|  |
| --- |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HỒ CHÍ MINH** |
| **KHOA CÔNG NGHỆ THÔNG TIN**  **🕯✡🕮🕮✡🕯**  **Logo ĐH Sư Phạm Kỹ Thuật TP Hồ Chí Minh-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 Tấn 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](#_Toc17119)

[1.1. Lý do chọn đề tài 1](#_Toc23871)

[1.2. Tổng quan về tập dữ liệu 1](#_Toc12714)

[1.2.1. Nguồn dữ liệu 1](#_Toc1078)

[1.2.2. Mô tả chi tiết tập dữ liệu 2](#_Toc11741)

[1.2.2.1. Bảng AdventureWorks\_Calendar 3](#_Toc23288)

[1.2.2.2. Bảng AdventureWorks\_Sales\_2015 3](#_Toc27558)

[1.2.2.3. Bảng AdventureWorks\_Customers 4](#_Toc28426)

[1.2.2.4. Bảng AdventureWorks\_Product 4](#_Toc3141)

[1.2.2.5. Bảng AdventureWorks\_Terrirories 5](#_Toc10066)

[1.2.2.6. Bảng AdventureWorks\_Product\_Categories 6](#_Toc798)

[1.2.2.7. Bảng AdventureWorks\_Product\_Subcategories 6](#_Toc6486)

[1.2.2.8. Bảng AdventureWorks\_Return 6](#_Toc2186)

[1.2.2.9. Bảng Bảng AdventureWorks\_Sales\_2016 6](#_Toc22856)

[1.2.2.10. Bảng Bảng AdventureWorks\_Sales\_2017 8](#_Toc12332)

[1.3. Giới thiệu các công cụ được sử dụng 8](#_Toc9110)

[CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP) 9](#_Toc9927)

[2.1. Xác định các Business Process và bảng Fact 9](#_Toc29657)

[2.1.1. Business Process: Sales Reporting 9](#_Toc31721)

[2.1.2. Business Process: Return 9](#_Toc31827)

[2.2. Xây dựng các bảng Dimension 10](#_Toc15860)

[2.2.1. DimCalendar 10](#_Toc19045)

[2.2.2. DimTerritory 10](#_Toc9022)

[2.2.3. DimProduct 10](#_Toc1953)

[2.2.4. DimCustomer 10](#_Toc4448)

[2.2.5. DimSubcategory 11](#_Toc1115)

[2.2.6. DimCategory 11](#_Toc7159)

[2.3. Star Schema (Lược đồ hình sao) 11](#_Toc24557)

[CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS) 13](#_Toc7771)

[3.1. Tạo ETL project và solution 13](#_Toc1041)

[3.2. Thiết lập các Connection 14](#_Toc28697)

[3.3. Import Dữ liệu vào các bảng stage: 18](#_Toc5109)

[3.3.1. Customer 18](#_Toc29235)

[3.3.1.1: Tạo Package 18](#_Toc13555)

[3.3.1.2: Thiết lập Control Flow 20](#_Toc25462)

[3.3.1.3: Load dữ liệu từ Source vào Stage 22](#_Toc1560)

[3.3.1.4: Thiết lập truncate table SQL task 27](#_Toc7267)

[3.3.2.Product 29](#_Toc1248)

[3.3.2.1. Tạo Package 29](#_Toc6014)

[3.3.2.2. Thiết lập Control Flow 30](#_Toc20688)

[3.3.2.3. Load dữ liệu từ Source vào Stage 31](#_Toc4358)

[3.3.3.Categories 33](#_Toc3281)

[3.3.3.1. Load dữ liệu từ Source vào Stage 33](#_Toc8536)

[3.3.4. SubCategory 36](#_Toc21629)

[3.3.4.1. Load dữ liệu từ Source vào Stage 36](#_Toc3986)

[3.3.5.Calendar 39](#_Toc11452)

[3.3.5.1. Load dữ liệu từ Source vào Stage 39](#_Toc17970)

[3.3.6. Territory 42](#_Toc21092)

[3.3.6.1. Load dữ liệu từ Source vào Stage 42](#_Toc25026)

[3.3.7. Returns 44](#_Toc7864)

[3.3.7.1. Load dữ liệu từ Source vào Stage 44](#_Toc20949)

[3.3.8. Sales 46](#_Toc14583)

[3.3.8.1. Load dữ liệu từ Source vào Stage 46](#_Toc1748)

[3.4. Import dữ liệu từ bảng stage vào bảng dimension 49](#_Toc17761)

[3.4.1. DimCustomer 49](#_Toc15228)

[3.4.1.1. Tạo Package 49](#_Toc1367)

[3.4.1.2. Thiết lập Control Flow 49](#_Toc7925)

[3.4.1.3. Load dữ liệu từ Stage vào Dimension 50](#_Toc14718)

[3.4.2. DimCalendar 52](#_Toc3302)

[3.4.2.1. Load dữ liệu từ Stage vào Dimension 52](#_Toc24087)

[3.4.3. DimCategory 55](#_Toc11002)

[3.4.3.1. Load dữ liệu từ Stage vào Dimension 55](#_Toc23265)

[3.4.4. DimSubCategory 57](#_Toc16298)

[3.4.4.1. Load dữ liệu từ Stage vào Dimension 57](#_Toc19267)

[3.4.5. DimTerritory 60](#_Toc29776)

[3.4.5.1. Load dữ liệu từ Stage vào Dimension 60](#_Toc26767)

[3.4.6. DimProduct 62](#_Toc1227)

[3.4.6.1. Load dữ liệu từ Stage vào Dimension 62](#_Toc12192)

[3.5. Import dữ liệu vào bảng các bảng fact 64](#_Toc12598)

[3.5.1. Fact Sales 64](#_Toc25633)

[3.5.1.1. Tạo Package 64](#_Toc10892)

[3.5.1.2. Thiết lập Control Flow 65](#_Toc26422)

[3.5.1.3. Load dữ liệu từ Stage vào Fact 66](#_Toc14633)

[3.5.2. Fact Returns 70](#_Toc17514)

[3.5.2.1. Load dữ liệu từ Stage vào Fact 70](#_Toc5253)

[CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS) 73](#_Toc12044)

[4.1. Quá trình xây dựng mô hình 73](#_Toc23968)

[4.1.1. Tạo Data Source 73](#_Toc15194)

[4.1.2. Tạo Data Source View 74](#_Toc31223)

[4.2. Quá trình xây dựng khối Cube 76](#_Toc14899)

[4.3. Cấu hình Hierarchy 78](#_Toc15098)

[4.3.1. Tạo Hierarchy cho Dem Calendar 78](#_Toc4898)

[4.3.2. Tạo Hierarchy cho Dim Customer 80](#_Toc24157)

[4.3.3. Tạo Hierarchy cho Dim Product 81](#_Toc6149)

[4.3.4. Tạo Hierarchy cho Dim Territory 82](#_Toc5557)

[4.4. Thực hiện phân tích dữ liệu 83](#_Toc26265)

[4.4.1. Câu hỏi: Cho biết sản phẩm nào có số lượng bán ra cao nhất 83](#_Toc30787)

[4.4.1.1. Sử dụng Power BI 83](#_Toc20399)

[4.4.2. Câu hỏi: Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất. 83](#_Toc6112)

[4.4.2.1. Sử dụng Power BI 83](#_Toc14559)

[4.4.3. Câu hỏi: Cho biết doanh số bán hàng theo từng khách hàng. 85](#_Toc31323)

[4.4.3.1. Sử dụng Power BI 85](#_Toc21741)

[CHƯƠNG 5: KẾT LUẬN 86](#_Toc723)

[5.1. Kết quả đạt được 86](#_Toc12352)

[5.2. Những hạn chế 86](#_Toc29390)

[5.3. Tài liệu tham khảo 86](#_Toc14869)

**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

## 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.

## Tổng quan về tập dữ liệu

### 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>

### 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 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 |

#### Bảng AdventureWorks\_Calendar

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| Date | Thời gian(khóa chính) |

#### 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 |

#### 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 |
| Annuallncome | Thu nhập hằng năm của khách hàng |
| TotalChildren | Tổng số con của mỗi khách hàng |

#### 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 |

#### Bảng AdventureWorks\_Terrirories

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| TerritoryKey | Mã khu vực (Khóa chính) |
| Region | Khu vực địa lí |
| Country | Quốc gia |
| Continent | Châu lục |

#### Bảng AdventureWorks\_Product\_Categories

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

#### 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) ) |

#### 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 |

#### Bảng 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(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 |

#### Bảng 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 |

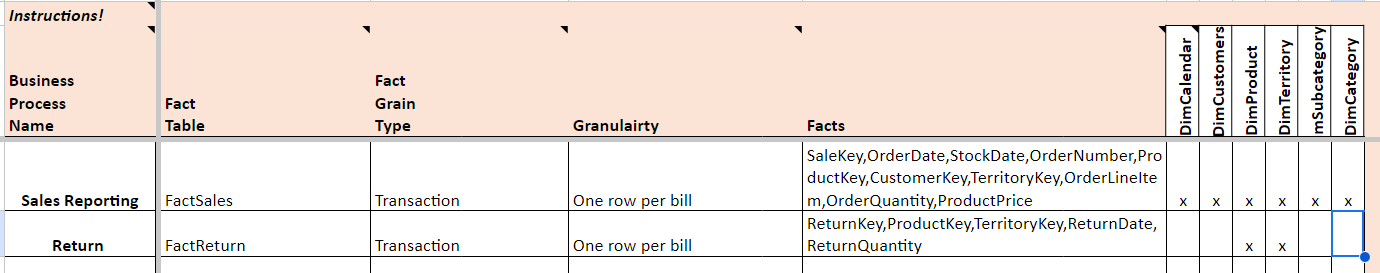
## Giới thiệu các công cụ được sử dụng

* 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.



### 2.1.1. Business Process: Sales Reporting

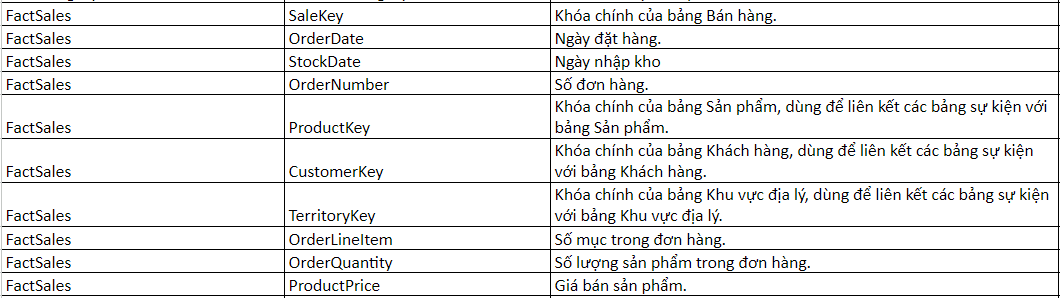
* Các câu hỏi cụ thể được đặt ra:

+ 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**

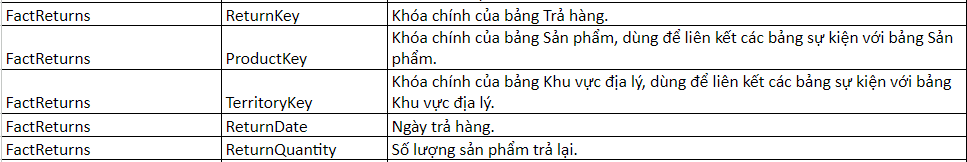
****

### 2.1.2. Business Process: Return

* Các câu hỏi cụ thể được đặt ra:

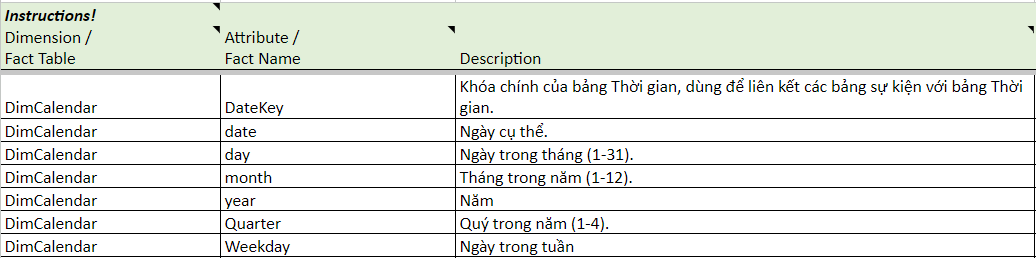
+ Cho biết tỷ lệ trả hàng theo từng sản phẩm.

* Bảng **FactReturns**

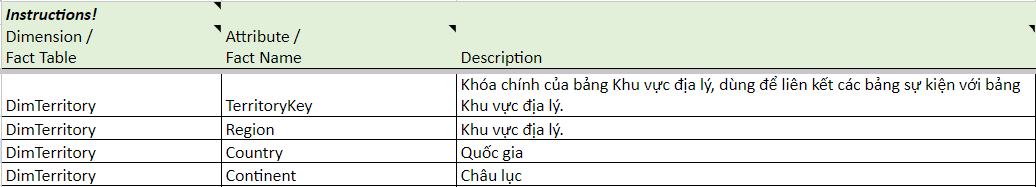
****

## 2.2. Xây dựng các bảng Dimension

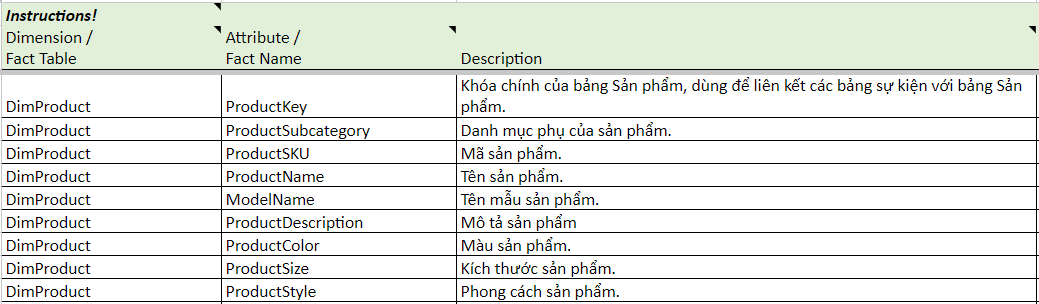
### 2.2.1. DimCalendar



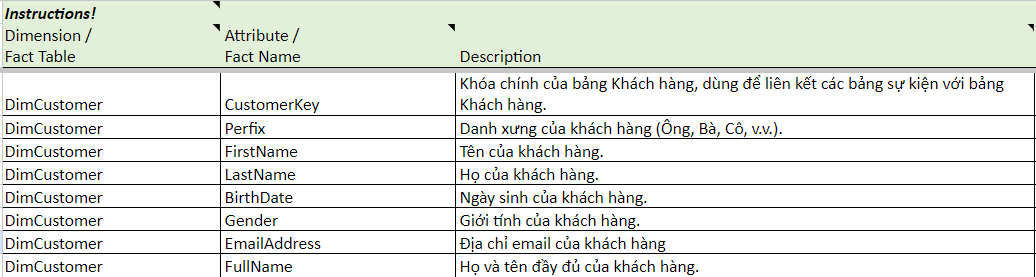
### 2.2.2. DimTerritory



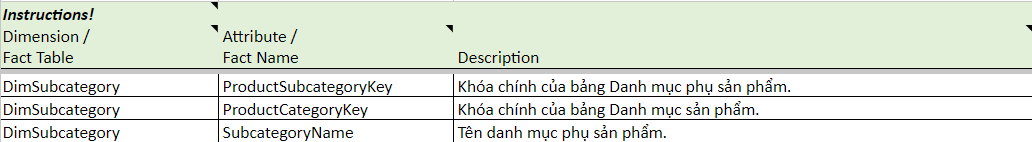
### 2.2.3. DimProduct



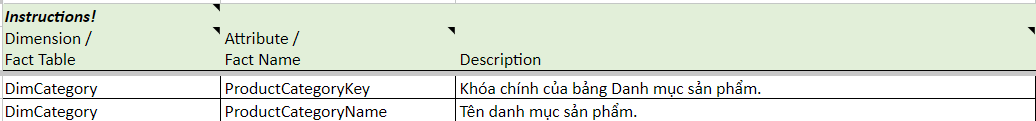
### 2.2.4. DimCustomer



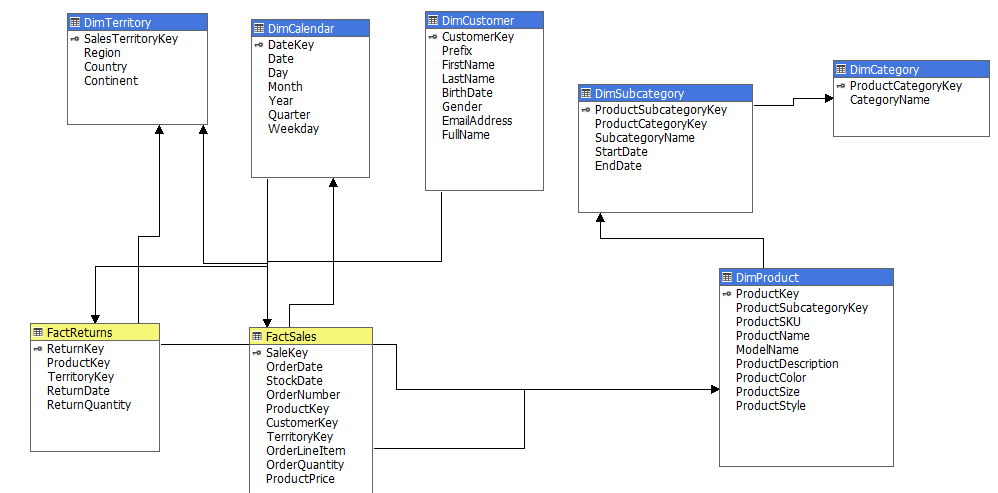
### 2.2.5. DimSubcategory



### 2.2.6. DimCategory



## 2.3. Star Schema (Lược đồ hình sao)



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

## 3.1. Tạo ETL project và solution

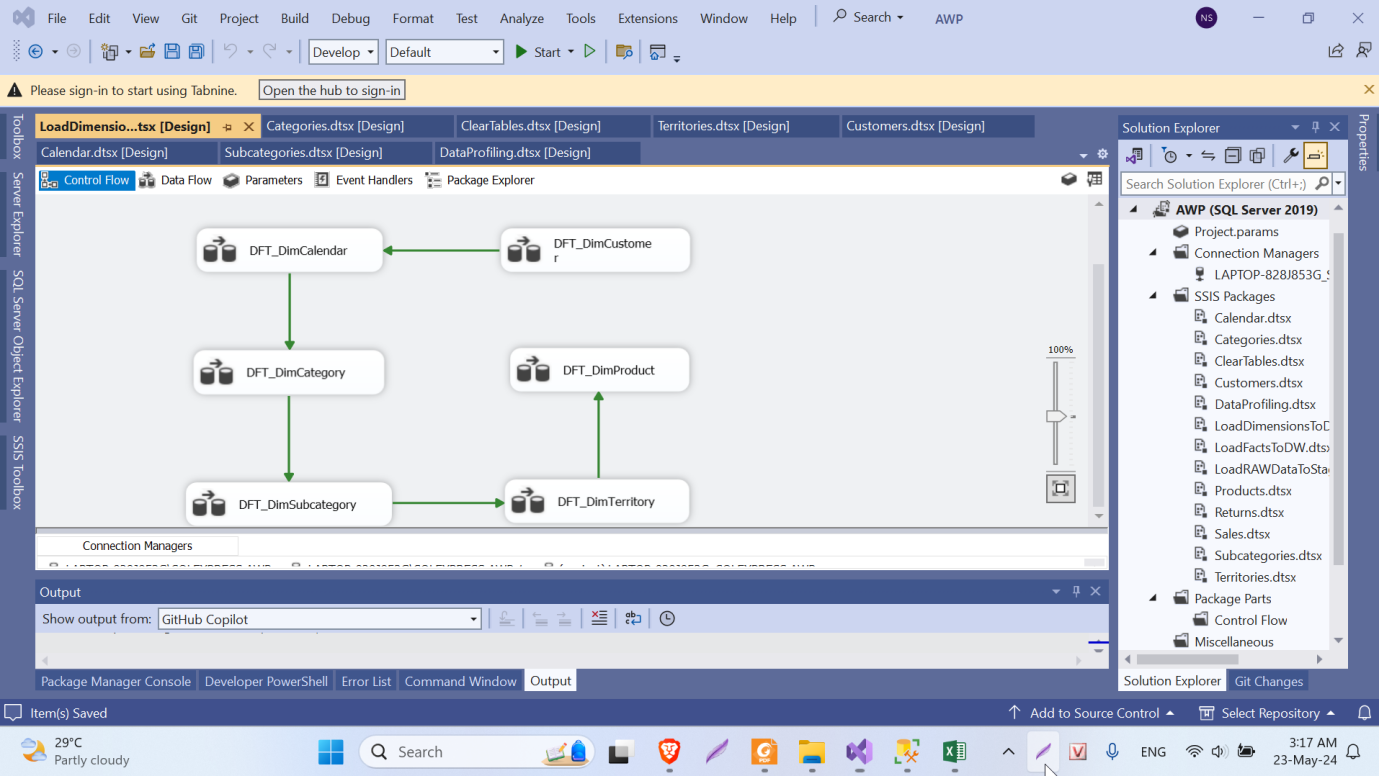
1. Từ **Visual Studio** menu, chọn **File** -> **New** -> **Project**.

