|  |
| --- |
| **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 CHO CỬA HÀNG BÁN XE ĐẠP** |
| **Môn học: KHO DỮ LIỆU**  **Mã lớp học phần:DAWH430784\_23\_2\_02**  **GVHD: ThS. Nguyễn Văn Thành** |
| **Nhóm sinh viên thực hiện: Nhóm 14**   |  |  | | --- | --- | | Phạm Quốc Vương | 21110731 | | Phún Khoắn Vồ | 21110942 | | Nguyễn Anh Dũng | 21110873 | |

**TP. Hồ Chí Minh, 21 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 | Phạm Quốc Vương | 21110731 | 100% |
| 2 | Phún Khoắn Vồ | 21110942 | 100% |
| 3 | Nguyễn Anh Dũng | 21110873 | 100% |

**Nhận xét của giảng viên:**

**……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………**

*Ngày … tháng 05 năm 2024*

*Giảng viên chấm điểm*

*Ths. Nguyễn Văn Thành*

# PHÂN CÔNG NHIỆM VỤ

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

**Nhóm 14**

| **Nhiệm vụ** | **Anh Dũng** | **Khoắn Vồ** | **Quốc Vương** |
| --- | --- | --- | --- |
| Tìm kiếm tập dữ liệu | **x** | **x** | **x** |
| **Mô hình hóa dữ liệu vào workbook** | **x** | **x** | **x** |
| + DimStaffs  + DimStores  + DimLocations |  | **x** |  |
| **+** DimProducts  + FactOrders  **+** DimDate | **x** |  |  |
| + DimCustomers  + FactSales |  |  | **x** |
| **Thực hiện ETL đưa dữ liệu vào các bảng dim và fact dùng SSIS** | **x** |  | **x** |
| + Tạo và đưa dữ liệu vào stage |  |  | **x** |
| + Đưa dữ liệu từ stage vào các bảng dimension | **x** |  |  |
| + Tạo và đưa dữ liệu vào FactSales | **x** |  |  |
| + Tạo và đưa dữ liệu vào FactOrders | **x** |  |  |
| **Data Cube Design with SSAS** | **x** | **x** | **x** |
| Nhập dữ liệu vào SSAS, tạo data source, data source view và cube |  | **x** |  |
| Tạo các phân cấp hierarchy cho các bảng dim và các attribute relationships. |  |  | **x** |
| + Hierarchy cho DimDate | **x** |  |  |
| + Hierarchy cho DimProduct |  | **x** |  |
| + Hierarchy cho DimLocation |  |  | **x** |
| Tạo Cubes |  | **x** |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Nhiệm vụ** | **Anh Dũng** | **Khoắn Vồ** | **Quốc Vương** |
| **Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Sales Reporting** | **x** | **x** | **x** |
| + Cho Biết xếp hạng KPI của Nhân viên trong Tháng/ Quý/ Năm |  | **x** |  |
| + Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất trong tháng/ quý/ năm | **x** |  |  |
| + Cho biết doanh số bán hàng theo từng State trong tháng/ quý/ năm |  |  | **x** |
| + Cho biết top những nhân viên bán được nhiều sản phẩm nhất Trong Tháng/ Quý/ Năm. |  | **x** |  |
| +Cho biết State có số lượng Sản phẩm tiêu thụ nhiều nhất | **x** |  |  |
| + Cho biết Cửa Hàng có nhiều đơn hàng nhất trong tháng/ quý/ năm |  | **x** |  |
| + Cho biết Nhãn Hàng được mua nhất trong tháng/ quý/ năm |  |  | **x** |
| + Chọn ra top 10 khách hàng mua nhiều sản phẩm nhất. |  |  | **x** |
| Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Order and delivery | **x** |  |  |
| + Cho biết Tổng tiền tất cả đơn hàng của từng của hàng trong tháng/ quý/ năm. | **x** |  |  |
| Trả lời câu hỏi dùng SSAS | **x** |  |  |
| Trả lời câu hỏi dùng Power BI Desktop |  | **x** | **x** |

**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**

[PHÂN CÔNG NHIỆM VỤ 8](#_Toc167253232)

[CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI 1](#_Toc167253233)

[1.1. Lý do chọn đề tài 1](#_Toc167253234)

[1.2. Tổng quan về tập dữ liệu 1](#_Toc167253235)

[1.2.1. Nguồn dữ liệu 1](#_Toc167253236)

[1.2.2. Mô tả chi tiết tập dữ liệu 2](#_Toc167253237)

[1.2.2.1. Bảng CUSTOMERS 3](#_Toc167253238)

[1.2.2.2. Bảng STAFFS 3](#_Toc167253239)

[1.2.2.3. Bảng STORES 4](#_Toc167253240)

[1.2.2.4. Bảng ORDERS 4](#_Toc167253241)

[1.2.2.5. Bảng ORDER\_ITEMS 5](#_Toc167253242)

[1.2.2.6. Bảng CATEGORIES 5](#_Toc167253243)

[1.2.2.7. Bảng PRODUCTS 5](#_Toc167253244)

[1.2.2.8. Bảng STOCKS 5](#_Toc167253245)

[1.2.2.9. Bảng BRANDS 6](#_Toc167253246)

[1.3. Giới thiệu các công cụ được sử dụng 7](#_Toc167253247)

[CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP) 8](#_Toc167253248)

[2.1. Xác định các Business Process và bảng Fact 8](#_Toc167253249)

[2.1.1. Business Process: Sales Analysis 8](#_Toc167253250)

[2.1.2. Business Process: Order Fulfillment 8](#_Toc167253251)

[2.2. Xây dựng các bảng Dimension 9](#_Toc167253252)

[2.2.1. DimDate 9](#_Toc167253253)

[2.2.2. DimCustomers 9](#_Toc167253254)

[2.2.3. DimProducts 9](#_Toc167253255)

[2.2.4. DimStaffs 10](#_Toc167253256)

[2.2.5. DimStores 10](#_Toc167253257)

[2.2.6. DimLocation 10](#_Toc167253258)

[2.3. Star Schema (Lược đồ hình sao) 10](#_Toc167253259)

[CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS) 11](#_Toc167253260)

[3.1. Tạo ETL project và solution 11](#_Toc167253261)

[**3.2.** **Thiết lập các Connection** 11](#_Toc167253262)

[3.3. Import Dữ liệu vào các bảng dimension: 16](#_Toc167253263)

[3.3.1 Date Dimension 16](#_Toc167253264)

[3.3.2 **Products** 40](#_Toc167253265)

[3.3.3 **DimStore** 48](#_Toc167253266)

[**3.3.4** **Staffs** 51](#_Toc167253267)

[3.3.5 **Location** 56](#_Toc167253268)

[3.3.6 **Date** 61](#_Toc167253269)

[3.4. Import dữ liệu vào bảng các bảng fact 66](#_Toc167253270)

[3.4.1. Fact Sales 66](#_Toc167253271)

[**3.4.2. Fact Orders** 72](#_Toc167253272)

[CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS) 80](#_Toc167253273)

[4.1. Quá trình xây dựng mô hình 80](#_Toc167253274)

[**4.1.1.** **Tạo một Analysis Services Project** 80](#_Toc167253275)

[**4.1.2.** **Kết nối đến Analysis Services Server** 81](#_Toc167253276)

[4.1.3. Tạo Data Source 83](#_Toc167253277)

[4.1.4. Tạo Data Source View 87](#_Toc167253278)

[4.2. Quá trình xây dựng khối Cube 91](#_Toc167253279)

[4.3. Cấu hình các dimension 94](#_Toc167253280)

[4.4. Thực hiện phân tích dữ liệu 99](#_Toc167253281)

[4.4.1. Câu hỏi: Cho Biết xếp hạng KPI của Nhân viên trong Tháng/ Quý/ Năm. 99](#_Toc167253282)

[4.4.2. Câu hỏi: Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất trong tháng/ quý/ năm. 100](#_Toc167253283)

[4.4.3. Câu hỏi: Cho biết doanh số bán hàng theo từng State trong tháng/ quý/ năm. 102](#_Toc167253284)

[4.4.4. Câu hỏi: Cho biết top những nhân viên bán được nhiều sản phẩm nhất Trong Tháng/ Quý/ Năm. 103](#_Toc167253285)

[4.4.5. Câu Hỏi: Chọn ra top 10 khách hàng mua nhiều sản phẩm nhất. 104](#_Toc167253286)

[4.4.6. Câu hỏi: Cho biết Tổng tiền tất cả đơn hàng của từng của hàng trong tháng/ quý/ năm. 106](#_Toc167253287)

[4.4.7. Câu hỏi: Cho biết Cửa Hàng có nhiều đơn hàng nhất trong tháng/ quý/ năm. 107](#_Toc167253288)

[4.4.8. Câu Hỏi: Cho biết State có số lượng Sản phẩm tiêu thụ nhiều nhất. 108](#_Toc167253289)

[4.4.9. Câu hỏi: Cho biết Nhãn Hàng được mua nhất trong tháng/ quý/ năm. 109](#_Toc167253290)

[CHƯƠNG 5: KẾT LUẬN 111](#_Toc167253291)

[5.1. Kết quả đạt được 111](#_Toc167253292)

[5.2. Những hạn chế 111](#_Toc167253293)

[5.3. Tài liệu tham khảo 111](#_Toc167253294)

Link video báo cáo :

# CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

## Lý do chọn đề tài

Trong những năm gần đây, nhu cầu sử dụng xe đạp đã tăng lên đáng kể do các yếu tố như ý thức về sức khỏe và lối sống lành mạnh, sự quan tâm đến việc bảo vệ môi trường, và xu hướng đô thị hóa khiến xe đạp trở thành phương tiện giao thông tiện lợi và thân thiện. Điều này đặc biệt quan trọng khi người tiêu dùng ngày càng ưu tiên các giải pháp di chuyển bền vững và giảm lượng khí thải CO2. Việc quản lý một cửa hàng bán xe đạp không chỉ đơn giản là việc bán sản phẩm, mà còn bao gồm nhiều khía cạnh phức tạp như quản lý hàng tồn kho để đảm bảo cung cấp đủ sản phẩm mà không bị tồn đọng quá nhiều, và chiến lược giá cả cùng các chương trình khuyến mãi để thu hút khách hàng. Ngoài ra, phân tích hành vi và nhu cầu của khách hàng là yếu tố quan trọng để cung cấp dịch vụ tốt hơn và tăng cường sự hài lòng của khách hàng. Vì vậy, việc chọn đề tài phân tích kho dữ liệu về cửa hàng bán xe đạp không chỉ giúp tối ưu hóa các hoạt động kinh doanh mà còn góp phần hiểu rõ hơn về xu hướng và nhu cầu thị trường, từ đó đưa ra các quyết định chiến lược hiệu quả.

