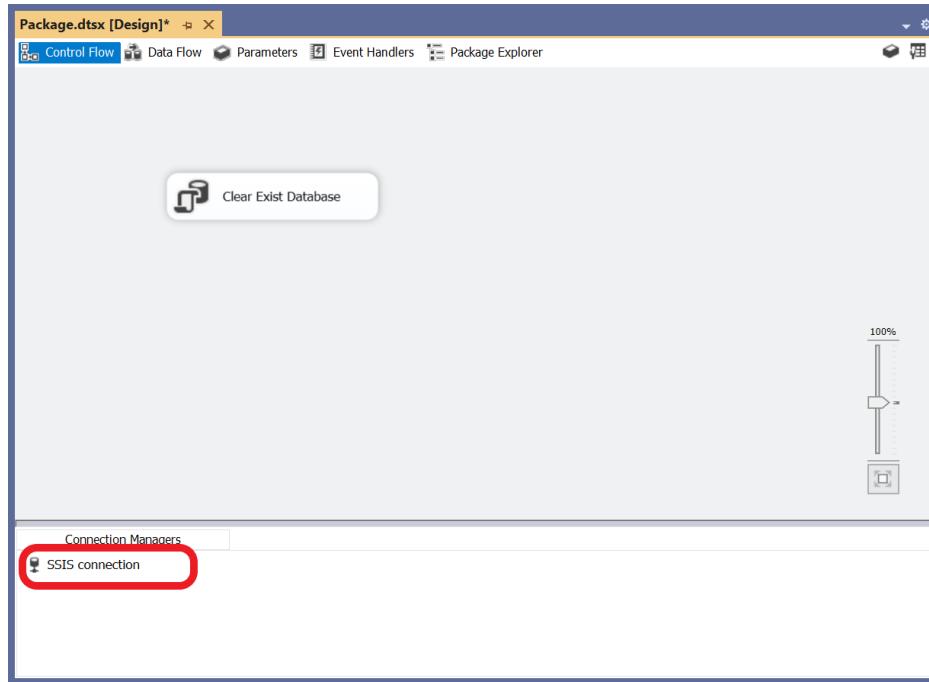
Bước 6: Dùng bàn phím nhập vào **Server name** của máy, tại phần connect to a database, chọn tên database đã tạo tại giai đoạn chuẩn bị. Sau đó chọn **OK** để đóng Connection Manager panel và **OK** lần nữa để kết nối với database.



Bước 7: Chọn “Browser...” để thêm sql script, chọn script để dọn sạch database đã chuẩn bị trước, **OK** để kết thúc.

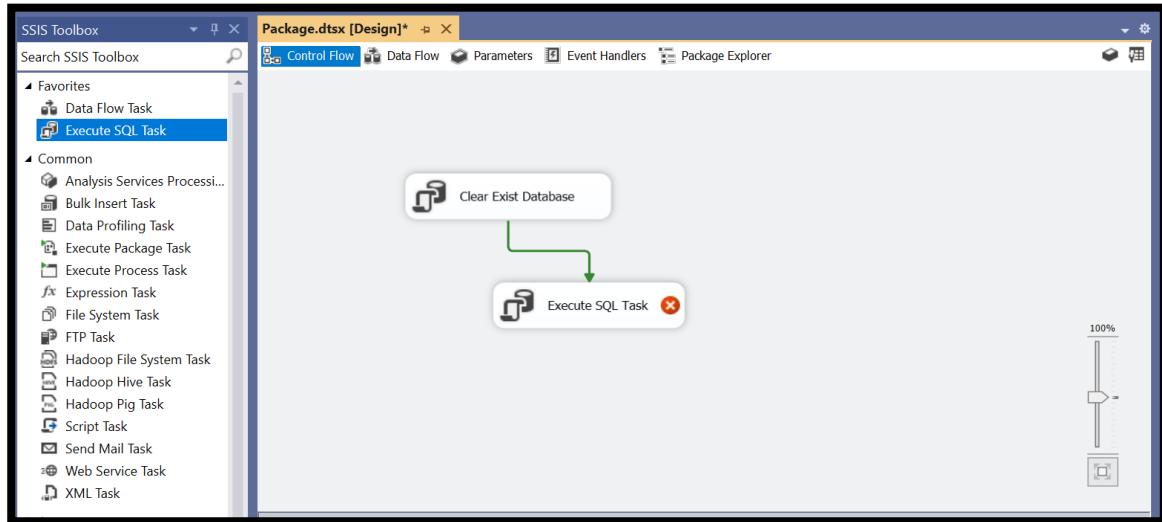


- Lưu ý: có thẻ Click và connection và f để đổi tên connection nhằm thuận tiện cho quá trình SSIS

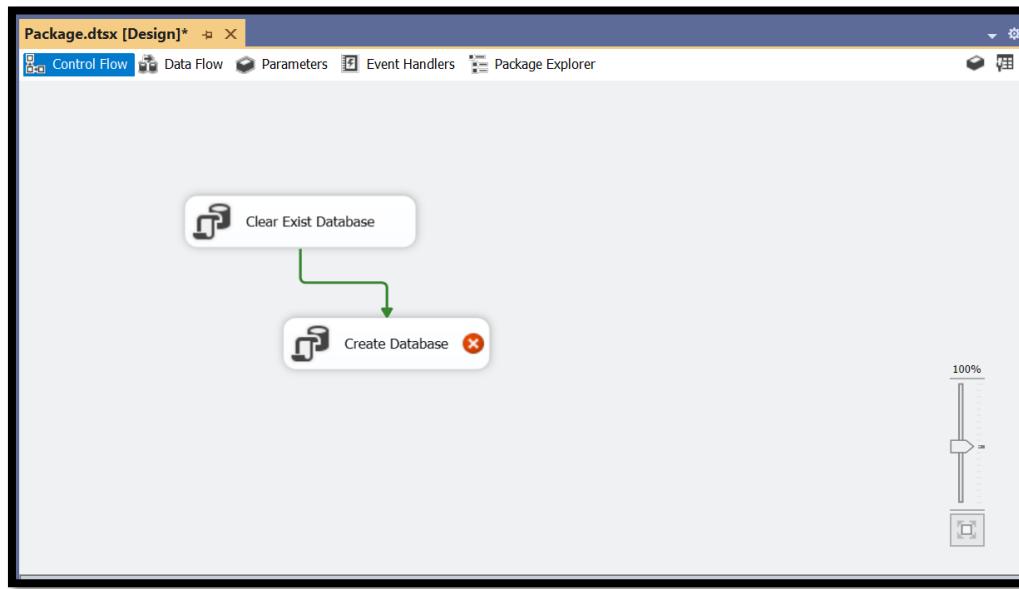


Tạo task để tạo mới database

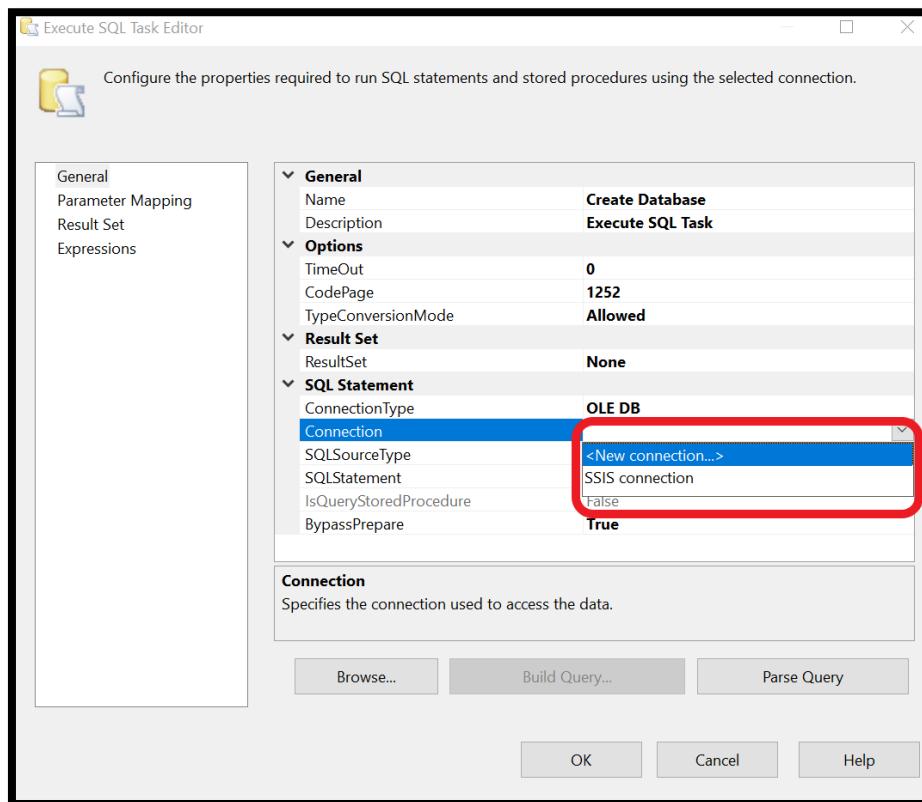
Bước 1: Tại SSIS toolbox kéo thả **Execute SQL Task** vào Control Flow panel. Sau đó nối “Clear Exist Database” vào **Execute SQL Task** vừa tạo.



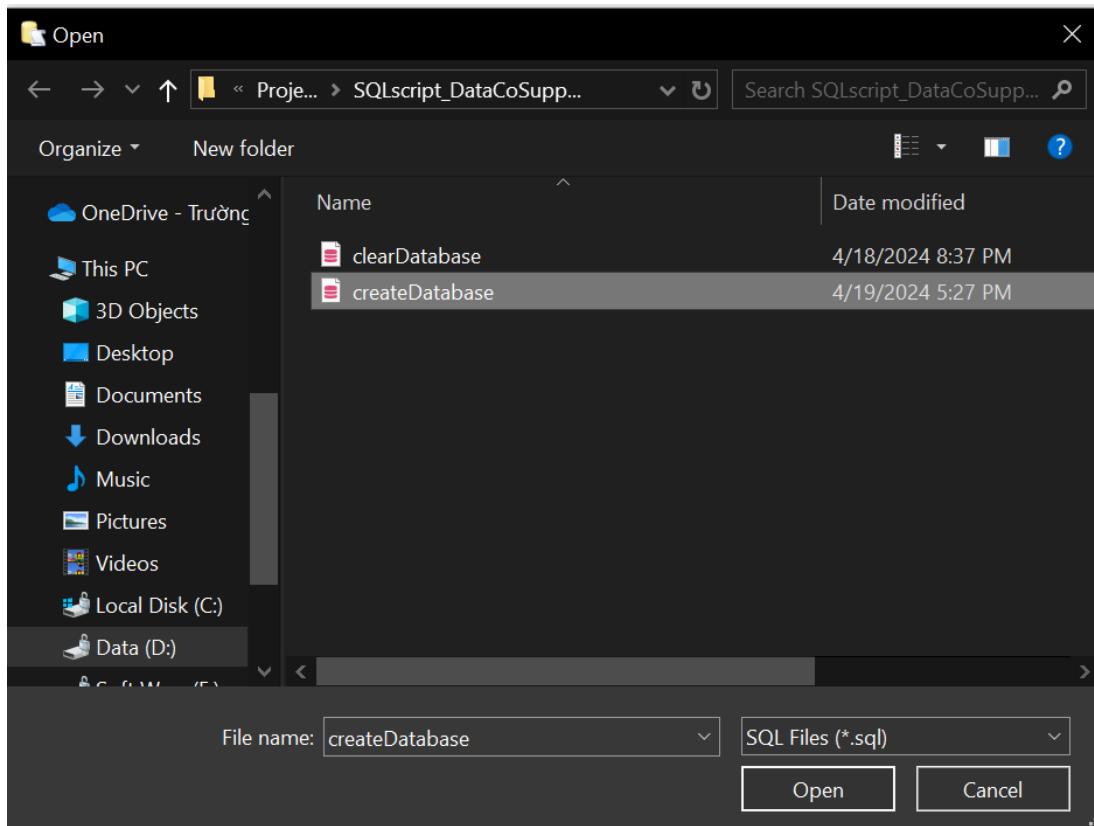
Bước 2: Click vào và chọn f2 để đổi tên task thành “Create Database”.



Bước 3: D-click vào “Create Database” để hiển thị dialog box, do đã tạo connection ở bước trước nên chỉ cần chọn lại.

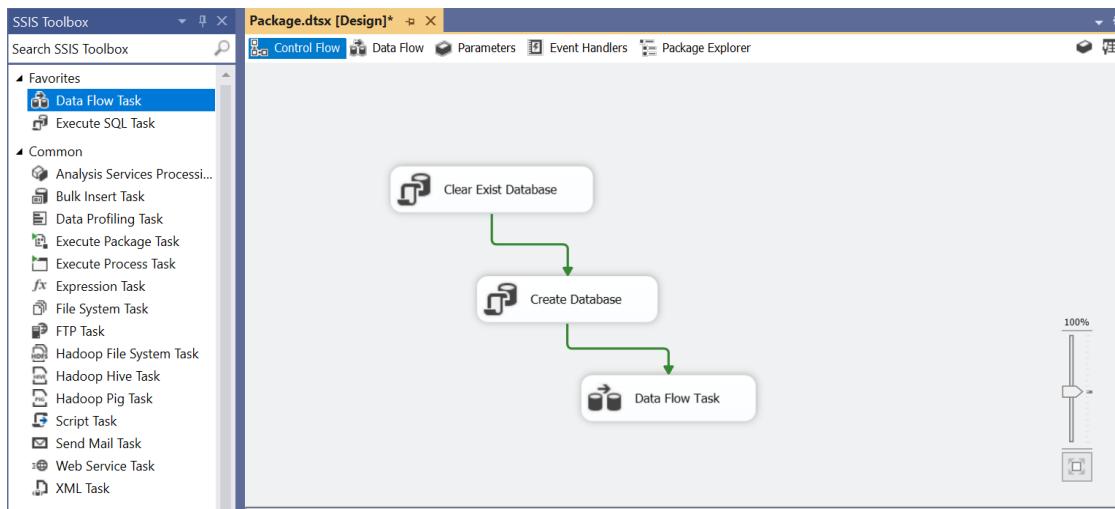


Bước 4: Chọn Browser... để thêm sql script, chọn script để tạo mới database đã chuẩn bị trước, chọn **Open** và **OK** để kết thúc.

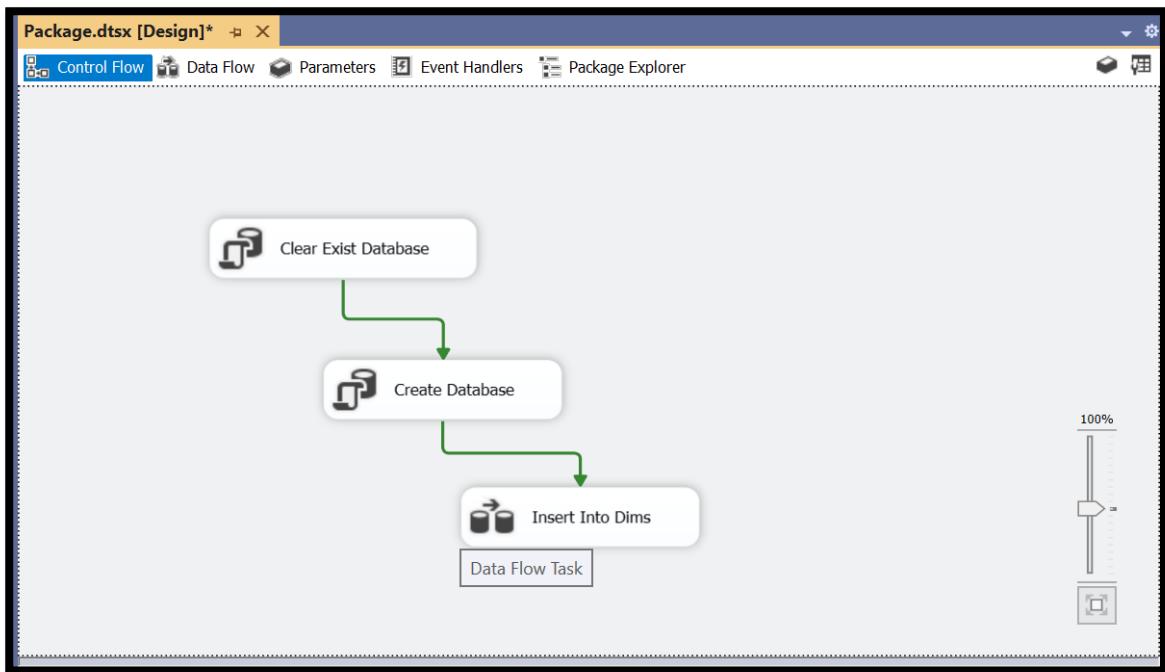


Tạo Data Flow Task để đổ data vào trong các dims

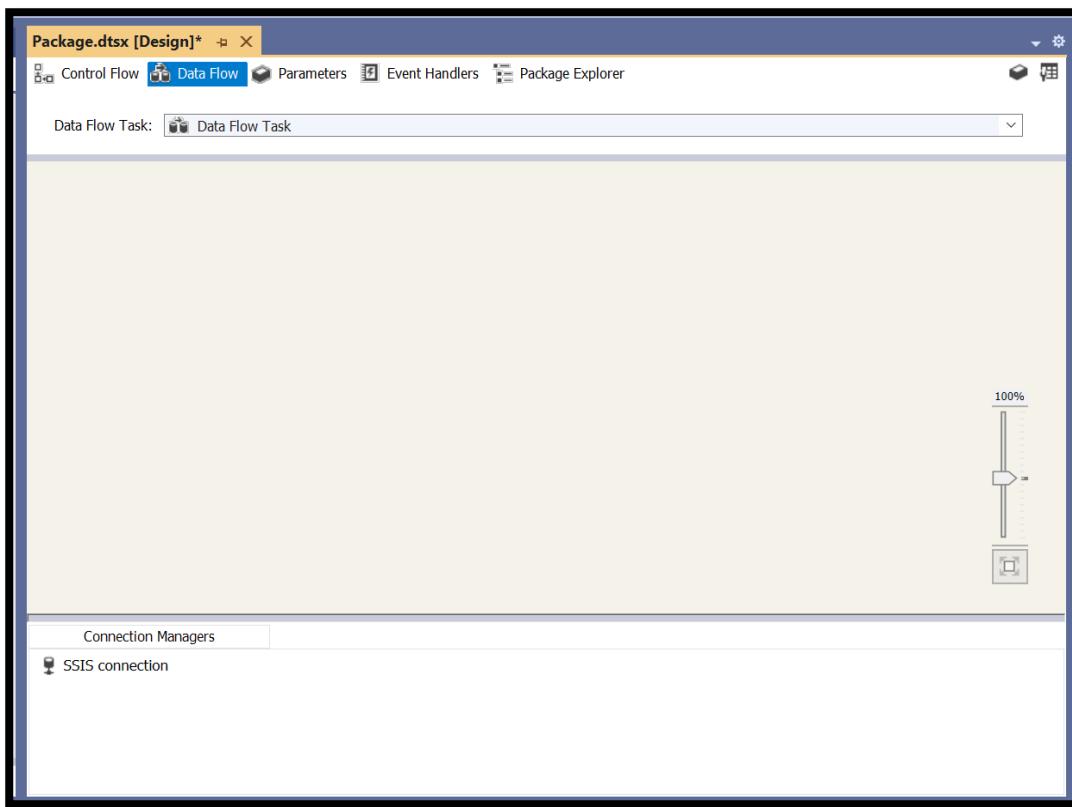
Bước 1: Tại SSIS toolbox kéo thả **Data Flow Task** vào Control Flow panel. Sau đó nối “Create Database” vào Execute SQL Task vừa tạo.



Bước 2: Click vào và chọn f2 để đổi tên task thành “Insert Into Dims”.

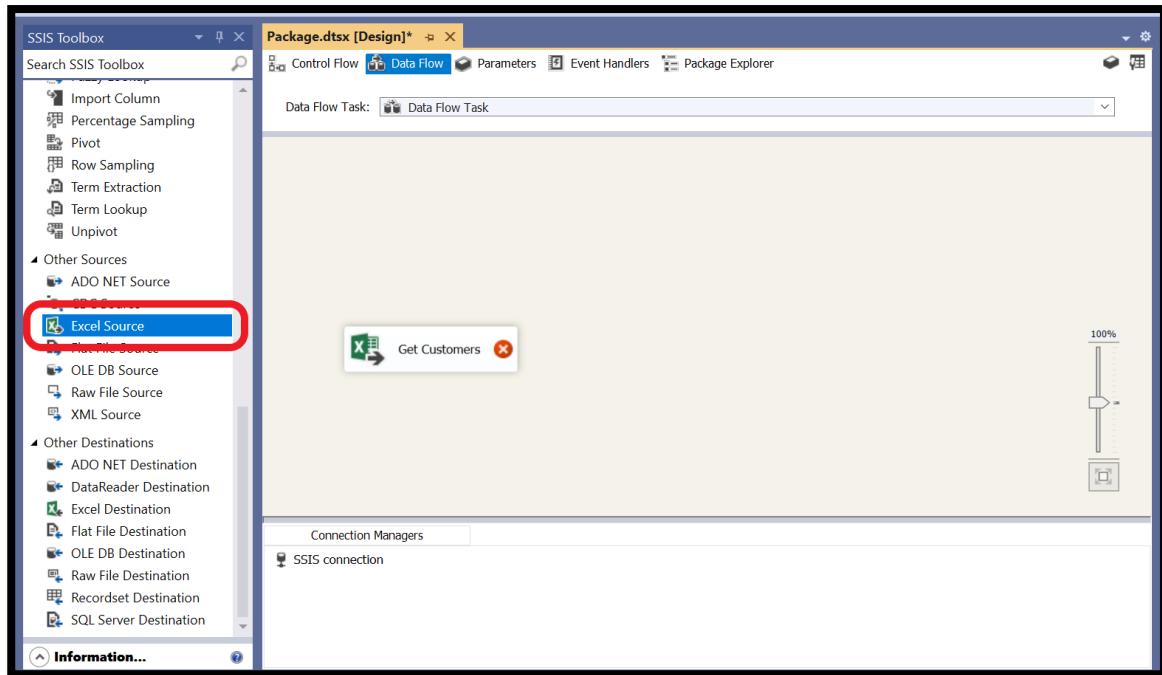


Bước 3: D-click vào Insert Into Dims để mở Data Flow panel

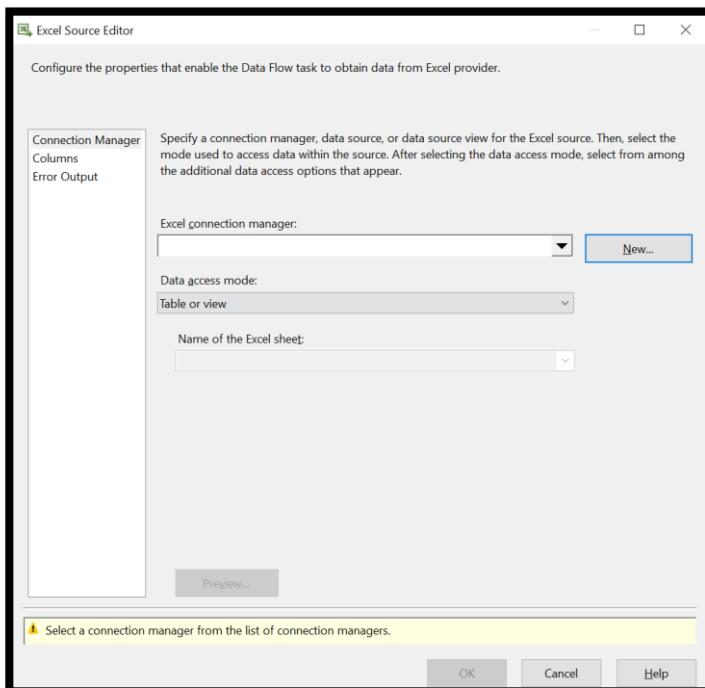


Đỗ dữ liệu vào dim customers

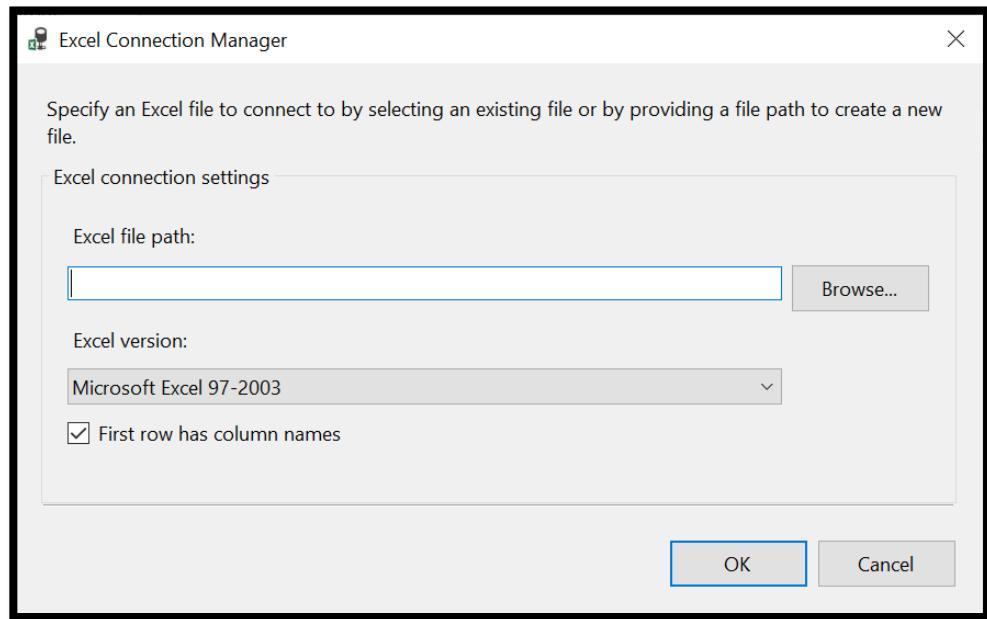
Bước 4: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Customers”.



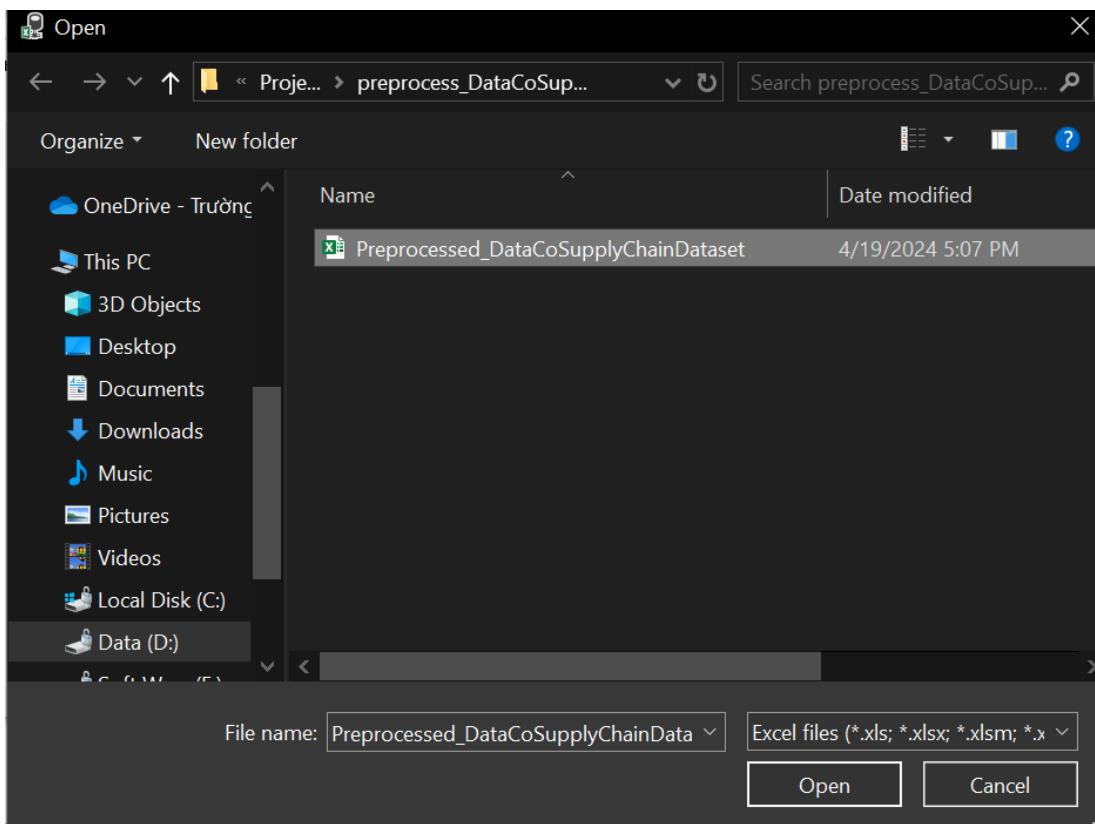
Bước 5: D-click vào **Get Customers** chọn New để tạo một Excel File Connection Manager. Chọn New để tạo một connection tới 1 file excel mới.



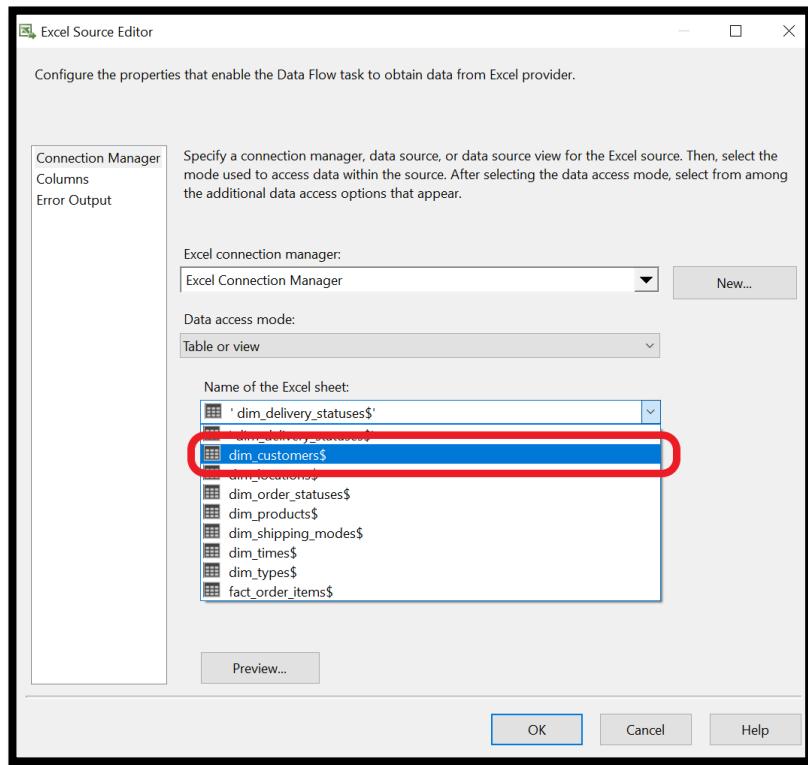
Bước 6: Chọn **Browser...**, sau đó chọn đường dẫn tới Excel.



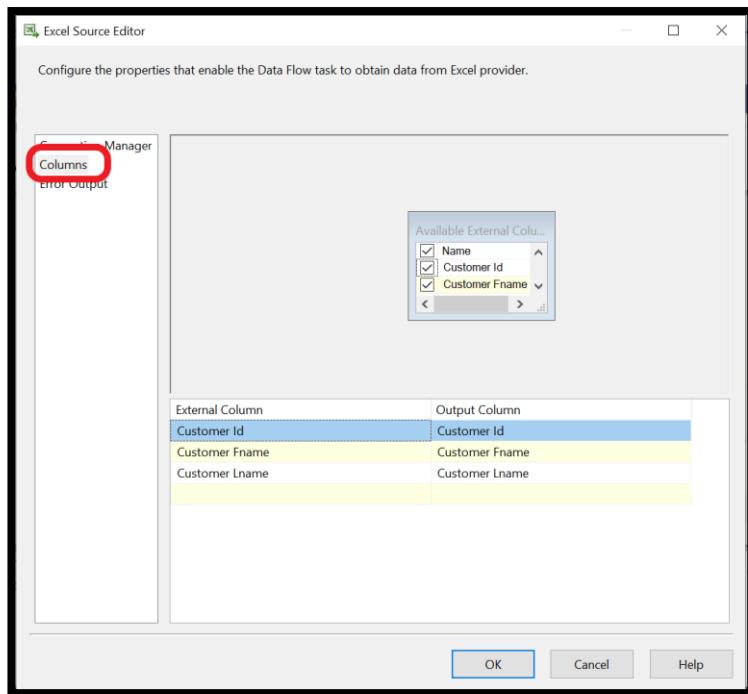
Bước 8: Chọn file dữ liệu .xlss và chọn **Open, OK** để kết thúc quá trình tạo Excel connection.



Bước 9: Vì dữ liệu được lưu vào các sheet con bên trong file excel nên cần chọn Customers để lấy bảng customers.

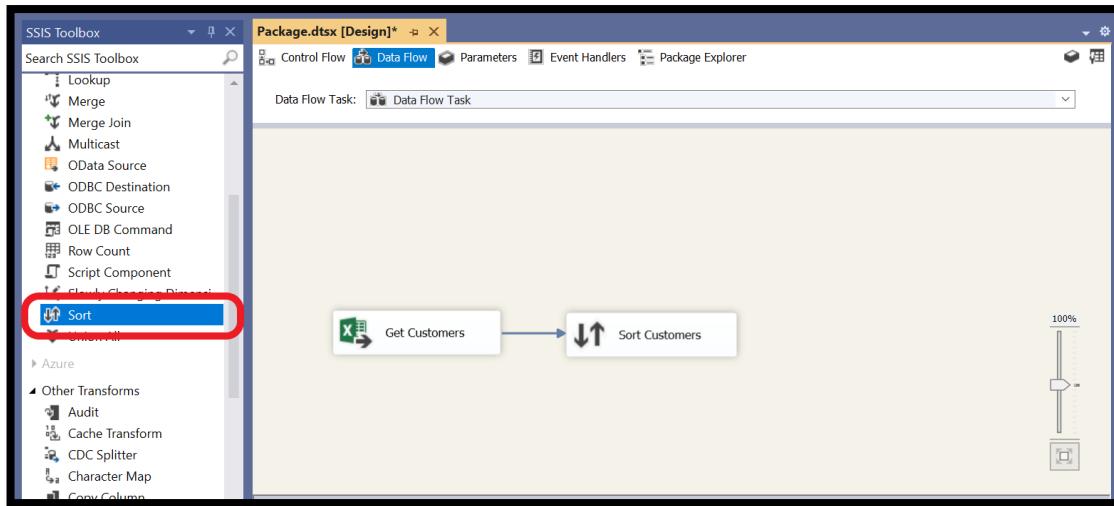


Bước 10: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu customers

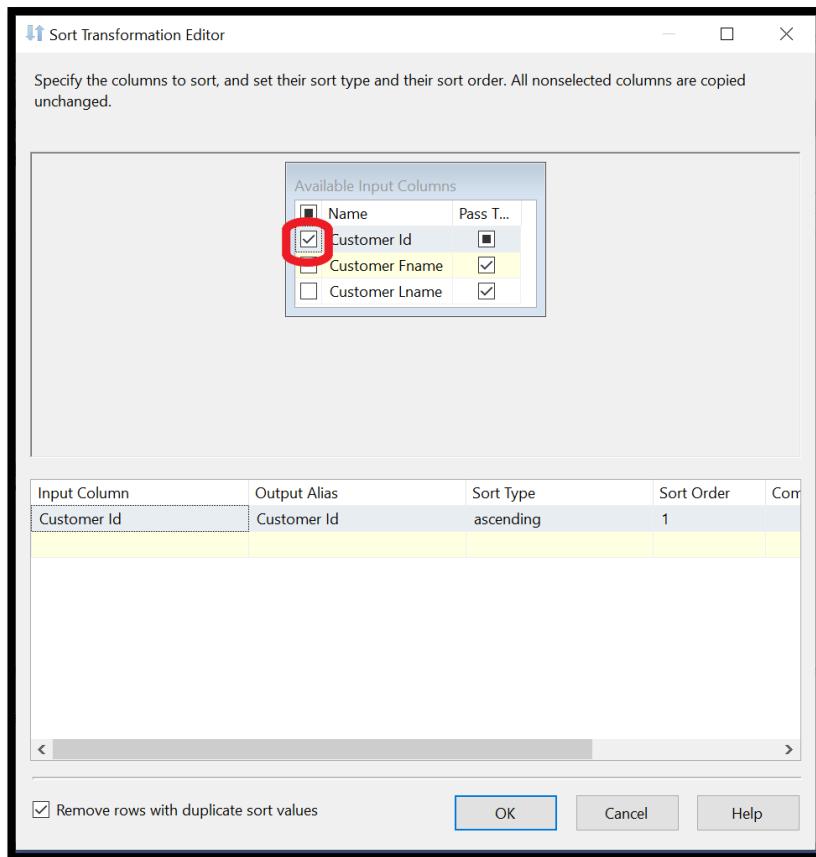


- Có thể D-click và f2 để đổi tên Excel connection nhằm dễ dàng hơn trong quá trình SSIS. Đồ án này không đổi tên Excel connection

Bước 11: Kéo thả một Sort mới click và f2 để đặt tên là Sort Customers, nối Get Customers tới Sort Customers.



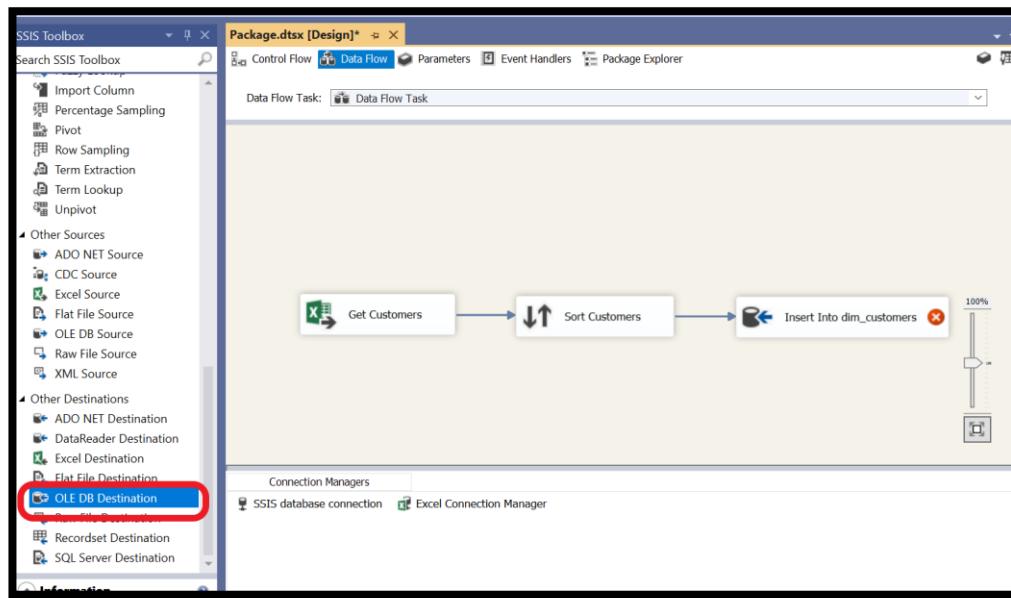
D-click **Sort Customers** và chọn sort theo Customer Id.



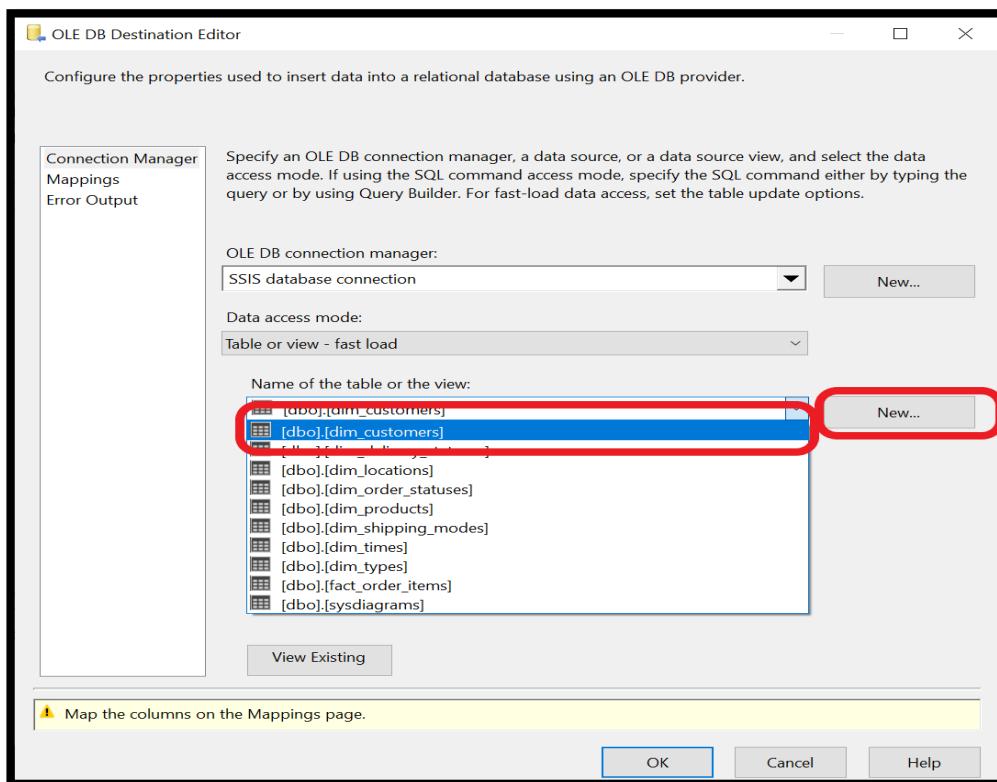
- Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tùy chọn.

Chọn **OK** để kết thúc sort.

Bước 12: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **Insert Into dim_customers**, nối **Sort Customers** tới **Insert Into dim_customers**.

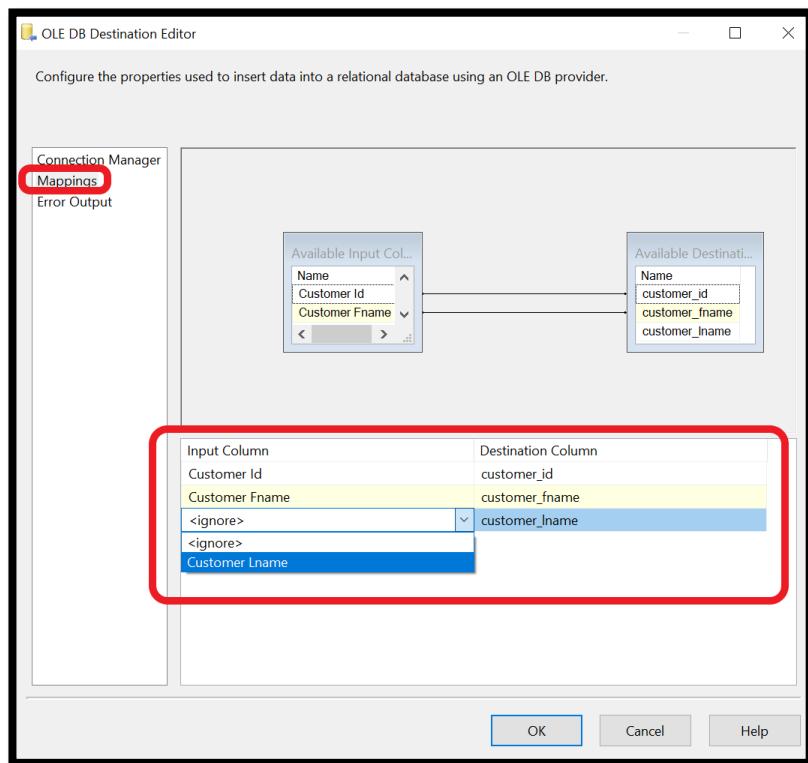


Bước 13: D-click và **Insert Into dim_customers**, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **dim_customers**



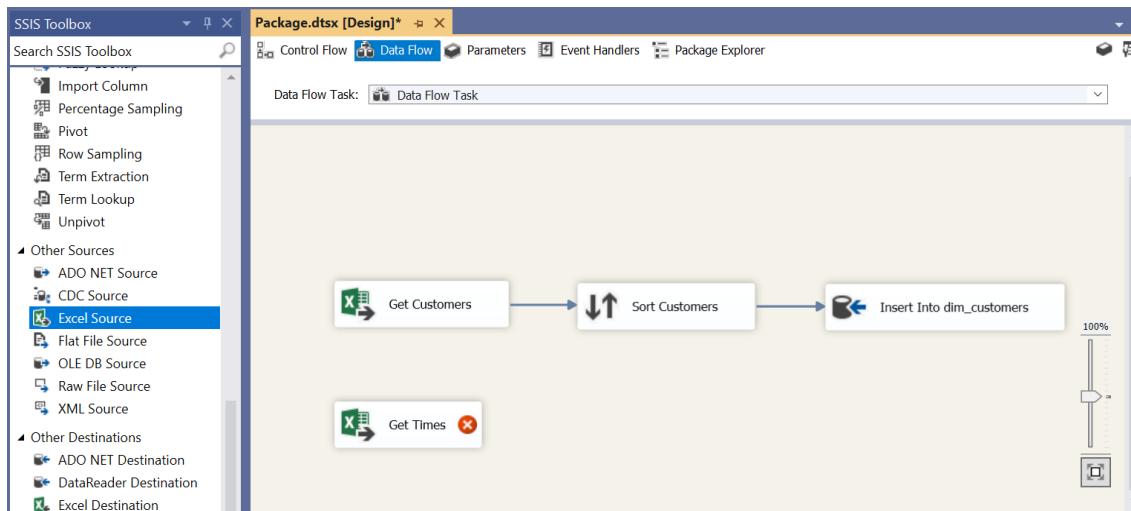
Bước 14: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_customers.

Chọn OK để kết thúc

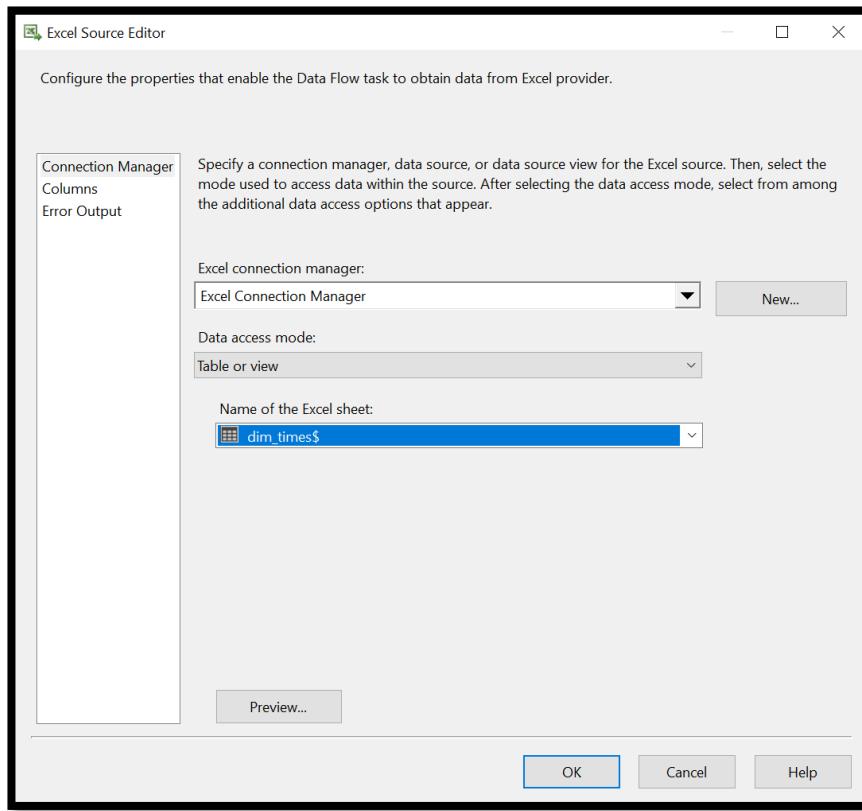


Đỗ dữ liệu vào dim_times

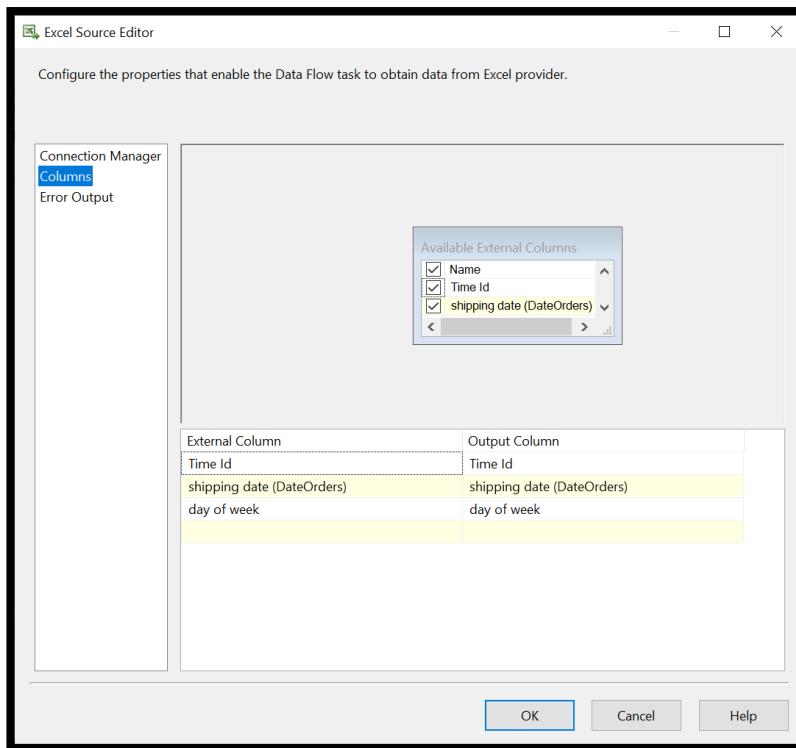
Bước 15: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Times”.



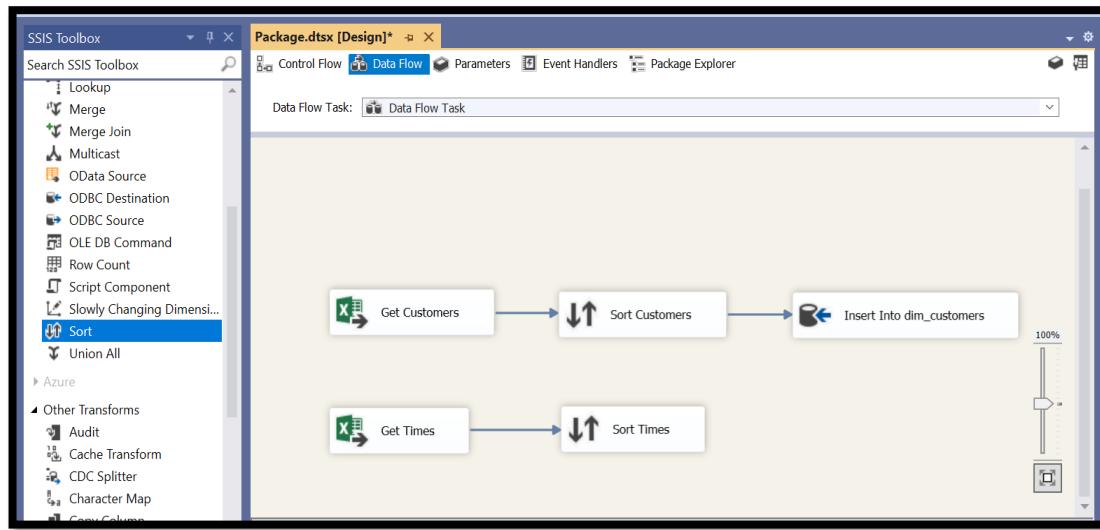
Bước 16: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn dim_times.



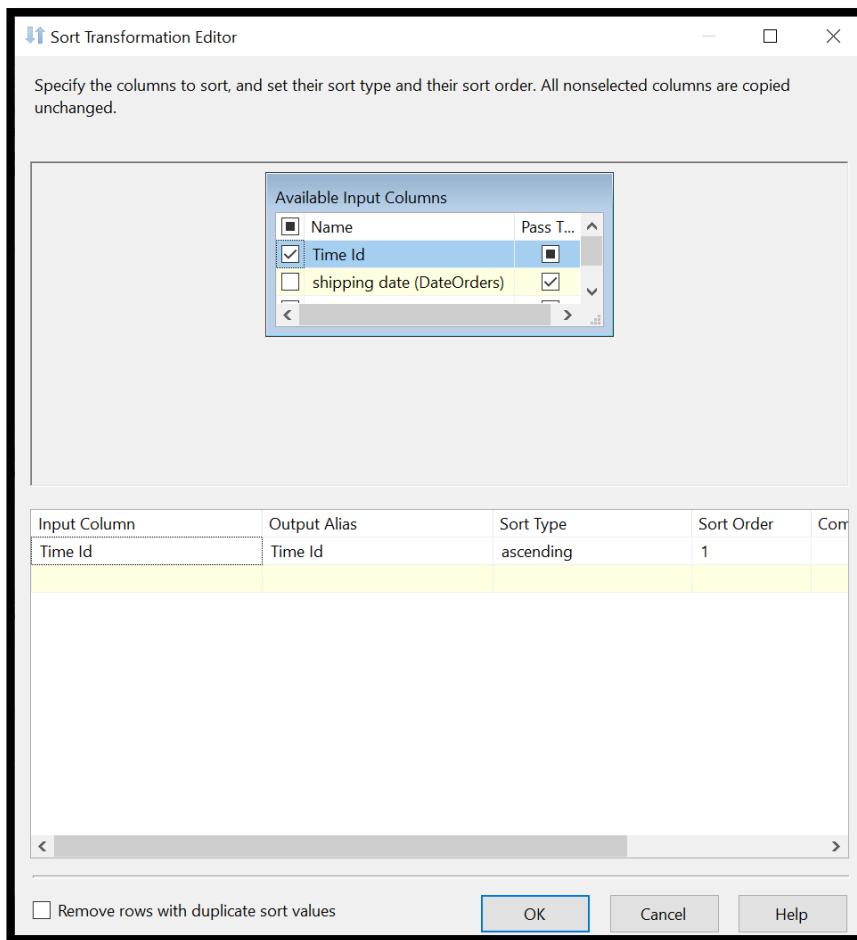
Bước 17: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu times.



Bước 18: Kéo thả một Sort mới click và f2 để đặt tên là Sort Customers, nối Get Customers tới Sort Customers.



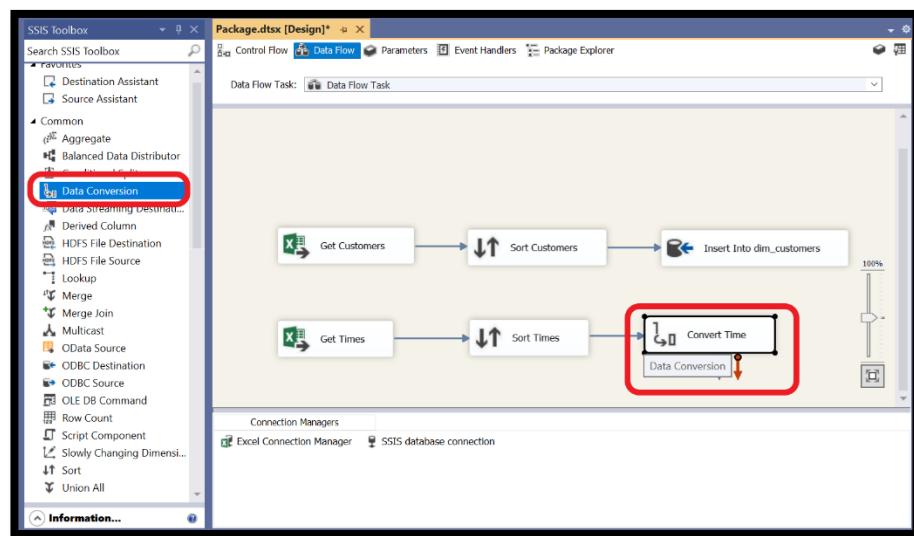
D-click **Sort Times** và chọn sort theo Time Id.



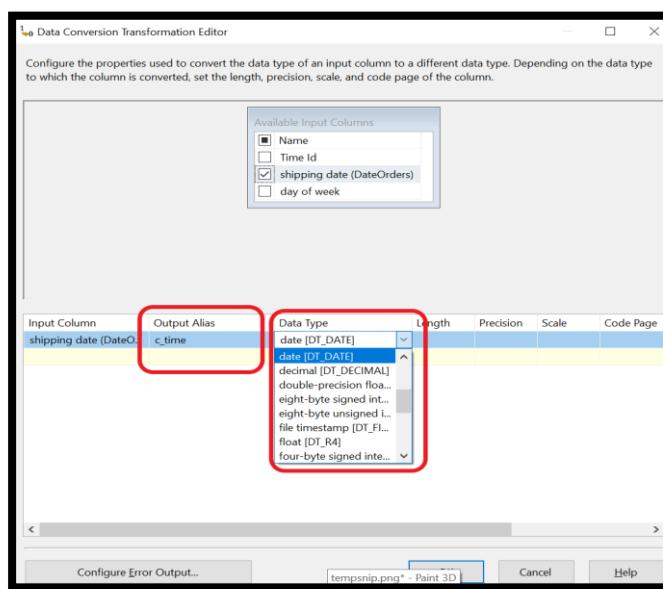
- Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tùy chọn.

Chọn **OK** để kết thúc sort.

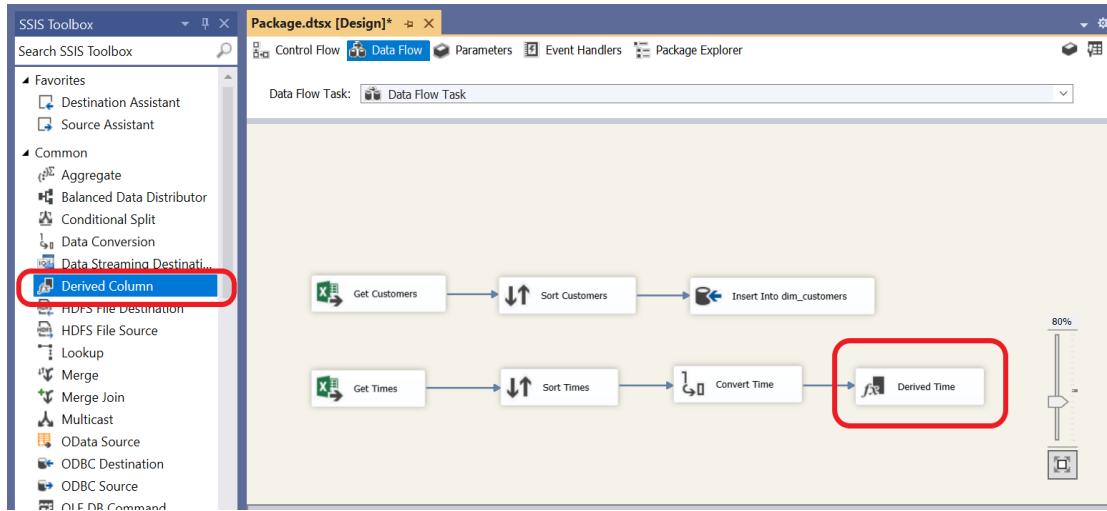
Bước 19: Kiểu dữ liệu thời gian ban đầu khi lấy từ excel sẽ được mặc định là kiểu string. Ta dùng Data Conversion để chuyển đổi kiểu string về dạng DateTime. Kéo thả **Data Conversion**, click và f2 để đổi tên thành **Convert Time**, nối **Sort Times** tới **Convert Time**.



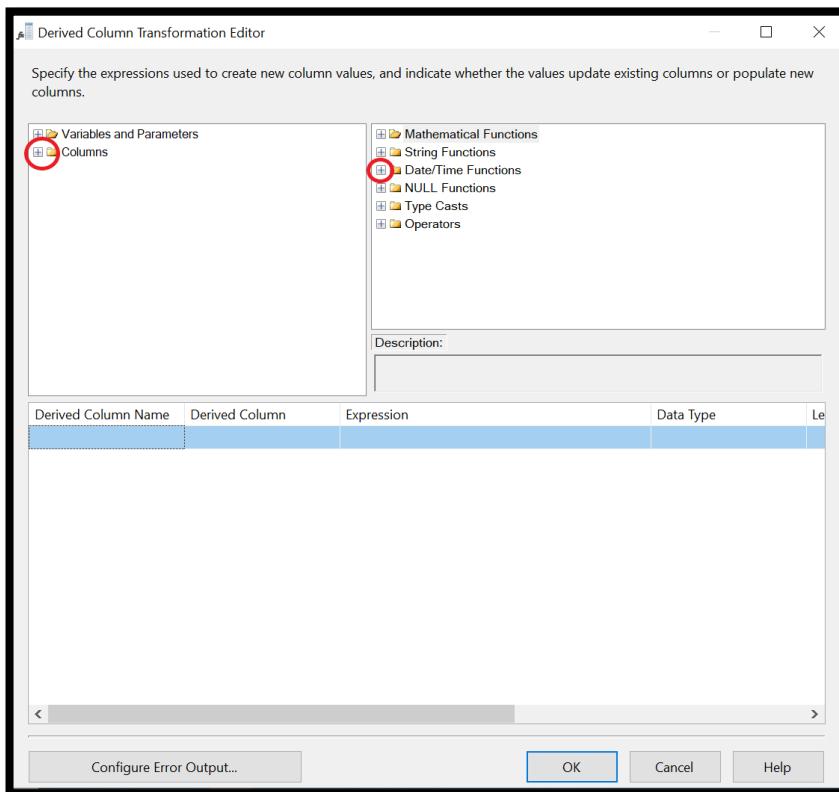
Bước 20: D-click vào **Convert Time**, chọn “shipping date (DateOrders)” cho việc convert. Tại **Output Alias**, nhập tên **c_time** cho time đã được convert kiểu dữ liệu. Tại **Data type** chọn **date [DT_DATE]**. Nhấn **OK** để kết thúc



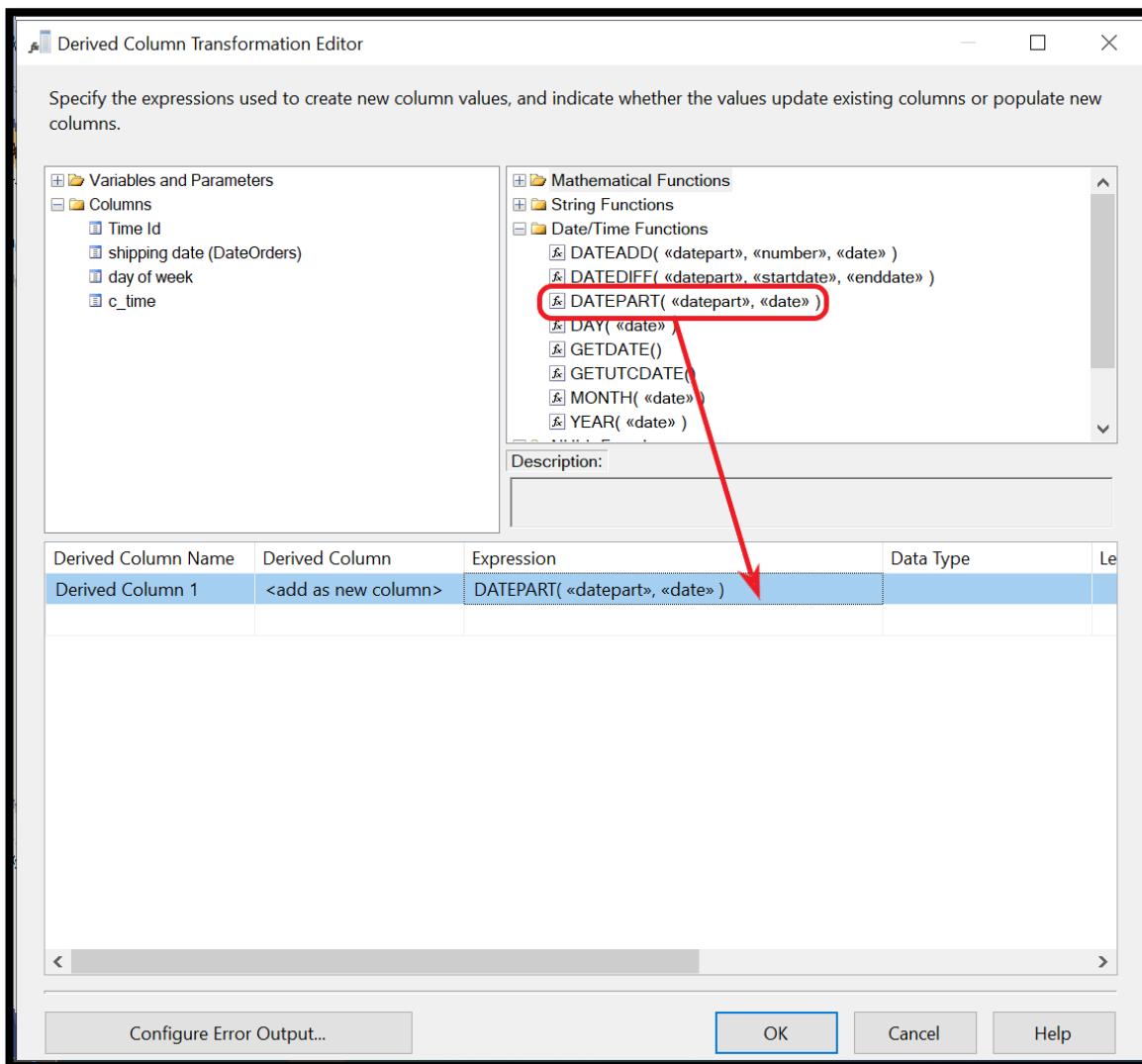
Bước 21: Từ c_time tách thành các data nhỏ hơn như day month year hour minute bằng cách kéo thả **Derived Column**, click và f2 để đổi tên **Derived Time**, nối Convert Time tới Derived Time.



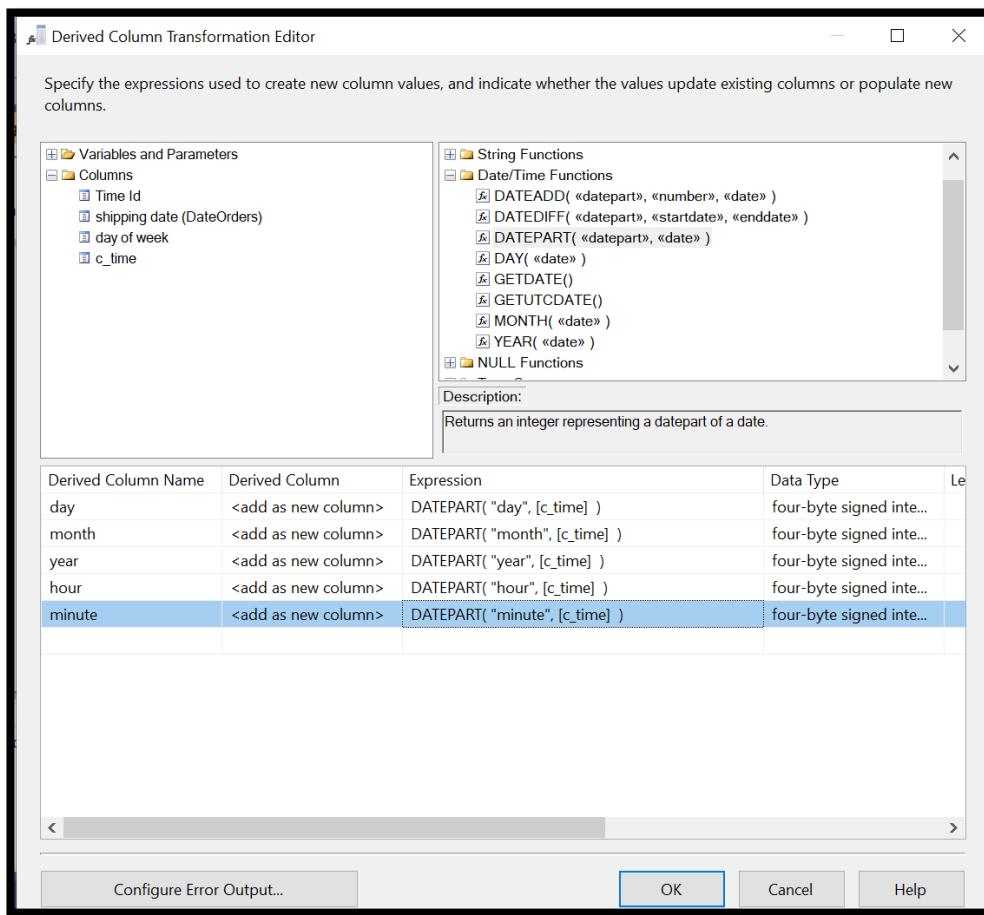
Bước 22: Tại đây ta cần chia dữ liệu từ cột c_time có kiểu dữ liệu dd/MM/yyyy HH:mm thành các cột day, month, year, hour, minute. D-click vào Derived Time. Mở rộng các tuỳ chọn phù hợp



Bước 23: kéo thả DATEPART vào bảng chọn.

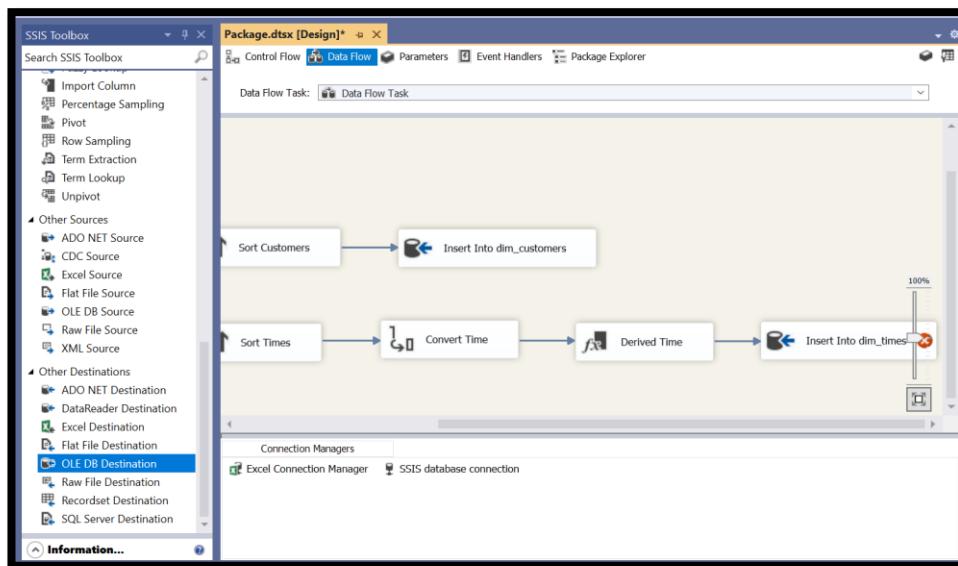


Bước 24: Đổi tên Derived Column 1 thành **day**. Kéo thả c_time vào đúng tham số bên phải trong Expression, tham số bên trái sửa thành “day”. Thực hiện tương tự với month, year, hour, minute ta được

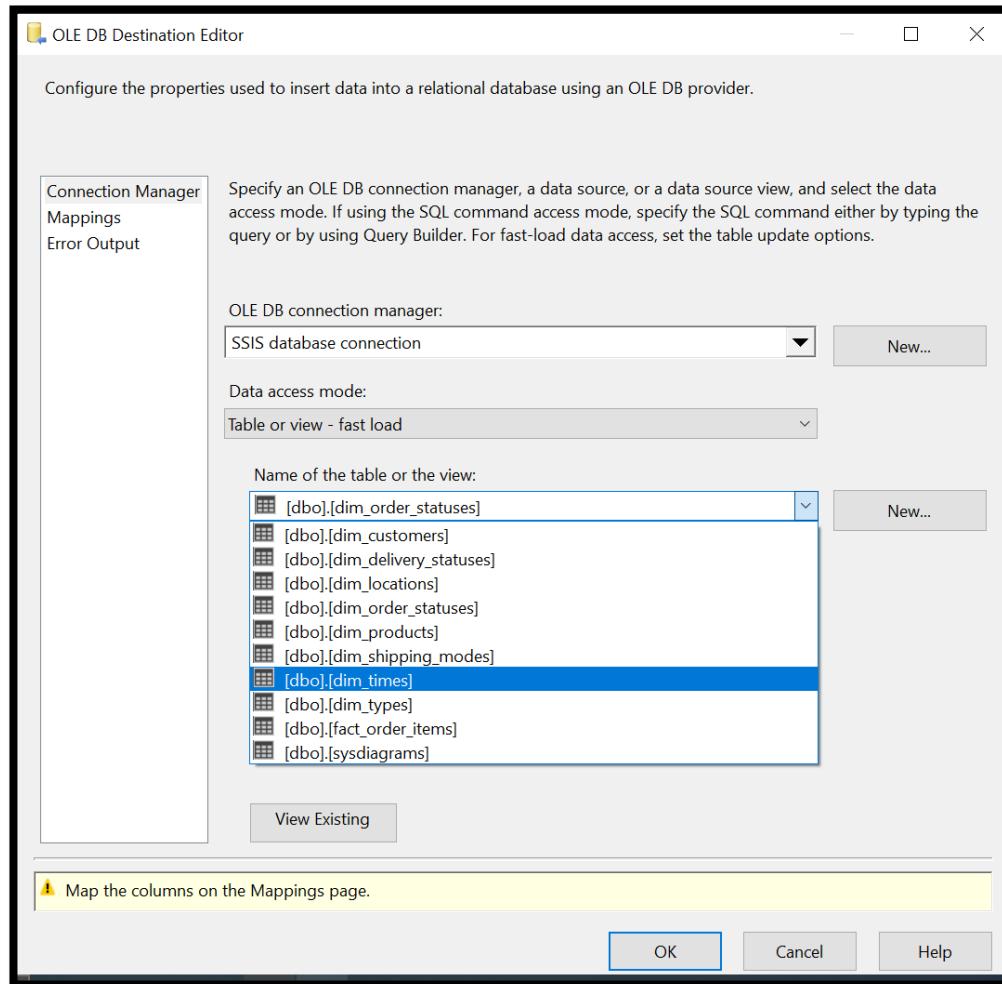


Nhấn OK để kết thúc devered.

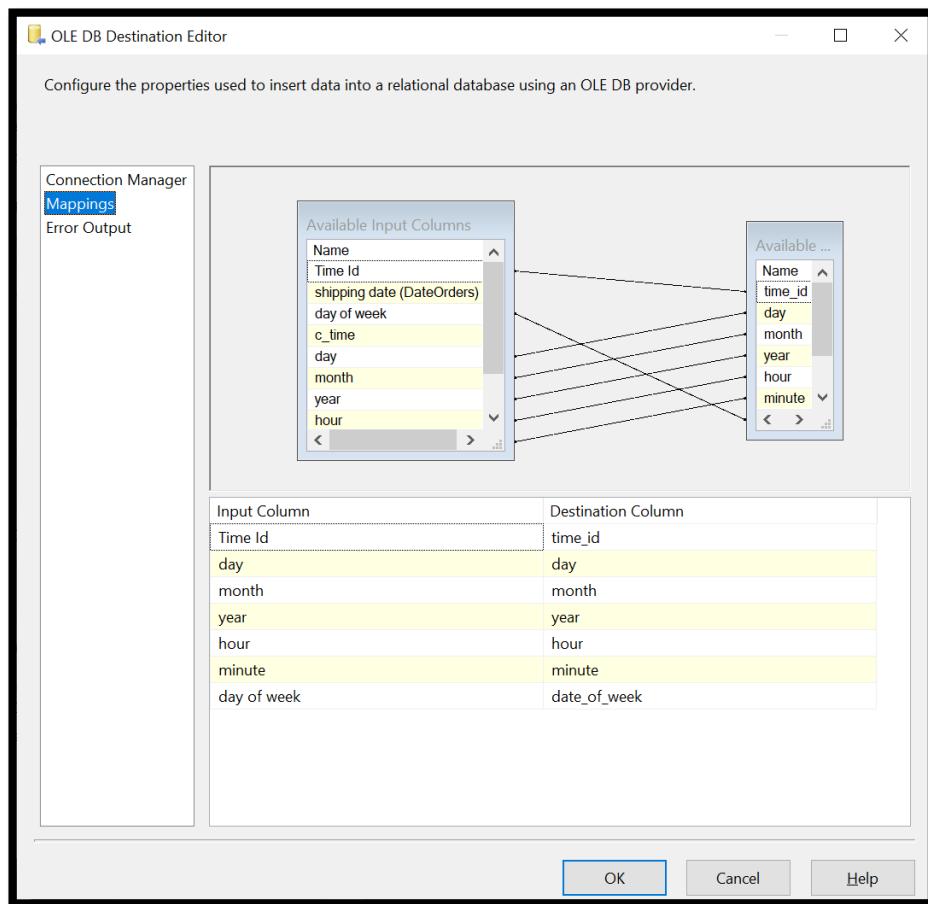
Bước 25: Kéo thả một **OLE DB Destionation** mới, click và f2 để đặt tên là **Insert Into dim_times**, nối **Derived Time** tới **Insert Into dim_times**.



Bước 26: D-click và **Insert Into dim_times**, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **dim_times**

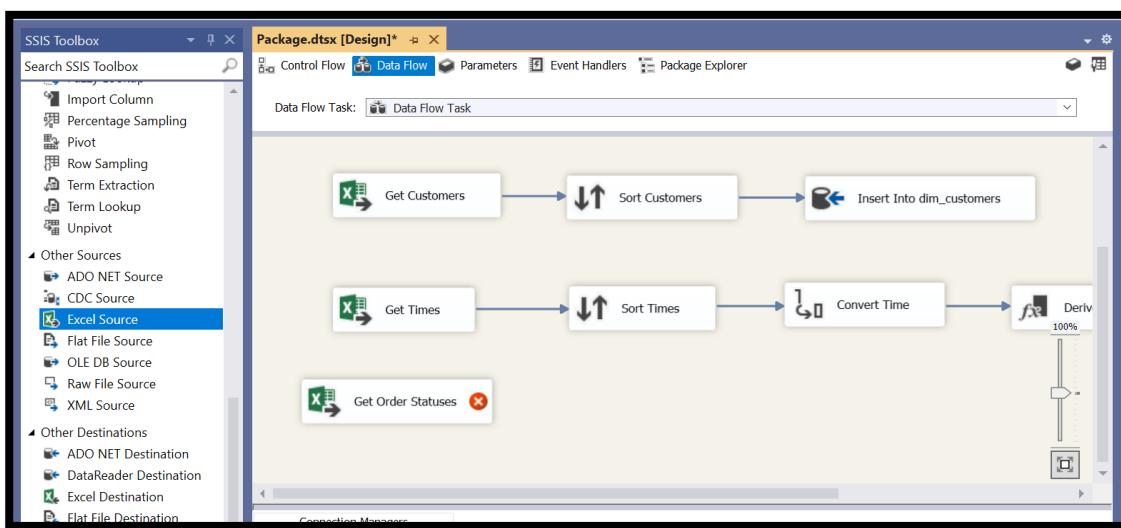


Bước 27: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_times. Chọn OK để kết thúc quá trình đổ dữ liệu vào dim_times

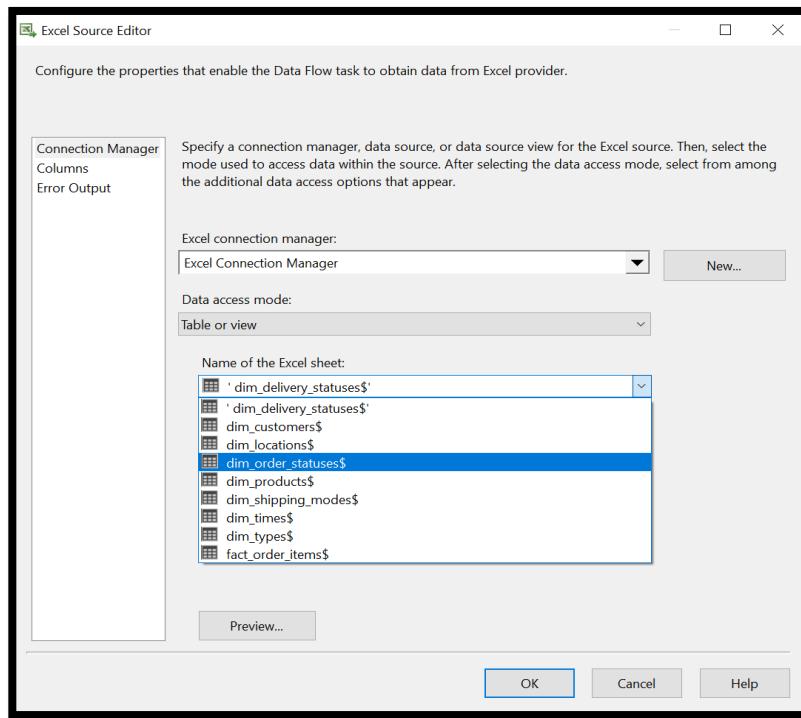


Đỗ dữ liệu vào dim_order_statuses

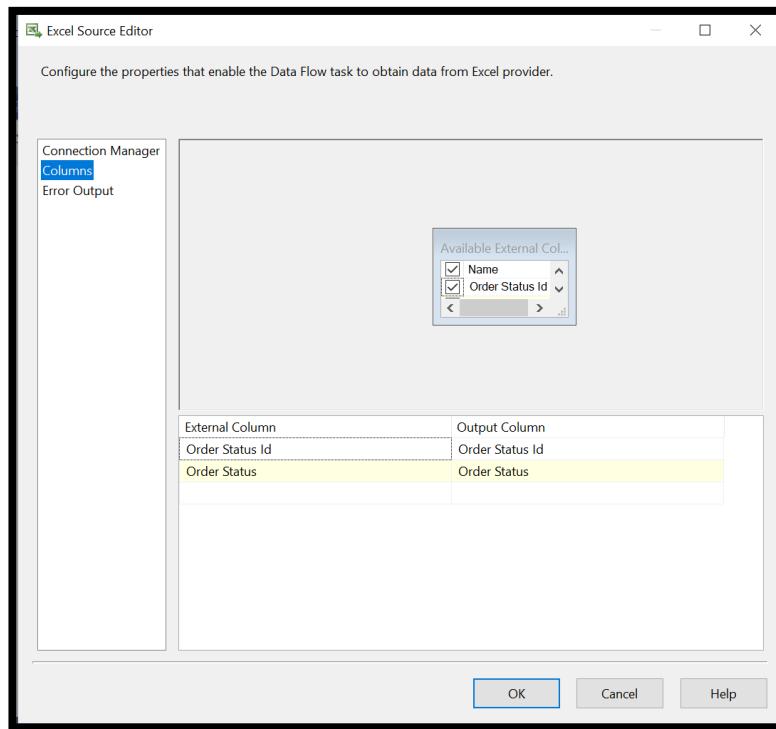
Bước 28: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Order statuses”.



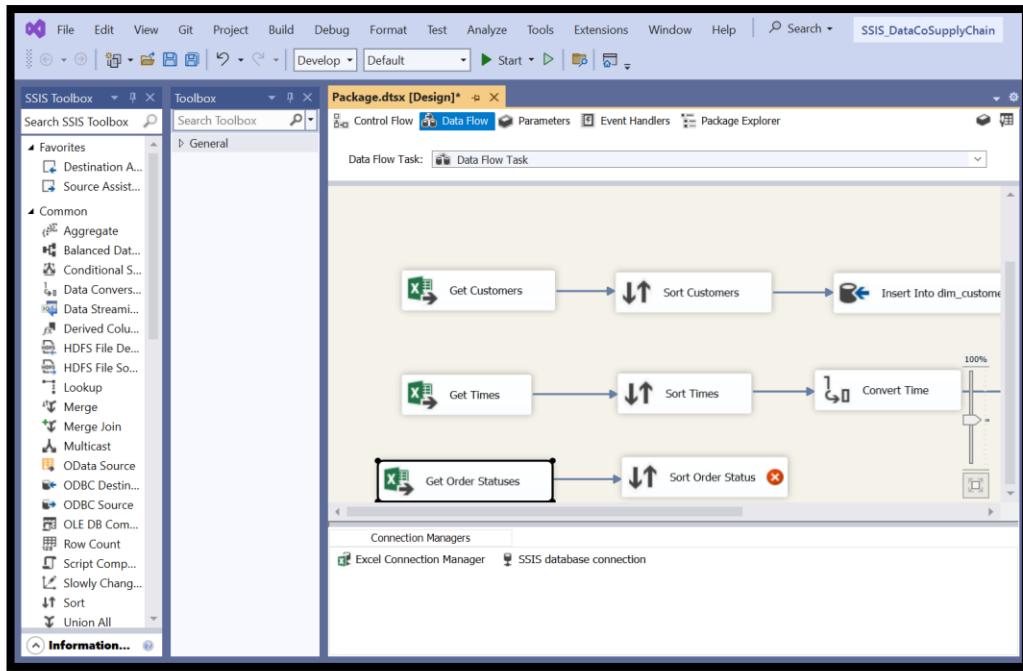
Bước 29: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn dim_order_statuses.



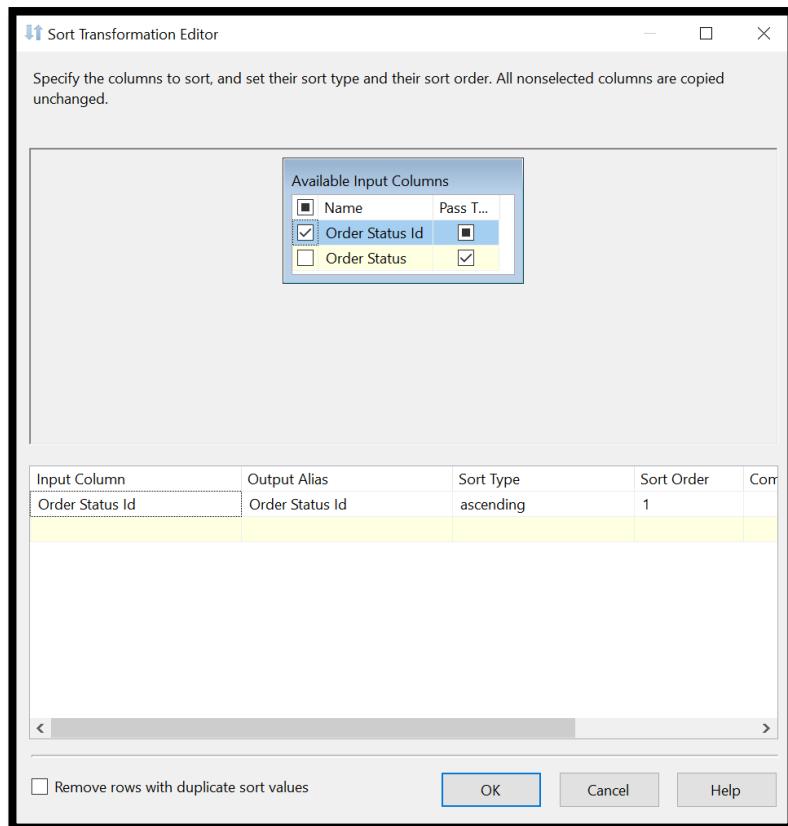
Bước 30: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn “OK” để kết thúc quá trình lấy dữ liệu order statuses.



Bước 31: Kéo thả một Sort mới click và f2 để đặt tên là Sort Order Statuses, nối Get Order Statuses tới Sort Order Statuses



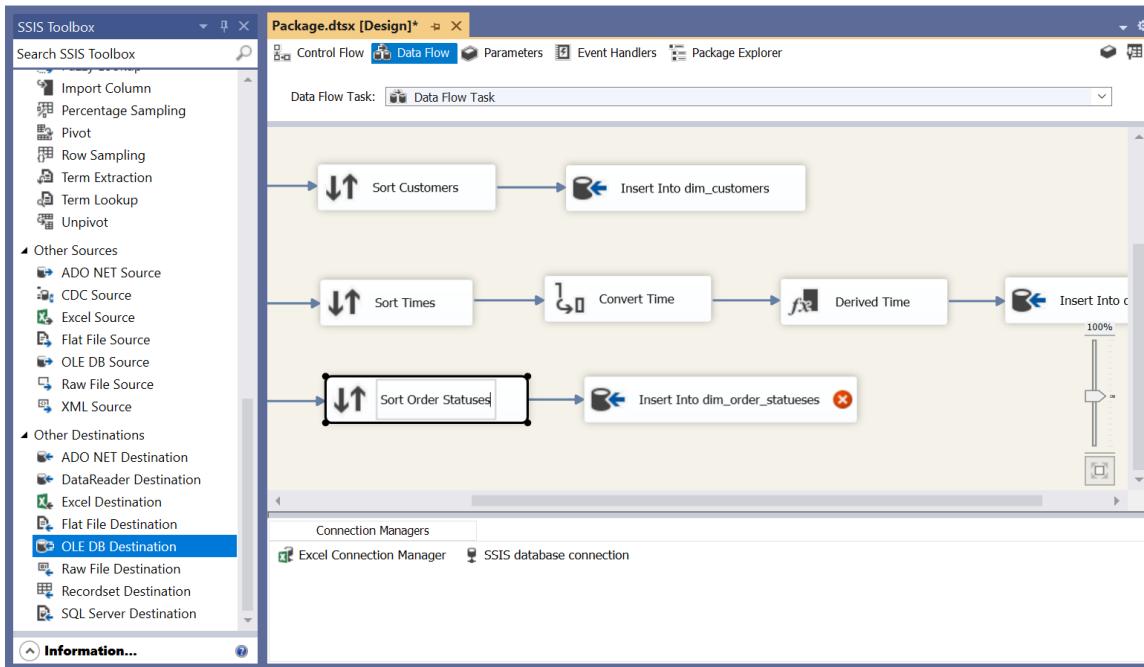
D-click Sort Order Statuses và chọn sort theo Order Statuses Id.



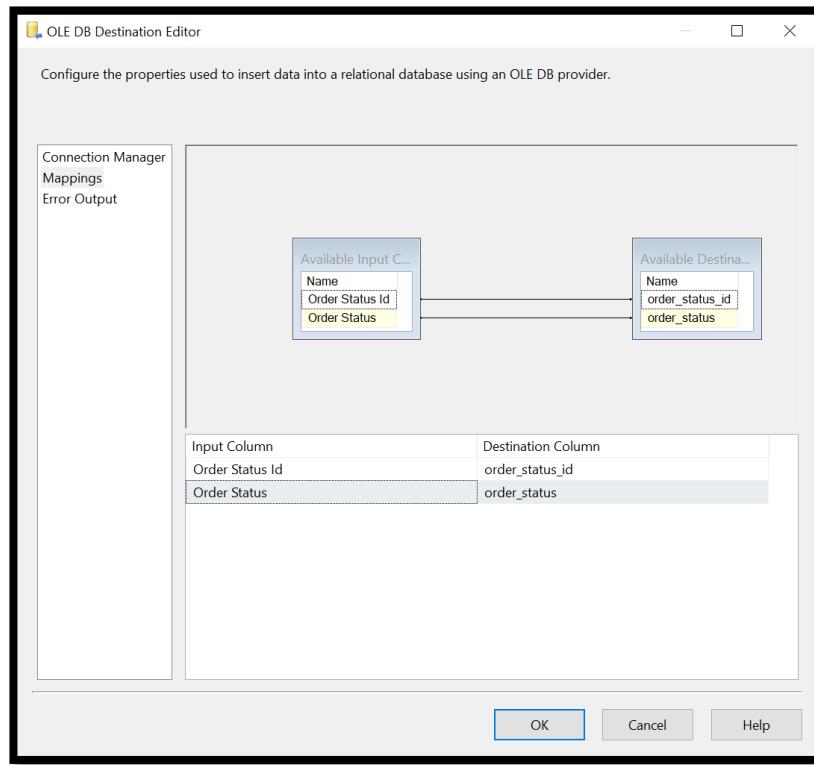
*Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tuỳ chọn.

Chọn **OK** để kết thúc sort.

Bước 33: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **Insert Into dim_order_statuses**, nối **Sort Order Statuses** tới **Insert Into dim_order_statuses**.

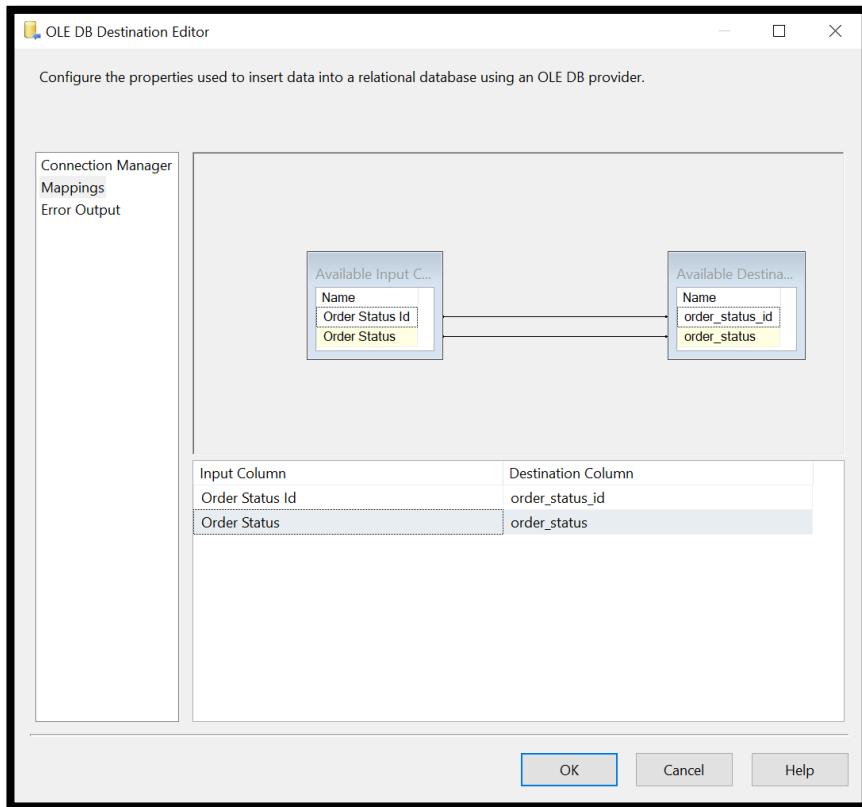


Bước 34: D-click và **Insert Into dim_order_statuses**, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **dim_order_statuses**



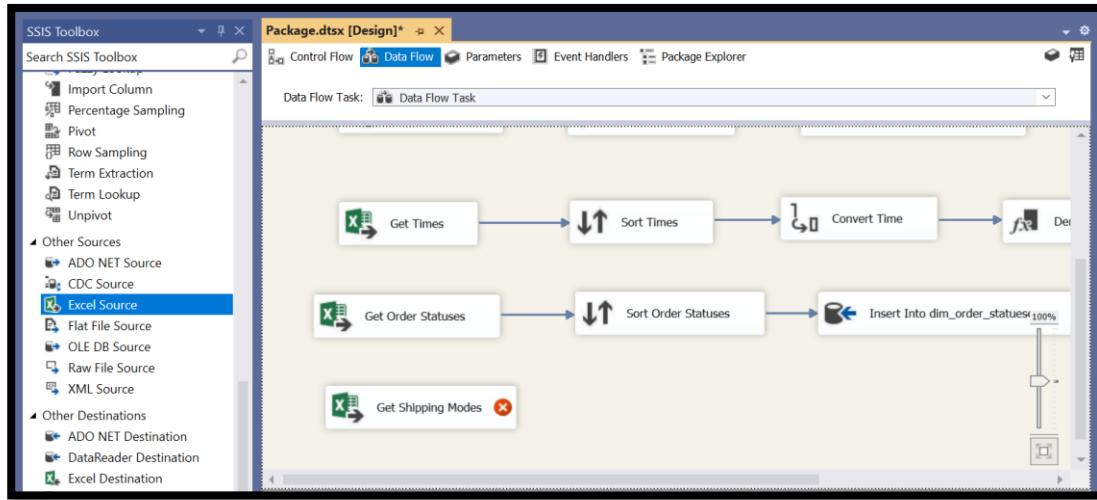
Bước 35: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_customers.

Chọn OK để kết thúc

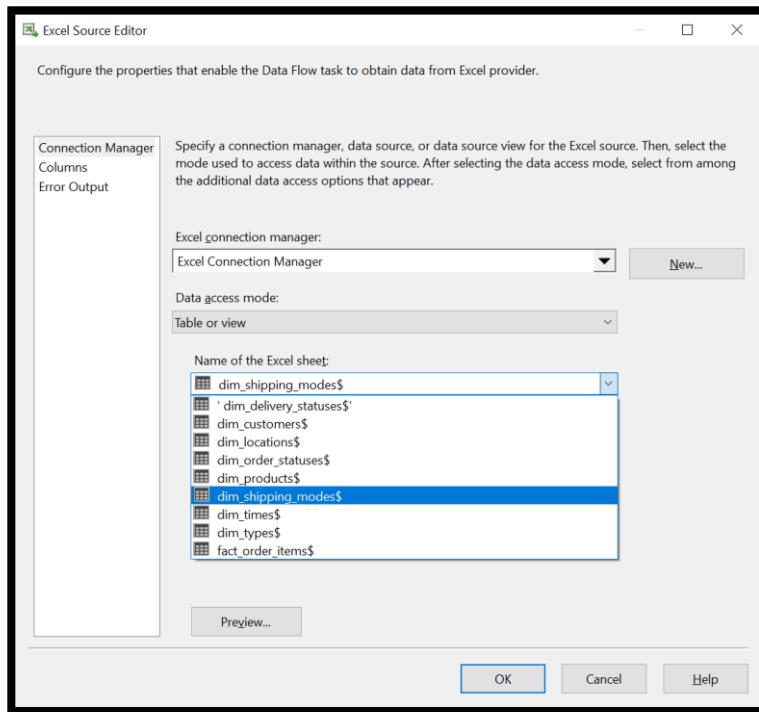


Đỗ dữ liệu vào dim_order_statuses

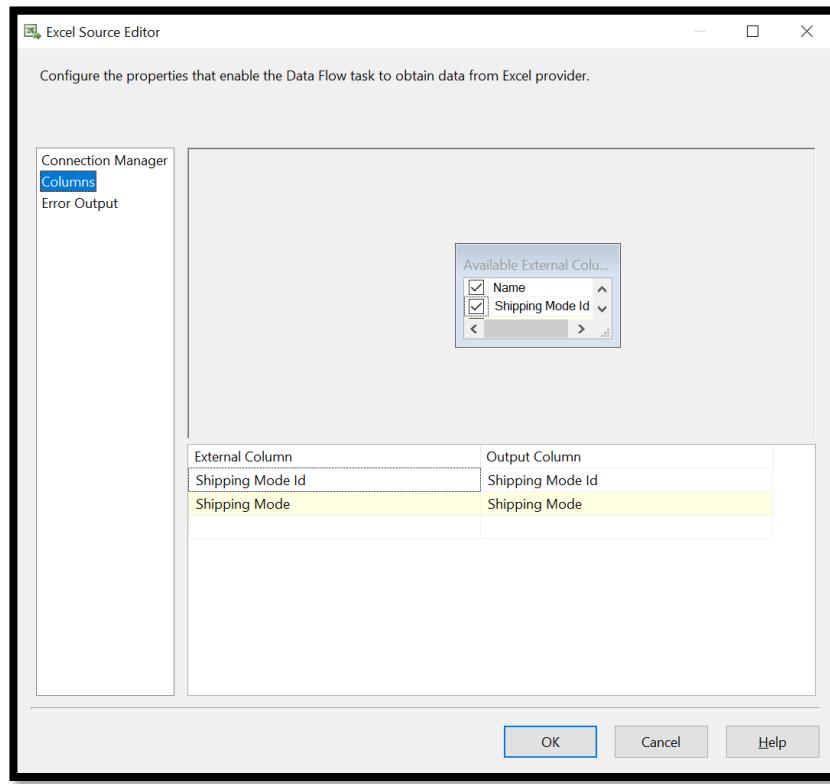
Bước 36: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Shipping Modes”.



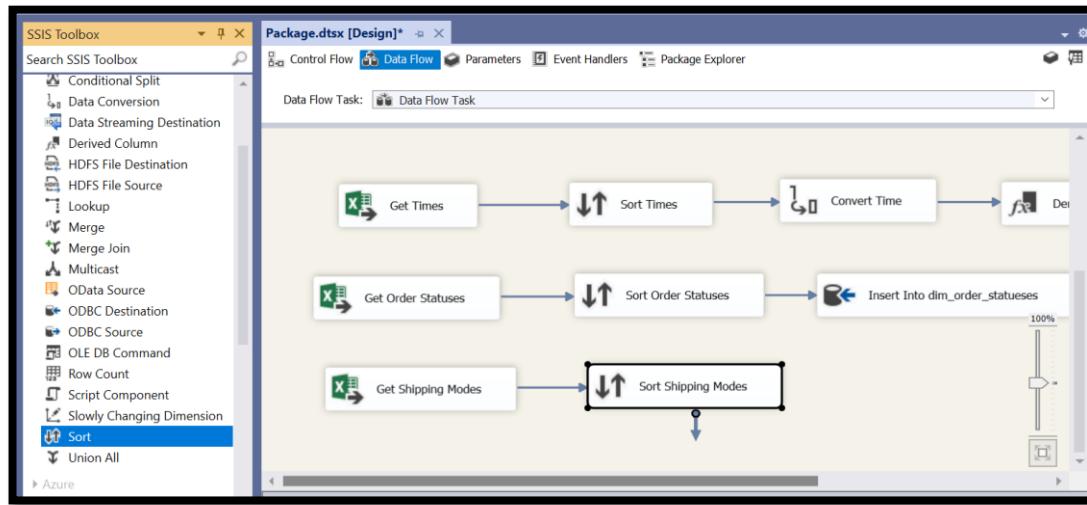
Bước 37: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn dim_shipping_modes.



