

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

ĐCCT * ĐCCT



KHO DỮ LIỆU VÀ OLAP

**Đề Tài: PHÂN TÍCH VÀ THIẾT KẾ KHO DỮ LIỆU VỀ
DOANH THU VÀ LỢI NHUẬN CỦA SIÊU THỊ**

GVHD: ThS. Nguyễn Thị Kim Phụng

Lớp: IS217.M21.HTCL

Sinh viên thực hiện: Huỳnh Hữu Hào – 19520521

Tp. Hồ Chí Minh – 6/2022

[illegible]

MỤC LỤC

1.	TỔNG QUAN ĐỀ TÀI.....	1
1.1.	Mô tả dữ liệu gốc.....	1
1.2.	Mô tả dữ liệu đã xử lý	2
1.3.	Thiết kế kho dữ liệu.....	3
1.3.1.	Lược đồ hình sao	3
1.3.2.	Bảng Fact.....	3
1.3.3.	Bảng Dim.....	4
2.	XÂY DỰNG KHO DỮ LIỆU (SSIS)	5
2.1.	Tạo Database	5
2.2.	Tạo Project và thiết lập kết nối.....	5
2.3.	Tạo Data Flow Task	8
2.4.	Tạo Flag File Source kết nối với file dữ liệu.....	9
2.5.	Rẽ nhánh và lọc data trùng	11
2.6.	Tạo các bảng Dimension	12
2.7.	Tạo bảng Fact	14
2.8.	Tạo Execute SQL Task khóa ngoại cho bảng Fact và các bảng Dim	15
2.9.	Tạo bảng Dim_Time và các ràng buộc	17
2.10.	Tạo Execute SQL Task để xóa ràng buộc, xóa bảng Fact và các bảng Dim ...	17
2.11.	Thực thi project	18
3.	PHÂN TÍCH KHO DỮ LIỆU (SSAS)	19
2.12.	Tạo Project.....	19
2.13.	Tạo Data Source	19
2.14.	Tạo Data Source View	21

2.15. Tạo Cube	23
2.16. Tạo thêm Measure	24
2.17. Chỉnh sửa thuộc tính các bảng Dim	25
2.18. Thực thi project	27
2.19. Truy vấn dữ liệu	28
2.19.1. Truy vấn loại Roll Up: tính tổng lợi nhuận qua các năm.....	28
2.19.2. Truy vấn loại Drill Down: tính tổng lợi nhuận của từng thành phố, từng bang. 29	
2.19.3. Truy vấn loại Slice and Dice: hình thức giao hàng phổ biến nhất ở Texas.29	
2.19.4. Truy vấn loại Pivot: số đơn hàng của từng bang qua các năm.	29
2.20. Truy vấn MDX	30
2.20.1. Tổng số lợi nhuận theo từng năm	30
2.20.2. Tổng số lợi nhuận của từng bang, từng thành phố.....	30
2.20.3. Top 3 thành phố có số hóa đơn nhiều nhất	31
2.20.4. Top 10 mặt hàng (danh mục hàng) bán chạy nhất (số lượng bán nhiều nhất) 31	
2.20.5. Số đơn hàng của từng bang qua các năm.....	31
2.20.6. Hình thức giao hàng phổ biến nhất ở Texas	32
2.20.7. Loại điện thoại (Phones) nào được bán nhiều nhất (sắp xếp giảm dần) 32	
2.20.8. Năm có doanh thu hàng công nghệ (Technology) cao nhất	33
2.20.9. Số tiền mua hàng của Claire Gute qua từng quý trong năm	33
2.20.10. Ở New York City thì Consumer nào mua nhiều đồ nội thất (Furniture) nhất 34	
2.20.11. Liệt kê các sản phẩm có số lượng bán > 100	34

2.20.12.	Top 5 sản phẩm được giảm giá nhiều nhất ở bang Washington.....	35
2.20.13.	Top 5 hóa đơn có đơn giá cao nhất ở thành phố Houston	35
2.20.14.	Tổng giá tiền các Home Office bỏ ra để mua văn phòng phẩm (Office Supplies), sắp xếp tăng dần	36
2.20.15.	Loại hàng nào được vận chuyển bằng phương thức First Class nhiều nhất (sắp xếp giảm dần).....	36
4.	KHAI PHÁ DỮ LIỆU (DATA MINING)	37
2.21.	Tạo Project.....	37
2.22.	Khai phá với Time Series	37
5.	POWER BI.....	41
2.23.	Import dữ liệu với Analysis Services Database	41
2.24.	Thực hiện các báo cáo	42
2.24.1.	Lợi nhuận qua các năm	42
2.24.2.	Lợi nhuận của từng bang, thành phố.....	45
2.24.3.	Hình thức giao hàng phổ biến ở Texas	46

1. TỔNG QUAN ĐỀ TÀI

1.1. Mô tả dữ liệu gốc

Với nhu cầu ngày càng tăng và các cuộc cạnh tranh gắt gao trên thị trường, một Superstore đang tìm kiếm một sự phân tích về dữ liệu của họ. Họ muốn hiểu những sản phẩm, khu vực, danh mục và phân khúc khách hàng nào họ nên đặt mục tiêu hoặc tránh.

Thuộc tính	Ý nghĩa
Row ID	Thứ tự mỗi hàng (là chi tiết của mỗi món hàng trong hóa đơn)
Order ID	Mã hóa đơn
Order Date	Ngày mua hàng
Ship Date	Ngày giao hàng
Ship Mode	Phương thức vận chuyển
Customer ID	Mã khách hàng (duy nhất cho mỗi khách hàng)
Customer Name	Tên khách hàng
Segment	Phân khúc khách hàng
Country	Quốc gia (ở đây chỉ có United States)
City	Thành phố
State	Bang
Postal Code	Mã bưu điện
Region	Khu vực
Product ID	Mã sản phẩm (duy nhất cho mỗi sản phẩm)
Category	Danh mục sản phẩm
Sub-Category	Danh mục chi tiết
Product Name	Tên sản phẩm
Sales	Giá bán
Quantity	Số lượng mua
Discount	Giảm giá
Profit	Lợi nhuận

1.2. Mô tả dữ liệu đã xử lý

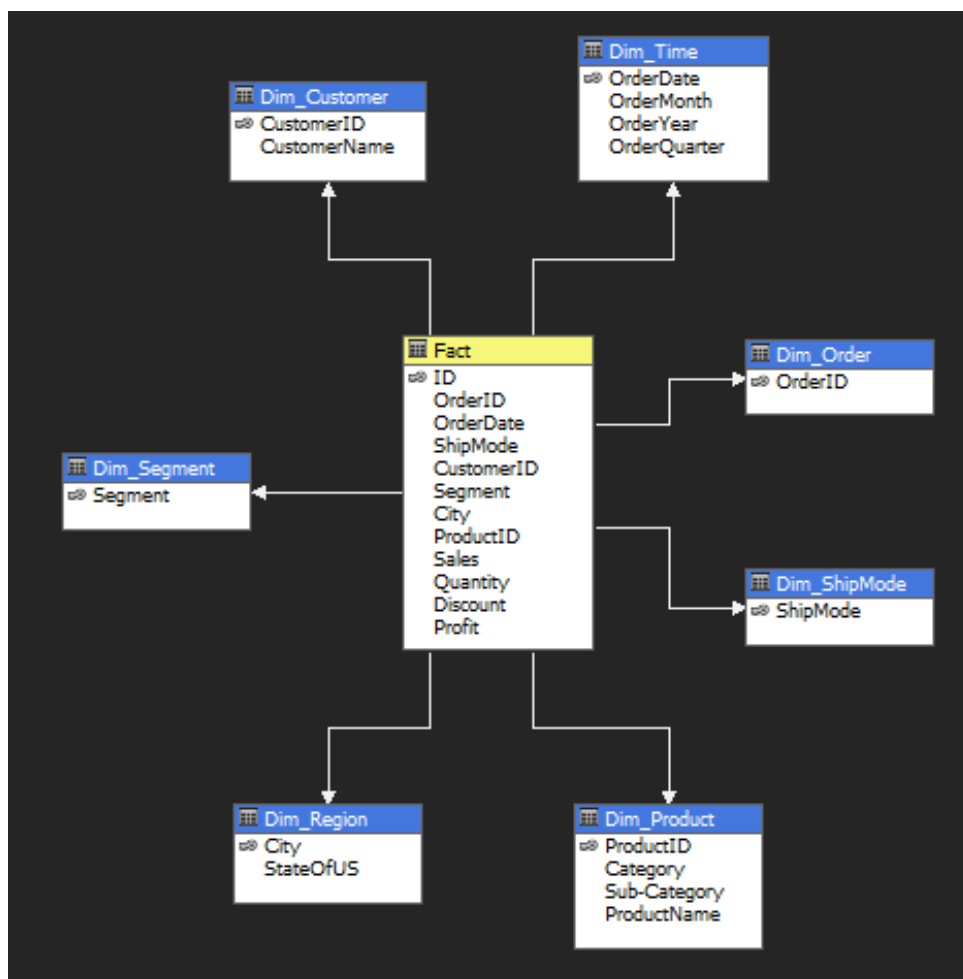
Lý do loại bỏ các cột dữ liệu:

- Ship Date: tránh phức tạp hóa dữ liệu khi phân tích
- Country: vì dữ liệu chỉ có duy nhất giá trị “United States”, đây cũng là lý do thuộc tính “State” được đổi thành “State Of US”
- Postal Code: vì mỗi mã bưu điện chỉ tồn tại ở 1 khu vực của thành phố, nhưng 1 khu vực đó lại có nhiều mã bưu điện. Khách hàng mỗi lúc lại đặt hàng ở những nơi khác nhau => không mang lại nhiều giá trị khi phân tích
- Region: chỉ mang các giá trị chung (đông, nam, trung tâm,) => không mang lại nhiều giá trị khi phân tích.

Thuộc tính	Ý nghĩa
ID	ID mỗi hàng
Order ID	Mã hóa đơn
Order Date	Ngày mua hàng
Ship Mode	Phương thức vận chuyển
Customer ID	Mã khách hàng (duy nhất cho mỗi khách hàng)
Customer Name	Tên khách hàng
Segment	Phân khúc khách hàng
City	Thành phố
State of US	Bang (ở US)
Product ID	Mã sản phẩm
Category	Danh mục sản phẩm
Sub-Category	Danh mục chi tiết
Product Name	Tên sản phẩm
Sales	Giá bán
Quantity	Số lượng mua
Discount	Giảm giá
Profit	Lợi nhuận

1.3. Thiết kế kho dữ liệu

1.3.1. Lược đồ hình sao



1.3.2. Bảng Fact

Thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
ID	Int	Khóa chính	ID mỗi hàng
OrderID	Varchar(50)	Khóa Ngoại	Mã hóa đơn
OrderDate	Datetime	Khóa Ngoại	Ngày đặt hàng
ShipMode	Varchar(50)	Khóa Ngoại	Phương thức vận chuyển
CustomerID	Varchar(50)	Khóa Ngoại	Mã khách hàng
Segment	Varchar(50)	Khóa Ngoại	Phân khúc khách hàng
City	Varchar(50)	Khóa Ngoại	Thành phố

ProductID	Varchar(50)	Khóa Ngoại	Mã sản phẩm
Sales	Float		Giá bán
Quantity	Int		Số lượng mua
Discount	Float		Giảm giá
Profit	float		Lợi nhuận

1.3.3. Bảng Dim

Dim_Order

Thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
OrderID	Varchar(50)	Khóa chính	Mã hóa đơn

Dim_Time

Thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
OrderDate	Datetime	Khóa chính	Ngày đặt hàng
OrderMonth	Int	Procedure	Tháng đặt hàng
OrderYear	Int	Procedure	Năm đặt hàng
OrderQuarter	Int	Procedure	Quý đặt hàng

Dim_ShipMode

Thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
ShipMode	Varchar(50)	Khóa chính	Phương thức giao hàng

Gồm 4 phương thức: First Class, Same Day, Second Class, Standard Class.

Dim_Customer

Thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
CustomerID	Varchar(50)	Khóa chính	Mã khách hàng
CustomerName	Varchar(50)		Tên khách hàng

Dim_Segment

Thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
Segment	Varchar(50)	Khóa chính	Phân khúc khách hàng

Gồm 3 phân khúc: Consumer, Corporate, Home Office.

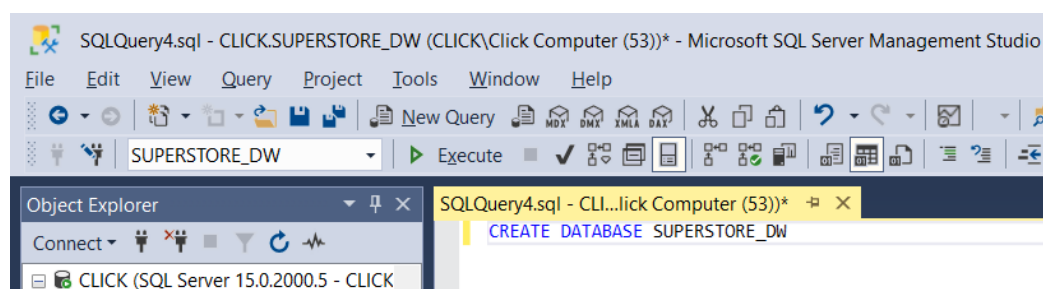
Dim_Region

Thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
City	Varchar(50)	Khóa chính	Thành phố
State Of US	Varchar(50)		Bang (ở US)

2.XÂY DỰNG KHO DỮ LIỆU (SSIS)

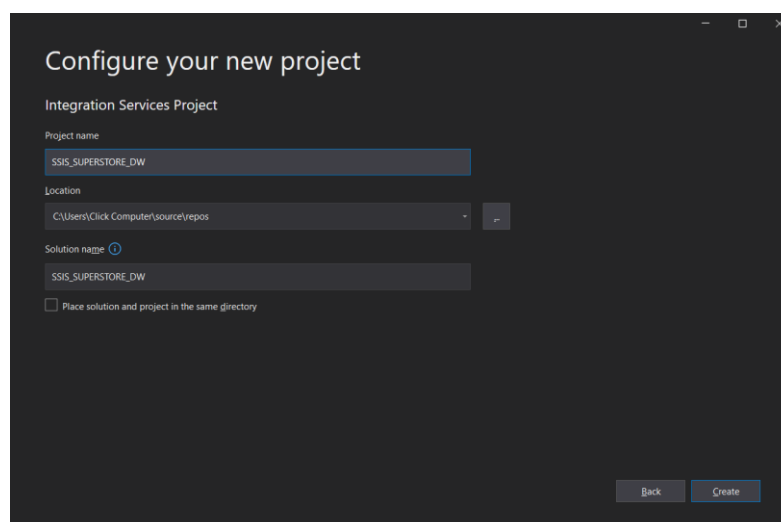
2.1. Tạo Database

Tạo mới một Database đặt tên là SUPERSTORE_DW.

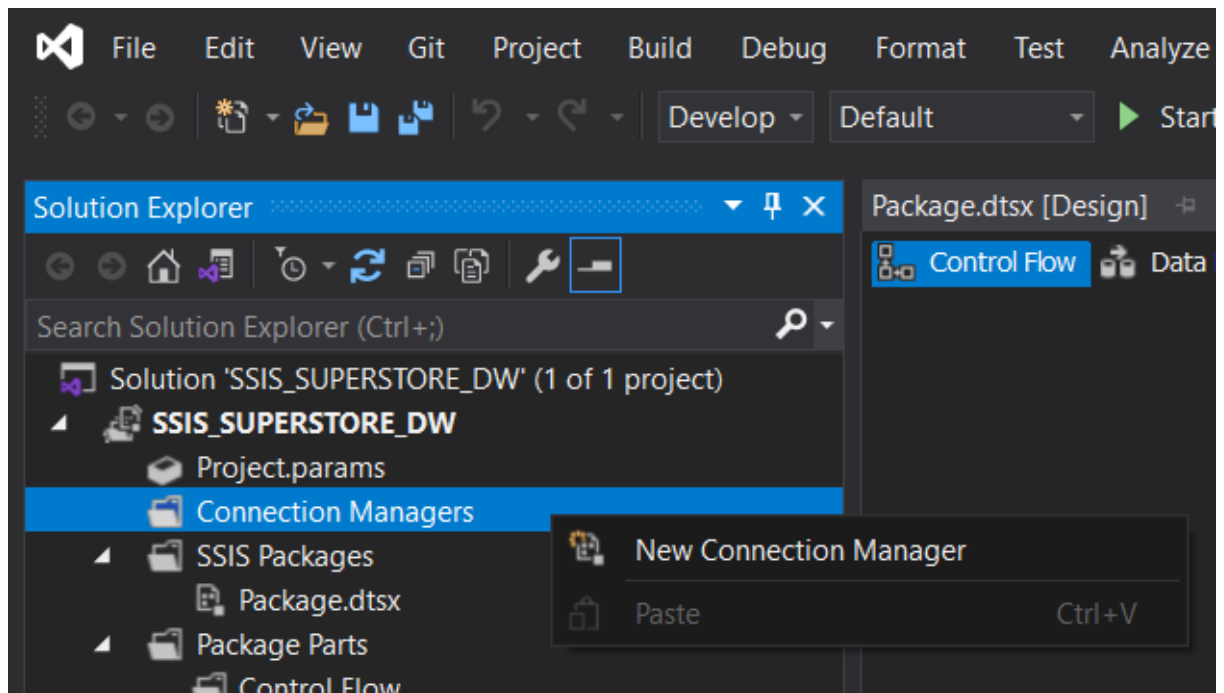


2.2. Tạo Project và thiết lập kết nối

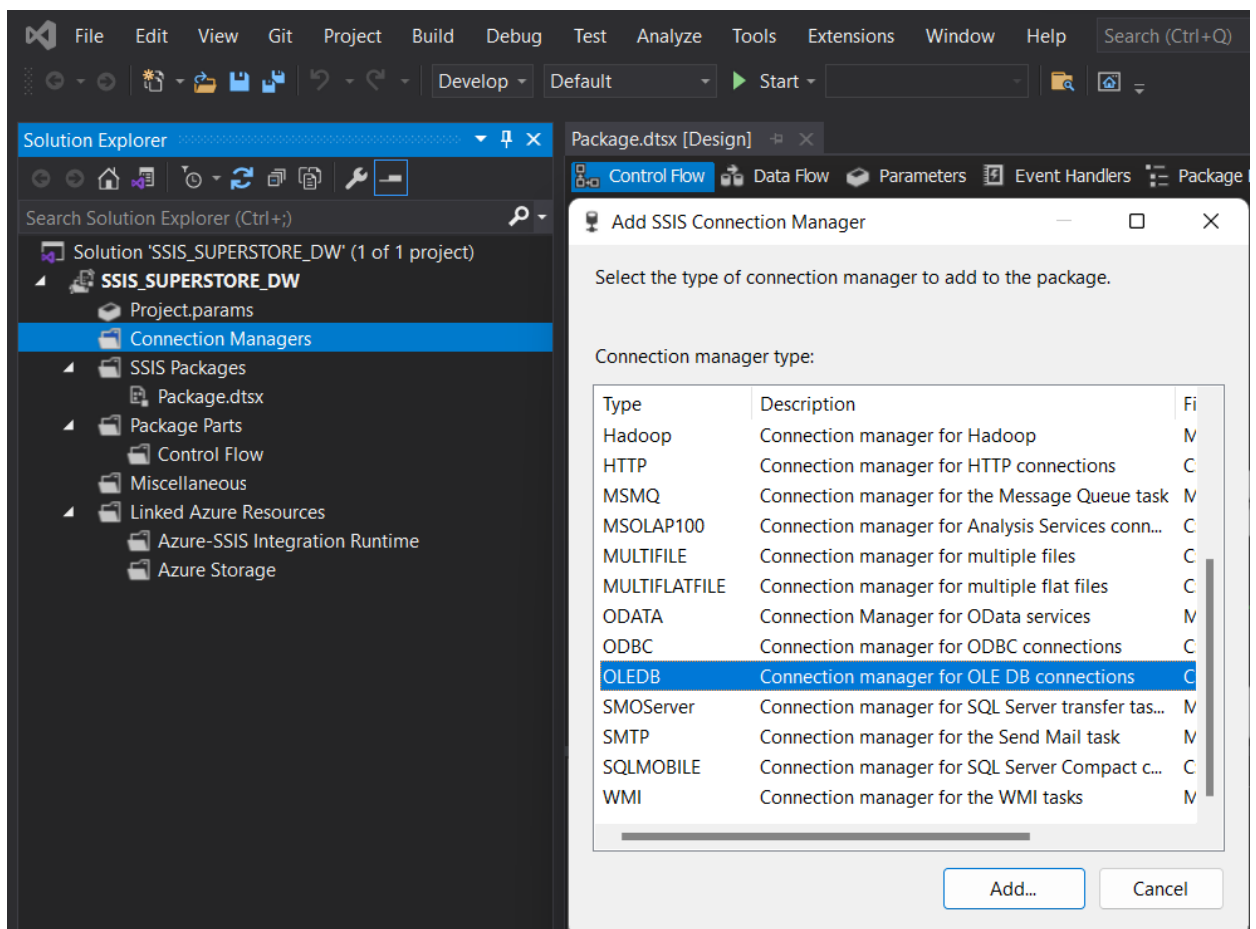
Vào Microsoft Visual Studio tạo một project SSIS đặt tên là SSIS_SUPERSTORE_DW.