2. Từ **dialog**:

a. Chọn **Integration Services Project**

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

3. Click **OK**

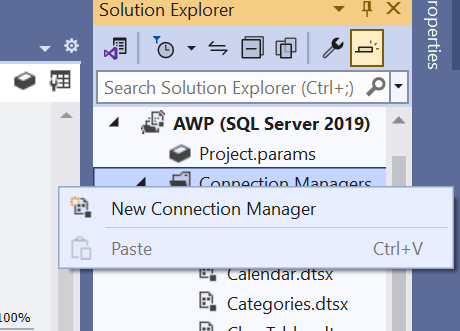


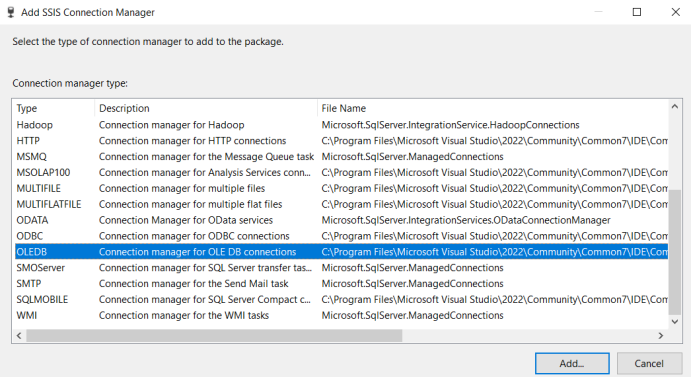
## 

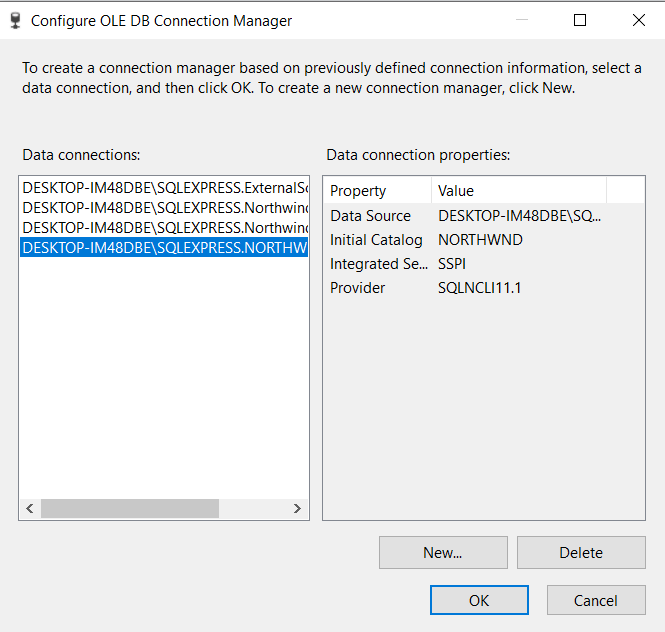
## 3.2. Thiết lập các Connection

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…**



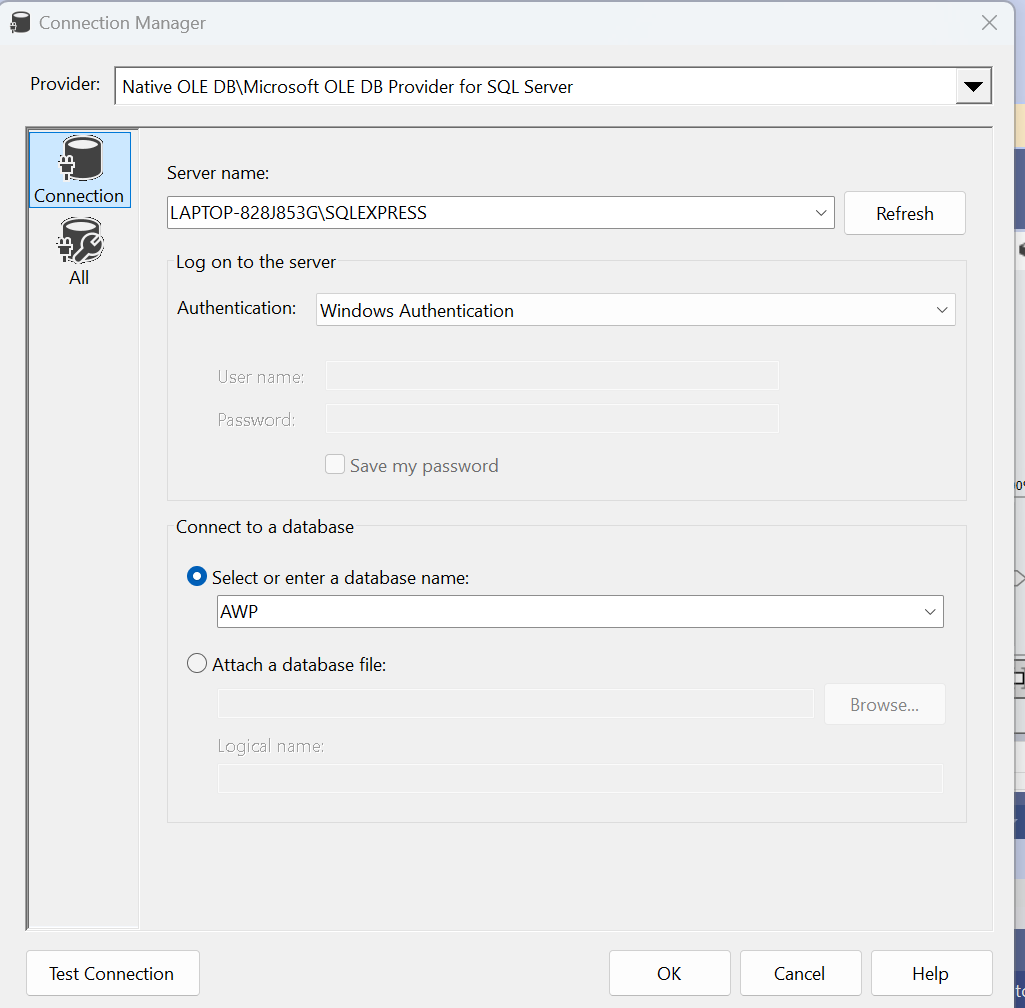
Đặt tên cho các thông tin như bên dưới:

Provider: **SQL Server Native Client**

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

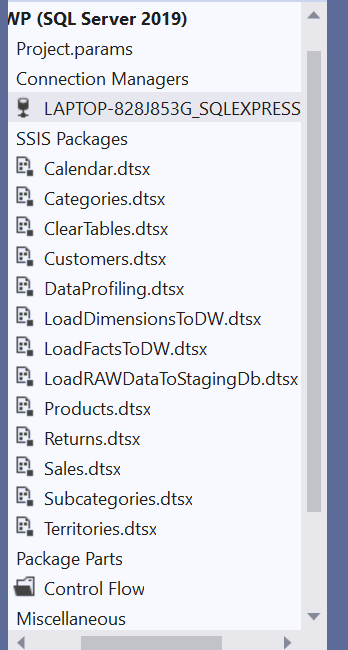
Log on to the Server: **Use Windows Authentication**

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.

## 

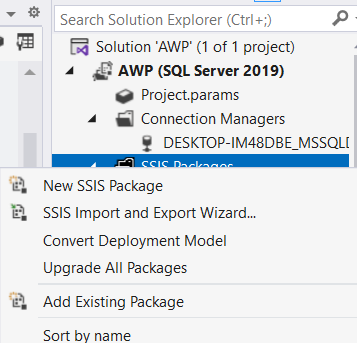
## 3.3. Import Dữ liệu vào các bảng stage:

### 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**

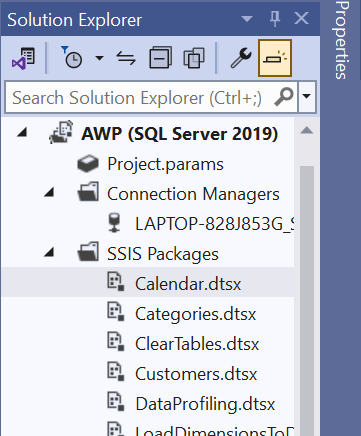


2. Right-Click vào package có tên **Package1.dstx** và chọn **Rename**

3. Đặt tên package là **Calendar**

4. Sau khi thực hiện xong ta được **Package.dtsx** và **Calendar.dtsx**

như hình sau:

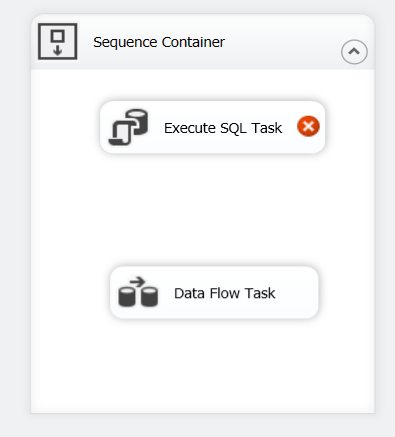


5. Double-click vào **Customer.dtsx** package để mở nó.

#### 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:

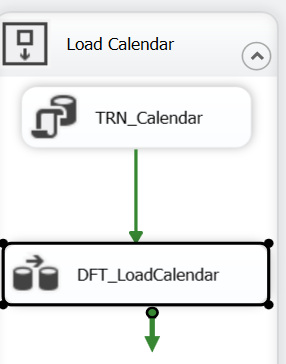


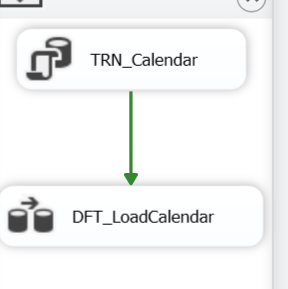
2. Đặt tên cho các **task**.

a. **Sequence Container - Load Calendar**

b. **Execute SQL Task - TRN\_Calendar**

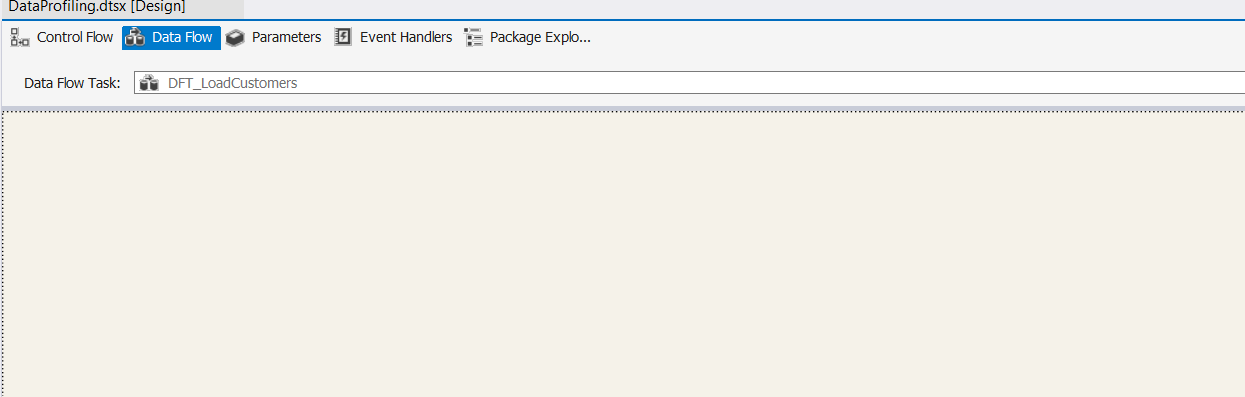
c. **Data flow tasks - DFT\_LoadCalendar**



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

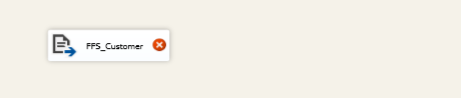
1. Double-click **DFT\_LoadCalendar** task để mở nó trong **Data Flow design surface**:



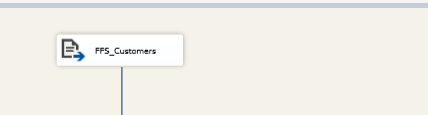
2. Từ **SSIS Toolbox**, kéo rê **Flat File Source** vào design surface

Click **OK**.

3. **Flat File Source** sẽ xuất hiện như sau:

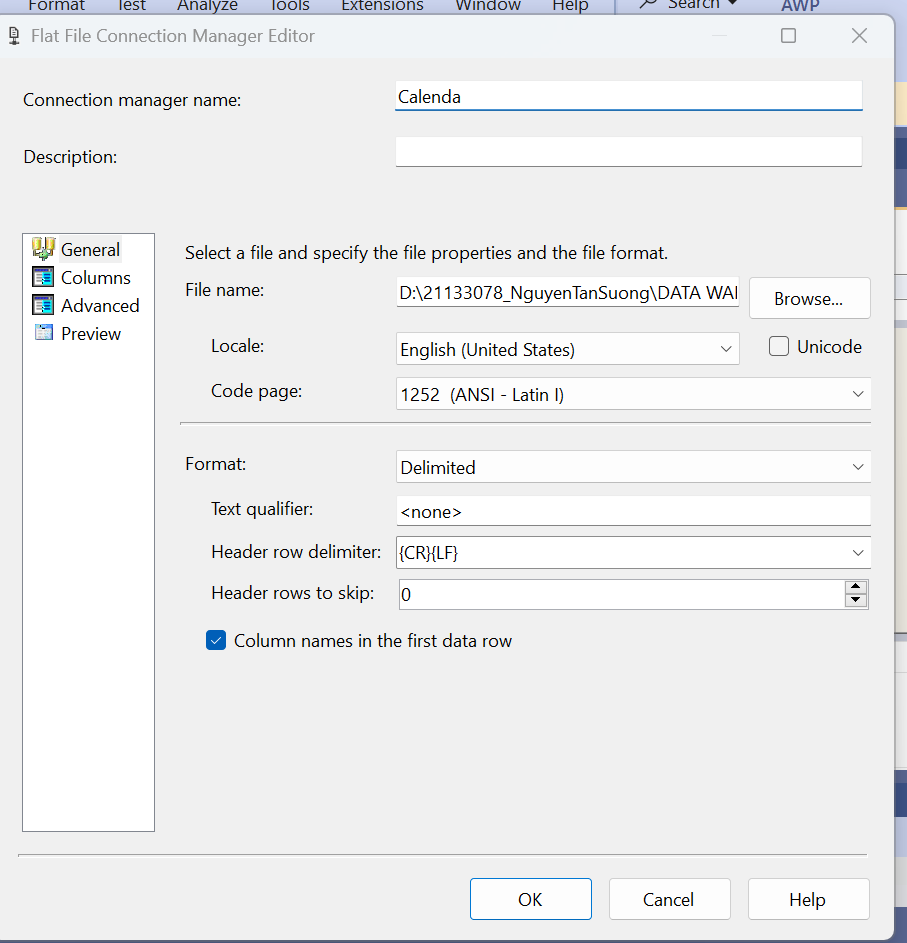


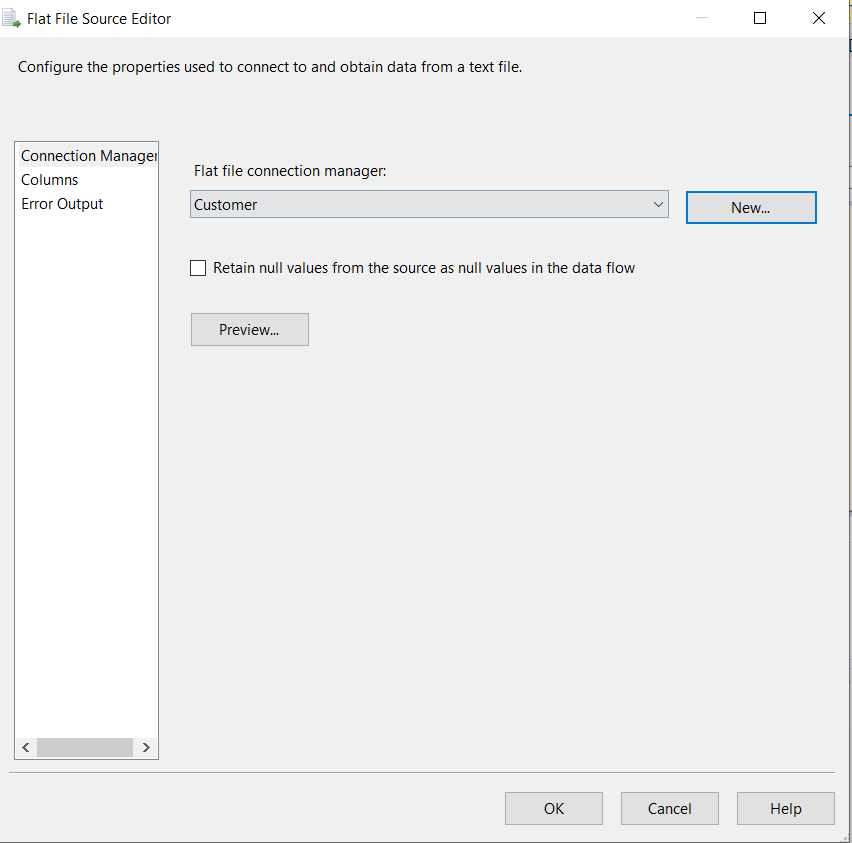
4. Đổi tên nó thành **FFS\_Customer**



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.**

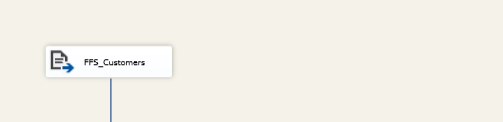
Click **Preview…** để xem source data.



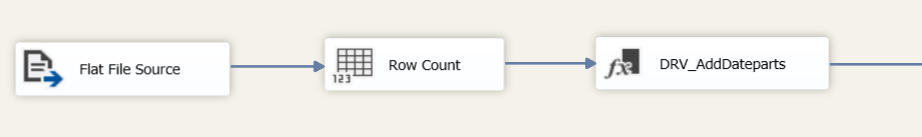


Click **OK** để lưu.

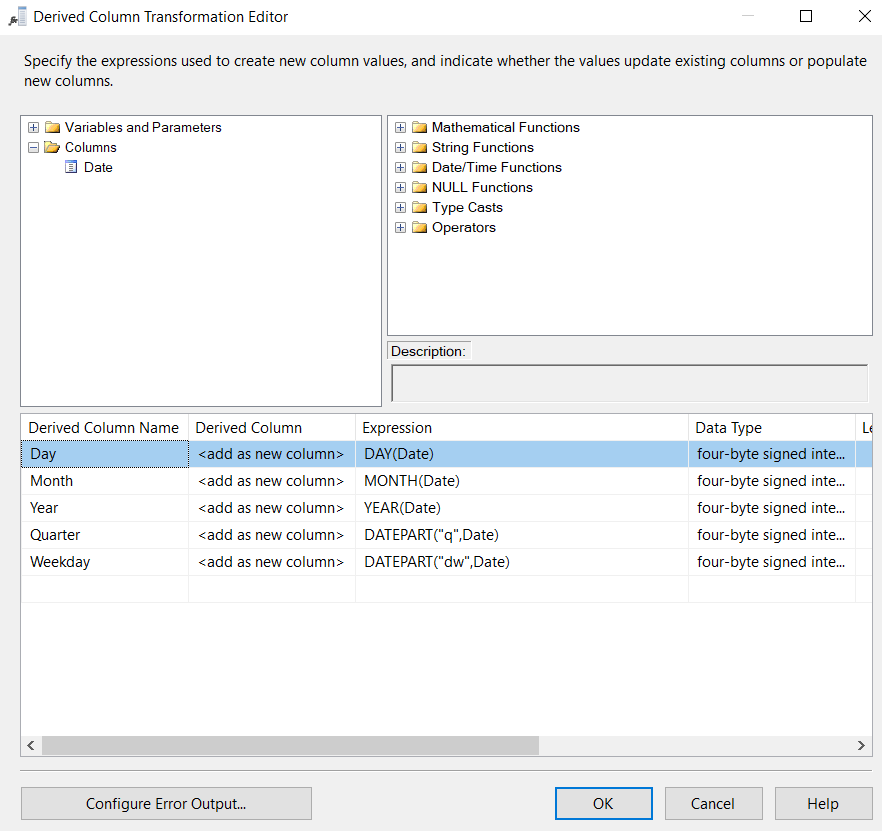
6. Sau khi cấu hình **source** xong, bạn sẽ không còn thấy dấu **[x]** màu đỏ (như bên dưới):



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.**



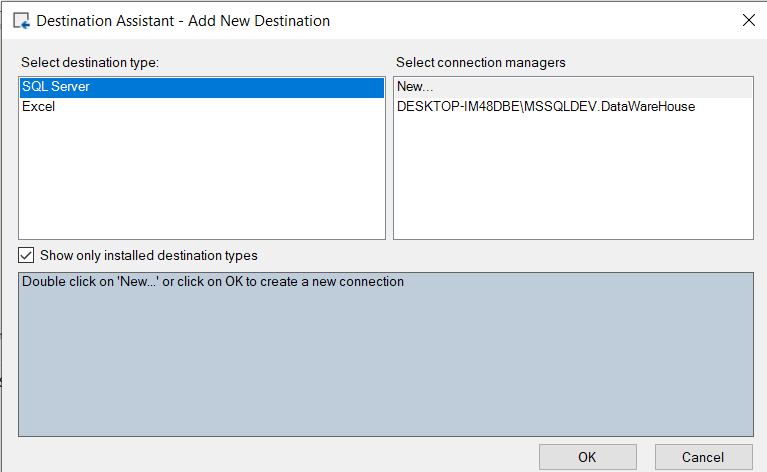
8. Double click vào **DRV\_AddDateparts.** Kéo rê cột Date vào ô Derived Column Name, ta được Derived Column 1, đổi tên lại thành Day và thay đổi Expression. Làm tương tự để tạo thành các cột mới là Month, Year, Quarter, Weekday.

****

Click OK

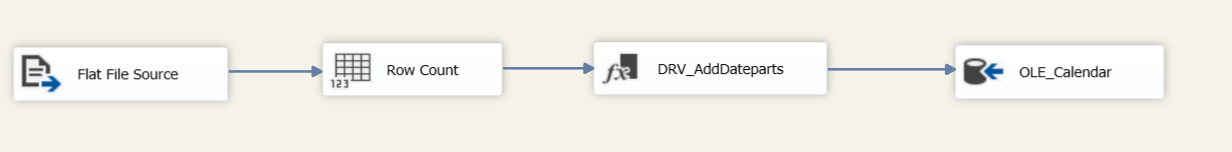
9. Tiếp theo, cấu hình destination. Từ **SSIS Toolbox**, kéo rê **Destination Assistant** vào

design surface, xuất hiện dialog



10. Chọn Connection DataWareHouse rồi click OK

11. Đổi tên **Destination Assistant** thành **OLE\_Calendar**,ta được design surface:

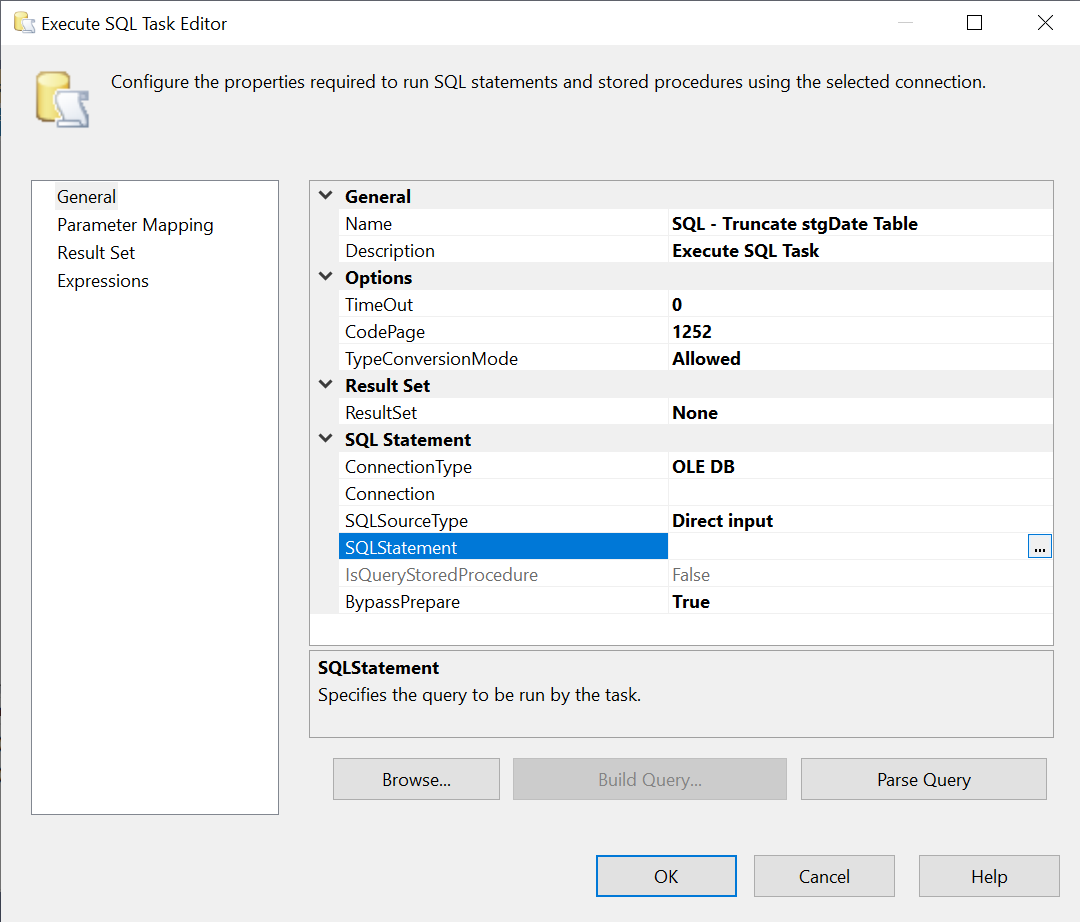


#### 3.3.1.4: Thiết lập truncate table SQL task

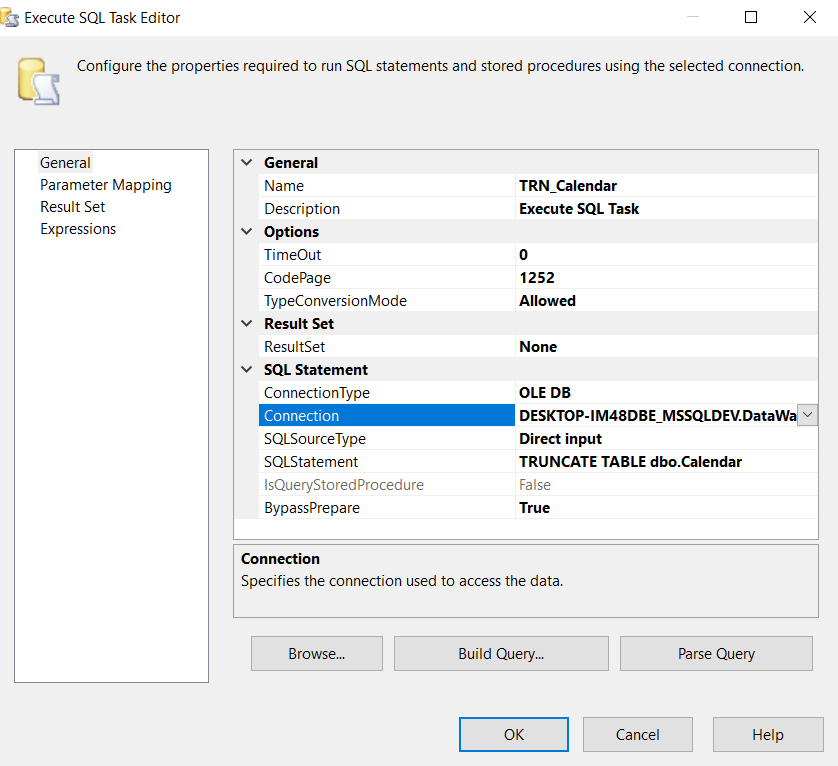
Click **Control Flow** tab.