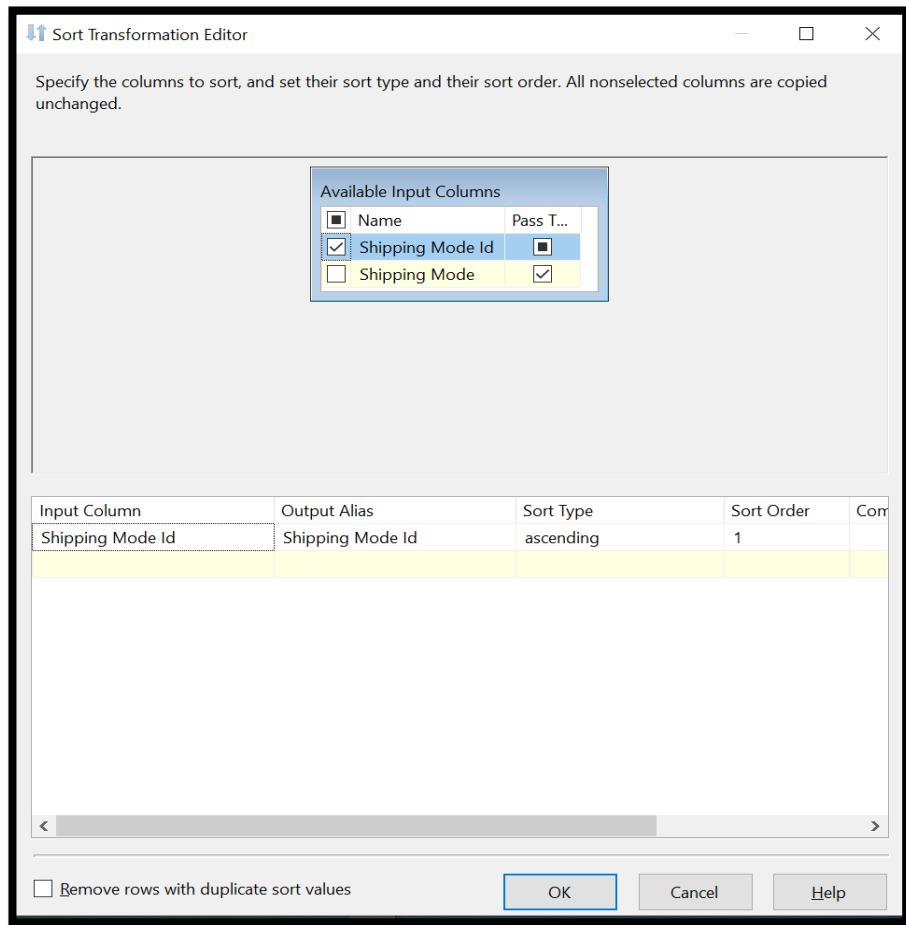
Bước 38: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu shipping modes



Bước 39: Kéo thả một Sort mới click và f2 để đặt tên là Sort Order Statuses, nối Get Shipping Modes tới Sort Shipping Modes.



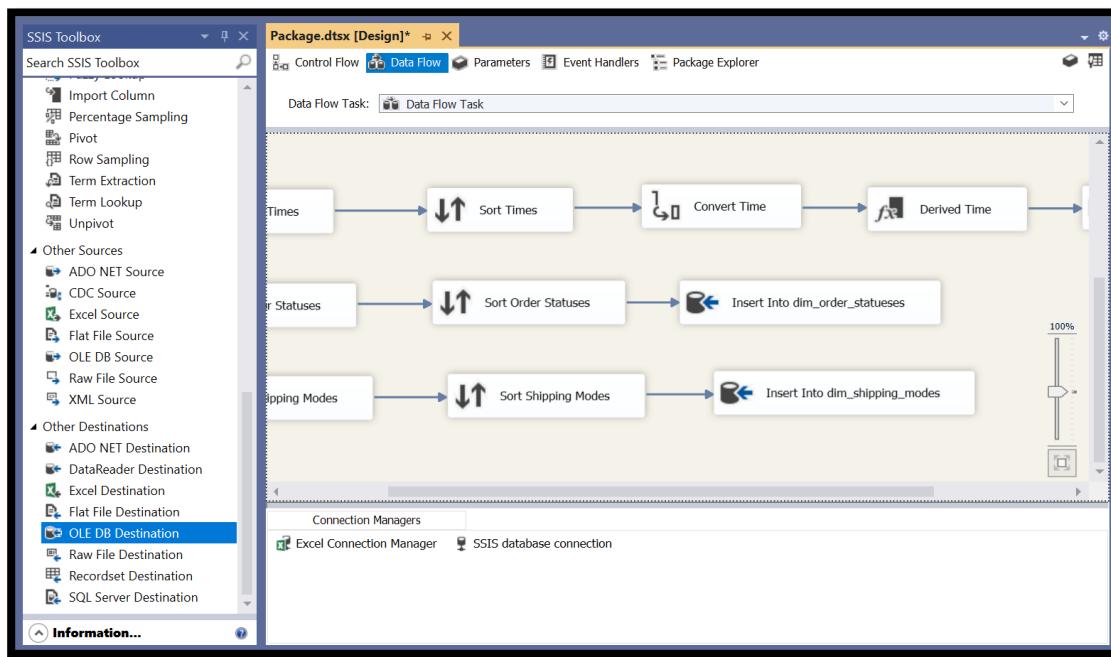
D-click **Sort Shipping Modes** và chọn sort theo **Shipping Mode Id**.



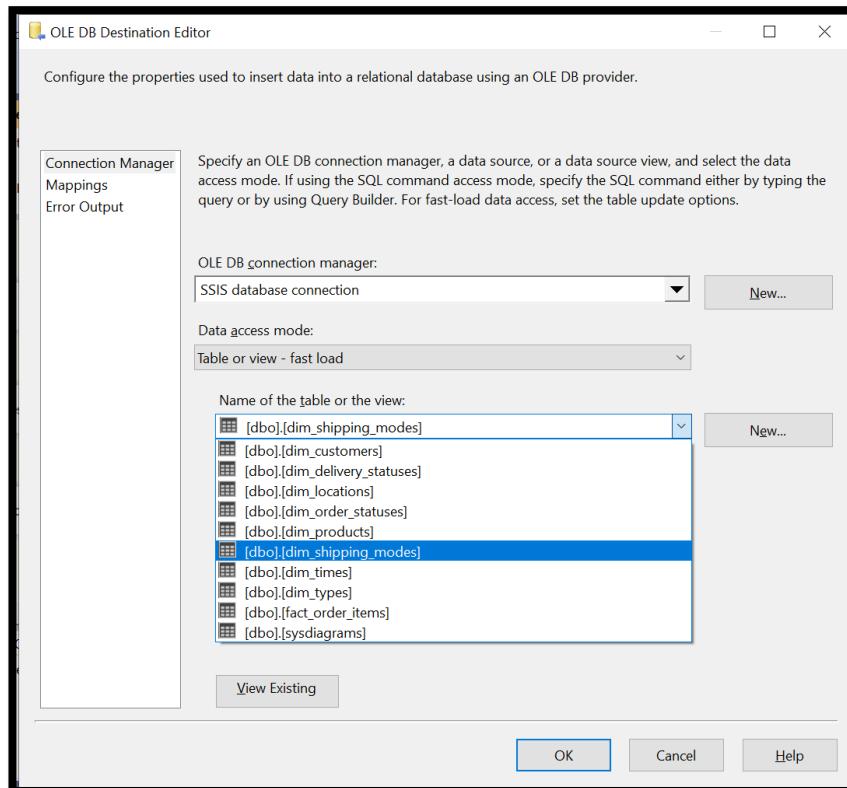
*Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tùy chọn.

Chọn **OK** để kết thúc sort.

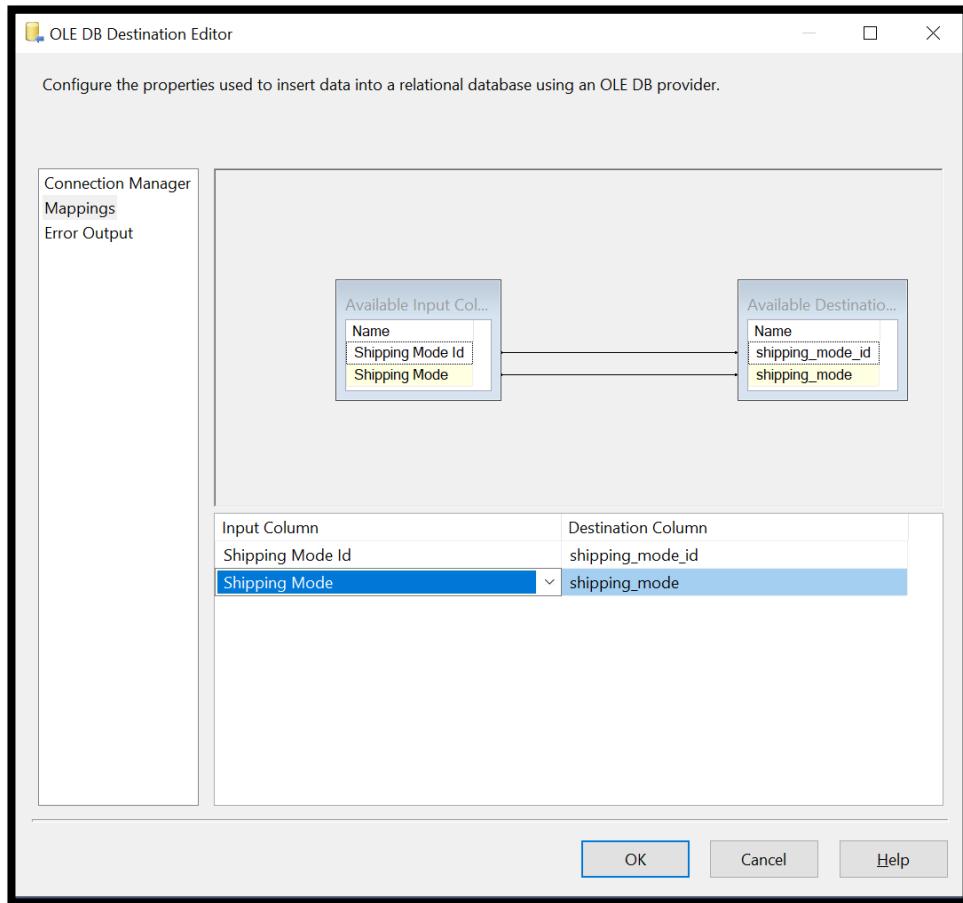
Bước 40: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **Insert Into dim_shipping_modes**, nối **Sort Shipping Modes** tới **Insert Into dim_shipping_modes**.



Bước 41: D-click và **Insert Into dim_shipping_modes**, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **dim_shipping_modes**

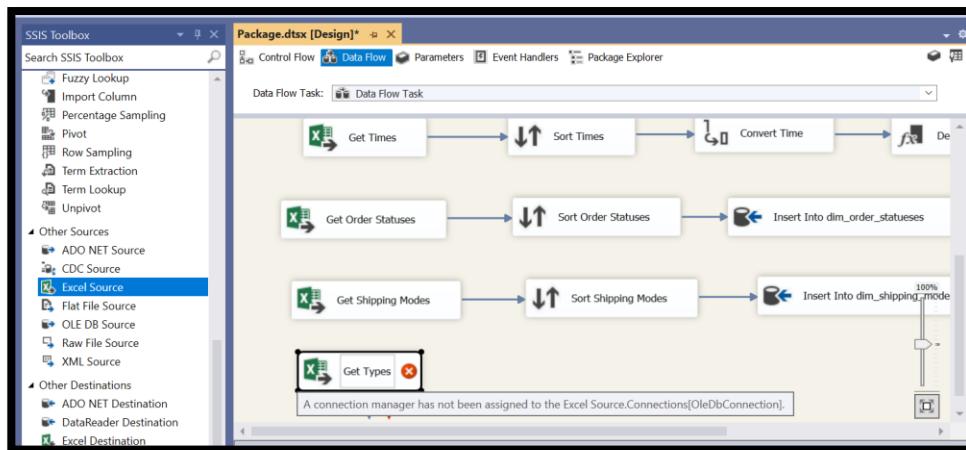


Bước 42: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_shipping_modes. Chọn OK để kết thúc

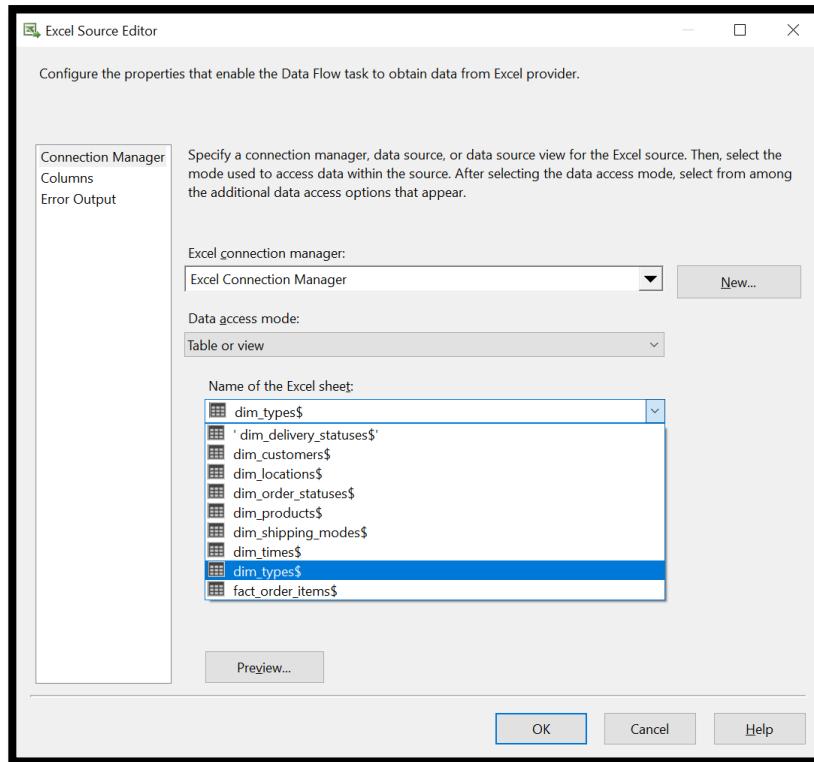


Đỗ dữ liệu vào dim_types

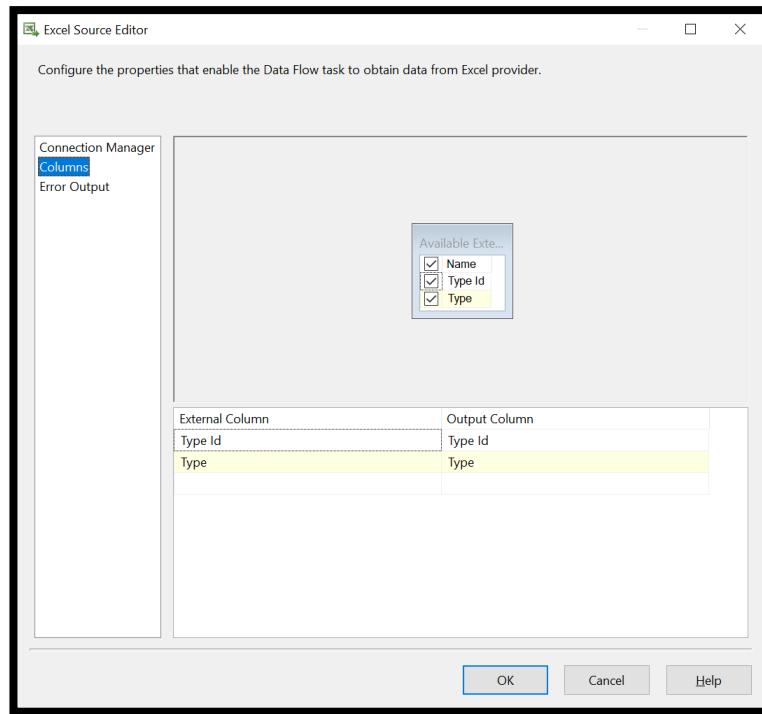
Bước 43: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Types”.



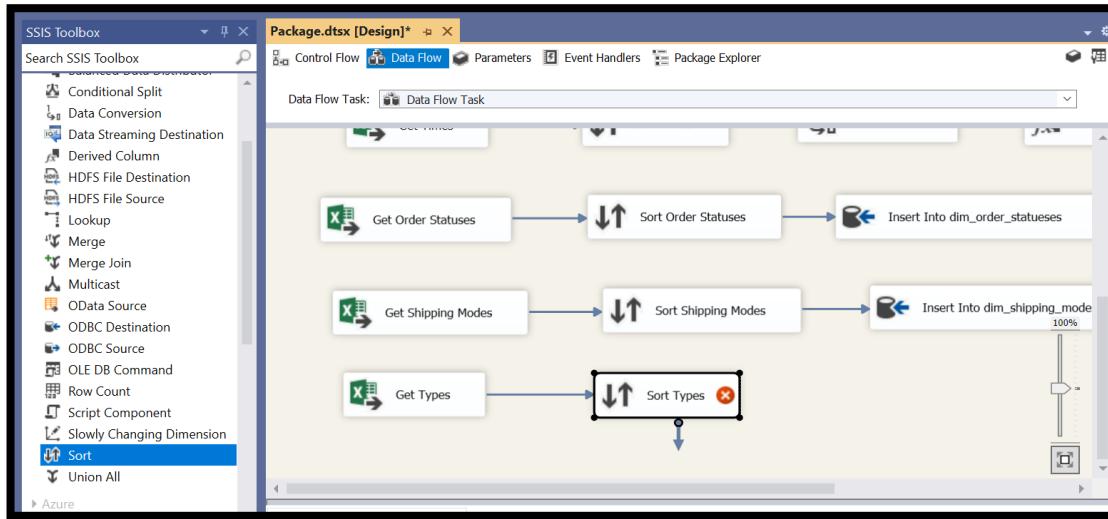
Bước 44: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn dim_types.



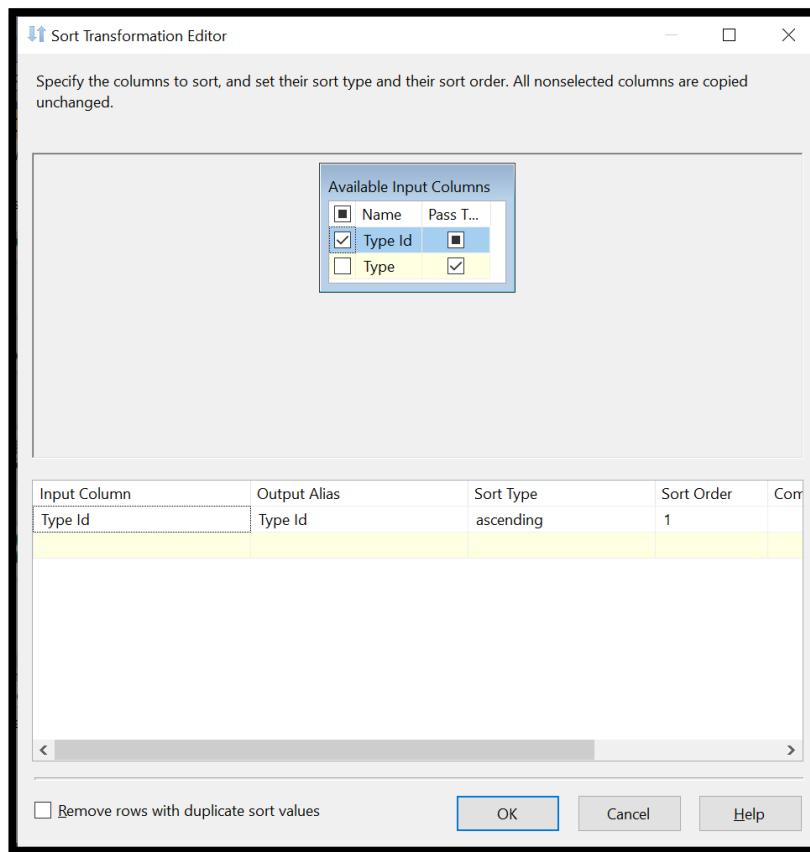
Bước 45: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu types.



Bước 46: Kéo thả một Sort mới click và f2 để đặt tên là Sort Types, nối Get Types tới Sort Types.



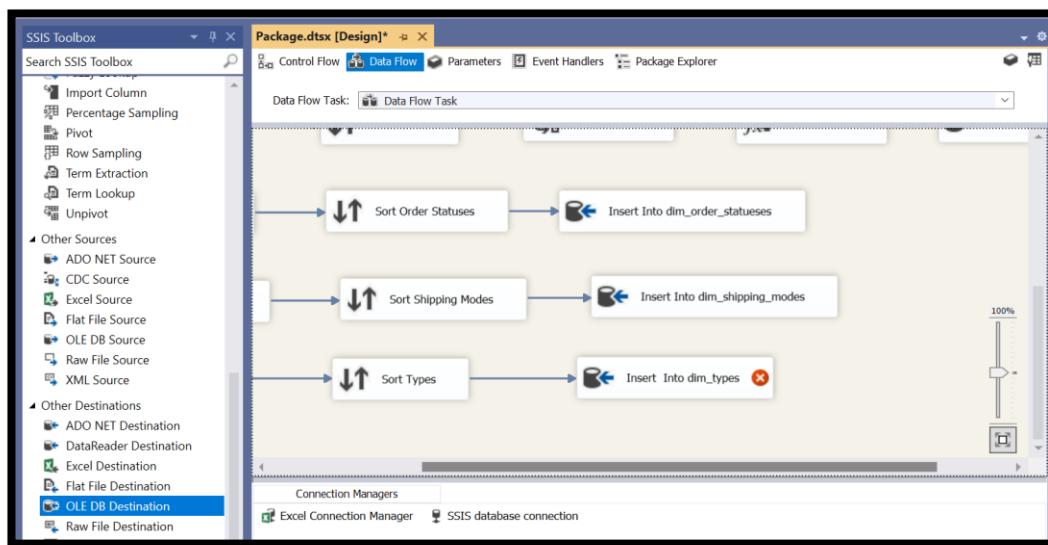
D-click Sort Shipping Modes và chọn sort theo Shipping Type Id.



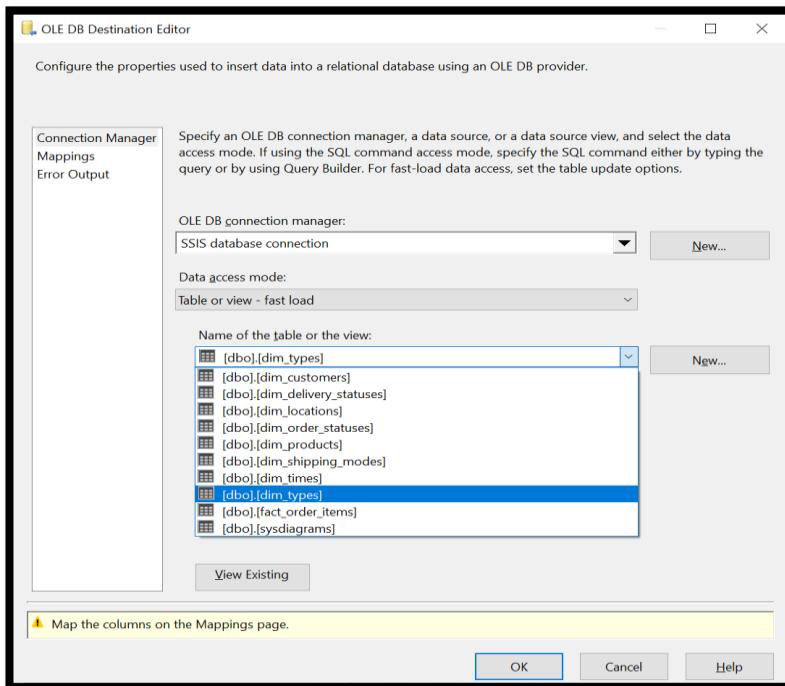
- Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tuỳ chọn.

Chọn **OK** để kết thúc sort.

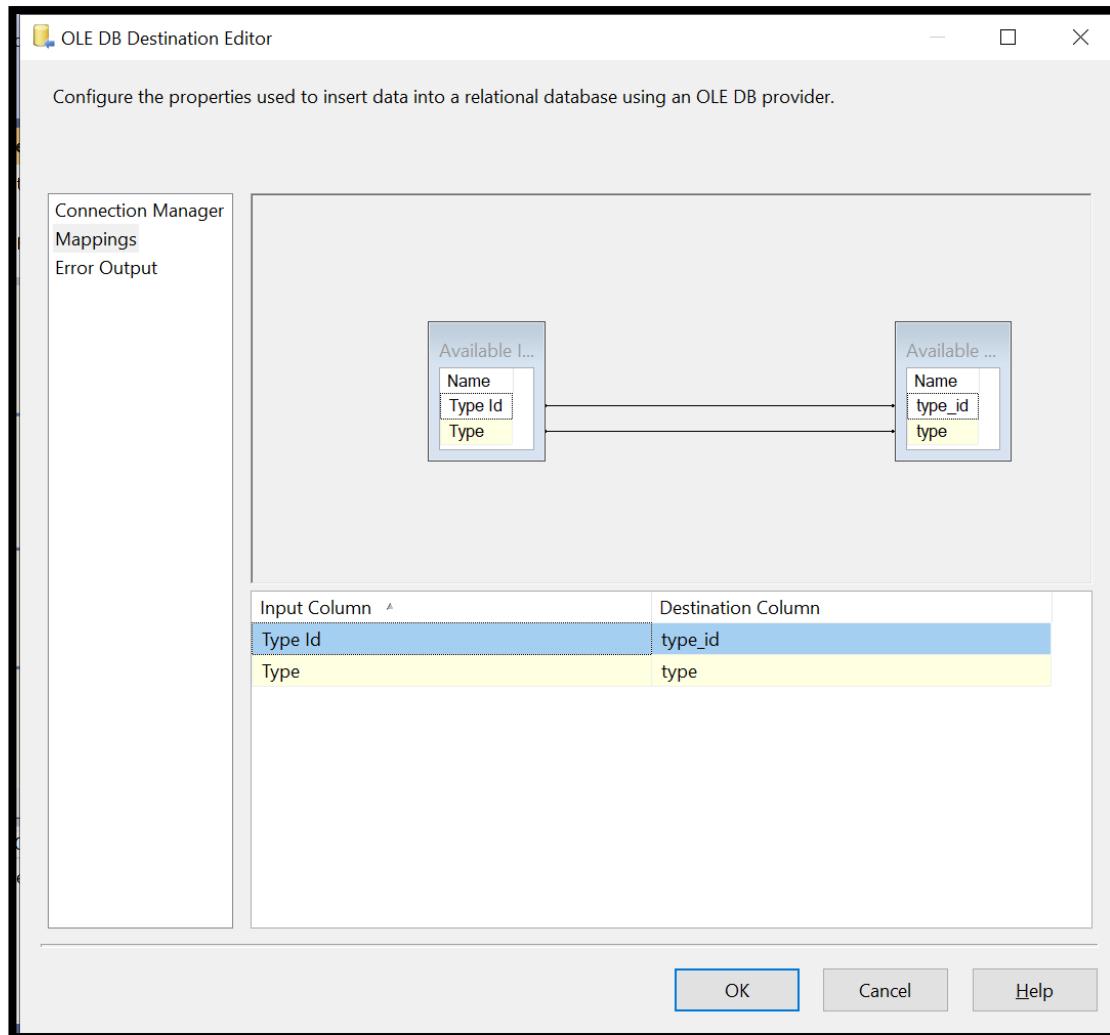
Bước 47: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **Insert Into dim_types**, nối **Sort Types** tới **Insert Into dim_types**.



Bước 48: D-click và **Insert Into dim_types**, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **dim_types**

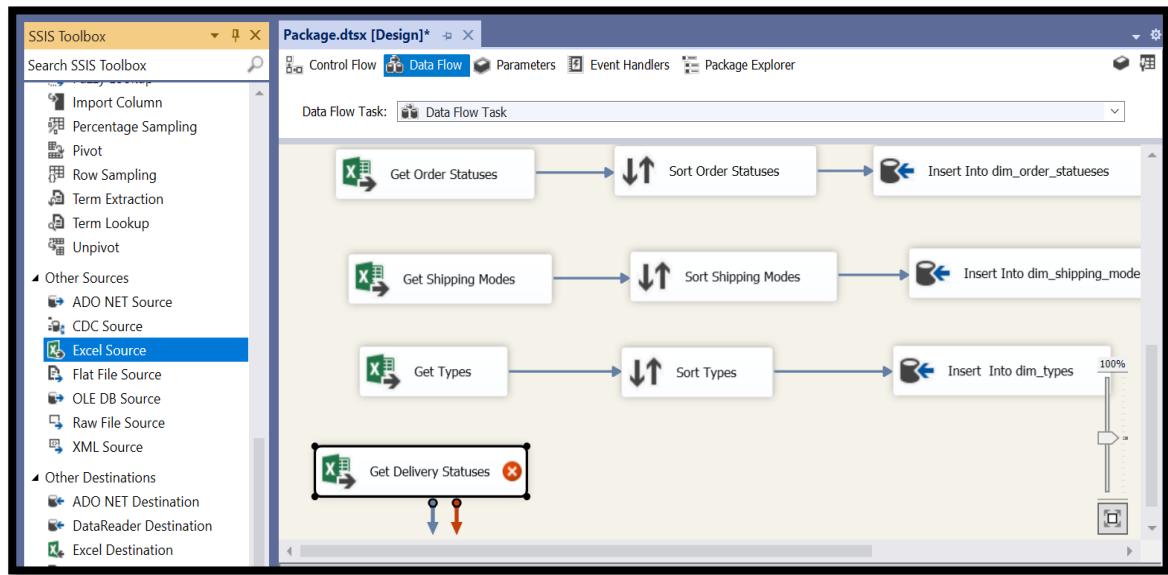


Bước 49: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_types. Chọn OK để kết thúc

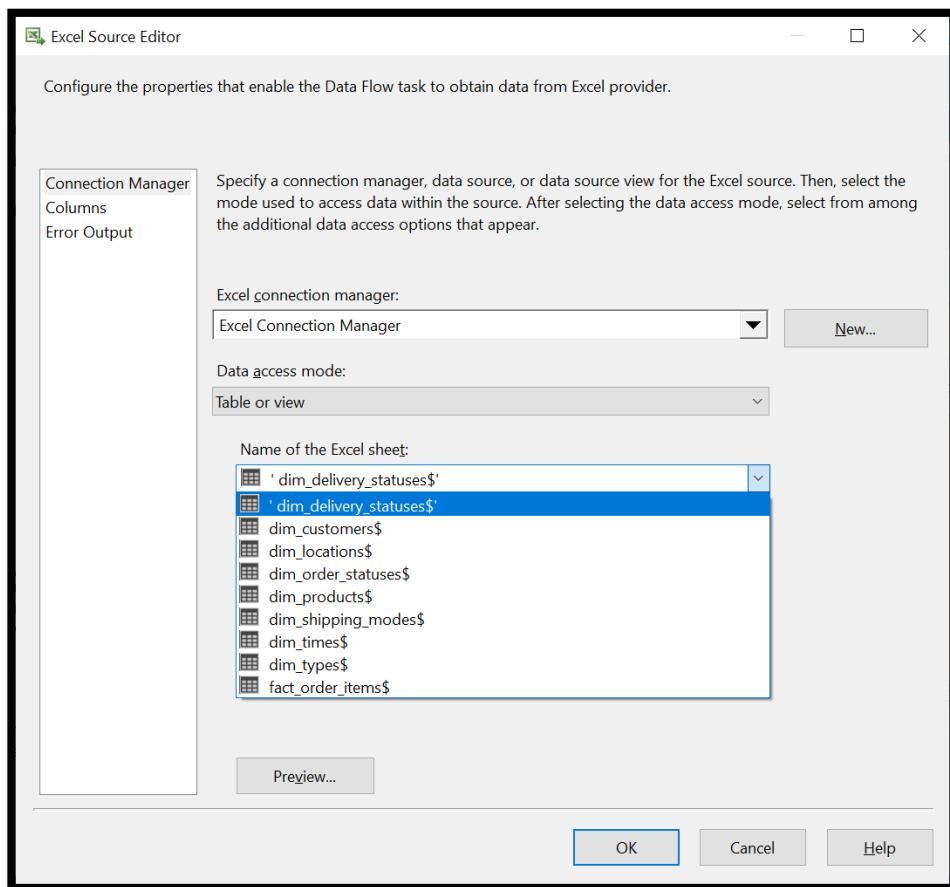


Đỗ dữ liệu vào dim delivery statuses

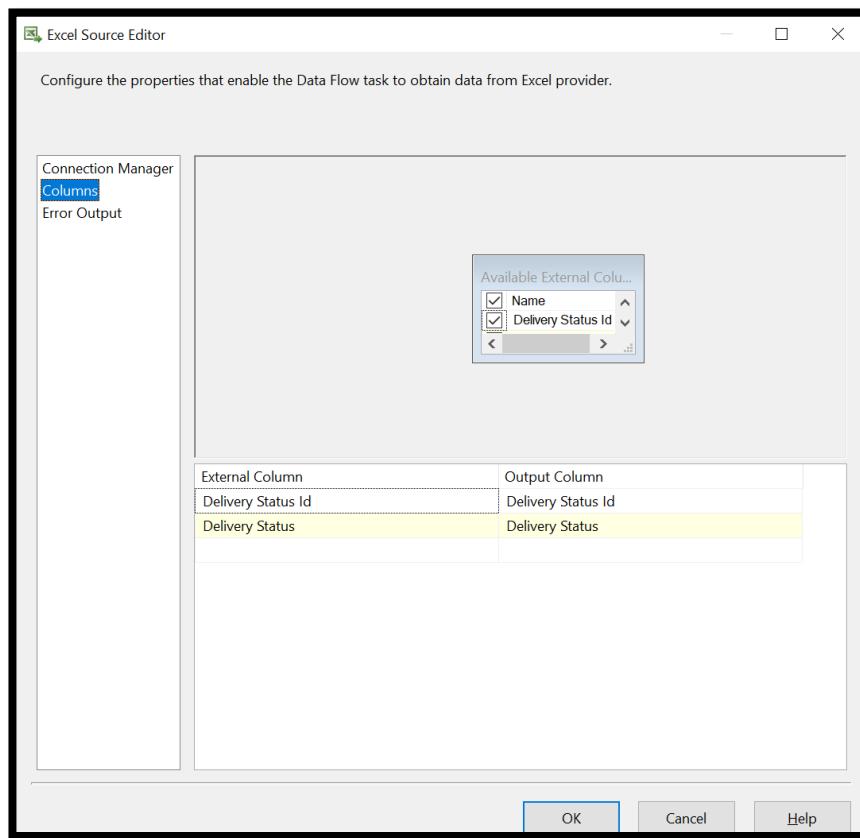
Bước 50: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Delivery Statuses”.



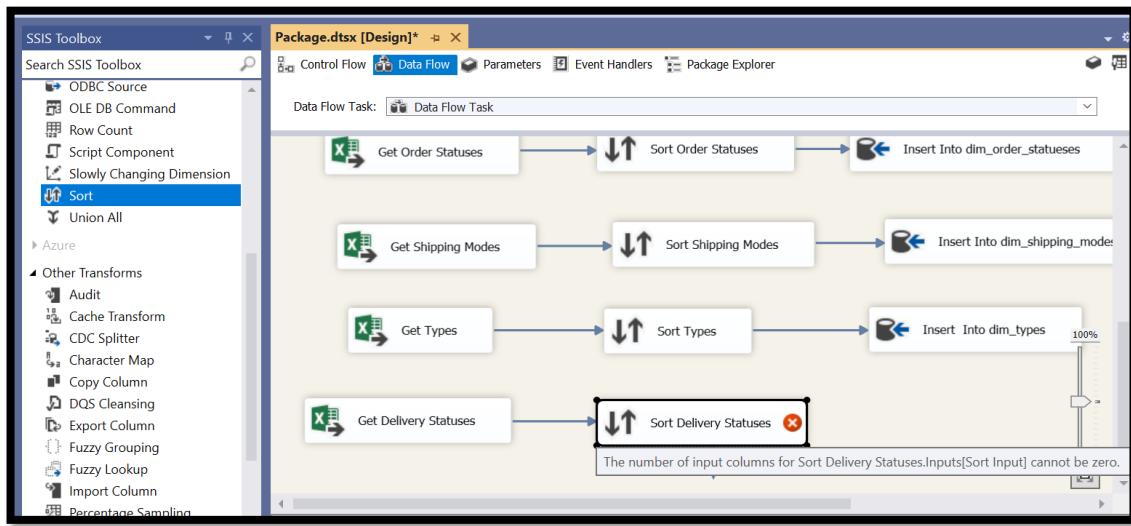
Bước 51: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn dim_delivery_statuses.



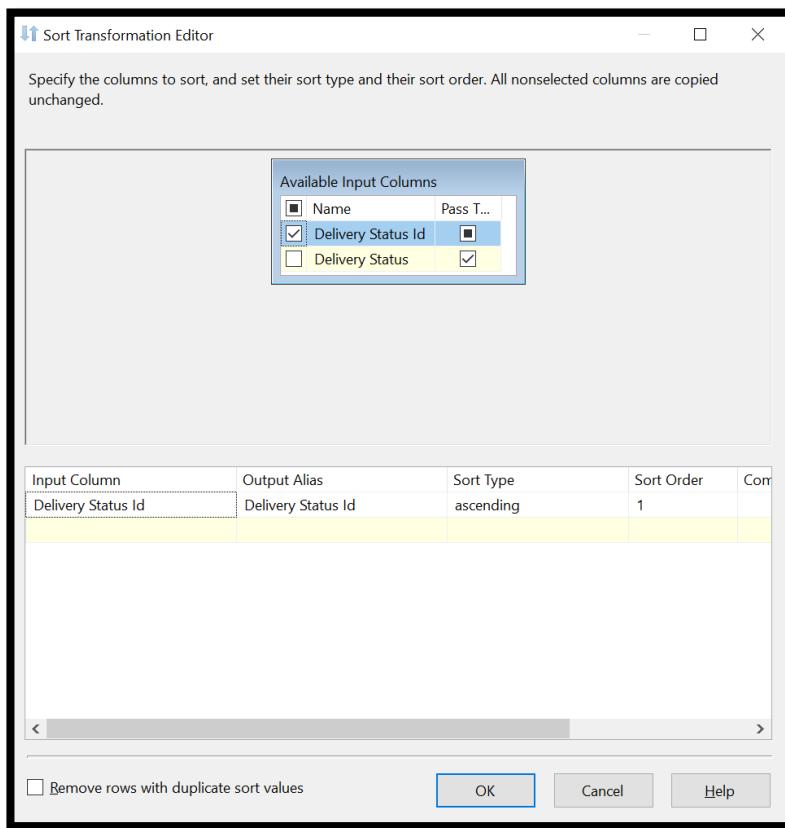
Bước 52: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu types



Bước 53: Kéo thả một Sort mới click và f2 để đặt tên là **Sort Delivery Statuses**, nối Get Delivery Statuses tới Sort Delivery Statuses.



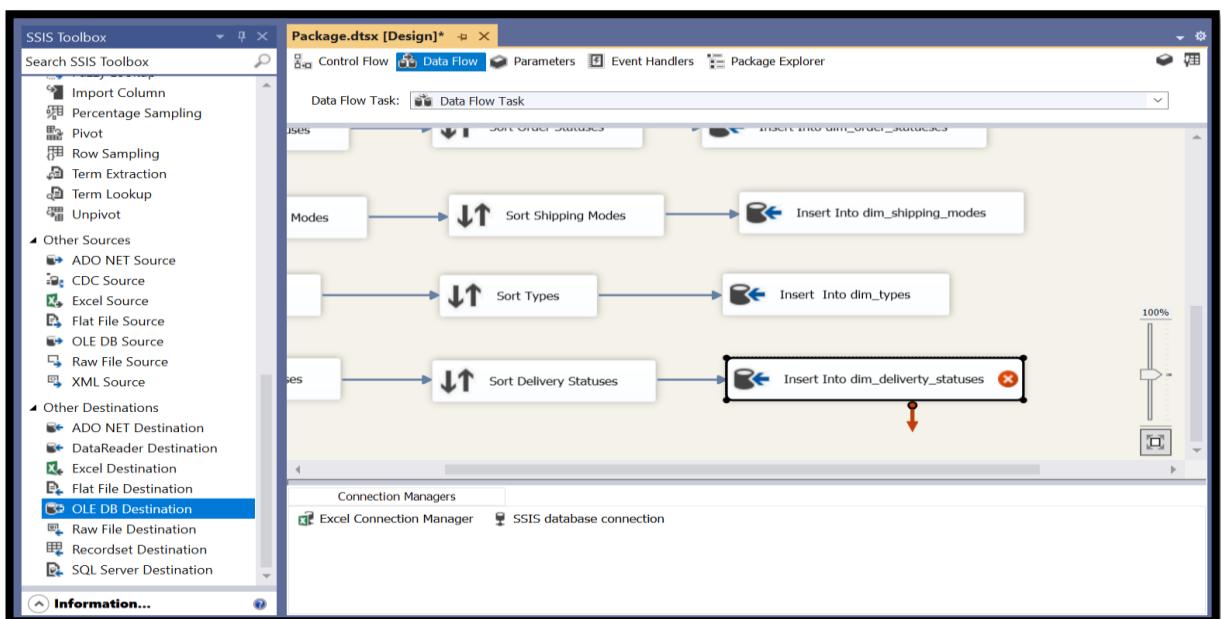
D-click **Sort Shipping Modes** và chọn sort theo **Shipping Mode Id**.



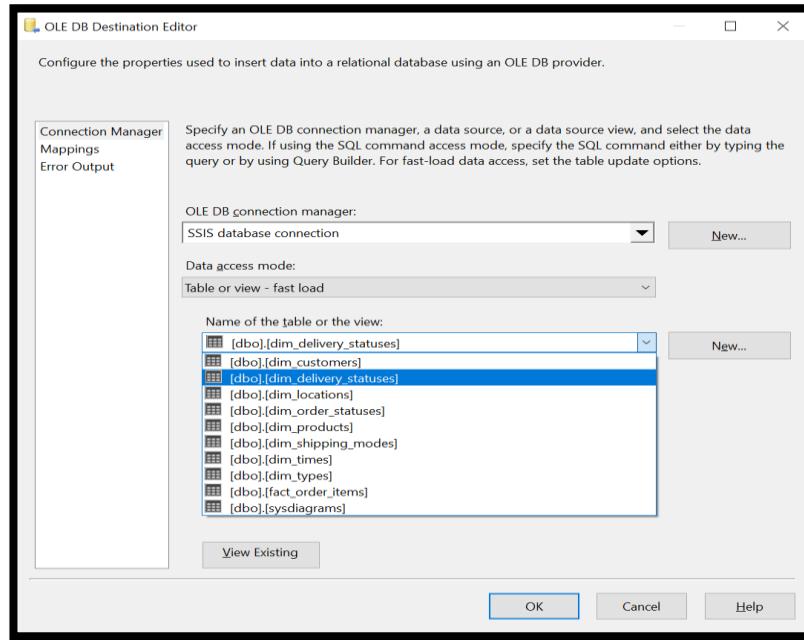
- Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tùy chọn.

Chọn **OK** để kết thúc sort.

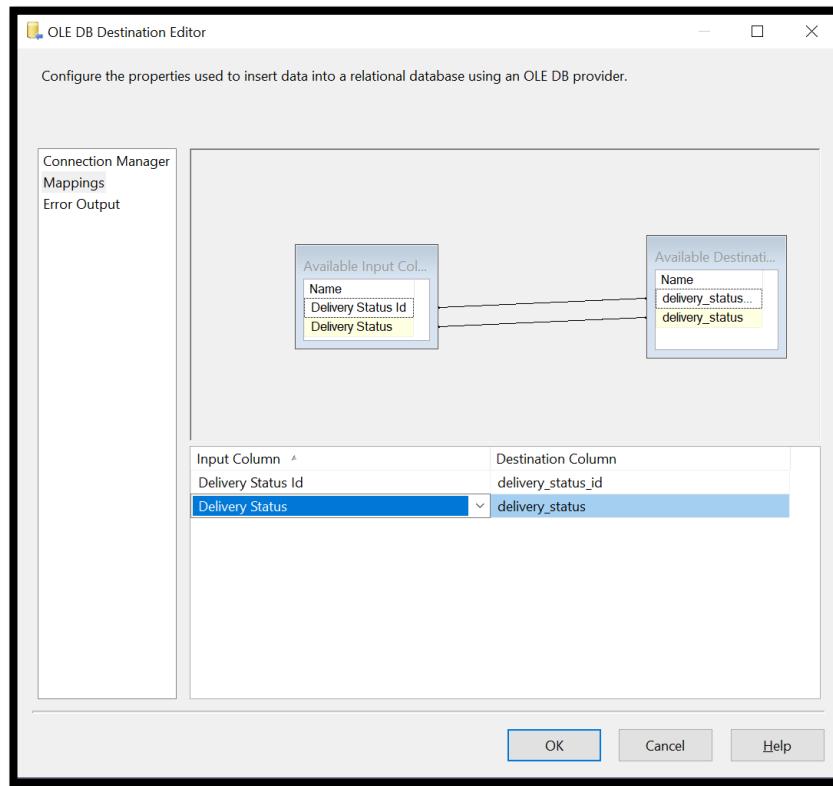
Bước 54: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **Insert Into dim_types**, nối **Sort Types** tới **Insert Into dim_types**.



Bước 55: D-click và **Insert Into dim_delivery_statuses**, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **dim_delivery_statuses**.

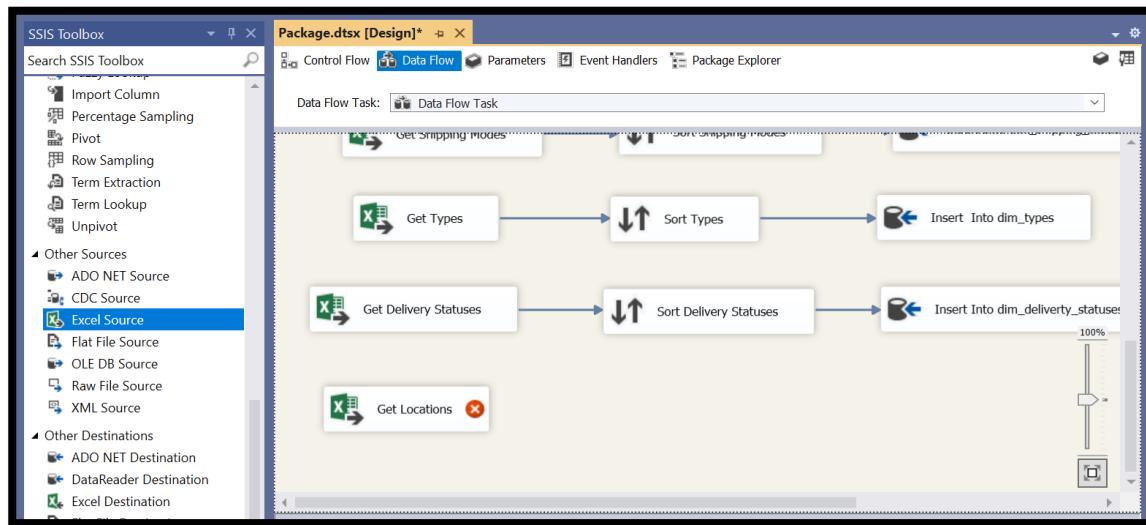


Bước 56: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_delivery_statuses. Chọn OK để kết thúc

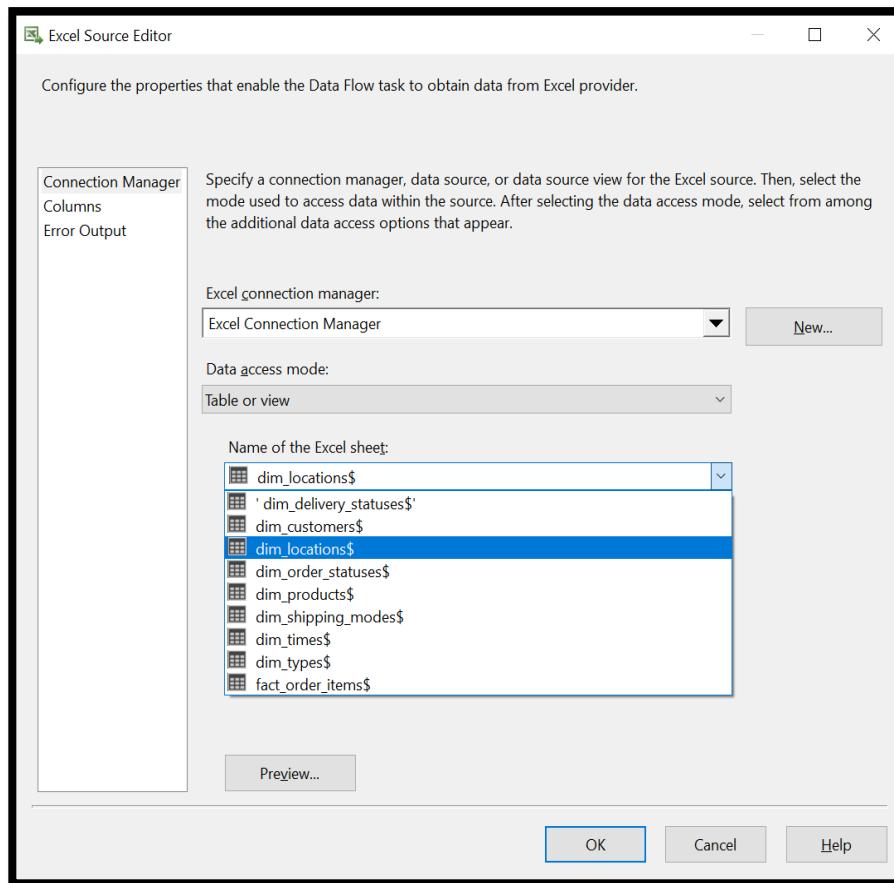


Đỗ dữ liệu vào dim_locations

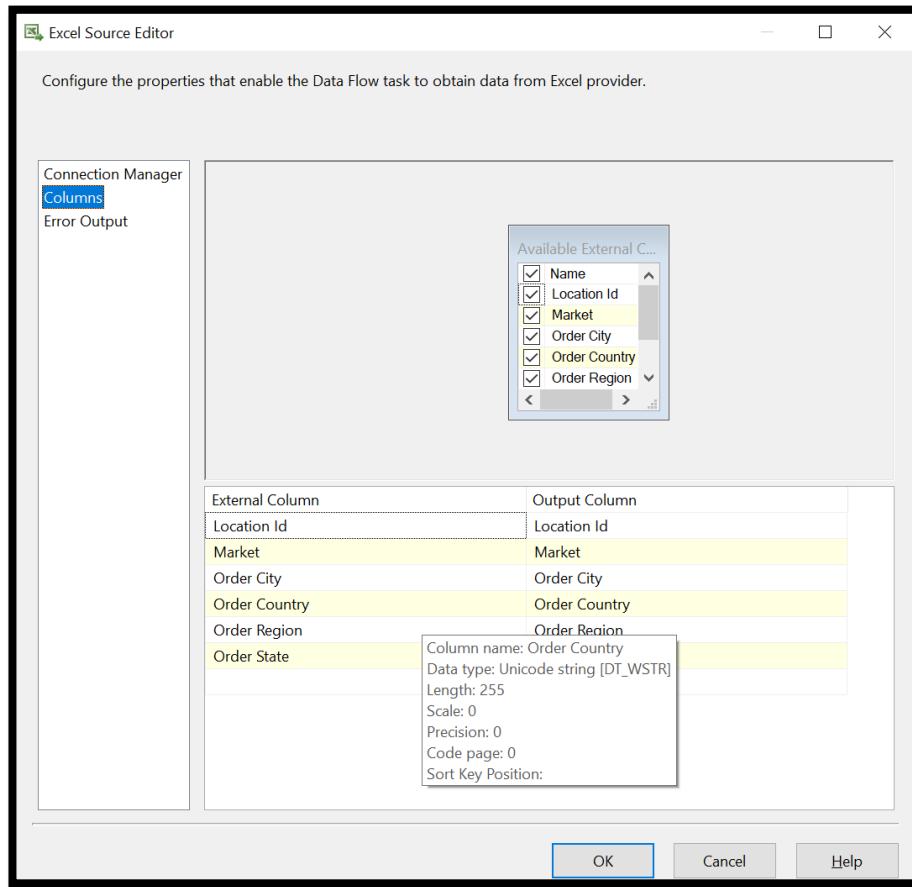
Bước 57: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Locations”.



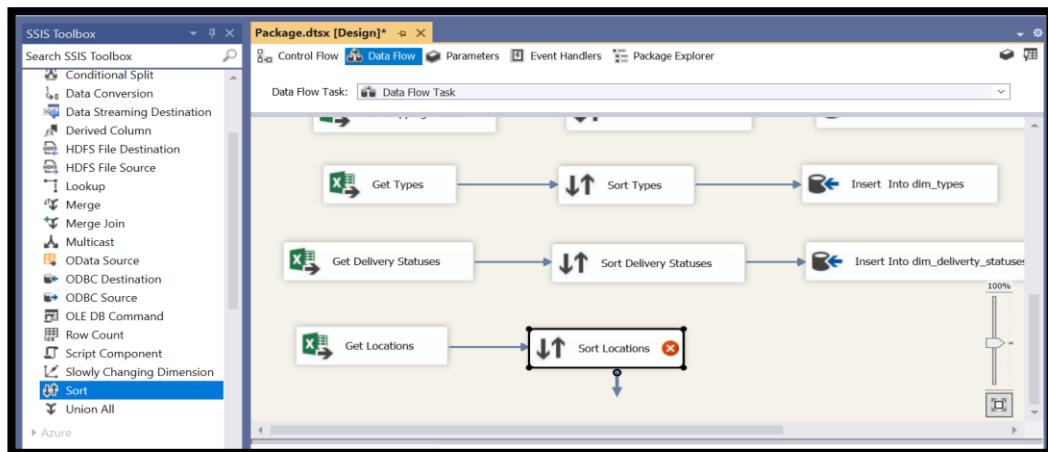
Bước 58: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn dim_locations.



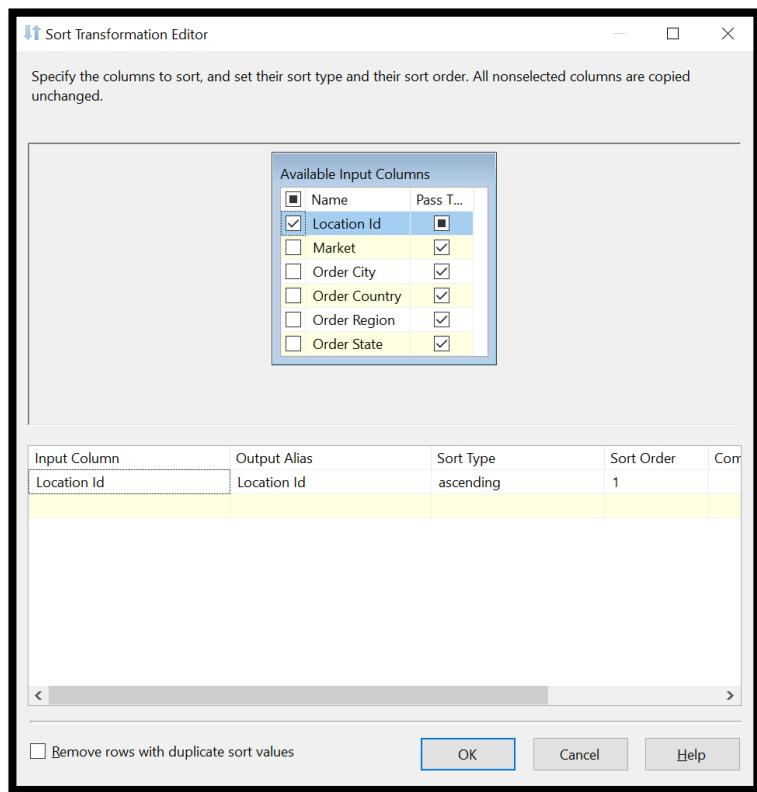
Bước 59: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu locations



Bước 60: Kéo thả một Sort mới click và f2 để đặt tên là **Sort Locations**, nối Get Locations tới Sort Locations.



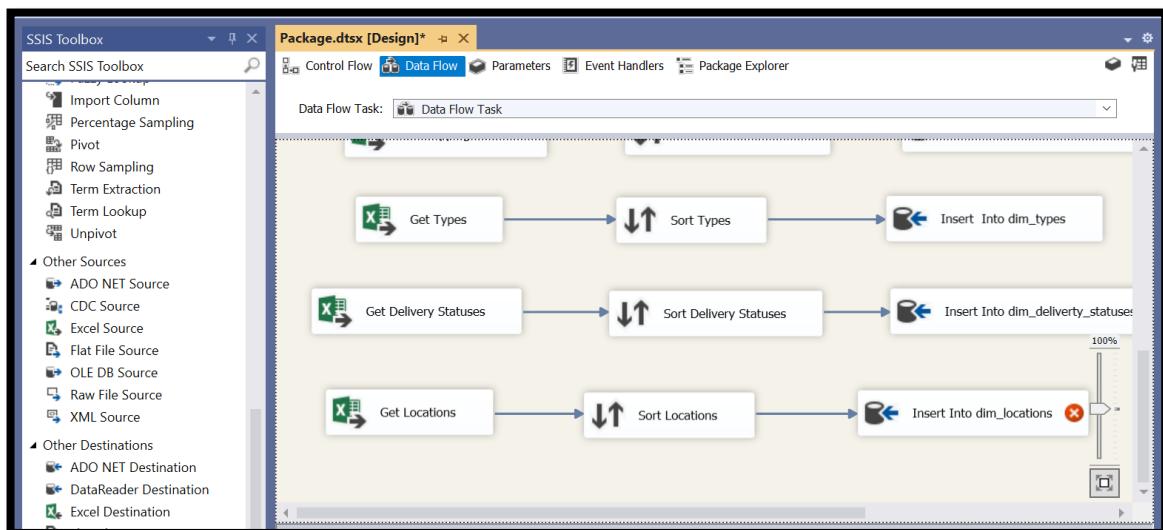
D-click **Sort Locations** và chọn sort theo **Location Id**.



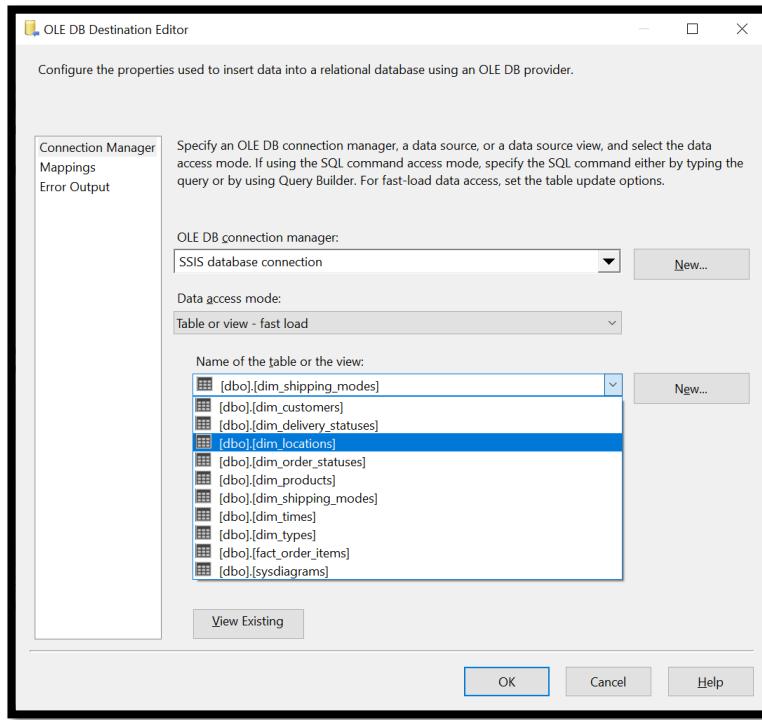
- Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tùy chọn.

Chọn **OK** để kết thúc sort.

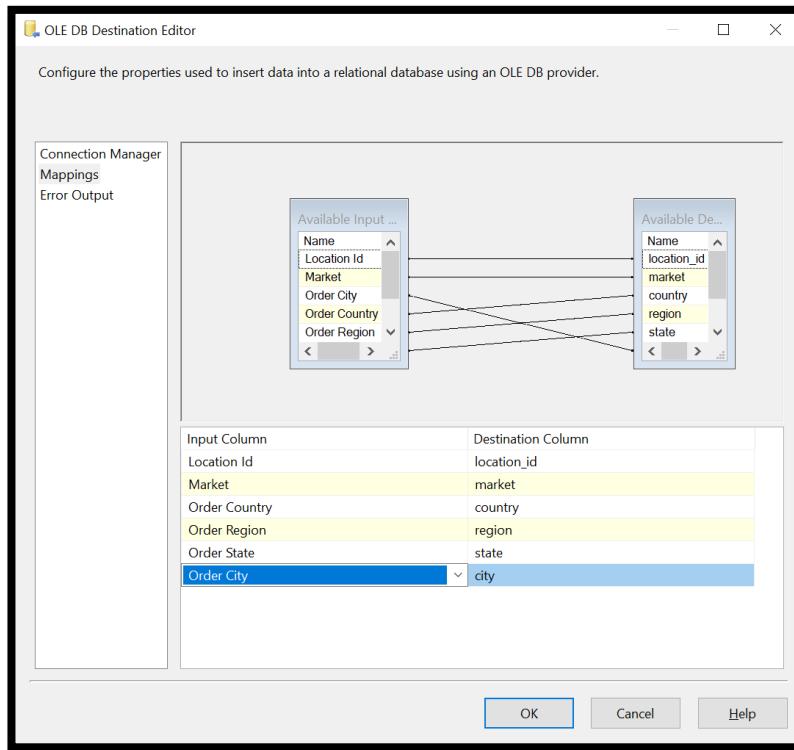
Bước 61: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **n Insert Into dim_locations**, nối **Sort Locations** tới **Insert Into dim_locations**.



Bước 62: D-click và **Insert Into dim_locations**, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **dim_locations**.

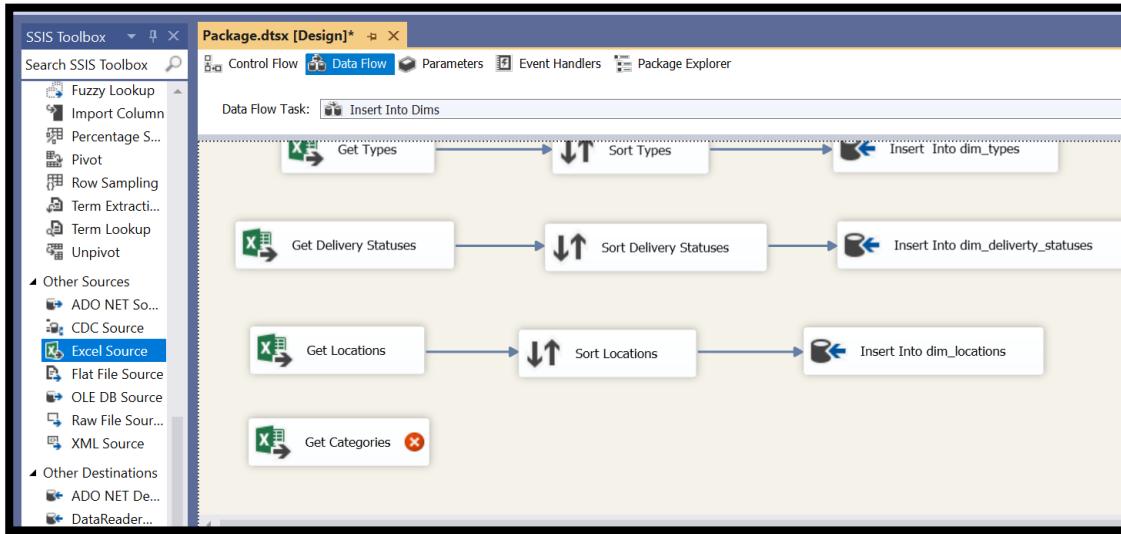


Bước 63: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_delivery_statuses. Chọn OK để kết thúc

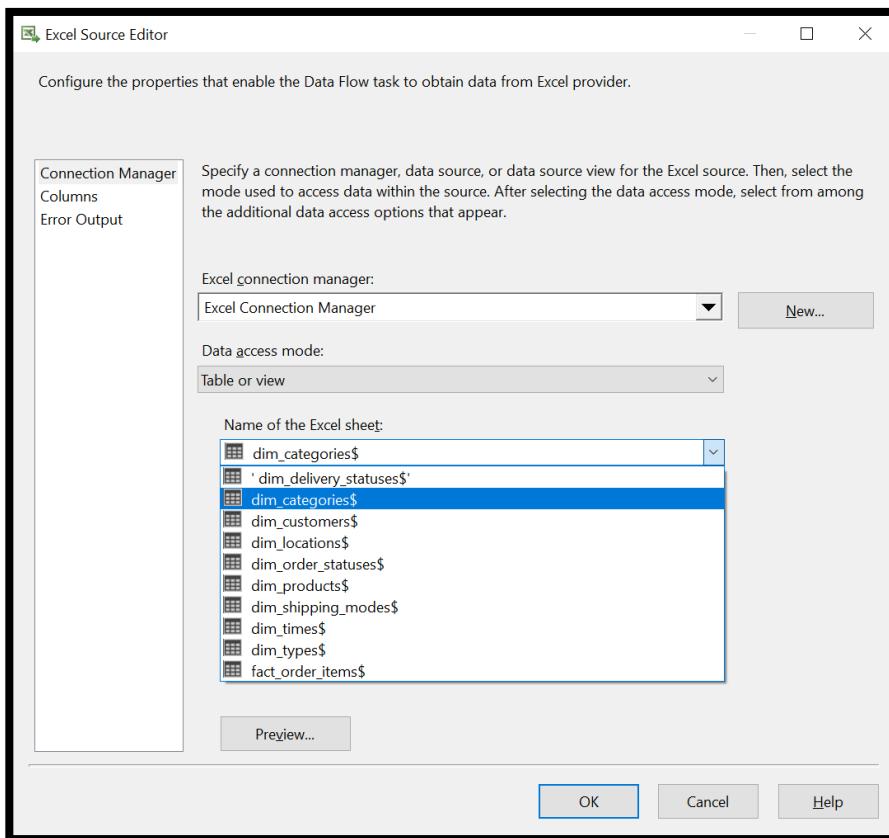


Đỗ dữ liệu vào dim_categories

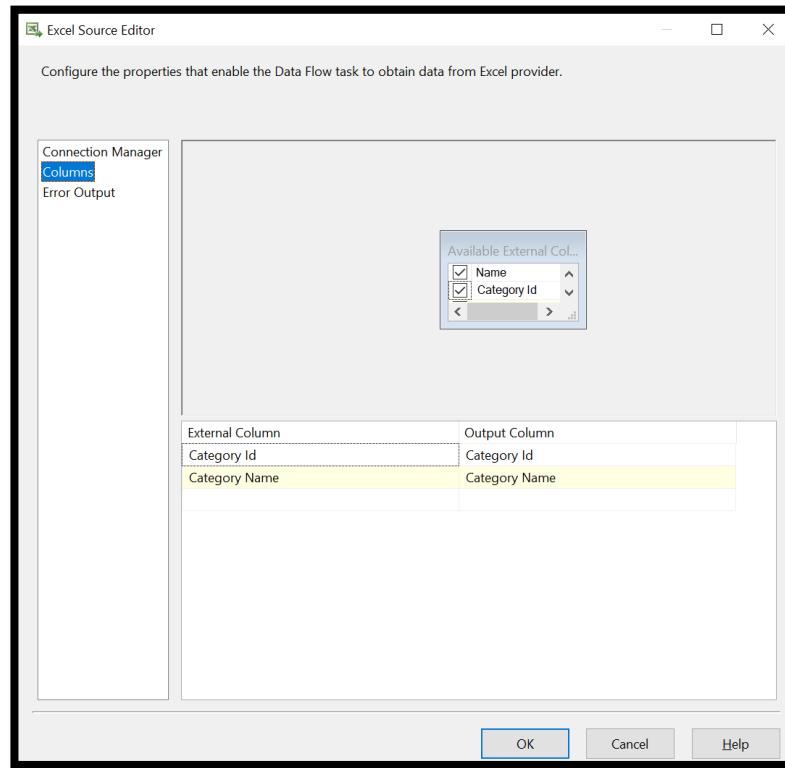
Bước 64: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Categories”.



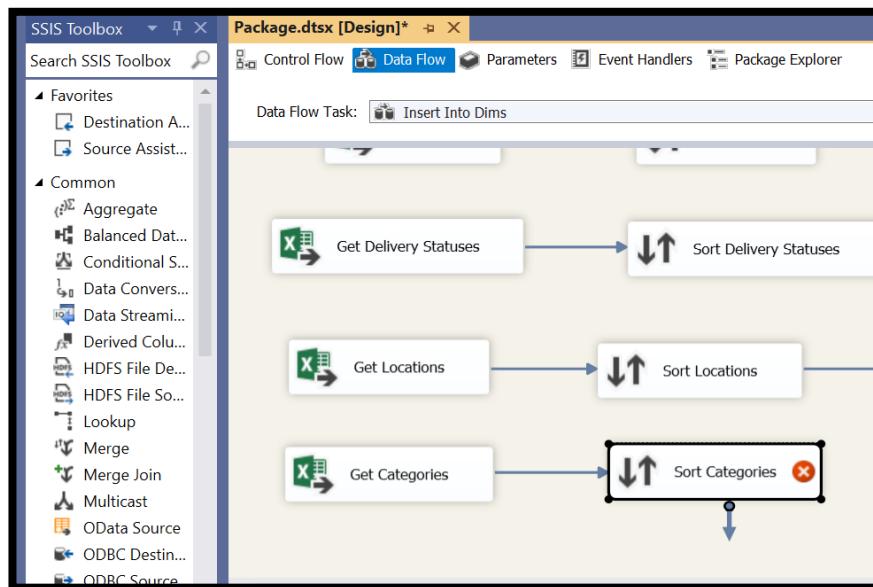
Bước 65: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn dim_categories.



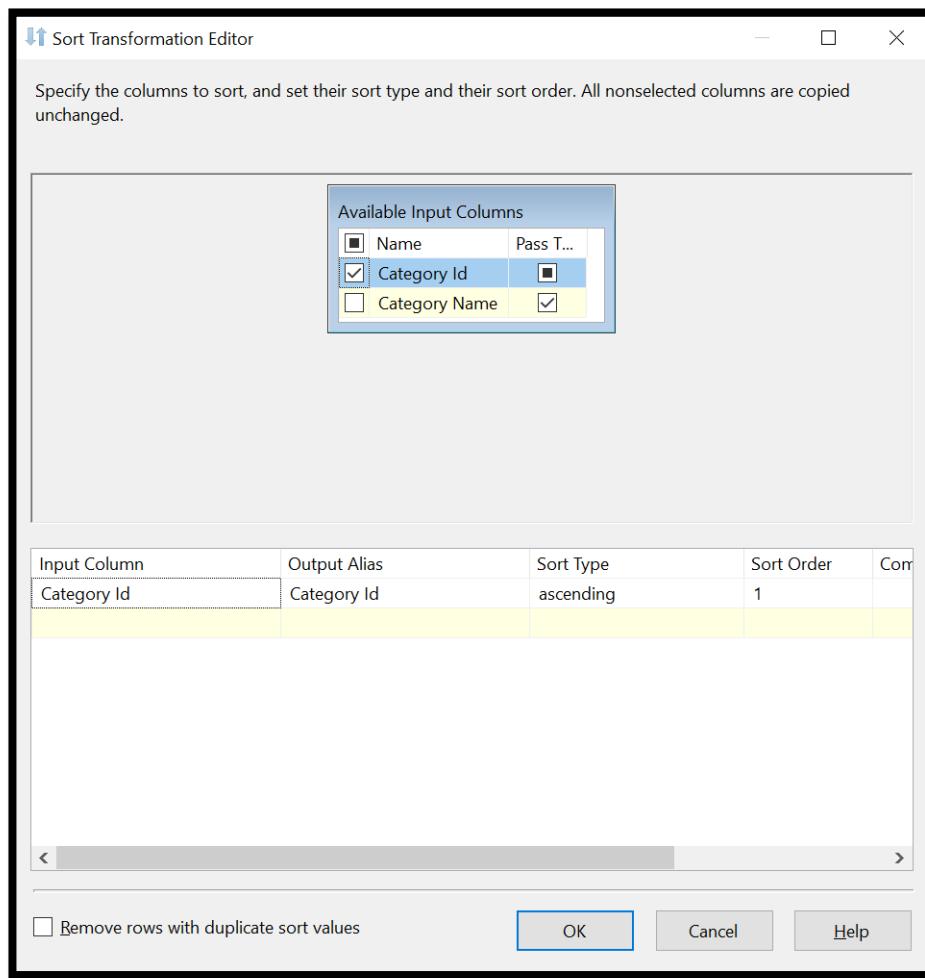
Bước 66: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu categories



Bước 67: Kéo thả một Sort mới click và f2 để đặt tên là **Sort Categories**, nối Get Categories tới Sort Categories.



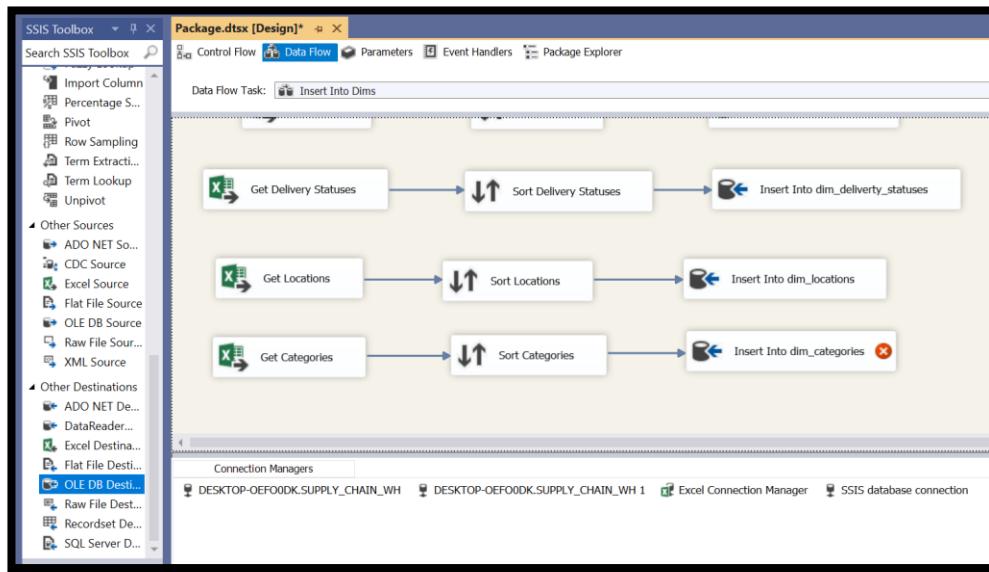
D-click **Sort Products** và chọn sort theo **Products Id**.



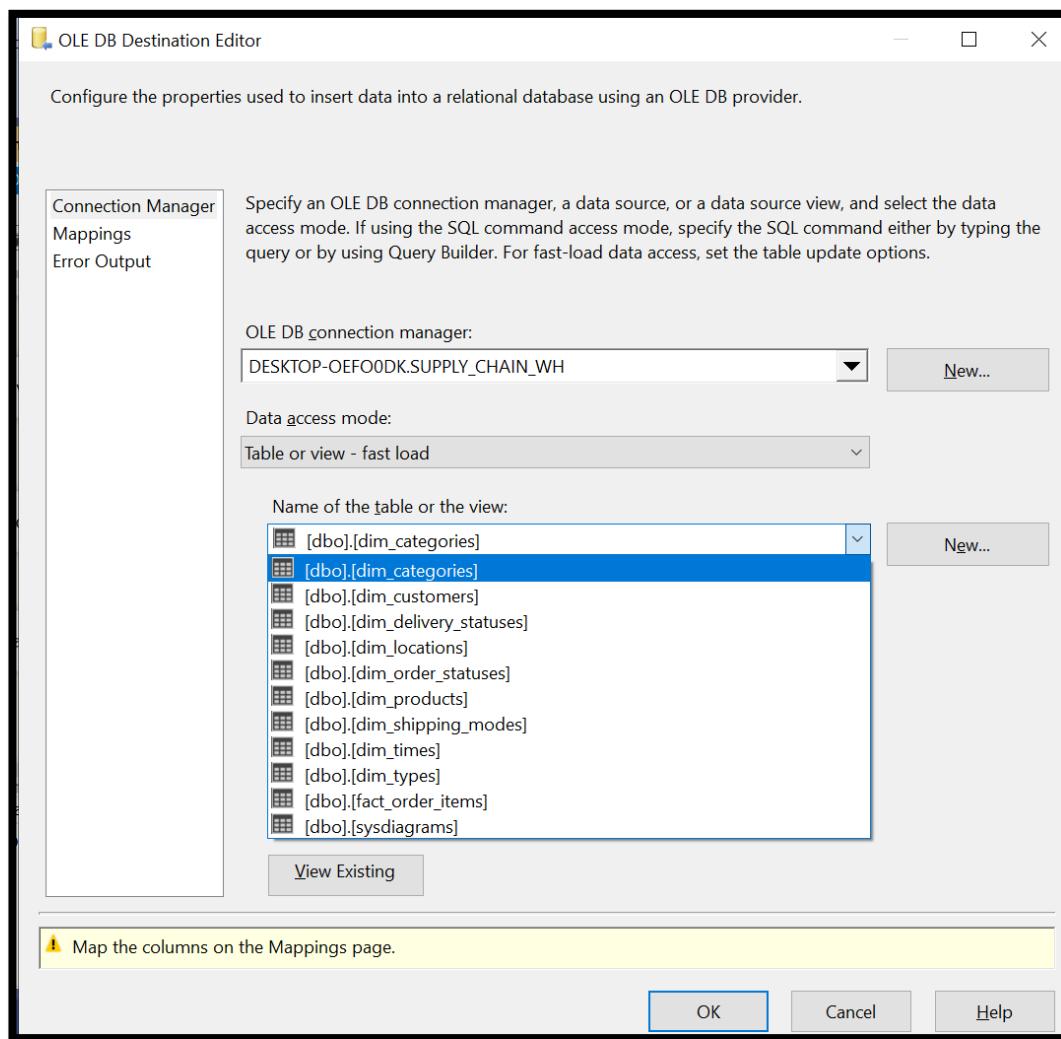
- Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tùy chọn.

Chọn **OK** để kết thúc sort.

Bước 68: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **Insert Into dim_categories**, nối **Sort Categories** tới **Insert Into dim_categories**.

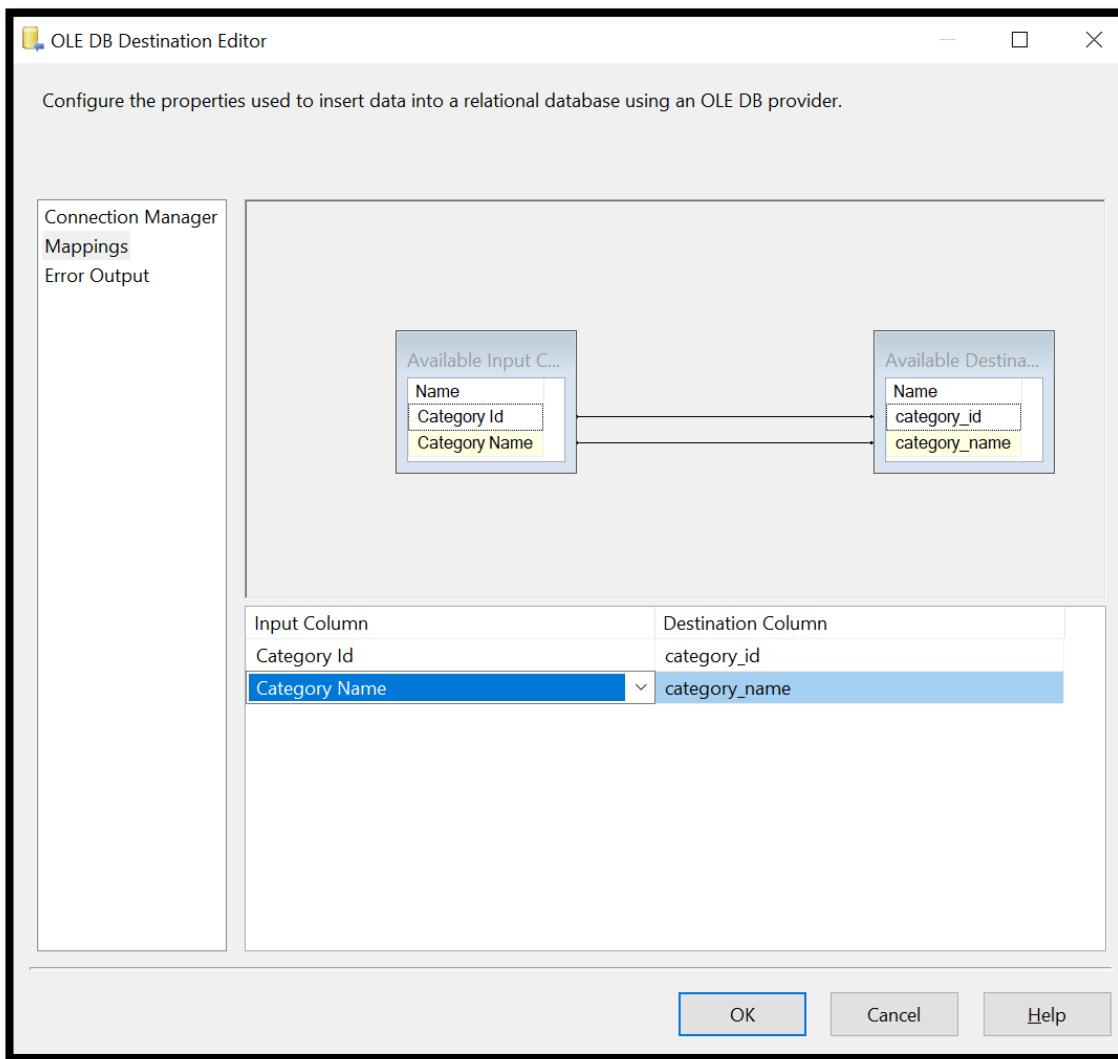


Bước 69: D-click và Insert Into dim-categories, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn dim- categories.



Bước 70: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_categories.

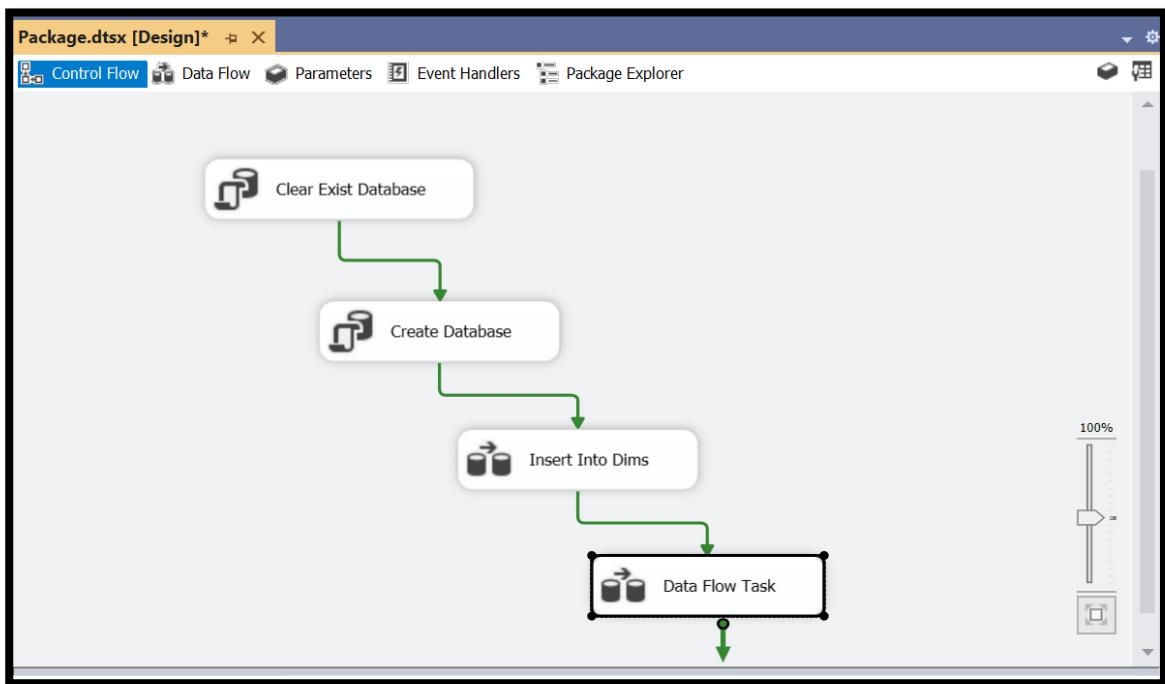
Chọn **OK** để kết thúc



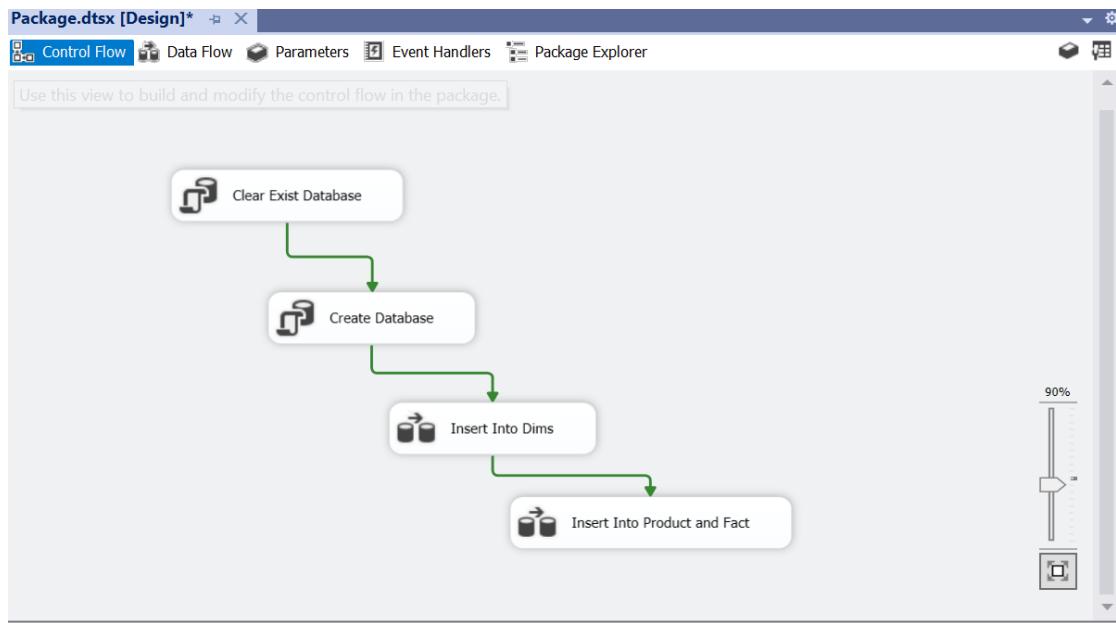
Tạo Data Flow Task để đổ data vào trong fact

Bước 1: Tại SSIS toolbox kéo thả **Data Flow Task** vào Control Flow panel.

Sau đó nối “Insert Into Dims” vào Execute SQL Task vừa tạo.



Bước 2: Click vào và chọn f2 để đổi tên task thành “Insert Into Product and Fact”.

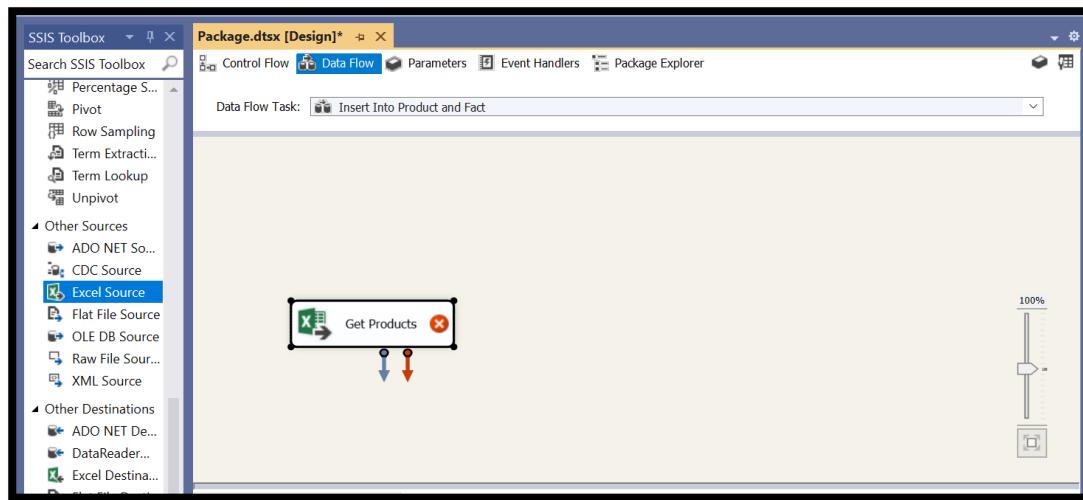


Bước 3: D-click vào **Insert Into Product and Fact** để mở **Data Flow** panel
Đỗ dữ liệu vào Product

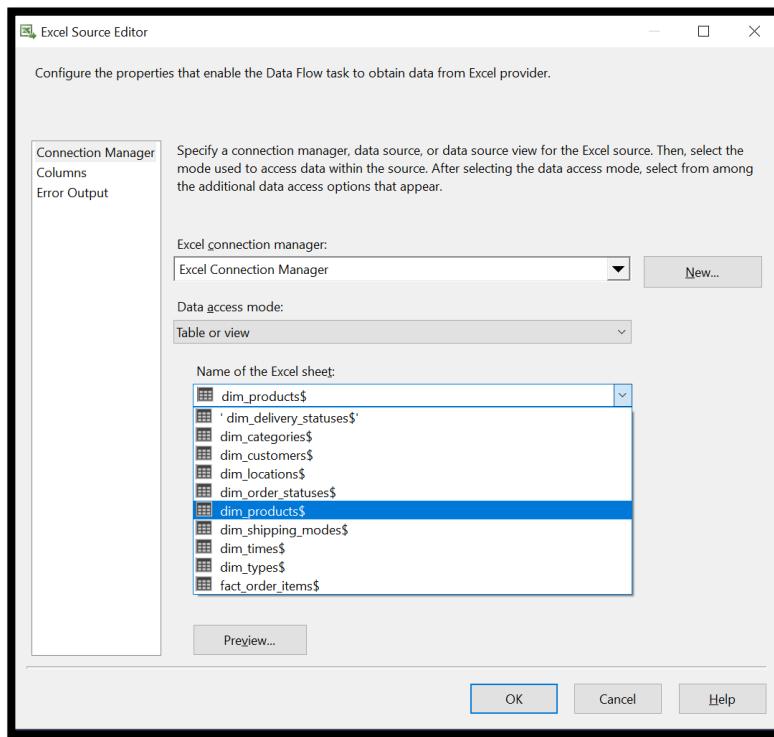
Vì Products sẽ có các foreign key tới category, vì vậy ta cần đỗ tuần tự vào categories rồi mới qua 1 task khác để đỗ vào products

Đỗ dữ liệu vào dim categories

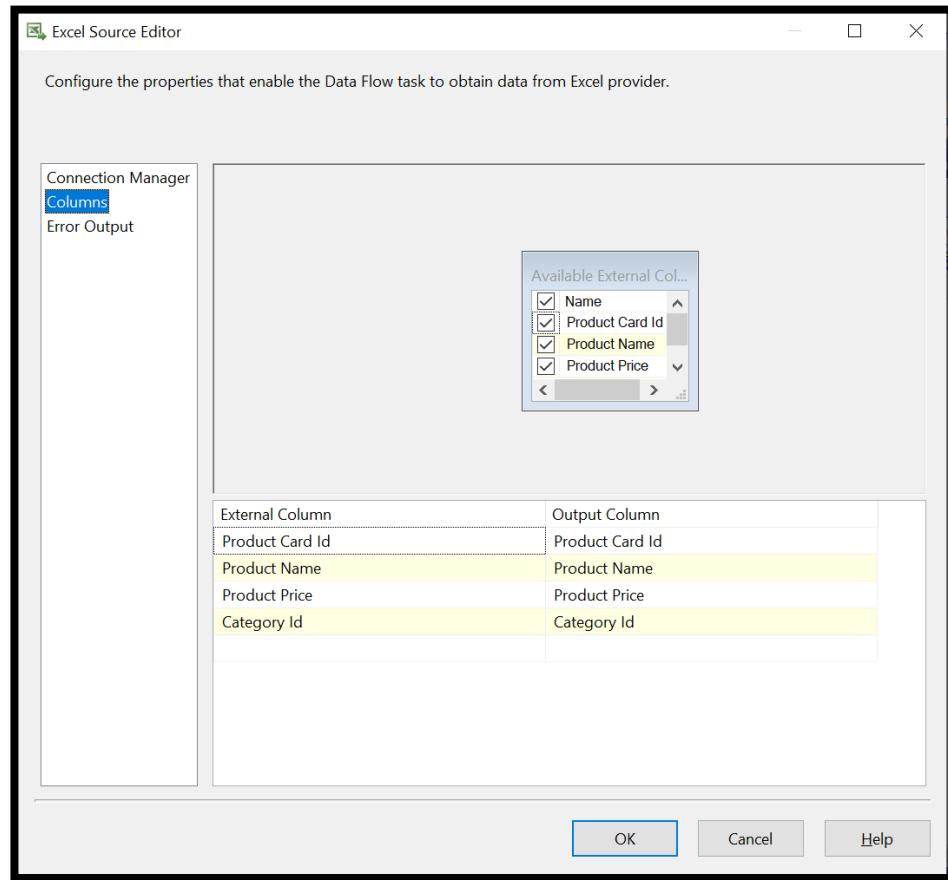
Bước 4: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get Product”.



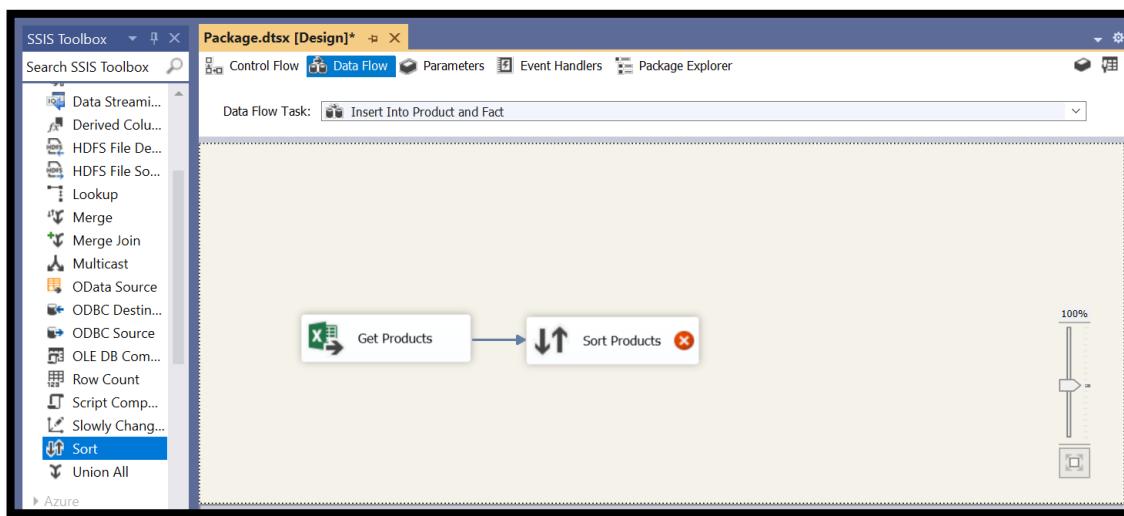
Bước 65: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn dim_products.



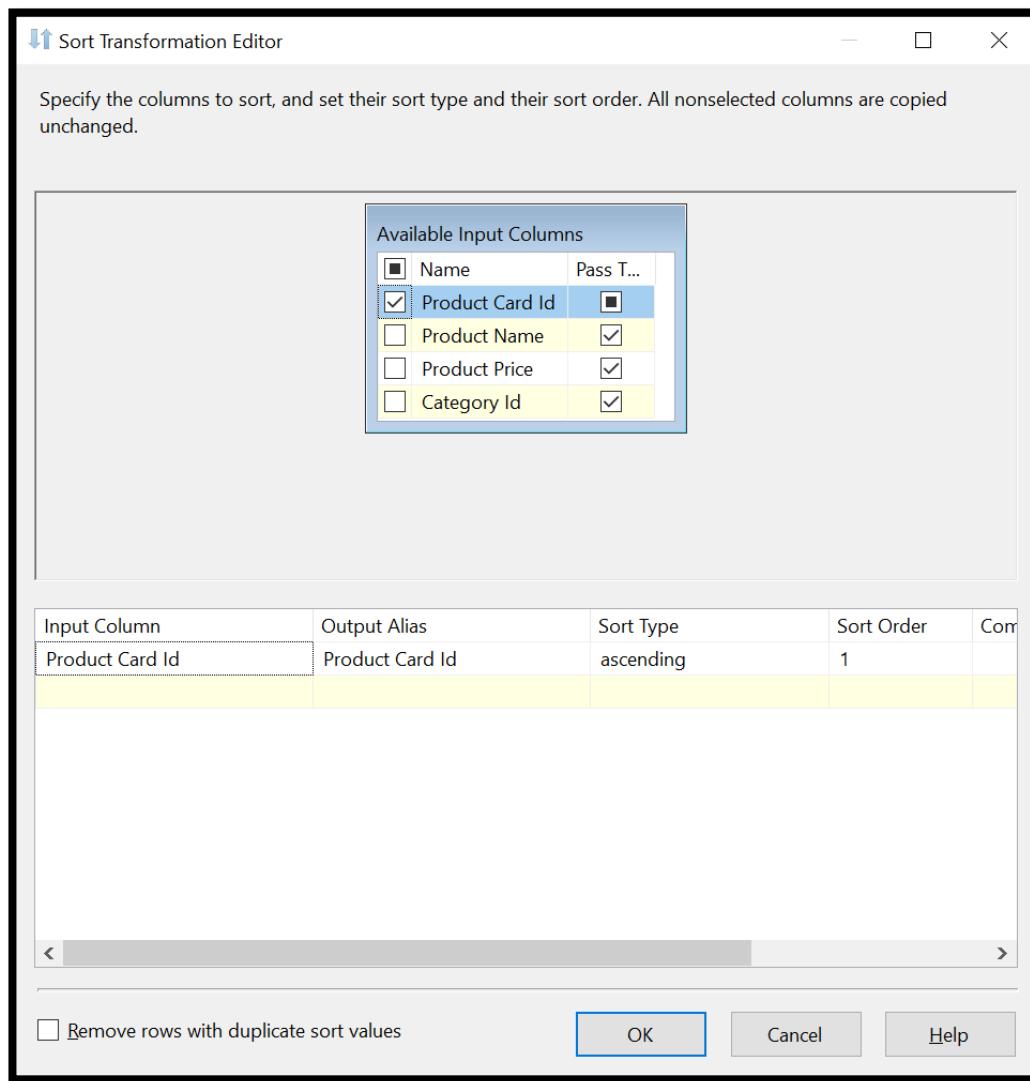
Bước 66: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu products.



Bước 67: Kéo thả một Sort mới click và f2 để đặt tên là **Sort Products**, nối Get Products tới Sort Products.



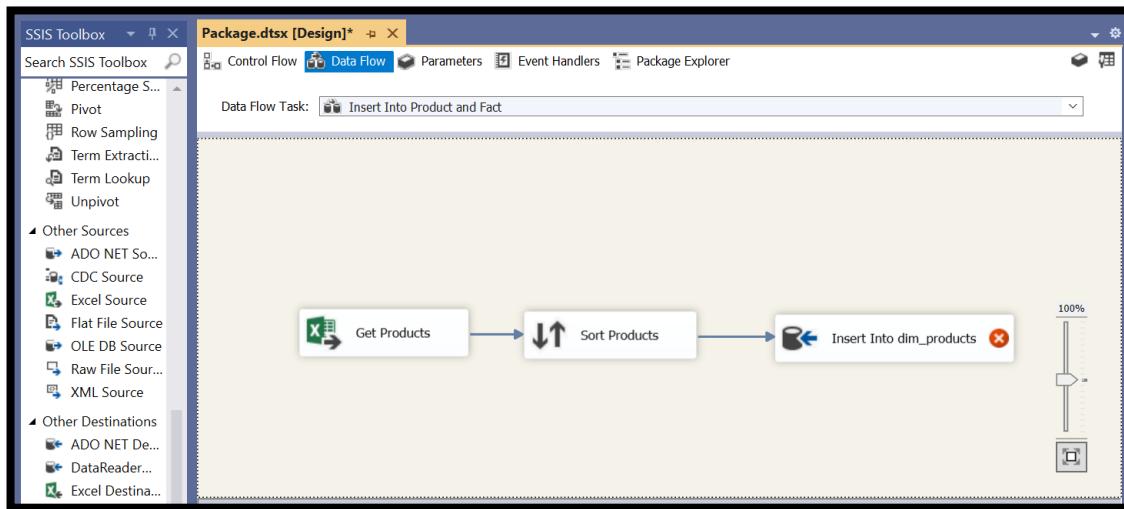
D-click **Sort Products** và chọn sort theo **Products Id**.



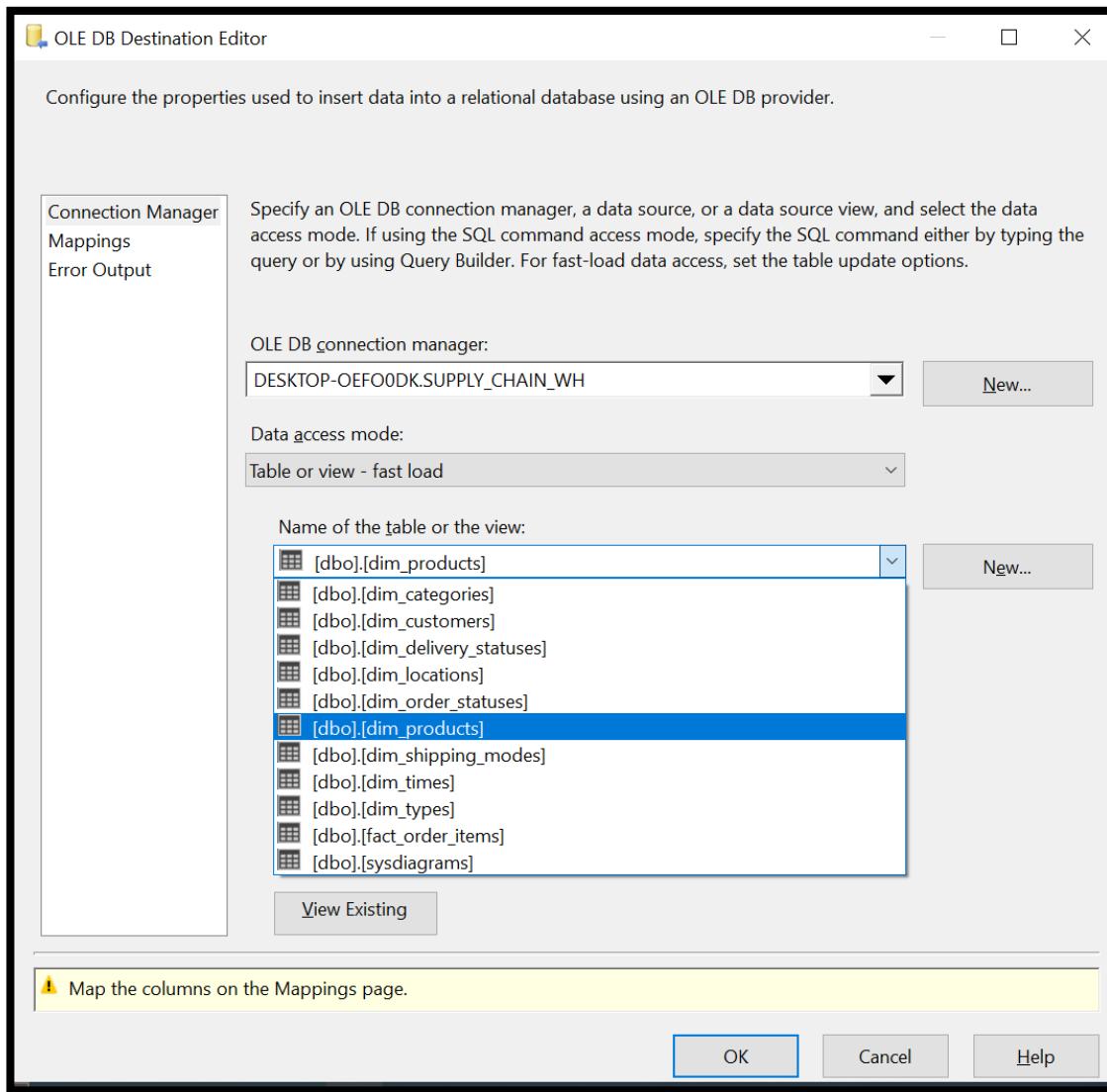
- Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tùy chọn.

