**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

🙤🙧🟍🙥🙦



**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**

**NHẬN XÉT CỦA GIẢNG VIÊN**

**…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………**

**…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………**

**…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………  
…………………………………………………………………………………**

MỤC LỤC

[1. TỔNG QUAN ĐỀ TÀI 1](#_Toc106277892)

[1.1. Mô tả dữ liệu gốc 1](#_Toc106277893)

[1.2. Mô tả dữ liệu đã xử lý 2](#_Toc106277894)

[1.3. Thiết kế kho dữ liệu 3](#_Toc106277895)

[1.3.1. Lược đồ hình sao 3](#_Toc106277896)

[1.3.2. Bảng Fact 3](#_Toc106277897)

[1.3.3. Bảng Dim 4](#_Toc106277898)

[2. XÂY DỰNG KHO DỮ LIỆU (SSIS) 5](#_Toc106277899)

[2.1. Tạo Database 5](#_Toc106277900)

[2.2. Tạo Project và thiết lập kết nối 5](#_Toc106277901)

[2.3. Tạo Data Flow Task 8](#_Toc106277902)

[2.4. Tạo Flag File Source kết nối với file dữ liệu 9](#_Toc106277903)

[2.5. Rẽ nhánh và lọc data trùng 11](#_Toc106277904)

[2.6. Tạo các bảng Dimension 12](#_Toc106277905)

[2.7. Tạo bảng Fact 14](#_Toc106277906)

[2.8. Tạo Execute SQL Task khóa ngoại cho bảng Fact và các bảng Dim 15](#_Toc106277907)

[2.9. Tạo bảng Dim\_Time và các ràng buộc 17](#_Toc106277908)

[2.10. Tạo Execute SQL Task để xóa ràng buộc, xóa bảng Fact và các bảng Dim 17](#_Toc106277909)

[2.11. Thực thi project 18](#_Toc106277910)

[3. PHÂN TÍCH KHO DỮ LIỆU (SSAS) 19](#_Toc106277911)

[2.12. Tạo Project 19](#_Toc106277912)

[2.13. Tạo Data Source 19](#_Toc106277913)

[2.14. Tạo Data Source View 21](#_Toc106277914)

[2.15. Tạo Cube 23](#_Toc106277915)

[2.16. Tạo thêm Measure 24](#_Toc106277916)

[2.17. Chỉnh sửa thuộc tính các bảng Dim 25](#_Toc106277917)

[2.18. Thực thi project 27](#_Toc106277918)

[2.19. Truy vấn dữ liệu 28](#_Toc106277919)

[2.19.1. Truy vấn loại Roll Up: tính tổng lợi nhuận qua các năm. 28](#_Toc106277920)

[2.19.2. Truy vấn loại Drill Down: tính tổng lợi nhuận của từng thành phố, từng bang. 29](#_Toc106277921)

[2.19.3. Truy vấn loại Slice and Dice: hình thức giao hàng phổ biến nhất ở Texas. 29](#_Toc106277922)

[2.19.4. Truy vấn loại Pivot: số đơn hàng của từng bang qua các năm. 29](#_Toc106277923)

[2.20. Truy vấn MDX 30](#_Toc106277924)

[2.20.1. Tổng số lợi nhuận theo từng năm 30](#_Toc106277925)

[2.20.2. Tổng số lợi nhuận của từng bang, từng thành phố 30](#_Toc106277926)

[2.20.3. Top 3 thành phố có số hóa đơn nhiều nhất 31](#_Toc106277927)

[2.20.4. Top 10 mặt hàng (danh mục hàng) bán chạy nhất (số lượng bán nhiều nhất) 31](#_Toc106277928)

[2.20.5. Số đơn hàng của từng bang qua các năm 31](#_Toc106277929)

[2.20.6. Hình thức giao hàng phổ biến nhất ở Texas 32](#_Toc106277930)

[2.20.7. Loại điện thoại (Phones) nào được bán nhiều nhất (sắp xếp giảm dần) 32](#_Toc106277931)

[2.20.8. Năm có doanh thu hàng công nghệ (Technology) cao nhất 33](#_Toc106277932)

[2.20.9. Số tiền mua hàng của Claire Gute qua từng quý trong năm 33](#_Toc106277933)

[2.20.10. Ở New York City thì Consumer nào mua nhiều đồ nội thất (Furniture) nhất 34](#_Toc106277934)

[2.20.11. Liệt kê các sản phẩm có số lượng bán > 100 34](#_Toc106277935)

[2.20.12. Top 5 sản phẩm được giảm giá nhiều nhất ở bang Washington 35](#_Toc106277936)

[2.20.13. Top 5 hóa đơn có đơn giá cao nhất ở thành phố Houston 35](#_Toc106277937)

[2.20.14. Tổng giá tiền các Home Office bỏ ra để mua văn phòng phẩm (Office Supplies), sắp xếp tăng dần 36](#_Toc106277938)

[2.20.15. Loại hàng nào được vận chuyển bằng phương thức First Class nhiều nhất (sắp xếp giảm dần) 36](#_Toc106277939)

[4. KHAI PHÁ DỮ LIỆU (DATA MINING) 37](#_Toc106277940)

[2.21. Tạo Project 37](#_Toc106277941)

[2.22. Khai phá với Time Series 37](#_Toc106277942)

[5. POWER BI 41](#_Toc106277943)

[2.23. Import dữ liệu với Analysis Services Database 41](#_Toc106277944)

[2.24. Thực hiện các báo cáo 42](#_Toc106277945)

[2.24.1. Lợi nhuận qua các năm 42](#_Toc106277946)

[2.24.2. Lợi nhuận của từng bang, thành phố 45](#_Toc106277947)

[2.24.3. Hình thức giao hàng phổ biến ở Texas 46](#_Toc106277948)

# 1. TỔNG QUAN ĐỀ TÀI

## 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 |

## Mô tả dữ liệu đã xử lý

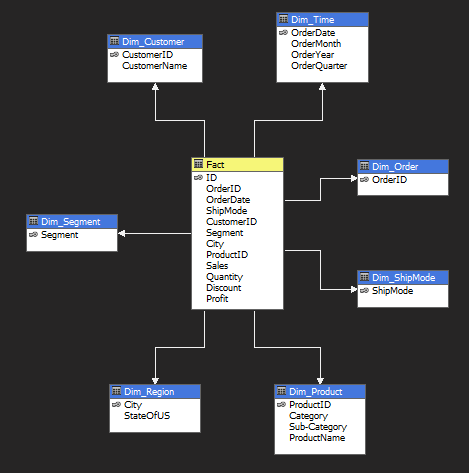
**Lý do loại bỏ các cột dữ liệu:**

* 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 |
| **Quatity** | Số lượng mua |
| **Discount** | Giảm giá |
| **Profit** | Lợi nhuận |

## Thiết kế kho dữ liệu

### Lược đồ hình sao



### 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 |

### 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.

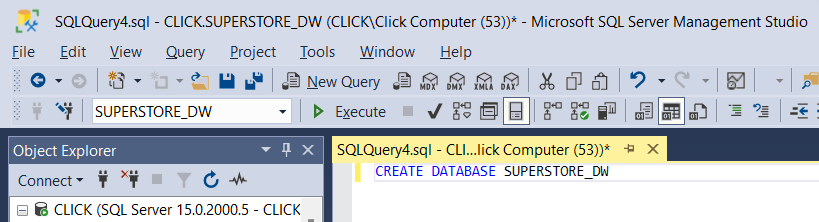
**Dim\_Region**

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

# XÂY DỰNG KHO DỮ LIỆU (SSIS)

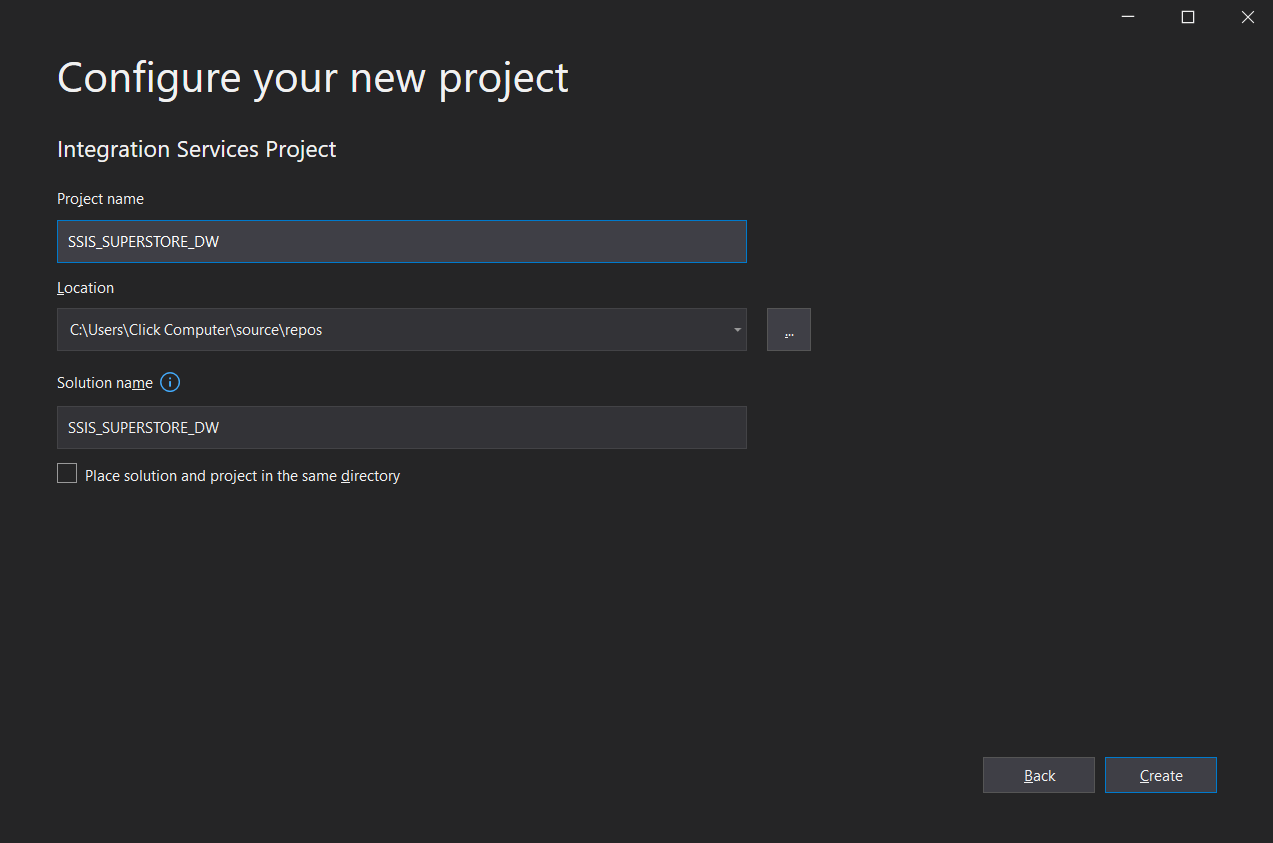
## Tạo Database

Tạo mới một Database đặt tên là SUPERSTORE\_DW.

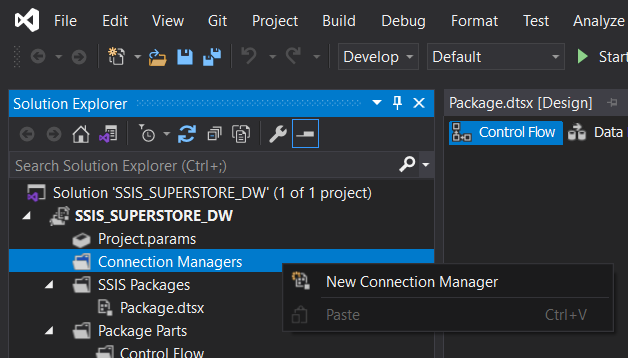


## Tạo Project và thiết lập kết nối

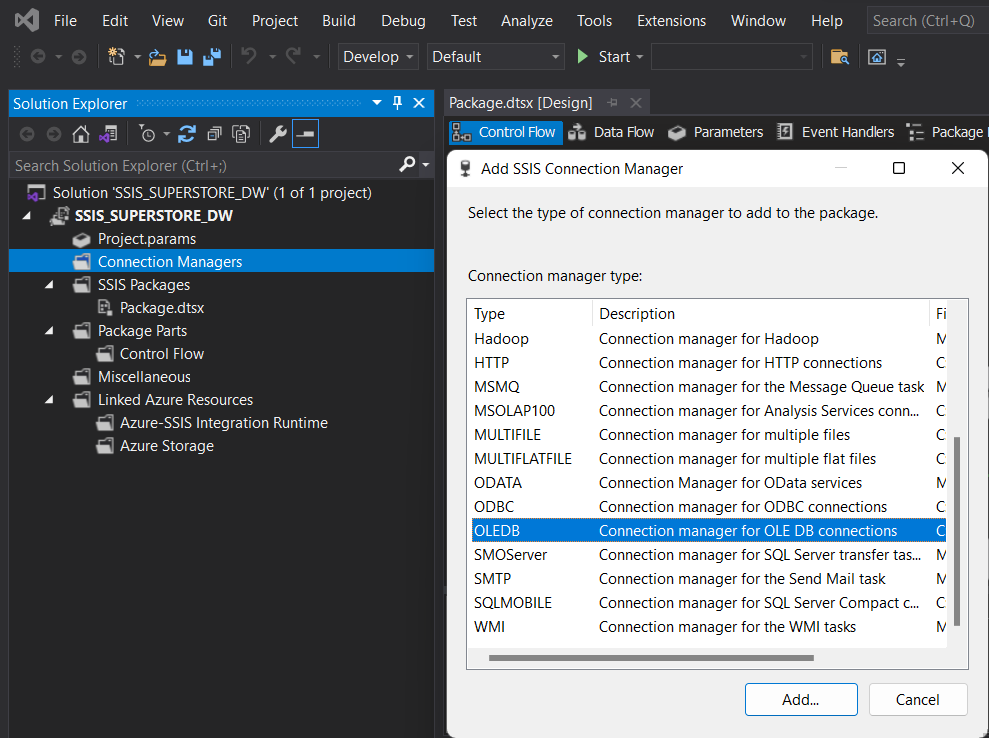
Vào Mircosoft Visual Studio tạo một project SSIS đặt tên là SSIS\_SUPERSTORE\_DW.



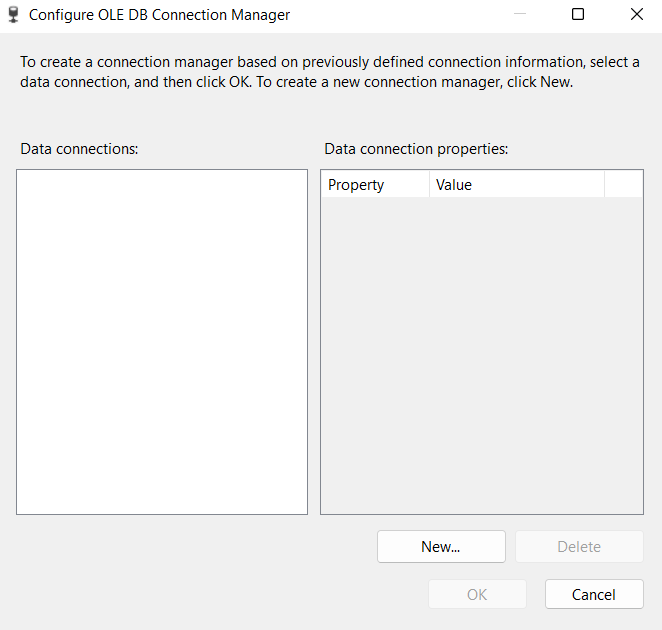
Click chuột phải vào Connection Managers chọn New Connection Managers.



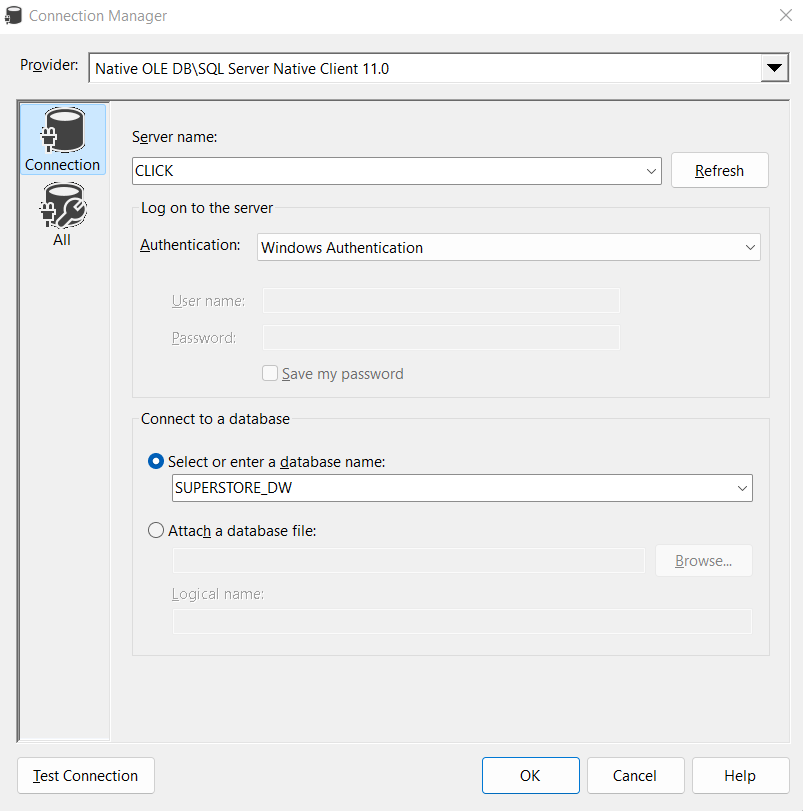
Chọn OLEDB và nhấn Add.



