

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 BÁN HÀNG
CHO CÔNG TY THIẾT BỊ PHẦN CỨNG**

**Môn học: KHO DỮ LIỆU
Mã lớp học phần: DAWH430784_22_2_01
GVHD: ThS. Nguyễn Văn Thành**

Nhóm sinh viên thực hiện: Nhóm 14

Trần Lê Ngọc Gia Hân	20133038
Nguyễn Thị Hoàng Trang	20133099
Nguyễn Thị Thanh Ngân	20133018

TP. Hồ Chí Minh, 13 tháng 05 năm 2023

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

Môn: Kho dữ liệu - HỌC KÌ II – NĂM HỌC 2022 – 2023

STT	HỌ VÀ TÊN	MSSV	TỶ LỆ ĐÓNG GÓP
1	Trần Lê Ngọc Gia Hân	20133038	100%
2	Nguyễn Thị Hoàng Trang	20133099	100%
3	Nguyễn Thị Thanh Ngân	20133018	100%

Nhận xét của giảng viên:

Ngày ... tháng 05 năm 2023

Giảng viên chấm điểm

Ths. Nguyễn Văn Thành

LỜI CẢM ƠN

Lời đầu tiên, nhóm chúng em xin gửi lời cảm ơn chân thành đến **Trường Đại học Sư phạm Kỹ thuật TPHCM** đã đưa môn học **Kho Dữ Liệu** vào chương trình giảng dạy. Đặc biệt, chúng em xin gửi lời cảm ơn sâu sắc đến giảng viên bộ môn – **Th.S Thầy Nguyễn Văn Thành** đã dạy dỗ, truyền đạt những kiến thức quý báu cho chúng em trong suốt thời gian học tập vừa qua. Trong thời gian tham gia lớp học Kho Dữ Liệu của Thầy, nhóm em đã có thêm cho mình nhiều kiến thức bổ ích, tinh thần học tập hiệu quả, nghiêm túc. Đây chắc chắn sẽ là những kiến thức quý báu, là hành trang để chúng em có thể vững bước sau này. Bộ môn Kho Dữ Liệu là môn học thú vị, vô cùng bổ ích và có tính thực tế cao. Đảm bảo cung cấp đủ kiến thức, gắn liền với nhu cầu thực tiễn của sinh viên. Mặc dù chúng em đã cố gắng hết sức nhưng chắc chắn bài đồ án này khó có thể tránh khỏi những thiếu sót và nhiều chỗ còn chưa chính xác, kính mong Thầy xem xét và góp ý để bài đồ án của chúng em được hoàn thiện hơn.

Nhóm chúng em xin chân thành cảm ơn!

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	2
1.2.2.1. Bảng CUSTOMERS	3
1.2.2.2. Bảng CONTACTS	3
1.2.2.3. Bảng EMPLOYEES	4
1.2.2.4. Bảng PRODUCTS	4
1.2.2.5. Bảng INVENTORIES	4
1.2.2.6. Bảng PRODUCT_CATEGORIES	5
1.2.2.7. Bảng WAREHOUSES	5
1.2.2.8. Bảng LOCATIONS	5
1.2.2.9. Bảng COUNTRIES	5
1.2.2.10. Bảng REGIONS	6
1.2.2.11. Bảng ORDERS	6
1.2.2.12. Bảng ORDER_ITEMS	6
1.3. Giới thiệu các công cụ được sử dụng	6
CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP)	7
2.1. Xác định các Business Process và bảng Fact	7
2.1.1. Business Process: Sales Analysis	7
2.1.2. Business Process: Inventory Management	7
2.2. Xây dựng các bảng Dimension	8
2.2.1. DimDate	8
2.2.2. DimGeography	8
2.2.3. DimCustomer	8
2.2.4. DimEmployee	9
2.2.5. DimInventory	9
2.2.6. DimProduct	9
2.2.6. DimCategory	9
2.3. Star Schema (Lược đồ hình sao)	10
CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS).....	11
3.1. Tạo ETL project và solution	11
3.2. Thiết lập các Connection	12
3.3. Import Dữ liệu vào các bảng dimension:	17
3.3.1. Date Dimension	17
3.3.1.1: Tạo Package	18
3.3.1.2: Thiết lập Control Flow	20
3.3.1.3: Load dữ liệu từ Source vào Stage	21
3.3.1.4: Thiết lập truncate table SQL task	31
3.3.1.5: Load dữ liệu từ Stage vào Dimension	32
3.3.2. Product Dimension	40
3.3.2.1. Tạo Package	40

3.3.2.2. Thiết lập Control Flow	40
3.3.2.3. Load dữ liệu từ Source vào Stage	41
3.3.2.4. Load dữ liệu từ Stage vào Dimension	43
3.3.3. Geography Dimension	43
3.3.3.1. Load dữ liệu từ Source vào Stage	43
3.3.3.2. Load dữ liệu từ Stage vào Dimension	45
3.3.4. Customer Dimension	48
3.3.4.1. Load dữ liệu từ Source vào Stage	48
3.3.4.2. Load dữ liệu từ Stage vào Dimension	51
3.3.5. Employee Dimension	53
3.3.5.1. Load dữ liệu từ Source vào Stage	53
3.3.5.2. Load dữ liệu từ Stage vào Dimension	56
3.3.6. Inventory Dimension	58
3.3.6.1. Load dữ liệu từ Source vào Stage	58
3.3.6.2. Load dữ liệu từ Stage vào Dimension	60
3.3.7. Category Dimension	63
3.3.7.1. Load dữ liệu từ Source vào Stage	63
3.3.7.2. Load dữ liệu từ Stage vào Dimension	66
3.4. Import dữ liệu vào bảng các bảng fact	68
3.4.1. Fact Sales	68
3.4.1.1. Tạo Package	68
3.4.1.2. Thiết lập Control Flow	68
3.4.1.3. Load dữ liệu từ Source vào Stage	70
3.4.1.4. Load dữ liệu từ Stage vào Fact	73
3.4.2. Fact Product Inventory	78
3.4.2.1. Load dữ liệu từ Source vào Stage	78
3.4.2.4. Load dữ liệu từ Stage vào Fact	80
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)	87
4.1. Quá trình xây dựng mô hình	87
4.1.1. Tạo Data Source	87
4.1.2. Tạo Data Source View	88
4.2. Quá trình xây dựng khối Cube	90
4.3. Cấu hình Hierarchy	92
4.3.1. Tạo Hierarchy cho Dim Date	92
4.3.2. Tạo Hierarchy cho Dim Geography	94
4.3.3. Tạo Hierarchy cho Dim Inventory	96
4.3.4. Tạo Hierarchy cho Dim Product	96
4.3.5. Tạo Hierarchy cho Dim Employee	99
4.4. Thực hiện phân tích dữ liệu	103
4.4.1. Câu hỏi: Cho biết xếp hạng KPI bán hàng của nhân viên trong tháng/ quý/ năm	103
4.4.1.1. Sử dụng công cụ SSAS	103
4.4.1.2. Sử dụng Pivot Table trong Excel	104
4.4.1.3. Sử dụng Power BI	104

4.4.2. Câu hỏi: Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất trong tháng/ quý/ năm.	105
4.4.2.1. Sử dụng công cụ SSAS	105
4.4.2.2. Sử dụng Pivot Table trong Excel	106
4.4.2.3. Sử dụng Power BI.....	106
4.4.3. Câu hỏi: Cho biết doanh số bánh hàng theo từng khách hàng trong tháng/ quý/ năm.	107
4.4.3.1. Sử dụng công cụ SSAS	107
4.4.3.2. Sử dụng Pivot Table trong Excel	109
4.4.3.3. Sử dụng Power BI	109
4.4.4. Câu hỏi: Quản lý số lượng sản phẩm phân bổ trong các kho.	111
4.4.4.1. Sử dụng công cụ SSAS	111
4.4.4.2. Sử dụng Pivot Table trong Excel	112
4.4.4.3. Sử dụng Power BI	113
CHƯƠNG 5: KẾT LUẬN	115
5.1. Kết quả đạt được	115
5.2. Những hạn chế	115
5.3. Tài liệu tham khảo	115

Link video báo cáo :

<https://drive.google.com/drive/folders/1QNDBXkR2jxrMryIa6gCJhvIBk3IWHMPW?fbclid=IwAR2ZMtYbPs3MSL-wp8NFV-0hRr61Uta1RTHlenJzsUpNgFNKE6f5B3fAfdk>

PHÂN CÔNG NHIỆM VỤ

Nhiệm vụ	Hoàng Trang	Gia Hân	Thanh Ngân
Tìm kiếm tập dữ liệu		x	
Mô hình hóa dữ liệu vào workbook	x	x	x
+ DimDate			
+ DimGeography	x		
+ DimCustomer			
+ DimEmployee			x
+ DimInventory			
+ FactProductInventory			
+ DimProduct		x	
+ DimCategory			
+ FactSales			
Thực hiện ETL đưa dữ liệu vào các bảng dim và fact dùng SSIS	x	x	x
+ Tạo và đưa dữ liệu vào các dim stages và dim			x
+ Tạo và đưa dữ liệu vào FactSales stage và FactSales		x	
+ Tạo và đưa dữ liệu vào FactProductInventory stage và FactProductInventory	x		
Nhập dữ liệu vào SSAS, tạo data source, data source view và cube	x		
Tạo các phân cấp hierarchy cho các bảng dim và các attribute relationships.	x	x	x
+ Hierarchy cho DimDate	x		
+ Hierarchy cho DimGeography, DimProduct			x
+ Hierarchy cho DimInventory, DimEmployee		x	

Nhiệm vụ	Hoàng Trang	Gia Hân	Thanh Ngân
Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Sales Analysis	x	x	x
+ Cho biết xếp hạng KPI bán hàng của nhân viên trong tháng/ quý/ năm	x		
+ Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất trong tháng/ quý/ năm		x	
+ Cho biết danh số bán hàng theo từng khách hàng theo tháng/ quý/ năm			x
Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Sales Analysis: quản lý số lượng sản phẩm phân bổ trong các kho	x	x	x
+ Trả lời câu hỏi dùng SSAS			x
+ Trả lời câu hỏi dùng Power BI Desktop	x		
+ Trả lời câu hỏi dùng Excel Pivot		x	

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

1.1. Lý do chọn đề tài

- Trong kỷ nguyên công nghệ ngày nay, hầu như ở bất cứ đâu, bất cứ khi nào cũng có sự xuất hiện và hoạt động của các thiết bị công nghệ. Vì nhu cầu cao về sử dụng thiết bị công nghệ, việc cung cấp, sửa chữa và bảo trì kịp thời đóng một vai trò rất quan trọng. Để những quá trình từ các khâu sản xuất, vận chuyển lưu trữ, phân phối sản phẩm được diễn ra một cách mượt mà, tự động, cần có một hệ thống quản lý tốt tất cả các khâu. Thông thường ở mỗi giai đoạn như vậy, dữ liệu lưu trữ cũng khác nhau, các doanh nghiệp nếu chuẩn bị kỹ càng từ bước kết hợp dữ liệu này thì sẽ có khả năng cao hơn trong việc đưa ra những quyết định. Chính vì lý do đó, nhóm sinh viên chúng em chọn đề tài "Xây dựng Kho dữ liệu bán hàng cho công ty thiết bị phần cứng" 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

- Nhóm sử dụng Tập dữ liệu OT Database được lấy từ trang web Oracle Tutorial (oracletutorial.com), trang web này cung cấp cơ sở dữ liệu mẫu theo chuẩn của Oracle để phục vụ cho việc luyện tập thao tác của người học Oracle.
- Đường dẫn tải tập dữ liệu: [OT Oracle Sample Database](#)

1.2.2. Mô tả chi tiết tập dữ liệu

- Tập dữ liệu ‘OT’ bao gồm thông tin về một công ty toàn cầu ảo chuyên kinh doanh các mặt hàng linh kiện điện tử. Dữ liệu bao gồm các 12 bảng với mô tả như sau:

Tên bảng	Mô tả	Số dòng
CONTACTS	lưu trữ thông tin người liên hệ của khách hàng	319 dòng
COUNTRIES	lưu trữ thông tin quốc gia	25 dòng
CUSTOMERS	lưu trữ thông tin khách hàng	319 dòng
EMPLOYEES	lưu trữ thông tin nhân viên	107 dòng
INVENTORIES	lưu trữ thông tin tồn kho của sản phẩm	1112 dòng
LOCATIONS	lưu trữ vị trí của kho hàng	23 dòng
ORDERS	lưu trữ thông tin chung của đơn hàng	105 dòng
ORDER_ITEMS	lưu trữ thông tin chi tiết sản phẩm trong đơn hàng	665 dòng
PRODUCT_CATEGORIES	lưu trữ thông tin loại hàng của sản phẩm	5 dòng
PRODUCTS	lưu trữ thông tin sản phẩm	288 dòng
REGIONS	lưu trữ khu vực nơi công ty hoạt động	4 dòng
WAREHOUSES	lưu trữ thông tin kho hàng	9 dòng

1.2.2.1. Bảng CUSTOMERS

Biến	Mô tả
CUSTOMER_ID	Mã khách hàng(Khóa chính): dùng để xác định và phân biệt các khách hàng với nhau.
NAME	Tên khách hàng
ADDRESS	Địa chỉ của khách hàng
WEBSITE	Địa chỉ website của khách hàng
CREDIT_LIMIT	Hạn mức tín dụng: đây là giới hạn số tiền mà khách hàng được cấp phép để mua sản phẩm hoặc dịch vụ từ công ty

1.2.2.2. Bảng CONTACTS

Biến	Mô tả
CONTACTS_ID	Mã liên lạc của khách hàng (Khóa chính)
FIRST_NAME	Họ của người đại diện khách hàng
LAST_NAME	Tên của người đại diện khách hàng
EMAIL	Email của người đại diện khách hàng
PHONE	Số điện thoại của người đại diện khách hàng
CUSTOMER_ID	Mã khách hàng (Khóa ngoại trỏ đến CUSTOMERS(CUSTOMER_ID))

1.2.2.3. Bảng EMPLOYEES

Biến	Mô tả
EMPLOYEE_ID	Mã nhân viên (Khóa chính)
FIRST_NAME	Họ nhân viên
LAST_NAME	Tên nhân viên
EMAIL	Email của nhân viên
PHONE	Số điện thoại của nhân viên
HIRE_DATE	Ngày tuyển dụng làm việc của nhân viên
MANAGER_ID	Mã người quản lý của nhân viên
JOB_TITLE	Vị trí công việc của nhân viên

1.2.2.4. Bảng PRODUCTS

Biến	Mô tả
PRODUCT_ID	Mã sản phẩm (Khóa chính)
PRODUCT_NAME	Tên sản phẩm
DESCRIPTION	Mô tả thông tin về sản phẩm
STANDARD_COST	Giá nhập của sản phẩm
LIST_PRICE	Giá bán của sản phẩm
CATEGORY_ID	Mã ngành hàng của sản phẩm (Khóa ngoại trỏ đến PRODUCT_CATEGORIES(CATEGORY_ID))

1.2.2.5. Bảng INVENTORIES

Biến	Mô tả
PRODUCT_ID	Mã sản phẩm (Khóa ngoại trỏ đến PRODUCTS(PRODUCT_ID))
WAREHOUSE_ID	Mã kho hàng (Khóa ngoại trỏ đến WAREHOUSES(WAREHOUSE_ID))
QUANTITY	Số lượng sản phẩm trong kho hàng

1.2.2.6. Bảng PRODUCT_CATEGORIES

Biến	Mô tả
CATEGORY_ID	Mã ngành hàng (Khóa chính)
CATEGORY_NAME	Tên ngành hàng

1.2.2.7. Bảng WAREHOUSES

Biến	Mô tả
WAREHOUSE_ID	Mã kho hàng (Khóa chính)
WAREHOUSE_NAME	Tên kho hàng
LOCATION_ID	Mã vị trí của kho

1.2.2.8. Bảng LOCATIONS

Biến	Mô tả
LOCATION_ID	Mã vị trí (Khóa chính)
ADDRESS	Địa chỉ cụ thể
POSTAL_CODE	Mã bưu chính
CITY	Tên thành phố
STATE	Tên bang
COUNTRY_ID	Mã quốc gia (Khóa ngoại trỏ đến COUNTRIES(COUNTRY_ID))

1.2.2.9. Bảng COUNTRIES

Biến	Mô tả
COUNTRY_ID	Mã quốc gia (Khóa chính)
COUNTRY_NAME	Tên quốc gia
REGION_ID	Mã vùng (Khóa ngoại trỏ đến REGIONS(REGION_ID))

1.2.2.10. Bảng REGIONS

Biên	Mô tả
REGION_ID	Mã vùng (Khóa chính)
REGION_NAME	Tên vùng

1.2.2.11. Bảng ORDERS

Biên	Mô tả
ORDER_ID	Mã đơn hàng (Khóa chính)
CUSTOMER_ID	Mã khách hàng (Khóa ngoại trỏ đến CUSTOMERS(CUSTOMER_ID))
STATUS	Tình trạng đơn hàng
SALESMAN_ID	Mã người bán hàng (Khóa ngoại trỏ đến EMPLOYEES(EMPLOYEE_ID))
ORDER_DATE	Ngày tạo đơn hàng

1.2.2.12. Bảng ORDER_ITEMS

Biên	Mô tả
ORDER_ID	Mã đơn hàng ((Khóa ngoại trỏ đến ORDERS(ORDER_ID)))
ITEM_ID	Thứ tự sản phẩm trong đơn hàng
PRODUCT_ID	Mã sản phẩm ((Khóa ngoại trỏ đến PRODUCTS(PRODUCT_ID)))
QUANTITY	Số lượng sản phẩm được đặt trong đơn hàng
UNIT_PRICE	Đơn giá sản phẩm

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
- Power BI.

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

2.1. Xác định các Business Process và bảng Fact

- Xây dựng Detailed Bus Matrix xác định các Business Process, bảng Fact, bảng Dim cần thiết.

Instructions!		Business Process Name	Fact Table	Fact Grain Type	Granularity	Facts	DimDate	DimGeography	DimCustomer	DimEmployee	DimInventory	DimCategory	DimProduct
Business Process	Fact Table	Granularity	Facts	DimDate	DimGeography	DimCustomer	DimEmployee	DimInventory	DimCategory	DimProduct			
Sales Analysis	FactSales	Transaction	one row per order detail (Unit Price (non-additive), Quantity(additive), Total amount (derived from order_items.unit_price, order_items.quantity))	x	x x					x			
Inventory Management	FactProductInventory	Periodic Snapshots	Quantity (additive)	x		x x x							

2.1.1. Business Process: Sales Analysis

- Các câu hỏi cụ thể được đặt ra:
 - + Cho biết xếp hạng KPI bán hàng của nhân viên trong tháng/ quý/ năm.
 - + Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất trong tháng/ quý/ năm.
 - + Cho biết doanh số bánh hàng theo từng khách hàng trong tháng/ quý/ năm.

- Bảng FactSales

FactSales	DateKey	ForeignKey to DimDate(DateKey)
FactSales	CustomerKey	ForeignKey to DimCustomer(CustomeKey)
FactSales	EmployeeKey	ForeignKey to DimEmployee(EmployeeKey)
FactSales	ProductKey	ForeignKey to DimProduct(ProductKey)
FactSales	UnitPrice	Unit Price of each product
FactSales	Quantity	Quantity of each product
FactSales	SalesAmount	Total amount sold

2.1.2. Business Process: Inventory Management

- Các câu hỏi cụ thể được đặt ra:
 - + Quản lý số lượng sản phẩm phân bổ trong các kho.

- Bảng FactProductInventory

FactProductInventory	DateKey	ForeignKey to DimDate(DateKey)
FactProductInventory	GeographyKey	ForeignKey to DimGeography(GeographyKey)
FactProductInventory	InventoryKey	ForeignKey to DimInventory(InventoryKey)
FactProductInventory	CategoryKey	ForeignKey to DimCategory(CategoryKey)
FactProductInventory	ProductKey	ForeignKey to DimProduct(ProductKey)
FactProductInventory	Quantity	Product quantity of each product

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. DimGeography

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description
	DimGeography	GeographyKey	A unique identifier for each geography area in the table
	DimGeography	LocationID	location id of store
	DimGeography	Region	store region where the company operates
	DimGeography	CountryCode	store country code in region
	DimGeography	CountryName	store country name
	DimGeography	State	store state in country
	DimGeography	City	store city
	DimGeography	PostalCode	Delivery Postal Code

2.2.3. DimCustomer

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description
	DimCustomer	CustomerKey	A unique identifier for each customer in the table
	DimCustomer	CustomerID	ID of each Customer
	DimCustomer	CustomerName	name of each Customer
	DimCustomer	Address	Address of each Customer
	DimCustomer	Website	Website of each Company
	DimCustomer	CreditLimit	Credit Limit of each customer

2.2.4. DimEmployee

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description
	DimEmployee	EmployeeKey	A unique identifier for each employee in the table
	DimEmployee	EmployeeID	ID of each employee
	DimEmployee	FirstName	First name of each employee
	DimEmployee	LastName	Last name of each employee
	DimEmployee	Email	Email of each employee
	DimEmployee	Phone	Phone number of each employee
	DimEmployee	HireDate	Hire date of each employee
	DimEmployee	ManagerID	ID of employee's manager
	DimEmployee	JobTitle	Job title of each employee

2.2.5. DimInventory

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description
	DimInventory	InventoryKey	A unique identifier for each inventory in the table
	DimInventory	ProductID	ID of a product
	DimInventory	WarehouseId	Id of each warehouse
	DimInventory	WarehouseName	Name of each warehouse

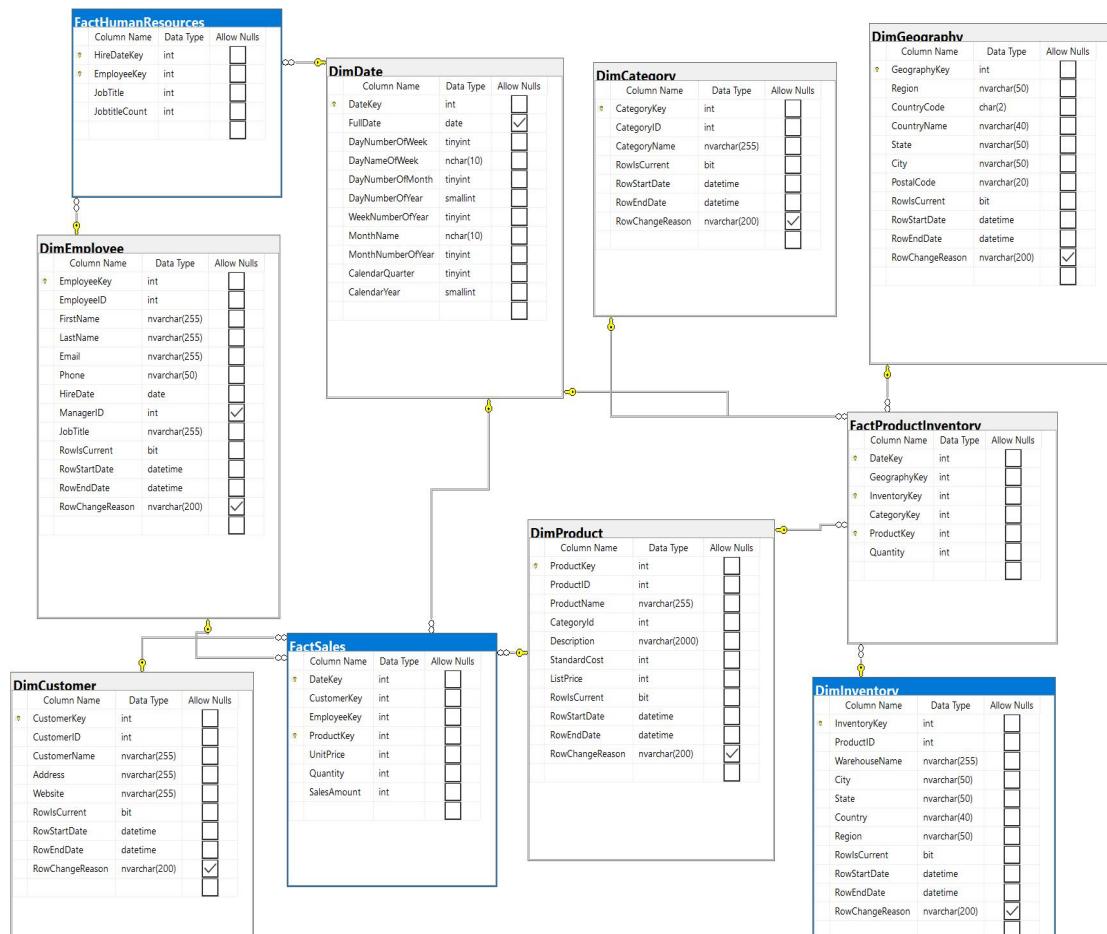
2.2.6. DimProduct

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	ProductName	Name of each product
	DimProduct	CategoryID	ID of product's category
	DimProduct	Description	Description of each product
	DimProduct	StandardCost	Standard cost of each product
	DimProduct	ListPrice	List price of each product

2.2.7. DimCategory

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description
	DimCategory	CategoryKey	A unique identifier for each category in the table
	DimCategory	CategoryID	ID of each Category
	DimCategory	CategoryName	Name of Category

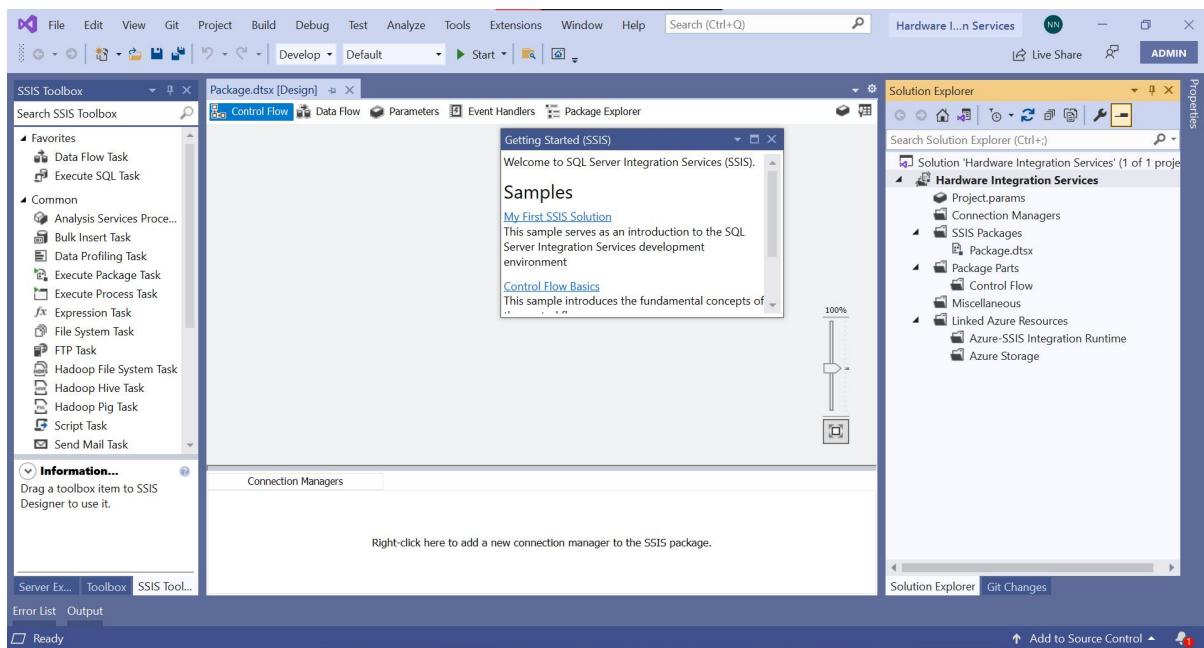
2.3. Star Schema (Lược đồ hình sao)



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

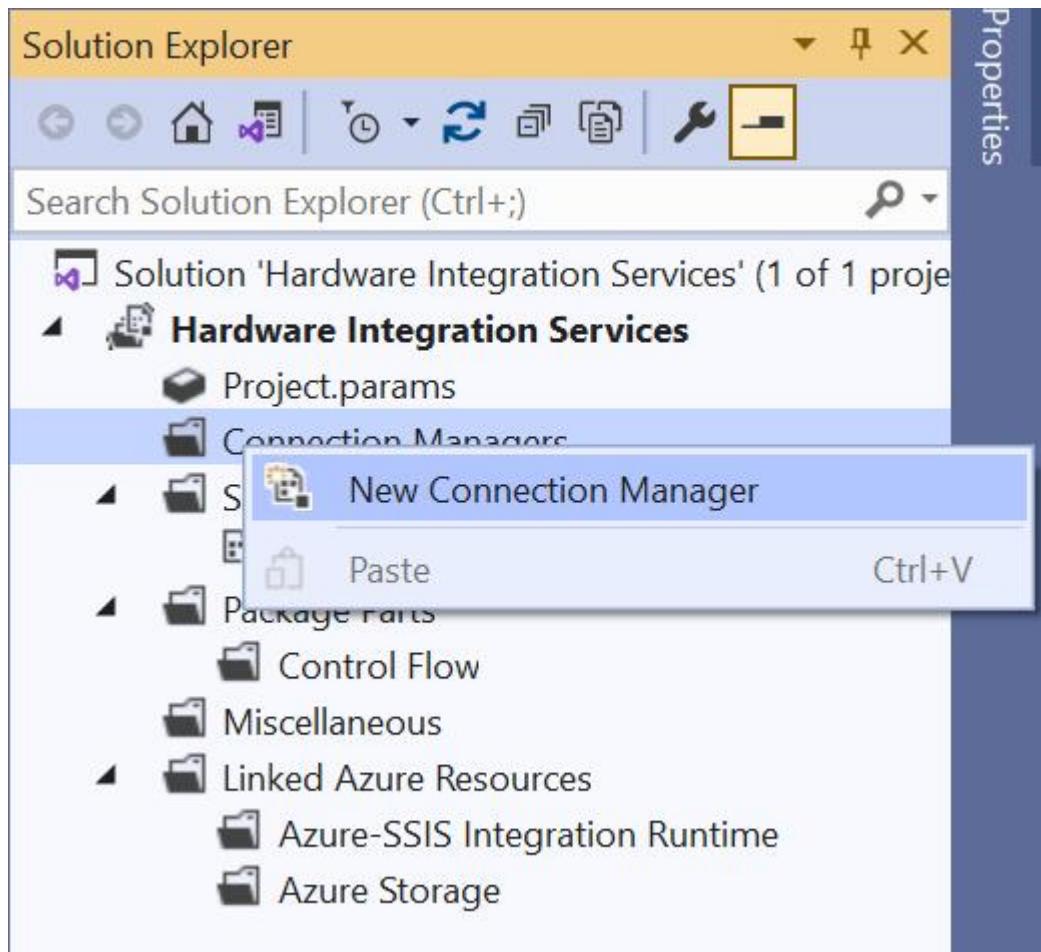
3.1. Tạo ETL project và solution

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

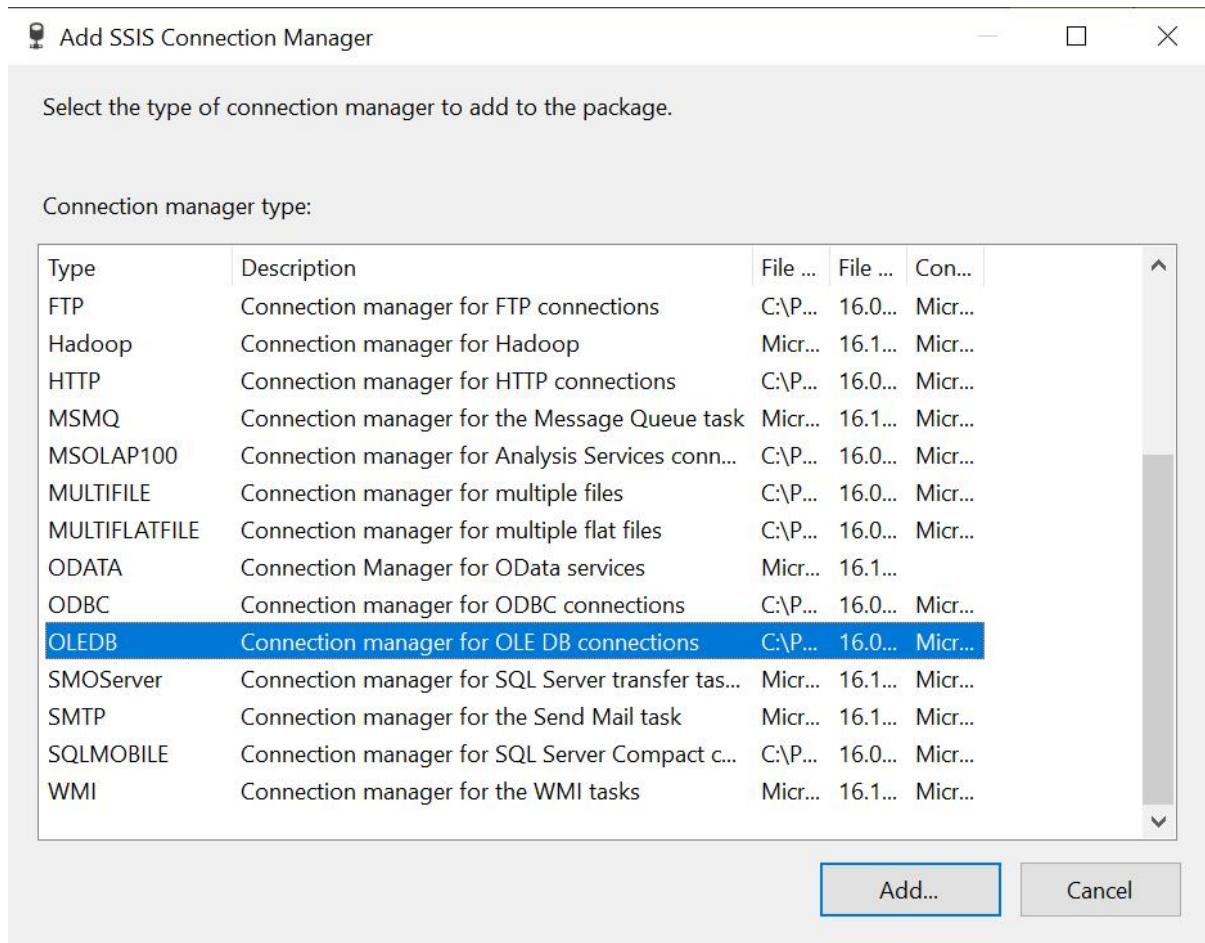


3.2. Thiết lập các Connection

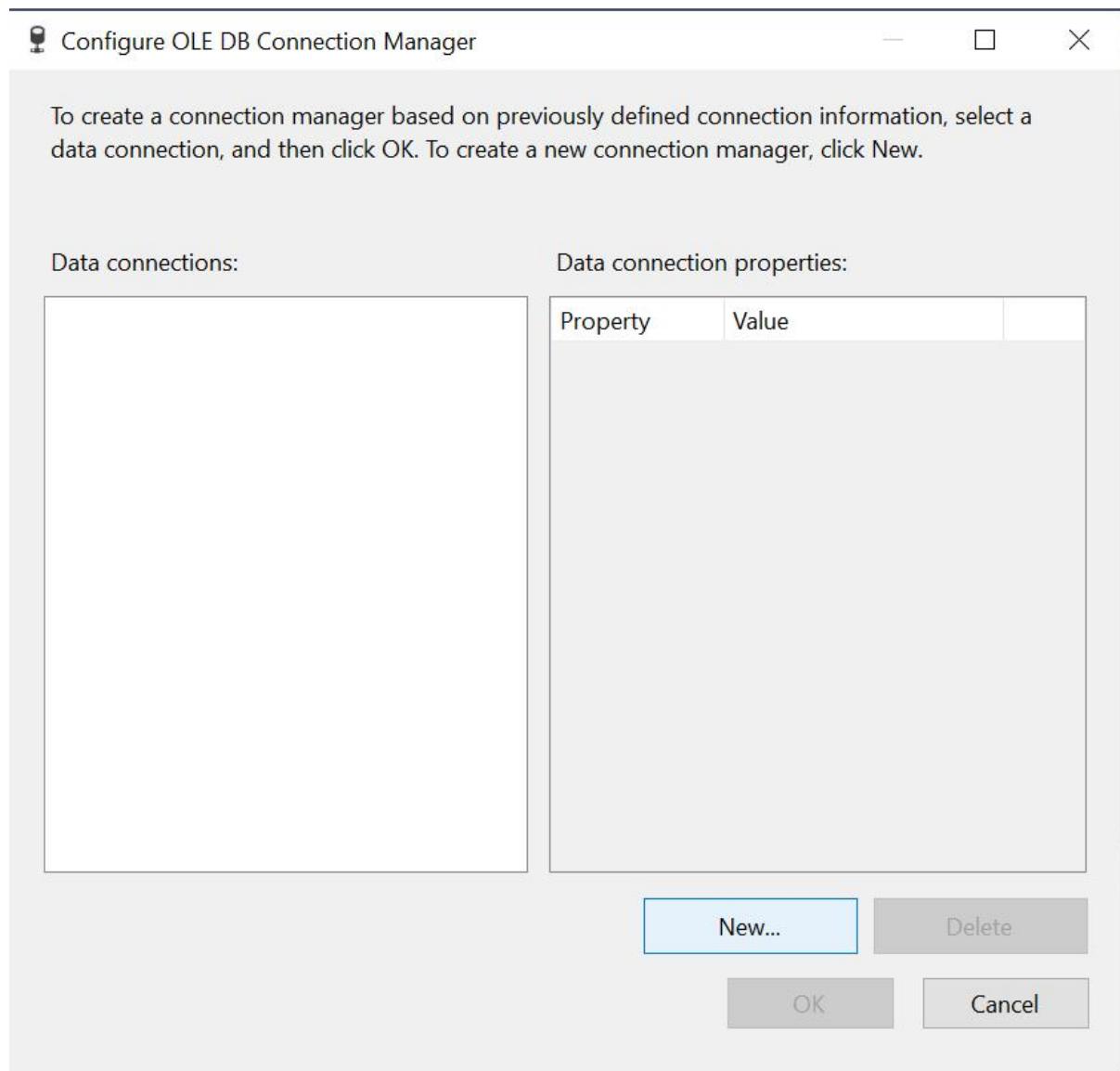
1. Tại **solution explorer**, right-click vào **Connection Managers**, chọn **New Connection Manager**.



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



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



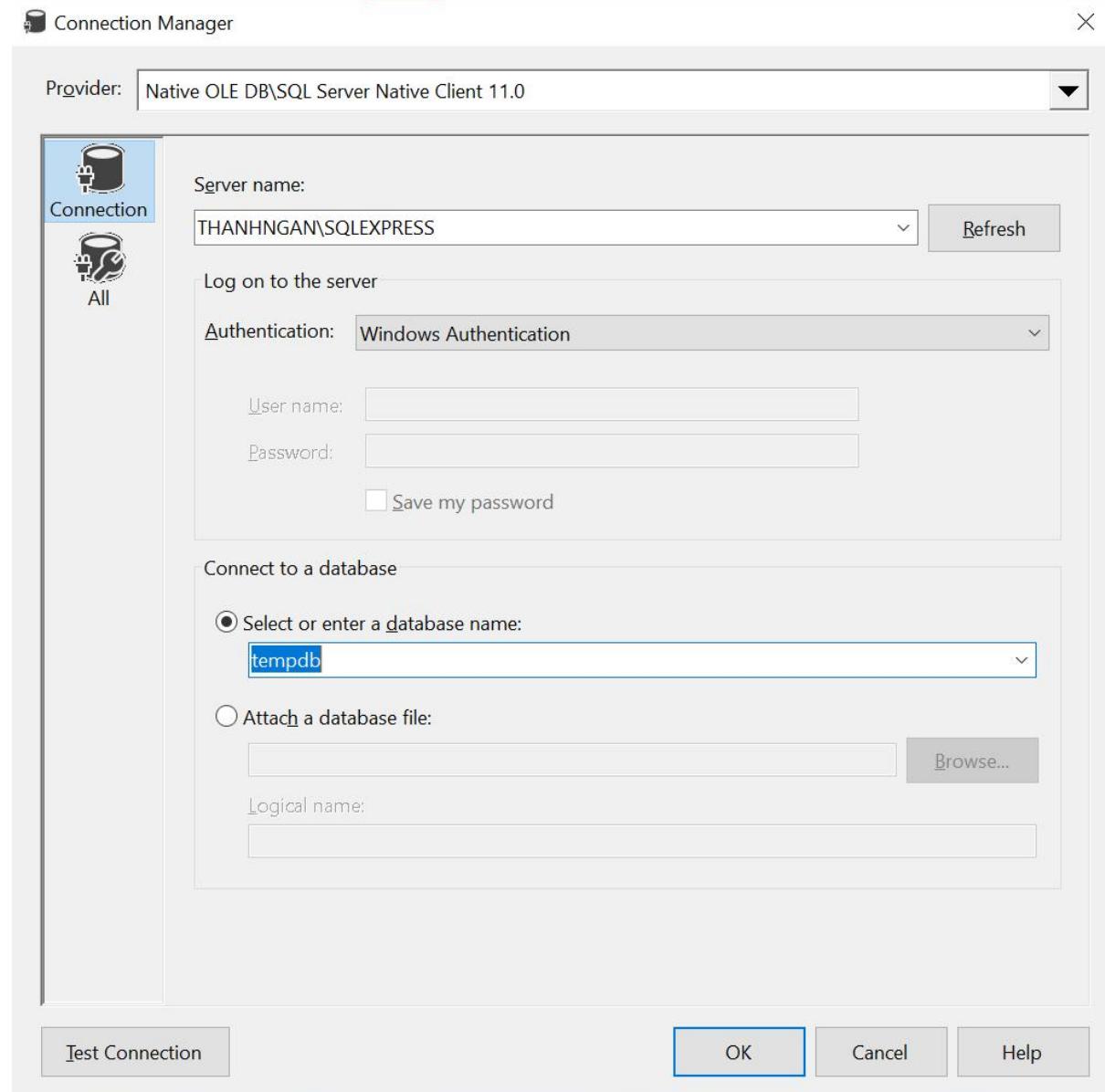
Đặt tên cho các thông tin như bên dưới:

Provider: **SQL Server Native Client**

Server Name: **THANHNGAN\SQLEXPRESS**

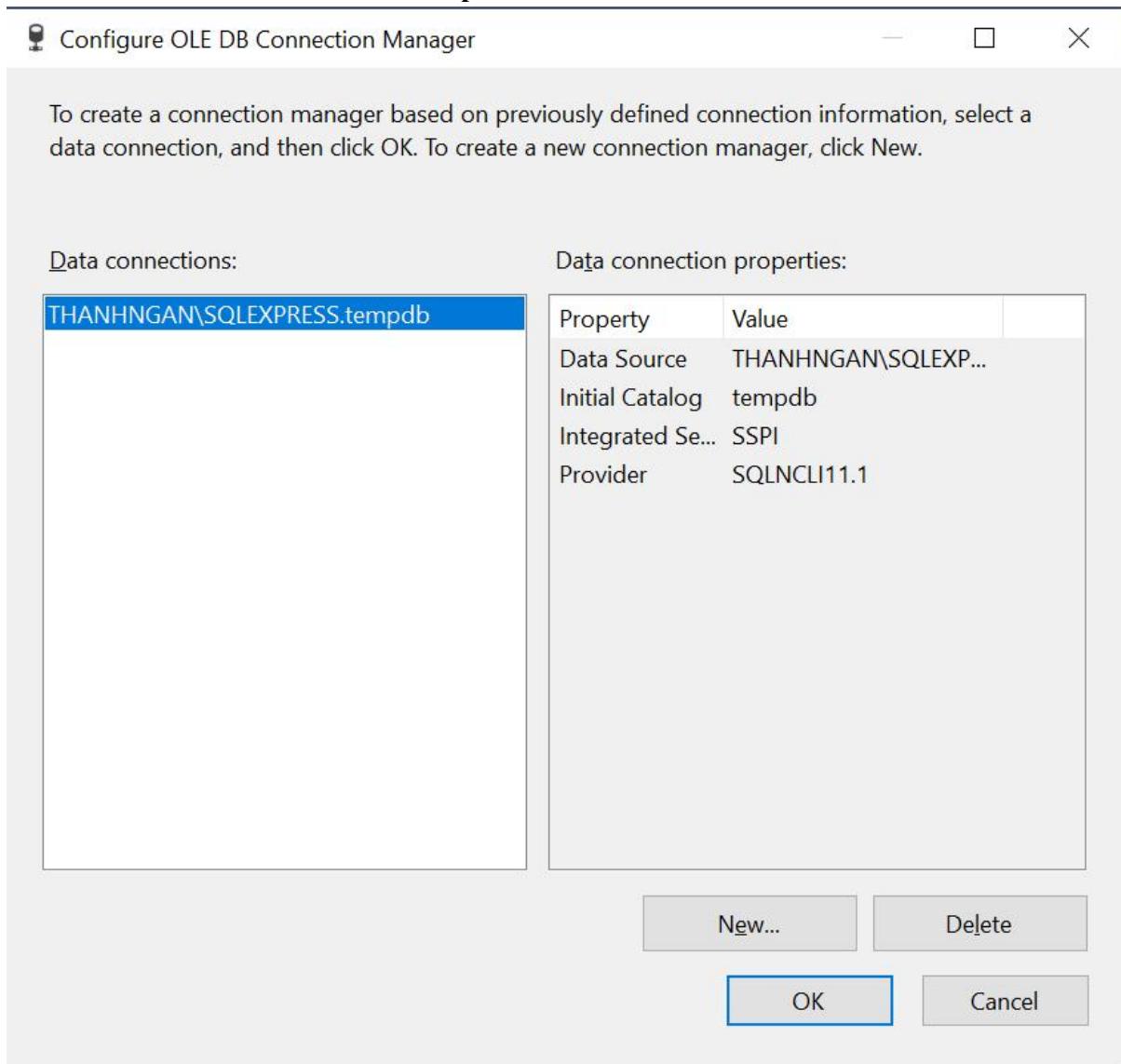
Log on to the Server: **Use Windows Authentication**

Connect to a database: **tempdb**



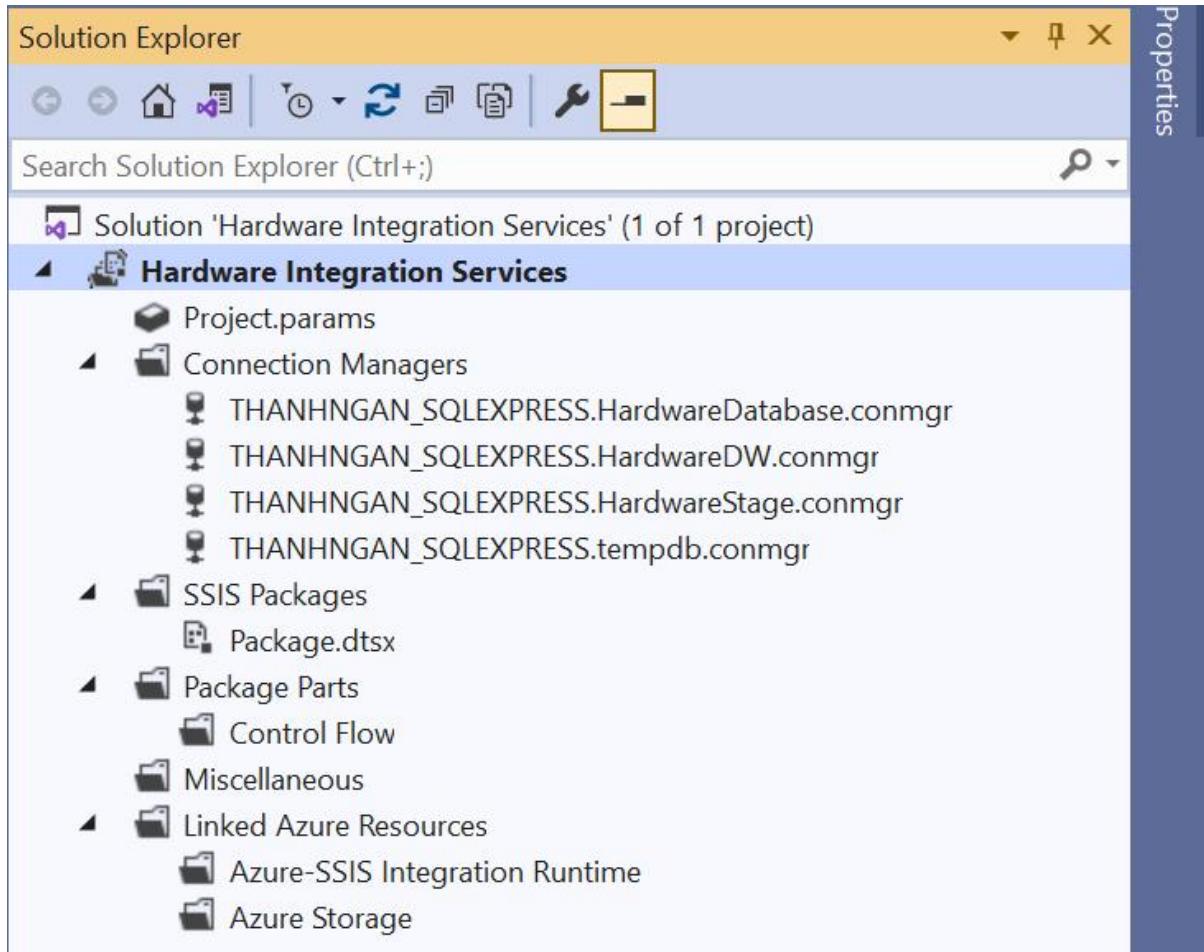
Click **Test Connection** để kiểm tra connection hoạt động, sau đó click **OK** để lưu connection.

4. Trở lại màn hình **Configure OLE DB Connection Manager**, ta được **THANHNGAN\SQLEXPRESS.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 **HardwareDW**, **HardwareStage** và **HardwareDatabase** source database. Sau khi tạo xong **Connection Managers** folder sẽ như bên dưới.