1. Double-click vào **TRN\_Calendar** task để cấu hình nó, **Execute SQL**

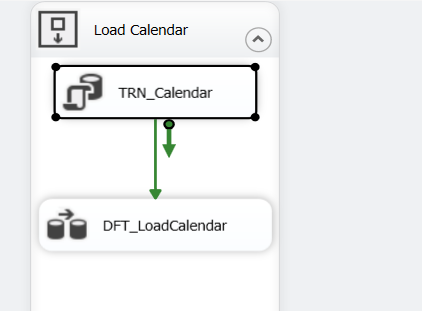
**Task Editor** sẽ mở ra.



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

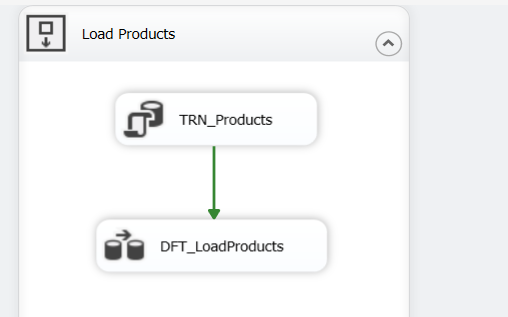
- Tạo package mới và đổi tên như hình bên dưới.



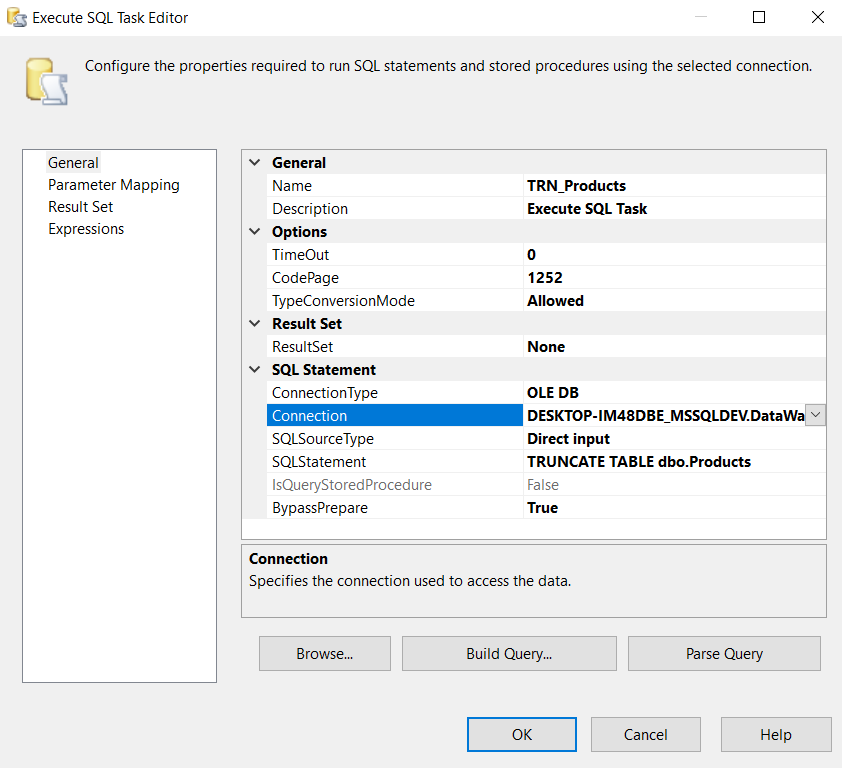
#### 

#### 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

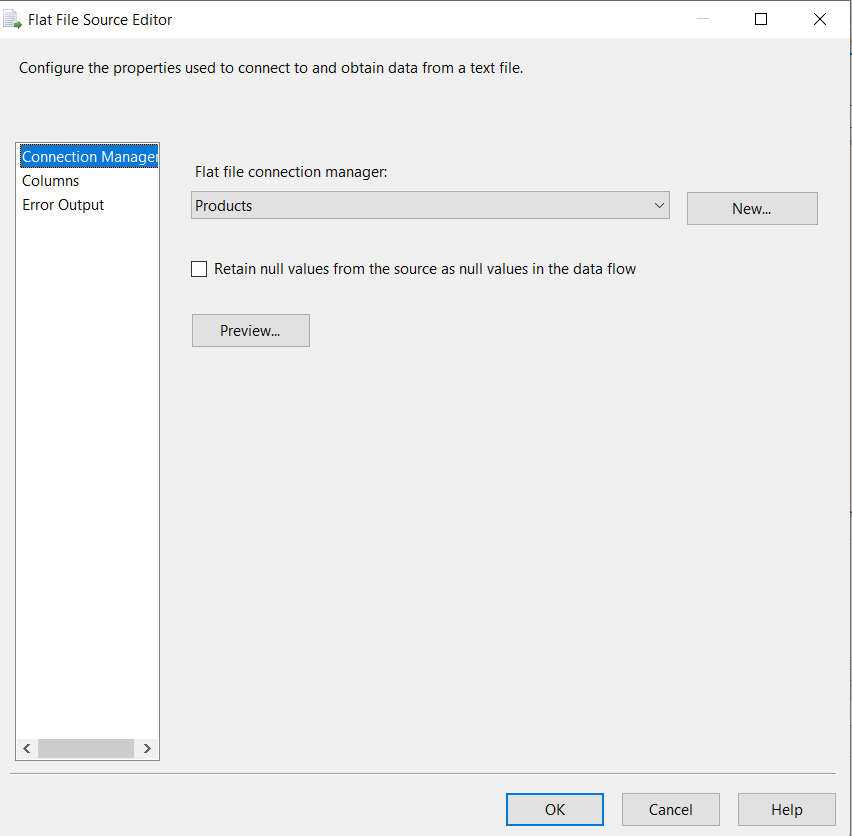


#### 3.3.2.3. Load dữ liệu từ Source vào Stage

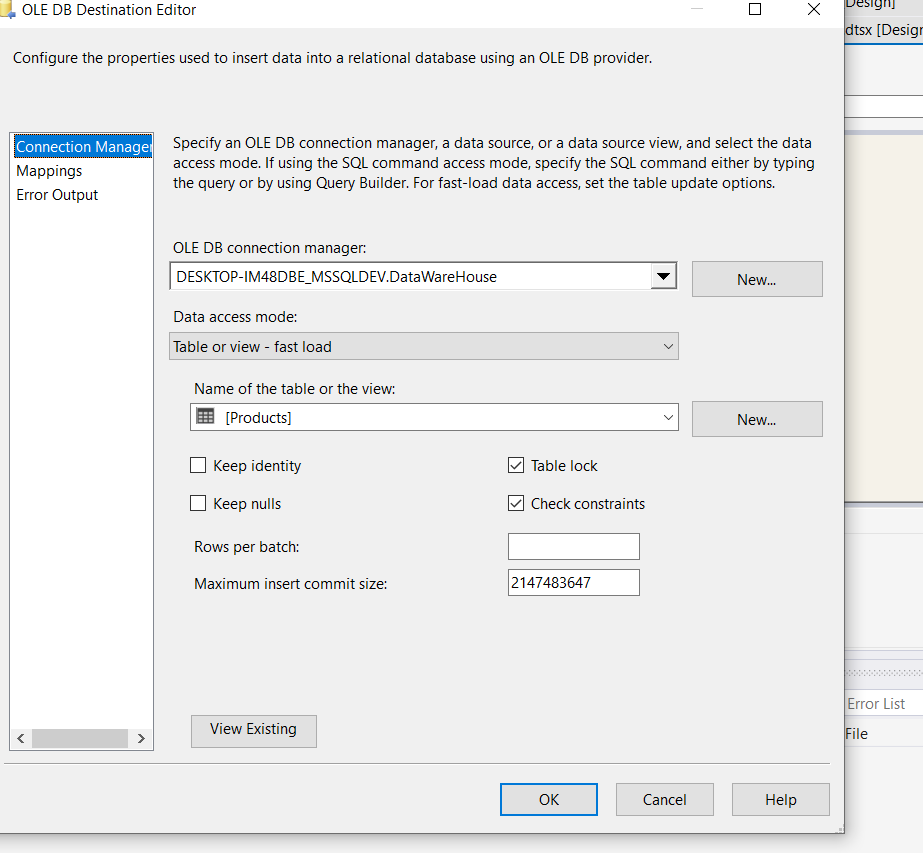
- Double-click **DFT\_LoadProducts** task để mở nó trong **Data Flow design surface:**

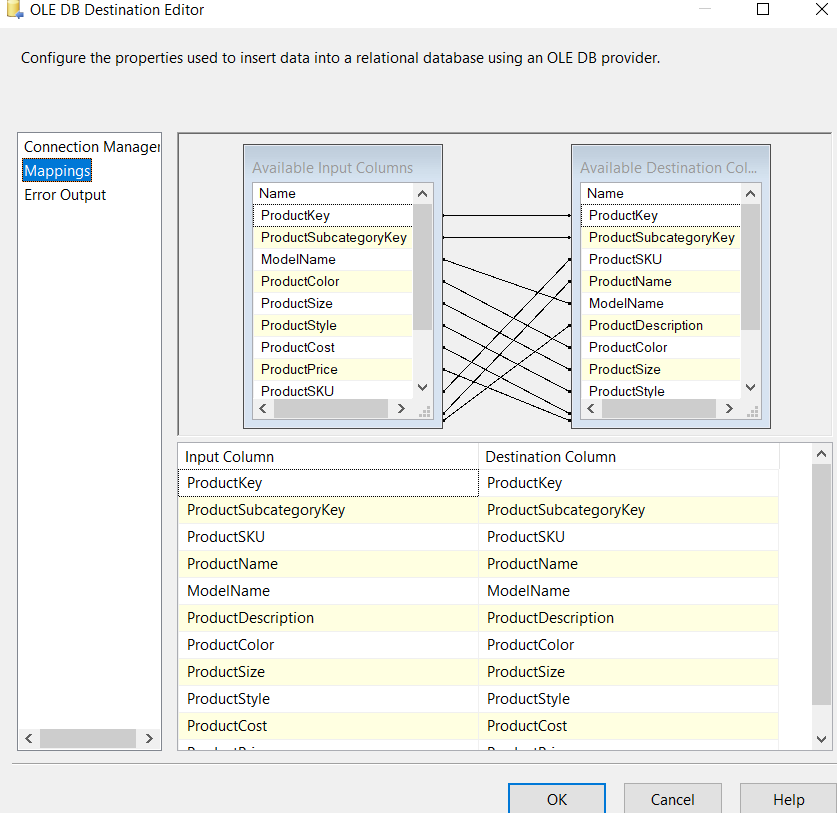
- Tạo **Flat File Source**, **Destination Assistant**, **Row Count** cấu hình chúng như sau:

+ **Flat File Source**:

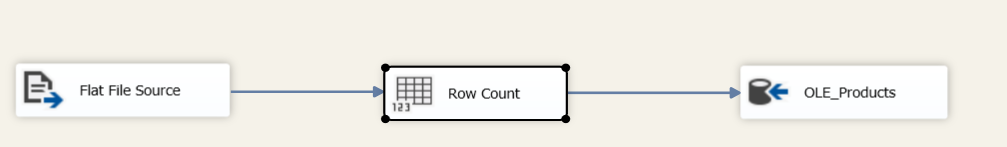


**+ Destination Assistant:**





+ Sau khi cấu hình xong ta được:



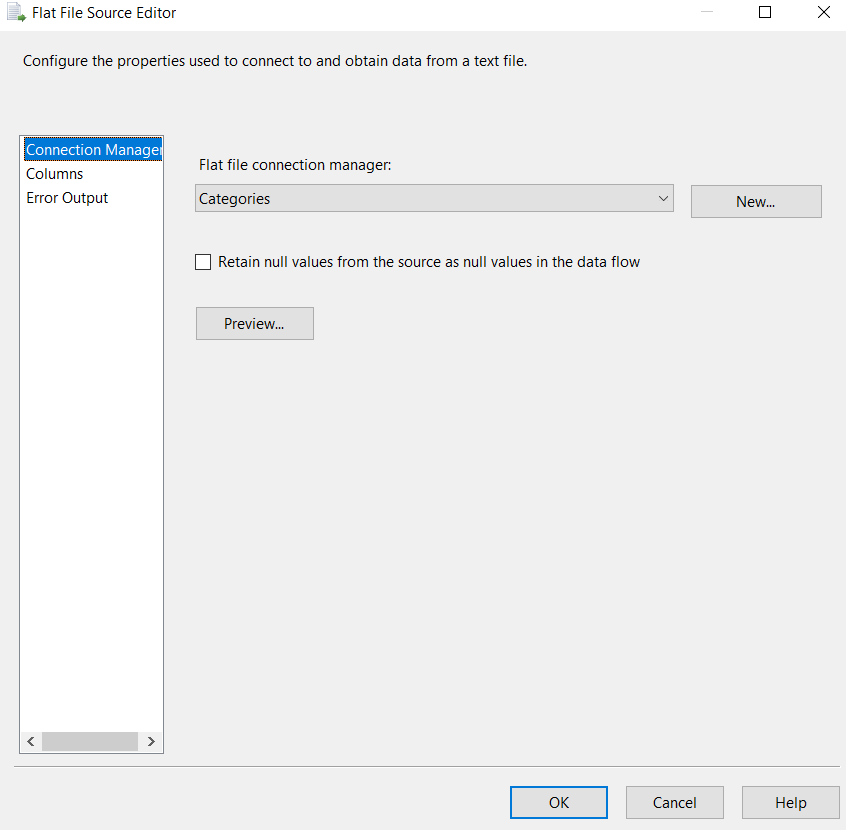
### 3.3.3.Categories

#### 3.3.3.1. Load dữ liệu từ Source vào Stage

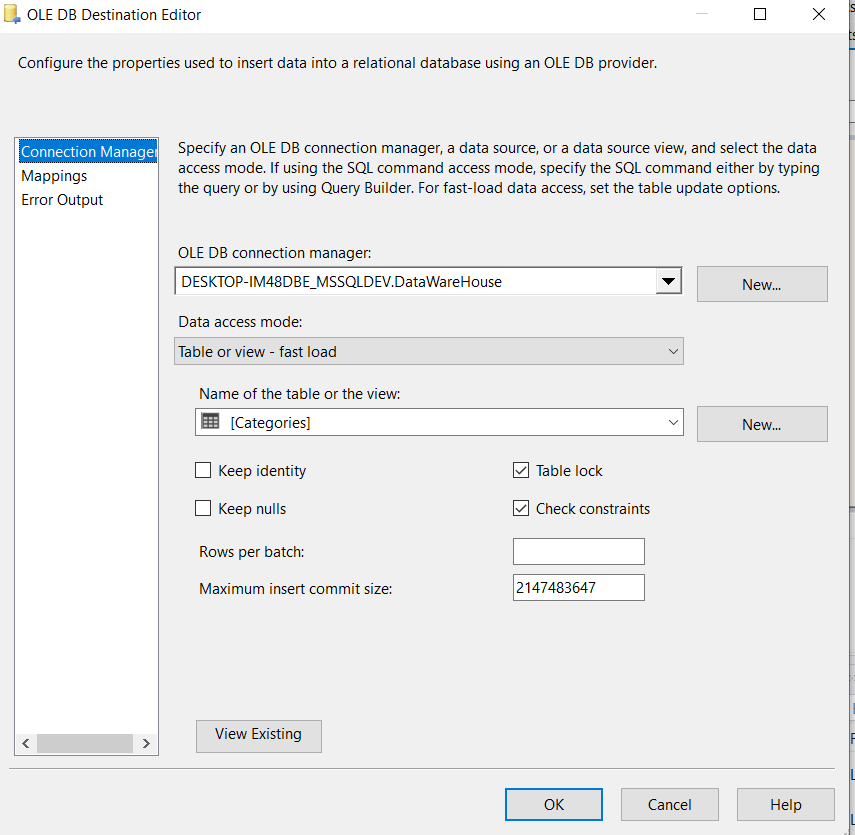
- Double-click **DFT\_LoadCategories** task để mở nó trong **Data Flow design surface:**

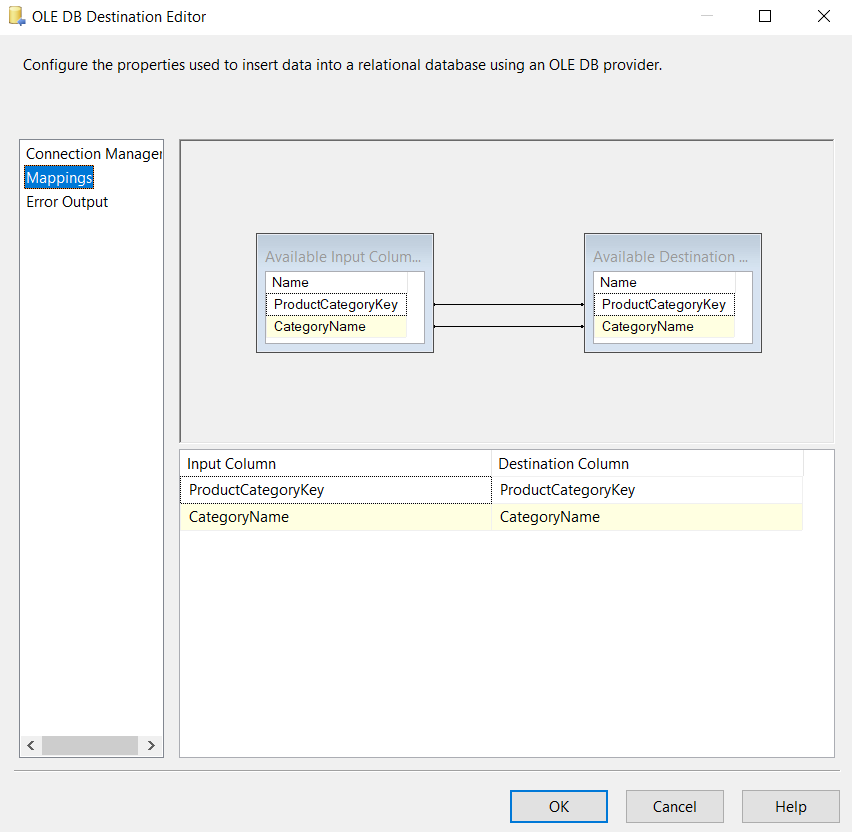
- Tạo **Flat File Source, RowCount** và **Destination Assistant** và cấu hình chúng như sau:

+ **Flat File Source**

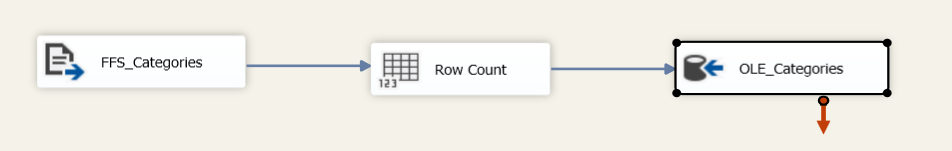


+ Destination Assistant:





+ Sau khi cấu hình xong ta được:



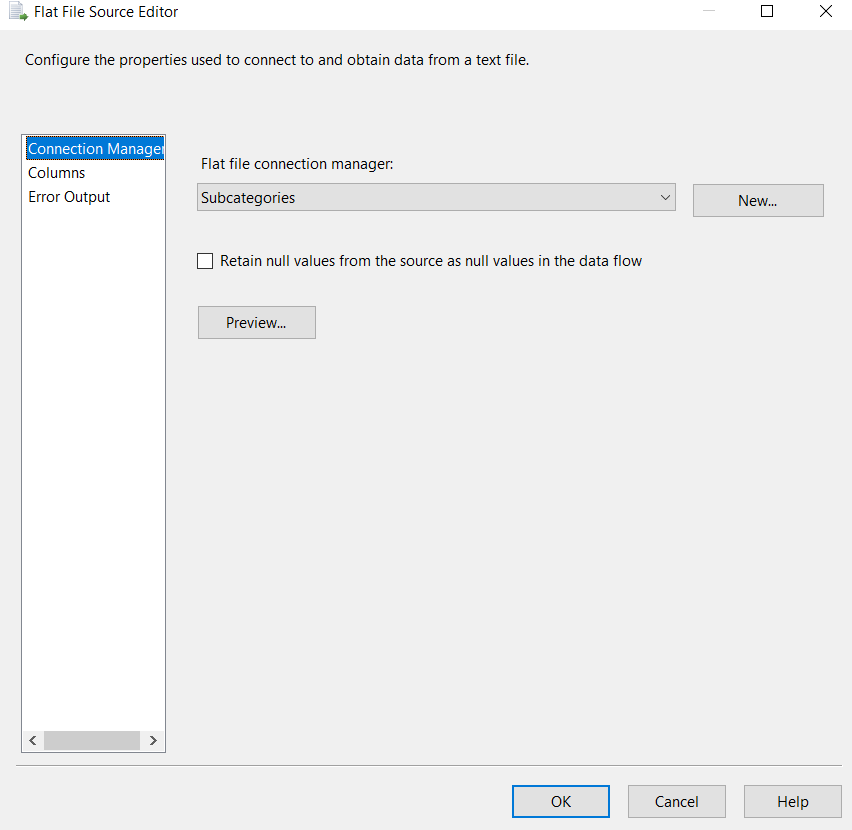
### 3.3.4. SubCategory

#### 3.3.4.1. Load dữ liệu từ Source vào Stage

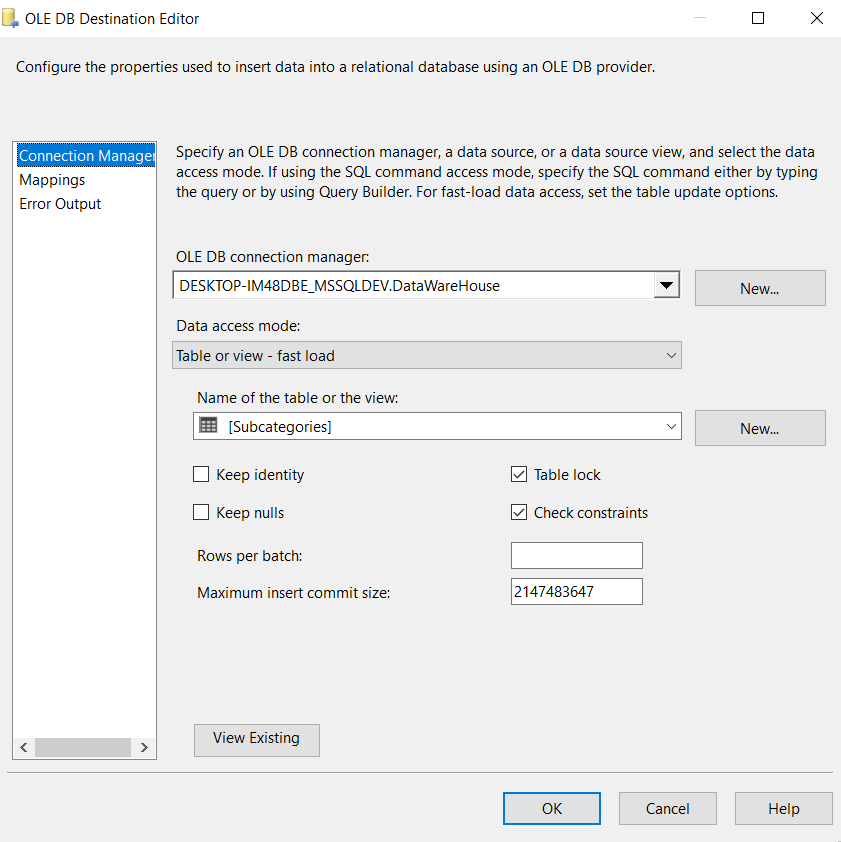
- Double-click **DFT\_LoadSubcategories** task để mở nó trong **Data Flow design surface:**