## Tổng quan về tập dữ liệu

### Nguồn dữ liệu

* Nhóm sử dụng Tập dữ liệu Bike Store Database được lấy từ trang web Kaggle (kaggle.com), trang web này cung cấp cơ sở dữ liệu mẫu để phục vụ cho việc luyện tập thao tác của người học .
* Đường dẫn tải tập dữ liệu:

https://www.kaggle.com/datasets/dillonmyrick/bike-store-sample-database/data?select=order\_items.csv

### Mô tả chi tiết tập dữ liệu

* Tập dữ liệu ‘Bike Store’ bao gồm thông tin về các cửa hàng bán xe đạp. Dữ liệu bao gồm các 9 bảng với mô tả như sau:

| **Tên bảng** | **Mô tả** | **Số dòng** |
| --- | --- | --- |
| CUSTOMERS | lưu trữ thông tin người liên hệ của khách hàng | 100 dòng |
| STAFFS | lưu trữ thông tin nhân viên | 10 dòng |
| STORES | lưu trữ thông tin cửa hàng | 3 dòng |
| ORDERS | lưu trữ thông tin đơn hàng | 1616 dòng |
| ORDER\_ITEMS | lưu trữ thông tin chi tiết đơn hàng | 9445 dòng |
| CATEGORIES | lưu trữ vị trí của ngành hàng | 15 dòng |
| PRODUCTS | lưu trữ thông tin chung sản phẩm | 322 dòng |
| STOCKS | lưu trữ thông tin kho hàng | 940 dòng |
| BRANDS | lưu trữ thông tin thương hiệu | 37 dòng |

#### Bảng CUSTOMERS

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| customer\_id | Mã khách hàng(Khóa chính): dùng để xác định và phân biệt các khách hàng với nhau. |
| first\_name | Họ khách hàng |
| last\_name | Tên khách hàng |
| phone | Số điện thoại khách hàng |
| email | Địa chỉ email của khách hàng |
| street | Đường mà nơi khách hàng ở |
| city | Thành phố nơi khách hàng ở |
| state | Bang nơi khách hàng ở |
| zip\_code | Mã bưu điện |

#### Bảng STAFFS

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| staff\_id | Mã nhân viên (Khóa chính): Dùng để xác minh và phân biệt nhân viên |
| first\_name | Họ của nhân viên |
| last\_name | Tên của nhân viên |
| email | Email của nhân viên |
| phone | Số điện thoại của nhân viên |
| active | Trạng thái làm việc(1. Đang làm, 2. Đã nghỉ) |
| store\_id | Mã cửa hàng(Khóa ngoại): Tham chiếu đến bảng STORES |
| manager\_id | Mã quản lý |

#### Bảng STORES

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| store\_id | Mã cửa hàng (Khóa chính): để xác định và phân biệt cửa hàng |
| store\_name | Tên của cửa hàng |
| phone | Số điện thoại của cửa hàng |
| email | Email của cửa hàng |
| street | Đường nơi đặt cửa hàng |
| city | Thành phố nơi đặt của cửa hàng |
| state | Bang nơi đặt cửa hàng |
| zip\_code | Mã bưu chính |

#### Bảng ORDERS

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| order\_id | Mã đơn hàng (Khóa chính) |
| customer\_id | Mã khách hàng(Khóa ngoại) tham chiếu đến bảng CUSTOMERS |
| order\_status | Trạng thái của đơn hàng |
| required\_date | Ngày yêu cầu đặt hàng |
| shipped\_date | Ngày giao hàng |
| order\_date | Ngày đặt hàng |
| store\_id | Mã cửa hàng |
| staff\_id | Mã nhân viên |

#### Bảng ORDER\_ITEMS

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| order\_id | Mã đơn hàng (Khóa chính) |
| item\_id | Mã mặt hàng(Khóa chính) |
| product\_id | Mã sản phẩm(Khóa ngoại) tham chiếu đến bảng PRODUCTS |
| quantity | Số lượng sản phẩm |
| list\_price | Bảng giá |
| discount | Mã khuyến mãi |

#### Bảng CATEGORIES

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| category\_id | Mã ngành hàng (Khóa chính) |
| category\_name | Tên ngành hàng |

#### Bảng PRODUCTS

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| product\_id | Mã sản phẩm (Khóa chính) |
| product\_name | Tên sản phẩm |
| brand\_id | Mã thương hiệu |
| category\_id | Mã ngành hàng |
| model\_year | mẫu sản phẩm |
| list\_price | Bảng giá |

#### Bảng STOCKS

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| store\_id | Mã cửa (Khóa chính) |
| product\_id | Mã sản phẩm |
| quantity | Số lượng sản phẩm |

#### Bảng BRANDS

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| brand\_id | Mã thương hiệu (Khóa chính) |
| brand\_name | Tên của thương hiệu |

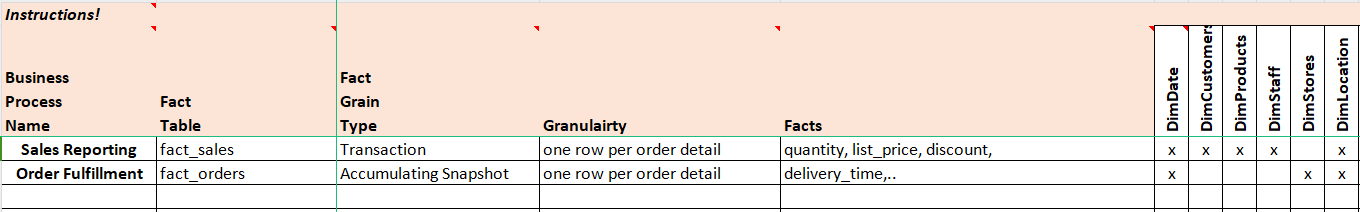
## Giới thiệu các công cụ được sử dụng

* 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

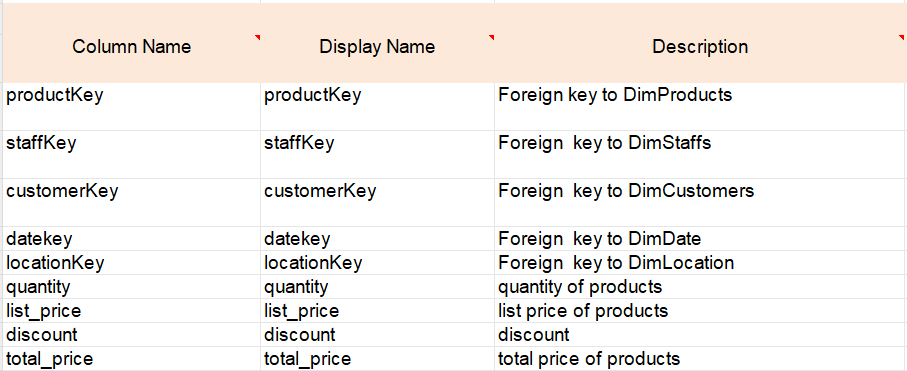
* Xây dựng Detailed Bus Matrix xác định các Business Process, bảng Fact, bảng Dim cần thiết.



### 2.1.1. Business Process: Sales Reporting

* Cho Biết xếp hạng KPI của Nhân viên trong Tháng/ Quý/ Năm.
* Cho biết danh mục các sản phẩm được tiêu thụ nhiều nhất trong tháng/ quý/ năm.
* Cho biết doanh số bán hàng theo từng State trong tháng/ quý/ năm.
* Cho biết top những nhân viên bán được nhiều sản phẩm nhất Trong Tháng/ Quý/ Năm.
* Chọn ra top 10 khách hàng mua nhiều sản phẩm nhất.
* Cho biết State có số lượng Sản phẩm tiêu thụ nhiều nhất.
* Cho biết Nhãn Hàng được mua nhất trong tháng/ quý/ năm.

Bảng **FactSales**



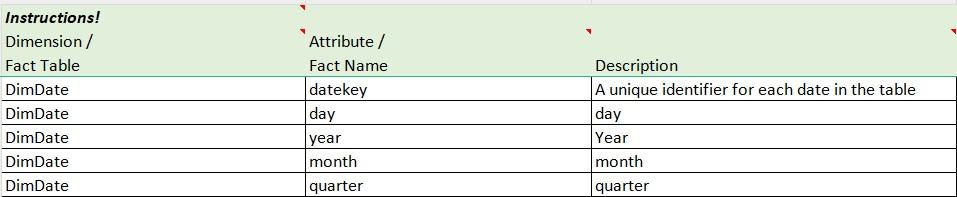
### 2.1.2. Business Process: Order Fulfillment

* Các câu hỏi cụ thể được đặt ra:
* Cho biết Tổng tiền tất cả đơn hàng của từng của hàng trong tháng/ quý/ năm.
* Cho biết Cửa Hàng có nhiều đơn hàng nhất trong tháng/ quý/ năm.
* Bảng **FactOrderFulfillment**