Chọn **OK** để kết thúc sort.

Bước 68: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **Insert Into dim_products**, nối **Sort Categories** tới **Insert Into dim_products**.

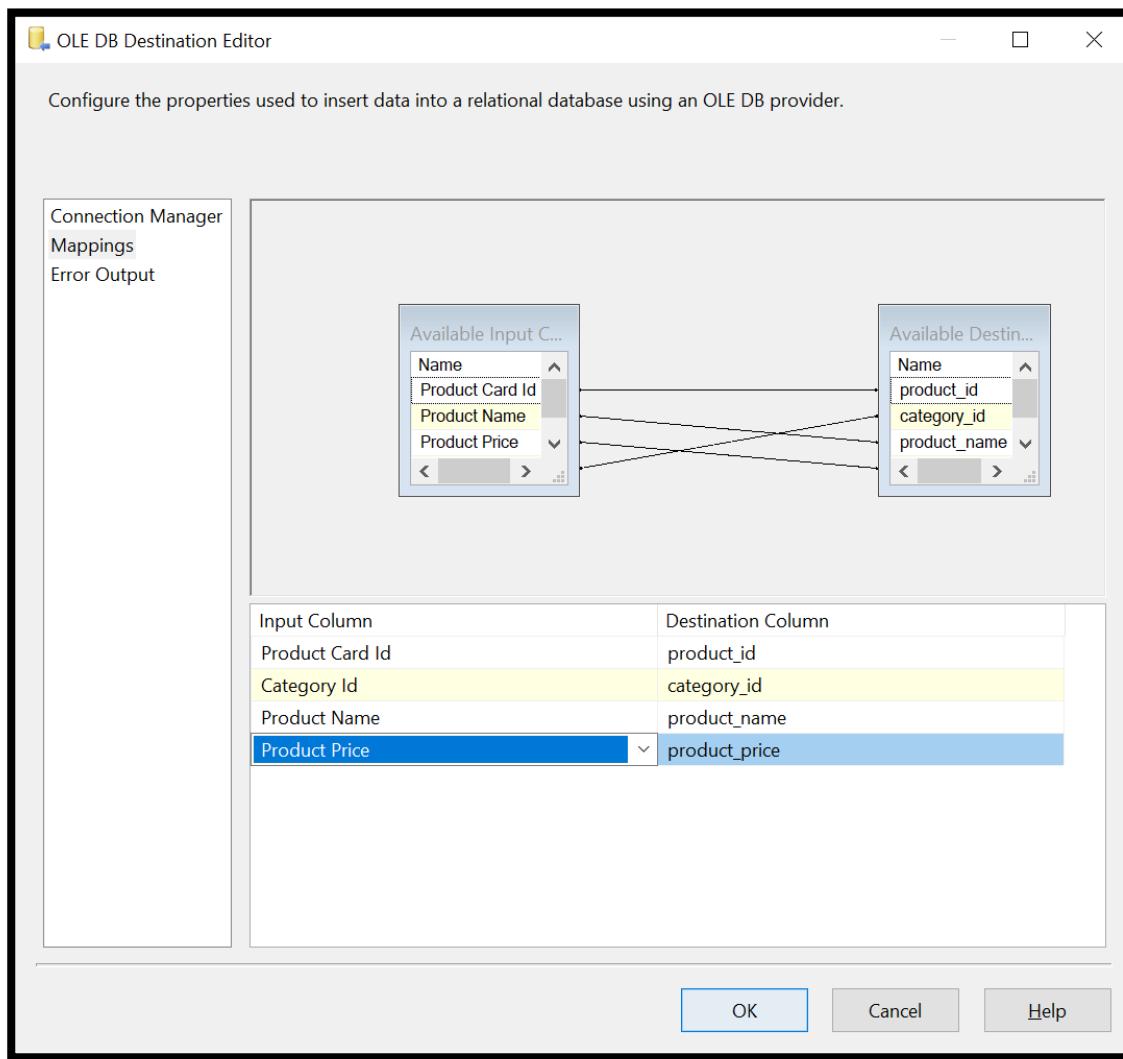


Bước 69: D-click và Insert Into dim-categories, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **dim- categories**.



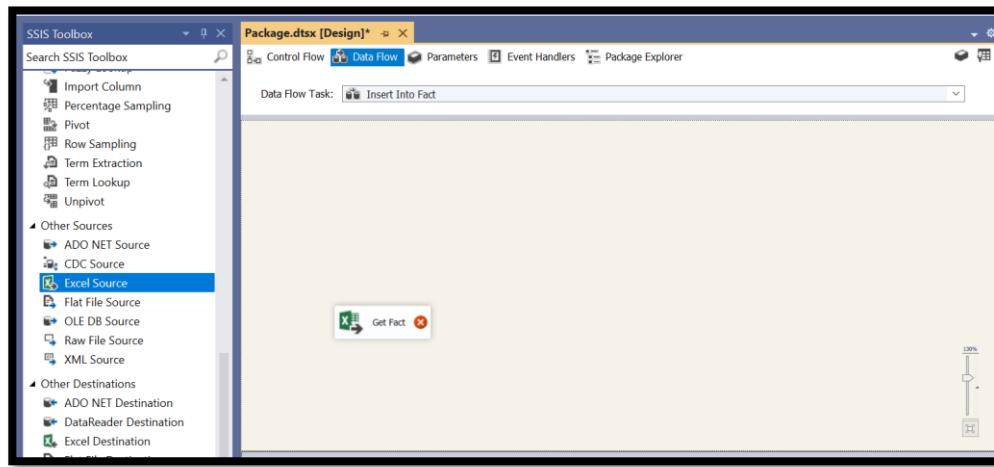
Bước 70: Chọn Mappings để kiểm tra việc ánh xạ các cột vào trong dim_categories.

Chọn **OK** để kết thúc

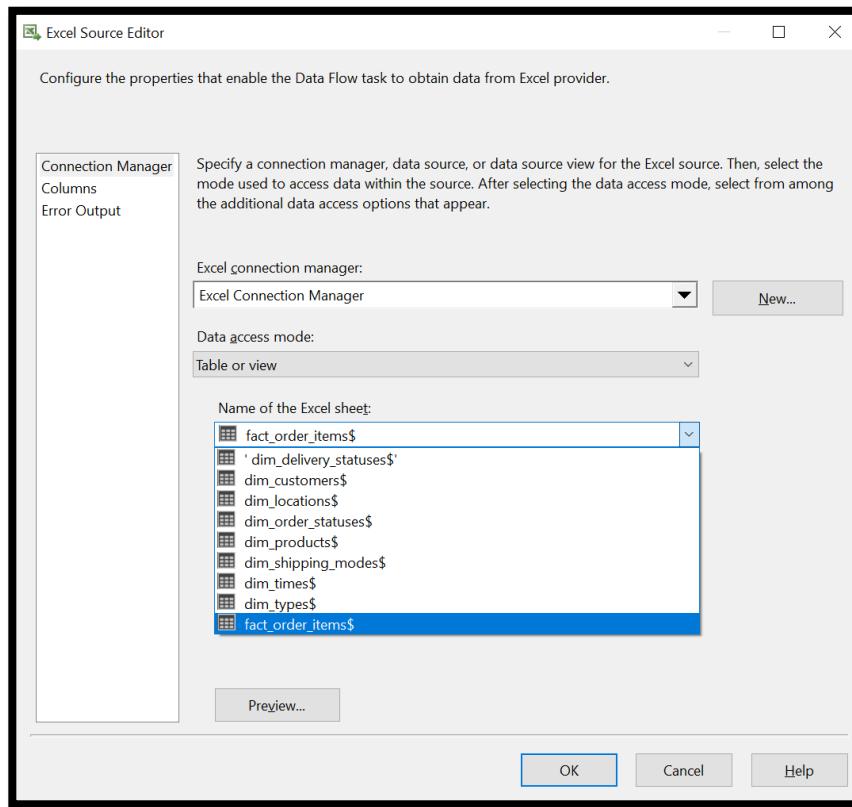


Đỗ dữ liệu vào fact

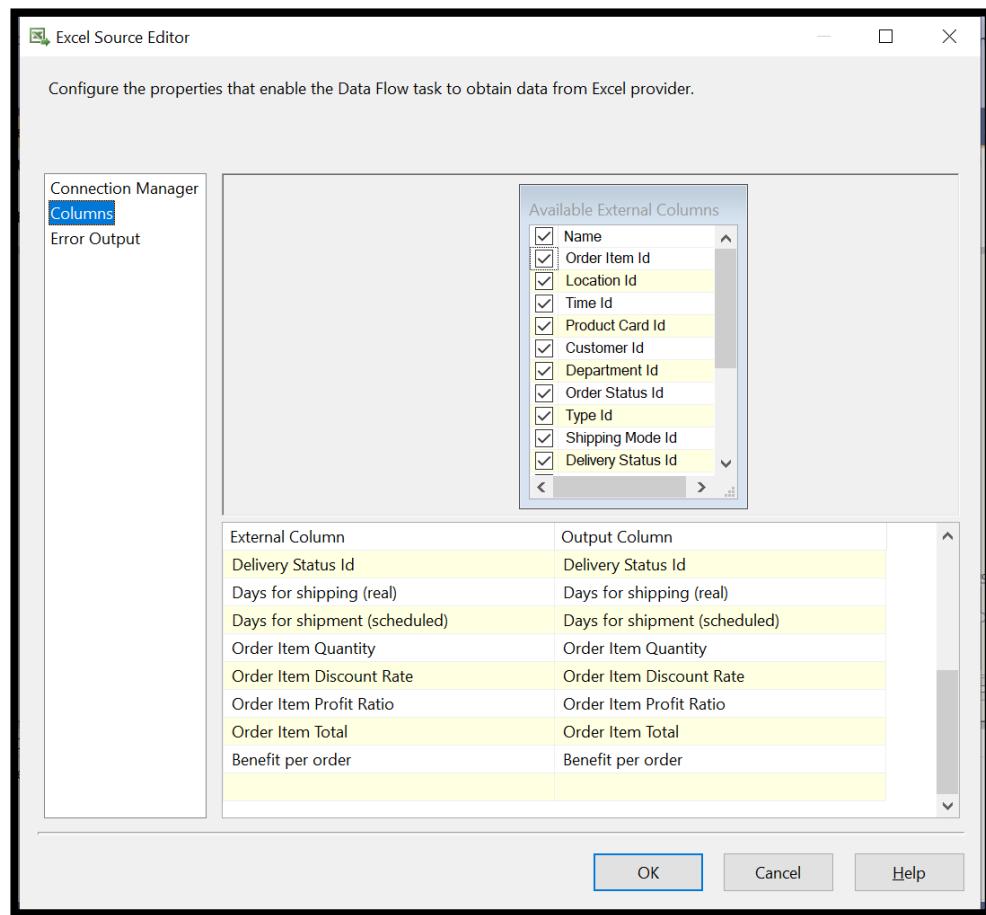
Bước 4: Trong Data Flow, kéo thả Excel Source để lấy dữ liệu gốc từ file .xlss đã preprocess trước đó. Click và f2 để đổi tên thành “Get fact”.



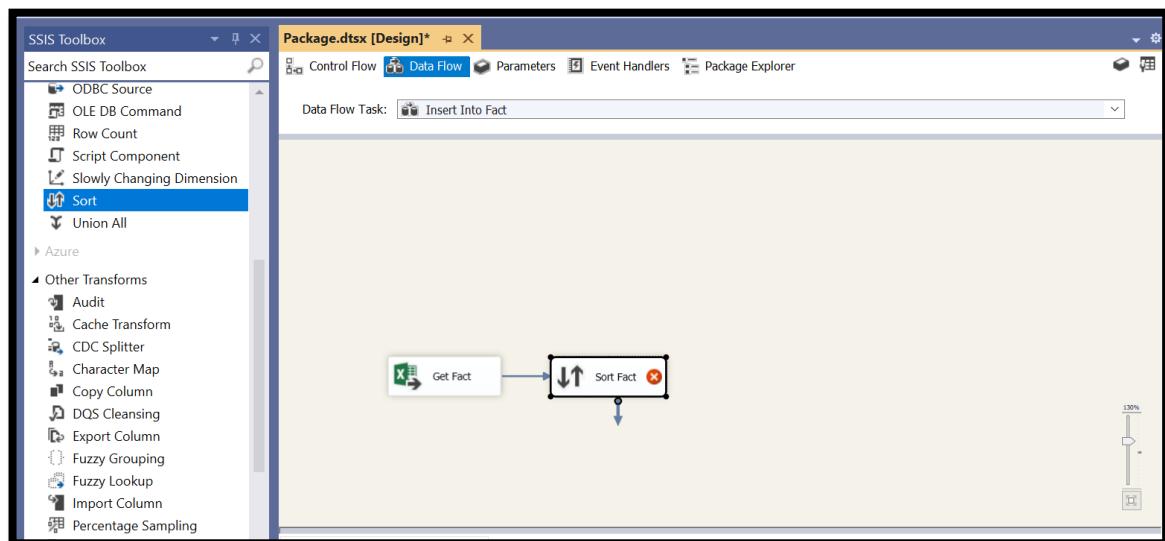
Bước 5: Mặc định đã chọn Excel connection đã tạo trước đó. Vì dữ liệu được lưu vào các sheet con bên trong file excel nên tại cần chọn fact_order_items



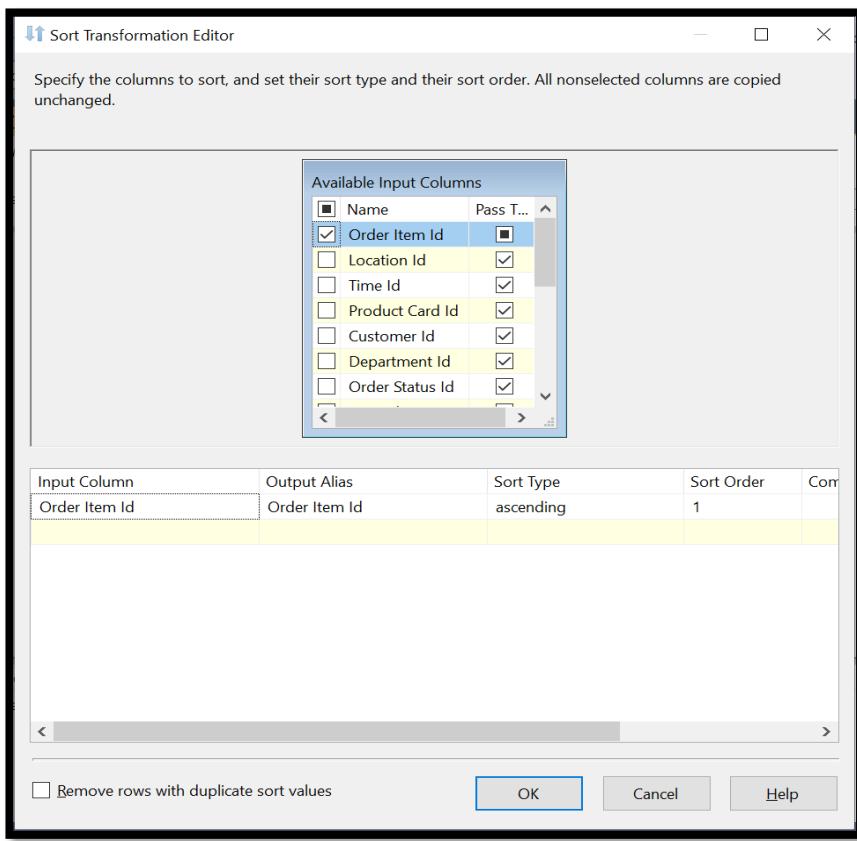
Bước 6: Chọn Columns để kiểm tra quá trình ánh xạ. Chọn OK để kết thúc quá trình lấy dữ liệu fact.



Bước 7: Kéo thả một Sort mới click và f2 để đặt tên là **Sort Fact**, nối Get Fact tới Sort Fact.



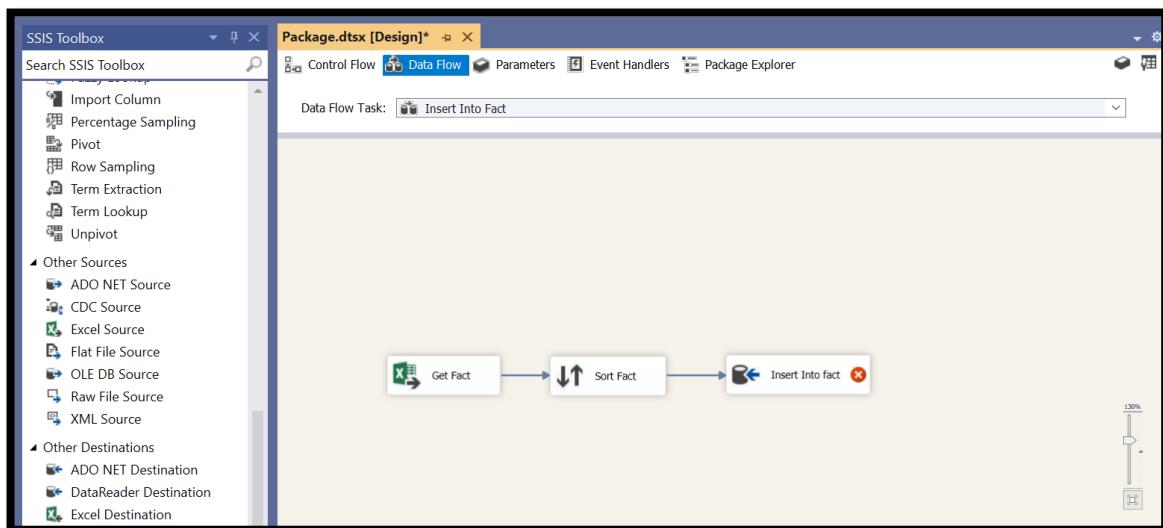
D-click **Sort Fact** và chọn sort theo **Order Item Id**.



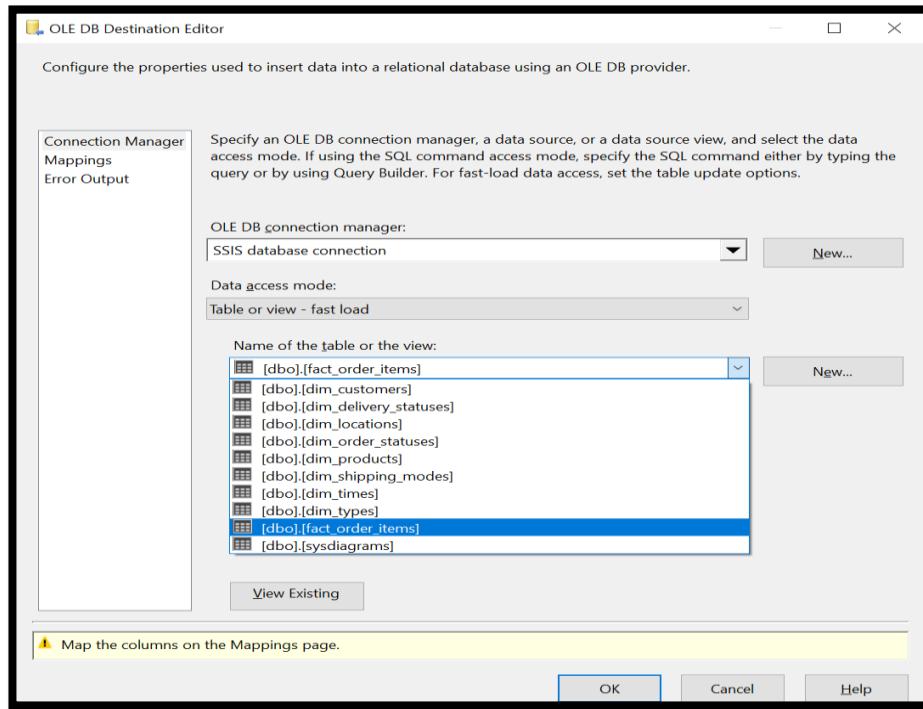
- Vì file excel này đã được xoá duplicate trong quá trình preprocess nên “Remove rows with duplicate sort values” là tùy chọn.

Chọn **OK** để kết thúc sort.

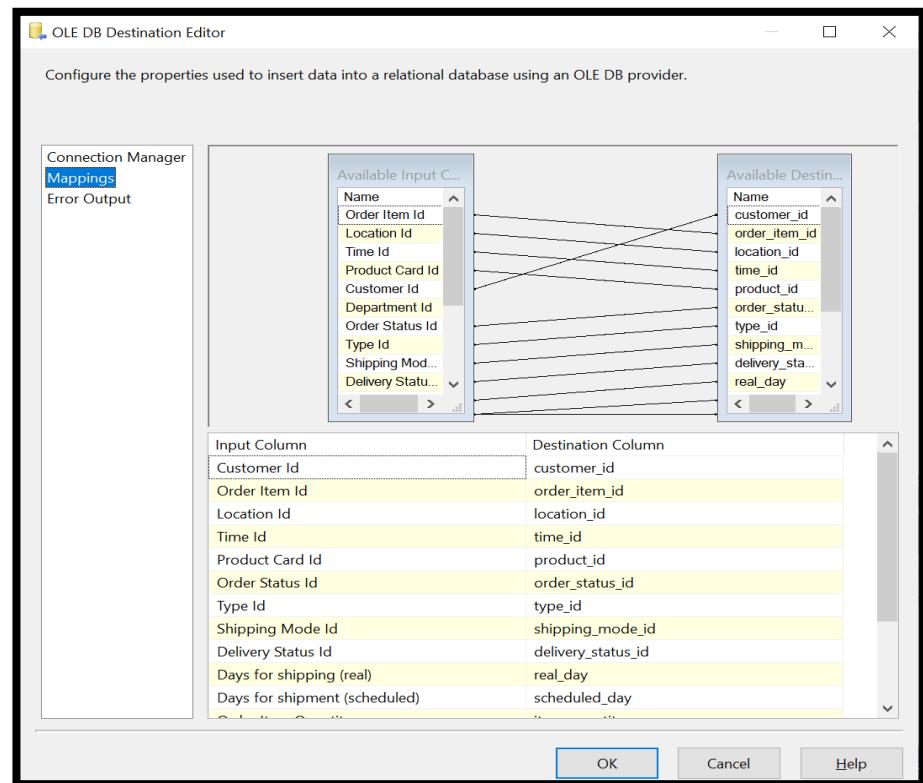
Bước 8: Kéo thả một **OLE DB Destination** mới, click và f2 để đặt tên là **Insert Into fact**, nối **Sort Fact** tới **Insert Into fact**.



Bước 9: D-click và **Insert Into fact**, vì đã tạo database connection trước đó nên data connection sẽ tự động kết nối. sau đó chọn **fact**.

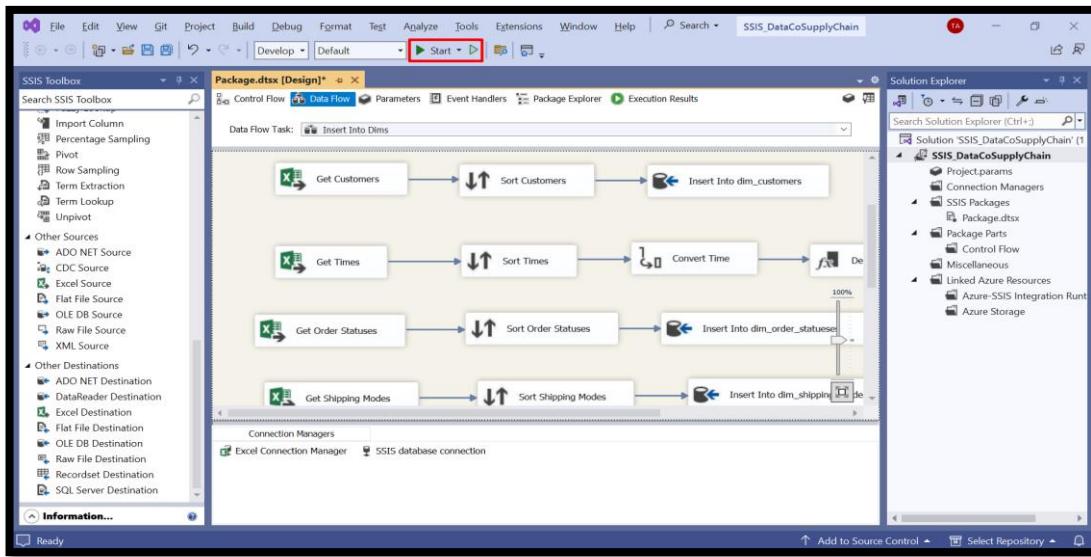


Bước 10: Chọn **Mappings** để kiểm tra việc ánh xạ các cột vào trong fact. Chọn **OK** để kết thúc

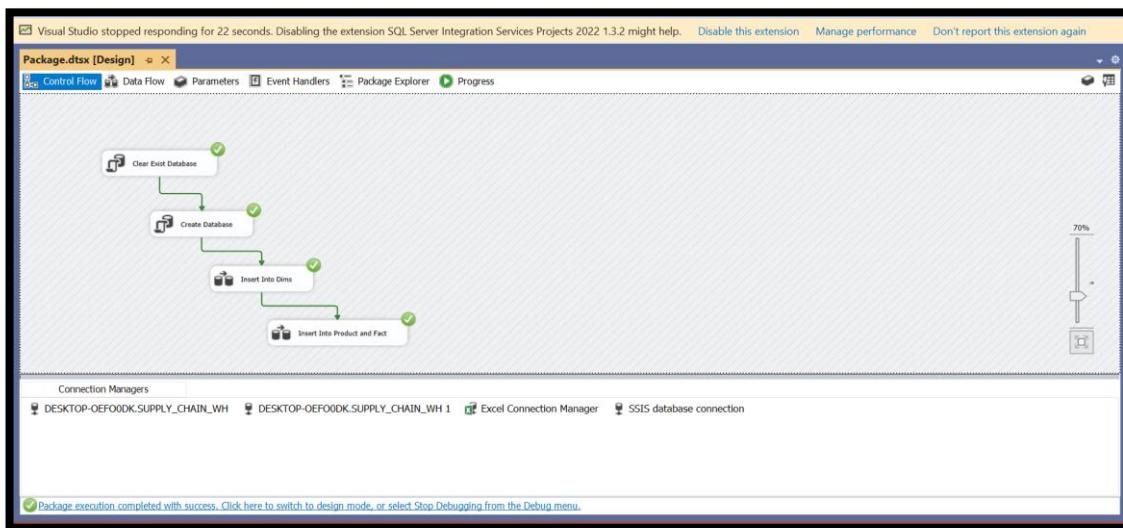


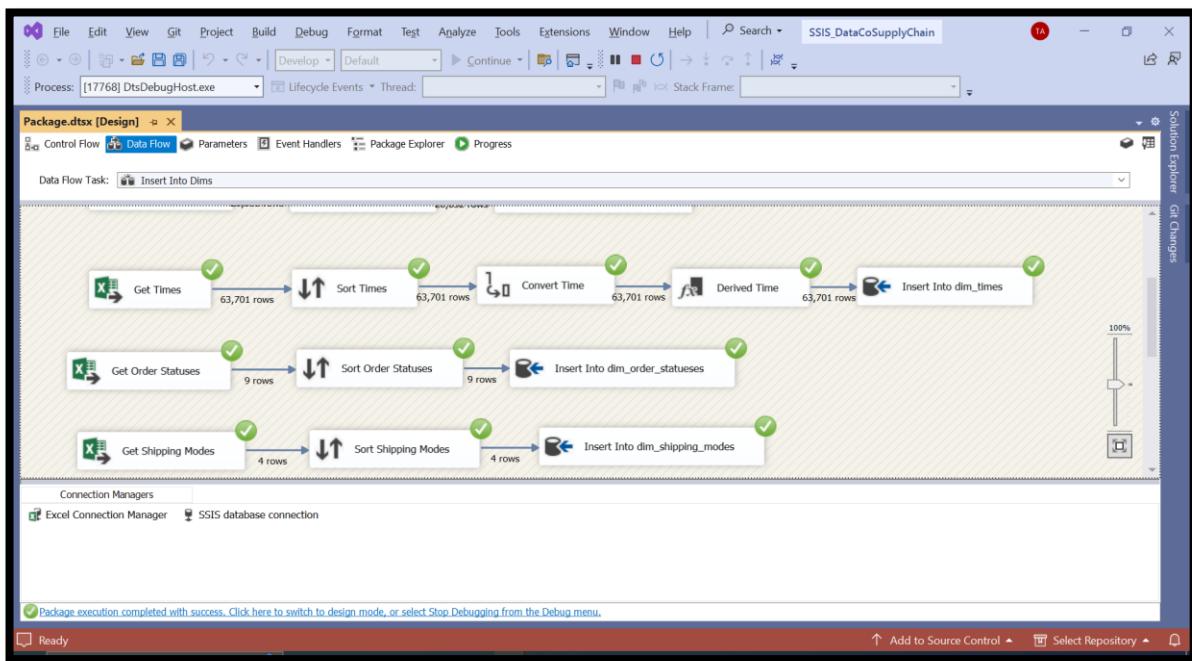
Chạy quá trình SSIS

Click start để thực thi SSIS



Kết quả





2.4. Kiểm tra dữ liệu các bảng

Kiểm tra dữ liệu fact

	order_item_id	location_id	time_id	product_id	customer_id	order_status_id	type_id	shipping_mode_id	delivery_status_id	real_day	scheduled_day	item_quantity	item_discount_rate
1	1	1942	4999	957	11599	2	1	4	1	2	4	1	0.20
2	2	1767	5665	1073	256	7	3	4	1	3	4	1	0.02
3	3	1767	5665	502	256	7	3	4	1	3	4	5	0.09
4	4	1767	5665	403	256	7	3	4	1	3	4	1	0.17
5	5	1767	6103	897	8827	2	1	4	2	5	4	2	0.18
6	6	1767	6103	365	8827	2	1	4	2	5	4	5	0.01
7	7	1767	6103	502	8827	2	1	4	2	5	4	3	0.18
8	8	1767	6103	1014	8827	2	1	4	2	5	4	4	0.20
9	9	1767	6352	957	11318	3	2	4	2	6	4	1	0.18
10	10	1767	6352	365	11318	3	2	4	2	6	4	5	0.00
11	11	1767	6352	1014	11318	3	2	4	2	6	4	2	0.17
12	12	1767	6352	957	11318	3	2	4	2	6	4	1	0.17
13	13	1767	6352	403	11318	3	2	4	2	6	4	1	0.15
14	14	2153	5681	1073	4530	3	2	3	2	3	2	1	0.02
15	15	2153	5681	957	4530	3	2	3	2	3	2	1	0.15
16	16	2153	5681	926	4530	3	2	3	2	3	2	5	0.02
17	17	2153	5898	365	2911	8	4	4	4	4	4	3	0.17
18	18	2153	5898	365	2911	8	4	4	4	4	4	5	0.25
19	19	2153	5898	1014	2911	8	4	4	4	4	4	4	0.18
20	20	2153	5898	502	2911	8	4	4	4	4	4	1	0.01
21	21	2153	6117	191	5657	7	3	4	2	5	4	2	0.05
22	22	2153	6117	1073	5657	7	3	4	2	5	4	1	0.01
23	23	2153	6117	1073	5657	7	3	4	2	5	4	1	0.00

Kiểm tra dữ liệu bảng dim_times

	time_id	day	month	year	hour	minute	date_of_week
1	1	1	1	2016	0	22	Friday
2	2	1	1	2016	0	32	Friday
3	3	1	1	2016	0	43	Friday
4	4	1	1	2016	0	53	Friday
5	5	1	1	2016	11	34	Friday
6	6	1	1	2016	12	38	Friday
7	7	1	1	2016	12	48	Friday
8	8	1	1	2016	12	59	Friday
9	9	1	1	2016	13	9	Friday
10	10	1	1	2016	13	19	Friday
11	11	1	1	2016	14	23	Friday
12	12	1	1	2016	14	44	Friday
13	13	1	1	2016	14	54	Friday
14	14	1	1	2016	15	4	Friday
15	15	1	1	2016	15	25	Friday
16	16	1	1	2016	16	8	Friday
17	17	1	1	2016	16	18	Friday
18	18	1	1	2016	16	39	Friday

Kiểm tra dữ liệu bảng dim_categories

	category_id	category_name
1	2	Soccer
2	3	Baseball & Softball
3	4	Basketball
4	5	Lacrosse
5	6	Tennis & Racquet
6	7	Hockey
7	9	Cardio Equipment
8	10	Strength Training
9	11	Fitness Accessories
10	12	Boxing & MMA
11	13	Electronics
12	16	As Seen on TV!

Kiểm tra dữ liệu bảng dim_products

	product_id	category_id	category_name	product_name	product_price
1	19	2	Soccer	Nike Men's Fingertrap Max Training Shoe	124.98
2	24	2	Soccer	Elevation Training Mask 2.0	79.98
3	35	3	Baseball & Softball	adidas Brazuca 2014 Official Match Ball	159.99
4	37	3	Baseball & Softball	adidas Kids' F5 Messi FG Soccer Cleat	34.99
5	44	3	Baseball & Softball	adidas Men's F10 Messi TRX FG Soccer Cleat	59.99
6	58	4	Basketball	Diamondback Boys' Insight 24 Performance Hybr	299.98
7	60	4	Basketball	SOLE E25 Elliptical	999.98
8	61	4	Basketball	Diamondback Girls' Clarity 24 Hybrid Bike 201	299.98
9	78	5	Lacrosse	Nike Kids' Grade School KD VI Basketball Shoe	99.98
10	93	5	Lacrosse	Under Armour Men's Tech II T-Shirt	24.98
11	116	6	Tennis & Racquet	Nike Men's Comfort 2 Slide	44.99
12	127	7	Hockey	Stiga Master Series ST3100 Competition Indoor	329.98
13	134	7	Hockey	Nike Women's Legend V-Neck T-Shirt	25.00
14	135	7	Hockey	Nike Dri-FIT Crew Sock 6 Pack	22.00
15	172	9	Cardio Equipment	Nike Women's Tempo Shorts	30.00
16	191	9	Cardio Equipment	Nike Men's Free 5.0+ Running Shoe	99.98
17	203	10	Strength Training	GoPro HERO3+ Black Edition Camera	399.98
18	208	10	Strength Training	SOLE E35 Elliptical	1999.98

Kiểm tra dữ liệu bảng dim_locations

	location_id	market	country	region	state	city
1	1	Africa	Nigeria	West Africa	Abia	Aba
2	2	Africa	Nigeria	West Africa	Ebonyi	Abakaliki
3	3	Africa	Nigeria	West Africa	Ogun	Abeokuta
4	4	Africa	Costa de Marfil	West Africa	Lagunes	Abidjan
5	5	Africa	Egipto	North Africa	Ash Sharqiyah	Abu Kabir
6	6	Africa	Ghana	West Africa	Gran Accra	Accra
7	7	Africa	Sudán	North Africa	Nilo Blanco	Ad Diwem
8	8	Africa	Marruecos	North Africa	Sus-Masa-Draa	Agadir
9	9	Africa	Argelia	North Africa	Djelfa	Ain Oussera
10	10	Africa	Nigeria	West Africa	Ondo	Akure
11	11	Africa	Libia	North Africa	Al Marqab	Al Khums
12	12	Africa	Egipto	North Africa	Gharbia	Al Mahallah al Kubra
13	13	Africa	Sudán	North Africa	Gezira	Al Manaqil
14	14	Africa	Egipto	North Africa	Alejandría	Alexandria
15	15	Africa	Argelia	North Africa	Alger	Algiers
16	16	Africa	Argelia	North Africa	Annaba	Annaba
17	17	Africa	Madagascar	East Africa	Analamanga	Antananarivo
18	18	Africa	Madagascar	East Africa	Diana	Antsiranana

Kiểm tra dữ liệu bảng dim_customer

	customer_id	customer_fname	customer_lname
1	1	Richard	Hernandez
2	2	Mary	Barrett
3	3	Ann	Smith
4	4	Mary	Jones
5	5	Robert	Hudson
6	6	Mary	Smith
7	7	Melissa	Wilcox
8	8	Megan	Smith
9	9	Mary	Perez
10	10	Melissa	Smith
11	11	Mary	Huffman
12	12	Christopher	Smith
13	13	Mary	Baldwin
14	14	Katherine	Smith
15	15	Jane	Luna
16	16	Tiffany	Smith
17	17	Mary	Robinson
18	18	Robert	Smith

Kiểm tra dữ liệu bảng dim_types

	type_id	type
1	1	CASH
2	2	DEBIT
3	3	PAYMENT
4	4	TRANSFER

Kiểm tra dữ liệu bảng dim_order_statuses

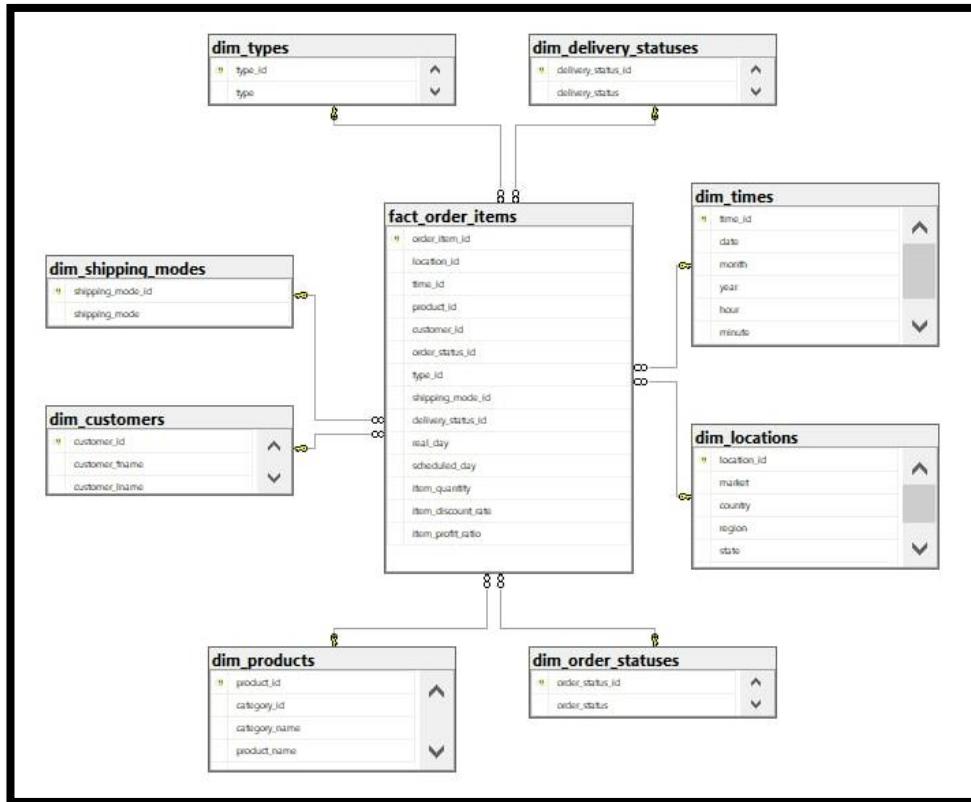
	order_status_id	order_status
1	1	CANCELED
2	2	CLOSED
3	3	COMPLETE
4	4	ON_HOLD
5	5	PAYMENT REVIEW
6	6	PENDING
7	7	PENDING PAYMENT
8	8	PROCESSING
9	9	SUSPECTED FRAUD

Kiểm tra dữ liệu bảng dim_delivery_statuses

	delivery_status_id	delivery_status
1	1	Advance shipping
2	2	Late delivery
3	3	Shipping canceled
4	4	Shipping on time

Kiểm tra dữ liệu bảng dim_shipping_modes

	shipping_mode_id	shipping_mode
1	1	First Class
2	2	Same Day
3	3	Second Class
4	4	Standard Class



CHƯƠNG 3. PHÂN TÍCH DỮ LIỆU

3.1. Cài đặt công cụ

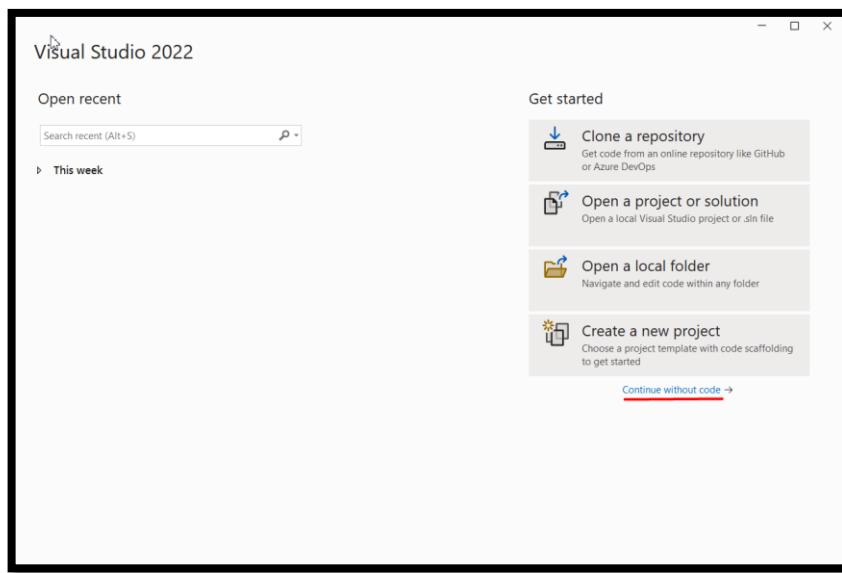
Để thực hiện được quá trình SSAS ta cần chuẩn bị và cài đặt các công cụ sau:

+ Microsoft SQL Server có cài đặt Analysis Services.

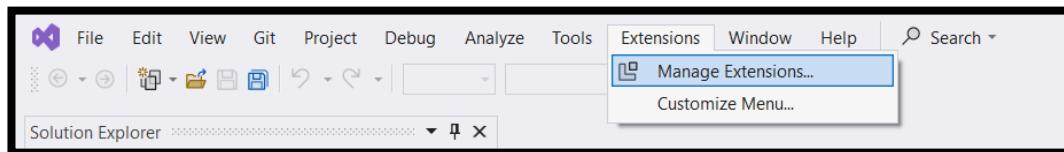
+ Microsoft Analysis Services Projects

3.1.1. Cài đặt Microsoft Analysis Services Projects extension

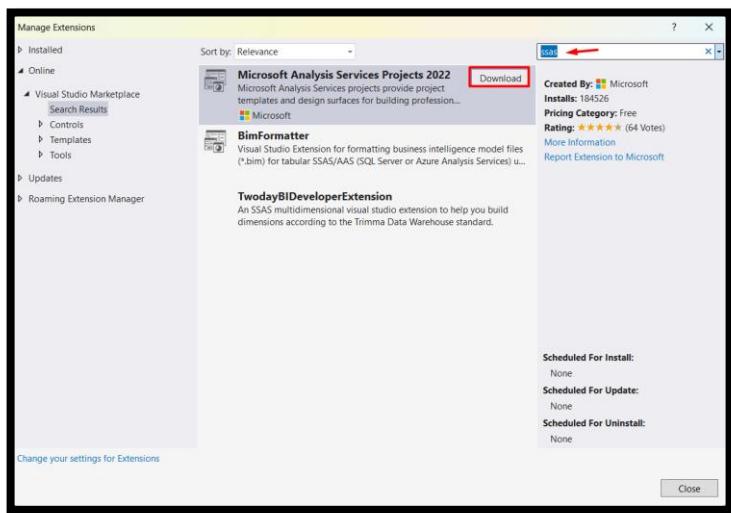
Bước 1: Mở Visual Studio và chọn "Continue without code".



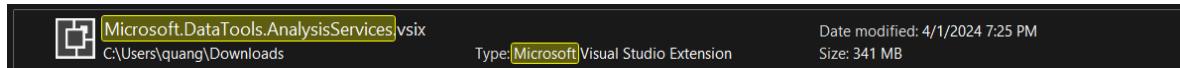
Bước 2: Trên thanh công cụ , ở tab “Extensions”, chọn “Manage Extensions...” ở phần pop-up bên dưới.



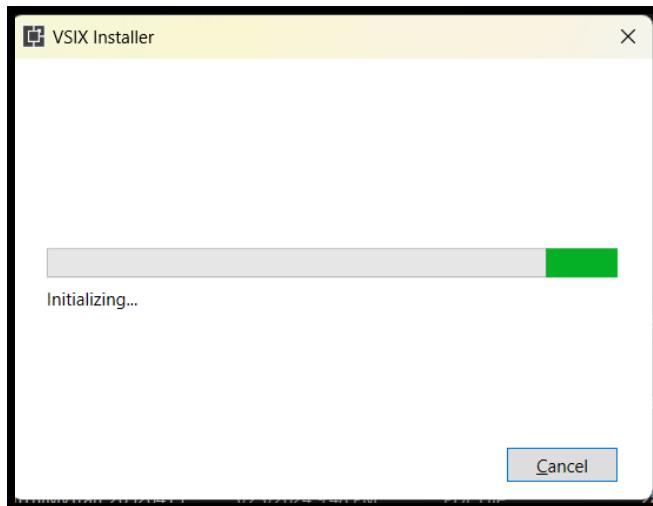
Bước 3: Tìm extension với keyword “ssas” và tải về công cụ Microsoft Analysis Services Projects.



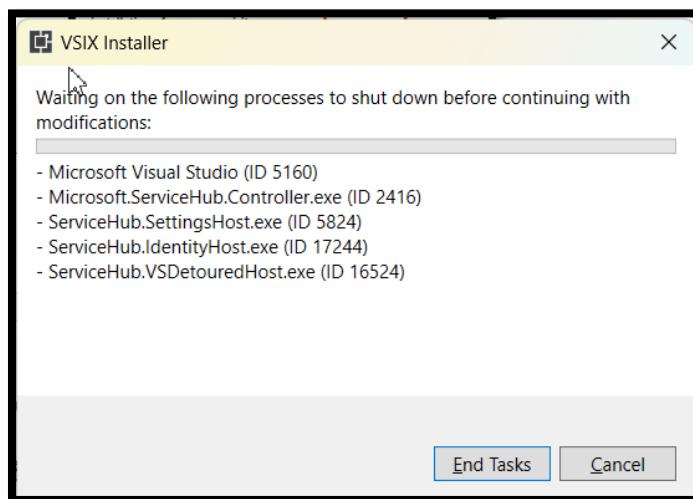
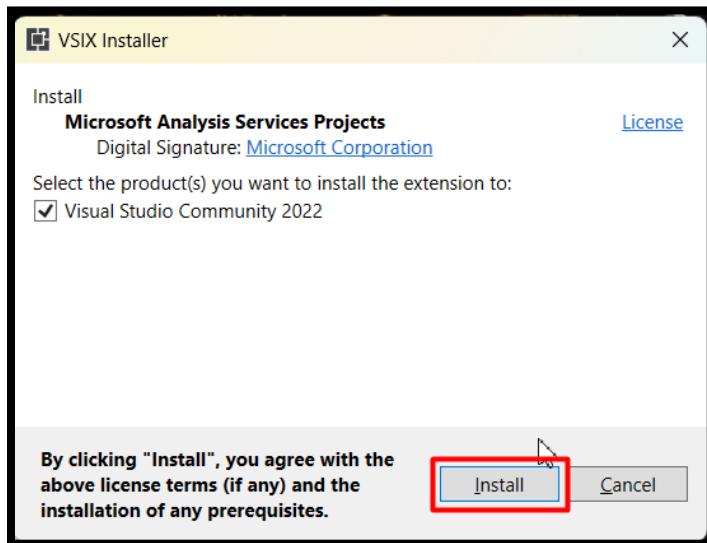
Sau khi chọn **Download**, công cụ sẽ được tự động cài đặt.



Bước 5: Đóng VS code và Mở file cài đặt vừa tải về.



Chọn “Install”



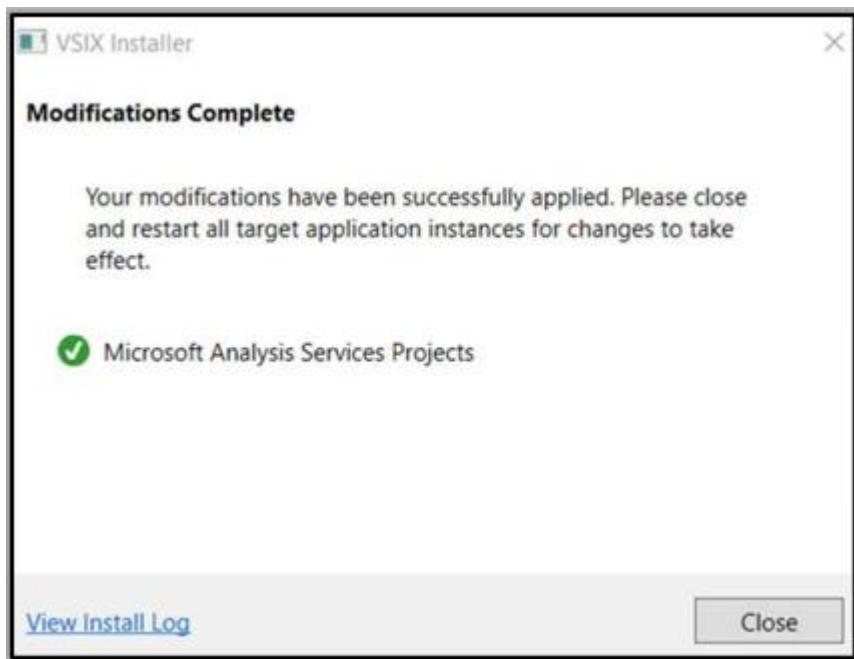
Bước 6: Sau khi đã khởi tạo xong, ta chọn **Modify** để đồng ý với các điều khoản khi cài đặt công cụ Microsoft Analysis Services Projects.



Sau đó, công cụ được tiếp tục cài đặt vào Visual Studio.

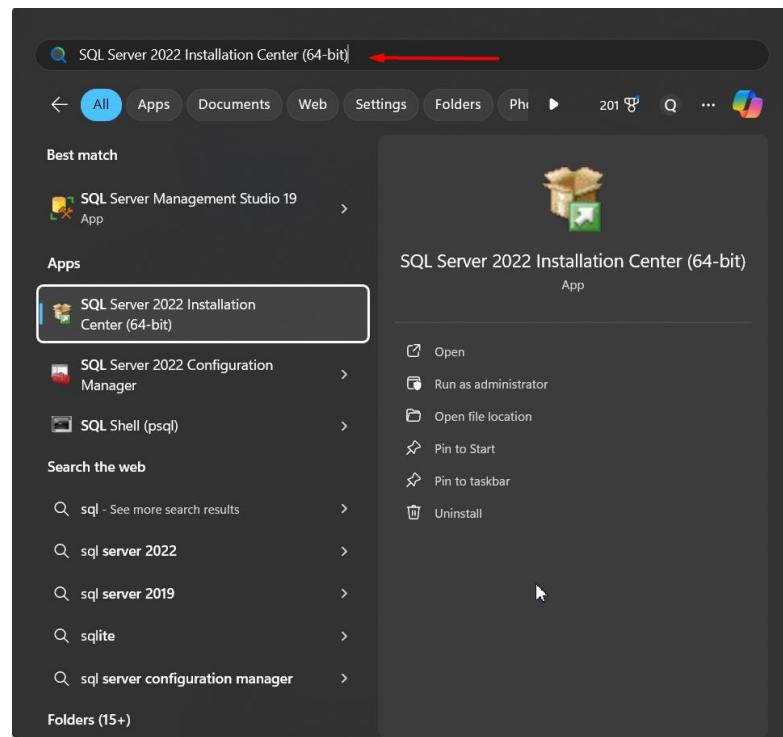


Bước 7: Sau khi cài đặt hoàn tất, ta chọn **Close** để đóng cửa sổ VSIX Installer.

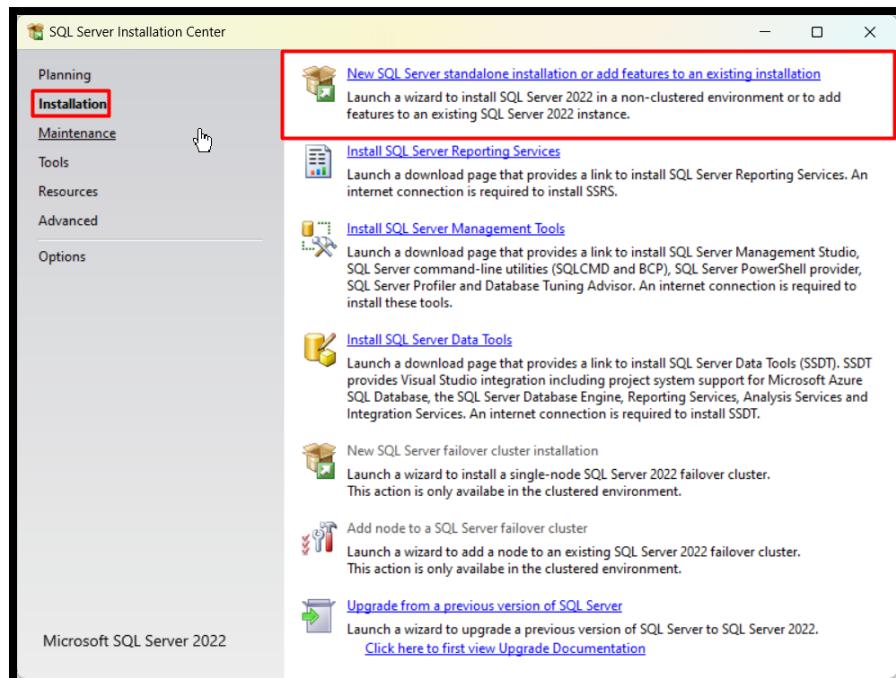


3.1.2. Cài đặt Analysis Services

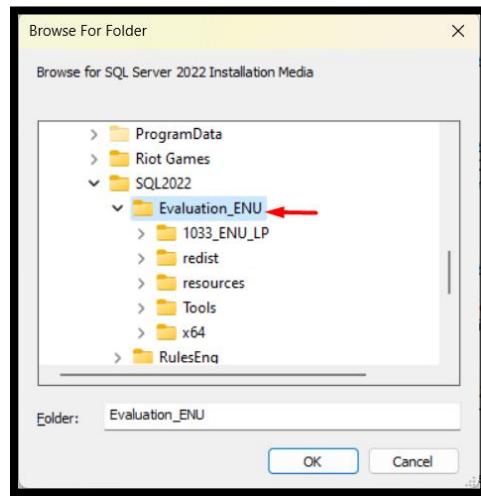
Bước 1: Tại thanh search window, tìm kiếm với keyword “**sql installation ...**”, tìm **SQL Server Installation Center**.



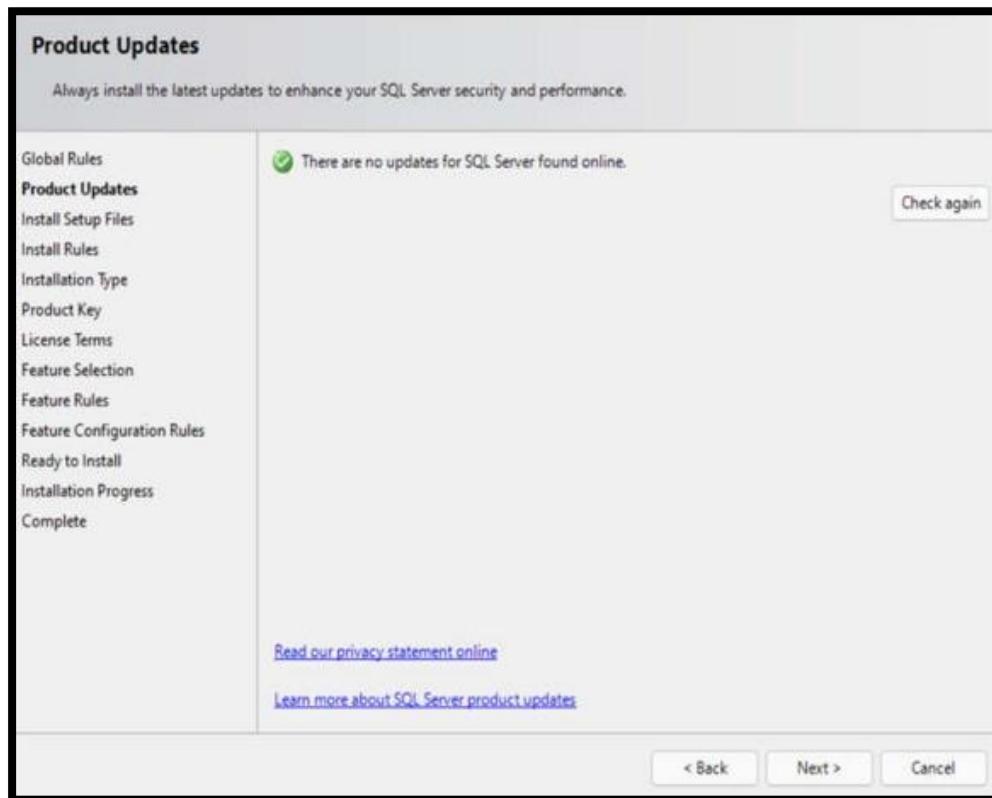
Bước 2: Chọn mục **Installation**, tại mục Installation chọn “**New SQL Server standalone or add features to an existing installation**”.



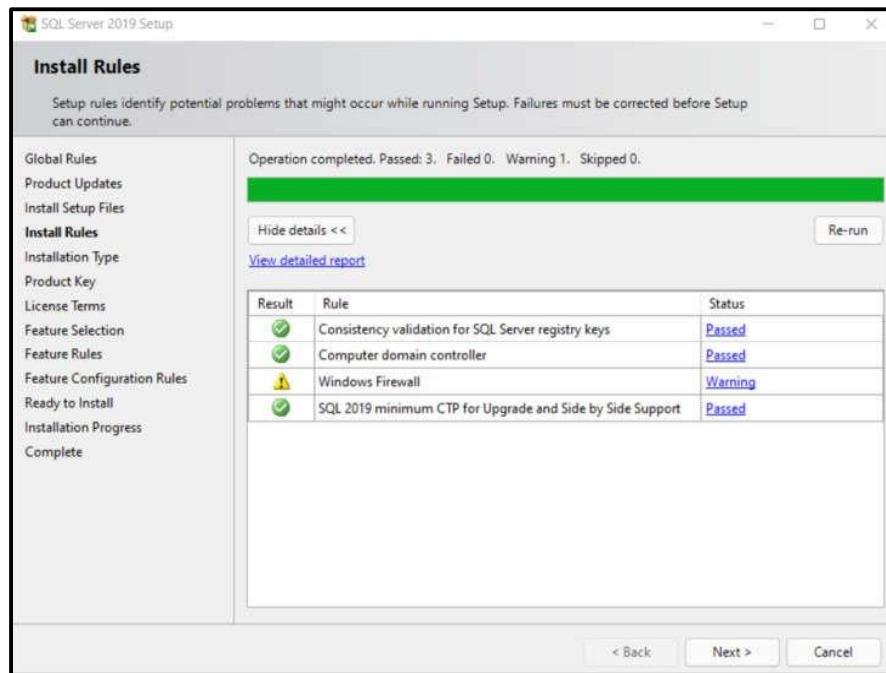
Sau đó, hộp thoại mới được mở ra để chọn thư mục chứa ứng dụng SQL.



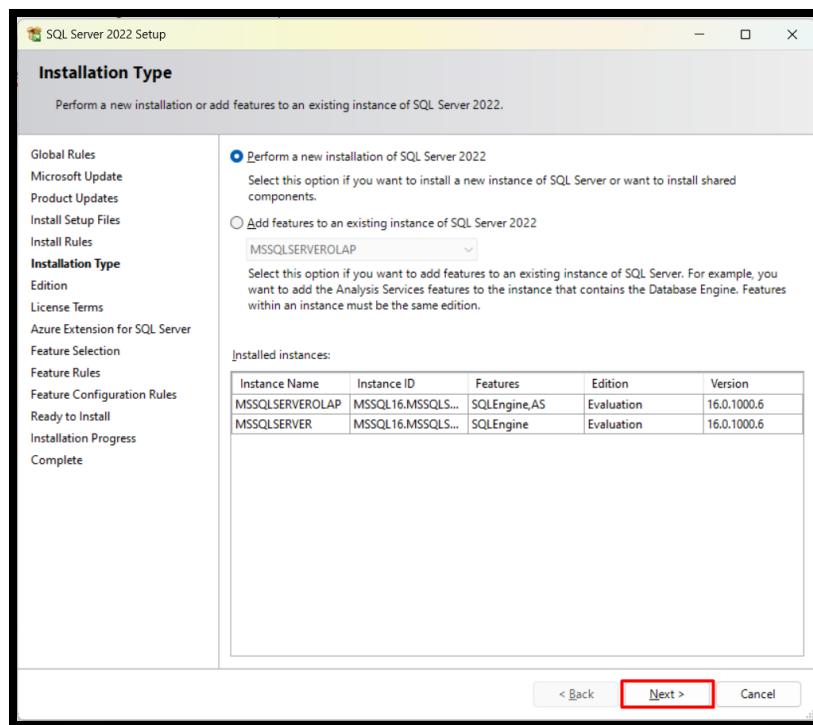
Bước 3: Hộp thoại mới xuất hiện, ta chọn Next



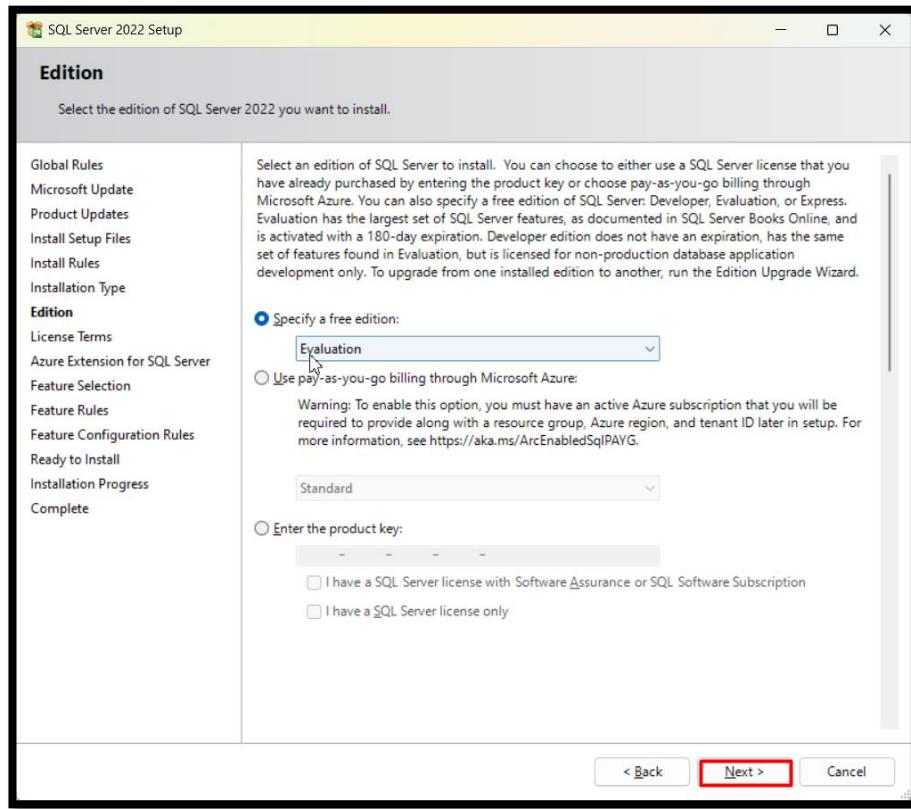
Bước 4: Tiếp tục chọn Next



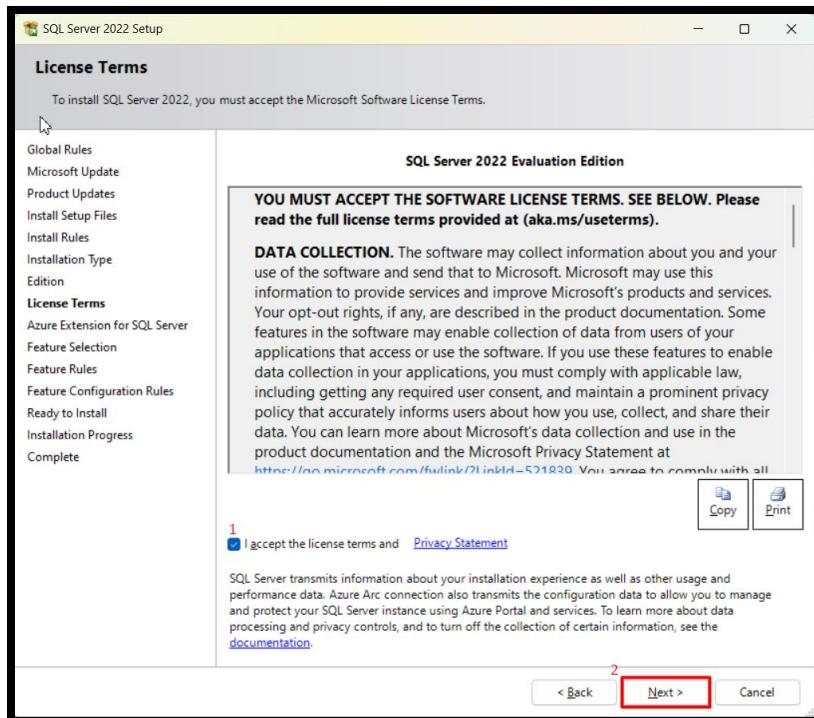
Bước 5: Tiếp tục chọn Next



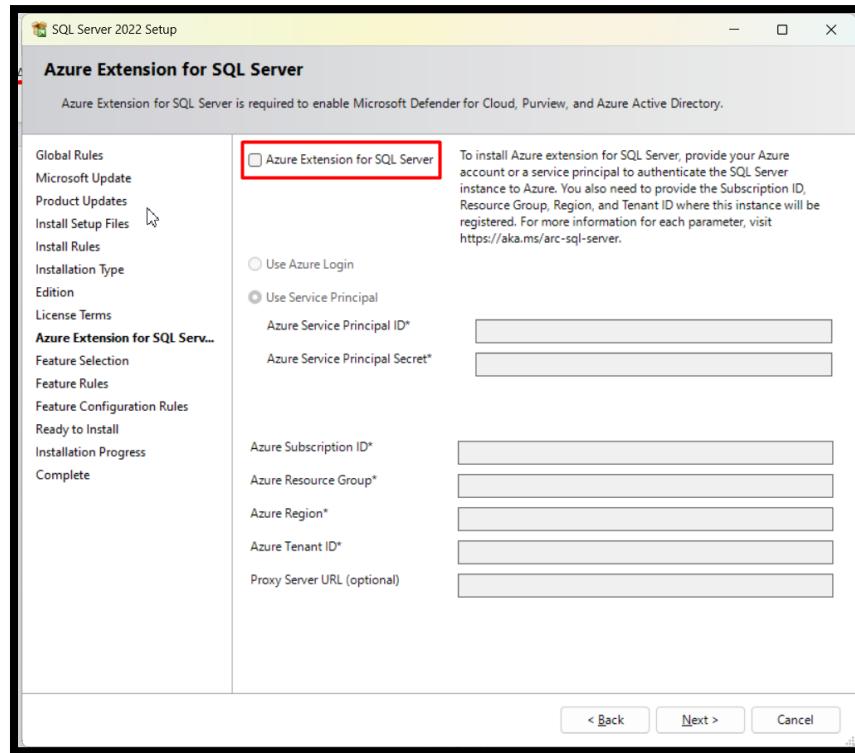
Bước 6: Tiếp tục chọn Next



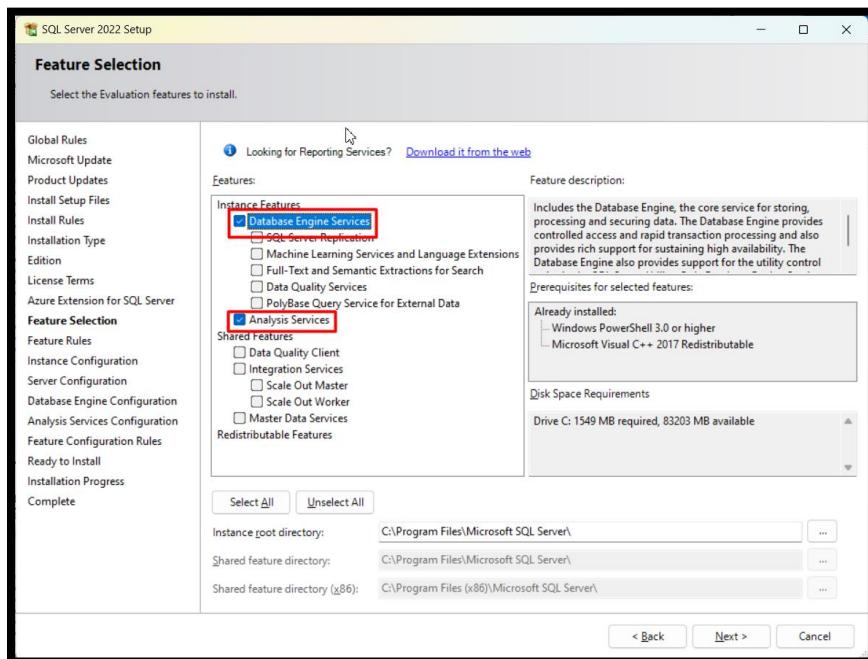
Bước 7: Chọn I accept license terms and Privacy Statement. Sau đó tiếp tục chọn Next.



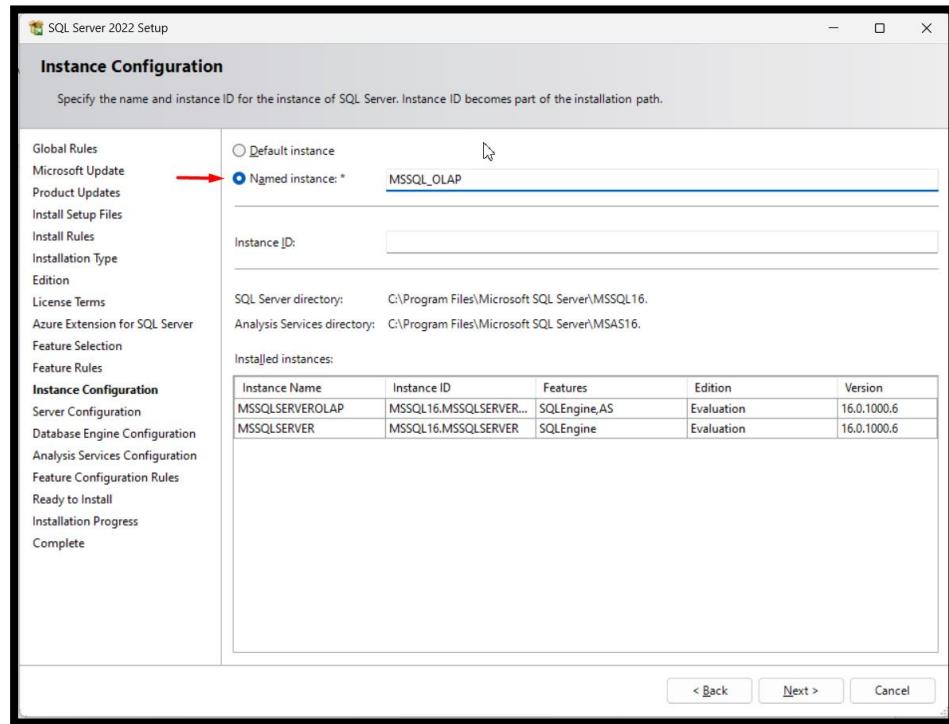
Bước 8: Do không sử dụng Azure nên bỏ chọn option: “Azure Extension for SQL Server”



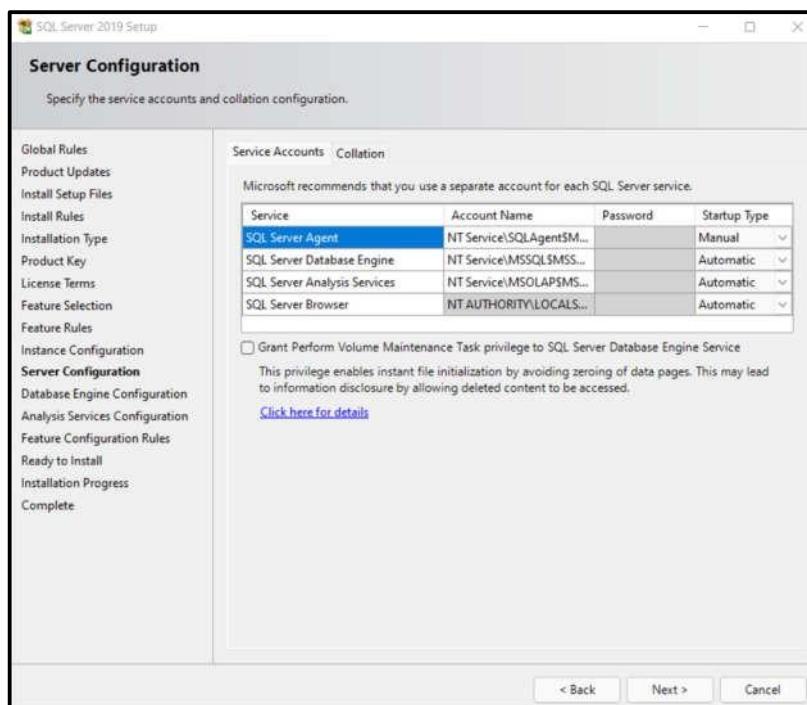
Bước 9: Ta chọn **Database Engine Services** và **Analysis Services**. Sau đó tiếp tục chọn Next.



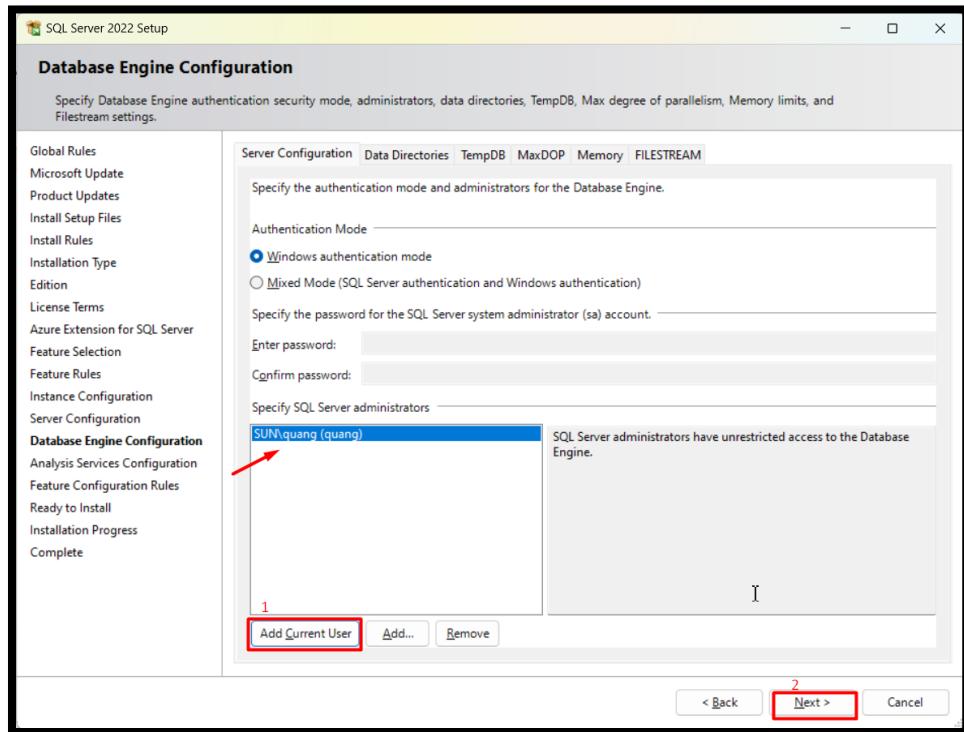
Bước 9: Chọn **Name instance**, sau đó ta đặt tên cho instance vừa tạo. Sau đó chọn Next.



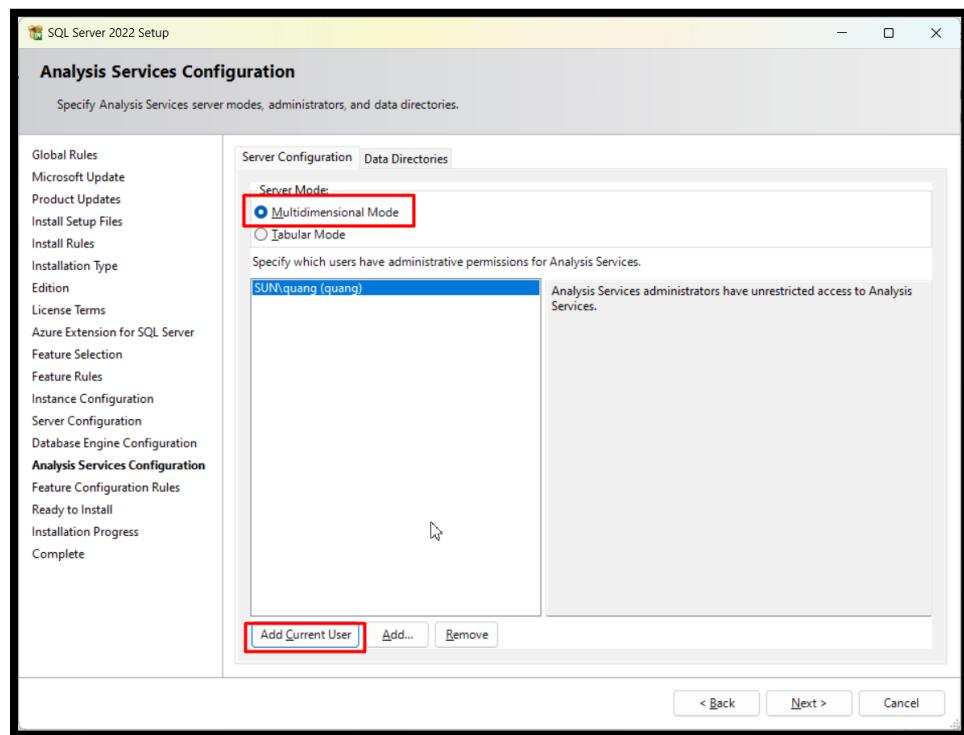
Bước 9: Chọn Next.



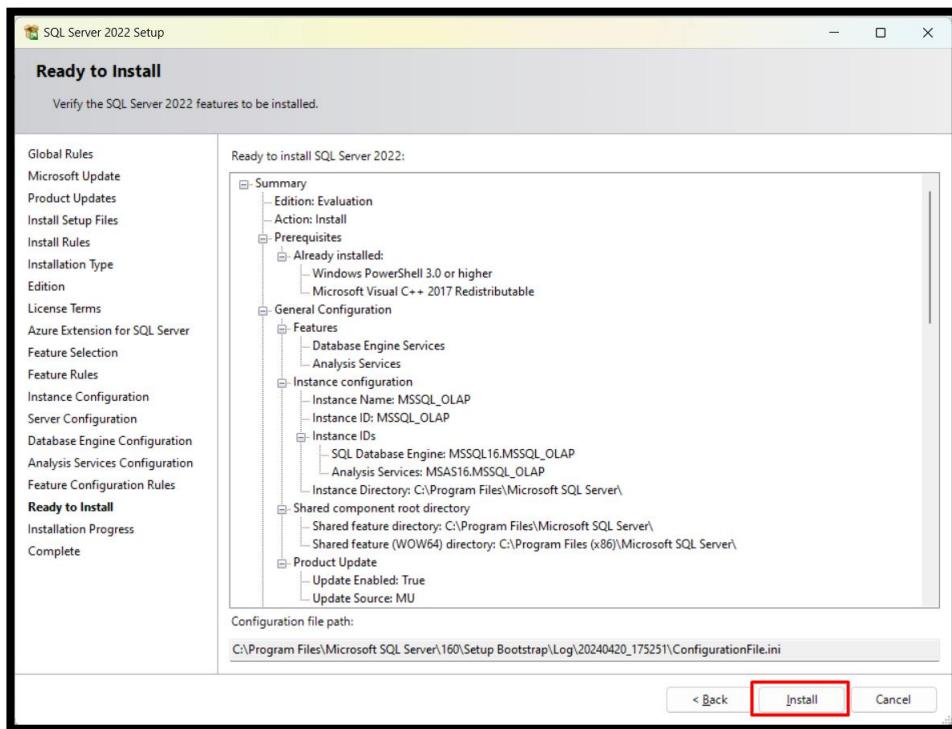
Bước 11: Chọn “Windows authentication mode”. Tiếp theo, chọn “Add Current User”. Đợi một lúc khi cho đến khi user xuất hiện, sau đó chọn Next.



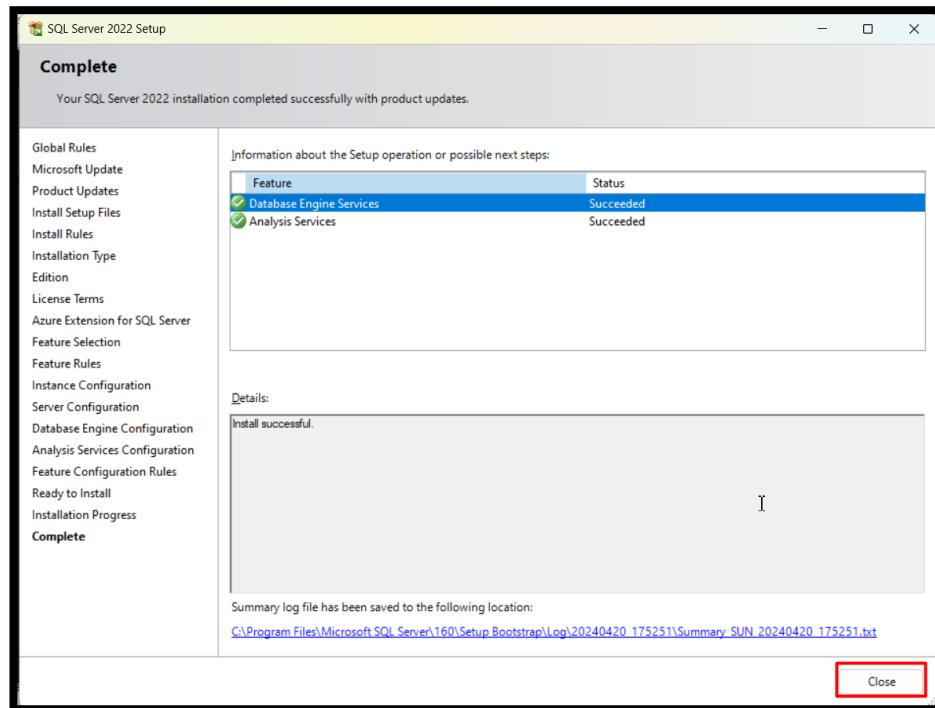
Bước 12: Chọn Multidimensional and Data Mining Mode. Tiếp theo, chọn Add Current User. Sau khi user xuất hiện, ta chọn vào tên user và chọn Next.



Bước 13: Chọn Install.

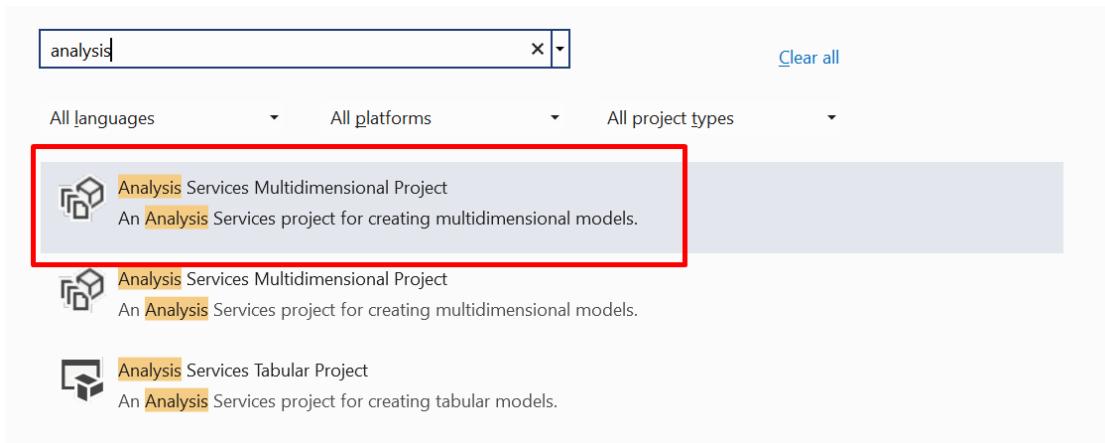


Bước 14: Sau khi cài đặt hoàn tất, chọn “Close”.

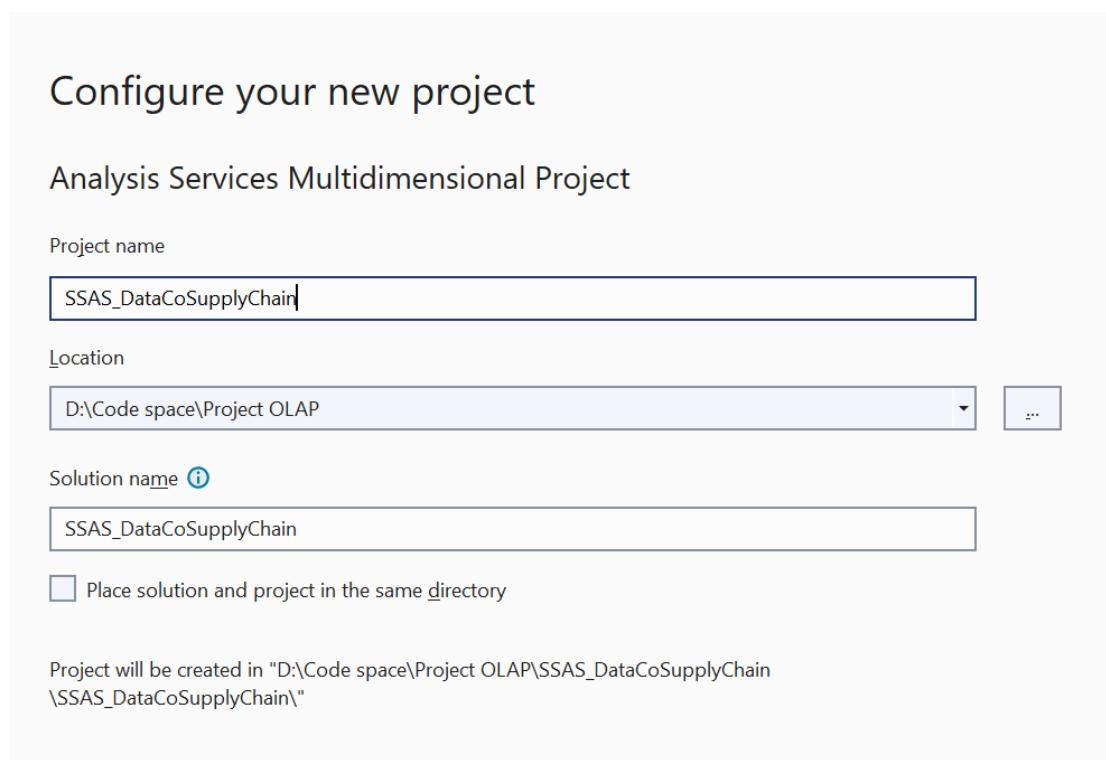


3.2. Tạo mới Project SSAS

Bước 1: Mở Visual Studio và chọn “Create a new project”. Chọn **Analysis Services Multidimensional Project** và chọn **Next**

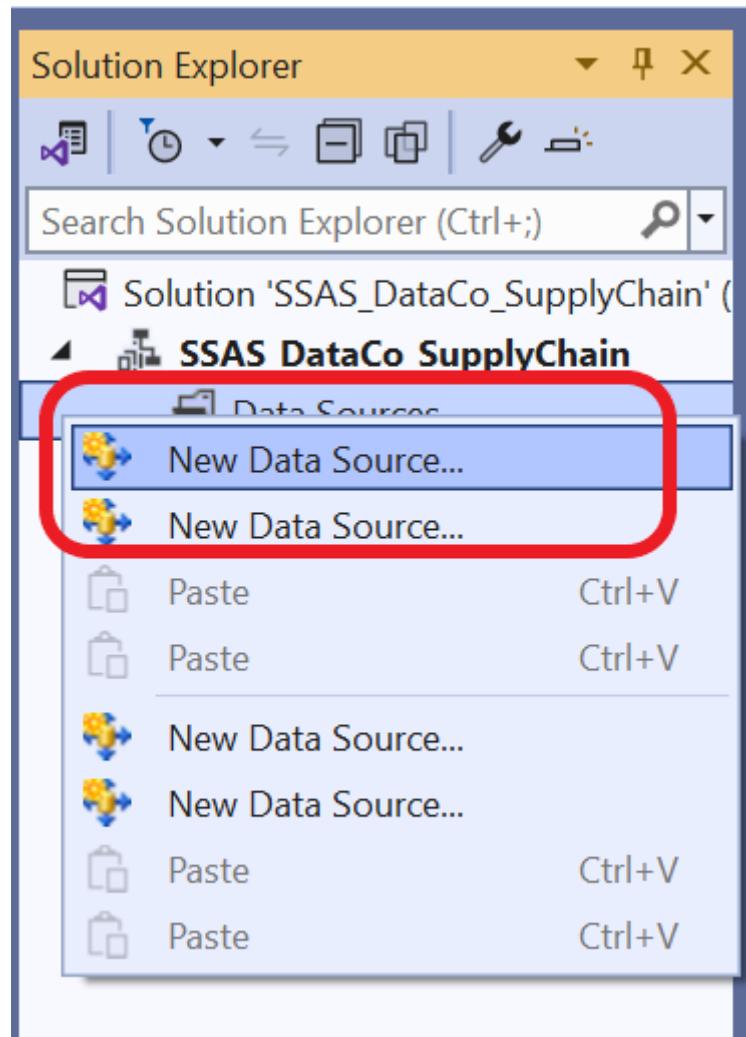


Bước 2: Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn **Create**.

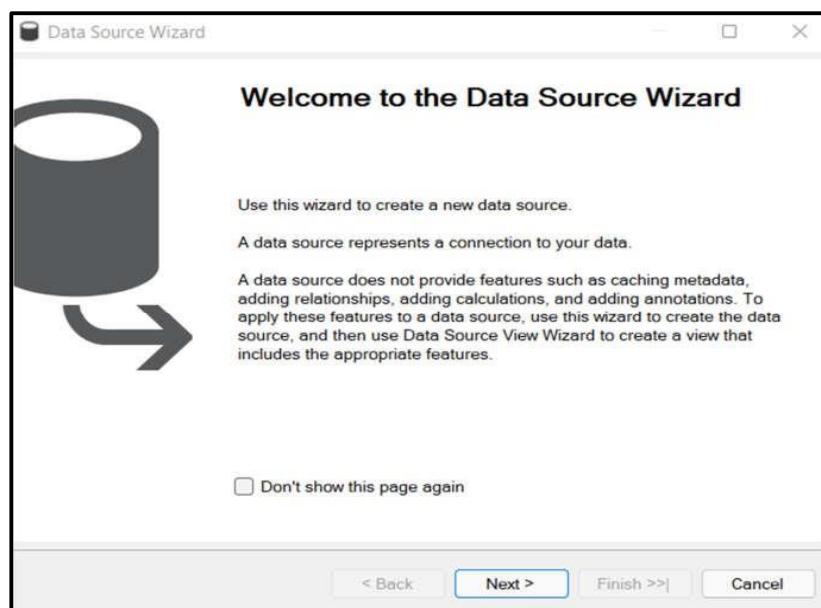


3.3. Xác định dữ liệu nguồn (Data Sources)

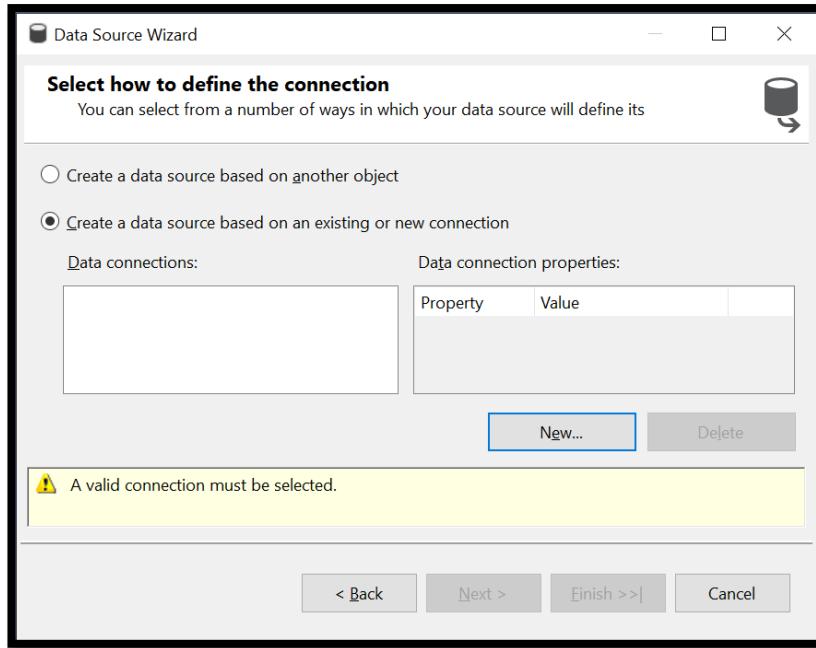
Bước 1: Tại Solution Explorer, ta click chuột phải vào thư mục Data Sources và chọn **New Data Source**.



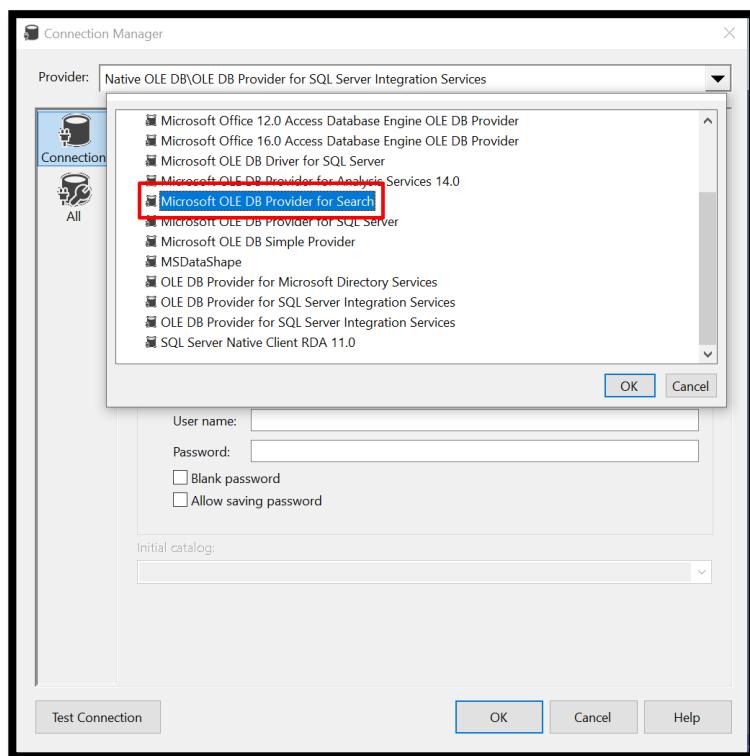
Bước 2: Hộp thoại Data Source Wizard xuất hiện, chọn Next để tiếp tục.



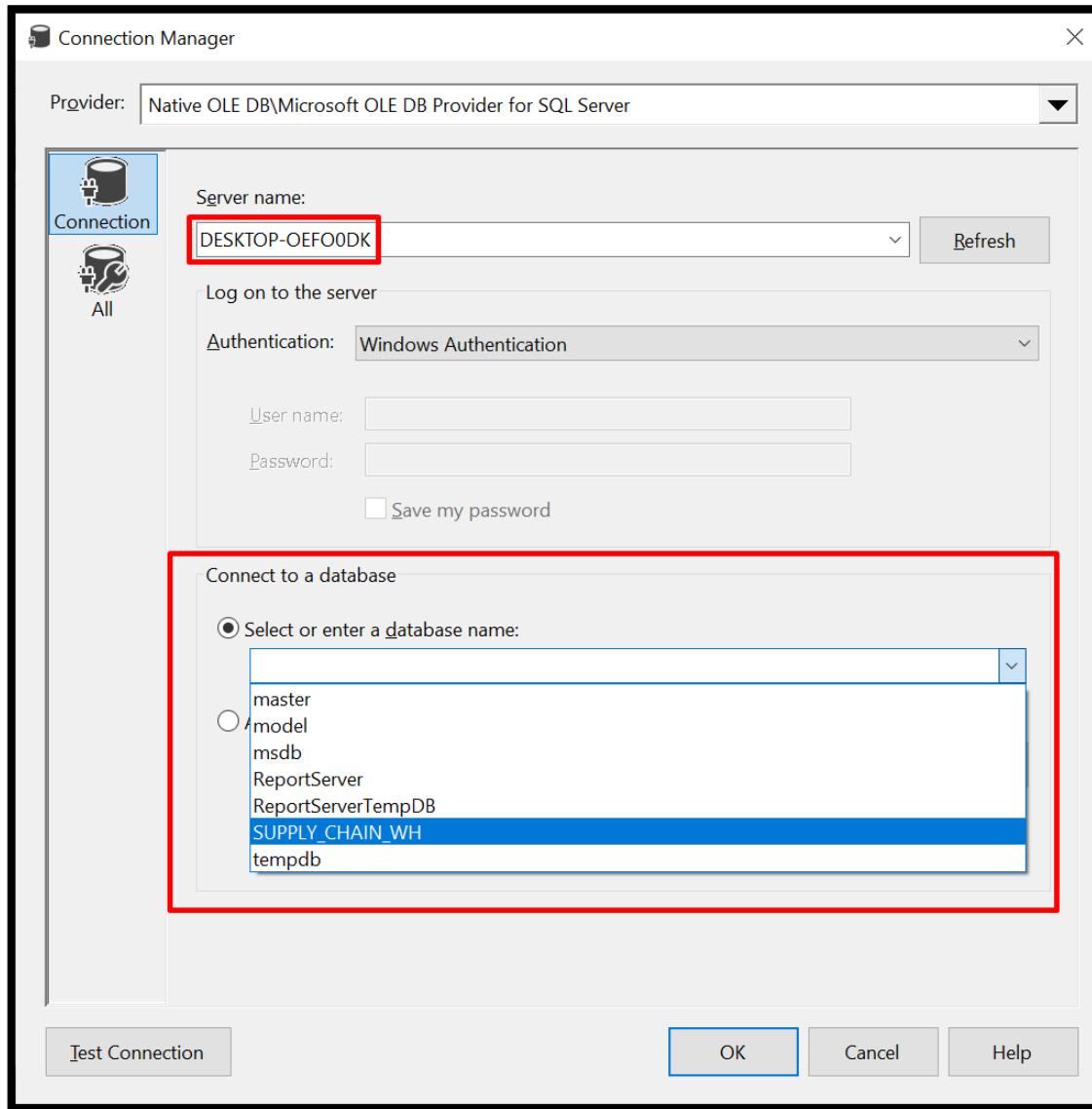
Bước 3: Chọn “Create a data source based on an existing or new connection” sau đó chọn New... để tạo kết nối với cơ sở dữ liệu đã được tạo từ quá trình SSIS.



Bước 4: Hộp thoại Connection Manager xuất hiện, tại phần Provider, chọn Microsoft OLE DB Provider for Search.

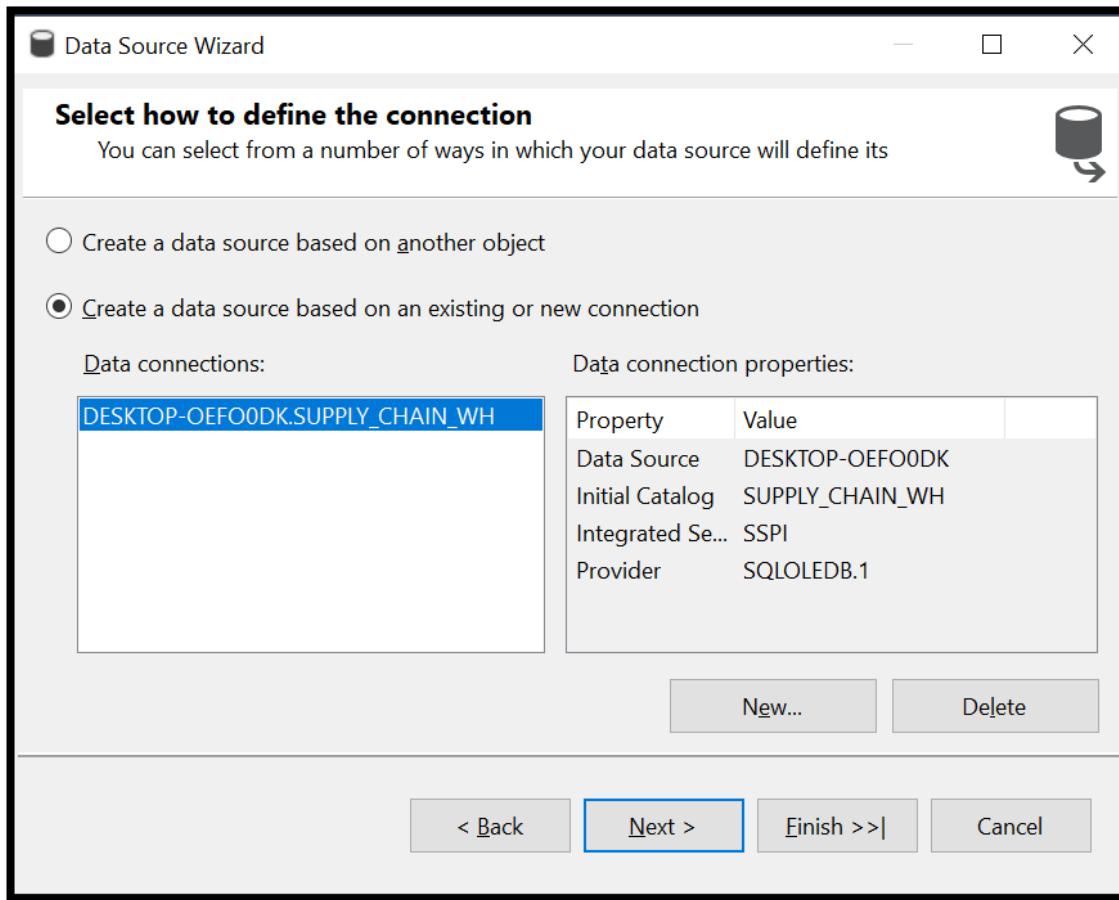


Nhập Server name bằng tay, chọn phương thức kết nối phù hợp (tại đây là Window Authentication) và chọn cơ sở dữ liệu mà ta đã tạo ra từ quá trình SSIS. Sau đó chọn OK.

