

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 ADVENTURE WORK CYCLES**

**Môn học: KHO DỮ LIỆU
Mã lớp học phần: DAWH430784_23_2_01**

GVHD: ThS. Nguyễn Văn Thành

Nhóm sinh viên thực hiện: Nhóm 15

Đào Lê Huy Giáp	21133028
Bùi Thị Huỳnh Hân	21133029
Phan Khải Huyền	21133041
Nguyễn Tấn Sương	21133078

TP. Hồ Chí Minh, 16 tháng 05 năm 2024

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

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

STT	HỌ VÀ TÊN	MSSV	TỶ LỆ ĐÓNG GÓP
1	Đào Lê huy Giáp	21133028	100%
2	Bùi Thị Huỳnh Hân	21133029	100%
3	Phan Khải Huyền	21133041	100%
4	Nguyễn Tân Sương	21133078	100%

Nhận xét của giảng viên:

Ngày ... tháng 05 năm 2024

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 AdventureWorks_Calendar	3
1.2.2.2. Bảng AdventureWorks_Sales_2015	3
1.2.2.3. Bảng AdventureWorks_Customers	4
1.2.2.4. Bảng AdventureWorks_Product	4
1.2.2.5. Bảng AdventureWorks_Territories	5
1.2.2.6. Bảng AdventureWorks_Product_Categories	6
1.2.2.7. Bảng AdventureWorks_Product_Subcategories	6
1.2.2.8. Bảng AdventureWorks_Return	6
1.2.2.9. Bảng AdventureWorks_Sales_2016	6
1.2.2.10. Bảng AdventureWorks_Sales_2017	8
1.3. Giới thiệu các công cụ được sử dụng	8
CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP)	9
2.1. Xác định các Business Process và bảng Fact	9
2.1.1. Business Process: Sales Reporting	9
2.1.2. Business Process: Return	9
2.2. Xây dựng các bảng Dimension	10
2.2.1. DimCalendar	10
2.2.2. DimTerritory	10
2.2.3. DimProduct	10
2.2.4. DimCustomer	10
2.2.5. DimSubcategory	11
2.2.6. DimCategory	11
2.3. Star Schema (Lược đồ hình sao)	11
CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)	13
3.1. Tạo ETL project và solution	13
3.2. Thiết lập các Connection	14
3.3. Import Dữ liệu vào các bảng stage:	18
3.3.1. Customer	18
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	22
3.3.1.4: Thiết lập truncate table SQL task	27
3.3.2. Product	29
3.3.2.1. Tạo Package	29
3.3.2.2. Thiết lập Control Flow	30
3.3.2.3. Load dữ liệu từ Source vào Stage	31

3.3.3. Categories	33
3.3.3.1. Load dữ liệu từ Source vào Stage	33
3.3.4. SubCategory	36
3.3.4.1. Load dữ liệu từ Source vào Stage	36
3.3.5. Calendar	39
3.3.5.1. Load dữ liệu từ Source vào Stage	39
3.3.6. Territory	42
3.3.6.1. Load dữ liệu từ Source vào Stage	42
3.3.7. Returns	44
3.3.7.1. Load dữ liệu từ Source vào Stage	44
3.3.8. Sales	46
3.3.8.1. Load dữ liệu từ Source vào Stage	46
3.4. Import dữ liệu từ bảng stage vào bảng dimension	49
3.4.1. DimCustomer	49
3.4.1.1. Tạo Package	49
3.4.1.2. Thiết lập Control Flow	49
3.4.1.3. Load dữ liệu từ Stage vào Dimension	50
3.4.2. DimCalendar	52
3.4.2.1. Load dữ liệu từ Stage vào Dimension	52
3.4.3. DimCategory	55
3.4.3.1. Load dữ liệu từ Stage vào Dimension	55
3.4.4. DimSubCategory	57
3.4.4.1. Load dữ liệu từ Stage vào Dimension	57
3.4.5. DimTerritory	60
3.4.5.1. Load dữ liệu từ Stage vào Dimension	60
3.4.6. DimProduct	62
3.4.6.1. Load dữ liệu từ Stage vào Dimension	62
3.5. Import dữ liệu vào bảng các bảng fact	64
3.5.1. Fact Sales	64
3.5.1.1. Tạo Package	64
3.5.1.2. Thiết lập Control Flow	65
3.5.1.3. Load dữ liệu từ Stage vào Fact	66
3.5.2. Fact Returns	70
3.5.2.1. Load dữ liệu từ Stage vào Fact	70
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)	73
4.1. Quá trình xây dựng mô hình	73
4.1.1. Tạo Data Source	73
4.1.2. Tạo Data Source View	74
4.2. Quá trình xây dựng khối Cube	76
4.3. Cấu hình Hierarchy	78
4.3.1. Tạo Hierarchy cho Dim Calendar	78
4.3.2. Tạo Hierarchy cho Dim Customer	80
4.3.3. Tạo Hierarchy cho Dim Product	81

4.3.4. Tạo Hierarchy cho Dim Territory	82
4.4. Thực hiện phân tích dữ liệu	83
4.4.1. Câu hỏi: Cho biết sản phẩm nào có số lượng bán ra cao nhất.....	83
4.4.1.1. Sử dụng Power BI.....	83
4.4.2. Câu hỏi: Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất.....	83
4.4.2.1. Sử dụng Power BI.....	83
4.4.3. Câu hỏi: Cho biết doanh số bán hàng theo từng khách hàng.....	85
4.4.3.1. Sử dụng Power BI.....	85
CHƯƠNG 5: KẾT LUẬN	86
5.1. Kết quả đạt được	86
5.2. Những hạn chế	86
5.3. Tài liệu tham khảo	86

PHÂN CÔNG NHIỆM VỤ NHÓM 15

Nhiệm vụ	Tấn Sương	Huy Giáp	Huỳnh Hân	Khải Huyền
Tìm kiếm tập dữ liệu	x	x	x	x
Mô hình hóa dữ liệu vào workbook			x	
+ DimCalendar + DimCustomer			x	
+DimSubCategory + DimTerritory			x	
+ DimProduct + DimCategory + FactSales +FactReturn			x	
Thực hiện ETL đưa dữ liệu vào các bảng dim và fact dùng SSIS		x		
+ Tạo và đưa dữ liệu vào các dim, stages		x		
+ Tạo và đưa dữ liệu vào FactSales		x		
+ Tạo và đưa dữ liệu vào FactReturn		x		
Nhập dữ liệu vào SSAS, tạo data source, data source view và cube	x	x		
Tạo các phân cấp hierarchy cho các bảng dim và các attribute relationships.	x			x
+ Hierarchy cho DimCalendar	x			
+ Hierarchy cho DimProduct	x			

+ Hierarchy cho DimTerritory, DimCustomer	x			
---	---	--	--	--

Nhiệm vụ	Tấn Sương	Huy Giáp	Huỳnh Hân	Khải Huyền
Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Sales Analysis	x			x
+ Cho biết sản phẩm nào bán ra có số lượng doanh số cao nhất	x			x
+ Cho biết danh mục các loại sản phẩm được người mua order nhiều nhất	x			x
+ Cho biết danh số bán hàng theo từng khu vực	x			x
+ Cho biết tỷ lệ trả hàng theo từng sản phẩm.	x			x
+ Trả lời câu hỏi dùng SSAS	x			
+ Trả lời câu hỏi dùng Power BI Desktop				x

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

1.1. Lý do chọn đề tài

Lý do chọn đề tài "Xây dựng và phân tích kho dữ liệu bán hàng từ AdventureWorks" như sau: Đây là một đề tài mang tính thực tế và ứng dụng cao, vì AdventureWorks là một mô hình cơ sở dữ liệu mẫu được sử dụng rộng rãi trong các ứng dụng về phân tích dữ liệu và báo cáo, vì vậy xây dựng và phân tích kho dữ liệu từ AdventureWorks sẽ giúp tích lũy những kỹ năng thực tế và có thể ứng dụng vào nhiều bài toán trong thực tế. Ngoài ra, AdventureWorks bao gồm nhiều bảng dữ liệu với các mối quan hệ phức tạp, vì vậy việc thiết kế và xây dựng kho dữ liệu từ nguồn dữ liệu này sẽ giúp tích lũy kinh nghiệm trong việc quản lý các nguồn dữ liệu lớn và phức tạp, điều này rất hữu ích cho sự phát triển nghề nghiệp trong tương lai. Sau khi xây dựng xong kho dữ liệu, có cơ hội thực hành các kỹ thuật phân tích dữ liệu để rút ra những thông tin, xu hướng và các chỉ số quan trọng từ dữ liệu bán hàng, điều này sẽ giúp nâng cao kỹ năng phân tích và báo cáo dữ liệu. Cuối cùng, các kỹ năng và kiến thức thu được từ đề tài này có thể áp dụng vào nhiều lĩnh vực khác nhau như quản trị kinh doanh, marketing, quản lý chuỗi cung ứng, v.v. điều này sẽ mang lại giá trị cao cho sự phát triển nghề nghiệp.

1.2. Tổng quan về tập dữ liệu

1.2.1. Nguồn dữ liệu

- Nhóm sử dụng Tập dữ liệu AdventureWorks Database
- Đường dẫn tải tập dữ liệu:
<https://www.kaggle.com/datasets/ukveteran/adventure-works/data>

1.2.2. Mô tả chi tiết tập dữ liệu

- AdventureWorks là một mô hình cơ sở dữ liệu mẫu được phát triển và cung cấp bởi Microsoft. Nó mô phỏng một công ty sản xuất và bán xe đạp.
- Cơ sở dữ liệu này bao gồm nhiều bảng thể hiện các khía cạnh của hoạt động kinh doanh như bán hàng, quản lý sản phẩm, quản lý nhân sự, v.v.

Tên bảng	Mô tả	Số dòng
AdventureWorks_Calendar		912 dòng
AdventureWorks_Customers	lưu trữ thông tin khách hàng	18148 dòng
AdventureWorks_Product_Categories	lưu trữ thông tin	4 dòng
AdventureWorks_Product_Subcategories	lưu trữ thông tin	37 dòng
AdventureWorks_Products	lưu trữ thông tin	293 dòng
AdventureWorks_Returns	lưu trữ	1809 dòng
AdventureWorks_Territories	lưu trữ thông tin chung của đơn hàng	10 dòng
AdventureWorks_Sales_2015	lưu trữ thông tin	2632 dòng
AdventureWorks_Sales_2016	lưu trữ thông tin	23935 dòng
AdventureWorks_Sales_2017	lưu trữ thông tin	29481 dòng

1.2.2.1. Bảng AdventureWorks_Calendar

Biến	Mô tả
Date	Thời gian(khóa chính)

1.2.2.2. Bảng AdventureWorks_Sales_2015

Biến	Mô tả
OrderDate	Ngày đặt hàng
StockDate	Ngày nhập kho
OrderNumber	Số đơn hàng
ProductKey	Mã sản phẩm(Khóa ngoại trỏ đến AdventureWorks_Product(ProductKey))
CustomerKey	Mã khách hàng (Khóa ngoại trỏ đến AdventureWorks_Customers(CustomerKey))
TerritoryKey	Mã khu vực (Khóa ngoại trỏ đến AdventureWorks_Territories(TerritoryKey))
OrderLineItem	Số mục trong đơn hàng
OrderQuantity	Số lượng sản phẩm trong đơn hàng

1.2.2.3. Bảng AdventureWorks_Customers

Biến	Mô tả
CUSTOMERSKEY	Mã khách hàng (Khóa chính)
PREFIX	Danh xưng của khách hàng
FIRSTNAME	Họ khách hàng
LASTNAME	Tên khách hàng
BIRTHDATE	Sinh nhật
MaritalStatus	Tình trạng hôn nhân
Gender	Giới tính
EmailAddress	Địa chỉ Email
AnnualIncome	Thu nhập hằng năm của khách hàng
TotalChildren	Tổng số con của mỗi khách hàng

1.2.2.4. Bảng AdventureWorks_Product

Biến	Mô tả
ProductKey	Mã sản phẩm (Khóa chính)
ProductSubcategoryKey	Mã danh mục phụ sản phẩm (Khóa ngoại trỏ đến AdventureWorks_Product_Subcategory(ProductSubcategoryKey))
ProductSKU	Mã sản phẩm
ProductName	Tên sản phẩm
ModelName	Mẫu sản phẩm
ProductDescription	Mô tả sản phẩm
ProductColor	Màu sản phẩm
ProductSize	Kích thước sản phẩm
ProductStyle	Phong cách sản phẩm
ProductCost	Chi phí sản xuất

1.2.2.5. Bảng AdventureWorks_Territories

Biển	Mô tả
TerritoryKey	Mã khu vực (Khóa chính)
Region	Khu vực địa lý
Country	Quốc gia
Continent	Châu lục

1.2.2.6. Bảng AdventureWorks_Product_Categories

Biến	Mô tả
ProductCatagoryKey	Mã danh mục sản phẩm (Khóa chính)
ProductCatagoryName	Tên danh mục sản phẩm

1.2.2.7. Bảng AdventureWorks_Product_Subcategories

Biến	Mô tả
ProductSubcatagoryKey	Mã danh mục phụ sản phẩm(Khóa chính)
SubcatagoryName	Tên danh mục phụ sản phẩm
ProductCatagoryKey	Mã danh mục sản phẩm(Khóa ngoại trỏ đến AdventureWorks_Product_Categories(ProductCatagoryKey))

1.2.2.8. Bảng AdventureWorks_Return

Biến	Mô tả
ReturnDate	Ngày trả
TerritoryKey	Mã khu vực (Khóa ngoại trỏ đến AdventureWorks_Territories(TerritoryKey))
ProductKey	Mã sản phẩm(Khóa ngoại trỏ đến AdventureWorks_Product(ProductKey))
ReturnQuantity	Số lượng sản phẩm trả lại

1.2.2.9. Bảng AdventureWorks_Sales_2016

Biến	Mô tả
OrderDate	Ngày đặt hàng
StockDate	Ngày nhập kho
OrderNumber	Số đơn hàng
ProductKey	Mã sản phẩm(Khóa ngoại trỏ đến AdventureWorks_Product(ProductKey))
CustomerKey	Mã khách hàng (Khóa ngoại trỏ đến AdventureWorks_Customers(CustomerK

	ey))
TerritoryKey	Mã khu vực (Khóa ngoại trỏ đến AdventureWorks_Territories(TerritoryKey))
OrderLineItem	Số mục trong đơn hàng
OrderQuantity	Số lượng sản phẩm trong đơn hàng

1.2.2.10. Bảng AdventureWorks_Sales_2017

Biến	Mô tả
OrderDate	Ngày đặt hàng
StockDate	Ngày nhập kho
OrderNumber	Số đơn hàng
ProductKey	Mã sản phẩm (Khóa ngoại trỏ đến AdventureWorks_Product(ProductKey))
CustomerKey	Mã khách hàng (Khóa ngoại trỏ đến AdventureWorks_Customers(CustomerKey))
TerritoryKey	Mã khu vực (Khóa ngoại trỏ đến AdventureWorks_Territories(TerritoryKey))
OrderLineItem	Số mục trong đơn hàng
OrderQuantity	Số lượng sản phẩm trong đơn hàng

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	DimCalendar	DimCustomer	DimProduct	DimTerritory	DimSubcategory	DimCategory
Sales Reporting	FactSales	Sales Reporting	FactSales	Transaction	One row per bill	SaleKey, OrderDate, StockDate, OrderNumber, ProductKey, CustomerKey, TerritoryKey, OrderLineItem, OrderQuantity, ProductPrice	x	x	x	x	x	x
Return	FactReturn	Return	FactReturn	Transaction	One row per bill	ReturnKey, ProductKey, TerritoryKey, ReturnDate, ReturnQuantity			x	x		

2.1.1. Business Process: Sales Reporting

- Các câu hỏi cụ thể được đặt ra:
 - + Cho biết sản phẩm nào có số lượng bán ra cao nhất
 - + Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất.
 - + Cho biết doanh số bán hàng theo từng khách hàng.
- Bảng FactSales

FactSales	SaleKey	Khóa chính của bảng Bán hàng.
FactSales	OrderDate	Ngày đặt hàng.
FactSales	StockDate	Ngày nhập kho
FactSales	OrderNumber	Số đơn hàng.
FactSales	ProductKey	Khóa chính của bảng Sản phẩm, dùng để liên kết các bảng sự kiện với bảng Sản phẩm.
FactSales	CustomerKey	Khóa chính của bảng Khách hàng, dùng để liên kết các bảng sự kiện với bảng Khách hàng.
FactSales	TerritoryKey	Khóa chính của bảng Khu vực địa lý, dùng để liên kết các bảng sự kiện với bảng Khu vực địa lý.
FactSales	OrderLineItem	Số mục trong đơn hàng.
FactSales	OrderQuantity	Số lượng sản phẩm trong đơn hàng.
FactSales	ProductPrice	Giá bán sản phẩm.

2.1.2. Business Process: Return

- Các câu hỏi cụ thể được đặt ra:
 - + Cho biết tỷ lệ trả hàng theo từng sản phẩm.

- Bảng FactReturns

FactReturns	ReturnKey	Khóa chính của bảng Trả hàng.
FactReturns	ProductKey	Khóa chính của bảng Sản phẩm, dùng để liên kết các bảng sự kiện với bảng Sản phẩm.
FactReturns	TerritoryKey	Khóa chính của bảng Khu vực địa lý, dùng để liên kết các bảng sự kiện với bảng Khu vực địa lý.
FactReturns	ReturnDate	Ngày trả hàng.
FactReturns	ReturnQuantity	Số lượng sản phẩm trả lại.

2.2. Xây dựng các bảng Dimension

2.2.1. DimCalendar

Instructions!	Attribute / Fact Name	Description
Dimension / Fact Table		
DimCalendar	DateKey	Khóa chính của bảng Thời gian, dùng để liên kết các bảng sự kiện với bảng Thời gian.
DimCalendar	date	Ngày cụ thể.
DimCalendar	day	Ngày trong tháng (1-31).
DimCalendar	month	Tháng trong năm (1-12).
DimCalendar	year	Năm
DimCalendar	Quarter	Quý trong năm (1-4).
DimCalendar	Weekday	Ngày trong tuần

2.2.2. DimTerritory

Instructions!	Attribute / Fact Name	Description
Dimension / Fact Table		
DimTerritory	TerritoryKey	Khóa chính của bảng Khu vực địa lý, dùng để liên kết các bảng sự kiện với bảng Khu vực địa lý.
DimTerritory	Region	Khu vực địa lý.
DimTerritory	Country	Quốc gia
DimTerritory	Continent	Châu lục

2.2.3. DimProduct

Instructions!	Attribute / Fact Name	Description
Dimension / Fact Table		
DimProduct	ProductKey	Khóa chính của bảng Sản phẩm, dùng để liên kết các bảng sự kiện với bảng Sản phẩm.
DimProduct	ProductSubcategory	Danh mục phụ của sản phẩm.
DimProduct	ProductSKU	Mã sản phẩm.
DimProduct	ProductName	Tên sản phẩm.
DimProduct	ModelName	Tên mẫu sản phẩm.
DimProduct	ProductDescription	Mô tả sản phẩm
DimProduct	ProductColor	Màu sản phẩm.
DimProduct	ProductSize	Kích thước sản phẩm.
DimProduct	ProductStyle	Phong cách sản phẩm.

2.2.4. DimCustomer

Instructions!	Attribute / Fact Name	Description
Dimension / Fact Table		
DimCustomer	CustomerKey	Khóa chính của bảng Khách hàng, dùng để liên kết các bảng sự kiện với bảng Khách hàng.
DimCustomer	Prefix	Danh xưng của khách hàng (Ông, Bà, Cô, v.v.).
DimCustomer	FirstName	Tên của khách hàng.
DimCustomer	LastName	Họ của khách hàng.
DimCustomer	BirthDate	Ngày sinh của khách hàng.
DimCustomer	Gender	Giới tính của khách hàng.
DimCustomer	EmailAddress	Địa chỉ email của khách hàng
DimCustomer	FullName	Họ và tên đầy đủ của khách hàng.

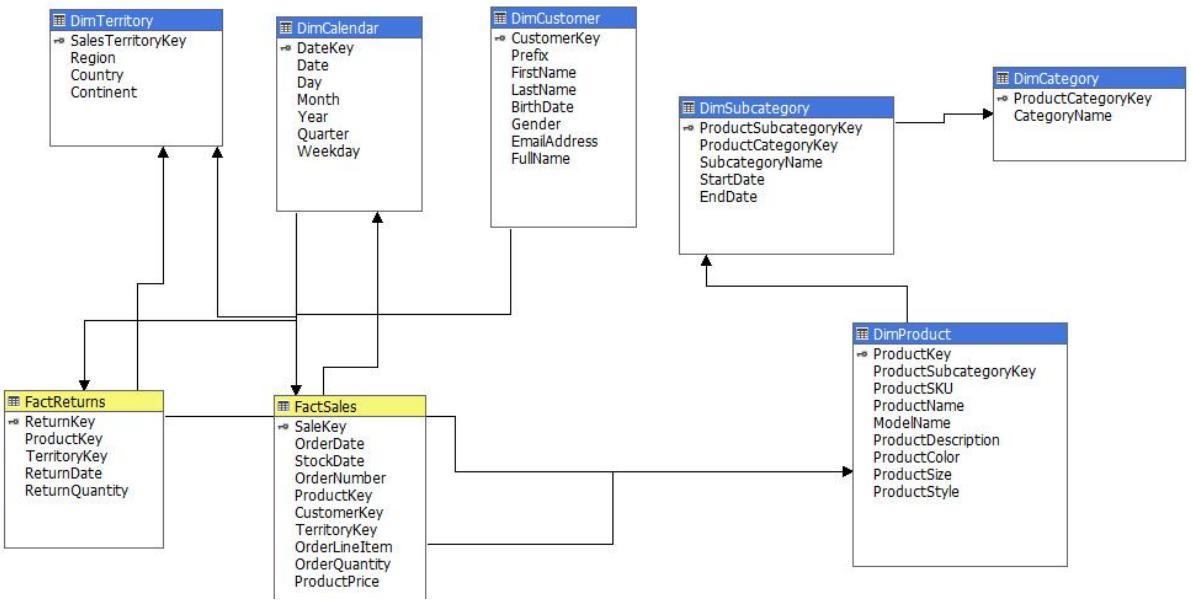
2.2.5. DimSubcategory

Instructions!		
Dimension / Fact Table	Attribute / Fact Name	Description
DimSubcategory	ProductSubcategoryKey	Khóa chính của bảng Danh mục phụ sản phẩm.
DimSubcategory	ProductCategoryKey	Khóa chính của bảng Danh mục sản phẩm.
DimSubcategory	Sub categoryName	Tên danh mục phụ sản phẩm.

2.2.6. DimCategory

Instructions!		
Dimension / Fact Table	Attribute / Fact Name	Description
DimCategory	ProductCategoryKey	Khóa chính của bảng Danh mục sản phẩm.
DimCategory	ProductCategoryName	Tên danh mục sản phẩm.

2.3. Star Schema (Lược đồ hình sao)



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

3.1. Tạo ETL project và solution

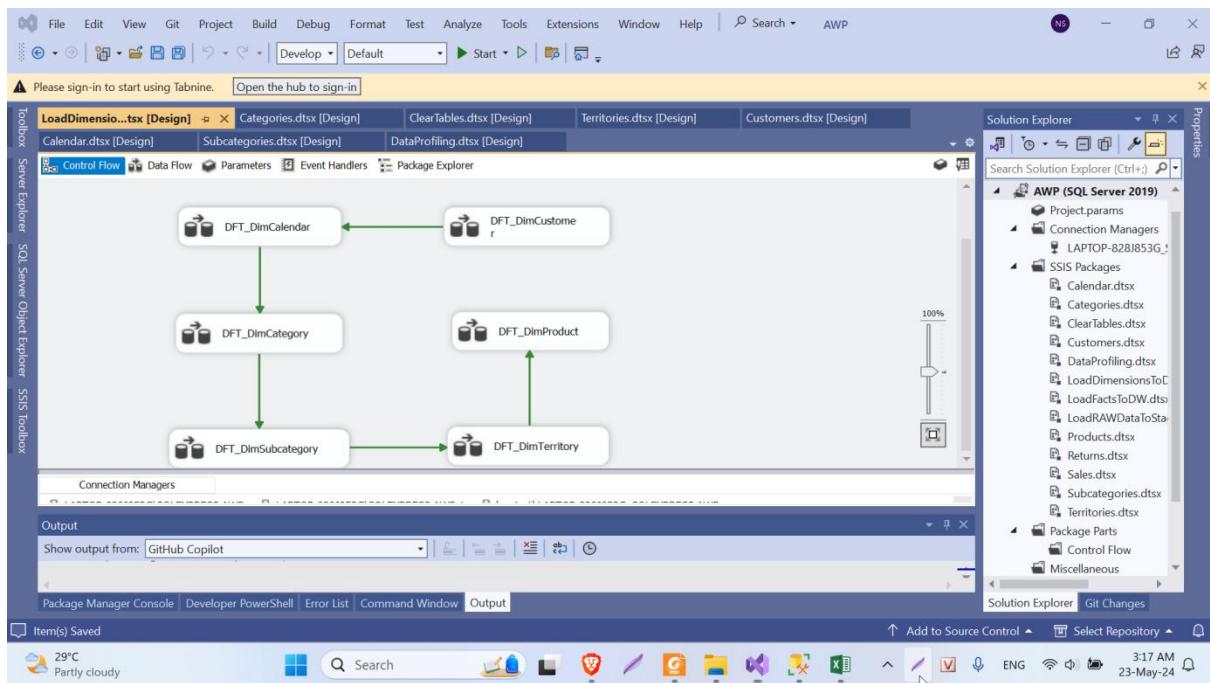
1. Từ Visual Studio menu, chọn File -> New -> Project.

2. Từ dialog:

a. Chọn **Integration Services Project**

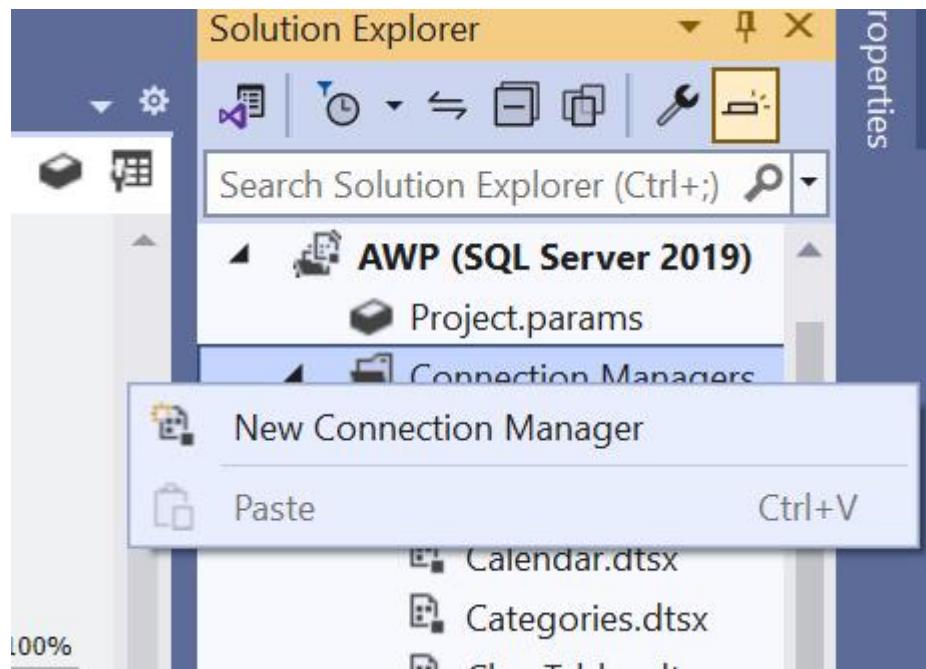
b. Đặt tên: **Hardware Integration Services**

3. Click **OK**

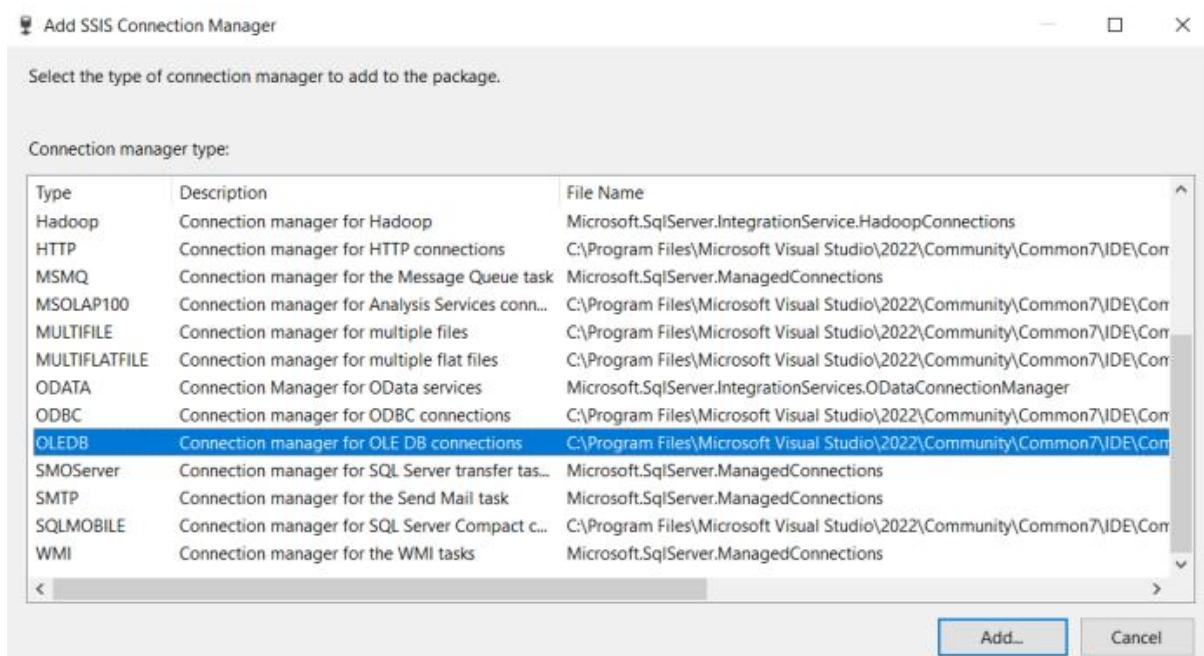


3.2. Thiết lập các Connection

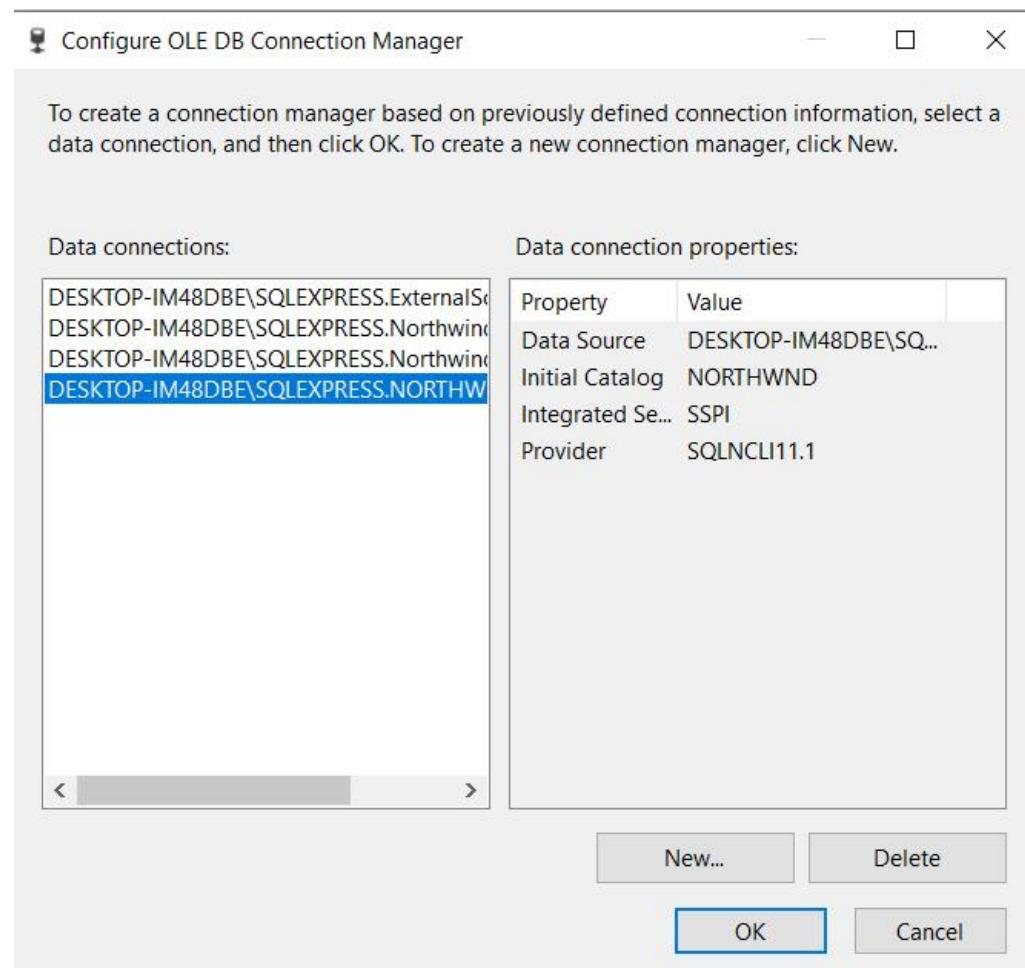
1. Tại **solution explorer**, right-click vào **Connection Managers**, chọn **New Connection Manager**.



