

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN



HCMUTE

**BÁO CÁO ĐỒ ÁN CUỐI KỲ
ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU CHO
HỆ THỐNG CHUỖI CUNG ỨNG DATACO**

Môn học: KHO DỮ LIỆU

Mã lớp học phần: DAWH430784_23_2_02

GVHD: ThS. Nguyễn Văn Thành

Nhóm sinh viên thực hiện: Nhóm 04

Ngô Thanh Thanh	21110643
Vũ Thị Bích Ngọc	21110905
Nguyễn Quốc Thịnh	21110661
Hoàng Anh Tuấn	21110710

TP. Hồ Chí Minh, 15 tháng 05 năm 2024

DANH SÁCH THÀNH VIÊN THAM GIA THỰC HIỆN ĐỀ TÀI VÀ VIẾT BÁO CÁO

Môn: Kho dữ liệu - HỌC KÌ II – NĂM HỌC 2023 – 2024

STT	HỌ VÀ TÊN	MSSV	TỶ LỆ ĐÓNG GÓP
1	Ngô Thanh Thanh	21110643	85%
2	Vũ Thị Bích Ngọc	21110905	85%
3	Nguyễn Quốc Thịnh	21110661	85%
4	Hoàng Anh Tuấn	21110710	80%

Nhận xét của giảng viên:

Ngày ... tháng 05 năm 2023

Giảng viên chấm điểm

Ths. Nguyễn Văn Thành

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành và sự tri ân sâu sắc đối với các thầy cô của Trường Đại học Sư phạm Kỹ thuật TP Hồ Chí Minh, đặc biệt là quý thầy cô bộ môn Hệ thống Thông tin của trường đã giúp cho chúng em trang bị các kiến thức cơ bản, các kỹ năng thực tế và tạo điều kiện để chúng em có thể hoàn thành đồ án môn học của mình.

Đặc biệt chúng em xin chân thành cảm ơn **thầy Nguyễn Văn Thành** đã nhiệt tình hướng dẫn, quan tâm truyền đạt những kiến thức và kinh nghiệm, trực tiếp hướng dẫn tận tình, sửa chữa và đóng góp ý kiến quý báu cho chúng em trong suốt thời gian học tập để chúng em có thể hoàn thành tốt môn học này.

Trong thời gian một học kỳ thực hiện đề tài, nhóm chúng em đã vận dụng những kiến thức nền tảng đã tích lũy đồng thời kết hợp với việc học hỏi và nghiên cứu những kiến thức mới. Từ đó, nhóm chúng em vận dụng tối đa những gì đã thu thập được để hoàn thành một báo cáo đồ án tốt nhất. Tuy nhiên, trong quá trình thực hiện, nhóm chúng em không tránh khỏi những thiếu sót. Chính vì vậy, nhóm chúng em rất mong nhận được những sự góp ý từ phía Thầy/Cô nhằm hoàn thiện những kiến thức mà nhóm chúng em đã học tập và là hành trang để nhóm tác giả thực hiện tiếp các đề tài khác trong tương lai.

Nhóm chúng em xin chân thành cảm ơn!

Nhóm thực hiện

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI	1
1.1. Lý do chọn đề tài.....	1
1.2. Tổng quan về tập dữ liệu	1
1.2.1. Nguồn dữ liệu	1
1.2.2. Mô tả chi tiết tập dữ liệu	3
1.3. Giới thiệu các công cụ được sử dụng.....	14
CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP (OLAP) ..	15
2.1. Xác định các Business Process và bảng Fact.....	15
2.1.1. Business Process: Sales Analysis.....	15
2.1.2. Business Process: Order FullFillment and Delivery	16
2.2. Xây dựng các bảng Dimension	16
2.2.1. DimDate	16
2.2.2. DimOrder	17
2.2.3. DimOrderDetail	17
2.2.4. DimCategory	17
2.2.5. DimCustomer.....	17
2.2.6. DimShipping	18
2.2.7. DimProduct.....	18
2.3. Star Schema (Lược đồ galaxy).....	18
CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS).....	19
3.1. Tạo ETL project và solution	19
3.2. Thiết lập các Connection	20
3.3. Import Dữ liệu vào các bảng dimension:	25
3.3.1.Date Dimension	25
3.3.1.1: Tạo Package	26
3.3.1.2: Thiết lập Control Flow.....	28
3.3.1.3: Load dữ liệu từ Source vào Stage	29
3.3.1.4: Thiết lập truncate table SQL task	43
3.3.1.5: Load dữ liệu từ Stage vào Dimension.....	45
3.3.2.Product Dimension.....	53
3.3.2.1. Tạo Package	53
3.3.2.2. Thiết lập Control Flow	53
3.3.2.3. Load dữ liệu từ Source vào Stage	54
3.3.2.4. Load dữ liệu từ Stage vào Dimension.....	58
3.3.3.Category Dimension	61
3.3.3.1. Load dữ liệu từ Source vào Stage	61
3.3.3.2. Load dữ liệu từ Stage vào Dimension.....	64
3.3.4.Order Dimension.....	66
3.3.4.1. Load dữ liệu từ Source vào Stage	66
3.3.4.2. Load dữ liệu từ Stage vào Dimension.....	69
3.3.5. Order Detail Dimension	73
3.3.5.1. Load dữ liệu từ Source vào Stage	73

3.3.5.2. Load dữ liệu từ Stage vào Dimension.....	76
3.3.6. Customer Dimension	78
3.3.6.1. Load dữ liệu từ Source vào Stage	78
3.3.6.2. Load dữ liệu từ Stage vào Dimension.....	82
3.3.7.Shipping Dimension.....	86
3.3.7.1. Load dữ liệu từ Source vào Stage	86
3.3.7.2. Load dữ liệu từ Stage vào Dimension.....	88
3.4. Import dữ liệu vào bảng các bảng fact.....	91
3.4.1. Fact Sales	91
3.4.1.1. Tạo Package	91
3.4.1.2. Thiết lập Control Flow	92
3.4.1.3. Load dữ liệu từ Source vào Stage	95
3.4.1.4. Load dữ liệu từ Stage vào Fact	98
3.4.2. Fact OrderFullfillment	103
3.4.2.1. Load dữ liệu từ Source vào Stage	103
3.4.2.4. Load dữ liệu từ Stage vào Fact	105
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU.....	112
4.1. Thực hiện phân tích dữ liệu bằng Pivot Table trong Excel	112
4.4.1. Câu hỏi: Cho biết sản phẩm có doanh thu cao nhất trong mỗi năm	112
4.4.2. Câu hỏi: Cho biết các sản phẩm có doanh thu cao nhất theo phân khúc khách hàng Corporate (Doanh nghiệp) trong 3 năm 2015,2016,2017.....	114
4.4.3. Câu hỏi: Cho biết doanh số thu được theo từng khách hàng để tìm ra khách hàng tiềm năng	116
CHƯƠNG 5: KẾT LUẬN	118
5.1. Kết quả đạt được	118
5.2. Những hạn chế	118
5.3. Tài liệu tham khảo	118

PHÂN CÔNG NHIỆM VỤ

Nhiệm vụ	Thanh Thanh	Bích Ngọc	Quốc Thịnh	Anh Tuấn
Tìm kiếm tập dữ liệu	x			
Xử lý tập dữ liệu			x	
Mô hình hóa dữ liệu vào workbook	x	x	x	x
+ DimDate + DimOrder + DimOrderDetail	x			
+ DimCustomer + DimProduct + DimCategory			x	x
+ DimShipping + FactOrderFulfillment + FactSales		x		
Thực hiện ETL đưa dữ liệu vào các bảng dim và fact dùng SSIS	x	x	x	x
+ Tạo và đưa dữ liệu vào các DimOrder, DimOrderDetail	x			x
+ Tạo và đưa dữ liệu vào DimProduct, DimCategory		x		
+ Tạo và đưa dữ liệu vào DimShipping, DimDate, DimCustomer			x	
Thực hiện phân tích dữ liệu	x	x	x	
Câu hỏi: Cho biết sản phẩm có doanh thu cao nhất trong mỗi năm			x	
Câu hỏi: Cho biết các sản phẩm có doanh thu cao nhất theo phân khúc khách hàng Corporate (Doanh nghiệp) trong 3 năm 2015,2016,2017	x			

Câu hỏi: Cho biết doanh số thu được theo từng khách hàng để tìm ra khách hàng tiềm năng		x		
---	--	---	--	--

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

1.1. Lý do chọn đề tài

Bên cạnh tốc độ phát triển vượt bậc về công nghệ hiện nay, nhiều lĩnh vực kinh tế được thúc đẩy mạnh, thị trường dần xuất hiện đa dạng các doanh nghiệp, cá nhân buôn bán với quy mô lớn nhỏ khác nhau. Cùng với đó, nhiều thử thách được đặt ra với các doanh nghiệp hoạt động trong lĩnh vực buôn bán và thương mại, đặc biệt nhất có thể kể tới là báo cáo, thống kê dữ liệu bán hàng để nắm bắt được kịp thời nhu cầu của khách hàng cũng như đưa ra những quyết định, chiến lược kinh doanh hiệu quả hơn.

Nhằm giúp doanh nghiệp tăng lợi nhuận một cách đáng kể, nhóm chúng em nhận thấy việc tổ chức xây dựng một hệ thống giúp quản lý và phân tích dữ liệu là vô cùng cần thiết. Từ hệ thống này, doanh nghiệp có thể đưa ra được cái nhìn tổng thể về kế hoạch kinh doanh, phân phối. Vì vậy nhóm sinh viên chúng em chọn đề tài "Xây dựng Kho dữ liệu cho hệ thống chuỗi cung ứng Dataco" với mong muốn tạo ra một Kho dữ liệu giúp công ty dễ dàng quản lý, thao tác và từ đó đưa ra những quyết định đúng đắn.

1.2. Tổng quan về tập dữ liệu

1.2.1. Nguồn dữ liệu

The screenshot shows the Kaggle website interface. On the left, there's a sidebar with navigation links: Create, Home, Competitions, Datasets (which is selected), Models, Code, Discussions, Learn, and More. The main content area displays a dataset titled "DataCo Supply Chain dataset" by "SAI CHARAN KOMATI - UPDATED 3 YEARS AGO". Below the title, it says "DataCo Market Analysis". There are tabs for Data Card, Code (2), Discussion (0), and Suggestions (0). A large blue decorative graphic of overlapping circles is on the right. At the top right, there are buttons for "New Notebook", "Download (19 MB)", and a menu icon. Below the main title, there's a section titled "About Dataset" with a description: "The dataset consists of roughly 180k transactions from supply chains used by the company DataCo Global". To the right of this, there are sections for "Usability" (rating 5.29) and "License".

- Bộ dữ liệu bao gồm khoảng 18 000 giao dịch từ chuỗi cung ứng được công ty DataCo Global sử dụng trong 3 năm
- Thông qua dữ liệu, người dùng có thể biết được thông tin của khách hàng, thông tin sản phẩm, thông tin hóa đơn, thông tin vận chuyển,...
- Dữ liệu bao gồm 53 cột và 18 000 dòng

Link dữ liệu gốc:

<https://www.kaggle.com/datasets/saicharankomati/dataco-supply-chain-dataset/data>

1.2.2. Mô tả chi tiết tập dữ liệu

Tập dữ liệu khi chưa qua tiền xử lý:

STT	Thuộc tính	Kiểu dữ liệu	Ý nghĩa	Giá trị
1	Type	object	Hình thức thanh toán	1. DEBIT (thẻ ghi nợ) 2. TRANSFER (chuyển khoản) 3. CASH (tiền mặt) 4. PAYMENT (thanh toán điện tử)
2	Days for shipping (real)	int64	Số ngày vận chuyển trong thực tế	Trong khoảng từ [0,6]
3	Days for shipment (scheduled)	int64	Số ngày vận chuyển dự kiến	Trong khoảng từ [0,4]
4	Benefit per order	Float64	Lợi nhuận trên mỗi hóa đơn	
5	Sales per customer	Float64	Doanh số trên mỗi khách hàng	
6	Delivery Status	object	Trạng thái giao hàng	1. advanced shipping (trước dự kiến) 2. late delivery (muộn) 3. shipping canceled (hủy) 4. shipping on time (đúng giờ)
7	Late_delivery_risk	Int64	Nguy cơ giao hàng muộn	1: giao hàng muộn

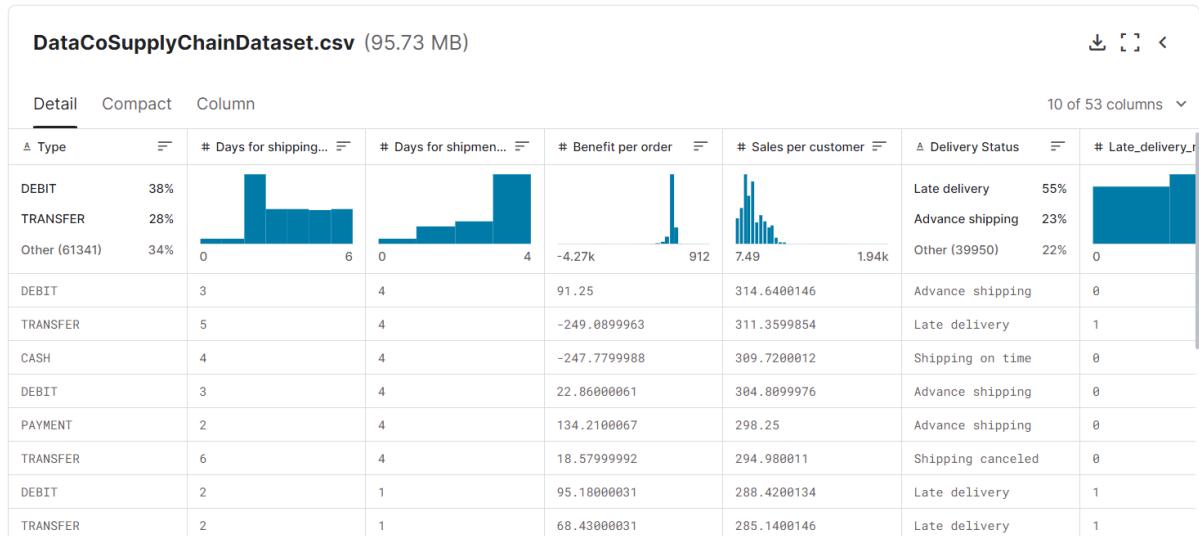
				0: Giao hàng đúng hẹn hoặc sớm hơn
8	Category Id	Int64	Mã loại sản phẩm	Trong khoảng [2,76]
9	Category Name	object	Tên loại sản phẩm	
10	Customer City	object	Thành phố khách hàng sinh sống	
11	Customer Country	object	Đất nước khách hàng sinh sống	1. EE.UU (Estados Unidos - United States) 2. Puerto Rico
12	Customer Email	object	Email của khách hàng	XXXXXXXXXX
13	Customer Fname	object	Tên (First name) của khách hàng	
14	Customer Id	Int64	Mã khách hàng	Trong khoảng [1, 50000]
15	Customer Lname	object	Họ của khách hàng (Last name)	
16	Customer Password	object	Mật khẩu của khách hàng	XXXXXXXXXX
17	Customer Segment	object	Phân khúc khách hàng	1. consumer (người tiêu dùng cá nhân) 2. corporate (doanh nghiệp) 3. home office (văn phòng)
18	Customer State	object	Tiểu bang khách hàng đang sinh sống	
19	Customer Street	object	Địa chỉ số nhà của khách hàng	

20	Customer Zipcode	Float64	Mã bưu điện của khách hàng	
21	Department Id	Int64	Mã phòng ban	Trong khoảng [2,12]
22	Department Name	object	Tên phòng ban	
23	Latitude	Float64	Vĩ độ của địa chỉ khách hàng	
24	Longitude	Float64	Kinh độ của địa chỉ khách hàng	
25	Market	object	Thị trường	1. Africa (Châu Phi) 2. Europe (Châu Âu) 3. LATAM (Châu Mỹ La Tinh) 4. Pacific Asia (Châu Á Thái Bình Dương) 5. USCA (Hoa Kỳ và Canada)
26	Order City	object	Thành phố đặt đơn hàng	
27	Order Country	object	Đất nước đặt đơn hàng	
28	Order Customer Id	Int64	Mã khách hàng đặt đơn hàng	Trong khoảng [1,50000]
29	order date (DateOrders)	object	Ngày đặt đơn hàng	Trong 4 năm từ 2015 đến 2018
30	Order Id	Int64	Mã đơn hàng	
31	Order Item Cardprod Id	Int64	Mã định danh duy nhất cho một mục sản phẩm cụ thể trong một đơn	Trong khoảng [19,1363]

			hàng (cũng là mã sản phẩm)	
32	Order Item Discount	Float64	Mức giảm giá cho một sản phẩm trong đơn hàng	
33	Order Item Discount Rate	Float64	Tỷ lệ phần trăm giảm giá được áp dụng cho một mục hàng cụ thể trong đơn hàng	Trong khoảng [0,0.25] Order Item Discount Rate = Order Item Discount / Order Item Product Price
34	Order Item Id	Int64	Mã định danh duy nhất cho mỗi mục hàng trong một đơn hàng	
35	Order Item Product Price	Float64	Giá bán ban đầu của một đơn vị sản phẩm trước khi áp dụng bất kỳ khoản giảm giá hay thuế nào	
36	Order Item Profit Ratio	Float64	Tỷ lệ phần trăm giữa lợi nhuận thu được từ một mục hàng và giá trị doanh số của mục hàng đó.	
37	Order Item Quantity	Int64	Số lượng của mỗi sản phẩm được mua trong đơn hàng	
38	Sales	Float64	Tổng đơn giá của từng mục sản phẩm khi đã nhân	Sales = Order Item Product

			số lượng, chưa tính ưu đãi	Price x Order Item Quantity
39	Order Item Total	Float64	Tổng đơn giá của từng mục sản phẩm khi đã tính ưu đãi	Order Item Total = Sales - Order Item Discount * Order Item Discount
40	Order Profit Per Order	Float64	Lợi nhuận trên mỗi mục sản phẩm trên đơn hàng	Order Profit Per Order = Order Item Total * Order Item Profit Ratio
41	Order Region	object	Khu vực đặt đơn hàng	
42	Order State	object	Tiểu bang đặt đơn hàng	
43	Order Status	object	Trạng thái đơn hàng	1. CANCELED (Hủy) 2. CLOSED (đơn hàng đã được xử lý hoàn toàn) 3. COMPLETE (đơn hàng đã được giao thành công) 4. ON_HOLD (Chờ xử lý) 5. PAYMENT_REVIE W (Đang được kiểm tra về thanh toán) 6.PENDING (chờ xác nhận đơn hàng)

				7.PENDING_PAYMENT (chờ thanh toán) 8. PROCESSING (đang được chuẩn bị để giao hàng) 9. SUSPECTED_FRAUD (có nghi ngờ về hoạt động gian lận)
44	Order Zipcode	Float64	Mã bưu chính của địa chỉ đặt đơn hàng	Có thể có hoặc không
45	Product Card Id	Int64	Mã sản phẩm	Trong khoảng [19,1363]
46	Product Category Id	Int64	Mã loại sản phẩm	Trong khoảng [2,76]
47	Product Description	Float64	Mô tả sản phẩm	
48	Product Image	object	Hình ảnh sản phẩm	Link hình ảnh
49	Product Name	object	Tên sản phẩm	
50	Product Price	Float64	Giá sản phẩm	
51	Product Status	Int64	Trạng thái sản phẩm	
52	shipping date (DateOrders)	object	Ngày đơn hàng giao cho đơn vị vẫn chuyển để bắt đầu vận chuyển	
53	Shipping Mode	object	Hình thức vận chuyển	



Tập dữ liệu khi chưa xử lý

Tập dữ liệu khi đã qua tiền xử lý:

Tên bảng	Mô tả	Số dòng
ORDER	lưu trữ thông tin về các đơn hàng	11288 dòng
ORDER_DETAIL	Lưu trữ thông tin về chi tiết các đơn hàng	17986 dòng
CATEGORY	lưu trữ thông tin phân loại sản phẩm	42 dòng
CUSTOMER	lưu trữ thông tin khách hàng	7603 dòng
SHIPPING	lưu trữ thông tin vận chuyển	11288 dòng
PRODUCT	lưu trữ thông tin sản phẩm	90 dòng

Bảng ORDER

Biến	Mô tả
Order Id	Mã đơn hàng (Khóa chính)
Type	Hình thức thanh toán
Market	Thị trường
Order City	Thành phố đặt đơn hàng
Order Country	Đất nước đặt đơn hàng
Order Customer Id	Mã khách hàng (Khóa ngoại trỏ đến CUSTOMER(CustomerId))
order date (DateOrders)	Ngày đặt hàng
Order Region	Khu vực đặt đơn hàng
Order State	Tiều bang đặt đơn hàng
Order Status	Trạng thái đơn hàng

Bảng ORDER_DETAIL

Biến	Mô tả
Order Id	Mã đơn hàng (Khóa chính)
Order Item Cardprod Id	Mã sản phẩm trên đơn hàng (Khóa chính) (Khóa ngoại - trỏ đến PRODUCT (Product Card Id))
Order Item Discount	Mức giảm giá trên một đơn vị sản phẩm
Order Item Discount Rate	Tỉ lệ phần trăm giảm giá được áp dụng cho sản phẩm
Order Item Id	Mã mục hàng trong một đơn hàng
Order Item Product Price	Giá gốc của một đơn vị sản phẩm
Order Item Profit Ratio	Tỉ lệ giữa lợi nhuận thu được từ một mục hàng và giá trị doanh số của mục hàng đó.
Order Item Quantity	Số lượng sản phẩm trên đơn hàng
Sales	Giá bán của sản phẩm đã tính số lượng
Order Item Total	Tổng đơn giá của từng mục sản phẩm khi đã tính ưu đãi
Order Profit Per Order	Lợi nhuận trên mỗi mục sản phẩm trên đơn hàng

Bảng CATEGORY

Biến	Mô tả
Category Id	Mã loại sản phẩm (Khóa chính)
Category Name	Tên loại sản phẩm

Bảng CUSTOMER

Biến	Mô tả
CustomerID	Mã khách hàng (Khóa chính): xác định và phân biệt các khách hàng với nhau

CustomerFName	Tên của khách hàng
CustomerLName	Họ của khách hàng
CustomerCity	Lưu trữ thông tin về thành phố của khách hàng đang sinh sống hoặc thực hiện giao dịch.
CustomerCountry	Lưu trữ thông tin về đất nước của khách hàng đang sinh sống hoặc thực hiện giao dịch.
CustomerSegment	Phân khúc của khách hàng
CustomerState	Lưu trữ thông tin về tiểu bang của khách hàng đang sinh sống hoặc thực hiện giao dịch.
CustomerStreet	Lưu trữ thông tin về địa chỉ chi tiết số nhà và đường của khách hàng đang sinh sống hoặc thực hiện giao dịch.
CustomerZipcode	Mã bưu điện của khách hàng.

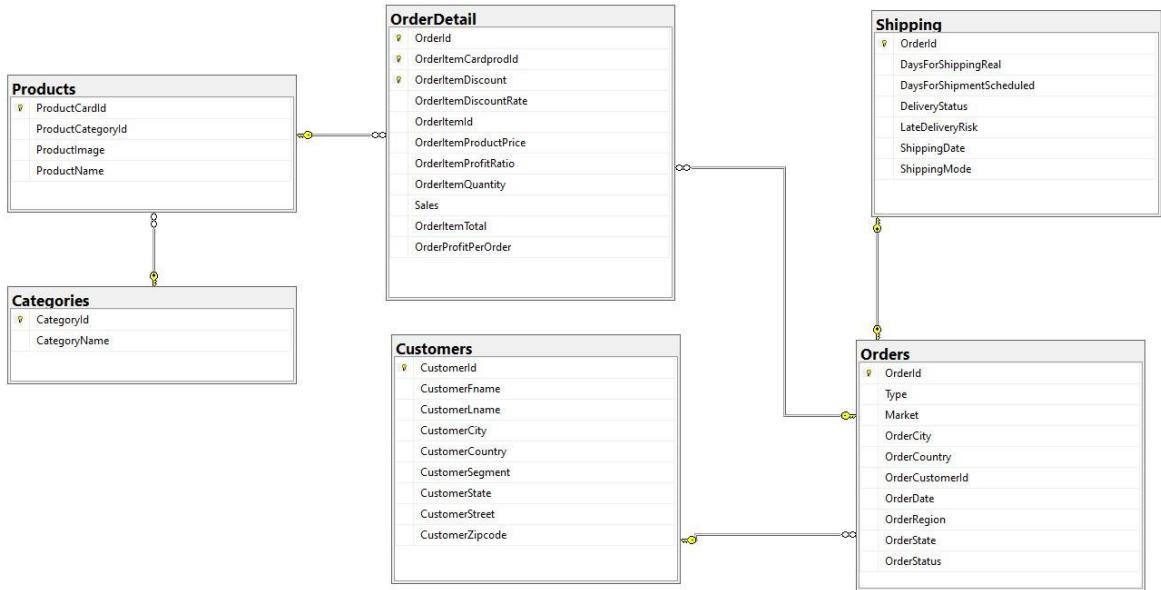
Bảng SHIPPING

Biến	Mô tả
OrderID	Mã đơn hàng (Khóa chính): xác định và phân biệt các đơn hàng với nhau (Khóa ngoại trỏ đến ORDER(Order Id))
DaysForShipping (real)	Số ngày đơn hàng được vận chuyển trong thực tế
DaysForShipment (scheduled)	Số ngày đơn hàng được vận chuyển trong dự kiến

DeliveryStatus	Trạng thái giao hàng của đơn hàng.
Late_delivery_risk	Nguy cơ đơn giao hàng sẽ giao hàng muộn của đơn hàng
ShippingDate (DateOrders)	Ngày đơn hàng được giao cho đơn vị vận chuyển
ShippingMode	Hình thức vận chuyển của đơn hàng

Bảng PRODUCT

Biến	Mô tả
ProductCardID	Mã sản phẩm (Khóa chính): xác định và phân biệt các sản phẩm với nhau
ProductCategoryID	Mã loại sản phẩm: xác định mã loại sản phẩm của từng sản phẩm (Khóa ngoại trỏ đến CATEGORY (CategoryId))
ProductImage	Chứa thông tin chi tiết về hình ảnh của sản phẩm
ProductName	Chứa thông tin về tên gọi của sản phẩm.



Lược đồ CSDL quan hệ trên SQL

1.3. Giới thiệu các công cụ được sử dụng

- Visual Studio 2019/2022 tích hợp SSIS (SQL Server Integration Services) và SSAS (SQL Server Analysis Services).
- SQL Server Management Studio 2019

CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP (OLAP)

2.1. Xác định các Business Process và bảng Fact

Business requirement:

- Phân tích doanh số (Sales Analysis): người quản lý muốn theo dõi doanh số bán hàng theo khách hàng (customer), sản phẩm (product) và và thị trường để có thể biết được sản phẩm nào bán chạy, mang lại nhiều lợi nhuận nhất, thị trường nào tiềm năng nhất, đặc điểm nhu cầu mua sắm của các phân khúc khách hàng để đưa ra các chiến lược kinh doanh, marketing hiệu quả.
- Việc đặt và giao hàng (Order Fulfillment and Delivery): người quản lý muốn phân tích quá trình đặt hàng để biết được khi nào đơn hàng có xu hướng giao chậm hay đúng hẹn để cải thiện việc giao hàng.

Xây dựng Detail Bus Matrix

Business Process Name	Fact Table	Fact Grain (the grain of the fact table)	Granularity	Facts	DimOrder	DimOrderDetail	DimCategory	DimCustomer	DimShipping	DimProduct	DimDate
Sale Analysis	FactSale	Transaction	one row per order detail	Order Item Cardprod Id, Order Id, UnitPrice, Order Item Discount, Quantity, Order Item Total, Order Profit Per Order	x	x	x	x	x	x	
Order Fulfillment and Delivery	FactOrderFulfillment	Accumulating Snapshot	one row per orders	Order Id, Order Date, Shipping Date, Delivery Status, Shipping mode	x			x		x	

2.1.1. Business Process: Sales Analysis

Các câu hỏi cụ thể được đặt ra:

- + Cho biết danh sách các sản phẩm thu lại lợi nhuận cao nhất trong tháng/quý/năm
- + Cho biết thị trường (market), khu vực (region) tiềm năng, mang lại doanh thu cao cho hệ thống chuỗi bán lẻ
- + Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất trong tháng/quý/ năm.
- + Cho biết doanh số bán hàng theo từng khách hàng trong tháng/ quý/ năm.

Bảng FactSales

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description
	FactSales	DateKey	ForeignKey to DimDate (DateKey)
	FactSales	CustomerKey	ForeignKey to DimCustomer (CustomerKey)
	FactSales	OrderKey	ForeignKey to DimOrder (OrderKey)
	FactSales	OrderDetailKey	ForeignKey to DimOrderDetail (OrderDetailKey)
	FactSales	CategoryKey	ForeignKey to DimCategory (CategoryKey)
	FactSales	ShippingKey	ForeignKey to DimShipping (ShippingKey)
	FactSales	ProductKey	ForeignKey to DimProduct (ProductKey)
	FactSales	UnitPrice	Unit price of each product
	FactSales	Quantity	Quantity of each product

2.1.2. Business Process: Order FullFfillment and Delivery

Các câu hỏi cụ thể được đặt ra:

+ Cho biết danh sách các địa điểm giao hàng có rủi ro giao trễ nhiều nhất

Bảng FactOrderFullfillment

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description
	FactOrderFullfillment	DateKey	ForeignKey to DimDate (DateKey)
	FactOrderFullfillment	OrderKey	ForeignKey to DimOrder (OrderKey)
	FactOrderFullfillment	ShippingKey	ForeignKey to DimShipping (ShippingKey)
	FactOrderFullfillment	DeliveryStatus	Order delivery status

2.2. Xây dựng các bảng Dimension

2.2.1. DimDate

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description
	DimDate	DateKey	A unique identifier for each date in the table
	DimDate	FullDate	Full date as a SQL date
	DimDate	DayNumberOfWeek	Number of the day of week
	DimDate	DayNameOfWeek	Day name of week
	DimDate	DayNumberOfMonth	Number of the day in the month
	DimDate	DayNumberOfYear	Number of the day in the year
	DimDate	WeekNumberOfYear	Week of year
	DimDate	MonthName	Month name of year
	DimDate	MonthNumberOfYear	Month of year
	DimDate	CalendarQuarter	Calendar quarter
	DimDate	CalendarYear	Year

2.2.2. DimOrder

Instructions!	Attribute / Fact Name	Description
Dimension / Fact Table		
DimOrder	OrderKey	A unique identifier for each order in the table
DimOrder	Order Id	Id of each order
DimOrder	Type	Payment method
DimOrder	Market	Market for the order
DimOrder	Order City	City of each order
DimOrder	Order Country	Country of each order
DimOrder	Order Customer Id	Customer order ID
DimOrder	Order date (DateOrders)	The date an order was placed
DimOrder	Order Region	Geographic area of order placement
DimOrder	Order State	State of order placement
DimOrder	Order Status	Status of each order

2.2.3. DimOrderDetail

Instructions!	Attribute / Fact Name	Description
Dimension / Fact Table		
DimOrderDetail	OrderDetailKey	A unique identifier for each order detail in the table
DimOrderDetail	Order Id	Id of each order
DimOrderDetail	Order Item Cardprod Id	Product code on the order
DimOrderDetail	Order Item Discount	Discount rate per unit of product
DimOrderDetail	Order Item Discount Rate	Percentage discount applied to the product
DimOrderDetail	Order Item Id	Product item code on an order
DimOrderDetail	Order Item Product Price	Original price per unit of product
DimOrderDetail	Order Item Profit Ratio	The ratio of profit obtained from an item to the sales value of that item.
DimOrderDetail	Order Item Quantity	Quantity of products on the order
DimOrderDetail	Sales	Selling price of the product calculated with quantity
DimOrderDetail	Order Item Total	Total unit price of each item after applying discounts
DimOrderDetail	Order Profit Per Order	Profit margin per item on the order

2.2.4. DimCategory

Instructions!	Attribute / Fact Name	Description
Dimension / Fact Table		
DimCategory	CategoryKey	A unique identifier for each order detail in the table
DimCategory	Category Id	Id of each category
DimCategory	Category Name	Name of category

2.2.5. DimCustomer

Instructions!	Attribute / Fact Name	Description
Dimension / Fact Table		
DimCustomers	CustomerKey	A unique identifier for each customer in the table
DimCustomers	CustomerID	Id of each customer
DimCustomers	CustomerFName	First name of each customer
DimCustomers	CustomerLName	Last name of each customer
DimCustomers	CustomerCity	City of each customer
DimCustomers	CustomerCountry	Country of customer
DimCustomers	CustomerSegment	Segment of each customer
DimCustomers	CustomerState	State of each customer
DimCustomers	CustomerStreet	Address of each customer
DimCustomers	CustomerZipcode	Zipcode of each customer

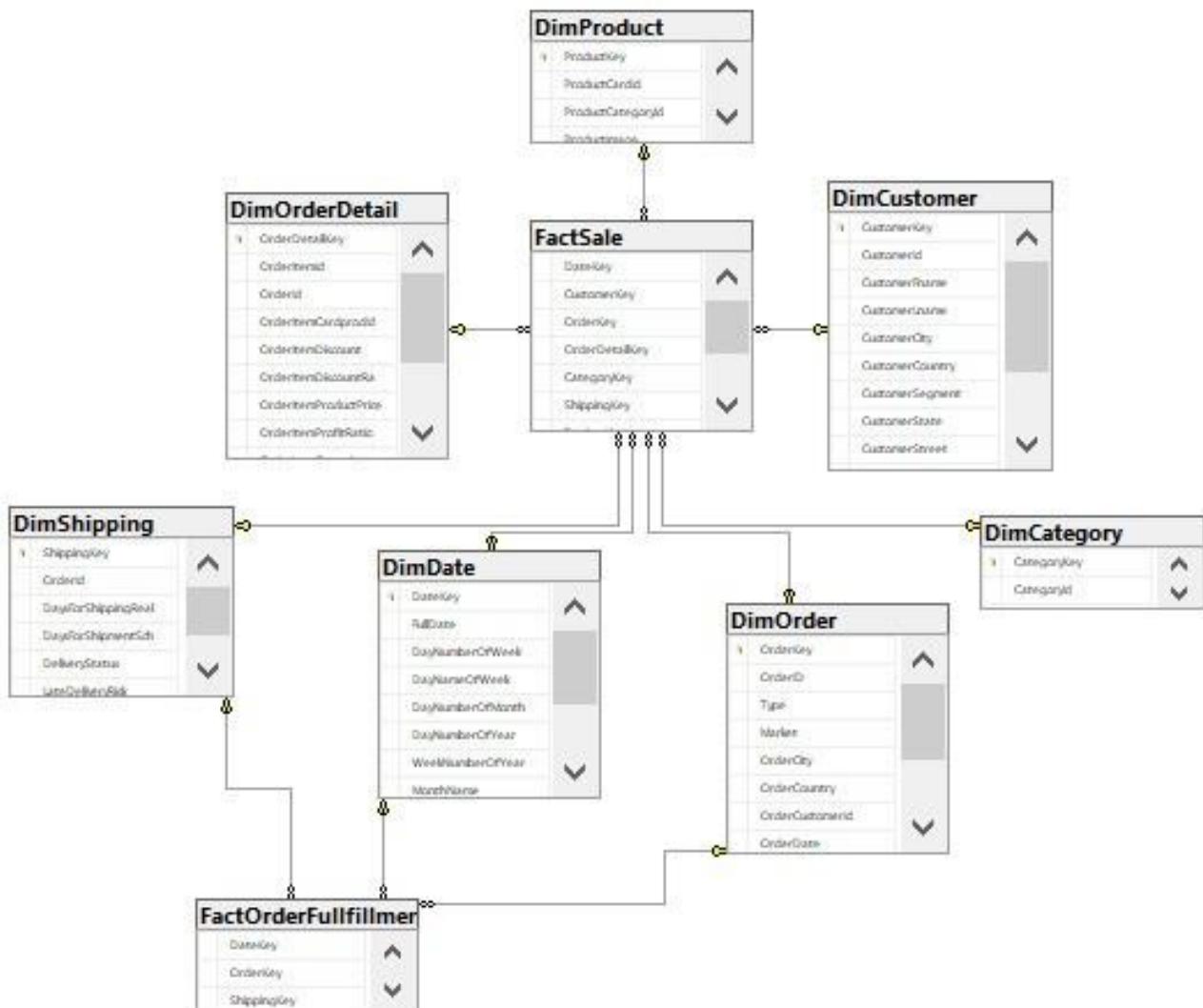
2.2.6. DimShipping

Instructions!		
Dimension / Fact Table	Attribute / Fact Name	Description
DimShipping	ShippingKey	A unique identifier for each product in the table
DimShipping	OrderID	Id of each order
DimShipping	DaysForShipping (real)	Number of actual shipping days
DimShipping	DaysForShipment (scheduled)	Scheduled shipping days
DimShipping	Delivery Status	Order delivery status
DimShipping	Late_delivery_risk	Risk of late delivery for the order
DimShipping	ShippingDate (DateOrders)	Date the order was handed over to the shipping carrier
DimShipping	ShippingMode	Shipping mode of each order

2.2.7. DimProduct

Instructions!		
Dimension / Fact Table	Attribute / Fact Name	Description
DimProduct	ProductKey	A unique identifier for each product in the table
DimProduct	ProductID	Id of each product
DimProduct	ProductCategoryID	Category id of each product
DimProduct	ProductImage	Image of each product
DimProduct	ProductName	Name of each product

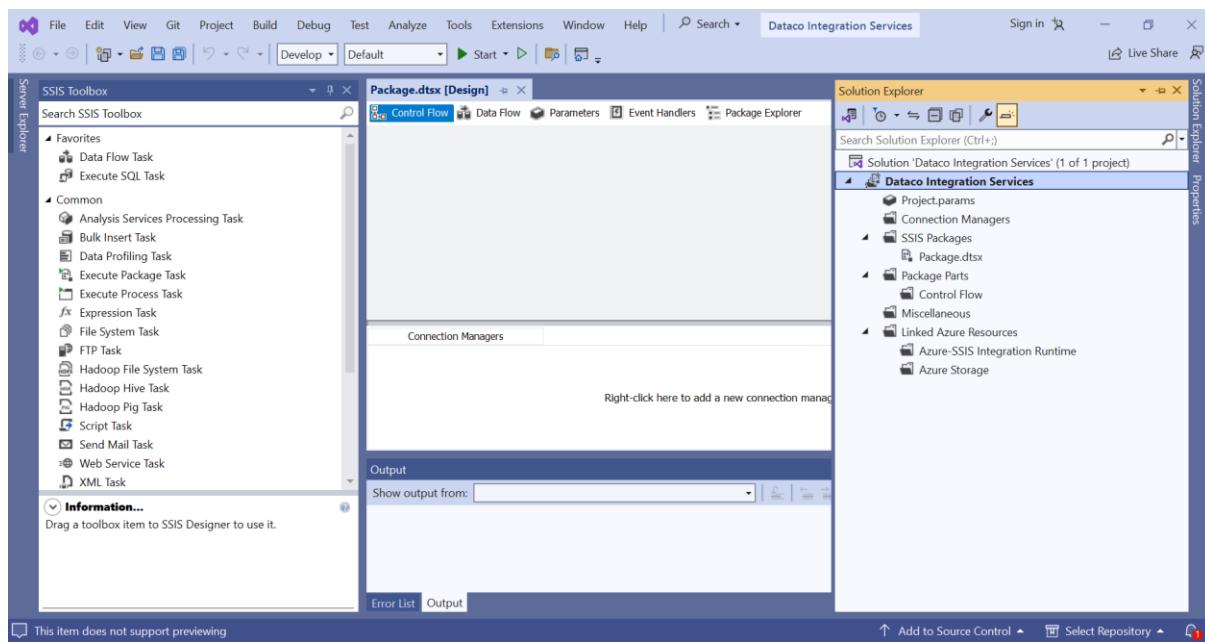
2.3. Galaxy Schema (Lược đồ galaxy)



CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

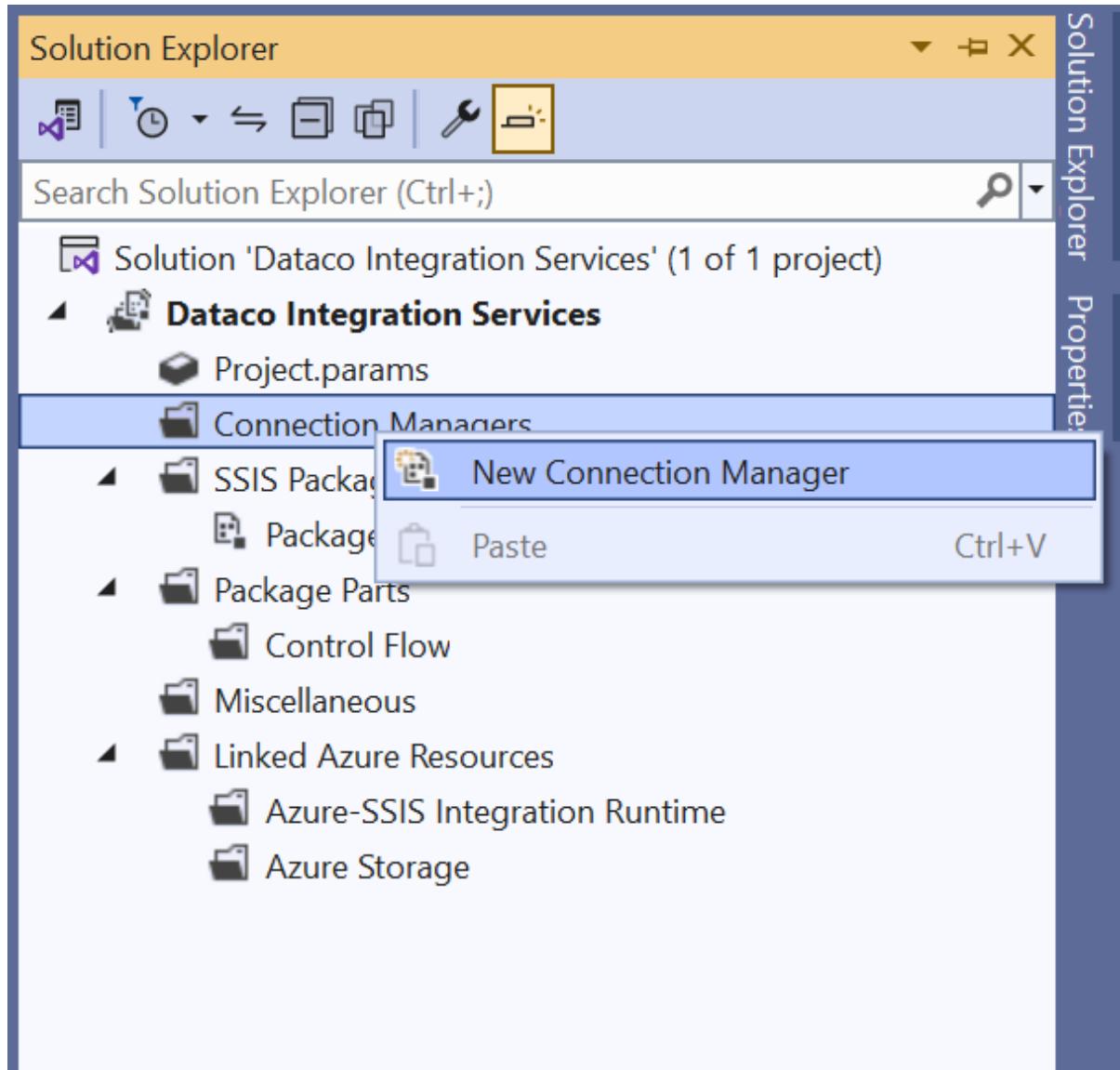
3.1. Tạo ETL project và solution

1. Từ Visual Studio menu, chọn File -> New -> Project.
2. Từ dialog:
 - a. Chọn Integration Services Project
 - b. Đặt tên: Dataco Integration Services
3. Click OK

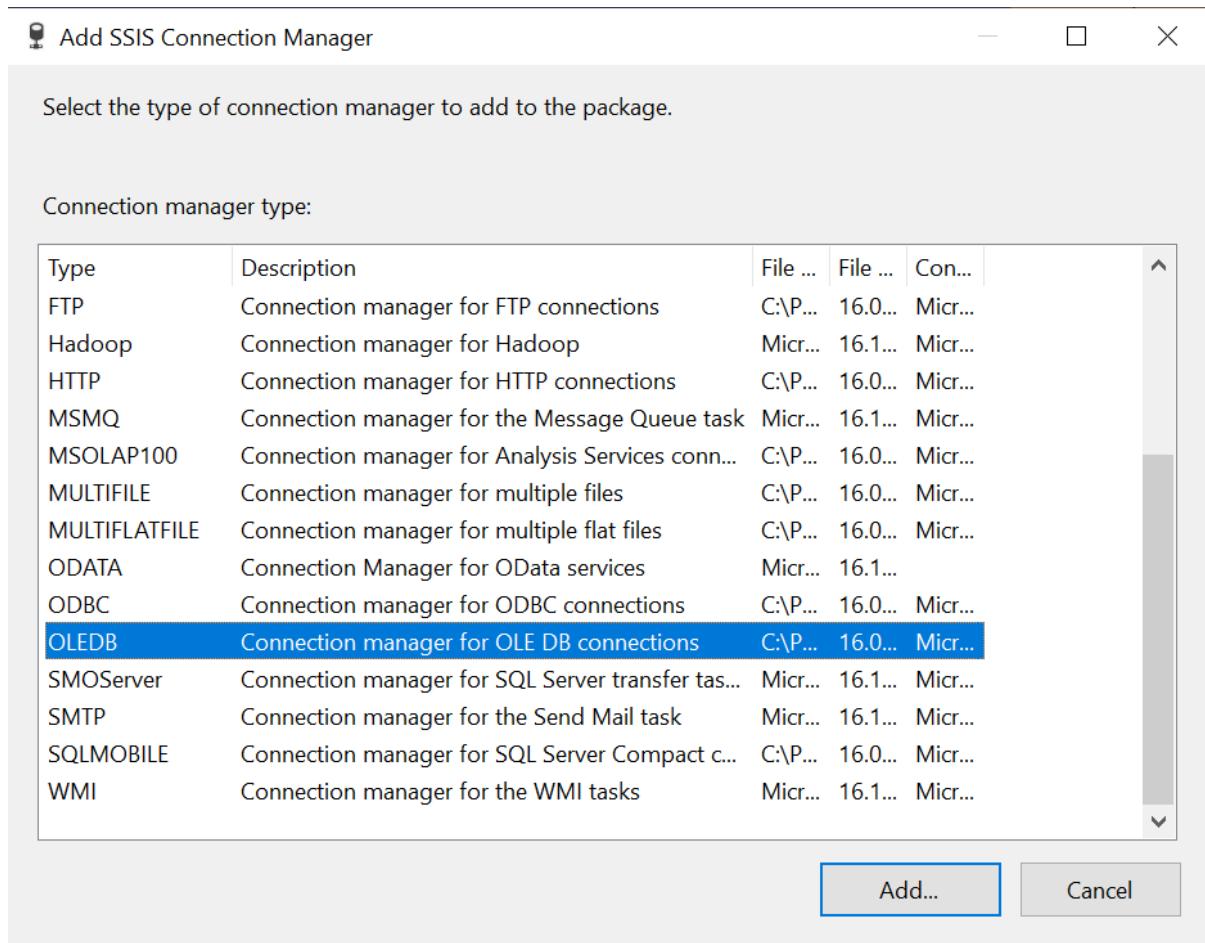


3.2. Thiết lập các Connection

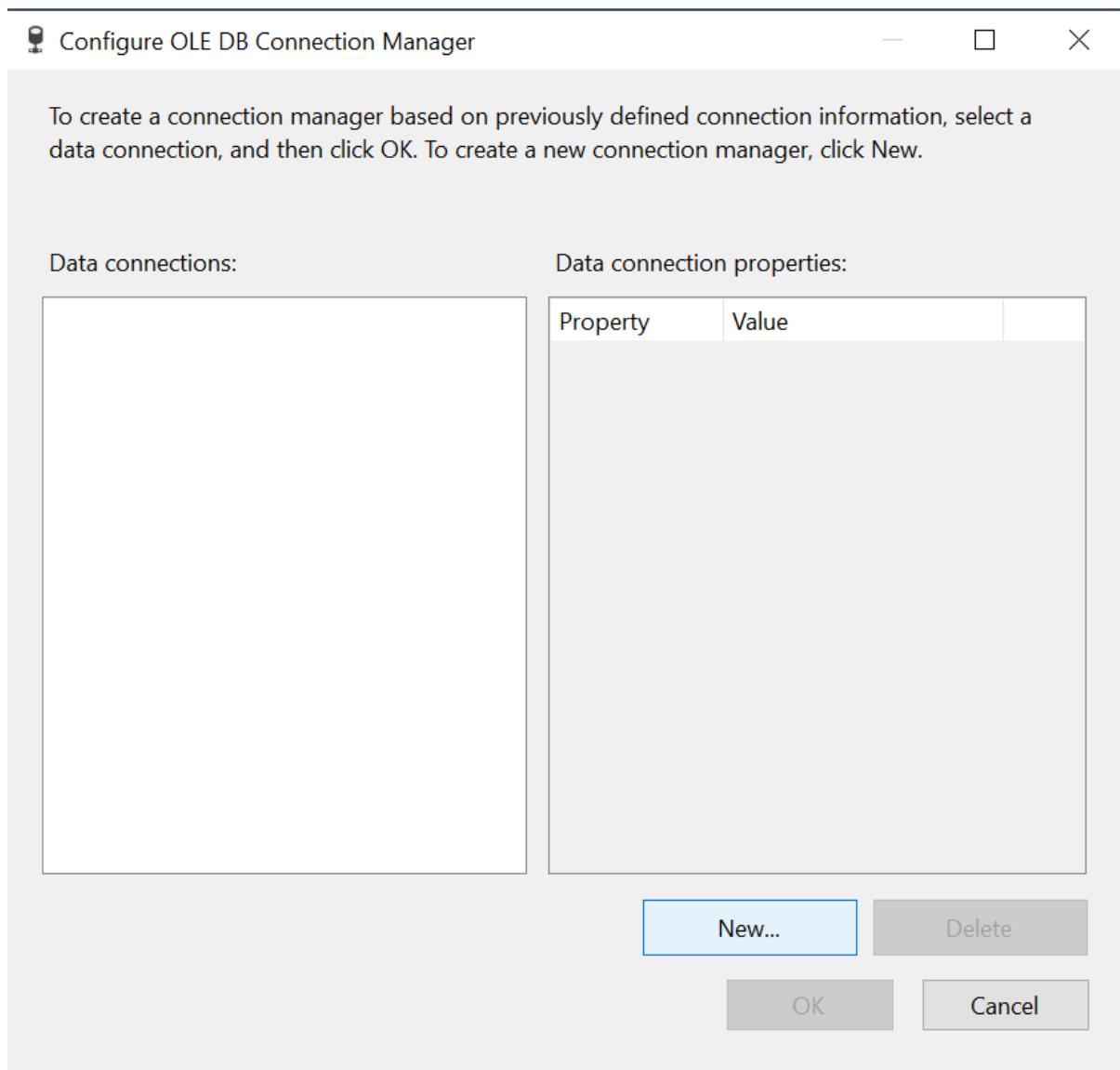
1. Tại **solution explorer**, right-click vào **Connection Managers**, chọn **New Connection Manager**.