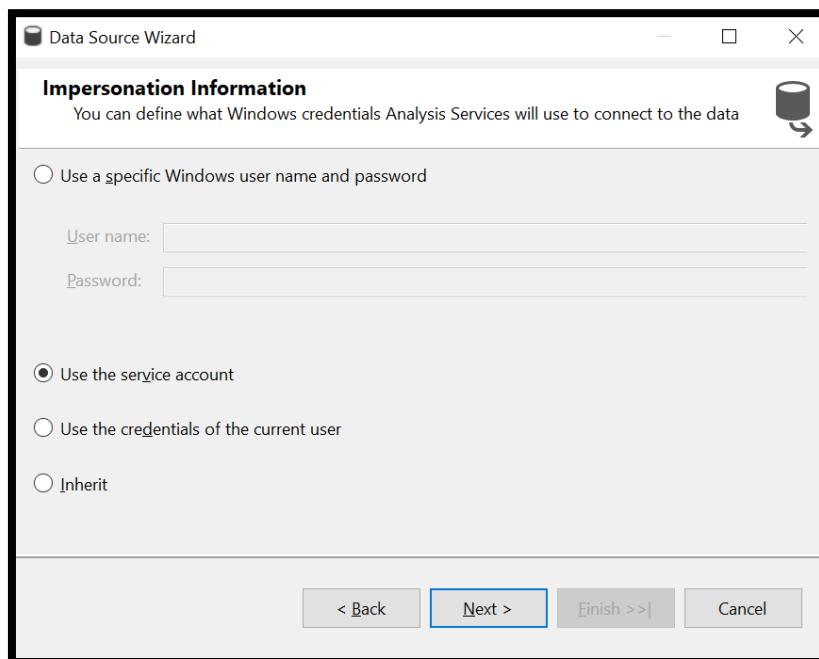


Chọn **OK**

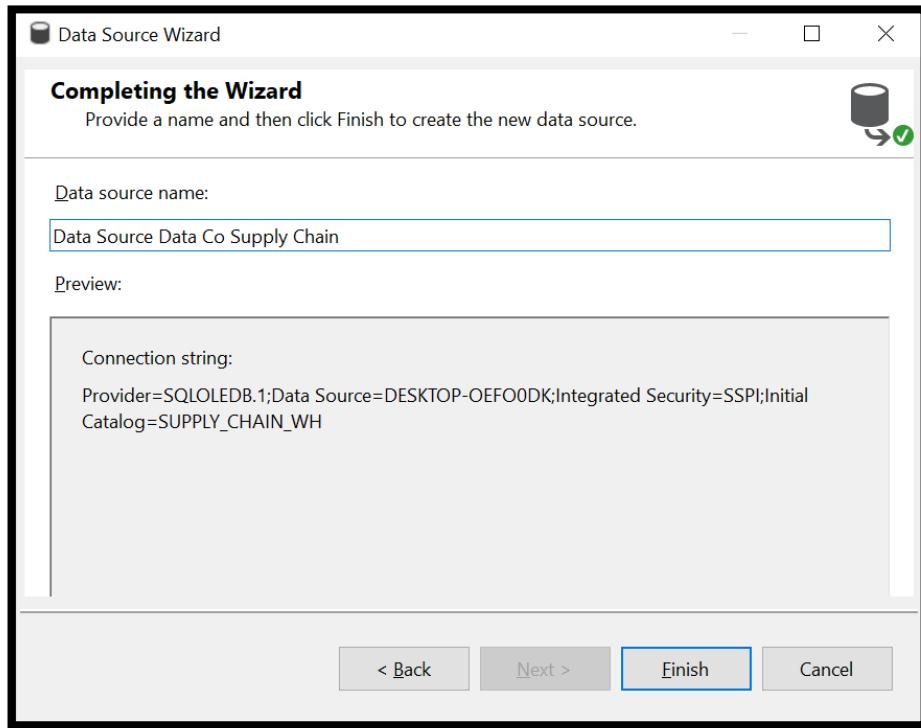
Chọn data source vừa tạo và chọn Next để tiếp tục.



Bước 5: Chọn “Use the service account”, sau đó chọn Next để tiếp tục.

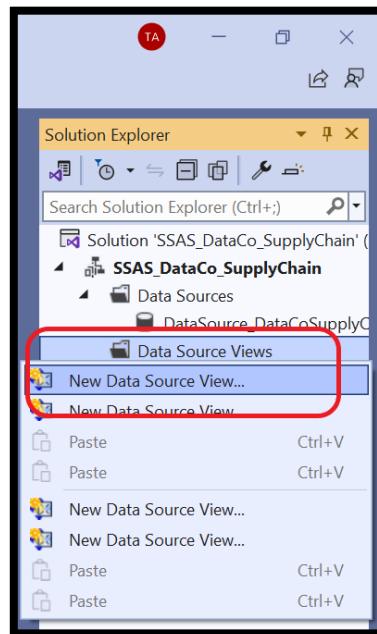


Bước 6: Cuối cùng chọn tên cho datasource và chọn Finish để hoàn tất quy trình định nghĩa nguồn dữ liệu.

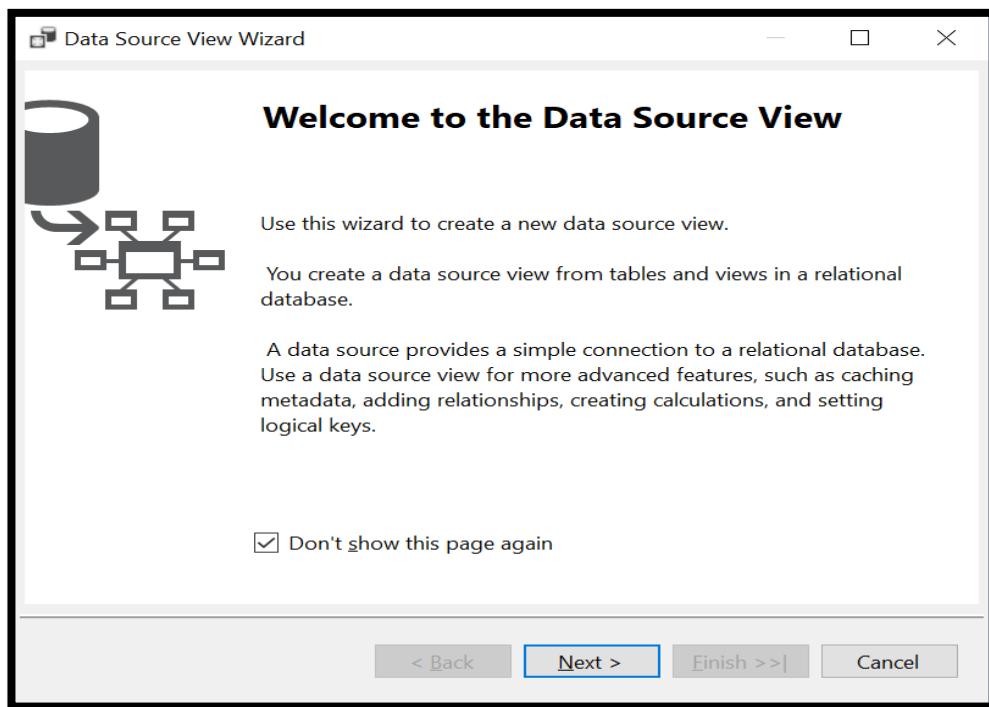


3.4. Xác định Data Source Views

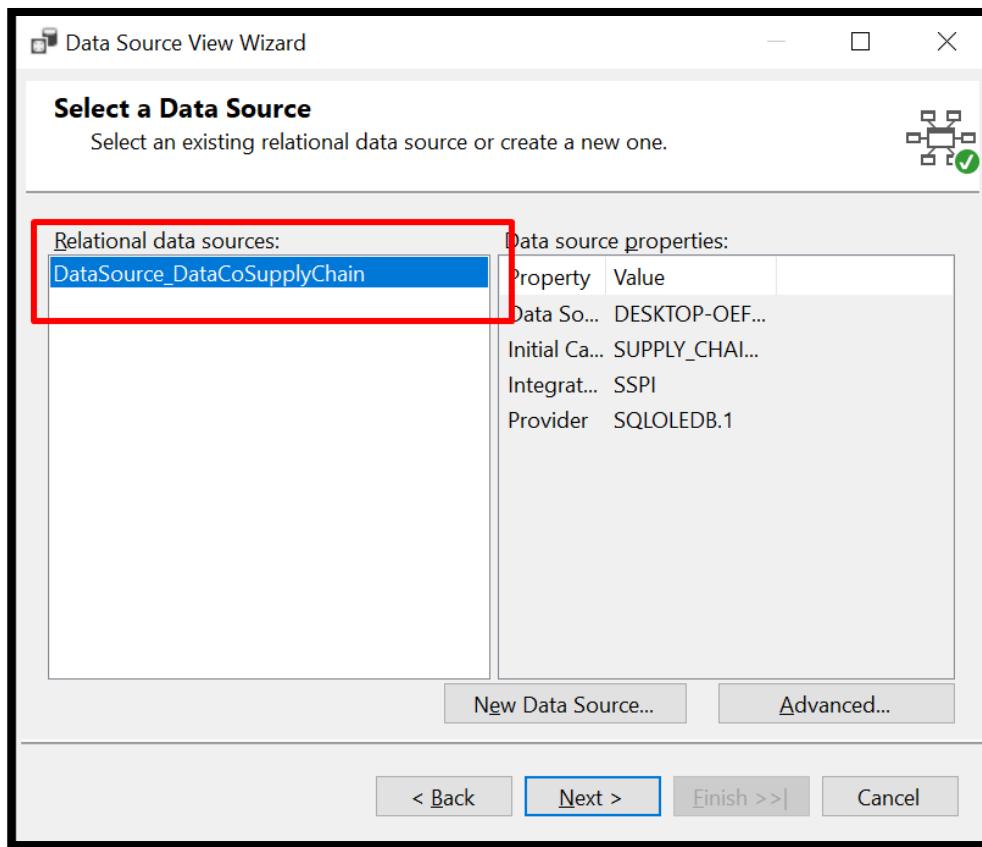
Bước 1: Tại Solution Explorer, r-click Data Source Views và chọn **New Data Source View**.

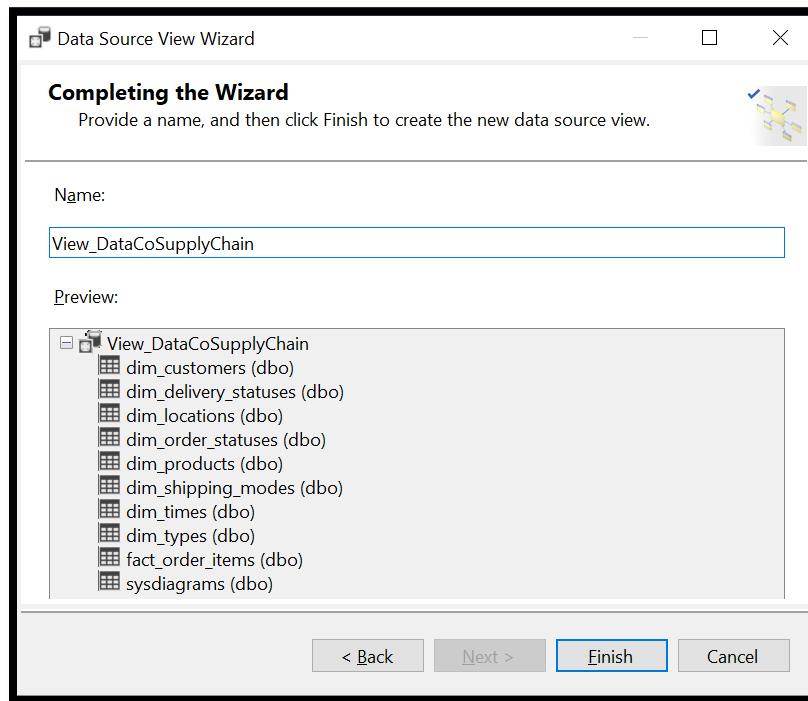


Bước 2: Hộp thoại Data Source View Wizard xuất hiện, chọn Next để tiếp tục.

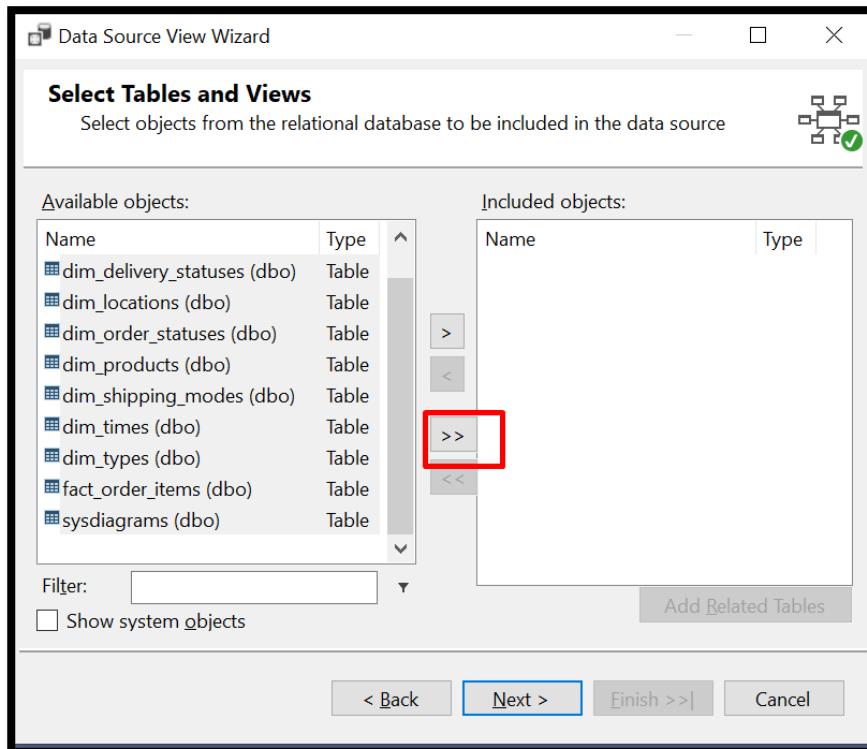


Bước 3: Chọn data source vừa tạo, sau đó chọn Next để tiếp tục.



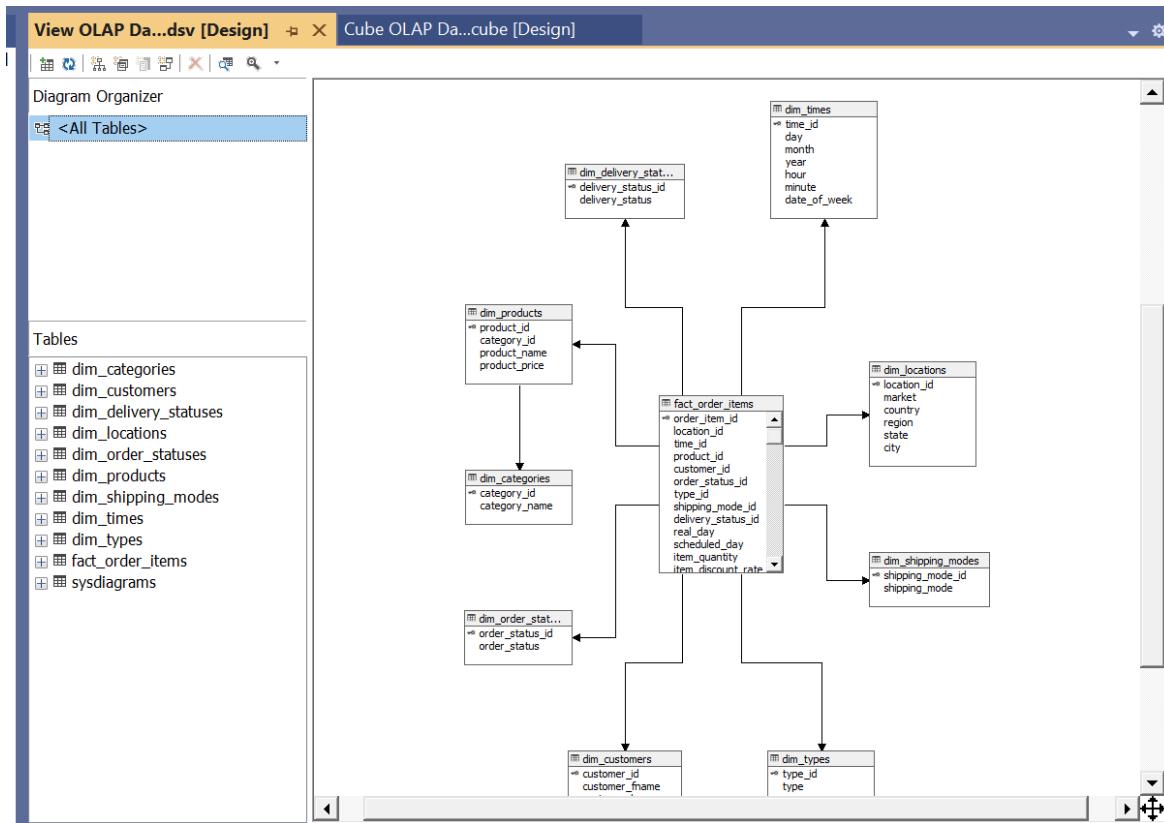


Bước 4: Chọn bảng Fact, sau đó chọn nút >> để thêm tất cả các bảng (tùy chọn bảng phù hợp nếu có) vào data source view.



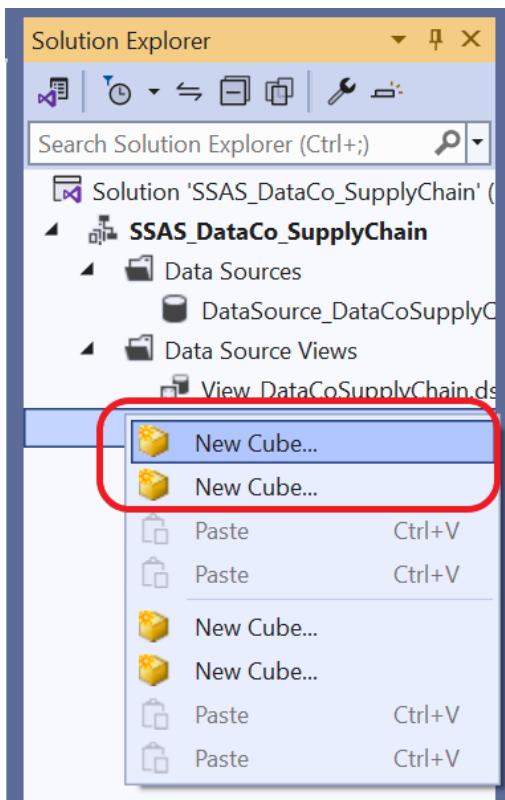
Bước 5: Chọn chọn tên cho views, Finish để hoàn tất quá trình.

Sau khi kết thúc quá trình này, ta sẽ được data source view như hình sau

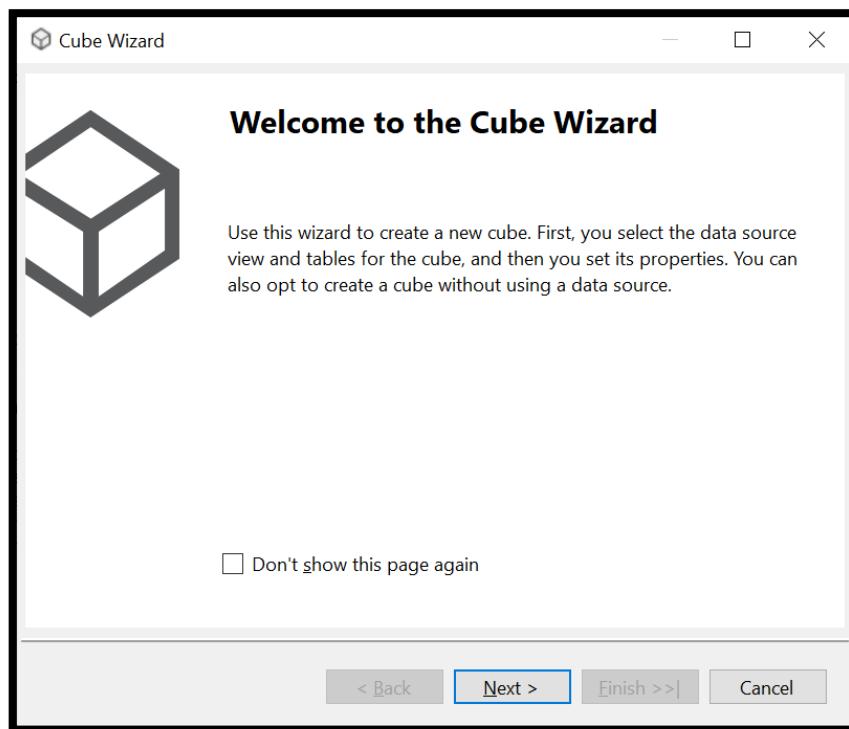


3.5. Xây dựng khối (Cubes) và xác định các độ đo (Measures).

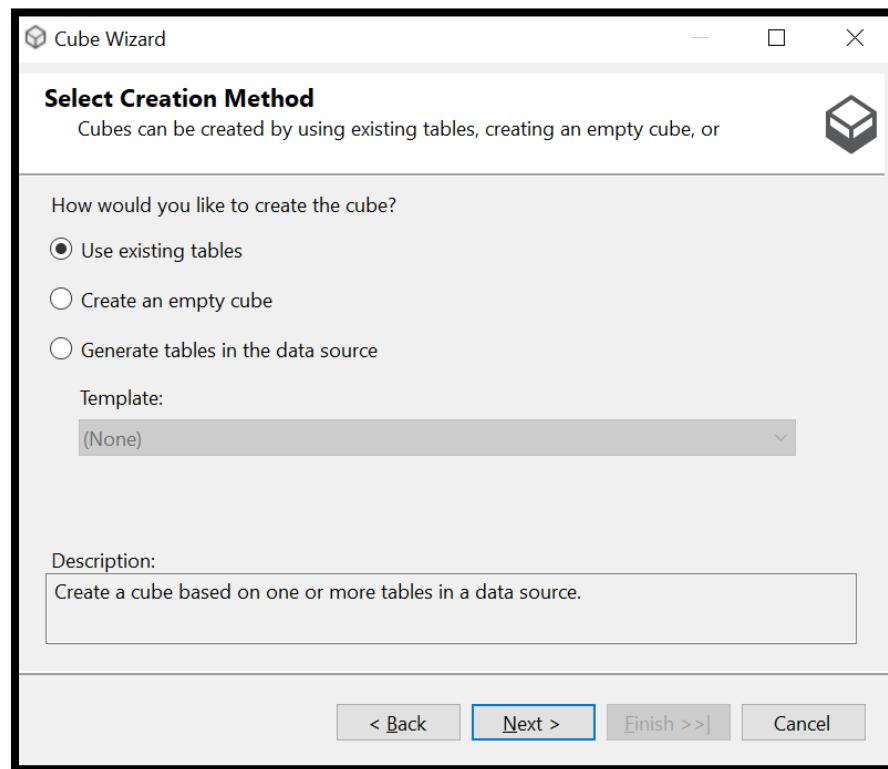
Bước 1: Tại Solution Explorer, ta click chuột phải vào thư mục Cubes và chọn New Cube.



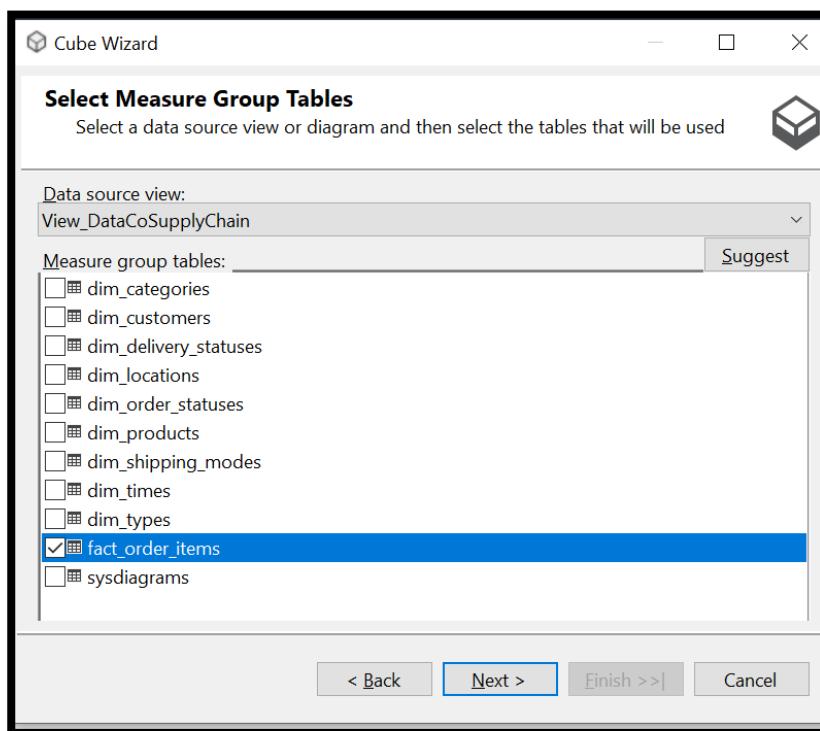
Bước 2: Hộp thoại Cube Wizard xuất hiện, chọn Next để tiếp tục.



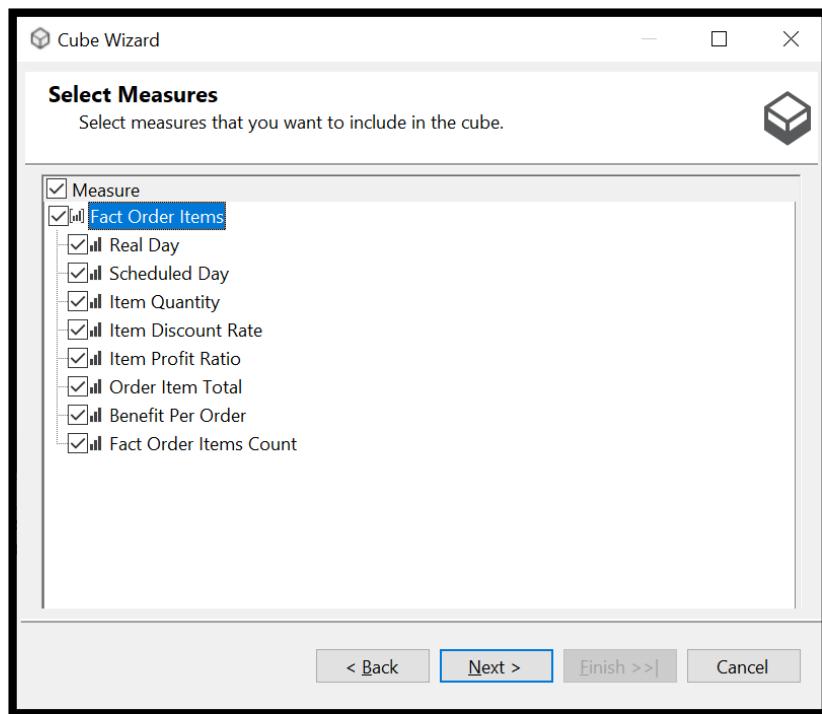
Bước 3: Chọn use existing tables, sau đó chọn Next để tiếp tục.



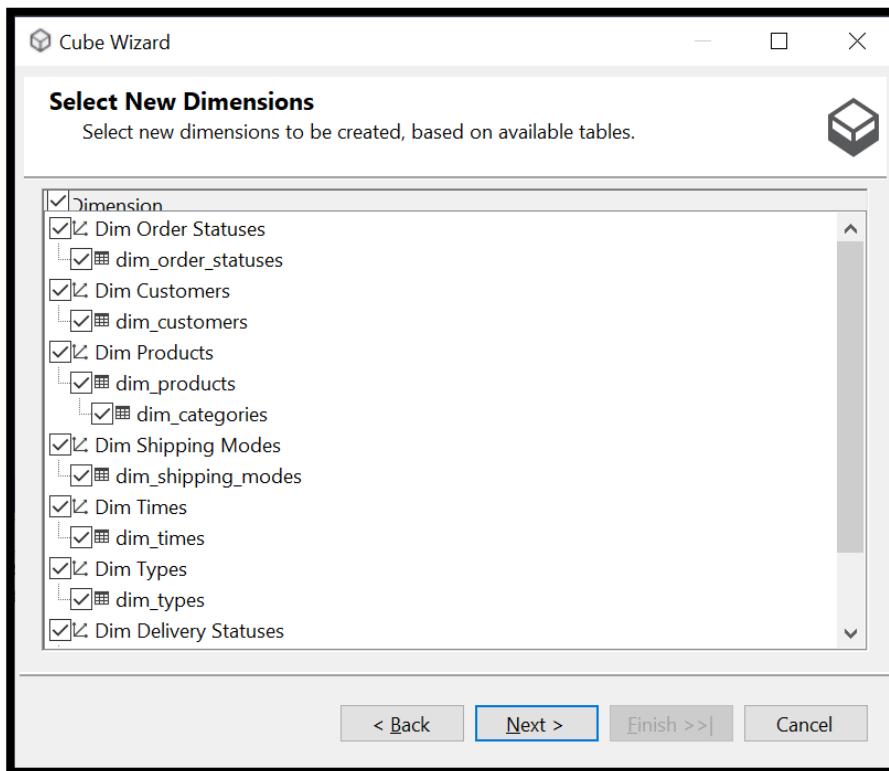
Bước 4: Chọn Fact để phân chia các measure group.



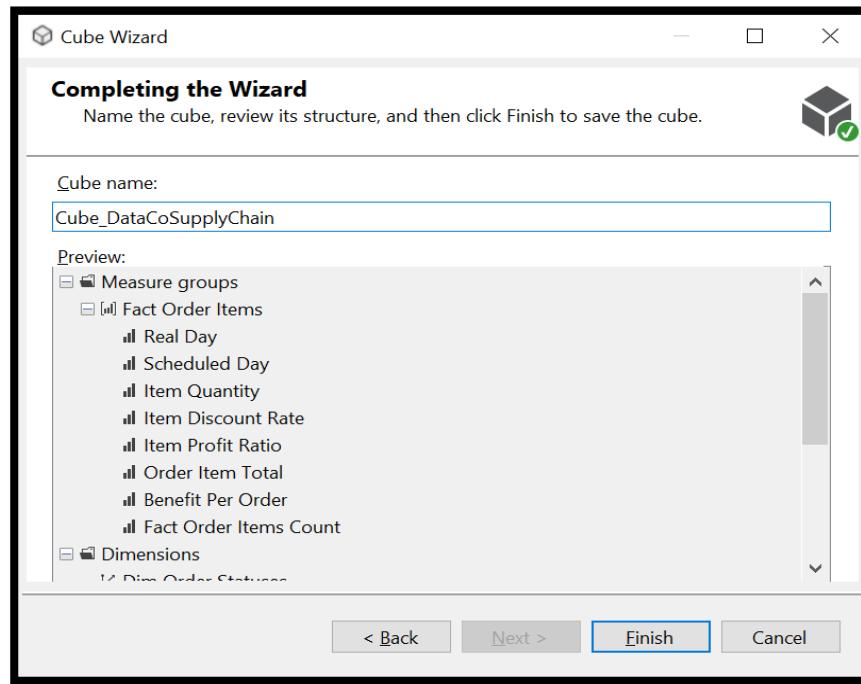
Bước 5: Chọn những độ đo để xuất, sau đó chọn Next để tiếp tục.



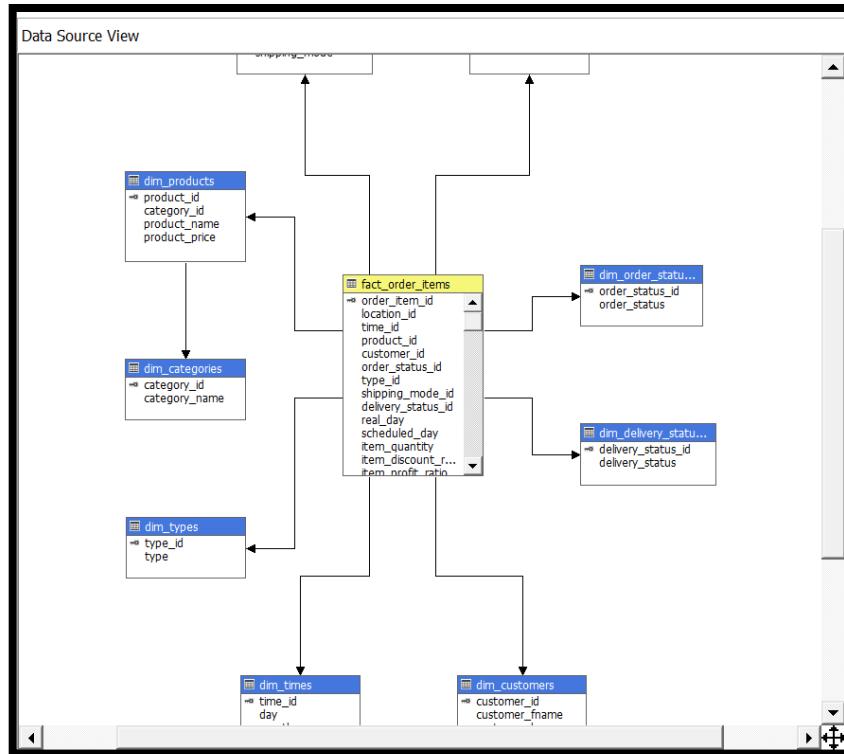
Bước 6: Chọn danh sách các bảng Dimension, sau đó chọn Next để tiếp tục.



Bước 7: Next, nhập tên và chọn Finish để hoàn tất quy trình xây dựng các khối (Cubes) và xác định các độ đo (Measures).

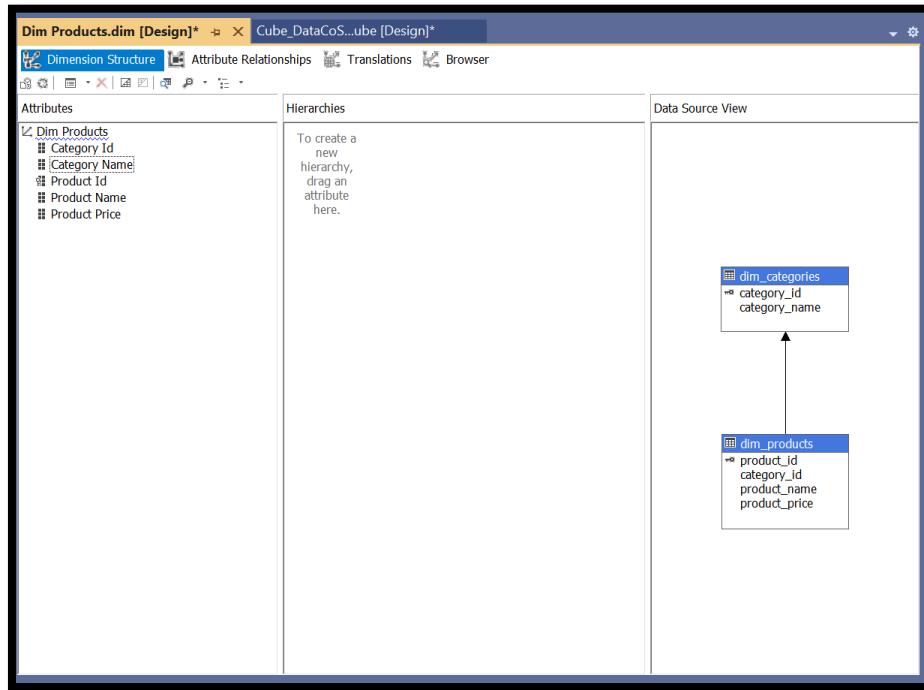


Sau khi kết thúc quá trình này, ta sẽ được kết quả như hình sau:

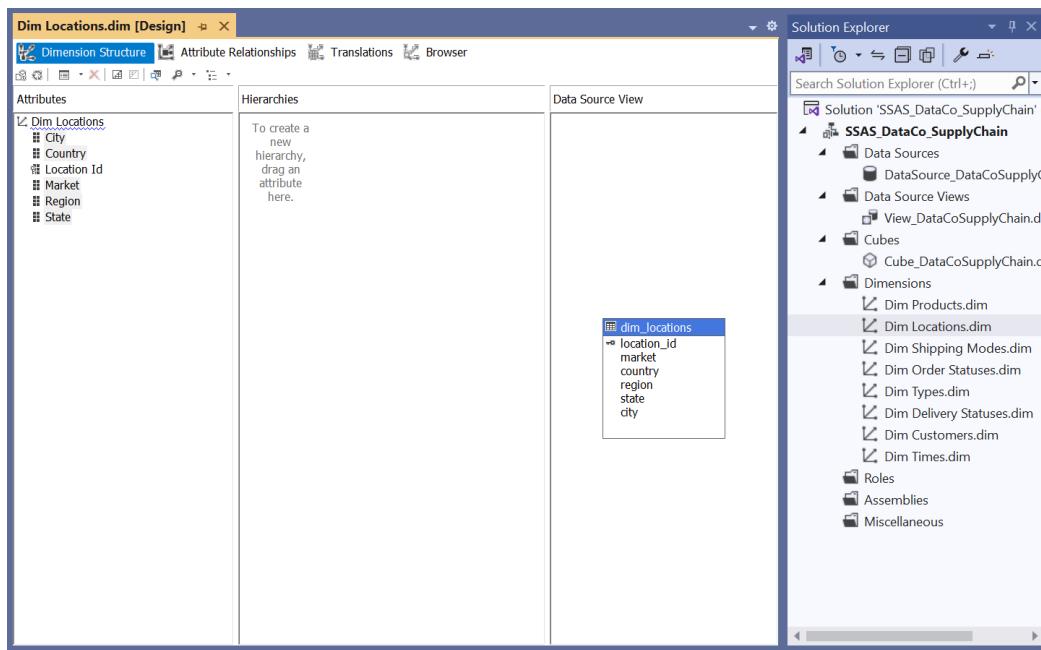


3.6. Xác định các chiều (Dimensions)

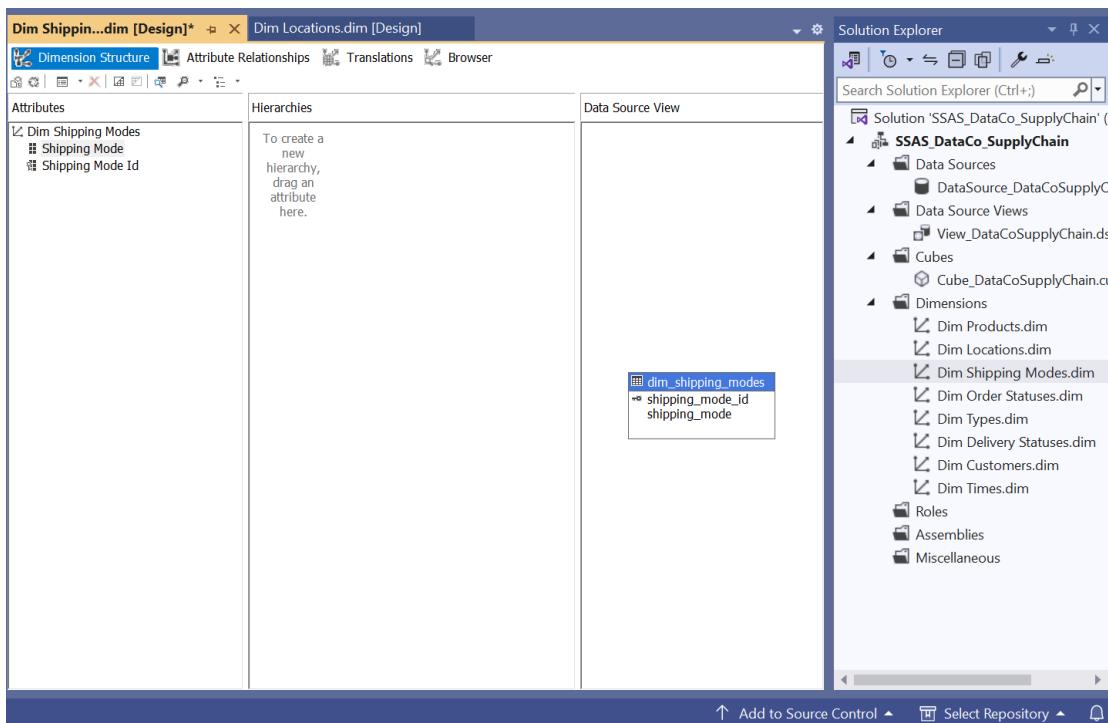
Bước 1: Tại folder **Dimensions** trong **Solution Explorer**, ta chọn Dim **Product.dim**. Giữ Ctrl để chọn nhiều thuộc tính, kéo thả các thuộc tính **category_id**, **category_name**, **product_name**, **product_price** vào Attributes.



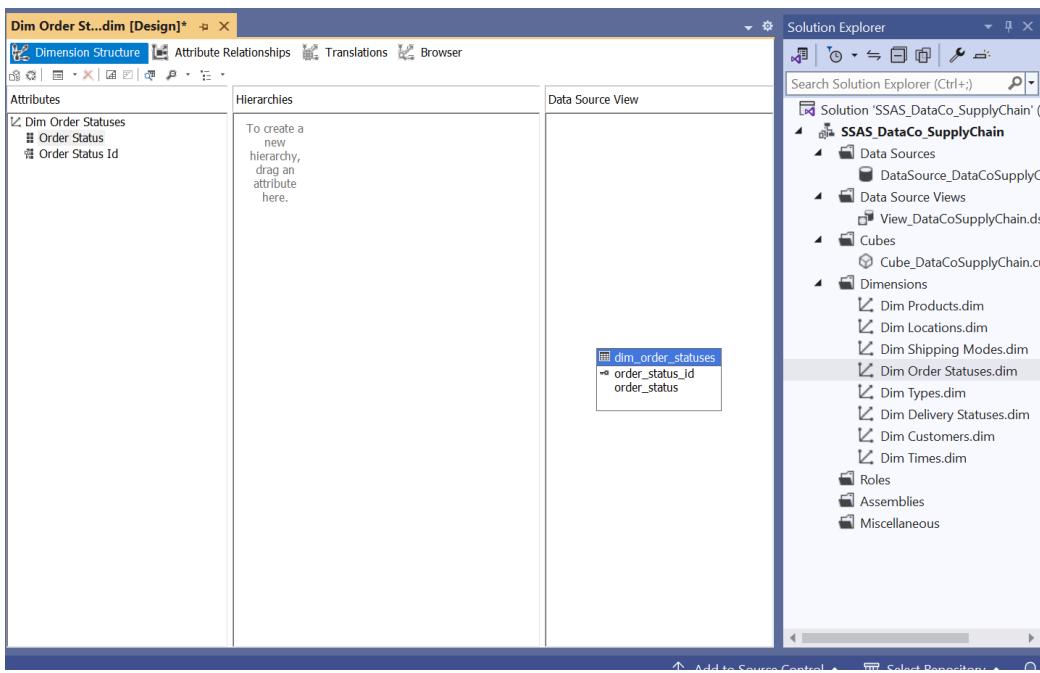
Bước 2: Tại folder Dimensions trong Solution Explorer, ta chọn Dim **Location.dim**. Giữ Ctrl để chọn nhiều thuộc tính, kéo thả các thuộc tính **market**, **country**, **region**, **state**, **city** vào Attributes.



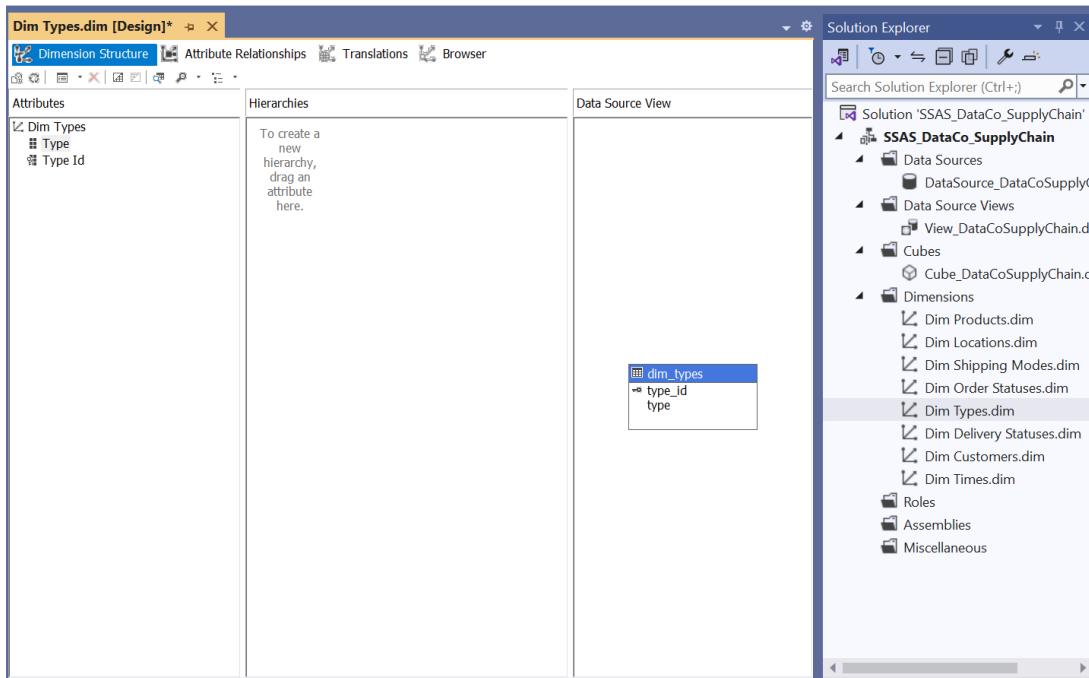
Bước 3. Tại folder Dimensions trong Solution Explorer, ta chọn **Dim Shipping Modes.dim**. Giữ Ctrl để chọn nhiều thuộc tính, kéo thả các thuộc tính **shipping mode** vào Attributes.



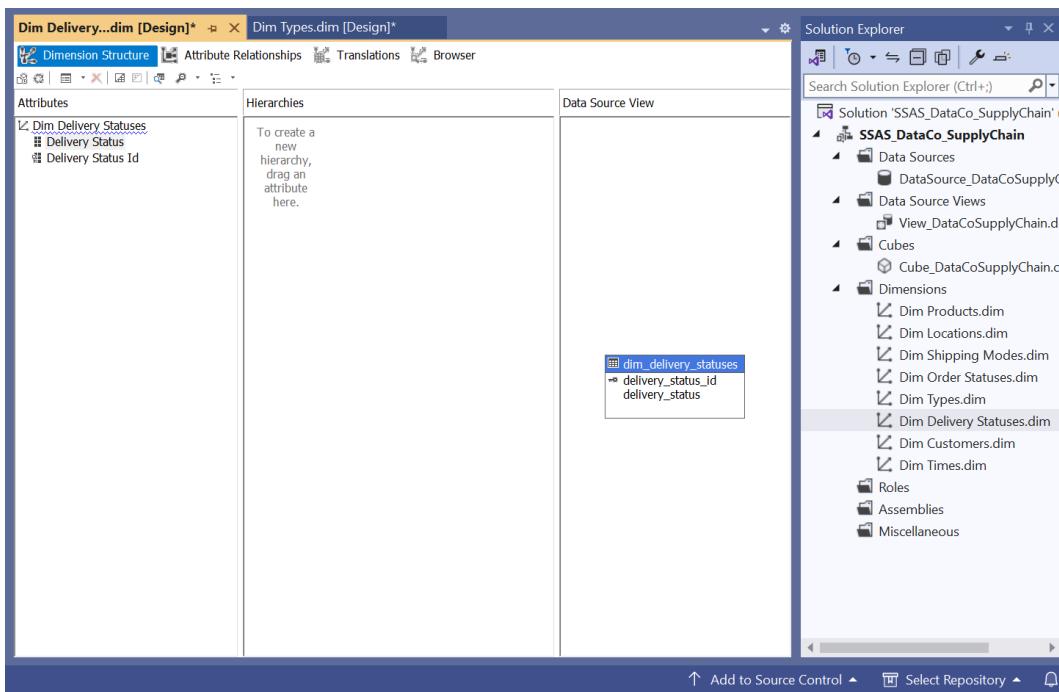
Bước 4: Tại folder Dimensions trong Solution Explorer, ta chọn **Dim Order Statuses.dim**. Giữ Ctrl để chọn nhiều thuộc tính, kéo thả các thuộc tính **order statuses** vào Attributes.



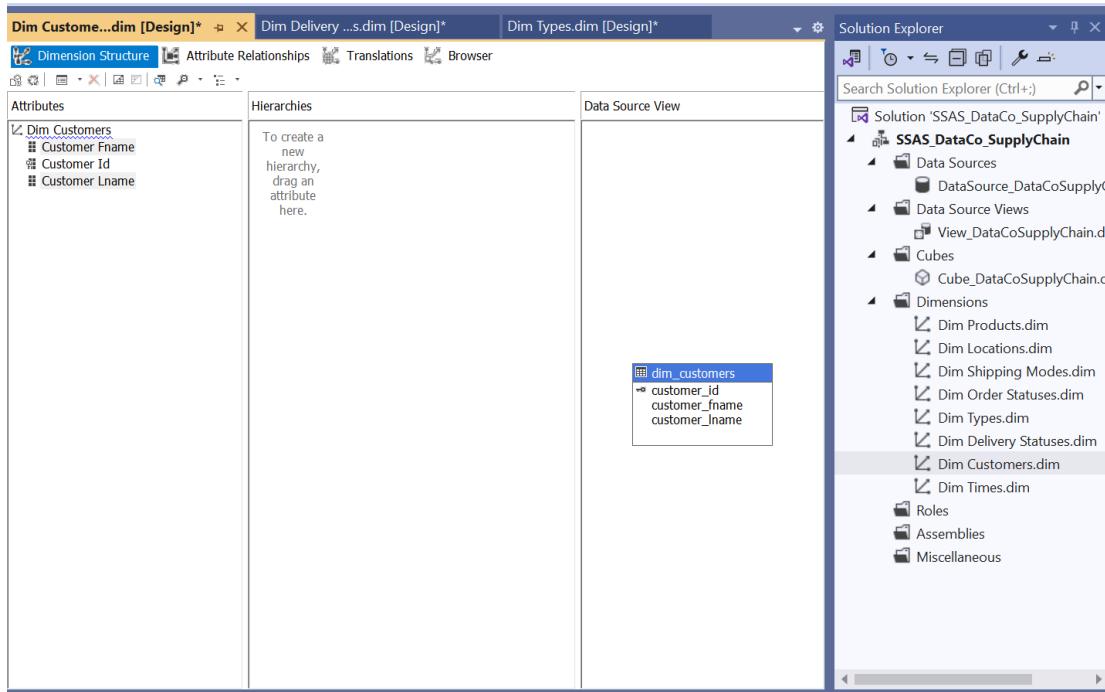
Bước 5: Tại folder Dimensions trong Solution Explorer, ta chọn **Dim Order Types.dim**. Giữ Ctrl để chọn nhiều thuộc tính, kéo thả các thuộc tính **order status** vào Attributes.



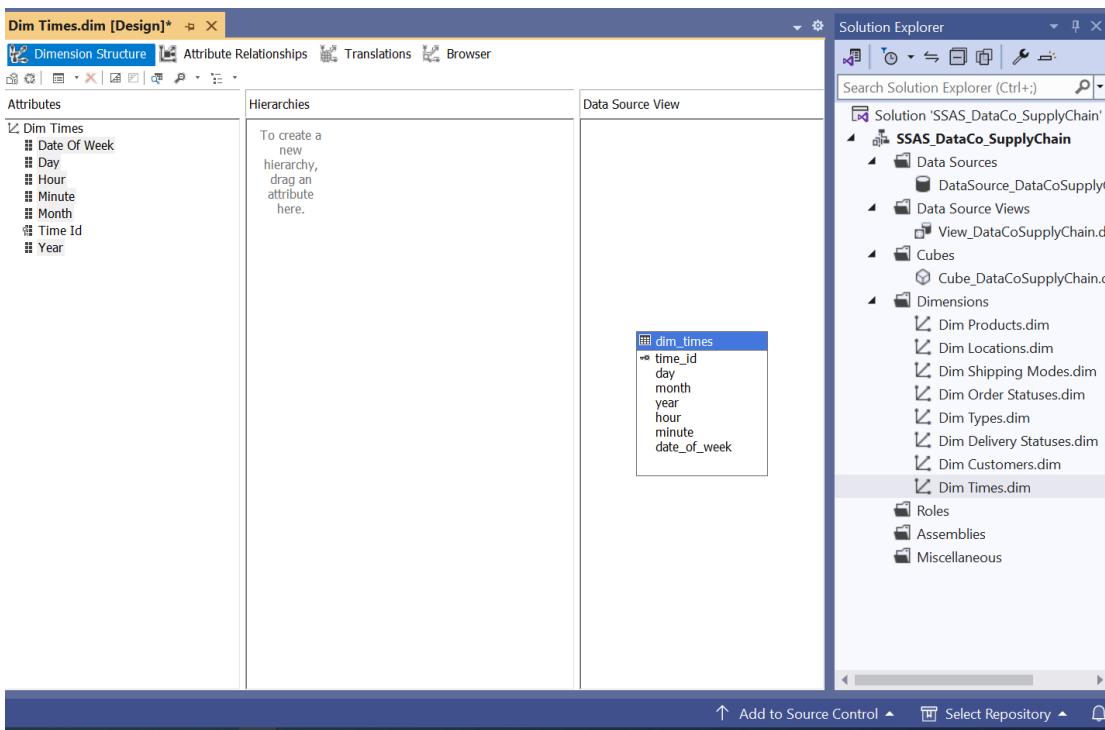
Bước 6: Tại folder Dimensions trong Solution Explorer, ta chọn **Dim Delivery Statuses.dim**. Giữ Ctrl để chọn nhiều thuộc tính, kéo thả các thuộc tính **delivery statuses** vào Attributes.



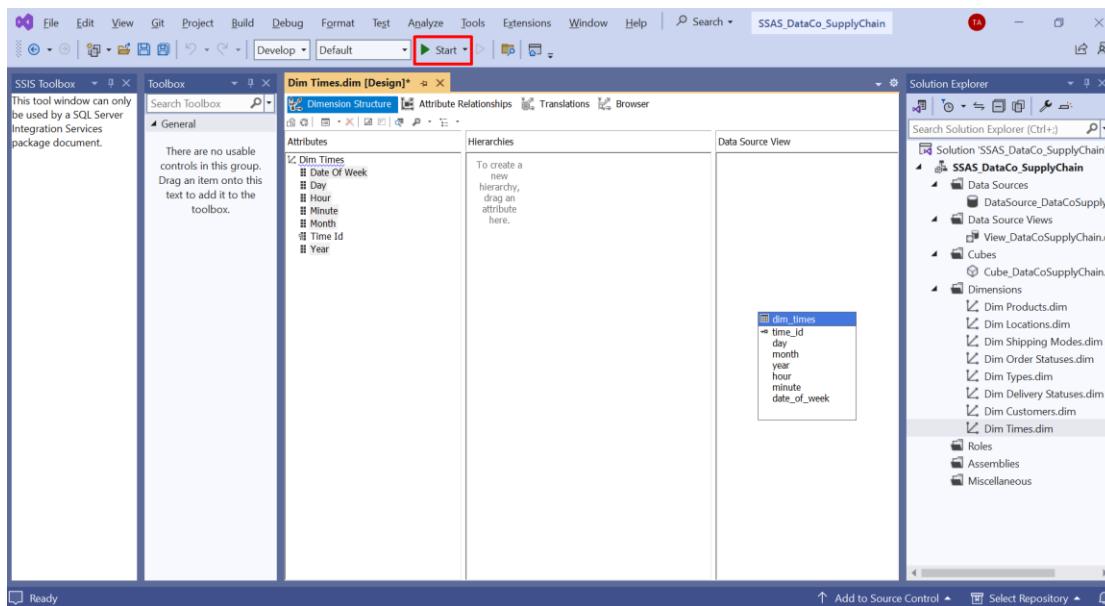
Bước 7: Tại folder Dimensions trong Solution Explorer, ta chọn Dim **Customers.dim**. Giữ Ctrl để chọn nhiều thuộc tính, kéo thả các thuộc tính **customer_fname, customer_lname** vào Attributes.



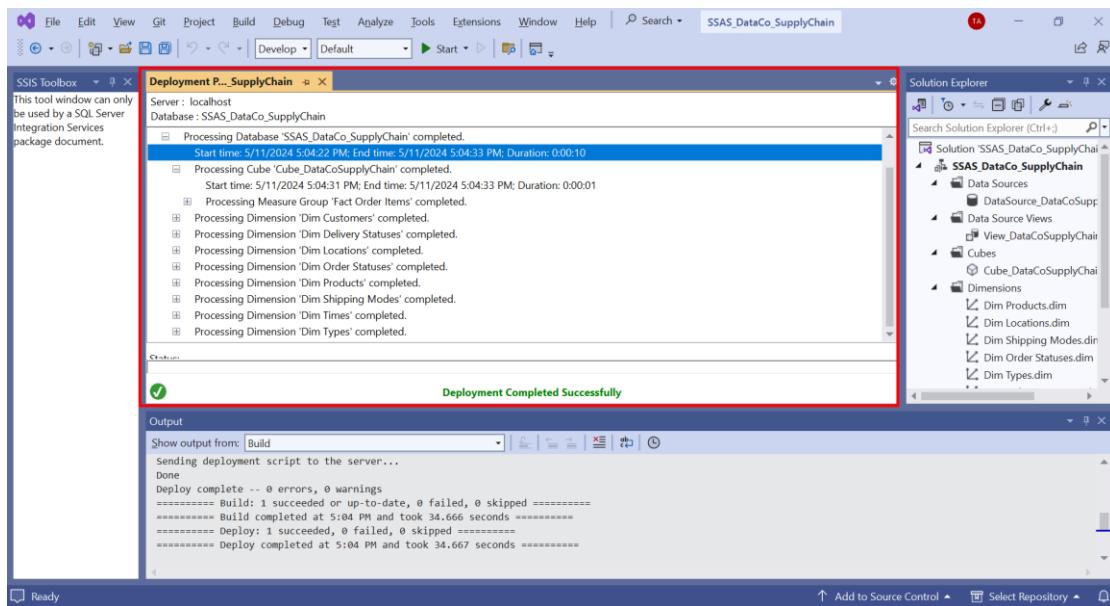
Bước 8: Tại folder Dimensions trong Solution Explorer, ta chọn Dim **Time.dim**. Giữ Ctrl để chọn nhiều thuộc tính, kéo thả các thuộc tính **day, month, year, hour, minute** vào Attributes.



Bước 10: Ta chọn Start để deploy project. Mục đích để kiểm tra xem SSAS cho đến hiện tại có gặp lỗi không.



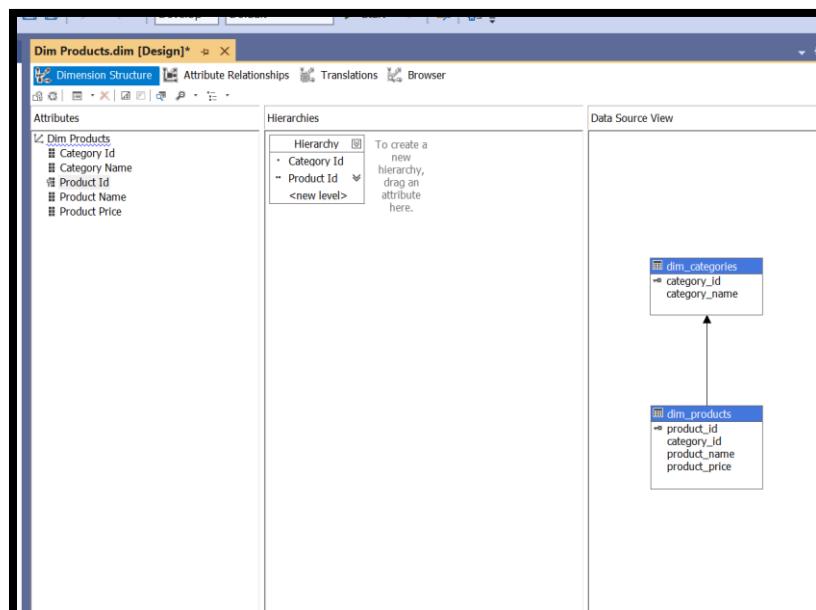
Kết quả



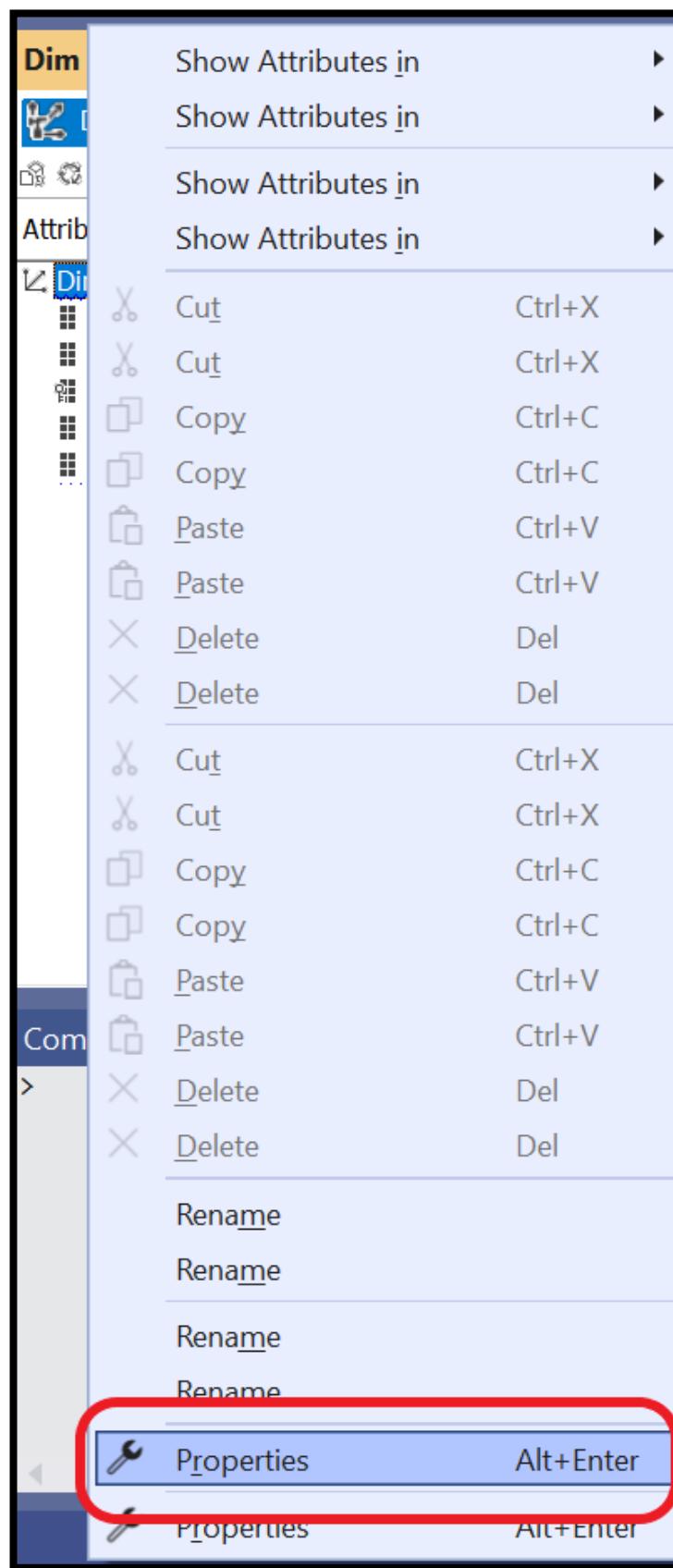
Cấu hình deployment. Cấu hình các dim và tạo các thuộc tính phân cấp (hierarchy)

3.6.1. Cấu hình deployment. Cấu hình và tạo thuộc tính phân cấp cho “Dim Products.dim”

Bước 1: d-click vào Dim Product.dim, kéo thả lần lược theo thứ tự từ cao đến thấp (trên xuống dưới) Category Id, Product Id vào Hierarchy panel. Đổi tên Hierarchy thành category_product.



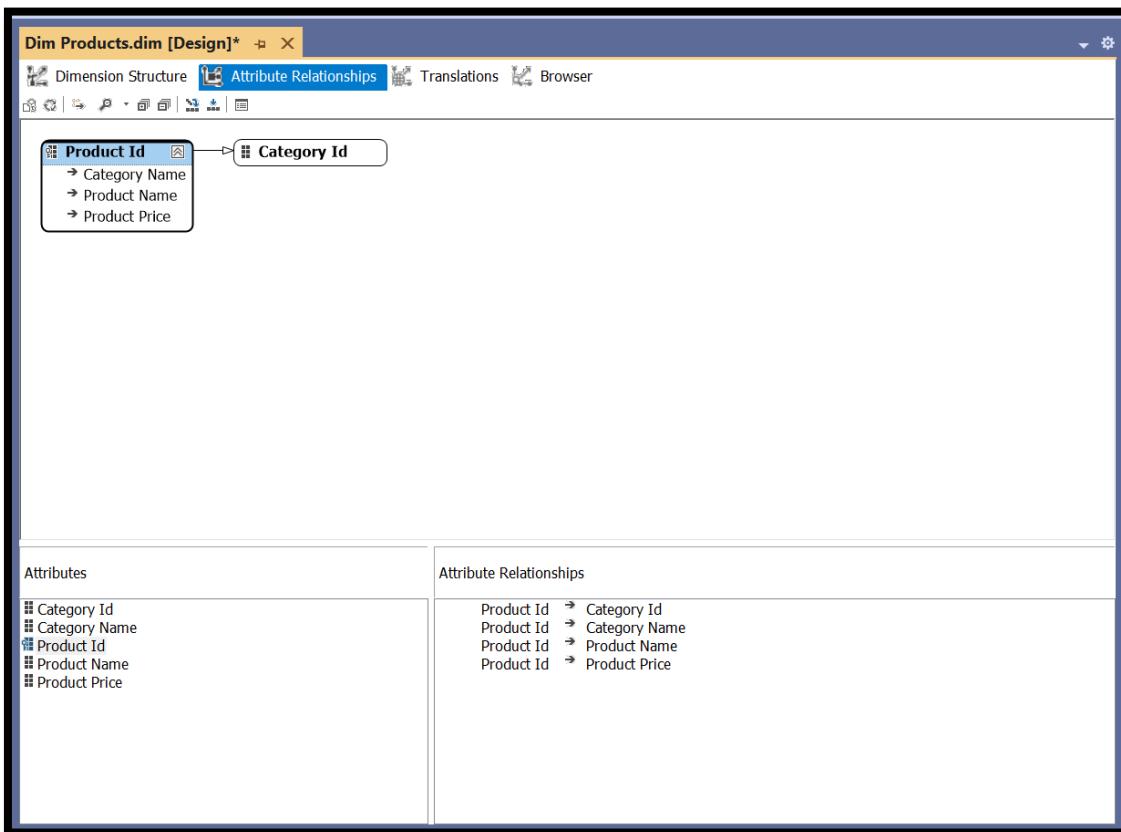
Bước 3: R-click vào Dim Products bên trong Attribute panel, chọn Properties.



Bước 4: tại UnknownMember chọn Hidden, tại Type, chọn Product.

UnknownMember	Hidden
UnknownMemberName	
WriteEnabled	False
Basic	
Description	
ID	Dim Products
Name	Dim Products
Type	Products

Bước 5: Click **Attribute Relationships** để mở bảng thuộc tính phân cấp, kéo thả để phân cấp các thuộc tính từ trái sang phải (nhỏ nhất bên trái) Product Id -> Category Id.



Bước 6: Cấu hình các thuộc tính con.

Bước 6.1: Cấu hình cho Product

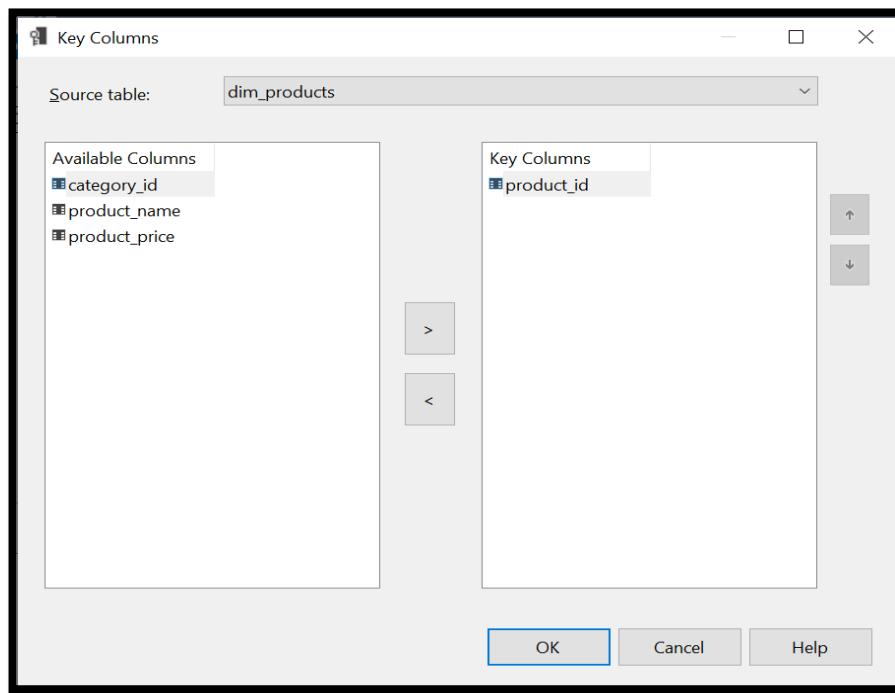
Bước 6.1.1 R-click vào **Product Id** để mở bảng properties



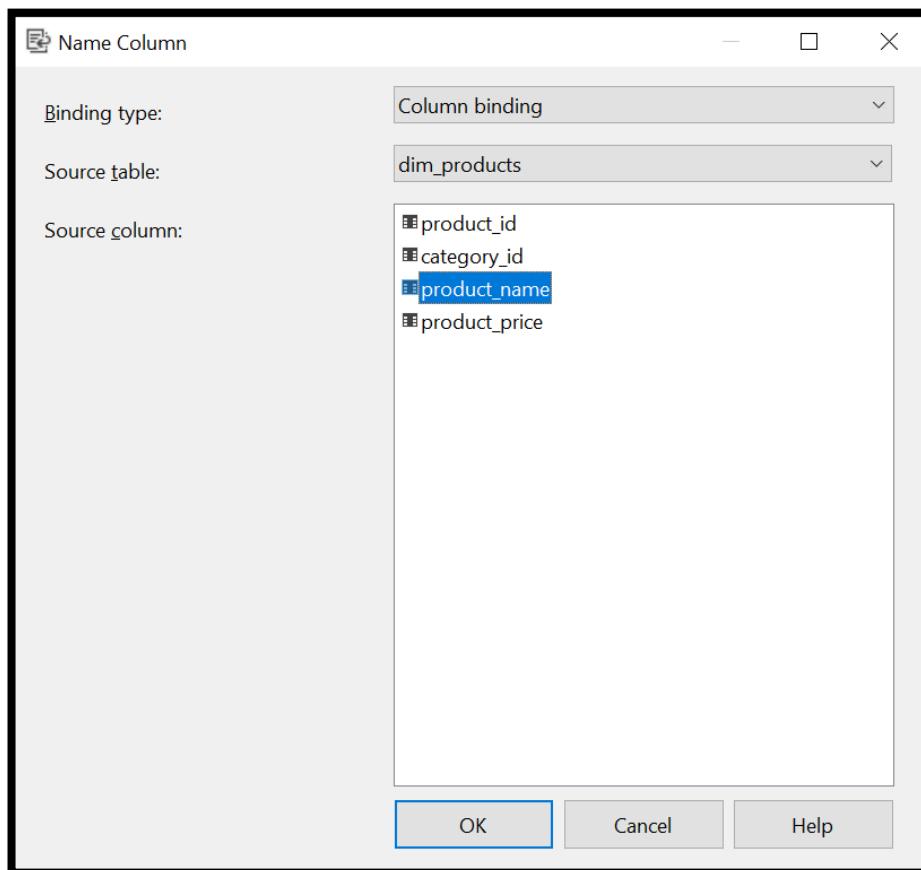
Bước 6.1.2 Trong bảng properties của Product Id, tại OrderBy chọn Key, tại Type chọn Product

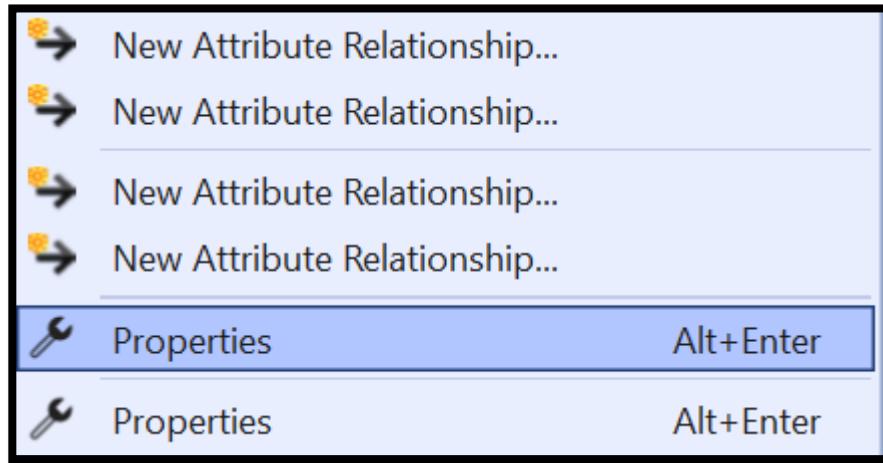
OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompressionH	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	Product Id
Name	Product Id
Type	Product
Usage	Key

Bước 6.1.3: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Product. Product phát triển theo mô hình bông tuyết (snowflake schema) sẽ lấy khóa cột là chính nó như hình bên dưới. Sau đó OK để kết thúc chọn cột cho Product



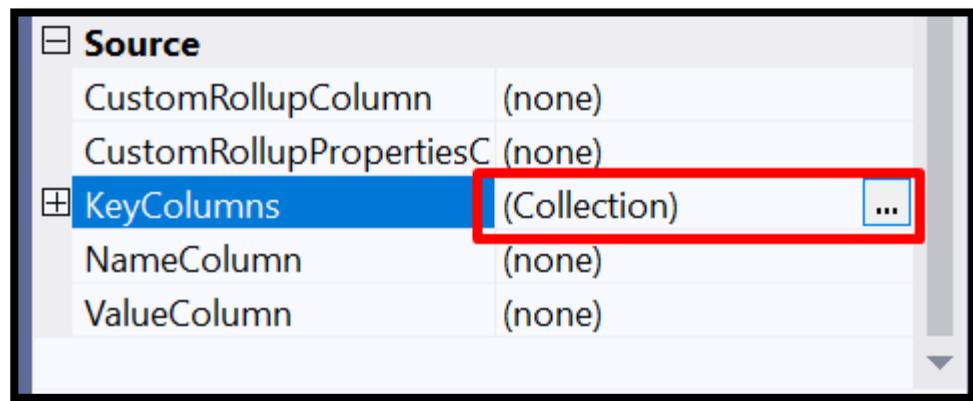
Bước 6.1.4: Tại cửa sổ Properties của thuộc tính Product, ta chọn Name Column và chọn tên thuộc tính là Product Name. OK để kết thúc



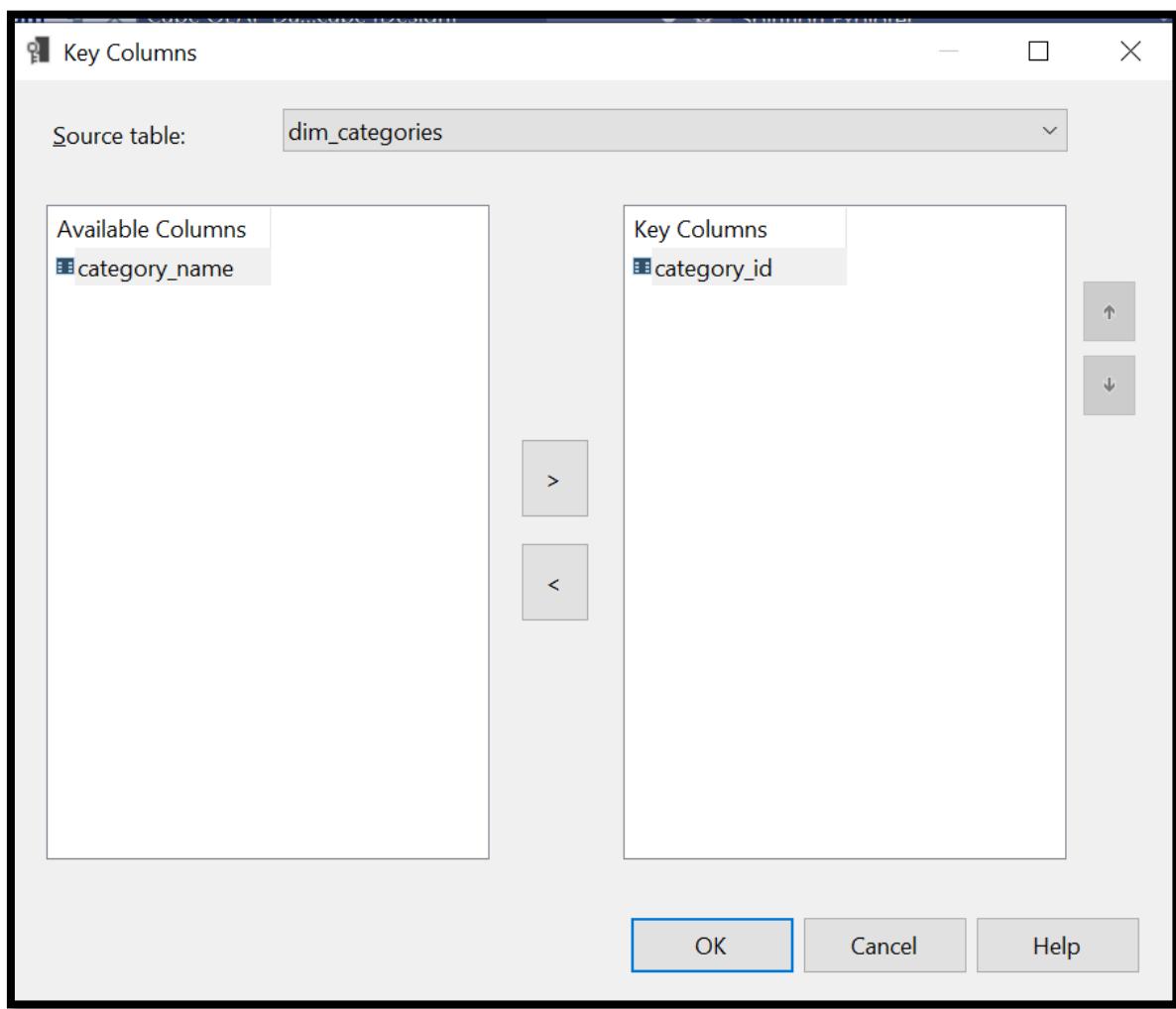
Bước 6.2: Cấu hình cho Category**Bước 6.2.1** R-click vào Category Id để mở bảng properties của Category Id**Bước 6.2.2:** Trong bảng properties của Category Id, tại OrderBy chọn Key, tại Type chọn ProductCategory

	OrderBy	Key
	OrderByAttribute	
ProcessingState		Unprocessed
TokenizationBehavior		TokenizationNone
UserEditFlag		0
VertipaqCompressionH		VertipaqAutomatic
Basic		
Description		
FormatString		
ID		Category Id
Name		Category Id
Type		ProductCategory

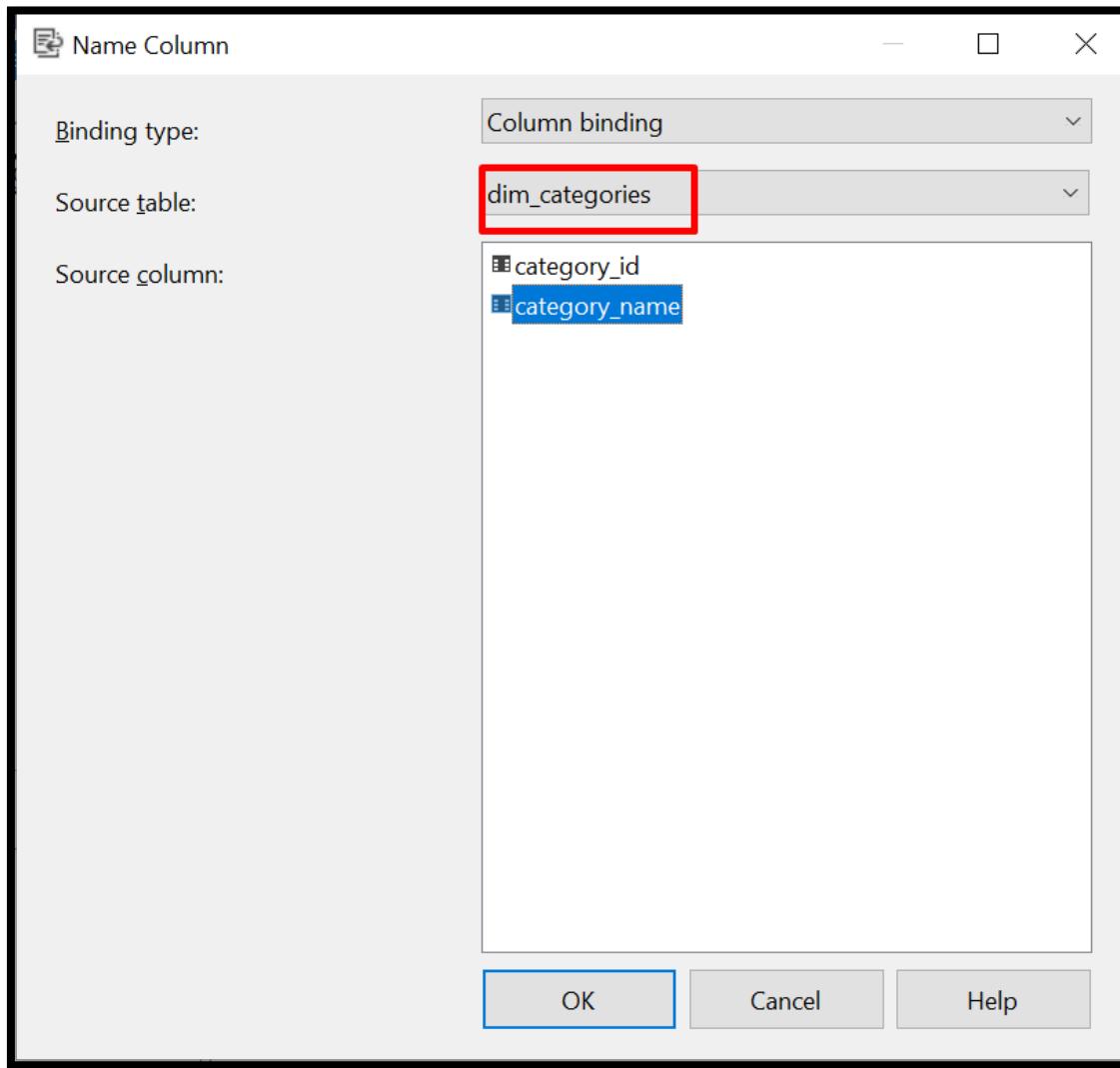
Bước 6.2.3: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Category. Lấy khóa cột cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho Category



Lấy Category

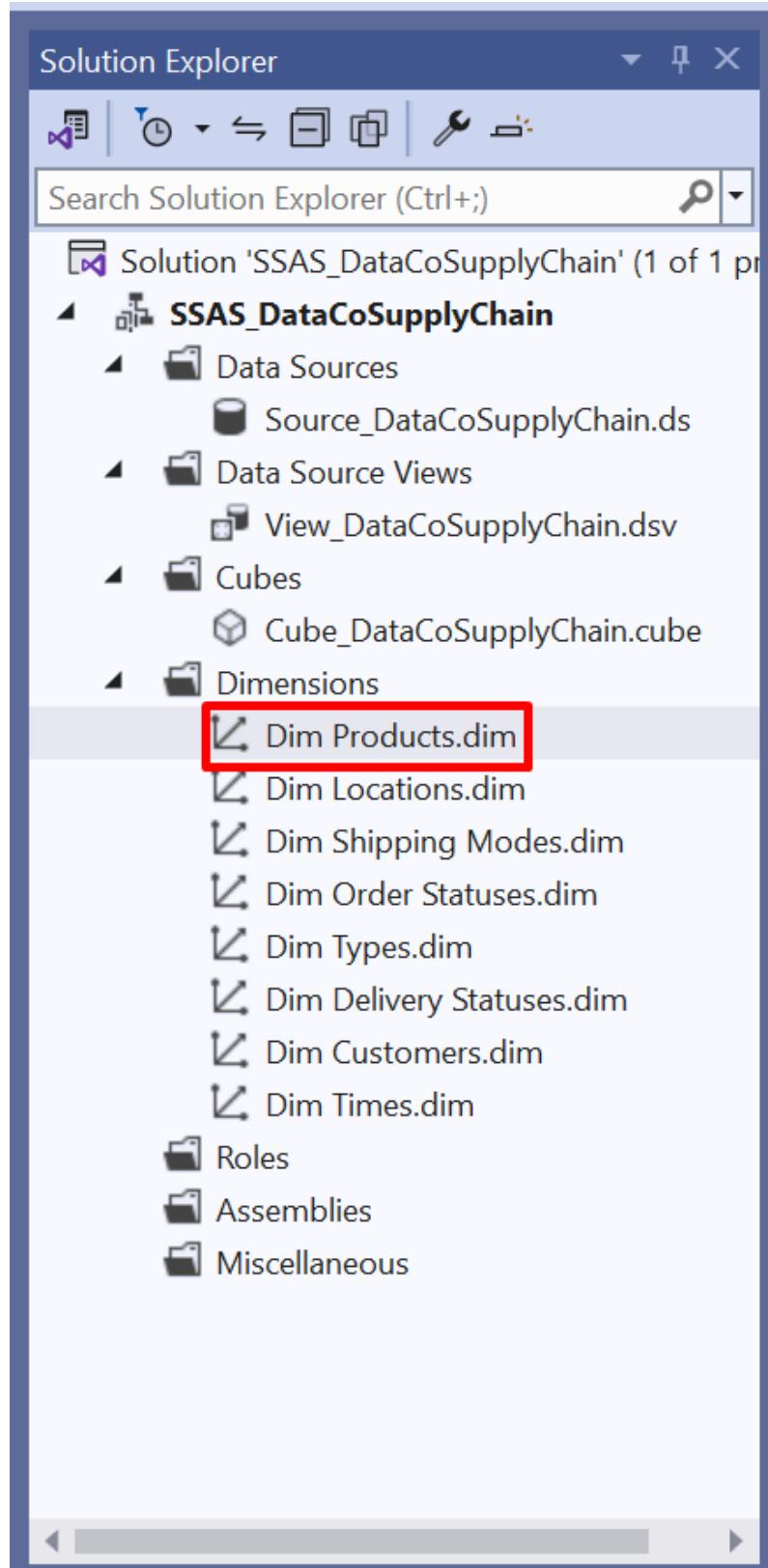


Bước 6.2.4: Tại cửa sổ Properties của thuộc tính Category, ta chọn Name Column và chọn tên thuộc tính là Category Name. OK để kết thúc

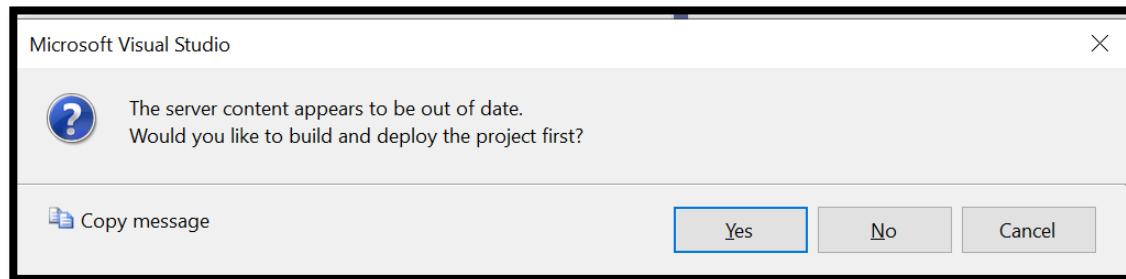


Bước 7: Tạo Process Dim Products.dim

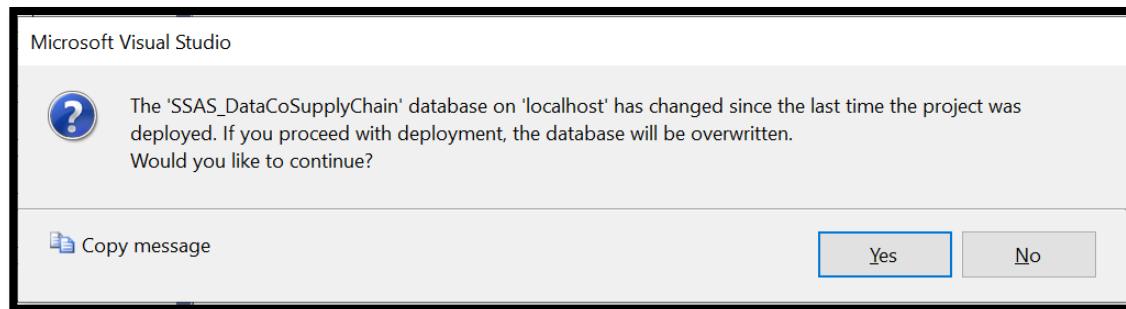
Tại Solution Explorer, r-click vào **Dim Products.dim** và chọn **Process** để chạy quá trình tạo dim product.



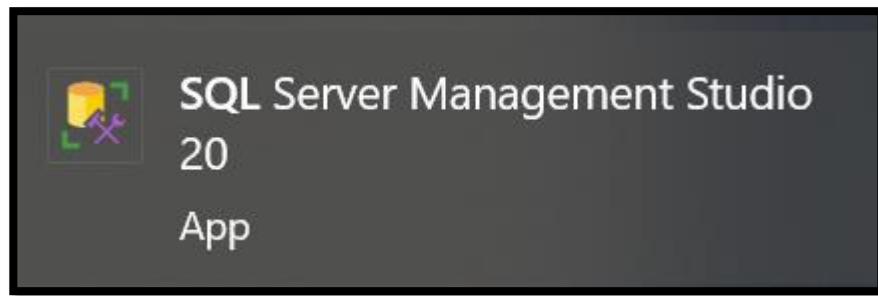
Hệ thống thông báo project bị outdated (Tức là ta đã thao tác, và project bị thay đổi, cần khởi động lại) chọn Yes.



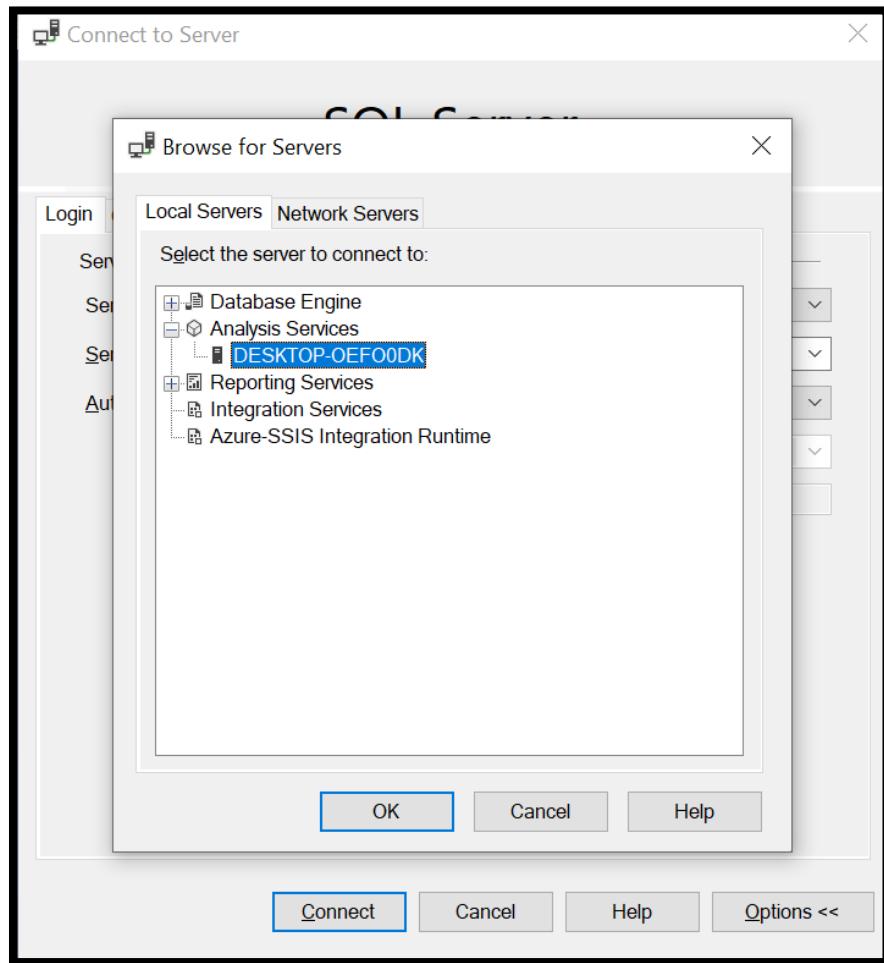
Hệ thống thông báo chúng ta đang deploy database tại localhost, nên dữ liệu sẽ không được ghi lại, ta cần chuyển sang SSAS server (lưu ý tại đây là Analysis services tức “SSAS” chứ không phải là Database Engine tức “SSIS”)



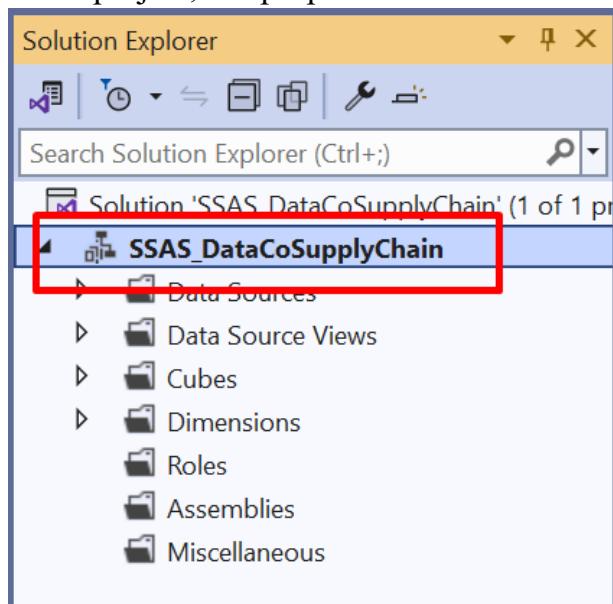
Ta chọn **No** để thoát ra. Sau đó mở SQL Server Management Studio



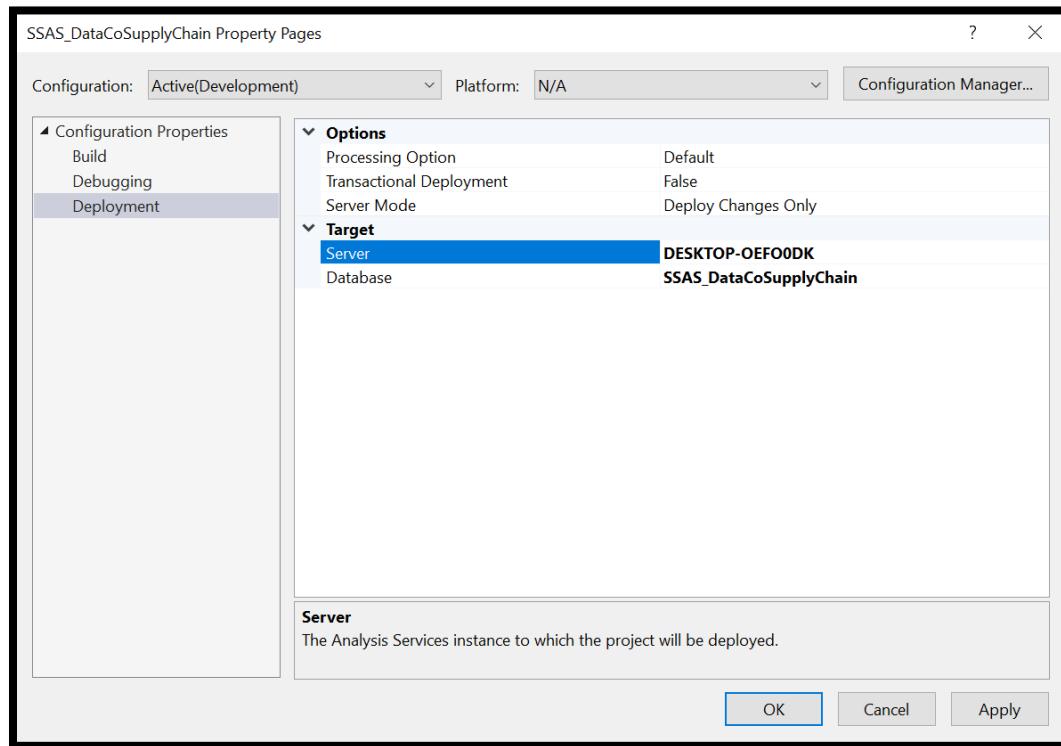
Lấy server name của Analysis services



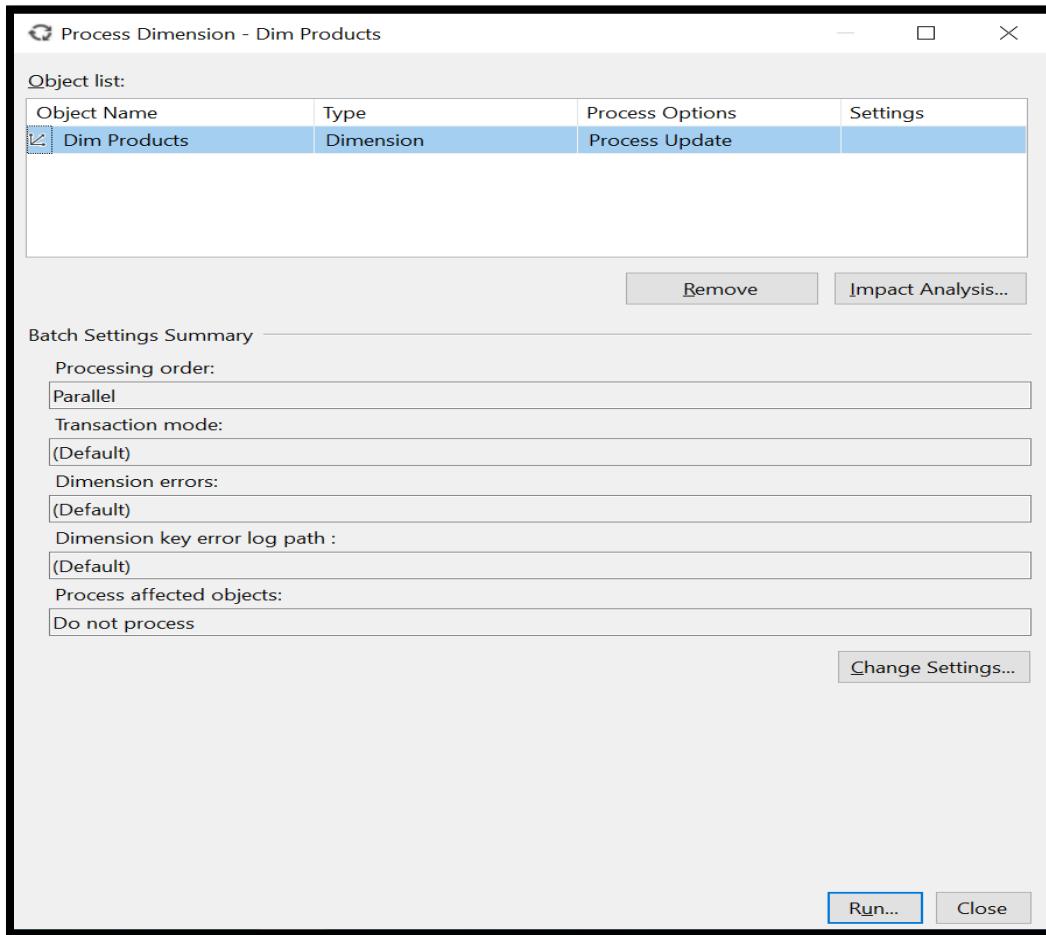
Trở lại SSAS, r-click vào project, mở properties



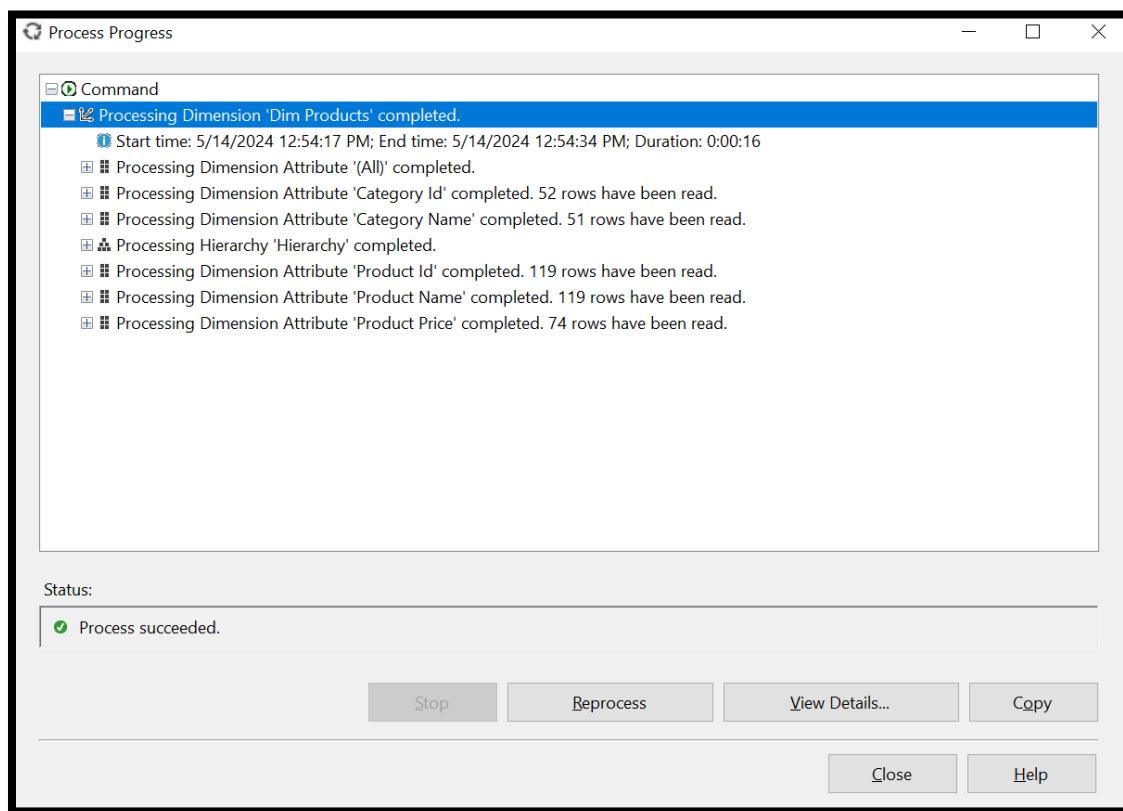
Chuyển sang tab Deployment, tại Server ta dùng bàn phím điền tên Analyst Services vào.