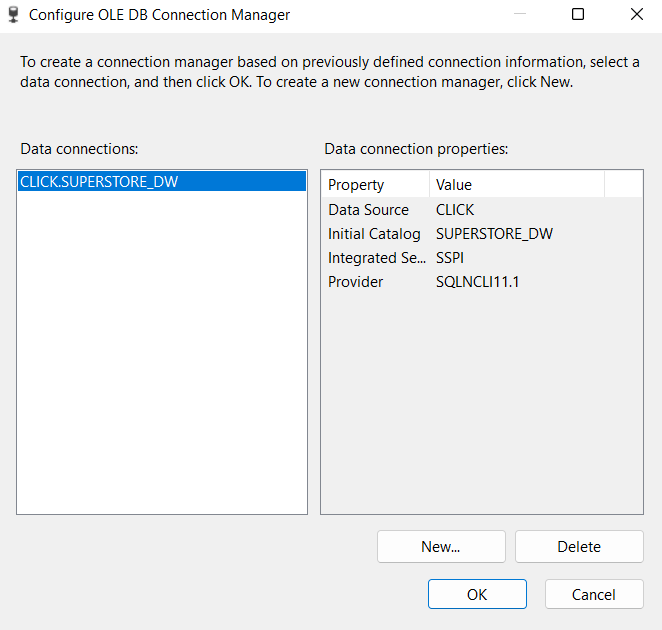
Tiếp tục nhấn New.



Nhập tên Sever name và connect tới Database đã tạo.

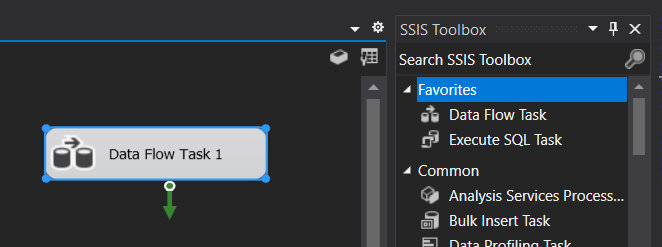


Nhấn Ok.



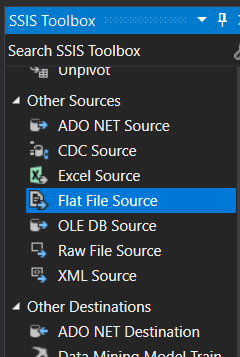
## Tạo Data Flow Task

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

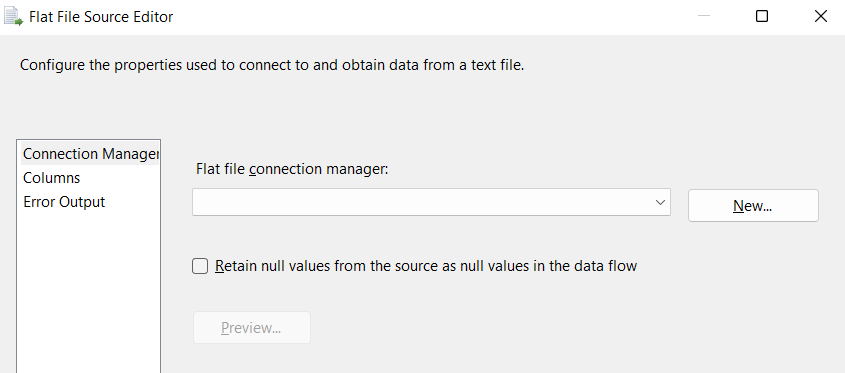


## Tạo Flag File Source kết nối với file dữ liệu

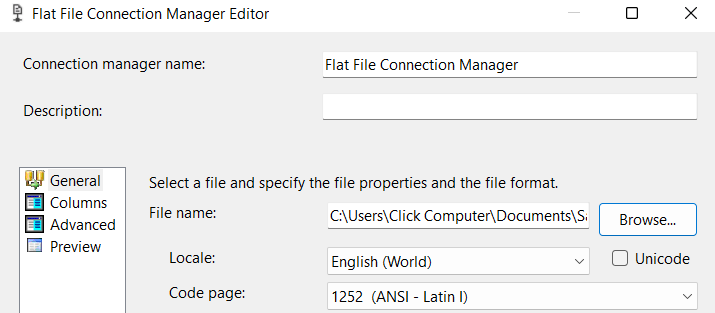
Chọn Flag File Source trên thanh công cụ và kết nối tới file dữ liệu.



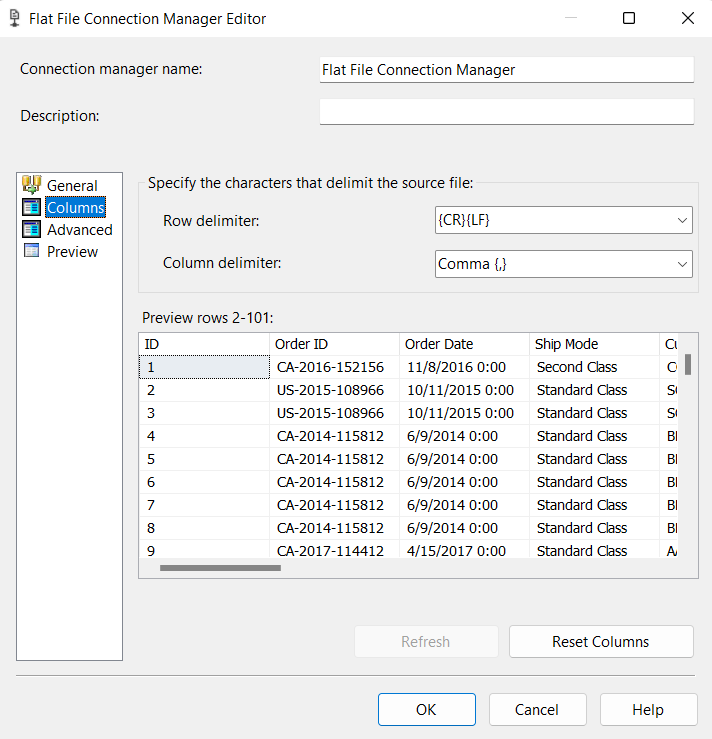
Chọn New.



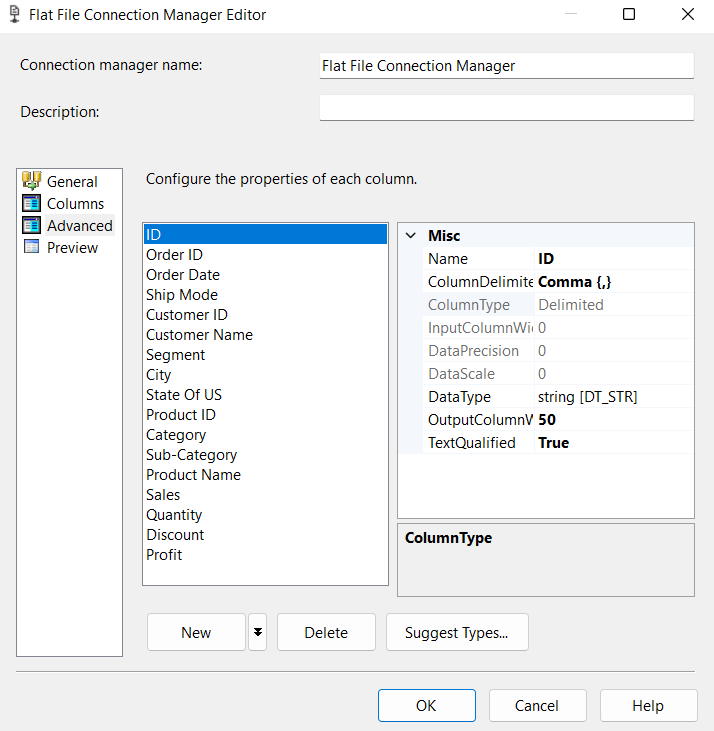
Chọn Browse và đường dẫn tới file dữ liệu



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

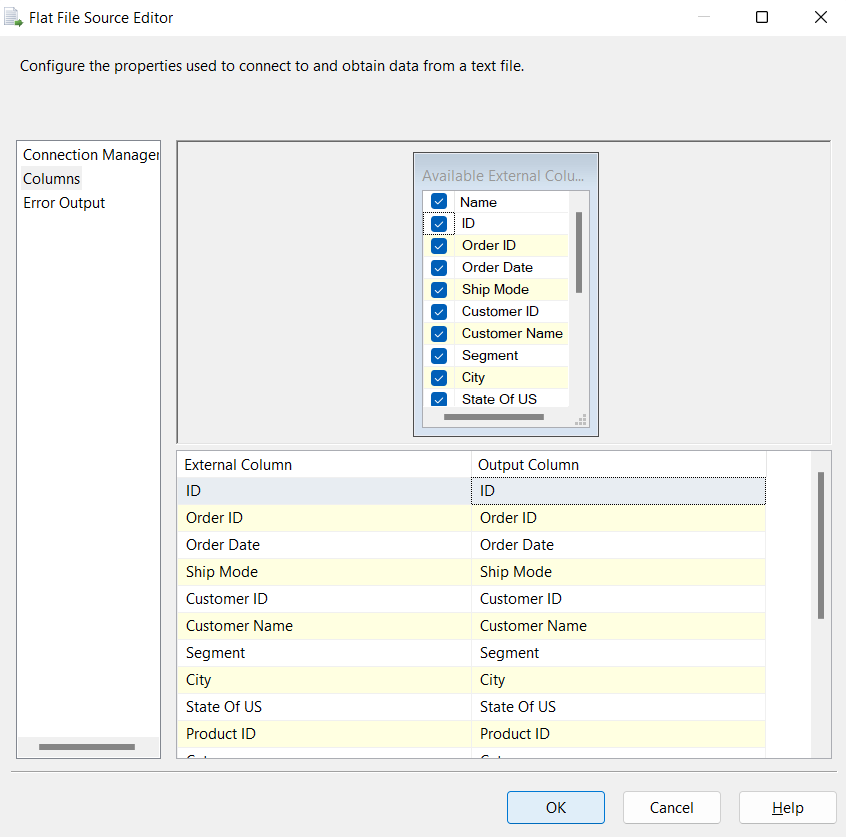


Chọn Advanced để chỉnh sửa dữ liệu (đổi tên các cột, xóa các cột không cần thiết)



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.

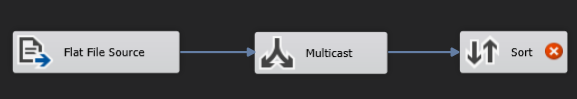


## Rẽ nhánh và lọc data trùng

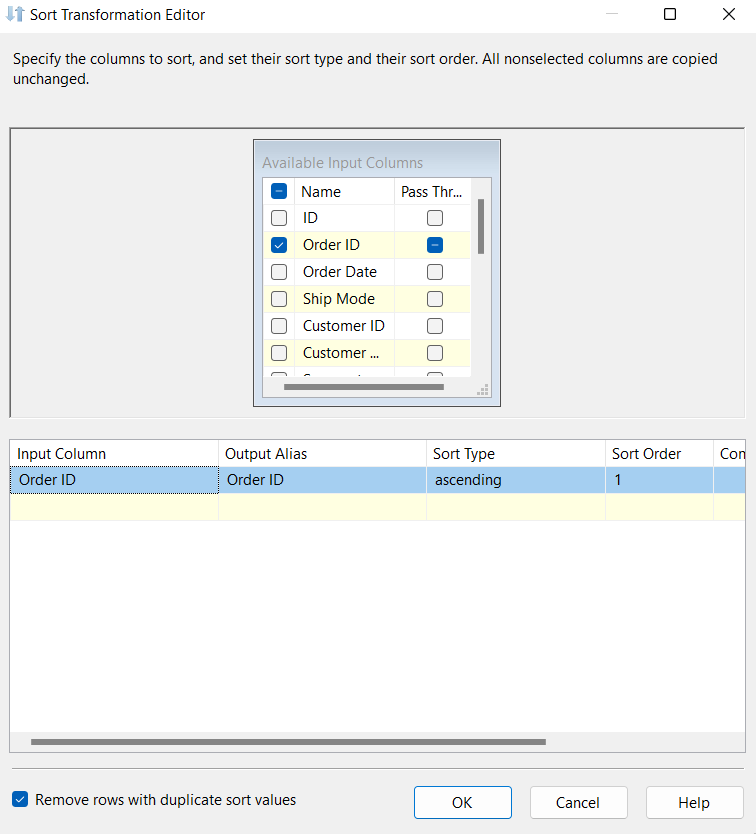
Chọn công cụ Multicast để có thể đẩy dữ liệu vào nhiều bảng



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.

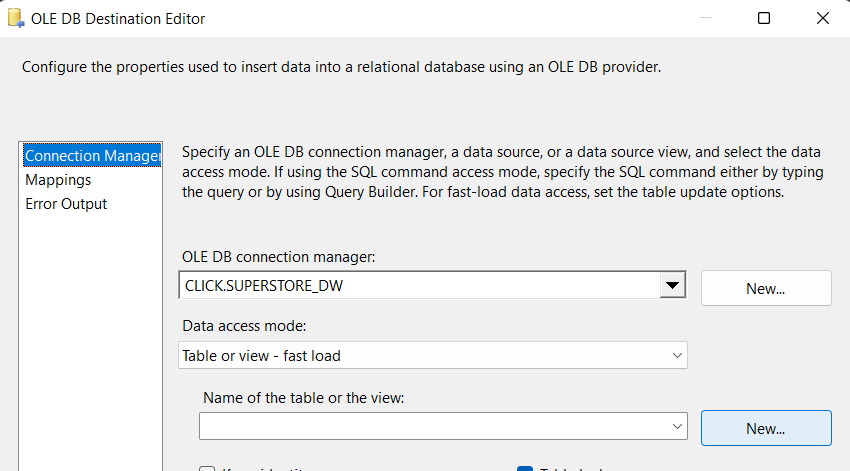


## Tạo các bảng Dimension

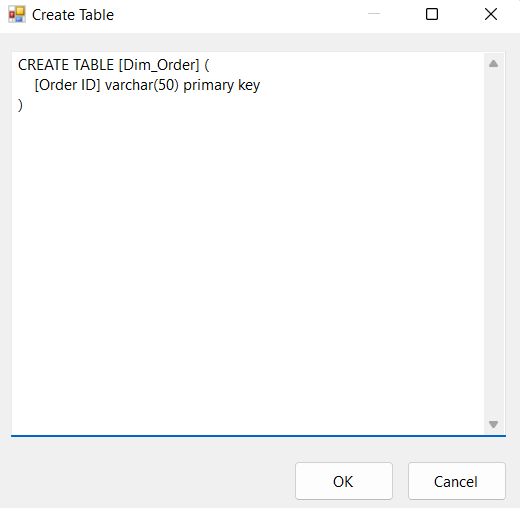
Chọn OLE DB Destination trên thanh công cụ. Click chuột phải, chọn Rename để đổi tên.



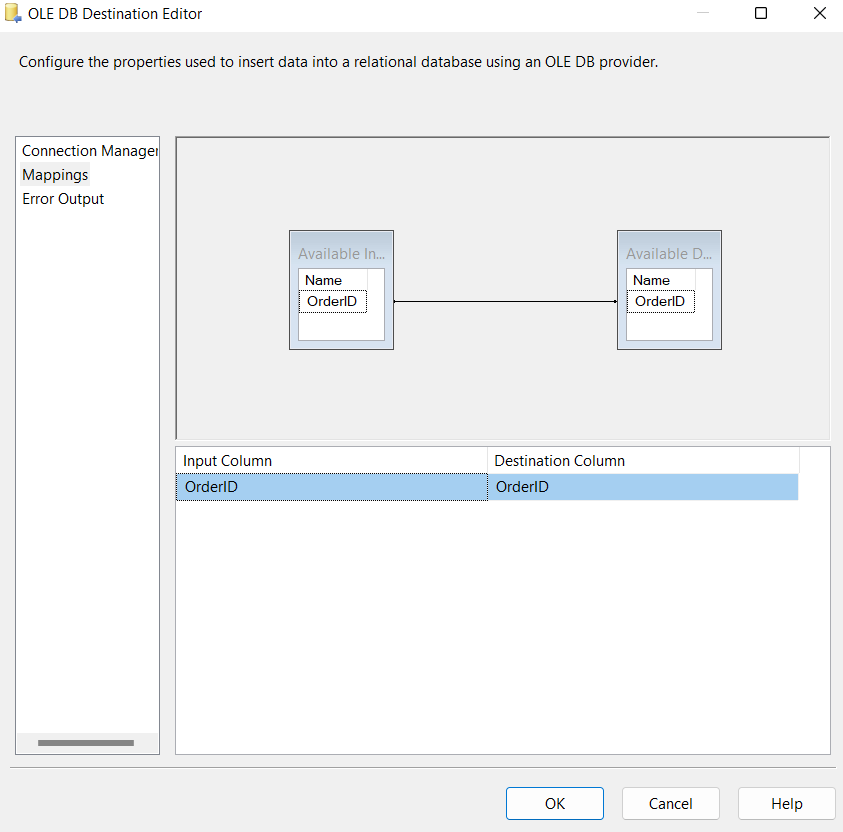
Click chuột vào Dim\_Order và chọn New để thực hiện việc tạo bảng.