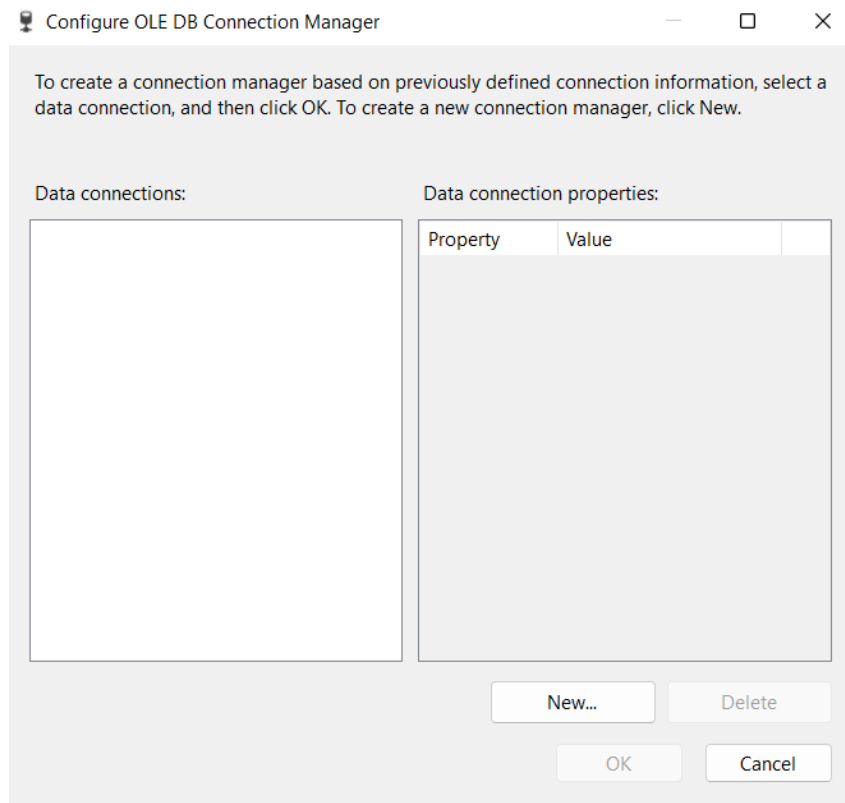
Click chuột phải vào Connection Managers chọn New Connection Managers.



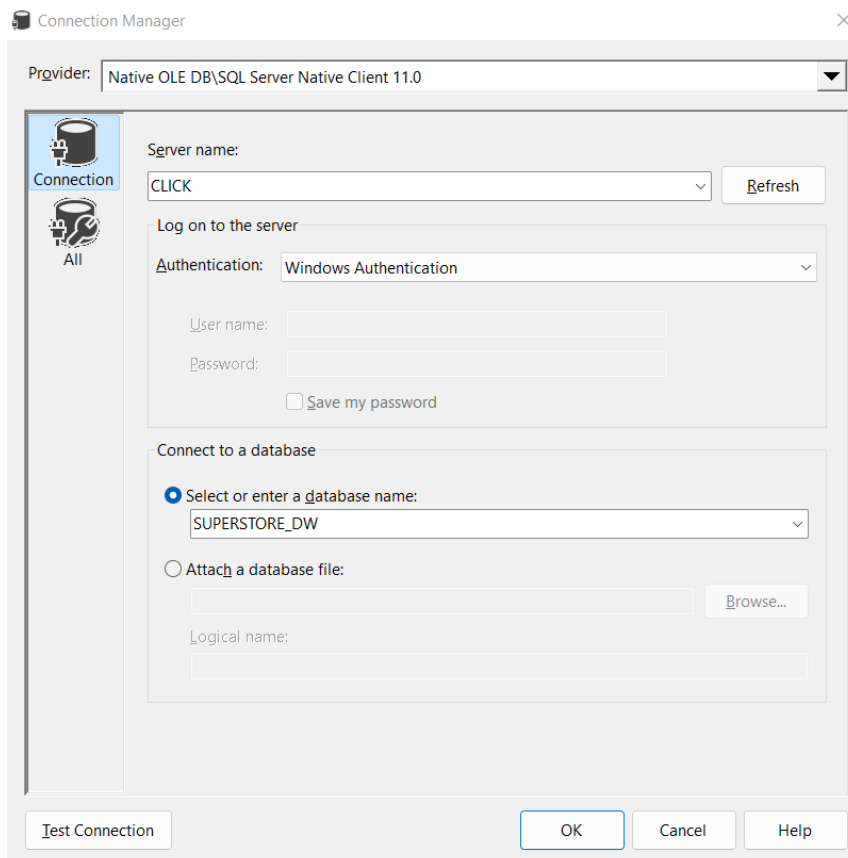
Chọn OLEDB và nhấn Add.



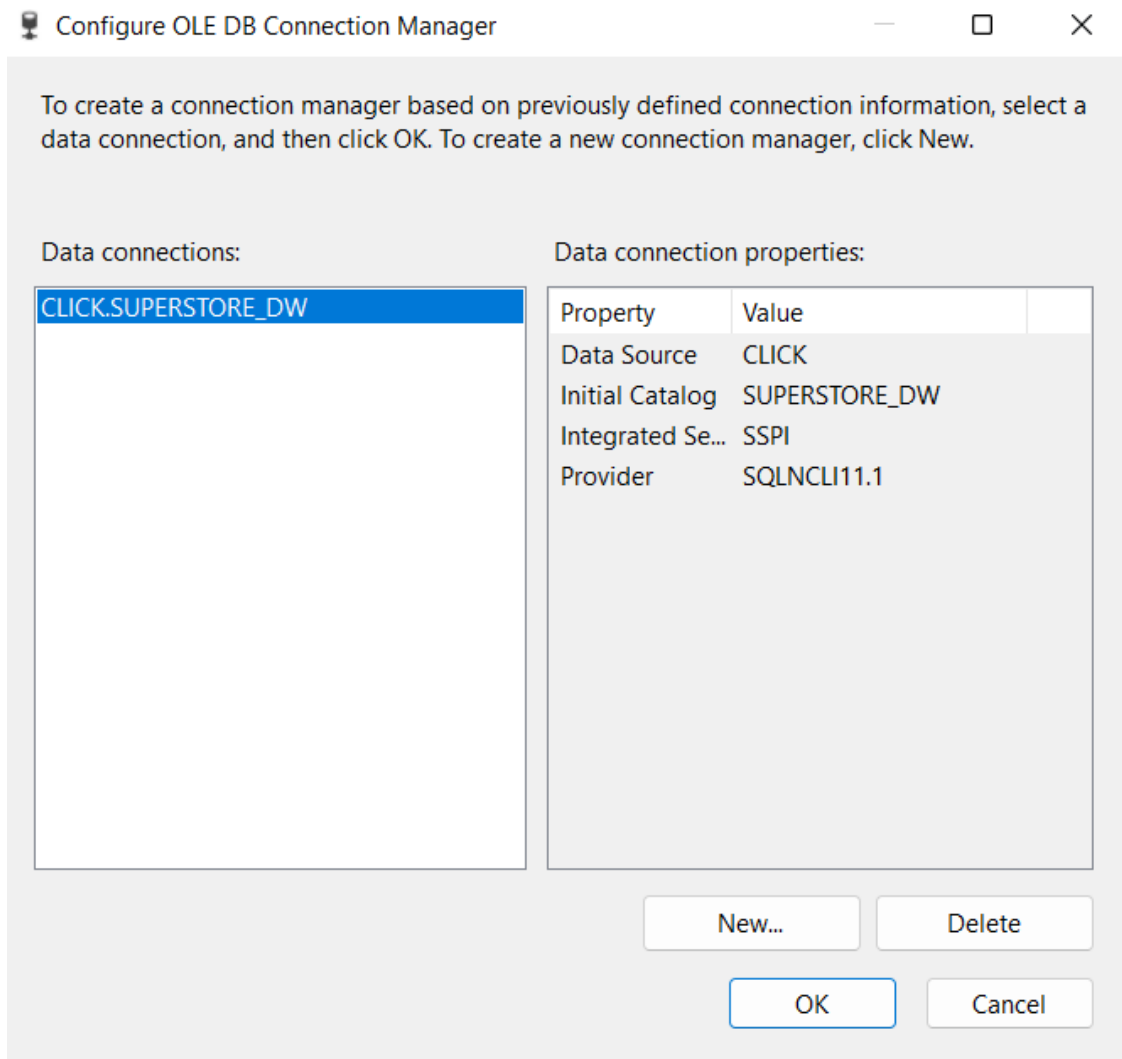
Tiếp tục nhấn New.



Nhập tên Sever name và connect tới Database đã tạo.

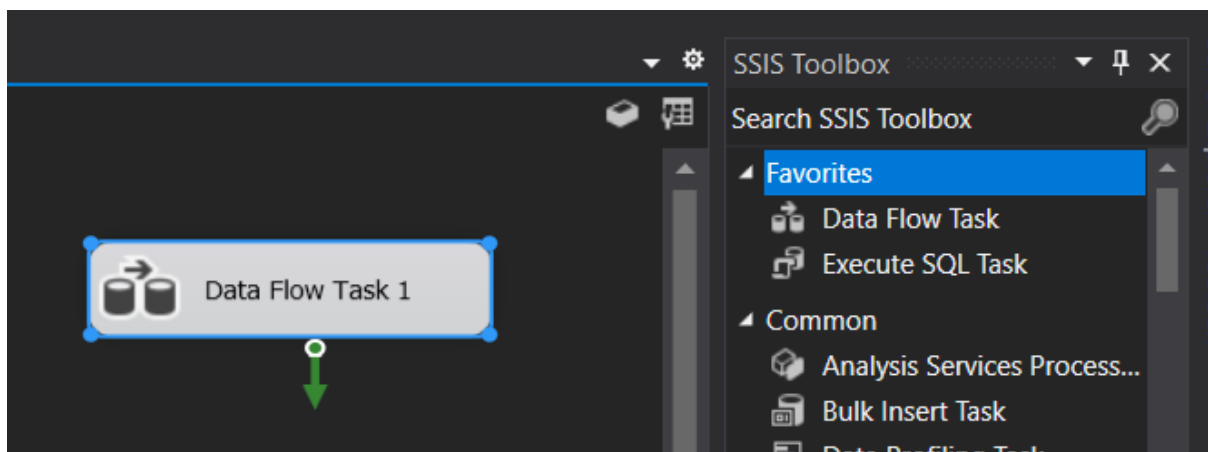


Nhấn Ok.



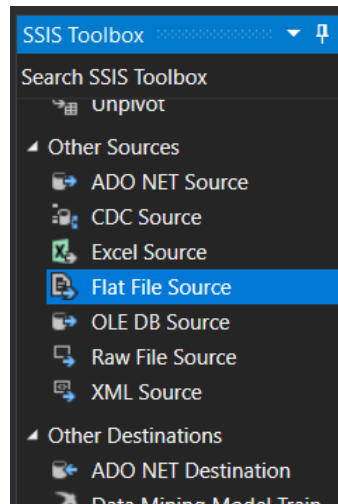
2.3. Tạo Data Flow Task

Đúp chuột vào Data Flow Task trên thanh công cụ, sau đó lại đúp vào Data Flow Task vừa tạo.

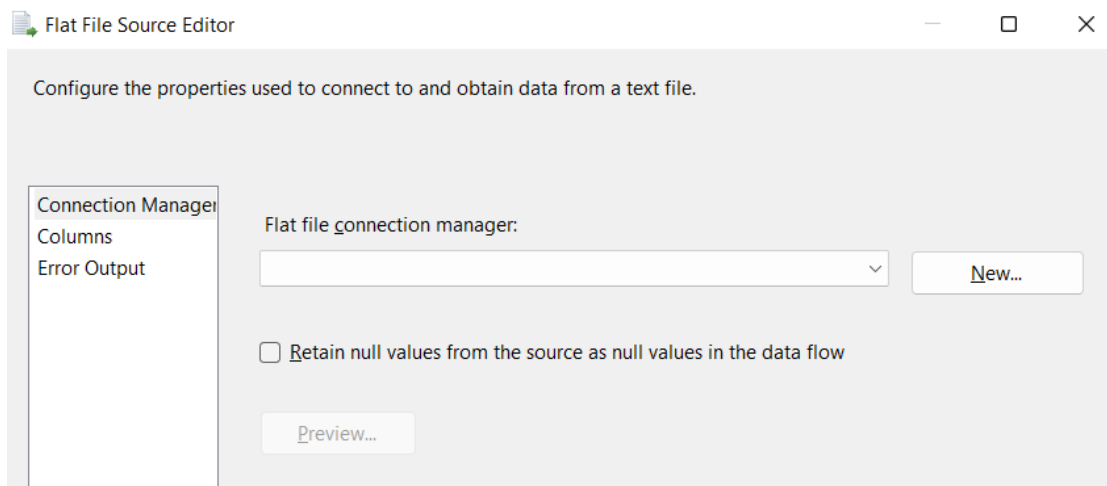


2.4. Tạo Flag File Source kết nối với file dữ liệu

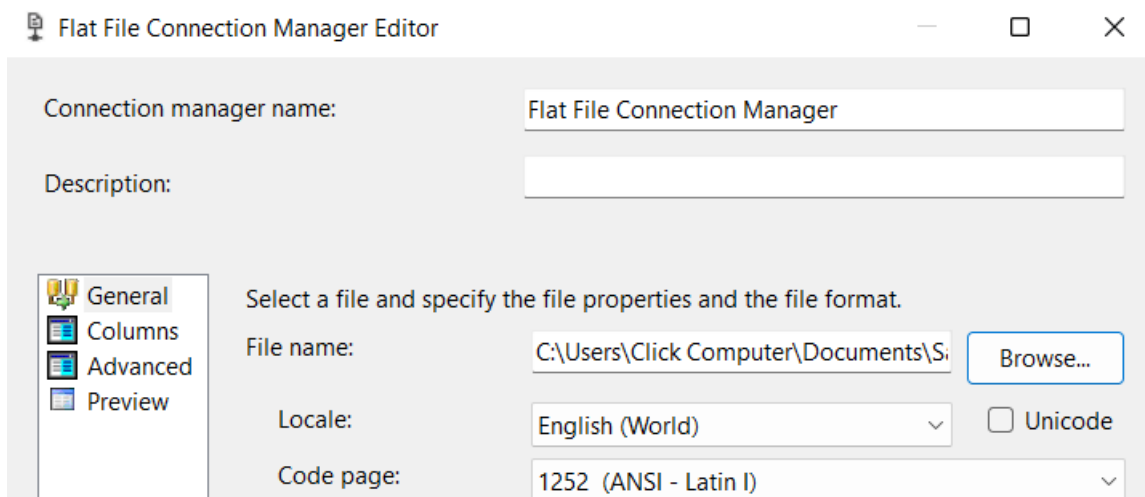
Chọn Flag File Source trên thanh công cụ và kết nối tới file dữ liệu.



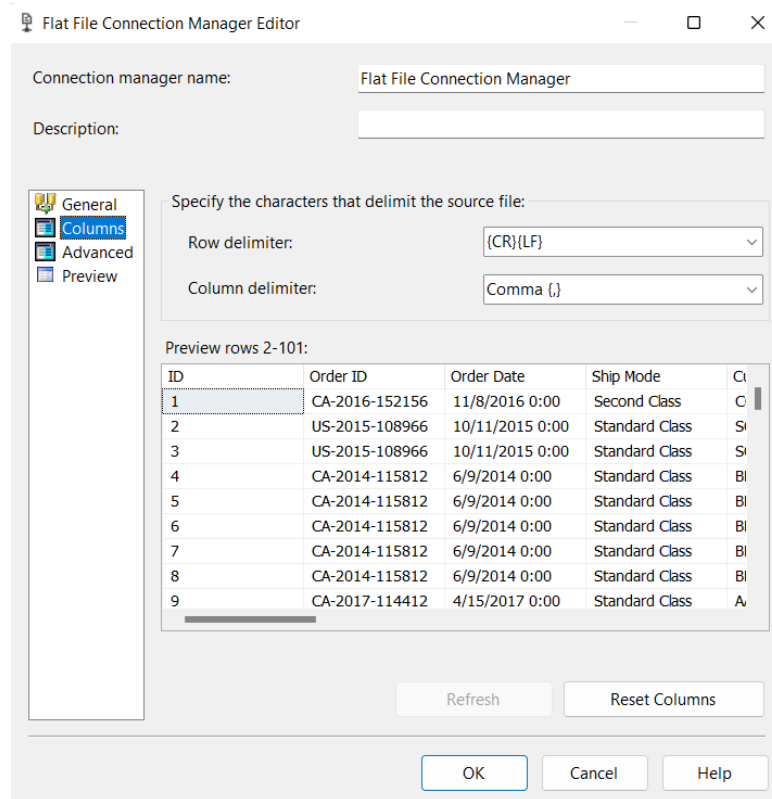
Chọn New.



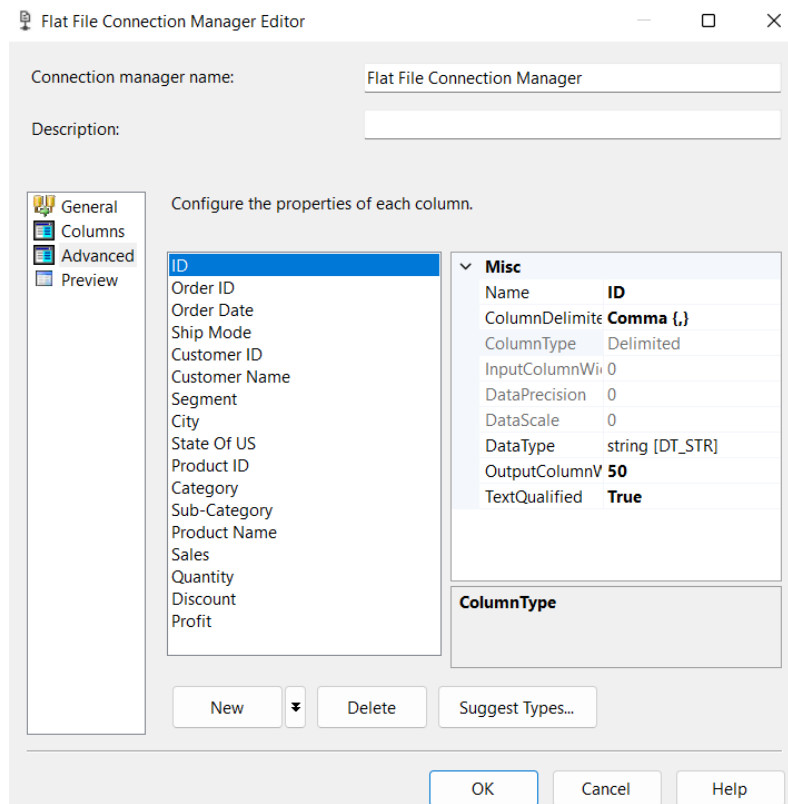
Chọn Browse và đường dẫn tới file dữ liệu



Chọn Columns để thực hiện chia dữ liệu theo cột hoặc hàng nếu dữ liệu chưa được chỉnh sửa.

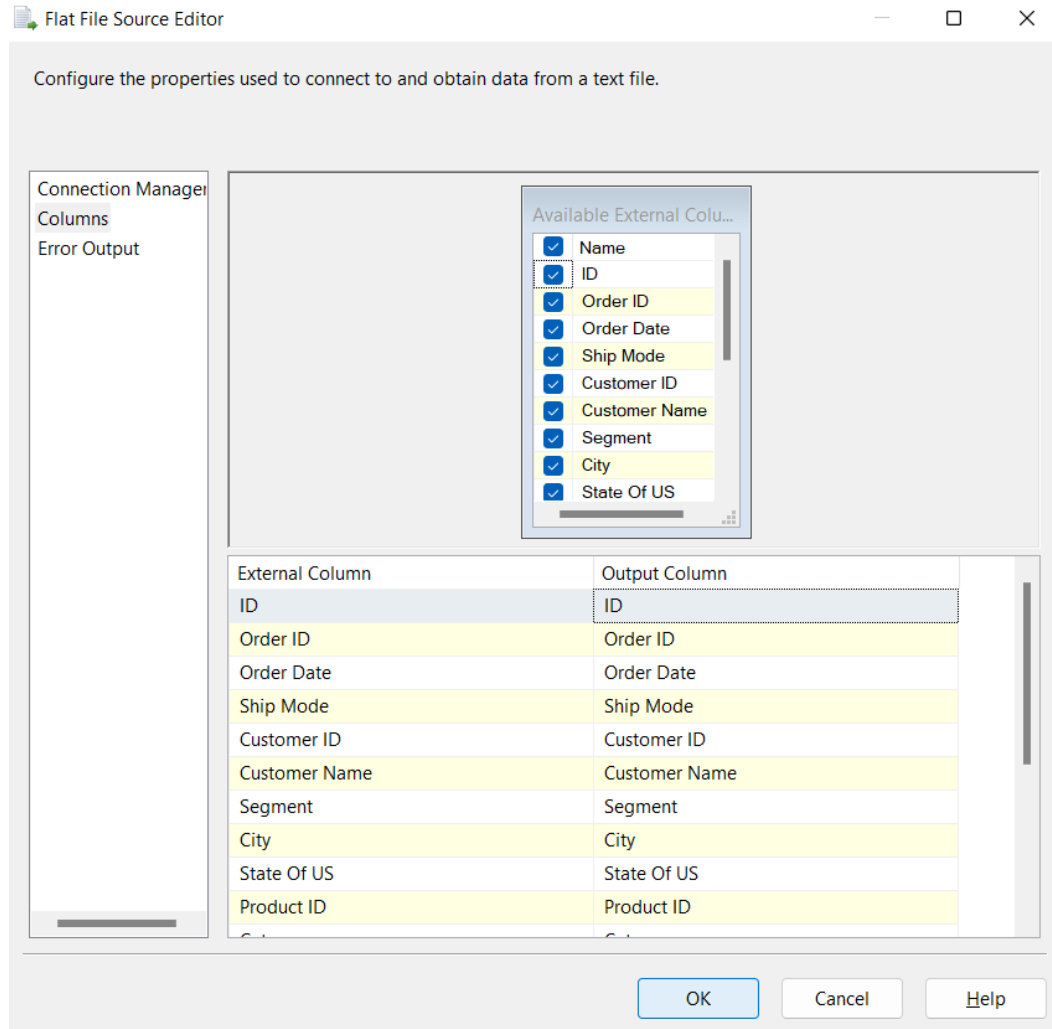


Chọn Advanced để chỉnh sửa dữ liệu (đổi tên các cột, xóa các cột không cần thiết)



Chọn Preview để kiểm tra lại và nhấn Ok.