2. Chọn **OLEDB** từ **Add SSIS Connection Manager** dialog và click **Add...**



3. Configure OLE DB Connection Manager dialog xuất hiện, click New...



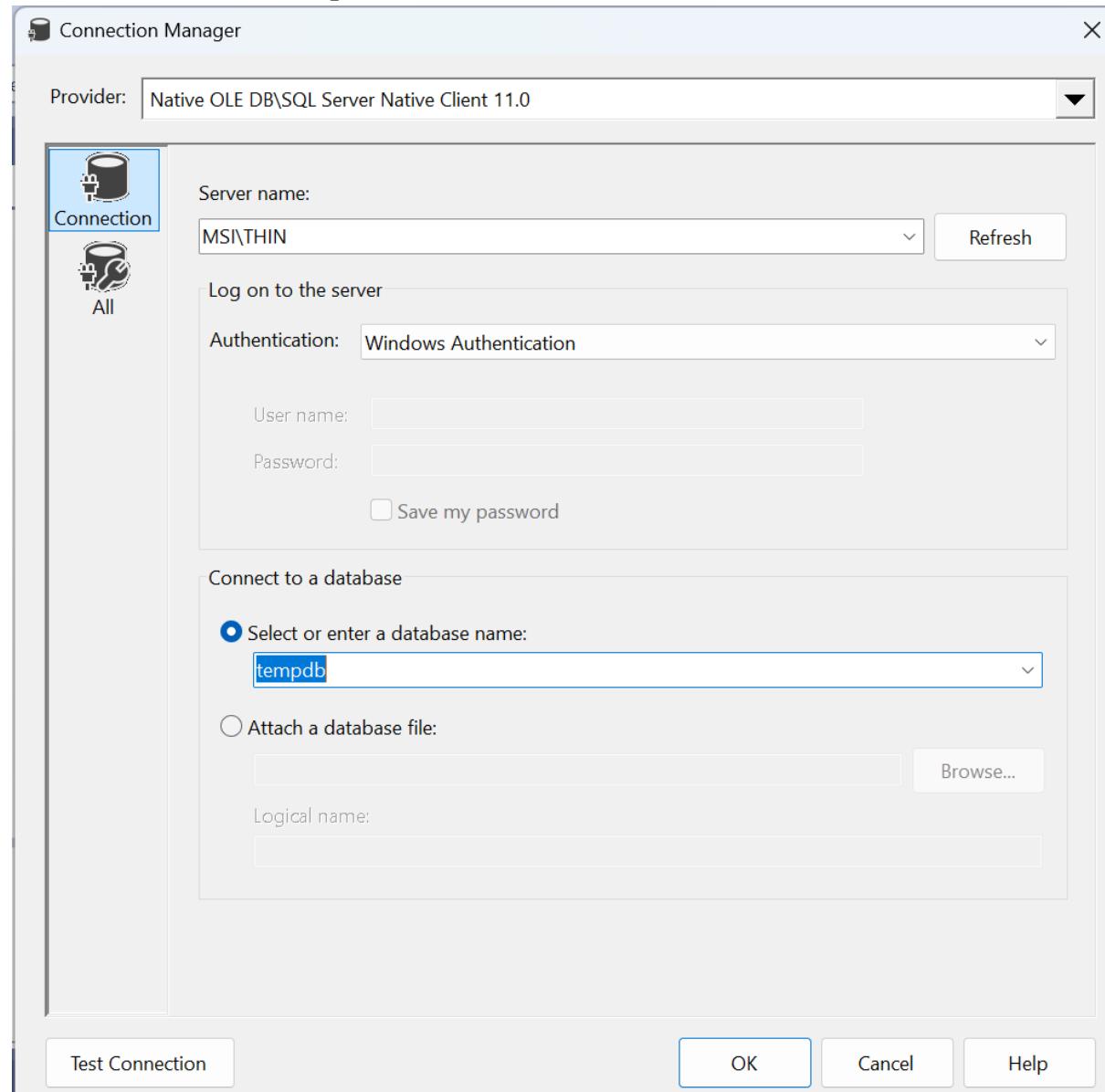
Đặt tên cho các thông tin như bên dưới:

Provider: **SQL Server Native Client**

Server Name: **MSI\THIN**

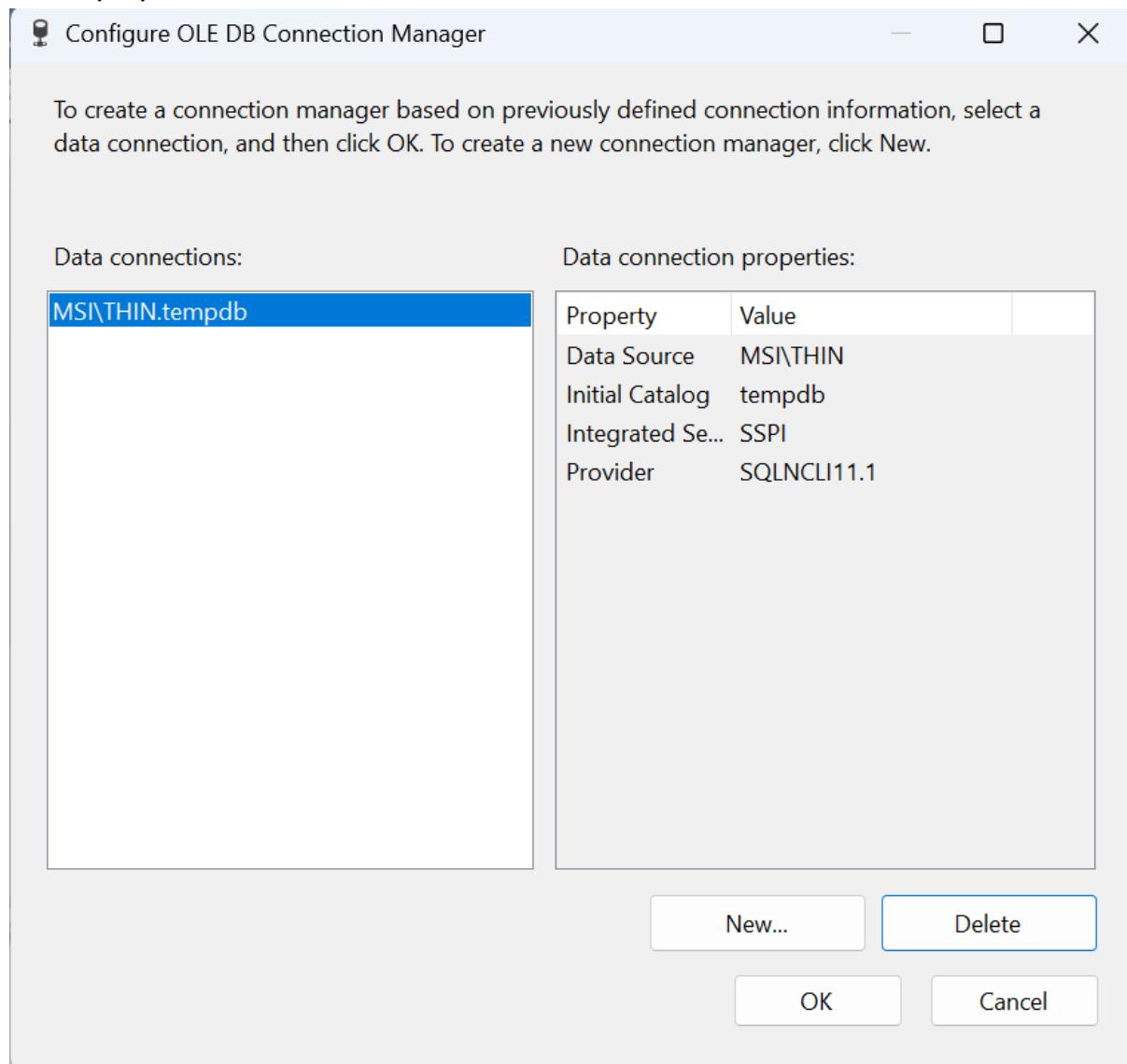
Log on to the Server: **Use Windows Authentication**

Connect to a database: **tempdb**



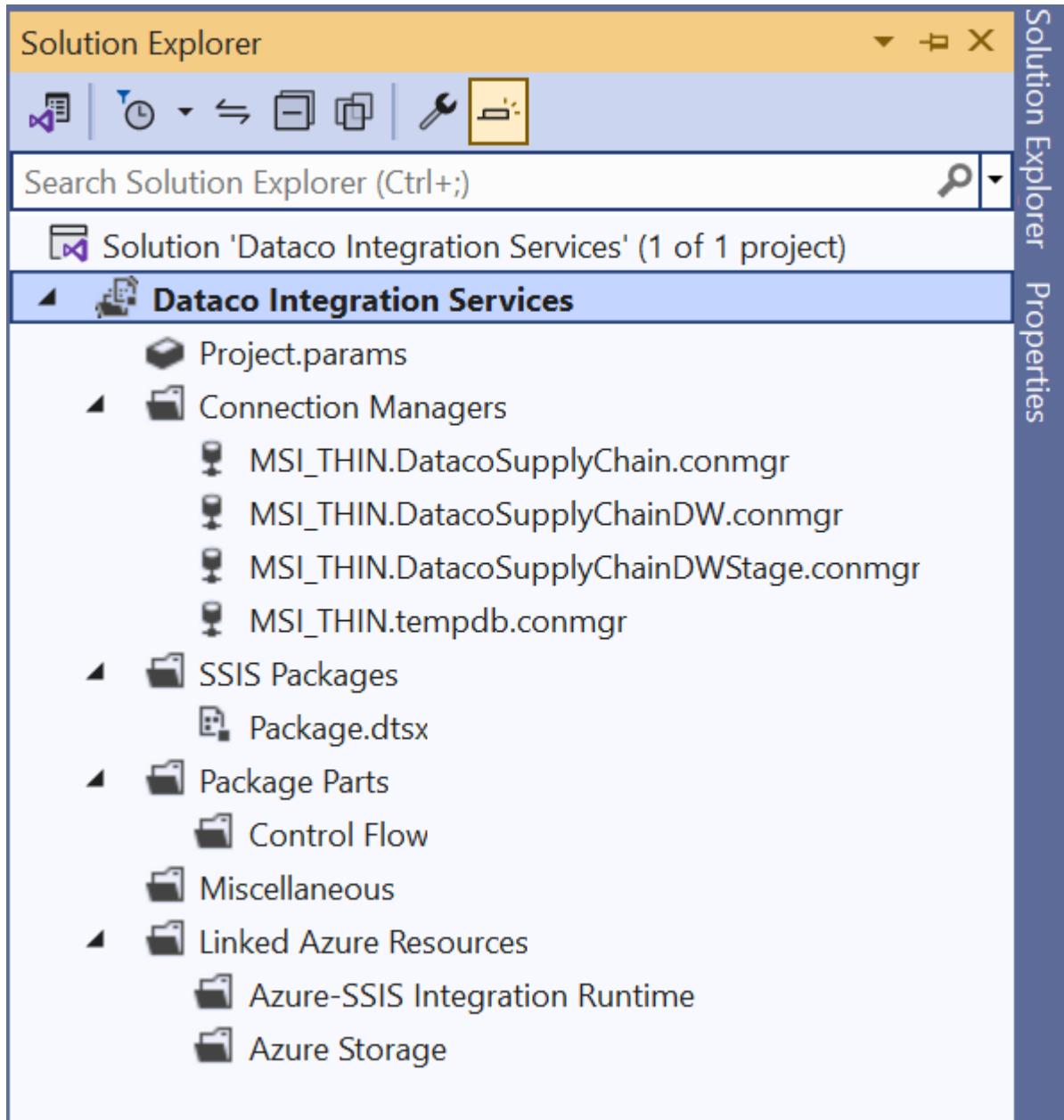
Click **Test Connection** để kiểm tra connection hoạt động, sau đó click **OK** để lưu connection.

4. Trở lại màn hình **Configure OLE DB Connection Manager**, ta được **MSI\THIN.tempdb** vừa được tạo:



5. Chọn **tempdb** connection và click **OK** để thêm connection này vào project.

6. Lặp lại các bước trên (bước 1->5) ba lần cho lần lượt **DatacoSupplyChainDW**, **DatacoSupplyChainDWStage** và **DatacoSupplyChainDatabase** source database. Sau khi tạo xong **Connection Managers** folder sẽ như bên dưới.



3.3. Import Dữ liệu vào các bảng dimension:

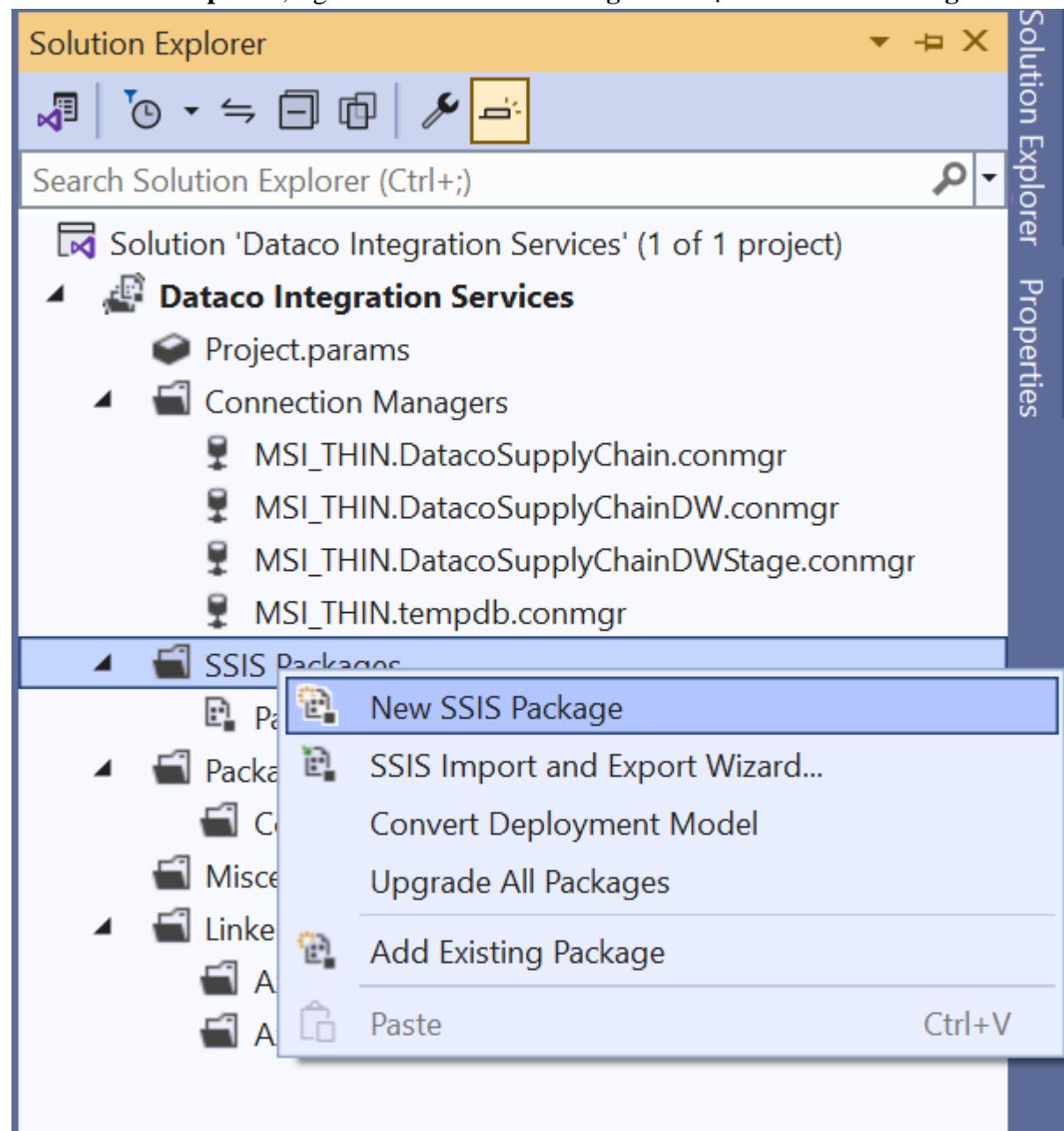
3.3.1.Date Dimension

- Đầu tiên chép dữ liệu từ bảng **[Date Dimension]** của CSDL **tempdb** sang bảng **stgDate** CSDL **DatacoSupplyChainDWStage**. Sử dụng SSIS để tạo destination table, truncate table trước khi staging dữ liệu.
- Ánh xạ các cột từ bảng **stgDate** sang bảng **DimDate**.

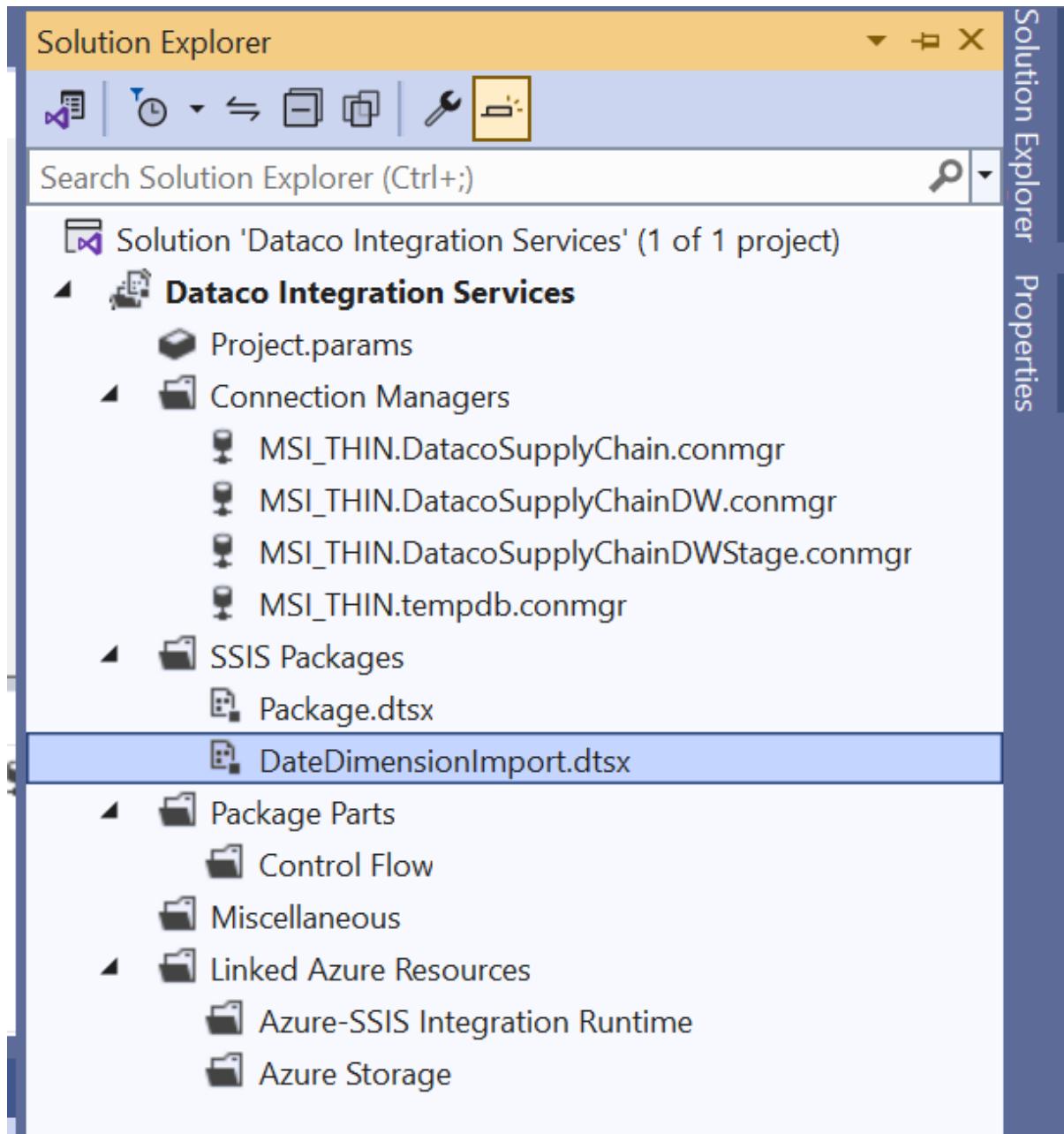
3.3.1.1: Tạo Package

Tạo một package mới, **DateDimensionImport**.

1. Ở **Solution Explorer**, right-click vào **SSIS Packages** và chọn **New SSIS Package**



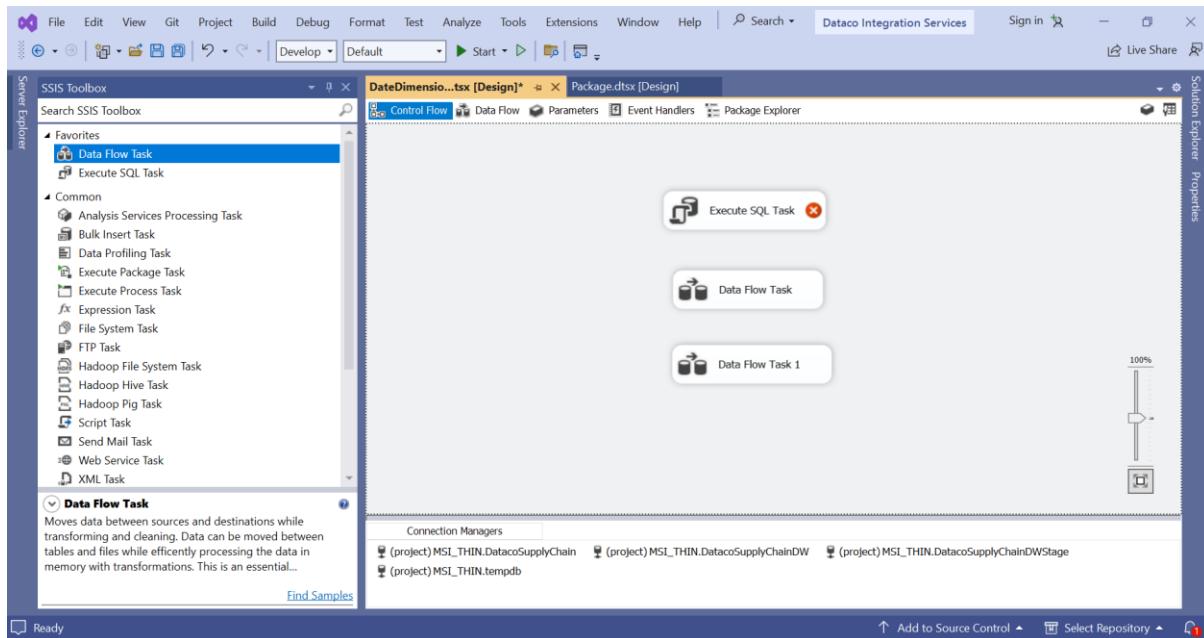
2. Right-Click vào package có tên **Package1.dtsx** và chọn **Rename**
3. Đặt tên package là **DateDimensionImport**
4. Sau khi thực hiện xong ta được **Package.dtsx** và **DateDimensionImport.dtsx** như hình sau:



5. Double-click vào **DateDimensionImport.dtsx** package để mở nó.

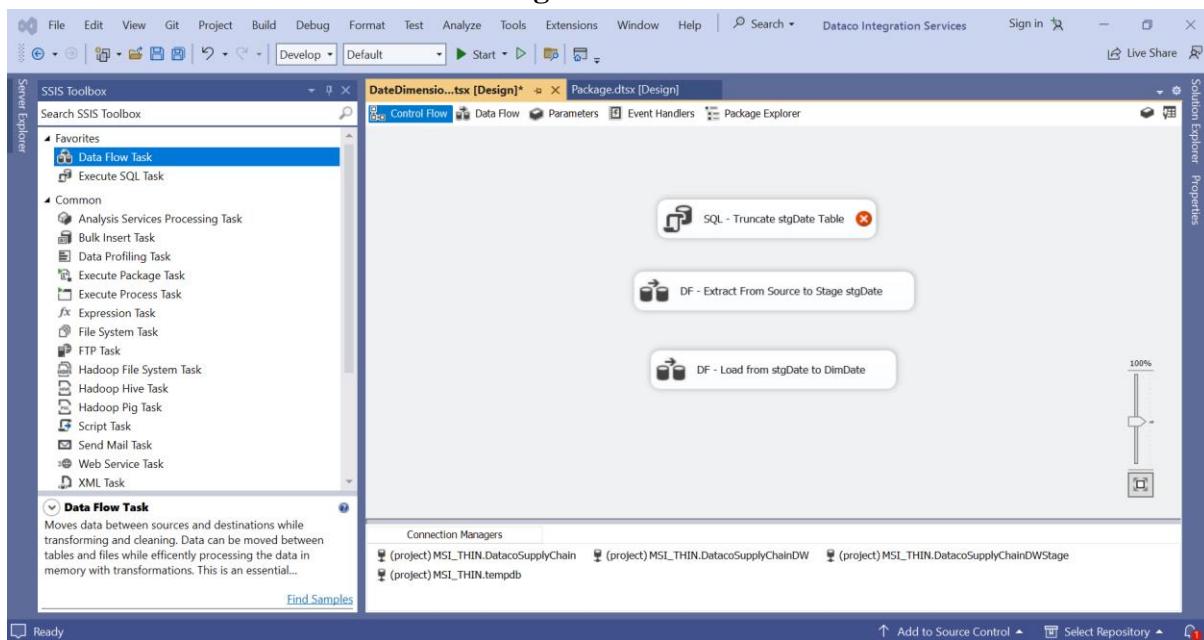
3.3.1.2: Thiết lập Control Flow

1. Từ SSIS Toolbox tạo 1 **Execute SQL task** và 2 **Data flow tasks** vào surface design như hình sau:

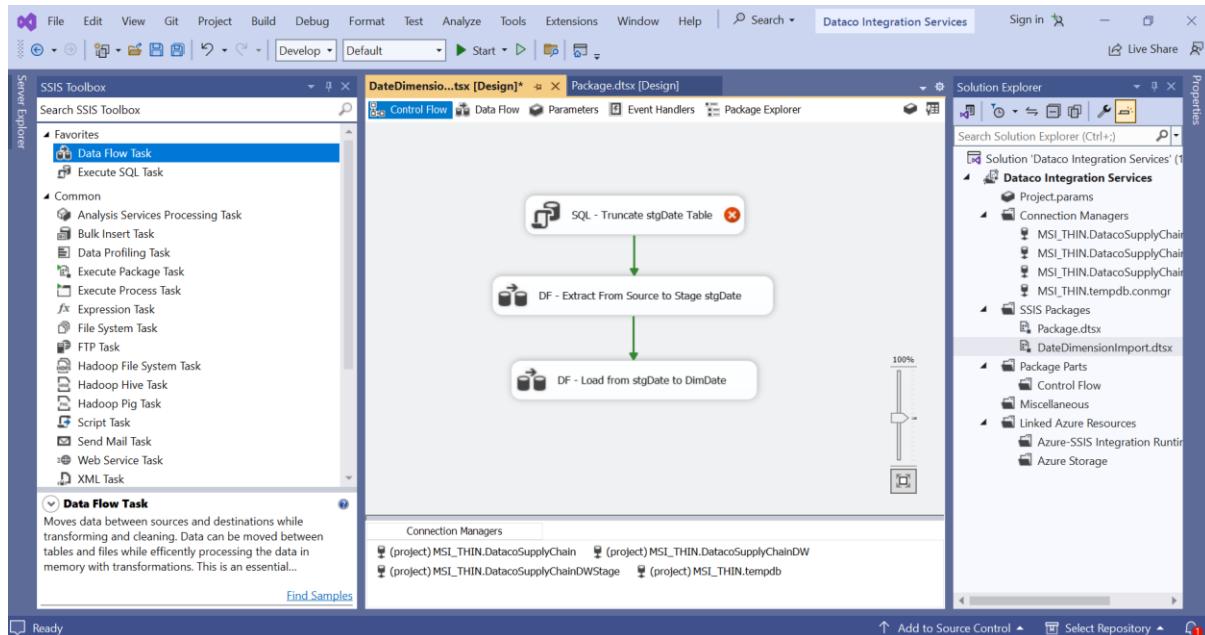


2. Đặt tên cho các task.

- Execute SQL Task SQL - Truncate stgDate Table**
- Data Flow Task DF - Extract From Source to Stage stgDate**
- Data Flow Task 1 DF - Load from stgDate to DimDate**

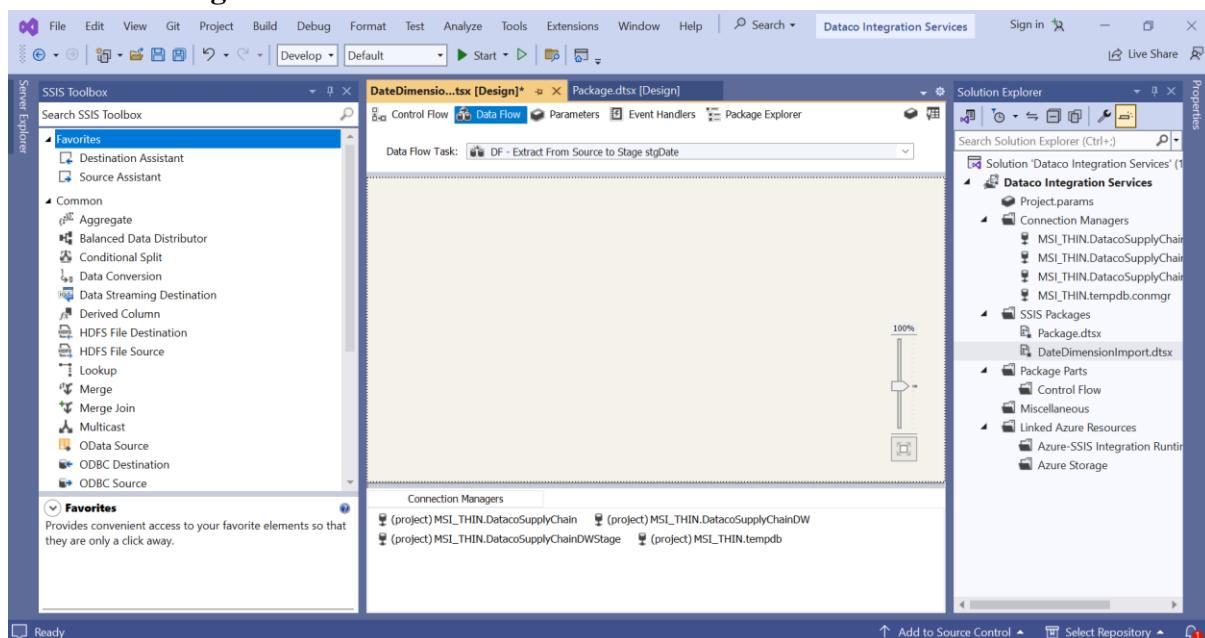


3. Connect các tasks để chúng có thể chạy theo thứ tự. Click vào **SQL - Truncate stgDate Table** task và một mũi tên xanh sẽ xuất hiện. Kéo rê nó vào **DF - Extract From Source To Stage stgDate** task. Tương tự, connect hai task còn lại.



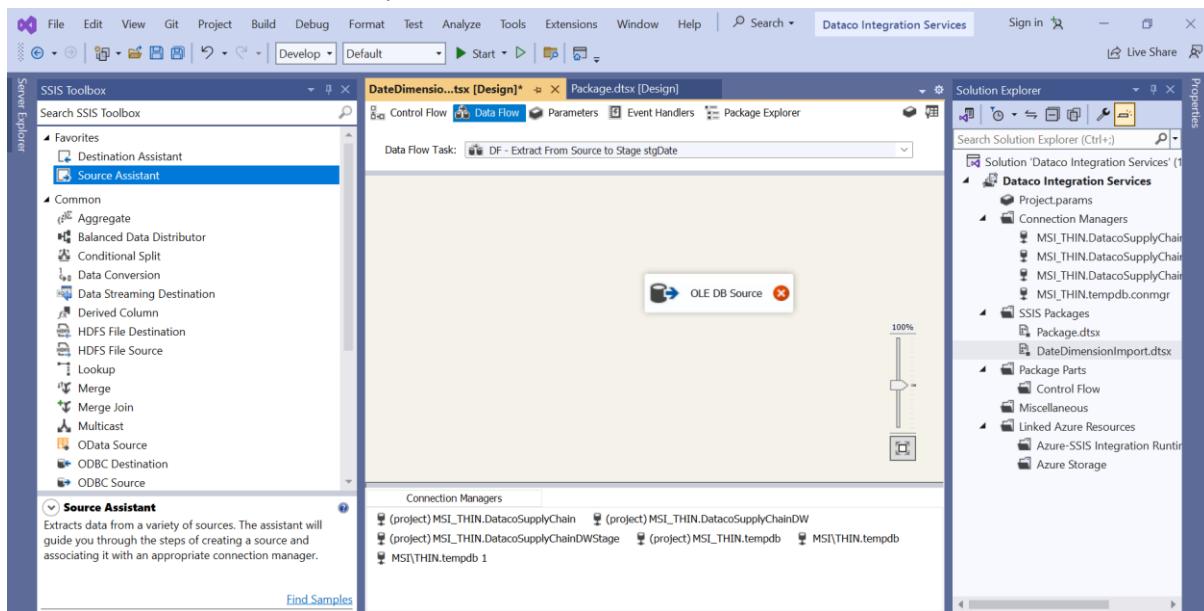
3.3.1.3: Load dữ liệu từ Source vào Stage

1. Double-click **DF - Extract From Source to Stage stgDate** task để mở nó trong **Data Flow design surface**:

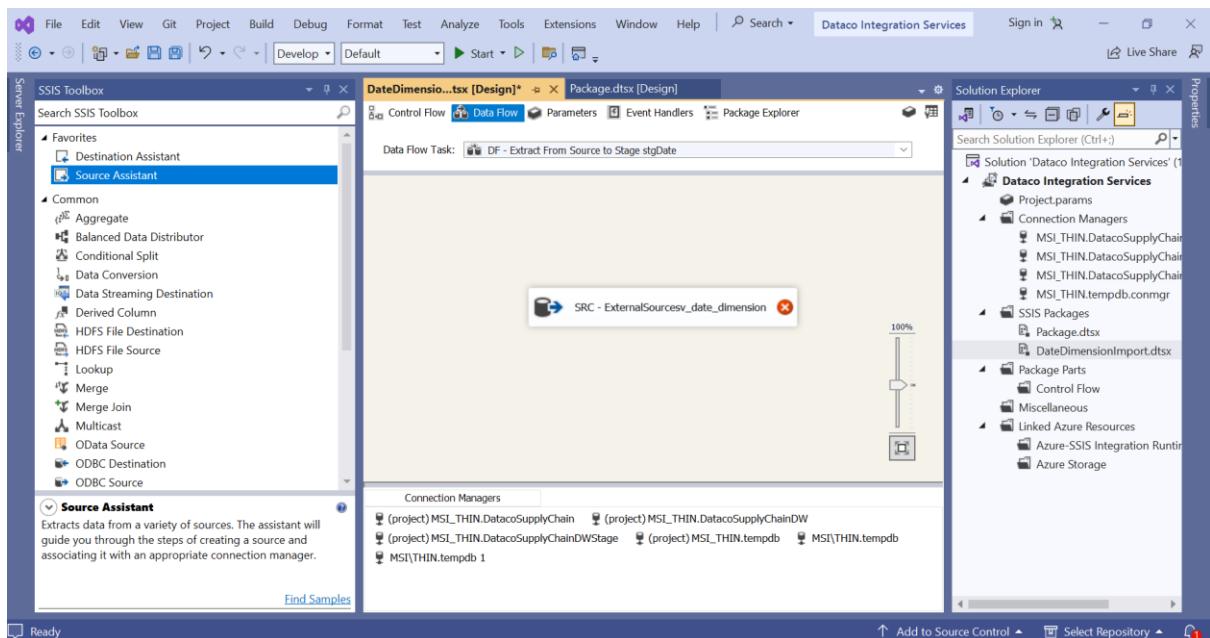


2. Từ **SSIS Toolbox**, kéo rê **Source Assistant** vào design surface, một dialog mở ra, Chọn source Type: **SQL Server** và **tempdb** connection manager. Click **OK**.

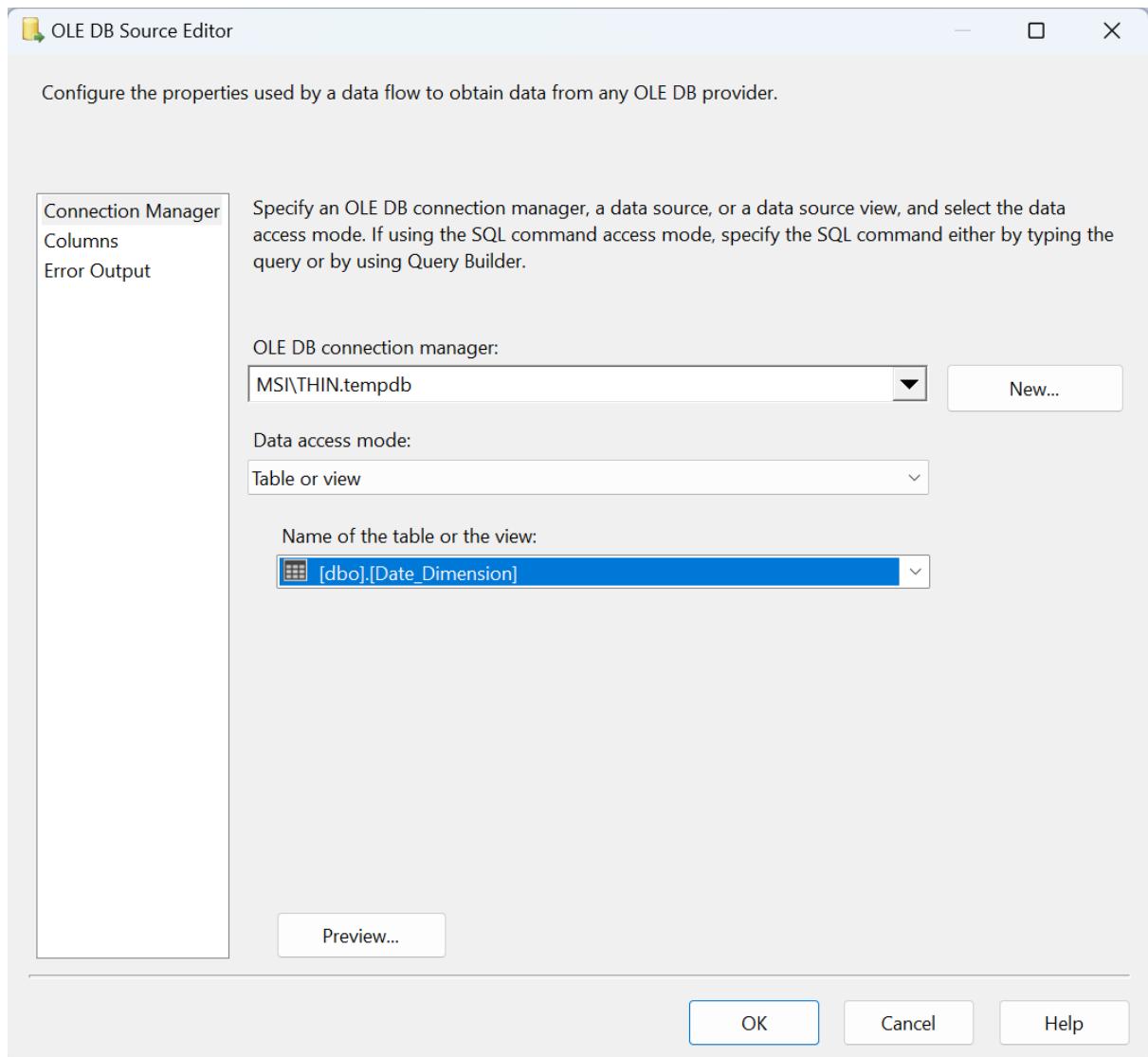
3. OLE DB Source sẽ xuất hiện như sau:



4. Đổi tên nó thành SRC - ExternalSourcesv_date_dimension

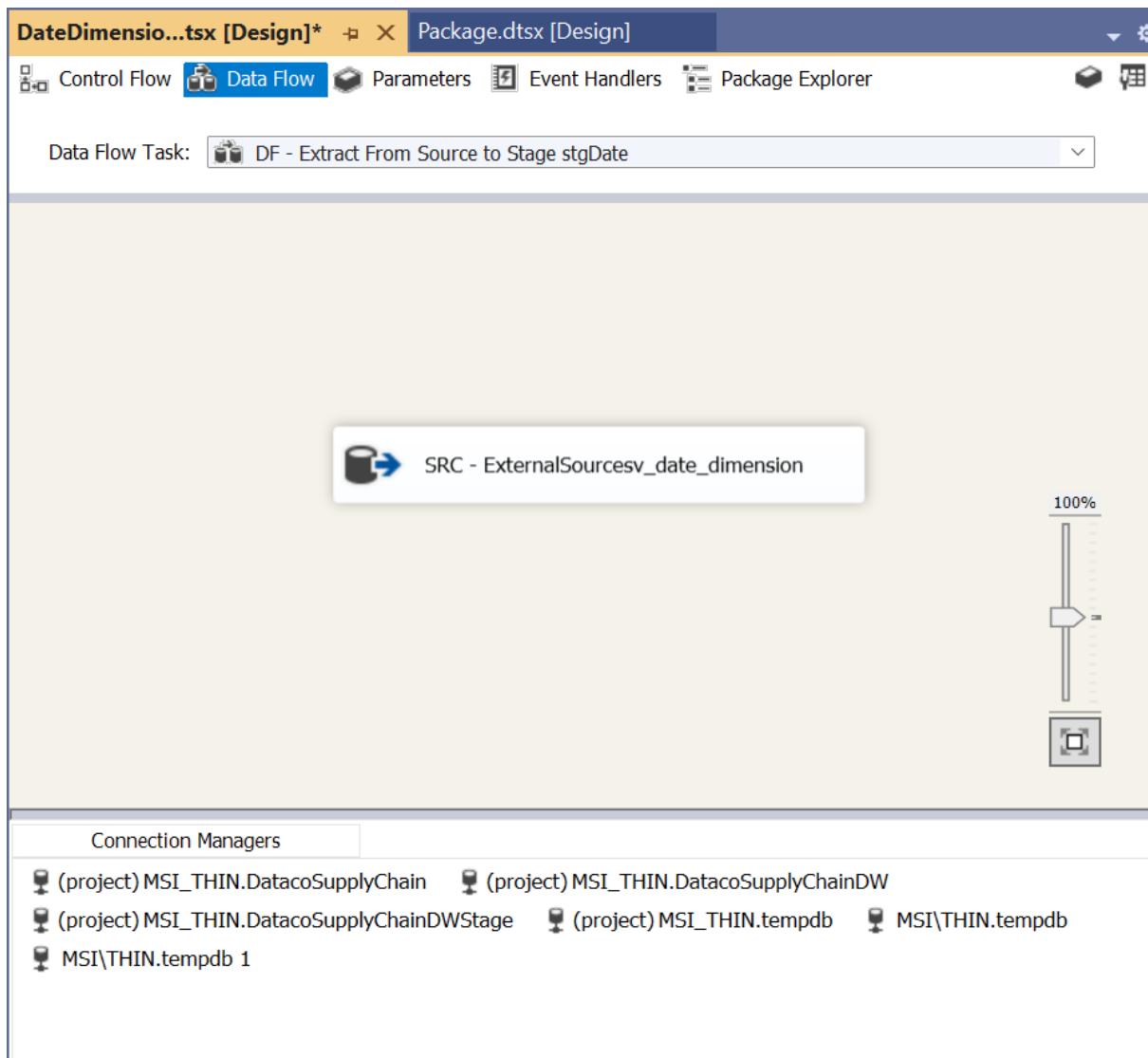


5. Double click vào source để cấu hình nó, **OLE DB Source Editor** sẽ mở ra. Từ dropdown **name of the table or view** chọn [**dbo].[Date Dimension**] là source. Click **Preview...** để xem source data.

