

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

BÁO CÁO ĐỒ ÁN MÔN HỌC



KHO DỮ LIỆU VÀ OLAP

**Đề tài: PHÂN TÍCH DOANH SỐ BÁN HÀNG
DATACO GLOBAL**

GVHD: Ths. Đỗ Thị Minh Phụng

Nhóm 25

Sinh viên thực hiện

Đoàn Ngọc Tuấn - 21521623

Đoãn Công Trí - 21520492

Thành phố Hồ Chí Minh, ngày 22 tháng 05 năm 2024

LỜI CẢM ƠN

Lời đầu tiên nhóm em xin cảm ơn tập thể cán bộ, giảng viên trường Đại học Công Nghệ Thông Tin – ĐHQG Tp. Hồ Chí Minh đã tạo một môi trường tích cực cho sinh viên chúng em được thỏa sức sáng tạo, học tập và làm việc, cũng như trau dồi kiến thức cho chính bản thân mình.

Và đặc biệt, nhóm xin gửi lời cảm ơn chân thành và sự tri ân sâu sắc đối với cô Đỗ Thị Kim Phụng - giảng viên lý thuyết môn Kho dữ liệu và OLAP. Cảm ơn vì sự nhiệt tình, chu đáo, tận tâm của cô và trong công tác giảng dạy cũng như hỗ trợ nhóm em trong quá trình thực hiện đồ án môn học này.

Ngoài ra không thể không cảm ơn tập thể lớp IS217.O21 nói chung và những thành viên trong nhóm nói riêng đã có những đóng góp, ý kiến đề nhóm có thể cải thiện chất lượng dự án. Cảm ơn vì các bạn đã đồng hành cùng chúng mình trong suốt quá trình thực hiện đồ án môn học.

Trong quá trình làm đồ án môn học, khó tránh khỏi sai sót, rất mong cô bỏ qua. Đồng thời do trình độ lý luận cũng như kinh nghiệm thực tiễn còn hạn chế nên bài báo cáo không thể tránh khỏi những thiếu sót, nhóm em rất mong nhận được ý kiến đóng góp từ cô để nhóm em học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt hơn những đồ án khác trong tương lai.

Em xin chân thành cảm ơn!

Nhóm sinh viên thực hiện

Nhóm 24

NHẬN XÉT CỦA GIẢNG VIÊN

Chương 1: TỔNG QUAN ĐỀ TÀI

1.1. LÝ DO CHỌN ĐỀ TÀI

Hiện nay, các ngành kinh doanh bán lẻ trở nên rất phổ biến, có sự cạnh tranh khốc liệt, một ngành đòi hỏi sự quản lý chặt chẽ và hiểu biết sâu rộng về hành vi mua sắm của khách hàng, xu hướng sản phẩm và hiệu suất bán hàng.

Phân tích doanh số trong lĩnh vực này cung cấp cái nhìn tổng quan về hiệu suất kinh doanh, các mô hình bán hàng, xu hướng của khách hàng và các cơ hội để tối ưu hóa doanh số, giúp các nhà quản lý và người kinh doanh hiểu rõ hơn về hoạt động kinh doanh của họ và đưa ra các quyết định chiến lược thông minh. Từ đó giúp nắm bắt được các cơ hội cạnh tranh và đáp ứng nhanh chóng với thay đổi trong thị trường, tăng doanh số bán hàng, cải thiện lợi nhuận, phát hiện và loại bỏ các vấn đề trong quy trình kinh doanh.

1.2. GIỚI THIỆU KHO DỮ LIỆU

1.2.1. Giới thiệu dữ liệu

1.2.1.1. Nguồn dữ liệu

- Tập dữ liệu về Chuỗi cung ứng được công ty **DataCo Global** sử dụng cho phân tích.
- Tập dữ liệu chủ yếu chứa thông tin về hoạt động vận hành của chuỗi cung ứng của DataCo. Tập dữ liệu bao gồm một số thông tin quan trọng, chẳng hạn như thông tin về khách hàng (người bán), đơn đặt hàng, vận chuyển và sản phẩm. Nó bao gồm dữ liệu từ các hoạt động kinh doanh như bán hàng, xử lý đơn đặt hàng và phân phối gói đơn hàng.
- Dataset có 180.000 dòng khảo sát và 53 thuộc tính. (tính đến tháng 3 năm 2024).
- Nguồn: <https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis>

1.2.1.2 Tiết lộ dữ liệu

- Import thư viện và đọc dữ liệu thô

```
import pandas as pd
df = pd.read_csv('DataCoSupplyChainDataset.csv', encoding='latin1')
df
```

- Tạo thêm cột name từ dữ liệu ban đầu

```
df['Customer Name'] = df['Customer Fname'] + ' ' + df['Customer Lname']
df['Customer Name'].head()
```

```
0      Cally Holloway
1      Irene Luna
2    Gillian Maldonado
3      Tana Tate
4      Orli Hendricks
Name: Customer Name, dtype: object
```

- Lọc ra columns cần thiết cho việc xử lý và phân tích

```
df.columns
Index(['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', 'Customer Country', 'Customer Email', 'Customer Fname', 'Customer Id', 'Customer Lname', 'Customer Password', 'Customer Segment', 'Customer State', 'Customer Street', 'Customer Zipcode', 'Department Id', 'Department Name', 'Latitude', 'Longitude', 'Market', 'Order City', 'Order Country', 'Order Customer Id', 'order date (DateOrders)', 'Order Id', 'Order Item Cardprod Id', 'Order Item Discount', 'Order Item Discount Rate', 'Order Item Id', 'Order Item Product Price', 'Order Item Profit Ratio', 'Order Item Quantity', 'Sales', 'Order Item Total', 'Order Profit Per Order', 'Order Region', 'Order State', 'Order Status', 'Order Zipcode', 'Product Card Id', 'Product Category Id', 'Product Description', 'Product Image', 'Product Name', 'Product Price', 'Product Status', 'shipping date (DateOrders)', 'Shipping Mode', 'Customer Name'], dtype='object')

columns = ['Order Id', 'Type', 'Order Status', 'order date (DateOrders)', 'Sales', 'Market', 'Order Item Discount Rate', 'Latitude', 'Longitude', 'Order Item Quantity', 'Category Id', 'Category Name', 'Days for shipping (real)', 'Days for shipment (scheduled)', 'Delivery Status', 'Shipping Mode', 'Customer Id', 'Customer Country', 'Customer Name', 'Customer Segment', 'Product Card Id', 'Product Name', 'Product Price']
]

# Table: Orders, Customers, Categories, Products, Shipping, Time
```

```
df_res = df[columns]
df_res.head()
```

- Xóa giá trị null trong tập dữ liệu

```
df_res.dropna(inplace=True)
df_res.info()
```

- Đổi lại tên cột để thuận tiện trong việc phân tích và xử lý dữ liệu

```
df_res = df_res.rename(columns={'Order Item Discount Rate': 'discount_rate', 'Order Item Quantity': 'product_quantity',
                                'order date (DateOrders)': 'order_date', 'Days for shipping (real)': 'day_shipping_real',
                                'Days for shipment (scheduled)': 'day_shipment_scheduled', 'Product Card Id': 'product_id'})
```

```
df_res.columns
```

```
Index(['Order Id', 'Type', 'Order Status', 'order_date', 'Sales', 'Market',
       'discount_rate', 'Latitude', 'Longitude', 'product_quantity',
       'Category Id', 'Category Name', 'day_shipping_real',
       'day_shipment_scheduled', 'Delivery Status', 'Shipping Mode',
       'Customer Id', 'Customer Country', 'Customer Name', 'Customer Segment',
       'product_id', 'Product Name', 'Product Price'],
       dtype='object')
```

```
df_res.columns = df_res.columns.str.lower().str.replace(' ', '_')
df_res
```

- Tạo thêm cột shipping_id, order_year, order_quarter, order_month để thuận tiện trong việc tách thành các bảng dimension

```
df['shipping_id'] = range(1, len(df) + 1)
```

```
df.order_date = pd.to_datetime(df.order_date)
```

```
df_date = df[['order_date']]
```

```
df_date['order_month'] = df.order_date.dt.month
df_date['order_quarter'] = df.order_date.dt.quarter
df_date['order_year'] = df.order_date.dt.year
```

1.2.2. Mô tả thuộc tính

STT	Tên thuộc tính	Kiểu dữ liệu	Mô tả
1	order_id	varchar(10)	Mã hóa đơn
2	order_date	datetime	Ngày tạo hóa đơn
3	type	varchar(20)	Loại thanh toán: - DEBIT - TRANSFER - CASH - PAYMENT

4	order_status	varchar(20)	Tình trạng hóa đơn: - COMPLETE - PENDING - LOSED - PENDING_PAYMENT - CANCELED - PROCESSING - SUSPECTED_FRAUD
5	sales	decimal(18,2)	Doanh số bán hàng
6	market	varchar(20)	Nơi đơn hàng được giao: - Africa - Europe - LATAM - Pacific Asia - USCA
7	discount_rate	decimal(18,2)	Phần trăm giảm giá của đơn hàng
8	product_quantity	int	Số lượng sản phẩm đặt mua
9	customer_id	varchar(10)	Mã khách hàng
10	customer_country	varchar(10)	Quốc gia của khách hàng
11	customer_name	varchar(20)	Tên khách hàng
12	customer_segment	varchar(20)	Loại khách hàng: - Consumer - Corporate - Home Office
13	product_id	varchar(10)	Mã sản phẩm
14	product_name	varchar(10)	Tên sản phẩm
15	product_price	decimal(18,2)	Giá sản phẩm
16	category_id	varchar(10)	Mã danh mục sản phẩm
17	category_name	varchar(50)	Tên danh mục sản phẩm
18	shipping_id	int	Mã giao hàng

19	shipping_mode	varchar(20)	Chế độ vận chuyển: - Standard Class - First Class - Second Class - Same Day
20	day_shipping_real	int	Thời gian giao hàng thật sự
21	day_shipping_scheduled	int	Thời gian giao hàng dự kiến
22	delivery_status	varchar(20)	Tình trạng giao hàng: - Advance shipping - Late delivery - Shipping canceled - Shipping on time

1.2.3. Thiết kế kho dữ liệu

1.2.3.1. Sơ đồ xây dựng dữ liệu

Lược đồ đa chiều (multidimensional schema) là một cấu trúc dữ liệu được sử dụng để tổ chức các thông tin trong các hệ thống quản lý cơ sở dữ liệu, nhằm giúp cho việc truy xuất và phân tích dữ liệu trở nên dễ dàng và nhanh chóng hơn.

Lược đồ đa chiều được sử dụng trong các hệ thống OLAP (Online Analytical Processing) để tổ chức dữ liệu theo các kết nối giữa các thuộc tính và các số liệu có liên quan.

Có 2 loại lược đồ đa chiều: Lược đồ hình sao (star schema) và lược đồ hình bông tuyết (snowflake schema).

a. Lược đồ hình sao (Star schema)

Lược đồ hình sao (Star schema) là một loại lược đồ đa chiều được sử dụng trong hệ thống quản lý cơ sở dữ liệu phân tích (OLAP) và các ứng dụng Business Intelligence (BI). Nó được xây dựng dựa trên mô hình hóa dữ liệu hướng tài nguyên

(Resource-oriented data modeling) và được sử dụng để tổ chức dữ liệu cho việc thực hiện các phân tích dữ liệu.

Trong lược đồ hình sao, dữ liệu được tổ chức thành một bảng trung tâm (fact table) và các bảng chi tiết (dimension tables) liên quan đến bảng trung tâm. Bảng trung tâm chứa các số liệu định lượng, trong khi các bảng chi tiết chứa các thông tin phân loại (thuộc tính). Mỗi bảng chi tiết đại diện cho một chiều của dữ liệu và được kết nối với bảng trung tâm thông qua một khóa ngoại.

- **Ưu điểm:**

- *Tốc độ truy vấn nhanh:* Do dữ liệu được tổ chức thành các bảng trung tâm và các bảng chi tiết liên quan, việc truy vấn và phân tích dữ liệu trở nên nhanh chóng và hiệu quả hơn => xử lý khối lượng dữ liệu lớn dễ dàng hơn.
- *Thích hợp cho phân tích dữ liệu:* Lược đồ hình sao được thiết kế để phục vụ cho các mục đích phân tích dữ liệu và đưa ra quyết định. Các bảng chi tiết đại diện cho các thuộc tính của dữ liệu, giúp cho người dùng có thể thực hiện các phân tích đa chiều và tìm kiếm các mối quan hệ giữa các thuộc tính khác nhau.
- *Linh hoạt và dễ dàng mở rộng:* Lược đồ hình sao linh hoạt và có thể dễ dàng mở rộng khi cần thiết. Nếu cần thêm các thuộc tính mới vào hệ thống, người dùng chỉ cần thêm các bảng chi tiết mới và kết nối chúng với bảng trung tâm.
- *Dễ dàng hiểu và sử dụng:* Lược đồ hình sao đơn giản và dễ hiểu. Các bảng trung tâm và bảng chi tiết được tổ chức theo cấu trúc phẳng, giúp cho người dùng dễ dàng hình dung và phân tích dữ liệu.

- **Nhược điểm:**

- *Dữ liệu trùng lặp:* Một số dữ liệu có thể bị trùng lặp trong các bảng chi tiết. Điều này có thể xảy ra khi các bảng chi tiết có liên kết với nhiều bảng trung tâm khác nhau => làm tăng khối lượng dữ liệu, ảnh hưởng đến hiệu suất.
- *Không phù hợp cho các tình huống phức tạp:* Lược đồ hình sao là một mô hình đơn giản, chỉ phù hợp với các tình huống phân tích dữ liệu đơn giản.
- *Không linh hoạt cho các loại truy vấn khác nhau:* Khi cần thực hiện các loại truy vấn như truy vấn thống kê hoặc truy vấn văn bản đầy đủ, lược đồ hình sao có thể không phù hợp.

b. Lược đồ hình bông tuyết (Snowflake schema)

Lược đồ hình bông tuyết (Snowflake schema) là một mô hình quản lý cơ sở dữ liệu phân tích, tương tự như lược đồ hình sao (Star schema). Tuy nhiên, lược đồ hình bông tuyết khác với lược đồ hình sao ở cách tổ chức bảng chi tiết.

Trong lược đồ hình bông tuyết, các bảng chi tiết được phân chia thành các bảng con, và các bảng con này được lưu trữ ở mức độ bậc thấp hơn. Điều này có nghĩa là lược đồ hình bông tuyết có nhiều mức độ lưu trữ dữ liệu hơn so với lược đồ hình sao. Lược đồ hình bông tuyết được sử dụng để xử lý các bộ dữ liệu lớn và phức tạp hơn, và cung cấp cho người dùng khả năng phân tích chi tiết hơn về các mối quan hệ giữa dữ liệu.

· **Ưu điểm:**

- *Tối ưu hóa khả năng truy vấn:* Tối ưu hóa khả năng truy vấn của cơ sở dữ liệu phân tích, giúp tăng tốc độ truy vấn và giảm thời gian phản hồi.
- *Hiệu quả trong việc lưu trữ dữ liệu:* Sử dụng một cấu trúc lưu trữ phân cấp, giúp tối ưu hóa việc lưu trữ dữ liệu và giảm thiểu khối lượng dữ liệu lưu trữ.

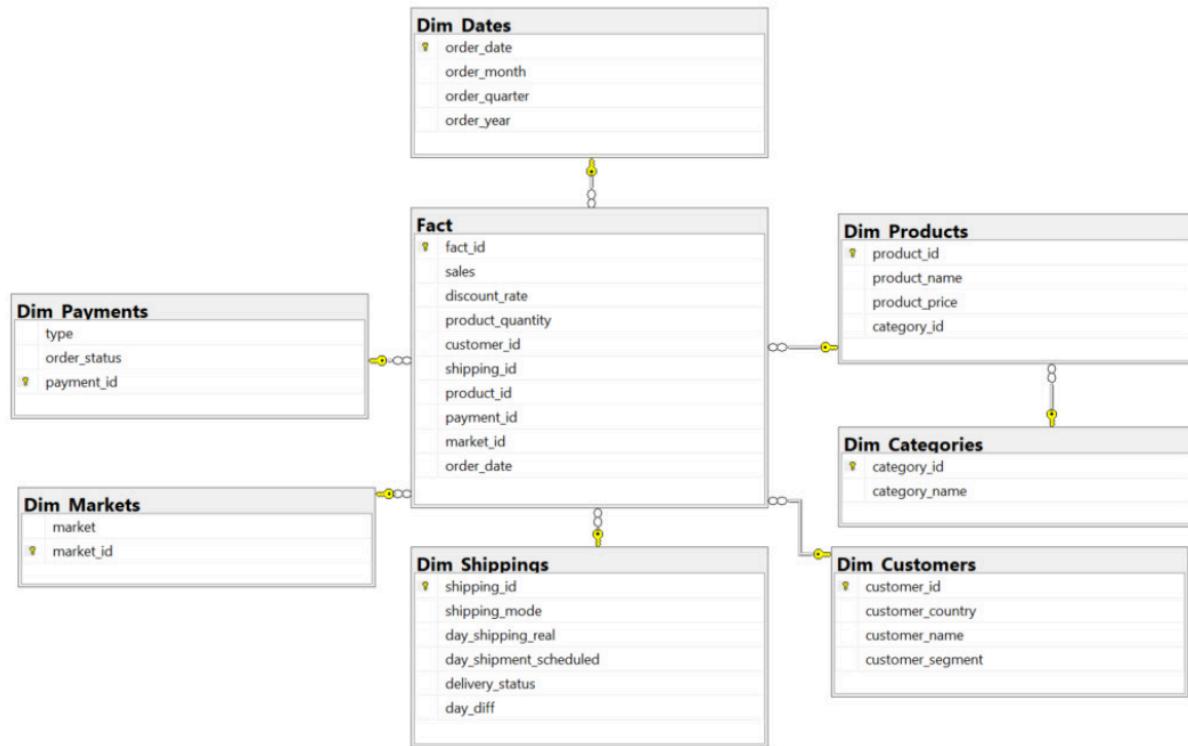
- *Khả năng phân tích chi tiết:* Cung cấp cho người dùng khả năng phân tích chi tiết hơn về các mối quan hệ giữa dữ liệu, vì các bảng con được phân tách ra từ các bảng chi tiết là các nhánh phụ của bông tuyêt.
- *Dễ dàng mở rộng:* Dễ dàng mở rộng và bổ sung thêm các bảng con mới vào cấu trúc hiện có.

- **Nhược điểm:**

- *Sử dụng tài nguyên lưu trữ lớn:* Vì các bảng con được phân tách ra từ các bảng chi tiết.
- *Khó khăn trong việc thiết kế và triển khai:* Vì cấu trúc phức tạp hơn.
- *Khó khăn trong việc truy vấn bảng con:* Có thể xảy ra hiện tượng thừa mật độ dữ liệu khi truy vấn, vì các bảng con có thể chứa nhiều thông tin trùng lặp với các bảng khác.
- *Khó khăn trong việc đáp ứng các yêu cầu thay đổi:* Việc thay đổi cấu trúc của lược đồ có thể phức tạp và đòi hỏi sự chú ý đến tính nhất quán của dữ liệu.

1.2.3.2 Thiết kế lược đồ

Lược đồ bông tuyêt gồm 8 bảng. Trong đó bảng Fact là bảng sự kiện chứa các khóa ngoại của các bảng Dimension và các thuộc tính khác. Dim_Customers, Dim_Products, Dim_Catgories, Dim_Shippings, Dim_Dates, Dim_Markets, Dim_Payments là các bảng chiều chứa thuộc tính chính và các thuộc tính khác của bảng.



1.2.3.3 Các bảng chiều

Bảng Dim_Customers

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
1	customer_id	varchar(10)	Primary key
2	customer_country	varchar(10)	
3	customer_name	varchar(20)	
4	customer_segment	varchar(20)	

Bảng Dim_Products

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
-----	----------------	--------------	-----------

1	product_id	varchar(10)	Primary key
2	product_name	varchar(10)	
3	product_price	decimal(18,2)	
4	category_id	varchar(10)	

Bảng Dim_Categories

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
1	category_id	varchar(10)	Primary key
2	category_name	varchar(50)	

Bảng Dim_Shippings

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
1	shipping_id	varchar(10)	Primary key
2	shipping_mode	varchar(20)	
3	day_shipping_real	int	
4	day_shipping_scheduled	int	

Bảng Dim_Dates

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
1	order_date	datetime	Primary key
2	order_year	int	
3	order_quarter	int	

4	order_month	int	
---	-------------	-----	--

Bảng Dim_Payments

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
1	payment_id	varchar(10)	Primary key
2	type	varchar(20)	
3	order_status	varchar(20)	

Bảng Dim_Markets

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
1	market_id	varchar(10)	Primary key
2	market	varchar(50)	

1.2.3.4 Bảng sự kiện

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
1	fact_id	varchar(10)	Primary key
2	product_id	varchar(10)	Foreign key
3	customer_id	varchar(10)	Foreign key
4	shipping_id	varchar(10)	Foreign key
5	payment_id	varchar(10)	Foreign key
6	market_id	varchar(10)	Foreign key
7	order_date	datetime	Foreign key

8	sales	decimal(18,2)	
9	discount_rate	decimal(18,2)	
10	product_quantity	int	

1.2.4. Các câu truy vấn

Câu 1: Liệt kê Top 5 sản phẩm bán chạy nhất theo các năm.

Câu 2: Tổng doanh số bán hàng theo từng khu vực trong năm 2016, sắp xếp tăng dần theo doanh số

Câu 3: Tìm ra danh mục có số lượng sản phẩm bán được cao nhất

Câu 4: Tính doanh thu trung bình trên 1 sản phẩm bán ra vào các quý của năm 2016, thuộc danh mục Fishing

Câu 5: Tìm ra khách hàng có số lần mua hàng cao nhất và thấp nhất trong năm 2017

Câu 6: Liệt kê Top 5 sản phẩm bán chạy nhất Africa

Câu 7: Cho biết khu vực nào có số lượng đơn hàng bị hủy cao nhất theo từng năm.

Câu 8: Cho biết loại hình thanh toán nào được ưa chuộng nhất ở LATAM

Câu 9: Liệt kê doanh số đóng góp của nhóm khách hàng ‘Corporate’ theo từng quý, năm.

Câu 10: Liệt kê những khách hàng có số lượng đơn hàng đã mua > 2 và < 5

Câu 11: Liệt kê số lượng đơn hàng bán ra theo từng năm ở tất cả thị trường, trừ thị trường ‘LATAM’.

Câu 12: Truy vấn ra các đơn vận chuyển có phần trăm giảm giá > 0.2 theo từng khu vực trong năm 2015, sắp xếp giảm dần theo phần trăm giảm giá.

Câu 13: Truy vấn ra các đơn hàng được giao đúng thời gian dự kiến

Câu 14: Liệt kê Top 5 sản phẩm có số lượng bán cao nhất với ký tự bắt đầu là ‘P’, sắp xếp theo thứ tự giảm dần.

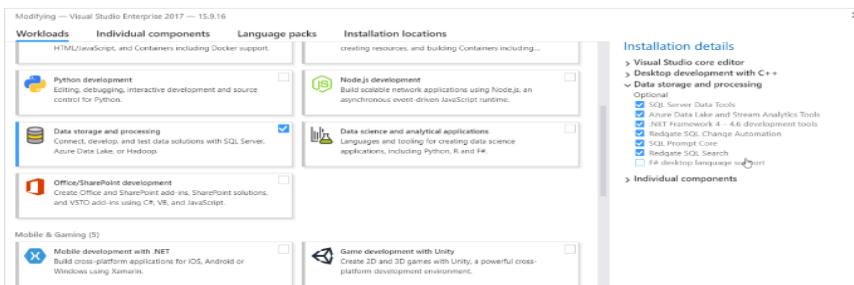
Câu 15: Với mỗi danh mục sản phẩm, liệt kê 3 sản phẩm có doanh số đóng góp cao nhất trong năm 2015

Chương 2: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

2.1. CHUẨN BỊ CÔNG CỤ VÀ DATA WAREHOUSE

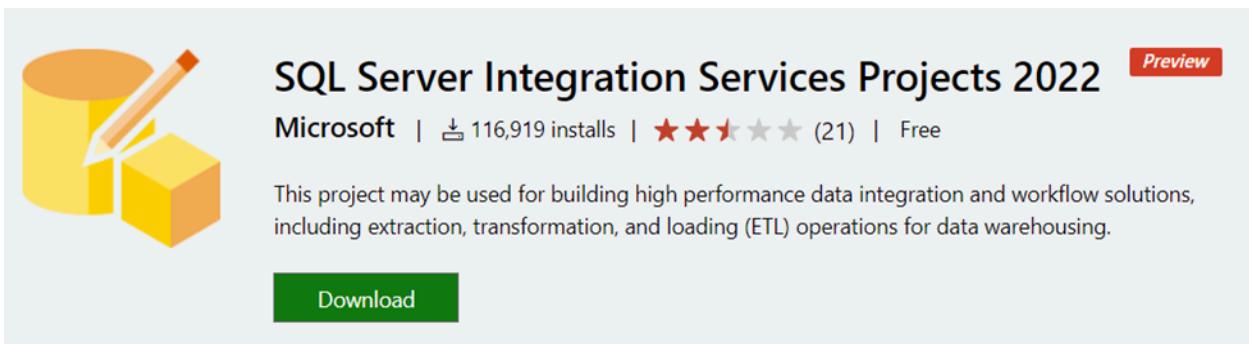
2.1.1 Chuẩn bị công cụ

- Download công cụ SQL Server Data Tools



Chọn **Data storage and processing** và chọn những mục như hình minh họa

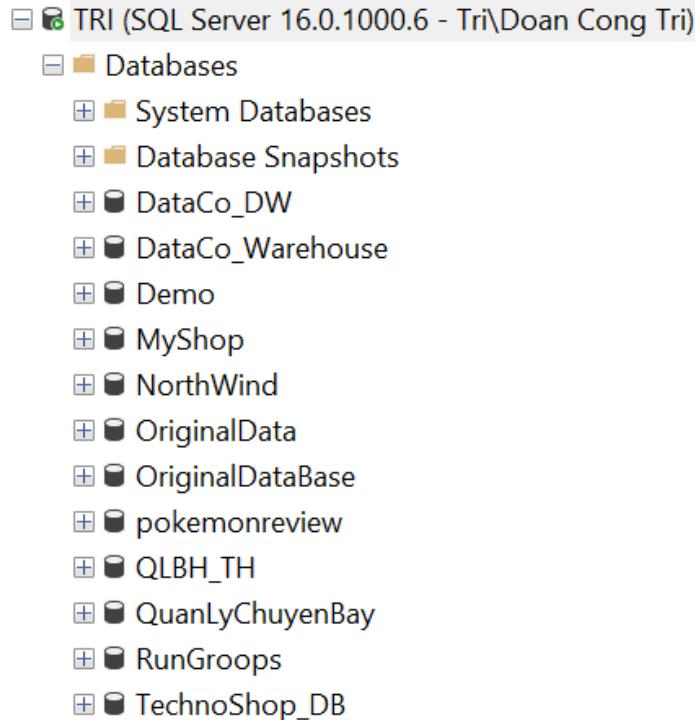
- Vào trang web [SQL Server Integration Services Projects – Visual Studio 2022](#) để download extension cho quá trình SSIS



Hình 2. 5: Create a new project trên Visual Studio 2022

2.1.2. Tạo cơ sở dữ liệu

- Tạo hai cơ sở dữ liệu trên Microsoft SQL Server Management Studio



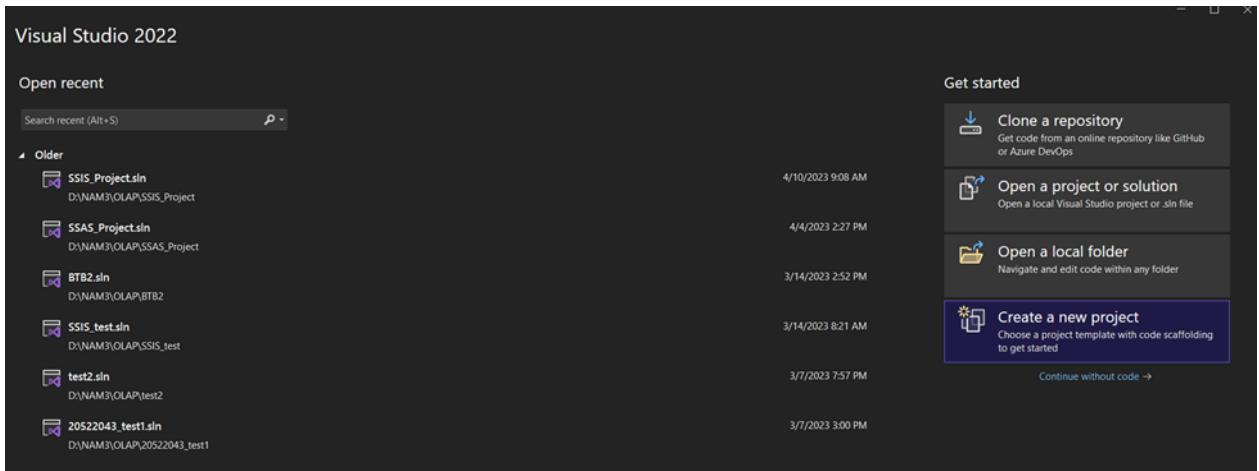
Hình 2. 5: Create a new project trên Visual Studio 2022

- OriginalDataBase: là database chứa dữ liệu gốc, gồm các bảng
 - RawData: Chứa dữ liệu trước khi làm sạch
 - CleanData: Chứa dữ liệu đã làm sạch
- DataCo_Warehouse: là database chứa các bảng Fact và bảng Dimensions được thiết kế

2.2. TẠO PROJECT SSIS VÀ THIẾT LẬP KẾT NỐI

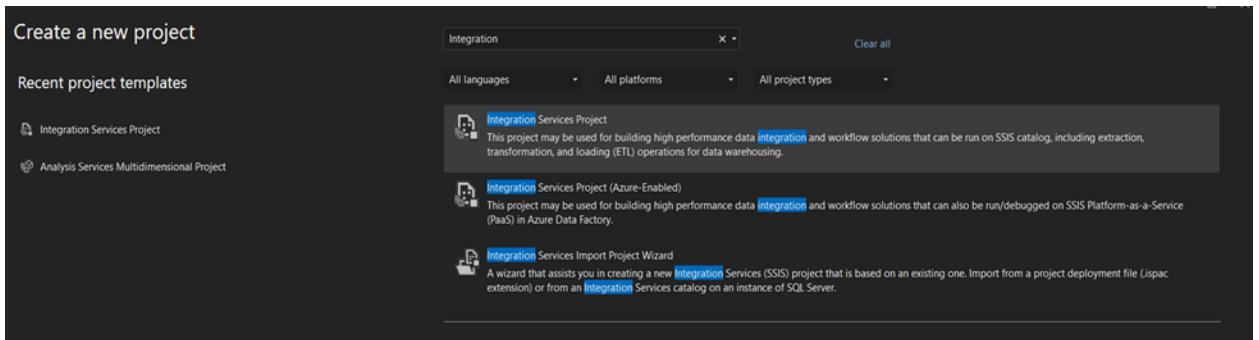
2.2.1. Tạo project SSIS và thiết lập kết nối

- Mở Visual Studio 2022 và chọn Create new project



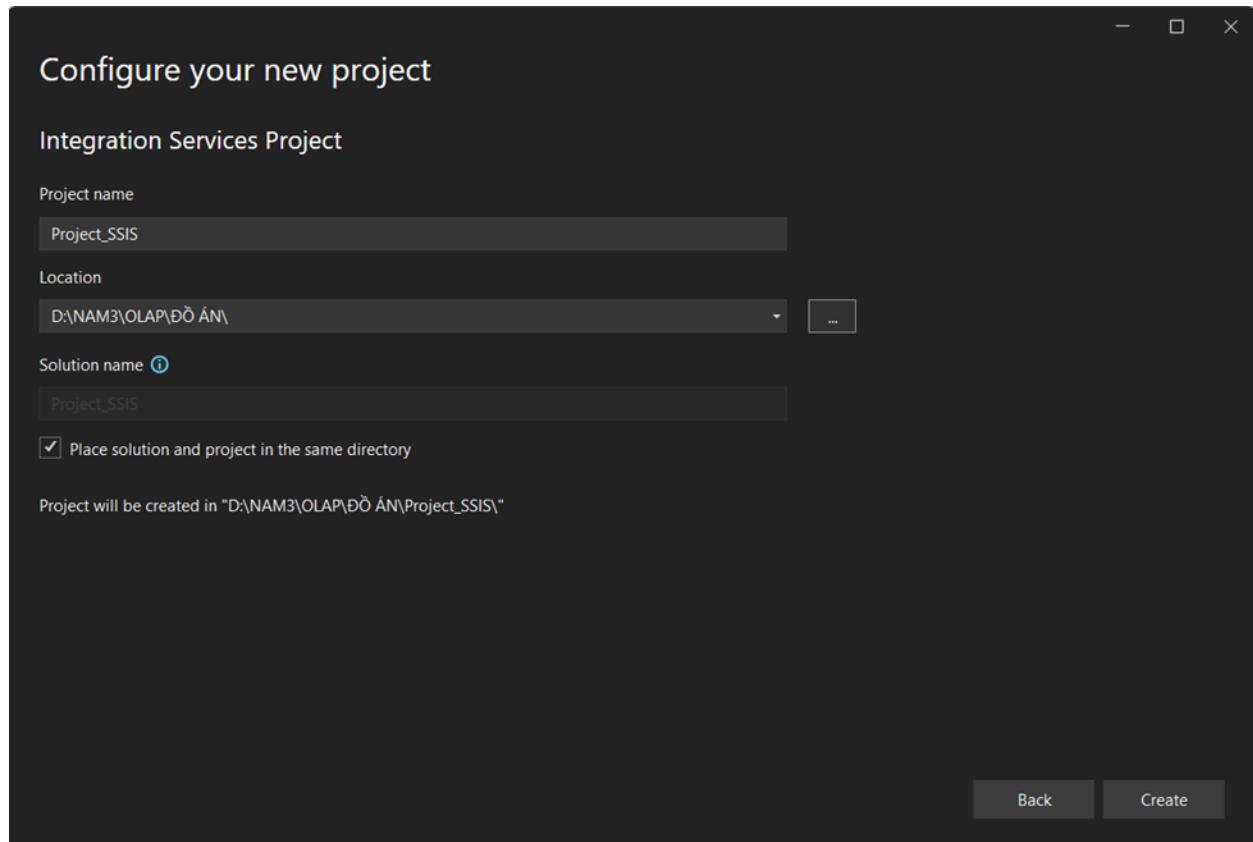
Hình 2. 5: Create a new project trên Visual Studio 2022

- Ở thanh tìm kiếm, tìm công cụ **Integration Services Project**



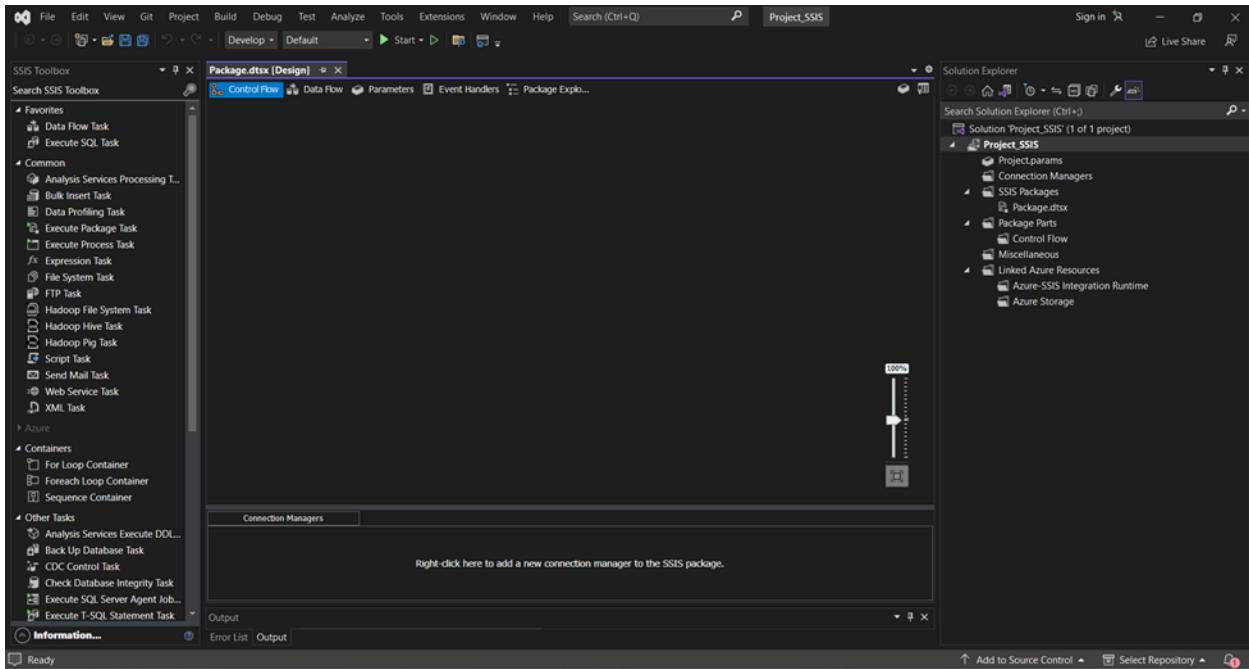
Hình 2. 5: Create a new project trên Visual Studio 2022

- Đặt tên file project và xác nhận nơi lưu.



Hình 2. 5: Create a new project trên Visual Studio 2022

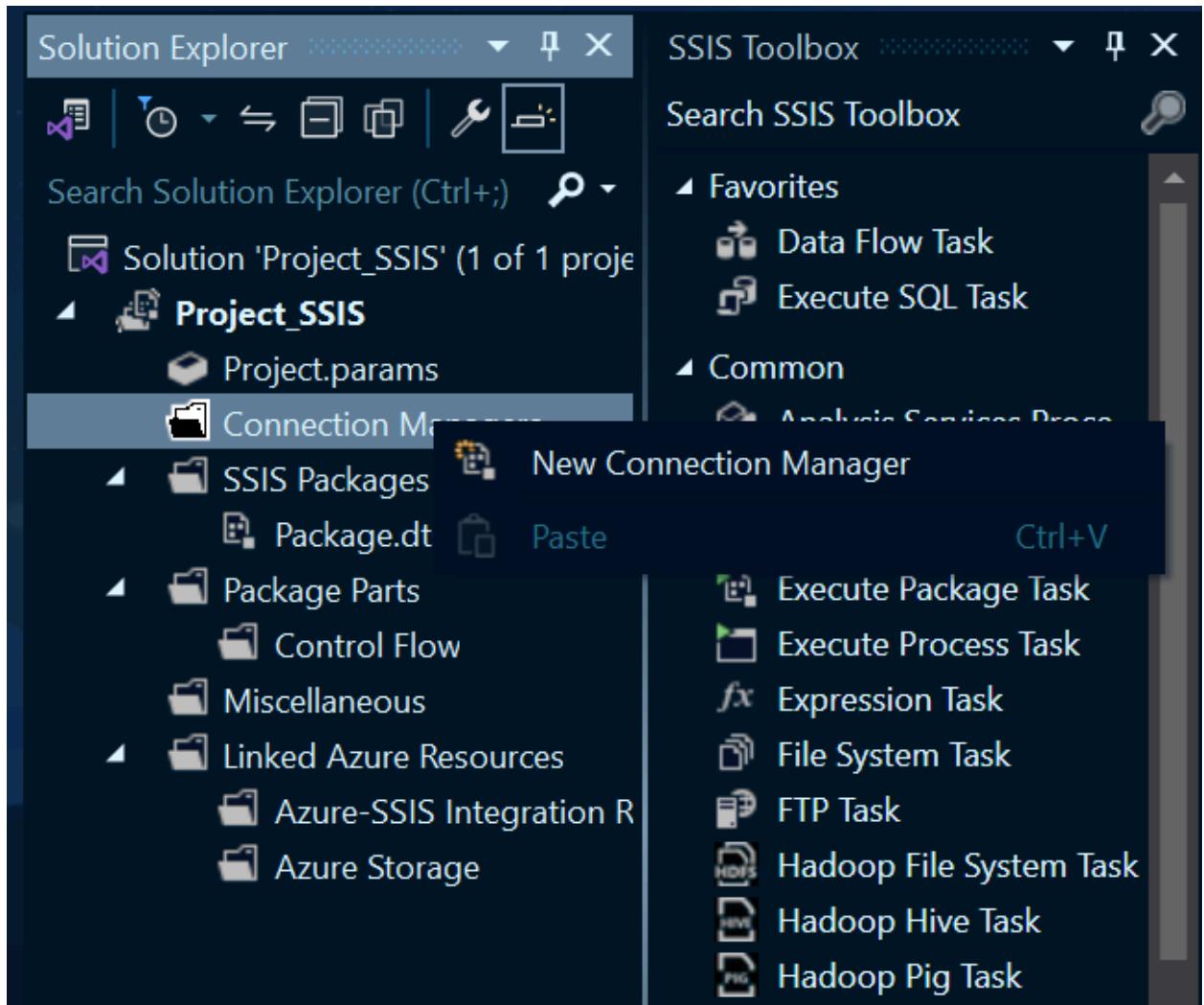
- Giao diện SSIS



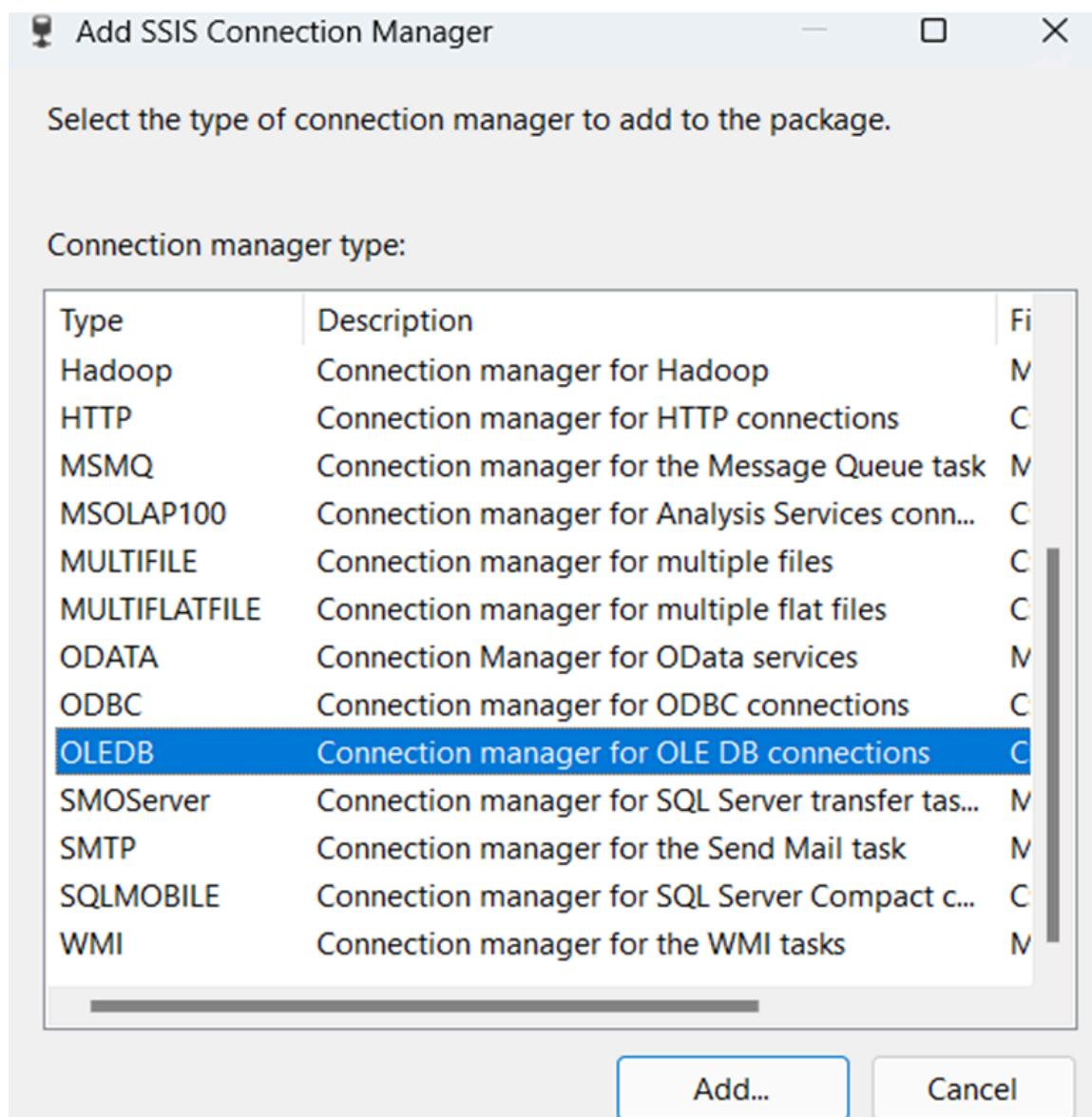
Hình 2. 5: Create a new project trên Visual Studio 2022

2.2.1. Thiết lập kết nối

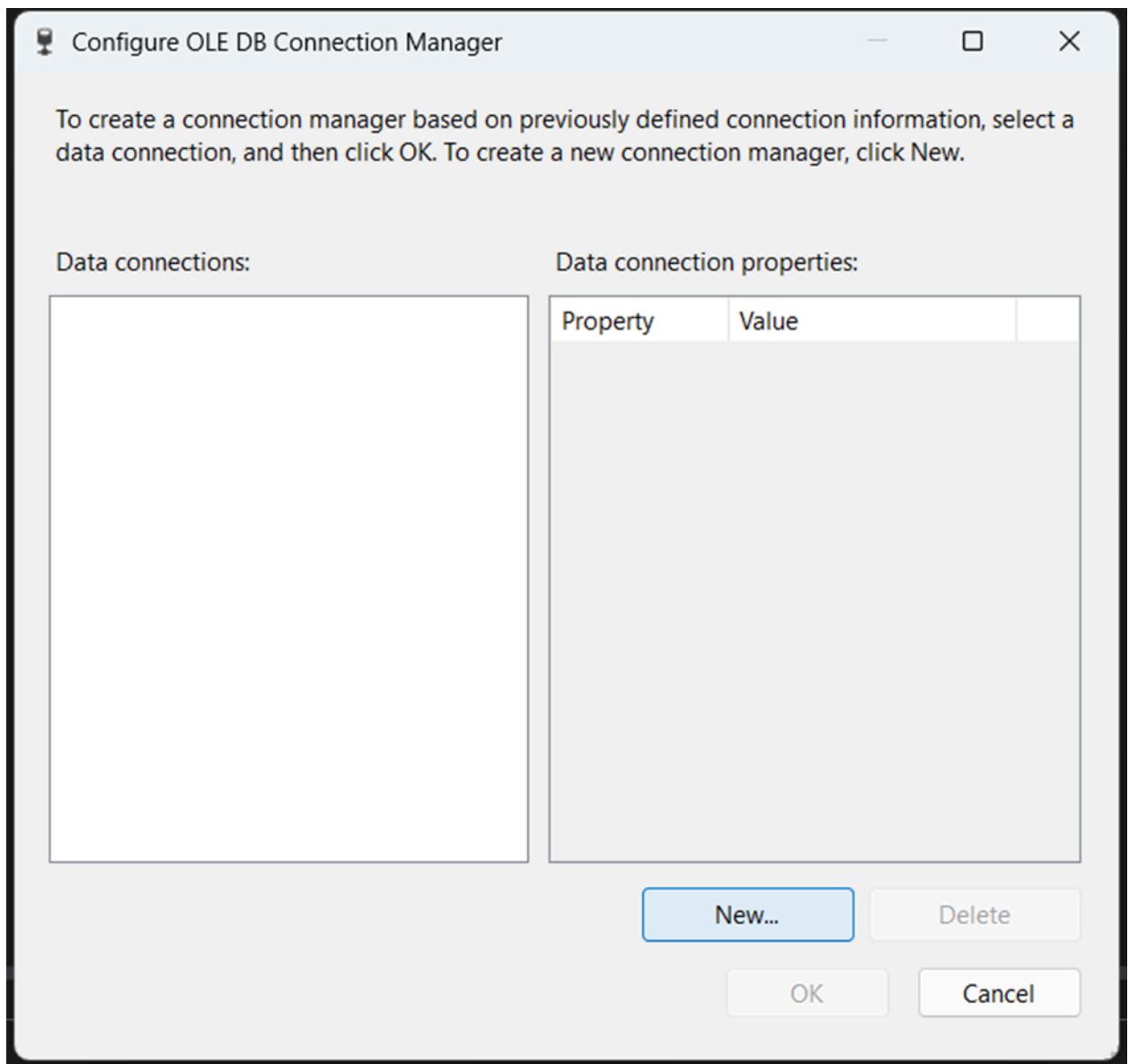
- Tại phần Solution Explorer, click chuột phải vào Connection Managers, chọn New Connection Manager để tiến hành tạo kết nối mới



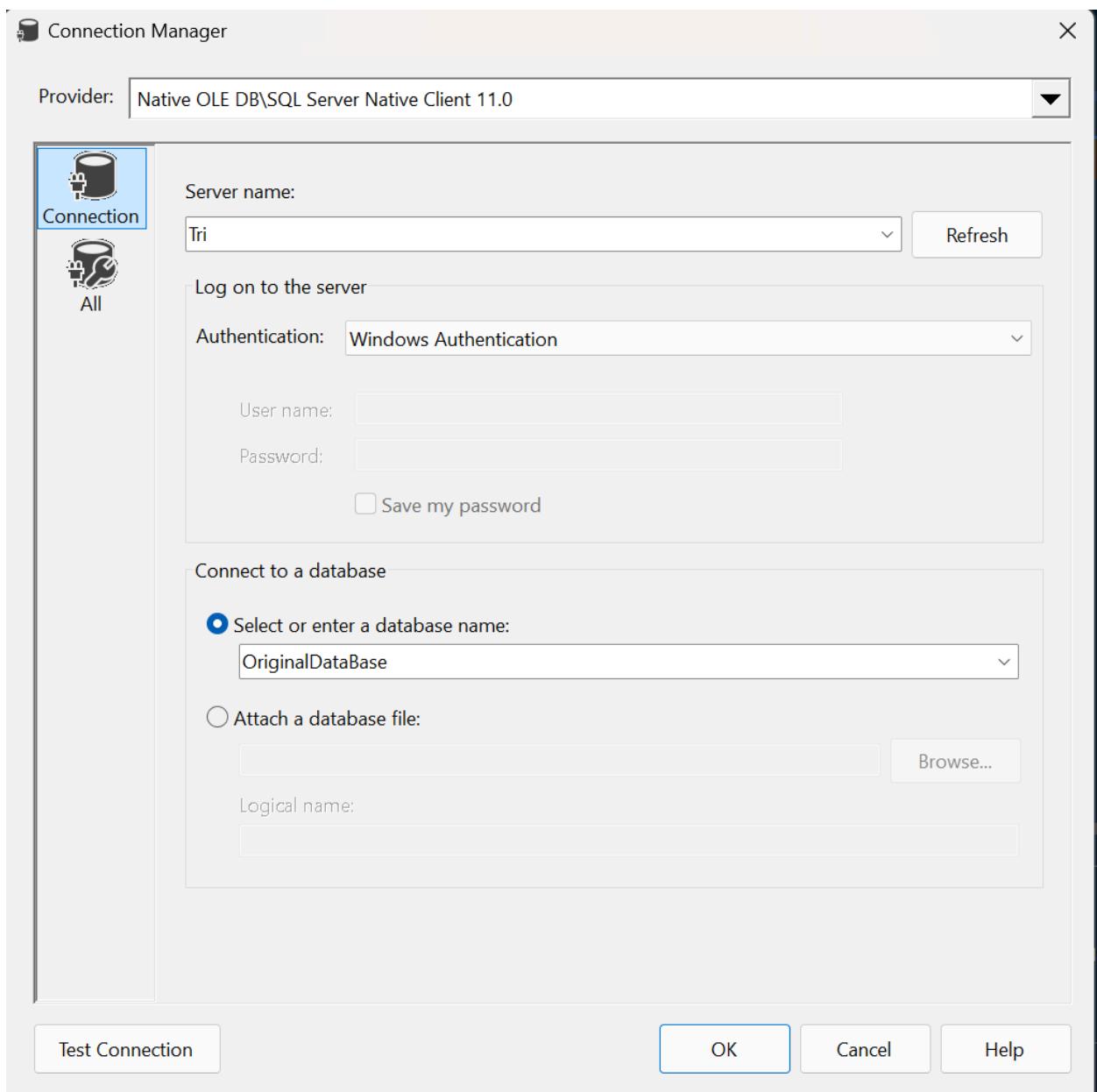
- Chọn Chọn OLEDB, sau đó chọn Add để kết nối tới hệ quản trị SQL Server



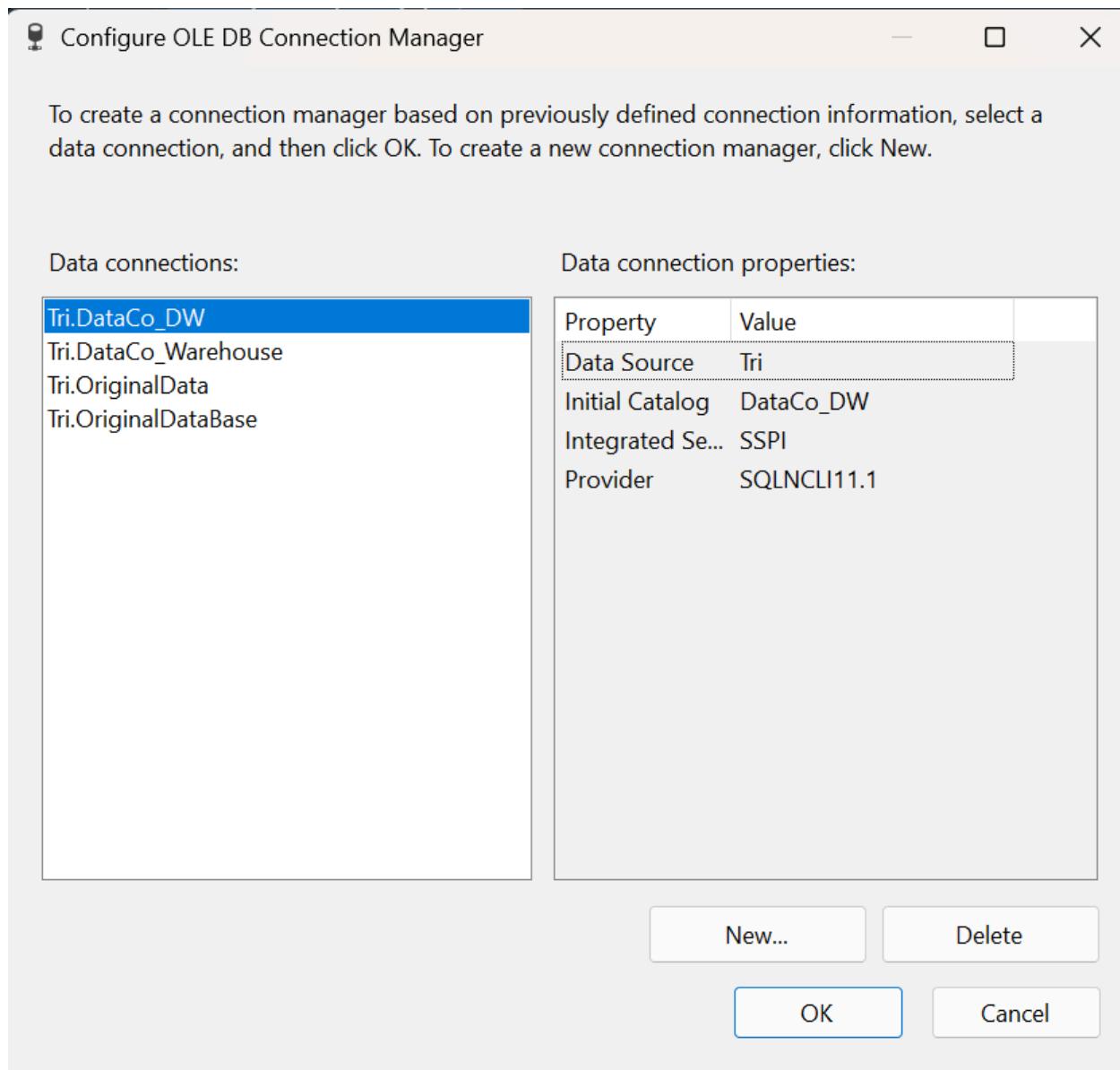
- Màn hình connection sẽ hiện lên, ta bấm New để tiến hành thêm 1 connect mới.



- Ở mục Server name chọn tên server cục bộ ở laptop. Chọn database OriginalDataBase sau đó chọn OK.



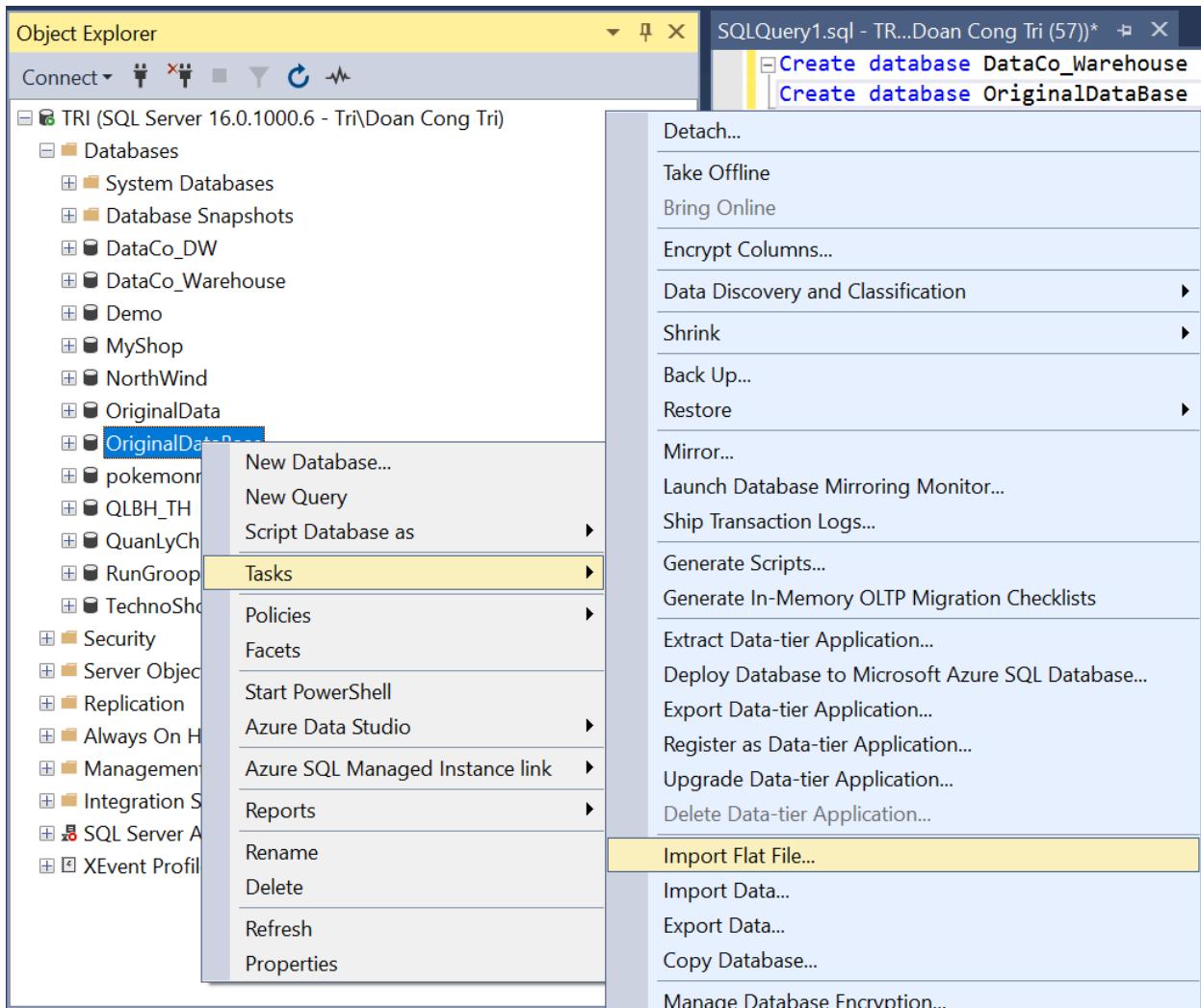
- Làm tương tự với database DataCo_Warehouse



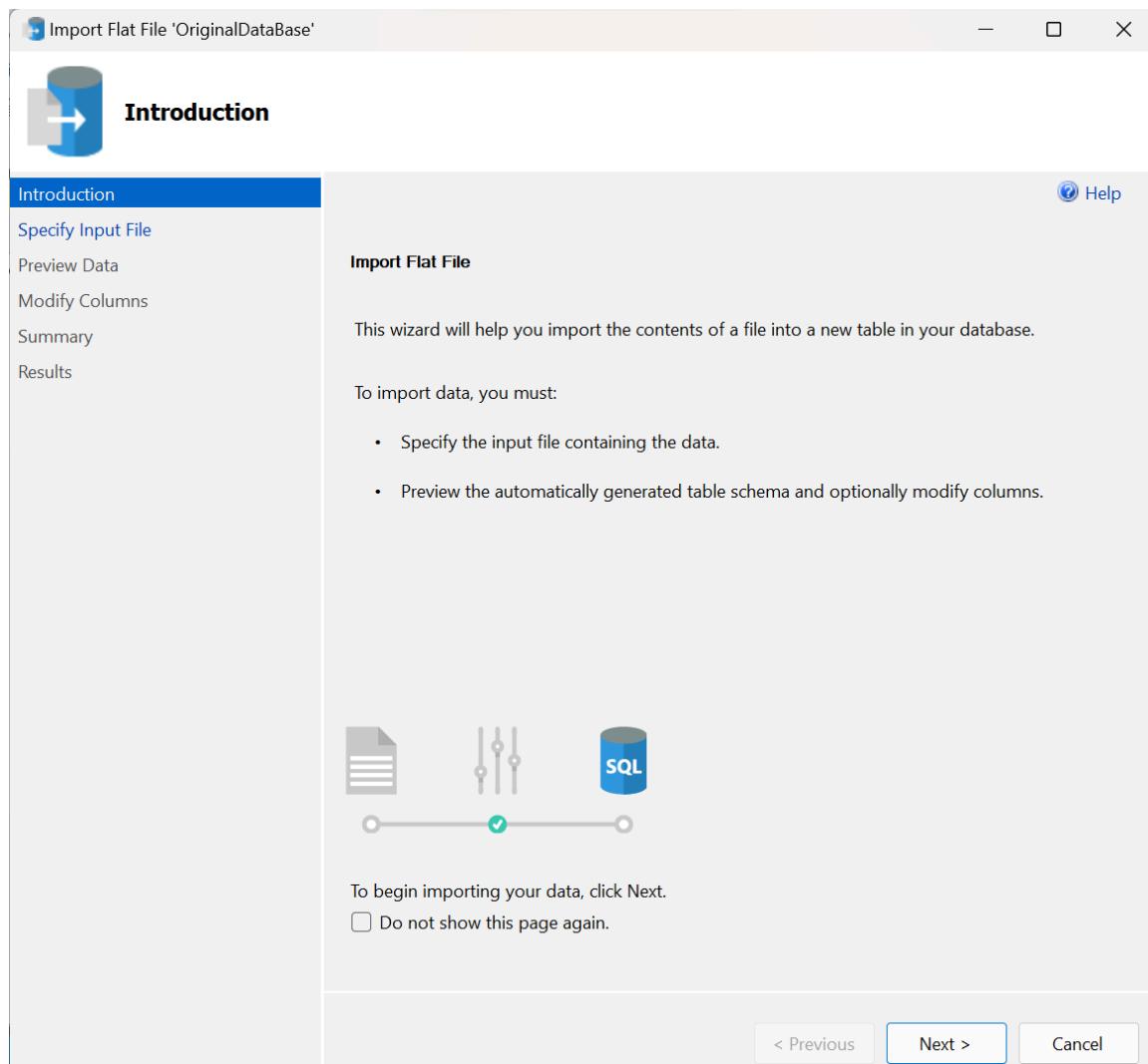
2.2.2. Chuẩn bị dữ liệu gốc và import vào dữ liệu gốc

Dữ liệu được sử dụng trong đồ án là một sheet trong file csv. Ta sẽ import và đồng thời tạo một bảng mới chứa dữ liệu gốc có tên là “DataCoSupplyChainDataset.csv”

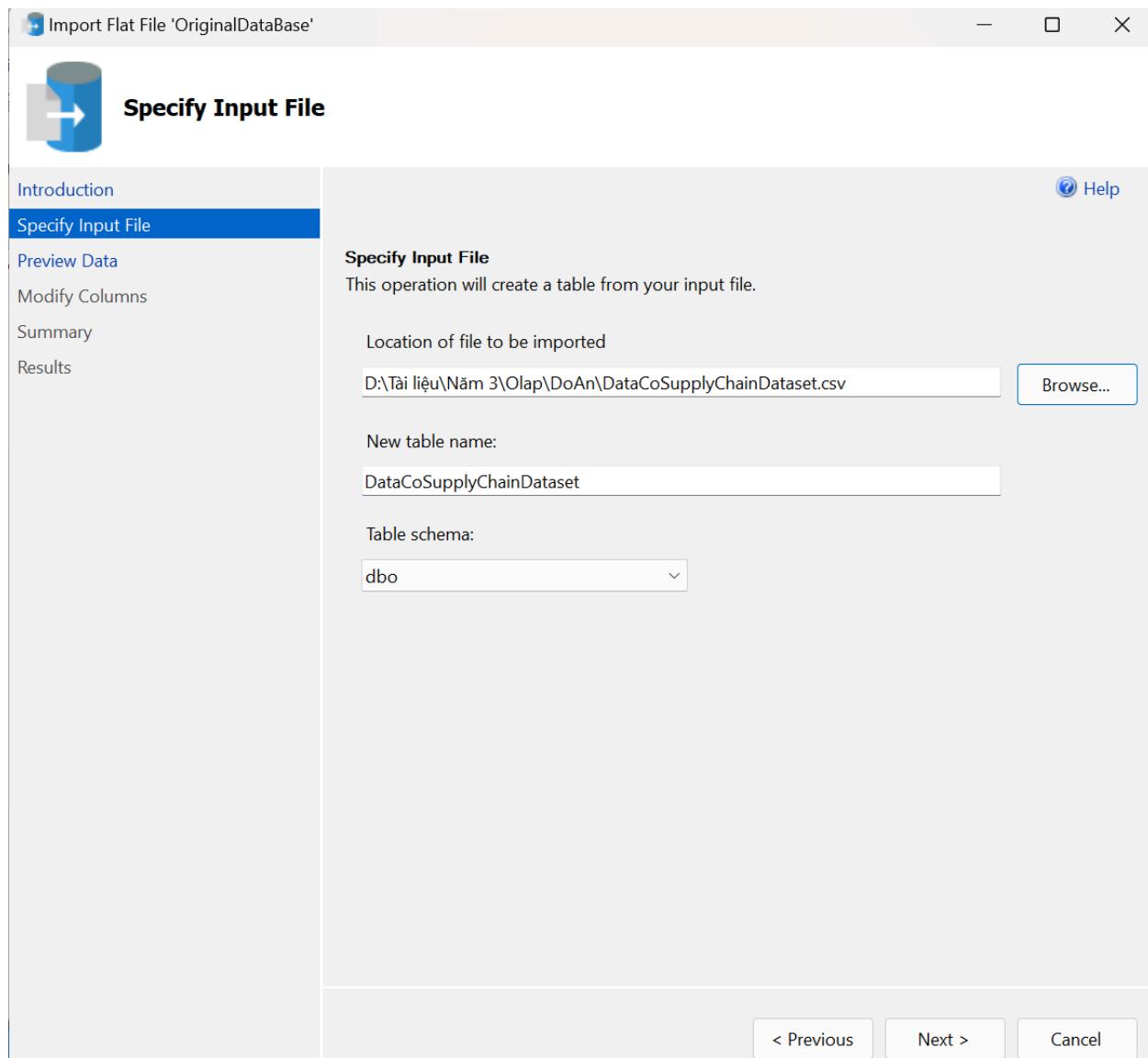
- Đầu tiên, ta nhập phải vào database OriginalData chọn Task => Import Flat File



- Hộp thoại hiện lên và bấm “Next”



- Tại mục tiếp theo chọn “Browse” dẫn đường dẫn đến file csv và chọn “Next”



- Thực hiện xem review lại bộ dữ liệu và chọn “Next”

Preview Data

This operation analyzed the input file structure to generate the preview below for up to the first 50 rows.

order_id	type	order_status	order_date	sales	market	discount_ra	product
77202	DEBIT	COMPLETE	1/31/201...	327.75	Pacific Asia	0.039999...	1
75939	TRANSFER	PENDING	1/13/201...	327.75	Pacific Asia	0.050000...	1
75938	CASH	CLOSED	1/13/201...	327.75	Pacific Asia	0.059999...	1
75937	DEBIT	COMPLETE	1/13/201...	327.75	Pacific Asia	0.07	1
75936	PAYMENT	PENDING...	1/13/201...	327.75	Pacific Asia	0.090000...	1
75935	TRANSFER	CANCELED	1/13/201...	327.75	Pacific Asia	0.100000...	1
75934	DEBIT	COMPLETE	1/13/201...	327.75	Pacific Asia	0.119999...	1
75933	TRANSFER	PROCESS...	1/13/201...	327.75	Pacific Asia	0.129999...	1
75932	CASH	CLOSED	1/13/201...	327.75	Pacific Asia	0.150000...	1
75931	CASH	CLOSED	1/13/201...	327.75	Pacific Asia	0.159999...	1
75930	TRANSFER	SUSPECT...	1/13/201...	327.75	Pacific Asia	0.170000...	1
75929	TRANSFER	PENDING	1/13/201...	327.75	Pacific Asia	0.180000...	1
75928	TRANSFER	PENDING	1/13/201...	327.75	Pacific Asia	0.200000...	1
75927	DEBIT	COMPLETE	1/13/201...	327.75	Pacific Asia	0.25	1
75926	TRANSFER	PROCESS...	1/13/201...	327.75	Pacific Asia	0.0	1
75925	DEBIT	COMPLETE	1/13/201...	327.75	Pacific Asia	0.01	1
75924	PAYMENT	PENDING...	1/13/201...	327.75	Pacific Asia	0.02	1

Use Rich Data Type Detection - may provide a closer type fit. However, cells with anomalous values may be dropped.

< Previous Next > Cancel

- Ở mục tiếp theo thực hiện chỉnh lại kiểu dữ liệu cho cột, đồng thời cho phép insert các cột được phép null và chọn “Next”

Modify Columns

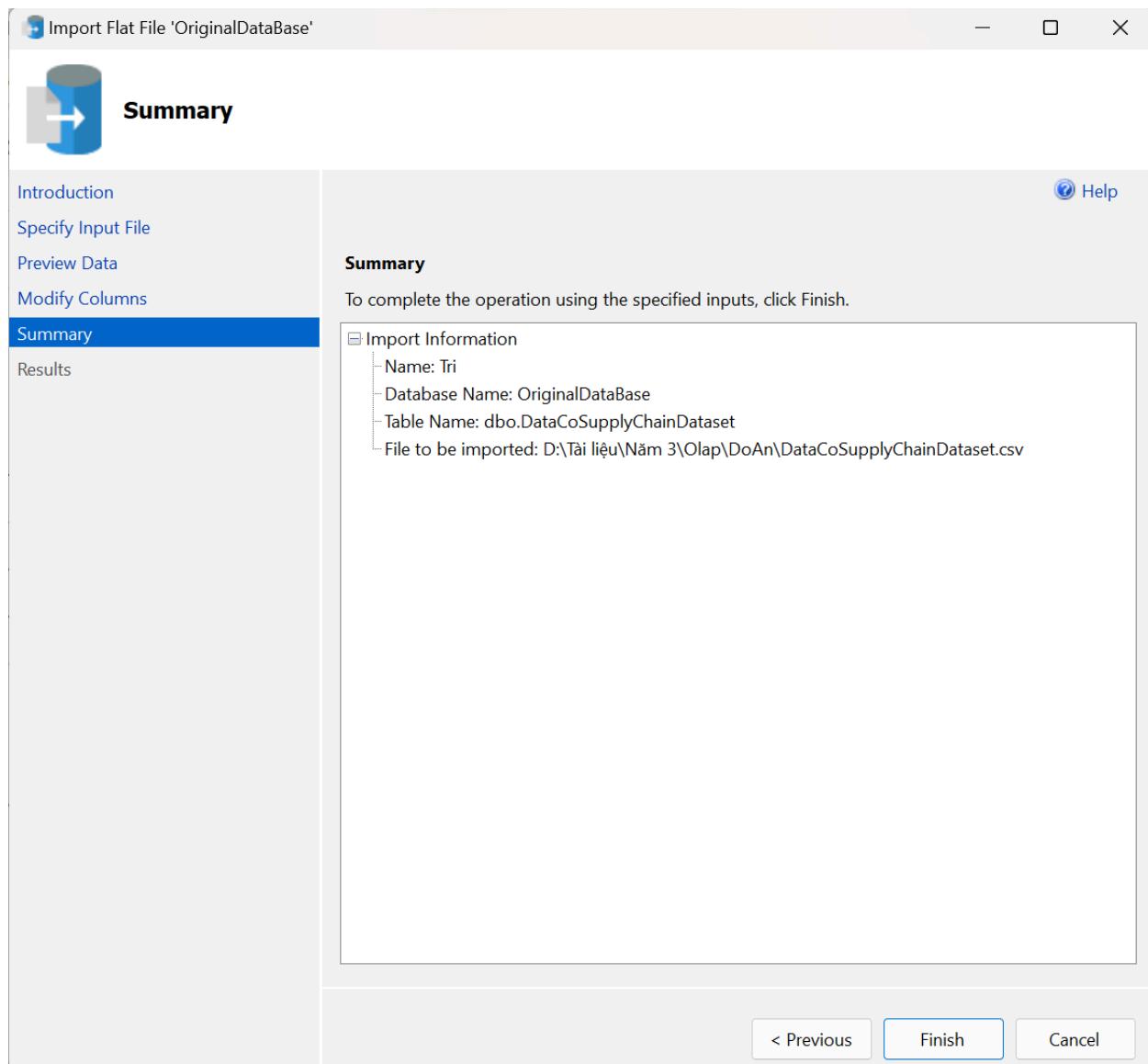
This operation generated the following table schema. Please verify if schema is accurate, and if not, please make any changes.

Column Name	Data Type	Primary Key	Allow Nulls
order_date	datetime2	<input type="checkbox"/>	<input checked="" type="checkbox"/>
sales	float	<input type="checkbox"/>	<input checked="" type="checkbox"/>
market	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
discount_rate	float	<input type="checkbox"/>	<input checked="" type="checkbox"/>
product_quantity	int	<input type="checkbox"/>	<input checked="" type="checkbox"/>
category_id	int	<input type="checkbox"/>	<input checked="" type="checkbox"/>
category_name	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
day_shipping_real	int	<input type="checkbox"/>	<input checked="" type="checkbox"/>
day_shipment_scheduled	int	<input type="checkbox"/>	<input checked="" type="checkbox"/>
delivery_status	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
shipping_mode	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
customer_id	int	<input type="checkbox"/>	<input checked="" type="checkbox"/>
customer_country	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
customer_name	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
customer_segment	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
product_id	int	<input type="checkbox"/>	<input checked="" type="checkbox"/>
product_name	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
product_price	float	<input type="checkbox"/>	<input checked="" type="checkbox"/>
shipping_id	int	<input type="checkbox"/>	<input checked="" type="checkbox"/>

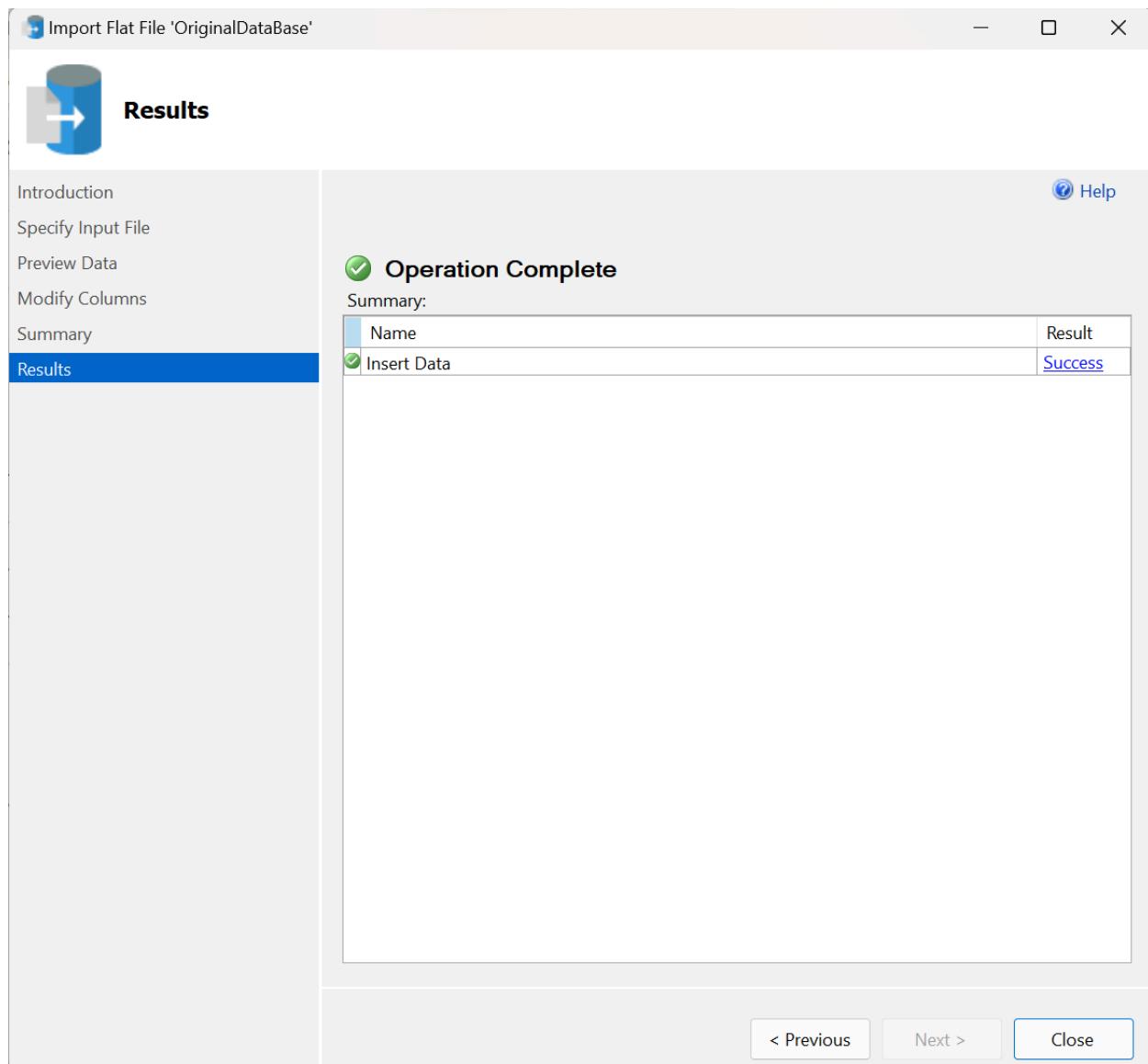
Row granularity of error reporting (performance impact with smaller ranges)

< Previous Next > Cancel

- Ở hộp thoại tiếp theo chọn “Finish”



- Thông báo import dữ liệu thành công



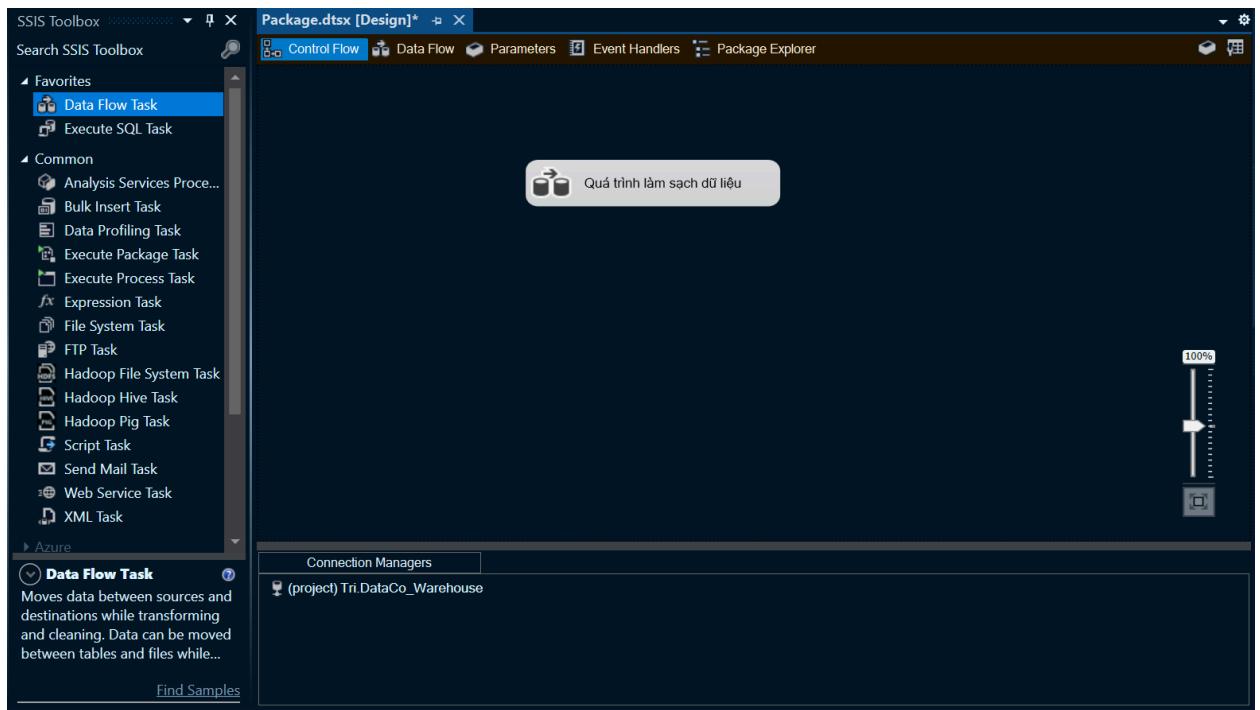
- Dữ liệu gốc đã được import vào SQL Server

	order_id	type	order_status	order_date	sales	market	discount_rate	product_quantity	category_id	cate
1	77202	DEBIT	COMPLETE	2018-01-31 22:56:00.0000000	327.75	Pacific Asia	0.039999991059303	1	73	Spc
2	75939	TRANSFER	PENDING	2018-01-13 12:27:00.0000000	327.75	Pacific Asia	0.0500000007450581	1	73	Spc
3	75938	CASH	CLOSED	2018-01-13 12:06:00.0000000	327.75	Pacific Asia	0.059999986588955	1	73	Spc
4	75937	DEBIT	COMPLETE	2018-01-13 11:45:00.0000000	327.75	Pacific Asia	0.070000002980232	1	73	Spc
5	75936	PAYMENT	PENDING_PAYMENT	2018-01-13 11:24:00.0000000	327.75	Pacific Asia	0.0900000035762787	1	73	Spc
6	75935	TRANSFER	CANCELED	2018-01-13 11:03:00.0000000	327.75	Pacific Asia	0.100000001490116	1	73	Spc
7	75934	DEBIT	COMPLETE	2018-01-13 10:42:00.0000000	327.75	Pacific Asia	0.119999997317791	1	73	Spc
8	75933	TRANSFER	PROCESSING	2018-01-13 10:21:00.0000000	327.75	Pacific Asia	0.12999995231628	1	73	Spc
9	75932	CASH	CLOSED	2018-01-13 10:00:00.0000000	327.75	Pacific Asia	0.150000005960464	1	73	Spc
10	75931	CASH	CLOSED	2018-01-13 09:39:00.0000000	327.75	Pacific Asia	0.15999996423721	1	73	Spc
11	75930	TRANSFER	SUSPECTED_FRAUD	2018-01-13 09:18:00.0000000	327.75	Pacific Asia	0.17000001788139	1	73	Spc
12	75929	TRANSFER	PENDING	2018-01-13 08:57:00.0000000	327.75	Pacific Asia	0.180000007152557	1	73	Spc
13	75928	TRANSFER	PENDING	2018-01-13 08:36:00.0000000	327.75	Pacific Asia	0.200000002980232	1	73	Spc
14	75927	DEBIT	COMPLETE	2018-01-13 08:15:00.0000000	327.75	Pacific Asia	0.25	1	73	Spc
15	75926	TRANSFER	PROCESSING	2018-01-13 07:54:00.0000000	327.75	Pacific Asia	0	1	73	Spc
16	75925	DEBIT	COMPLETE	2018-01-13 07:33:00.0000000	327.75	Pacific Asia	0.0099999977648...	1	73	Spc
17	75924	PAYMENT	PENDING_PAYMENT	2018-01-13 07:12:00.0000000	327.75	Pacific Asia	0.019999995529652	1	73	Spc
18	75923	CASH	CLOSED	2018-01-13 06:51:00.0000000	327.75	Pacific Asia	0.029999993294477	1	73	Spc
19	75922	DEBIT	COMPLETE	2018-01-13 06:30:00.0000000	327.75	Pacific Asia	0.039999991059303	1	73	Spc
20	75921	PAYMENT	PENDING_PAYMENT	2018-01-13 06:09:00.0000000	327.75	Pacific Asia	0.0500000007450581	1	73	Spc
21	75920	TRANSFER	PROCESSING	2018-01-13 05:48:00.0000000	327.75	Pacific Asia	0.0599999986588955	1	73	Spc
22	75919	TRANSFER	PENDING	2018-01-13 05:27:00.0000000	327.75	Pacific Asia	0.070000002980232	1	73	Spc

2.3. QUÁ TRÌNH LÀM SẠCH DỮ LIỆU

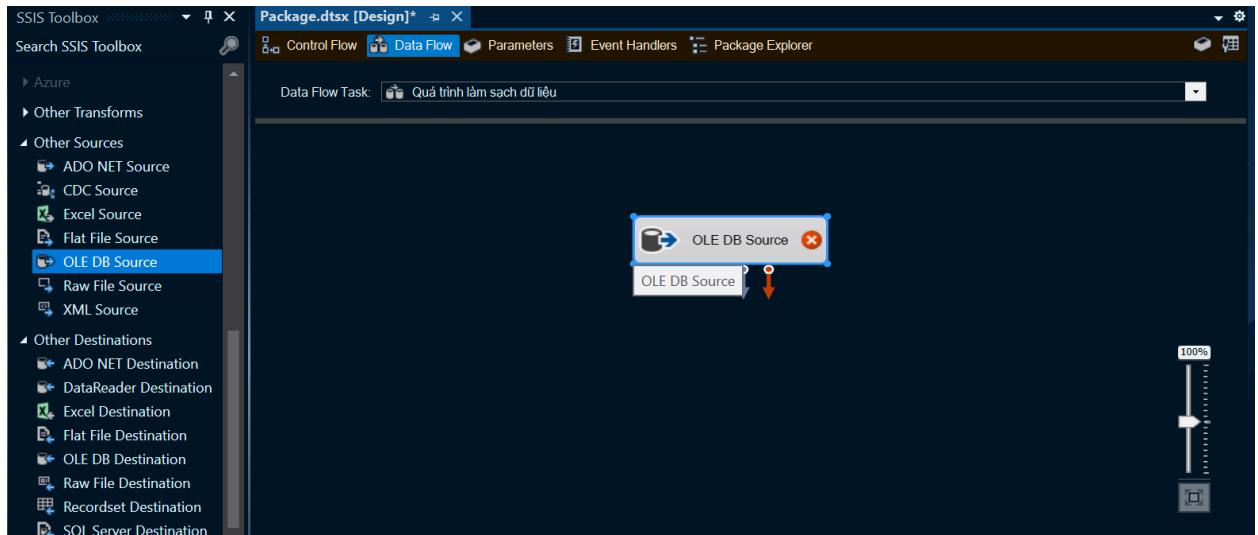
- Tạo Data Flow Task

Kéo thả công cụ Data Flow Task ở SSIS Toolbox sang vùng làm việc của Control Flow và đổi tên thành “Quá trình làm sạch dữ liệu”

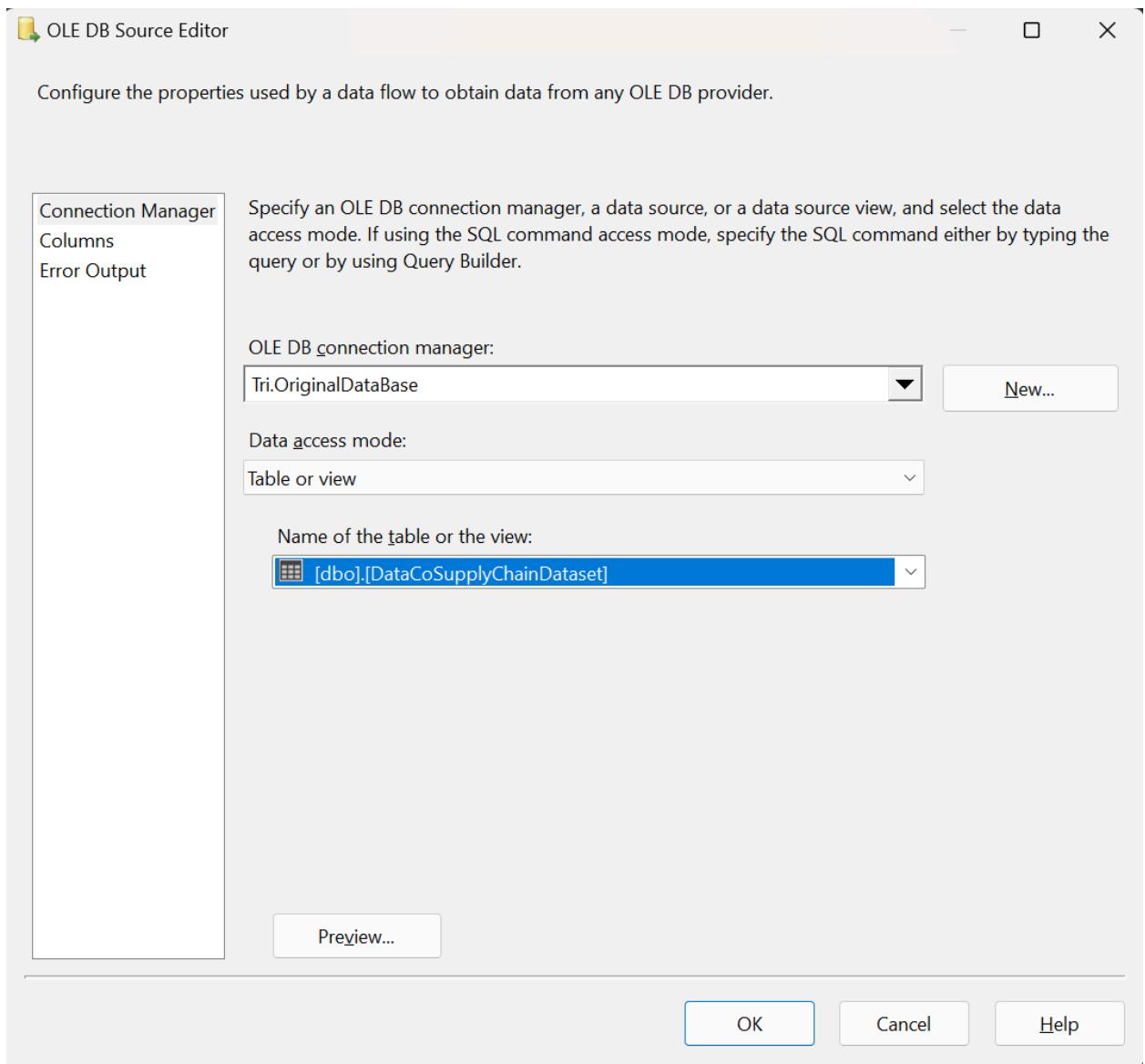


- Thực hiện cấu hình “Quá trình làm sạch dữ liệu”

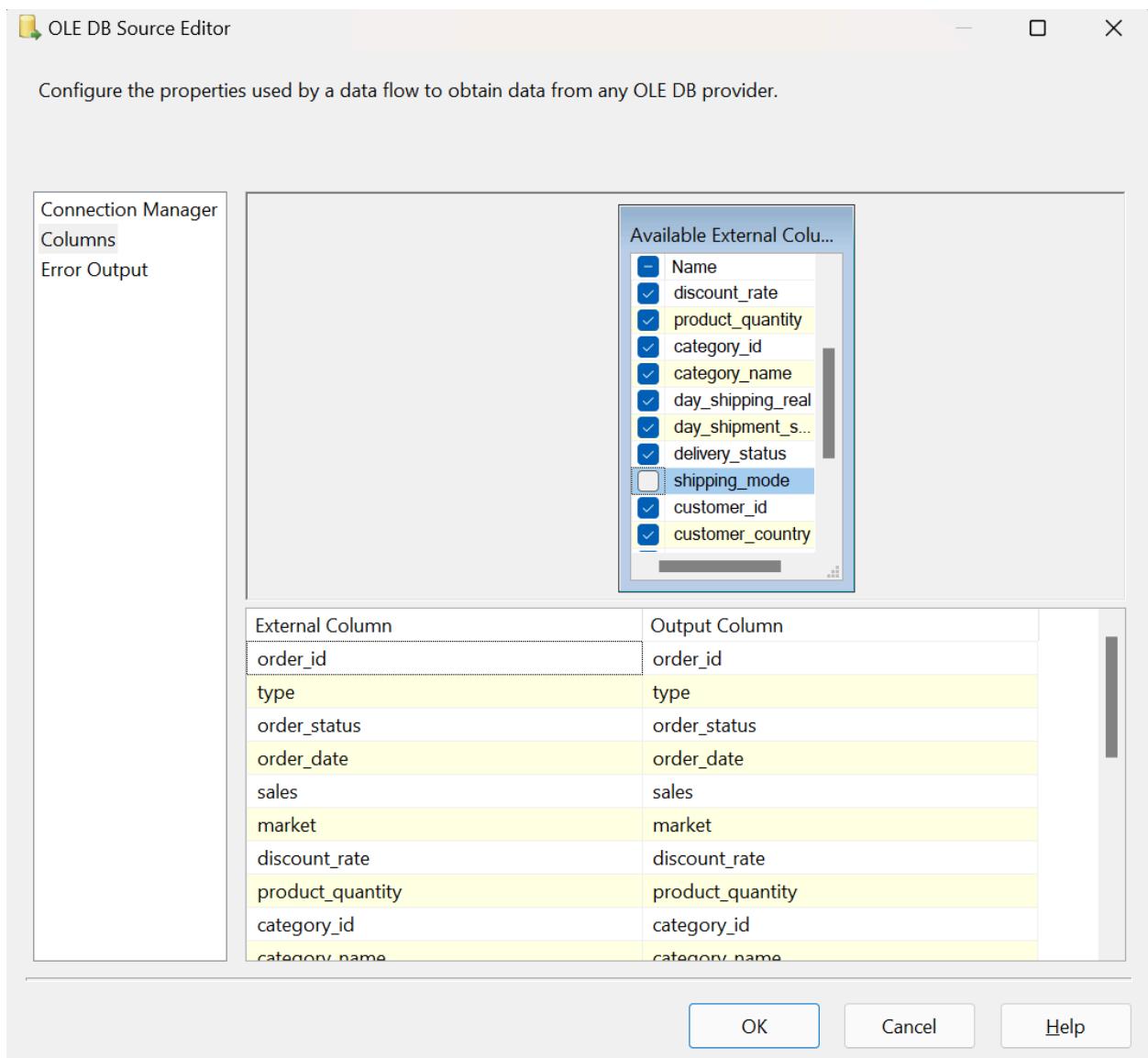
Tạo và cấu hình “OLE DB Source” để import dữ liệu từ database: Kéo thả OLE DB Source từ SSIS Toolbox



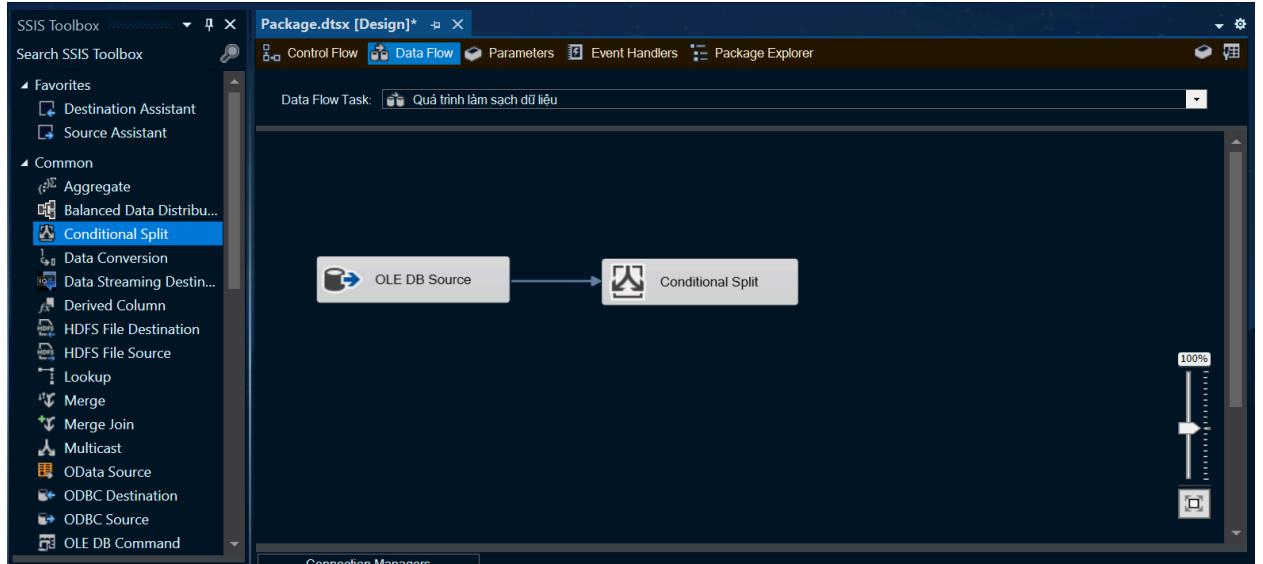
- Edit OLE DB Source, thực hiện nhấn phải chuột và chọn Edit. Chọn new và Connect đến bảng **DataCoSupplyChainDataset** trong **OriginalDatabase**



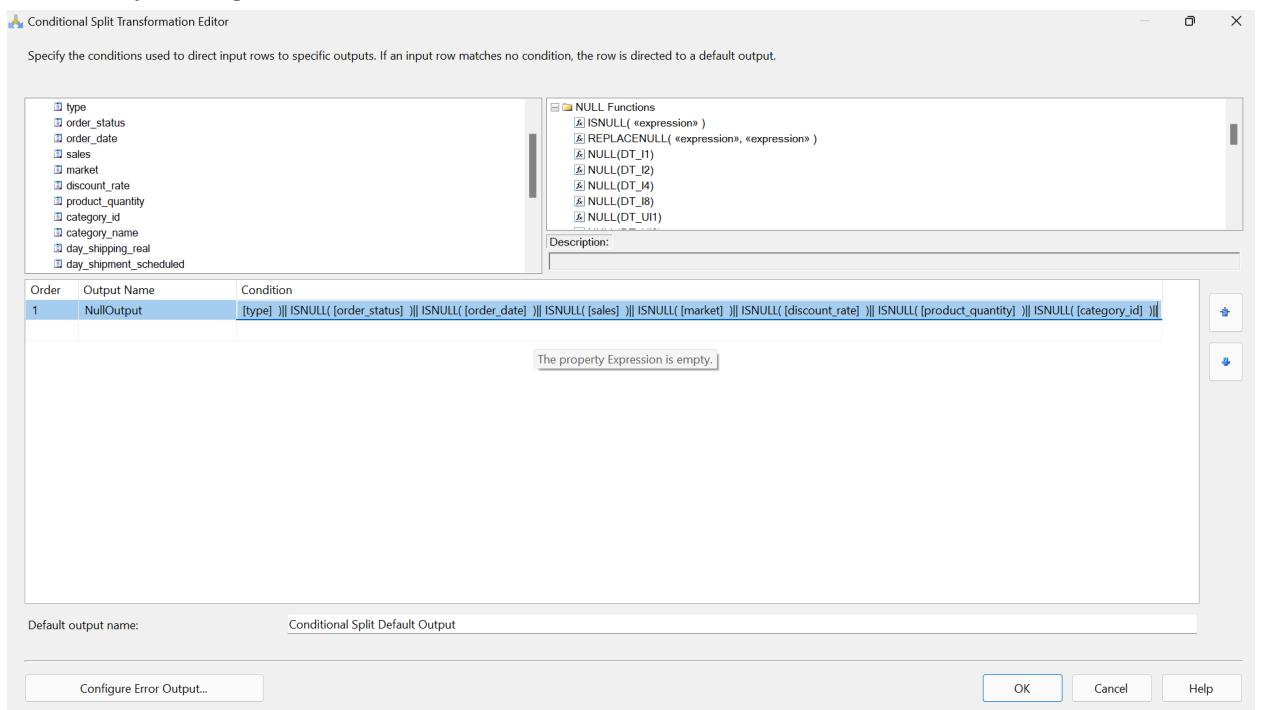
- Tiếp theo, chọn qua tab Columns và thực hiện bỏ những thuộc tính không sử dụng và chọn Ok



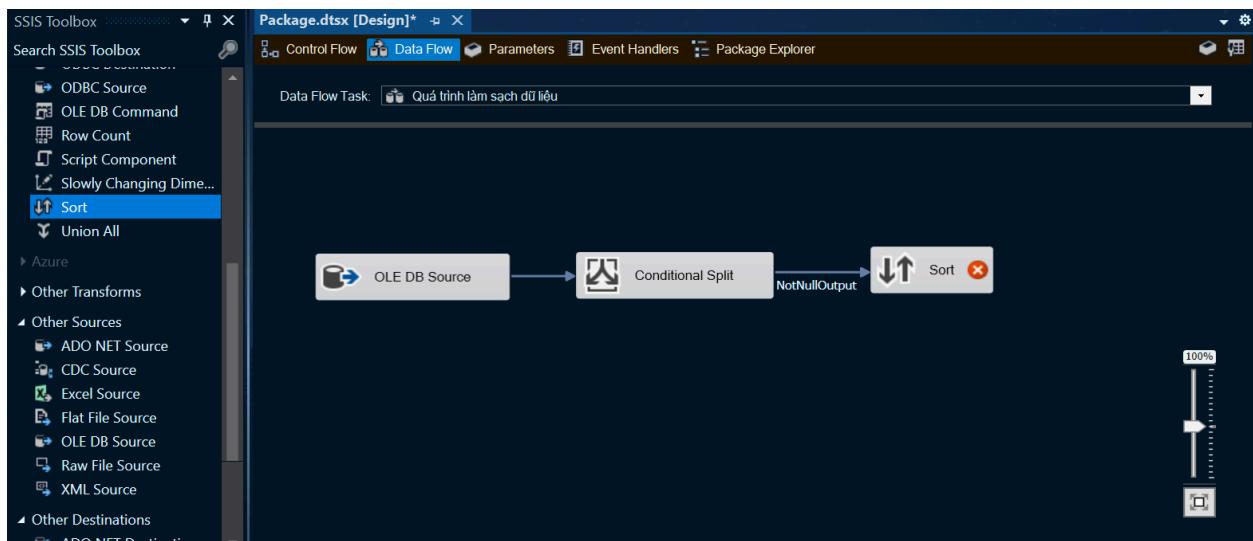
- Kéo thả “Condition Split” từ SSIS Toolbox, nhận đầu vào là OLE DB Source, tiến hành edit



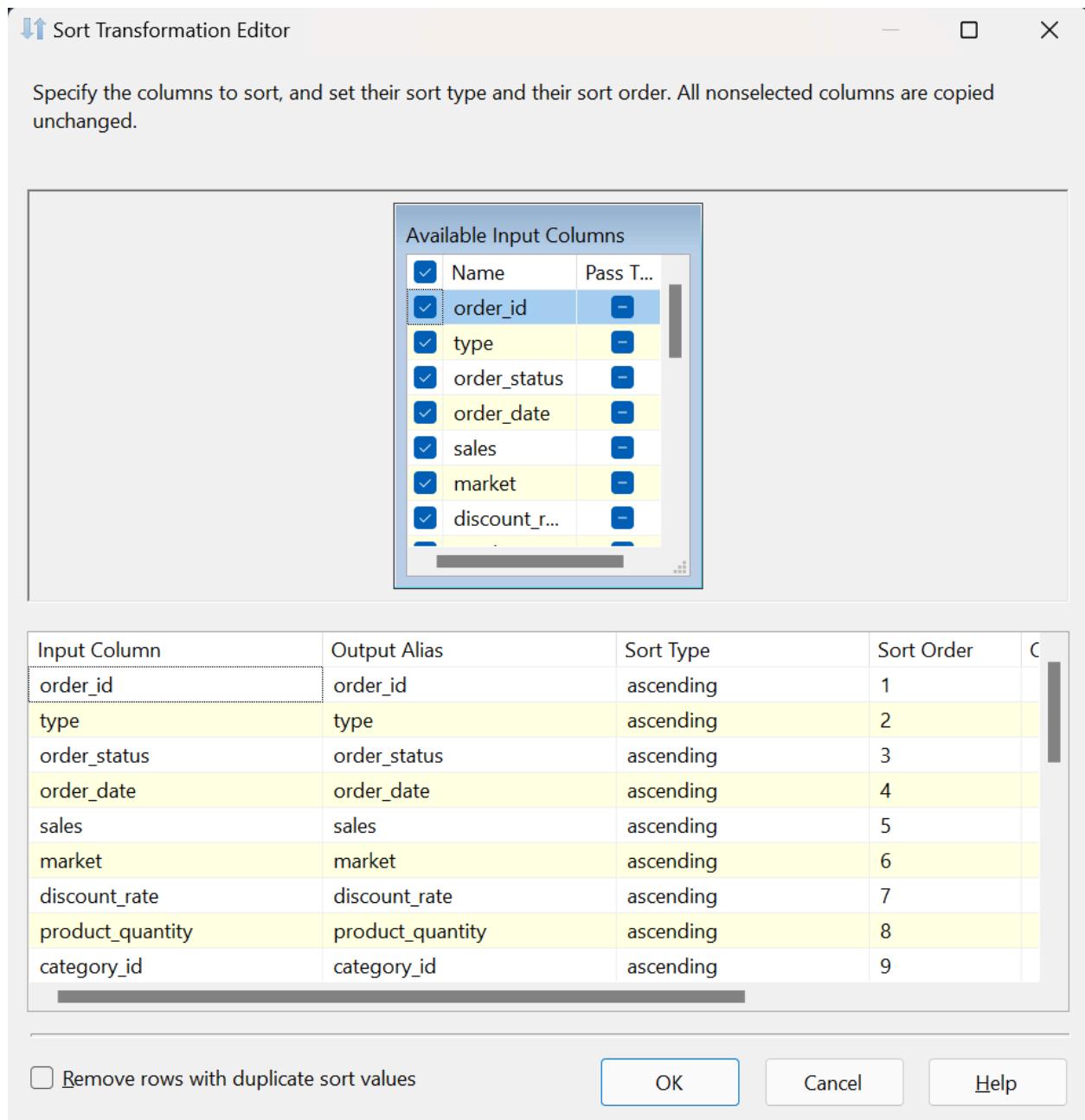
- Thực hiện edit “Condition Split”, thêm điều kiện kiểm tra các dòng các dữ liệu có bị null hay không



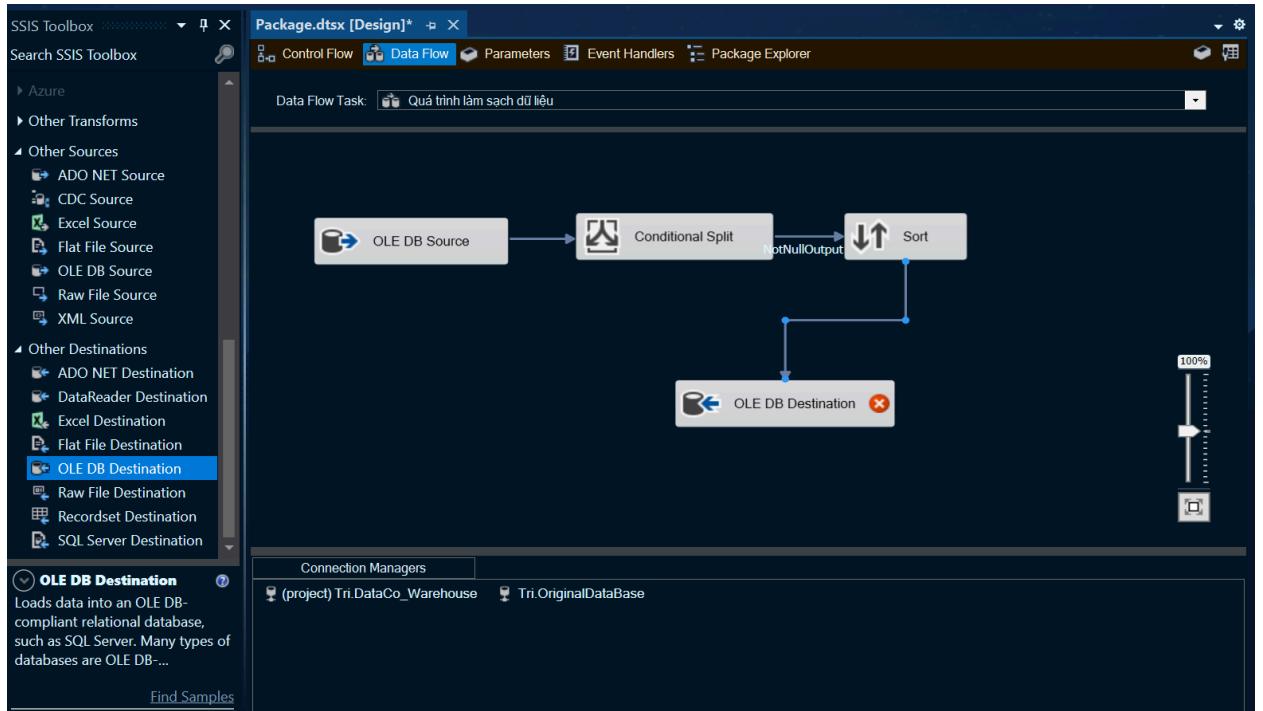
- Kéo thả “Sort” từ SSIS Toolbox, công cụ này chọn đầu vào từ “Condition Split” với điều kiện là NotNullOutput, tiến hành Edit.



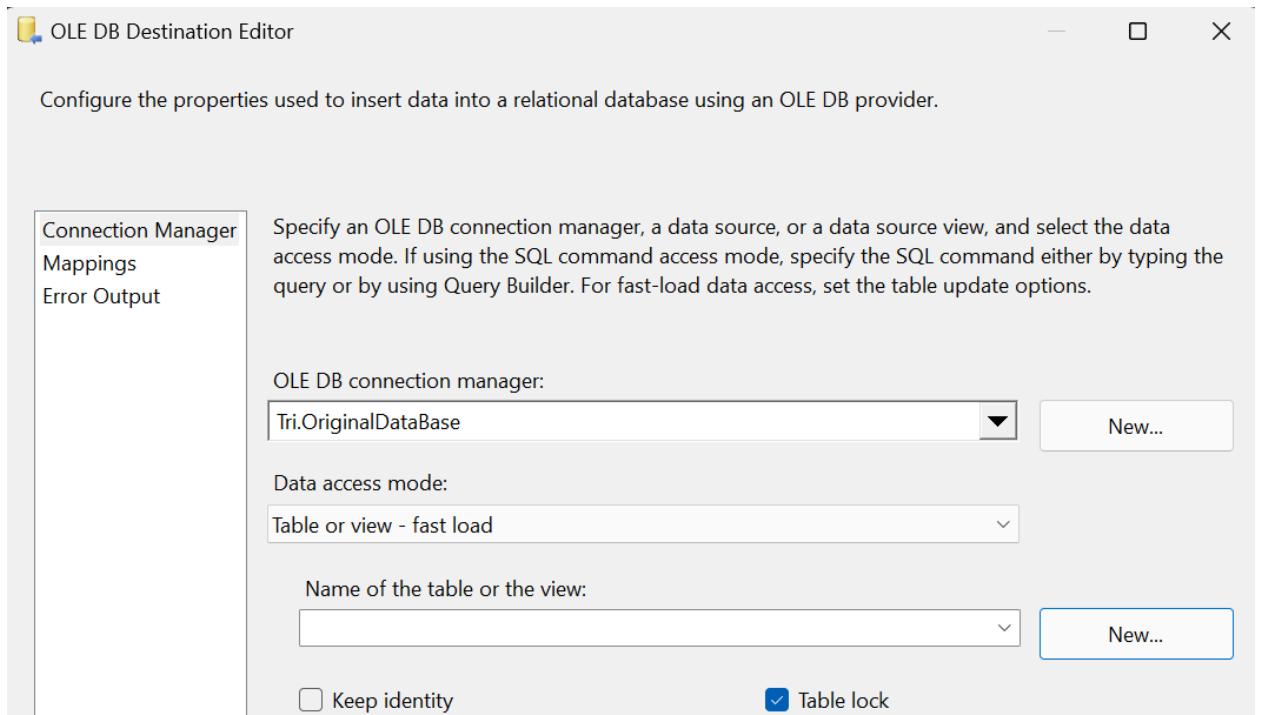
- Tiếp theo nháy đúp chuột vào “Sort” để tiến hành Edit, chọn tích vào các thuộc tính cần sắp xếp theo thứ tự tăng dần.

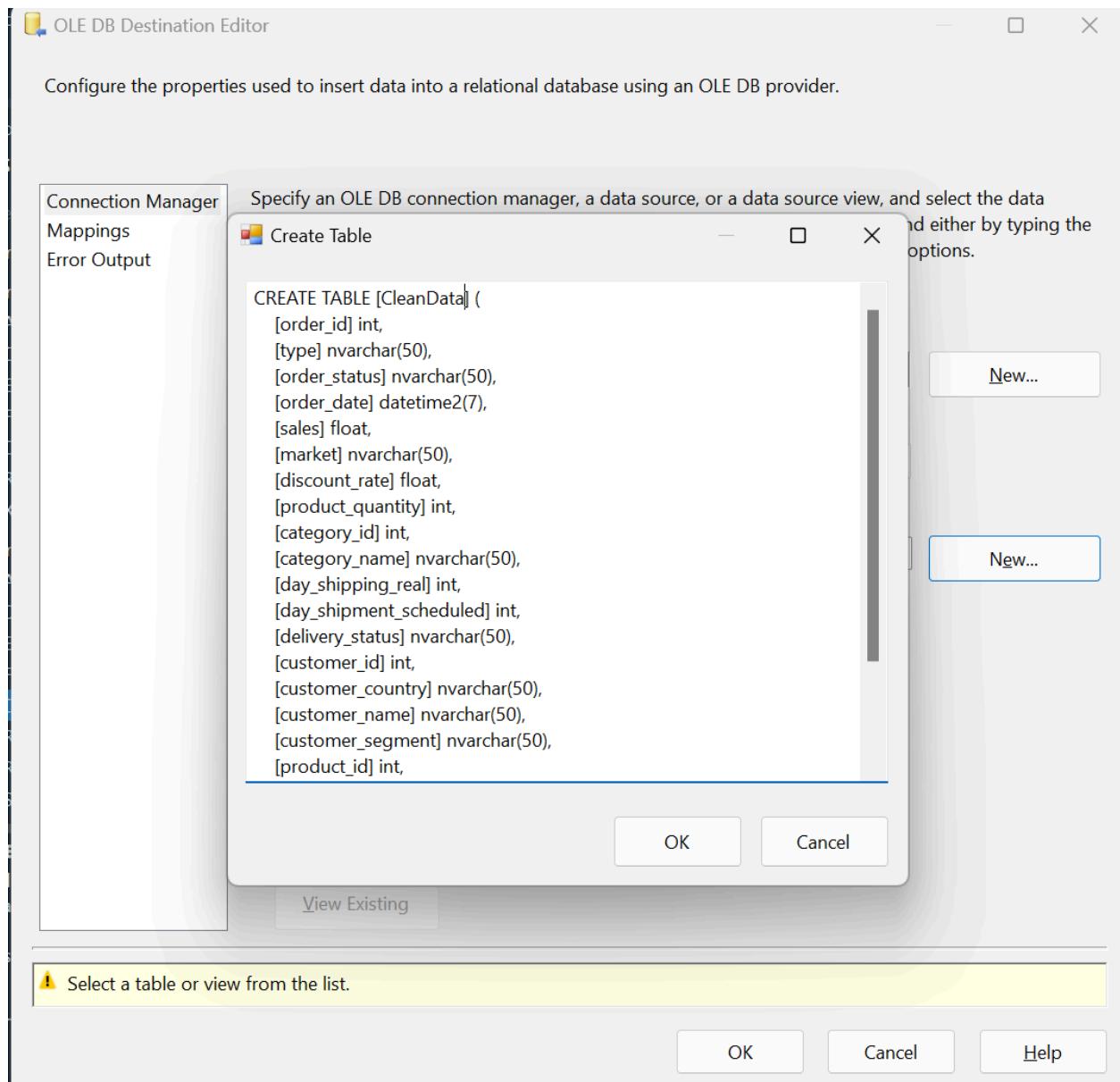


- Tạo và cấu hình OLE DB Destination, kéo thả OLE DB Destination từ Toolbox, đầu vào là “Sort”

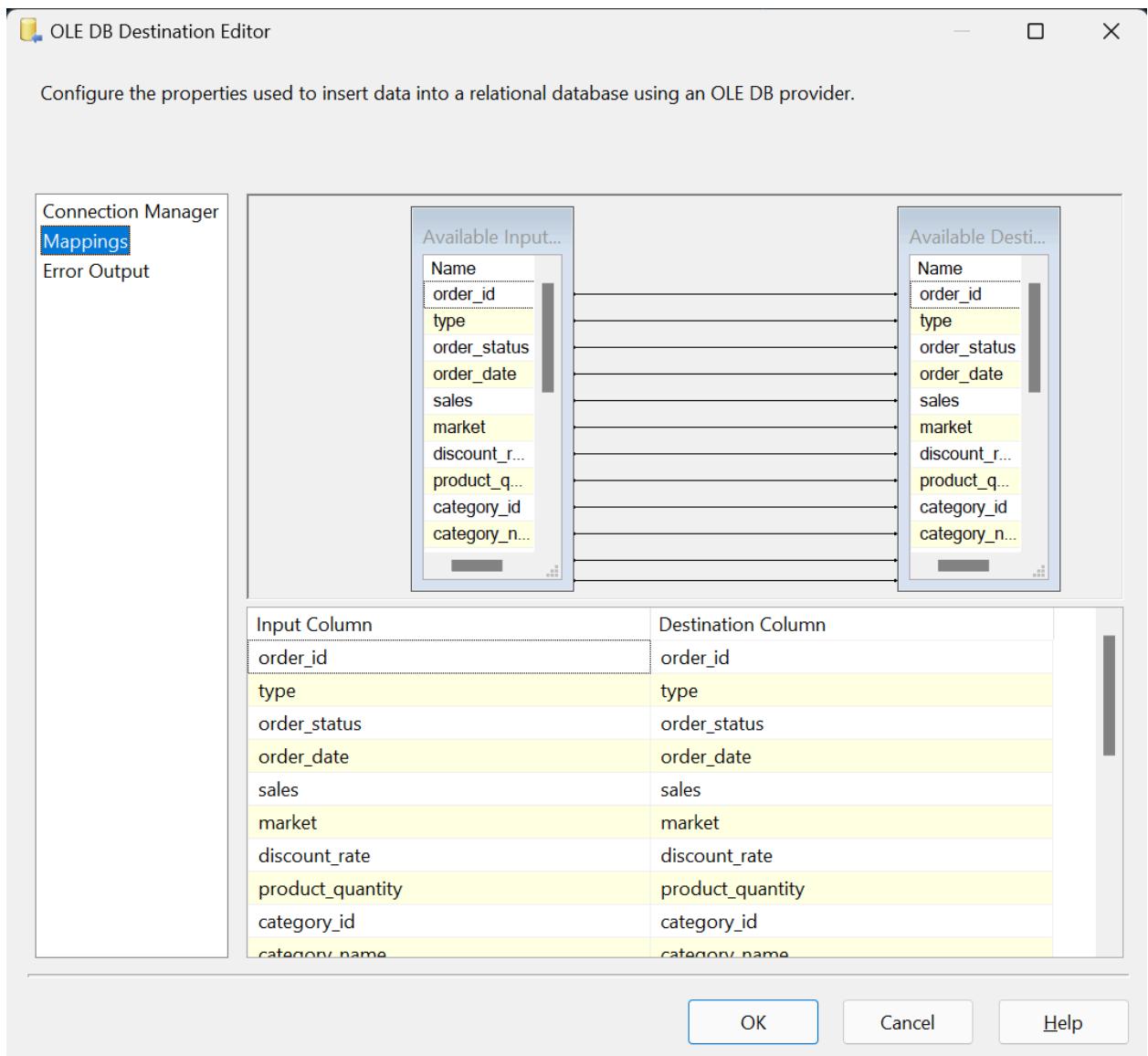


- Tiến hành edit OLE DB Destination, thực hiện kết nối tới database OriginalDatabase và tạo bảng CleanData

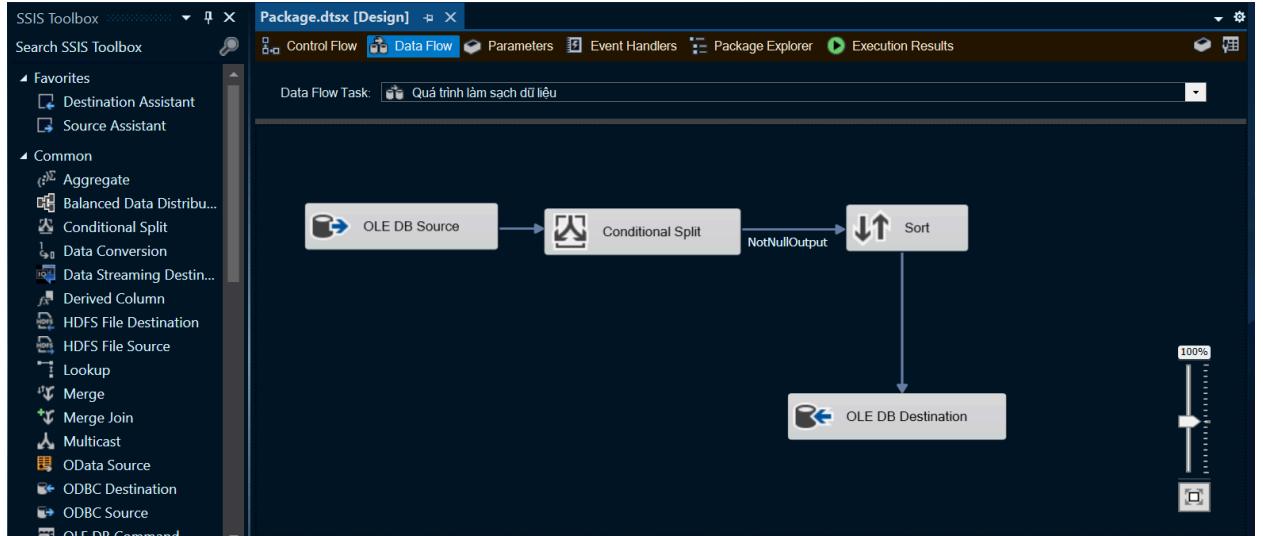




- Sau đó chuyển qua tab Mapping, thực hiện map các thuộc tính cần thiết và chọn OK



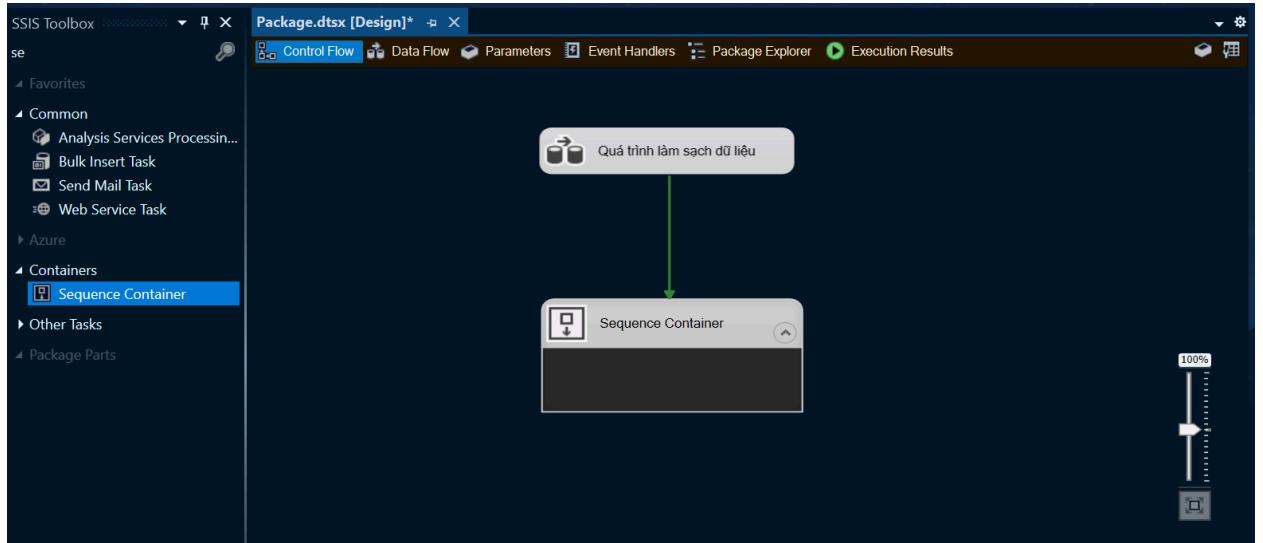
- Data Flow quá trình làm sạch dữ liệu hoàn thành



2.3. TẠO BẢNG DIMENSION

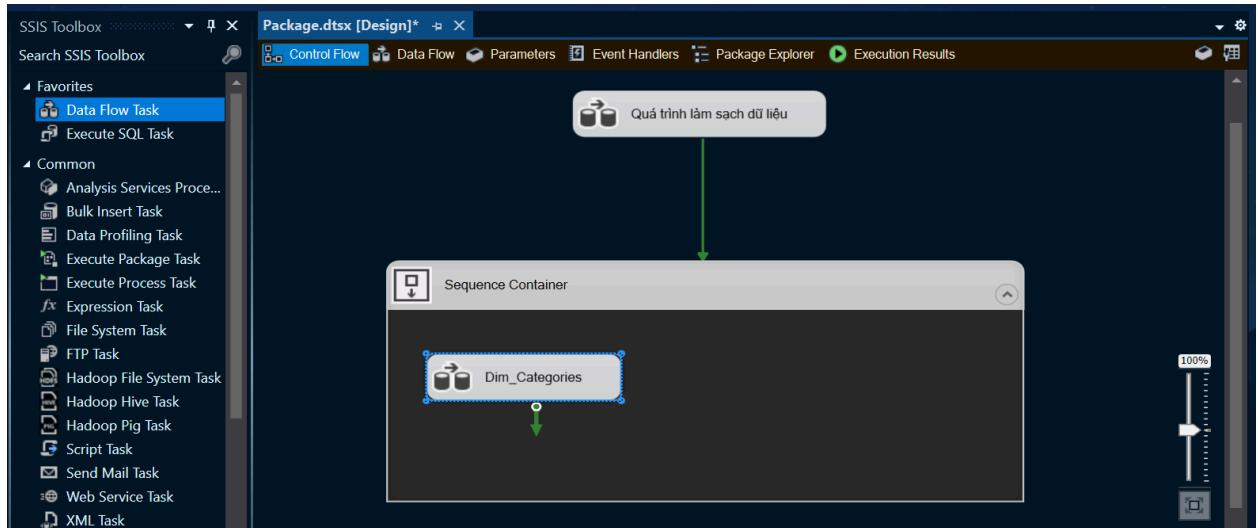
2.3.1. Tạo Sequence Container

- Kéo thả công cụ Sequence Container từ SSIS Toolbox, công cụ này giúp chúng ta có thể chạy đồng thời các Flow Task

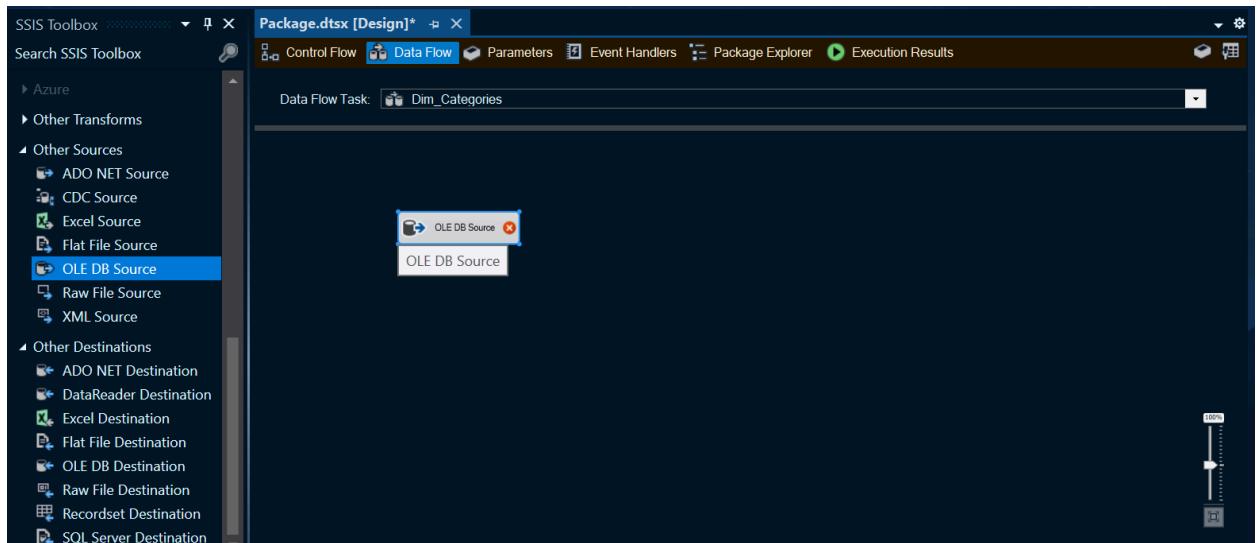


2.3.2. Tạo các bảng Dim_Categories

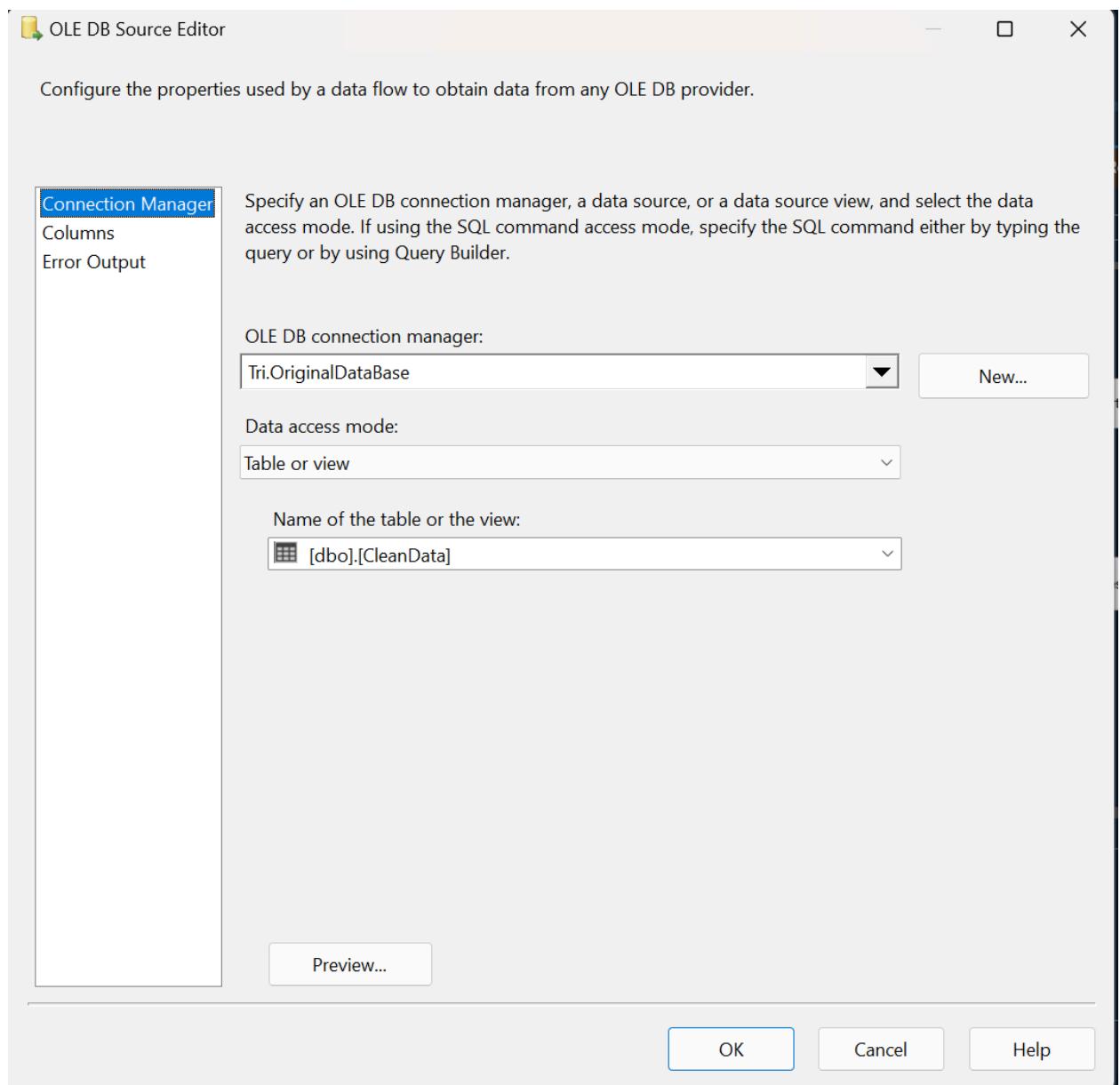
- Kéo thả công cụ Data Flow Task từ Toolbox. Sau đó tiến hành đổi tên thành “Dim_Categories”.