Tiếp theo có thể chọn Columns để lựa chọn dữ liệu một lần nữa. Nếu hoàn tất, nhấn Ok.



2.5. Rẽ nhánh và lọc data trùng

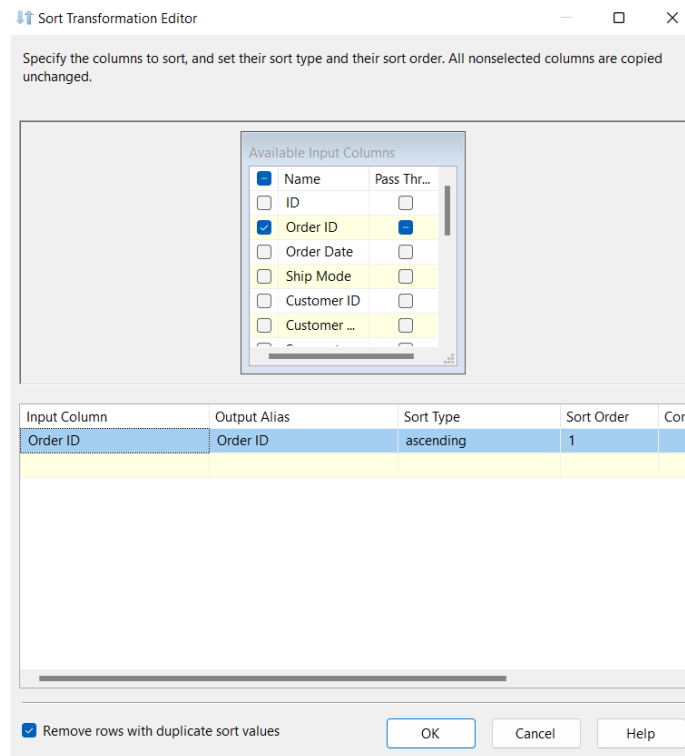
Chọn công cụ Multicast để có thể đẩy dữ liệu vào nhiều bảng



Sử dụng công cụ Sort để sắp xếp và lọc dữ liệu trùng trước khi đưa vào các bảng



Click chuột vào Sort, chọn thuộc tính cần lọc (tùy theo bảng Dim), click vào Remove rows with duplicate sort values để lọc dữ liệu trùng. Sau đấy hấn Ok.

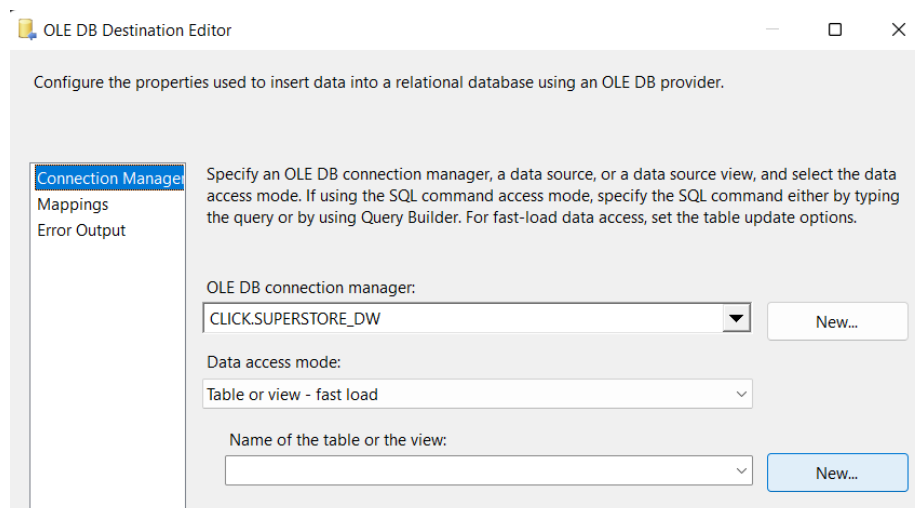


2.6. Tạo các bảng Dimension

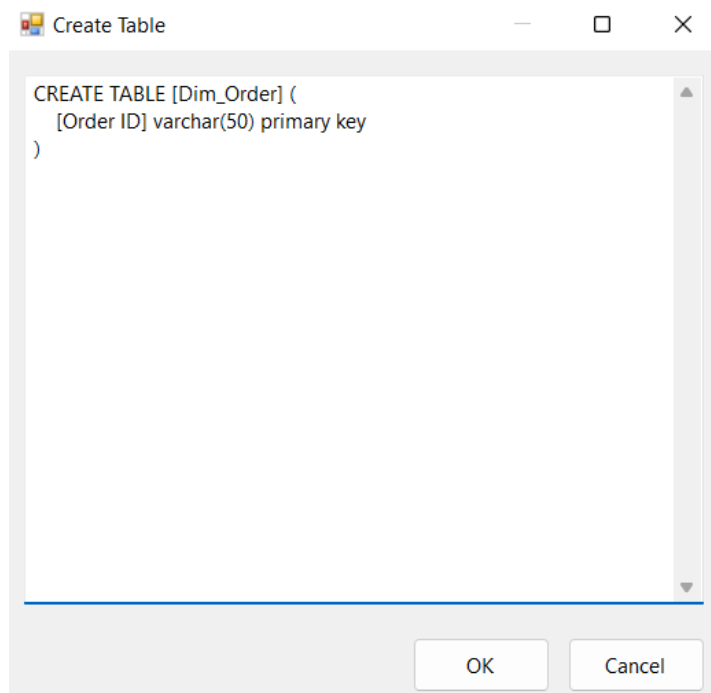
Chọn OLE DB Destination trên thanh công cụ. Click chuột phải, chọn Rename để đổi tên.



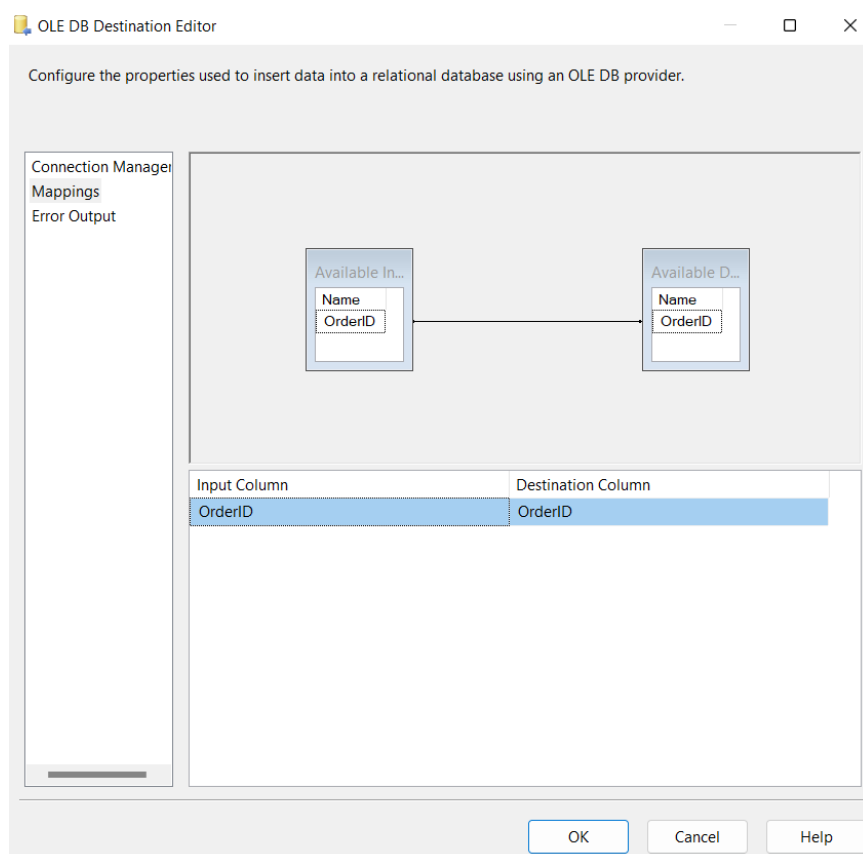
Click chuột vào Dim_Order và chọn New để thực hiện việc tạo bảng.



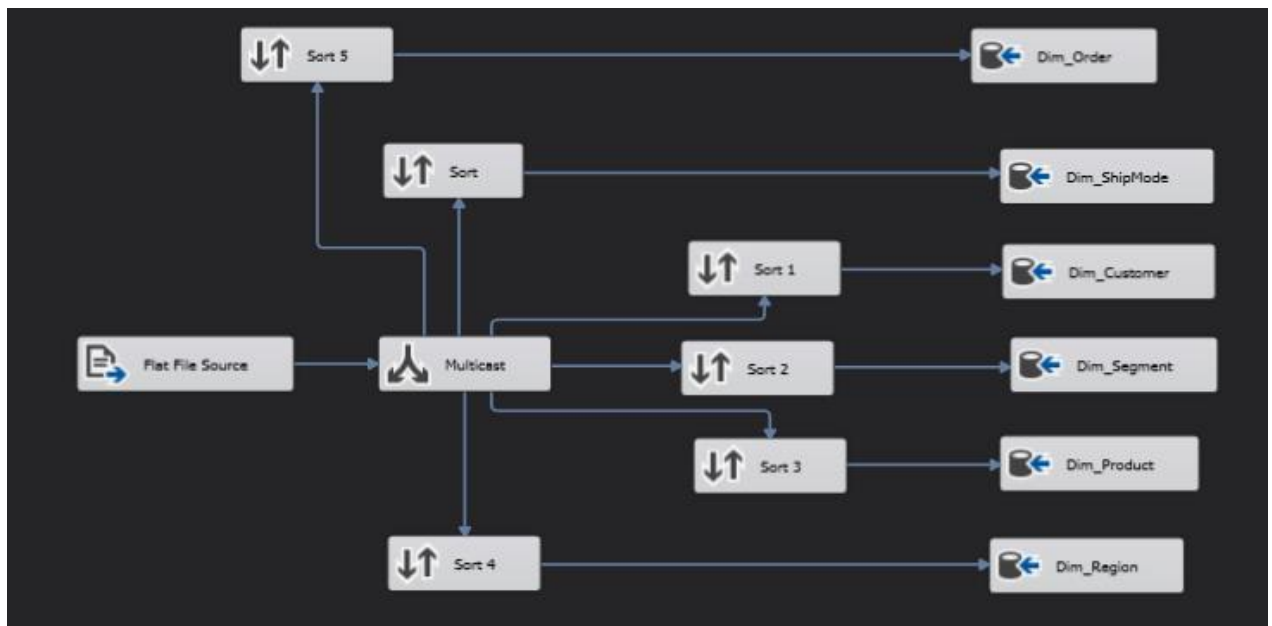
Tiến hành chỉnh sửa thuộc tính của bảng, thêm các khóa cần thiết. Nhấn Ok để hoàn tất.



Chọn Mappings để xem các thuộc tính đã liên kết đúng hay chưa và chỉnh sửa lại. Nhấn Ok để kết thúc.

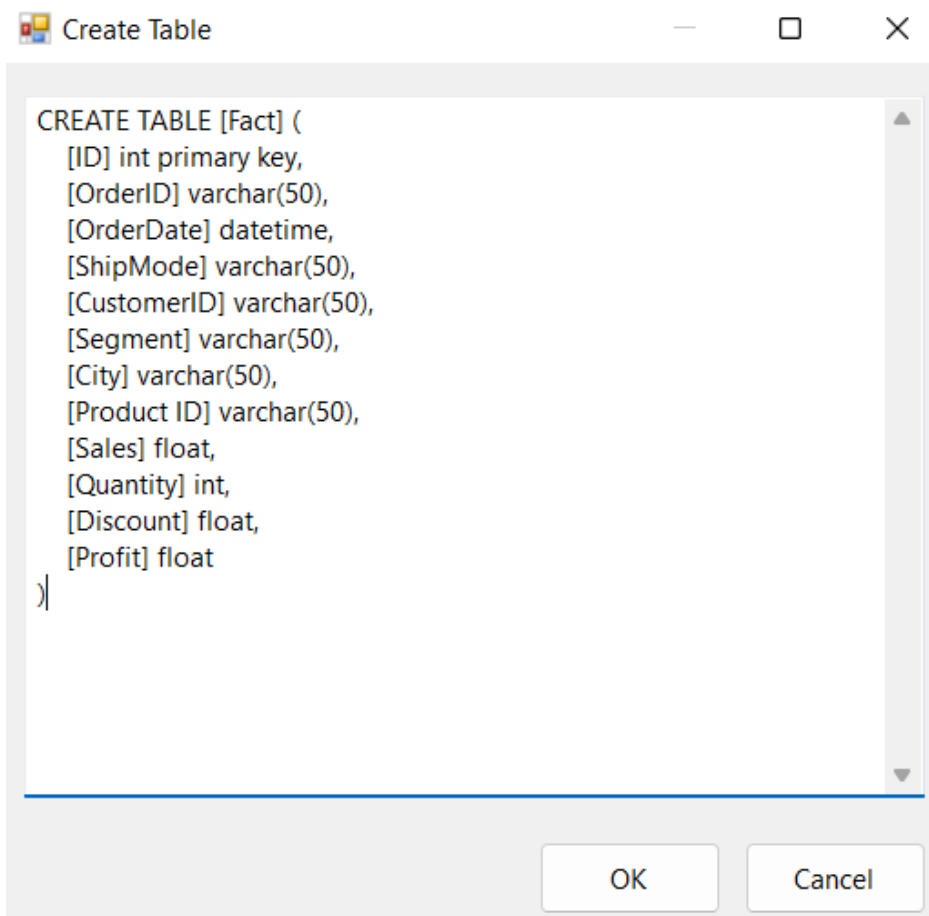


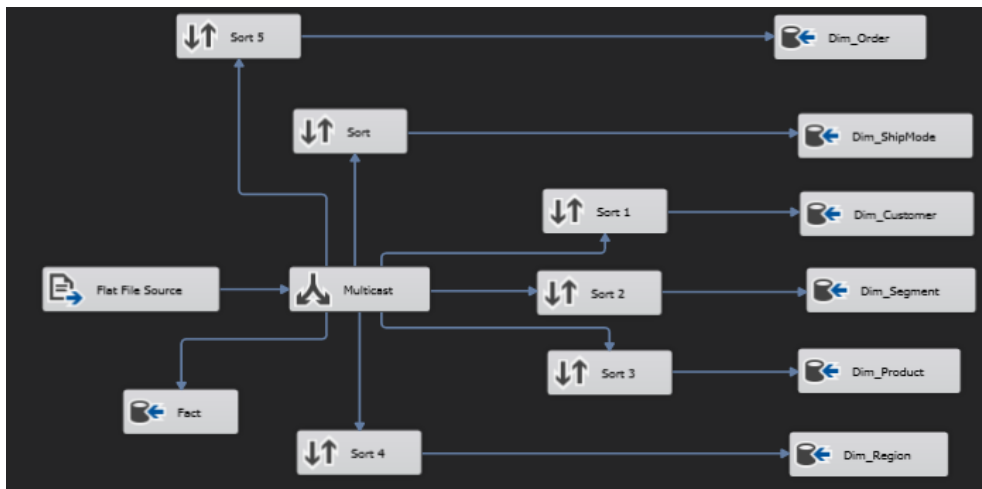
Thực hiện tương tự cho các bảng Dim khác



2.7. Tạo bảng Fact

Tạo mới một OLE DB Destination và đổi tên là Fact. Thực hiện thao tác tương tự như tạo bảng Dim.



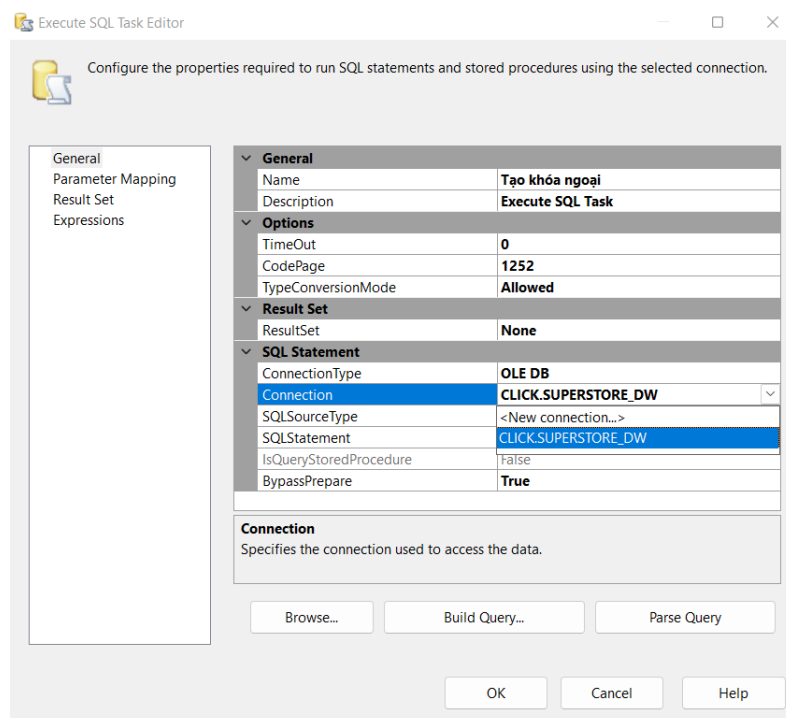


2.8. Tạo Execute SQL Task khóa ngoại cho bảng Fact và các bảng Dim

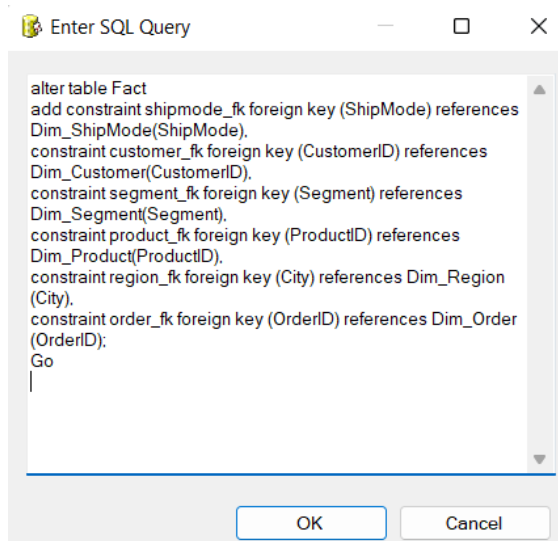
Quay lại giao diện Control Flow, tạo mới Execute SQL Task và đặt tên là Tạo khóa ngoại cho bảng Fact.



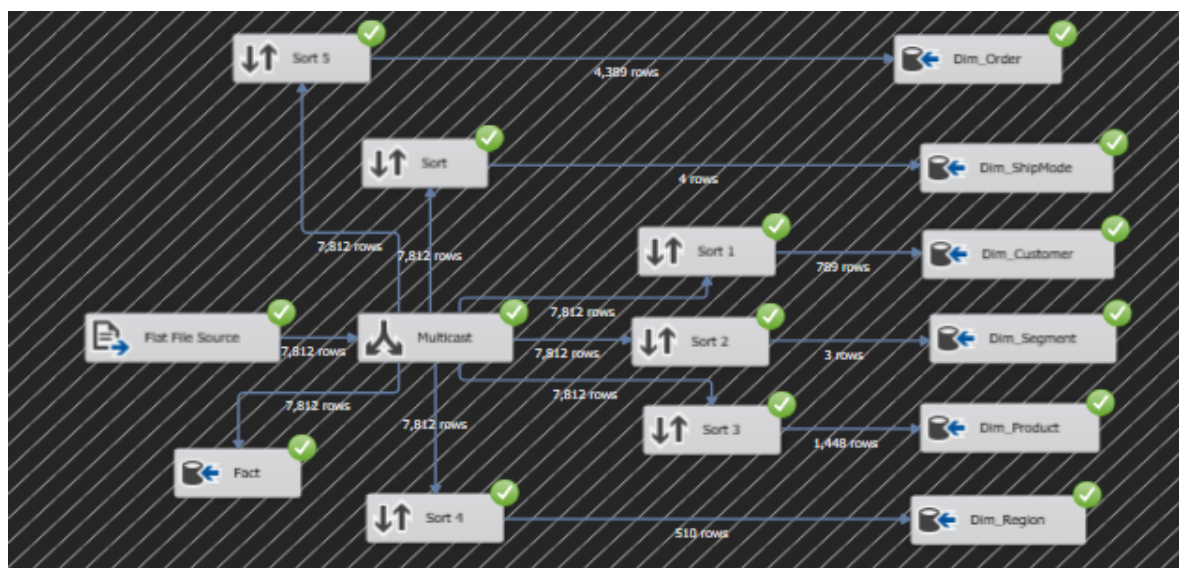
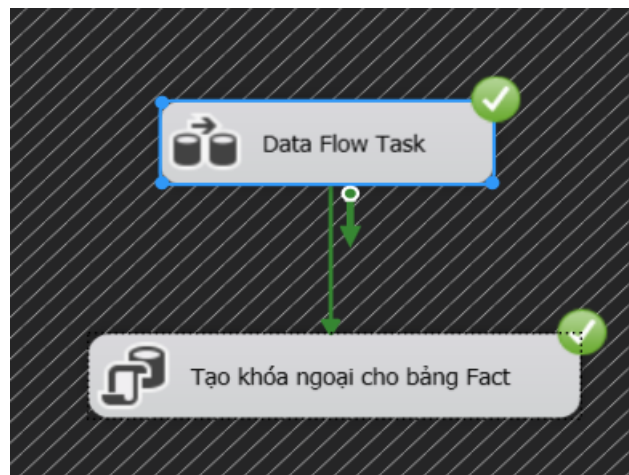
Tiếp tục Click chuột vào, ở dòng Connection chọn kết nối tới Database.