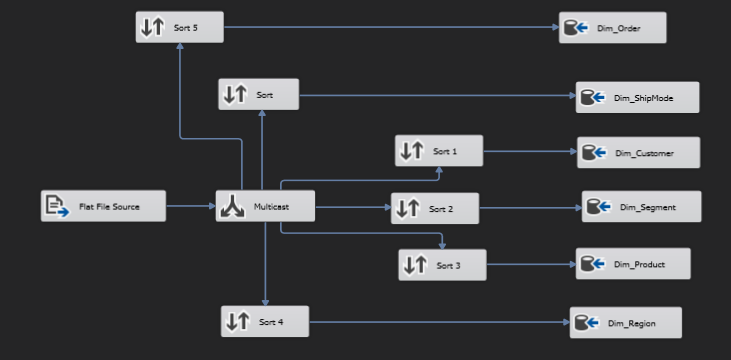
Tiến hành chỉnh sửa thuộc tính của bảng, thêm các khóa cần thiết. Nhấn Ok để hoàn tất.



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

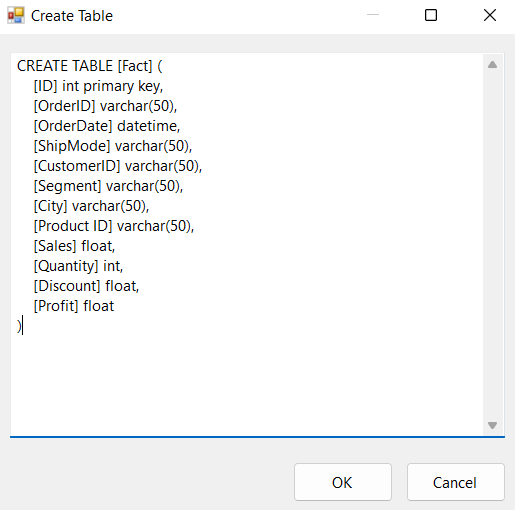


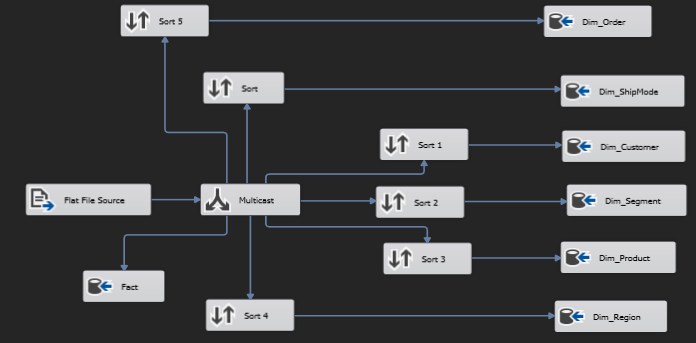
Thực hiện tương tự cho các bảng Dim khác



## Tạo bảng Fact

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



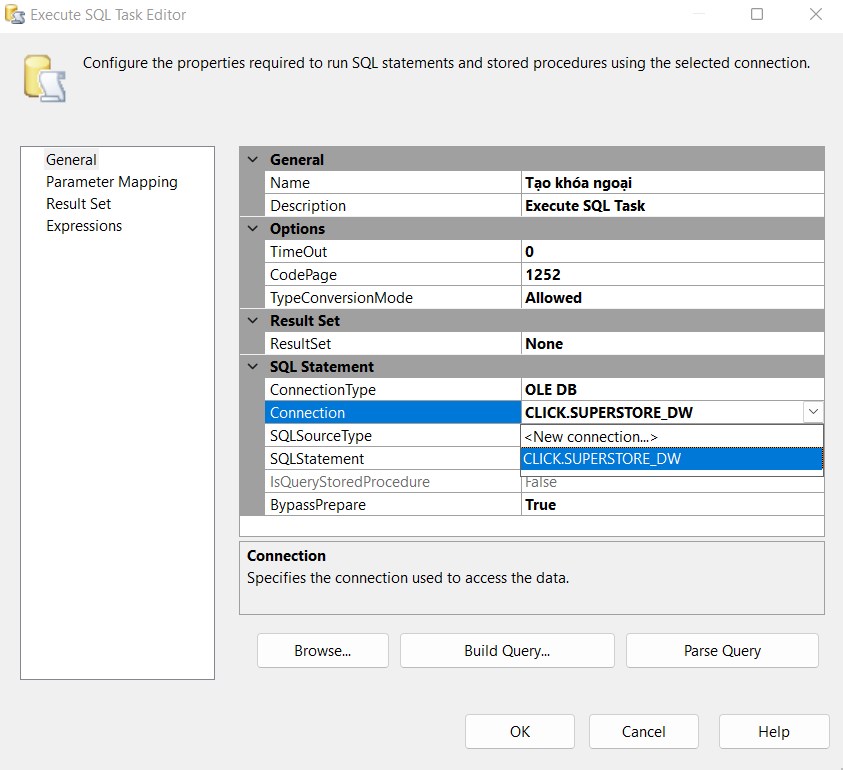


## Tạo Execute SQL Task khóa ngoại cho bảng Fact và các bảng Dim

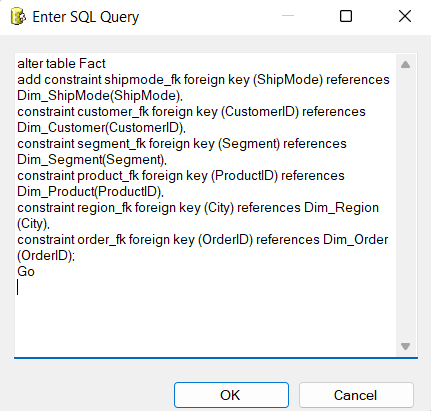
Quay lại giao diện Control Flow, tạo mới Execute SQL Task và đặt tên là Tạo khóa ngoại cho bảng Fact.



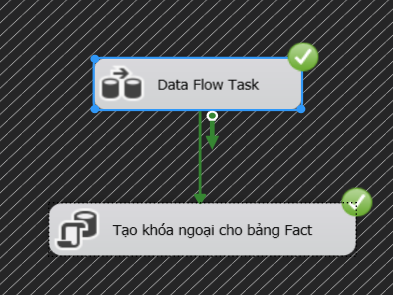
Tiếp tục Clicl chuột vào, ở dòng Connection chọn kết nối tới Database.

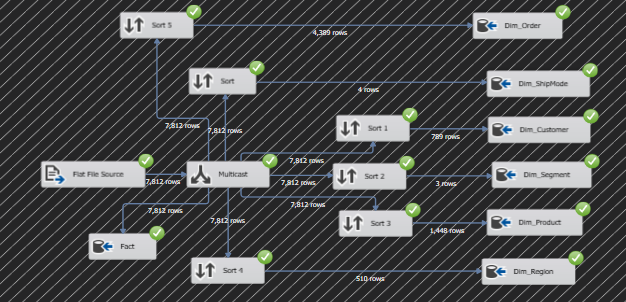


Click chuột vào dấu ... ở dòng SQLStatement và thực hiện nhập các câu lệnh SQL để tạo khóa ngoại. Nhấn 2 lần OK để kết thúc.



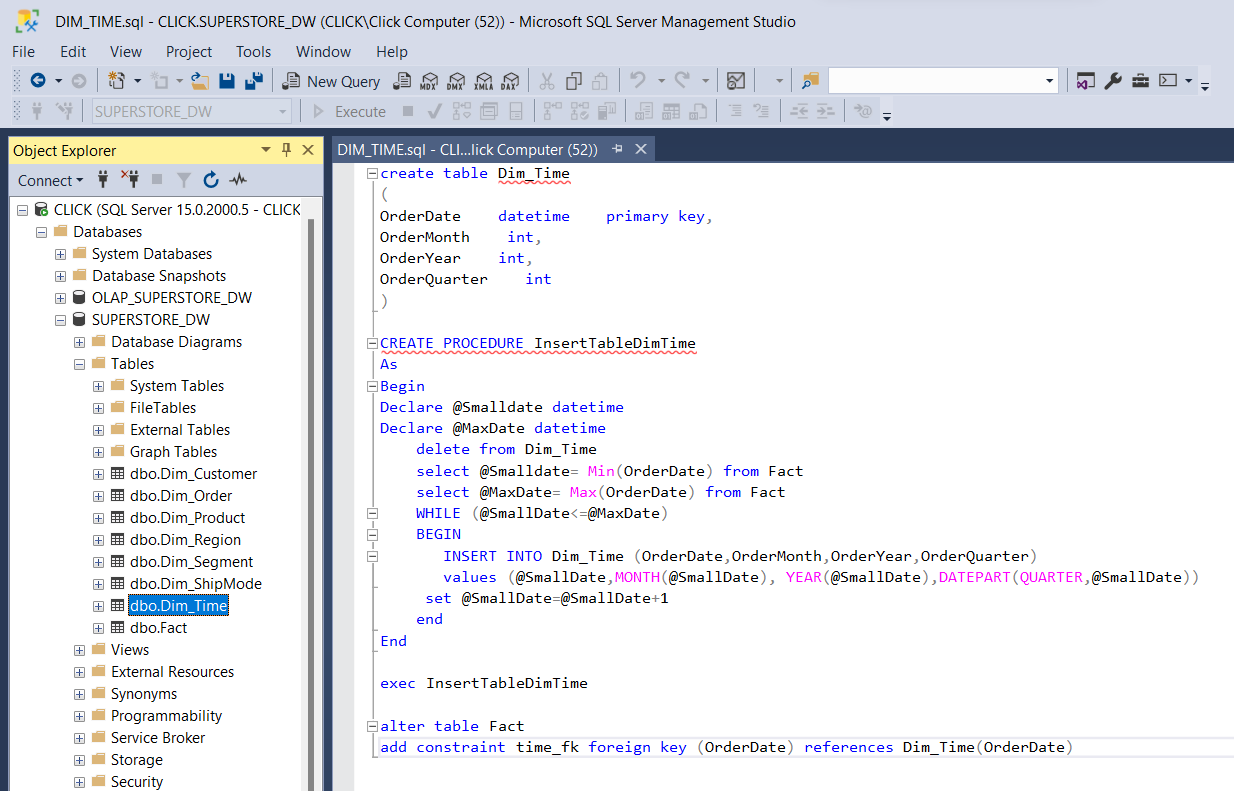
Nhấn Start để thực hiện chạy Project lần đầu tiên.





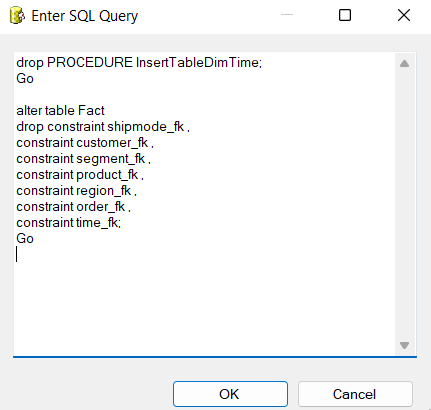
## Tạo bảng Dim\_Time và các ràng buộc

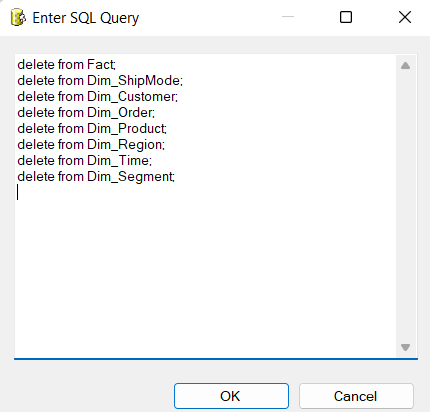
Vào SQL Server chọn New Query và thực thi lần lượt các lệnh: tạo bảng Dim\_Time, tạo Procedure, tạo khóa ngoại OrderDate cho bảng Fact. (\*)



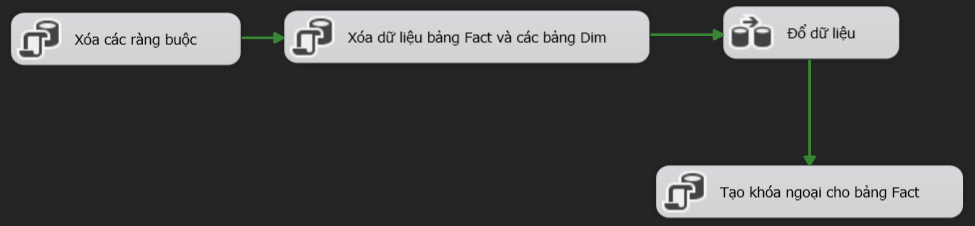
## Tạo Execute SQL Task để xóa ràng buộc, xóa bảng Fact và các bảng Dim

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



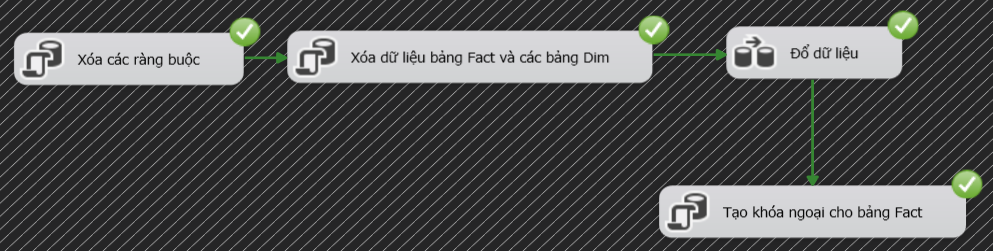


Nối các khối lệnh lại với nhau



## Thực thi project

Chạy Project

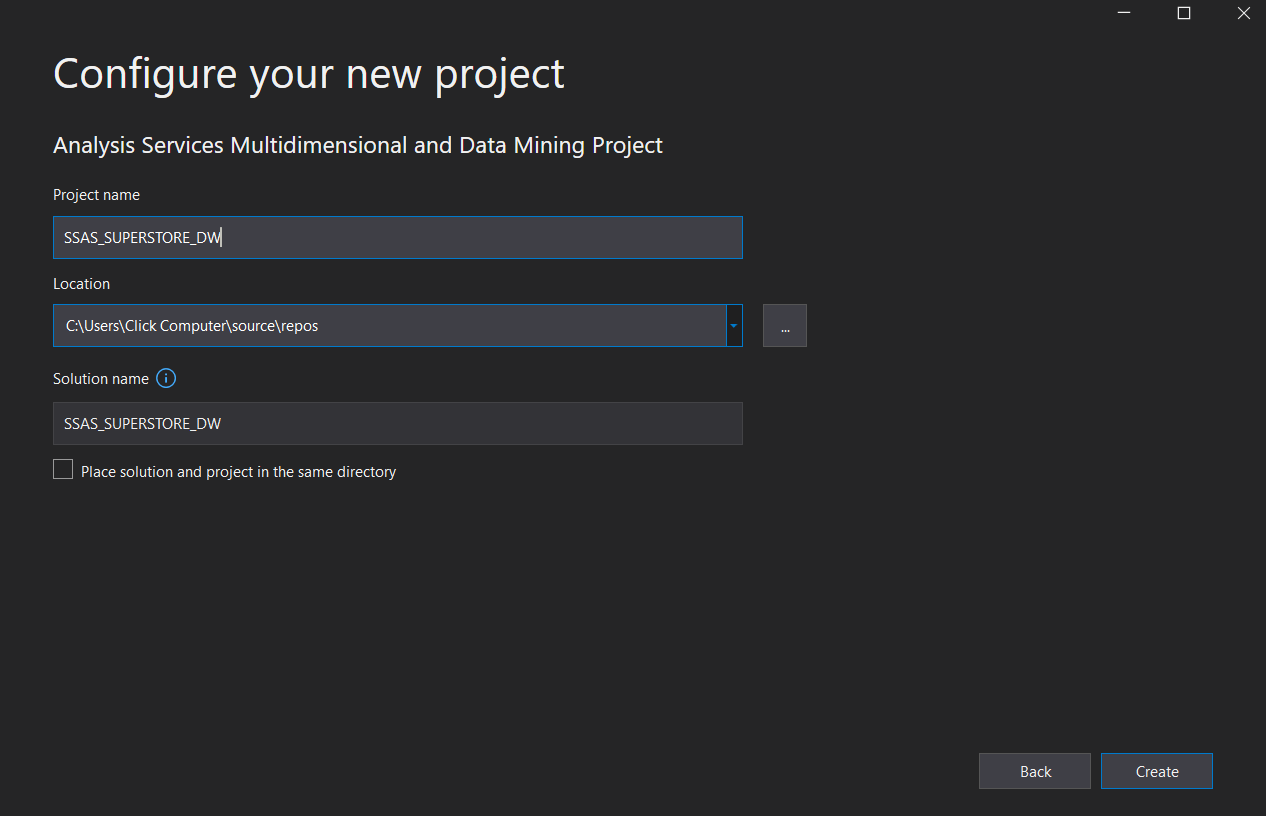


**Lưu ý:** kể từ lần chạy thứ 2, ta phải vào SQL Server để thực hiện lại bước (\*) (trừ lệnh tạo bảng Dim\_Time).

# 3. PHÂN TÍCH KHO DỮ LIỆU (SSAS)

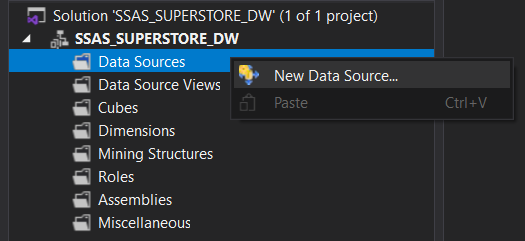
## Tạo Project

Tạo một project SSAS đặt tên là SSAS\_SUPERSTORE\_DW.