Bước 8: Run process lại Dim Product.dim 1 lần nữa tương tự bước 7



Kết quả



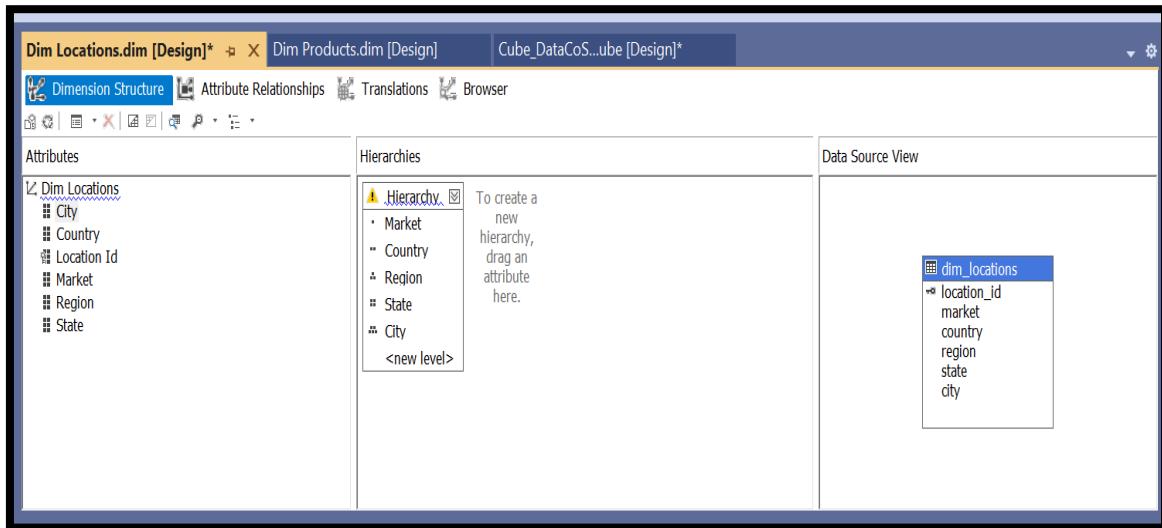
Mở lại Dim Products.dim, chọn tab Browser, kiểm tra Hierarchy

The screenshot shows the 'Dim Products.dim [Design]' window with the 'Browser' tab selected. The hierarchy is set to 'category_product'. The tree view displays the following hierarchy levels:

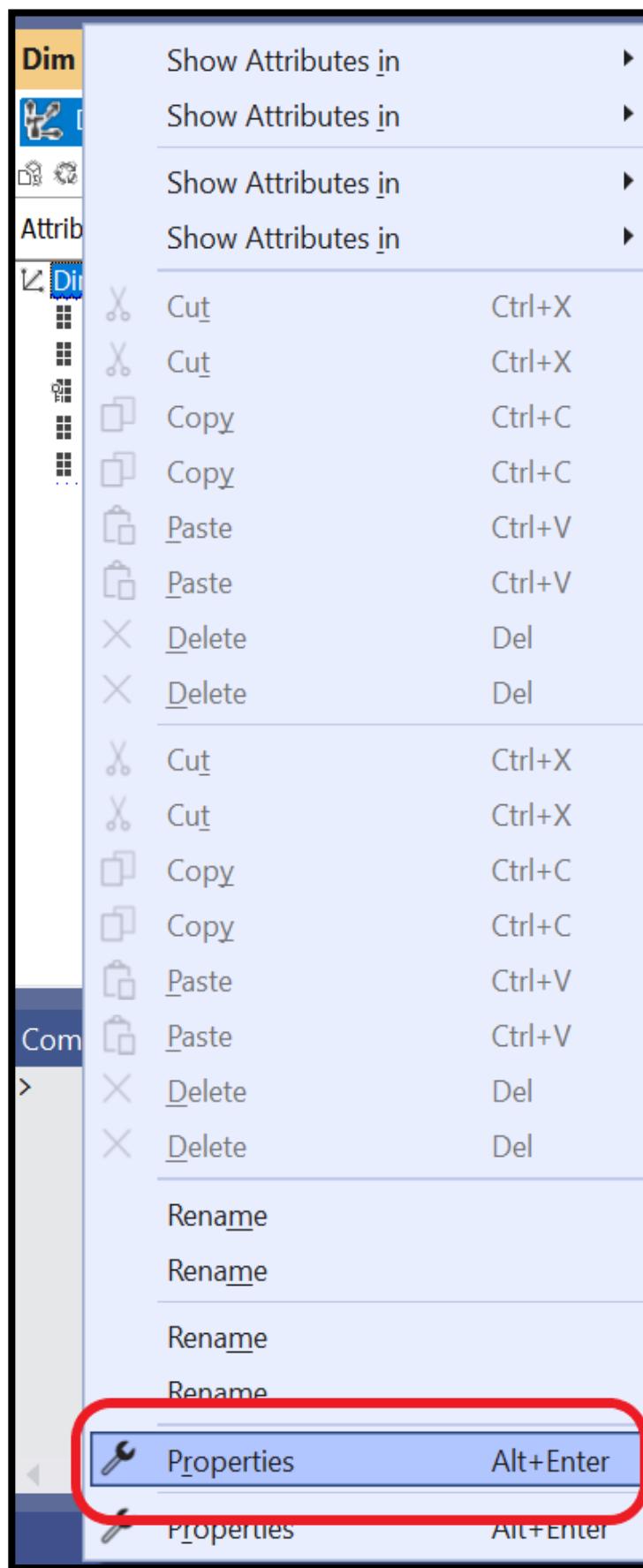
- All
 - Soccer
 - Baseball & Softball
 - Basketball
 - Lacrosse
 - Nike Kids' Grade School KD VI Basketball Shoe
 - Under Armour Men's Tech II T-Shirt
 - Tennis & Racquet
 - Hockey
 - Stiga Master Series ST3100 Competition Indoor
 - Nike Women's Legend V-Neck T-Shirt
 - Nike Dri-FIT Crew Sock 6 Pack
 - Cardio Equipment
 - Strength Training
 - Fitness Accessories
 - Boxing & MMA
 - Electronics
 - As Seen on TV!
 - Cleats
 - Men's Footwear
 - Women's Apparel
 - Girls' Apparel
 - Shop By Sport
 - Men's Golf Clubs
 - Women's Golf Clubs
 - Golf Apparel
 - Golf Shoes
 - Golf Bags & Carts
 - Golf Gloves

3.6.2. Cấu hình và tạo thuộc tính phân cấp cho “Dim Locations.dim”

Bước 1: d-click vào Dim Locations.dim, kéo thả lần lược theo thứ tự từ cao đến thấp (trên xuống dưới) Market, Country, Region, State, City vào Hierarchy panel. Đổi tên Hierarchy thành location_hierarchy



Bước 2: R-click vào Dim Locations bên trong Attribute panel, chọn Properties.



Bước 3: tại UnknownMember chọn Hidden, tại Type, chọn Geography.

Properties

Dim Locations Dimension

Advanced

- AttributeAllMemberN
- CurrentStorageMode Molap
- CurrentStringStoresCo 1050
- DependsOnDimensions

ErrorConfiguration (custom)

- Language
- MiningModelID
- ProcessingGroup ByAttribute
- ProcessingMode Regular
- ProcessingPriority 0
- ProcessingRecommen None
- ProcessingState Unprocessed
- Source View_DataCoSupplyChai
- StringStoresCompatibl 1050
- UnknownMember **Hidden**
- UnknownMemberName
- WriteEnabled False

Basic

- Description
- ID Dim Locations
- Name Dim Locations
- Type **Geography**

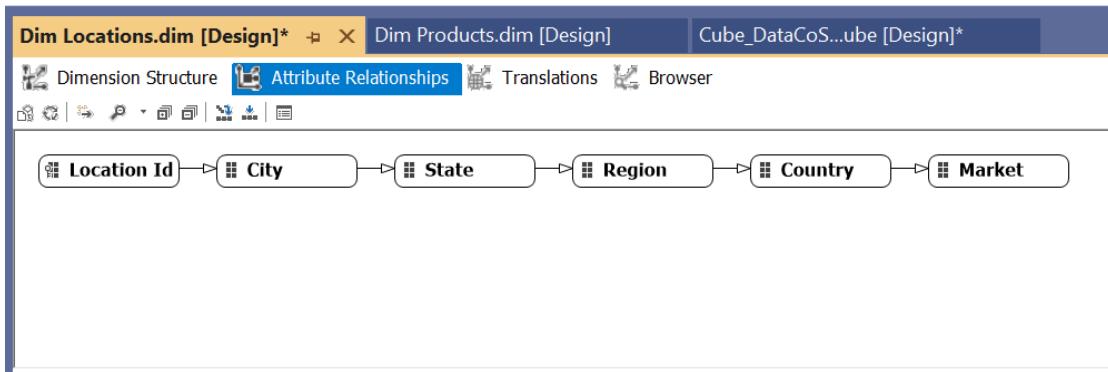
Misc

- Collation

Type

Specifies the type of information contained by the dimension.

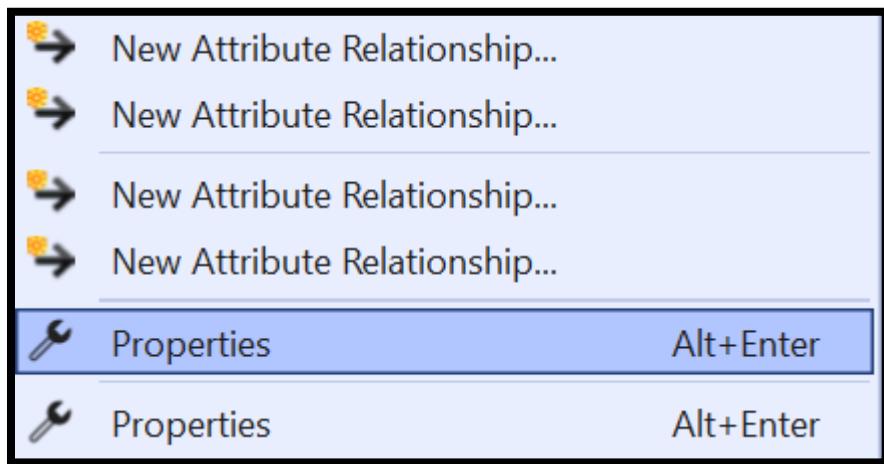
Bước 4: Click **Attribute Relationships** để mở bảng thuộc tính phân cấp, click giữ và kéo thả thuộc tính nhỏ hơn vào thuộc tính lớn hơn để phân cấp các thuộc tính từ trái sang phải (nhỏ nhất bên trái)



Bước 6: Cấu hình các thuộc tính con.

Bước 6.1: Cấu hình cho City

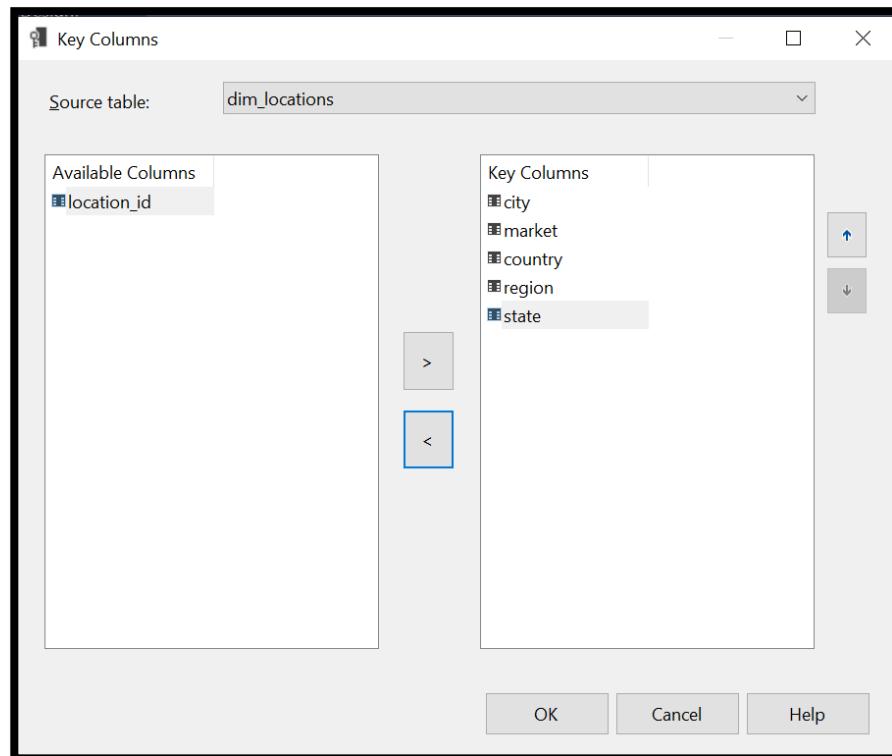
Bước 6.1.1: R-click vào City để mở bảng properties



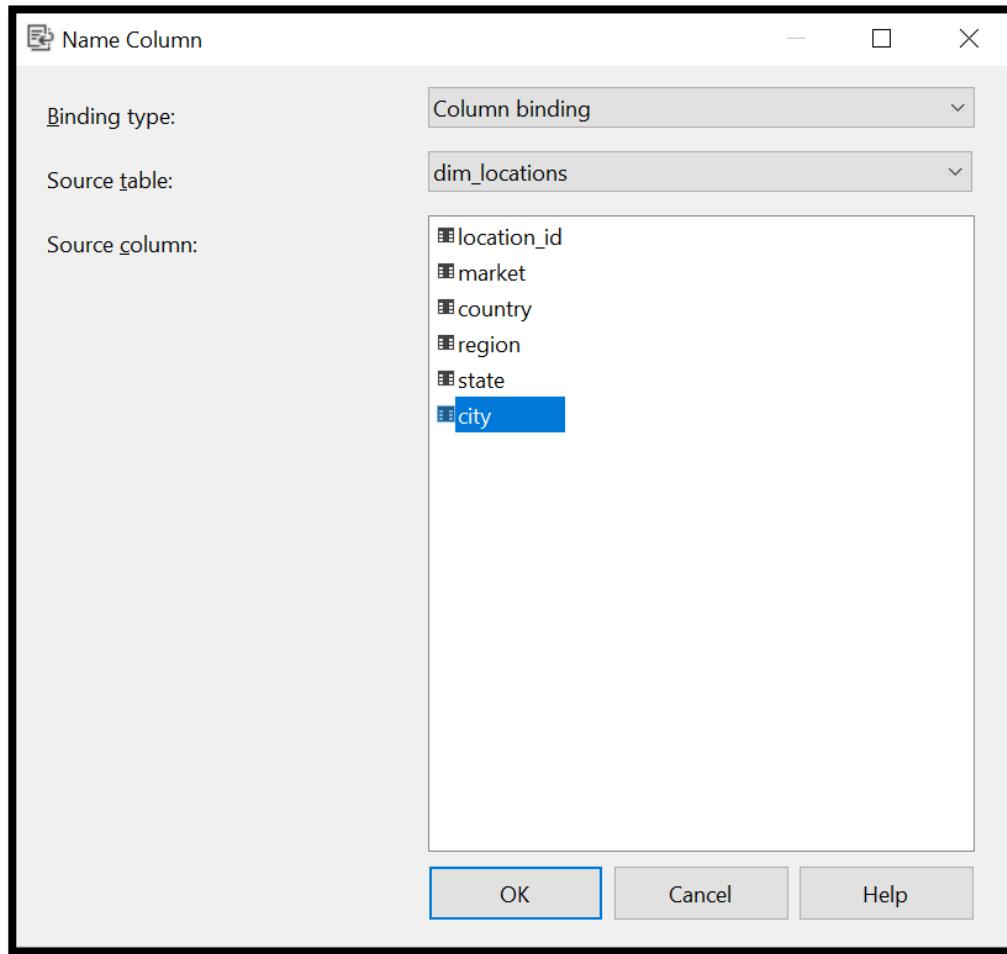
Bước 6.1.2: Trong bảng properties, tại OrderBy chọn Key, tại Type chọn City

OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	City
Name	City
Type	City
Basic	

Bước 6.1.3 Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính City. Vì thuộc City là thuộc tính cấp nhỏ nhất sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho City

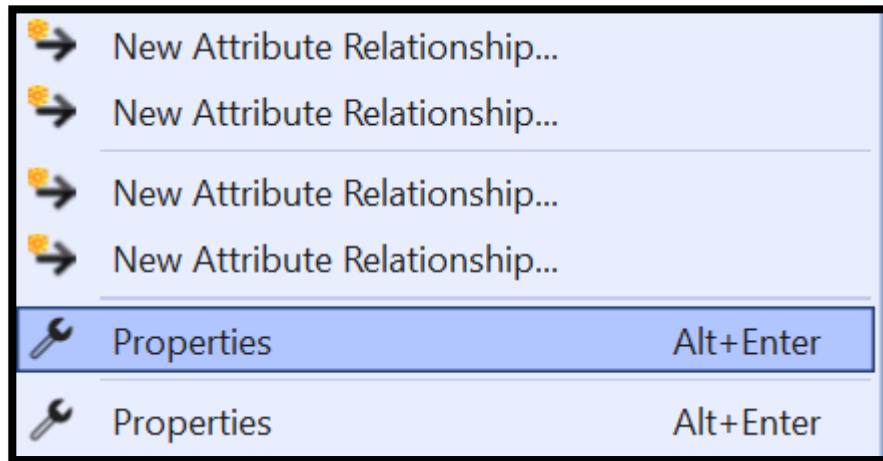


Bước 6.1.4: Tại cửa sổ Properties của City, ta chọn Name Column và chọn tên thuộc tính là City.



Bước 6.2: Cấu hình cho State

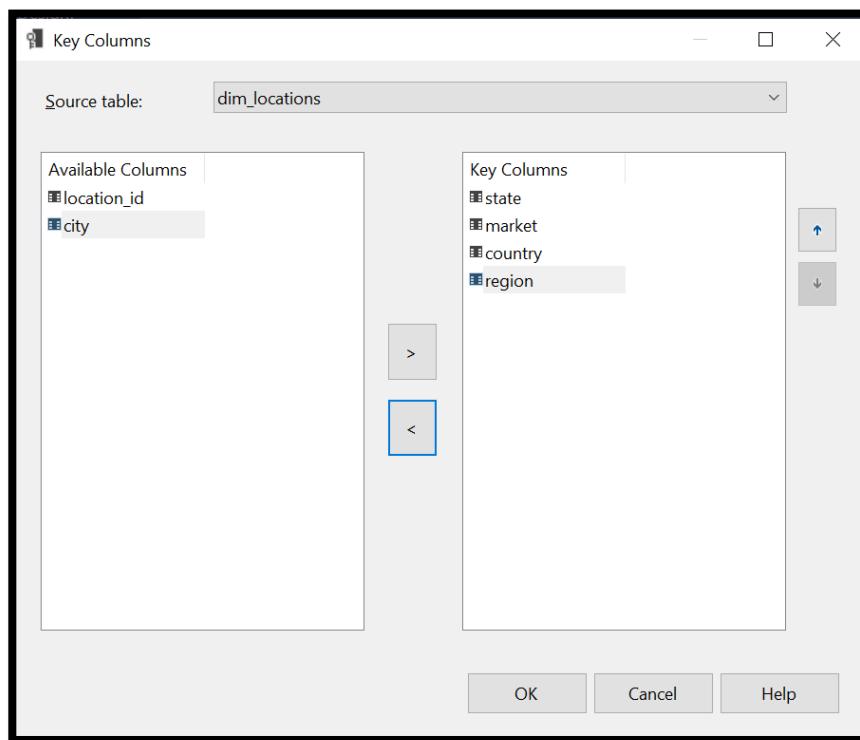
Bước 6.2.1: R-click vào State để mở bảng properties



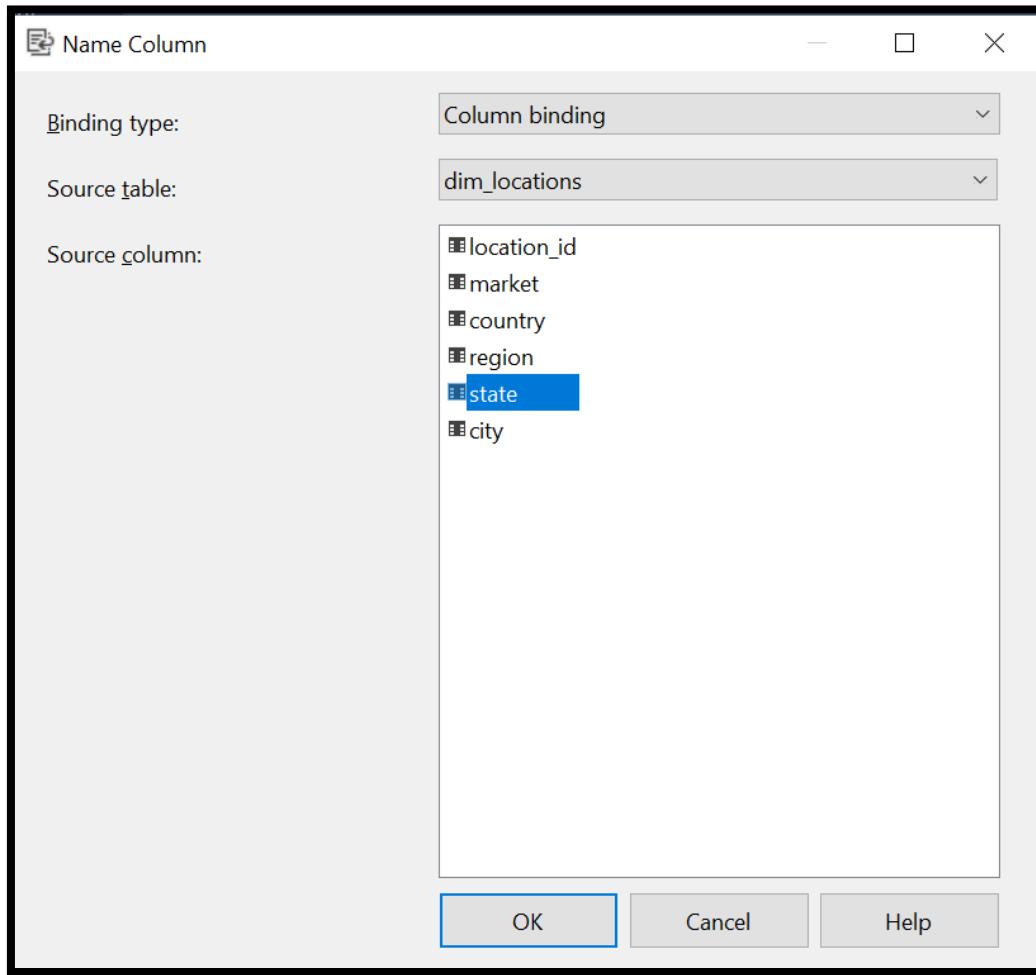
Bước 6.2.2: Trong bảng properties, tại OrderBy chọn Key, tại Type chọn StateOrProvince

OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	State
Name	State
Type	StateOrProvince
Usage	Regular

Bước 6.2.3: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính State. Vì thuộc State là thuộc tính cấp nhỏ sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho State



Bước 6.2.4: Tại cửa sổ Properties của City, ta chọn Name Column và chọn tên thuộc tính là State.



Bước 6.3: Cấu hình cho State

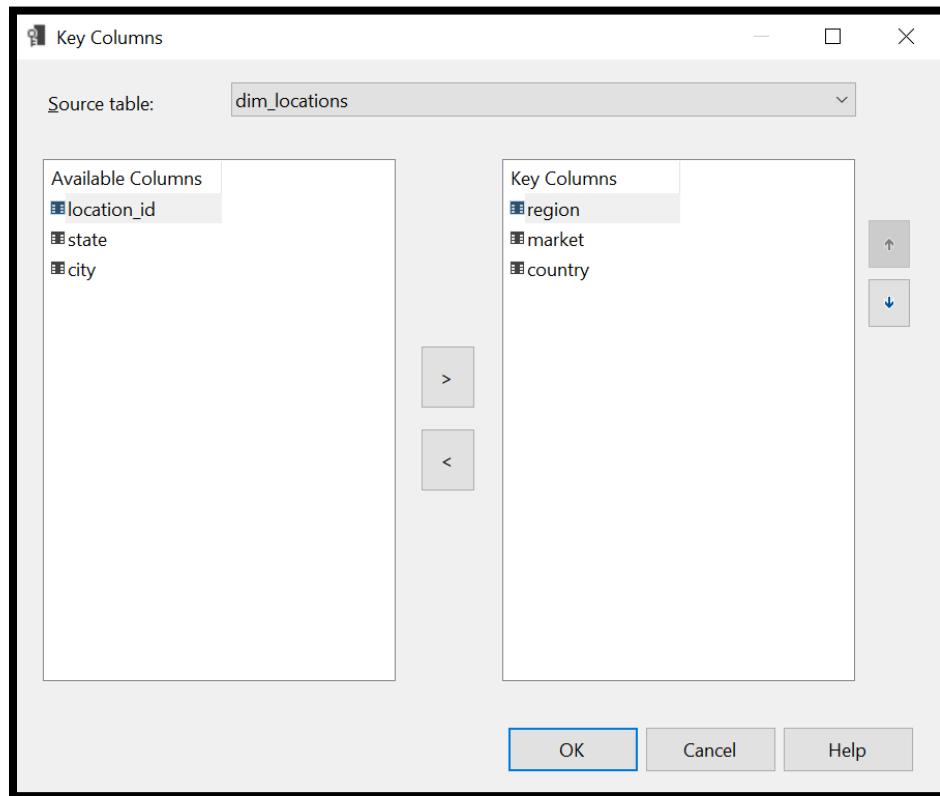
Bước 6.3.1: R-click vào **Region** để mở bảng properties



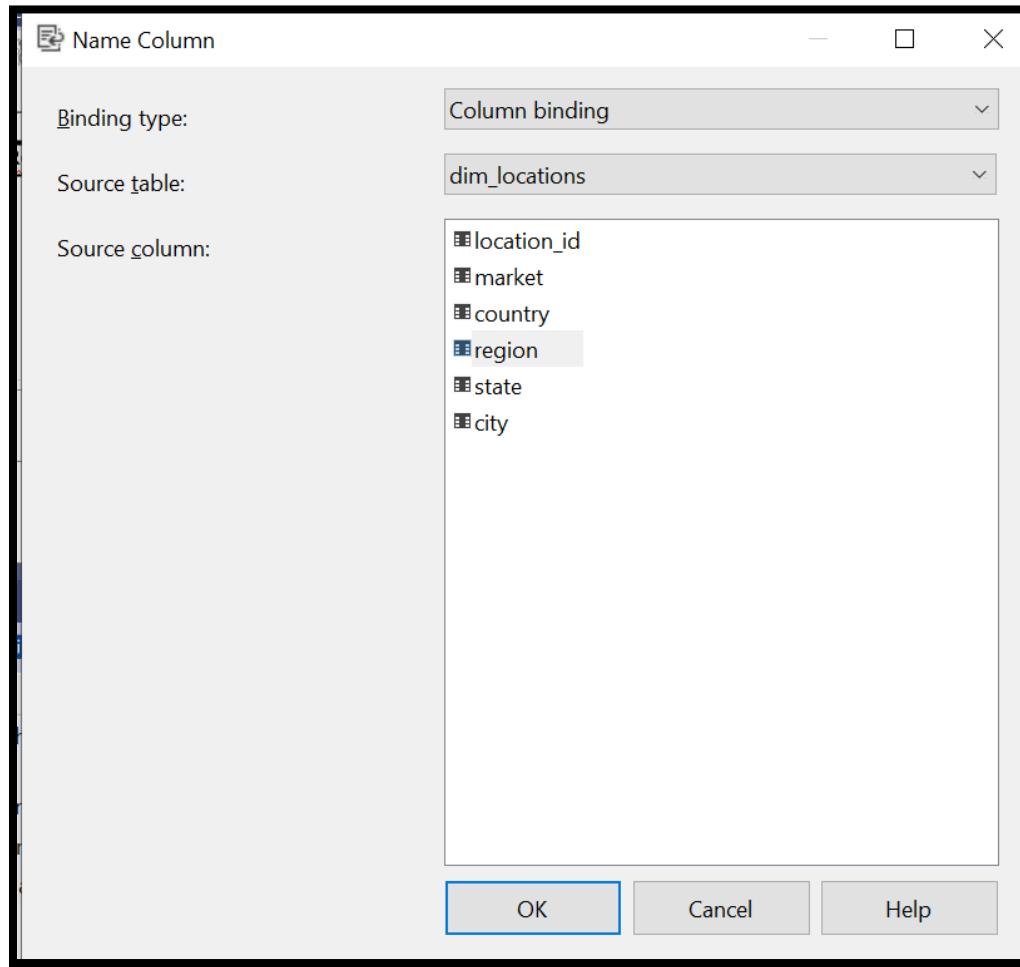
Bước 6.3.2: Trong bảng properties, tại OrderBy chọn Key, tại Type chọn Region

OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	Region
Name	Region
Type	Region
Usage	Regular
Misc	

Bước 6.2.3: Chỉnh khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Region. Vì thuộc Region là thuộc tính cấp nhỏ sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho Region.



Bước 6.3.4: Tại cửa sổ Properties của Region, ta chọn Name Column và chọn tên thuộc tính là Region.



Bước 6.4: Cấu hình cho Country

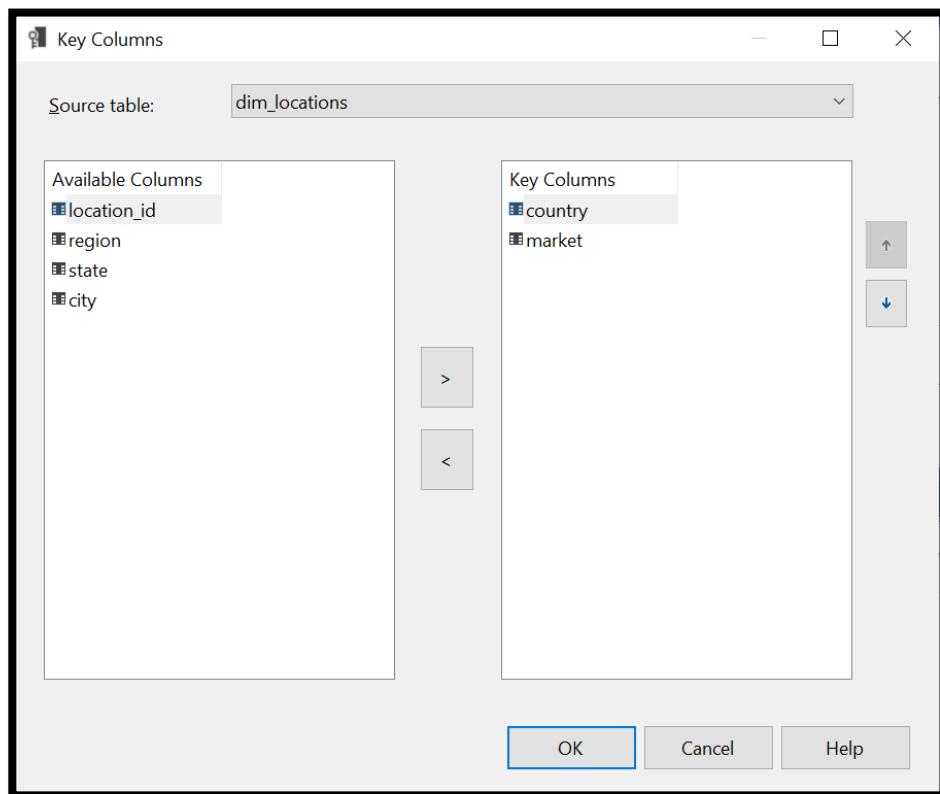
Bước 6.4.1: R-click vào Country để mở bảng properties



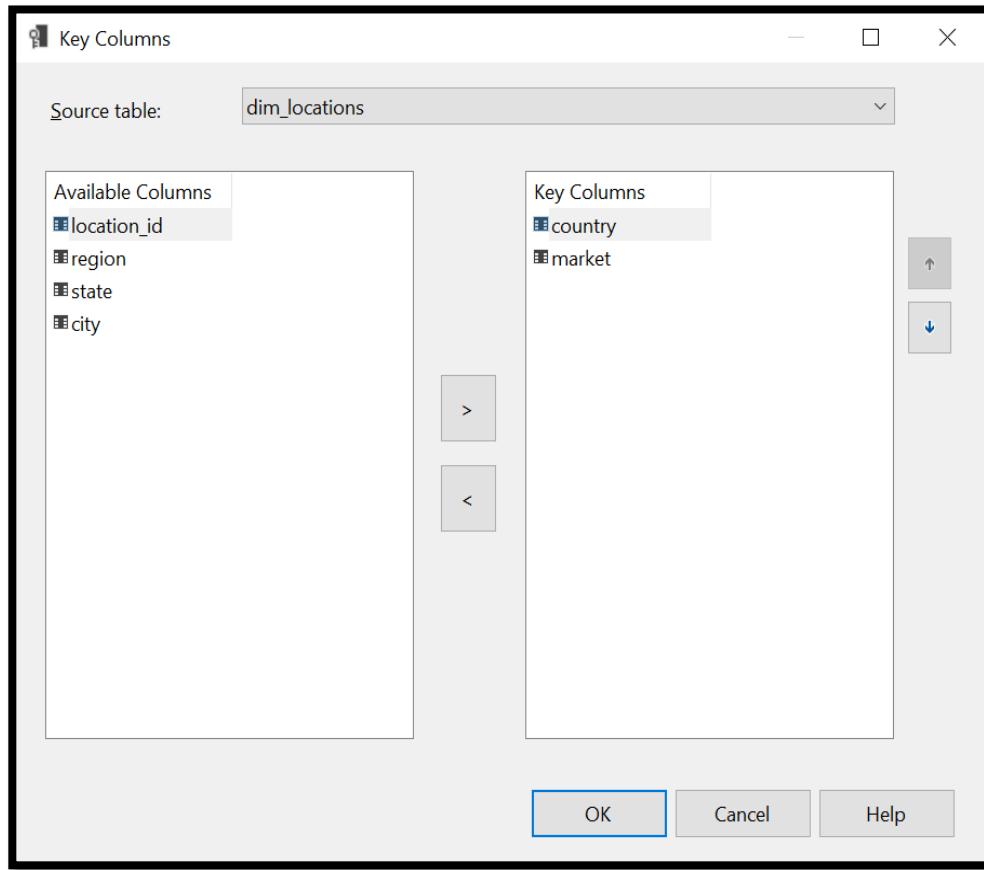
Bước 6.4.2: Trong bảng properties, tại OrderBy chọn Key, tại Type chọn Country

OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	Country
Name	Country
Type	Country
Usage	Regular

Bước 6.4.3 Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Country. Vì thuộc Country là thuộc tính cấp nhỏ sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho Country



Bước 6.4.4: Tại cửa sổ Properties của Country, ta chọn Name Column và chọn tên thuộc tính là Country.



Bước 6.5: Cấu hình cho Market

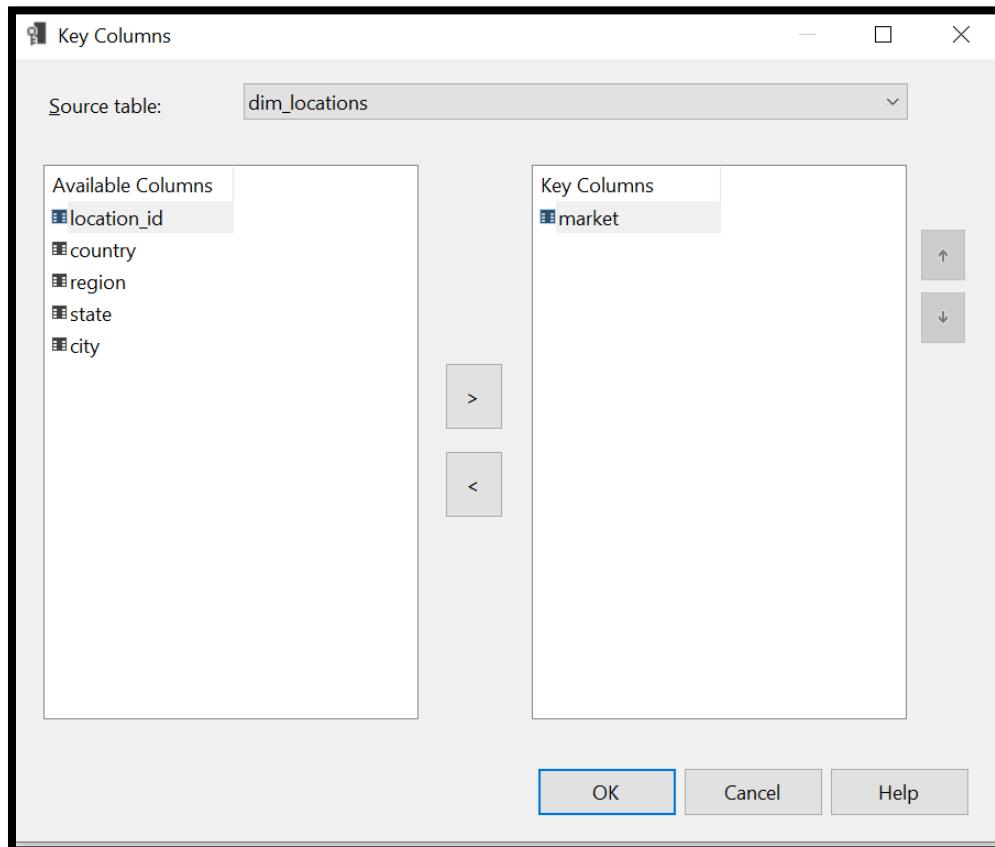
Bước 6.5.1: R-click vào **Market** để mở bảng properties



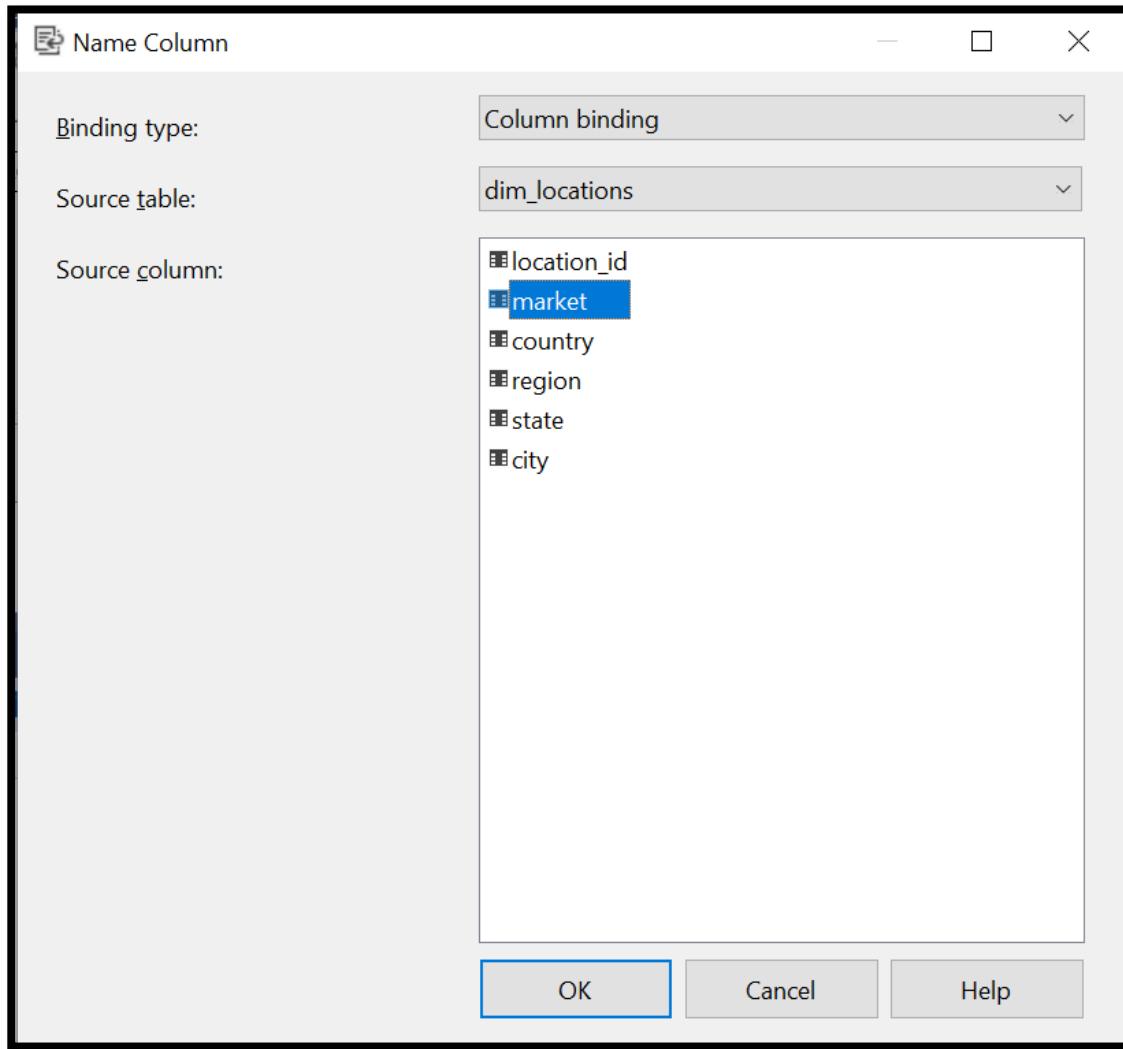
Bước 6.5.2: Trong bảng properties, tại OrderBy chọn **Key**, tại Type chọn **Continent**

DiscretizationMethod	None
EstimatedCount	0
HasLineage	False
IsAggregatable	True
OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	Market
Name	Market
Type	Continent
Usage	Regular

Bước 6.5.3: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Market. Vì thuộc Region là thuộc tính cấp lớn nhất sẽ lấy khóa cột là chính nó.

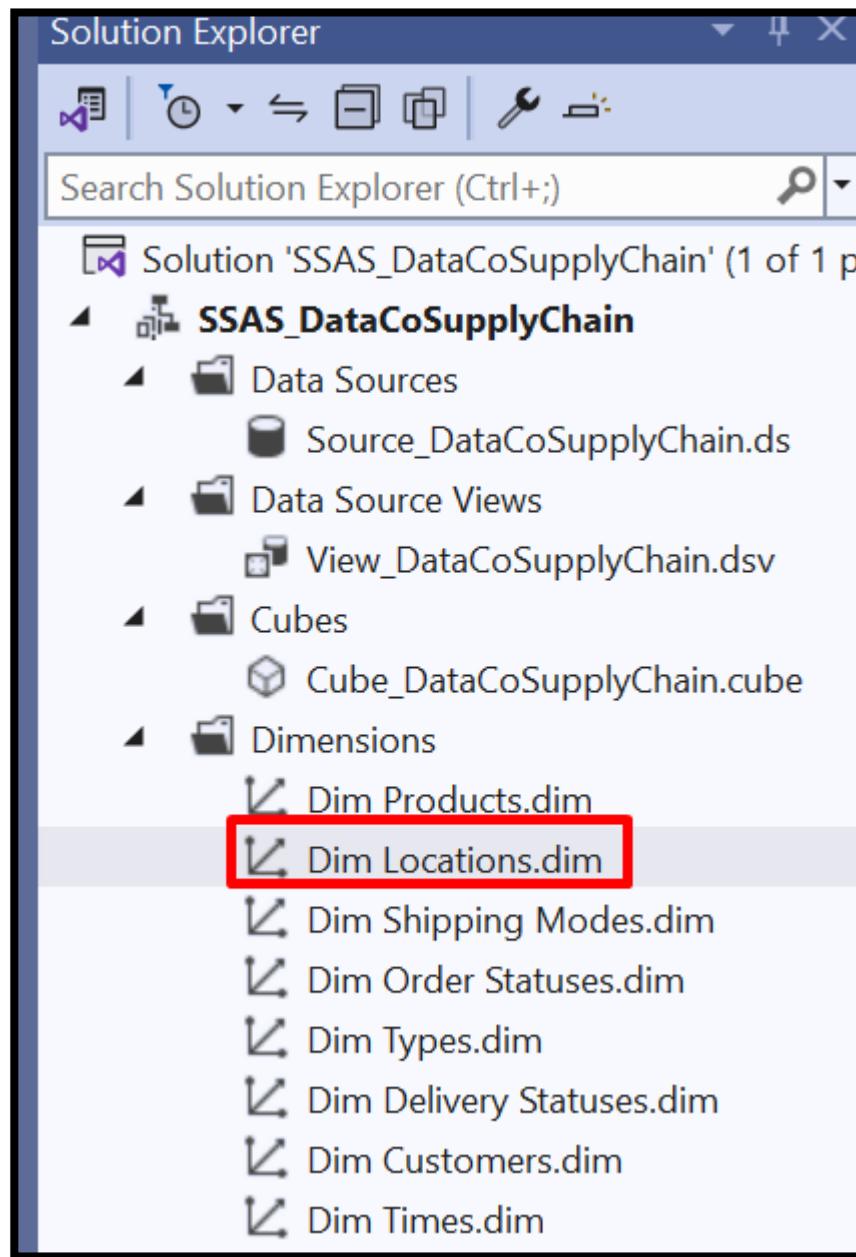


Bước 6.5.4: Tại cửa sổ Properties của Region, ta chọn Name Column và chọn tên thuộc tính là Market.

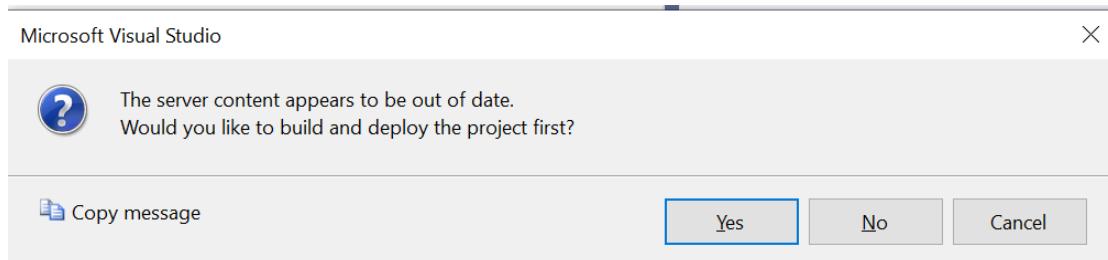


Bước 7: Ta Process Dim Locations.dim

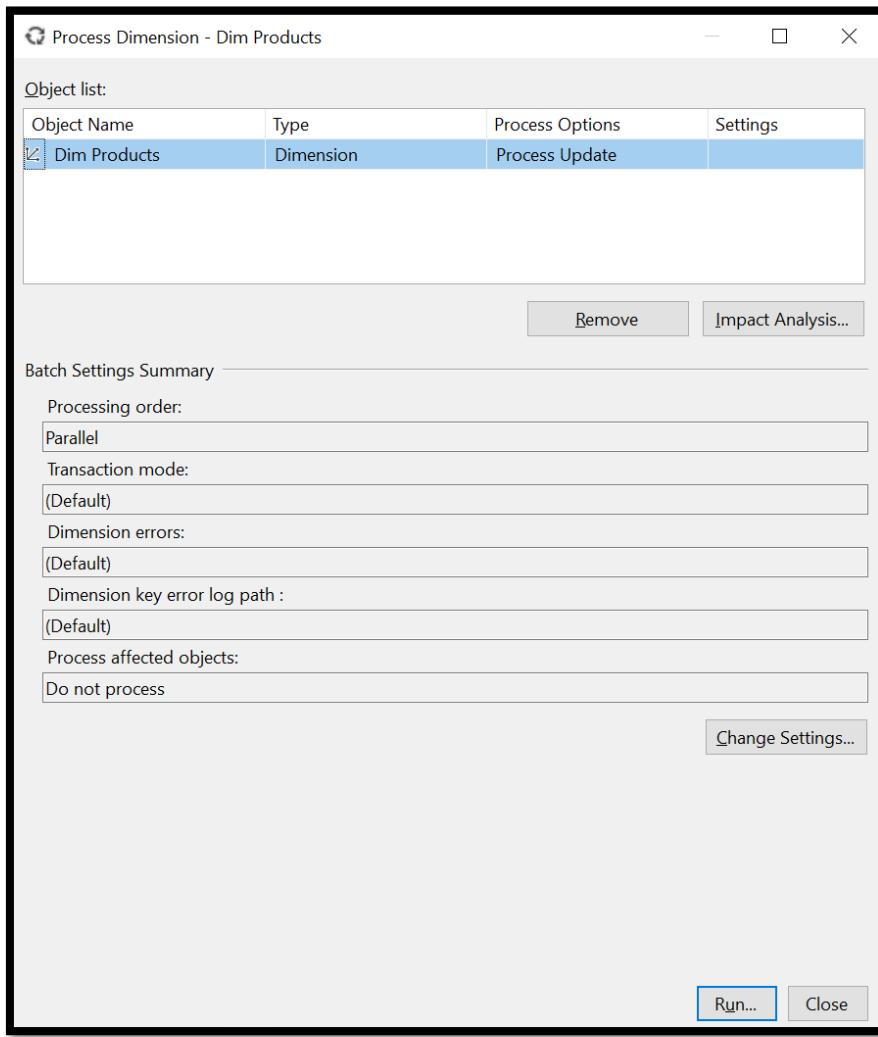
Tại Solution Explorer, r-click vào **Dim Locations.dim** và chọn **Process** để chạy quá trình tạo dim locations.



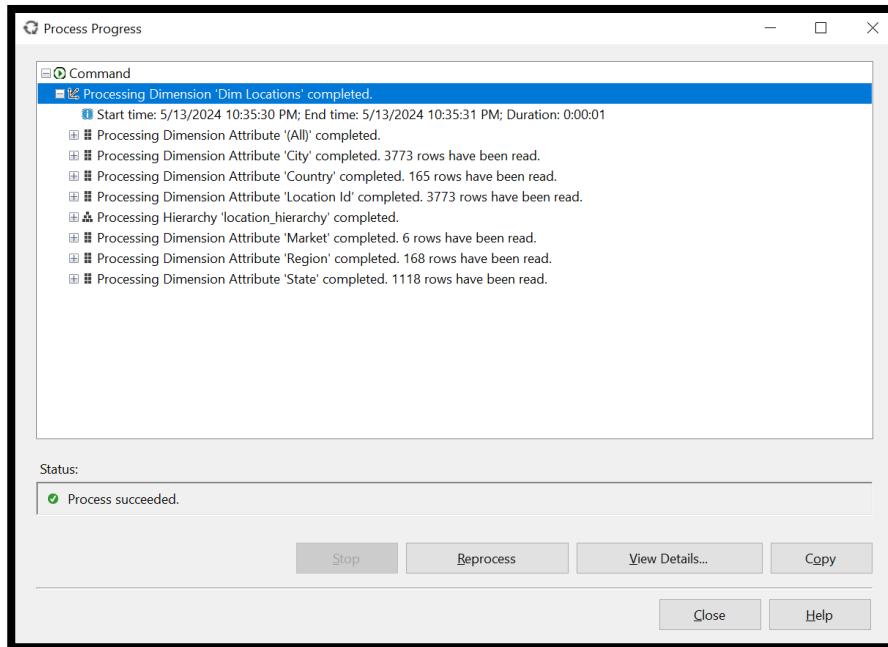
Hệ thống thông báo project bị outdated (Tức là ta đã thao tác, và project bị thay đổi, cần khởi động lại) chọn Yes.



Sau đó chọn **Run**.



Kết quả



Mở lại Dim Locations.dim, chọn tab Browser, kiểm tra Hierarchy

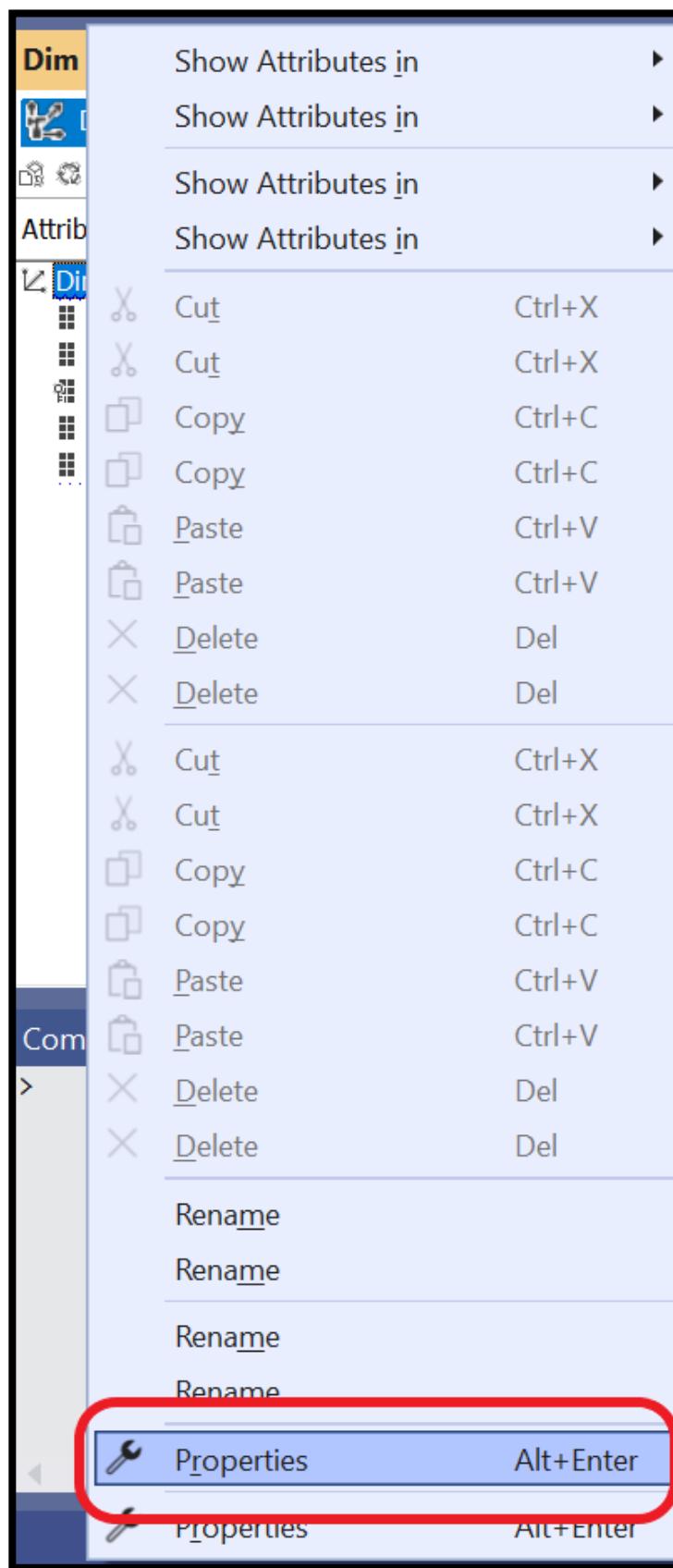
The screenshot shows the 'Dim Locations.dim [Design]' tab selected. In the top navigation bar, the 'Browser' tab is highlighted. The 'Hierarchy' dropdown is set to 'location_hierarchy'. The main pane displays a hierarchical tree under 'Current level: (All)'. The tree starts with 'Europe', which branches into 'Albania', 'Alemania', 'Austria', and 'Bélgica'. 'Bélgica' further branches into 'Western Europe', which includes 'Amberes', 'Brabante Flamenco', 'Bruselas', 'Flanders Occidental', 'Flanders Oriental', 'Henao', 'Lieja', 'Limburgo', and 'Namur'.

3.6.3. Cấu hình và tạo thuộc tính phân cấp cho “Dim Times.dim”

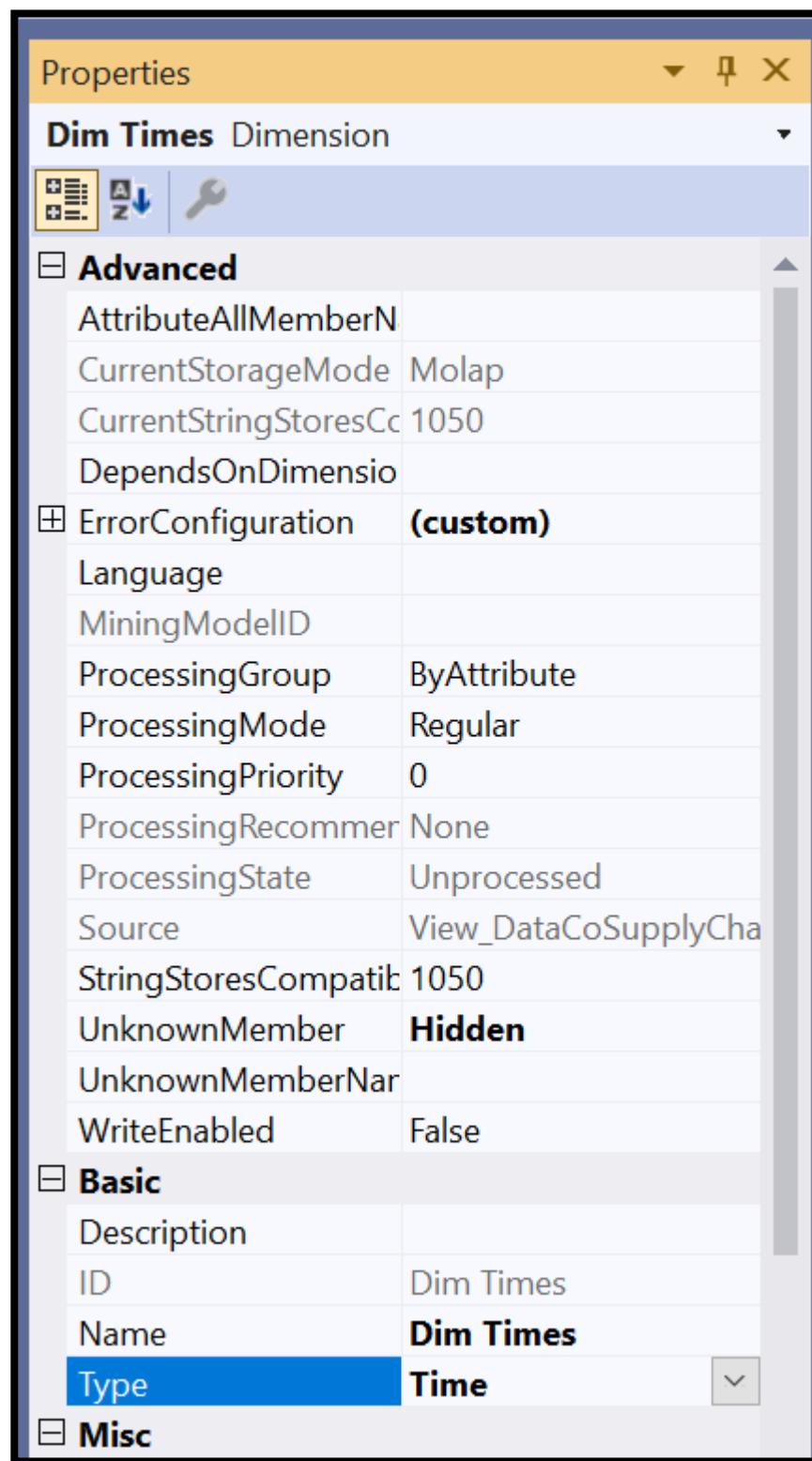
Bước 1: d-click vào Dim Times.dim, kéo thả lần lược theo thứ tự từ cao đến thấp (trên xuống dưới) day, month, year, hour, minute vào Hierarchy panel. Đổi tên Hierarchy thành time_hierarchy

The screenshot shows the 'Dim Times.dim [Design]' tab selected. In the top navigation bar, the 'Hierarchy' tab is highlighted. The 'Hierarchies' panel on the right shows a new hierarchy named 'time_hierarchy' being created. The hierarchy structure is shown as follows: Year, Month, Day, Hour, Minute. A tooltip indicates: 'To create a new hierarchy, drag an attribute here.' The 'Attributes' panel on the left lists attributes for 'Dim Times' such as Date Of Week, Day, Hour, Minute, Month, Time Id, and Year. The 'Data Source View' panel on the right shows the physical table 'dim_time' with columns time_id, day, month, year, hour, minute, and date_of_week.

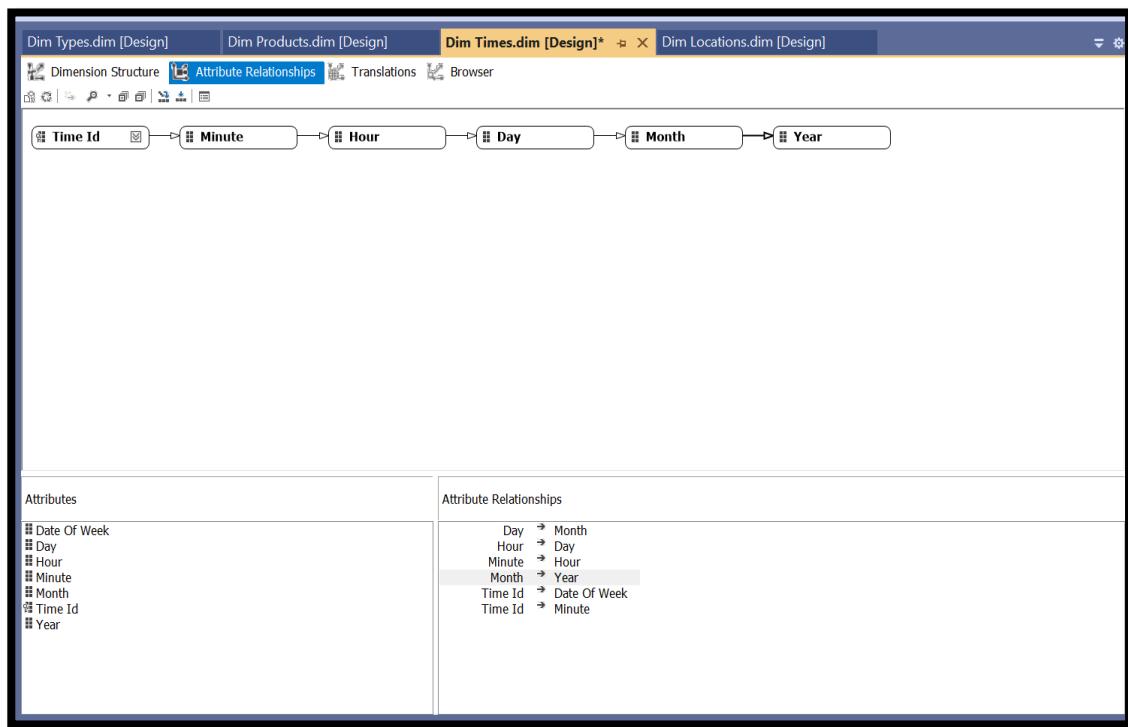
Bước 3: R-click vào Dim Time bên trong Attribute panel, chọn Properties.



Bước 4: tại UnknownMember chọn Hidden, tại Type, chọn Time.



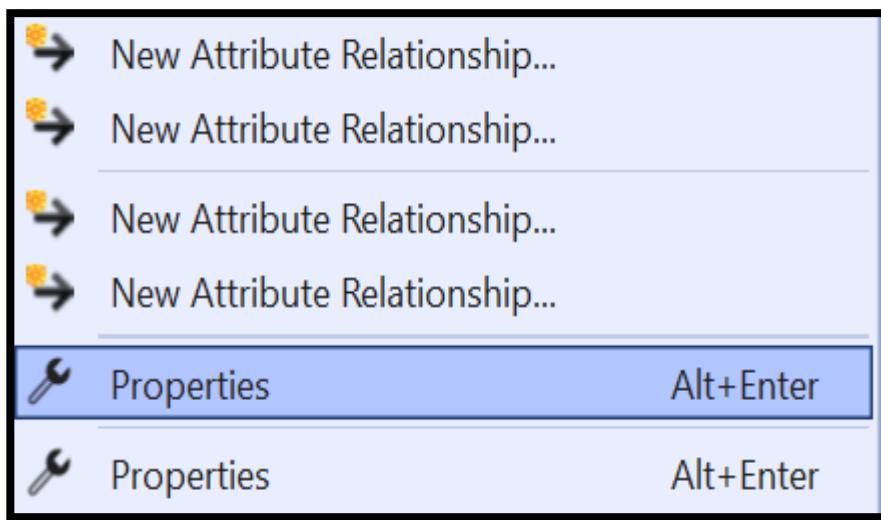
Bước 5: Click **Attribute Relationships** để mở bảng thuộc tính phân cấp, click giữ và kéo thả thuộc tính nhỏ hơn vào thuộc tính lớn hơn để phân cấp các thuộc tính từ trái sang phải (nhỏ nhất bên trái).



Bước 6: Cấu hình các thuộc tính con.

Bước 6.1: Cấu hình cho Minute

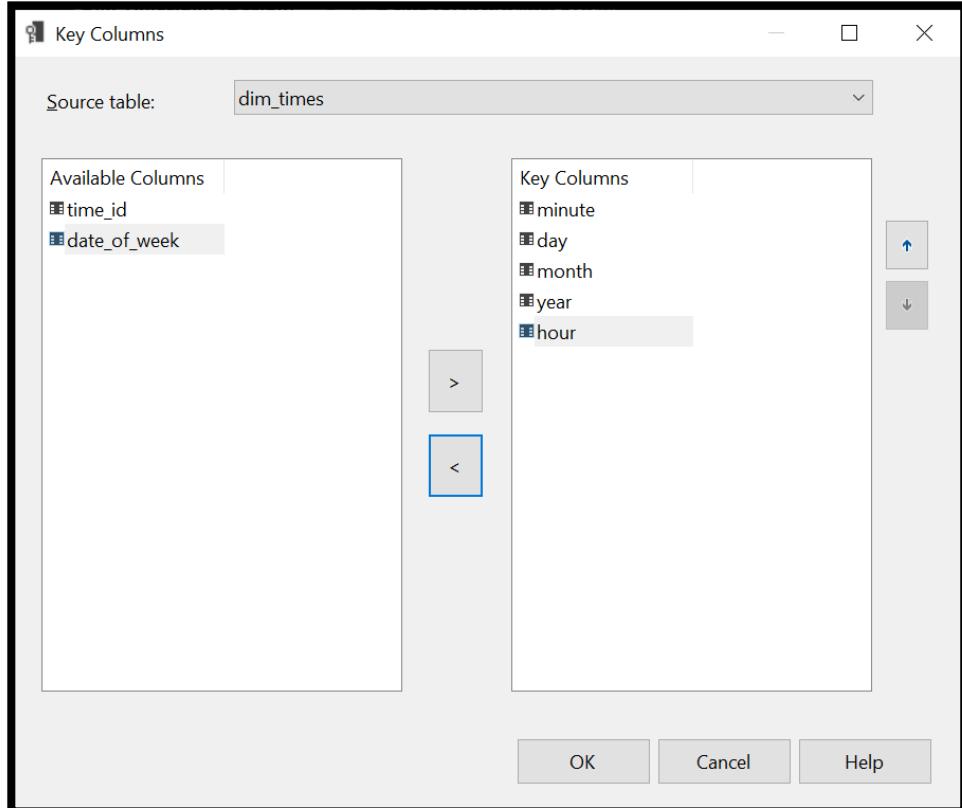
Bước 6.1.1 R-click vào **Minute** để mở bảng properties



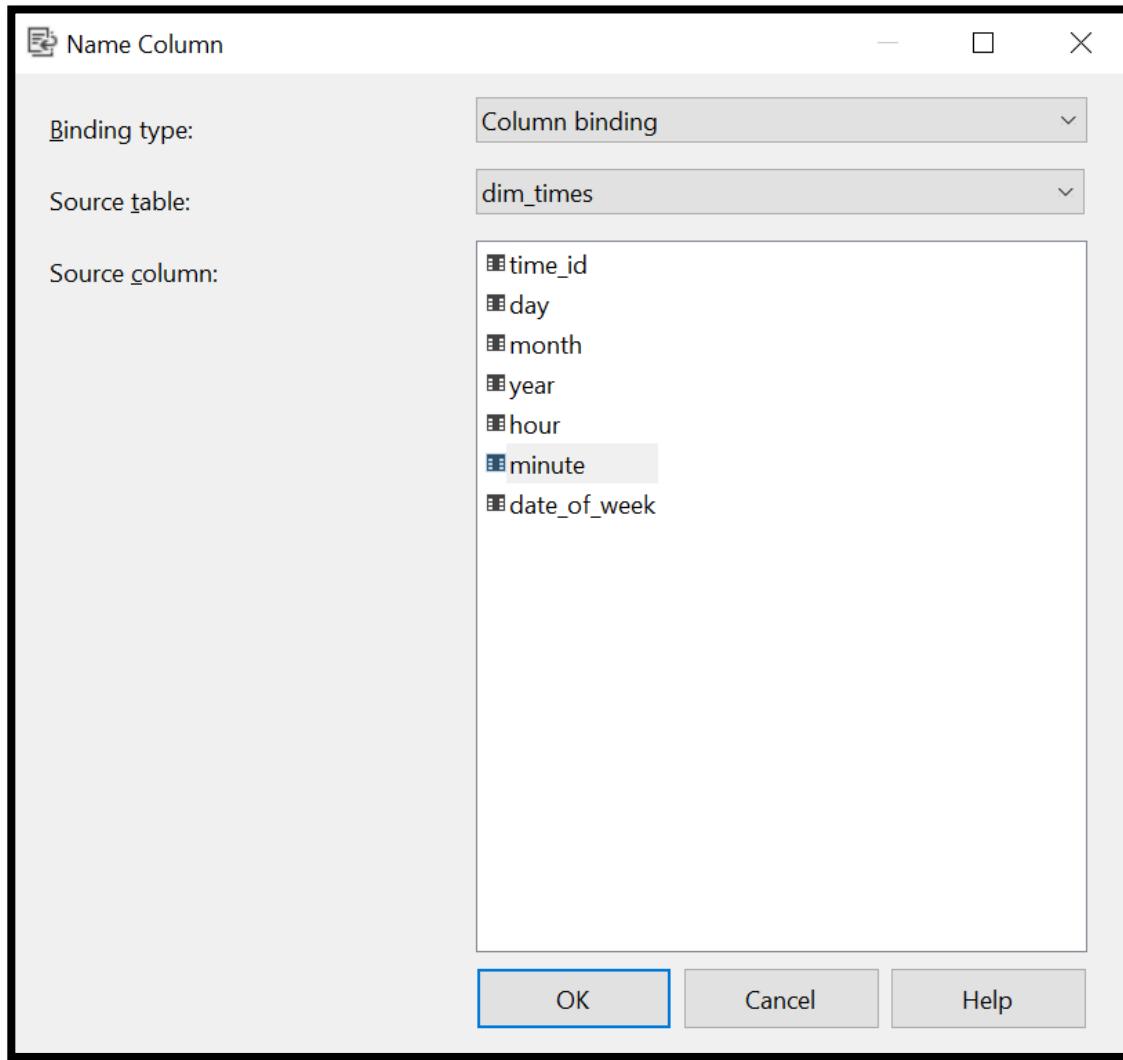
Bước 6.1.2: Trong bảng properties, tại OrderBy chọn Key, tại Type chọn Minute

OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	Minute
Name	Minute
Type	Minutes
Usage	Regular

Bước 6.1.3 Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Minute. Vì thuộc Minute là thuộc tính cấp nhỏ nhất sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho Minute



Bước 6.1.4: Tại cửa sổ Properties của minute, ta chọn Name Column và chọn tên thuộc tính là minute.



Bước 6.2: Cấu hình cho hour

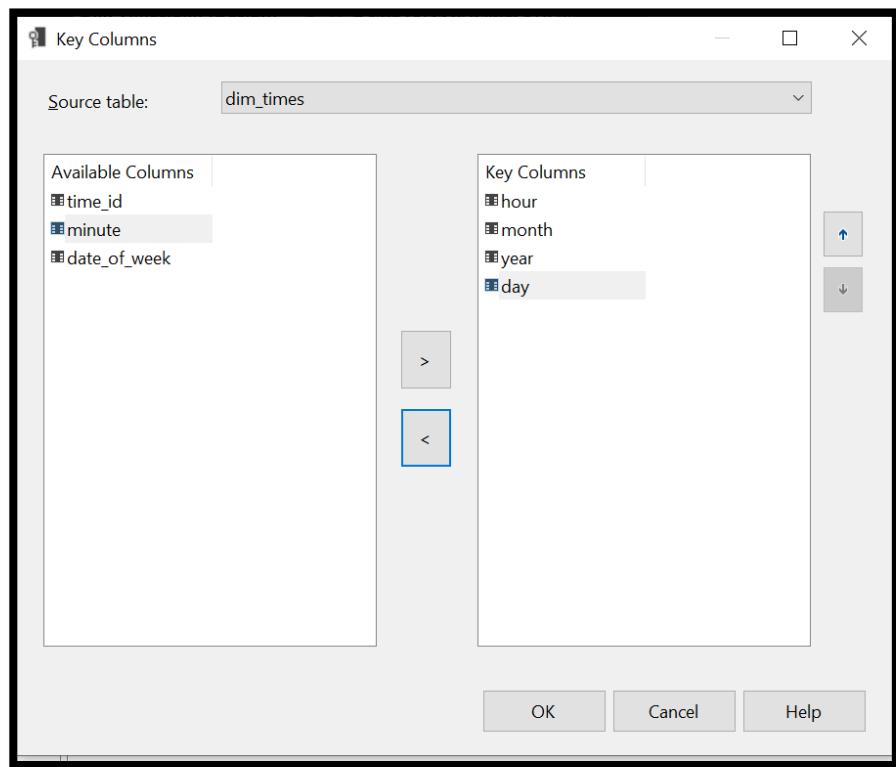
Bước 6.2.1: R-click vào **hour** để mở bảng properties



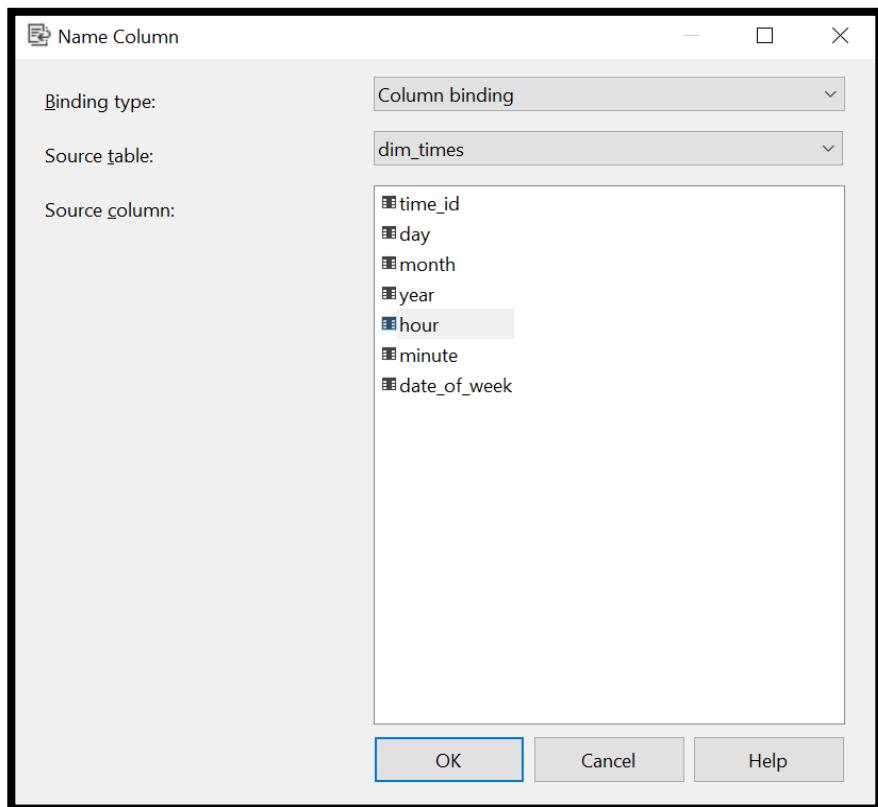
Bước 6.2.2: Trong bảng properties, tại OrderBy chọn Key, tại Type chọn hour

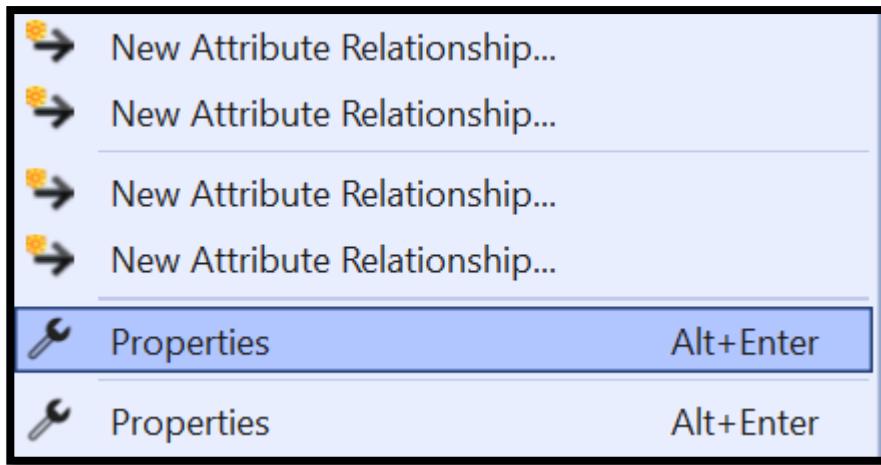
IsAggregatable	True
OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	Hour
Name	Hour
Type	Hours
Usage	Regular

Bước 6.2.3: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính hour. Vì thuộc hour là thuộc tính cấp nhỏ sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho hour



Bước 6.2.4: Tại cửa sổ Properties của hour, ta chọn Name Column và chọn tên thuộc tính là hour.

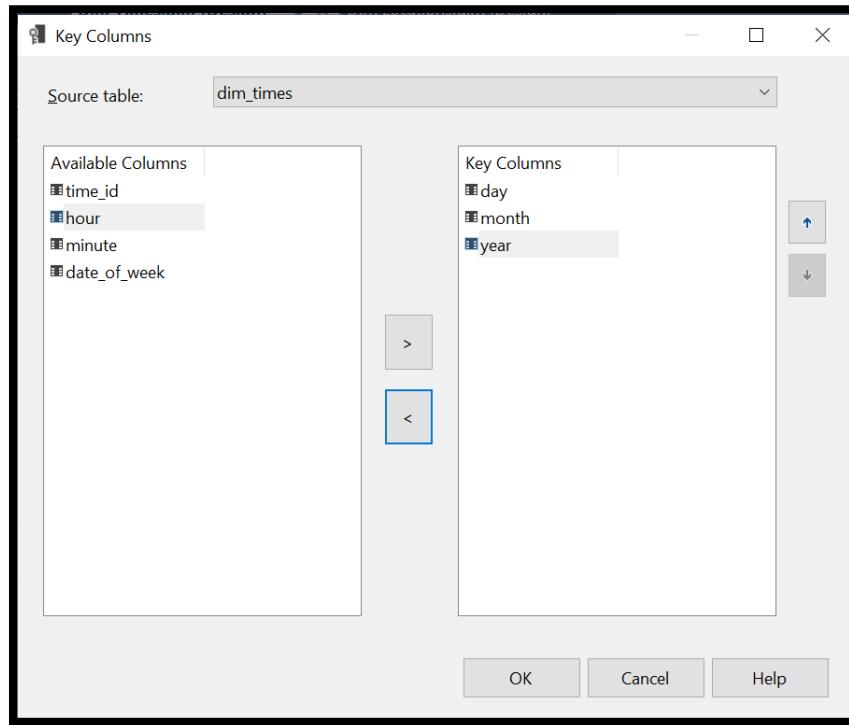


Bước 6.3: Cấu hình phần “day”**Bước 6.3.1:** R-click vào day để mở bảng properties**Bước 6.3.2:** Trong bảng properties, tại OrderBy chọn Key, tại Type chọn DayOfMonth.

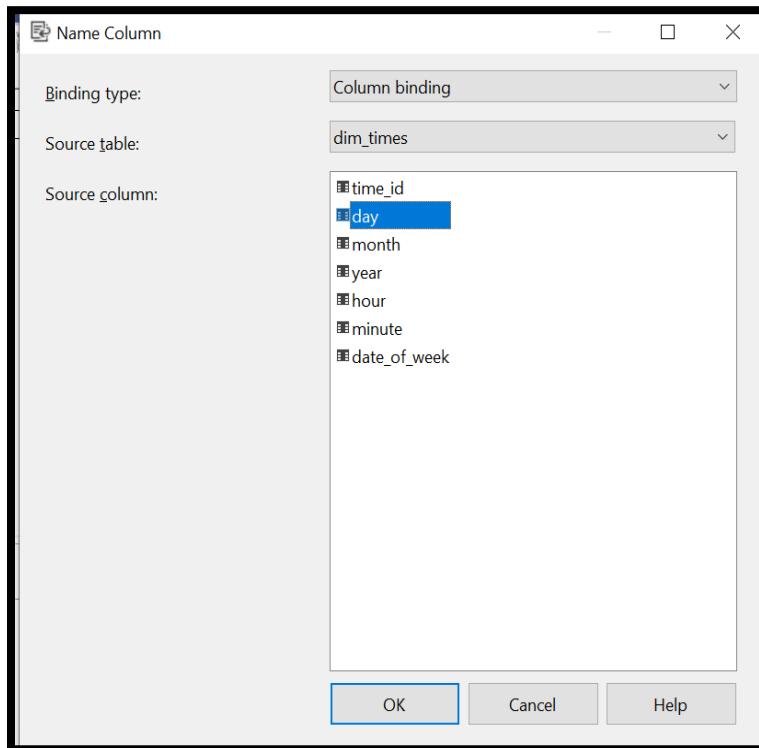
OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	Day
Name	Day
Type	DayOfMonth
Usage	Regular

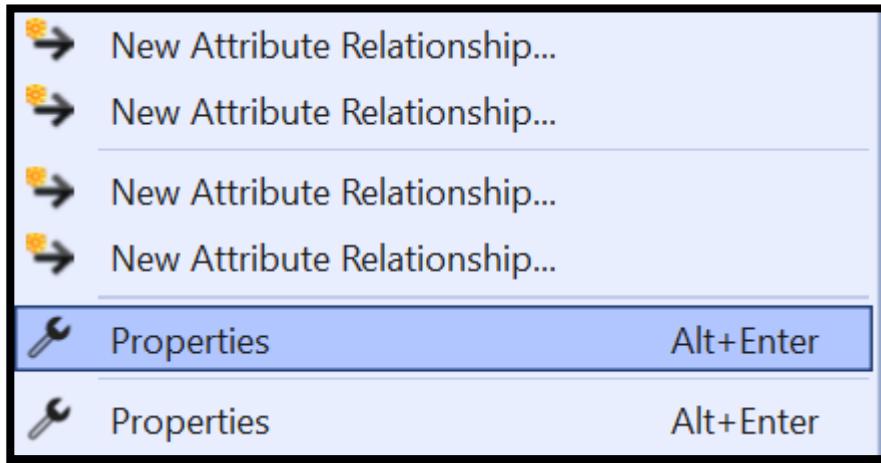
Bước 6.2.3: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính day. Vì day là thuộc tính cấp nhỏ sẽ lấy khóa cột gồm chính nó và những thuộc

tính cấp cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho day



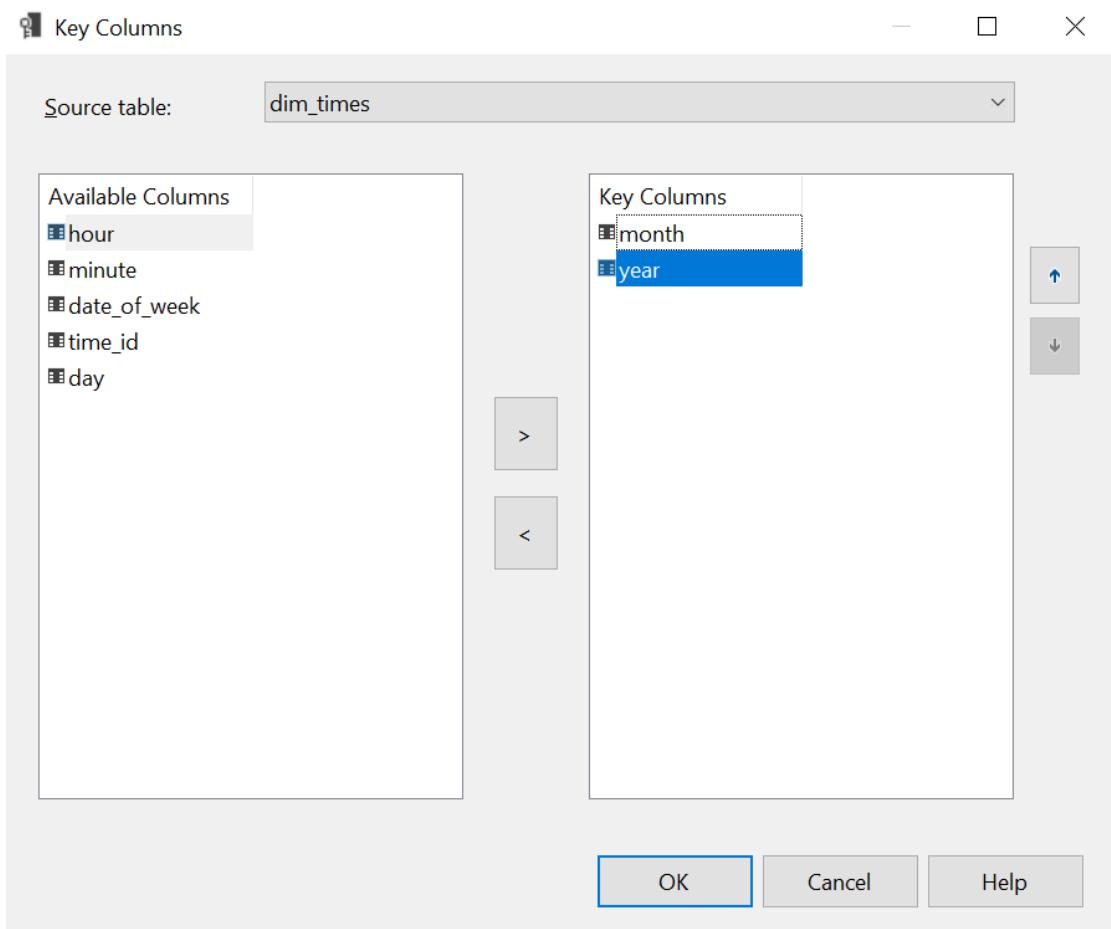
Bước 6.3.4: Tại cửa sổ Properties của day, ta chọn Name Column và chọn tên thuộc tính là day.



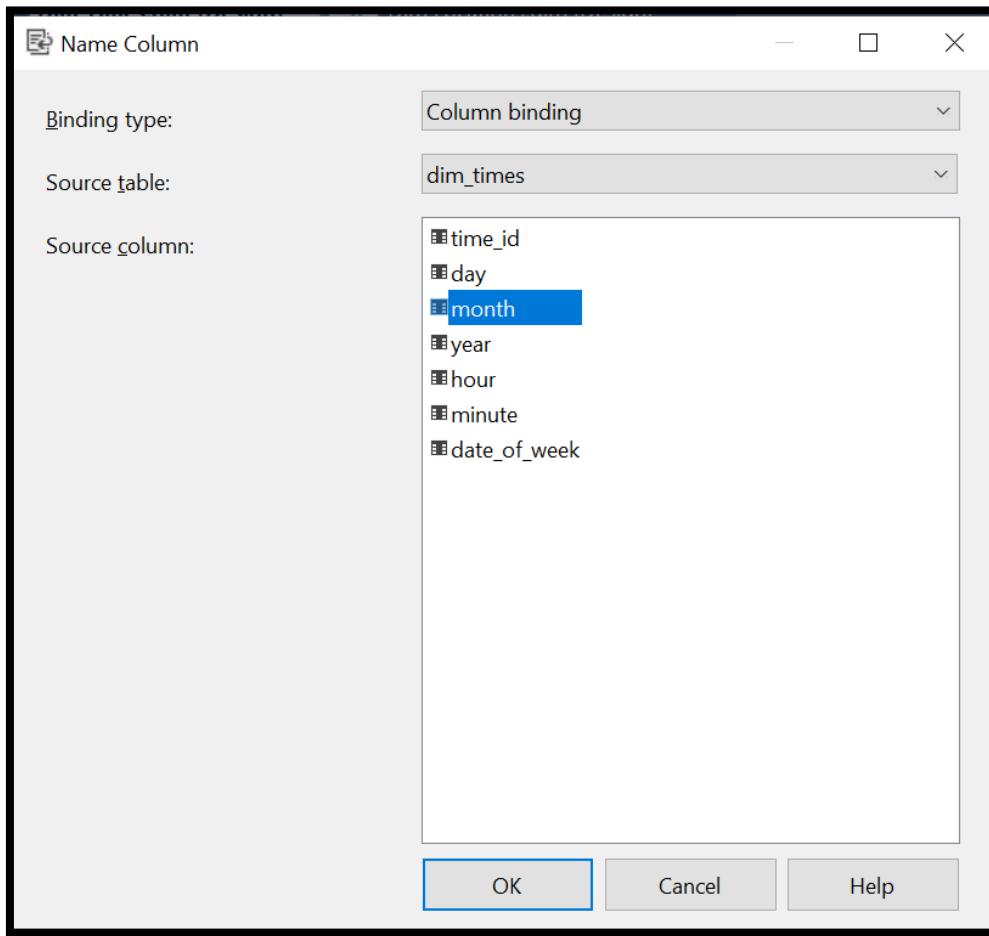
Bước 6.4: Cấu hình cho month**Bước 6.4.1:** R-click vào month để mở bảng properties**Bước 6.4.2:** Trong bảng properties, tại OrderBy chọn Key, tại Type chọn MonthOfYear.

HasLineage	False
IsAggregatable	True
OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
⊖ Basic	
Description	
FormatString	
ID	Month
Name	Month
Type	MonthOfYear

Bước 6.4.3: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính month. Vì month là thuộc tính cấp nhỏ sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn như hình bên dưới. Sau đó OK để kết thúc chọn cột cho month.



Bước 6.4.4: Tại cửa sổ Properties của Country, ta chọn Name Column và chọn tên thuộc tính là month.



Bước 6.5: Cấu hình cho year

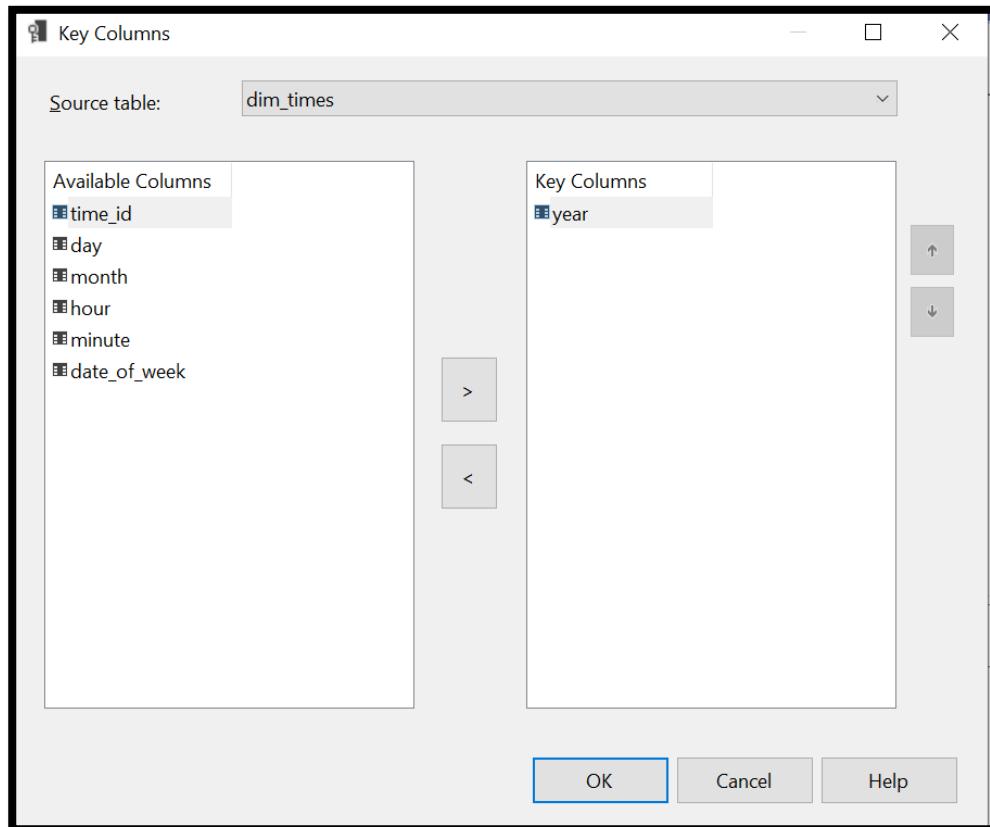
Bước 6.5.1: R-click vào year để mở bảng properties



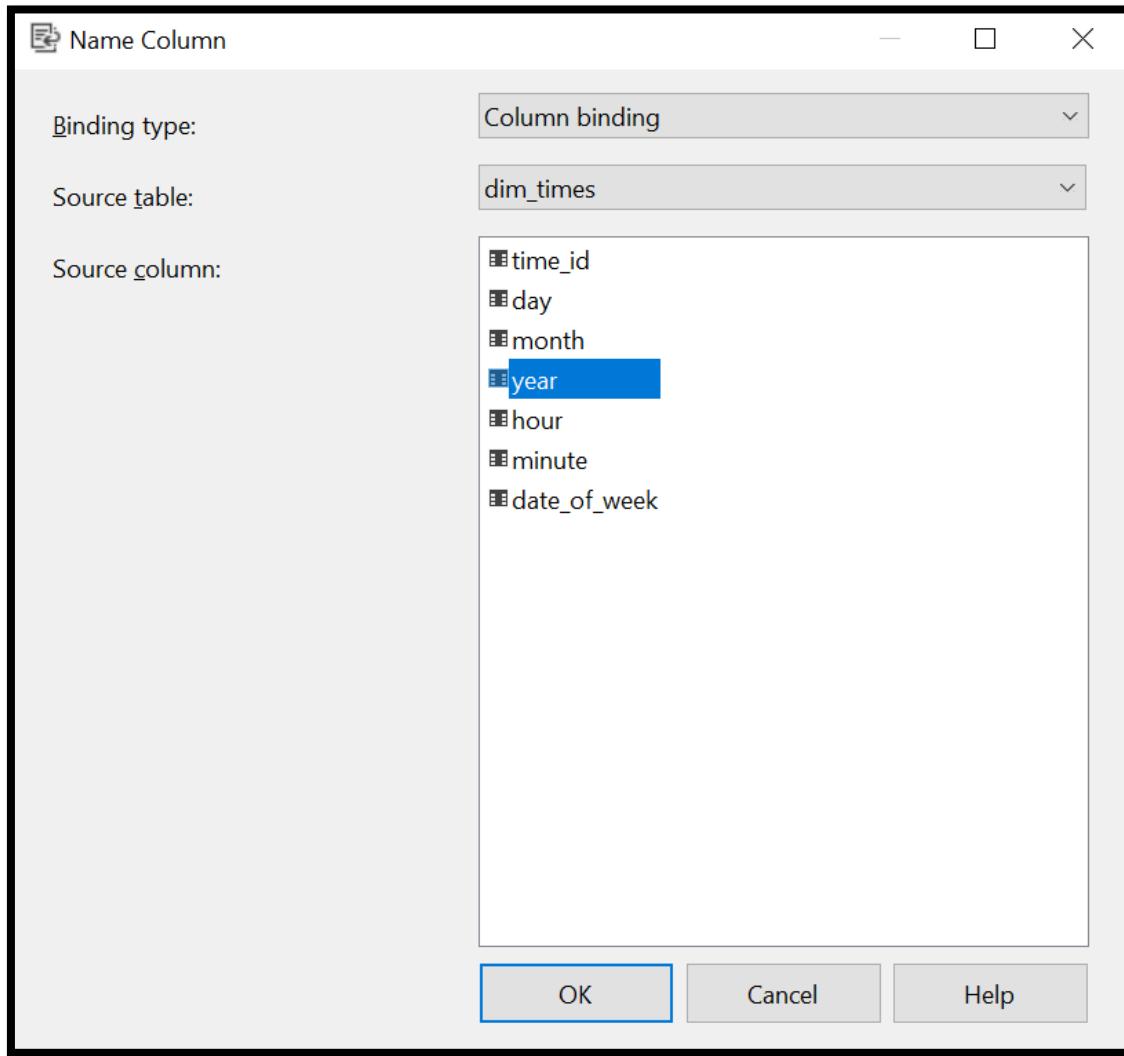
Bước 6.5.2 Trong bảng properties, tại OrderBy chọn Key, tại Type chọn year

OrderBy	Key
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompression	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	Year
Name	Year
Type	Years
Usage	Regular

Bước 6.5.3: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính year. Vì year là thuộc tính cấp lớn nhất sẽ lấy khóa cột là chính nó.

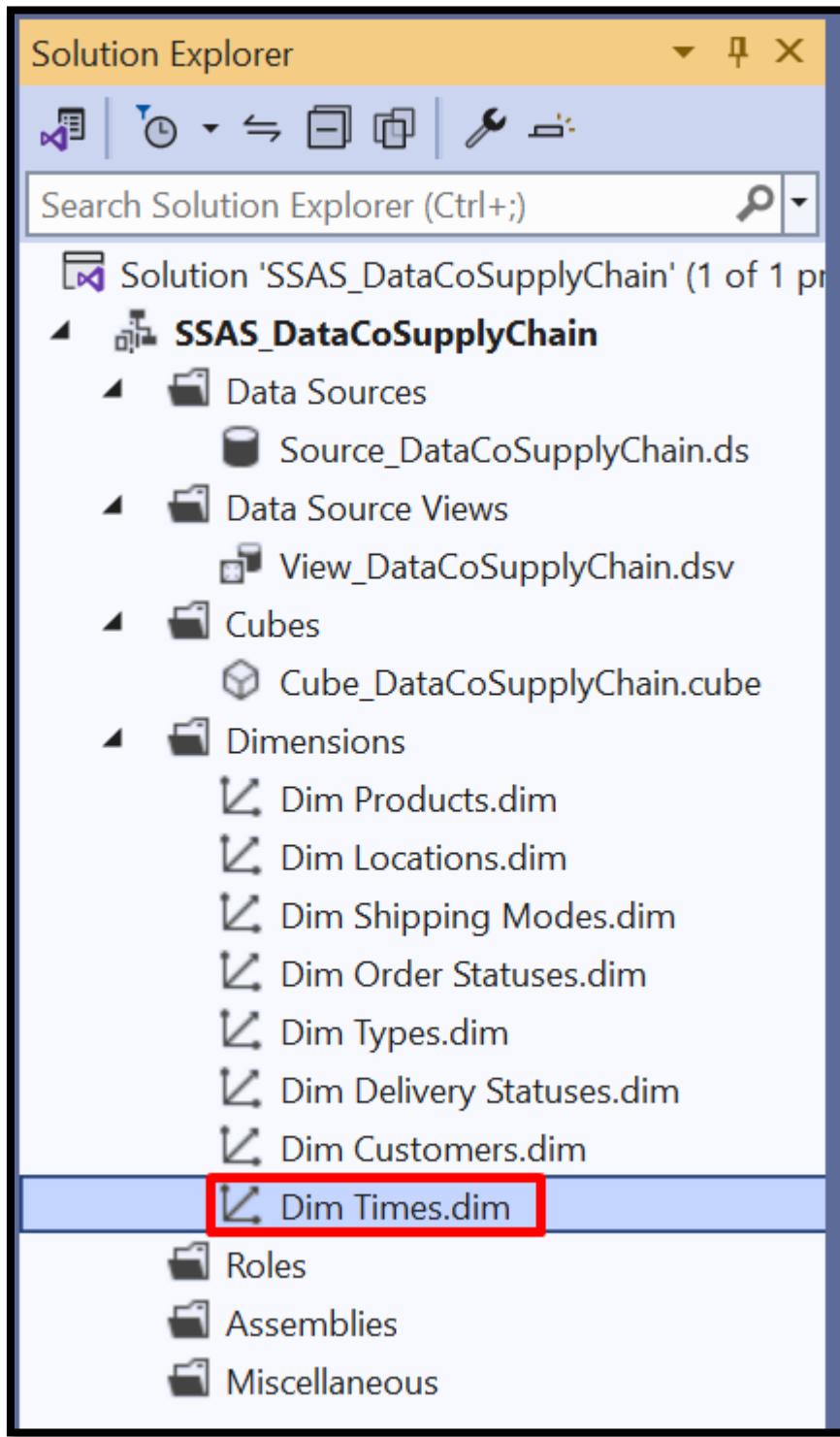


Bước 6.5.4: Tại cửa sổ Properties của Region, ta chọn Name Column và chọn tên thuộc tính là Market.

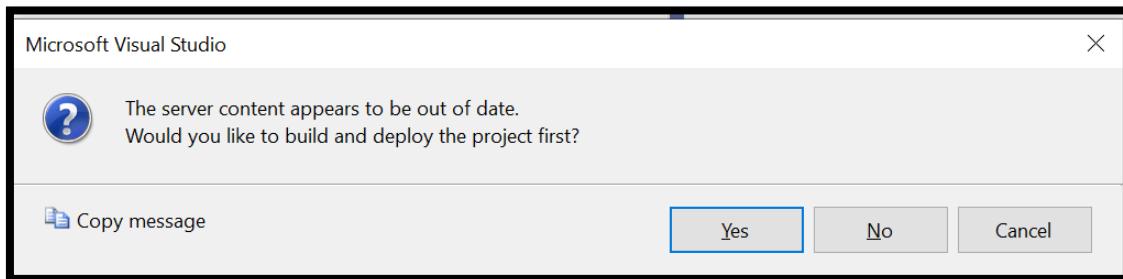


Bước 7: Tạo Process Dim Times.dim

Tại Solution Explorer, r-click vào **Dim Times.dim** và chọn **Process** để chạy quá trình tạo dim times.