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



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



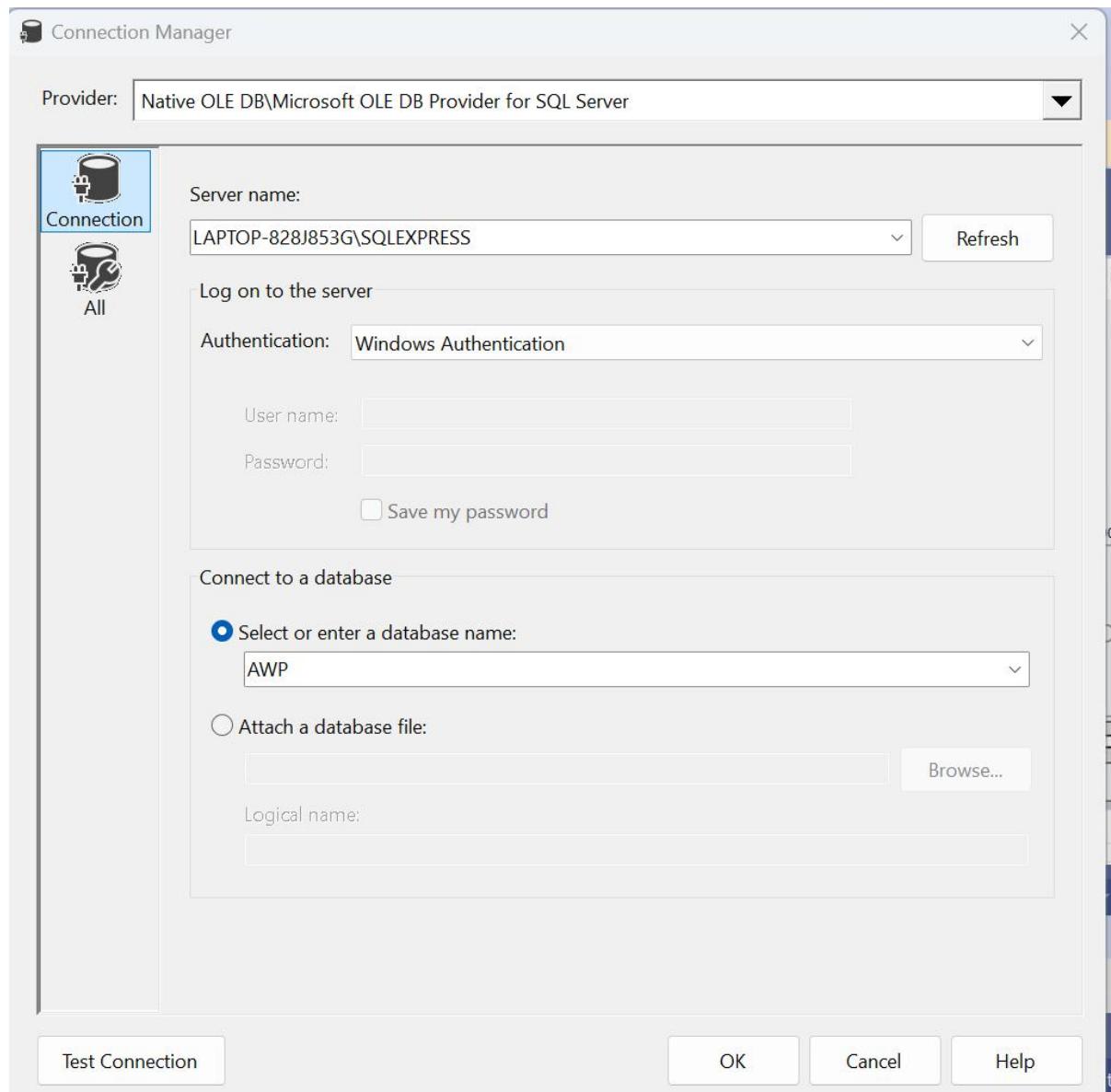
Đặt tên cho các thông tin như bên dưới:

Provider: **SQL Server Native Client**

Server Name: **LAPTOP-828J853G\SQLEXPRESS**

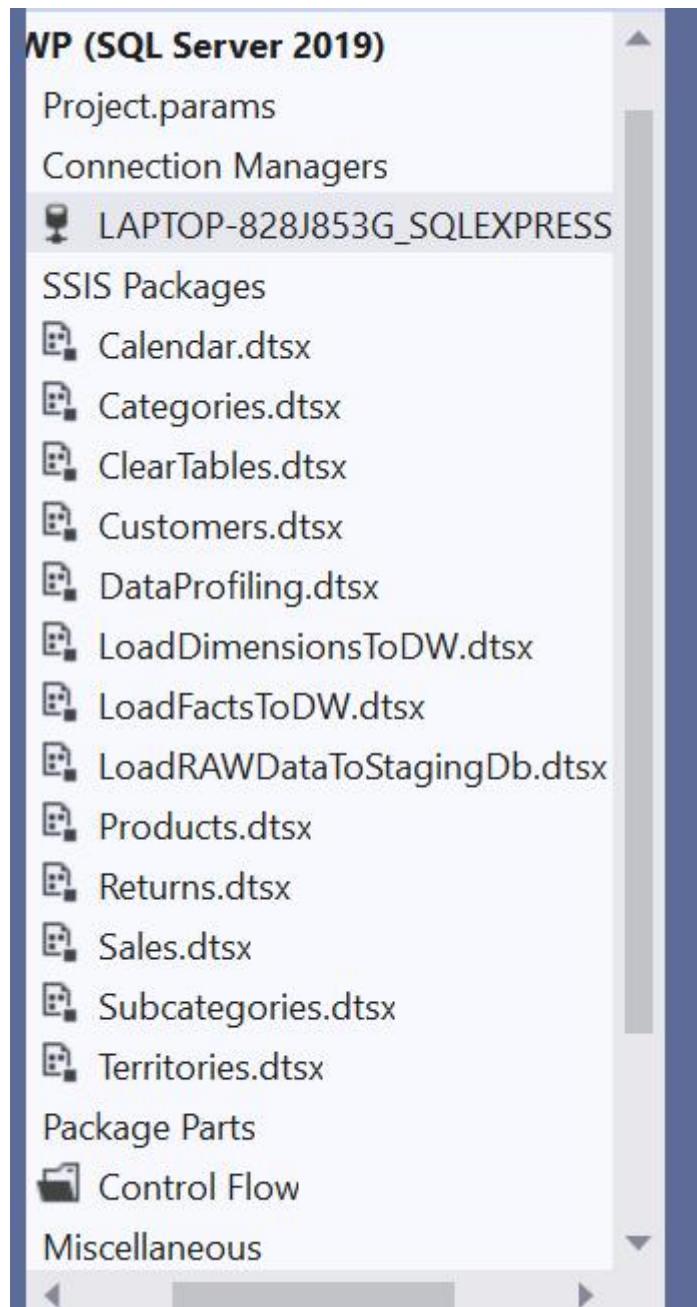
Log on to the Server: **Use Windows Authentication**

Connect to a database: **AWP**



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 **DESKTOP-IM48DBE_MSSQLDEV.DataWareHouse.conmgr** vừa được tạo:



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

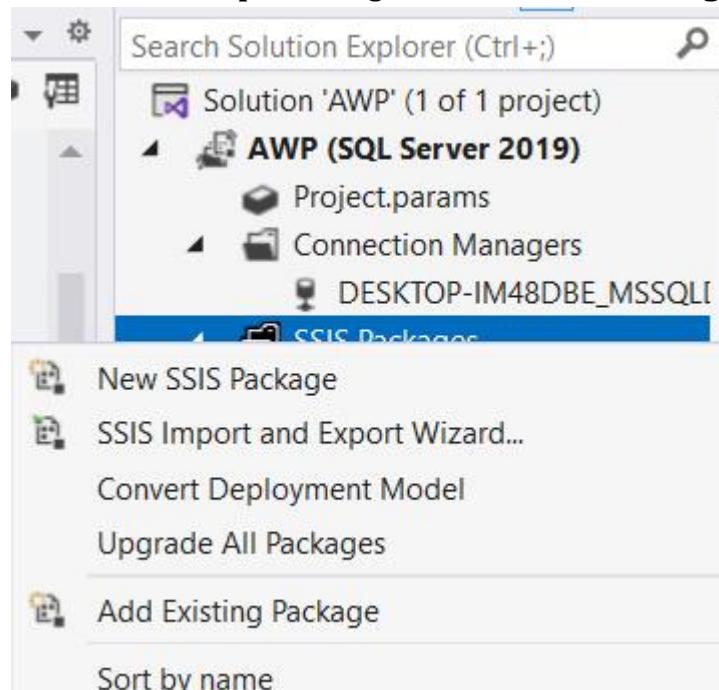
3.3. Import Dữ liệu vào các bảng stage:

3.3.1. Customer

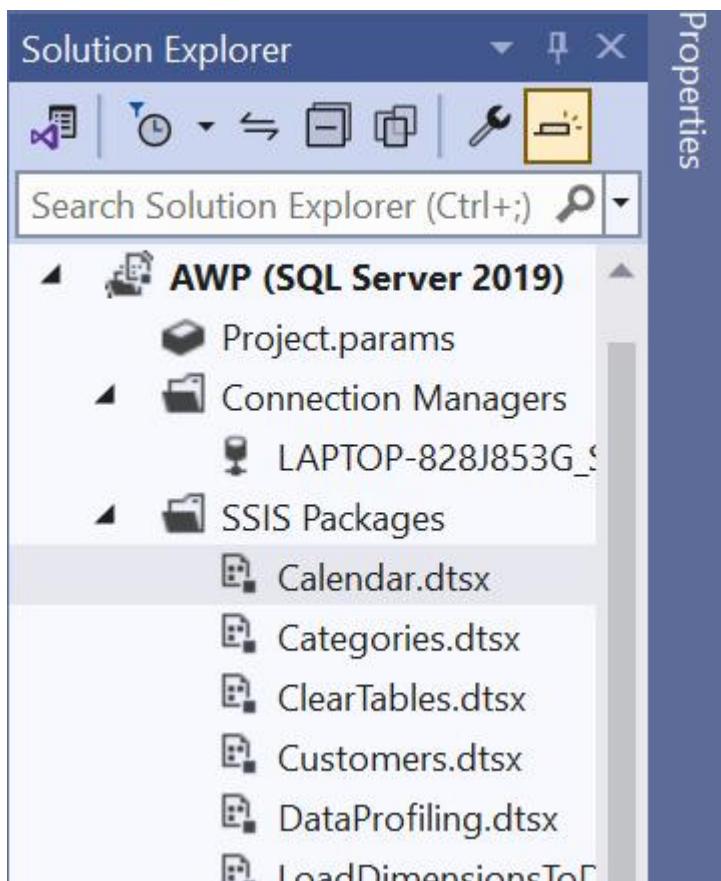
3.3.1.1: Tạo Package

Tạo một package mới.

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



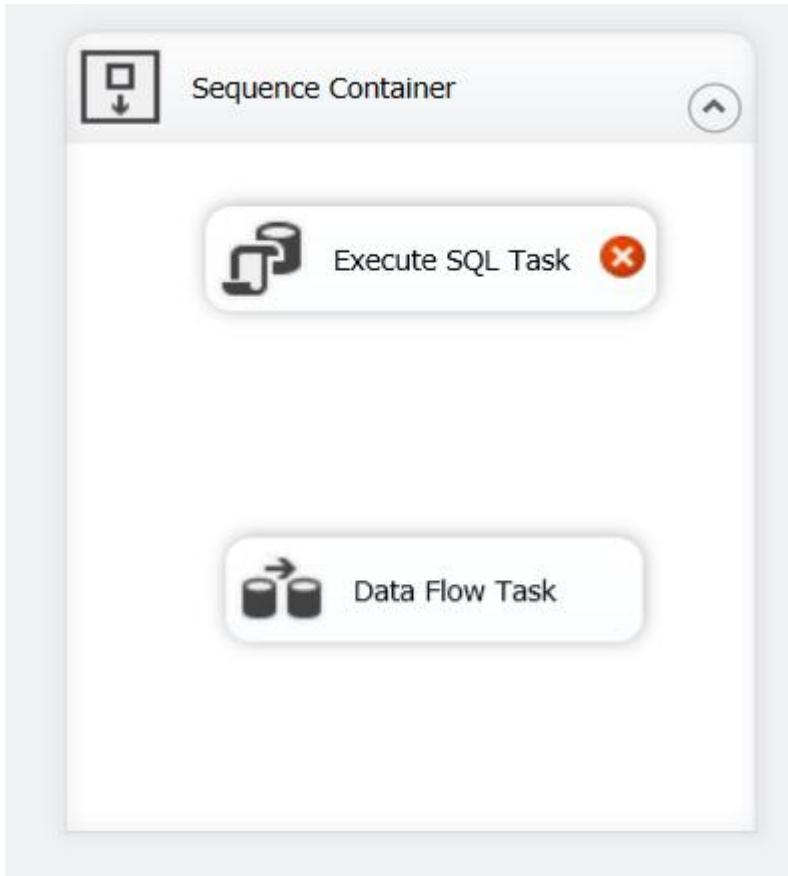
2. Right-Click vào package có tên **Package1.dtsx** và chọn **Rename**
3. Đặt tên package là **Calendar**
4. Sau khi thực hiện xong ta được **Package.dtsx** và **Calendar.dtsx** như hình sau:



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

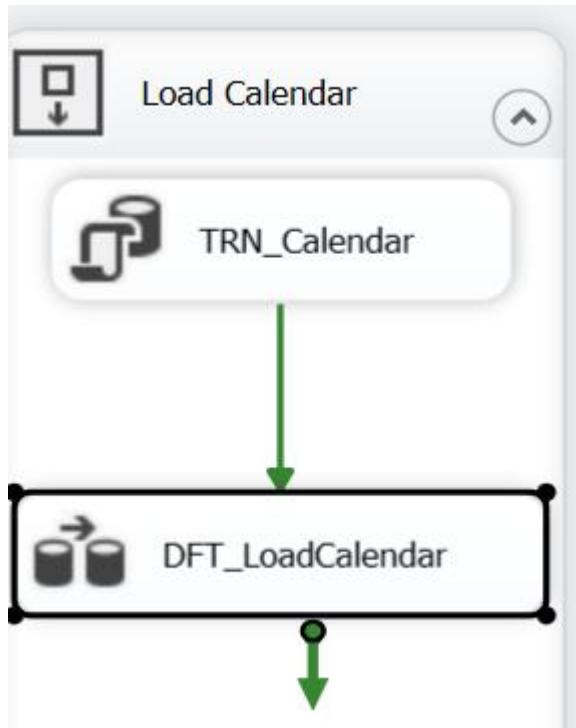
3.3.1.2: Thiết lập Control Flow

1. Từ SSIS Toolbox tạo 1 **Sequence Container**, 1 **Execute SQL Task** và 1 **Data flow tasks** vào surface design như hình sau:

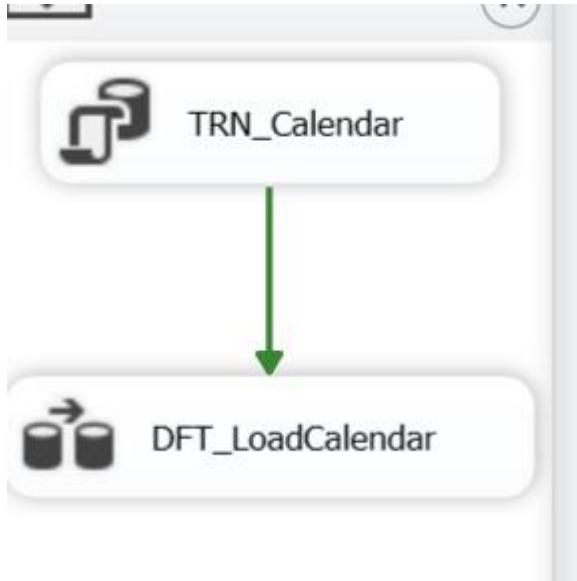


2. Đặt tên cho các task.

- a. Sequence Container - Load Calendar
- b. Execute SQL Task - TRN_Calendar
- c. Data flow tasks - DFT_LoadCalendar

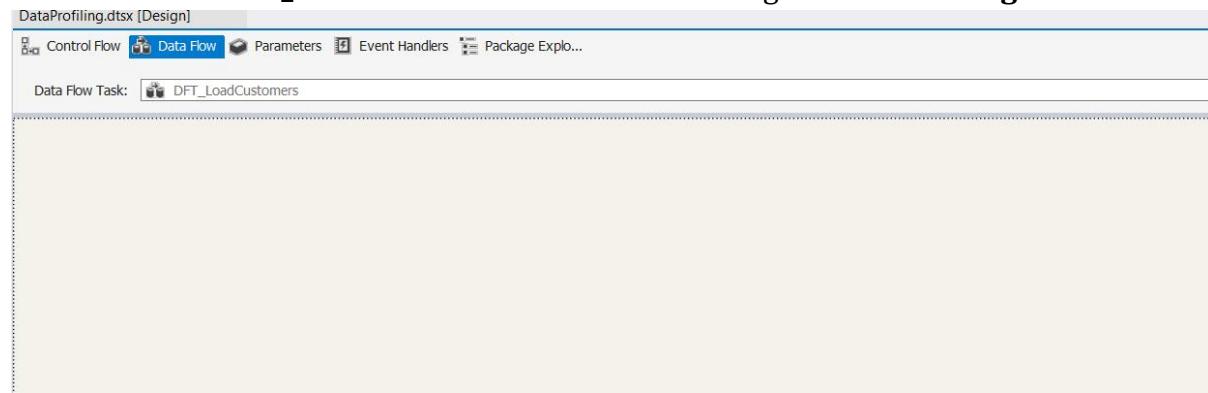


3. Connect các **tasks** để chúng có thể chạy theo thứ tự. Click vào **TRN_Calendar** task và một mũi tên xanh sẽ xuất hiện. Kéo rê nó vào **DFT_LoadCalendar** task.



3.3.1.3: Load dữ liệu từ Source vào Stage

1. Double-click **DFT_LoadCalendar** task để mở nó trong **Data Flow design surface**:



2. Từ **SSIS Toolbox**, kéo rê **Flat File Source** vào design surface

Click **OK**.

3. Flat File Source sẽ xuất hiện như sau:

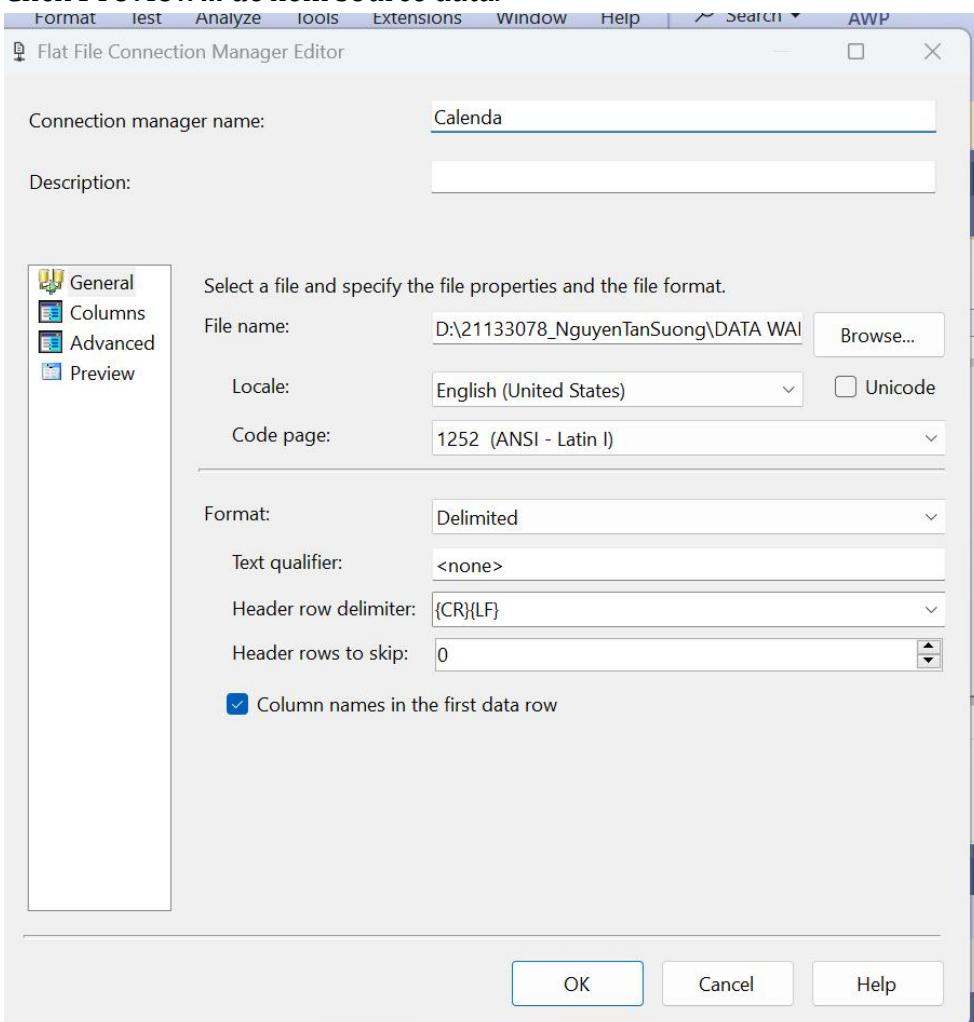


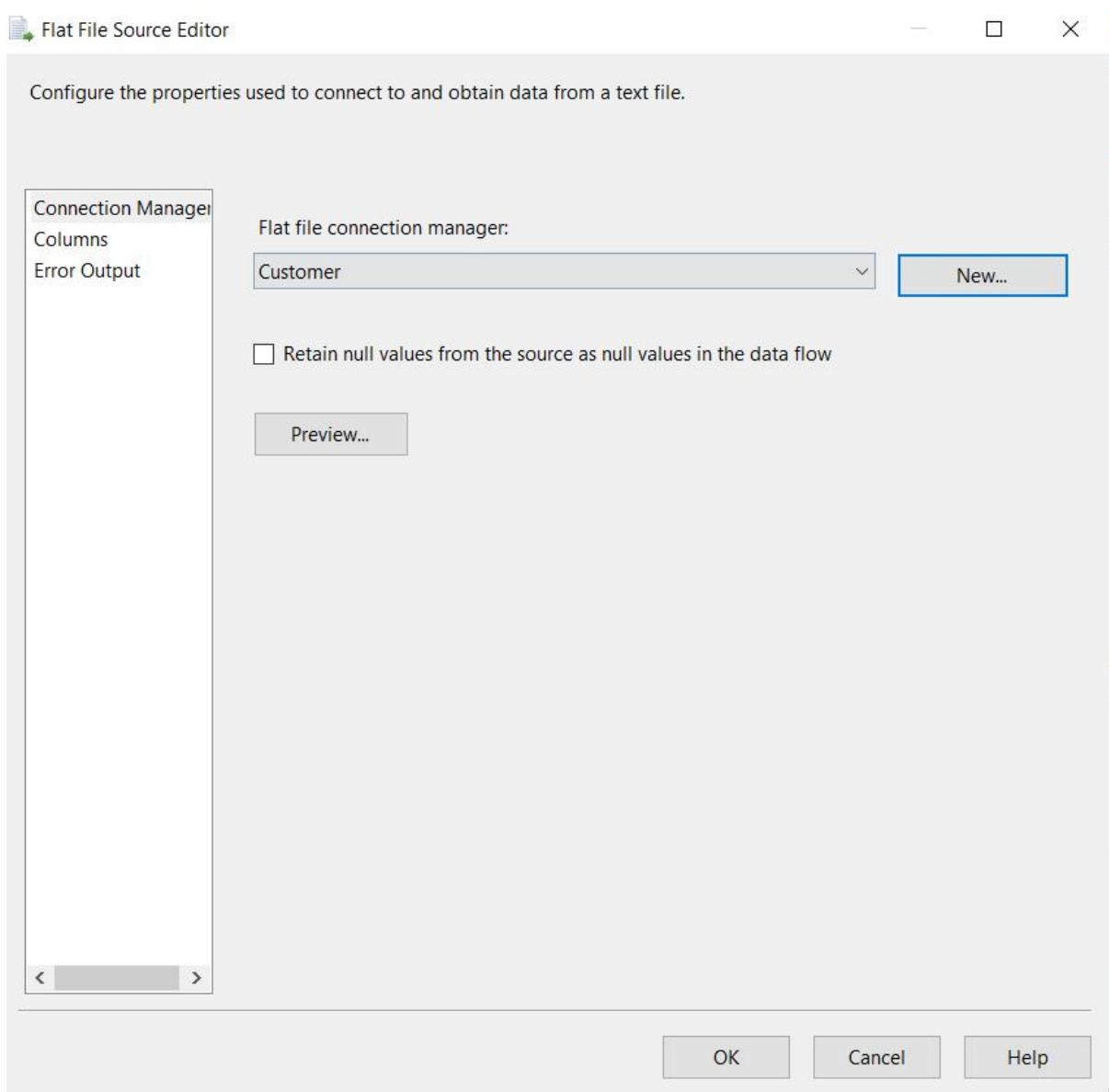
4. Đổi tên nó thành FFS_Customer



5. Double click vào source để cấu hình nó, **Flat File Source Editor** sẽ mở ra. Chọn **New**, **Flat File Connection Manager Editor** hiện ra . Chọn **FileName Customer**. Chọn **Column** sau đó chọn **OK**.

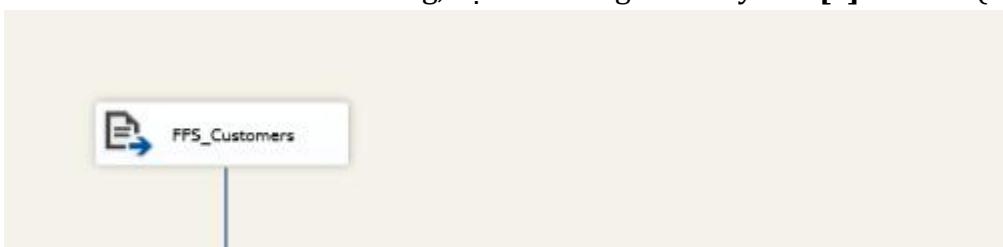
Click **Preview...** để xem source data.





Click **OK** để lưu.

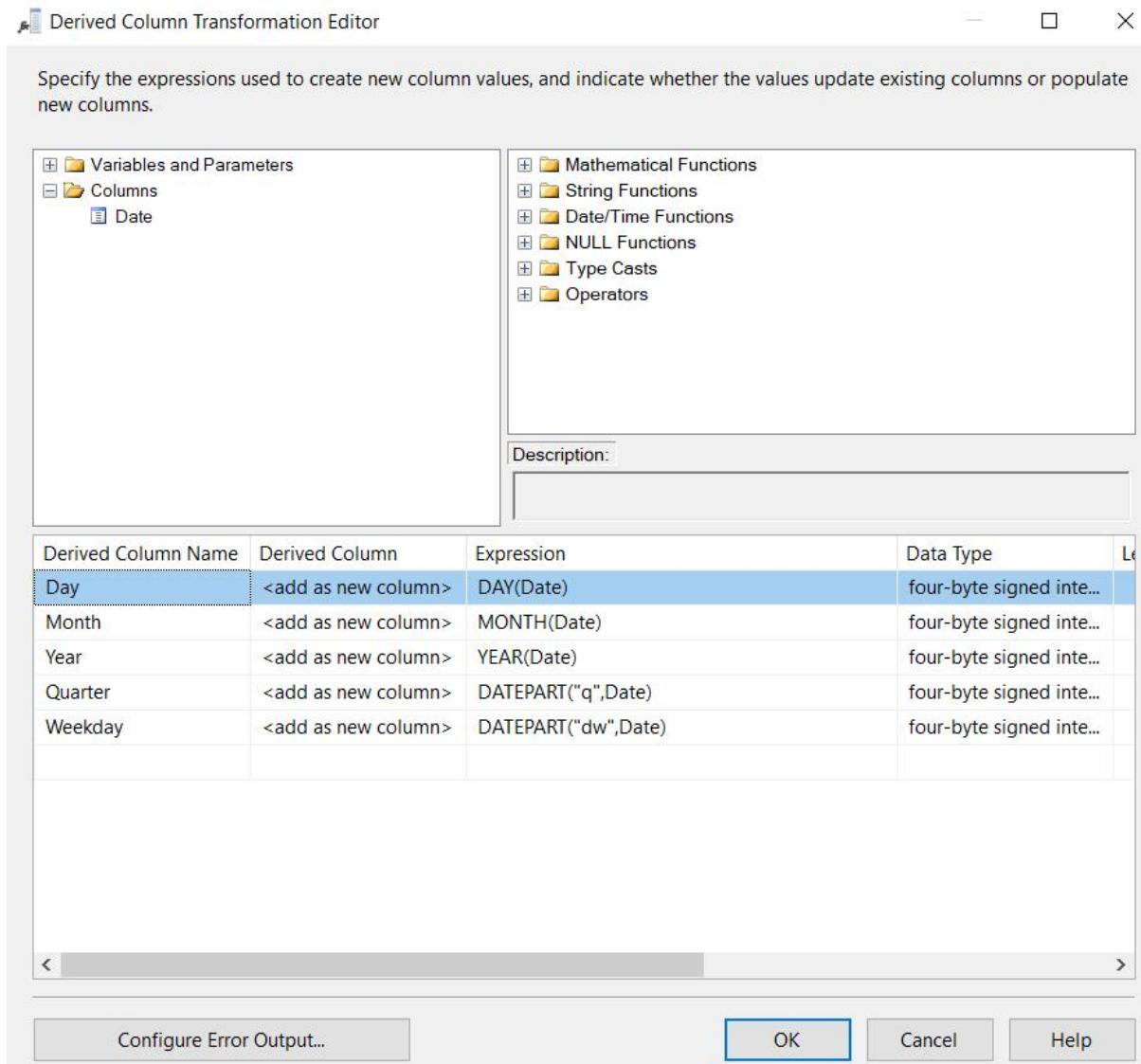
6. Sau khi cấu hình **source** xong, bạn sẽ không còn thấy dấu [x] màu đỏ (như bên dưới):



7. Tiếp tục, cấu hình destination. Từ **SSIS Toolbox**, kéo rê **Row Count**, **Derived Column** vào design surface. Đổi tên **Derived Column** thành **DRV_AddDateparts**.

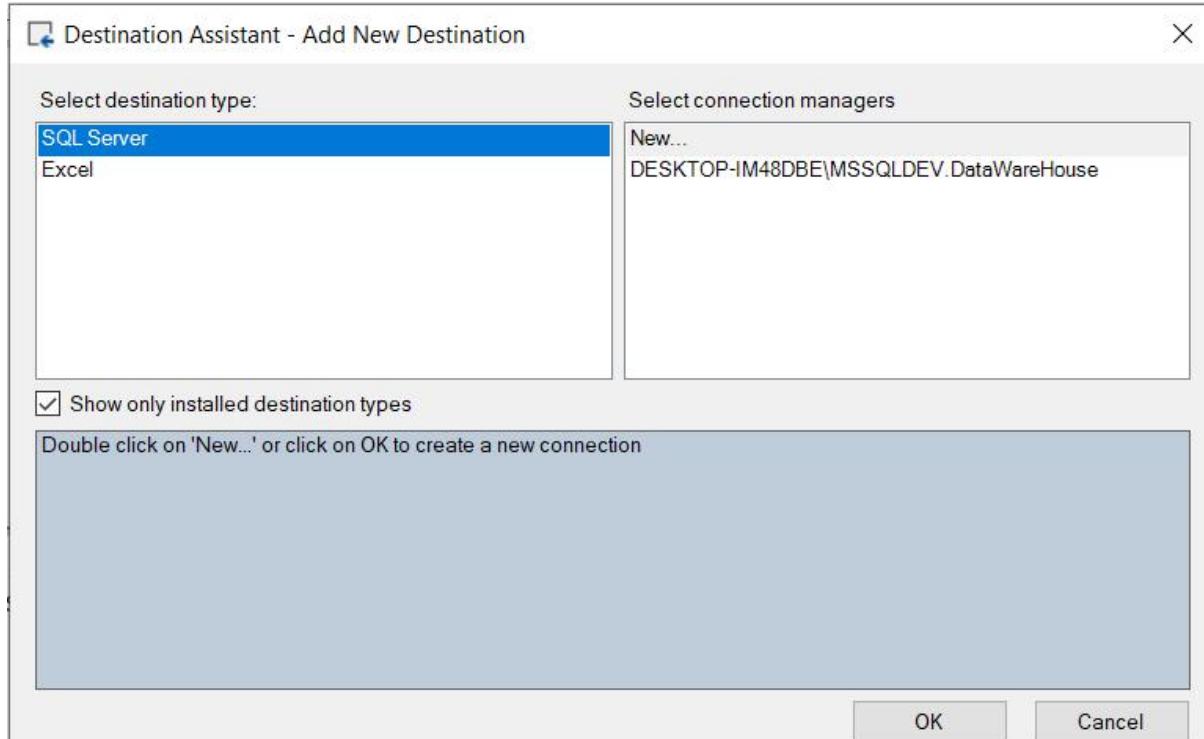


8. Double click vào **DRV_AddDateparts**. Kéo rê cột Date vào ô Derived Column Name, ta được Derived Column 1, đổi tên lại thành Day và thay đổi Expression. Làm tương tự để tạo thành các cột mới là Month, Year, Quarter, Weekday.