Click chuột vào dấu ... ở dòng SQLStatement và thực hiện nhập các câu lệnh SQL để tạo khóa ngoại. Nhấn 2 lần OK để kết thúc.

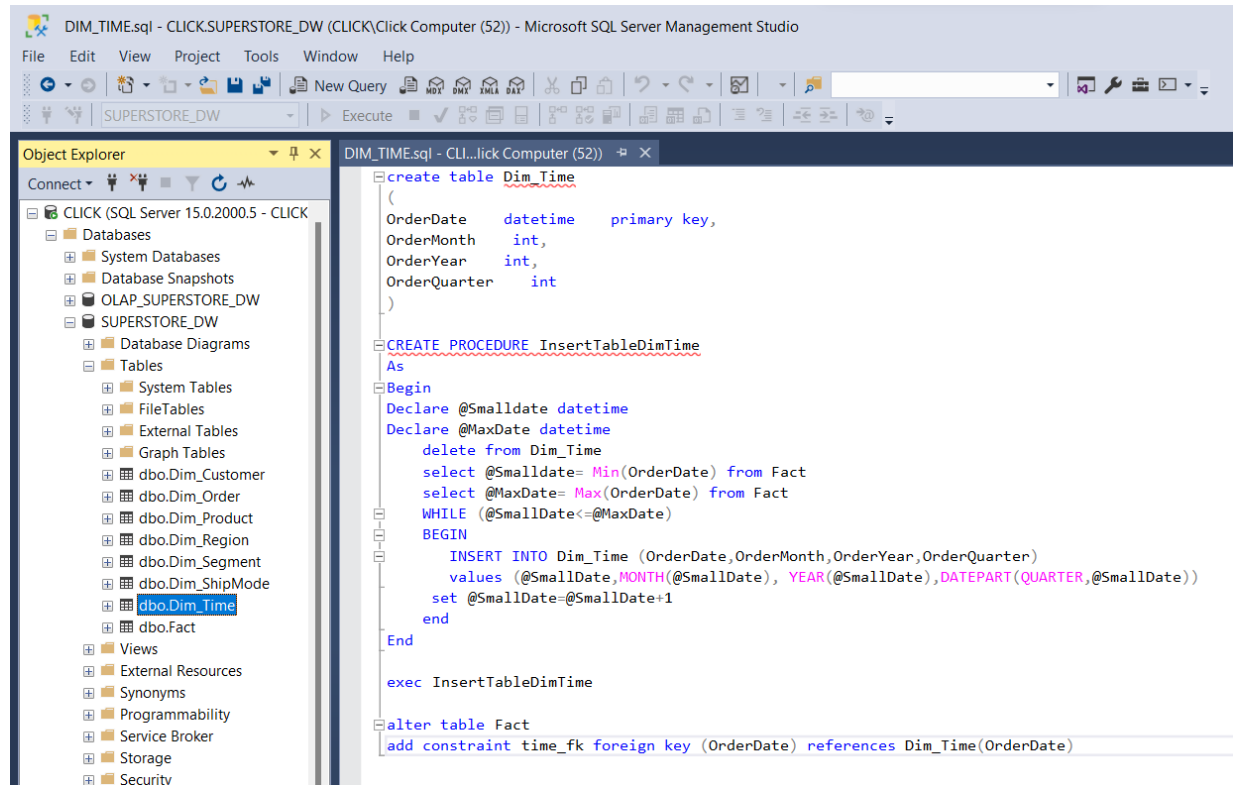


Nhấn Start để thực hiện chạy Project lần đầu tiên.



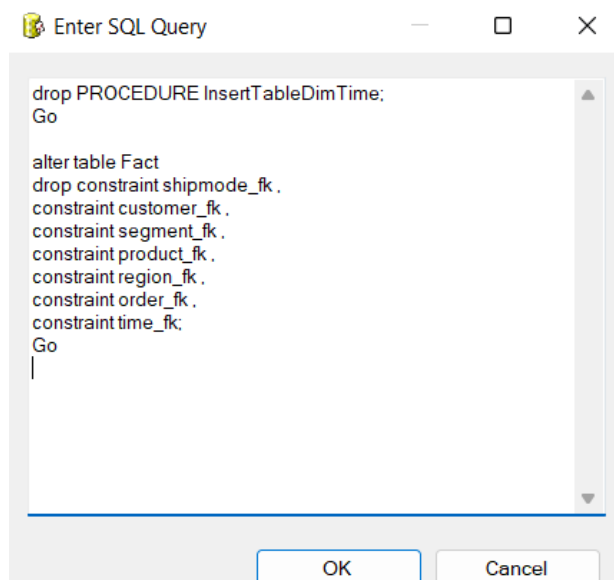
2.9. Tạo bảng Dim_Time và các ràng buộc

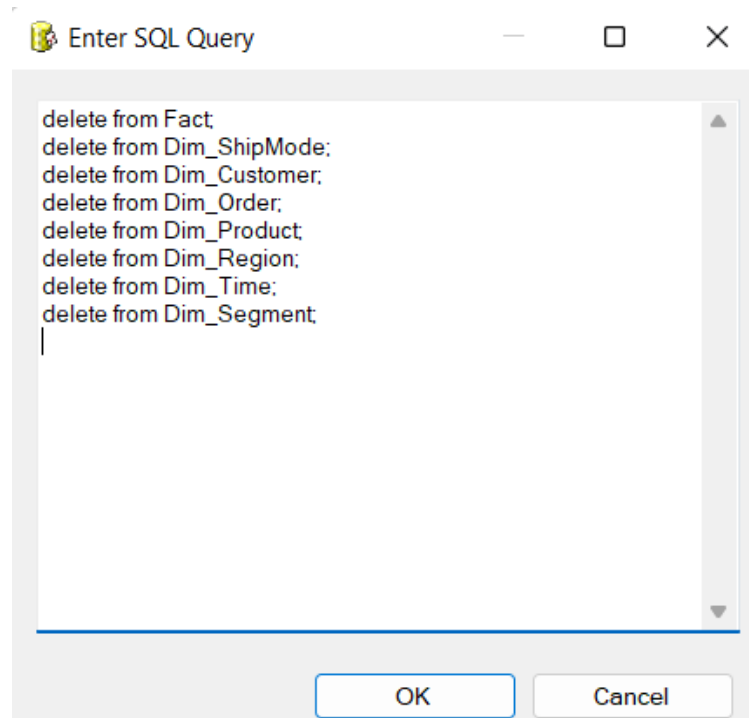
Vào SQL Server chọn New Query và thực thi lần lượt các lệnh: tạo bảng Dim_Time, tạo Procedure, tạo khóa ngoại OrderDate cho bảng Fact. (*)



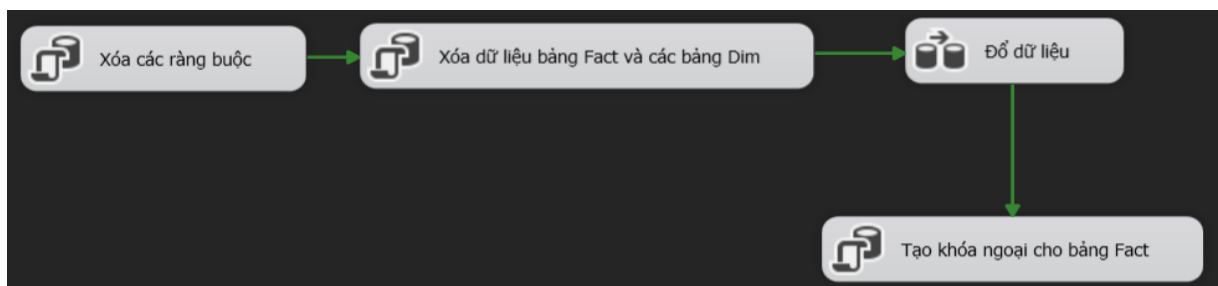
2.10. Tạo Execute SQL Task để xóa ràng buộc, xóa bảng Fact và các bảng Dim

Quay lại project SSIS tạo lần lượt 2 Execute SQL Task để xóa ràng buộc, xóa bảng Fact và các bảng Dim.



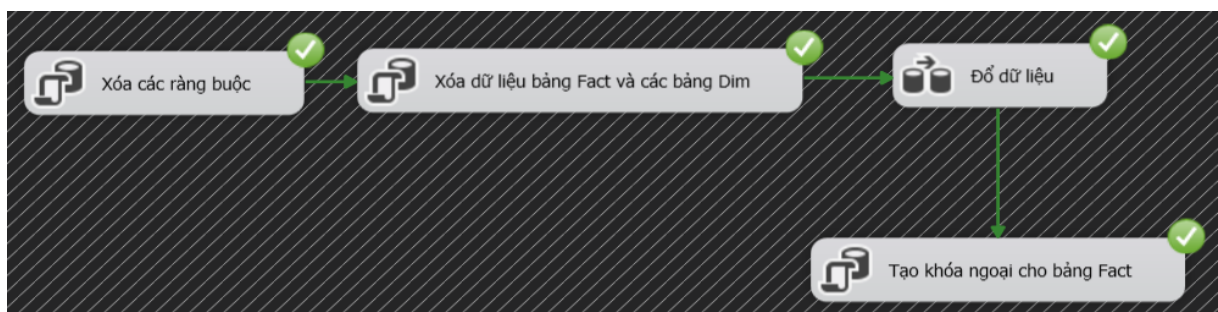


Nối các khối lệnh lại với nhau



2.11. Thực thi project

Chạy Project

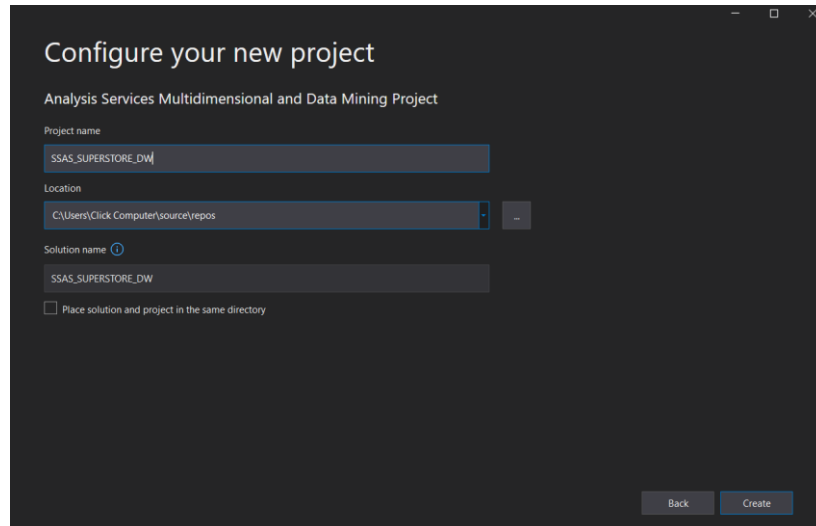


Lưu ý: kể từ lần chạy thứ 2, ta phải vào SQL Server để thực hiện lại bước (*) (trừ lệnh tạo bảng Dim_Time).

3. PHÂN TÍCH KHO DỮ LIỆU (SSAS)

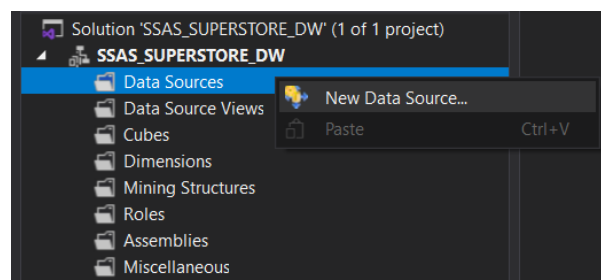
2.12.Tạo Project

Tạo một project SSAS đặt tên là SSAS_SUPERSTORE_DW.

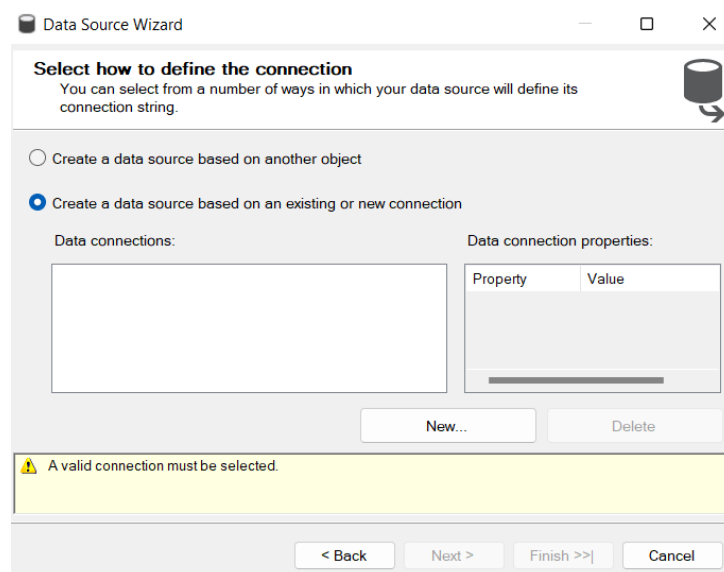


2.13.Tạo Data Source

Click chuột phải và Data Sources chọn New Data Source



Chọn Create a data source based on an existing or new connection, sau đó chọn New.



Nhập Server Name và chọn Database SUPERSTORE_DW.

Connection Manager

Provider: Native OLE DB\SQL Server Native Client 11.0

Server name: CLICK Refresh

Log on to the server

Authentication: Windows Authentication

User name: Password: Save my password

Connect to a database

☒ Select or enter a database name: SUPERSTORE_DW

☐ Attach a database file: Browse... Logical name:

Test Connection OK Cancel Help

Nhấn OK và nhấn Next.

Data Source Wizard

Select how to define the connection
You can select from a number of ways in which your data source will define its connection string.

☐ Create a data source based on another object

☒ Create a data source based on an existing or new connection

Data connections:

CLICK.SUPERSTORE_DW

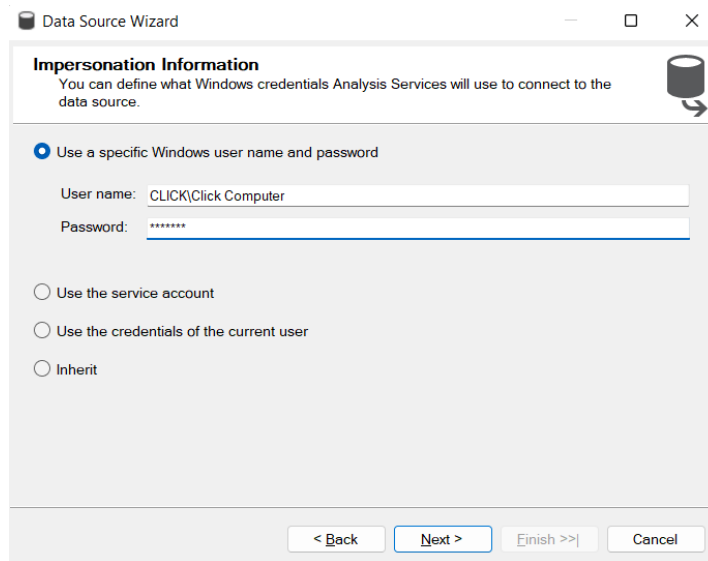
Data connection properties:

Property	Value
Data Source	CLICK
Initial Catalog	SUPERSTORE_DW
Integrated Sec...	SSPI
Provider	SQLNCLI11.1

New... Delete

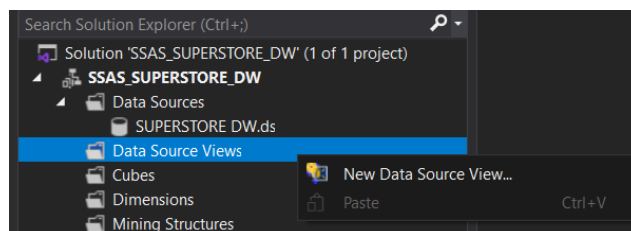
< Back Next > Finish >> Cancel

Nhập tên tài khoản trong phần Security của SQL Server và nhập mật khẩu đăng nhập Window. Nhấn Next. Nhấn Finish để kết thúc.

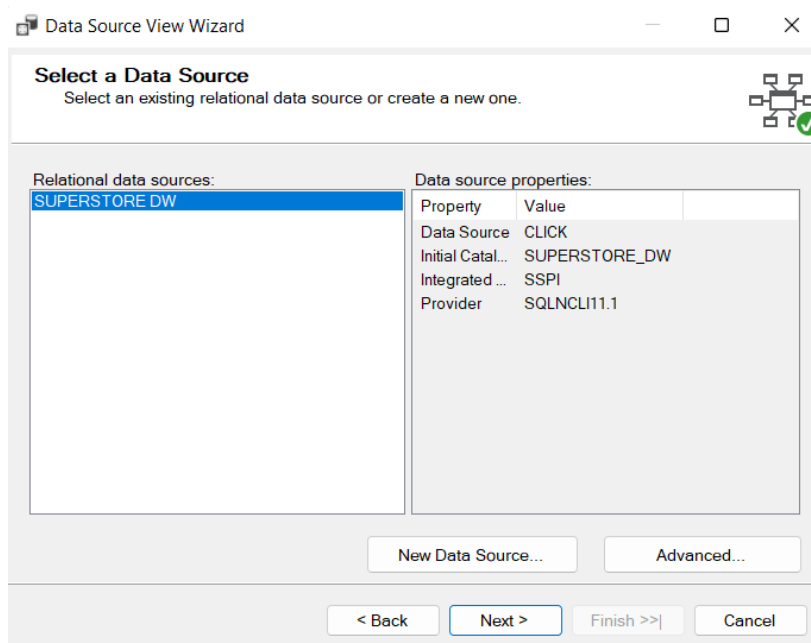


2.14. Tạo Data Source View

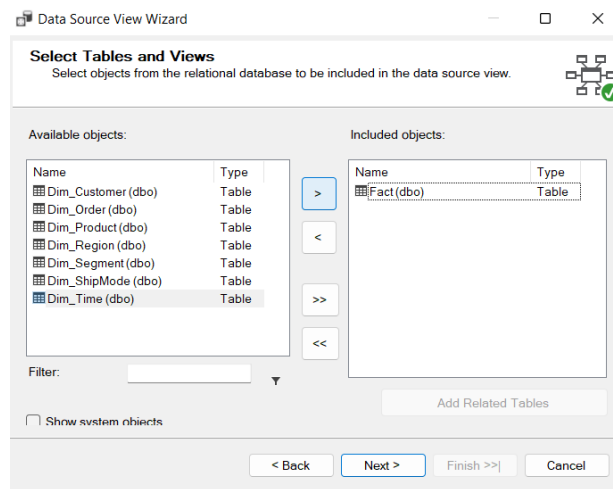
Click chuột phải và Data Source View chọn New Data Source View.



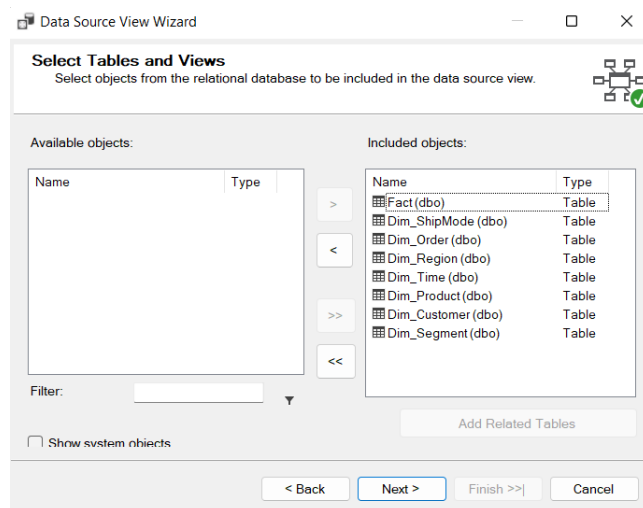
Chọn Data Source đã tạo và nhấn Next.