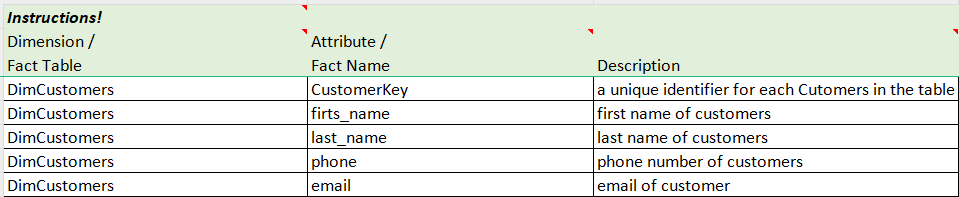


## 2.2. Xây dựng các bảng Dimension

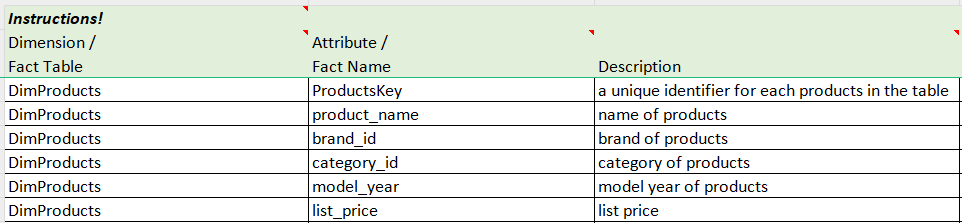
### 2.2.1. DimDate



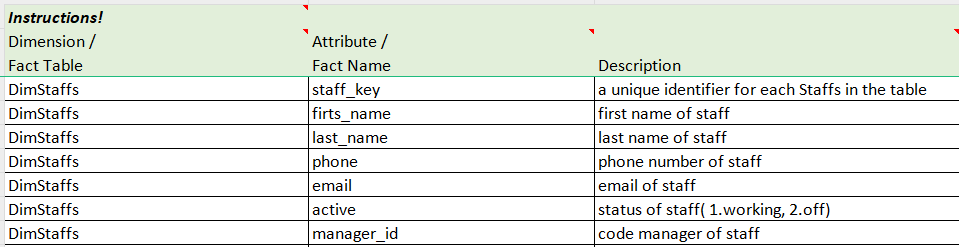
### 2.2.2. DimCustomers



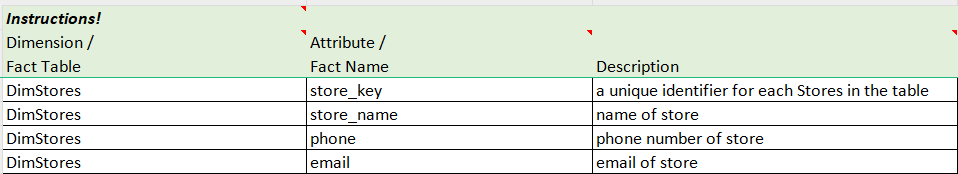
### 2.2.3. DimProducts



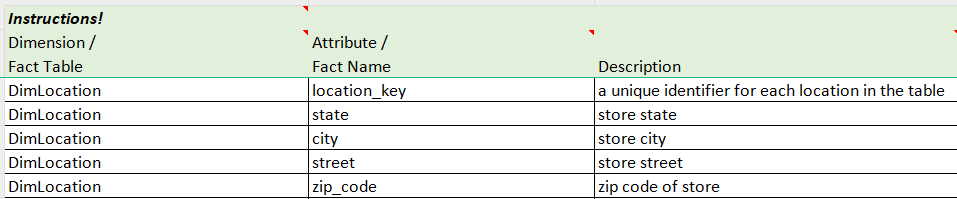
### 2.2.4. DimStaffs



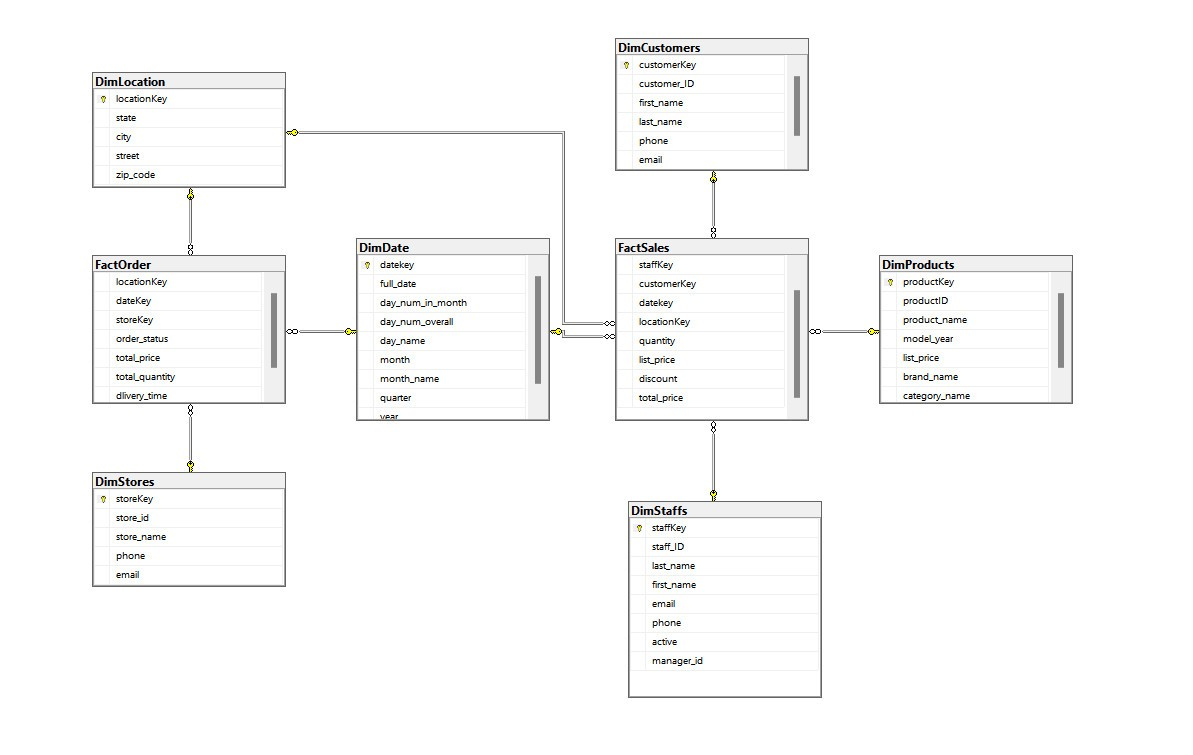
### 2.2.5. DimStores



### 2.2.6. DimLocation



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



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

## Tạo ETL project và solution

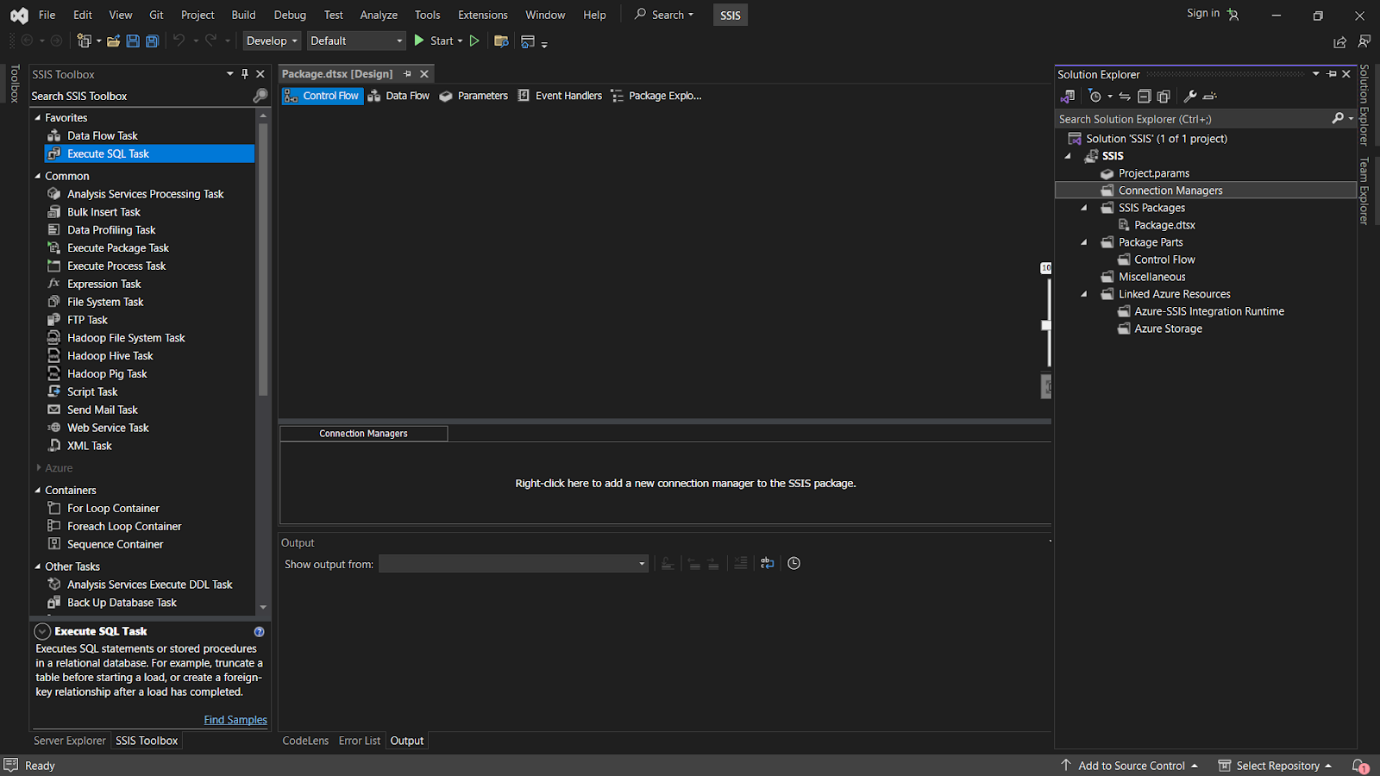
1. Từ **Visual Studio** menu, chọn **File** -> **New** -> **Project**.

2. Từ **dialog**:

a. Chọn **Integration Services Project**

b. Đặt tên: **SSIS**

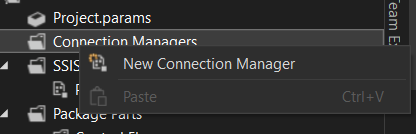
3. Click **OK**



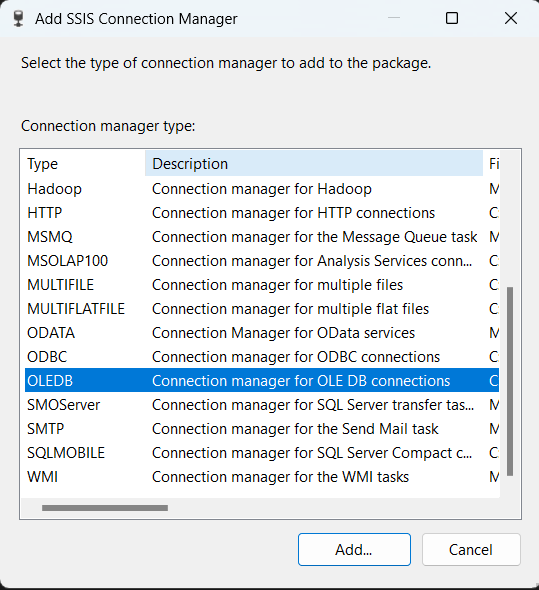
1. **Thiết lập các Connection**

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

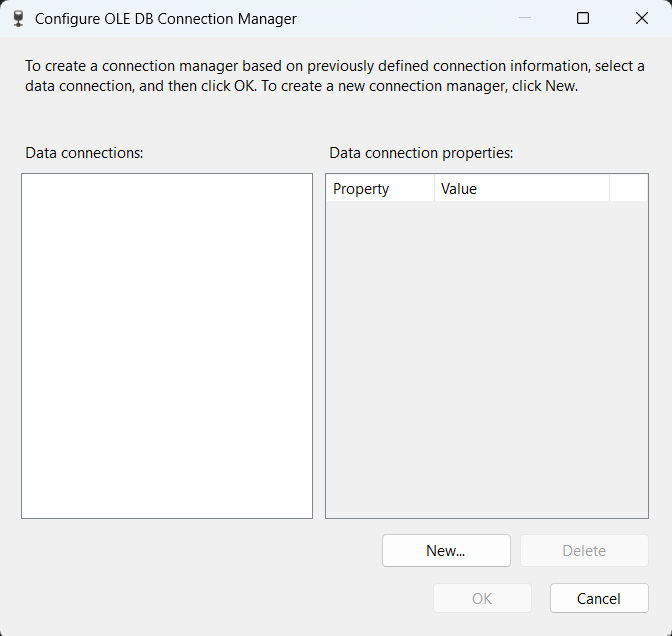
**Connection Manager**.



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



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

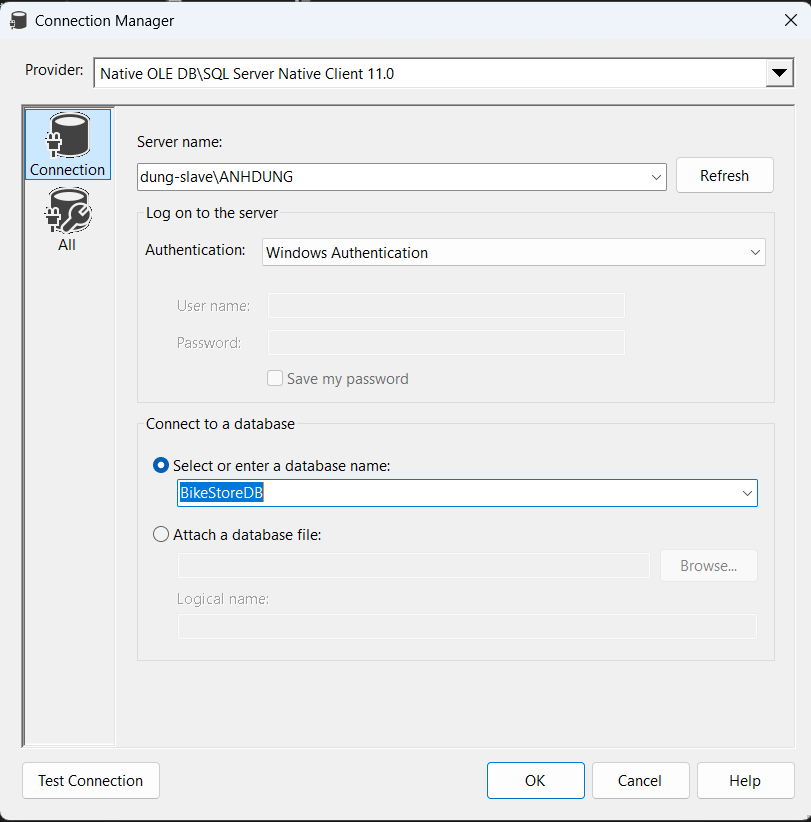
Đặt tên cho các thông tin như bên dưới:

Provider: **SQL Server Native Client**

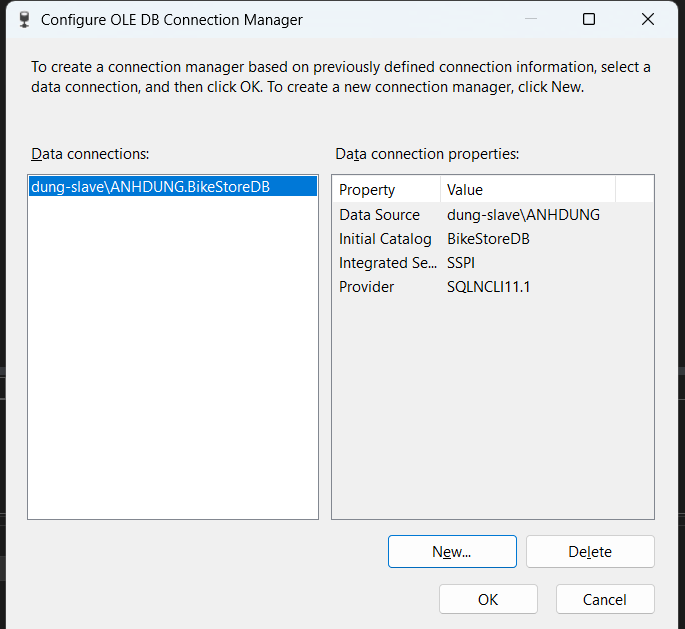
Server Name: **dung-slave\ANHDUNG**

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

Connect to a database: **BikeStoreDB**

****

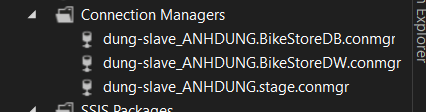
**Click Test Connection để kiểm tra connection hoạt động, sau đó click OK để lưu connection.**

****

4. Trở lại màn hình **Configure OLE DB Connection Manager**, ta được **dung-slave\ANHDUNG.BikeStoreDB** vừa được tạo:



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



## Import Dữ liệu vào các bảng dimension:

#### Date Dimension

1. Đầu tiên chép dữ liệu từ bảng [**Customers Dimension]** của CSDL **BikeStoreDB**

sang bảng **stgCustomers** CSDL **stage**. Sử dụng SSIS để tạo destination

table, truncate table trước khi staging dữ liệu.