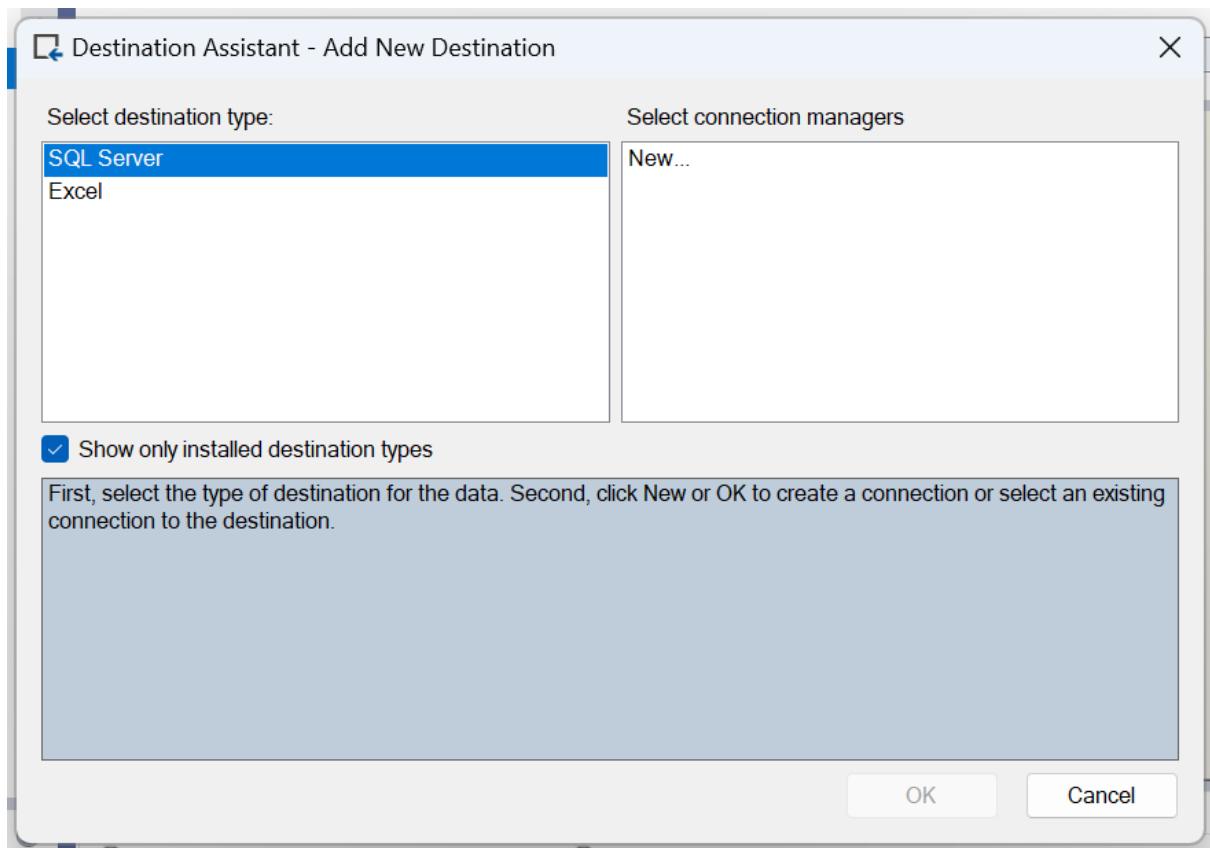


Click **OK** để lưu.

6. Sau khi cấu hình **source** xong, bạn sẽ không còn thấy dấu [x] màu đỏ (như bên dưới):

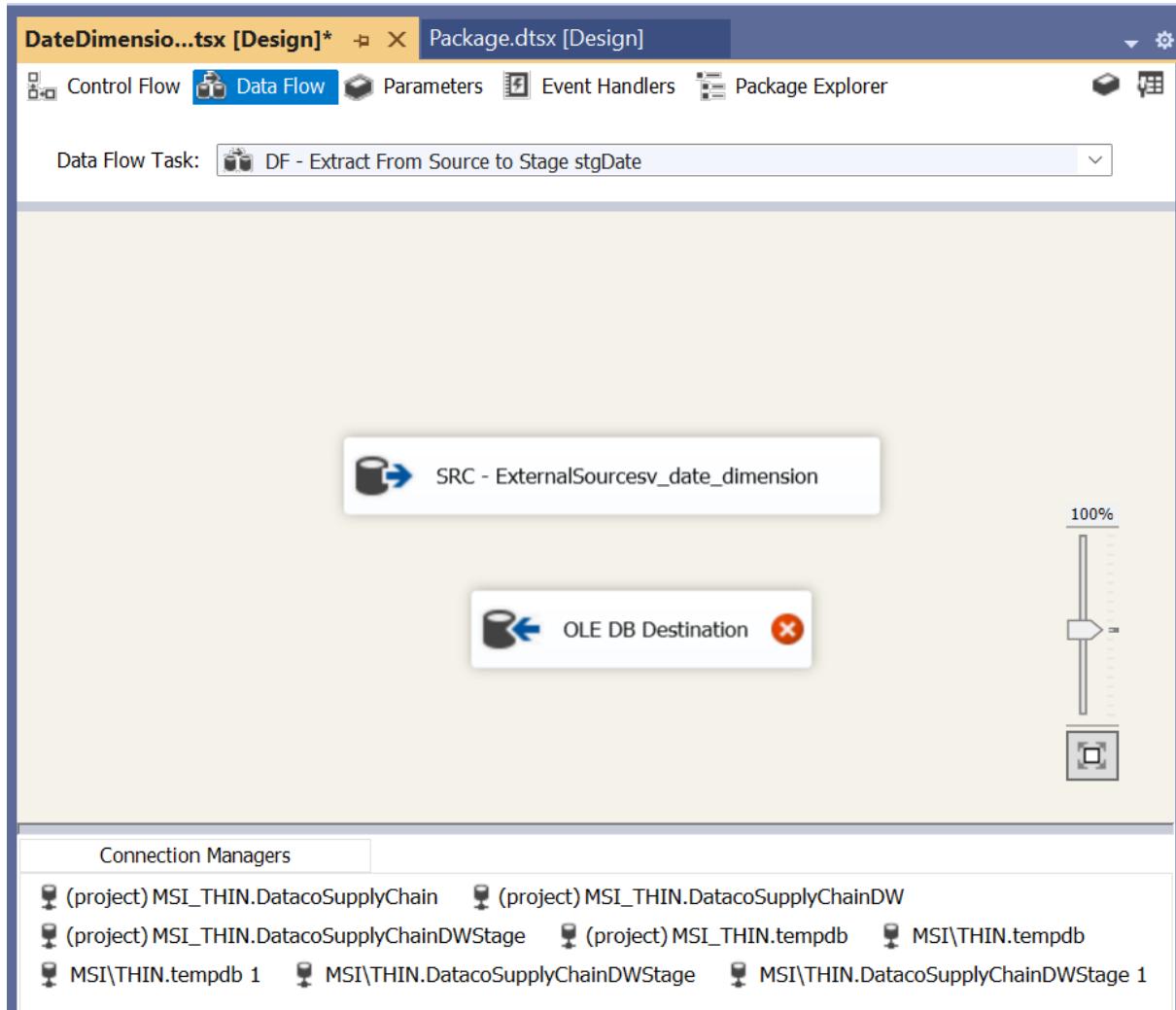


7. Tiếp tục, cấu hình destination. Từ **SSIS Toolbox**, kéo rê **Destination Assistant** vào design surface, một dialog xuất hiện. Chọn source Type: **SQL Server** và **HardwareStage** là connection manager.

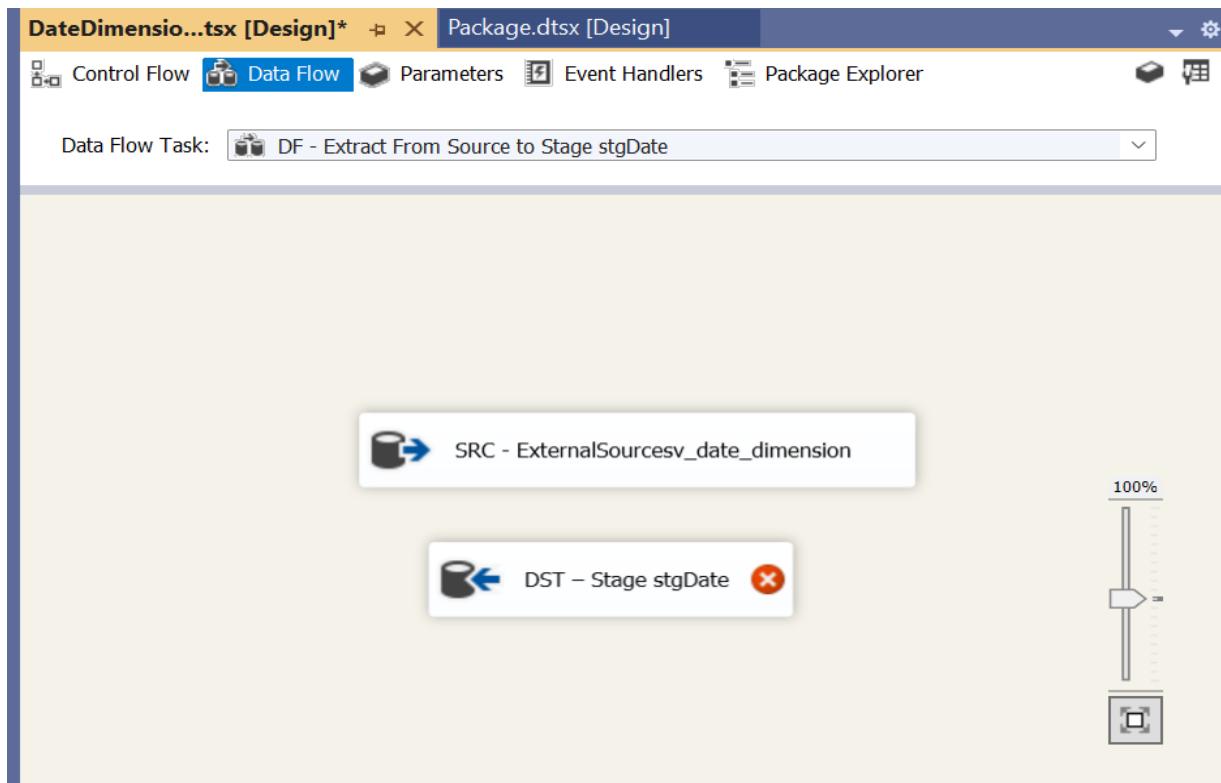


Click **OK**.

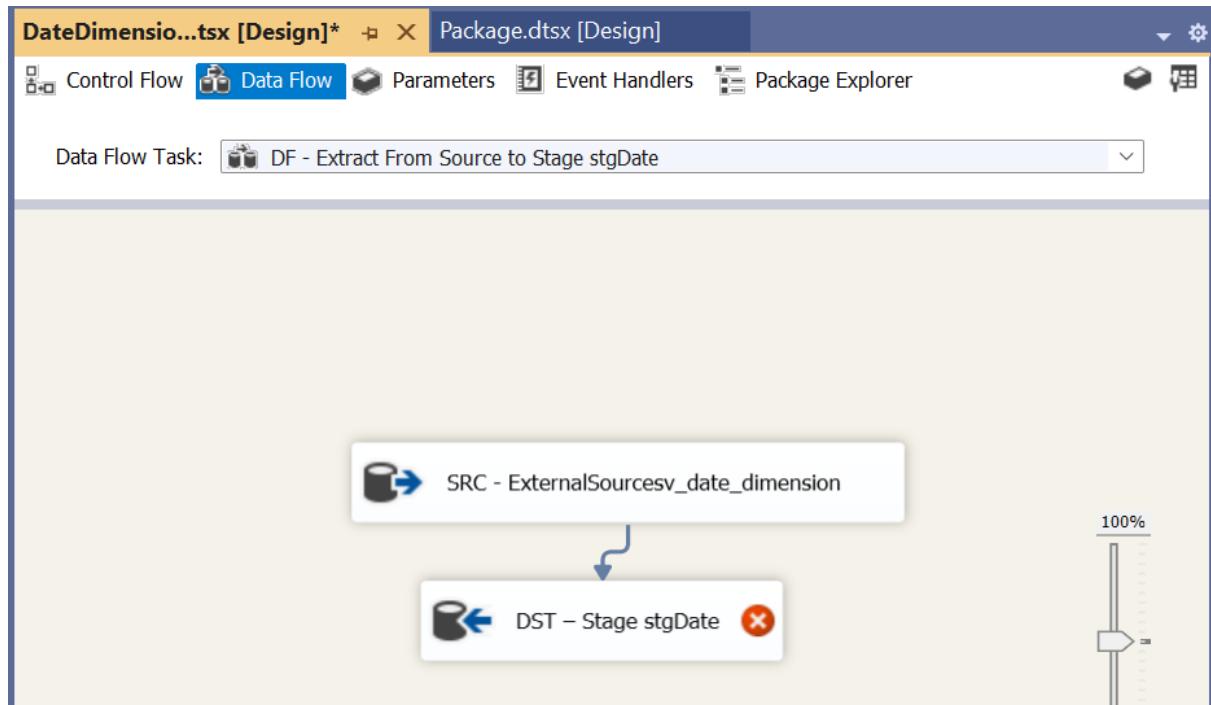
8. Ta có một **destination** như bên dưới:



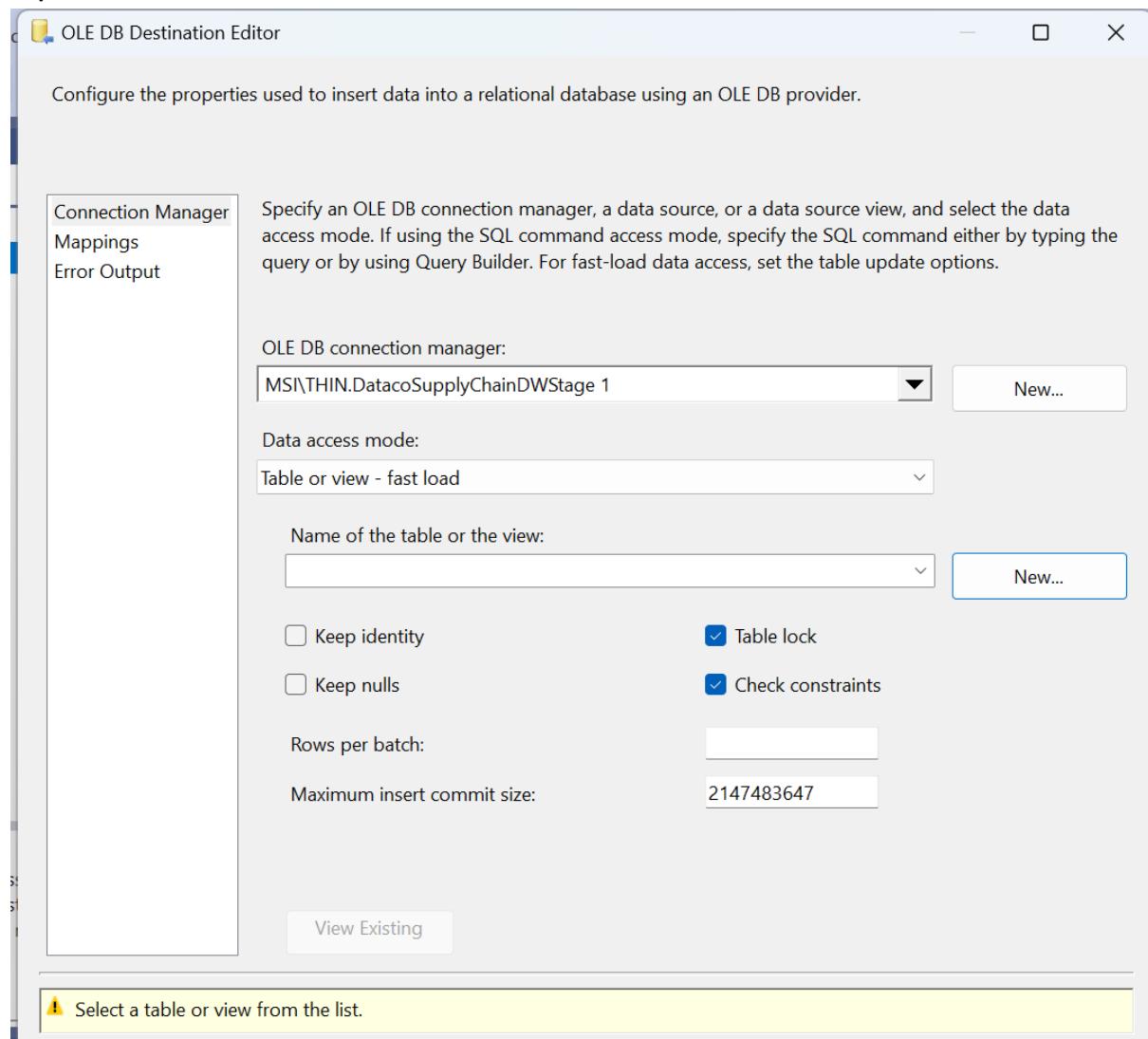
9. Đổi tên lại thành **DST – Stage stgDate**



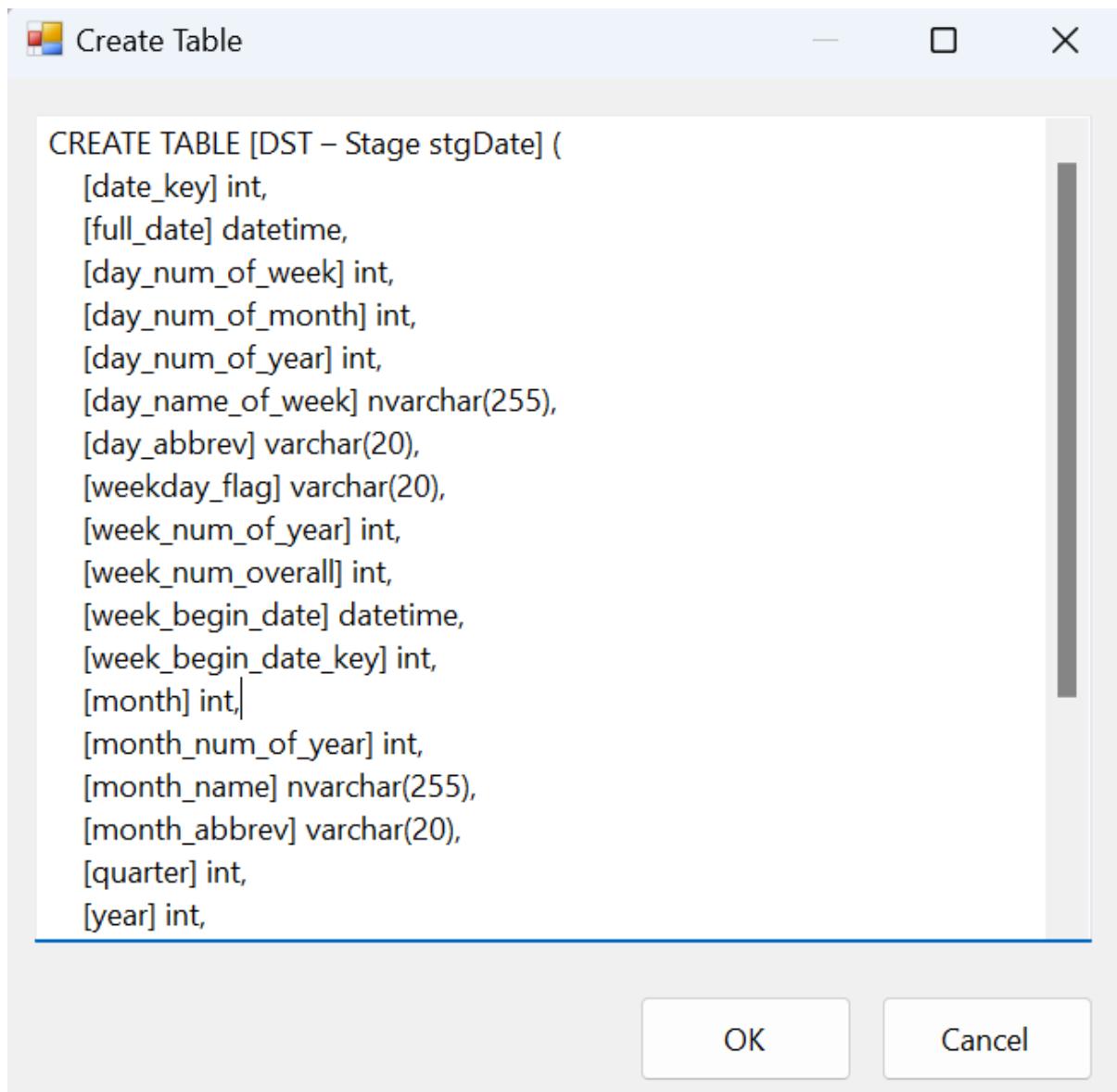
10. Click vào **source** và kéo rê nó vào **destination**.



11. Double click vào destination để cấu hình nó, **OLE DB Destination Editor** sẽ xuất hiện như sau:



Click nút **New...** để tạo bảng trong CSDL stage, ta sẽ thấy window **Create Table** như sau:



Đổi tên bảng lại thành **stgDate**. (**CREATE TABLE [stgDate]**).



Create Table

```
CREATE TABLE [stgDate] (
    [date_key] int,
    [full_date] datetime,
    [day_num_of_week] int,
    [day_num_of_month] int,
    [day_num_of_year] int,
    [day_name_of_week] nvarchar(255),
    [day_abbrev] varchar(20),
    [weekday_flag] varchar(20),
    [week_num_of_year] int,
    [week_num_overall] int,
    [week_begin_date] datetime,
    [week_begin_date_key] int,
    [month] int,
    [month_num_of_year] int,
    [month_name] nvarchar(255),
    [month_abbrev] varchar(20),
    [quarter] int,
    [year] int,
```

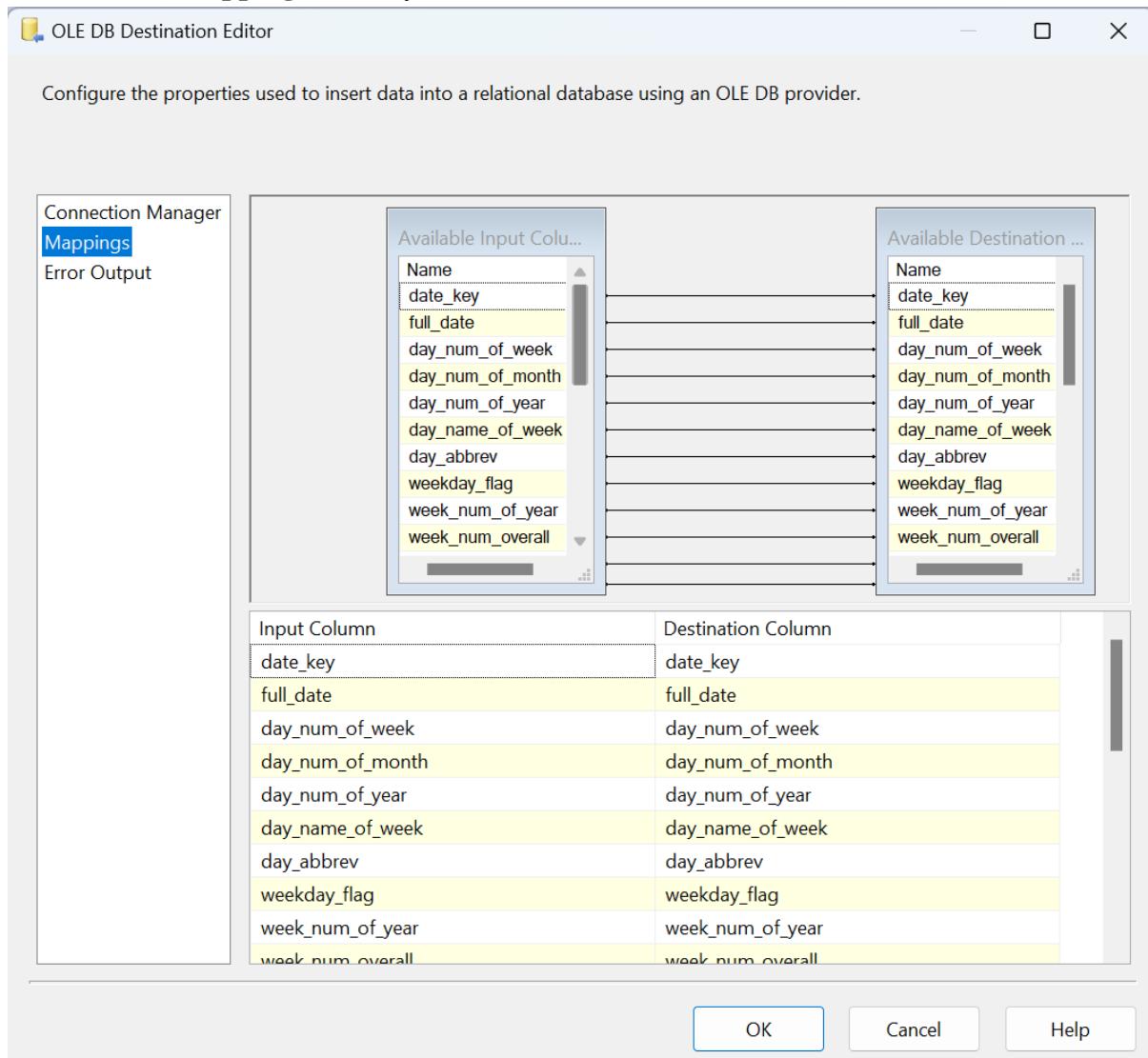
OK

Cancel

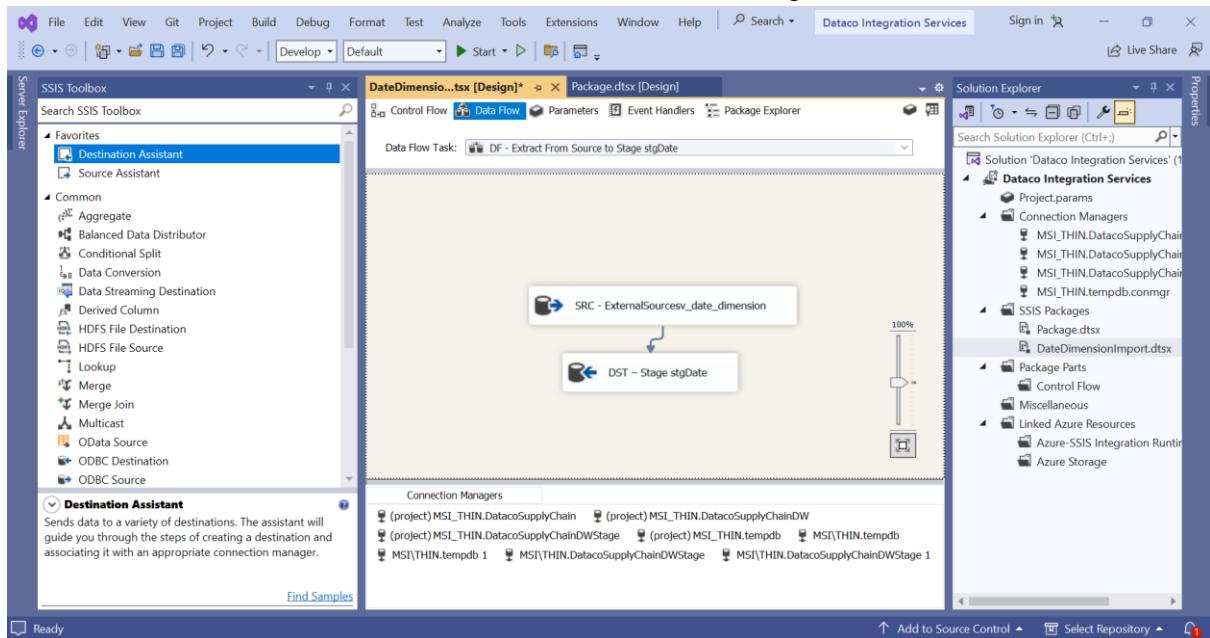
Click OK.

Trở lại màn hình **OLEDB Destination Editor**, tại mục **Name of the table or view** bạn sẽ thấy **[stgDate]**.

Click vào tab **Mapping** ta sẽ thấy như hình bên dưới:



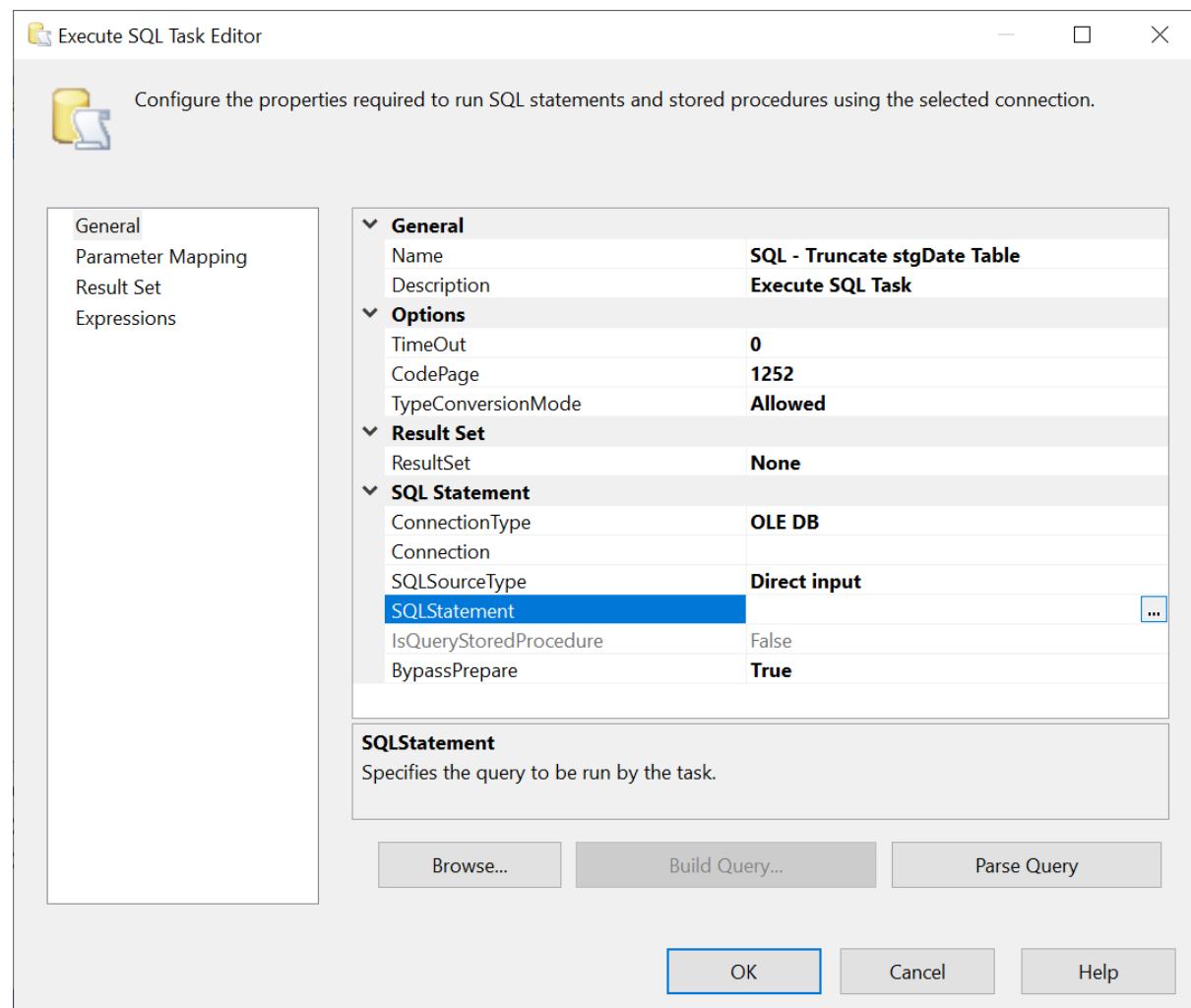
Click **OK** để hoàn tất việc cấu hình. Màn hình data flow design surface sẽ như bên dưới.



3.3.1.4: Thiết lập truncate table SQL task

Click **Control Flow** tab.

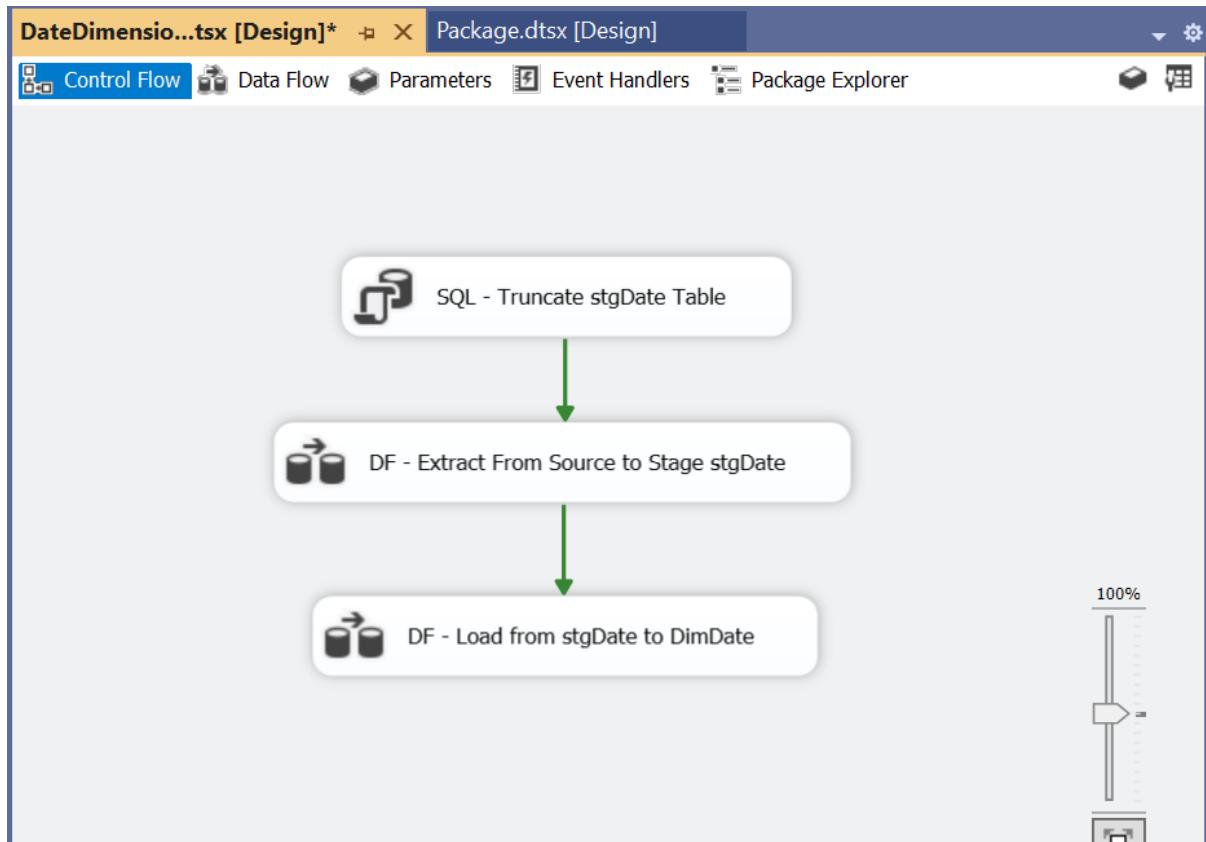
- Double-click vào **SQL - Truncate stgDate Table** task để cấu hình nó, **Execute SQL Task Editor** sẽ mở ra.



- Ở mục **SQL Statement**, chọn **Connection** là **MSI\THIN.DatacoSupplyChainDWStage** và **SQLStatement** là **truncate table stgDdate;**

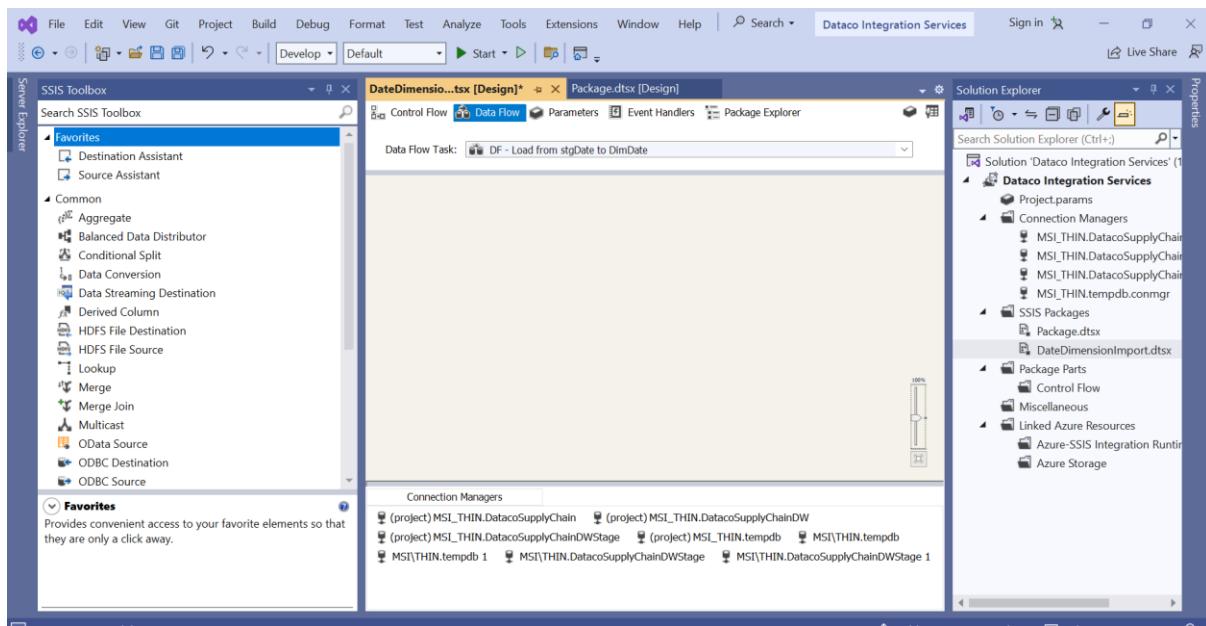
General	
Name	SQL - Truncate stgDate Table
Description	Execute SQL Task
Options	
TimeOut	0
CodePage	1252
TypeConversionMode	Allowed
Result Set	
ResultSet	None
SQL Statement	
ConnectionType	OLE DB
Connection	MSI\THIN\DatacoSupplyChainDWStage
SQLSourceType	Direct input
SQLStatement	truncate table stgDdate; ...
IsQueryStoredProcedure	False
BypassPrepare	True

3. Click **OK** và trở lại màn hình **Control Flow**

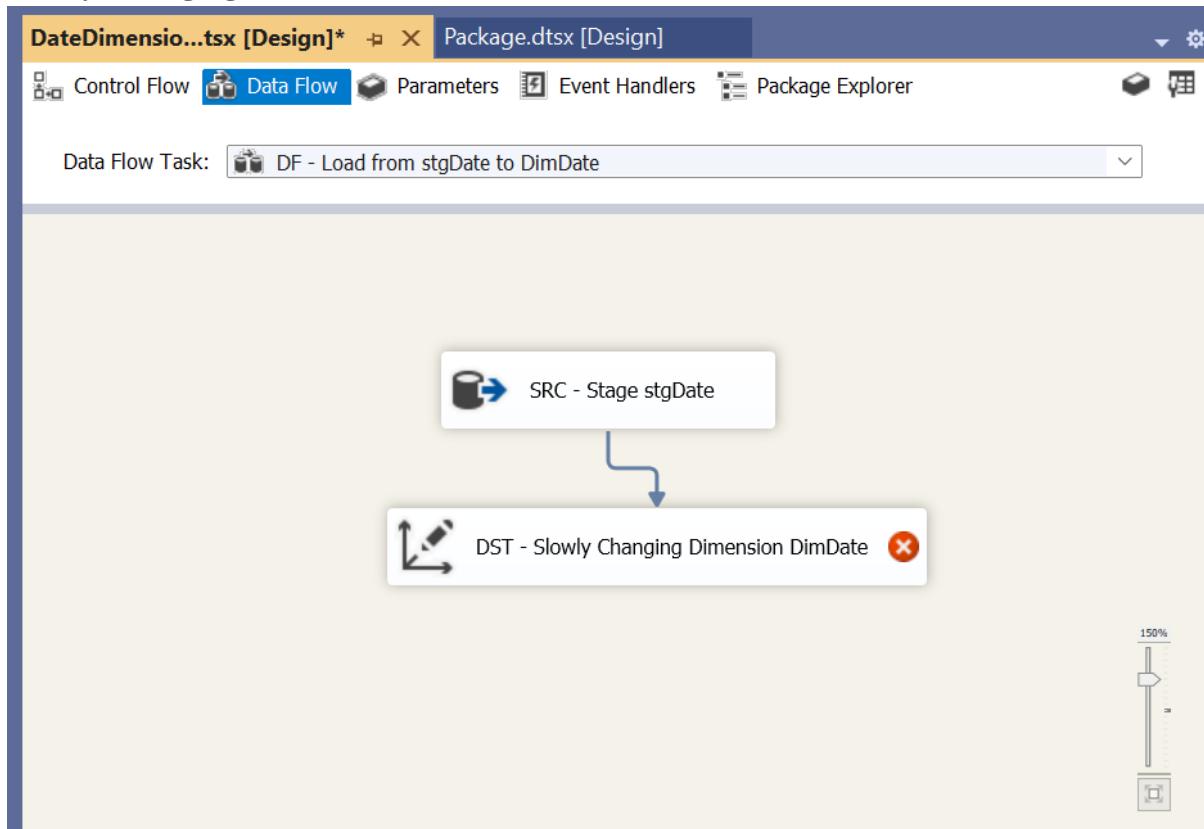


3.3.1.5: Load dữ liệu từ Stage vào Dimension

1. Double-click vào **DF - Load from stgDate to DimDate** task để mở data flow design surface.

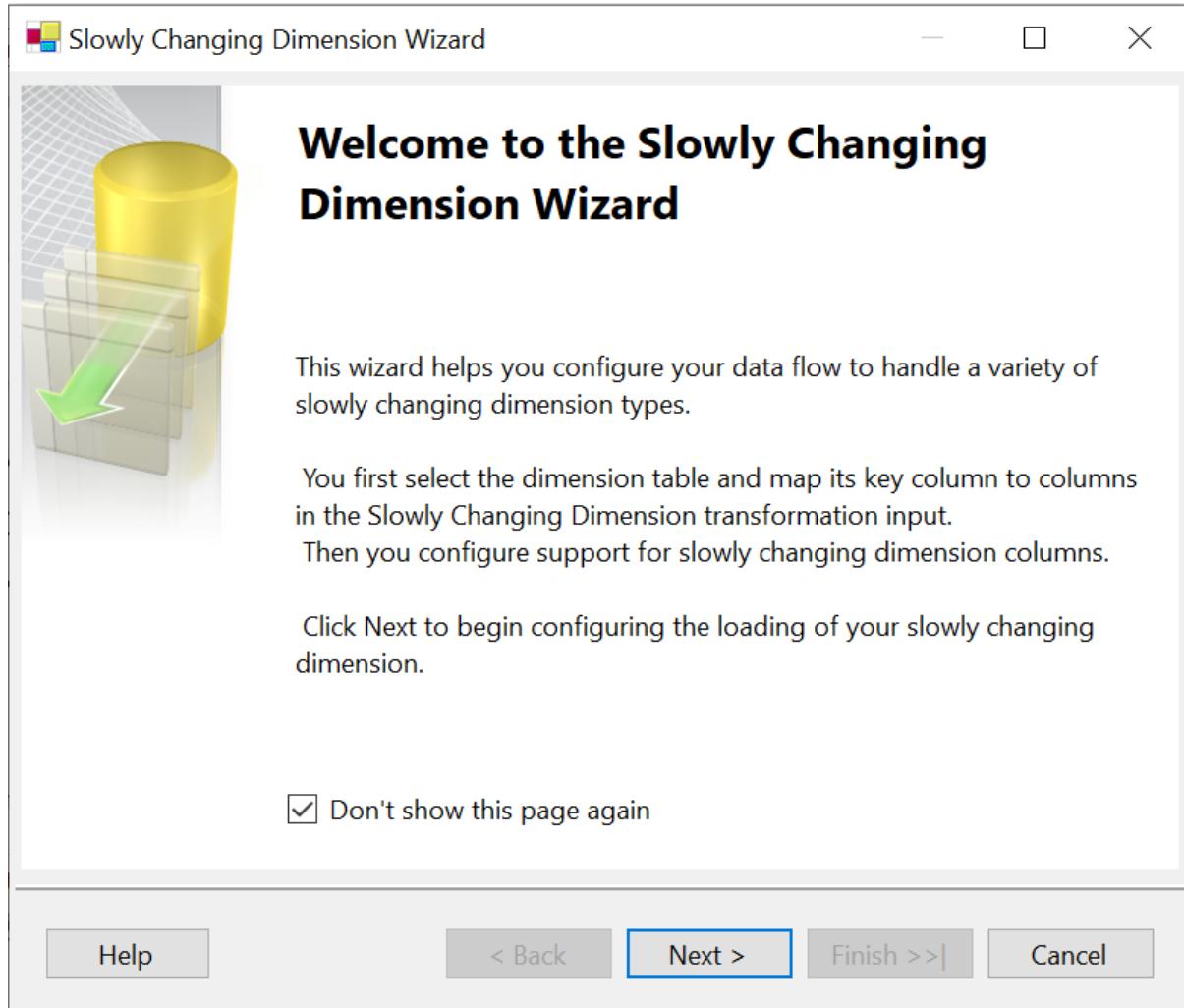


2. Kéo rê **Source Assistant** vào design surface, chọn **source type** là **SQL Server** và connection manager là **HardwareStage**.
3. Đổi tên source thành **SRC - Stage stgDate**, và double-click vào nó để cấu hình.
4. Từ **OLE DB Source Editor** chọn **[dbo].[stgDate]** table ở mục **name of the table or view**. Click OK để hoàn tất việc cấu hình.
5. Kéo rê **Slowly Changing Dimension** vào design surface. Đặt tên nó lại thành **DST - Slowly Changing Dimension DimDate** và connect **SRC -> DST**.