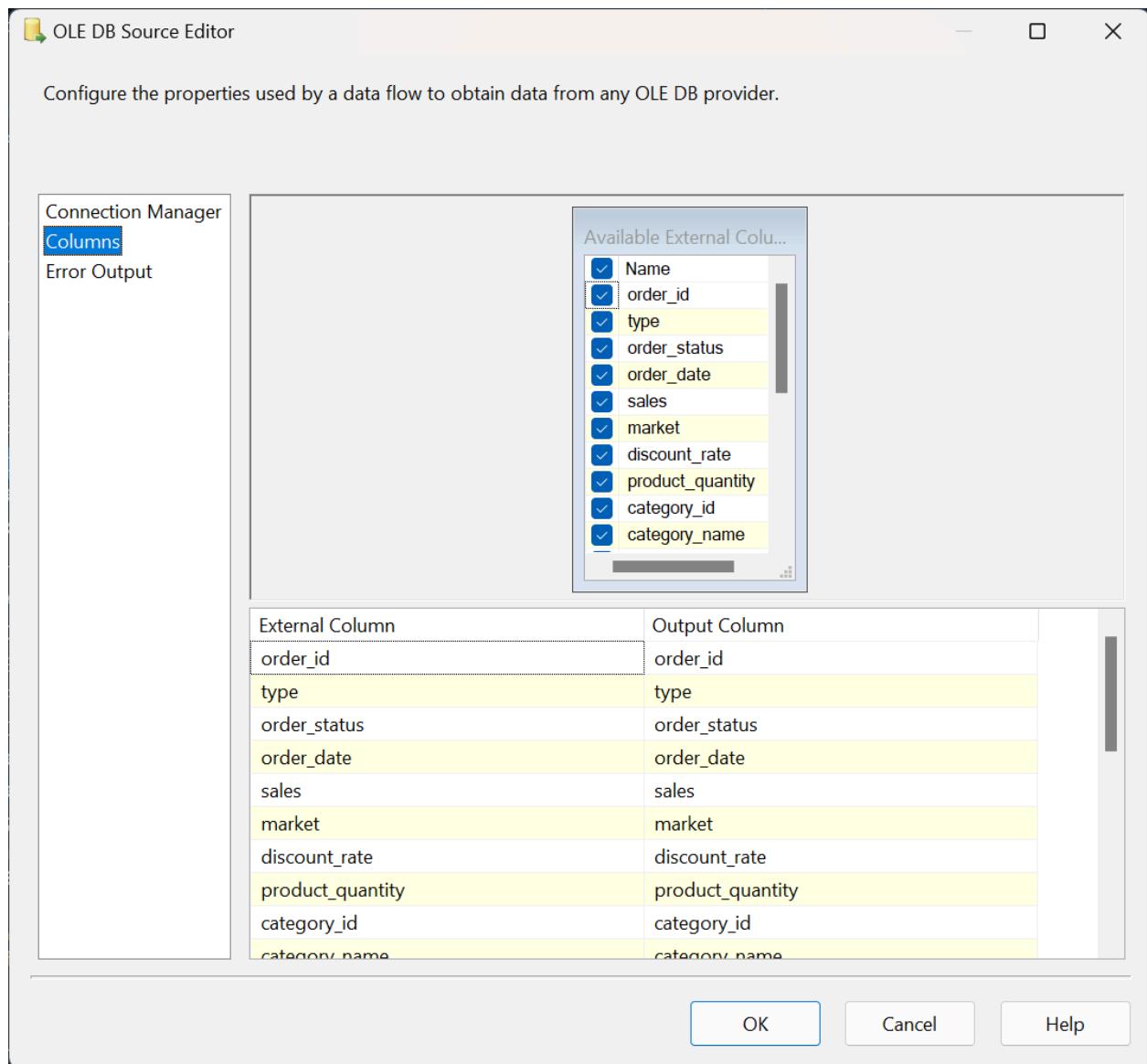
- Thực hiện edit Dim_Categories, kéo thả công cụ OLE DB Source từ Toolbox



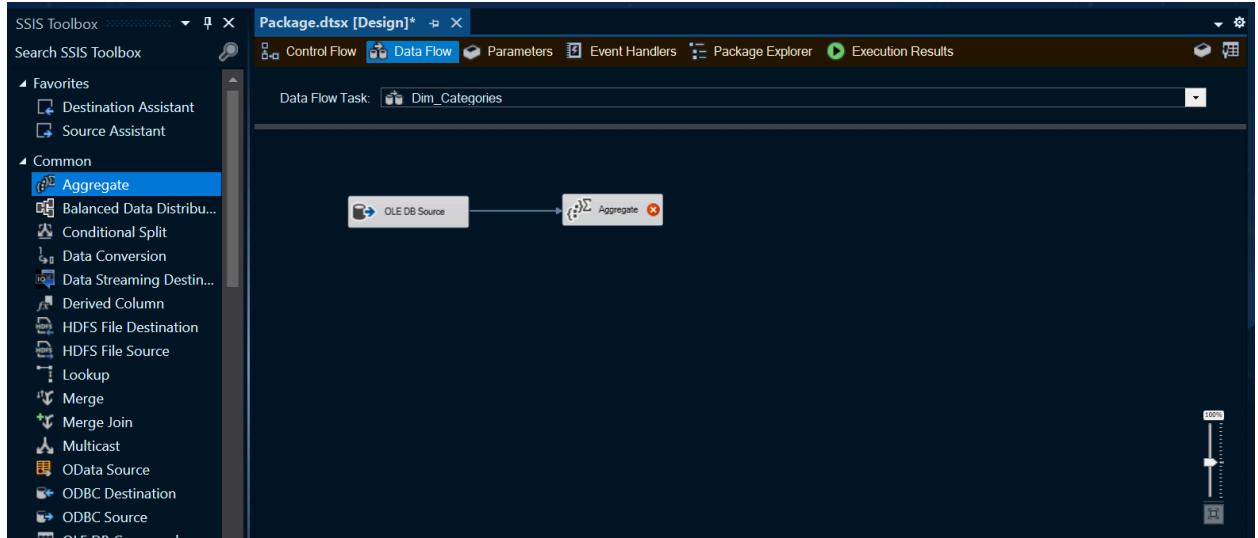
- Thực hiện edit OLE DB Source, kết nối tới bảng CleanData trong database OriginalDatabase.



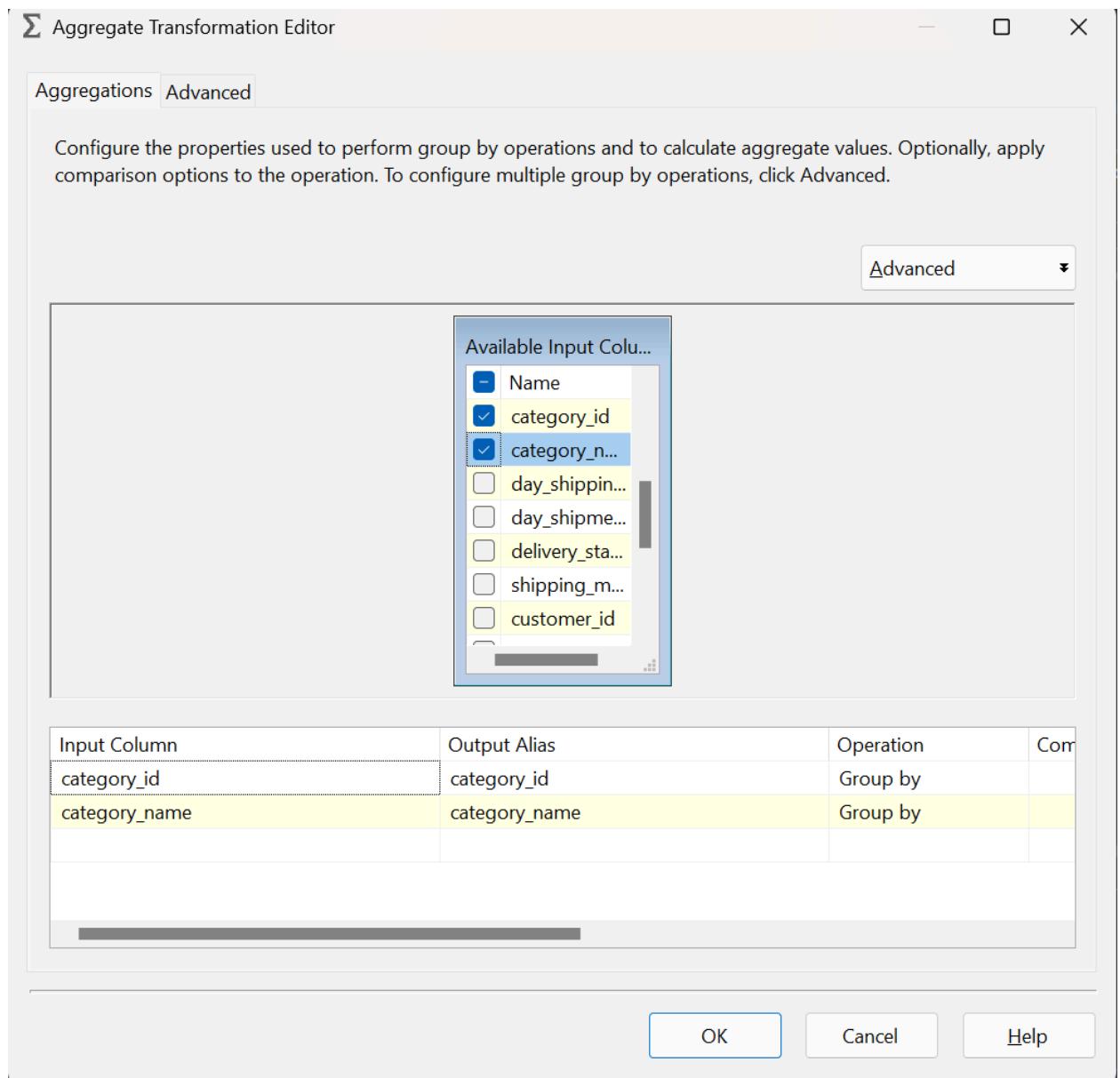
- Chọn tab Column và tích chọn các thuộc tính cần thiết và chọn OK



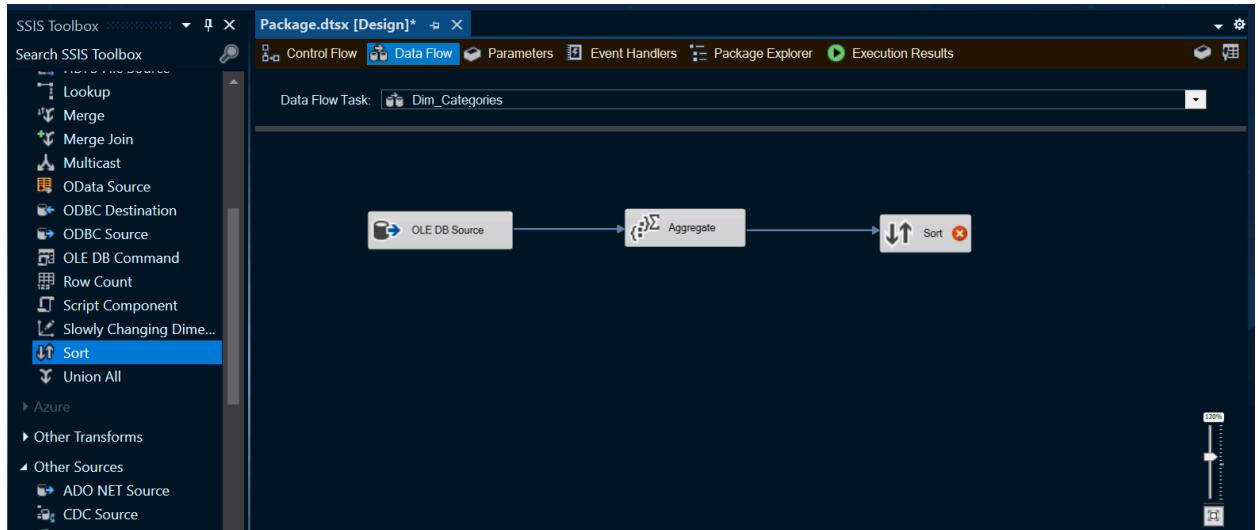
- Thực hiện kéo thả công cụ Aggregate từ SSIS Toolbox với đầu vào là OLE DB Source để chọn các thuộc tính cần thiết cho bảng Dim_Categories



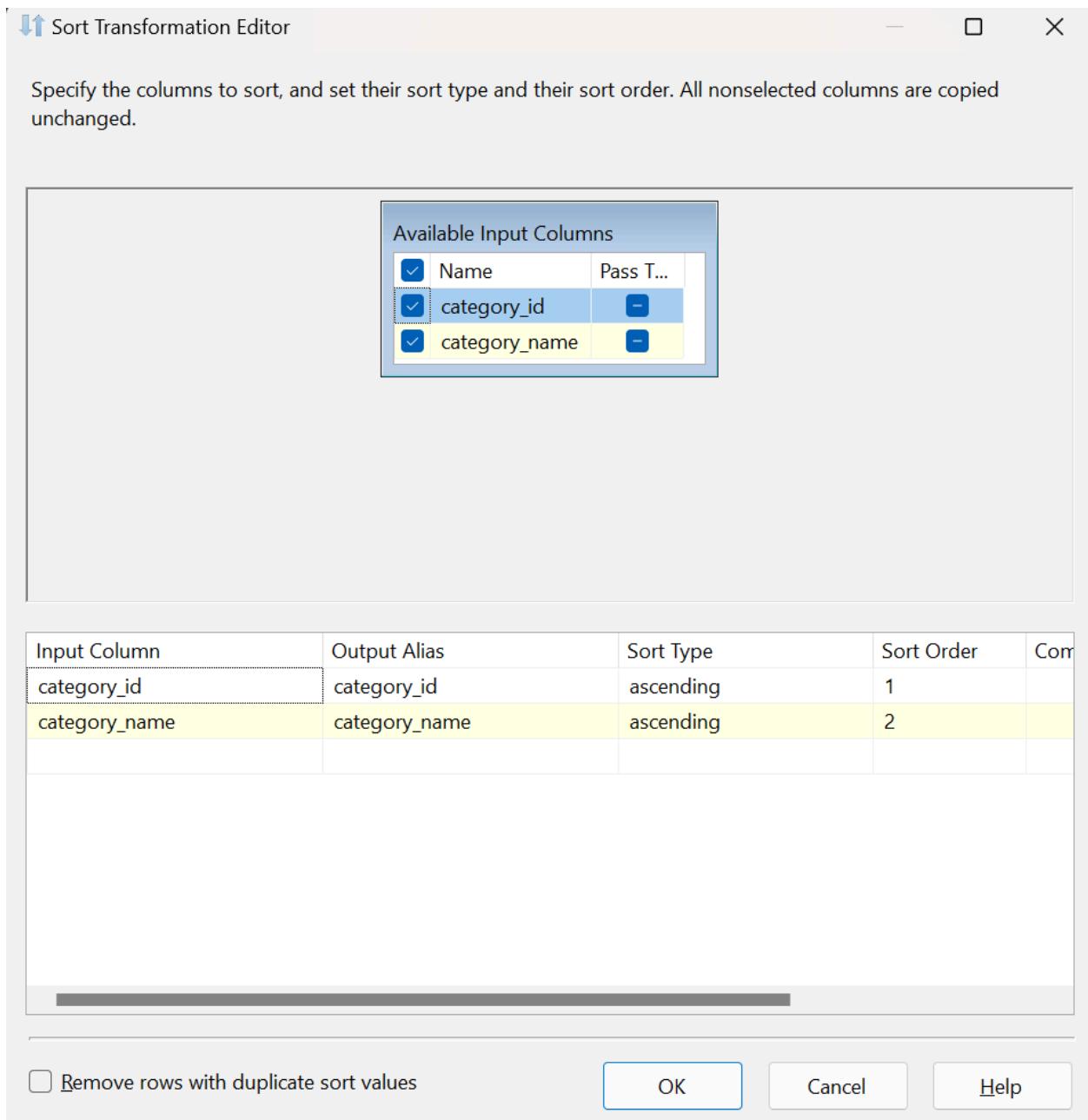
- Tiến hành edit cho Aggregate, tích chọn các thuộc tính cần thiết cho bảng Dim_Categories và chọn OK



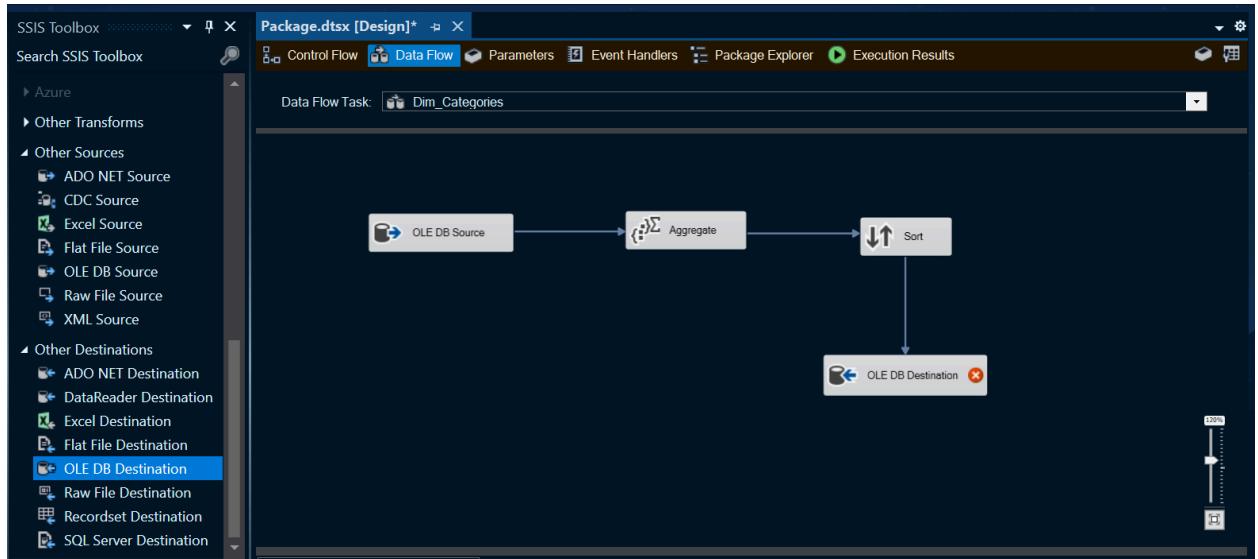
- Tiếp theo, thực hiện kéo thả công cụ Sort từ Toolbox để sắp xếp dữ liệu.



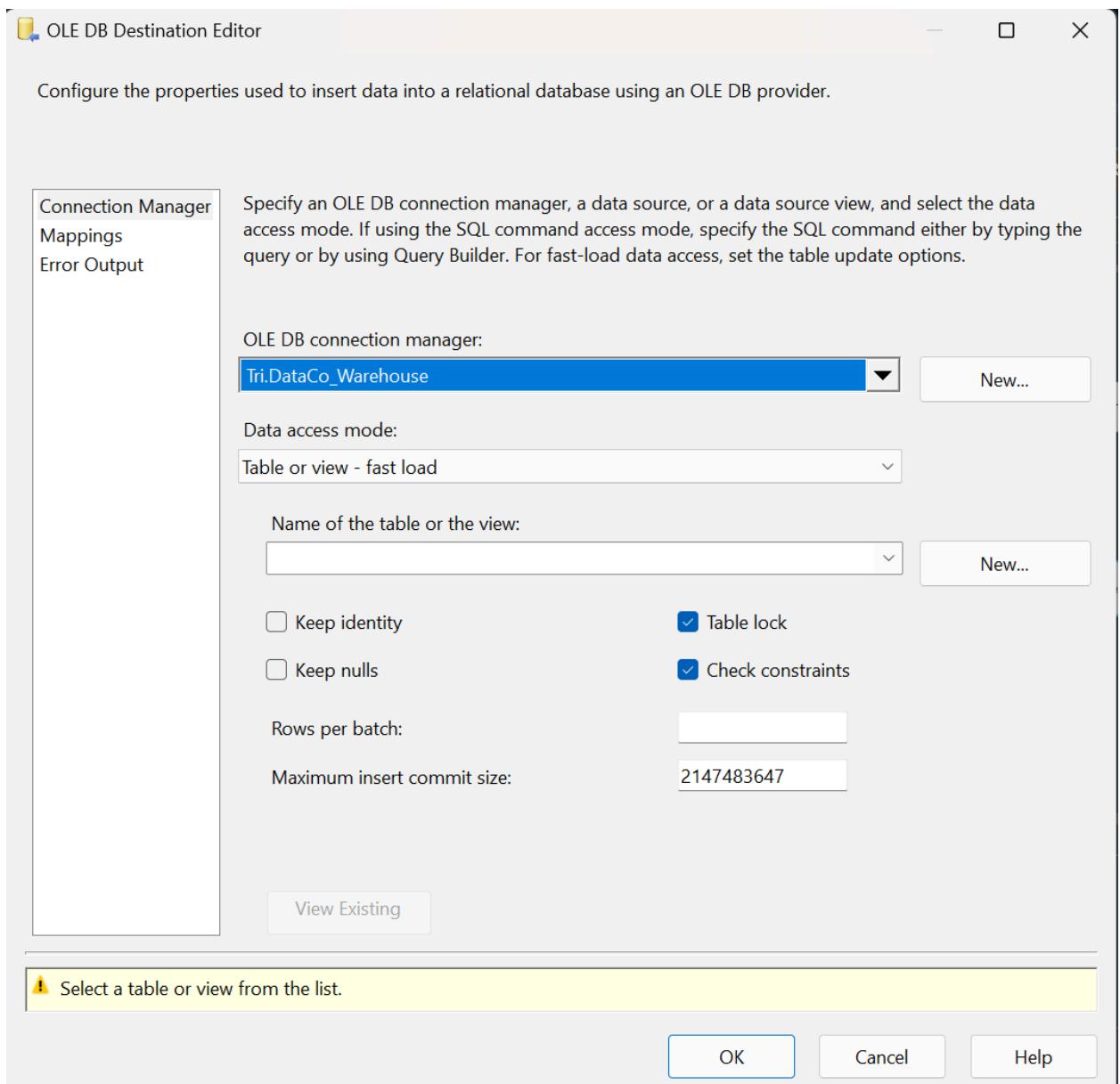
- Thực hiện edit Sort, tích chọn các thuộc tính cần sắp xếp theo thứ tự tăng dần và chọn OK.



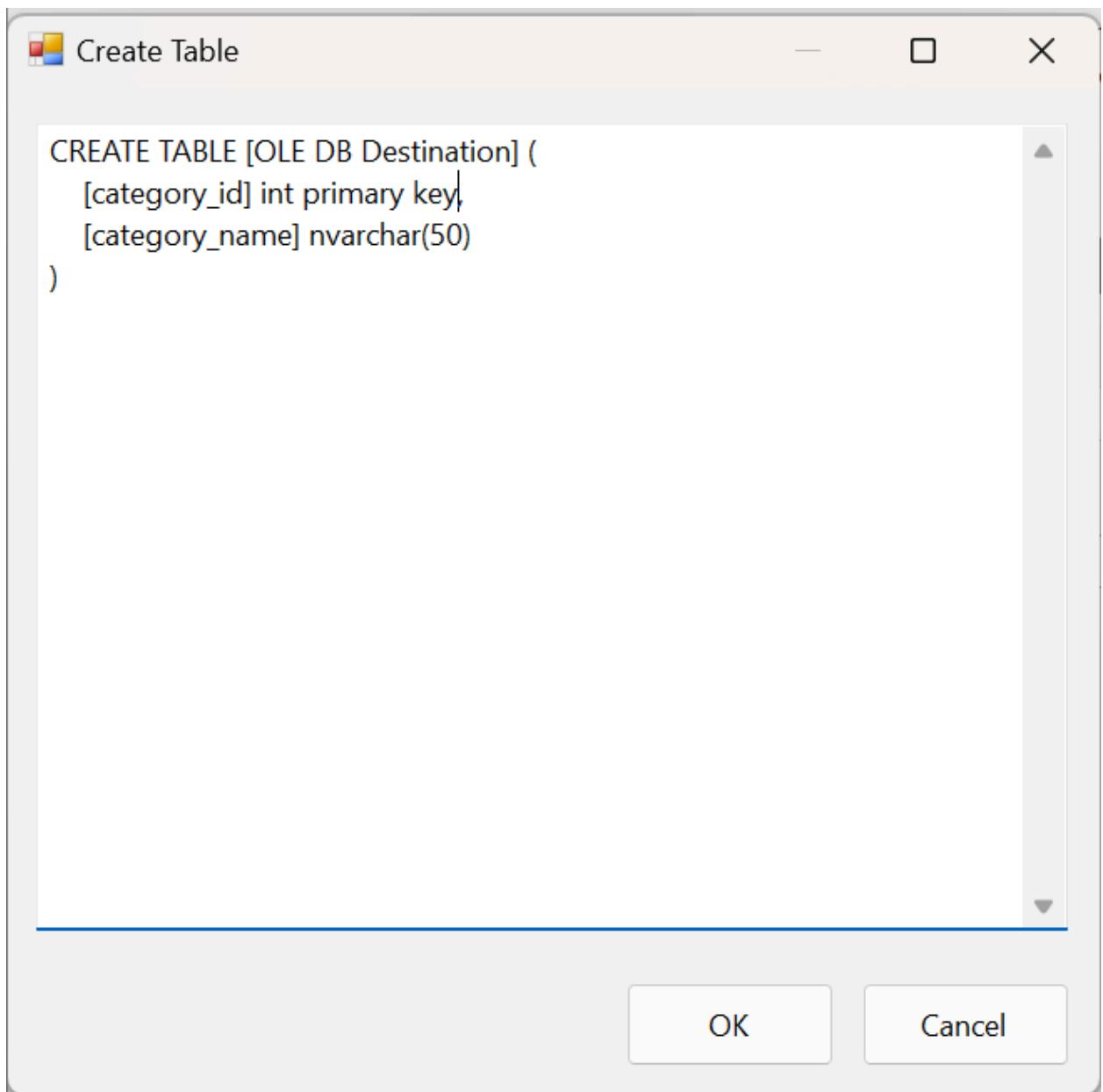
- Kéo thả công cụ OLE DB Destination từ Toolbox, đầu vào từ công cụ Sort.



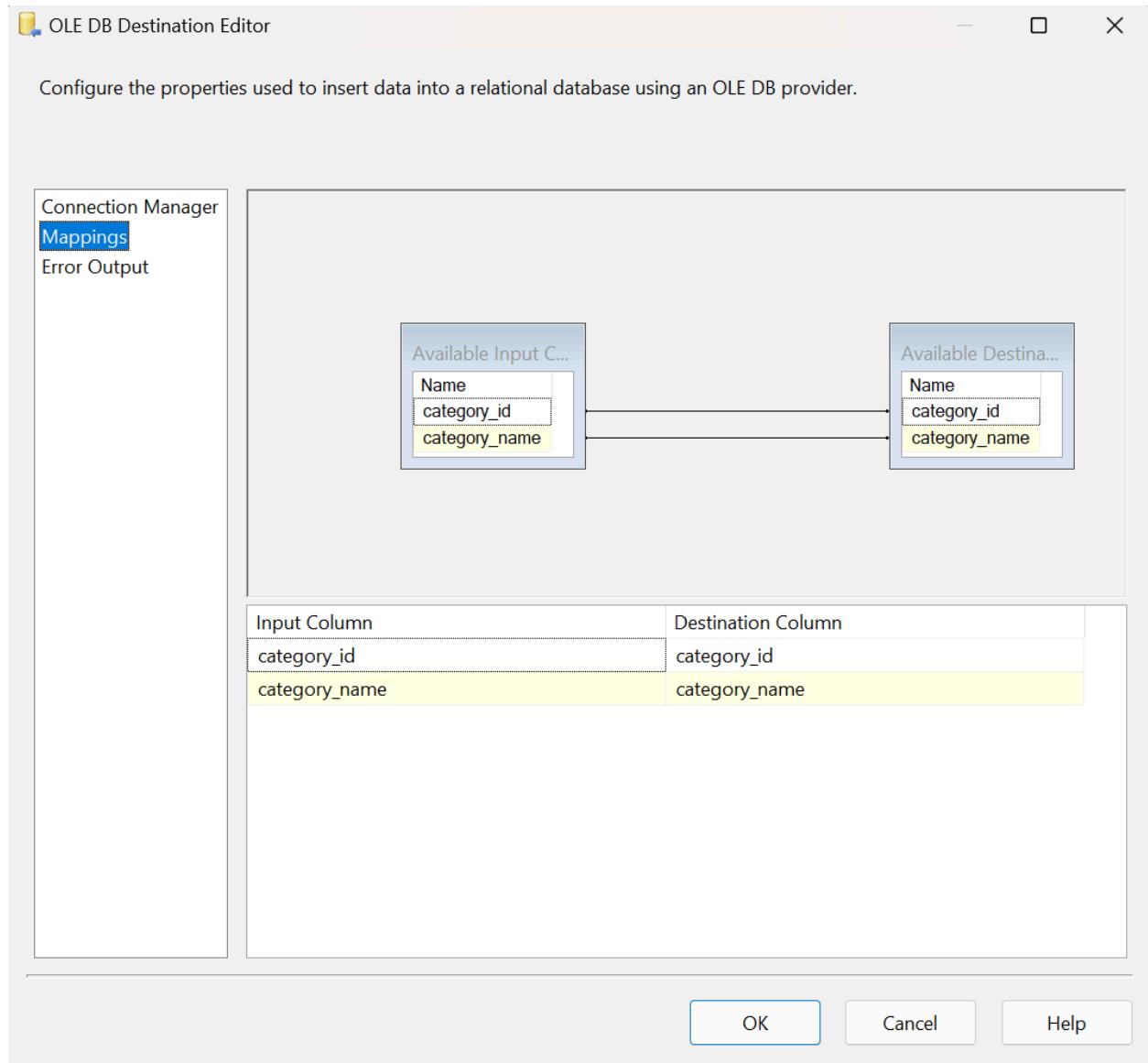
- Thực hiện edit OLE DB Source, chọn kết nối tới CoData_Warehouse



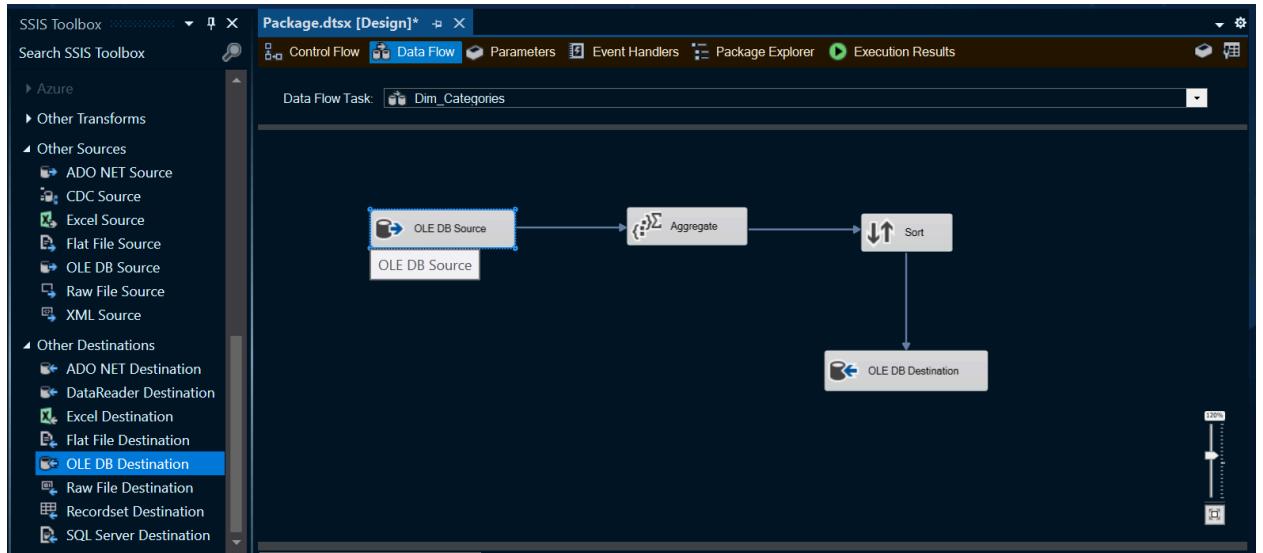
- Chọn New thực hiện tạo bảng mới với các thuộc tính và khóa chính cần thiết cho Dim_Categories, sau đó chọn OK



- Tiếp theo, chuyển qua tab Mapping và map các thuộc tính cần thiết.

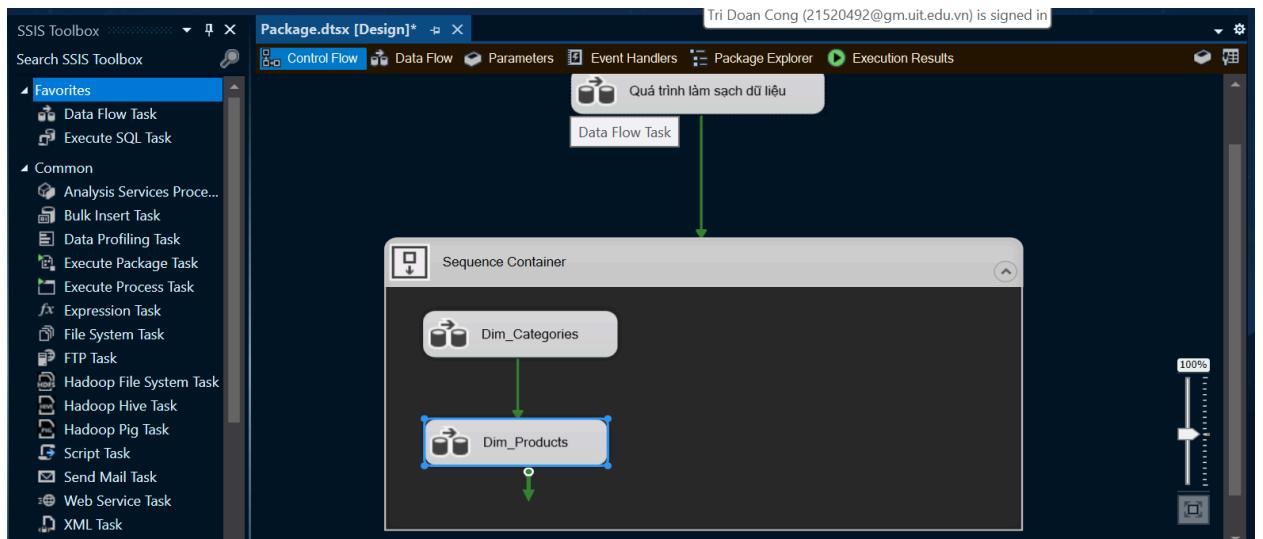


- Data Flow Dim_Categories hoàn thành.

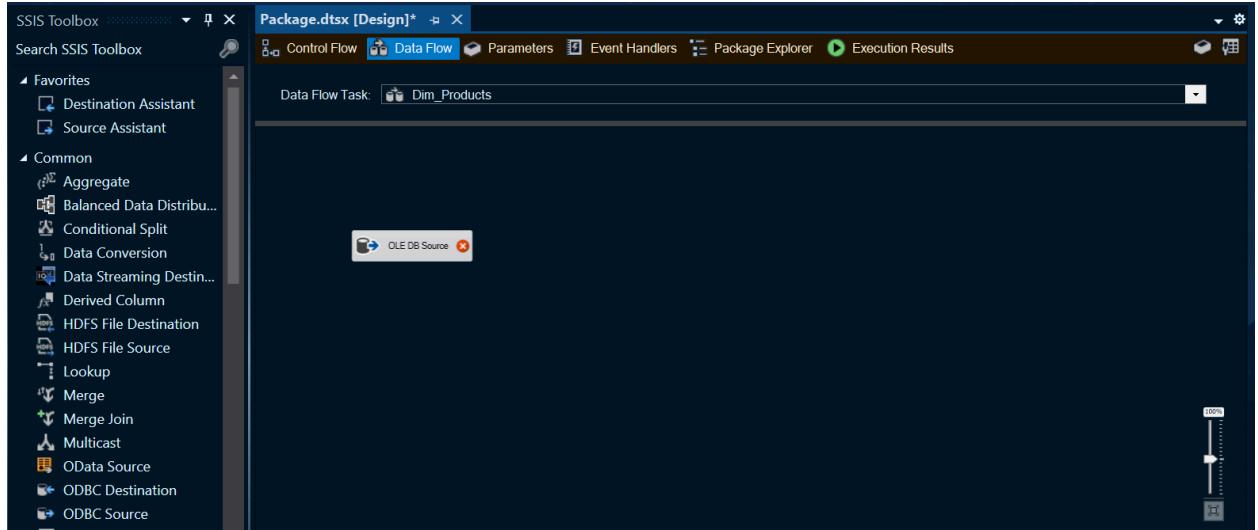


2.3.3. Tạo bảng Dim_Products

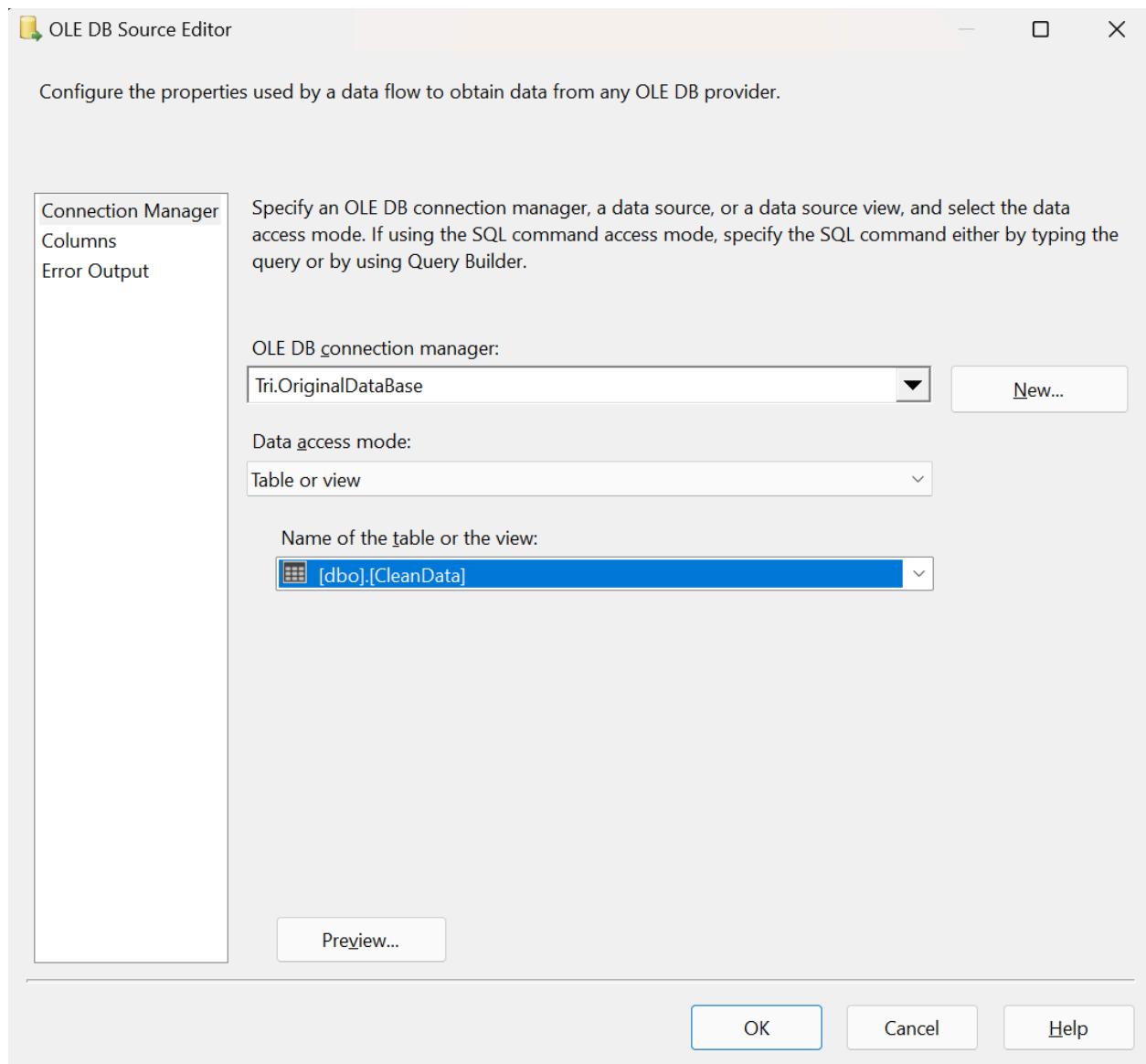
- Kéo thả công cụ Data Flow Task từ Toolbox, sau đó tiến hành đổi tên thành “Dim_Products”



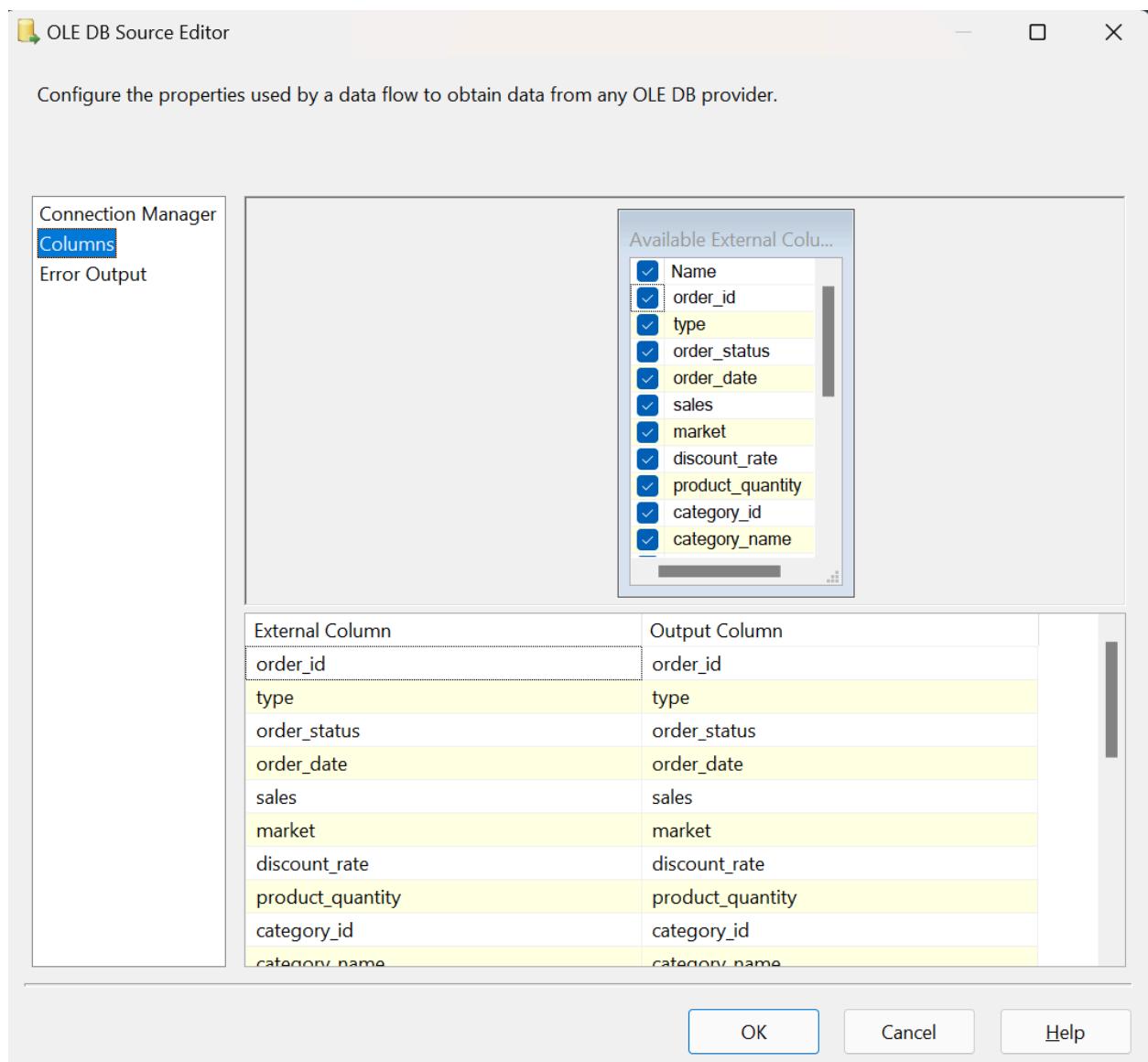
- Thực hiện edit Dim_Products, kéo thả công cụ OLE DB Source từ Toolbox.



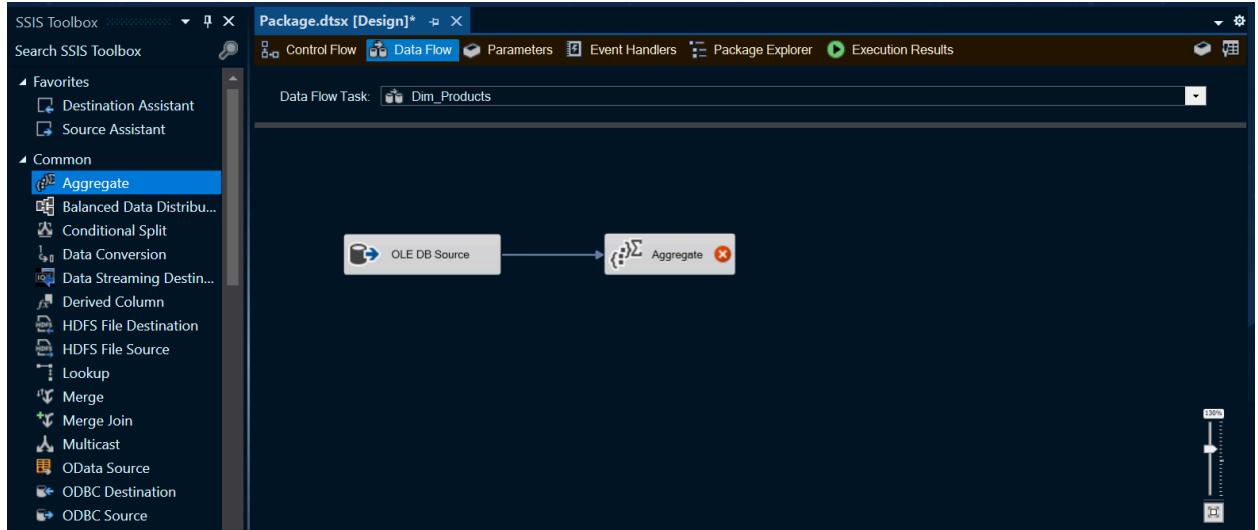
- Thực hiện edit OLE DB Source, cấu hình kết nối tới bảng CleanData trong database OriginalDatabase.



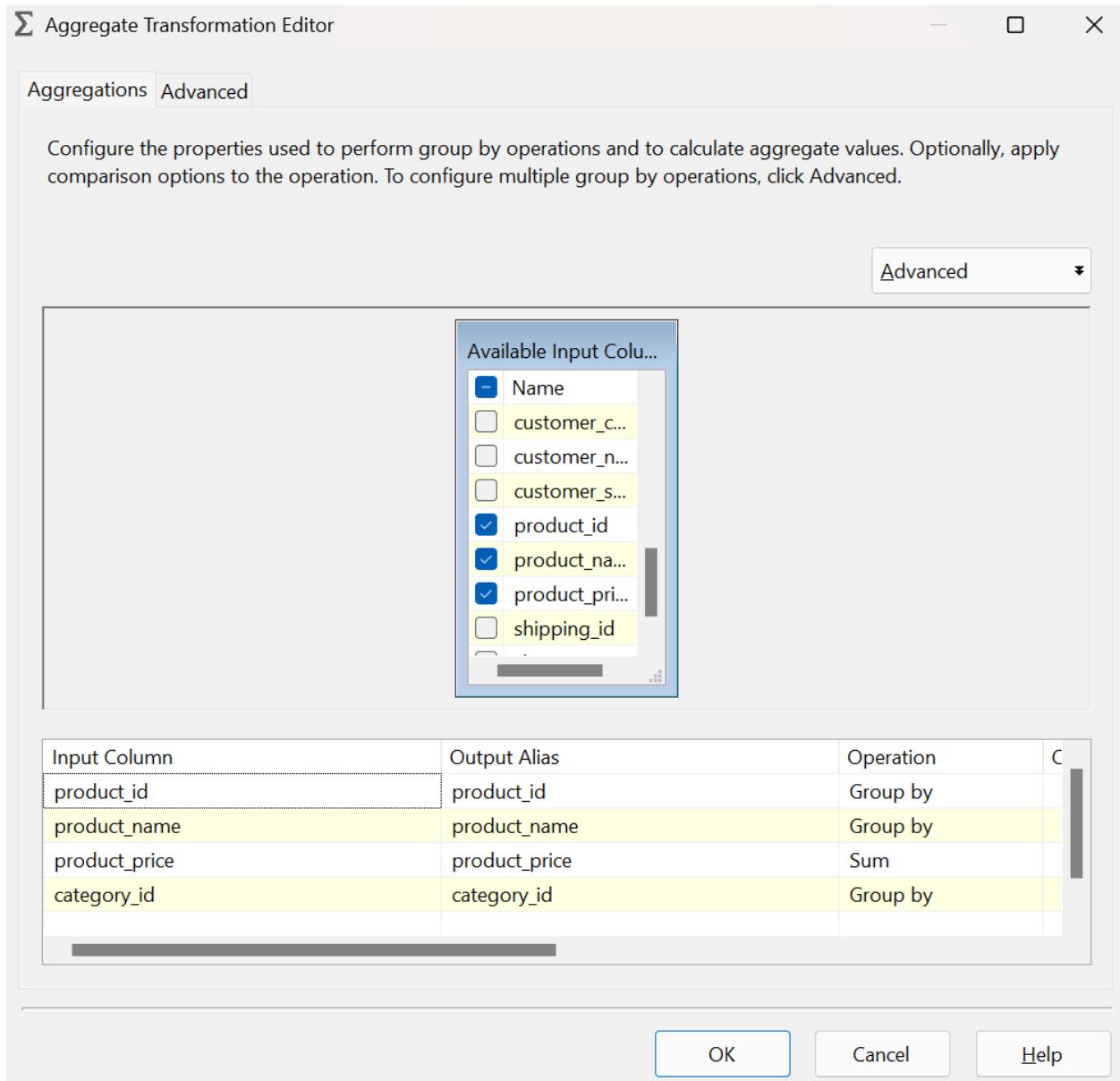
- Chuyển qua tab Column và tích chọn các thuộc tính cần thiết.



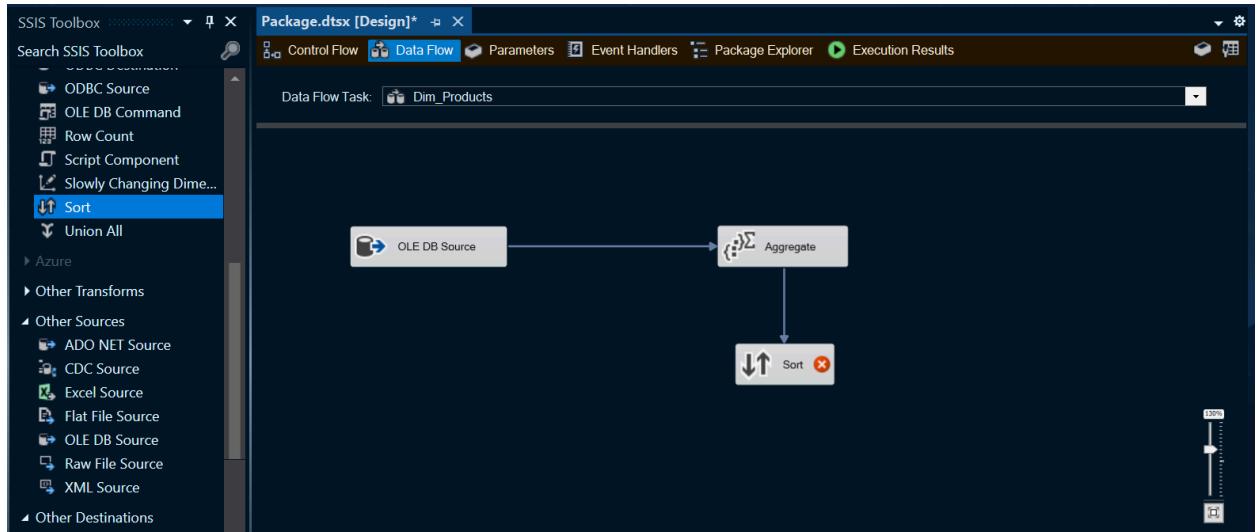
- Thực hiện kéo thả công cụ Aggregate từ Toolbox, với đầu vào là OLE DB Source để lọc những thuộc tính cần thiết cho bảng Dim_Products.



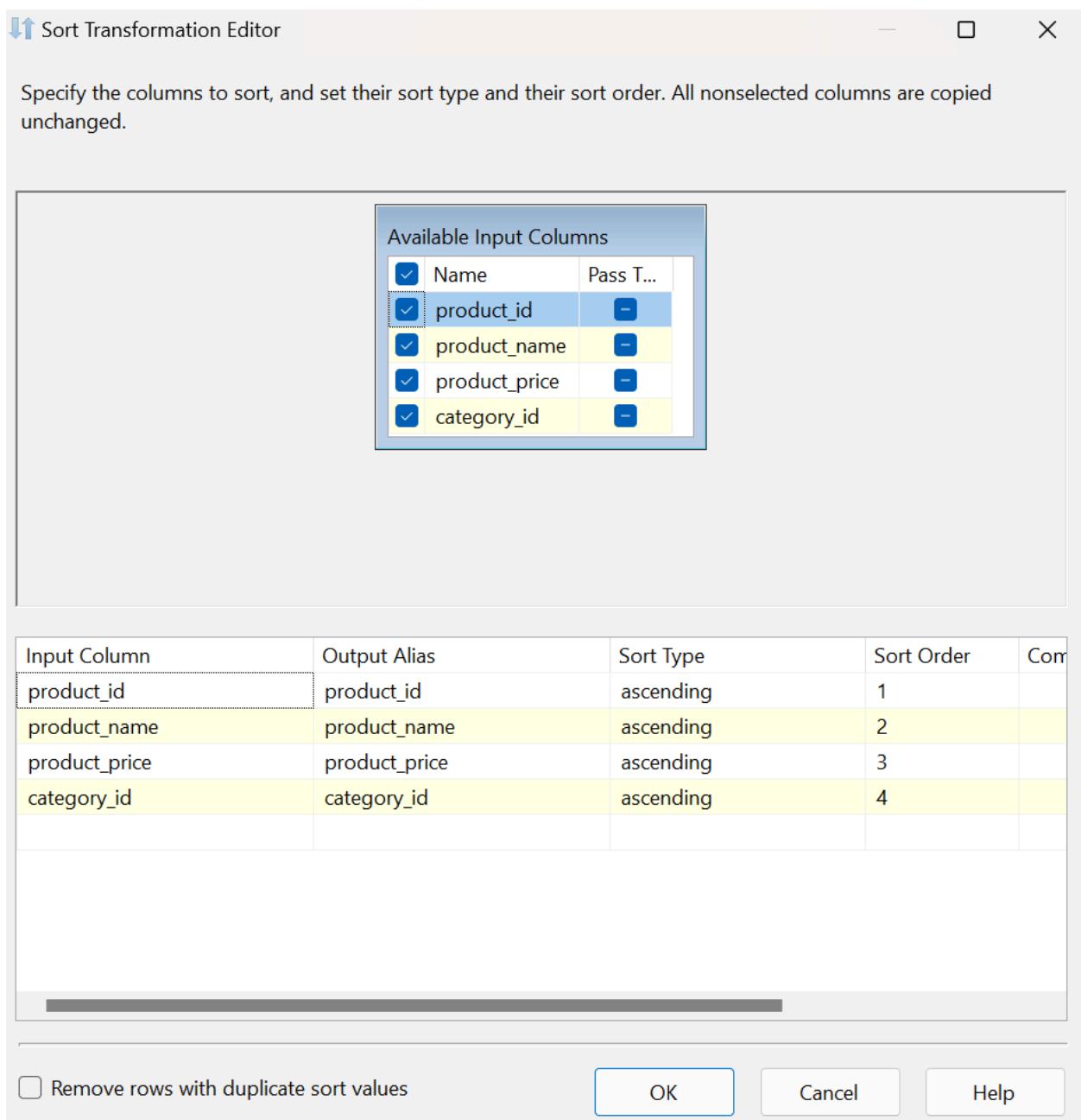
- Tiến hành Edit Aggregate, tích chọn những thuộc tính cần thiết cho bảng Dim_Products.



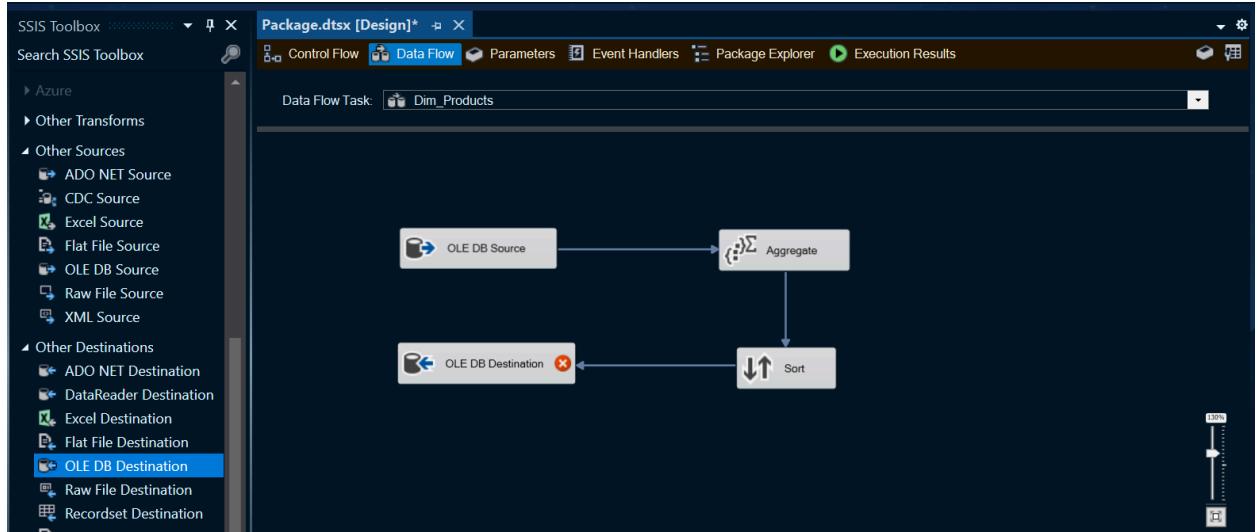
- Kéo thả công cụ Sort từ Toolbox để sắp xếp lại dữ liệu cho bảng Dim_Products.



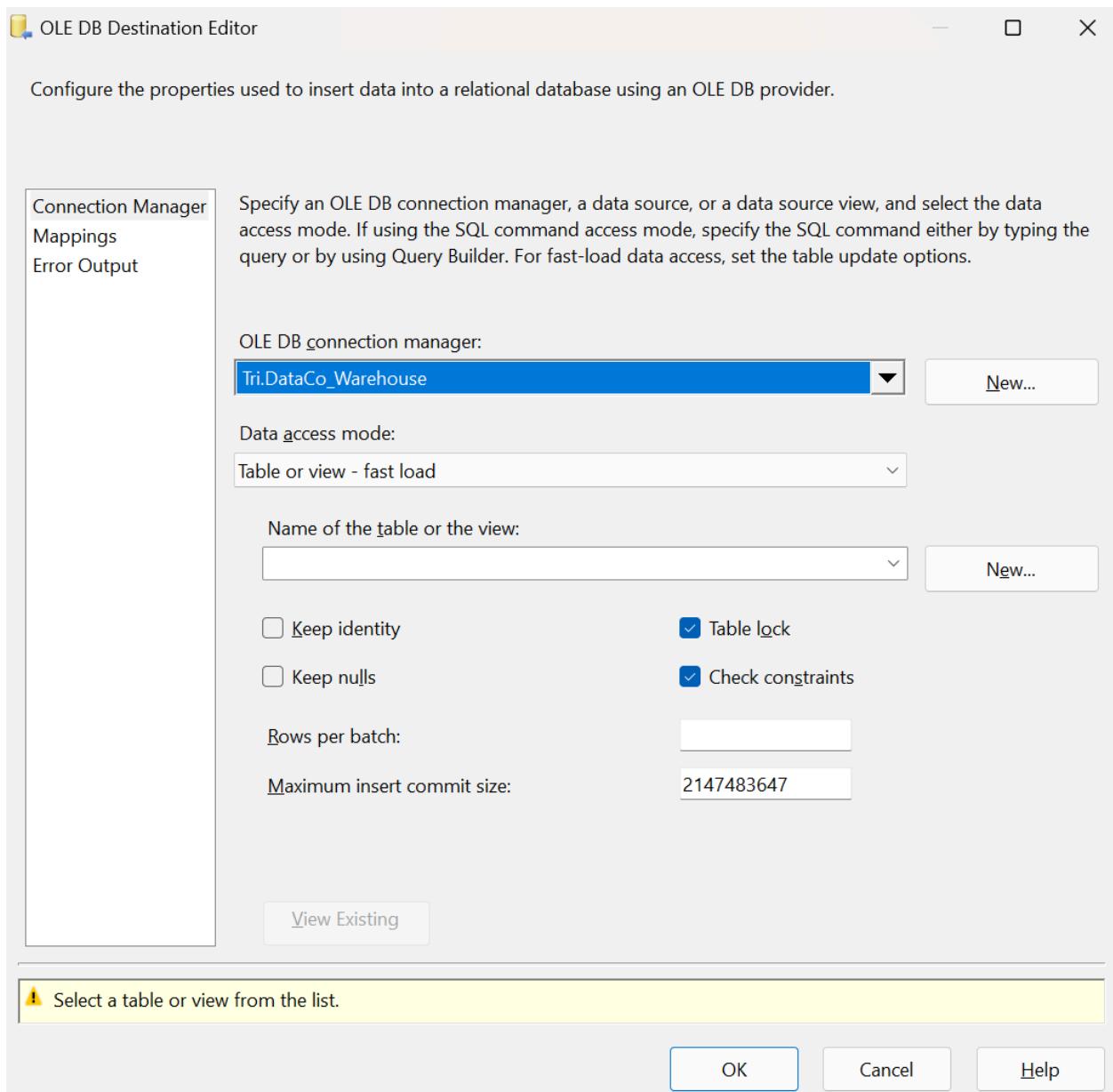
- Tiến hành edit Sort, tích chọn các thuộc tính sắp xếp dần cho bảng Dim_Products.



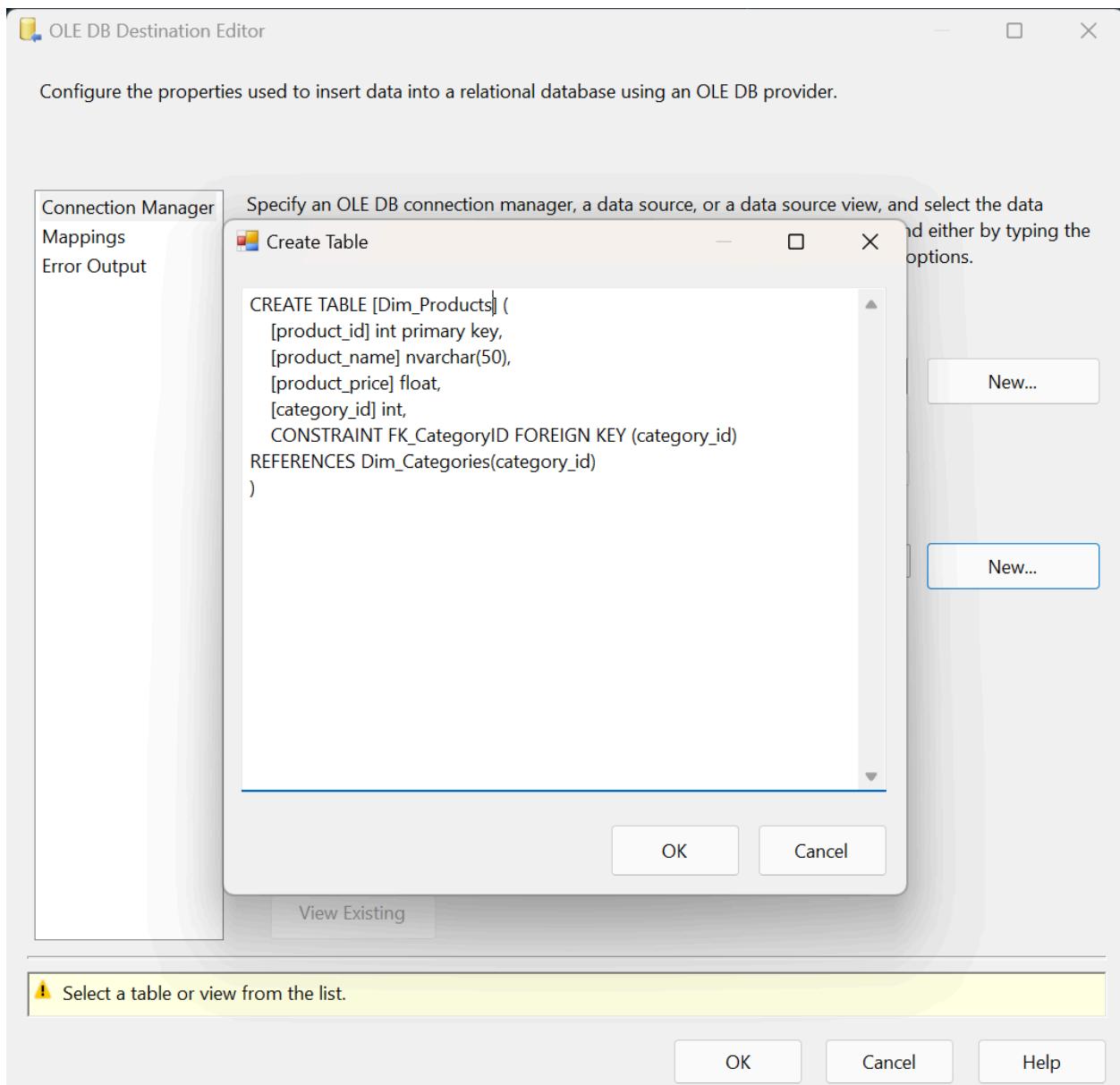
- Kéo thả công cụ OLE DB Destination từ Toolbox và tiến hành edit.



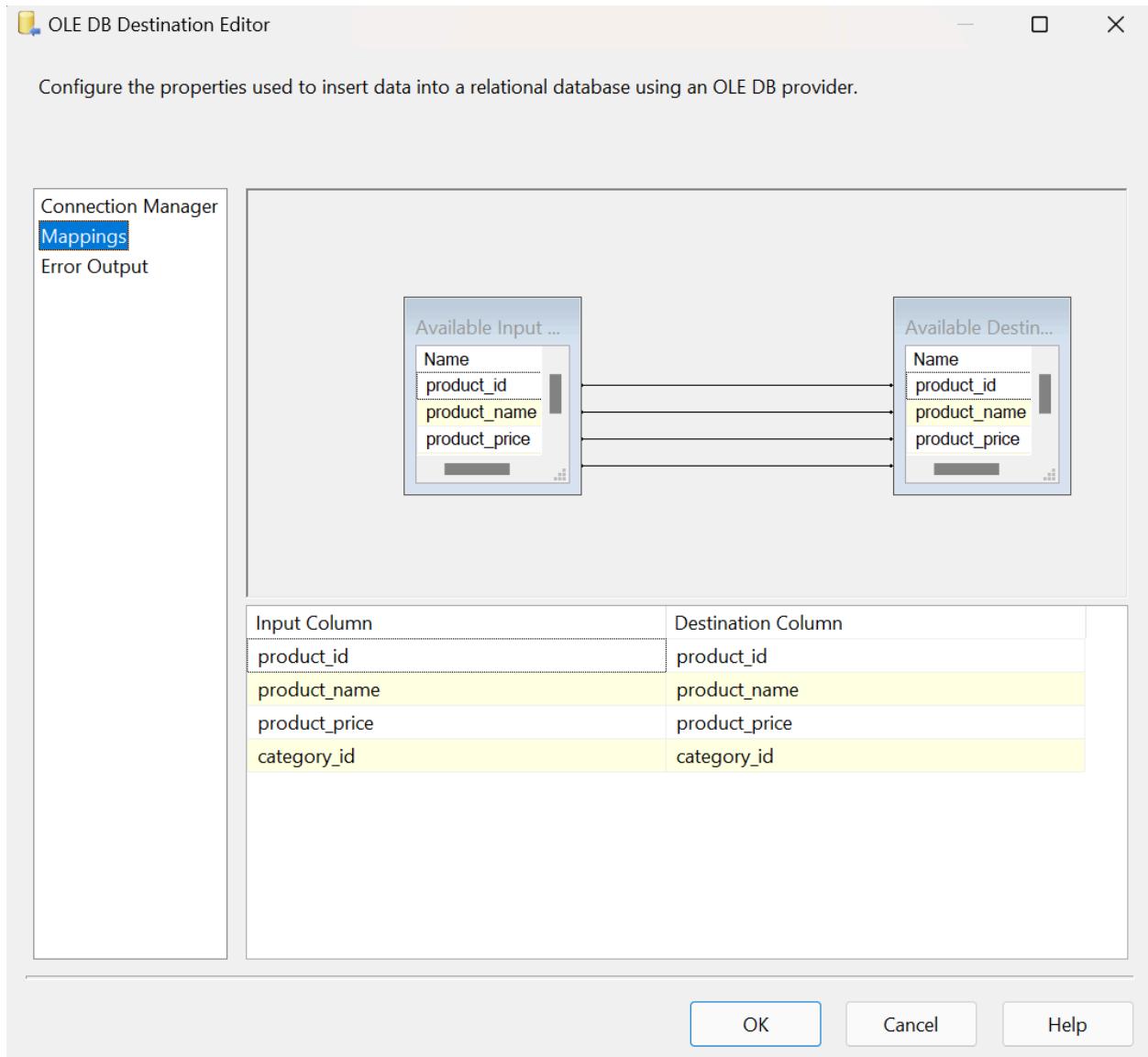
- Thực hiện edit OLE DB Source, chọn kết nối tới database CoData_Warehouse.



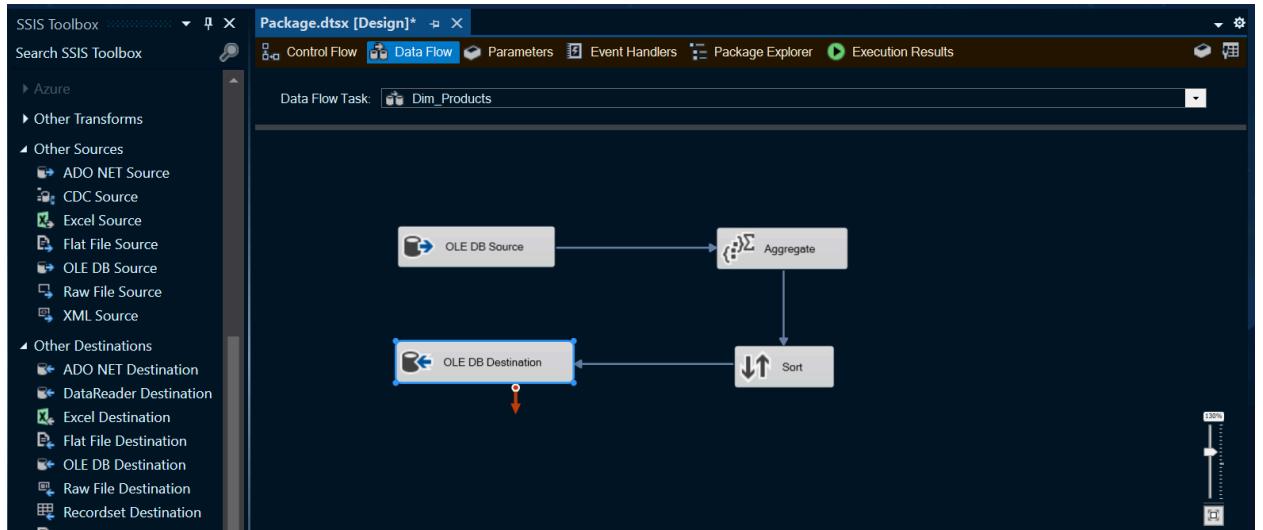
- Tiếp theo, chọn New để tạo bảng Dim_Products



- Chọn qua tab Mapping và tiến hành map các thuộc tính cần thiết cho bảng Dim_Products.

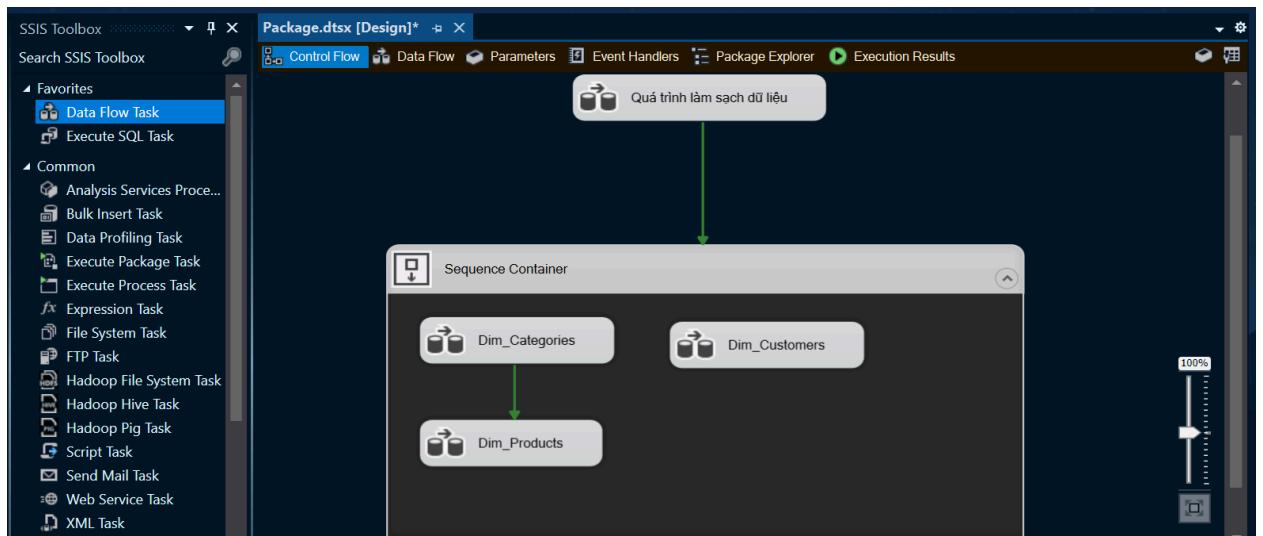


- Data Flow Task Dim_Products khi hoàn thành.

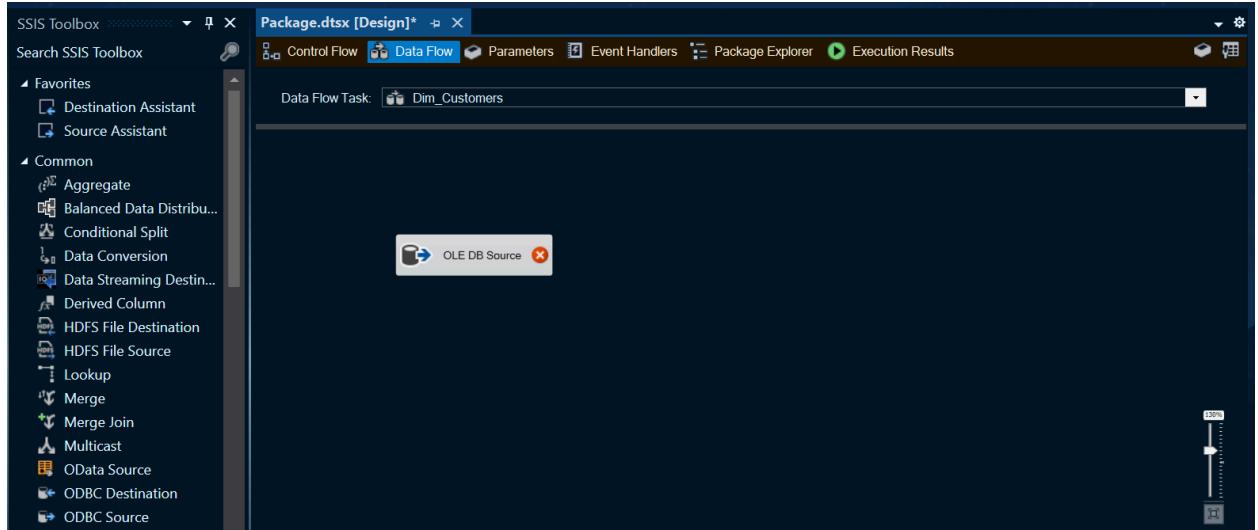


2.3.4. Tạo bảng Dim_Customers

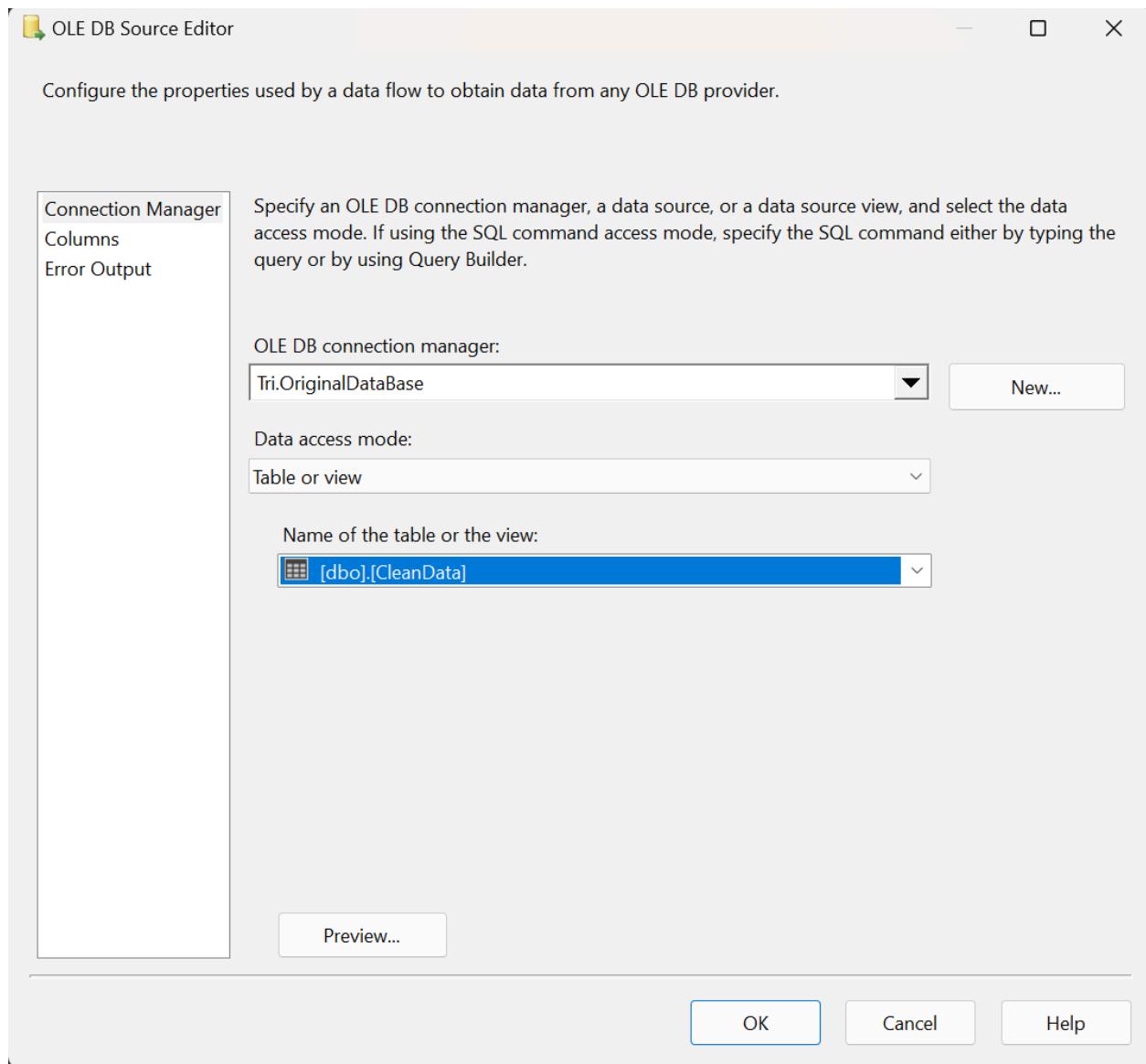
- Kéo thả công cụ Data Flow Task từ Toolbox, sau đó đổi tên thành Dim_Customers.



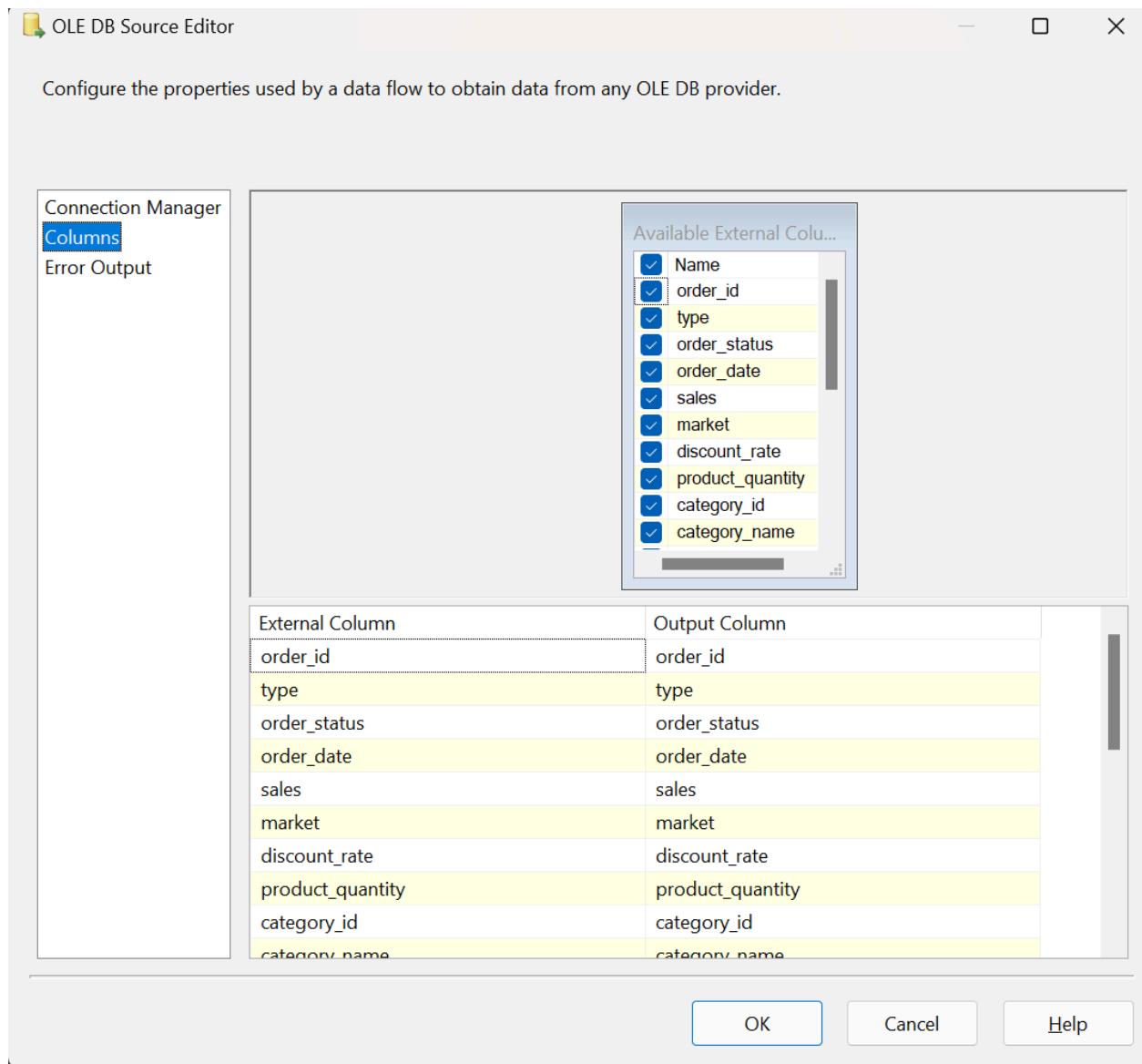
- Thực hiện edit Dim_Customers, kéo thả công cụ OLE DB Source từ Toolbox.



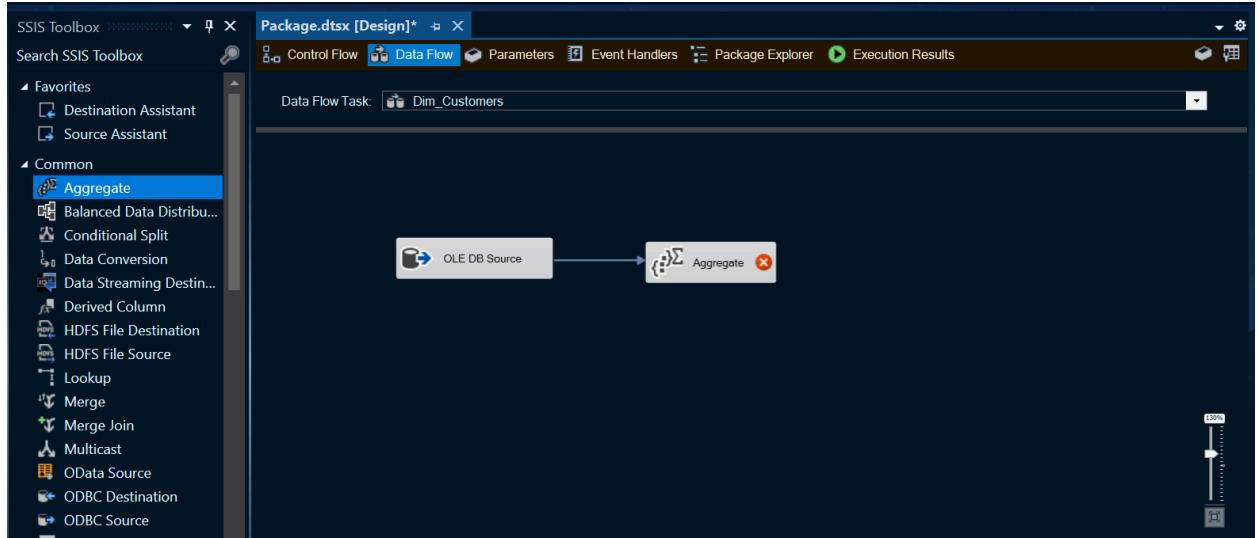
- Sau đó edit OLE DB Source, thực hiện kết nối tới bảng CleanData trong database OriginalDatabase.



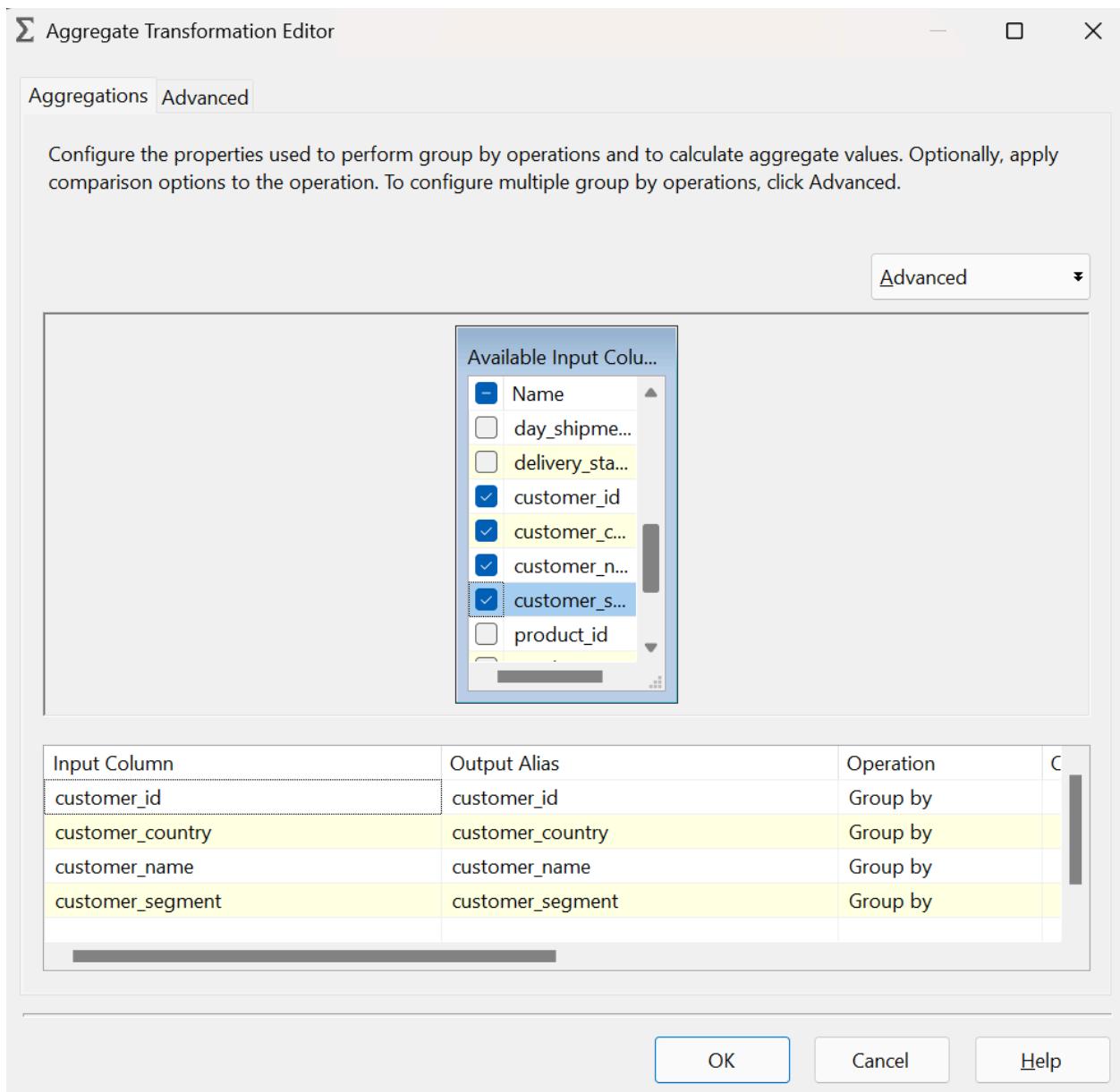
- Chuyển qua tab Column và tích chọn các thuộc tính cần thiết.



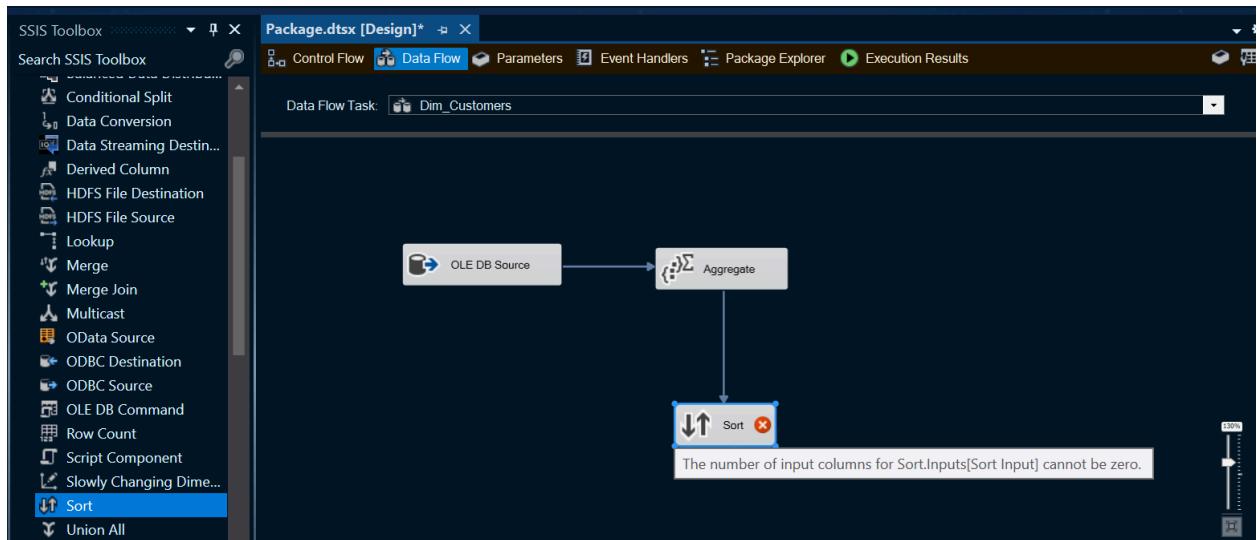
- Kéo thả công cụ Aggregate từ Toolbox, với đầu vào là OLE DB Source để chọn các thuộc tính cần thiết cho bảng Dim_Customers.



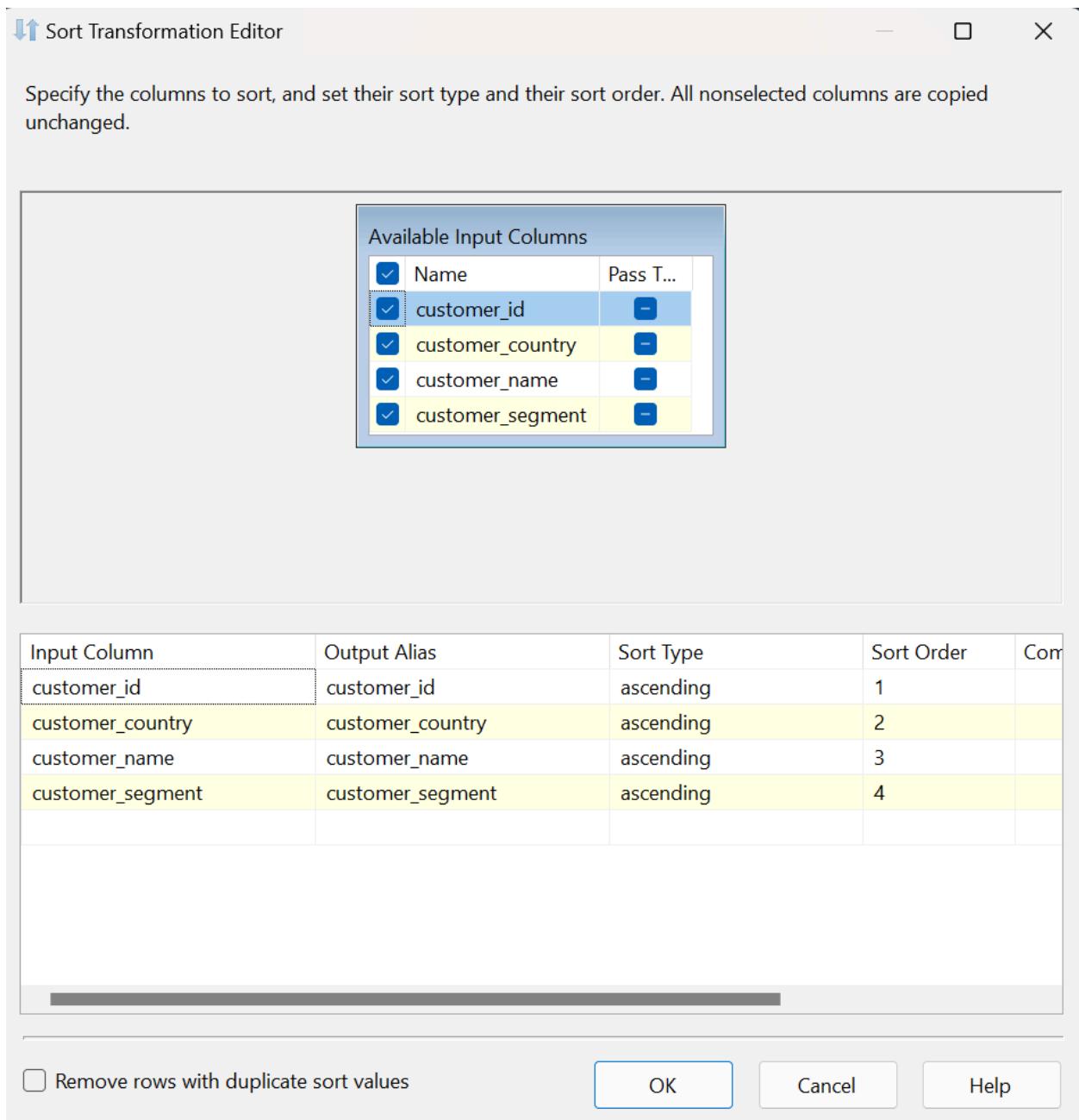
- Thực hiện edit Aggregate, tích chọn các thuộc tính cần lọc ra cho bảng Dim_Customers, sau đó chọn OK



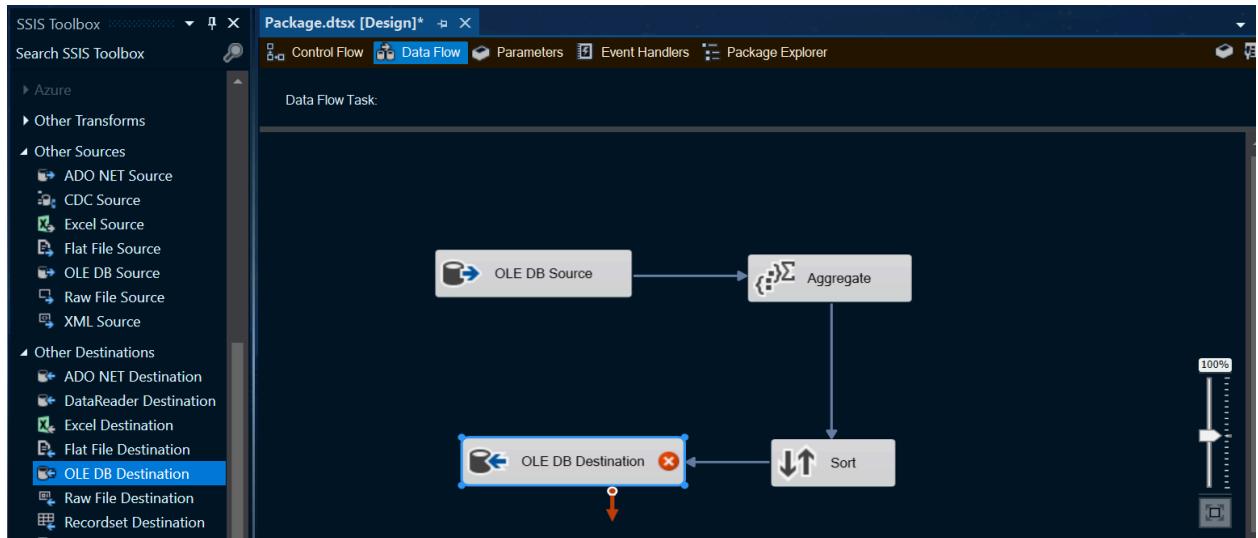
- Kéo thả công cụ Sort từ Toolbox để sắp xếp dữ liệu cho bảng Dim_Customers, đầu vào là dữ liệu được lọc ra từ AggreGate.



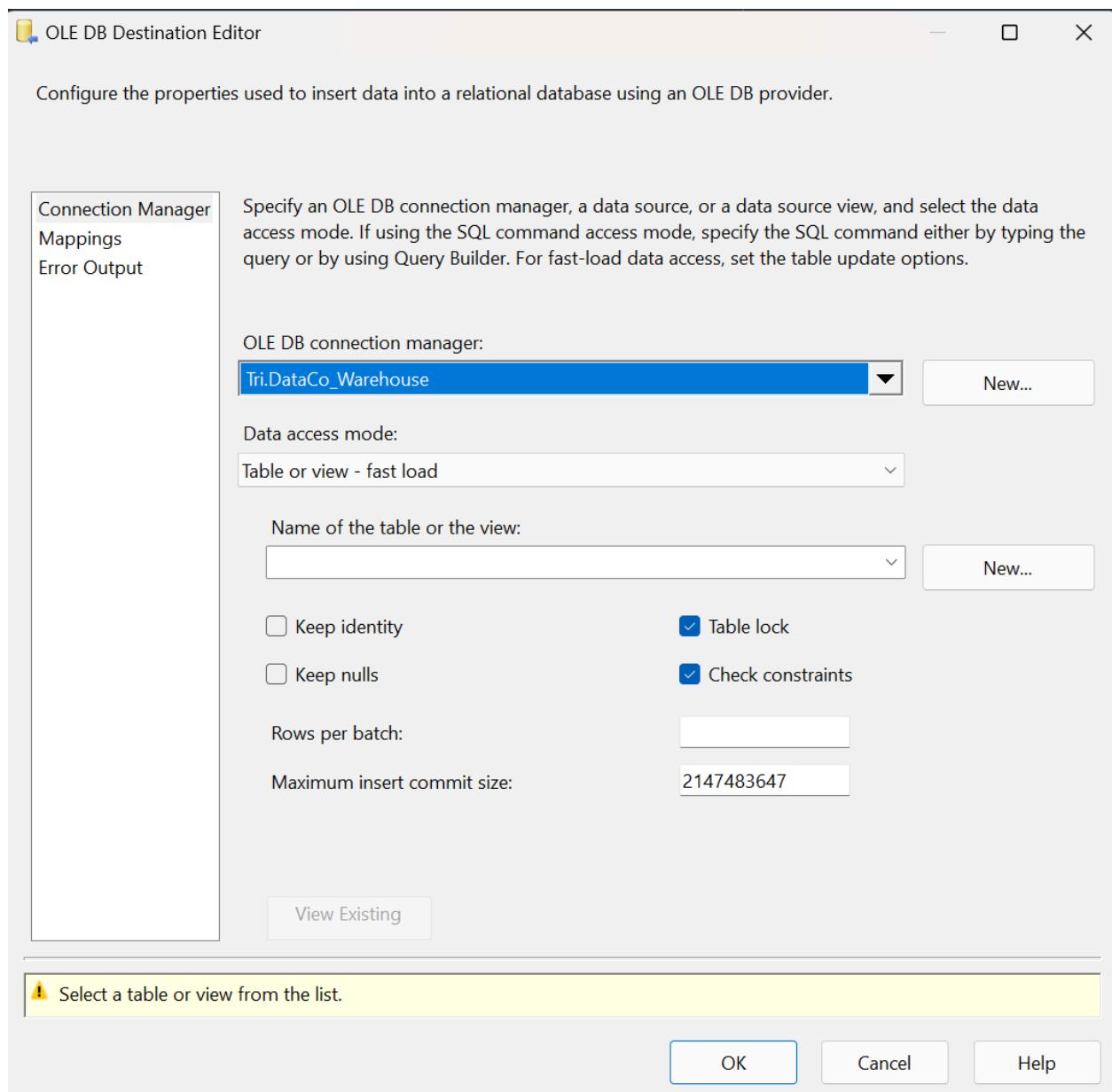
- Thực hiện Edit Sort, tích chọn các thuộc tính cần sắp xếp tăng dần trong bảng Dim_Customers.



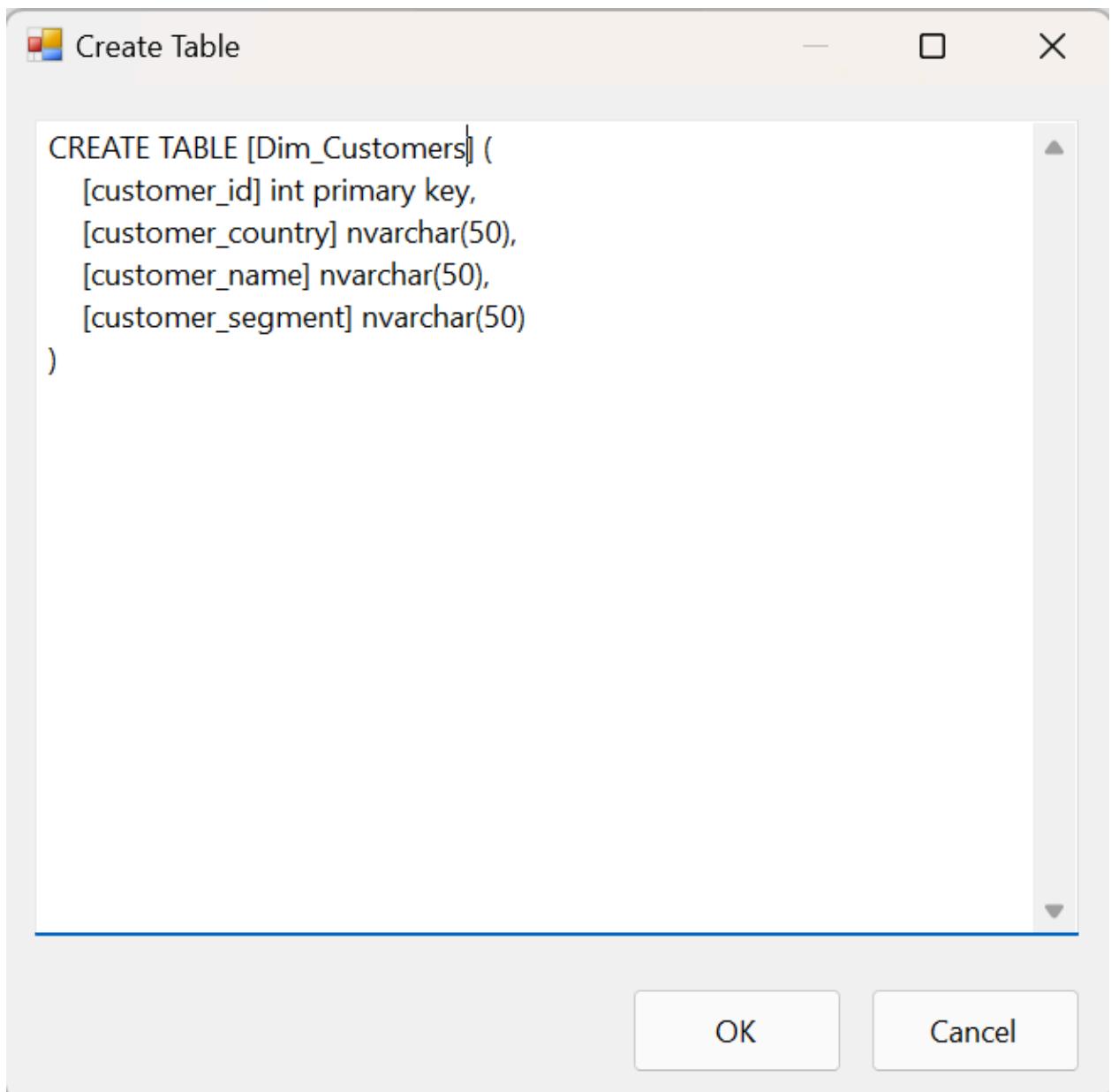
- Kéo thả công cụ OLE DB Destination từ Toolbox, với đầu vào là dữ liệu từ Sort, sau đó tiến hành edit



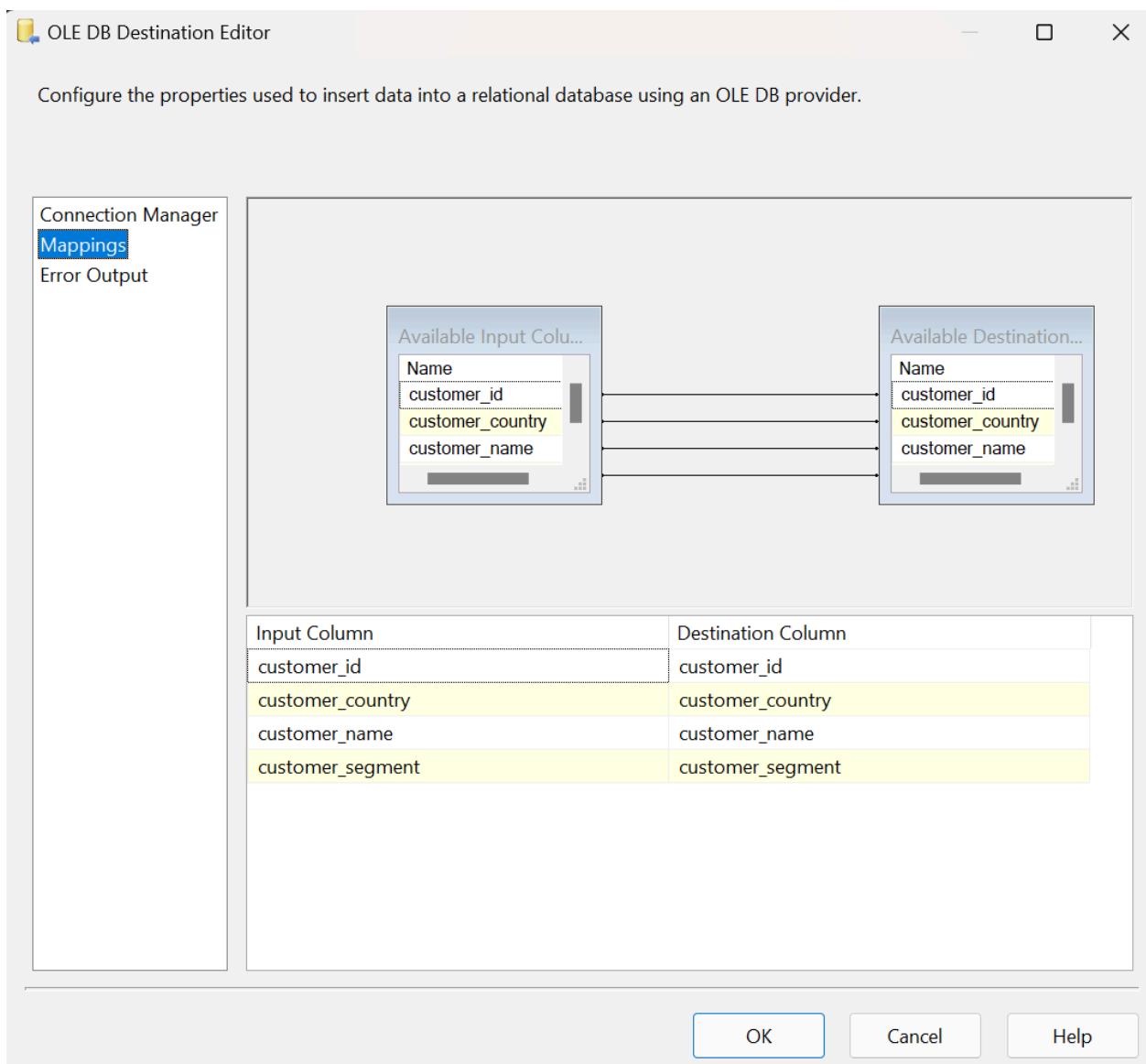
- Tiến hành Edit OLE DB Destination, thực hiện kết nối tới database DataCo_Warehouse.



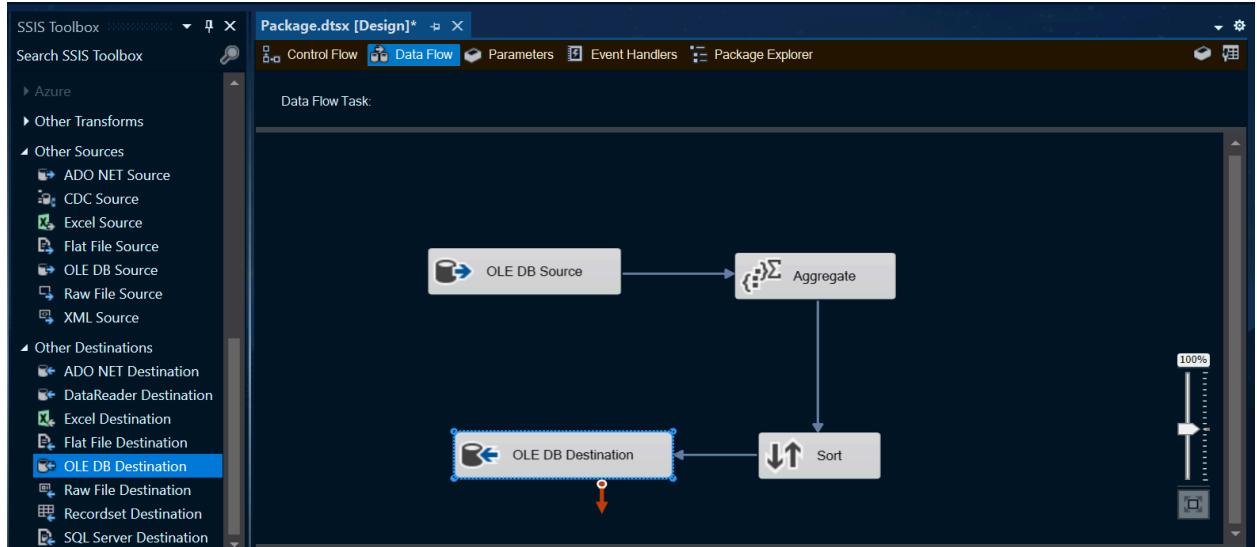
- Sau đó chọn New để tạo bảng Dim_Customers.



- Chuyển qua tab Mapping và thực hiện map các thuộc tính cần thiết

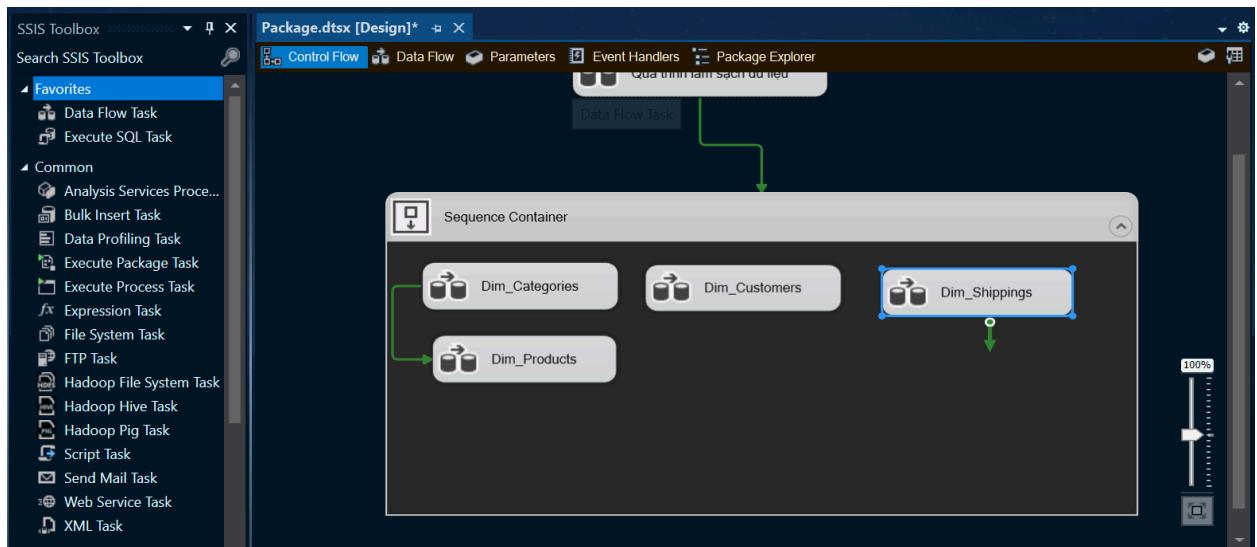


- Data Flow Task Dim_Customers khi hoàn thành.

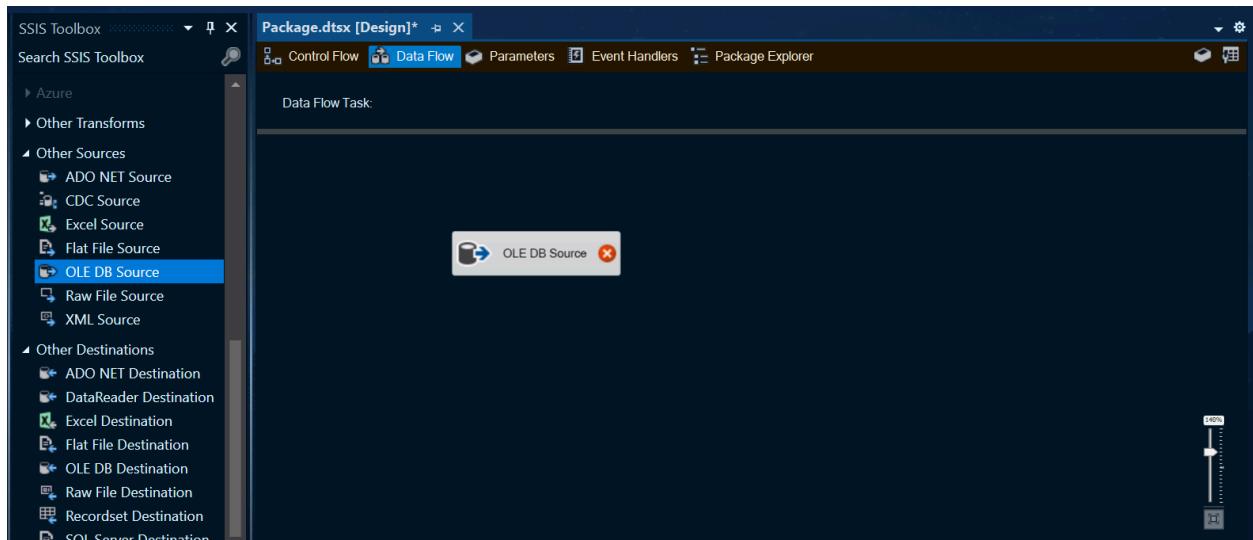


2.3.4. Tạo bảng Dim_Shippings

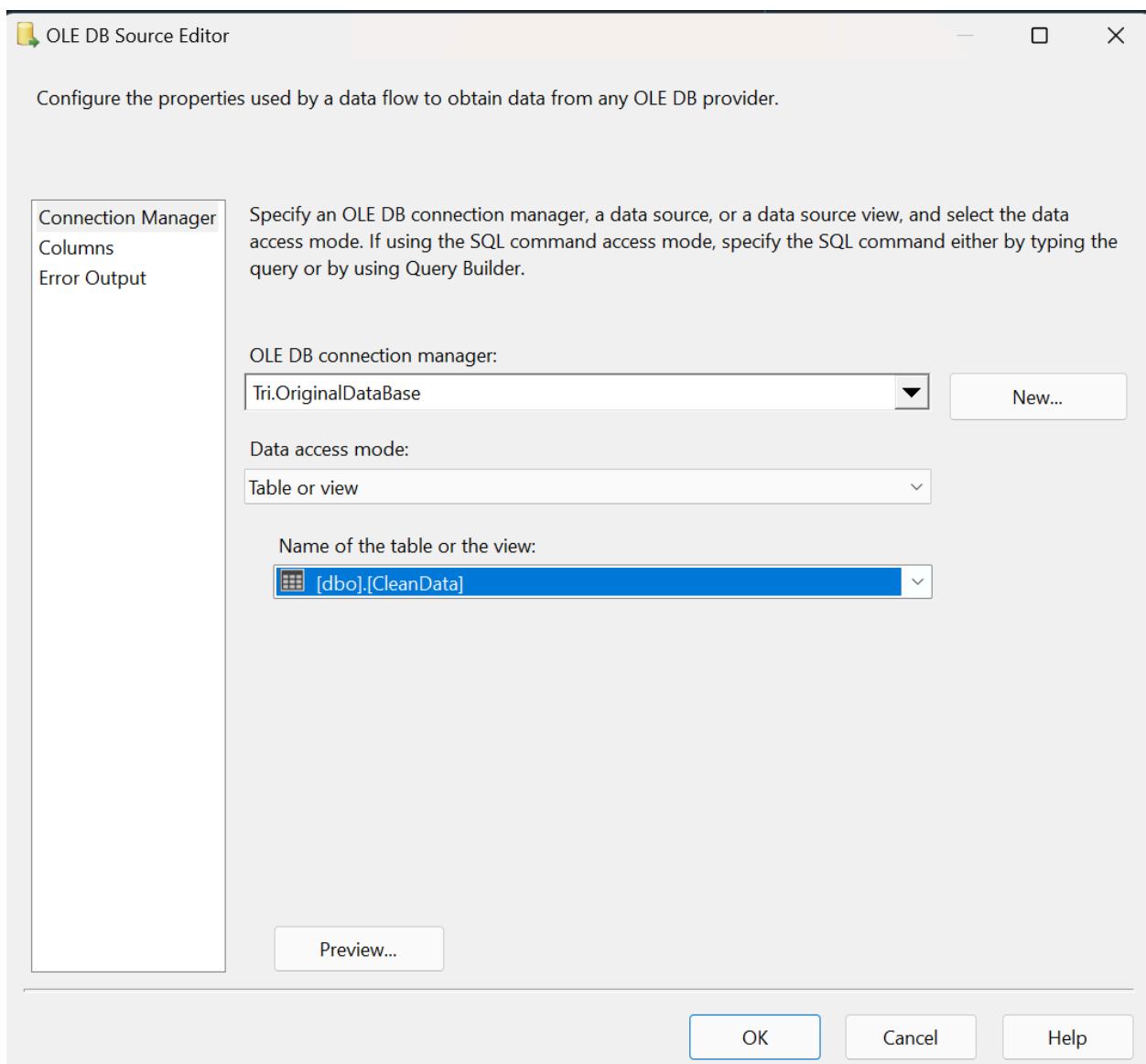
- Kéo thả công cụ Data Flow Task từ Toolbox và tiến hành đổi tên thành “Dim_Shippings”



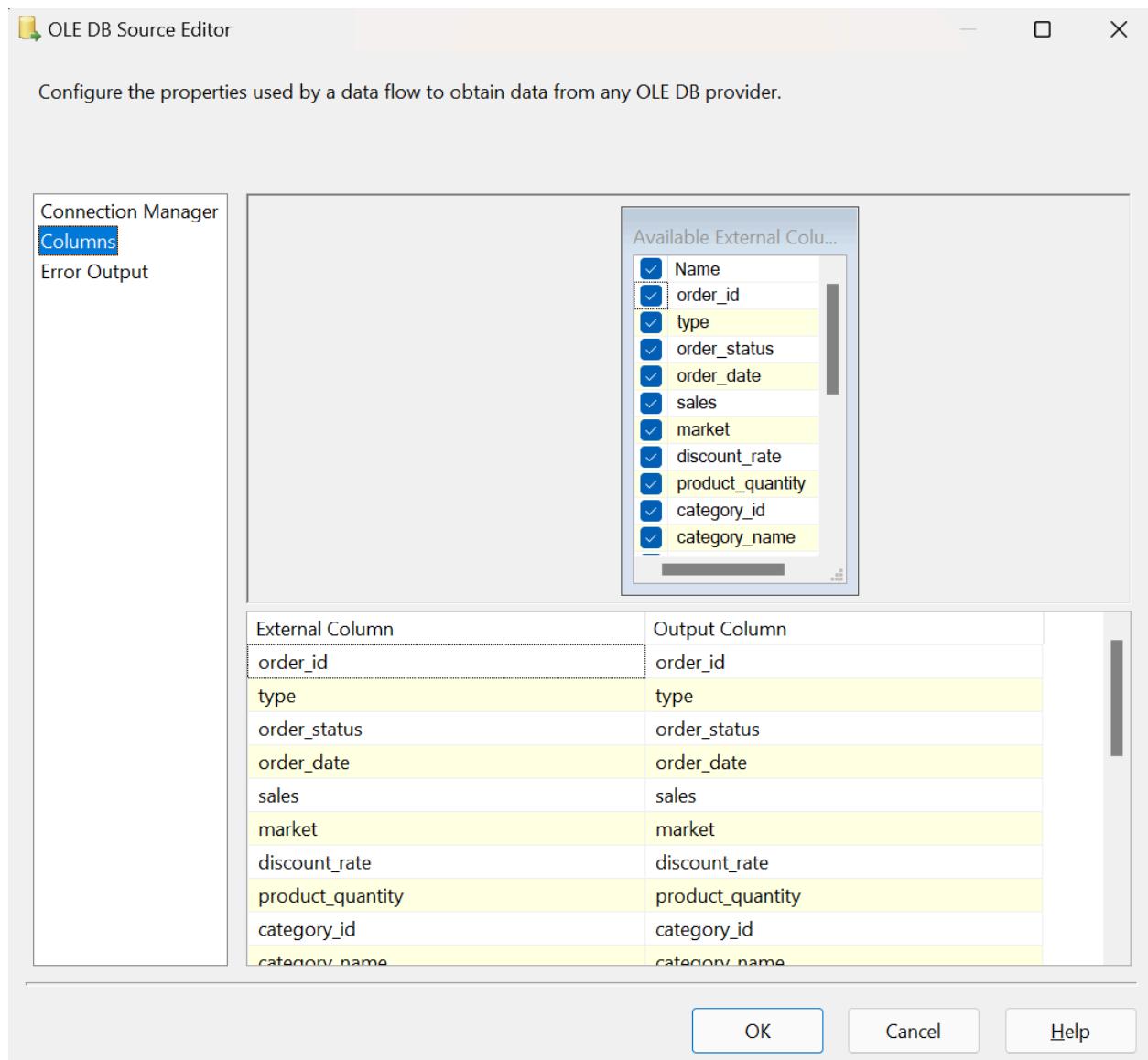
- Thực hiện cấu hình cho Dim_Shippings, kéo thả công cụ OLE DB Source từ Toolbox.



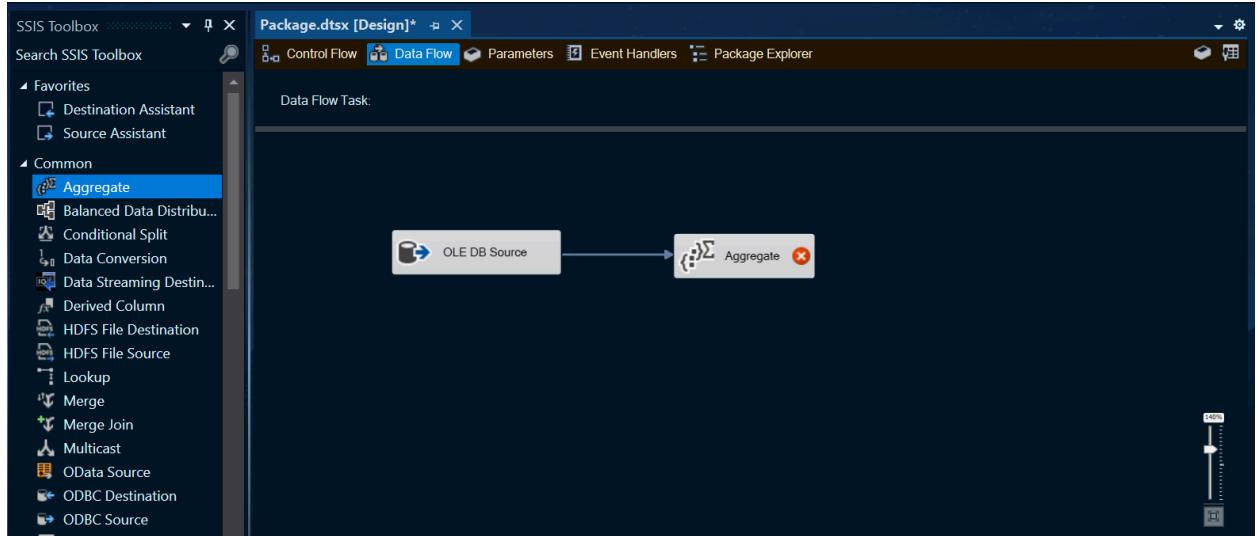
- Tiến hành edit, OLE DB Source, thực hiện kết nối tới bảng CleanData trong database OriginalDatabase.