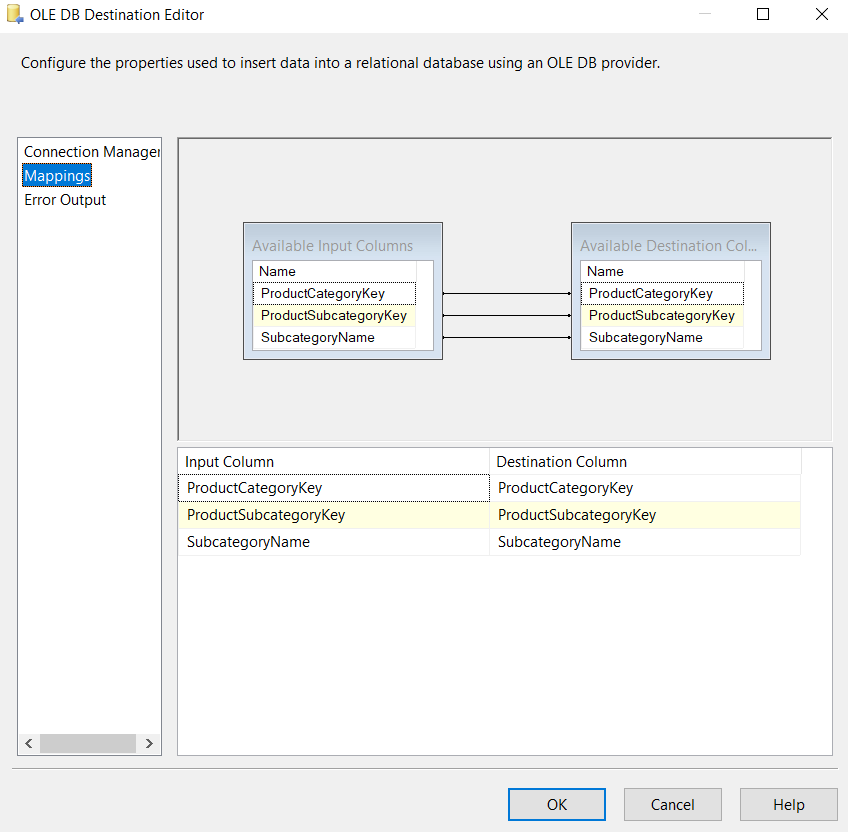
- Tạo **Flat File Source** và **Destination Assistant** và cấu hình chúng như sau:

+ **Flat File Source**:

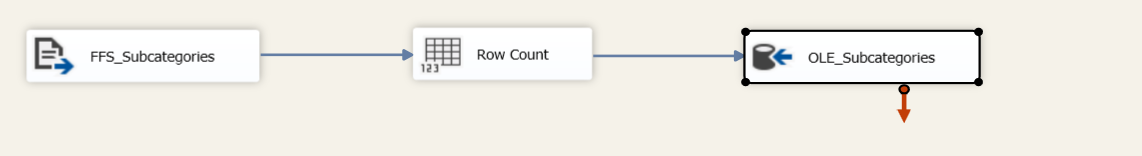


+ Destination Assistant:





+ Sau khi cấu hình xong ta được:



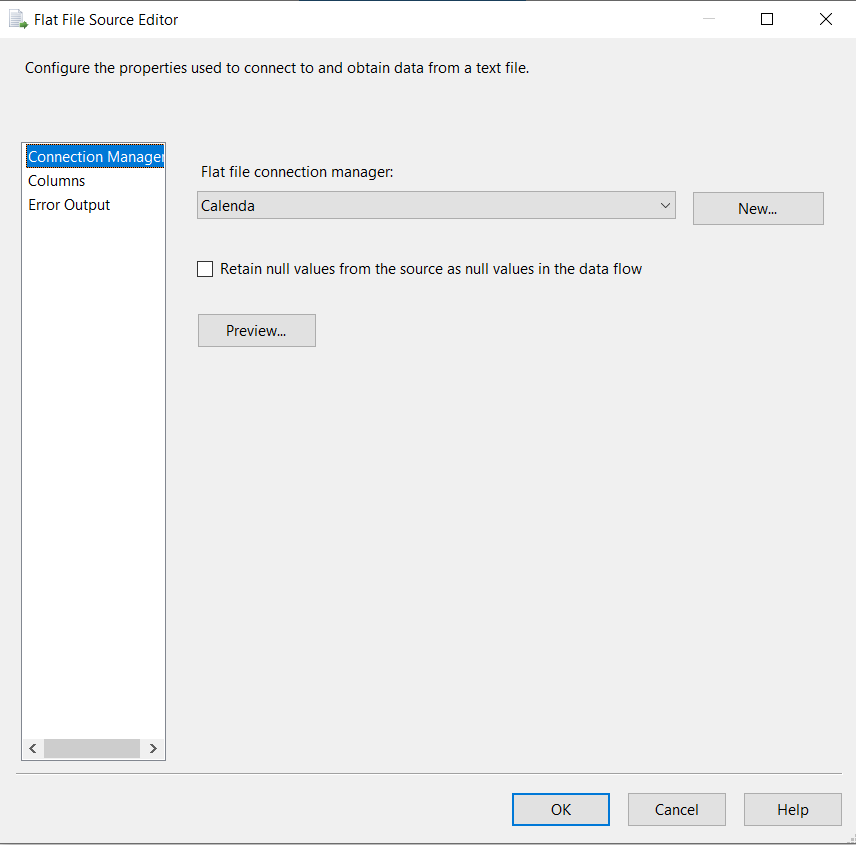
### 3.3.5.Calendar

#### 3.3.5.1. Load dữ liệu từ Source vào Stage

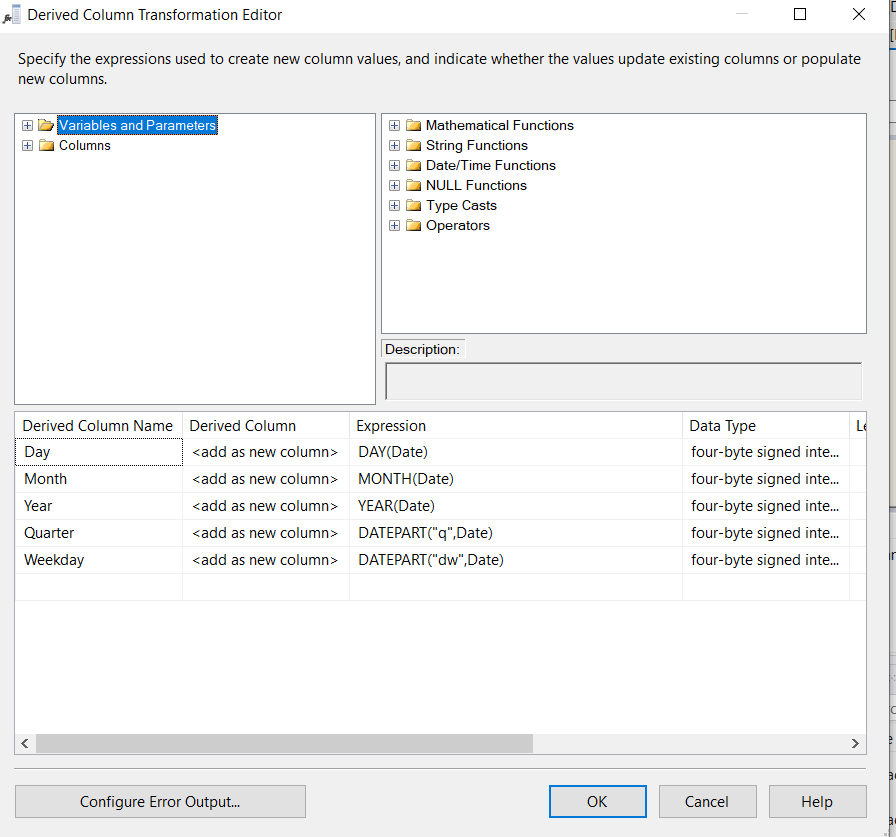
- Double-click **DF - Stage Employee** task để mở nó trong **Data Flow design surface:**

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

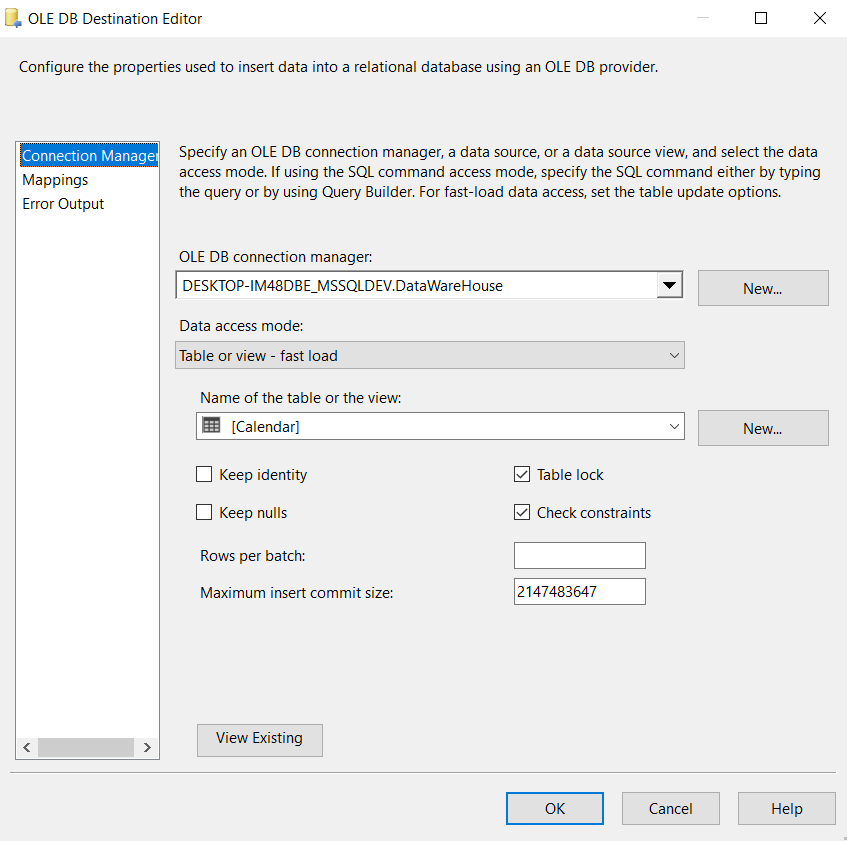
+ **Flat File Source**:

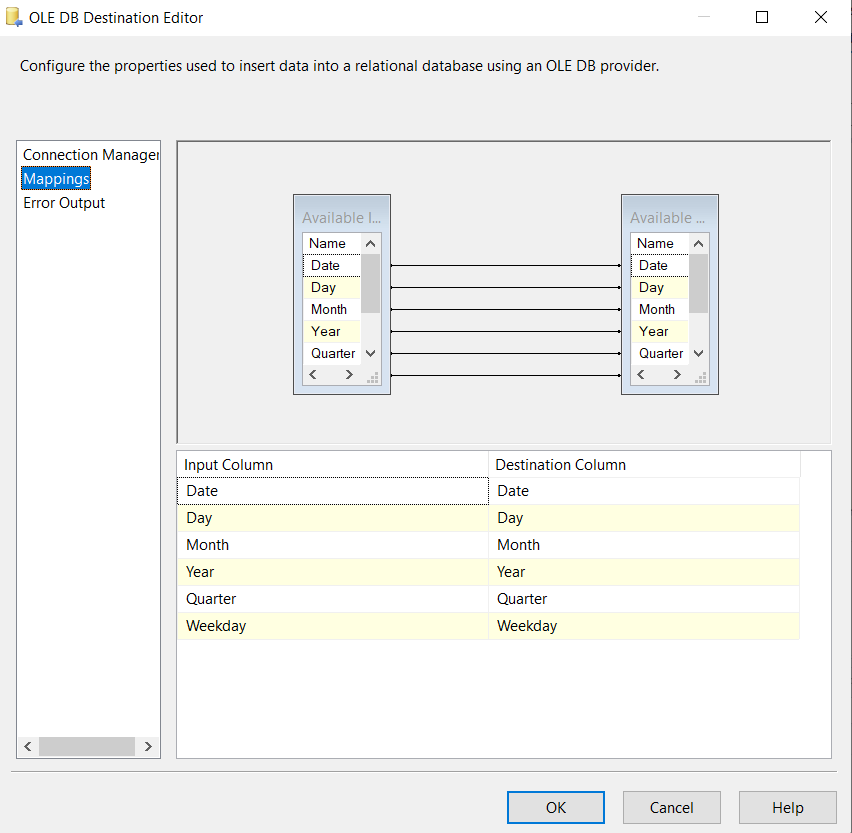


* Derived Column:

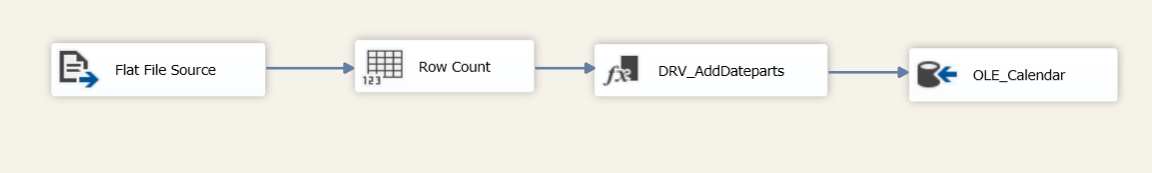


+ Destination Assistant:





+ Sau khi cấu hình xong ta được:



### 3.3.6. Territory

#### 3.3.6.1. Load dữ liệu từ Source vào Stage

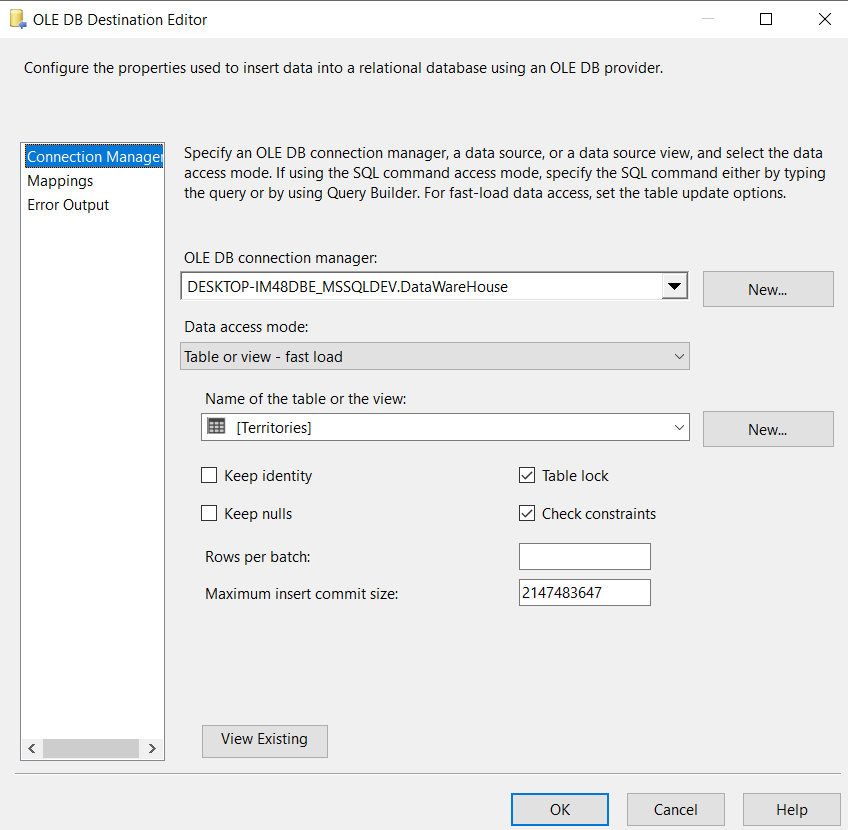
- Double-click **DFT\_LoadTerritories** task để mở nó trong **Data Flow design surface:**

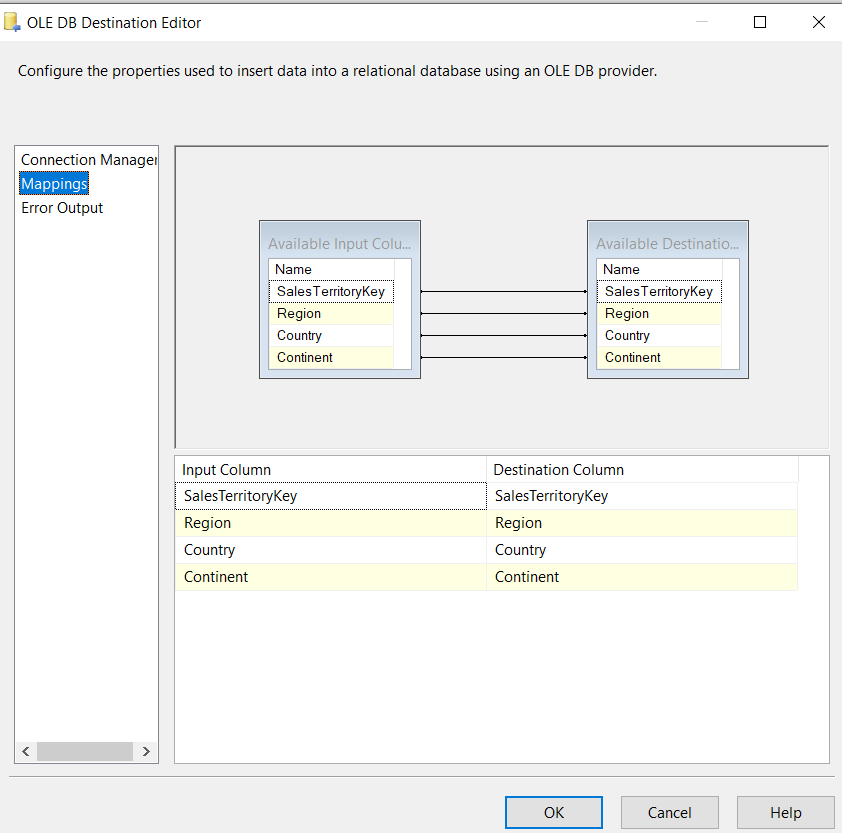
- Tạo **Flat File Source, RowCount** và **Destination Assistant** và cấu hình chúng như sau:

+ **Flat File Source**:

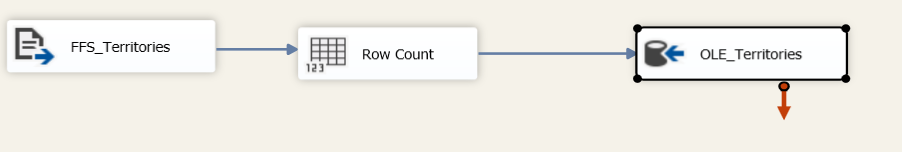


+ Destination Assistant:





+ Sau khi cấu hình xong ta được:



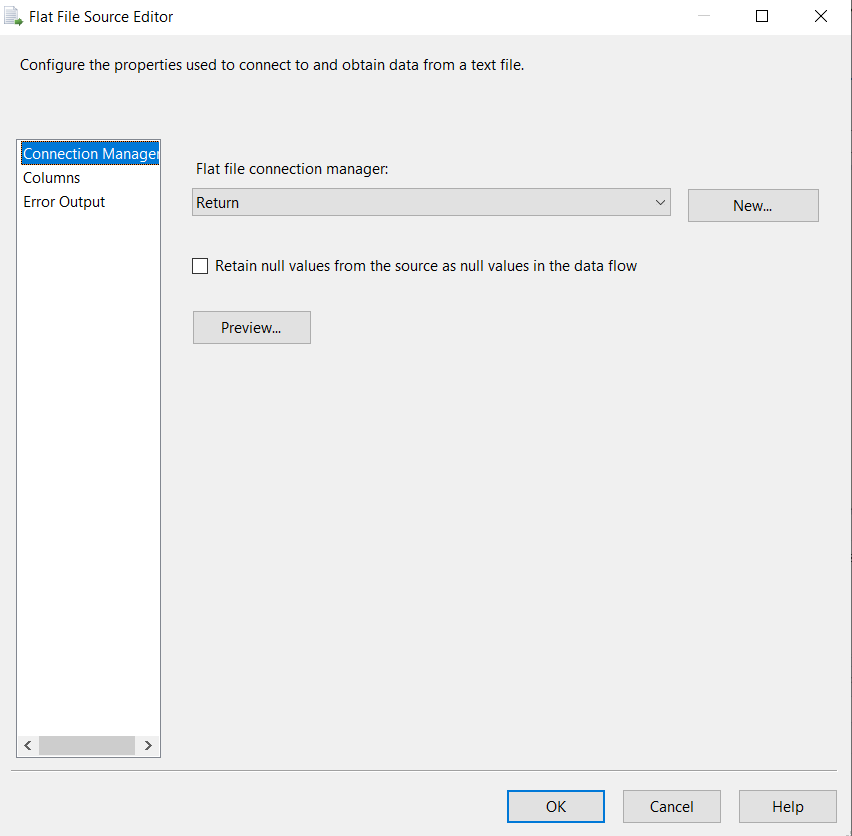
### 3.3.7. Returns

#### 3.3.7.1. Load dữ liệu từ Source vào Stage

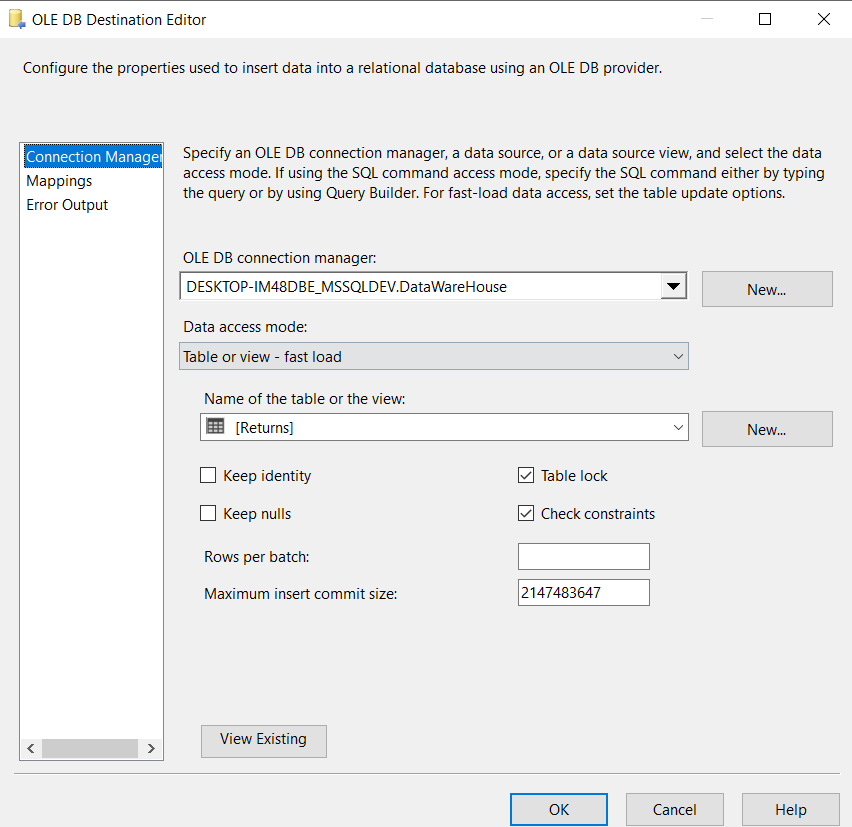
- Double-click **DFT\_LoadReturns** task để mở nó trong **Data Flow design surface:**

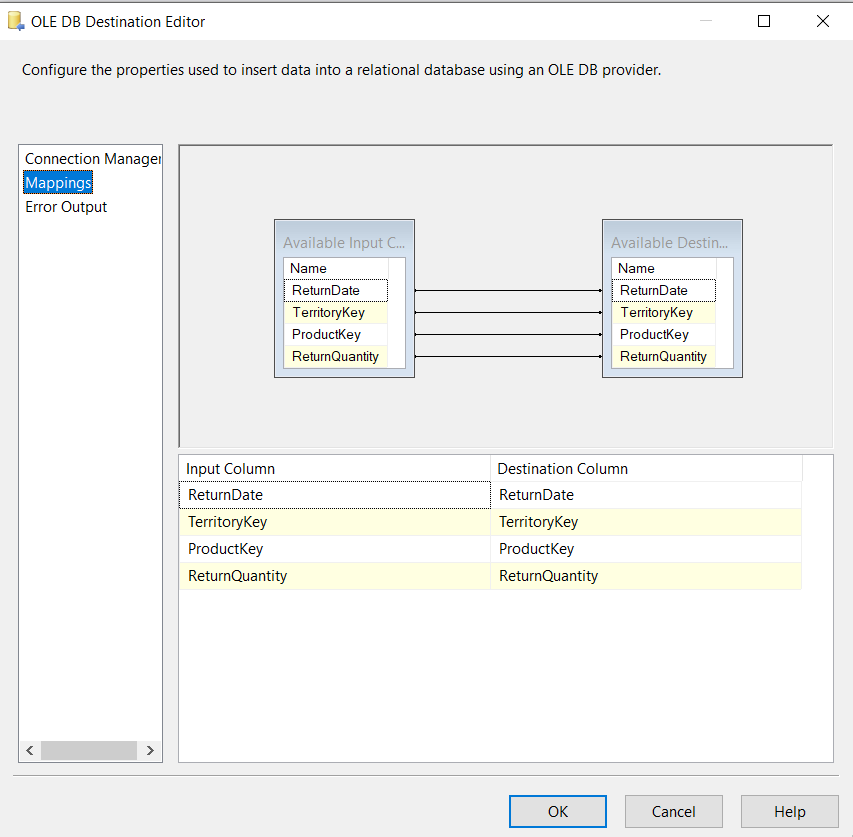
- Tạo **Flat File Source, RowCount** và **Destination Assistant** và cấu hình chúng như sau:

+ **Flat File Source**:



+ Destination Assistant:





+ Sau khi cấu hình xong ta được:

****

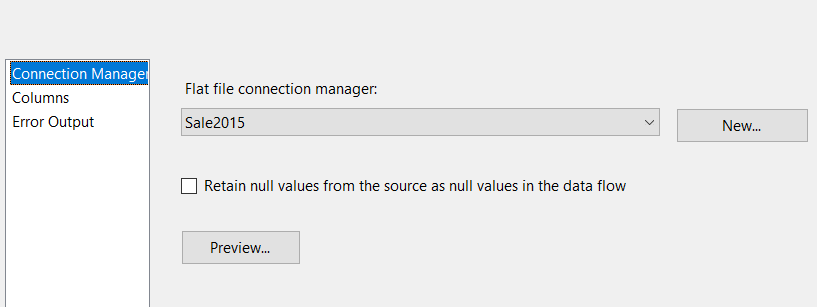
### 3.3.8. Sales

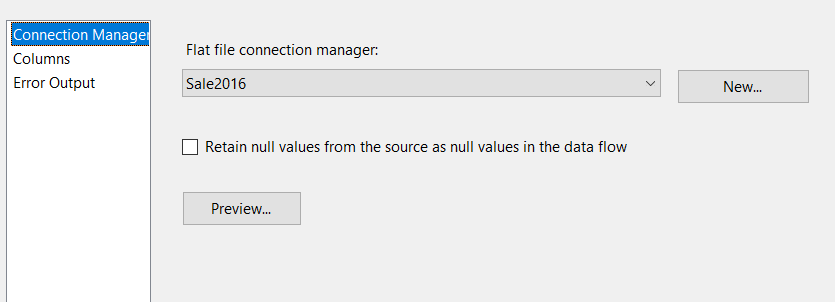
#### 3.3.8.1. Load dữ liệu từ Source vào Stage

- 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:

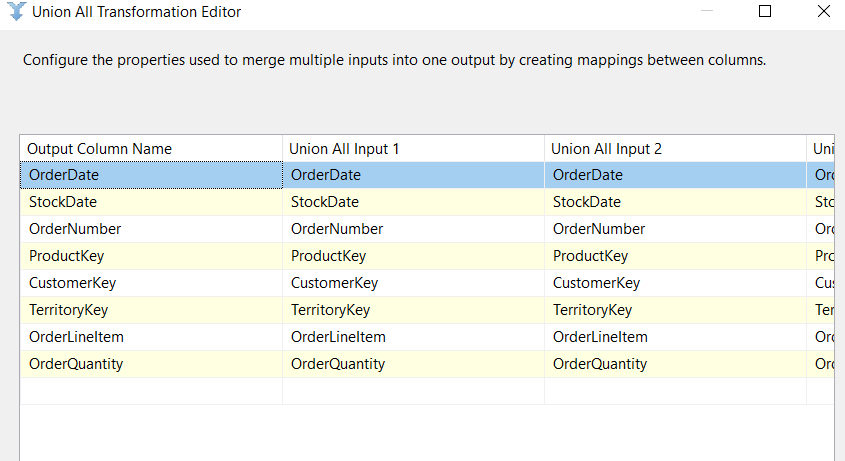
+ **Flat File Source**:

****

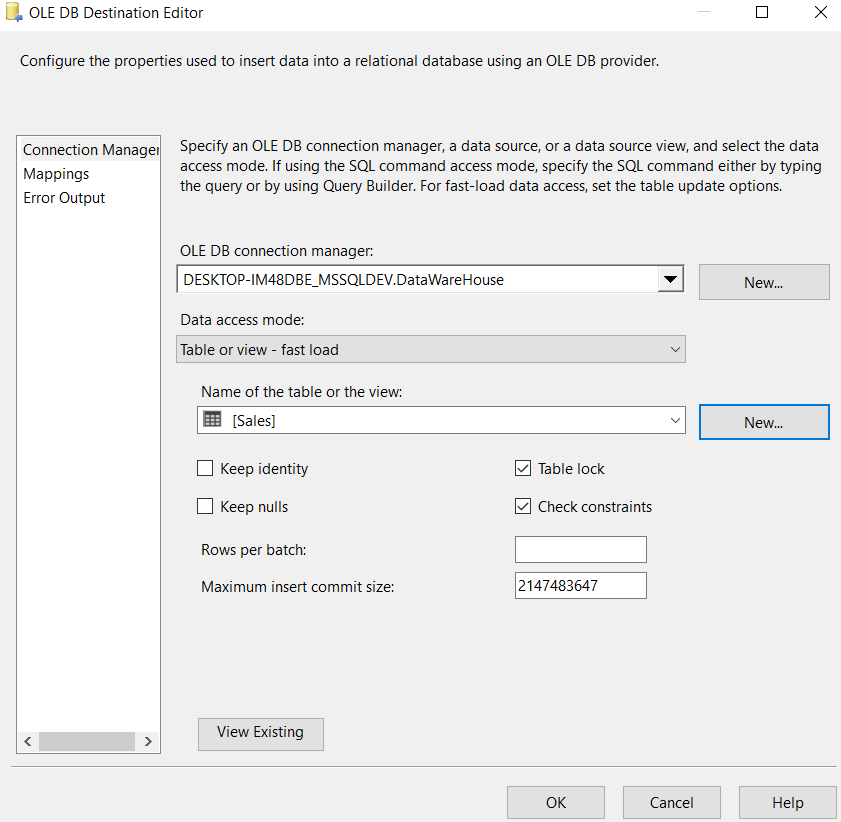
****

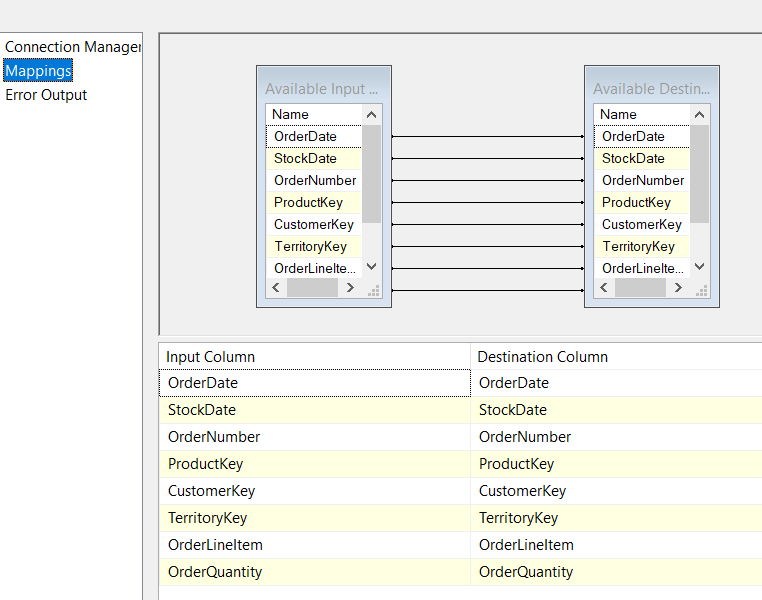
****

* **Union All:**

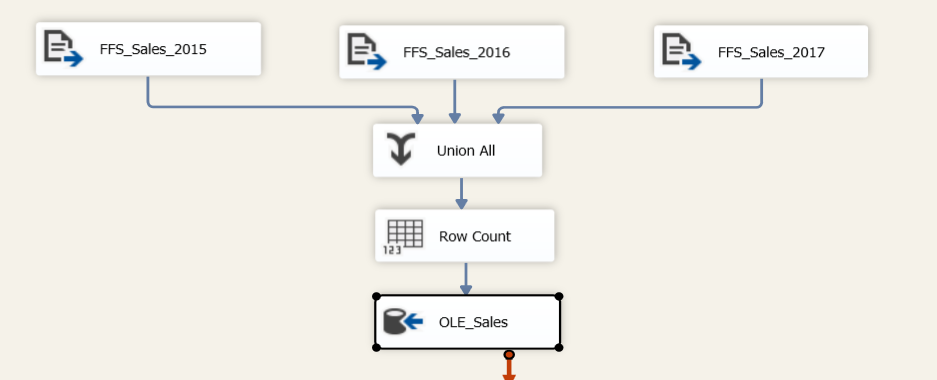
****

* **Destination Assistant:**

****

****

+ Sau khi cấu hình xong ta được:

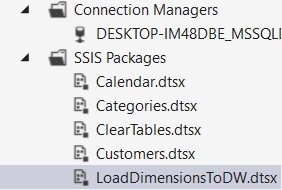


**3.4. Import dữ liệu từ bảng stage vào bảng dimension**

**3.4.1. DimCustomer**

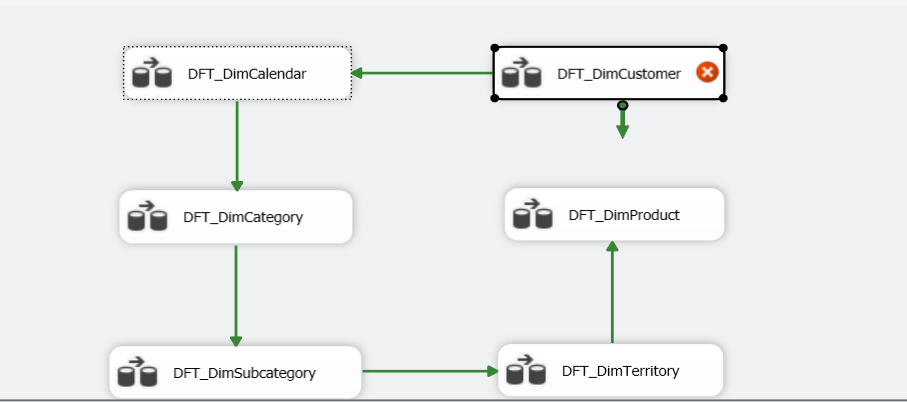
#### 3.4.1.1. Tạo Package

- 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.

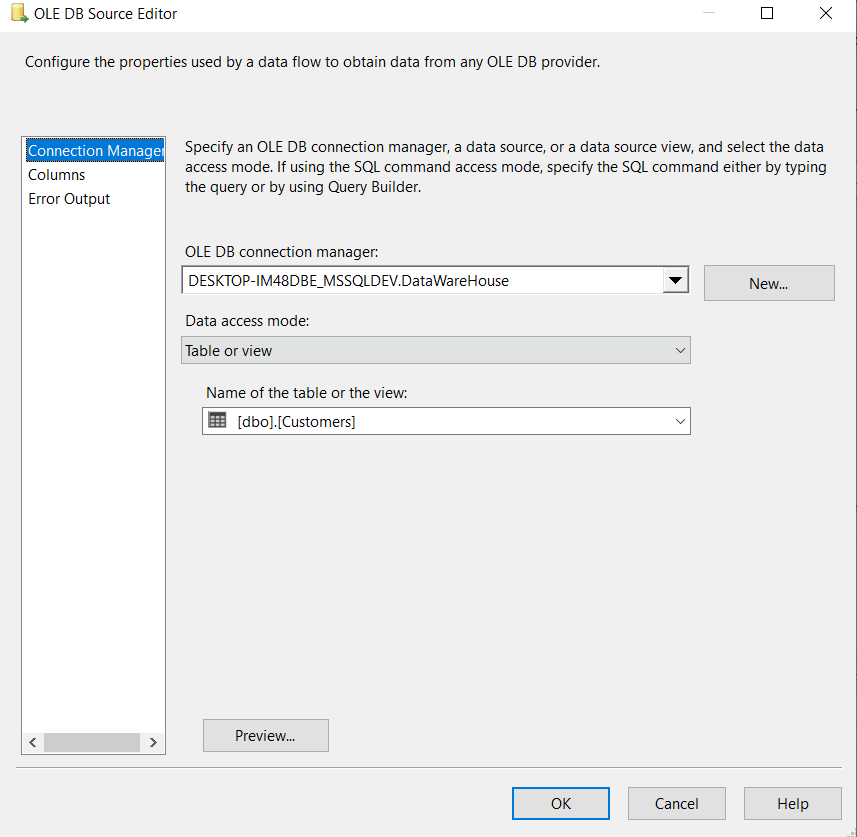
****

#### 3.4.1.3. Load dữ liệu từ Stage vào Dimension

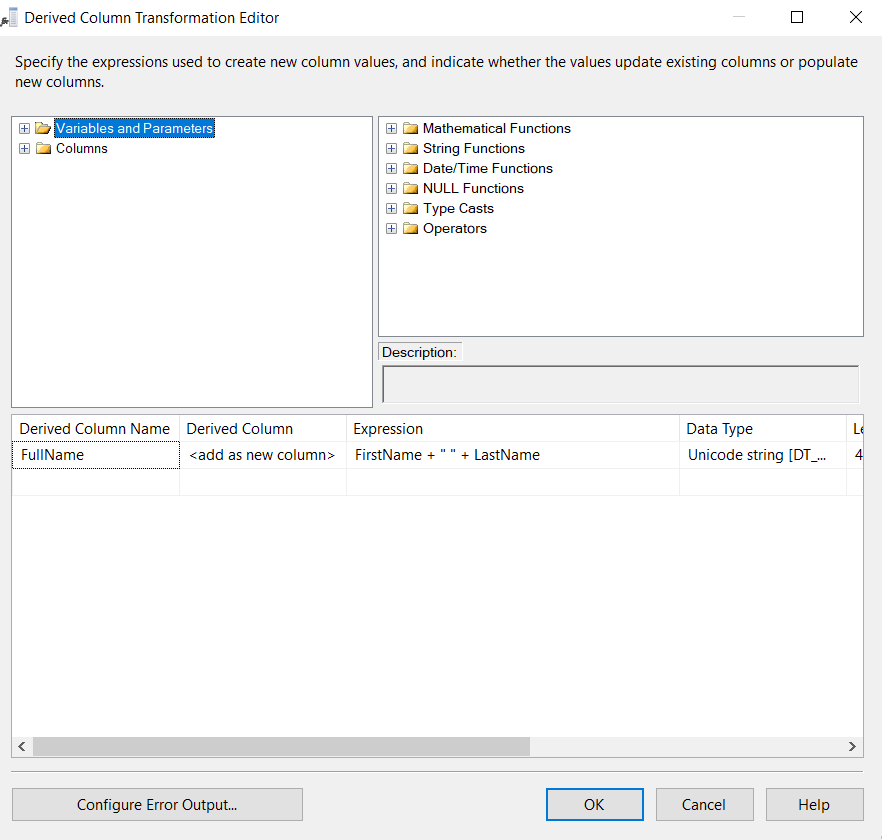
- Double-click **DFT\_DimCustomer** task để mở nó trong **Data Flow design surface:**

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

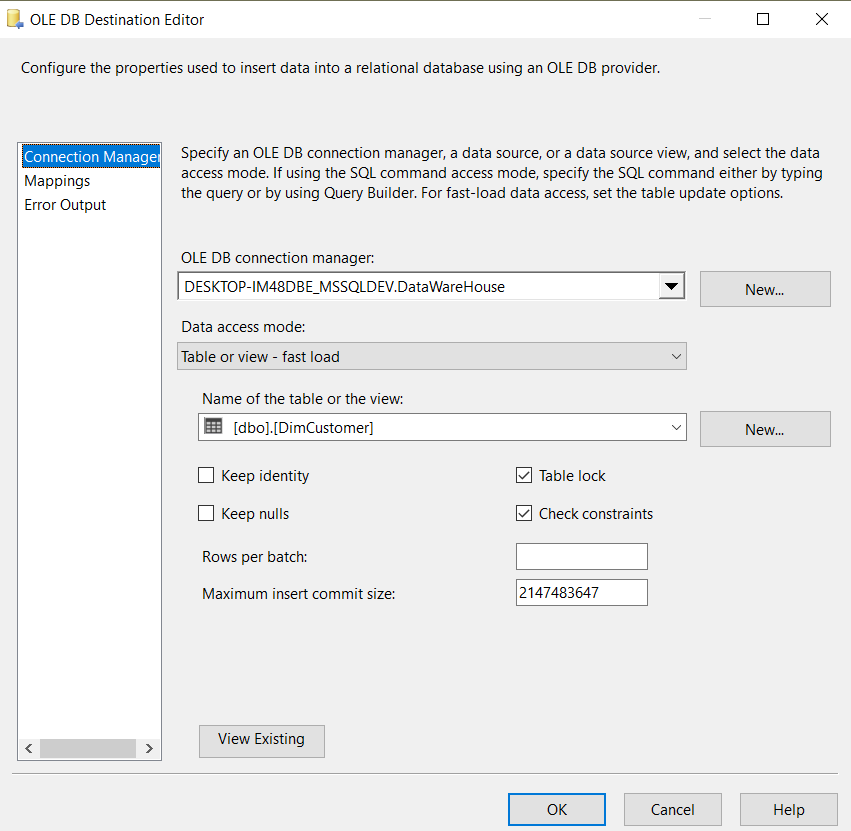
+ **Flat File Source**:

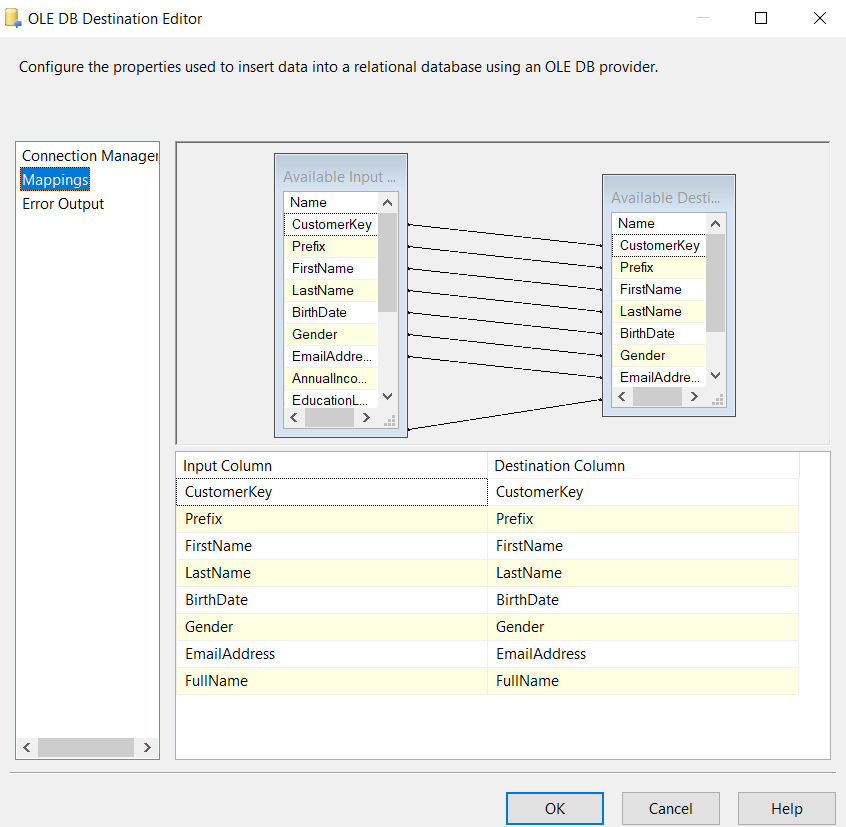


+ Derived Column:



+ Destination Assist





+ Sau khi cấu hình xong ta được:



**3.4.2. DimCalendar**

#### 3.4.2.1. Load dữ liệu từ Stage vào Dimension

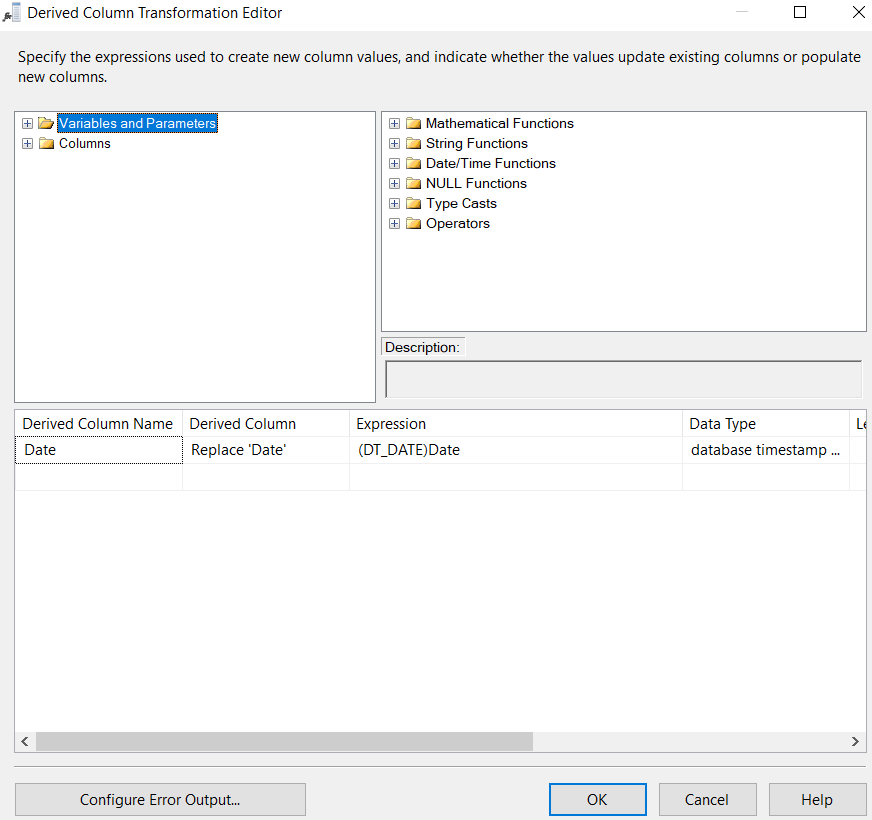
- Double-click **DFT\_DimCalendar** task để mở nó trong **Data Flow design surface:**

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

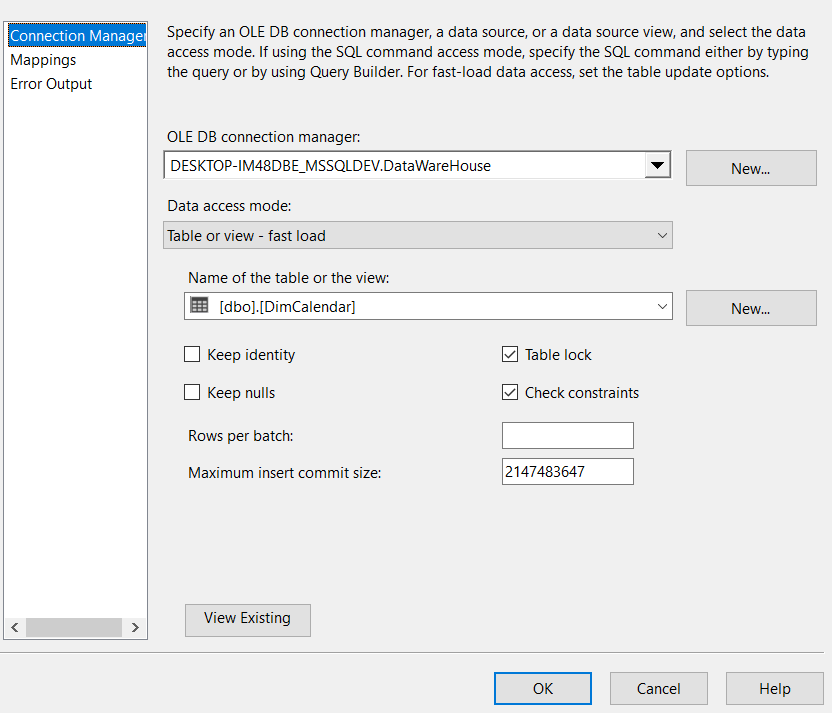
+ Source Assistant:

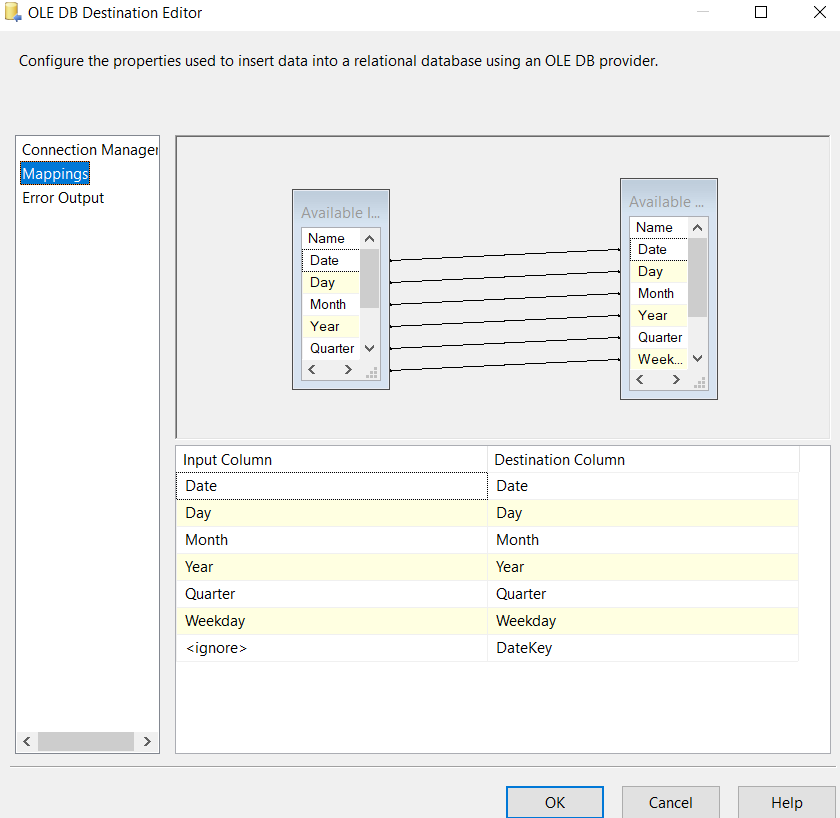


+ Derived Column:



+ Destination Assist





+ Sau khi cấu hình xong ta được:



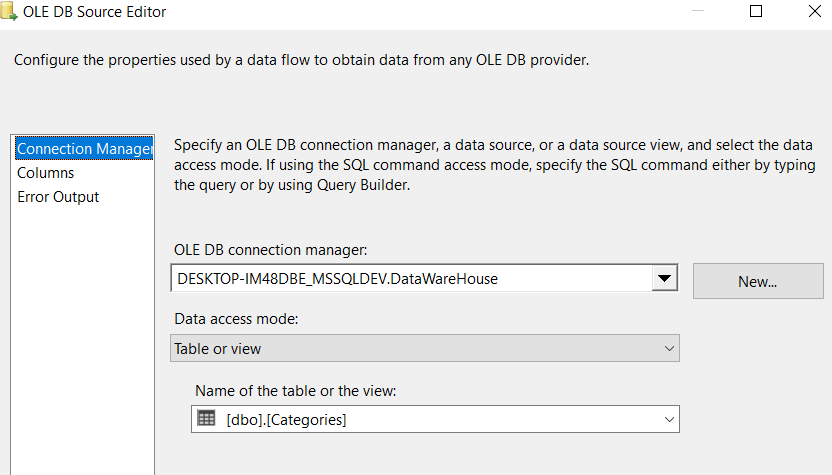
**3.4.3. DimCategory**

#### 3.4.3.1. Load dữ liệu từ Stage vào Dimension

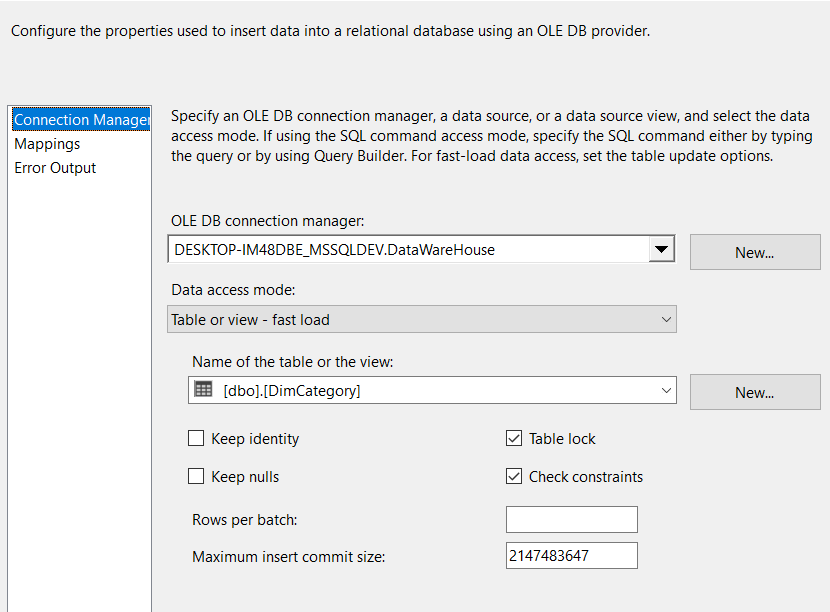
- Double-click **DFT\_DimCustomer** task để mở nó trong **Data Flow design surface:**

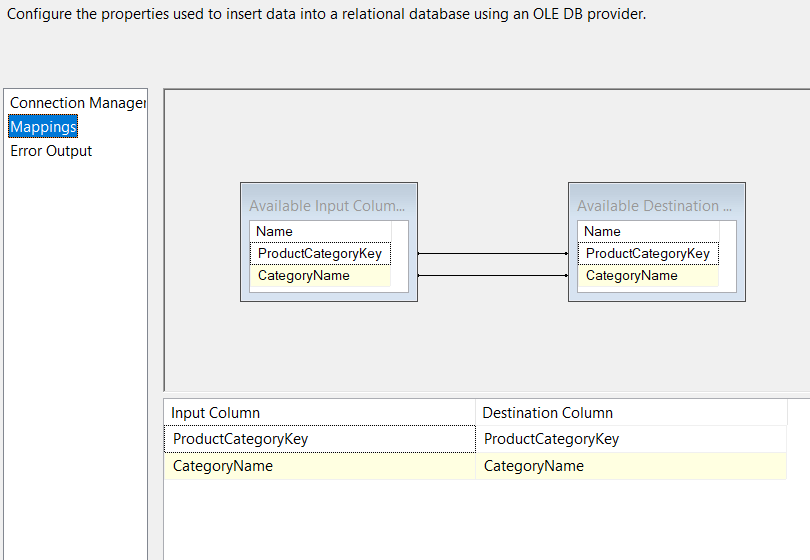
- Tạo **Flat File Source** và **Destination Assistant** và cấu hình chúng như sau:

+**Flat File Source**:



+ Destination Assist





+ Sau khi cấu hình xong ta được:

****

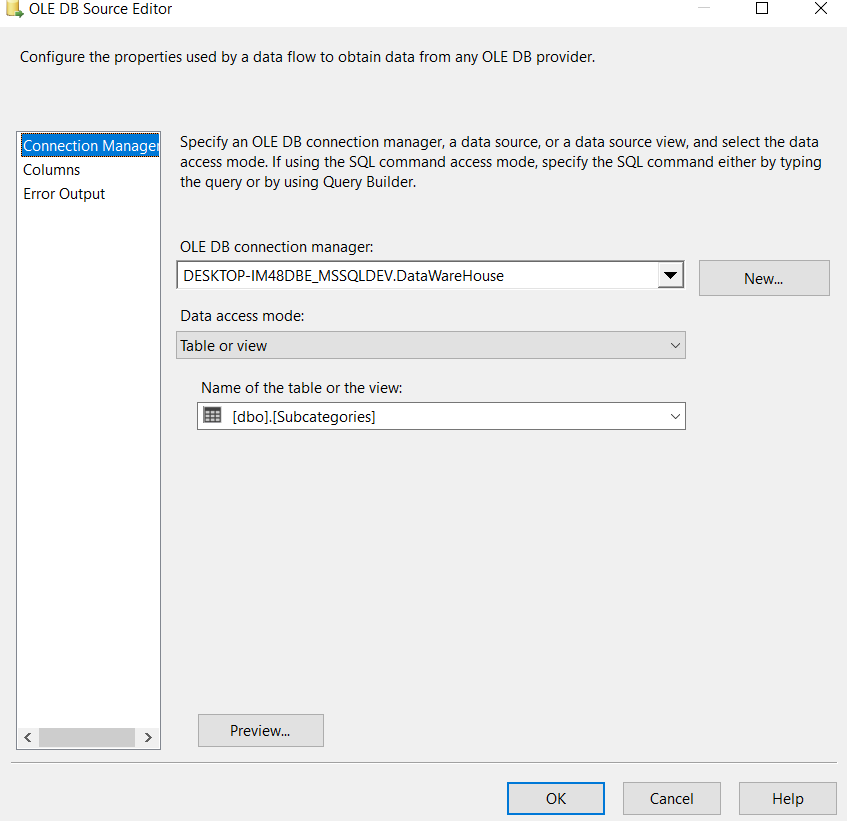
**3.4.4. DimSubCategory**

#### 3.4.4.1. Load dữ liệu từ Stage vào Dimension

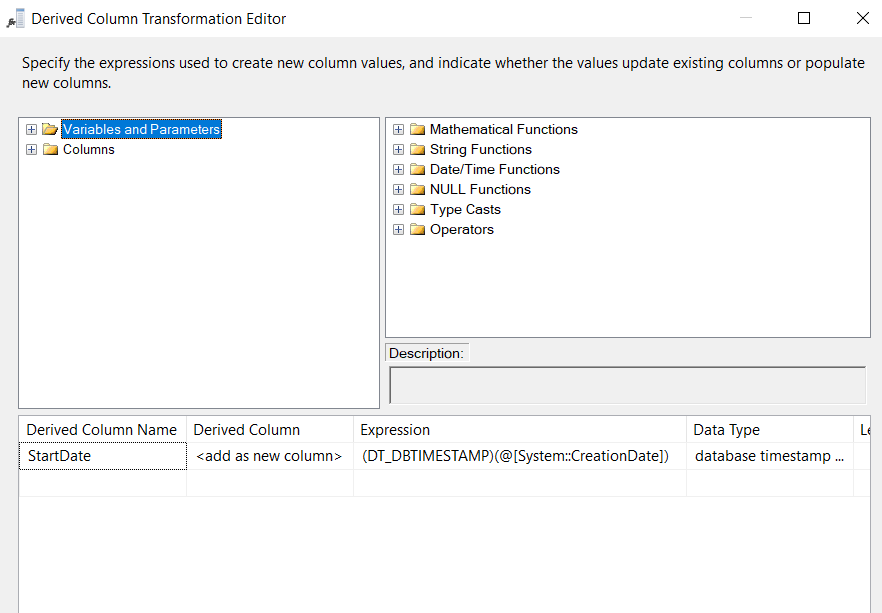
- Double-click **DFT\_DimCustomer** task để mở nó trong **Data Flow design surface:**

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

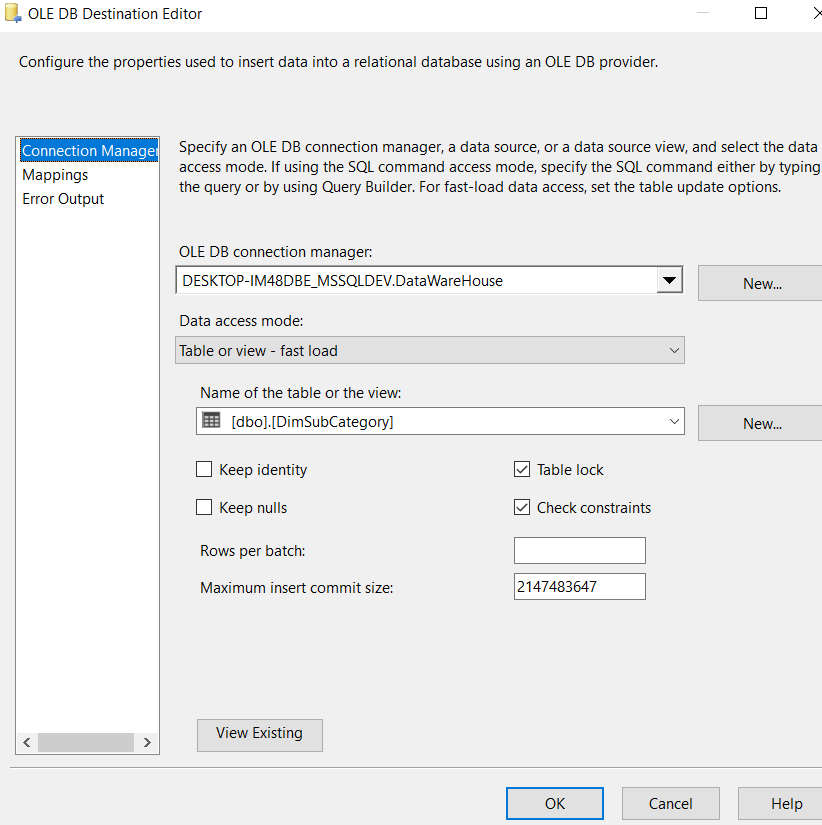
+ **Flat File Source**:

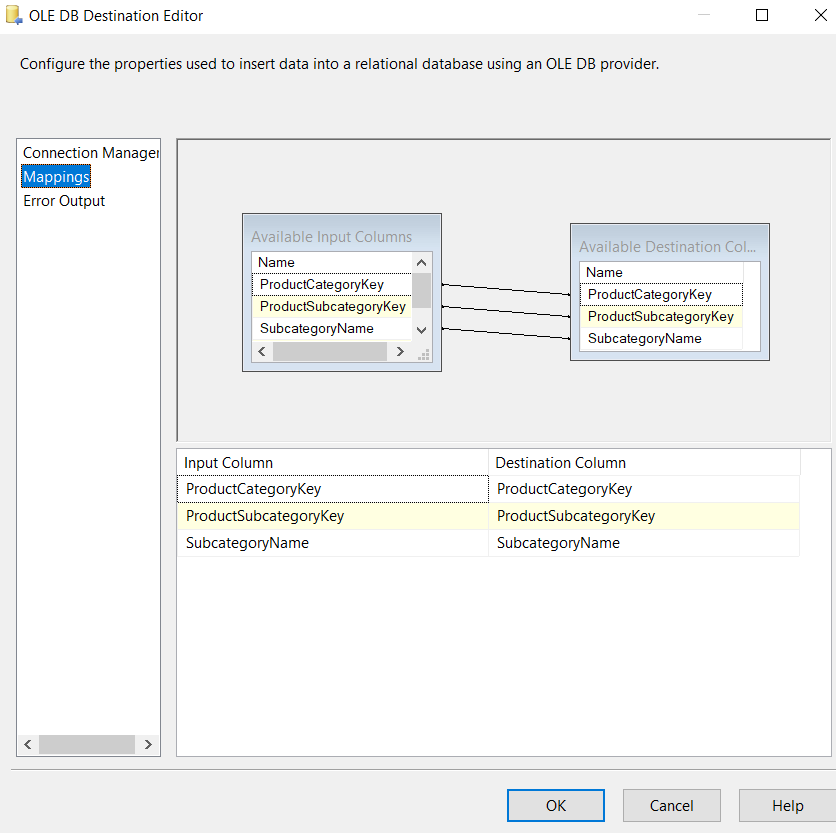


+ Derived Column:

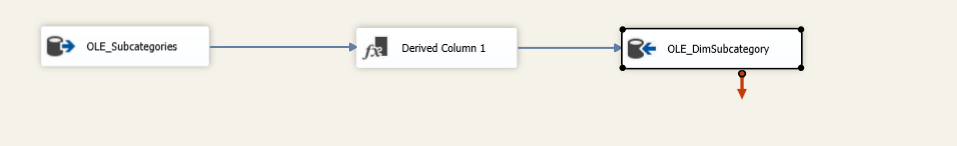


+ Destination Assist





+ Sau khi cấu hình xong ta được:

****

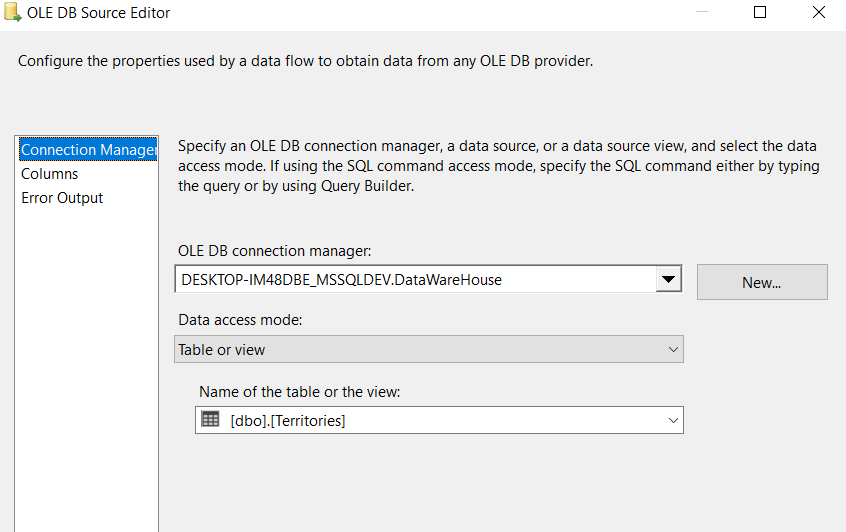
**3.4.5. DimTerritory**

#### 3.4.5.1. Load dữ liệu từ Stage vào Dimension

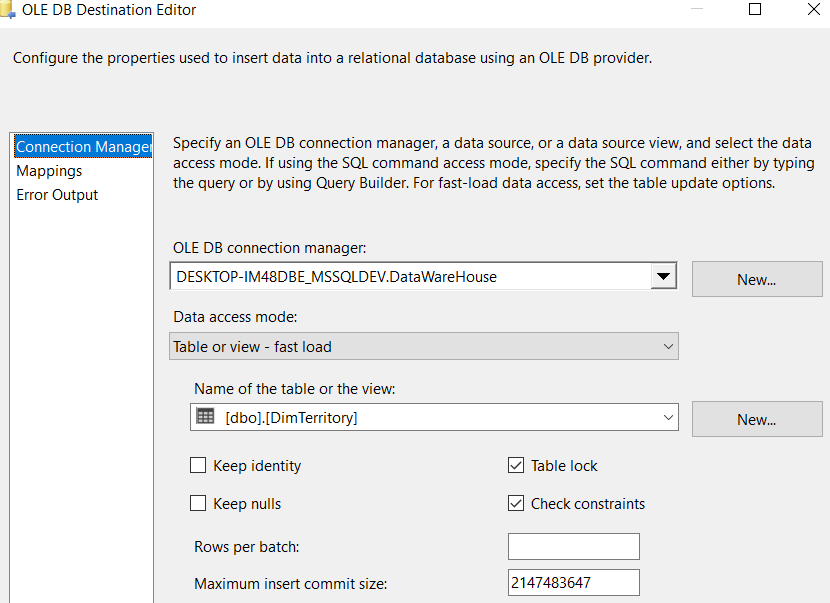
- Double-click **DFT\_DimTerritory** task để mở nó trong **Data Flow design surface:**

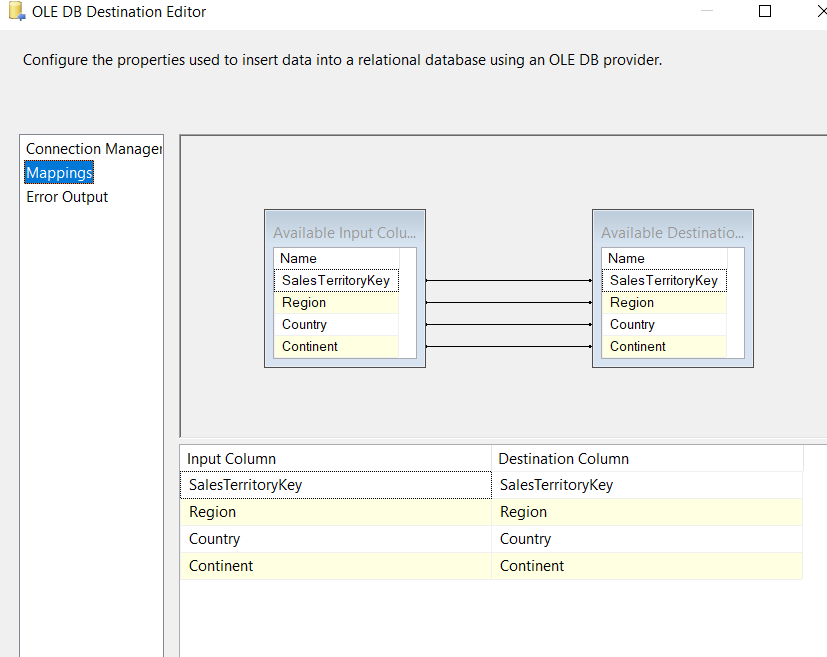
- Tạo **Flat File Source, Derived Column** và **Destination Assistant** và cấu hình chúng như sau:

+ **Flat File Source**:

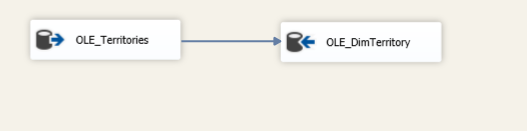


+ Destination Assist





+ Sau khi cấu hình xong ta được:

****

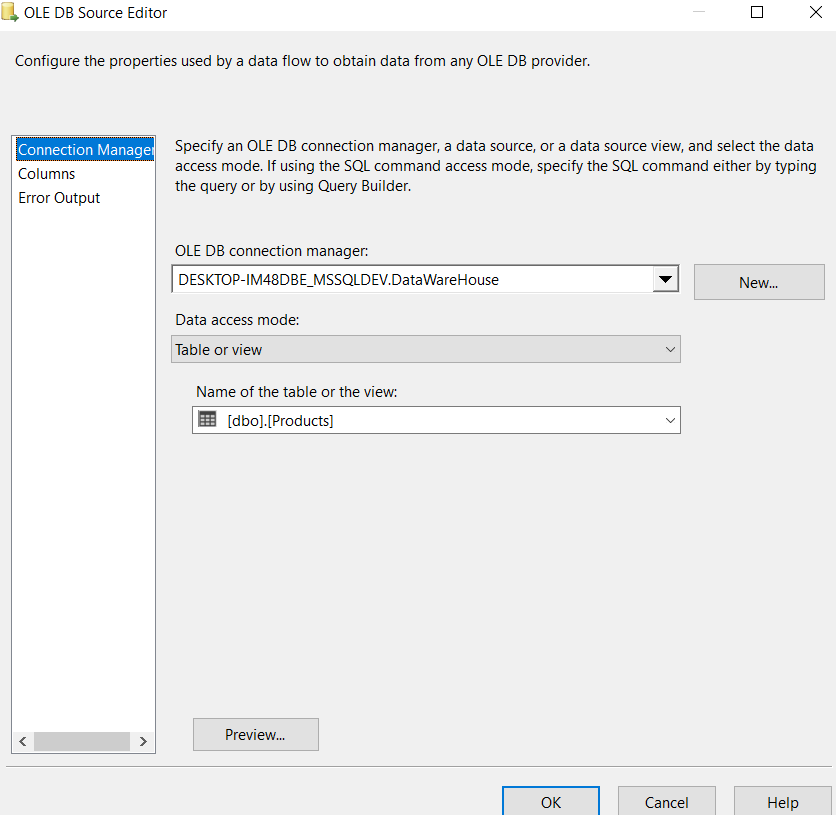
**3.4.6. DimProduct**

#### 3.4.6.1. Load dữ liệu từ Stage vào Dimension

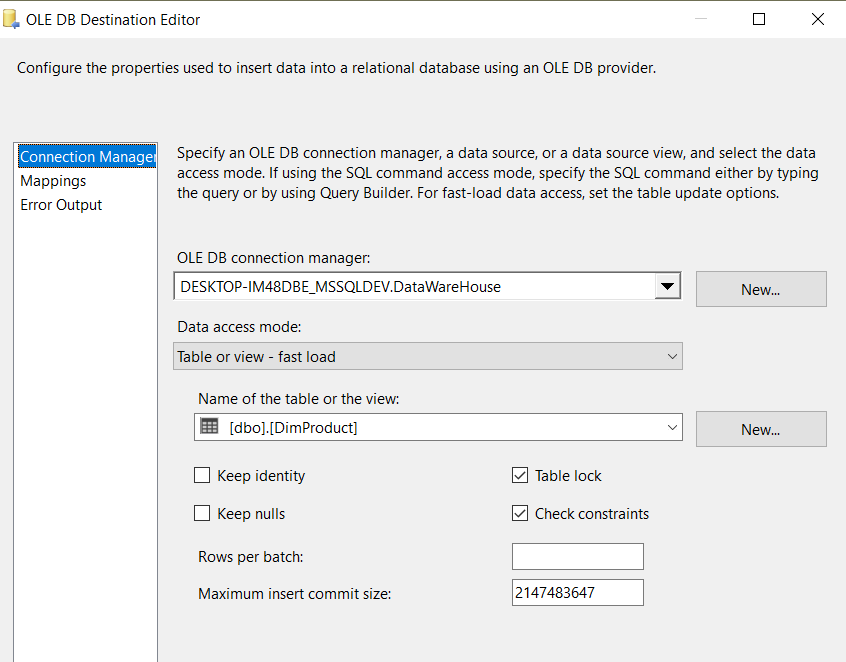
- Double-click **DFT\_DimProduct** task để mở nó trong **Data Flow design surface:**