6. Double click vào **DST - Slowly Changing Dimension DimDate** để cấu hình.

7. Slowly Changing Dimension Wizard xuất hiện



Click **Next**.

8. Chọn bảng **DimDate** ở CSDL **DatacoSupplyChainDW**. Ở mục **connection manager** chọn CSDL **DatacoSupplyChainDW**.

9. Ở mục **Table or view** chọn bảng [**DimDate**].

10. Chọn **DateKey** là **business key**. Tất cả các cột khác là “**Not a key column**”.

Slowly Changing Dimension Wizard

Select a Dimension Table and Keys

Select a dimension table to load and map columns in the transformation input to

Connection manager:

MSI\THIN.DatacoSupplyChainDWStage

Table or view:

[dbo].[stgDate]

Input Columns	Dimension Columns	Key Type
date_key	date_key	Business key
day_abbrev	day_abbrev	Not a key column
day_name_of...	day_name_of_week	Not a key column
day_num_of...	day_num_of_month	Not a key column
day_num_of...	day_num_of_week	Not a key column
day_num_of...	day_num_of_year	Not a key column
fiscal_month	fiscal_month	Not a key column
fiscal_quarter	fiscal_quarter	Not a key column

Help < Back Next >| Finish >> Cancel

Click **Next**.

11. Cấu hình SCD type cho dimension. Các lựa chọn là:

- **Fixed → Không thay đổi**
- **Changing → Type 1 (Update)**
- **Historical → Type 2 (thêm dòng mới, dòng cũ inactive)**

Slowly Changing Dimension Wizard

Slowly Changing Dimension Columns

Manage the changes to column data in your slowly changing dimensions by setting the change type for dimension columns.

Fixed Attribute

Select this type when the value in a column should not change. Changes are treated as errors.

Changing Attribute

Select this type when changed values should overwrite existing values. This is a Type 1 change.

Historical Attribute

Select this type when changes in column values are saved in new records. Previous values are saved in records marked as outdated. This is a Type 2 change.

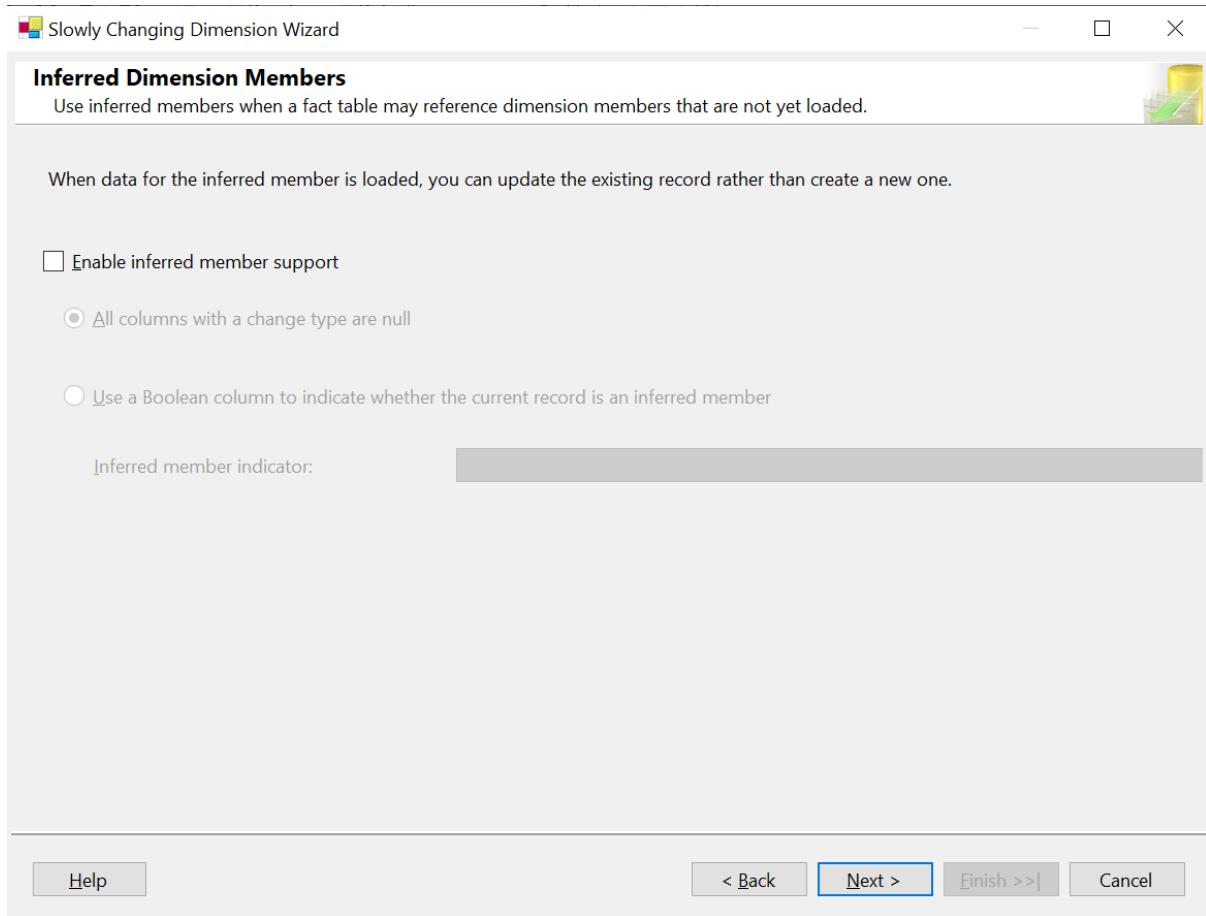
Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
day_abbrev	Changing attribute
day_name_of_week	Changing attribute
day_num_of_month	Changing attribute
day_num_of_week	Changing attribute
day_num_of_year	Changing attribute
fiscal_month	Changing attribute
fiscal_quarter	Changing attribute
fiscal_year	Changing attribute
full_date	Changing attribute
last_day_in_month...	Changing attribute
month	Changing attribute
month_abbrev	Changing attribute
month_name	Changing attribute
month_num_of_year	Changing attribute
quarter	Changing attribute
same_day_year_ag...	Changing attribute
week_begin_date	Changing attribute
week_begin_date_...	Changing attribute
week_num_of_year	Changing attribute
week_num_overall	Changing attribute
weekday_flag	Changing attribute
year	Changing attribute
yearmo	Changing attribute

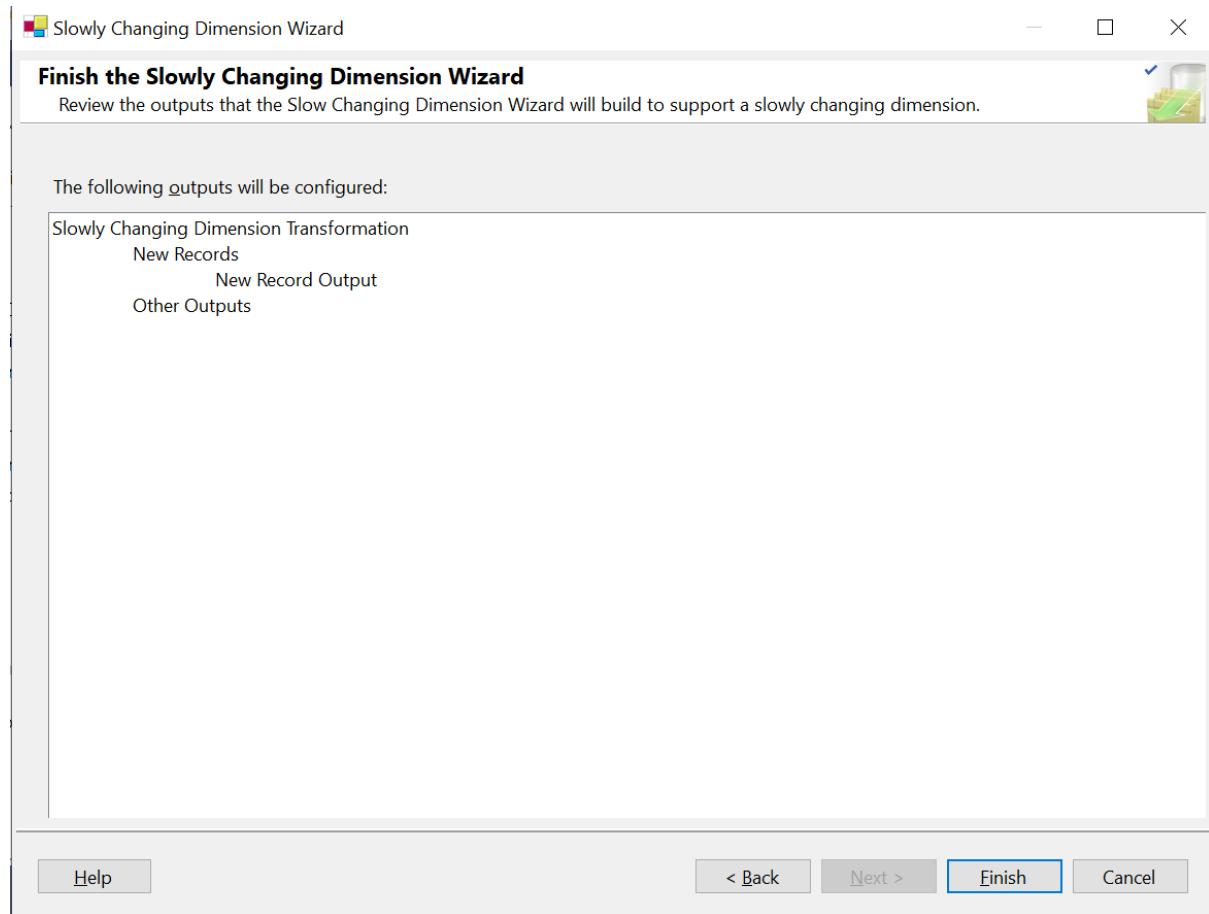
Cấu hình tất cả non-business key là changing attribute.

12. Click Next, xuất hiện màn hình **Fixed and Changing Attribute Options**. Click **Next**.

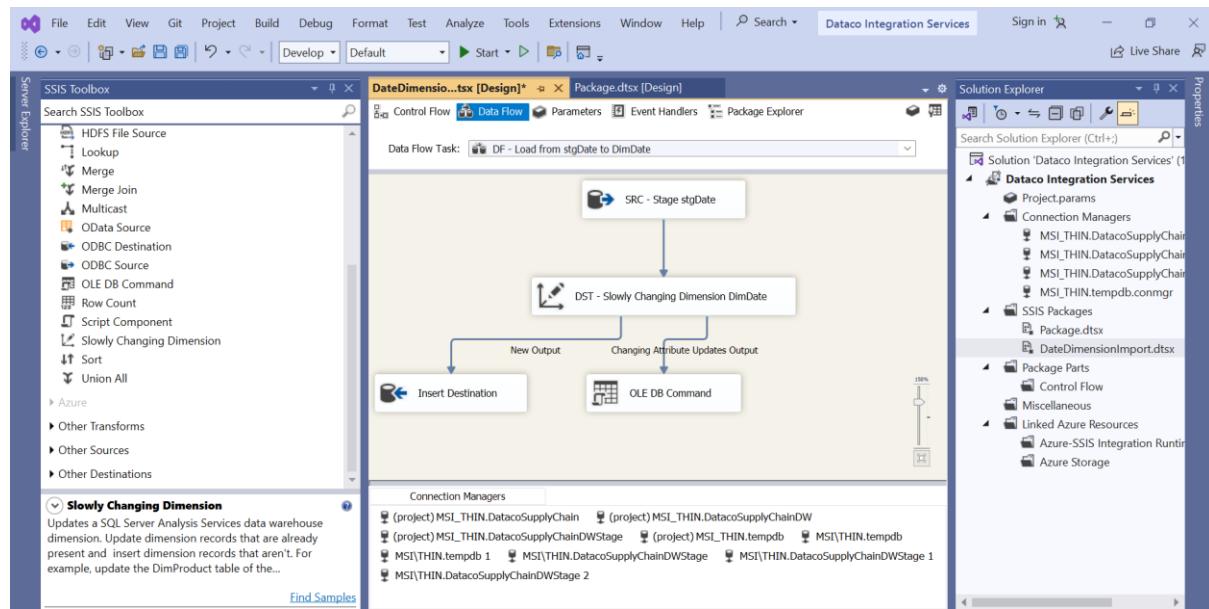
13. Xuất hiện màn hình **Inferred Dimension Members**. Bỏ chọn **check box** và click **Next**.



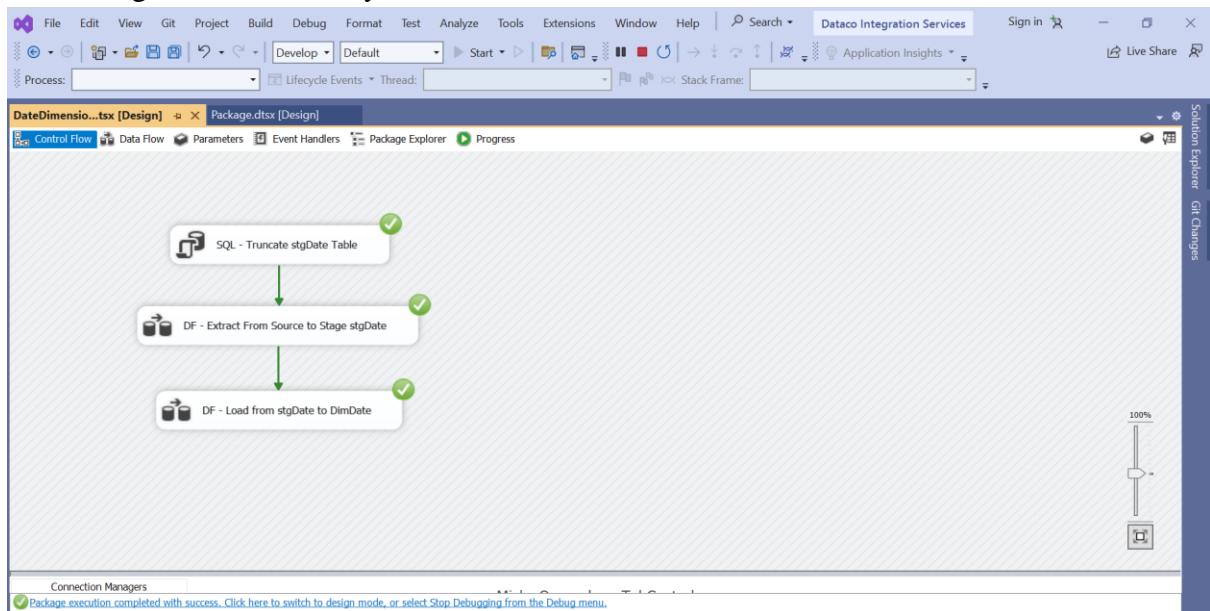
14. Xuất hiện màn hình Finish



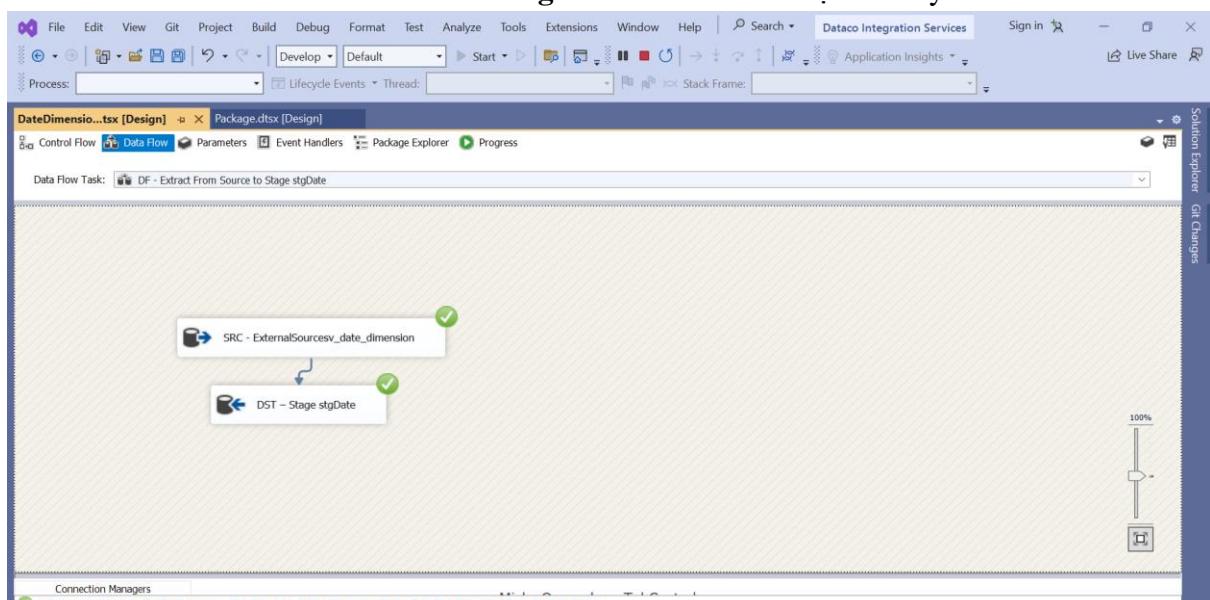
Click **Finish**.



15. Bấm **F5** để chạy. Nếu xảy ra lỗi, có găng đọc hiểu thông báo lỗi để tìm cách sửa. Nếu không có lỗi bạn sẽ thấy màn hình như sau:

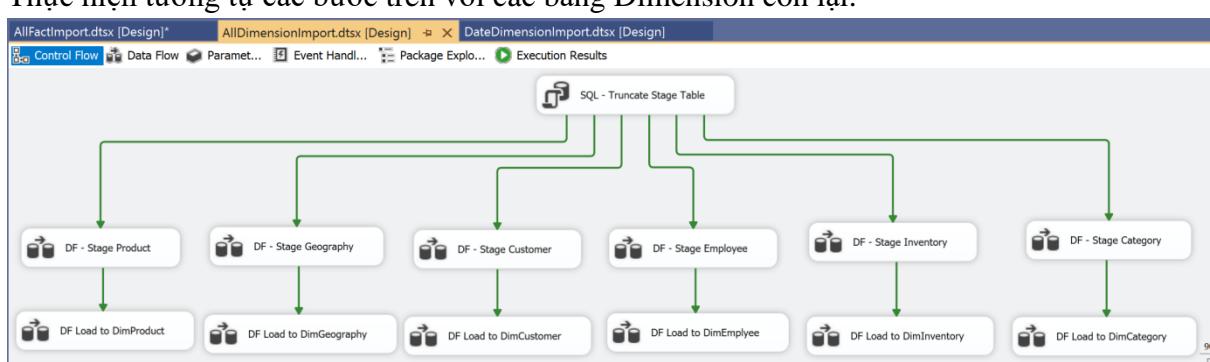


16. Double click vào **DF - Load From stgDate To DimDate** bạn sẽ thấy màn hình sau:



17. Nhấn **[Shift] + [F5]** để kết thúc thực thi.

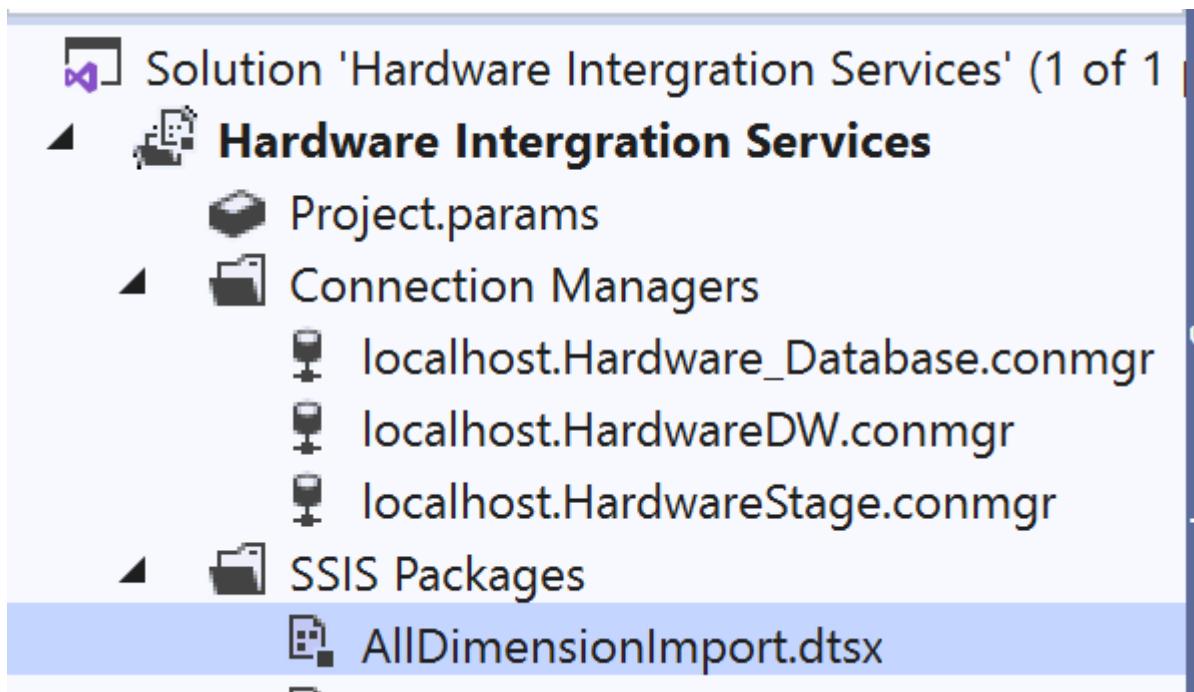
Thực hiện tương tự các bước trên với các bảng Dimension còn lại.



3.3.2. Product Dimension

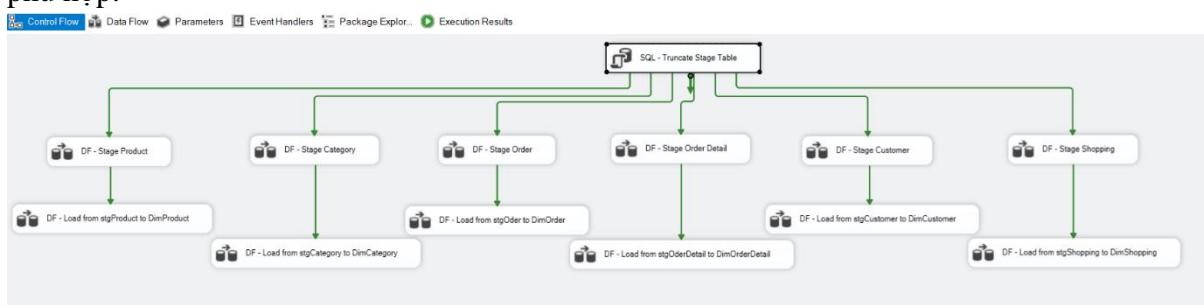
3.3.2.1. Tạo Package

- Tạo package mới và đổi tên như hình bên dưới.

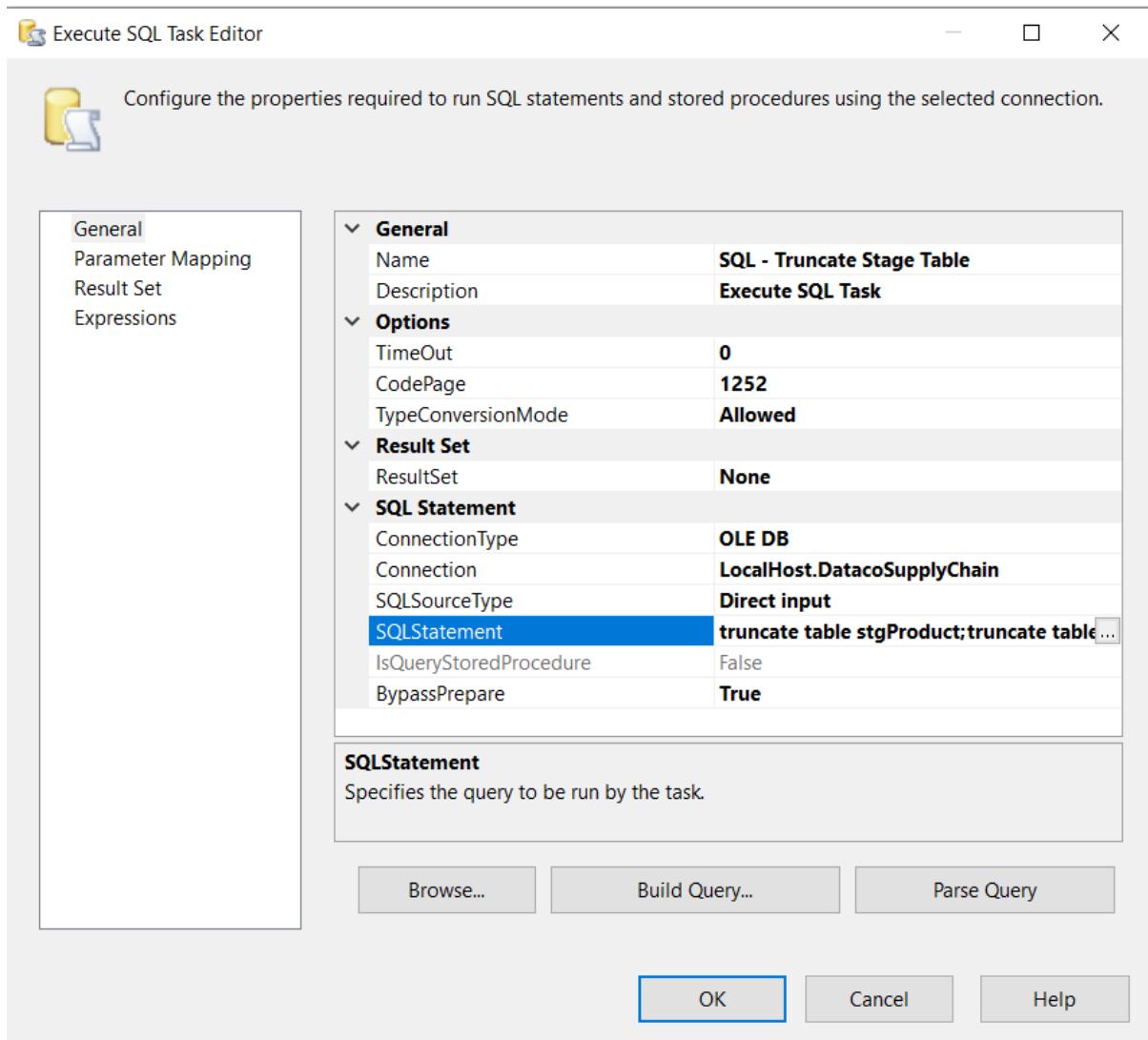


3.3.2.2. Thiết lập Control Flow

- Tạo và đặt tên tương ứng cho các task, tạo các connect bằng cách kéo mũi tên vào các task phù hợp.

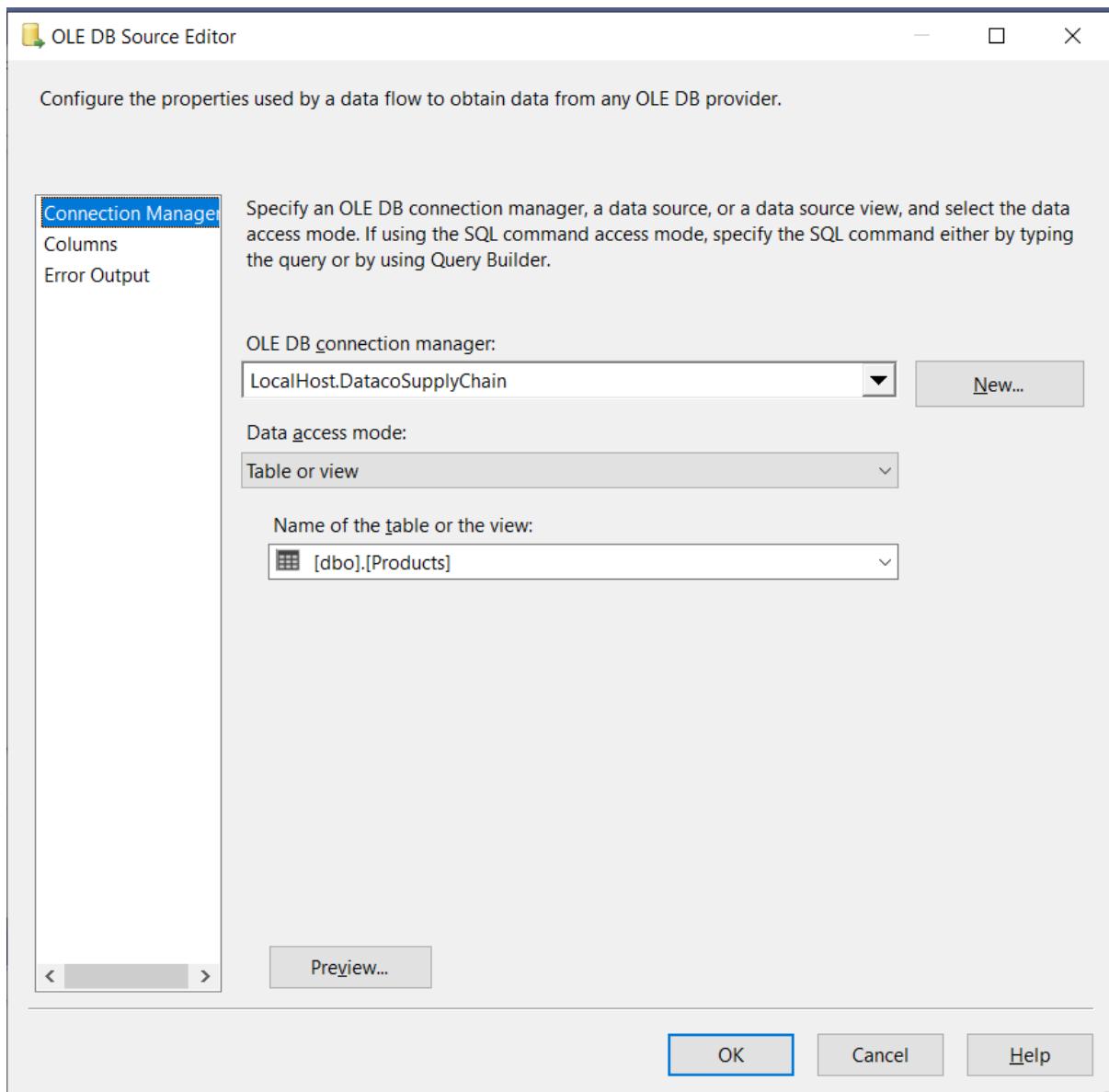


- Thiết lập truncate table SQL task

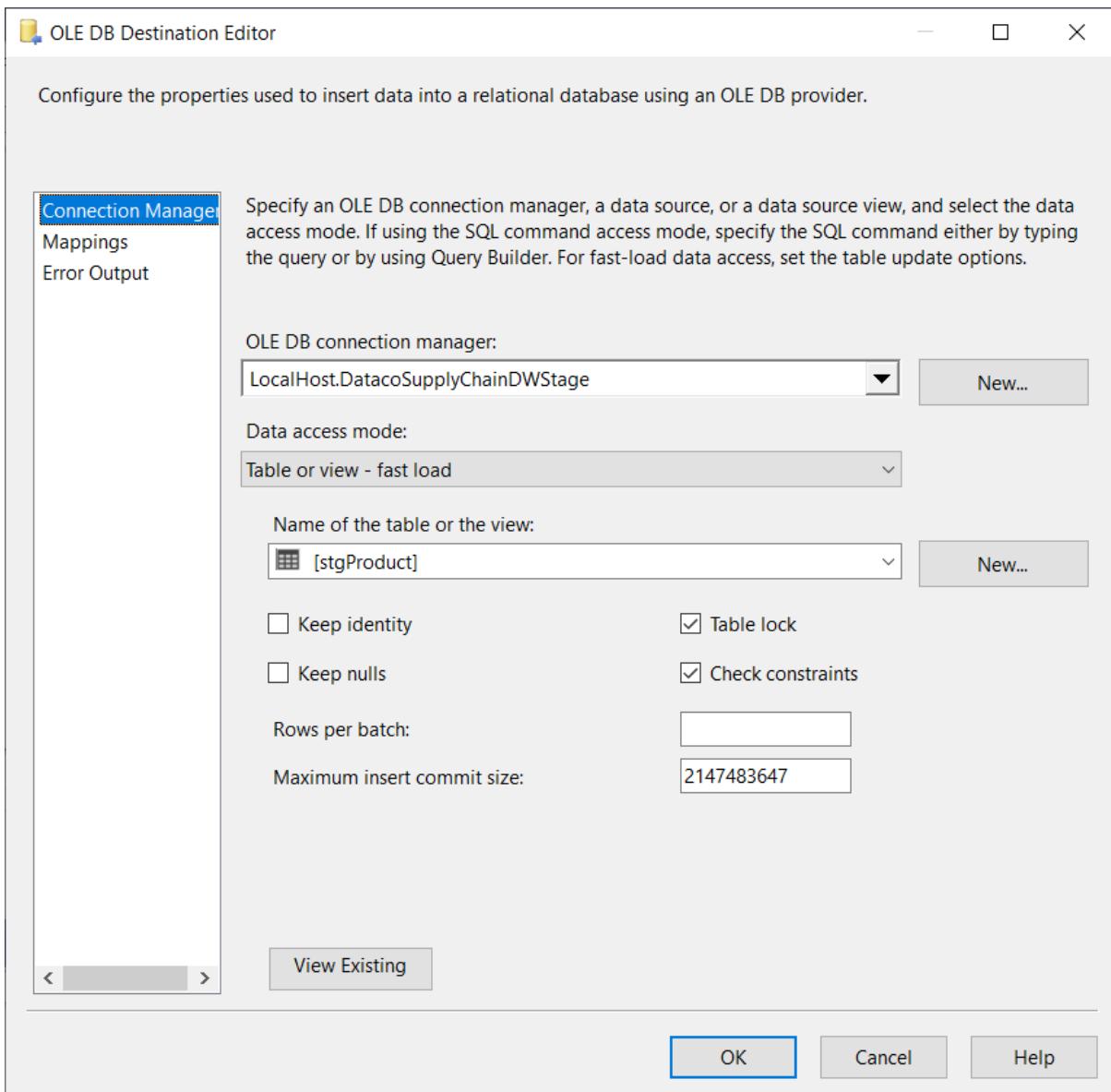


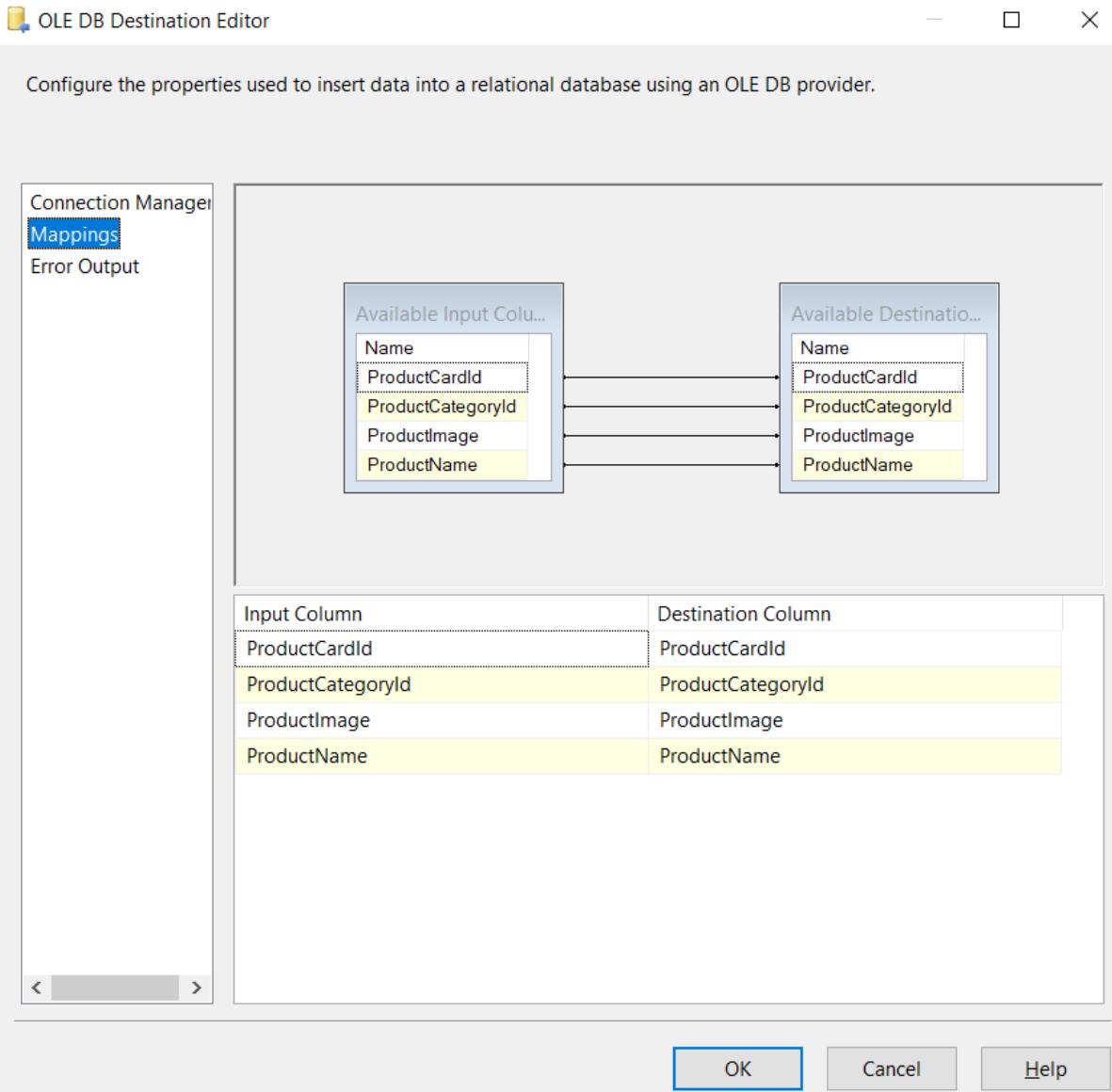
3.3.2.3. Load dữ liệu từ Source vào Stage

- Double-click DF - Stage Product task để mở nó trong Data Flow design surface:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
 - + Source Assistant:

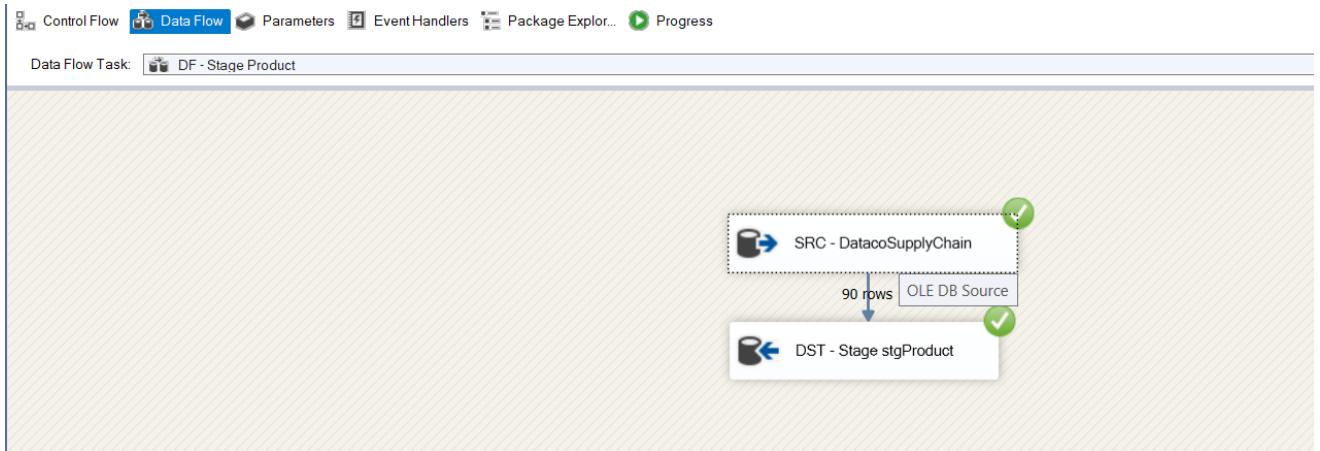


+ Destination Assistant:





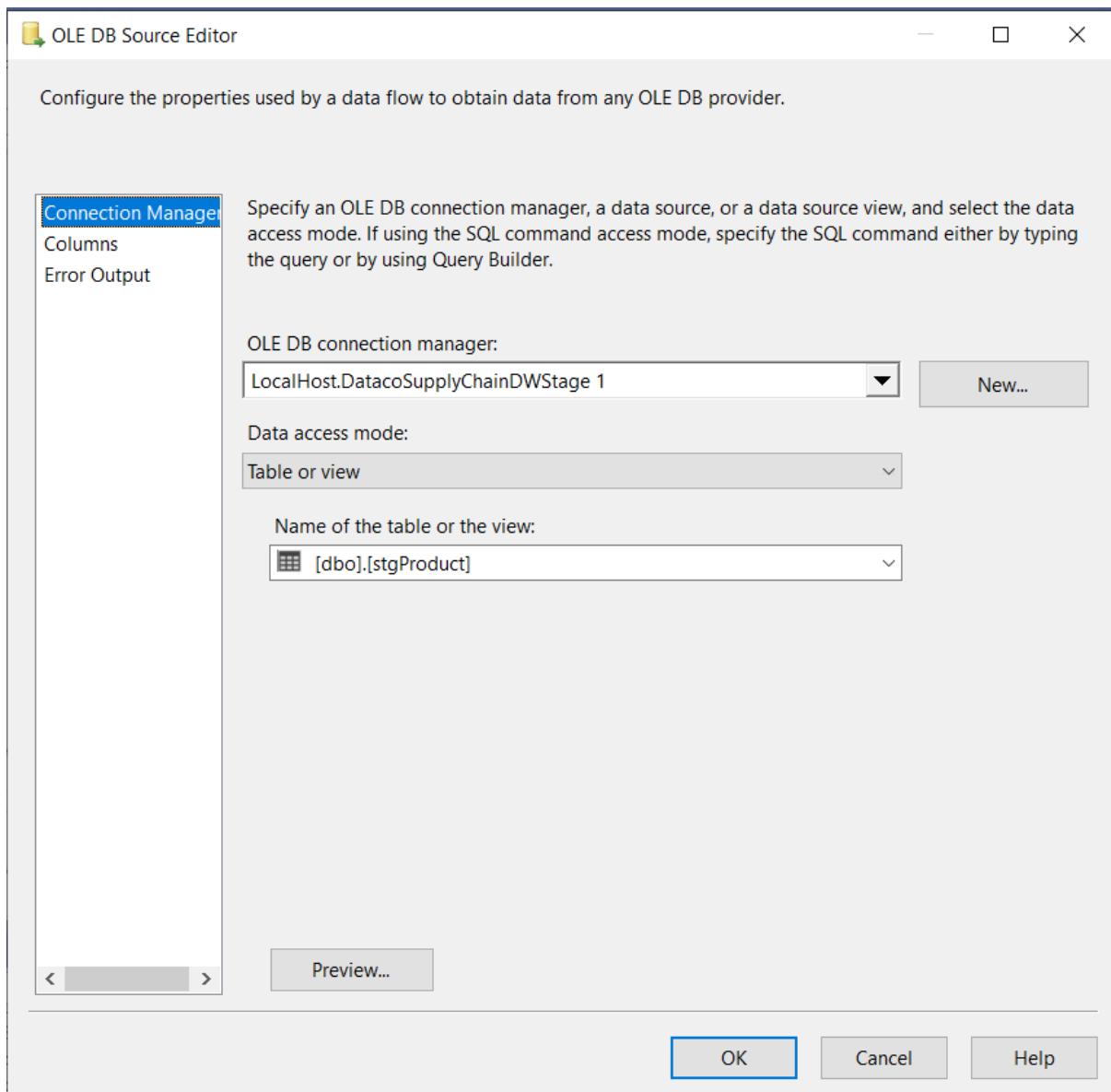
+ Sau khi cấu hình xong ta được:



3.3.2.4. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load from stgProduct to DimProduct** task để mở data flow design surface.