3.3. Import Dữ liệu vào các bảng dimension:

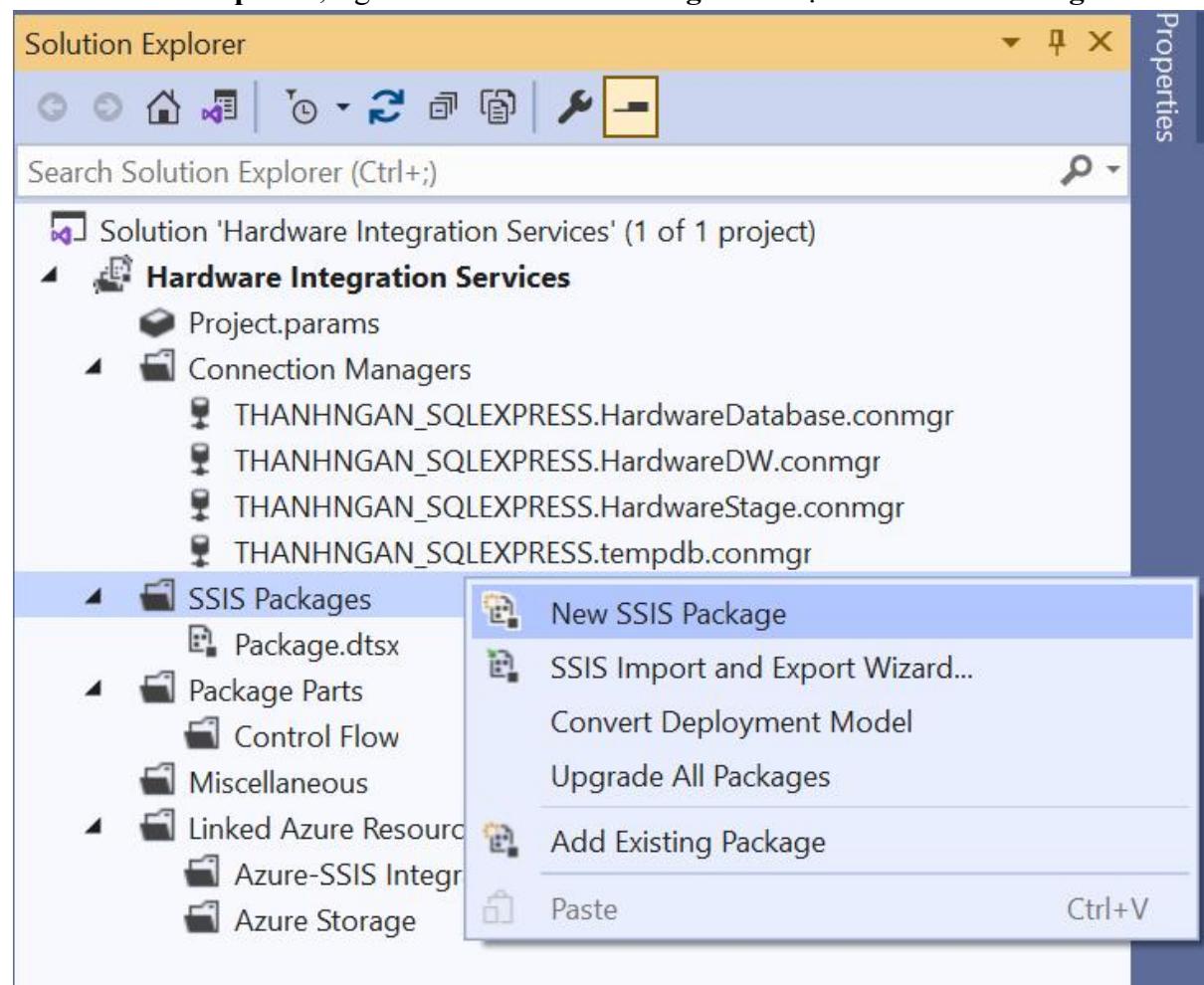
3.3.1. Date Dimension

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

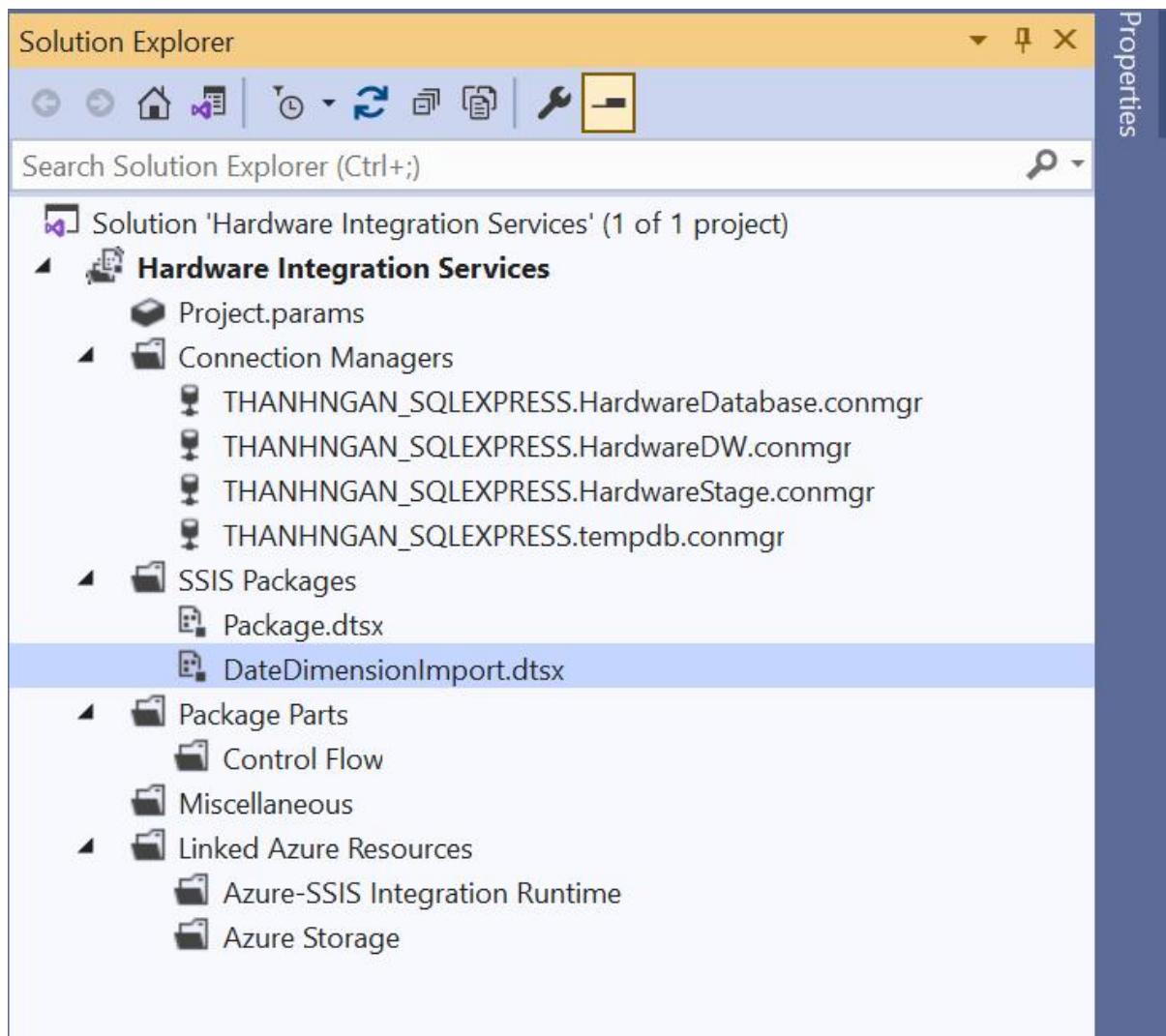
3.3.1.1: Tạo Package

Tạo một package mới, **DateDimensionImport**.

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



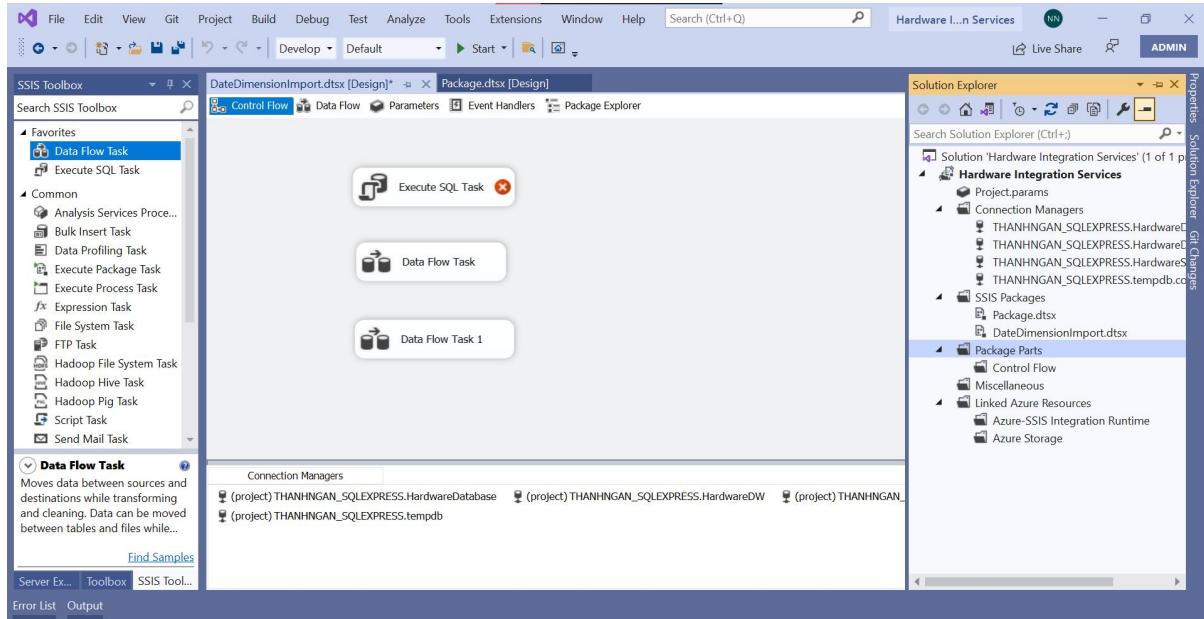
2. Right-Click vào package có tên **Package1.dtsx** và chọn **Rename**
3. Đặt tên package là **DateDimensionImport**
4. Sau khi thực hiện xong ta được **Package.dtsx** và **DateDimensionImport.dtsx** như hình sau:



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

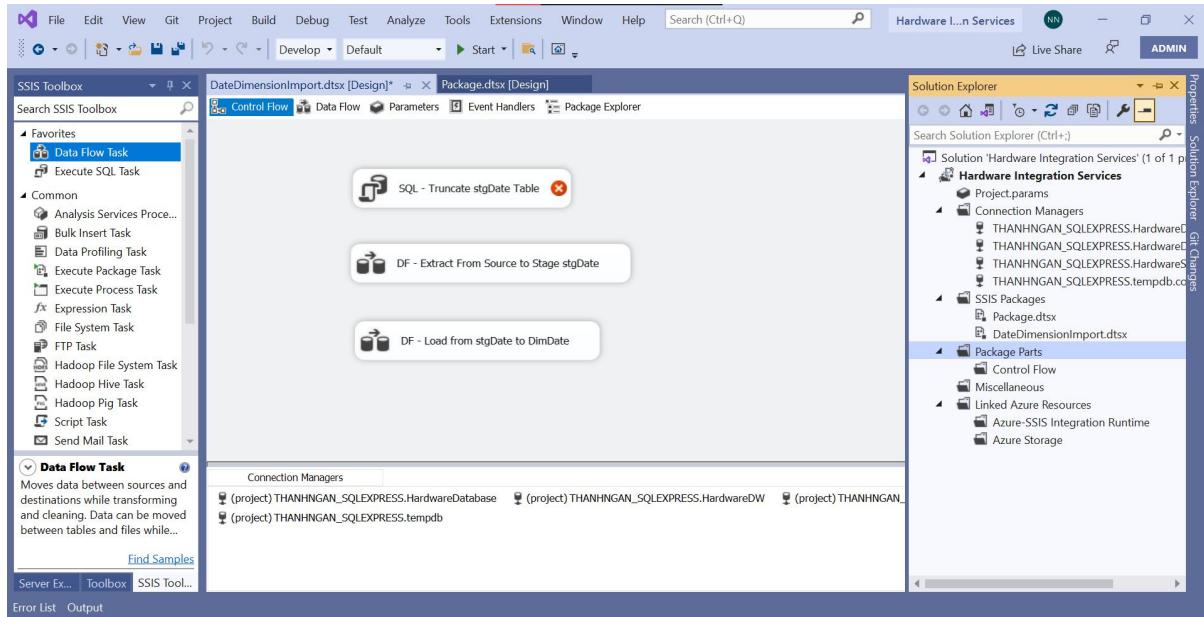
3.3.1.2: Thiết lập Control Flow

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

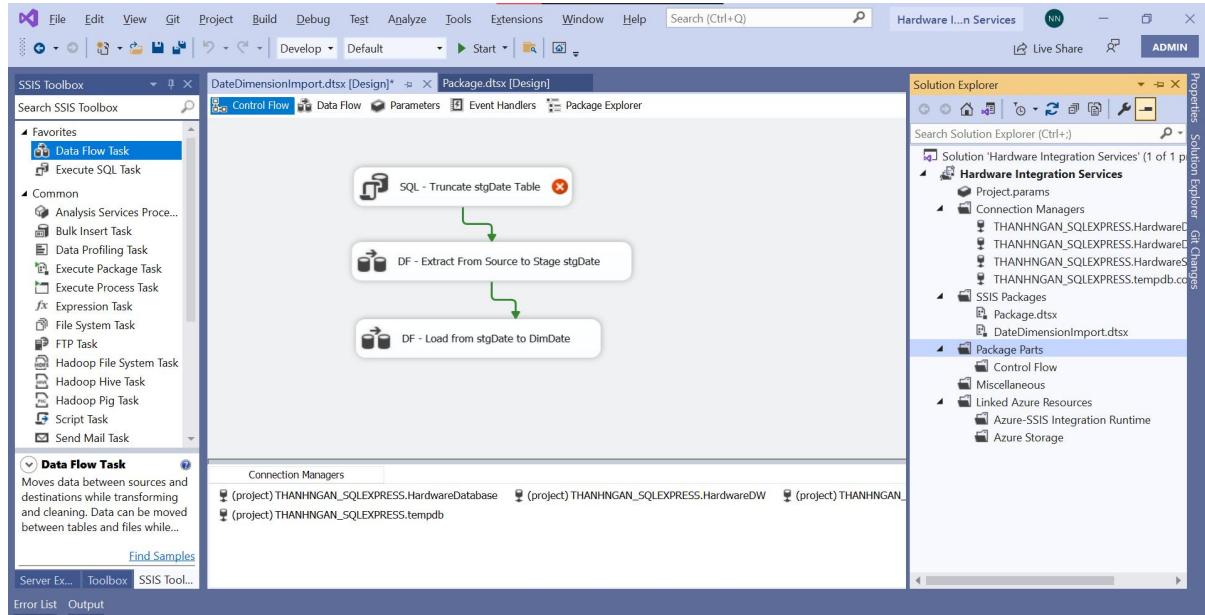


2. Đặt tên cho các task.

- 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**

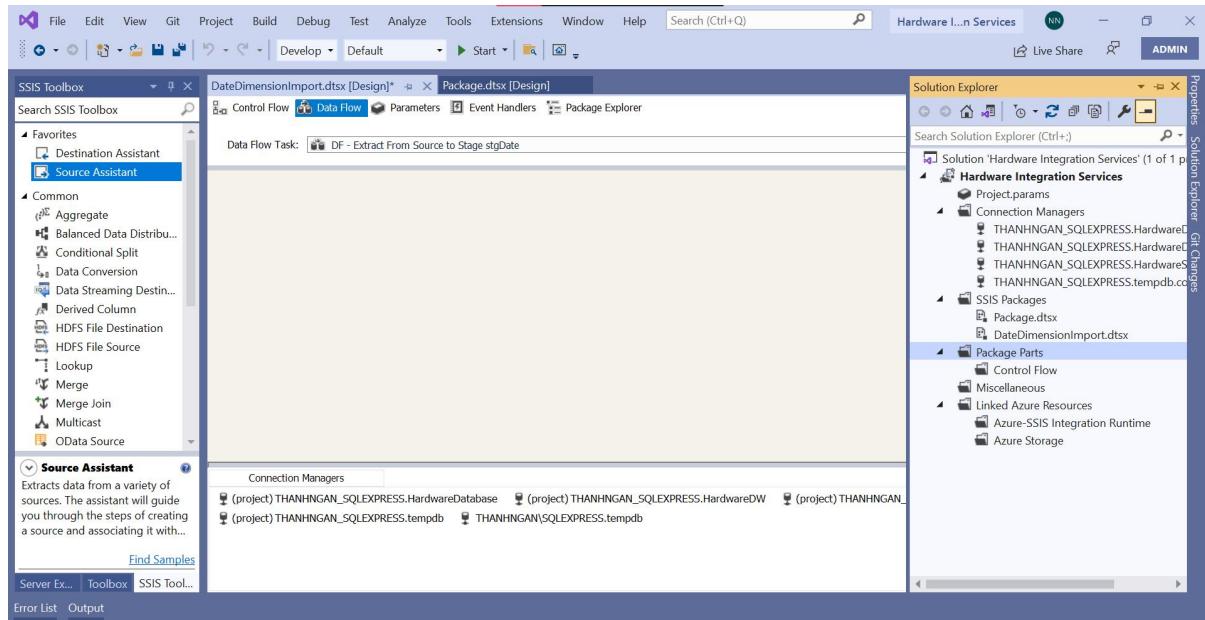


3. Connect các tasks để chúng có thể chạy theo thứ tự. Click vào **SQL - Truncate stgDate Table** task và một mũi tên xanh sẽ xuất hiện. Kéo rê nó vào **DF - Extract From Source To Stage stgDate** task. Tương tự, connect hai task còn lại.

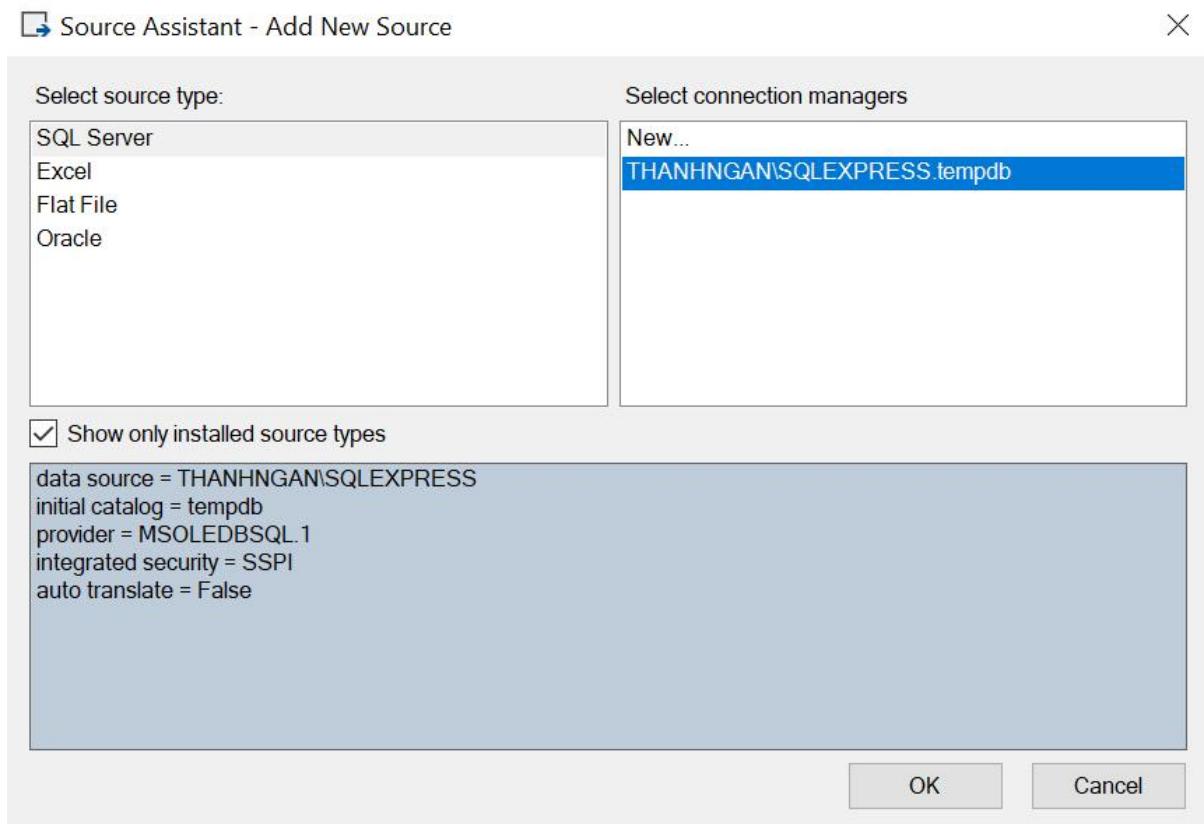


3.3.1.3: Load dữ liệu từ Source vào Stage

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

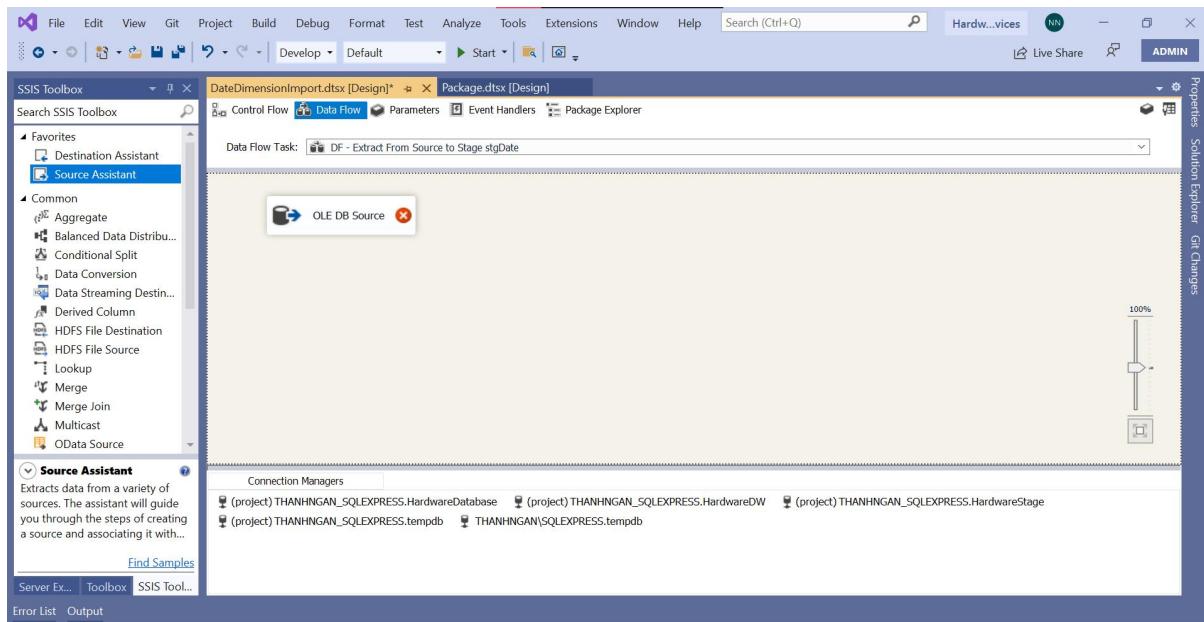


2. Từ **SSIS Toolbox**, kéo rê **Source Assistant** vào design surface, một dialog mở ra, Chọn source Type: **SQL Server** và **tempdb** connection manager.

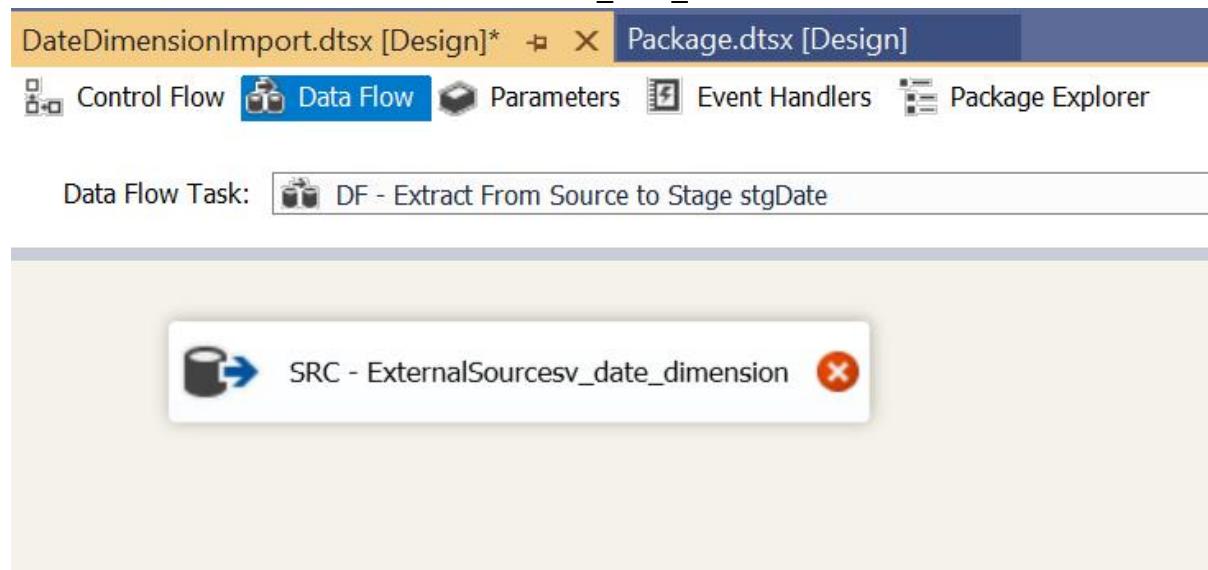


Click **OK**.

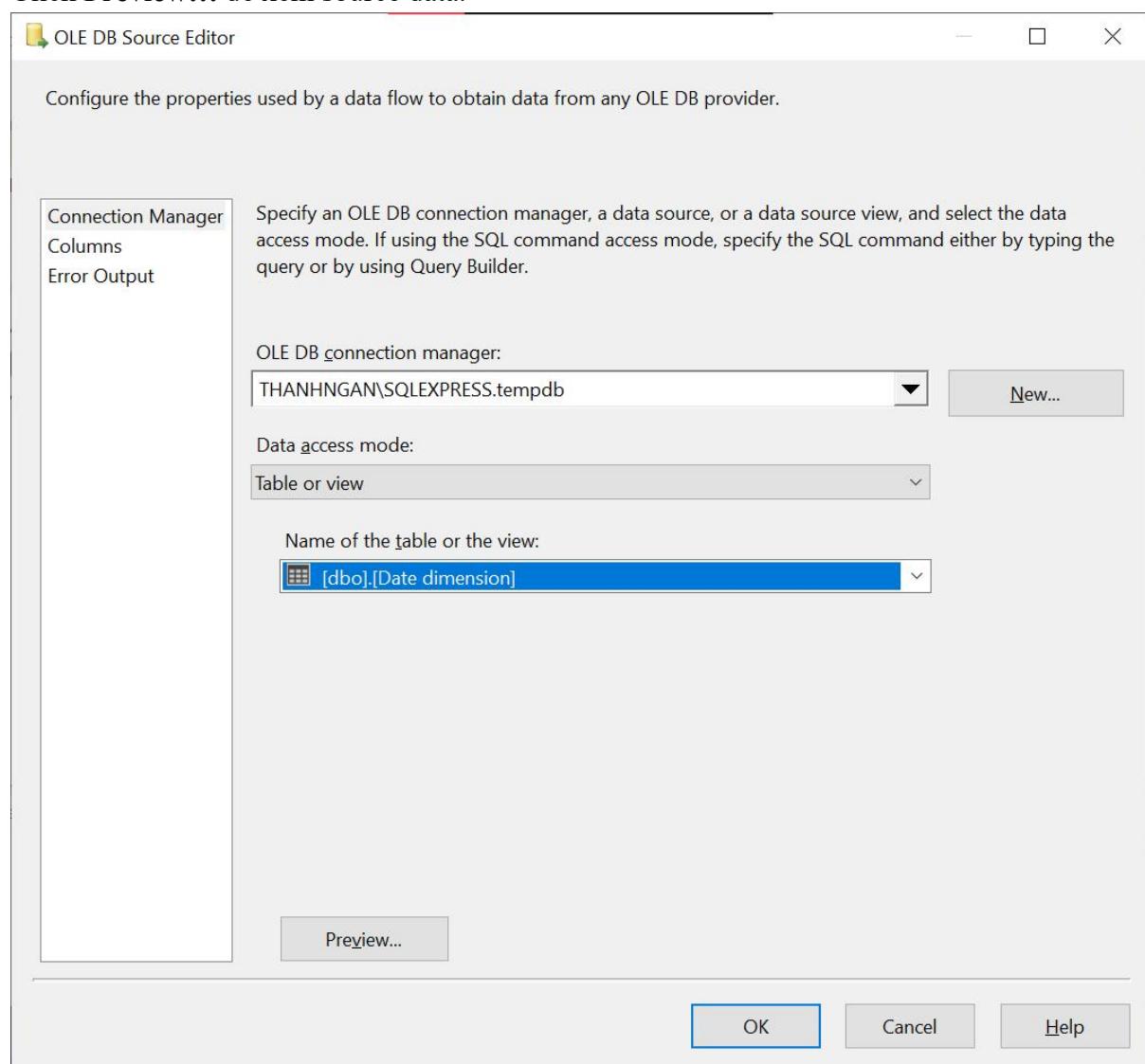
3. OLE DB Source sẽ xuất hiện như sau:



4. Đổi tên nó thành SRC - ExternalSourcesv_date_dimension

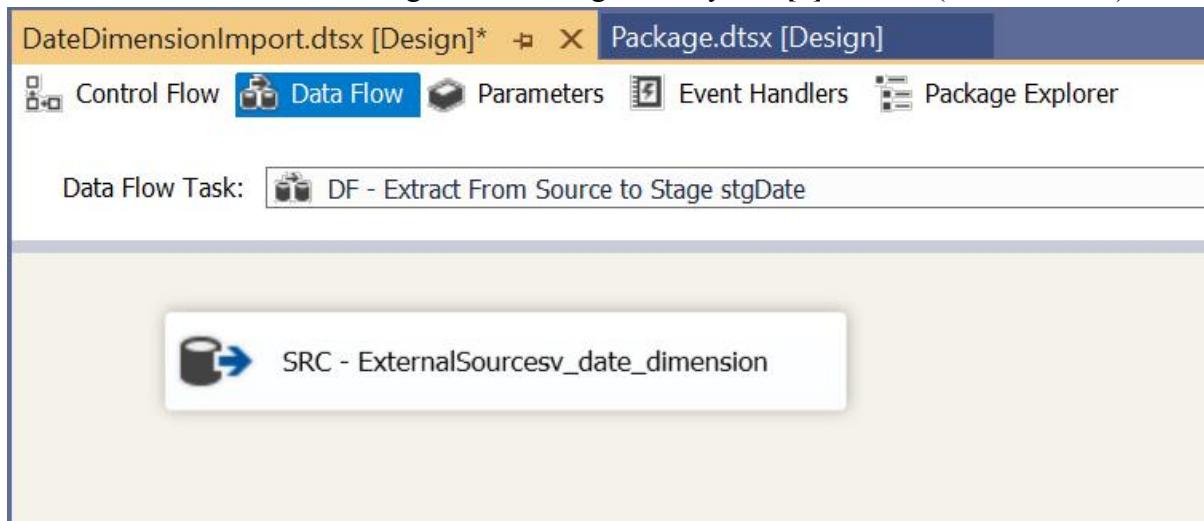


5. Double click vào source để cấu hình nó, **OLE DB Source Editor** sẽ mở ra. Từ dropdown **name of the table or view** chọn [**dbo].[Date Dimension**] là source. Click **Preview...** để xem source data.

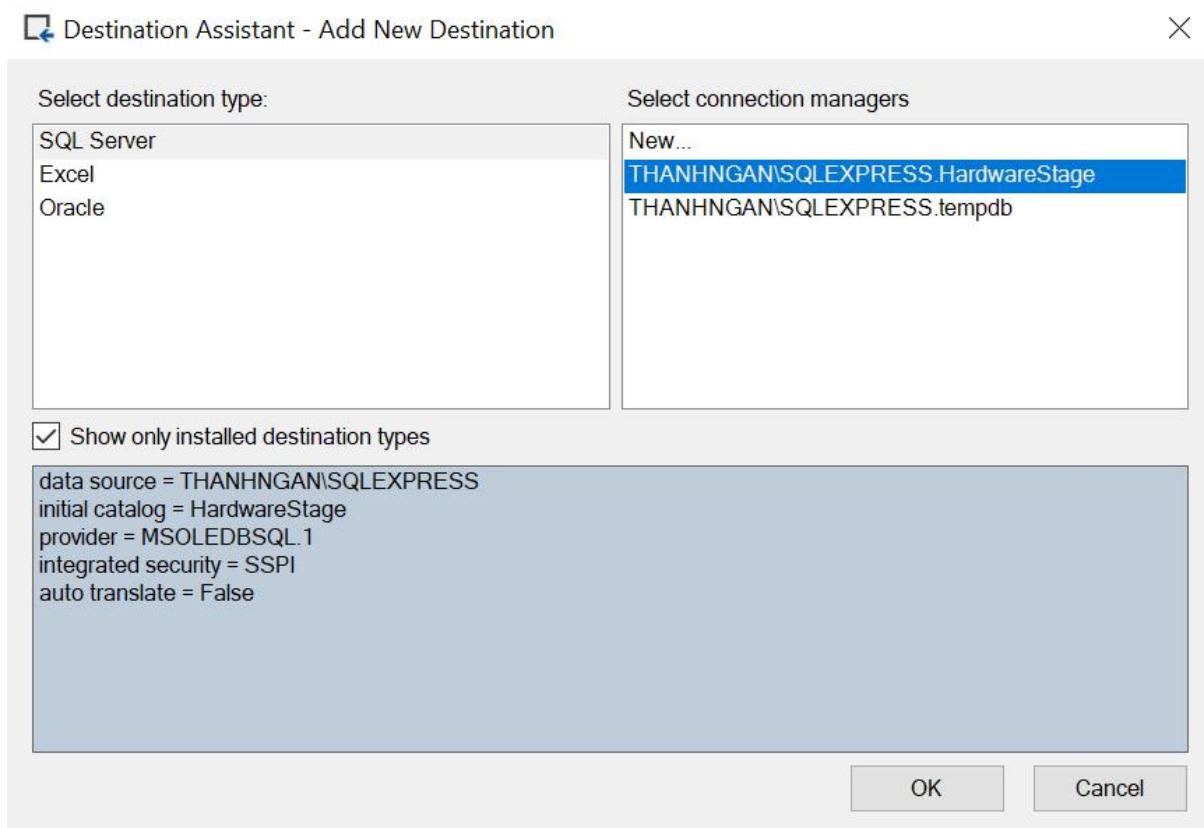


Click **OK** để lưu.

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

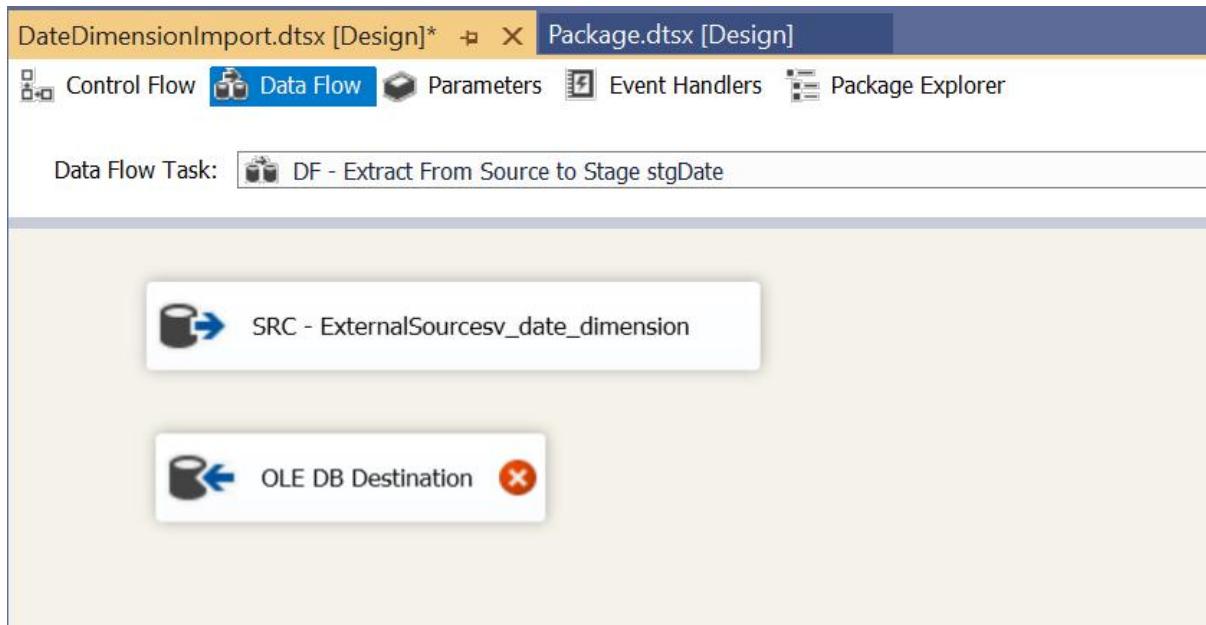


7. Tiếp tục, cấu hình destination. Từ **SSIS Toolbox**, kéo rê **Destination Assistant** vào design surface, một dialog xuất hiện. Chọn source Type: **SQL Server** và **HardwareStage** là connection manager.

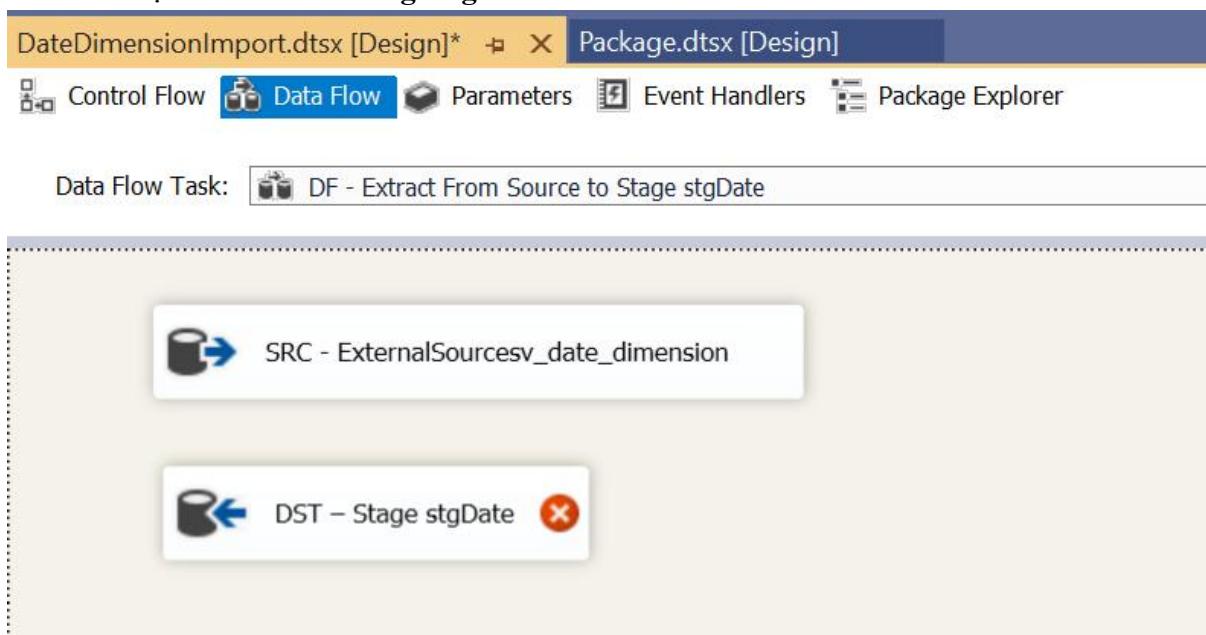


Click **OK**.

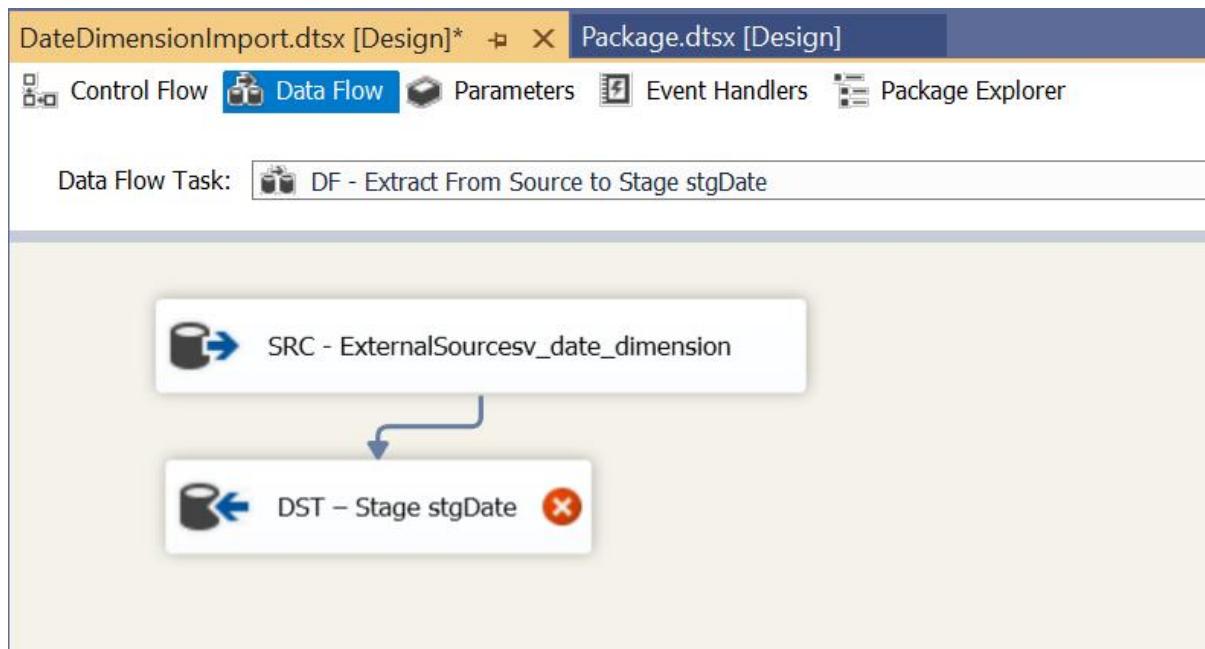
8. Ta có một **destination** như bên dưới:



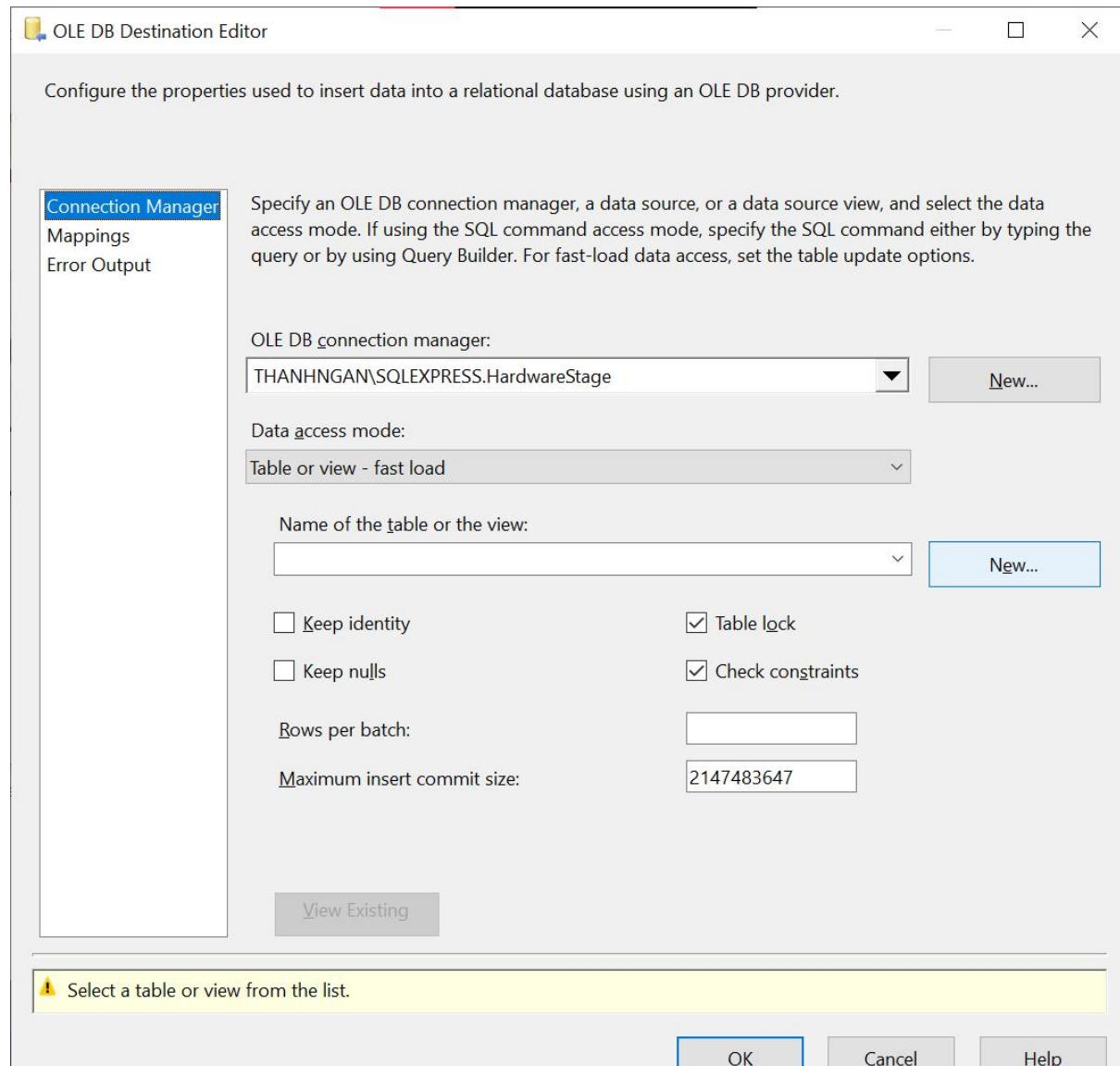
9. Đổi tên lại thành **DST – Stage stgDate**



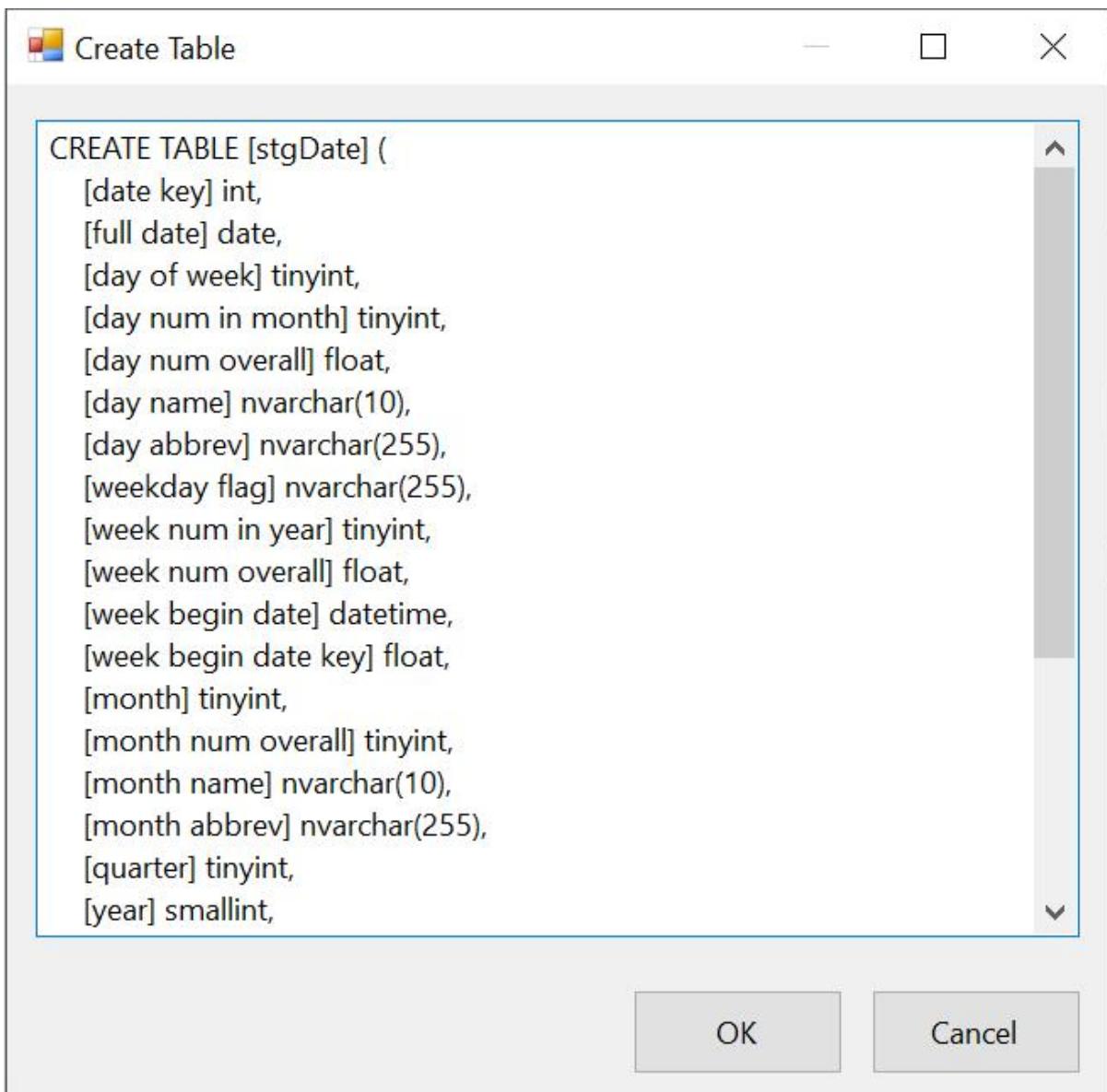
10. Click vào **source** và kéo rê nó vào **destination**.