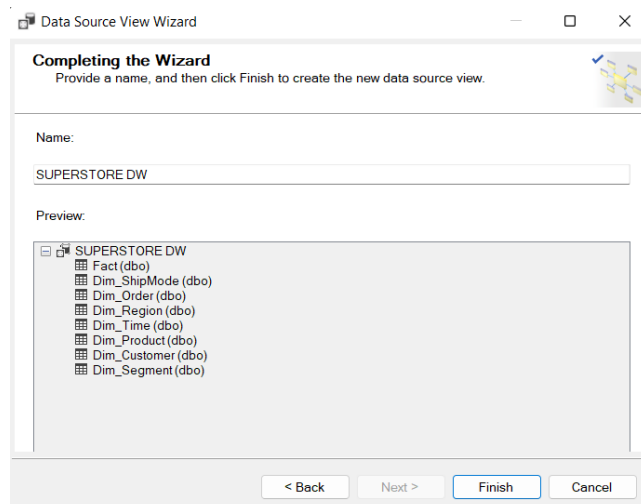
Chọn bảng Fact và bấm mũi tên để đẩy bảng Fact qua.



Chọn Add Related Tables để đẩy các bảng có liên kết với bảng Fact qua. Nhấn một lần nữa để đẩy các bảng có liên kết với các bảng Dim qua. Nhấn Next.

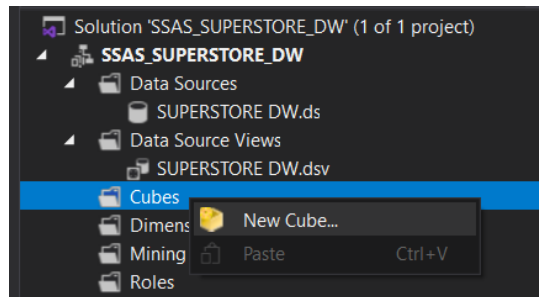


Nhấn Finish để kết thúc.

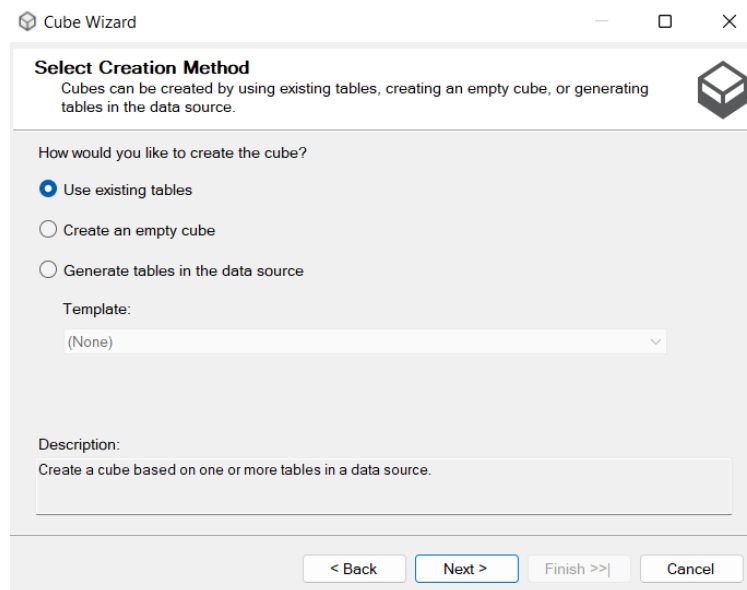


2.15.Tạo Cube

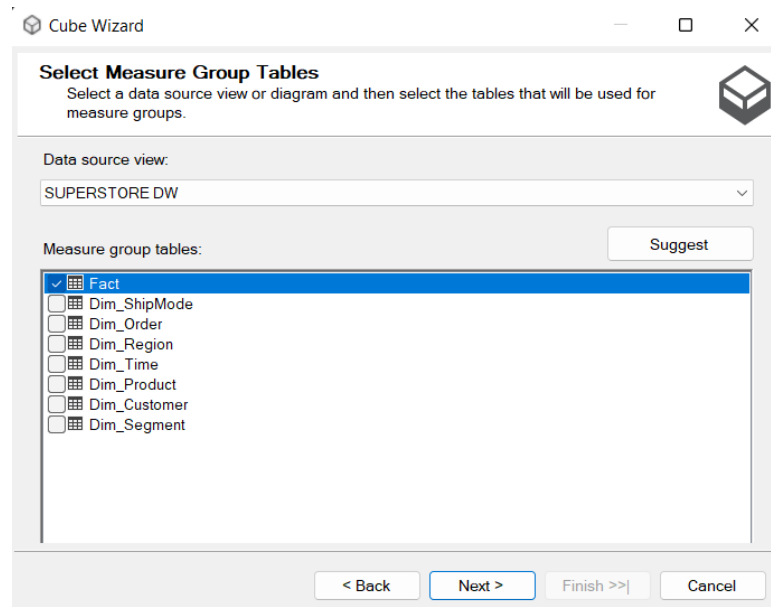
Chuột phải vào Cubes chọn New Cube.



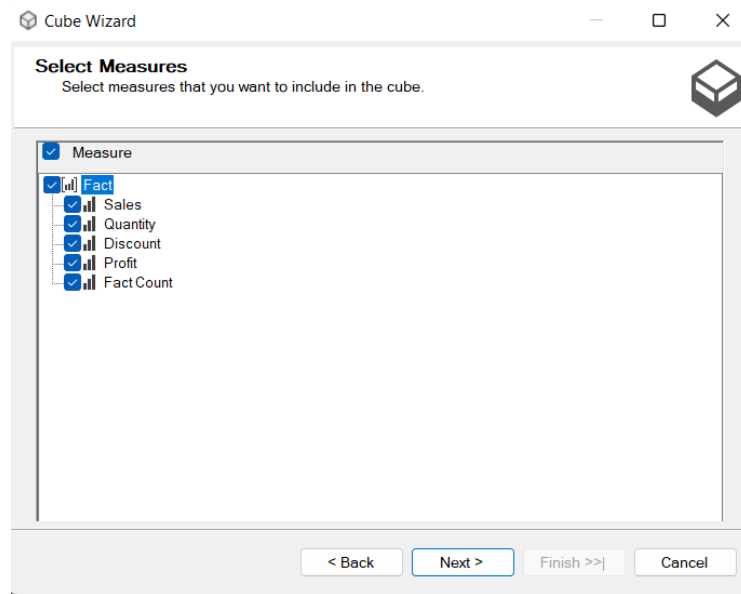
Nhấn Next.



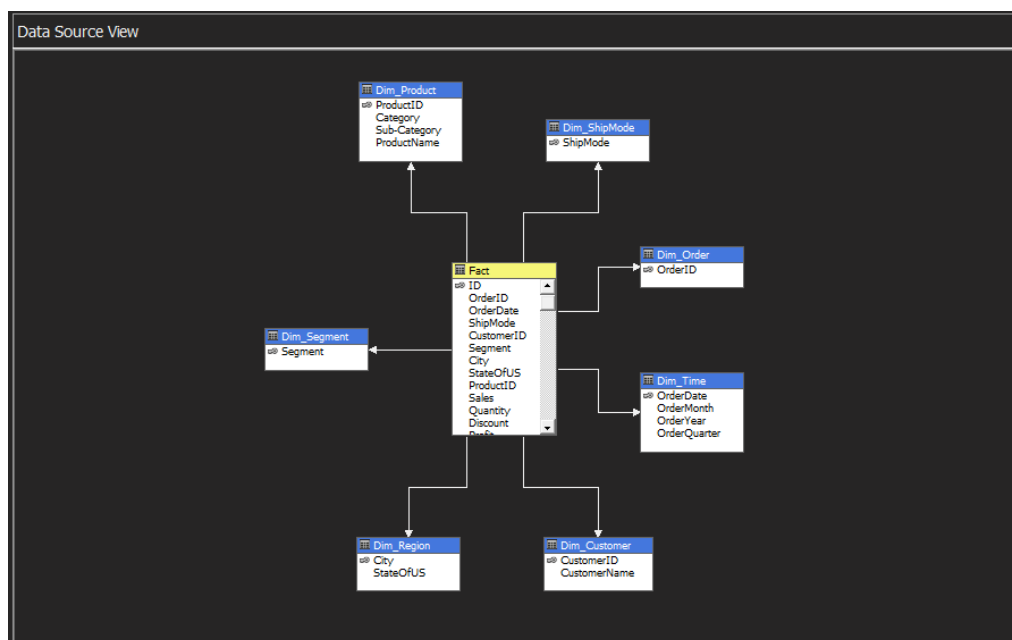
Chọn bảng Fact và nhấn Next.



Tiếp tục nhấn Next và chọn Finish để kết thúc.

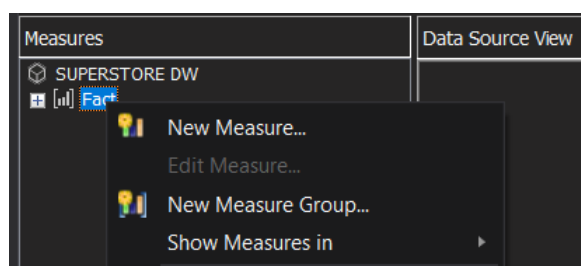


Lược đồ hình sao khi được tạo hoàn tất.

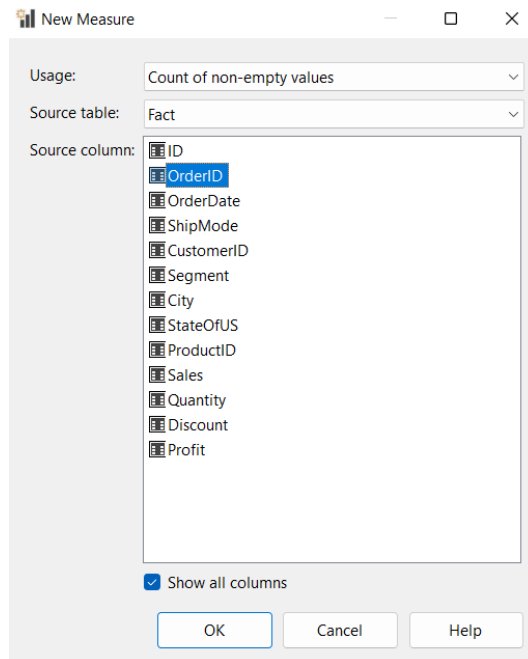


2.16. Tạo thêm Measure

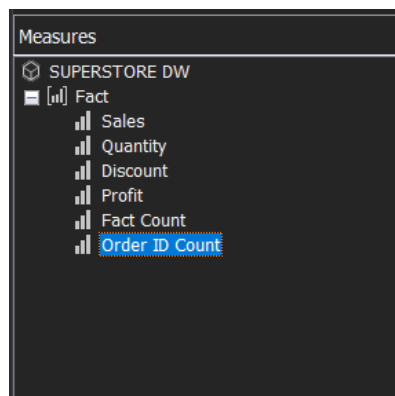
Ở cột Measure click chuột phải vào bảng Fact và chọn New Measure.



Ta cần tạo Measure để đếm số hóa đơn. Nên ở dòng hàm ta chọn Count of non-empty values chọn OrderID để đếm số hóa đơn không trùng. Nhấn OK.

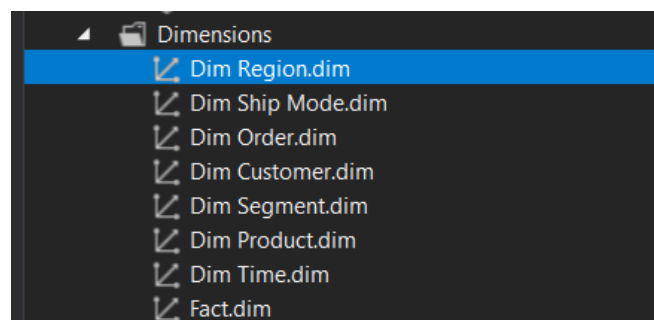


Kết quả



2.17. Chỉnh sửa thuộc tính các bảng Dim

Ở phần Dimensions lần lượt chọn các bảng Dim, sau đấy thực hiện kéo thả các thuộc tính chưa có ở phần Data Source View sang Attributes.



Attributes	Hierarchies	Data Source View
<div> <div> <div></div> <div>Dim Region</div> </div> <div> <div></div> <div>City</div> </div> <div> <div></div> <div>State Of US</div> </div> </div>	<div> <div></div> <div>To create a new hierarchy, drag an attribute here.</div> </div>	<div> <div> <div></div> <div>Dim Region</div> </div> <div> <div></div> <div>City</div> </div> <div> <div></div> <div>StateOfUS</div> </div> </div>

Thực hiện tương tự cho các bảng Dim còn lại.

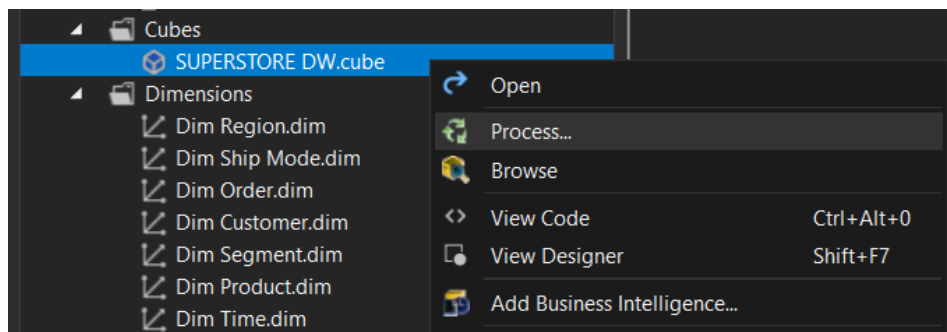
Attributes	Hierarchies	Data Source View
<div> <div> <div></div> <div>Dim Customer</div> </div> <div> <div></div> <div>Customer ID</div> </div> <div> <div></div> <div>Customer Name</div> </div> </div>	<div> <div></div> <div>To create a new hierarchy, drag an attribute here.</div> </div>	<div> <div> <div></div> <div>Dim Customer</div> </div> <div> <div></div> <div>CustomerID</div> </div> <div> <div></div> <div>CustomerName</div> </div> </div>

Attributes	Hierarchies	Data Source View
<div> <div> <div></div> <div>Dim Product</div> </div> <div> <div></div> <div>Category</div> </div> <div> <div></div> <div>Product ID</div> </div> <div> <div></div> <div>Product Name</div> </div> <div> <div></div> <div>Sub-Category</div> </div> </div>	<div> <div></div> <div>To create a new hierarchy, drag an attribute here.</div> </div>	<div> <div> <div></div> <div>Dim Product</div> </div> <div> <div></div> <div>ProductID</div> </div> <div> <div></div> <div>Category</div> </div> <div> <div></div> <div>Sub-Category</div> </div> <div> <div></div> <div>ProductName</div> </div> </div>

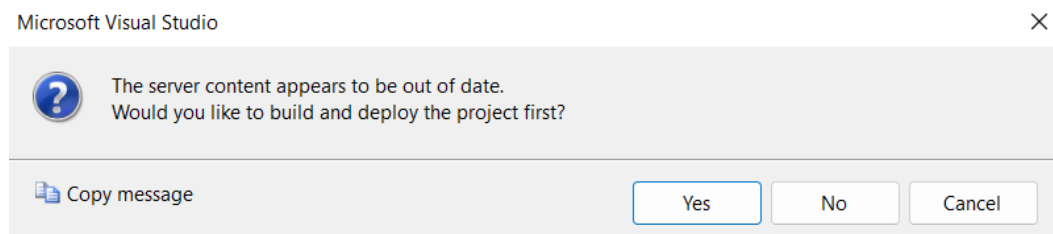
Attributes	Hierarchies	Data Source View
<div> <div> <div></div> <div>Dim Time</div> </div> <div> <div></div> <div>Order Date</div> </div> <div> <div></div> <div>Order Month</div> </div> <div> <div></div> <div>Order Quarter</div> </div> <div> <div></div> <div>Order Year</div> </div> </div>	<div> <div></div> <div>To create a new hierarchy, drag an attribute here.</div> </div>	<div> <div> <div></div> <div>Dim Time</div> </div> <div> <div></div> <div>OrderDate</div> </div> <div> <div></div> <div>OrderMonth</div> </div> <div> <div></div> <div>OrderYear</div> </div> <div> <div></div> <div>OrderQuarter</div> </div> </div>

2.18. Thực thi project

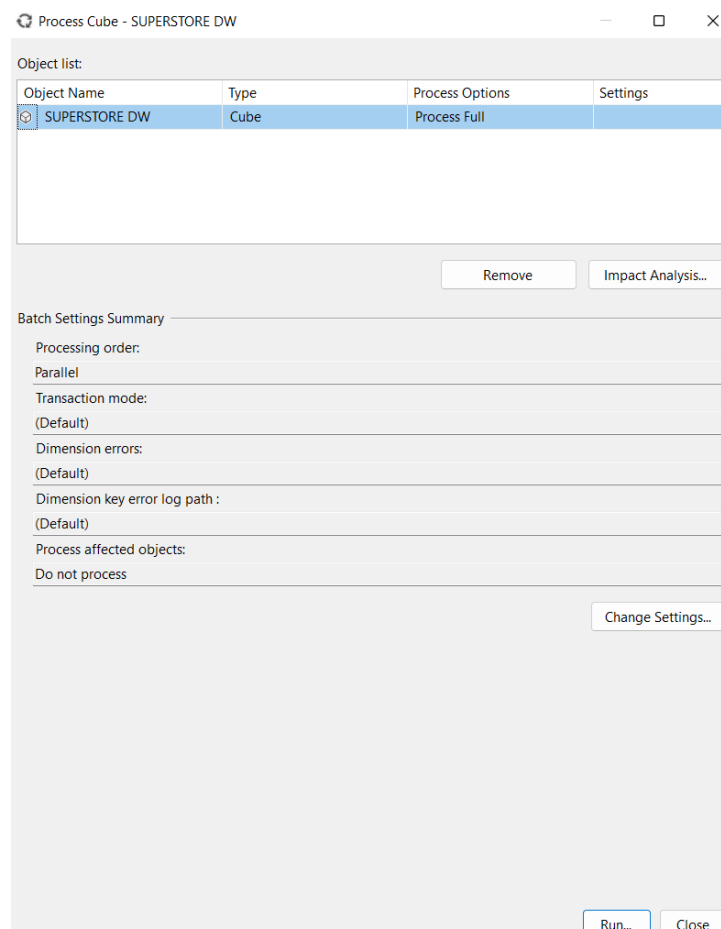
Click chuột phải vào Cube đã tạo và chọn Process.



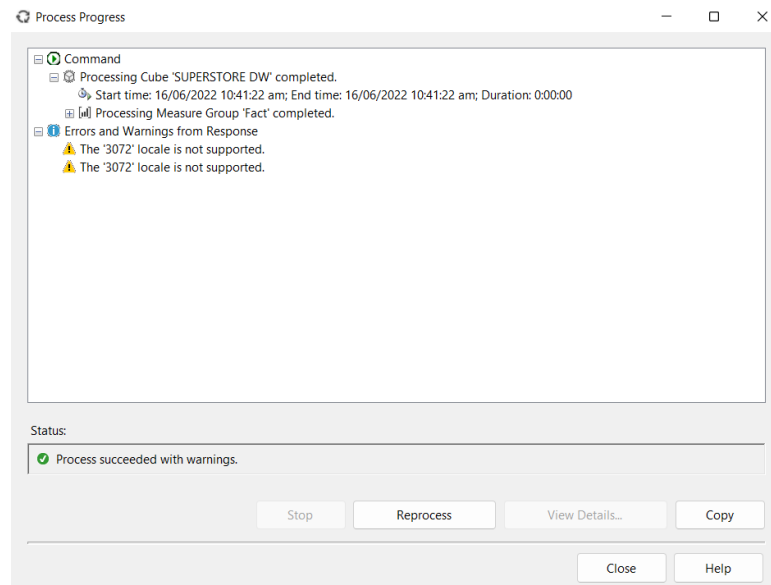
Chọn Yes để Deploy Project.



Chọn Run để chạy.

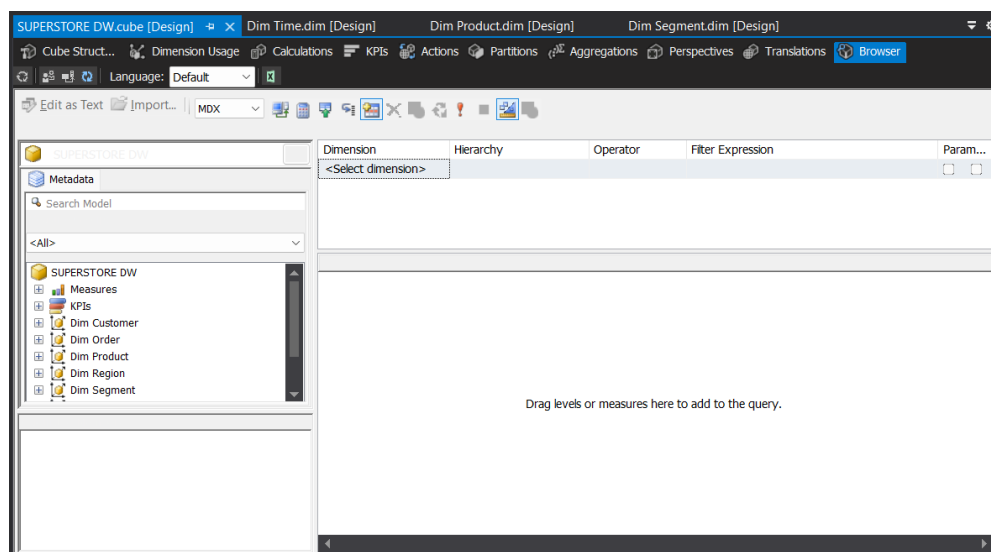


Hoàn tất. Chọn Close để đóng.