+ Source Assistant:



+ Slowly Changing Dimension:

Slowly Changing Dimension Wizard

Select a Dimension Table and Keys

Select a dimension table to load and map columns in the transformation input to

Connection manager:

localhost.DatacoSupplyChainDWStage ▼ New...

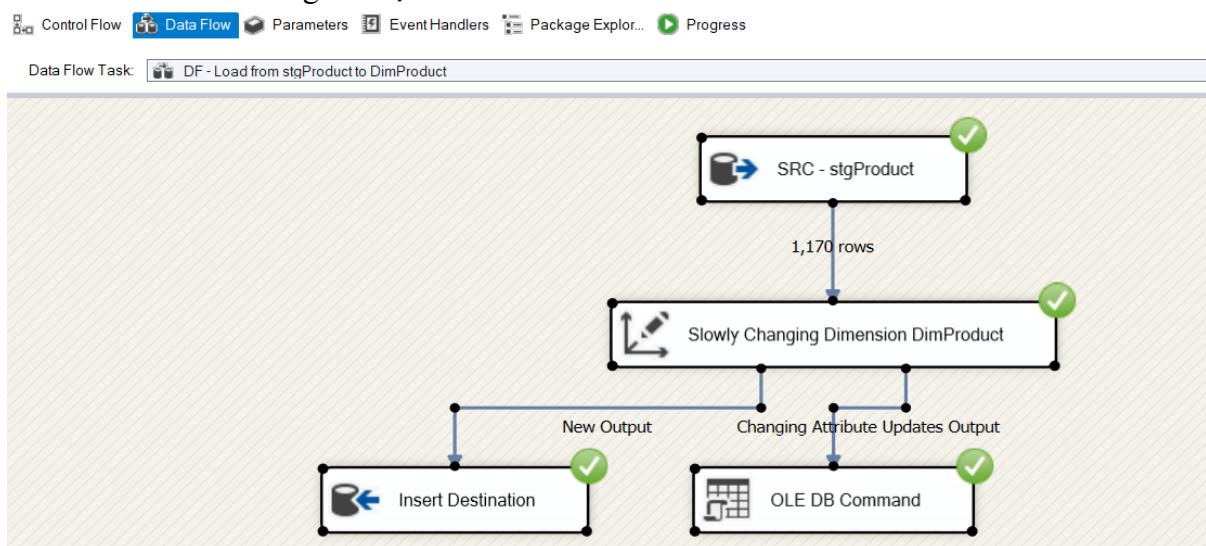
Table or view:

[dbo].[stgProduct] ▼

Input Columns	Dimension Columns	Key Type
ProductCardId	ProductCardId	Business key
ProductCate...	ProductCategoryId	Not a key column
ProductImage	ProductImage	Not a key column
ProductName	ProductName	Not a key column

Help Next > Finish >> Cancel

+ Sau khi cấu hình xong ta được:



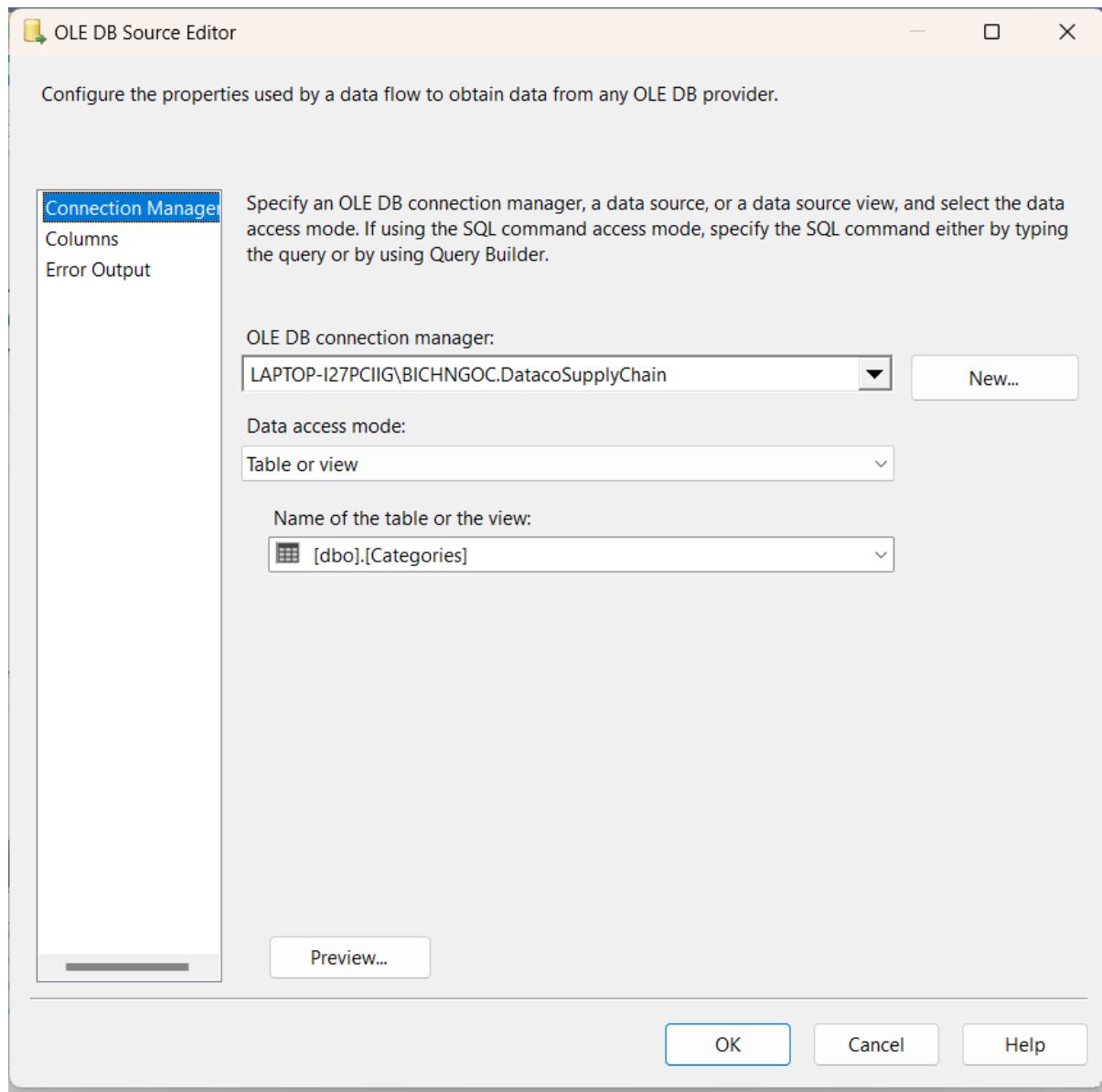
3.3.3.Category Dimension

3.3.3.1. Load dữ liệu từ Source vào Stage

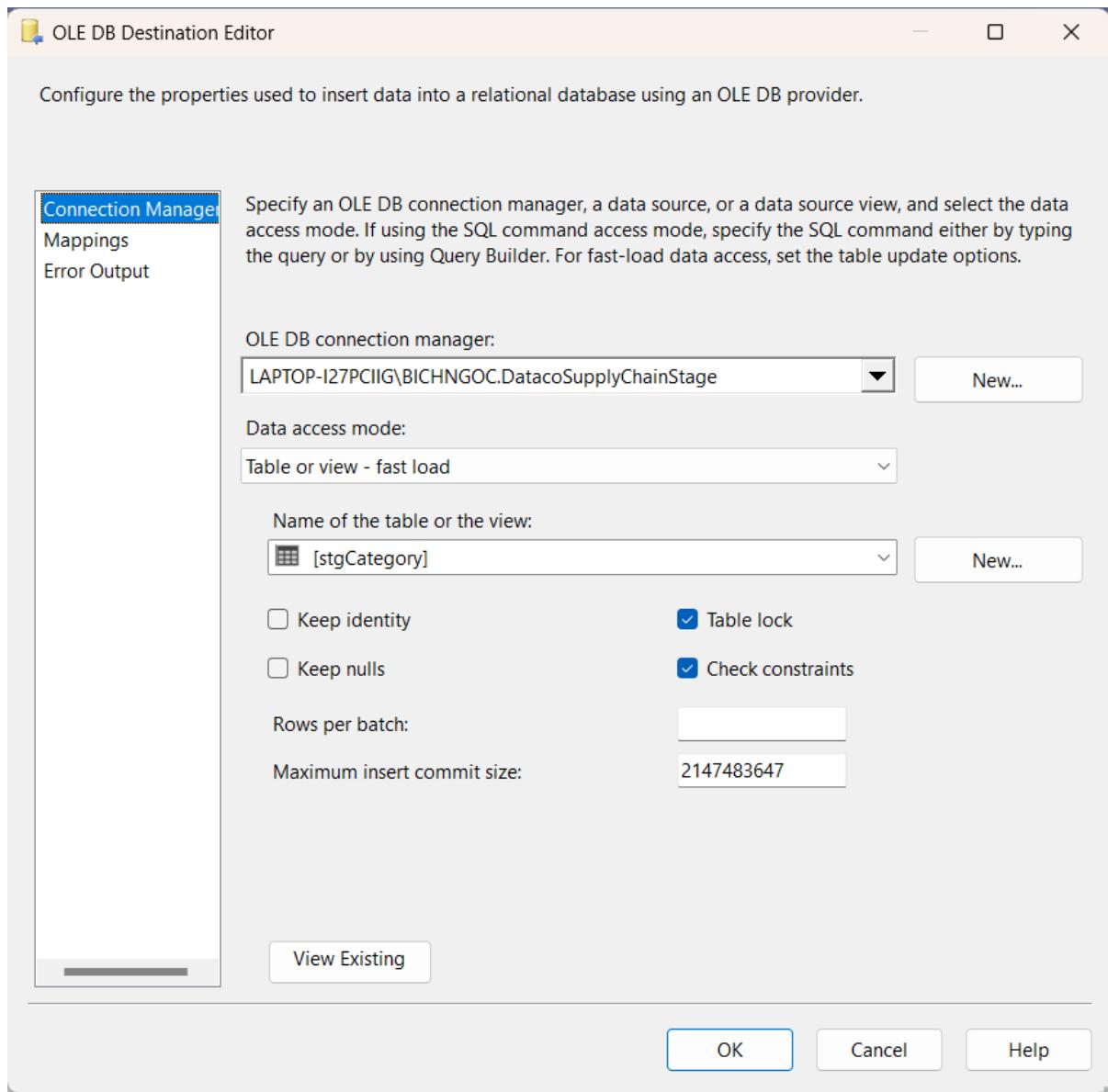
- Double-click DF - Extract From Source to Stage Category task để mở nó trong **Data Flow design surface**:

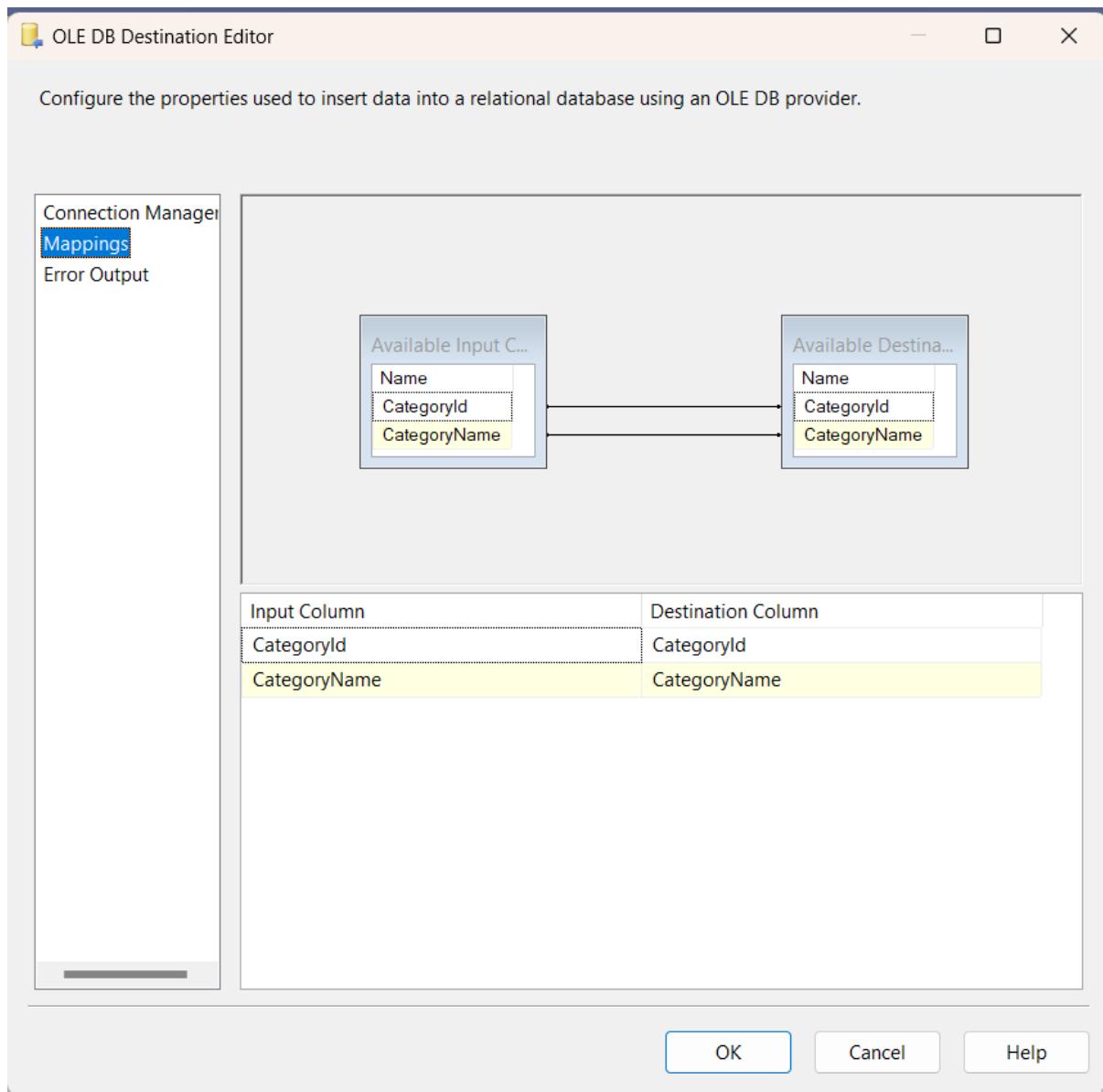
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

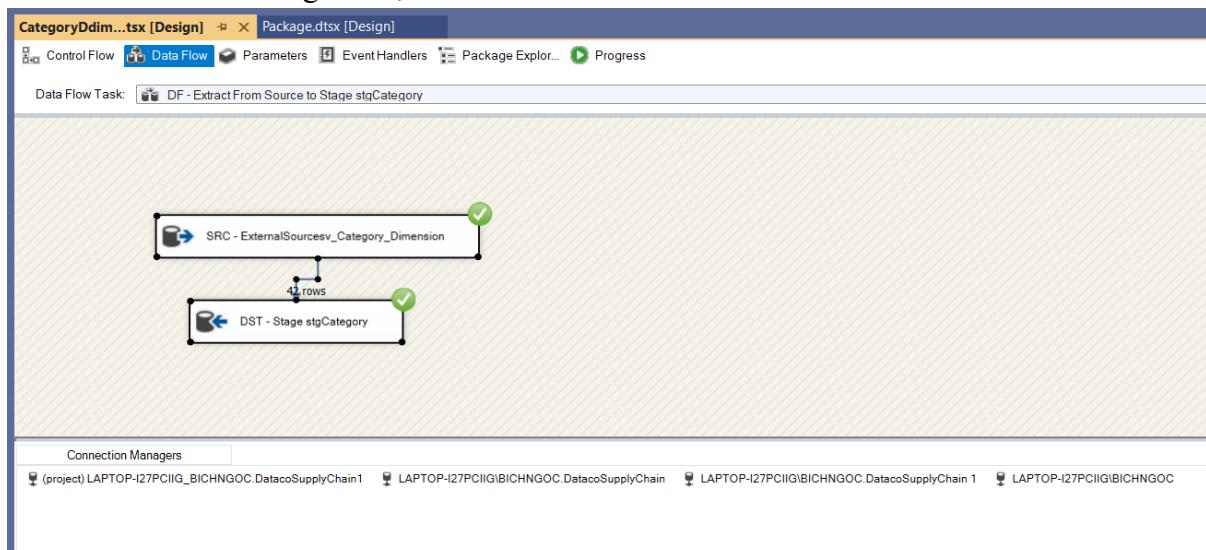


+ Destination Assistant:





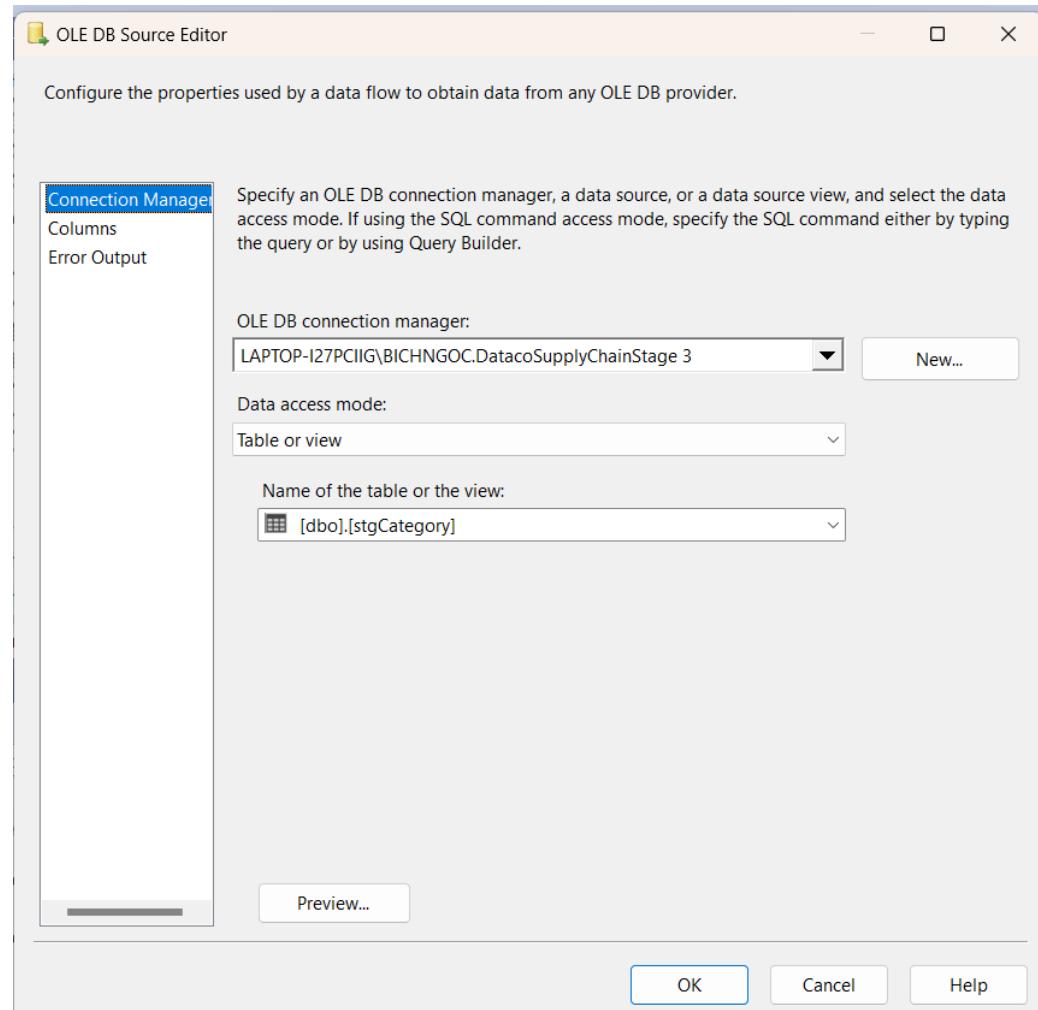
+ Sau khi cấu hình xong ta được:



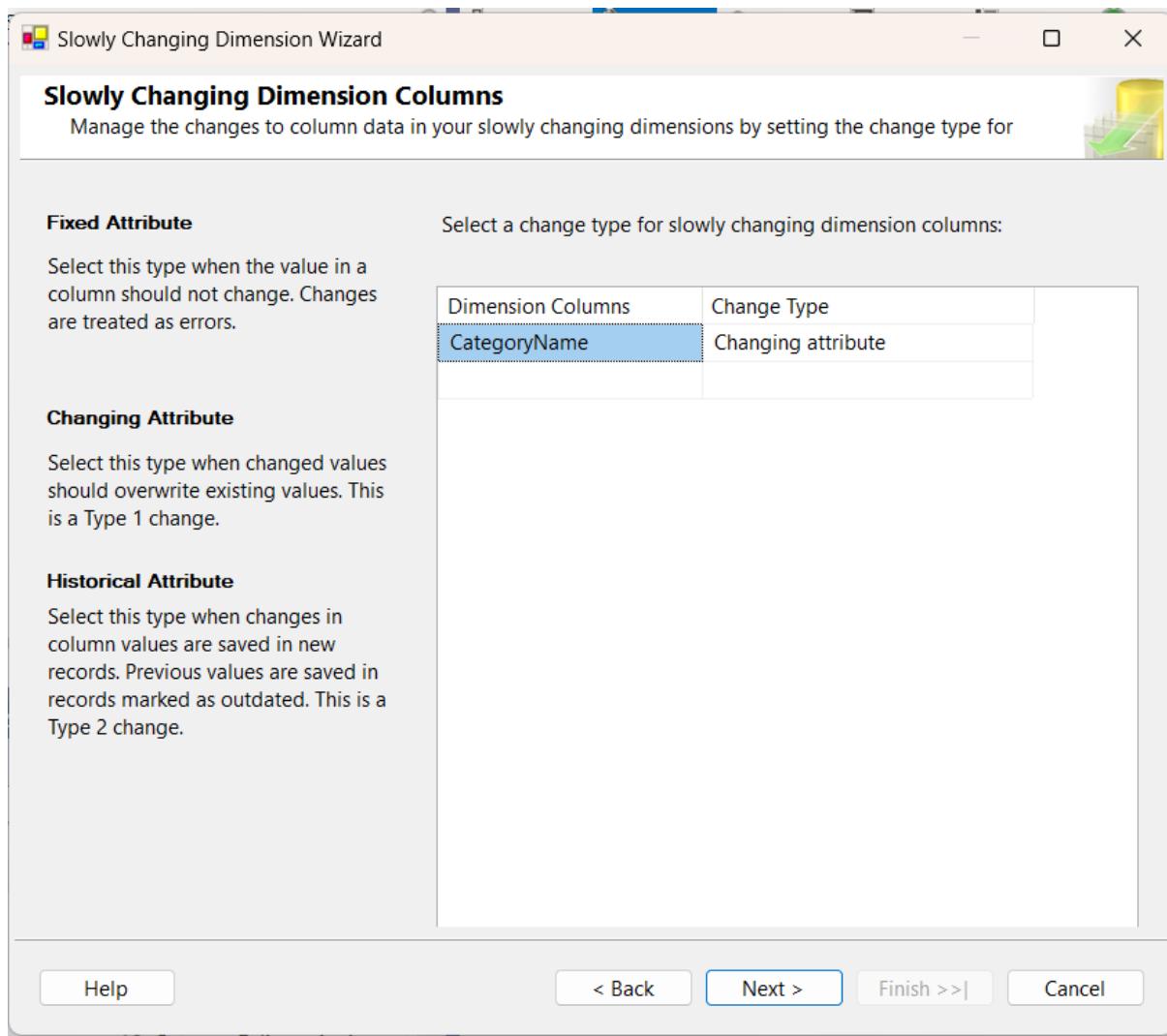
3.3.3.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load from stgCategory to DimCategory** task để mở data flow design surface.

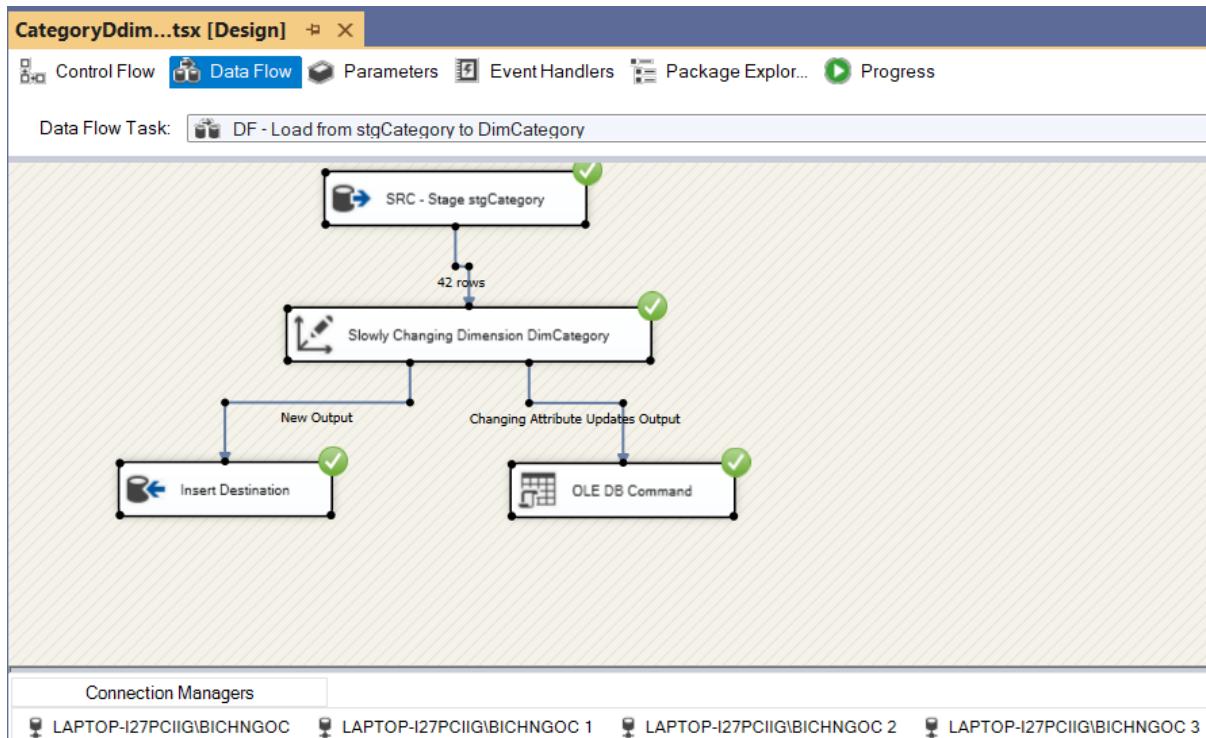
+ Source Assistant:



+ Slowly Changing Dimension:



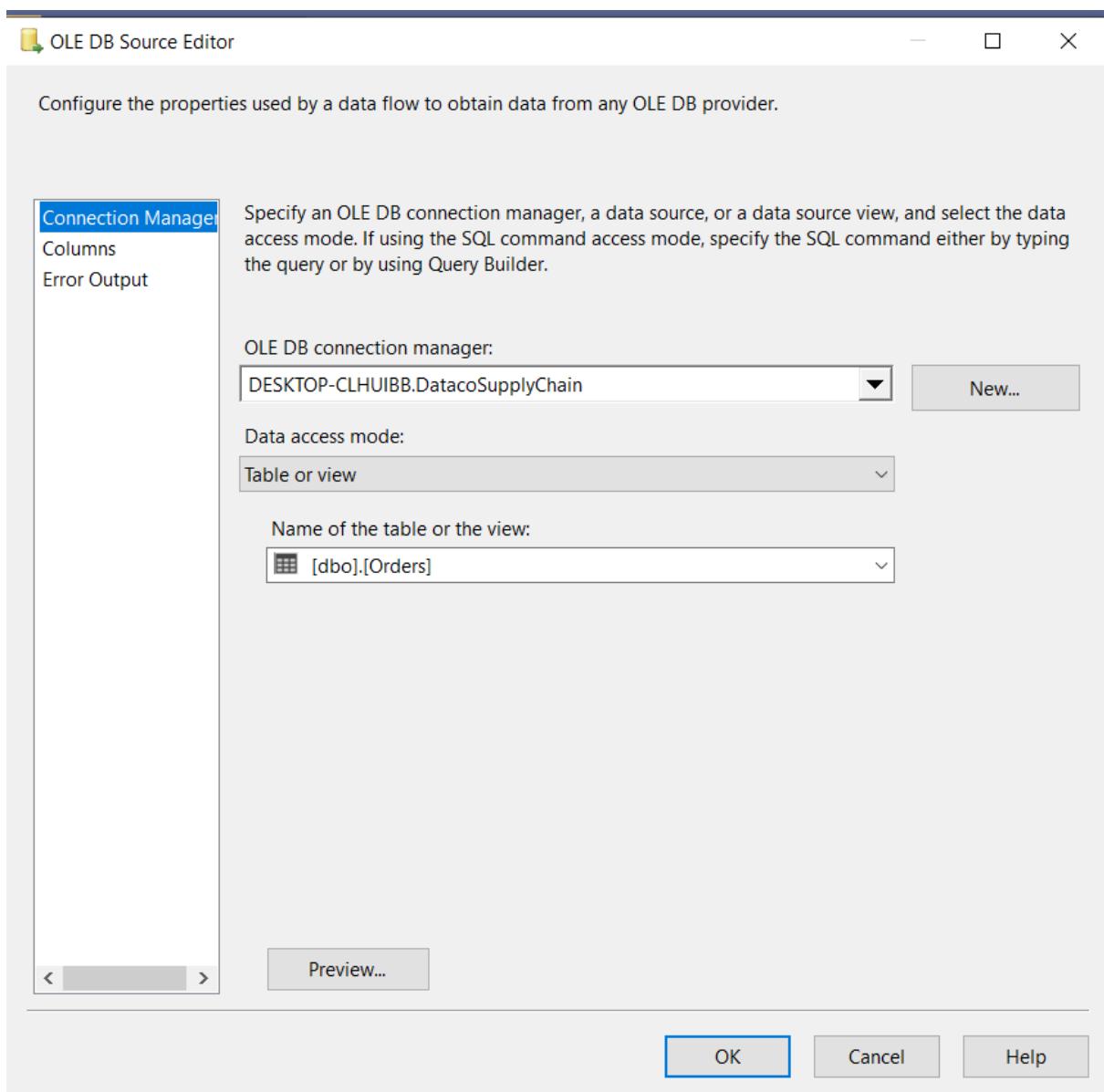
+ Sau khi cấu hình xong ta được:



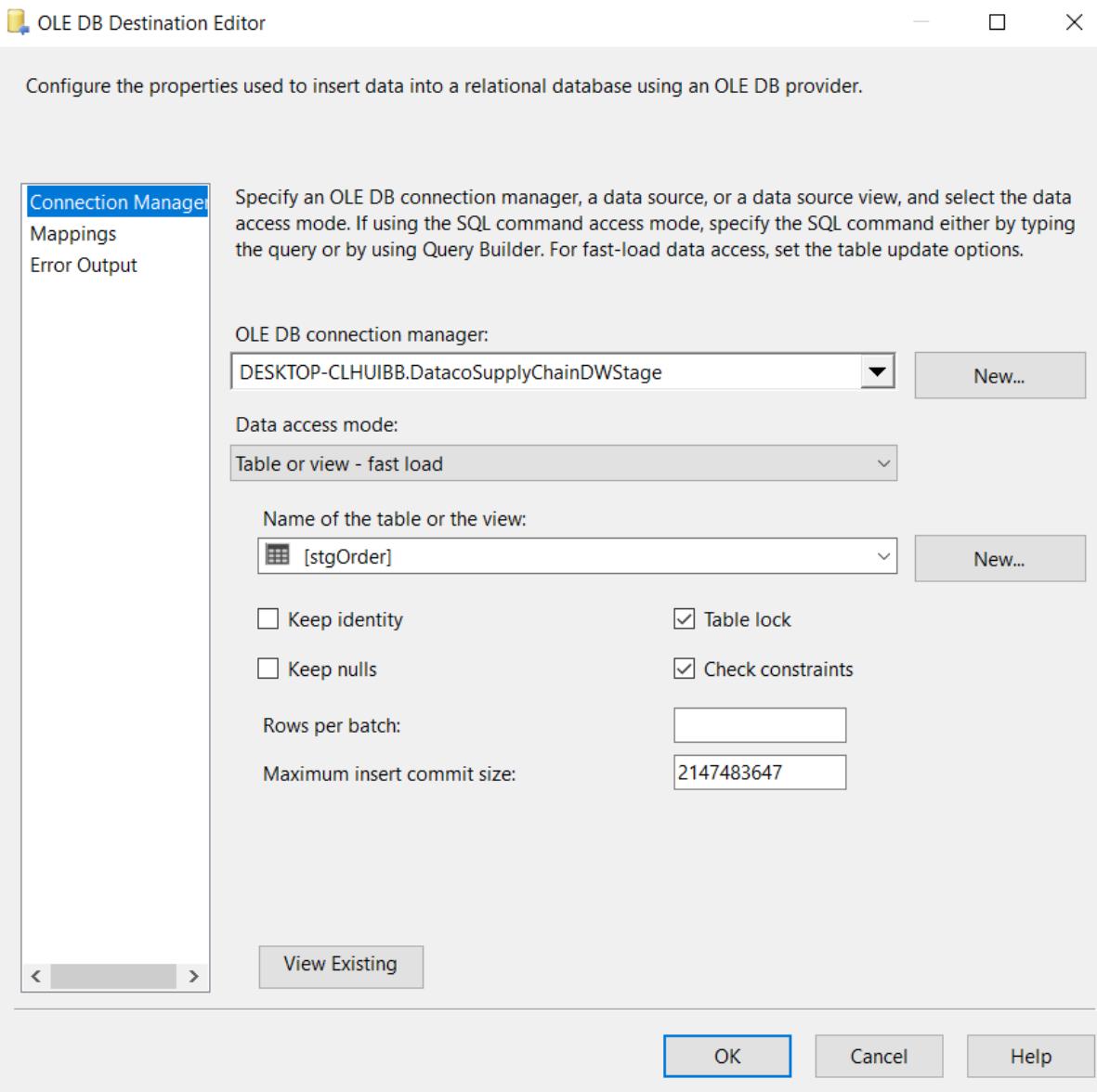
3.3.4. Order Dimension

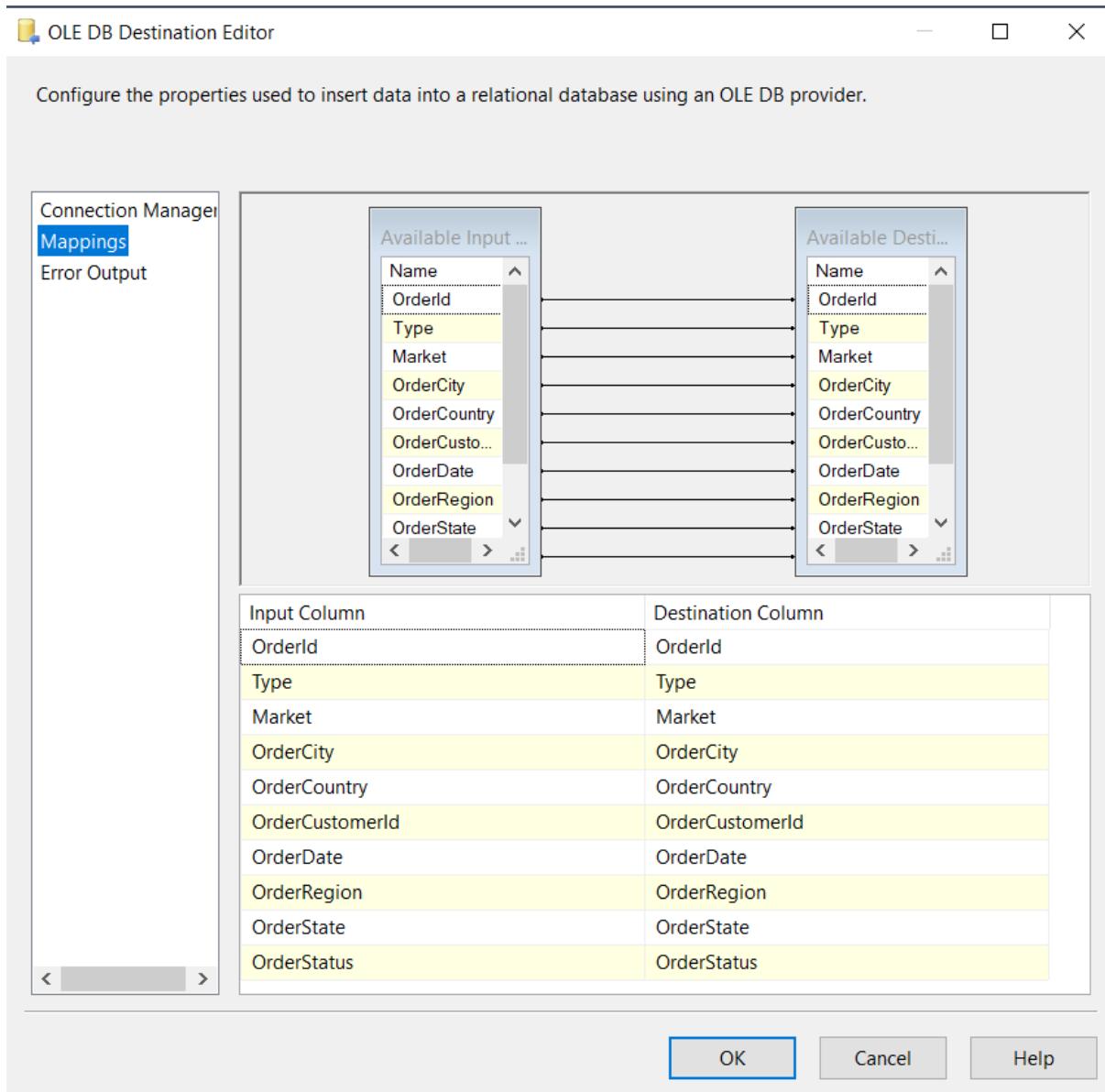
3.3.4.1. Load dữ liệu từ Source vào Stage

- Double-click **DF - Stage Customer** task để mở nó trong **Data Flow design surface**:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
- + **Source Assistant**:

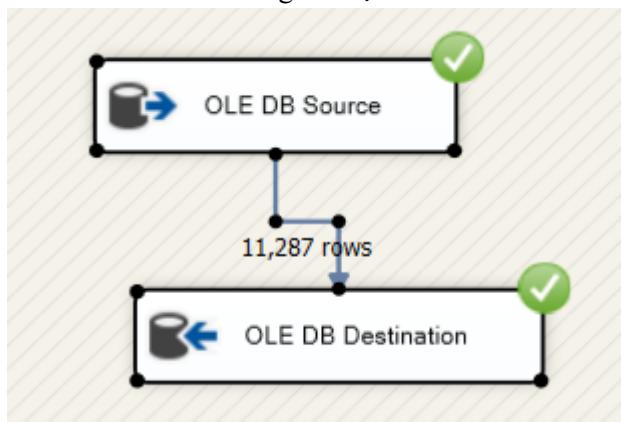


+ Destination Assistant:





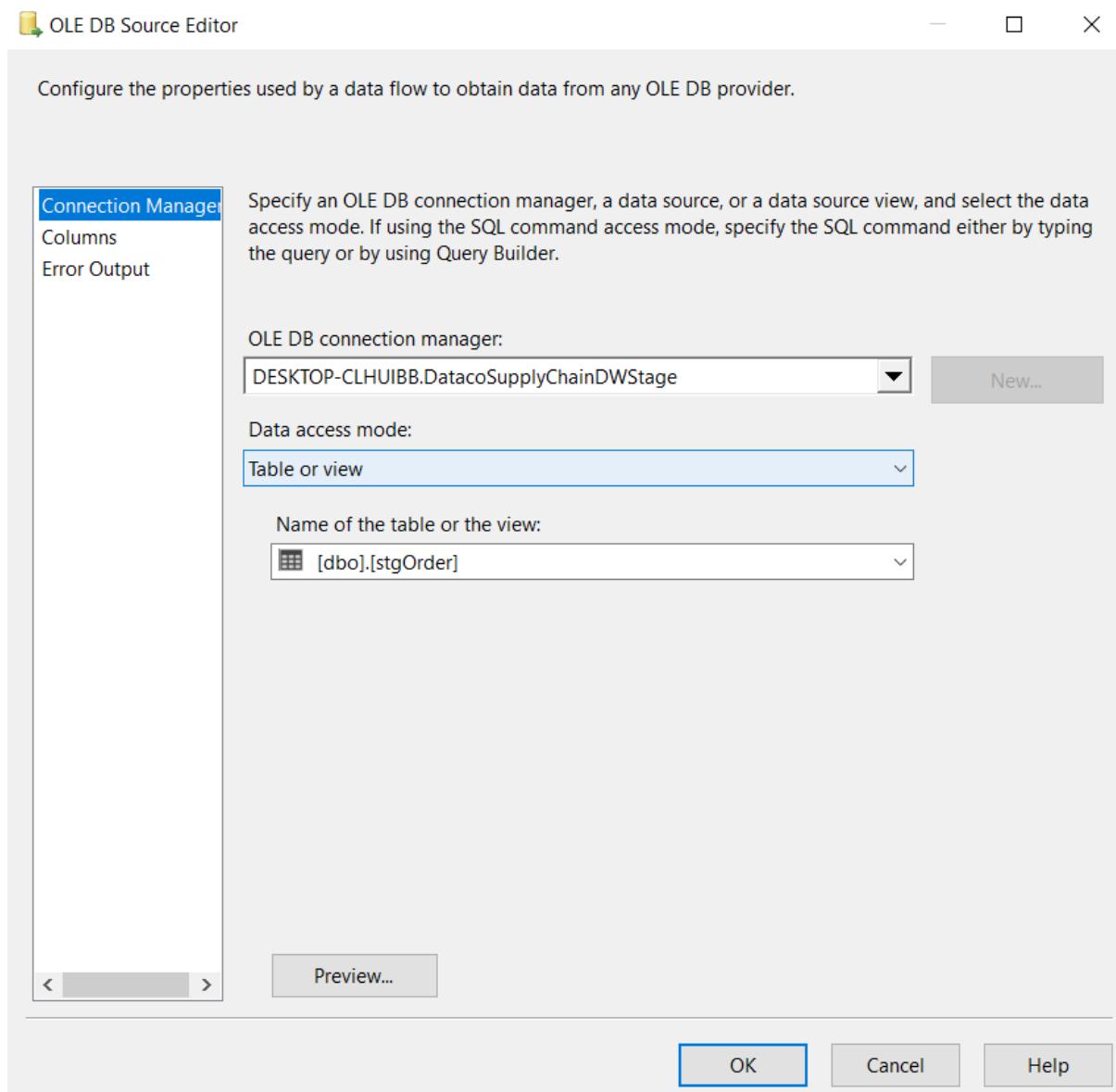
+ Sau khi cấu hình xong ta được:



3.3.4.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào DF - Load to DimOrder task để mở data flow design surface.

+ Source Assistant:



+ Slowly Changing Dimension:

Slowly Changing Dimension Wizard

Select a Dimension Table and Keys

Select a dimension table to load and map columns in the transformation input to

Connection manager:

DESKTOP-CLHUIBB.DatacoSupplyChainDW

Table or view:

[dbo].[DimOrder]

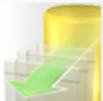
Input Columns	Dimension Columns	Key Type
Market	Market	Not a key column
OrderCity	OrderCity	Not a key column
OrderCountry	OrderCountry	Not a key column
OrderCustom...	OrderCustomerId	Not a key column
OrderDate	OrderDate	Not a key column
OrderId	OrderID	Business key

Help

 Slowly Changing Dimension Wizard

Slowly Changing Dimension Columns

Manage the changes to column data in your slowly changing dimensions by setting the



Fixed Attribute

Select this type when the value in a column should not change. Changes are treated as errors.

Changing Attribute

Select this type when changed values should overwrite existing values. This is a Type 1 change.

Historical Attribute

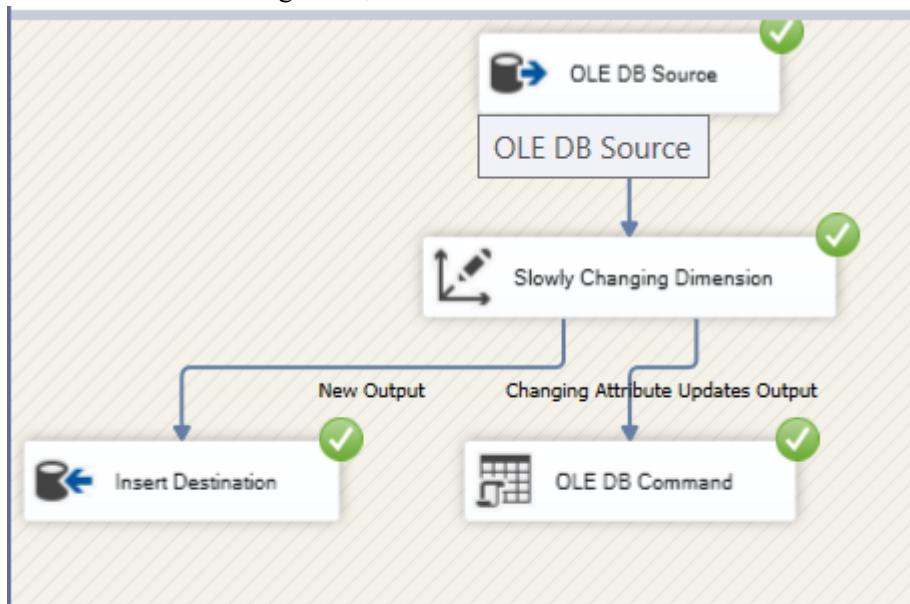
Select this type when changes in column values are saved in new

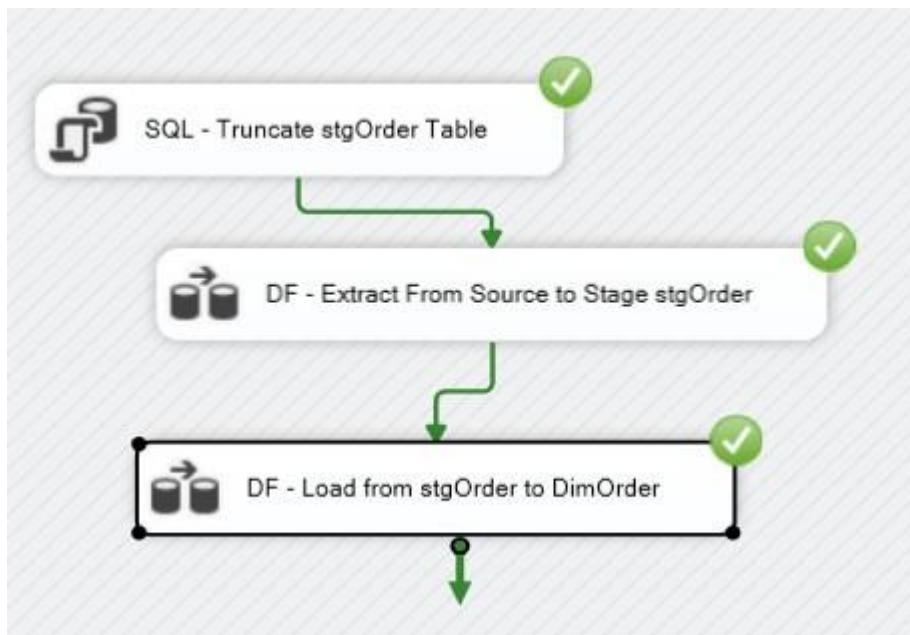
Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
Market	Changing a...
OrderCity	Changing a...
OrderCountry	Changing a...
OrderCustomerId	Changing a...
OrderDate	Changing a...
OrderRegion	Changing a...
OrderState	Changing a...
OrderStatus	Changing a...
Type	Changing a...