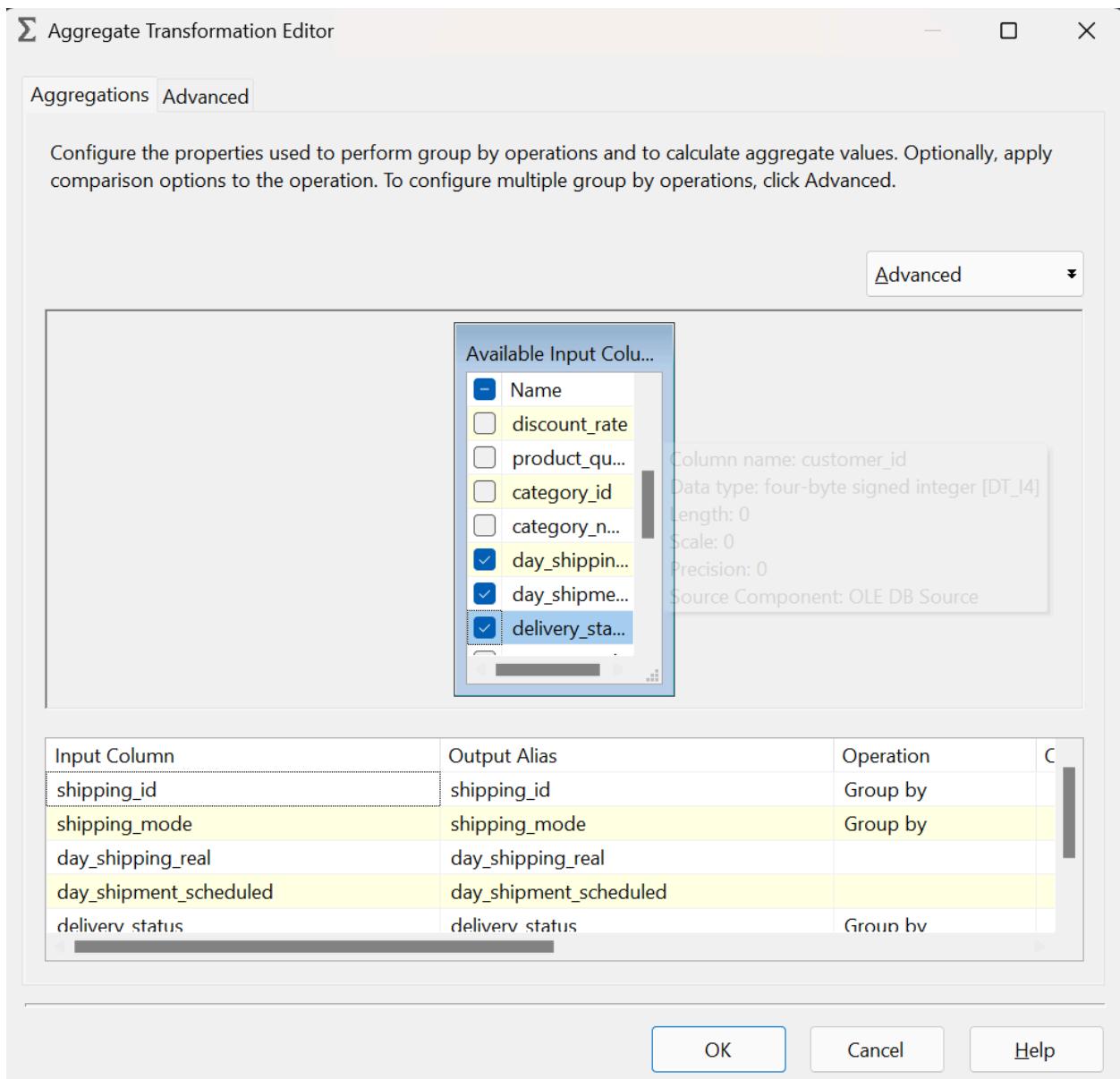
- Chọn qua bảng Column và chọn các thuộc tính cần thiết.



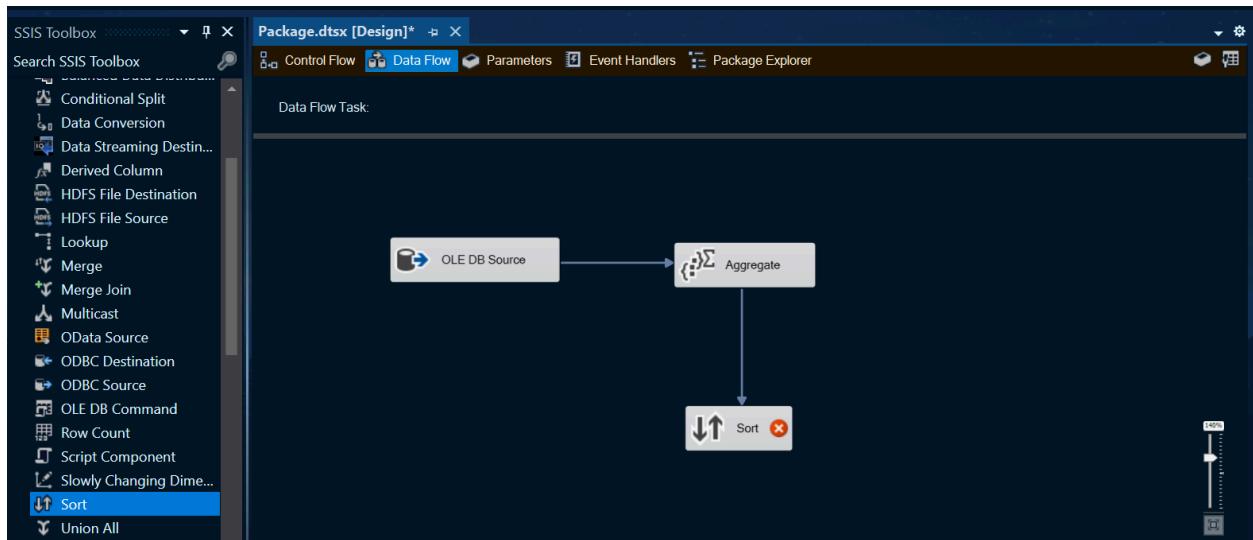
- Kéo thả công cụ Aggregate từ Toolbox để lọc ra những thuộc tính cần thiết cho bảng Dim_Shippings



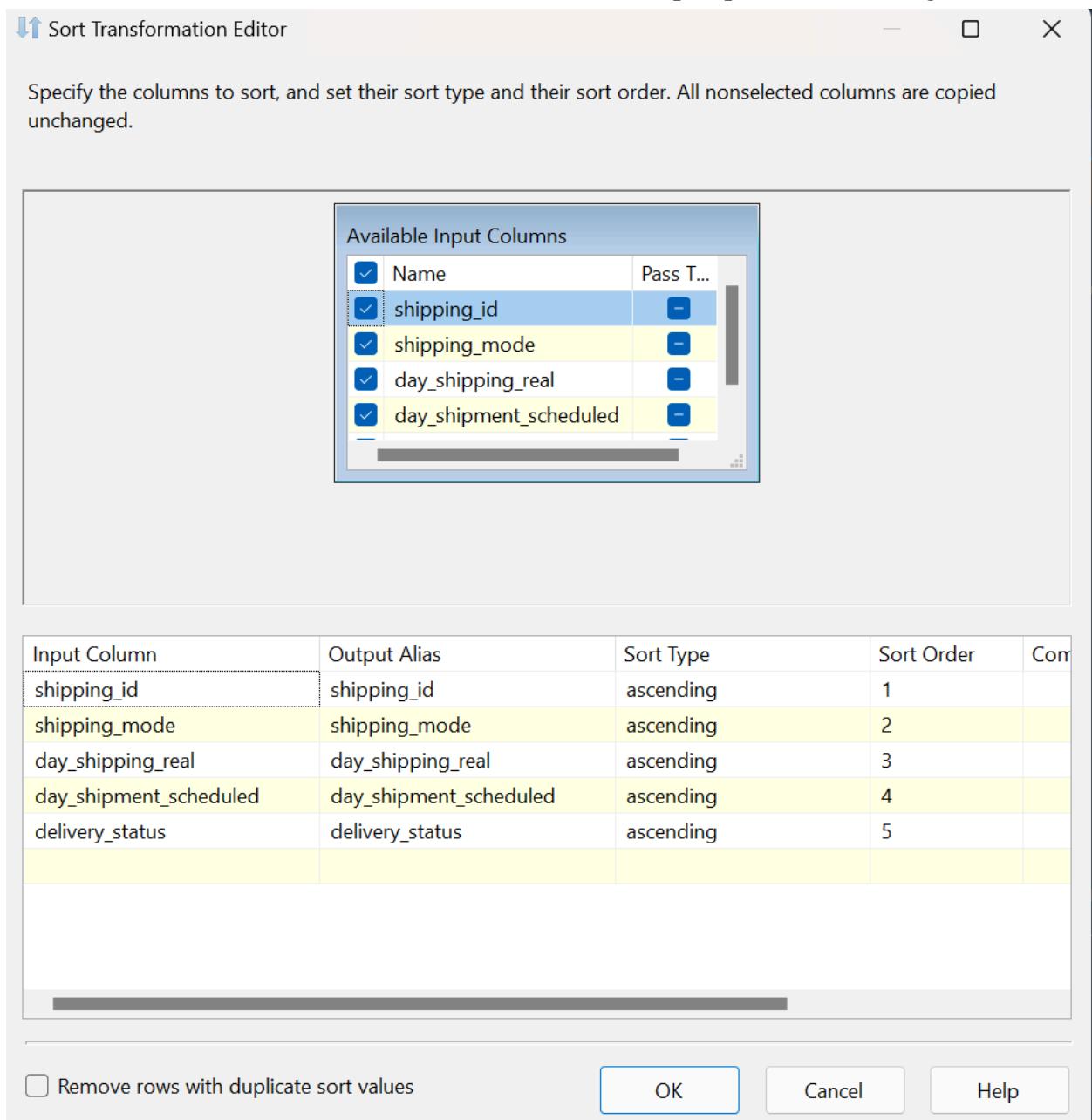
- Sau đó edit Aggregate, tích chọn các thuộc tính cần lọc trong bảng Dim_Shippings.



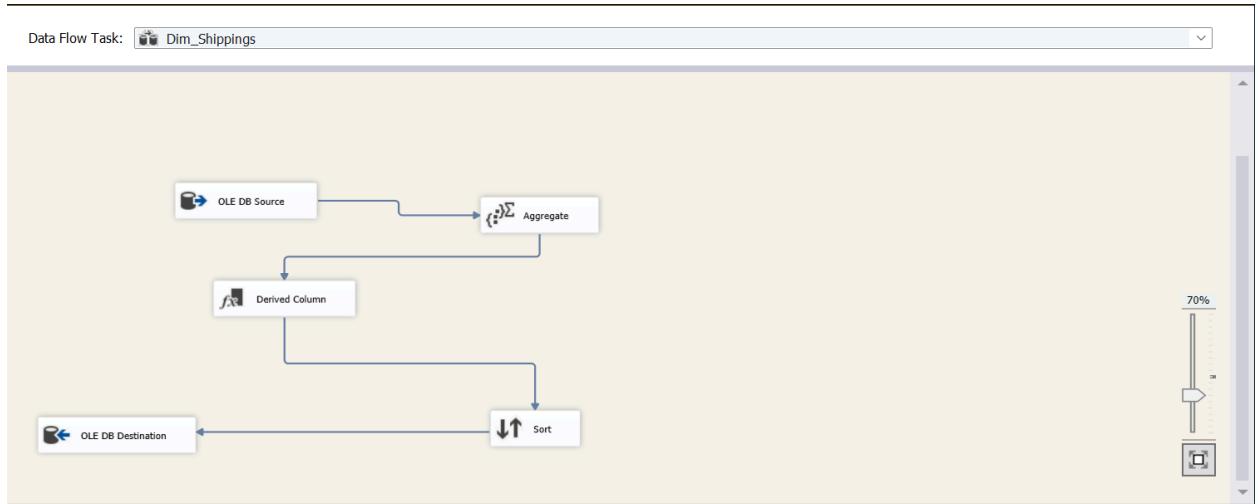
- Thực hiện kéo thả công cụ Sort từ Toolbox để sắp xếp lại dữ liệu cho bảng Dim_Shippings.



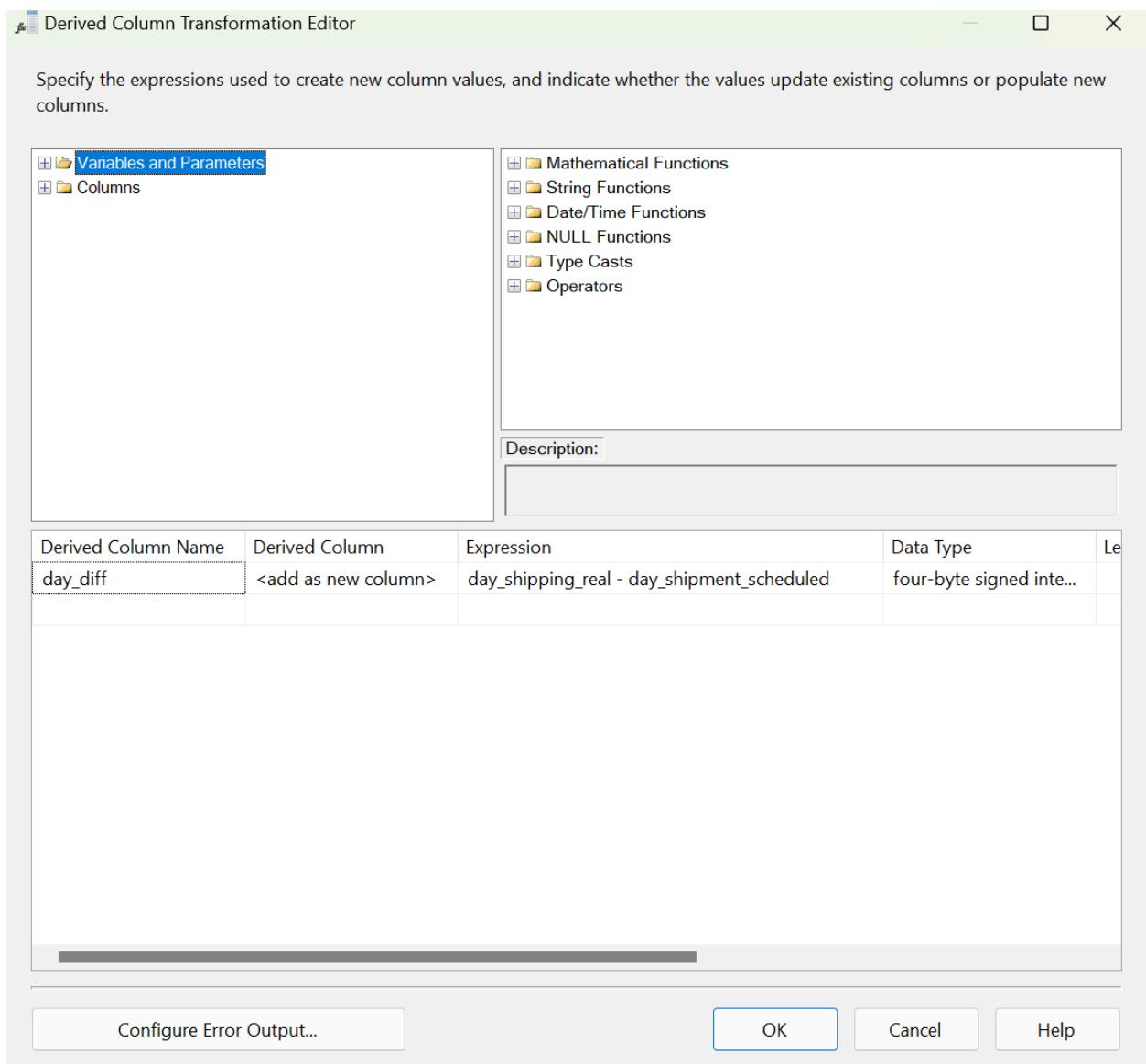
- Thực hiện edit Sort, chọn tích vào các thuộc tính cần sắp xếp theo thứ tự tăng dần



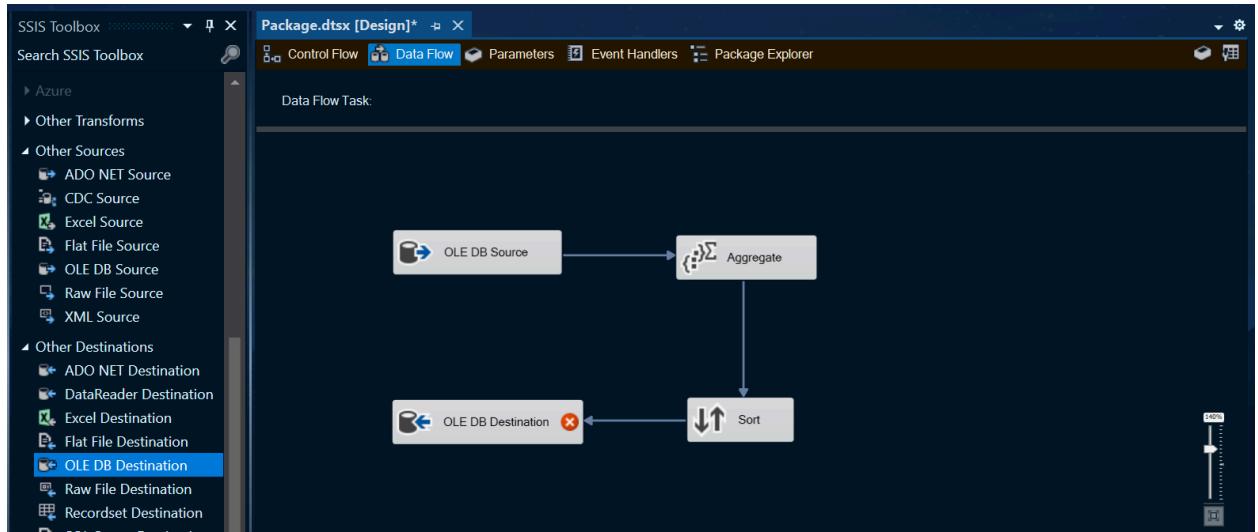
- Kéo thả công cụ và tạo thêm cột day_diff = day_shipping_real - day_shipment_scheduled



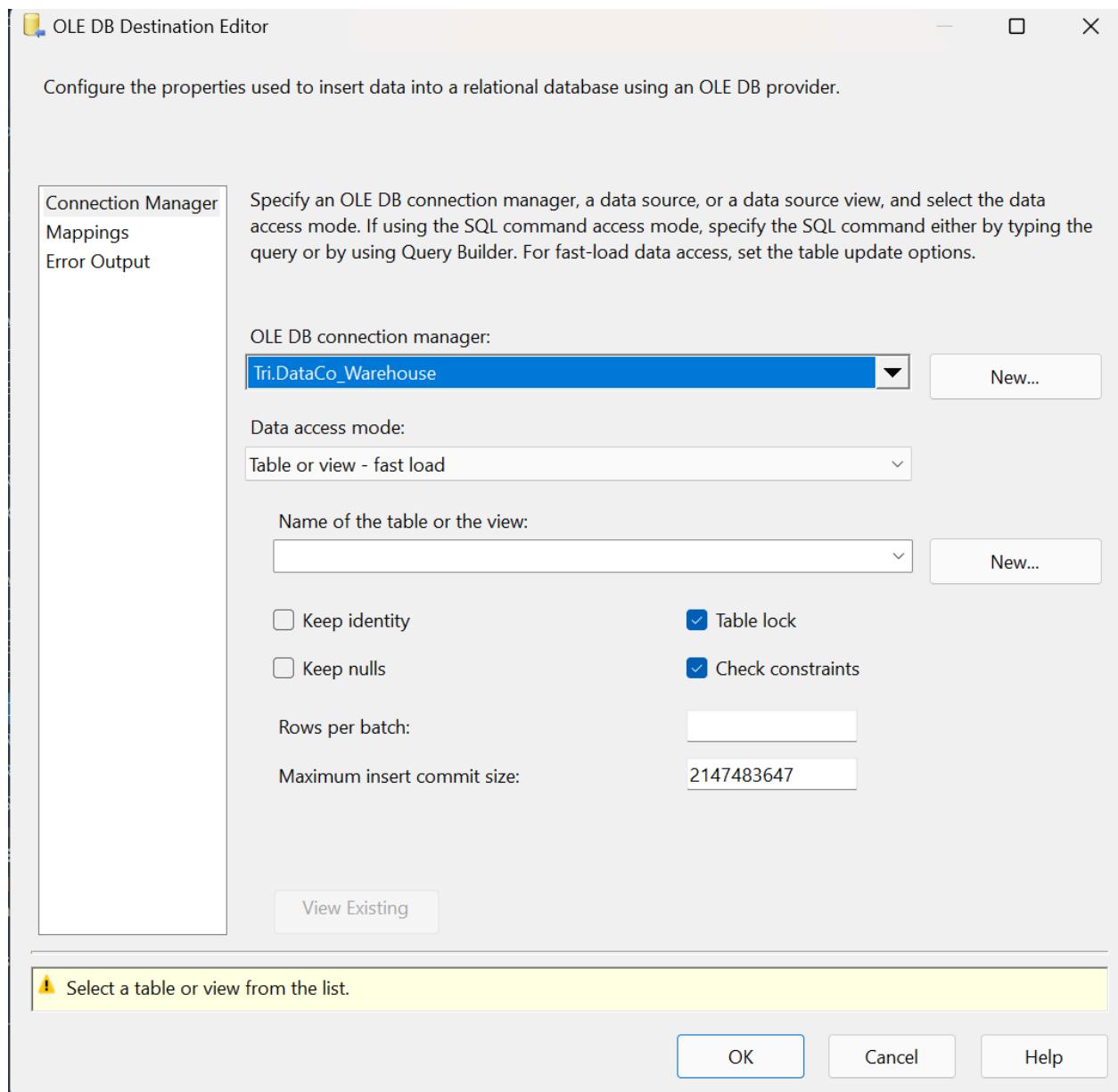
- Thực hiện edit công cụ Derived Column



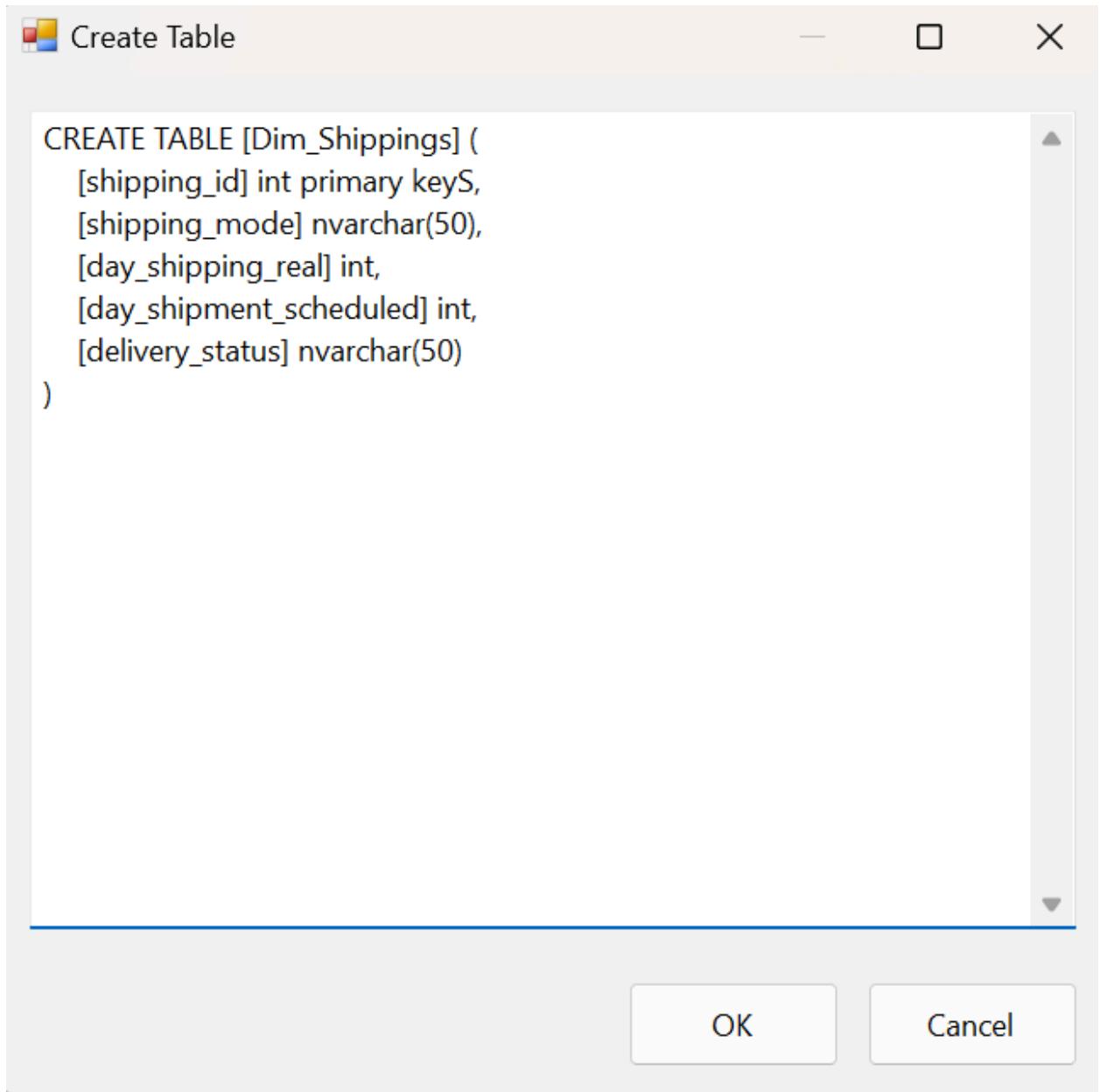
- Kéo thả công cụ OLE DB Destination từ Toolbox.



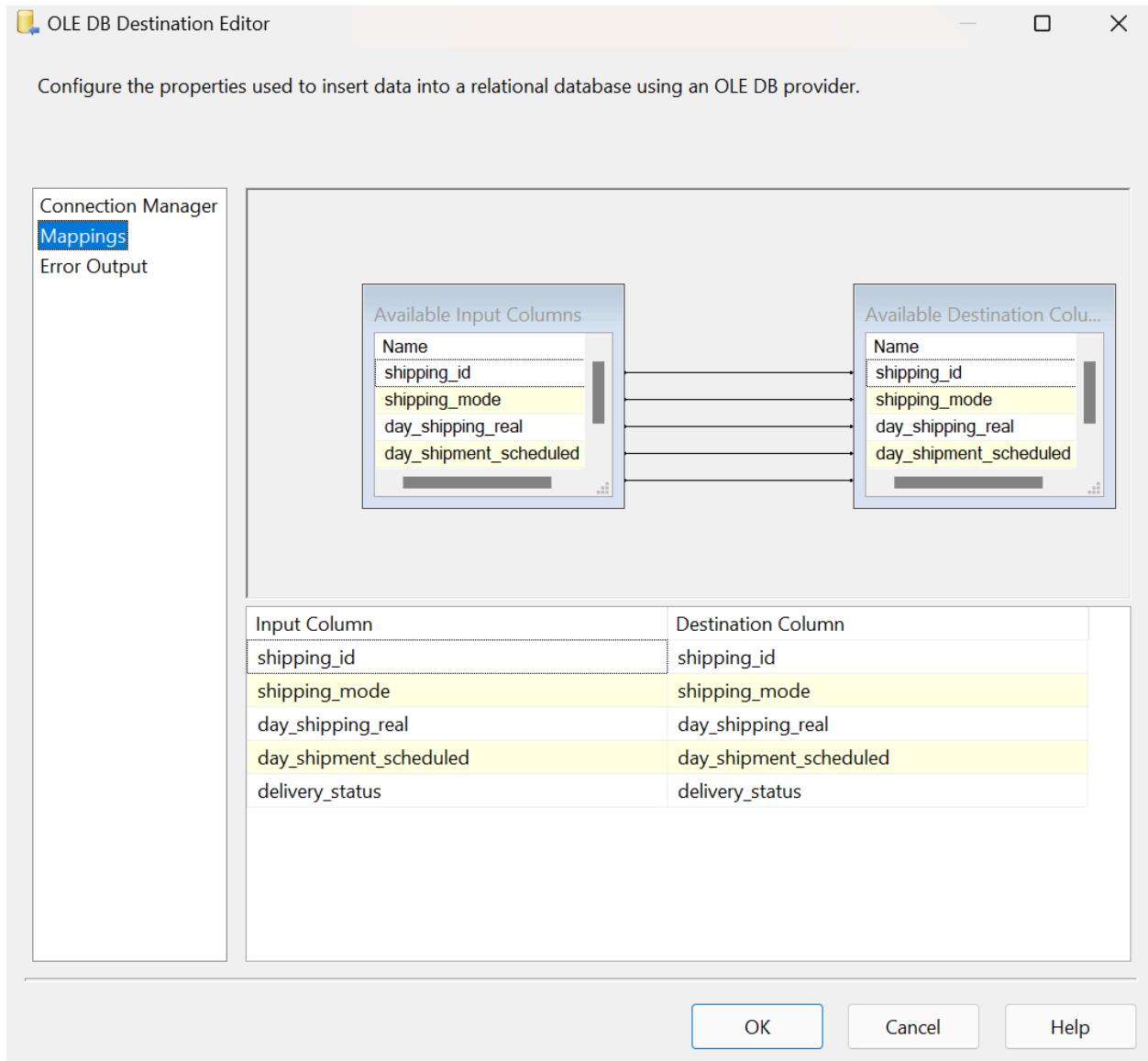
- Thực hiện edit công cụ OLE DB Source, kết nối tới database DataCo_Warehouse.



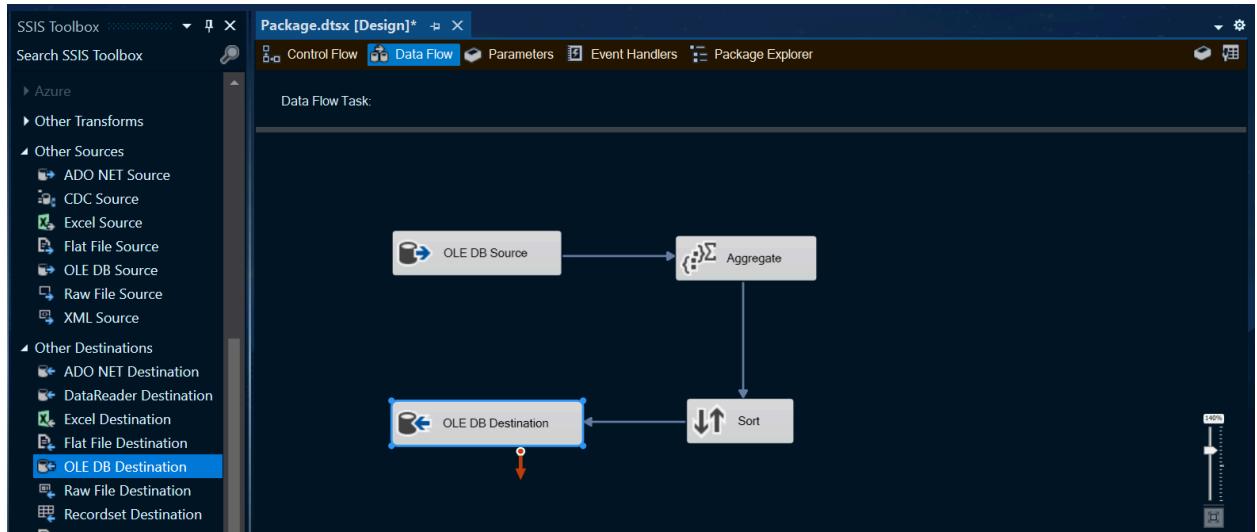
- Nhấn New để tạo bảng Dim_Shippings trong DataCo_Warehouse.



- Chọn tab Mapping và map những thuộc tính cần thiết cho bảng Dim_Shippings.

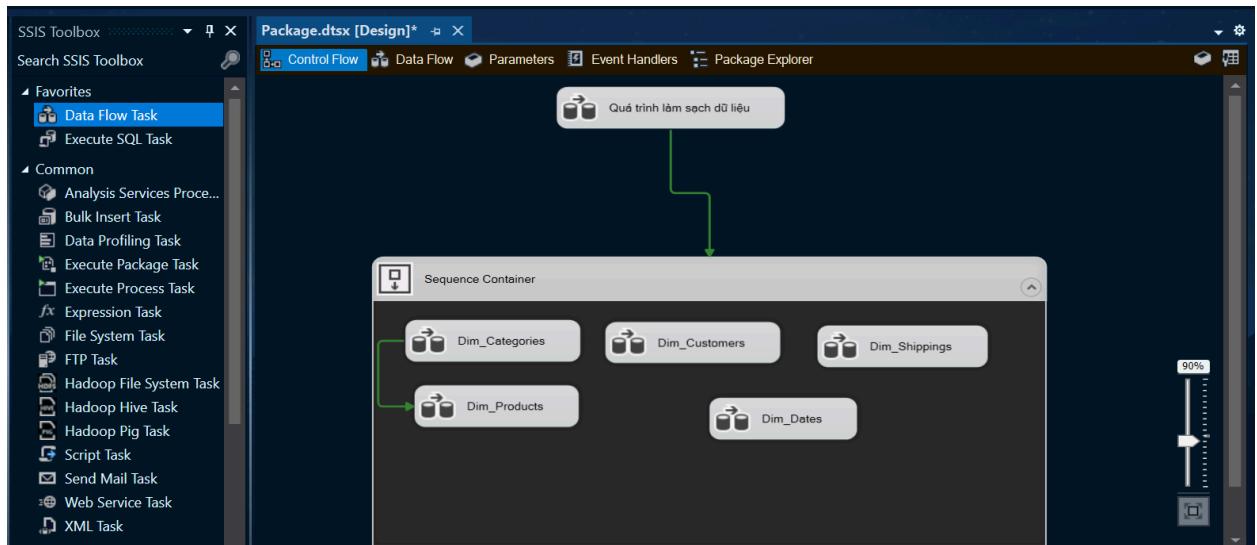


- Data Flow Task Dim_Shippings khi hoàn thành.

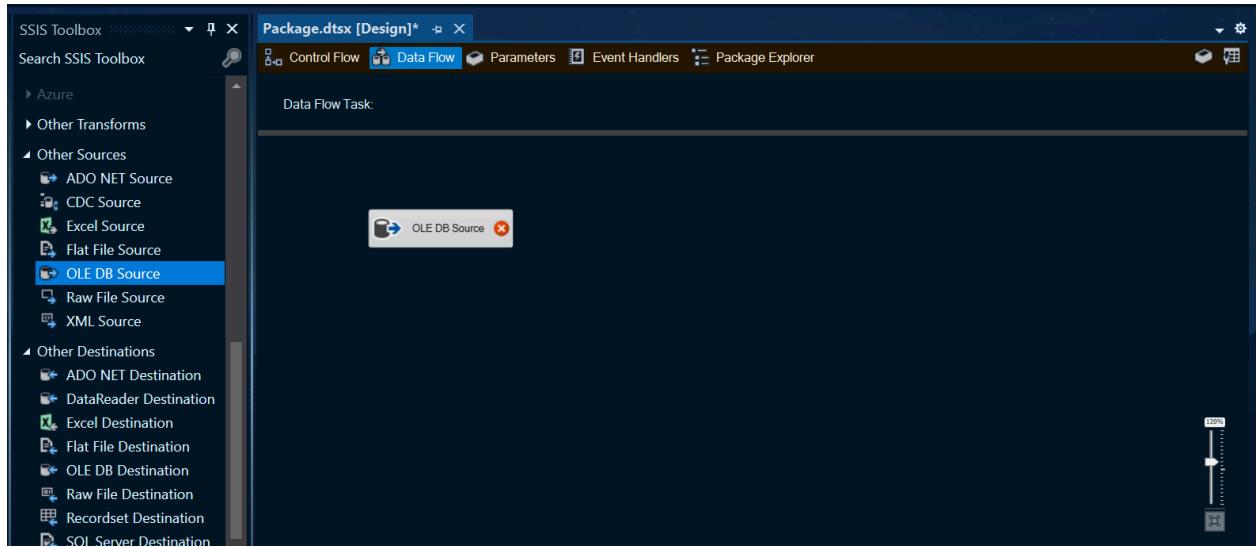


2.3.5. Tạo bảng Dim_Dates.

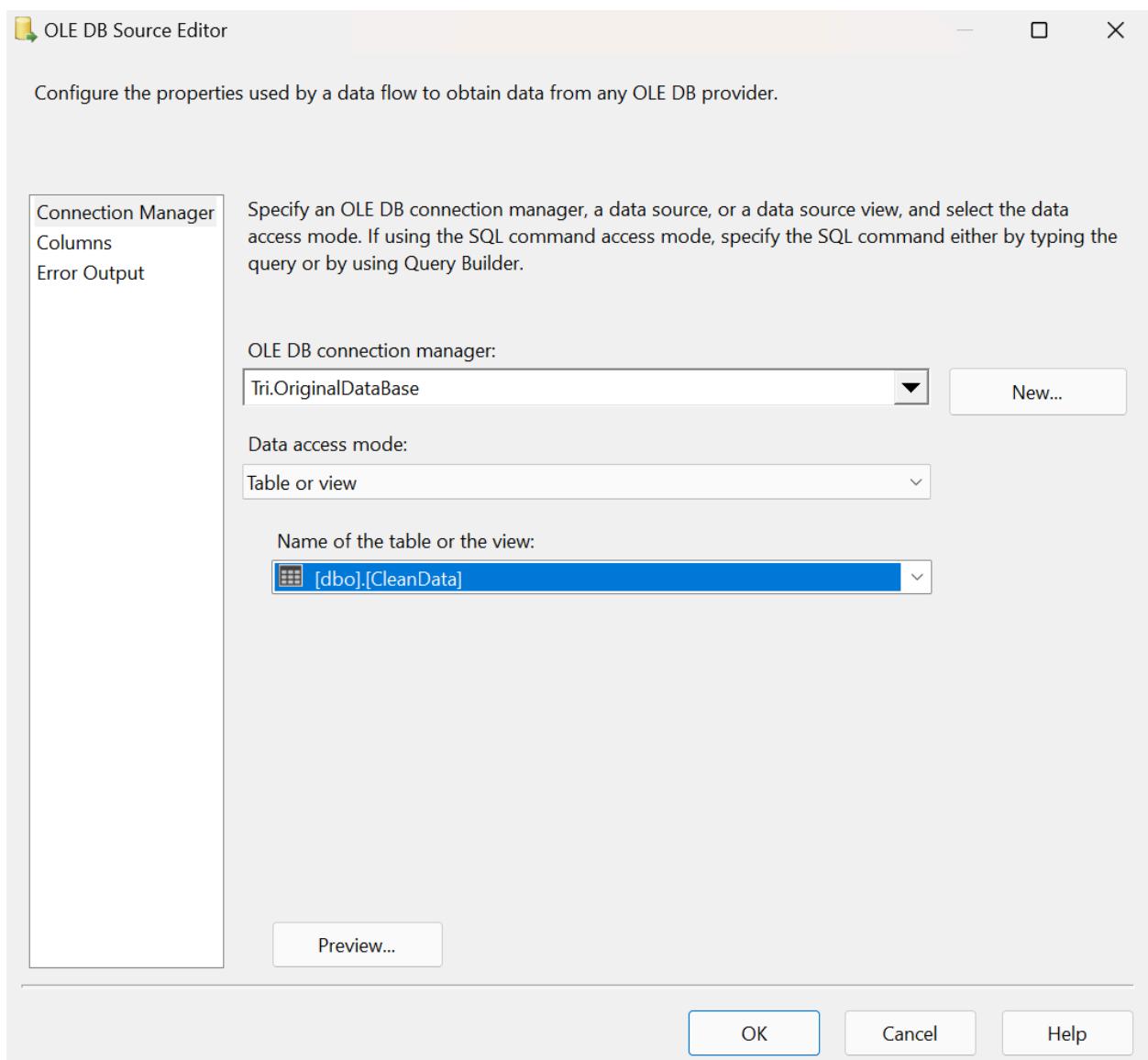
- Kéo thả công cụ Data Flow Task từ Toolbox, sau đó tiến hành đổi tên thành “Dim_Dates”



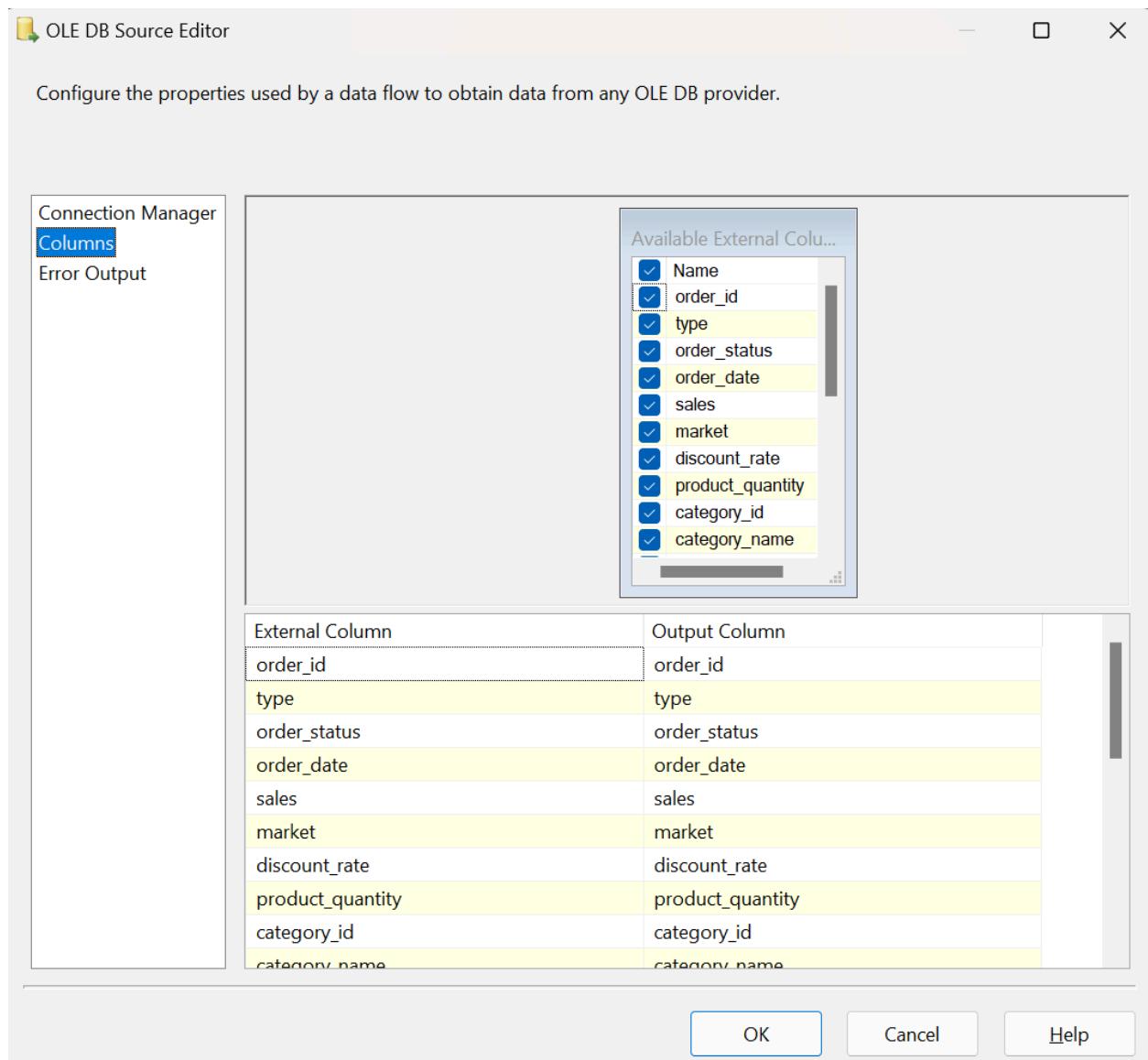
- Edit cho Dim_Dates, kéo thả công OLE DB Source từ Toolbox.



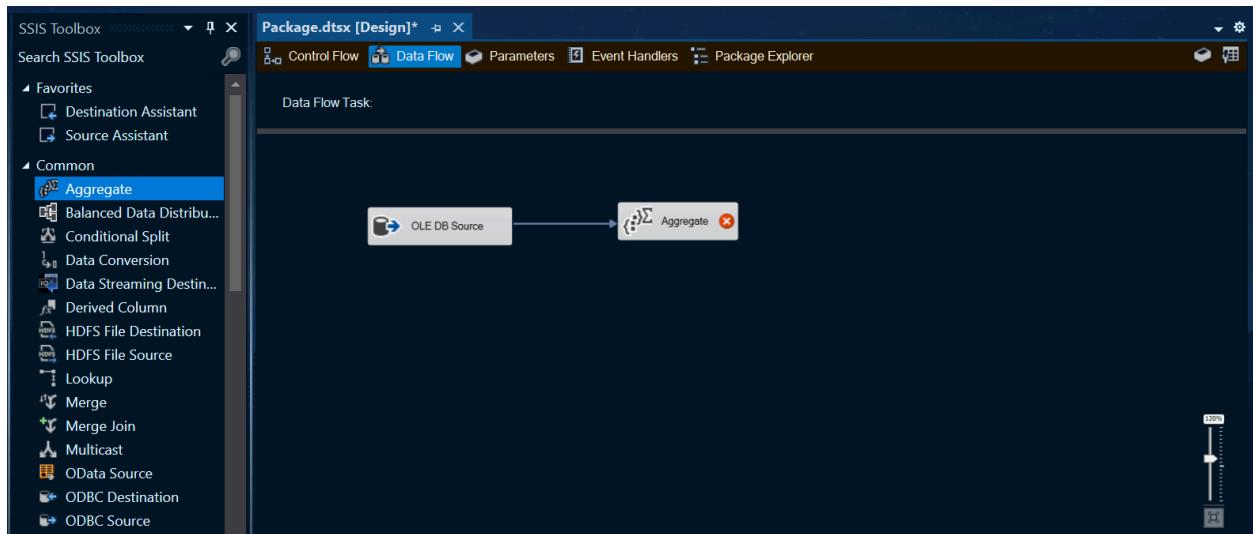
- Thực hiện edit OLE DB Source, chọn kết nối tới bảng CleanData trong database OriginalDatabase.



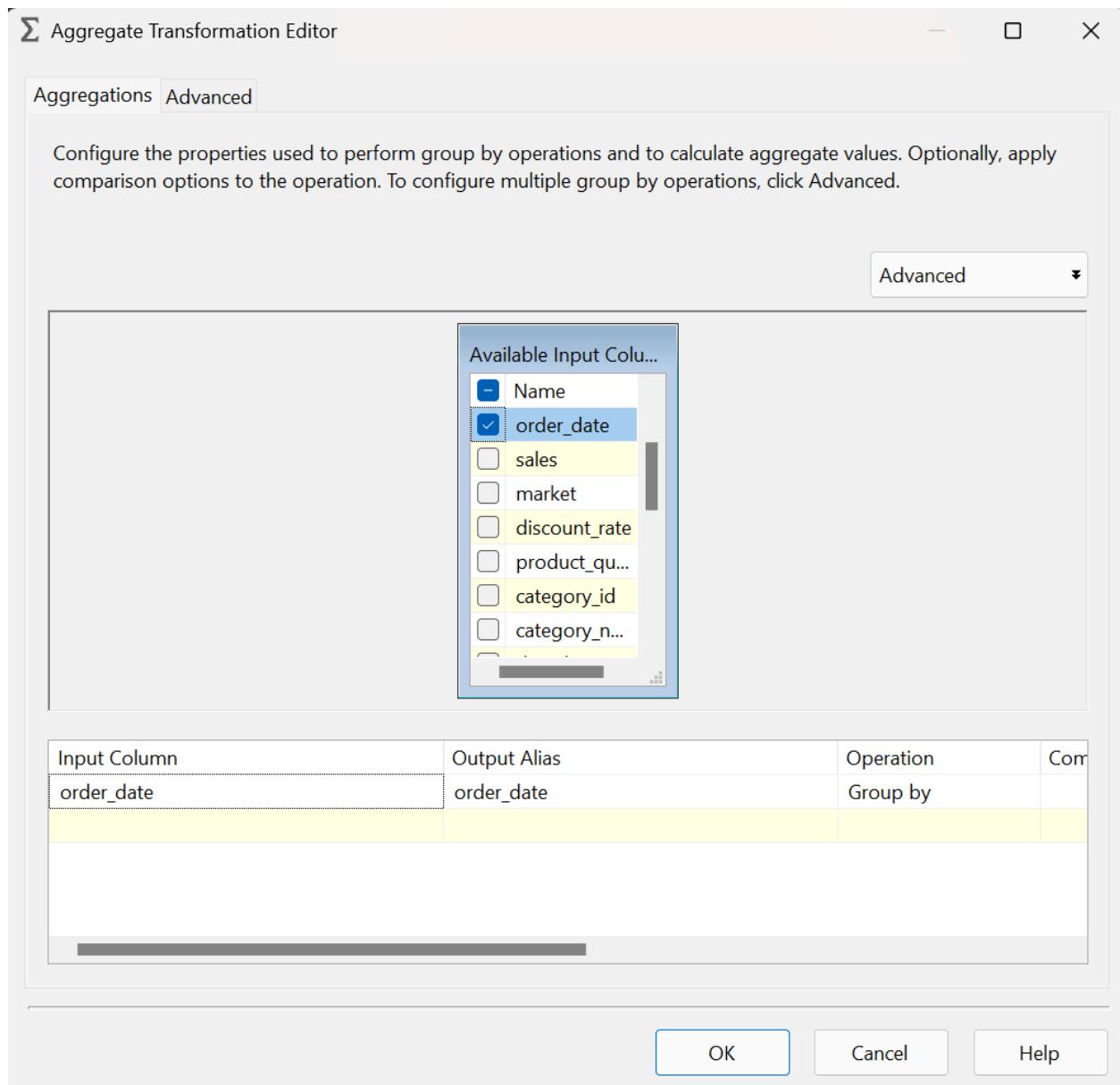
- Chuyển sang tab Columns và tích chọn thuộc tính cần thiết cho bảng Dim_Dates.



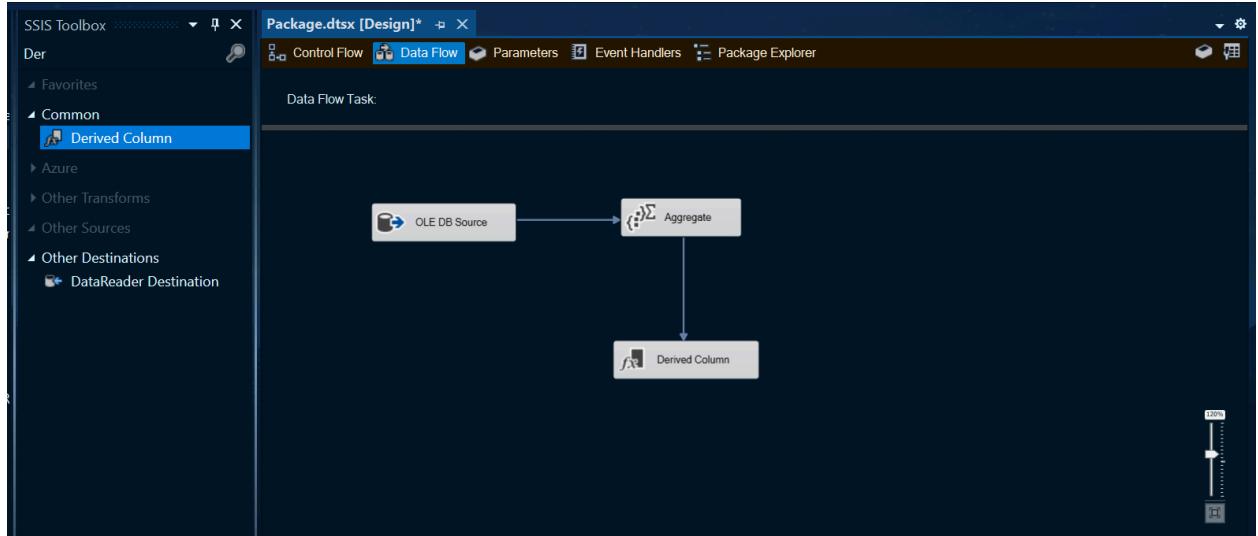
- Kéo thả công cụ Aggregate từ Toolbox để lọc dữ liệu cho bảng Dim_Dates.



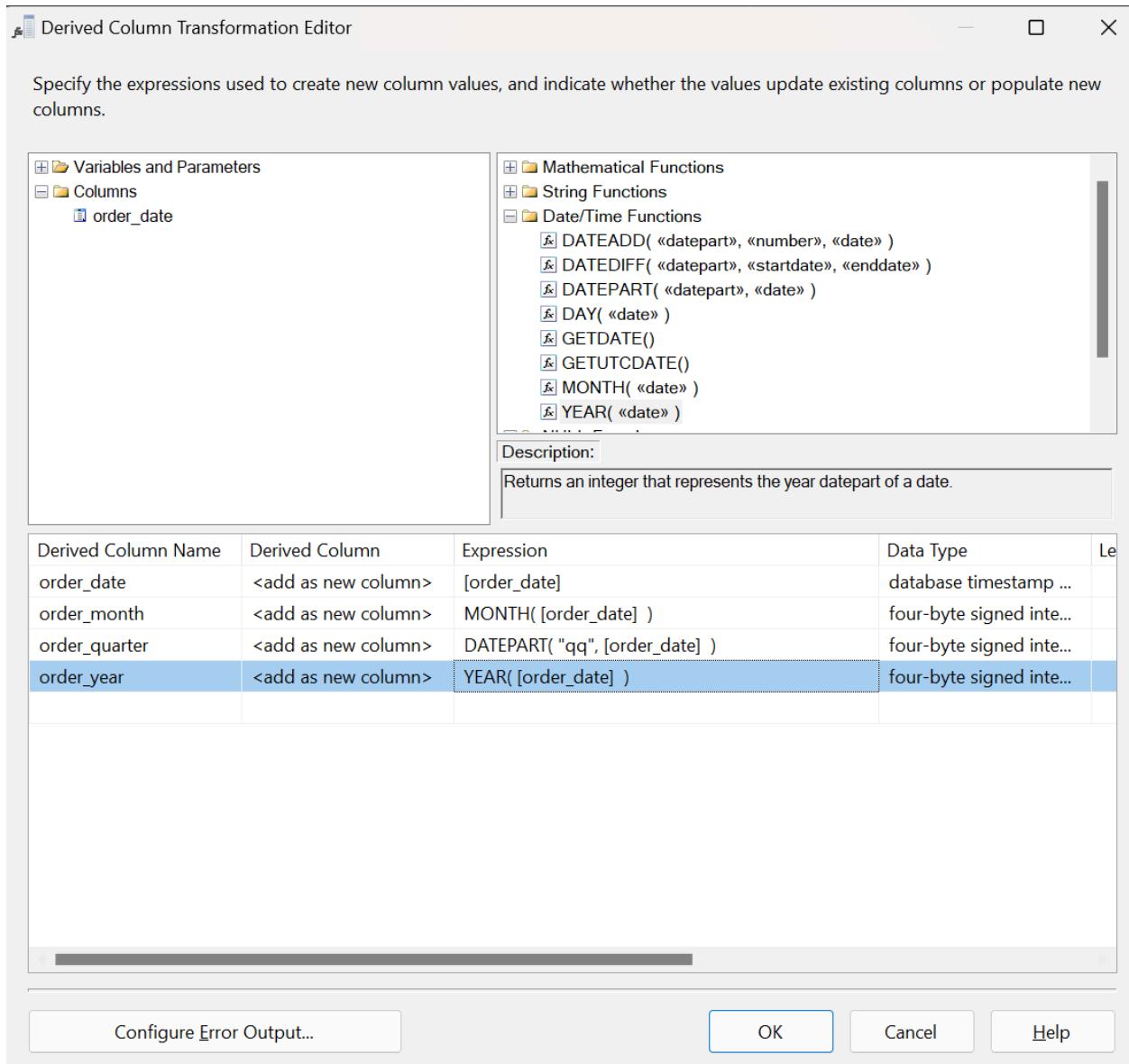
- Thực hiện edit Aggregate, tích chọn thuộc tính cần thiết cho bảng Dim_Dates.



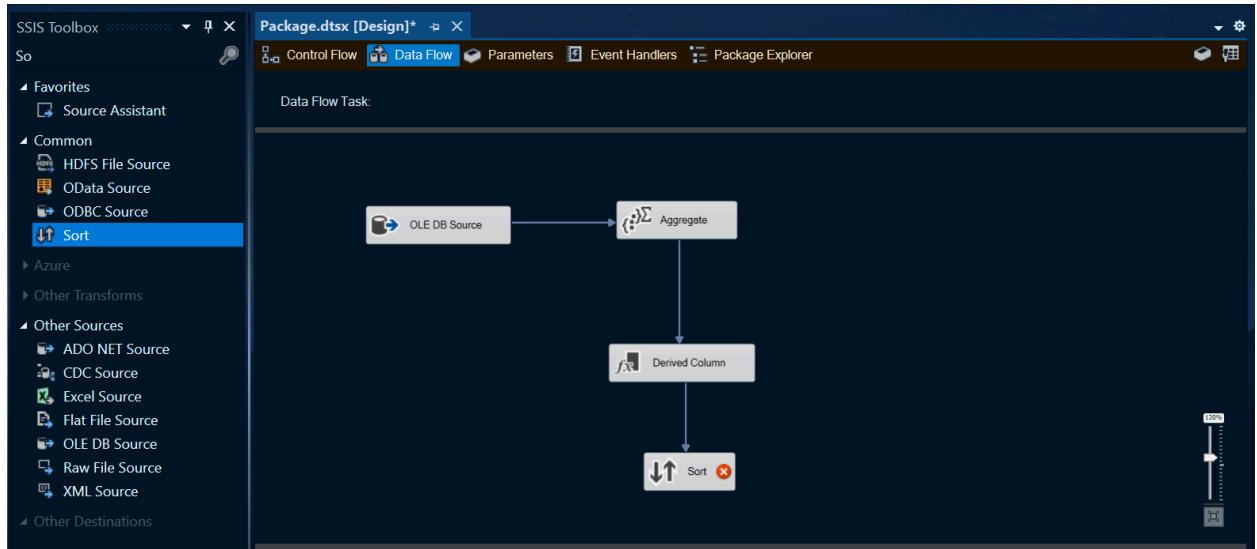
- Kéo thả công cụ Derived Column để tách cột order_date thành các thuộc tính cần cho bảng Dim_Dates.



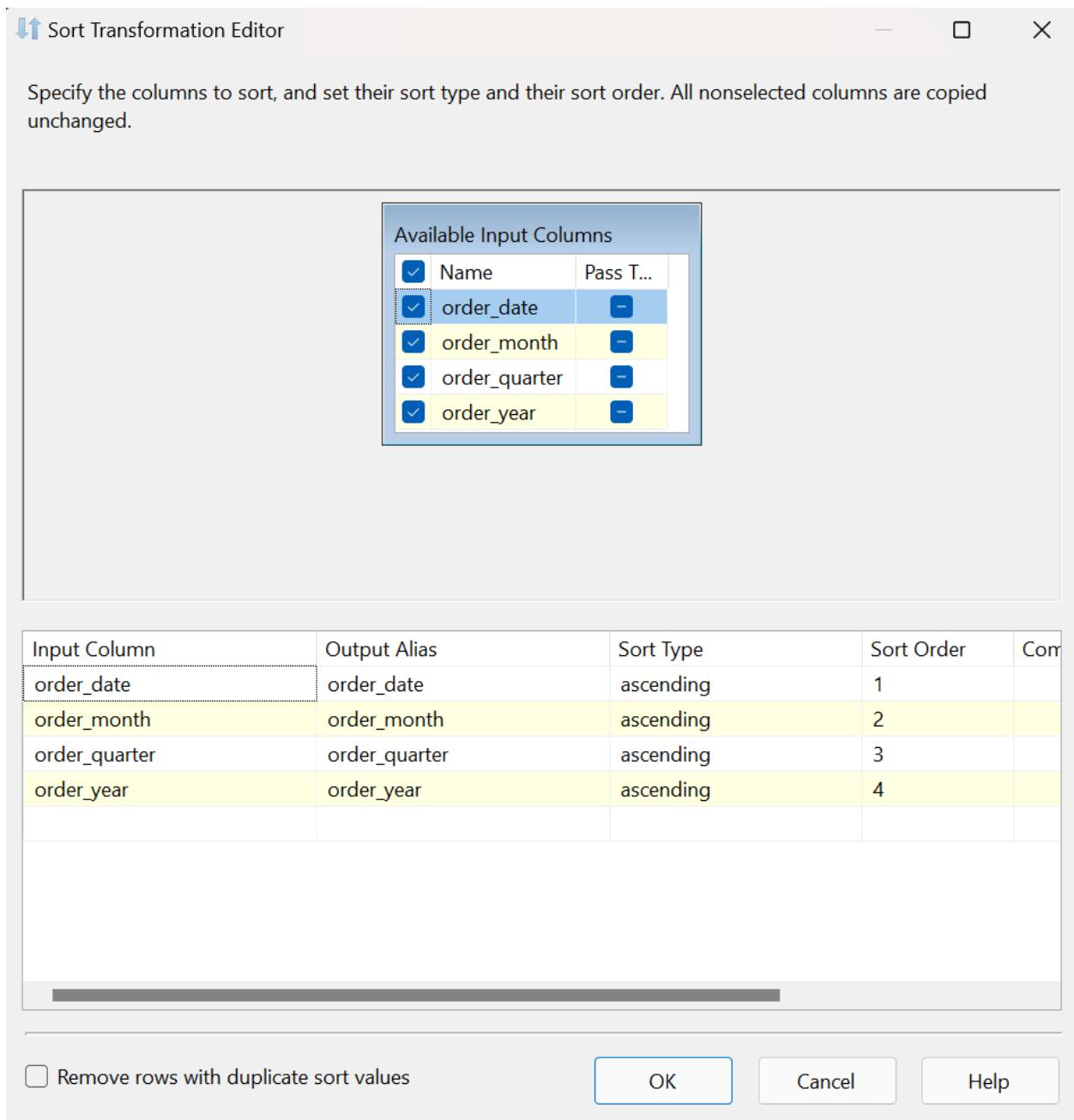
- Thực hiện xây dựng các thuộc tính cần thiết cho Dim_Dates từ thuộc tính order_date, sử dụng các hàm được xây dựng sẵn.



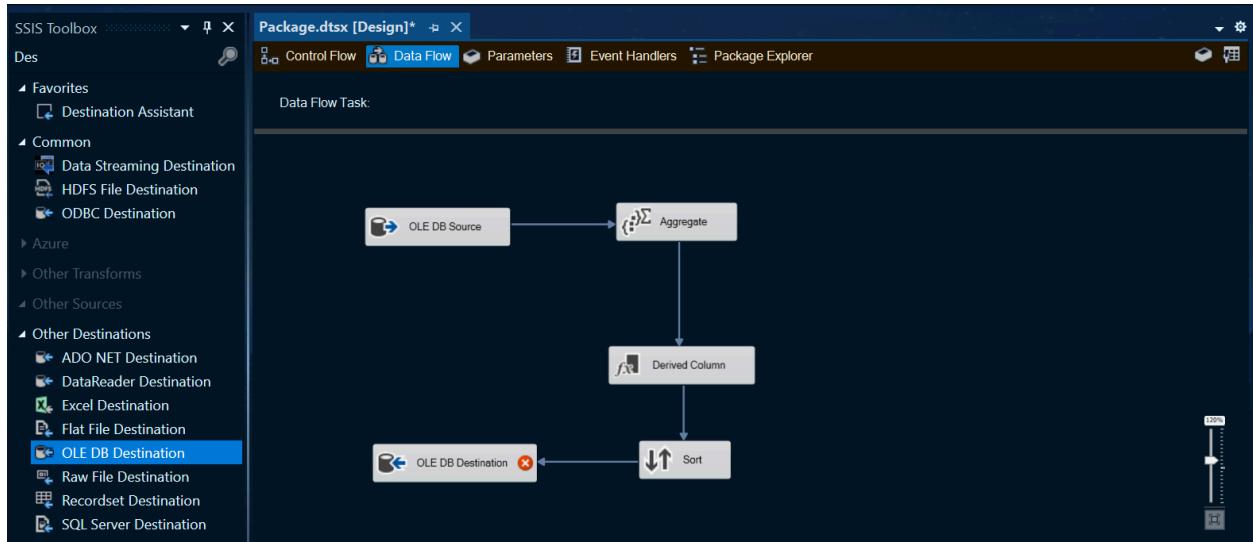
- Kéo thả công cụ Sort từ Toolbox để sắp xếp lại dữ liệu cho bảng Dim_Dates.



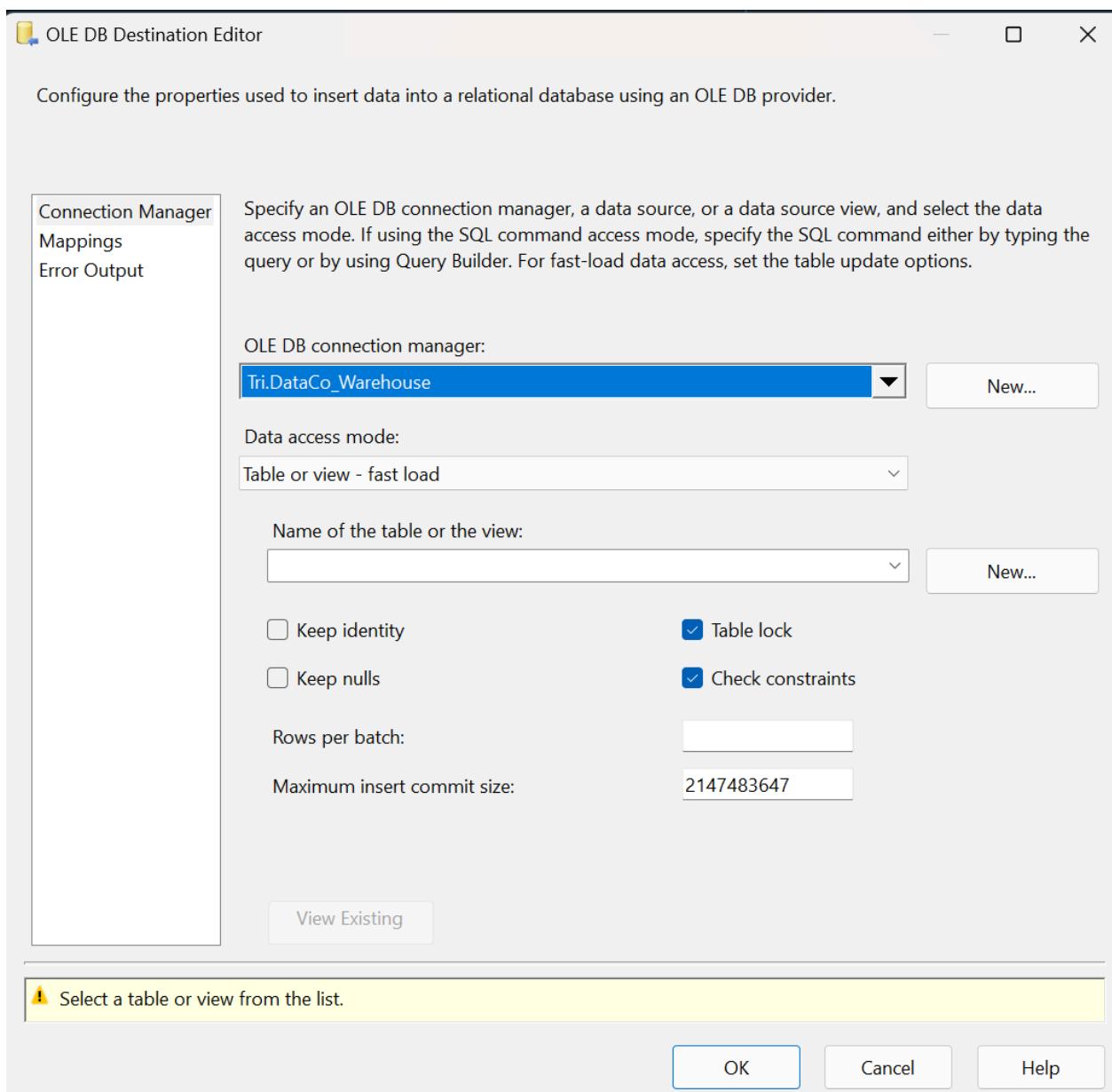
- Thực hiện edit Sort, tích chọn các thuộc tính cần sắp xếp tăng dần.



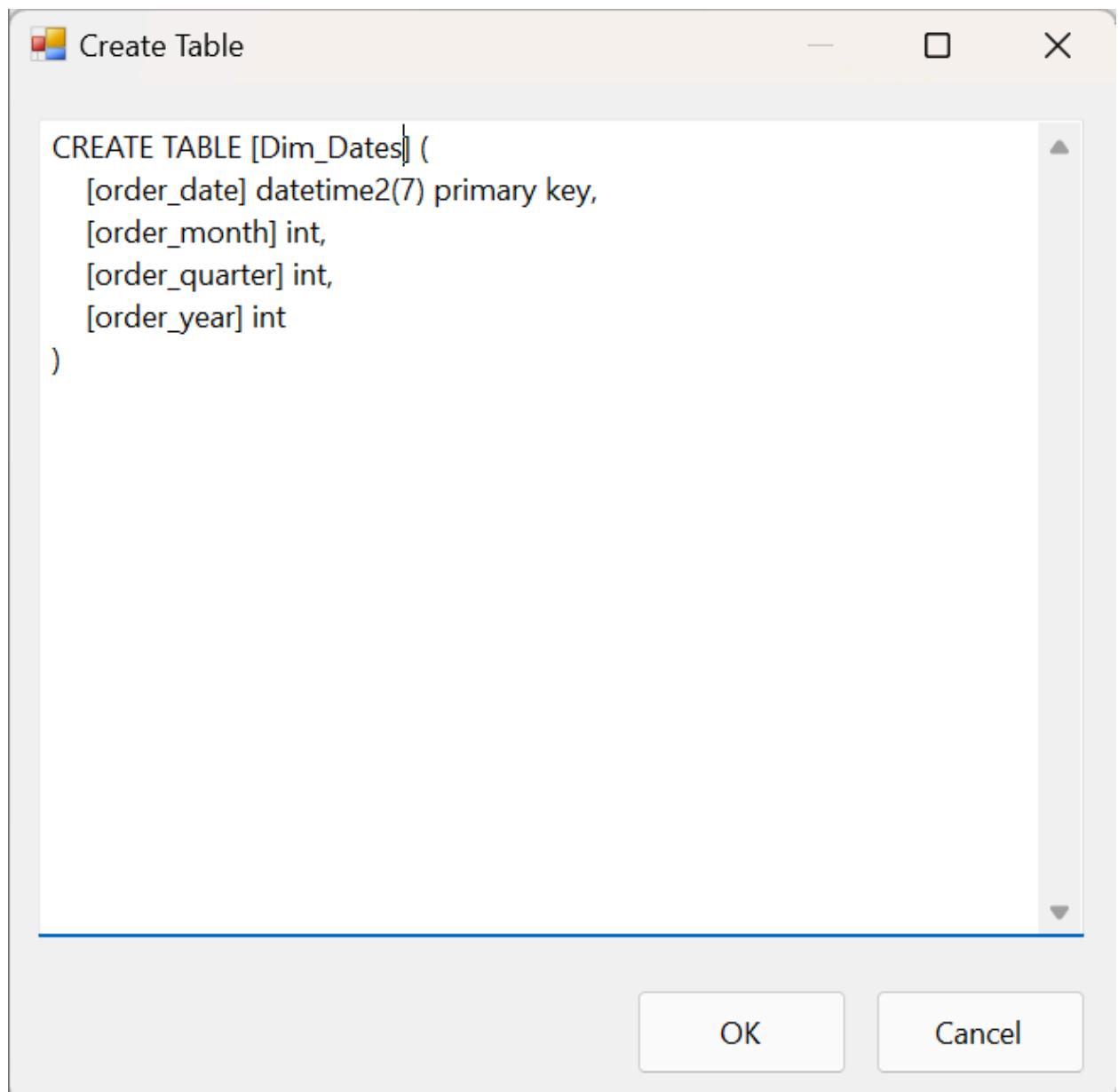
- Thực hiện kéo thả công cụ OLE DB Destination từ Toolbox



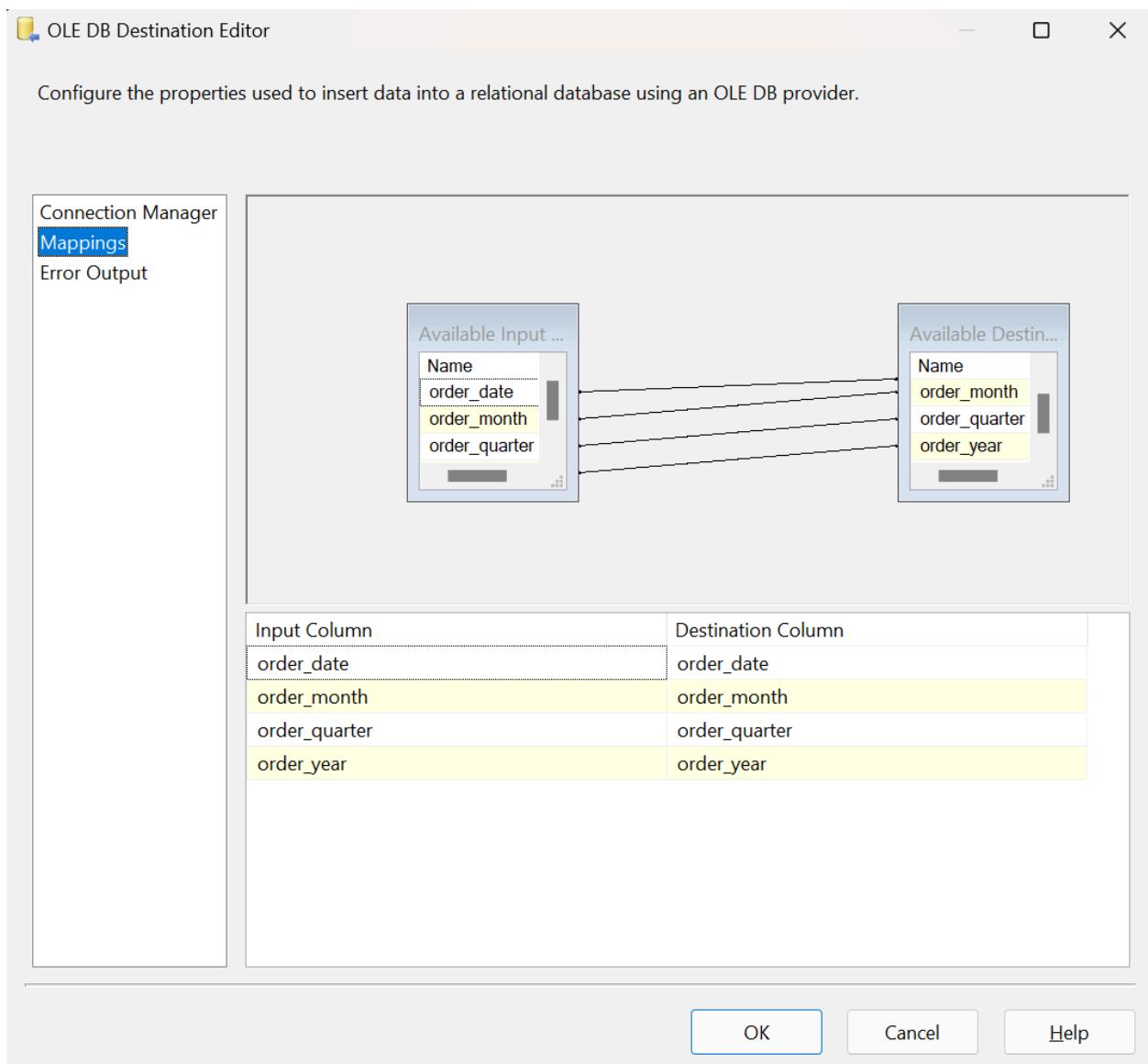
- Sau đó edit OLE DB Source, thực hiện kết nối tới DataCo_Warehouse



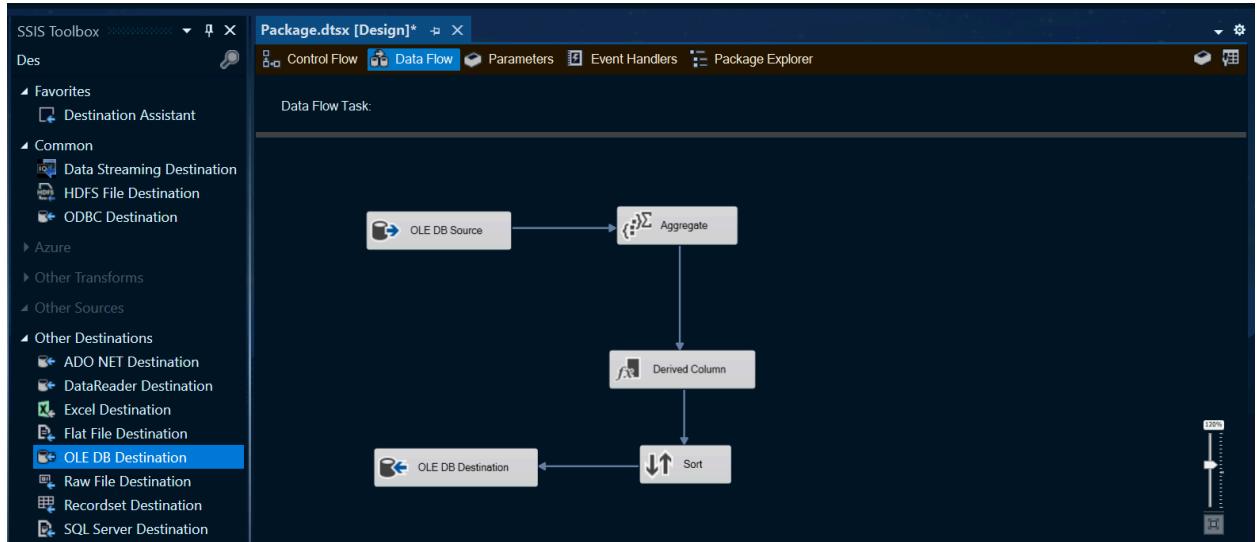
- Chọn New để tạo bảng Dim_Dates.



- Chuyển qua tab Mapping để map các thuộc tính cần thiết cho bảng Dim_Dates.

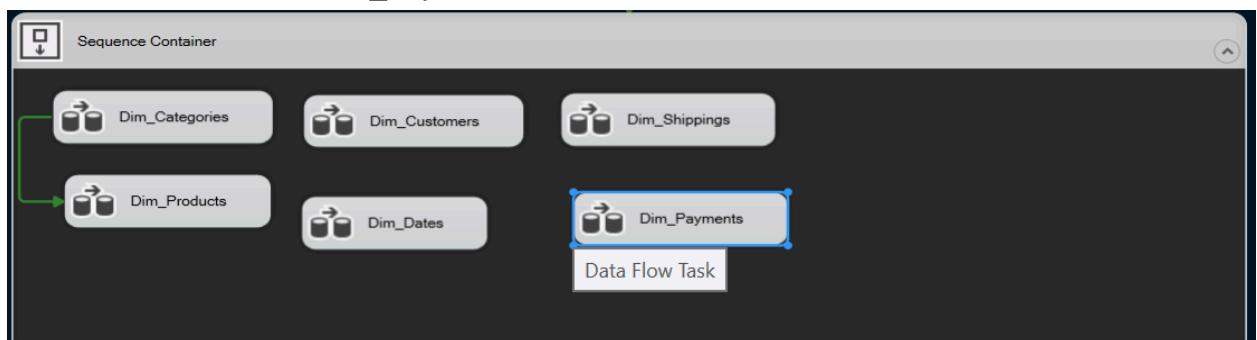


- Date Flow Task Dim_Dates khi hoàn thành.



2.3.6. Tạo bảng Dim_Payments

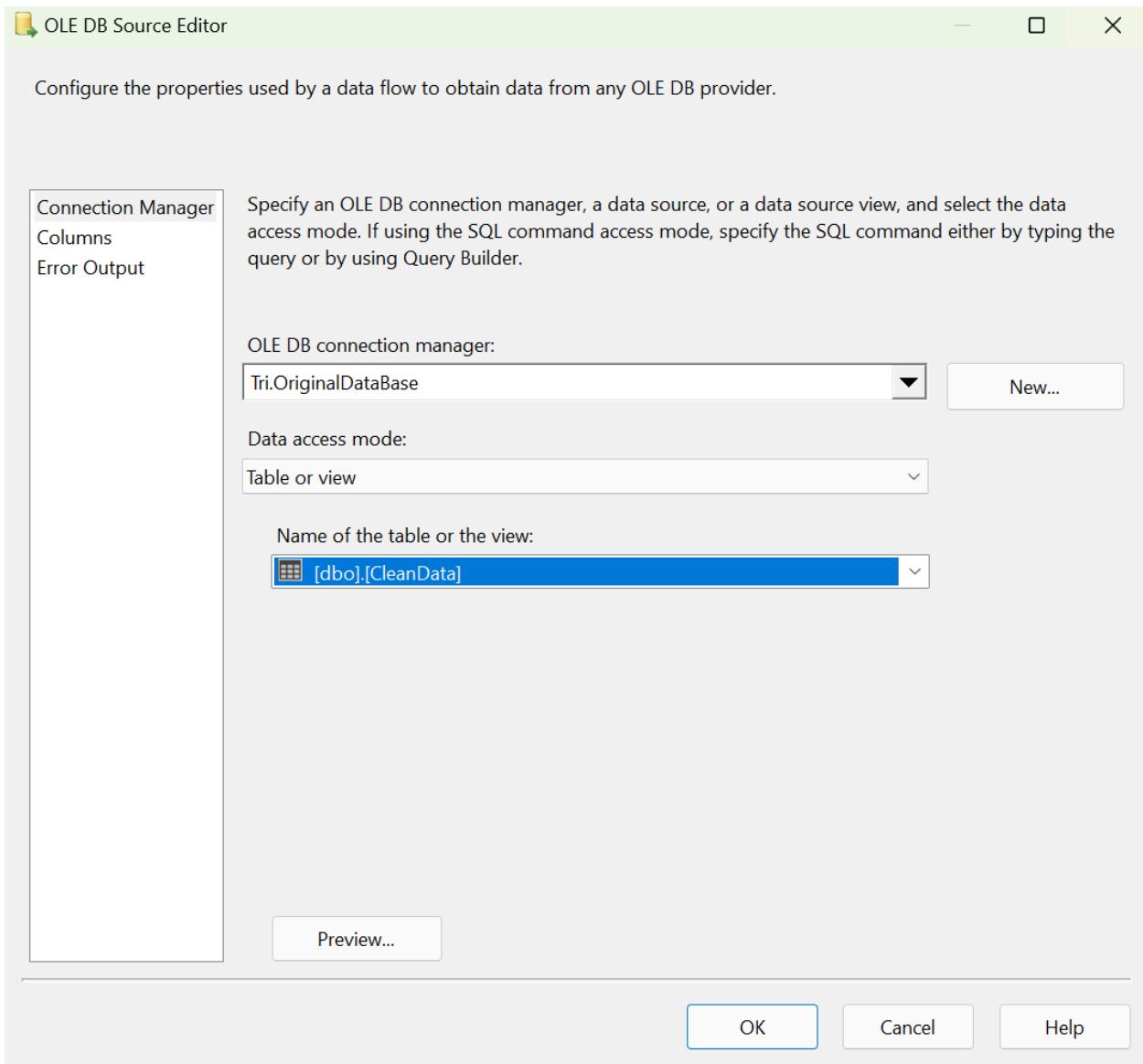
- Kéo thả tạo flow task Dim_Payments



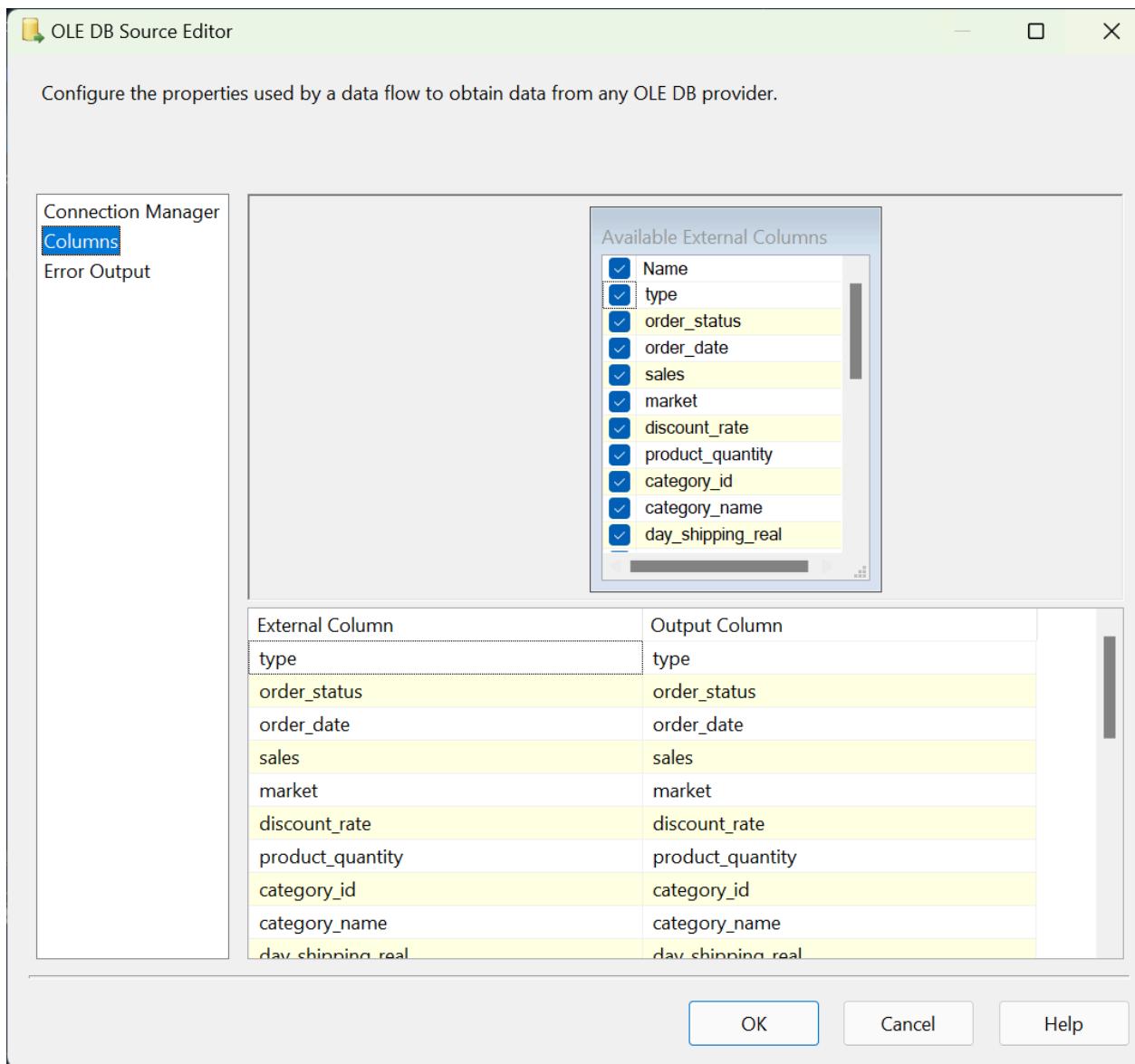
- Tạo OLE DBSource



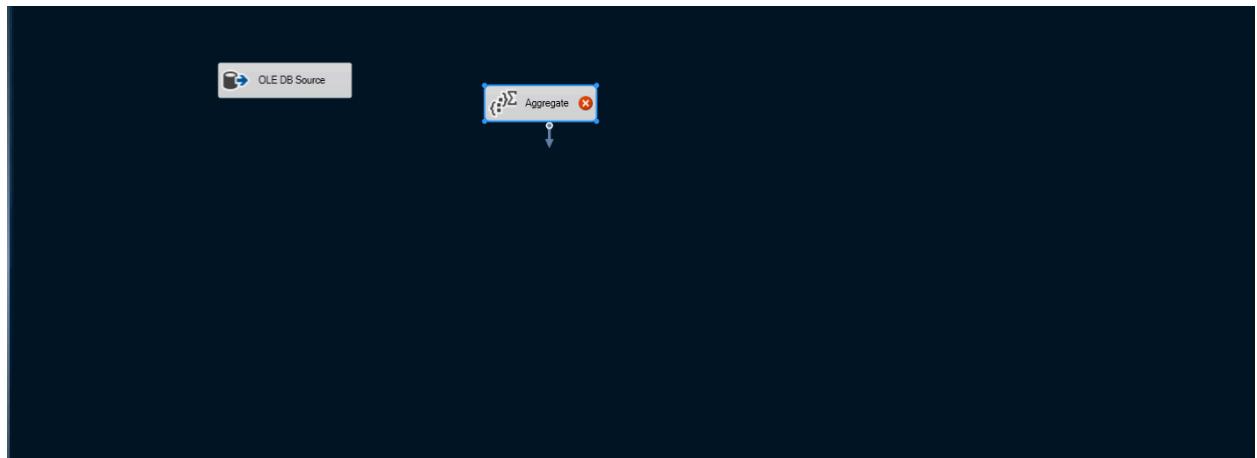
- Thực hiện Edit OLE DB Source, kết nối tới bảng Clean Data



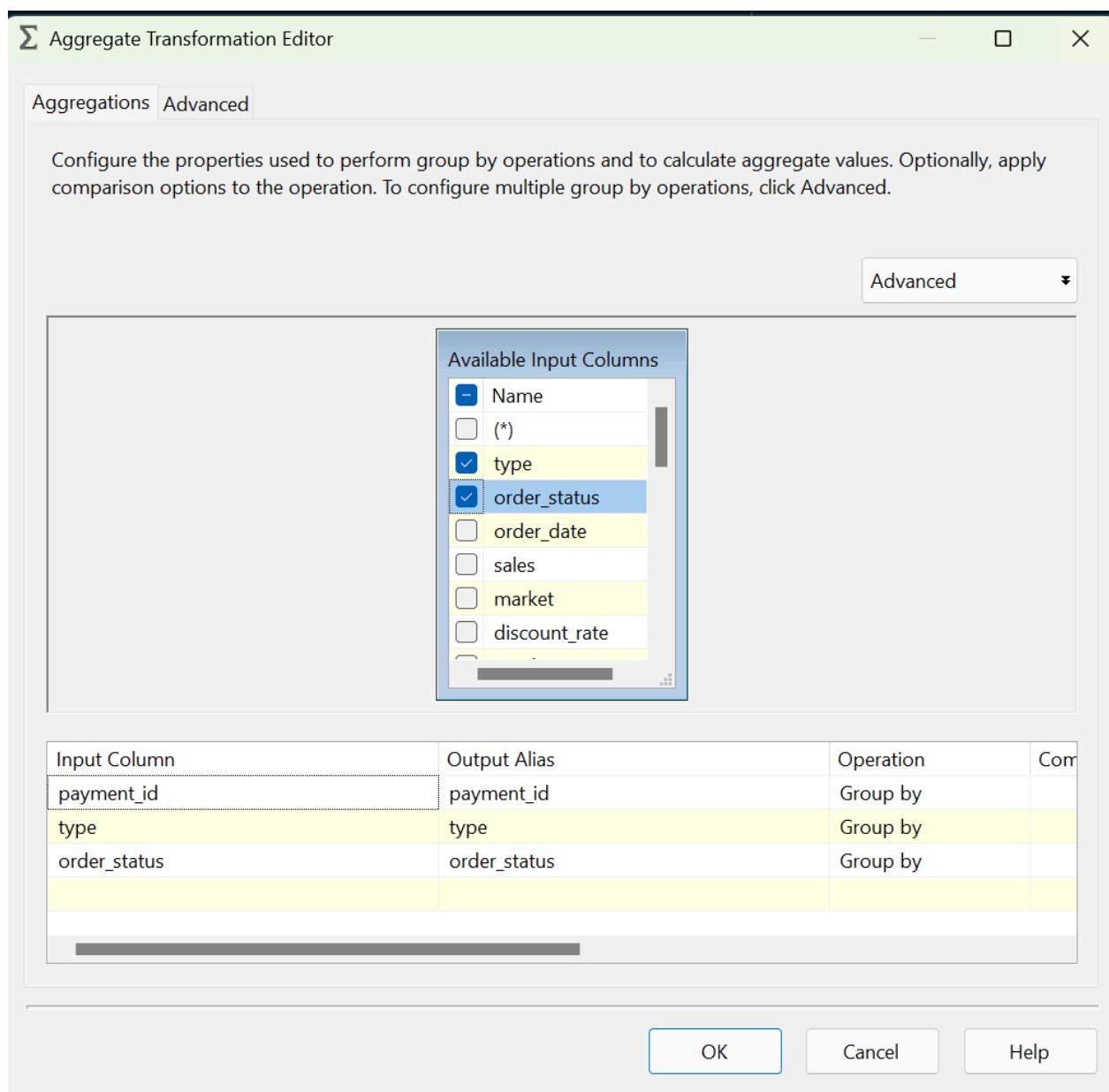
- Tích chọn các column cần thiết



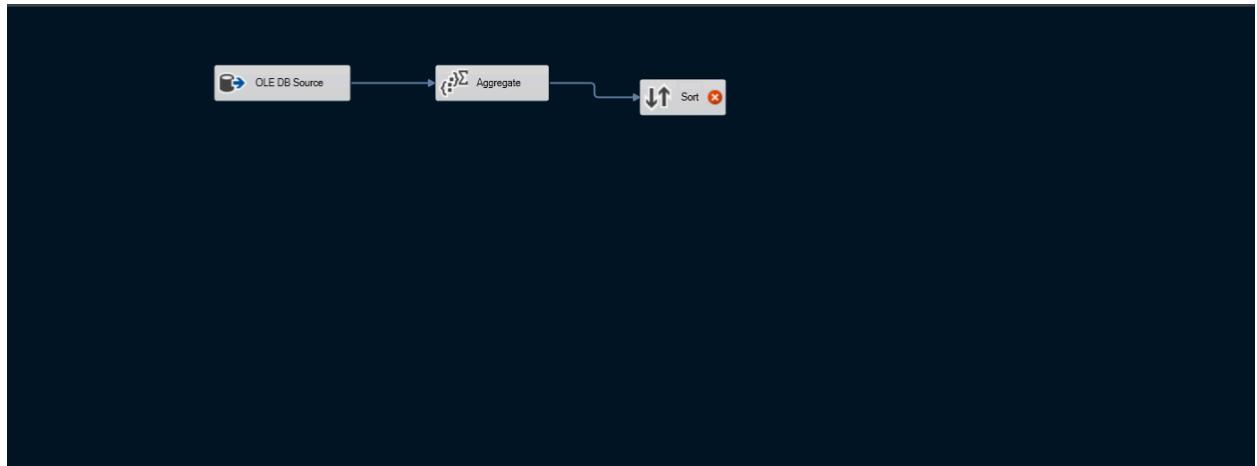
- Thực hiện kéo thả công cụ Aggregate



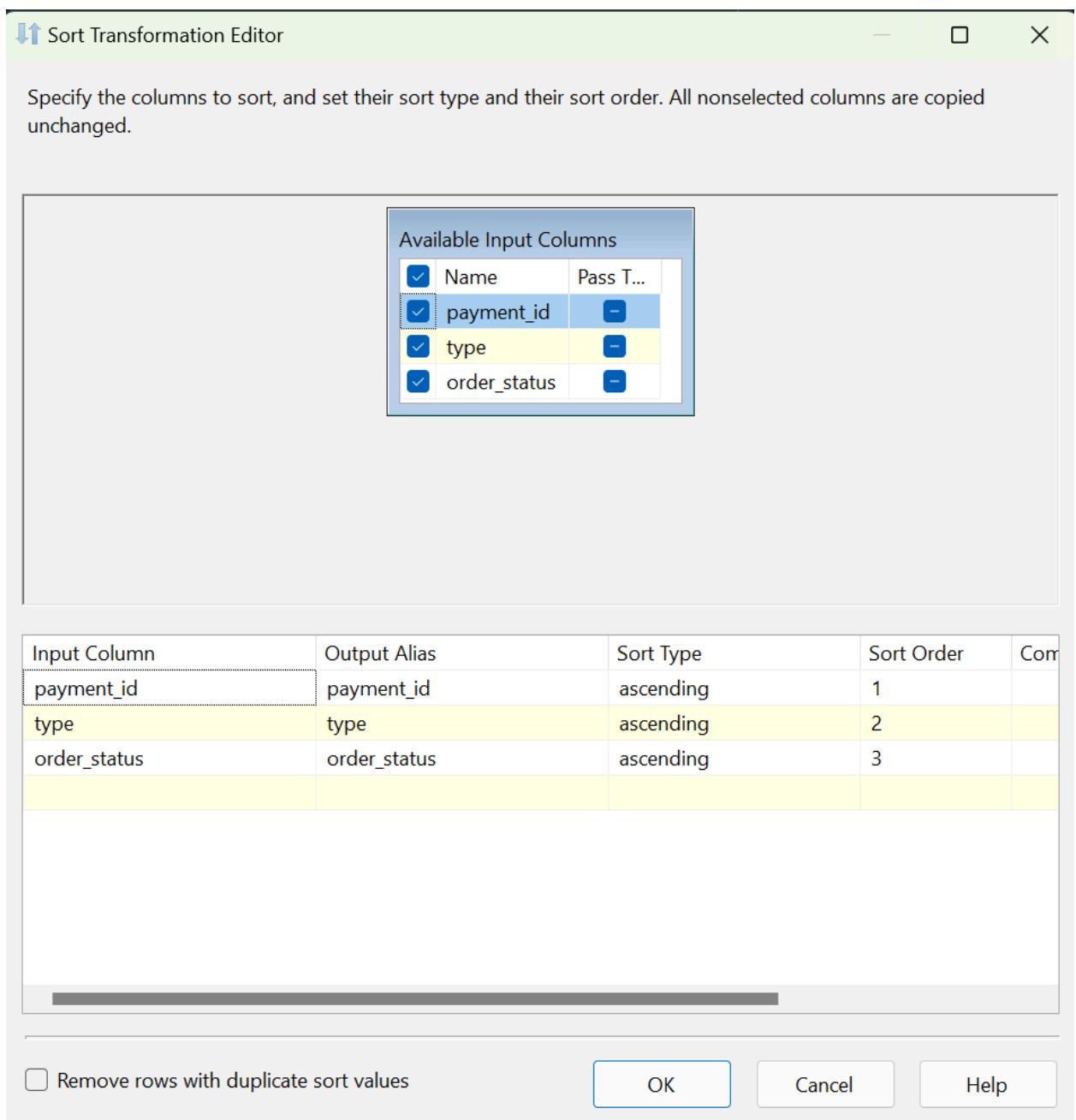
- Thực hiện edit Aggregate, tích chọn các thuộc tính cần thiết cho bảng Dim_Payments



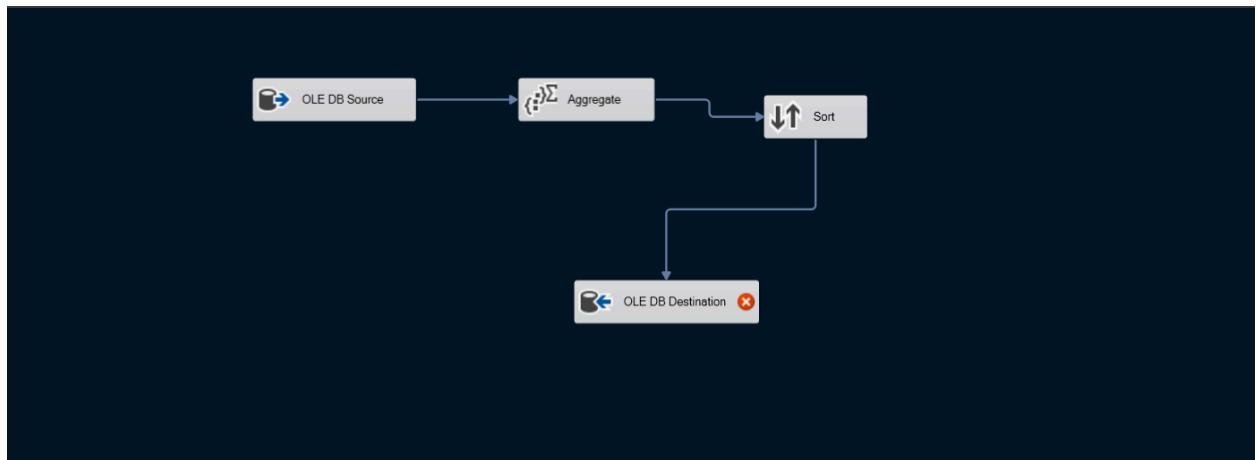
- Thực hiện kéo thả công cụ Sort



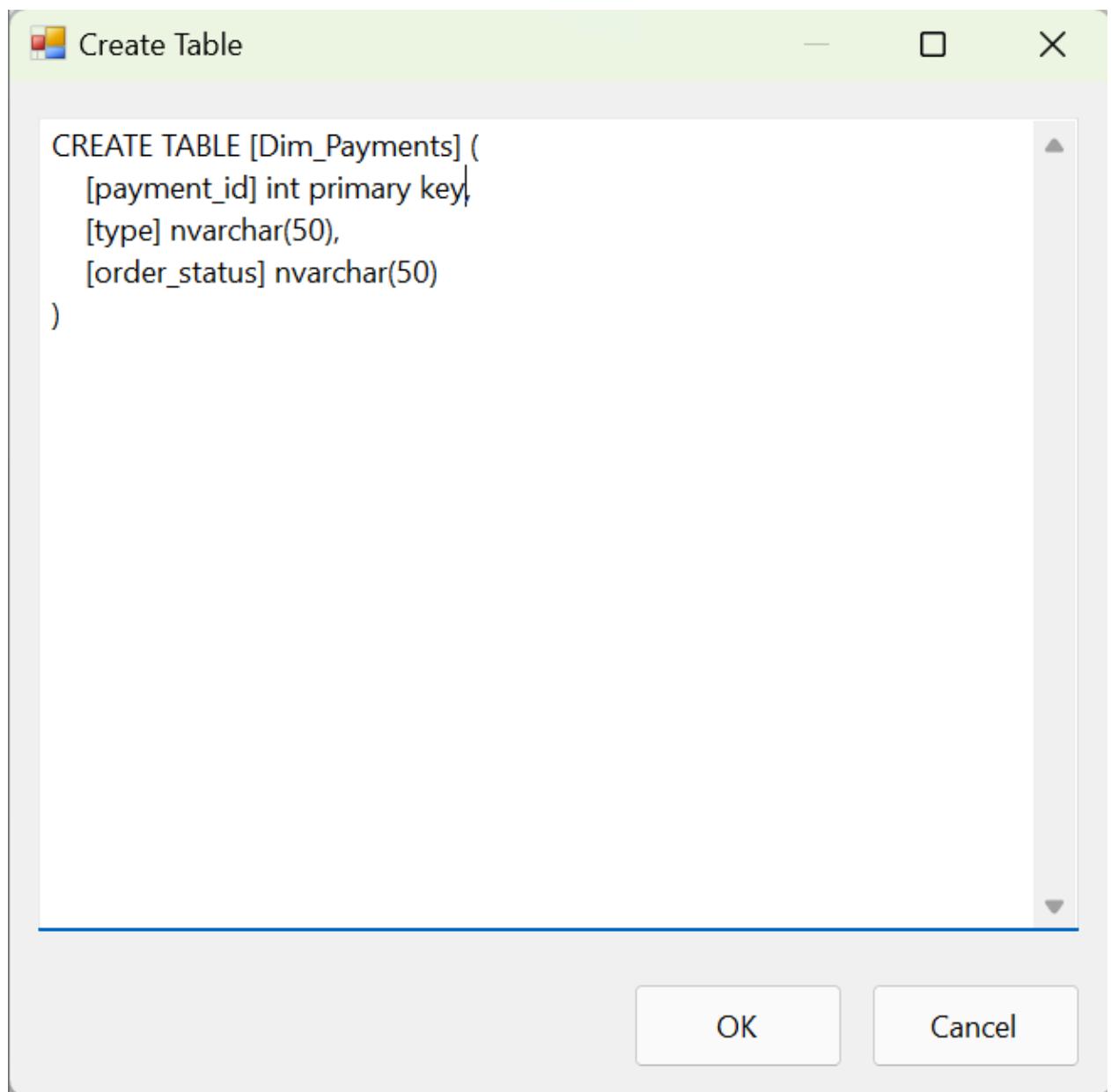
- Thực hiện edit công cụ Sort, tích chọn các thuộc tính cần sắp xếp



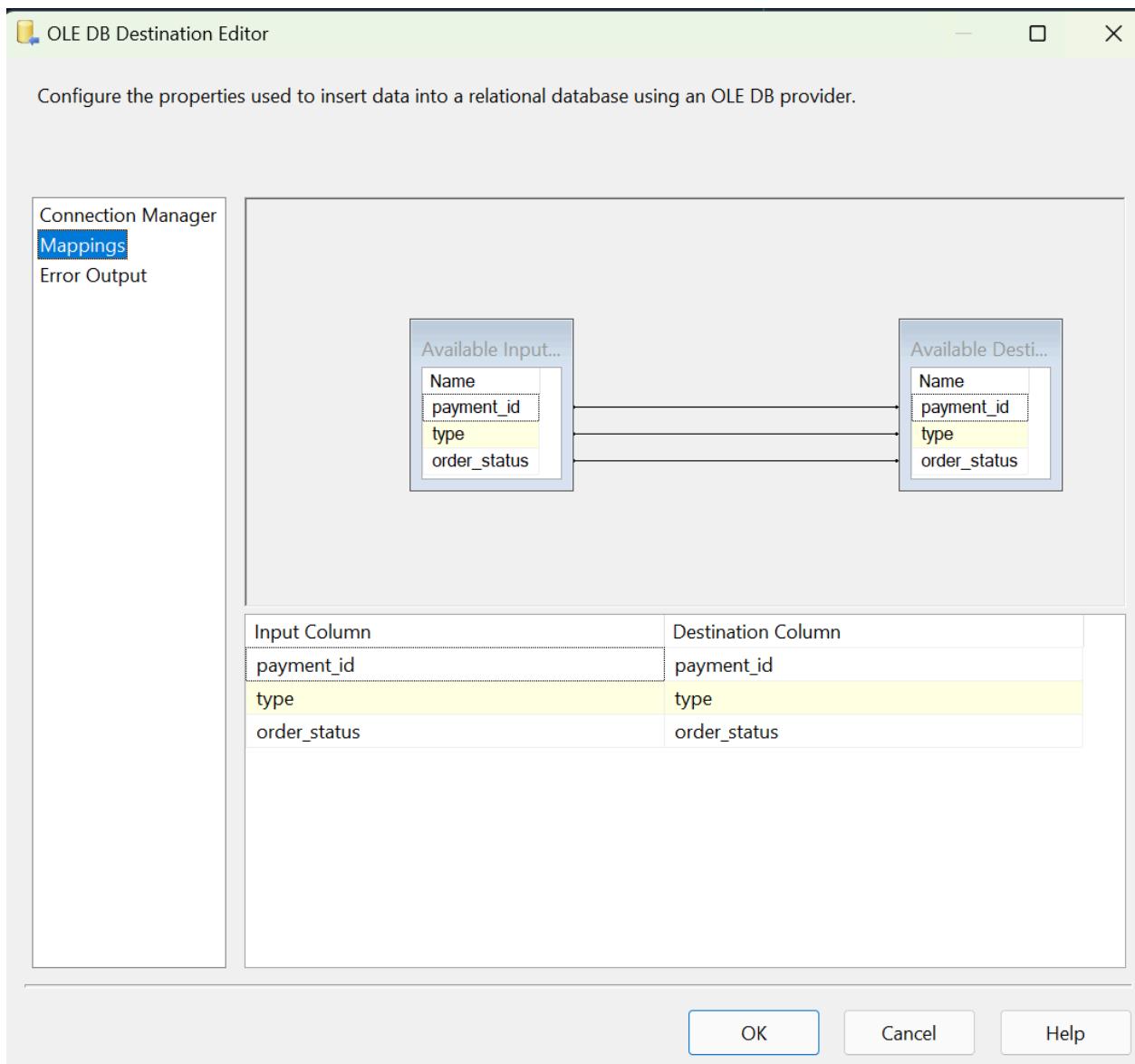
- Thực hiện kéo thả công cụ OLE DB Destination



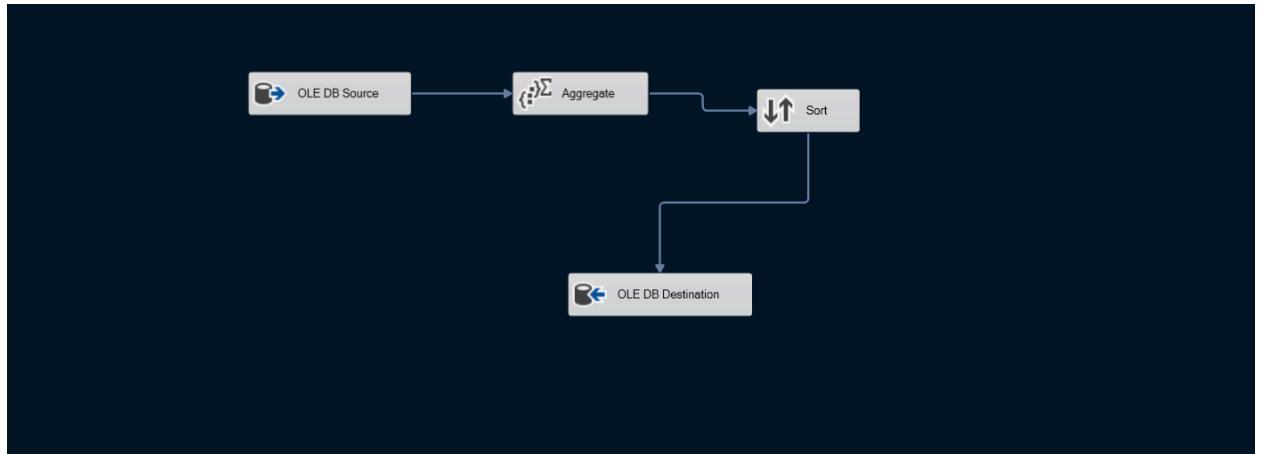
- Thực hiện edit công cụ OLE DB Destination, kết nối tới DatacoWareHouse và tạo bảng Dim_Payments



- Chọn qua tab Mapping và map các thuộc tính



- Data flow khi hoàn thành

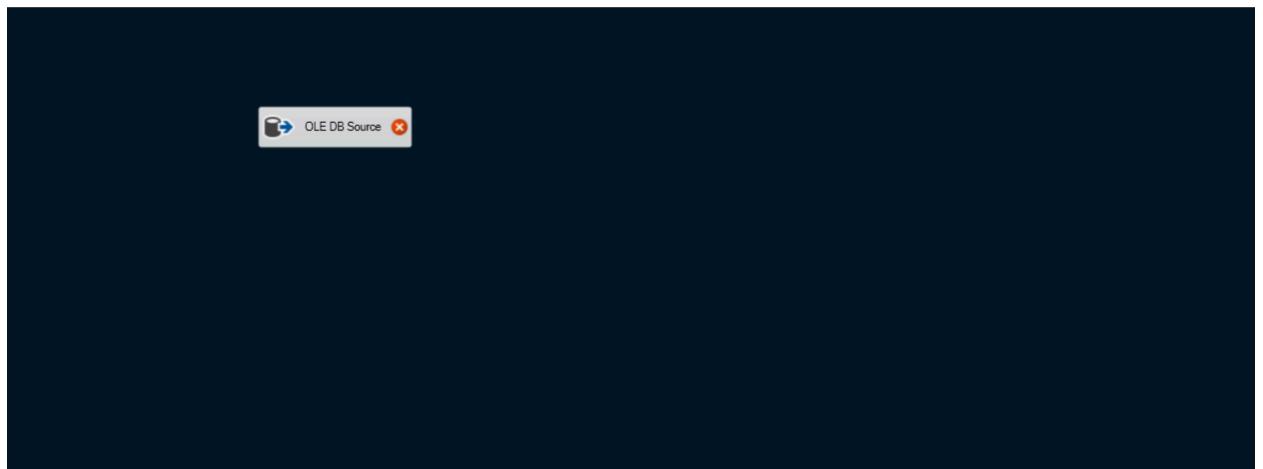


2.3.7. Tạo bảng Dim_Markets

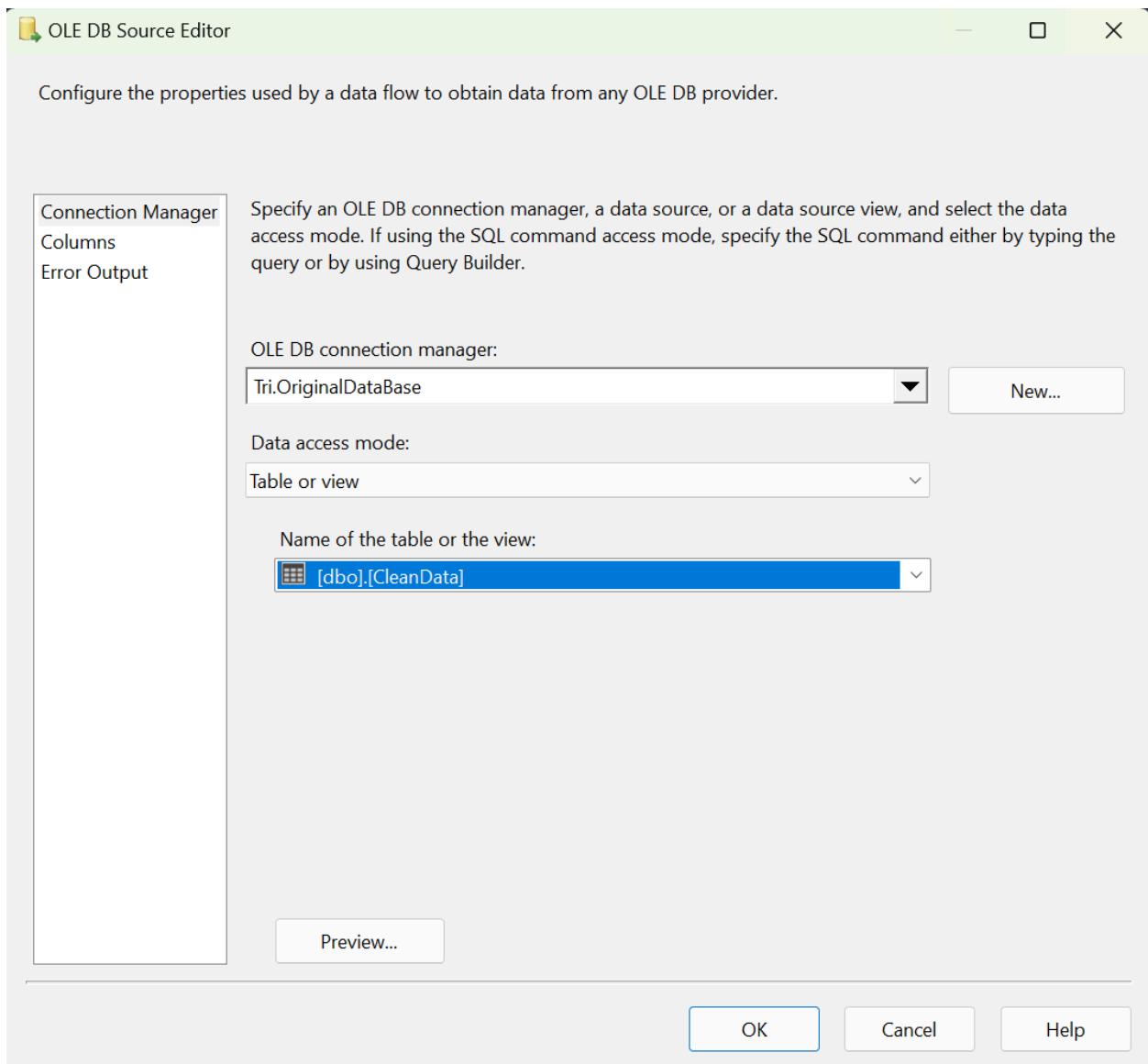
- Thực hiện kéo thả Data flow task và đổi tên thành Dim_Markets



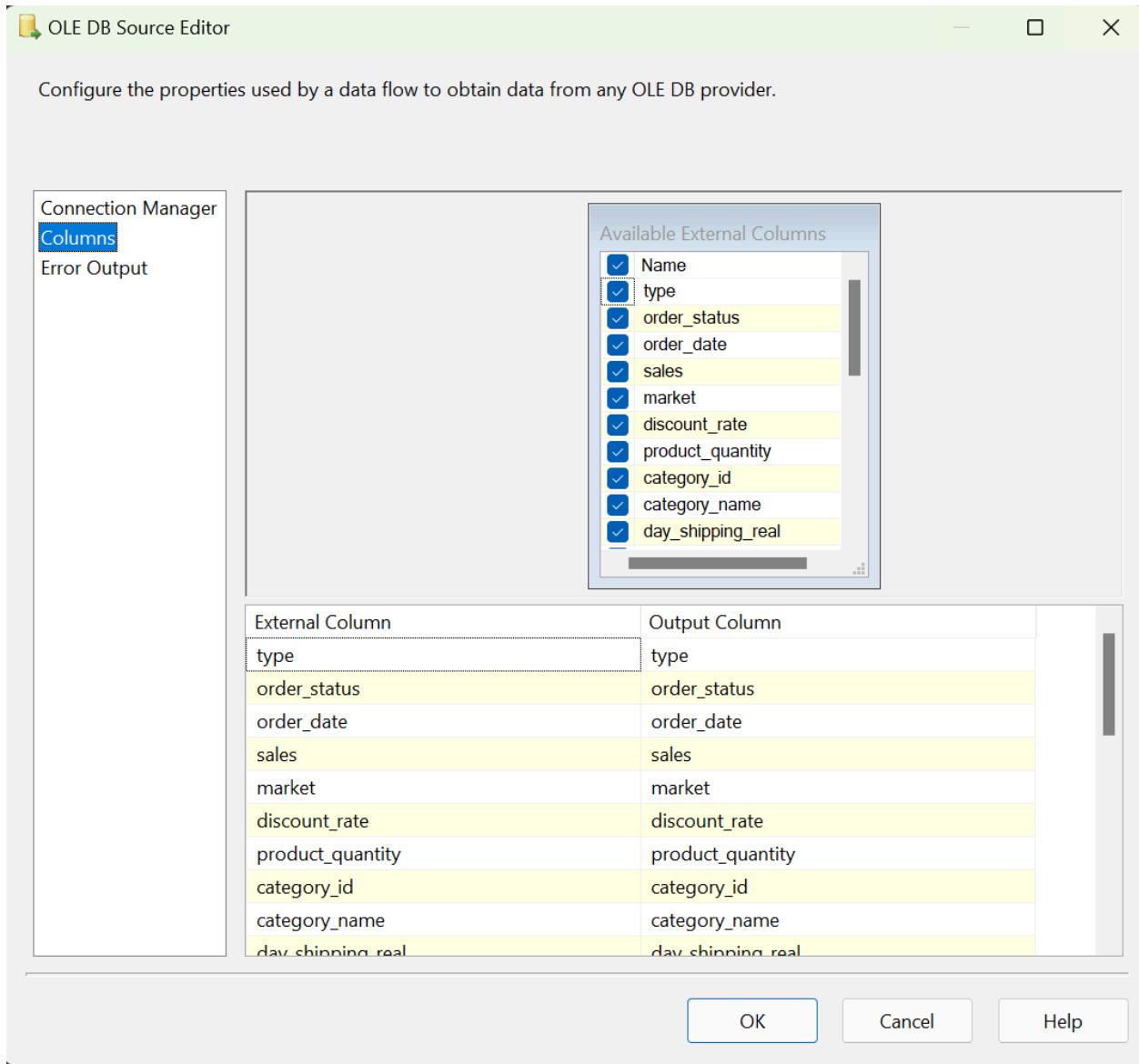
- Thực hiện edit data flow task trên, kéo thả công cụ OLE DB Source



- Thực hiện kết nối tới bảng CleanData



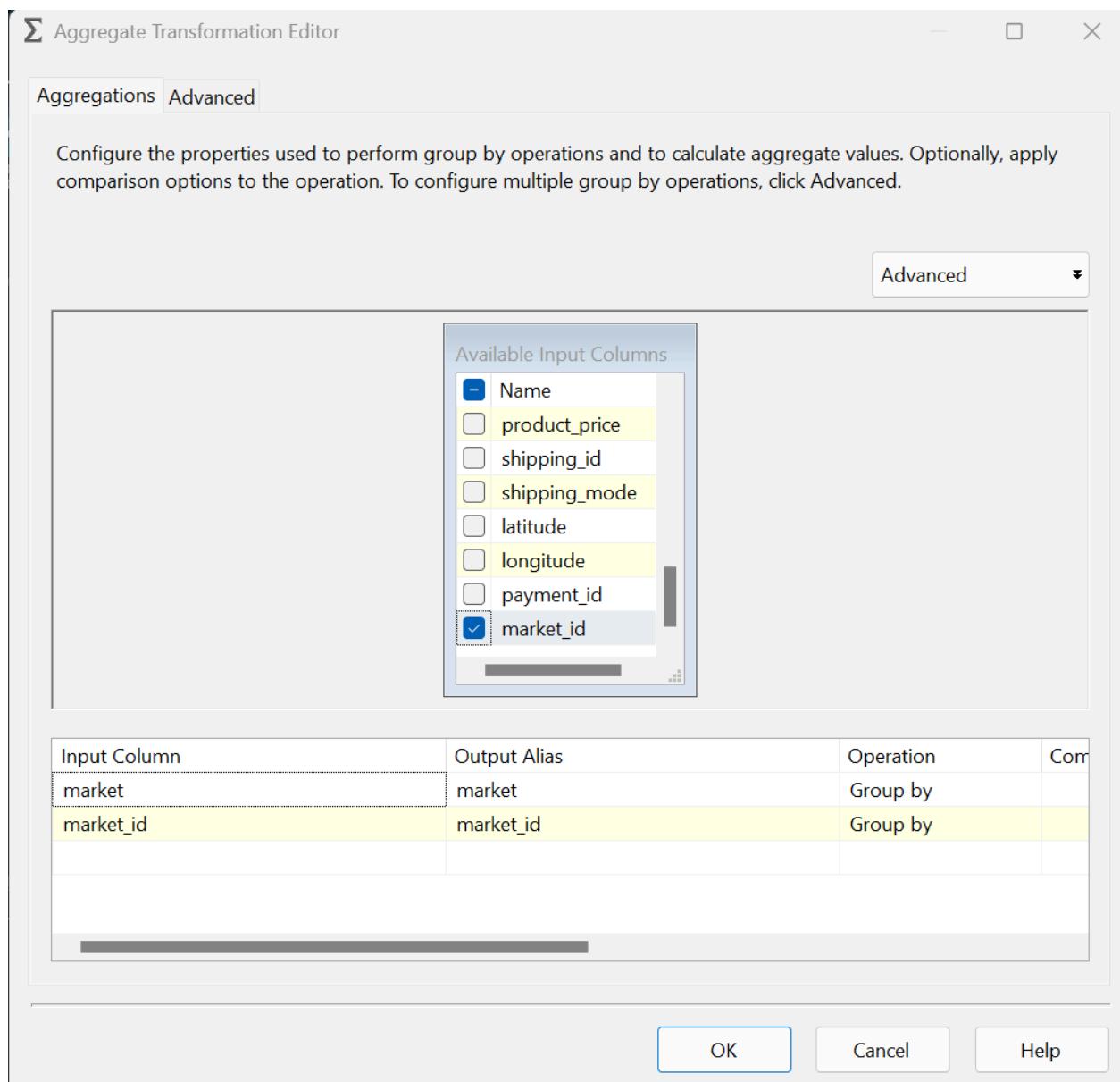
- Chọn qua tab Columns và tích chọn các thuộc tính cần thiết



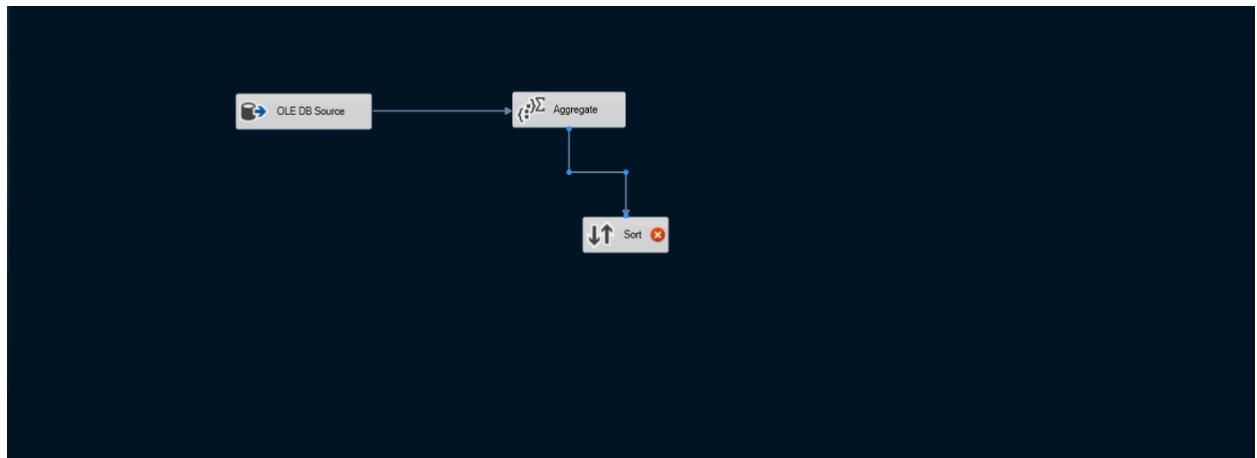
- Thực hiện kéo thả công cụ Aggregate để tách chọn các thuộc tính cho bảng



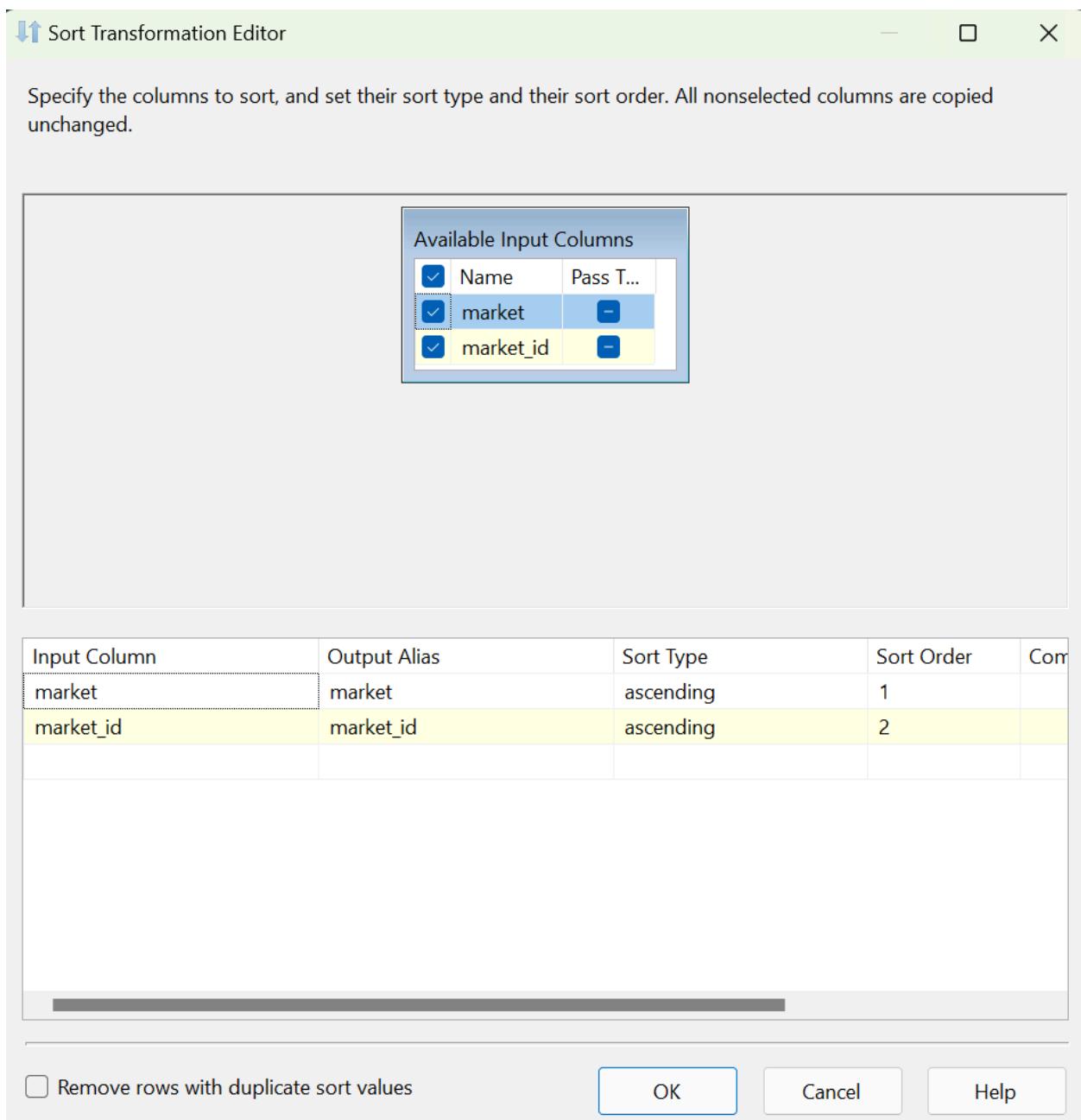
- Thực hiện edit công cụ, tích chọn các thuộc tính cần thiết



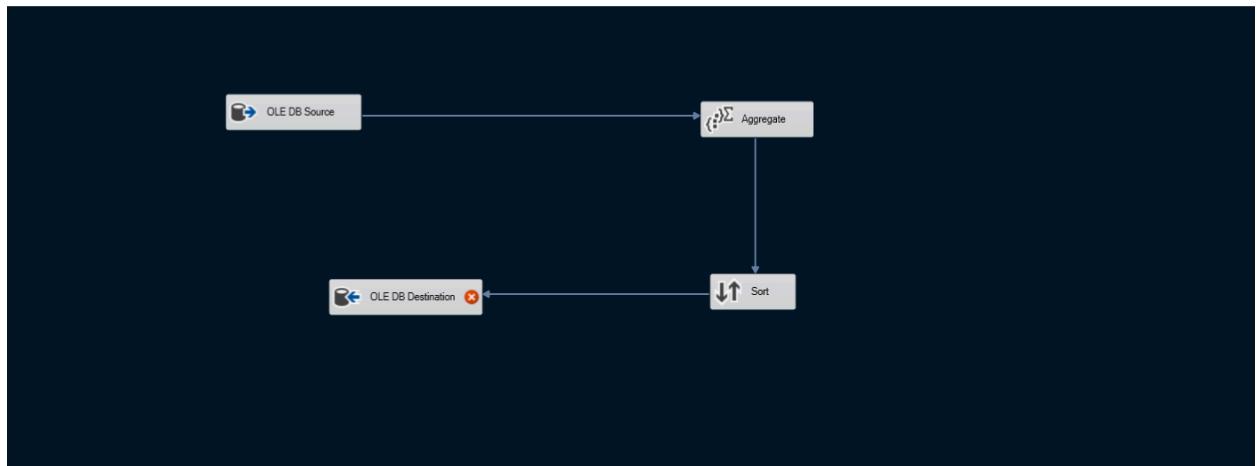
- Thực hiện kéo thả công cụ Sort để sắp xếp lại dữ liệu



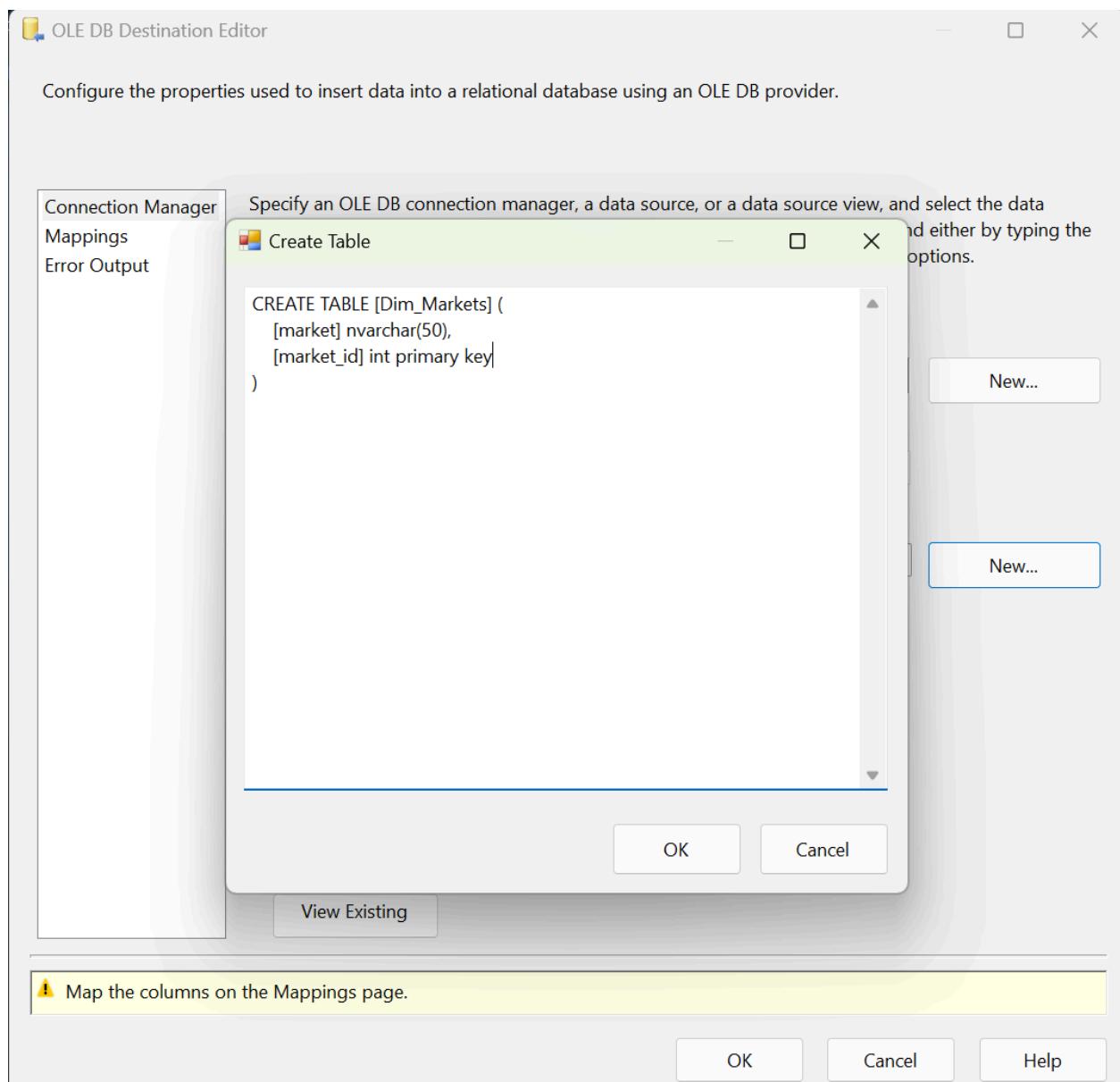
- Thực hiện edit, tích chọn các thuộc tính cần sắp xếp lại dữ liệu



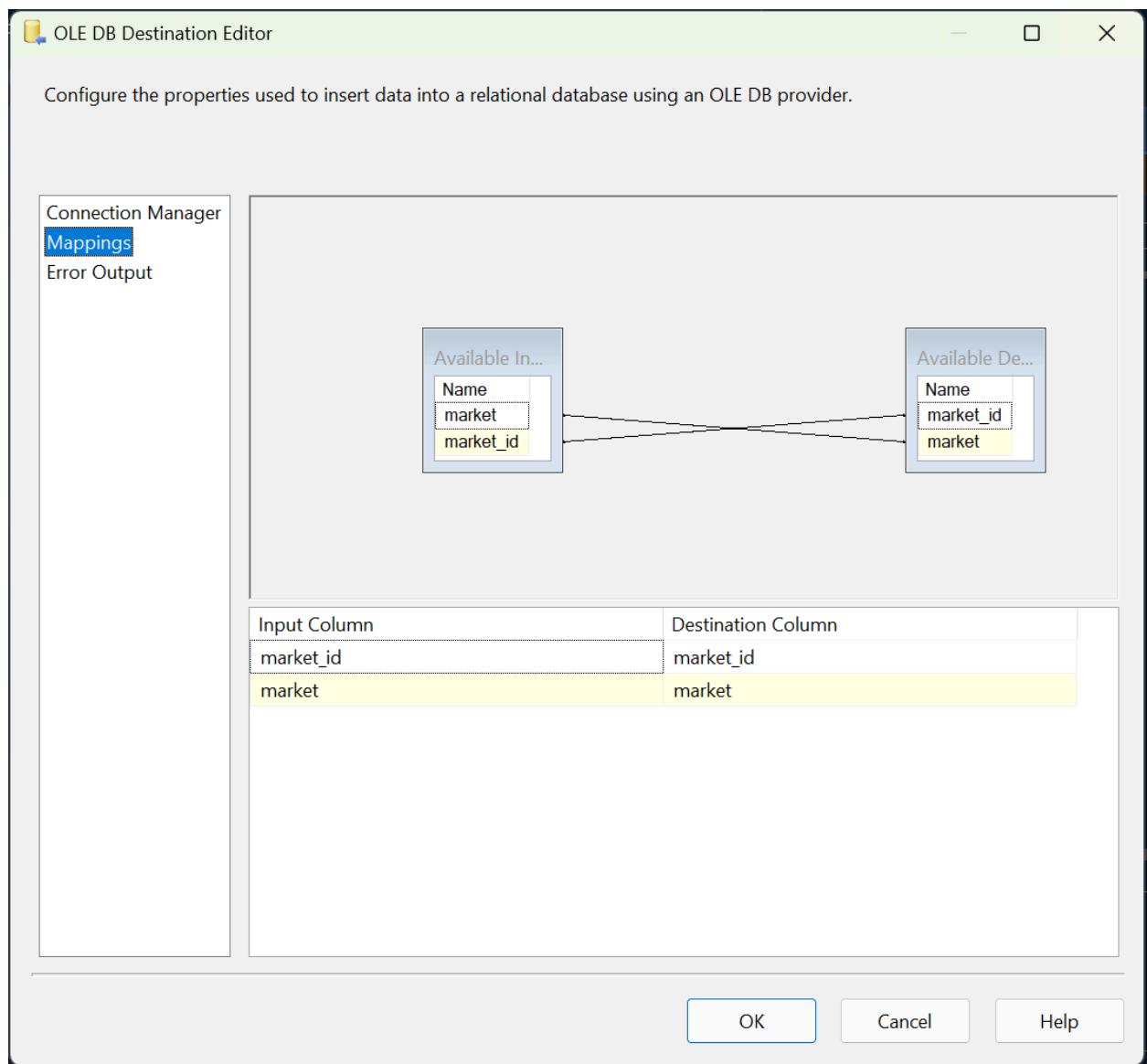
- Thực hiện kéo thả công cụ OLE DB Destination và tiến hành edit



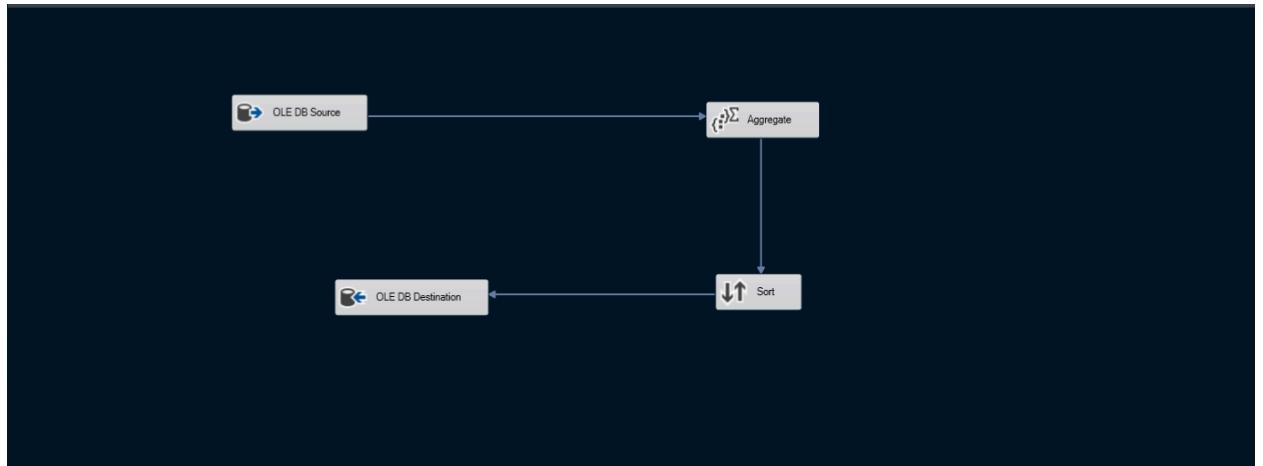
- Thực hiện kết nối tới Warehouse và tạo bảng Dim_Markets



- Chuyển qua tab map và map các thuộc tính cần thiết

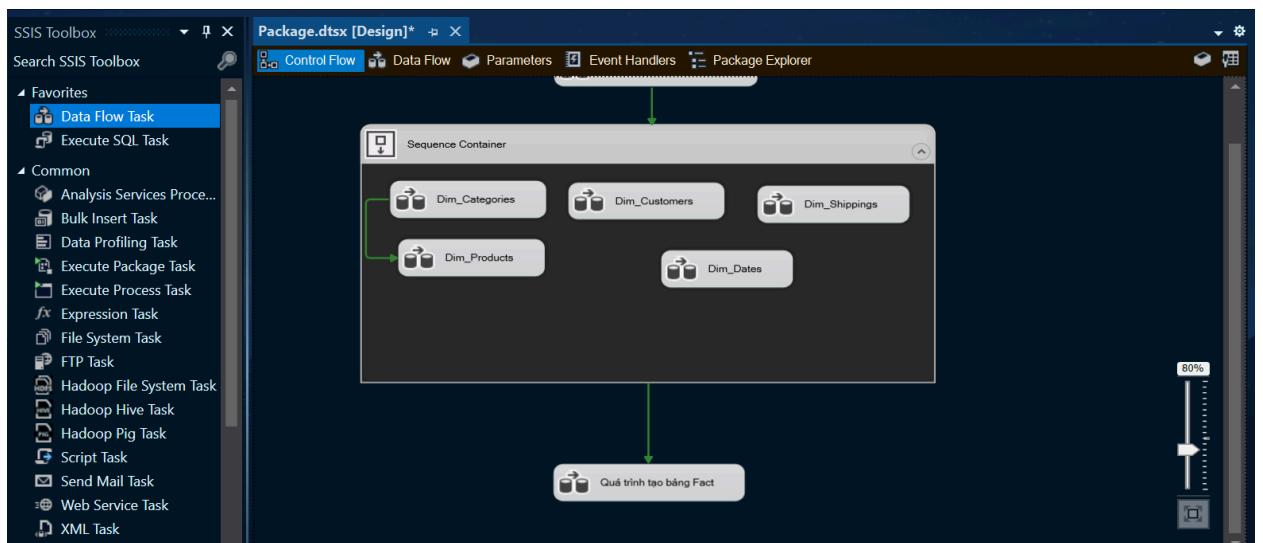


- Data flow task khi hoàn thành

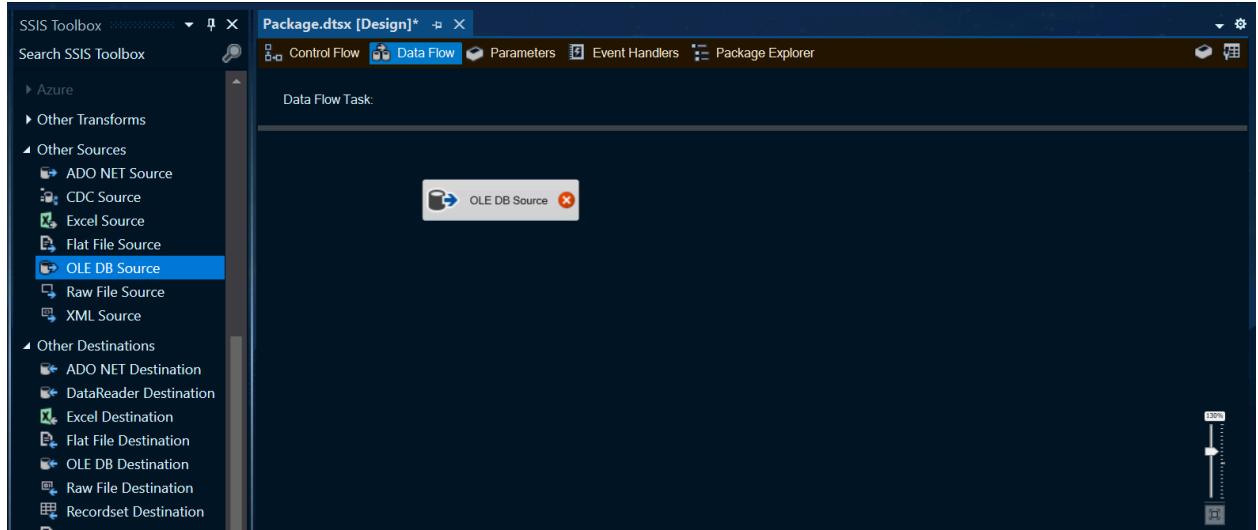


2.4. TẠO BẢNG FACT

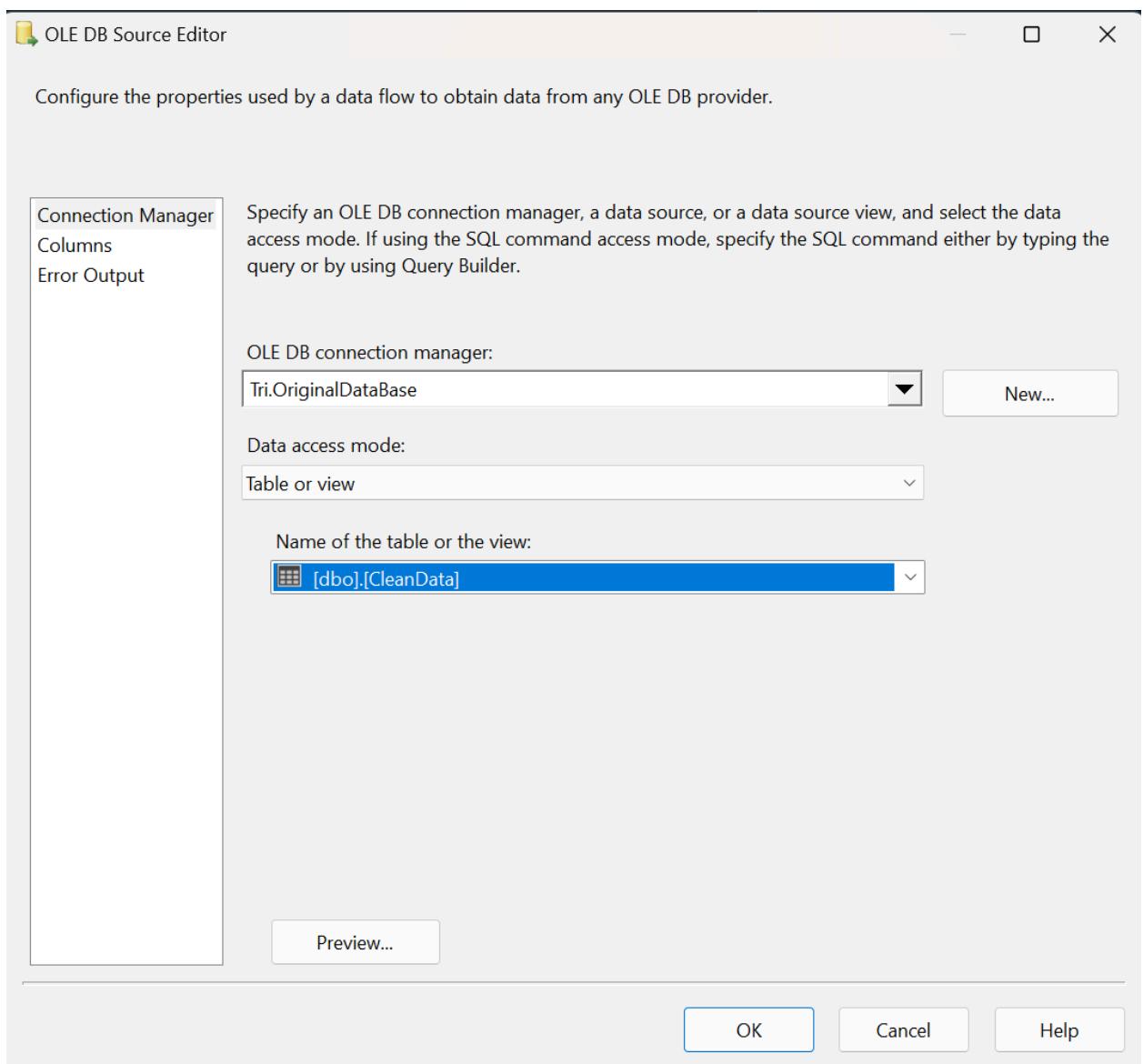
- Kéo thả công cụ Data Flow Task từ Toolbox sau đó đổi tên thành “Quá trình tạo bảng Fact”.