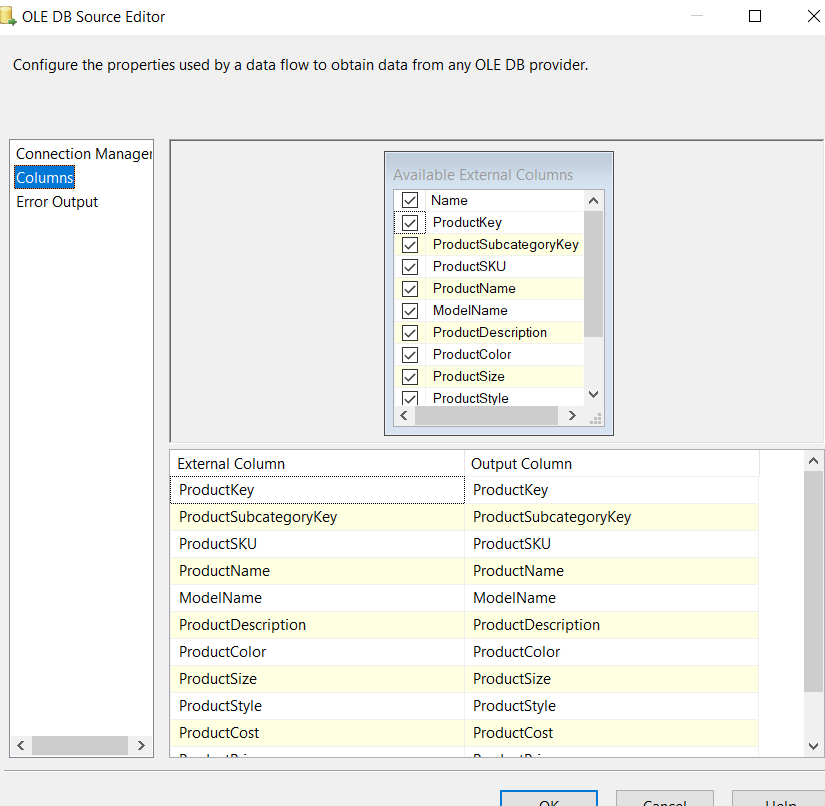
- Tạo **Flat File Sourcet, Derived Column** và **Destination Assistant** và cấu hình chúng như sau:

+ **Flat File Source**:

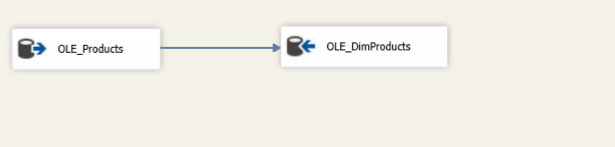


+ Destination Assist





+ Sau khi cấu hình xong ta được:

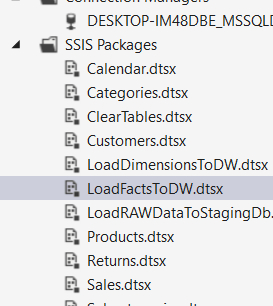


## 3.5. Import dữ liệu vào bảng các bảng fact

### 3.5.1. Fact Sales

#### 3.5.1.1. Tạo Package

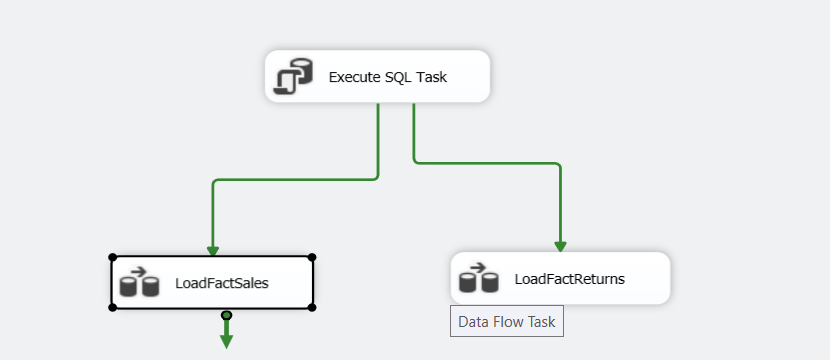
- Tạo package mới và đổi tên như hình bên dưới.



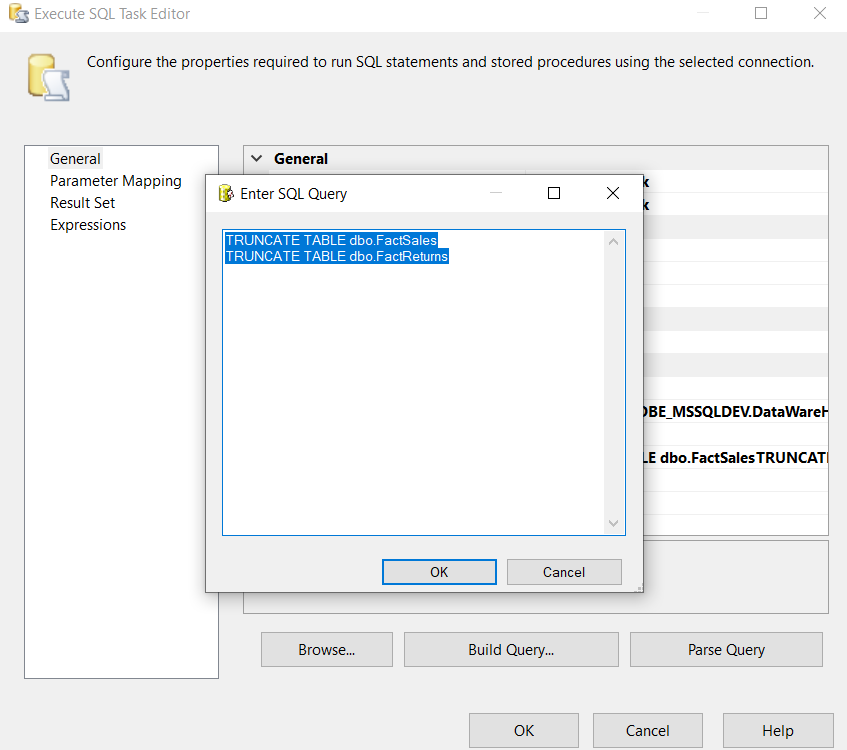
#### 

#### 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.

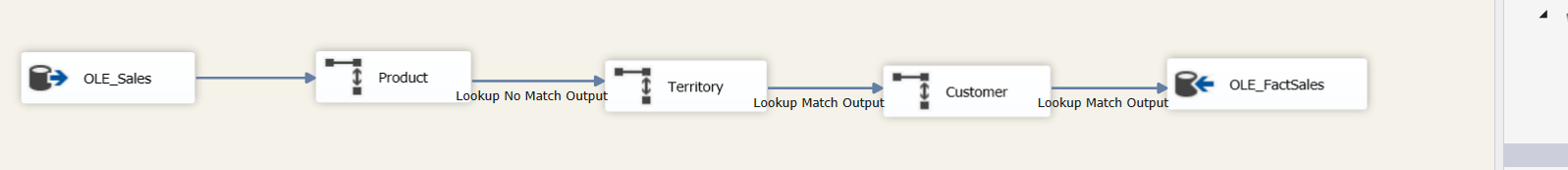


- Thiết lập truncate table SQL task

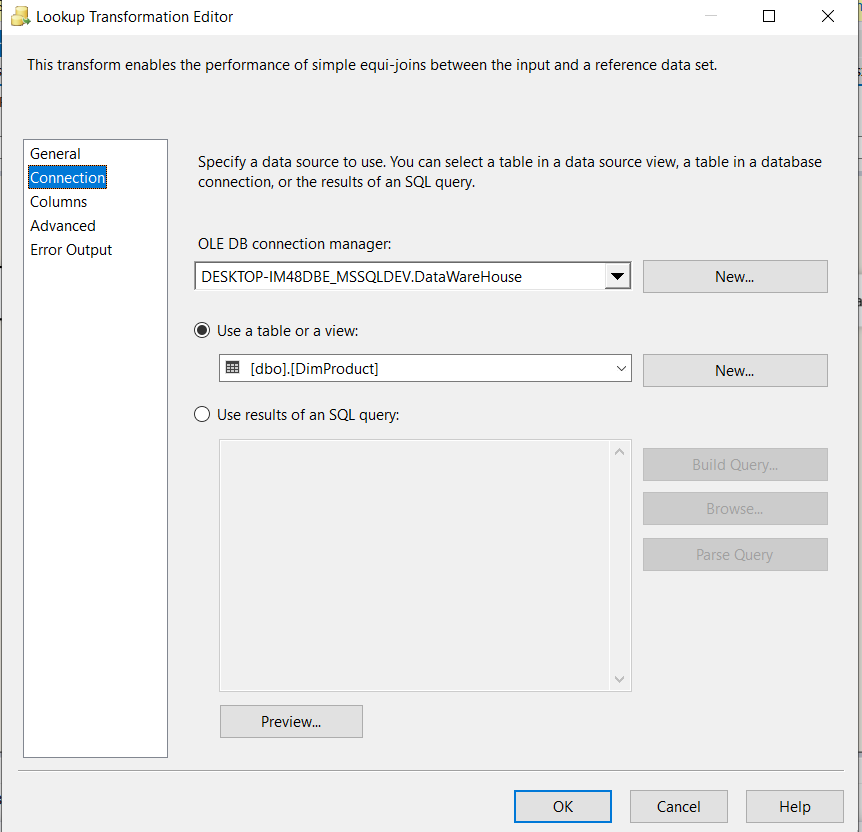


#### 3.5.1.3. Load dữ liệu từ Stage vào Fact

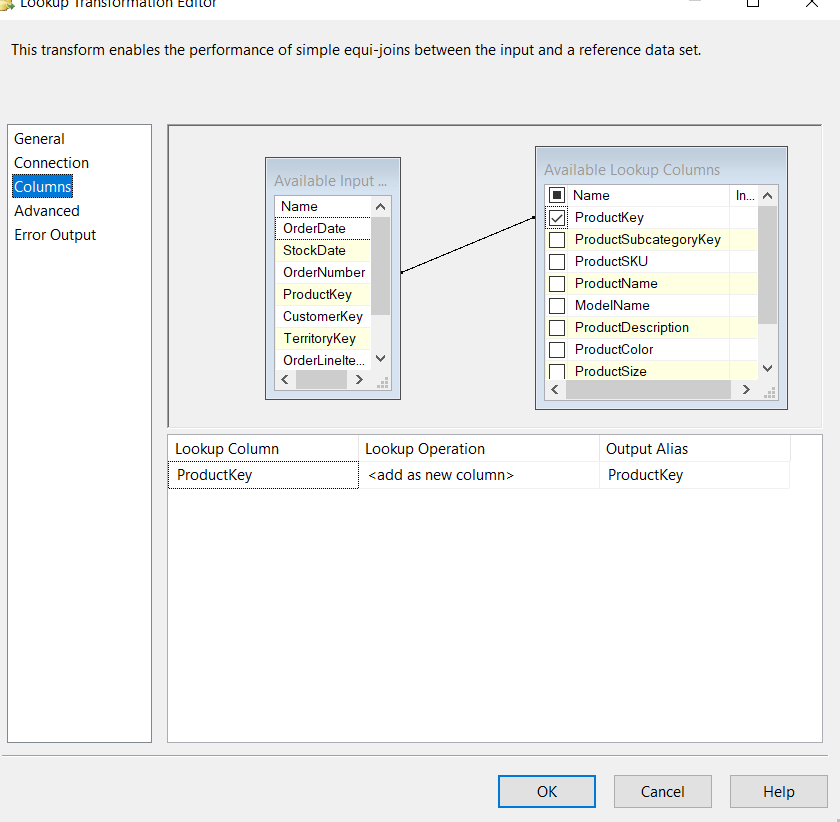
- Double-click vào **LoadFactSales**task để mở data flow design Surface và thiết lập các cấu hình sau



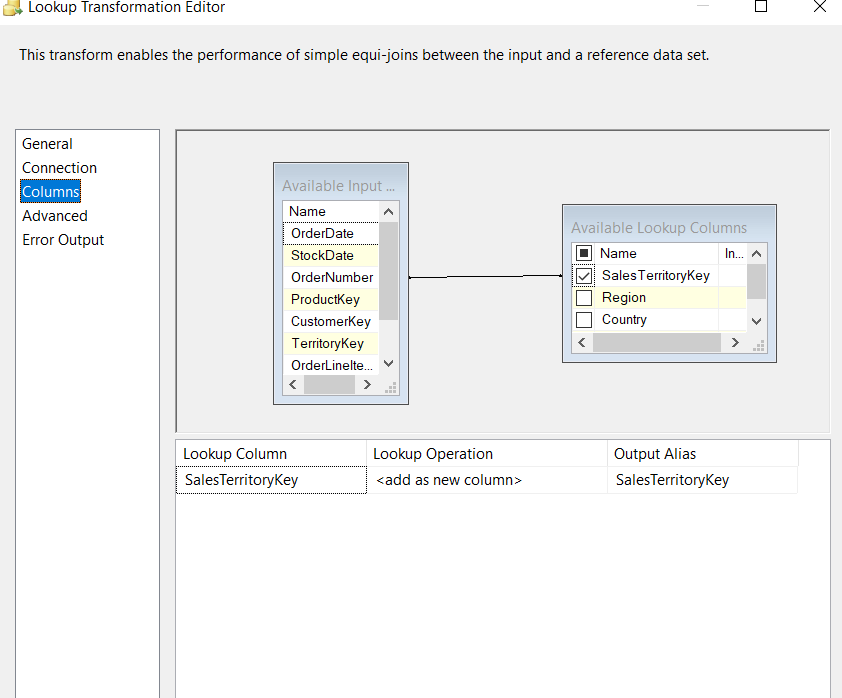
- **Lookup** qua bảng **DimProduct**



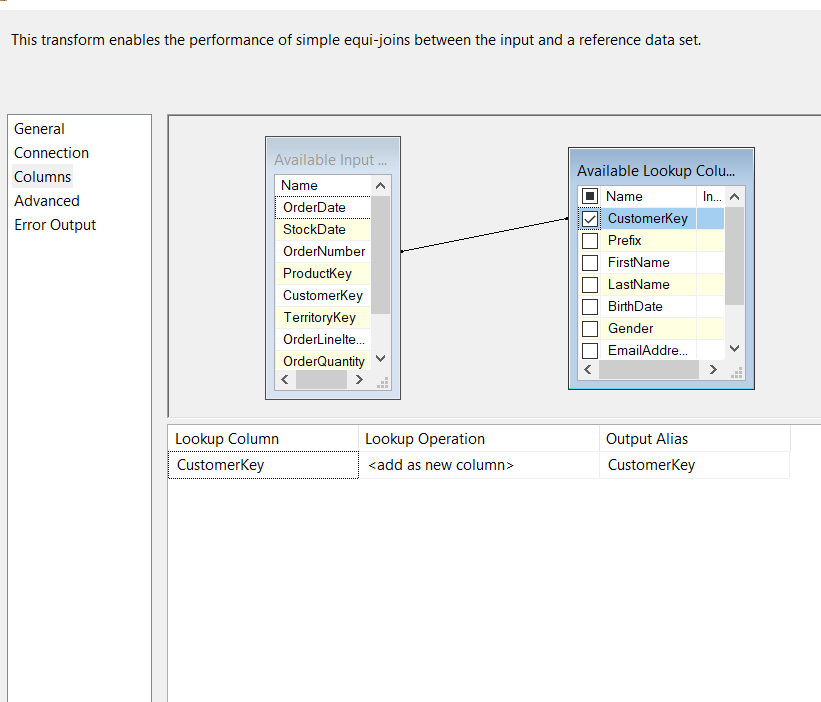
- Thiết lập **mapping** và lấy ra **ProductKey**



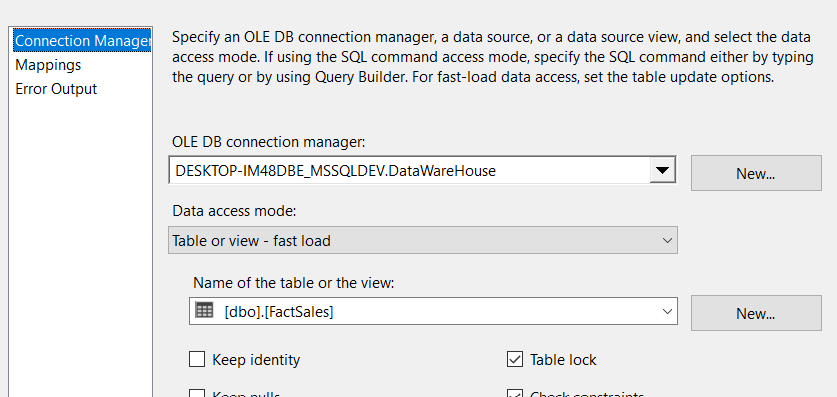
- **Lookup** qua bảng **DimTerritory**



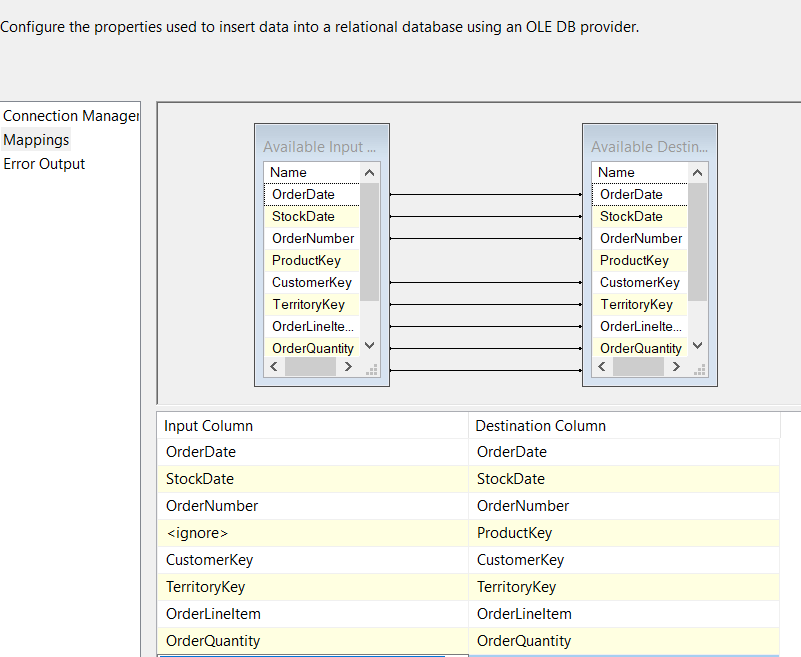
- **Lookup** qua bảng **DimCustomer**



- Cấu hình Destination chọn bảng **FactSales**



- Mapping các cột



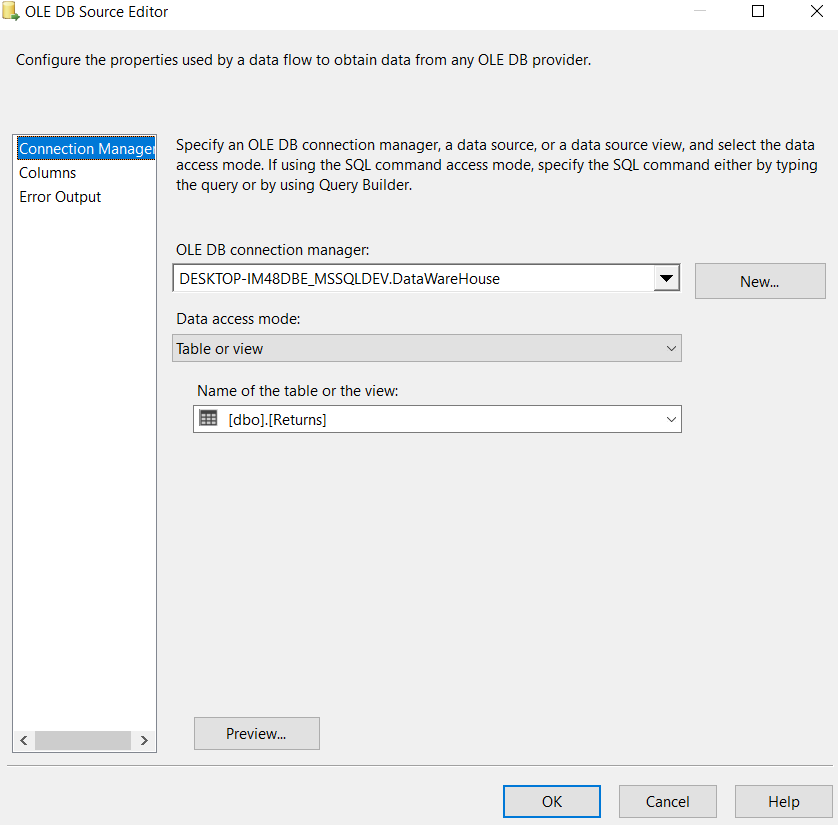
### 3.5.2. Fact Returns

#### 3.5.2.1. Load dữ liệu từ Stage vào Fact

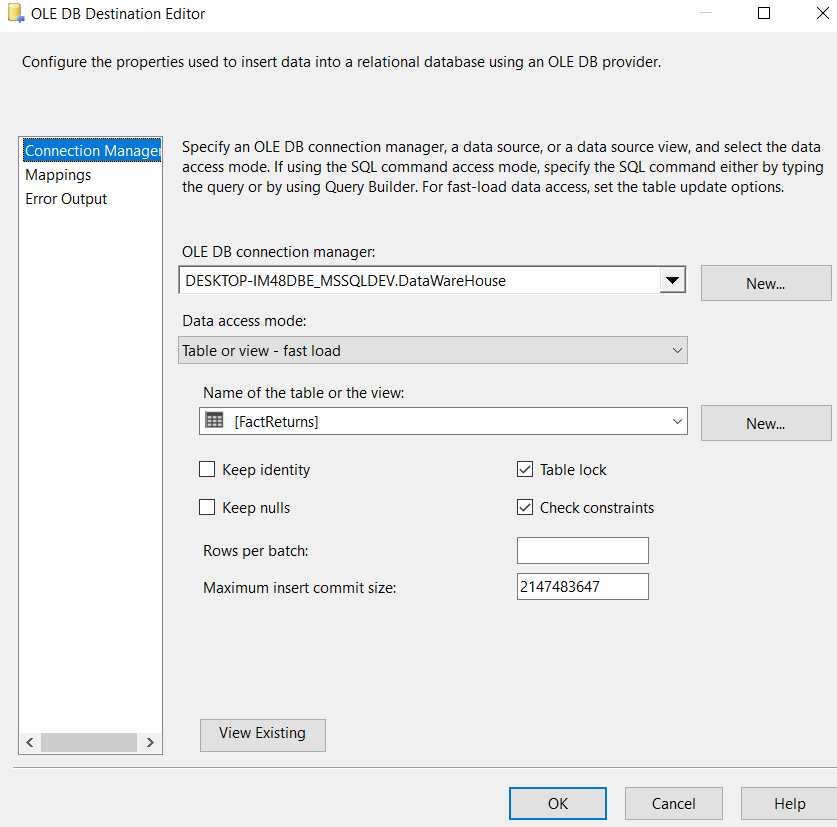
- Double-click vào **LoadFactReturns** task để mở data flow design Surface và thiết lập các cấu hình sau