Help < Back Next >| Finish >> Cancel

+ Sau khi cấu hình xong ta được:

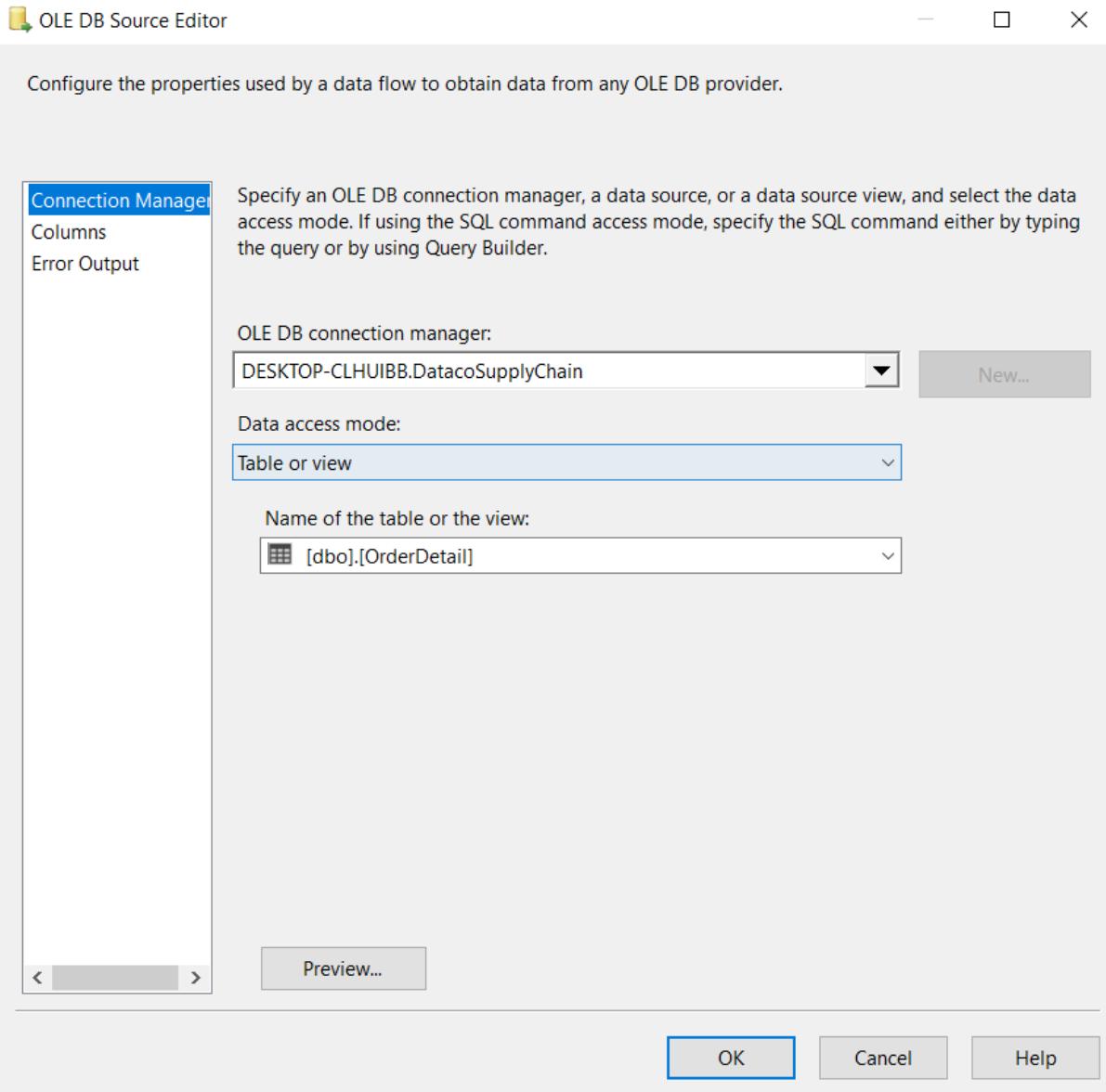




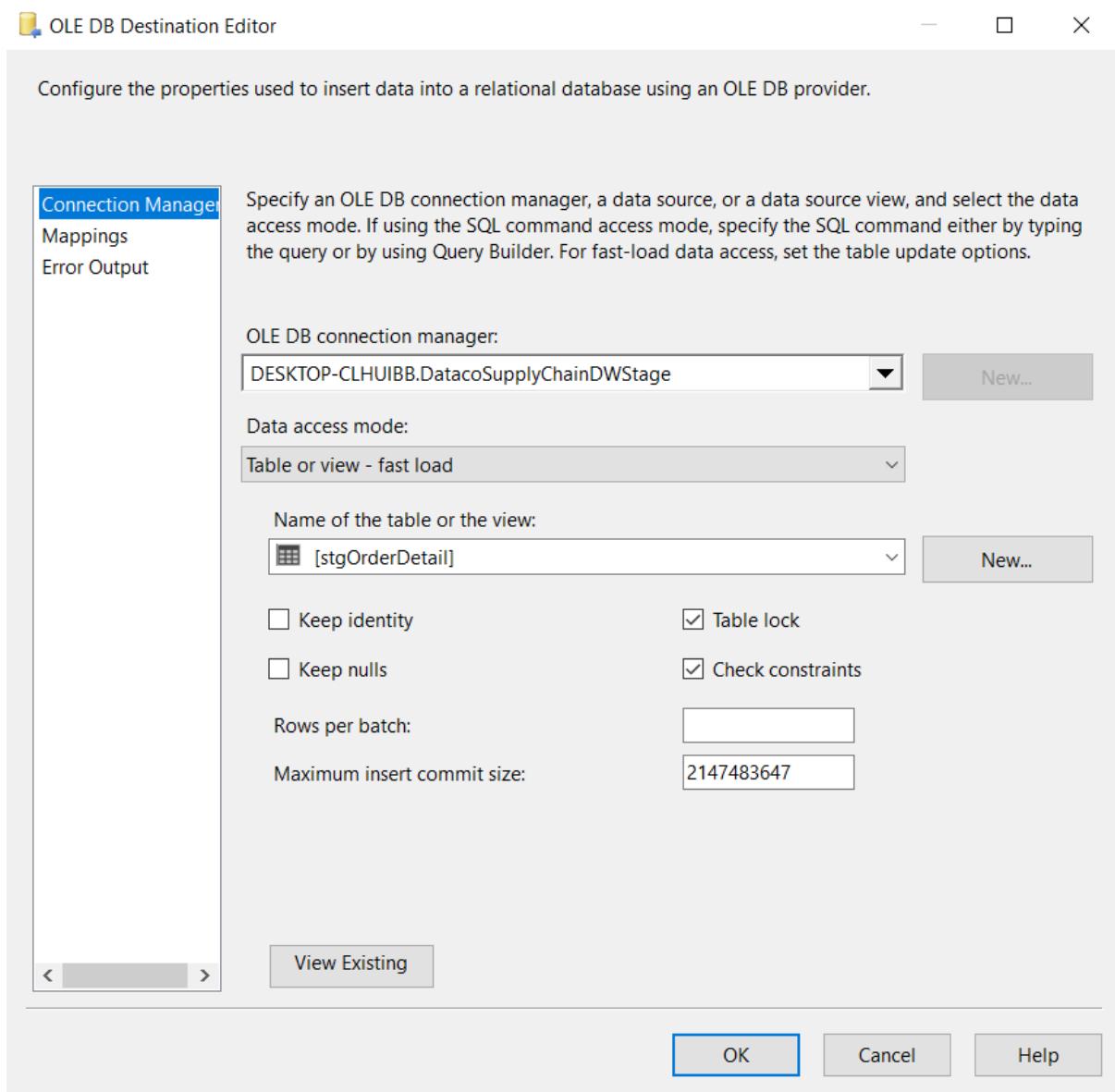
3.3.5. Order Detail Dimension

3.3.5.1. Load dữ liệu từ Source vào Stage

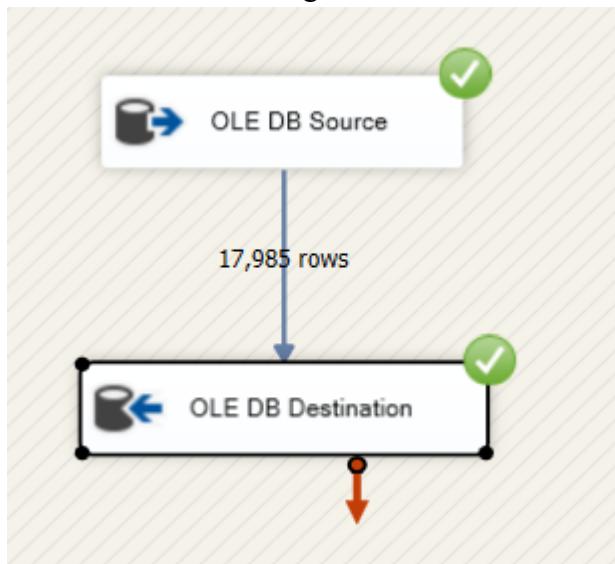
- Double-click **DF - Stage Employee** task để mở nó trong **Data Flow design surface**:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
- + **Source Assistant:**



+ Destination Assistant:

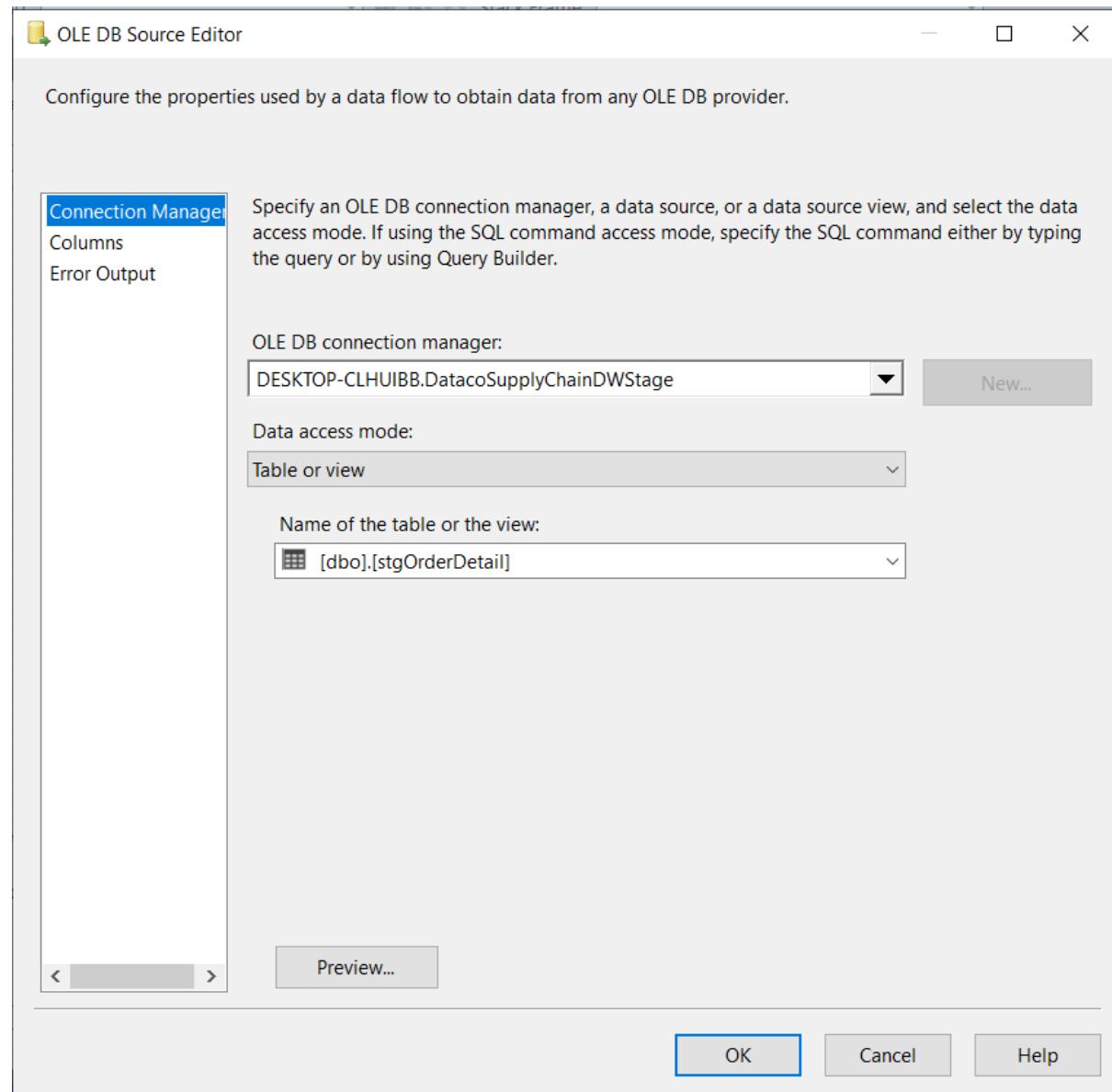


+ Sau khi cấu hình xong ta được:



3.3.5.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load to DimOrderDetail** task để mở data flow design surface.
- + Source Assistant:



- + Slowly Changing Dimension:

Select a Dimension Table and Keys

Select a dimension table to load and map columns in the transformation input to

Connection manager:

DESKTOP-CLHUIBB.DatacoSupplyChainDW

Table or view:

[dbo].[DimOrderDetail]

Input Columns	Dimension Columns	Key Type
OrderId	OrderId	Business key
OrderItemCa...	OrderItemCardpro...	Business key
OrderItemDis...	OrderItemDiscount	Business key
OrderItemDis...	OrderItemDiscount...	Not a key column
OrderItemId	OrderItemId	Not a key column
OrderItemPr...	OrderItemProduct...	Not a key column

Slowly Changing Dimension Wizard

Manage the changes to column data in your slowly changing dimensions by setting the

Fixed Attribute
Select this type when the value in a column should not change. Changes are treated as errors.

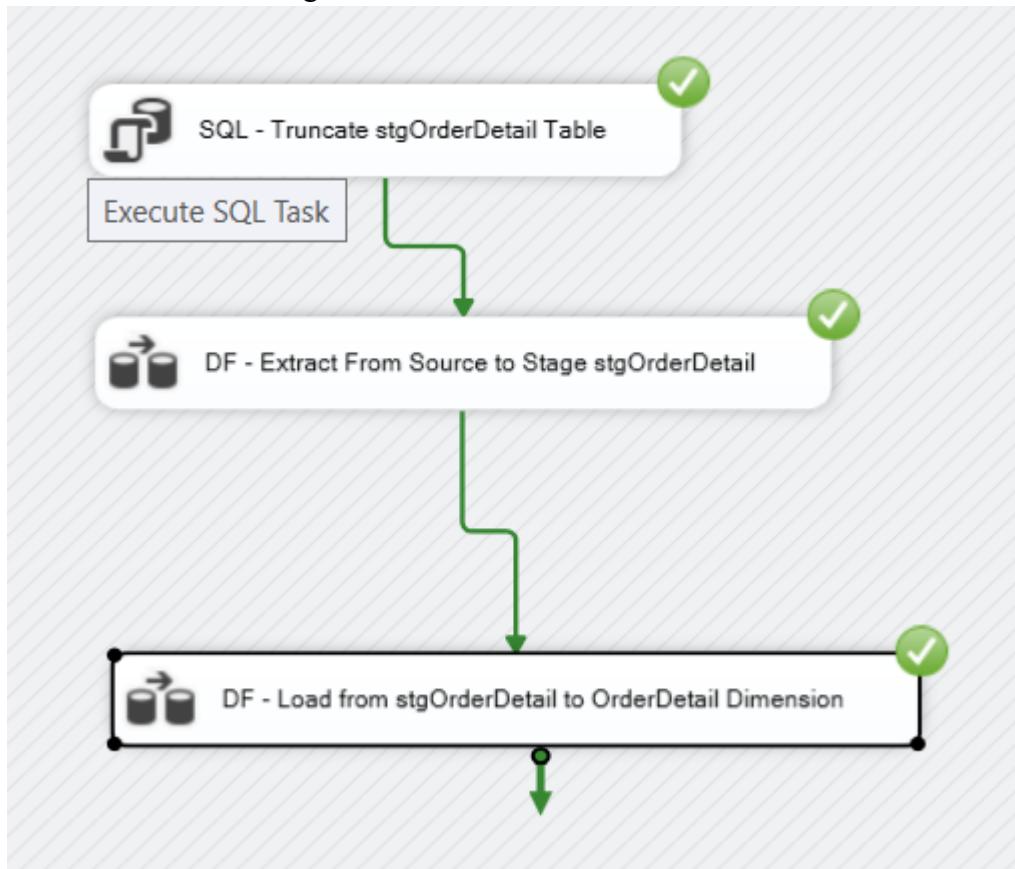
Changing Attribute
Select this type when changed values should overwrite existing values. This is a Type 1 change.

Historical Attribute
Select this type when changes in column values are saved in new

Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
OrderItemDiscount...	Changing a...
OrderItemId	Changing a...
OrderItemProduct...	Changing a...
OrderItemProfitRat...	Changing a...
OrderItemQuantity	Changing a...
OrderItemTotal	Changing a...
OrderProfitPerOrder	Changing a...
Sales	Changing a...

+ Sau khi cấu hình xong ta được:



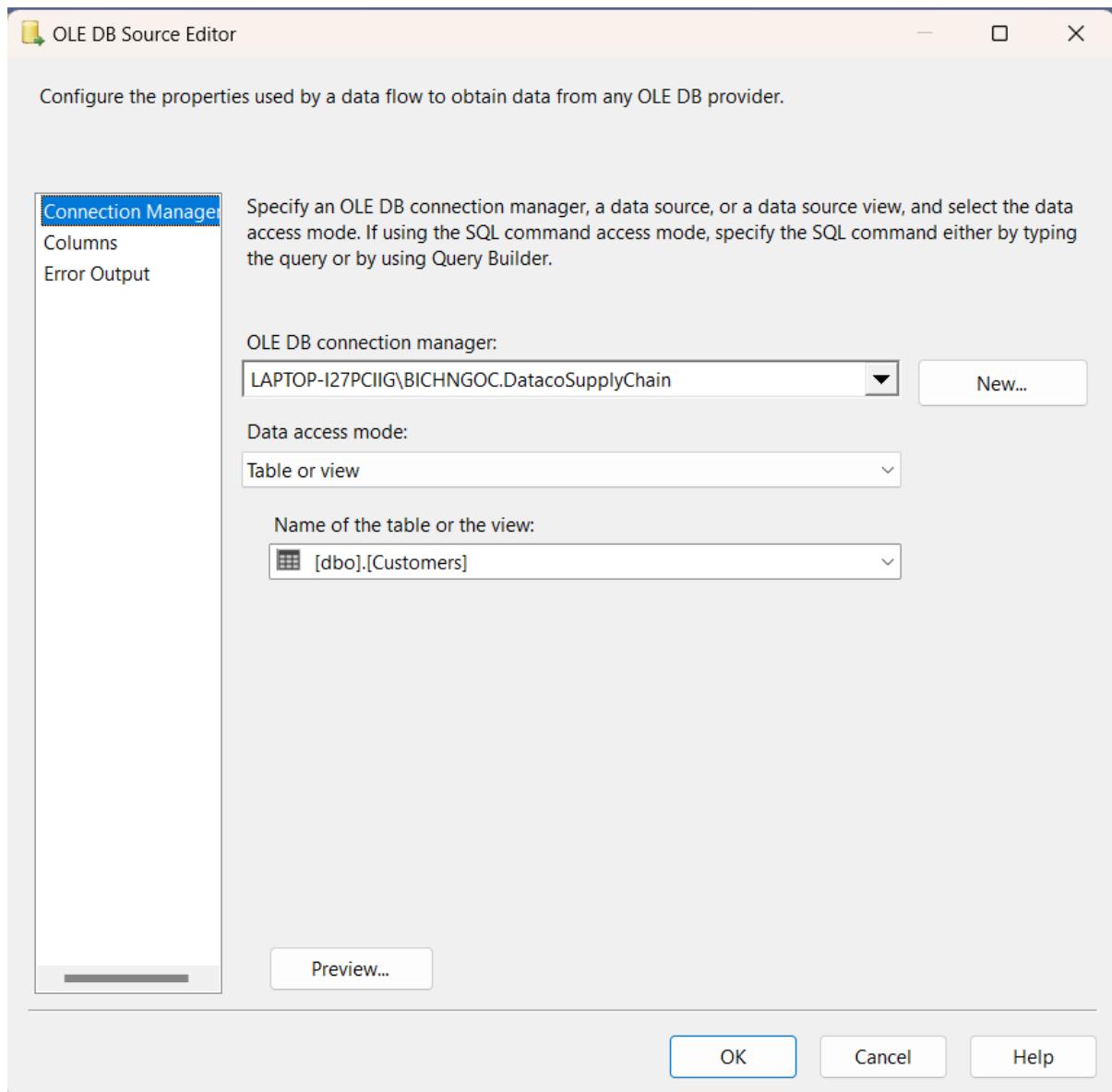
3.3.6. Customer Dimension

3.3.6.1. Load dữ liệu từ Source vào Stage

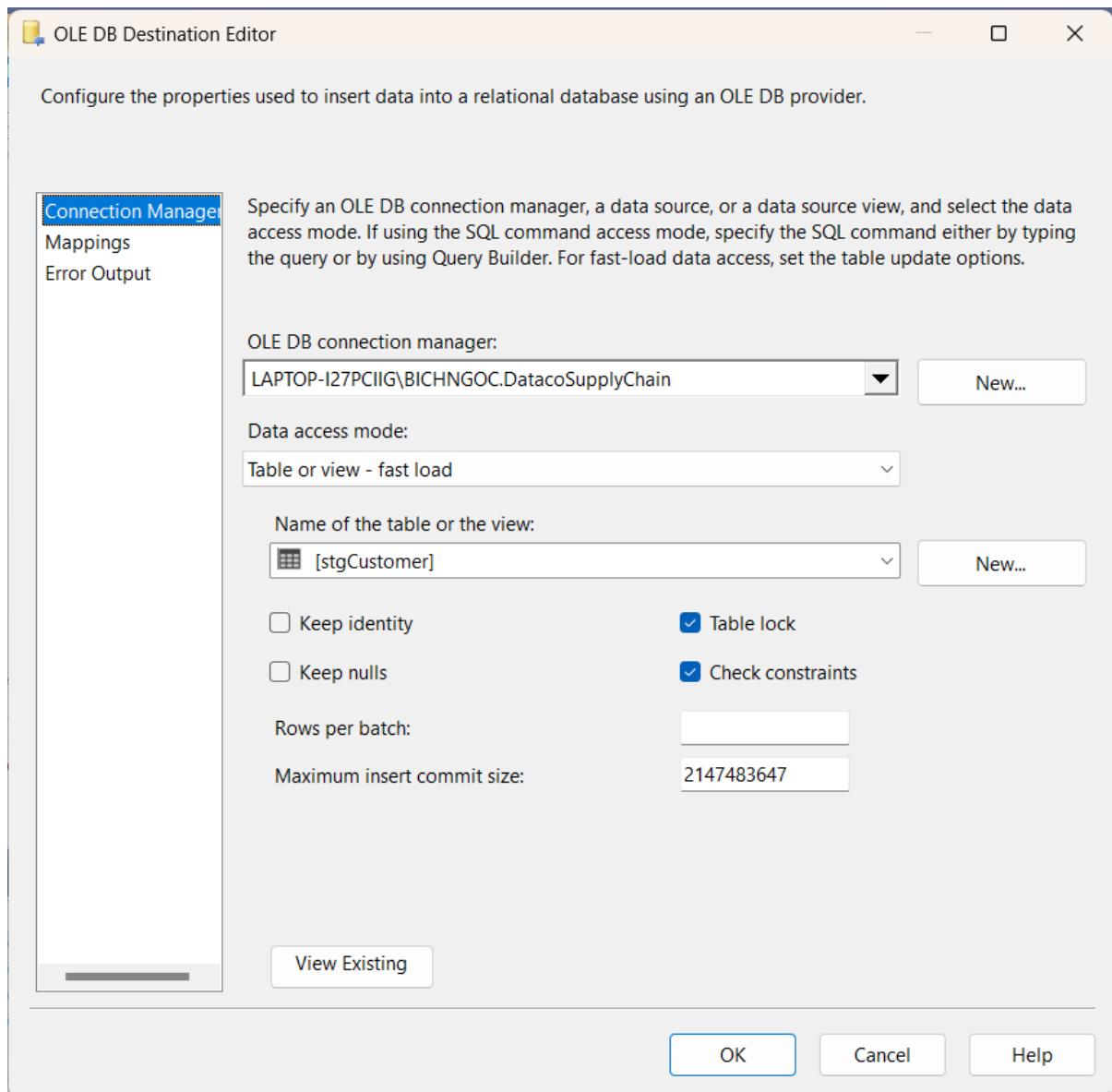
- Double-click **DF - Extract From Source to Stage stgCustomer** task để mở nó trong **Data Flow design surface**:

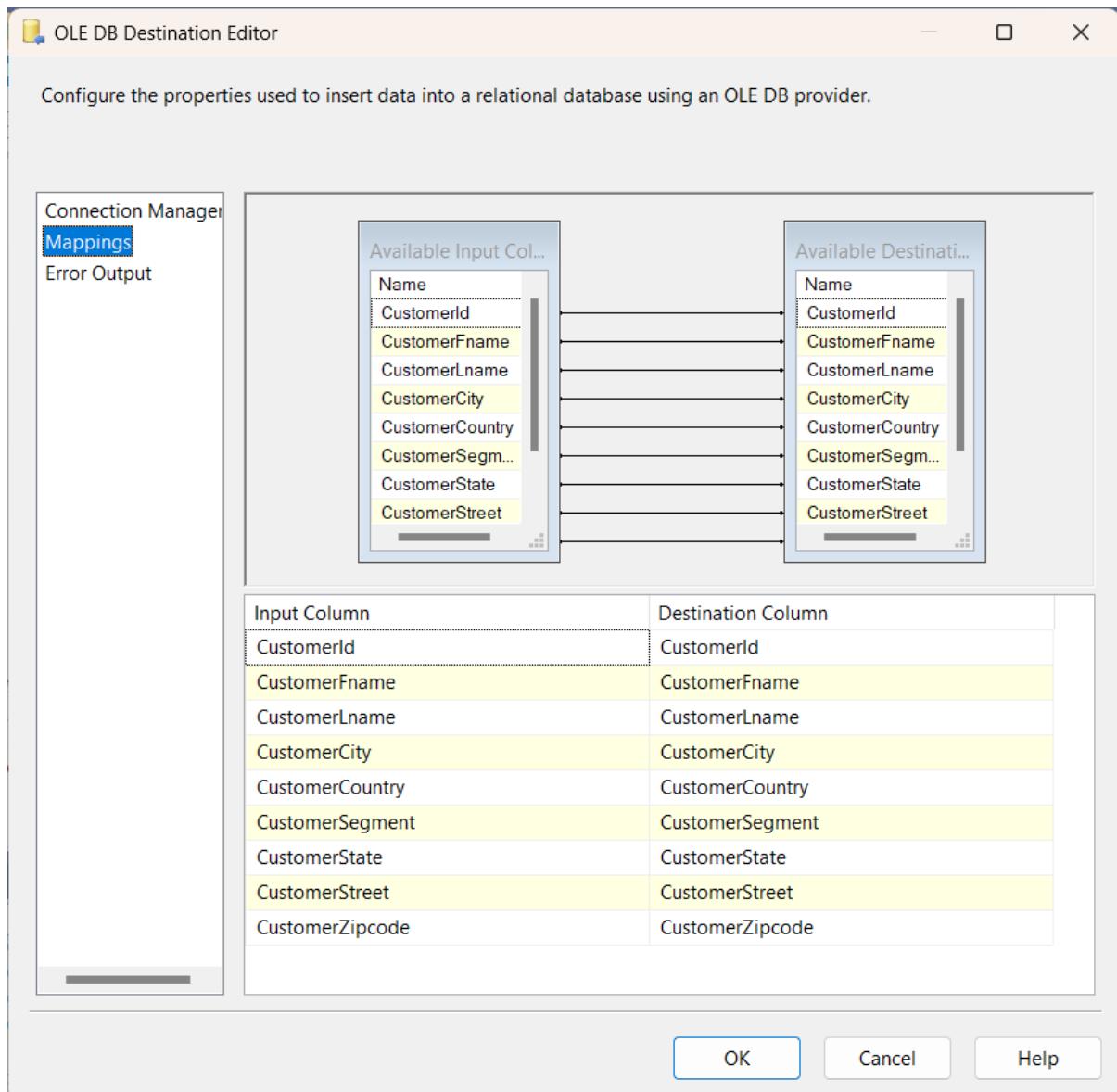
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

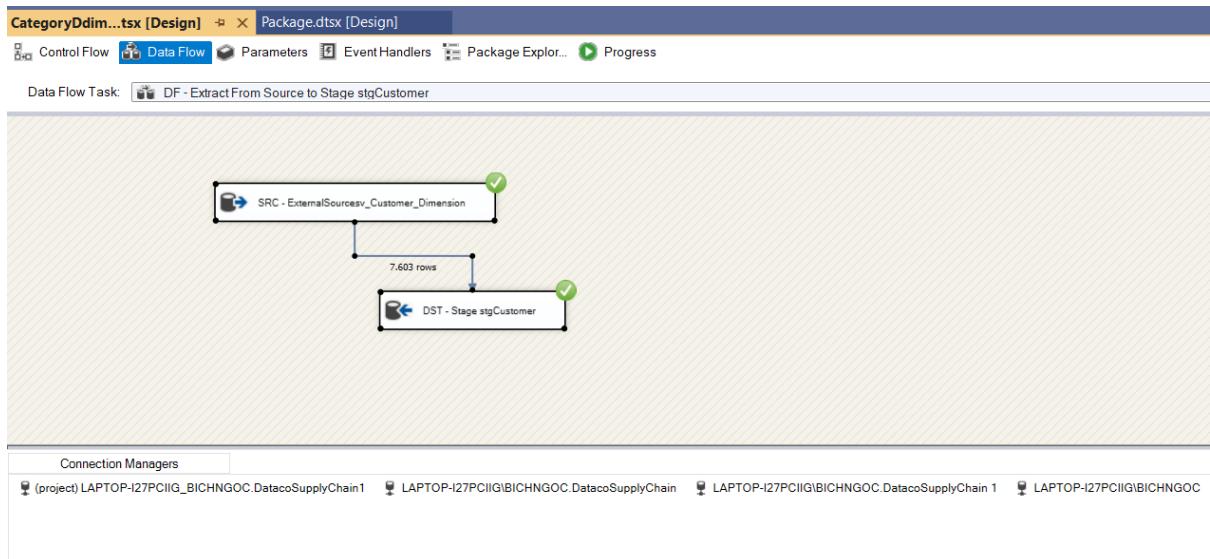


+ Destination Assistant:



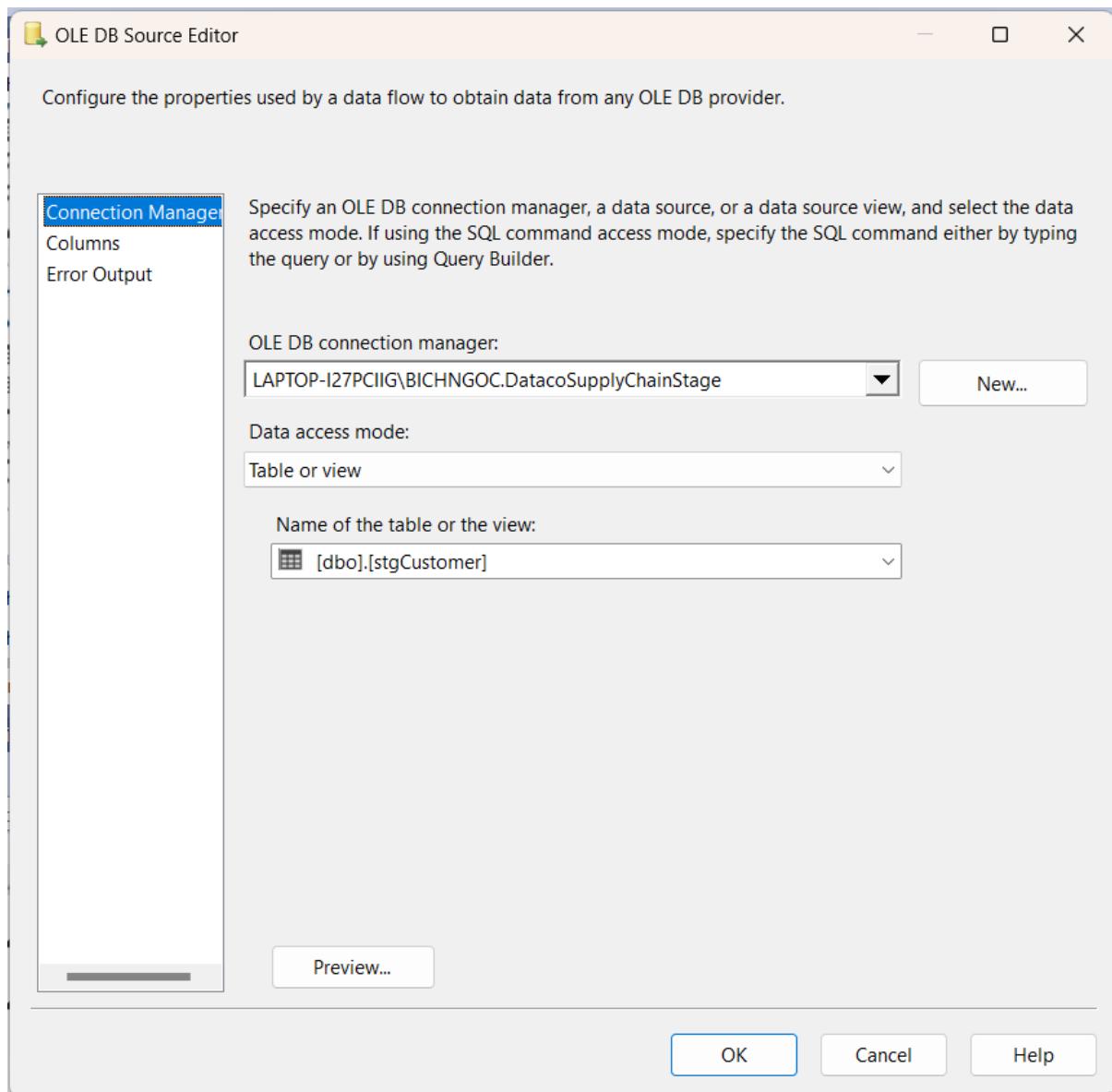


+ Sau khi cấu hình xong ta được:

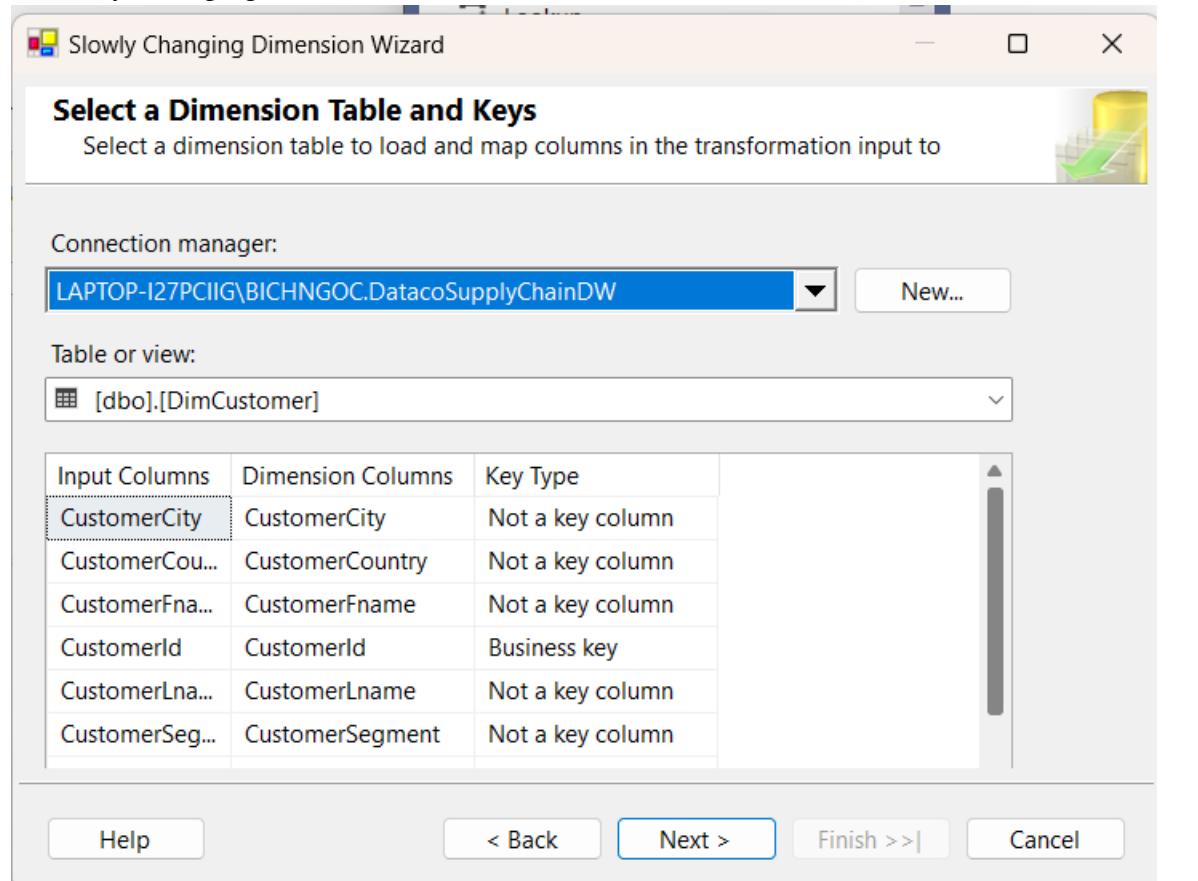


3.3.6.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load from stgCustomer to DimCustomer** task để mở data flow design surface.
- + Source Assistant:



+ Slowly Changing Dimension:

Slowly Changing Dimension Wizard

Select a Dimension Table and Keys

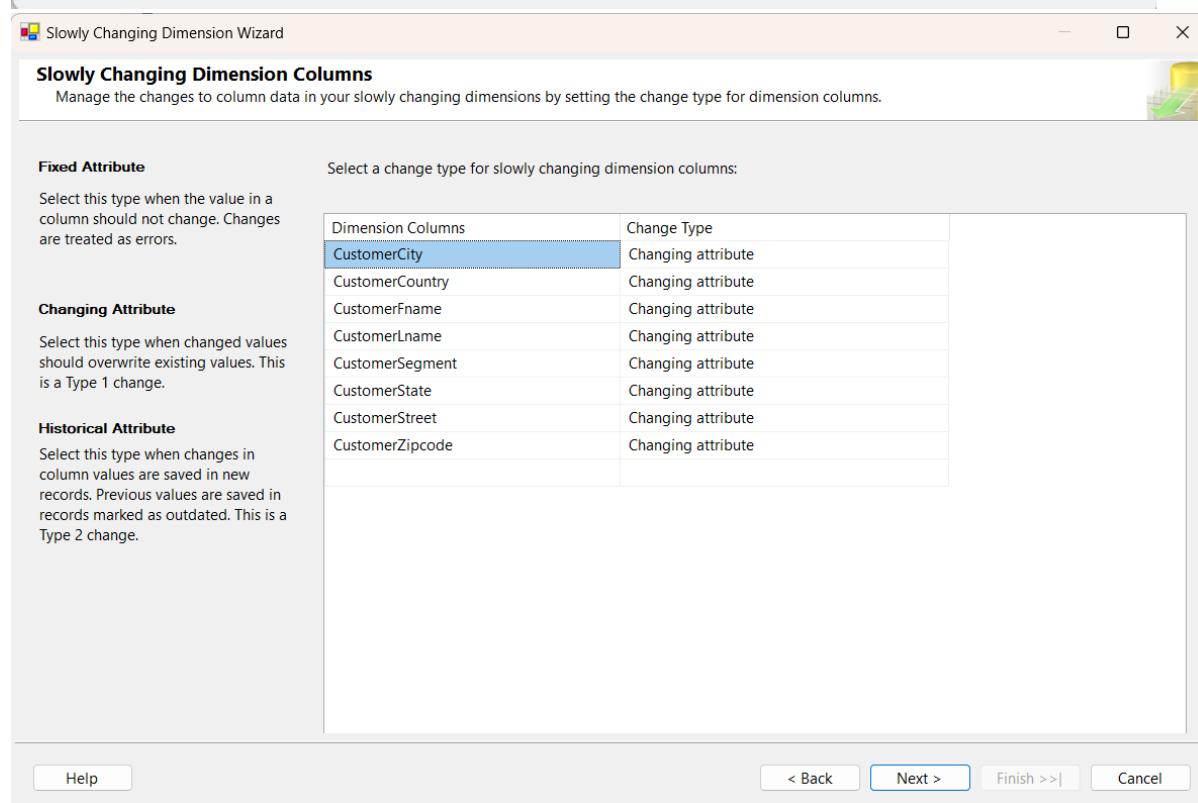
Select a dimension table to load and map columns in the transformation input to

Connection manager: LAPTOP-I27PCIIG\BICHNGOC.DatacoSupplyChainDW ▼ New...

Table or view: [dbo].[DimCustomer]

Input Columns	Dimension Columns	Key Type
CustomerCity	CustomerCity	Not a key column
CustomerCou...	CustomerCountry	Not a key column
CustomerFna...	CustomerFname	Not a key column
CustomerId	CustomerId	Business key
CustomerLna...	CustomerLname	Not a key column
CustomerSeg...	CustomerSegment	Not a key column

Help < Back Next > Finish >> Cancel

Slowly Changing Dimension Wizard

Slowly Changing Dimension Columns

Manage the changes to column data in your slowly changing dimensions by setting the change type for dimension columns.

Fixed Attribute
Select this type when the value in a column should not change. Changes are treated as errors.

Changing Attribute
Select this type when changed values should overwrite existing values. This is a Type 1 change.

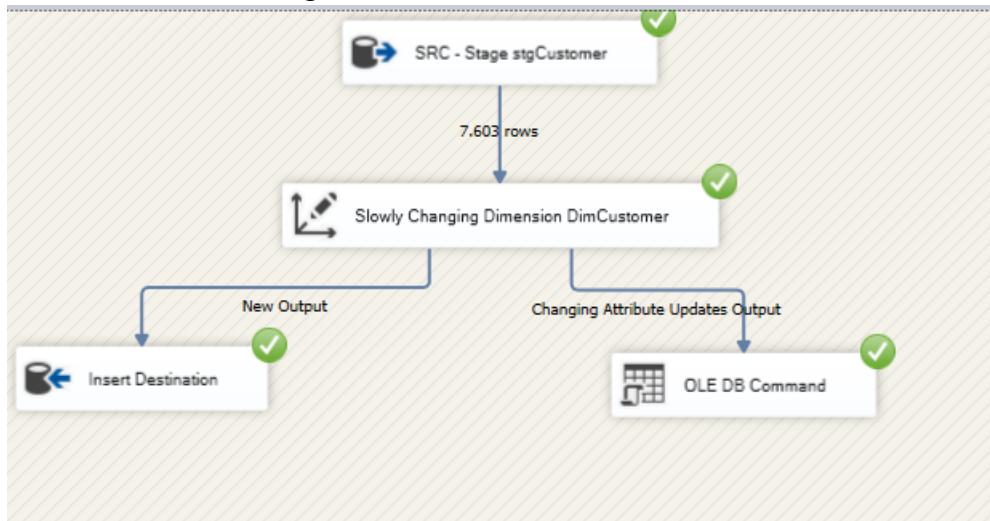
Historical Attribute
Select this type when changes in column values are saved in new records. Previous values are saved in records marked as outdated. This is a Type 2 change.

Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
CustomerCity	Changing attribute
CustomerCountry	Changing attribute
CustomerFname	Changing attribute
CustomerLname	Changing attribute
CustomerSegment	Changing attribute
CustomerState	Changing attribute
CustomerStreet	Changing attribute
CustomerZipcode	Changing attribute

Help < Back Next > Finish >> Cancel

+ Sau khi cấu hình xong ta được:



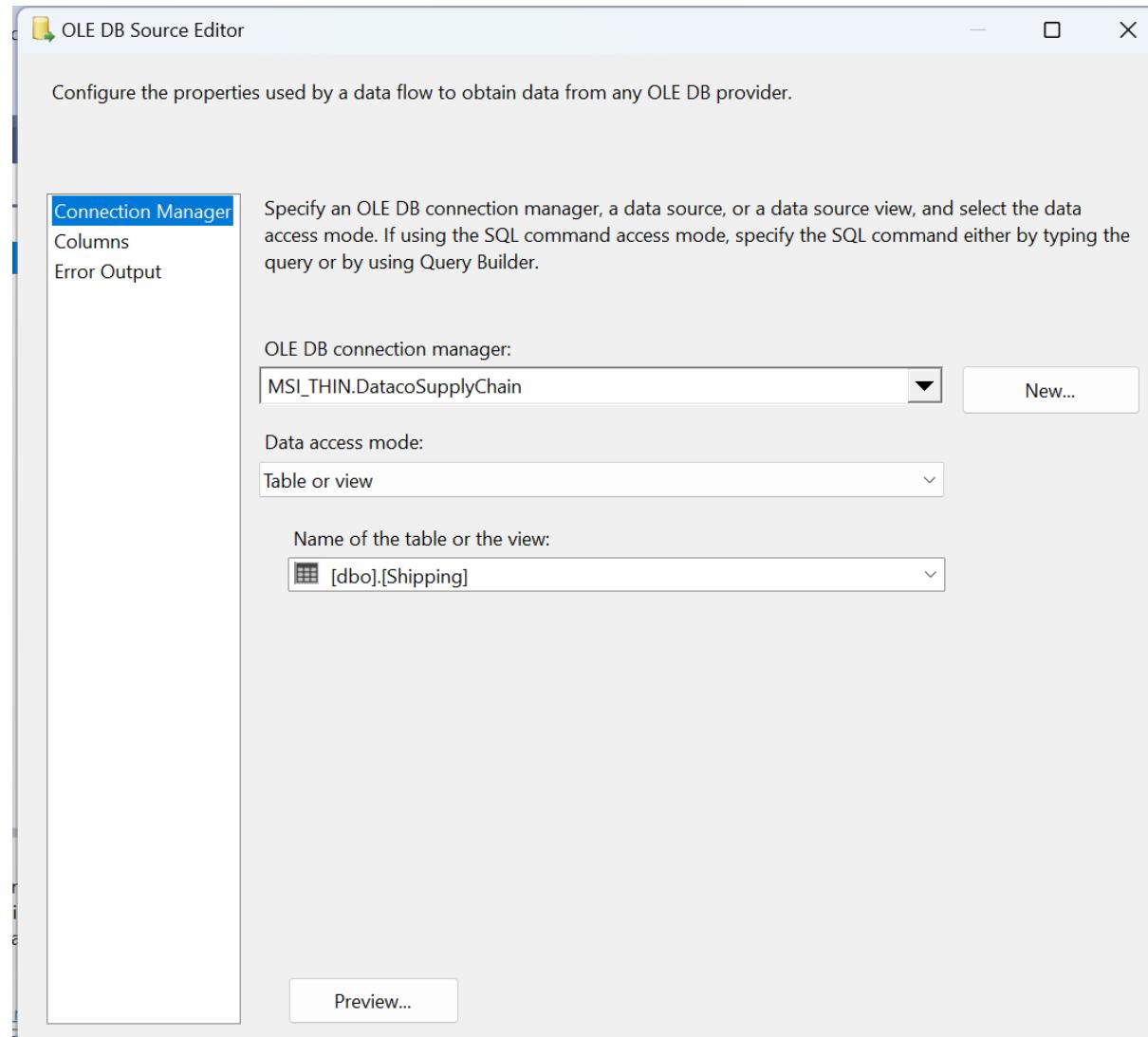
3.3.7.Shipping Dimension

3.3.7.1. Load dữ liệu từ Source vào Stage

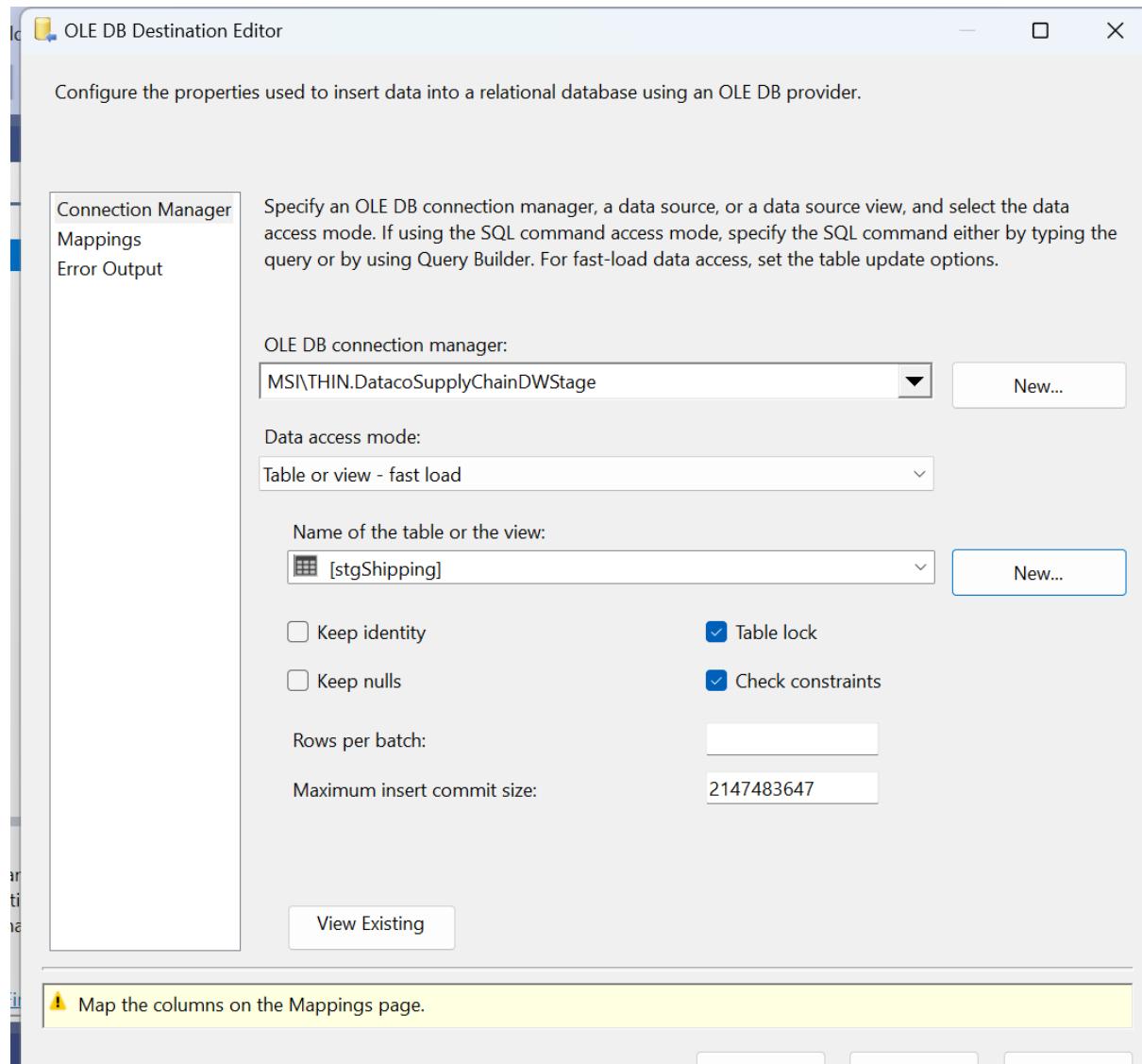
- Double-click DF - Stage Employee task để mở nó trong **Data Flow design surface**:

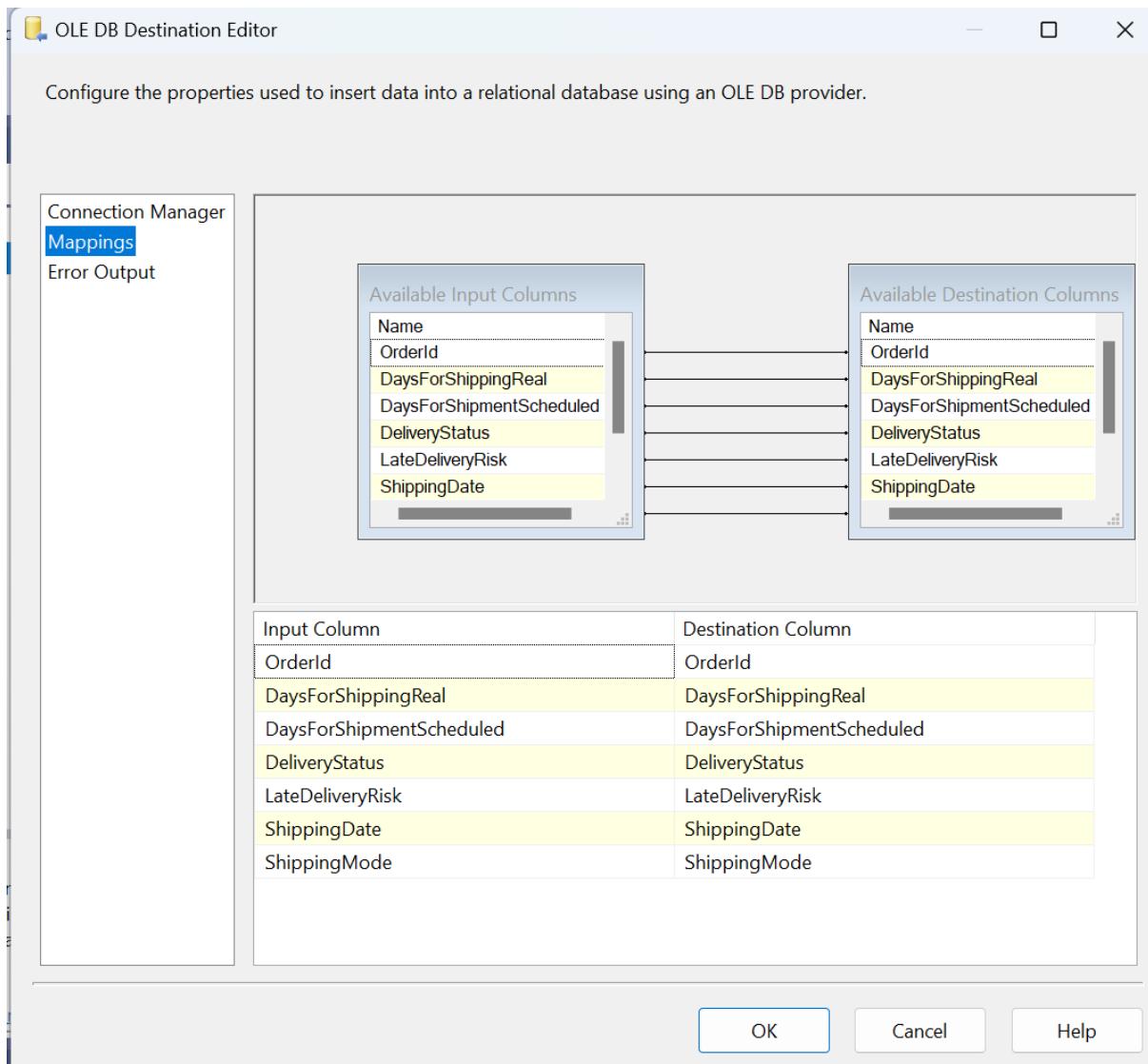
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

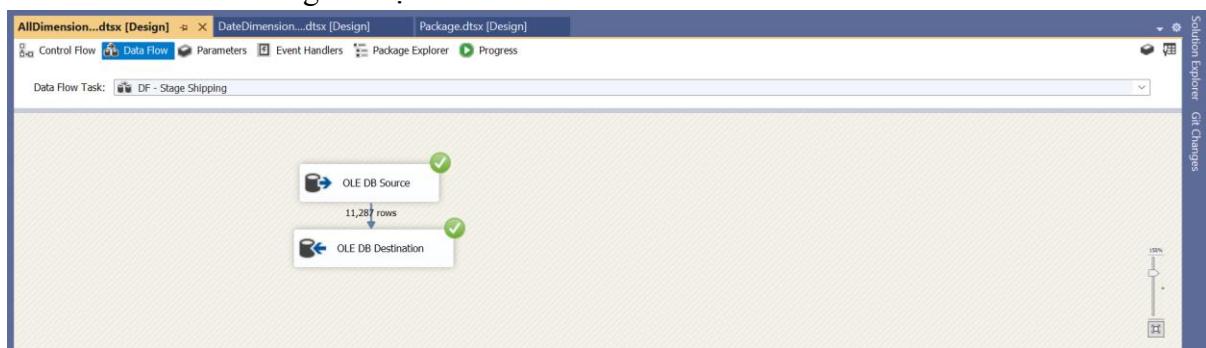


+ Destination Assistant:



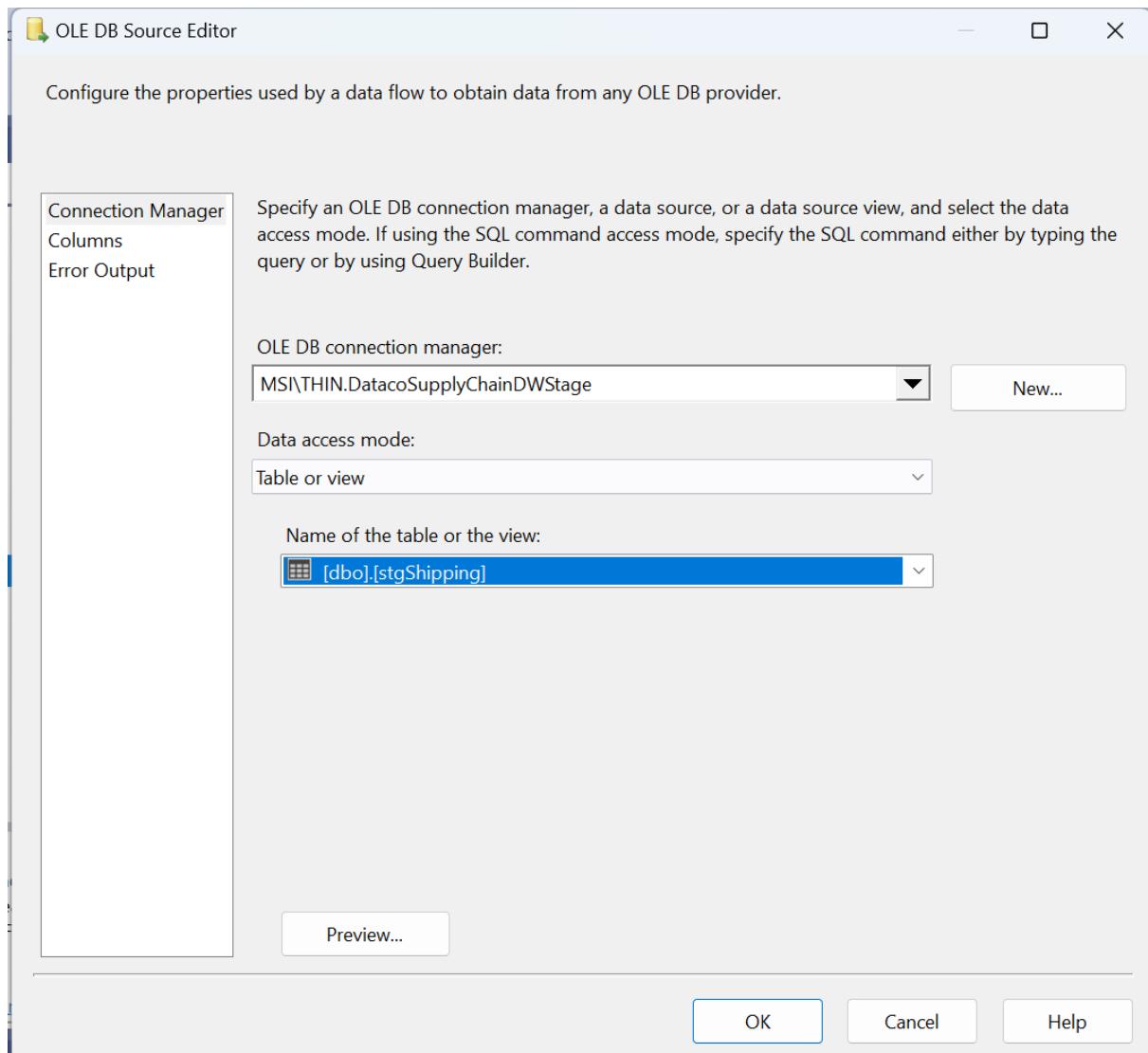


+ Sau khi cấu hình xong ta được:



3.3.7.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load to DimEmployee** task để mở data flow design surface.
- + Source Assistant:



+ Slowly Changing Dimension:

Slowly Changing Dimension Wizard

Select a Dimension Table and Keys

Select a dimension table to load and map columns in the transformation input to

Connection manager:

MSI\THIN.DatacoSupplyChainDW ▼ New...

Table or view:

[dbo].[DimShipping]

Input Columns	Dimension Columns	Key Type
DaysForShipmentSch...	DaysForShipment...	Not a key column
DaysForShippingReal...	DaysForShippingR...	Not a key column
DeliveryStatus	DeliveryStatus	Not a key column
LateDeliveryRisk	LateDeliveryRisk	Not a key column
OrderId	OrderId	Business key
ShippingDate	ShippingDate	Not a key column
ShippingMode	ShippingMode	Not a key column

[Help](#)[< Back](#)[Next >](#)[Finish >>](#)[Cancel](#)

Slowly Changing Dimension Wizard

Slowly Changing Dimension Columns

Manage the changes to column data in your slowly changing dimensions by setting the change type for dimension columns.

Fixed Attribute

Select this type when the value in a column should not change. Changes are treated as errors.

Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
DaysForShipmentSch...	Changing attribute
DaysForShippingReal...	Changing attribute
DeliveryStatus	Changing attribute
LateDeliveryRisk	Changing attribute
ShippingDate	Changing attribute
ShippingMode	Changing attribute

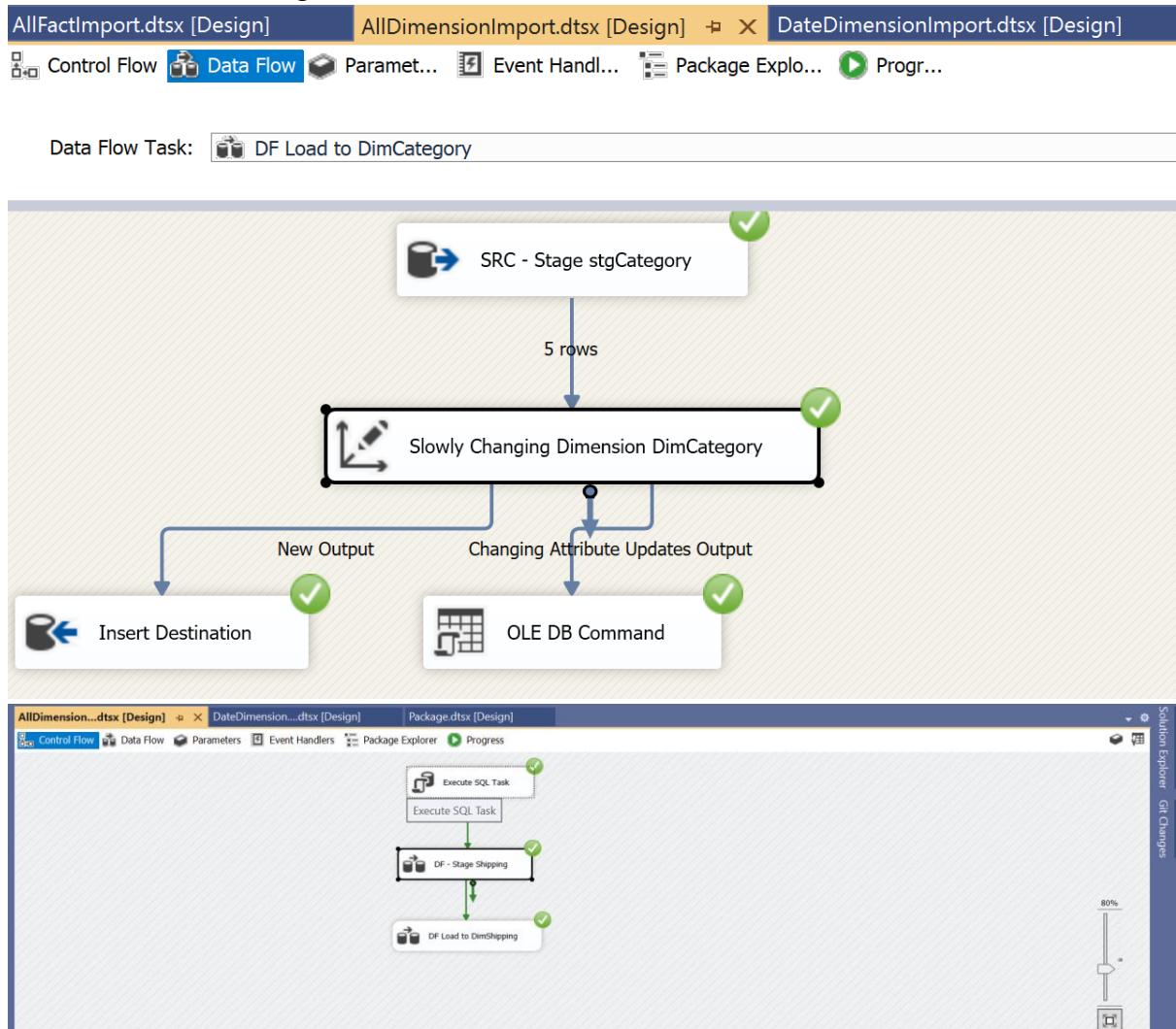
Changing Attribute

Select this type when changed values should overwrite existing values. This is a Type 1 change.

Historical Attribute

Select this type when changes in column values are saved in new records. Previous values are saved in records marked as outdated. This is a Type 2 change.

+ Sau khi cấu hình xong ta được:

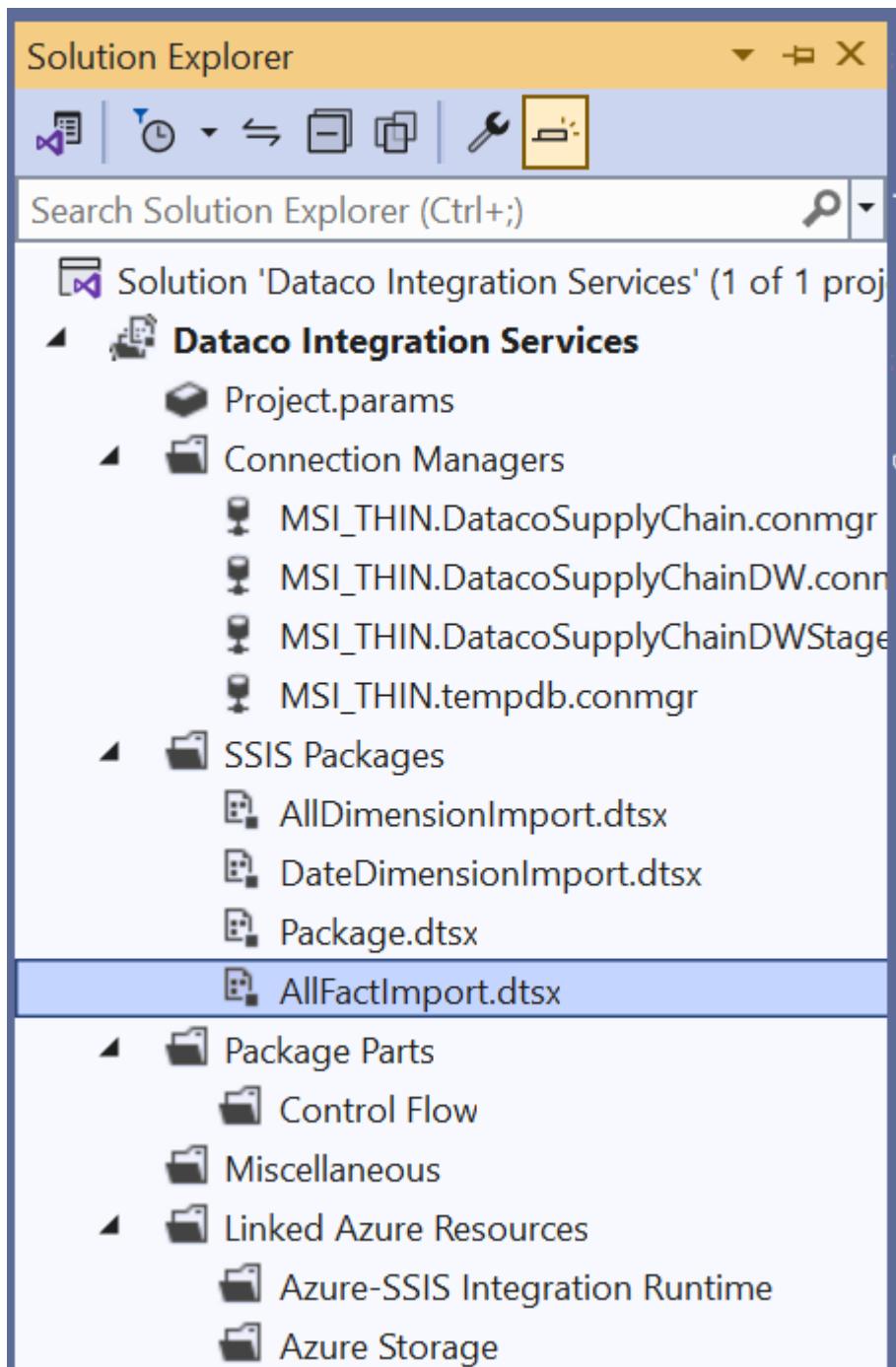


3.4. Import dữ liệu vào bảng các bảng fact

3.4.1. Fact Sales

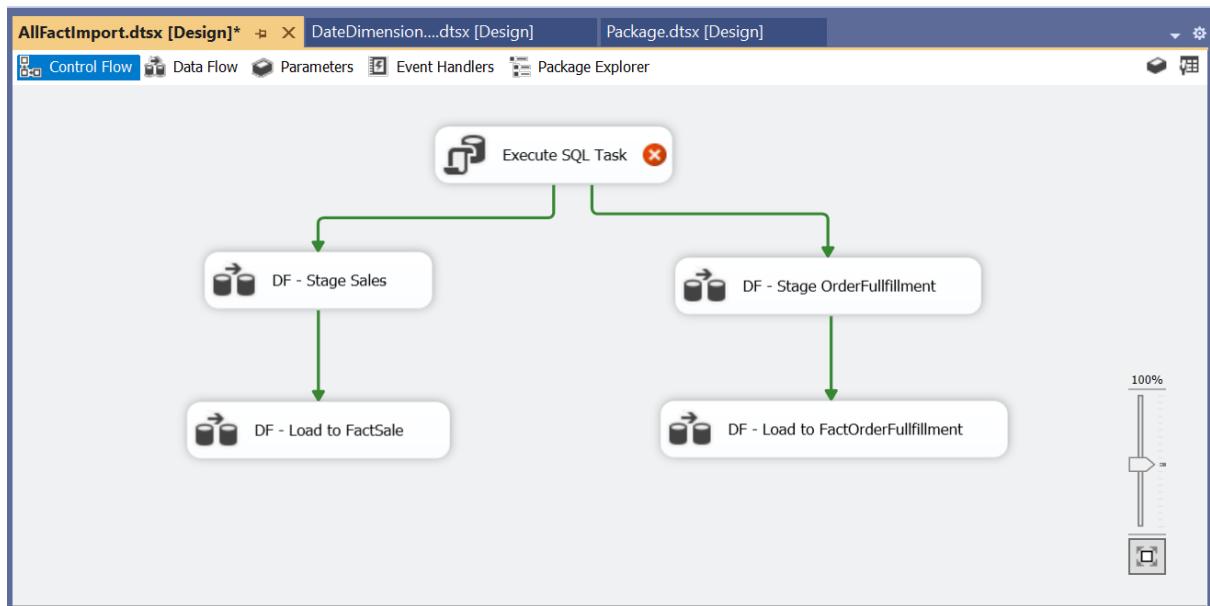
3.4.1.1. Tạo Package

- Tạo package mới và đổi tên như hình bên dưới.

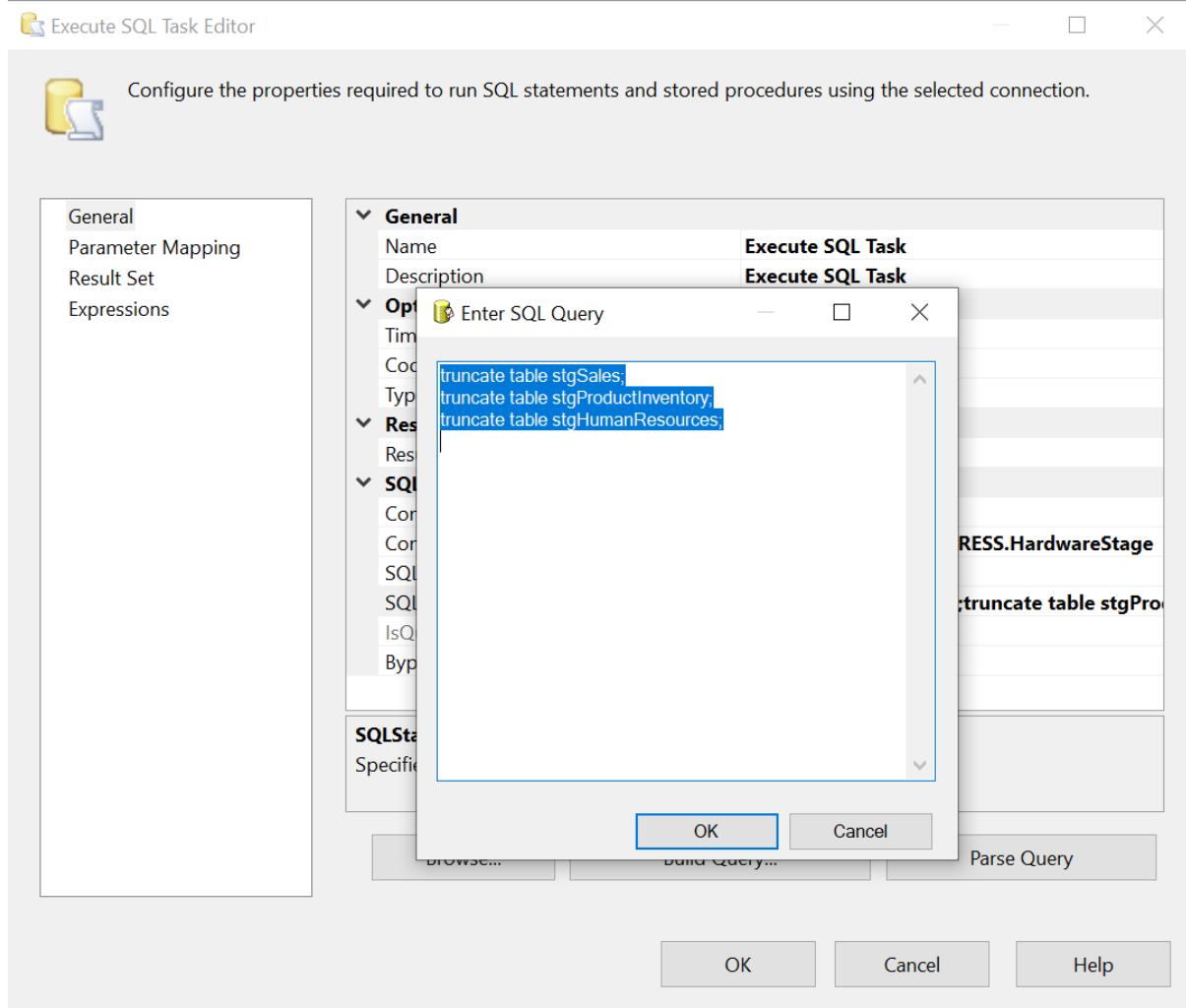


3.4.1.2. Thiết lập Control Flow

- Tạo và đặt tên tương ứng cho các task, tạo các connect bằng cách kéo mũi tên vào các task phù hợp.

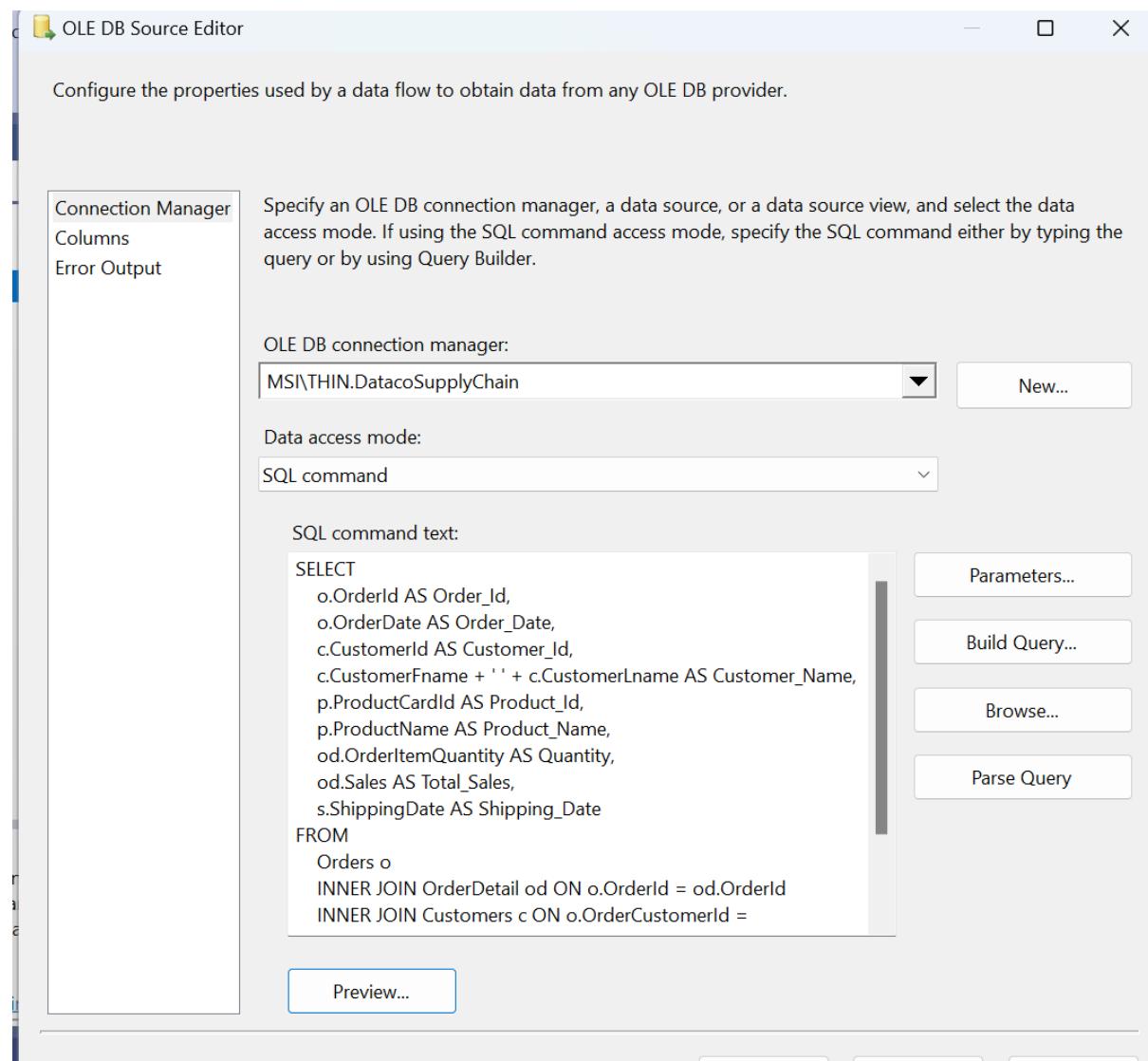


- Thiết lập truncate table SQL task

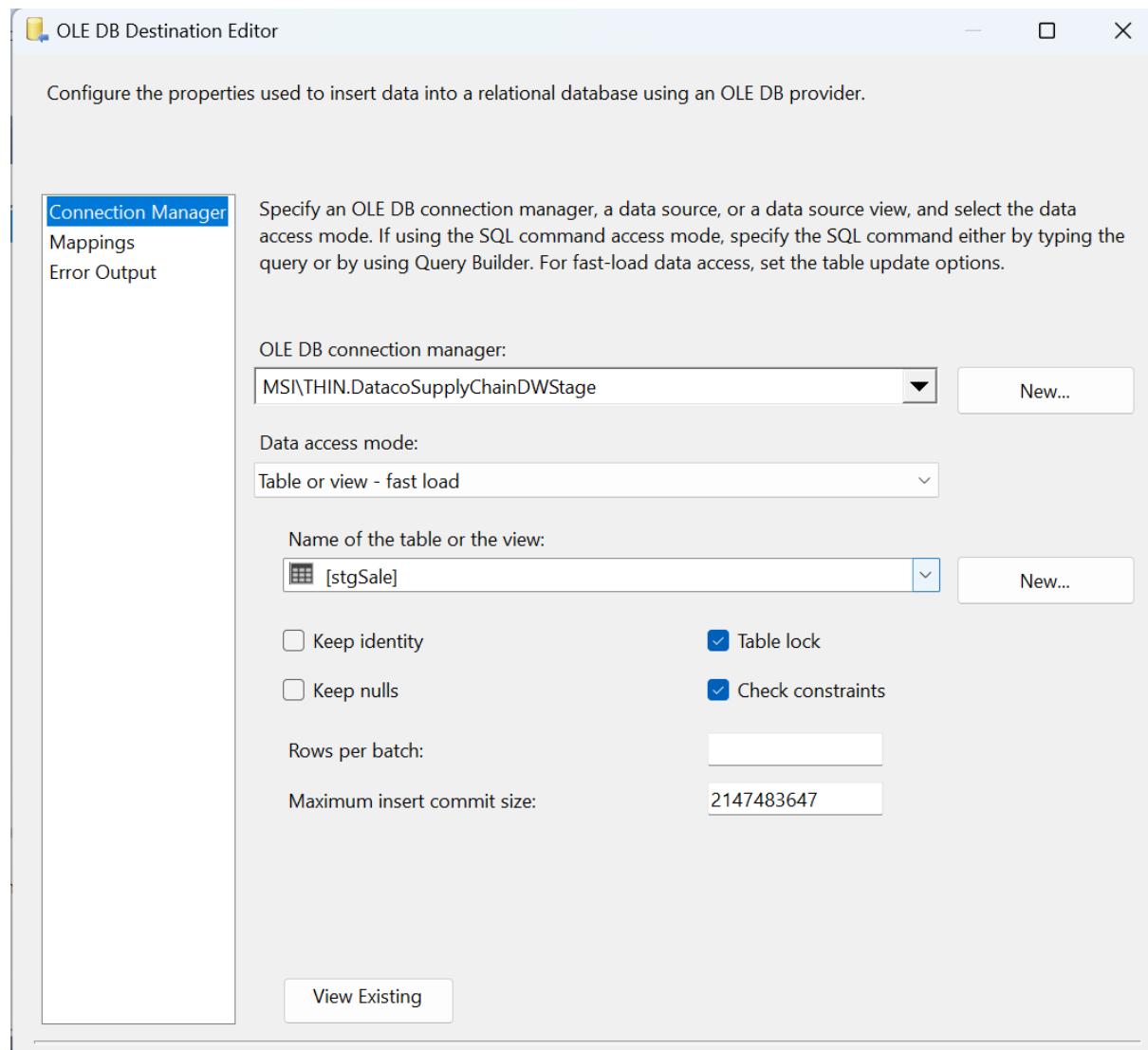


3.4.1.3. Load dữ liệu từ Source vào Stage

- Double-click **DF - Stage Sales** task để mở nó trong **Data Flow design surface**:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
 - + **Source Assistant:**



+ Destination Assist



OLE DB Destination Editor

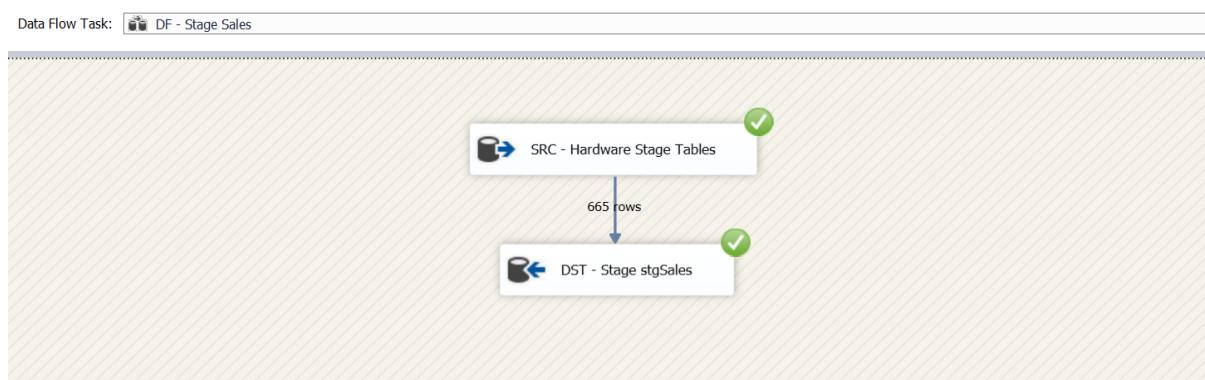
Configure the properties used to insert data into a relational database using an OLE DB provider.

Mappings

Available Input Column	Available Destination Column
Name	Name
Order_Id	Order_Id
Order_Date	Order_Date
Customer_Id	Customer_Id
Customer_Name	Customer_Name
Product_Id	Product_Id
Product_Name	Product_Name
Quantity	Quantity
Total_Sales	Total_Sales

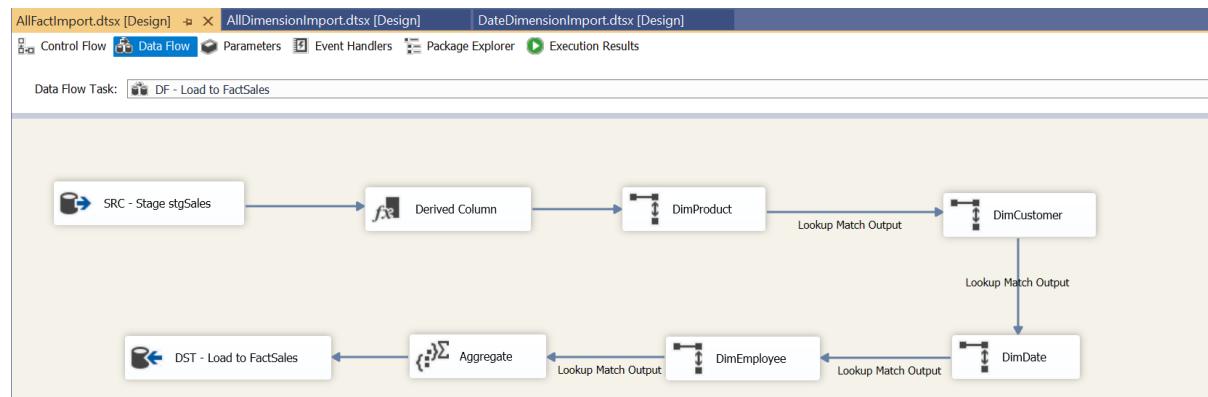
Input Column	Destination Column
Order_Id	Order_Id
Order_Date	Order_Date
Customer_Id	Customer_Id
Customer_Name	Customer_Name
Product_Id	Product_Id
Product_Name	Product_Name
Quantity	Quantity
Total_Sales	Total_Sales
Shipping_Date	Shipping_Date

+ Sau khi cấu hình xong ta được:



3.4.1.4. Load dữ liệu từ Stage vào Fact

- Double-click vào **DF - Load to FactSales** task để mở data flow design Surface và thiết lập các cấu hình sau



- Cấu hình **Derived Column** thêm thuộc tính **SalesAmount**

Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Derived Column Name	Derived Column	Expression	Data Type	Length
SalesAmount	<add as new column>	unit_price * quantity	four-byte signed integer	10

The Derived Column Transformation Editor shows the configuration for adding a new column "SalesAmount". The expression is set to "unit_price * quantity". The data type is "four-byte signed integer" with a length of 10.

- **Lookup qua bảng DimProduct**

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection (highlighted)
Columns
Advanced
Error Output

Specify a data source to use. You can select a table in a data source view, a table in a database connection, or the results of an SQL query.

OLE DB connection manager:
THANHNGAN\SQLEXPRESS.HardwareDW

Use a table or a view:
[dbo].[DimProduct]

- Thiết lập mapping và lấy ra ProductKey

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Available Input Columns:

Name
product_id
order_id
customer_id
quantity
unit_price
order_date
salesman...

Available Lookup Columns:

Name
<input checked="" type="checkbox"/> ProductKey
<input type="checkbox"/> ProductID
<input type="checkbox"/> ProductName
<input type="checkbox"/> CategoryId
<input type="checkbox"/> Description
<input type="checkbox"/> StandardCo...
<input type="checkbox"/> ListPrice
<input type="checkbox"/> RowIsCurrent
<input type="checkbox"/> RowStartDate
<input type="checkbox"/> RowEndDate

Lookup Column: ProductKey
Lookup Operation: <add as new column>
Output Alias: ProductKey

- Lookup qua bảng DimCustomer

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Available Input Columns:

Name
product_id
order_id
customer_id
quantity
unit_price
order_date
salesman...
SalesAmo...

Available Lookup Columns:

Name
<input checked="" type="checkbox"/> CustomerKey
<input type="checkbox"/> CustomerID
<input type="checkbox"/> CustomerName
<input type="checkbox"/> Address
<input type="checkbox"/> CreditLimit
<input type="checkbox"/> Website
<input type="checkbox"/> RowIsCurrent
<input type="checkbox"/> RowStartDate
<input type="checkbox"/> RowEndDate

Lookup Column: CustomerKey
Lookup Operation: <add as new column>
Output Alias: CustomerKey

- Lookup qua bảng DimDate

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Available Input Columns

Name
product_id
order_id
customer_id
quantity
unit_price
order_date
salesman...
SalesAmo...
ProductKey

Available Lookup Columns

Name
<input checked="" type="checkbox"/> DateKey
FullDate
DayNumberOfWeek
DayNameOfWeek
DayNumberOfMo...
DayNumberOfYear
WeekNumberOfY...
MonthName
MonthNumberOfY...
CalendarQuarter

Lookup Column Lookup Operation Output Alias

DateKey	<add as new column>	DateKey
---------	---------------------	---------

- Lookup qua bảng DimEmployee

- Thiết lập query để lấy những đơn hàng không có người bán hàng

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Specify a data source to use. You can select a table in a data source view, a table in a database connection, or the results of an SQL query.