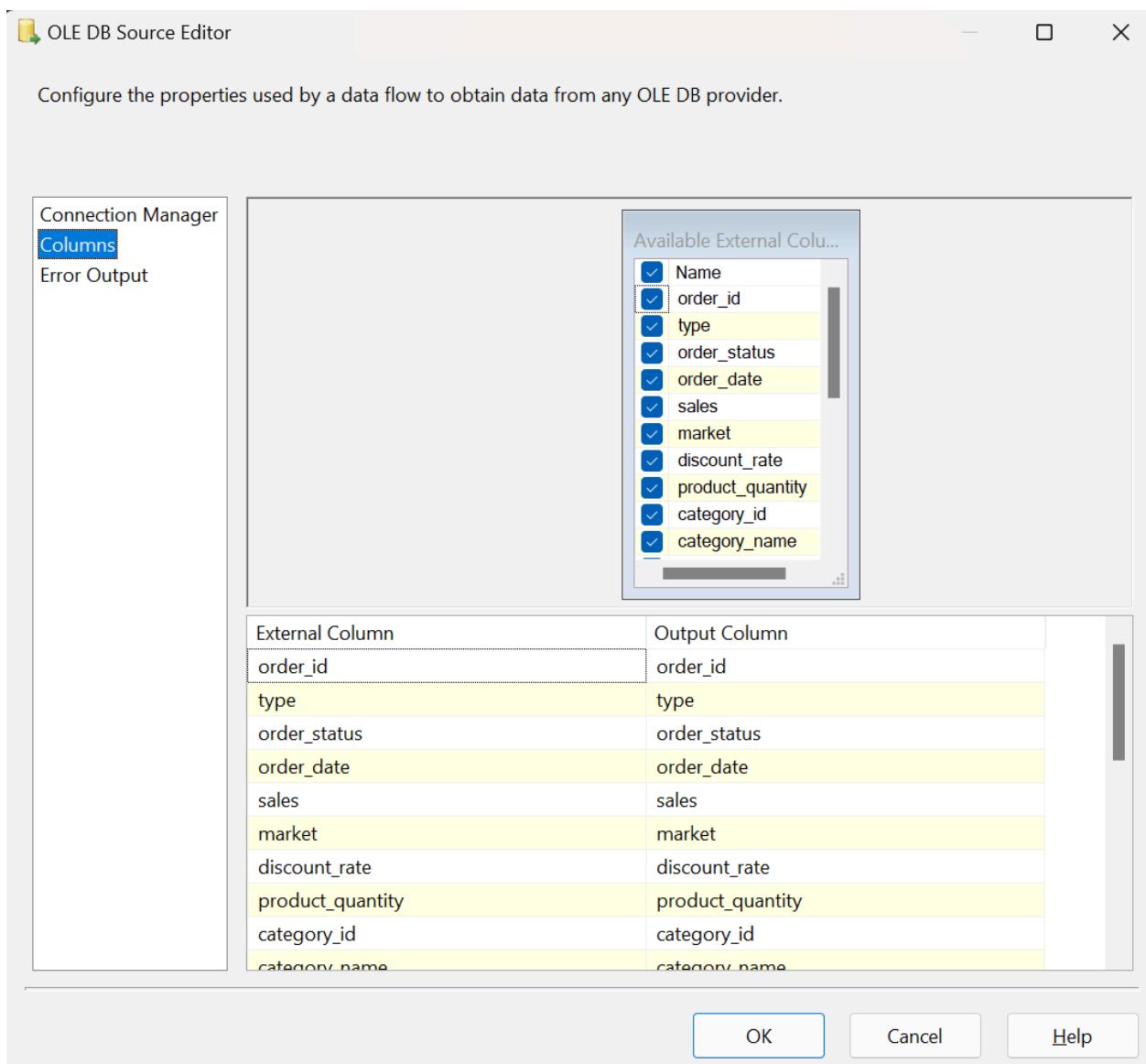
- Tiếp theo thực hiện edit Flow Task vừa tạo, kéo thả công cụ OLE DB Source từ Toolbox.



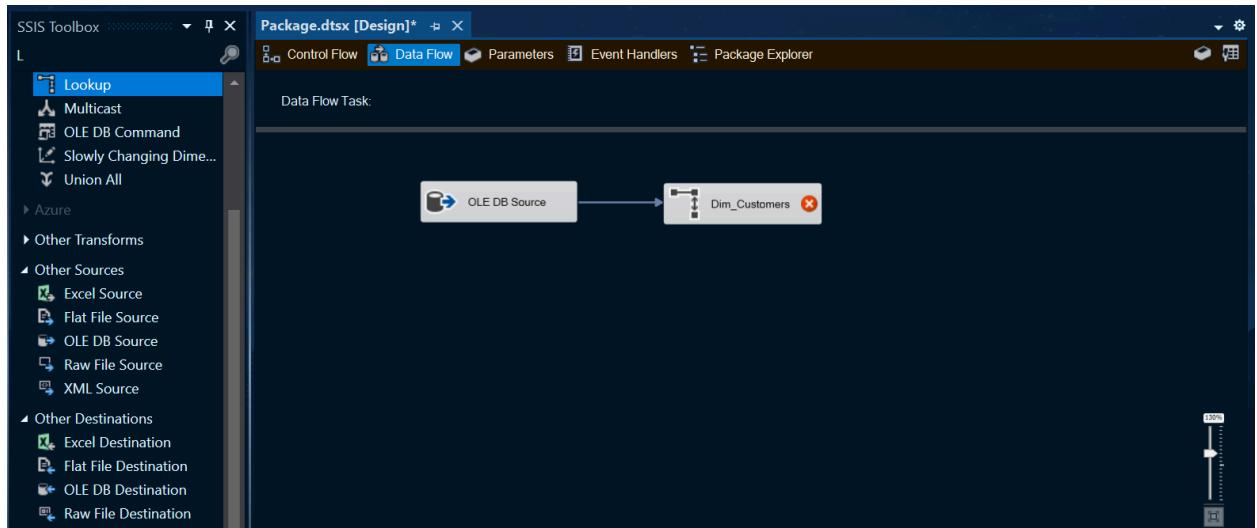
- Thực hiện edit OLE DB Source, kết nối tới bảng CleanData trong database OriginalDatabase



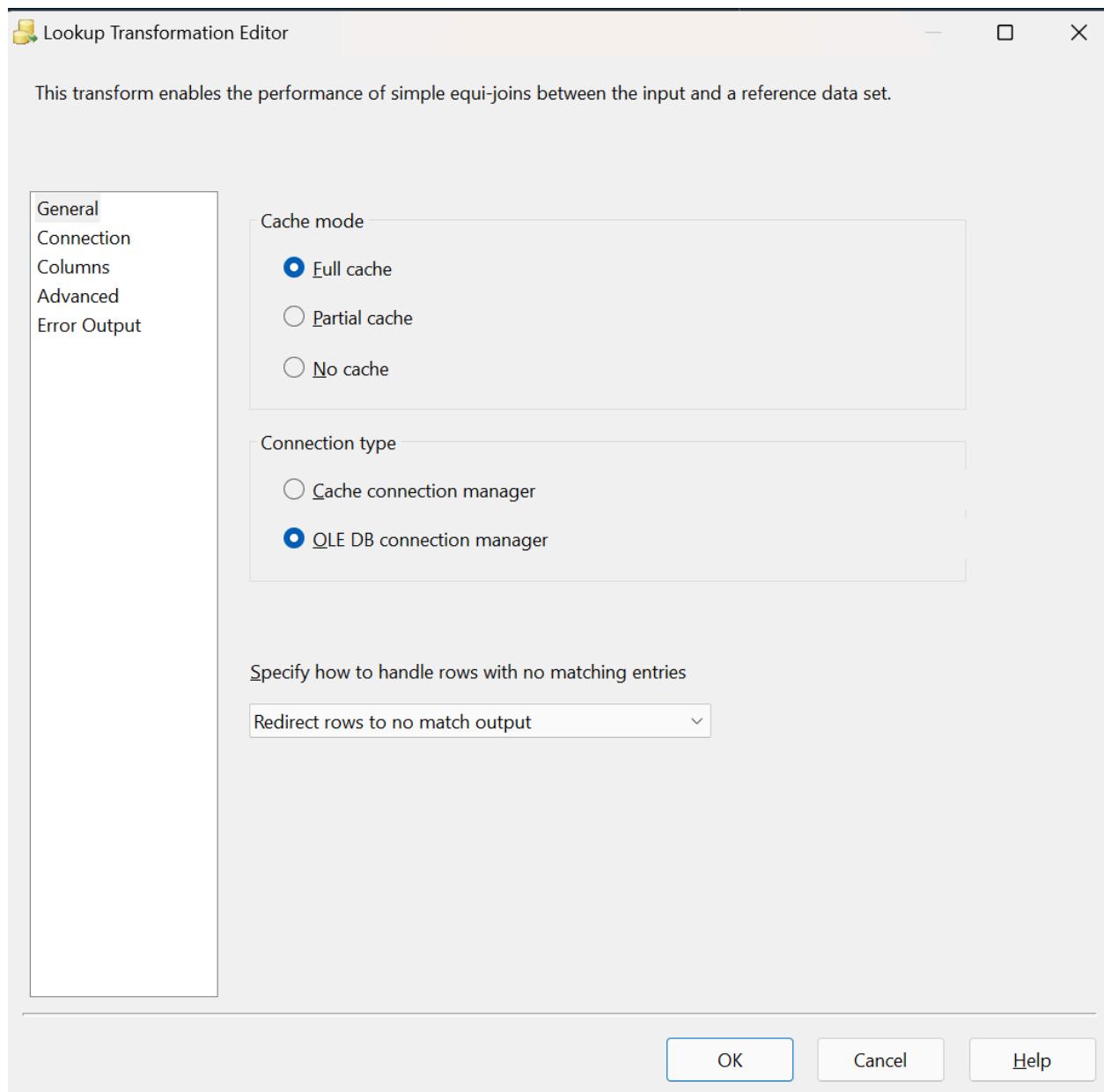
- Chuyển qua tab Column và tích chọn các thuộc tính cần thiết.



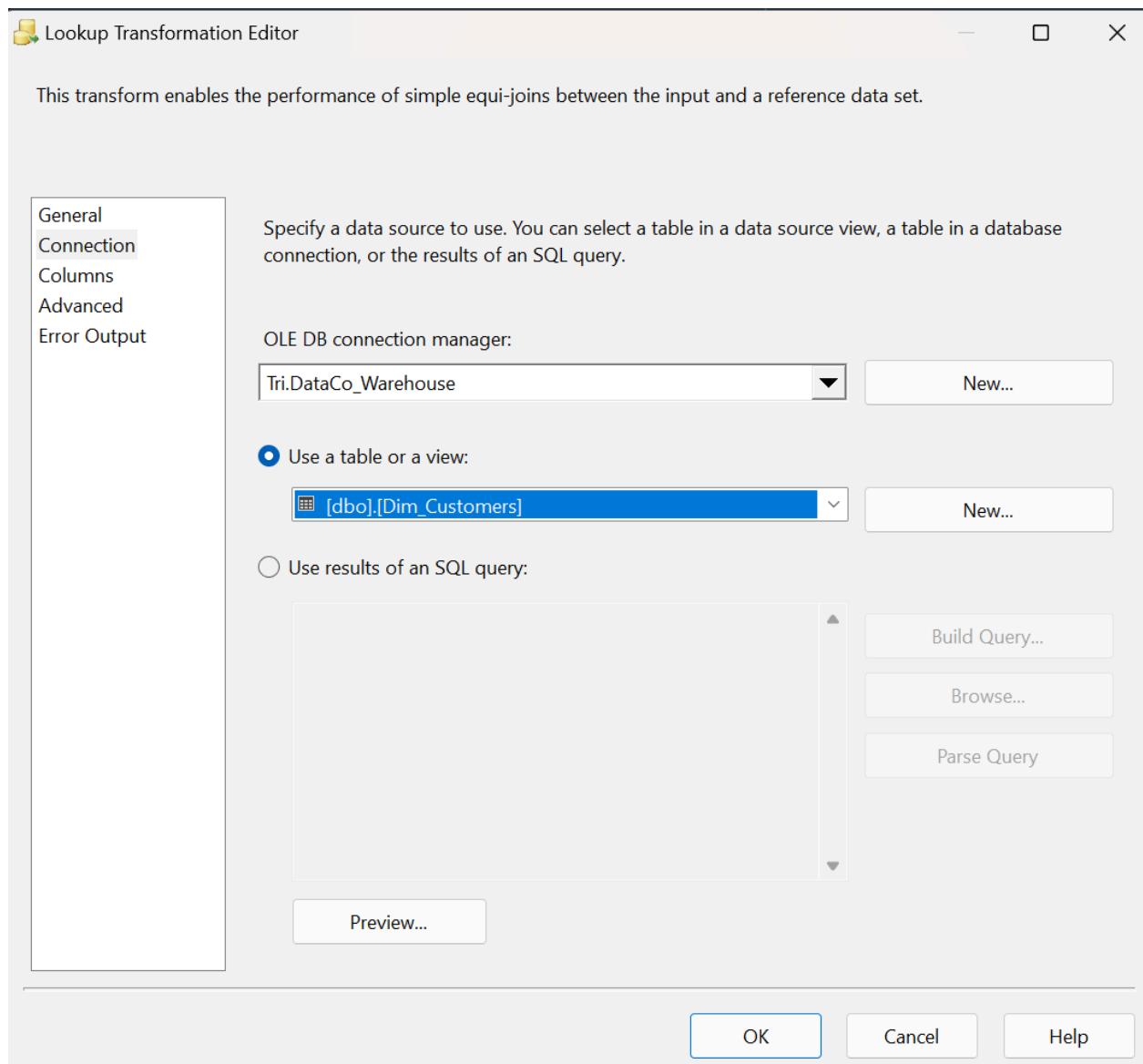
- Kéo thả công cụ Lookup từ Toolbox, sau đó đổi tên thành Dim_Customers.



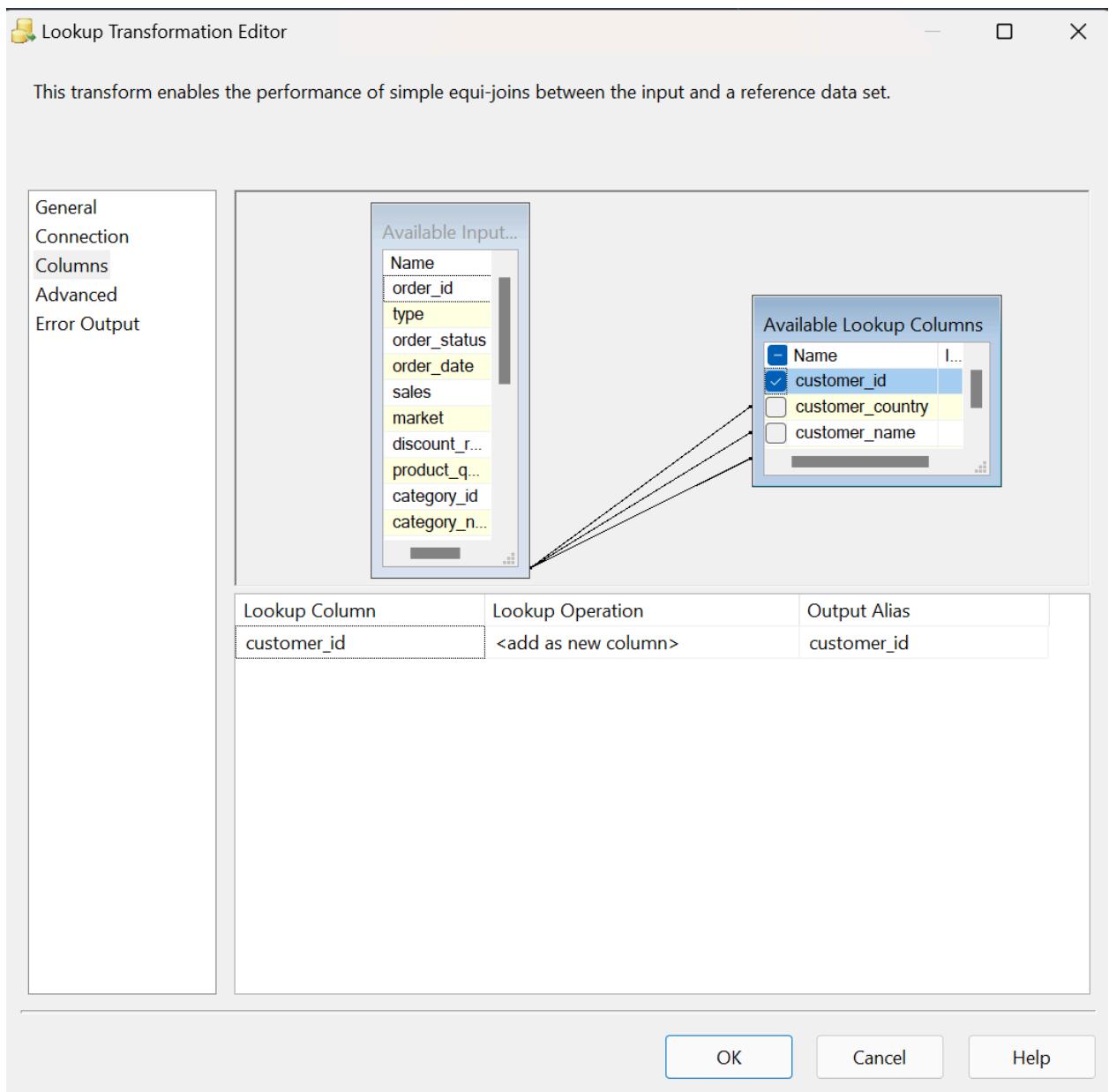
- Thực hiện edit Lookup Dim_Customers, ở mục **Specify how to handle rows with no matching entries** => chọn **Redirect rows to no match output**.



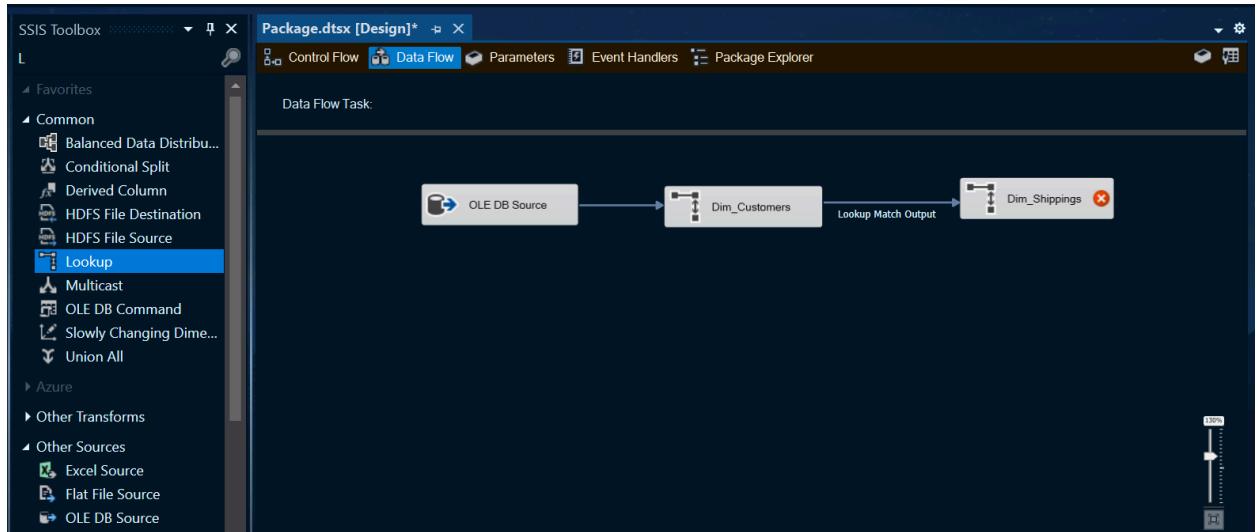
- Chuyển qua tab Connect và thực hiện kết nối tới bảng Dim_Customers trong database DataCo_Warehouse.



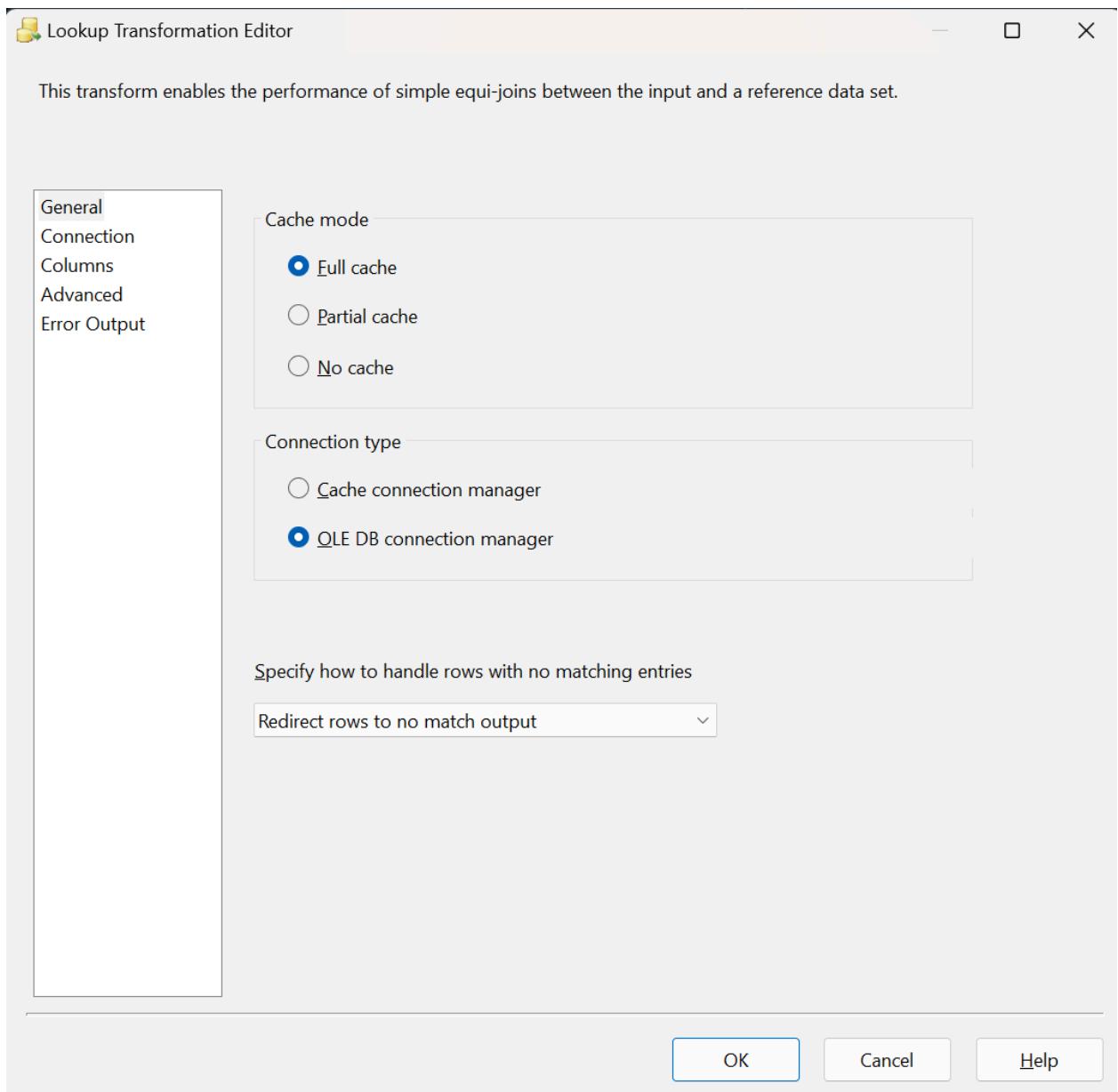
- Chuyển qua tab Column và thực hiện map các thuộc tính cần thiết.



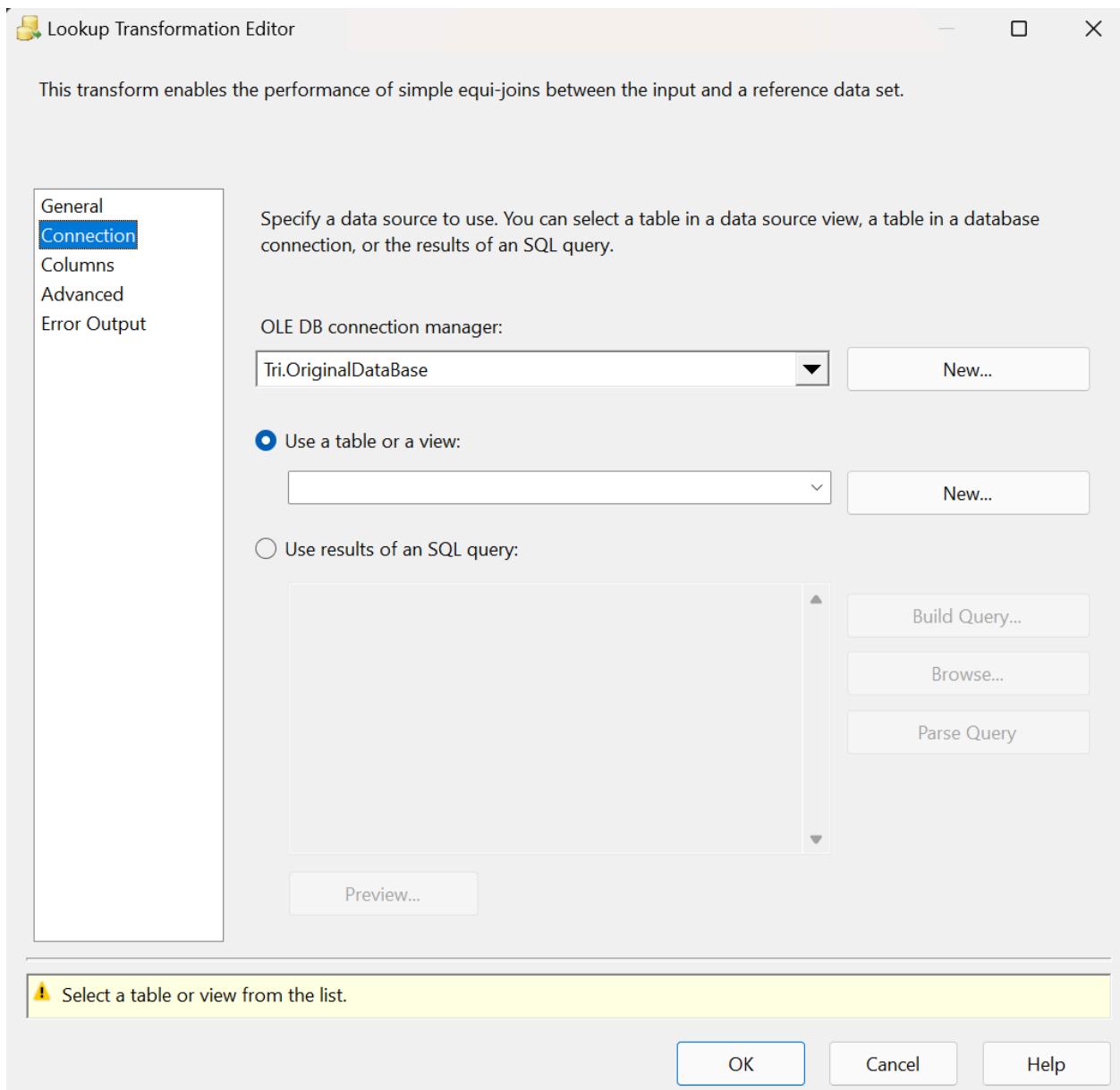
- Kéo thả công cụ Lookup và đổi tên thành Dim_Shippings.



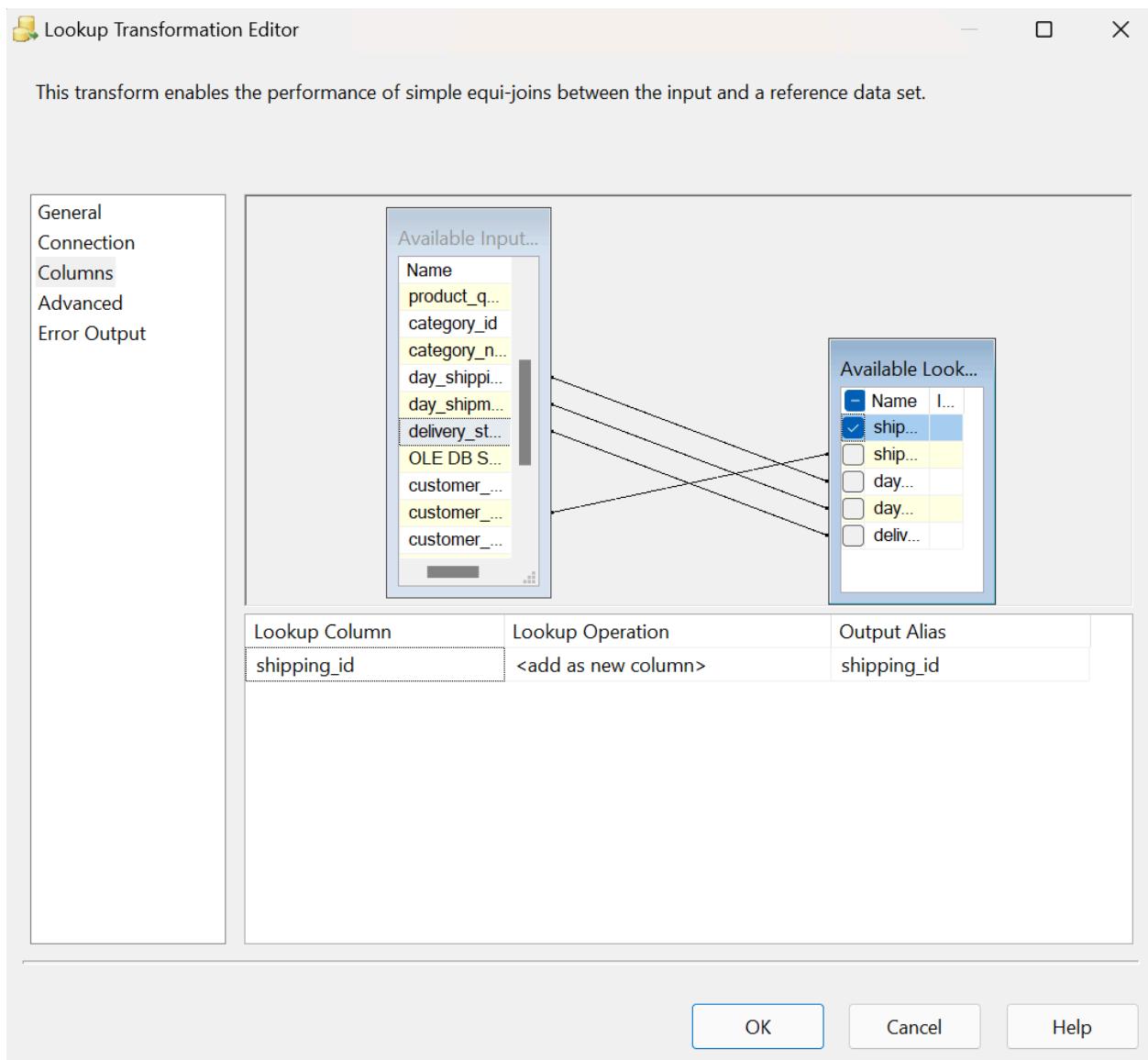
- Thực hiện edit Lookup Dim_Shippings, ở mục **Specify how to handle rows with no matching entries** => chọn **Redirect rows to no match output**.



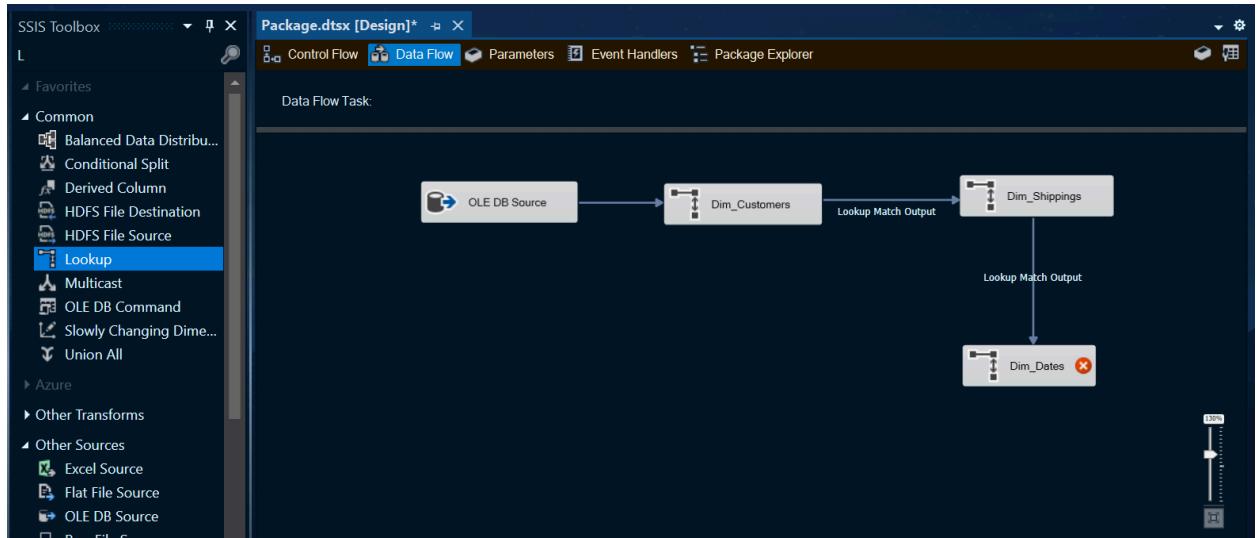
- Chuyển qua tab Connect và thực hiện kết nối tới bảng Dim_Shippings trong database DataCo_Warehouse.



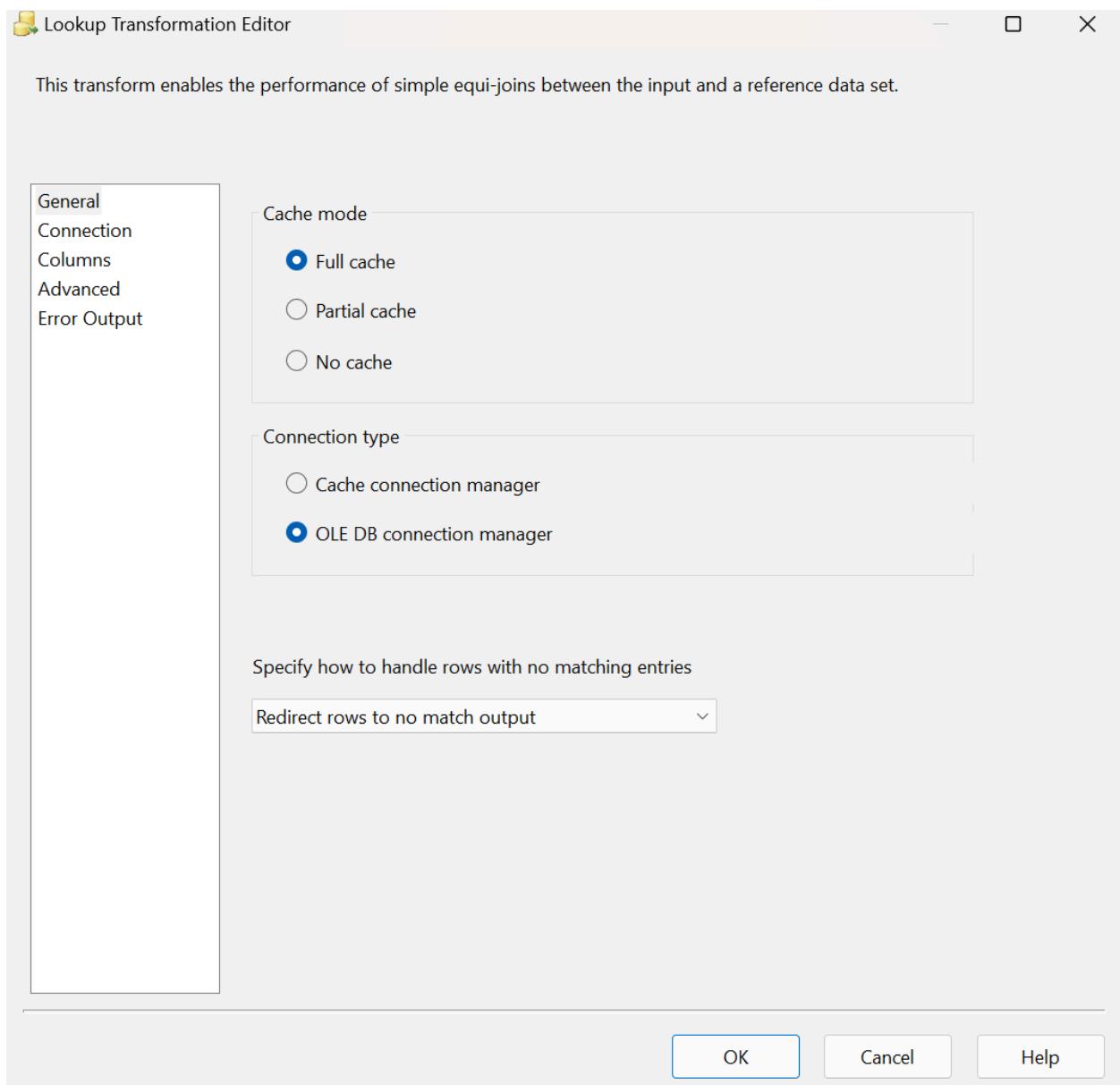
- Chuyển qua tab Column và map các thuộc tính cần thiết, đồng thời tích vào thuộc tính khóa.



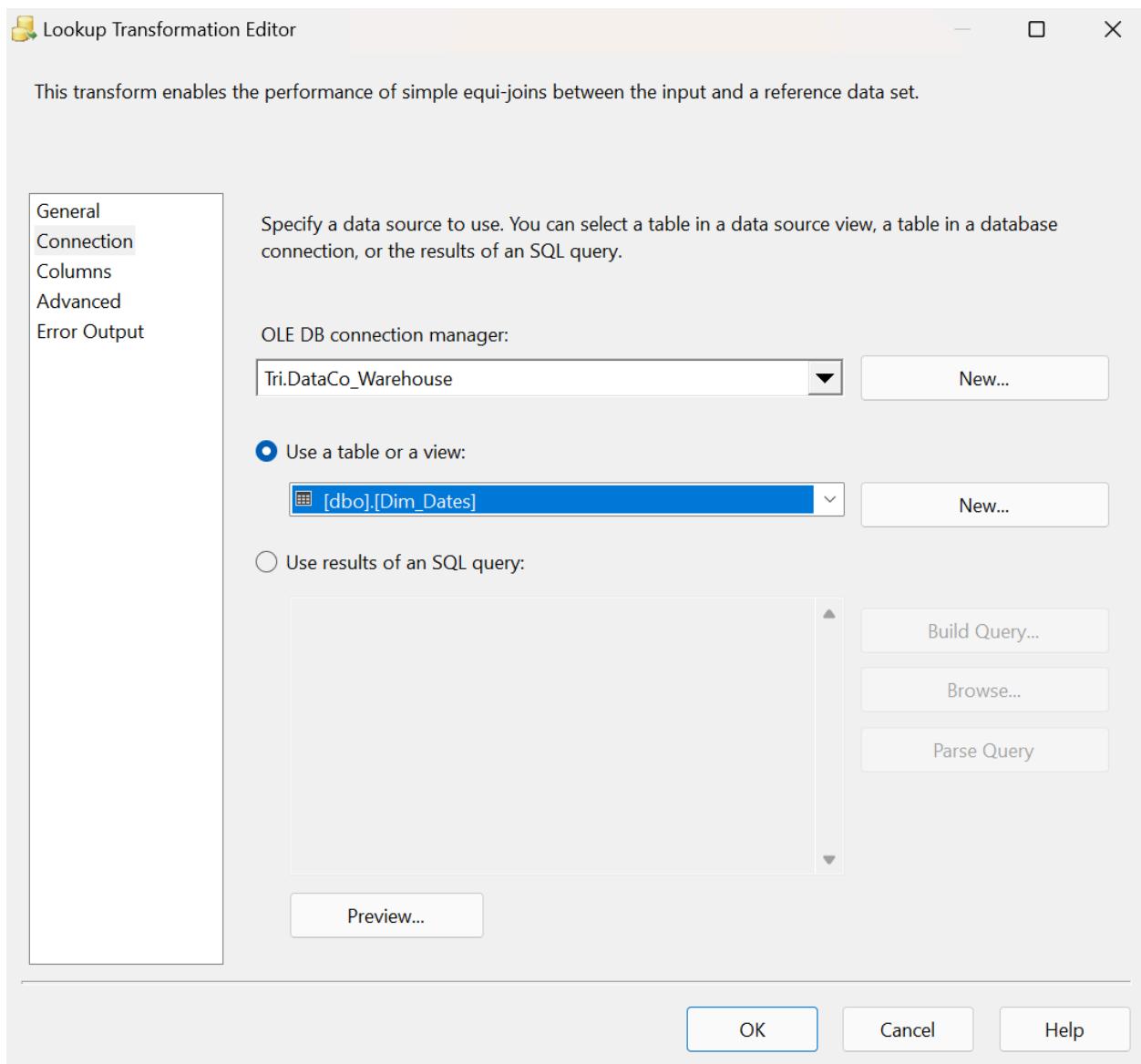
- Kéo thả công cụ Lookup và đổi tên thành Dim_Dates



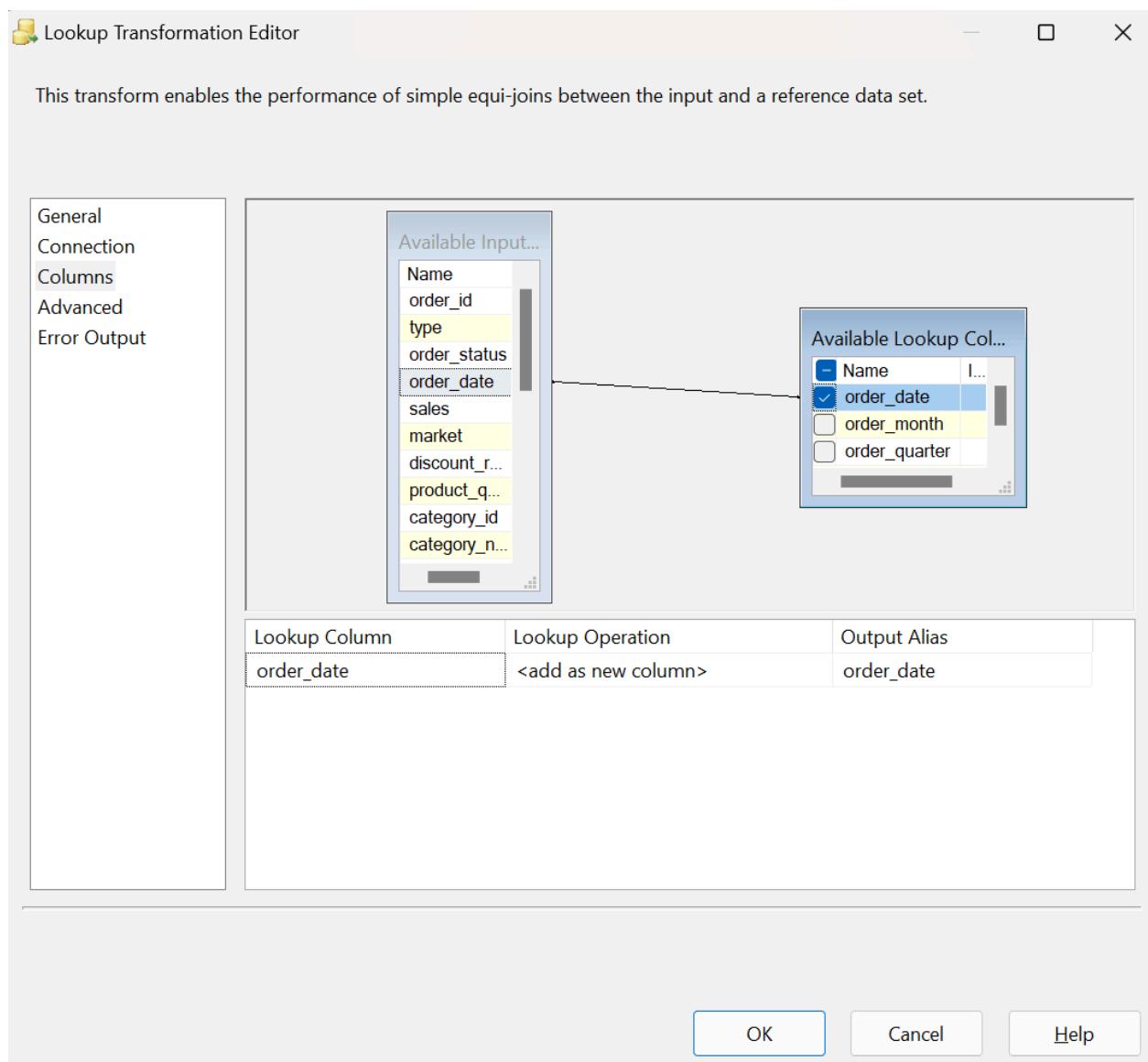
- Thực hiện edit Lookup Dim_Dates, ở mục **Specify how to handle rows with no matching entries** => chọn **Redirect rows to no match output**.



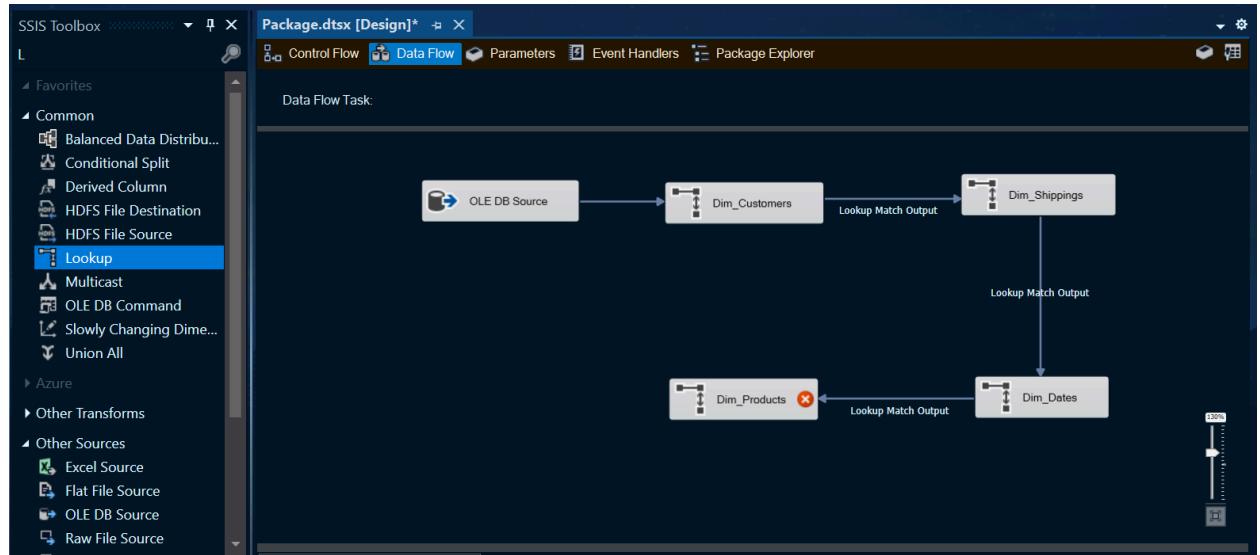
- Chuyển qua tab Connection và kết nối tới bảng Dim_Dates trong database DataCo_Warehouse.



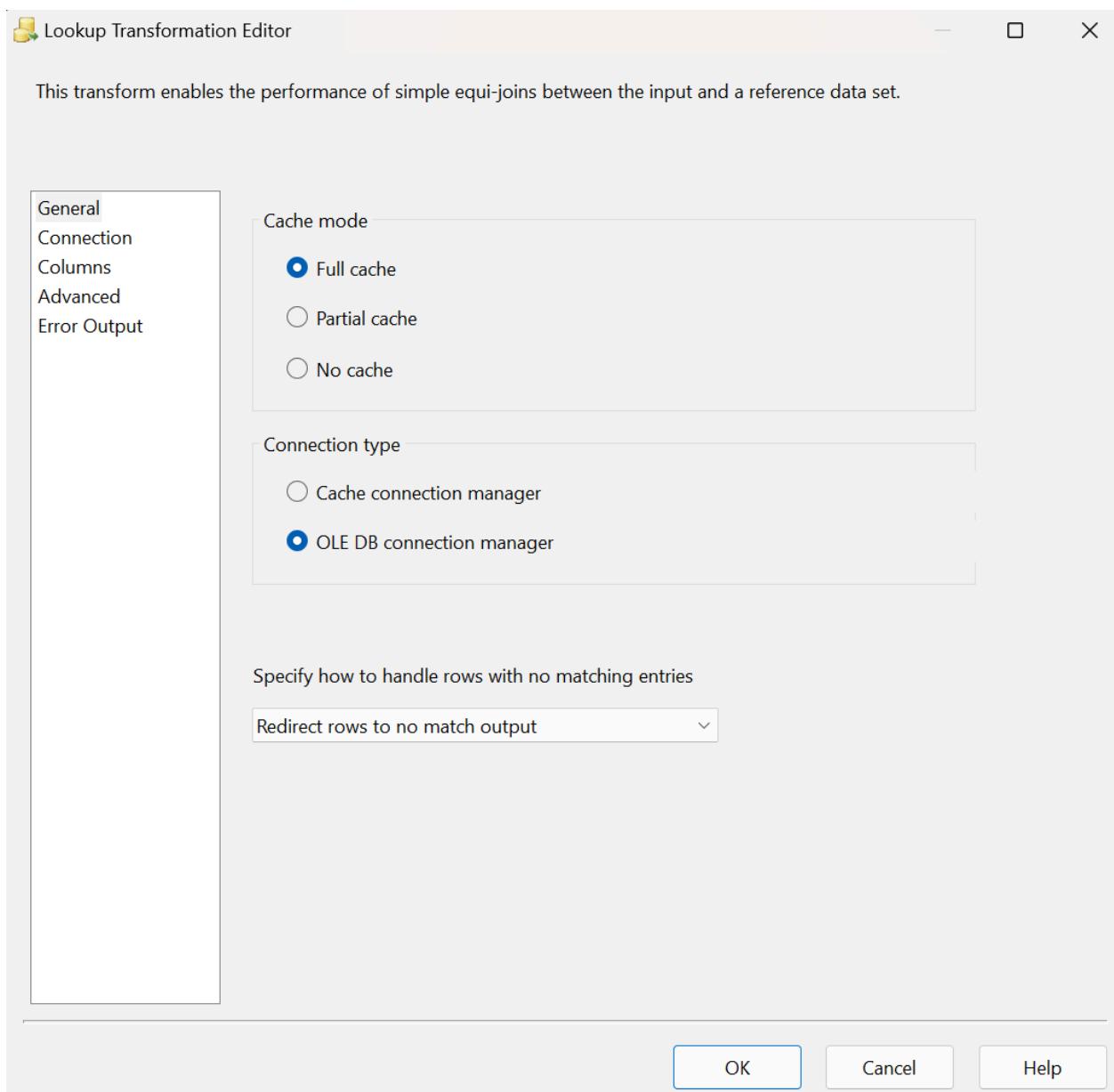
- Chuyển qua tab Column và map các thuộc tính cần thiết, đồng thời tích chọn vào khóa chính.



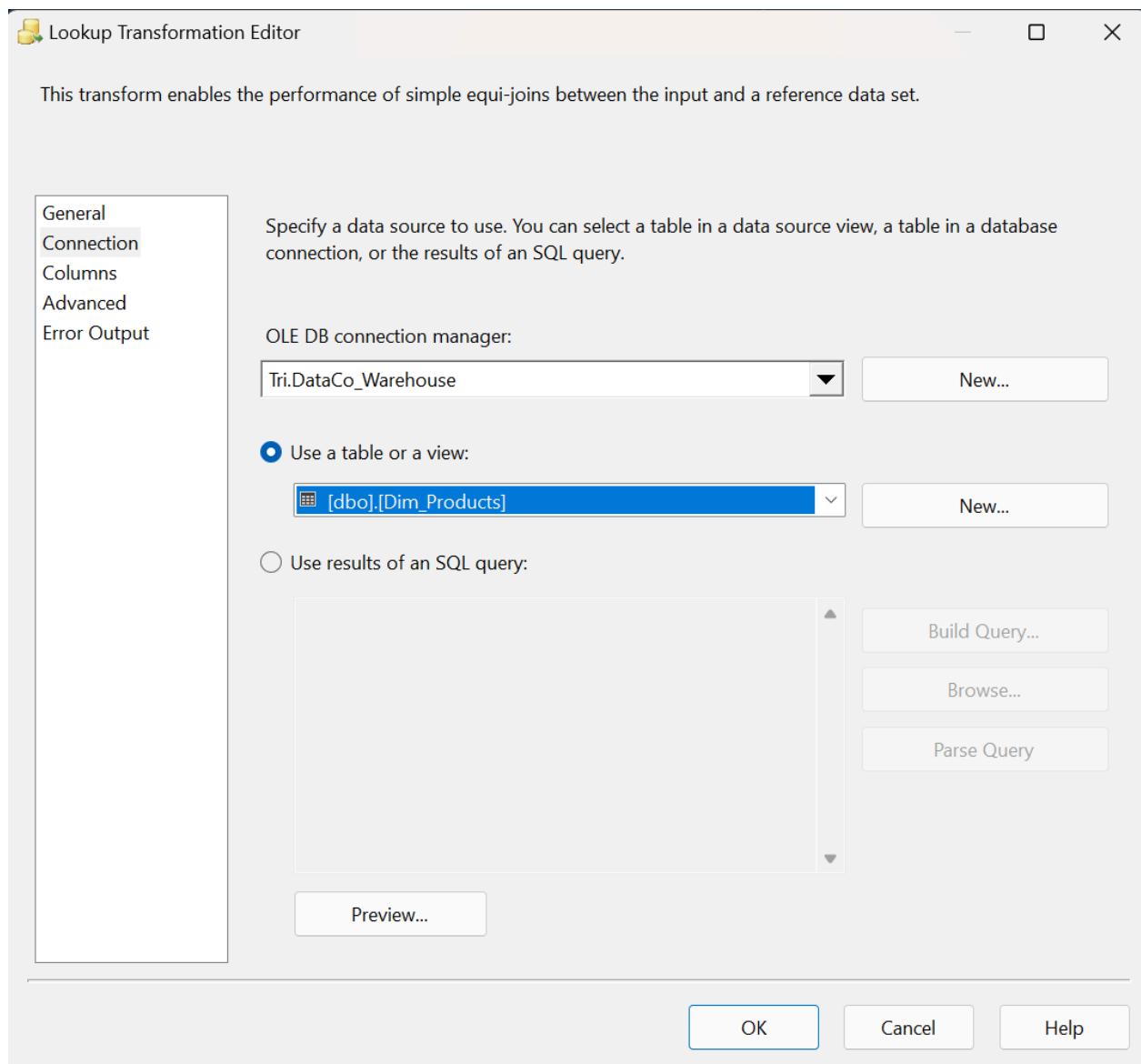
- Kéo thả công cụ Lookup từ Toolbox, sau đó đổi tên Dim_Products.



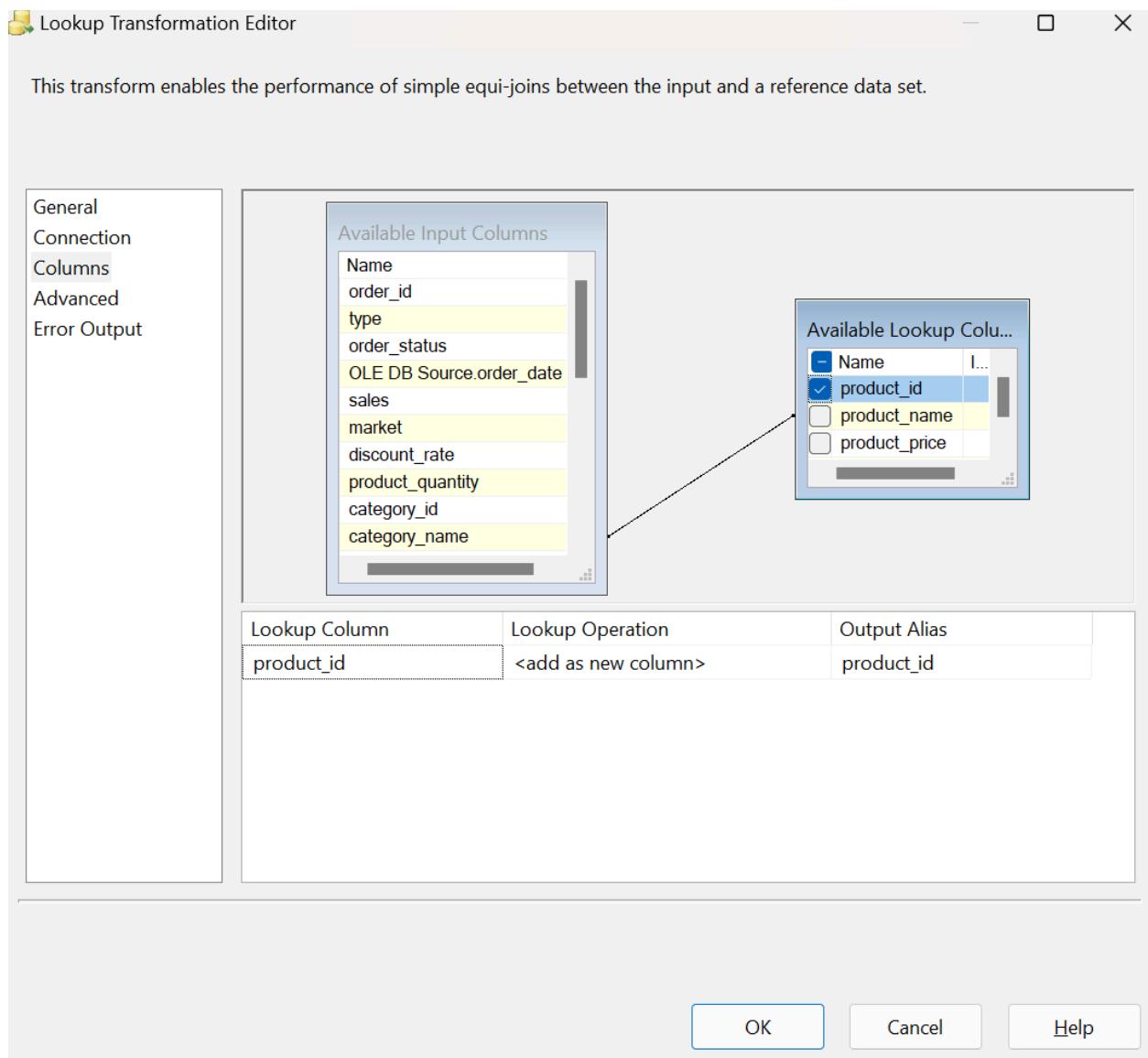
- Thực hiện edit Lookup Dim_Products, ở mục **Specify how to handle rows with no matching entries** => chọn **Redirect rows to no match output**.



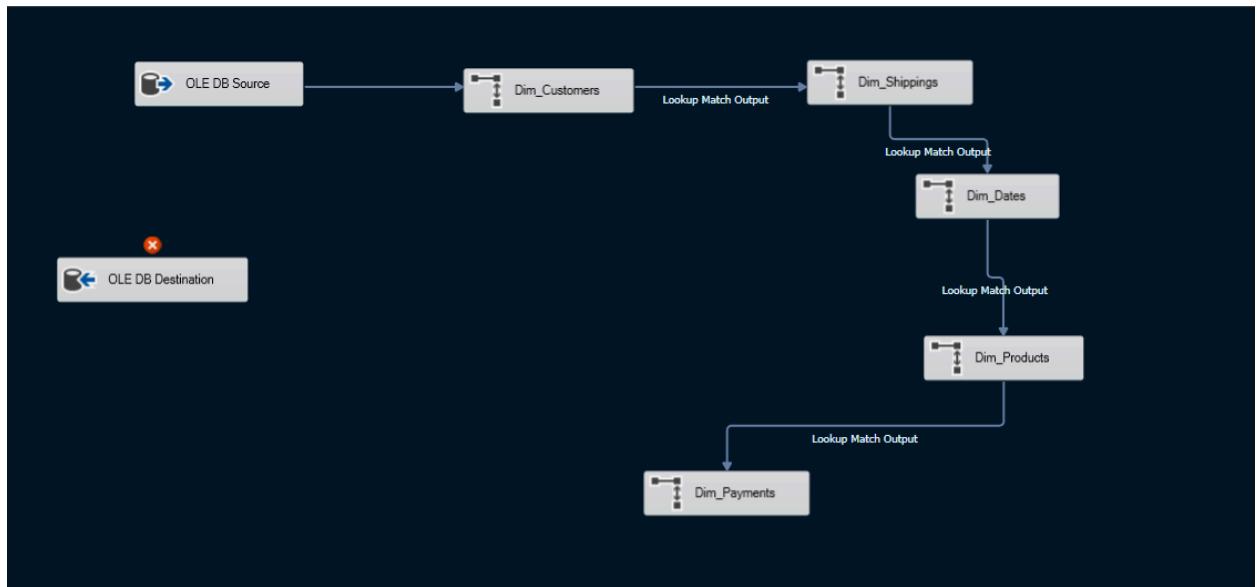
- Chuyển tab Connection và thực hiện kết nối tới bảng Dim_Products trong database DataCo_Warehouse.



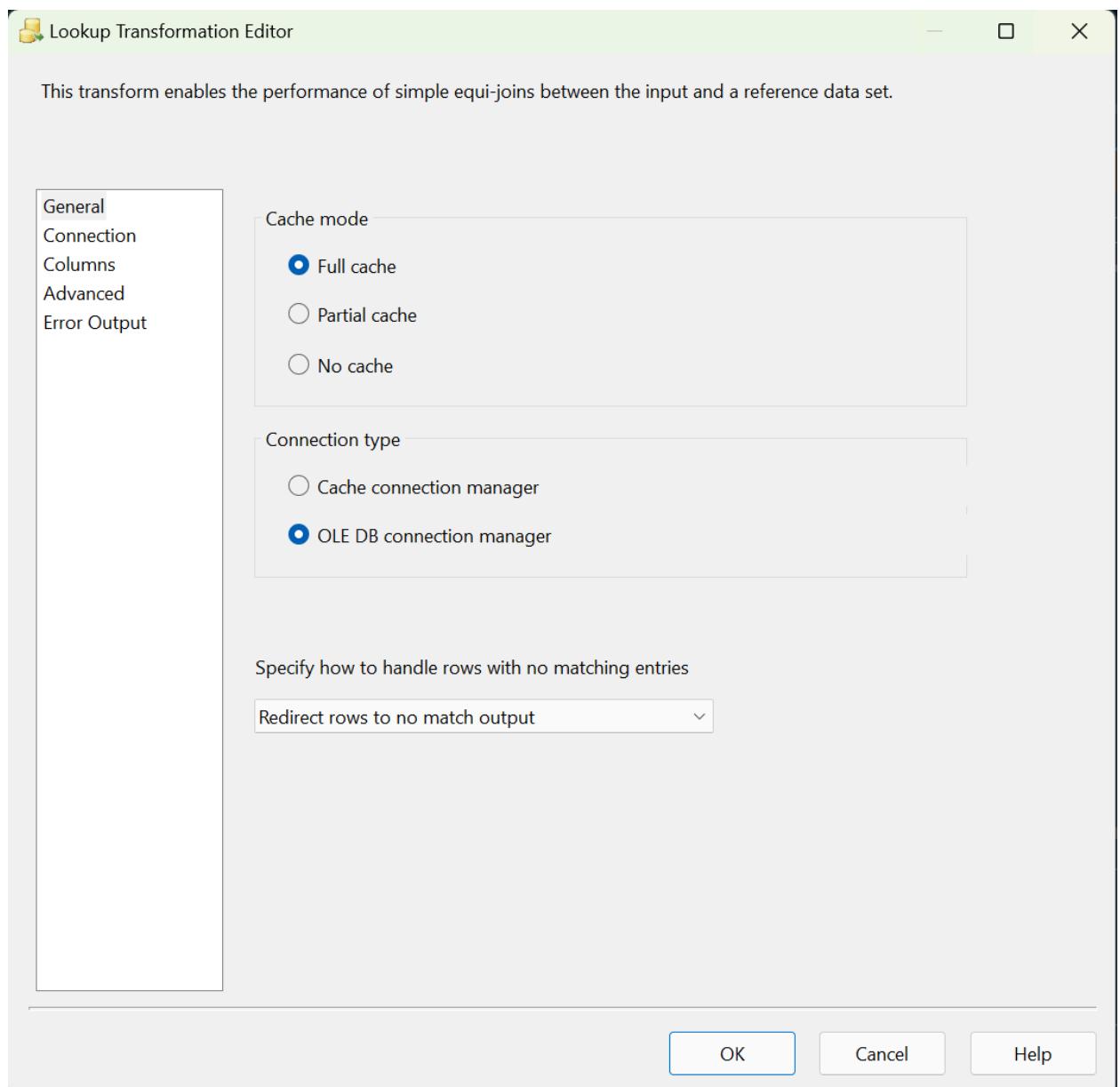
- Chuyển tab Column và map những thuộc tính cần thiết, dòng thời tích vào khóa chính.



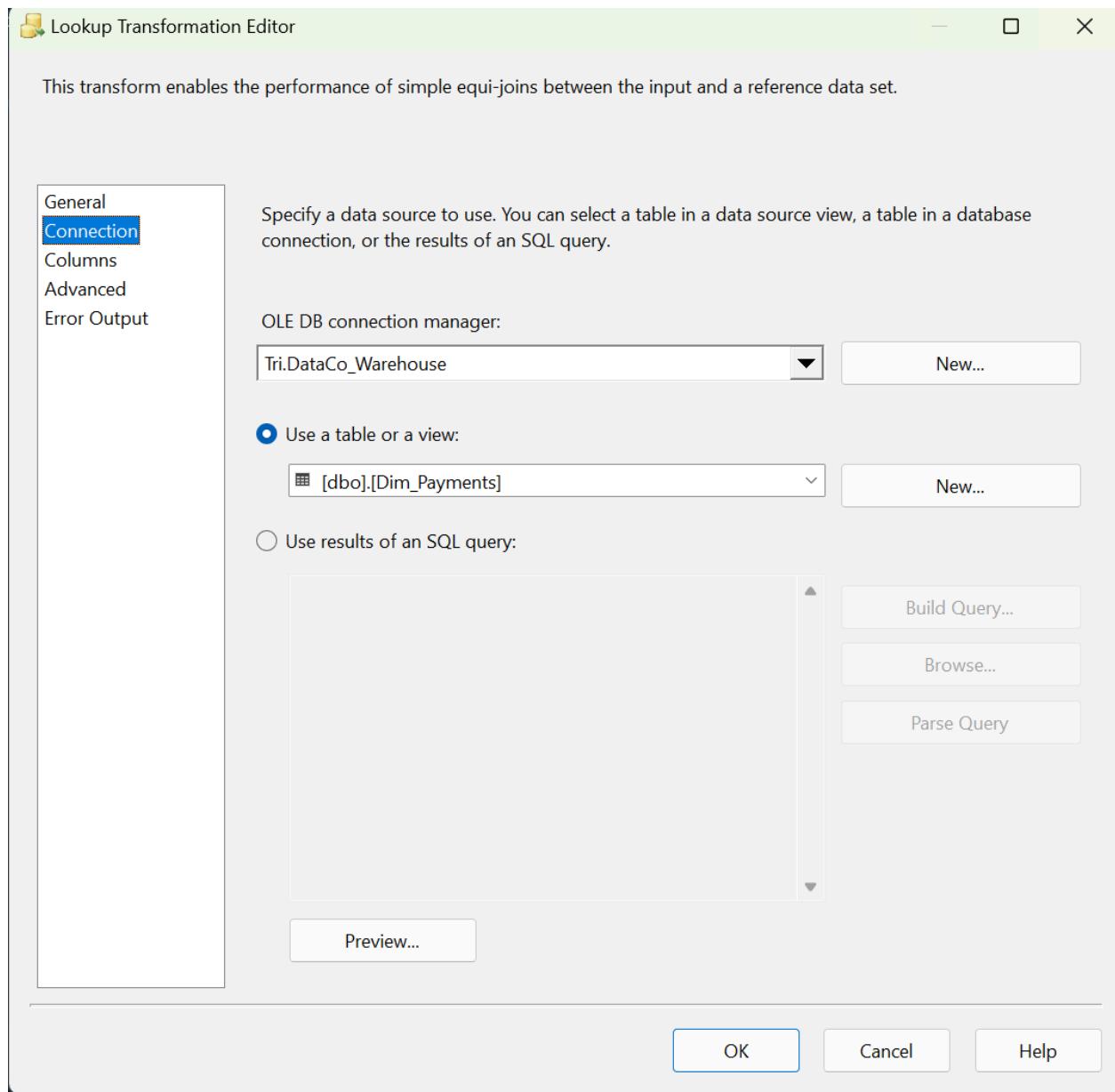
- Kéo thả công cụ Lookup



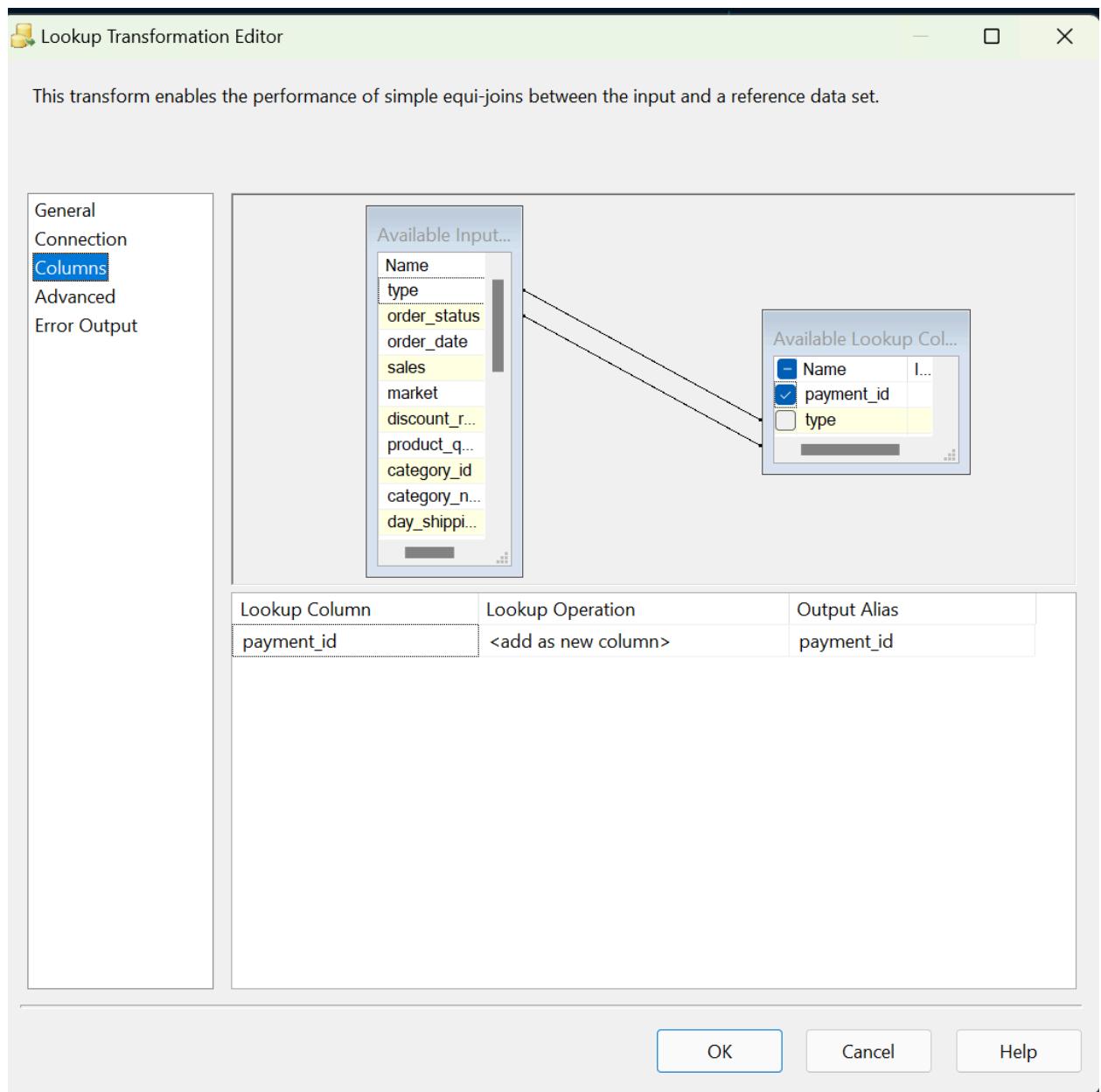
- Thực hiện edit công cụ



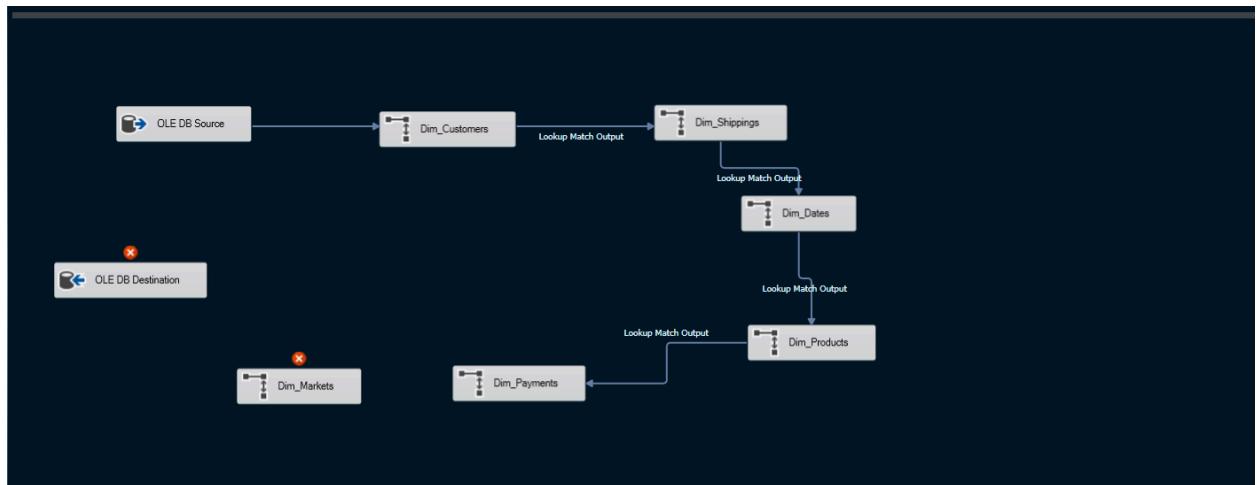
- Kết nối tới bảng Dim_Payments trong Warehouse



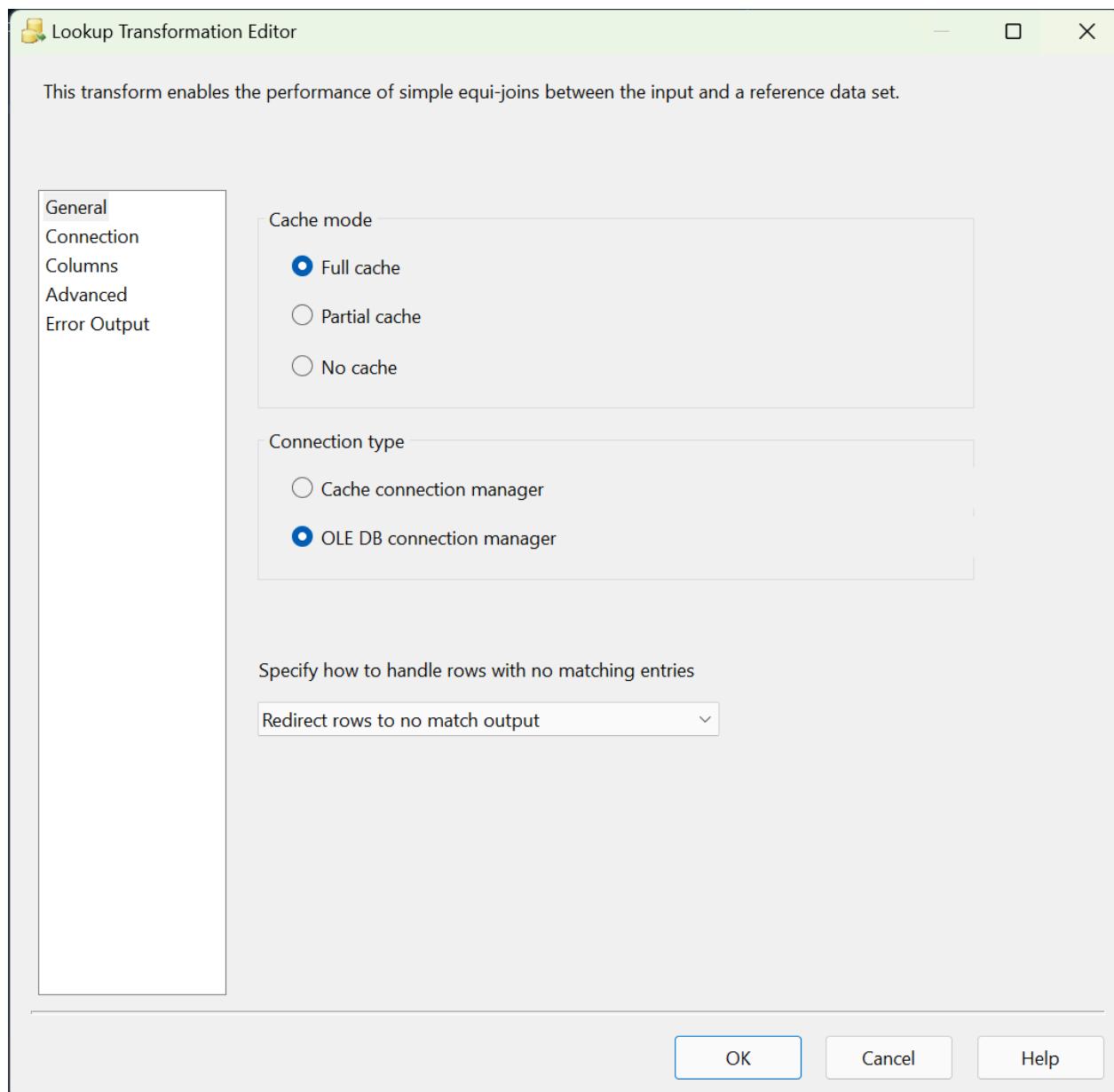
- Map các column và tích chọn thuộc tính cần thiết



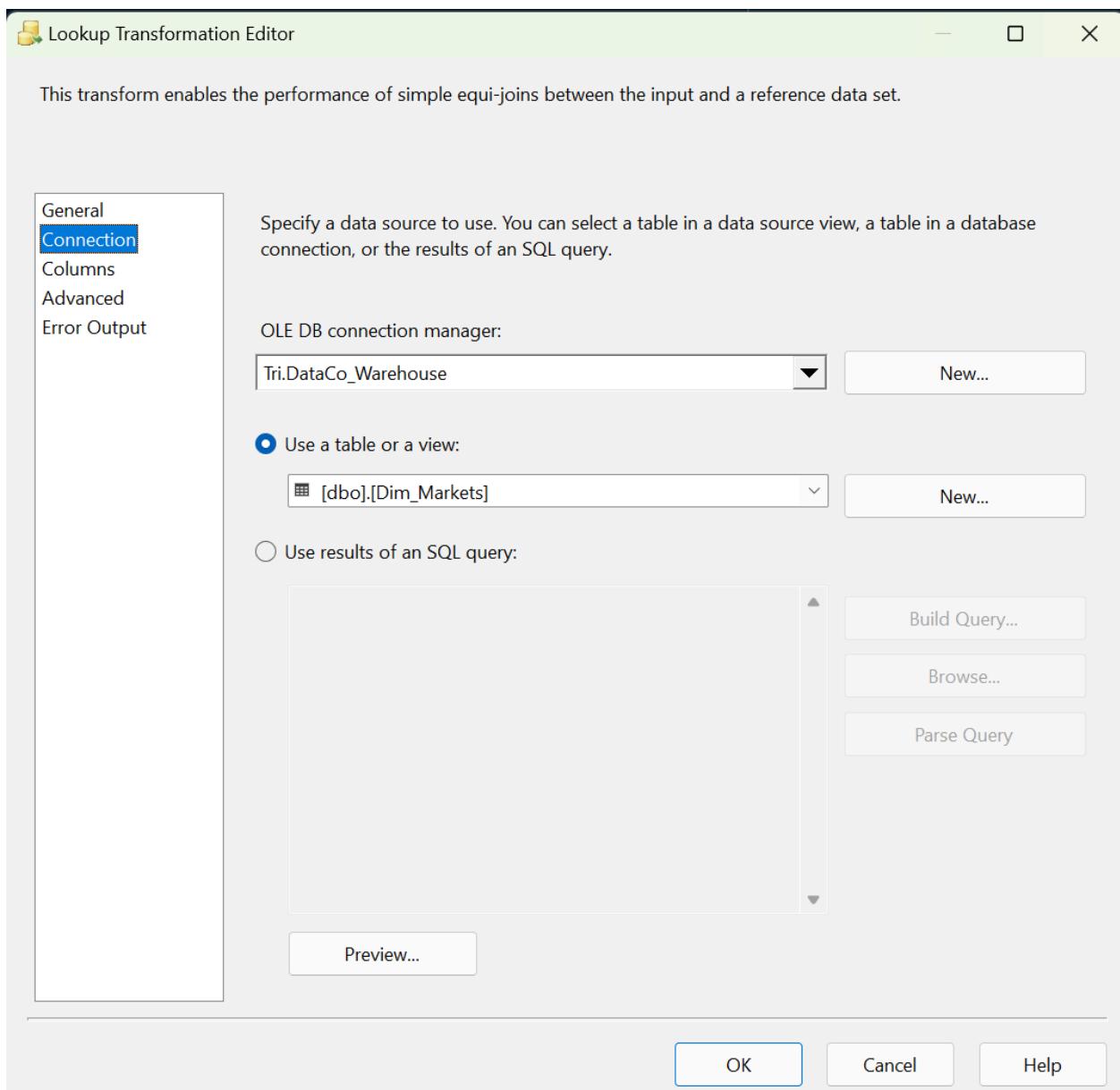
- Kéo thả công cụ Lookup và đổi tên thành Dim_Markets



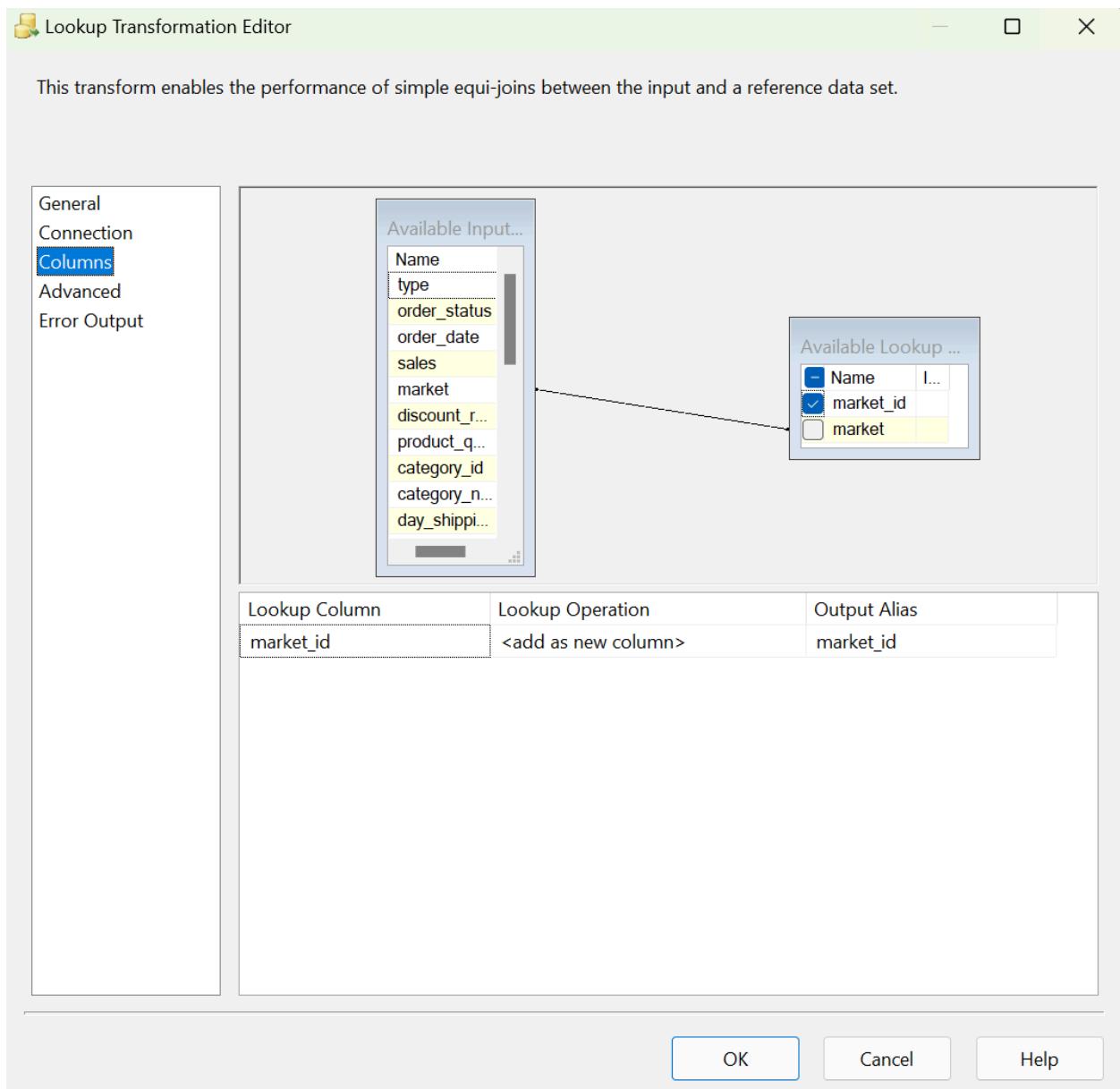
- Thực hiện edit công cụ



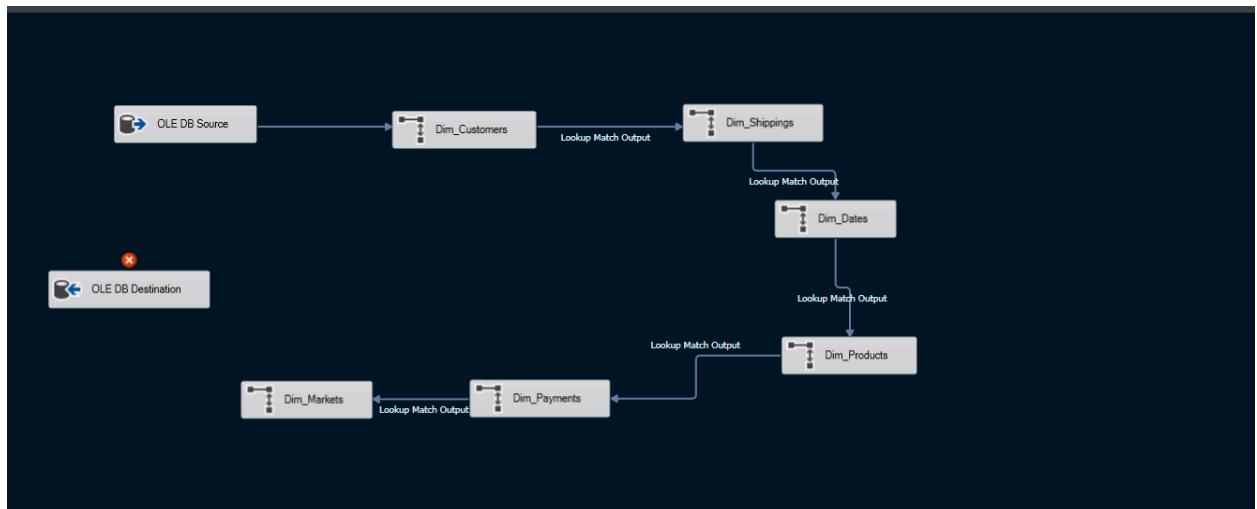
- Thực hiện kết nối tới bảng Dim_Markets



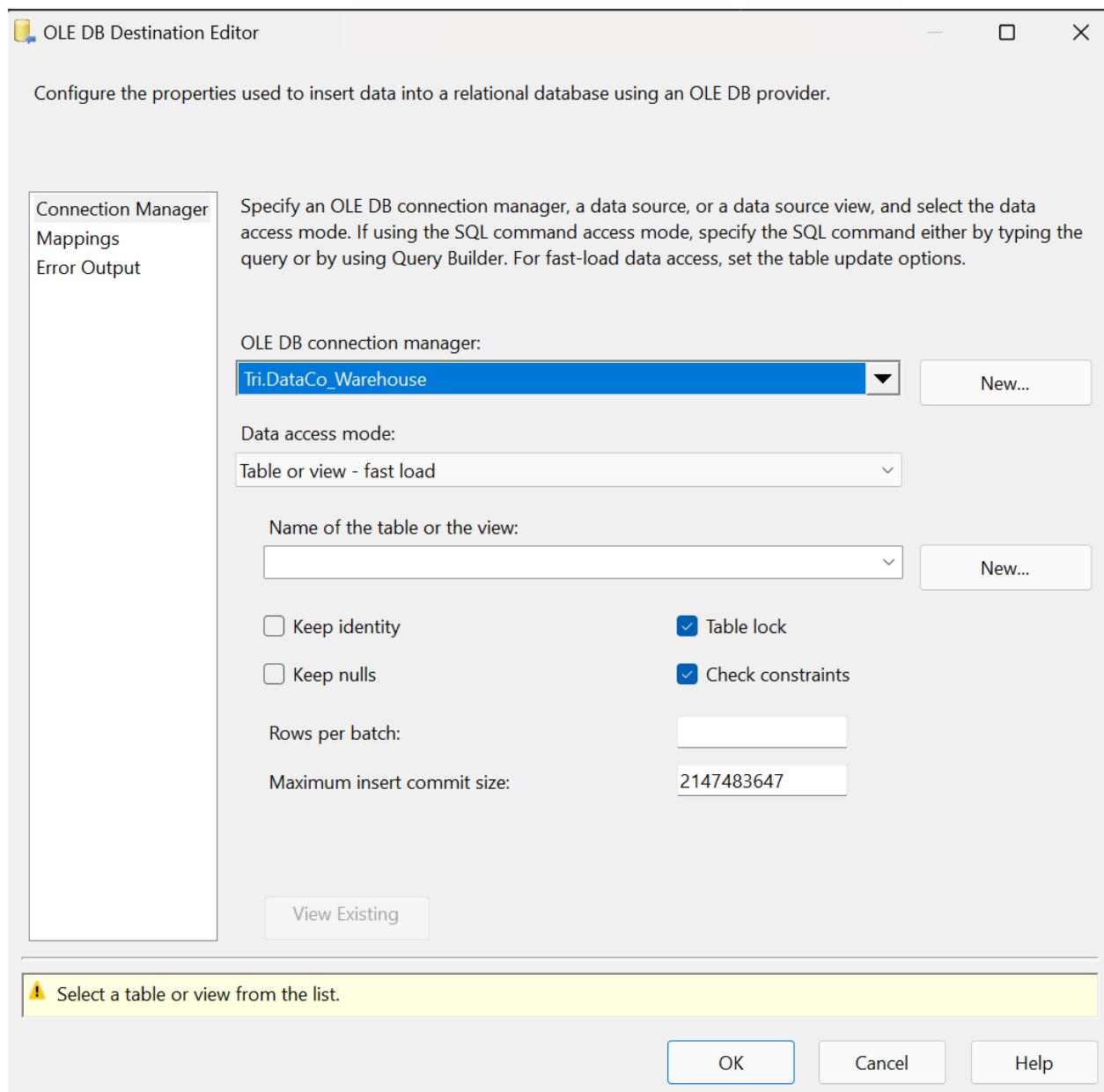
- Chọn vào tab column và map các thuộc tính



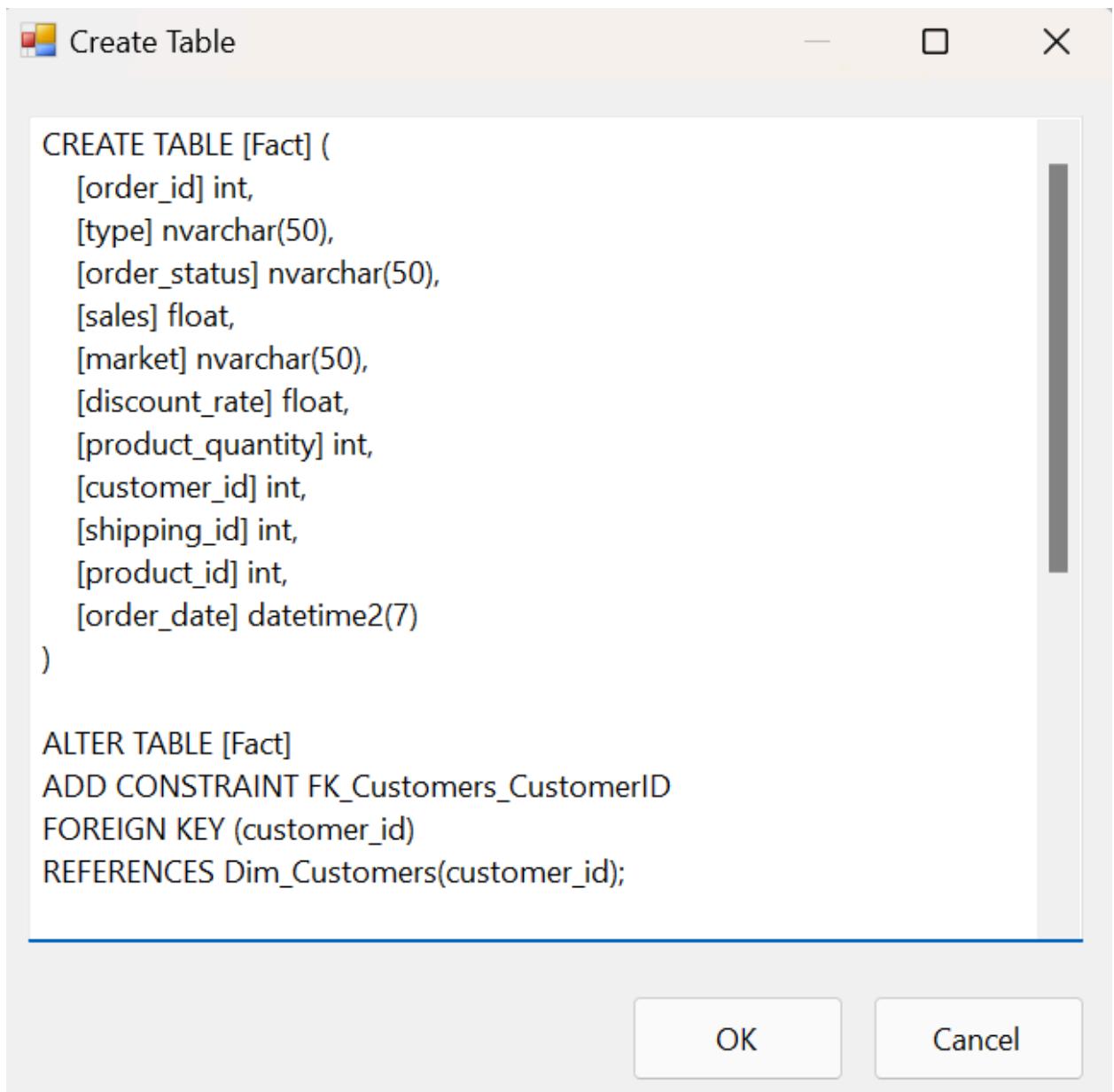
- Kéo thả OLE DB Destination từ Toolbox và edit.