Click OK

9. Tiếp theo, cấu hình destination. Từ **SSIS Toolbox**, kéo rê **Destination Assistant** vào design surface, xuất hiện dialog



10. Chọn Connection DataWarehouse rồi click OK

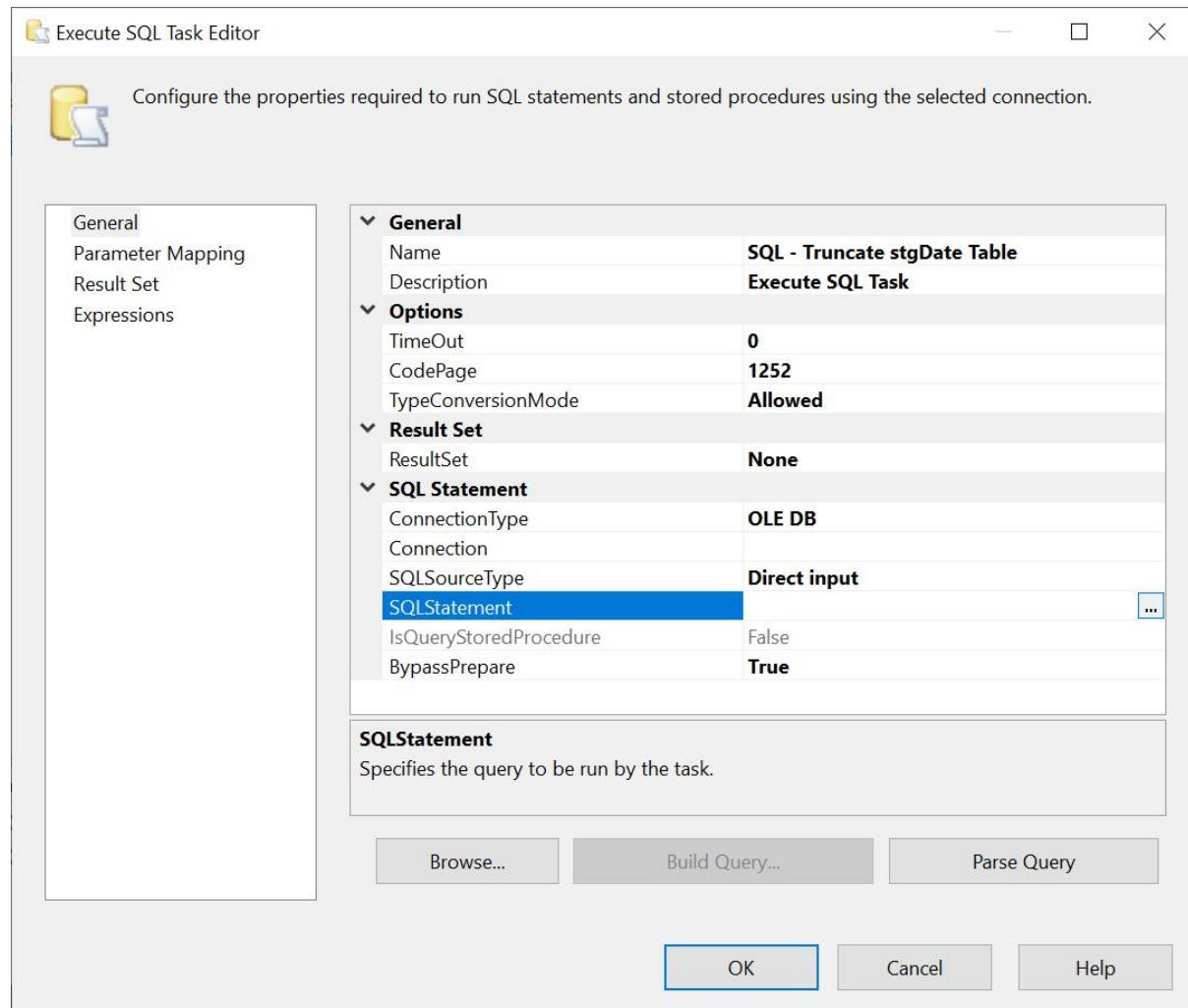
11. Đổi tên **Destination Assistant** thành **OLE_Calendar**, ta được design surface:



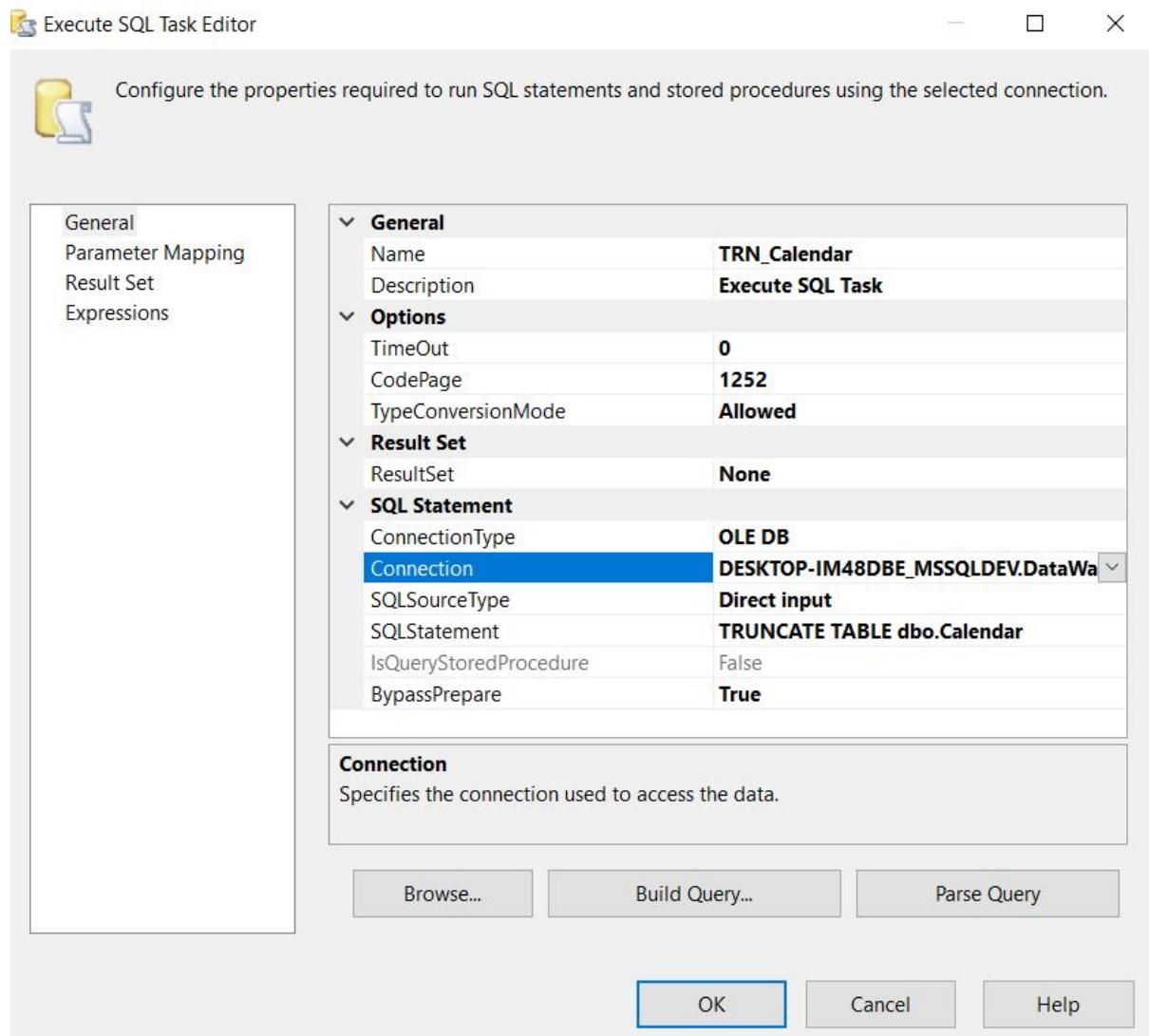
3.3.1.4: Thiết lập truncate table SQL task

Click **Control Flow** tab.

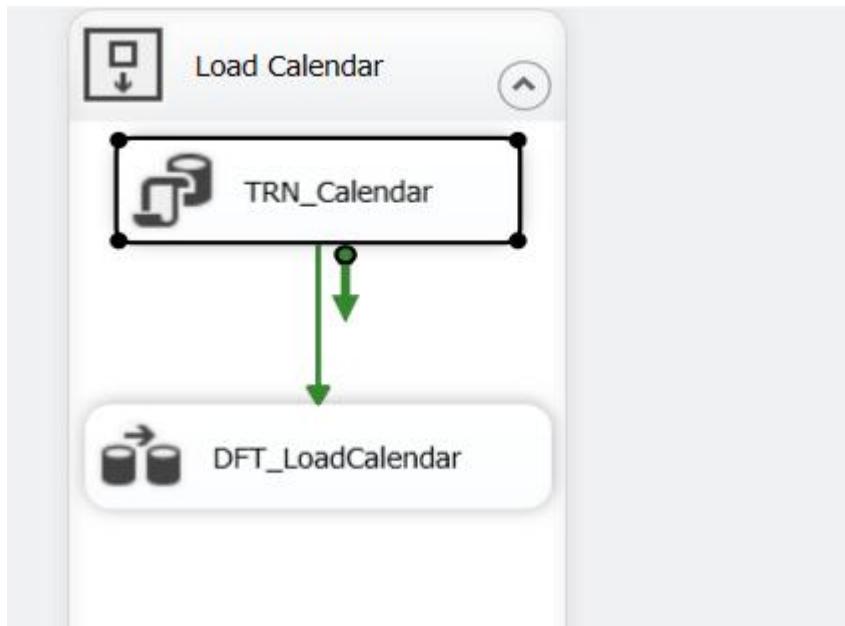
1. Double-click vào **TRN_Calendar** task để cấu hình nó, **Execute SQL Task Editor** sẽ mở ra.



2. Ở mục **SQL Statement**, chọn **Connection** là **DESKTOP-IM48DBE_MSSQLDEV.DataWareHouse** và **SQLStatement** là **TRUNCATE TABLE dbo.Calendar**



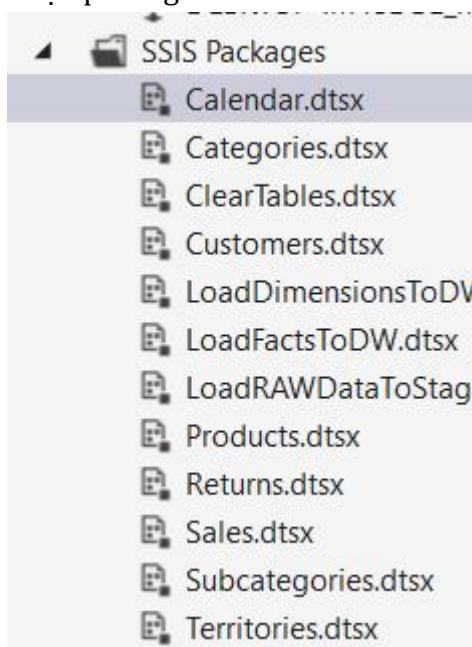
3. Click **OK** và trở lại màn hình **Control Flow**



3.3.2.Product

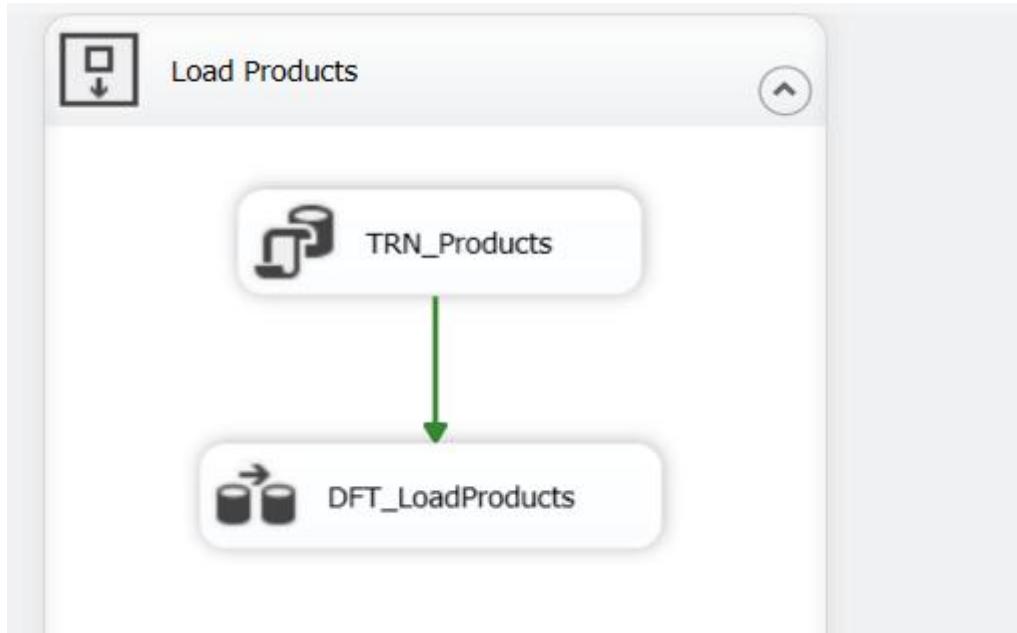
3.3.2.1. Tạo Package

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

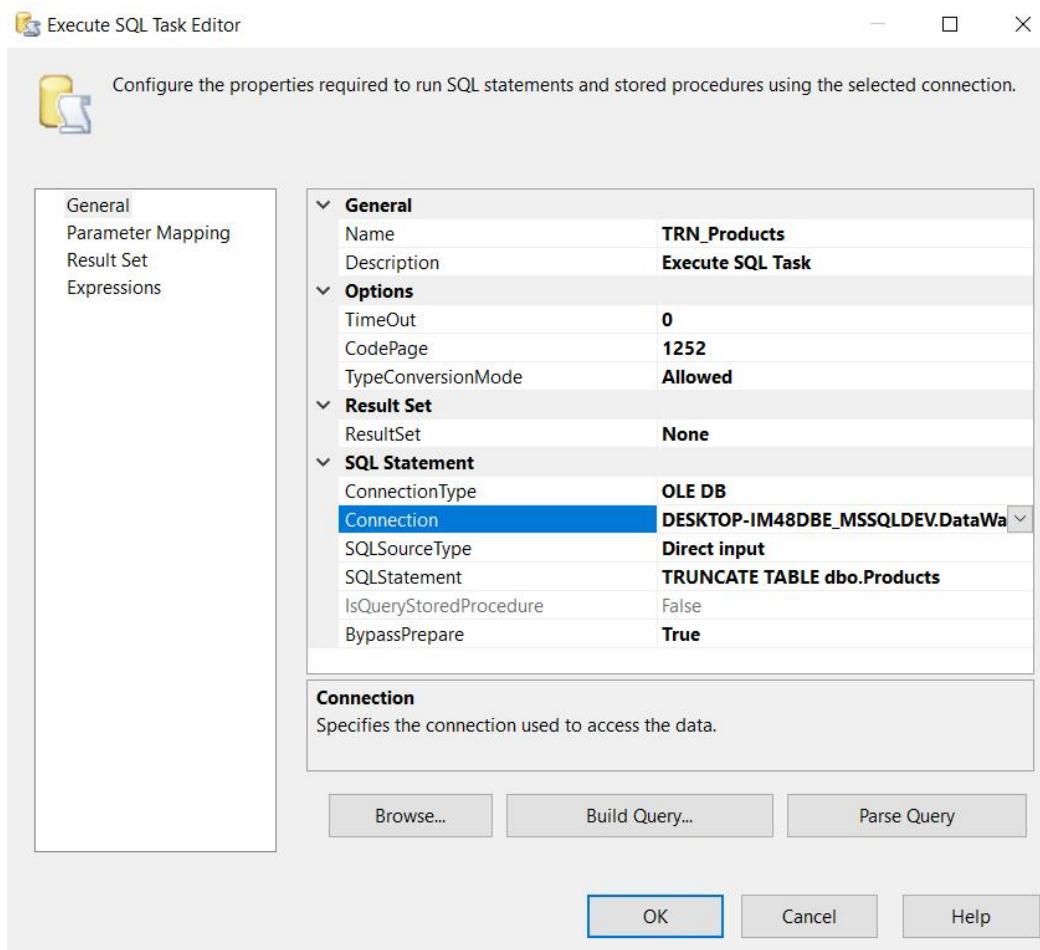


3.3.2.2. Thiết lập Control Flow

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

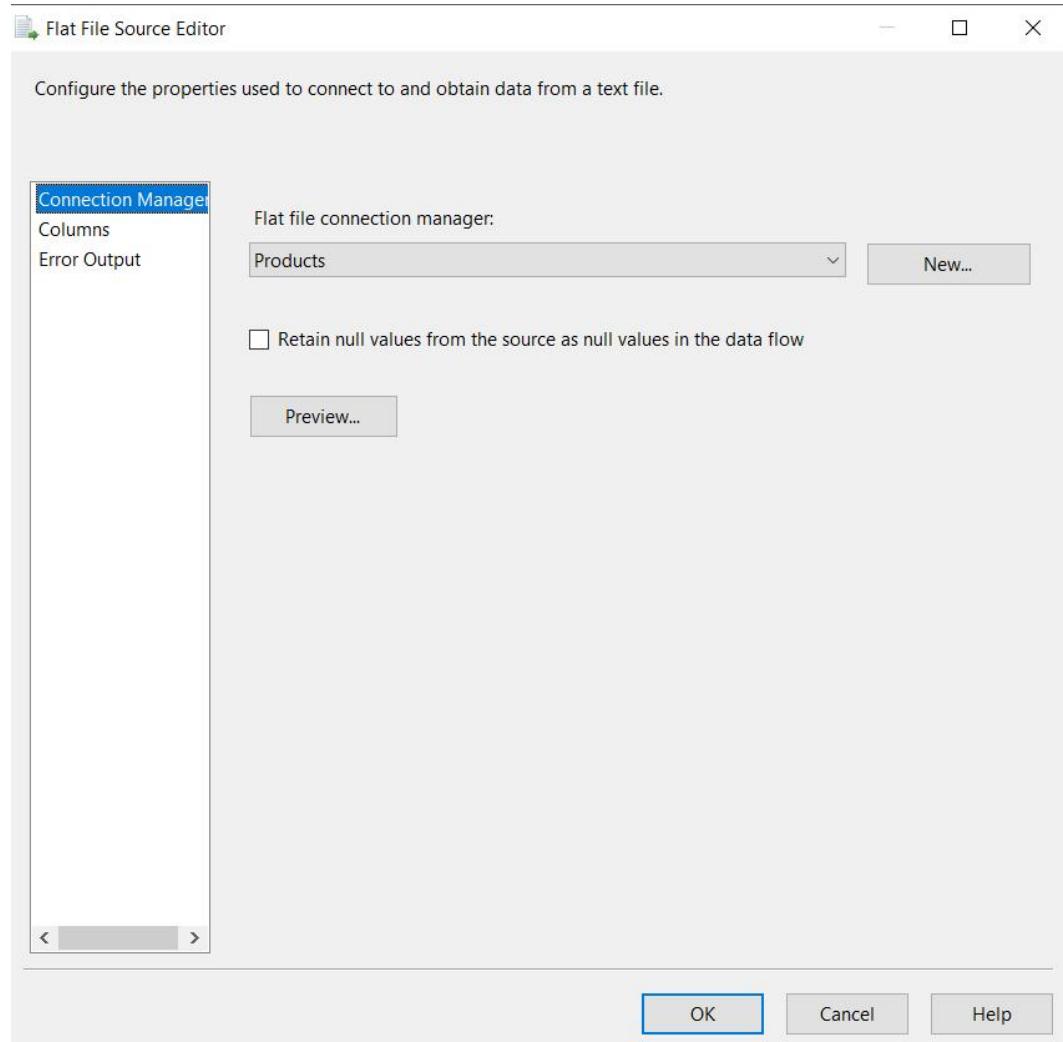


- Thiết lập truncate table SQL task

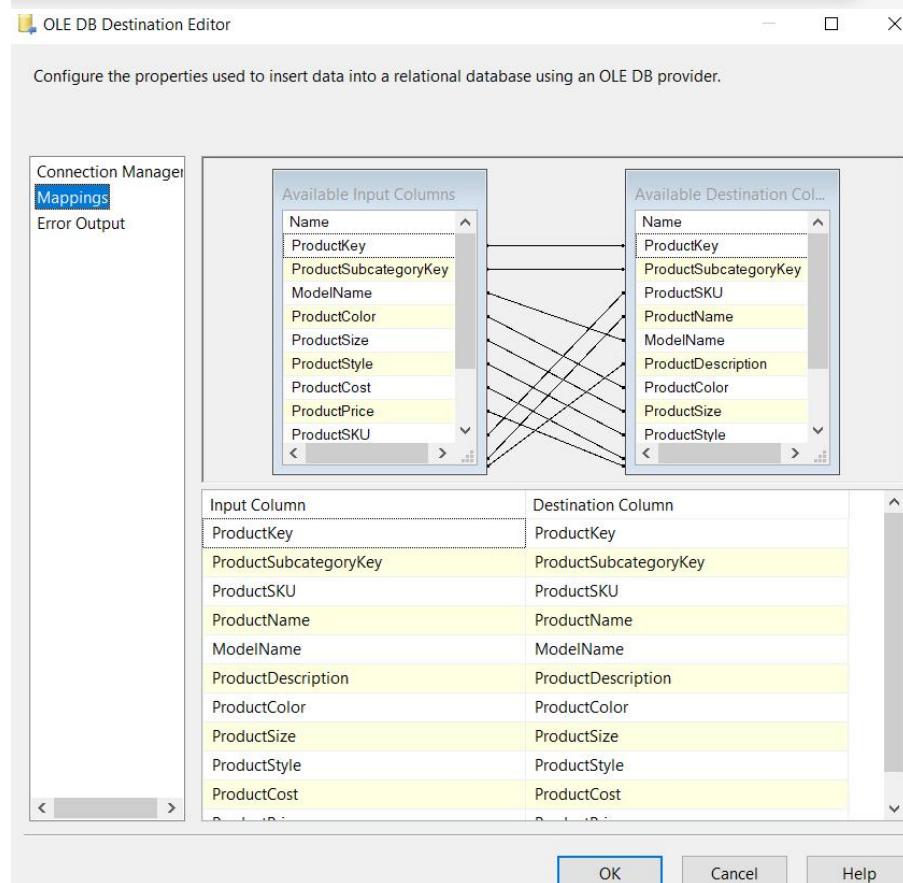
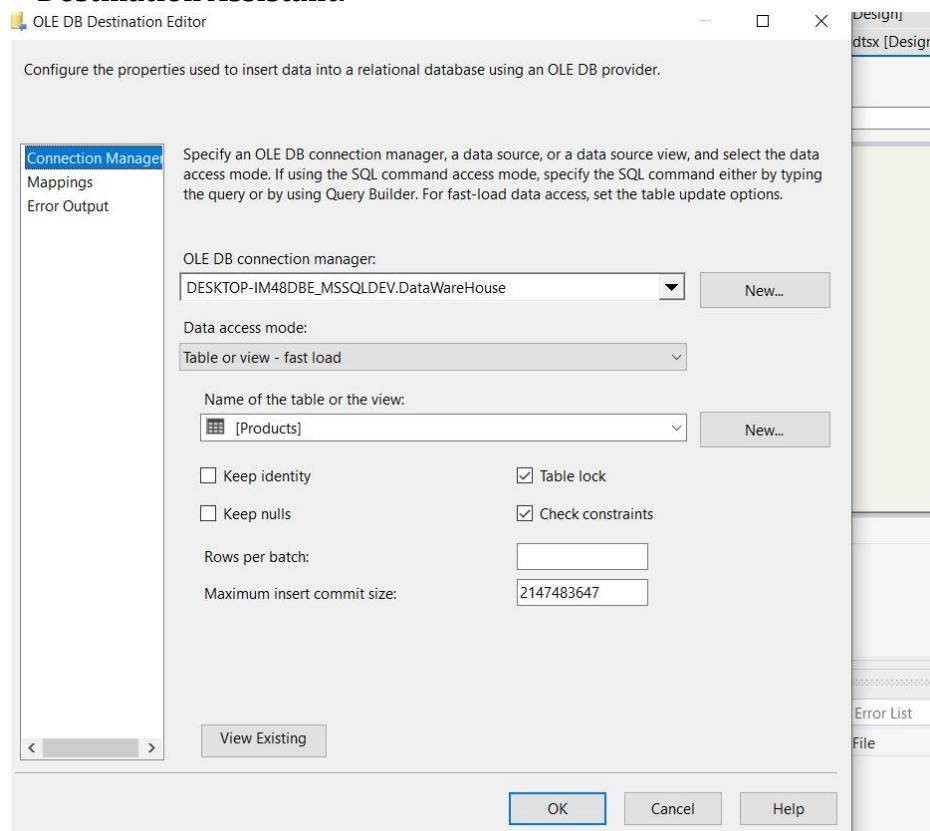


3.3.2.3. Load dữ liệu từ Source vào Stage

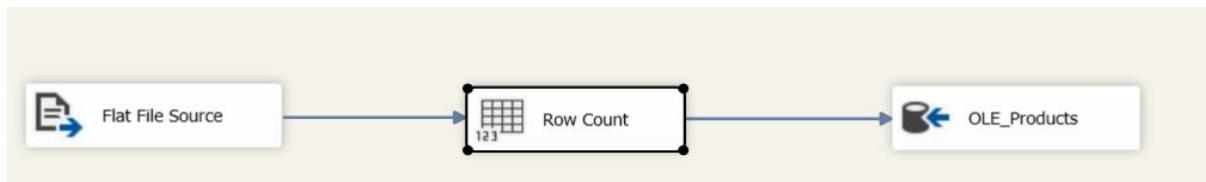
- Double-click **DFT_LoadProducts** task để mở nó trong **Data Flow design surface**:
- Tạo **Flat File Source, Destination Assistant, Row Count** cấu hình chúng như sau:
 - + **Flat File Source:**



+ Destination Assistant:



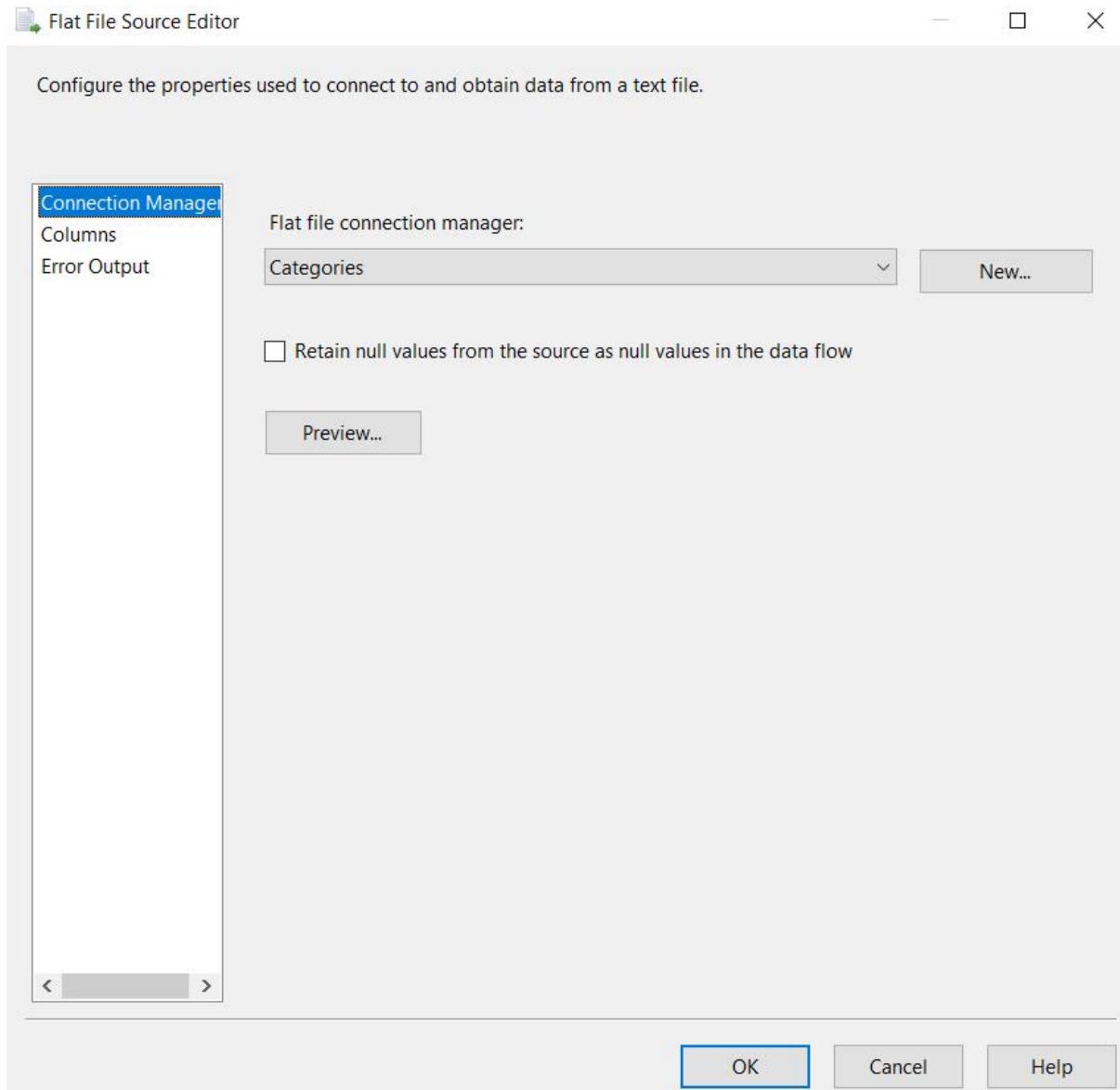
+ Sau khi cấu hình xong ta được:



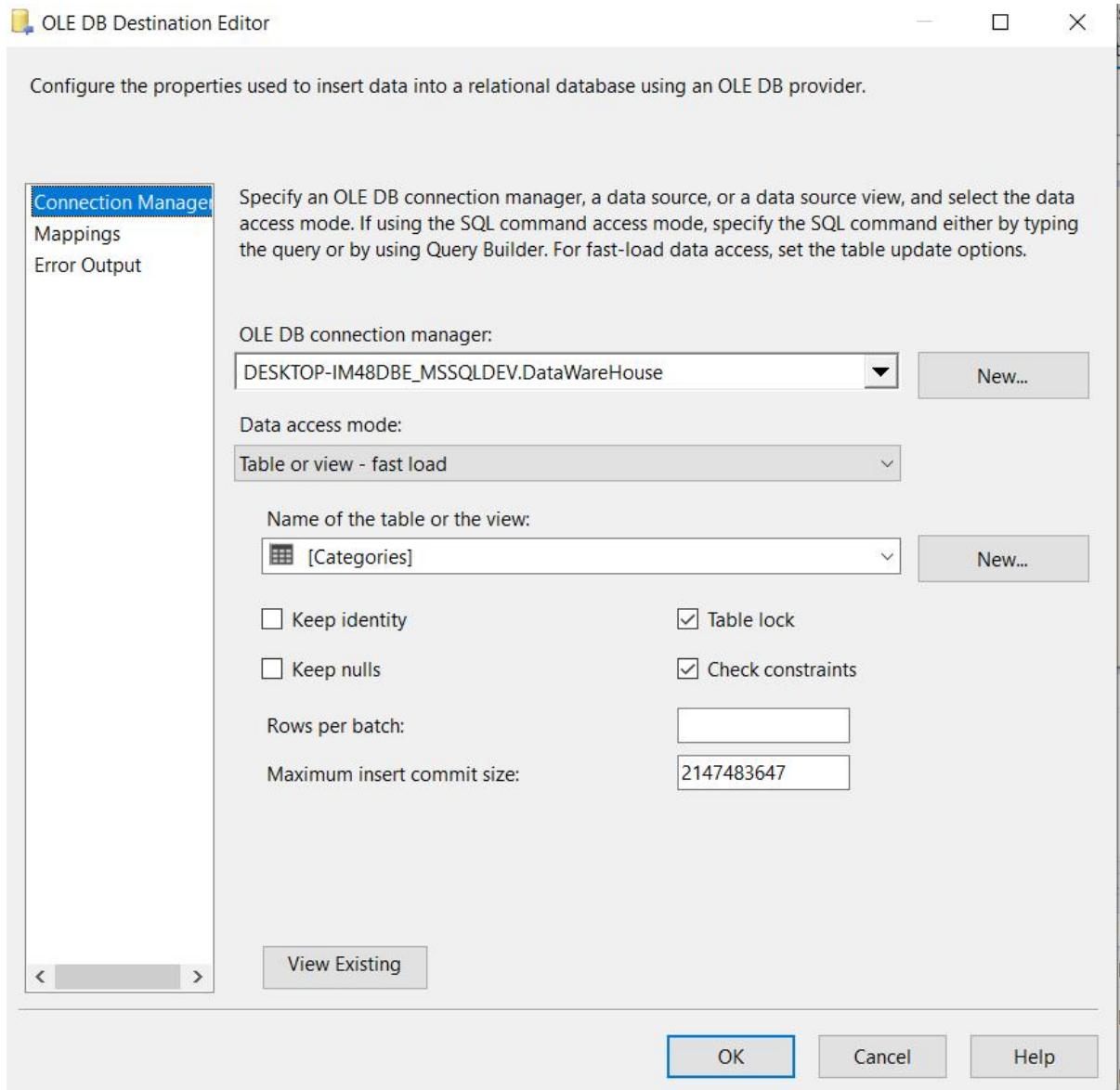
3.3.3.Categories

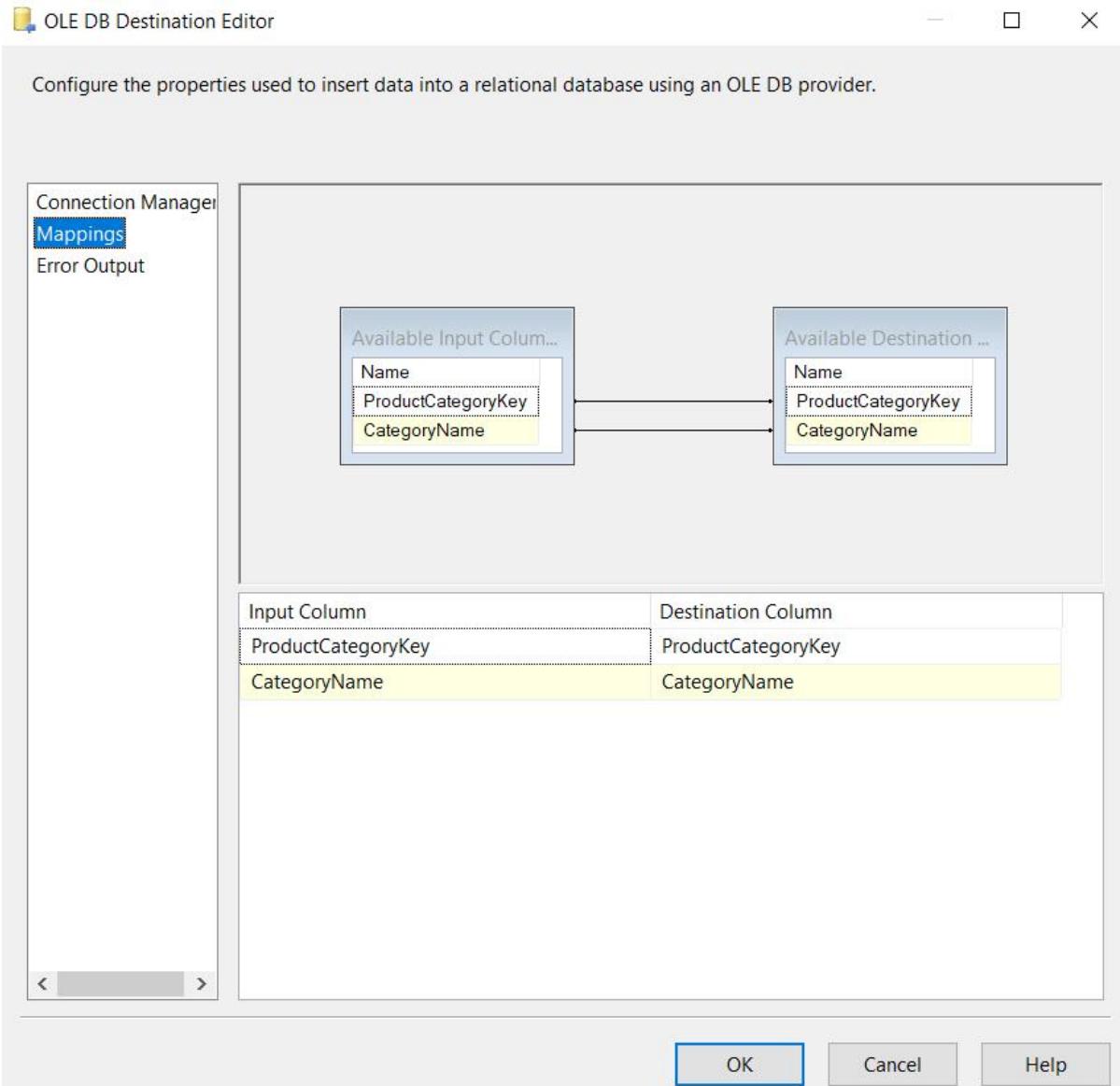
3.3.3.1. Load dữ liệu từ Source vào Stage

- Double-click DFT_LoadCategories task để mở nó trong Data Flow design surface:
- Tạo Flat File Source, RowCount và Destination Assistant và cấu hình chúng như sau:
 - + Flat File Source



+ Destination Assistant:





+ Sau khi cấu hình xong ta được:



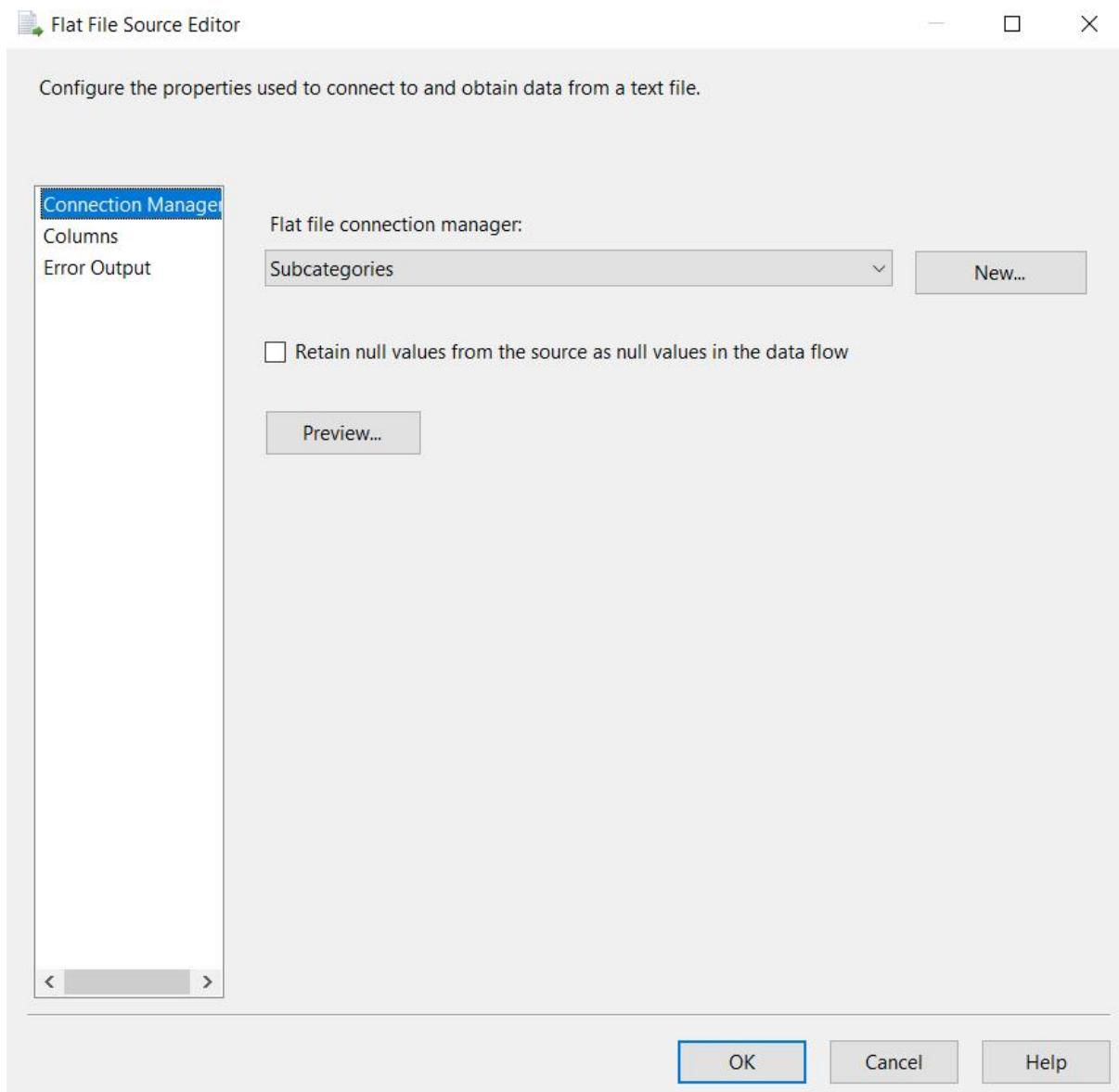
3.3.4. SubCategory

3.3.4.1. Load dữ liệu từ Source vào Stage

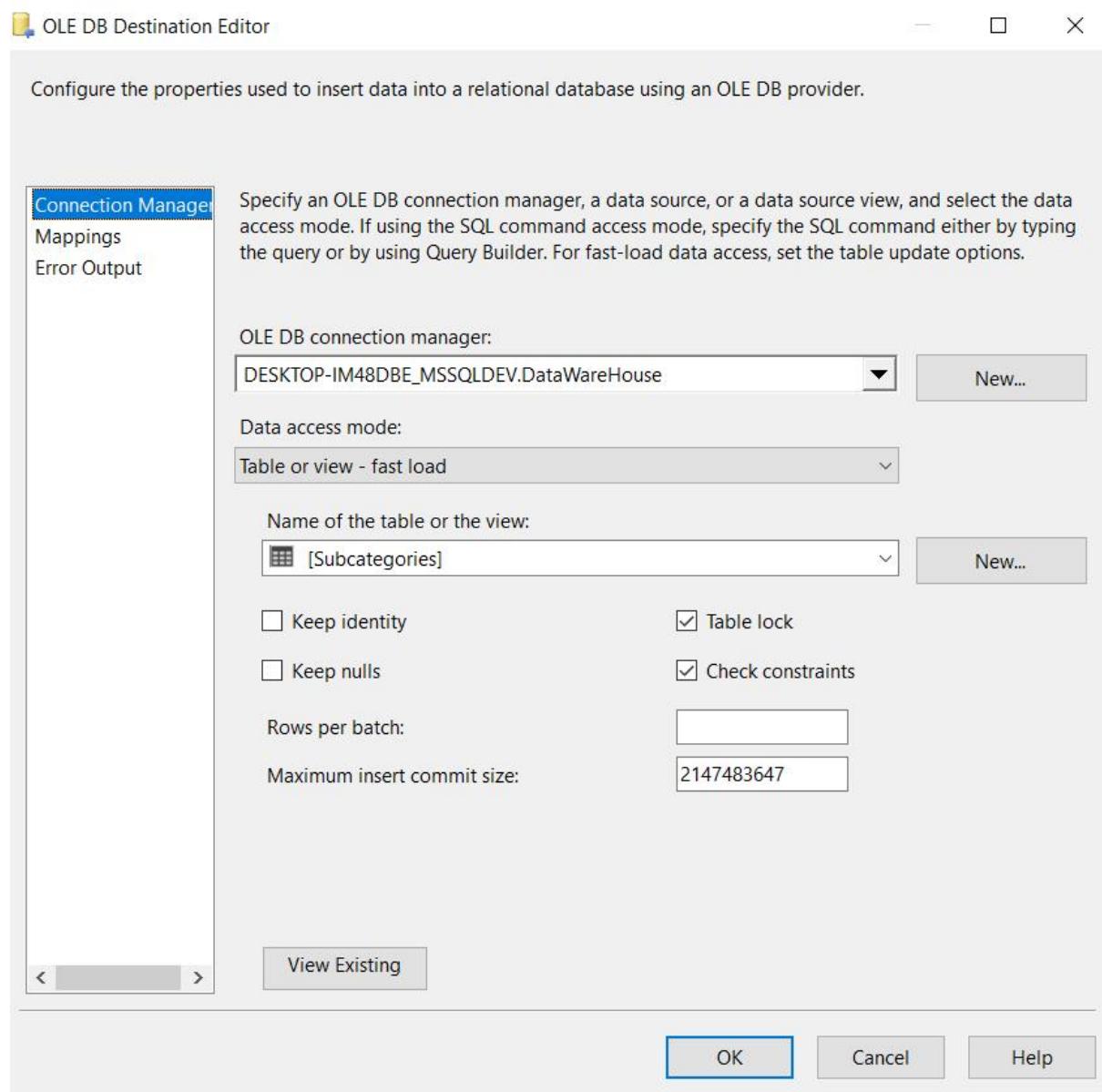
- Double-click **DFT_LoadSubcategories** task để mở nó trong **Data Flow design surface**:

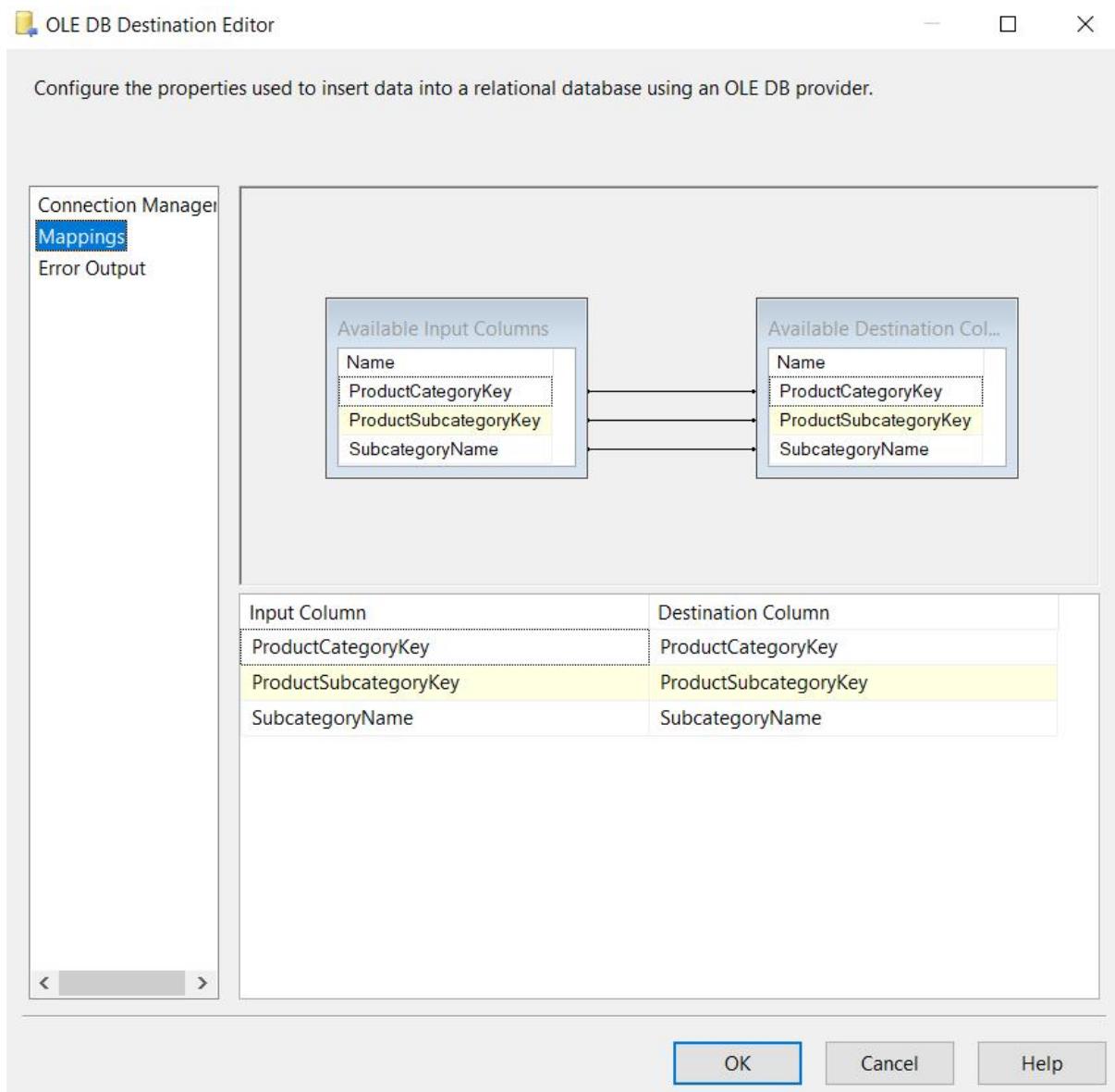
- Tạo **Flat File Source** và **Destination Assistant** và cấu hình chúng như sau:

+ **Flat File Source**:

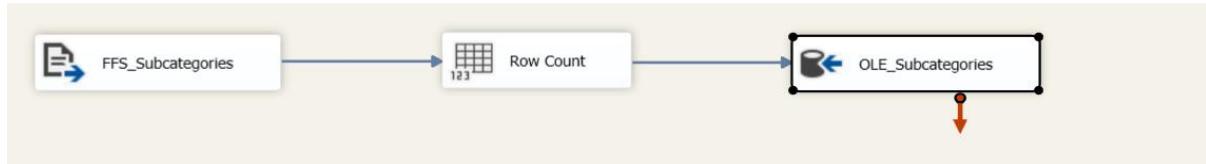


+ Destination Assistant:





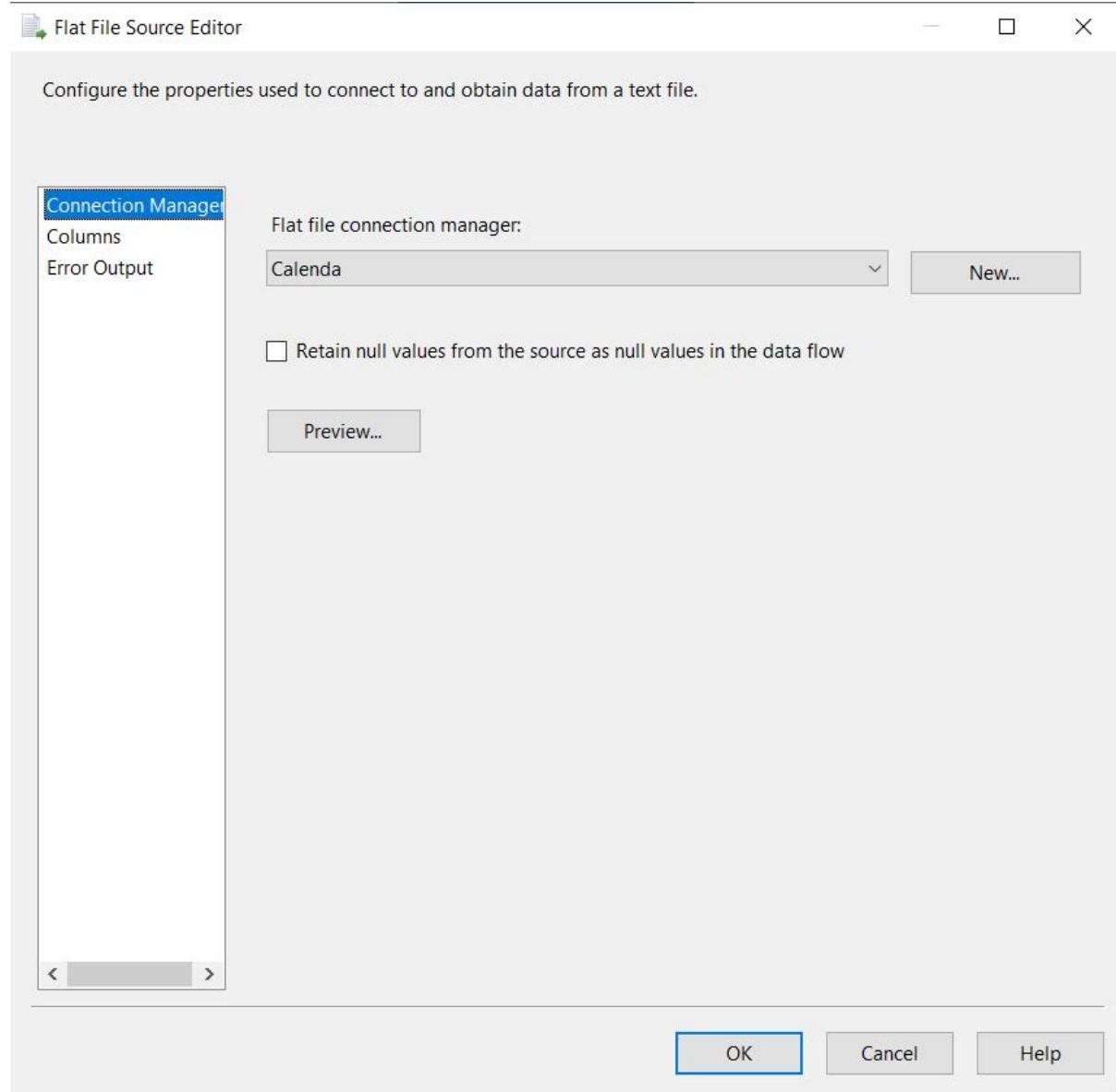
+ Sau khi cấu hình xong ta được:



3.3.5.Calendar

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 **Flat File Source**, **RowCount**, **Destination Assistant** và cấu hình chúng như sau:
- + **Flat File Source**:



+ Derived Column:

Derived Column Transformation Editor

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

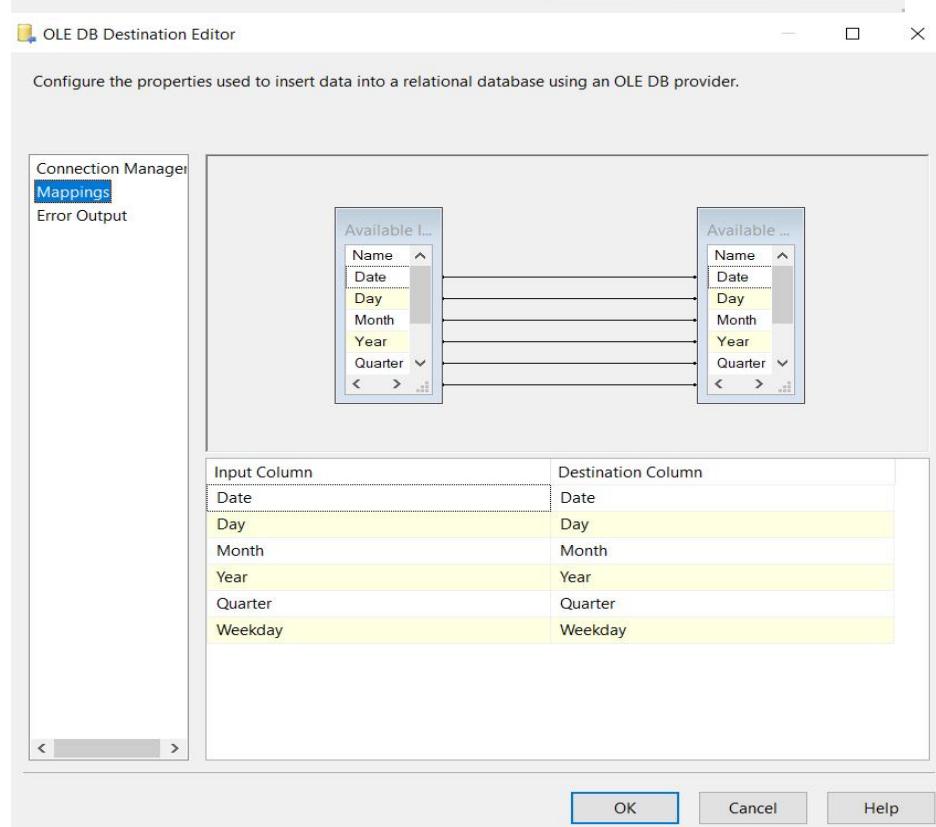
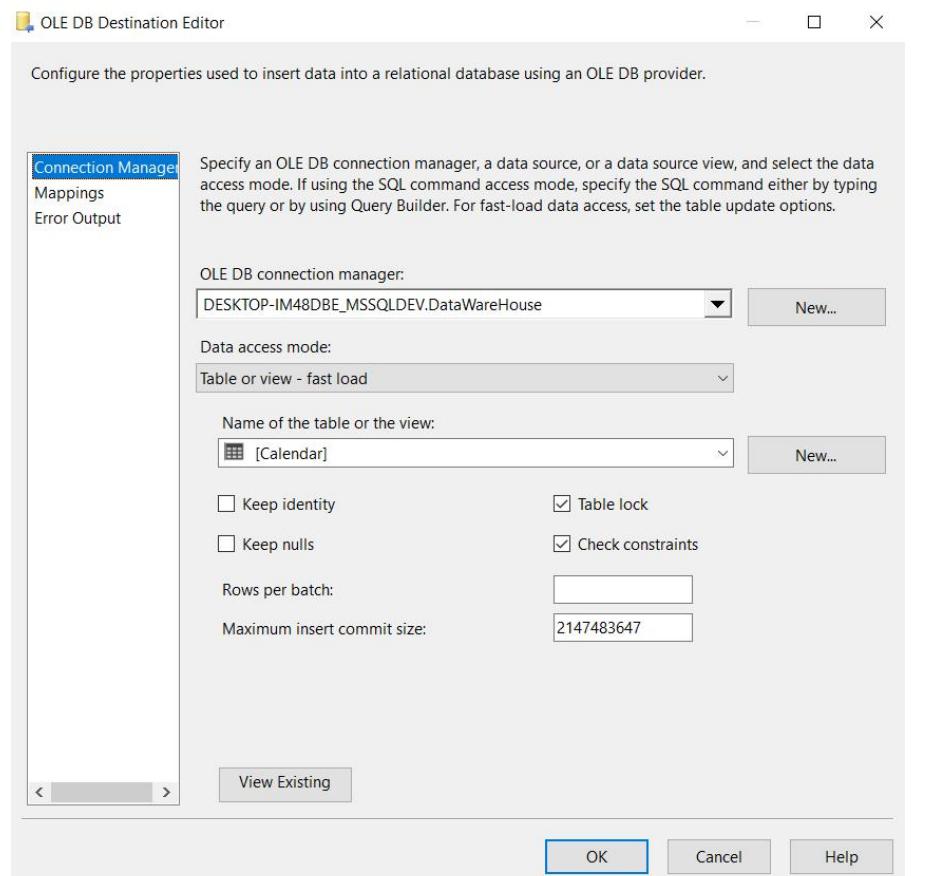
Variables and Parameters Mathematical Functions
Columns String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators

Description:

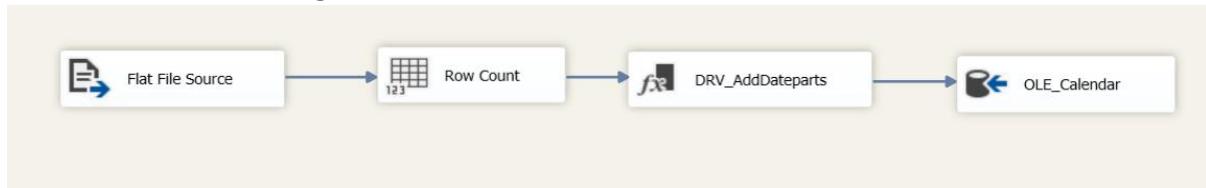
Derived Column Name	Derived Column	Expression	Data Type
Day	<add as new column>	DAY(Date)	four-byte signed integer
Month	<add as new column>	MONTH(Date)	four-byte signed integer
Year	<add as new column>	YEAR(Date)	four-byte signed integer
Quarter	<add as new column>	DATEPART("q",Date)	four-byte signed integer
Weekday	<add as new column>	DATEPART("dw",Date)	four-byte signed integer

Configure Error Output... OK Cancel Help

+ Destination Assistant:



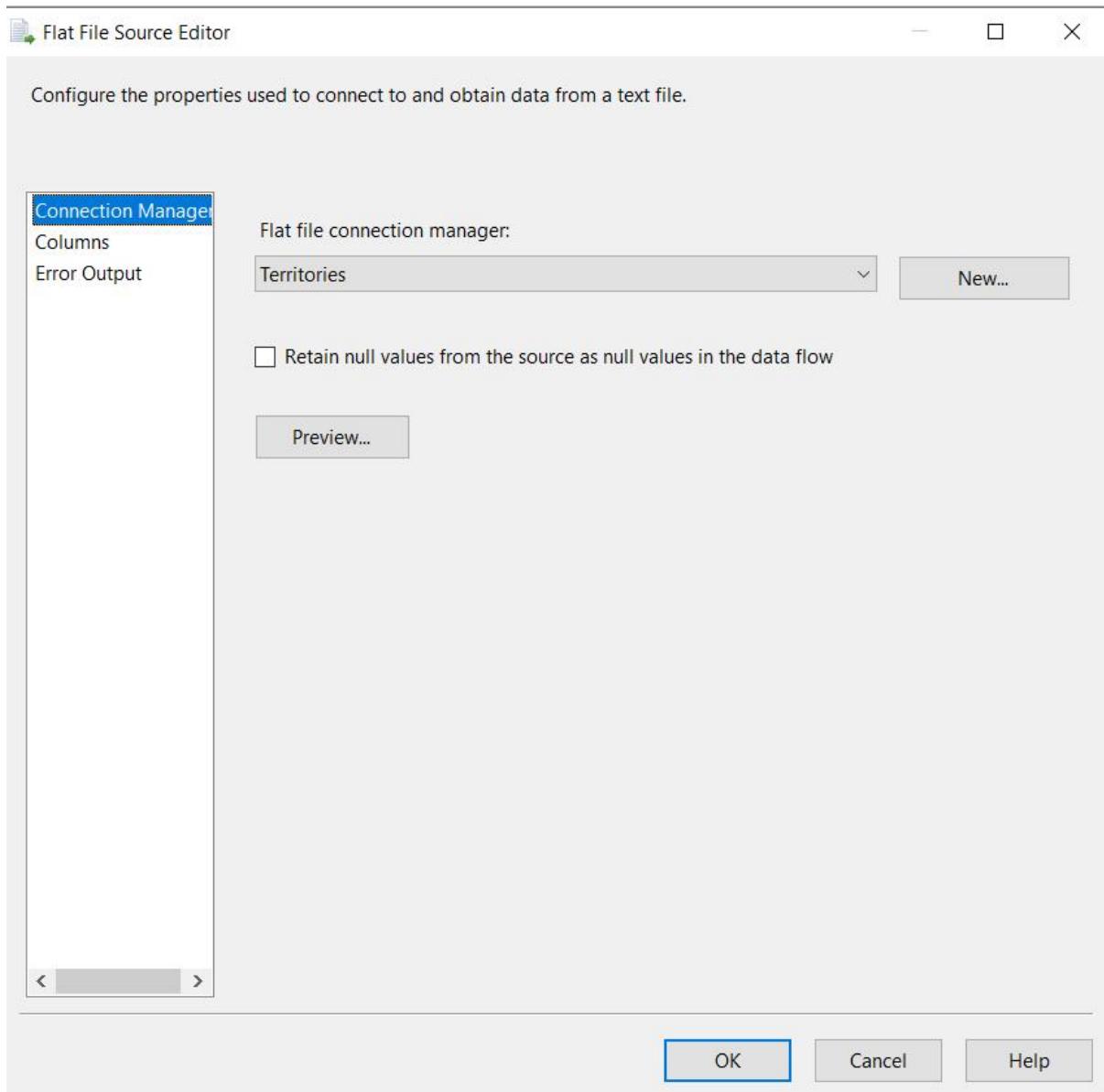
+ Sau khi cấu hình xong ta được:



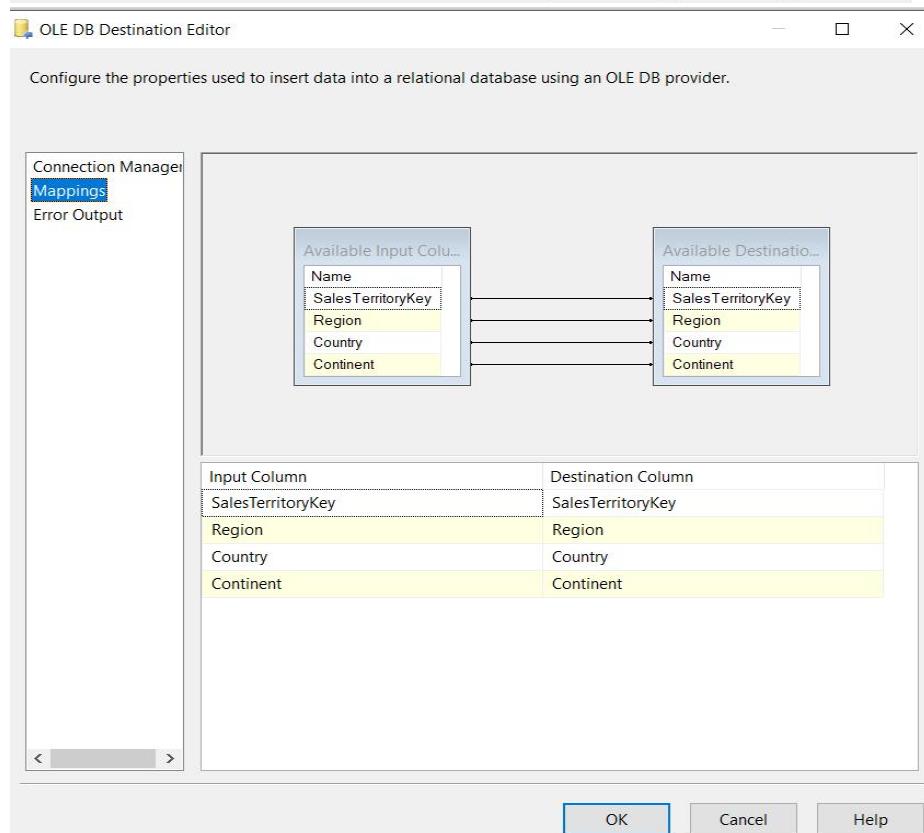
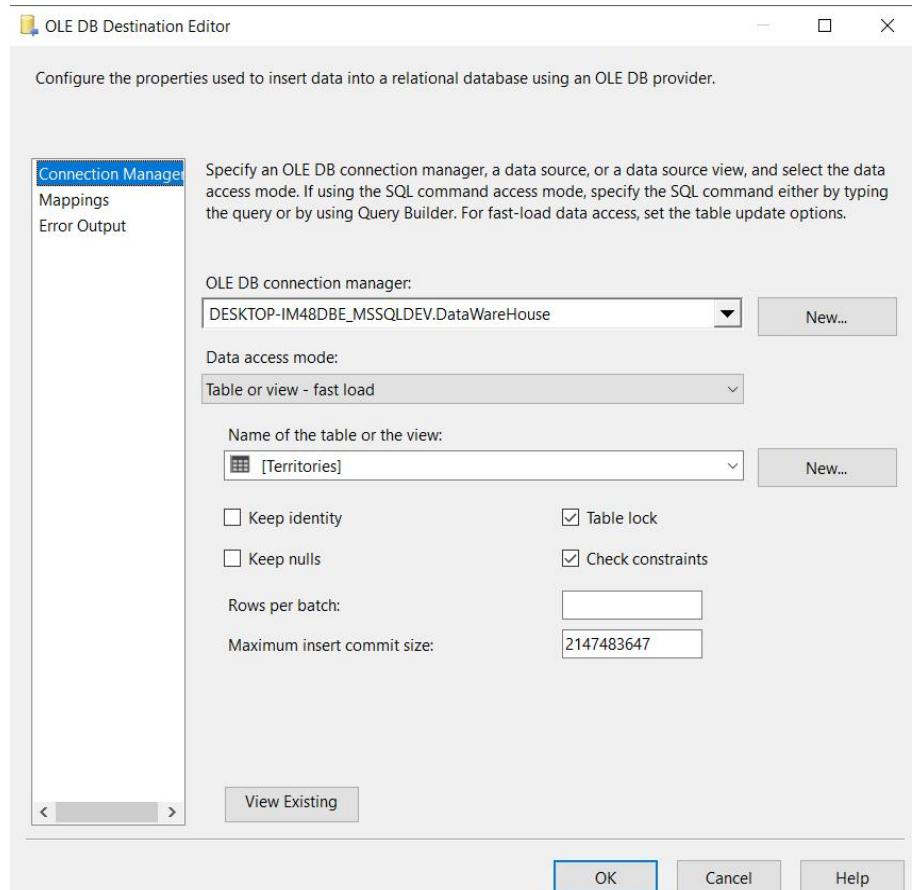
3.3.6. Territory

3.3.6.1. Load dữ liệu từ Source vào Stage

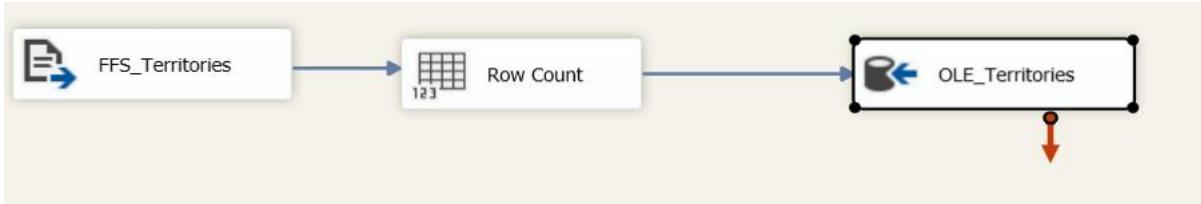
- Double-click **DFT_LoadTerritories** task để mở nó trong **Data Flow design surface**:
- Tạo **Flat File Source**, **RowCount** và **Destination Assistant** và cấu hình chúng như sau:
 - + **Flat File Source**:



+ Destination Assistant:



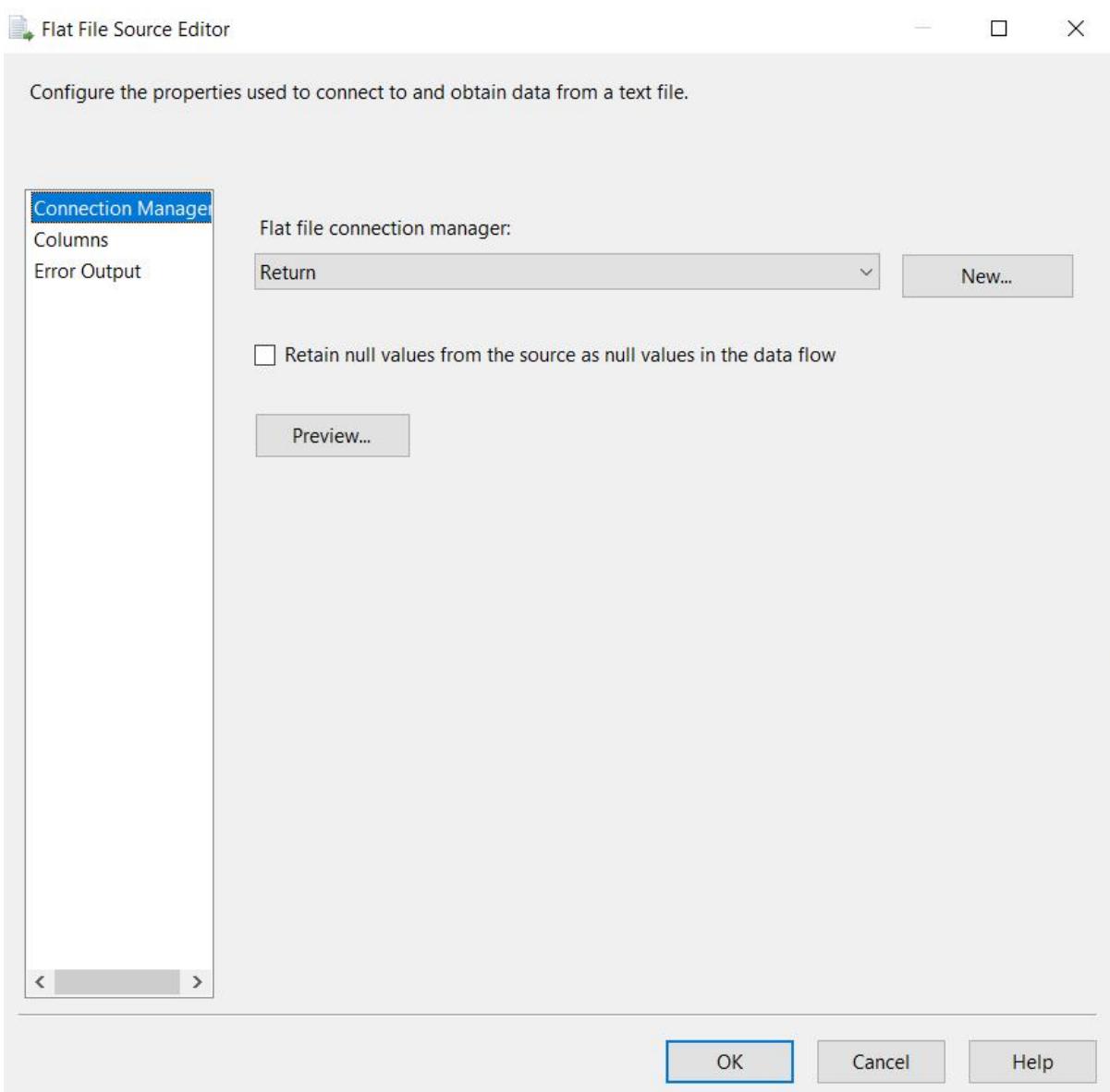
+ Sau khi cấu hình xong ta được:



3.3.7. Returns

3.3.7.1. Load dữ liệu từ Source vào Stage

- Double-click **DFT_LoadReturns** task để mở nó trong **Data Flow design surface**:
- Tạo **Flat File Source**, **RowCount** và **Destination Assistant** và cấu hình chúng như sau:
 - + **Flat File Source**:



+ Destination Assistant:

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

OLE DB Destination Editor

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: DESKTOP-IM48DBE_MSSQLDEV.DataWareHouse **New...**

Data access mode: Table or view - fast load

Name of the table or the view: [Returns] **New...**

Keep identity Table lock

Keep nulls Check constraints

Rows per batch:

Maximum insert commit size:

View Existing

OK **Cancel** **Help**

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

OLE DB Destination Editor

Mappings

Available Input Columns: Name, ReturnDate, TerritoryKey, ProductKey, ReturnQuantity

Available Destination Columns: Name, ReturnDate, TerritoryKey, ProductKey, ReturnQuantity

Input Column	Destination Column
ReturnDate	ReturnDate
TerritoryKey	TerritoryKey
ProductKey	ProductKey
ReturnQuantity	ReturnQuantity

OK **Cancel** **Help**

+ Sau khi cấu hình xong ta được:



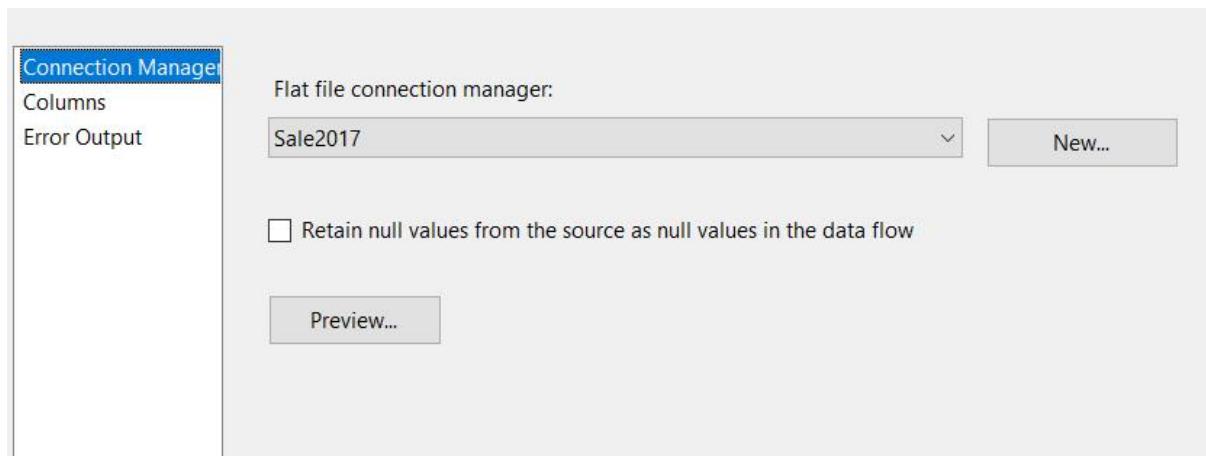
3.3.8. Sales

3.3.8.1. Load dữ liệu từ Source vào Stage

- Double-click DFT_LoadReturns task để mở nó trong Data Flow design surface:
- Tạo Flat File Source, Union All, RowCount và Destination Assistant và cấu hình chúng như sau:
 - + Flat File Source:

The screenshot shows two instances of the 'Flat file connection manager' dialog box side-by-side. Both instances have 'Connection Manager' selected in the left sidebar. The right pane contains the following fields:

- 'Flat file connection manager:' dropdown: 'Sale2015' in the top instance and 'Sale2016' in the bottom instance.
- 'New...' button.
- A checkbox labeled 'Retain null values from the source as null values in the data flow'.
- 'Preview...' button.



+ Union All:

Union All Transformation Editor

Configure the properties used to merge multiple inputs into one output by creating mappings between columns.

Output Column Name	Union All Input 1	Union All Input 2	Uni
OrderDate	OrderDate	OrderDate	Or
StockDate	StockDate	StockDate	Sto
OrderNumber	OrderNumber	OrderNumber	Ord
ProductKey	ProductKey	ProductKey	Pro
CustomerKey	CustomerKey	CustomerKey	Cus
TerritoryKey	TerritoryKey	TerritoryKey	Ter
OrderLineItem	OrderLineItem	OrderLineItem	Ord
OrderQuantity	OrderQuantity	OrderQuantity	Ord

+ Destination Assistant:

OLE DB Destination Editor

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

OLE DB connection manager: DESKTOP-IM48DBE_MSSQLDEV.DataWarehouse

Data access mode: Table or view - fast load

Name of the table or the view: [Sales]

Keep identity Table lock

Keep nulls Check constraints

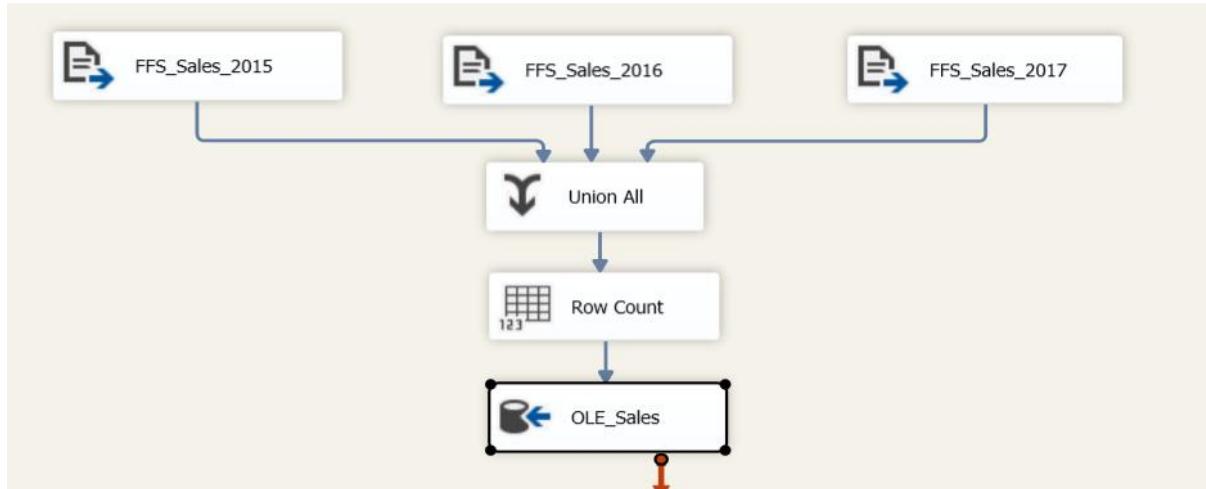
Rows per batch:

Maximum insert commit size: 2147483647

Mappings

Input Column	Destination Column
OrderDate	OrderDate
StockDate	StockDate
OrderNumber	OrderNumber
ProductKey	ProductKey
CustomerKey	CustomerKey
TerritoryKey	TerritoryKey
OrderLineItem	OrderLineItem
OrderQuantity	OrderQuantity

+ Sau khi cấu hình xong ta được:

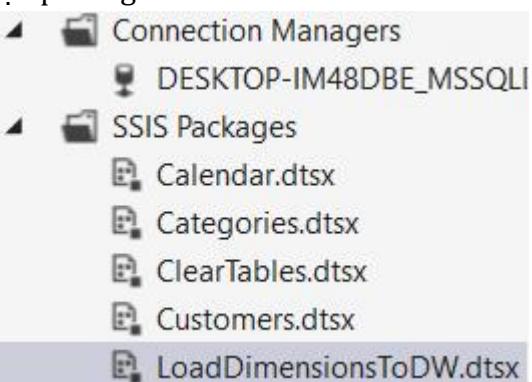


3.4. Import dữ liệu từ bảng stage vào bảng dimension

3.4.1. DimCustomer

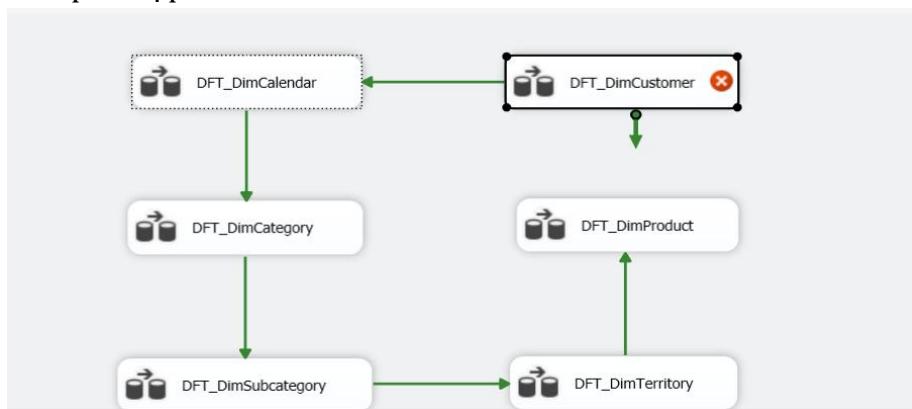
3.4.1.1. Tạo Package

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



3.4.1.2. Thiết lập Control Flow

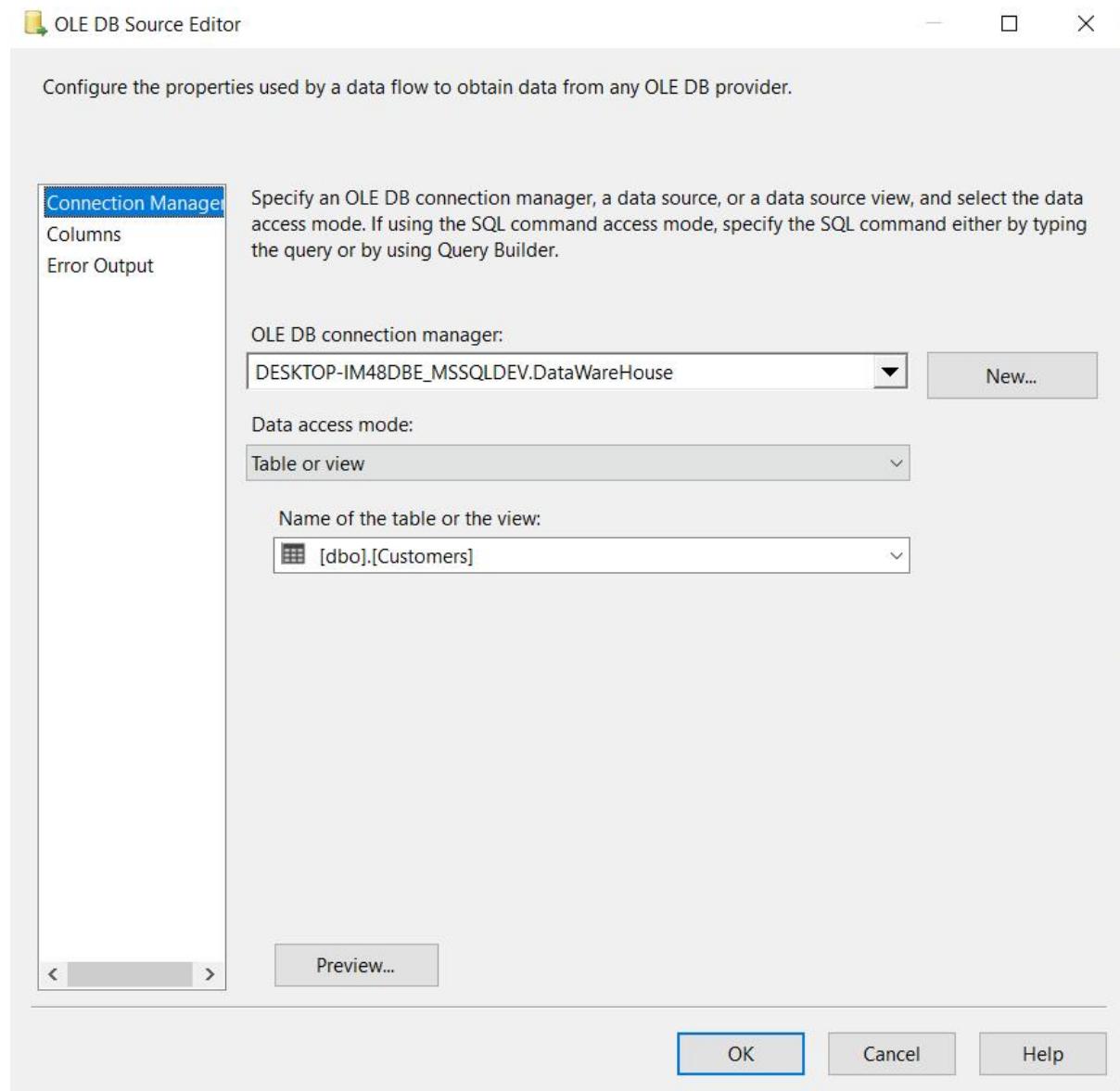
- Tạo và đặt tên tương ứng cho các task, tạo các connect bằng cách kéo mũi tên vào các task phù hợp.



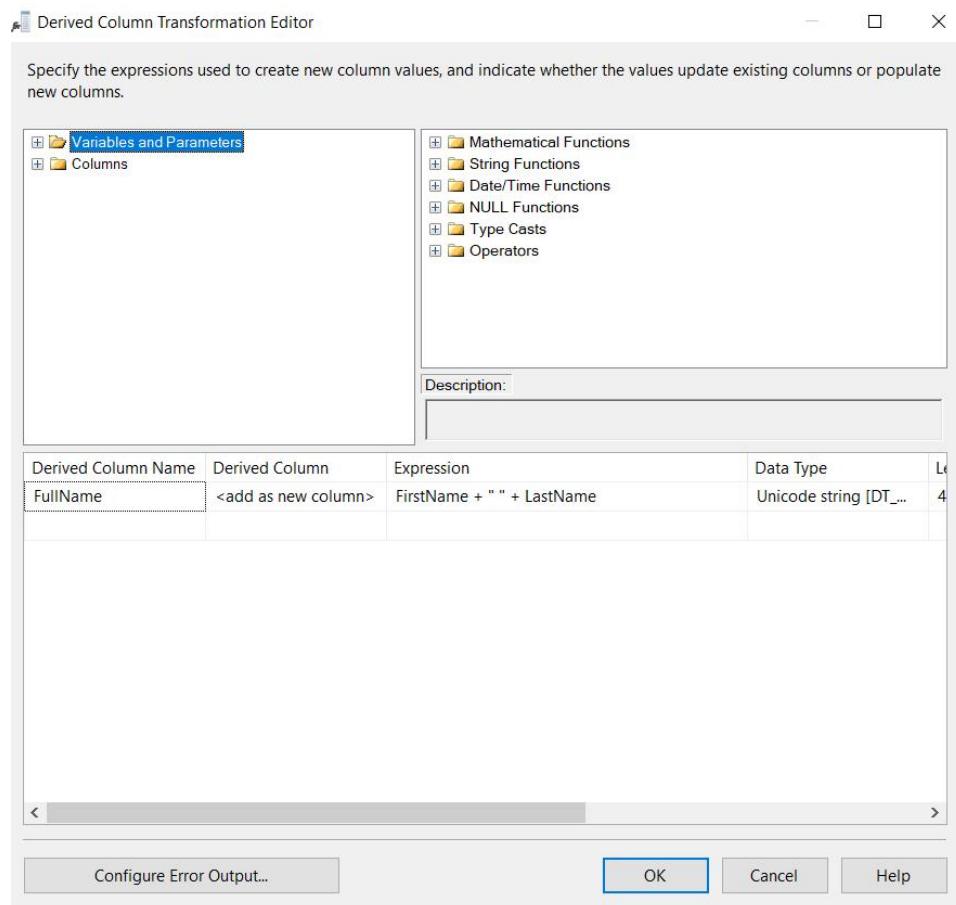
3.4.1.3. Load dữ liệu từ Stage vào Dimension

- Double-click **DFT_DimCustomer** task để mở nó trong **Data Flow design surface**:
- Tạo **Flat File Source**, **Derived Column** và **Destination Assistant** và cấu hình chúng như sau:

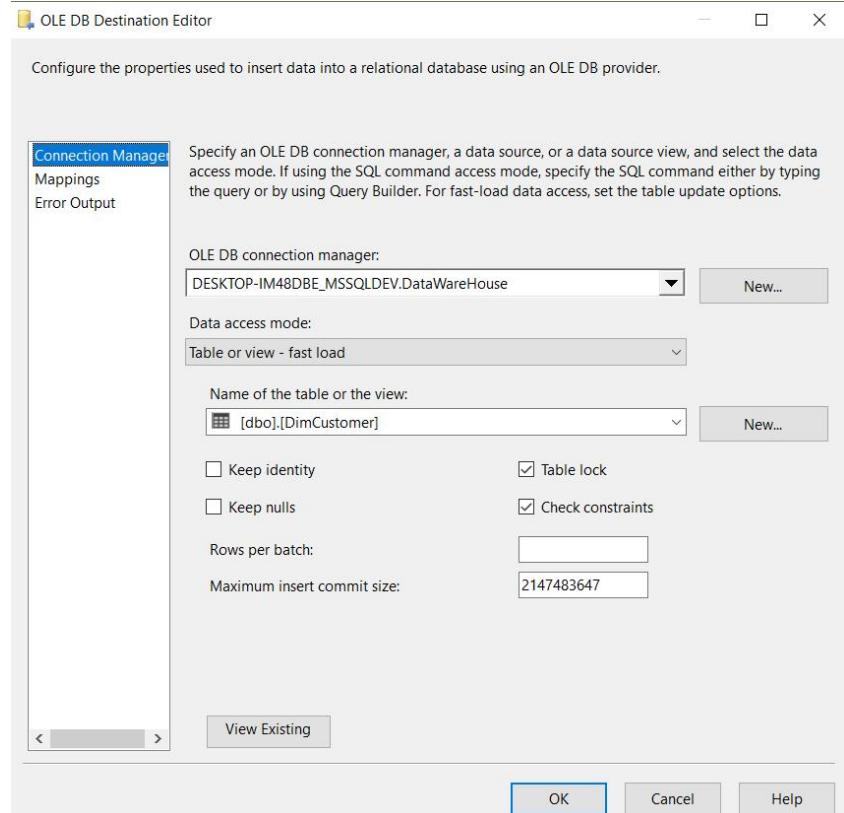
+ Flat File Source:

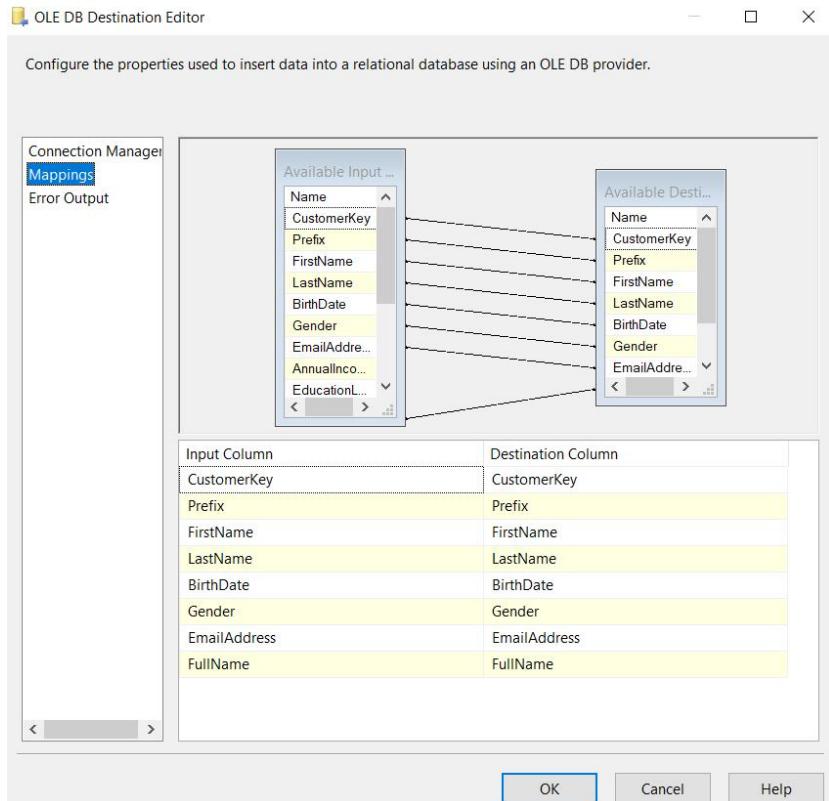


+ Derived Column:



+ Destination Assist





+ Sau khi cấu hình xong ta được:



3.4.2. DimCalendar

3.4.2.1. Load dữ liệu từ Stage vào Dimension

- Double-click DFT_DimCalendar task để mở nó trong Data Flow design surface:
- Tạo Source Assistant, Derived Column và Destination Assistant và cấu hình chúng như sau:

+ Source Assistant:

OLE DB Source Editor

Configure the properties used by a data flow to obtain data from any OLE DB provider.

Connection Manager

Columns

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.

OLE DB connection manager:

DESKTOP-IM48DBE_MSSQLDEV.DataWareHouse

New...

Data access mode:

Table or view

Name of the table or the view:

[dbo].[Calendar]

< >

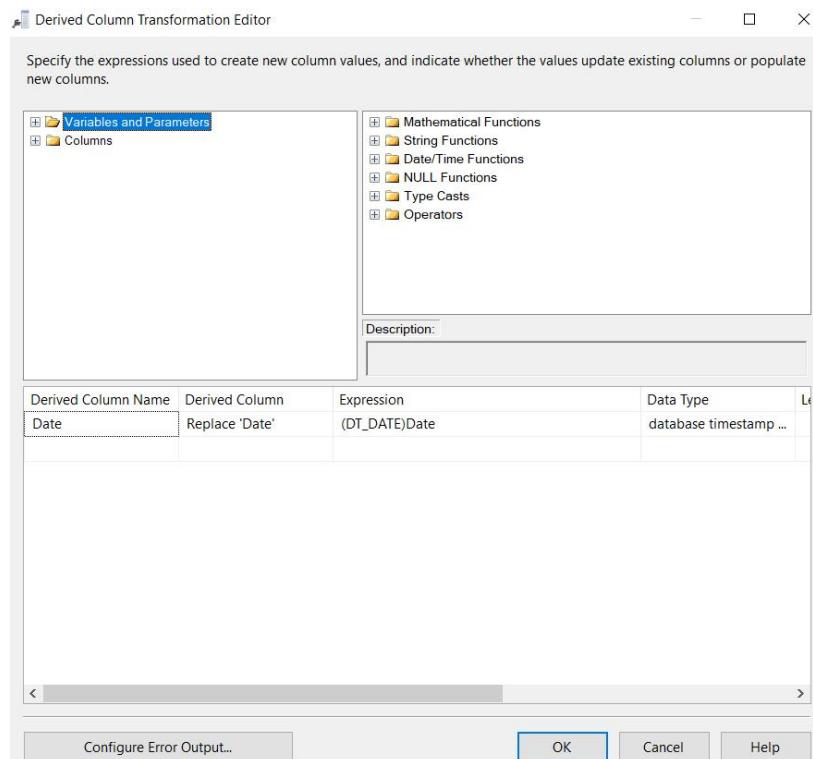
Preview...