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:

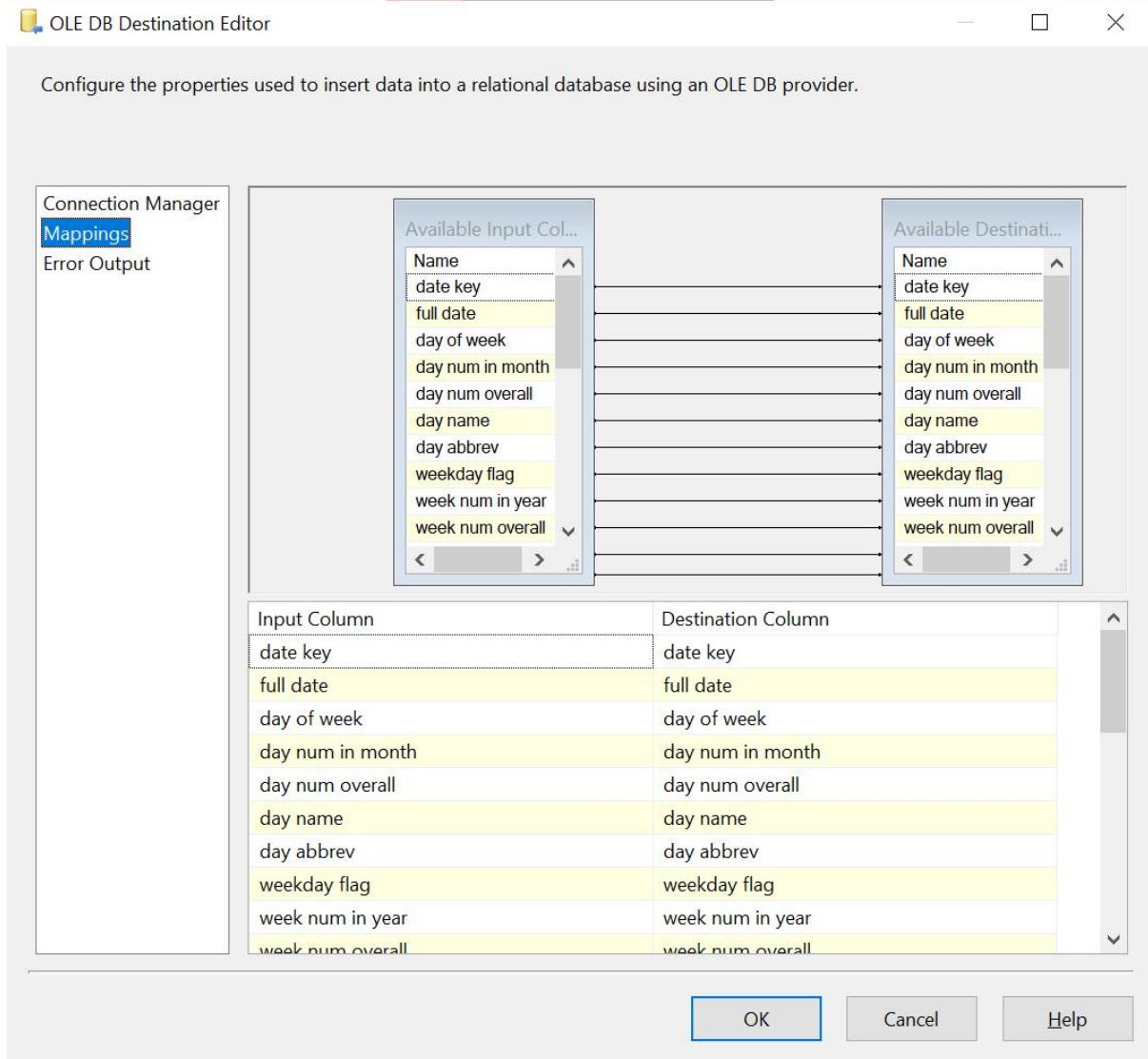


Đổi tên bảng lại thành **stgDate**. (**CREATE TABLE [stgDate]**).

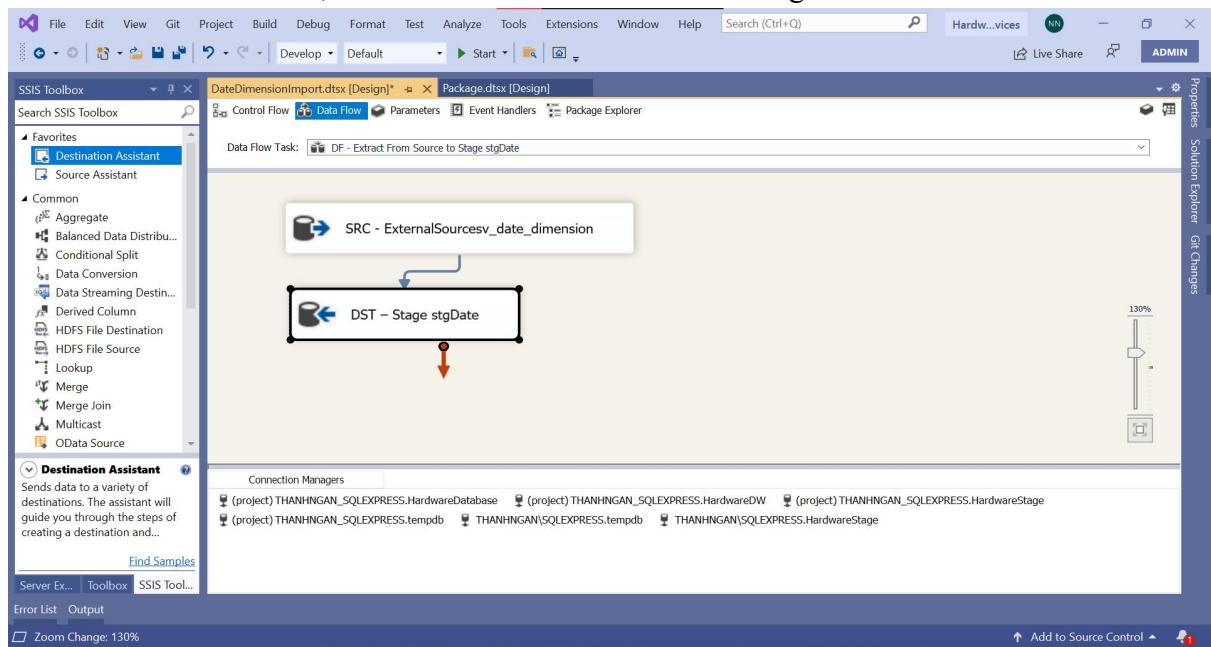
Click OK.

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:



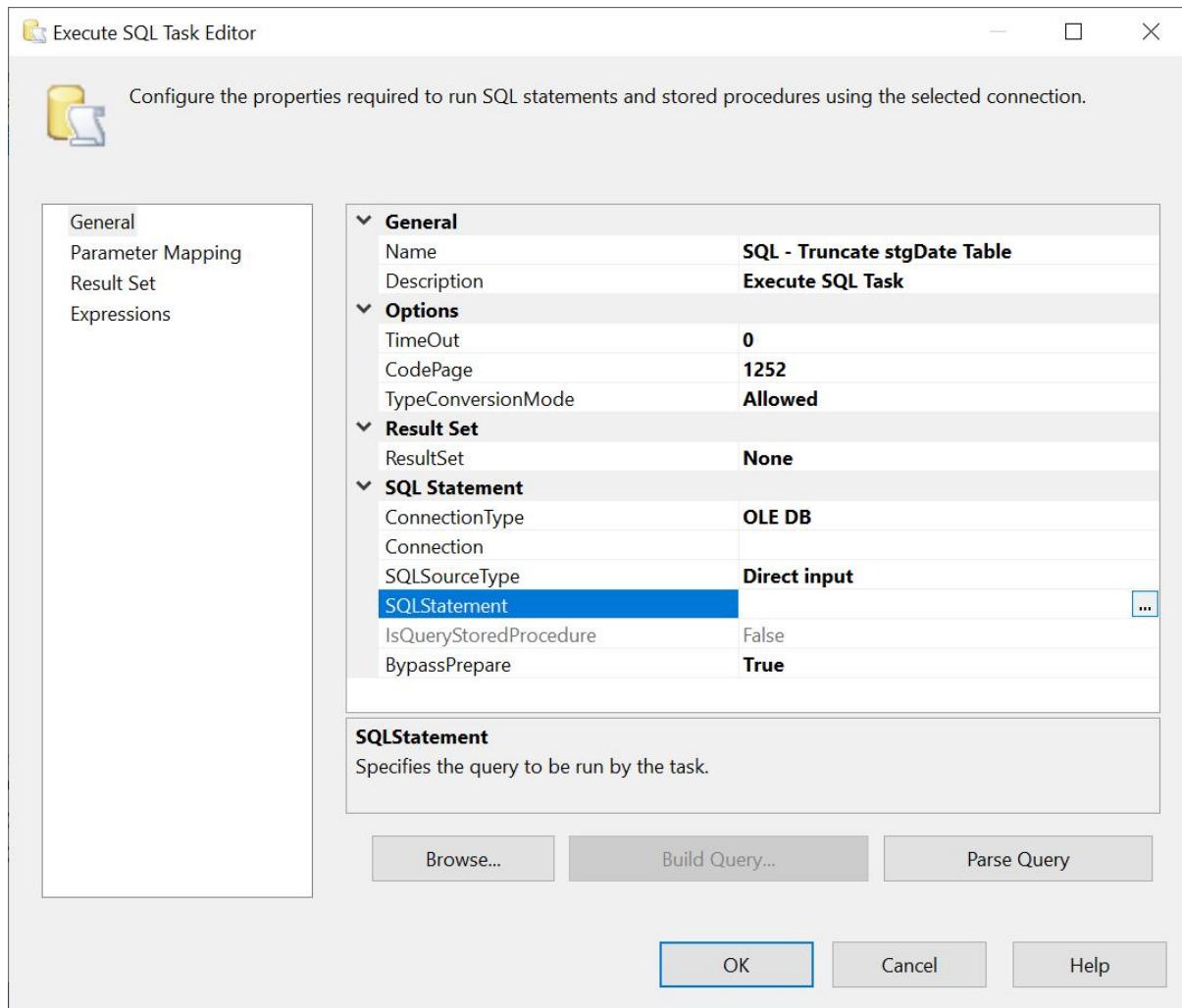
Click **OK** để hoàn tất việc cấu hình. Màn hình data flow design surface sẽ như bên dưới.



3.3.1.4: Thiết lập truncate table SQL task

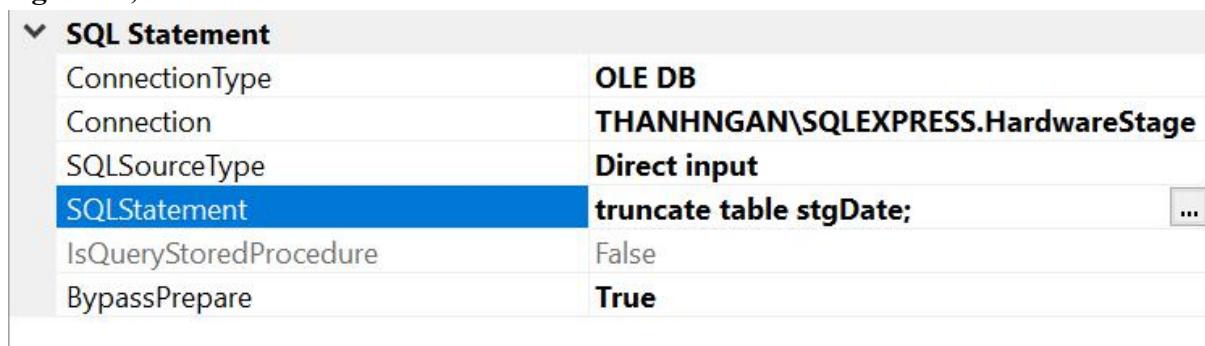
Click Control Flow tab.

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

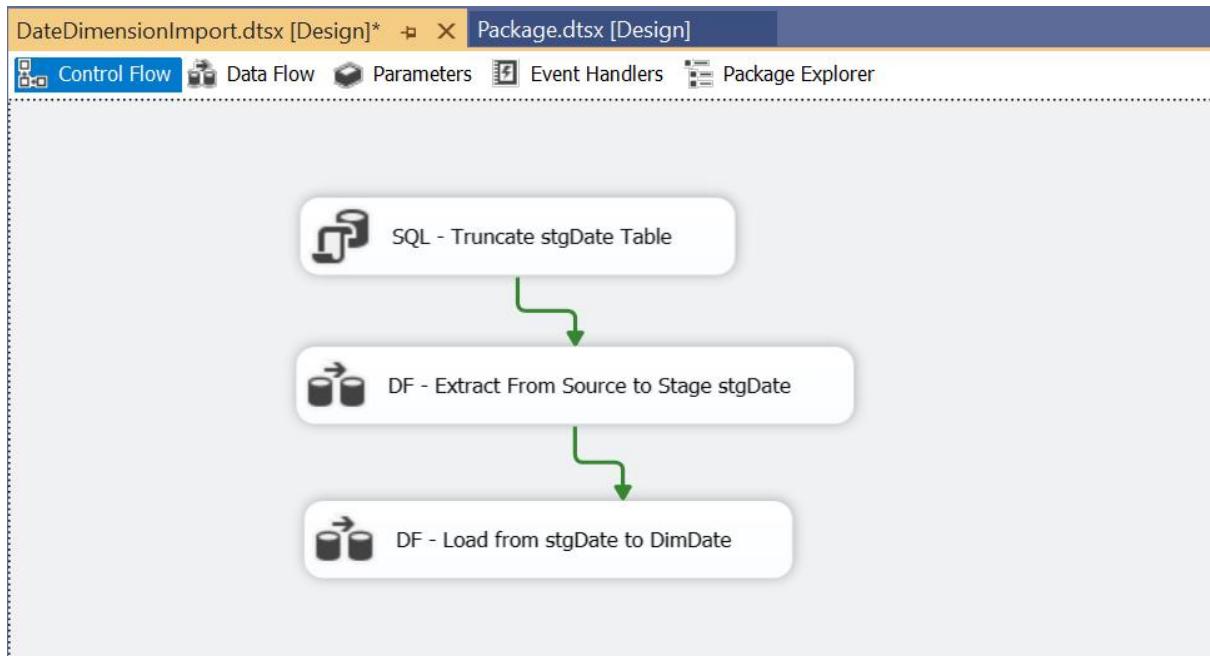


2. Ở mục SQL Statement, chọn Connection là

THANHNGAN\SQLEXPRESS.HardwareStage và SQLStatement là **truncate table stgDdate;**

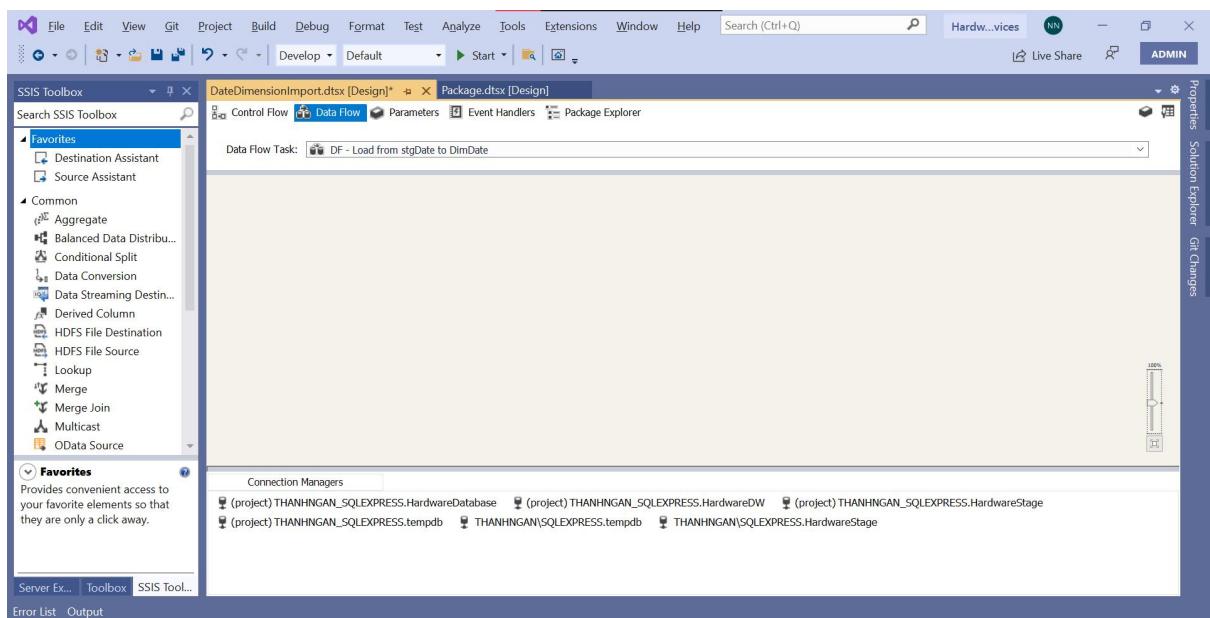


3. Click OK và trở lại màn hình Control Flow

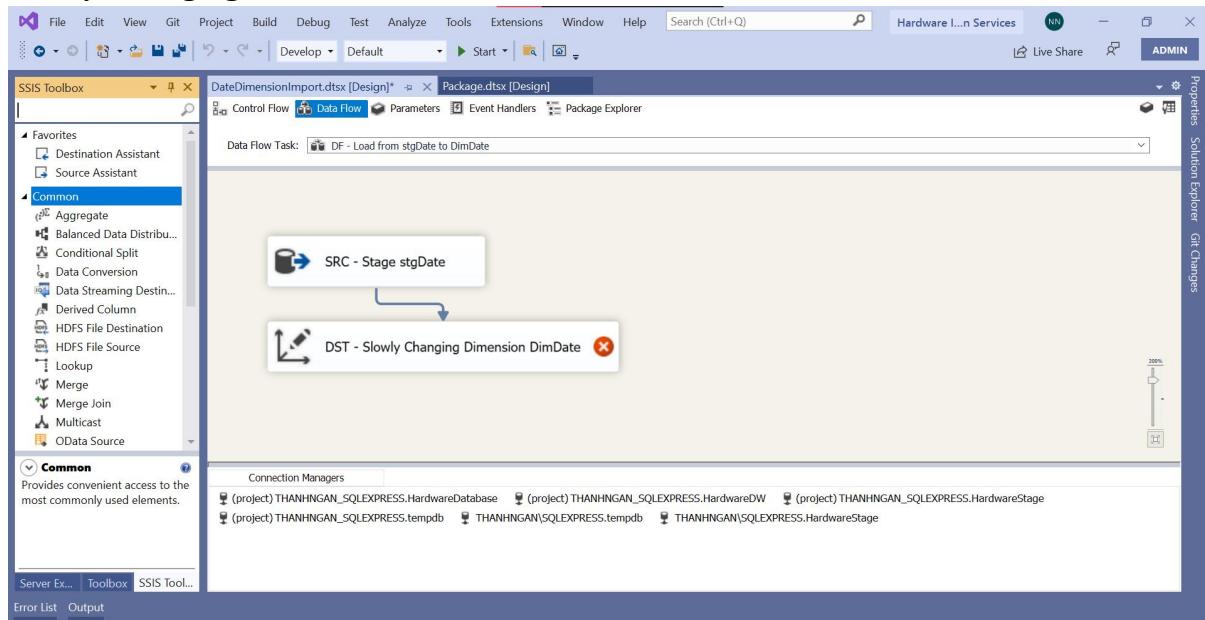


3.3.1.5: Load dữ liệu từ Stage vào Dimension

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

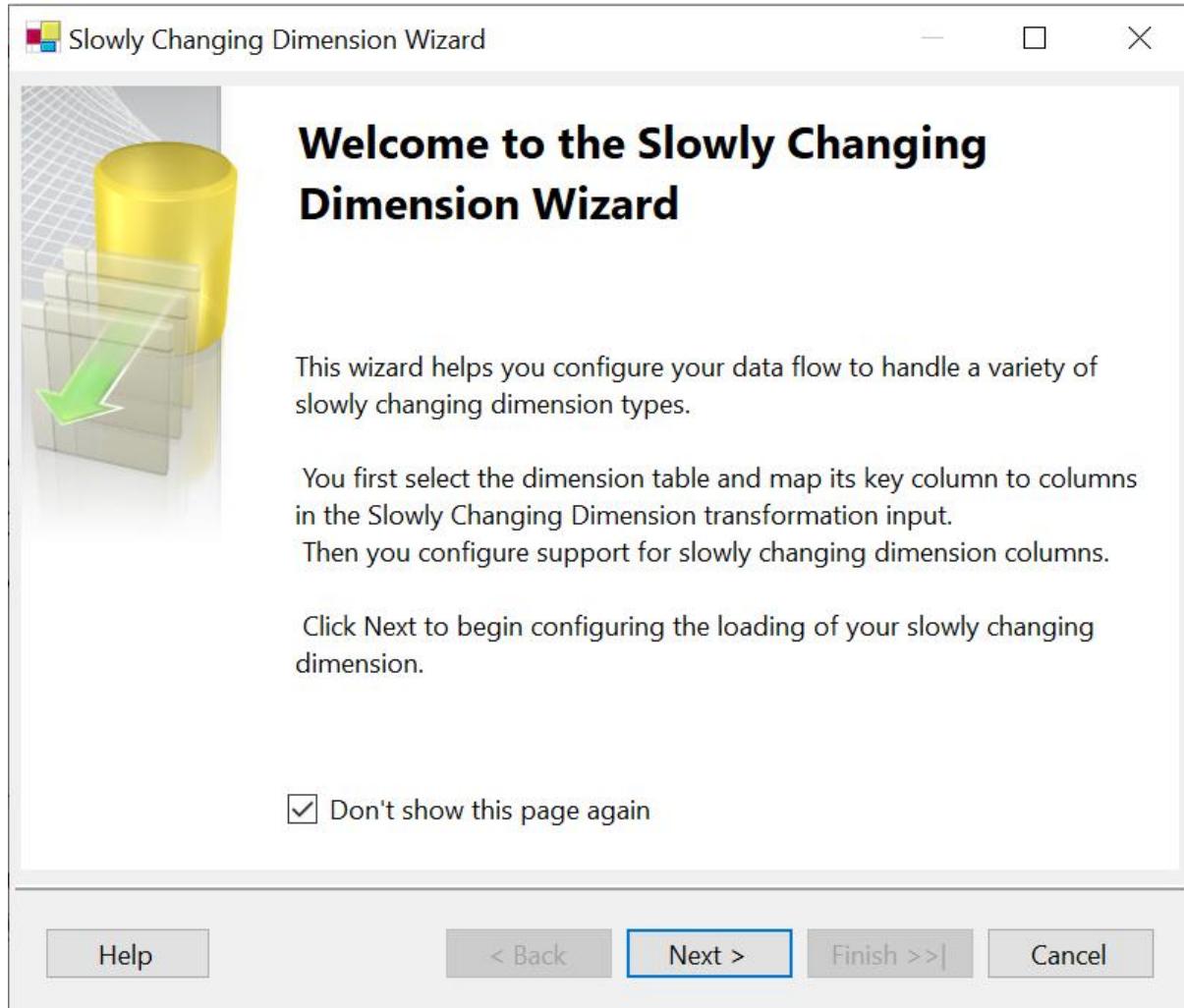


2. Kéo rê **Source Assistant** vào design surface, chọn **source type** là **SQL Server** và connection manager là **HardwareStage**.
3. Đổi tên source thành **SRC - Stage stgDate**, và double-click vào nó để cấu hình.
4. Từ **OLE DB Source Editor** chọn **[dbo].[stgDate]** table ở mục **name of the 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**.



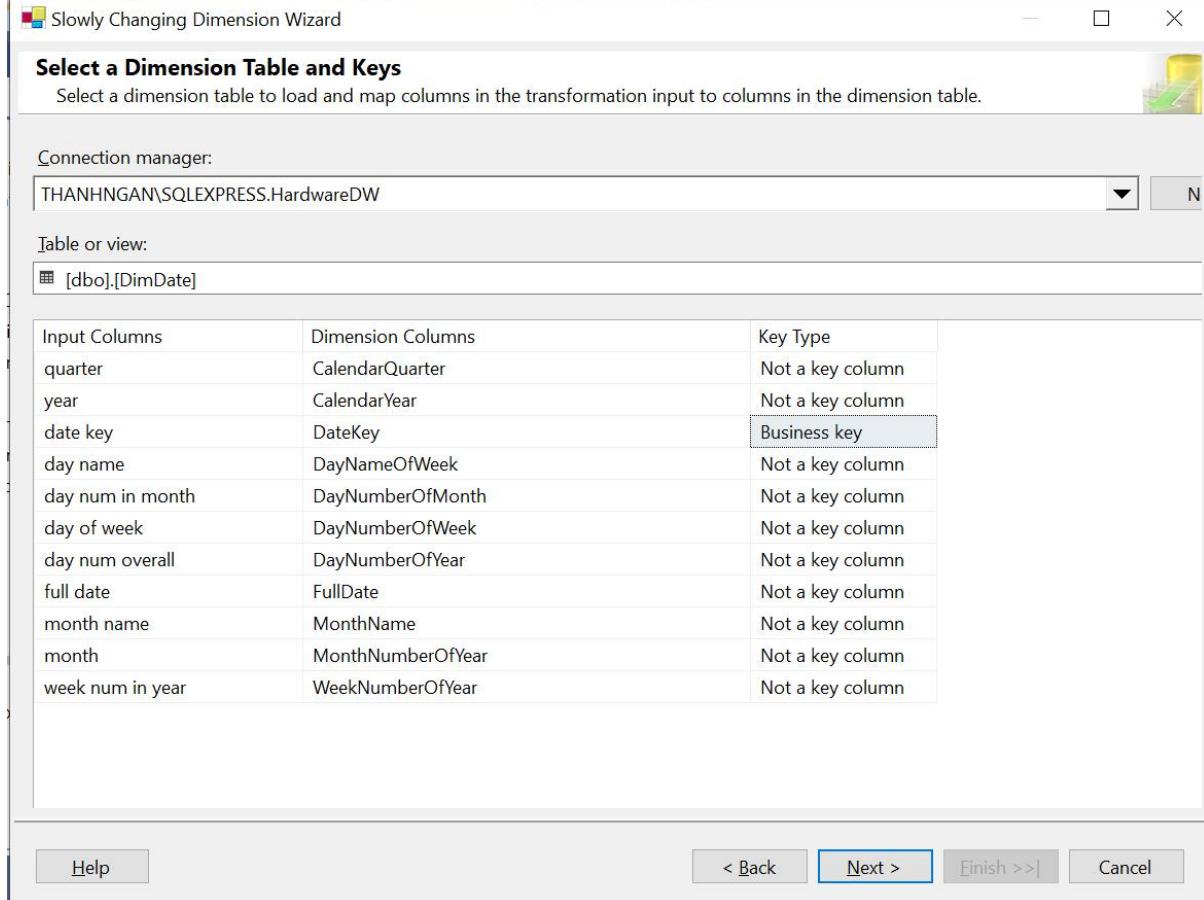
6. Double click vào **DST - Slowly Changing Dimension DimDate** để cấu hình.

7. Slowly Changing Dimension Wizard xuất hiện



Click **Next**.

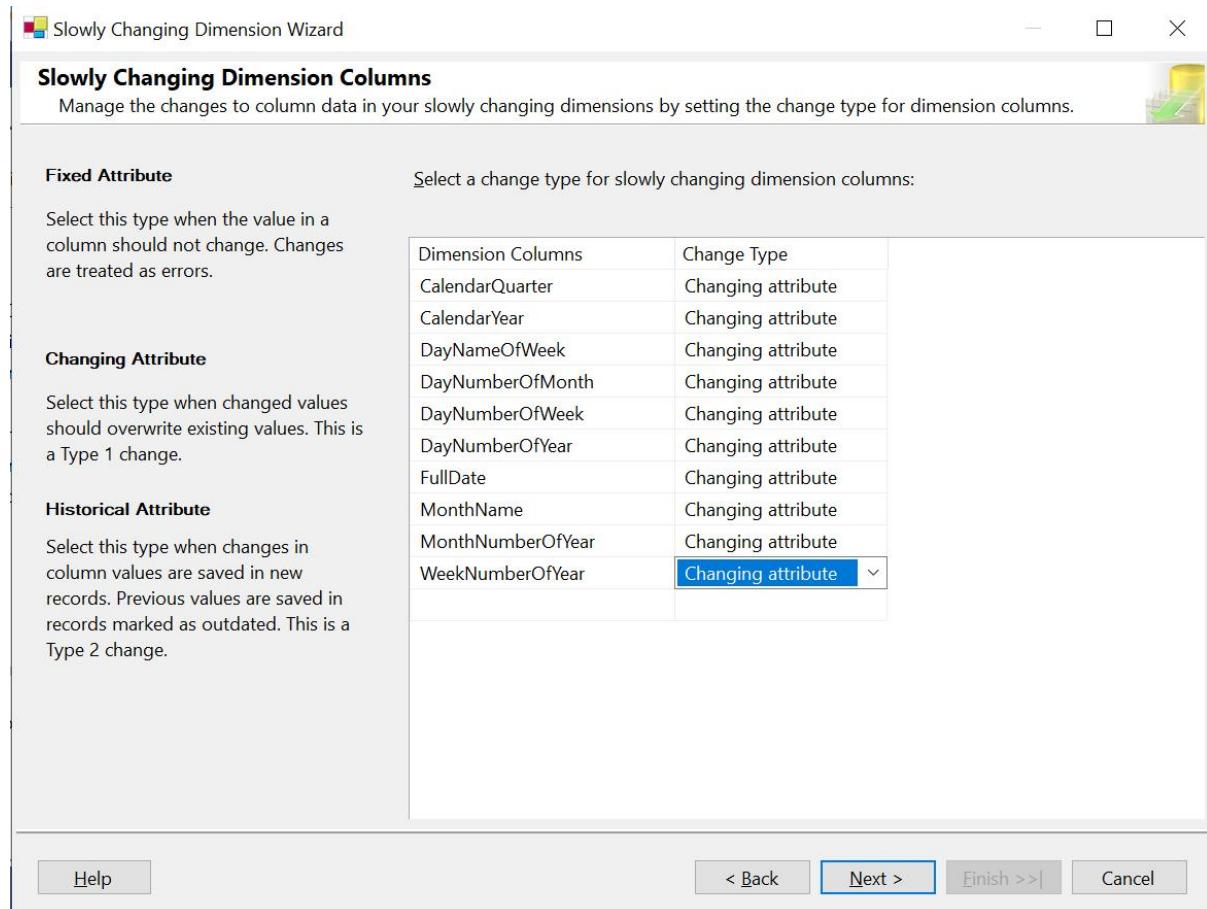
8. Chọn bảng **DimDate** ở CSDL **HardwareDW**. Ở mục **connection manager** chọn CSDL **HardwareDW**.
9. Ở mục **Table or view** chọn bảng **[DimDate]**.
10. Chọn **DateKey** là **business key**. Tất cả các cột khác là “**Not a key column**”.



Click **Next**.

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

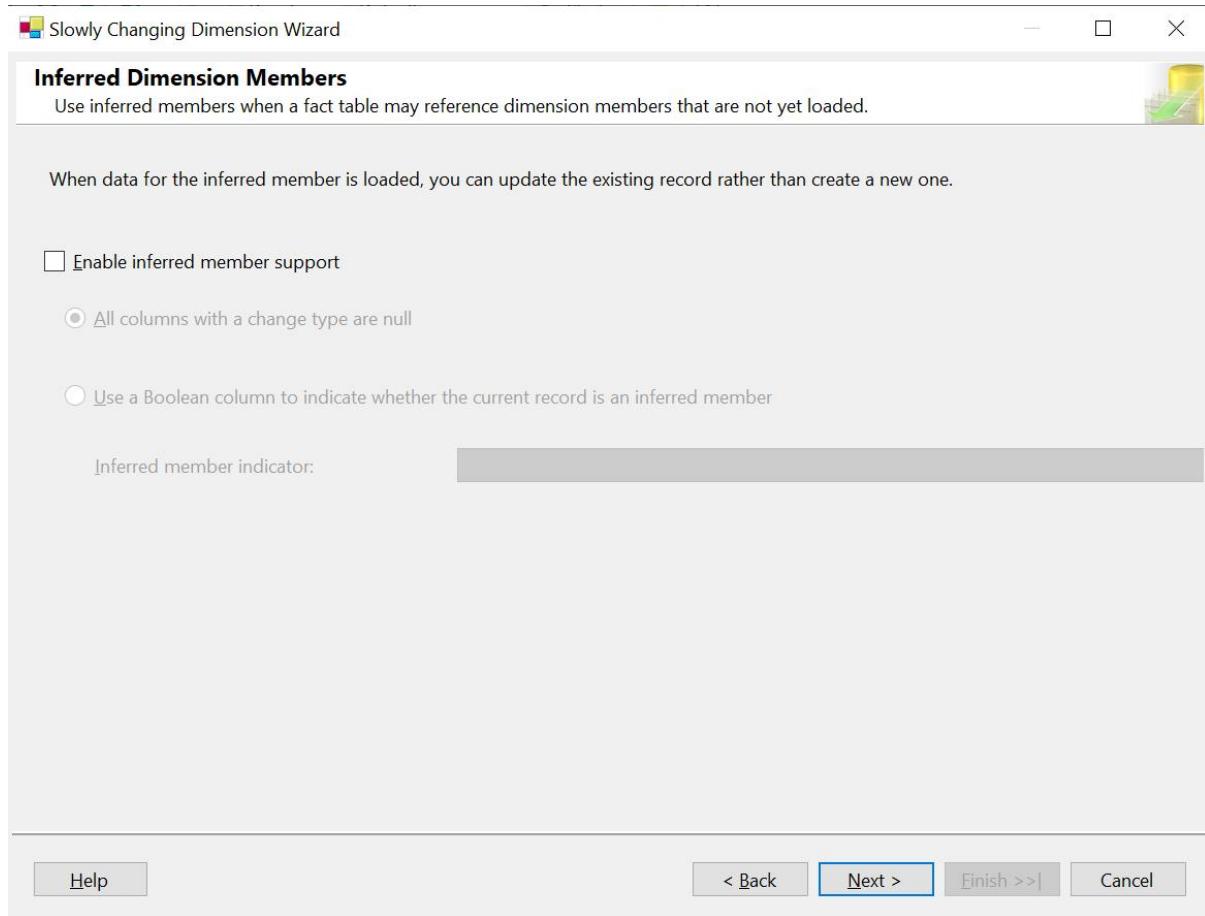
- **Fixed** → Không thay đổi
- **Changing** → Type 1 (Update)
- **Historical** → Type 2 (thêm dòng mới, dòng cũ inactive)



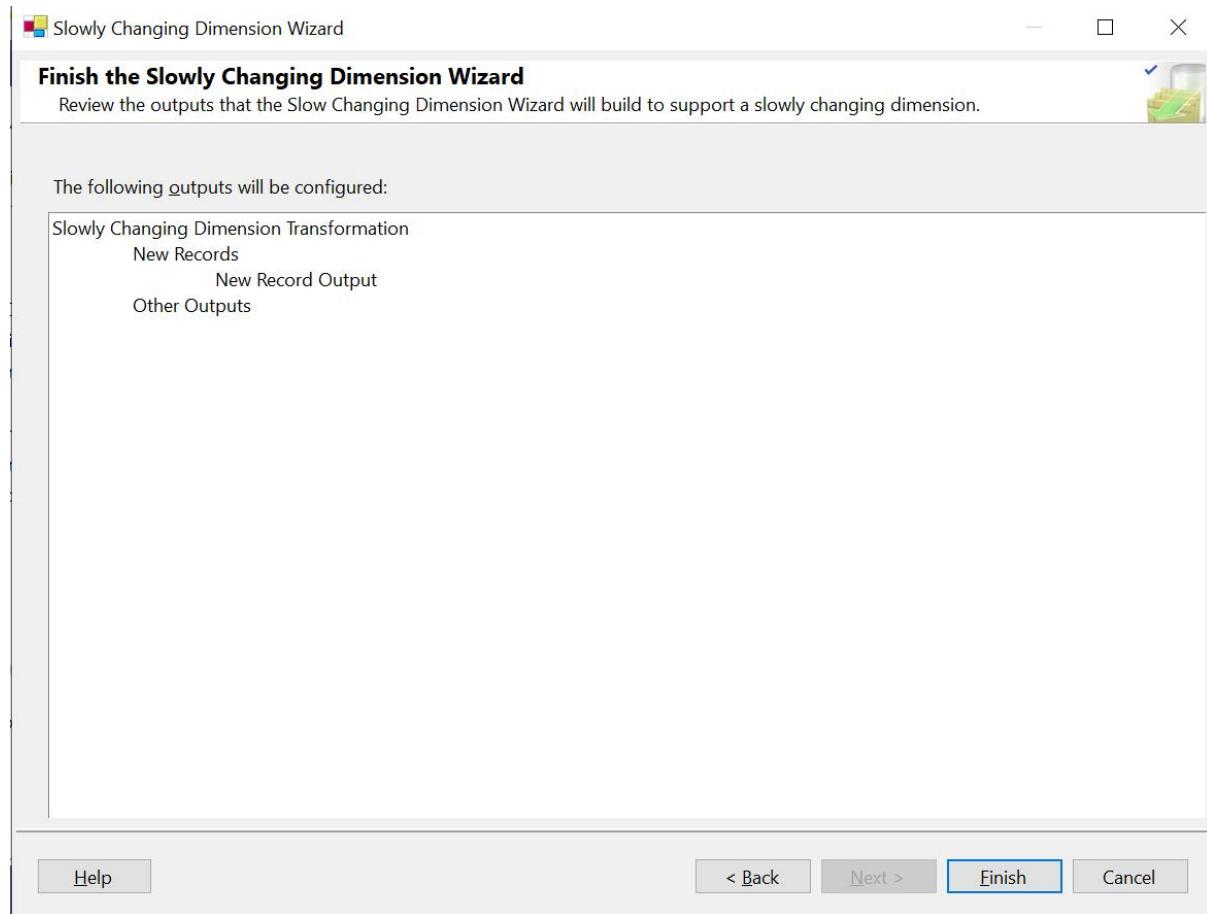
Cấu hình tất cả non-business key là changing attribute.

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

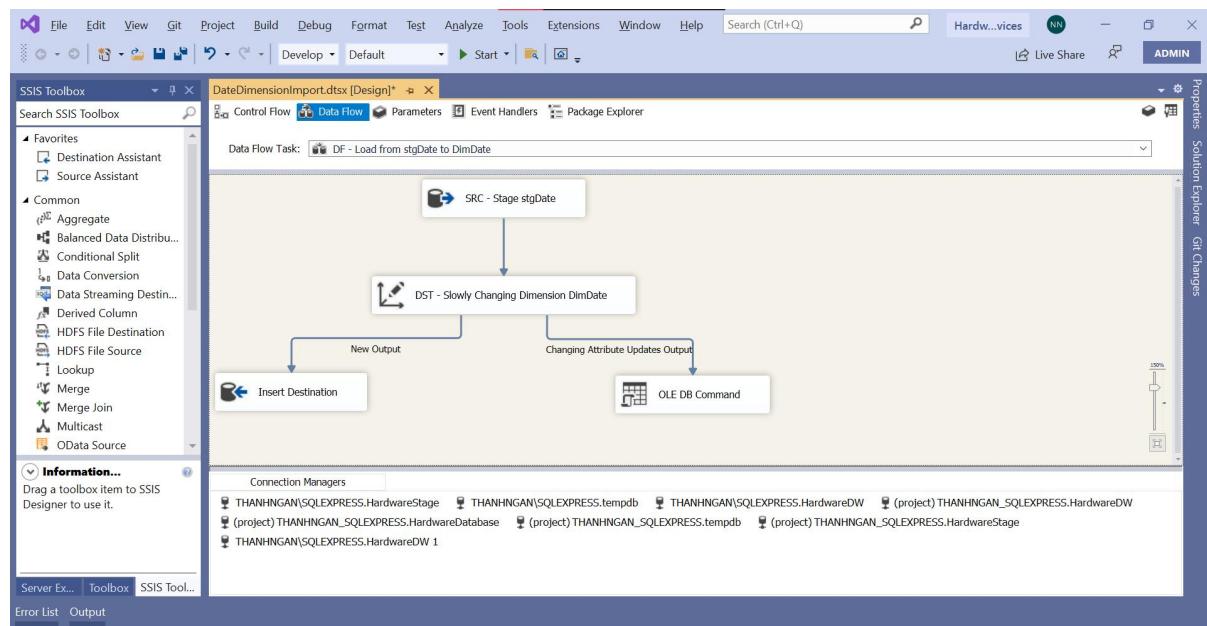
13. Xuất hiện màn hình **Inferred Dimension Members**. Bỏ chọn **check box** và click **Next**.



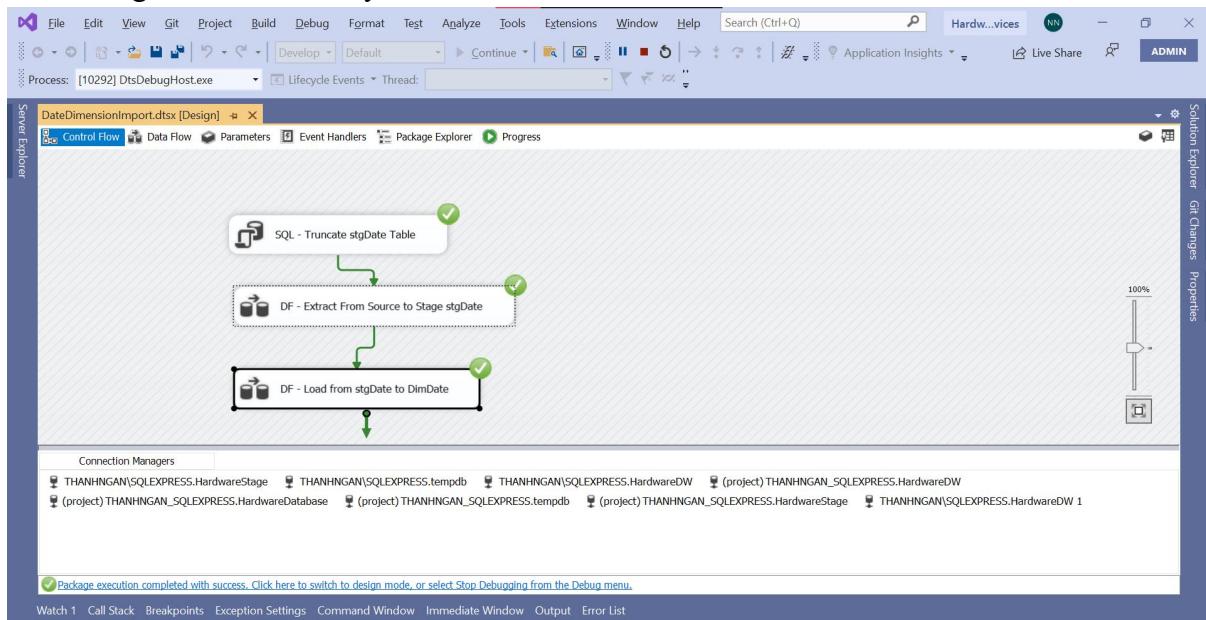
14. Xuất hiện màn hình Finish



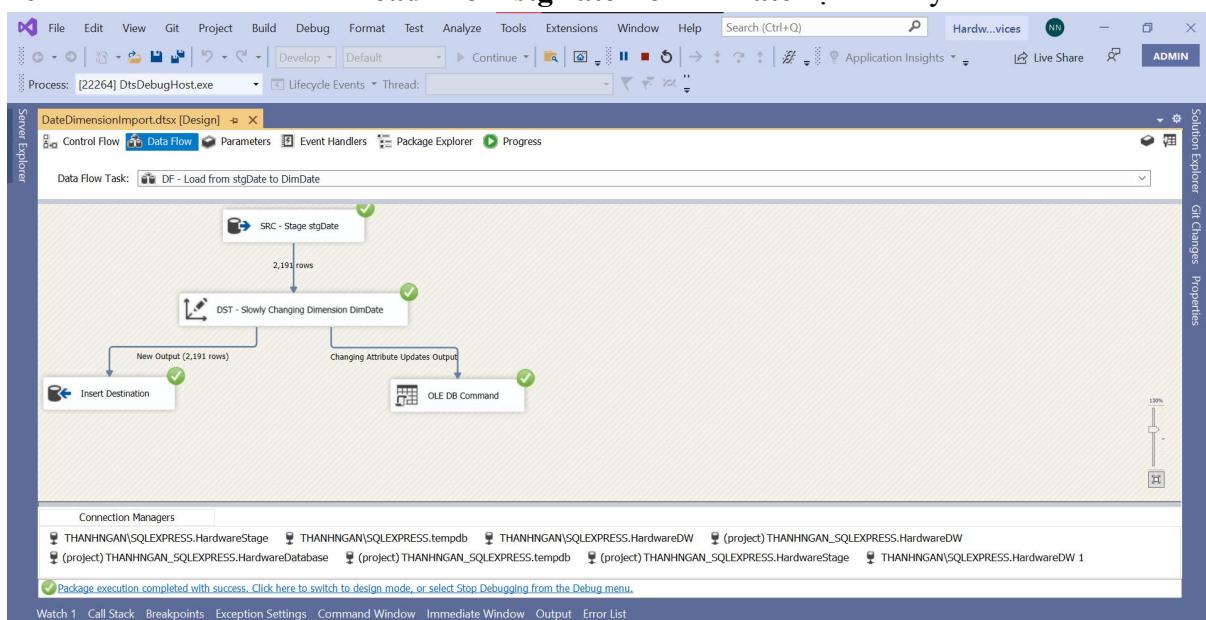
Click Finish.



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:

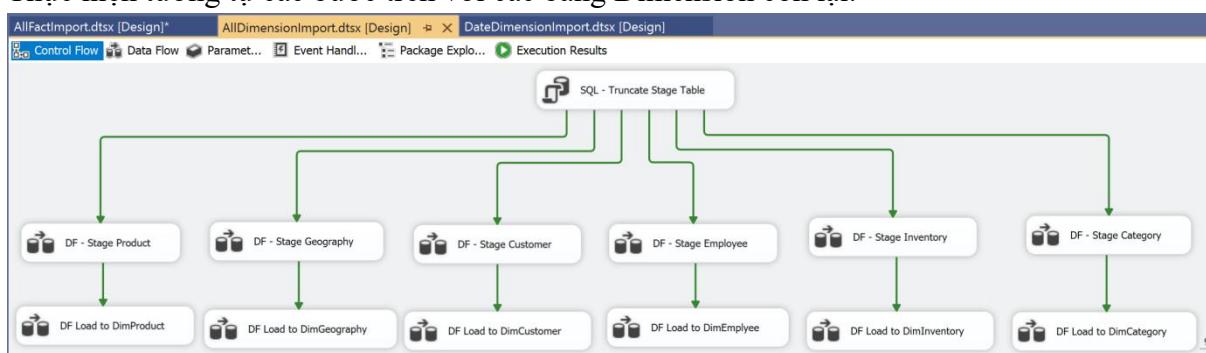


16. Double click vào DF - Load From stgDate To DimDate bạn sẽ thấy màn hình sau:



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

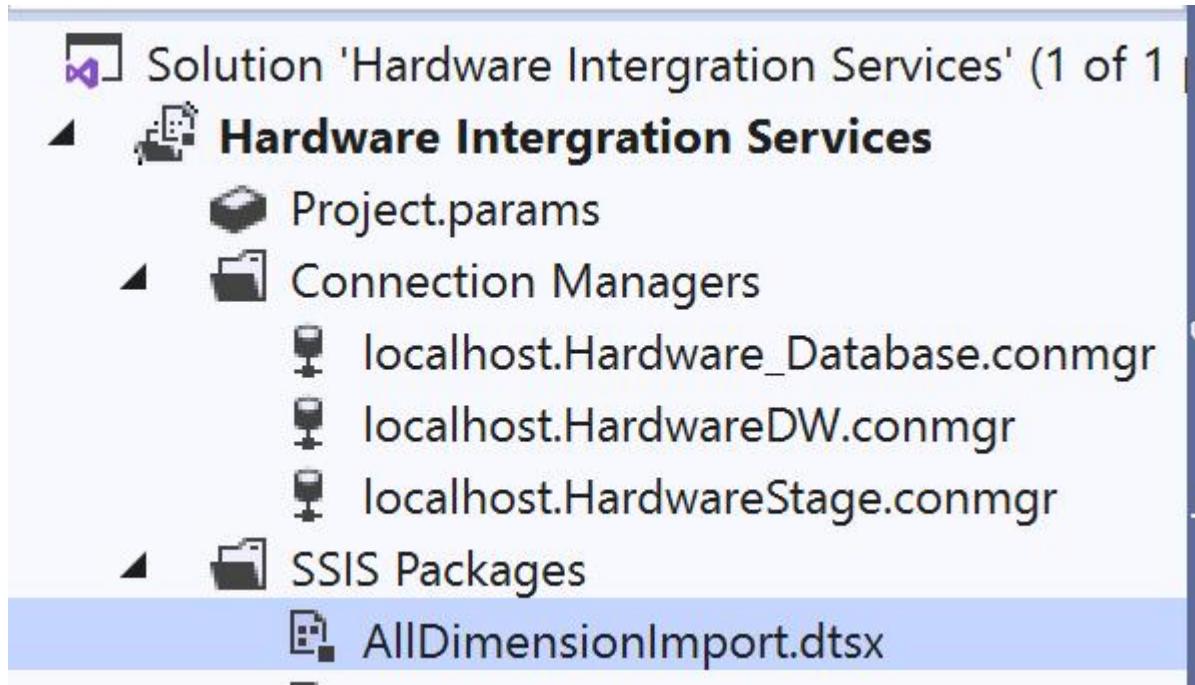
Thực hiện tương tự các bước trên với các bảng Dimension còn lại.