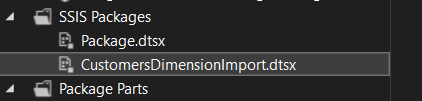
2. Ánh xạ các cột từ bảng **stgCustomers** sang bảng **DimCustomers**.

##### 3.3.1.1: Tạo Package

Tạo một package mới, **CustomersDimensionImport**.

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

Sau khi thực hiện xong ta được Package.dtsx và CustomersDimensionImport.dtsx như hình sau:

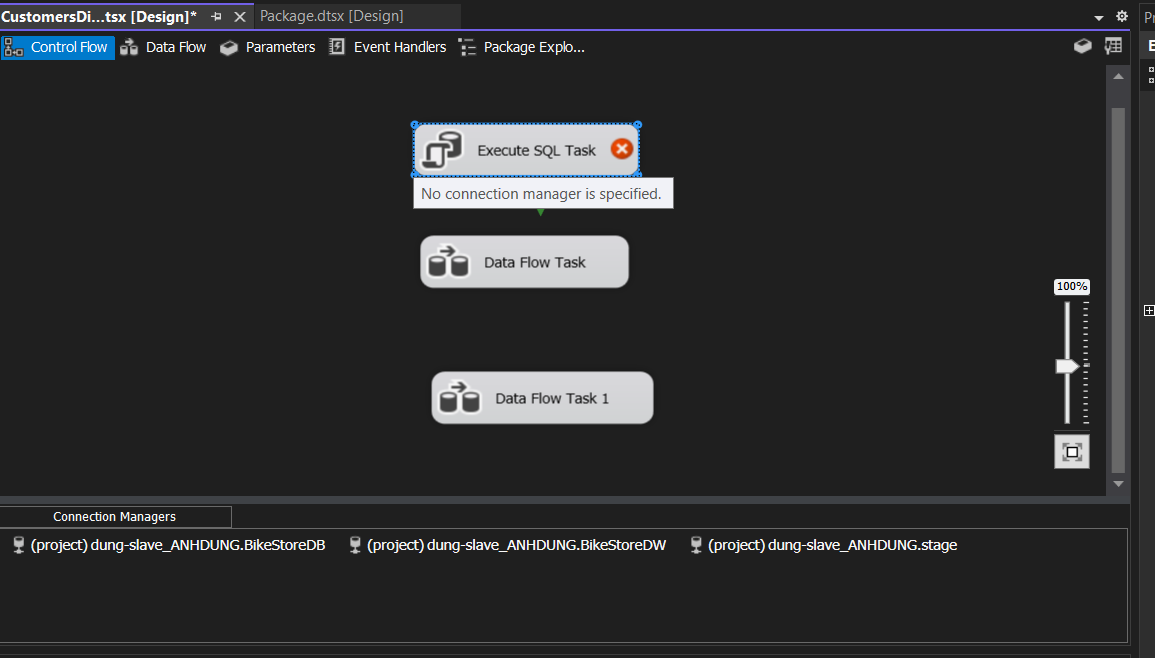
****

Double-click vào CustomersDimensionImport.dtsx package để mở nó**.**

##### **3.3.1.2: Thiết lập Control Flow**

1. Từ SSIS Toolbox tạo 1 Execute SQL task và 2 Data flow tasks vào surface design

như hình sau:

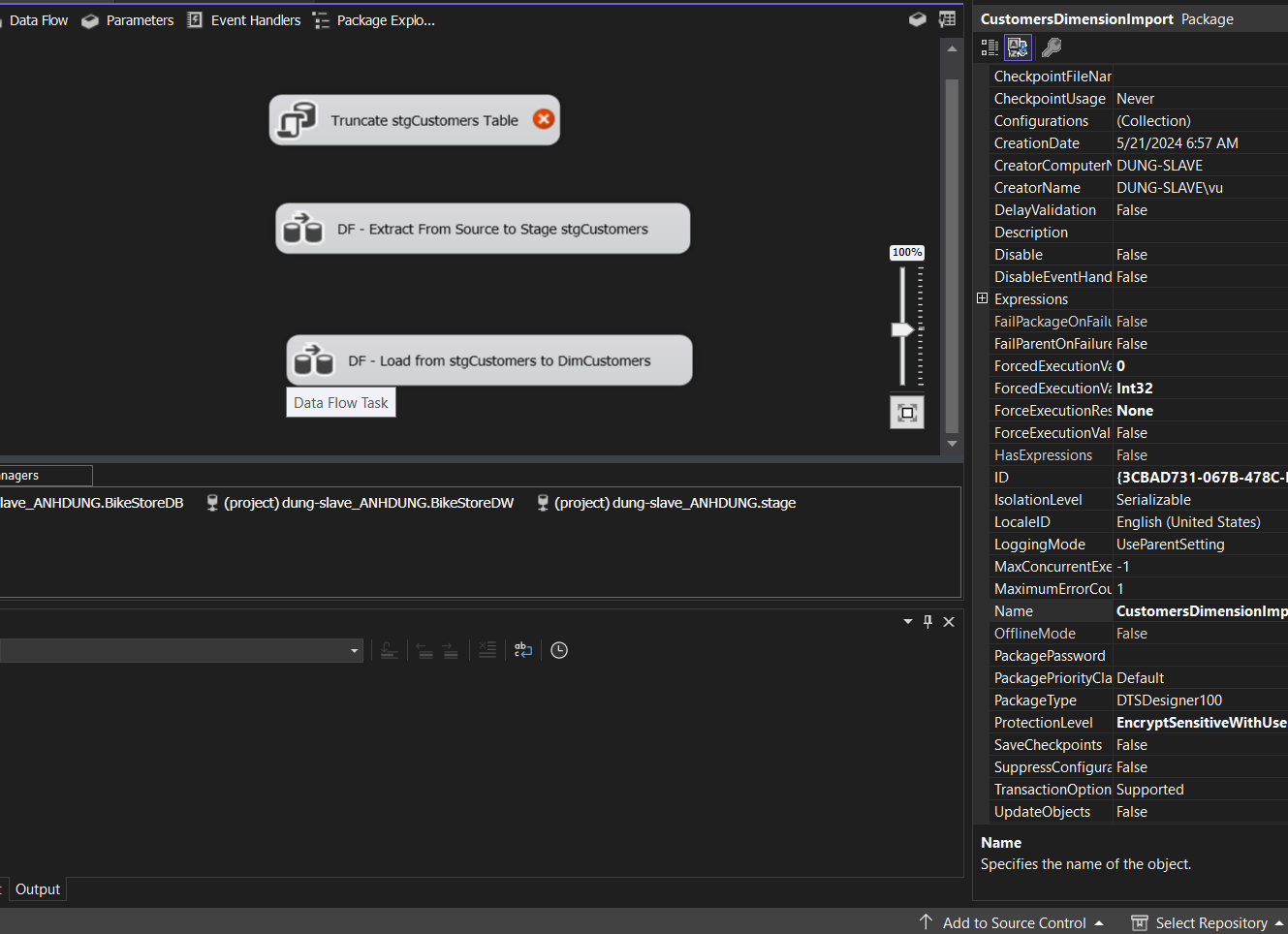


2. Đặt tên cho các **task**.

a. **Execute SQL Task SQL - Truncate stgCustomers Table**

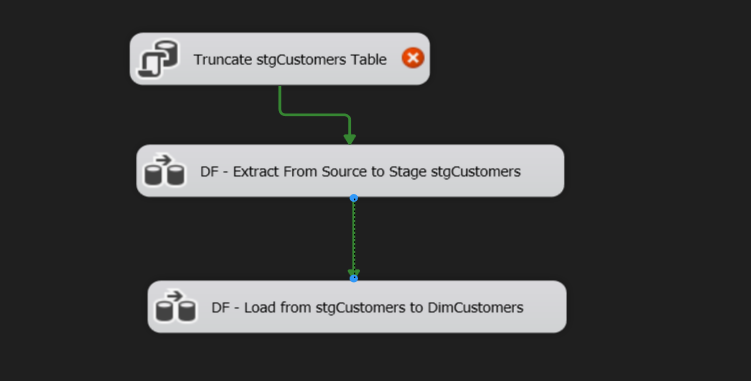
b. **Data Flow Task DF - Extract From Source to Stage stgCustomers**

c. **Data Flow Task 1 DF - Load from stgCustomers to DimCustomers**



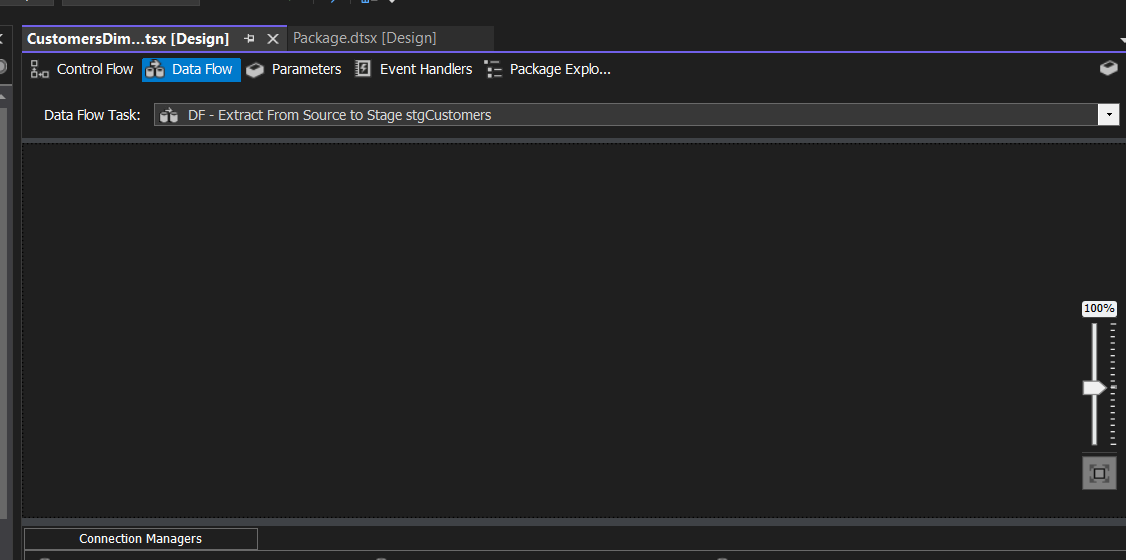
3. Connect các **tasks** để chúng có thể chạy theo thứ tự. Click vào **SQL - Truncate**

**stgCustomers Table** task và một mũi tên xanh sẽ xuất hiện. Kéo rê nó vào **DF - Extract From Source To Stage stgCustomers** task. Tương tự, connect hai task còn lại.