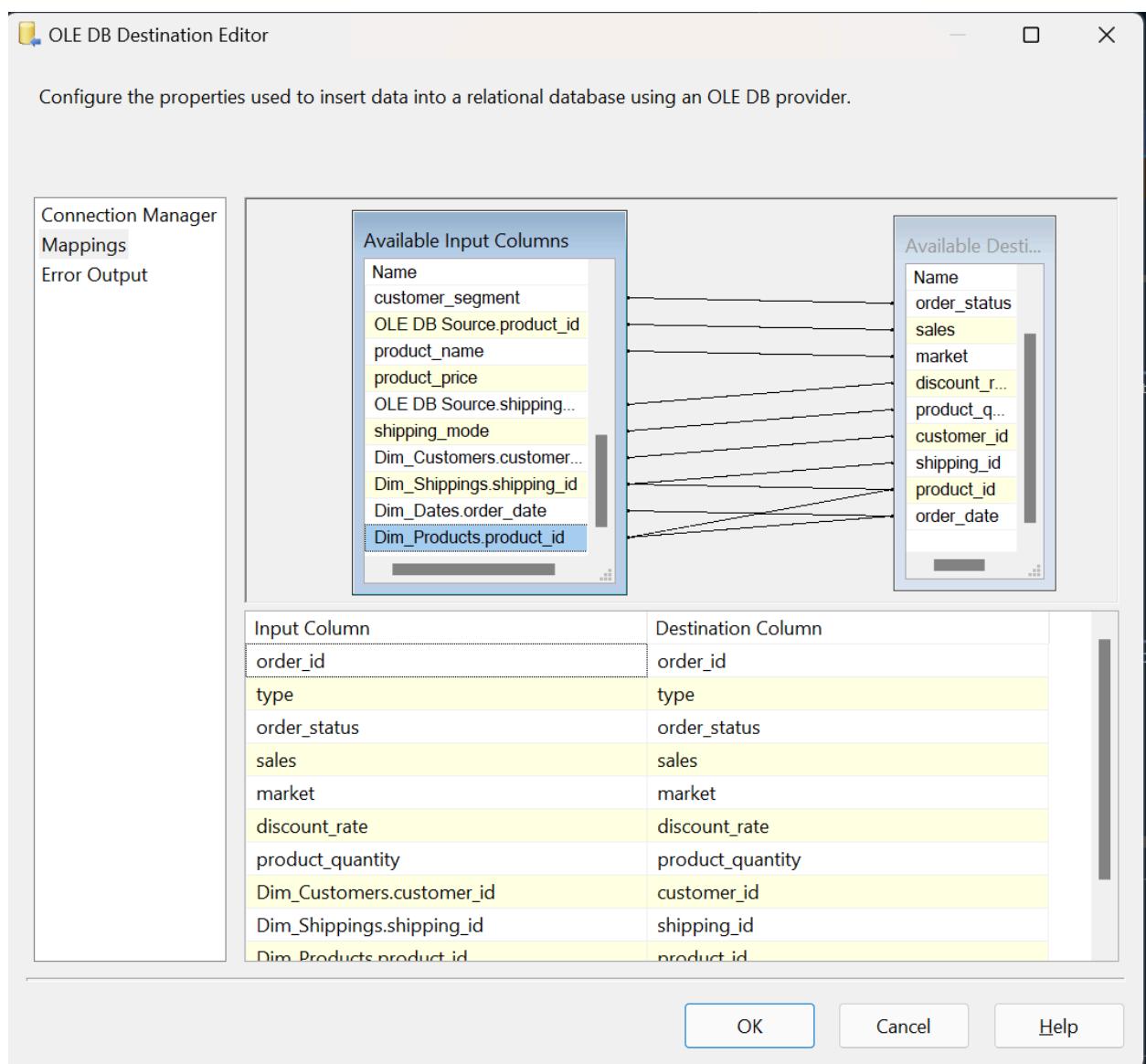
- Tiếp theo Edit OLE DB Source, thực hiện kết nối tới database DataCo_Warehouse



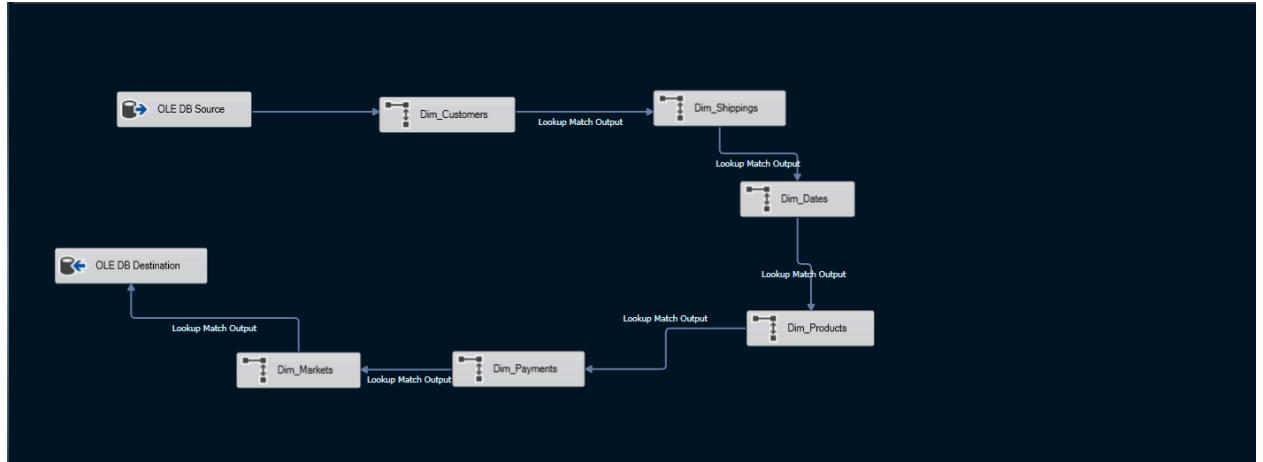
- Sau đó chọn New để tạo bảng Fact với các thuộc tính cần thiết



- Chuyển qua tab Mapping và map các thuộc tính cần thiết



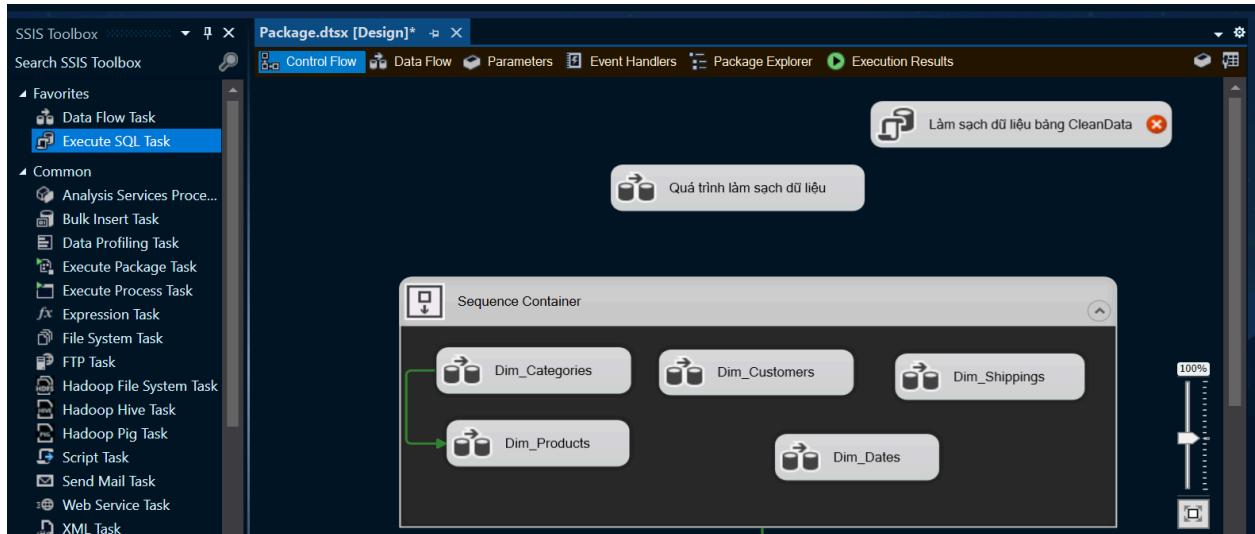
- Data Flow Task “Quá trình tạo bảng Fact” hoàn thành



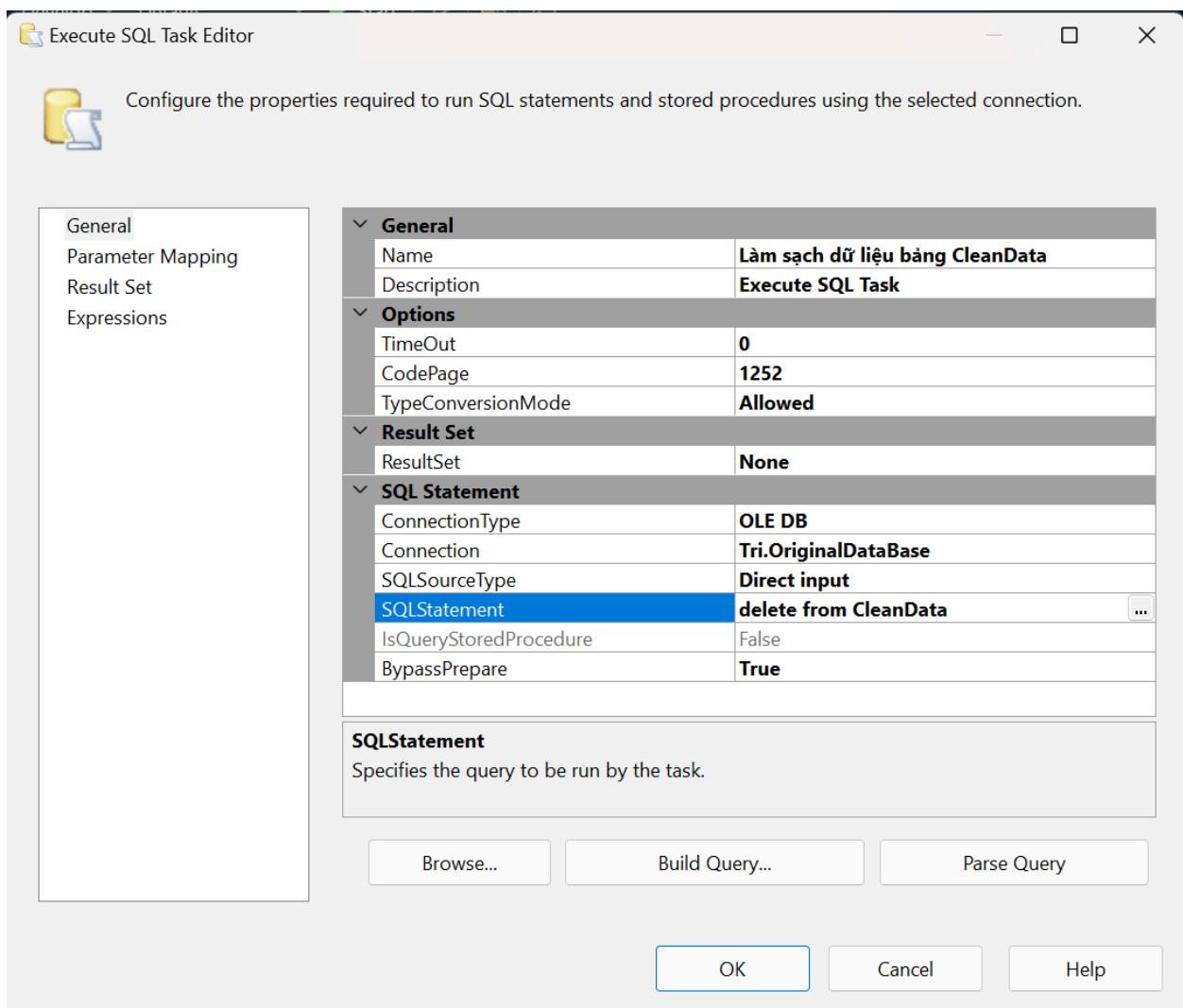
2.5. TẠO CÁC EXECUTE SQL TASK ĐỂ TRÁNH TRÙNG LẮP DỮ LIỆU

2.5.1 Tạo Execute SQL Task “Làm sạch bảng CleanData”

- Kéo thả công cụ Execute SQL Task từ Toolbox và đổi tên “Làm sạch dữ liệu bảng CleanData”

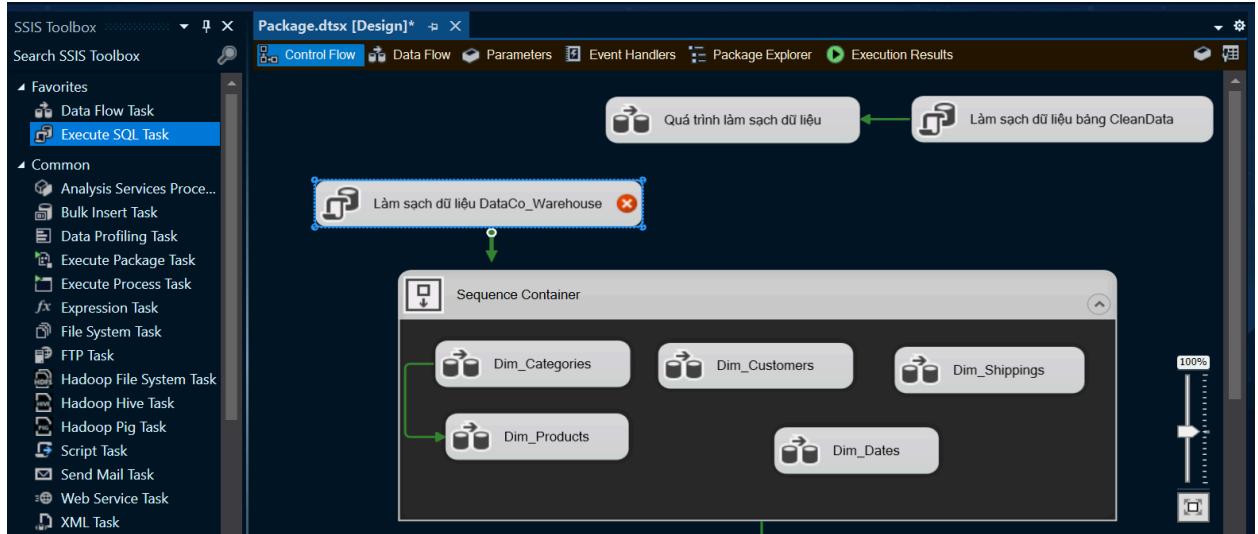


- Thực hiện edit Execute SQL Task “Làm sạch dữ liệu bảng CleanData”, kết nối tới database OriginalDatabase và viết lệnh xóa dữ liệu bảng “CleanData”

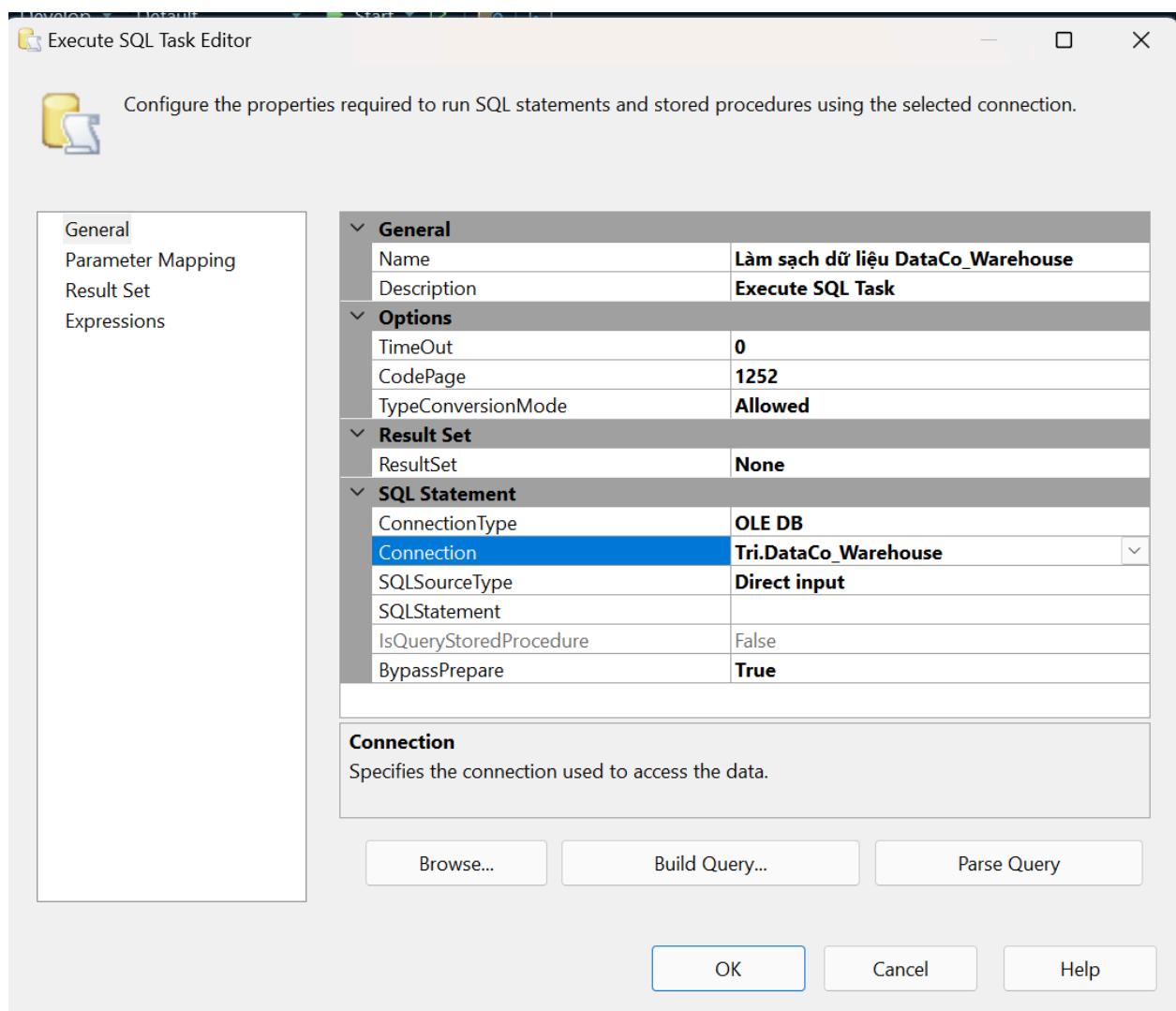


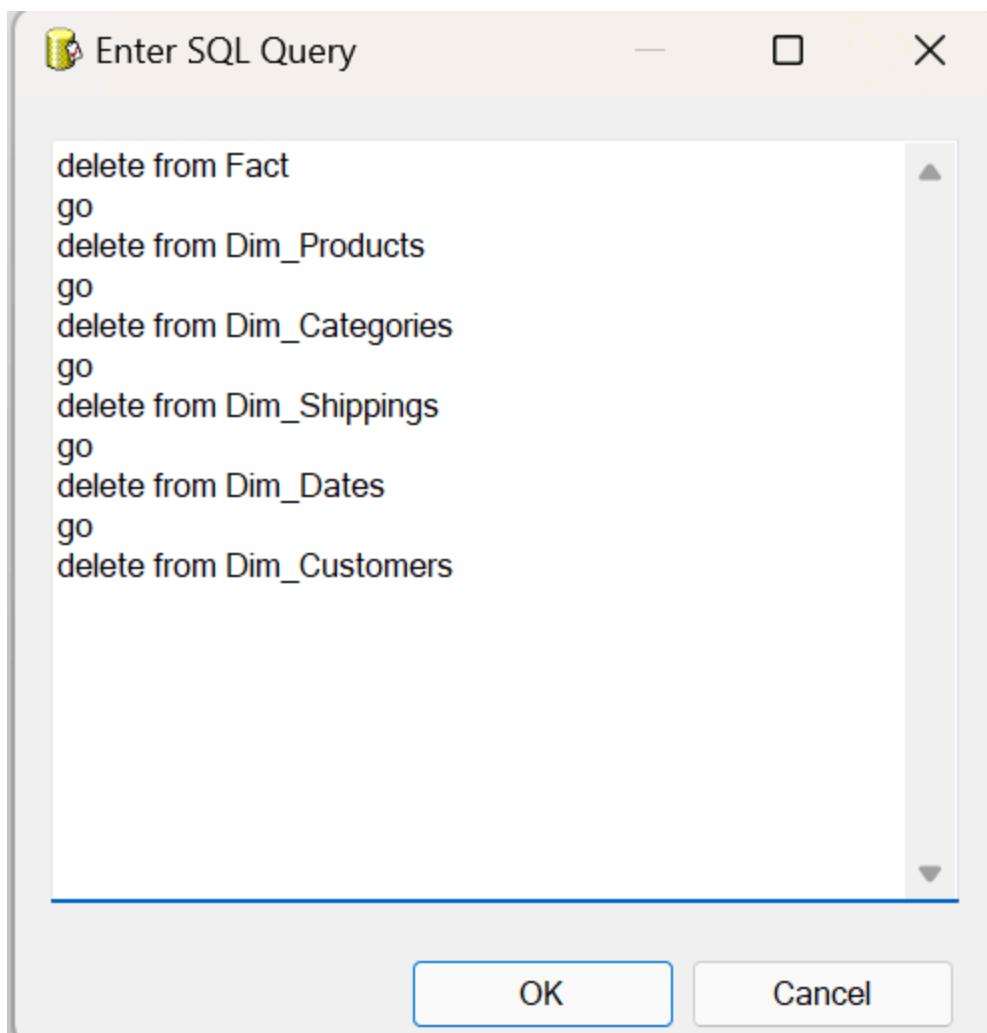
2.5.2. Tạo Execute SQL Task “Làm sạch dữ liệu DataCo_warehouse”

- Kéo thả công cụ Execute SQL Task và đổi tên thành “Làm sạch dữ liệu DataCo_Warehouse”



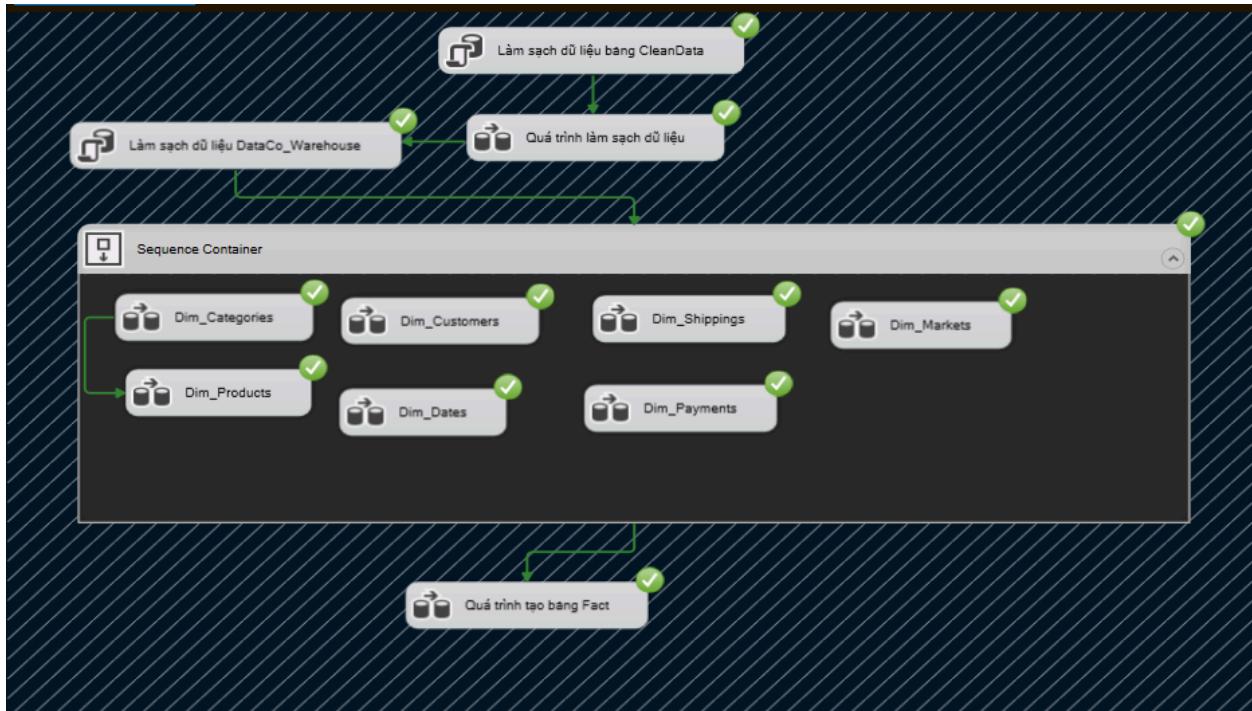
- Thực hiện edit Execute SQL Task “Làm sạch dữ liệu DataCo_Warehouse”, kết nối tới database DataCo_Warehouse và viết lệnh xóa dữ liệu các bảng





2.6. THỰC THI PACKAGE

2.6.1. Thực thi Package



2.6.2. Dữ liệu các bảng

- Bảng Dim_Customers

100 %

Results Messages

	customer_id	customer_country	customer_name	customer_segment
1	1	EE. UU.	Richard Hernandez	Consumer
2	2	EE. UU.	Mary Barrett	Consumer
3	3	Puerto Rico	Ann Smith	Consumer
4	4	EE. UU.	Mary Jones	Consumer
5	5	Puerto Rico	Robert Hudson	Home Office
6	6	EE. UU.	Mary Smith	Consumer
7	7	Puerto Rico	Melissa Wilcox	Corporate
8	8	EE. UU.	Megan Smith	Corporate
9	9	Puerto Rico	Mary Perez	Consumer
10	10	EE. UU.	Melissa Smith	Corporate
11	11	Puerto Rico	Mary Huffman	Consumer
12	12	EE. UU.	Christopher Smith	Corporate
13	13	Puerto Rico	Mary Baldwin	Home Office
14	14	EE. UU.	Katherine Smith	Corporate
15	15	EE. UU.	Jane Luna	Corporate
16	16	Puerto Rico	Tiffany Smith	Corporate
17	17	EE. UU.	Mary Robinson	Consumer
18	18	EE. UU.	Robert Smith	Consumer
19	19	Puerto Rico	Stephanie Mitchell	Home Office
20	20	EE. UU.	Mary Ellis	Consumer
21	21	Puerto Rico	William Zimmerman	Consumer
22	22	EE. UU.	Joseph Smith	Corporate
23	23	Puerto Rico	Benjamin Duarte	Corporate
24	24	Puerto Rico	Mary Smith	Home Office
25	25	EE. UU.	Paul Richardson	Home Office
26	26	EE. UU.	Johnny Hood	Consumer
27	27	Puerto Rico	Mary Vincent	Consumer
28	28	EE. UU.	Timothy Smith	Consumer

- Bảng Dim_Products

Results Messages

	product_id	product_name	product_price	category_id
1	19	Nike Men's Fingertrap Max Training Shoe	7999.35986328125	2
2	24	Elevation Training Mask 2.0	5919.25984191895	2
3	35	adidas Brazuca 2014 Official Match Ball	10399.3503570557	3
4	37	adidas Kids' F5 Messi FG Soccer Cleat	9167.3804397583	3
5	44	adidas Men's F10 Messi TRX FG Soccer Cleat	18296.9505119324	3
6	58	Diamondback Boys' Insight 24 Performance Hybr	8699.70971679688	4
7	60	SOLE E25 Elliptical	9999.89990234375	4
8	61	Diamondback Girls' Clarity 24 Hybrid Bike 201	8399.7197265625	4
9	78	Nike Kids' Grade School KD VI Basketball Shoe	6299.36986541748	5
10	93	Under Armour Men's Tech II T-Shirt	6997.19993591309	5
11	116	Nike Men's Comfort 2 Slide	14756.7205505371	6
12	127	Stiga Master Series ST3100 Competition Indoor	8909.72973632813	7
13	134	Nike Women's Legend V-Neck T-Shirt	6975	7
14	135	Nike Dri-FIT Crew Sock 6 Pack	6776	7
15	172	Nike Women's Tempo Shorts	9540	9
16	191	Nike Men's Free 5.0+ Running Shoe	1216778.28400421	9
17	203	GoPro HERO3+ Black Edition Camera	12799.6796875	10
18	208	SOLE E35 Elliptical	29999.8498535156	10
19	216	Yakima DoubleDown Ace Hitch Mount 4-Bike ...	12096	10
20	226	Bowflex SelectTech 1090 Dumbbells	5999.89990234375	11
21	235	Under Armour Hustle Storm Medium Duffle Bag	10462.0105018616	11
22	249	Under Armour Women's Micro G Skulpt Runnin...	15776.3903503418	12
23	251	Brooks Women's Ghost 6 Running Shoe	5939.33985900879	12

- 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!
13	17	Cleats
14	18	Men's Footwear
15	24	Women's Apparel
16	26	Girls' Apparel
17	29	Shop By Sport
18	30	Men's Golf Clubs
19	31	Women's Golf Clu...
20	32	Golf Apparel
21	33	Golf Shoes
22	34	Golf Bags & Carts
23	35	Golf Gloves

- Bảng Dim_Shippings

	shipping_id	shipping_mode	day_shipping_real	day_shipment_scheduled	delivery_status
1	1	Standard Class	3	4	Advance shipping
2	2	Standard Class	5	4	Late delivery
3	3	Standard Class	4	4	Shipping on time
4	4	Standard Class	3	4	Advance shipping
5	5	Standard Class	2	4	Advance shipping
6	6	Standard Class	6	4	Shipping canceled
7	7	First Class	2	1	Late delivery
8	8	First Class	2	1	Late delivery
9	9	Second Class	3	2	Late delivery
10	10	First Class	2	1	Late delivery
11	11	Second Class	6	2	Shipping canceled
12	12	Second Class	5	2	Late delivery
13	13	Second Class	4	2	Late delivery
14	14	First Class	2	1	Late delivery
15	15	First Class	2	1	Late delivery
16	16	First Class	2	1	Late delivery
17	17	Second Class	5	2	Late delivery
18	18	First Class	2	1	Late delivery
19	19	First Class	2	1	Late delivery
20	20	Same Day	0	0	Shipping on time
21	21	Same Day	0	0	Shipping on time
22	22	Standard Class	5	4	Late delivery
23	23	Second Class	4	3	Late delivery

- Bảng Dim_Payments

	payment_id	type	order_status
1	1	DEBIT	COMPLETE
2	2	TRANSFER	PENDING
3	3	CASH	CLOSED
4	4	DEBIT	COMPLETE
5	5	PAYMENT	PENDING_PAYMENT
6	6	TRANSFER	CANCELED
7	7	DEBIT	COMPLETE
8	8	TRANSFER	PROCESSING
9	9	CASH	CLOSED
10	10	CASH	CLOSED
11	11	TRANSFER	SUSPECTED_FRAUD
12	12	TRANSFER	PENDING
13	13	TRANSFER	PENDING

- Bảng Dim_Markets

Results		Messages	
	market_id	market	
1	1	Pacific Asia	
2	2	USCA	
3	3	Africa	
4	4	Europe	
5	5	LATAM	

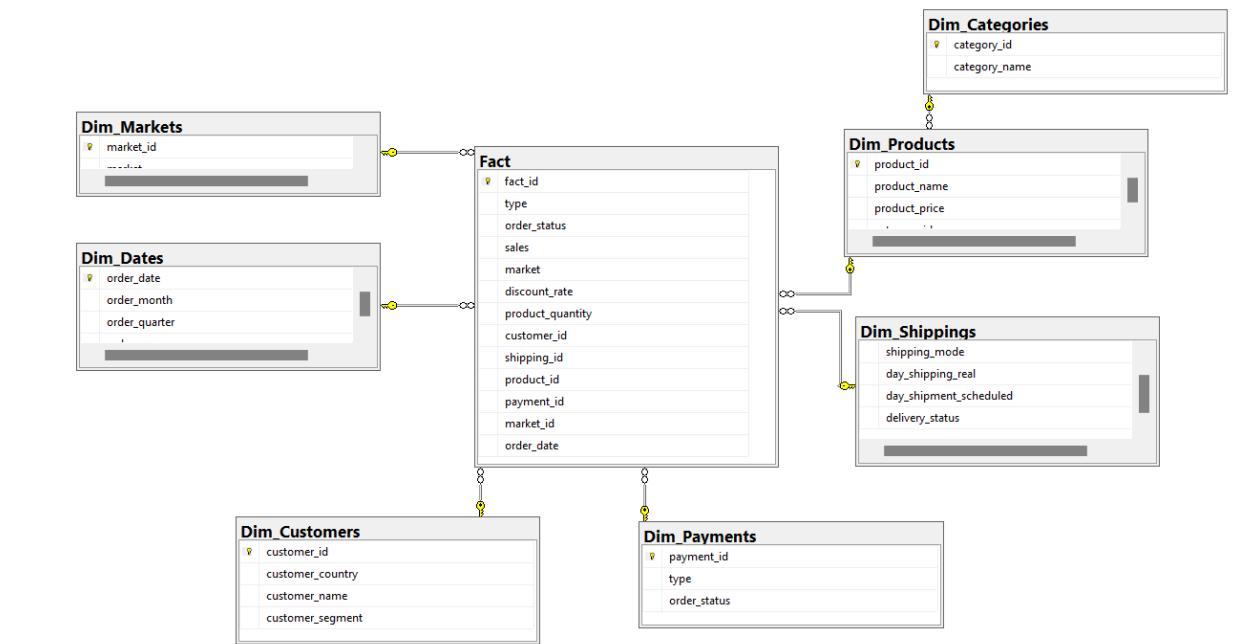
- Bảng Dim_Dates

Results		Messages		
	order_date	order_month	order_quarter	order_year
1	2015-01-01 00:00:00.000000	1	1	2015
2	2015-01-01 00:21:00.000000	1	1	2015
3	2015-01-01 01:03:00.000000	1	1	2015
4	2015-01-01 01:24:00.000000	1	1	2015
5	2015-01-01 02:06:00.000000	1	1	2015
6	2015-01-01 02:27:00.000000	1	1	2015
7	2015-01-01 02:48:00.000000	1	1	2015
8	2015-01-01 03:09:00.000000	1	1	2015
9	2015-01-01 03:30:00.000000	1	1	2015
10	2015-01-01 03:51:00.000000	1	1	2015
11	2015-01-01 04:12:00.000000	1	1	2015
12	2015-01-01 04:33:00.000000	1	1	2015
13	2015-01-01 04:54:00.000000	1	1	2015
14	2015-01-01 05:15:00.000000	1	1	2015
15	2015-01-01 05:36:00.000000	1	1	2015
16	2015-01-01 05:57:00.000000	1	1	2015
17	2015-01-01 06:18:00.000000	1	1	2015
18	2015-01-01 06:39:00.000000	1	1	2015
19	2015-01-01 07:00:00.000000	1	1	2015
20	2015-01-01 07:42:00.000000	1	1	2015
21	2015-01-01 08:03:00.000000	1	1	2015
22	2015-01-01 08:24:00.000000	1	1	2015
23	2015-01-01 09:06:00.000000	1	1	2015

- Bảng Fact

fact_id	type	order_status	sales	market	discount_rate	product_quantity	customer_id	shipping_id	product_id	
1	360918	CASH	CLOSED	299.980010986328	LATAM	0.200000002980232	1	11599	33834	957
2	360919	CASH	CLOSED	49.9799995422363	LATAM	0.180000007152557	2	8827	63763	897
3	360920	CASH	CLOSED	150	LATAM	0.180000007152557	3	8827	95935	502
4	360921	CASH	CLOSED	199.919998168945	LATAM	0.200000002980232	4	8827	114911	1014
5	360922	CASH	CLOSED	299.950012207031	LATAM	0.0099999977648258	5	8827	96128	365
6	360923	CASH	CLOSED	100	LATAM	0.129999995231628	4	1837	5695	134
7	360924	CASH	CLOSED	149.940002441406	LATAM	0.119999997317791	3	1837	114905	1014
8	360925	CASH	CLOSED	250	LATAM	0.0700000002980232	5	1837	141062	502
9	360926	CASH	CLOSED	299.980010986328	LATAM	0.150000005960464	1	1837	140734	957
10	360927	CASH	CLOSED	499.950012207031	LATAM	0.019999995529652	5	1837	141051	191
11	360928	CASH	CLOSED	119.980003356934	LATAM	0.0599999986588955	2	1205	23829	365
12	360929	CASH	CLOSED	129.990005493164	LATAM	0.129999995231628	1	1205	23252	403
13	360930	CASH	CLOSED	199.990005493164	LATAM	0.159999996423721	1	1205	77002	1073
14	360931	CASH	CLOSED	50	LATAM	0.200000002980232	1	11441	16342	502
15	360932	CASH	CLOSED	129.990005493164	LATAM	0.100000001490116	1	11441	16310	403
16	360933	CASH	CLOSED	199.990005493164	LATAM	0.129999995231628	1	11441	77000	1073
17	360934	CASH	CLOSED	199.990005493164	LATAM	0.150000005960464	1	11441	77001	1073
18	360935	CASH	CLOSED	250	LATAM	0.0500000007450581	5	11441	16911	502
19	360936	CASH	CLOSED	399.980010986328	LATAM	0.150000005960464	1	9503	62999	1004
20	360937	CASH	CLOSED	59.9900016784668	LATAM	0.170000001788139	1	5863	18108	365
21	360938	CASH	CLOSED	99.9599990844727	LATAM	0.119999997317791	4	5863	18415	906
22	360939	CASH	CLOSED	50	LATAM	0.180000007152557	1	12271	33712	502
23	360940	CASH	CLOSED	200.000010008229	LATAM	0.0500000026589055	1	12271	22771	1004

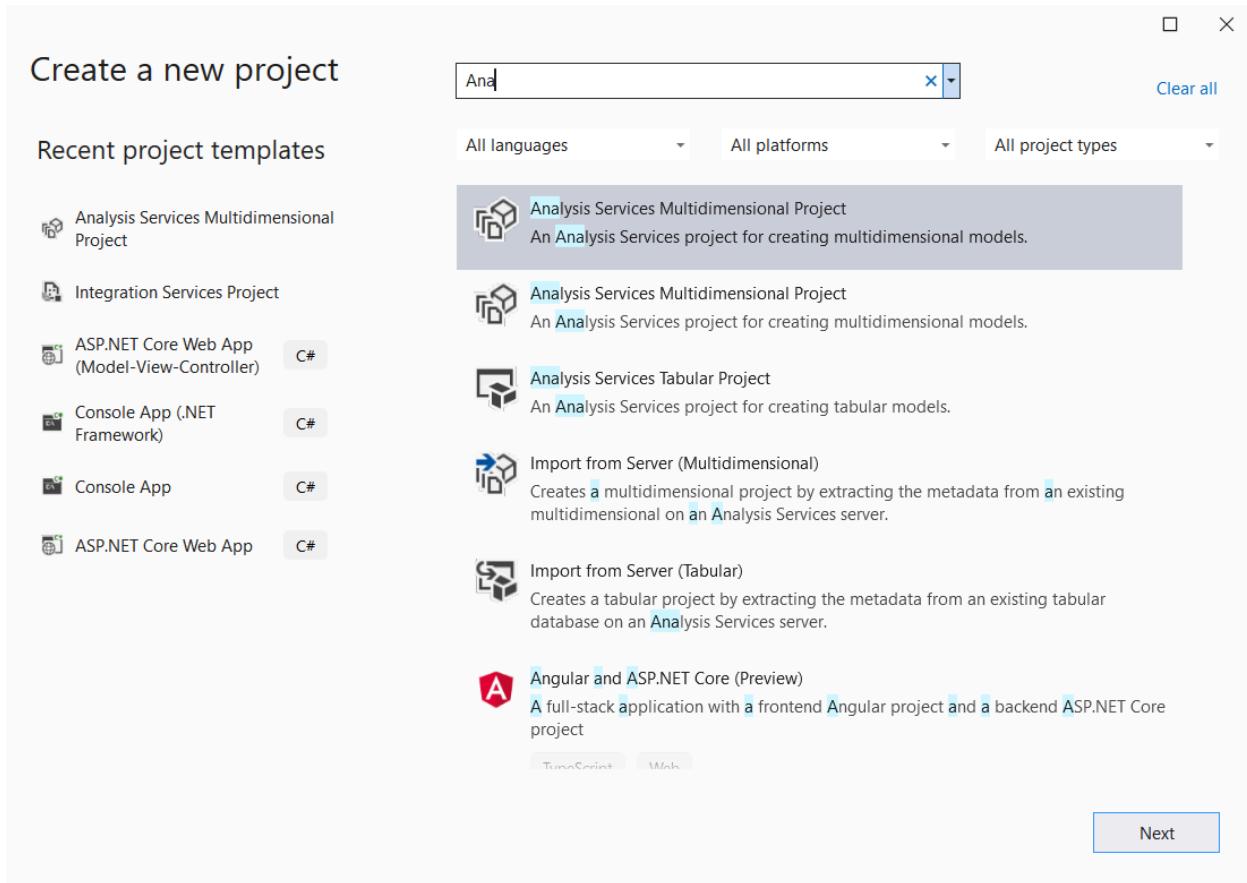
2.6. DIAGRAM SAU KHI HOÀN THÀNH



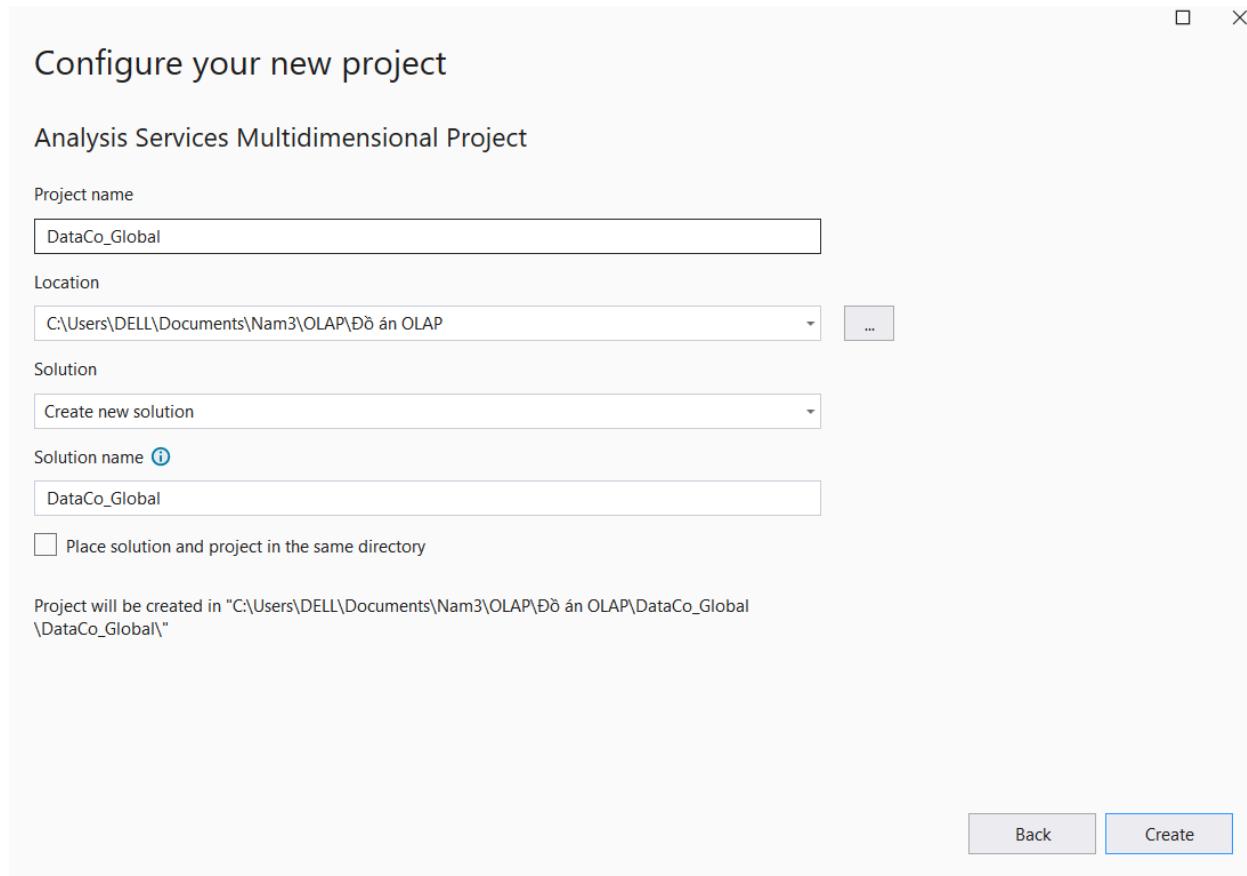
Chương 3: PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS)

3.1. TẠO PROJECT SSAS

- Chọn File -> New -> Project -> Analysis Services

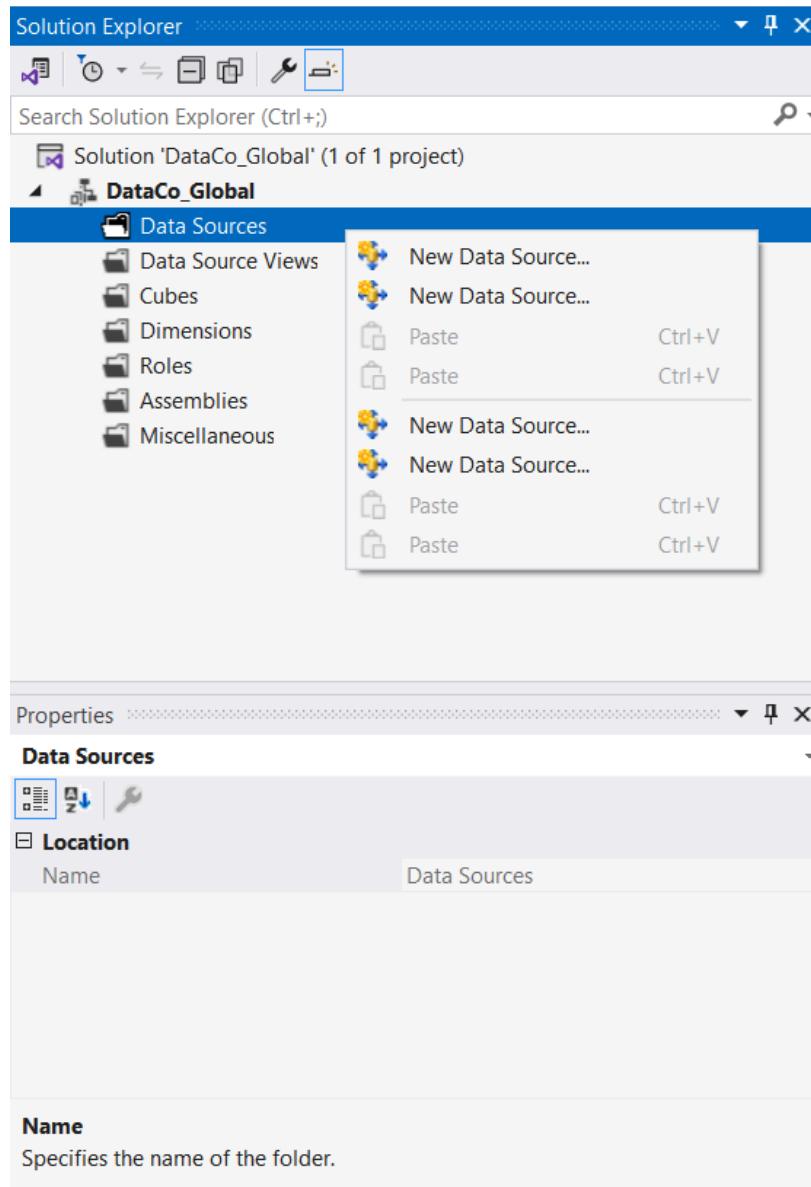


- Đặt tên cho Project và chọn nơi lưu trữ -> Create

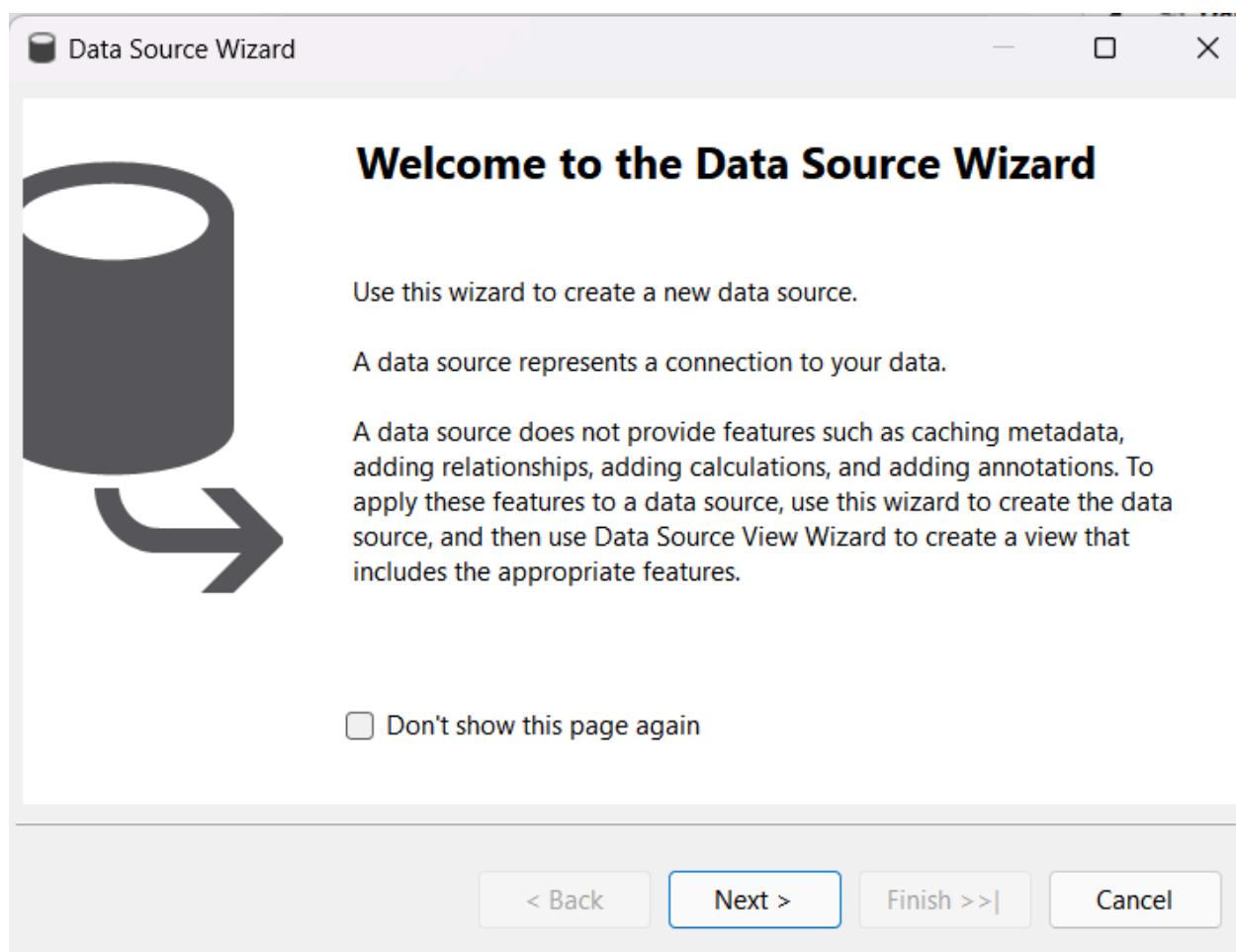


3.2. TẠO DATA SOURCE

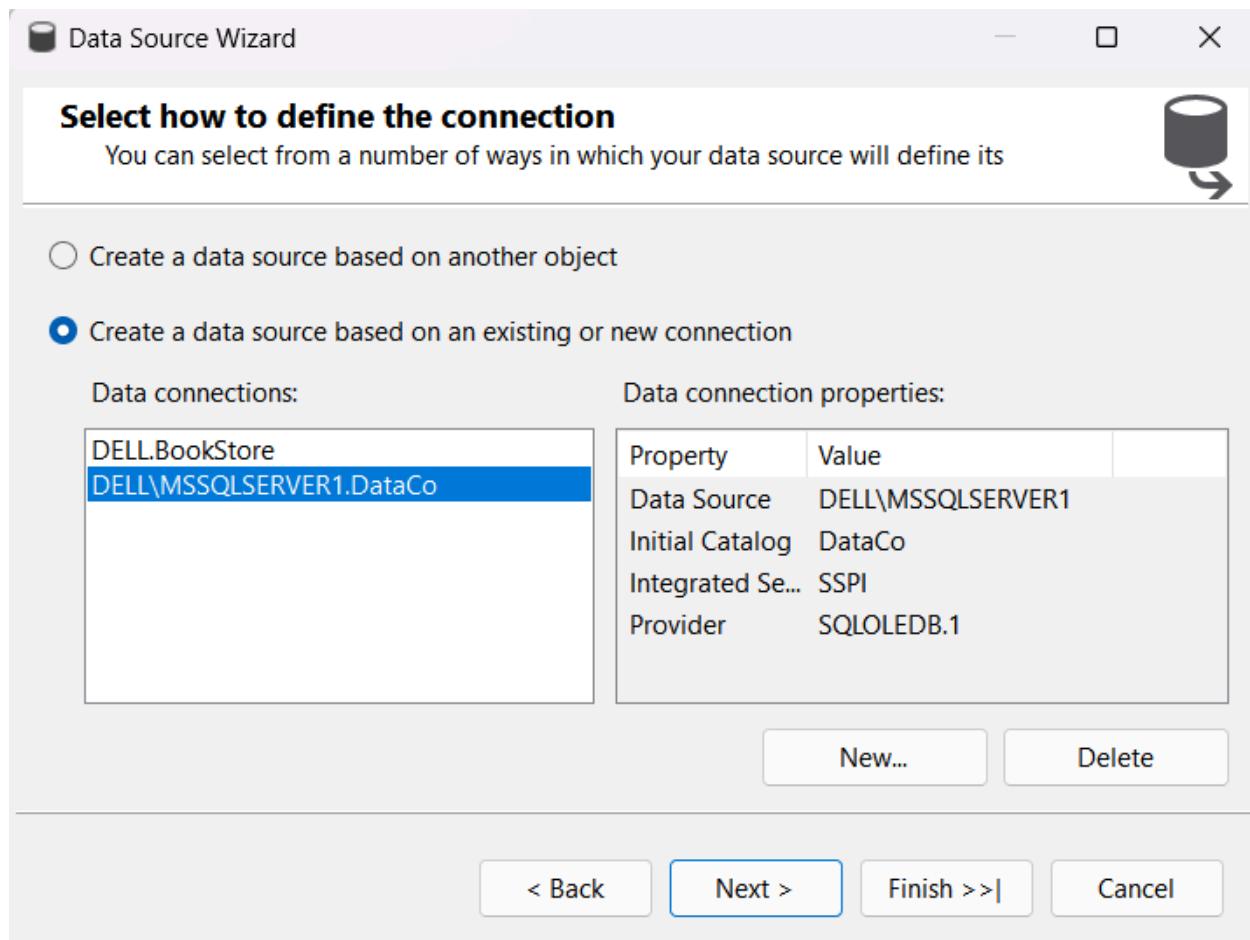
- Ở tab Solution Explorer bên phải màn hình -> **Data Sources** -> **New Data Sources**



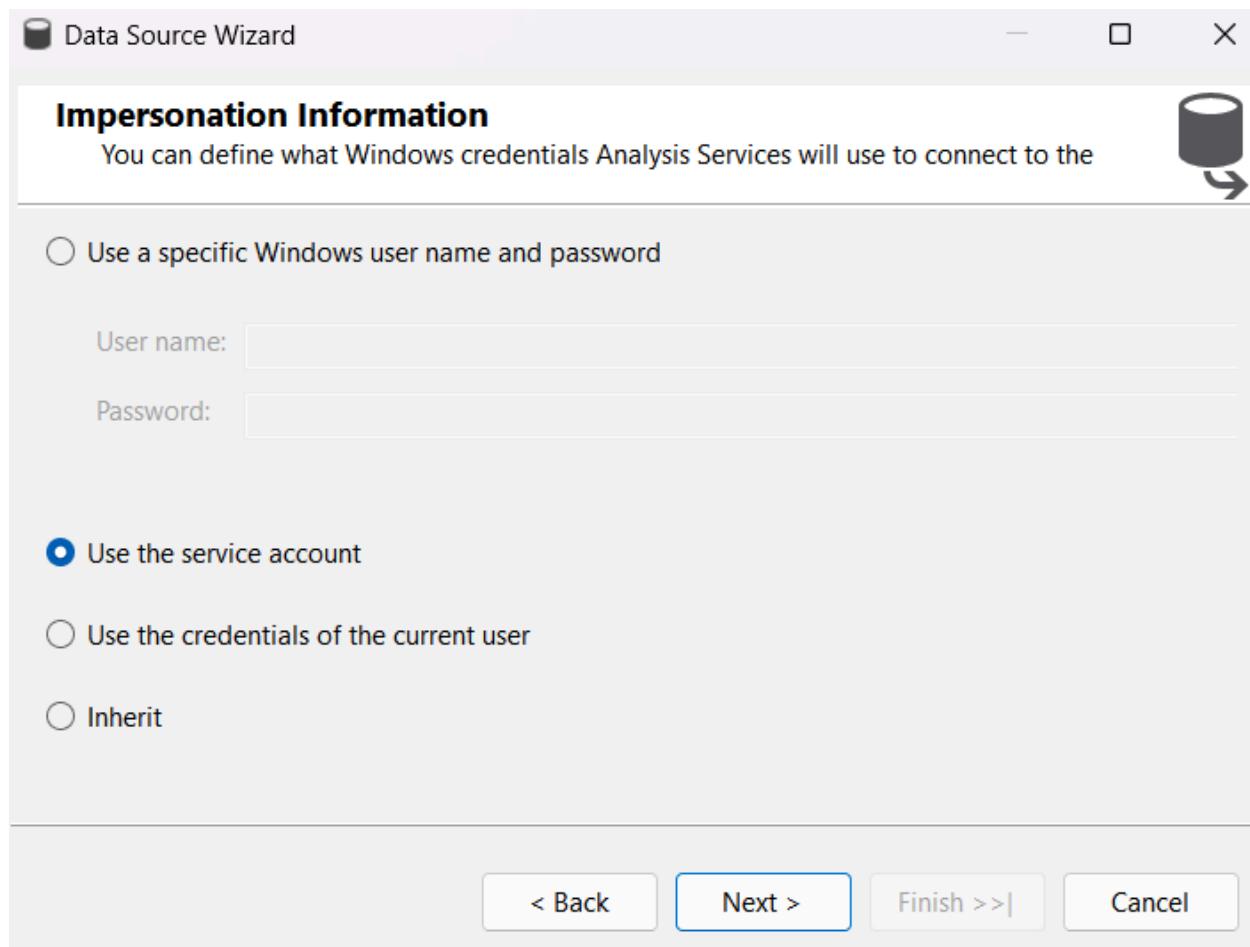
- Cửa sổ **Data Source Wizard** xuất hiện -> **Next**



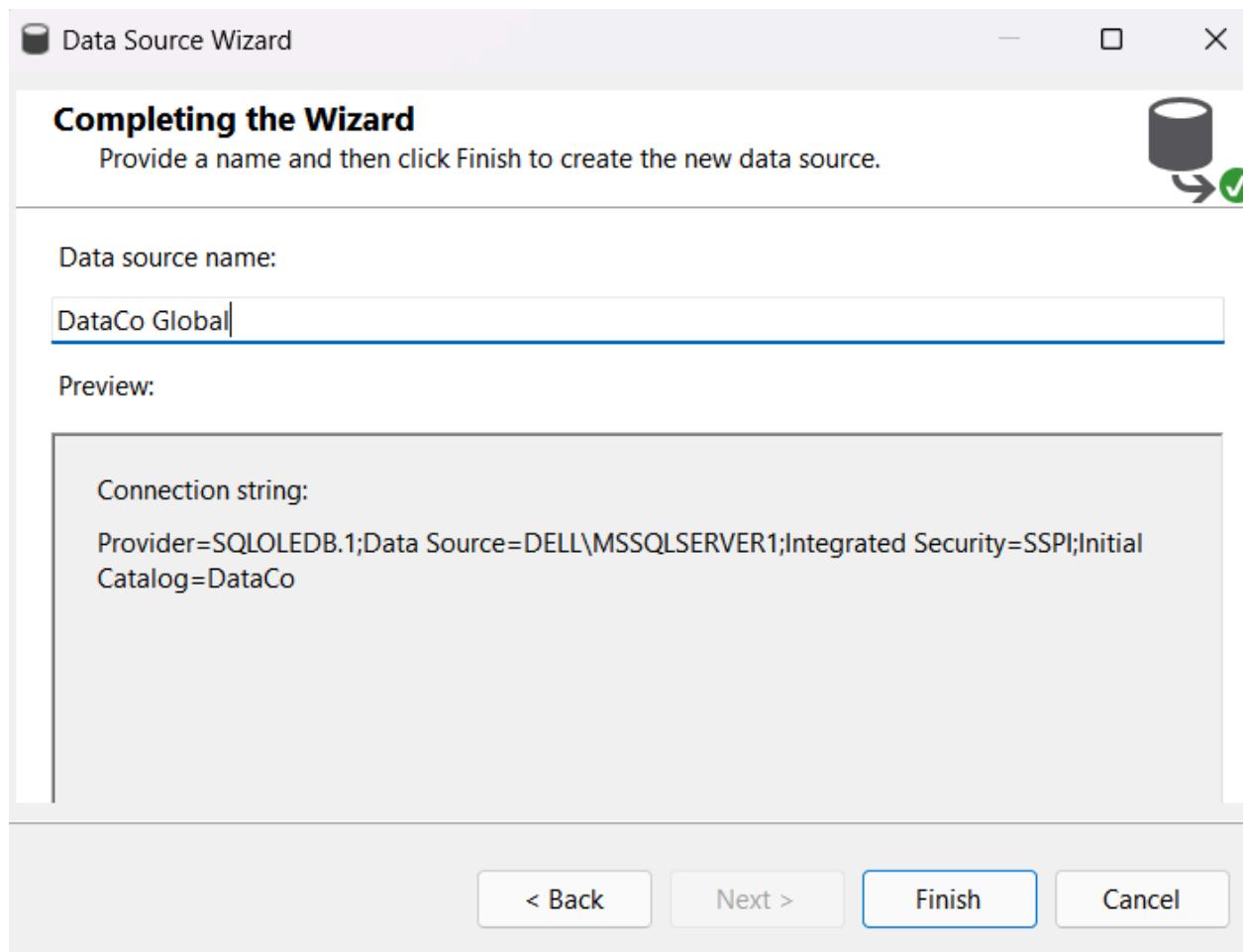
- Chọn connection đã có sẵn hoặc tạo mới connection -> **Next**



- Chọn Use the service account -> Next

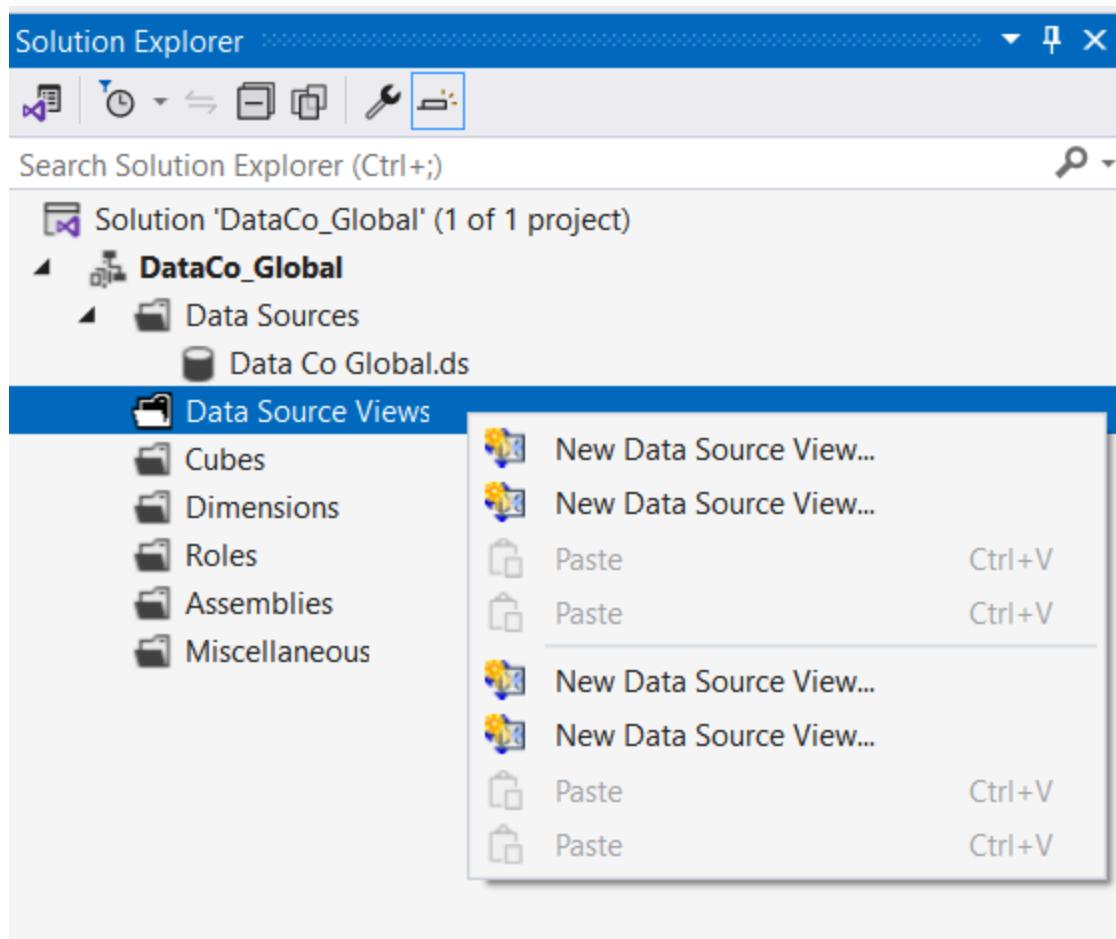


- Đặt tên cho Data Source -> Finish

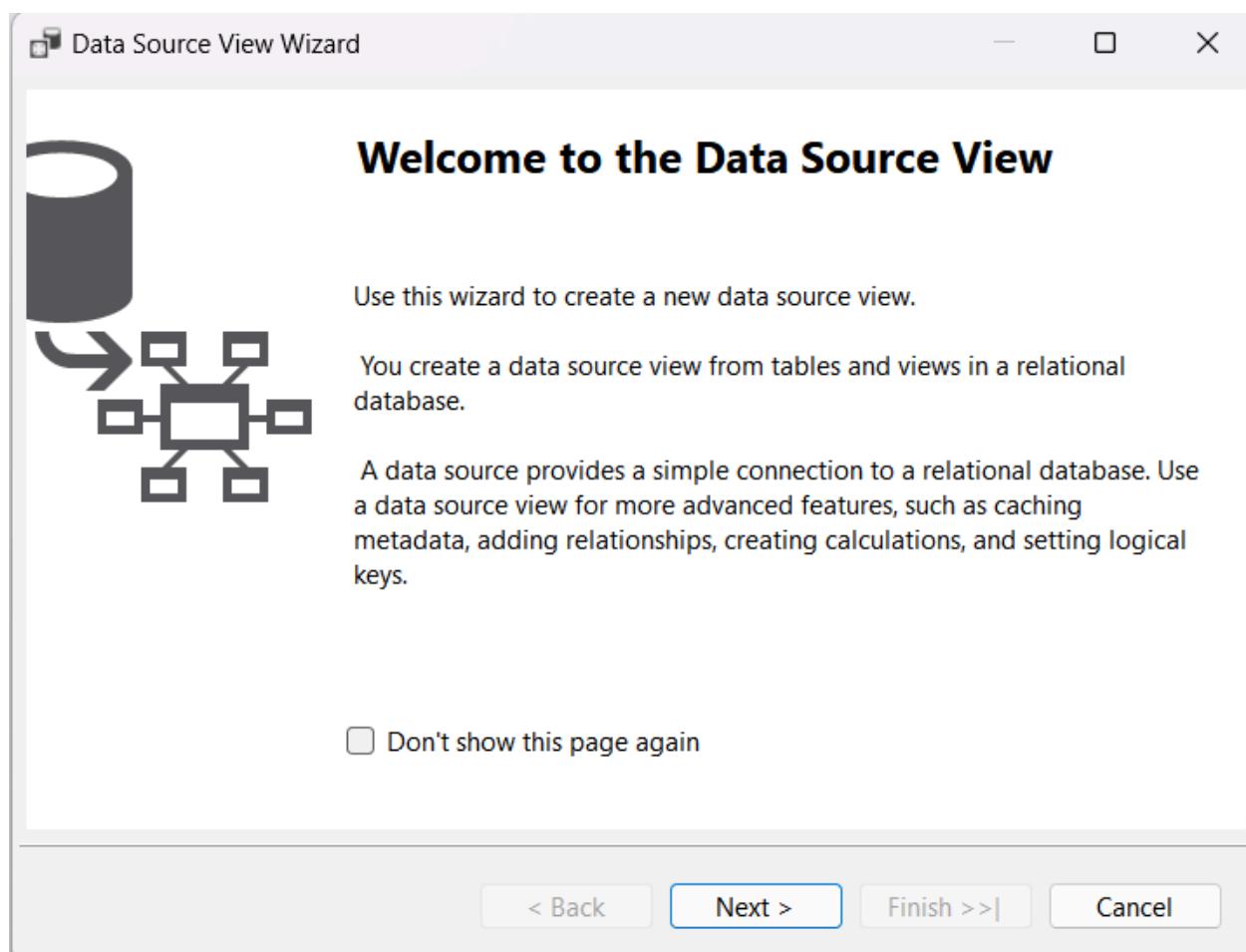


3.3. TẠO DATA SOURCE VIEW

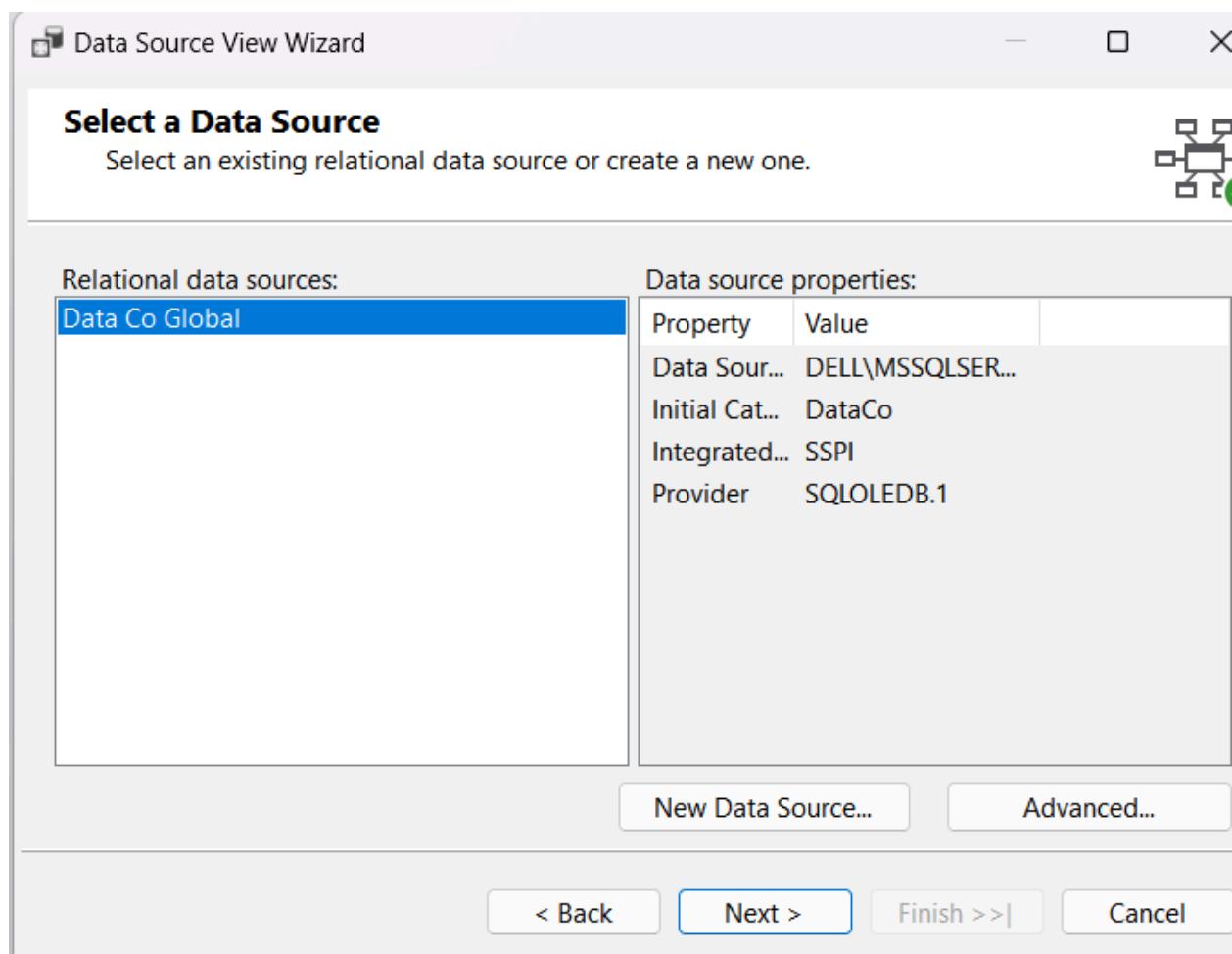
- Ở tab Solution Explorer bên phải màn hình -> **Data Sources View** -> **New Data Sources View**.



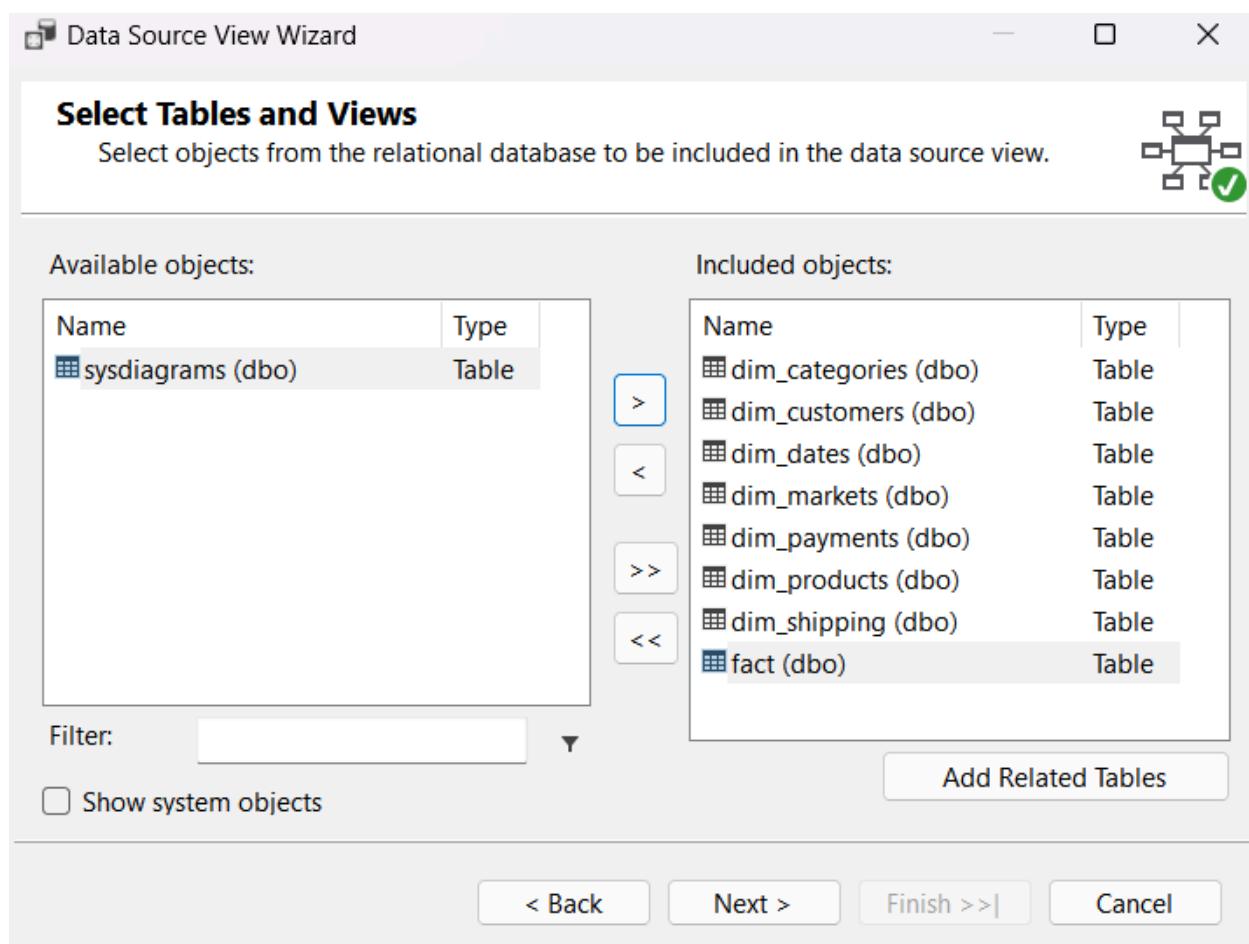
- Cửa sổ Data Source View Wizard xuất hiện -> Next



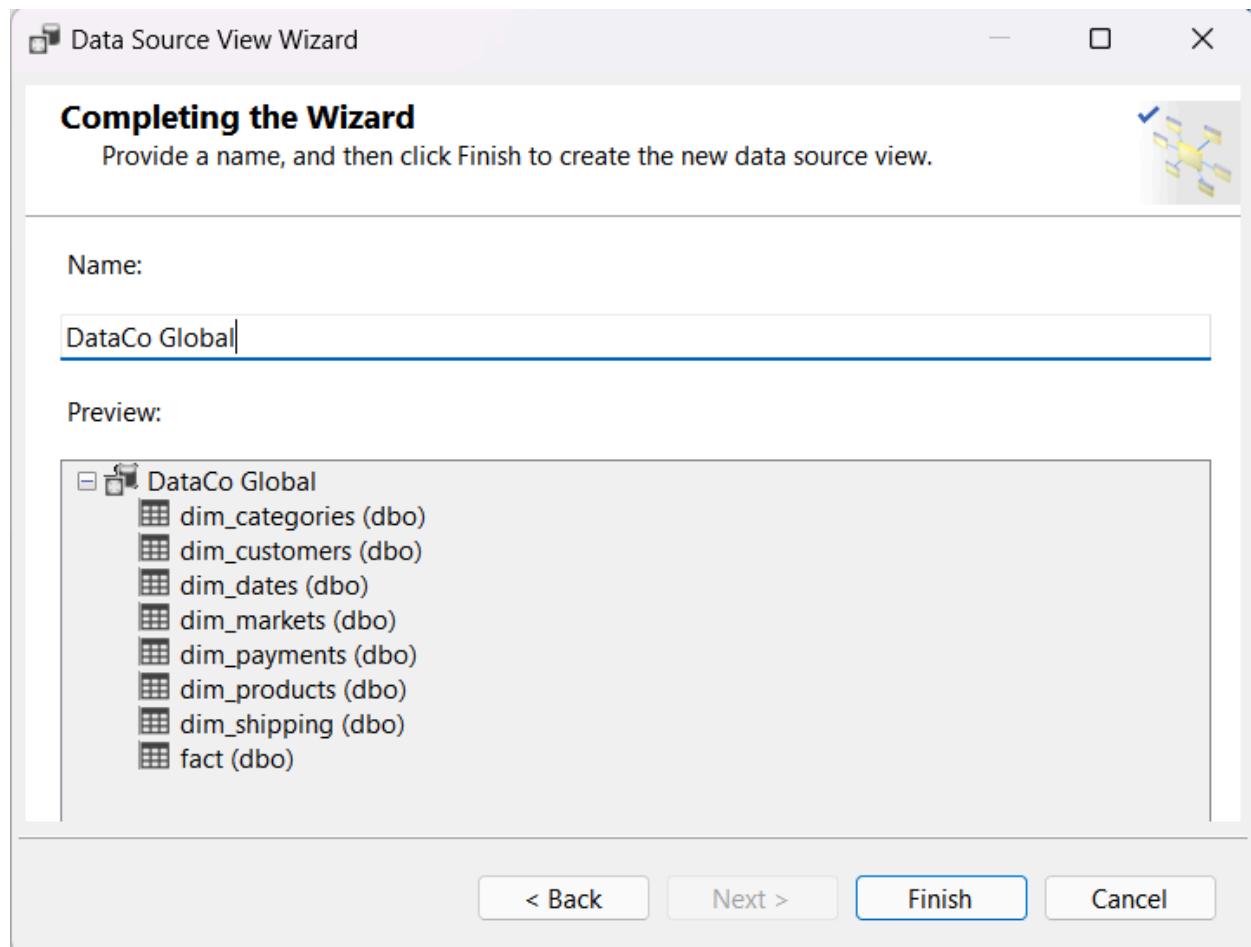
- Chọn Data Sources đã tạo ở bước trên -> Next



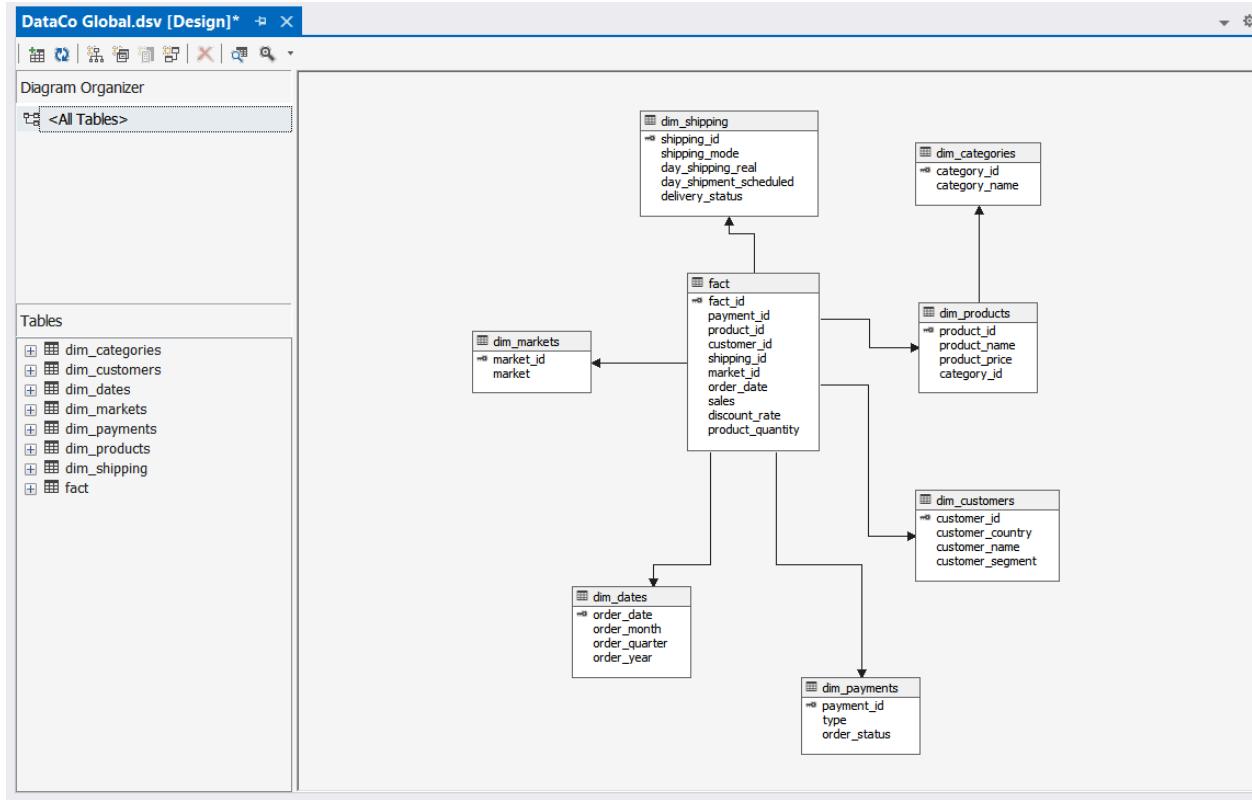
- Chọn bảng Fact và các bảng Dimensions -> Next



- Án Finish để hoàn tất quá trình tạo Data Sources View

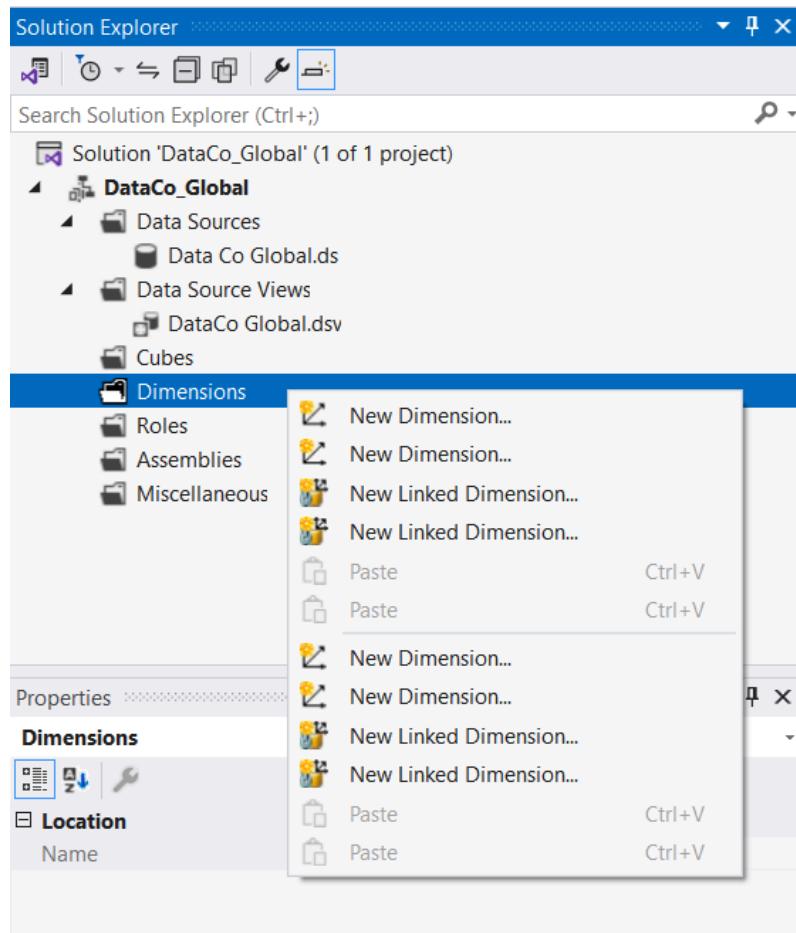


- Data Sources View sau khi hoàn thành.

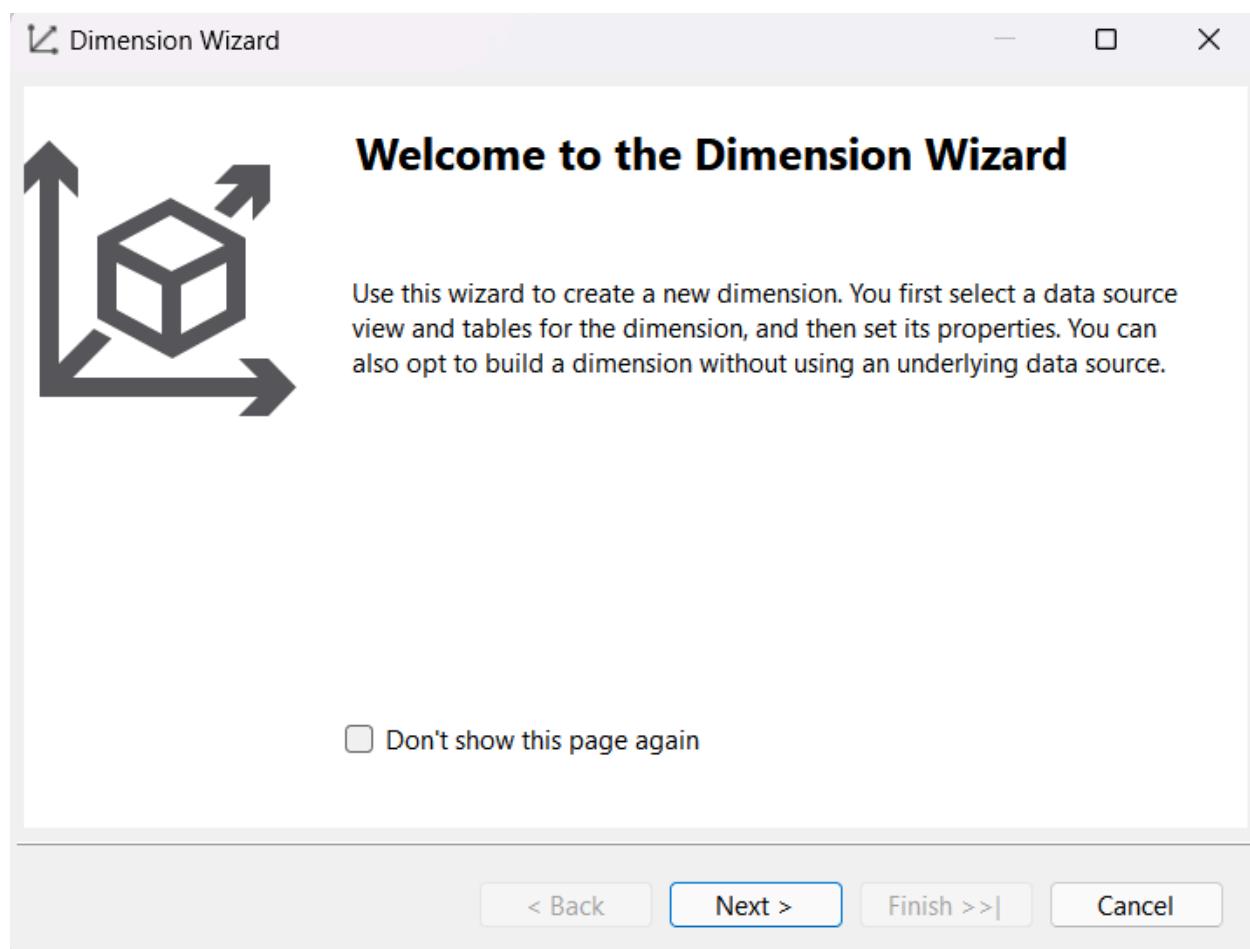


3.4. XÁC ĐỊNH CÁC CHIỀU

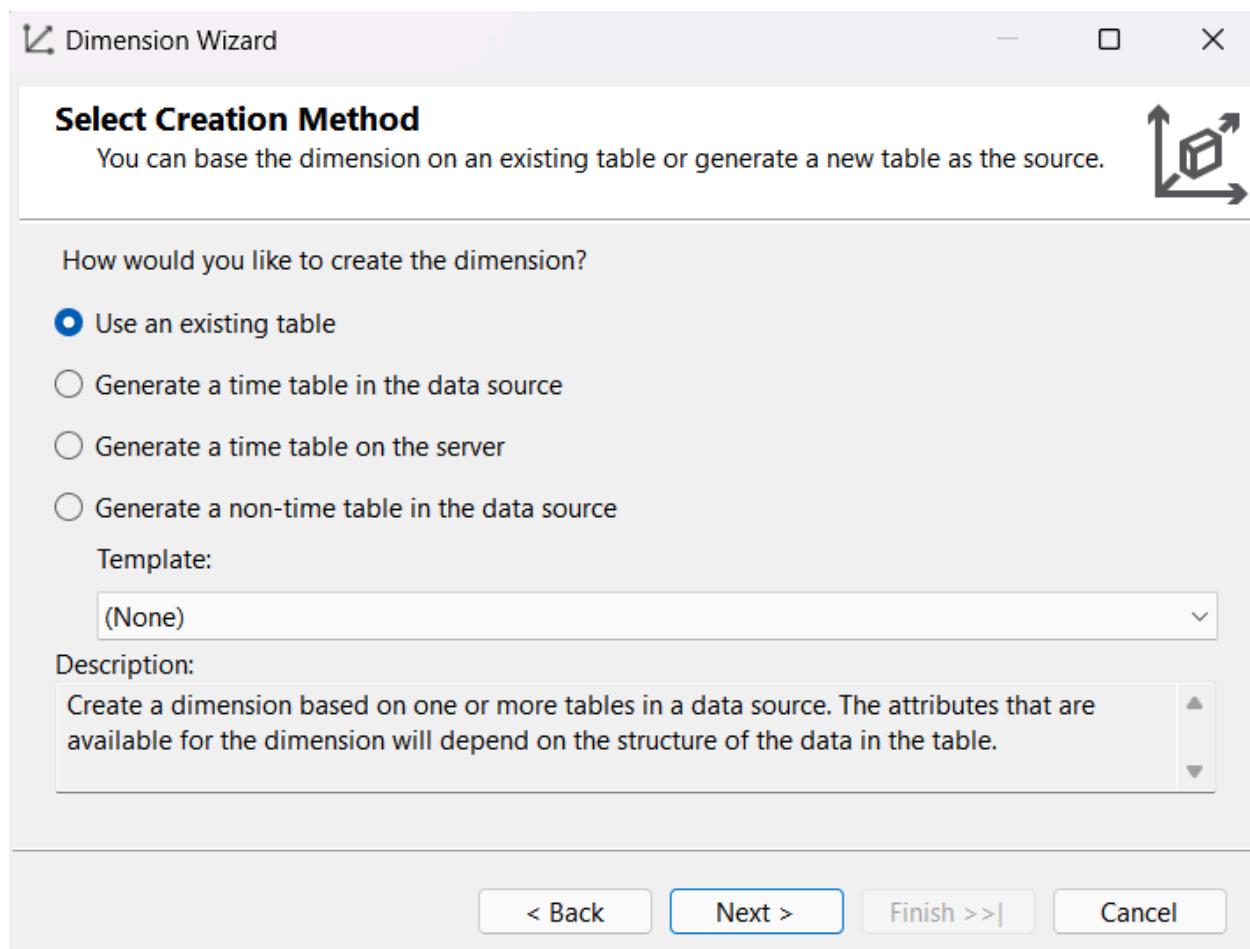
- Ở tab Solution Explorer bên phải màn hình -> **Dimensions** -> **New Dimensions**.



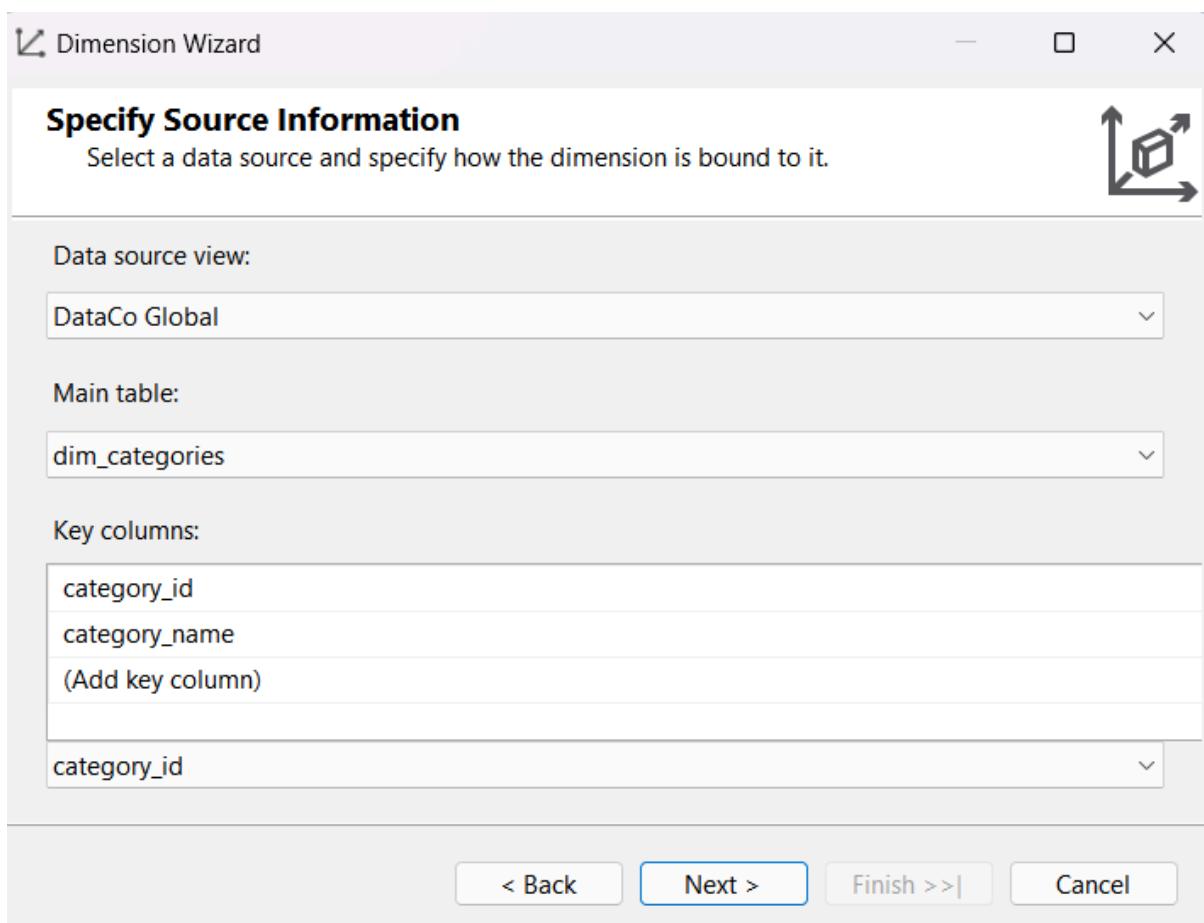
- Cửa sổ Dimension Wizard xuất hiện -> Next.



- Chọn Use an existing table -> Next



- Tiến hành tạo các bảng chiều DIM_CATEGORIES, DIM_PRODUCTS, DIM_SHIPPINGS, DIM_DATES, DIM_MARKETS, DIM_CUSTOMERS, DIM_PAYMENST
 - DIM_CATEGORIES



Dimension Wizard

Select Dimension Attributes

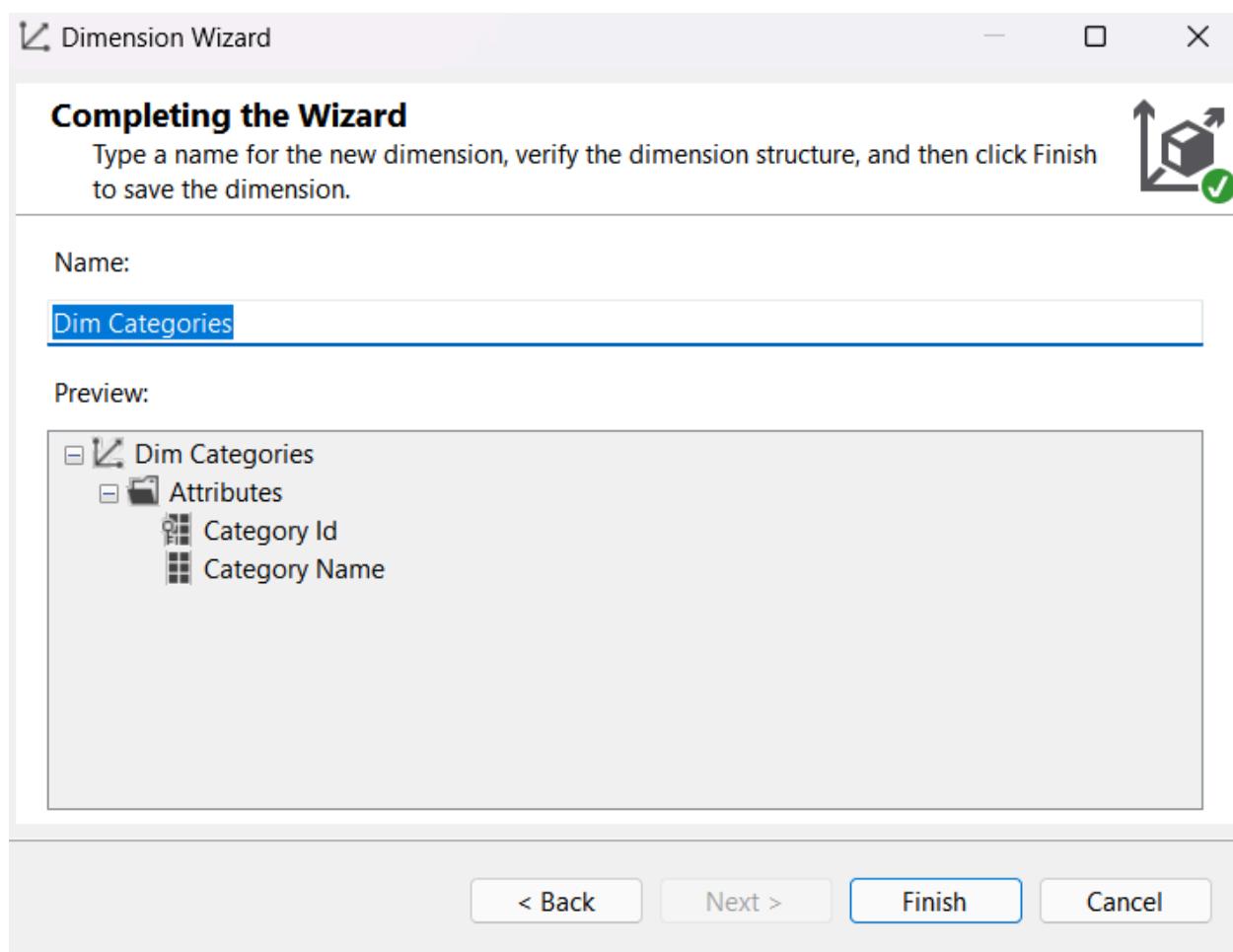
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.



Available attributes:

Attribute Name	Enable Browsing	Attribute Type
Category Id	<input checked="" type="checkbox"/>	Regular
Category Name	<input checked="" type="checkbox"/>	Regular

< Back Next > Finish >> Cancel



Hoàn tất tạo **Dim_Categories**

Attributes	Hierarchies	Data Source View
<input checked="" type="checkbox"/> Dim_Categories Category Id Category Name	To create a new hierarchy, drag an attribute here.	dim_categories category_id category_name

- **DIM_PRODUCTS**

Dimension Wizard

Specify Source Information

Select a data source and specify how the dimension is bound to it.



Data source view:

DataCo Global

Main table:

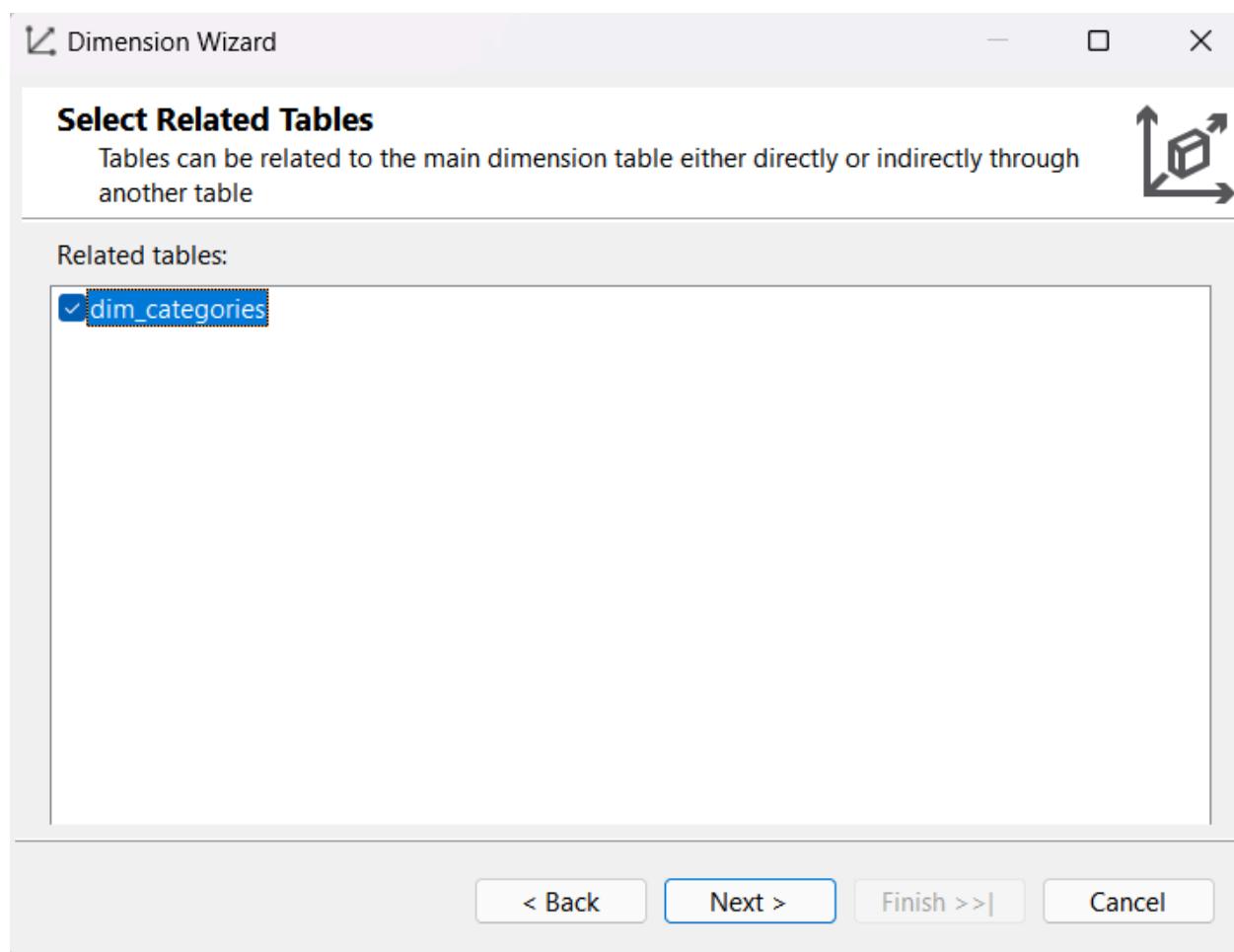
dim_products

Key columns:

product_id
(Add key column)

product_id

< Back Next > Finish >> Cancel



Dimension Wizard

Select Dimension Attributes

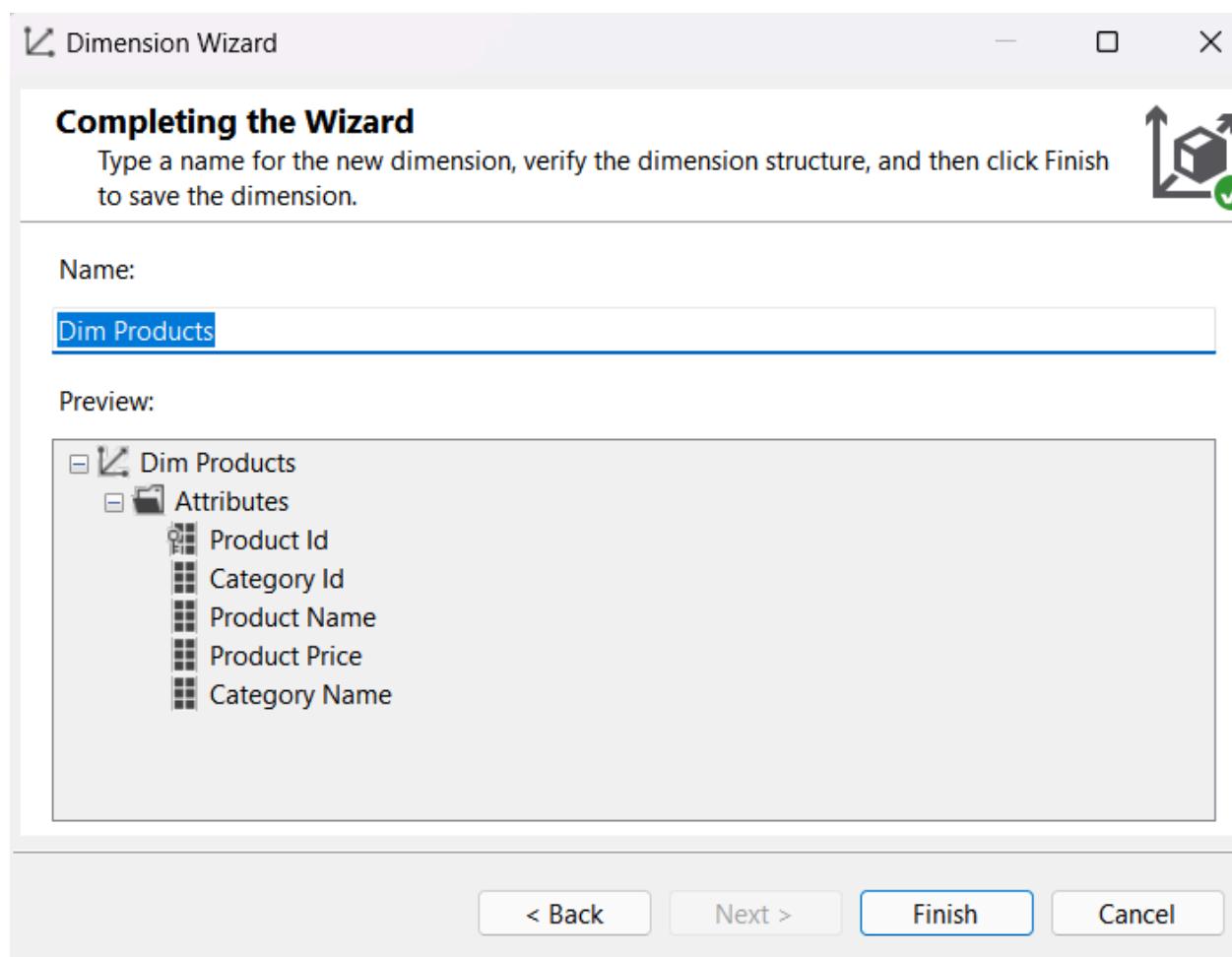
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.

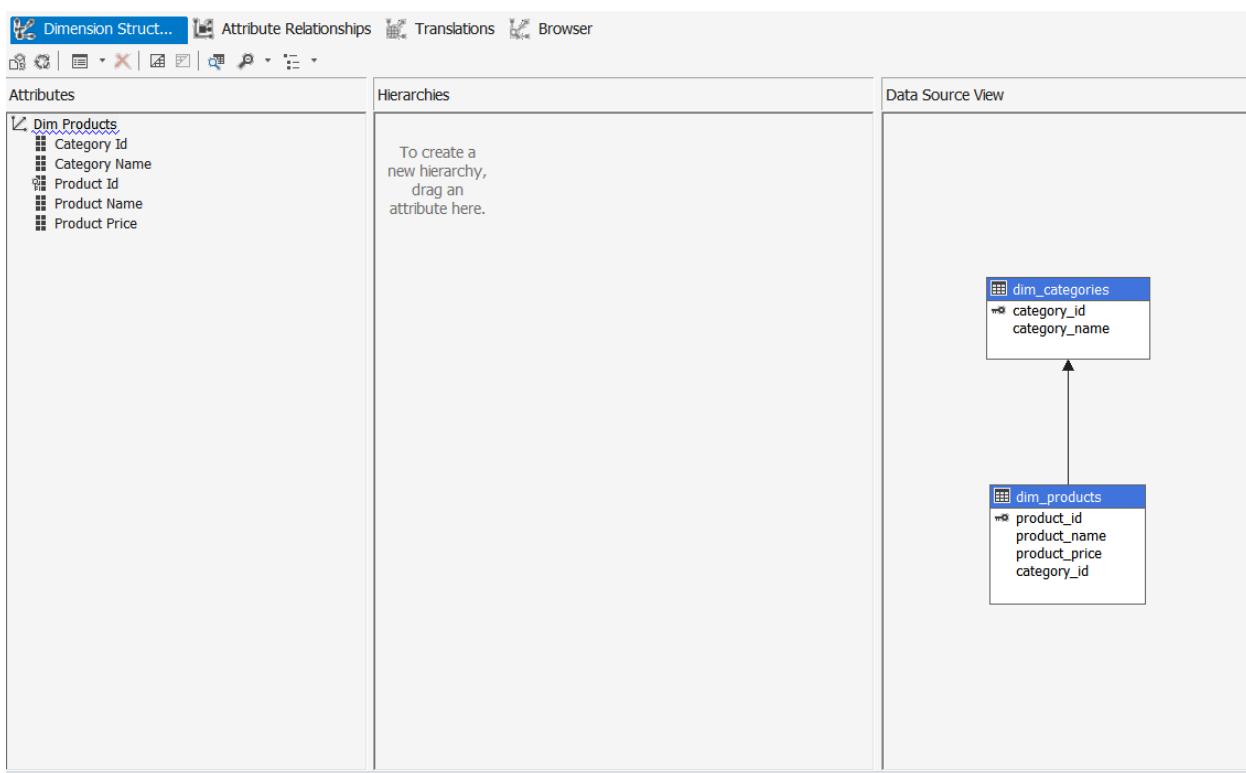


Available attributes:

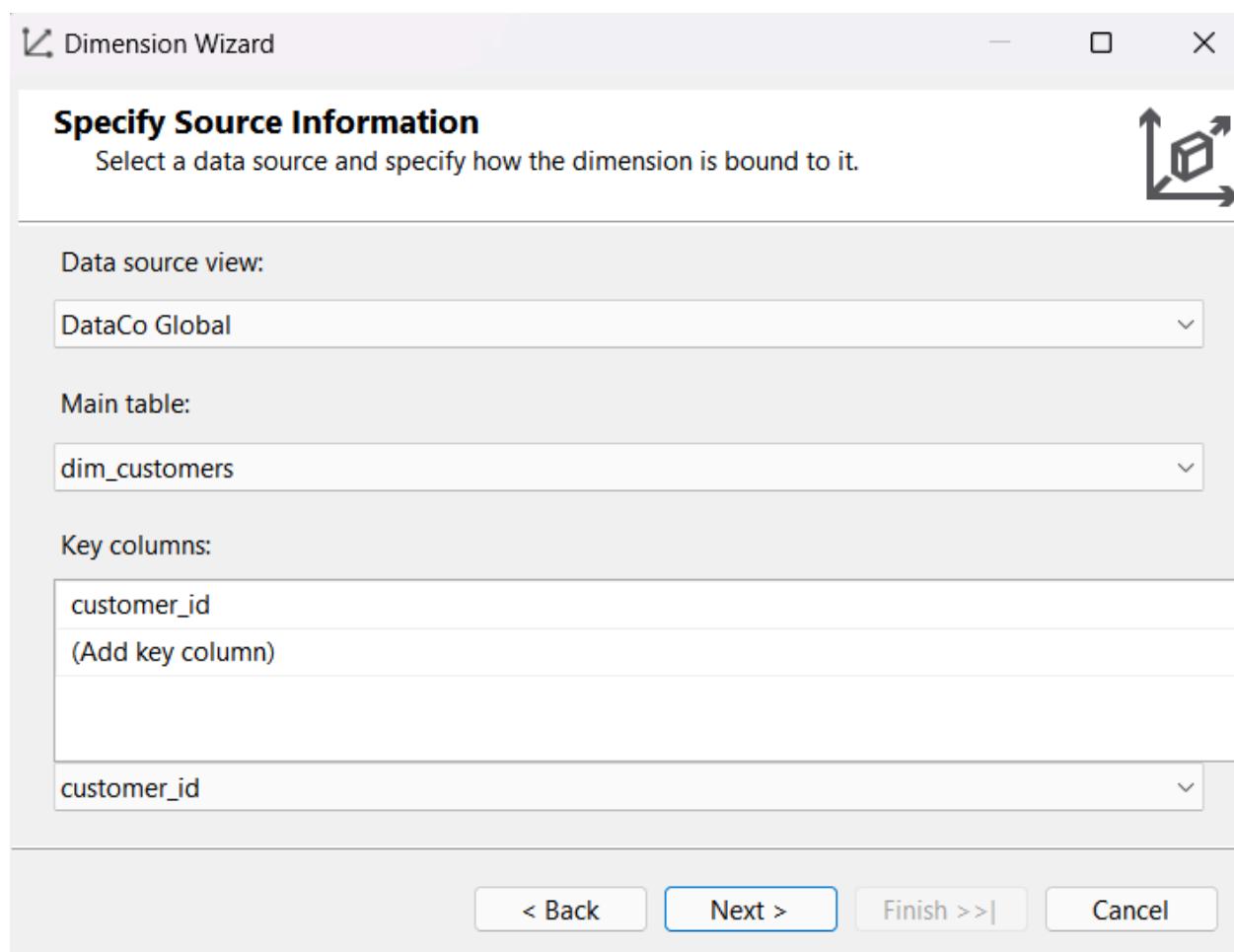
Attribute Name	Enable Browsing	Attribute Type
Product Id	<input checked="" type="checkbox"/>	Regular
Product Name	<input checked="" type="checkbox"/>	Regular
Product Price	<input checked="" type="checkbox"/>	Regular
Category Id	<input checked="" type="checkbox"/>	Regular
Category Name	<input checked="" type="checkbox"/>	Regular

< Back Next > Finish >> Cancel





- **DIM_CUSTOMERS**



Dimension Wizard

Select Dimension Attributes

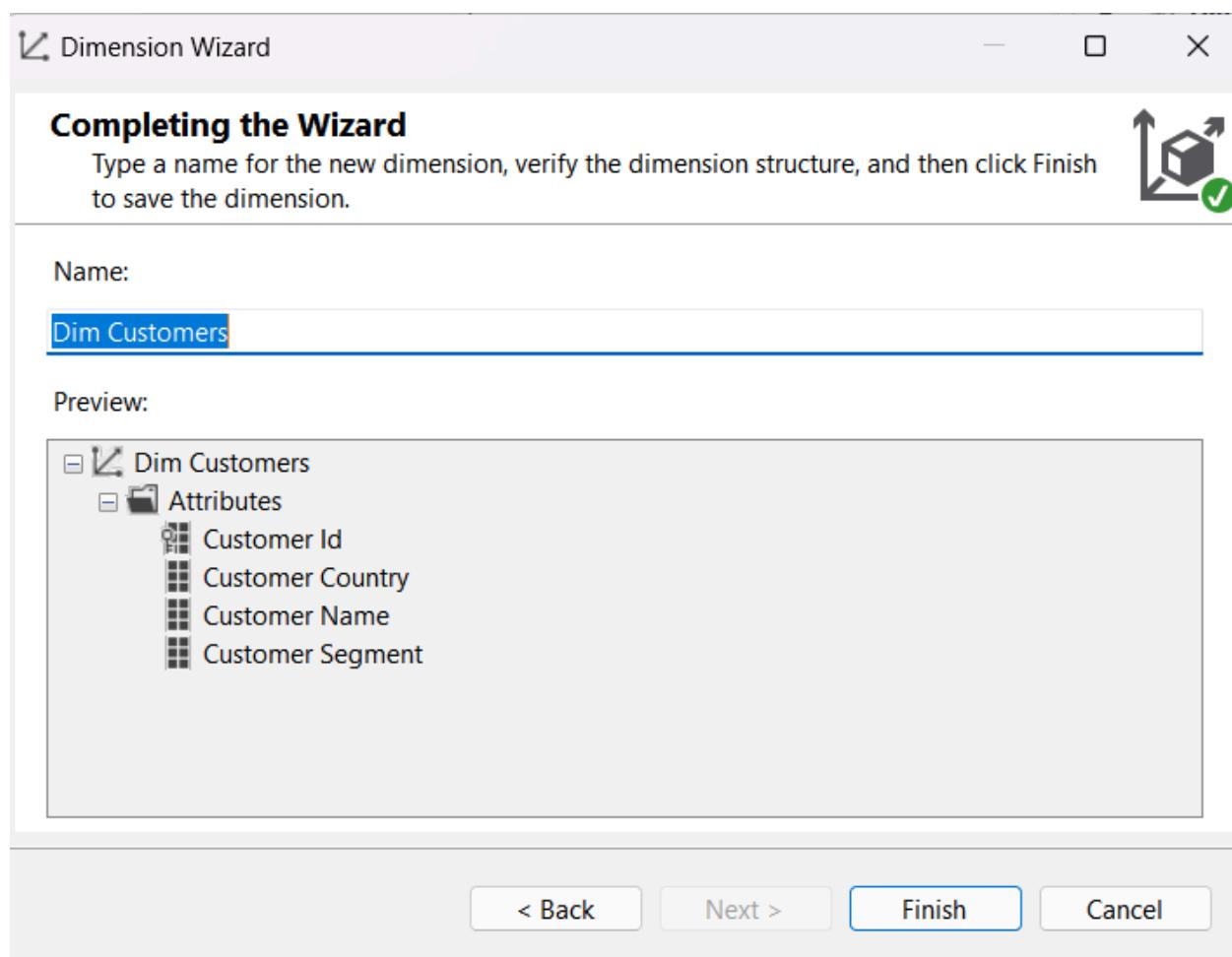
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.



Available attributes:

Attribute Name	Enable Browsing	Attribute Type
Attribute Name	<input checked="" type="checkbox"/>	
Customer Id	<input checked="" type="checkbox"/>	Regular
Customer Country	<input checked="" type="checkbox"/>	Regular
Customer Name	<input checked="" type="checkbox"/>	Regular
Customer Segment	<input checked="" type="checkbox"/>	Regular

< Back Next > Finish >> Cancel



The screenshot shows the Microsoft Analysis Services Dimension Structure tool interface. The top navigation bar includes tabs for Dimension Structure, Attribute Relationships, Translations, and Browser. Below the tabs are standard window controls (Minimize, Maximize, Close). The main area is divided into three panes:

- Attributes** pane: Displays a tree structure for the 'Dim Customers' dimension. It includes attributes: Customer Country, Customer Id, Customer Name, and Customer Segment.
- Hierarchies** pane: A placeholder area with the instruction: "To create a new hierarchy, drag an attribute here."
- Data Source View** pane: Shows a data source view for the 'dim_customers' table. The table has four columns: customer_id, customer_country, customer_name, and customer_segment.

- **DIM_MARKETS**

Dimension Wizard

Specify Source Information

Select a data source and specify how the dimension is bound to it.



Data source view:

DataCo Global

Main table:

dim_markets

Key columns:

market_id
(Add key column)
market_id

< Back Next > Finish >>| Cancel

Dimension Wizard

Select Dimension Attributes

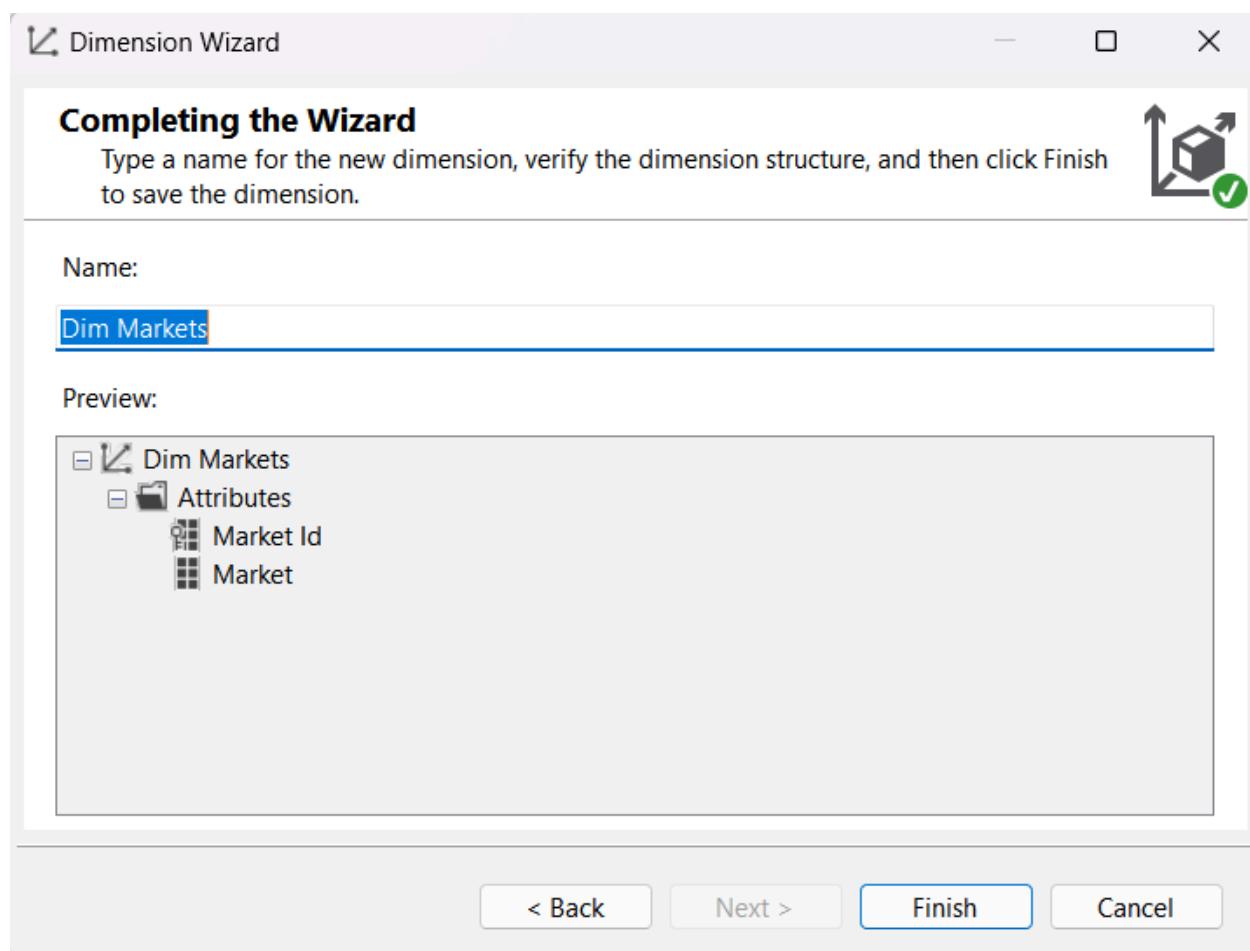
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.



Available attributes:

Attribute Name	Enable Browsing	Attribute Type
Attribute Name	<input checked="" type="checkbox"/>	
Market Id	<input checked="" type="checkbox"/>	Regular
Market	<input checked="" type="checkbox"/>	Regular

< Back Next > Finish >> Cancel



The screenshot shows the Microsoft Analysis Services Dimension Structure tool interface. The top navigation bar includes tabs for Dimension Structure, Attribute Relationships, Translations, and Browser. Below the tabs is a toolbar with various icons for filtering, sorting, and search.

The main area is divided into three columns:

- Attributes:** A tree view showing a single node "Dim Markets" expanded to show "Market" and "Market Id".
- Hierarchies:** A large empty area with the placeholder text: "To create a new hierarchy, drag an attribute here."
- Data Source View:** A list box containing the table "dim_markets" with columns "market_id" and "market".

- **DIM_SHIPPINGS**

Dimension Wizard

Specify Source Information

Select a data source and specify how the dimension is bound to it.



Data source view:

DataCo Global

Main table:

dim_shipping

Key columns:

shipping_id
(Add key column)

shipping_id

< Back Next > Finish >>| Cancel

Dimension Wizard

Select Dimension Attributes

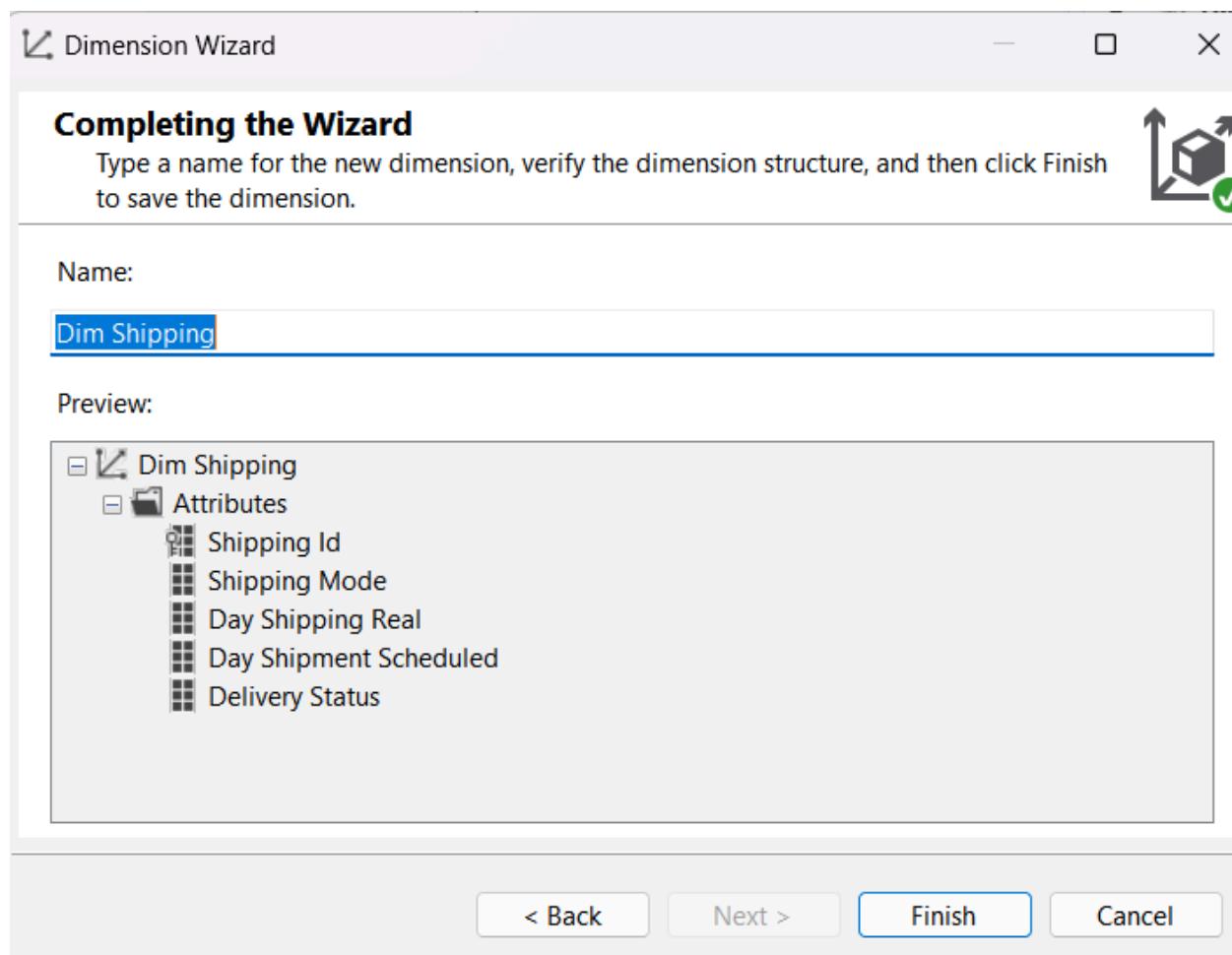
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.



Available attributes:

Attribute Name	Enable Browsing	Attribute Type
Shipping Id	<input checked="" type="checkbox"/>	Regular
Shipping Mode	<input checked="" type="checkbox"/>	Regular
Day Shipping Real	<input checked="" type="checkbox"/>	Regular
Day Shipment Scheduled	<input checked="" type="checkbox"/>	Regular
Delivery Status	<input checked="" type="checkbox"/>	Regular

< Back Next > Finish >> Cancel



The screenshot shows the Microsoft SQL Server Analysis Services (SSAS) Dimension Structure tool interface. The top navigation bar includes tabs for Dimension Structure, Attribute Relationships, Translations, and Browser. Below the tabs, there are three main sections: Attributes, Hierarchies, and Data Source View.

- Attributes:** This section lists the attributes of the Dim Shipping dimension. It includes:
 - Dim Shipping
 - Day Shipment Scheduled
 - Day Shipping Real
 - Delivery Status
 - Shipping Id
 - Shipping Mode
- Hierarchies:** This section contains a placeholder message: "To create a new hierarchy, drag an attribute here."
- Data Source View:** This section displays the Data Source View for the dim_shipping dimension, which contains the following columns:

dim_shipping
shipping_id
shipping_mode
day_shipping_real
day_shipment_scheduled
delivery_status

- **DIM_DATES**

Dimension Wizard

Specify Source Information

Select a data source and specify how the dimension is bound to it.



Data source view:

DataCo Global

Main table:

dim_dates

Key columns:

order_date
(Add key column)

order_date

< Back Next > Finish >>| Cancel

Dimension Wizard

Select Dimension Attributes

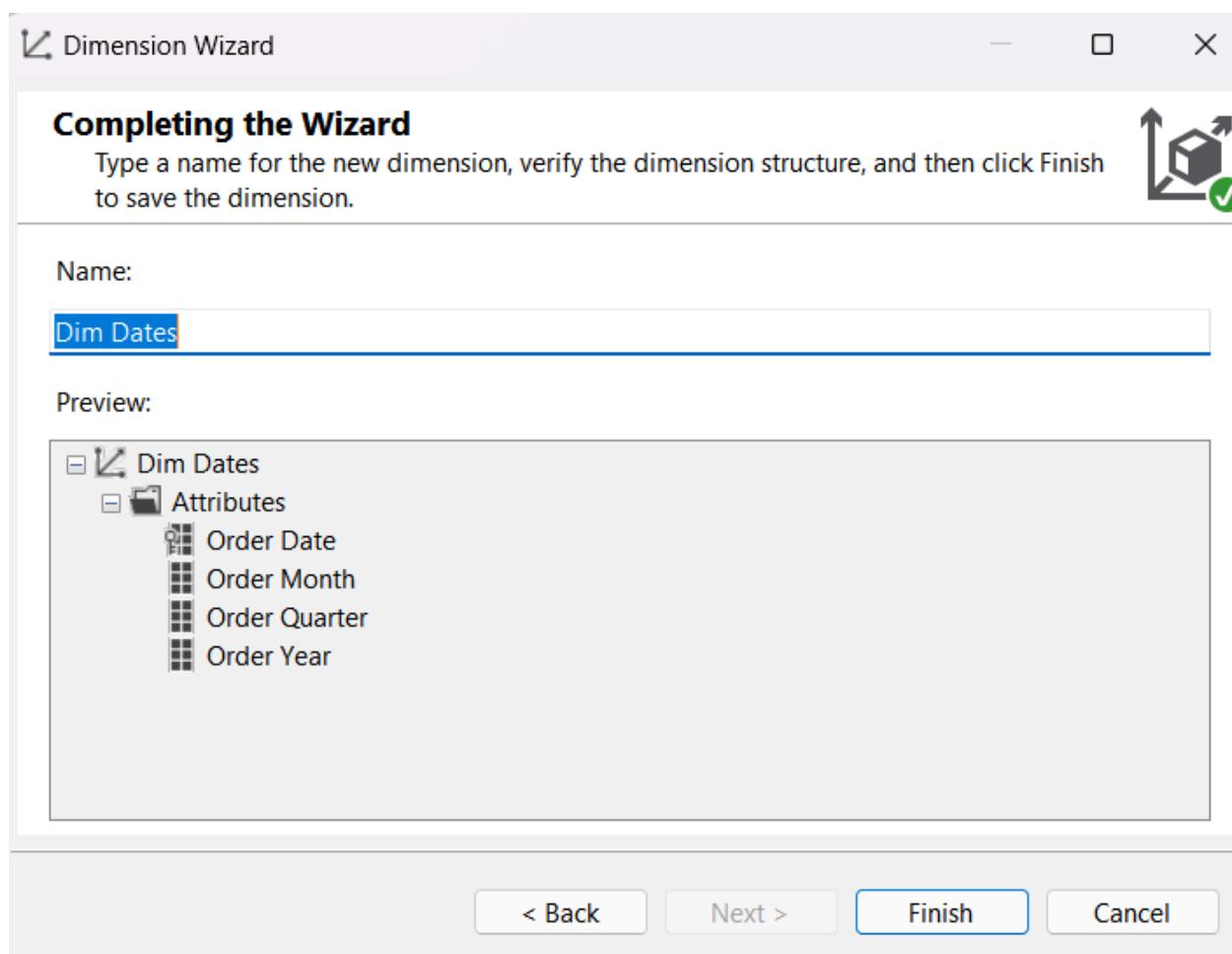
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.



Available attributes:

Attribute Name	Enable Browsing	Attribute Type
Order Date	<input checked="" type="checkbox"/>	Regular
Order Month	<input checked="" type="checkbox"/>	Regular
Order Quarter	<input checked="" type="checkbox"/>	Regular
Order Year	<input checked="" type="checkbox"/>	Regular

< Back Next > Finish >> Cancel



The screenshot shows the Microsoft Analysis Services Dimension Structure tool interface. The top navigation bar includes tabs for 'Dimension Struct...', 'Attribute Relationships', 'Translations', and 'Browser'. Below the navigation bar are standard toolbar icons.

Attributes pane:

- Dim Dates
 - Order Date
 - Order Month
 - Order Quarter
 - Order Year

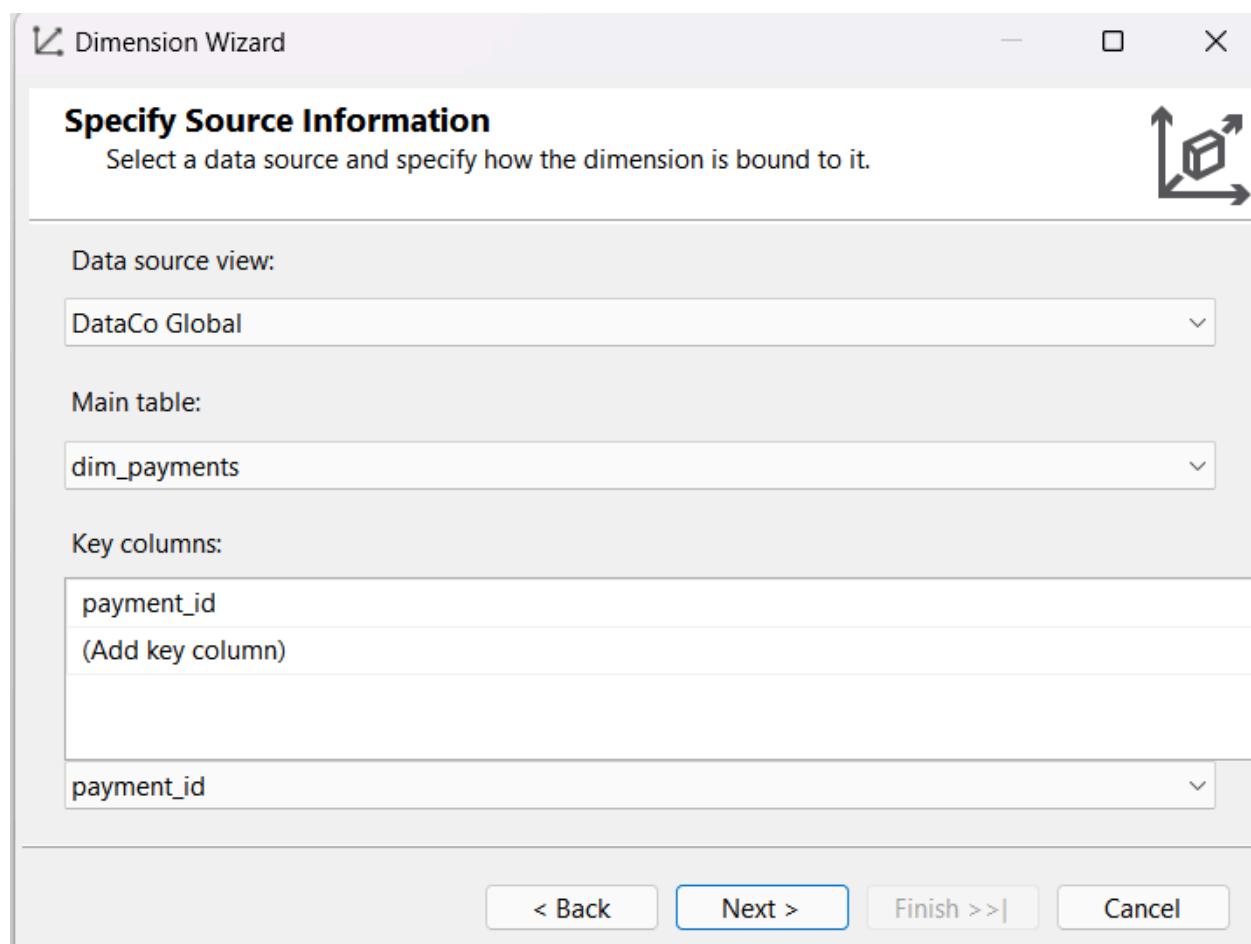
Hierarchies pane:

To create a new hierarchy, drag an attribute here.

Data Source View pane:

- dim_dates
 - order_date
 - order_month
 - order_quarter
 - order_year

- DIM_PAYMENTS



Dimension Wizard

Select Dimension Attributes

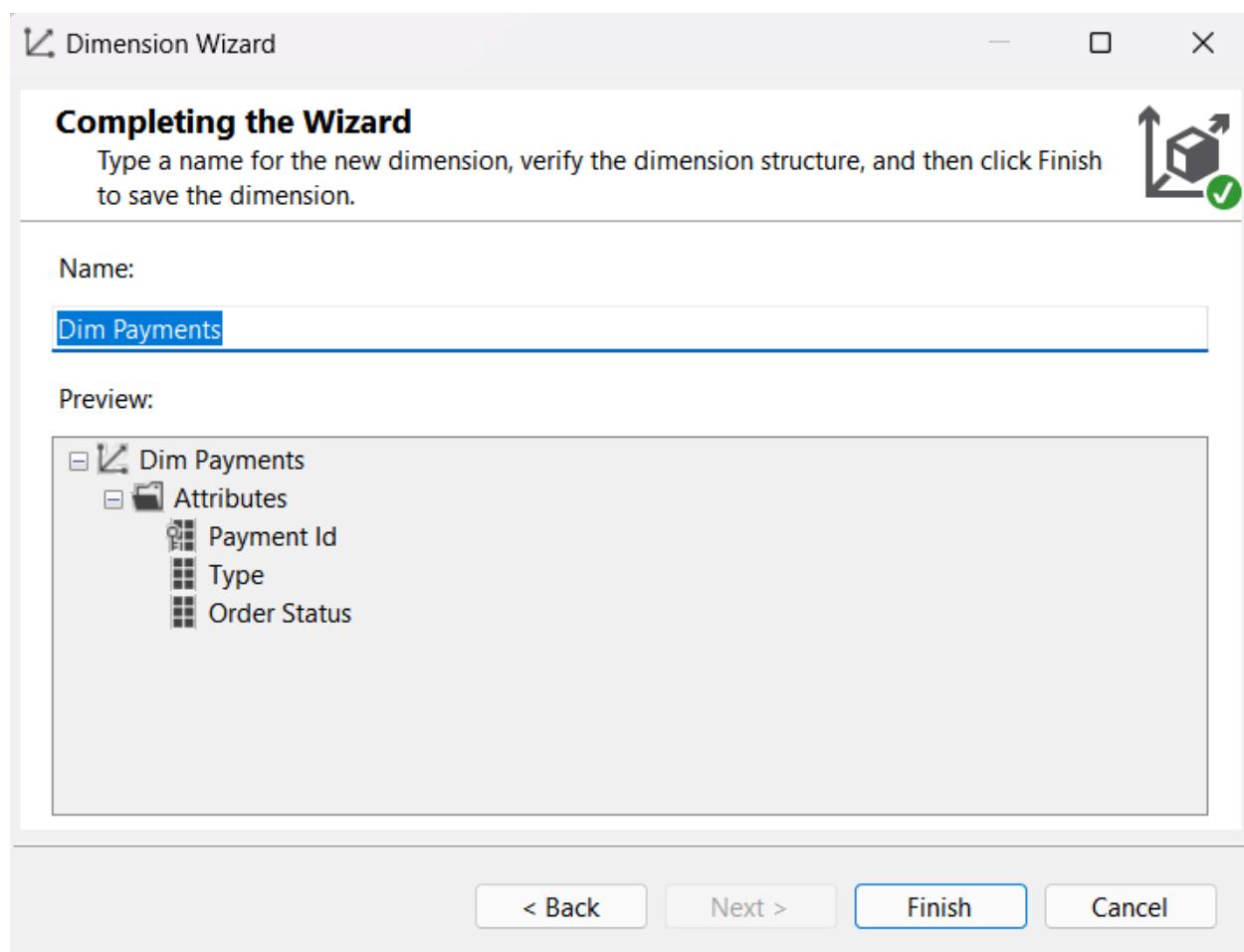
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.



Available attributes:

Attribute Name	Enable Browsing	Attribute Type
Payment Id	<input checked="" type="checkbox"/>	Regular
Type	<input checked="" type="checkbox"/>	Regular
Order Status	<input checked="" type="checkbox"/>	Regular

< Back Next > Finish >>| Cancel



Attributes

- Dim Payments
 - Order Status
 - Payment Id
 - Type

Hierarchies

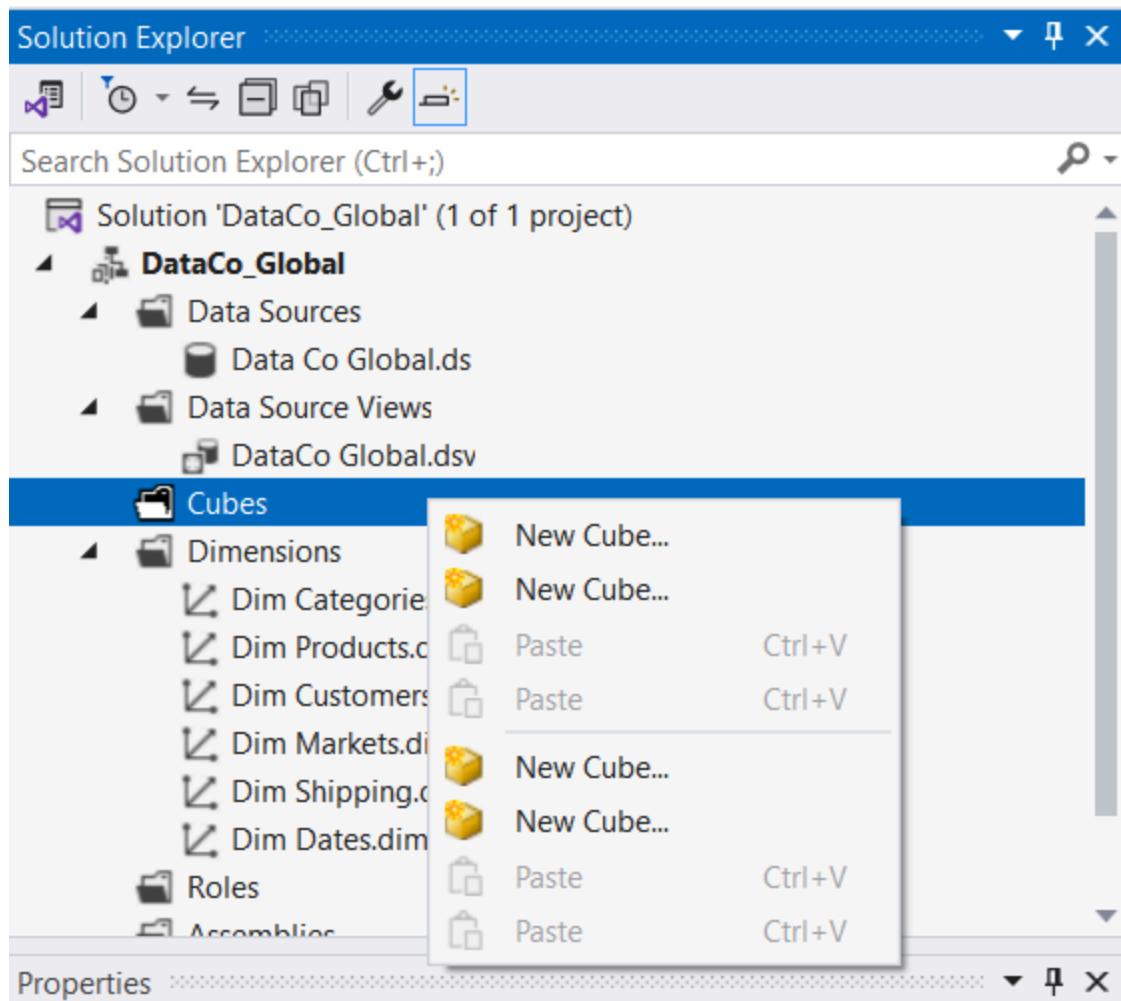
To create a new hierarchy, drag an attribute here.