3.3.2. Product Dimension

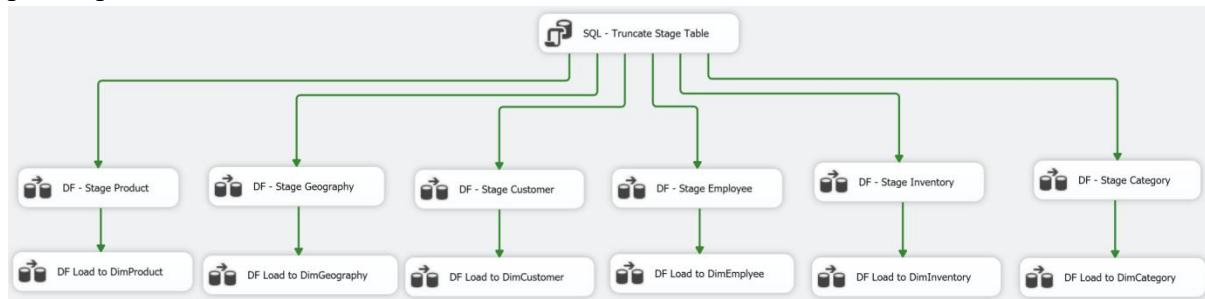
3.3.2.1. Tạo Package

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

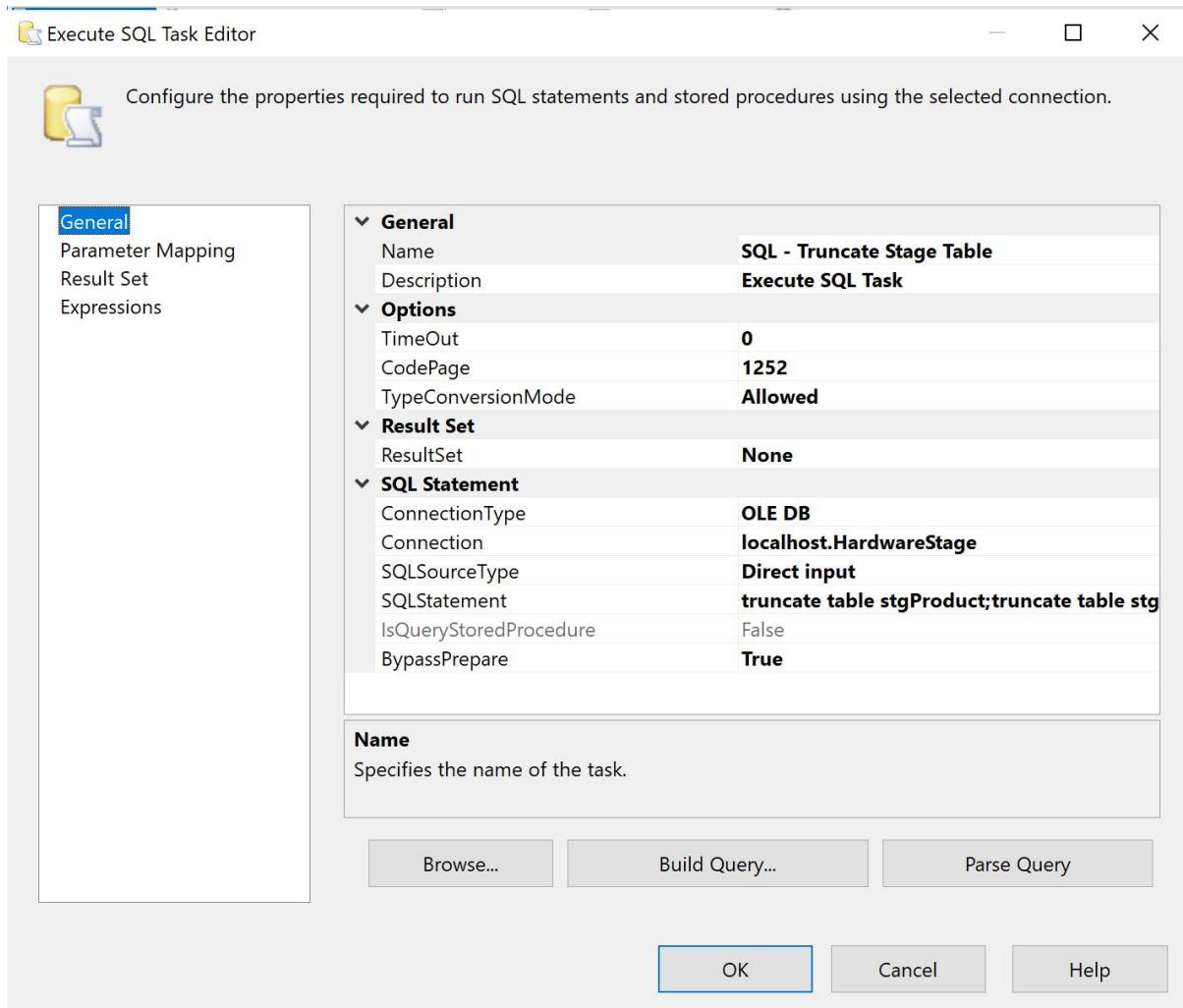


3.3.2.2. Thiết lập Control Flow

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

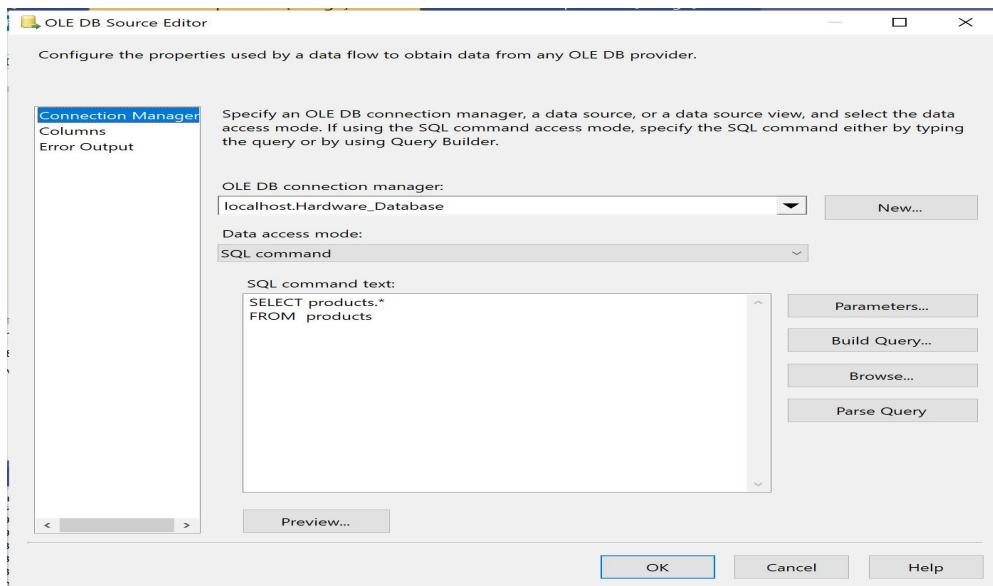


- Thiết lập truncate table SQL task



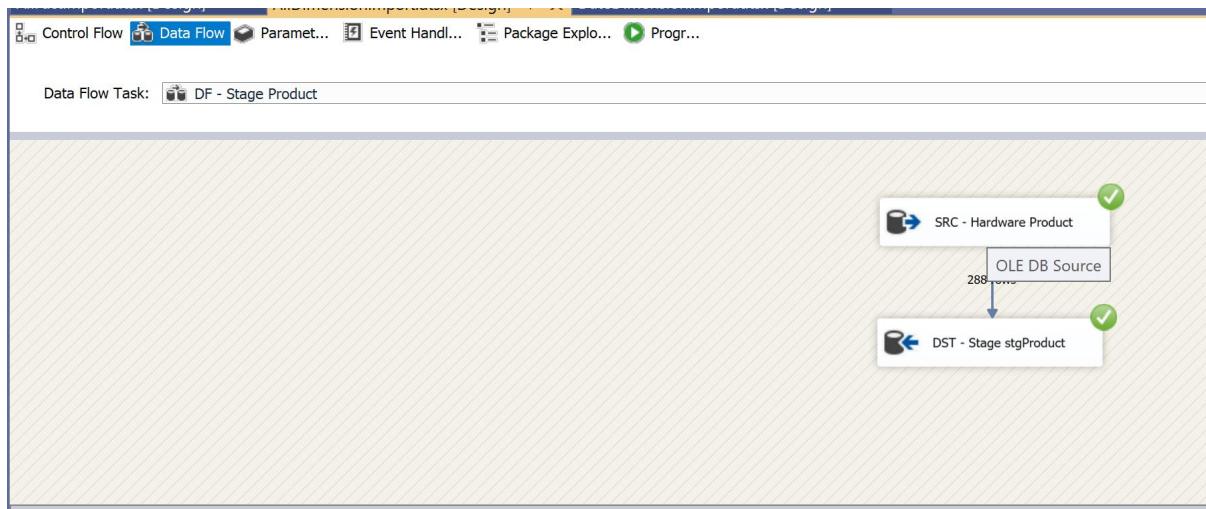
3.3.2.3. Load dữ liệu từ Source vào Stage

- 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:



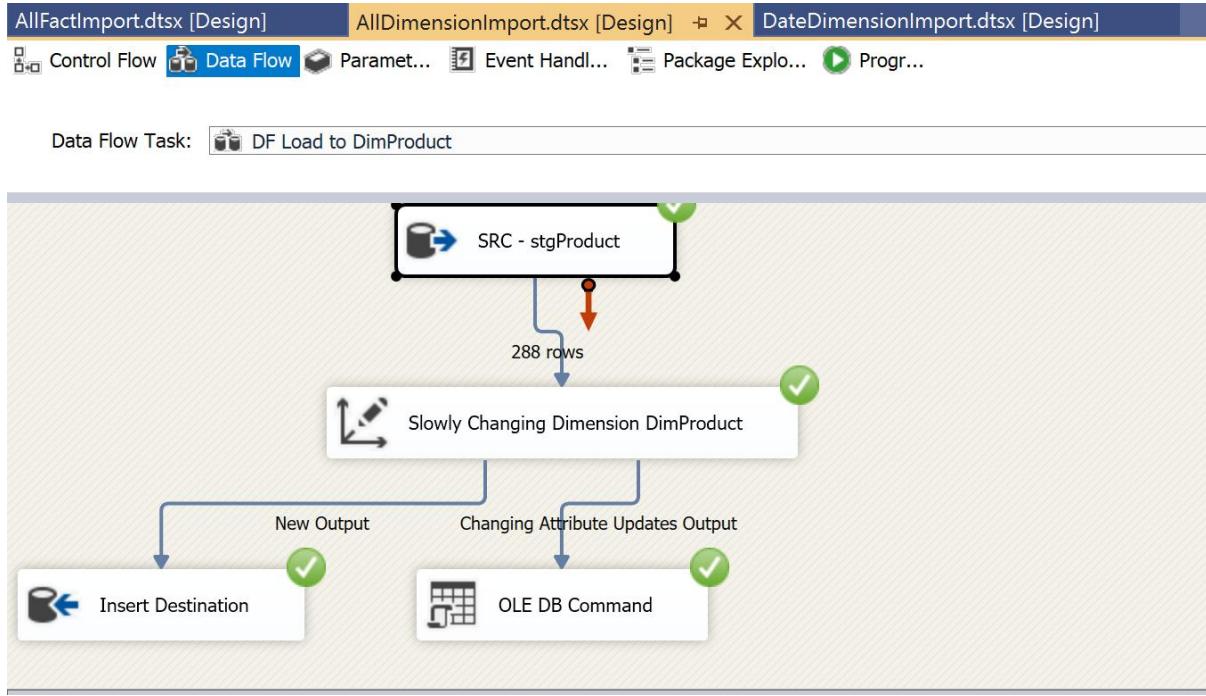
+ Destination Assistant:

+ Sau khi cấu hình xong ta được:



3.3.2.4. Load dữ liệu từ Stage vào Dimension

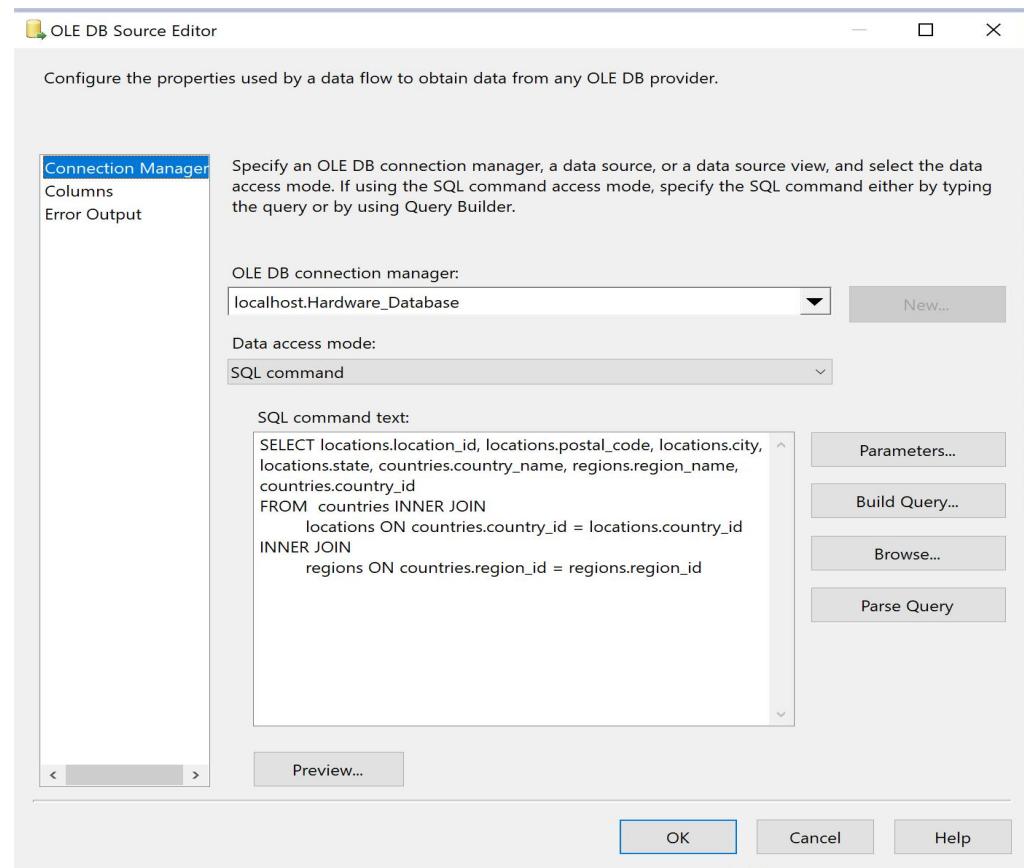
- Double-click vào **DF - Load to DimProduct** task để mở data flow design Surface và thực hiện các bước cấu hình tương tự mục 3.3.



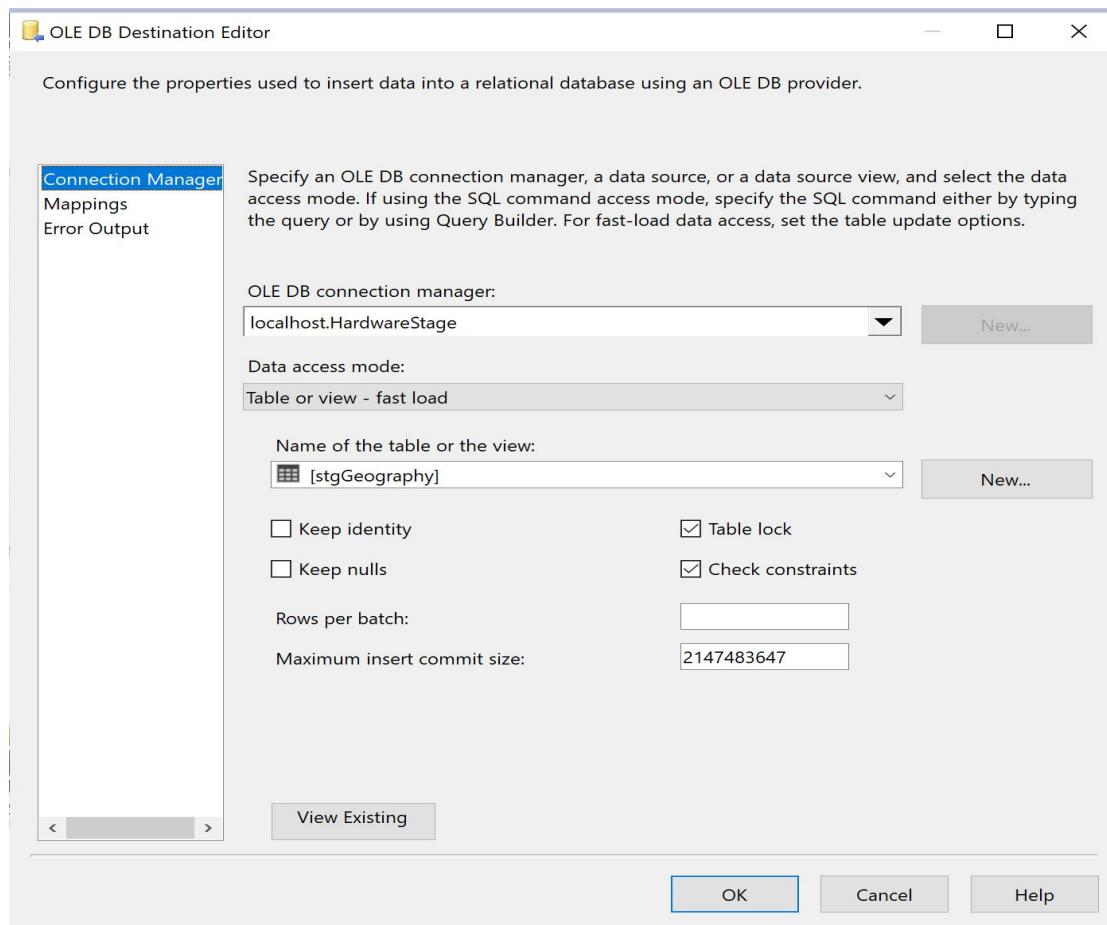
3.3.3. Geography Dimension

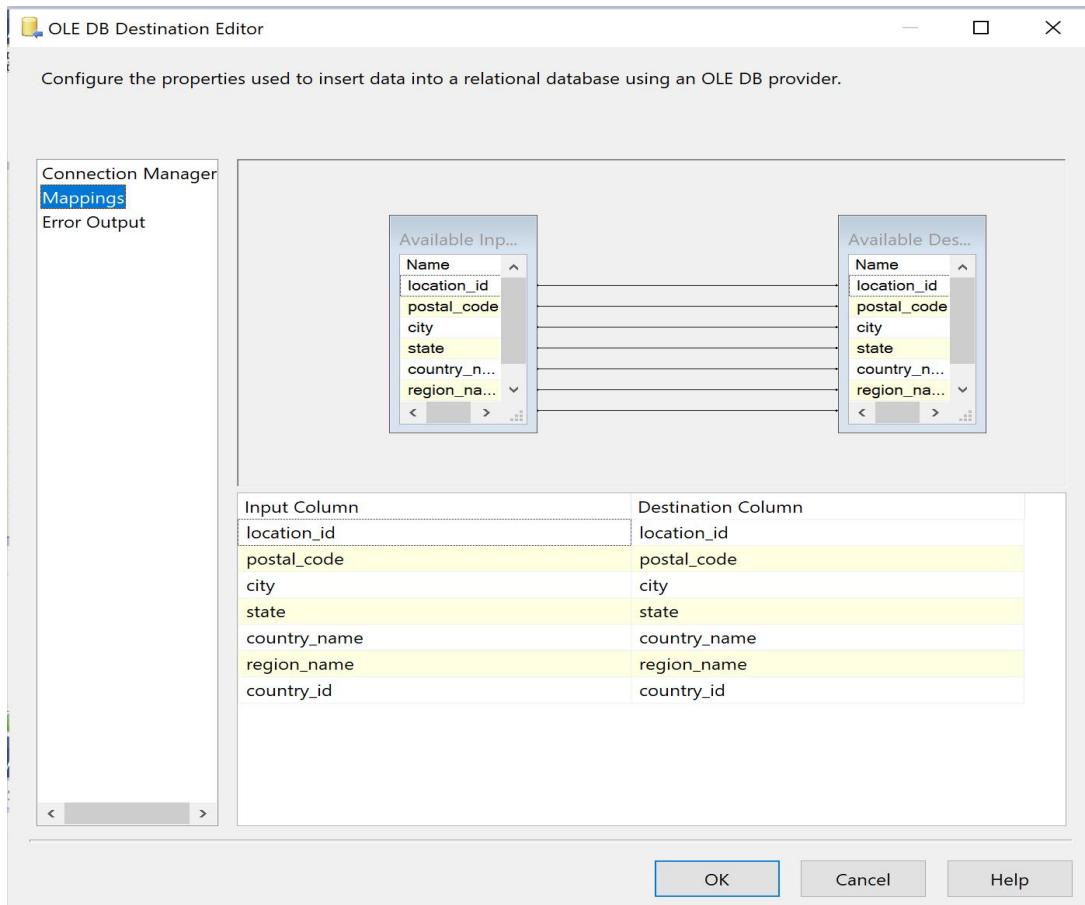
3.3.3.1. Load dữ liệu từ Source vào Stage

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

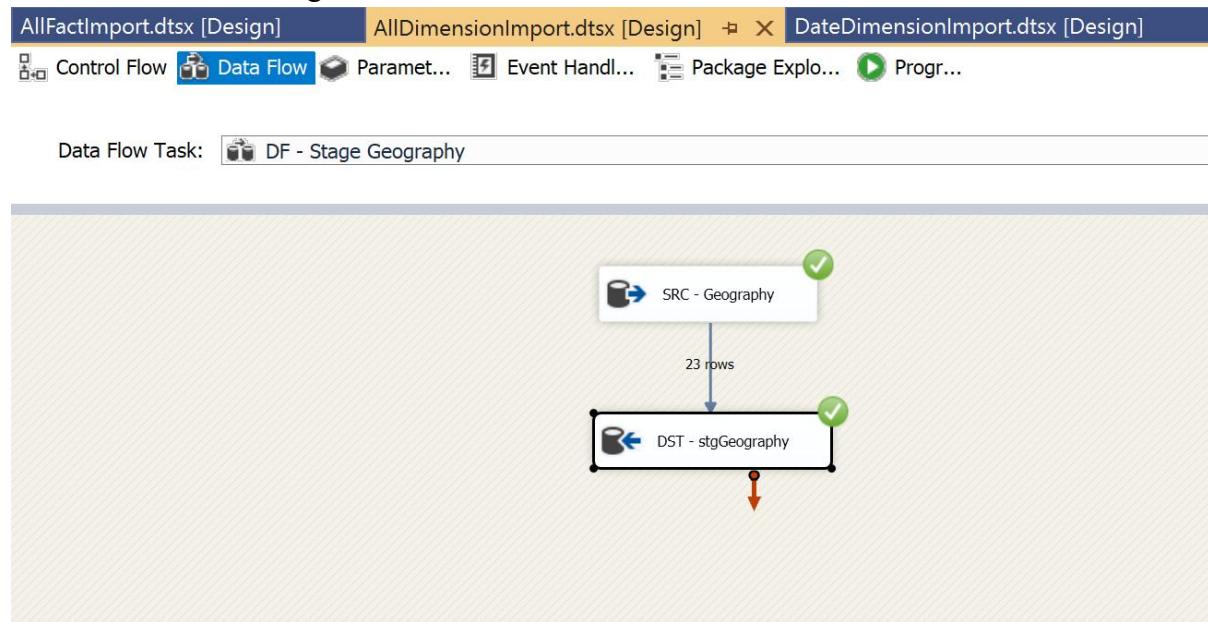


+ Destination Assistant:



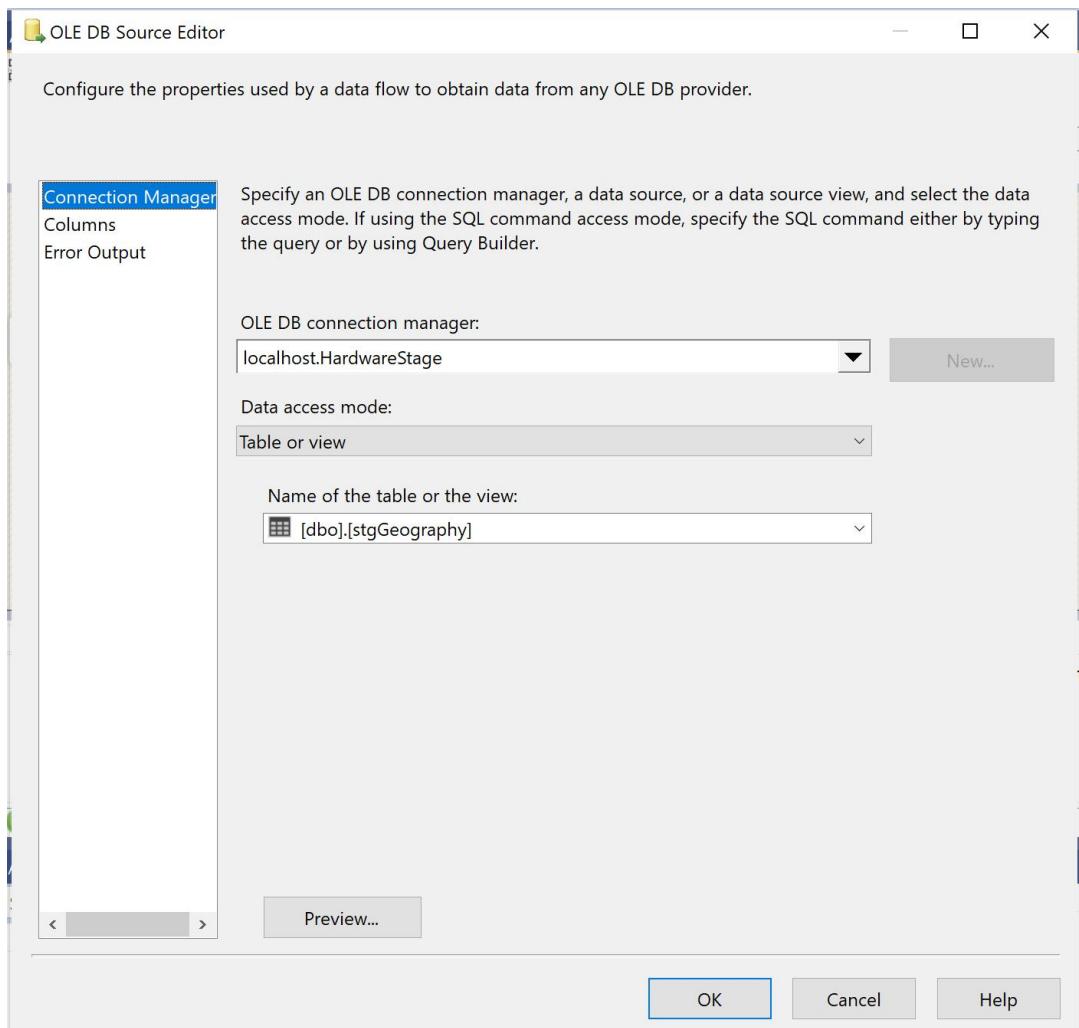


+ Sau khi cấu hình xong ta được:



3.3.3.2. Load dữ liệu từ Stage vào Dimension

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



+ Slowly Changing Dimension:

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.

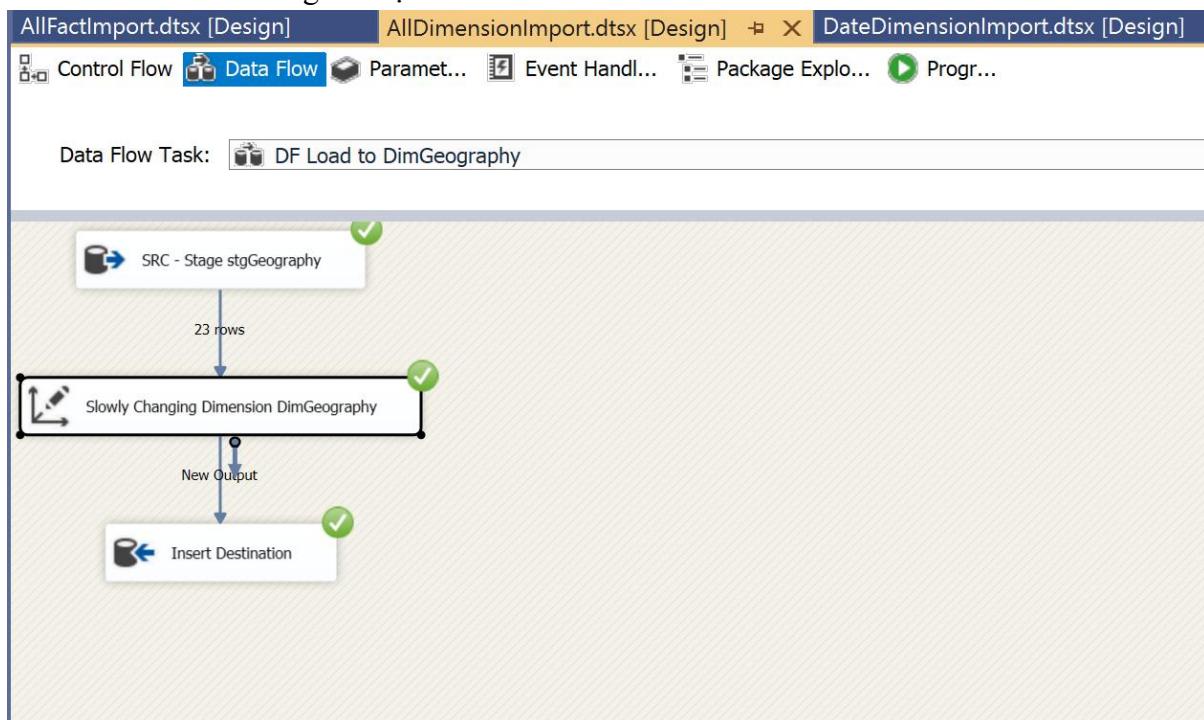
Historical Attribute

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
City	Fixed attrib...
CountryCode	Fixed attrib...
CountryName	Fixed attrib...
PostalCode	Fixed attrib...
Region	Fixed attrib...
State	Fixed attrib...

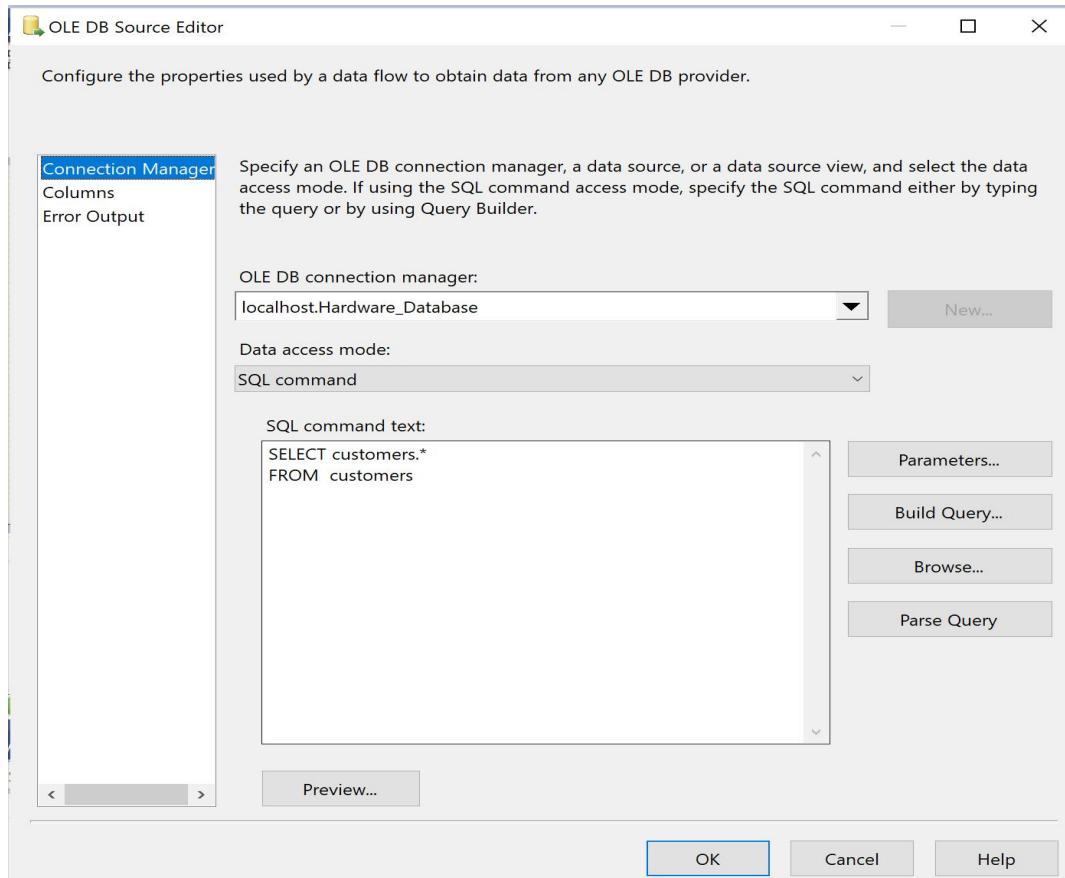
+ Sau khi cấu hình xong ta được:



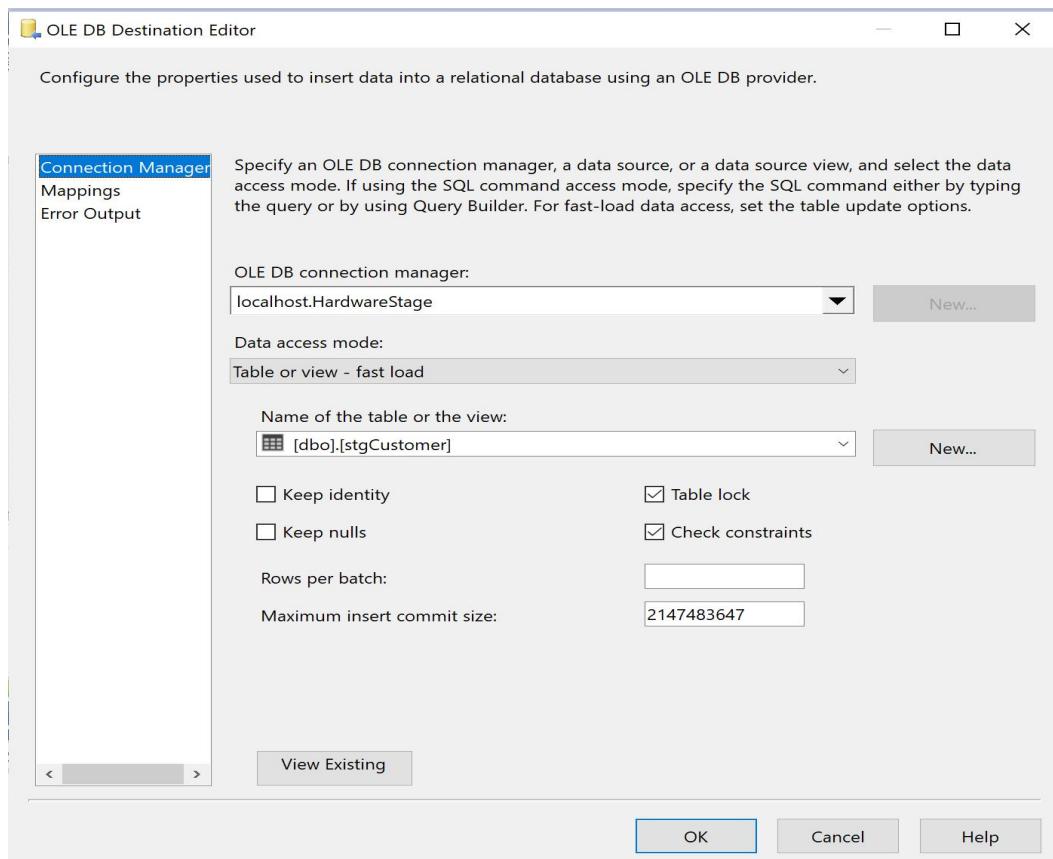
3.3.4.Customer Dimension

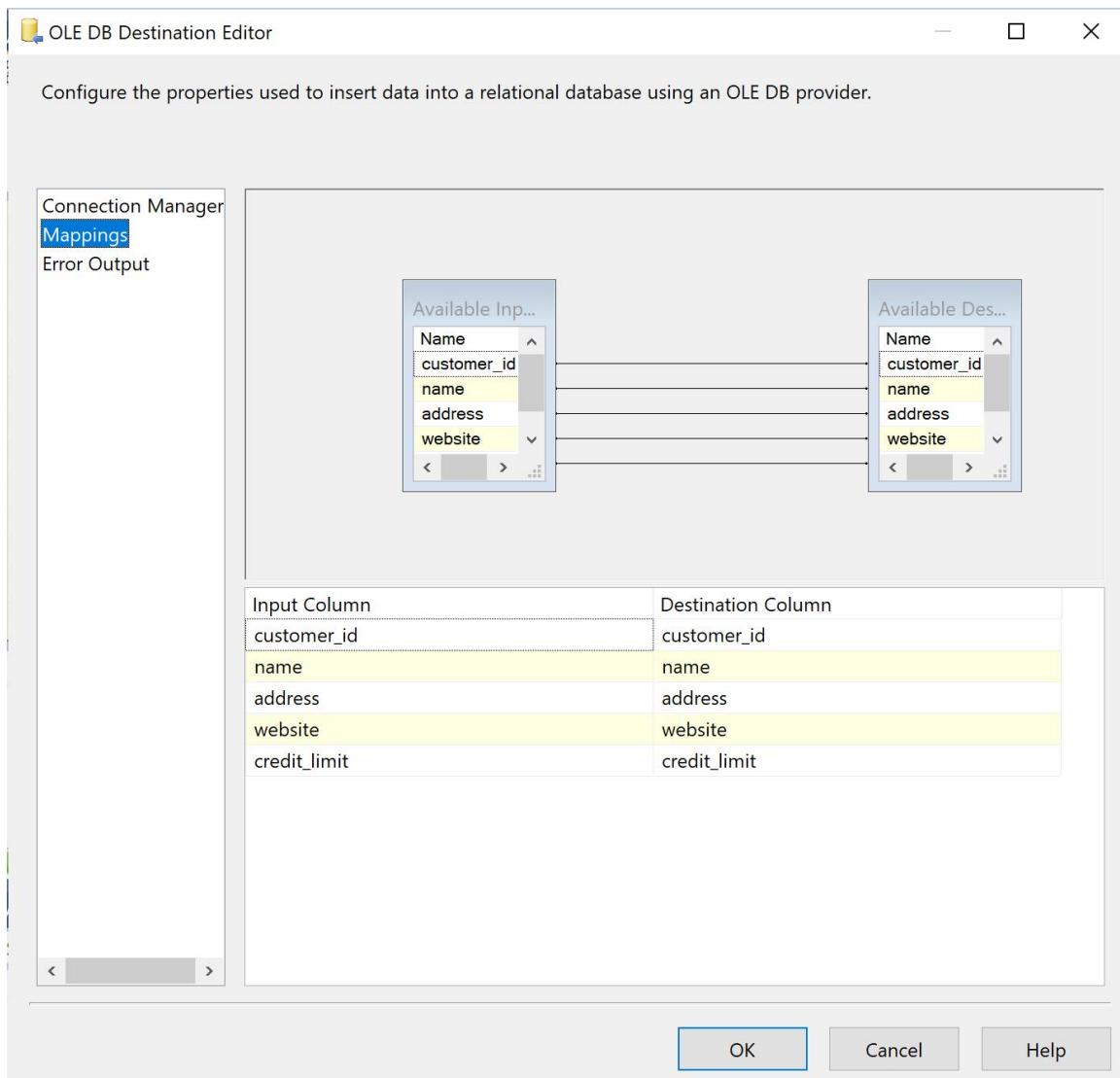
3.3.4.1. Load dữ liệu từ Source vào Stage

- 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**:

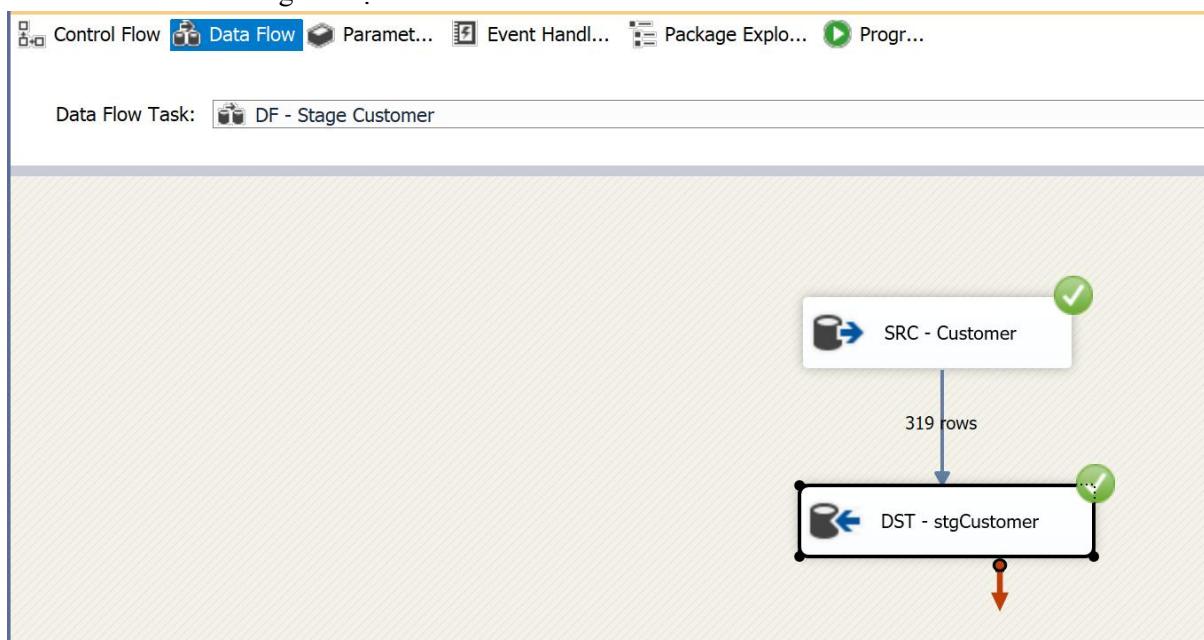


- + **Destination Assistant**:



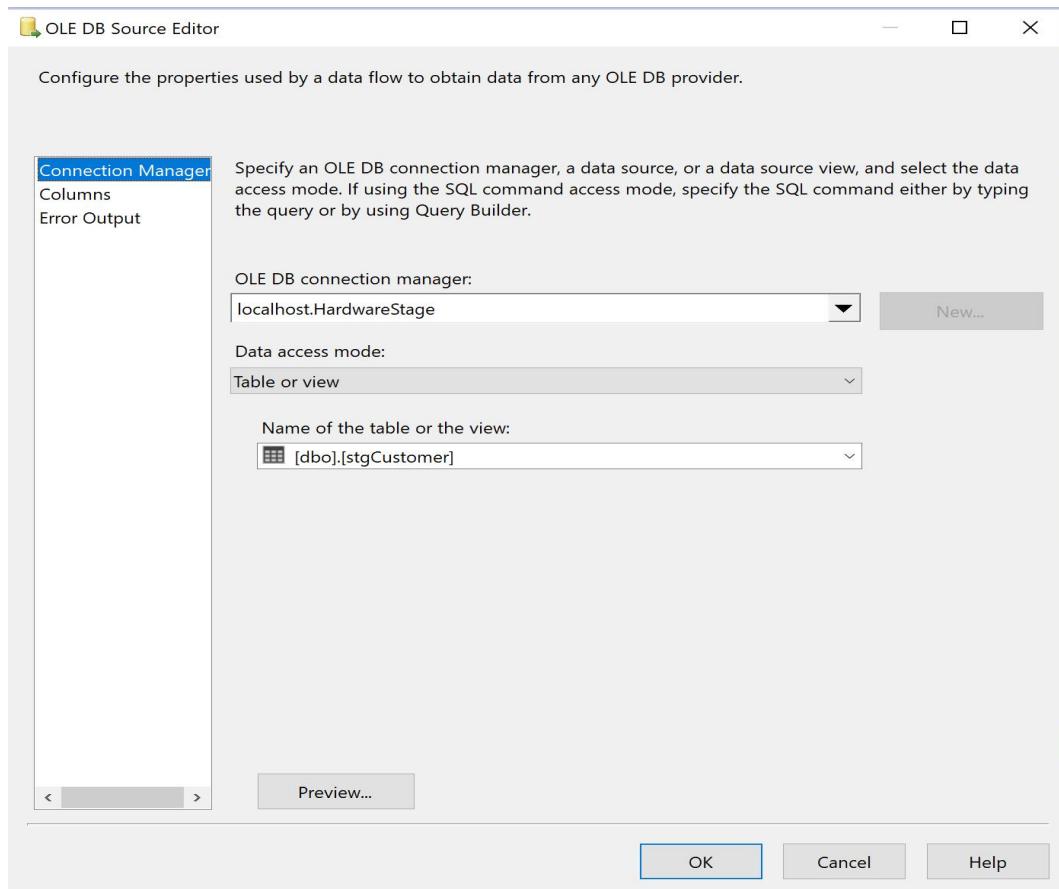


+ Sau khi cấu hình xong ta được:



3.3.4.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load to 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:

localhost.HardwareDW ▼ New

Table or view:

[dbo].[DimCustomer]

Input Columns	Dimension Columns	Key Type
address	Address	Not a key column
credit_limit	CreditLimit	Not a key column
customer_id	CustomerID	Business key
name	CustomerName	Not a key column
	RowChangeReason	
	RowEndDate	
	RowIsCurrent	
	RowStartDate	

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 a change type for each dimension column.

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.

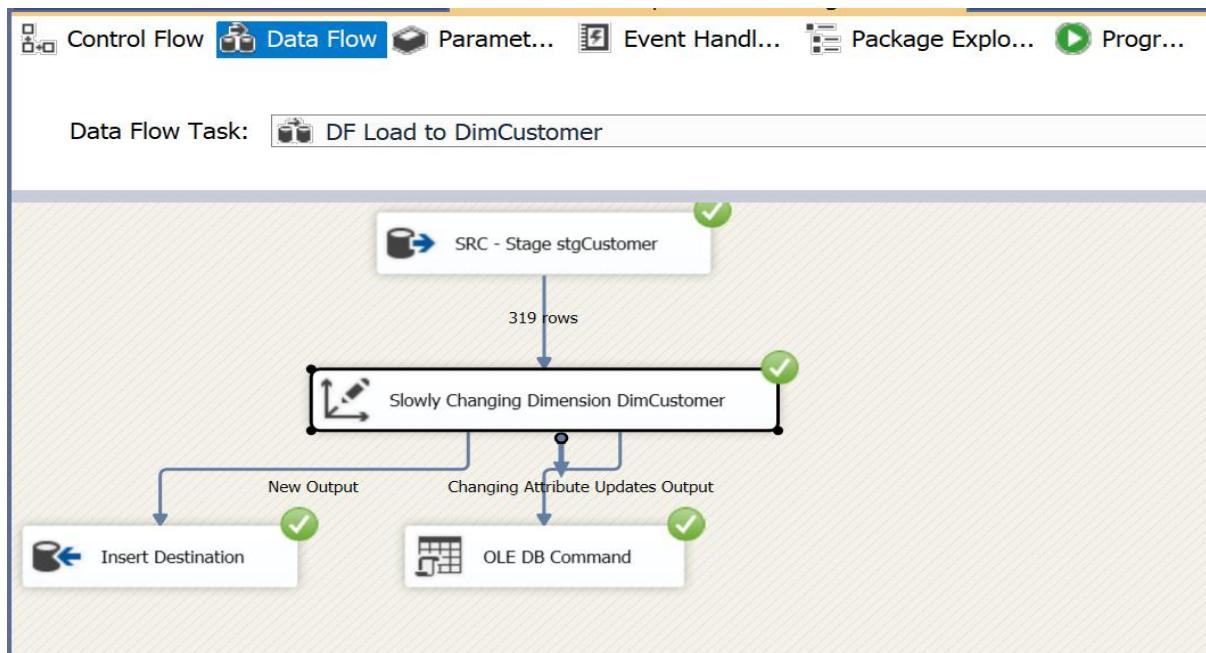
Historical Attribute
Select this type when changes in column values are saved in new records. Previous values are saved in history records.

Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
Address	Changing a...
CreditLimit	Changing a...
CustomerName	Changing a...
Website	Changing a...