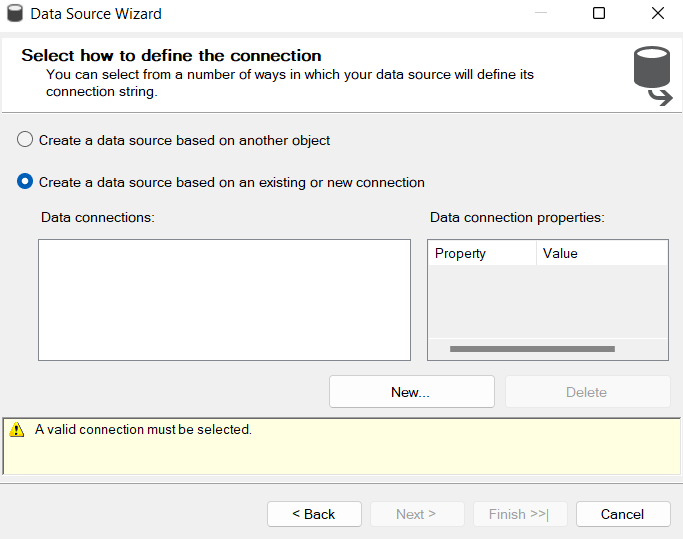


## Tạo Data Source

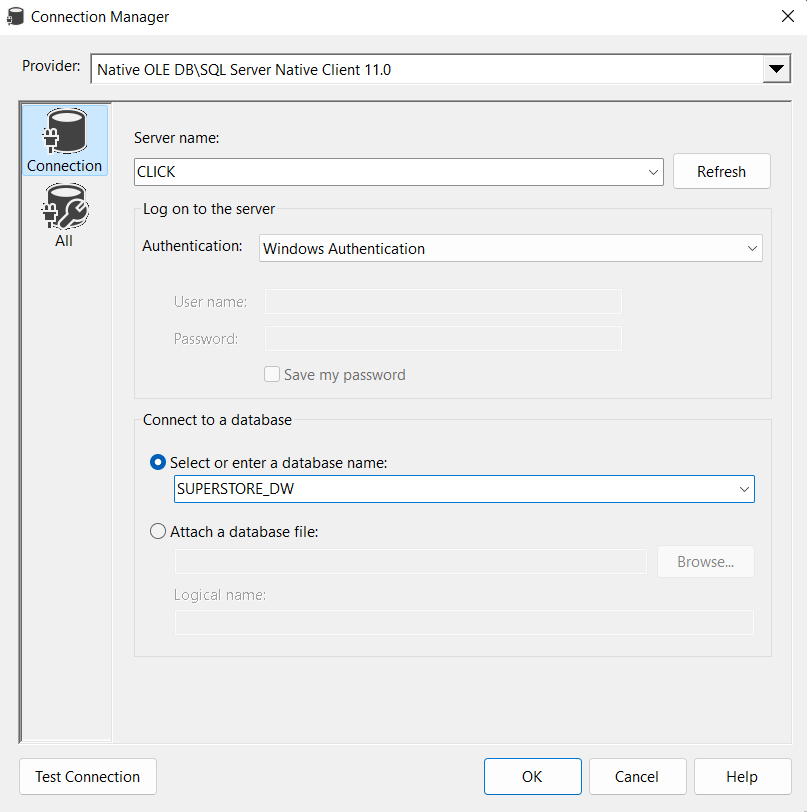
Click chuột phải và Data Sources chọn New Data Source



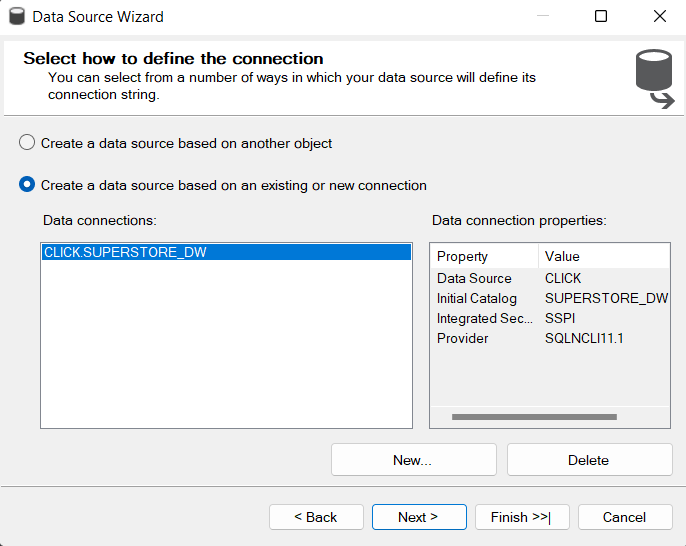
Chọn Create a data source based on an existing or new connection, sau đó chọn New.



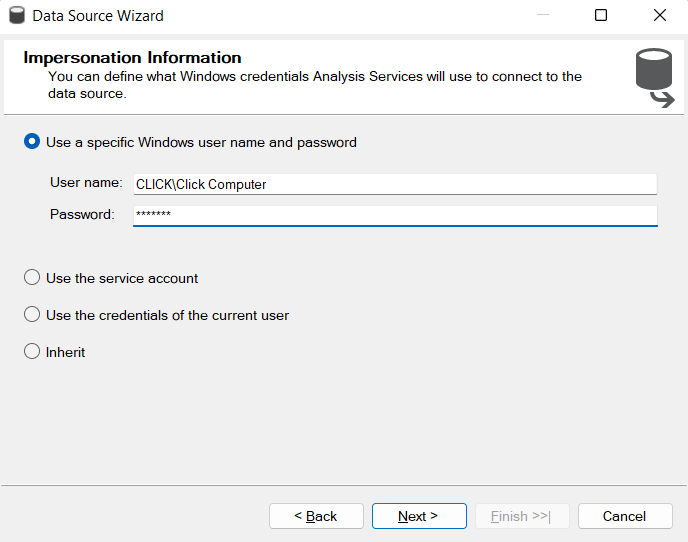
Nhập Server Name và chọn Database SUPERSTORE\_DW.



Nhấn OK và nhấn Next.

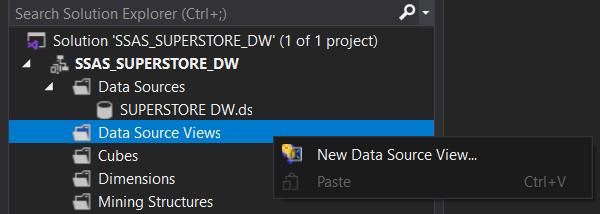


Nhập tên tài khoản trong phần Security của SQL Server và nhập mật khẩu đăng nhập Window. Nhấn Next. Nhấn Finish để kết thúc.

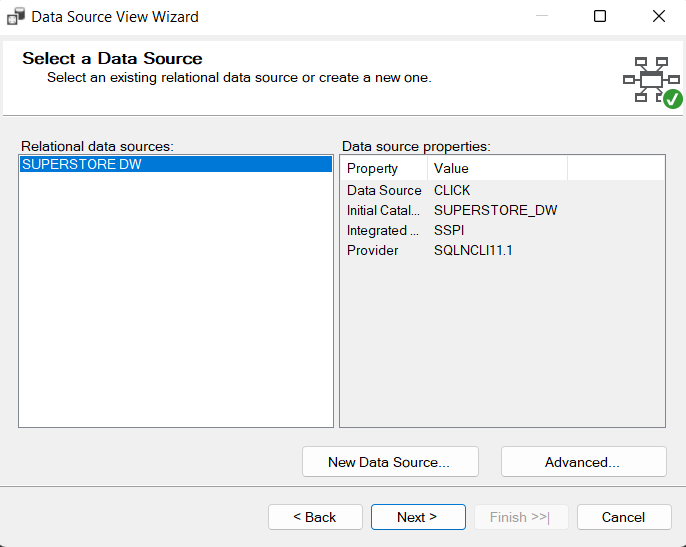


## Tạo Data Source View

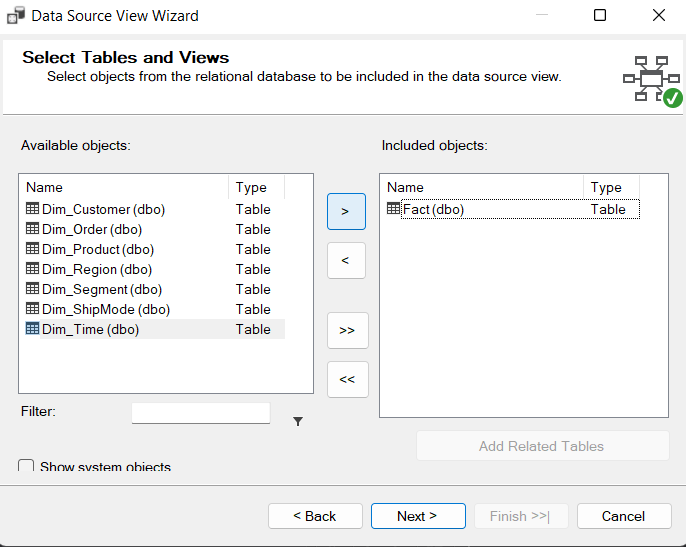
Click chuột phải và Data Source View chọn New Data Source View.



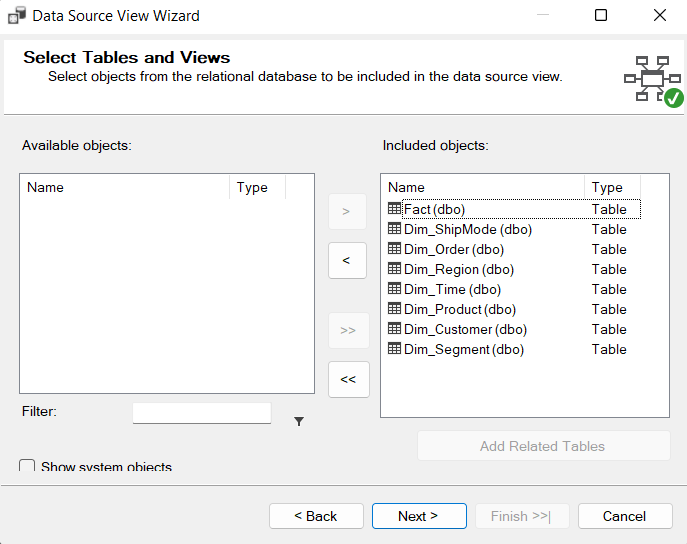
Chọn Data Source đã tạo và nhấn Next.



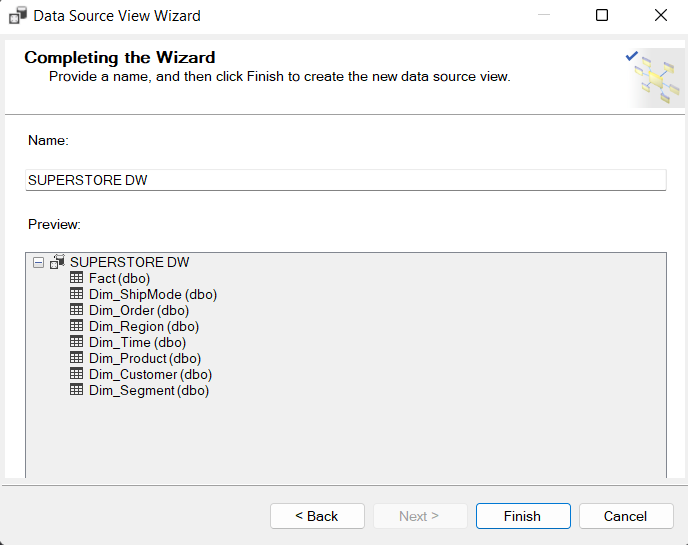
Chọn bàng Fact và bấm mũi tên để đẩy bàng Fact qua.



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.

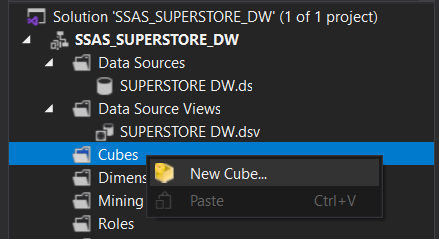


Nhấn Finish để kết thúc.

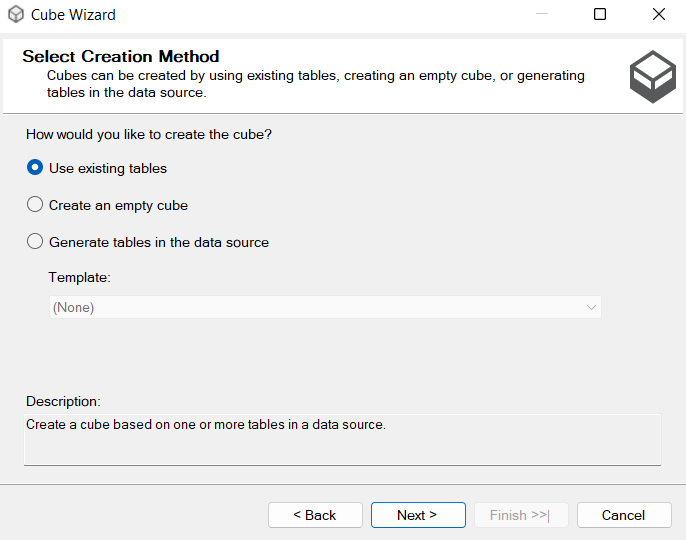


## Tạo Cube

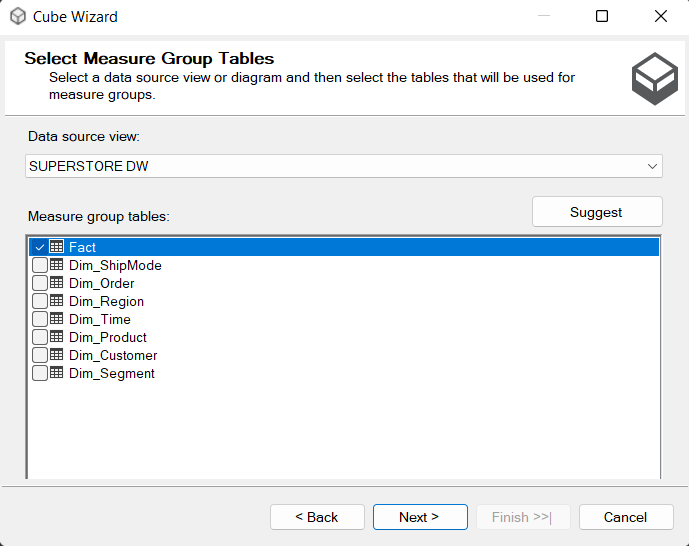
Chuột phải vào Cubes chọn New Cube.



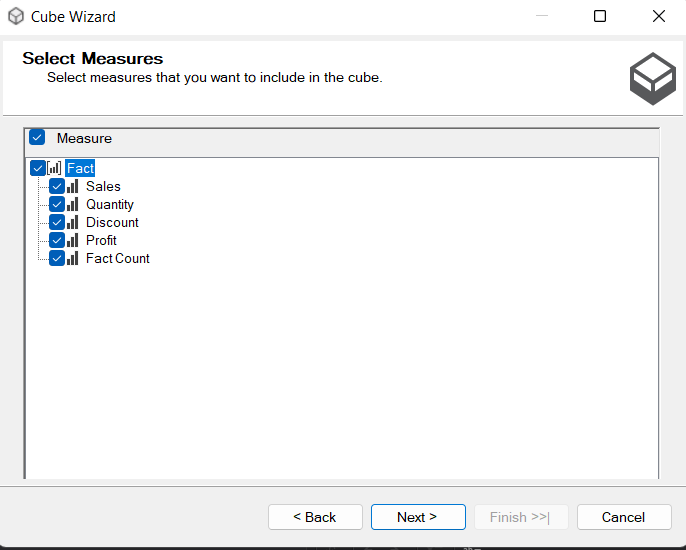
Nhấn Next.