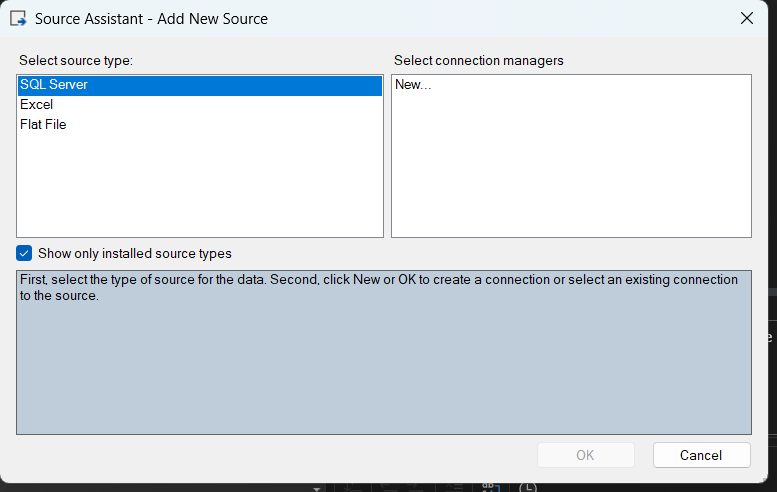


##### 3.3.1.3: Load dữ liệu từ Source vào Stage

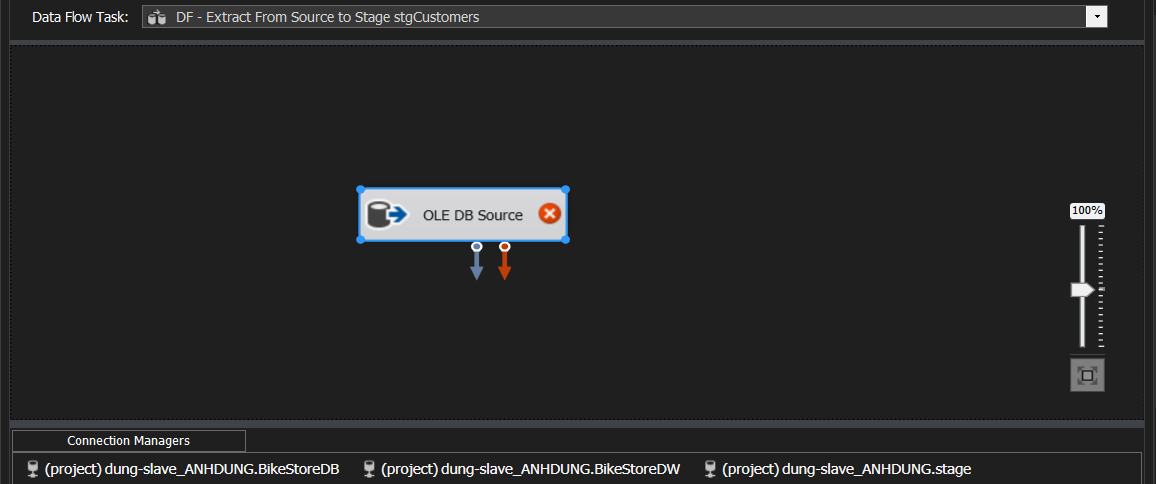
1. Double-click **DF - Extract From Source to Stage stgCustomers** task để mở nó trong **Data Flow design surface**:



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



3. OLE DB Source sẽ xuất hiện như sau:



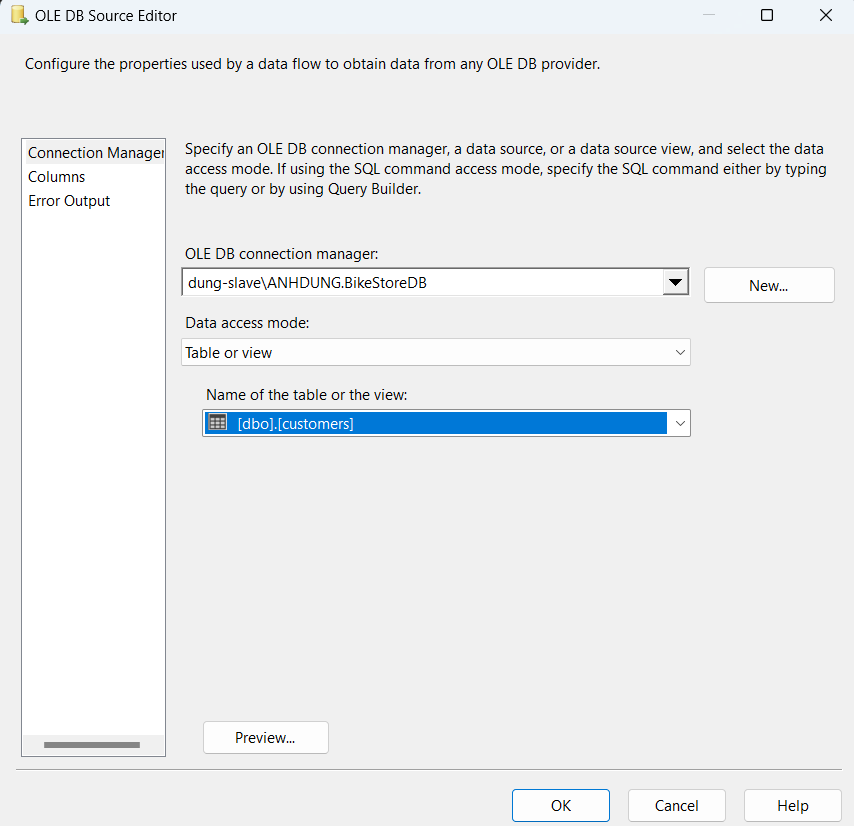
4. Đổi tên nó thành **SRC - BikeStore\_Customers\_dimension**

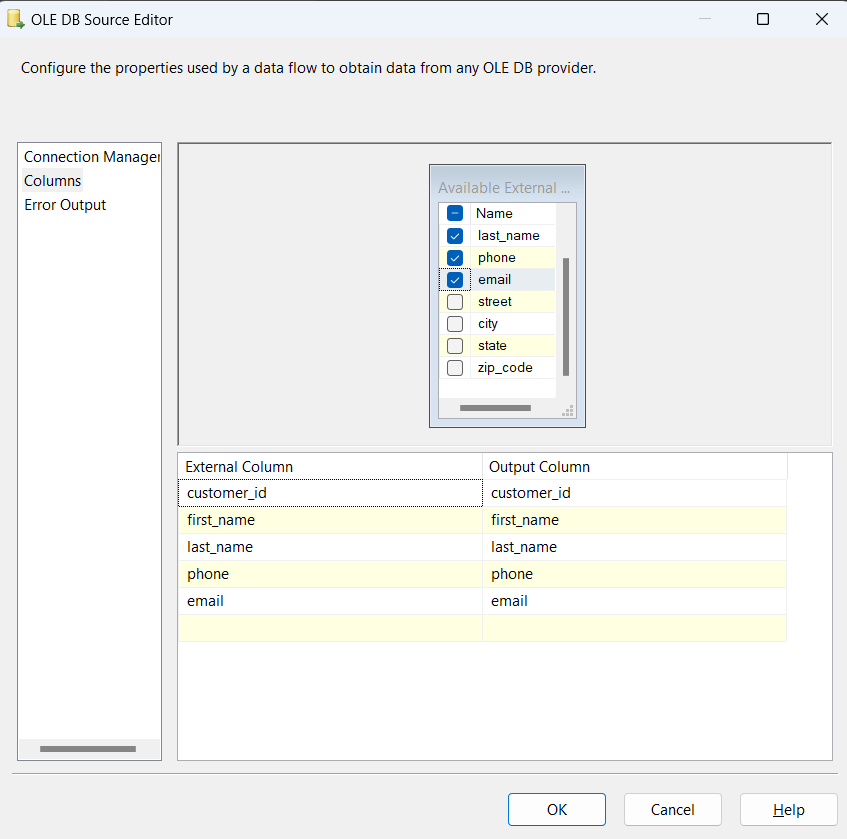


5. Double click vào source để cấu hình nó, **OLE DB Source Editor** sẽ mở ra. Từ

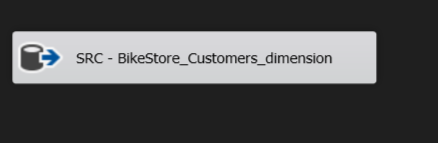
dropdown **name of the table or view** chọn **[dbo].[Customers Dimension]** là source.

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

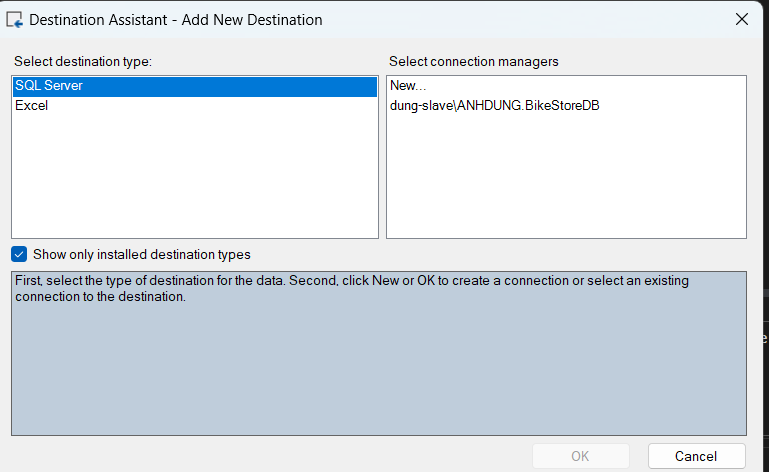




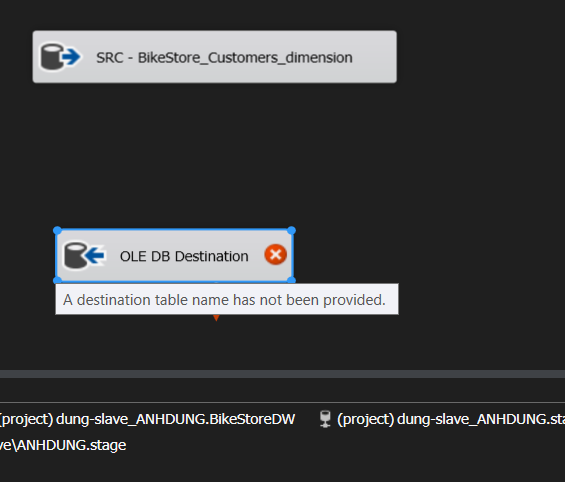
6. Sau khi cấu hình **source** xong, bạn sẽ không còn thấy dấu **[x]** màu đỏ (như bên dưới):



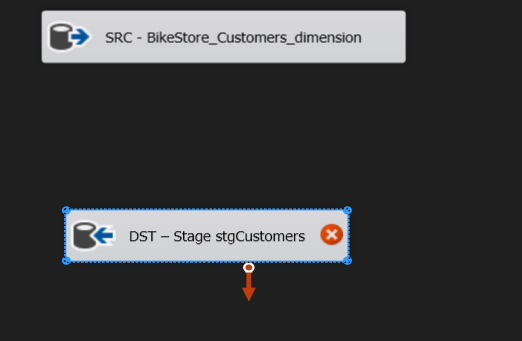
7. Tiếp tục, cấu hình destination. Từ **SSIS Toolbox**, kéo rê **Destination Assistant** vào design surface, một dialog xuất hiện. Chọn source Type: **SQL Server** và **stage** là connection manager.