Help < Back Next > Finish >> Cancel

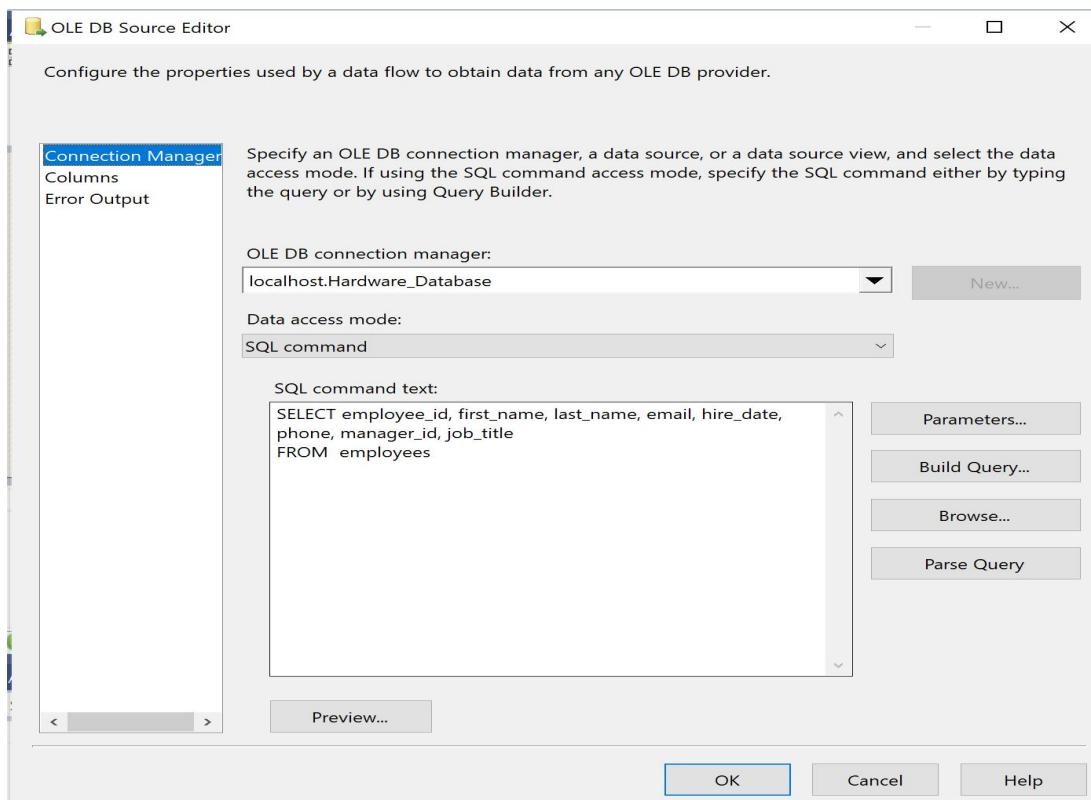
+ Sau khi cấu hình xong ta được:



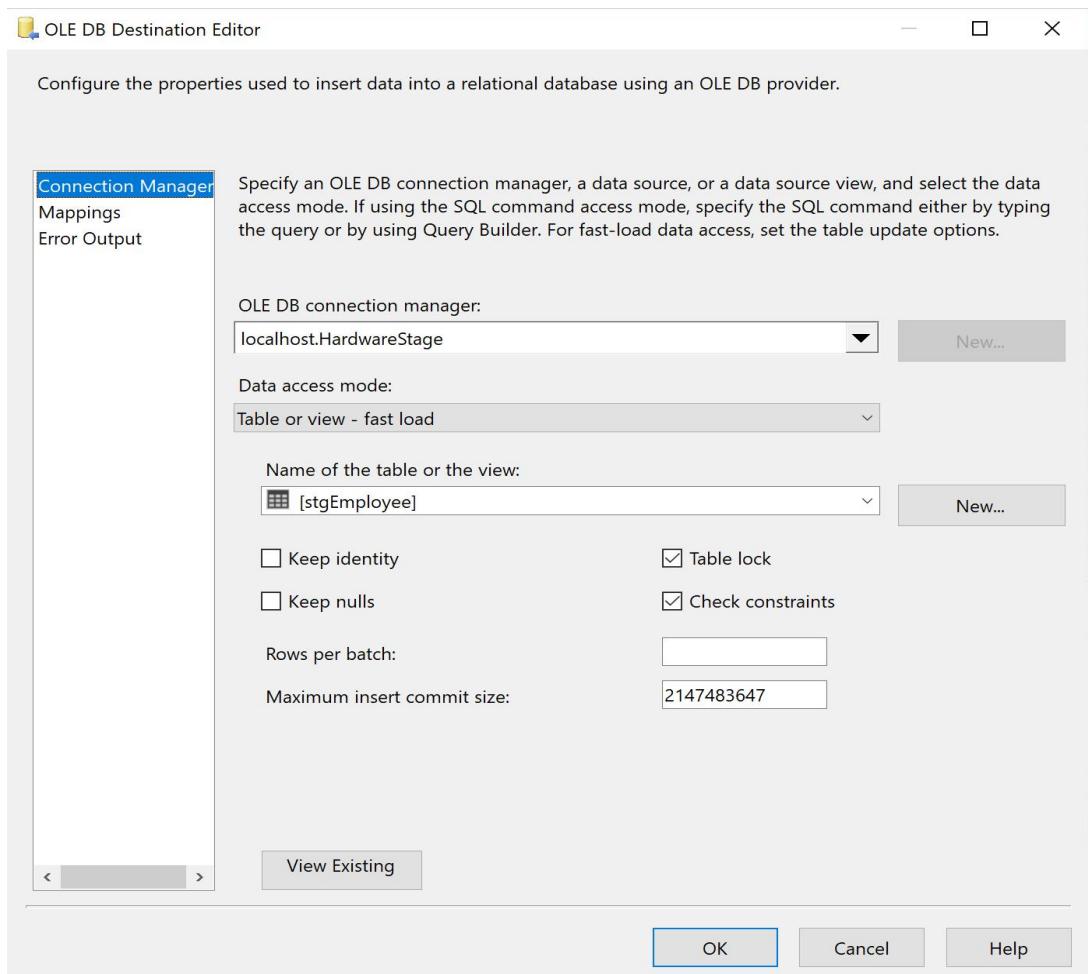
3.3.5. Employee Dimension

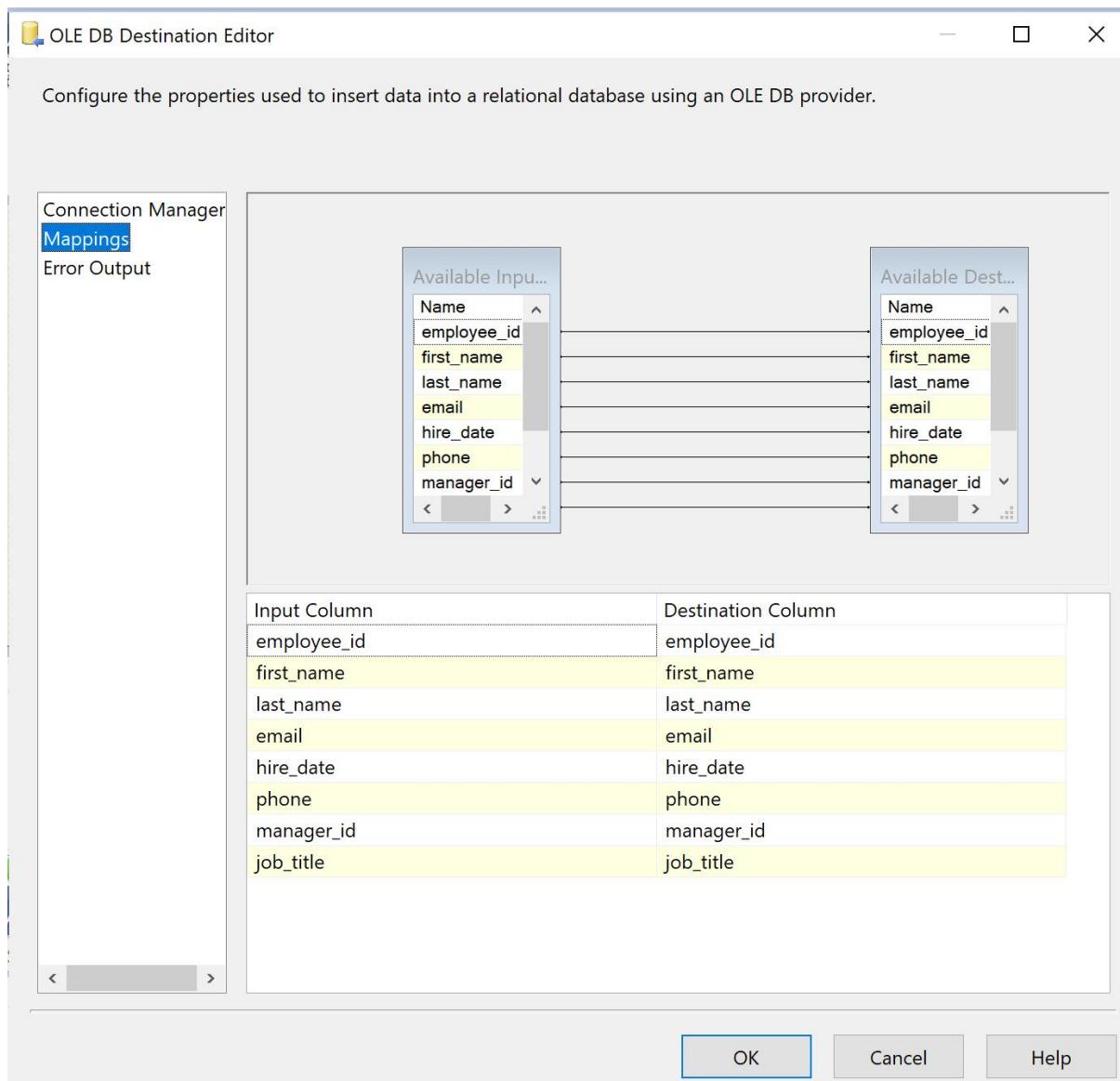
3.3.5.1. Load dữ liệu từ Source vào Stage

- 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:

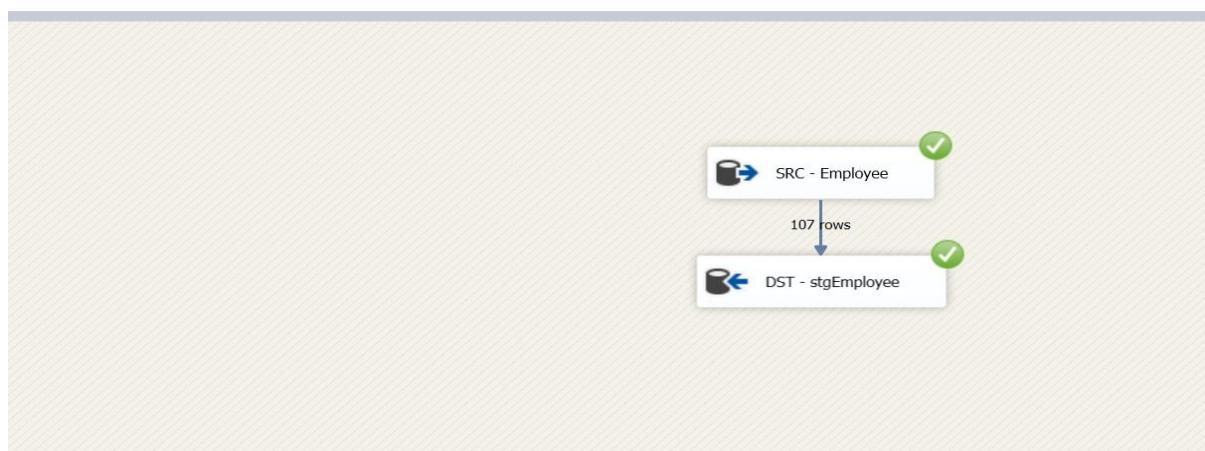




+ Sau khi cấu hình xong ta được:

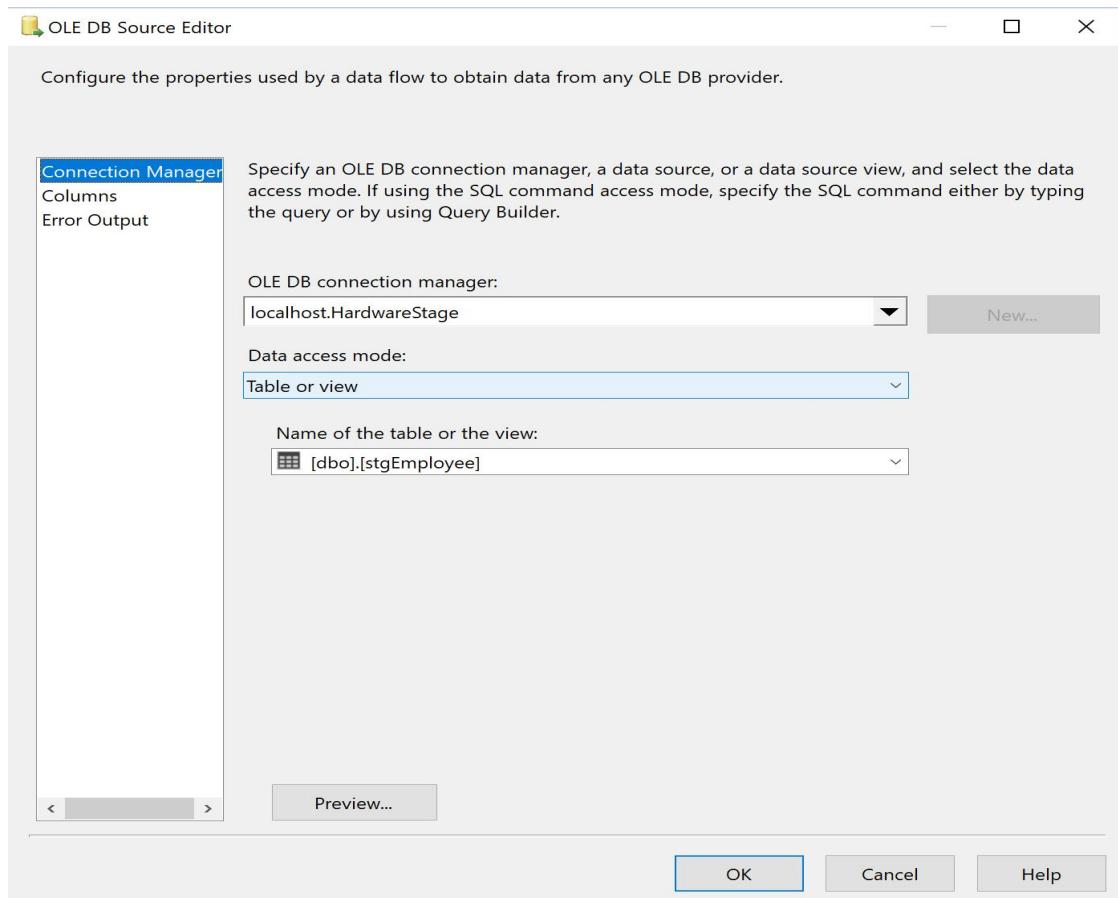
Control Flow Data Flow Parameter Event Handler Package Explorer Program...

Data Flow Task: DF - Stage Employee



3.3.5.2. 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 columns in the target table.

Connection manager:

localhost.HardwareDW

Table or view:

[dbo].[DimEmployee]

Input Columns	Dimension Columns	Key Type
email	Email	Not a key column
employee_id	EmployeeID	Business key
first_name	FirstName	Not a key column
hire_date	HireDate	Not a key column
job_title	JobTitle	Not a key column
last_name	LastName	Not a key column
manager_id	ManagerID	Not a key column
phone	Phone	Not a key column
	RowChangeReason	
	RowEndDate	
	RowIsCurrent	
	RowStartDate	

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.

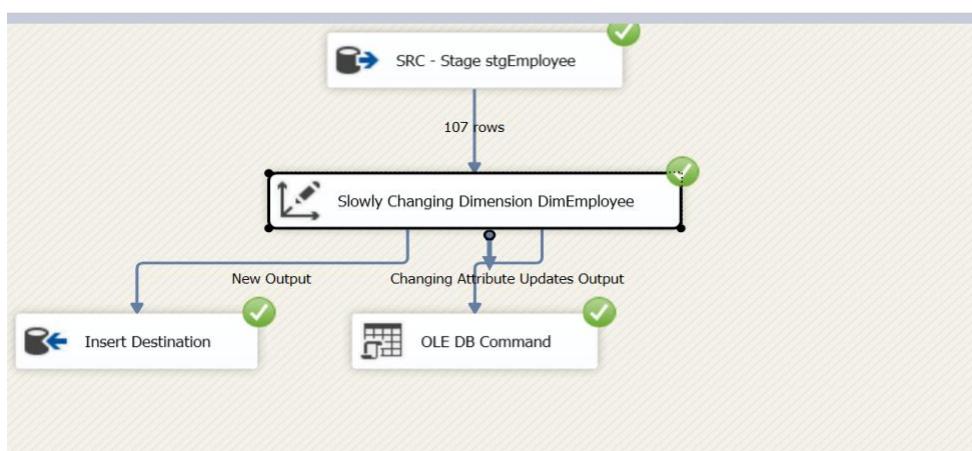
Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
Email	Changing a...
FirstName	Changing a...
HireDate	Changing a...
JobTitle	Changing a...
LastName	Changing a...
ManagerID	Changing a...
Phone	Changing a...

+ Sau khi cấu hình xong ta được:



Data Flow Task: DF Load to DimEmployee



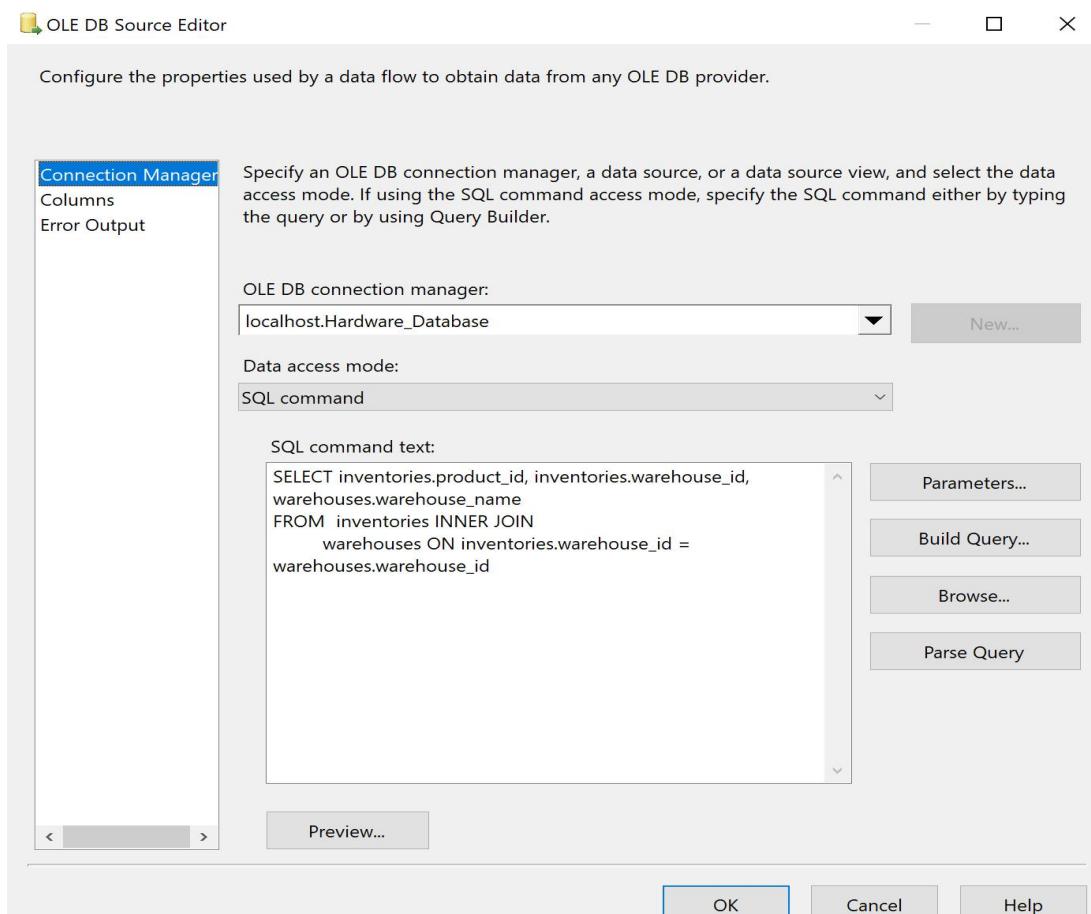
3.3.6. Inventory Dimension

3.3.6.1. Load dữ liệu từ Source vào Stage

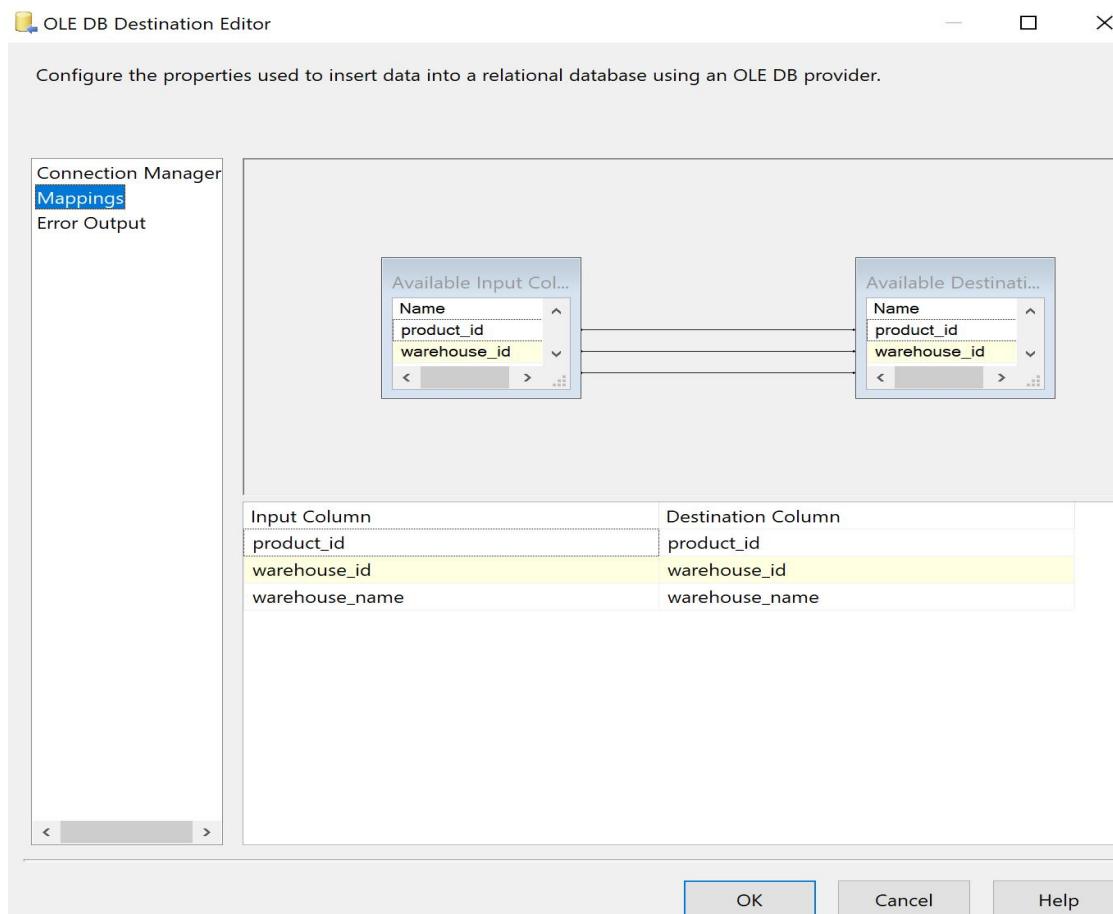
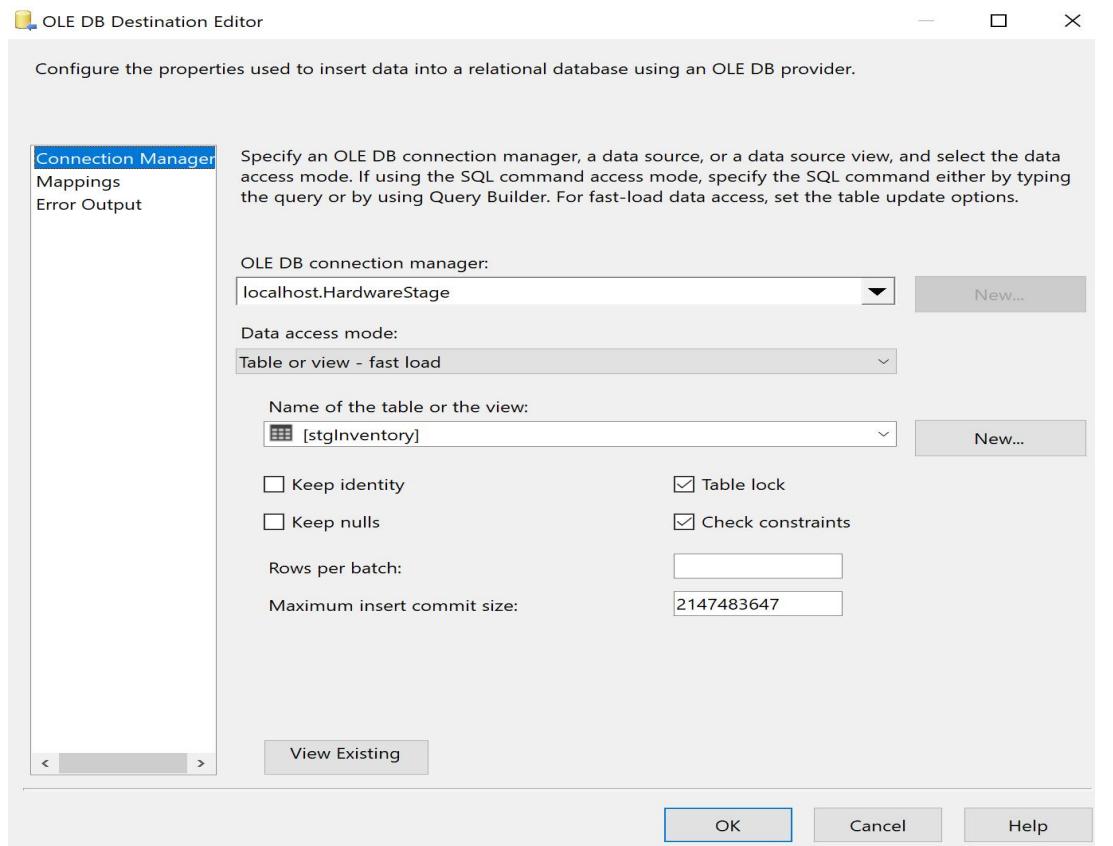
- Double-click **DF - Stage Inventory** task để mở nó trong **Data Flow design surface**:

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

+ Source Assistant:



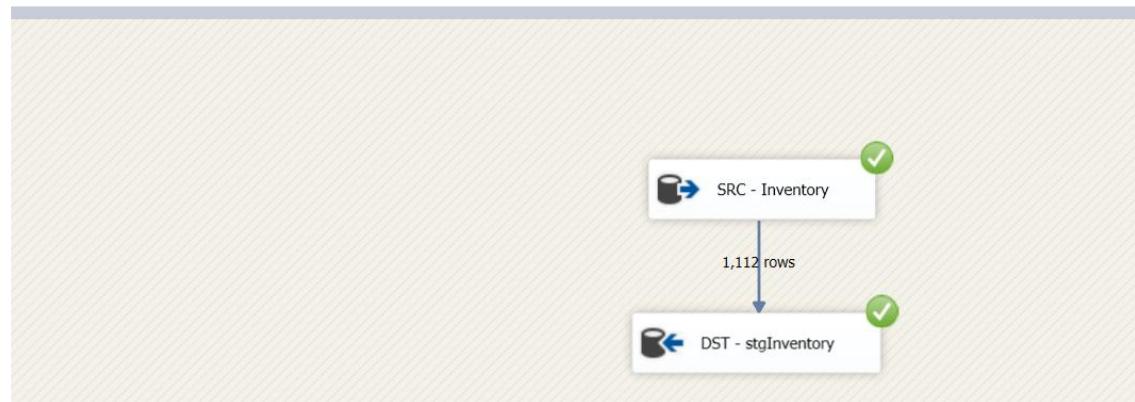
+ Destination Assistant:



+ Sau khi cấu hình xong ta được:



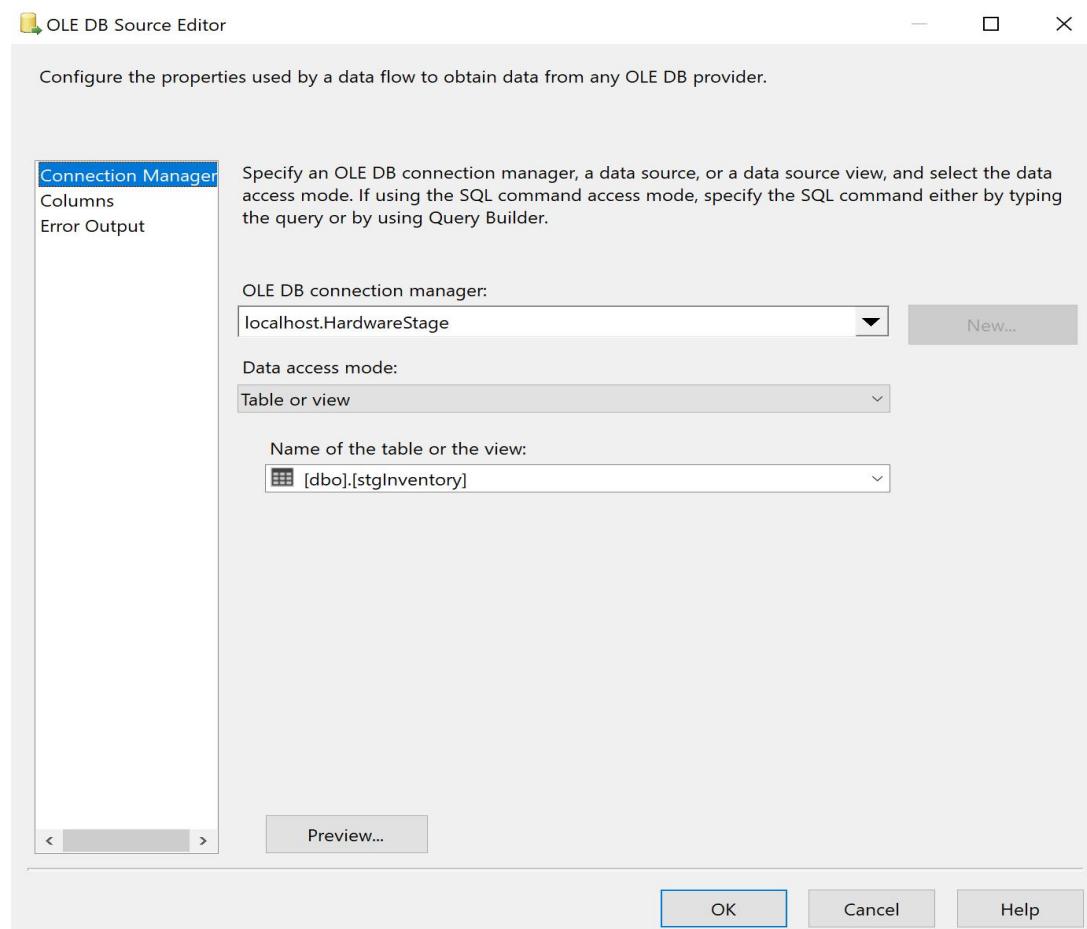
Data Flow Task: DF - Stage Inventory



3.3.6.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load to DimInventory** 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.HardwareDW

Table or view:

[dbo].[DimInventory]

Input Columns	Dimension Columns	Key Type
product_id	ProductID	Business key
	RowChangeReason	
	RowEndDate	
	RowIsCurrent	
	RowStartDate	
warehouse_id	WarehouseId	Business key
warehouse_n...	WarehouseName	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

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.

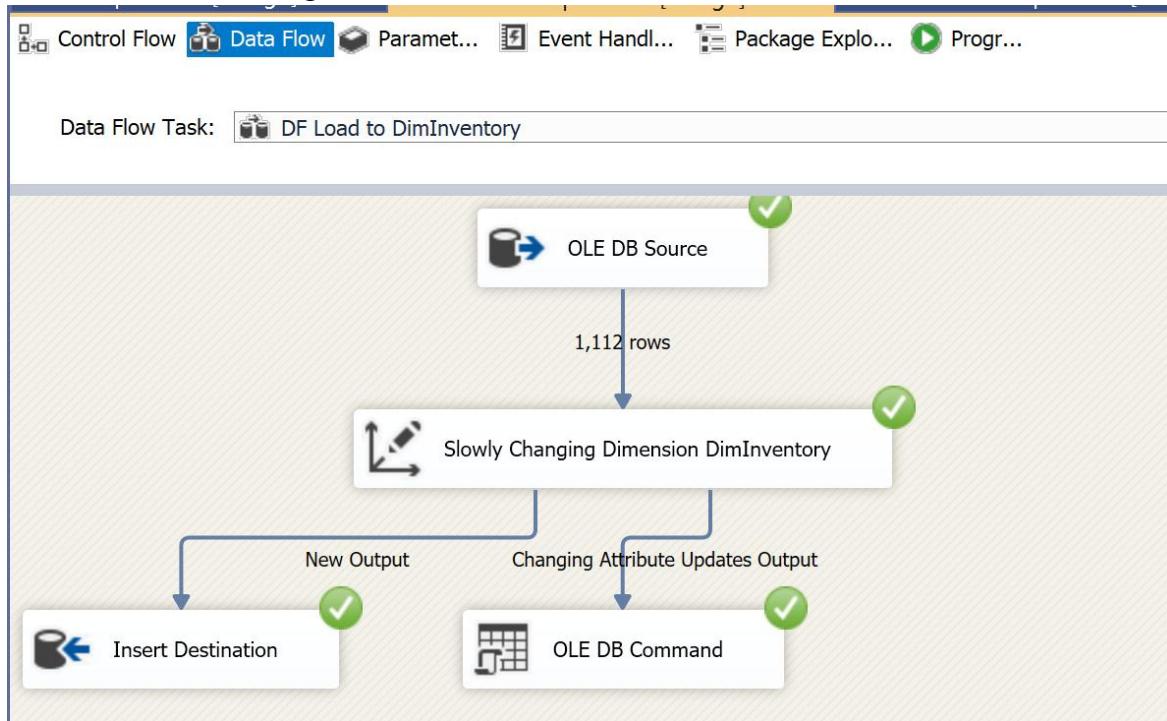
Historical Attribute
Select this type when changes in column values are saved in new records. Previous values are saved in

Select a change type for slowly changing dimensions:

Dimension Columns	Change Type
WarehouseName	Changing a...

Help < Back Next > Finish >> Cancel

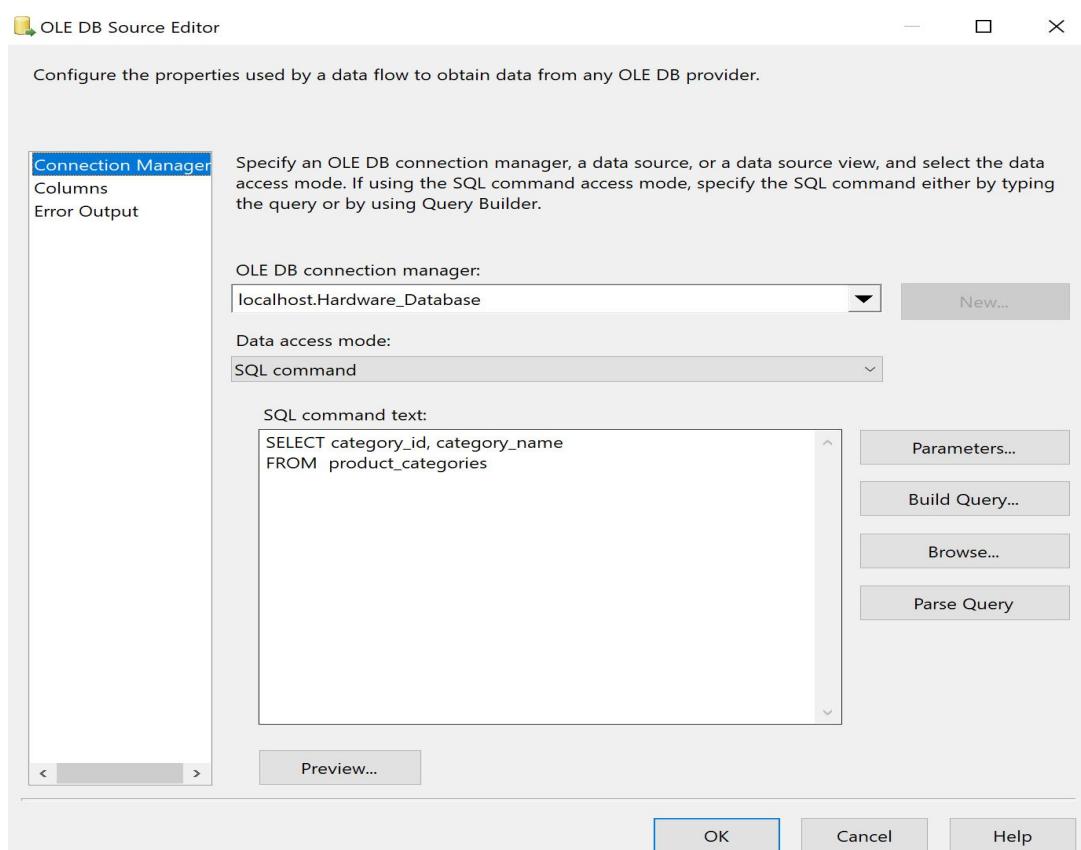
+ Sau khi cấu hình xong ta được:



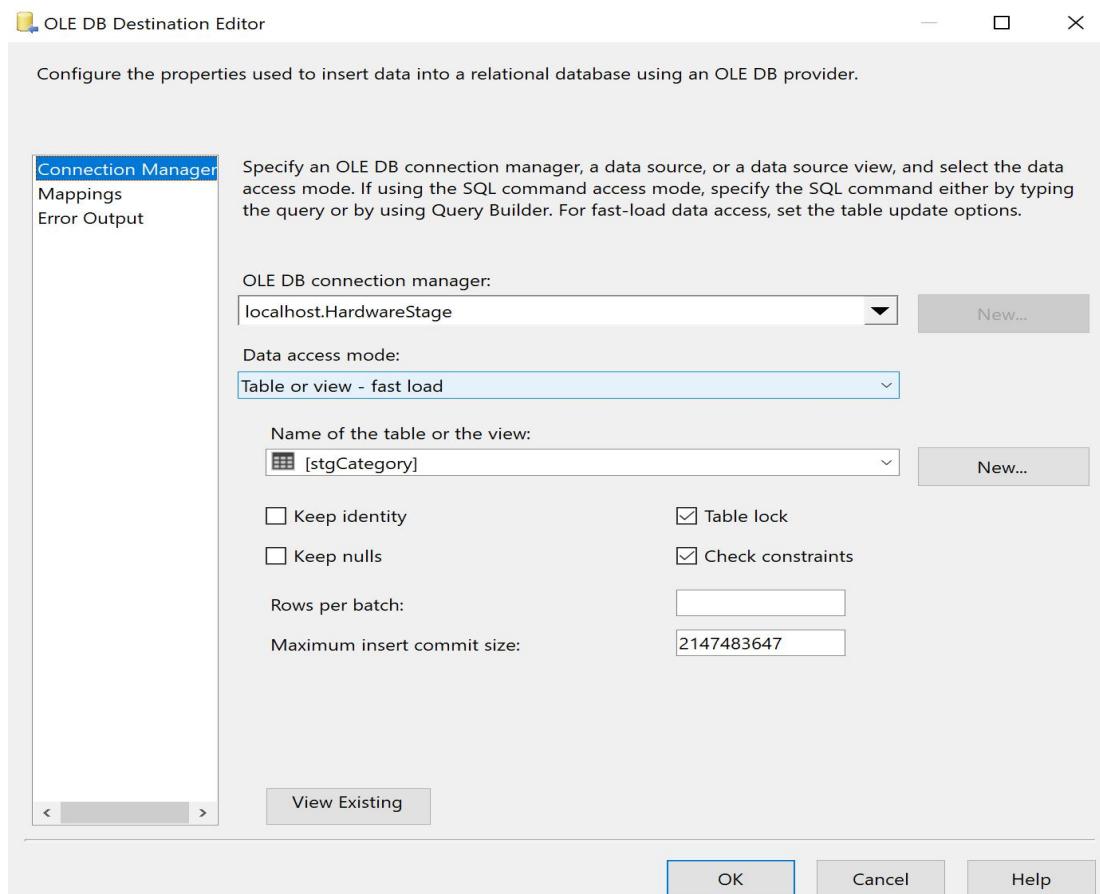
3.3.7.Category Dimension

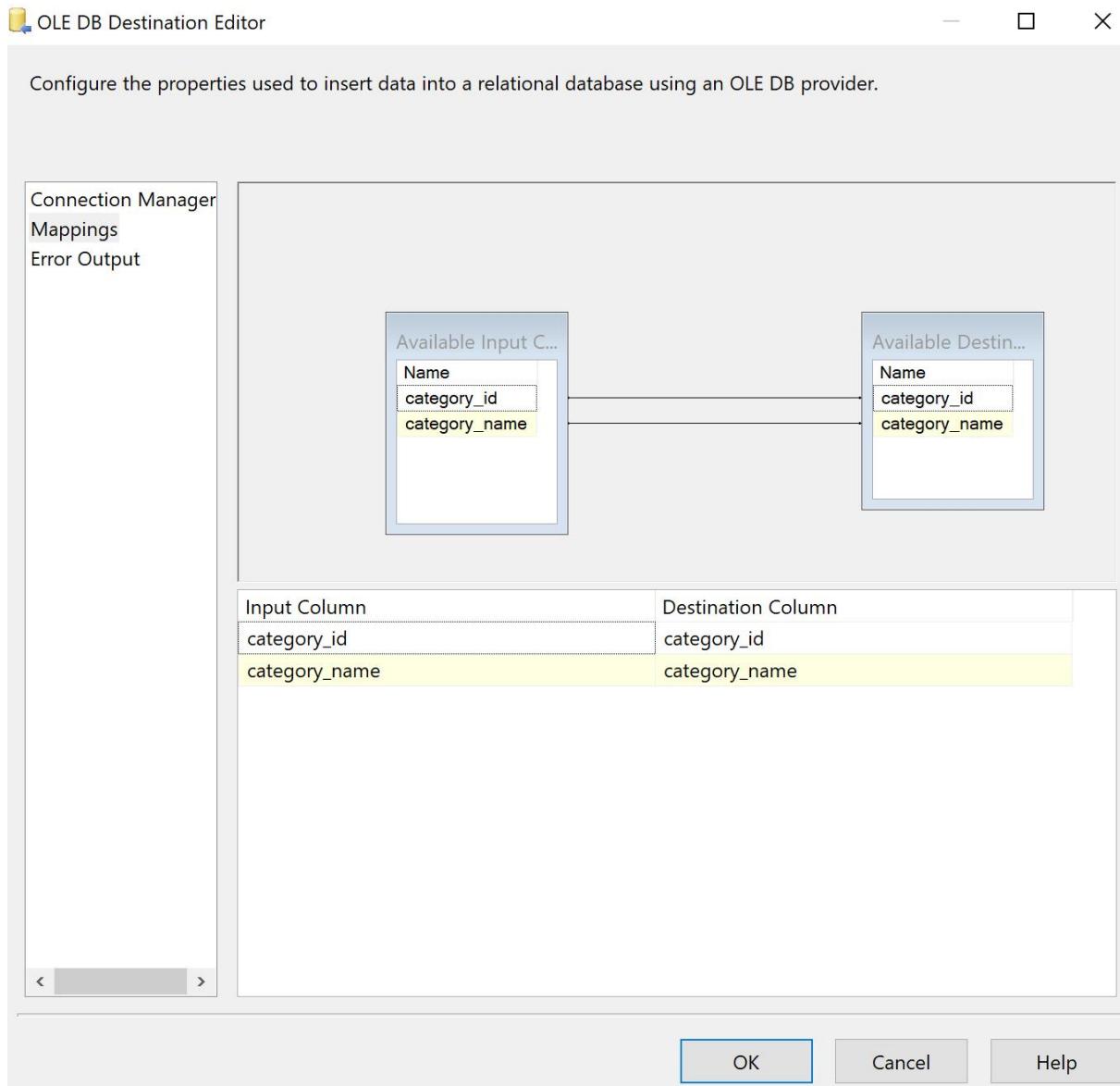
3.3.7.1. Load dữ liệu từ Source vào Stage

- 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:

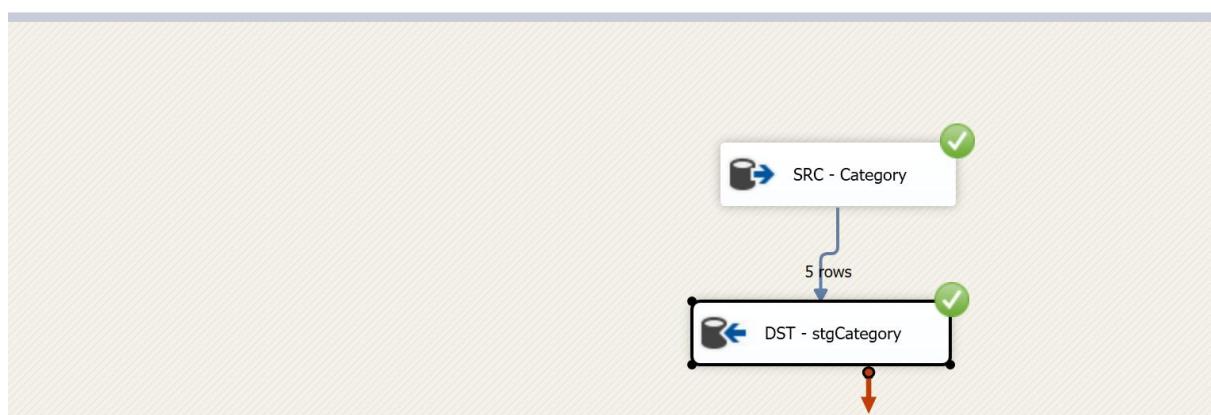




+ Sau khi cấu hình xong ta được:

Control Flow Data Flow Paramet... Event Handl... Package Explor... Program...

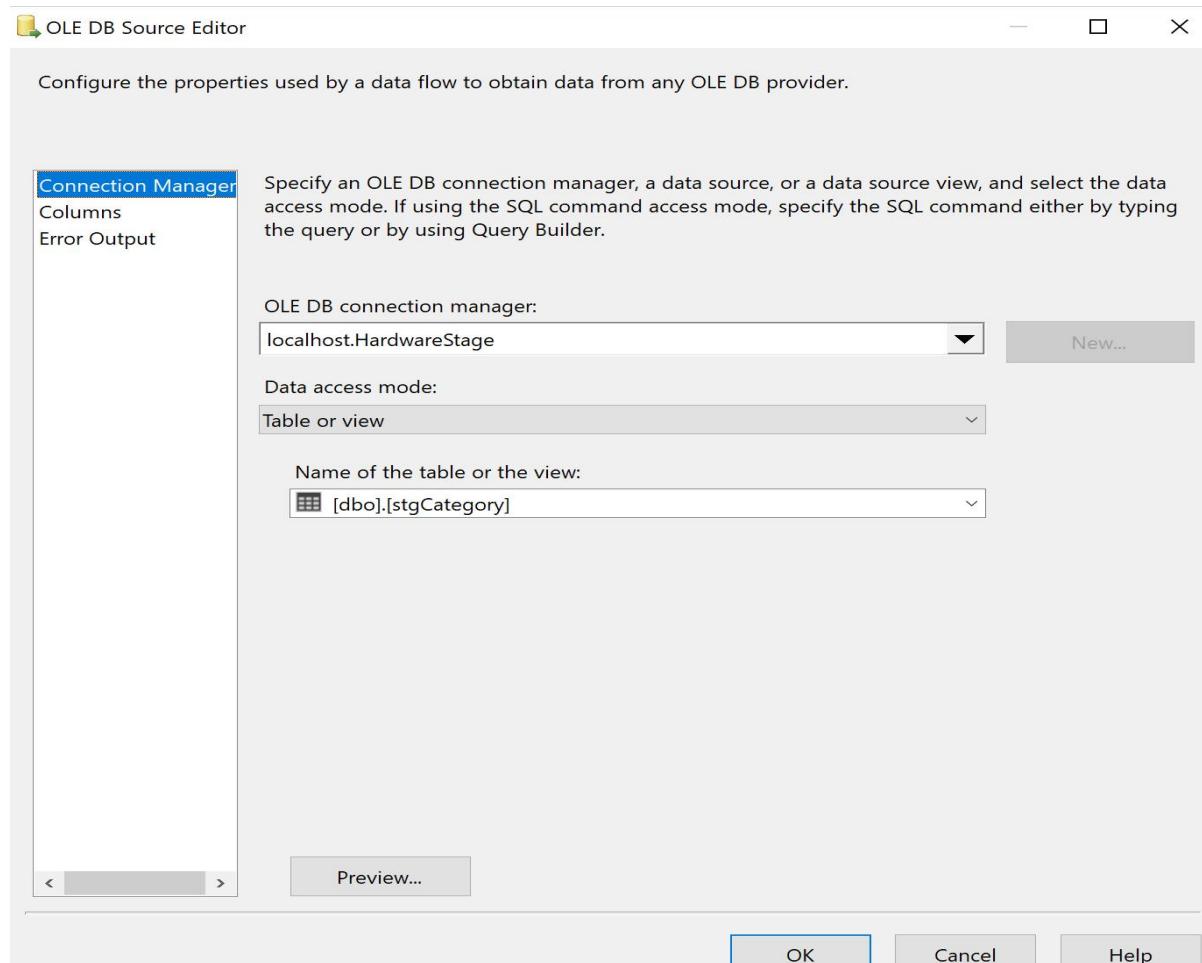
Data Flow Task: DF - Stage Category



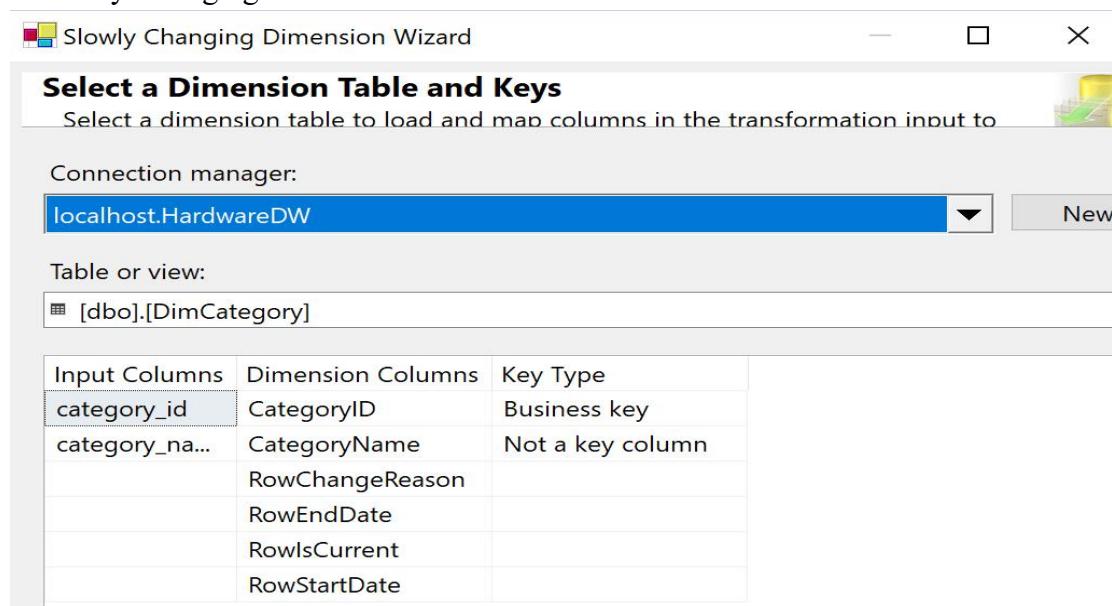
3.3.7.2. Load dữ liệu từ Stage vào Dimension

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

+ Source Assistant:



+ Slowly Changing Dimension:



 Slowly Changing Dimension Wizard

Slowly Changing Dimension Columns

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

Fixed Attribute

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

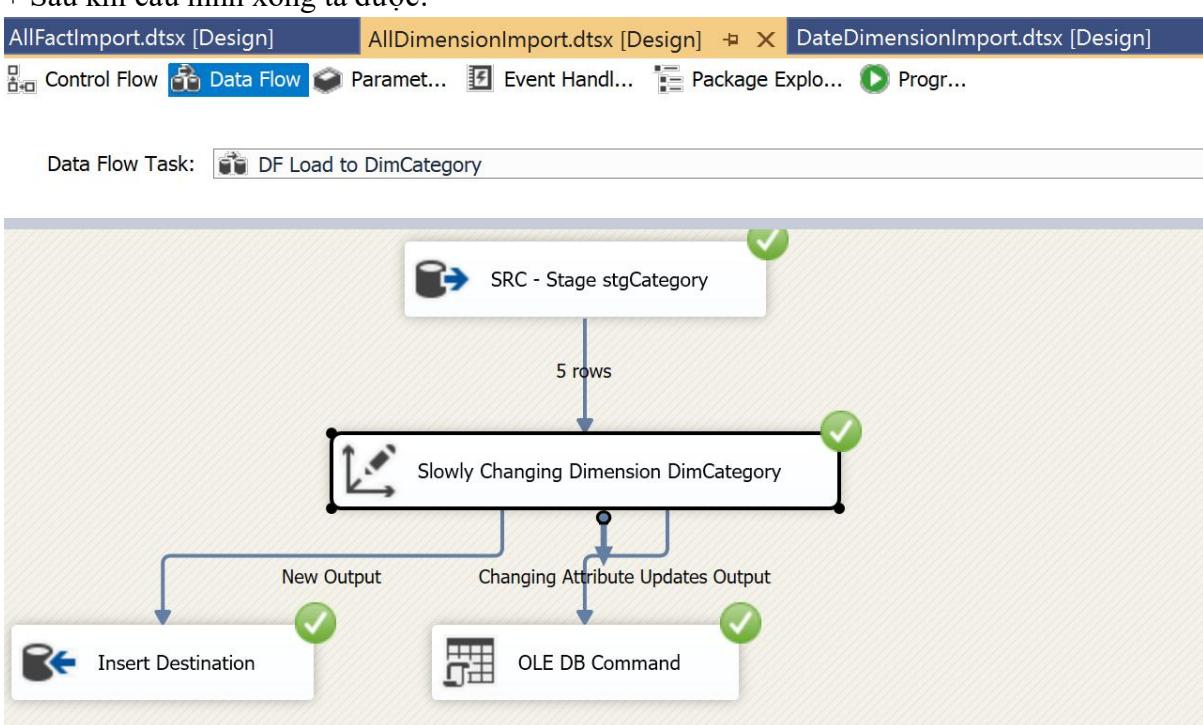
Changing Attribute

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

+ Sau khi cấu hình xong ta được:

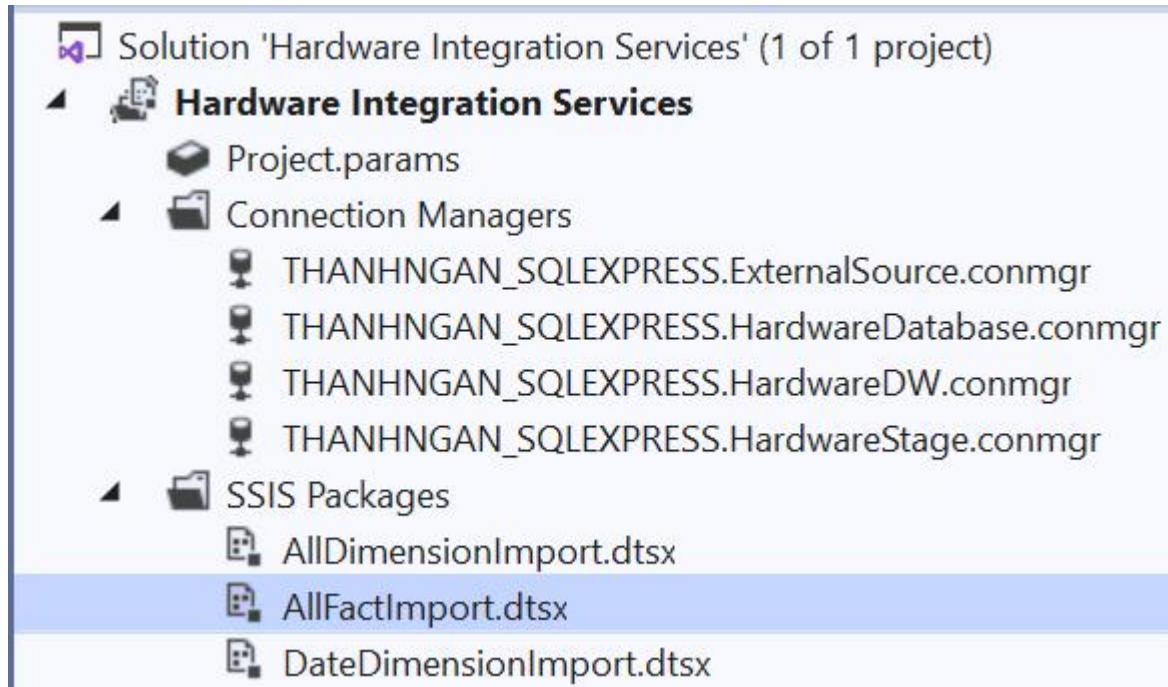


3.4. Import dữ liệu vào bảng các bảng fact

3.4.1. Fact Sales

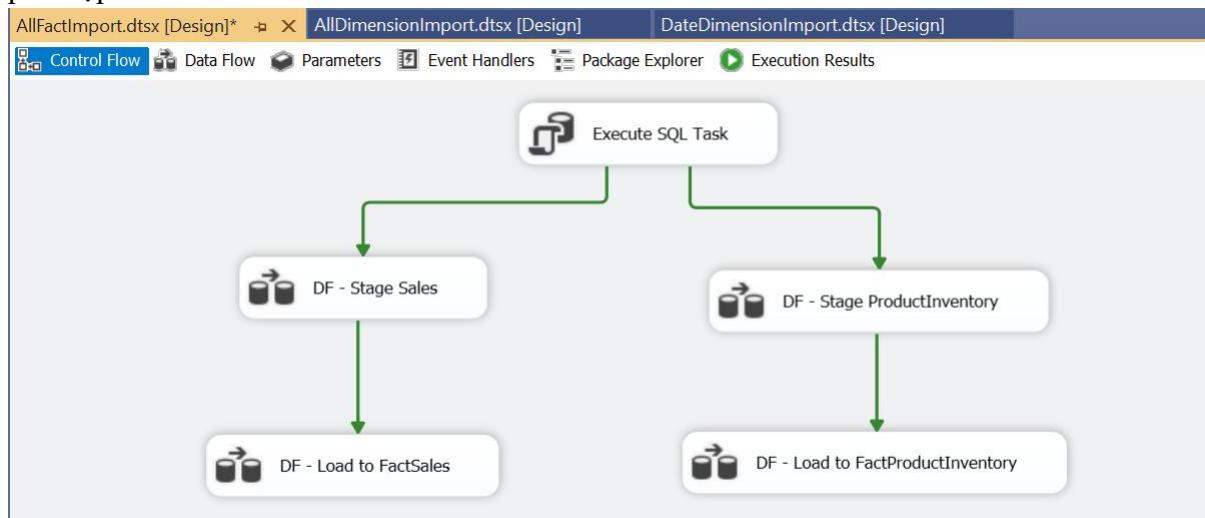
3.4.1.1. Tạo Package

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

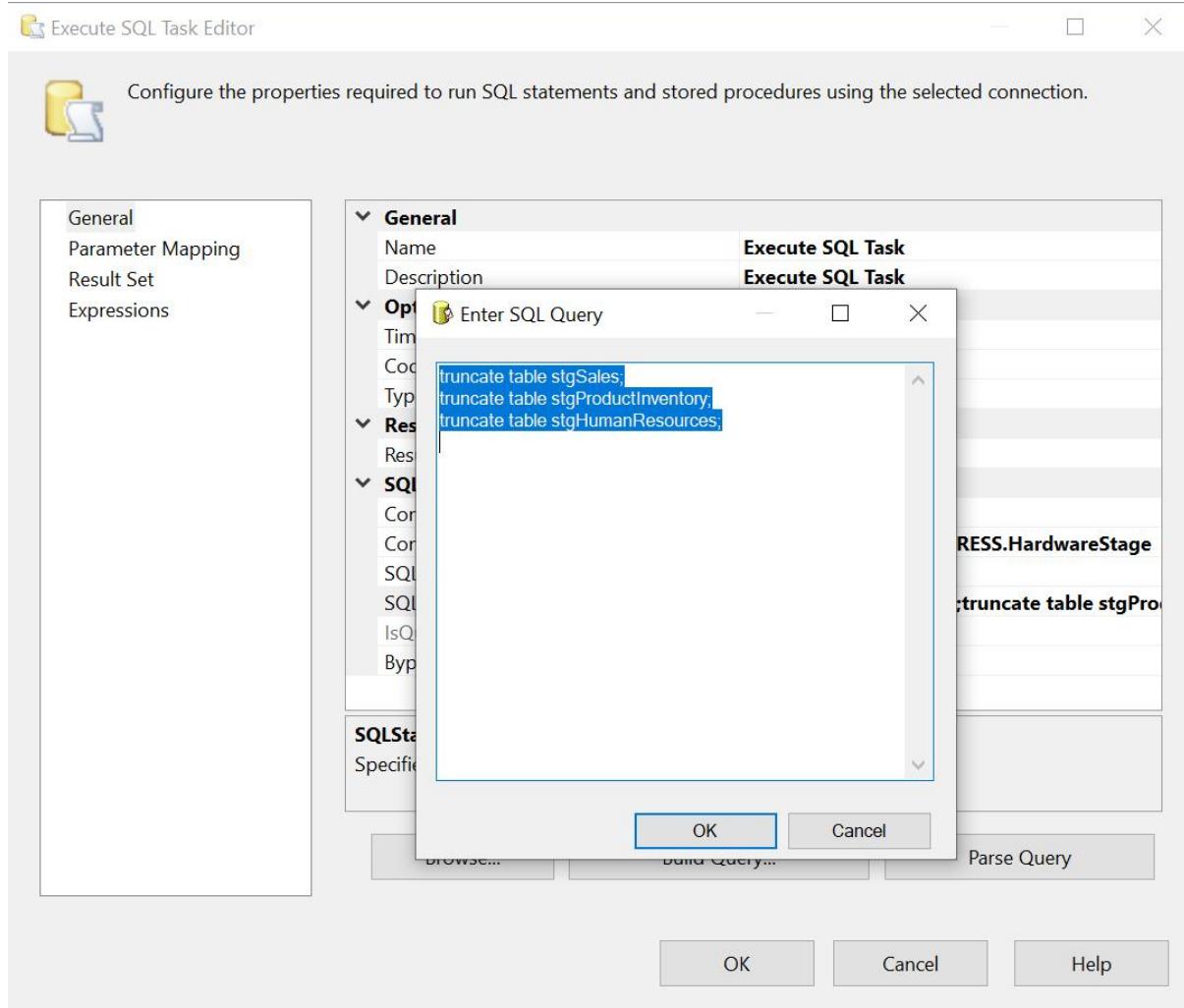


3.4.1.2. Thiết lập Control Flow

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

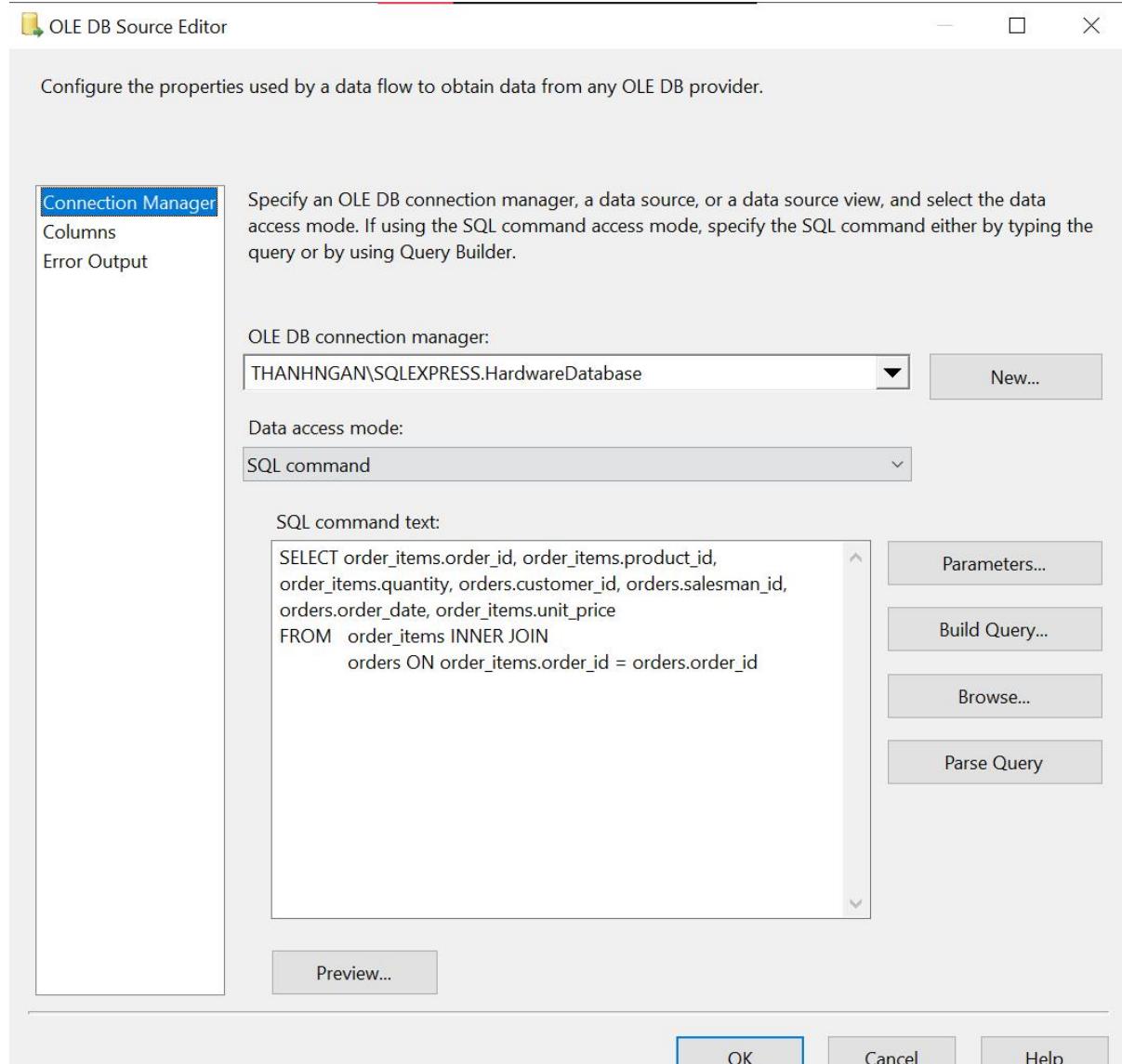


- Thiết lập truncate table SQL task

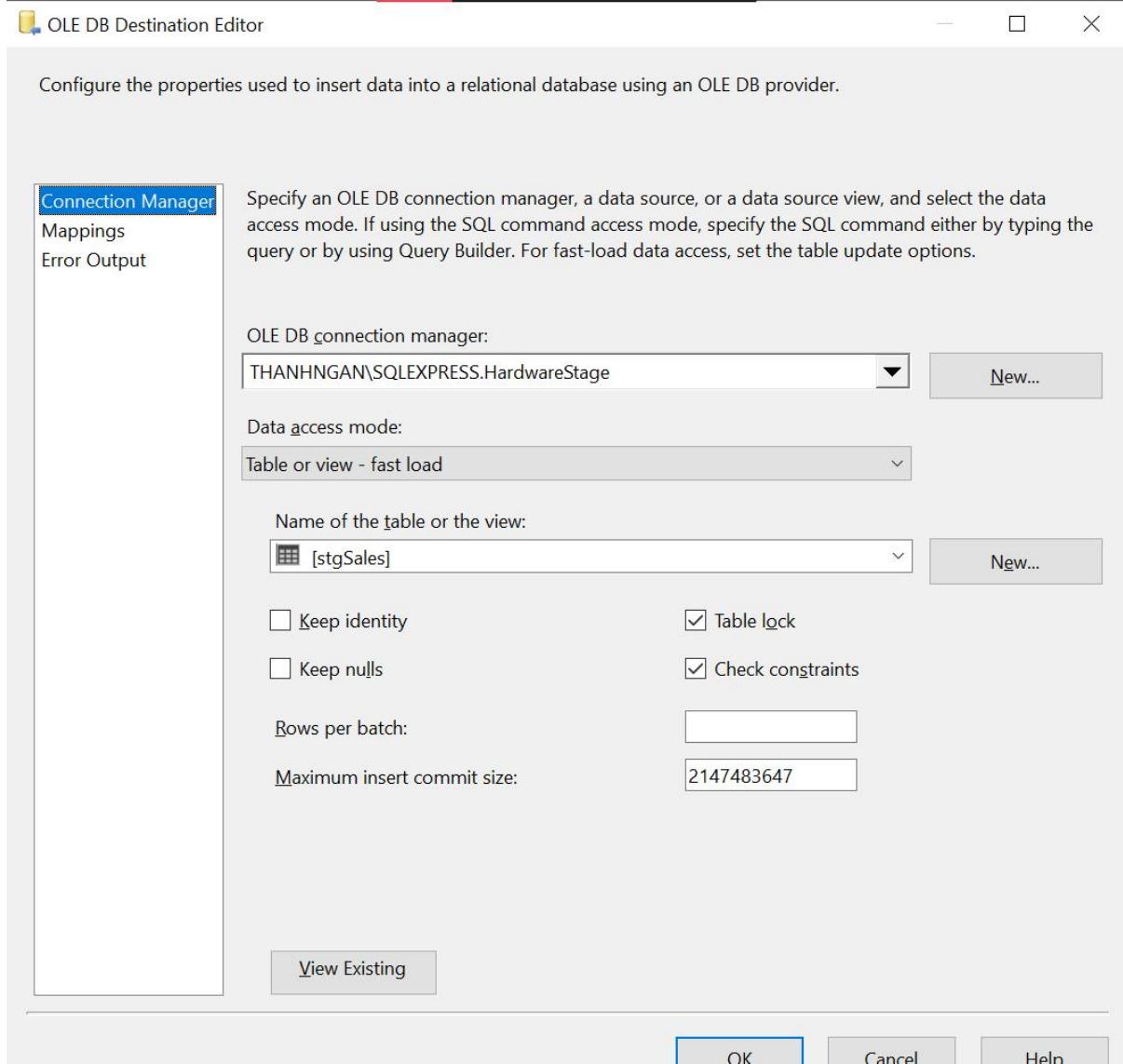


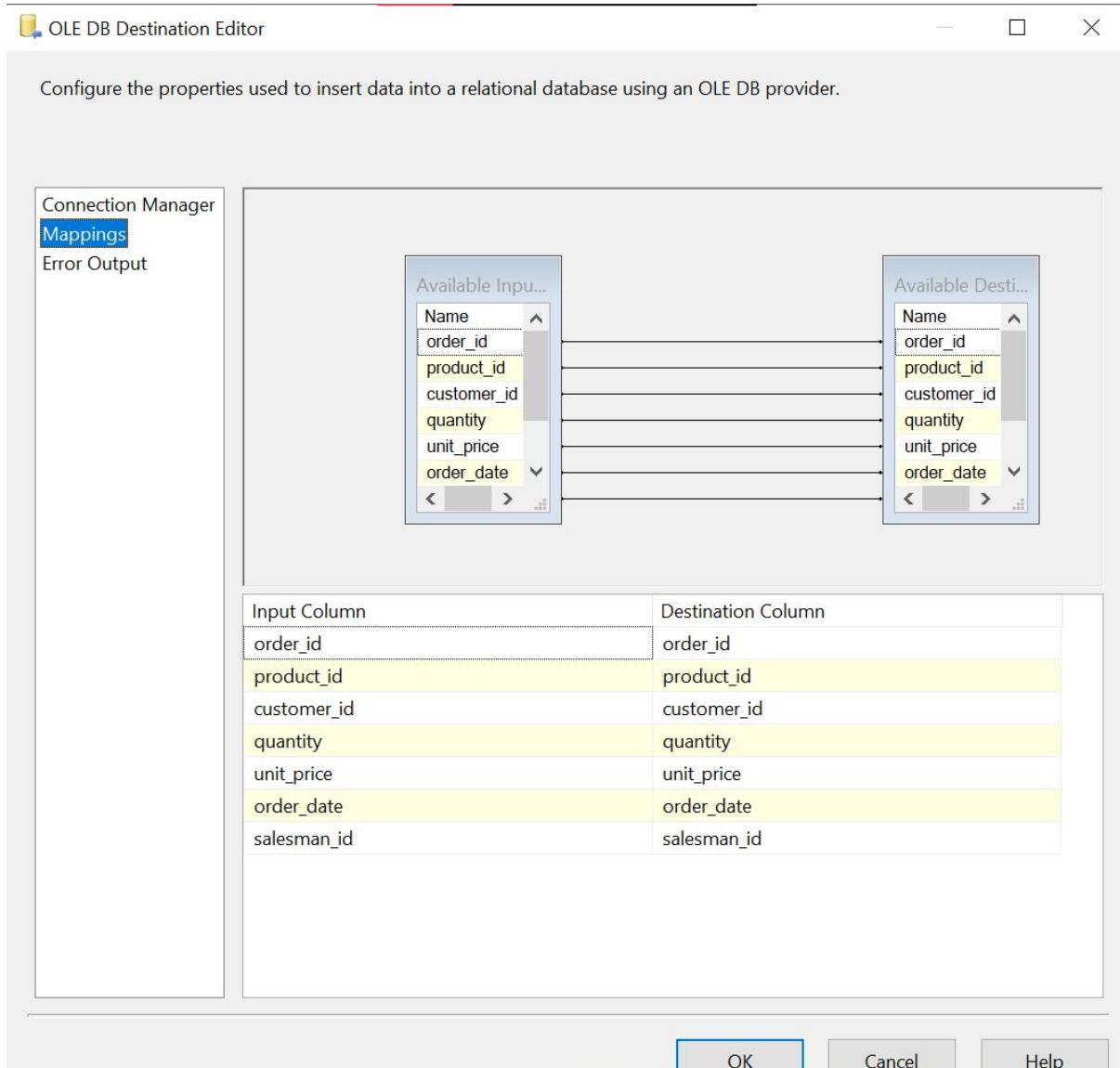
3.4.1.3. Load dữ liệu từ Source vào Stage

- 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:**

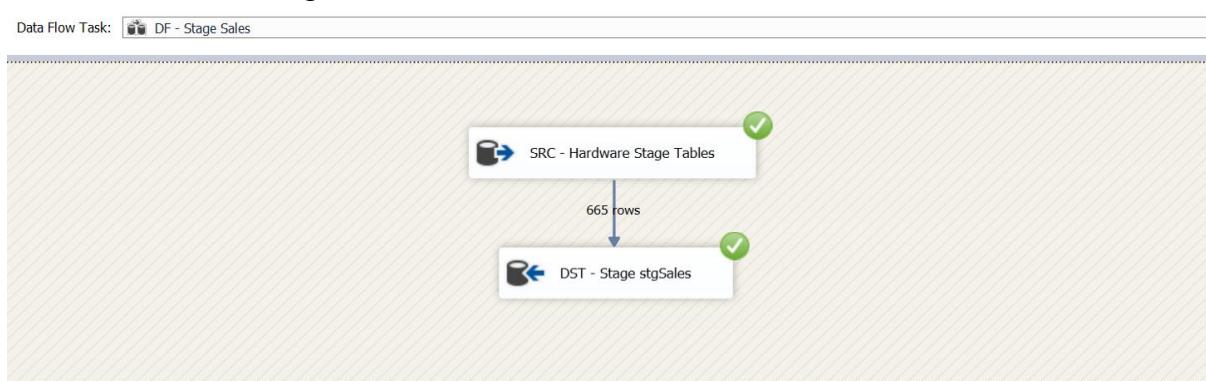


+ Destination Assist



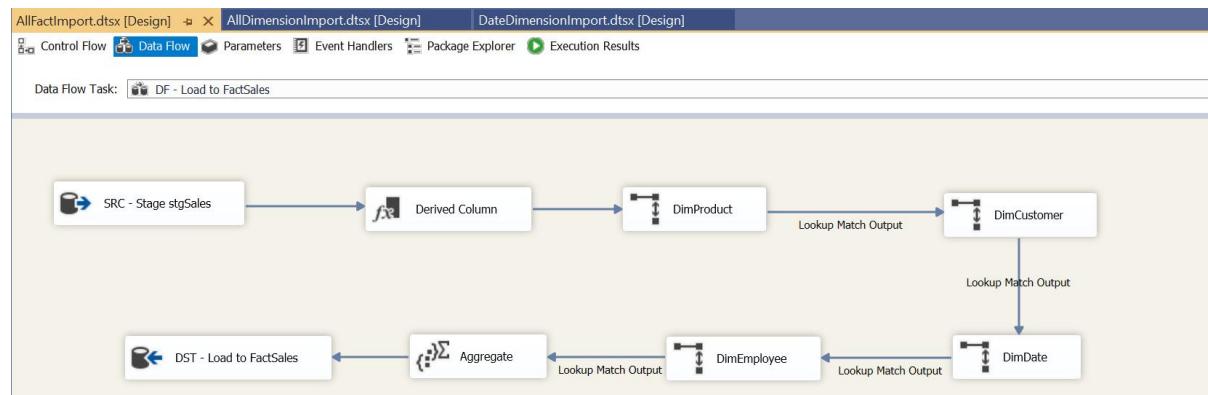


+ Sau khi cấu hình xong ta được:



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**

The Derived Column Transformation Editor window shows the configuration for adding a new column "SalesAmount". The left pane lists "Variables and Parameters" and "Columns". The right pane lists various functions: Mathematical Functions, String Functions, Date/Time Functions, NULL Functions, Type Casts, and Operators. A "Description:" text area is present. The bottom table details the new column:

Derived Column Name	Derived Column	Expression	Data Type	Length
SalesAmount	<add as new column>	unit_price * quantity	four-byte signed integer	9

- **Lookup qua bảng DimProduct**

The Lookup Transformation Editor window shows the configuration for the "DimProduct" lookup. The left pane has tabs for General, Connection (selected), Columns, Advanced, and Error Output. The main area describes the transform as performing simple equi-joins between input and reference data sets. It shows an OLE DB connection manager set to "THANHNGAN\SQLEXPRESS.HardwareDW" and a dropdown for selecting a table or view, currently set to "[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 Lookup Operation Output Alias

ProductKey	<add as new column>	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 Lookup Operation Output Alias

CustomerKey	<add as new column>	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:

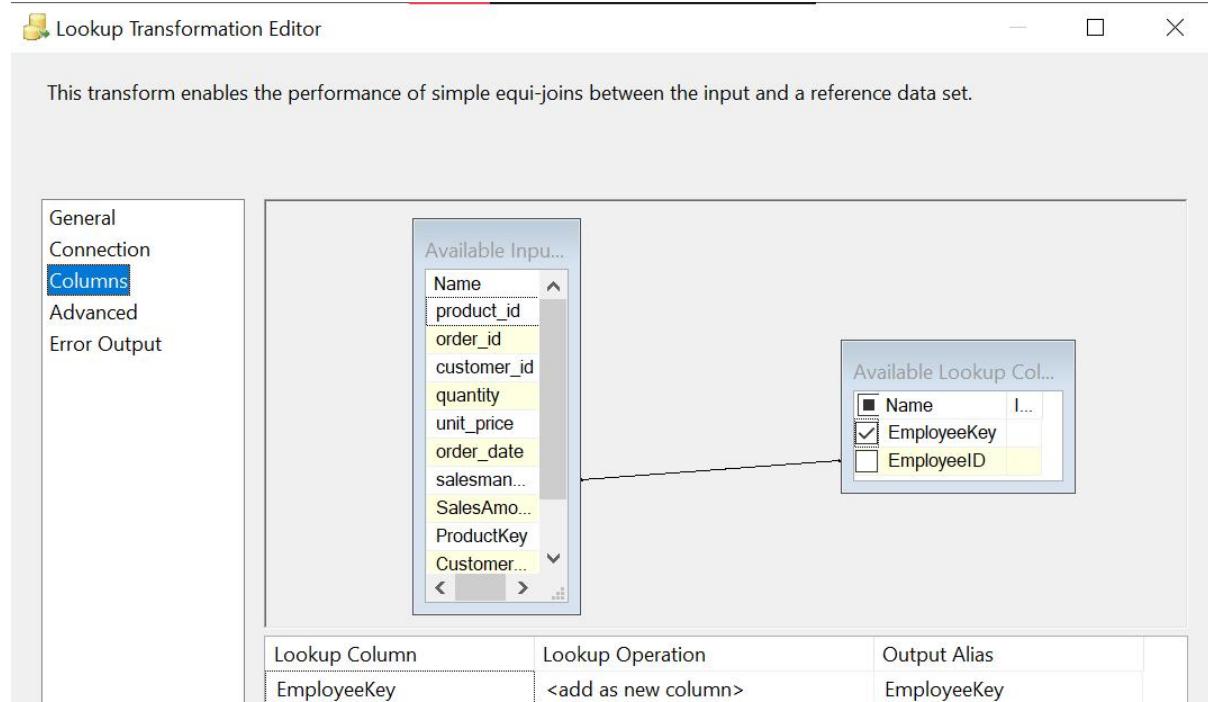
THANHNGAN\SQLEXPRESS.HardwareDW

Use a table or a view:

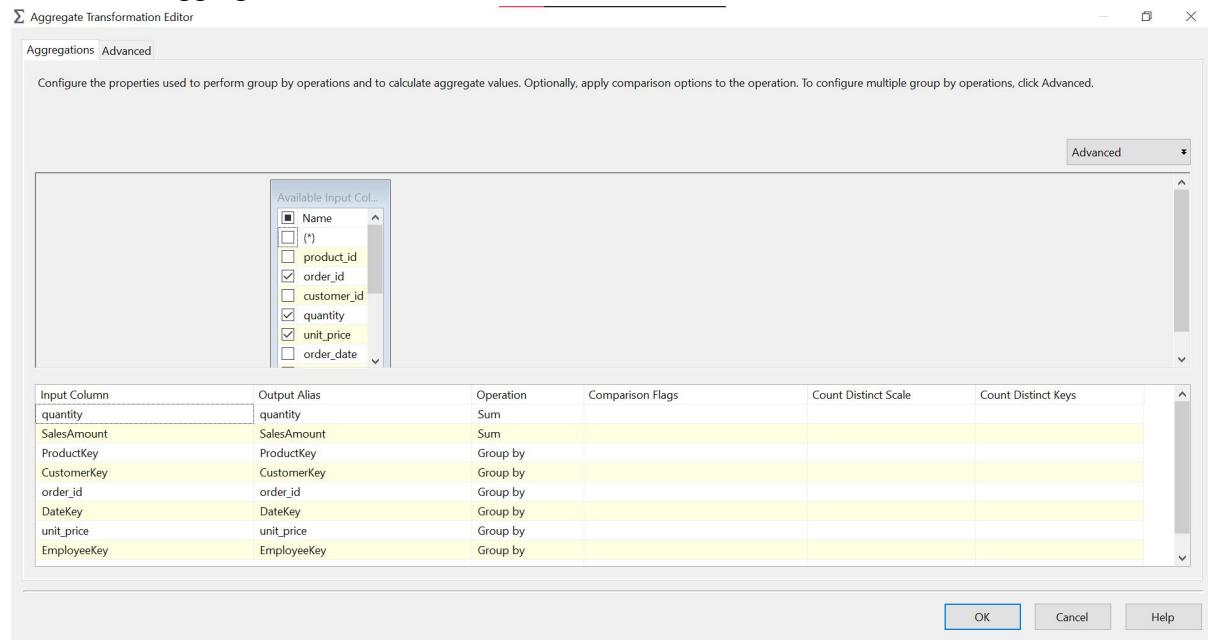
Use results of an SQL query:

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

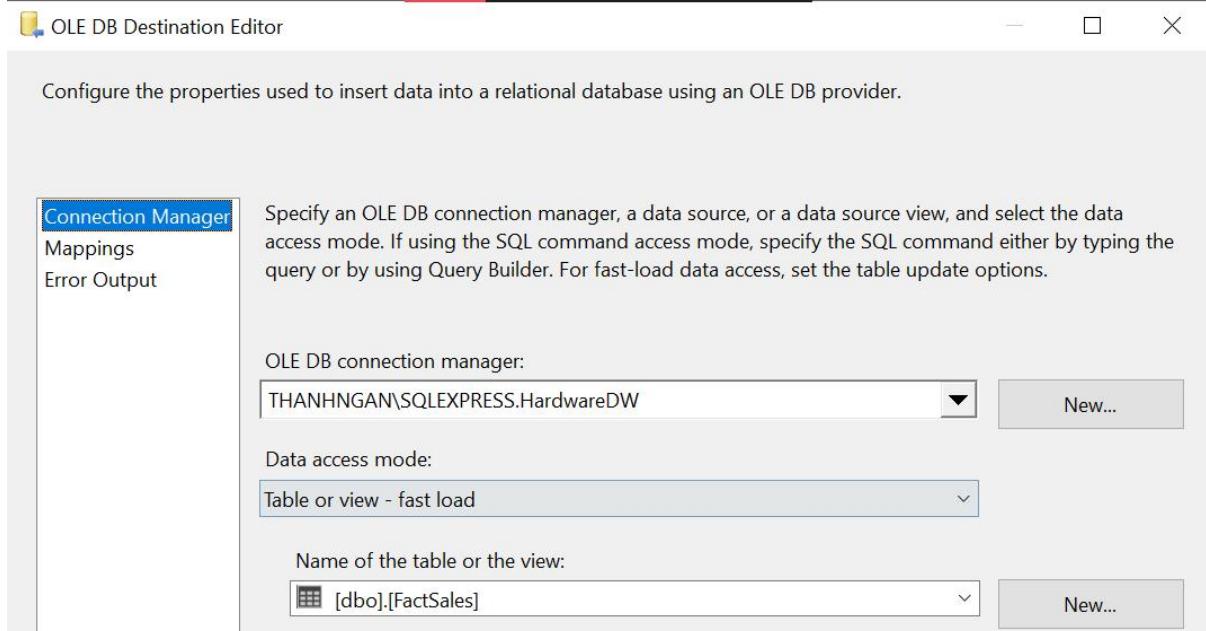
- Mapping hai bảng và lấy ra EmployeeKey



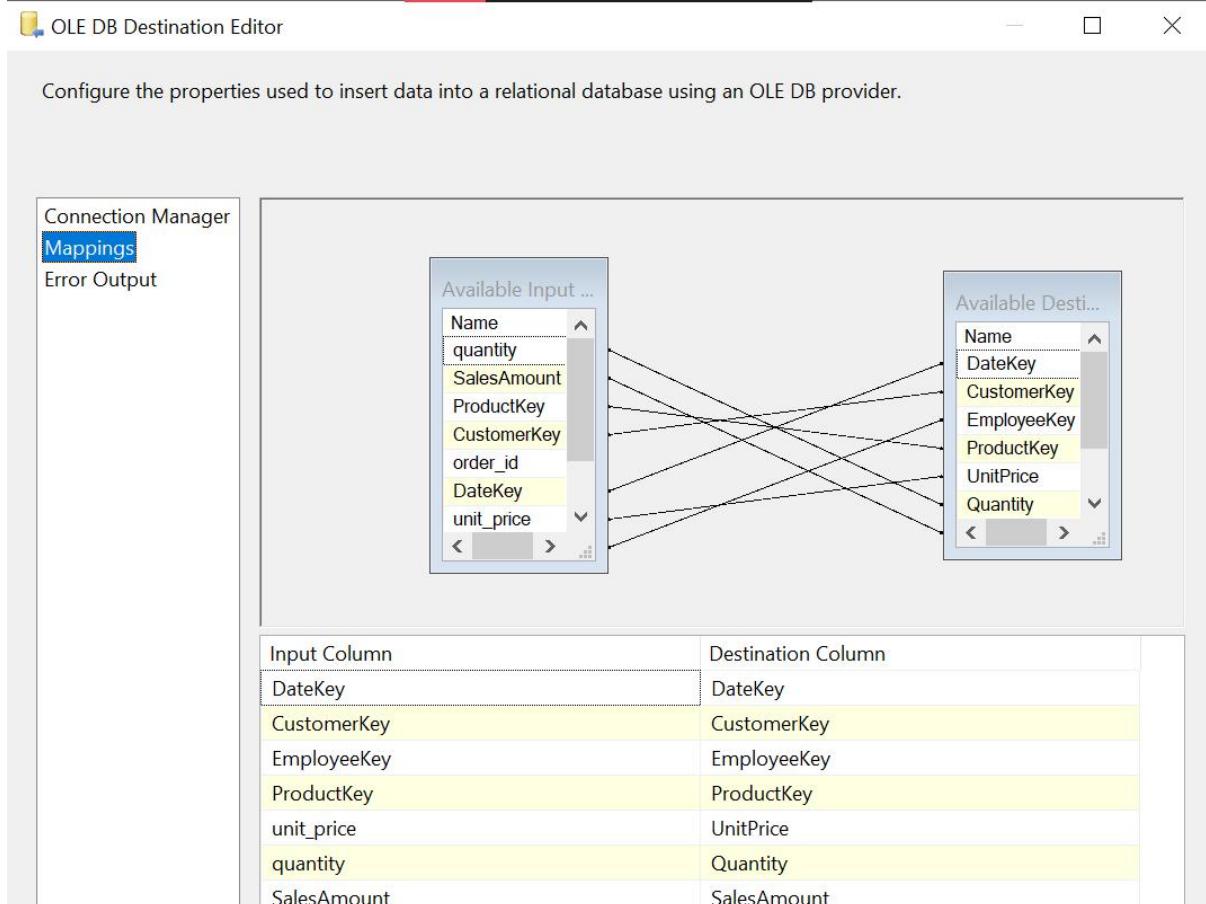
- Cấu hình Aggregate chọn các thuộc tính cần thiết và các measures



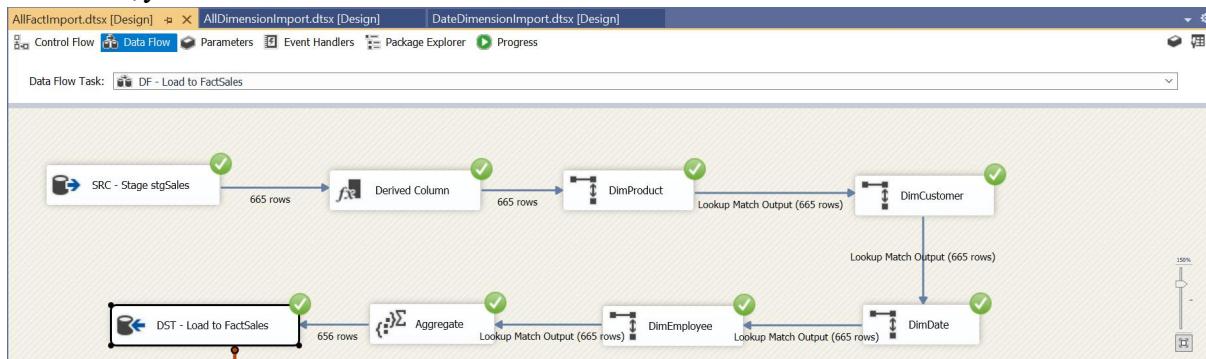
- Cấu hình Destination chọn bảng FactSales



- Mapping các cột



- Khởi chạy



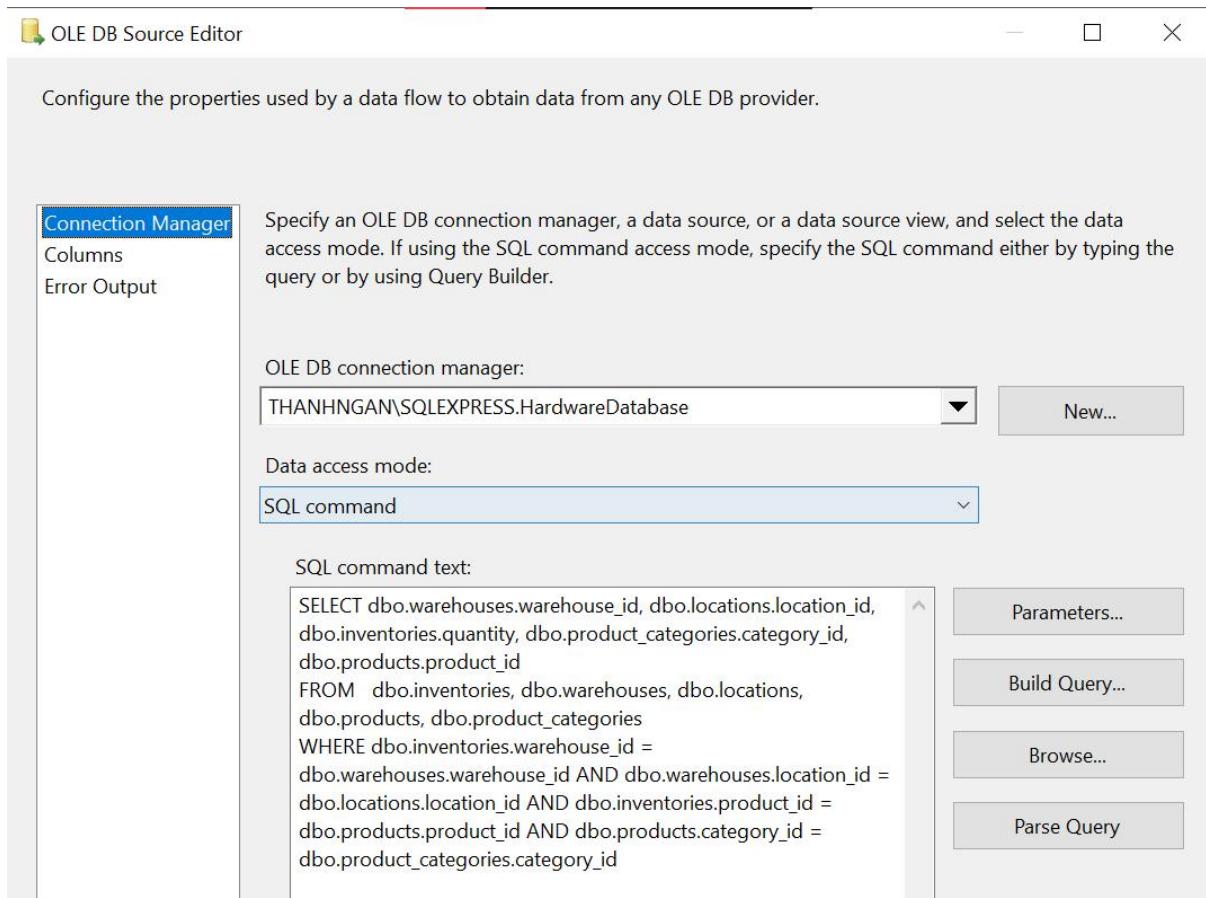
3.4.2. Fact Product Inventory

3.4.2.1. Load dữ liệu từ Source vào Stage

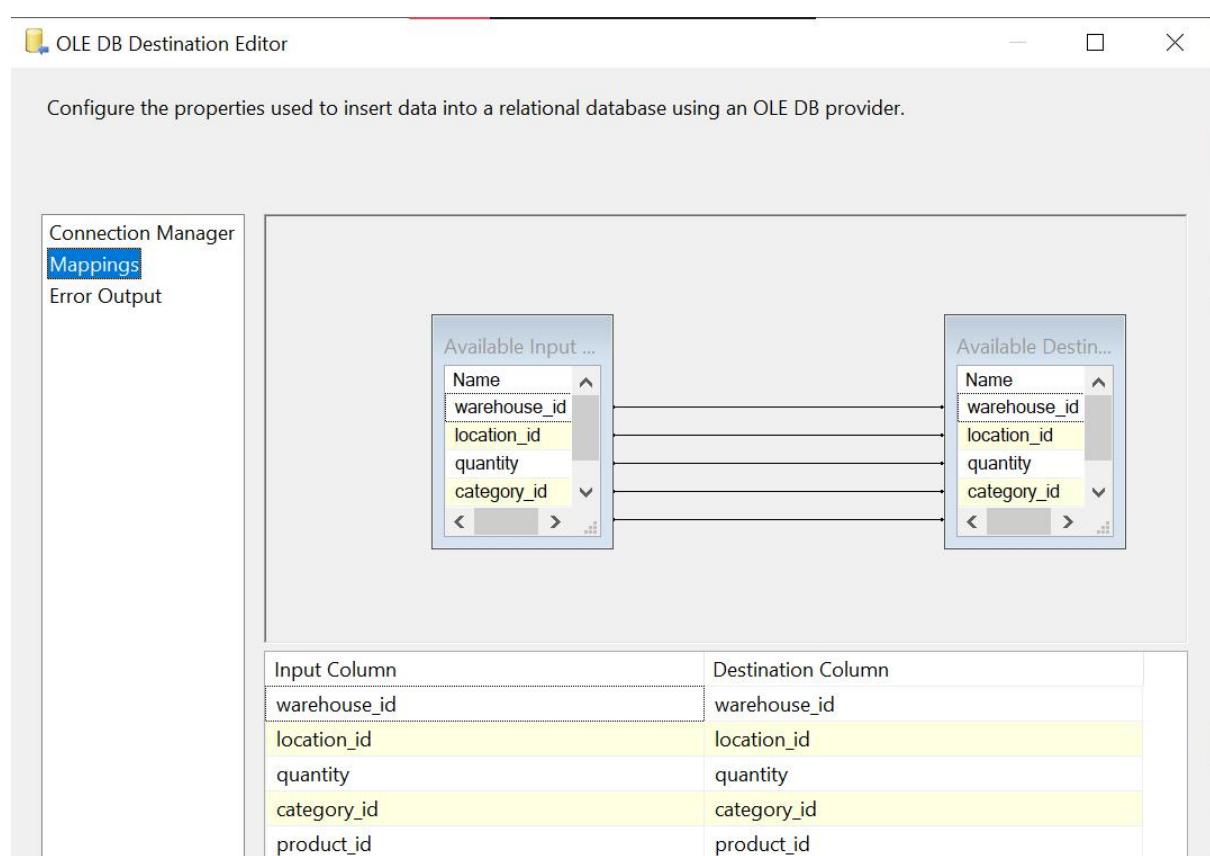
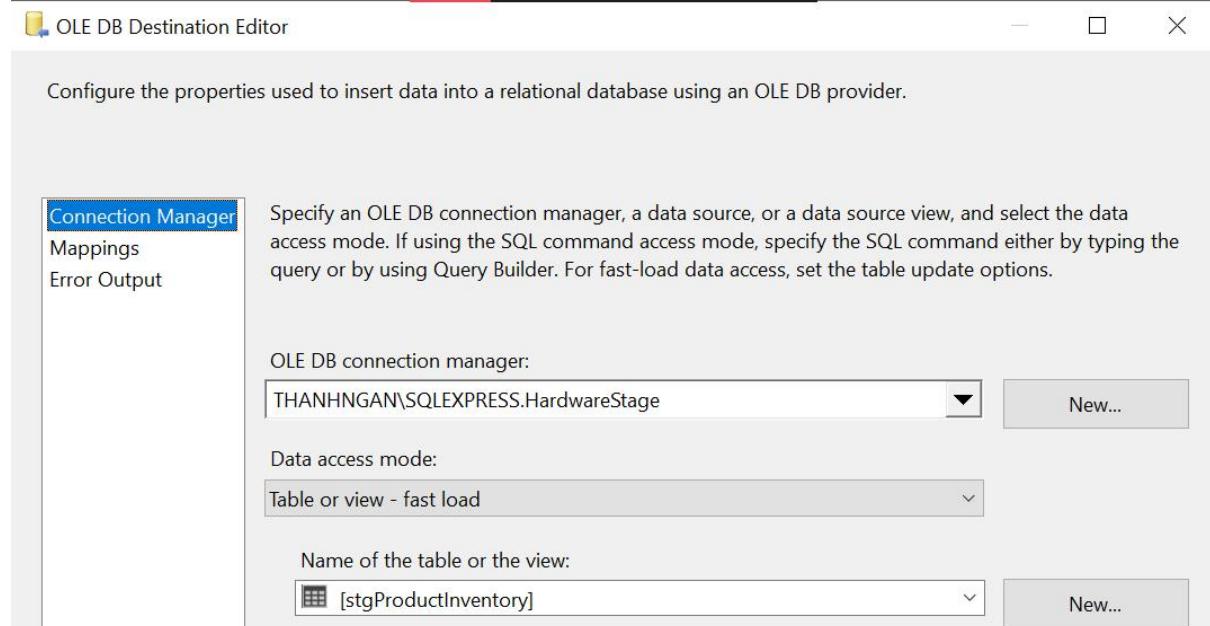
- Double-click DF - Stage Product Inventory task để mở nó trong Data Flow design surface:

- Tao Source Assistant và Destination Assistant và cấu hình chúng như sau:

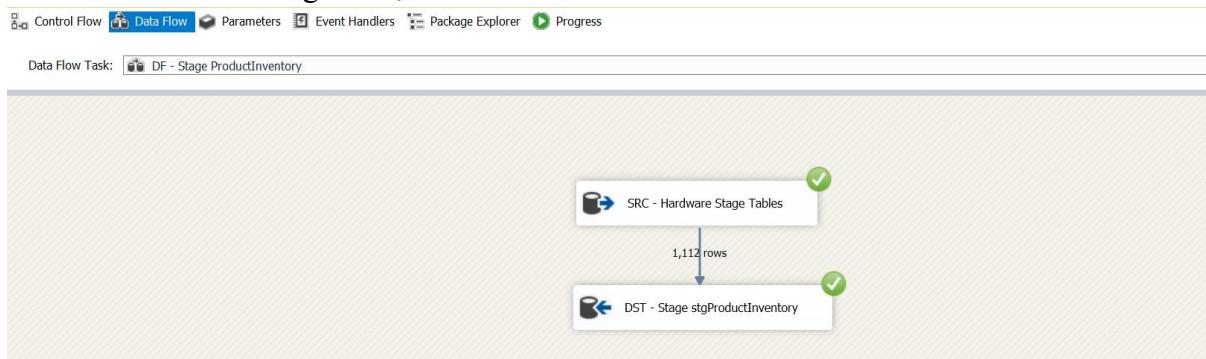
+ Source Assistant:



+ Destination Assist

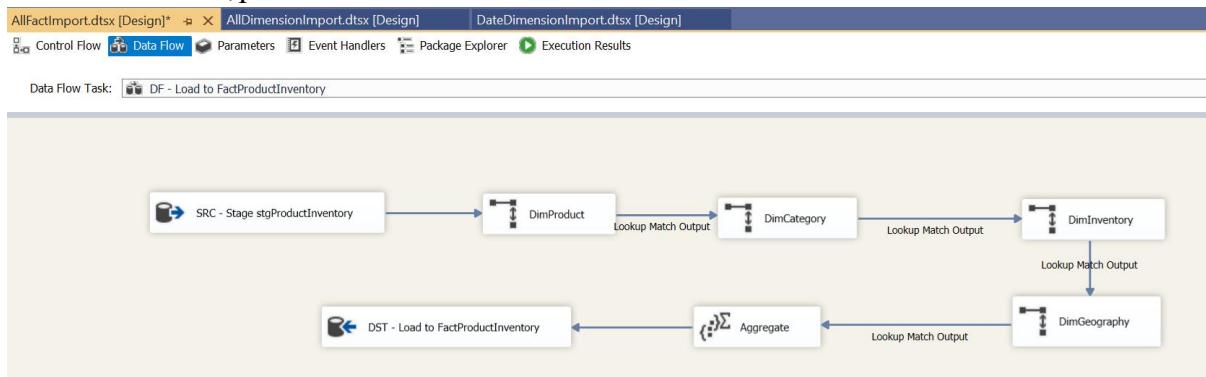


+ Sau khi cấu hình xong ta được:



3.4.1.4. Load dữ liệu từ Stage vào Fact

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

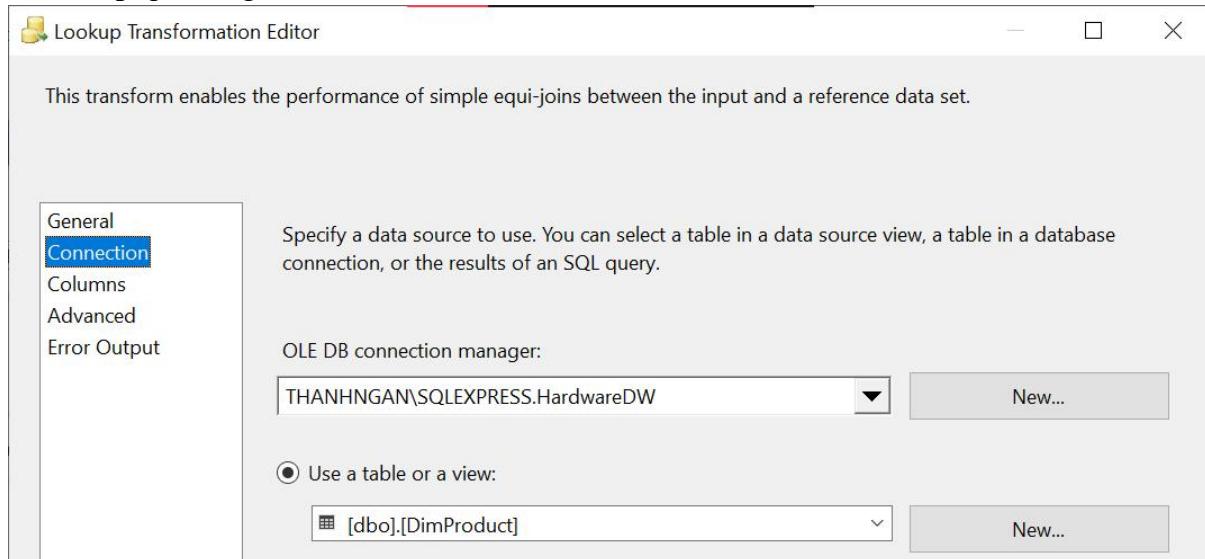


- Cấu hình Source chọn bảng **stgProductInventory**

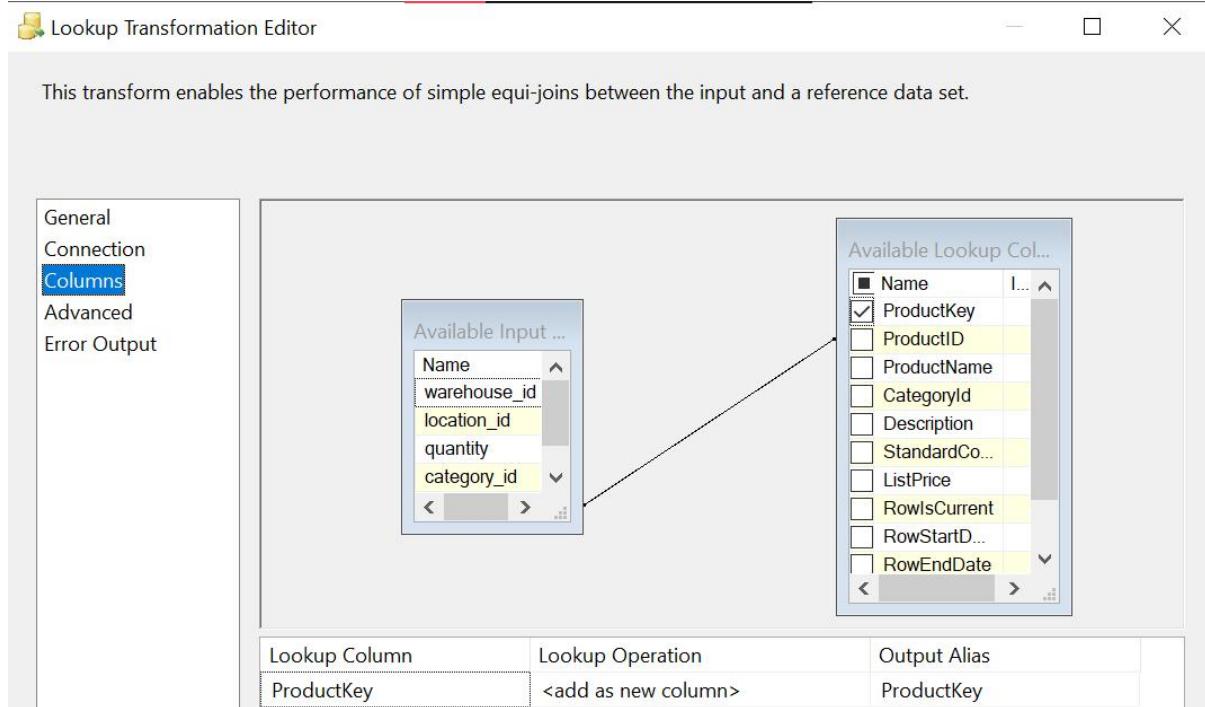
The screenshot shows the 'OLE DB Source Editor' configuration. On the left, there's a sidebar with tabs: 'Connection Manager' (selected), 'Columns', and 'Error Output'. The main area contains the following settings:

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

- Lookup qua bảng DimProduct



- Thiết lập mapping và lấy ra 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.