OLE DB connection manager:

THANHNGAN\SQLEXPRESS.HardwareDW

Use a table or a view:

Use results of an SQL query:

```
SELECT EmployeeKey, EmployeeID FROM DimEmployee
UNION SELECT NULL, NULL
```

- Mapping hai bảng và lấy ra EmployeeKey

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Available Input Columns:

- Name
- product_id
- order_id
- customer_id
- quantity
- unit_price
- order_date
- salesman...
- SalesAmo...
- ProductKey
- Customer...

Available Lookup Columns:

- Name
- EmployeeKey
- EmployeeID

Lookup Column	Lookup Operation	Output Alias
EmployeeKey	<add as new column>	EmployeeKey

- Cấu hình Aggregate chọn các thuộc tính cần thiết và các measures

Aggregate Transformation Editor

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

Advanced

Available Input Columns:

- Name
- (*)
- product_id
- order_id
- customer_id
- quantity
- unit_price
- order_date

Input Column	Output Alias	Operation	Comparison Flags	Count Distinct Scale	Count Distinct Keys
quantity	quantity	Sum			
SalesAmount	SalesAmount	Sum			
ProductKey	ProductKey	Group by			
CustomerKey	CustomerKey	Group by			
order_id	order_id	Group by			
DateKey	DateKey	Group by			
unit_price	unit_price	Group by			
EmployeeKey	EmployeeKey	Group by			

OK Cancel Help

- Cấu hình Destination chọn bảng FactSales

OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

Connection Manager

Mappings

Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager: THANHNGAN\SQLEXPRESS.HardwareDW

Data access mode: Table or view - fast load

Name of the table or the view: [dbo].[FactSales]

- Mapping các cột

OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

Mappings

Connection Manager

Error Output

Available Input ...

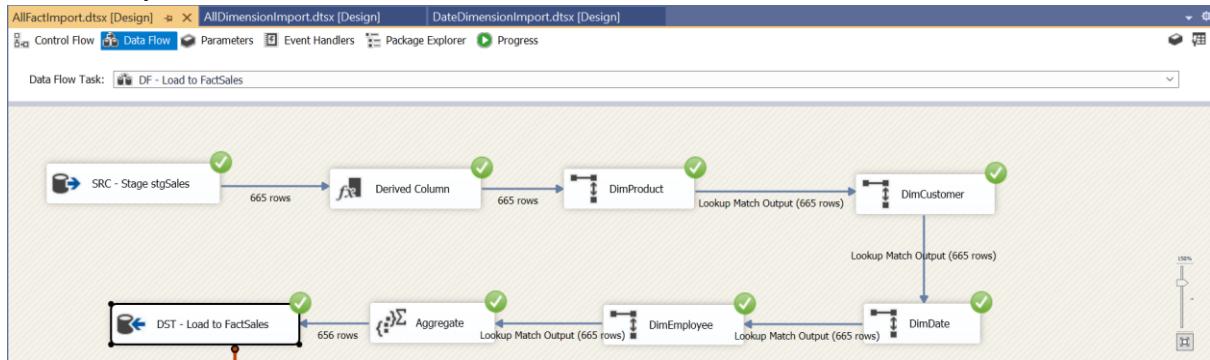
Available Desti...

Input Column

Destination Column

Input Column	Destination Column
DateKey	DateKey
CustomerKey	CustomerKey
EmployeeKey	EmployeeKey
ProductKey	ProductKey
unit_price	UnitPrice
quantity	Quantity
SalesAmount	SalesAmount

- Khởi chạy



3.4.2. Fact OrderFullfillment

3.4.2.1. Load dữ liệu từ Source vào Stage

- Double-click **DF - Stage Product Inventory** task để mở nó trong **Data Flow design surface**:

- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

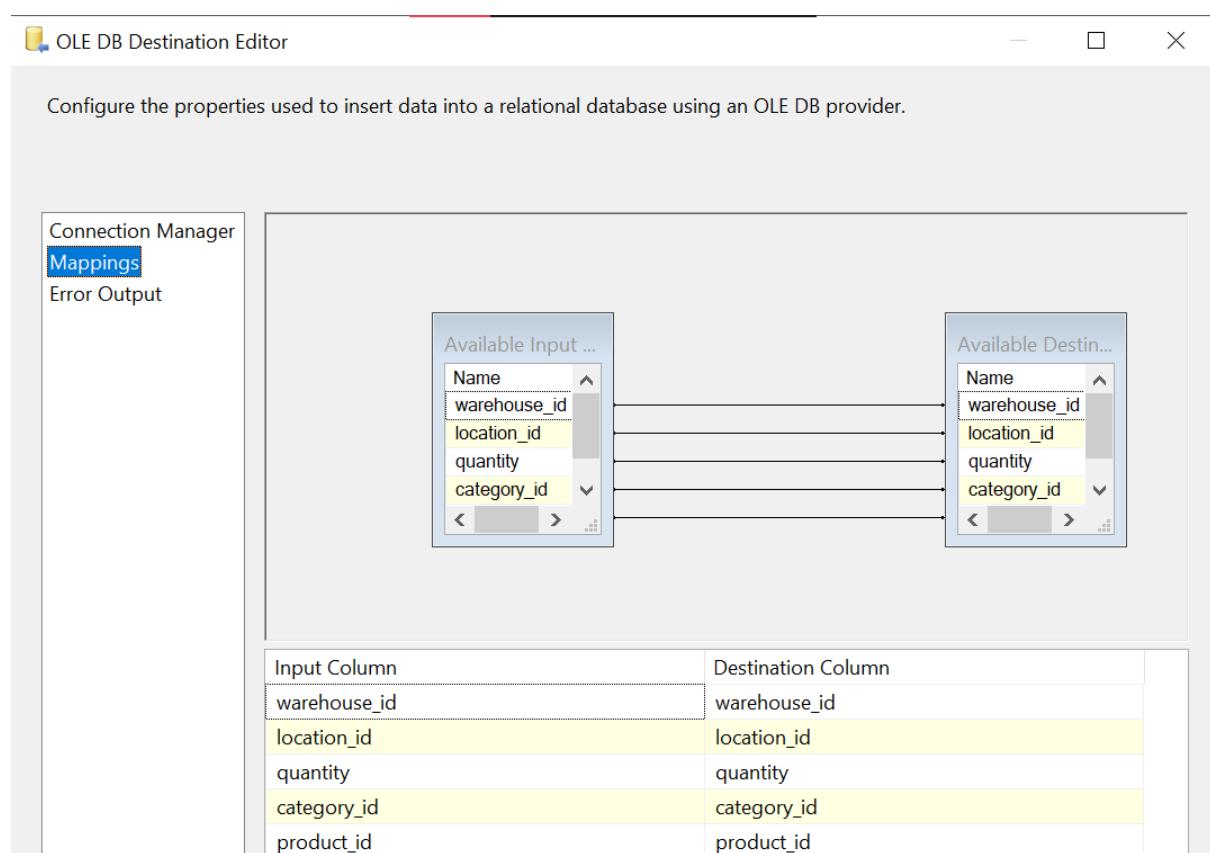
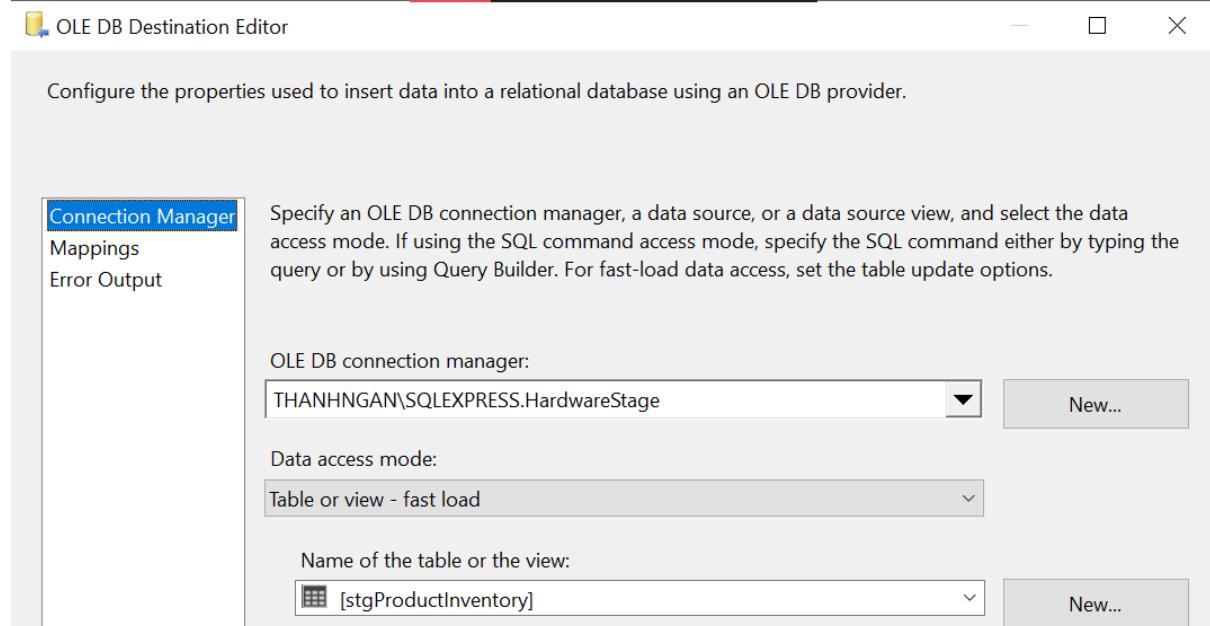
+ Source Assistant:

The screenshot shows the "OLE DB Source Editor" configuration window. On the left, there's a sidebar with tabs: "Connection Manager" (selected), "Columns", and "Error Output". The main area contains the following settings:

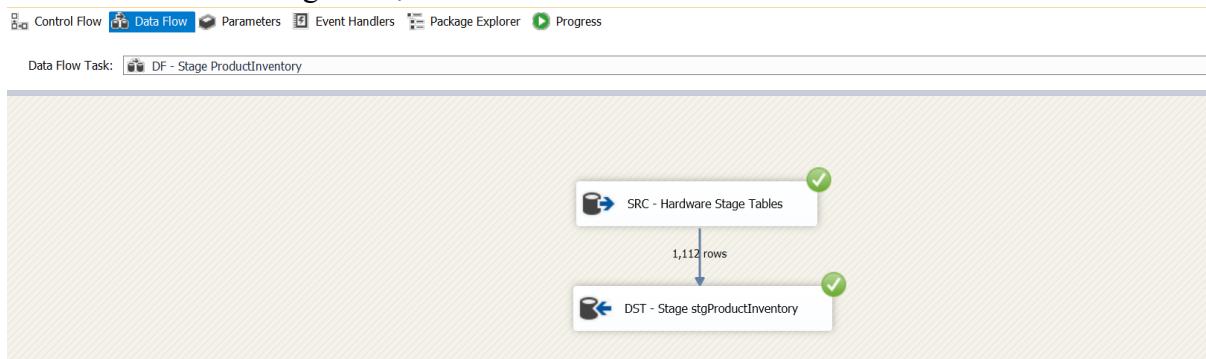
- OLE DB connection manager:** THANHNGAN\SQLEXPRESS.HardwareDatabase
- Data access mode:** SQL command
- SQL command text:**

```
SELECT dbo.warehouses.warehouse_id, dbo.locations.location_id,
dbo.inventories.quantity, dbo.product_categories.category_id,
dbo.products.product_id
FROM dbo.inventories, dbo.warehouses, dbo.locations,
dbo.products, dbo.product_categories
WHERE dbo.inventories.warehouse_id =
dbo.warehouses.warehouse_id AND dbo.warehouses.location_id =
dbo.locations.location_id AND dbo.inventories.product_id =
dbo.products.product_id AND dbo.products.category_id =
dbo.product_categories.category_id
```
- On the right side, there are four buttons:
 - Parameters...
 - Build Query...
 - Browse...
 - Parse Query

+ Destination Assist

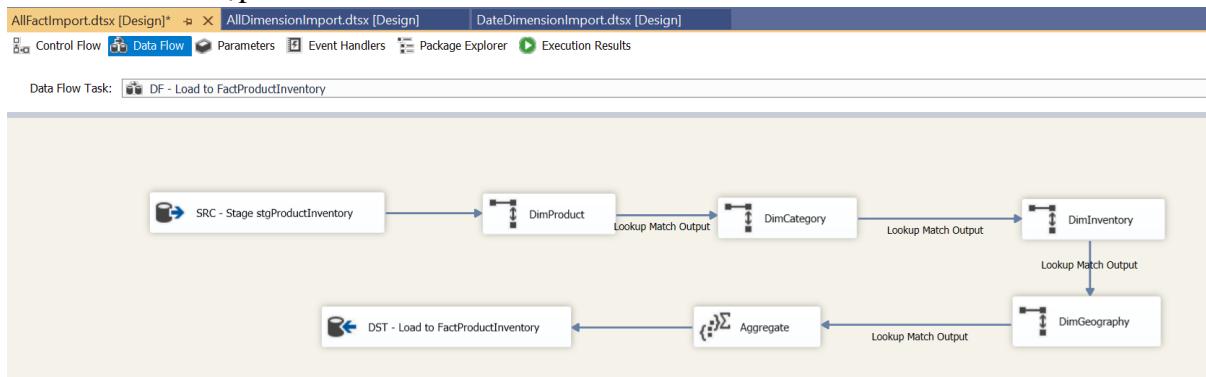


+ Sau khi cấu hình xong ta được:



3.4.1.4. Load dữ liệu từ Stage vào Fact

- Double-click vào **DF - Load to FactOrderFullfillment** task để mở data flow design Surface và thiết lập các cấu hình sau



- Cấu hình Source chọn bảng **stgProductInventory**

The screenshot shows the 'OLE DB Source Editor' configuration for the 'stgProductInventory' source. On the left, there is a navigation pane with tabs: 'Connection Manager' (selected), 'Columns', and 'Error Output'. The main area contains the following settings:

- OLE DB connection manager:** A dropdown menu set to 'THANHNGAN\SQLEXPRESS.HardwareStage'.
- Data access mode:** A dropdown menu set to 'Table or view'.
- Name of the table or the view:** A dropdown menu set to '[dbo].[stgProductInventory]'.

- Lookup qua bảng DimProduct

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Specify a data source to use. You can select a table in a data source view, a table in a database connection, or the results of an SQL query.

OLE DB connection manager:
THANHNGAN\SQLEXPRESS.HardwareDW

Use a table or a view:
[dbo].[DimProduct]

- Thiết lập mapping và lấy ra ProductKey

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Available Input ...

Name	warehouse_id
location_id	quantity
category_id	

Available Lookup Col...

Name	ProductKey
ProductID	ProductName
CategoryId	Description
StandardCo...	ListPrice
ListPrice	RowIsCurrent
RowStartD...	RowEndDate

Lookup Column Lookup Operation Output Alias

ProductKey	<add as new column>	ProductKey
------------	---------------------	------------

- Lookup qua bảng DimCategory

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Specify a data source to use. You can select a table in a data source view, a table in a database connection, or the results of an SQL query.

OLE DB connection manager:
THANHNGAN\SQLEXPRESS.HardwareDW

Use a table or a view:
[dbo].[DimCategory]

- Mapping và lấy CategoryKey

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

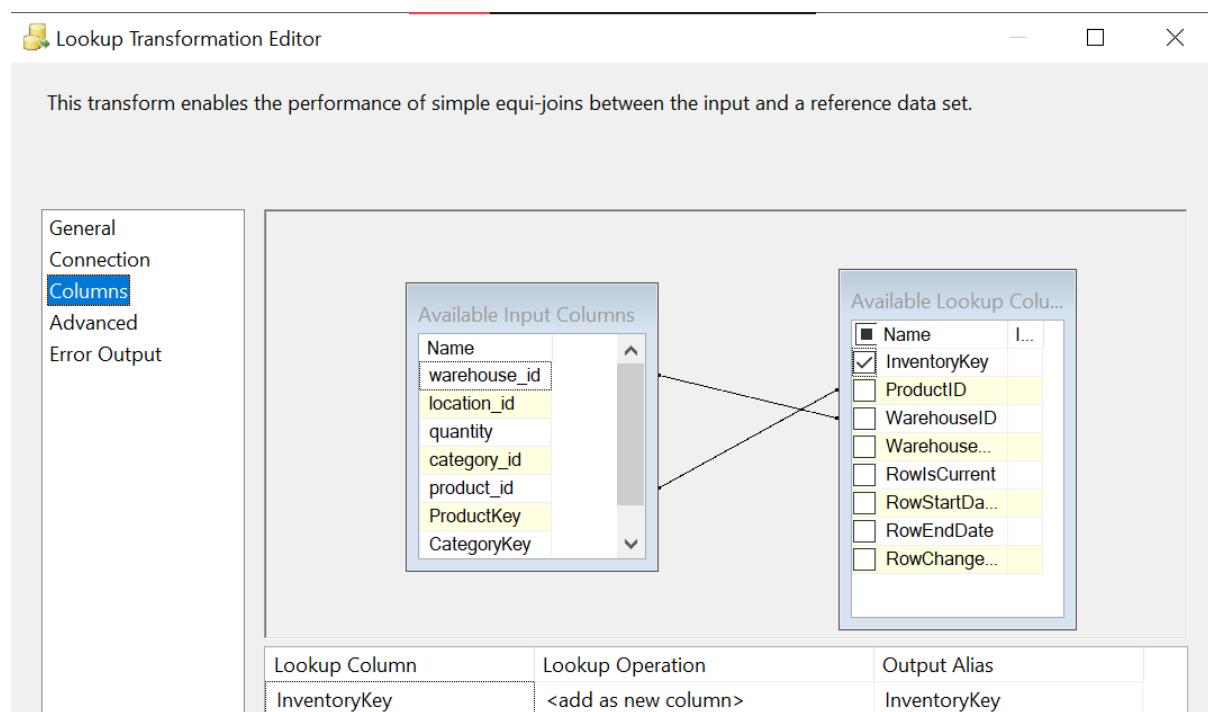
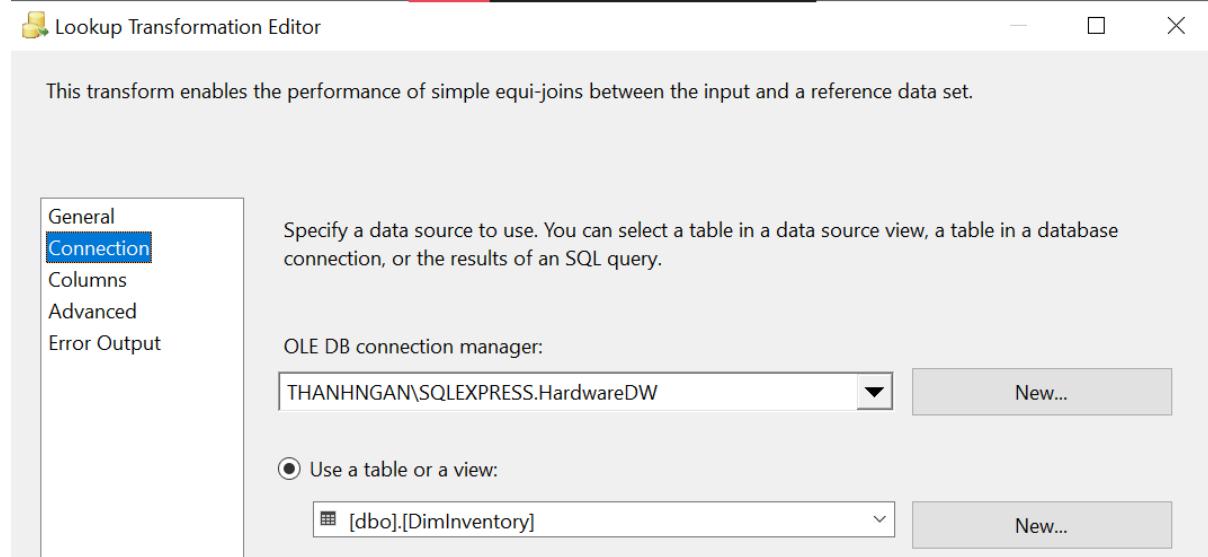
General
Connection
Columns
Advanced
Error Output

Available Input ...
Name
warehouse_id
location_id
quantity
category_id
product_id

Available Lookup Colu...
Name
CategoryKey
CategoryID
CategoryName
RowIsCurrent
RowStartDate
RowEndDate

Lookup Column Lookup Operation Output Alias
CategoryKey <add as new column> CategoryKey

- Lookup qua bảng DimShipping



- Lookup qua bảng DimGeography

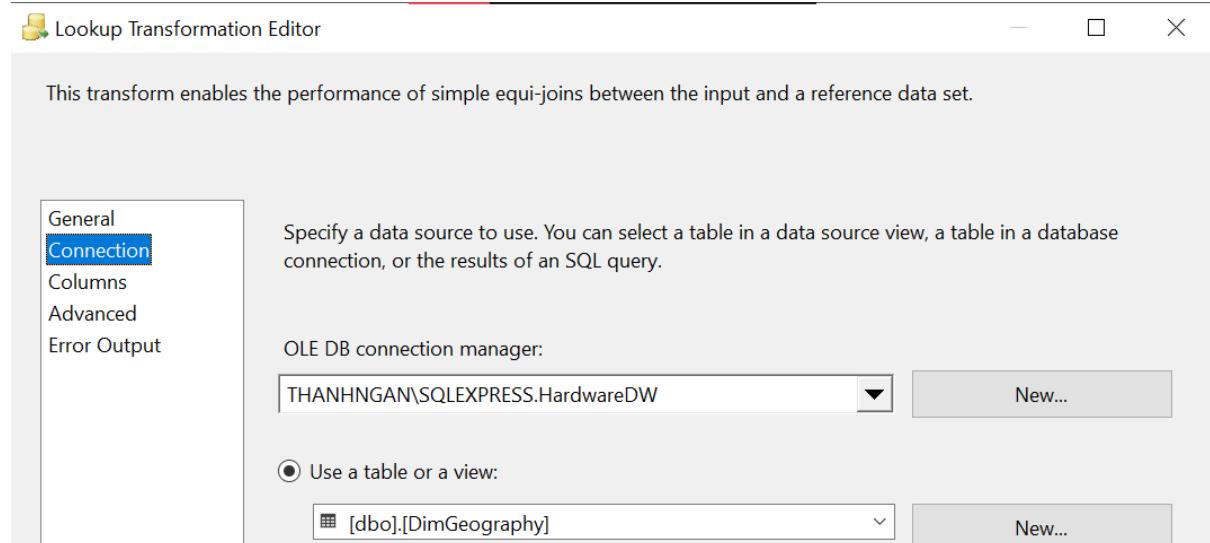
This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Specify a data source to use. You can select a table in a data source view, a table in a database connection, or the results of an SQL query.

OLE DB connection manager:
THANHNGAN\SQLEXPRESS.HardwareDW

(Use a table or a view:
[dbo].[DimGeography]



- Mapping hai bảng và lấy ra GeographyKey

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Available Input ...

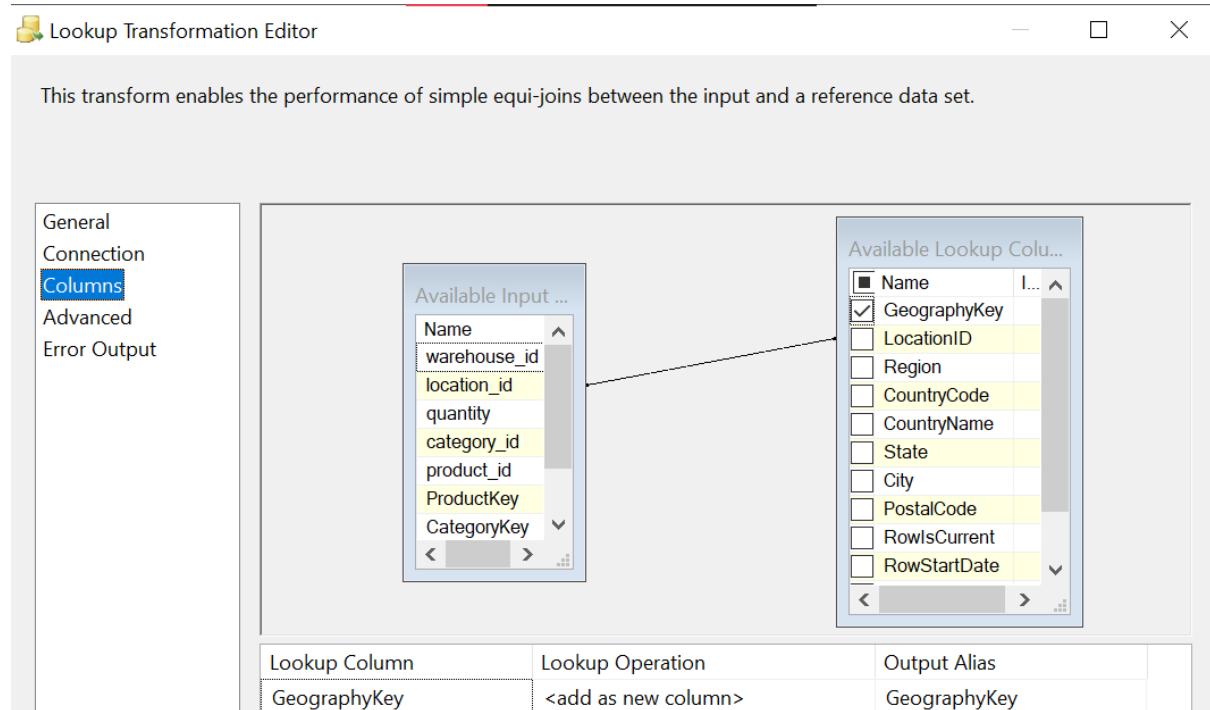
Name	I... ^
warehouse_id	
location_id	
quantity	
category_id	
product_id	
ProductKey	
CategoryKey	▼
< >	

Available Lookup Colu...

<input checked="" type="checkbox"/> Name	I... ^
<input checked="" type="checkbox"/> GeographyKey	
<input type="checkbox"/> LocationID	
<input type="checkbox"/> Region	
<input type="checkbox"/> CountryCode	
<input type="checkbox"/> CountryName	
<input type="checkbox"/> State	
<input type="checkbox"/> City	
<input type="checkbox"/> PostalCode	
<input type="checkbox"/> RowIsCurrent	
<input type="checkbox"/> RowStartDate	▼
< >	

Lookup Column Lookup Operation Output Alias

GeographyKey	<add as new column>	GeographyKey
--------------	---------------------	--------------



- Cấu hình Aggregate chọn các thuộc tính cần thiết và các measures

Σ Aggregate Transformation Editor

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

Advanced

Available Input Columns

Input Column	Output Alias	Operation	Comparison Operator
ProductKey	ProductKey	Group by	
CategoryKey	CategoryKey	Group by	
InventoryKey	InventoryKey	Group by	
GeographyKey	GeographyKey	Group by	
quantity	quantity	Sum	

- Cấu hình Destination chọn bảng FactSales

OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

Connection Manager

Mappings

Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:

THANHNGAN\SQLEXPRESS.HardwareDW

New...

Data access mode:

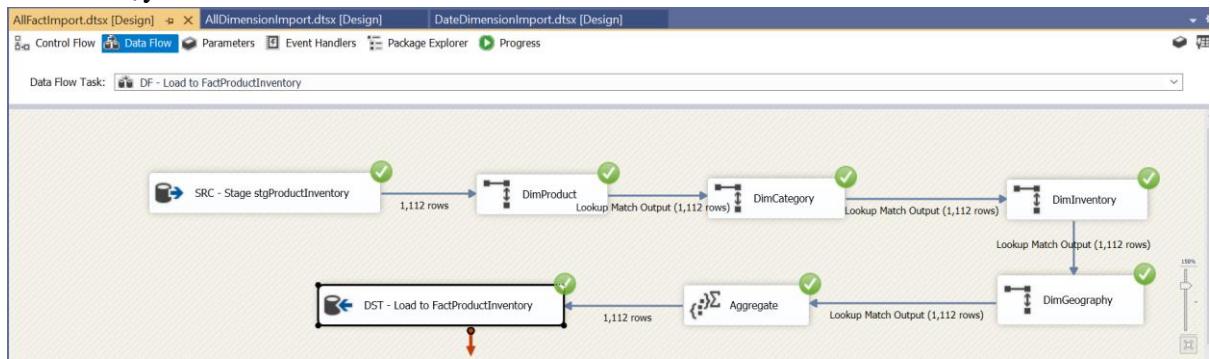
Table or view - fast load

Name of the table or the view:

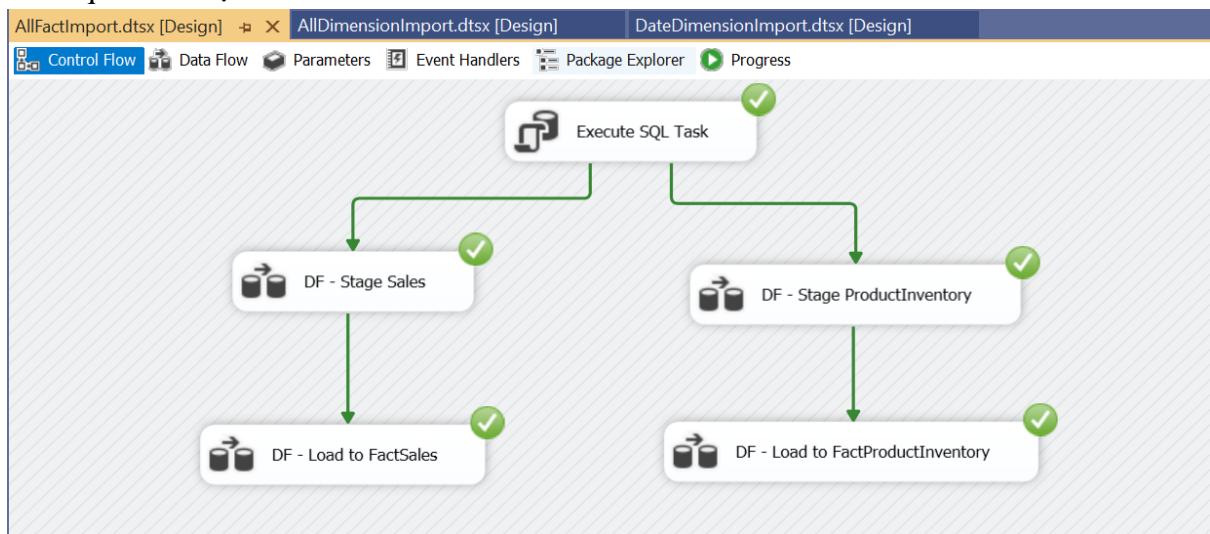
[dbo].[FactProductInventory]

New...

- Khởi chạy



- Kết quả thu được

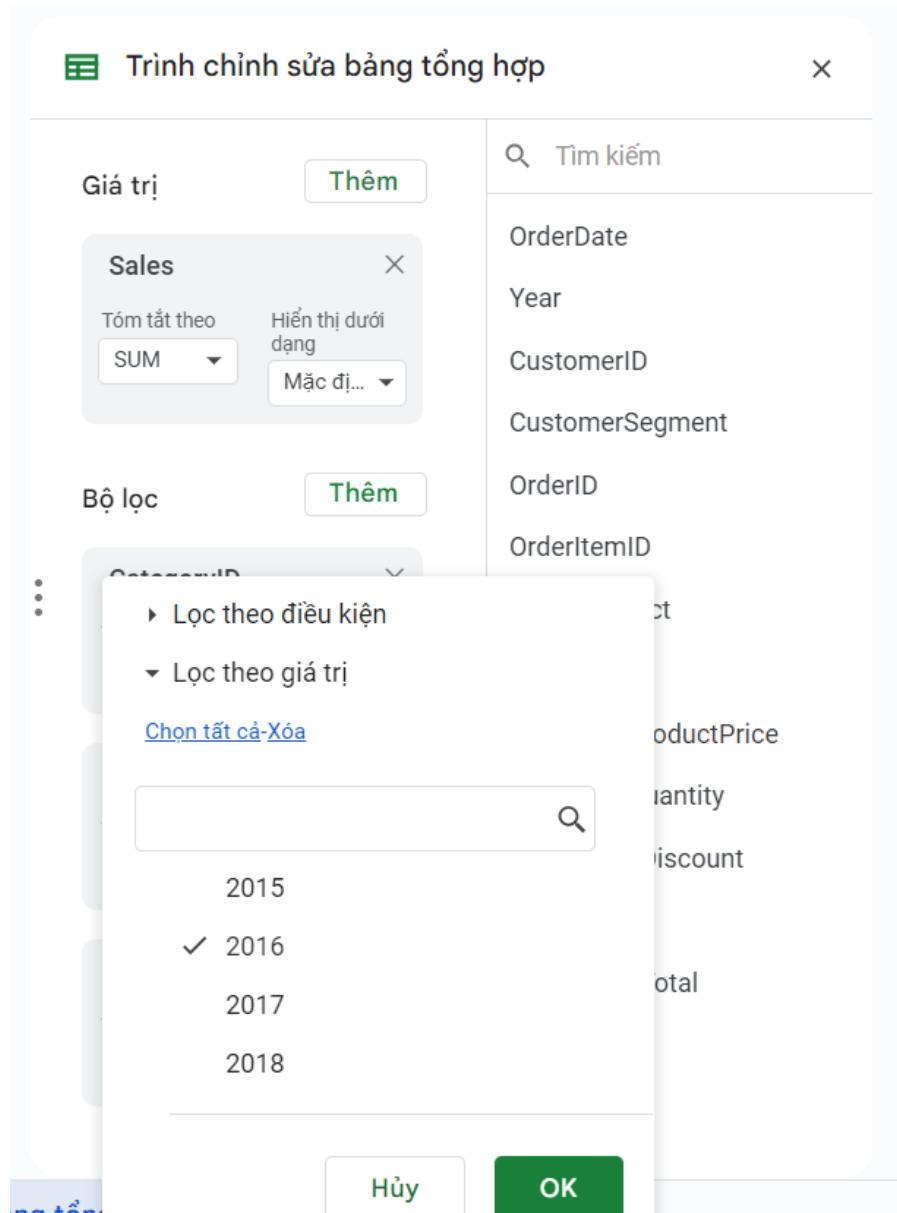


CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU

4.1. Thực hiện phân tích dữ liệu bằng Pivot Table trong Excel

4.4.1. Câu hỏi: Cho biết sản phẩm có doanh thu cao nhất trong mỗi năm

Sử dụng bộ lọc để chọn năm để biết được tổng doanh thu của từng sản phẩm trong từng năm

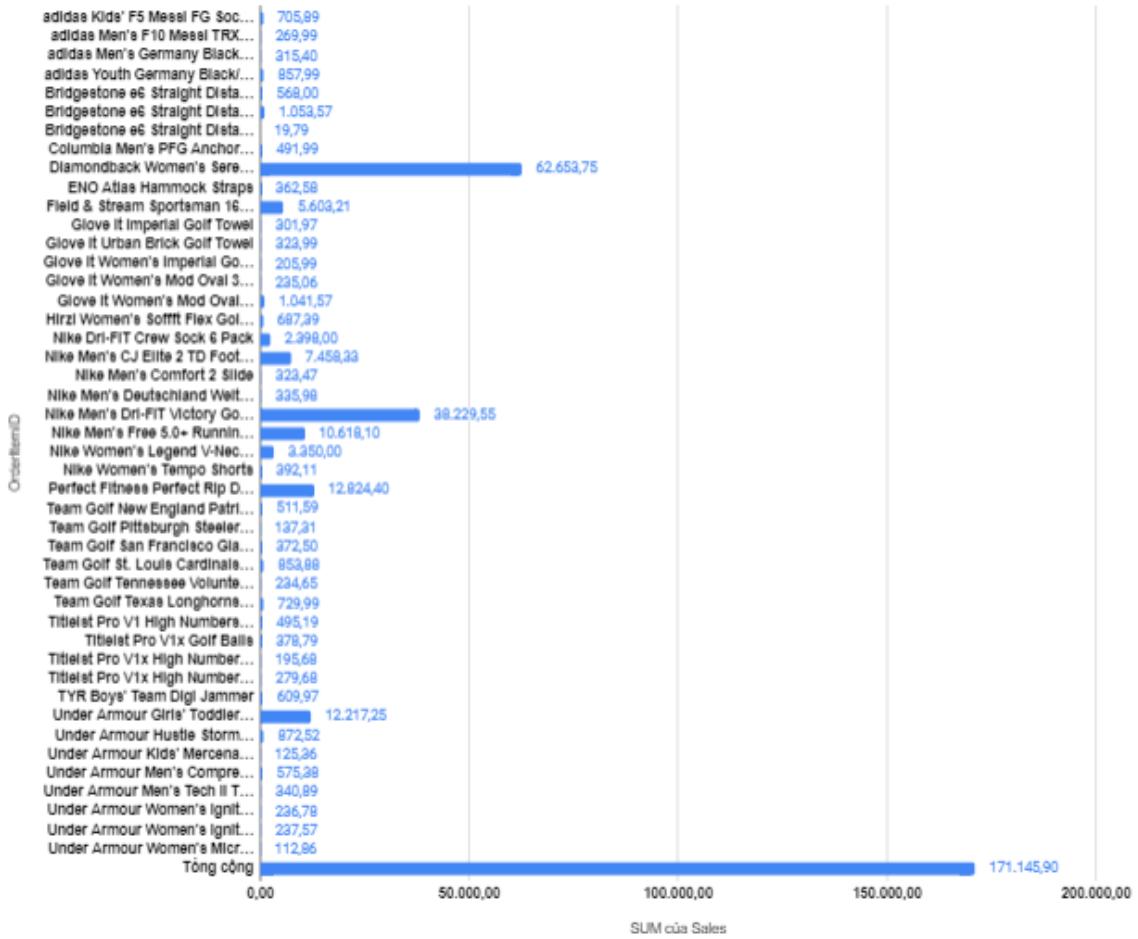


- Sau khi chọn xong có thể xem kết quả ở bảng số liệu thống kê

NameProduct	SUM của Sales
adidas Kids' F5 Messi FG Soccer Cleat	705,89
adidas Men's F10 Messi TRX FG Soccer Cleat	269,99
adidas Men's Germany Black Crest Away Tee	315,40
adidas Youth Germany Black/Red Away Match Soc	857,99
Bridgestone e6 Straight Distance NFL Carolina	568,00
Bridgestone e6 Straight Distance NFL San Dieg	1.053,57
Bridgestone e6 Straight Distance NFL Tennesse	19,79
Columbia Men's PFG Anchor Tough T-Shirt	491,99
Diamondback Women's Serene Classic Comfort Bi	62.653,75
ENO Atlas Hammock Straps	362,58
Field & Stream Sportsman 16 Gun Fire Safe	5.603,21
Glove It Imperial Golf Towel	301,97
Glove It Urban Brick Golf Towel	323,99
Glove It Women's Imperial Golf Glove	205,99
Glove It Women's Mod Oval 3-Zip Carry All Gol	235,06
Glove It Women's Mod Oval Golf Glove	1.041,57
Hirzl Women's Soffit Flex Golf Glove	687,39
Nike Dri-FIT Crew Sock 6 Pack	2.398,00
Nike Men's CJ Elite 2 TD	7.450,00

Và xem biểu đồ thống kê trực quan

TỔNG DOANH THU CỦA TỪNG SẢN PHẨM



Từ biểu đồ có thể thấy tổng doanh thu của năm 2016 là 171.145,90 USD trong đó, sản phẩm Diamon Back Women Serence Classic Comfort BI đạt doanh thu cao nhất với 62 653,75 USD

4.4.2. Câu hỏi: Cho biết các sản phẩm có doanh thu cao nhất theo phân khúc khách hàng Corporate (Doanh nghiệp) trong 3 năm 2015,2016,2017

Chọn bộ lọc

Giá trị

Thêm

CategoryID

Tóm tắt theo

SUM

Hiển thị dưới dạng

Lọc theo điều kiện

Lọc theo giá trị

Chọn tất cả-Xóa

OK

Hủy

CustomerID

CustomerSegment

OrderID

OrderItemID

2015

2016

2017

2018

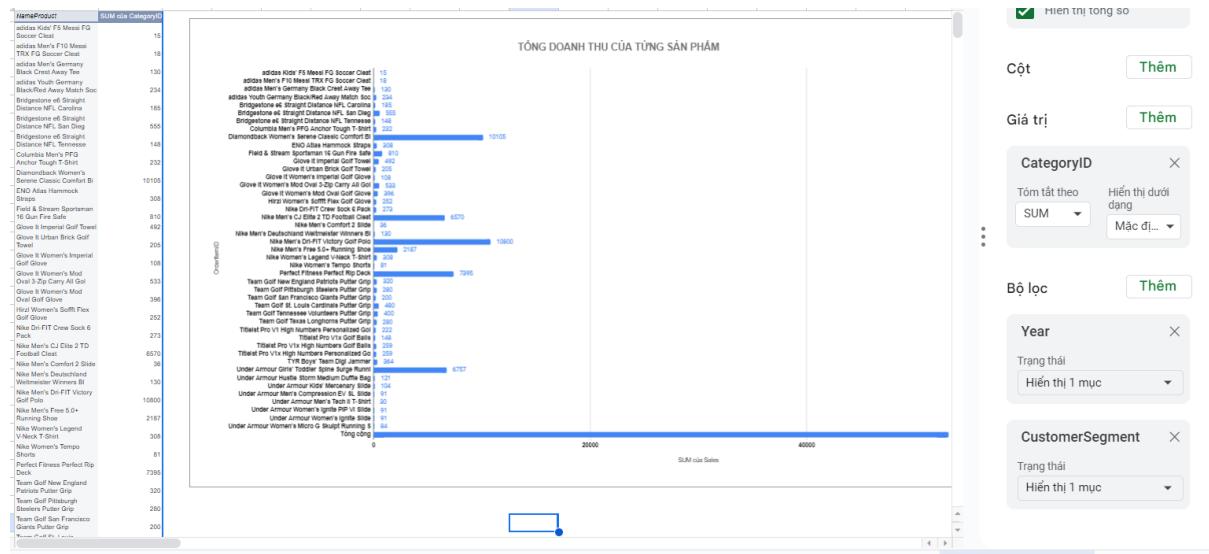
Consumer

Corporate

Home Office

OK

Hủy



Theo kết quả thống kê, có thể thấy sản phẩm có doanh thu cao nhất trong 3 năm 2015, 2016, 2017 đối với phân khúc khách hàng Coporate là Nike Men Victory Fit Golf Polo với 109 900 USD.

4.4.3. Câu hỏi: Cho biết doanh số thu được theo từng khách hàng để tìm ra khách hàng tiềm năng

Chọn thứ tự sắp xếp giảm dần theo SumOfOrderItem

FACT_SALE!1:17986	<input type="button" value="Xoá"/>	<input type="text" value="Tìm kiếm"/>
	<input type="button" value="Xoá tất cả"/>	
Được đề xuất	<input type="button" value="▼"/>	
Hàng	<input type="button" value="Thêm"/>	
CustomerID	<input type="button" value="X"/>	
Thứ tự	Sắp xếp theo	
Giảm ...	SUM tr...	
<input checked="" type="checkbox"/> Hiển thị tổng	CustomerID	
Cột	<input type="button" value="Thêm"/>	SUM trong tổng số Order Item Total
Giá trị	<input type="button" value="Thêm"/>	
Order Item Total	<input type="button" value="X"/>	
Tóm tắt theo	Hiển thị dưới dạng	
SUM	Mặc định...	
Bộ lọc	<input type="button" value="Thêm"/>	
g tổng hợp 1	<input type="button" value="▼"/>	

CustomerID	SUM của Order Item
496	1.909,23
9349	1.838,04
8380	1.833,01
4257	1.603,70
1948	1.548,82
8688	1.495,49
8027	1.470,57
6994	1.451,23
6320	1.435,11
9494	1.382,79
11733	1.345,72
823	1.334,39
10492	1.334,34
11223	1.325,63
1868	1.304,16
2927	1.301,34
7027	1.297,45
2851	1.269,31
335	1.255,80
5877	1.253,91
1503	1.239,79
4706	1.227,99
1423	1.225,67

Từ bảng danh sách thu được có thể thấy những khách hàng ở vị trí trên ưng chính là những khách hàng tiềm năng của chuỗi cung ứng. Dựa vào đó có thể áp dụng ưu đãi đặc biệt cho những khách hàng tiềm năng này để kích cầu lượng mua sắm và gắn bó lâu dài với hệ thống chuỗi cung ứng Dataco.

CHƯƠNG 5: KẾT LUẬN

5.1. Kết quả đạt được

Thực hiện xây dựng ‘*KHO DỮ LIỆU CHO HỆ THỐNG CHUỖI CUNG ỨNG DATACO*’ giúp chúng em hiểu sâu hơn về kiến thức môn học Kho Dữ liệu và việc sử dụng các công cụ hỗ trợ quá trình ETL đưa dữ liệu vào kho dữ liệu như công cụ SSIS. Nâng cao kỹ năng nhìn nhận và tiếp thu, xử lý thông tin đầu vào, đặt những câu hỏi tiền đề và thảo luận hướng đi để giải quyết và trả lời câu hỏi.

5.2. Những hạn chế

Trong quá trình thực hiện xây dựng kho dữ liệu, nhóm chúng em không tránh khỏi gặp một số những khó khăn vì còn hạn chế về mặt kiến thức. Tập dữ liệu còn nhiều dữ kiện chưa được khai thác hết. Hơn nữa, vì một phần chủ quan mà nhóm chúng em gặp nhiều vấn đề trong quá trình triển khai đồ án dẫn đến chưa được hoàn chỉnh như đúng kỳ vọng.

5.3. Tài liệu tham khảo

- [1] Tài liệu các file PDF hướng dẫn về Data Integration với SQL Server, SSIS, SSAS của GVHD Ths. Nguyễn Văn Thành.
- [2] Minewiskan(no date), *Lesson 3: Modifying Measures, Attributes and Hierarchies*. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/lesson-3-modifying-measures-attributes-and-hierarchies?view=asallproducts-allversions>. [Accessed 15 May 2024].
- [3]Minewiskan(no date), *Lesson 4: Defining Advanced Attribute and Dimension Properties*. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/lesson-4-defining-advanced-attribute-and-dimension-properties?view=asallproducts-allversions>. [Accessed 12 May 2024].
- [4] Nguyễn Văn Chúc (no date). *ETL Project From Excel Data Source to Star Schema with SSIS*. [online] youtube.com. Available at: https://www.youtube.com/watch?v=Yp8fXLnVCp8&ab_channel=ChucNguyenVan. [Accessed 15 May 2024].