* Source Assistant:



* Destination Assistant:

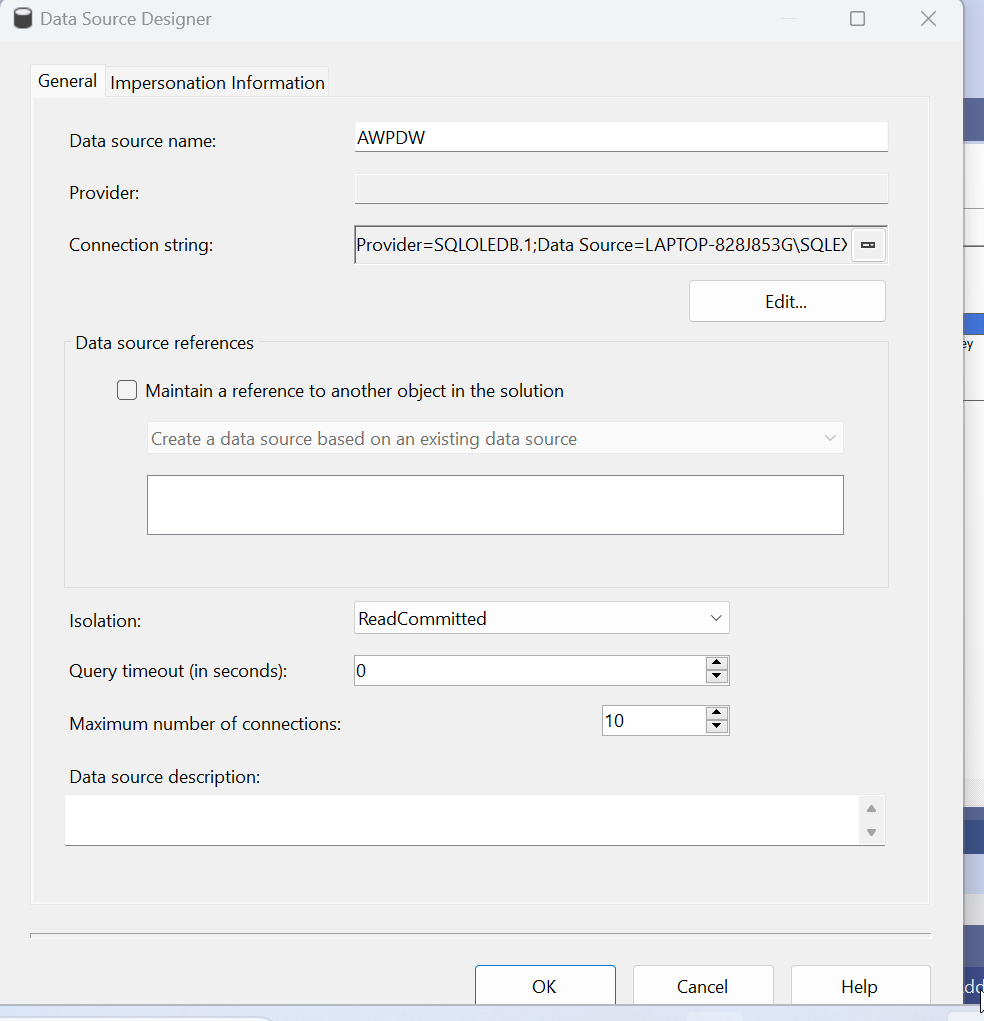


# CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)

## 4.1. Quá trình xây dựng mô hình

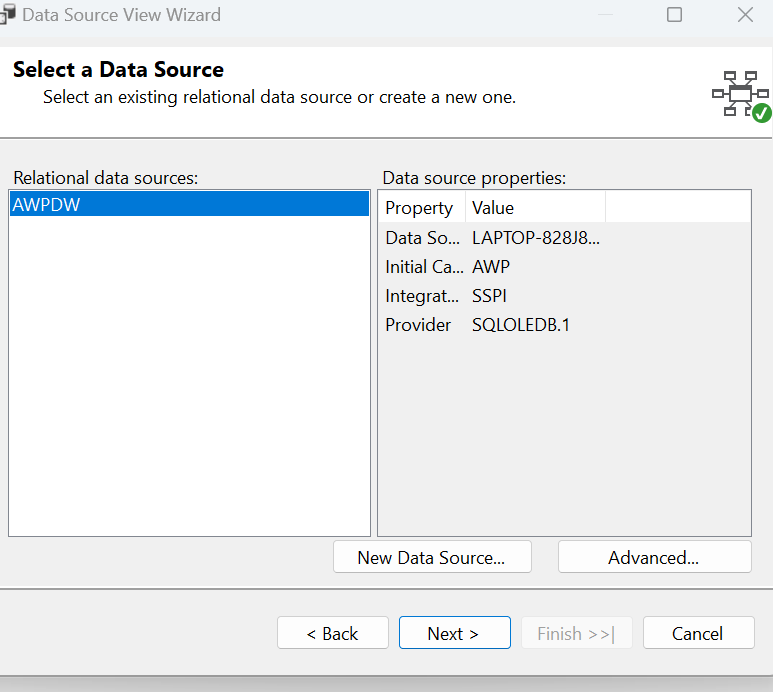
### 4.1.1. Tạo Data Source

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

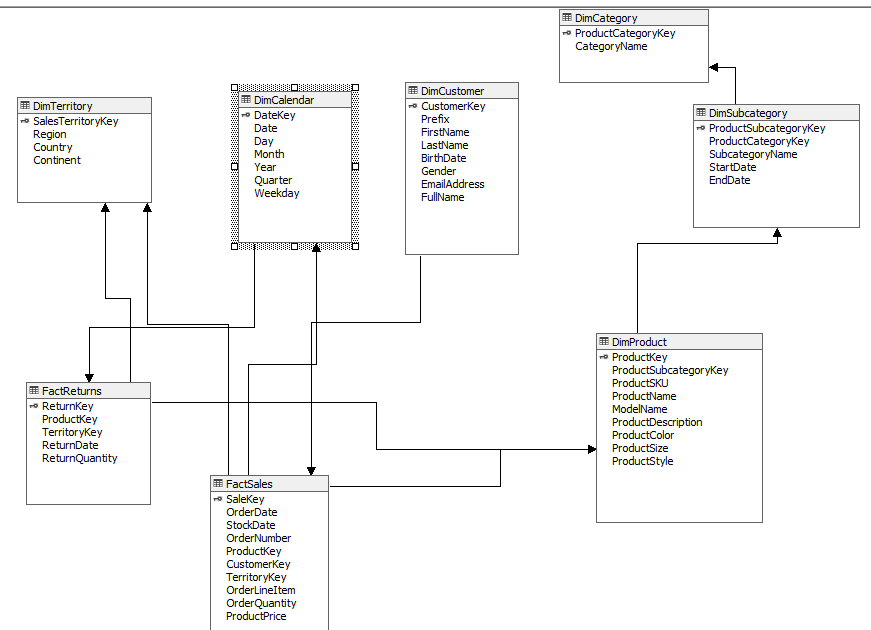


### 4.1.2. Tạo Data Source View

* 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**.

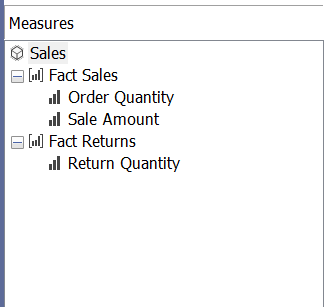


* Đặt tên cho Data Source View và nhấn Finish để hoàn tất việc tạo Data Source View
* Double-Click vào Data Source View vừa tạo để hiển thị diagram

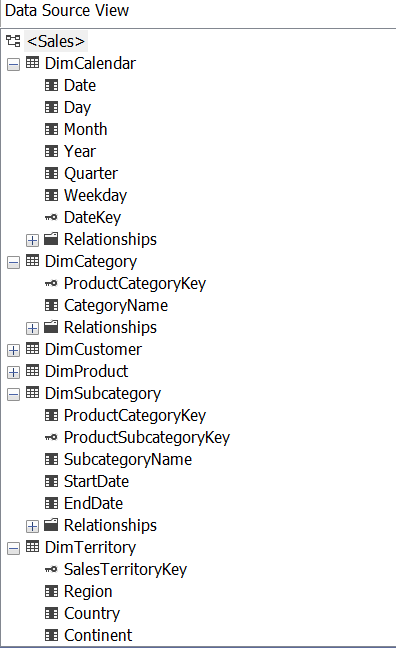


## 4.2. Quá trình xây dựng khối Cube

* 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



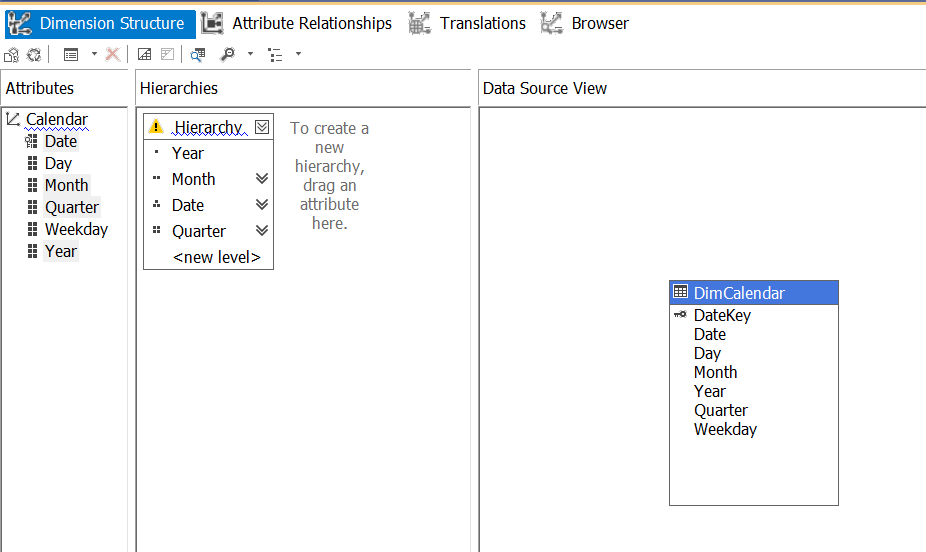
* Đặt tên **Sales.cube** để hoàn tất khởi tạo.
* Các bảng Fact, Dimension và quan hệ giữa chúng sẽ được thể hiện trong Diagram



## 4.3. Cấu hình Hierarchy

### 4.3.1. Tạo Hierarchy cho Dem Calendar

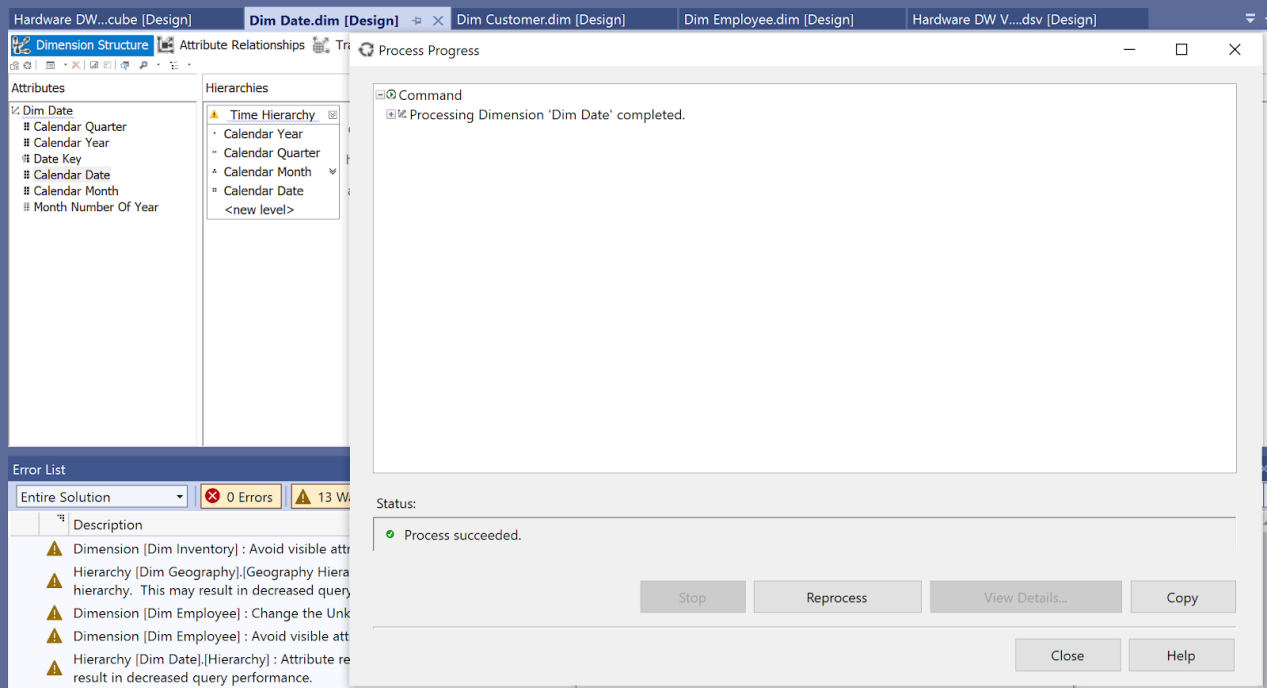
* Thêm thuộc tính vào bảng Dim: Tại table Dim Calendar trong panel Data Source View, kéo thuộc tính cần thiết vào panel Attributes.
* Kéo thả 4 thuộc tính *Year, Quarter, Name, Date Key* từ Attributes vào panel Hierarchies để tạo ra Time Hierarchy
* Đặt lại tên cho các thuộc tính trong Time Hierarchy



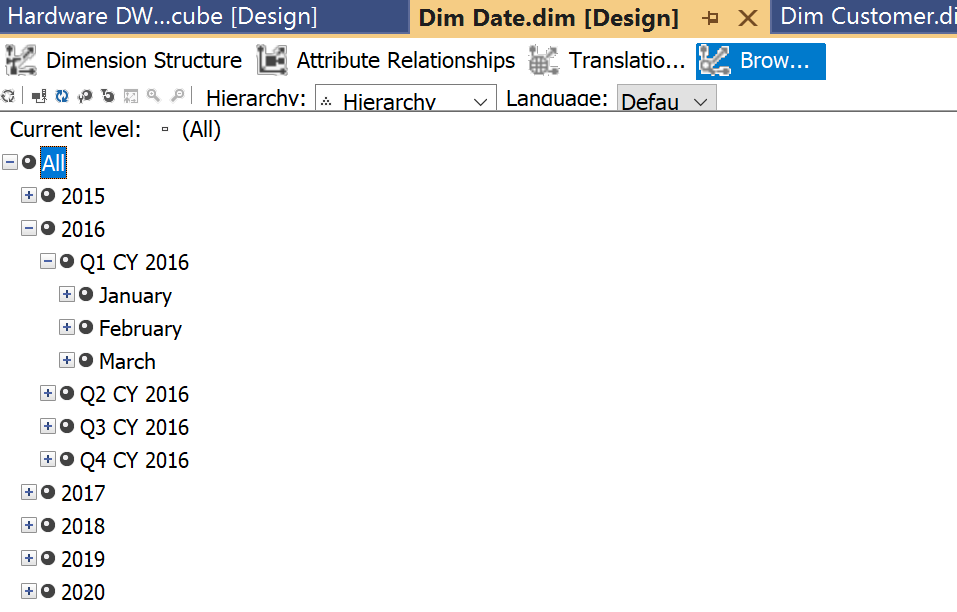
* Tạo Attribute Relationship



* 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.



### 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

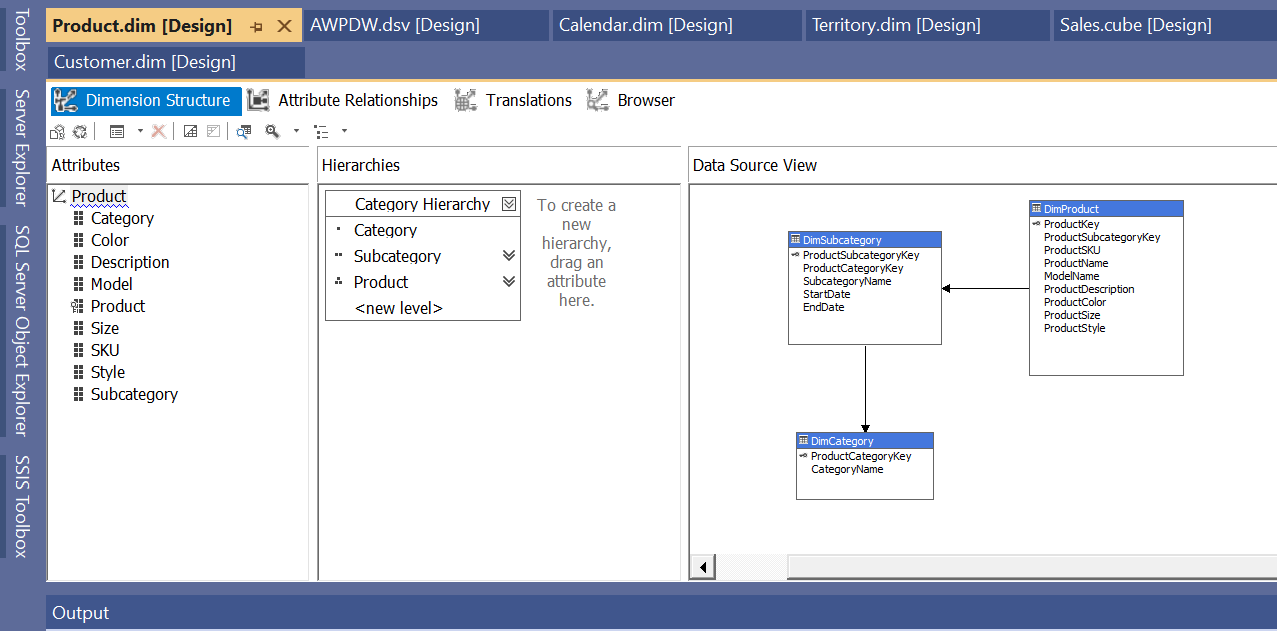


* Process bảng Dim Customer

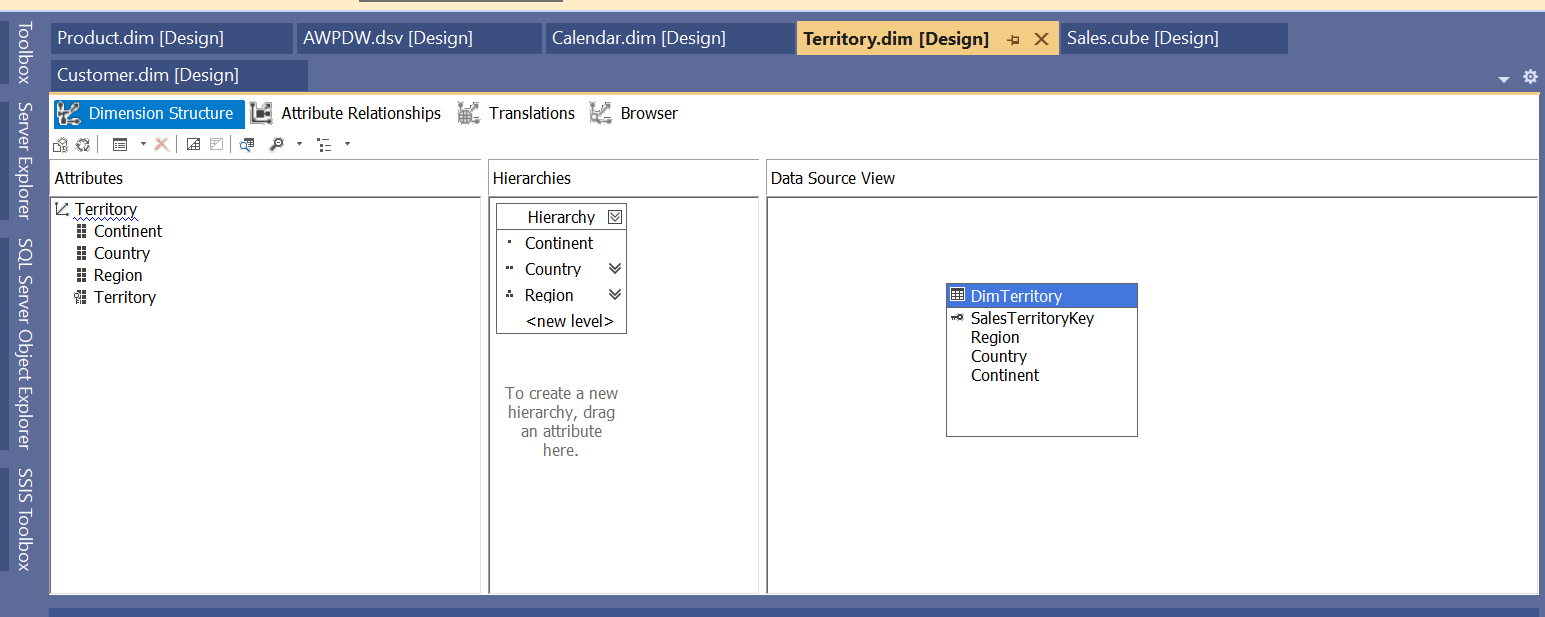


### 4.3.3. Tạo Hierarchy cho Dim Product

* Tạo Product Hierarchy



### 4.3.4. Tạo Hierarchy cho Dim Territory

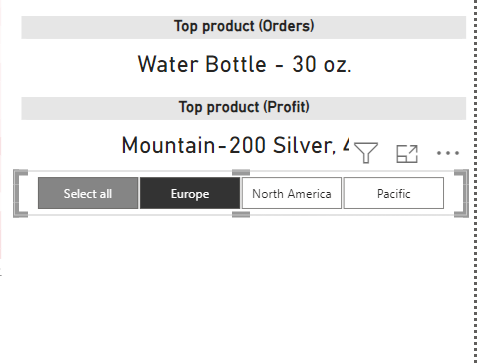


## 4.4. Thực hiện phân tích dữ liệu

### 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

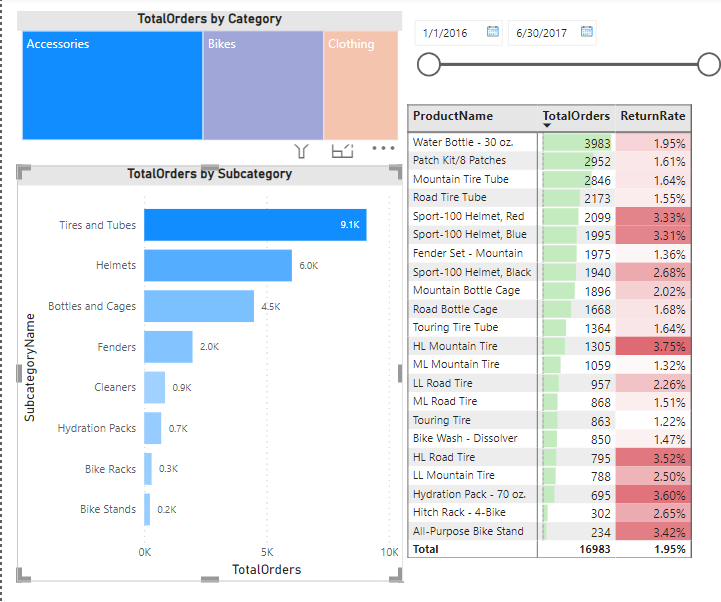
* Biểu đồ sản phẩm nào có số lượng bán ra cao nhất ở Europe từ 1/1/2016 đến 30/6/2017



### 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.

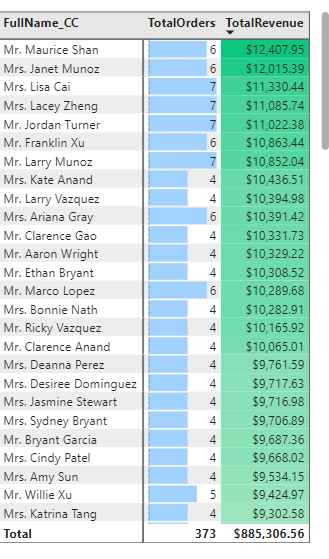


* Nhìn vào biểu đồ ngang, ta có thể thấy sản phẩm Water Bottle là sản phẩm có lượt tiêu thụ cao nhất với gần bốn nghìn lượt bán ra.

### 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



# 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](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].