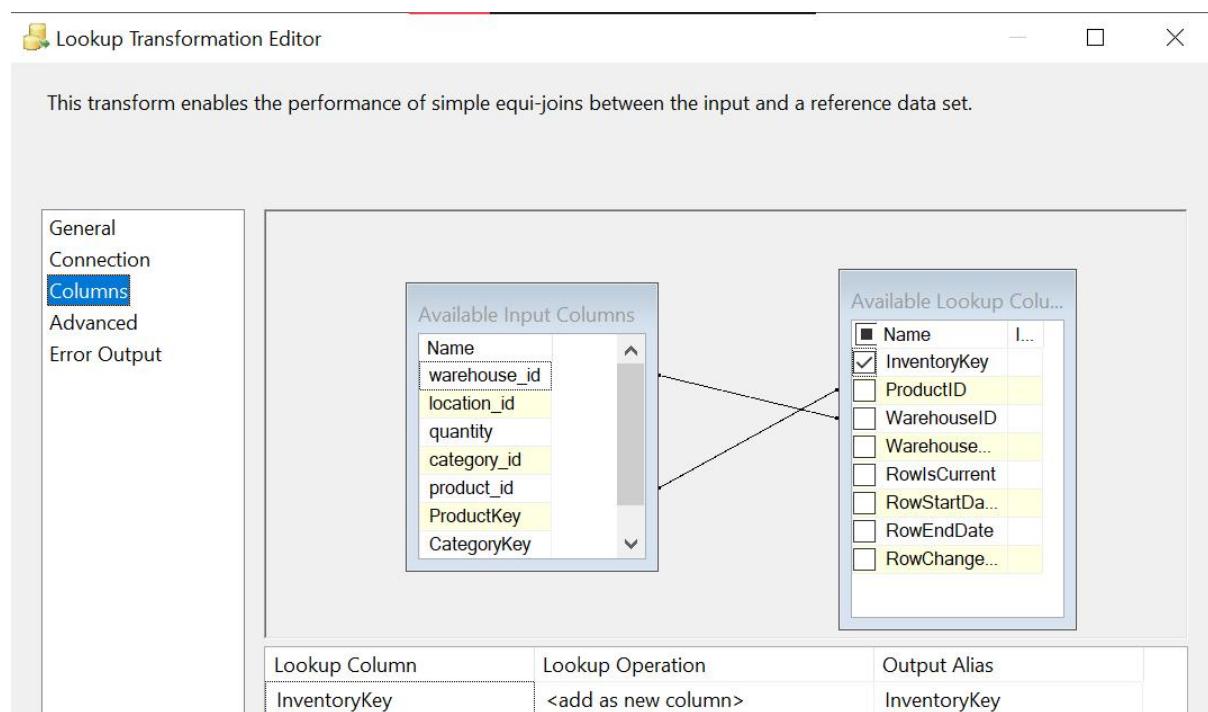
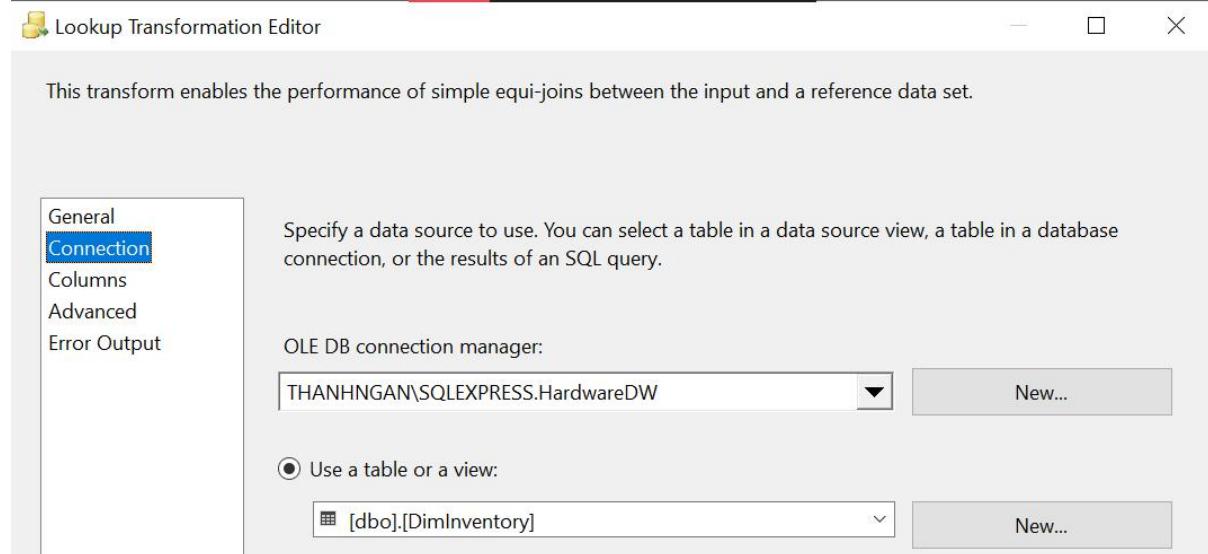
General
Columns
Advanced
Error Output

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 DimInventory



- Lookup qua bảng DimGeography

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].[DimGeography]

- Mapping hai bảng và lấy ra GeographyKey

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
ProductKey
CategoryKey

Available Lookup Colu...
Name
GeographyKey
LocationID
Region
CountryCode
CountryName
State
City
PostalCode
RowIsCurrent
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 Col...

Input Column	Output Alias	Operation	Compa
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:

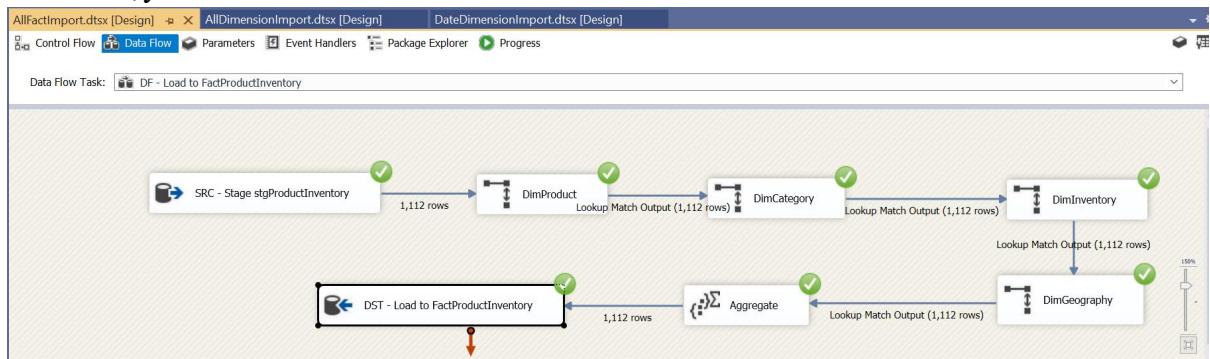
Table or view - fast load

Name of the table or the view:

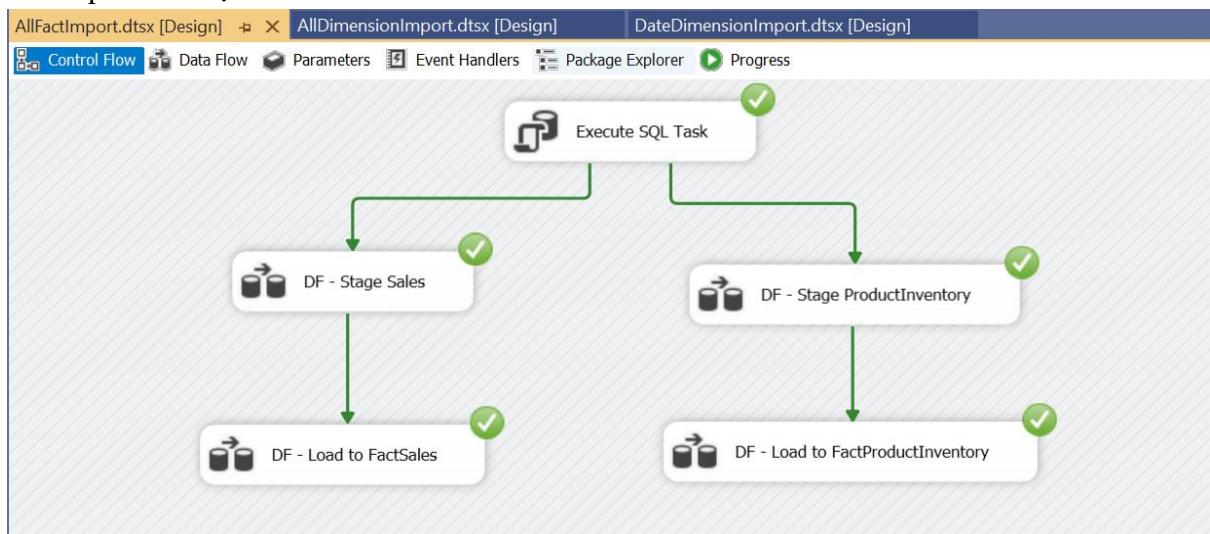
[dbo].[FactProductInventory]

New...

- Khởi chạy



- Kết quả thu được

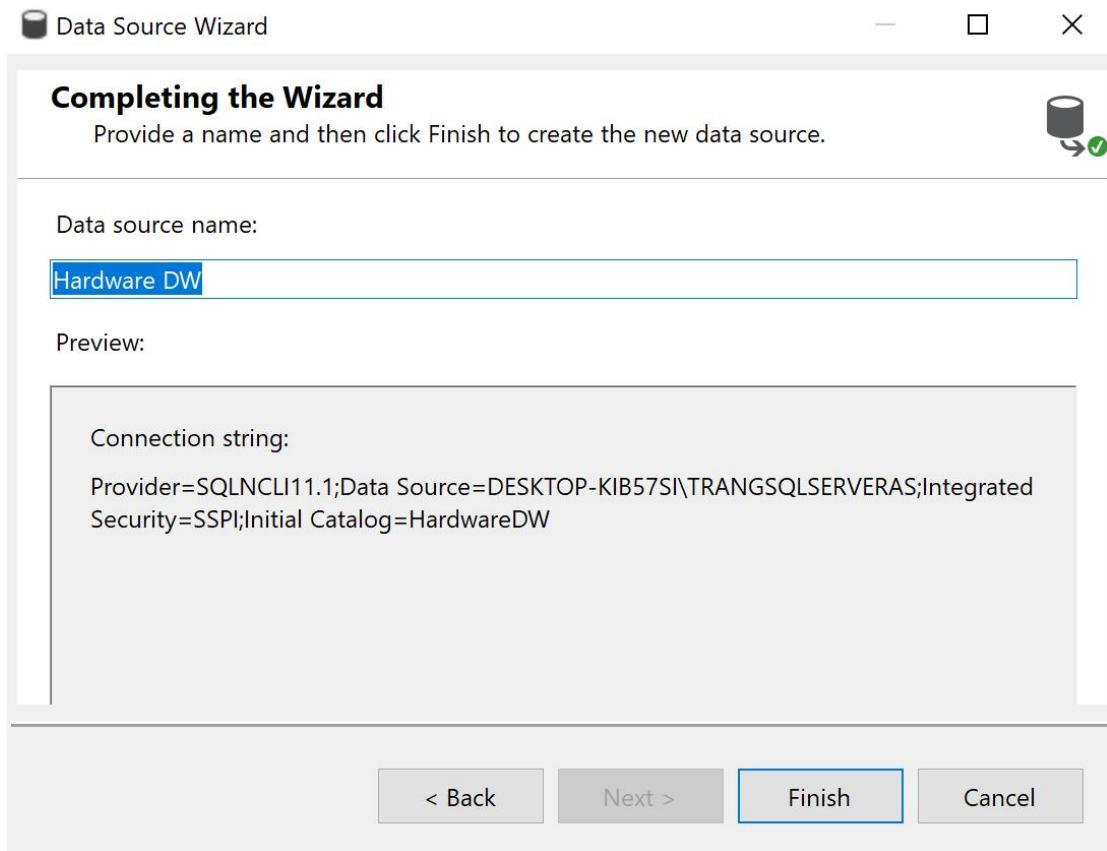


CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)

4.1. Quá trình xây dựng mô hình

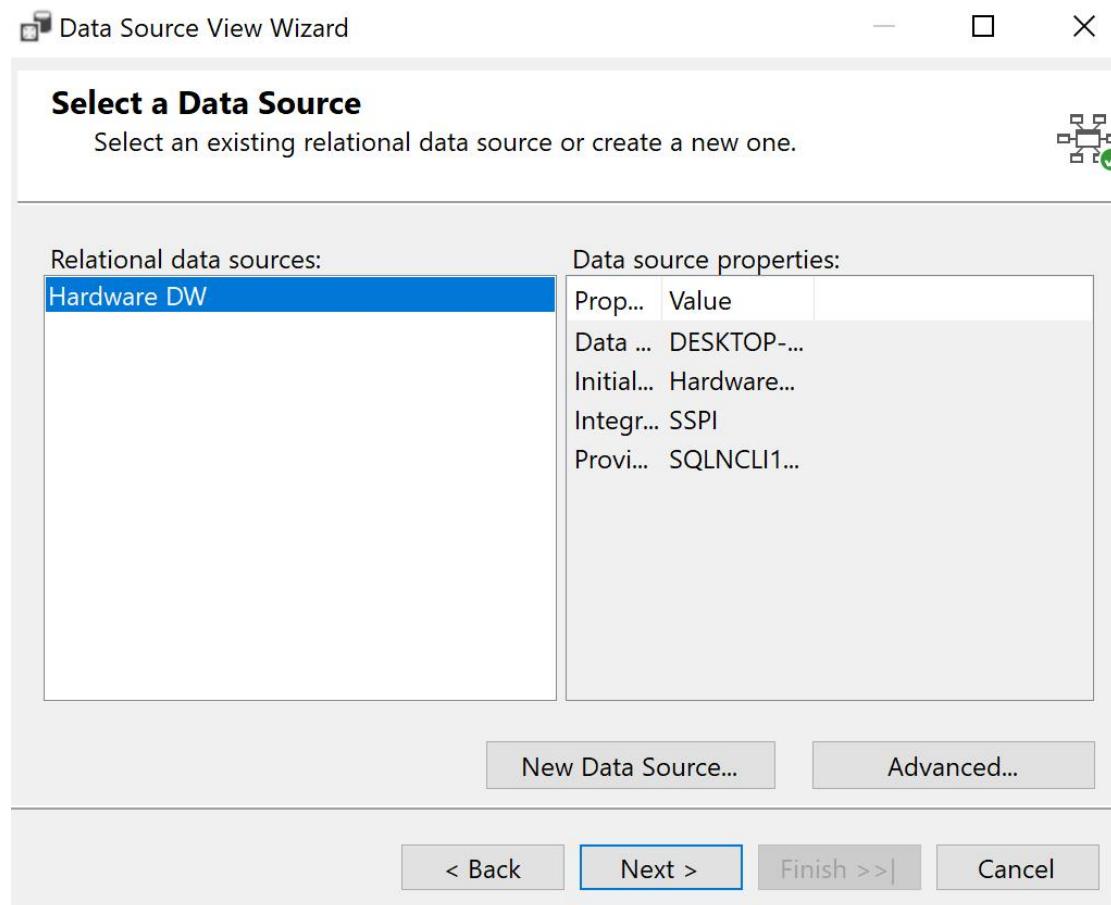
4.1.1. Tạo Data Source

- Right-click vào Data Sources ở khung Solution Explorer. Chon New Data Source để tạo Data Source mới. Kết nối đến Database **HardwareDW**

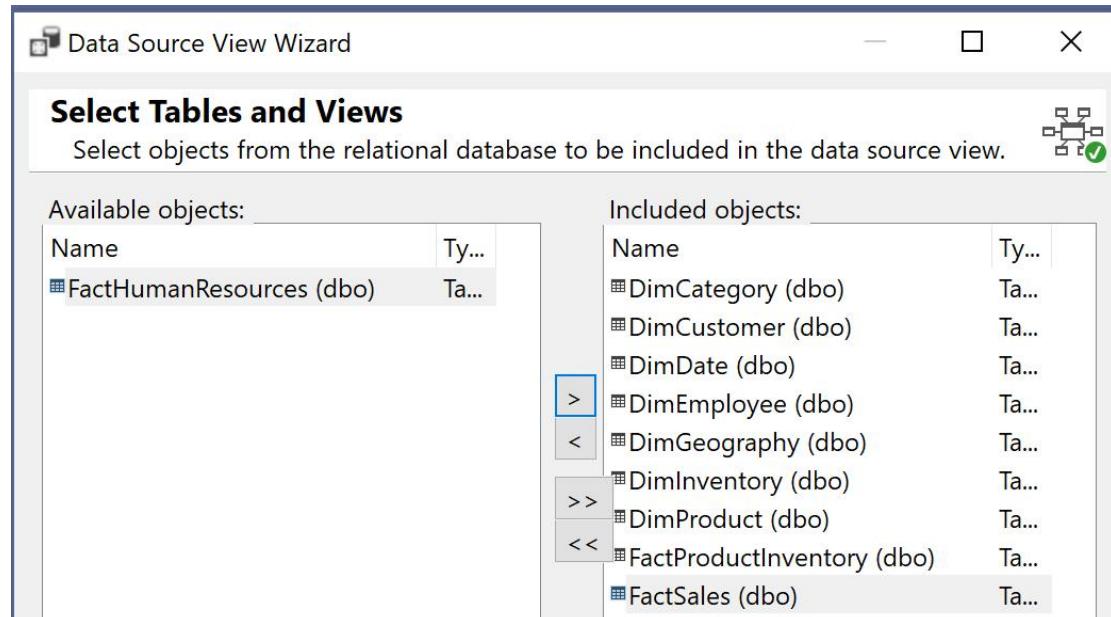


4.1.2. Tạo Data Source View

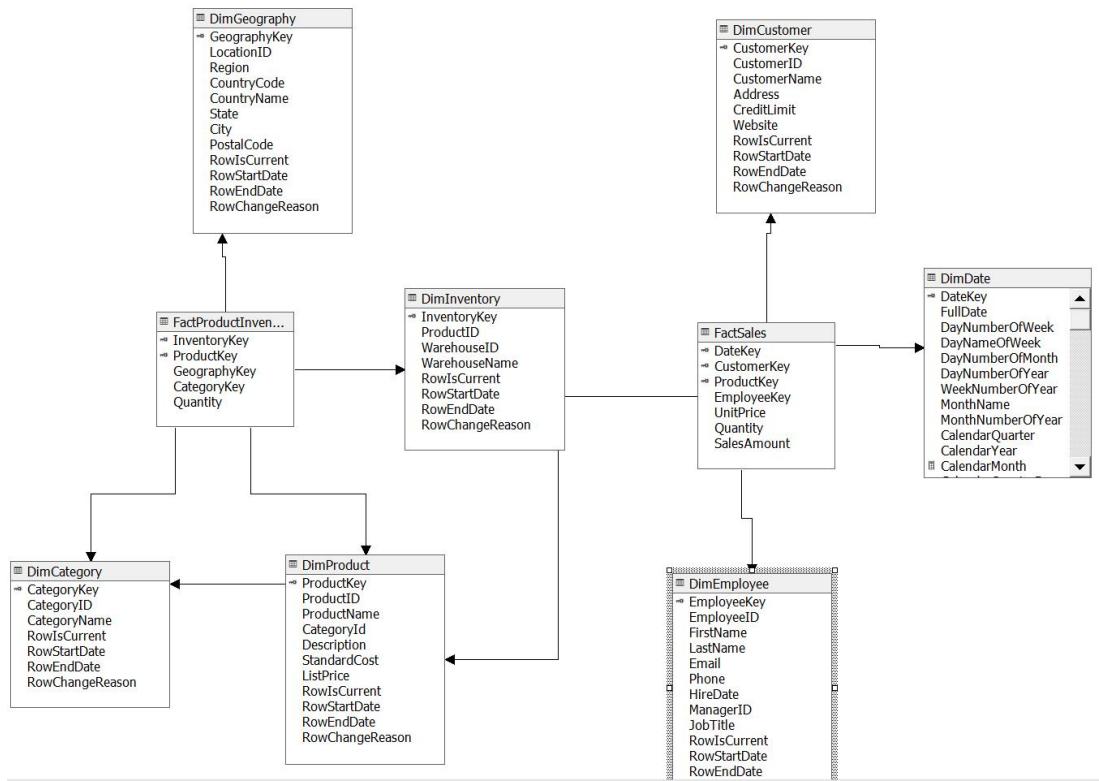
- Right-click vào Data Source Views ở khung Solution Explorer. Chon New Data Source View để tạo Data Source View mới. Chọn Data Soure là **Hardware DW** và chọn **Next**.



- Trong mục **Included objects**, chọn các bảng Dim và Fact cần thiết, nhấn Next

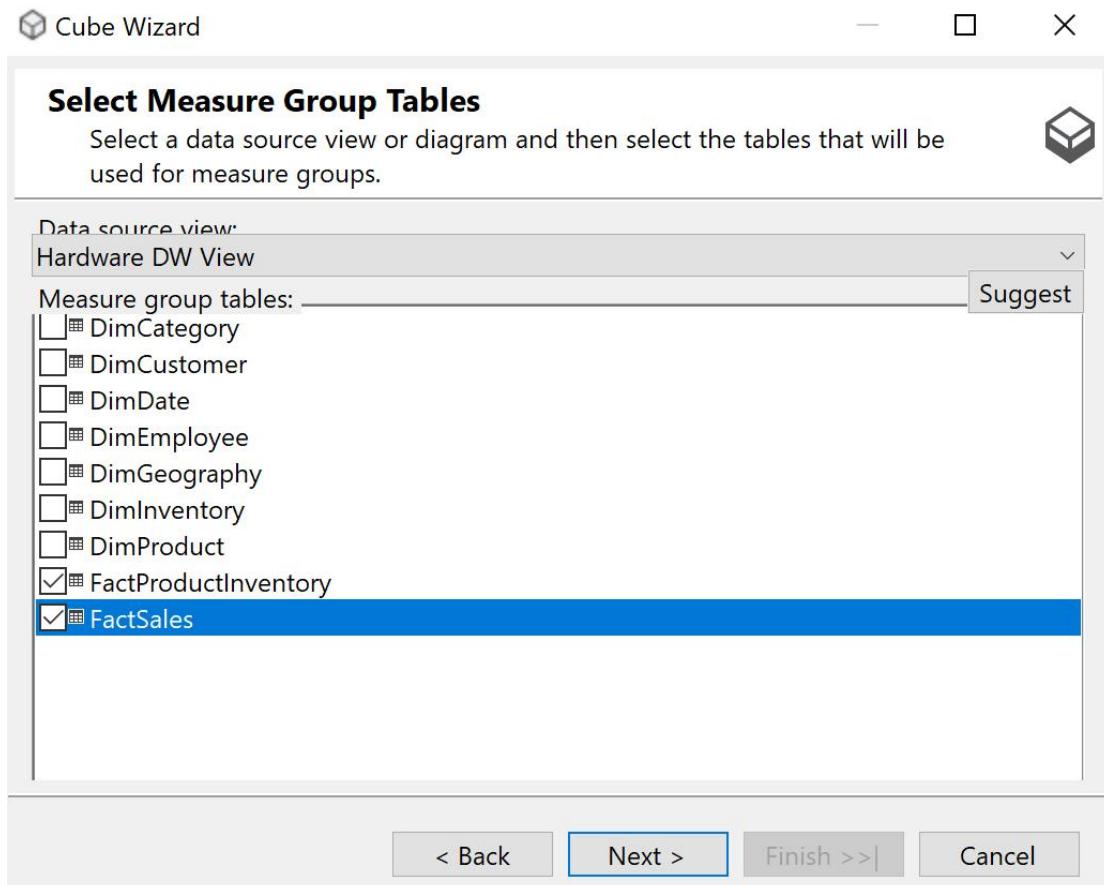


- Đặt tên cho Data Source View và nhấn Finish để hoàn tất việc tạo Data Source View
- Double-Click vào Data Source View vừa tạo để hiển thị diagram

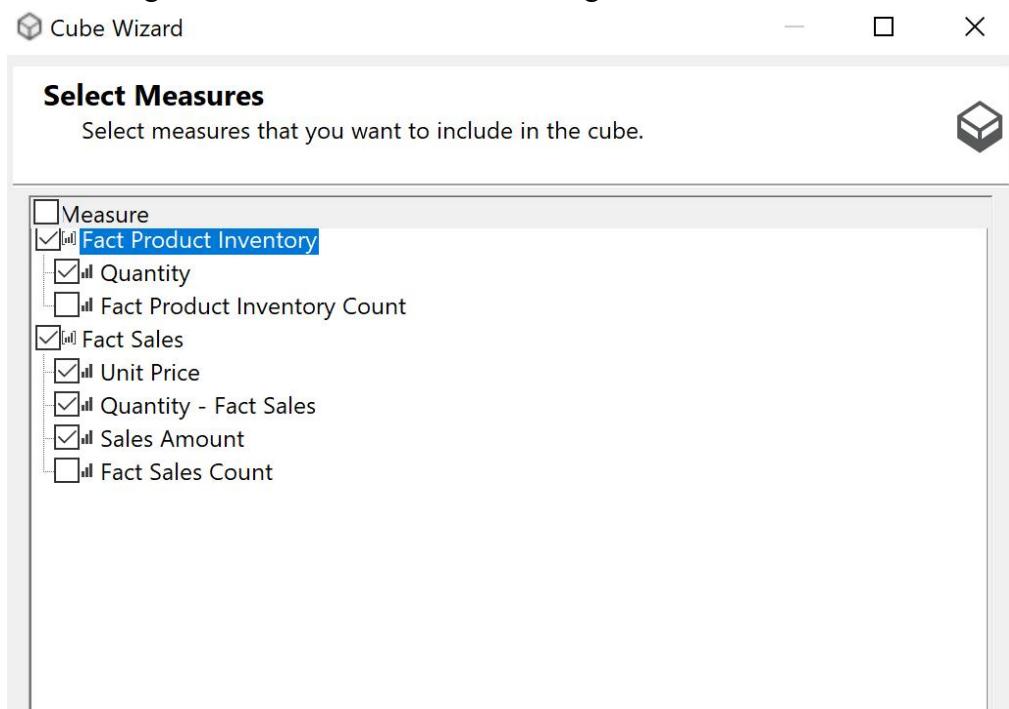


4.2. Quá trình xây dựng khối Cube

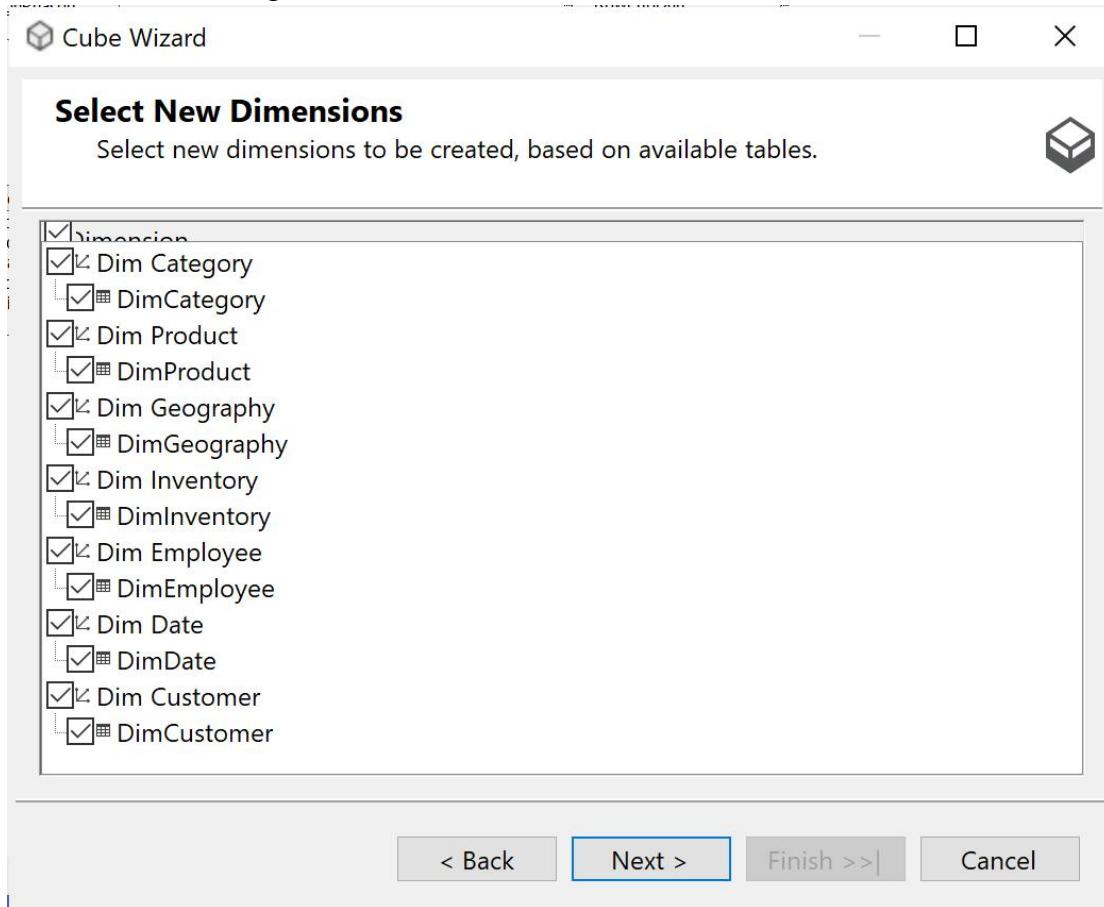
- Right-click vào **Cubes** ở khung Solution Explorer. Chon New Cube để tạo Cube mới.
- Chọn Data Source View là **Hardware DW View**. Tích chọn 2 bảng **FactSales** và **FactProductInventory**, nhấn Next



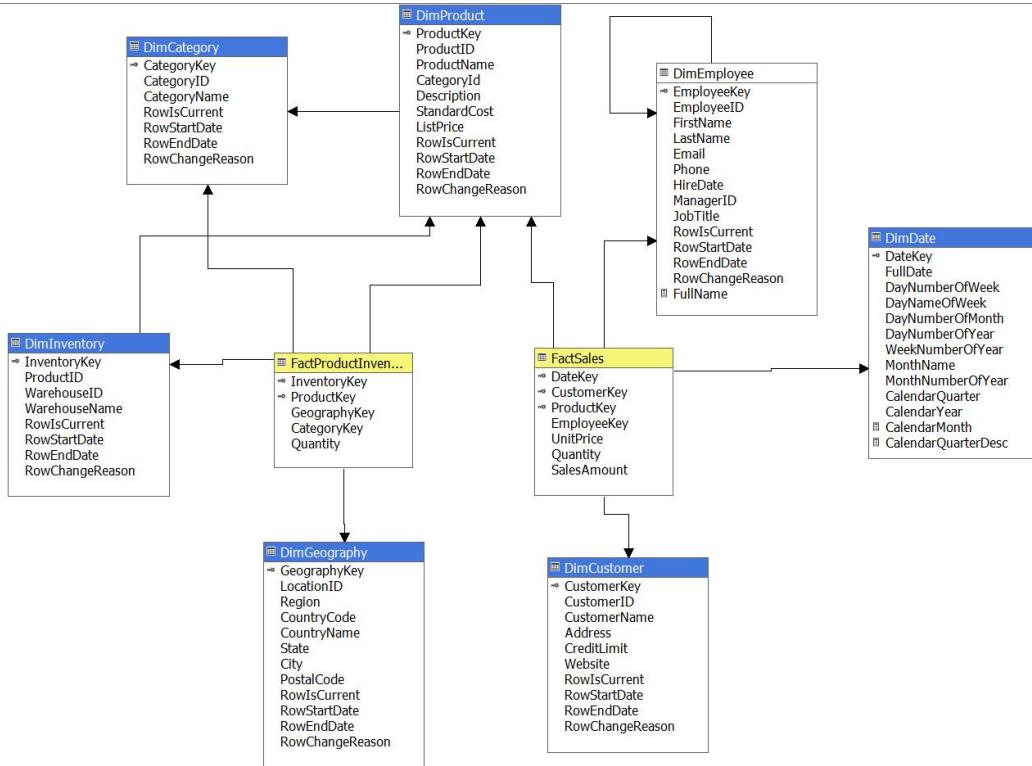
- Ở dialog Select Measures, tích chọn những measure cần thiết các mục sau



- Khởi tạo các bảng Dimension



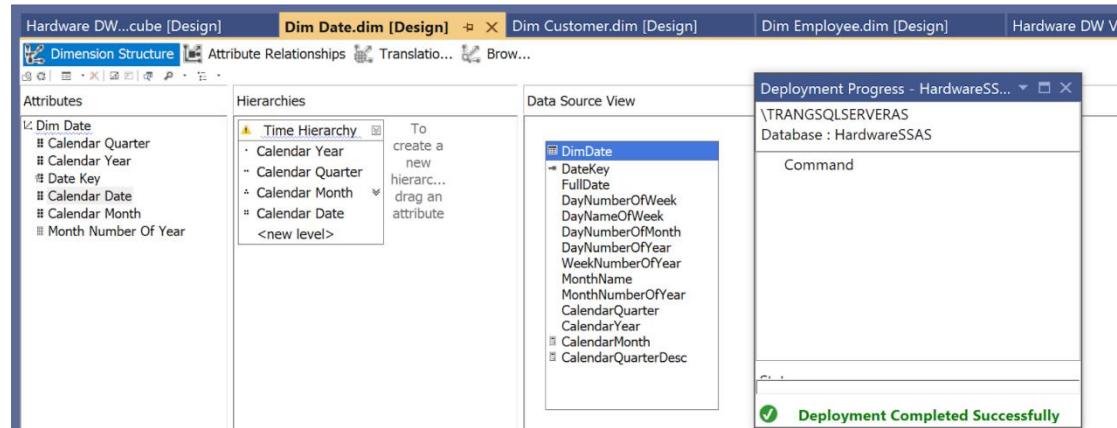
- Đặt tên **Hardware DW Cube** để hoàn tất khởi tạo.
- Các bảng Fact, Dimension và quan hệ giữa chúng sẽ được thể hiện trong Diagram



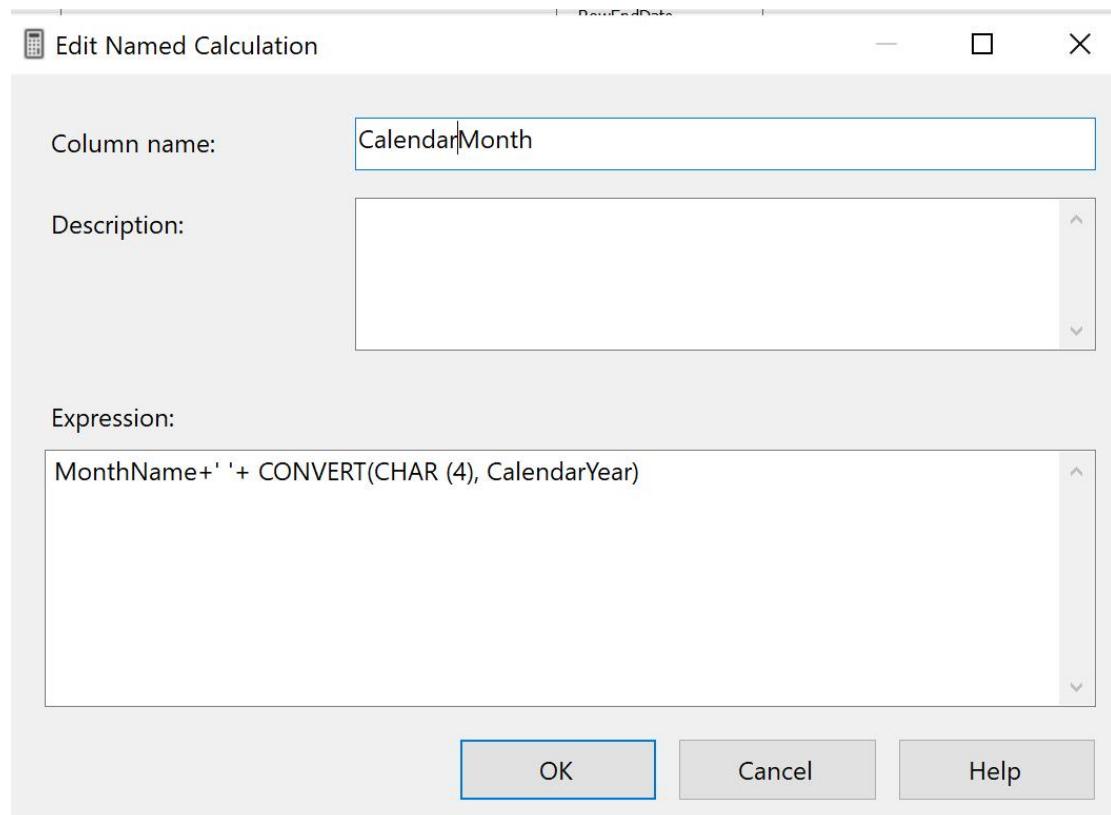
4.3. Cấu hình Hierarchy

4.3.1. Tạo Hierarchy cho Dim Date

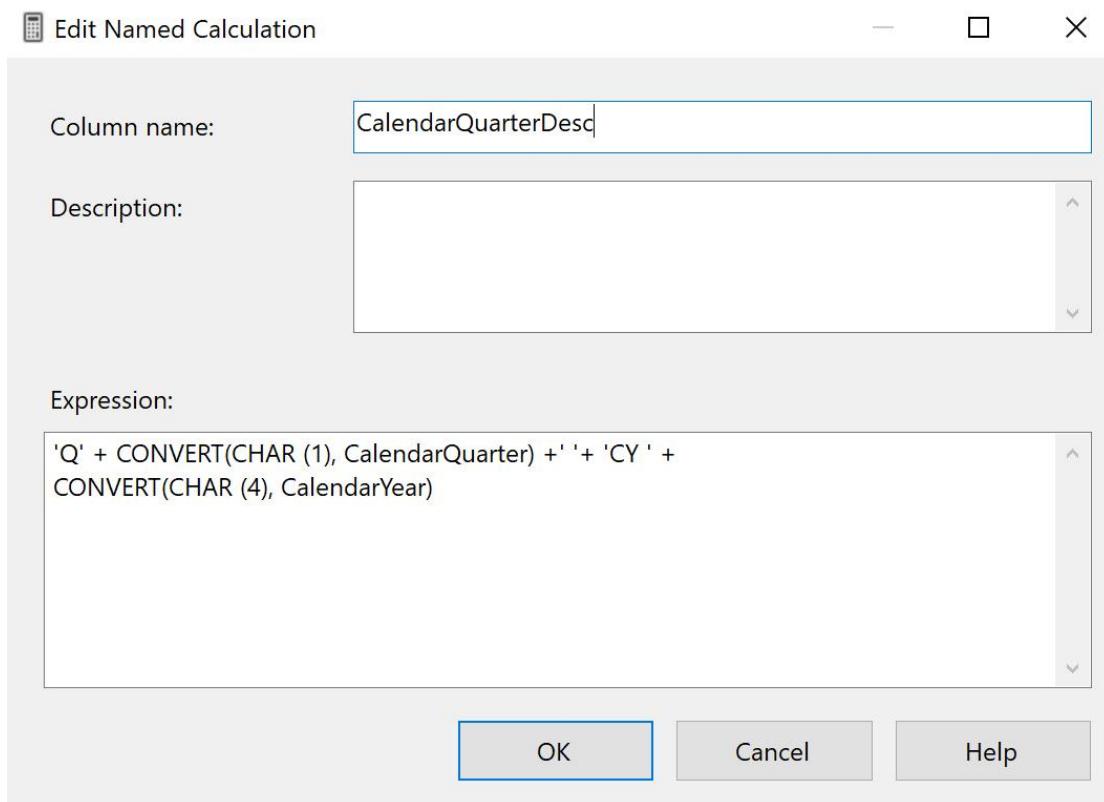
- Thêm thuộc tính vào bảng Dim: Tại table DimDate trong panel Data Source View, kéo thuộc tính cần thiết vào panel Attributes.
- Kéo thả 4 thuộc tính *Calendar Year*, *Calendar Quarter*, *Month Name*, *Date Key* từ Attributes vào panel Hierarchies để tạo ra Time Hierarchy
- Đặt lại tên cho các thuộc tính trong Time Hierarchy



- Tạo Caculator Đặt lại tên tháng kèm theo năm cho các giá trị trong thuộc tính MonthName và CalendarMonth



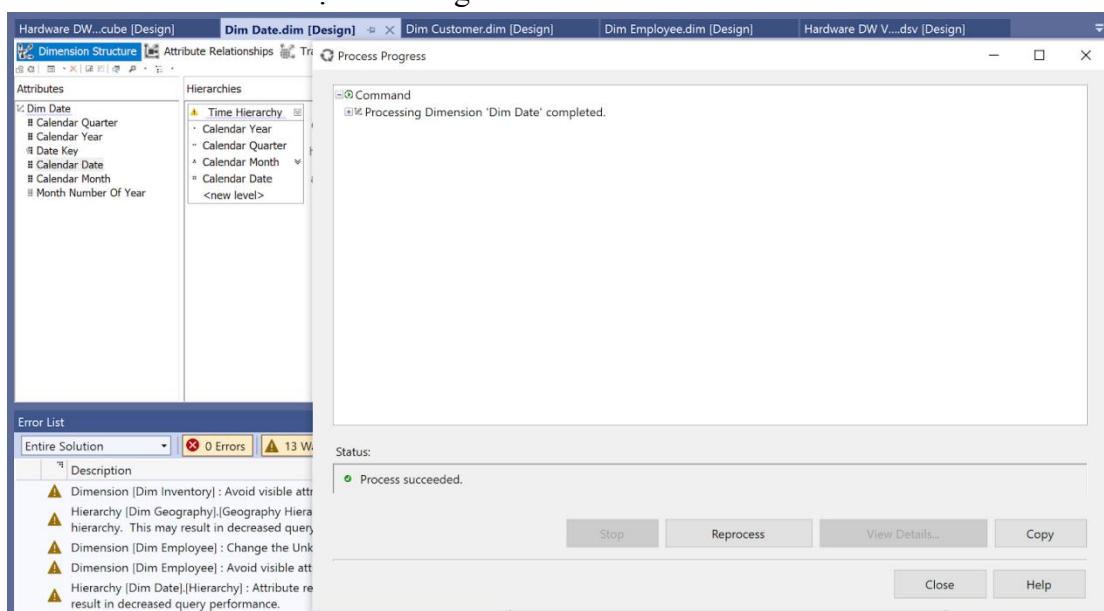
- Tạo thuộc tính mới CalendarQuaterDesc thể hiện quý trong năm được sắp xếp tăng dần



- Tạo Attribute Relationship



- Tạo Composite KeyColumn và đặt Column Name cho các thuộc tính *Calendar Quarter*, *Calendar Month*.
- Process để load dữ liệu cho bảng Dim Date



- Chuyển đến tab Browser. Ta thấy thời gian được phân cấp theo năm → quý → tháng.

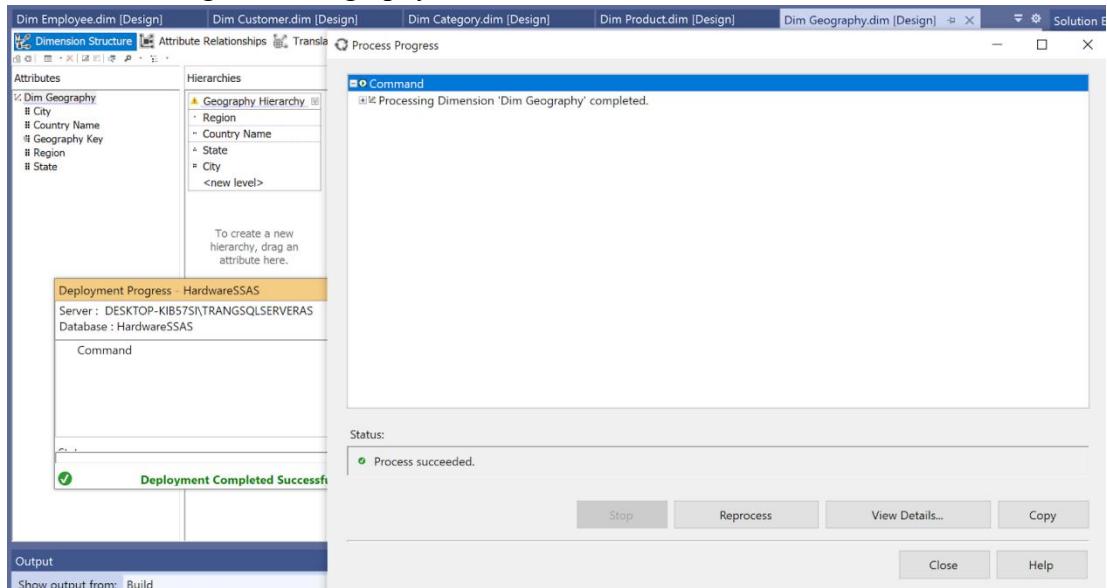
The screenshot shows the 'Dim Date.dim [Design]' tab selected in the top navigation bar. Below it, the 'Dimension Structure' tab is active. The hierarchy tree on the left shows levels from 'All' down to '2020'. The 'Hierarchies' section in the center displays a 'Geography Hierarchy' with levels: Region, Country Name, State, City, and <new level>. The 'Data Source View' section on the right lists attributes from the Dim Geography table, including GeographyKey, LocationID, Region, CountryCode, CountryName, State, City, PostalCode, RowIsCurrent, RowStartDate, RowEndDate, and RowChangeReason.

4.3.2. Tạo Hierarchy cho Dim Geography

- Tại table Dim Geography trong panel Data Source View, kéo 4 thuộc tính như hình vào panel Attributes. Sau đó kéo thả lần lượt Region, State, Country Name, City vào Hierarchies để tạo Geography Hierarchy

This screenshot shows the 'Dim Geography.dim [Design]' tab selected. In the 'Attributes' pane, 'Dim Geography' is expanded, showing attributes like City, Country Name, Geography Key, Postal Code, Region, and State. In the 'Hierarchies' pane, a 'Geography Hierarchy' is selected, displaying a tree structure with levels: Region, Country Name, State, City, and <new level>. A note below says 'To create a new hierarchy, drag an attribute here.' In the 'Data Source View' pane, the 'Dim Geography' table is shown with its columns: GeographyKey, LocationID, Region, CountryCode, CountryName, State, City, PostalCode, RowIsCurrent, RowStartDate, RowEndDate, and RowChangeReason.