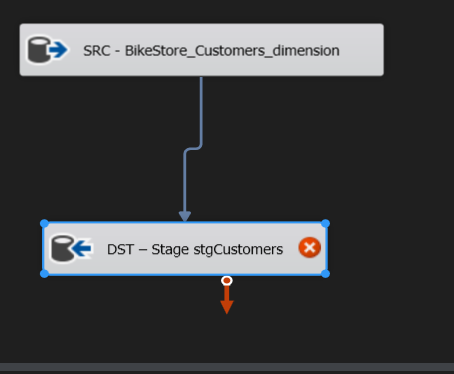
8. Ta có một ***destination*** như bên dưới:



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

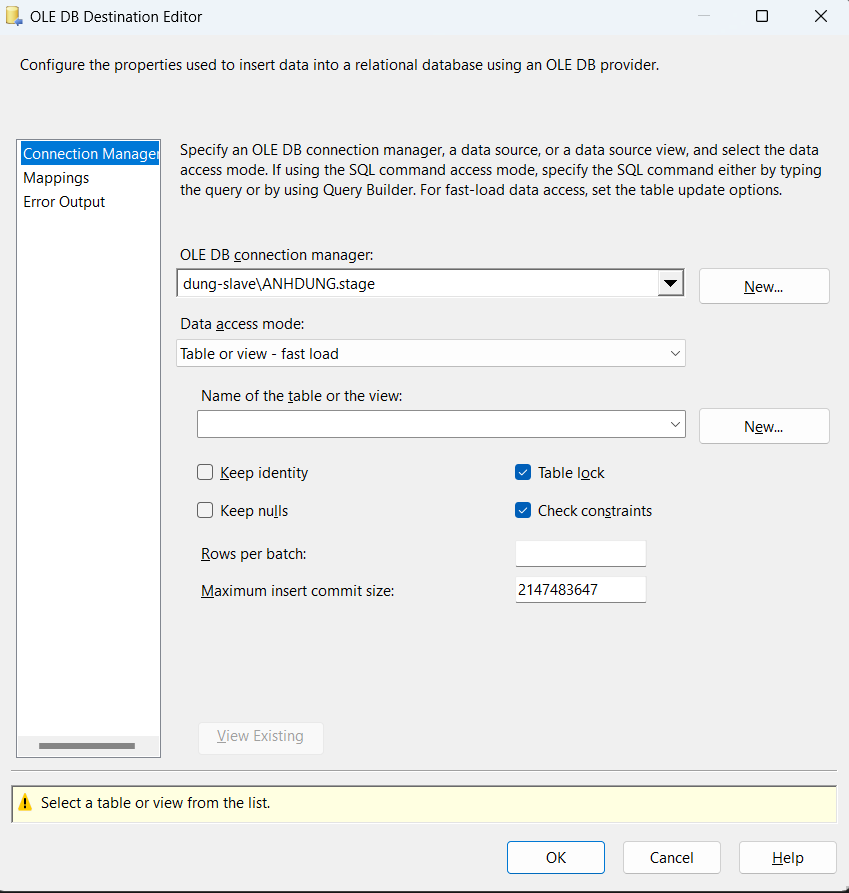
****

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

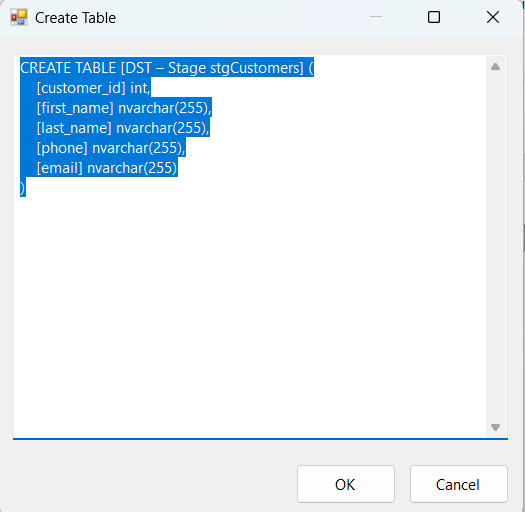
****

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

**hiện như sau:**

****

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

****

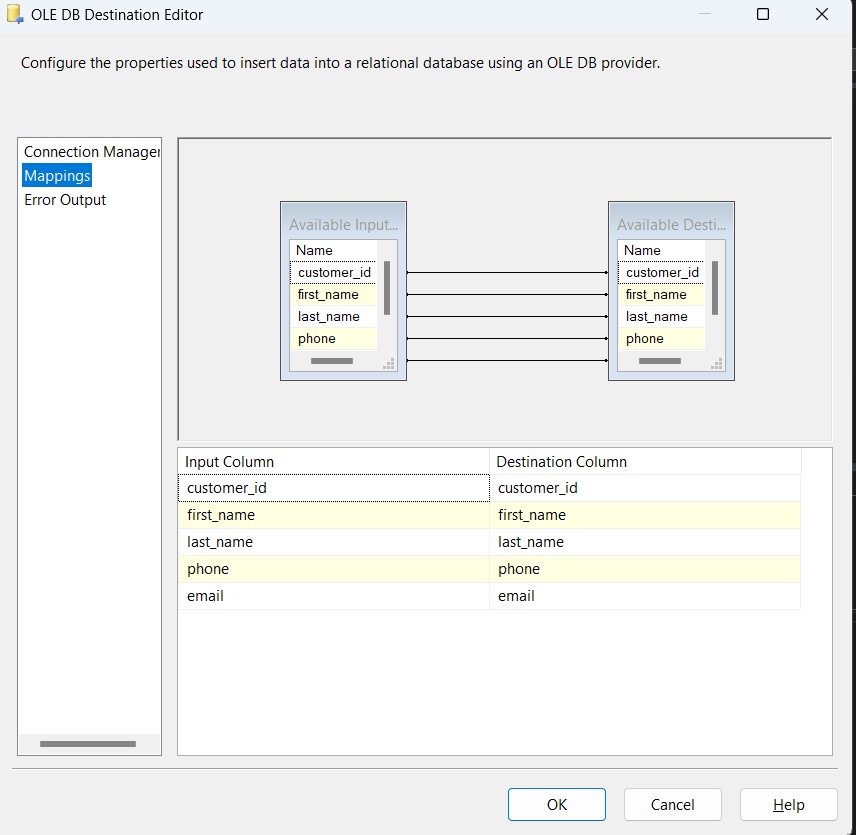
**Đổi tên bảng lại thành stgCustomers. (CREATE TABLE [stgCustomers]).**

**Click OK.**

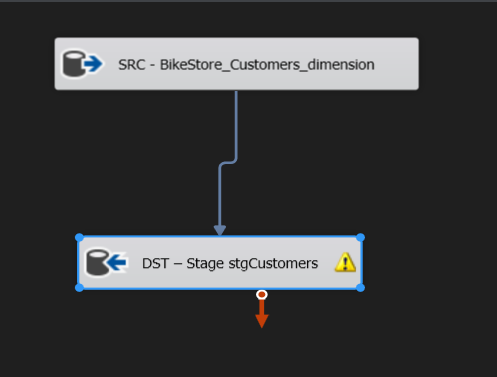
**Trở lại màn hình OLEDB Destination Editor, tại mục Name of the table or view bạn**

**sẽ thấy [stgCustomers].**

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

****

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

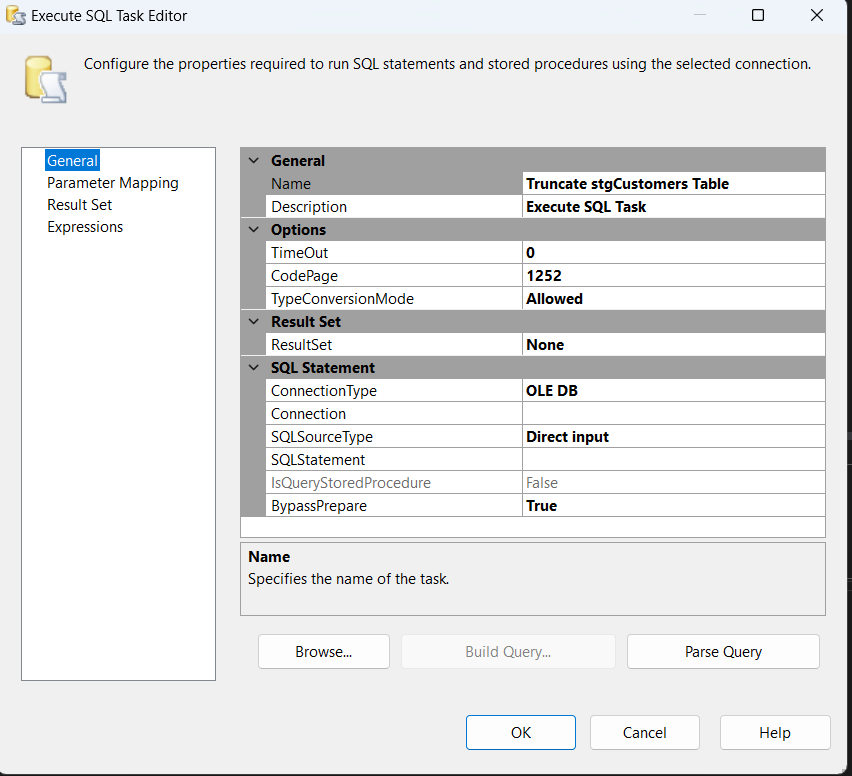
****

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

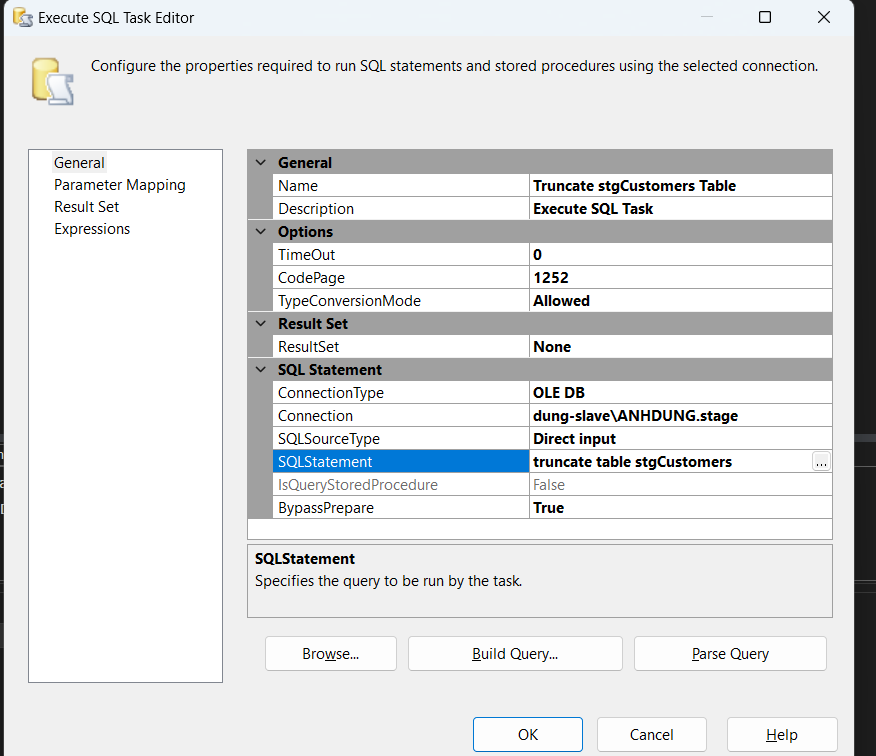
**Click Control Flow tab.**

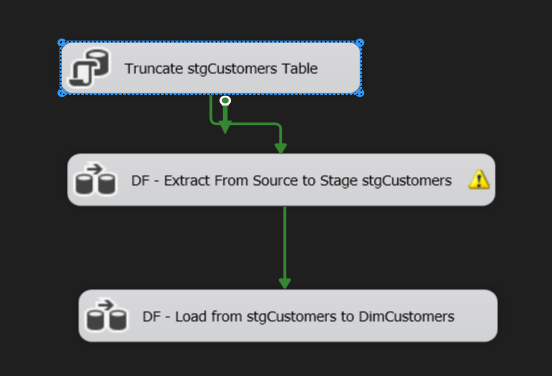
**1. Double-click vào SQL - Truncate stgCustomers Table task để cấu hình nó, Execute SQL**