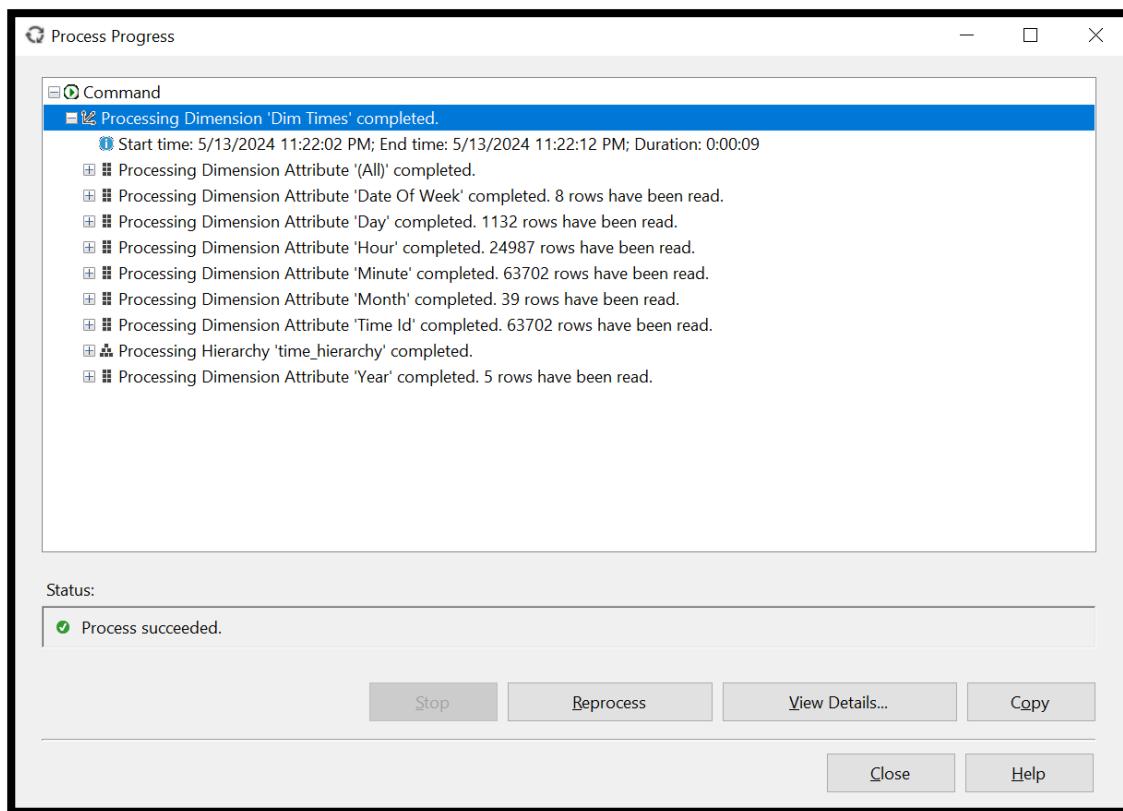


Hệ thống thông báo project bị outdated (Tức là ta đã thao tác, và project bị thay đổi, cần khởi động lại) chọn Yes.

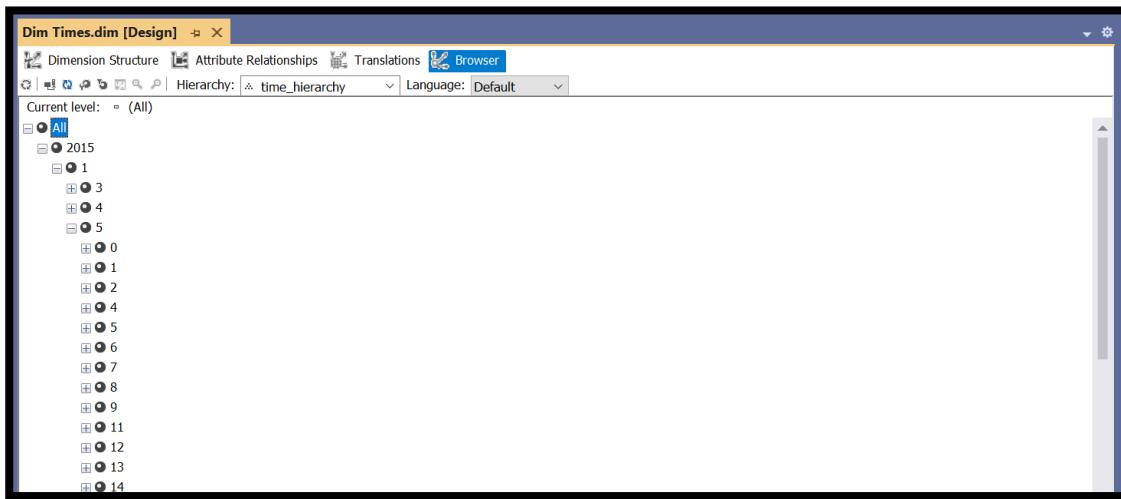


Sau đó chọn **Run**

Kết quả

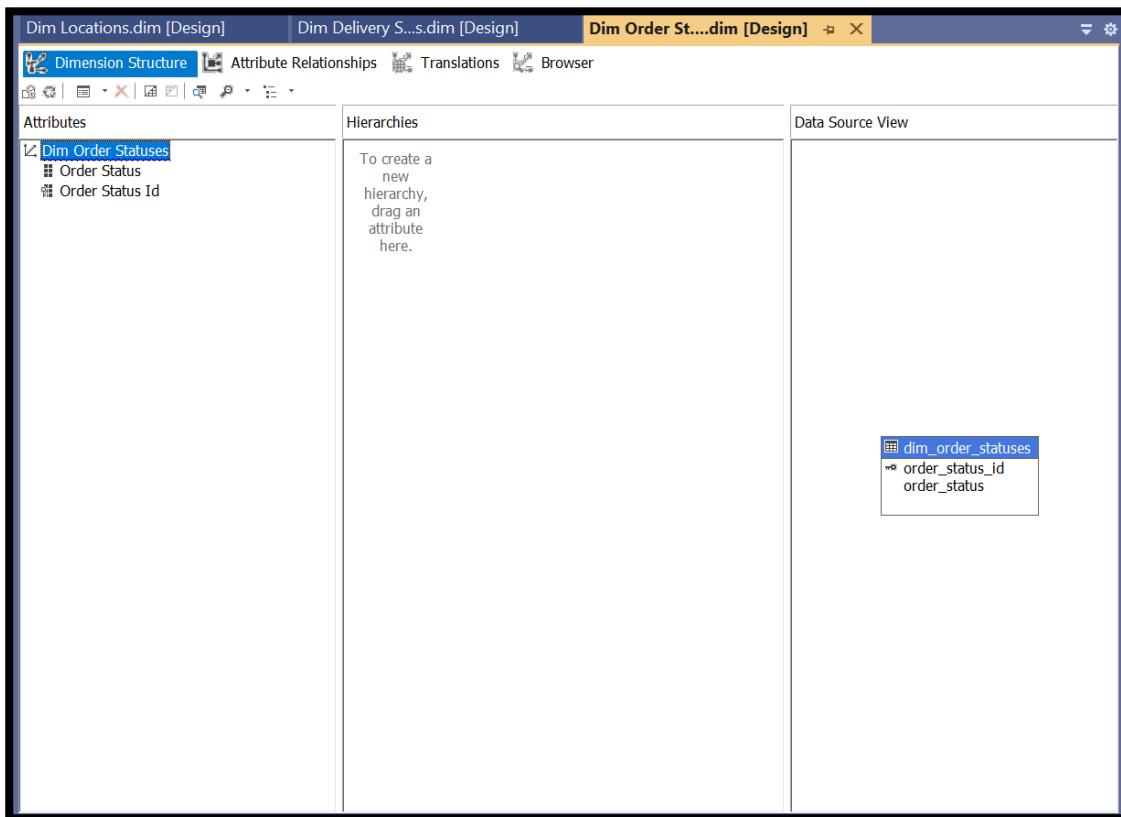


Mở lại Dim Times.dim, chọn tab Browser, kiểm tra Hierarchy

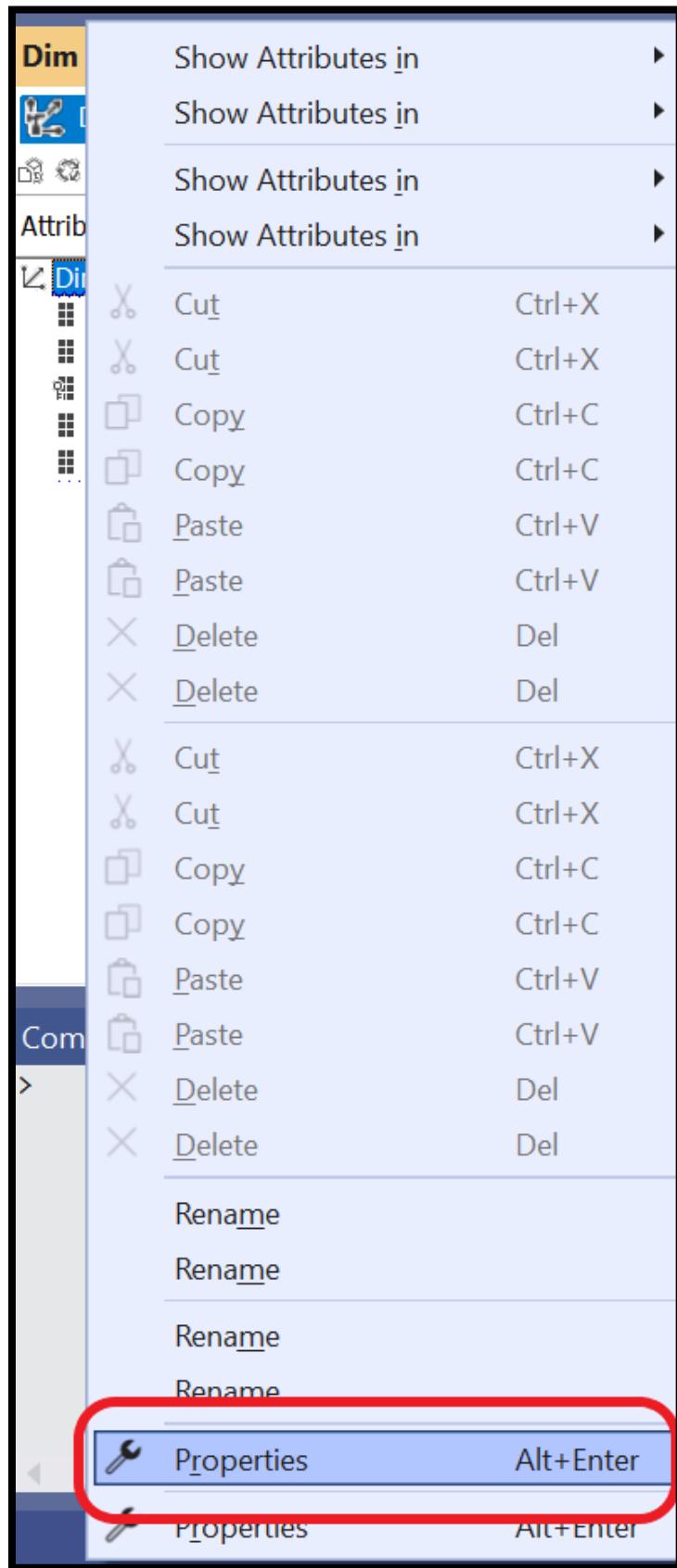


3.6.4. Cấu hình cho “Dim Statuses.dim”

Bước 1: d-click vào Dim Order Statuses.dim



Bước 2: R-click vào Dim Order Statuses bên trong Attribute panel, chọn Properties.



Bước 4: tại UnknownMember chọn Hidden.

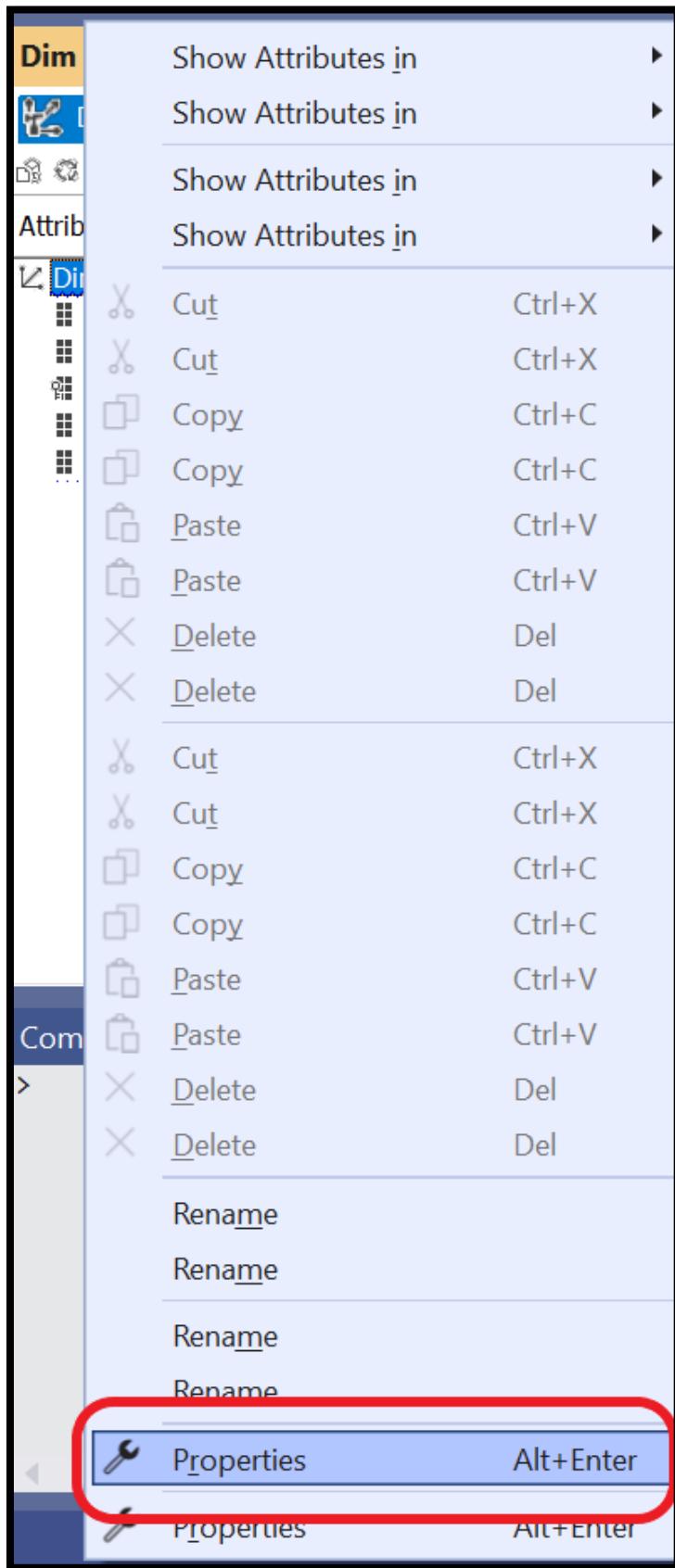
ProcessingState	Unprocessed
Source	View_DataCoSupplyChain
StringStoresCompatibility	1050
UnknownMember	Hidden
UnknownMemberName	
WriteEnabled	False
Basic	

3.6.5. Cấu hình cho “Dim Customers.dim”

Bước 1: d-click vào Dim Order Customers.dim

The screenshot shows the 'Dimension Structure' interface. On the left, under 'Attributes', the 'Dim Customers' attribute is selected and highlighted with a red box. It has three children: 'Customer Fname', 'Customer Id', and 'Customer Lname'. In the center, the 'Hierarchies' section contains a placeholder message: 'To create a new hierarchy, drag an attribute here.' On the right, the 'Data Source View' pane shows a table named 'dim_customers' with columns 'customer_id', 'customer_fname', and 'customer_lname'.

Bước 3: R-click vào Dim Order Statuses bên trong Attribute panel, chọn Properties.



Bước 4: tại UnknownMember chọn Hidden.

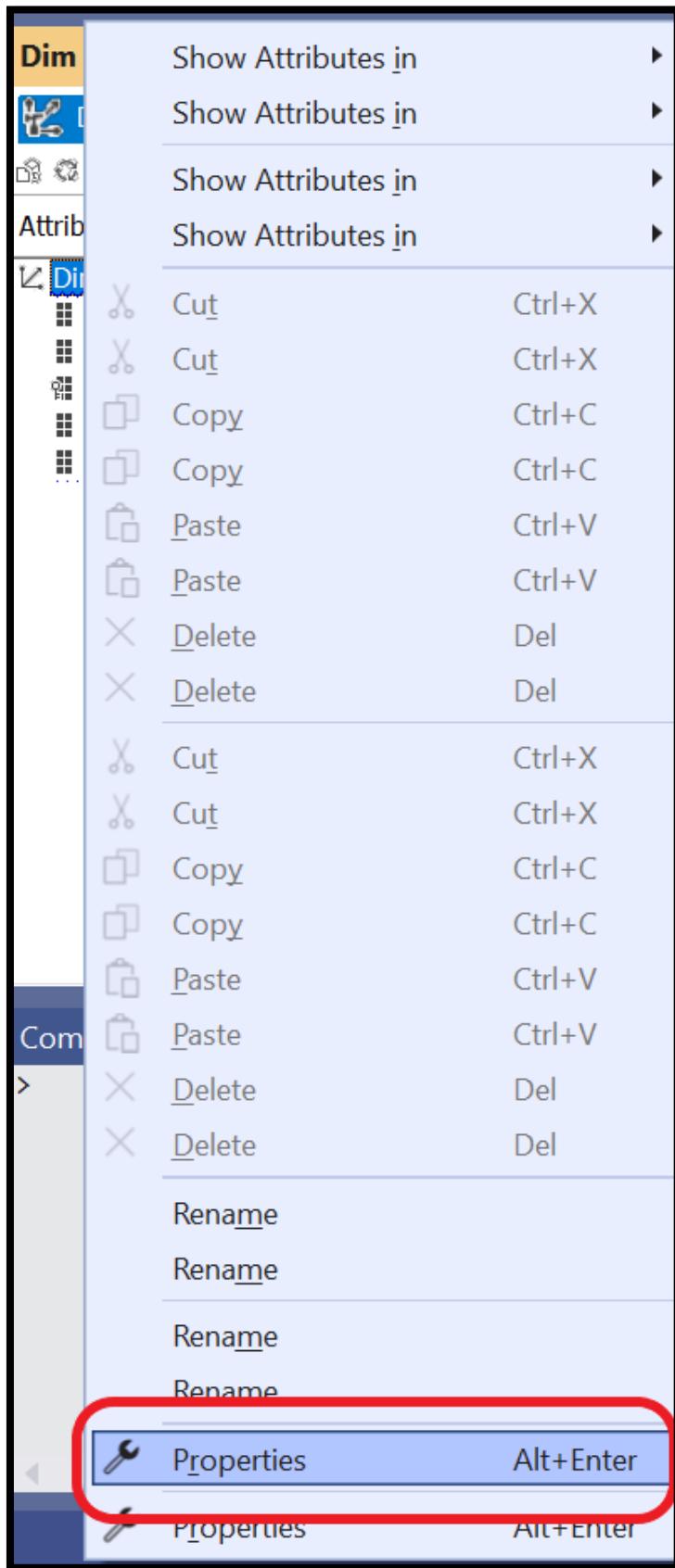
ProcessingRecommendation	None
ProcessingState	Unprocessed
Source	View_DataCoSupplyChain
StringStoresCompatibility	1050
UnknownMember	Hidden
UnknownMemberName	
WriteEnabled	False
<input type="checkbox"/> Basic	
Description	

3.6.6. Cấu hình cho “Dim Shipping Modes.dim”

Bước 1: d-click vào Dim Shipping Modes.dim

The screenshot shows the Microsoft Analysis Services Dimension Structure tool interface. The top menu bar includes 'Dimension Structure', 'Attribute Relationships', 'Translations', and 'Browser'. The ribbon below has tabs for 'Attributes', 'Hierarchies', and 'Data Source View'. The 'Attributes' tab is active, showing a list of attributes: 'Dim Shipping Modes' (selected and highlighted with a red box), 'Shipping Mode', and 'Shipping Mode Id'. The 'Hierarchies' tab contains a placeholder message: 'To create a new hierarchy, drag an attribute here.' The 'Data Source View' tab shows a data source view named 'dim_shipping_modes' with two columns: 'shipping_mode_id' and 'shipping_mode'.

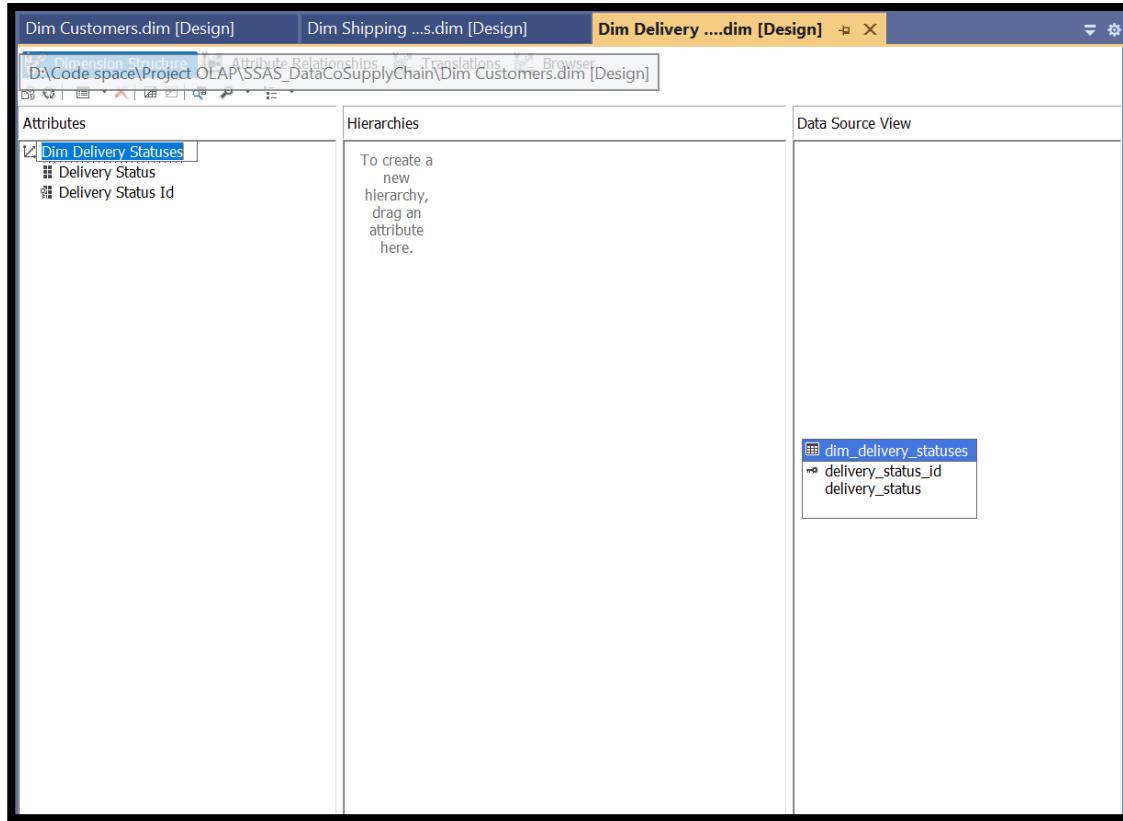
Bước 3: R-click vào **Dim Shipping Modes** bên trong Attribute panel, chọn Properties.



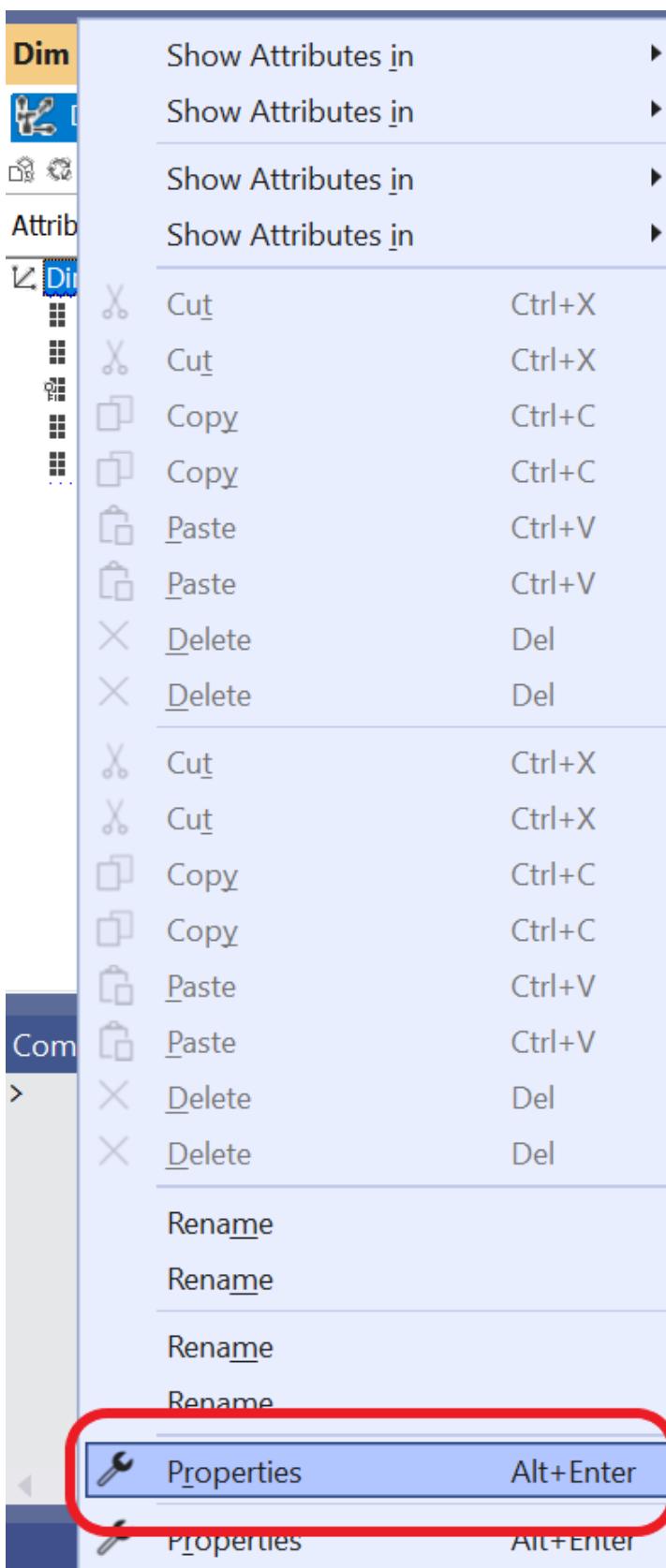
Bước 4: tại UnknownMember chọn Hidden.

3.6.7. Cấu hình cho “Dim Delivery Statuses.dim”

Bước 1: d-click vào Dim Statuses.dim



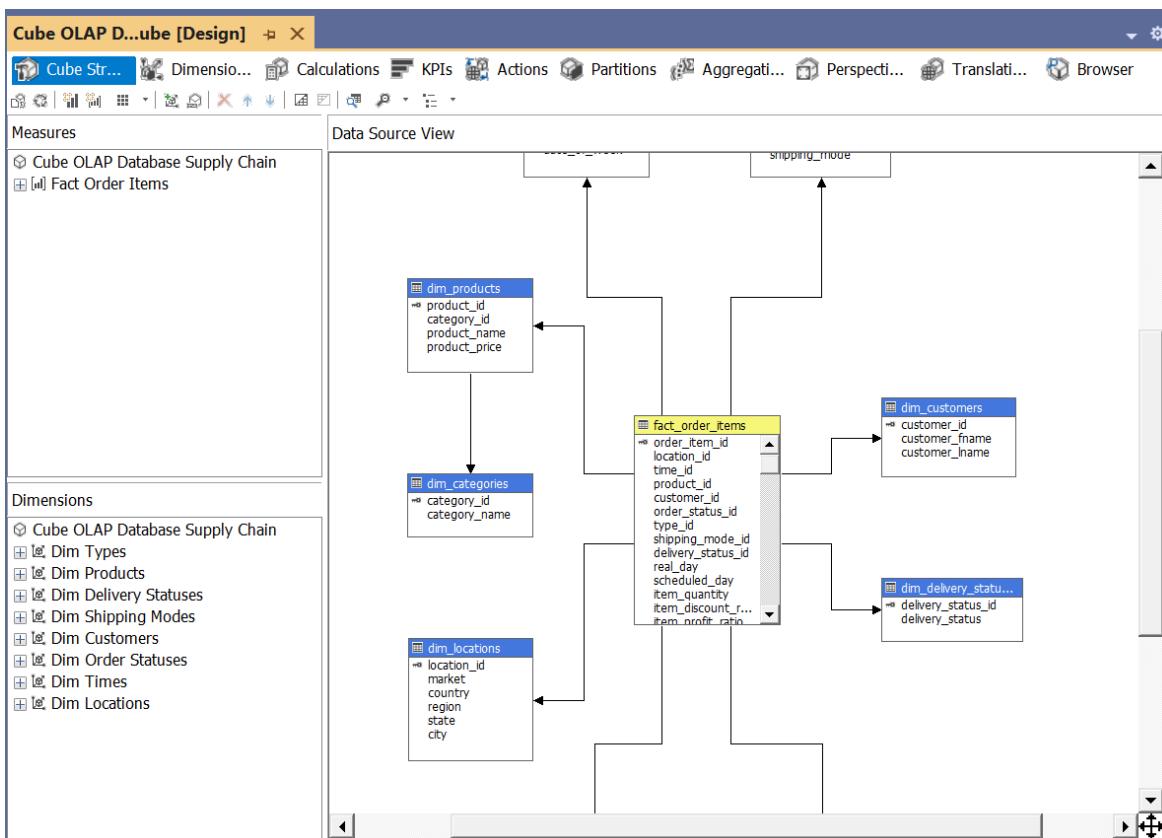
Bước 3: R-click vào **Dim Delivery Statuses** bên trong Attribute panel, chọn Properties.



Bước 4: tại **UnknownMember** chọn **Hidden**.

ProcessingState	Unprocessed
Source	View_DataCoSupplyChain
StringStoresCompatibility	1050
UnknownMember	Hidden
UnknownMemberName	
WriteEnabled	False
Basic	
Description	
ID	Dim_Delivery_Statuses

Cuối cùng ta start để chạy lại Cube và Project



3.7. Độ đo, KPI, Name Set

Độ đo (measures)

3.7.1. Đổi tên và thuộc tính các độ đo ban đầu.

Bước 1: D-click vào cube, chọn Show Measures Grid để hiện thị chi tiết các độ đo.

Chi tiết các độ đo sẽ hiển thị dưới dạng bảng, dễ dàng để tương tác.

The screenshot shows the Microsoft Analysis Services Cube Designer interface. The title bar says "Cube_DataCoS...ube [Design]". The ribbon menu has tabs: "Cube Structure" (selected), "Dimension Usage", "Calculations", "KPIs", and "Actions". Below the ribbon is a toolbar with various icons. The main area is titled "Measures" and contains a table:

	Name	Measure Group	Data Type	Aggregation
	Real Day	Fact Order Items	Integer	Sum
	Scheduled Day		Integer	Sum
	Item Quantity		Integer	Sum
	Item Discount Rate		Double	Sum
	Item Profit Ratio		Double	Sum
	Order Item Total		Double	Sum
	Benefit Per Order		Double	Sum
	Fact Order Items Count		Integer	Count

Bước 2: Ta đổi tên và thuộc tính các độ đo hiện tại theo các hàm tổng hợp (aggregation) như sau:

- Real Day (thời gian giao hàng thực tế) đổi thành AVG Real Day For Shipping (trung bình thời gian giao hàng thực tế), sau đó đổi Aggregation thành **AverageOfChildren**, và đổi Data type thành **Double**
- Schedule Day (thời gian giao hàng kế hoạch) đổi thành AVG Schedule Day For Shipping (trung bình thời gian giao hàng kế hoạch), sau đó đổi Aggregation thành **AverageOfChildren**, và đổi Data type thành **Double**
- Item Quantity (Số lượng item trong mỗi đơn hàng) đổi thành Sum Item Quantity (Tổng số item), sau đó đổi Aggregation thành **Sum**, và đổi Data type thành **Intger**
- Discount Rate (Tỷ lệ giảm giá của item trong order) đổi thành AVG Discount Rate (trung bình tỷ lệ giảm giá của item trong order), sau đó đổi **AverageOfChildren**, và đổi Data type thành **Double**.

- Item Profit Ratio (Tỷ lệ lê nhuận trong mỗi đơn hàng) đổi thành AVG Item Profit Ratio (trung bình tỷ lệ giảm giá của item trong order), sau đó đổi **AverageOfChildren**, và đổi Data type thành **Double**
- Order Item Total (Tổng giá tiền mỗi order) đổi thành AVG Order Item Total (trung bình tỷ lệ giảm giá của item trong order), sau đó đổi **AverageOfChildren**, và đổi Data type thành **Double**.
- Benefit Per Order (Tổng lợi nhuận mỗi order) đổi thành AVG Benefit Per Order (Tổng lợi nhuận mỗi order), sau đó đổi **AverageOfChildren**, và đổi Data type thành **Double**.
- Fact Order Items Count (số lượng item order) đổi thành Count Fact Order Items (Đếm số lượng Item order), sau đó đổi **Sum**, và đổi Data type thành **Integer**.

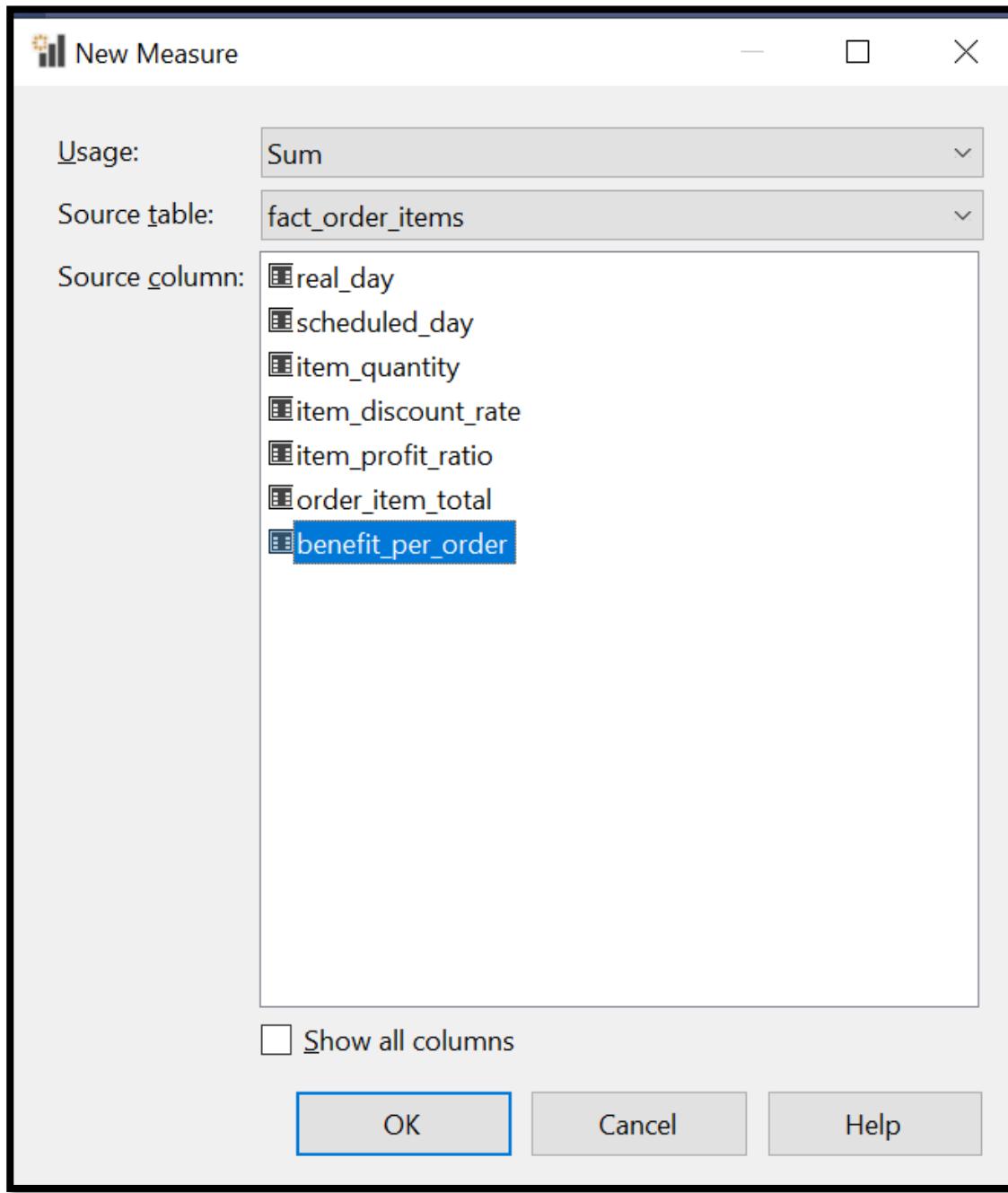
The screenshot shows the Microsoft Analysis Services (SSAS) Cube Designer interface. The title bar reads "Cube_DataCo...ube [Design]". The main area is a table titled "Measures" with the following data:

Name	Measure Group	Data Type	Aggregation
AVG Real Day For Shipping		Double	Average...
AVG Schedule Day For Shipping		Double	Average...
Sum Item Quantity		Integer	Sum
AVG Item Discount Rate		Double	Average...
AVG Item Profit Ratio		Double	Average...
AVG Order Item Total		Double	Average...
AVG Benefit Per Order	Fact Order Items	Double	Average...
Fact Order Items Count		Integer	Count

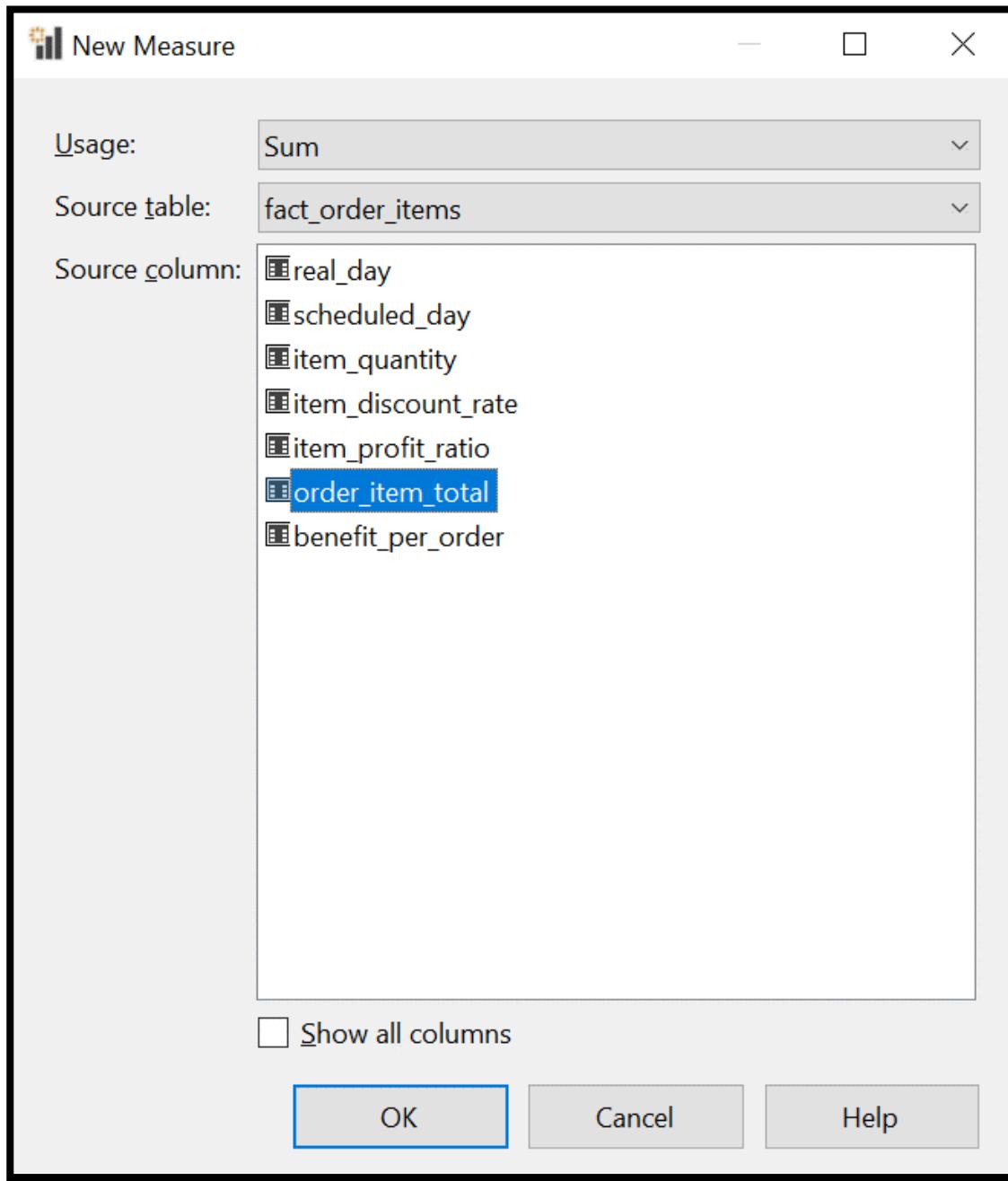
3.7.2. Tạo các độ đo mới

Để tạo ra các độ đo mới, ta chọn Add new measure..., sau đó tạo ra thêm một độ đo mới như sau:

- Sum Benefit Per Order (Tổng lợi nhuận trên mỗi đơn hàng)



- Sum Benefit Per Order (Tổng Doanh trên mỗi đơn hàng)



Kết quả

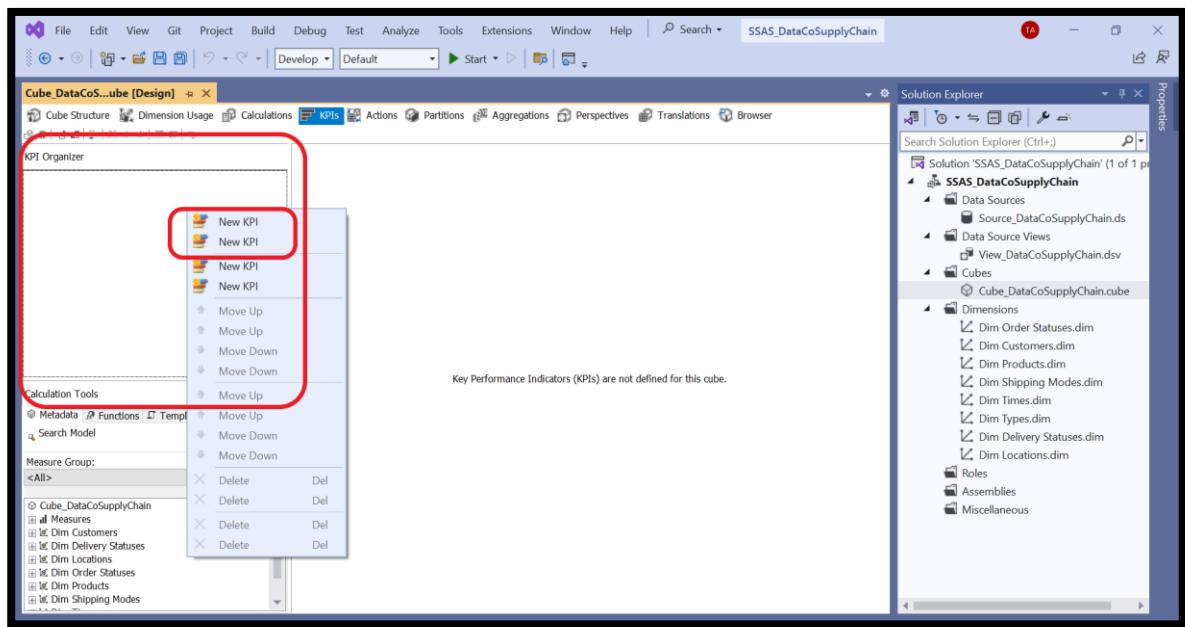
Name	Measure Group	Data Type	Aggregation
AVG Real Day For Shipping		Double	Average...
AVG Schedule Day For Shipping		Double	Average...
Sum Item Quantity		Integer	Sum
AVG Item Discount Rate		Double	Average...
AVG Item Profit Ratio		Double	Average...
AVG Order Item Total		Double	Average...
AVG Benefit Per Order		Double	Average...
Fact Order Items Count		Integer	Count
Sum Benefit Per Order		Double	Sum
Sum Order Item Total		Double	Sum
Add new measure...			

3.7.3. Tạo KPI

3.7.3.1. Tạo KPI về sự tăng trưởng lợi nhuận. Giá trị là trung bình lợi nhuận trên toàn thế giới.

Bước 1: D-click vào Cube, chọn KPIs để tạo mới KPI

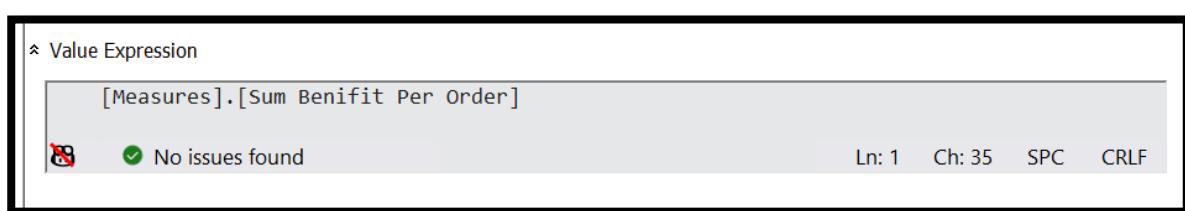
Bước 2: R-click vào ô **KPI Organizer**, chọn New KPI để tạo mới KPI



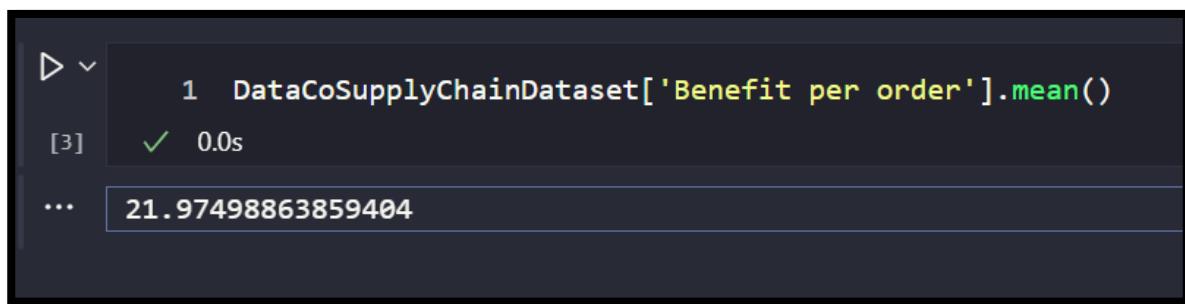
Bước 3: Nhập tên KPI và chọn loại độ đo (Tại đây chọn độ đo trong fact)



Bước 4: Tại **Value Expression**, nhập vào độ đo được sử dụng để tính toán.

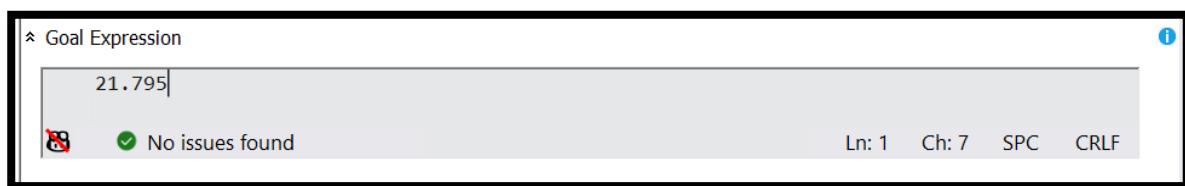


Bước 5: Tại **Goal Expression**, nhập vào giá trị **goal**.



```
1 DataCoSupplyChainDataset['Benefit per order'].mean()
[3] ✓ 0.0s
...
21.97498863859404
```

Dựa vào Jupyter, tính toán được giá trị lợi nhuận trung bình trên toàn thế giới trong mọi thời gian.



Bước 6: **Status expression** (status sẽ tính ra giá trị KPI tại thời điểm đó, kiểm tra so sánh với goal), status bên dưới kiểm tra giá trị tại thời điểm đó có lớn hơn mức trung bình thế giới không

CASE

WHEN IsEmpty([Measures].[Sum Benefit Per Order])

THEN "no data"

WHEN KPIVALUE("Benefit KPI") > KPIGOAL("Benefit KPI")

THEN "greater than AVG benefit"

WHEN KPIVALUE("Benefit KPI") = KPIGOAL("Benefit KPI")

THEN "Is AVG benefit"

WHEN KPIVALUE("Benefit KPI") < KPIGOAL("Benefit KPI")

THEN "less than AVG benefit"

END

```
 CASE
    WHEN IsEmpty([Measures].[Sum Benefit Per Order])
    THEN "no data"
    WHEN KPIVALUE("Growth KPI") > KPIGOAL("Growth KPI")
    THEN "greater than AVG benefit"
    WHEN KPIVALUE("Growth KPI") = KPIGOAL("Growth KPI")
    THEN "Is AVG benefit"
    WHEN KPIVALUE("Growth KPI") < KPIGOAL("Growth KPI")
    THEN "less than AVG benefit"
END
```

No issues found

Ln: 7 Ch: 24 SPC CRLF

Bước 7: Trend expression (Trend sẽ tính ra xu hướng tại thời điểm đó), bên dưới là trên để so sánh năm tiếp theo có tăng trưởng so với năm trước hay không

CASE

WHEN IsEmpty([Measures].[Sum Benefit Per Order])

THEN "no data"

WHEN KPIVALUE("Benefit KPI") > ([Dim Times].[Year].PREV MEMBER,
KPIVALUE("Benefit KPI"))

THEN "Growth"

WHEN KPIVALUE("Benefit KPI") < ([Dim Times].[Year].PREV MEMBER,
KPIVALUE("Benefit KPI"))

THEN "Reduce Growth"

END

Trend indicator: ▼ Reversed status arrow

Trend expression:

```

CASE
    WHEN IsEmpty([Measures].[Sum Benefit Per Order])
    THEN "no data"
    WHEN KPIVALUE("Growth KPI") > ([Dim Times].[Year].PREV MEMBER, KPIVALUE("Growth KPI"))
    THEN "Growth"
    WHEN KPIVALUE("Growth KPI") < ([Dim Times].[Year].PREV MEMBER, KPIVALUE("Growth KPI"))
    THEN "Reduce Growth"
END

```

✖ ✓ No issues found

Ln: 8 Ch: 4 SPC CRLF

*Ví dụ về Trend tính toán dựa tên [Measure][Sum Benefit Per Order]

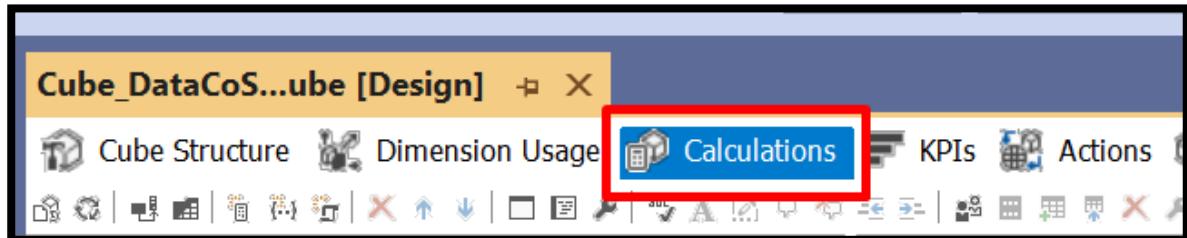
The screenshot shows the Analysis Manager application window. On the left, there's a navigation pane with 'Cube_DataCoSupplyChain' selected. Below it, under 'Measure Group', is a tree view with 'All' expanded, showing various measures like 'AVG Real Day For Shipping', 'Fact Order Items Count', etc., and 'KPIs' expanded, showing 'Growth KPI' with its sub-items: 'Value', 'Goal', 'Status', and 'Trend'. On the right, there's a large table grid with columns: Dimension, Hierarchy, Operator, Filter Expression, and Parameters. The table data is as follows:

Year	Sum Benefit Per Order	Growth KPI Value	Growth KPI Goal	Growth KPI Status	Growth KPI Trend
2015	1306027.73000001	1306027.73000...	21.795	greater than AVG ...	Growth
2016	1311771.53	1311771.53	21.795	greater than AVG ...	Growth
2017	1307124.66	1307124.66	21.795	greater than AVG ...	Reduce Growth
2018	41470.5000000001	41470.5000000...	21.795	greater than AVG ...	Reduce Growth

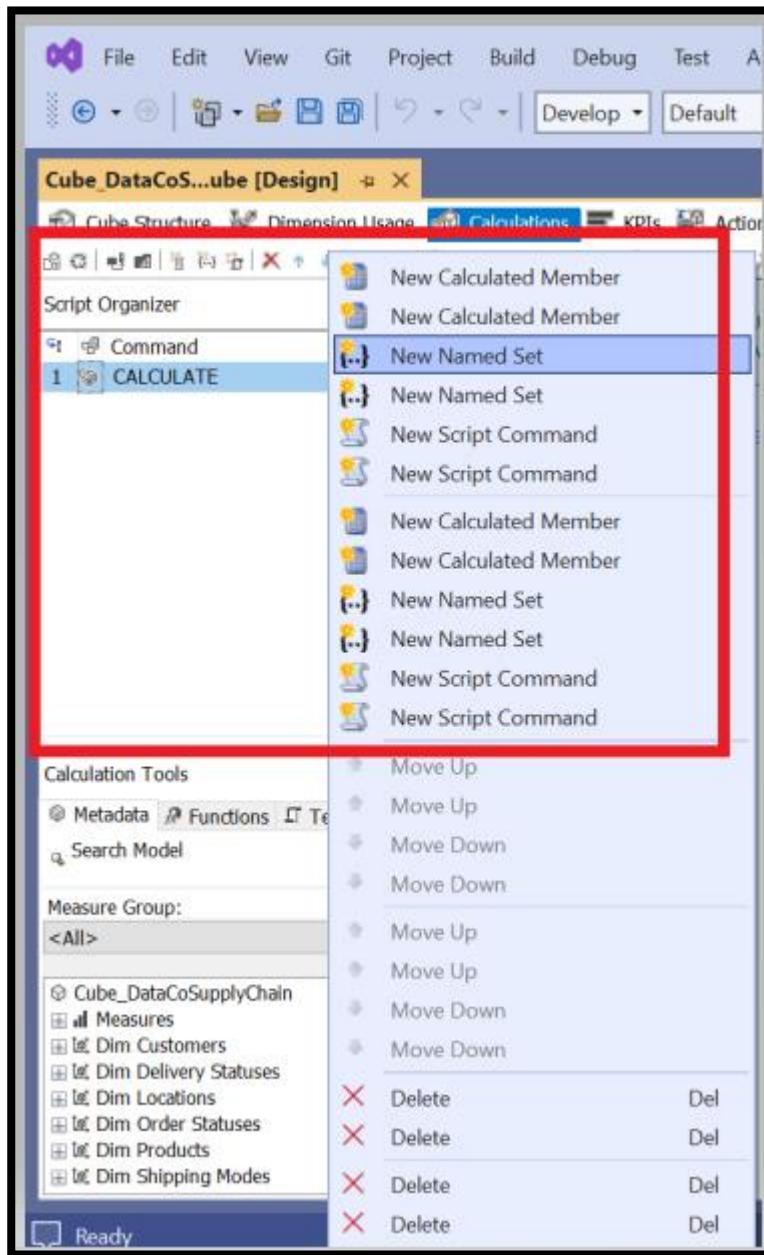
3.7.4. Tạo nameset

3.7.4.1. Top 5 sản phẩm được mua nhiều nhất

Bước 1: D-click vào cube, chọn Calculations

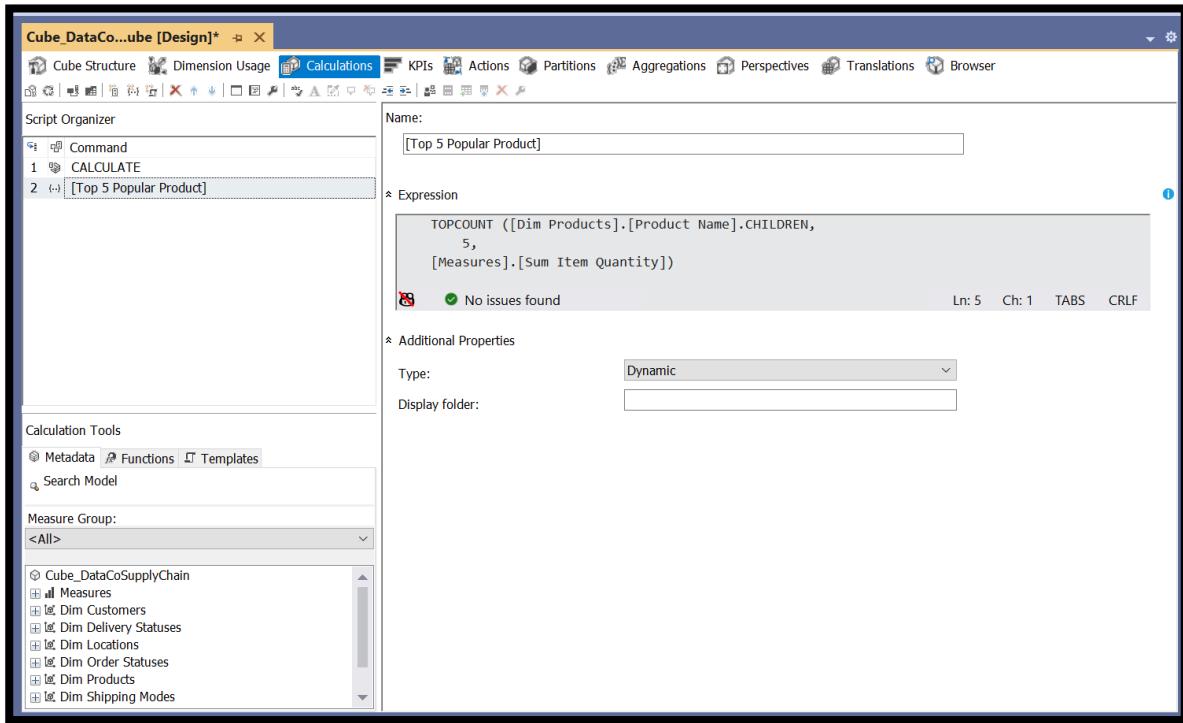


Bước 2: R-click vào script organization, chọn New Named Set



Bước 3: Nhập tên và MDX script

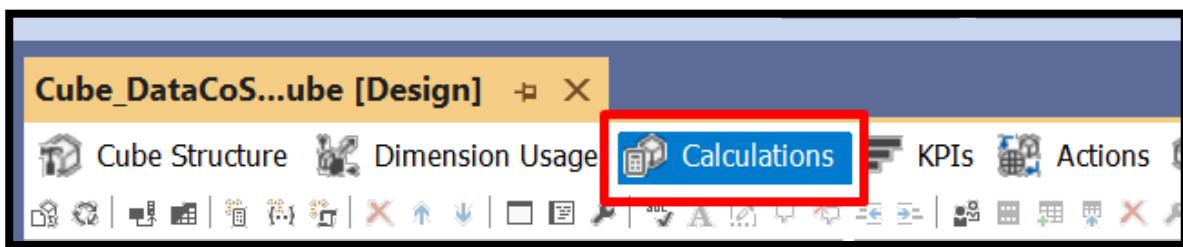
```
TOPCOUNT ( [Dim Products].[Product Name].CHILDREN,
            5,
            [Measures].[Sum Item Quantity])
```



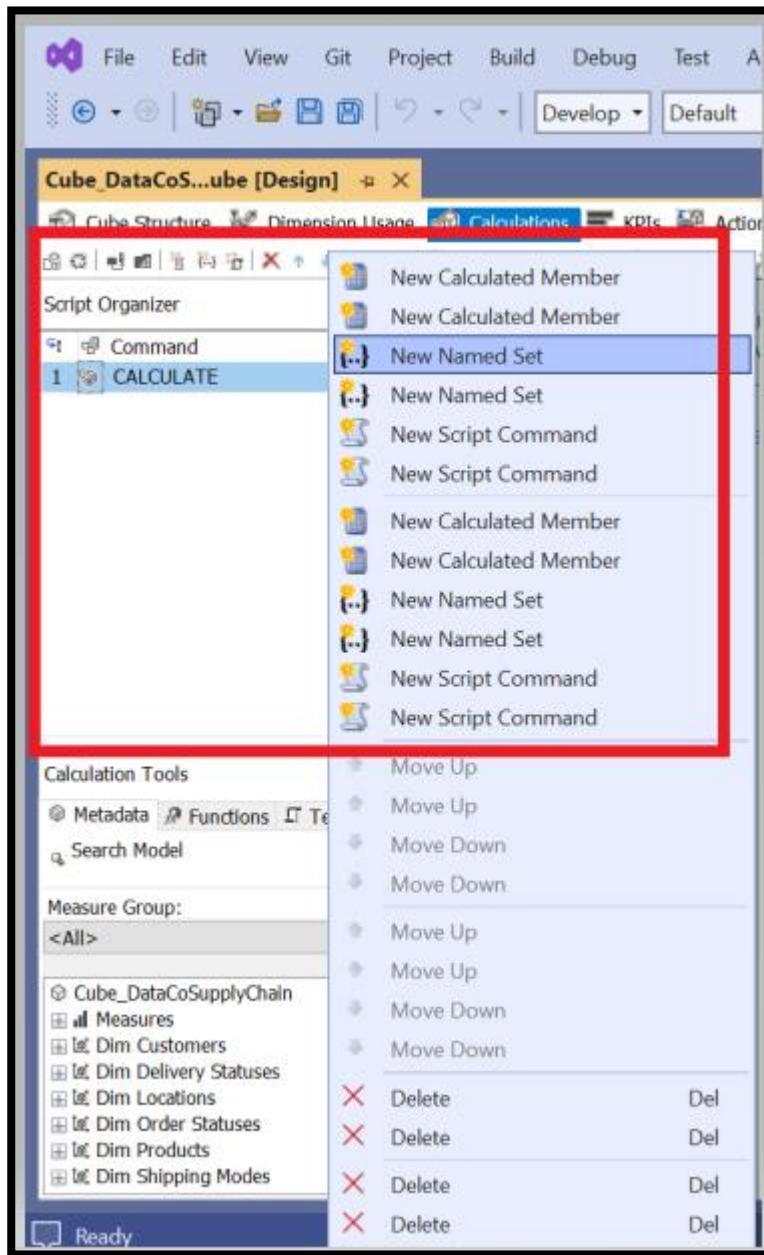
Bước 4: process lại cube

3.7.4.2. Danh sách các nước có doanh thu âm

Bước 1: D-click vào cube, chọn Calculations



Bước 2: R-click vào script organization, chọn New Named Set



Bước 3: Nhập tên và MDX script

```

ORDER (
    FILTER ( [Dim Locations].[Country].CHILDREN,
        [Measures].[Sum Benefit Per Order] <> 0),
    [Measures].[Sum Benefit Per Order],
    ASC)

```

The screenshot shows the 'Calculations' dialog box in Microsoft Analysis Services. At the top, there is a 'Name:' field containing '[Negative Benefit]'. Below it is an 'Expression' pane displaying the following MDX script:

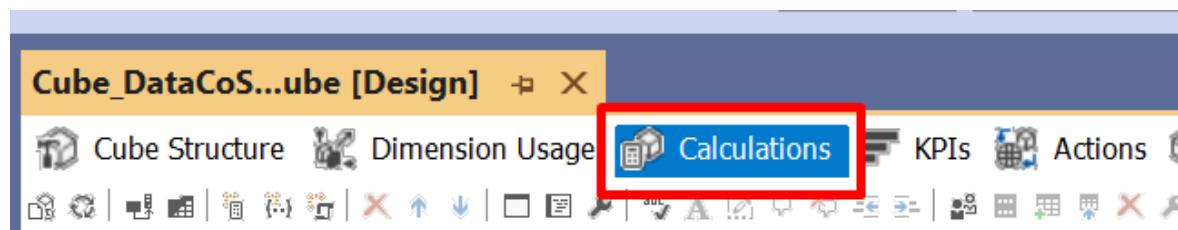
```
ORDER (
    FILTER ( [Dim Locations].[Country].CHILDREN,
        [Measures].[Sum Benefit Per Order] <0),
    [Measures].[Sum Benefit Per Order],
    ASC)
```

Below the expression, there is a status message 'No issues found' and some system information: Ln: 7 Ch: 1 TABS CRLF. Under the 'Additional Properties' section, the 'Type:' dropdown is set to 'Dynamic'.

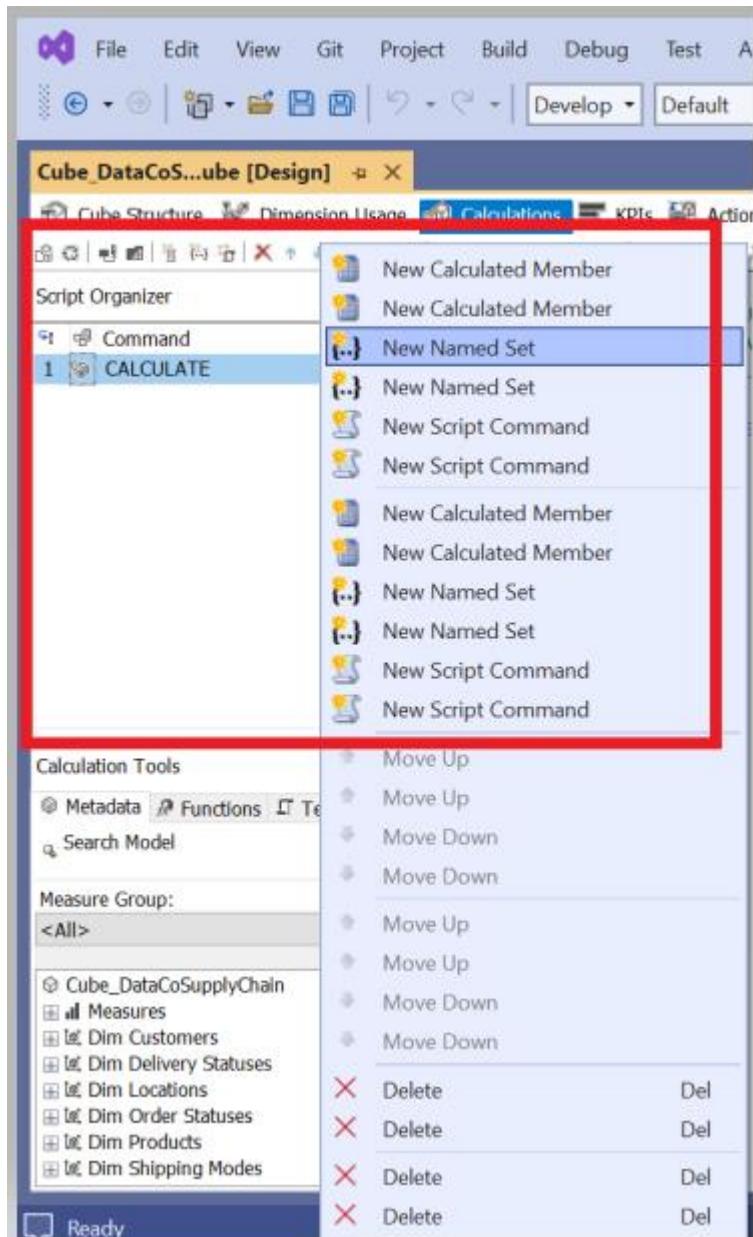
Bước 4: process lại cube

3.7.4.3. Top 5 quốc gia có lợi nhuận cao nhất không phải Mỹ

Bước 1: D-click vào cube, chọn Calculations



Bước 2: R-click vào script organization, chọn New Named Set



Bước 3: Nhập tên và MDX script

TOPCOUNT (

ORDER (

EXCEPT(

[Dim Locations].[Country].CHILDREN,

[Dim

Locations].[Country].&[Estados Unidos]&[USCA]

```
),
[Measures].[Sum Benefit Per Order],
ASC
),
5,
[Measures].[Sum Benefit Per Order]
)
```

Name: [Top 5 Countries not USA]

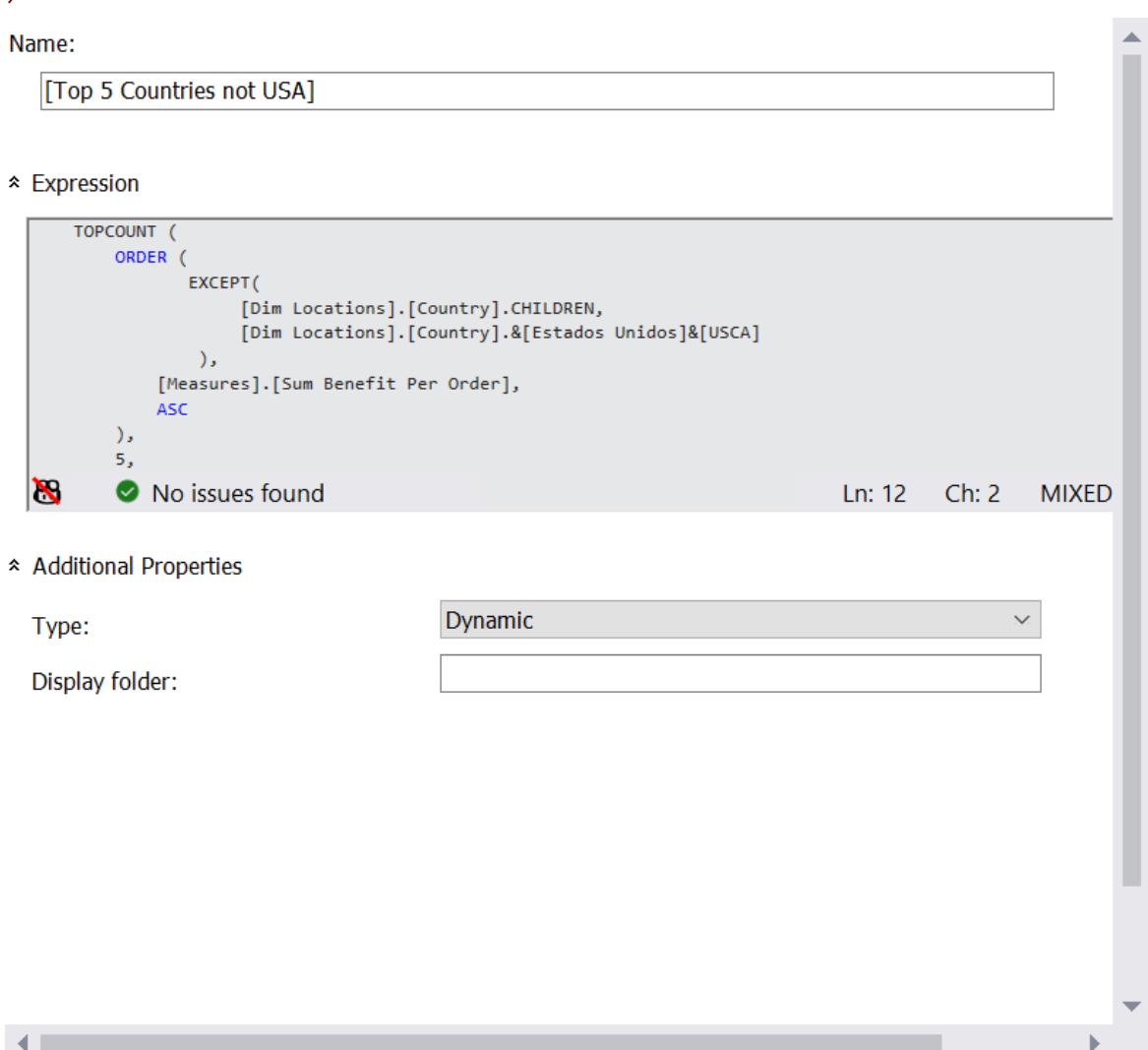
Expression

```
TOPCOUNT (
    ORDER (
        EXCEPT(
            [Dim Locations].[Country].CHILDREN,
            [Dim Locations].[Country].&[Estados Unidos]&[USCA]
        ),
        [Measures].[Sum Benefit Per Order],
        ASC
    ),
    5,
    No issues found
)
Ln: 12 Ch: 2 MIXED
```

Additional Properties

Type: Dynamic

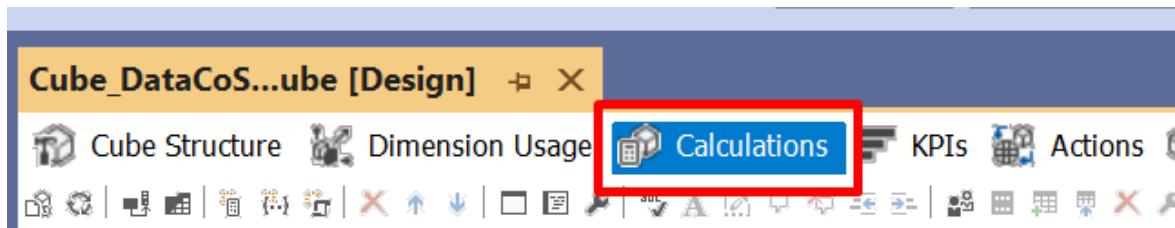
Display folder:



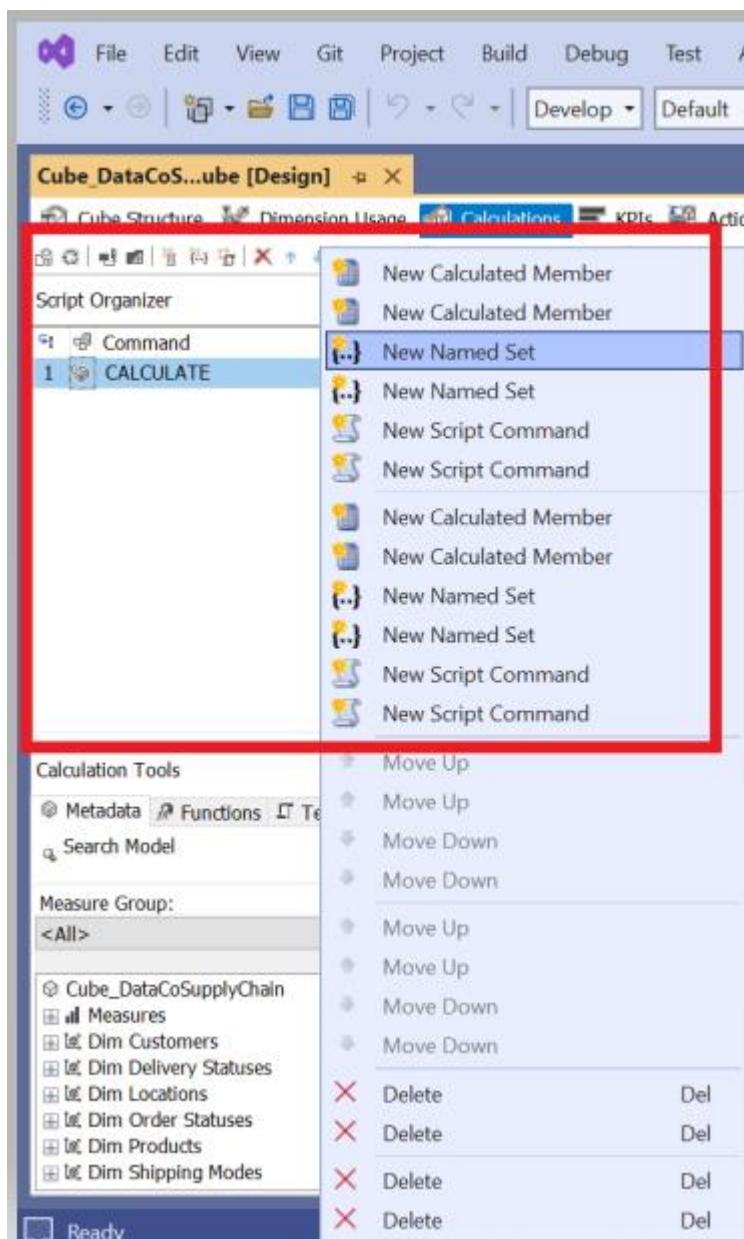
Bước 4: process lại cube

Danh sách các nước có lợi nhuận thấp nhất (lợi nhuận không âm)

Bước 1: D-click vào cube, chọn Calculations



Bước 2: R-click vào script organization, chọn New Named Set



Bước 3: Nhập tên và MDX script

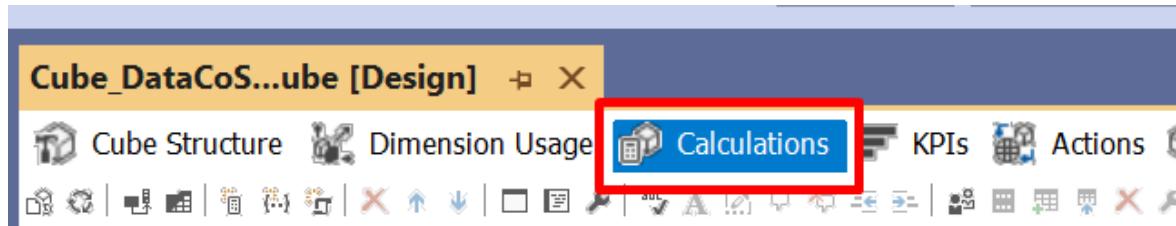
```
TAIL(
    ORDER(
        FILTER(
            [Dim Locations].[Country].CHILDREN,
            [Measures].[Sum Benefit Per Order] > 0
        ),
        [Measures].[Sum Benefit Per Order],
        DESC
    ),
    5
)
```

The screenshot shows the Microsoft Analysis Services MDX Editor interface. At the top, there is a toolbar with various icons. Below the toolbar, the main area has a title bar with the text "Name:" followed by a text input field containing "[Tail 5 Country Lowest Benefit no Negative]". To the right of the input field is an information icon (i). Below the title bar, there is a section titled "Expression" which contains the MDX code shown in the previous code block. At the bottom left of the expression editor, there is a status bar with the text "No issues found". To the right of the status bar are buttons for "Ln: 11", "Ch: 2", "TABS", and "CRLF". On the left side of the main area, there is a tree view with nodes like "Additional Properties", "Type: Dynamic", and "Display folder:". The background of the editor is light gray.

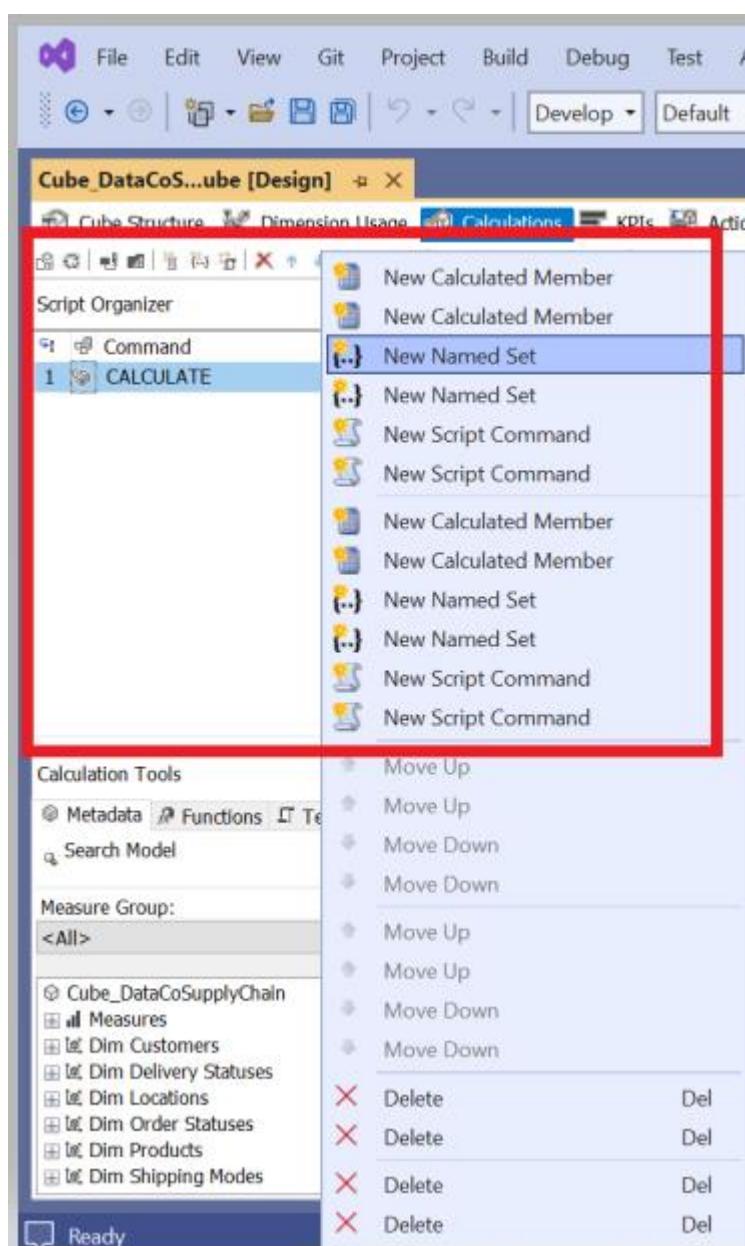
Bước 4: process lại cube

3.7.4.4. Danh sách các khách hàng đem lại lợi nhuận > 700 từ Cash hoặc > 1200 bằng Debit

Bước 1: D-click vào cube, chọn Calculations



Bước 2: R-click vào script organization, chọn New Named Set



Bước 3: Nhập tên và MDX script

UNION

(

FILTER(

([Dim Customers].[Customer Id].CHILDREN,[Dim Customers].[Customer Fname].CHILDREN,[Dim Customers].[Customer Lname].CHILDREN),

([Measures].[Sum Benefit Per Order], [Dim Types].[Type].&[CASH]) > 700

),

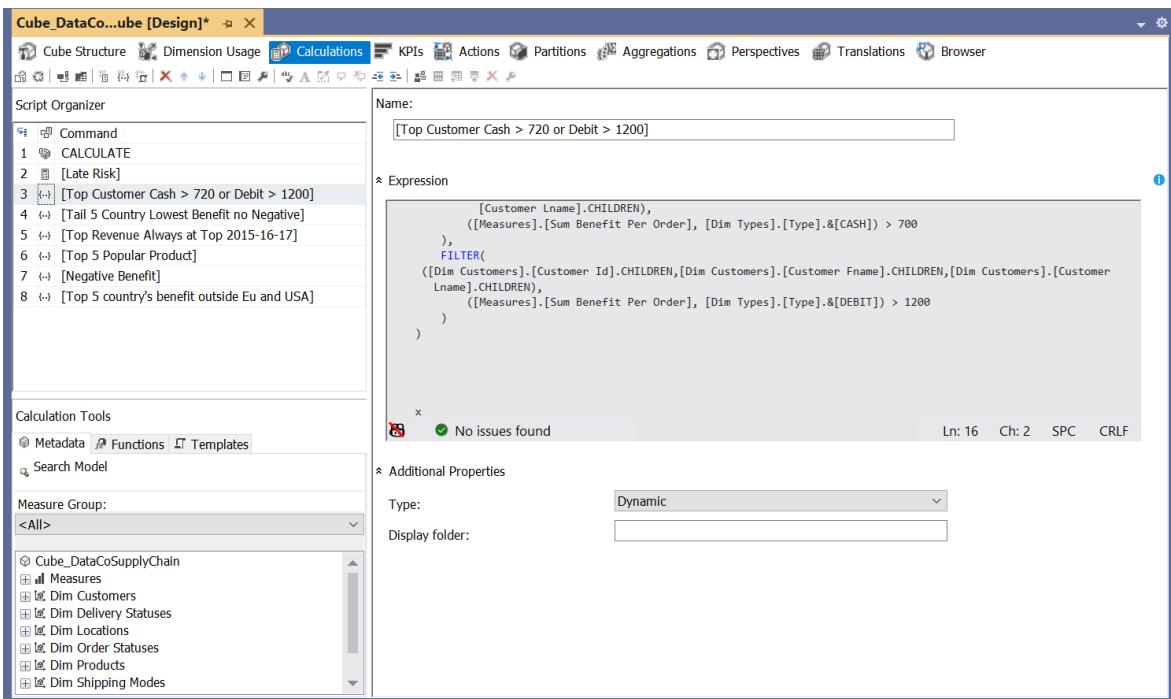
FILTER(

([Dim Customers].[Customer Id].CHILDREN,[Dim Customers].[Customer Fname].CHILDREN,[Dim Customers].[Customer Lname].CHILDREN),

([Measures].[Sum Benefit Per Order], [Dim Types].[Type].&[DEBIT]) > 1200

)

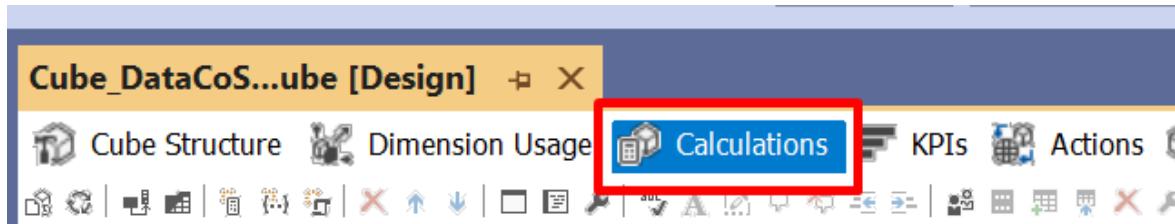
)



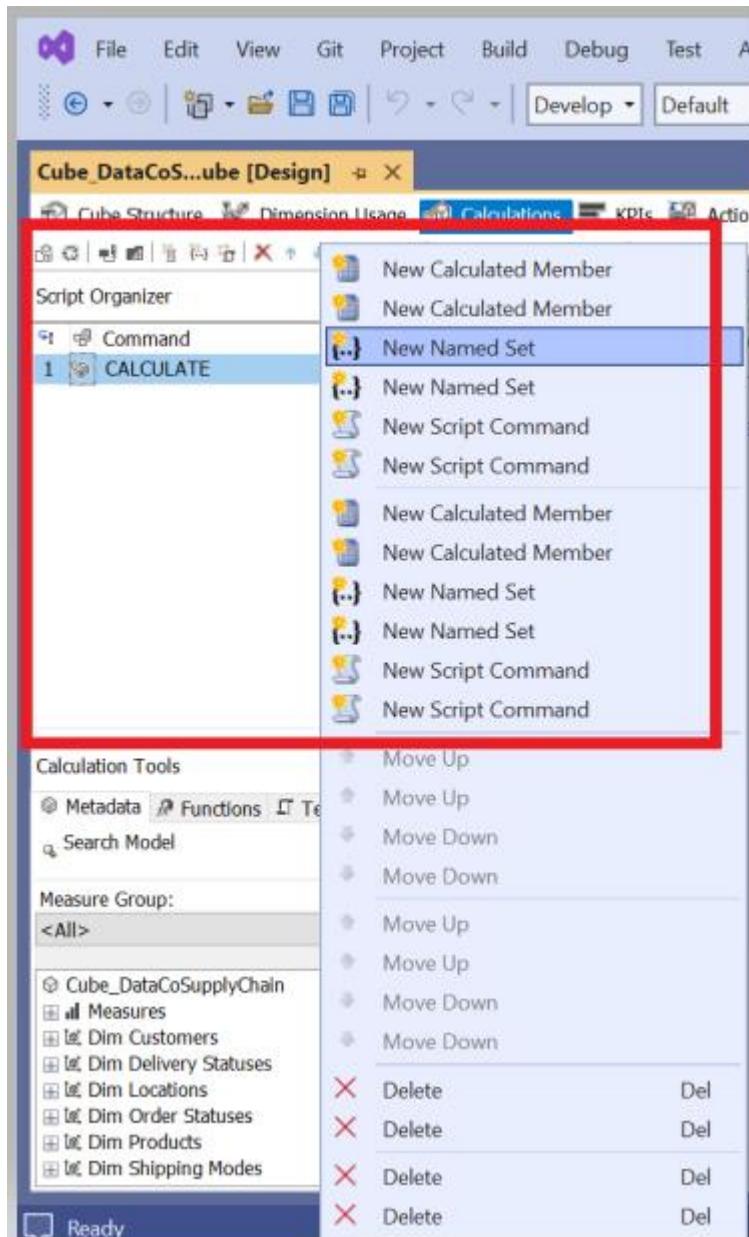
Bước 4: process lại cube

3.7.4.5. Danh sách Top 10 khách hàng đem lại lợi nhuận nhiều nhất

Bước 1: D-click vào cube, chọn Calculations



Bước 2: R-click vào script organization, chọn New Named Set



Bước 3: Nhập tên và MDX script

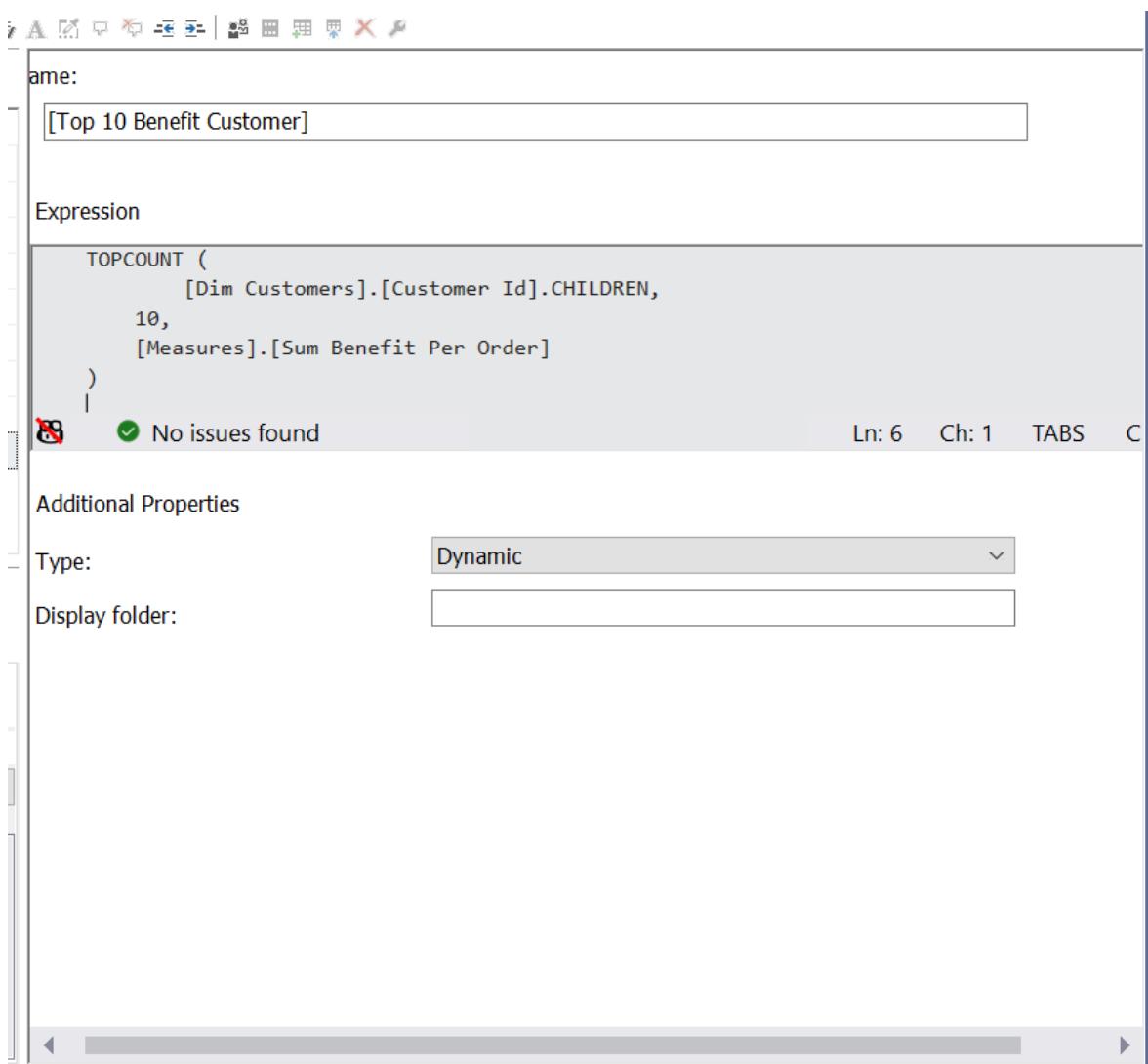
TOPCOUNT (

[Dim Customers].[Customer Id].CHILDREN,

10,

[Measures].[Sum Benefit Per Order]

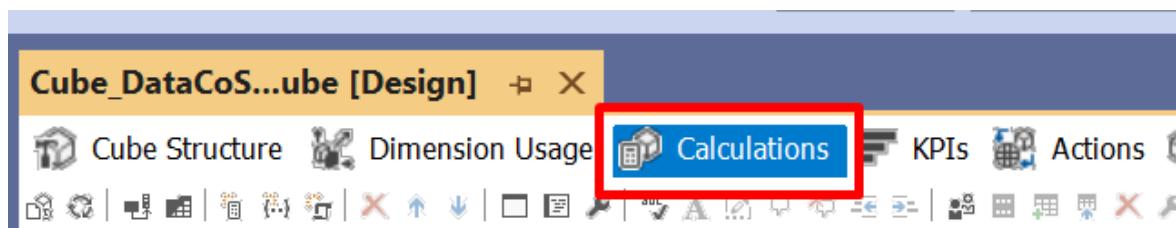
)



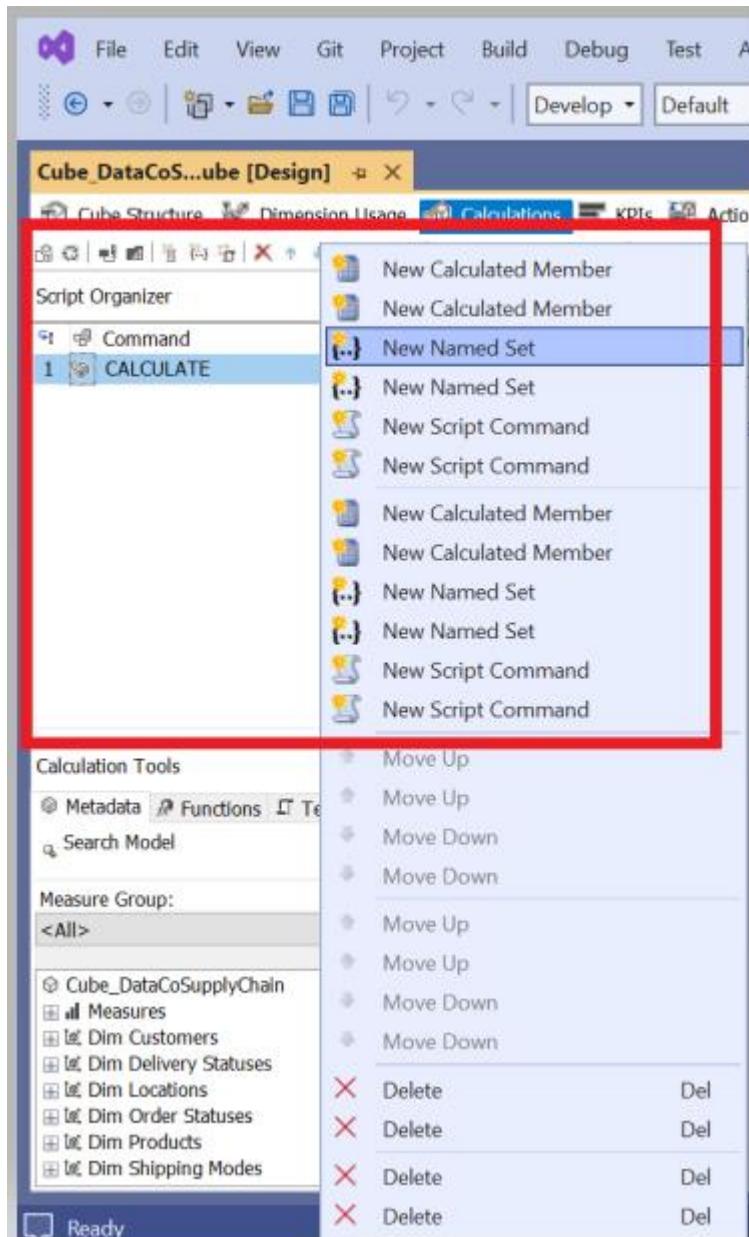
Bước 4: process lại cube

3.7.4.6. Danh sách Top 1 quốc gia đem lại nhiều lợi nhuận nhất

Bước 1: D-click vào cube, chọn Calculations



Bước 2: R-click vào script organization, chọn New Named Set



Bước 3: Nhập tên và MDX script

TOPCOUNT (

[Dim Locations].[Country].CHILDREN,

1,

[Measures].[Sum Benefit Per Order]

)

The screenshot shows the Microsoft Analysis Services (MAS) Cube OLAP Designer in design mode. The top menu bar includes 'File', 'Edit', 'View', 'Tools', 'Help', and a 'Cube OLAP D...ube [Design]' tab. The ribbon tabs include 'Calculations' (which is selected), 'KPIs', 'Actions', 'Partitions', 'Aggregates', 'Perspectives', 'Translations', and 'Browser'. The left sidebar contains sections for 'Script Organizer' (listing various calculated members like '[Late Risk]', '[Top 5 Product's Quantity]', etc.), 'Calculation Tools' (with tabs for 'Metadata', 'Functions', and 'Templates'), and 'Measure Group' (listing categories like 'Cube OLAP Database Supply Chain', 'Measures', 'Dim Customers', etc.). The main workspace on the right shows a named set named '[Top 1 Benefit Country]' with the following expression:

```
TOPCOUNT(
    [Dim Locations].[Country].CHILDREN,
    1,
    [Measures].[Sum Benefit Per Order]
)
```

The status bar at the bottom right indicates 'Ln: 7 Ch: 1 SPC C'.

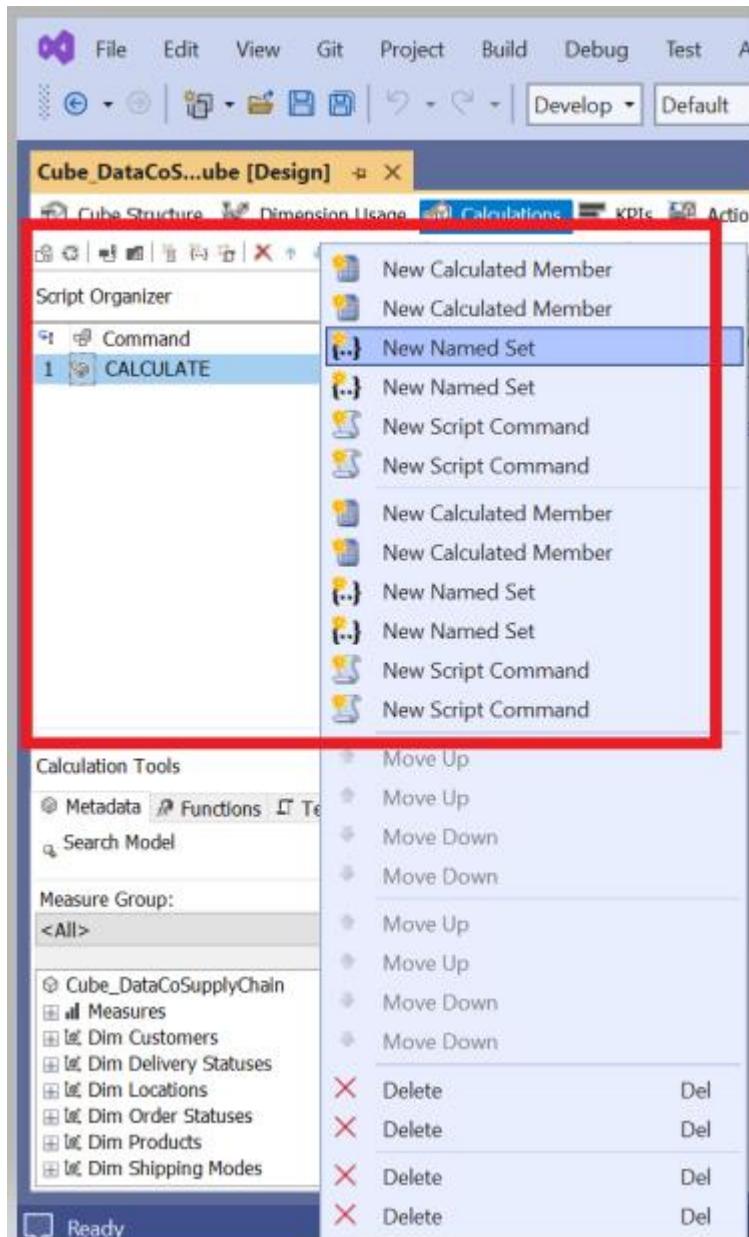
Bước 4: process lại cube

3.7.4.7. Danh sách top 10 quốc gia đem lại ít lợi nhuận nhất (không âm)

Bước 1: D-click vào cube, chọn Calculations

The screenshot shows the Microsoft Analysis Services (MAS) Cube DataCoS...ube in design mode. The top menu bar includes 'File', 'Edit', 'View', 'Tools', 'Help', and a 'Cube DataCoS...ube [Design]' tab. The ribbon tabs include 'Calculations' (which is highlighted with a red box), 'KPIs', 'Actions', and others. The left sidebar contains sections for 'Cube Structure', 'Dimension Usage', and 'Script Organization'. The main workspace on the right is currently empty.