Data Source View

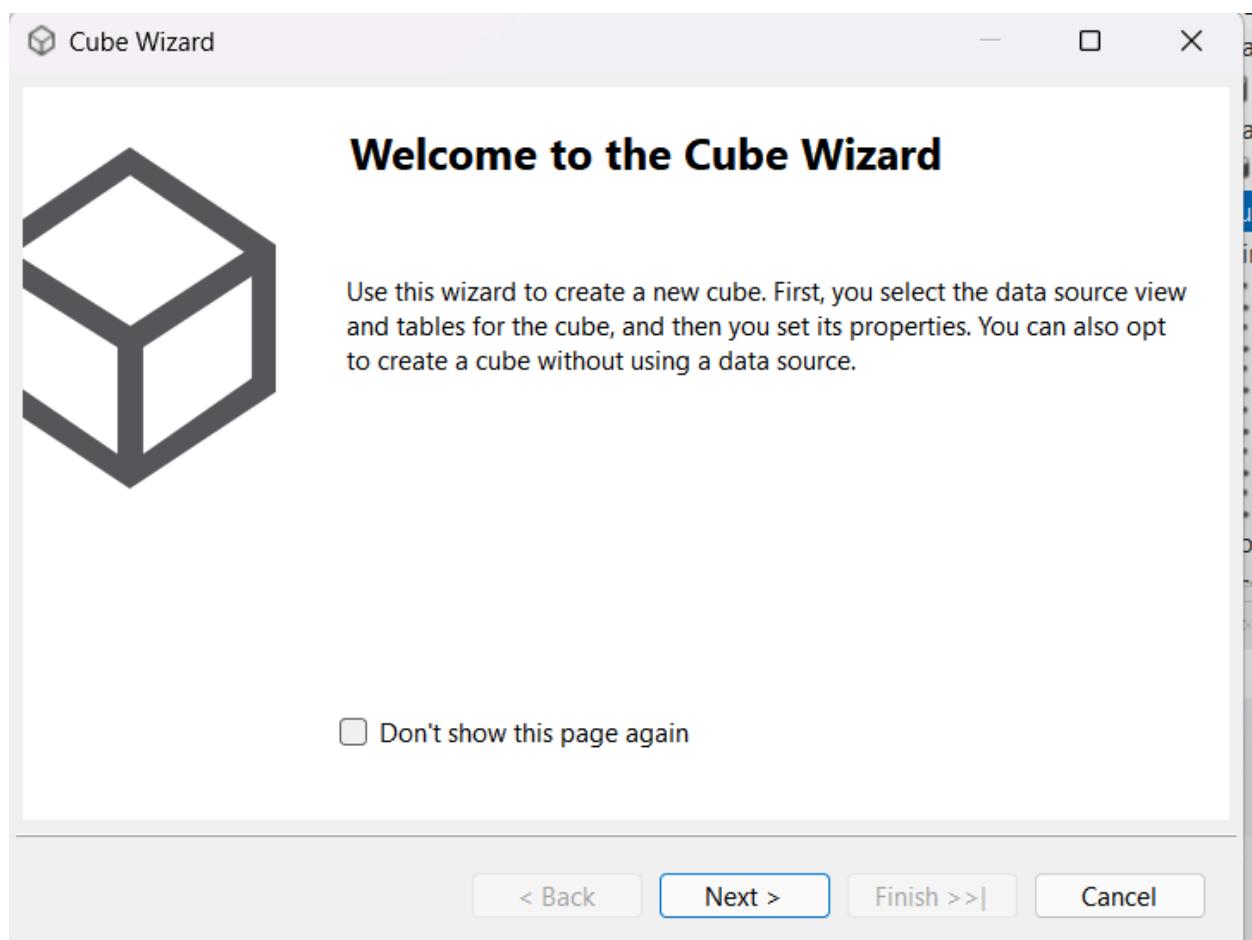
	dim_payments
payment_id	
type	
order_status	

3.5. XÂY DỰNG CÁC KHỐI VÀ MEASURES

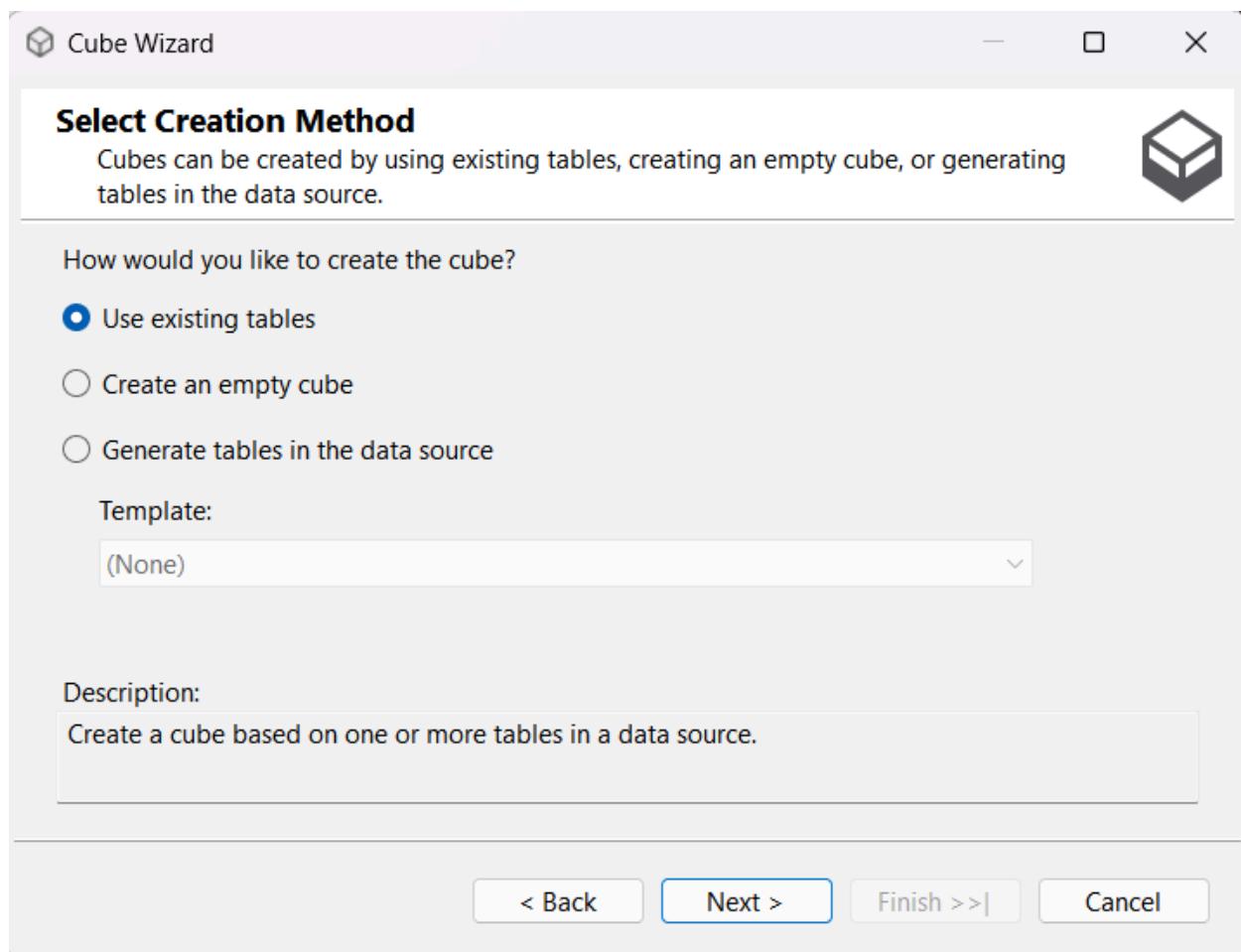
- Ở tab Solution Explorer bên phải màn hình -> **Cubes** -> **New Cubes** .



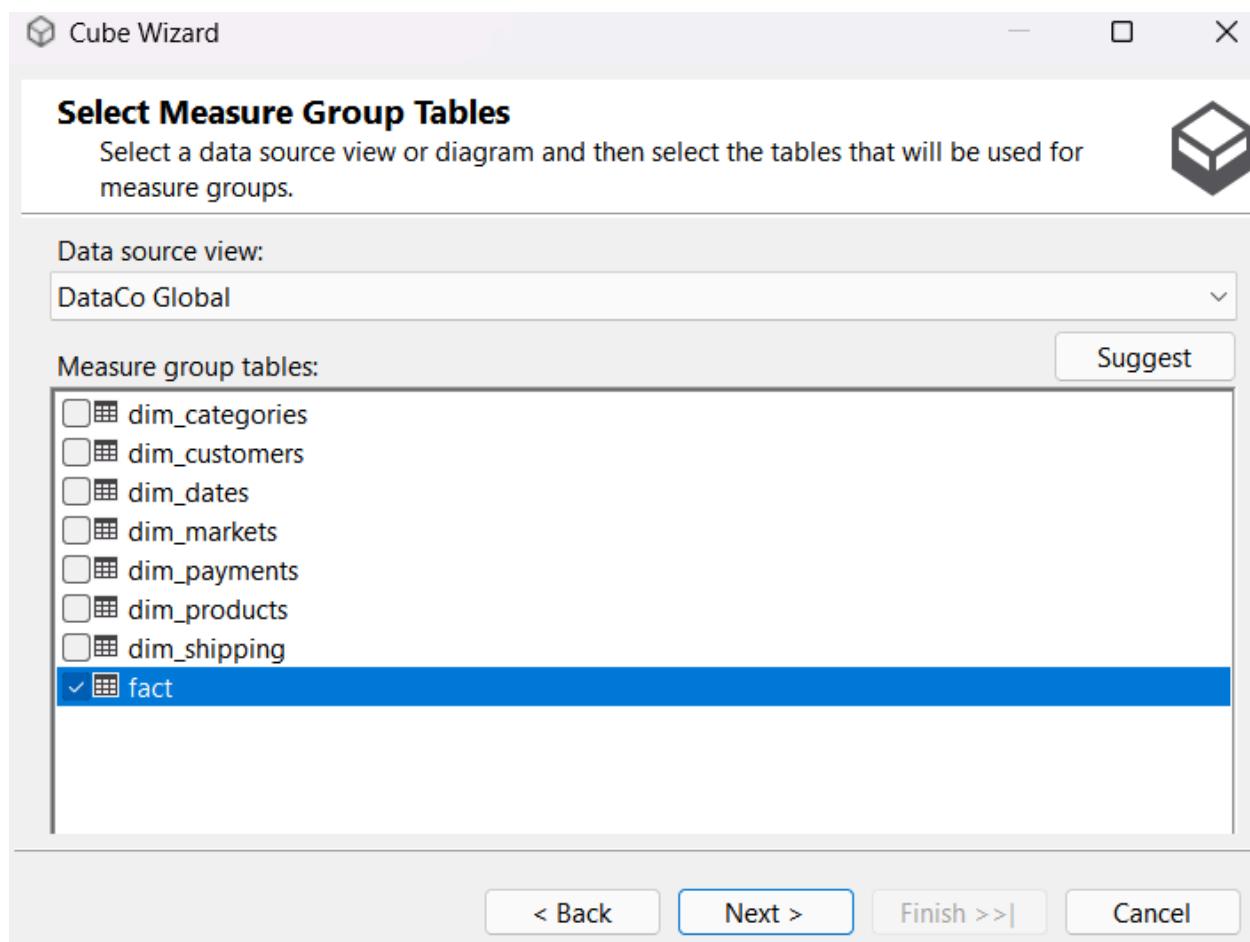
- Hộp thoại Cube Wizard xuất hiện -> Next.



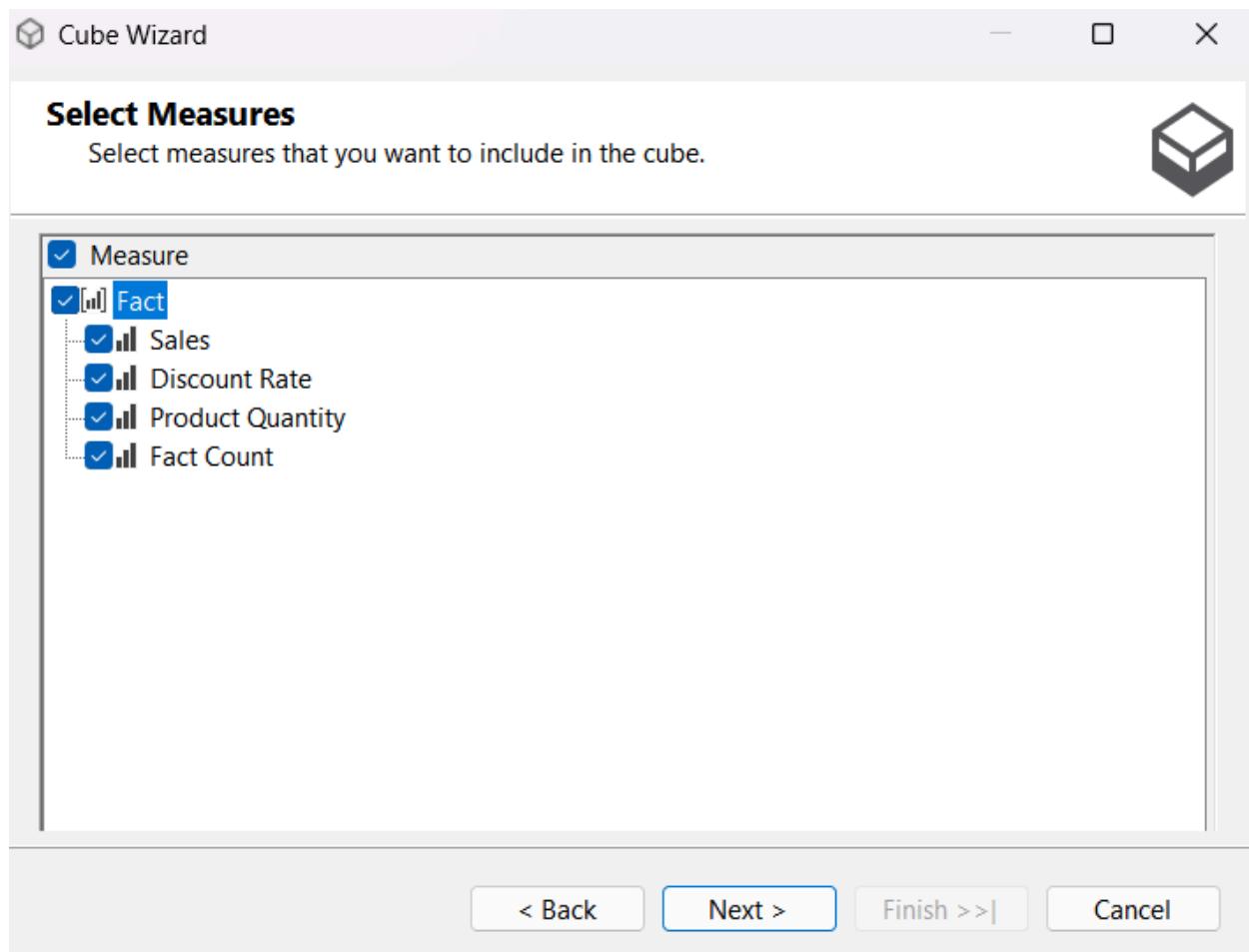
- Chọn Use existing tables -> Next.



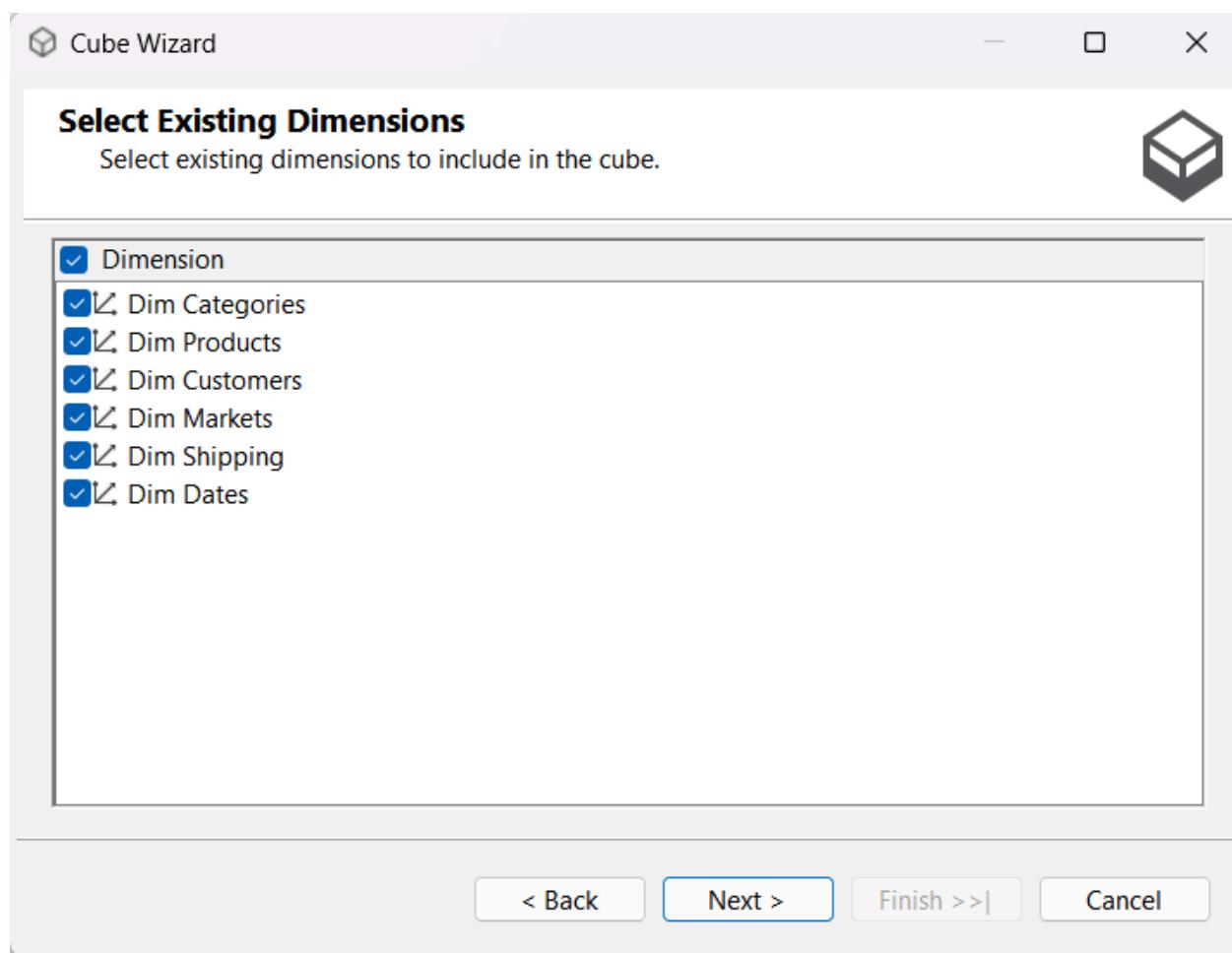
- Chọn Fact để phân chia các Measures Group



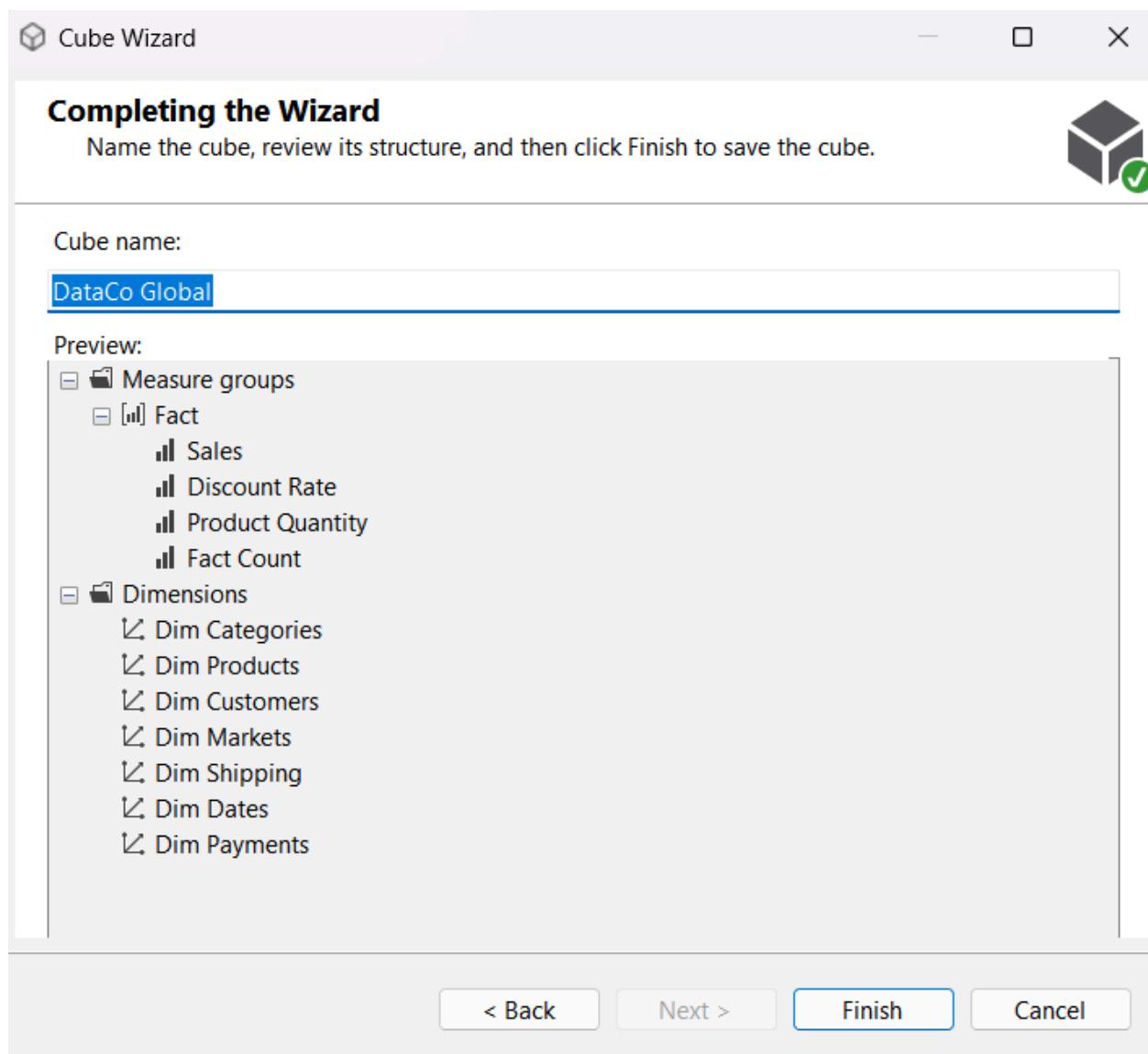
- Chọn các thuộc tính độ đo được đề xuất -> Next



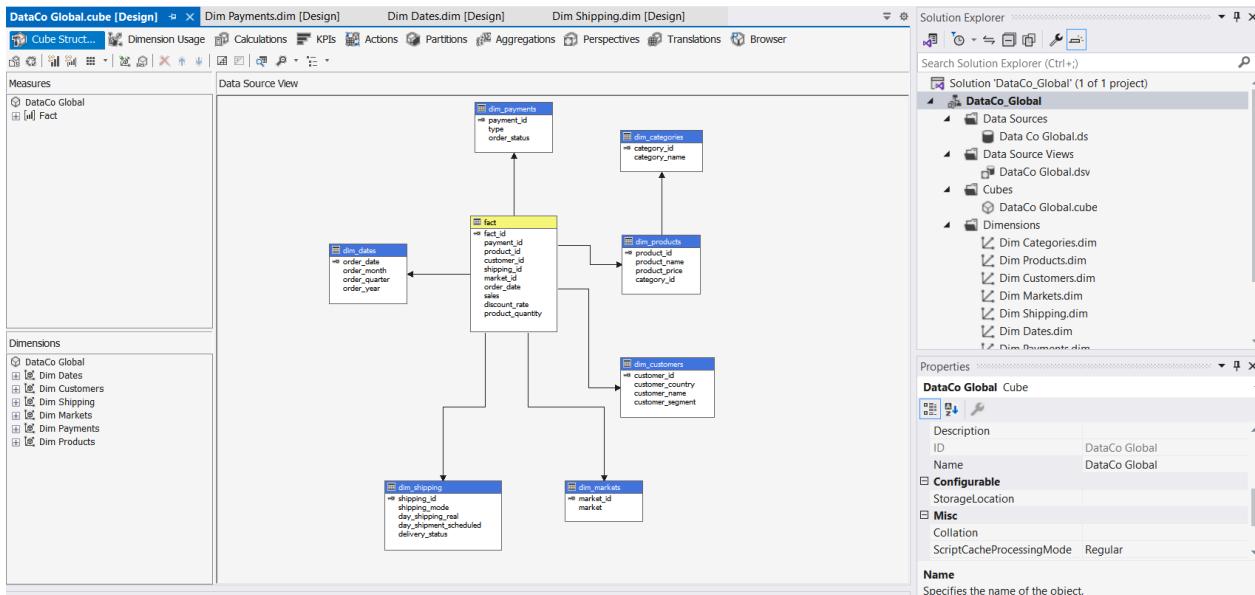
- Chọn các Dimensions -> Next



- Chọn Finish để hoàn thành quá trình tạo Cube



- Hoàn tất quá trình tạo Cube



3.6. ĐIỀU CHỈNH CÁC THUỘC TÍNH

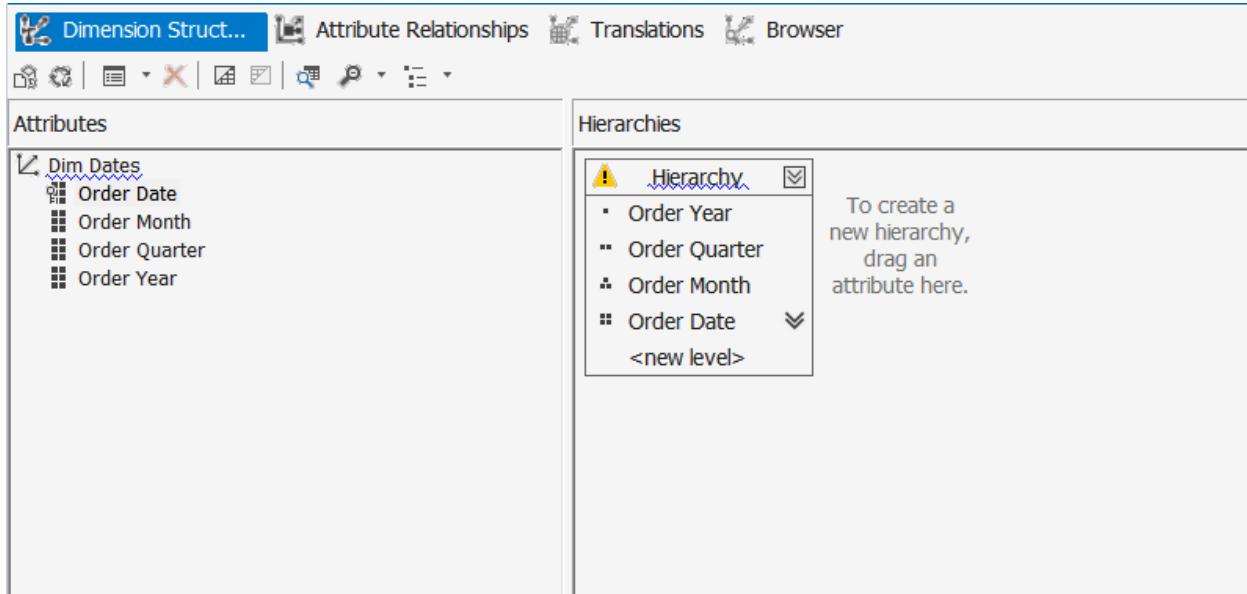
3.6.1. Đổi tên thuộc tính

Chọn Show Measures Group (dấu mũi tên hướng xuống) -> Chọn thuộc tính cần đổi tên

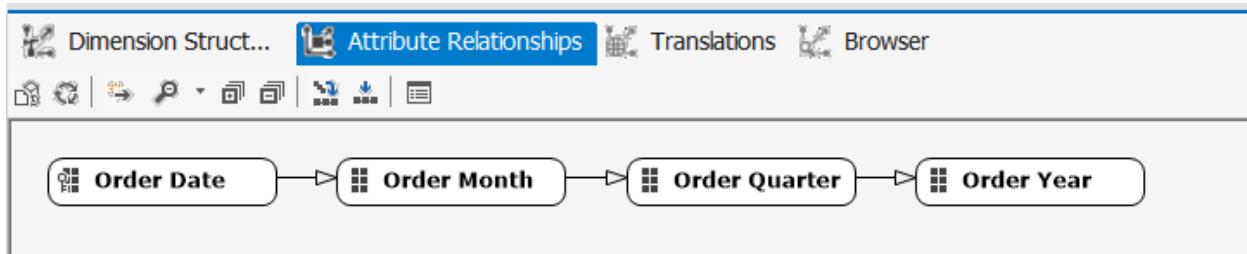
	Name	Measure Group	Data Type	Aggregation
Bar chart icon	Sales	Fact	Double	Sum
Bar chart icon	Discount Rate	Fact	Double	Sum
Bar chart icon	Product Quantity	Fact	UnsignedInt	Sum
Bar chart icon	Fact Count	Fact	Integer	Count

3.6.2. Tạo Hierarchies

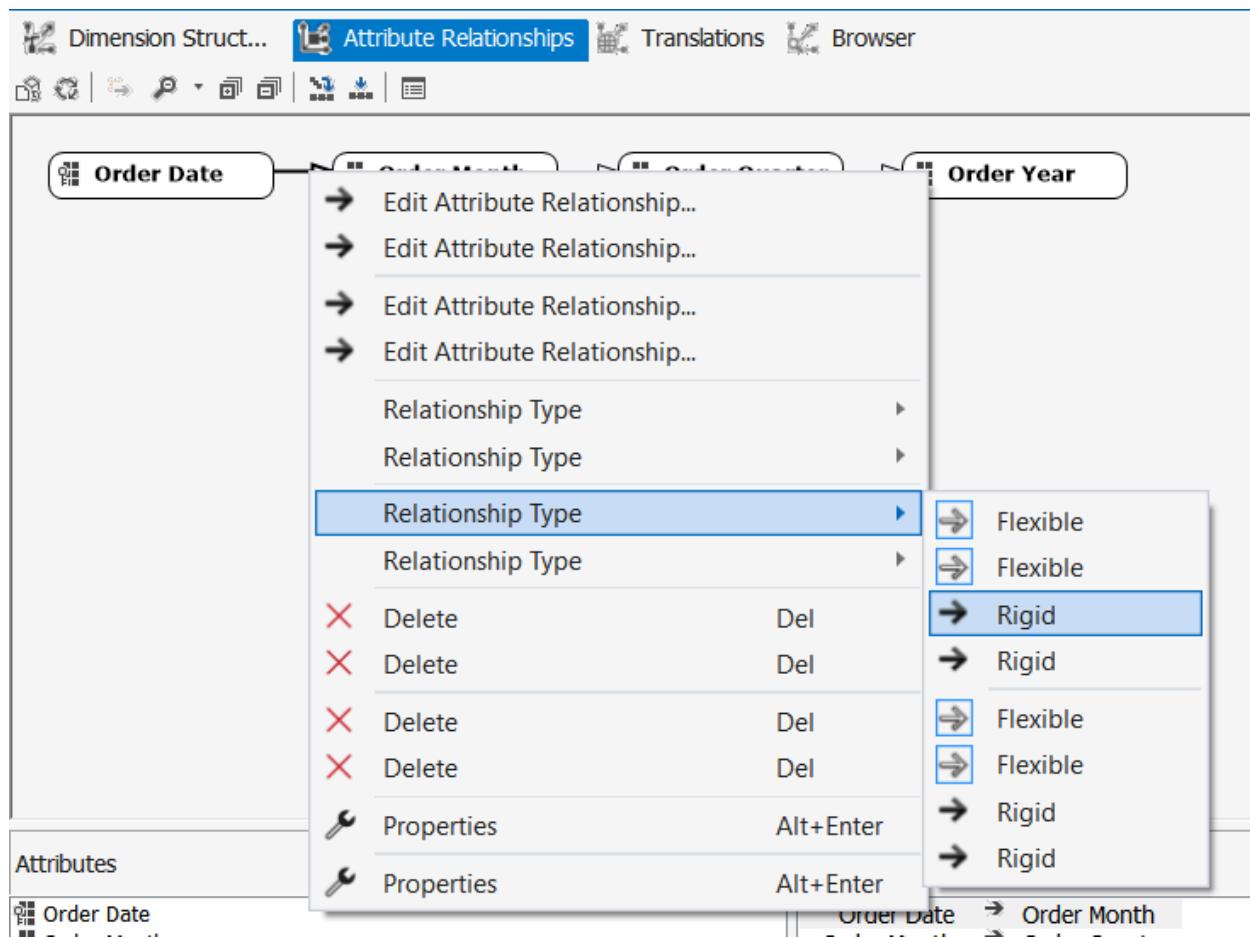
- Kéo các thuộc tính cần phân cấp qua tab **Hierarchies**



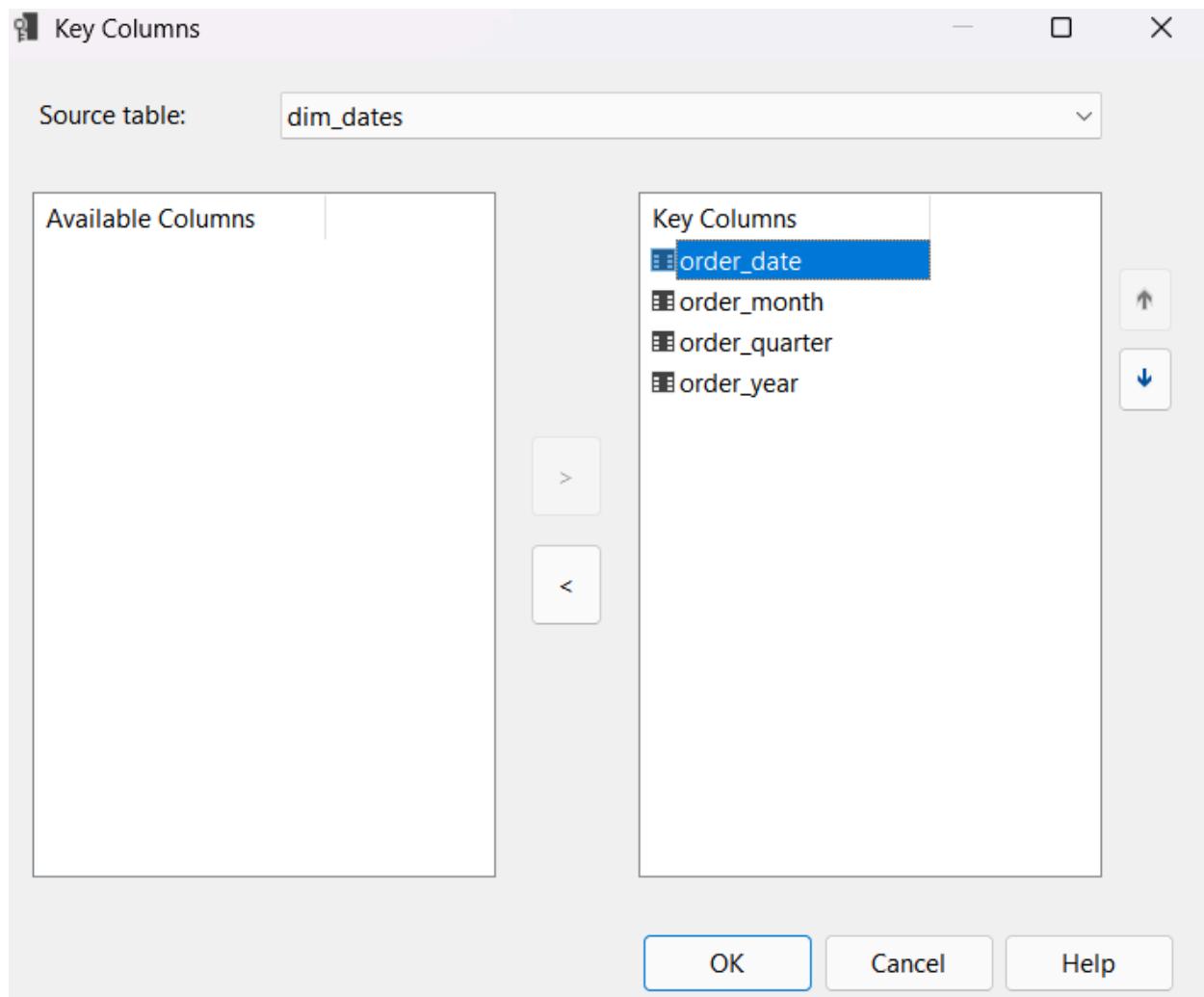
- Tại panel **Attribute Relationships**, điều chỉnh các relationship như sau:



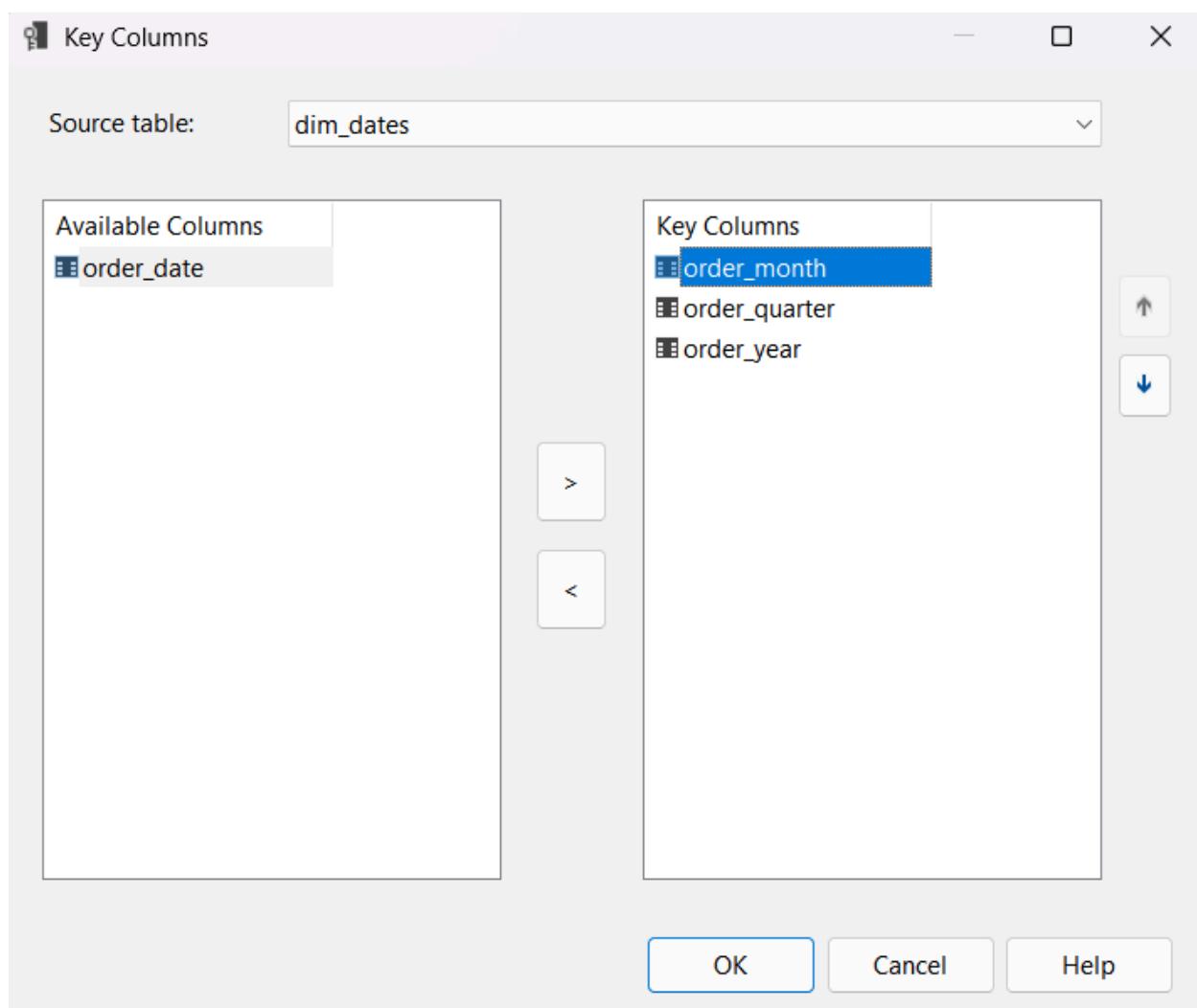
- Chọn các đường dẫn -> **Edit Attribute Relationships ->Rigid**



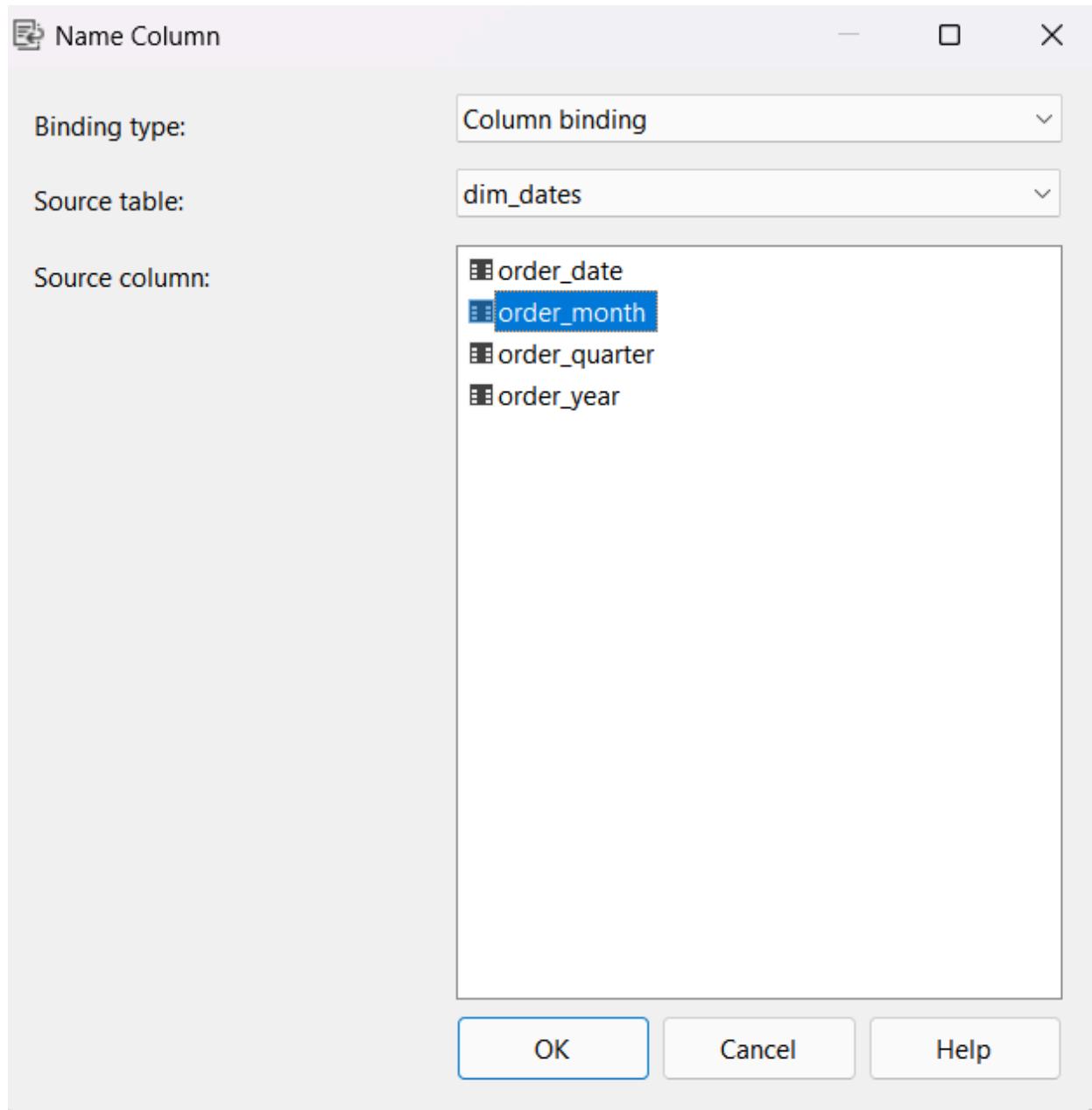
- Chính Key Columns cho Order_Date. Vì Order_Date là thuộc tính cấp nhỏ nhất nên lấy Key là chính nó và các thuộc tính cấp cao hơn.



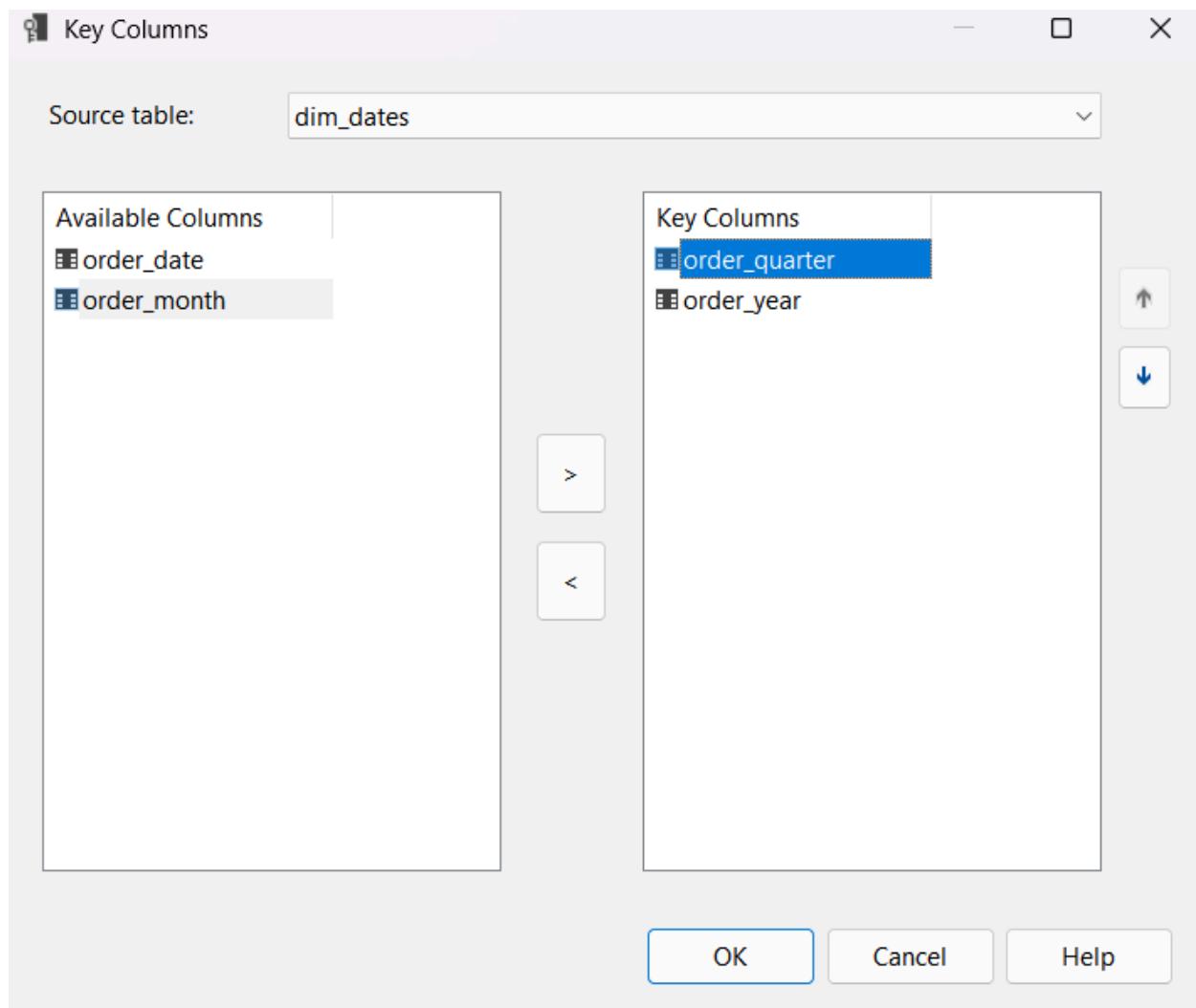
- Chính Key Columns cho Order_Month. Lấy Key là chính nó và các thuộc tính cấp cao hơn.



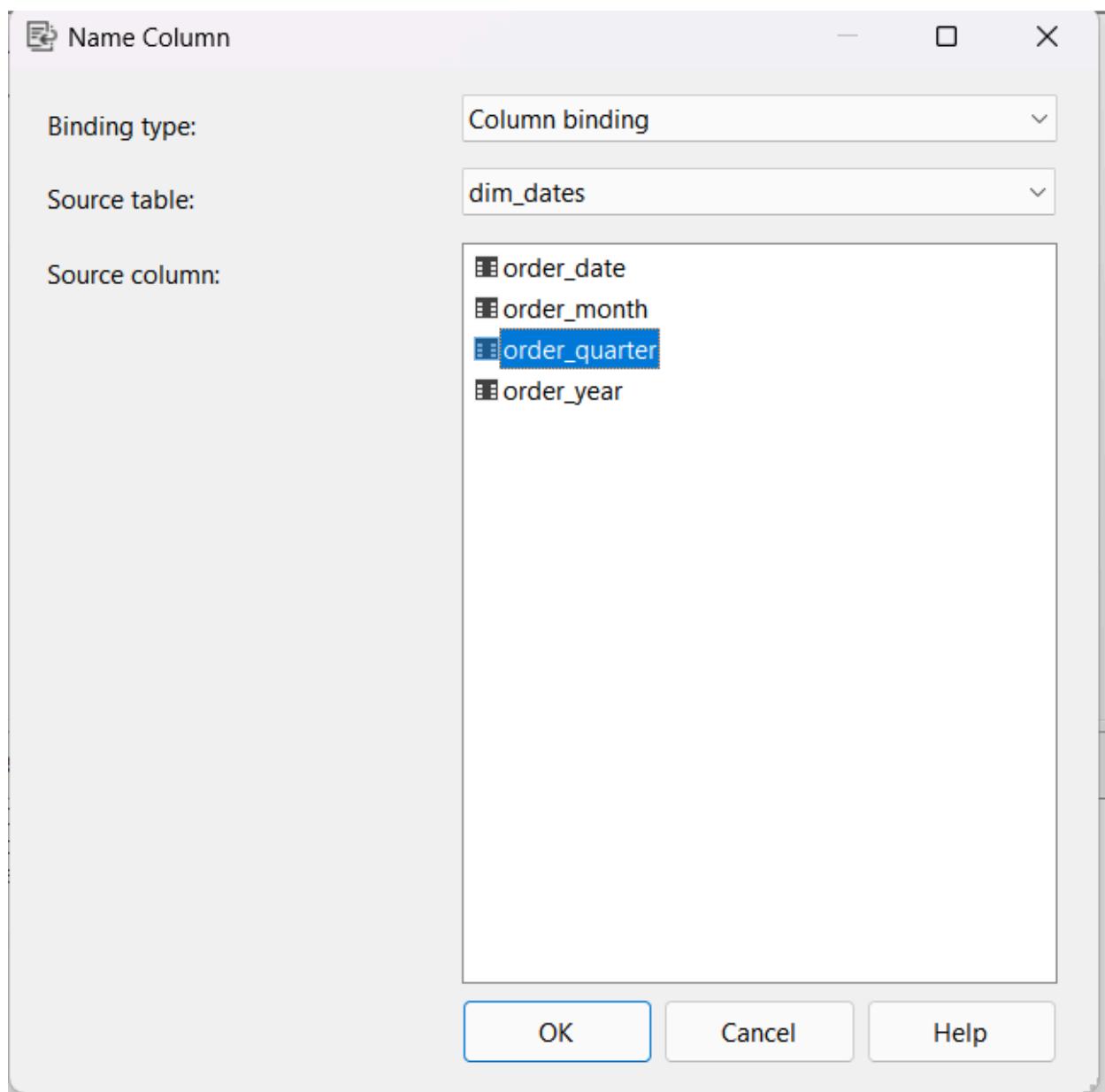
- Tạo Name Column của thuộc tính Order_Month



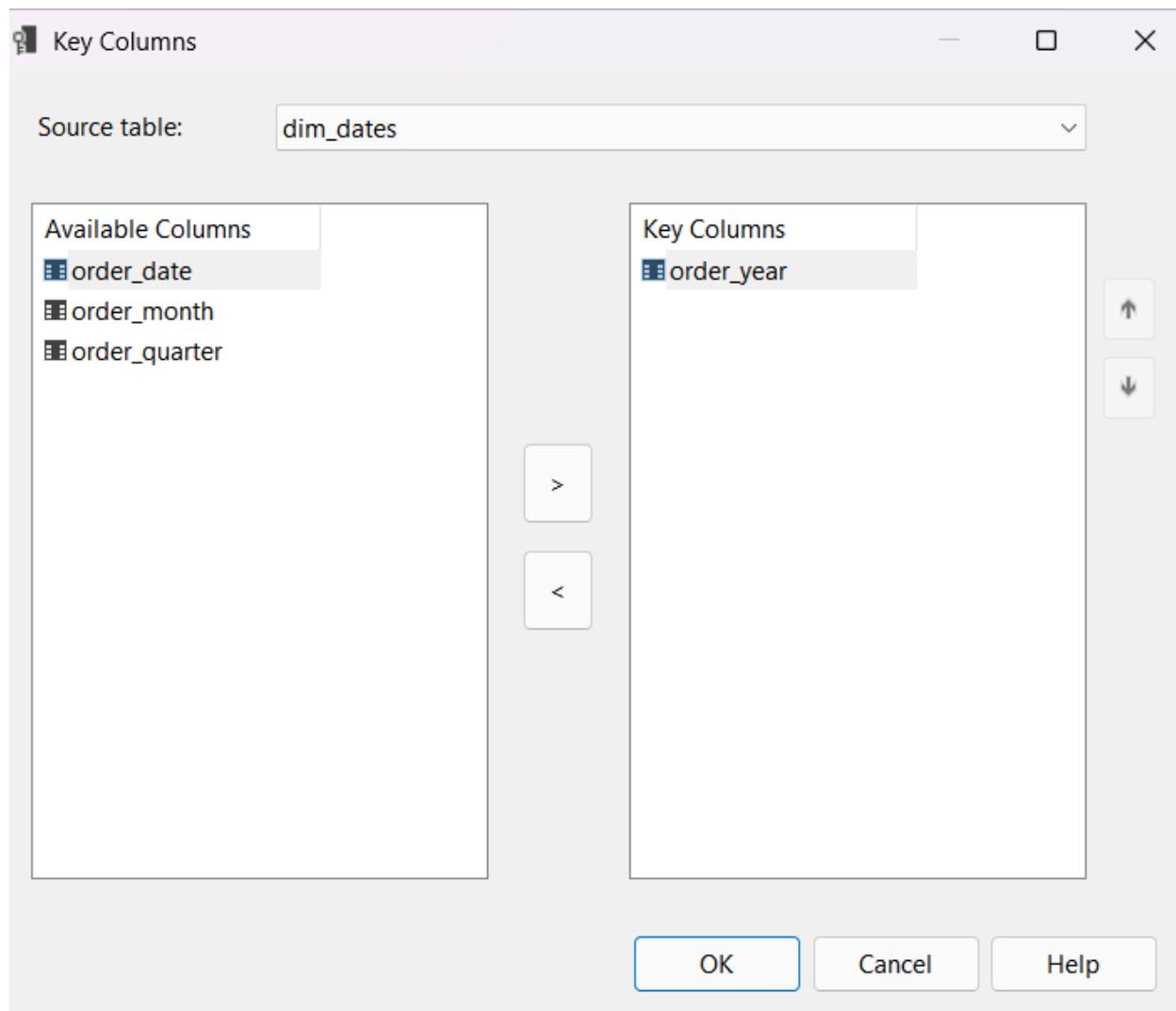
- Chính Key Columns cho Order_Quarter. Lấy Key là chính nó và các thuộc tính cấp cao hơn.



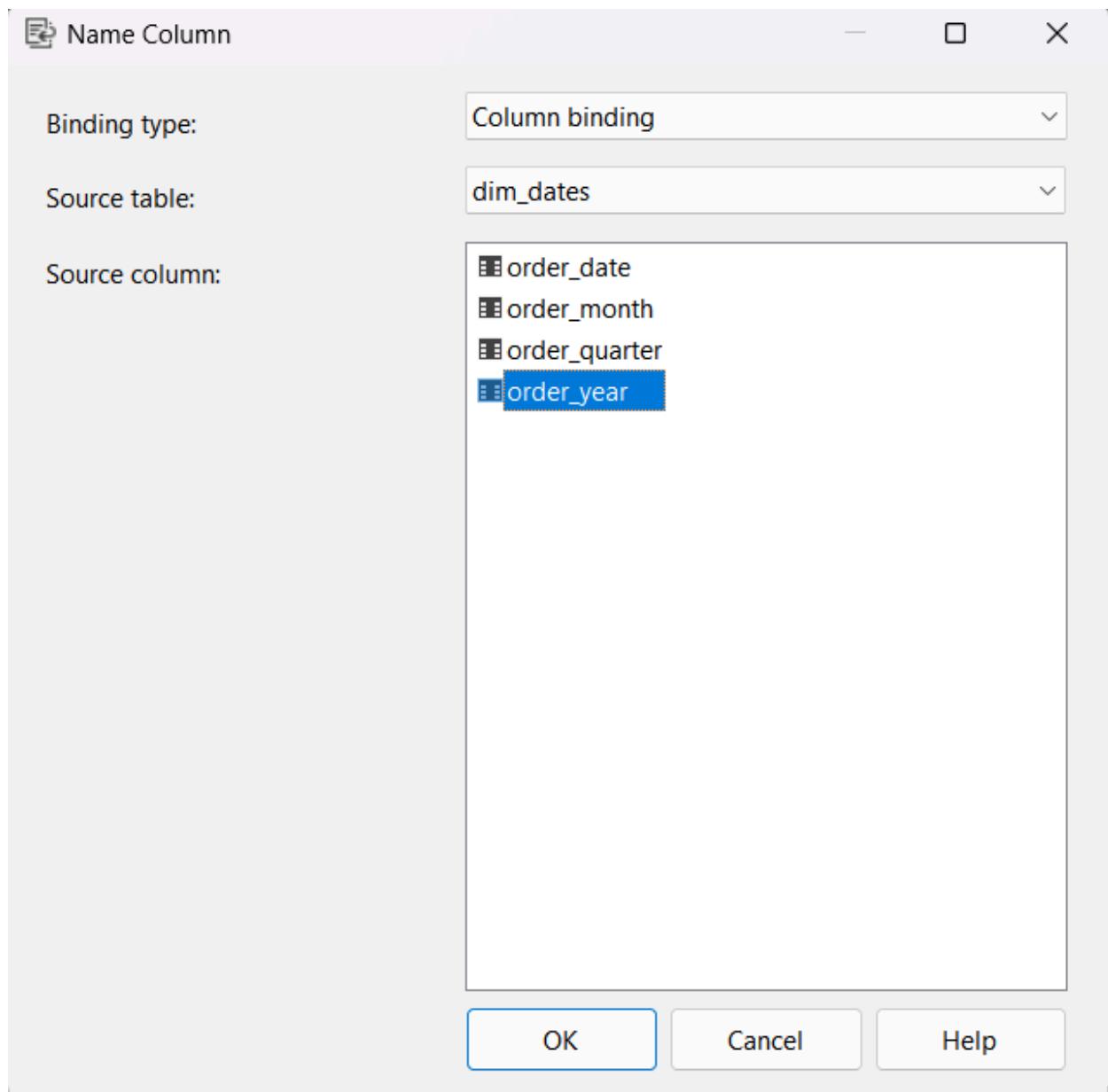
- Tạo Name Column của thuộc tính Order_Quarter



- Chính Key Columns cho Order_Year. Lấy Key là chính nó

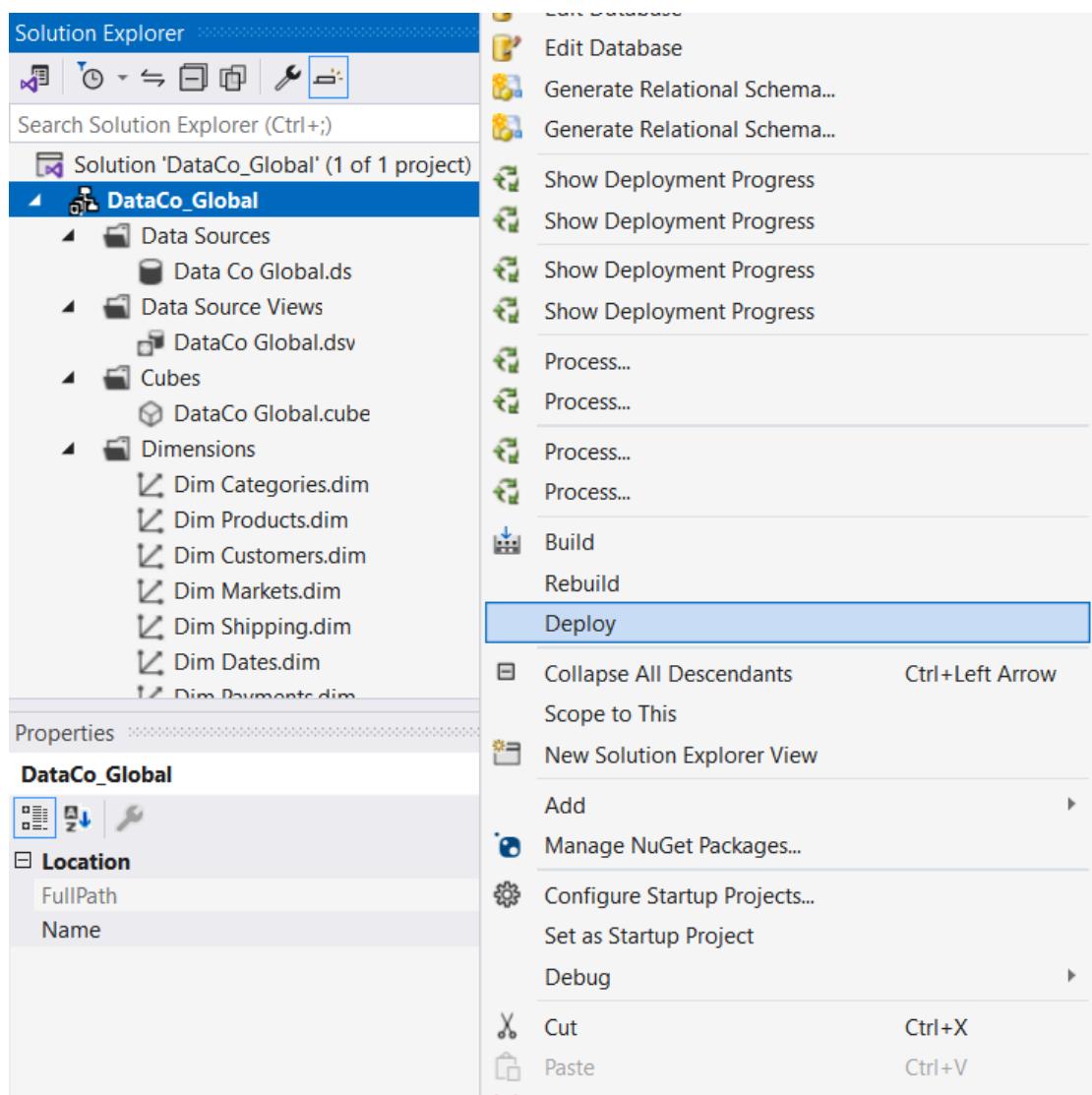


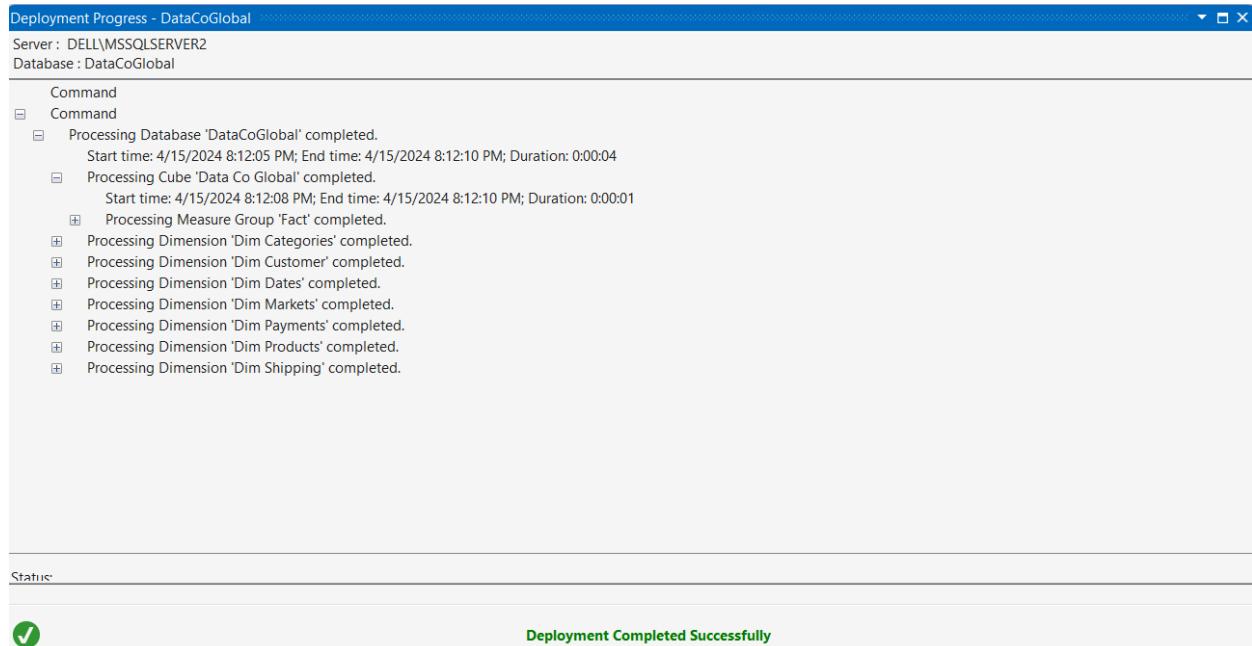
- Tạo Name Column của thuộc tính Order_Year



3.7. CHẠY PROJECT SSAS

Chọn Project -> Deploy





3.7. THỰC HIỆN CÁC CÂU TRUY VẤN

3.7.1. Liệt kê Top 5 sản phẩm bán chạy nhất theo các năm.

a. MDX

```
Select [Measures].[Product Quantity] on Columns
, NON EMPTY generate (
    [Dim Dates].[Order Year].[Order Year],
    TOPCount (
        [Dim Dates].[Order Year].CurrentMember * [Dim Products].[Product Name].[Product Name]
        , 5
        , [Measures].[Product Quantity]
    )
) on Rows
From [Data Co Global];
```

		Product Quantity
2015	Perfect Fitness Perfect Rip Deck	26755
2015	Nike Men's Dri-FIT Victory Golf Polo	22840
2015	O'Brien Men's Neoprene Life Vest	21226
2015	Nike Men's Free 5.0+ Running Shoe	13576
2015	Under Armour Girls' Toddler Spine Surge Runni	11761
2016	Perfect Fitness Perfect Rip Deck	27136
2016	Nike Men's Dri-FIT Victory Golf Polo	22859
2016	O'Brien Men's Neoprene Life Vest	21047
2016	Nike Men's Free 5.0+ Running Shoe	12982
2016	Under Armour Girls' Toddler Spine Surge Runni	11235
2017	Perfect Fitness Perfect Rip Deck	19807
2017	Nike Men's Dri-FIT Victory Golf Polo	17257
2017	O'Brien Men's Neoprene Life Vest	15530
2017	Nike Men's Free 5.0+ Running Shoe	10122
2017	Under Armour Girls' Toddler Spine Surge Runni	8739
2018	Fighting video games	610
2018	Summer dresses	377
2018	Toys	276
2018	Adult dog supplies	246
2018	Smart watch	212

b. Manual

Tạo NameSet [Cau 1]

Name: [Cau 1]

Expression

```
generate (
    [Dim Dates].[Order Year].[Order Year],
    TOPCount (
        [Dim Dates].[Order Year].CurrentMember * [Dim Products].[Product Name].[Product Name]
        , 5
        , [Measures].[Product Quantity]
    )
)
No issues found
```

Ln: 8 Ch: 2 TABS CRLI

Additional Properties

Type:	Dynamic
Display folder:	Caculated Sets

Kết quả

c. Pivot Excel

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable Fields pane on the right indicates that 'Product Quantity' is selected as a value field. The main table area shows sales data for various products across four years (2015, 2016, 2017, 2018). The data includes items like Perfect Fitness Perfect Rip Deck, Nike Men's Dri-FIT Victory Golf Polo, O'Brien Men's Neoprene Life Vest, etc., with their respective quantities.

	Product	Quantity
2015	Perfect Fitness Perfect Rip Deck	26755
	Nike Men's Dri-FIT Victory Golf Polo	22840
	O'Brien Men's Neoprene Life Vest	21226
	Nike Men's Free 5.0+ Running Shoe	13576
	Under Armour Girls' Toddler Spine Surge Runni	11761
2016	Perfect Fitness Perfect Rip Deck	27136
	Nike Men's Dri-FIT Victory Golf Polo	22859
	O'Brien Men's Neoprene Life Vest	21047
	Nike Men's Free 5.0+ Running Shoe	12982
	Under Armour Girls' Toddler Spine Surge Runni	11235
2017	Perfect Fitness Perfect Rip Deck	19807
	Nike Men's Dri-FIT Victory Golf Polo	17257
	O'Brien Men's Neoprene Life Vest	15530
	Nike Men's Free 5.0+ Running Shoe	10122
	Under Armour Girls' Toddler Spine Surge Runni	8739
2018	Fighting video games	610
	Summer dresses	377
	Toys	276
	Adult dog supplies	246
	Smart watch	212

3.7.2. Tổng doanh số bán hàng theo từng khu vực trong năm 2017, sắp xếp tăng dần theo doanh số

a. MDX

```
Select [Measures].[Sales] on Columns
, Order(
    [Dim Markets].[Market].[Market].Members
    , [Measures].[Sales]
    , asc
) on Rows
From [Data Co Global]
Where [Dim Dates].[Order Year].&[2017]
```

	Sales
USCA	23063.1904983521
Africa	247830.164852142
Pacific Asia	1133058.90028477
Europe	5196492.17457485
LATAM	5205424.37261772

b. Manual

Tạo NameSet [Cau 2]

Name: [Cau 2]

Expression:

```
Order(
    [Dim Markets].[Market].[Market].Members
    , [Measures].[Sales]
    , ASC
)
No issues found
```

Ln: 7 Ch: 1 TABS CRLF

Additional Properties:

Type: Dynamic

Display folder: Calculated Sets

Kết quả

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Markets	Market	In	Cau 2	<input type="checkbox"/> <input type="checkbox"/>
Dim Dates	Order Year	Equal	{ { 2017 } }	<input checked="" type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>
Order Year	Market	Sales		
2017	Africa	247830.164852142		
2017	Europe	5196492.17457485		
2017	LATAM	5205424.37261772		
2017	Pacific Asia	1133058.90028477		
2017	USCA	23063.1904983521		

c. Pivot Excel

The screenshot shows a Microsoft Excel spreadsheet titled "Book1 - E...". The ribbon menu is visible at the top, with the "PivotTable Analyze" tab selected. The main area contains a table with the following data:

	PhanCapDate	Sales
1	2017	
2		
3	Row Labels	Sales
4	USCA	23063.1905
5	Africa	247830.1649
6	Pacific Asia	1133058.9
7	Europe	5196492.175
8	LATAM	5205424.373
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		

The "PivotTable Fields" pane on the right side of the screen lists fields under the "Fact" category, with "Sales" checked. The "Filters" section shows "PhanCapDate" as a filter. The "Rows" section shows "Cau 2" as a row label, and the "Values" section shows "Sales" as a value.

3.7.3. Tìm ra danh mục có số lượng sản phẩm bán được cao nhất

a. MDX

```
Select [Measures].[Product Quantity] on Columns
, TOPCount(
    [Dim Products].[Category Name].[Category Name]
, 1
, [Measures].[Product Quantity]
) on Rows
From [Data Co Global]
```

	Product Quantity
Cleats	73734

b. Manual

Tạo NameSet [Cau 3]

Name: [Cau 3]

Expression: TOPCOUNT([Dim Products].[Category Name].[Category Name].ALLMEMBERS, 1 , [Measures].[Product Quantity])

No issues found

Ln: 1 Ch: 103 SPC CRLF

Additional Properties:

Type: Dynamic

Display folder: Calculated Sets

Kết quả

Dimension	Hierarchy	Operator	Filter Expression
Dim Products	Category Name	In	Cau 3
<Select dimension>			
Category Name			Product Quantity
Cleats			73734

c. Pivot Excel

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable Fields pane on the right indicates that the calculated set 'Cau 3' has been selected. The main area displays a single data point: 'Cleats' with a value of 73734.

Row Labels	Product Quantity
Cleats	73734

3.7.4. Tính doanh thu trung bình trên 1 sản phẩm bán ra vào các quý của năm 2016, thuộc danh mục Fishing

a. MDX

```
With member [Measures].[AvgSales]
as (SUM([Measures].[Sales]) / SUM([Measures].[Product Quantity]))
Select [Measures].[AvgSales] on Columns,
Non empty Generate([Dim Dates].[Order Quarter].[Order Quarter].Members,
      Filter([Dim Dates].[Order Quarter].CurrentMember,
      [Dim Products].[Category Name].&[Fishing])) on Rows
From [Data Co Global]
Where [Dim Dates].[Order Year].&[2016]
```

	AvgSales
1	90.199170013115
2	89.703164803065
3	89.4396981970683
4	89.001757281262

b. Manual

Tạo độ đo [AvgSales]

Name: [AvgSales]

Parent Properties

Parent hierarchy: Measures

Parent member: Change

Expression

```
SUM([Measures].[Sales]) / SUM([Measures].[Product Quantity])
```

No issues found Ln: 1 Ch: 61 SPC CRLF

Additional Properties

Format string:

Visible: True

Non-empty behavior:

Associated measure group: (Undefined)

Display folder:

Color Expressions

Font Expressions

Tạo NameSet [Cau 4]

Name: [Cau 4]

Expression

```
generate([Dim Dates].[Order Quarter].[Order Quarter].Members,
        Filter([Dim Dates].[Order Quarter].CurrentMember,
        [Dim Products].[Category Name].&[Fishing]))
```

No issues found Ln: 6 Ch: 1 TABS CRLF

Additional Properties

Type: Dynamic

Display folder:

Kết quả

Dimension	Hierarchy	Operator	Filter Expression
Dim Dates	Order Quarter	In	Cau 4
Dim Dates	PhanCapDate	Equal	{ 2016 }
<Select dimension>			
Order Quarter	AvgSales		
1	90.199170013115		
2	89.703164803065		
3	89.4396981970683		
4	89.001757281262		

c. Pivot Excel

The screenshot shows a Microsoft Excel interface with a PivotTable set up. The PivotTable Fields pane on the right indicates that 'PhanCapDate' is selected under the 'Rows' category. The main worksheet displays data for 'AvgSales' across four categories of 'PhanCapDate'. The data is as follows:

Row Labels	AvgSales
1	90.19917001
2	89.70316480
3	89.4396982
4	89.00175728

3.7.5. Tìm ra khách hàng có số lần mua hàng cao nhất và thấp nhất trong năm 2017

a. MDX

```
Select [Measures].[Fact Count] on Columns,
Union(
    TopCount(
        [Dim Customer].[Customer ID].Children*[Dim Customer].[Customer Name].Children
        ,1
        ,[Measures].[Fact Count]
    ),
    BottomCount(
        [Dim Customer].[Customer ID].Children*[Dim Customer].[Customer Name].Children
        ,1
        ,[Measures].[Fact Count]
    )
)on Rows
From [Data Co Global]
Where [Dim Dates].[Order Year].&[2017]
```

		Fact Count
5004	Dorothy Friedman	28
20757	Katelyn Gonzales	(null)

b. Manual

Tạo NameSet [Cau 5]

Name: [Cau 5]

Expression

```
Union(
    TopCount(
        [Dim Customer].[Customer ID].Children*[Dim Customer].[Customer Name].Children
        ,1
        ,[Measures].[Fact Count]
    ),
    BottomCount(
        [Dim Customer].[Customer ID].Children*[Dim Customer].[Customer Name].Children
        ,1
        ,[Measures].[Fact Count]
    )
)

```

No issues found

Ln: 13 Ch: 1 TABS

Additional Properties

Type: Dynamic

Display folder: Calculated Sets

Kết quả

c. Pivot Excel

Row Labels	Fact Count
5654	
Jerry Smith	47
20757	
Katelyn Gonzales	1

Đề tài: Phân tích doanh số bán hàng Data Global

3.7.6. Liệt kê Top 5 sản phẩm bán chạy nhất Africa

a. MDX

```
Select [Measures].[Product Quantity] on Columns  
, Head(  
Order(  
    [Dim Products].[Product Name].[Product Name].Members  
    , [Measures].[Product Quantity]  
    , desc  
)  
, 5  
) on Rows  
From [Data Co Global]  
Where [Dim Markets].[Market].&[Africa]
```

	Product Quantity
Perfect Fitness Perfect Rip Deck	4975
Nike Men's Dri-FIT Victory Golf Polo	4182
O'Brien Men's Neoprene Life Vest	3975
Nike Men's Free 5.0+ Running Shoe	2551
Under Armour Girls' Toddler Spine Surge Runni	2060

b. Manual

Tạo NameSet [Câu 6]

Name: [Cau 6]

Expression

```
Head(
    Order(
        [Dim Products].[Product Name].[Product Name].Members
        , [Measures].[Product Quantity]
        , desc
    ), 5
)
```

No issues found

Ln: 8 Ch: 1 MIXED CRLF

Additional Properties

Type: Dynamic

Display folder: Calculated Sets

Kết quả

Product Name	Product Quantity
Nike Men's Dri-FIT Victory Golf Polo	4182
Nike Men's Free 5.0+ Running Shoe	2551
O'Brien Men's Neoprene Life Vest	3975
Perfect Fitness Perfect Rip Deck	4975
Under Armour Girls' Toddler Spine Surge Runni	2060

c. Pivot Excel

A screenshot of Microsoft Excel showing a PivotTable setup. The PivotTable Fields pane on the right shows 'Market' selected under 'Dim Markets'. The main table displays product quantities by market.

Market	Product Quantity
Perfect Fitness Perfect Rip Deck	4975
Nike Men's Dri-FIT Victory Golf Polo	4182
O'Brien Men's Neoprene Life Vest	3975
Nike Men's Free 5.0+ Running Shoe	2551
Under Armour Girls' Toddler Spine Surge Runni	2060

3.7.7. Cho biết khu vực nào có số lượng đơn hàng bị hủy cao nhất theo từng năm.

a. MDX

```
Select [Measures].[Fact Count] on Columns
, Generate (
    [Dim Dates].[Order Year].[Order Year].Members
    , TopCount(
        [Dim Dates].[Order Year].CurrentMember * [Dim Markets].[Market].Members
        , 1
        , [Measures].[Fact Count]
    )
) on Rows
From [Data Co Global]
Where [Dim Payments].[Order status].&[CANCELED]
```

		Fact Count
2015	LATAM	582
2016	USCA	562
2017	LATAM	521
2018	Pacific Asia	45

b. Manual

Tạo NameSet [Cau 7]

Name: [Cau 7]

Expression:

```
generate(
    [Dim Dates].[Order Year].[Order Year].Members
    ,TopCount(
        [Dim Dates].[Order Year].CurrentMember * [Dim Markets].[Market].[Market].Members
        , 1
        , [Measures].[Fact Count]
    )
)
```

No issues found

Ln: 8 Ch: 4 Col: 10 TABS CRLF

Additional Properties

Type: Dynamic

Display folder: Calculated Sets

Kết quả

c. Pivot Excel

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable Fields pane on the right indicates that 'Order Status' is selected from the 'Dim Payments' dimension. The main table shows the following data:

	Order Status	Fact Count
2015	LATAM	582
2016	USCA	562
2017	LATAM	521
2018	Pacific Asia	45

3.7.8. Cho biết loại hình thanh toán nào được ưa chuộng nhất ở LATAM

a. MDX

```

SELECT [Measures].[Fact Count] on Columns,
TopCount(
    [Dim Payments].[Type].[Type].MEMBERS
    ,1
    ,[Measures].[Fact Count]
) on Rows
FROM [Data Co Global]
Where [Dim Markets].[Market].&[LATAM]

```



b. Manual

Tạo NameSet

Name: [Cau 8]

Expression:

```
TopCount(
    [Dim Payments].[Type].[Type].MEMBERS
    ,1
    ,[Measures].[Fact Count]
)
```

No issues found

Ln: 7 Ch: 1 TABS CRLF

Additional Properties:

Type: Dynamic

Display folder:

Kết quả

Dimension	Hierarchy	Operator	Filter Expression
Dim Payments	Type	In	Cau 8
Dim Markets	Market	Equal	{ LATAM }
<Select dimension>			
Type	Fact Count		
DEBIT	19441		

c. Pivot Excel

A screenshot of Microsoft Excel showing a PivotTable setup. The PivotTable Fields pane on the right shows 'Dim Markets' expanded, with 'Market' selected. The main area shows a table with Row Labels 'DEBIT' and Fact Count '19441'. The PivotTable Fields pane also shows 'Rows' set to 'Cau 8' and 'Values' set to 'Fact Count'.

3.7.9. Liệt kê doanh số đóng góp của nhóm khách hàng ‘Corporate’ theo từng quý, năm

a. MDX

```
Select [Measures].[Sales] on Columns
, DrillDownLevel(
    [Dim Dates].[PhanCapDate].[Order Year].Members
) on Rows
From [Data Co Global]
Where [Dim Customer].[Customer Segment].&[Corporate]
```

	Sales
2015	3778905.37383842
1	913965.647462845
2	939523.018264771
3	959261.848901749
4	966154.859209061
2016	3671523.60243225
1	911352.918087006
2	944247.39881897
3	886173.437475204
4	929749.848051071
2017	3611751.7950058
1	968105.558855057
2	970404.819301605
3	997340.727539063
4	675900.689310074
2018	104730.191583633
1	104730.191583633

b. Manual

Tạo NameSet [Cau 9]

Name:

Expression

```
DrillDownLevel(
    [Dim Dates].[PhanCapDate].[Order Year].Members
)
```

No issues found

Ln: 6

Additional Properties

Type: Dynamic

Display folder:

Kết quả

Order Year	Order Quarter	Sales
2015	1	913965.647462845
2015	2	939523.018264771
2015	3	959261.848901749
2015	4	966154.859209061
2016	1	911352.918087006
2016	2	944247.39881897
2016	3	886173.437475204
2016	4	929749.848051071
2017	1	968105.558855057
2017	2	970404.819301605
2017	3	997340.727539063
2017	4	675900.689310074
2018	1	104730.191583633

c. Pivot Excel

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable Fields pane on the right indicates that 'Customer Segment' is selected under the 'Dim Customer' category. The main table displays sales data for various years and customer segments. The data includes:

	Row Labels	Sales
2015	Corporate	3778905.374
1	Corporate	913965.6475
2	Corporate	939523.0183
3	Corporate	959261.8489
4	Corporate	966154.8592
2016	Corporate	3671523.602
1	Corporate	911352.9181
2	Corporate	944247.3988
3	Corporate	886173.4375
4	Corporate	929749.8481
2017	Corporate	3611751.795
1	Corporate	968105.5589
2	Corporate	970404.8193
3	Corporate	997340.7275
4	Corporate	675900.6893
2018	Corporate	104730.1916
1	Corporate	104730.1916

3.7.10. Liệt kê những khách hàng có số lượng đơn hàng đã mua > 2 và < 5

a. MDX

```
Select [Measures].[Fact Count] on Columns
, Intersect(
    Filter([Dim Customer].[Customer Name].Children, [Measures].[Fact Count] > 2),
    Filter([Dim Customer].[Customer Name].Children, [Measures].[Fact Count] < 5)
) on Rows
From [Data Co Global]
```

	Fact Count
Aaron Cordova	3
Alan Avila	4
Alan Padilla	3
Alexander Harris	3
Alexander Jordan	3
Alexander Williams	4
Amanda Mann	3
Amanda Moore	3
Amber Armstrong	4
Amber Carr	4
Amber Peterson	4
Amber Zimmerman	3
Andrea Alexander	4
Andrea Diaz	3
Andrea Melton	3
Andrea Randall	4
Angela English	4
Ann Zimmerman	3
Anthony Spence	4
Arthur Hunter	4
Austin Diaz	3
Barbara Hogan	3
Barbara Jordan	4
Barbara Kemp	4
Benjamin Davenport	3
Betty Cortez	3

b. **Manual**

Tạo NameSet [Câu 10]

Name: [Cau 10]

Expression

```
Intersect(
    Filter([Dim Customer].[Customer Name].Children, [Measures].[Fact Count] > 2),
    Filter([Dim Customer].[Customer Name].Children, [Measures].[Fact Count] < 5)
)
```

No issues found

Ln: 6 Ch: 1 TABS

Additional Properties

Type: Dynamic

Display folder:

Kết quả:

Dimension	Hierarchy	Operator	Filter Expression
Dim Customer	Customer Name	In	Cau 10
<Select dimension>			

Customer Name	Fact Count
Aaron Cordova	3
Alan Avila	4
Alan Padilla	3
Alexander Harris	3
Alexander Jordan	3
Alexander Willia...	4
Amanda Mann	3
Amanda Moore	3
Amber Armstrong	4
Amber Carr	4
Amber Peterson	4
Amber Zimmerman	3
Andrea Alexander	4
Andrea Diaz	3
Andrea Melton	3
Andrea Randall	4
Angela English	4
Ann Zimmerman	3
Anthony Spence	4
Arthur Hunter	4
Austin Diaz	3
Barbara Hogan	3
Barbara Jordan	4
Barbara Kemn	4

c. Pivot Excel

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable Fields pane on the right indicates that 'Fact Count' is selected under the 'Fact' category. The main table area shows rows of names from Aaron Cordova to Arthur Hunter, each associated with a value of 3 or 4 in column B.

	Fact Count
Aaron Cordova	3
Alan Avila	4
Alan Padilla	3
Alexander Harris	3
Alexander Jordan	3
Alexander Williams	4
Amanda Mann	3
Amanda Moore	3
Amber Armstrong	4
Amber Carr	4
Amber Peterson	4
Amber Zimmerman	3
Andrea Alexander	4
Andrea Diaz	3
Andrea Melton	3
Andrea Randall	4
Angela English	4
Ann Zimmerman	3
Anthony Spence	4
Arthur Hunter	4

3.7.11. Liệt kê số lượng đơn hàng bán ra theo từng năm ở tất cả thị trường, trừ thị trường ‘LATAM’.

a. MDX

```

SELECT [Dim Dates].[Order Year].[Order Year].Members ON COLUMNS,
EXCEPT(
    [Dim Markets].[Market].[Market].Members,
    {[Dim Markets].[Market].&[LATAM]}
) ON ROWS
FROM [Data Co Global]
Where [Measures].[Fact Count]

```

	2015	2016	2017	2018
Africa	(null)	10356	1258	(null)
Europe	24914	3872	21463	(null)
Pacific Asia	11961	22641	4531	2122
USCA	(null)	25681	118	(null)

b. Manual

Tạo NameSet [Câu 11]

Name: [Cau 11]

Expression:

```

EXCEPT(
    [Dim Markets].[Market].[Market].Members,
    {[Dim Markets].[Market].&[LATAM]}
)

```

No issues found

Additional Properties:

Type: Dynamic

Display folder:

Kết quả

Dimension	Hierarchy	Operator	Filter Expression
Dim Markets	Market	In	Cau 11
<Select dimension>			

Market	Order Year	Fact Count
Africa	2016	10356
Africa	2017	1258
Europe	2015	24914
Europe	2016	3872
Europe	2017	21463
Pacific...	2015	11961
Pacific...	2016	22641
Pacific...	2017	4531
Pacific...	2018	2122
USCA	2016	25681
USCA	2017	118

c. Pivot Excel

The screenshot shows a Microsoft Excel spreadsheet titled "tmpE884". The PivotTable Fields pane on the right side of the ribbon indicates the following fields:

- Row Labels: PhanCapDate, Order Year, Order Quarter, Order Month
- Columns: Fact Count
- Values: Fact Count

The main worksheet area displays the following data:

	2015	2016	2017	2018	Grand Total
Africa		10356	1258		11614
Europe	24914	3872	21463		50249
Pacific Asia	11961	22641	4531	2122	41255
USCA		25681	118		25799

3.7.12. Truy vấn ra các đơn vận chuyển có phần trăm giảm giá > 0.2 theo từng khu vực trong năm 2015, sắp xếp giảm dần theo phần trăm giảm giá.

a. MDX

```
Select [Measures].[Discount Rate] on Columns  
, Generate(  
    [Dim Markets].[Market].[Market].Members  
, Order(  
        Filter(  
            [Dim Markets].[Market].CurrentMember *  
            [Dim Shipping].[Shipping ID].[Shipping ID].Members  
, [Measures].[Discount Rate] > 0.2)  
, [Measures].[Discount Rate]  
, DESC  
)) on Rows  
From [Data Co Global]  
Where [Dim Dates].[Order Year].&[2015];
```

		Discount Rate
Europe	933.00000000000	0.25
Europe	1389.00000000000	0.25
Europe	1579.00000000000	0.25
Europe	1580.00000000000	0.25
Europe	1581.00000000000	0.25
Europe	1646.00000000000	0.25
Europe	1798.00000000000	0.25
Europe	1808.00000000000	0.25
Europe	1810.00000000000	0.25
Europe	2001.00000000000	0.25
Europe	2016.00000000000	0.25
Europe	2017.00000000000	0.25
Europe	2139.00000000000	0.25
Europe	2140.00000000000	0.25
Europe	2141.00000000000	0.25
Europe	2143.00000000000	0.25
Europe	2144.00000000000	0.25
Europe	2157.00000000000	0.25
Europe	2158.00000000000	0.25
Europe	2532.00000000000	0.25
Europe	2763.00000000000	0.25
Europe	2769.00000000000	0.25
Europe	2771.00000000000	0.25
Europe	2797.00000000000	0.25
Europe	2801.00000000000	0.25

b. Manual

Tạo NameSet [Cau 12]

Name: [Cau 12]

Expression

```
Generate(
    [Dim Markets].[Market].[Market].Members
    , Order(
        Filter(
            [Dim Markets].[Market].CurrentMember * [Dim Shipping].[Shipping ID].[Shipping ID].Members
            , [Measures].[Discount Rate] > 0.2)
            , [Measures].[Discount Rate]
            , DESC
        )))

```

No issues found

Ln: 13 Ch: 1 MIXED CRLF

Additional Properties

Type: Dynamic

Display folder:

c. Pivot Excel

Row Labels	Discount Rate
1 Africa	
2 143.0000000000	0.200000003
3 205.0000000000	0.200000003
4 218.0000000000	0.200000003
5 219.0000000000	0.25
6 285.0000000000	0.200000003
7 286.0000000000	0.25
8 287.0000000000	0.25
9 294.0000000000	0.200000003
10 299.0000000000	0.200000003
11 750.0000000000	0.200000003
12 755.0000000000	0.200000003
13 1321.0000000000	0.25
14 1322.0000000000	0.25
15 1336.0000000000	0.25
16 1337.0000000000	0.25
17 1338.0000000000	0.25
18 1339.0000000000	0.25
19 1353.0000000000	0.200000003
20 2619.0000000000	0.200000003
21 2625.0000000000	0.25
22 2628.0000000000	0.200000003
23 2634.0000000000	0.25
24 3819.0000000000	0.200000003
25 3857.0000000000	0.200000003
26	

3.7.13. Truy vấn ra các đơn hàng được giao đúng thời gian dự kiến

a. MDX

```
Select [Measures].[Sales] on Columns
, TopCount (
    Filter (
        [Dim Shipping].[Shipping ID].[Shipping ID] * [Dim Shipping].[Day Shipping Real].[Day Shipping Real]
        , [Dim Shipping].[Day Shipping Real] = [Dim Shipping].[Day Shipment Scheduled]
    )
    , 10
    , [Measures].[Sales]
) on Rows
From [Data Co Global]
```

		Sales
12607.000000000000	4	1999.98999023438
13344.000000000000	5	1999.98999023438
48281.000000000000	2	1999.98999023438
59846.000000000000	3	1999.98999023438
63989.000000000000	5	1999.98999023438
82116.000000000000	4	1999.98999023438
82129.000000000000	3	1999.98999023438
82160.000000000000	5	1999.98999023438
83213.000000000000	6	1999.98999023438
141112.000000000000	2	1999.98999023438

b. Manual

Tạo NameSet [Cau 13]

Name:

Expression i

```
TopCount (
    Filter (
        [Dim Shipping].[Shipping ID].[Shipping ID] * [Dim Shipping].[Day Shipping Real].[Day
            Shipping Real]
        , [Dim Shipping].[Day Shipping Real] = [Dim Shipping].[Day Shipment Scheduled]
    )
    , 10
    , [Measures].[Sales]
)
```

No issues found Ln: 11 Ch: 1 TABS CRLF

Additional Properties

Type:

Display folder:

c. Pivot Excel

A screenshot of Microsoft Excel showing a PivotTable setup. The PivotTable Fields pane on the right shows the 'Fact' category with 'Sales' selected. The main area displays a list of sales values from 12607.0000000000 down to 141112.0000000000. The PivotTable Fields pane also shows the 'Rows' section with 'Cau 13' and the 'Values' section with 'Sales'.

3.7.14. Liệt kê Top 5 sản phẩm có số lượng bán cao nhất với ký tự bắt đầu là 'P', sắp xếp theo thứ tự giảm dần.

a. MDX

```
Select [Measures].[Fact Count] on Columns
, TopCount(
    Filter(
        [Dim Products].[Product Name].[Product Name].Members,
        Left([Dim Products].[Product Name].CurrentMember.Name, 1) = "P"
    )
    , 5
    , [Measures].[Fact Count]
) on Rows
From [Data Co Global]
```

	Fact Count
Perfect Fitness Perfect Rip Deck	24515
Pelican Sunstream 100 Kayak	15500
Porcelain crafts	481
Polar FT4 Heart Rate Monitor	60
Polar Loop Activity Tracker	54

b. Manual

Tạo NameSet

Name: [Cau 14]

Expression:

```
TopCount(
    Filter(
        [Dim Products].[Product Name].[Product Name].Members,
        Left([Dim Products].[Product Name].CurrentMember.Name, 1) = "P"
    ),
    5
    , [Measures].[Fact Count]
)
```

No issues found

Ln: 10 Ch: 1 TABS CRLF

Additional Properties

Type: Dynamic

Display folder:

Kết quả

Dimension	Hierarchy	Operator	Filter Expression
Dim Products	# Product Name	In	Cau 14
<Select dimension>			
Product Name			Fact Count
Pelican Sunstream 100 Kayak			15500
Perfect Fitness Perfect Rip Deck			24515
Polar FT4 Heart Rate Monitor			60
Polar Loop Activity Tracker			54
Porcelain crafts			481

c. Pivot Excel

A screenshot of Microsoft Excel showing a PivotTable. The PivotTable Fields pane on the right indicates that 'Fact Count' is selected under the 'Fact' category. The main table shows the following data:

	Fact Count
Perfect Fitness Perfect Rip Deck	24515
Pelican Sunstream 100 Kayak	15500
Porcelain crafts	481
Polar FT4 Heart Rate Monitor	60
Polar Loop Activity Tracker	54

3.7.15. Với mỗi danh mục sản phẩm, liệt kê 3 sản phẩm có doanh số đóng góp cao nhất trong năm 2015

a. MDX

```
Select [Measures].[Sales] on Columns
, Non Empty (
    Generate(
        [Dim Products].[Category Name].[Category Name].Members
        , TopCount(
            [Dim Products].[Category Name].CurrentMember *
            [Dim Products].[Product Name].[Product Name].Members
            , 3
            ,[Measures].[Sales]
        ))
    ) on Rows
From [Data Co Global]
Where [Dim Dates].[Order Year].&[2015];
```

		Sales
Accessories	Team Golf San Francisco Giants Putter Grip	10720.7099151611
Accessories	Team Golf Texas Longhorns Putter Grip	10120.9499282837
Accessories	Team Golf St. Louis Cardinals Putter Grip	9371.24992370605
Baseball & Softball	adidas Men's F10 Messi TRX FG Soccer Cleat	24355.940700531
Baseball & Softball	adidas Kids' F5 Messi FG Soccer Cleat	10636.960231781
Boxing & MMA	Under Armour Women's Micro G Skulpt Running S	18030.1604003906
Camping & Hiking	Diamondback Women's Serene Classic Comfort Bi	1480101.37420654
Cardio Equipment	Nike Men's Free 5.0+ Running Shoe	1357464.23821259
Cardio Equipment	Nike Women's Tempo Shorts	12180
Cleats	Perfect Fitness Perfect Rip Deck	1605032.4948616
Electronics	Titleist Pro V1x High Numbers Personalized Go	20951.9706344604
Electronics	Titleist Pro V1 High Numbers Personalized Gol	19236.3005981445
Electronics	Under Armour Men's Compression EV SL Slide	18535.8803482056
Fishing	Field & Stream Sportsman 16 Gun Fire Safe	2496275.24856567
Fitness Accessories	Under Armour Hustle Storm Medium Duffle Bag	13821.0503234863
Girls' Apparel	adidas Youth Germany Black/Red Away Match Soc	30450
Girls' Apparel	TYR Boys' Team Digi Jammer	17755.5603103638
Girls' Apparel	Nike Men's Deutschland Weltmeister Winners Bl	10740
Golf Apparel	Top Flite Women's 2014 XL Hybrid	7576.20992660522
Golf Balls	Glove It Women's Imperial Golf Glove	8995.49988746643
Golf Balls	Hirzl Women's Soffit Flex Golf Glove	7681.72988700867
Golf Balls	Glove It Women's Mod Oval Golf Glove	6896.54991531372
Golf Gloves	Clicgear Rovic Cooler Bag	14796.3001976013
Golf Gloves	Bag Boy Beverage Holder	9896.03988647461
Golf Gloves	Clicgear 8.0 Shoe Brush	4095.89997291565
Golf Shoes	LIJA Women's Eyelet Sleeveless Golf Polo	28860
Hockey	Nike Women's Legend V-Neck T-Shirt	9525
Hockey	Nike Dri-FIT Crew Sock 6 Pack	9174
Hunting & Shooting	FNO Atlas Hammock Straps	12925.6898994446

b. Manual

Tạo NameSet [Cau 15]

Name: [Cau 15]

Expression

```
generate(
    [Dim Products].[Category Name].[Category Name].Members
    , TopCount(
        [Dim Products].[Category Name].CurrentMember * [Dim Products].[Product Name].[Product Name].Members
        , 3
        ,[Measures].[Sales]
    ))
```

No issues found

Ln: 9 Ch: 1 SPC CRLF

Additional Properties

Type: Dynamic

Display folder:

Kết quả

c. Pivot Excel

The screenshot shows a Microsoft Excel interface with a PivotTable set up. The PivotTable Fields pane on the right indicates that the 'Sales' field has been selected. The main worksheet area displays a PivotTable with data from the 'PhanCapDate' range, filtered to show results for 'Cau 15'.

	Category	Product	Sales
1	Row Labels		Sales
2	Accessories		
3	Team Golf San Francisco Giants Putter Grip	10720.70992	
4	Team Golf Texas Longhorns Putter Grip	10120.94993	
5	Team Golf St. Louis Cardinals Putter Grip	9371.249924	
6	Baseball & Softball		
7	adidas Men's F10 Messi TRX FG Soccer Cleat	24355.9407	
8	adidas Kids F5 Messi FG Soccer Cleat	10536.96023	
9	Boxing & MMA		
10	Under Armour Women's Micro G Skulpt Running S	18030.1604	
11	Camping & Hiking		
12	Diamondback Women's Serene Classic Comfort Bi	1480101.374	
13	Cardio Equipment		
14	Nike Men's Free 5.0+ Running Shoe	1357464.238	
15	Nike Women's Tempo Shorts	12180	
16	Cleats		
17	Perfect Fitness Perfect Rip Deck	1605032.495	
18	Electronics		
19	Titleist Pro V1x High Numbers Personalized Go	20951.97063	
20	Titleist Pro V1 High Numbers Personalized Gol	19236.3006	
21	Under Armour Men's Compression EV SL Slide	18535.88035	
22	Fishing		
23	Field & Stream Sportsman 16 Gun Fire Safe	2496275.249	
24	Fitness Accessories		
25	Under Armour Hustle Storm Medium Duffle Bag	13821.05032	
26	Girls' Apparel		
27	adidas Youth Germany Black/Red Away Match Soc	30450	
28	TYR Boys' Team Digi Jammer	17755.56031	
29	Nike Men's Deutschland Weltmeister Winners Bl	10740	
30	Golf Apparel		
31	Top Flite Women's 2014 XL Hybrid	7576.209927	
32	Golf Balls		
33	Glove It Women's Imperial Golf Glove	8995.499887	
34	Hirzl Women's Softflex Golf Glove	7681.729887	
35	Glove It Women's Mod Oval Golf Glove	6896.549915	
36	Golf Gloves		
37	Clicgear Rovic Cooter Bag	14796.3002	
38		20000.000000	

Chương 4: QUÁ TRÌNH LẬP BÁO BIỂU (SSRS)

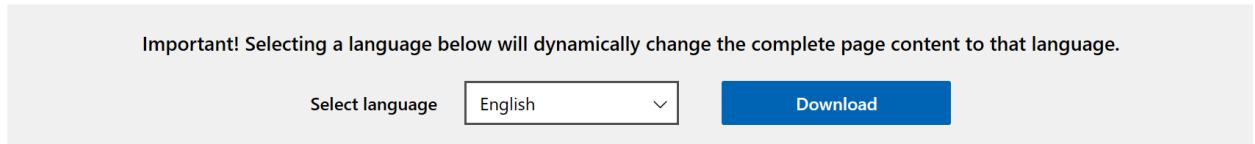
4.1. REPORT VỚI VISUAL STUDIO

4.1.1. Cài đặt Microsoft Sql Server 2019 Reporting Services

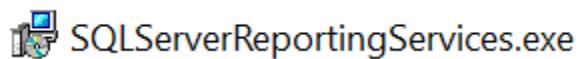
- Download công cụ Microsoft SQL Server 2019 Reporting Services

Microsoft SQL Server 2019 Reporting Services

SQL Server Reporting Services is a server-based reporting platform that provides comprehensive reporting functionality.



- Cài đặt công cụ Microsoft SQL Server 2022 Reporting Services
- Sau khi tải về ta có được một file exe như hình



- Bước 1: Chọn Install ta được giao diện cài đặt như sau



- Bước 2: Chuyển qua màn hình Choose an edition to install, thực hiện chọn free edition và tiếp tục chọn Developer sau đó nhấn Next



- Bước 3: Tại giao diện Review the license terms, thực hiện chọn I accept the license terms và chọn Next



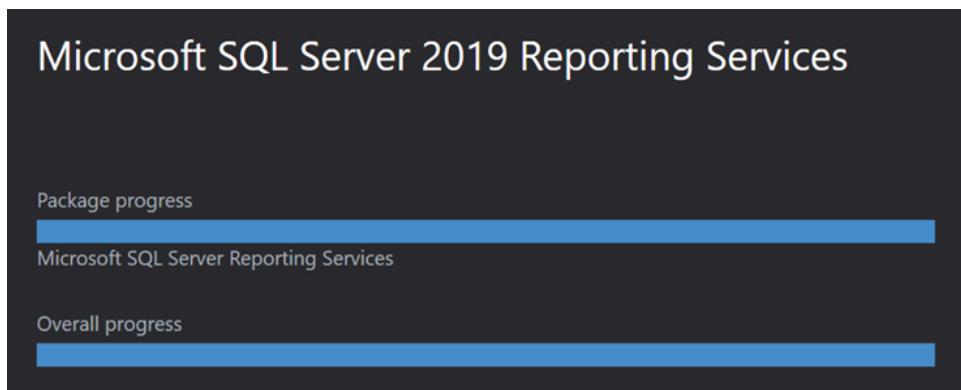
- Bước 4: Tiếp tục chọn Next



- Bước 5: Tiếp tục chọn Next



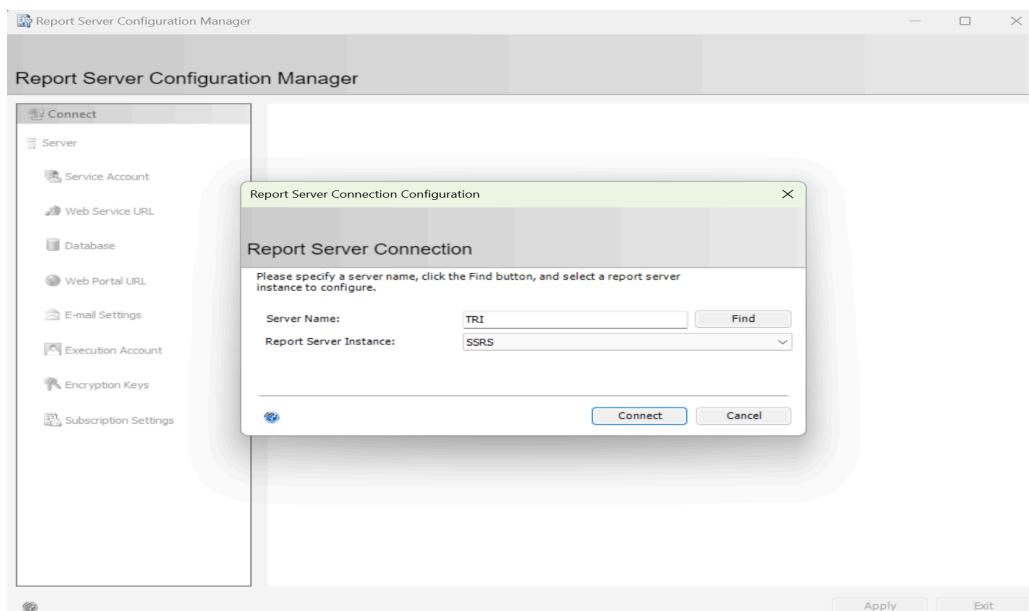
- Bước 6: Tiếp tục chọn Install và tiến hành cài đặt công cụ



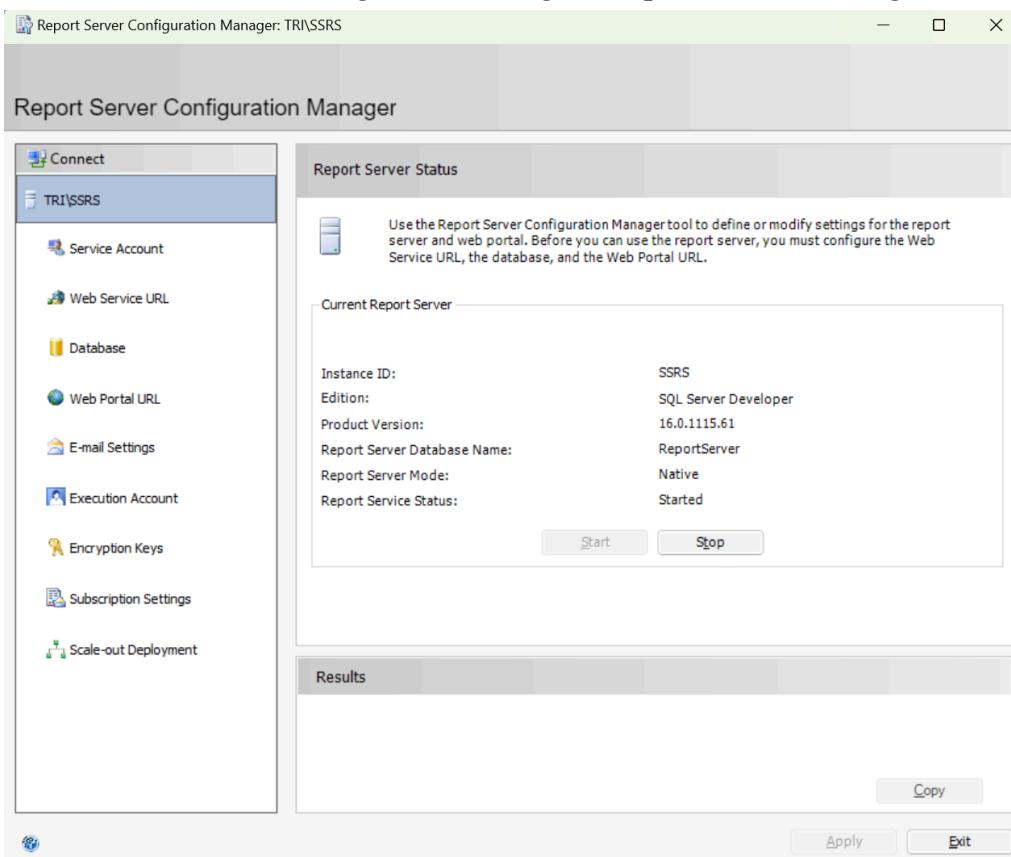
- Bước 7: Sau khi cài đặt thành công, chọn Configure Report Server để mở giao diện Setup Completed



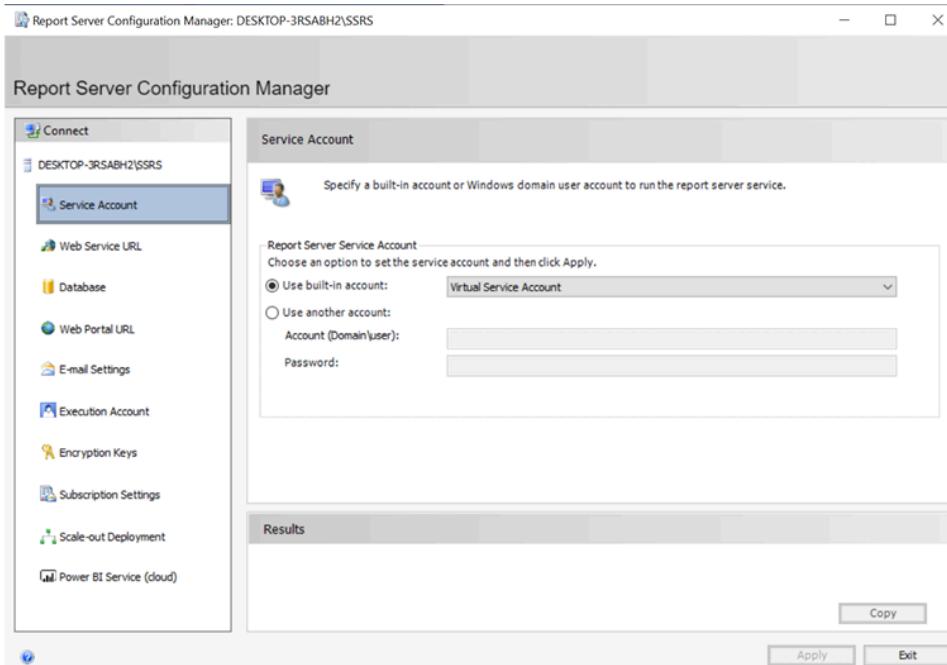
- Bước 8: Thực hiện cấu hình Report Server Configuration Manager.
 - 8.1: Chọn Configure report server hoặc mở công cụ Report Server Configuration Connection trong thanh tìm kiếm



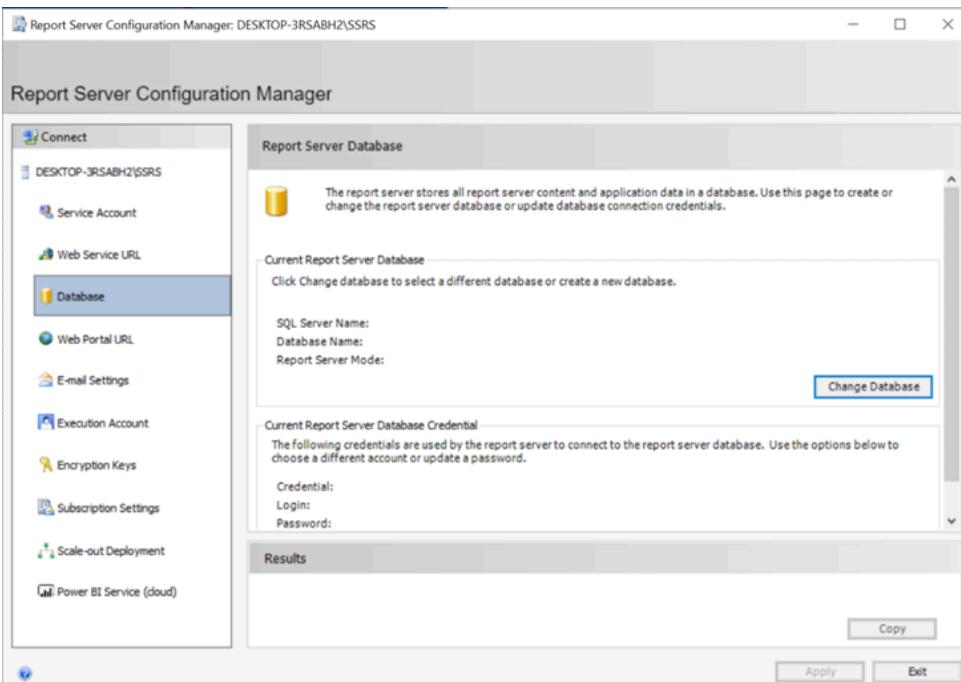
- 8.2: Giao diện tổng thể của công cụ Report Server Configuration Manager.



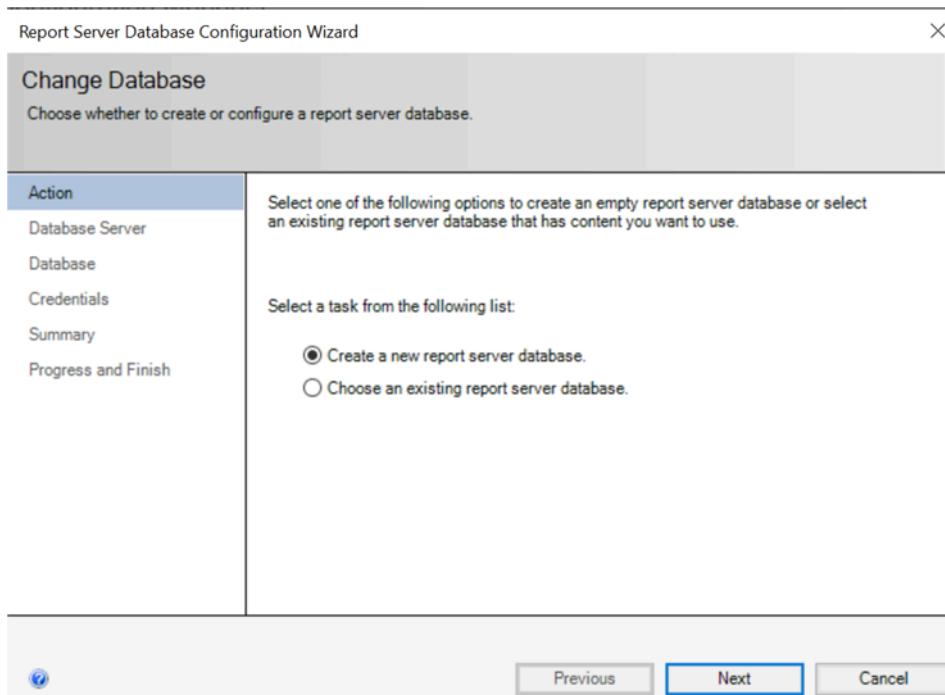
- 8.3: Tại tab Service Account, chuyển đổi giá trị trong Use a built - in account và chọn Apply



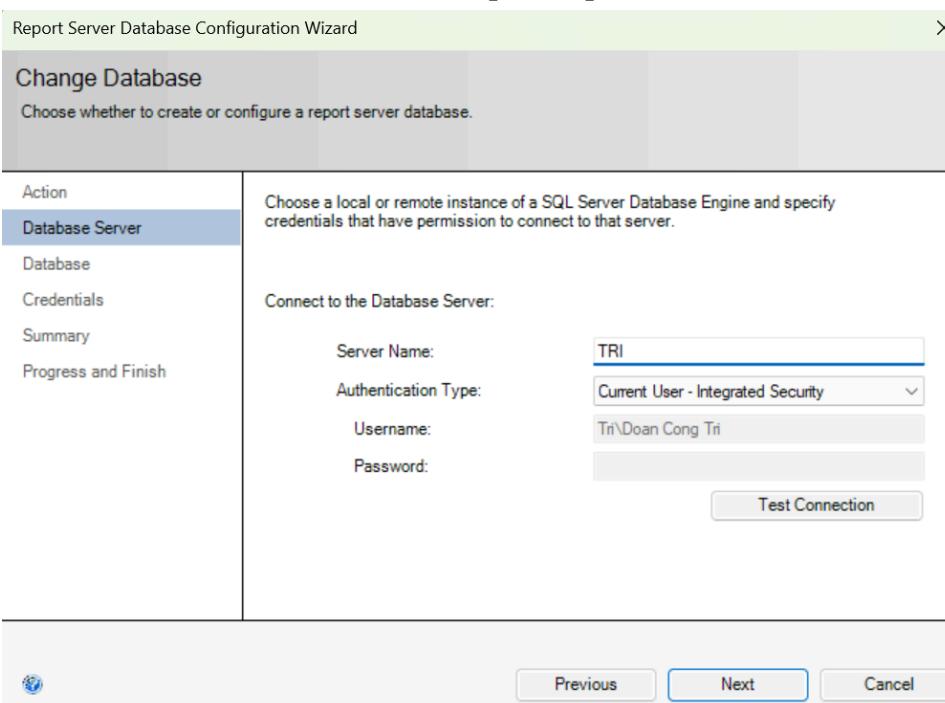
- 8.4: Tại tab Database, tạo mới một dataset bằng cách chọn Change Database



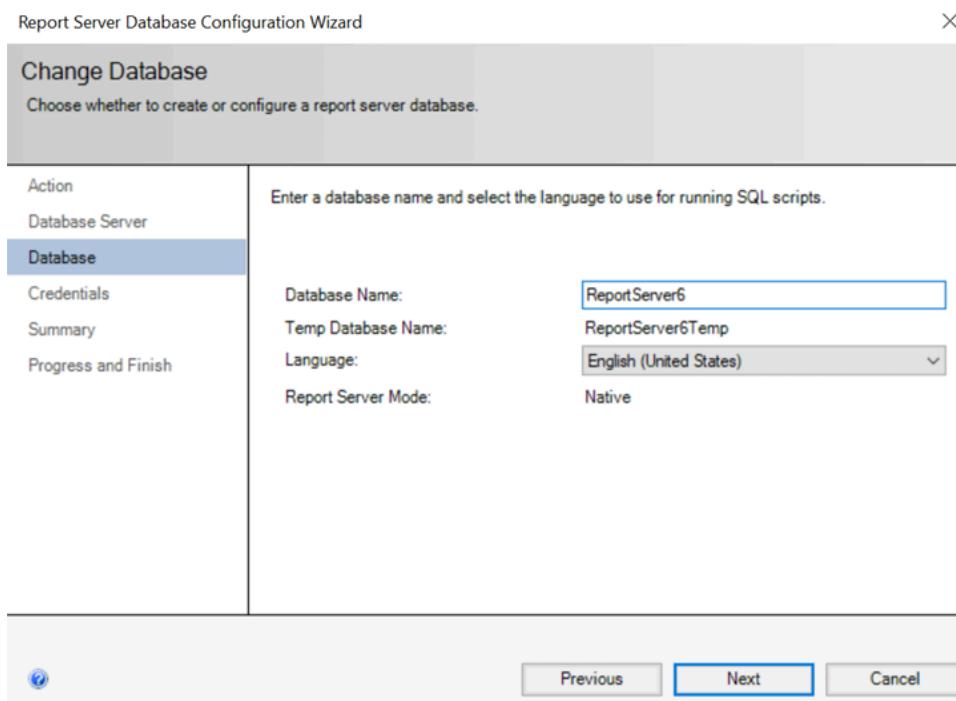
- 8.5: Chọn Create a new report server database và nhấn Next.



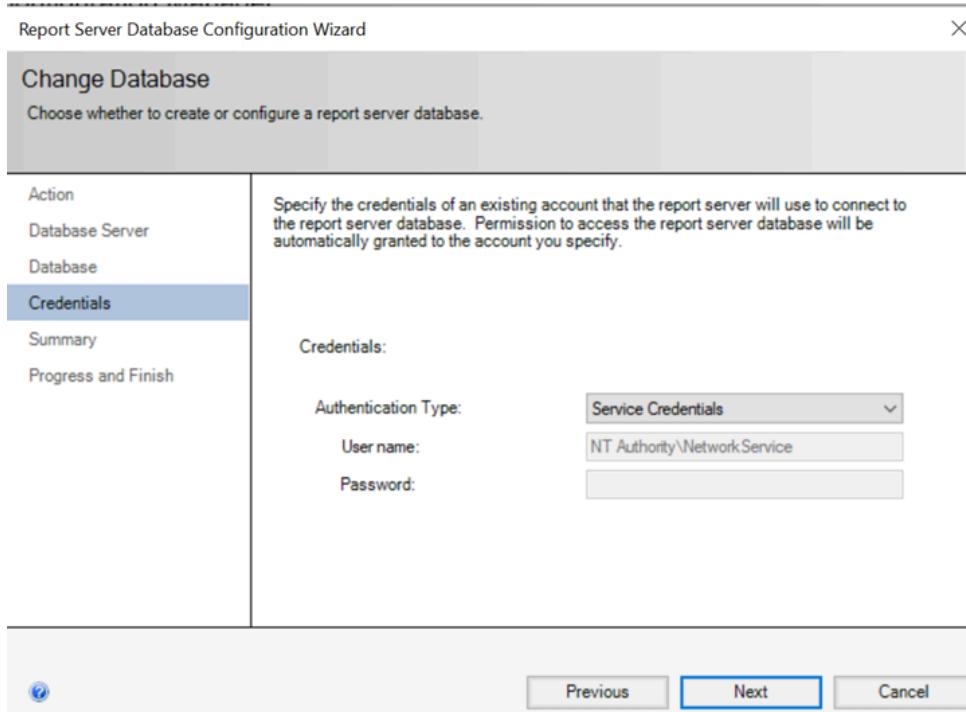
- 8.6: Điền vào Server Name phù hợp và nhấn Next.



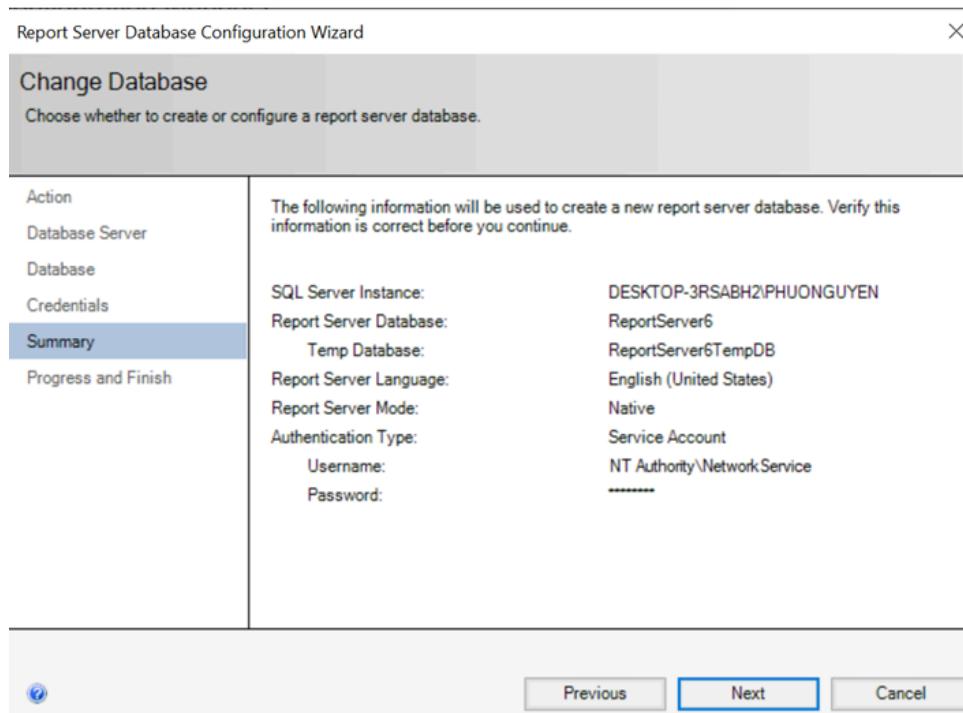
- 8.7: Đặt tên cho Database và nhấn Next.



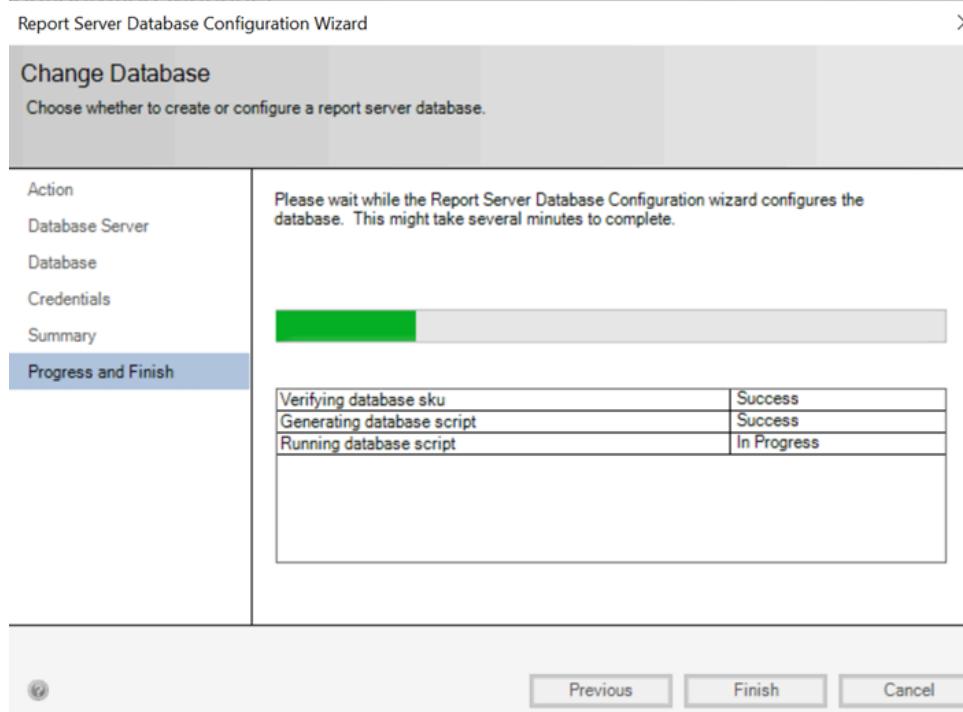
- 8.8: Trong tab Credentials tiếp tục nhấn Next.



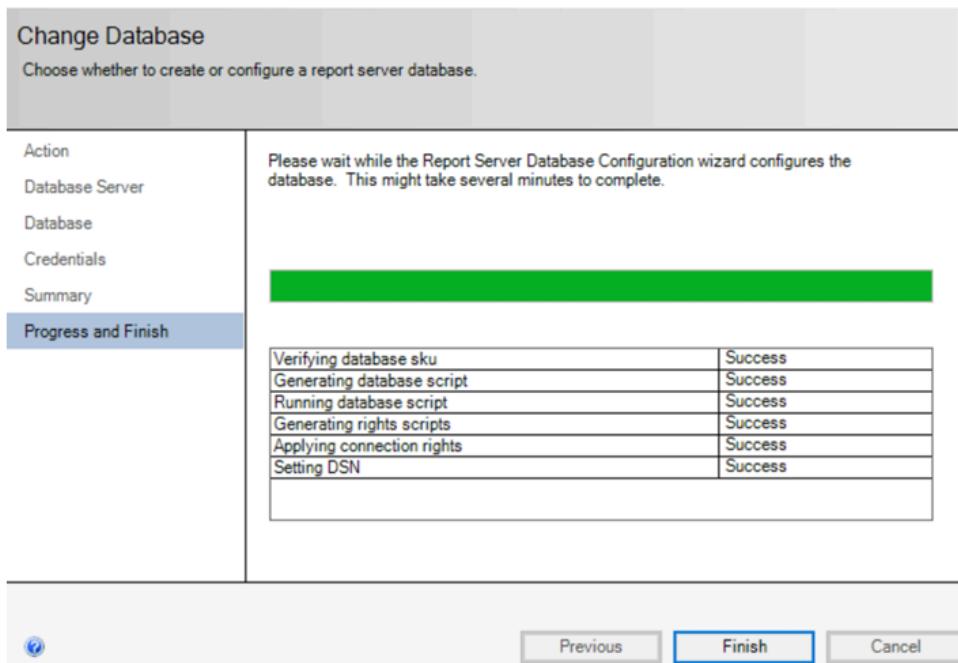
- 8.9: Trong tab Summary, tiếp tục nhấn Next.



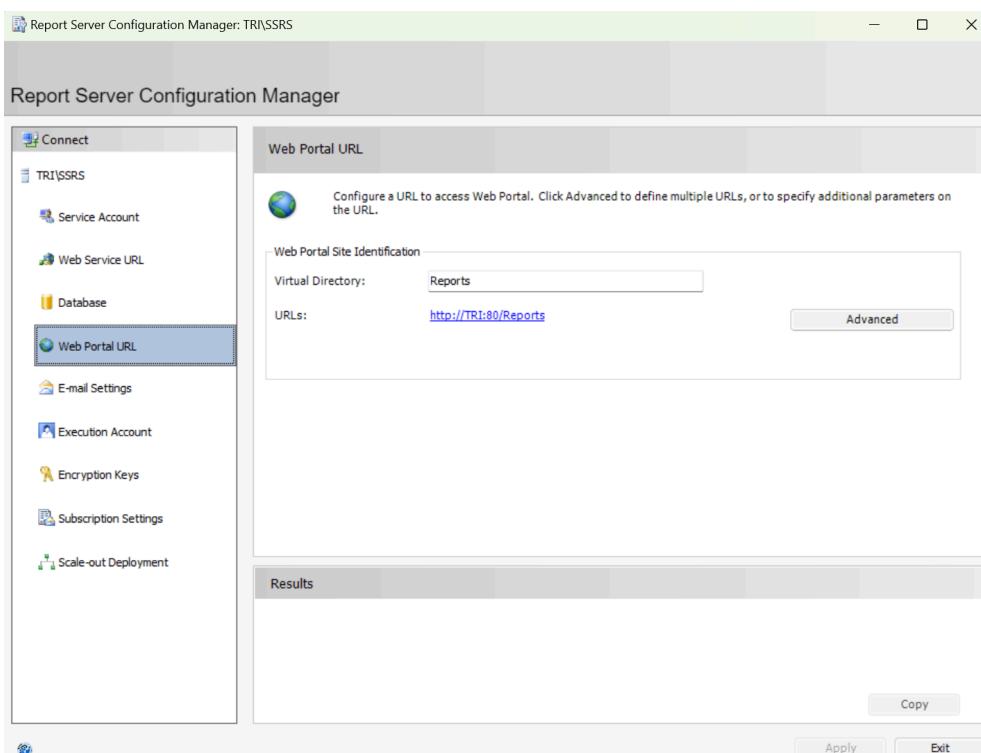
● 8.10: Hệ thống tiến hành tạo Database.



- 8.11: Tạo database thành công.



- 8.12: Tại giao diện tab Web Portal URL, chọn Apply để xây dựng trang web.