**Task Editor sẽ mở ra.**

****

**2. Ở mục SQL Statement, chọn Connection là dung-slave\ANHDUNG.stage và SQLStatement là truncate table stgCustomers;**

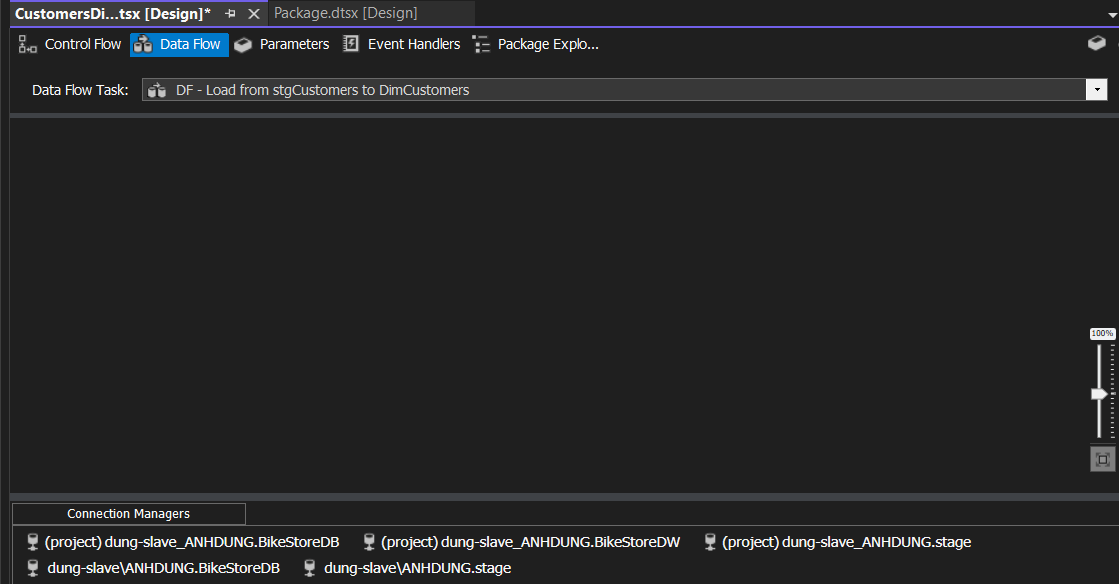
****

****

##### 3.3.1.5: Load dữ liệu từ Stage vào Dimension

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

**surface.**

****

**2. Kéo rê Source Assistant vào design surface, chọn source type là SQL Server và**

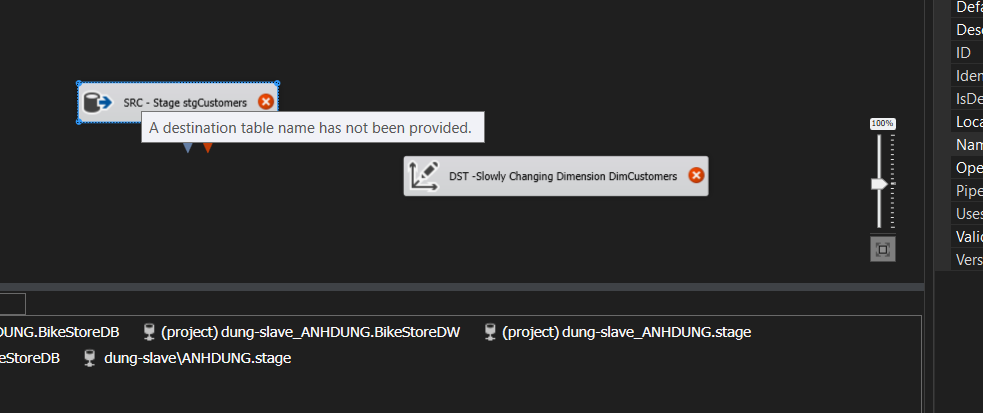
**connection manager là stage .**

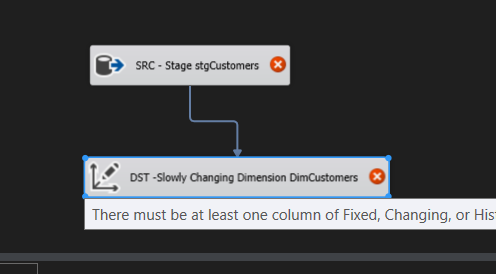
**3. Đổi tên source thành SRC - Stage stgCustomers, và double-click vào nó để cấu hình.**

**4. Từ OLE DB Source Editor chọn [dbo].[stgCustomers] table ở mục name of the table or**

**view. Click OK để hoàn tất việc cấu hình.**

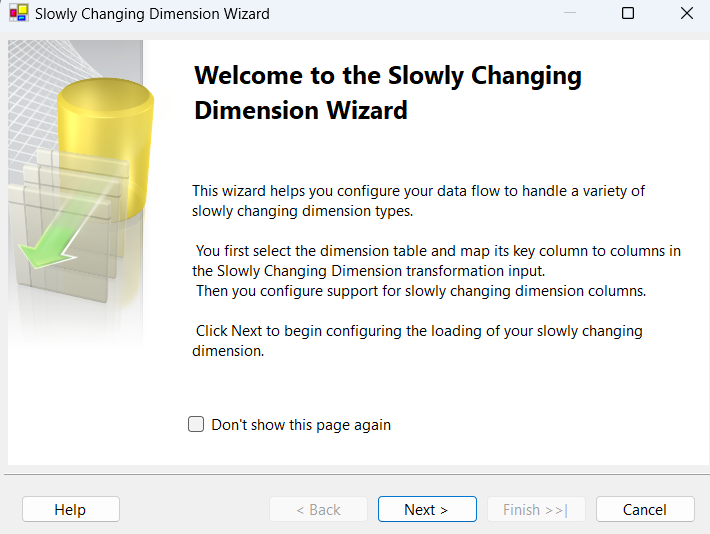
**5. Kéo rê Slowly Changing Dimension vào design surface. Đặt tên nó lại thành DST -Slowly Changing Dimension DimCustomers và connect SRC -> DST.**

****

****

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

**7. Slowly Changing Dimension Wizard xuất hiện**

****

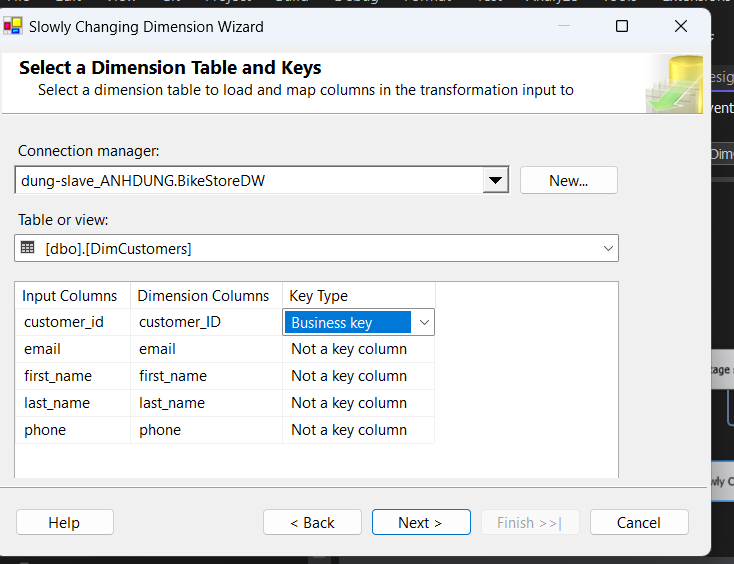
**Click Next.**

**8. Chọn bảng DimCustomers ở CSDL BikeStoreDW. Ở mục connection manager chọn**

**CSDL BikeStoreDW.**

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

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

****

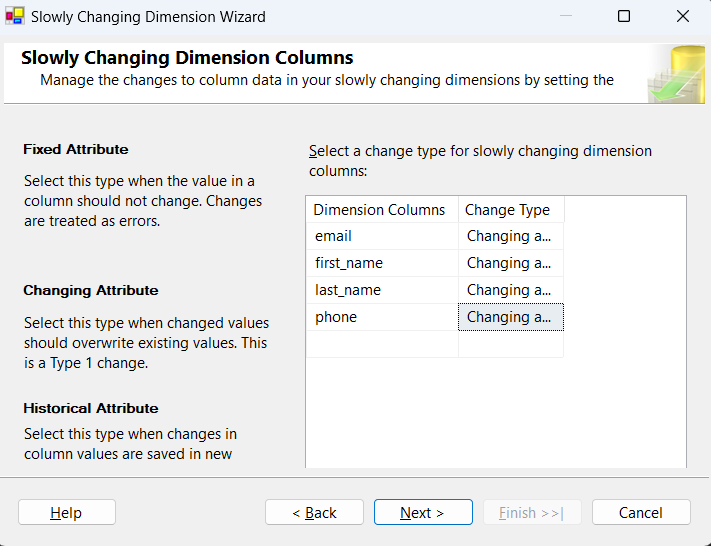
**Click Next.**

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

**- Fixed à Không thay đổi**

**- Changing à Type 1 (Update)**

**- Historical à Type 2 (thêm dòng mới, dòng cũ inactive)**

****

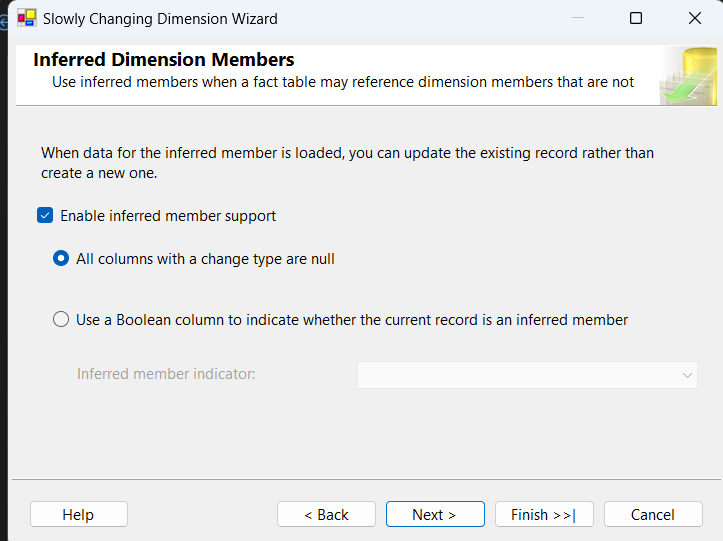
**Cấu hình tất cả non-business key là changing attribute.**