- Process bảng Dim Geography



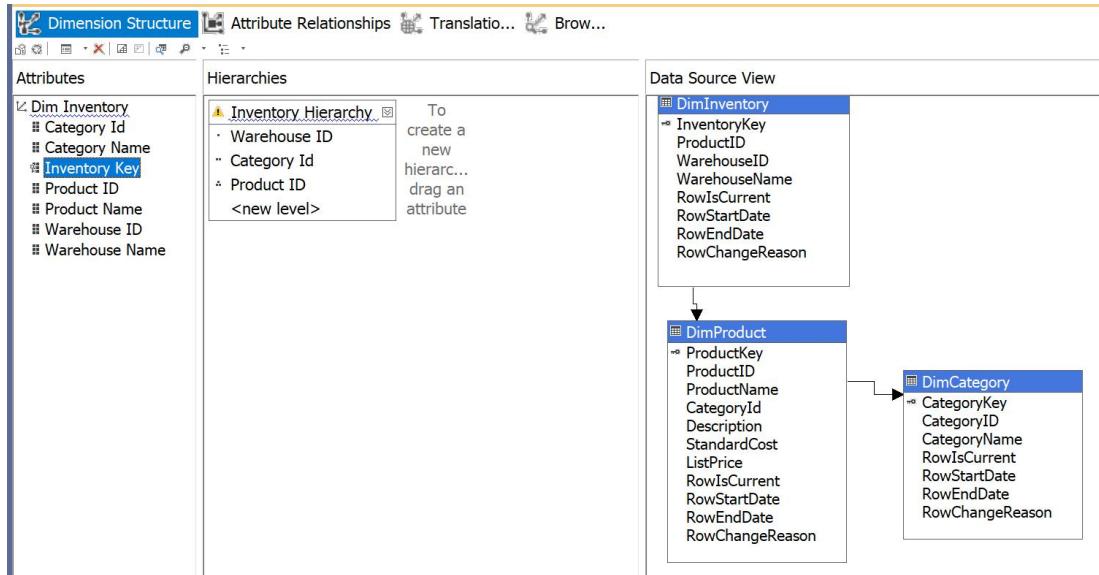
- Chuyển đến tab Browers. Vị trí được sắp xếp theo khu vực → quốc gia → thành phố

The screenshot shows the 'Dimension Structure' browser for the 'Geography' hierarchy. The tree view displays the following levels:

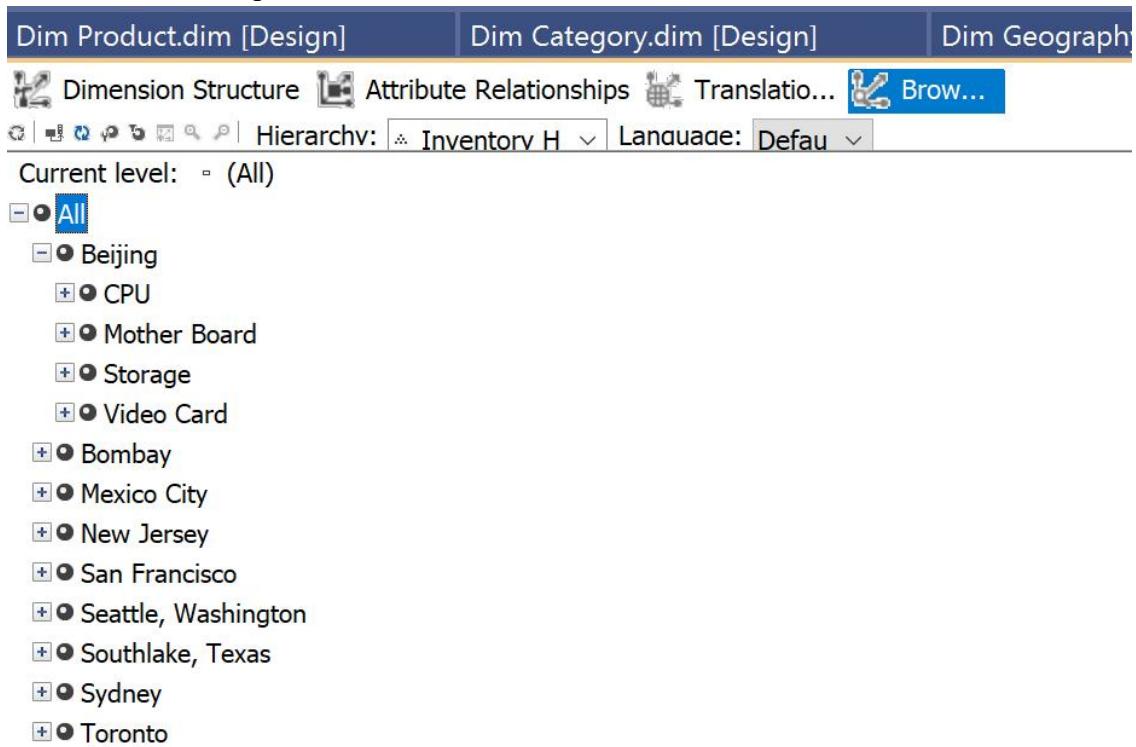
- All** (selected)
- Americas**
 - Brazil**
 - Sao Paulo**
 - Canada**
 - Ontario**
 - Yukon**
 - Mexico**
 - Distrito Federal,**
 - United States of America**
 - California**
 - New Jersey**
 - Texas**
 - Washington**
- Asia**
 - Australia**
 - China**
 - India**
 - Japan**
 - Singapore**
- Europe**

4.3.3. Tạo Hierarchy cho Dim Inventory

- Tạo Inventory Hierarchy



- Chuyển qua tab Browser sẽ thấy các sản phẩm được phân loại theo thứ tự sau: kho hàng - danh mục - sản phẩm.



4.3.4. Tạo Hierarchy cho Dim Product

- Tạo Product Category Hierarchy

Dim Product.dim [Design] * Dim Date.dim [Design] Dim Customer.dim [Design]

Dimension Structure Attribute Relationships Translation... Brow...

Attributes

- Dim Product
 - # Category ID
 - # Category Name
 - # List Price
 - # Product ID
 - # Product Key
 - # Product Name
 - # Standard Cost

Hierarchies

Product Category Hierarchy

- Category Name
- Product Name
- <new level>

To create a new hierarchy, drag an attribute here.

Data Source View

```

graph TD
    DimCategory[DimCategory] --> DimProduct[DimProduct]
    DimCategory[DimCategory]
    DimCategory[DimCategory] --> CategoryKey[CategoryKey]
    DimCategory[DimCategory] --> CategoryID[CategoryID]
    DimCategory[DimCategory] --> CategoryName[CategoryName]
    DimCategory[DimCategory] --> RowIsCurrent[RowIsCurrent]
    DimCategory[DimCategory] --> RowStartDate[RowStartDate]
    DimCategory[DimCategory] --> RowEndDate[RowEndDate]
    DimCategory[DimCategory] --> RowChangeReason[RowChangeReason]
    DimProduct[DimProduct]
    DimProduct[DimProduct] --> ProductKey[ProductKey]
    DimProduct[DimProduct] --> ProductID[ProductID]
    DimProduct[DimProduct] --> ProductName[ProductName]
    DimProduct[DimProduct] --> CategoryId[CategoryId]
    DimProduct[DimProduct] --> Description[Description]
    DimProduct[DimProduct] --> StandardCost[StandardCost]
    DimProduct[DimProduct] --> ListPrice[ListPrice]
    DimProduct[DimProduct] --> RowIsCurrent[RowIsCurrent]
    DimProduct[DimProduct] --> RowStartDate[RowStartDate]
    DimProduct[DimProduct] --> RowEndDate[RowEndDate]
    DimProduct[DimProduct] --> RowChangeReason[RowChangeReason]
  
```

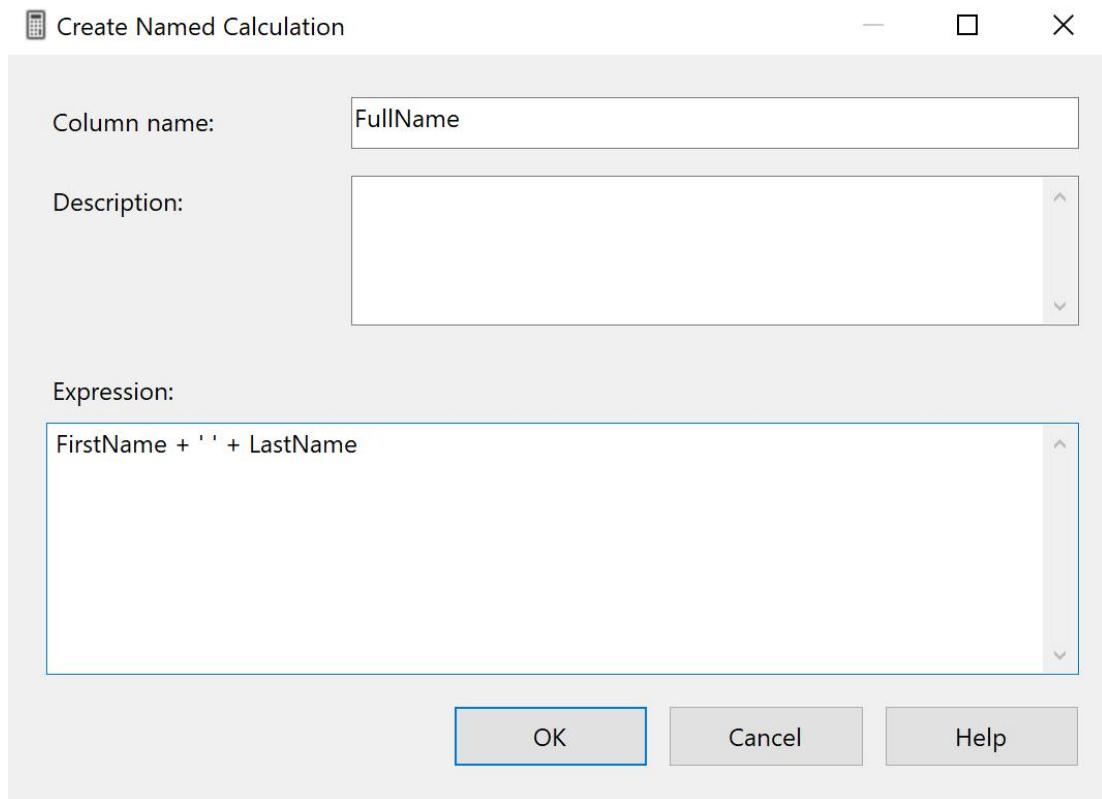
- Chuyển qua tab Browser sẽ thấy các sản phẩm được phân loại theo 4 mặt hàng chính.

The screenshot shows the 'Dimension Structure' tab of a data modeling application. At the top, there are tabs for 'Dim Employee.dim [Design]', 'Hardware DW...cube [Design]', and 'Dim Category.dim [Des]'. Below the tabs is a toolbar with icons for dimension structure, attribute relationships, translation, browser, and search. A dropdown menu for 'Hierarchv:' is set to 'Product Cat' and a language dropdown is set to 'Defau'. The main area displays a hierarchical tree under 'Current level: (All)'. The tree structure is as follows:

- All**
 - CPU**
 - Mother Board**
 - RAM**
 - Storage**
 - Video Card**
 - AMD**
 - 100-5056062
 - 100-505989
 - 100-506061
 - AMD FirePro S7000
 - AMD FirePro W9100
 - Asus GTX780TI-3GD5
 - Asus ROG-POSEIDON-GTX1080TI-P11G-GAMING
 - Asus STRIX-GTX1080TI-O11G-GAMING
 - ATI FirePro R5000
 - ATI FirePro S9050
 - ATI FirePro S9150
 - ATI FirePro W9000
 - Corsair CB-9060011-WW
 - EVGA 06G-P4-4998-KR
 - EVGA 11G-P4-6598-KR
 - EVGA 11G-P4-6696-KR
 - EVGA 12G-P4-1999-KR
 - EVGA 12G-P4-3992-KR
 - Gigabyte GV-N1070WF2OC-8GD
 - Gigabyte GV-N108TAORUS X-11GD
 - Gigabyte GV-N108TAORUS-11GD
 - Gigabyte GV-N108TAORUSX W-11GD
 - Gigabyte GV-N98TWF3OC-6GD
 - HP C2J95AT

4.3.5. Tạo Hierarchy cho Dim Employee

- Tạo cột mới để nối họ và tên của nhân viên



- Tạo Parent-Child Hierarchy với Usage của thuộc tính ManagerID là Parent.

- Chuyển đến tab Browser để xem danh sách nhân viên theo người quản lý.

Dim Employee.dim [Design] Hardware DW...cube [Design]

Dimension Structure Attribute Relationships Translatio... Brow...

Hierarchy: Manager ID Language: Default

Current level: (All)

All

- Tommy Bailey
- Jude Rivera
 - Mohammad Peterson
 - Annabelle Dunn
 - Harper Spencer
 - Gracie Gardner
 - Rose Stephens
- Blake Cooper
 - Louie Richardson
- Rory Kelly
 - Alex Sanders
 - Frederick Price
 - Ollie Bennett
 - Louis Wood
 - Dexter Barnes
 - Jaxon Ross
 - Liam Henderson
 - Jackson Coleman
 - Callum Jenkins
 - Ronnie Perry
 - Ava Sullivan
 - Ella Wallace
 - Jessica Woods
 - Isabella Cole
 - Mia West
 - Emma Perkins

- Tạo Job Title Hierarchy

The screenshot shows the Microsoft Analysis Services Dimension Structure tool interface. The title bar displays three tabs: "Dim Employee.dim [Design]" (selected), "Hardware DW...cube [Design]", and "Dim Category.dim [Design]". The main area is divided into three sections: "Attributes" on the left, "Hierarchies" in the center, and "Data Source View" on the right.

Attributes: A list of attributes for the "Dim Employee" dimension, including Email, Employee ID, Employee Key, First Name, Full Name, Hire Date, Job Title, Last Name, Manager ID, and Phone.

Hierarchies: A section titled "Job Title Hierarchy" with a warning icon. It contains two levels: "Job Title" and "Employee ID". A tooltip indicates: "To create a new hierarchy... drag an attribute".

Data Source View: A list of columns from the "DimEmployee" table, including EmployeeKey, EmployeeID, FirstName, LastName, Email, Phone, HireDate, ManagerID, JobTitle, RowIsCurrent, RowStartDate, RowEndDate, RowChangeReason, and FullName.

- Chuyển đến tab browser để xem danh sách nhân viên theo vị trí làm việc

Dim Employee.dim [Design] **Hardware DW...cube [Design]** Dim Catego

Dimension Structure Attribute Relationships Translation Brow...

Hierarchy: Hierarchy Language: Default

Current level: (All)

- All
- Accountant
 - Ryan Gray
 - Tyler Ramirez
 - Elliott James
 - Albert Watson
 - Elliot Brooks
- Accounting Manager
 - Rose Stephens
- Administration Assistant
 - Annabelle Dunn
- Administration Vice President
 - Jude Rivera
 - Blake Cooper
- Finance Manager
- Human Resources Representative
- Marketing Manager
- Marketing Representative
- President
- Programmer
- Public Accountant
- Public Relations Representative
- Purchasing Clerk
- Purchasing Manager
- Sales Manager
- Sales Representative
- Shipping Clerk
- Stock Clerk
- Stock Manager

4.4. Thực hiện phân tích dữ liệu

4.4.1. Câu hỏi: Cho biết xếp hạng KPI bán hàng của nhân viên trong tháng/quý/năm.

4.4.1.1. Sử dụng công cụ SSAS

- Hiển thị doanh thu của các nhân viên bán hàng từ trước đến nay

Dimension	Hierarchy	Operator	Filter Expression
Dim Date	# Time Hierarchy	Equal	{ All }
<Select dimension>			

Employee ID	Sales Amount
Lily Fisher	1882890
Grace Ellis	3522039
Evie Harrison	2752265
Scarlett Gib...	3519494
Chloe Cruz	3897078
Isabelle Ma...	3231078
Daisy Ortiz	3249434
Freya Gomez	8074231
Florence Fr...	4337771
Unknown	18227090

- Có 9 nhân viên bán hàng còn làm việc đến thời điểm hiện tại. Và những nhân viên đã không còn làm việc thì sẽ được gán cho biến ‘Unknown’
- Hiển thị doanh thu theo năm 2017

Calendar Year	Employee ID	Sales Amount
2017	Lily Fisher	1159720
2017	Grace Ellis	1988899
2017	Evie Harrison	1334254
2017	Scarlett Gibson	1309537
2017	Chloe Cruz	276996
2017	Isabelle Marshall	2090620
2017	Daisy Ortiz	1585976
2017	Freya Gomez	3765893
2017	Florence Freeman	544271
2017	Unknown	7837954

- Ta thấy top 5 nhân viên có doanh số bán hàng cao nhất trong năm 2017 là Freya Gomez, Isabelle Marshall, Grace Ellis, Daisy, Ortiz, Evie Harrison.

- Bằng cách này, chủ doanh nghiệp dễ dàng chọn ra những nhân viên có thành tích tốt để thực hiện công tác khen thưởng cuối năm.

4.4.1.2. Sử dụng Pivot Table trong Excel

- Hiển thị theo thời gian năm/ quý/ tháng

Row Labels	Column Labels	Grand Total		
		2015	2016	2017
Tommy Bailey	Q2 CY 2015	3353066		
	December 2015	1352695	180445	1988899
Ava Sullivan	October 2015	1418011	1334254	2752265
		723170	1159720	1882890
Ella Wallace	630317	1050294	1579640	3519494
	1655712	914076	276996	3897078
Jessica Woods		1140458	2090620	3231078
	955272	1663458	1585976	3249434
Freya Gomez	1089553	1585289	1118658	544271
Grand Total	4330854	1585289	2402989	12090982
		14056166	34466280	

- Với hình trên thì ta cũng có thể thấy được số liệu bán hàng của các nhân viên và đồng thời hiển thị người quản lý của nhân viên đó. Nhìn vào đây có thể thấy rằng doanh thu công ty vào năm 2015 còn khá ít và sau đó tăng dần theo thời gian.

4.4.1.3. Sử dụng Power BI

- Biểu đồ xếp hạng các nhân viên có doanh thu cao nhất theo năm/ quý/ tháng.



4.4.2. Câu hỏi: Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất trong tháng/ quý/ năm.

4.4.2.1. Sử dụng công cụ SSAS

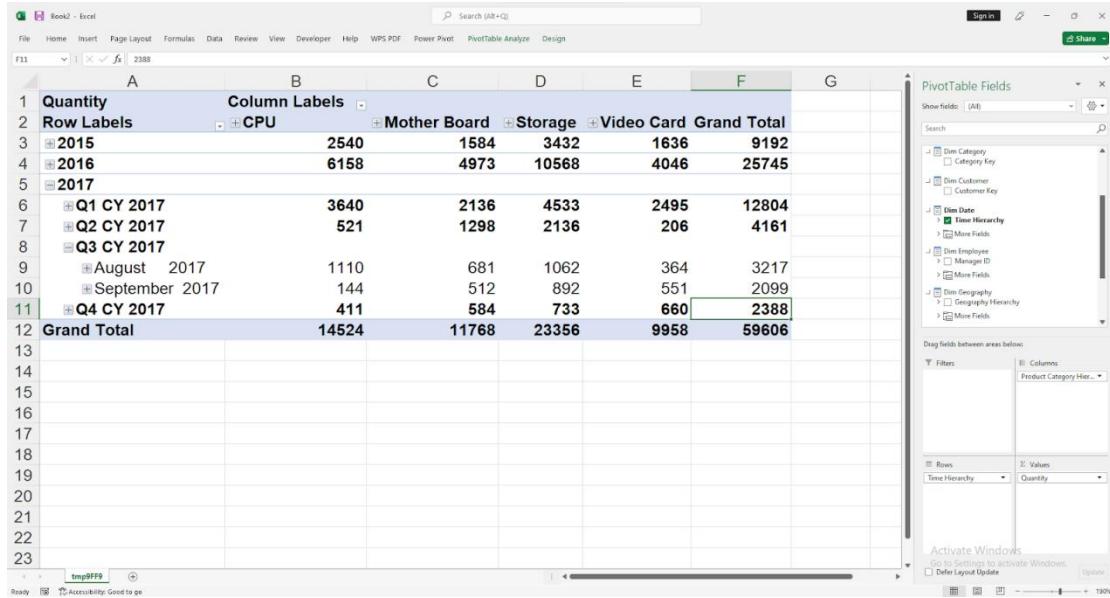
- Số lượng sản phẩm được phân loại theo ngành hàng trong các kho.

The screenshot shows the Microsoft Analysis Services (SSAS) Management Studio interface. The top menu bar includes tabs for Dim Employee.dim [Design], Dim Product.dim [Design], Dim Category.dim [Design], Dim Inventory.dim [Design], Dim Geography.dim [Design], and Dim Date [Design]. Below the menu is a toolbar with various icons for cube structure, dimension usage, calculations, KPIs, actions, partitions, aggregations, perspectives, translations, and browser. A language dropdown is set to Default. The main area has a grid titled "Hardware_DW" with columns for Dimension, Hierarchy, Operator, and Filter Expression. A specific row for "Dim Date" is selected with a value of "Calendar Year" under "Hierarchy", "Equal" under "Operator", and a filter expression of "{2016}" in the "Filter Expression" column. To the left of the grid is a navigation pane titled "Hardware_DW" containing sections for Metadata, Measures (Fact Product Inventory, Fact Sales), KPIs, Dim Category, Dim Customer, Dim Date (with sub-sections for Calendar Quarter, Calendar Year, Date Key, Month Name, Time Hierarchy), Dim Employee, Dim Geography, Dim Inventory, Dim Product (with sub-sections for Category Name, List Price, Product ID), and Calculated Members. The right side of the interface shows a large grid of data with columns for Category Name, Product Name, and Quantity. The data lists various CPU models and their quantities, such as AMD Opteron 6378 (166), Intel Core 2 Extreme QX6800 (128), Intel Core 2 Extreme QX9775 (32), Intel Core i7-3930K (300), Intel Core i7-4790K (207), Intel Core i7-4930K (122), Intel Core i7-4960X Extreme Edition (80), Intel Core i7-6900K (OEM/Tray) (102), Intel Core i7-7820X (104), Intel Core i7-990X Extreme Edition (80), Intel Core i9-7900X (117), Intel Xeon E5-1650 V3 (129), Intel Xeon E5-1650 V4 (89), Intel Xeon E5-1660 V3 (OEM/Tray) (47), Intel Xeon E5-1680 V3 (OEM/Tray) (67), Intel Xeon E5-2630 V2 (271), Intel Xeon E5-2630 V3 (96), Intel Xeon E5-2630 V4 (61), Intel Xeon E5-2637 V2 (OEM/Tray) (78), Intel Xeon E5-2640 V4 (106), Intel Xeon E5-2643 V2 (OEM/Tray) (105), Intel Xeon E5-2643 V3 (OEM/Tray) (301), Intel Xeon E5-2650 (210), and Intel Xeon E5-2650 V2 (108).

Category Name	Product Name	Quantity
CPU	AMD Opteron 6378	166
CPU	Intel Core 2 Extreme QX6800	128
CPU	Intel Core 2 Extreme QX9775	32
CPU	Intel Core i7-3930K	300
CPU	Intel Core i7-4790K	207
CPU	Intel Core i7-4930K	122
CPU	Intel Core i7-4960X Extreme Edition	80
CPU	Intel Core i7-6900K (OEM/Tray)	102
CPU	Intel Core i7-7820X	104
CPU	Intel Core i7-990X Extreme Edition	80
CPU	Intel Core i9-7900X	117
CPU	Intel Xeon E5-1650 V3	129
CPU	Intel Xeon E5-1650 V4	89
CPU	Intel Xeon E5-1660 V3 (OEM/Tray)	47
CPU	Intel Xeon E5-1680 V3 (OEM/Tray)	67
CPU	Intel Xeon E5-2630 V2	271
CPU	Intel Xeon E5-2630 V3	96
CPU	Intel Xeon E5-2630 V4	61
CPU	Intel Xeon E5-2637 V2 (OEM/Tray)	78
CPU	Intel Xeon E5-2640 V4	106
CPU	Intel Xeon E5-2643 V2 (OEM/Tray)	105
CPU	Intel Xeon E5-2643 V3 (OEM/Tray)	301
CPU	Intel Xeon E5-2650	210
CPU	Intel Xeon E5-2650 V2	108

4.4.2.2. Sử dụng Pivot Table trong Excel

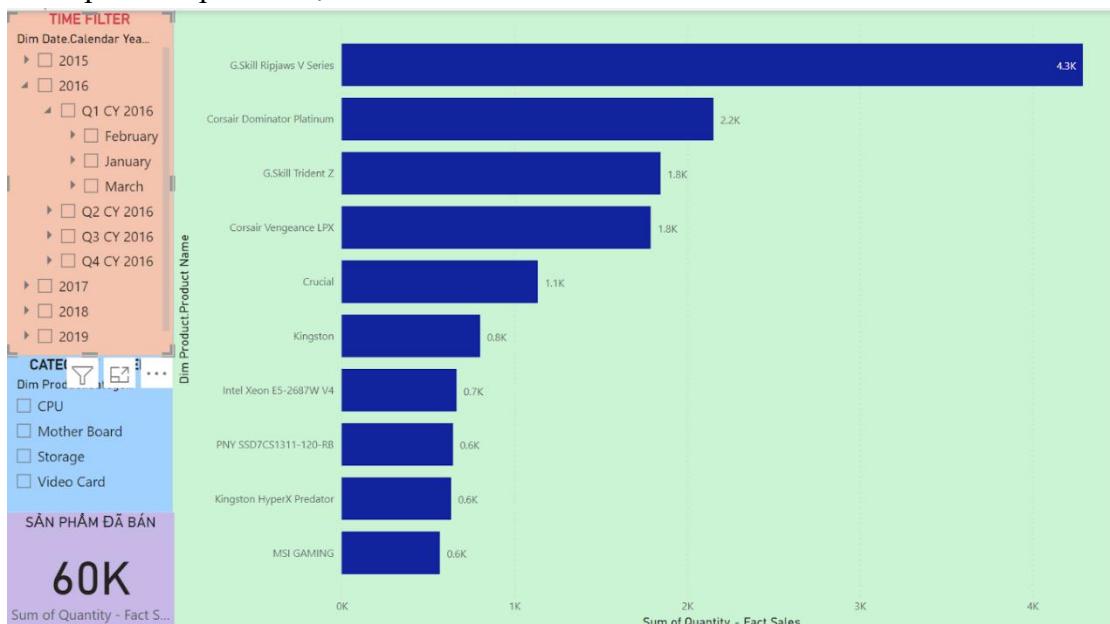
- Số lượng mặt hàng bán ra theo thời gian



- Hình trên thể hiện số lượng các mặt hàng được bán ra theo thời gian gồm có 4 mặt hàng chính là CPU, Mother Board, Storage, Video Card. Chú ý năm 2016 là năm có số lượng sản phẩm bán ra nhiều nhất với hơn hai mươi lăm nghìn sản phẩm, chiếm chủ yếu là mặt hàng Storage.

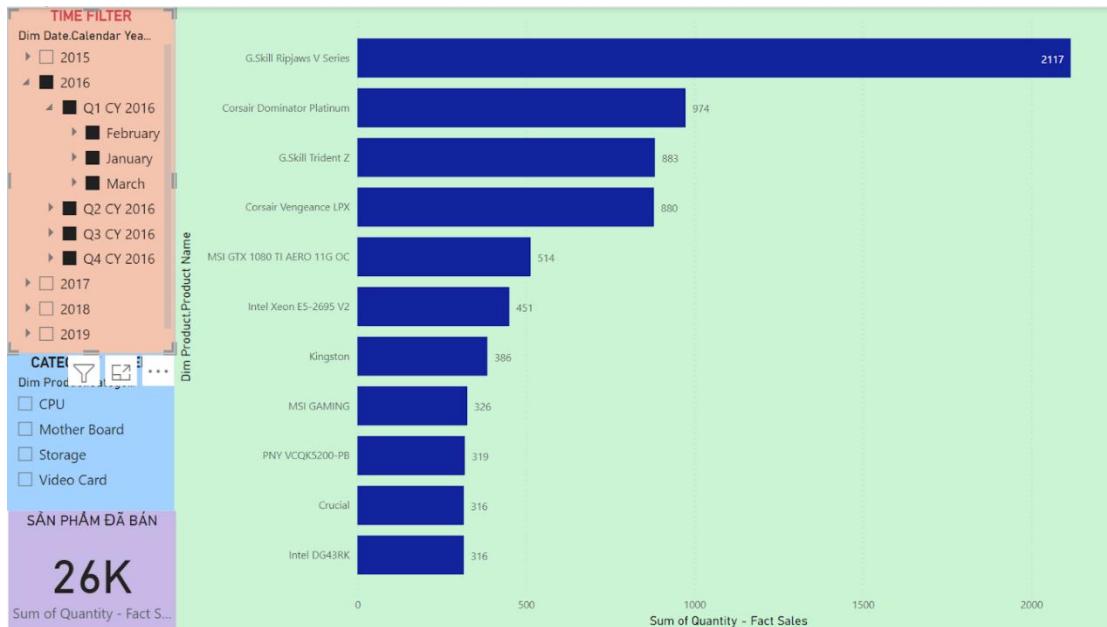
4.4.2.3. Sử dụng Power BI

- Top 10 sản phẩm được bán nhiều nhất



- Nhìn vào biểu đồ ngang, ta có thể thấy sản phẩm G.Skill Ripjaws V Series là sản phẩm có lượt tiêu thụ cao nhất với hơn hai mươi một nghìn lượt bán ra.

- Ta cũng có thể chọn xem số liệu theo Tháng/ Quý/ Năm bằng cách sử dụng bộ lọc Time Filter.



- Từ biểu đồ, ta có thể thấy sản phẩm *G.Skill Ripjaws V Series* có lượng tiêu thụ khá tốt, vượt hẳn những sản phẩm còn lại.

4.4.3. Câu hỏi: Cho biết doanh số bán hàng theo từng khách hàng trong tháng/ quý/ năm.

4.4.3.1. Sử dụng công cụ SSAS

- Số liệu thể hiện tổng số tiền mua sắm từ các khách hàng tính từ trước cho đến nay trên toàn thế giới.

Screenshot of the Analysis Services Dimension Designer interface showing the Dim Date dimension.

Top Navigation: Dim Customer.dim [Design], Dim Employee.dim [Design], Dim Product.dim [Design], Dim Category.dim [Design], Dim Inventory.dim [Design], Dim Geography

Toolbar: Cube Structure, Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, Browser

Language: Default

Dimension Properties:

Dimension	Hierarchy	Operator	Filter Expression
Dim Date	Time Hierarchy	Equal	{ 2017, Q2 CY 2017 }

Tree View (Left):

- Hardware_DW
 - Measures
 - Fact Product Inventory
 - Fact Sales
 - Quantity
 - Sales Amount
 - Unit Price
 - KPIs
 - Dim Category
 - Dim Customer
 - Address
 - Credit Limit
 - Customer ID
 - Customer Key
 - Customer Name
 - Website
 - Dim Date
 - Calendar Quarter
 - Calendar Year
 - Date Key
 - Month Name
 - Time Hierarchy
 - Dim Employee

Calculated Members:

Customer Name	Sales Amount
AbbVie	1142458
Aflac	708198
Alcoa	29234
American Electric Power	273054
AutoNation	967207
AutoZone	645025
Centene	608296
CenturyLink	1692386
Community Health Systems	972205
Dollar General	90522
Emerson Electric	951187
General Mills	1081228
Goodyear Tire & Rubber	675181
Health Net	1268440
International Paper	353473
Jabil Circuit	1882919
NextEra Energy	1447572
PG&E Corp.	336727
Plains GP Holdings	433467
PNC Financial Services Group	147825
Progressive	949548
Raytheon	2404317
Southern	1303954
Supervalu	699887

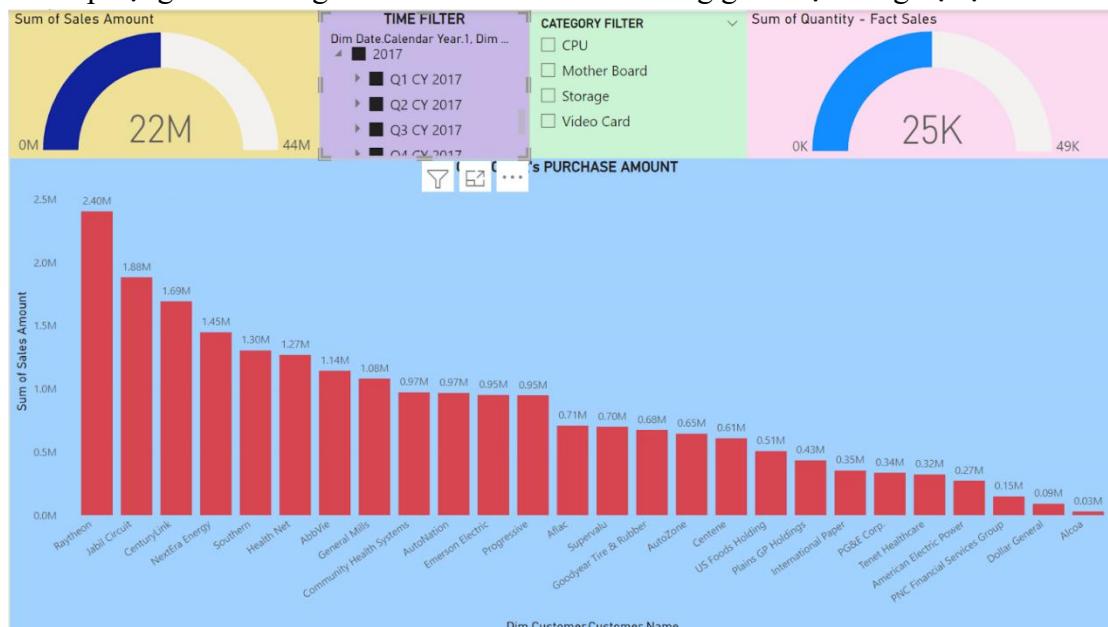
4.4.3.2. Sử dụng Pivot Table trong Excel

	A	B	C	D	E
2	Row Labels	2015	2016	2017	Grand Total
3	Abbott Laboratories		696602		696602
4	AbbVie	589663	170805	1142458	1902926
5	AECOM		376689		376689
6	Aflac		993582	708198	1701780
7	Alcoa	498821	1382328	29234	1910383
8	American Electric Power			273054	273054
9	AutoNation		343576	967207	1310783
10	AutoZone			645025	645025
11	Baker Hughes		557235		557235
12	Bank of New York Mellon Corp.		313257		313257
13	Becton Dickinson		483602		483602
14	Bristol-Myers Squibb		864195		864195
15	Centene		193664	608296	801960
16	CenturyLink		710482	1692386	2402868
17	Colgate-Palmolive		273518		273518
18	Community Health Systems	935551		972205	1907756

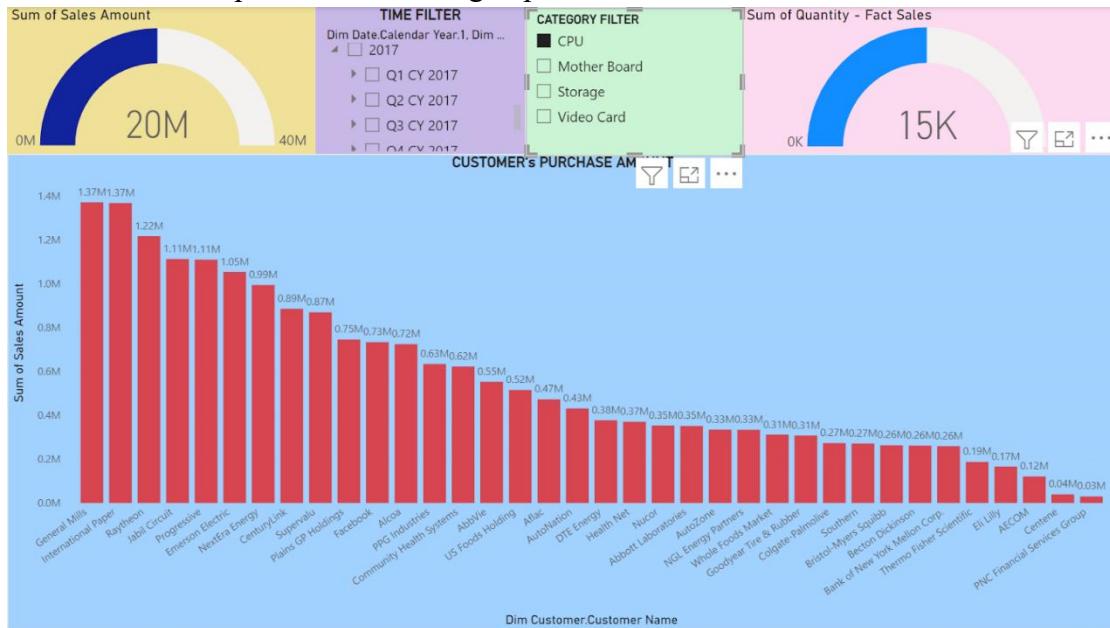
- Hình trên biểu diễn doanh thu bán hàng theo khách hàng với năm 2015 thì có thể thấy được không có quá nhiều khách hàng sử dụng sản phẩm tuy nhiên sau một năm thì lượng khách hàng tăng dần và doanh thu cũng tăng đáng kể. Dựa vào đây ta có thể xem xét những khách hàng đem lại doanh thu cao cho công ty từ đó thực hiện những chương trình ưu đãi hay quà tặng cho những khách hàng thân thiết.

4.4.3.3. Sử dụng Power BI

- Xếp hạng khách hàng có mức chi tiêu lớn theo từng giai đoạn bằng bộ lọc Time Filter



- Đối với từng loại danh mục sản phẩm, cho biết doanh thu theo khách hàng bằng bộ lọc Category Filter. Bằng cách này có thể xác định được tập khách hàng tiềm năng cho từng danh mục sản phẩm của doanh nghiệp.



- Chọn ra top 10 khách hàng có mức chi tiêu cao nhất, ta thấy không có sự chênh lệch đáng kể giữa các khách hàng.



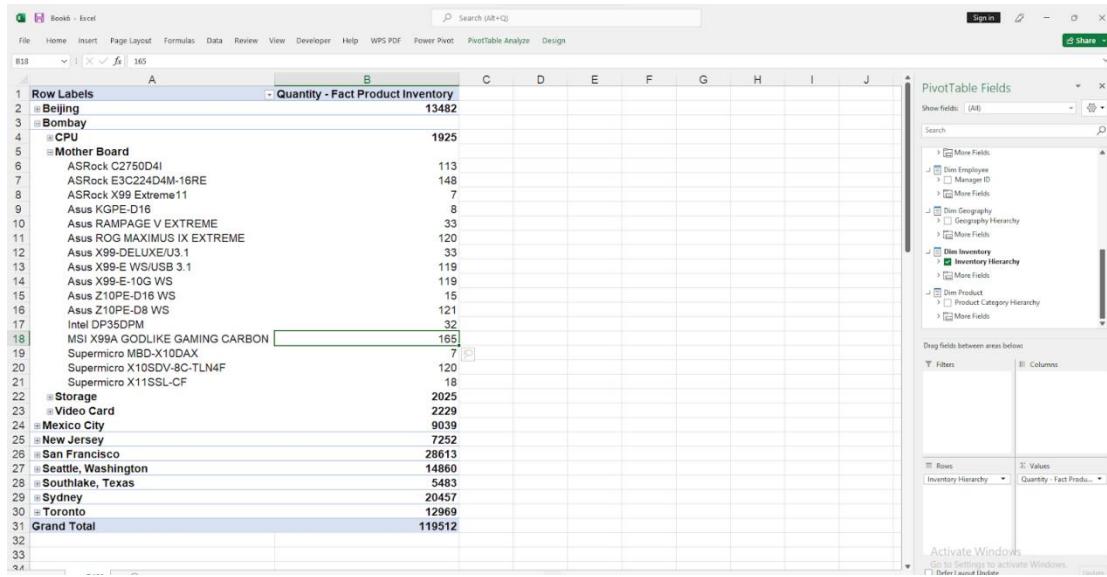
4.4.4. Câu hỏi: Quản lý số lượng sản phẩm phân bổ trong các kho.

4.4.4.1. Sử dụng công cụ SSAS

- Số liệu biểu diễn tổng số lượng sản phẩm của một ngành hàng được phân bổ theo các kho khác nhau.

Warehouse ID	Category Name	Quantity - Fact Product Inventory
Beijing	CPU	2472
Beijing	Mother Board	1779
Beijing	Storage	6543
Beijing	Video Card	2688
Bombay	CPU	1925
Bombay	Mother Board	1178
Bombay	Storage	2025
Bombay	Video Card	2229
Mexico City	CPU	892
Mexico City	Mother Board	544
Mexico City	Storage	4418
Mexico City	Video Card	3185
New Jersey	CPU	1581
New Jersey	Video Card	5671
San Francisco	CPU	4508
San Francisco	Mother Board	3459
San Francisco	Storage	14167
San Francisco	Video Card	6479
Seattle, Was...	CPU	3284
Seattle, Was...	Mother Board	1734
Seattle, Was...	Storage	5084
Seattle, Was...	Video Card	4758
Southlake, Te...	CPU	258
Southlake, Te...	Video Card	5225

4.4.4.2. Sử dụng Pivot Table trong Excel



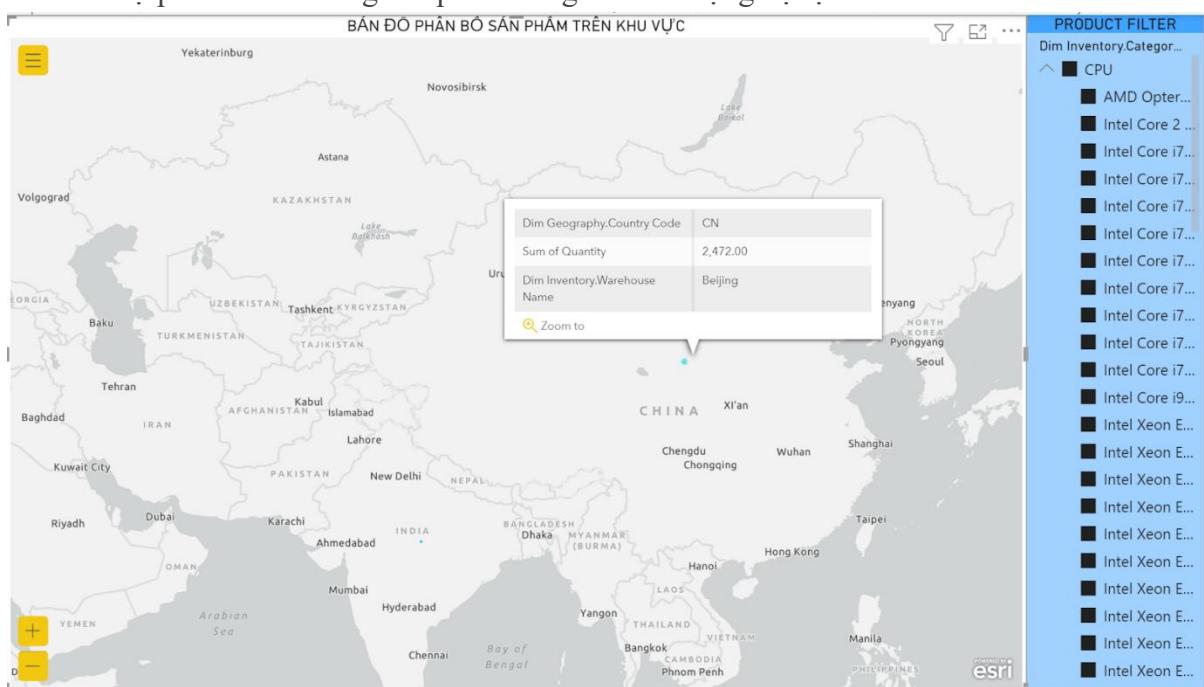
- Nhìn vào hình này sẽ thấy được số lượng sản phẩm theo mặt hàng được phân bổ trong các thành phố trên thế giới. Cụ thể công ty có 9 kho hàng ở các thành phố: Beijing, Mexico, New Jersey, San Francisco, Washington, Texas, Sydney và Toronto.
- Trong đó Bombay thuộc thành phố Beijing Trung Quốc là một kho hàng cũng khá lớn với hơn mười ba nghìn sản phẩm thuộc bốn mặt hàng chính. Và đối với mặt hàng Mother Board thì gồm có hơn 15 sản phẩm khác nhau.

4.4.4.3. Sử dụng Power BI

- Tạo bảng đồ thể hiện sự phân bố sản phẩm.
- Mức độ to nhỏ của các chấm tròn thể hiện số lượng sản phẩm được phân bố ở vị trí đó.
- Mỗi một màu sắc đại diện cho một warehouse trên khu vực đó.



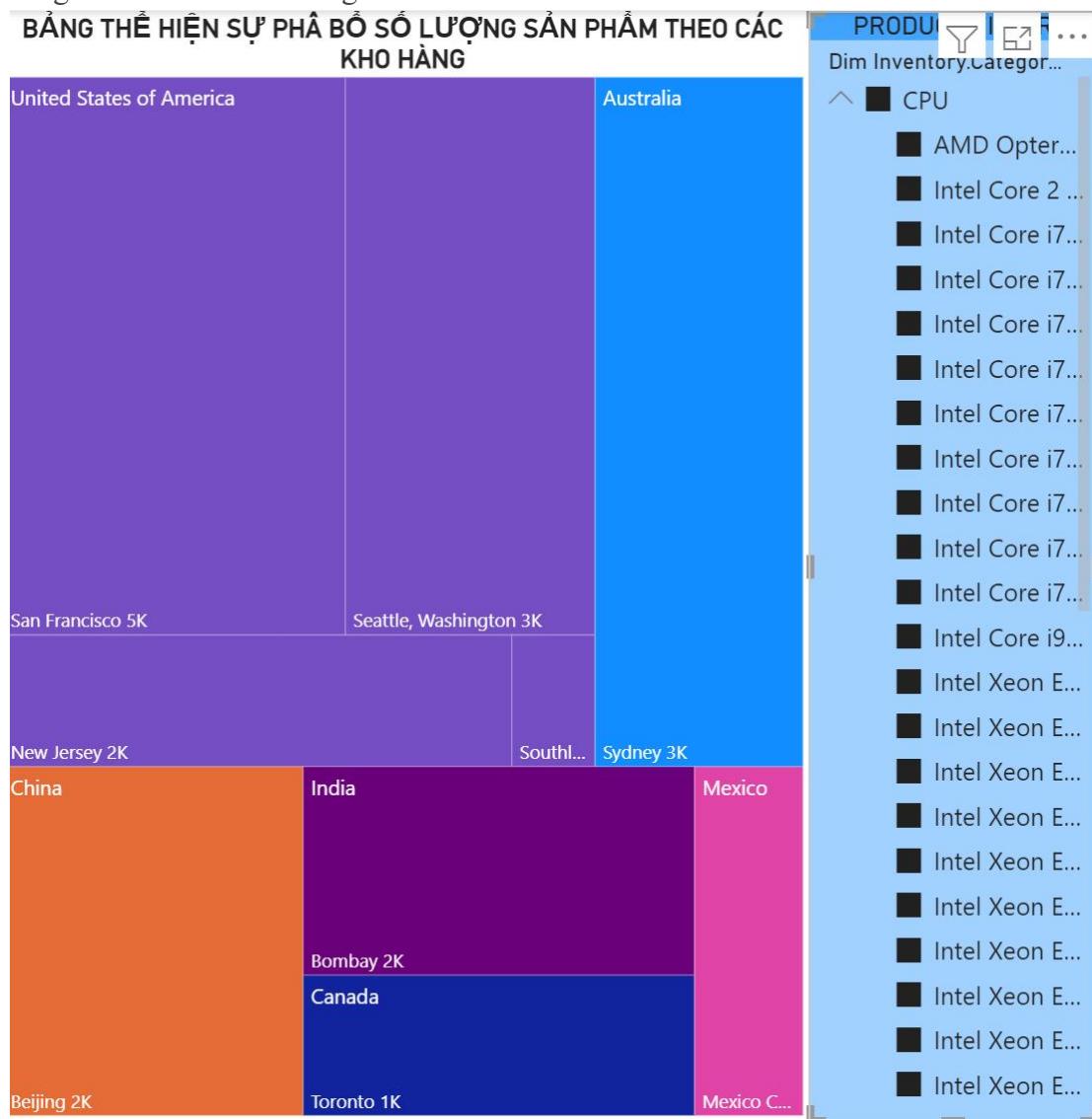
- Có thể thấy khu vực hoạt động của công ty sẽ tập trung ở hai châu lục lớn là Châu Mỹ và châu Á
- Xem sự phân bố của từng sản phẩm bằng cách sử dụng bộ lọc.



- Các sản phẩm trong danh mục CPU có 2472 các loại sản phẩm trong kho Beijing.

- Biểu đồ so sánh sự phân bố số lượng sản phẩm phân loại theo danh mục trong các kho giữa các nước trên thế giới

BẢNG THỂ HIỆN SỰ PHÂN BỐ SỐ LƯỢNG SẢN PHẨM THEO CÁC KHO HÀNG



- Nhìn vào đây ta có thể thấy ở Mỹ có tổng cộng bốn kho hàng và đây cũng là khu vực hoạt động lớn nhất của công ty khi chiếm hơn một nửa số lượng sản phẩm trên thế giới, tiếp đến là kho hàng Sydney ở Úc và Beijing của China, cùng thuộc khu vực châu Á có kho hàng Bombay của Ấn Độ, và hai kho hàng còn lại là ở Toronto của Canada và thành phố Mexico.
- Dựa vào những số liệu này thì ta có thể theo dõi được số lượng sản phẩm tồn kho để có thể phân phối tới những kho hàng lận cận.

CHƯƠNG 5: KẾT LUẬN

5.1. Kết quả đạt được

Thực hiện xây dựng ‘*KHO DỮ LIỆU BÁN HÀNG CHO CÔNG TY THIẾT BỊ PHẦN CỨNG*’ giúp chúng em hiểu sâu hơn về kiến thức môn học Kho Dữ liệu và thành thạo hơn trong việc sử dụng các công cụ hỗ trợ quá trình ETL đưa dữ liệu vào kho dữ liệu như công cụ SSIS, SSAS, các công cụ để trực quan hóa dữ liệu như Excel Pivot, Power BI Desktop. 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. Những thực hiện của chúng em chưa đạt đến mức nâng cao trong tổng thể quá trình xây dựng một kho dữ liệu hoàn chỉnh.

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 12 May 2023].
- [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 2023].
- [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 10 May 2023].