2.19. Truy vấn dữ liệu

Chọn Browser trên thanh công cụ để thực hiện việc truy vấn dữ liệu.



2.19.1. Truy vấn loại Roll Up: tính tổng lợi nhuận qua các năm.

Order Year	Profit
2014	37124.4334
2015	44722.5695
2016	63829.7175000001
2017	78407.2324999999

2.19.2. Truy vấn loại Drill Down: tính tổng lợi nhuận của từng thành phố, từng bang.

State Of US	City	Profit
Alabama	Auburn	477.4261
Alabama	Hoover	18.7173
Alabama	Mobile	645.2394
Alabama	Montgomery	774.3883
Alabama	Tuscaloosa	35.49
Arizona	Avondale	89.1836
Arizona	Bullhead City	6.7232
Arizona	Chandler	81.4605
Arizona	Gilbert	6.77640000000004
Arizona	Glendale	89.9372
Arizona	Mesa	-555.8948
Arizona	Peoria	-144.8053
Arizona	Phoenix	-1638.9157
Arizona	Scottsdale	-139.8533
Arizona	Sierra Vista	12.8127

2.19.3. Truy vấn loại Slice and Dice: hình thức giao hàng phổ biến nhất ở Texas.

Ship Mode	Order ID Count
First Class	117
Same Day	40
Second Class	172
Standard Class	463

2.19.4. Truy vấn loại Pivot: số đơn hàng của từng bang qua các năm.

Order Year	State Of US	Order ID Count
2014	Alabama	4
2014	Arizona	39
2014	Arkansas	11
2014	California	346
2014	Colorado	19
2014	Connecticut	4
2014	Delaware	24
2014	Florida	61
2014	Georgia	61
2014	Idaho	2
2014	Illinois	82
2014	Indiana	27
2014	Iowa	7
2014	Kansas	6
2014	Kentucky	22

2.20. Truy vấn MDX

2.20.1. Tổng số lợi nhuận theo từng năm

Lệnh truy vấn

```
--Câu 1 Tổng số lợi nhuận theo từng năm
select non empty [Measures].[Profit] on columns,
non empty [Dim Time].[Order Year].children on rows
from [SUPERSTORE DW]
```

Kết quả

Messages Results	
	Profit
2014	37124.4334
2015	44722.5695
2016	63829.7175
2017	78407.2325

2.20.2. Tổng số lợi nhuận của từng bang, từng thành phố

Lệnh truy vấn

```
--Câu 2 Tổng số lợi nhuận của từng bang, từng thành phố
select [Measures].[Profit] on columns,
non empty ([Dim Region].[State Of US].children, [Dim Region].[City].children) on rows
from [SUPERSTORE DW]
```

Kết quả

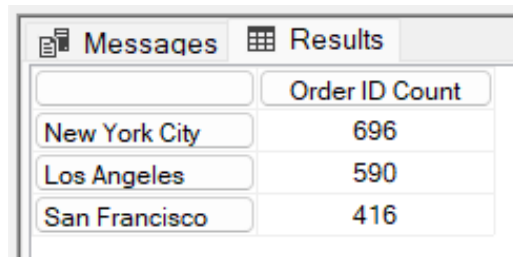
Messages Results		
		Profit
Alabama	Auburn	477.4261
Alabama	Hoover	18.7173
Alabama	Mobile	645.2394
Alabama	Montgomery	774.3883
Alabama	Tuscaloosa	35.49
Arizona	Avondale	89.1836
Arizona	Bullhead City	6.7232
Arizona	Chandler	81.4605
Arizona	Gilbert	6.77640000000004
Arizona	Glendale	89.9372
Arizona	Mesa	-555.8948
Arizona	Peoria	-144.8053
Arizona	Phoenix	-1638.9157
Arizona	Scottsdale	-139.8533
Arizona	Sierra Vista	13.8137
Arizona	Tempe	87.4108

2.20.3. Top 3 thành phố có số hóa đơn nhiều nhất

Lệnh truy vấn

```
--Câu 3 Top 3 thành phố có số hóa đơn nhiều nhất
select [Measures].[Order ID Count] on columns,
topcount([Dim Region].[City].children,3,[Measures].[Order ID Count]) on rows
from [SUPERSTORE DW]
```

Kết quả



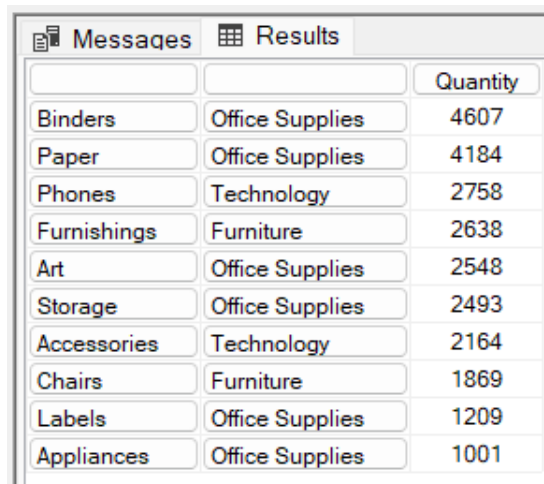
	Order ID Count
New York City	696
Los Angeles	590
San Francisco	416

2.20.4. Top 10 mặt hàng (danh mục hàng) bán chạy nhất (số lượng bán nhiều nhất)

Lệnh truy vấn

```
--Câu 4 Top 10 mặt hàng (danh mục hàng) bán chạy nhất (số lượng bán nhiều nhất)
select [Measures].[Quantity] on columns,
(topcount([Dim Product].[Sub- Category].children,10,[Measures].[Quantity]),[Dim Product].[Category].children) on rows
from [OLAP SUPERSTORE DW]
```

Kết quả



		Quantity
Binders	Office Supplies	4607
Paper	Office Supplies	4184
Phones	Technology	2758
Furnishings	Furniture	2638
Art	Office Supplies	2548
Storage	Office Supplies	2493
Accessories	Technology	2164
Chairs	Furniture	1869
Labels	Office Supplies	1209
Appliances	Office Supplies	1001

2.20.5. Số đơn hàng của từng bang qua các năm

Lệnh truy vấn

```
--Câu 5 Số đơn hàng của từng bang qua các năm
select non empty [Measures].[Order ID Count] on columns,
non empty ([Dim Time].[Order Year].children, [Dim Region].[State Of US].children) on rows
from [OLAP SUPERSTORE DW]
```

Kết quả

Messages		Results
		Order ID Count
2014	Alabama	4
2014	Arizona	39
2014	Arkansas	11
2014	California	346
2014	Colorado	19
2014	Connecticut	4
2014	Delaware	24
2014	Florida	61
2014	Georgia	61
2014	Idaho	2
2014	Illinois	82
2014	Indiana	27
2014	Iowa	7
2014	Kansas	6
2014	Kentucky	22
2014	Louisiana	10

2.20.6. Hình thức giao hàng phổ biến nhất ở Texas

Lệnh truy vấn

```
--Câu 6 Hình thức giao hàng phổ biến nhất ở Texas
select [Measures].[Order ID Count] on columns,
non empty [Dim Ship Mode].[Ship Mode].children on rows
from [OLAP SUPERSTORE DW]
where [Dim Region].[State Of US].&[Texas];
```

Kết quả

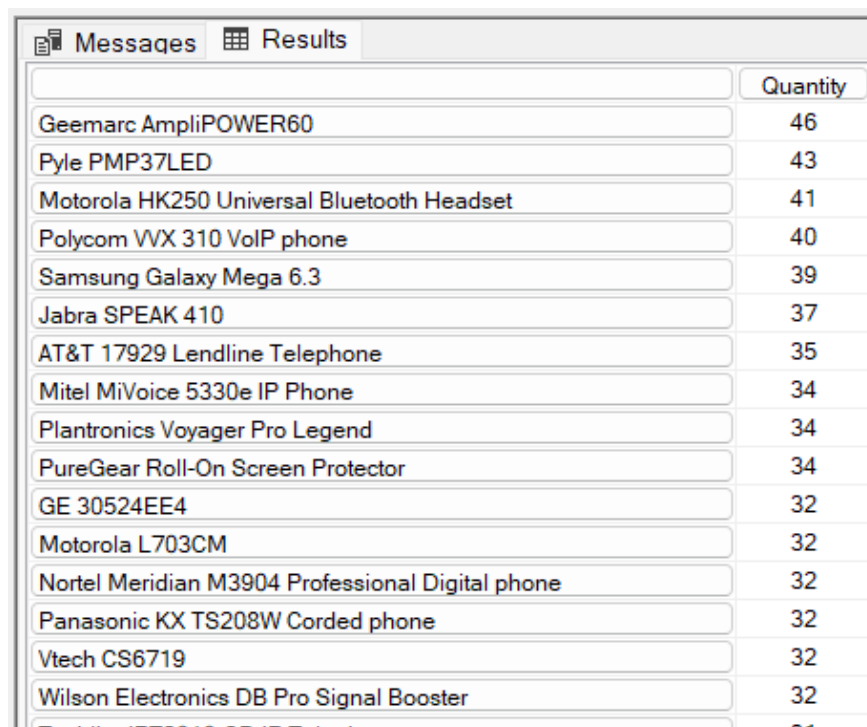
Messages		Results
		Order ID Count
	First Class	117
	Same Day	40
	Second Class	172
	Standard Class	463

2.20.7. Loại điện thoại (Phones) nào được bán nhiều nhất (sắp xếp giảm dần)

Lệnh truy vấn

```
--Câu 7 Loại điện thoại (Phones) nào được bán nhiều nhất (sx giảm dần)
select [Measures].[Quantity] on columns,
non empty order([Dim Product].[Product Name].children,[Measures].[Quantity],DESC) on rows
from [OLAP SUPERSTORE DW]
where [Dim Product].[Sub- Category].&[Phones];
```

Kết quả



The screenshot shows a BI tool interface with two tabs: 'Messages' and 'Results'. The 'Results' tab is active, displaying a table with two columns: an unnamed column for product names and a column named 'Quantity'. The table lists 18 products with their corresponding quantities.

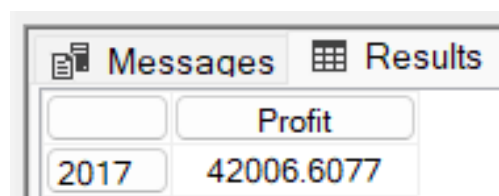
	Quantity
Geemarc AmpliPOWER60	46
Pyle PMP37LED	43
Motorola HK250 Universal Bluetooth Headset	41
Polycom VVX 310 VoIP phone	40
Samsung Galaxy Mega 6.3	39
Jabra SPEAK 410	37
AT&T 17929 Lendline Telephone	35
Mitel MiVoice 5330e IP Phone	34
Plantronics Voyager Pro Legend	34
PureGear Roll-On Screen Protector	34
GE 30524EE4	32
Motorola L703CM	32
Nortel Meridian M3904 Professional Digital phone	32
Panasonic KX TS208W Corded phone	32
Vtech CS6719	32
Wilson Electronics DB Pro Signal Booster	32

2.20.8. Năm có doanh thu hàng công nghệ (Technology) cao nhất

Lệnh truy vấn

```
--Câu 8 Năm có doanh thu hàng công nghệ (Technology) cao nhất
select [Measures].[Profit] on columns,
topcount([Dim Time].[Order Year].children,1,[Measures].[Profit]) on rows
from [OLAP SUPERSTORE DW]
where [Dim Product].[Category].&[Technology];
```

Kết quả



The screenshot shows a BI tool interface with two tabs: 'Messages' and 'Results'. The 'Results' tab is active, displaying a table with two columns: 'Profit' and an unnamed column for the year. The table shows the top profit for Technology products in 2017.

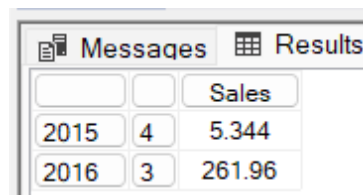
Profit	
42006.6077	2017

2.20.9. Số tiền mua hàng của Claire Gute qua từng quý trong năm

Lệnh truy vấn

```
--Câu 9 Số tiền mua hàng của Claire Gute qua từng quý trong năm
select [Measures].[Sales] on columns,
non empty ([Dim Time].[Order Year].children, [Dim Time].[Order Quarter].children) on rows
from [OLAP SUPERSTORE DW]
where [Dim Customer].[Customer Name].&[Claire Gute];
```

Kết quả



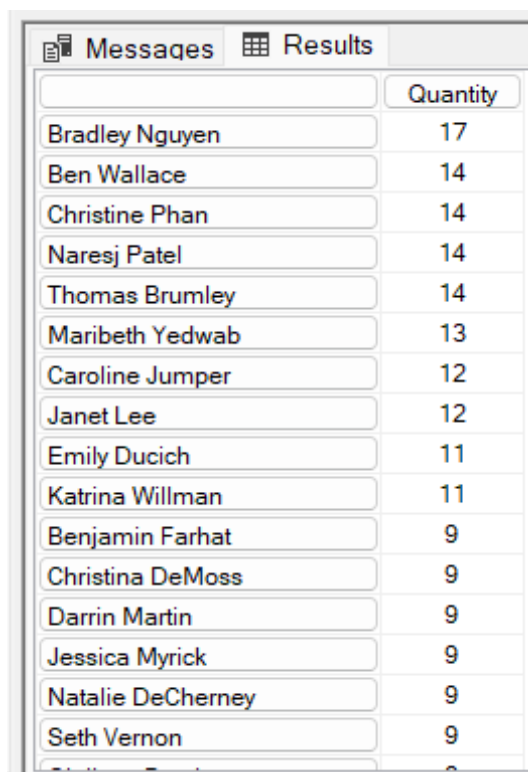
		Sales
2015	4	5.344
2016	3	261.96

2.20.10. Ở New York City thì Consumer nào mua nhiều đồ nội thất (Furniture) nhất

Lệnh truy vấn

```
--Câu 10 Ở New York City thì Consumer nào mua nhiều đồ nội thất (Furniture) nhất
select [Measures].[Quantity] on columns,
non empty order([Dim Customer].[Customer Name].children,[Measures].[Quantity],DESC) on rows
from [OLAP SUPERSTORE DW]
where ([Dim Region].[City].&[New York City],[Dim Product].[Category].&[Furniture]);
```

Kết quả



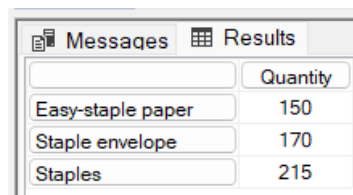
	Quantity
Bradley Nguyen	17
Ben Wallace	14
Christine Phan	14
Naresj Patel	14
Thomas Brumley	14
Maribeth Yedwab	13
Caroline Jumper	12
Janet Lee	12
Emily Ducich	11
Katrina Willman	11
Benjamin Farhat	9
Christina DeMoss	9
Darrin Martin	9
Jessica Myrick	9
Natalie DeCherney	9
Seth Vernon	9

2.20.11. Liệt kê các sản phẩm có số lượng bán > 100

Lệnh truy vấn

```
--Câu 11 Liệt kê các sản phẩm có số lượng bán > 100
select [Measures].[Quantity] on columns,
[Dim Product].[Product Name].children
having [Measures].[Quantity]>100 on rows
from [OLAP SUPERSTORE DW]
```

Kết quả



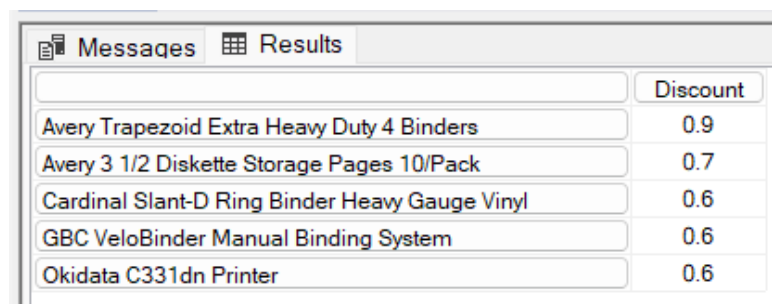
	Quantity
Easy-staple paper	150
Staple envelope	170
Staples	215

2.20.12. Top 5 sản phẩm được giảm giá nhiều nhất ở bang Washington

Lệnh truy vấn

```
--Câu 12 Top 5 sản phẩm được giảm giá nhiều nhất ở bang Washington
select [Measures].[Discount] on columns,
topcount([Dim Product].[Product Name].children,5,[Measures].[Discount]) on rows
from [OLAP SUPERSTORE DW]
where [Dim Region].[State Of US].&[Washington];
```

Kết quả



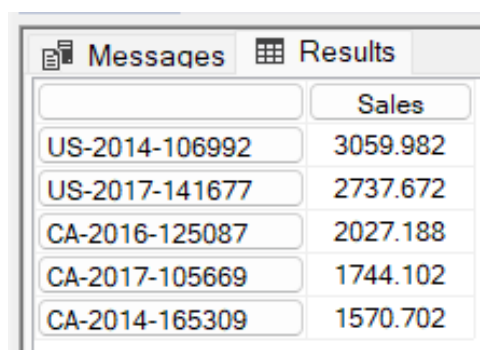
	Discount
Avery Trapezoid Extra Heavy Duty 4 Binders	0.9
Avery 3 1/2 Diskette Storage Pages 10/Pack	0.7
Cardinal Slant-D Ring Binder Heavy Gauge Vinyl	0.6
GBC VeloBinder Manual Binding System	0.6
Okidata C331dn Printer	0.6

2.20.13. Top 5 hóa đơn có đơn giá cao nhất ở thành phố Houston

Lệnh truy vấn

```
--Câu 13 Top 5 hóa đơn có đơn giá cao nhất ở thành phố Houston
select [Measures].[Sales] on columns,
topcount([Dim Order].[Order ID].children,5,[Measures].[Sales]) on rows
from [OLAP SUPERSTORE DW]
where [Dim Region].[City].&[Houston];
```

Kết quả



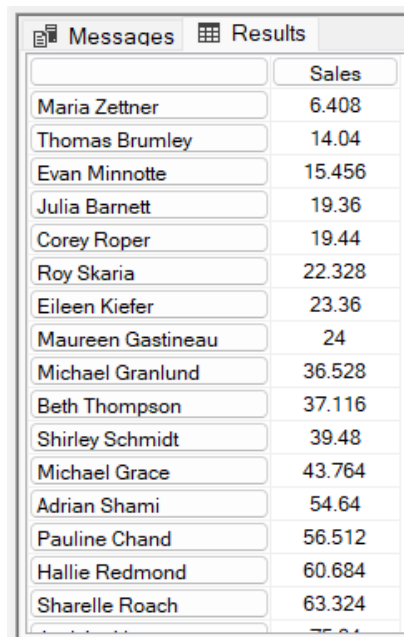
	Sales
US-2014-106992	3059.982
US-2017-141677	2737.672
CA-2016-125087	2027.188
CA-2017-105669	1744.102
CA-2014-165309	1570.702

2.20.14. Tổng giá tiền các Home Office bỏ ra để mua văn phòng phẩm (Office Supplies), sắp xếp tăng dần

Lệnh truy vấn

```
--Câu 14 Tổng giá tiền các Home Office bỏ ra để mua văn phòng phẩm (Office Supplies), sắp xếp tăng dần
select [Measures].[Sales] on columns,
non empty order([Dim Customer].[Customer Name].children,[Measures].[Sales],ASC) on rows
from [OLAP SUPERSTORE DW]
where ([Dim Segment].[Segment].&[Home Office],[Dim Product].[Category].&[Office Supplies]);
```

Kết quả



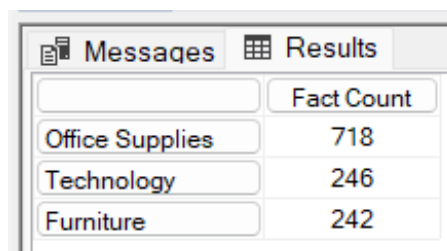
Messages	Results
	Sales
Maria Zettner	6.408
Thomas Brumley	14.04
Evan Minnotte	15.456
Julia Barnett	19.36
Corey Roper	19.44
Roy Skaria	22.328
Eileen Kiefer	23.36
Maureen Gastineau	24
Michael Granlund	36.528
Beth Thompson	37.116
Shirley Schmidt	39.48
Michael Grace	43.764
Adrian Shami	54.64
Pauline Chand	56.512
Hallie Redmond	60.684
Sharelle Roach	63.324

2.20.15. Loại hàng nào được vận chuyển bằng phương thức First Class nhiều nhất (sắp xếp giảm dần)

Lệnh truy vấn

```
--Câu 15 Loại hàng nào được vận chuyển bằng phương thức First Class nhiều nhất (sx giảm dần)
select [Measures].[Fact Count] on columns,
non empty order([Dim Product].[Category].children,[Measures].[Fact Count],DESC) on rows
from [OLAP SUPERSTORE DW]
where [Dim Ship Mode].[Ship Mode].&[First Class];
```

Kết quả

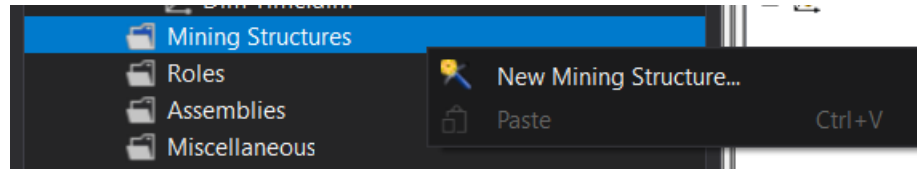


Messages	Results
	Fact Count
Office Supplies	718
Technology	246
Furniture	242

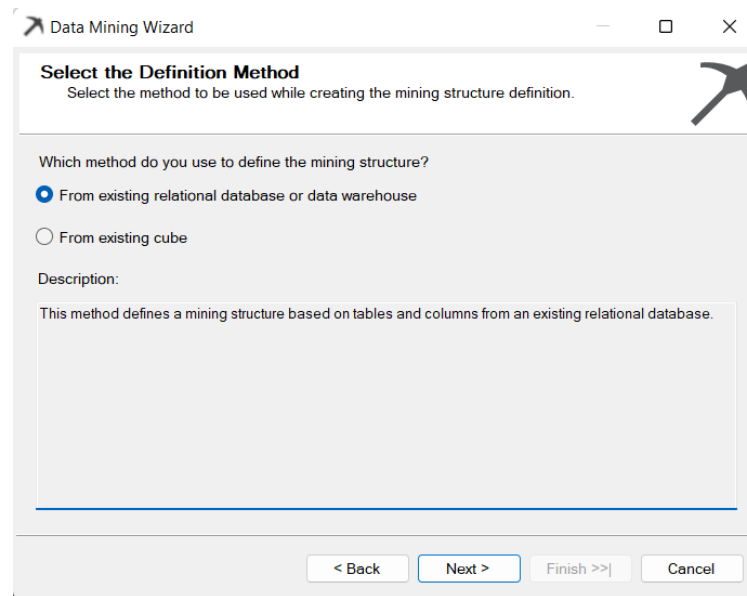
4. KHAI PHÁ DỮ LIỆU (DATA MINING)

2.21. Tạo Project

Chuột phải vào Mining Structures và chọn New Mining Structure.