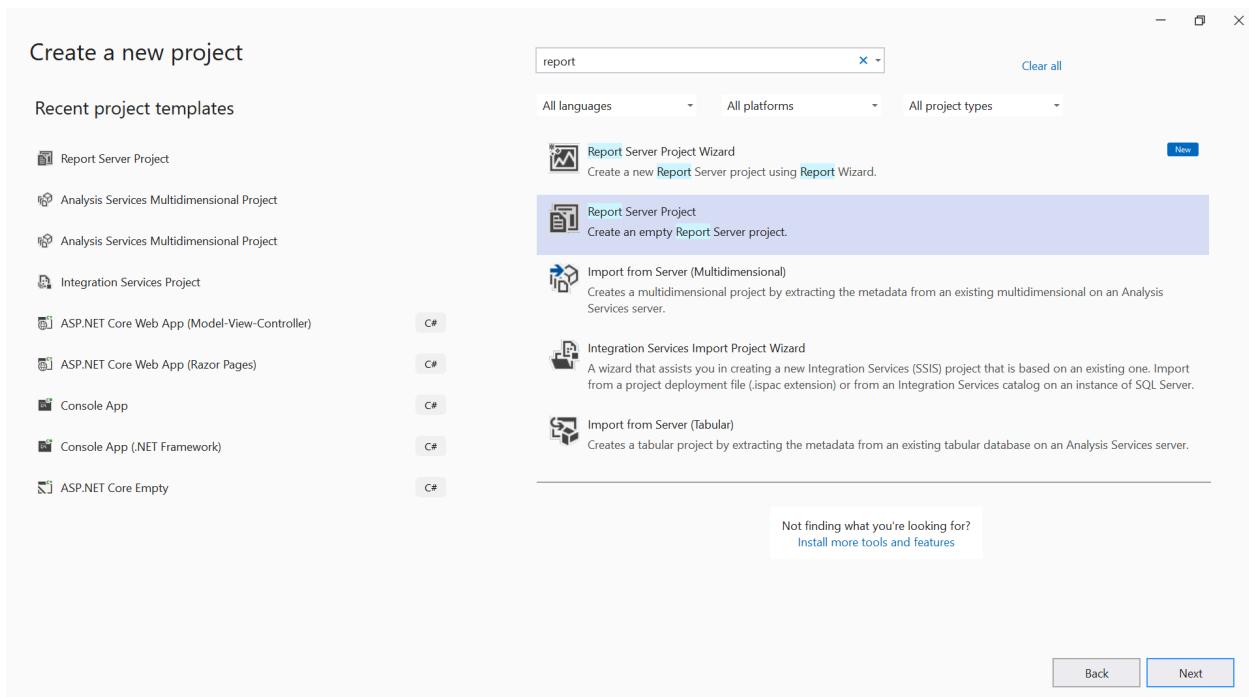


4.1.2 Quá trình tạo báo biểu bằng công cụ Report builder

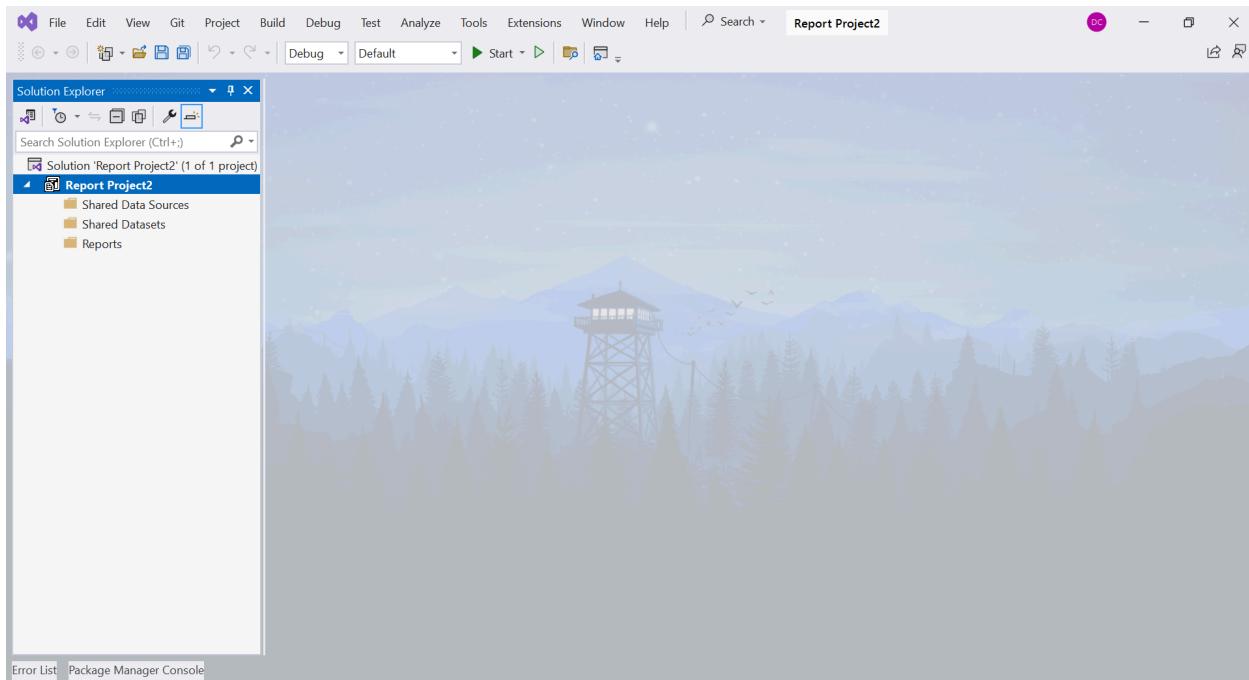
- Thực hiện tải và cài đặt



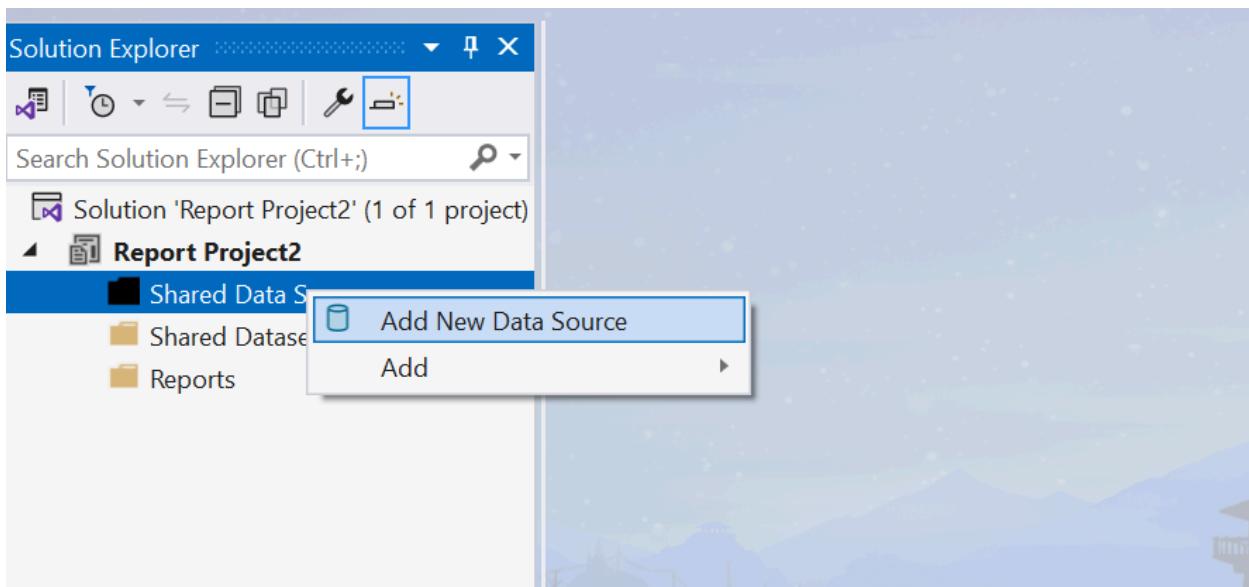
- Tạo dự án Report Server project



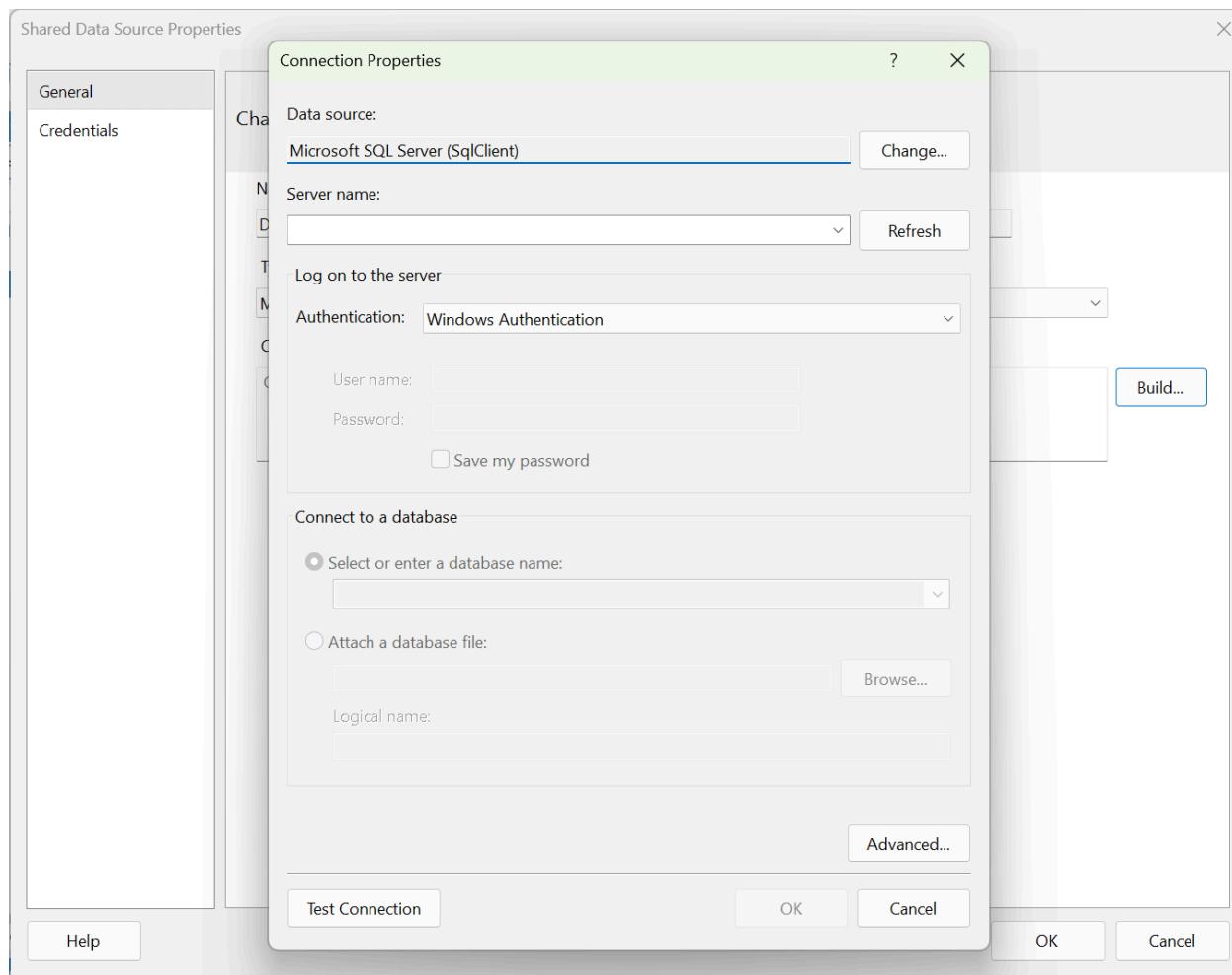
- Giao diện của dự án



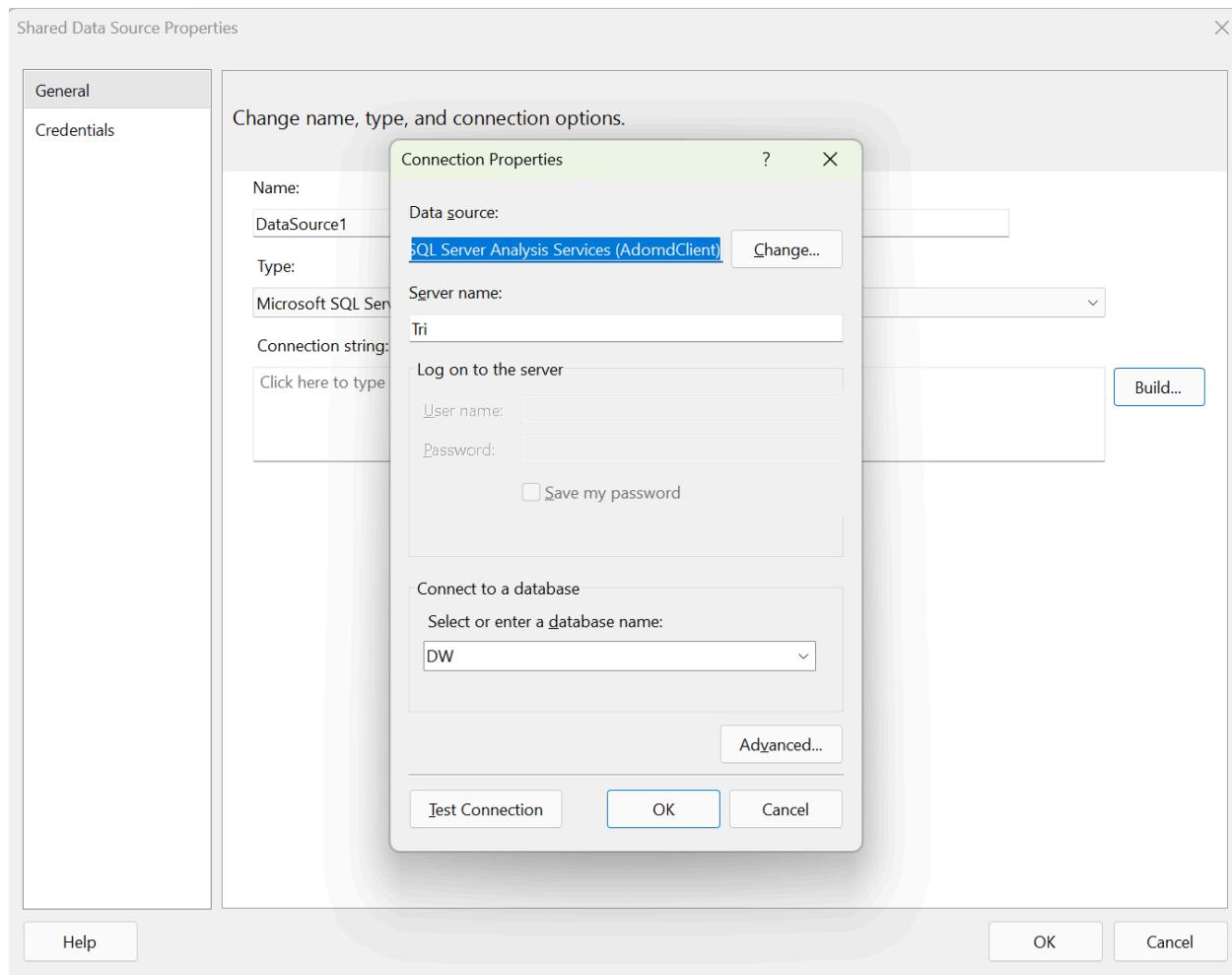
- Thực hiện thêm một Shared Data Sources vào dự án



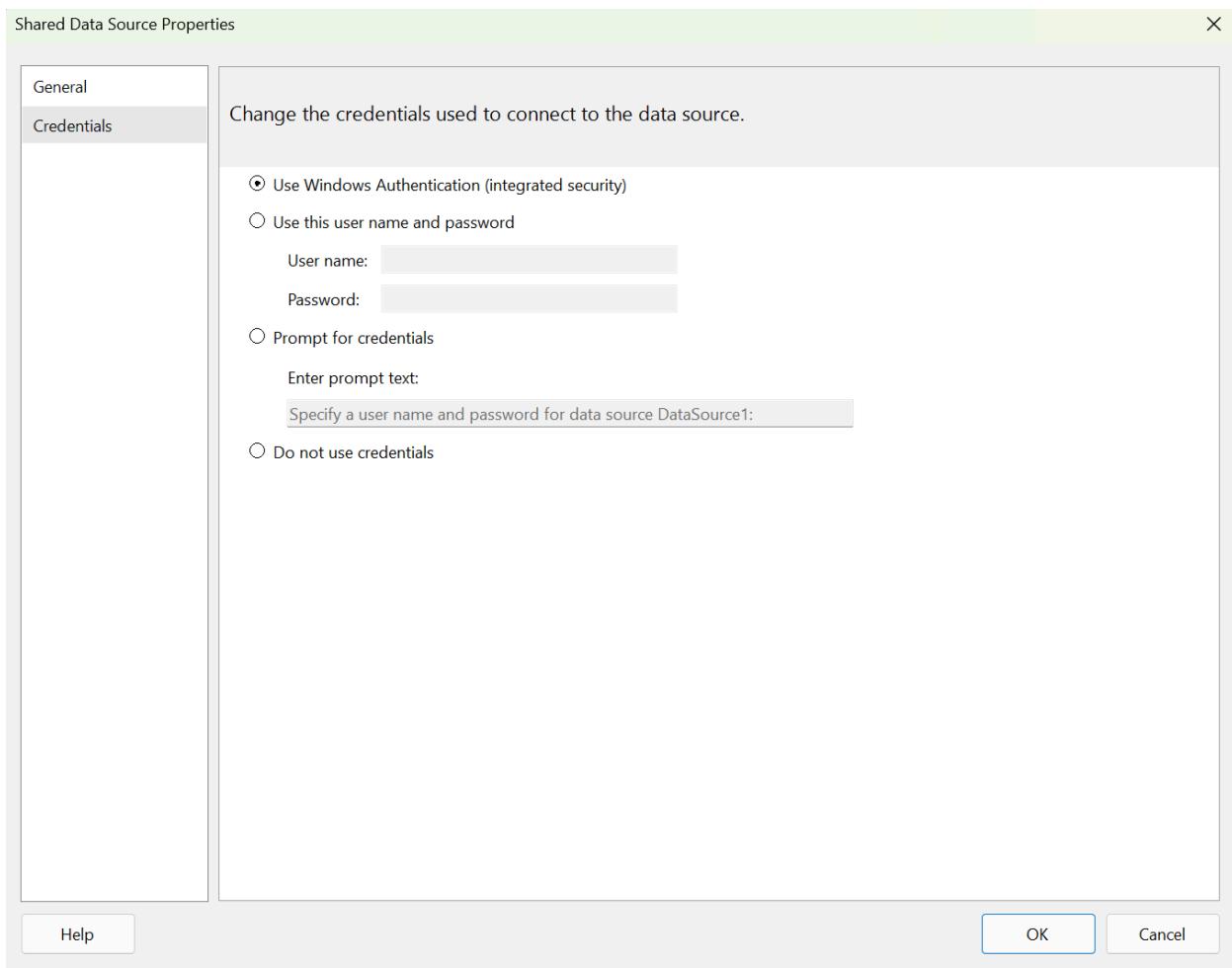
- Sau đó chọn vào Build



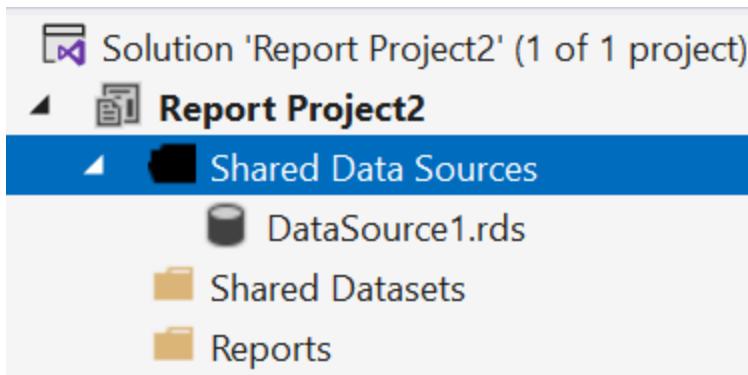
- Thực hiện chọn Data source là **Microsoft SQL Server Analysis Services (AdomdClient)**, sau đó điền tên server name và chọn khôi dữ liệu



- Chuyển qua tab Credential và chọn **Use Windows Authentication**

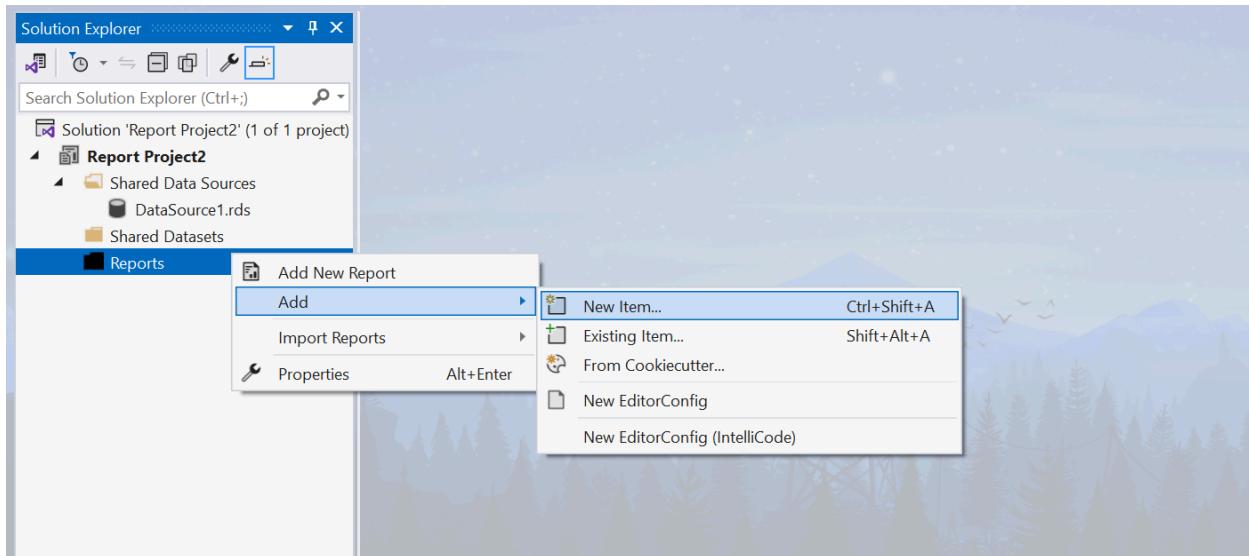


- Sau đó chọn OK và kiểm tra

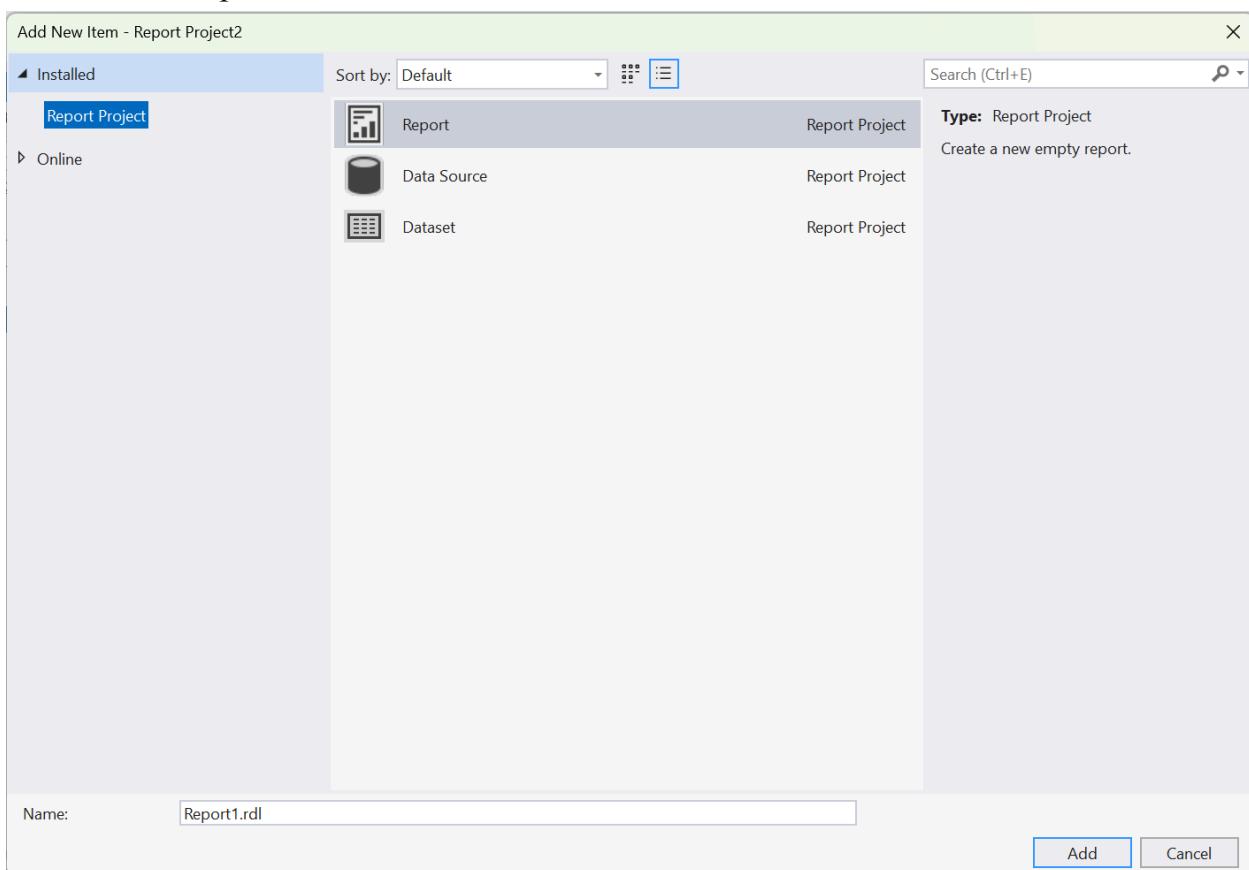


4.1.2.1 Report - Câu truy vấn 2

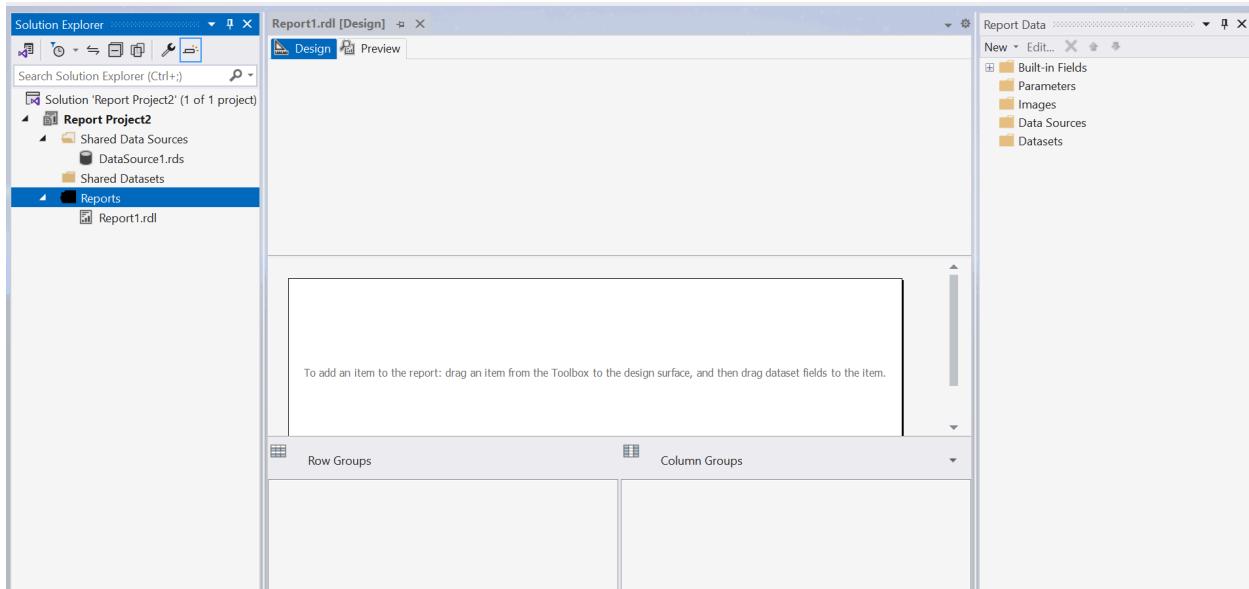
- Thực hiện tạo mới một Report trong dự án, chọn Add và chọn New Item



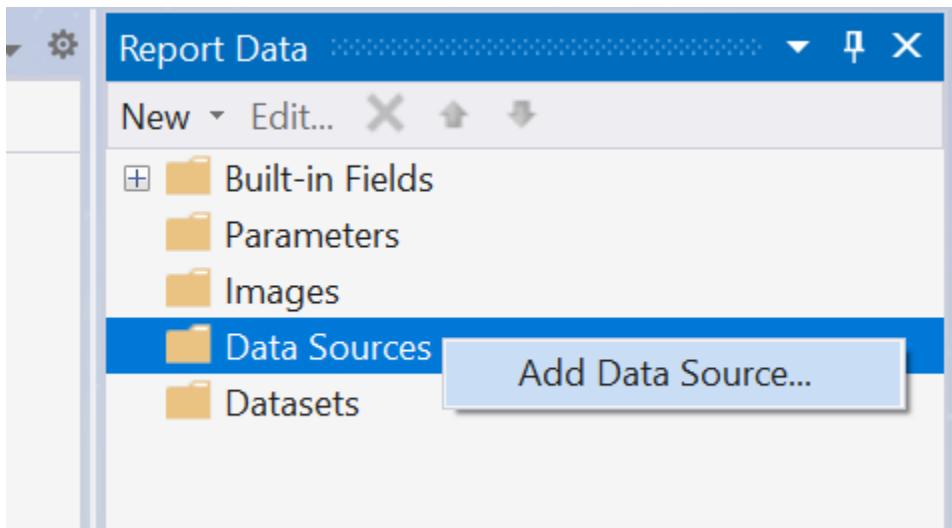
- Chọn Report, đổi tên và chọn Add



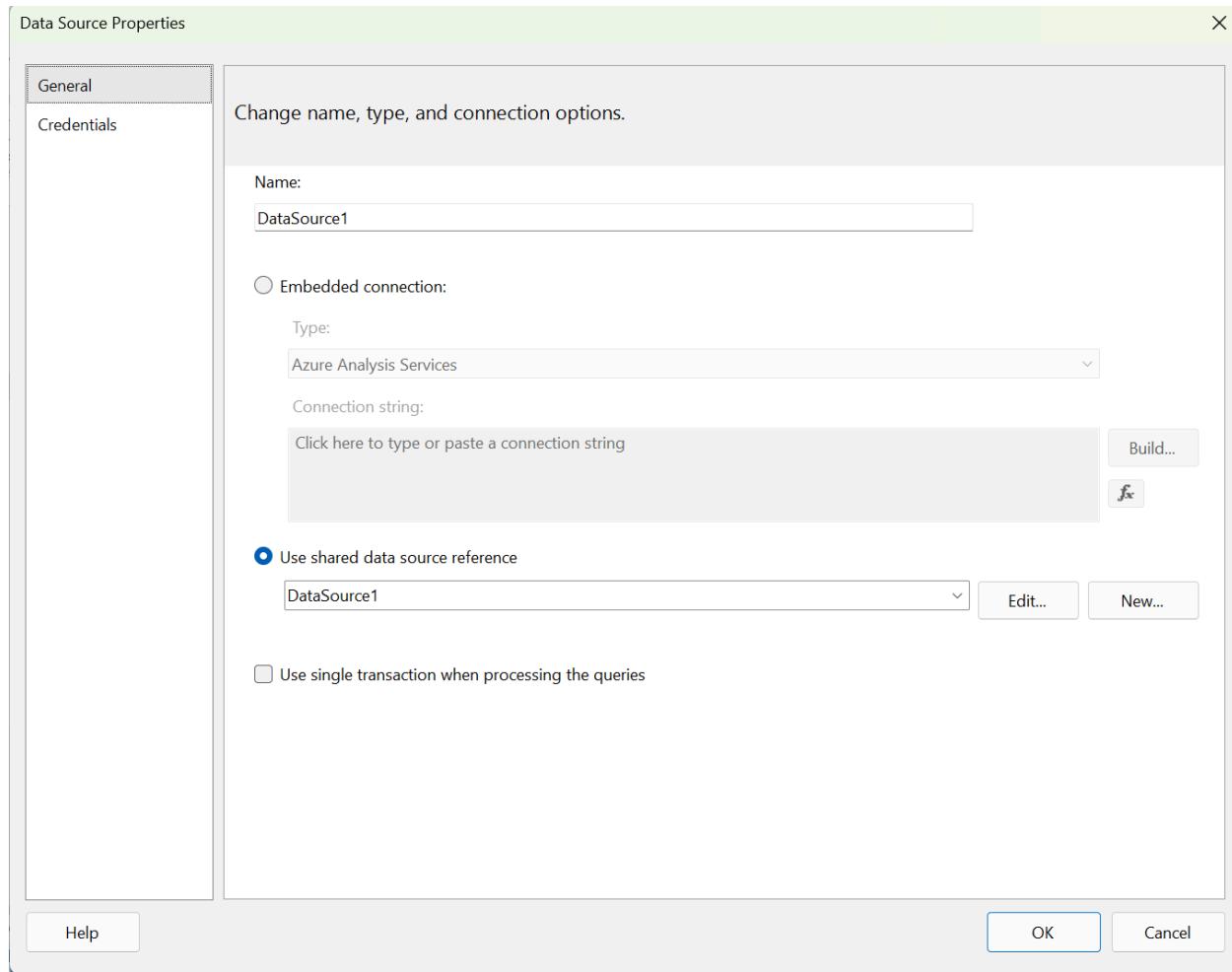
- Giao diện Report đã được tạo



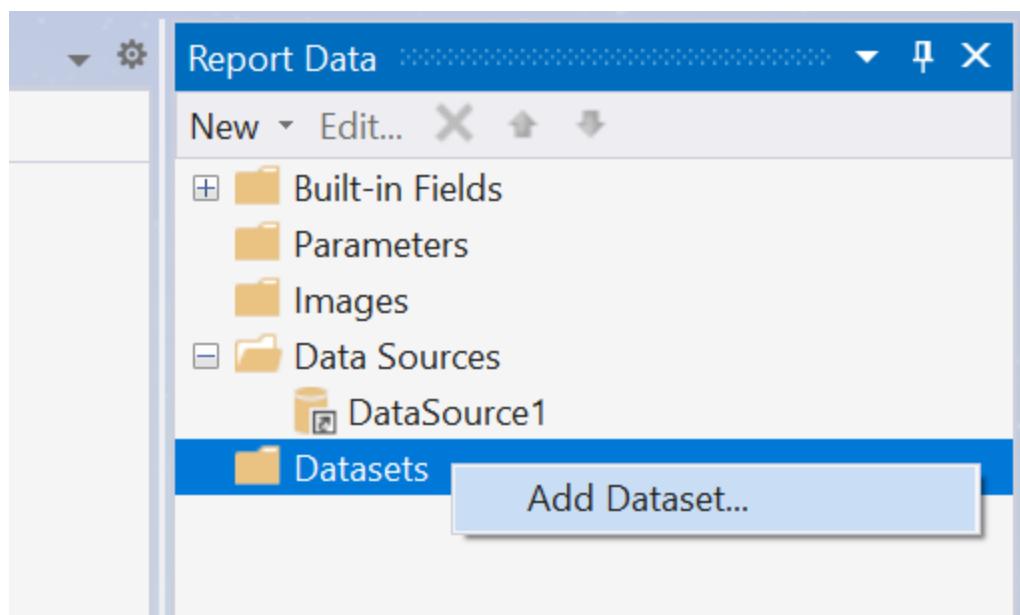
- Thực hiện thêm một Data Sources



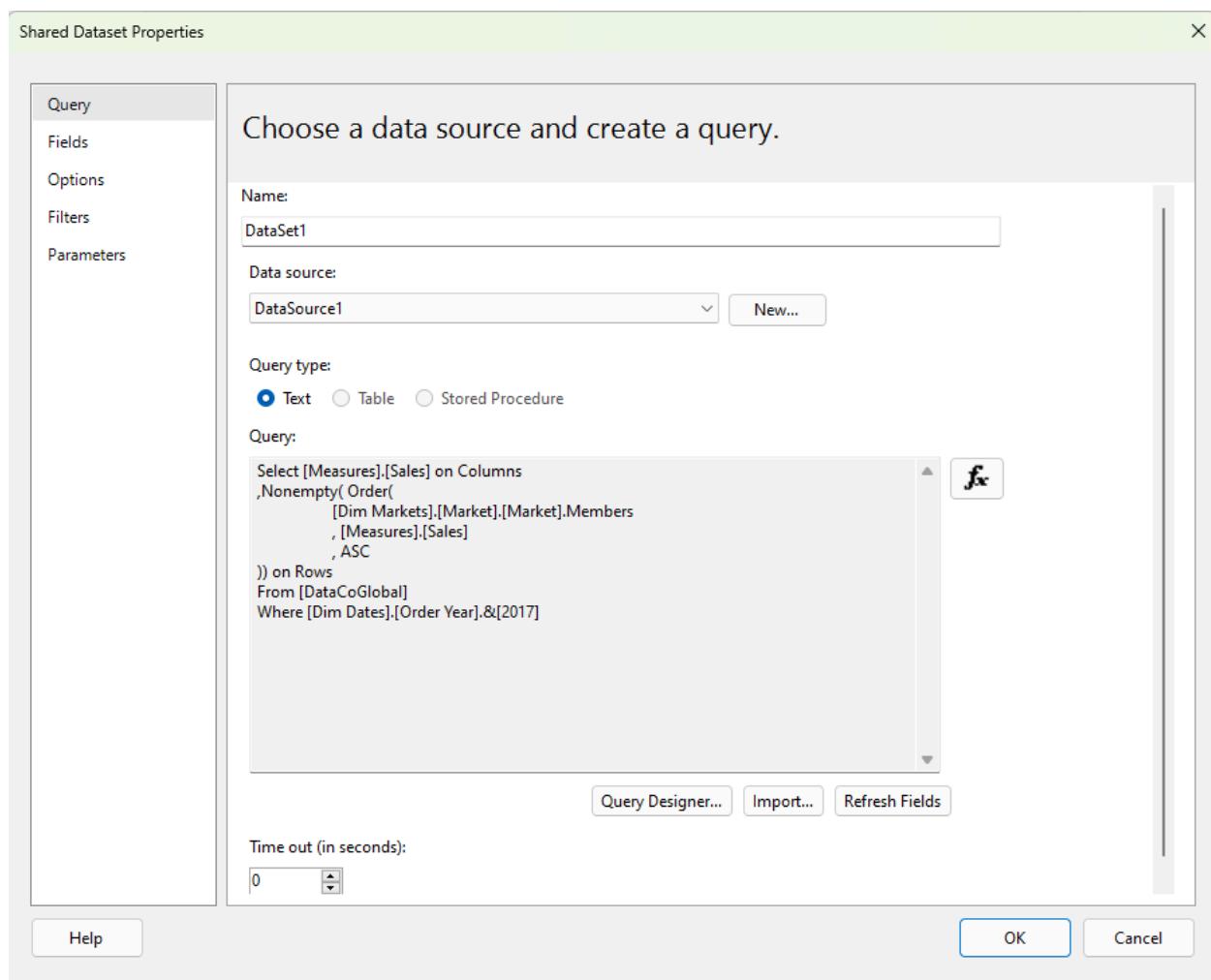
- Chọn Data Sources đã thêm khi nãy và tiến hành chọn Ok



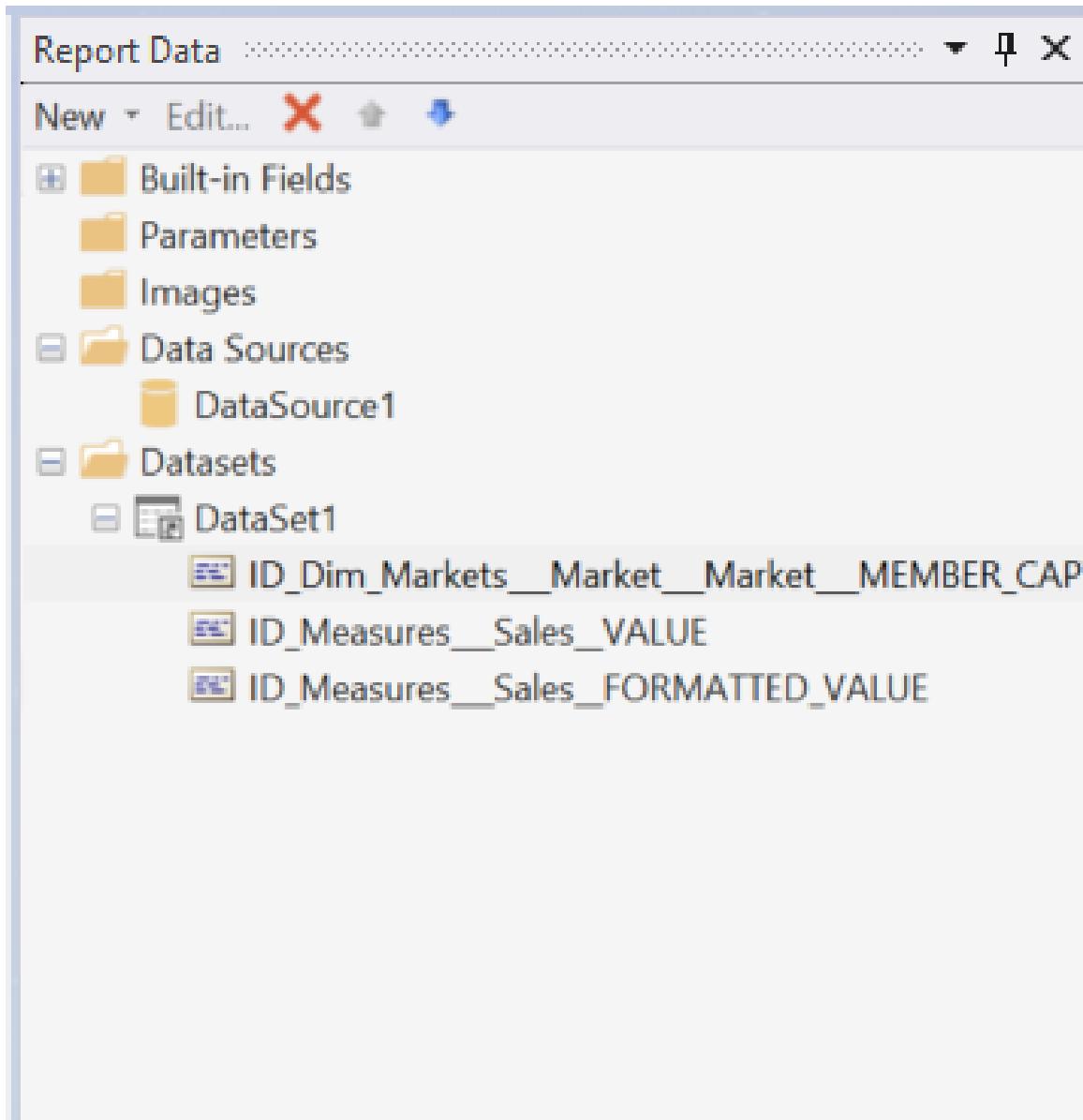
- Thêm Datasets



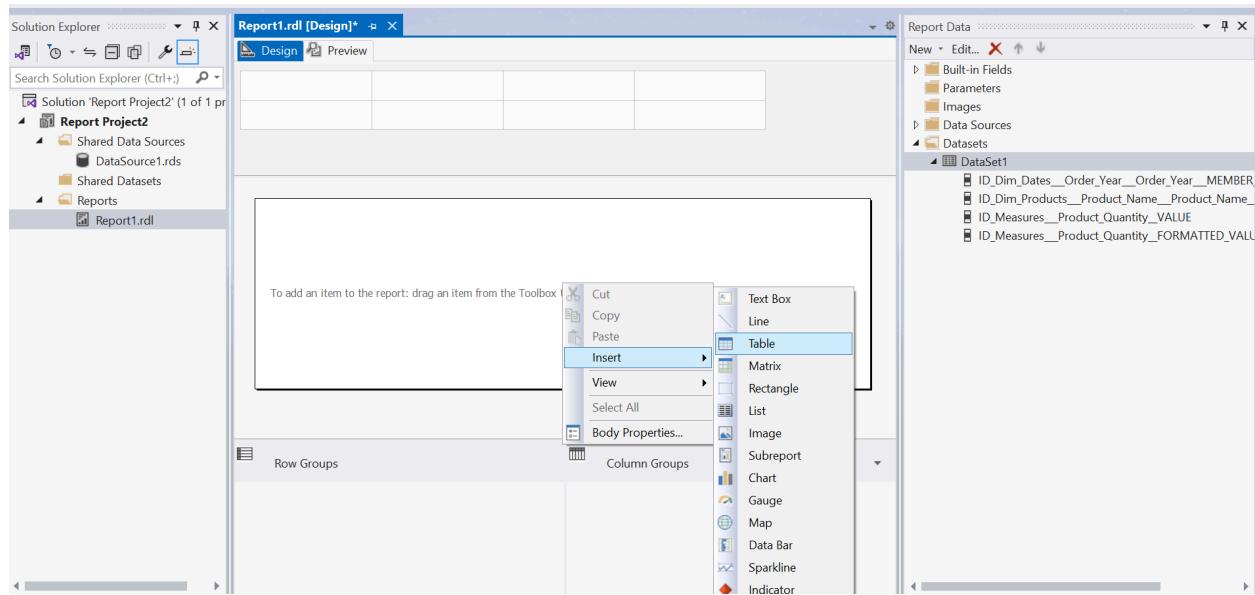
- Thực hiện dán câu truy vấn và chọn Data source



- Dataset đã được thêm vào



- Tiến hành tạo báo cáo theo ý muốn
- Đầu tiên ta sử dụng một bảng để mô tả báo cáo, tiến hành chọn Insert Table



- Đổi tên cột và kéo các trường thông tin cần thiết vào bảng

Markets	Sales
[ID_Dim_Markets__Market__Market_	[ID_Measures__Sales__FORMATTED.

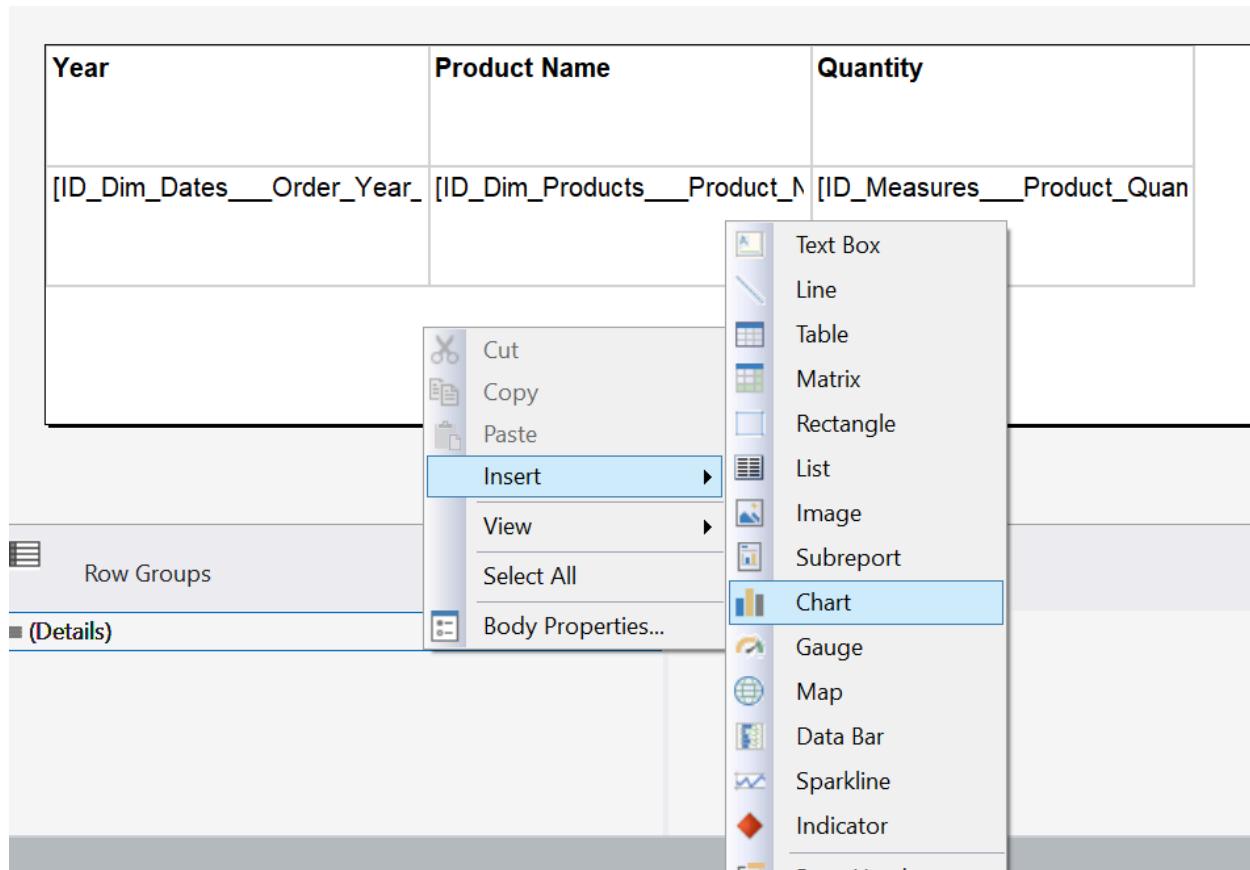
- Thực hiện xem Review

Markets	Sales
USCA	23063.1904983521
Africa	247830.164852142
Pacific Asia	1133058.90028477
Europe	5196492.17457485
LATAM	5205424.37261772

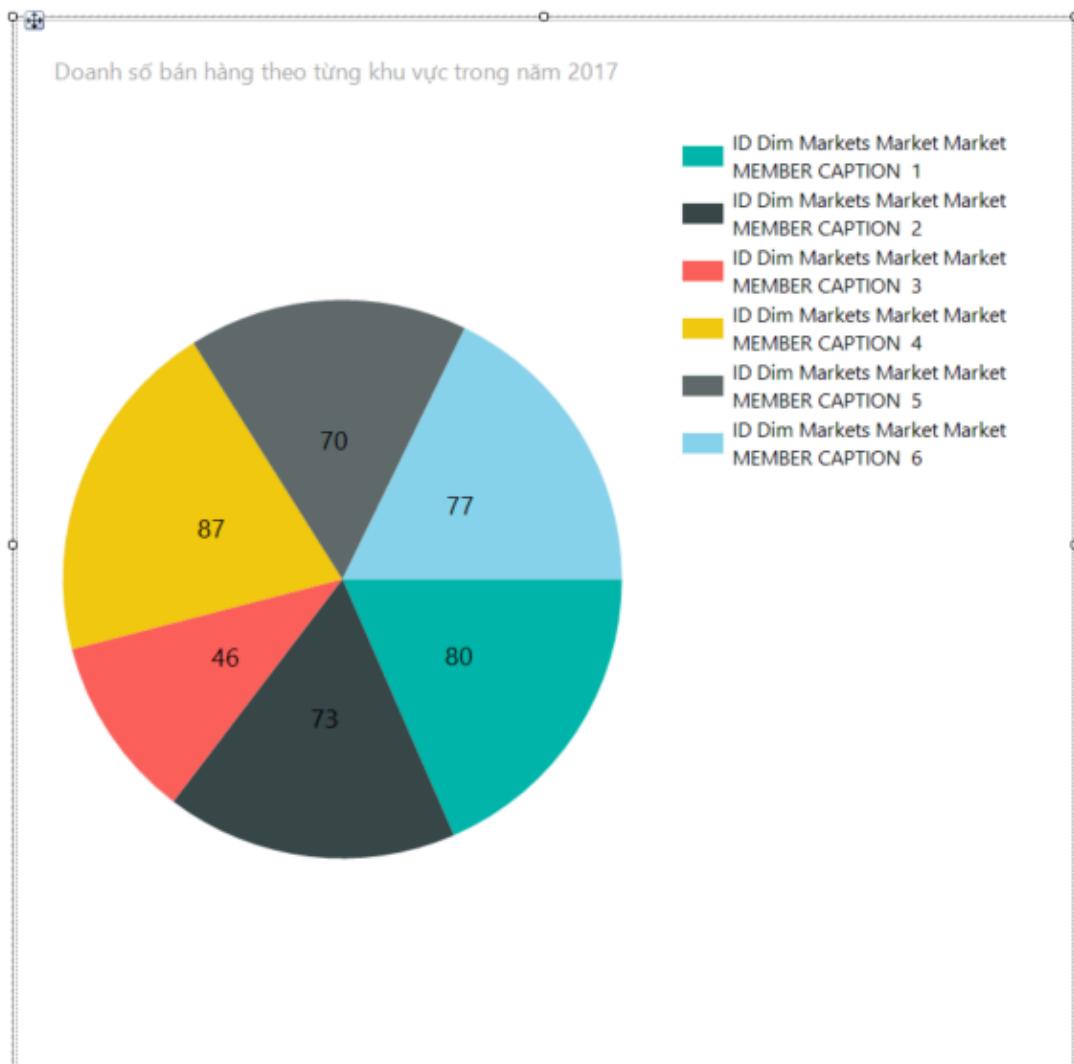
- Thực hiện tùy chỉnh trang trí table tùy ý

Markets	Sales
USCA	23063.1904983521
Africa	247830.164852142
Pacific Asia	1133058.90028477
Europe	5196492.17457485
LATAM	5205424.37261772

- Tiếp theo thực hiện Insert chart để biểu diễn thông tin trực quan

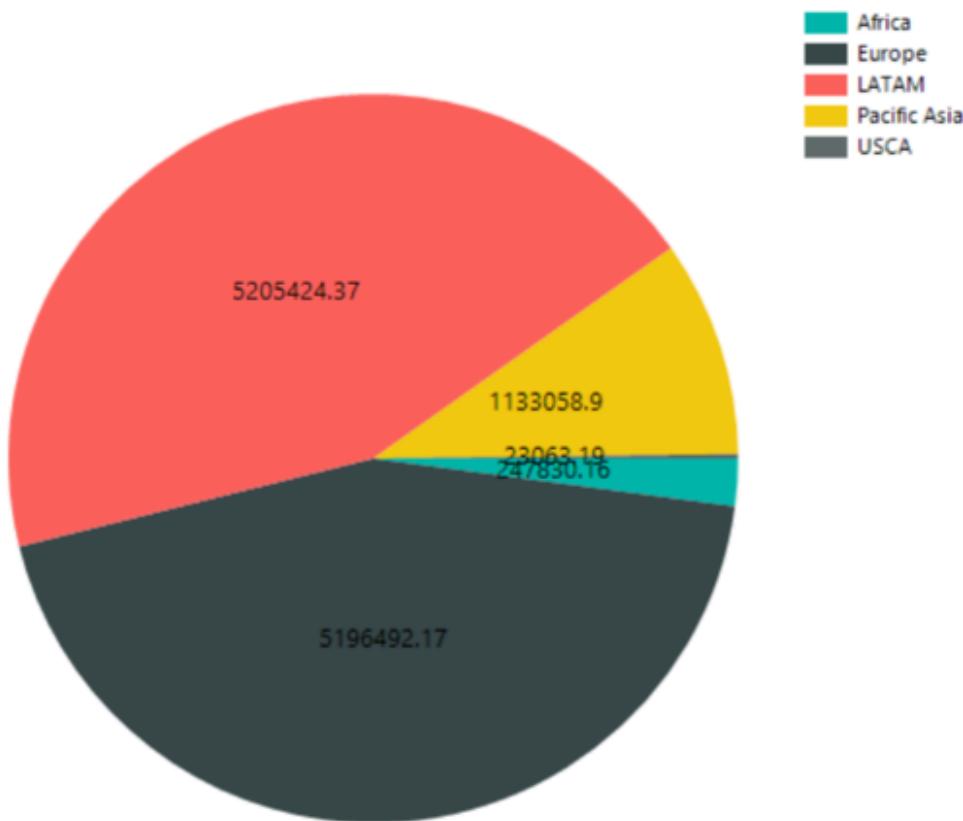


- Thực hiện thêm vào các thông tin của biểu đồ sao cho phù hợp

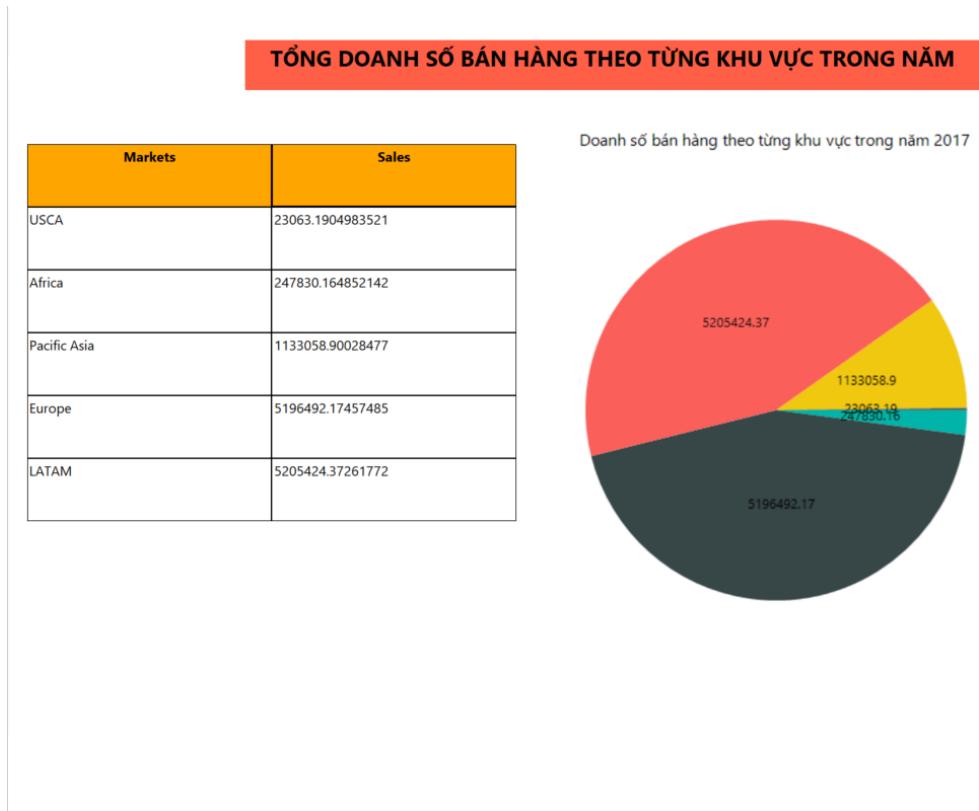


- Tiến hành xem review biểu đồ

Doanh số bán hàng theo từng khu vực trong năm 2017



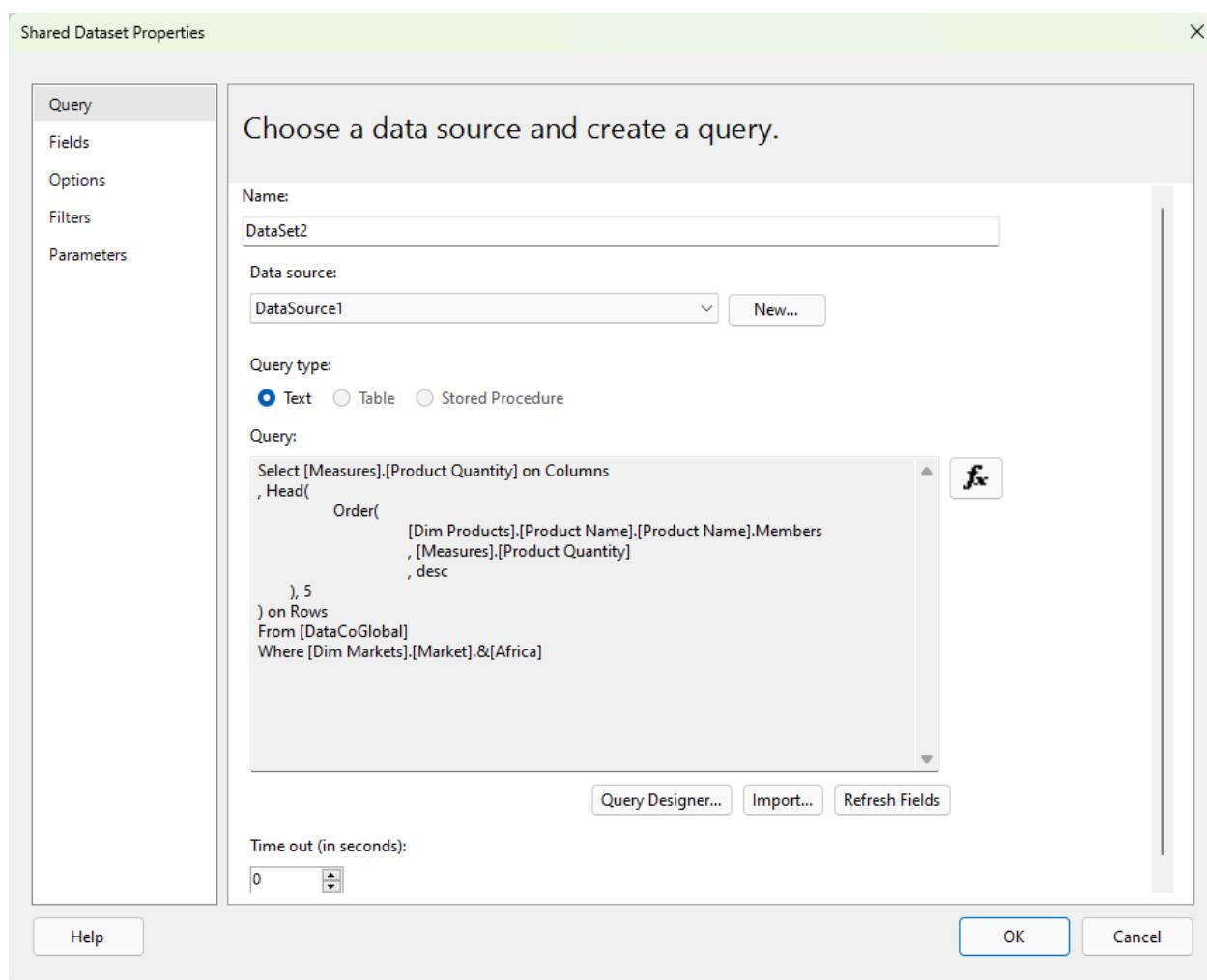
- Báo cáo sau khi hoàn thành



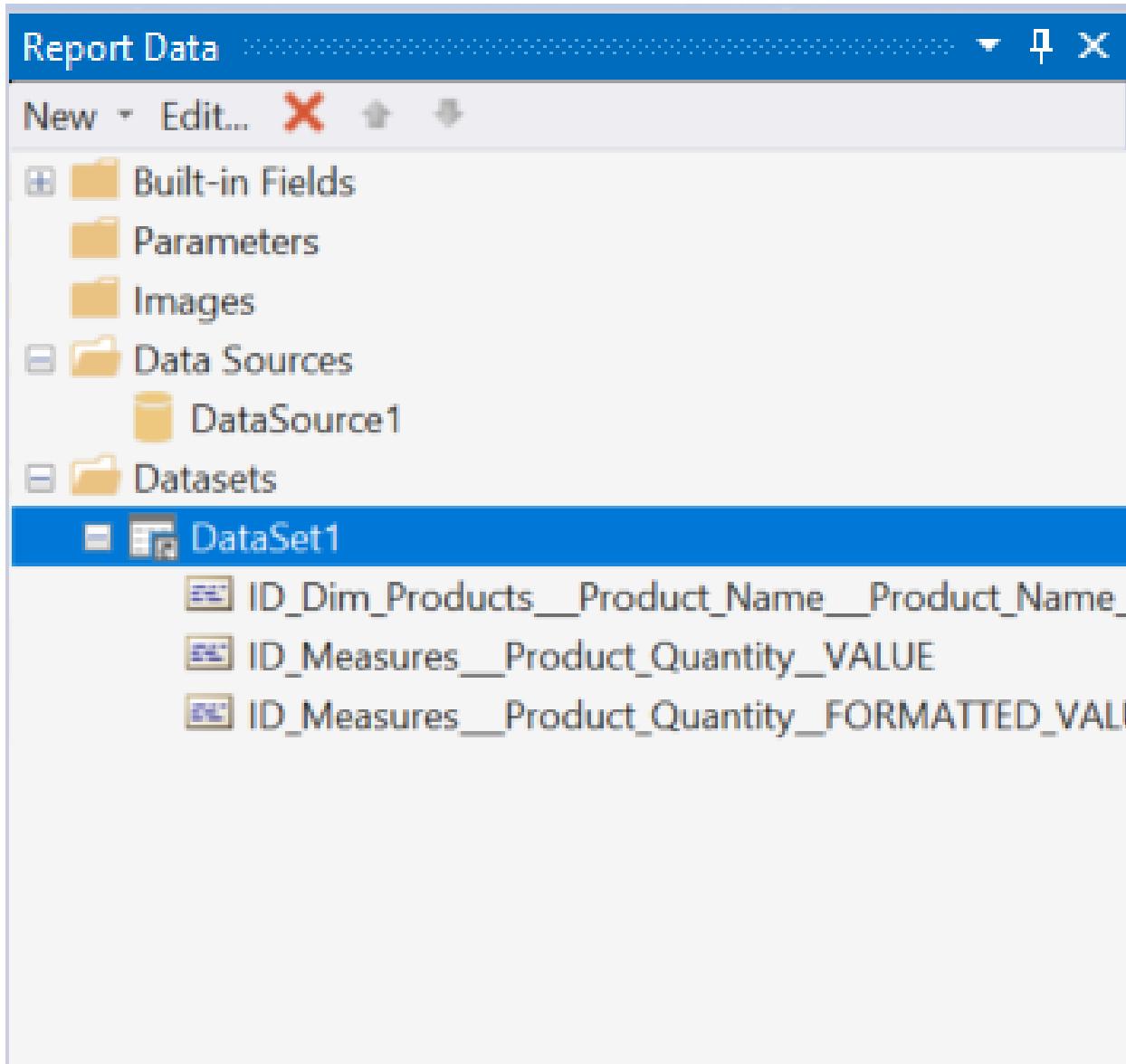
=> Doanh số bán hàng của khu vực LATAM chiếm tỉ trọng cao nhất, gần 50% tổng doanh thu với 5,2M \$, theo sau đó là khu vực Pacific Asia với doanh số tương đương 5,19M \$, và các khu vực còn lại chiếm doanh thu tương đối thấp cho thấy thị trường kinh doanh chưa phát triển mạnh ở khu vực này, cần áp dụng những chính sách phù hợp để kích thích nhu cầu của khách hàng.

4.1.2.2 Report - Câu truy vấn 6

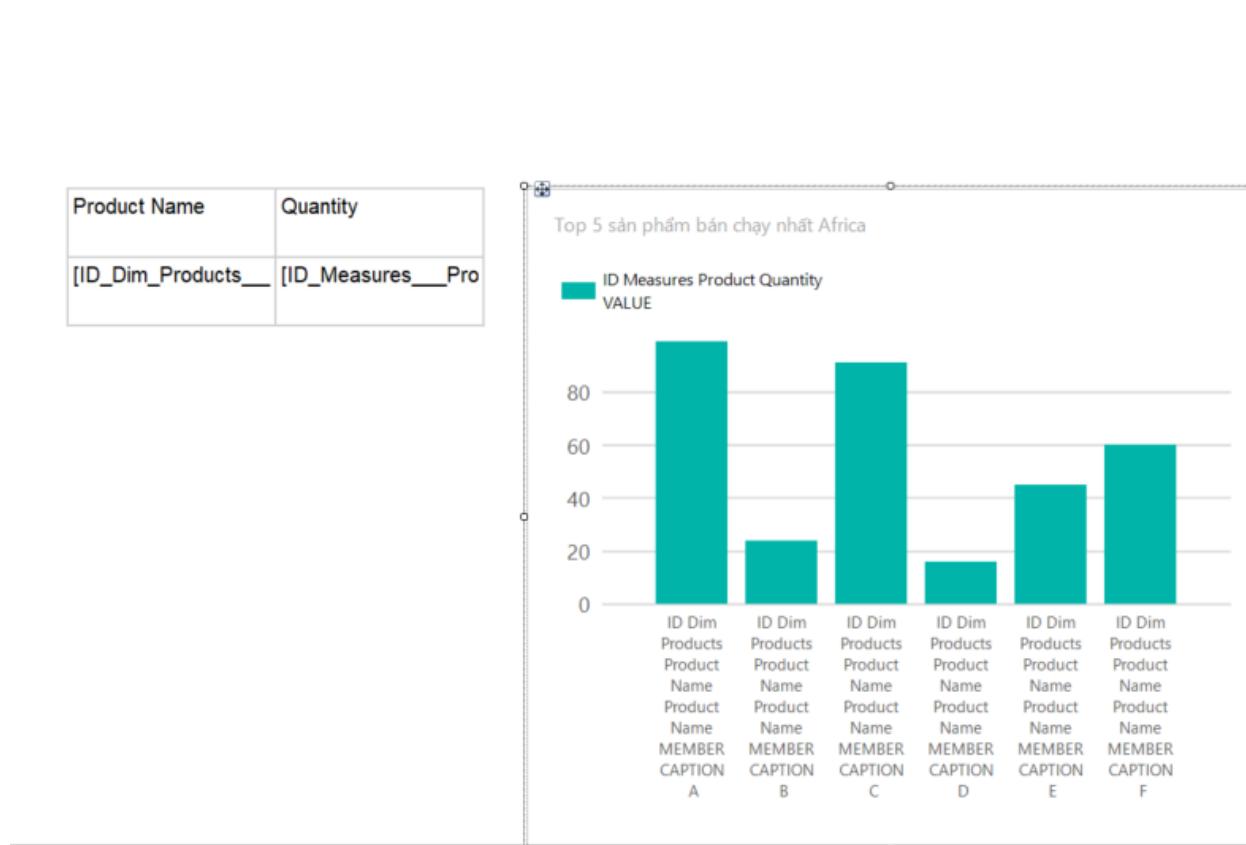
- Tương tự ta thực hiện thêm Dataset như Report - Câu truy vấn 2



- Tiến hành kiểm tra lại

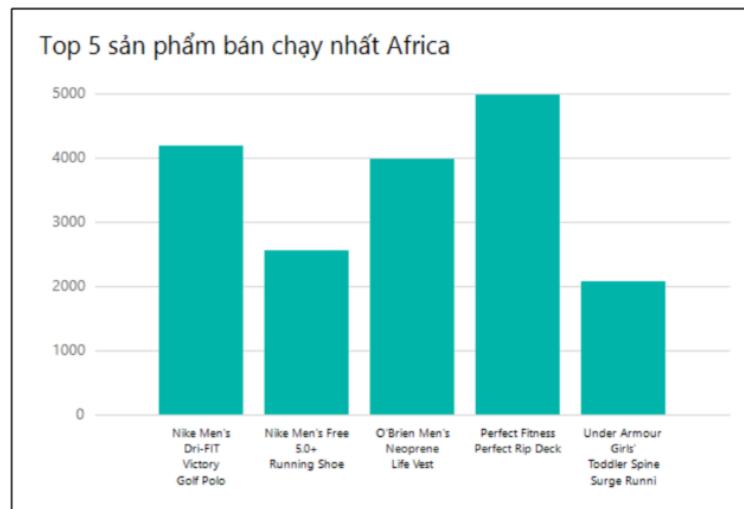


- Thực hiện Insert Table và chart để mô tả thông tin câu truy vấn, đồng thời tiến hành chỉnh sửa lại định dạng cho đẹp mắt



TOP 5 SẢN PHẨM BÁN CHẠY NHẤT AFRICA

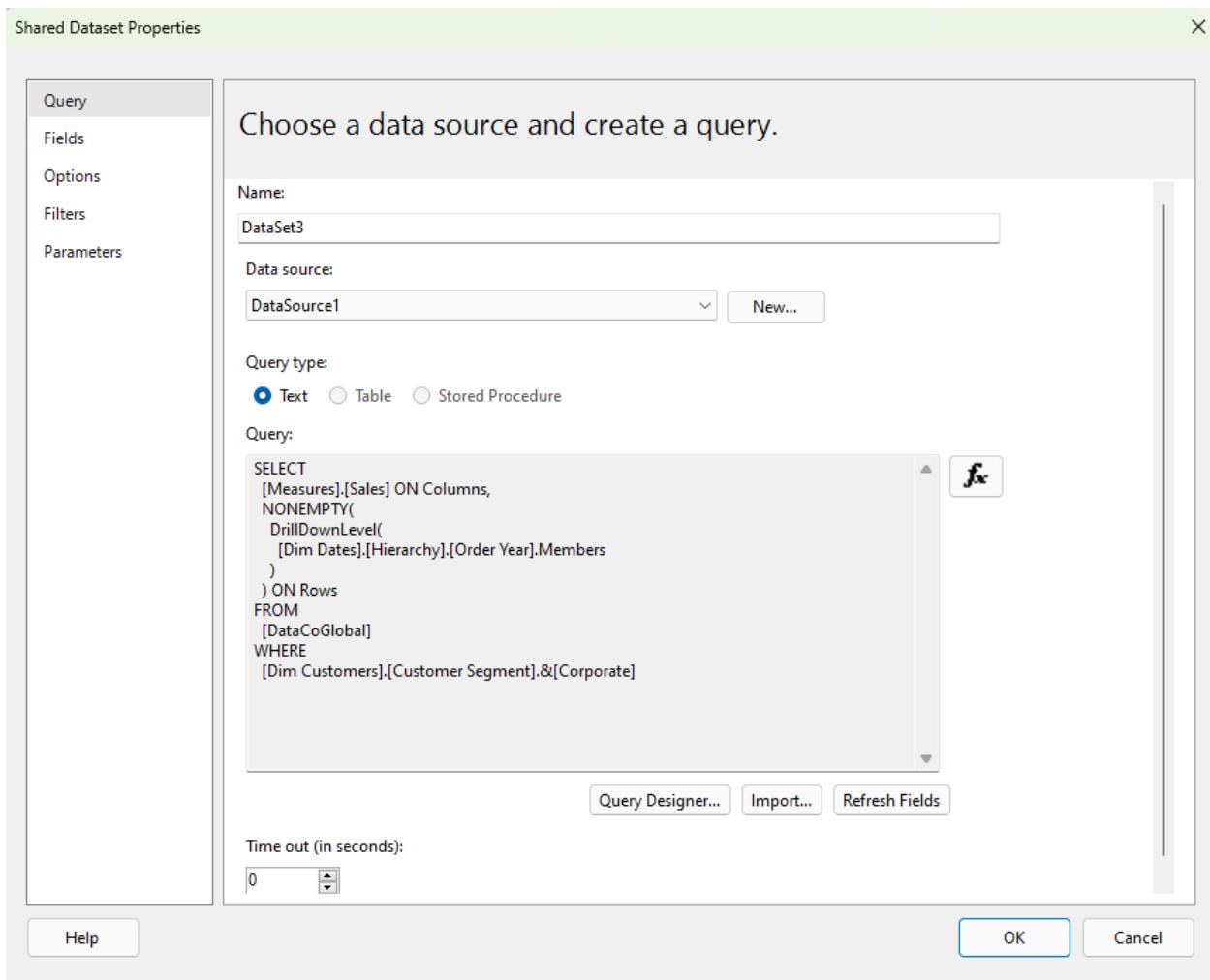
Product Name	Quantity
Perfect Fitness Perfect Rip Deck	4975
Nike Men's Dri-FIT Victory Golf Polo	4182
O'Brien Men's Neoprene Life Vest	3975
Nike Men's Free 5.0+ Running Shoe	2551
Under Armour Girls' Toddler Spine Surge Runni	2060



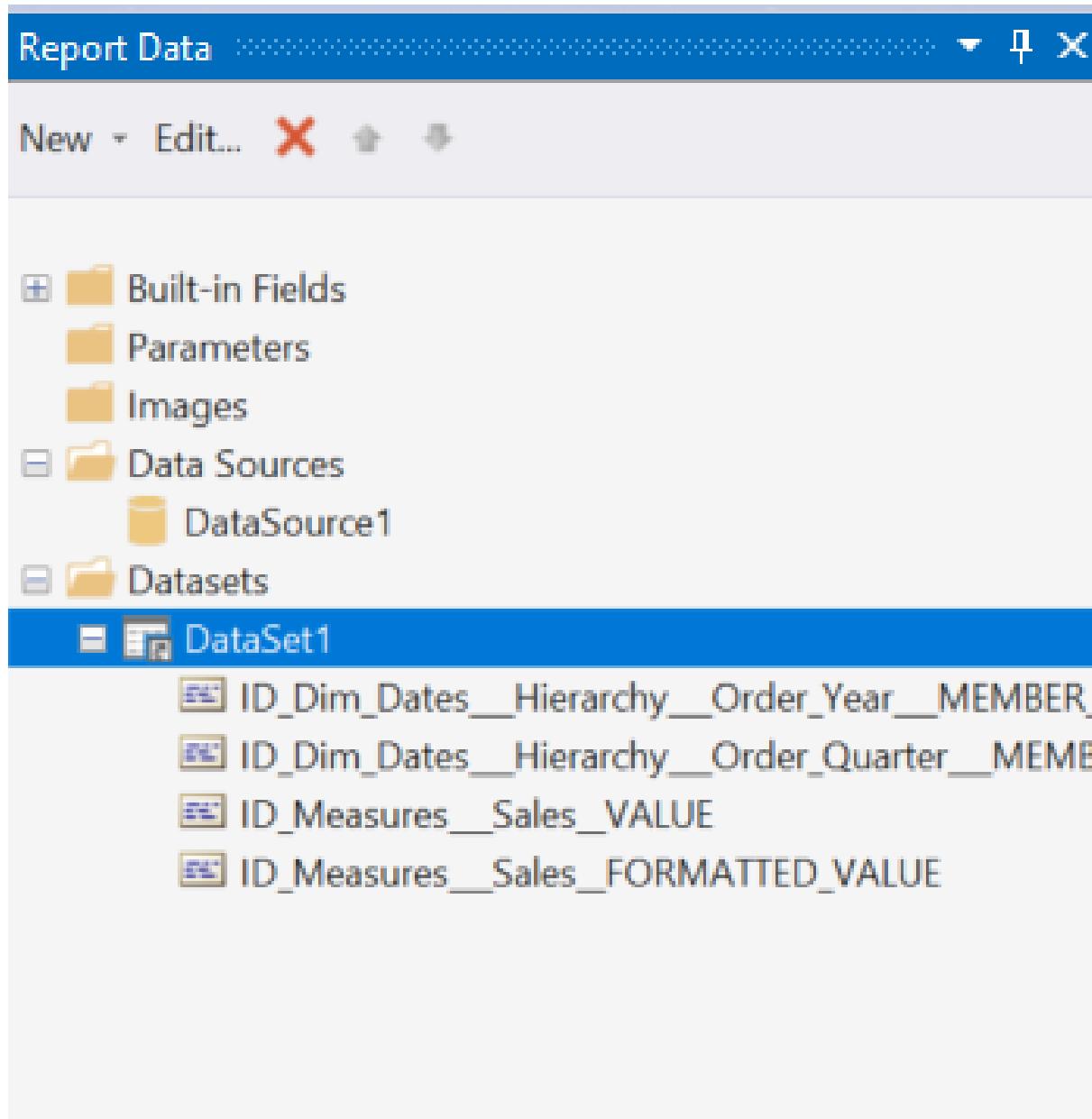
=> Dòng sản phẩm Rip Deck có số lượng bán được cao nhất với gần 5000 sản phẩm được bán ra, tiếp theo là Golf Polo và Life Vest có số lượng bán được tương đồng nhau. Và cuối cùng là các dòng sản phẩm Running Shoe, Surge Runni có số lượng sản phẩm bán được tương đối thấp hơn nhiều so với 3 dòng sản phẩm được đề cập ở trên.

4.1.2.3 Report - Câu truy vấn 9

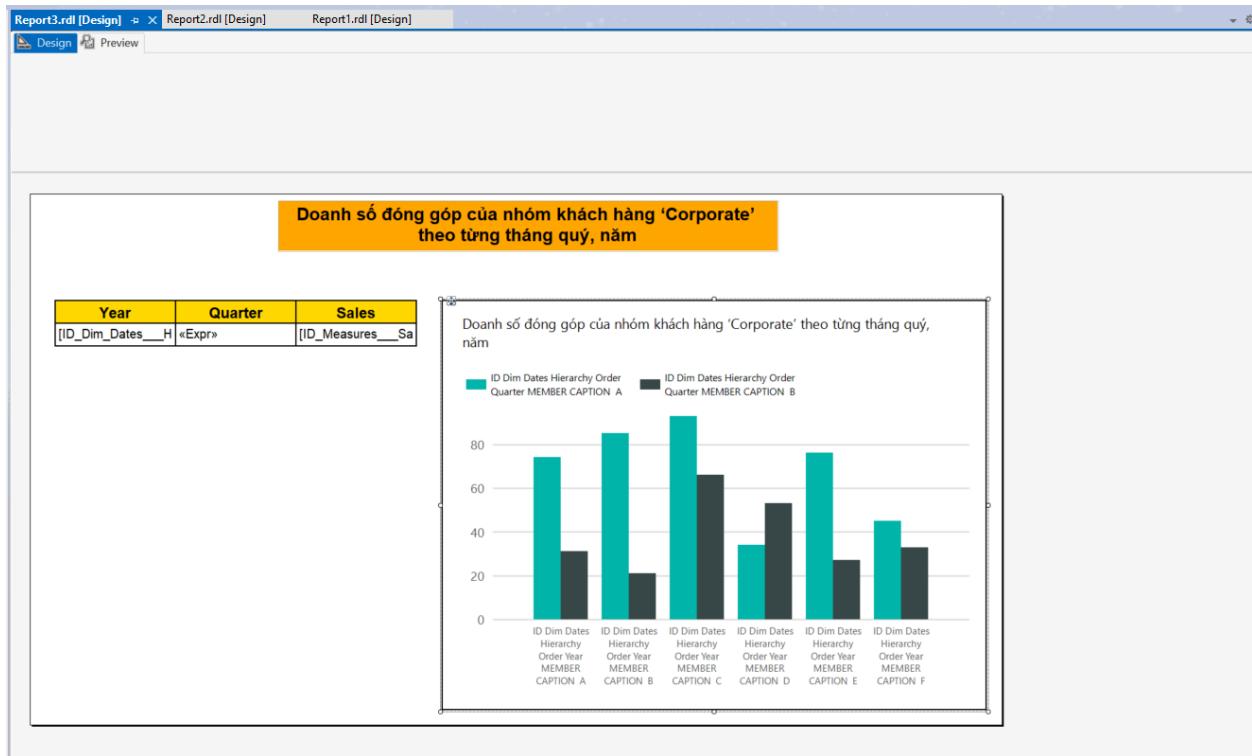
- Tiến hành thêm mới một dataset từ câu truy vấn



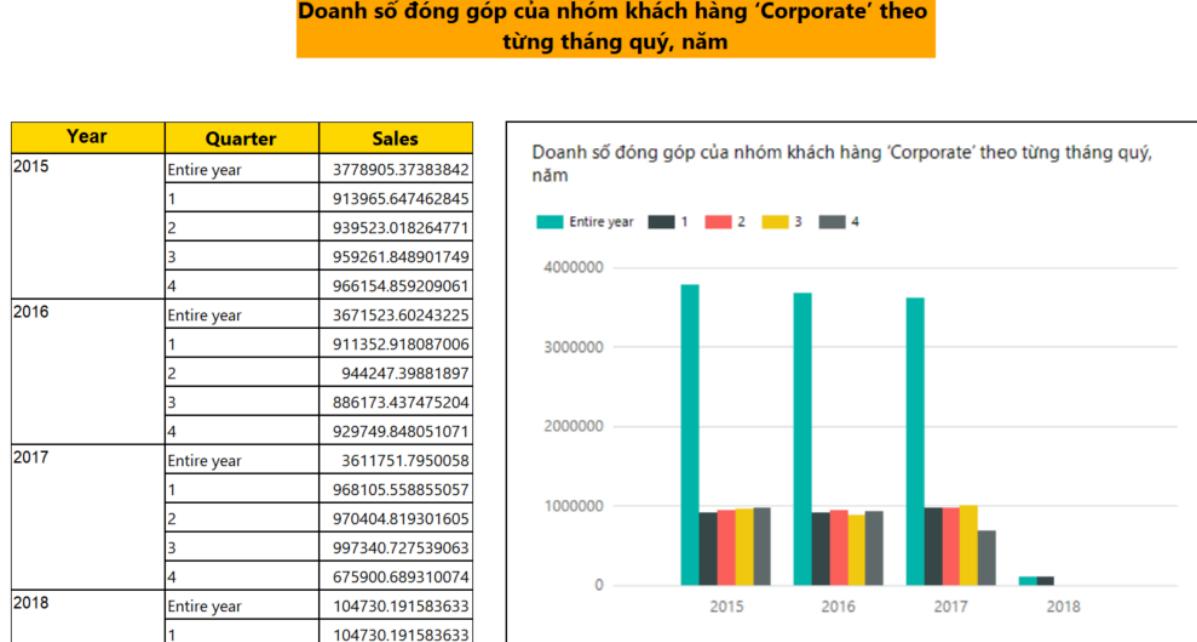
- Kiểm tra lại xem dataset đã được tạo



- Thực hiện Insert Table và chart cần thiết để mô tả thông tin từ câu truy vấn



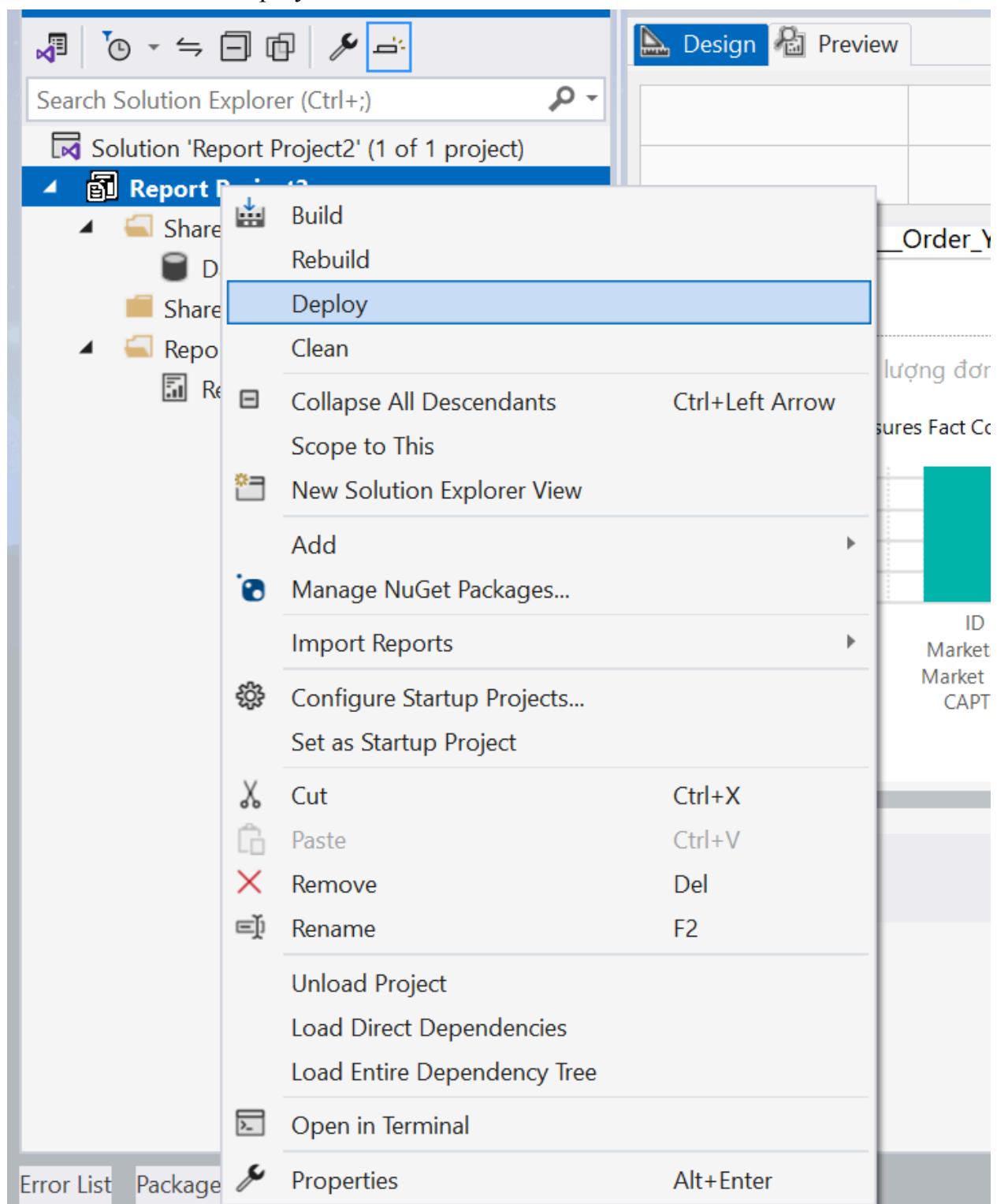
- Thực hiện xem review



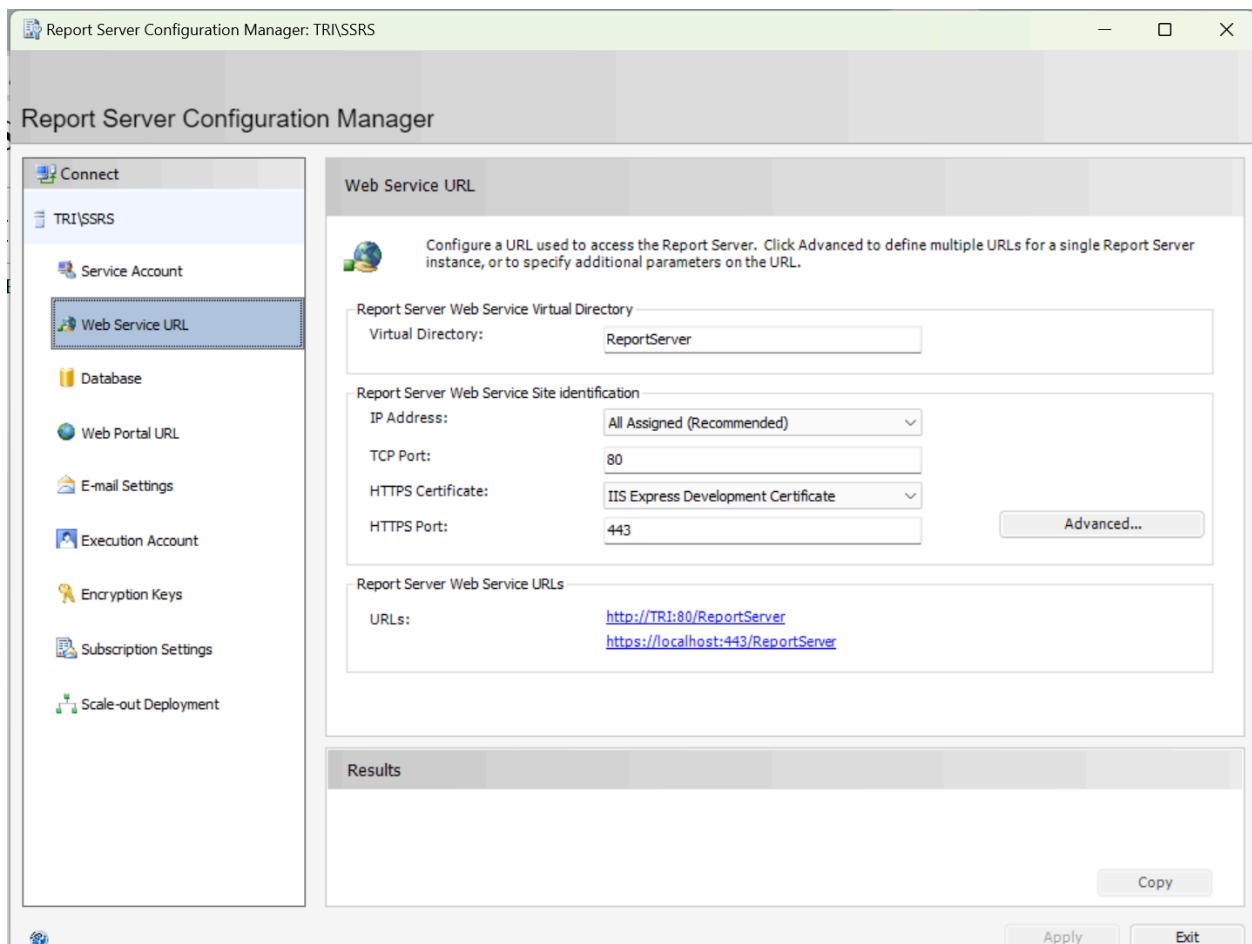
=> Doanh số đóng góp của nhóm khách hàng 'Corporate' qua các quý khá tương đồng nhau, tuy nhiên giảm mạnh ở những quý đầu của năm 2018.

4.1.3. Deploy Report

- Bước 1: ta thực hiện deploy



- Truy cập vào đường dẫn sau



localhost/ReportServer - /

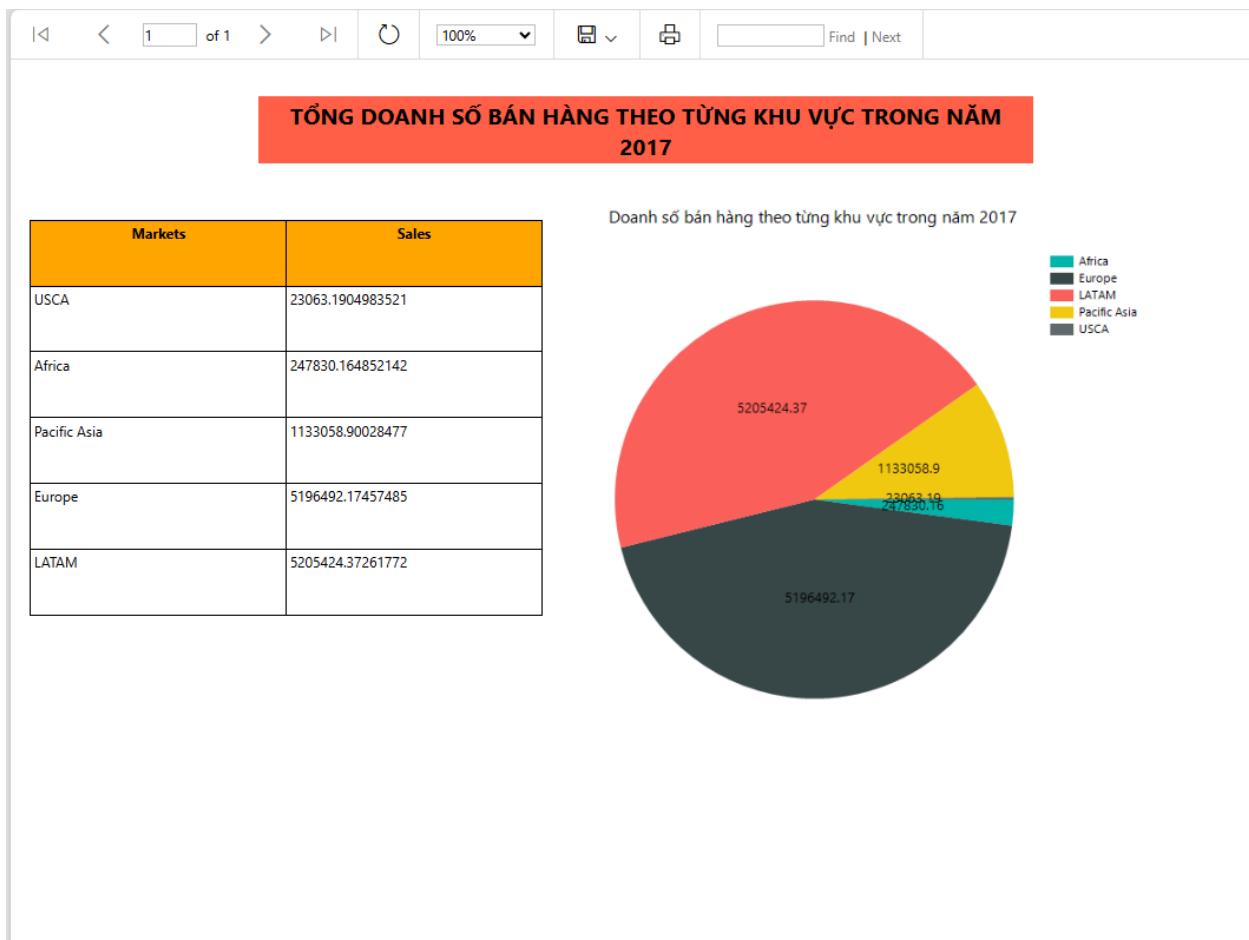
```

Monday, May 20, 2024 7:24 AM <dir> Data Sources
Monday, May 20, 2024 7:24 AM <dir> Datasets
Monday, May 20, 2024 7:24 AM <dir> Report Projects

```

Microsoft SQL Server Reporting Services Version 15.0.1102.1167

- Truy cập vào report đã deploy và xem



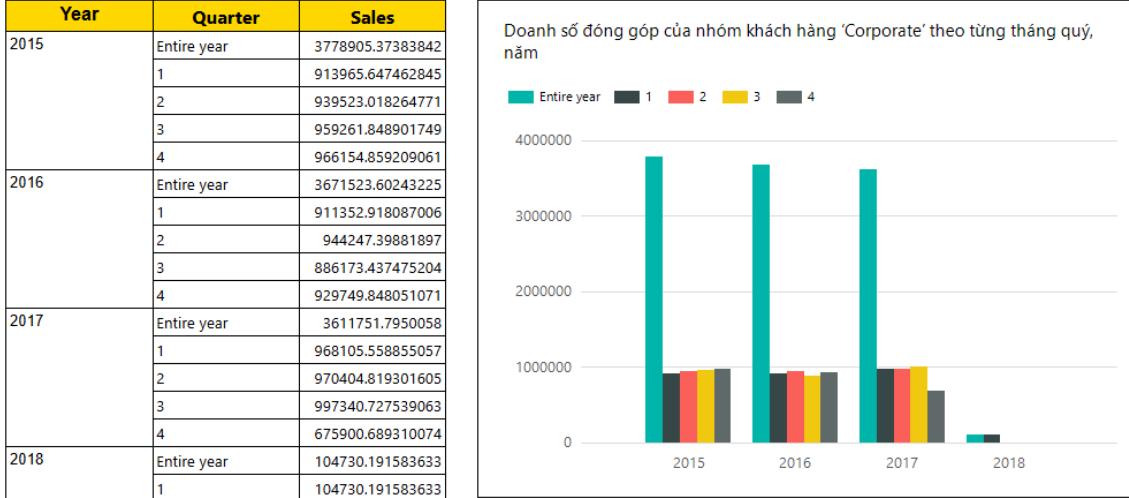
1 of 1 | 100% | Find | Next

TOP 5 SẢN PHẨM BÁN CHẠY NHẤT AFRICA

Product Name	Quantity
Perfect Fitness Perfect Rip Deck	4975
Nike Men's Dri-FIT Victory Golf Polo	4182
O'Brien Men's Neoprene Life Vest	3975
Nike Men's Free 5.0+ Running Shoe	2551
Under Armour Girls' Toddler Spine Surge Runni	2060

Top 5 sản phẩm bán chạy nhất Africa

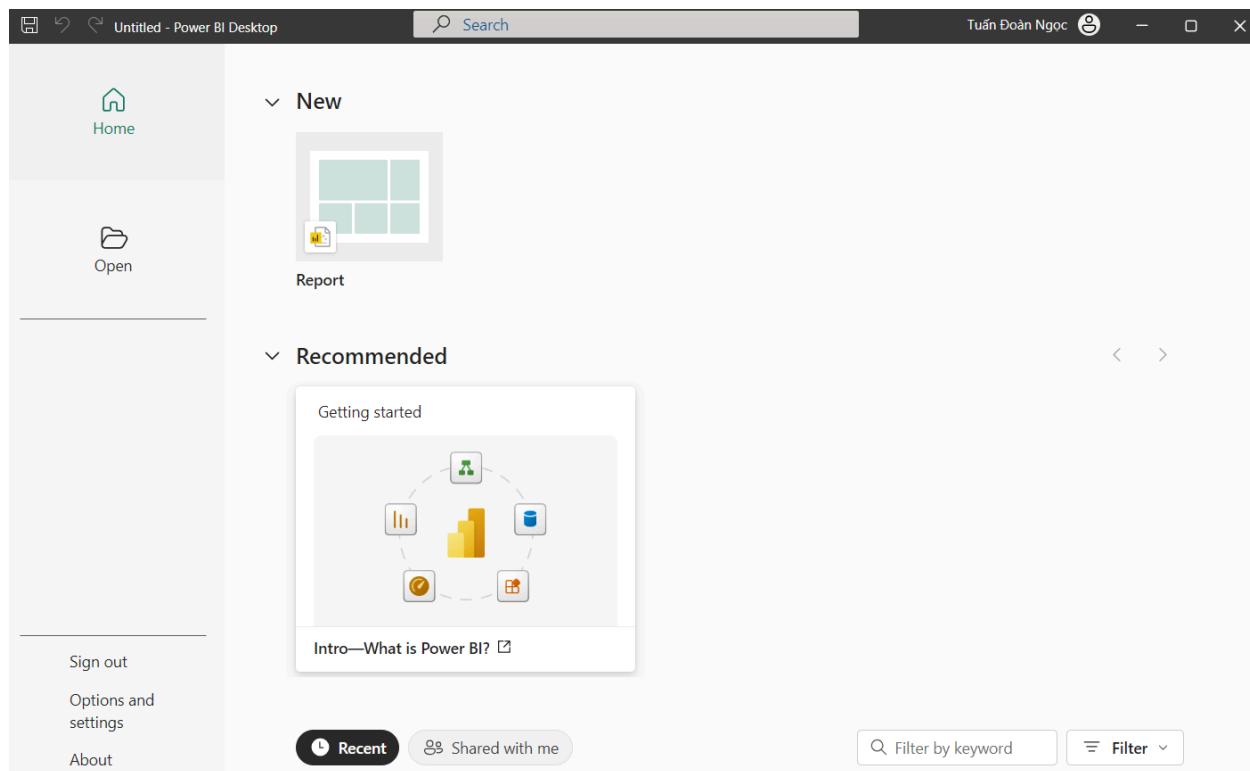
Product	Quantity
Nike Men's Dri-FIT Victory Golf Polo	4182
Nike Men's Free 5.0+ Running Shoe	2551
O'Brien Men's Neoprene Life Vest	3975
Perfect Fitness Perfect Rip Deck	4975
Under Armour Girls' Toddler Spine Surge Runni	2060



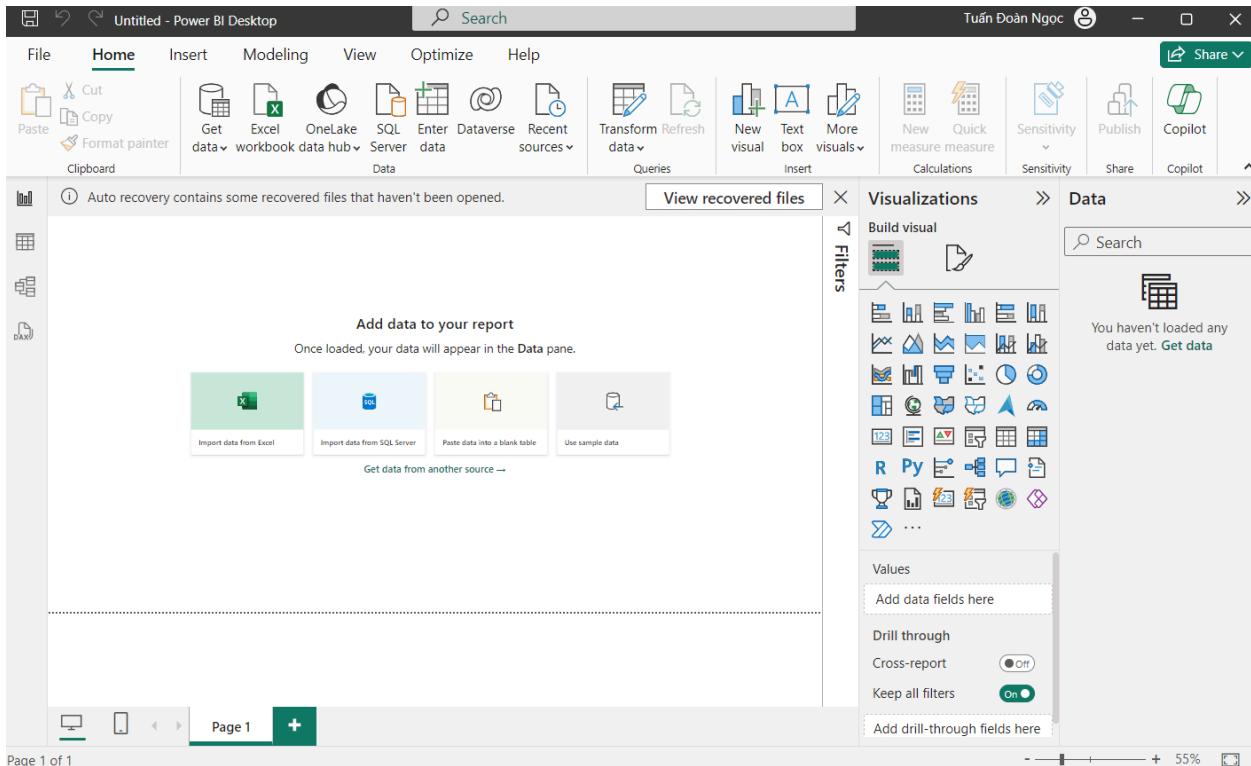
4.2. REPORT VỚI POWER BI

4.2.1. Tạo project SSRS trên Power BI

- Mở phần mềm Power BI -> New để tạo mới Project

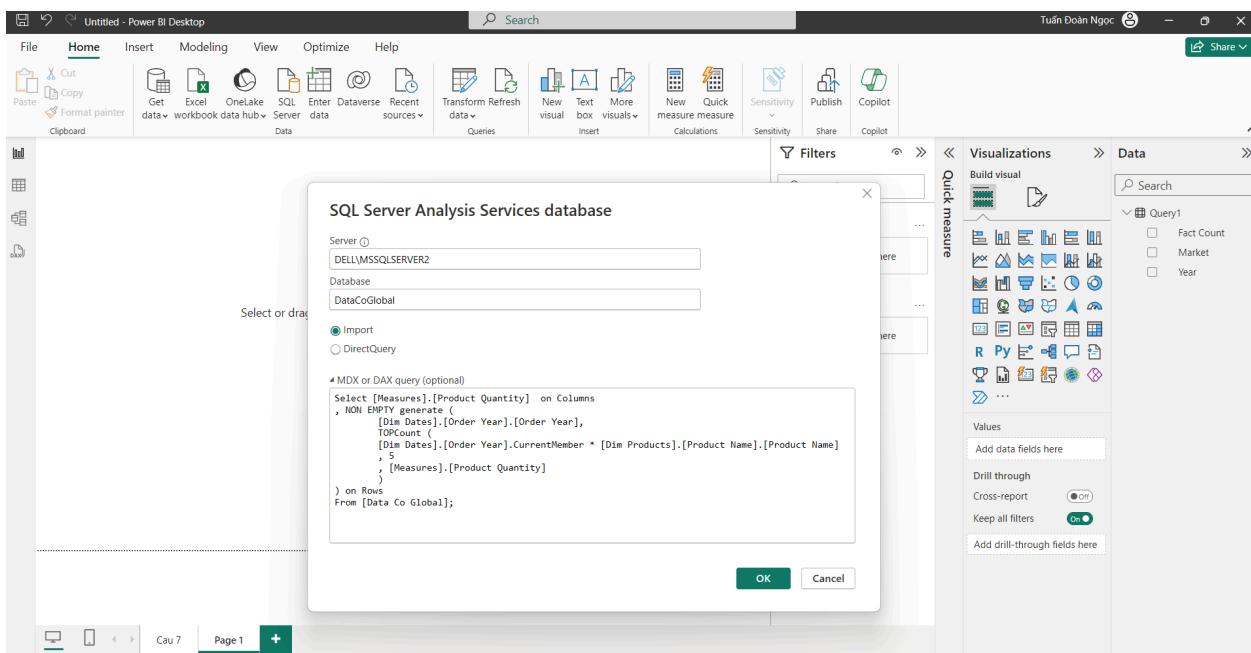


- Giao diện project mới sau khi tạo.



4.2.2. Report - Câu truy vấn 1

- Chọn Get Data -> Analysis Service -> Nhập thông tin Server, Database, MDX Query để lấy dữ liệu.



Đề tài: Phân tích doanh số bán hàng Data Global

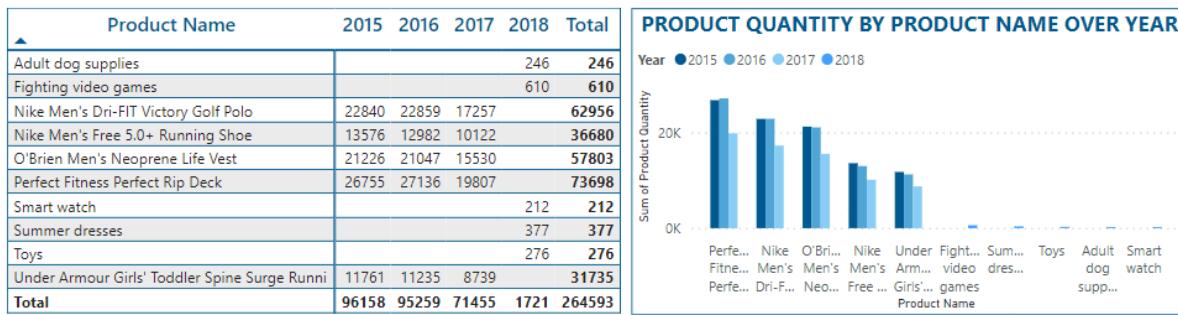
- Kiểm tra dữ liệu, chọn Transform Data để đổi tên thuộc tính và kiểu dữ liệu.

- Sau khi đổi thông tin, chọn Close & Apply

The screenshot shows the Power BI Desktop interface. On the left, there is a table visualization titled "Product Name" with columns for 2015, 2016, 2017, 2018, and Total. The table includes rows for various products like Under Armour Girls' Toddler Spine Surge Runni, Toys, Summer dresses, Smart watch, Perfect Fitness Perfect Rip Deck, O'Brien Men's Neoprene Life Vest, Nike Men's Free 5.0+ Running Shoe, Nike Men's Dri-FIT Victory Golf Polo, Fighting video games, and Adult dog supplies. The total value for all products is 264593. On the right, the "Visualizations" pane is open, showing a list of available chart types such as Bar, Line, Scatter, and Map. The "Data" pane is also visible, displaying the data source "Query1" with fields for Fact Count, Market, Year, Product Name, Product Quantity, and Year.

- Chọn Visualize theo Columns Chart

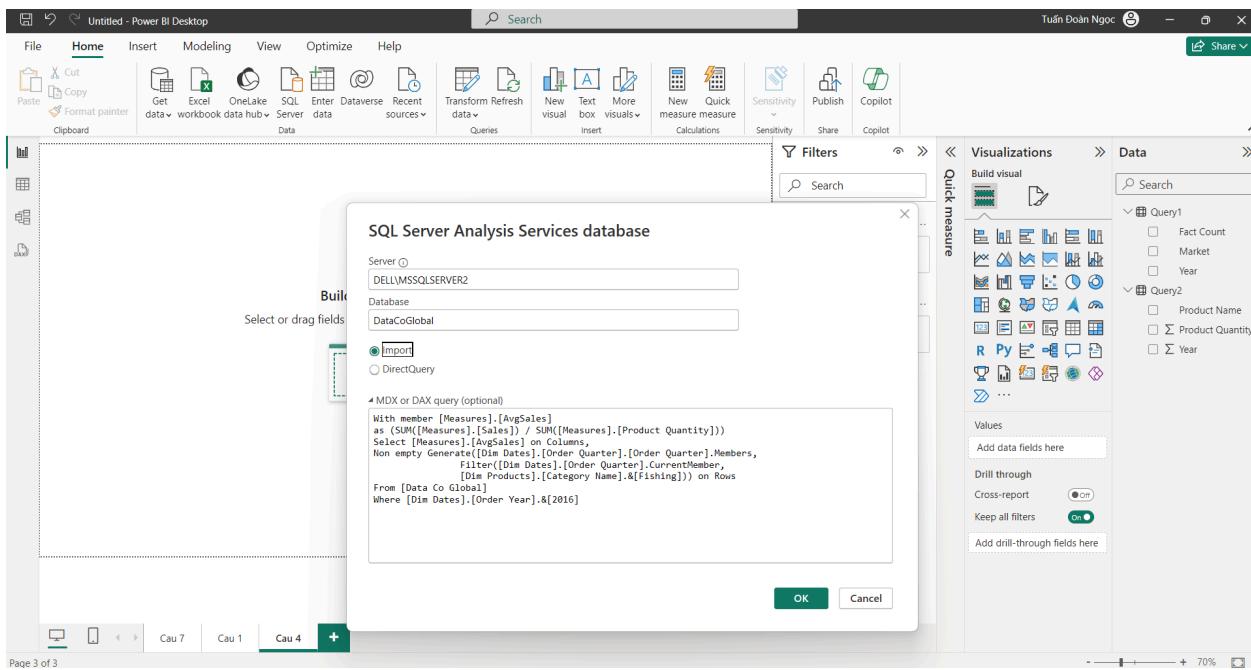
This screenshot shows the same Power BI Desktop environment after selecting a column chart. A bar chart titled "PRODUCT QUANTITY BY PRODUCT NAME OVER YEAR" is displayed. The Y-axis is labeled "Quantity" and ranges from 0K to 30K. The X-axis is labeled "Product Name". The chart shows the quantity for each product across four years (2015, 2016, 2017, 2018). The legend indicates the years: 2015 (blue), 2016 (orange), 2017 (green), and 2018 (red). The data points for each product are grouped by year. The "Visualizations" and "Data" panes are visible on the right side of the screen.



=> Chart thể hiện Top 5 sản phẩm bán chạy nhất theo các năm từ 2015 đến 2018. Dòng sản phẩm Rip Deck, Victory Golf và Life Vest có sự tăng trưởng vượt bậc so với các dòng sản phẩm còn lại, nhưng có xu hướng giảm dần theo các năm.

4.2.3. Report - Câu truy vấn 4

- Chọn Get Data -> Analysis Service -> Nhập thông tin Server, Database, MDX Query để lấy dữ liệu.



- Kiểm tra dữ liệu, chọn Transform Data để đổi tên thuộc tính và kiểu dữ liệu.

DELL\MSSQLSERVER2: DataCoGlobal

[Dim Dates].[Order Quarter].[Order Quarter].[MEMBER]	[Measures].[AvgSales]
1	90.19917001
2	89.7031648
3	89.4396982
4	89.00175728

Load Transform Data Cancel

- Sau khi đổi thông tin, chọn Close & Apply

Quarter	Average Sales Per Product
1	90
2	90
3	89
4	89

 The 'APPLIED STEPS' pane shows the steps taken: 'Renamed Columns' and 'Changed Type'. The 'Properties' pane shows the query is named 'Query3'.

- Chọn Matrix ở tab **Visualizations** và biểu diễn dữ liệu

Quarter	Sum of Average Sales Per Product
1	90
2	90
3	89
4	89
Total	358

- Biểu diễn theo Columns Chart

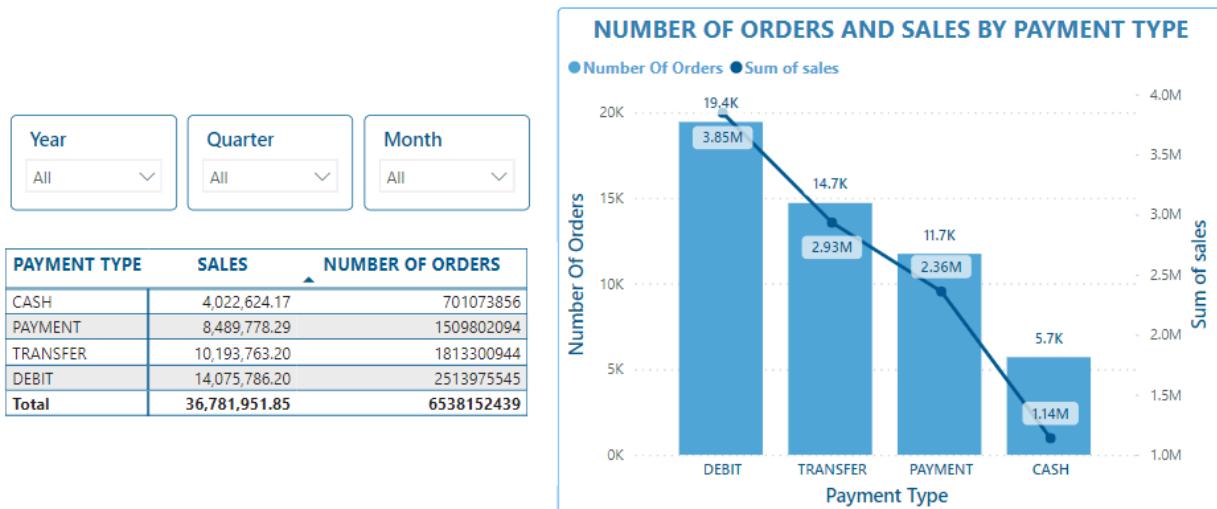
Quarter	Average Sales Per Product
1	90
2	90
3	89
4	89



=> Doanh thu trung bình trên 1 sản phẩm bán ra vào các quý của năm 2016, thuộc danh mục Fishing không có sự thay đổi lớn giữa các quý.

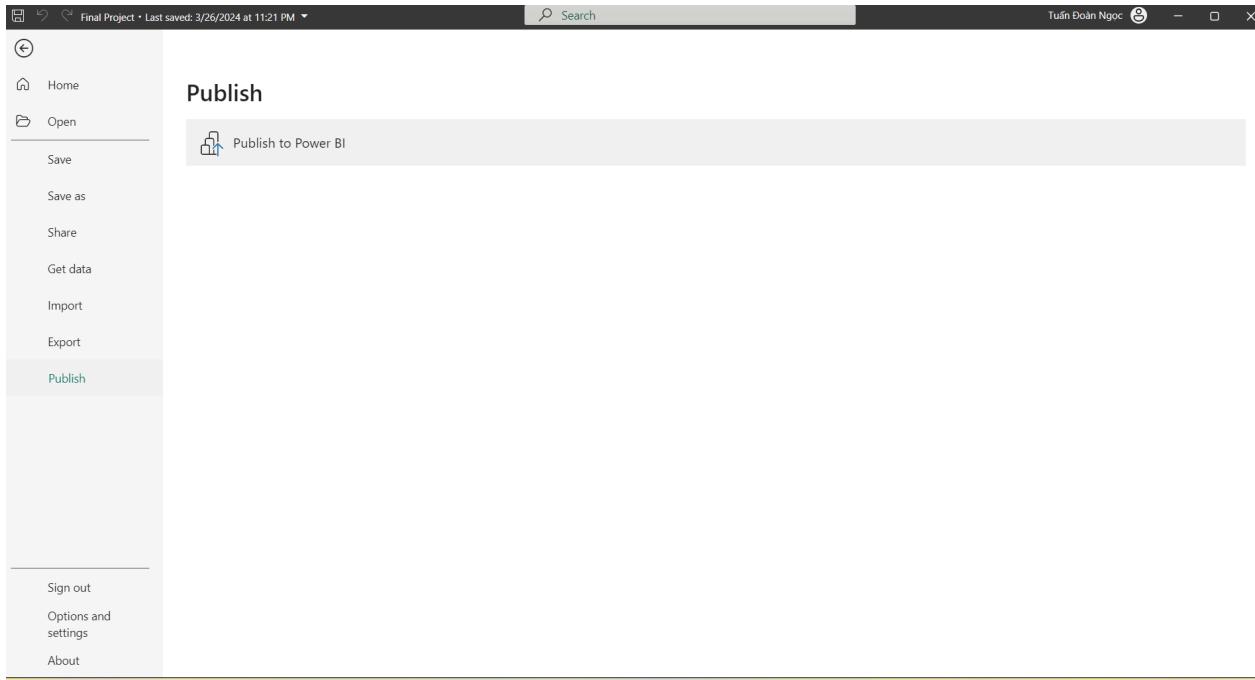
4.2.4. Report - Câu truy vấn 8

- Thực hiện các bước như 2 câu truy vấn trên



4.2.5. Triển khai SSRS lên POWER BI SERVICE

- Chọn File-> Publish -> Publish to Power BI để đưa report lên Power BI Service



- Tại cửa sổ Publish to Power BI -> My workspace -> Select để xác nhận nơi lưu trữ Report trên Server.

Đè tài: Phân tích doanh số bán hàng Data Global

- Sau khi Public thành công thì sẽ hiển thị màn hình như bên dưới, chọn Open 'Final Project.pbix' in Power BI để mở report vừa public.

- Report sau khi được đẩy lên Power BI Service

Chương 5: QUÁ TRÌNH THỰC HIỆN DATA MINING

5.1. MÔ TẢ DATASET VÀ ĐỒ DỮ LIỆU

5.1.1. Mô tả dataset

- Tên dataset: **Diabetes prediction dataset**
- Link dataset:
<https://www.kaggle.com/datasets/jammustafatz/diabetes-prediction-dataset>
- Dữ liệu gồm 100.000 dòng, 10 thuộc tính
 - + **id**: mã bệnh nhân
 - + **gender**: giới tính bệnh nhân
 - + **age**: tuổi của bệnh nhân
 - + **hypertension**: cho biết bệnh nhân có bị huyết áp cao hay không (0 hoặc 1)
 - + **heart_disease**: cho biết bệnh nhân có bị bệnh tim hay không (0 hoặc 1)
 - + **smoking_history**: tiền sử hút thuốc của bệnh nhân (not current,former,No Info,current,never and ever)
 - + **bmi**: chỉ số BMI
 - + **hbA1c_level**: lượng đường trung bình trong máu
 - + **blood_glucose_level**: mức đường huyết của bệnh nhân
 - + **diabetes**: có bị bệnh tiểu đường hay không (0: không, 1: có)

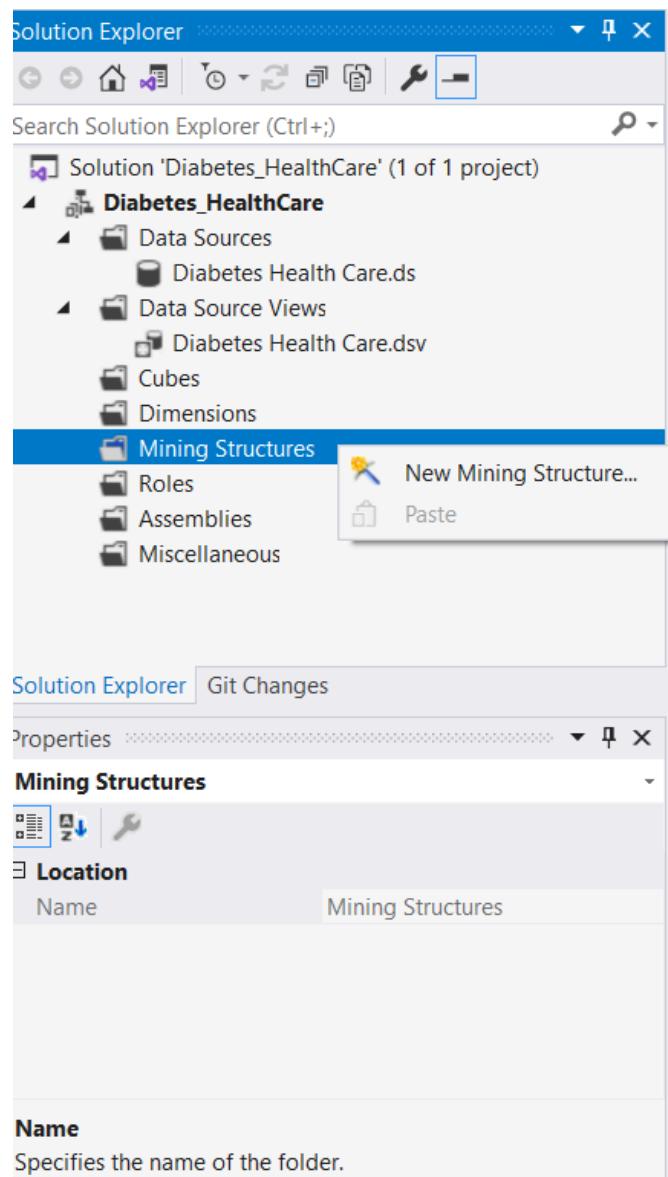
id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
1	Female	80	0	1	never	25.19	6.6	140	0
2	Female	54	0	0	No Info	27.32	6.6	80	0
3	Male	28	0	0	never	27.32	5.7	158	0
4	Female	36	0	0	current	23.45	5	155	0
5	Male	76	1	1	current	20.14	4.8	155	0
6	Female	20	0	0	never	27.32	6.6	85	0
7	Female	44	0	0	never	19.31	6.5	200	1
8	Female	79	0	0	No Info	23.86	5.7	85	0
9	Male	42	0	0	never	33.64	4.8	145	0
10	Female	32	0	0	never	27.32	5	100	0
11	Female	53	0	0	never	27.32	6.1	85	0
12	Female	54	0	0	former	54.7	6	100	0
13	Female	78	0	0	former	36.05	5	130	0
14	Female	67	0	0	never	25.69	5.8	200	0
15	Female	76	0	0	No Info	27.32	5	160	0
16	Male	78	0	0	No Info	27.32	6.6	126	0
17	Male	15	0	0	never	30.36	6.1	200	0
18	Female	42	0	0	never	24.48	5.7	158	0
19	Female	42	0	0	No Info	27.32	5.7	80	0
20	Male	37	0	0	ever	25.72	3.5	159	0
21	Male	40	0	0	current	36.38	6	90	0
22	Male	5	0	0	No Info	18.8	6.2	85	0
23	Female	69	0	0	never	21.24	4.8	85	0
24	Female	72	0	1	former	27.94	6.5	130	0
25	Female	4	0	0	No Info	13.99	4	140	0
26	Male	30	0	0	never	33.76	6.1	126	0
27	Male	67	0	1	not current	27.32	6.5	200	1
28	Male	40	0	0	former	27.85	5.8	80	0

5.1.2. Đỗ dữ liệu vào kho

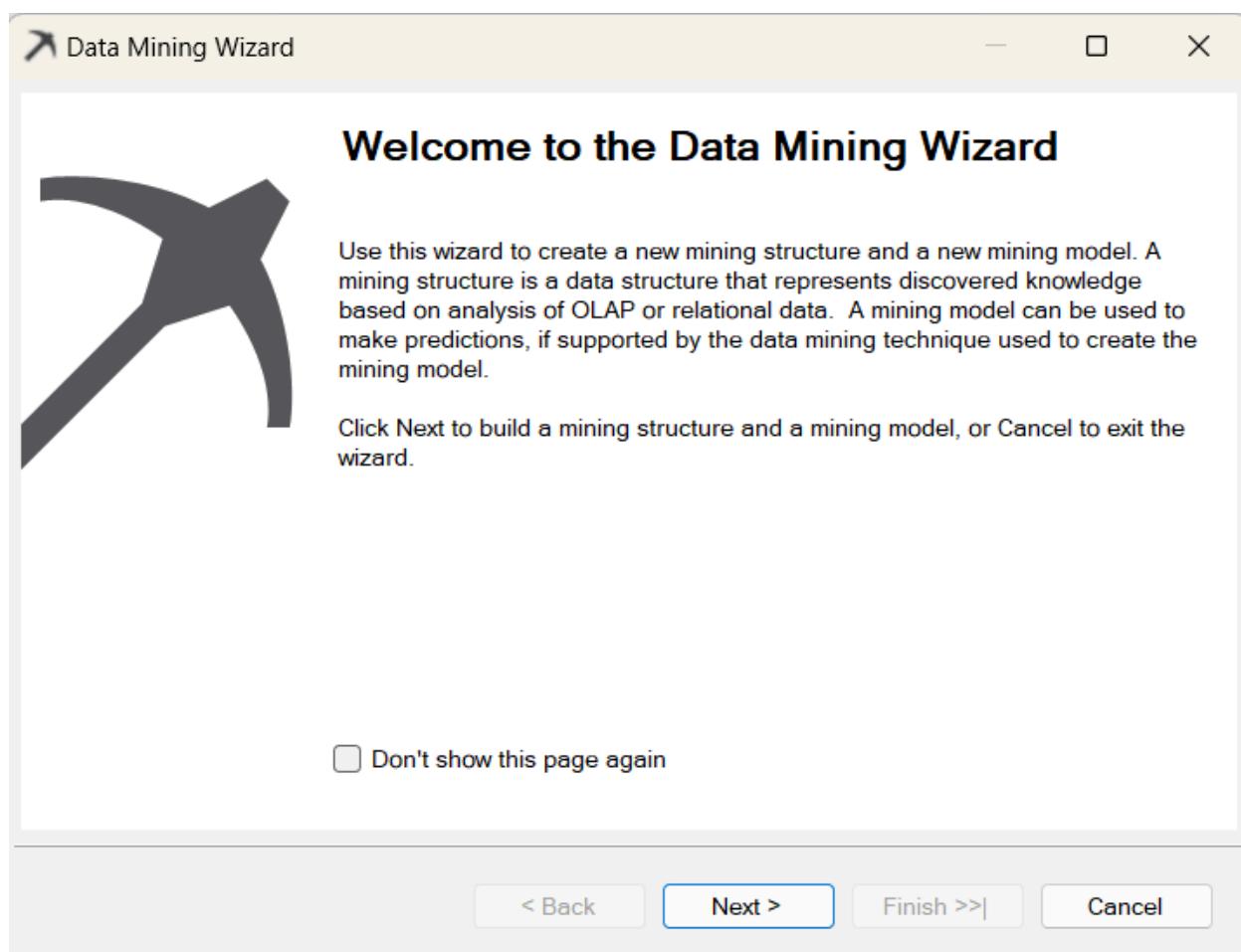
- Thực hiện như quy trình SSIS, SSAS

5.2. TẠO PROJECT DATA MINING

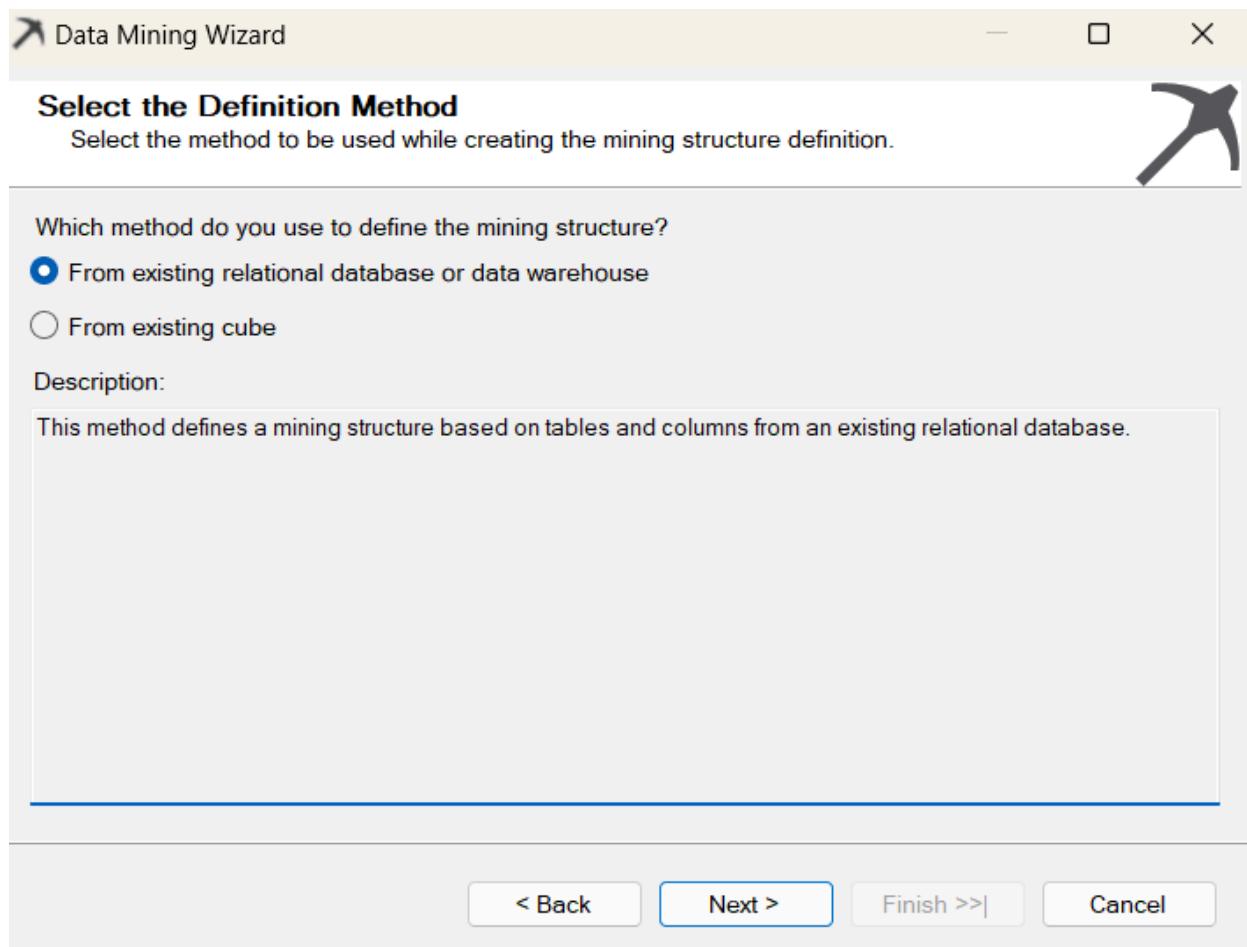
- Ở Solution Explorer, chọn Mining Structure -> New Mining Structure



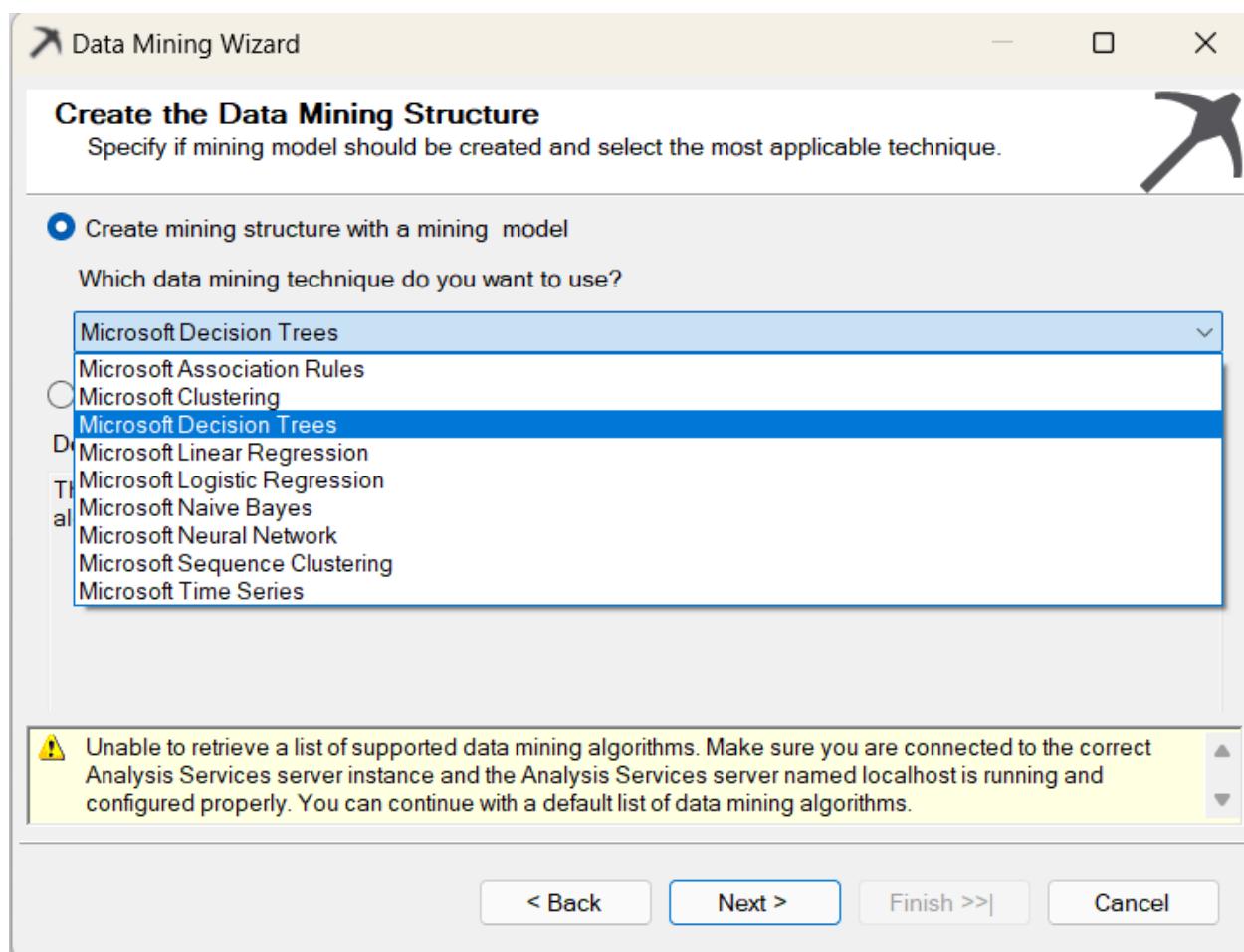
- Chọn Next ở cửa sổ **Data Mining Wizard**



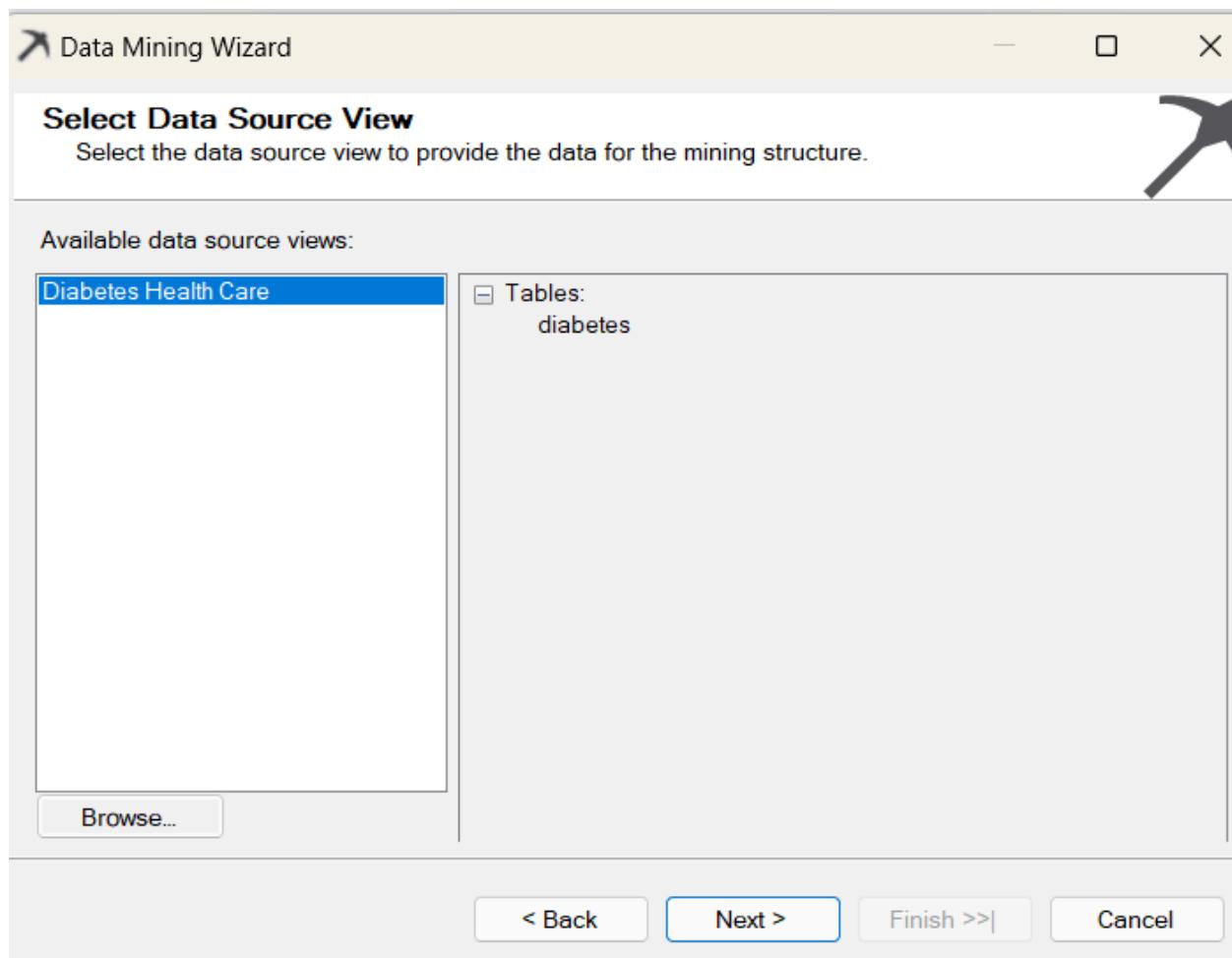
- Chọn **From existing relational database or database warehouse** -> **Next**



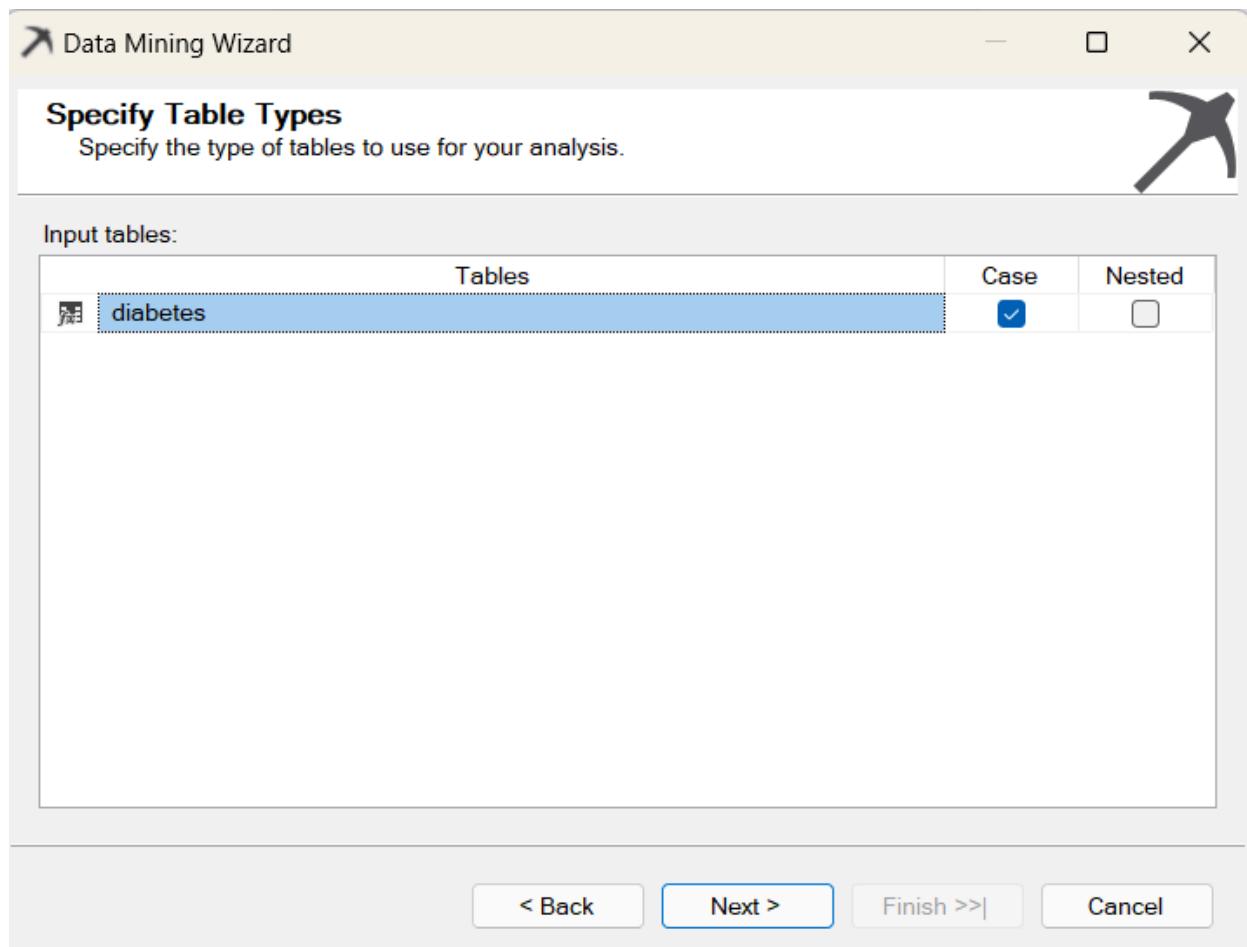
- Ở **Create mining structure with a mining model**, chọn thuật toán muôn phân tích -> **Next**



- Add Data Source Views -> Next



- Chọn tables làm case để chạy thuật toán -> Next



- Chọn **id** làm key và chọn **diabetes** làm thuộc tính dự đoán (0: không bệnh tiểu đường và 1: bệnh tiểu đường).