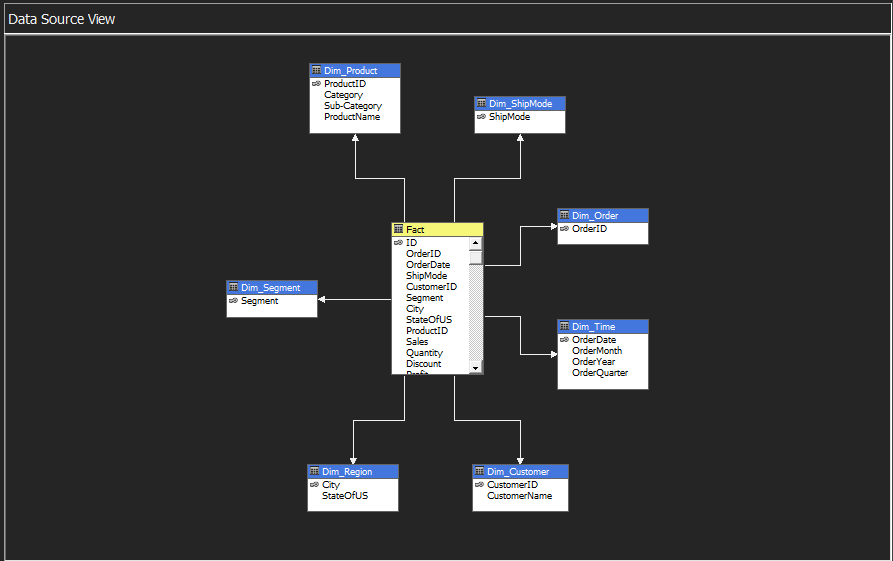
Chọn bảng Fact và nhấn Next.



Tiếp tục nhấn Next và chọn Finish để kết thúc.

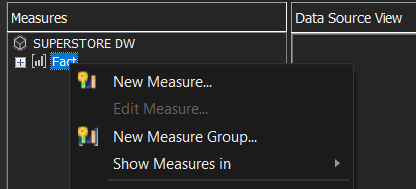


Lược đồ hình sao khi được tạo hoàn tất.

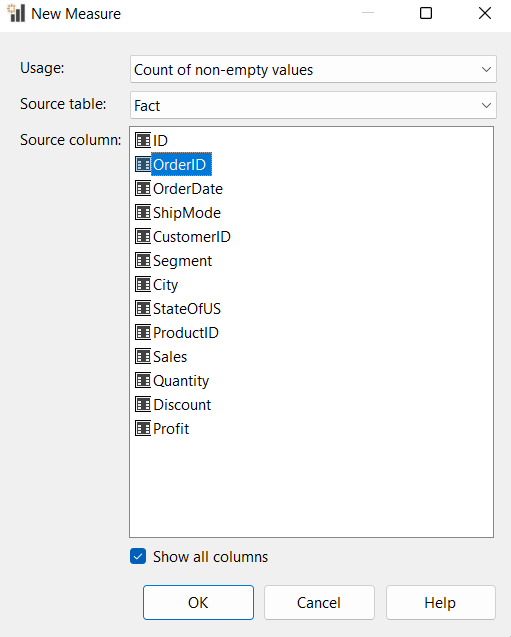


## Tạo thêm Measure

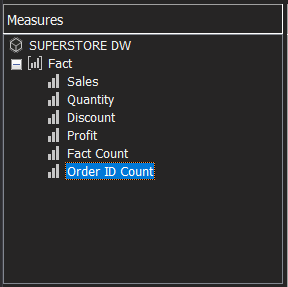
Ở cột Measure click chuột phải vào bảng Fact và chọn New Measure.



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.

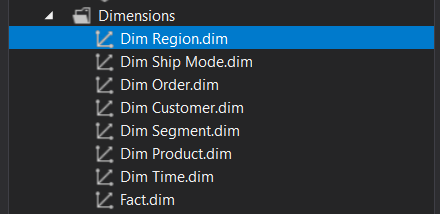


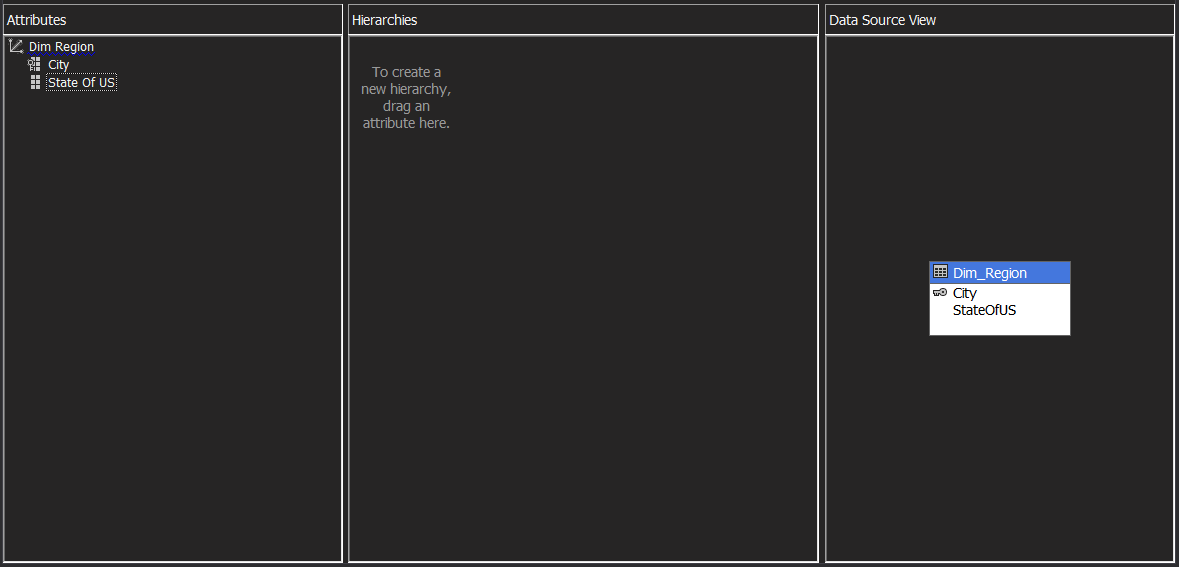
Kết quả



## Chỉnh sửa thuộc tính các bảng Dim

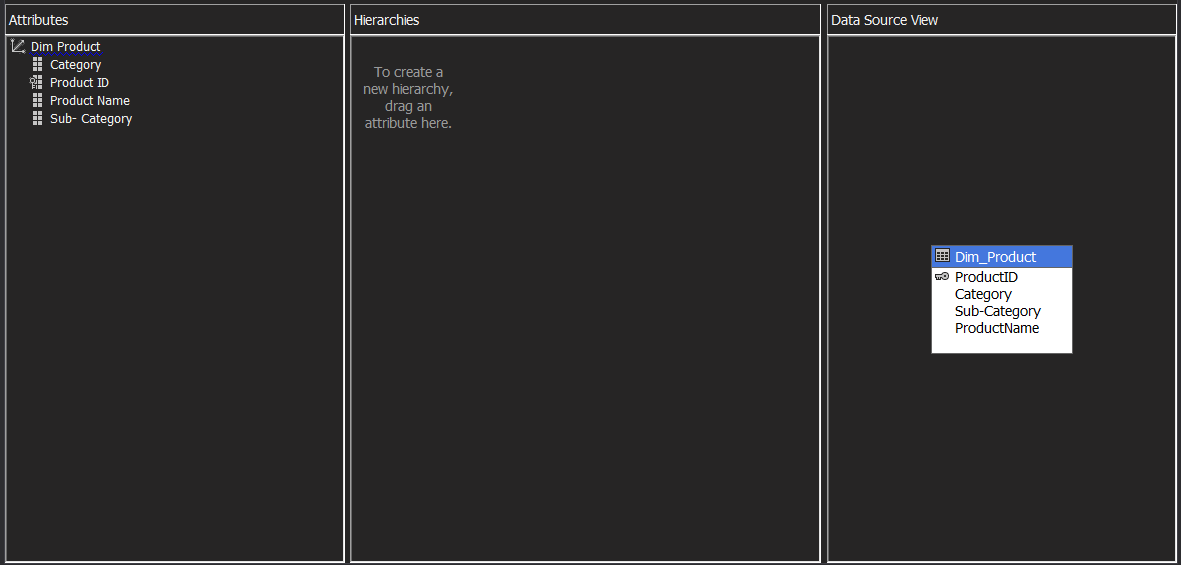
Ở phần Dimensions lần lượt chọn các bảng Dim, sau đấy thực hiện kéo thả các thuộc tính chưa có ở phần Data Source View sang Attributes.





Thực hiên tương tự cho các bảng Dim còn lại.

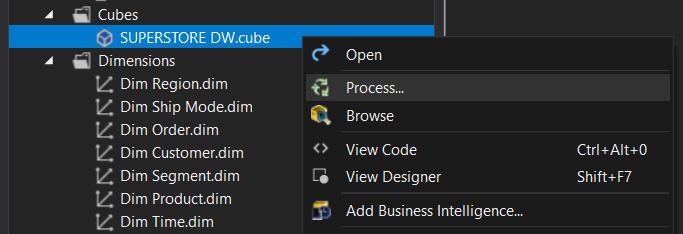




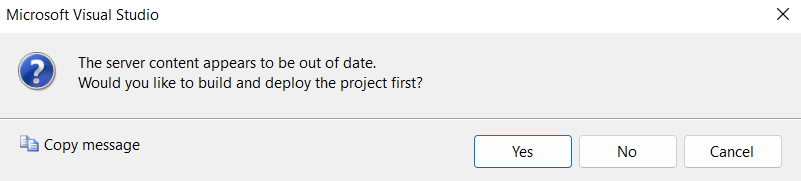


## Thực thi project

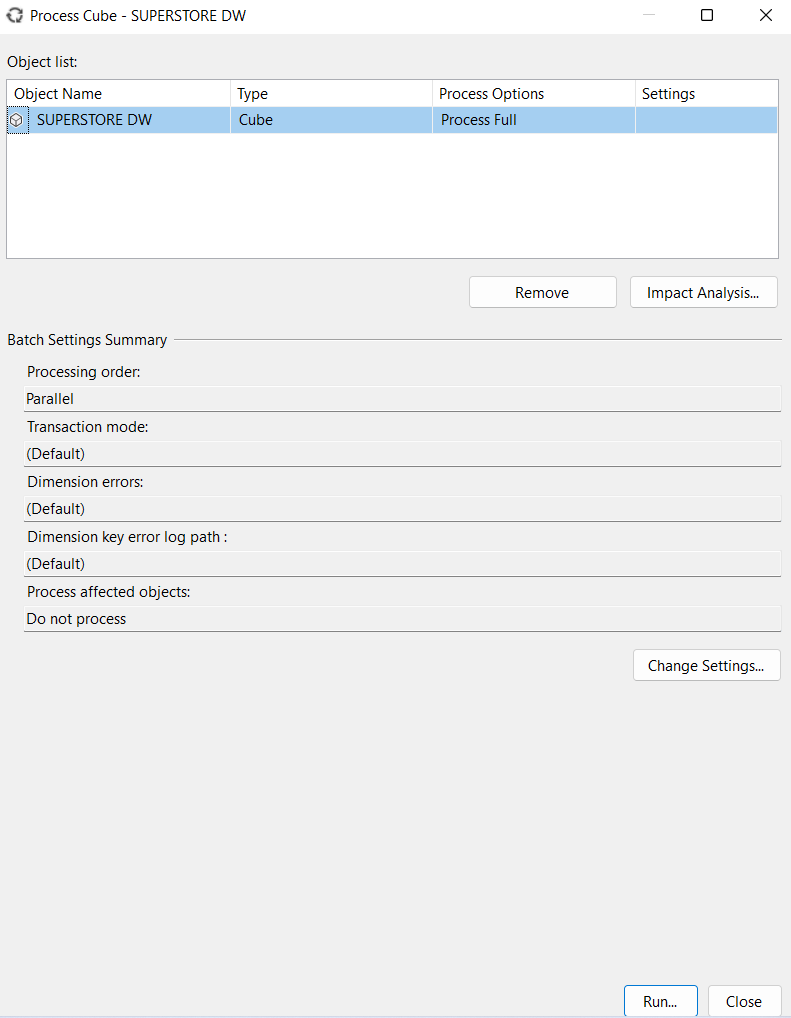
Click chuột phải vào Cube đã tạo và chọn Process.



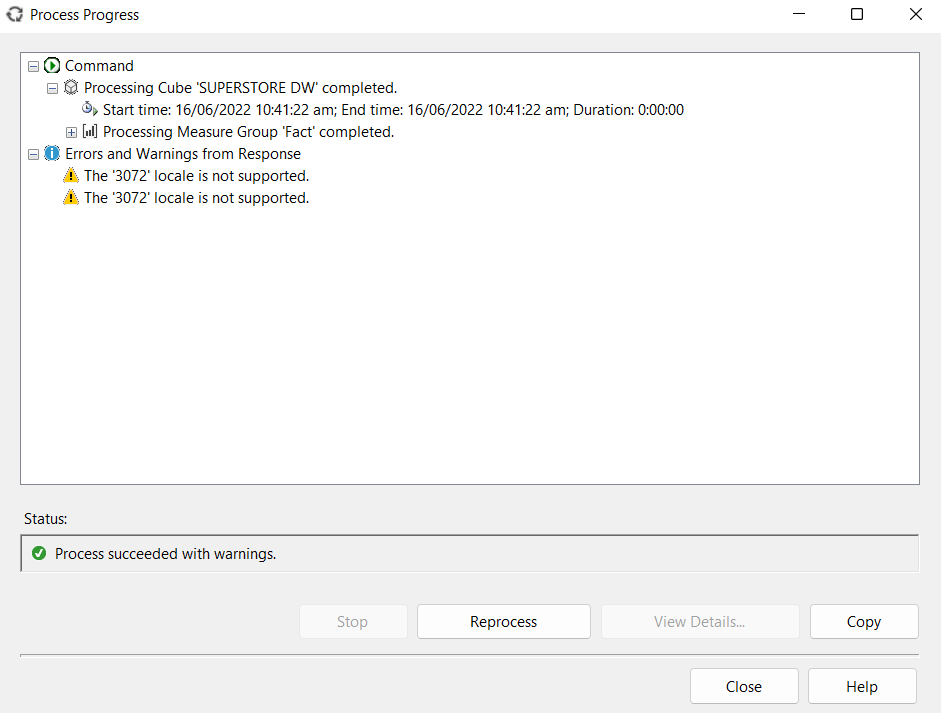
Chọn Yes để Deploy Project.



Chọn Run để chạy.

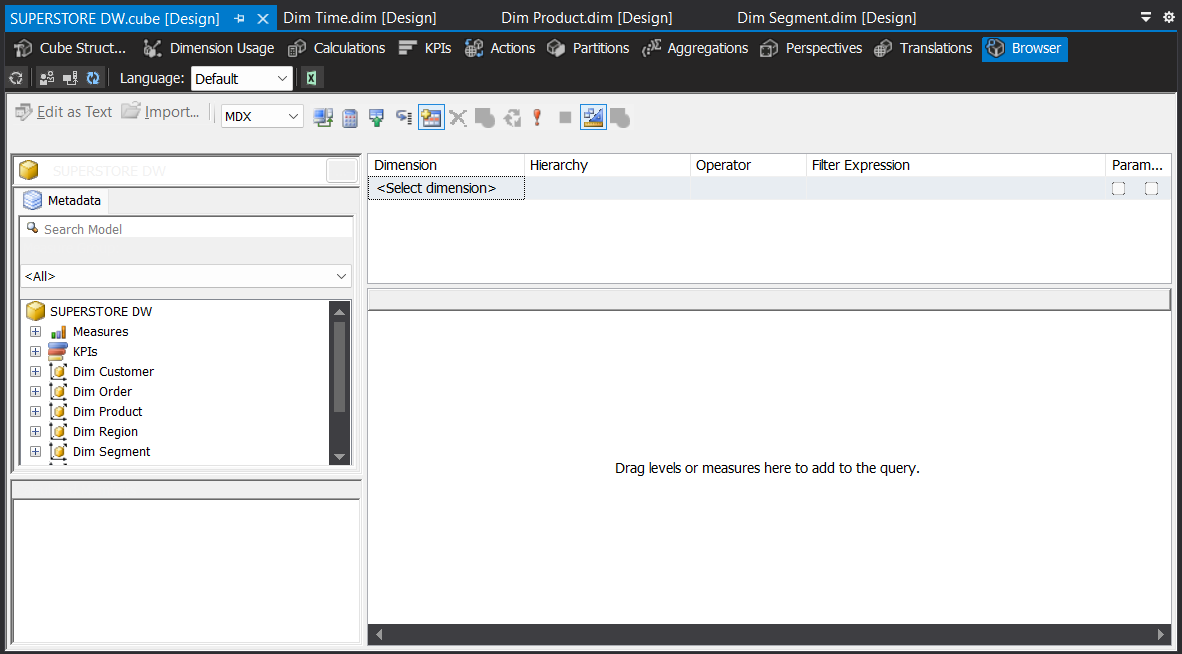


Hoàn tất. Chọn Close để đóng.