OK

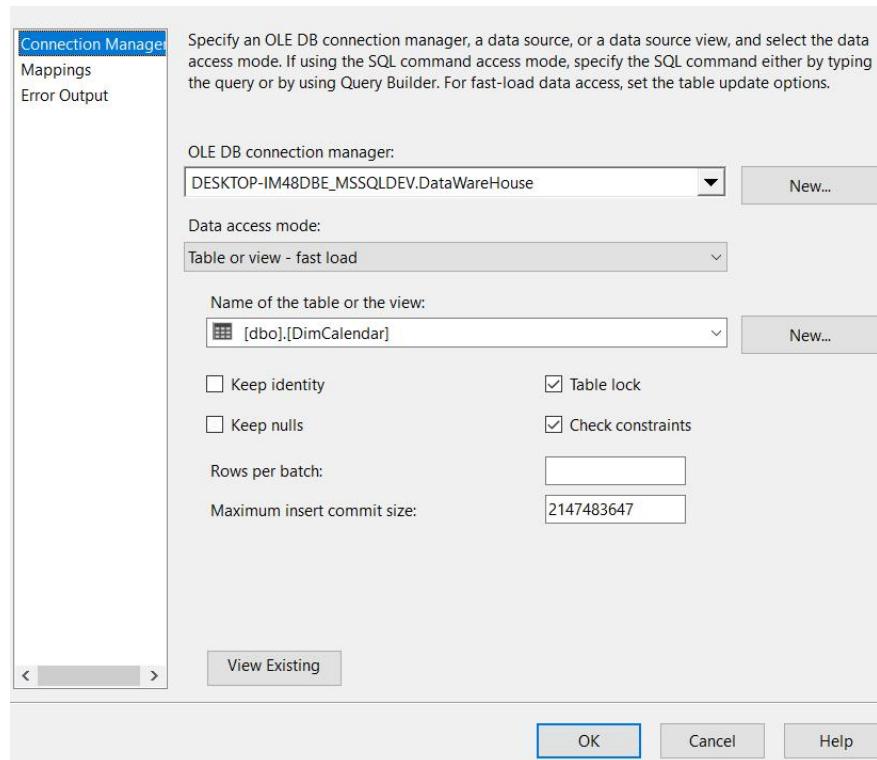
Cancel

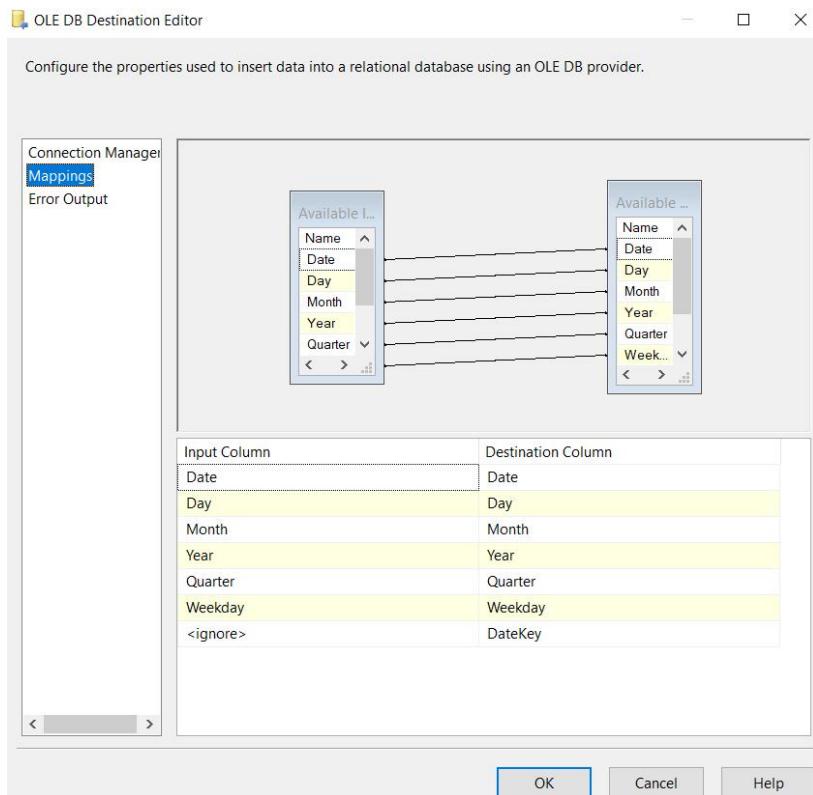
Help

+ Derived Column:



+ Destination Assist





+ Sau khi cấu hình xong ta được:



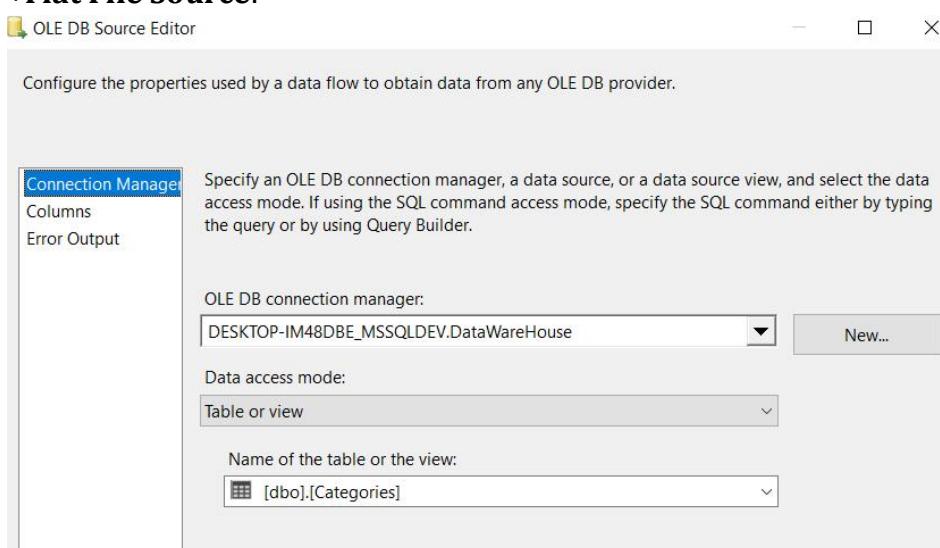
3.4.3. DimCategory

3.4.3.1. Load dữ liệu từ Stage vào Dimension

- Double-click **DFT_DimCustomer** task để mở nó trong **Data Flow design surface**:

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

+ **Flat File Source**:



+ Destination Assist

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

Connection Manager: DESKTOP-IM48DBE_MSSQLDEV.DataWareHouse New...

Data access mode: Table or view - fast load

Name of the table or the view: [dbo].[DimCategory] New...

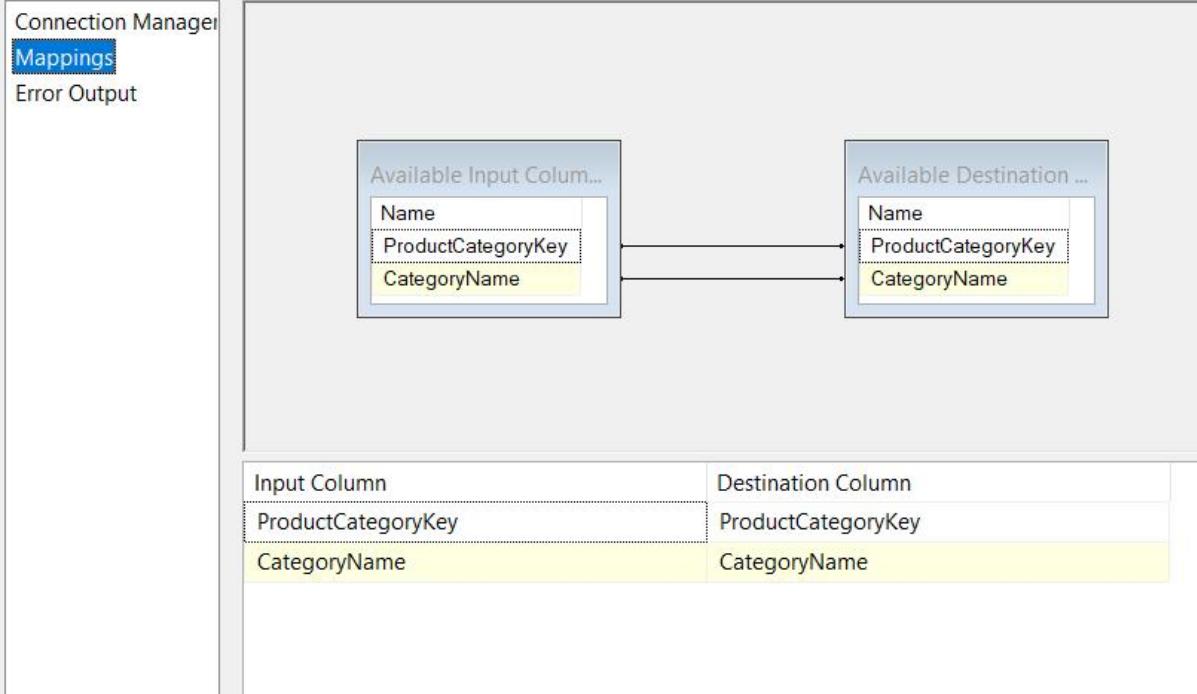
Keep identity Table lock

Keep nulls Check constraints

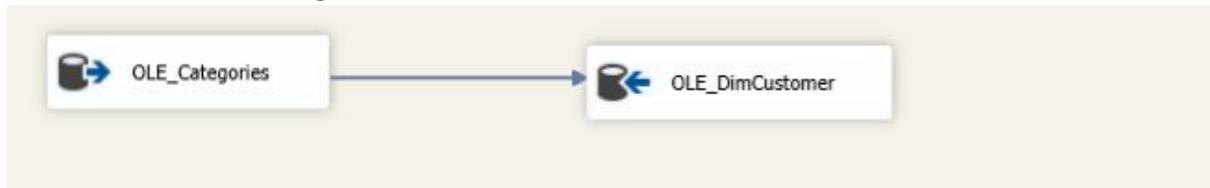
Rows per batch:

Maximum insert commit size:

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



+ Sau khi cấu hình xong ta được:

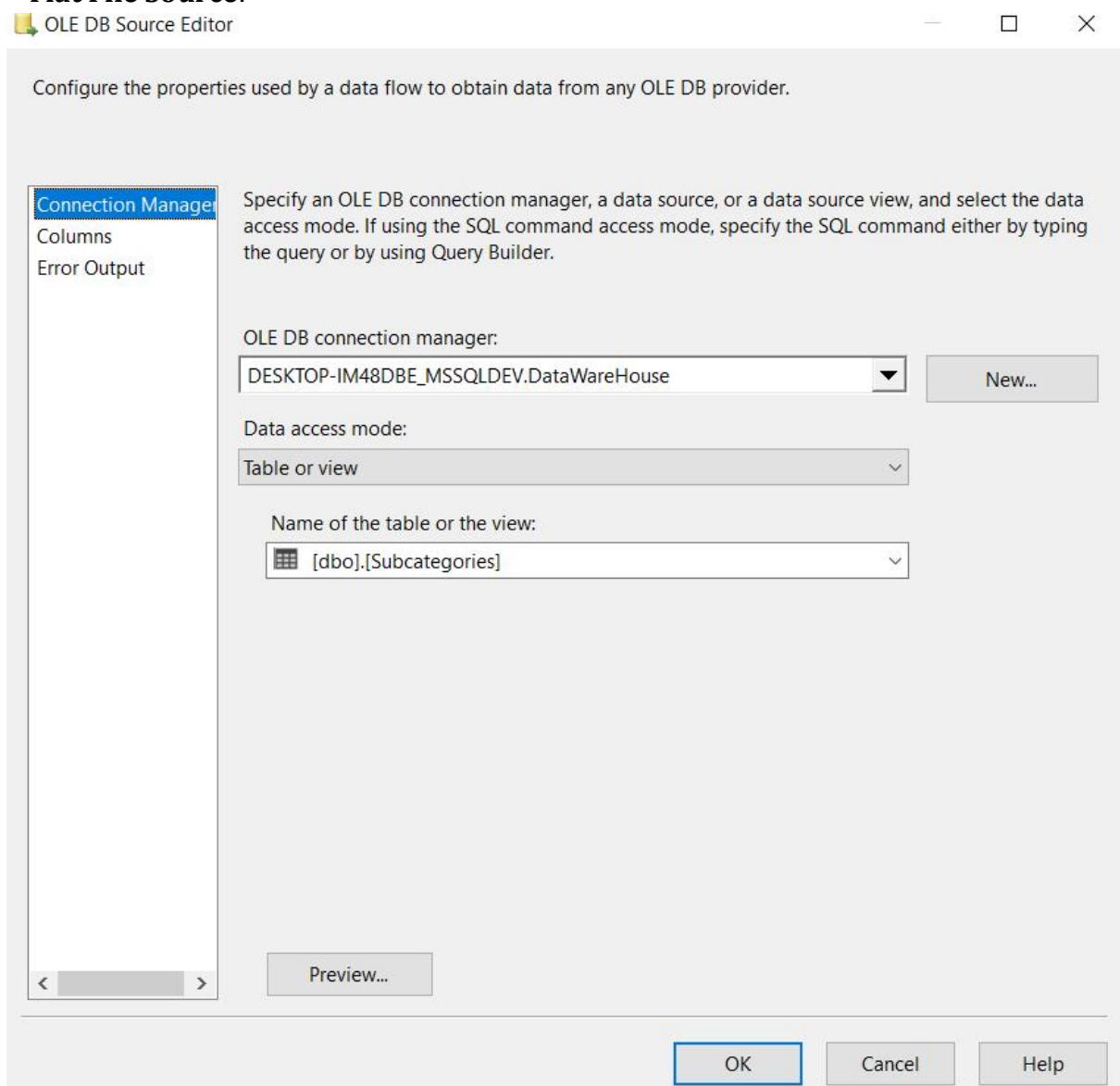


3.4.4. DimSubCategory

3.4.4.1. Load dữ liệu từ Stage vào Dimension

- Double-click **DFT_DimCustomer** task để mở nó trong **Data Flow design surface**:
- Tạo **Flat File Source**, **Derived Column** và **Destination Assistant** và cấu hình chúng như sau:

+ **Flat File Source:**



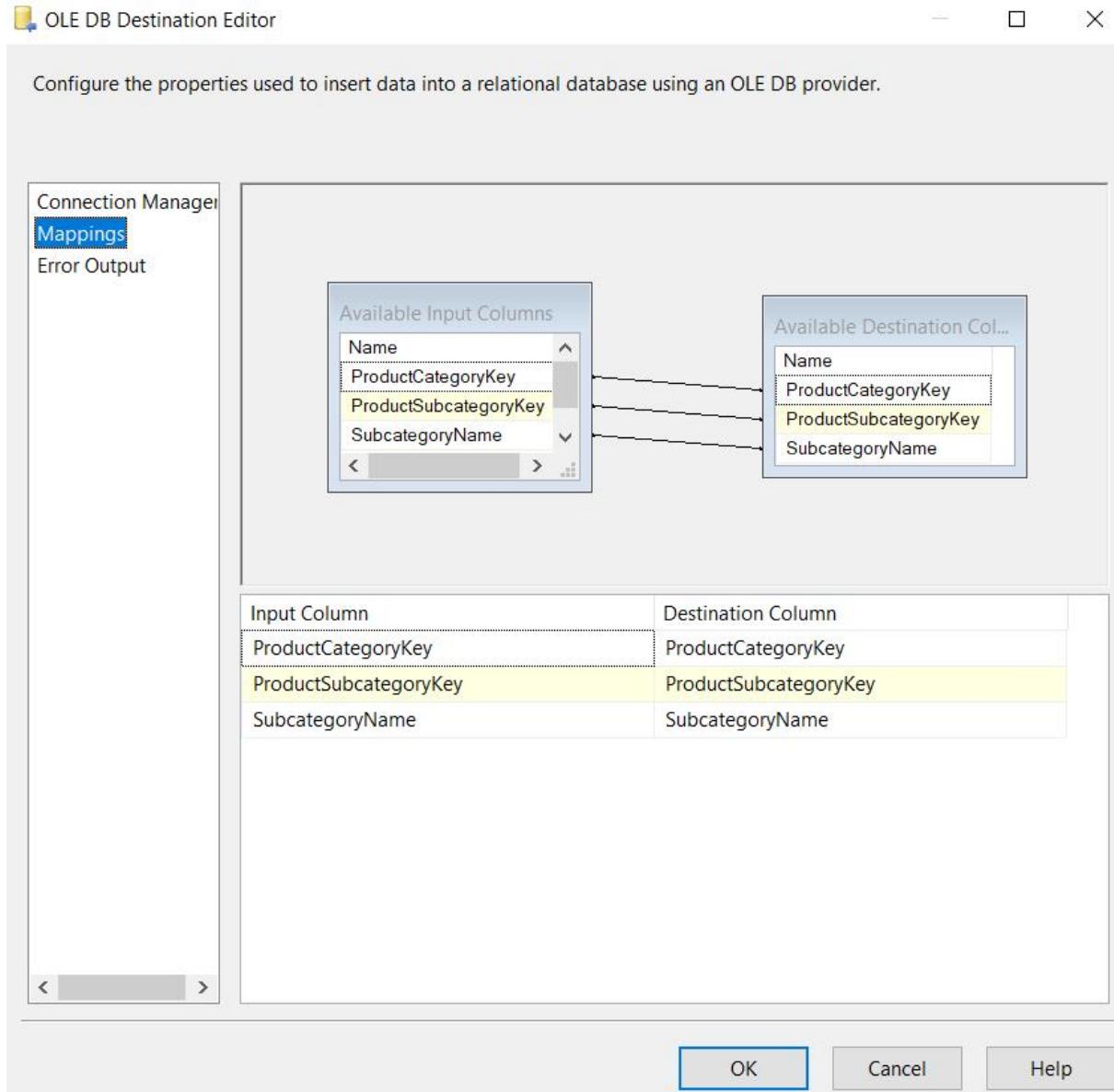
+ Derived Column:

The screenshot shows the Derived Column Transformation Editor window. On the left, there's a navigation pane with 'Variables and Parameters' selected, followed by 'Columns'. To the right of the pane is a list of function categories: Mathematical Functions, String Functions, Date/Time Functions, NULL Functions, Type Casts, and Operators. Below these is a 'Description:' text input field. The main area contains a table for defining derived columns:

Derived Column Name	Derived Column	Expression	Data Type	Length
StartDate	<add as new column>	(DT_DBTIMESTAMP)(@System::CreationDate])	database timestamp	8

+ Destination Assist

The screenshot shows the OLE DB Destination Editor window. On the left, there's a sidebar with 'Connection Manager' selected, followed by 'Mappings' and 'Error Output'. The main area is titled 'Configure the properties used to insert data into a relational database using an OLE DB provider.' It includes fields for 'OLE DB connection manager' (set to 'DESKTOP-IM48DBE_MSSQLDEV.DataWarehouse'), 'Data access mode' (set to 'Table or view - fast load'), 'Name of the table or the view' (set to '[dbo].[DimSubCategory]'), and various checkboxes for data handling like 'Keep identity', 'Table lock', 'Keep nulls', 'Check constraints', etc. At the bottom are buttons for 'View Existing', 'OK', 'Cancel', and 'Help'.



+ Sau khi cấu hình xong ta được:

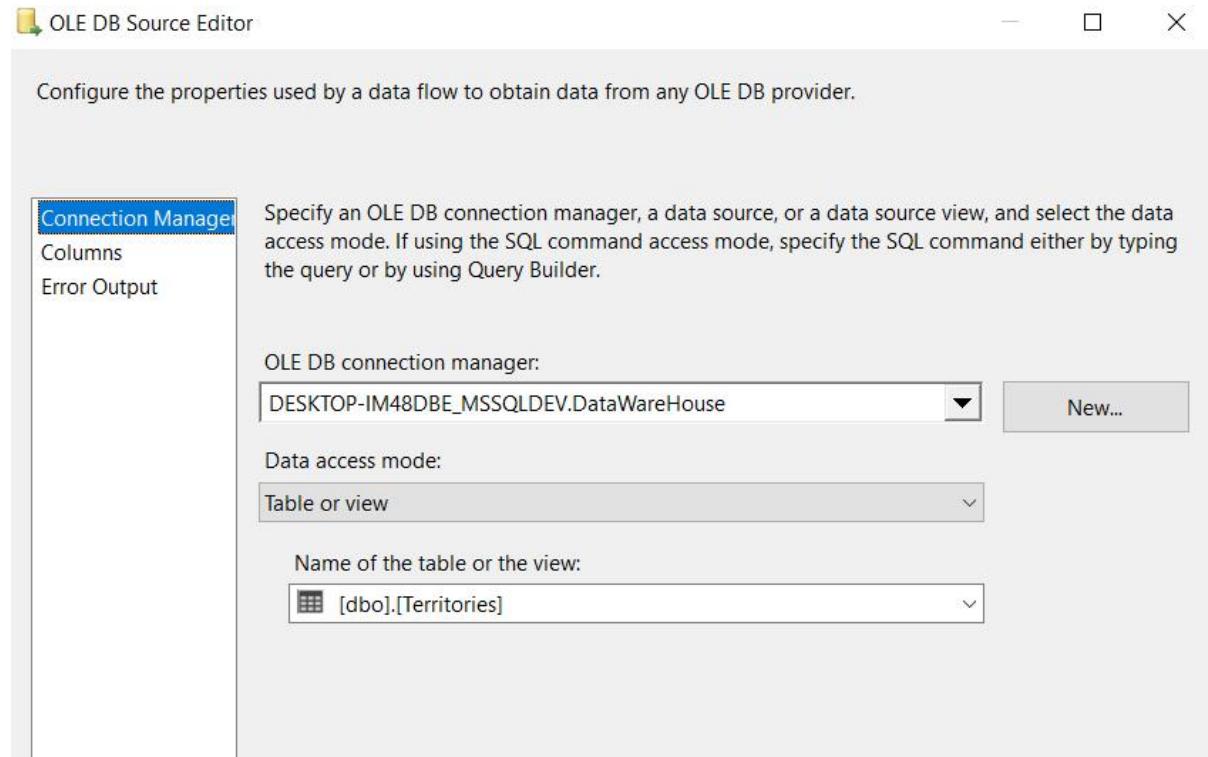


3.4.5. DimTerritory

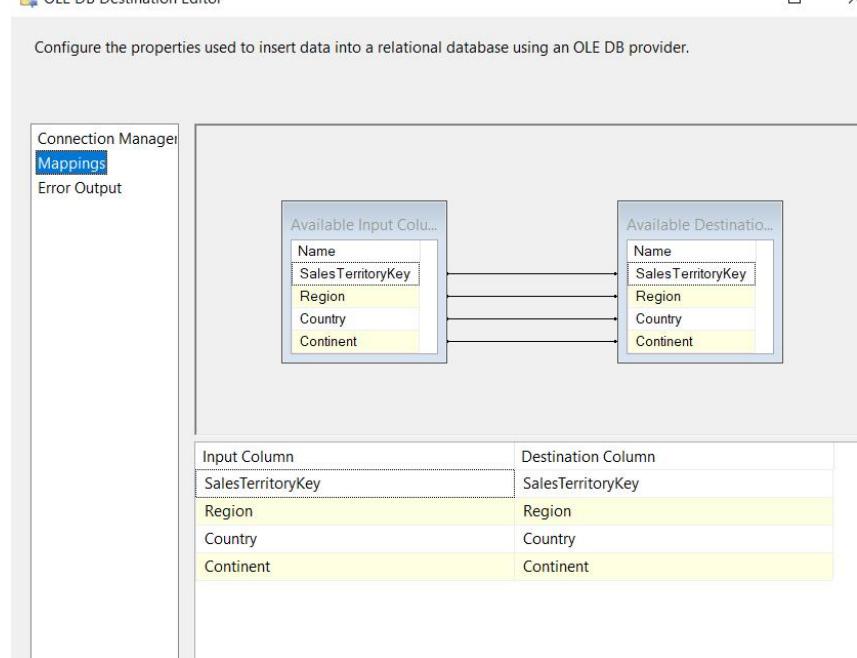
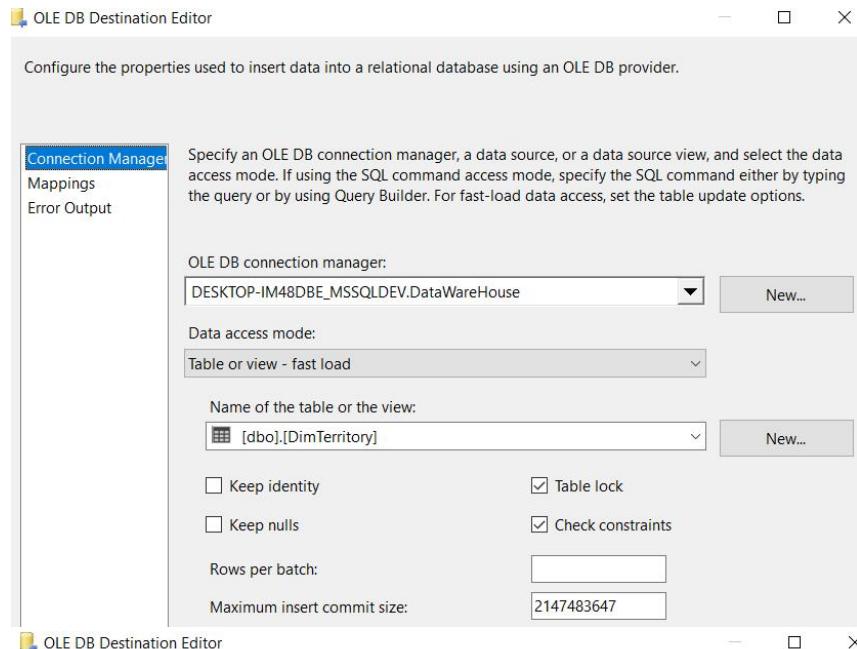
3.4.5.1. Load dữ liệu từ Stage vào Dimension

- Double-click **DFT_DimTerritory** task để mở nó trong **Data Flow design surface**:
- Tạo **Flat File Source**, **Derived Column** và **Destination Assistant** và cấu hình chúng như sau:

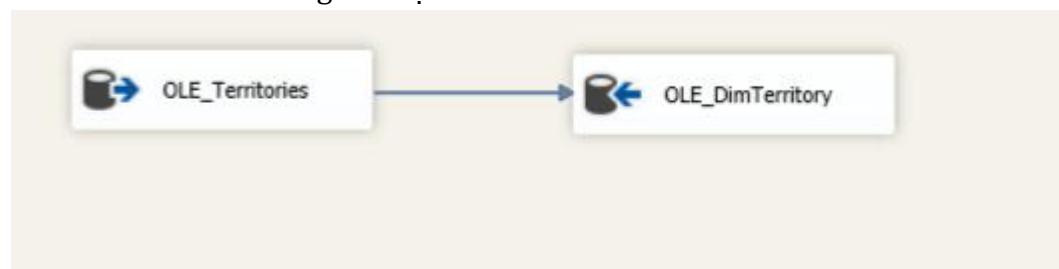
+ Flat File Source:



+ Destination Assist



+ Sau khi cấu hình xong ta được:

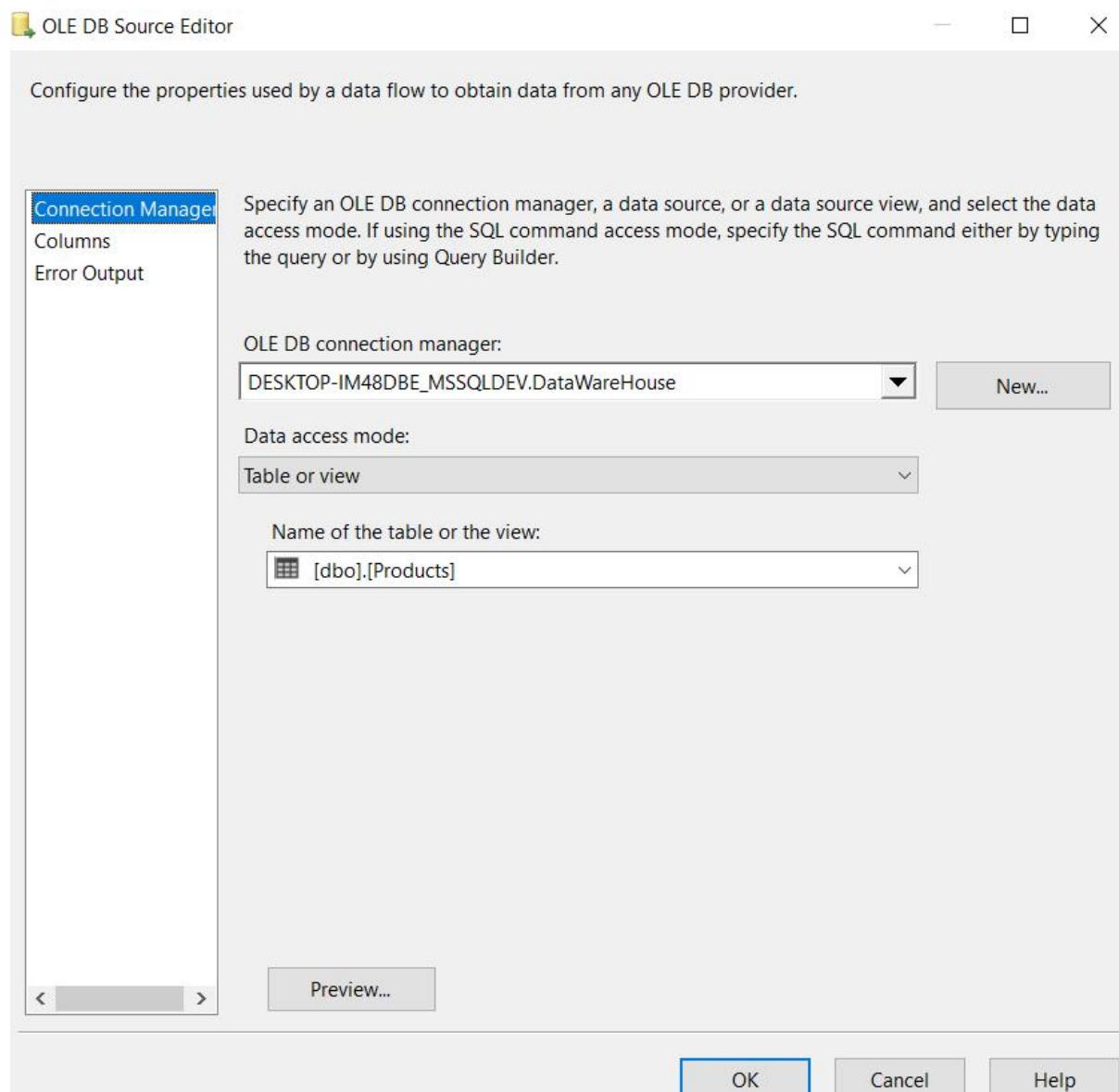


3.4.6. DimProduct

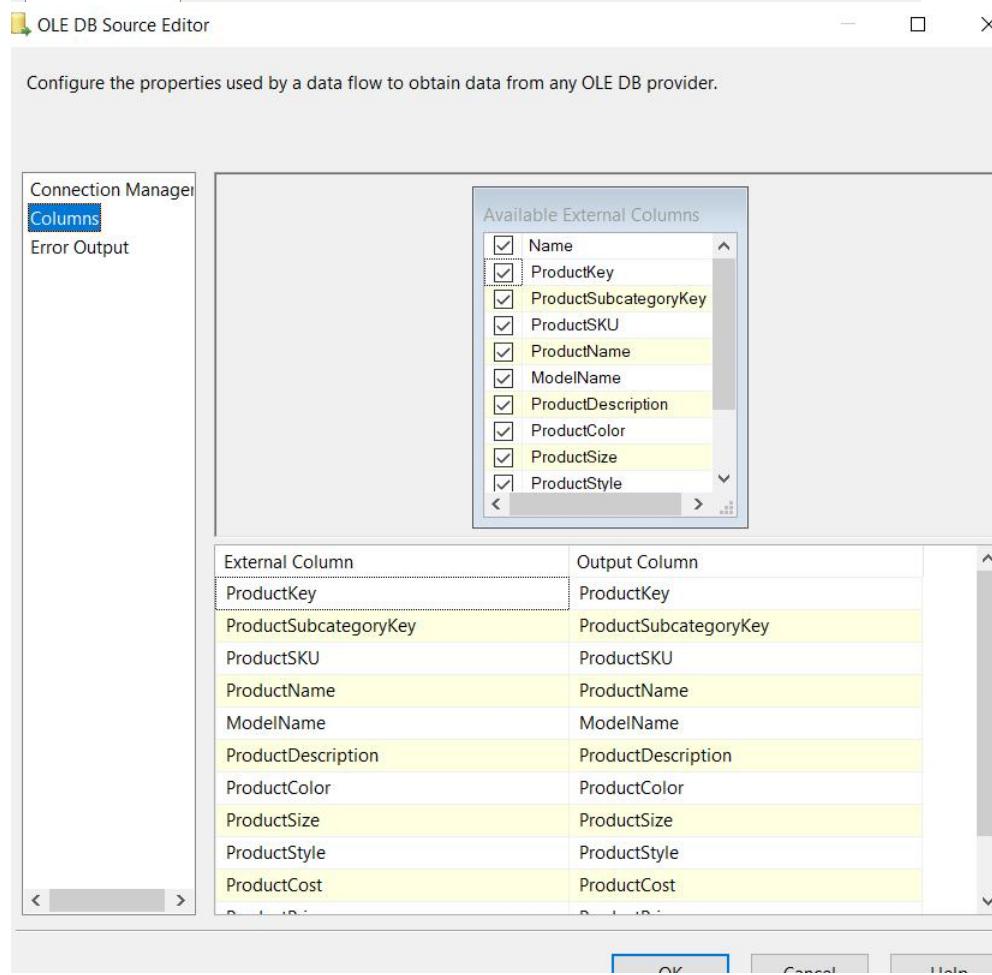
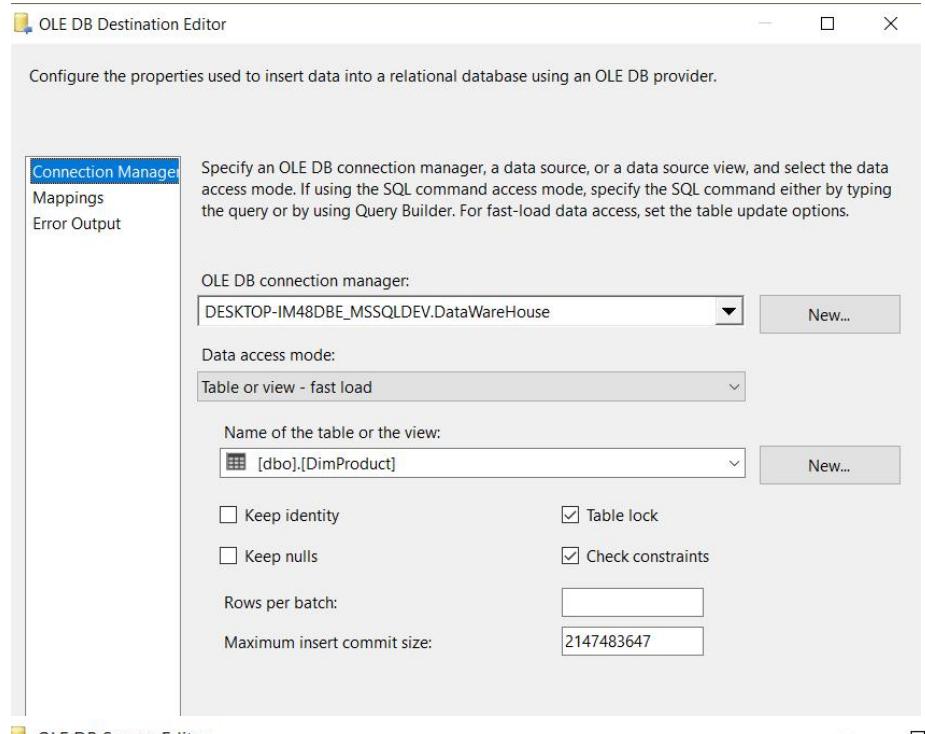
3.4.6.1. Load dữ liệu từ Stage vào Dimension