Bước 2: R-click vào script organization, chọn New Named Set



Bước 3: Nhập tên và MDX script

TAIL(

ORDER(

FILTER(

[Dim Locations].[Country].CHILDREN,

[Measures].[Sum Benefit Per Order] > 0

),

[Measures].[Sum Benefit Per Order],

DESC

),

5

)

Name: [Top Lowest Benefit not Negative]

Expression

```
TAIL(
    ORDER(
        FILTER(
            [Dim Locations].[Country].CHILDREN,
            [Measures].[Sum Benefit Per Order] > 0
        ),
        [Measures].[Sum Benefit Per Order],
        DESC
    ),
    5
)
```

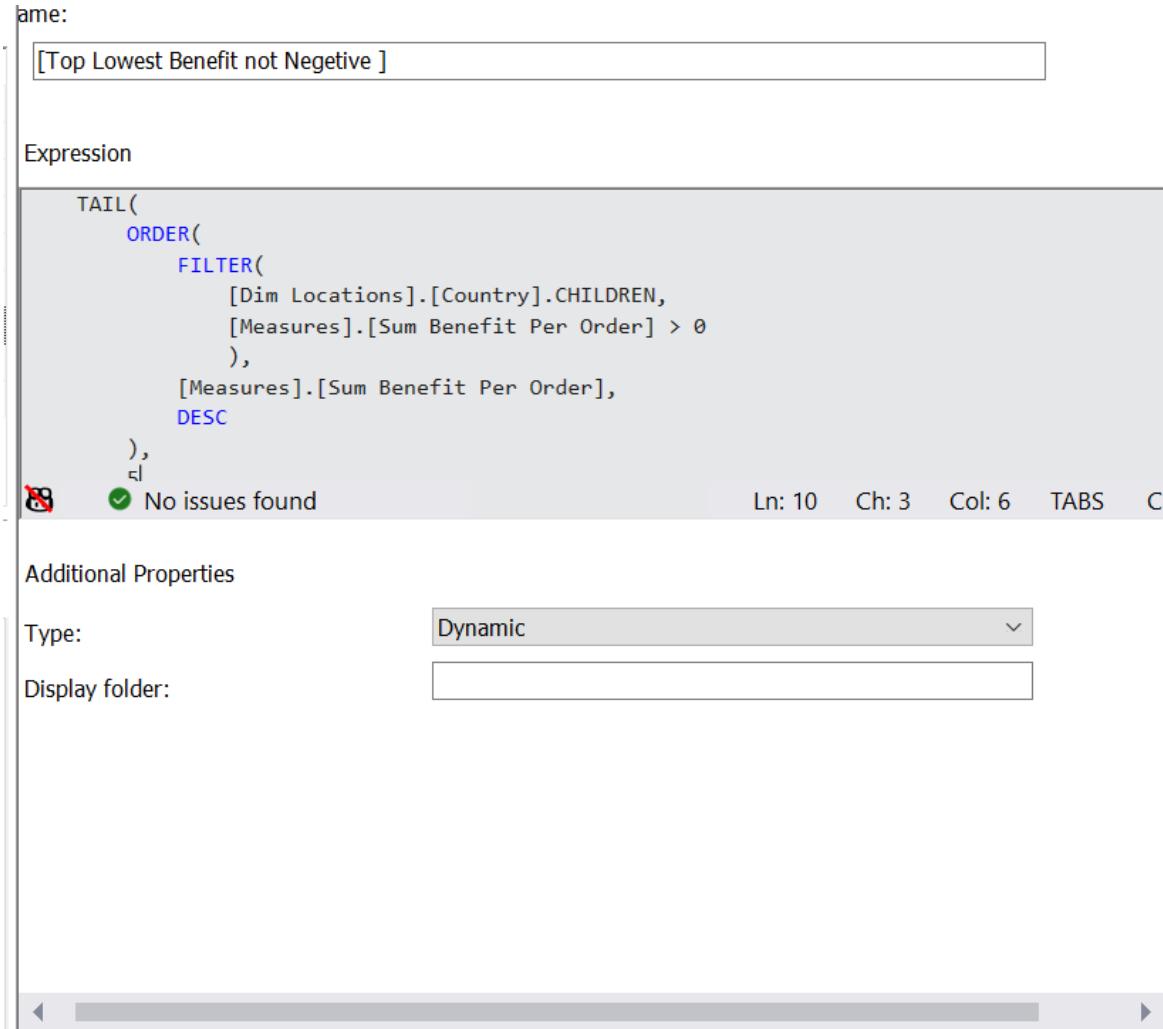
No issues found

Ln: 10 Ch: 3 Col: 6 TABS C

Additional Properties

Type: Dynamic

Display folder:

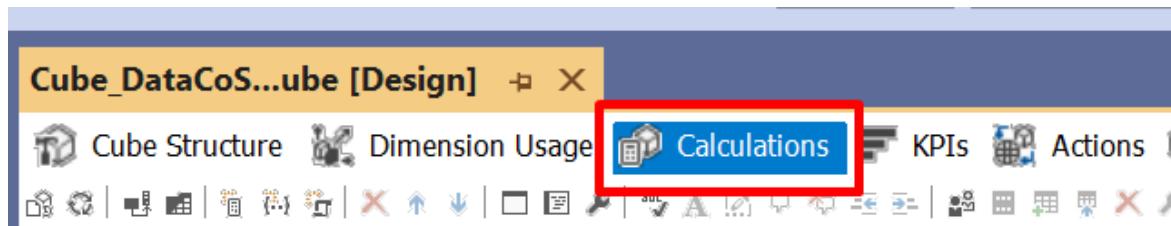


Bước 4: process lại cube

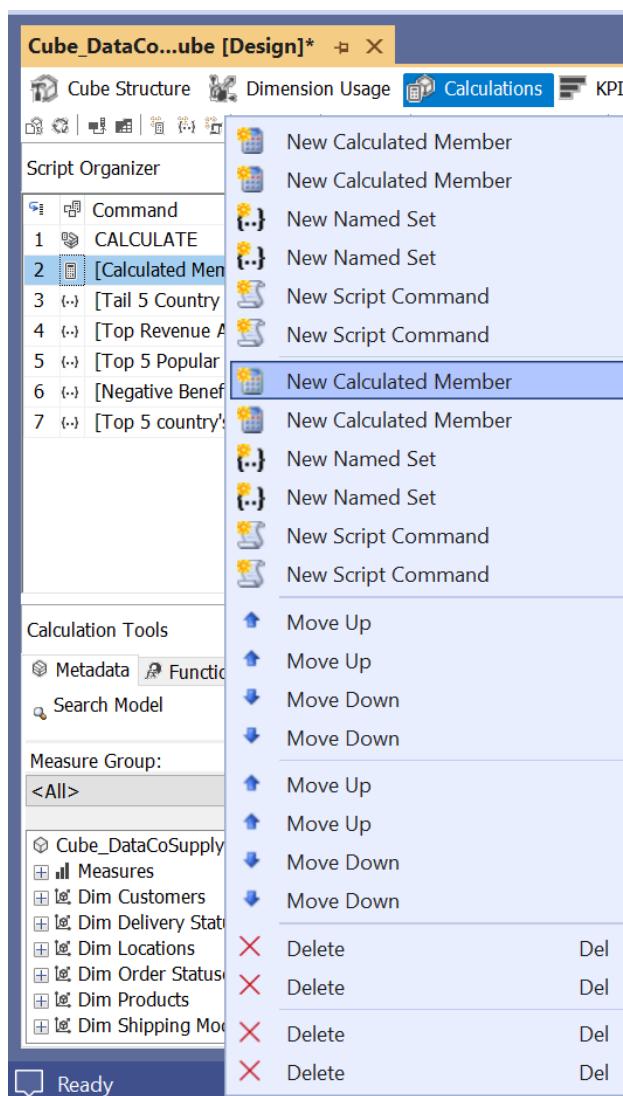
3.7.5. Tạo Calculated Member

Tạo Calculated “Late Risk”

Bước 1: D-click vào cube, chọn Calculations



Bước 2: R-click vào script organization, chọn New Calculator Member



Bước 3: đặt tên cho và nhập MDX script vào expression\

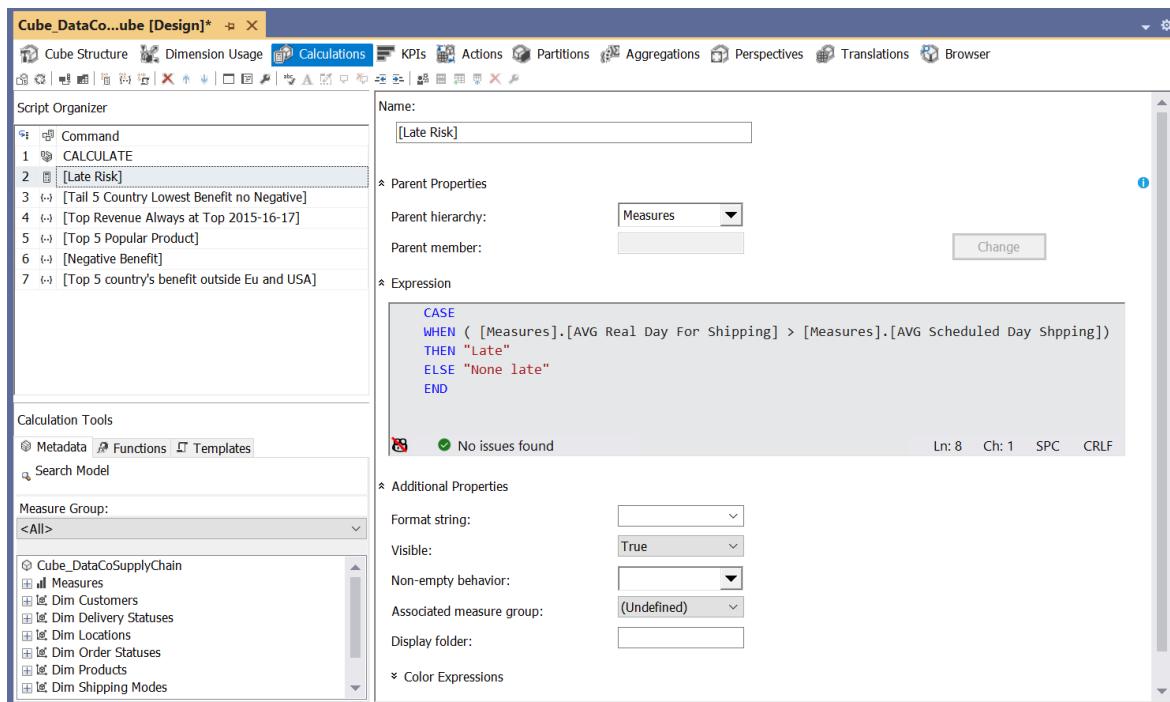
CASE

WHEN ([Measures].[AVG Real Day For Shipping] > [Measures].[AVG Scheduled Day Shpping])

THEN "Late"

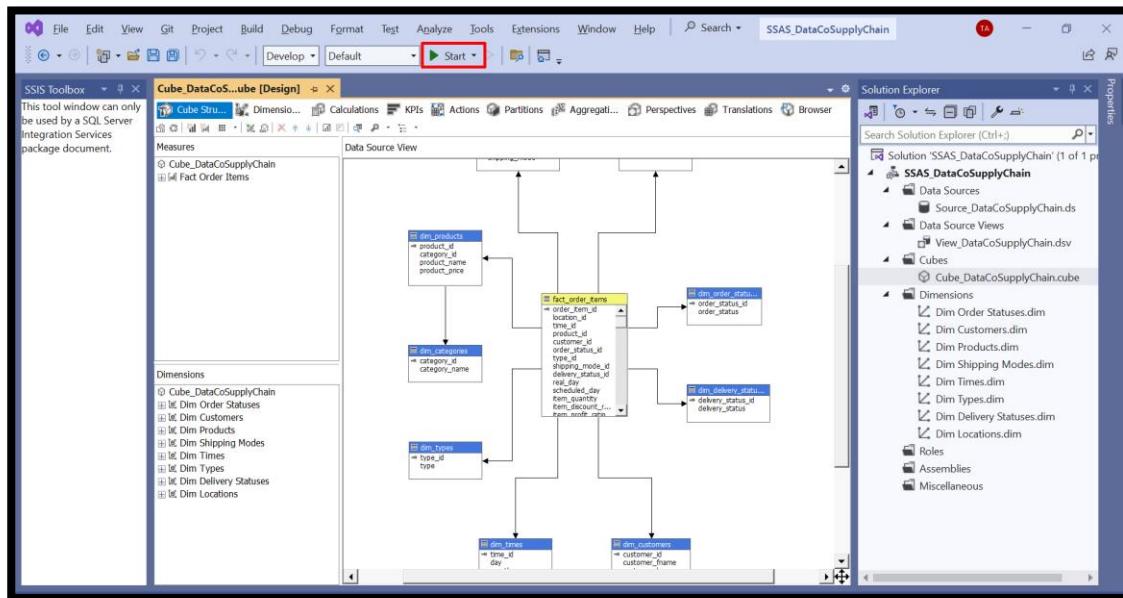
ELSE "None late"

END

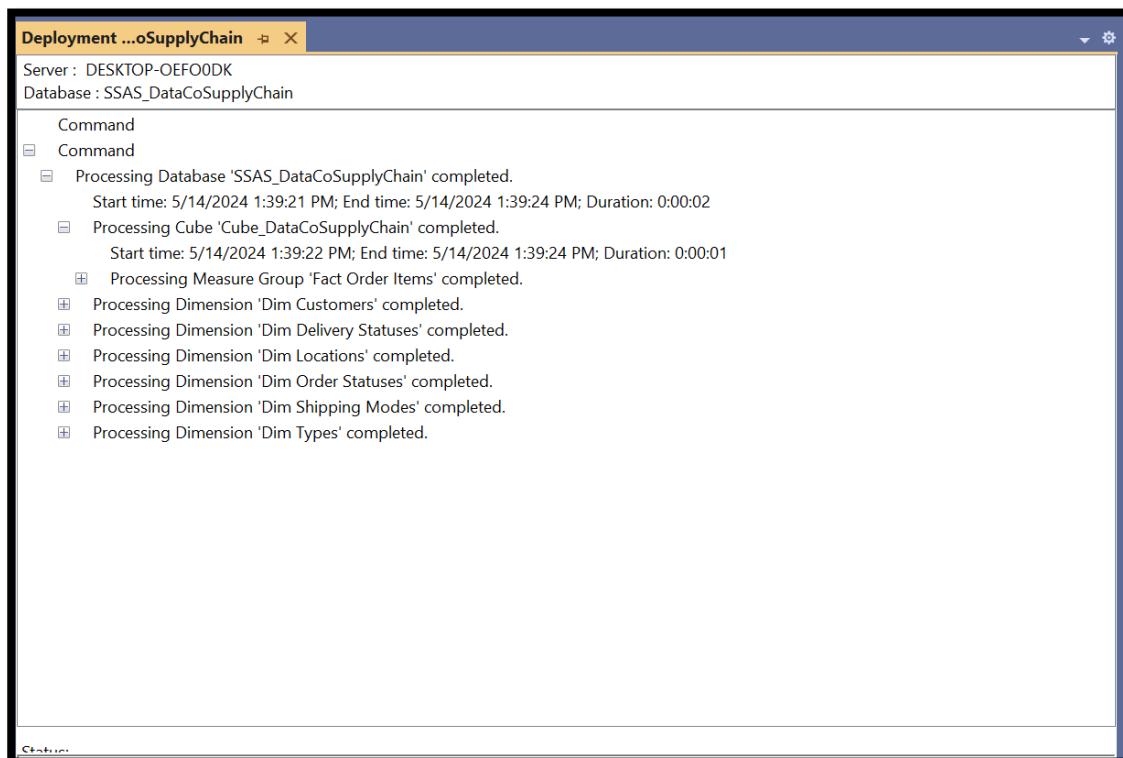


Deploy Process

Tại bước này, ta chính thức deploy project, từ đó tạo ra một data analysis bằng cách chọn start để chạy tất cả các thiết lập, cấu hình trong project bằng cách click start.



Kết quả:



3.8. 15 câu truy vấn

3.8.1. Các truy vấn về Việt Nam

Câu truy vấn 1: Thống kê số lượng item được bán ra và tổng lợi nhuận theo tháng và băn tại Việt Nam.

Ý nghĩa: Biết được tổng quan tình hình các sản phẩm được bán ra, và lợi nhuận thu được theo thời gian (tháng và năm)

Truy vấn MDX

```
-- Vietnam Analysis
-- Query 1: Thong ke so so item duoc ban ra va tong loi nhuan theo thang va nam
SELECT
NON EMPTY { [Measures].[Sum Item Quantity], [Measures].[Sum Benefit Per Order] } ON COLUMNS,
NON EMPTY { [Dim Times].[Year].CHILDREN * [Dim Times].[Month].CHILDREN } ON ROWS
FROM [Cube_DataCoSupplyChain]
WHERE [Dim Locations].[Country].&[Vietnam]&[Pacific Asia];
```

	Sum Item Quantity	Sum Benefit Per Order
2015 10	46	270.49
2015 11	143	920.01
2015 12	198	1982.86
2016 1	418	3468.55
2016 2	329	3345.59
2016 3	236	3755.73
2017 11	14	668.63
2017 12	44	168.62
2018 1	52	1700.15
2018 2	10	168.33

Truy vấn thủ công từ cube

Year	Month	Sum Item Quantity	Sum Benefit Per Order
2015	10	46	270.49
2015	11	143	920.01
2015	12	198	1982.86
2016	1	418	3468.55
2016	2	329	3345.59
2016	3	236	3755.73
2017	11	14	668.63
2017	12	44	168.62
2018	1	52	1700.15
2018	2	10	168.33

Truy vấn từ Excel

Country	Vietnam	Sum Item Quantity	Sum Benefit Per Order
2015			
10	46	270.49	
11	143	920.01	
12	198	1982.86	
2016			
1	418	3468.55	
2	329	3345.59	
3	236	3755.73	
2017			
11	14	668.63	
12	44	168.62	
2018			
1	52	1700.15	
2	10	168.33	
Grand Total	1490	16448.96	

Truy vấn từ Power BI

The screenshot shows the Microsoft Power BI desktop application. The ribbon at the top has tabs: File, Home (selected), Insert, Modeling, View, Optimize, and Help. The Home tab has icons for Paste, Cut, Copy, Format painter, Get data (with sub-options: workbook, data hub, OneLake, SQL Server, Enter data, Dataverse, Recent sources), Transform Refresh data, New visual, Text box, More visuals, New measure, Quick measure, and Calculations.

The main area displays a data grid with the following columns: Country, Sum Order Item Total, and Sum B. The data is hierarchically organized by Country (Vietnam) and Year (2016, 2015, 2017). The total for Vietnam is 149,436.43. The breakdown for 2016 shows three items with values 35,833.25, 25,069.47, and 18,787.64. The breakdown for 2015 shows three items with values 16,773.62, 11,928.94, and 3,709.55. The breakdown for 2017 shows two items with values 25,538.80 and 11,705.16. The total for all years is 149,436.43.

On the right side, there is a 'Filters' pane with sections for 'Filters on this page' and 'Filters on all pages'. Below the filters is a search bar labeled 'Search'.

At the bottom, there are navigation icons for desktop and mobile devices, followed by a row of query buttons labeled Query 1 through Query 8, and a green '+' button.

Country	Sum Order Item Total	Sum B
Vietnam	149,436.43	
2016	79,690.36	
1	35,833.25	
2	25,069.47	
3	18,787.64	
2015	32,412.11	
12	16,773.62	
11	11,928.94	
10	3,709.55	
2017	25,538.80	
2018	11,705.16	
Total	149,436.43	

Câu truy vấn 2: Thống kê doanh thu, số lượng item bán ra, số ngày giao hàng trung bình trên kế hoạch và thực tế tại Việt Nam, Thái Lan và Trung Quốc trong năm 2017

Truy vấn MDX

The screenshot shows the MDXQueryLab application window. The title bar says "project.mdx - DES...P-OEFOODK\anhthy". The main area displays an MDX query:

```
-- Query 2: Tong doanh thu, so san pham ban ra, so ngay giao hang tren
-- ke hoach/thuc te cua Vietnam, Thailand, China trong nam 2017
SELECT
NON EMPTY { [Dim Locations].[Country].&[Vietnam]&[Pacific Asia],
[Dim Locations].[Country].&[Tailandia]&[Pacific Asia],
[Dim Locations].[Country].&[China]&[Pacific Asia]} ON ROWS,
NON EMPTY { [Measures].[Sum Benefit Per Order],
[Measures].[Sum Item Quantity],
[Measures].[AVG Scheduled Day Shipping],
[Measures].[AVG Real Day For Shipping] } ON COLUMNS
FROM [Cube_DataCoSupplyChain]
WHERE [Dim Times].[Year].&[2017];
```

Below the query is a table titled "Messages Results" showing the results:

	Sum Benefit Per Order	Sum Item Quantity	Avg Scheduled Day Shipping	Avg Real Day For Shipping
Vietnam	837.25	58	3.12068965517241	3.79310344827586
Tailandia	5545.5	95	2.92631578947368	3.41052631578947
China	26534.49	625	2.89838709677419	3.46129032258065

At the bottom, a message says "Query executed successfully." and shows the execution time "00:00:01".

Truy vấn thủ công từ cubu

The screenshot shows the SSAS DataCoSupplyChain cube design interface. The title bar says "Cube_DataCoSupplyChain.ube [Design]". The main area has tabs for Cube Structure, Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser.

The browser tab is active, showing a manual MDX query setup:

```
MDX
Edit as Text Import... MDX
```

Below the query editor is a table showing the results:

Country	Sum Benefit Per Order	Sum Item Quantity	Avg Real Day For Shipp...	Avg Scheduled Day Shp...
China	26534.49	625	3.46129032258065	2.89838709677419
Tailandia	5545.5	95	3.41052631578947	2.92631578947368
Vietnam	837.25	58	3.79310344827586	3.12068965517241

Truy vấn từ Excel

PivotTable Fields

Choose fields to add to report:

- Region
- State
- City
- More Fields
 - City
 - Country
 - Location Id

Drag fields between areas below:

Filters	Columns
Year	Σ Values
Rows	Σ Values
Country	Sum Item Quantity Sum Benefit Per Order

Defer Layout Update Update

	Sum Item Quantity	Sum Benefit Per Order	AVG Scheduled Day Shipping	AVG Real Day For Shipping
Year	2017			
China	625	26534.49	2.898387097	3.461290323
Tailandia	95	5545.5	2.926315789	3.410526316
Vietnam	58	837.25	3.120689655	3.793103448
Grand Total	778	32917.24	2.92987013	3.493506494

Truy vấn từ Power BI

Year	2017			
Country	Sum Order Item Total	AVG Order Item Total	AVG Real Day For Shipping	Sum Item Quantity
China	157,511.91	254.05	3.46	625
Tailandia	30,789.67	324.10	3.41	95
Vietnam	25,538.80	440.32	3.79	58
Total	213,840.38	277.71	3.49	778

Câu truy vấn 3: Thống kê KPI doanh thu mua hàng tại Việt Nam trong tháng trong năm có vượt qua mức trung bình chung, sử dụng KPI Growth

Ý nghĩa: kiểm tra liệu Vietnam có vượt qua mức chung

Truy vấn MDX

The screenshot shows the SSMS MDX Query Editor interface. The query window contains the following MDX code:

```
-- Query 3: Vietnam KPI, thong ke tong loi nhuan hang thang cua VN co hon nhan
SELECT
NON EMPTY KPIStatus("Growth KPI") ON COLUMNS,
NON EMPTY {[Dim Times].[Year].CHILDREN
* [Dim Times].[Month].CHILDREN } ON ROWS
FROM [Cube_DataCoSupplyChain]
WHERE [Dim Locations].[Country].&[Vietnam]&[Pacific Asia];
```

The results pane displays a table titled "Growth KPI (Status)" with data from 2015 to 2016. The data is as follows:

Year	Month	Growth KPI Status
2015	4	no data
2015	5	no data
2015	6	no data
2015	7	no data
2015	8	no data
2015	9	no data
2015	10	greater than AVG benefit
2015	11	greater than AVG benefit
2015	12	greater than AVG benefit
2016	1	greater than AVG benefit
2016	2	greater than AVG benefit
2016	3	greater than AVG benefit
2016	4	no data
2016	5	no data
2016	6	no data
2016	7	no data
2016	8	no data

At the bottom left, a green checkmark indicates "Query executed successfully." The bottom right shows the session details: DESKTOP-OEFOODK, DESKTOP-OEFOODK\anhthy, SSAS_DataCoSupplyChain, and 00:00:01.

Truy vấn từ Excel

The screenshot shows the SSAS Cube Designer in Design mode. The top navigation bar includes options like Cube Structure, Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser. The main area displays the cube structure for "Cube_DataCoSupplyChain".

A filter expression is being applied to the "Growth KPI" dimension under the "Dim Locations" hierarchy, set to "Equal" with the value "{ Vietnam }".

The data grid below shows the results of the filtered query:

Year	Month	Growth KPI Status
2015	1	no data
2015	2	no data
2015	3	no data
2015	4	no data
2015	5	no data
2015	6	no data
2015	7	no data
2015	8	no data
2015	9	no data
2015	10	greater than AVG ...
2015	11	greater than AVG ...
2015	12	greater than AVG ...
2016	1	greater than AVG ...
2016	2	greater than AVG ...
2016	3	greater than AVG ...
2016	4	no data
2016	5	no data
2016	6	no data
2016	7	no data

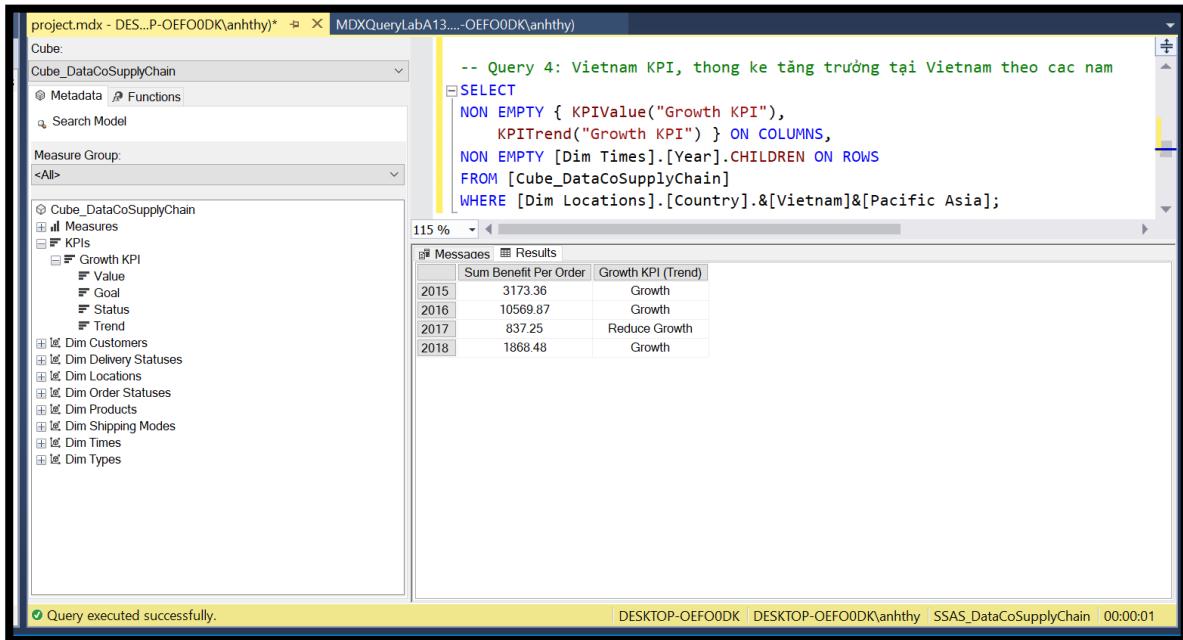
Pivot excel:

Truy vấn từ Power BI

	Country	Year	Sum Item Quantity	Benefit KPI Status
2015	1	387	no data	
	2	10	no data	
	3	11	no data	
	4	12	no data	
	5	104	no data	
	6	143	no data	
	7	198	no data	
	8	46	greater than AVG benefit	
	9	198	greater than AVG benefit	
	10	1490	greater than AVG benefit	
	Total	983	greater than AVG benefit	
		1490	greater than AVG benefit	

Câu truy vấn 4: Thống kê KPI tăng trưởng doanh thu mua hàng tại Việt Nam trong năm.

Truy vấn MDX



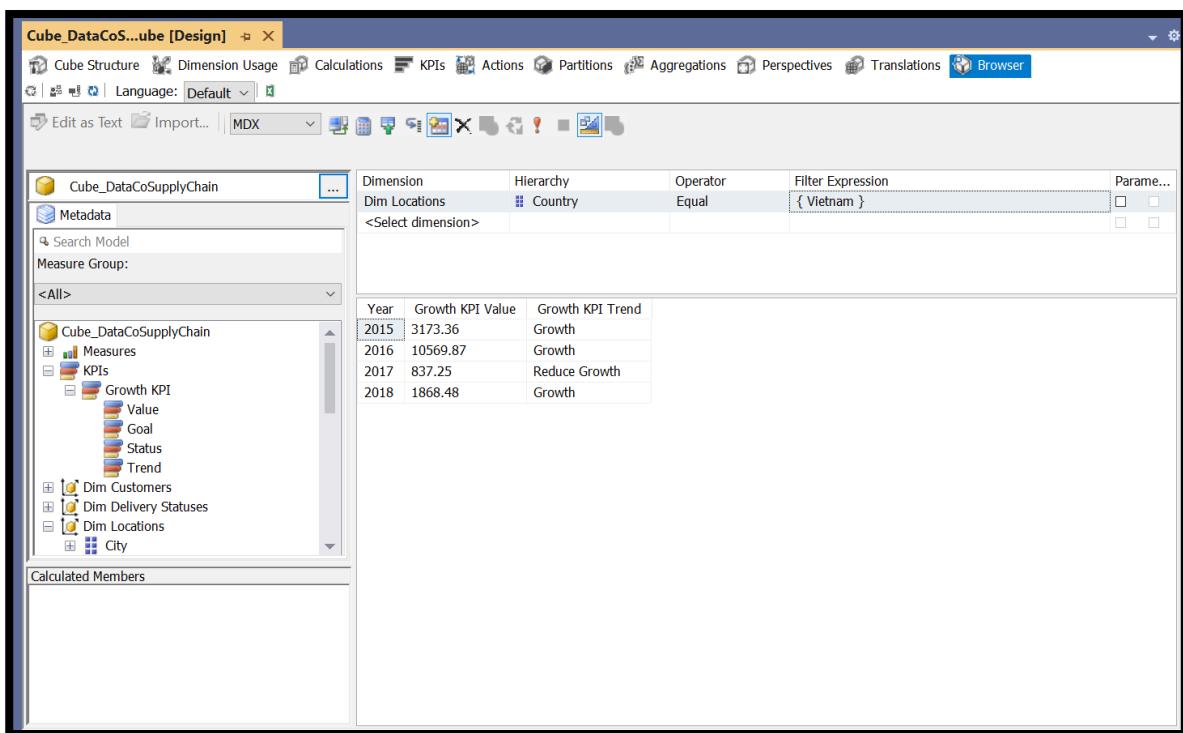
project.mdx - DES...P-OEFO0DK\anhthy* X MDXQueryLabA13....OEFO0DK\anhthy

```
-- Query 4: Vietnam KPI, thong ke tang truong tai Vietnam theo cac nam
SELECT
NON EMPTY { KPITValue("Growth KPI"),
KPITrend("Growth KPI") } ON COLUMNS,
NON EMPTY [Dim Times].[Year].CHILDREN ON ROWS
FROM [Cube_DataCoSupplyChain]
WHERE [Dim Locations].[Country].&[Vietnam]&[Pacific Asia];
```

Year	Sum Benefit Per Order	Growth KPI (Trend)
2015	3173.36	Growth
2016	10569.87	Growth
2017	837.25	Reduce Growth
2018	1868.48	Growth

Query executed successfully. DESKTOP-OEFO0DK DESKTOP-OEFO0DK\anhthy SSAS_DataCoSupplyChain 00:00:01

Truy vấn từ Excel



Cube_DataCoSupplyChain.ube [Design]

Dimension Hierarchy Operator Filter Expression
Dim Locations Country Equal { Vietnam }

Year	Growth KPI Value	Growth KPI Trend
2015	3173.36	Growth
2016	10569.87	Growth
2017	837.25	Reduce Growth
2018	1868.48	Growth

Pivot Excel

A1 File Home Insert Page Layout Formulas Data Review View Automate Help Power Pivot PivotTable Analyze Design Comments Share

Row Labels Growth KPI Growth KPI Trend

	Country	Vietnam
3	4	3173.36 Growth
4	2015	10569.87 Growth
5	2016	837.25 Reduce Growth
6	2017	1868.48 Growth
7	2018	
8	Grand Total	16448.98 Growth

PivotTable Fields

Choose fields to add to report:

Search

- Hour
- Minute
- Month
- Time Id
- Year

Drag fields between areas below:

Filters Columns

Country Values

Rows Values

Year Growth KPI Growth KPI Trend

Defer Layout Update Update

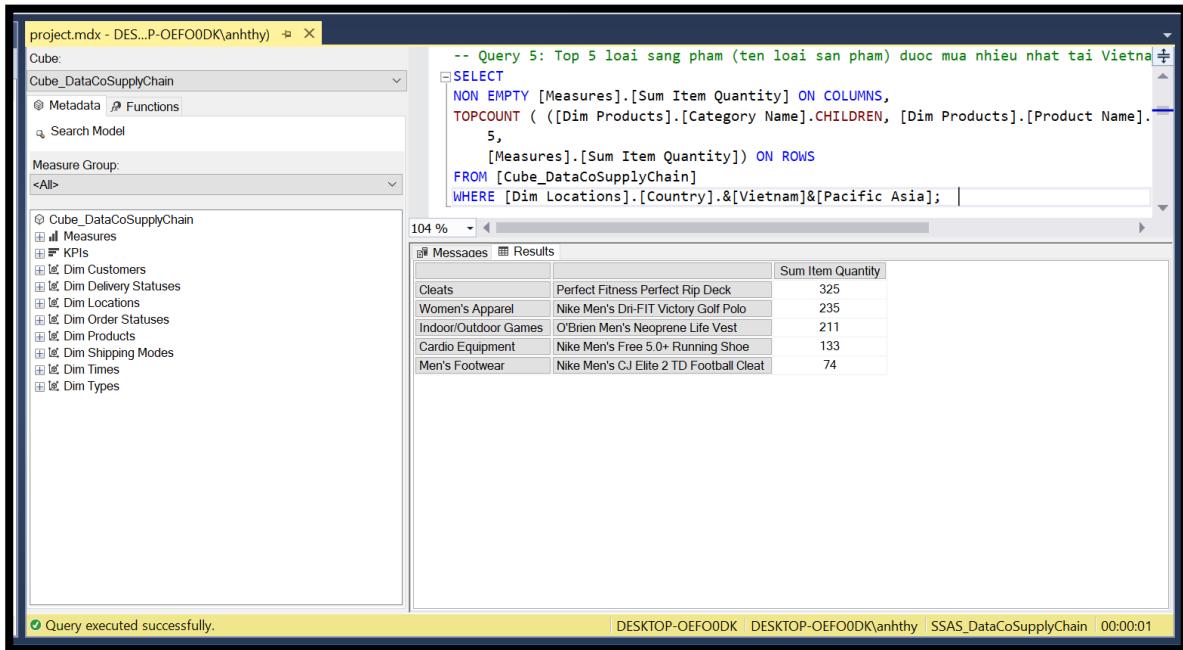
Truy vấn từ Power BI

Year	Vietnam	Total
	no data	no data
2015	Growth	Growth
2016	Growth	Growth
2017	Reduce Growth	Reduce Growth
2018	Growth	Growth
Total	Growth	Growth

Query 1 Query 2 Query 3 Query 4 Query 5 Query 6 Query 7

Câu truy vấn 5: Thống kê top 5 sản phẩm (cùng với tên loại sản phẩm) được mua nhiều nhất tại Việt Nam

Truy vấn MDX



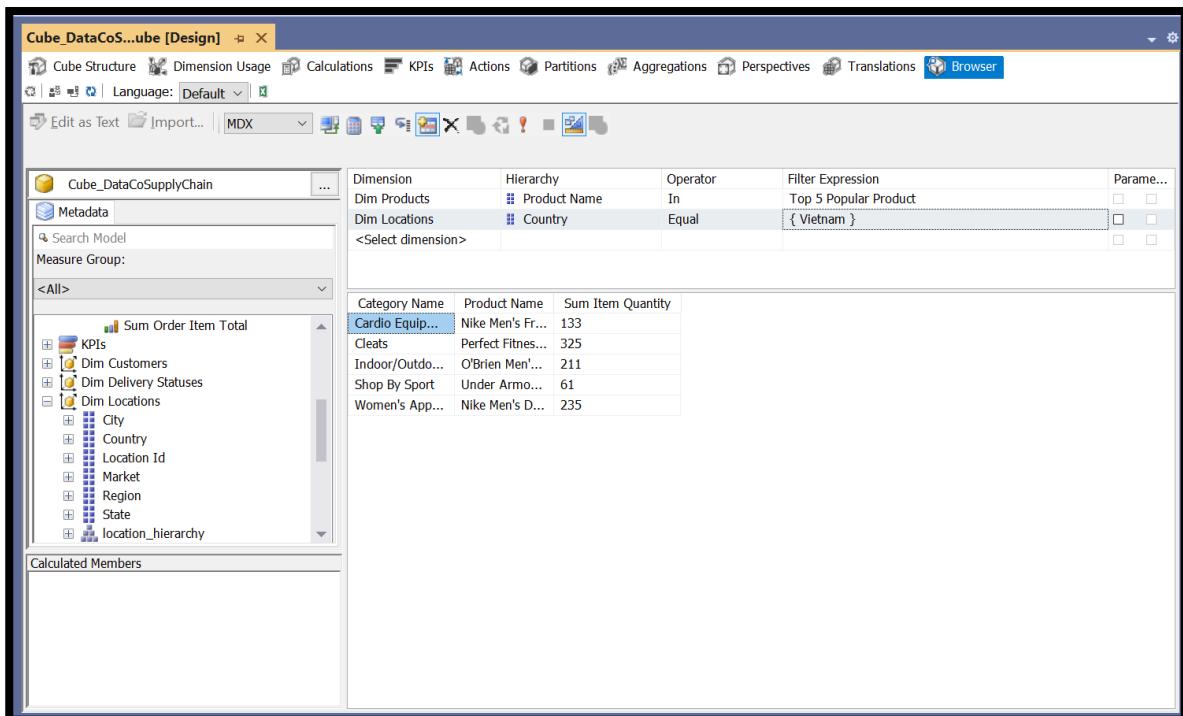
The screenshot shows the SSMS interface with the following details:

- Title Bar:** project.mdx - DES...P-OEFO0DK\anhthy
- Left pane:** Cube browser for Cube_DataCoSupplyChain. It shows Measure Group <All> and various dimensions like Dim Products, Dim Locations, etc.
- Center pane:** An MDX query window containing the following code:


```
-- Query 5: Top 5 loai sang pham (ten loai san pham) duoc mua nhieu nhat tai Viet Nam
SELECT
NON EMPTY [Measures].[Sum Item Quantity] ON COLUMNS,
TOPCOUNT ( ([Dim Products].[Category Name].CHILDREN, [Dim Products].[Product Name]),
5,
[Measures].[Sum Item Quantity]) ON ROWS
FROM [Cube_DataCoSupplyChain]
WHERE [Dim Locations].[Country].&[Vietnam]&[Pacific Asia];
```
- Right pane:** Results grid showing the top 5 products by quantity:

		Sum Item Quantity
Cleats	Perfect Fitness Perfect Rip Deck	325
Women's Apparel	Nike Men's Dri-FIT Victory Golf Polo	235
Indoor/Outdoor Games	O'Brien Men's Neoprene Life Vest	211
Cardio Equipment	Nike Men's Free 5.0+ Running Shoe	133
Men's Footwear	Nike Men's CJ Elite 2 TD Football Cleat	74
- Status Bar:** Query executed successfully. DESKTOP-OEFO0DK DESKTOP-OEFO0DK\anhthy SSAS_DataCoSupplyChain 00:00:01

Truy vấn thủ công từ cube



The screenshot shows the BIDS interface with the following details:

- Title Bar:** Cube_DataCoS...ube [Design]
- Toolbar:** Includes Cube Structure, Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser.
- Left pane:** Cube browser for Cube_DataCoSupplyChain. It shows Measure Group <All> and various dimensions like Dim Products, Dim Locations, etc.
- Center pane:**
 - Top section:** A table for defining dimensions and filters:

Dimension	Hierarchy	Operator	Filter Expression	Parameter
Dim Products	# Product Name	In	Top 5 Popular Product	<input type="checkbox"/>
Dim Locations	# Country	Equal	{ Vietnam }	<input type="checkbox"/>
 - Bottom section:** A results grid showing the top 5 products by quantity:

Category Name	Product Name	Sum Item Quantity
Cardio Equip...	Nike Men's Fr...	133
Cleats	Perfect Fitnes...	325
Indoor/Outdo...	O'Brien Men'...	211
Shop By Sport	Under Armo...	61
Women's App...	Nike Men's D...	235

Truy vấn từ Excel

A1 Country
1 Country Vietnam
2
3 Row Labels Sum Item Quantity
4 Nike Men's CJ Elite 2 TD Football Cleat
5 Men's Footwear 74
6 Nike Men's Dri-FIT Victory Golf Polo
7 Women's Apparel 235
8 Nike Men's Free 5.0+ Running Shoe
9 Cardio Equipment 133
10 O'Brien Men's Neoprene Life Vest
11 Indoor/Outdoor Games 211
12 Perfect Fitness Perfect Rip Deck
13 Cleats 325
14
15
16
17
18
19
20
21
22
23
24
25
26
27

Ready Accessibility: Good to go

PivotTable Fields:
Choose fields to add to report:
Search
Country
Location Id
Market
Region
State

Drag fields between areas below:
Filters: Country
Columns: Country
Rows: Top 5 Popular Product
Values: Sum Item Quantity
Category Name
Defer Layout Update
Update 100%

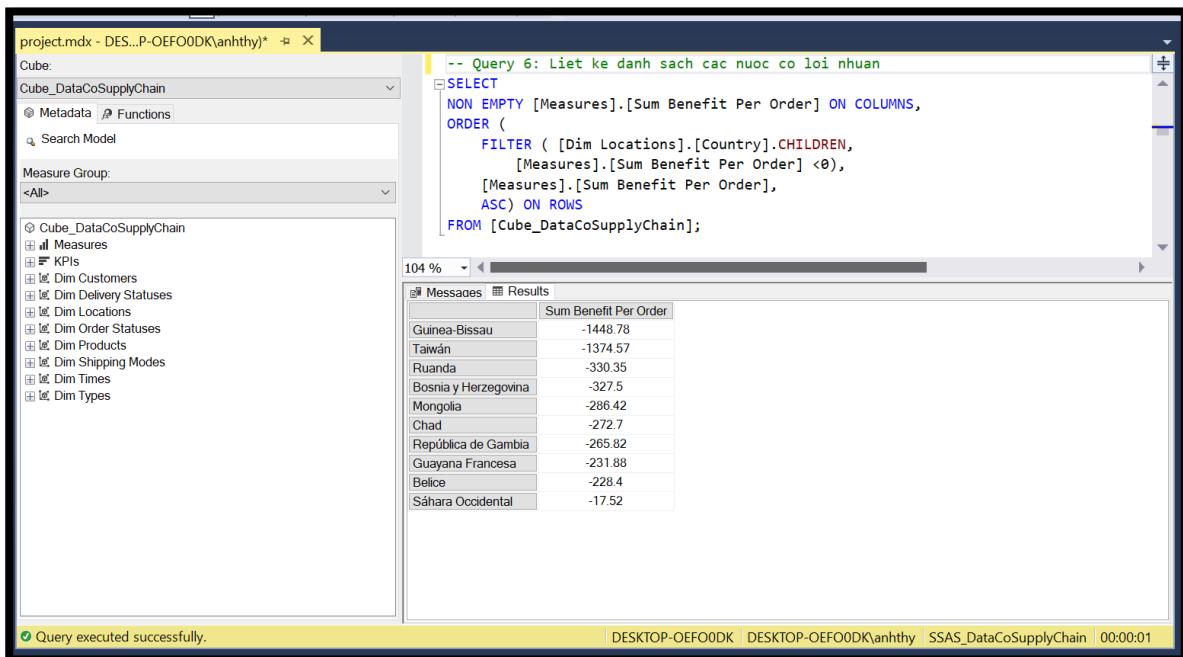
Truy vấn từ Power BI

Category Name	Sum Item Quantity
Accessories	24
Baby	2
Baby sweater	2
Baseball & Softball	8
adidas Kids' F5 Messi FG Soccer Cleat	1
adidas Men's F10 Messi TRX FG Soccer Cleat	7
Camping & Hiking	44
Cardio Equipment	139
CDs	2
Children's Clothing	7
Cleats	325
Total	1490

3.8.2. Các truy vấn về sales

Câu truy vấn 6: Liệt kê danh sách các nước có lợi nhuận, sắp xếp từ nhỏ đến lớn nhất

Truy vấn MDX

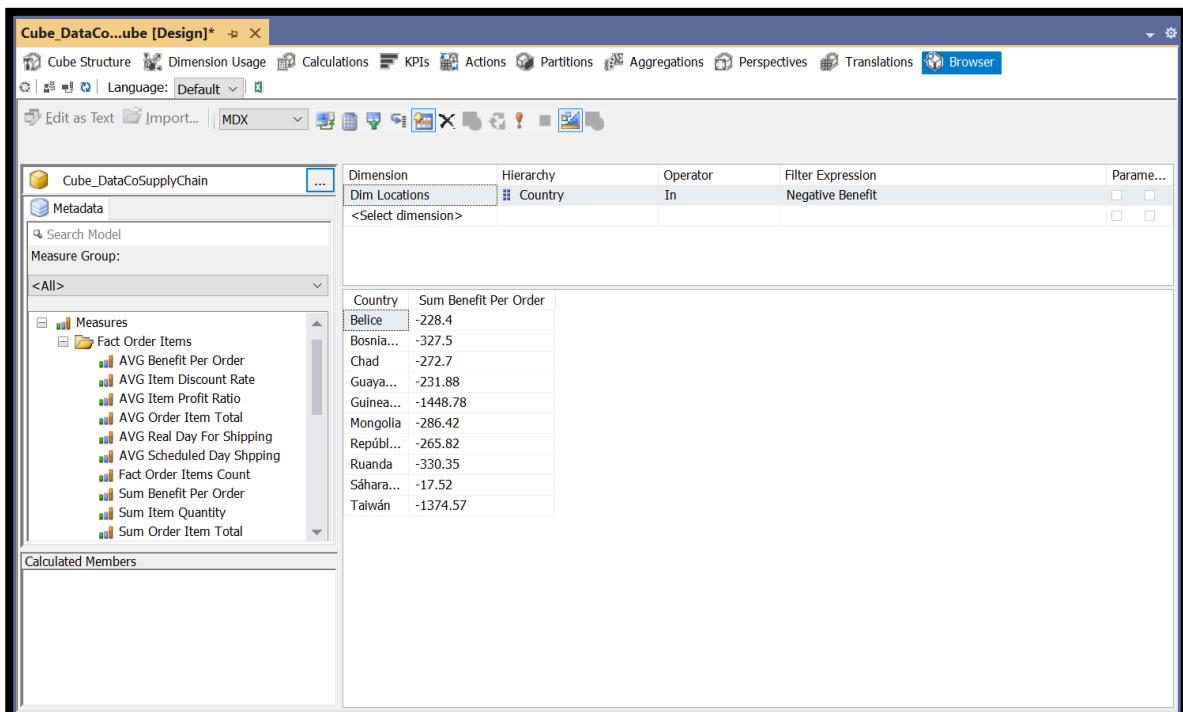


```
-- Query 6: Liet ke danh sach cac nuoc co loi nhuuan
SELECT
NON EMPTY [Measures].[Sum Benefit Per Order] ON COLUMNS,
ORDER (
    FILTER ( [Dim Locations].[Country].CHILDREN,
        [Measures].[Sum Benefit Per Order] < 0),
    [Measures].[Sum Benefit Per Order],
    ASC) ON ROWS
FROM [Cube_DataCoSupplyChain];
```

Country	Sum Benefit Per Order
Guinea-Bissau	-1448.78
Taiwan	-1374.57
Ruanda	-330.35
Bosnia y Herzegovina	-327.5
Mongolia	-286.42
Chad	-272.7
República de Gambia	-265.82
Guayana Francesa	-231.88
Belice	-228.4
Sáhara Occidental	-17.52

Query executed successfully.

Truy vấn thủ công từ cube



The screenshot shows the SSAS cube design interface for the 'Cube_DataCoSupplyChain'. The top navigation bar includes options like Cube Structure, Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser. The main workspace displays the cube's structure. On the left, the 'Measures' group under 'Fact Order Items' is expanded, showing various measures such as AVG Benefit Per Order, AVG Item Discount Rate, AVG Item Profit Ratio, AVG Order Item Total, AVG Real Day For Shipping, AVG Scheduled Day For Shipping, Fact Order Items Count, Sum Benefit Per Order, Sum Item Quantity, and Sum Order Item Total. On the right, a 'Dimension' pane shows a filter for 'Dim Locations' with 'Country' selected, 'Operator' set to 'In', and 'Filter Expression' set to 'Negative Benefit'. Below this, a table lists countries and their sum benefit per order:

Country	Sum Benefit Per Order
Belice	-228.4
Bosnia...	-327.5
Chad	-272.7
Guaya...	-231.88
Guinea...	-1448.78
Mongolia	-286.42
Repúbl...	-265.82
Ruanda	-330.35
Sáhara...	-17.52
Taiwan	-1374.57

Truy vấn từ Excel

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable Fields pane on the right side is configured with the following settings:

- Choose fields to add to report:** Country, Location Id, Market, Region, State.
- Sets:** Negative Benefit (selected).
- Drag fields between areas below:** Filters, Columns, Rows, Values.
- Rows:** Negative Benefit.
- Values:** Sum Benefit Per Order.

The data grid displays the following information:

	Row Labels	Sum Benefit Per Order
1	Row Labels	
2	Belice	-228.4
3	Bosnia y Herzegovina	-327.5
4	Chad	-272.7
5	Guayana Francesa	-231.88
6	Guinea-Bissau	-1448.78
7	Mongolia	-286.42
8	República de Gambia	-265.82
9	Ruanda	-330.35
10	Sáhara Occidental	-17.52
11	Taiwán	-1374.57

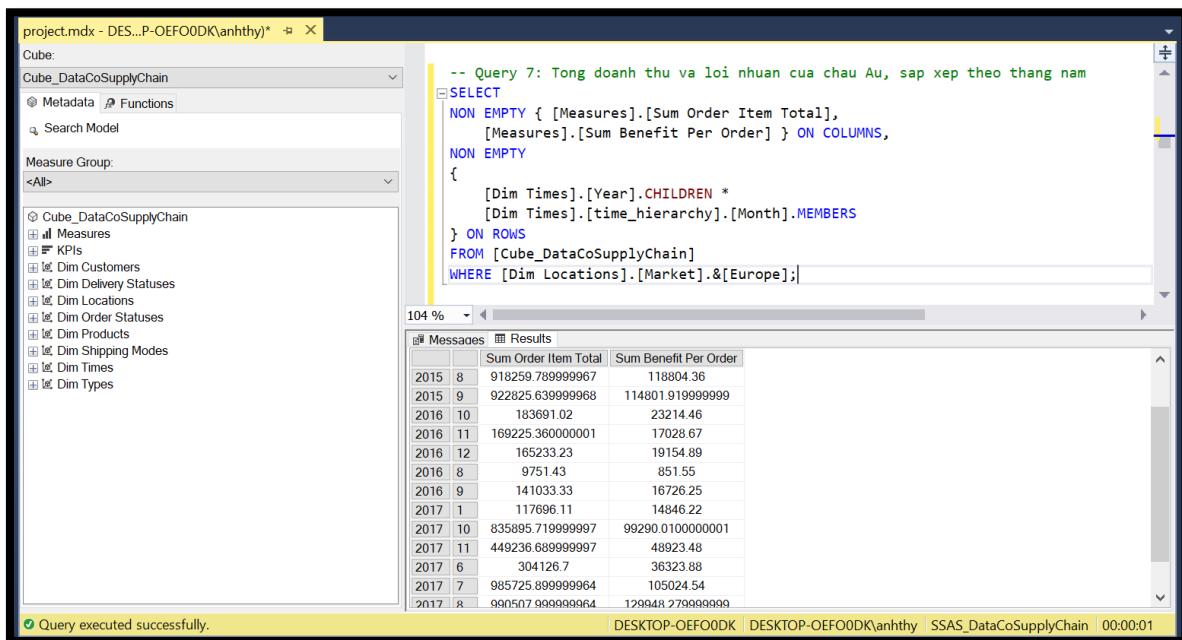
Truy vấn từ Power BI

The screenshot shows a Power BI table visualization with the following data:

Country	Sum Benefit Per Order
Sáhara Occidental	-17.52
Belice	-228.40
Guayana Francesa	-231.88
República de Gambia	-265.82
Chad	-272.70
Mongolia	-286.42
Bosnia y Herzegovina	-327.50
Ruanda	-330.35
Taiwán	-1,374.57
Guinea-Bissau	-1,448.78
Total	-4,783.94

Câu truy vấn 7: Tổng doanh thu và lợi nhuận của châu Âu, sắp xếp theo tháng năm

Truy vấn MDX:



The screenshot shows the SSMS interface with an MDX query window. The query is:

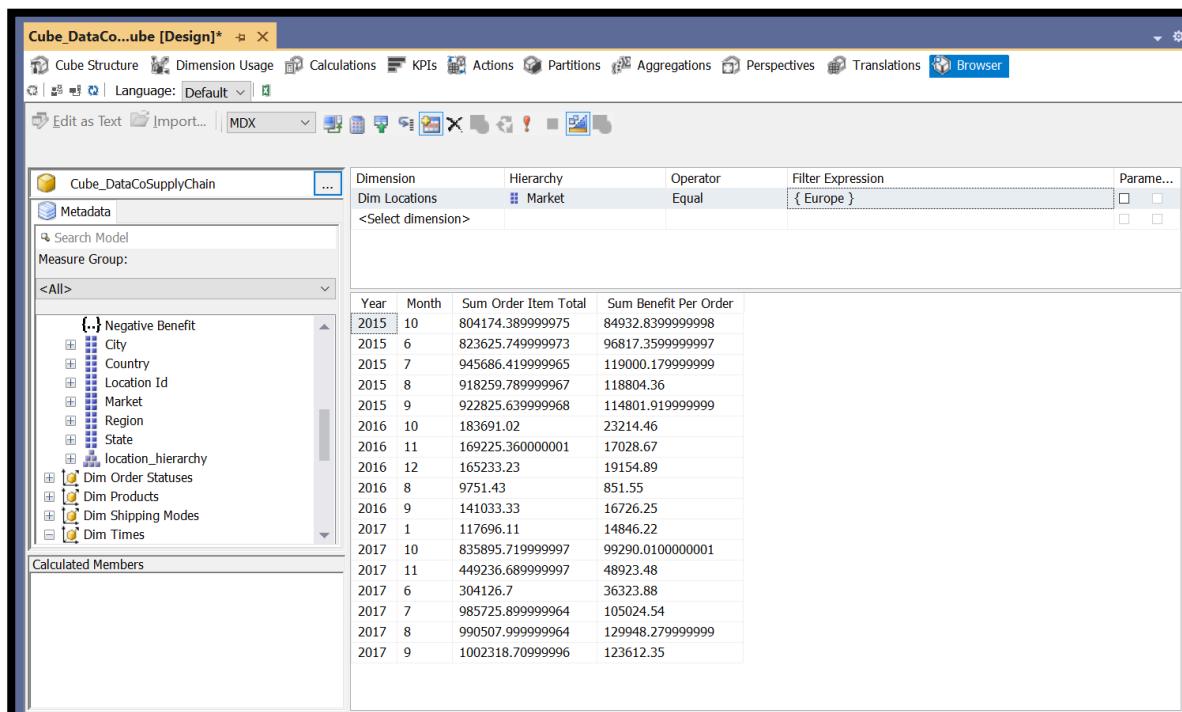
```
-- Query 7: Tong doanh thu va loi nhuan cua chau Au, sap xep theo thang nam
SELECT
NON EMPTY { [Measures].[Sum Order Item Total],
[Measures].[Sum Benefit Per Order] } ON COLUMNS,
NON EMPTY
{
    [Dim Times].[Year].CHILDREN *
    [Dim Times].[time_hierarchy].[Month].MEMBERS
} ON ROWS
FROM [Cube_DataCoSupplyChain]
WHERE [Dim Locations].[Market].&[Europe];
```

The results grid displays the following data:

	Sum Order Item Total	Sum Benefit Per Order
2015 8	918259.789999967	118804.36
2015 9	922825.639999968	114801.919999999
2016 10	183691.02	23214.46
2016 11	169225.360000001	17028.67
2016 12	165233.23	19154.89
2016 8	9751.43	851.55
2016 9	141033.33	16726.25
2017 1	117696.11	14846.22
2017 10	835895.719999997	99290.0100000001
2017 11	449236.689999997	48923.48
2017 6	304126.7	36323.88
2017 7	985725.899999964	105024.54
2017 8	990507.999999964	129948.279999999

Query executed successfully.

Truy vấn thủ công từ cube



The screenshot shows the BIDS interface with the 'Cube_Design' view open. The query is:

```
SELECT
NON EMPTY { [Measures].[Sum Order Item Total],
[Measures].[Sum Benefit Per Order] } ON COLUMNS,
NON EMPTY
{
    [Dim Times].[Year].CHILDREN *
    [Dim Times].[time_hierarchy].[Month].MEMBERS
} ON ROWS
FROM [Cube_DataCoSupplyChain]
WHERE [Dim Locations].[Market].&[Europe];
```

The results grid displays the same data as the previous screenshot:

	Sum Order Item Total	Sum Benefit Per Order
2015 8	918259.789999967	118804.36
2015 9	922825.639999968	114801.919999999
2016 10	183691.02	23214.46
2016 11	169225.360000001	17028.67
2016 12	165233.23	19154.89
2016 8	9751.43	851.55
2016 9	141033.33	16726.25
2017 1	117696.11	14846.22
2017 10	835895.719999997	99290.0100000001
2017 11	449236.689999997	48923.48
2017 6	304126.7	36323.88
2017 7	985725.899999964	105024.54
2017 8	990507.999999964	129948.279999999

Truy vấn từ Excel

PivotTable Fields

- Choose fields to add to report:
- Search: location_hierarchy
- location_hierarchy
 - Market
 - Country
 - Region
 - State
 - City
 - > More Fields

Drag fields between areas below:

Filters	Columns
location_hierarchy	Σ Values
Rows	Σ Values
time_hierarchy	Sum Benefit Per Order
	Sum Order Item Total

Defer Layout Update Update

Truy vấn từ Power BI

Year	Sum Order Item Total	Sum Benefit Per Order
2015	4,414,571.99	534,356.66
6	823,625.75	96,817.36
7	945,686.42	119,000.18
8	918,259.79	118,804.36
9	922,825.64	114,801.92
10	804,174.39	84,932.84
2016	668,934.37	76,975.82
2017	4,685,507.83	557,968.76
Total	9,769,014.19	1,169,301.24

Câu truy vấn 8: Liệt kê danh sách các nước không phải Mỹ đem lại nhiều lợi nhuận nhất

Truy vấn MDX

The screenshot shows the SSMS interface with the following details:

- Title Bar:** project.mdx - DES...P-OEFOOD\anhthy*
- Left pane:** Cube: Cube OLAP Database Supply Chain. It shows the Metadata section and a Measure Group dropdown set to <All>. Below it is a tree view of countries.
- Middle pane:** The MDX query window contains two queries:
 - Query 8: Liet ke 5 nuoc khong phai My dem lai nhieu loi nhuuan nhat
 - Query 9: Liet ke 5 nuoc co doanh thu thap nhat (khong am doanh thu)
- Right pane:** Results grid showing the top 5 countries by benefit:

Country	Sum Benefit Per Order
Francia	327790.059999998
México	303240.10999997
Alemania	194801.3
Brasil	186691.37
Reino Unido	180922.72
- Bottom status bar:** Query executed successfully. DESKTOP-OEFOODK DESKTOP-OEFOOD\anhthy OLAP_SSAS_SupplyChain 00:00:01

Truy vấn thủ công từ cube

The screenshot shows the Cube OLAP Designer interface with the following details:

- Title Bar:** Cube OLAP D...ube [Design]
- Toolbar:** Cube Str..., Dimension, Calculations, KPIs, Actions, Partitions, Aggregati..., Perspective..., Translati..., Browser.
- Language:** Default
- Left pane:** Cube OLAP Database Supply Chain. It shows the Metadata section and a Measure Group dropdown set to <All>. Below it is a tree view of measures.
- Middle pane:** A query editor window with the following content:


```
SELECT [Measures].[Sum Benefit Per Order] ON COLUMNS,
TOPCOUNT (
    ORDER (
        EXCEPT(
            [Dim Locations].[Country].CHILDREN,
            [Dim Locations].[Country].&[Estados Unidos]&[USCA]
        ),
        [Measures].[Sum Benefit Per Order],
        ASC
    ),
    5,
    [Measures].[Sum Benefit Per Order]
) ON ROWS
FROM [Cube OLAP Database Supply Chain];
```
- Right pane:** Results grid showing the top 5 countries by benefit:

Country	Sum Benefit Per ...
Alemania	194801.3
Brasil	186691.37
Francia	327790.05999999...
México	303240.110000...
Reino Unido	180922.72

Truy vấn từ Excel

A3 Row Labels Sum Benefit Per Order

	Row Labels	Sum Benefit Per Order
4	Alemania	194801.3
5	Brasil	186691.37
6	Francia	327790.06
7	México	303240.11
8	Reino Unido	180922.72

Truy vấn từ Power BI

Country	Sum Benefit Per Order
Francia	327,790.06
México	303,240.11
Alemania	194,801.30
Brasil	186,691.37
Reino Unido	180,922.72
Total	1,193,445.56

Câu truy vấn 9: Danh sách 5 nước có doanh thu thấp nhất (không âm)

Truy vấn MDX

The screenshot shows the SSMS interface with the following details:

- Title Bar:** project.mdx - DES...P-OEFOODK\anhthy*
- Left pane:** Cube browser for "Cube OLAP Database Supply Chain". It shows hierarchies for Country, Location Id, Market, Region, State, and City.
- Middle pane:** MDX query window containing the following code:


```
-- Query 9: Liet ke 5 nuoc co doanh thu thap nhat (khong am doanh thu)
SELECT {[Measures].[Sum Benefit Per Order]} ON COLUMNS,
TAIL(
  ORDER(
    FILTER(
      [Dim Locations].[Country].CHILDREN,
      [Measures].[Sum Benefit Per Order] > 0
    ),
    [Measures].[Sum Benefit Per Order],
    DESC
  ),
  5
) ON ROWS
FROM [Cube OLAP Database Supply Chain];
```
- Right pane:** Results grid showing the top 5 countries with the lowest benefit per order:

Country	Sum Benefit Per Order
Burundi	77.75
Estonia	60.61999999999999
Armenia	42.44
Serbia	42
Eslovenia	23
- Status Bar:** Query executed successfully. DESKTOP-OEFOODK | DESKTOP-OEFOODK\anhthy | OLAP_SSAS_SupplyChain | 00:00:01

Truy vấn thủ công từ cube

The screenshot shows the Cube Designer interface with the following details:

- Title Bar:** Cube OLAP Database [Design]
- Toolbar:** Cube Str..., Dimension..., Calculations, KPIs, Actions, Partitions, Aggregat..., Perspective..., Translation..., Browser
- Language:** Default
- Left pane:** Cube browser for "Cube OLAP Database Supply Chain". It shows Measure Groups like Metadata, AVG Item Profit Ratio, AVG Order Item Total, etc.
- Middle pane:** A table showing the results of a query:

Dimension	Hierarchy	Operator	Filter Expression
Dim Locations	Country	In	Top Lowest Benefit not Negative
- Right pane:** Results grid showing the top 5 countries with the lowest benefit per order:

Country	Sum Benefit Per ...
Armenia	42.44
Burundi	77.75
Eslovenia	23
Estonia	60.6199999999...
Serbia	42

Truy vấn từ Excel

A1 Row Labels Sum Benefit Per Order

	Row Labels	Sum Benefit Per Order
1	Alemania	194801.3
2	Brasil	186691.37
3	Francia	327790.06
4	México	303240.11
5	Reino Unido	180922.72
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		

Truy vấn từ Power BI

Country	Sum Benefit Per Order
Armenia	42.44
Burundi	77.75
Eslovenia	23.00
Estonia	60.62
Serbia	42.00
Total	245.81

Câu truy vấn 10: Danh sách các sản phẩm luôn nằm trong top 10 doanh thu trong các năm 2015-16-17

Truy vấn MDX

The screenshot shows the SSMS interface with the following details:

- Title Bar:** project.mdx - DES...P-OEFOODK\anhthy
- Left Panel:** Shows the cube structure for "Cube OLAP Database Supply Chain" under "Measure Group: <All>".
- Center Panel:** Displays an MDX query result set titled "Messages" with the following data:

	Sum Order Item Total
Field & Stream Sportsman 16 Gun Fire Safe	6226925.49999974
Perfect Fitness Perfect Rip Deck	3973106.8099999
Diamondback Women's Serene Classic Comfo...	3700745.3899998
Nike Men's Free 5.0+ Running Shoe	3295613.87999979
Nike Men's Dri-FIT Victory Golf Polo	2828708.5
Pelican Sunstream 100 Kayak	2785517.9999989
O'Brien Men's Neoprene Life Vest	2596314.53999984
Nike Men's CJ Elite 2 TD Football Cleat	2598333.87999988
Under Armour Girls' Toddler Spine Surge Runni	1140723.56999998

- Bottom Status Bar:** Query executed successfully. DESKTOP-OEFOODK\DESKTOP-OEFOODK\anhthy OLAP_SSAS_SupplyChain | 00:00:01

Truy vấn thủ công từ cube

The screenshot shows the SSDT interface with the following details:

- Title Bar:** Cube OLAP Database Supply Chain [Design]
- Toolbar:** Includes icons for Cube Structure, Dimension, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translation, and Browser.
- Language:** Default
- Query Editor:** Shows an MDX query and its results.
- Results Grid:** Displays the same data as the previous screenshot, listing products and their sum order item total.

Truy vấn Excel

The screenshot shows a Microsoft Excel-like interface with a PivotTable Fields pane on the right and a data grid on the left.

PivotTable Fields:

- Choose fields to add to report:
- Search:
- Sum Item Quantity:
- Sum Order Item Total:**
- KPIs:
- Other Customizations:

Drag fields between areas below:

- Filters
- Columns
- Rows: Products Always in To...
- Values

Data Grid (A1 to I22):

Row Labels	Sum Order Item Total
1 Diamondback Women's Serene Classic Comfort Bi	3700745.39
2 Field & Stream Sportsman 16 Gun Fire Safe	6226925.5
3 Nike Men's CJ Elite 2 TD Football Cleat	2598333.88
4 Nike Men's Dri-FIT Victory Golf Polo	2828708.5
5 Nike Men's Free 5.0+ Running Shoe	3295613.88
6 O'Brien Men's Neoprene Life Vest	2596314.54
7 Pelican Sunstream 100 Kayak	2785518
8 Perfect Fitness Perfect Rip Deck	3973106.81
9 Under Armour Girls' Toddler Spine Surge Runni	1140723.57

Bottom navigation bar: Query 4, Query 5, Query 6, Query 7, Query 8, Query 9, Qu ..., +, ...

Truy vấn từ Power BI

The screenshot shows a Power BI report with a table visualization.

Table Headers:

- Product Name
- Sum Order Item Total

Data Rows:

Diamondback Women's Serene Classic Comfort Bi	3,700,745.39
Field & Stream Sportsman 16 Gun Fire Safe	6,226,925.50
Nike Men's CJ Elite 2 TD Football Cleat	2,598,333.88
Nike Men's Dri-FIT Victory Golf Polo	2,828,708.50
Nike Men's Free 5.0+ Running Shoe	3,295,613.88
O'Brien Men's Neoprene Life Vest	2,596,314.54
Pelican Sunstream 100 Kayak	2,785,518.00
Perfect Fitness Perfect Rip Deck	3,973,106.81
Under Armour Girls' Toddler Spine Surge Runni	1,140,723.57
Total	29,145,990.07

Bottom right icons: Filter, Refresh, More

Câu truy vấn 11: Tình trạng giao hàng tại châu Á – Thái Bình Dương

Truy vấn MDX

The screenshot shows the SSMS interface with the following details:

- Title Bar:** project.mdx - DES...P-OEFOODK\anhthy
- Left Panel:** Cube: Cube OLAP Database Supply Chain. It includes sections for Metadata, Functions, and Search Model. A tree view lists various dimensions: Dim Locations, Dim Customers, Dim Delivery Statuses, Dim Order Statuses, Dim Products, Dim Shipping Modes, Dim Times, and Dim Types.
- Query Editor:** Contains two MDX queries:
 - Logistic Analysis


```
-- Query 11: Tinh trang giao hang dung hen tai chau A - Thai Binh Duong
SELECT
{
    [Measures].[Measures].[AVG Scheduled Day For Shipping],
    [Measures].[Measures].[AVG Real Day For Shipping],
    [Measures].[Late Risk]
} ON COLUMNS,
[Dim Locations].[Country].CHILDREN ON ROWS
FROM [Cube OLAP Database Supply Chain]
WHERE [Dim Locations].[Market].&[Pacific Asia];
```
 - Query 12: Thong ke so luong don hang duoc tao trong cac thu trong tuan


```
-- Query 12: Thong ke so luong don hang duoc tao trong cac thu trong tuan
SELECT
{
    COUNT([Measures].[Measures])
} ON COLUMNS,
[Dim Locations].[Country].CHILDREN ON ROWS
FROM [Cube OLAP Database Supply Chain]
WHERE [Dim Locations].[Market].&[Pacific Asia];
```
- Results Grid:** Shows the results of the second query, displaying columns for AVG Scheduled Day For Shipping, AVG Real Day For Shipping, and Late Risk across various countries.
- Status Bar:** Query executed successfully. DESKTOP-OEFOODK | DESKTOP-OEFOODK\anhthy | OLAP_SSAS_SupplyChain | 00:00:01

Truy vấn thủ công từ cube

The screenshot shows the Cube Designer in SSDT with the following details:

- Title Bar:** Cube OLAP D...ube [Design]
- Toolbar:** Includes Cube Structure, Dimensions, Calculations, KPIs, Actions, Partitions, Aggregates, Perspectives, Translations, and Browser.
- Query Editor:** Shows an MDX query being edited:


```
MDX
```
- Filter Configuration:** A table for setting filters on dimensions. For Dim Locations, it shows a Market filter set to Equal with the value { Pacific Asia }.
- Dimensions View:** Shows the structure of the Dim Locations dimension, including City, Country, Location Id, Market, and Region.
- Calculated Members View:** An empty list.
- Results Grid:** Displays the results of the MDX query, matching the data shown in the SSMS results grid above.

Truy vấn từ Excel

Market	Pacific Asia	AVG Scheduled Day For Shipping	AVG Real Day For Shipping	Late Risk
Row Labels				
Afganistán	6.410956904	7.767123288	Late	
Arabia Saudí	8.495016611	10.09302326	Late	
Armenia	5	10	Late	
Australia	6.603397929	7.889567295	Late	
Azerbaiyán	8.05	10	Late	
Bangladés	6.315533981	7.810679612	Late	
Baréin	16	12	Not late	
Bután	10	15	Late	
Camboya	6.46969697	7.606060606	Late	
China	6.542066846	7.781790242	Late	
Corea del Sur	6.076363636	7.669090909	Late	
Emiratos Árabes Unidos	6.090090901	7.636363636	Late	
Filipinas	6.598532495	7.857442348	Late	
Georgia	7.966666667	9.9	Late	
Hong Kong	5.702702703	6.405405405	Late	
India	6.485287249	7.860345633	Late	
Indonesia	6.362924282	7.636553525	Late	
Irak	9.116719243	10.52365931	Late	
Irán	8.975855131	10.55734406	Late	

Truy vấn từ Power Bi

Country	AVG Scheduled Day For Shipping	AVG Real Day For Shipping	Late Risk
Afganistán	6.41	7.77	Late
Arabia Saudí	8.50	10.09	Late
Armenia	5.00	10.00	Late
Australia	6.60	7.89	Late
Azerbaiyán	8.05	10.00	Late
Bangladés	6.32	7.81	Late
Baréin	16.00	12.00	Not late
Bután	10.00	15.00	Late
Camboya	6.47	7.61	Late
China	6.54	7.78	Late
Corea del Sur	6.08	7.67	Late
Total	7.07	8.45	Late

Câu truy vấn 12: Thống kê số lượng item trong đơn hàng được tạo trong các thứ trong tuần

Truy vấn MDX

IS217 – Kho dữ liệu và OLAP

project.mdx - DES...P-OEFOODK\anhthy

Cube: Cube OLAP Database Supply Chain

Measure Group: <All>

```
[Measures].[Measures].[AVG Real Day For Shipping],  
[Measures].[Late Risk]  
} ON COLUMNS,  
[Dim Locations].[Country].CHILDREN ON ROWS  
FROM [Cube OLAP Database Supply Chain]  
WHERE [Dim Locations].[Market].&[Pacific Asia];  
  
-- Query 12: Thong ke so luong don hang duoc tao trong cac thu trong tuan  
SELECT  
NON EMPTY [Measures].[Fact Order Items Count] ON COLUMNS,  
NON EMPTY [Dim Times].[Date Of Week].[Date Of Week].MEMBERS ON ROWS  
FROM [Cube OLAP Database Supply Chain];  
  
-- Query 13: Thong ke tinh trang giao hang tai Trung Quoc
```

	Fact Order Items Count
Friday	25670
Monday	25923
Saturday	25808
Sunday	25846
Thursday	25558
Tuesday	25813
Wednesday	25901

Truy vấn thủ công từ cube

Cube OLAP Database Supply Chain [Design]

Dimension Hierarchy Operator Filter Expression

Date Of Week	Fact Order Items ...
Friday	25670
Monday	25923
Saturday	25808
Sunday	25846
Thursday	25558
Tuesday	25813
Wednesday	25901

Calculated Members

Truy vấn Excel

Row Labels	Fact Order Items Count
Friday	25670
Monday	25923
Saturday	25808
Sunday	25846
Thursday	25558
Tuesday	25813
Wednesday	25901
Grand Total	180519

Truy vấn từ Power Bi

Date Of Week	Fact Order Items Count
Friday	25670
Monday	25923
Saturday	25808
Sunday	25846
Thursday	25558
Tuesday	25813
Wednesday	25901
Total	180519

Câu truy vấn 13: Thống kê tình trạng giao hàng của 4 loại vận chuyển tại Trung Quốc

Truy vấn MDX

The screenshot shows the Microsoft Analysis Services Management Studio (SSMS) interface. On the left, the 'Cube' pane displays the 'Cube OLAP Database Supply Chain' with various dimensions and measures selected. The main area contains an MDX query window with the following content:

```

NON EMPTY [Measures].[Fact Order Items Count] ON COLUMNS,
NON EMPTY [Dim Times].[Date Of Week].[Date Of Week].MEMBERS ON ROWS
FROM [Cube OLAP Database Supply Chain];
-- Query 13: Thong ke tinh trang giao hang tai Trung Quoc
SELECT
NON EMPTY {[Measures].[Fact Order Items Count], [Dim Shipping Modes].[Shipping Mode].CHILDREN}
NON EMPTY {[Dim Order Statuses].[Order Status].CHILDREN} ON ROWS
FROM [Cube OLAP Database Supply Chain]
WHERE [Dim Locations].[Country].&[China]&[Pacific Asia];
-- Customer Analysis
-- Query 14: Danh sach id cac customer thanh toan cash > 700 hoac debit > 1200

```

Below the query window is a results grid titled 'Results' showing the count of fact order items across different shipping modes and order statuses. The data is as follows:

	Fact Order Items Count			
	First Class	Same Day	Second Class	Standard Class
CANCELED	12	5	34	84
CLOSED	98	45	130	384
COMPLETE	234	89	371	1082
ON_HOLD	28	22	68	182
PAYMENT REVIEW	9	6	12	43
PENDING	96	60	134	398
PENDING PAYMENT	184	66	235	862
PROCESSING	121	32	103	405
SUSPECTED FRAUD	39	9	15	61

At the bottom, a message bar indicates 'Query executed successfully.' and shows the session details: DESKTOP-OEFOODK / DESKTOP-OEFOODK\anhthy / OLAP_SSAS_SupplyChain | 00:00:01.

Truy vấn thủ công từ cube

The screenshot shows the Microsoft Analysis Services Cube Designer in 'Design' mode. The top navigation bar includes options like Cube Structure, Dimensions, Calculations, KPIs, Actions, Partitions, Aggregates, Perspectives, Translations, and Browser. The main area has tabs for 'Edit as Text' and 'Import...', and a dropdown for 'MDX'.

The left pane displays the cube structure with dimensions: Dim Locations, Dim Order Statuses, Dim Products, Dim Shipping Modes, and Dim Times. The right pane shows a query editor and a results grid.

Query Editor:

```

SELECT
NON EMPTY {[Measures].[Fact Order Items Count], [Dim Shipping Modes].[Shipping Mode].CHILDREN}
NON EMPTY {[Dim Order Statuses].[Order Status].CHILDREN} ON ROWS
FROM [Cube OLAP Database Supply Chain]
WHERE [Dim Locations].[Country].&[China]&[Pacific Asia];

```

Results Grid:

Order Status	Shipping Mode	Fact Order Items ...
CANCELED	First Class	12
CANCELED	Same Day	5
CANCELED	Second Class	34
CANCELED	Standard Class	84
CLOSED	First Class	98
CLOSED	Same Day	45
CLOSED	Second Class	130
CLOSED	Standard Class	384
COMPLETE	First Class	234
COMPLETE	Same Day	89
COMPLETE	Second Class	371
COMPLETE	Standard Class	1082
ON_HOLD	First Class	28
ON_HOLD	Same Day	22
ON_HOLD	Second Class	68
ON_HOLD	Standard Class	182
PAYMENT ...	First Class	9
PAYMENT ...	Same Day	6

Truy vấn từ Excel

The screenshot shows a Power BI interface with a PivotTable Fields pane on the right and a data grid on the left.

PivotTable Fields:

- Choose fields to add to report: Country
- Search:
- More Fields:
 - City
 - Country
 - Location Id
 - Market

Drag fields between areas below:

- Filters: Country
- Columns:
- Rows: Order Status
- Values: Fact Order Items Count
- Defer Layout Update

Data Grid:

A	B	C	D	E	F	G	H	I	J	K
1	Country	China								
2										
3	Row Labels	Fact Order Items Count								
4	CANCELED									
5	First Class	12								
6	Same Day	5								
7	Second Class	34								
8	Standard Class	84								
9	CLOSED									
10	First Class	98								
11	Same Day	45								
12	Second Class	130								
13	Standard Class	384								
14	COMPLETE									
15	First Class	234								
16	Same Day	89								
17	Second Class	371								
18	Standard Class	1082								
19	ON HOLD									
20	First Class	28								
21	Same Day	22								
22	Second Class	68								

Bottom navigation bar: Ready, Accessibility: Good to go, ... | Query 8 | Query 9 | Query 10 | Query 11 | Query 12 | **Query 13 ...** | + | : | ← | → | Update | 100%

Truy vấn từ Power BI

Order Status	Shipping Mode	Fact Order Items Count
CANCELED	First Class	615
CANCELED	Same Day	225
CANCELED	Second Class	661
CANCELED	Standard Class	2191
CLOSED	First Class	3017
CLOSED	Same Day	1124
CLOSED	Second Class	3804
CLOSED	Standard Class	11671
COMPLETE	First Class	9307
COMPLETE	Same Day	3164
COMPLETE	Second Class	11655
COMPLETE	Standard Class	35365
ON_HOLD	First Class	1455
ON_HOLD	Same Day	522
ON_HOLD	Second Class	1815
ON_HOLD	Standard Class	6012
Total		180519

3.8.3. Các câu truy vấn về khách hàng

Câu truy vấn 14: Danh sách khách hàng đem lại lợi nhuận từ Cash > 700 hoặc từ Debit trên 1200.

Truy vấn MDX

The screenshot shows the SSMS interface with the following details:

- Query Editor:** Contains an MDX query to filter customers based on their benefit per order being greater than 700 or debit being greater than 1200.
- Results Grid:** Displays a table with columns: Customer ID, Customer Fname, Customer Lname, Sum Benefit Per Order (CASH), and Sum Benefit Per Order (DEBIT). The data includes rows for various customers like Laura Smith, Robert Sawyer, etc.
- Status Bar:** Shows "Query executed successfully." and the execution time "00:00:01".

Truy vấn thủ công từ cube

The screenshot shows the Cube Designer interface with the following details:

- Query Editor:** Shows a manual query using MDX to filter customers based on their cash or debit values.
- Results Grid:** Displays a table with columns: Customer Id, Customer Fname, Customer Lname, Type, and Sum Benefit Per ... (partially visible). The data includes rows for various customers like Jesse Matthews, Phillip Smith, etc.

Truy vấn từ Excel

The screenshot shows a Microsoft Power BI interface. On the left is a data grid with columns A through M. Row 3 is labeled "Row Labels" and "Sum Benefit Per Order". Rows 4 through 22 list customer data. The right side features a "PivotTable Fields" pane with fields like "Customer Fname", "Customer Lname", and "Sets". Below it are sections for "Filters", "Columns", "Rows", and "Values". The "Values" section shows "Customers Cash > ... Sum Benefit Per Order". At the bottom, there are tabs for "Query 12", "Query 13", "Query 14", and "Ready Recovered Accessibility Good to go".

Truy vấn từ Power Bi

Clipboard

Customer Id	Customer Lname	Customer Fname	Type	Sum Benefit Per Order
2626	Smith	Laura	CASH	750.88
5470	Sawyer	Robert	CASH	838.71
5691	Brown	James	CASH	808.44
5755	Shah	Marie	CASH	749.13
6054	Reilly	Richard	CASH	723.06
6491	Cruz	Bruce	CASH	723.17
7254	Smith	Mary	CASH	713.94
9119	Smith	Mary	CASH	716.35
9541	Leach	Richard	CASH	879.48
11452	Orr	Kathleen	CASH	746.43
Total				8,407.97

Customer Id	Customer Lname	Customer Fname	Type	Sum Benefit Per Order
8290	Jennings	Benjamin	DEBIT	1,227.67
4258	Rogers	Charles	DEBIT	1,211.89
5186	Robinson	Jason	DEBIT	1,325.94
749	Matthews	Jesse	DEBIT	1,325.74
1148	Anderson	Mary	DEBIT	1,543.39
4568	Hodge	Mary	DEBIT	1,285.11
9557	Figueroa	Mary	DEBIT	1,283.76
10672	Smith	Mary	DEBIT	1,258.95
2564	Smith	Phillip	DEBIT	1,212.09
3868	Smith	Rachel	DEBIT	1,255.53
Total				15,580.24

Câu truy vấn 15: Top 10 khách hàng đem lại lợi nhuận nhiều nhất từ quốc gia đem lại lợi nhuận nhiều nhất

Truy vấn MDX

```
-- Query 15: Lấy 10 khách hàng đem lại nhiều doanh thu nhất trong nước có nhiều doanh thu nhất
SELECT
(
[Measures].[Sum Benefit Per Order]
) ON COLUMNS,
TOPCOUNT (
(
[Dim Customers].[Customer Id].CHILDREN,
[Dim Customers].[Customer Fname].CHILDREN,
[Dim Customers].[Customer Lname].CHILDREN
),
10,
[Measures].[Sum Benefit Per Order]
) ON ROWS
FROM
(
SELECT
TOPCOUNT (
[Dim Locations].[Country].CHILDREN,
1,
[Measures].[Sum Benefit Per Order]
) ON COLUMNS
FROM [Cube OLAP Database Supply Chain]
)
```

Messages | Results

			Sum Benefit Per Order
11430	Mary	Miller	1157.36
1636	Julie	Petersen	876.75
1948	Peter	Smith	875.02
5897	Mary	Griffin	838.17
10340	Nicholas	Stevens	832.48
7960	Ethan	Thomas	817.85
1025	Megan	Todd	804.09
4929	Janet	Harvey	789.88
3979	Mary	Wagner	788.58
9665	Shawn	Smith	781.74

Query executed successfully.

Truy vấn thủ công từ cube

Cube OLAP D...ube [Design]

Dimension Hierarchy Operator Filter Expression

Dim Locations	Country	In	Top 1 Benefit Country
Dim Customers	Customer Id	In	Top 10 Benefit Customer

Customer Id Customer Fname Customer Lname Sum Benefit Per ...

1025	Megan	Todd	804.09
1636	Julie	Petersen	876.75
1948	Peter	Smith	875.02
3979	Mary	Wagner	788.58
4929	Janet	Harvey	789.88
5897	Mary	Griffin	838.17
7960	Ethan	Thomas	817.85
9665	Shawn	Smith	781.74
10340	Nicholas	Stevens	832.48
11430	Mary	Miller	1157.36

Truy vấn từ Excel

The screenshot shows a Power BI interface with a PivotTable on the left and a report card on the right.

PivotTable Fields:

- Choose fields to add to report: Customer Fname, Customer Lname
- Search: Search bar
- Customer Fname (checked), Customer Id, Customer Lname
- Sets: Customers Cash > 700 or Debit > 1200

Drag fields between areas below:

- Filters
- Columns
- Rows: Top 1 Benefit Cou...
- Values: Sum Benefit Per Order
- Defer Layout Update, Update

Truy vấn từ Power BI

Country	Sum Benefit Per Order
Estados Unidos	540,343.26
Total	540,343.26

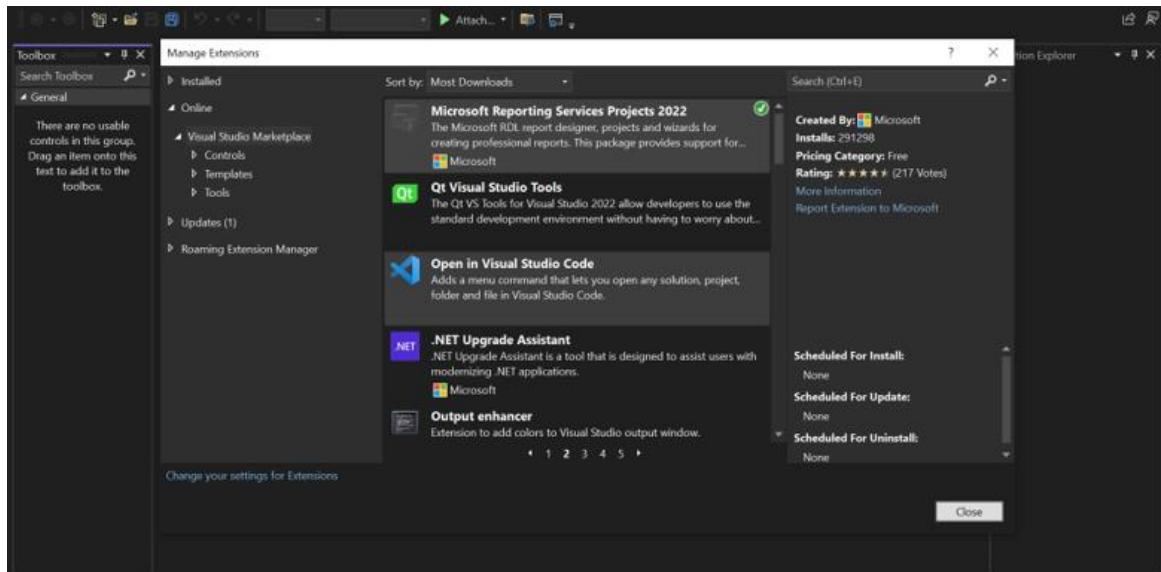
Customer Id	Customer Fname	Customer Lname	Sum Benefit Per Order
749	Jesse	Matthews	1,854.99
1657	Betty	Phillips	2,196.76
2626	Laura	Smith	1,928.45
2641	Betty	Spears	2,441.83
3735	Mary	Lewis	1,906.26
5004	Dorothy	Friedman	1,917.82
5053	Mary	Smith	1,813.21
5560	Mary	Rodriguez	1,831.36
9833	Jacob	Smith	1,938.29
10967	Alexander	Cunningham	1,822.22
Total			19,651.19

Chương 4: SSRS

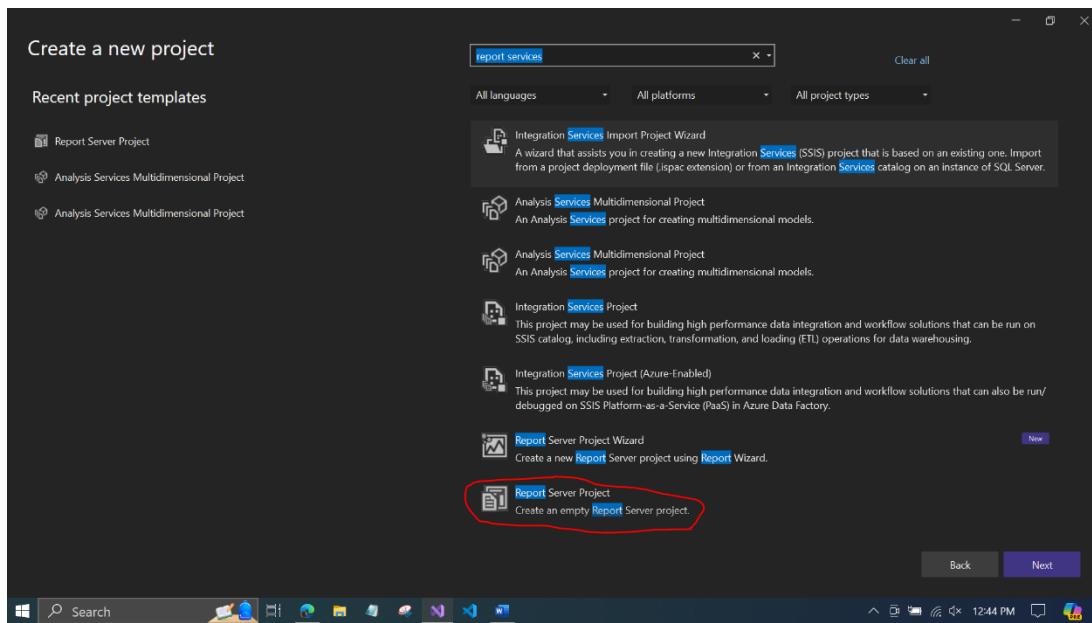
4.1. Cài đặt các công cụ cần thiết

4.2. Cài đặt Report Server Project

Bước 1: Cài đặt “Extension Ms Reporting Services Projects 2022”.



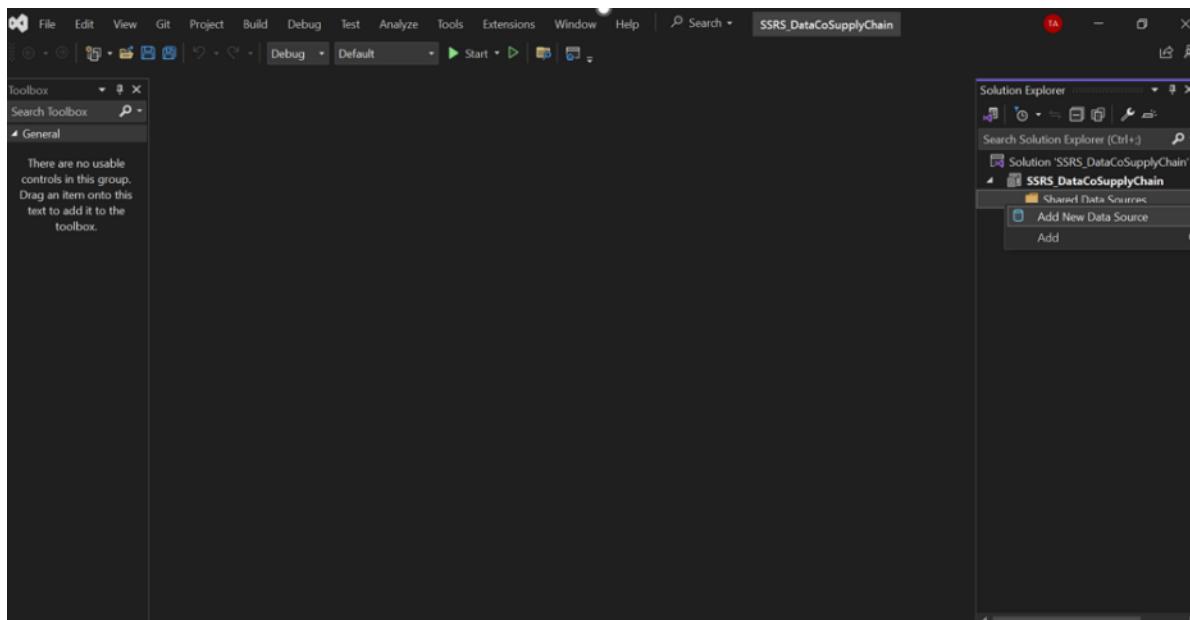
Bước 2: Tạo project SSRS mới.



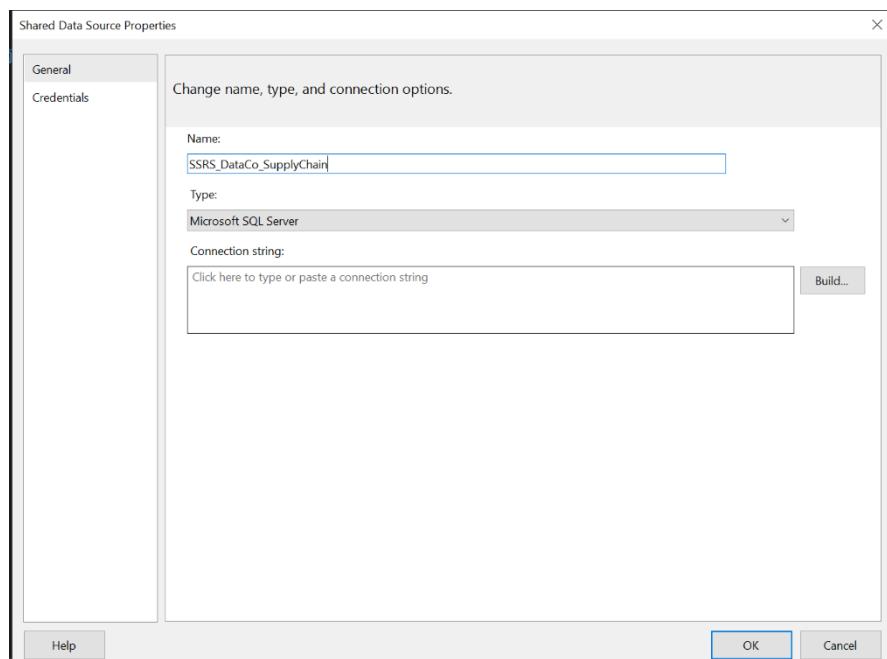
Bước 3: Chọn, đặt tên, và tạo mới dự án.

4.3. Tạo và cấu hình cho SSRS Project

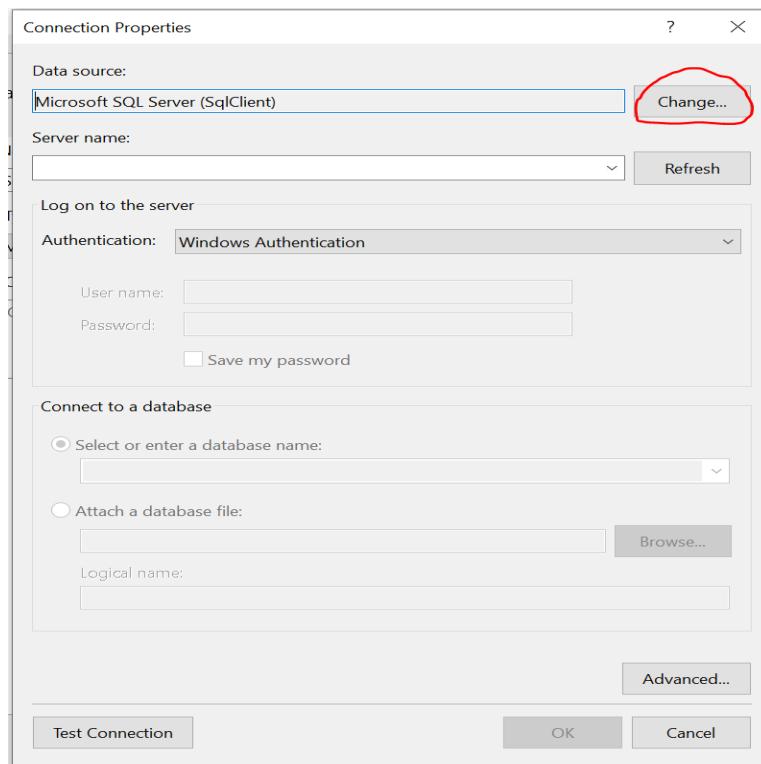
Bước 1: Bên trong “*Solution Explorer*”, chuột phải vào “*Shared Data Sources*”. Chọn “*Add New Data Sources*”.



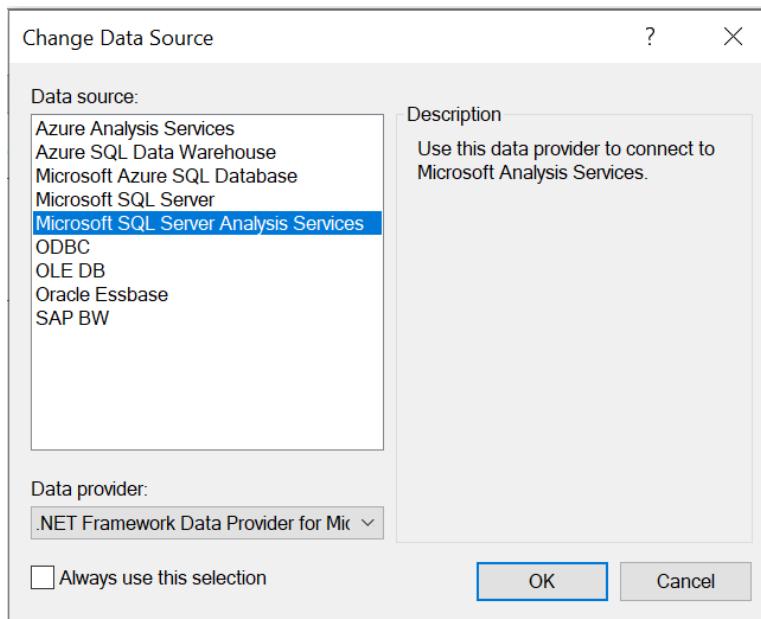
Bước 2: Đặt tên vào chọn loại nguồn data. Sau đó chọn “Build...” để tạo Connection.



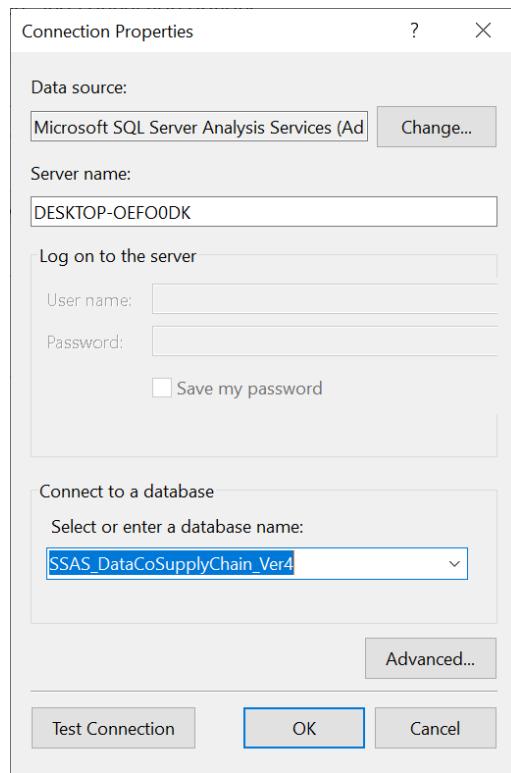
Bước 3: Chon “Change” để chuyển sang lấy data từ SSAS



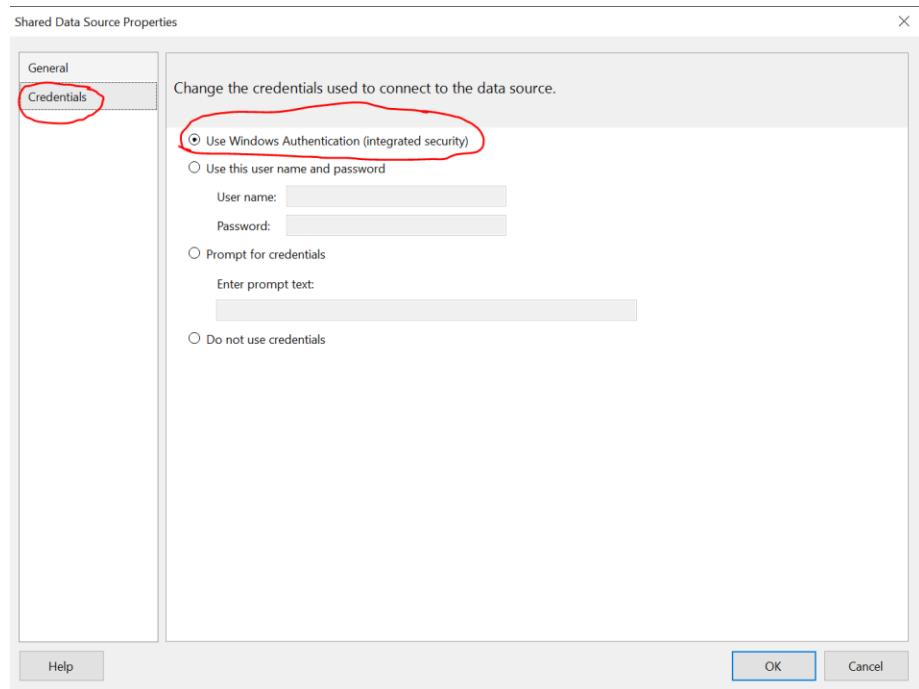
Bước 4: Chọn “*Microsoft SQL Server Analysis Services*”, nhấn “OK” để kết thúc.



Bước 5: Cấu hình cho connection, sau đó ấn “OK”.

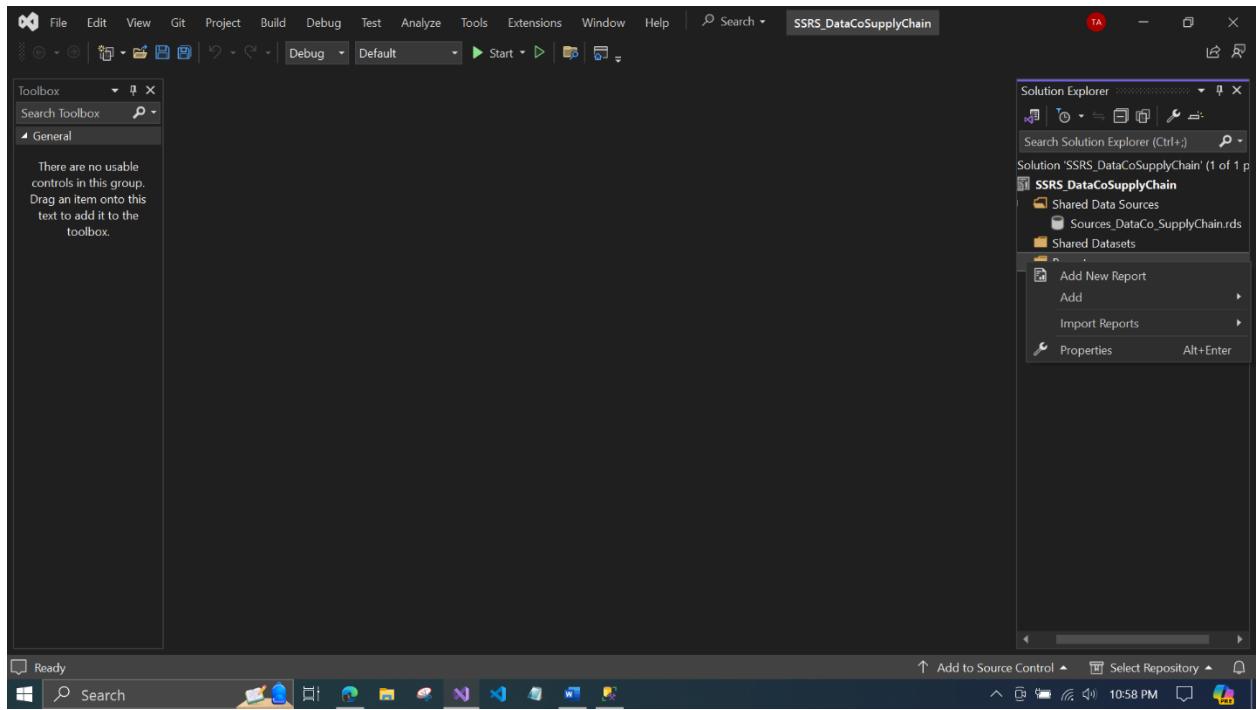


Bước 6: Chọn “*Credentials*”, và chọn phương thức kết nối phù hợp. Cuối cùng, chọn “OK” để hoàn thành tạo Data Source.