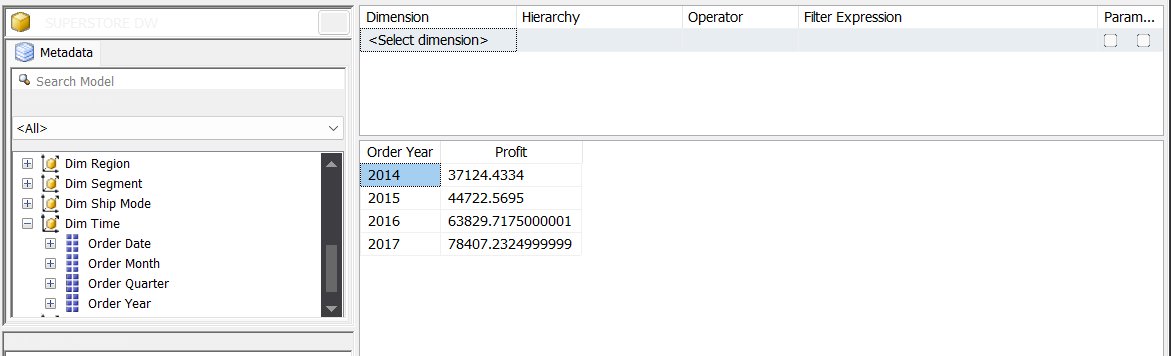


## Truy vấn dữ liệu

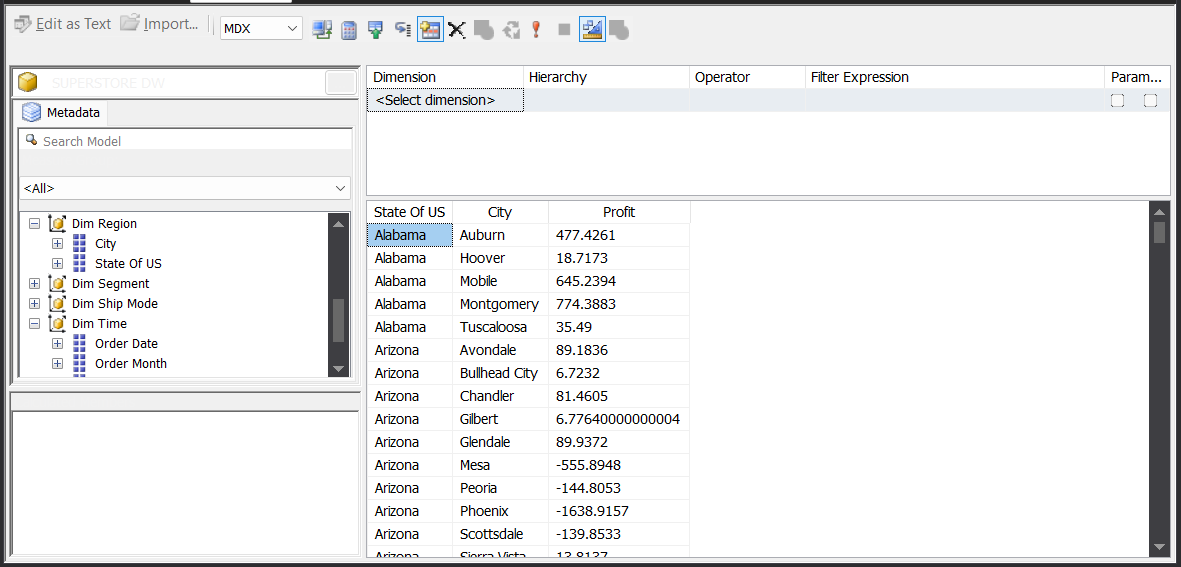
Chọn Browser trên thanh công cụ để thực hiện việc truy vấn dữ liệu.



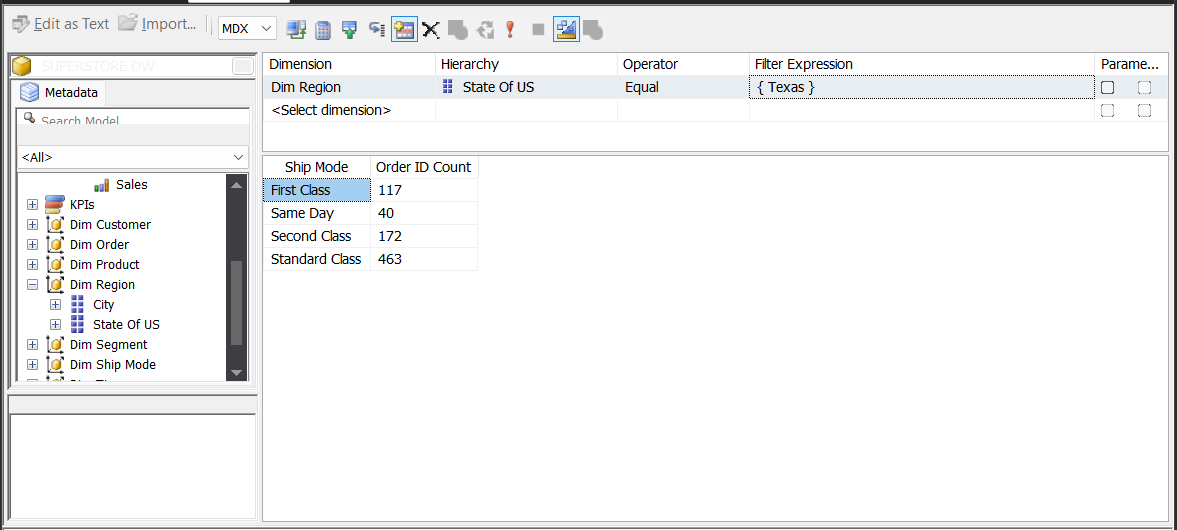
### Truy vấn loại Roll Up: tính tổng lợi nhuận qua các năm.



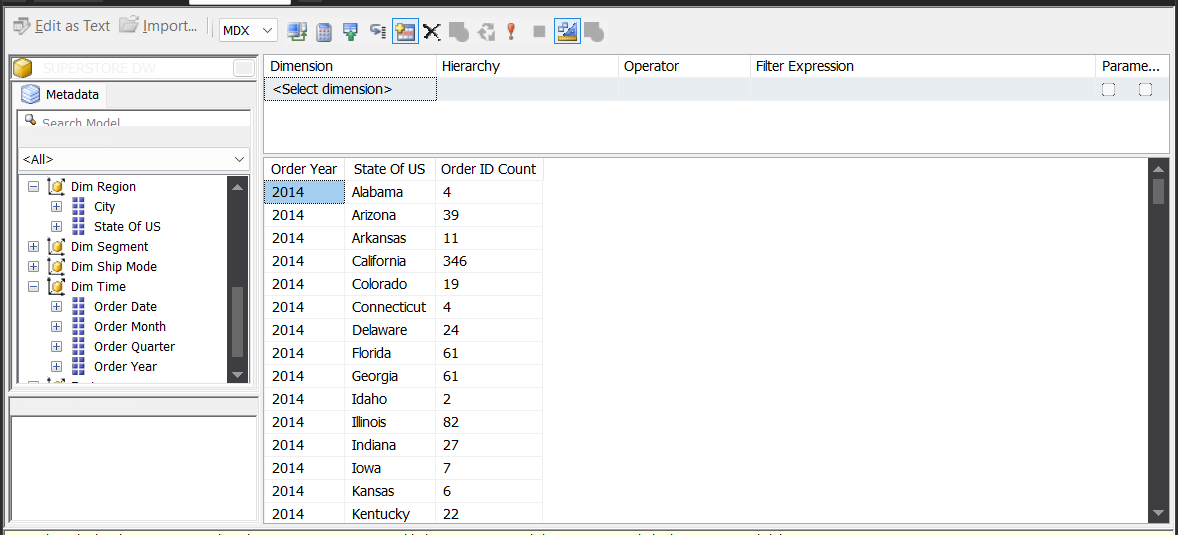
### Truy vấn loại Drill Down: tính tổng lợi nhuận của từng thành phố, từng bang.



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



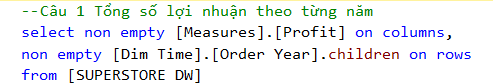
### Truy vấn loại Pivot: số đơn hàng của từng bang qua các năm.



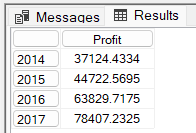
## Truy vấn MDX

### Tổng số lợi nhuận theo từng năm

Lệnh truy vấn

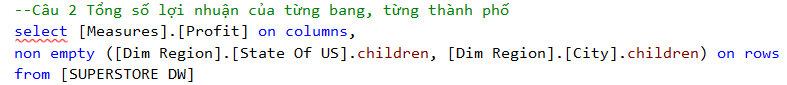


Kết quả

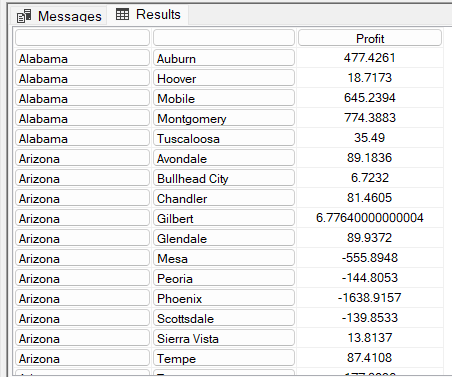


### Tổng số lợi nhuận của từng bang, từng thành phố

Lệnh truy vấn

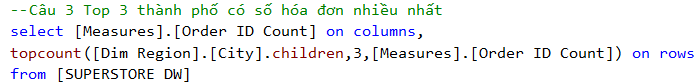


Kết quả

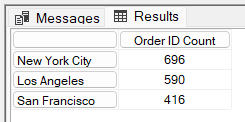


### Top 3 thành phố có số hóa đơn nhiều nhất

Lệnh truy vấn

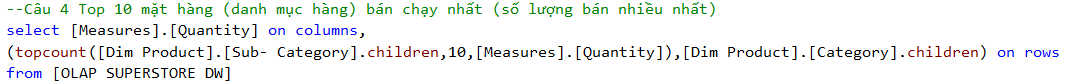


Kết quả

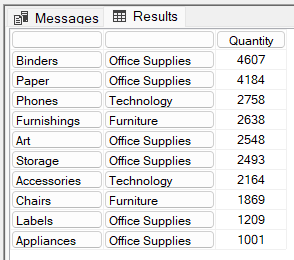


### Top 10 mặt hàng (danh mục hàng) bán chạy nhất (số lượng bán nhiều nhất)

Lệnh truy vấn

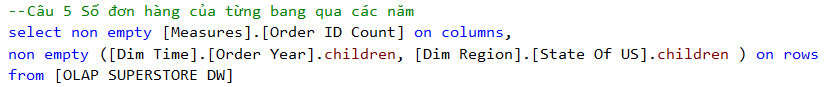


Kết quả

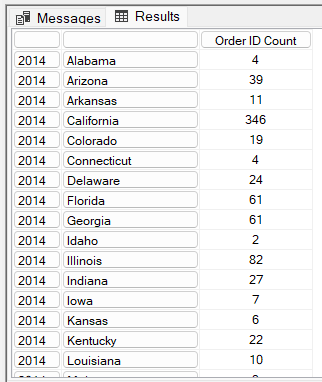


### Số đơn hàng của từng bang qua các năm

Lệnh truy vấn

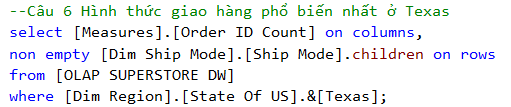


Kết quả

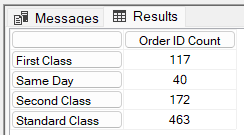


### Hình thức giao hàng phổ biến nhất ở Texas

Lệnh truy vấn

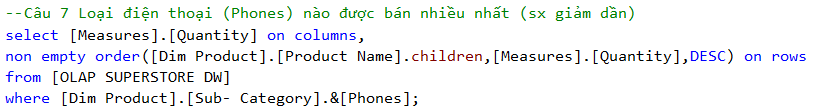


Kết quả

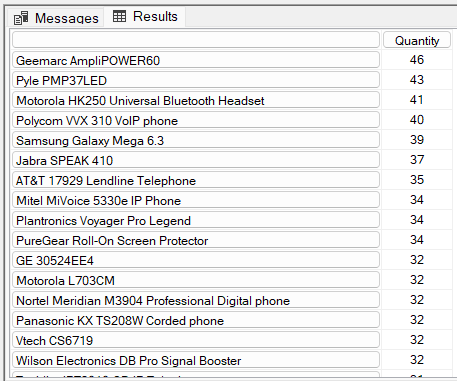


### Loại điện thoại (Phones) nào được bán nhiều nhất (sắp xếp giảm dần)

Lệnh truy vấn

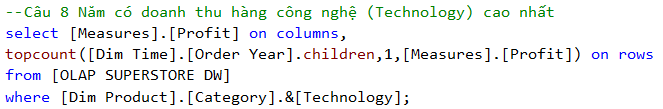


Kết quả

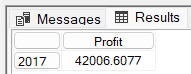


### Năm có doanh thu hàng công nghệ (Technology) cao nhất

Lệnh truy vấn

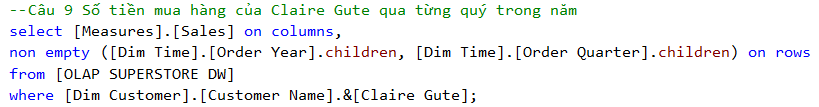


Kết quả

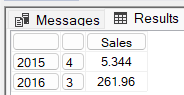


### Số tiền mua hàng của Claire Gute qua từng quý trong năm

Lệnh truy vấn

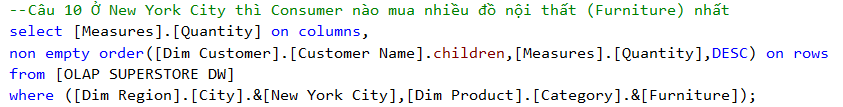


Kết quả

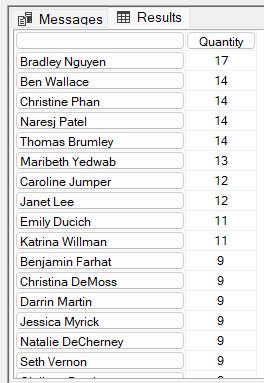


### Ở New York City thì Consumer nào mua nhiều đồ nội thất (Furniture) nhất

Lệnh truy vấn

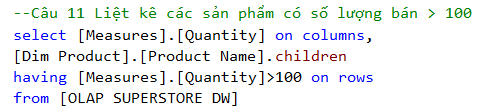


Kết quả

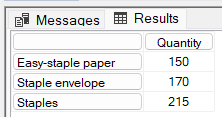


### Liệt kê các sản phẩm có số lượng bán > 100

Lệnh truy vấn

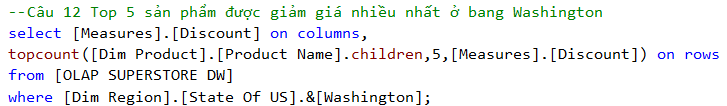


Kết quả

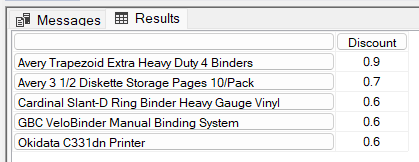


### Top 5 sản phẩm được giảm giá nhiều nhất ở bang Washington

Lệnh truy vấn

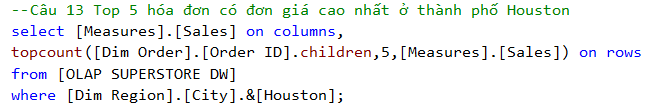


Kết quả

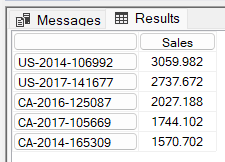


### Top 5 hóa đơn có đơn giá cao nhất ở thành phố Houston

Lệnh truy vấn

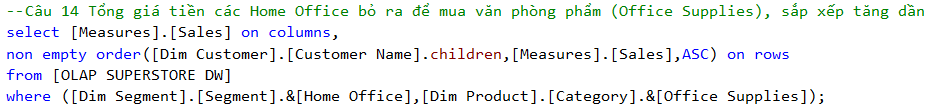


Kết quả

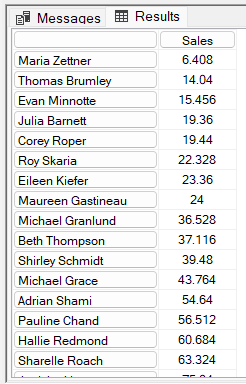


### Tổng giá tiền các Home Office bỏ ra để mua văn phòng phẩm (Office Supplies), sắp xếp tăng dần

Lệnh truy vấn

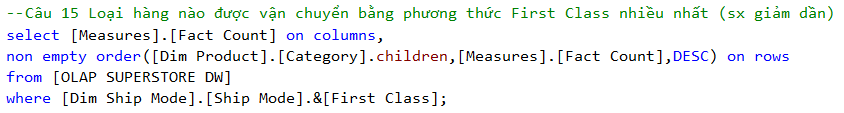


Kết quả



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

Lệnh truy vấn



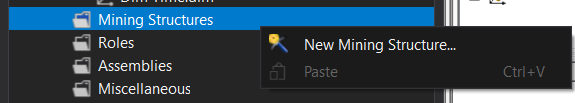
Kết quả



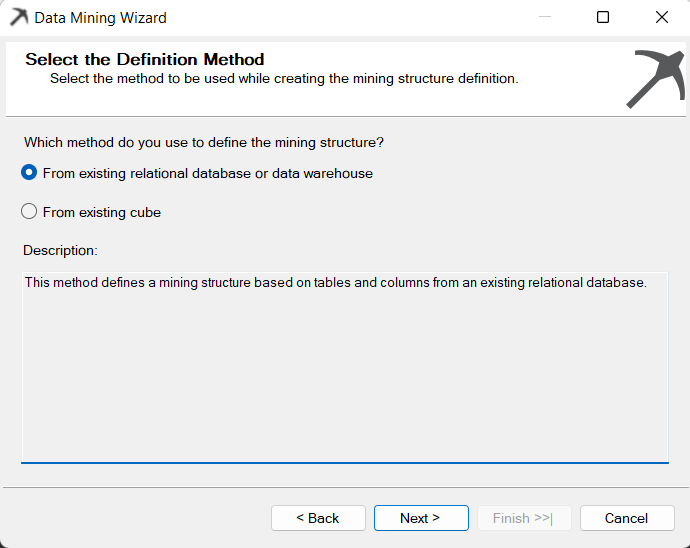
# 4. KHAI PHÁ DỮ LIỆU (DATA MINING)

## Tạo Project

Chuột phải vào Mining Structures và chọn New Mining Structure.