Data Mining Wizard

Specify the Training Data

Specify the columns used in your analysis.

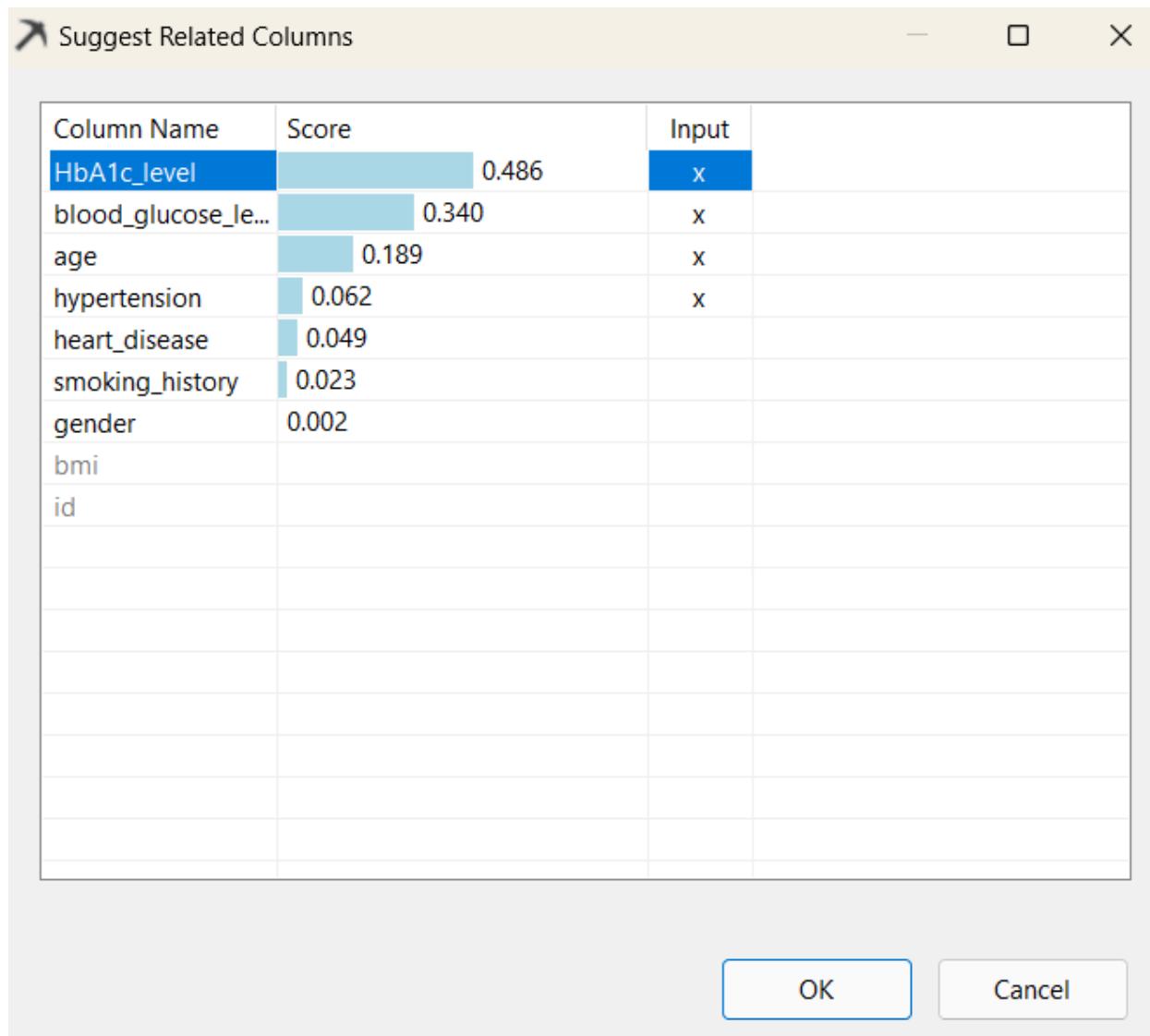
Mining model structure:

	Tables/Columns	Key	<input type="checkbox"/> Input	<input checked="" type="checkbox"/> Predictive
-	diabetes			
<input type="checkbox"/>	age	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	blood_glucose_level	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	bmi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	diabetes	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	gender	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	HbA1c_level	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	heart_disease	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	hypertension	<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 type="checkbox"/>	smoking_history	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Recommend inputs for currently selected predictable:

⚠ No input column is defined for this mining model.

- Có thể chọn **Suggest** để phần đưa ra những gợi ý về các thuộc tính chọn để dự đoán.



- Chọn các thuộc tính làm input: **gender, age, hypertension, heart_disease, smoking_history, bmi, hbA1c_level, blood_glucose_level.** -> Chọn Next

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

Mining model structure:

	Tables/Columns	Key	Input	Predi...
-	diabetes			
	age	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	blood_glucose_level	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	bmi	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	diabetes	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
	gender	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	HbA1c_level	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	heart_disease	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	hypertension	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	id	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	smoking_history	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Recommend inputs for currently selected predictable:

- Chọn **detect** để điều chỉnh lại kiểu dữ liệu.

 Data Mining Wizard

Specify Columns' Content and Data Type

Specify mining structure columns' content and data type.

Mining model structure:

Columns	Content Type	Data Type
Age	Continuous	Double
Blood Glucose Level	Continuous	Long
Bmi	Continuous	Double
Diabetes	Discrete	Boolean
Gender	Discrete	Text
Hb A1c Level	Continuous	Double
Heart Disease	Discrete	Boolean
Hypertension	Discrete	Boolean
Id	Key	Double
Smoking History	Discrete	Text

Detect continuous or discrete for numeric columns:

- Sau khi detect -> Next

Data Mining Wizard

Specify Columns' Content and Data Type

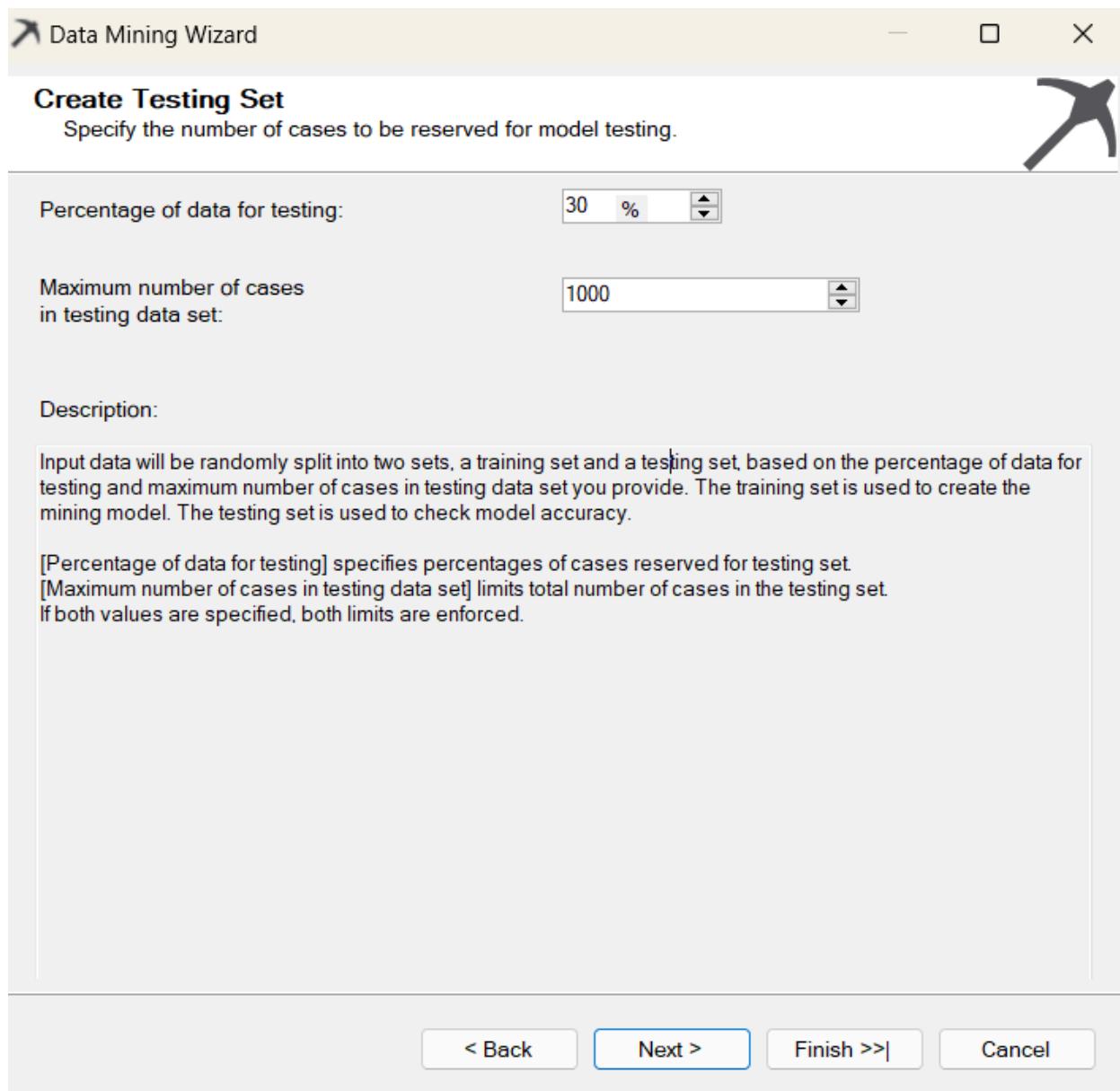
Specify mining structure columns' content and data type.

Mining model structure:

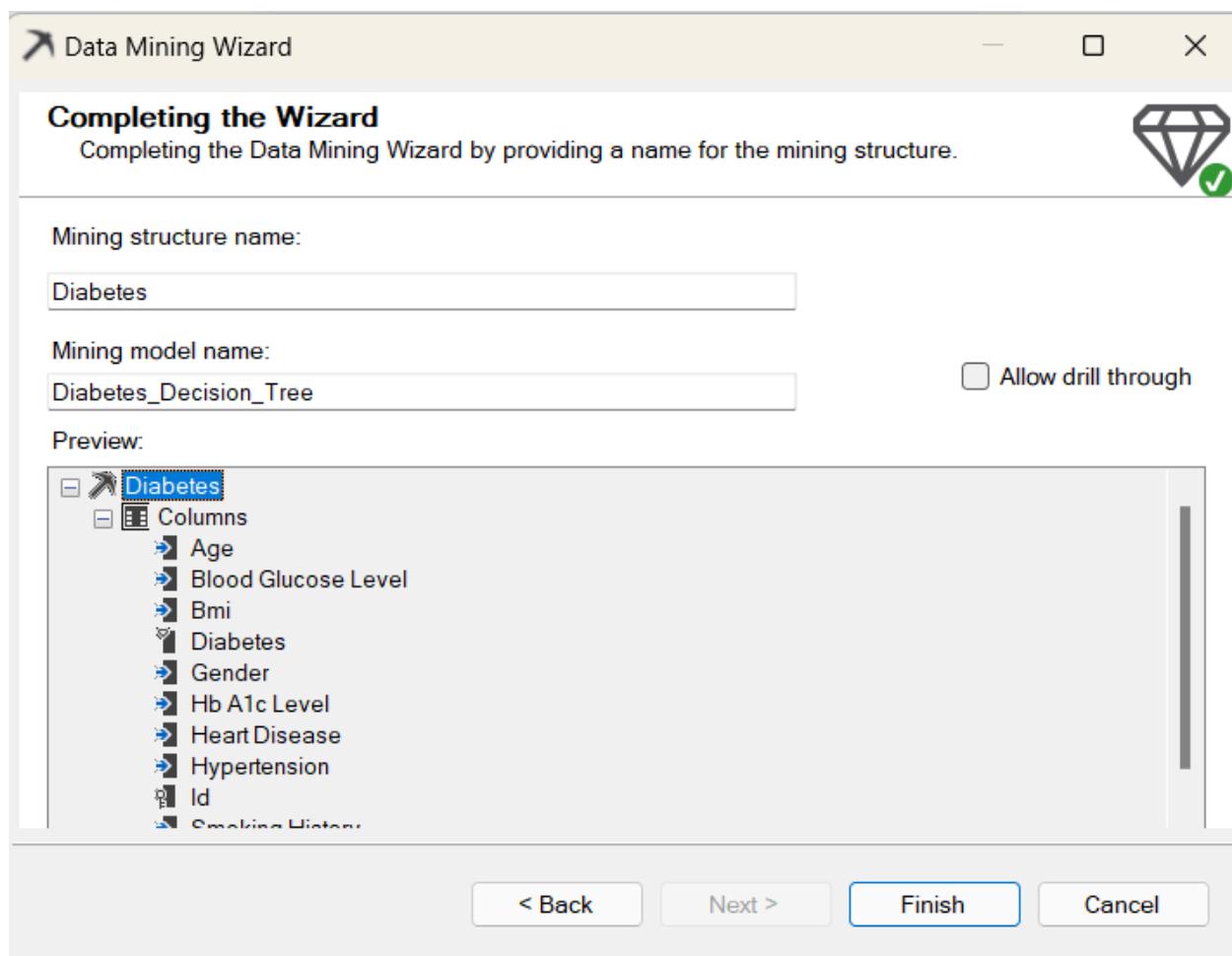
Columns	Content Type	Data Type
Age	Discrete	Double
Blood Glucose Level	Continuous	Long
Bmi	Continuous	Double
Diabetes	Discrete	Boolean
Gender	Discrete	Text
Hb A1c Level	Continuous	Double
Heart Disease	Discrete	Boolean
Hypertension	Discrete	Boolean
Id	Key	Double
Smoking History	Discrete	Text

Detect continuous or discrete for numeric columns:

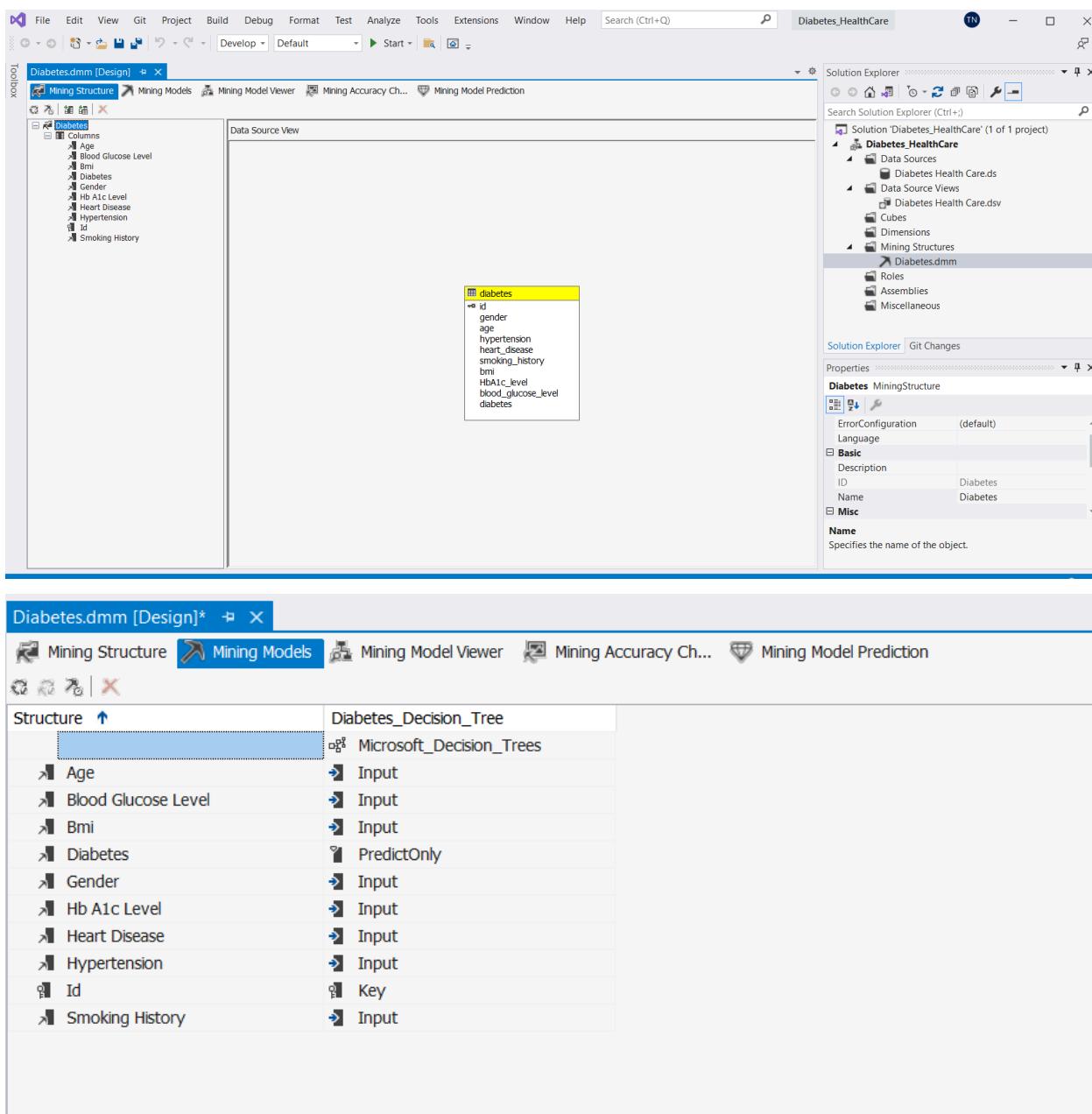
- Chọn **70%** dữ liệu để train và **30%** dữ liệu để train, chọn số case test tối đa là 1000 -> **Next**.



- Đặt tên cho Mining Structure và Mining Model -> Finish



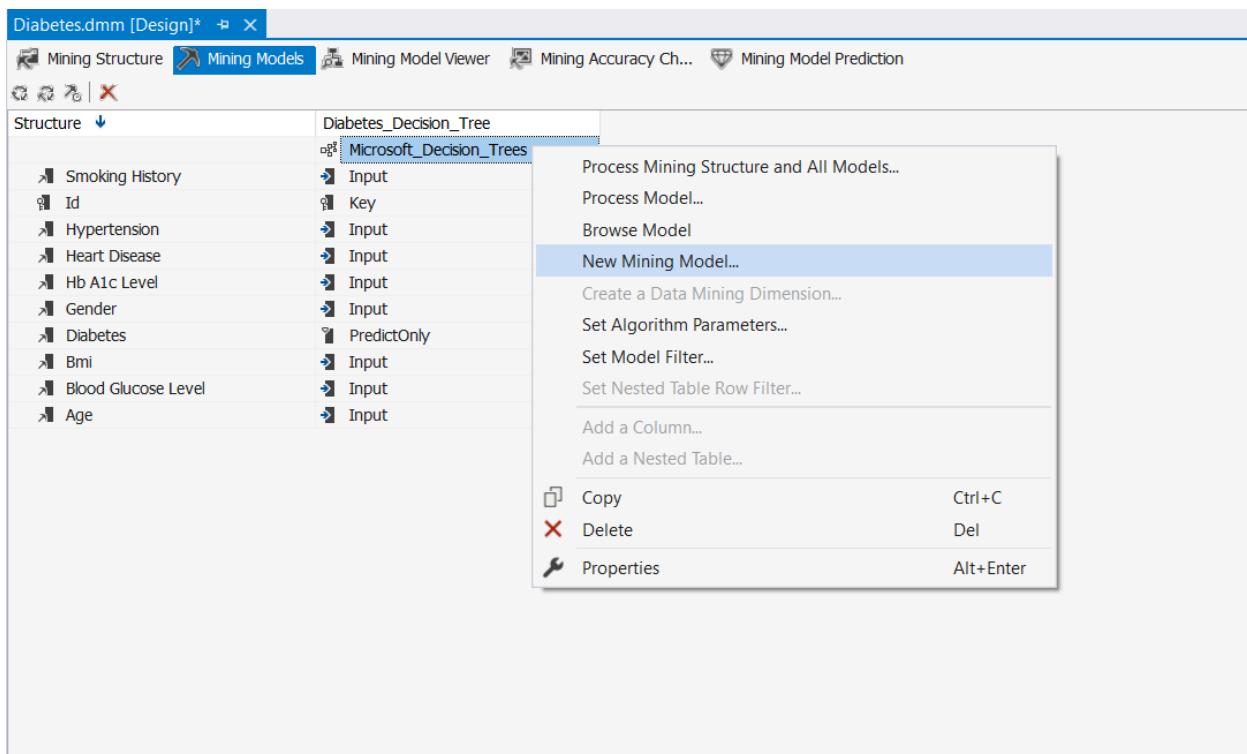
- **Mining Structure** sau khi được tạo



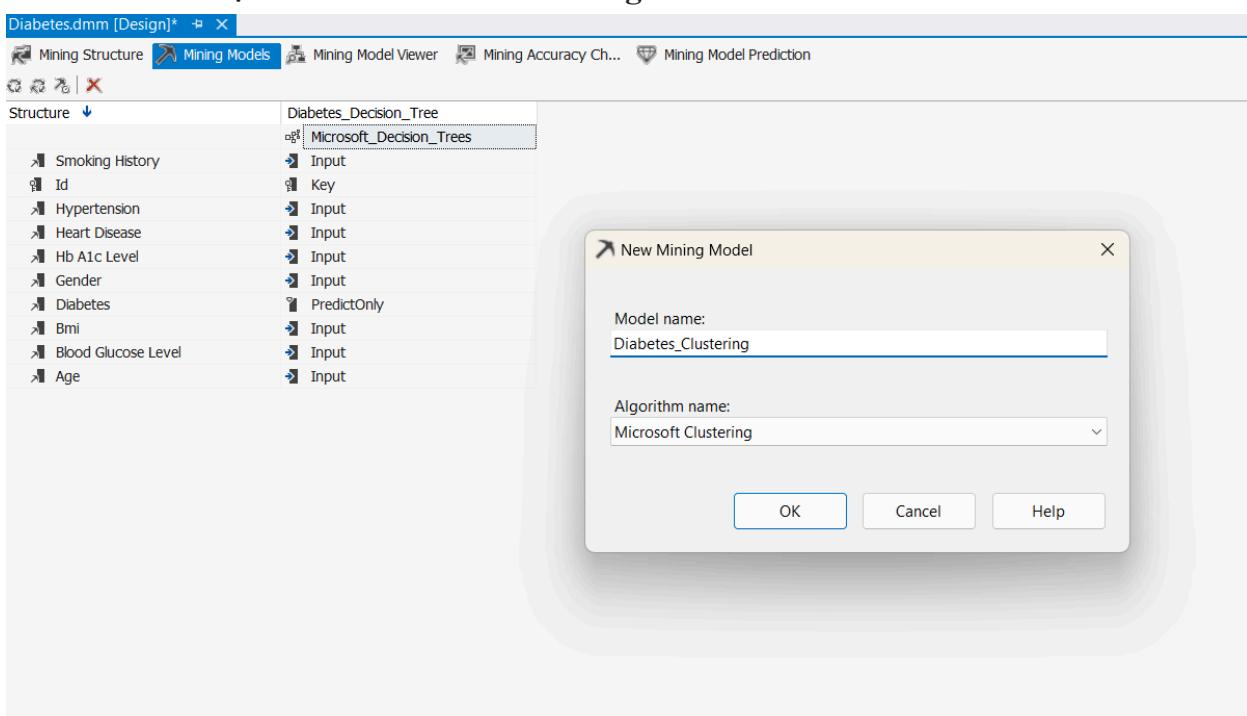
- Ở tab **Properties** -> Điều chỉnh **HoldoutSeed** thành 12



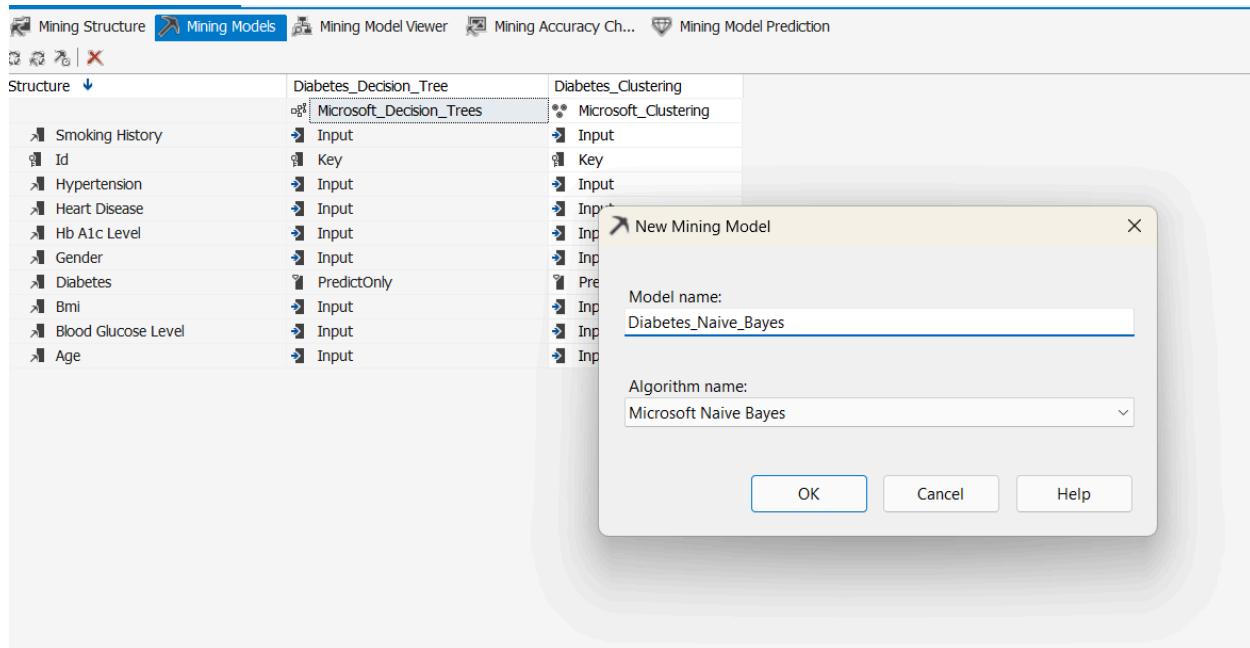
- Chọn **New Mining Model** để thêm các thuật toán mới.



- Thêm thuật toán Microsoft Clustering



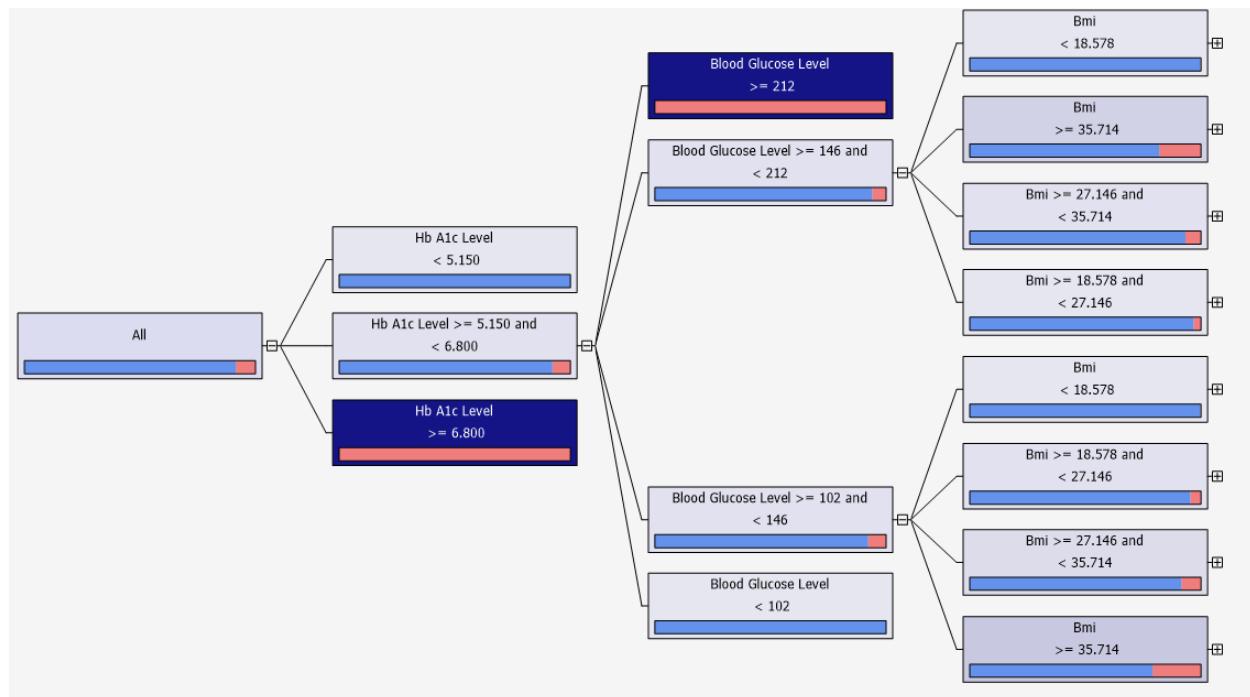
- Thêm thuật toán Naive Bayes



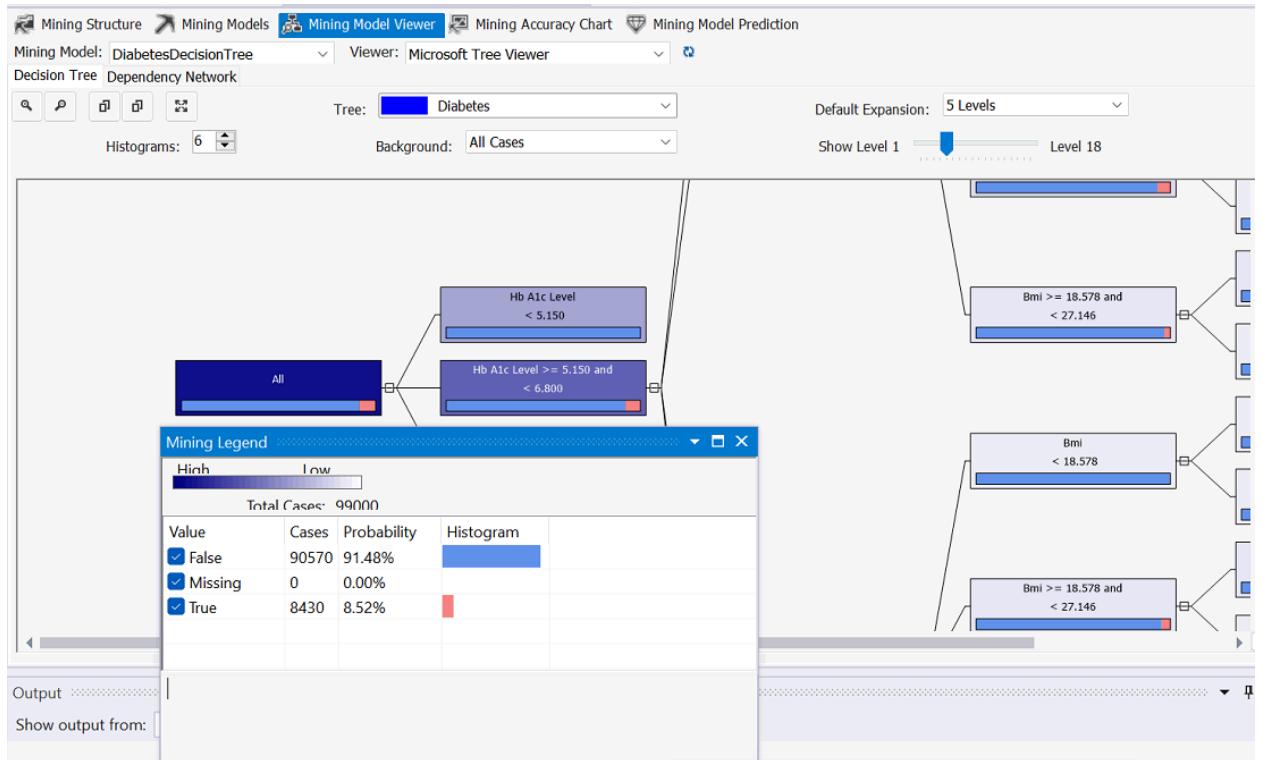
- Sau đó tiến hành **Run Process**

5.3. NHẬN XÉT CÁC THUẬT TOÁN

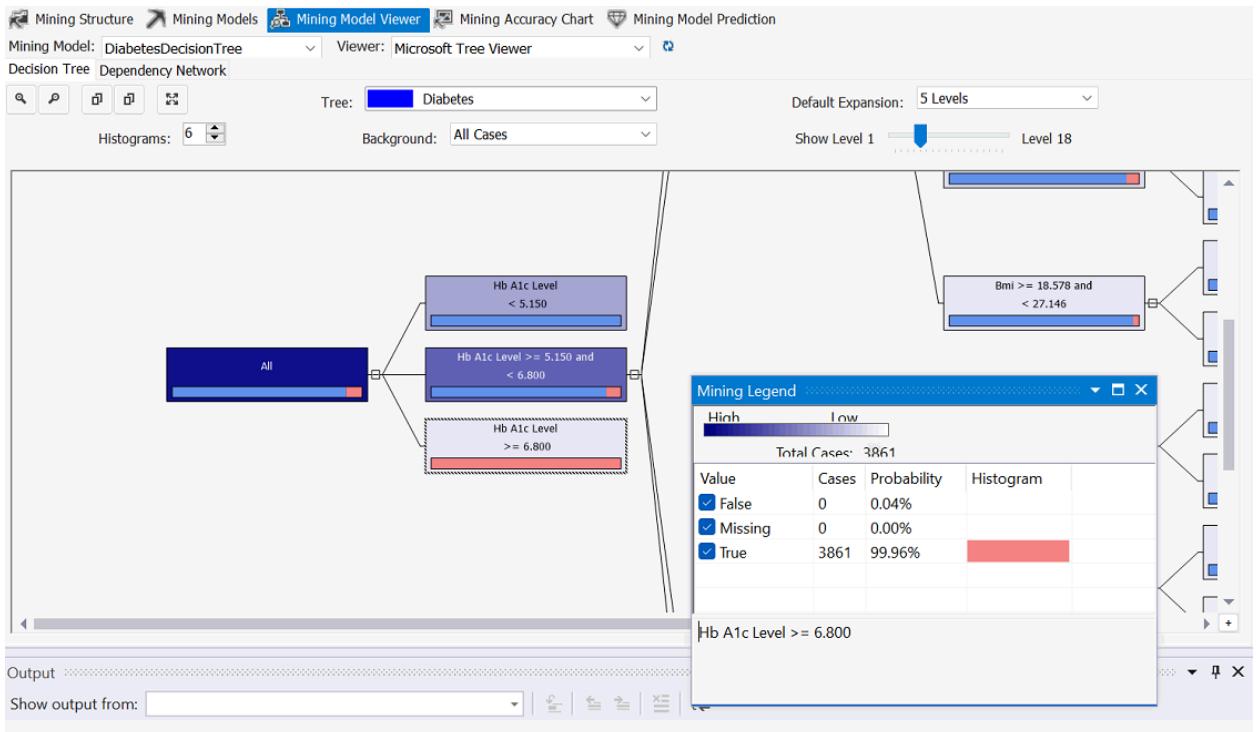
5.3.1. Thuật toán Decision Tree sử dụng Microsoft Decision Tree



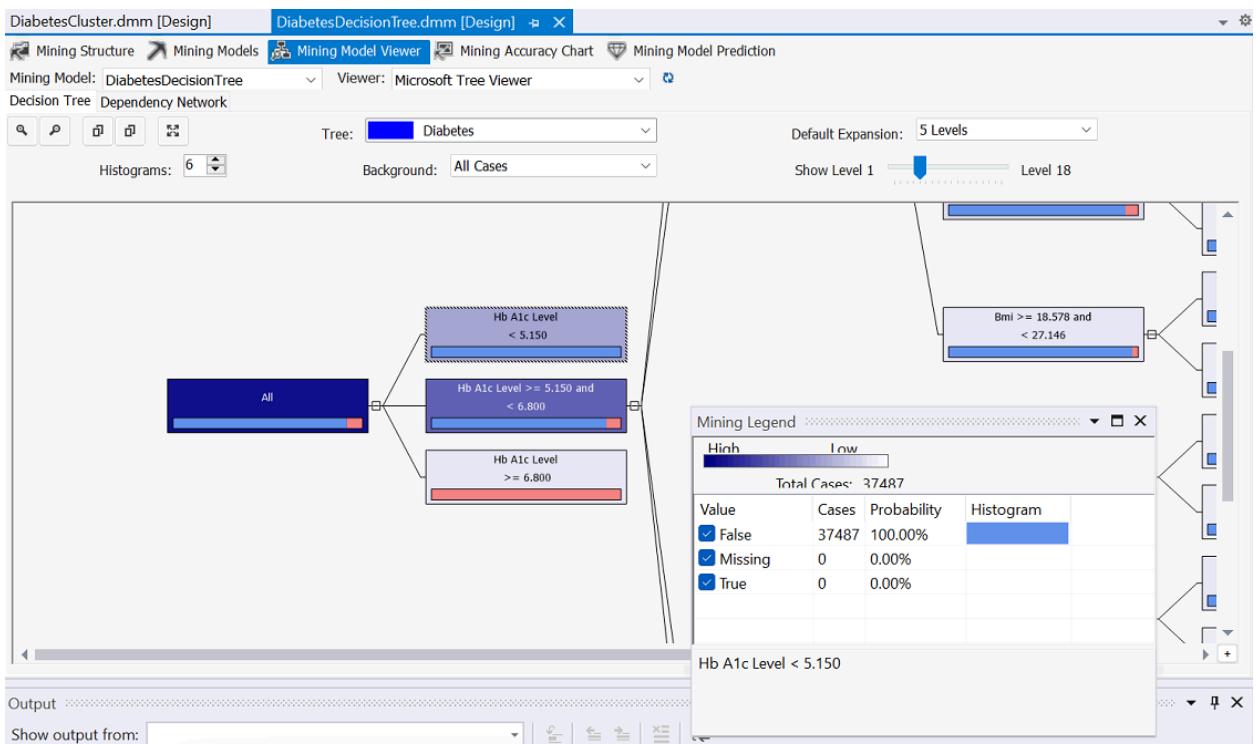
- Tỷ lệ bệnh nhân bị bệnh tiểu đường (1) là 8.52 %
- Tỷ lệ bệnh nhân không bị tiểu đường (0) là 91.48%



- Những bệnh nhân có lượng đường trung bình trong máu HbA1c ≥ 6800 có tỉ lệ 99.96% bị bệnh tiểu đường.

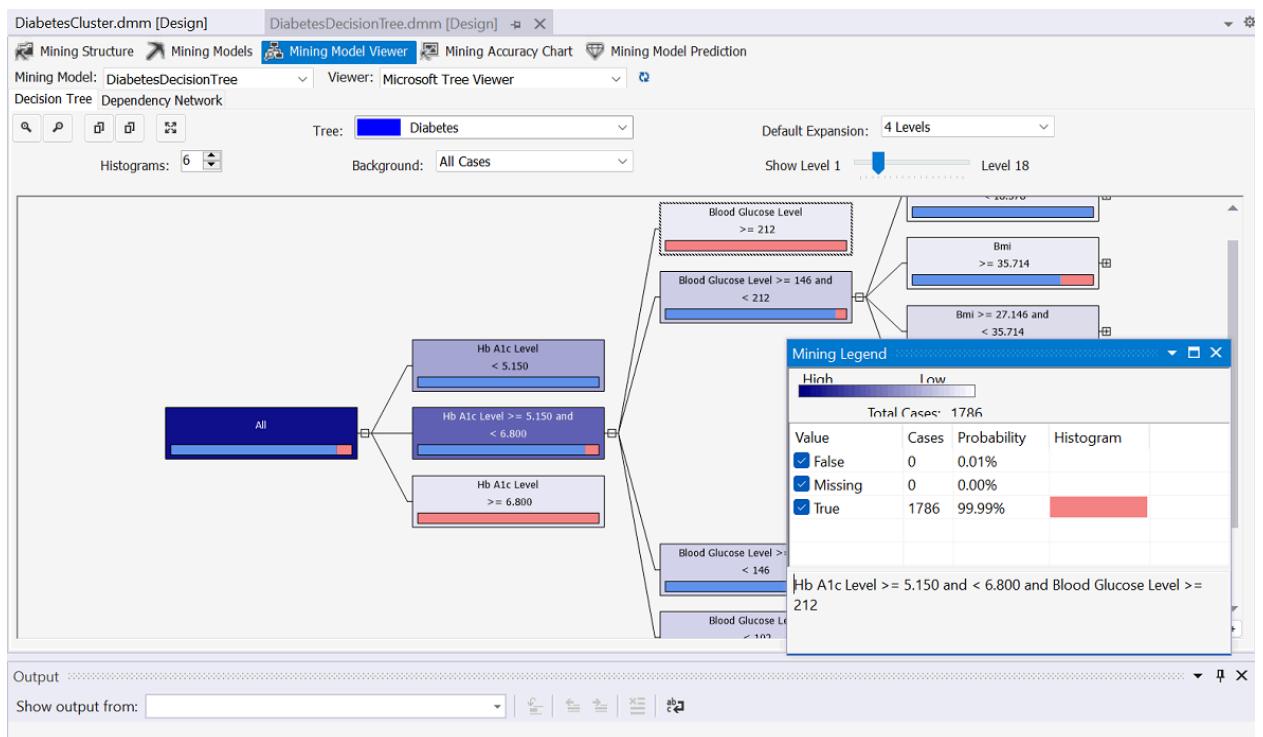


- Những bệnh nhân có lượng đường trung bình trong máu HbA1c < 5150 có tỉ lệ 100% **không bị** bệnh tiểu đường.

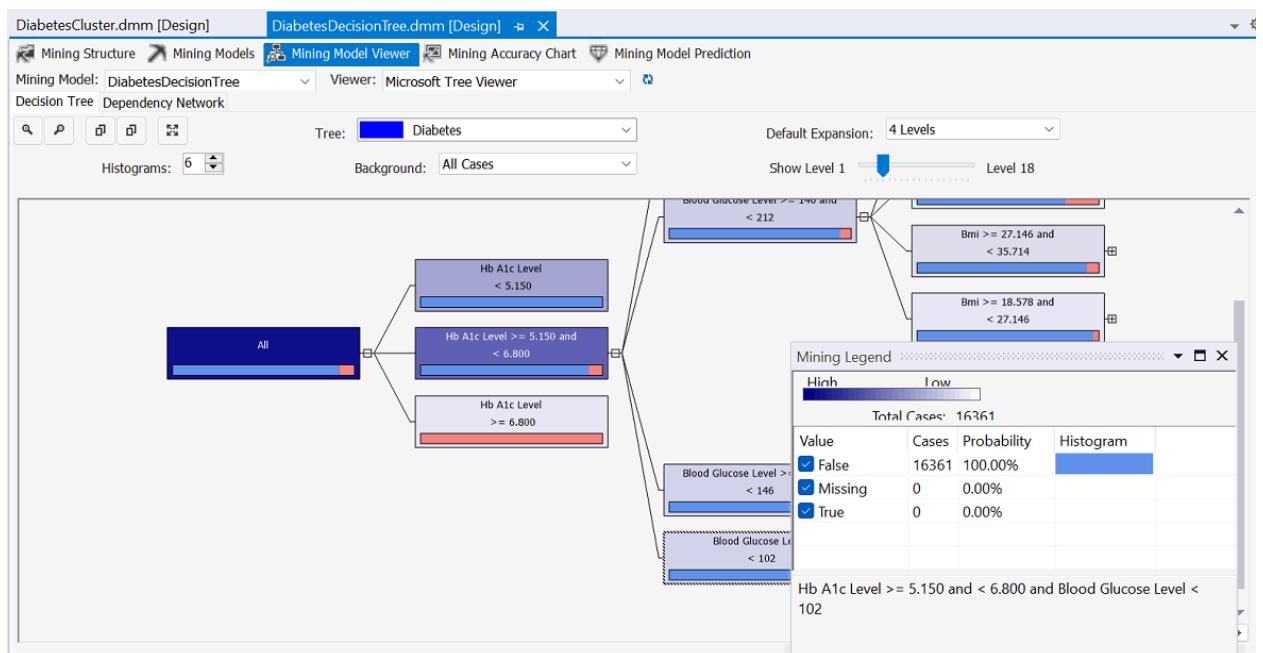


*Xét những bệnh nhân có $5150 \leq \text{HbA1c} < 6800$

- Những bệnh nhân có tỉ lệ bị bệnh tiểu đường 99.99% là những người có $5150 \leq \text{HbA1c} < 6800$ và mức đường huyết (Blood Glucose Level) rất cao, ≥ 212

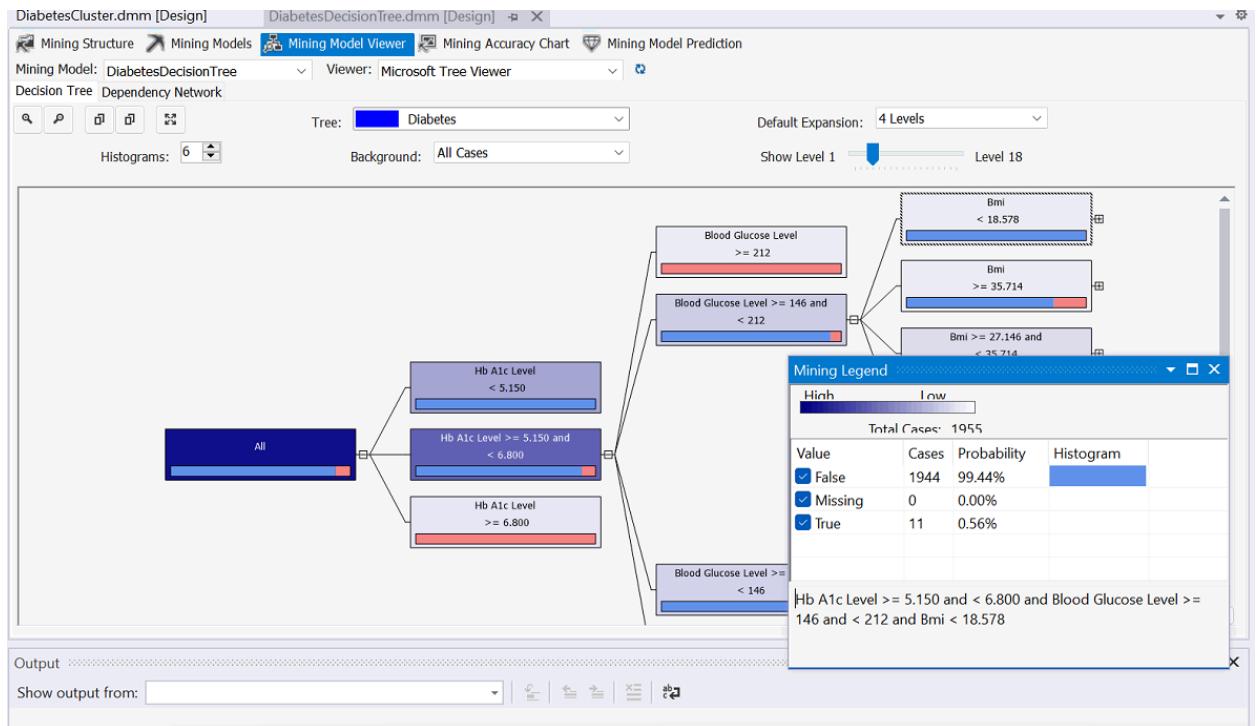


- Những bệnh nhân có tỉ lệ **không bị** bệnh tiểu đường 100% là những người có $5150 \leq \text{HbA1c} < 6800$ và mức đường huyết (Blood Glucose Level) < 102

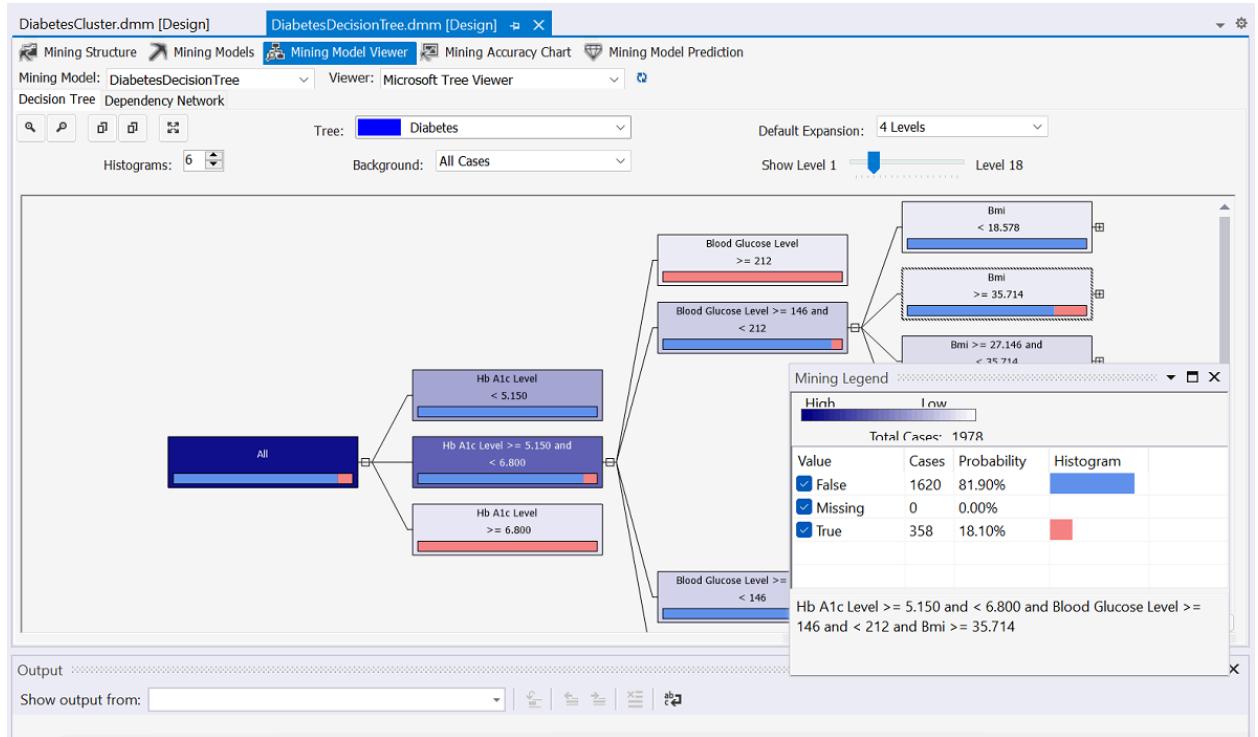


*Xét những bệnh nhân có $5150 \leq \text{HbA1c} < 6800$ và $146 \leq \text{Blood Glucose Level} < 212$

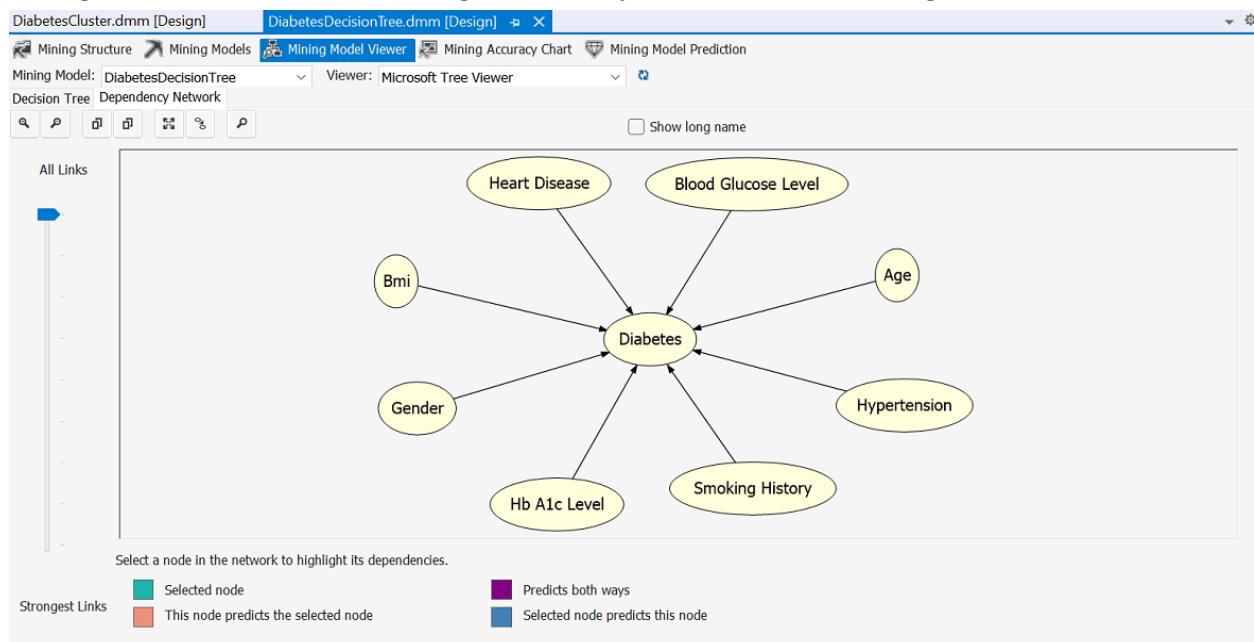
- Những bệnh nhân có tỉ lệ **không bị** bệnh tiểu đường 99.44% là những người có $5150 \leq \text{HbA1c} < 6800$ và mức đường huyết $146 \leq \text{Blood Glucose Level} < 212$, chỉ số BMI < 18.578 .



- Những bệnh nhân có tỉ lệ **không bị** bệnh tiểu đường 81.9% là những người có $5150 \leq \text{HbA1c} < 6800$ và mức đường huyết $146 \leq \text{Blood Glucose Level} < 212$, chỉ số BMI ≥ 35.714

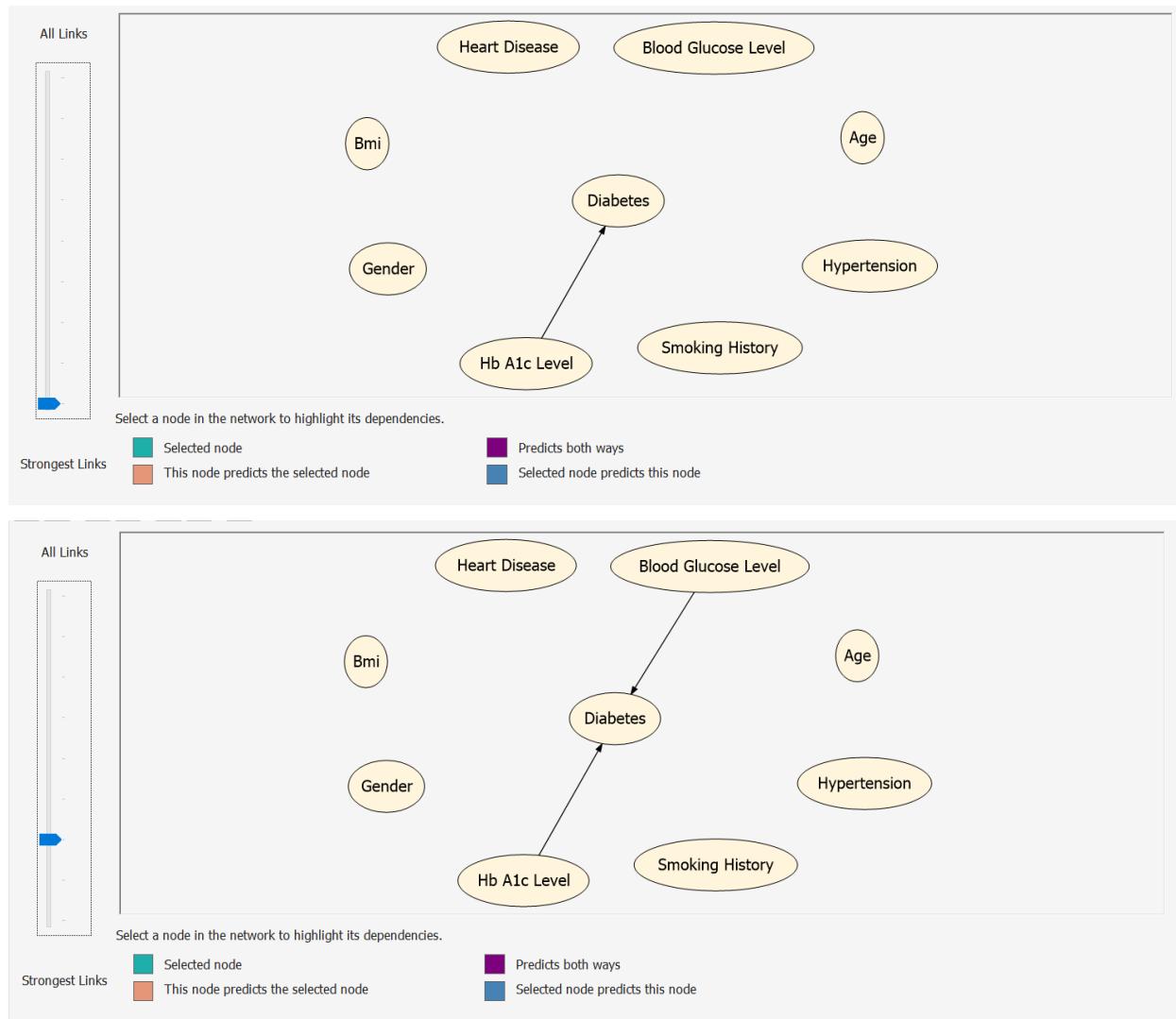


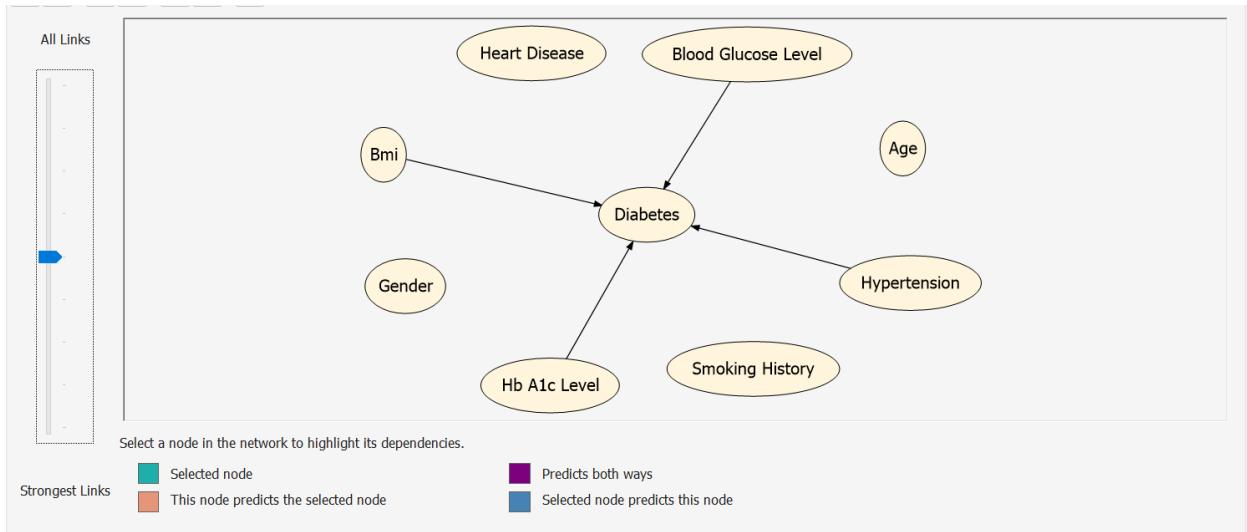
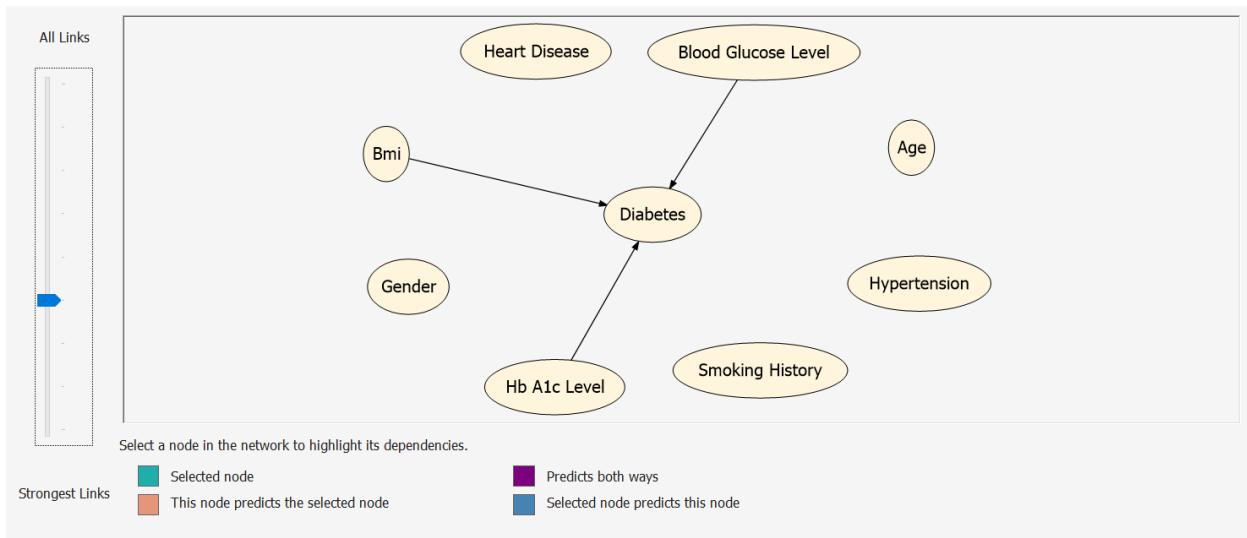
*Sử dụng **Dependency Network** của **Decision Tree** để kiểm tra xem thuộc tính nào ảnh hưởng nhiều nhất và ít ảnh hưởng nhất đến tỷ lệ bị bệnh tiểu đường **diabetes**.

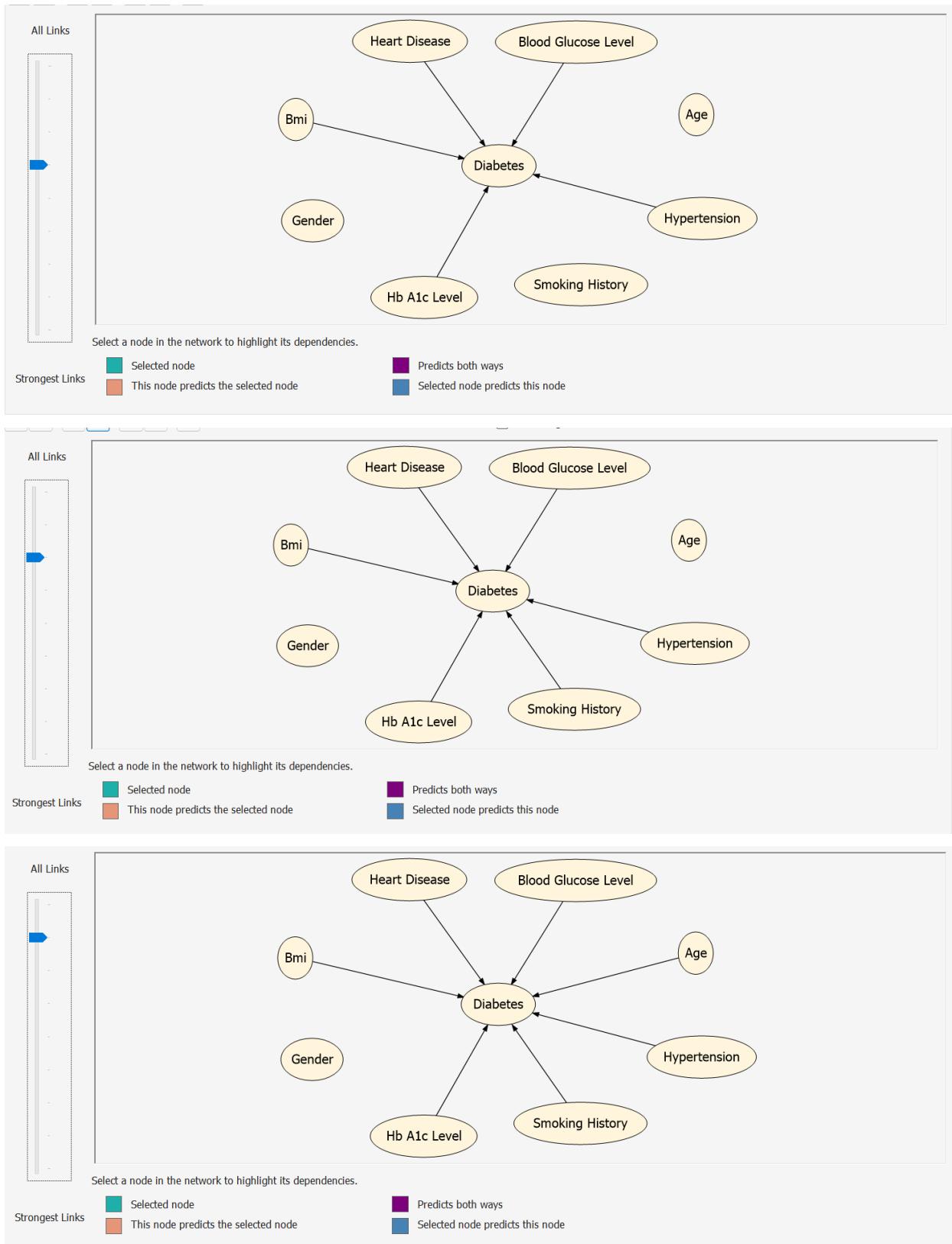


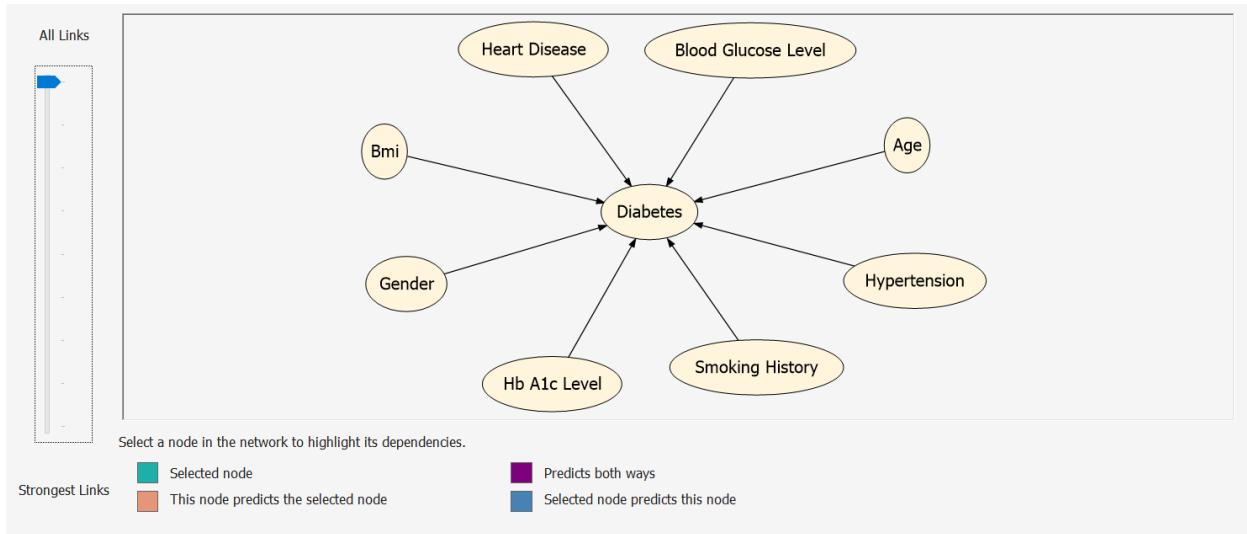
- Dependency Network của thuật toán Microsoft Decision Tree cho thấy những thuộc tính có tính ảnh hưởng tới việc bị bệnh tiểu đường.

- Những thuộc tính ảnh hưởng giảm dần là: Hb A1c Level, Blood Glucose Level, Bmi, Hypertension, Heart Disease, Smoking History, Age, Gender

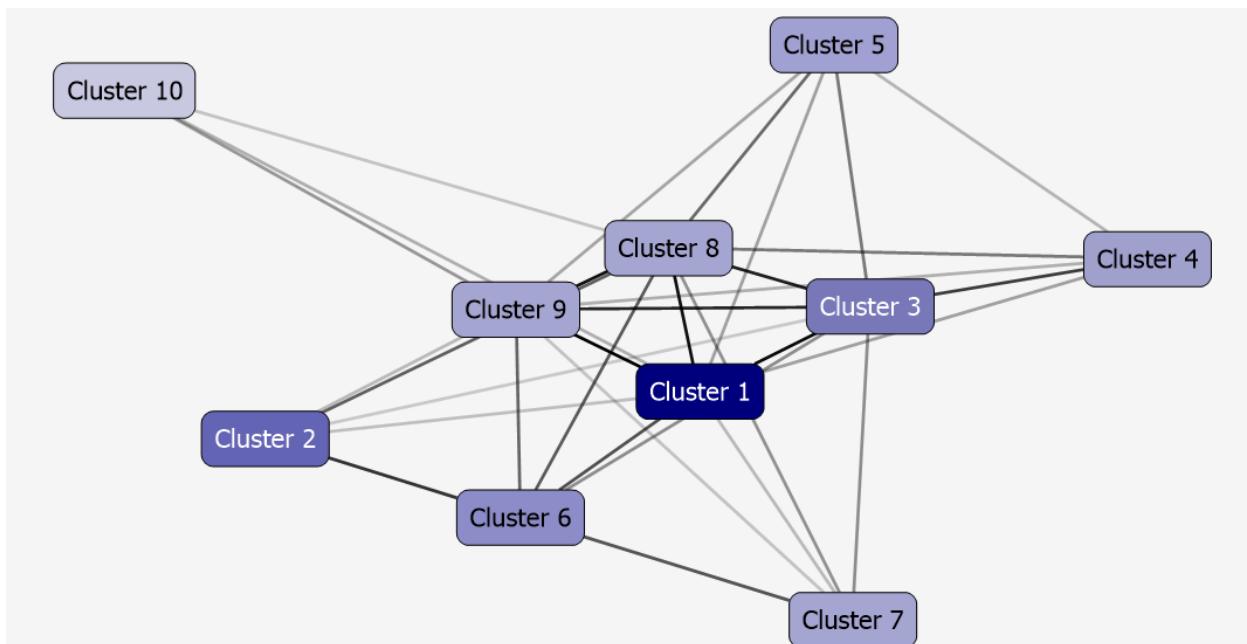




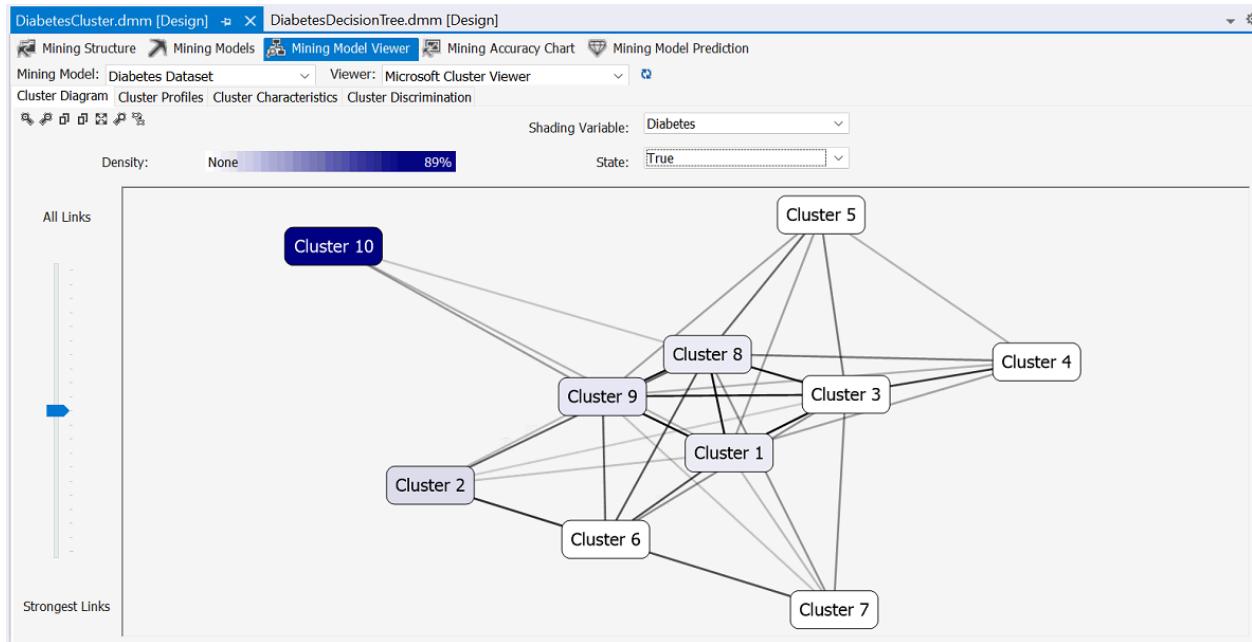




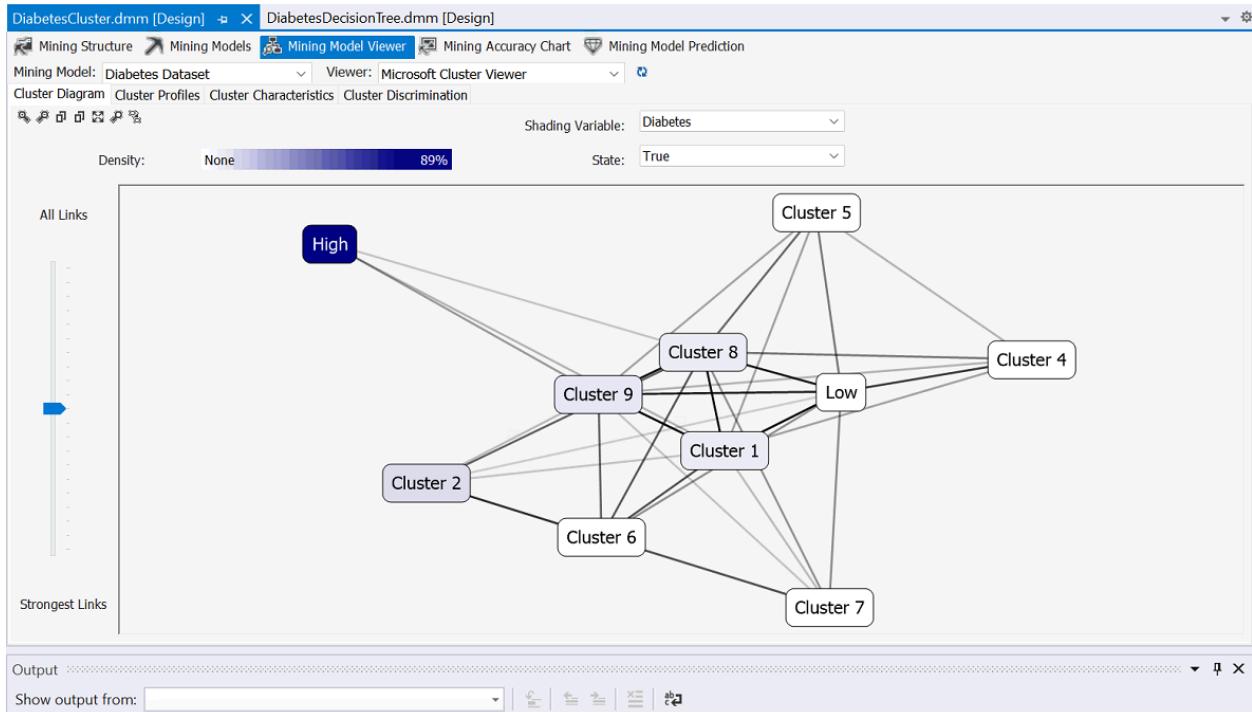
5.3.2. Thuật toán Clustering



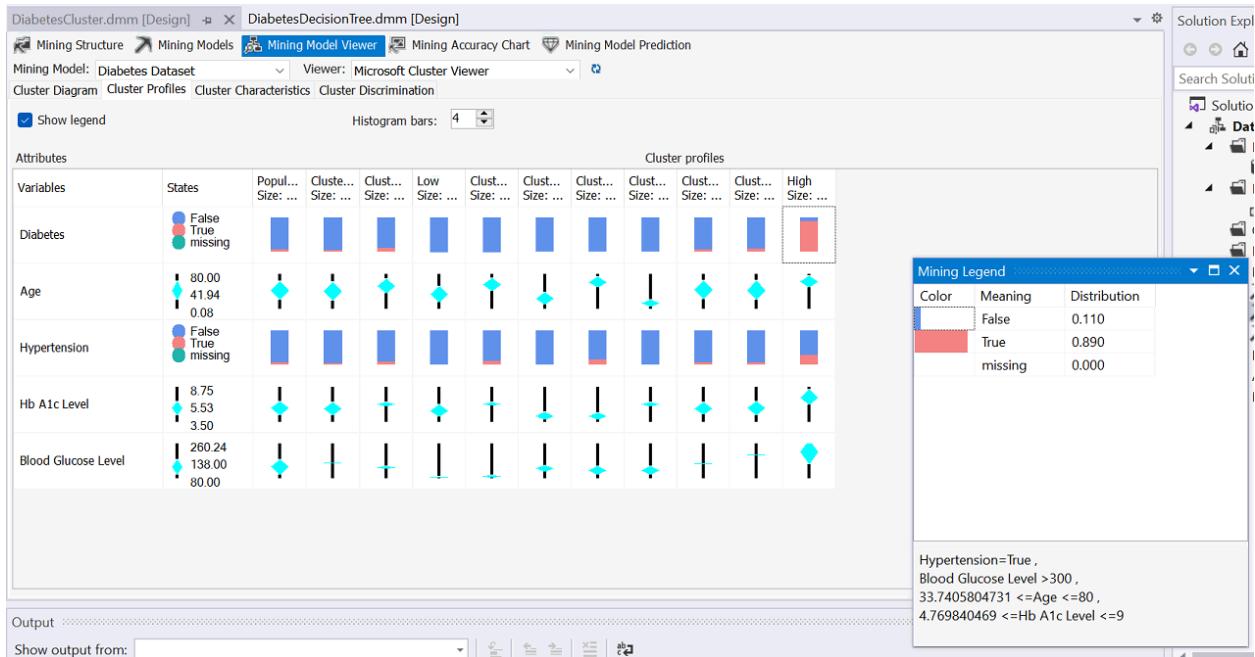
- Chọn Shading Variable và State



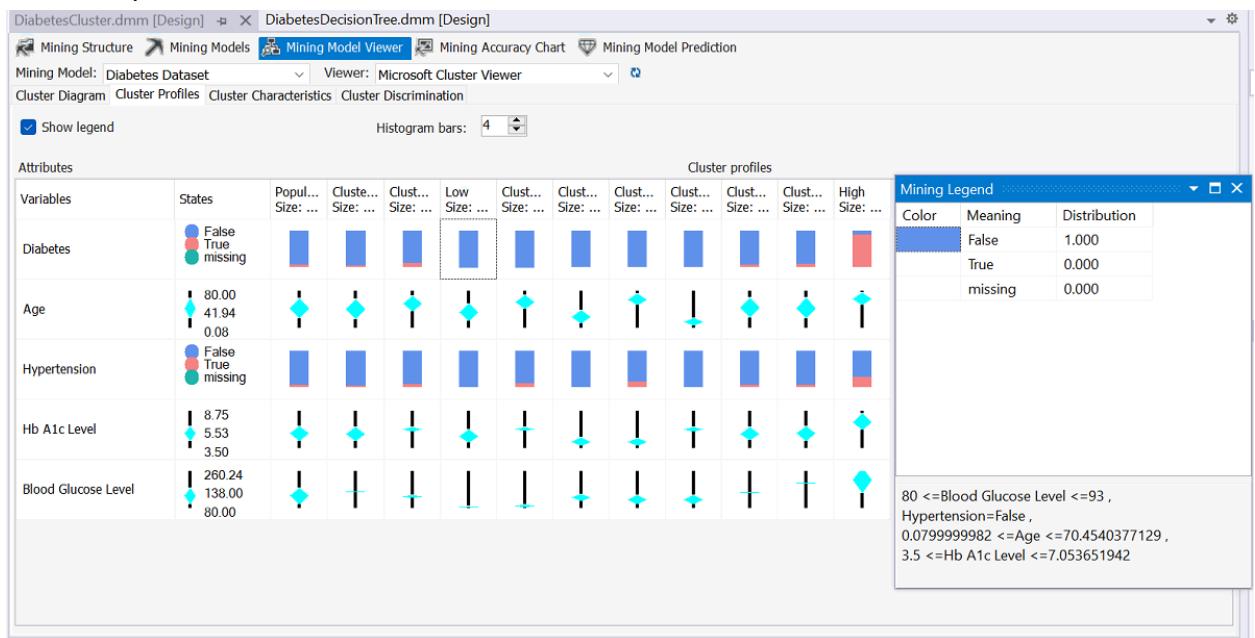
- Tiến hành đổi tên Cluster đậm nhất thành High, Cluster nhạt nhất thành Low



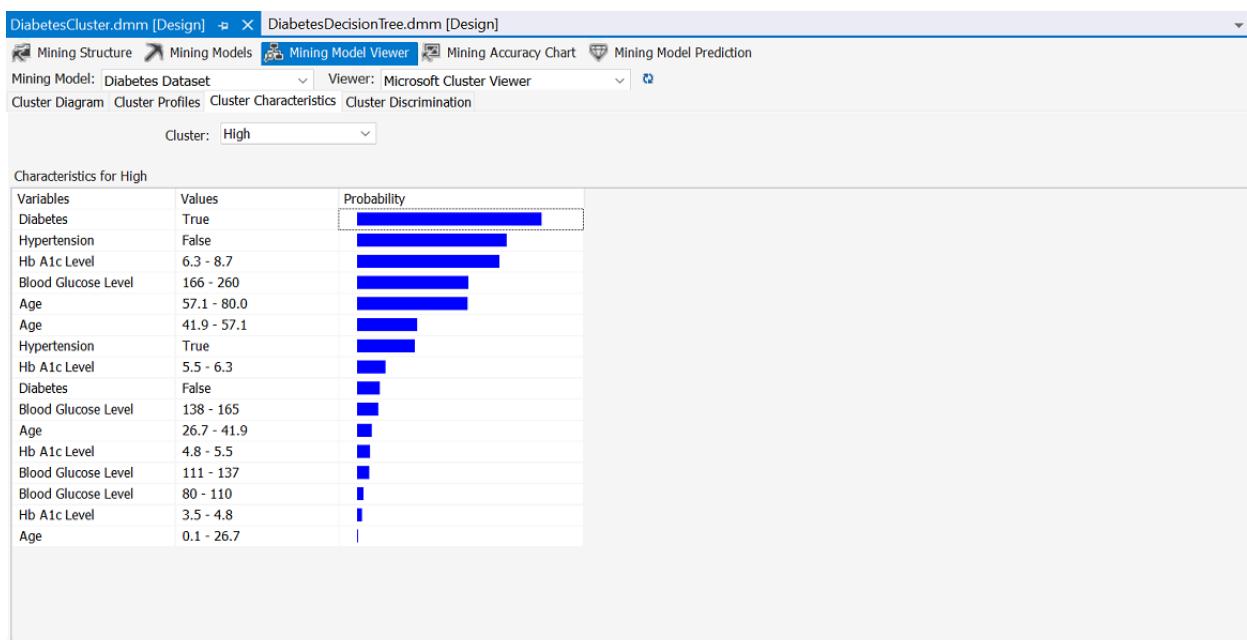
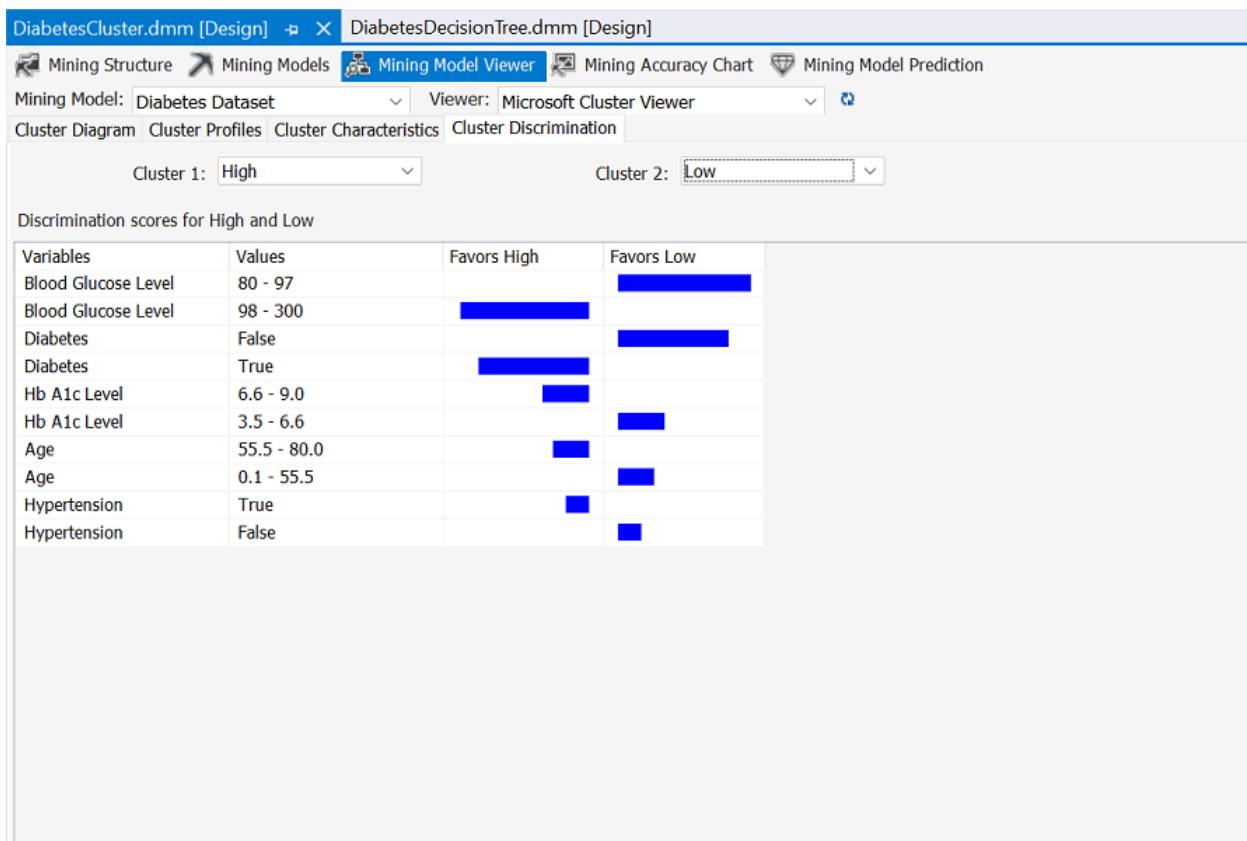
- Ta thấy Cluster High có 89% bệnh nhân mắc bệnh tiểu đường.
- Tập lục:



- Ta thấy Cluster Low có 100% bệnh nhân không bị tiểu đường
- Tập lục

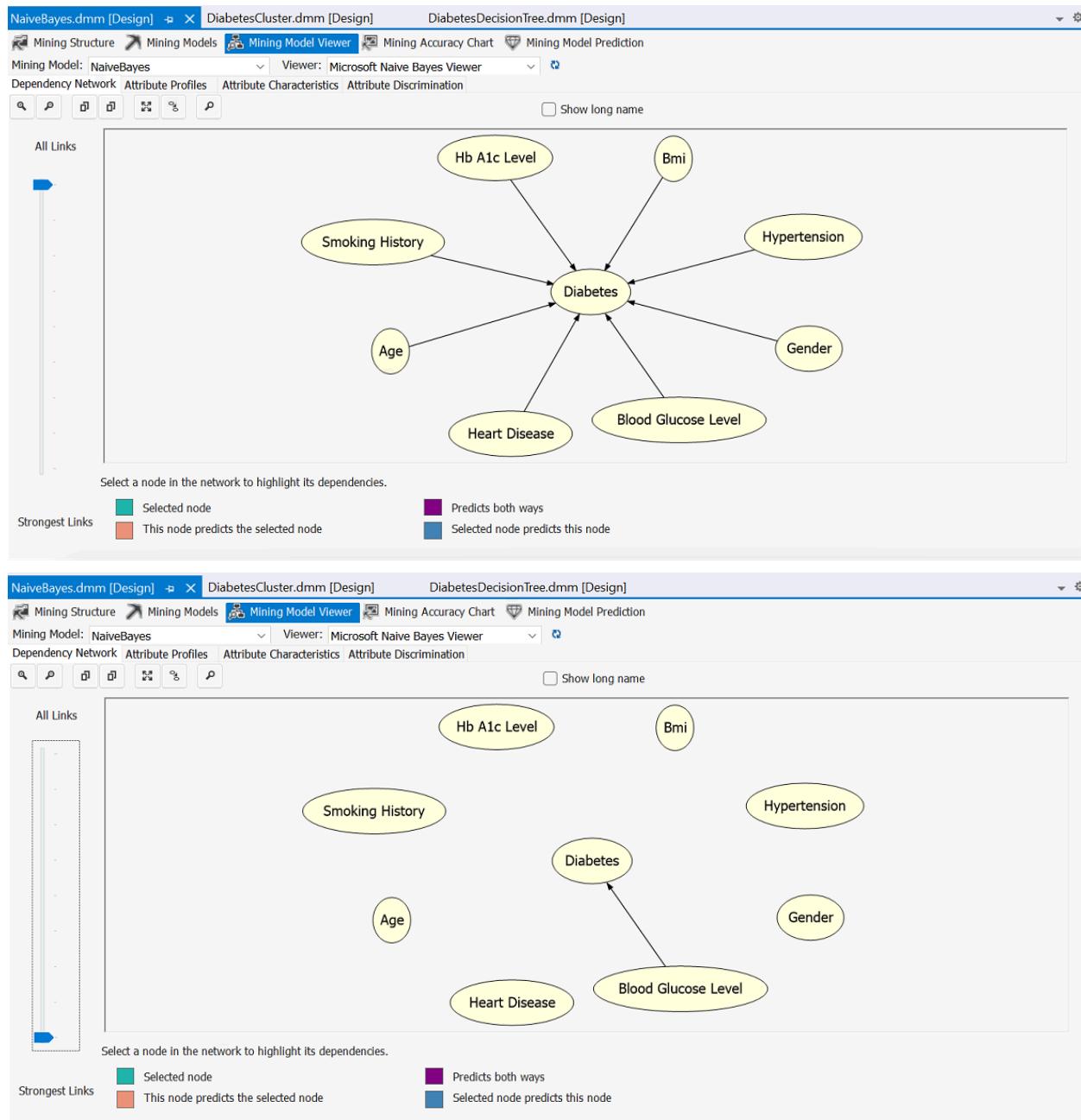


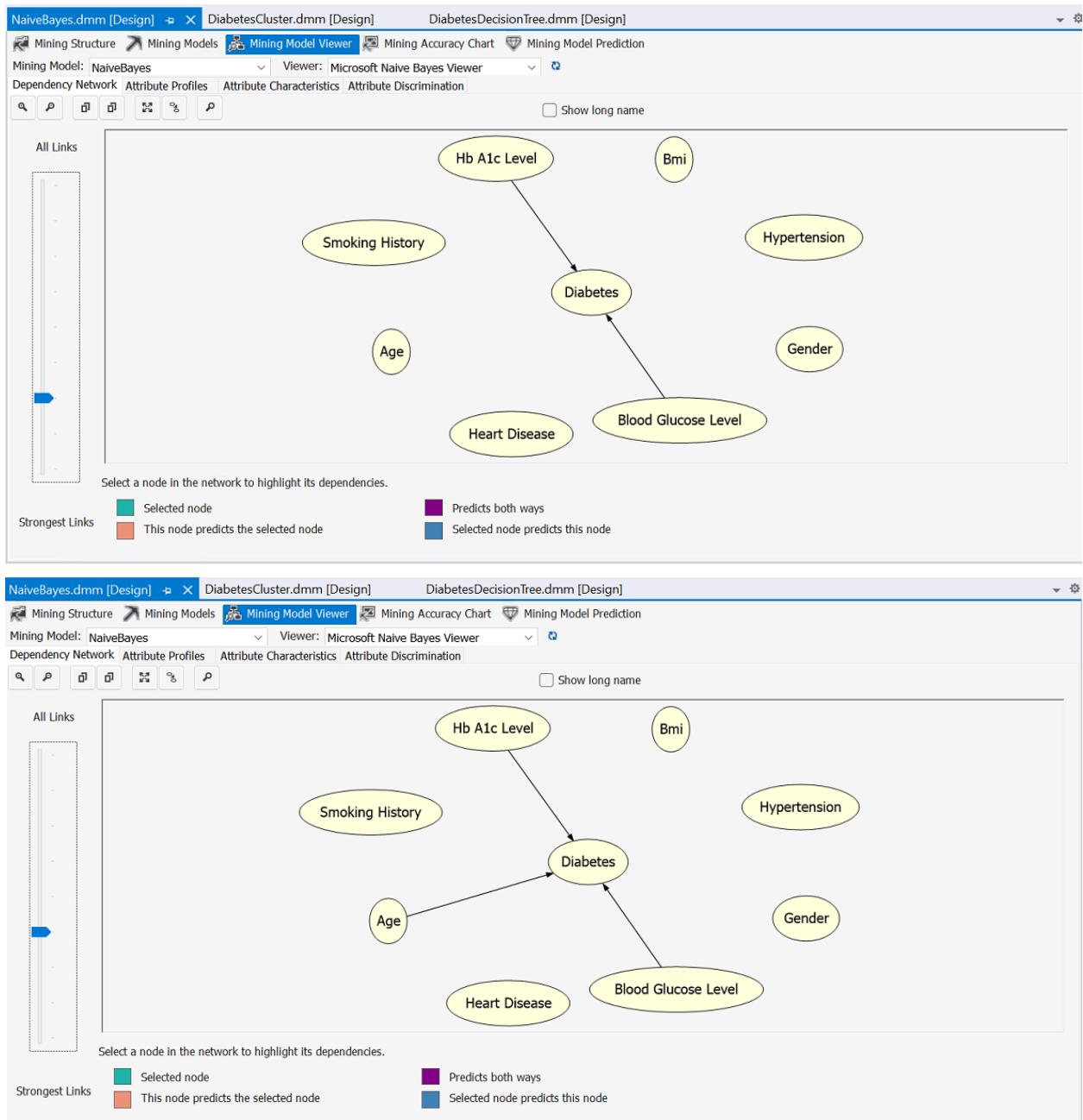
- Cluster Discrimination cho thấy sự tương quan giữa 2 cluster

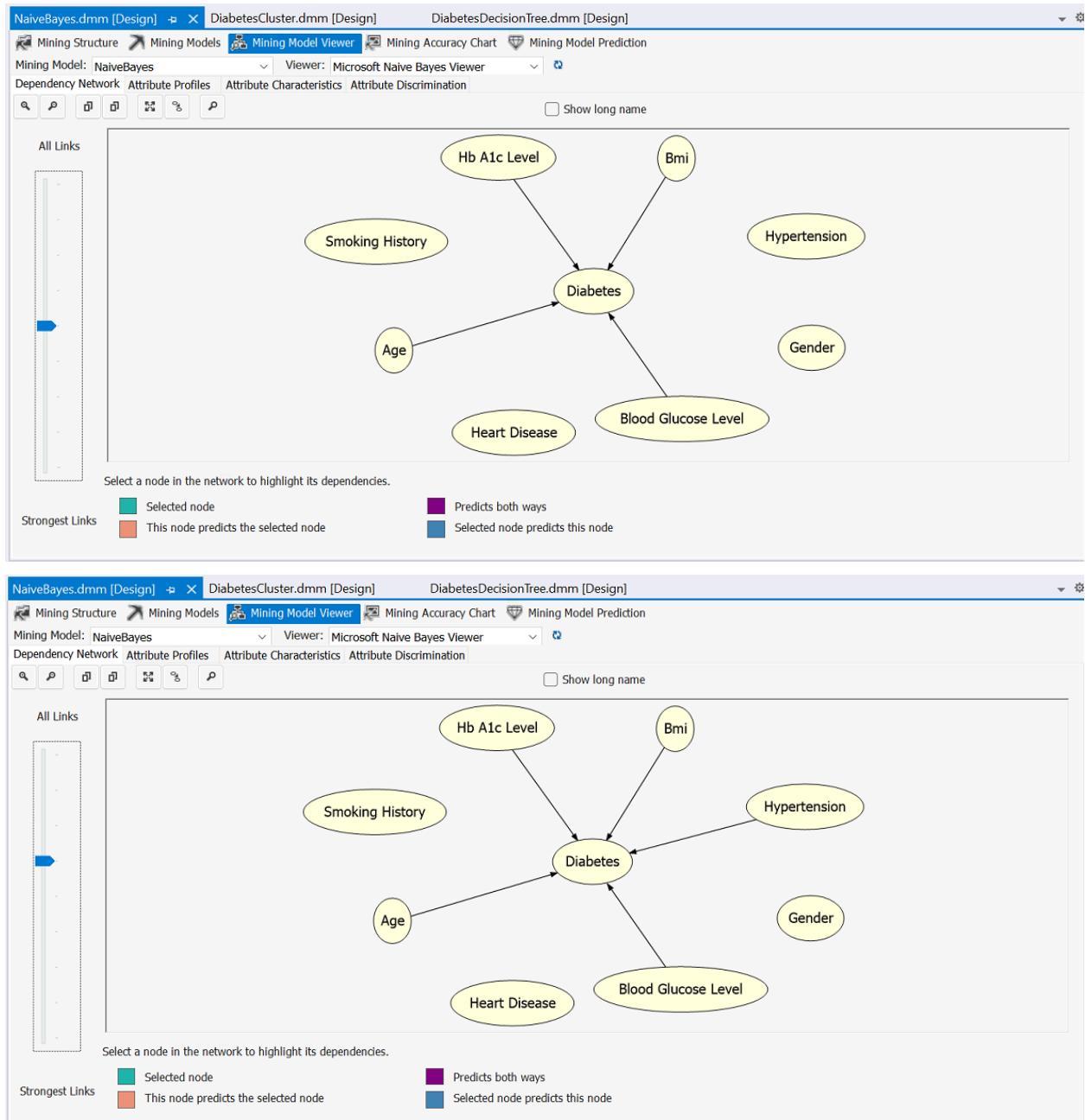


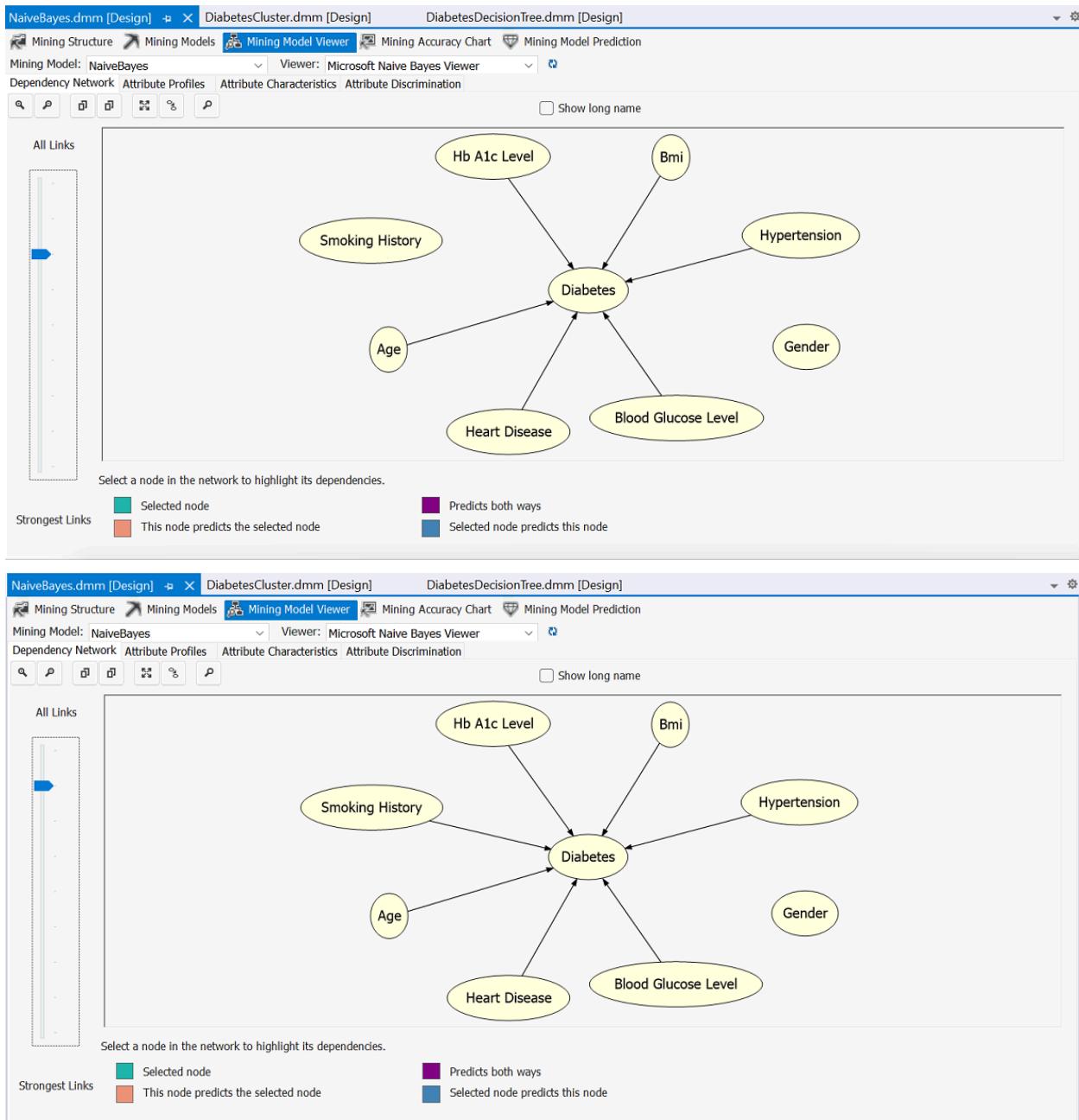
5.3.3. Thuật toán Microsoft Naive Bayes

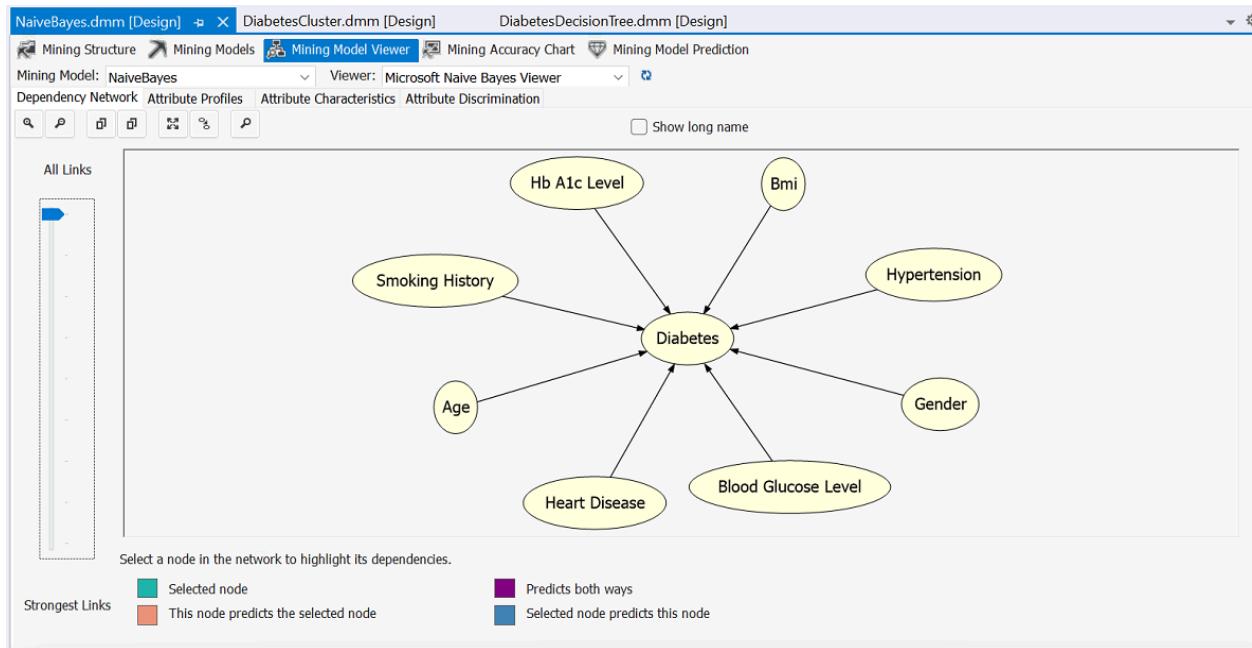
- Dependency Network của thuật toán Naive Bayes cho thấy những thuộc tính có tính ảnh hưởng tới việc bị bệnh tiểu đường (Diabetes)
- Những thuộc tính ảnh hưởng giảm dần là: Blood Glucose Level, Hb A1c Level, Age, Bmi, Hypertension, Heart Disease, Smoking History, Gender.





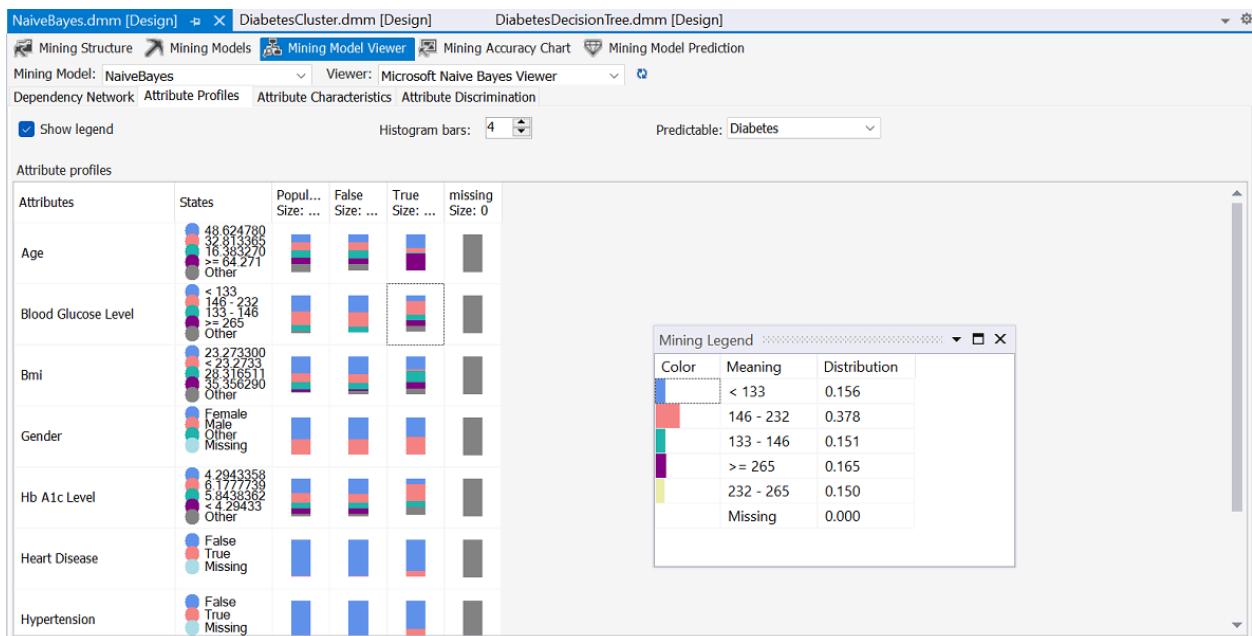






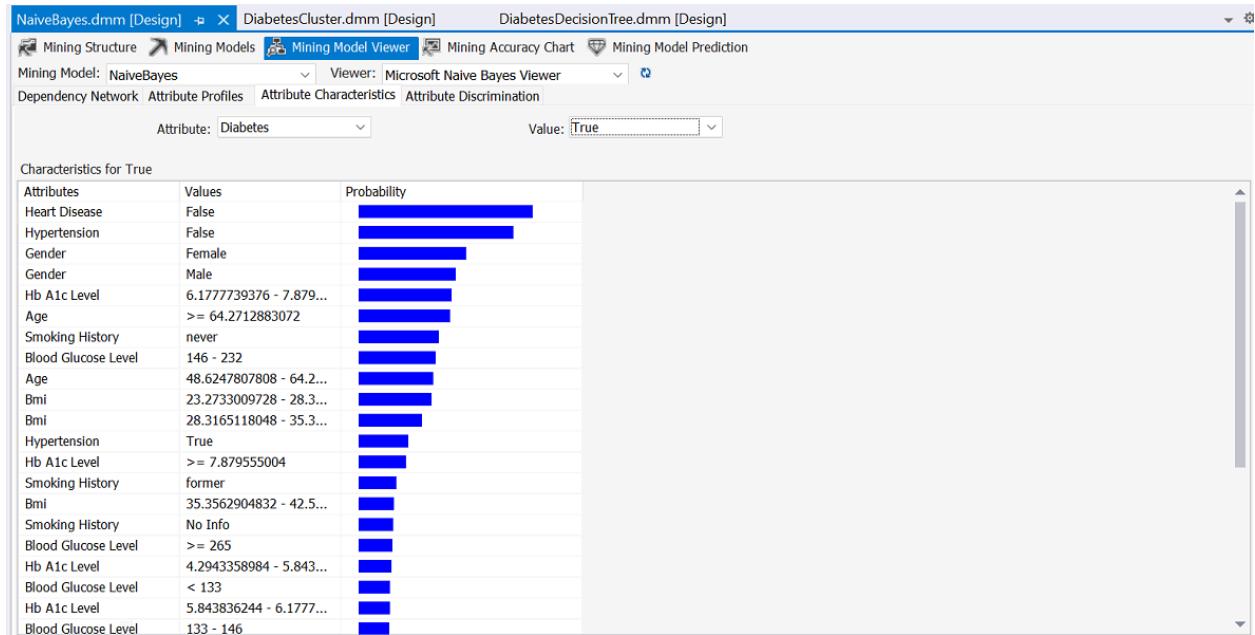
*Attribute Profiles

- Trong các bệnh nhân bị bệnh tiểu đường, có khoảng 37,8% bệnh nhân có Blood Glucose Level trong khoảng 146-232



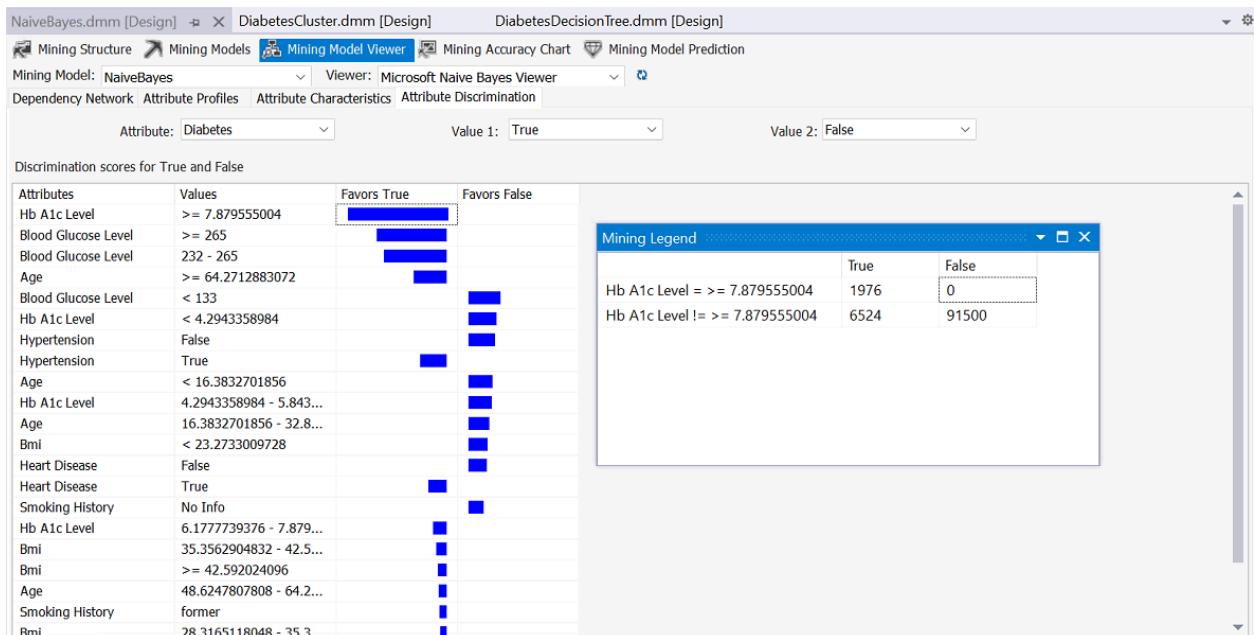
*Attribute Characteristics

- Hb A1c Level từ 6.1 đến 7.88 chiếm 45,4% các ca bị bệnh tiểu đường
- Nhóm tuổi ≥ 64 chiếm 44,48% các ca bệnh tiểu đường

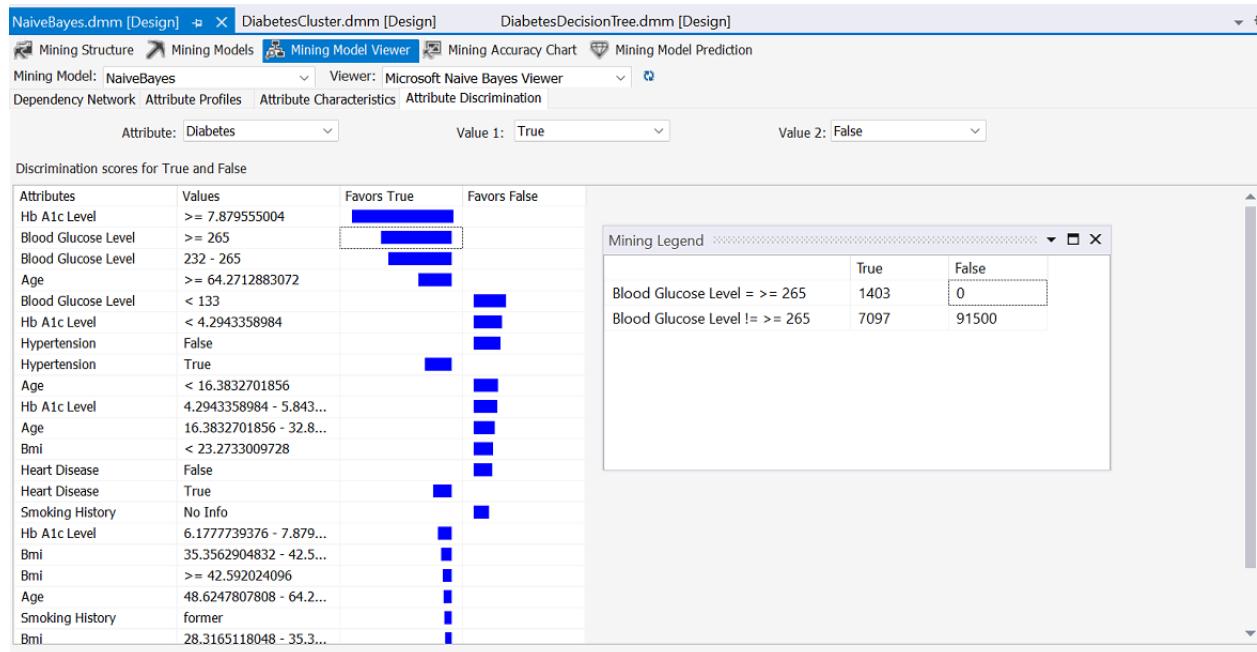


*Attribute Discrimination

- Những bệnh nhân có Hb A1c Level ≥ 7.88 hầu như sẽ mắc bệnh tiểu đường



- Những bệnh nhân có Blood Glucose Level ≤ 265 khả năng cao sẽ không mắc bệnh tiểu đường.



5.3.4. Đánh giá các thuật toán

- Xét lớp 1(True - bị tiểu đường)

Select predictable mining model columns to show in the lift chart:

Synchronize Prediction Columns and Values

Sh...	Mining Model	Predictable Column Name	Predict Value
<input checked="" type="checkbox"/>	NaiveBayes	Diabetes	True
<input checked="" type="checkbox"/>	Cluster	Diabetes	True
<input checked="" type="checkbox"/>	DecisionTree	Diabetes	True

*Theo line chart, độ chính xác của 3 thuật toán khi mining trên lớp 1 (True - bị bệnh tiểu đường):

- Decision Tree: 0.94
- Clustering: 0.90
- Naive Bayes: 0.94



- Xét trên lớp 0 (False: không bị tiểu đường)

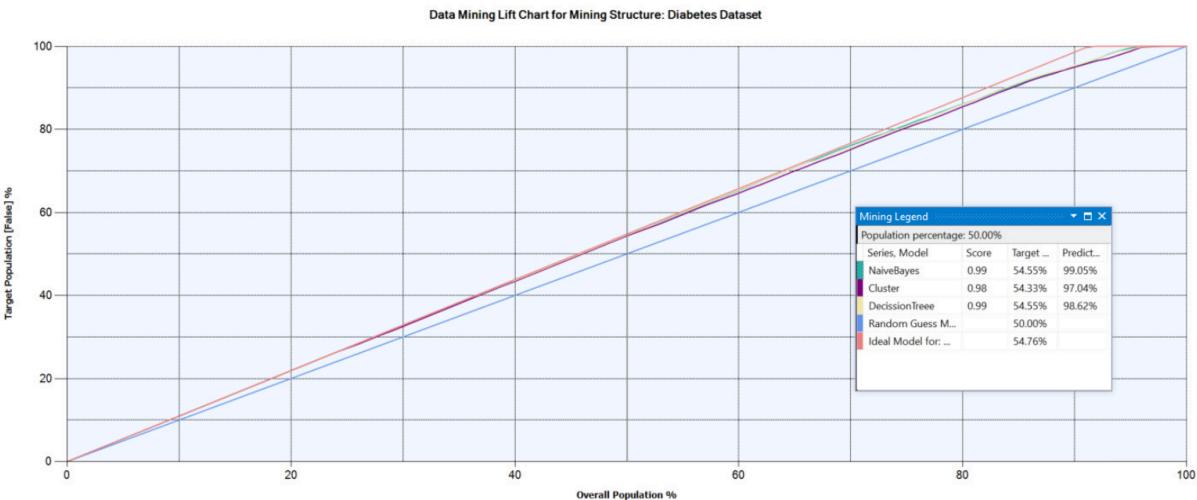
Select predictable mining model columns to show in the lift chart:

Synchronize Prediction Columns and Values

Show	Mining Model	Predictable Column Name	Predict Value
<input checked="" type="checkbox"/>	NaiveBayes	Diabetes	False
<input checked="" type="checkbox"/>	Cluster	Diabetes	False
<input checked="" type="checkbox"/>	DecisionTree	Diabetes	False

*Theo line chart, độ chính xác của 3 thuật toán khi mining trên lớp 1 (True - bị bệnh tiểu đường):

- Decision Tree: 0.99
- Clustering: 0.98
- Naive Bayes: 0.99



TÀI LIỆU THAM KHẢO

1. Ths. Đỗ Thị Minh Phụng (2024), Tài liệu lý thuyết và thực hành OLAP
[<https://courses.uit.edu.vn/course/view.php?id=12196>]
2. SQL Server 2012 Tutorials: Analysis Services - Multidimensional Modeling
[https://courses.uit.edu.vn/pluginfile.php/509409/mod_resource/content/1/2.SQL%20Server%202012%20Tutorials%20-%20Analysis%20Services%20Multidimensional%20Modeling%20%281%29.pdf]
3. SQL Server Integration Service - SSIS (03/01/2023)
[<https://learn.microsoft.com/en-us/sql/integration-services/lesson-1-1-creating-a-new-integration-services-project?view=sql-server-ver16>]
4. Analysis Services - SSAS
[<https://learn.microsoft.com/en-us/analysis-services/multidimensional-models/create-an-analysis-services-project-ssdt?view=asallproducts-allversions>]
5. Jin Do (21/03/2023), Hướng dẫn tạo project SSIS SSAS đơn giản với SQL server và Visual Studio 2022
[ [Hướng dẫn tạo project SSIS SSAS đơn giản với SQL server và Visual Studio 2022](#)]
6. Power BI Documentation (21-23/03/2024)
[<https://learn.microsoft.com/en-us/power-bi/>]