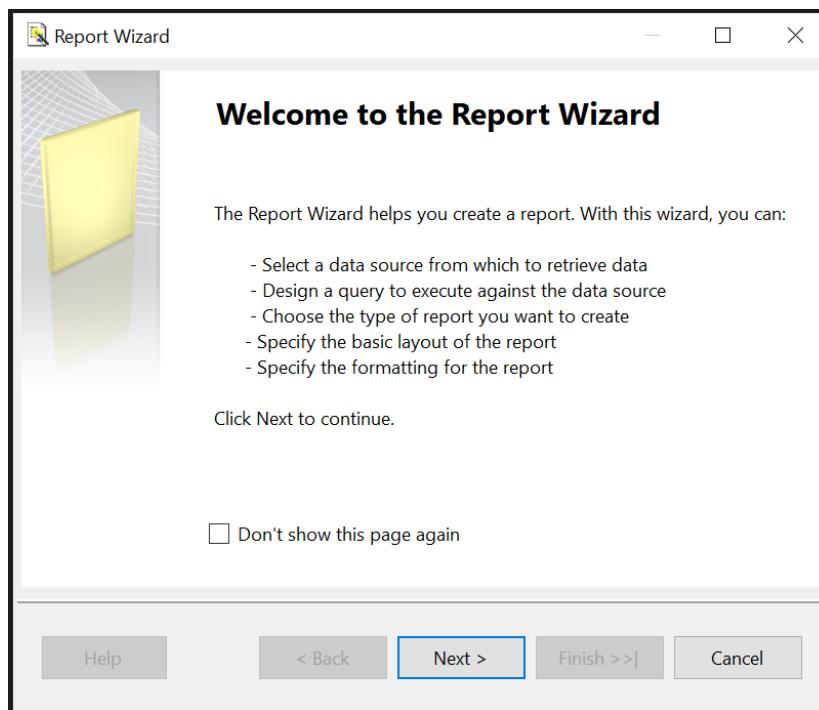


4.4. Tạo và cấu hình cho Power BI

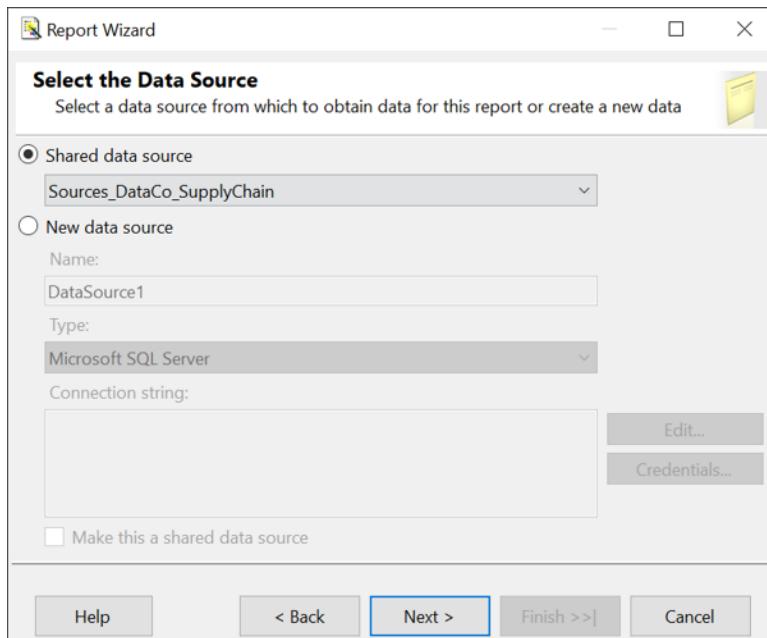
Bước 1: Bên trong “Solution Explorer”, chuột phải vào “Report”. Chọn “Add New Report”.



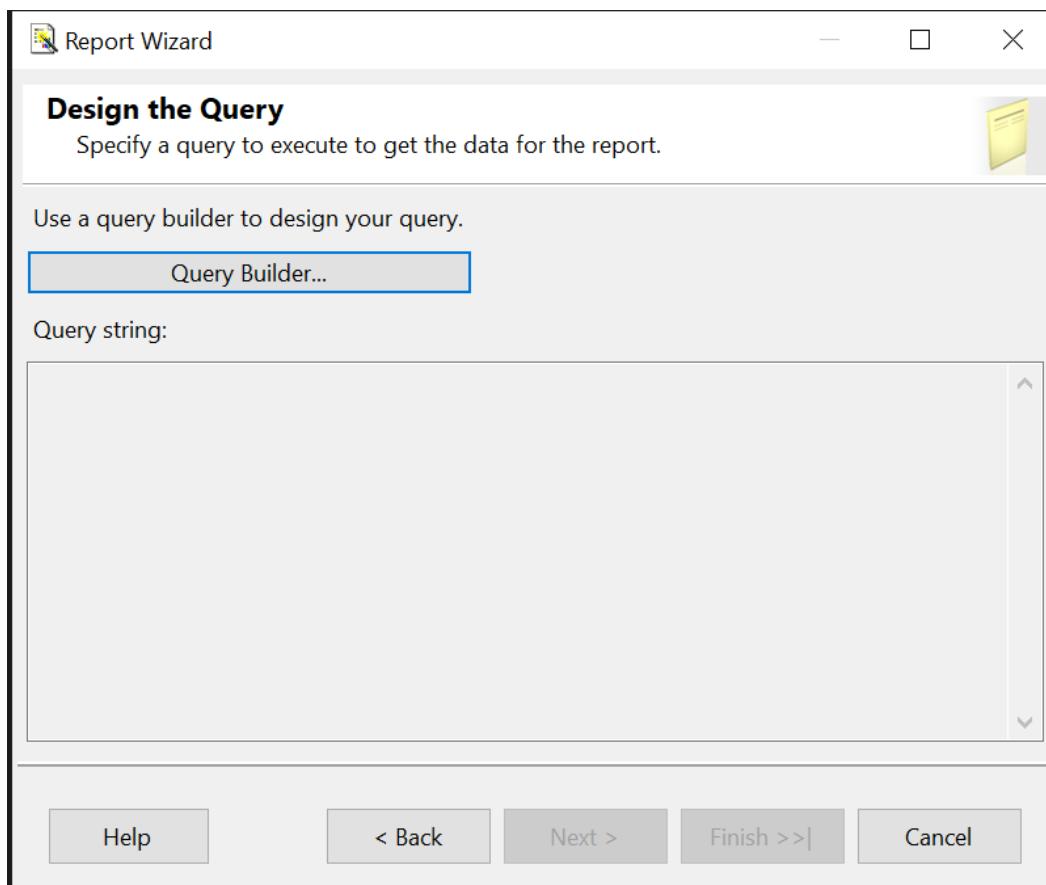
Bước 2: chọn “Next”.



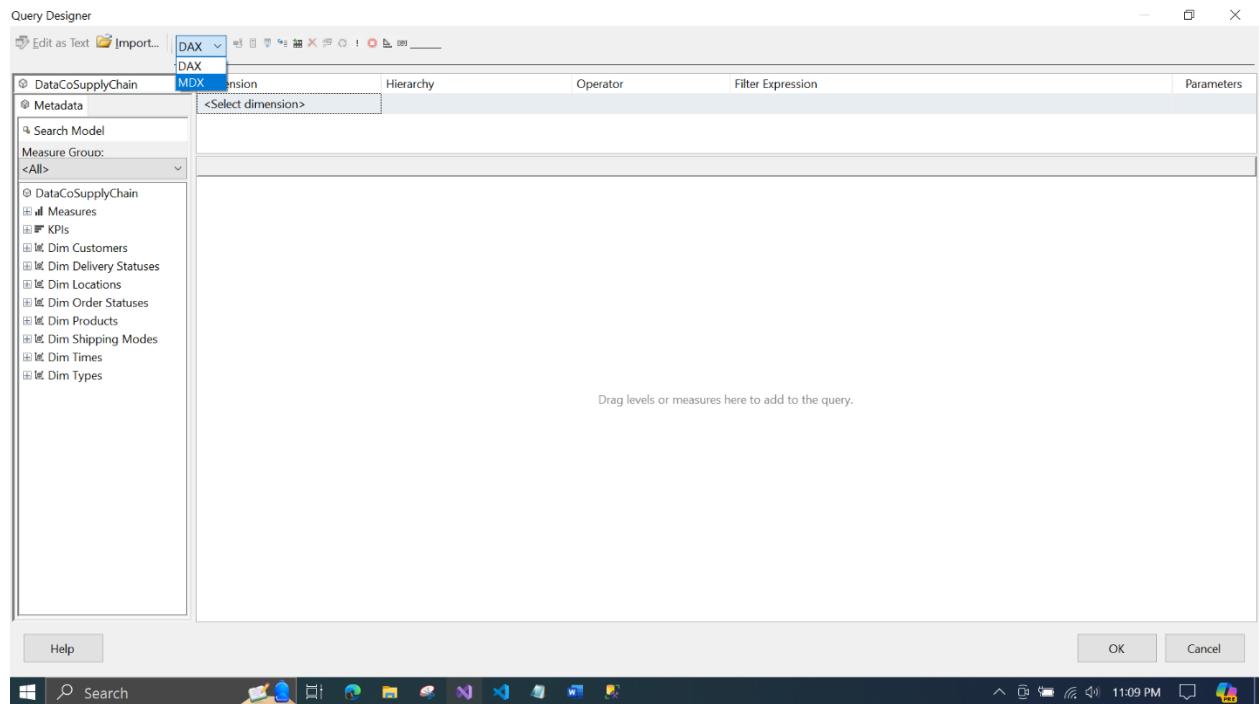
Bước 3: Chọn data sources phù hợp và nhấn “Next”.



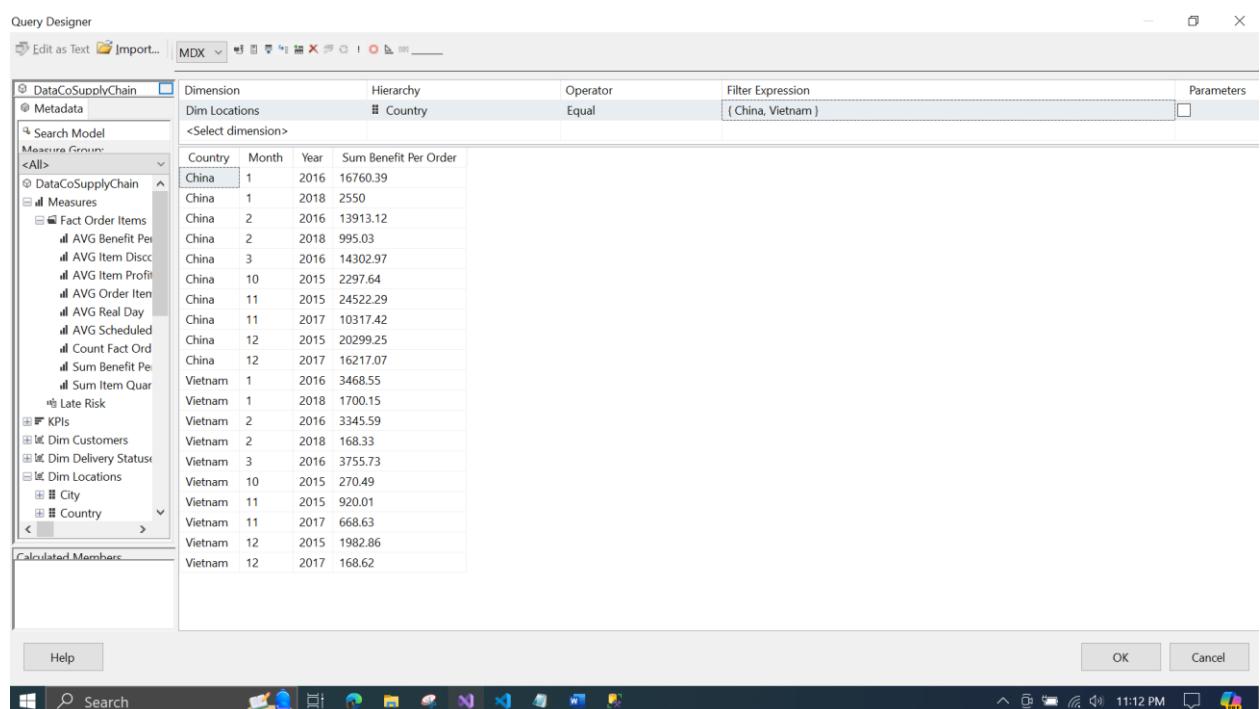
Bước 4: Chọn “Query Builder...” để chọn dữ liệu dùng chính cho report.



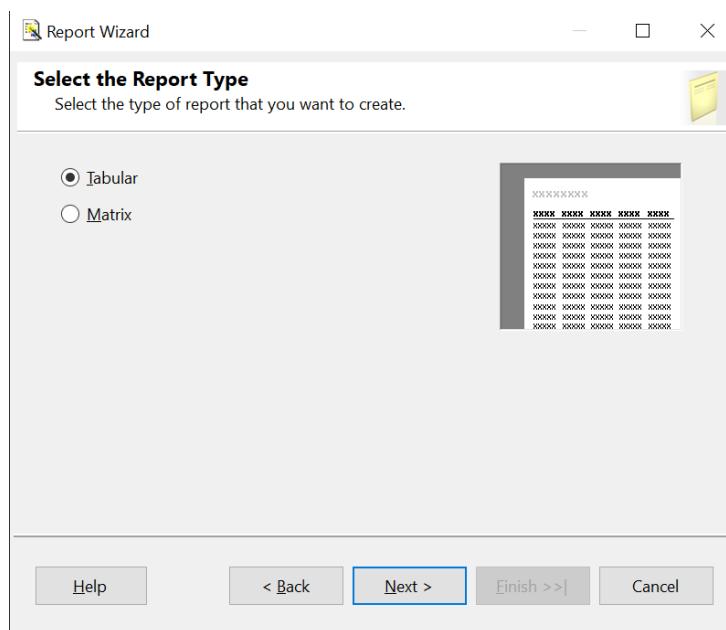
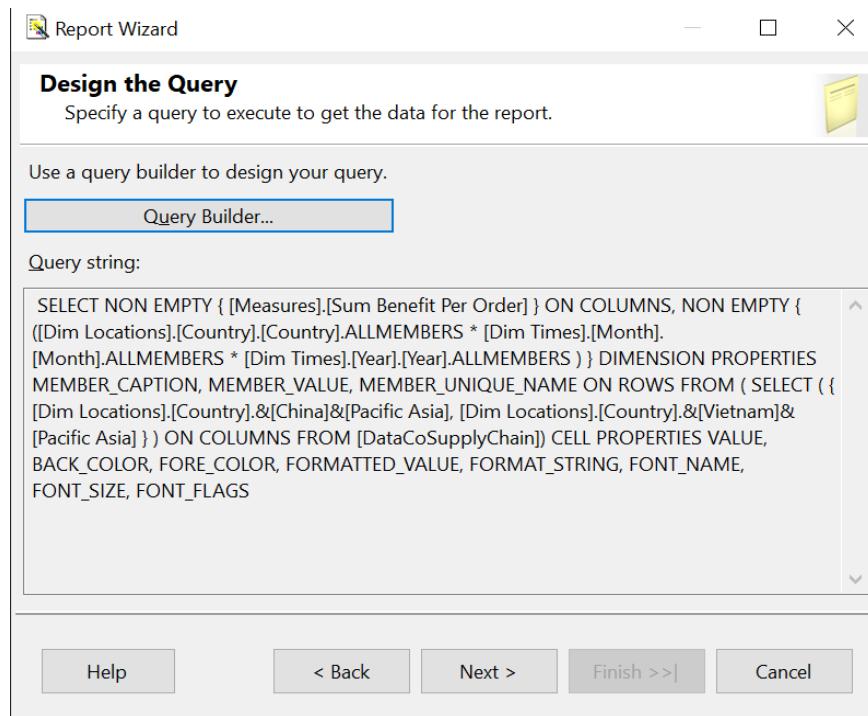
Bước 5: Tại cửa sổ “*Query Builder*”, chuyển sang MDX và tiến hành truy vấn tương tự trong quá trình SSIS.

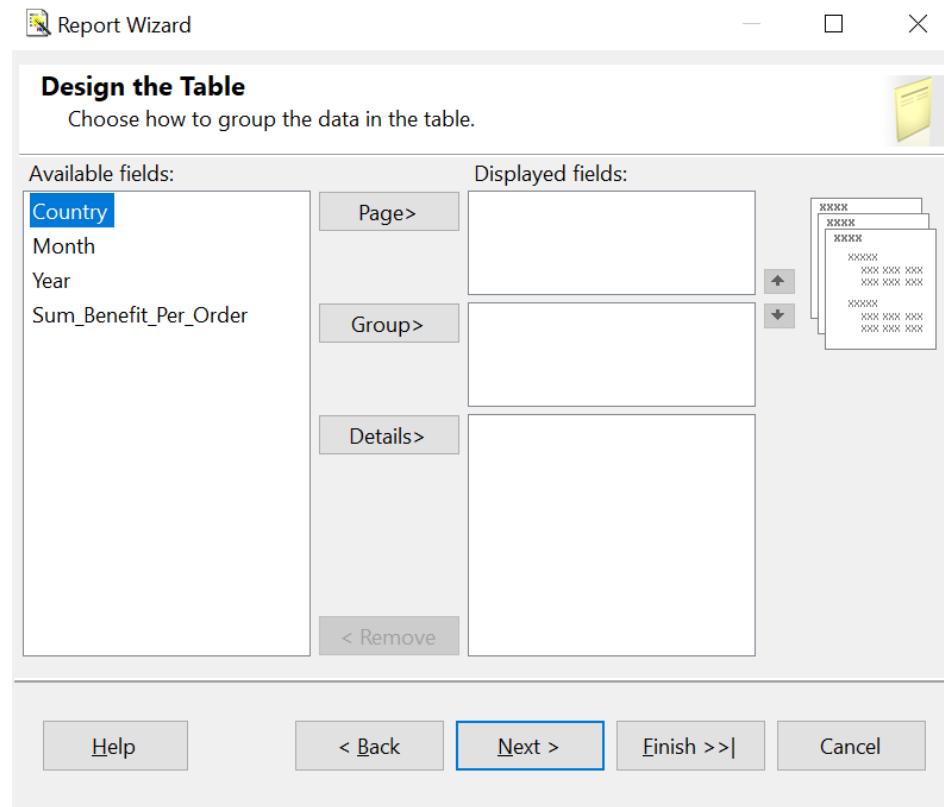


Bước 6: Nhấn “OK” để trở về cửa sổ tạo report.

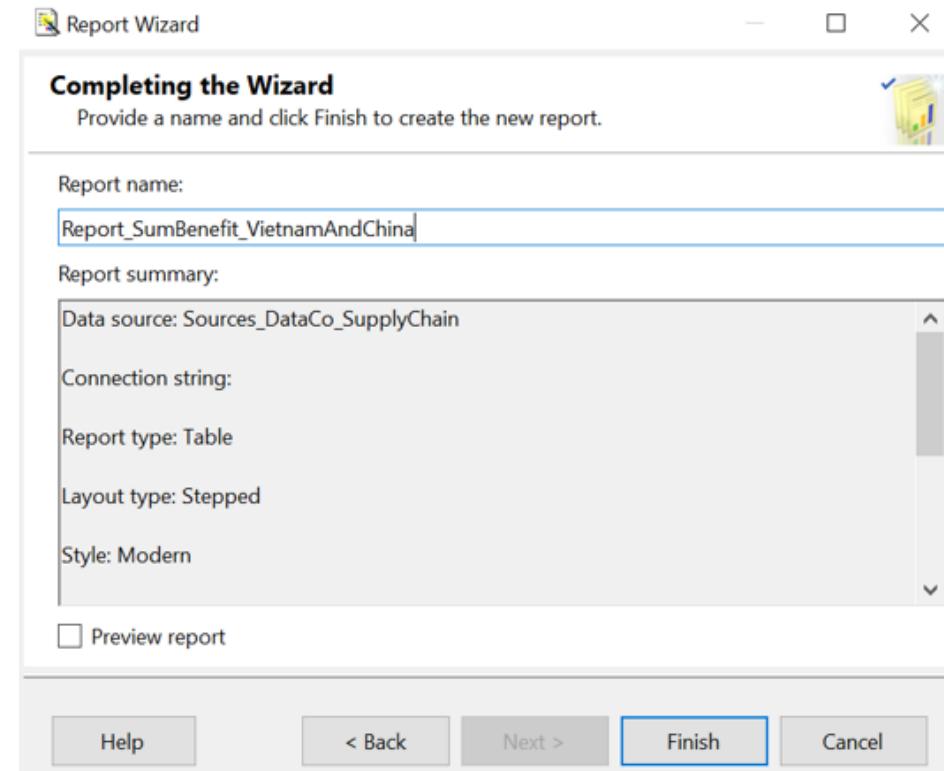


Bước 7: Nhấn “Next” để chuyển qua các tab khác nhau, cấu hình các tab phù hợp với yêu cầu.





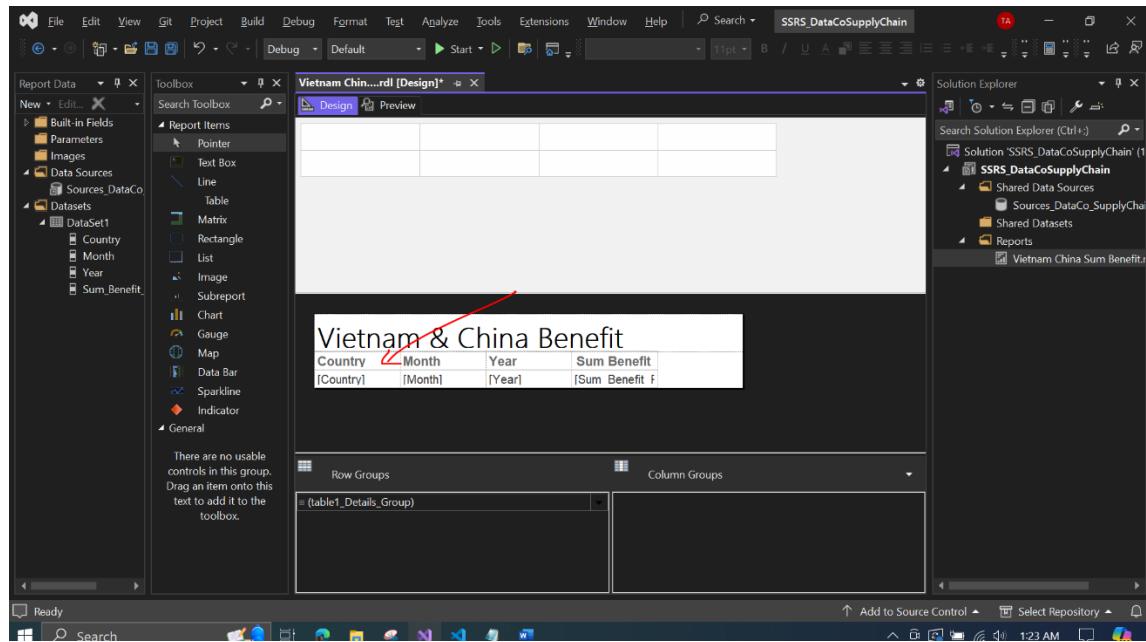
Bước 8: Chọn “Finish” để kết thúc.



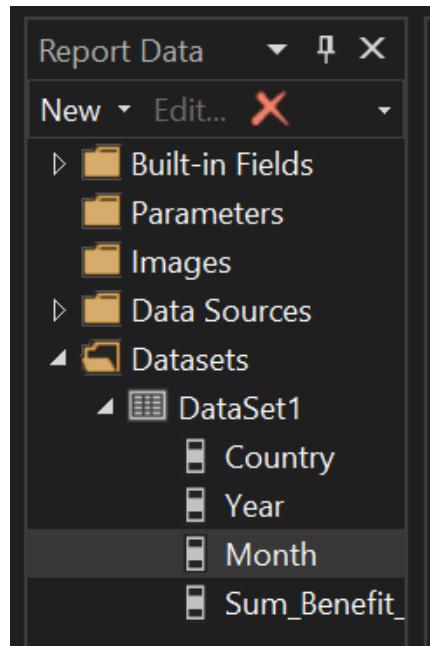
Bước 9: Thiết kế report tại tab “Design” và xem tại “Preview”.

Tạo Group

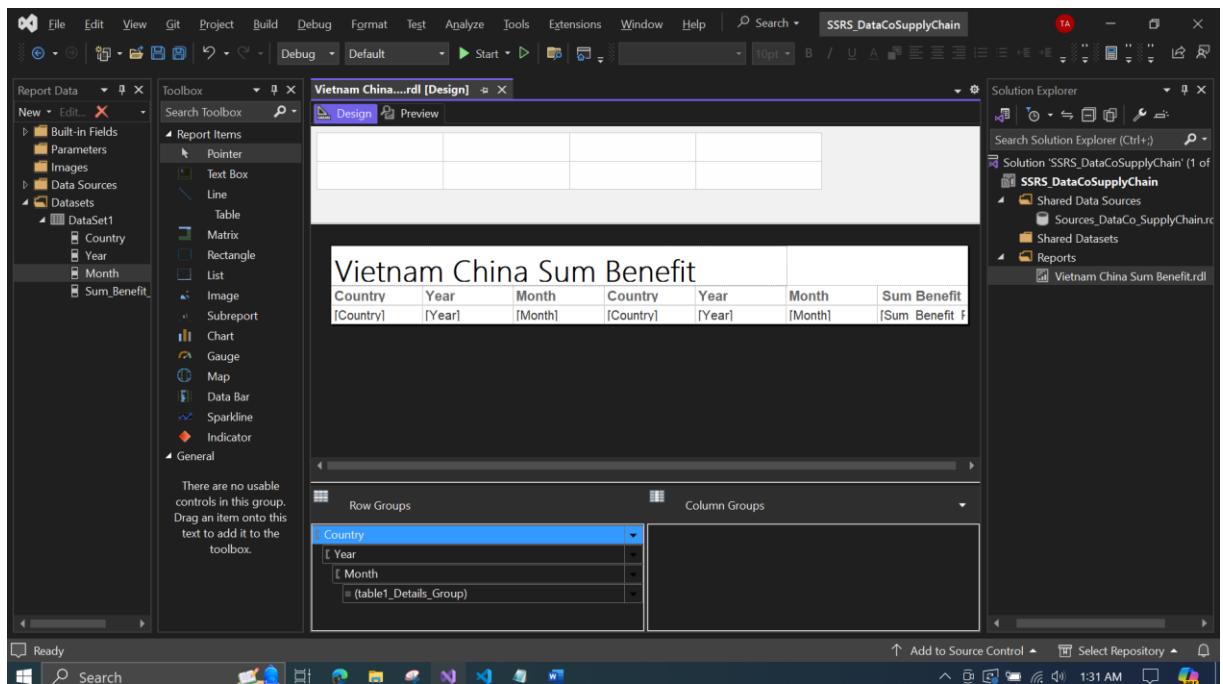
Bước 10: Chọn table cần tạo group.



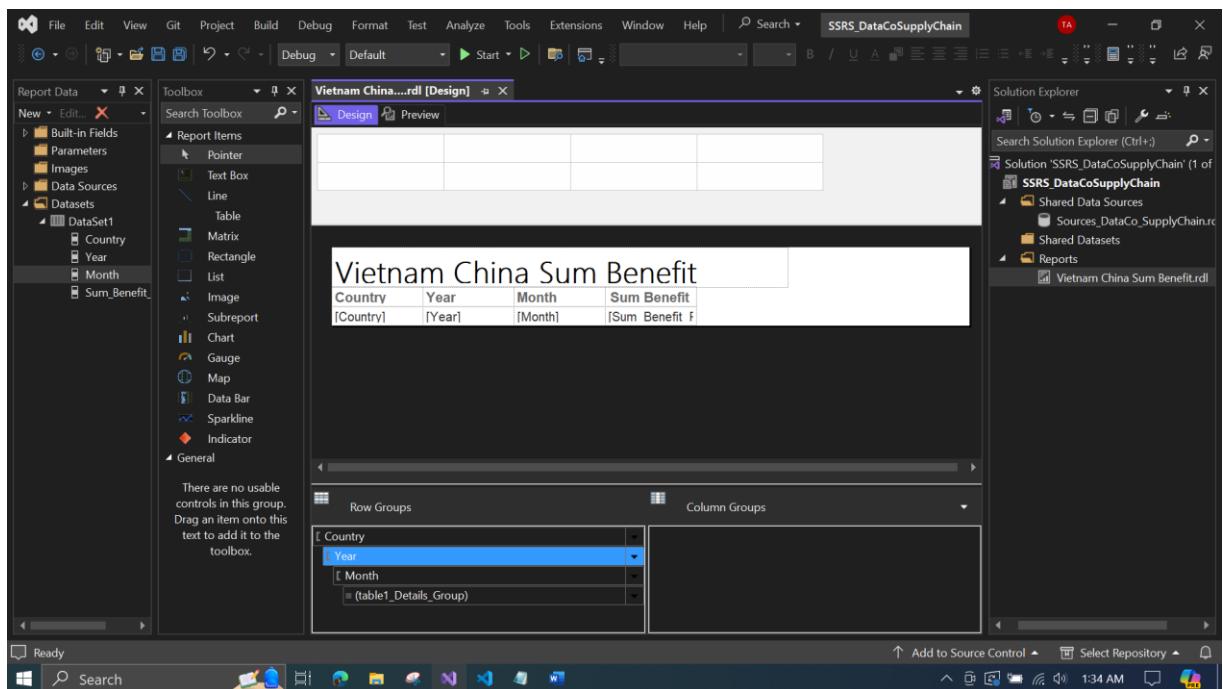
Bước 11: Tại Toolbox, mở rộng Datasets, mở rộng dataset muốn group.



Bước 12: Kéo thả thuộc tính muốn group lên thuộc tính con của group (ví dụ: detail là mặc định, sẽ nằm dưới cùng, kéo “Month” lên trên “detail”, kéo “Year” lên trên “Month”, “Country” lên trên “Year”) bên trong “Row Groups”.



Bước 13: Xoá các cột cũ (nếu không còn cần thiết).



⇒ So sánh trước và sau khi group

Trước

Sau

The image shows two side-by-side Microsoft Excel windows. Both windows have the title "Vietnam China Sum Benefit" and contain the same data. The data is organized into four columns: Country, Year, Month, and Sum Benefit Per Order. The rows list data for China from 2015 to 2018 and Vietnam from 2015 to 2018. The data is identical in both windows.

Country	Year	Month	Sum Benefit Per Order
China	2015	10	2297.64
China	2015	11	24522.29
China	2015	12	20299.25
China	2016	1	16760.39
China	2016	2	13913.12
China	2016	3	14302.97
China	2017	11	10317.42
China	2017	12	16217.07
China	2018	1	2550
China	2018	2	995.03
Vietnam	2015	10	270.49
Vietnam	2015	11	920.01
Vietnam	2015	12	1982.86
Vietnam	2016	1	3468.55
Vietnam	2016	2	3345.59
Vietnam	2016	3	3755.73
Vietnam	2017	11	668.63
Vietnam	2017	12	168.62
Vietnam	2018	1	1700.15
Vietnam	2018	2	168.33

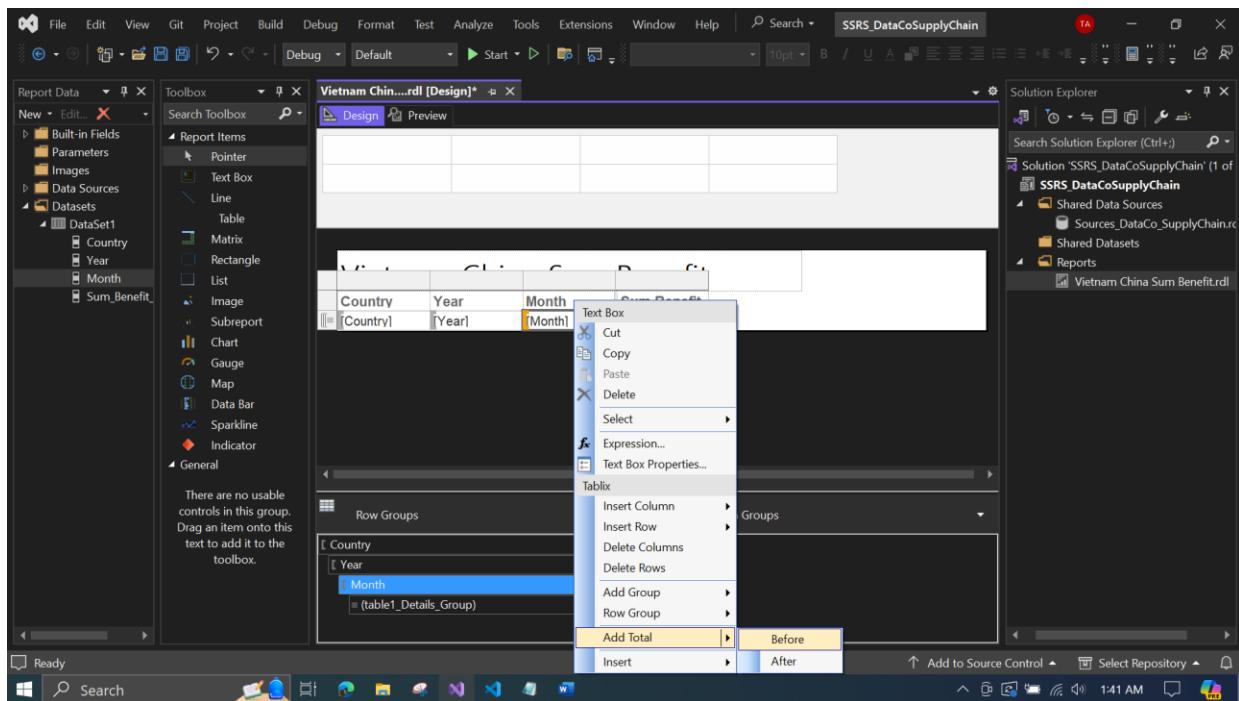
Tạo Total:

Bước 14: Chuột trái 1 lần vào ô cần tạo total.

The image shows a screenshot of the "Vietnam China Sum Benefit" table. The table has four columns: Country, Year, Month, and Sum Benefit. The "Sum Benefit" column header is highlighted with a red arrow. The table also contains placeholder text like "[Country]" and "[Year]" in the first few rows.

Country	Year	Month	Sum Benefit
[Country]	[Year]	[Month]	[Sum Benefit F]

Bước 15: Chuột phải lăn nữa để mở bảng chức năng, chọn “Add total” → “Before” hoặc “After” mặc định là before.



- Lưu ý: add total tại ô nào, thì có nghĩa là tính Sum cho cột đó, tương đương với việc tính cho cột mang thuộc tính cha. Ví dụ đặt total tại month thì sẽ hiện total tại year.

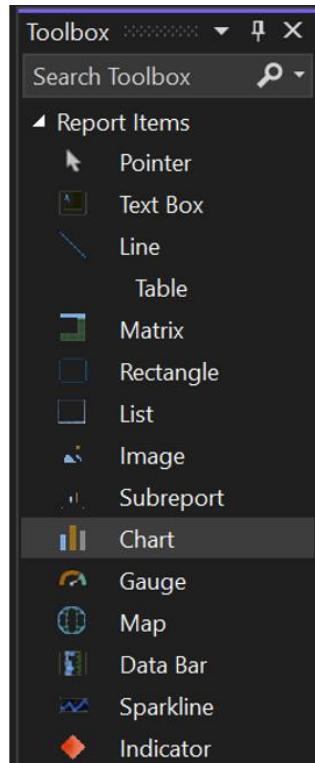
Vietnam China Sum Benefit			
Country	Year	Month	Sum Benefit
[Country]	[Year]	[Month]	[Sum Benefit F]
		Total Months	[Sum(Sum_Ben)]
	Total Years in		[Sum(Sum_Ben)]

Kết quả:

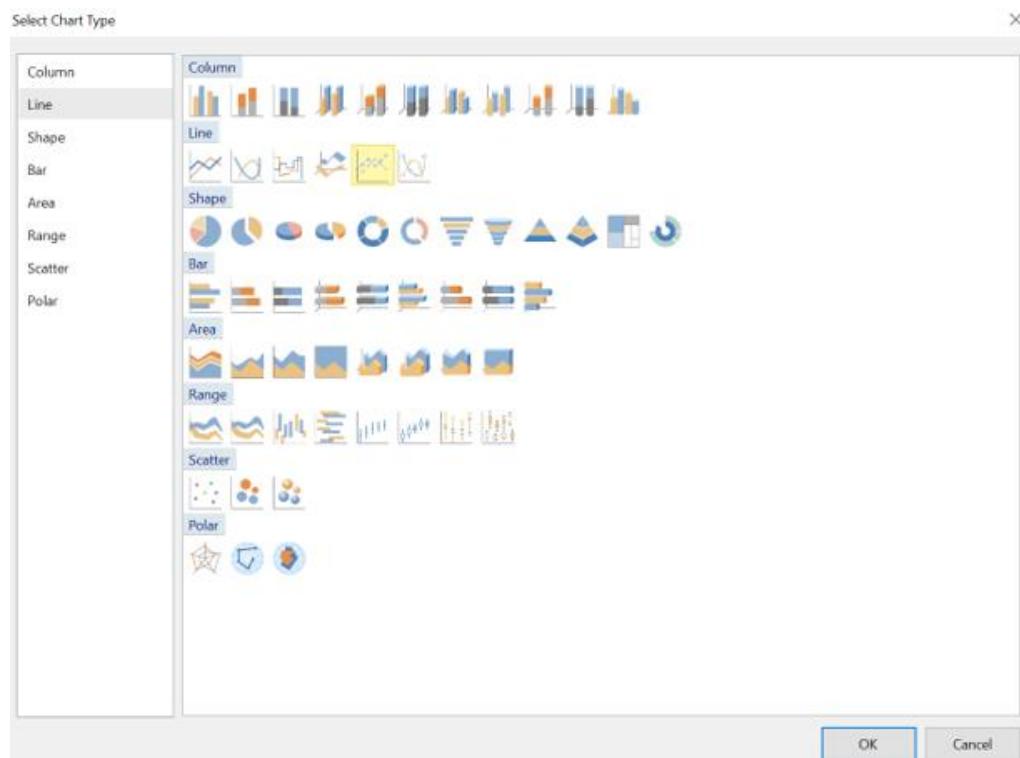
Vietnam	2015	10	270.49
		11	920.01
		12	1982.86
		Total Months of Year	3173.36
	2016	1	3468.55
		2	3345.59
		3	3755.73
		Total Months of Year	10569.87
	2017	11	668.63
		12	168.62
		Total Months of Year	837.25
	2018	1	1700.15
		2	168.33
		Total Months of Year	1868.48
	Total Years in Country		16448.96

Tạo Chart

Bước 16: Double Click vào “Chart” bên trong toolbox.



Bước 17: Chọn loại chart phù hợp. Sau đó, chọn “OK”.

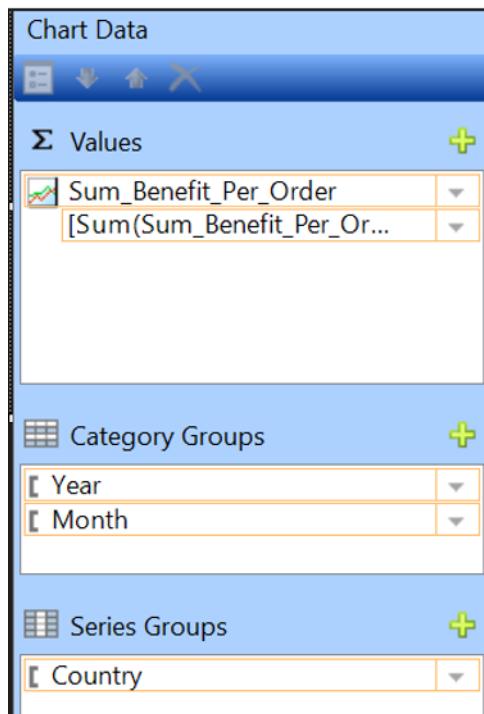


Bước 18: Kéo chart đến vị trí phù hợp.

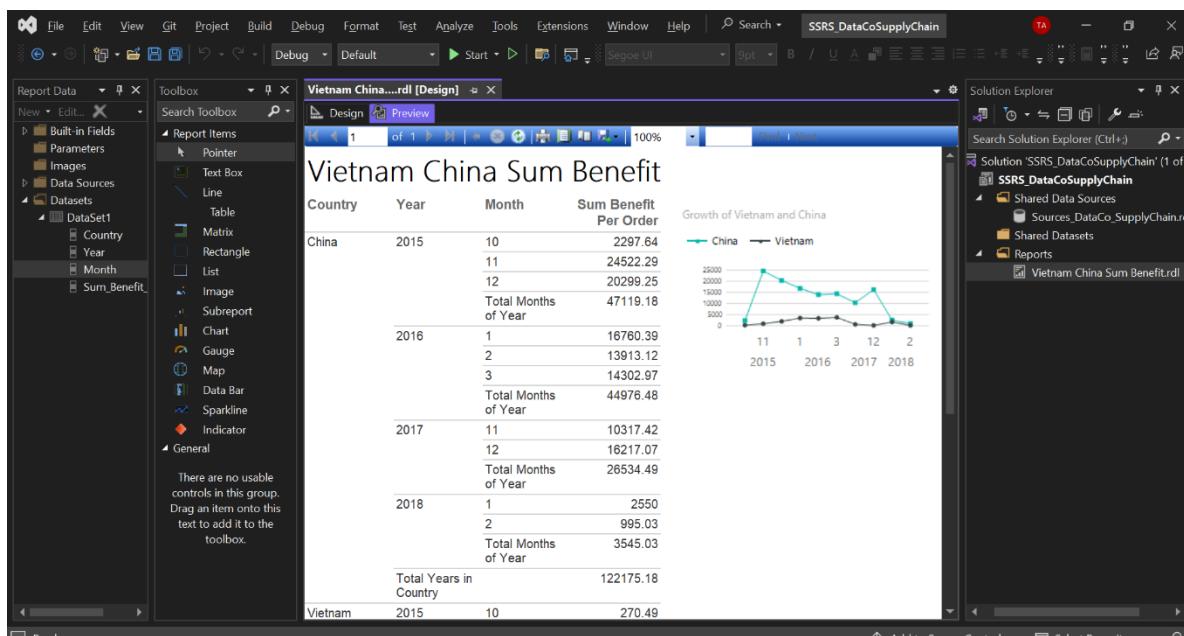
Bước 19: Double click vào chart để mở các câu hình cho chart.

- Lưu ý: Kéo các thanh trượt để hiển thị nhiều thông tin hơn.

Bước 20: Tiến hành cấu hình cho chart

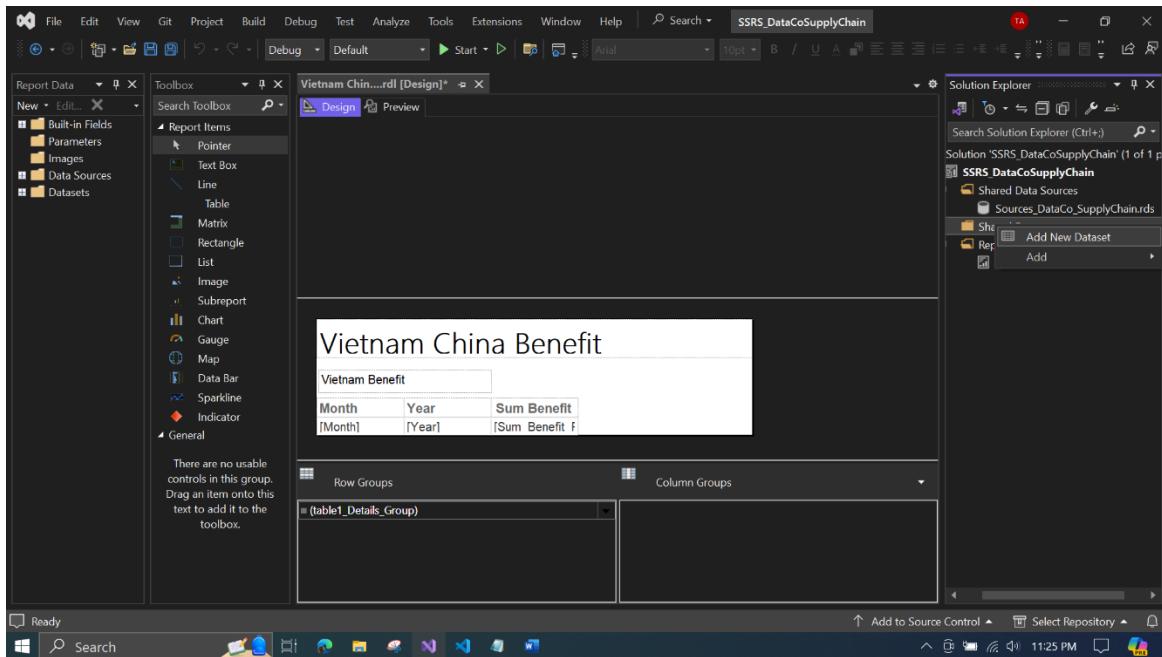


Kết quả

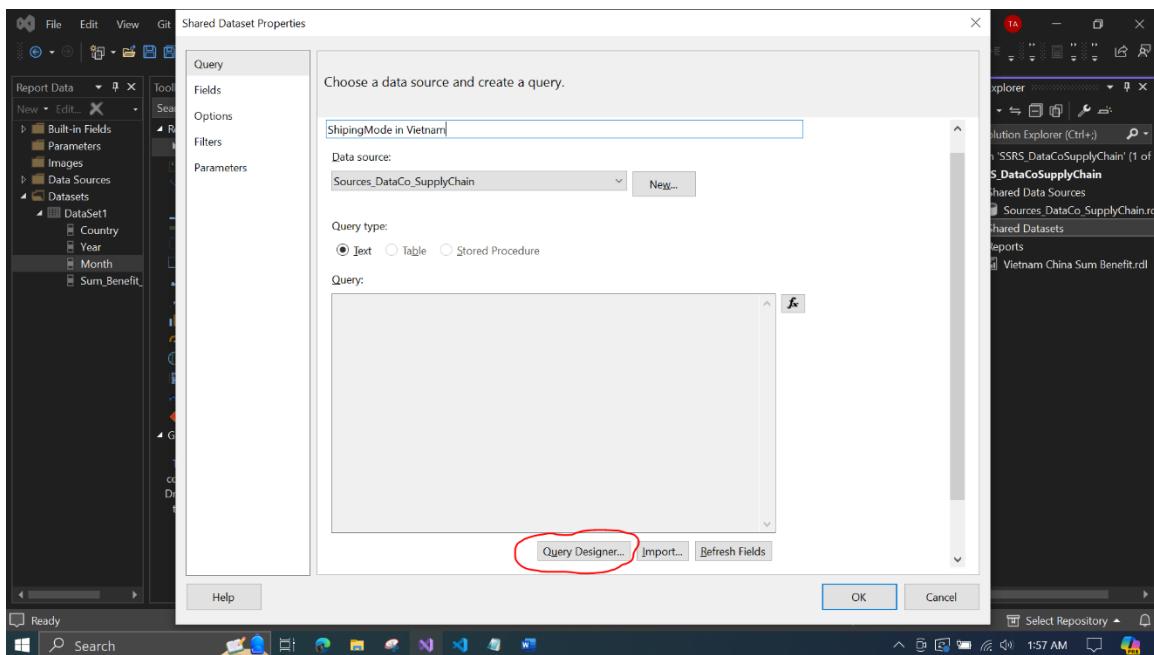


Tạo dataset

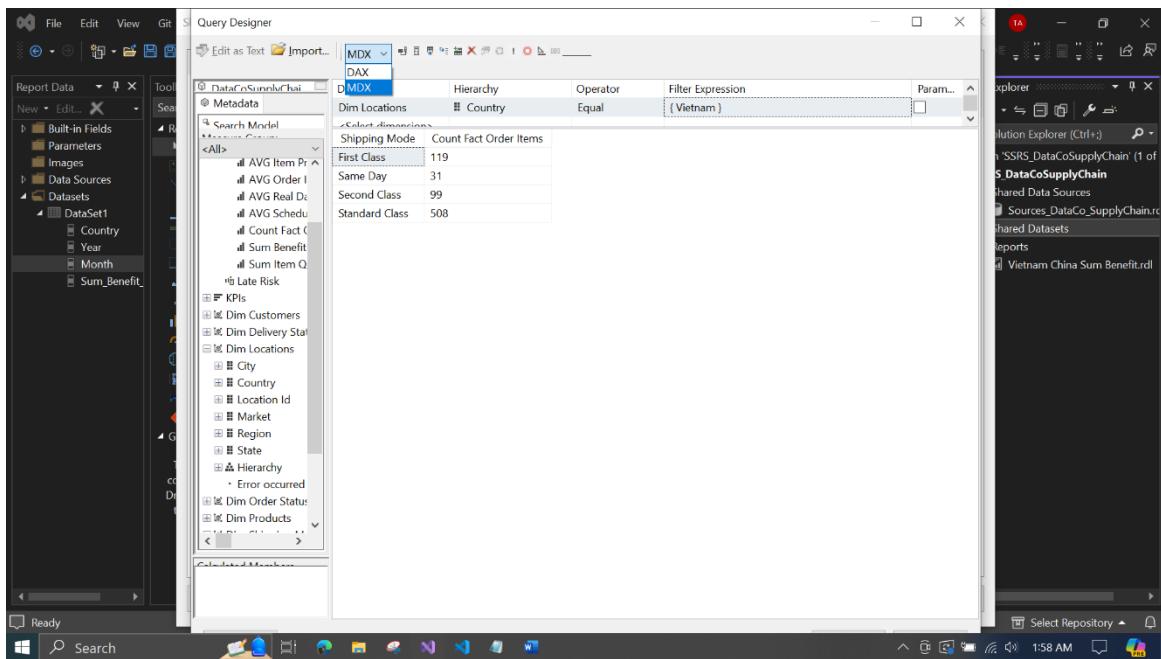
Bước 21: Bên trong “Solution Explorer”, chuột phải vào “Shared Datasets”. Chọn “Add New Datasets”.



Bước 22: Đặt tên và chọn các cấu hình cần thiết. Sau đó chọn “Query Designer” để tiến hành tạo dataset.



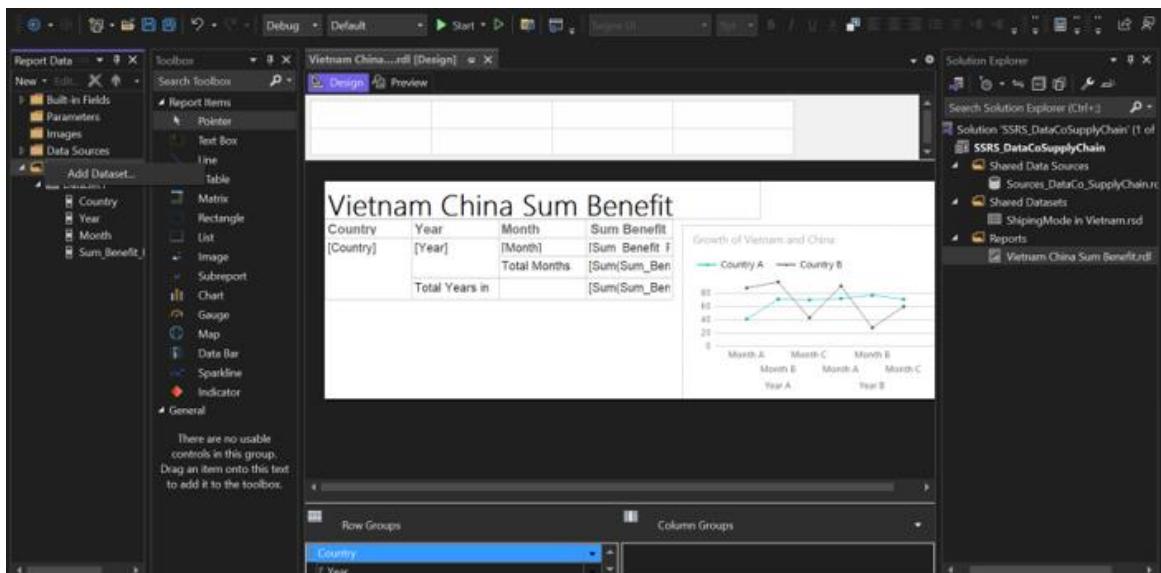
Bước 23: Chuyển sang MDX và truy vấn tương tự quá trình SSAS. Sau đó chọn “OK” để kết thúc.



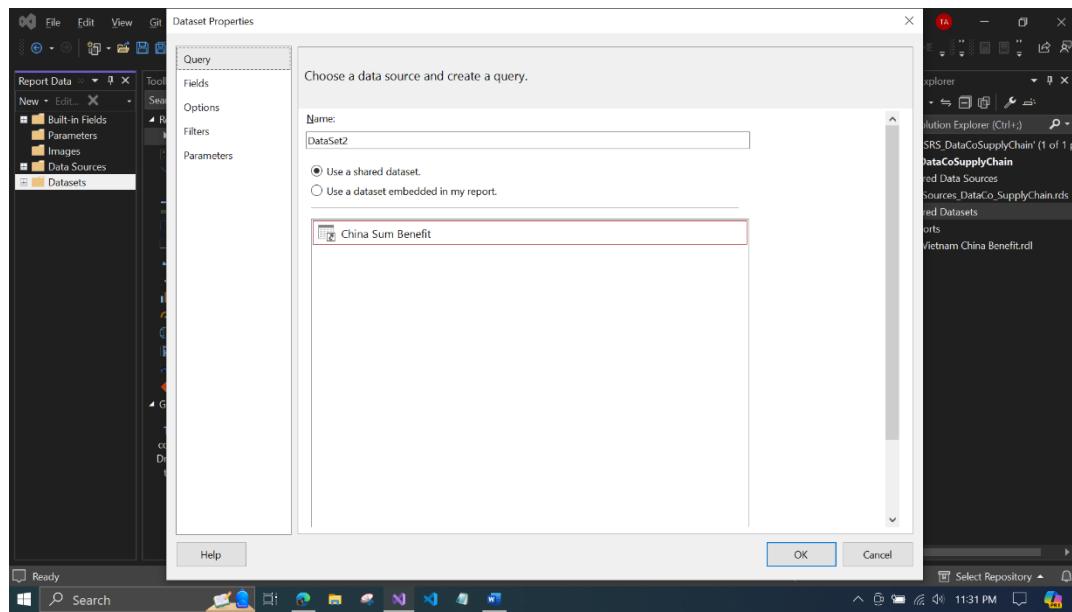
Bước 24: Chọn “OK” để kết thúc tạo dataset.

Thêm dataset vào trong report

Bước 25: Tại cửa sổ “Report Data” chuột phải vào Datasets → “Add Dataset” để thêm data cần thiết cho quá trình tạo report.

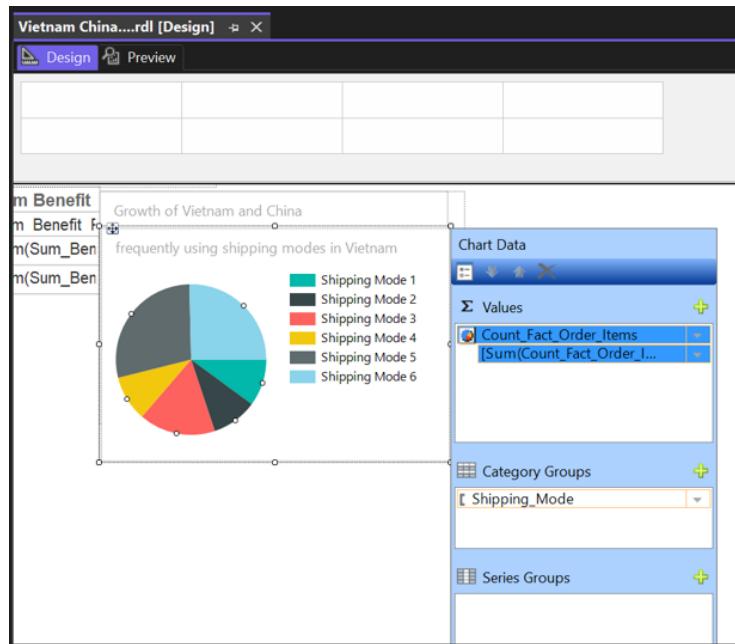


Bước 26: Đặt tên dataset. Tạo mới dataset hoặc thêm vào từ dataset đã tạo, chọn “OK” để kết thúc thêm dataset vào report.

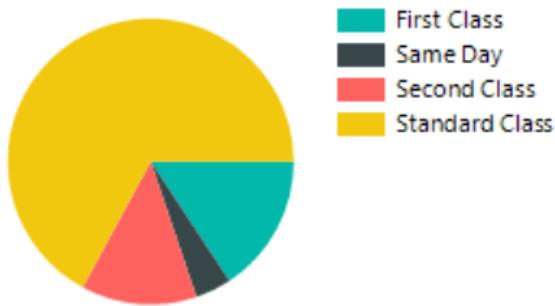


Bước 27: Dùng dataset đã thêm.

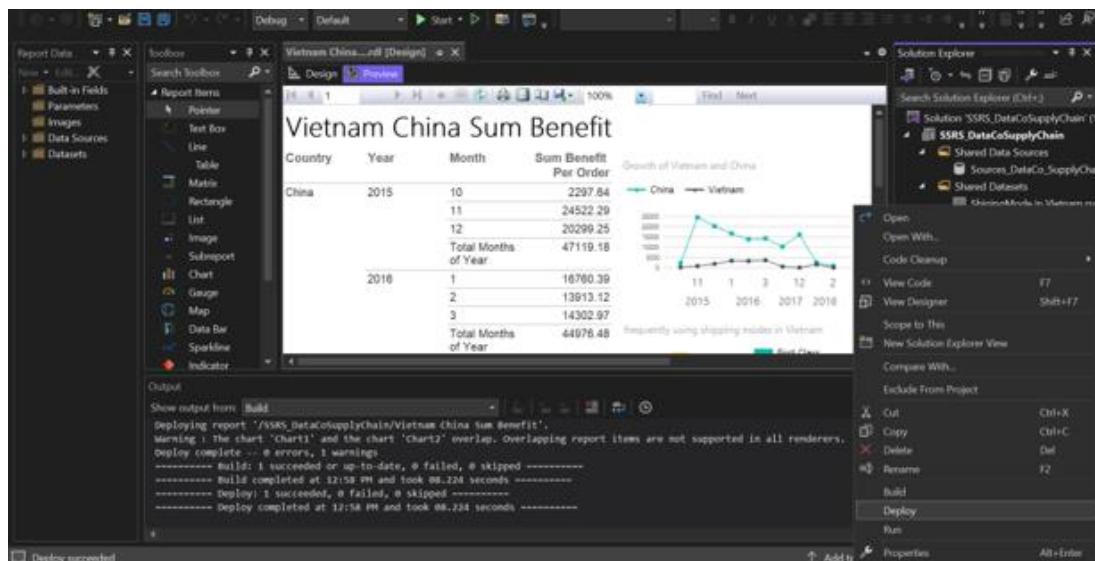
Ví dụ: dùng shipping mode in Vietnam để tạo biểu đồ tròn hiển thị tỷ lệ sử dụng các loại shipping tại Việt Nam



frequently using shipping modes in Vietnam



Bước 28: Deploy report lên “Report server” bằng cách nhấn start (hoặc chọn phải vào process report).



Bước 29: Truy cập vào localhost để xem các report

```

Output
Show output from: Build
Build started at 12:57 PM...
----- Build started: Project: SSRS_DataCoSupplyChain, Configuration: Debug -----
Skipping 'Sources_DataCo_SupplyChain.rds'. Item is up to date.
Skipping 'ShippingMode in Vietnam.rsd'. Item is up to date.
Skipping 'Vietnam China sum Benefit.rdl'. Item is up to date.
Build complete -- 0 errors, 0 warnings
----- Deploy started: Project: SSRS_DataCoSupplyChain, Configuration: Debug -----
Deploying to http://localhost/ReportServer.
Deploying data source 'Data Sources/Sources DataCo_SupplyChain'.
Deploying data set '/Datasets/ShippingMode in Vietnam'.
Deploying report '/SSRS_DataCoSupplyChain/Vietnam China Sum Benefit'.
Warning : The chart 'Chart1' and the chart 'Chart2' overlap. Overlapping report items are not supported in all renderers.
Deploy complete -- 0 errors, 1 warnings
----- Build: 1 succeeded or up-to-date, 0 failed, 0 skipped -----
----- Build completed at 12:58 PM and took 08.224 seconds -----
----- Deploy: 1 succeeded, 0 failed, 0 skipped -----
----- Deploy completed at 12:58 PM and took 08.224 seconds -----

```

- Lưu ý: có thể phát sinh lỗi nếu người dùng chưa cài đặt Report Server (là 1 server tương tự như SSAS)

The screenshot shows two browser windows. The top window displays the Microsoft SQL Server Reporting Services directory structure at localhost/ReportServer, showing items like Data_Sources, Datasets, and SSRS_DataCoSupplyChain. The bottom window shows a report titled "Vietnam China Sum Benefit" with a table of data and two charts: a line chart of "Growth of Vietnam and China" and a pie chart of shipping modes in Vietnam.

Country	Year	Month	Sum Benefit Per Order
China	2015	10	2297.64
		11	24522.29
		12	20299.25
		Total Months of Year	47119.18
2016	1	16760.39	
	2	13913.12	
	3	14302.97	
		Total Months of Year	44976.48
2017	11	10317.42	
	12	16217.07	
		Total Months of Year	26534.49
2018	1	2550	
	2	995.03	
		Total Months of Year	3545.03
	Total Years in Country	122175.18	
Vietnam	2015	10	270.49
		11	920.01
		12	1982.86

Growth of Vietnam and China

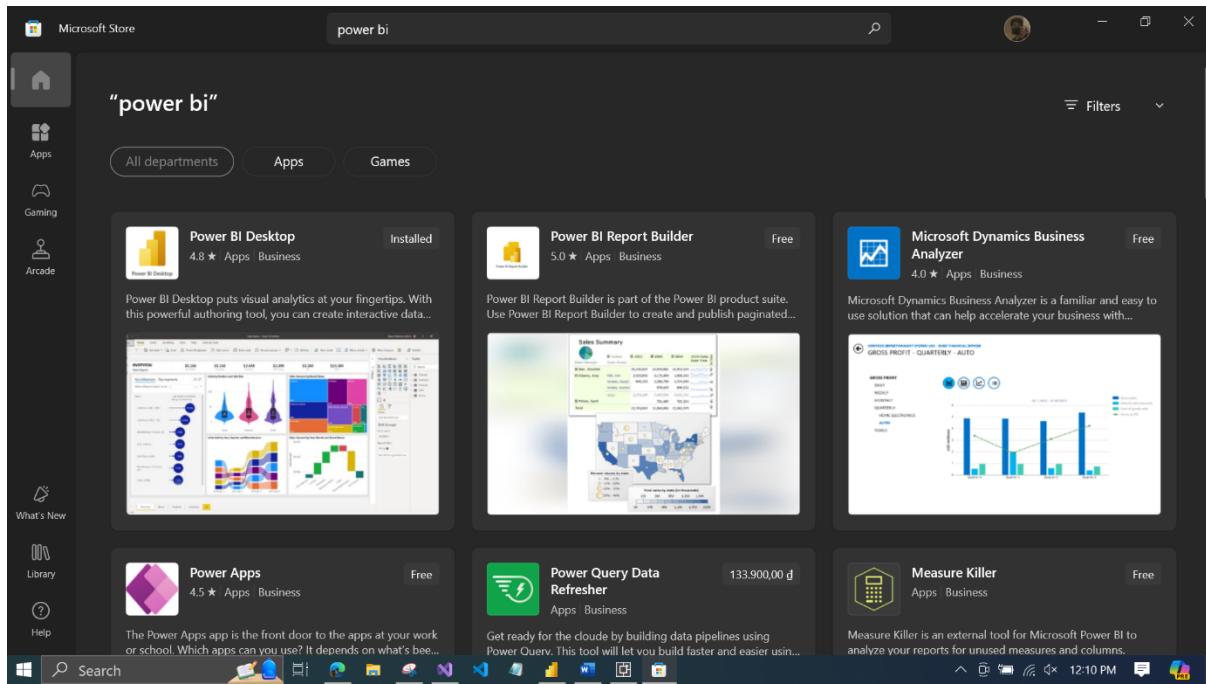
The line chart tracks the "Sum Benefit Per Order" over time. The Y-axis ranges from 0 to 35000. The X-axis shows months from November 2015 to February 2018. The China series (blue line) starts at approximately 25000 in Nov 2015, peaks at about 30000 in Jan 2016, and then fluctuates between 10000 and 20000. The Vietnam series (black line) starts at 0 in Nov 2015, remains low until Jan 2016, then rises steadily to about 10000 by Feb 2018.

frequently using shipping modes in Vietnam

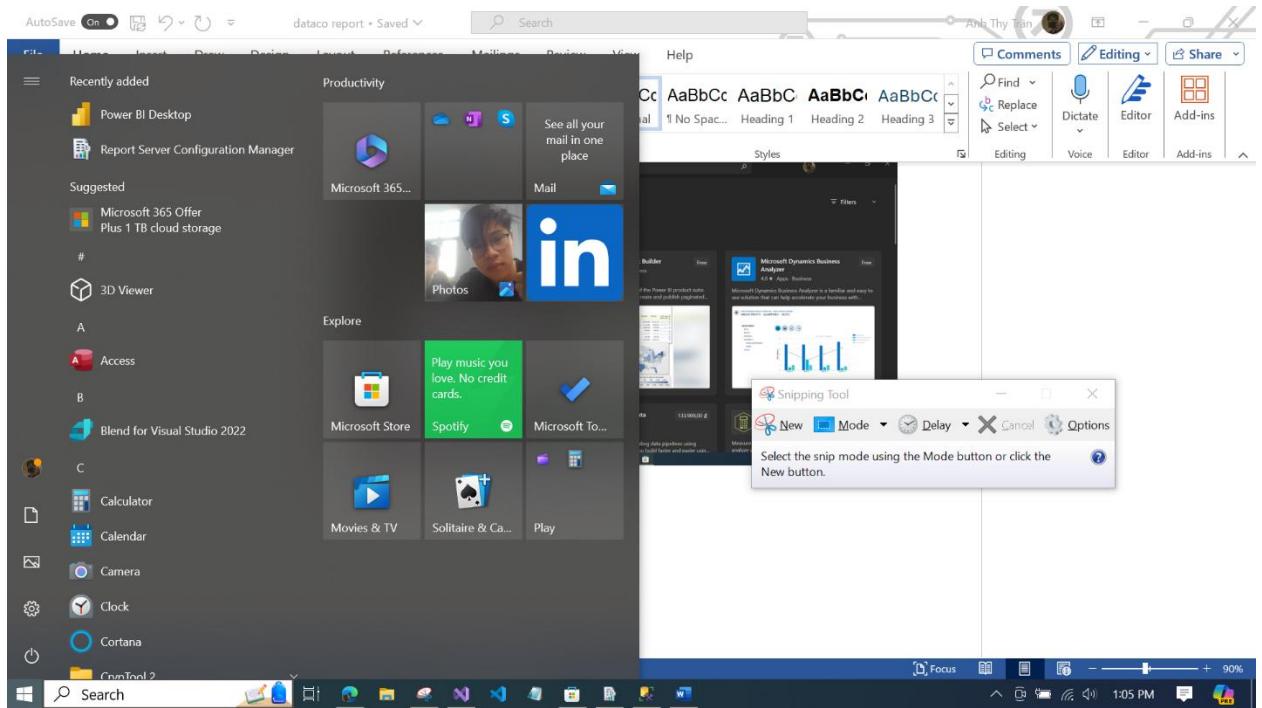
The pie chart illustrates the proportion of different shipping classes used in Vietnam. The segments are: Standard Class (yellow, largest), Second Class (red), First Class (teal), and Same Day (black).

Report mẫu bằng Power BI

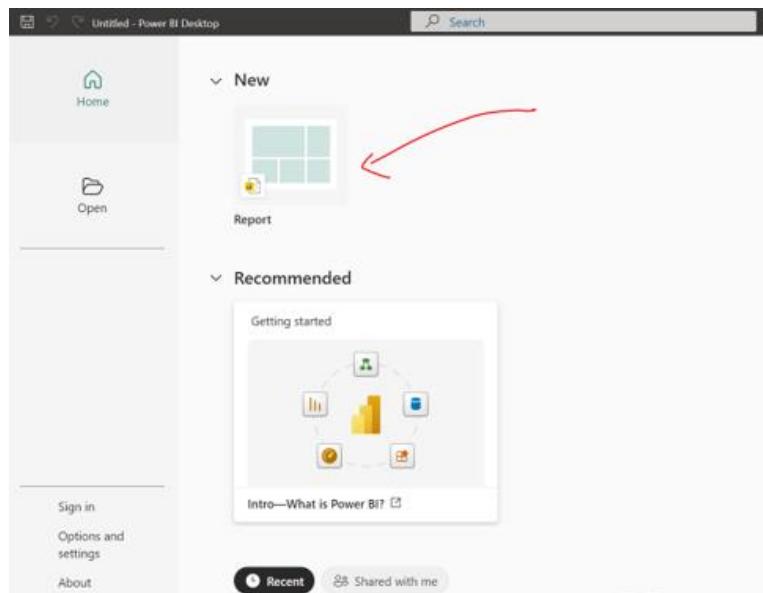
Cài đặt Power BI Desktop: truy cập Microsoft Store về Power BI. Tiến hành cài đặt Power BI Desktop.



Bước 1: mở Power BI Desktop.



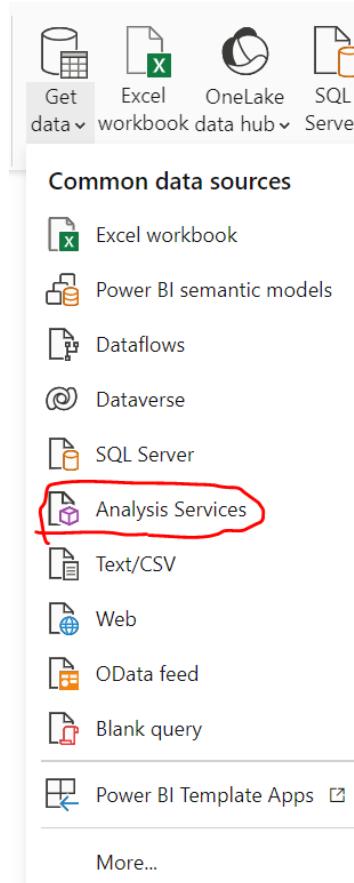
Bước 2: Chọn “New” để tạo một report trống mới.



Bước 3: Chọn Icon phía trên hoặc chữ “Get data” để thêm data sources vào Power BI.

The screenshot shows the Microsoft Power BI Desktop interface. The top navigation bar includes 'File', 'Home' (selected), 'Insert', 'Modeling', 'View', 'Optimize', and 'Help'. The 'Home' tab has sections for 'New' (Report icon) and 'Recommended' (Getting started). The 'Data' tab in the ribbon is selected, with icons for 'Clipboard', 'Get data', 'Excel', 'OneLake', 'SQL Server', 'Enter Data', 'Recent sources', 'Transform data', 'Refresh', 'Queries', 'New visual', 'Text box', 'More visuals', 'Insert', 'Calculations', 'Quick measure', 'Sensitivity', 'Share', 'Publish', and 'Copilot'. A red box highlights the 'Get data' icon. The main area shows the 'Get Data' dialog box with a sidebar for 'All' data sources and a list of options including 'Excel Workbook', 'Text/CSV', 'XML', 'JSON', 'Folder', 'PDF', 'Parquet', 'SharePoint folder', 'SQL Server database', 'Access database', 'SQL Server Analysis Services database' (highlighted with a red box), 'Oracle database', 'IBM Db2 database', 'IBM Informix database (Beta)', 'IBM Netezza', and 'MySQL database'. At the bottom are buttons for 'Certified Connectors', 'Template Apps', 'Connect' (green), and 'Cancel'.

Hoặc



Bước 4: Nhập bằng tay các thông số cho connection. Sau đó chọn “OK”.

SQL Server Analysis Services database

Server ⓘ

DESKTOP-OEFO0DK

Database (optional)

SSAS_DataCoSupplyChain_Ver4

Import

Connect live

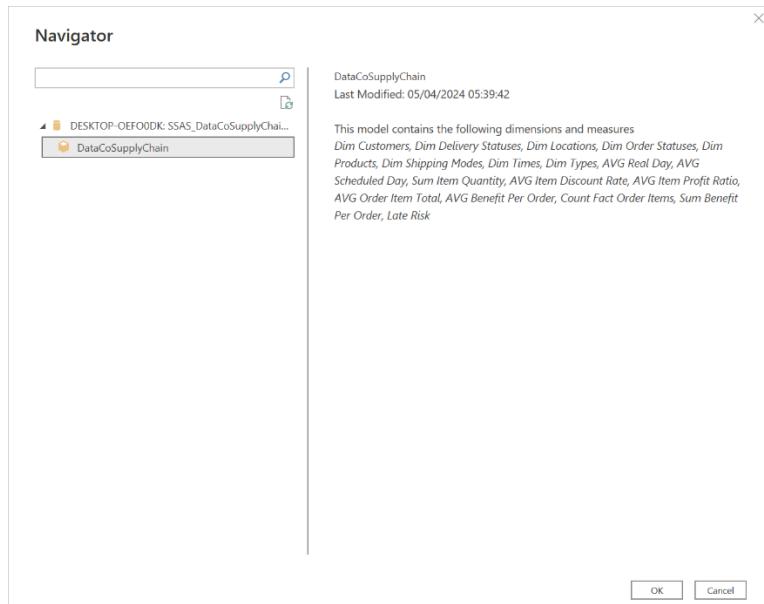
▷ MDX or DAX query (optional)

OK

Cancel

- Lưu ý: đây là server của SSAS không phải server của MS SQL.

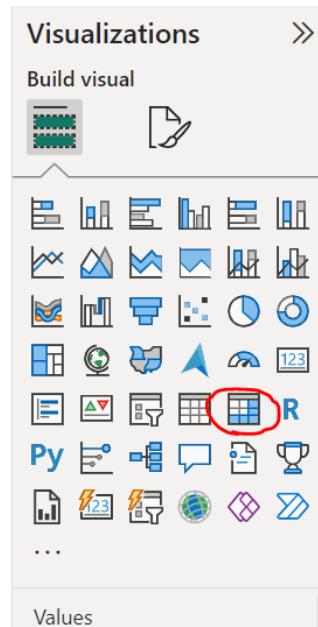
Bước 5: Chọn Cube để lấy dữ liệu, sau đó chọn “OK” để kết thúc.



Tạo Visualizations

Tạo matrix (bảng gồm group và total)

Bước 6: Chọn loại visualization là matrix (matrix là loại bảng có chiều, có thể hiểu là đã được group, còn table là bảng thông thường)



- So sánh matrix và table

Matrix	Table
--------	-------

Country	Sum Benefit Per Order
China	122,175.18
2015	47,119.18
10	2,397.64
11	24,522.28
12	30,399.25
2016	44,976.48
2017	26,534.49
2018	3,545.03
Vietnam	16,448.96
2015	3,173.36
10	270.49
Total	138,624.14

Country	Year	Month	Sum Benefit Per Order
China	2017	11	10,317.42
China	2017	12	16,217.07
China	2018	1	2,395.00
China	2018	2	399.63
Vietnam	2017	10	377.49
Vietnam	2017	11	920.01
Vietnam	2017	12	1,602.66
Vietnam	2018	1	3,685.55
Vietnam	2018	2	3,314.29
Vietnam	2018	3	3,733.75
Vietnam	2017	11	668.63
Vietnam	2017	12	1,688.62
Vietnam	2018	1	1,700.15
Vietnam	2018	2	1,688.62
Total			138,624.14

Bước 7: Click chuột vào Table cần làm, sau đó chọn các thuộc tính cần thiết, filter theo nhu cầu, ...

Country	Sum Benefit Per Order
China	122,175.18
2015	47,119.18
10	2,397.64
11	24,522.28
12	30,399.25
2016	44,976.48
2017	26,534.49
2018	3,545.03
Vietnam	16,448.96
2015	3,173.36
10	270.49
Total	138,624.14

- Chọn độ đo là tổng doanh thu.

The screenshot shows the Power BI Desktop interface with a table visual titled "Country Sum Benefit Per Order". The table contains the following data:

Country	Sum Benefit Per Order
China	122,175.18
Vietnam	16,448.96
Total	138,624.14

The "Filters" pane on the right shows a filter for "Country" set to "is Vietnam or China". Under "Basic filtering", "China" and "Vietnam" are selected. The "Data" pane on the right lists various dimensions and measures, with "Dim Locations" expanded to show "Country", "Hierarchy", "Region", and "State".

- Chọn country, lọc để chọn Việt Nam và Trung Quốc.

The screenshot shows the Power BI Desktop interface with a table visual titled "Country Sum Benefit Per Order". The table contains the same data as the previous screenshot:

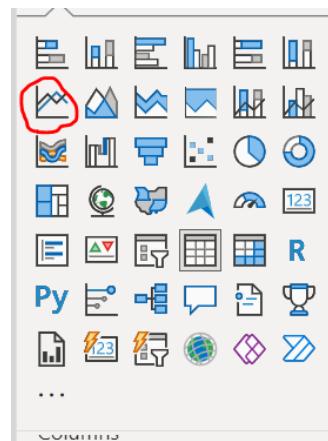
Country	Sum Benefit Per Order
China	122,175.18
Vietnam	16,448.96
Total	138,624.14

The "Filters" pane on the right shows a filter for "Country" set to "is China or Vietnam". Under "Basic filtering", "Select all" is checked. The "Data" pane on the right shows "Dim Hierarchy" selected under "Dim Times", with "Year", "Month", "Day", "Hour", and "Minute" listed. Other dimensions like "Dim Products" and "Dim Shipping Modes" are also visible.

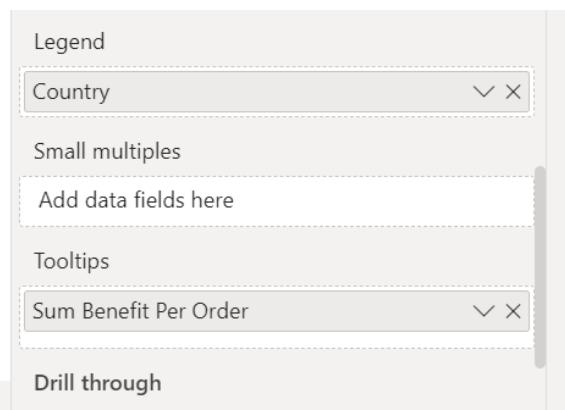
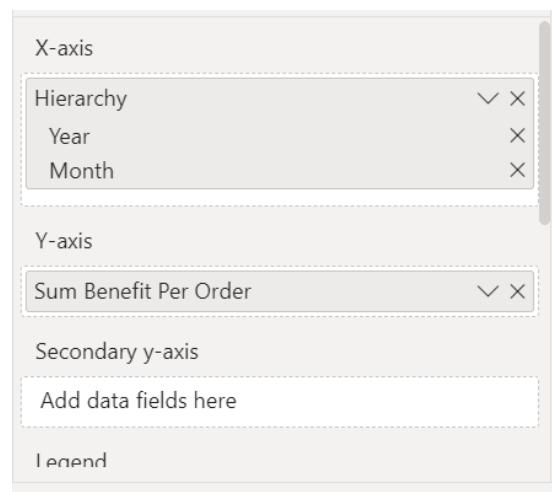
- Chọn Month và Year bên trong Hierarchy vì chúng cần phân cấp theo matrix.

Tạo Line chart

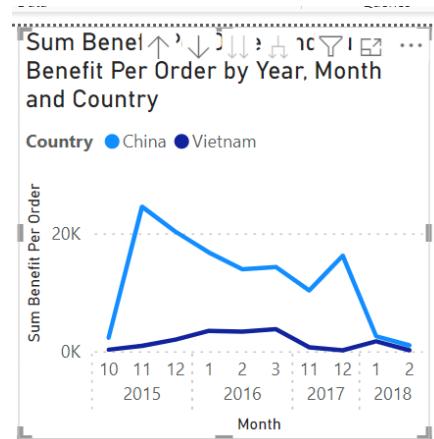
Bước 8: Chọn Line chart



Bước 9: Click chọn vào line chart vừa được tạo, cấu hình cho line chart

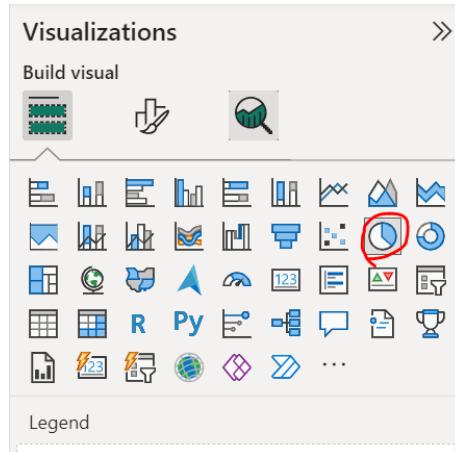


Kết quả

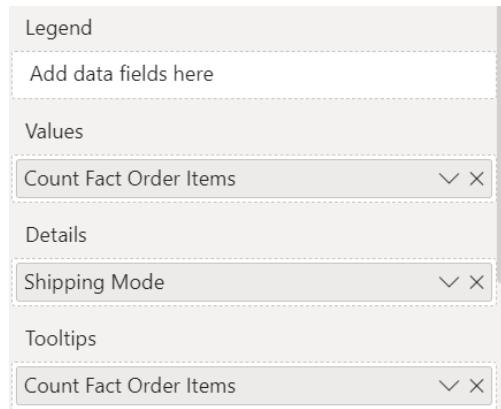


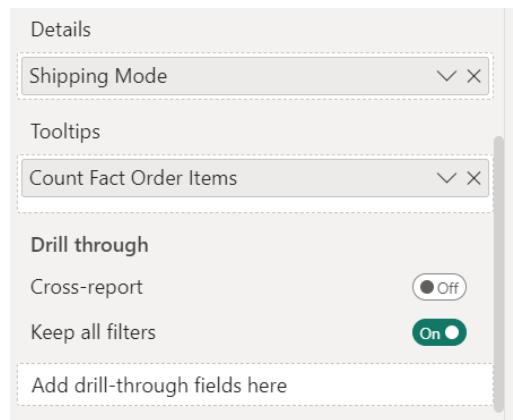
Tạo Pie chart thống kê phần trăm dùng các shipping modes

Bước 10: Chọn Pie chart.

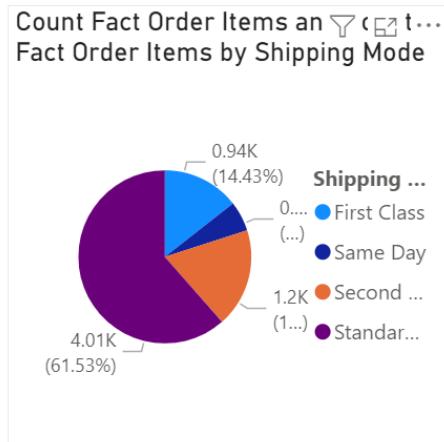


Bước 11: Click chọn vào pie chart vừa được tạo, cấu hình cho pie chart.





Kết quả



4.5. Các Reports

4.5.1. Report Builder

4.5.1.1. Vietnam China Benefit

The Report Builder interface shows a report titled "Vietnam China Sum Benefit". The report includes the following components:

- Table:** A table showing the "Sum Benefit Per Order" for China and Vietnam from 2015 to 2018, broken down by month.
- Line Chart:** A line chart titled "Growth of Vietnam and China" comparing the growth of China (blue line) and Vietnam (red line) over time.
- Pie Chart:** A pie chart titled "frequently using shipping modes in Vietnam" showing the distribution of shipping modes.

Country	Year	Month	Sum Benefit Per Order
China	2015	10	2297.64
		11	24522.29
		12	20299.25
		Total Months of Year	47119.18
2016	1	16760.39	
	2	13913.12	
	3	14302.97	
		Total Months of Year	44976.48
2017	11	10317.42	
	12	16217.07	
		Total Months of Year	26534.49
		Total Years in Country	122175.18
2018	1	2550	
	2	995.03	
		Total Months of Year	3545.03
		Total Years in Country	122175.18

4.5.1.2. Revenue of the markets

The screenshot shows the Microsoft Visual Studio interface with the following details:

- File, Edit, View, Git, Project, Build, Debug, Format, Test, Analyze, Tools, Extensions, Window, Help** menu items.
- SSRS DataCoSupplyChain** tab in the top right corner.
- Toolbox** on the left side, expanded to show **Report Items**. The **Pointer** item is selected. Other items include Text Box, Line, Table, Matrix, Rectangle, List, Image, Subreport, Chart, Gauge, Map, Data Bar, Sparkline, Indicator, and General.
- Report Data** ribbon tab.
- Search Toolbox** ribbon tab.
- Design** and **Preview** tabs are visible at the top of the main workspace.
- Revenue of the markets.rdl [Design]** is the active report.
- Market** and **Sum Benefit Per Order** columns are displayed in a table.
- Market's revenue** chart showing benefit per order for four regions: Africa, Europe, LATAM, and USCA.
- Solution Explorer** on the right shows the project structure with files like **SSRS DataCoSupplyChain**, **Shared Data Sources**, **Shared Datasets**, and **Reports**.
- Ready** status bar at the bottom.

4.5.1.3. Top sale each shipping mode through year

The screenshot shows the Microsoft Visual Studio interface with the following details:

- Title Bar:** SSRS_DataCoSupplyChain
- Toolbars:** Standard, Debug, Start, File, Edit, View, Project, Build, Format, Test, Analyze, Tools, Extensions, Window, Help.
- Solution Explorer:** Shows the solution 'SSRS_DataCoSupplyChain' with its projects, shared data sources, and reports.
- Toolbox:** Expanded to show 'Report Items' including Pointer, Text Box, Line, Table, Matrix, Rectangle, List, Image, Subreport, Chart, Gauge, Map, Data Bar, Sparkline, Indicator, and General.
- Report Designer:** Displays a chart titled 'mode through year' comparing shipping modes (First Class, Second Class, Same Day, Standard Class) across years (2015-2018). The chart uses stacked bars to show the count of fact order items for each mode per year.
- Data:** A table below the chart provides the total count of fact order items for each shipping mode and year.

Year	Shipping Mode	Count Fact Order Items
2015	First Class	9595
	Same Day	3315
	Second Class	12095
	Standard Class	36997
Total	62002	
2016	First Class	9821
	Same Day	3407
	Second Class	12197
	Standard Class	37177
Total	62602	
2017	First Class	8056
	Same Day	2894
	Second Class	10426
	Standard Class	32170
Total	53546	
2018	First Class	342
	Same Day	121
	Second Class	498
	Standard Class	1408
Total	2369	

4.5.2. Power BI

4.5.2.1. Vietnam China Benefit

The screenshot shows a Power BI desktop interface with the following details:

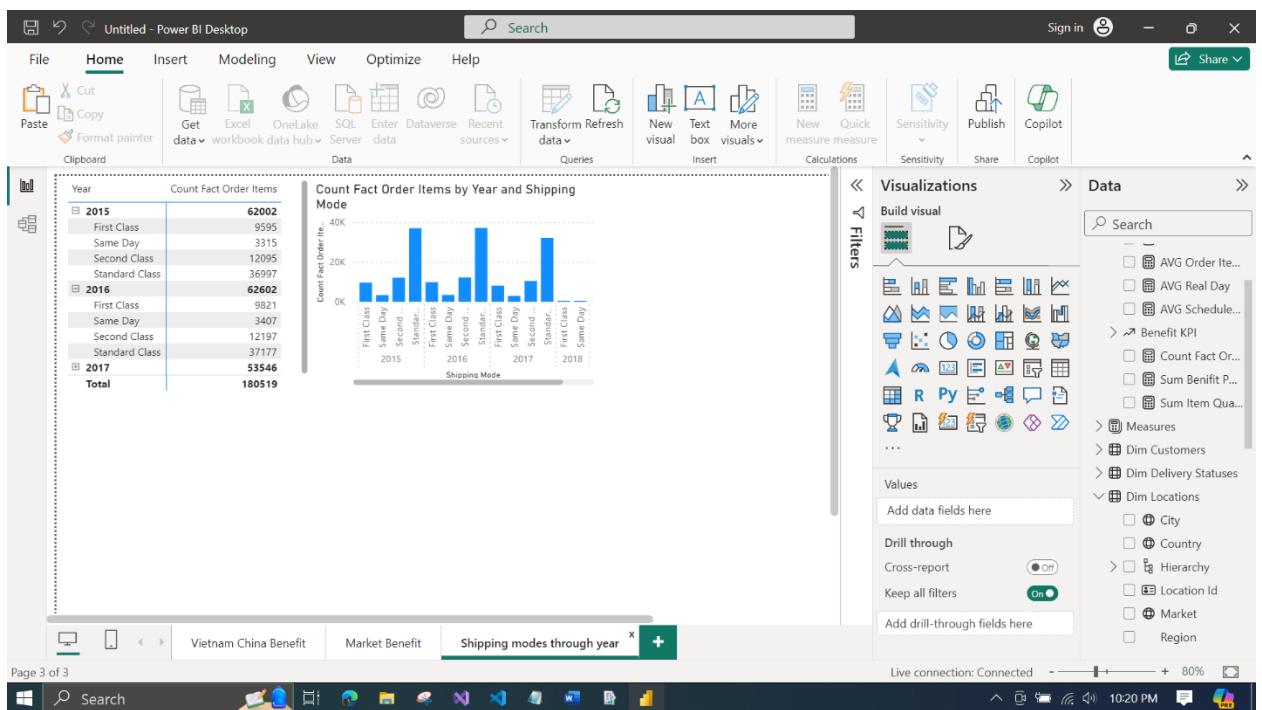
- Home tab selected:** Home, Insert, Modeling, View, Optimize, Help.
- Visualizations pane:** Shows various chart and table icons under 'Build visual'.
- Data pane:** Shows a list of measures and dimensions, including 'Benefit KPI' and 'Dim Locations'.
- Table:** 'Sum Benefit Per Order' by Country and Year.
- Line Chart:** 'Sum Benefit Per Order and Sum Benefit Per Order by Year, Month and Country'.
- Pie Chart:** 'Count Fact Order Items by Country and Shipping Mode'.
- Page navigation:** Page 1 of 3.
- System status:** Live connection: Connected, 80% zoom, 10:19 PM.

4.5.2.2. Revenue of the markets

The screenshot shows a Power BI desktop interface with the following details:

- Home tab selected:** Home, Insert, Modeling, View, Optimize, Help.
- Visualizations pane:** Shows various chart and table icons under 'Build visual'.
- Data pane:** Shows a list of measures and dimensions, including 'Benefit KPI' and 'Dim Locations'.
- Table:** 'Sum Benefit Per Order' by Market.
- Bar Chart:** 'Sum Benefit Per Order by Market'.
- Page navigation:** Page 2 of 3.
- System status:** Live connection: Connected, 80% zoom, 10:20 PM.

4.5.2.3. Top sale each shipping mode through year



CHƯƠNG 5: QUÁ TRÌNH KHAI THÁC DỮ LIỆU (DATA MINING)

5.1. Import các thư viện cần thiết

```
[1] import pandas as pd
2 giây   from sklearn.ensemble import RandomForestClassifier
   from sklearn.preprocessing import LabelEncoder
   from sklearn.model_selection import train_test_split
   from sklearn.tree import plot_tree, DecisionTreeClassifier, export_text
   import seaborn as sns
   import matplotlib.pyplot as plt
```

5.2. Đọc file dataset

```

[2] data_path='/content/Fact.csv'
giáy
[35] df = pd.read_csv(data_path,encoding='latin-1',thousands=',')
giáy
      fact_id  order_id   product_id order_date  ship_date  ship_mode customer_name          state  sales  quantity  discount  profit order_priority
0        1  AG-2011-2040  OFF-TEN-1000025  01/01/2011  06/01/2011  Standard Class  Toby Braunhardt  Constantine  408       2     0.1  106.1400  Medium
1        2  IN-2011-47883  OFF-SU-10000618  01/01/2011  08/01/2011  Standard Class  Joseph Holt  New South Wales  120       3     0.1  36.0360  Medium
2        3  HU-2011-1220  OFF-TEN-10001585  01/01/2011  05/01/2011  Second Class  Annie Thurman  Budapest  66       4     0.0  29.6400  High
3        4  IT-2011-3647632  OFF-PA-10001492  01/01/2011  05/01/2011  Second Class  Eugene Moren  Stockholm  45       3     0.5 -26.0550  High
4        5  IN-2011-47883  FUR-FU-10003447  01/01/2011  08/01/2011  Standard Class  Joseph Holt  New South Wales  114       5     0.1  37.7700  Medium
...
...
...
...
...
51285  51286  CA-2014-115427  OFF-BI-10002103  31/12/2014  04/01/2015  Standard Class  Erica Bern  California  14       2     0.2  4.5188  Medium
51286  51287  MO-2014-2560  OFF-WIL-10001069  31/12/2014  05/01/2015  Standard Class  Liz Preis  Sousse-Massa-Draâa  4       1     0.0  0.4200  Medium
51287  51288  MX-2014-110527  OFF-LA-10004182  31/12/2014  02/01/2015  Second Class  Charlotte Melton  Managua  26       3     0.0  12.3600  Medium
51288  51289  MX-2014-114783  OFF-LA-10000413  31/12/2014  05/01/2015  Standard Class  Tamara Dahlén  Chihuahua  7       1     0.0  0.5600  Medium
51289  51290  CA-2014-156720  OFF-FA-10003472  31/12/2014  04/01/2015  Standard Class  Jill Matthias  Colorado  3       3     0.2 -0.6048  Medium
51290 rows x 13 columns

```

5.3. Kiểm tra giá trị null

```

df.isna().sum()
fact_id      0
order_id      0
product_id    0
order_date    0
ship_date     0
ship_mode     0
customer_name 0
state         0
sales         0
quantity      0
discount      0
profit        0
order_priority 0
dtype: int64

```

Kết quả trả về là bộ dữ liệu đầy đủ, không có giá trị bị thiếu.

5.4. Thay đổi dữ liệu và chọn ra những cột cần thiết

Tạo index cho các hàng trong bảng.

```

df.reset_index(names="fact_id", inplace=True)
✓ 0.0s

```

Ánh xạ các giá trị text sang số (enum) để có thể thuật toán có thể hiểu và đưa ra kết quả.

```

✓ [6] df['order_priority'] = df['order_priority'].replace({'Low' : 1,'Medium':2,'High':3,'Critical':4})
giây

✓ [15] features=['sales', 'quantity', 'discount', 'profit', 'order_priority']
0
giây
    target = ['ship_mode']
    X = df[features]
    y = df[target]

```

Với cột order_priority dùng phương pháp LabelEncoder để thể hiện đặc trưng thứ tự ưu tiên của từng loại, tăng dần từ Low đến Critical .

5.5. Chia data test và train theo tỉ lệ 7:3

```

# Chia train test theo tỉ lệ 7:3
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

```

5.6. Khởi chạy và đánh giá thuật toán

- Thuật toán Random Forest:

```

rf_classifier = RandomForestClassifier()
rf_classifier.fit(X_train, y_train)
[23]   ✓ 21.0s
...
c:\Users\quang\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\base.py:1152:
    return fit_method(estimator, *args, **kwargs)

...
    RandomForestClassifier
RandomForestClassifier()

```

```

✓ [34] accuracy = rf_classifier.score(X_test, y_test)
giây
    print("Accuracy:", accuracy*100, '%')

```

Accuracy: 56.38526028465588 %

```

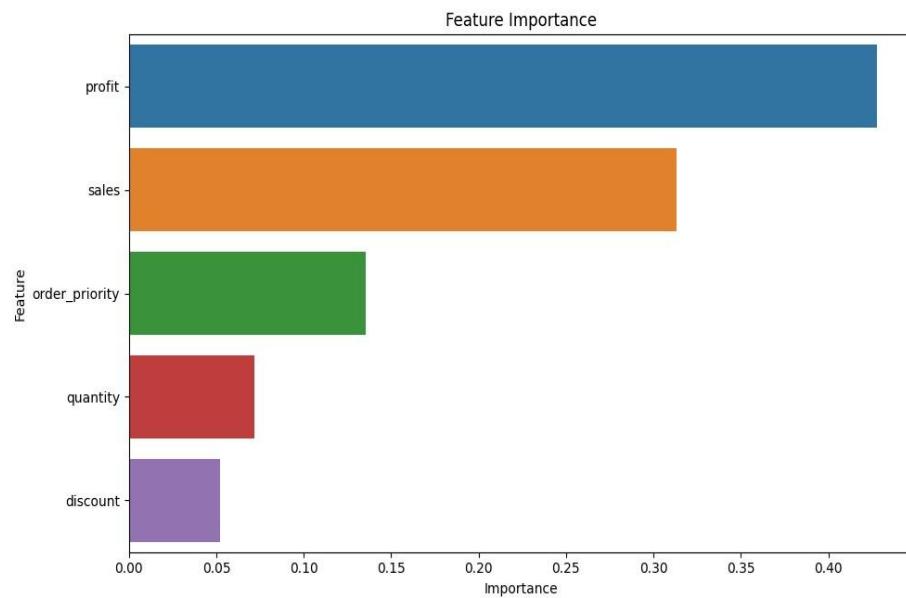
# Get the feature importances
importances = rf_classifier.feature_importances_
feature_names = X_train.columns

# Create a DataFrame with feature importances
feature_importances = pd.DataFrame({'Feature': feature_names, 'Importance': importances})

# Sort the DataFrame by importance score in descending order
feature_importances = feature_importances.sort_values('Importance', ascending=False)

# Plot the feature importances using Seaborn
plt.figure(figsize=(10, 6))
sns.barplot(data=feature_importances, x='Importance', y='Feature')
plt.xlabel('Importance')
plt.ylabel('Feature')
plt.title('Feature Importance')
plt.tight_layout()
plt.show()

```



Sau khi sử dụng thuật toán Random Forest, nhóm có kết quả về các thứ tự quan trọng của các đặc trưng ảnh hưởng đến phân loại. Kết quả ta thấy rằng việc khách hàng chọn loại vận chuyển phụ thuộc phần lớn vào lợi nhuận của đơn hàng đó mang lại cho cửa hàng.

- Thuật toán Decision Tree:

```

✓ [20] dtree_classifier = DecisionTreeClassifier(max_depth = 3)
        dtree_classifier.fit(X_train, y_train)

        □ DecisionTreeClassifier
        DecisionTreeClassifier(max_depth=3)

```

```

✓ [21] accuracy = dtree_classifier.score(X_test, y_test)
        print("Accuracy:", accuracy*100, '%')

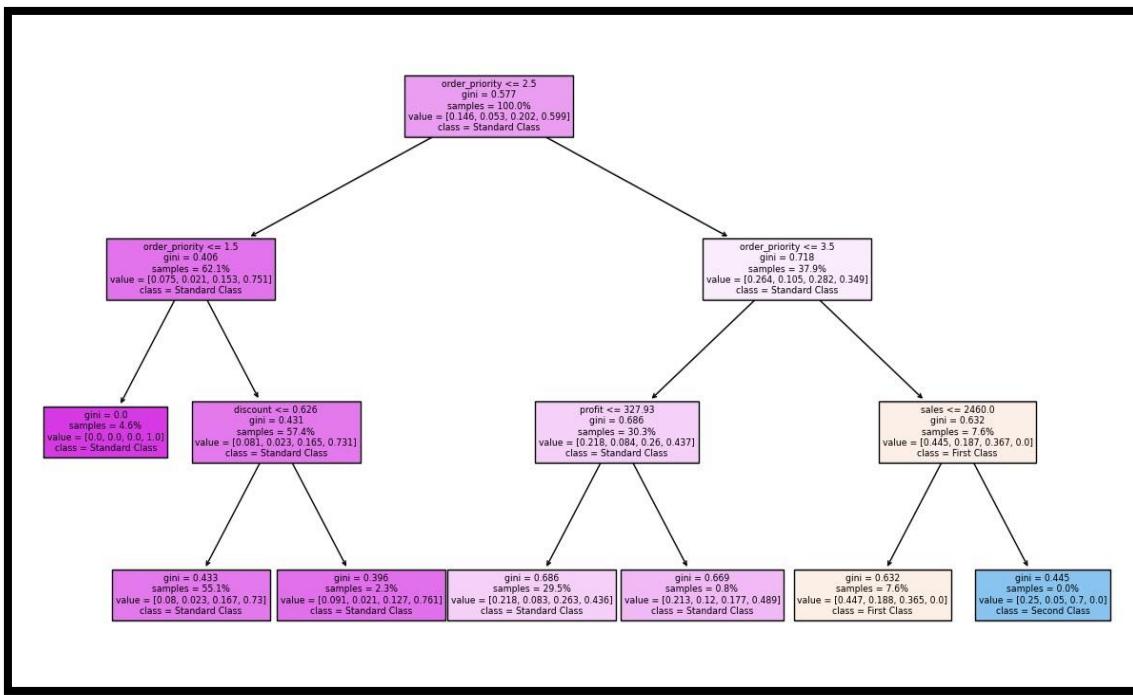
Accuracy: 63.43666731656593 %

```

```

✓ [22] X_train.columns
        Index(['sales', 'quantity', 'discount', 'profit', 'order_priority'], dtype='object')

```



Sơ đồ cây thể hiện quy luật phân lớp của bộ dữ liệu như sau:

- Nhánh bên phải $order_priority > 2.5$
 - + Khi $order_priority \leq 3.5$ và $profit \leq 327.93$ hoặc > 327.93 thì sẽ vào nhóm Standard Class

- + Khi $\text{order_priority} \leq 3.5$ và $\text{sales} \leq 2460$ sẽ vào nhóm First Class
- + Khi $\text{order_priority} \leq 3.5$ và $\text{sales} > 2460$ sẽ vào nhóm Second Class
- Nhánh bên trái $\text{order_priority} \leq 2.5$
 - + Khi $\text{order_priority} \leq 1.5$ sẽ vào nhóm Standard Class
 - + Khi $\text{order_priority} > 1.5$ và $\text{discount} \leq 0.63$ hoặc > 0.63 sẽ vào Standard Class

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] GeeksforGeeks, "Difference between Star Schema and Snowflake Schema," 21 02 2023. [Online]. Available:
<https://www.geeksforgeeks.org/differencebetween-star-schema-and-snowflake-schema/>. [Accessed 10 04 2023].
- [2] L. Breiman, "Random Forests," 10 2001. [Online]. Available:
<https://link.springer.com/article/10.1023/a:1010933404324>. [Accessed 12 06 2023].
- [3] V. K. Xindong Wu, "CART: Classification and Regression Trees," in *The Top Ten Algorithms in Data Mining*, CRC Press Taylor&Francis Group, 2009, pp. 179-183.
- [4] phamdinhkhanh, "Bài 46 - Đánh giá mô hình phân loại trong ML," 13 08 2020. [Online]. Available:
<https://phamdinhkhanh.github.io/2020/08/13/ModelMetric.html>. [Accessed 18 06 2023].

- [5] G. M. L. Education, "Classification: Accuracy," 19 07 2022. [Online]. Available: <https://developers.google.com/machine-learning/crashcourse/classification/accuracy>. [Accessed 20 06 2023].
- [6] G. M. L. Education, "Classification: Precision and Recall," 19 07 2022. [Online]. Available: <https://developers.google.com/machinelearning/crash-course/classification/precision-and-recall>. [Accessed 21 06 2023].
- [7] G. M. L. Education, "Classification: ROC Curve and AUC," 19 07 2022. [Online]. Available: <https://developers.google.com/machinelearning/crash-course/classification/roc-and-auc>. [Accessed 20 06 2023].
- [8] L. Breiman, "SpringerLink," 10 2021. [Online]. Available: <https://link.springer.com/article/10.1023/a:1010933404324>. [Accessed 15 06 2023].