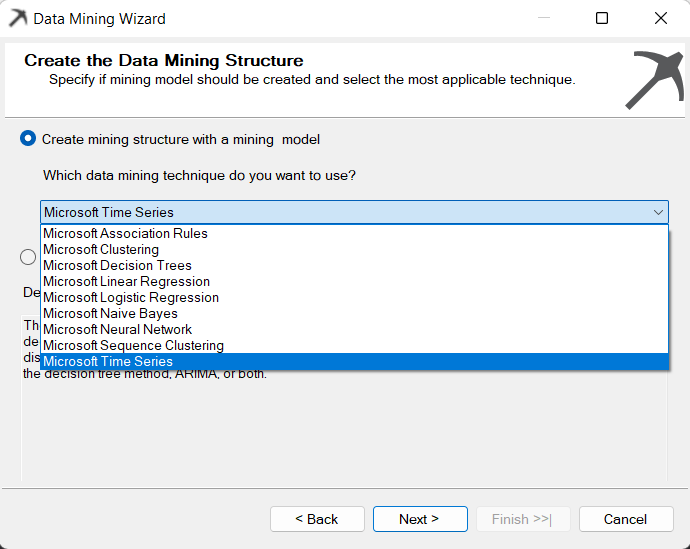


Chọn Next

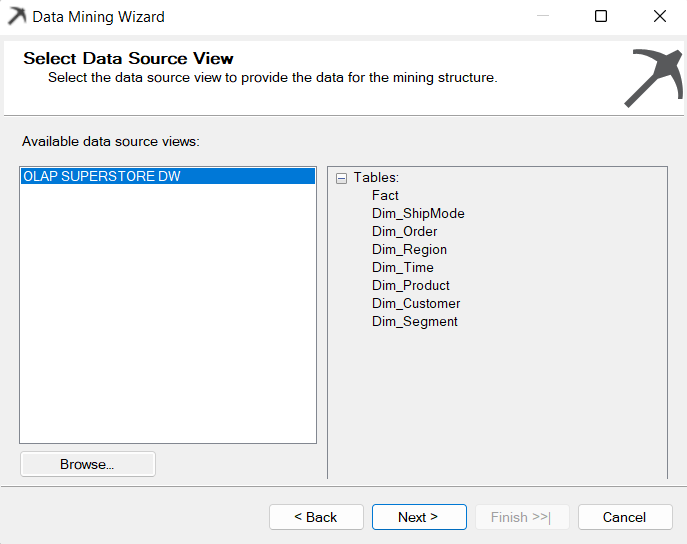


## Khai phá với Time Series

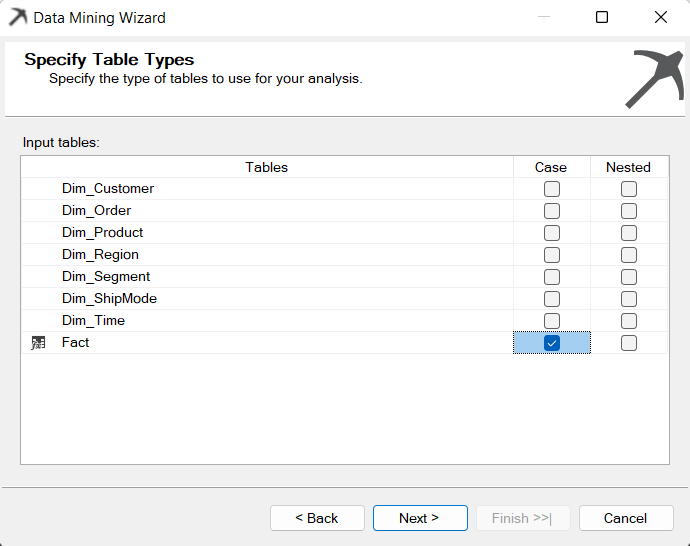
Chọn Time Series và nhấn Next.



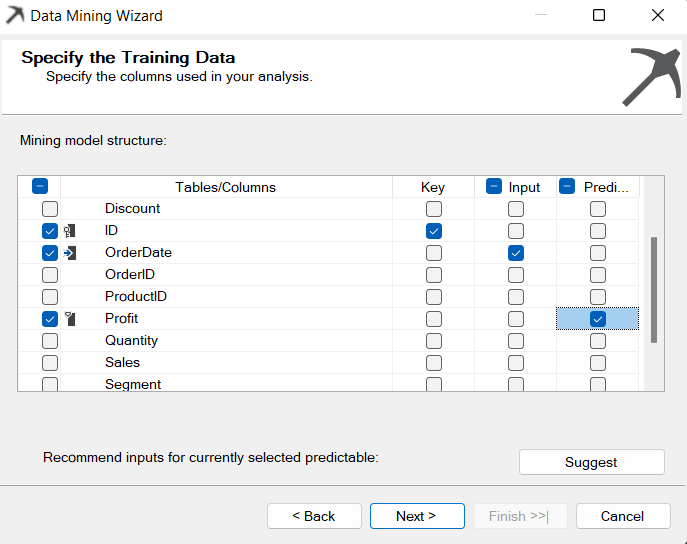
Chọn Next.



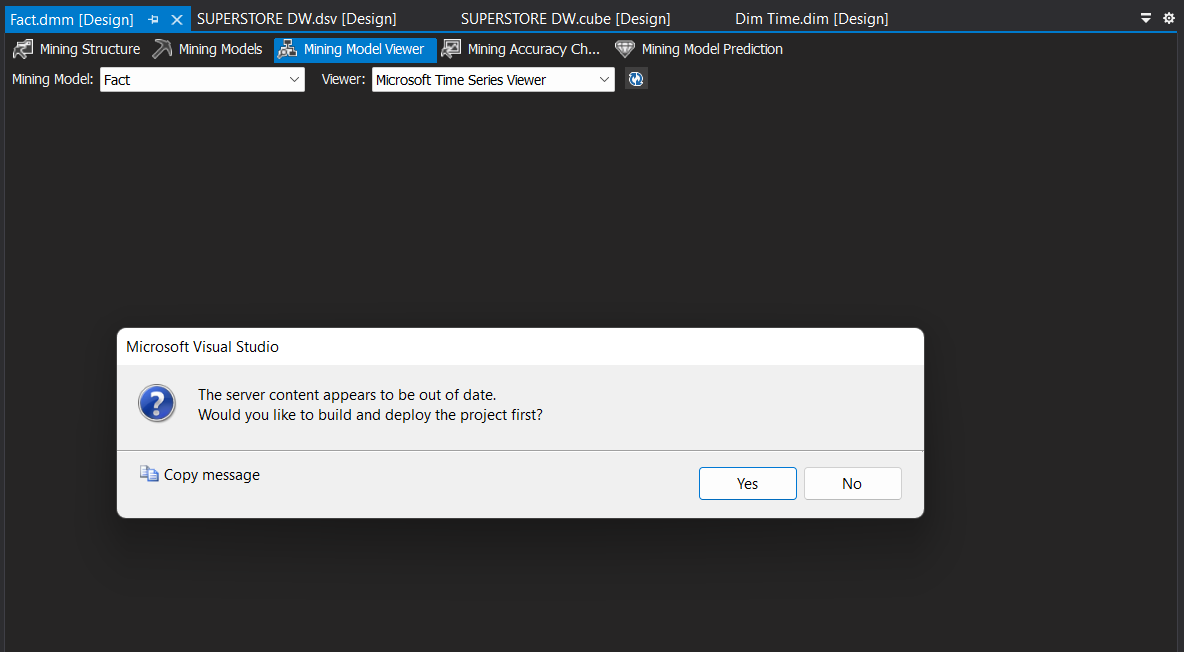
Input table ta chọn bảng Fact và nhấn Next



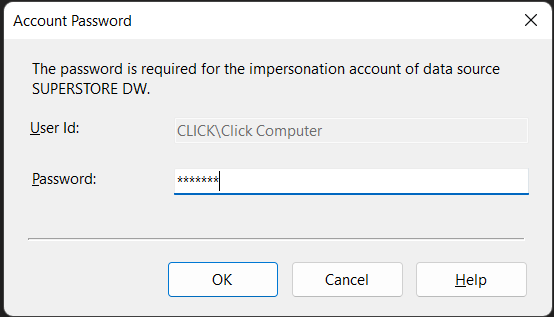
Ở cột Input chọn OrderDate làm dữ liệu đầu vào và chọn Profit là Predictable. Tiếp tục nhấn Next và Finish để kết thúc.



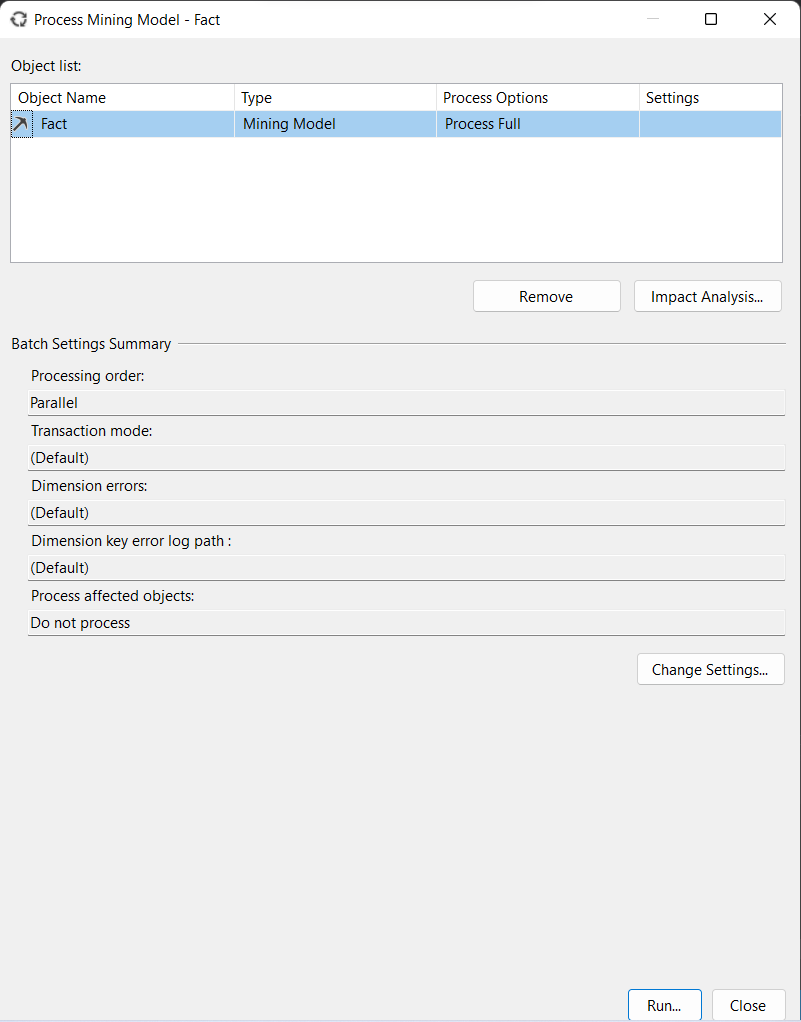
Chọn Mining Model Viewer trên thanh công cụ và nhấn Yes.



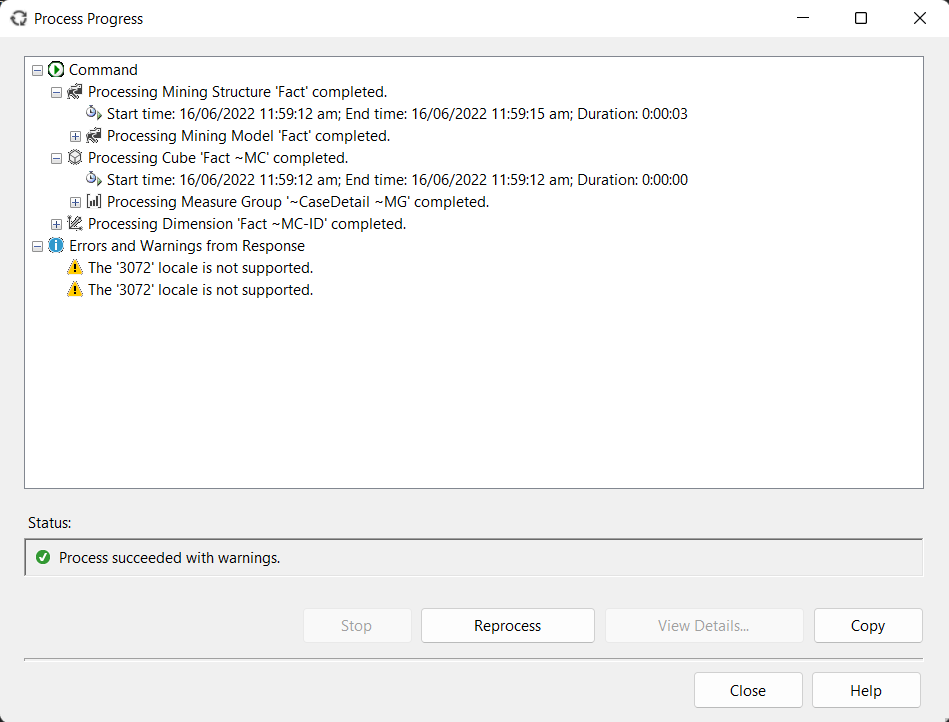
Nhập mật khẩu nếu có yêu cầu và nhấn OK.



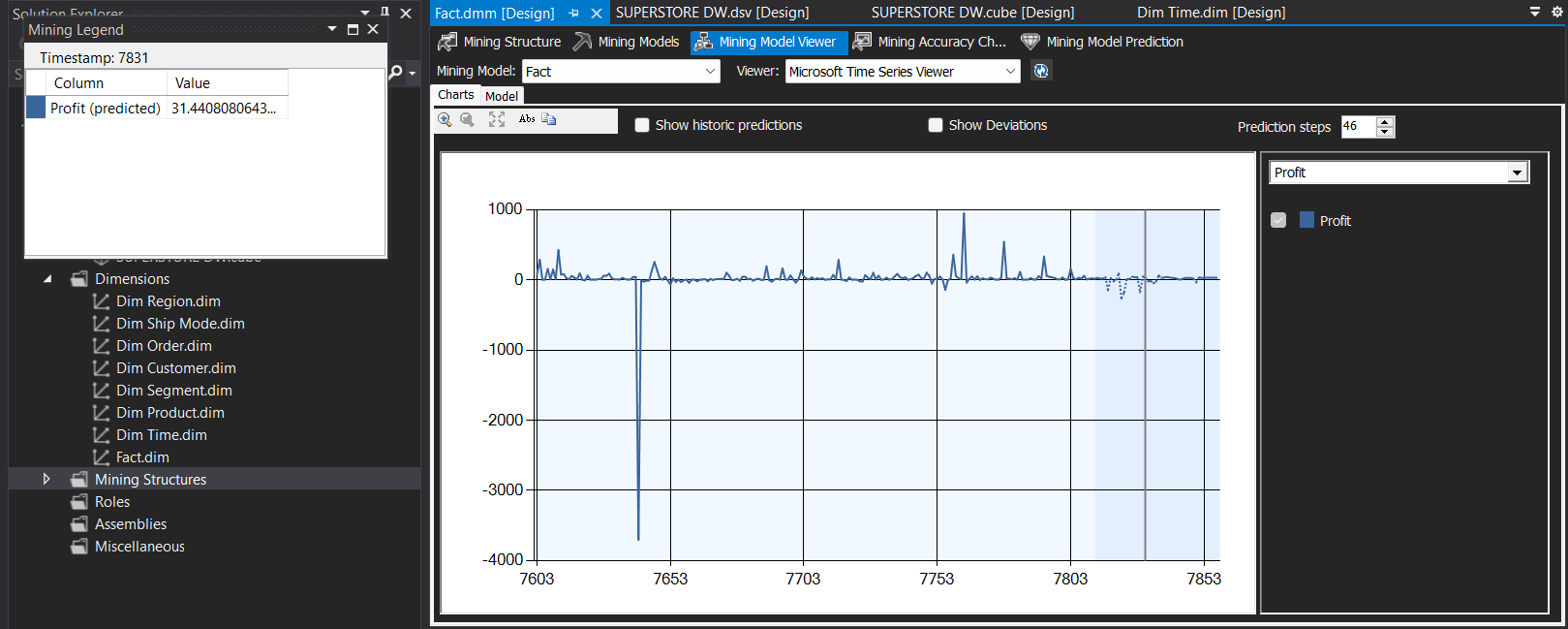
Tiếp tục nhấn Yes và chọn Run để chạy.