- Double-click **DFT_DimProduct** task để mở nó trong **Data Flow design surface**:
- Tạo **Flat File Source**, **Derived Column** và **Destination Assistant** và cấu hình chúng như sau:

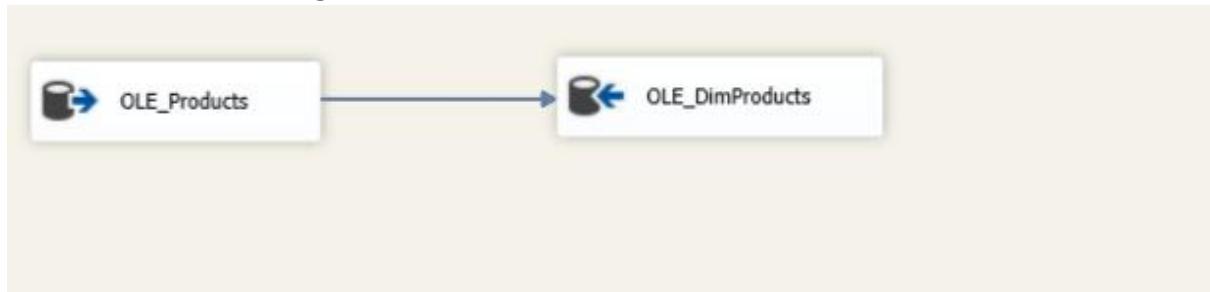
+ Flat File Source:



+ Destination Assist



+ Sau khi cấu hình xong ta được:

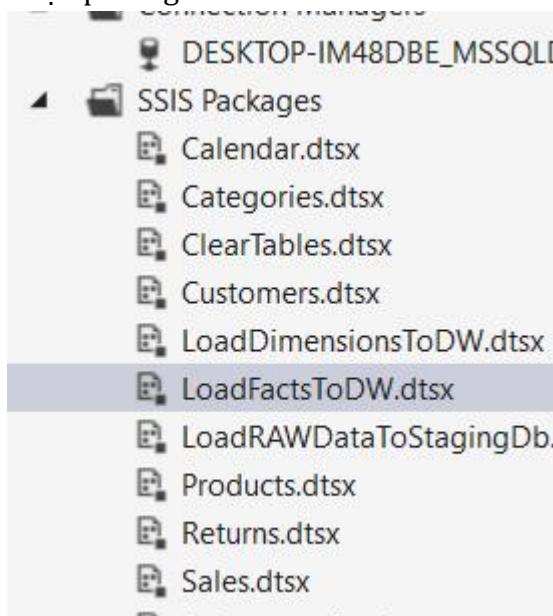


3.5. Import dữ liệu vào bảng các bảng fact

3.5.1. Fact Sales

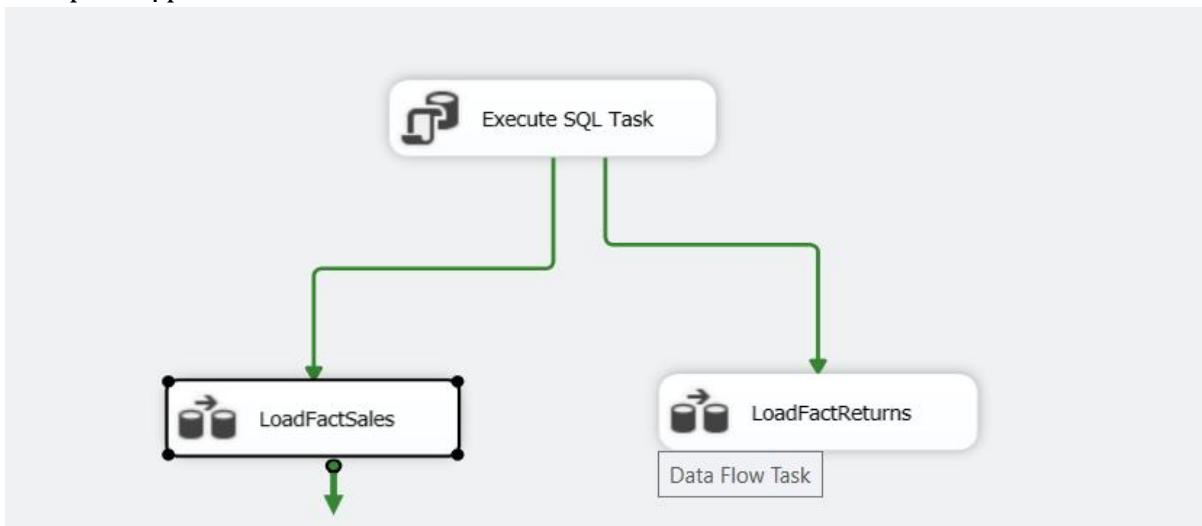
3.5.1.1. Tạo Package

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

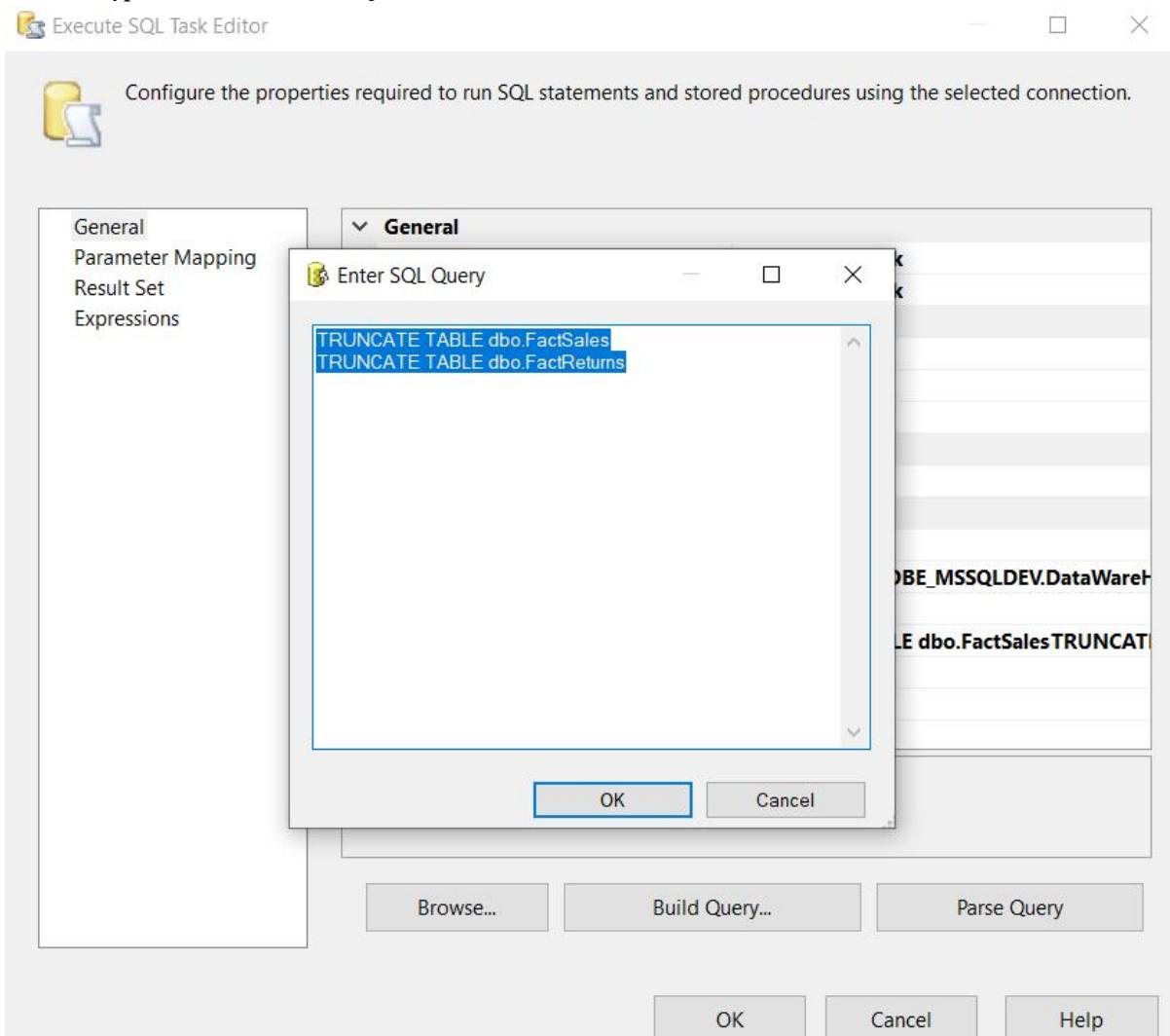


3.5.1.2. Thiết lập Control Flow

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

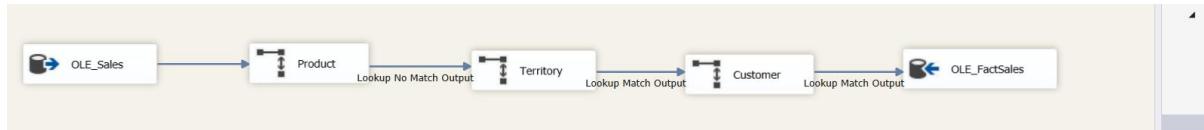


- Thiết lập truncate table SQL task

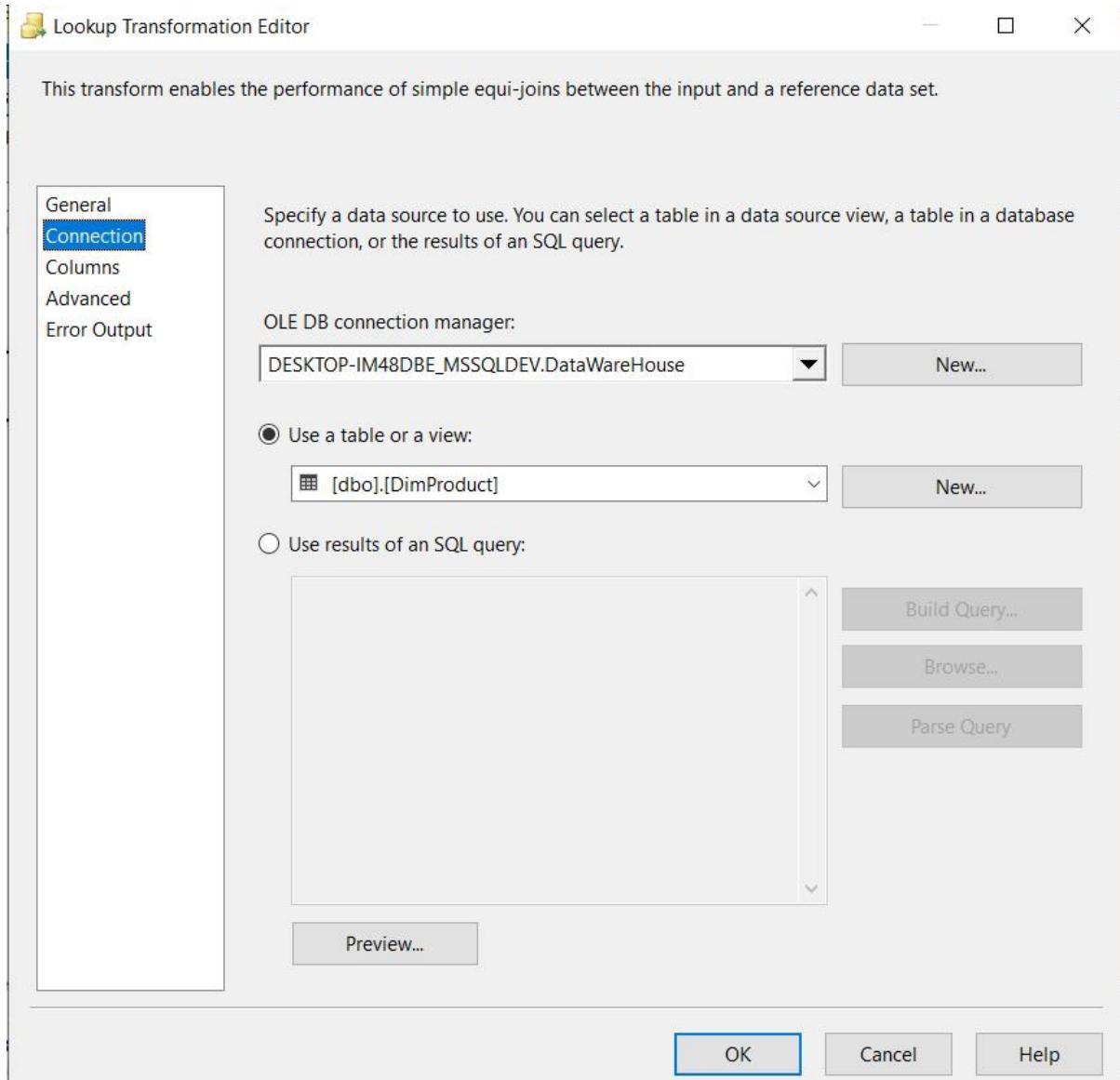


3.5.1.3. Load dữ liệu từ Stage vào Fact

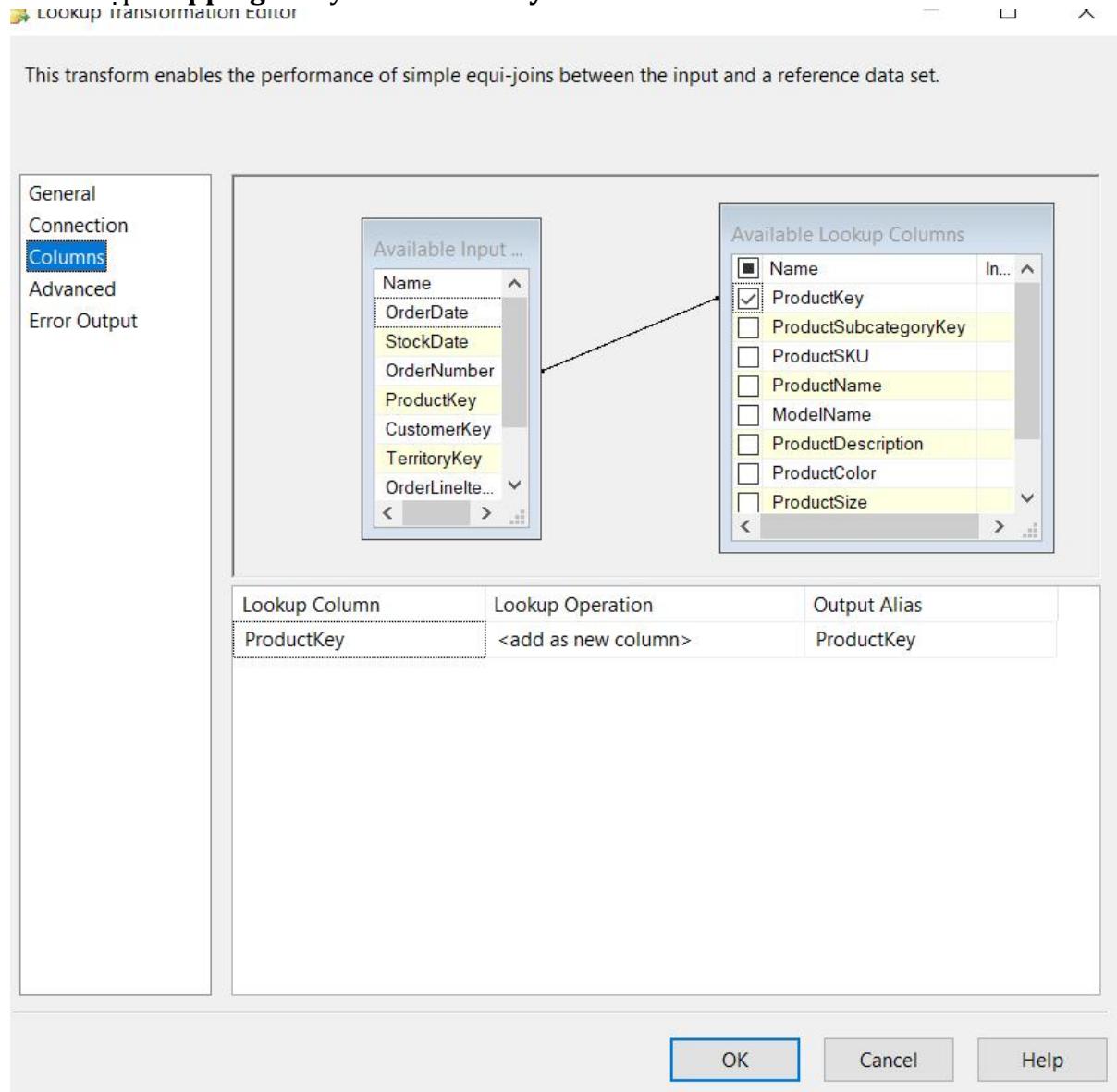
- Double-click vào **LoadFactSales** task để mở data flow design Surface và thiết lập các cấu hình sau



- Lookup qua bảng DimProduct



- Thiết lập mapping và lấy ra ProductKey



- Lookup qua bảng DimTerritory

 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
OrderDate
StockDate
OrderNumber
ProductKey
CustomerKey
TerritoryKey
OrderLineItem...

Available Lookup Columns

Name	In...
<input type="checkbox"/> Name	In...
<input checked="" type="checkbox"/> SalesTerritoryKey	In...
<input type="checkbox"/> Region	In...
<input type="checkbox"/> Country	In...

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

- Lookup qua bảng DimCustomer

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

The screenshot shows the configuration of a Lookup transformation. On the left, a sidebar lists General, Connection, Columns, Advanced, and Error Output. The main area has two 'Available' panes: 'Available Input Columns' (containing Name, OrderDate, StockDate, OrderNumber, ProductKey, CustomerKey, TerritoryKey, OrderLineItem..., and OrderQuantity) and 'Available Lookup Columns' (containing Name, CustomerKey, Prefix, FirstName, LastName, BirthDate, Gender, and EmailAddress...). A line connects CustomerKey in the input pane to CustomerKey in the lookup pane. Below these panes is a mapping grid:

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

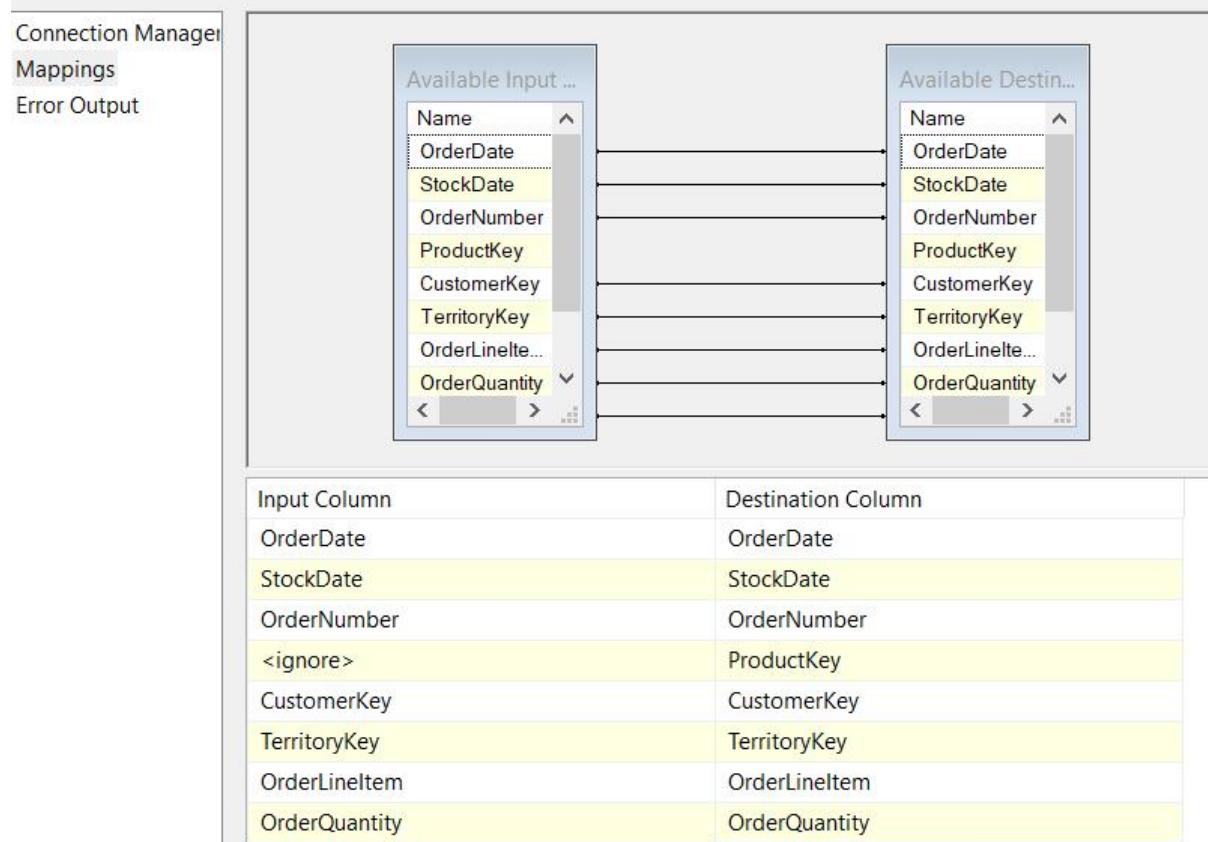
- Cấu hình Destination chọn bảng FactSales

The screenshot shows the configuration of a Destination component. The left sidebar includes Connection Manager, Mappings, and Error Output. The main area is titled 'Connection Manager' and contains the following settings:

- 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: DESKTOP-IM48DBE_MSSQLDEV.DataWarehouse
- Data access mode: Table or view - fast load
- Name of the table or the view: [dbo].[FactSales]
- Checkboxes at the bottom:
 - Keep identity
 - Table lock
 - Keep nulls
 - Check constraints

- Mapping các cột

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



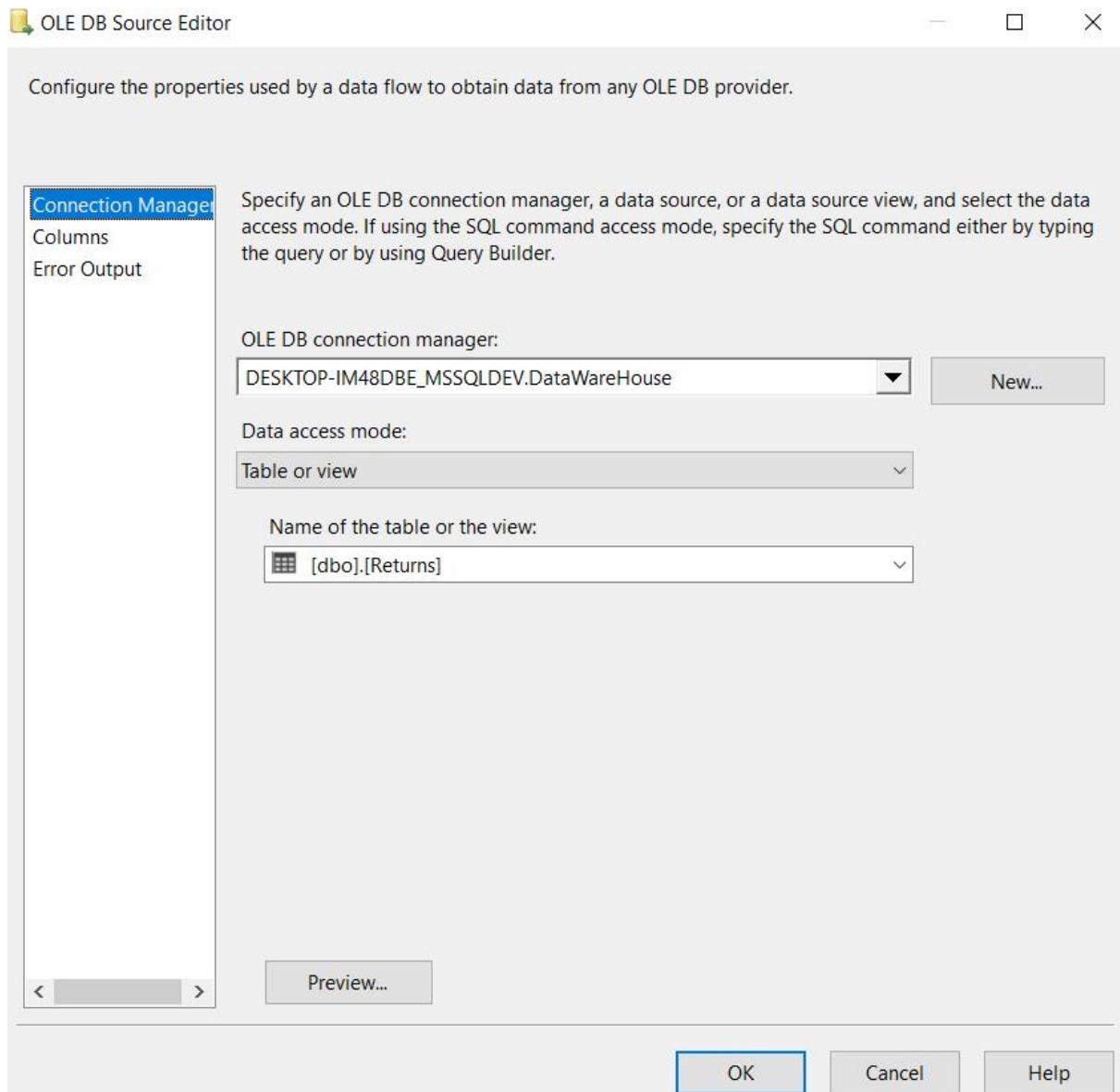
3.5.2. Fact Returns

3.5.2.1. Load dữ liệu từ Stage vào Fact

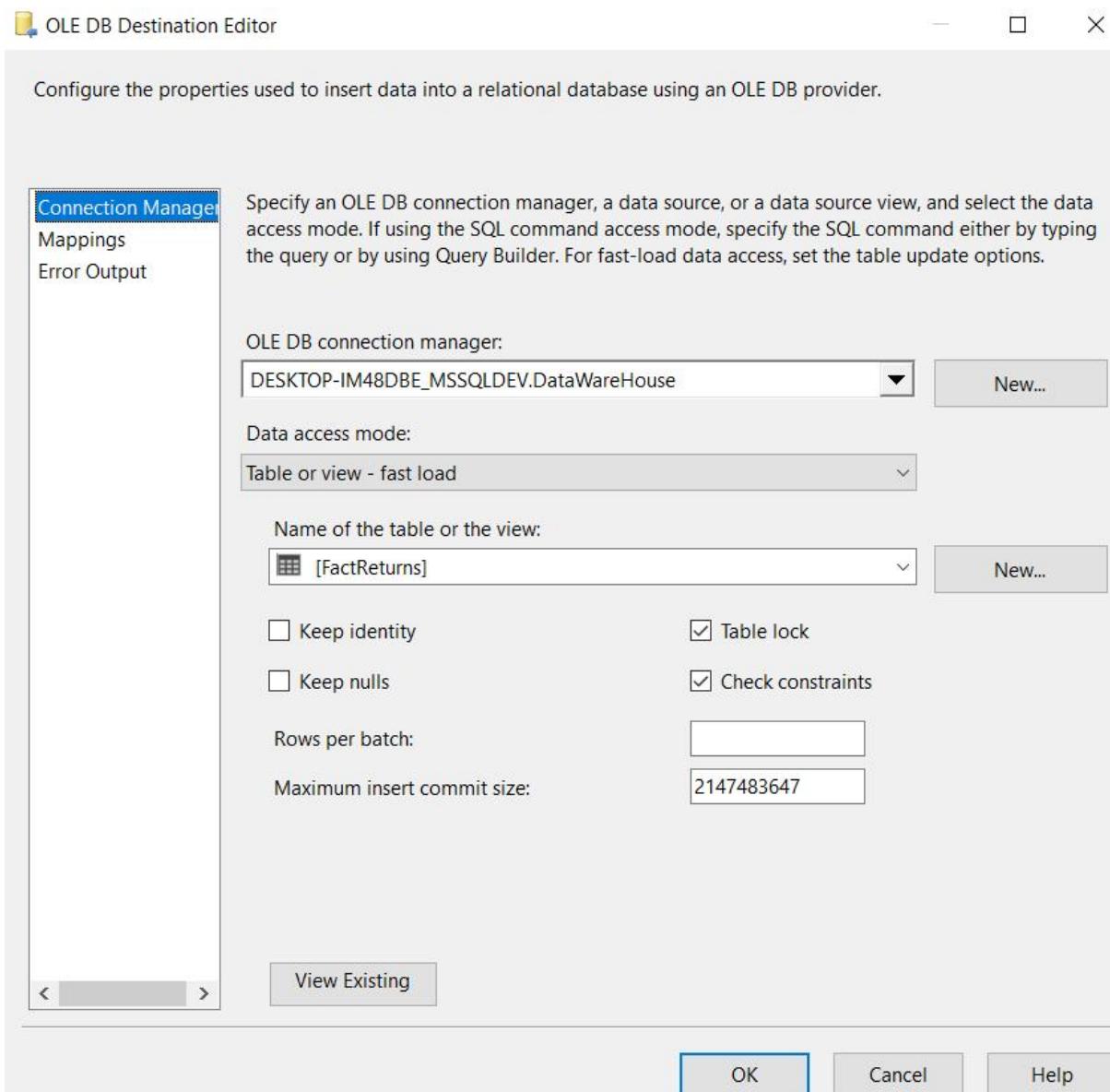
- Double-click vào **LoadFactReturns** task để mở data flow design Surface và thiết lập các cấu hình sau



+ Source Assistant:



+ Destination Assistant:

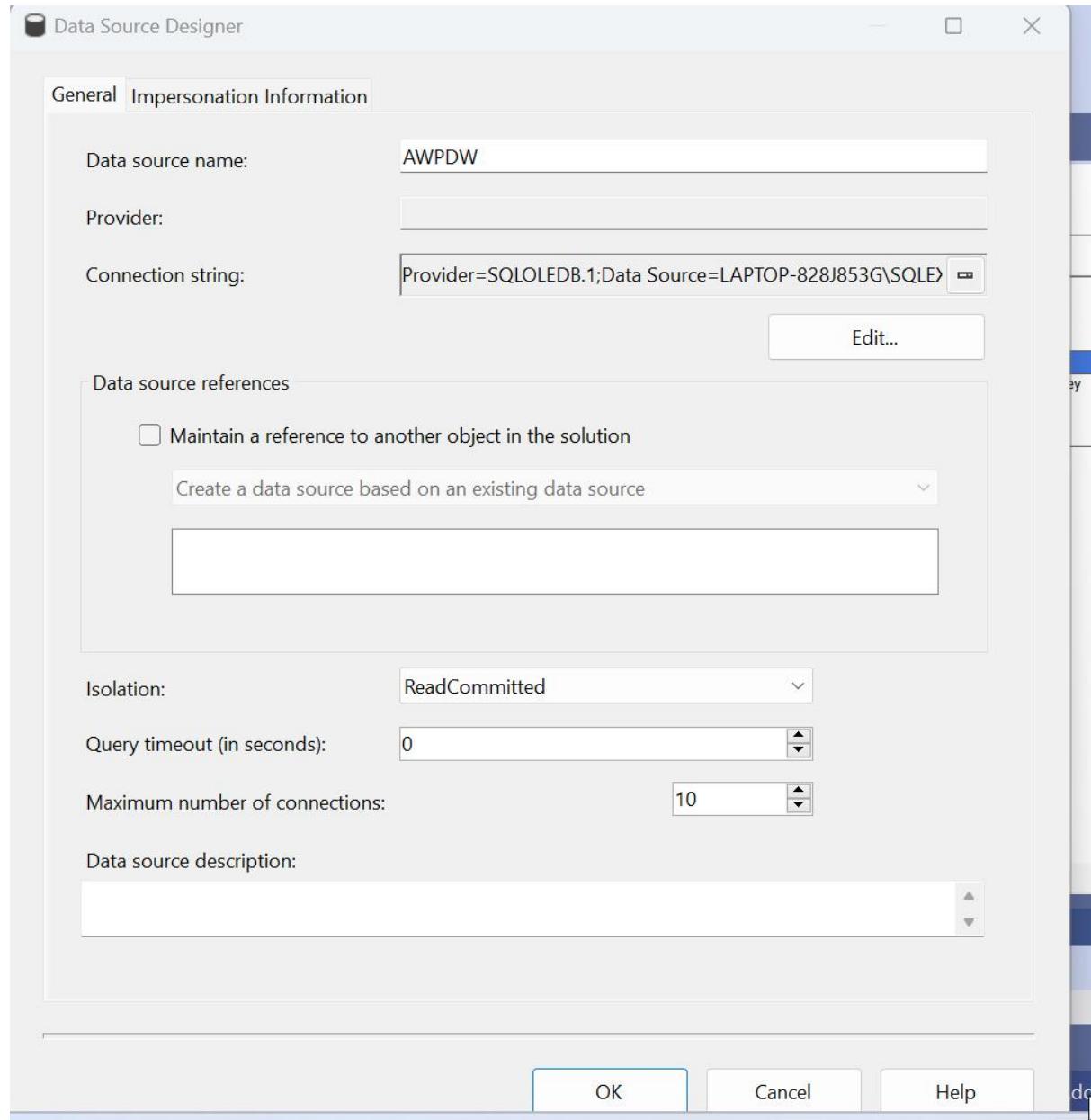


CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)