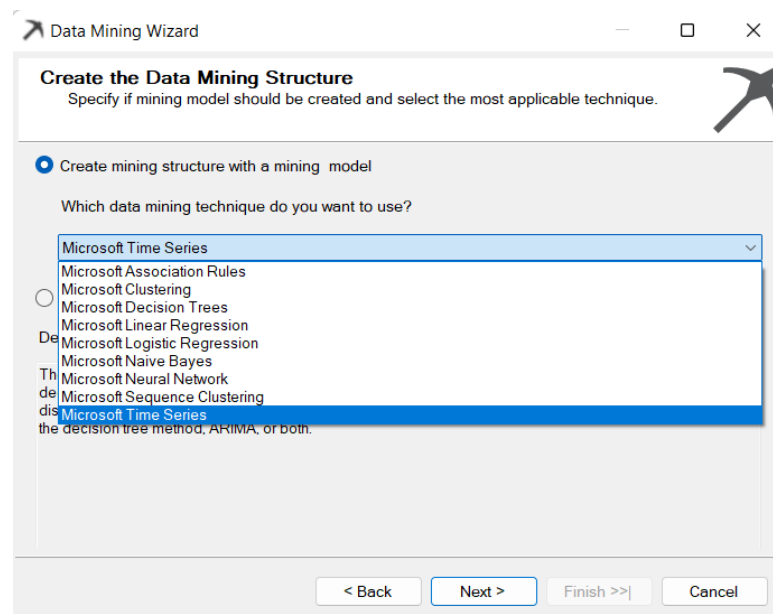


Chọn Next

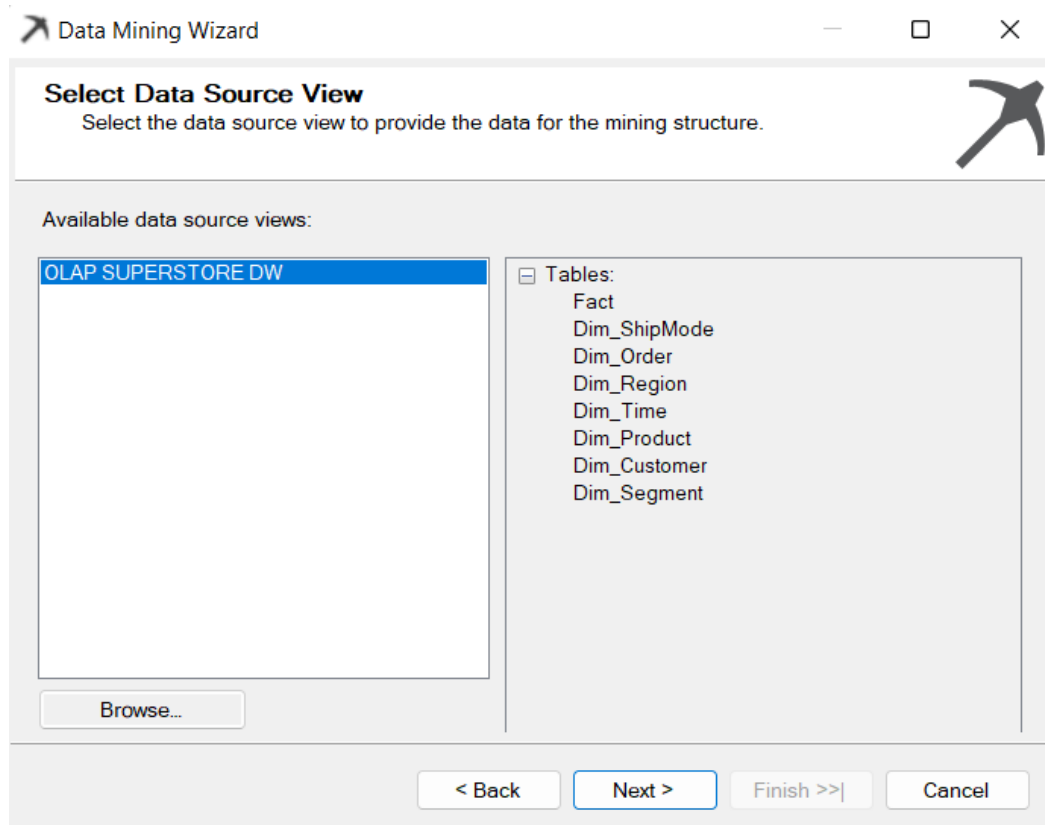


2.22. Khai phá với Time Series

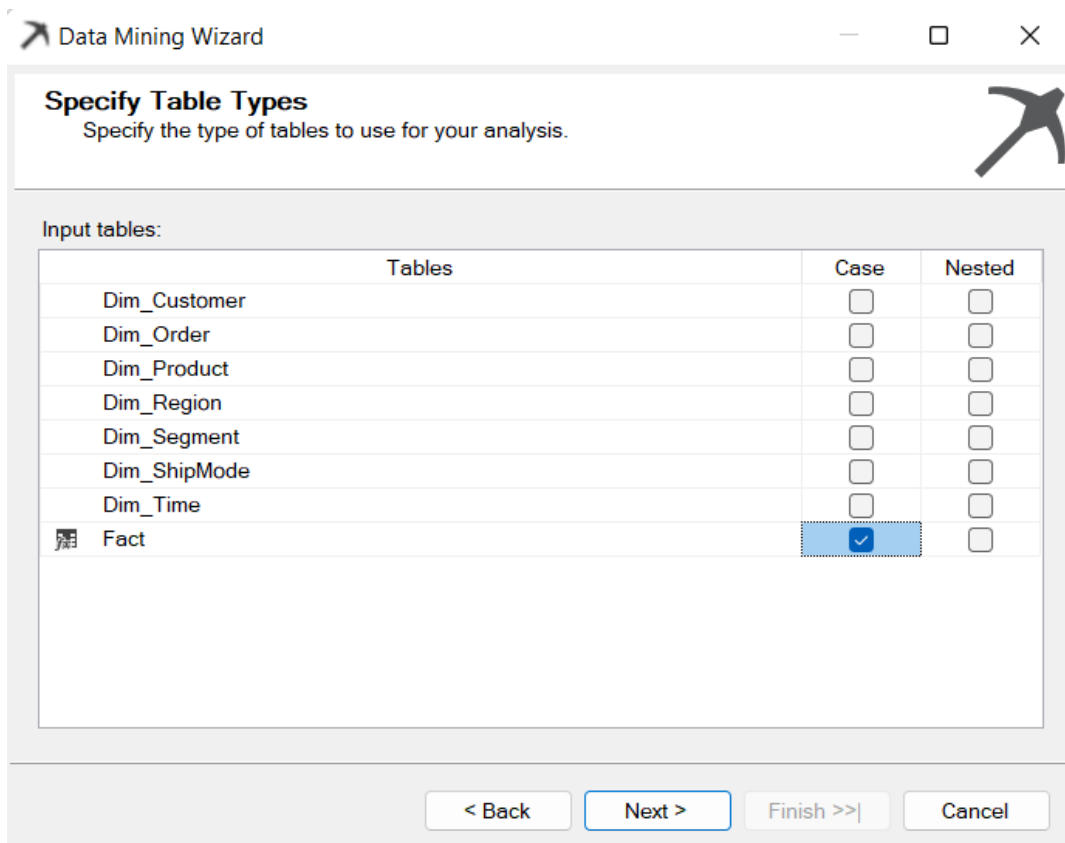
Chọn Time Series và nhấn Next.



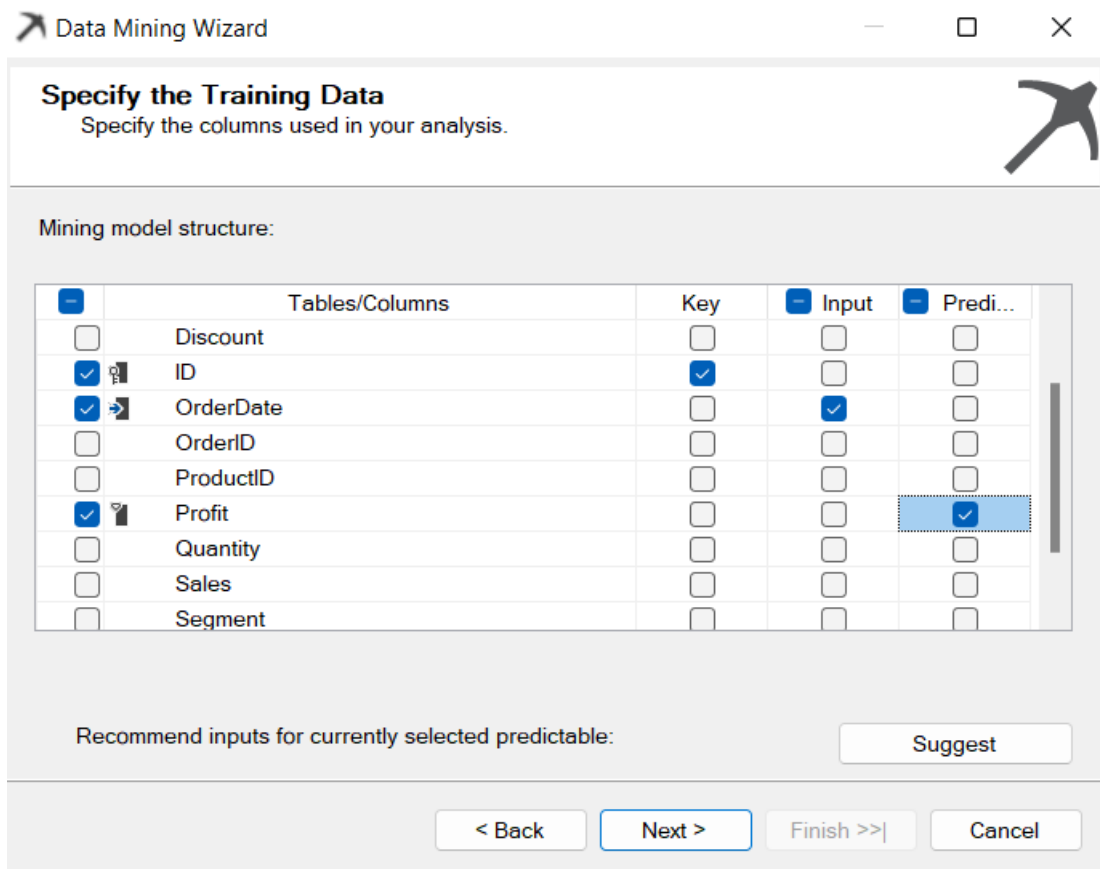
Chọn Next.



Input table ta chọn bảng Fact và nhấn Next



Ở cột Input chọn OrderDate làm dữ liệu đầu vào và chọn Profit là Predictable. Tiếp tục nhấn Next và Finish để kết thúc.



Data Mining Wizard

Specify the Training Data
Specify the columns used in your analysis.

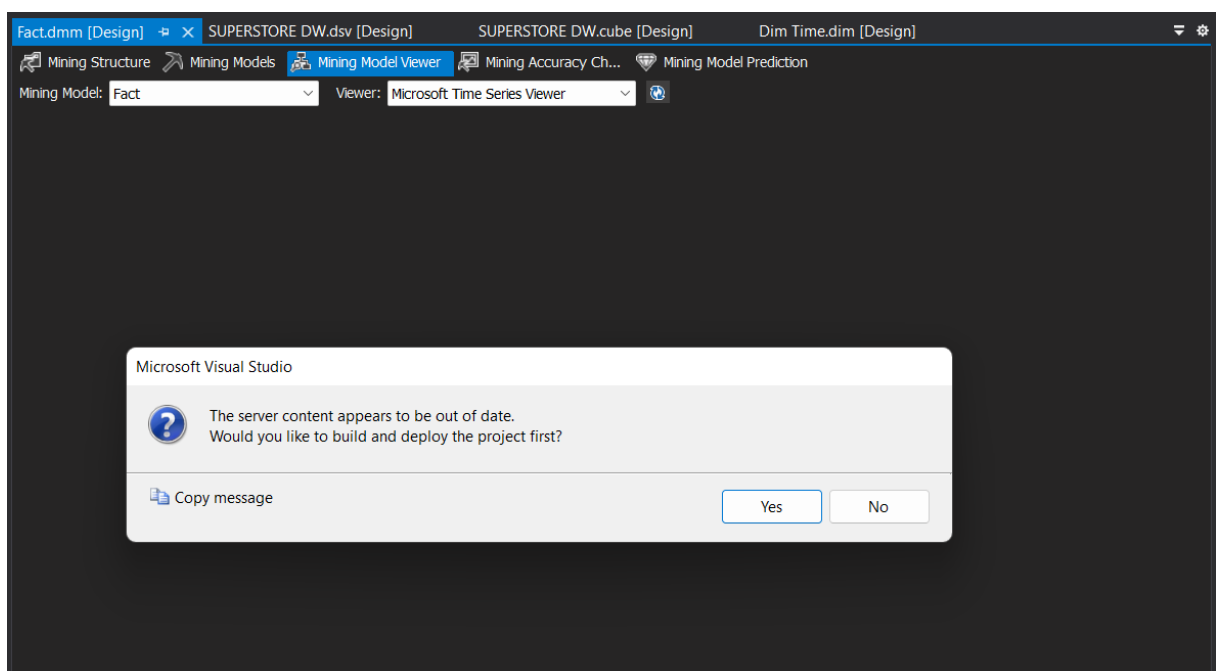
Mining model structure:

	Tables/Columns	Key	Input	Predictable
<input type="checkbox"/>	Discount	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	ID	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	OrderDate	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	OrderID	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	ProductID	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Profit	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Quantity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Sales	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Segment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

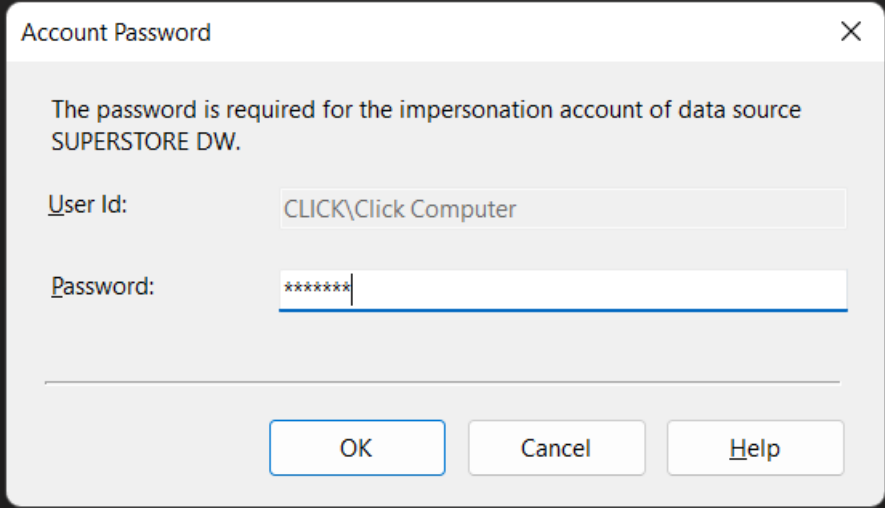
Recommend inputs for currently selected predictable: Suggest

< Back Next > Finish >> Cancel

Chọn Mining Model Viewer trên thanh công cụ và nhấn Yes.



Nhập mật khẩu nếu có yêu cầu và nhấn OK.



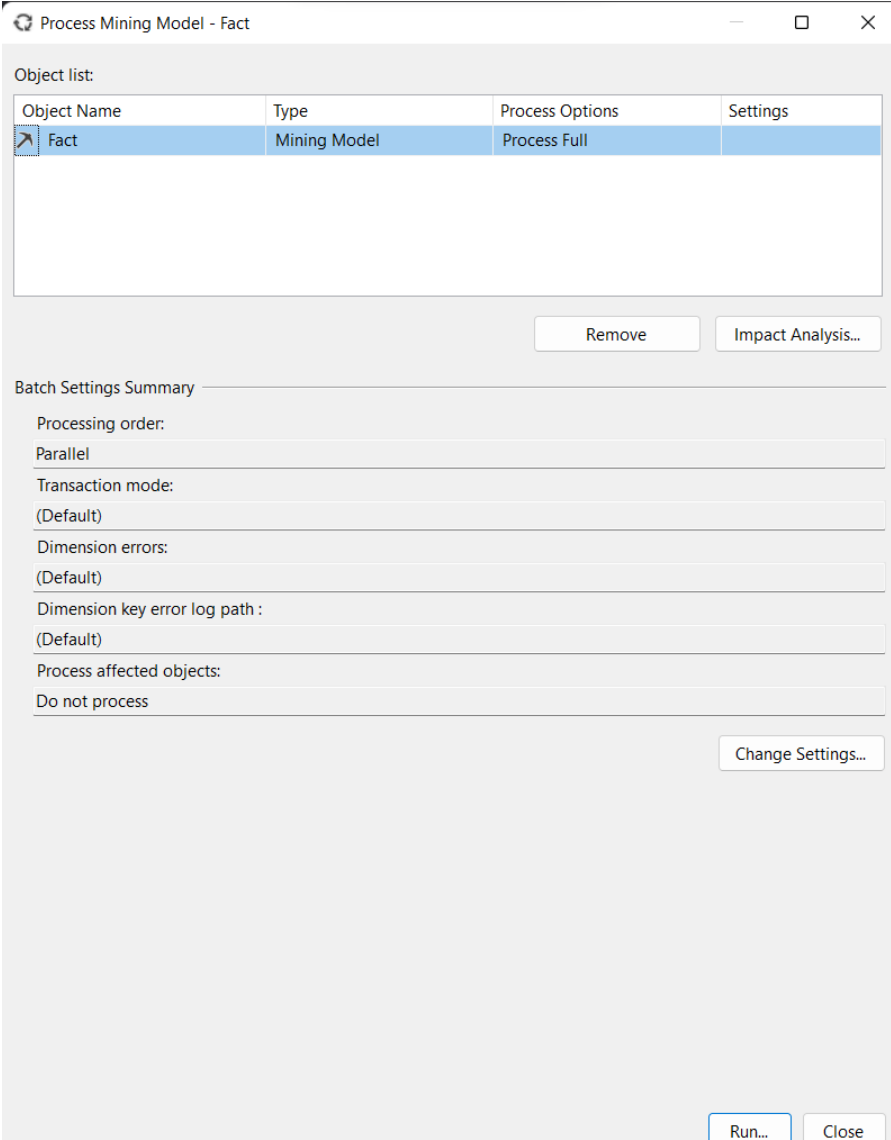
The password is required for the impersonation account of data source SUPERSTORE DW.

User Id: CLICK\Click Computer

Password: *****

OK Cancel Help

Tiếp tục nhấn Yes và chọn Run để chạy.



Process Mining Model - Fact

Object list:

Object Name	Type	Process Options	Settings
Fact	Mining Model	Process Full	

Remove Impact Analysis...