Chạy thành công, nhấn Close để đóng.



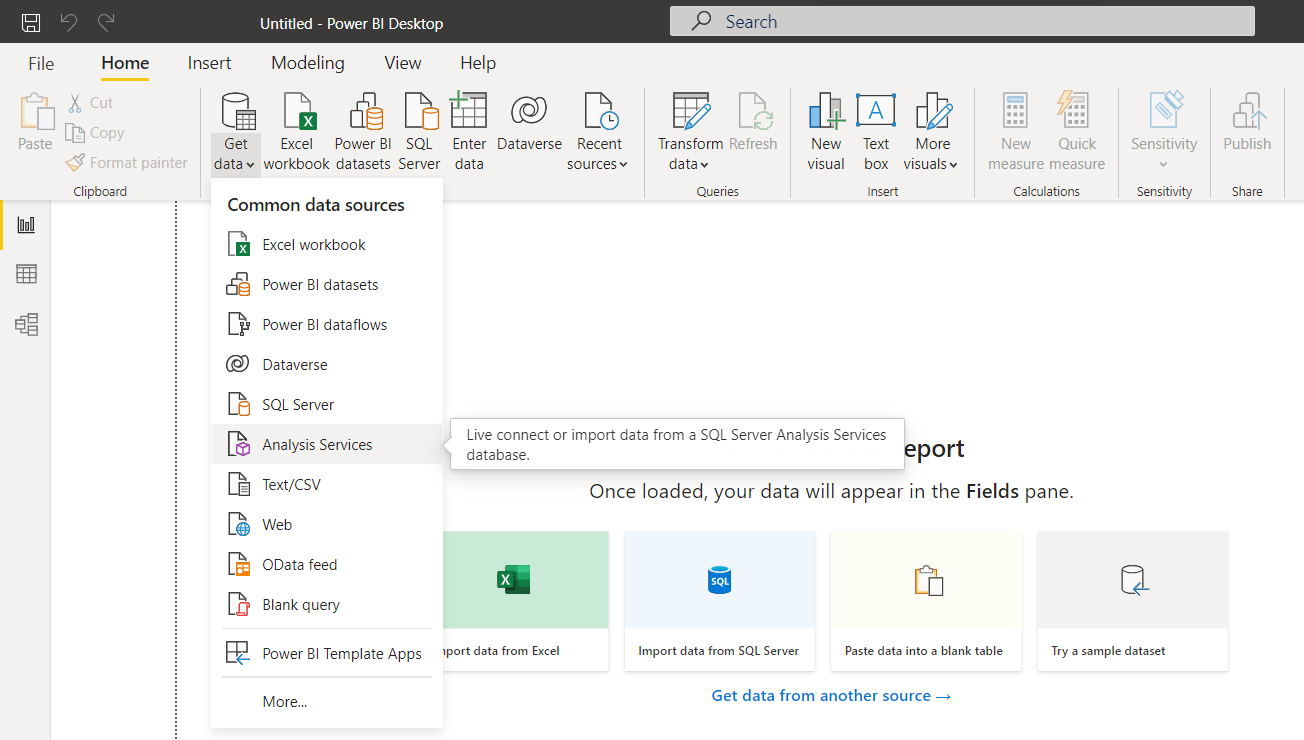
Tăng Prediction Steps để xem dự đoán lợi nhuận của các ngày kế tiếp.



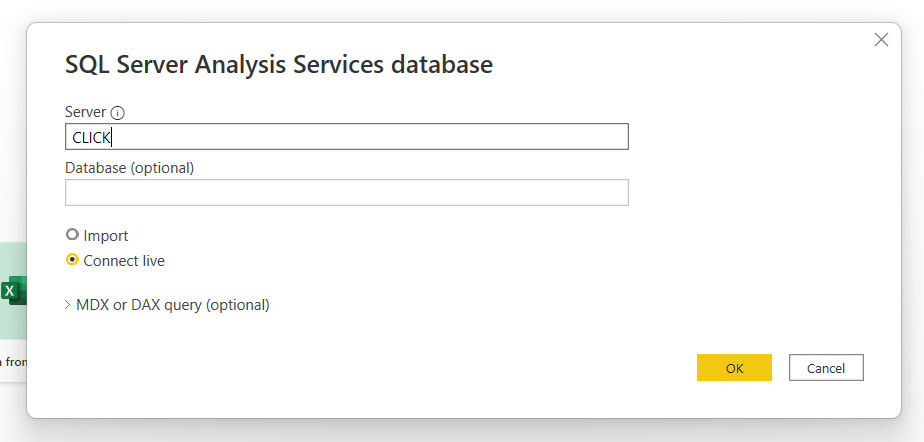
# 5. POWER BI

## Import dữ liệu với Analysis Services Database

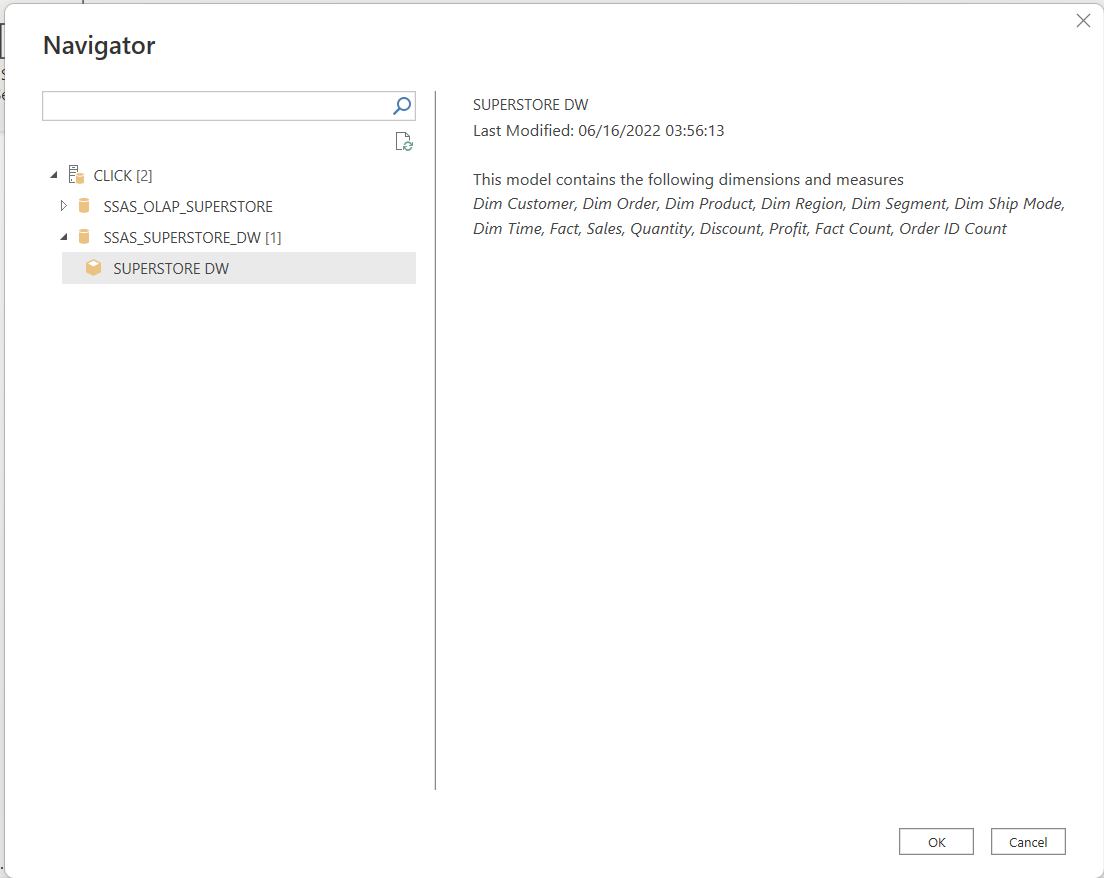
Mở phần mềm Power BI, thực hiện Import Data với project SSAS.



Nhập tên Server và nhấn OK.



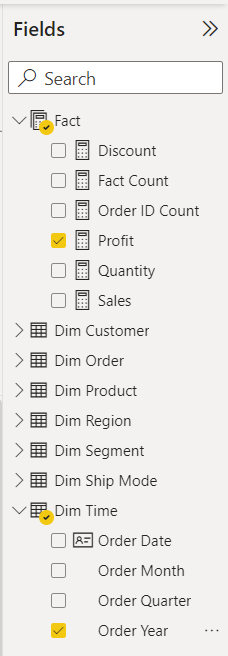
Chọn database SSAS và nhấn OK.



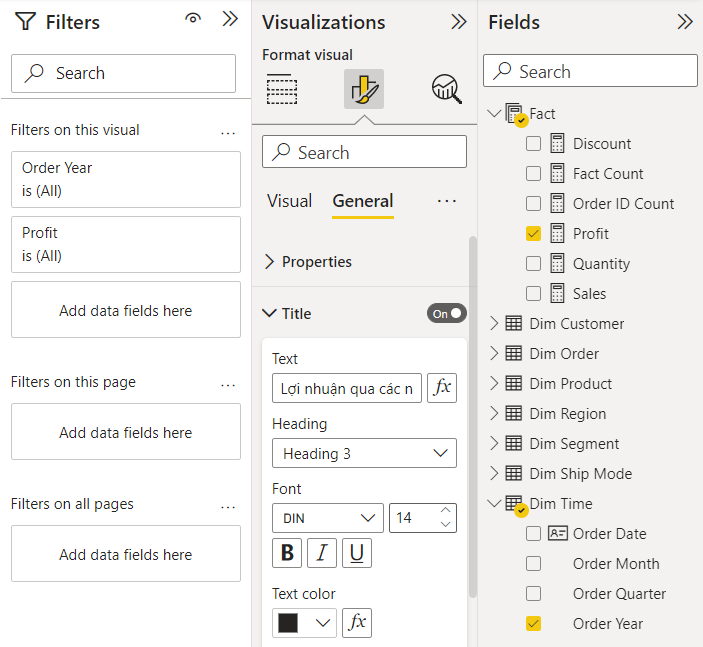
## Thực hiện các báo cáo

### Lợi nhuận qua các năm

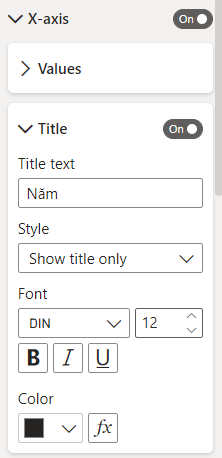
Lựa chọn các thuộc tính cần thiết. Ở đây là Profit và Order Year.

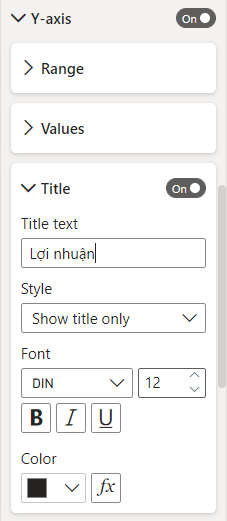


Đổi tên biểu đồ thành Lợi nhuận qua các năm

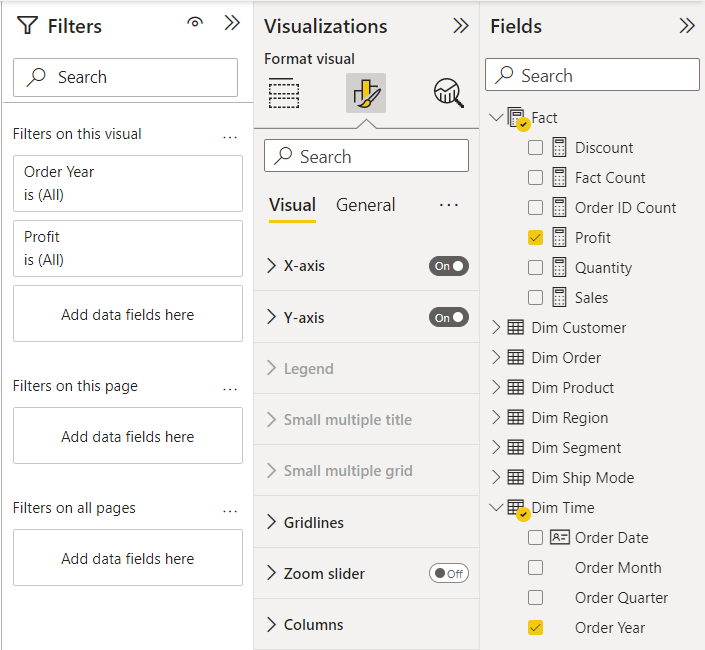


Tương tự thực hiện đổi tên cột X và cột Y

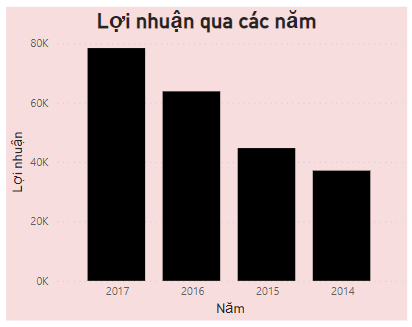




Chọn Format visual để thiết kế chỉnh sửa lại biểu đồ (thay đổi font chữ, màu sắc, .....)

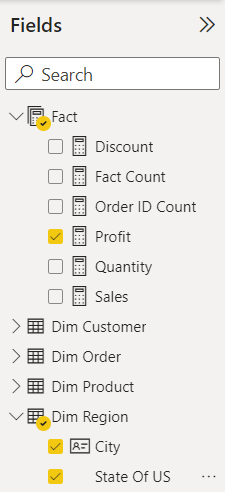


Kết quả

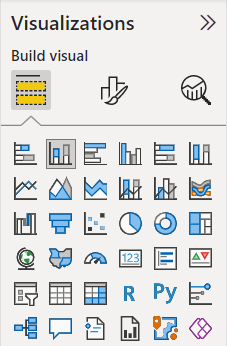


### Lợi nhuận của từng bang, thành phố

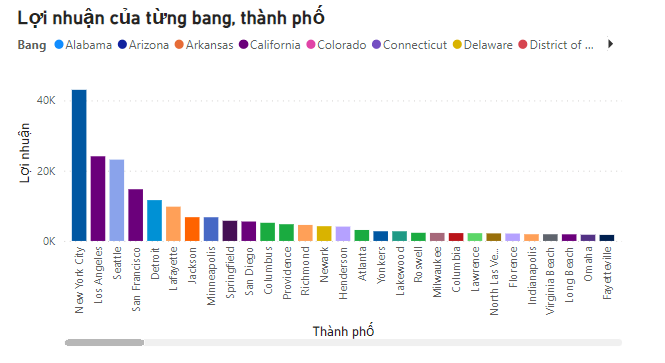
Lựa chọn các thuộc tính.



Lựa chọn loại biểu đồ cho phù hợp

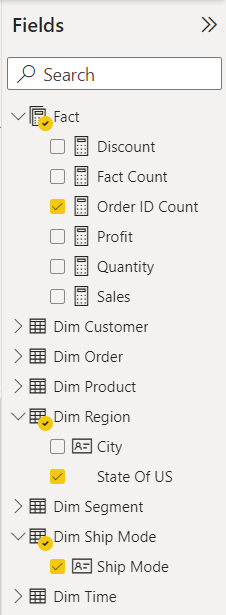


Thực hiện chỉnh sửa lại cho phù hợp. Kết quả.

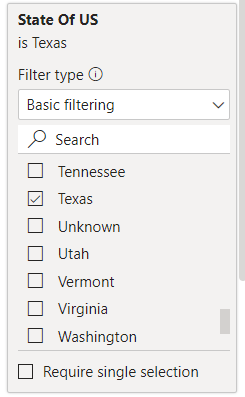


### Hình thức giao hàng phổ biến ở Texas

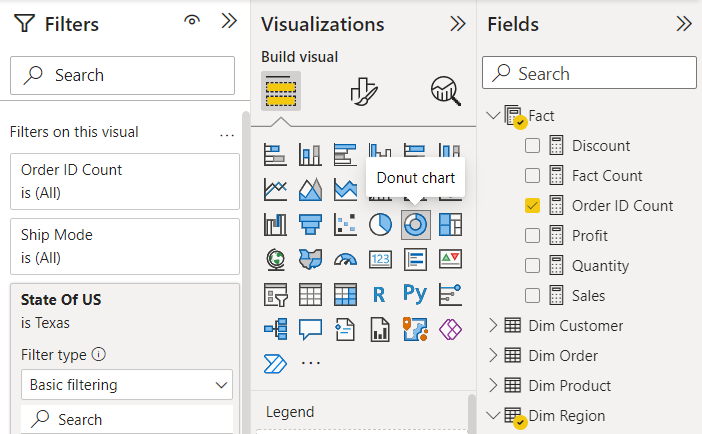
Lựa chọn các thuộc tính cần thiết.



Lọc theo bang Texas.



Chọn loại biểu đồ và chỉnh sửa cho phù hợp.



Kết quả.