4.1. Quá trình xây dựng mô hình

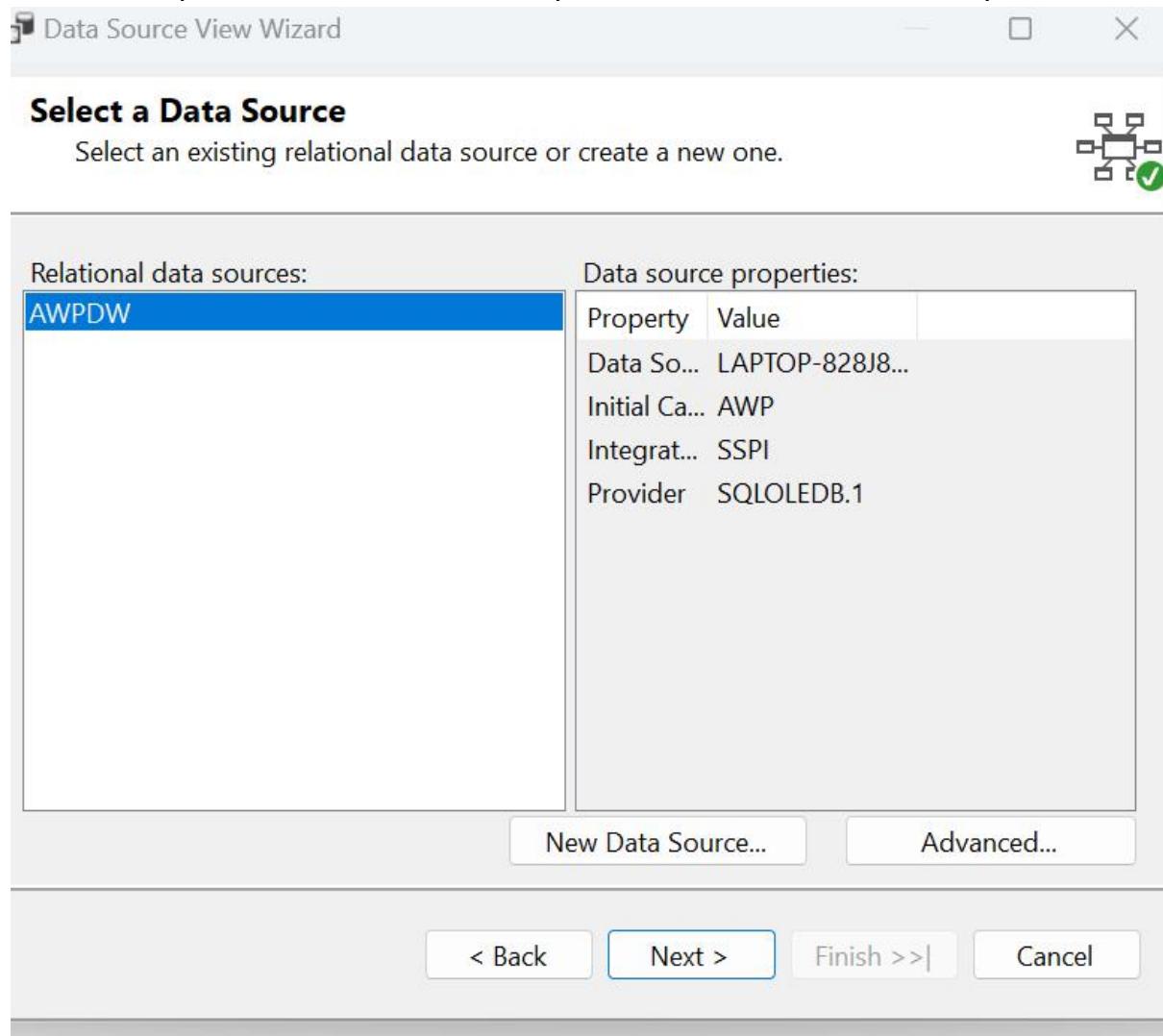
4.1.1. Tạo Data Source

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

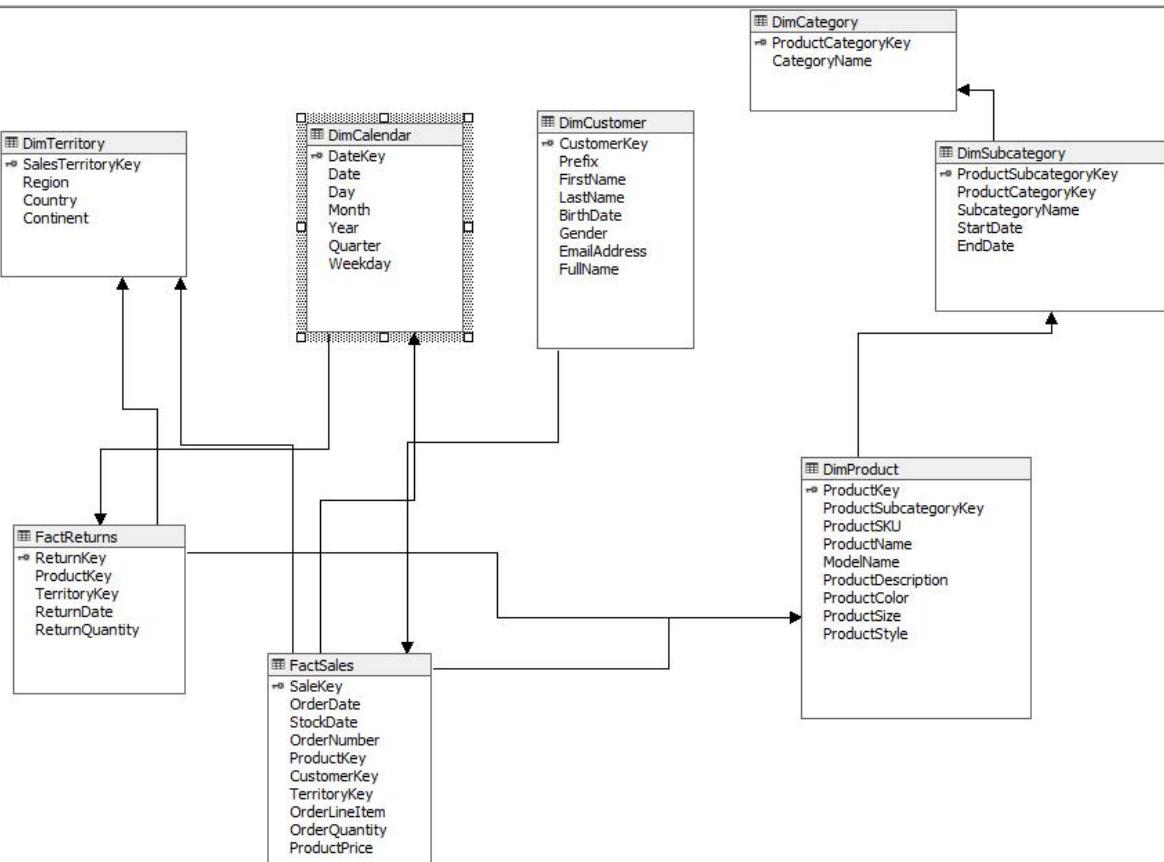


4.1.2. Tạo Data Source View

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

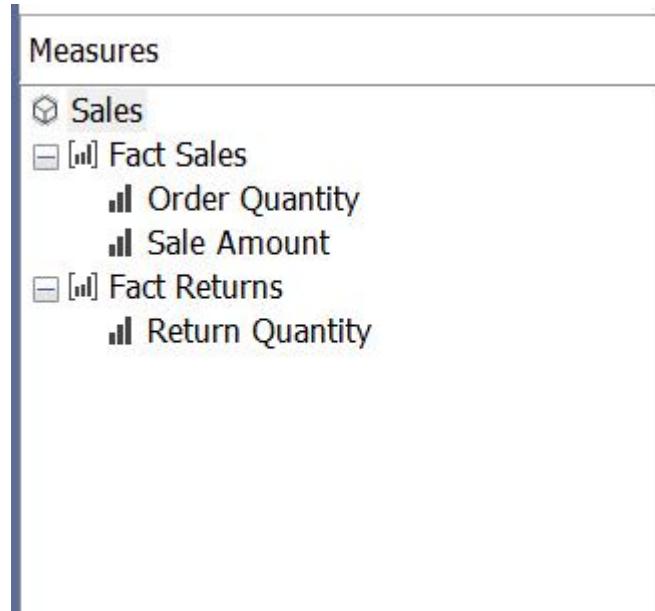


- Đặt tên cho Data Source View và nhấn Finish để hoàn tất việc tạo Data Source View
- Double-Click vào Data Source View vừa tạo để hiển thị diagram



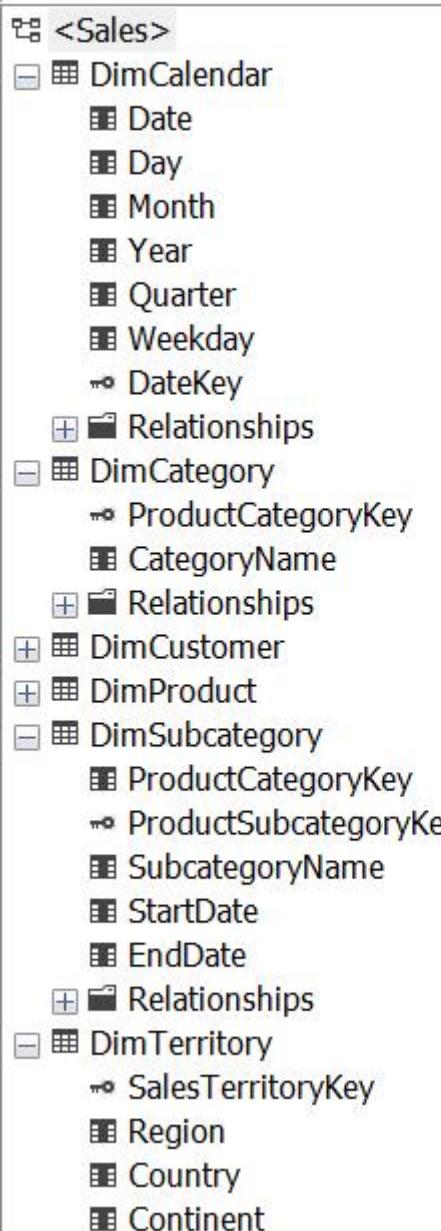
4.2. Quá trình xây dựng khối Cube

- Right-click vào **Cubes** ở khung Solution Explorer. Chon New Cube để tạo Cube mới.
- Chọn Data Source View là **AWPDW**. Tích chọn 2 bảng **FactSales** và **FactReturns**, nhấn Next
- Ở dialog Select Measures, tích chọn những measure cần thiết các mục sau

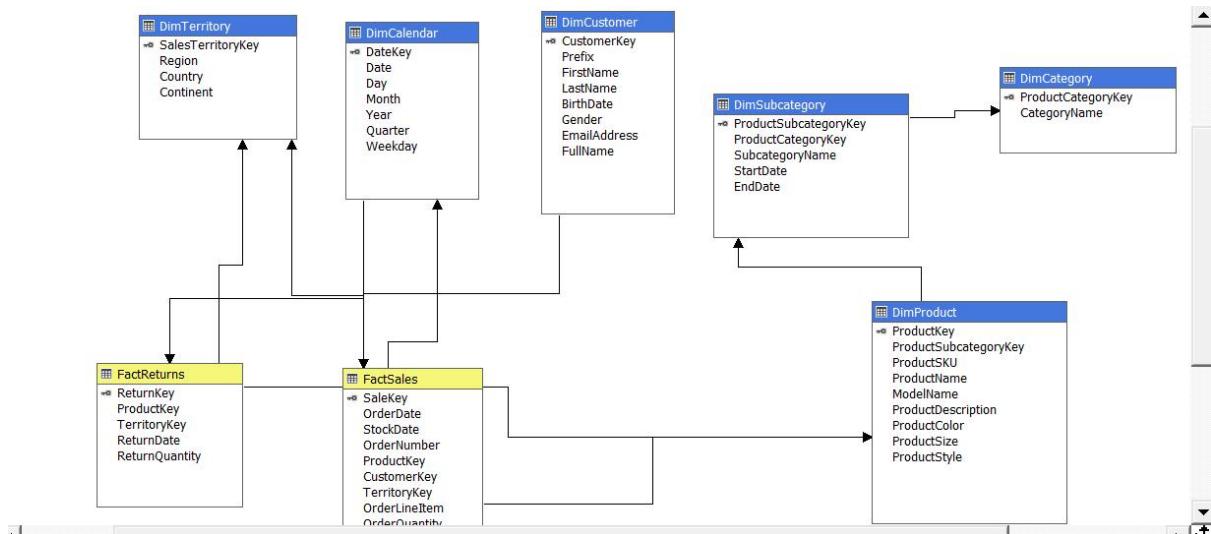


- Khởi tạo các bảng Dimension

Data Source View



- Đặt tên **Sales.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 Dem Calendar

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

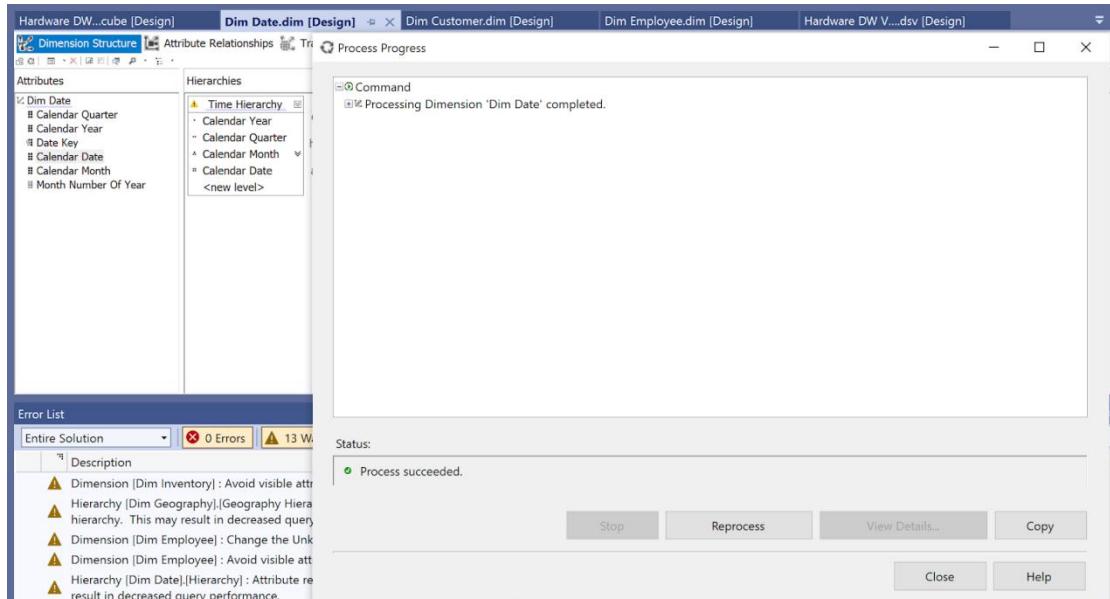
The screenshot shows the Analysis Services Dimension Structure tab. The top navigation bar includes tabs for Dimension Structure, Attribute Relationships, Translations, and Browser. Below the tabs, there are buttons for search, refresh, and other navigation functions.

Attributes	Hierarchies	Data Source View
Calendar <ul style="list-style-type: none"> Date Day Month Quarter Weekday Year 	Hierarchy <ul style="list-style-type: none"> Year Month Date Quarter <new level> <p>To create a new hierarchy, drag an attribute here.</p>	DimCalendar <ul style="list-style-type: none"> DateKey Date Day Month Year Quarter Weekday

- 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.

Current level: (All)

- All
- + 2015
 - + January
 - + February
 - + March
- + 2016
 - + Q1 CY 2016
 - + January
 - + February
 - + March
 - + Q2 CY 2016
 - + Q3 CY 2016
 - + Q4 CY 2016
- + 2017
- + 2018
- + 2019
- + 2020

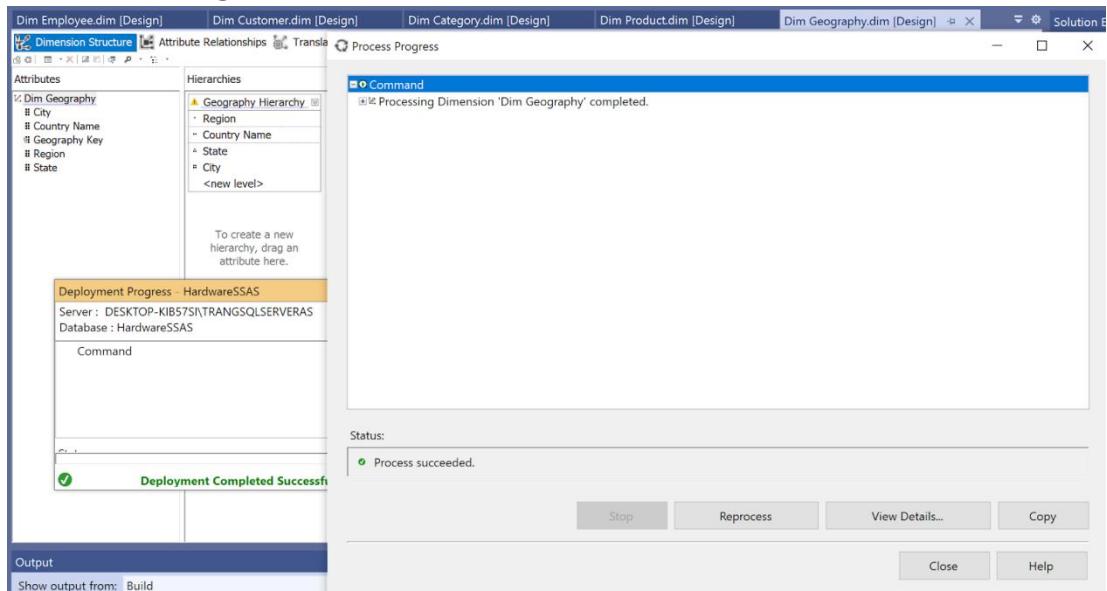
4.3.2. Tạo Hierarchy cho Dim Customer

- Tại table Dim Customer 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 Customer Hierarchy

The screenshot shows the Microsoft Analysis Services Dimension Designer. The title bar displays multiple open projects: Product.dim [Design], AWPDW.dsv [Design], Calendar.dim [Design], Territory.dim [Design], and Sales.cube [Design]. The active tab is Customer.dim [Design]. The interface has three main panes: Attributes, Hierarchies, and Data Source View.

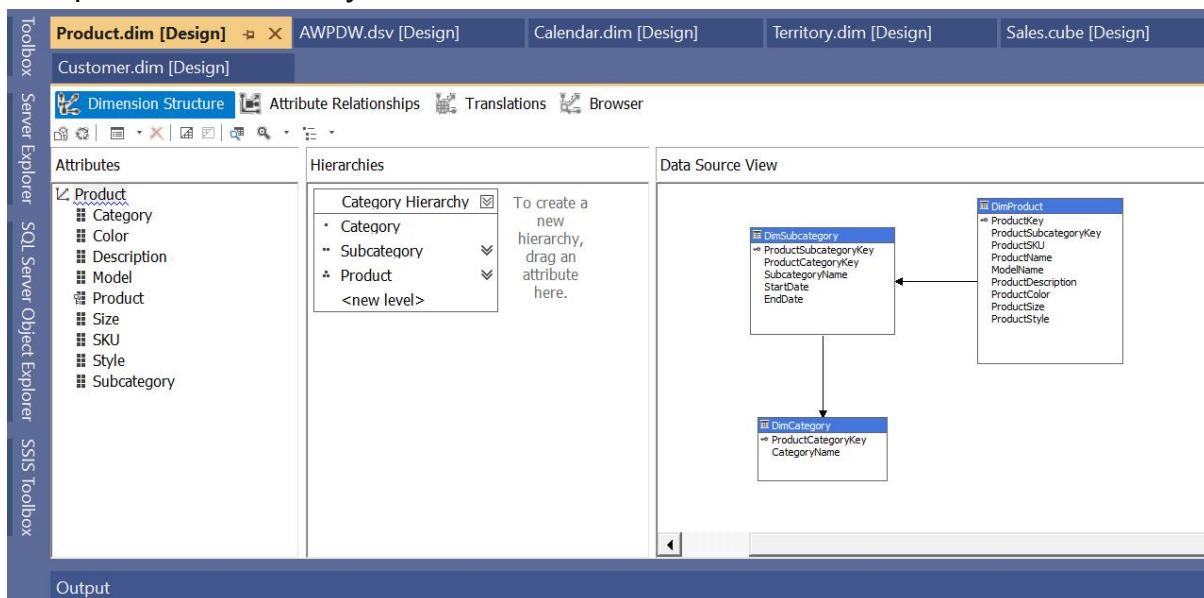
- Attributes pane:** Shows a tree structure under the Customer node, listing attributes: Birth Date, Customer, Email Address, First Name, Gender, and Last Name.
- Hierarchies pane:** Displays a hierarchy structure with a root node '_Hierarchy'. Underneath it are levels: First Name, Birth Date, Gender, Last Name, and Email Address, followed by '<new level>'. A tooltip below the hierarchy pane states: "To create a new hierarchy, drag an attribute here."
- Data Source View pane:** Shows the DimCustomer table with columns: CustomerKey, Prefix, FirstName, LastName, BirthDate, Gender, EmailAddress, and FullName. The 'CustomerKey' column is highlighted with a blue selection bar.

- Process bảng Dim Customer



4.3.3. Tạo Hierarchy cho Dim Product

- Tạo Product Hierarchy



4.3.4. Tao Hierarchy cho Dim Territory

The screenshot shows the Microsoft SQL Server Analysis Services (SSAS) Dimension Designer. The title bar indicates the current workspace is "Territory.dim [Design]". The left sidebar lists other dimensions: Product.dim [Design], AWPDW.dsv [Design], Calendar.dim [Design], Customer.dim [Design], and Sales(cube [Design]). The main area has tabs for "Dimension Structure", "Attribute Relationships", "Translations", and "Browser". The "Dimension Structure" tab is selected.

The "Attributes" pane on the left shows a tree structure under the "Territory" attribute, which includes "Continent", "Country", "Region", and "Territory".

The "Hierarchies" pane in the center contains a tree view of existing hierarchies. One hierarchy is shown, containing levels for "Continent", "Country", "Region", and "<new level>". A tooltip below the tree states: "To create a new hierarchy, drag an attribute here."

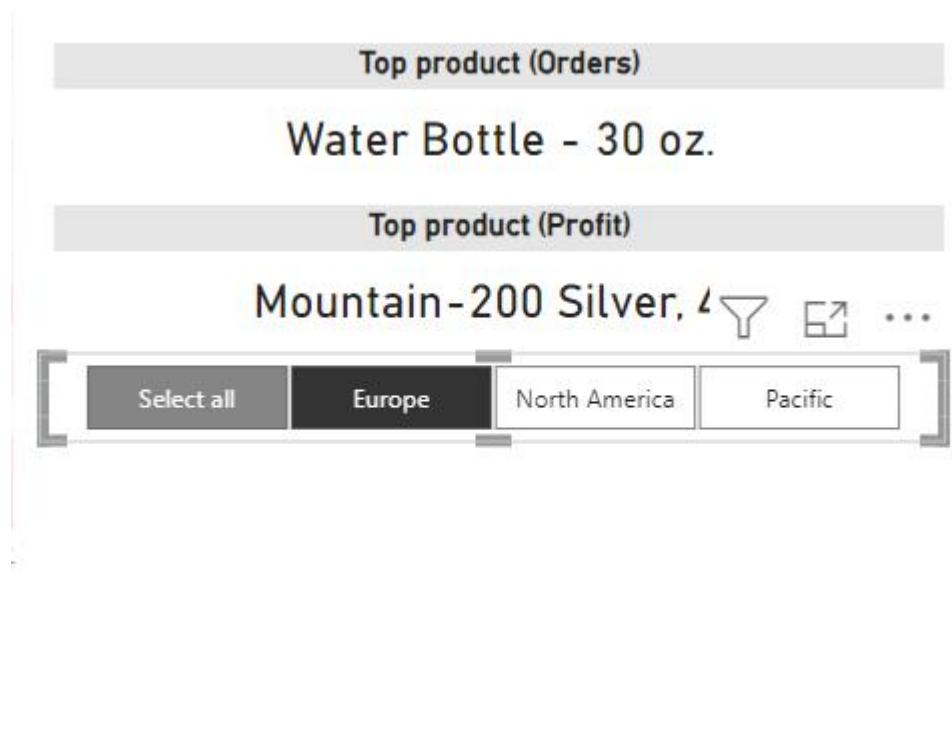
The "Data Source View" pane on the right displays the structure of the "Dim Territory" table, which includes attributes: SalesTerritoryKey, Region, Country, and Continent.

4.4. Thực hiện phân tích dữ liệu

4.4.1. Câu hỏi: Cho biết sản phẩm nào có số lượng bán ra cao nhất

4.4.1.1. Sử dụng Power BI

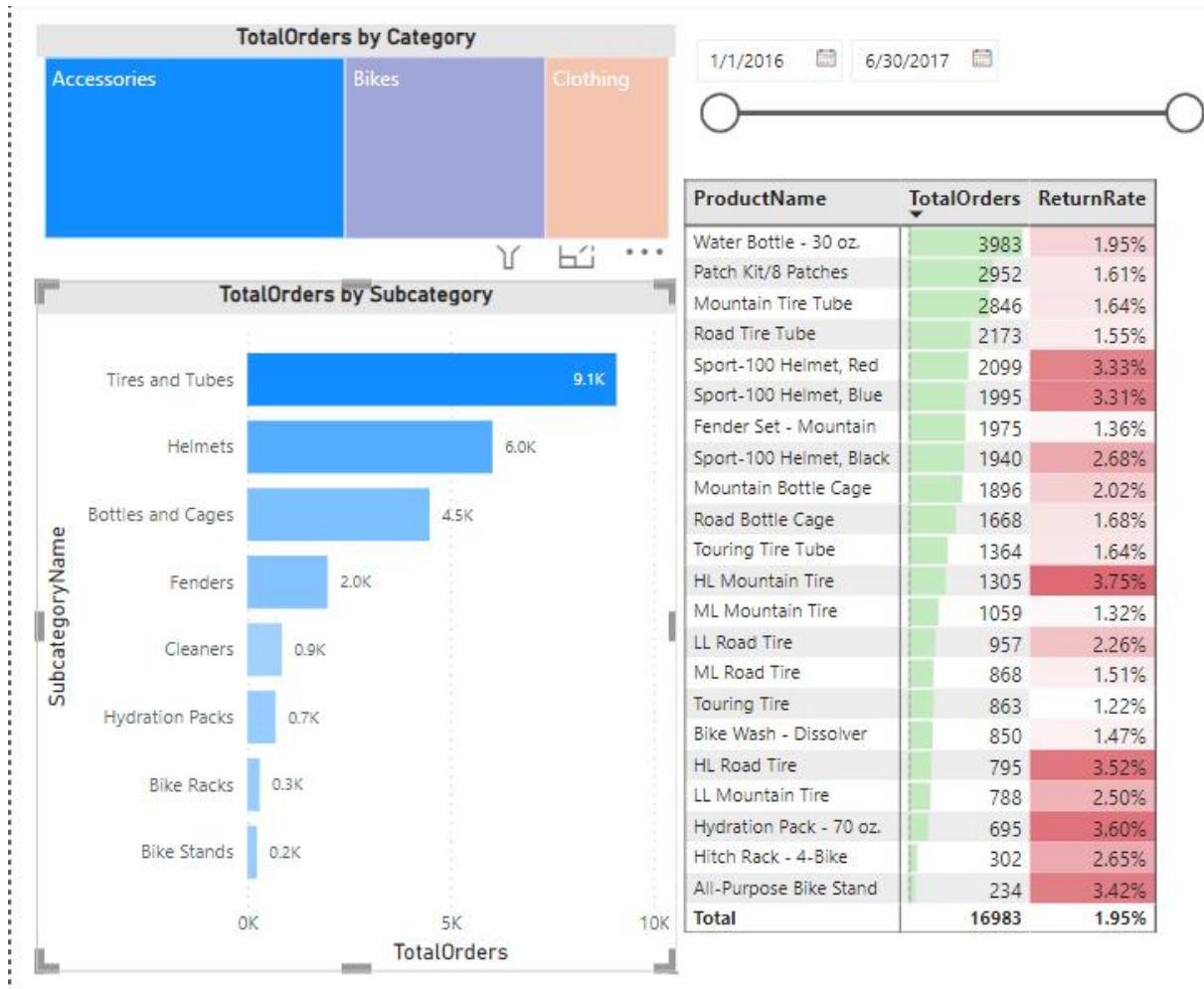
- Biểu đồ sản phẩm nào có số lượng bán ra cao nhất ở Europe từ 1/1/2016 đến 30/6/2017



4.4.2. Câu hỏi: Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất.

4.4.2.1. Sử dụng Power BI

- Danh mục các sản phẩm được tiêu thụ nhiều nhất từ 1/1/2016 đến 30/6/2017.



- Nhìn vào biểu đồ ngang, ta có thể thấy sản phẩm Water Bottle là sản phẩm có lượt tiêu thụ cao nhất với gần bốn nghìn lượt bán ra.

4.4.3. Câu hỏi: Cho biết doanh số bán hàng theo từng khách hàng.

4.4.3.1. Sử dụng Power BI

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

FullName_CC	TotalOrders	TotalRevenue
Mr. Maurice Shan	6	\$12,407.95
Mrs. Janet Munoz	6	\$12,015.39
Mrs. Lisa Cai	7	\$11,330.44
Mrs. Lacey Zheng	7	\$11,085.74
Mr. Jordan Turner	7	\$11,022.38
Mr. Franklin Xu	6	\$10,863.44
Mr. Larry Munoz	7	\$10,852.04
Mrs. Kate Anand	4	\$10,436.51
Mr. Larry Vazquez	4	\$10,394.98
Mrs. Ariana Gray	6	\$10,391.42
Mr. Clarence Gao	4	\$10,331.73
Mr. Aaron Wright	4	\$10,329.22
Mr. Ethan Bryant	4	\$10,308.52
Mr. Marco Lopez	6	\$10,289.68
Mrs. Bonnie Nath	4	\$10,282.91
Mr. Ricky Vazquez	4	\$10,165.92
Mr. Clarence Anand	4	\$10,065.01
Mrs. Deanna Perez	4	\$9,761.59
Mrs. Desiree Dominguez	4	\$9,717.63
Mrs. Jasmine Stewart	4	\$9,716.98
Mrs. Sydney Bryant	4	\$9,706.89
Mr. Bryant Garcia	4	\$9,687.36
Mrs. Cindy Patel	4	\$9,668.02
Mrs. Amy Sun	4	\$9,534.15
Mr. Willie Xu	5	\$9,424.97
Mrs. Katrina Tang	4	\$9,302.58
Total	373	\$885,306.56

CHƯƠNG 5: KẾT LUẬN

5.1. Kết quả đạt được

Thực hiện xây dựng '*KHO DỮ LIỆU BÁN HÀNG CHO CÔNG TY ADVENTURE WORK CYCLES*' giúp chúng em hiểu sâu hơn về kiến thức môn học Kho Dữ liệu và nâng cao khả năng sử dụng các công cụ hỗ trợ quá trình ETL như SSIS, SSAS, cũng như các công cụ dùng để trực quan hóa dữ liệu như Excel Pivot và Power BI Desktop. Bằng việc thực hiện dự án này, chúng em được thành thạo hơn về kỹ năng nhìn nhận và tiếp thu thông tin đầu vào, đặt câu hỏi tiền đề và thảo luận để tìm ra hướng giải quyết và trả lời các câu hỏi liên quan đến dữ liệu.

5.2. Những hạn chế

Trong quá trình chúng em thực hiện xây dựng kho dữ liệu, đã có những khó khăn và thách thức xuất hiện do kiến thức của chúng em còn hạn chế. Chúng em nhận thức rằng để xây dựng một kho dữ liệu đáng tin cậy và toàn diện, chúng em cần phải nắm vững các nguyên tắc, phương pháp và công cụ trong lĩnh vực này.

Một trong những khó khăn quan trọng mà chúng em đã gặp là tập dữ liệu chưa được khai thác hết tiềm năng. Dữ liệu có sẵn chứa đựng nhiều thông tin quan trọng mà chúng em chưa thể tận dụng một cách tối đa. Chúng em nhận thấy rằng việc xác định và khai thác các dữ kiện này là một yếu tố quan trọng để xây dựng một kho dữ liệu đầy đủ và có giá trị.

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 9 May 2024].
- [3] Minewiskan(no date), *Lesson 4: Defining Advanced Attribute and Dimension Properties*. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/lesson-4-defining-advanced-attribute-and-dimension-properties?view=asallproducts-allversions>. [Accessed 9 May 2024].

[4] Nguyễn Văn Chúc (no date). *ETL Project From Excel Data Source to Star Schema with SSIS*. [online] youtube.com. Available at: https://www.youtube.com/watch?v=Yp8fXLnVCp8&ab_channel=ChucNguyenVan. [Accessed 9 May 2024].