Batch Settings Summary

Processing order:
Parallel

Transaction mode:
(Default)

Dimension errors:
(Default)

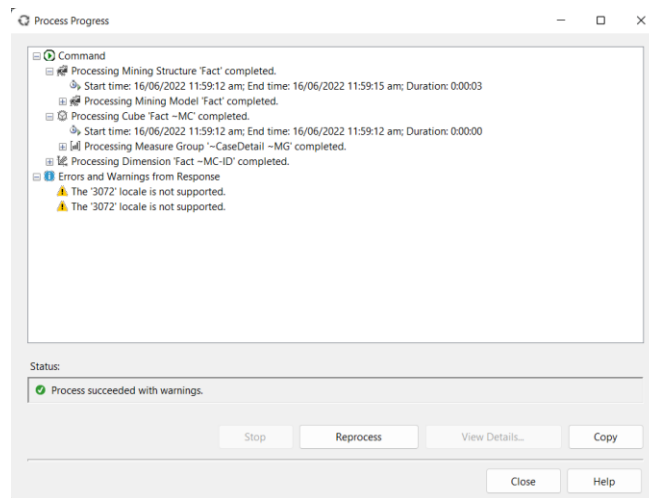
Dimension key error log path :
(Default)

Process affected objects:
Do not process

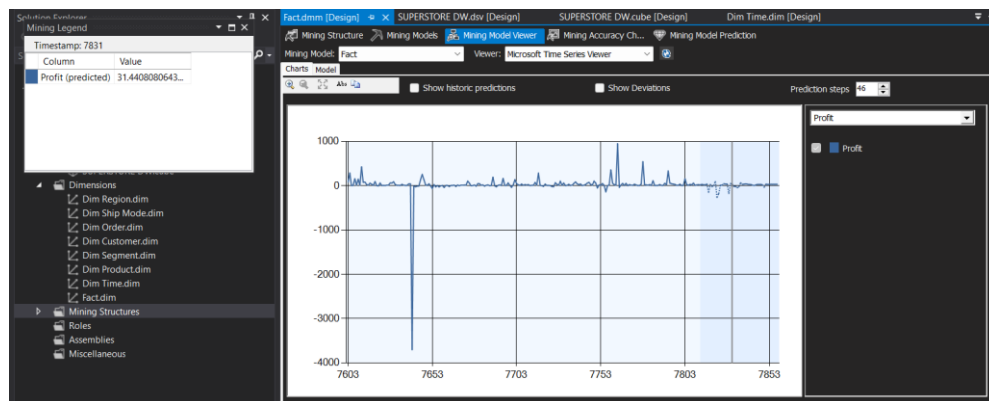
Change Settings...

Run... Close

Chạy thành công, nhấn Close để đóng.



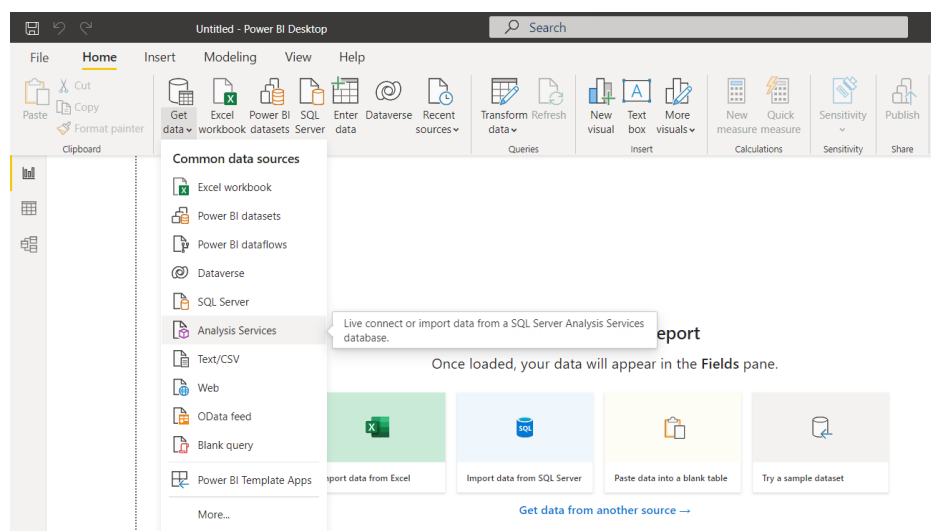
Tăng Prediction Steps để xem dự đoán lợi nhuận của các ngày kế tiếp.



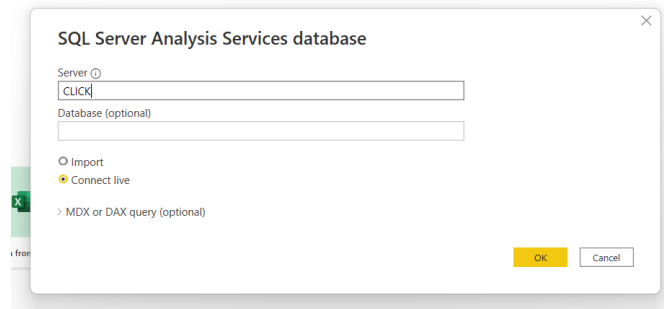
5. POWER BI

2.23. Import dữ liệu với Analysis Services Database

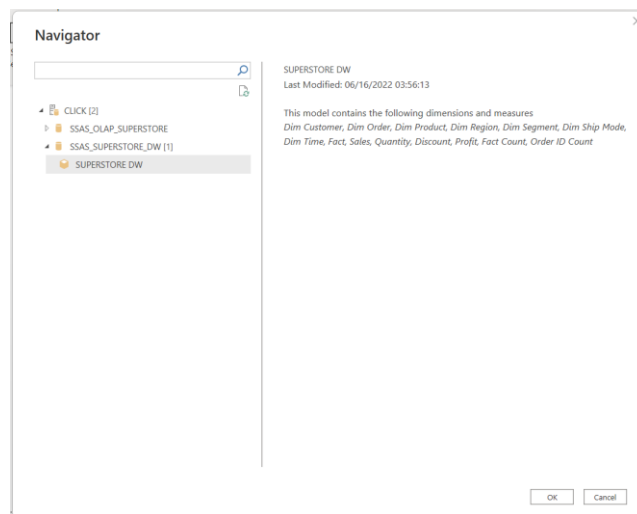
Mở phần mềm Power BI, thực hiện Import Data với project SSAS.



Nhập tên Server và nhấn OK.



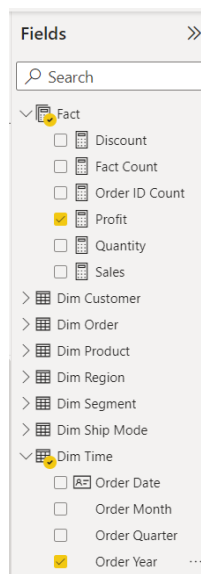
Chọn database SSAS và nhấn OK.



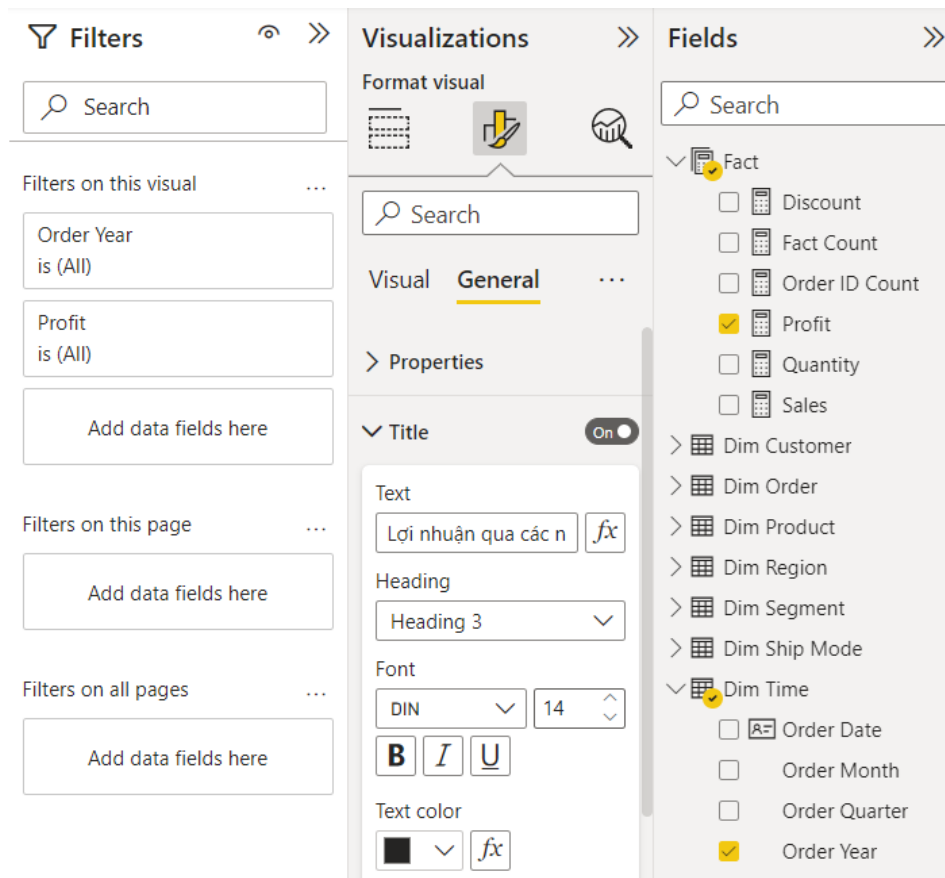
2.24. Thực hiện các báo cáo

2.24.1. Lợi nhuận qua các năm

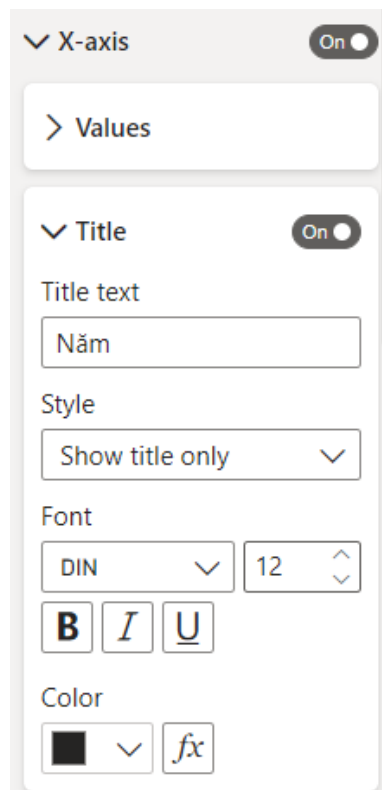
Lựa chọn các thuộc tính cần thiết. Ở đây là Profit và Order Year.

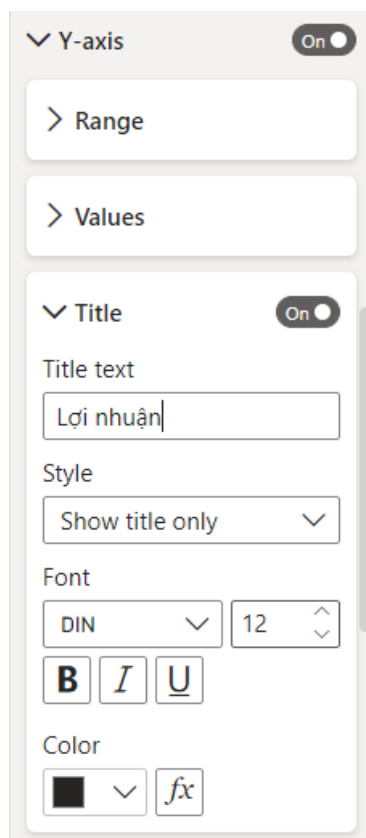


Đổi tên biểu đồ thành Lợi nhuận qua các năm

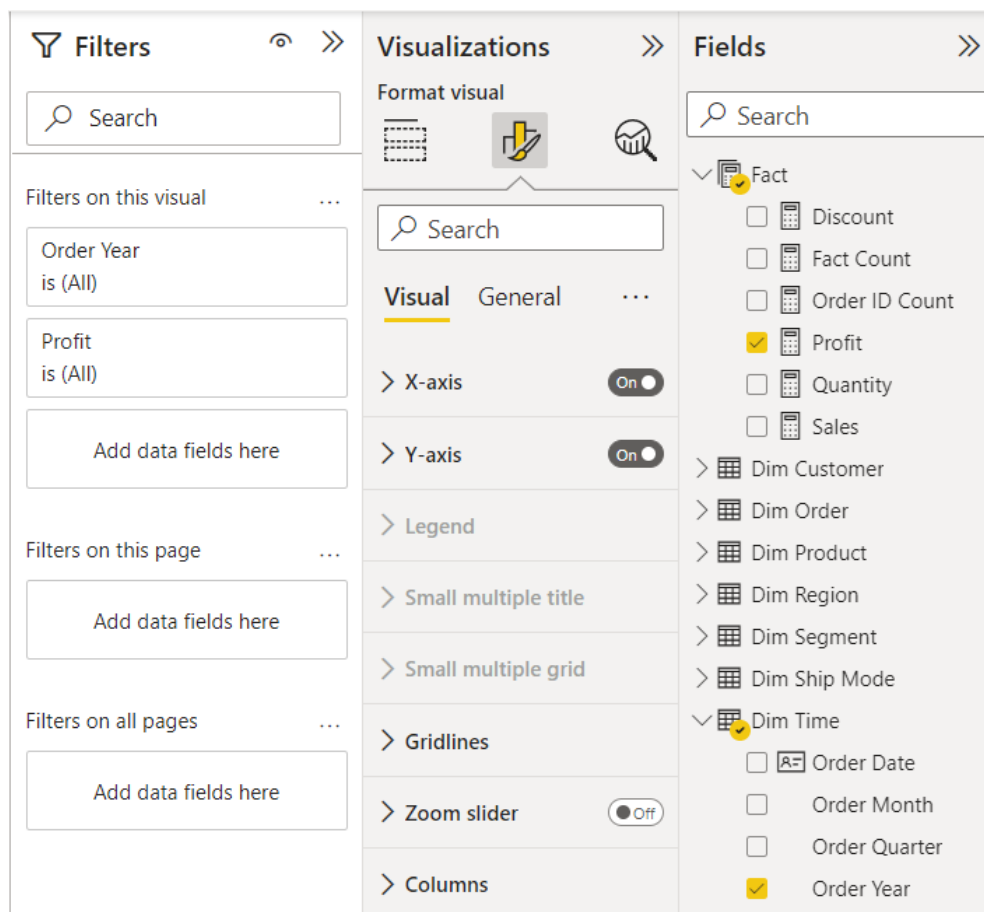


Tương tự thực hiện đổi tên cột X và cột Y

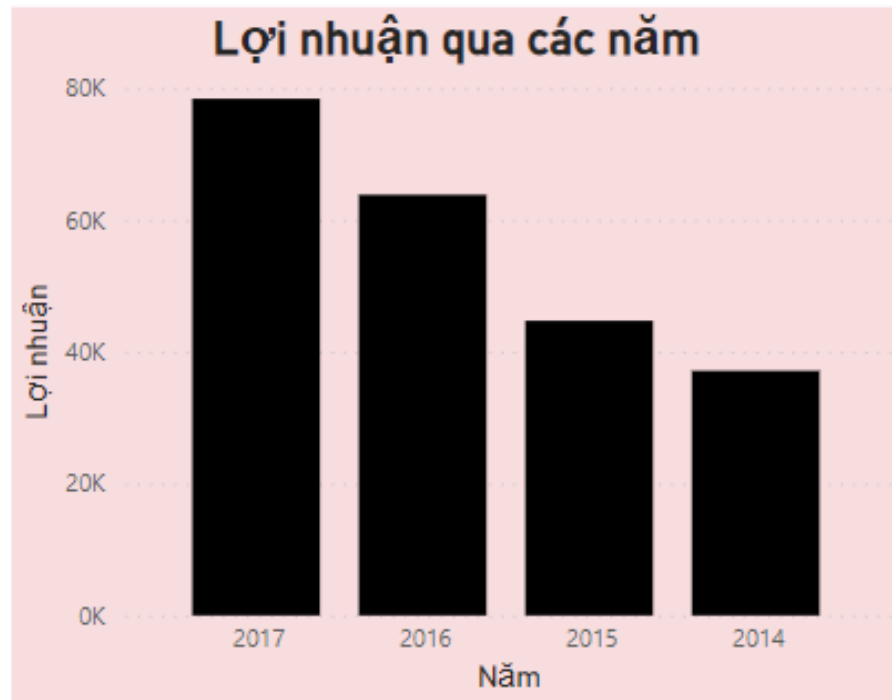




Chọn Format visual để thiết kế chỉnh sửa lại biểu đồ (thay đổi font chữ, màu sắc,)

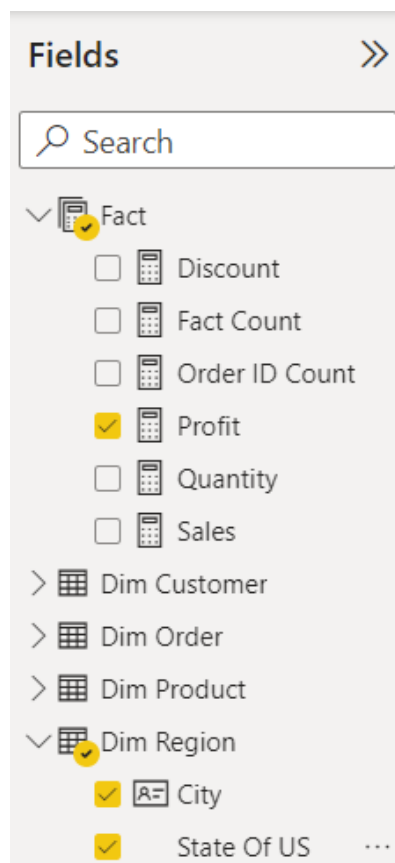


Kết quả

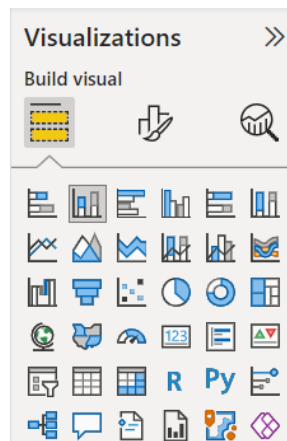


2.24.2. Lợi nhuận của từng bang, thành phố

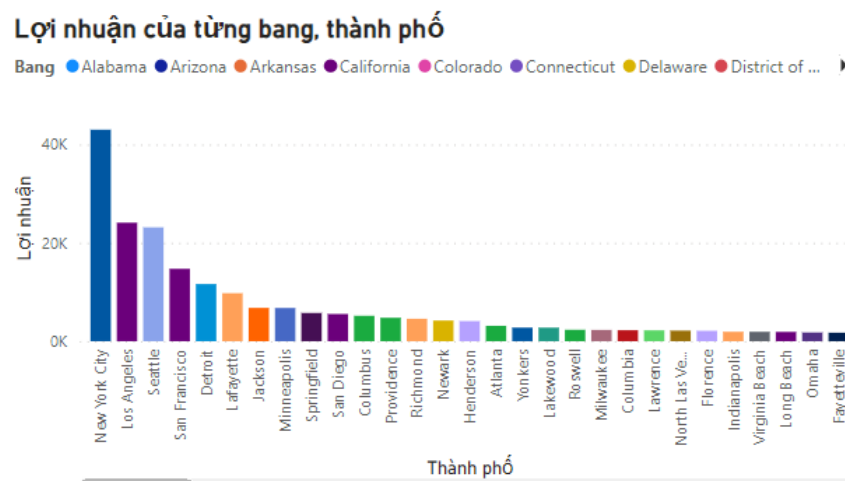
Lựa chọn các thuộc tính.



Lựa chọn loại biểu đồ cho phù hợp

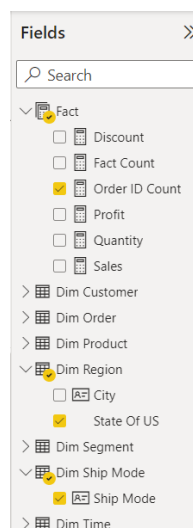


Thực hiện chỉnh sửa lại cho phù hợp. Kết quả.

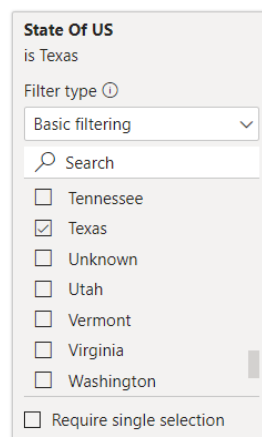


2.24.3. Hình thức giao hàng phổ biến ở Texas

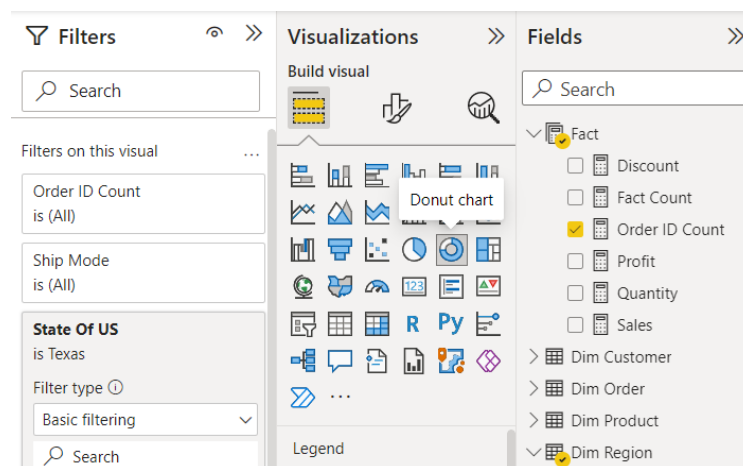
Lựa chọn các thuộc tính cần thiết.



Lọc theo bang Texas.



Chọn loại biểu đồ và chỉnh sửa cho phù hợp.



Kết quả.