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

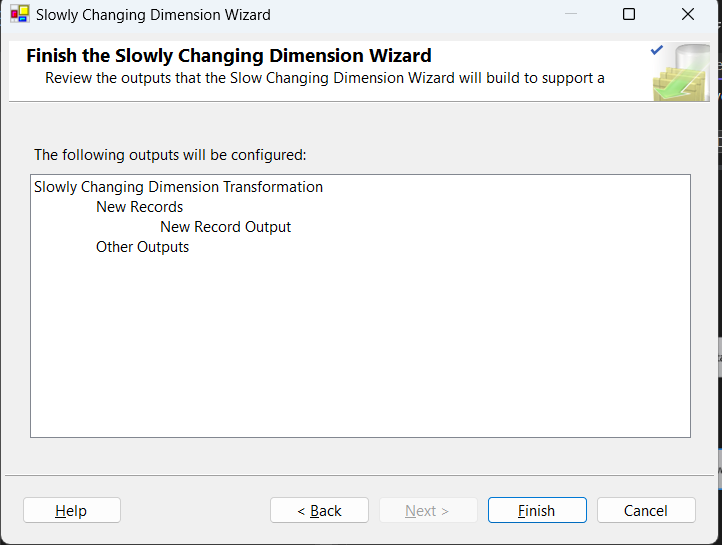
**Next.**

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

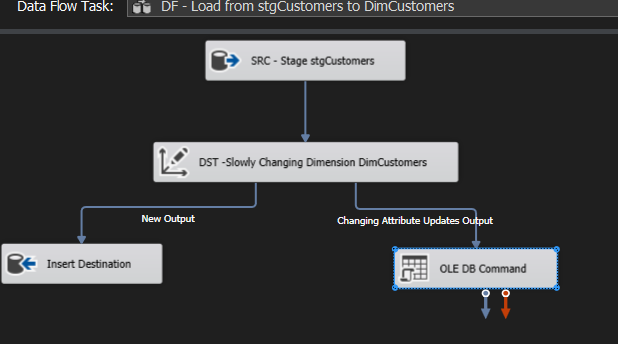
**Next.**

****

**14. Xuất hiện màn hình Finish**

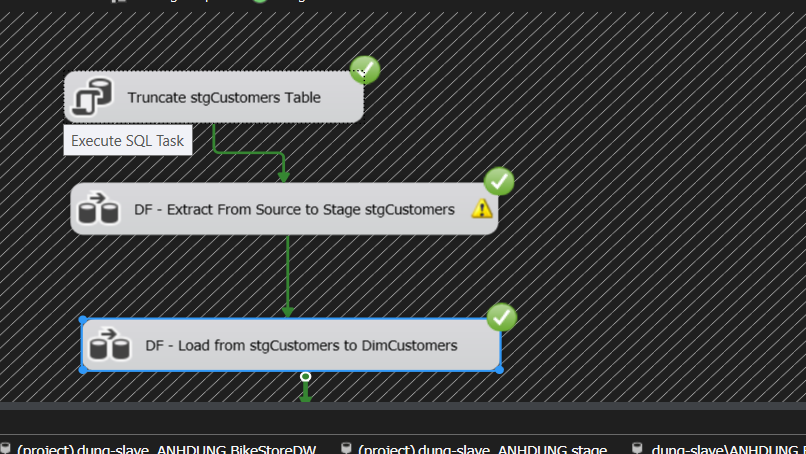
****

**Click Finish.**

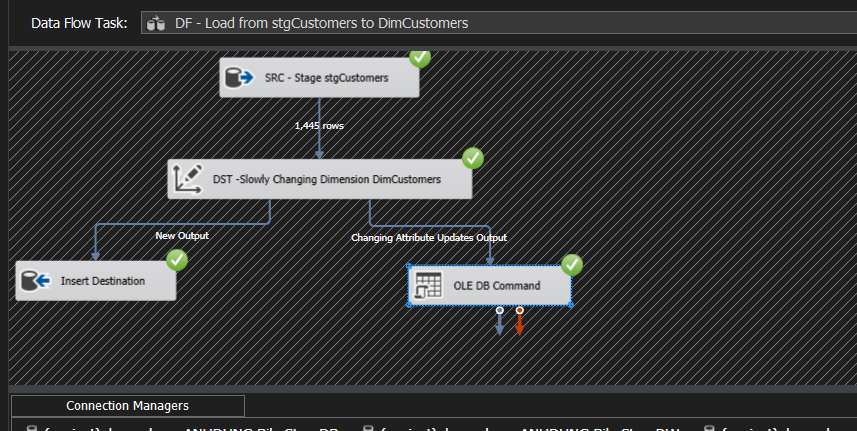
****

**15. Bấm F5 để chạy. Nếu xảy ra lỗi, cố gắng đọc hiểu thông báo báo lỗi để tìm cách sửa.**

**Nếu không có lỗi bạn sẽ thấy màn hình như sau:**

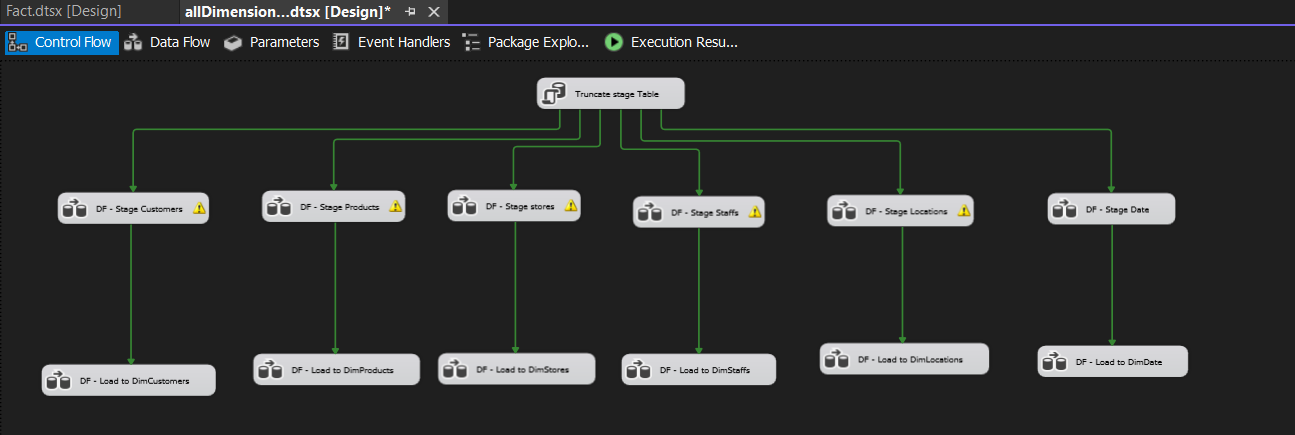
****

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

****

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

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

****

1. **Products**

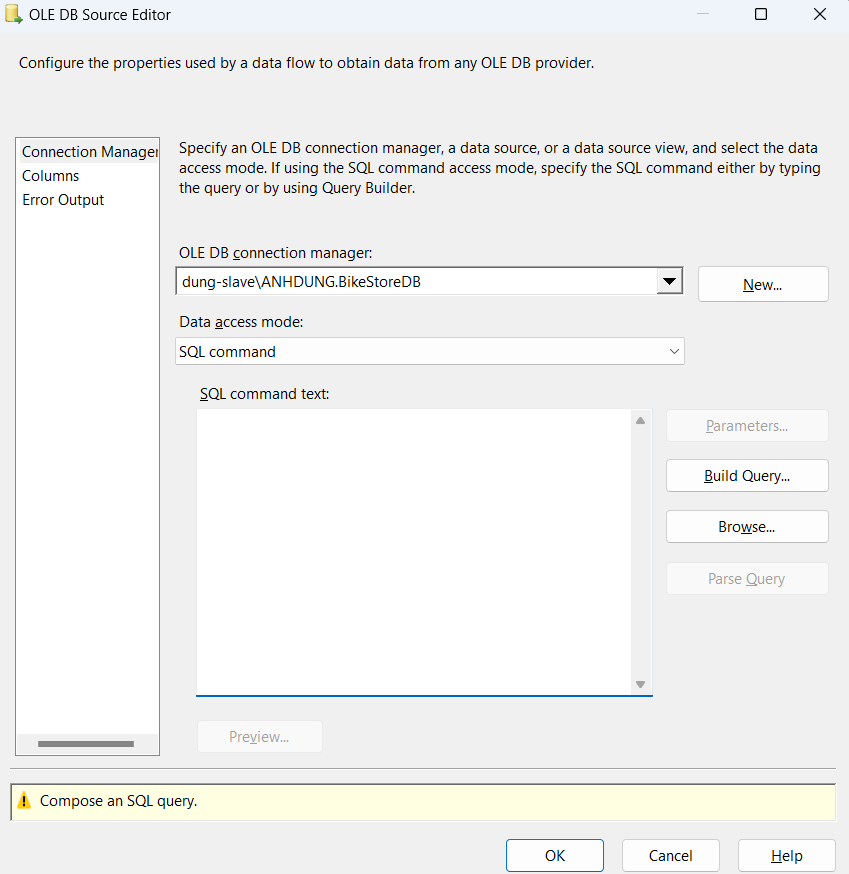
##### **Load dữ liệu từ Source vào Stage**

Ta cần các thuộc tính ProductID, ProductName,Moder\_year, list\_price, brand\_name,category\_name. Các thuộc tính này không thuộc cùng một bảng Product. Do vậy, ta phải viết truy vấn để lấy được các trường này. Double click vào DF – Stage Product, tạo một Source

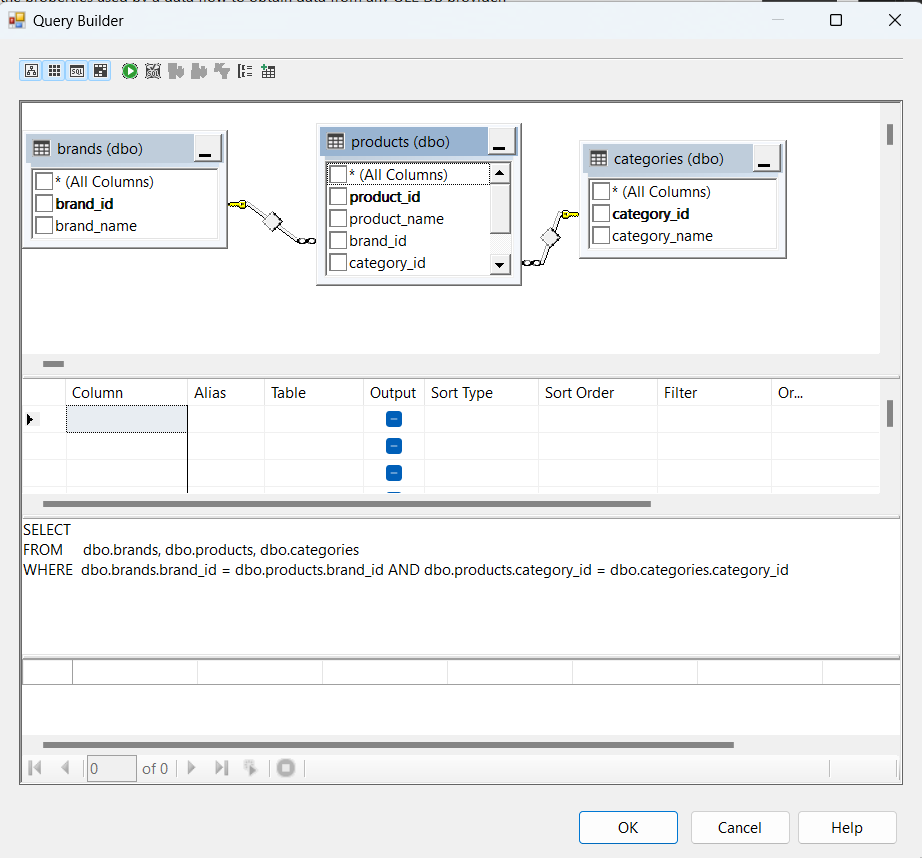
Assistant có tên SRC - Northwind Products. Double click vào nó để cấu hình. Ở màn hình

OLEDB Source Editor, mục Data access mode, chọn SQL Command. Sau đó click Query

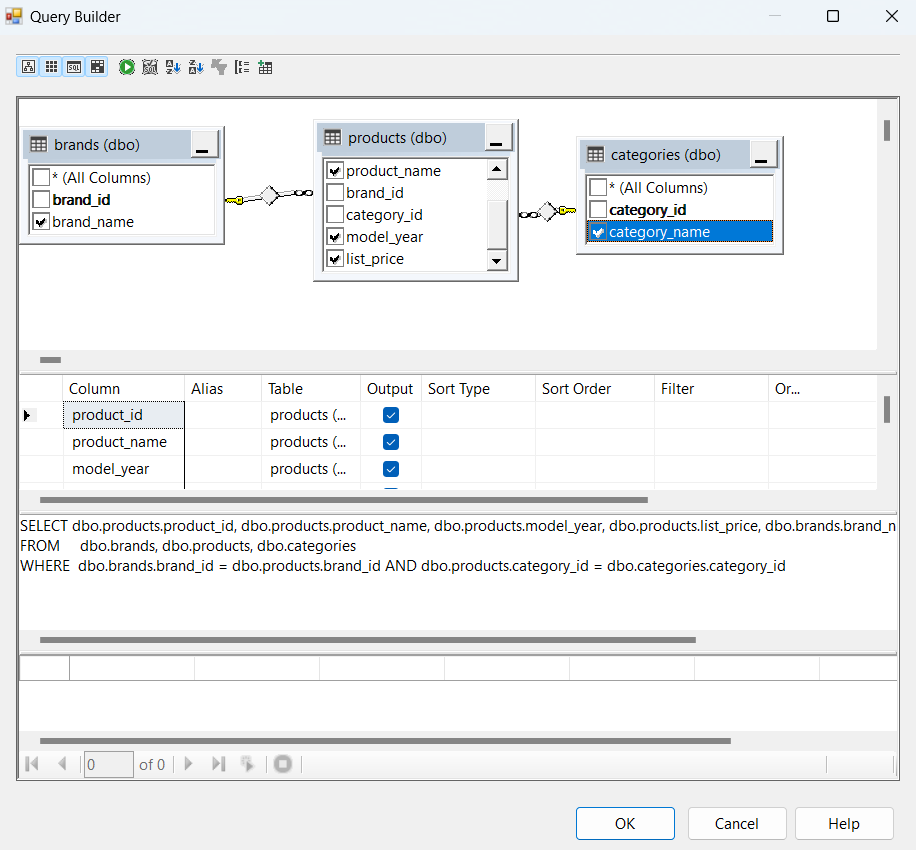
Builder …

****

Click biểu tượng Add Table

****

Chọn các thuộc tính cần cho bảng DimProduct

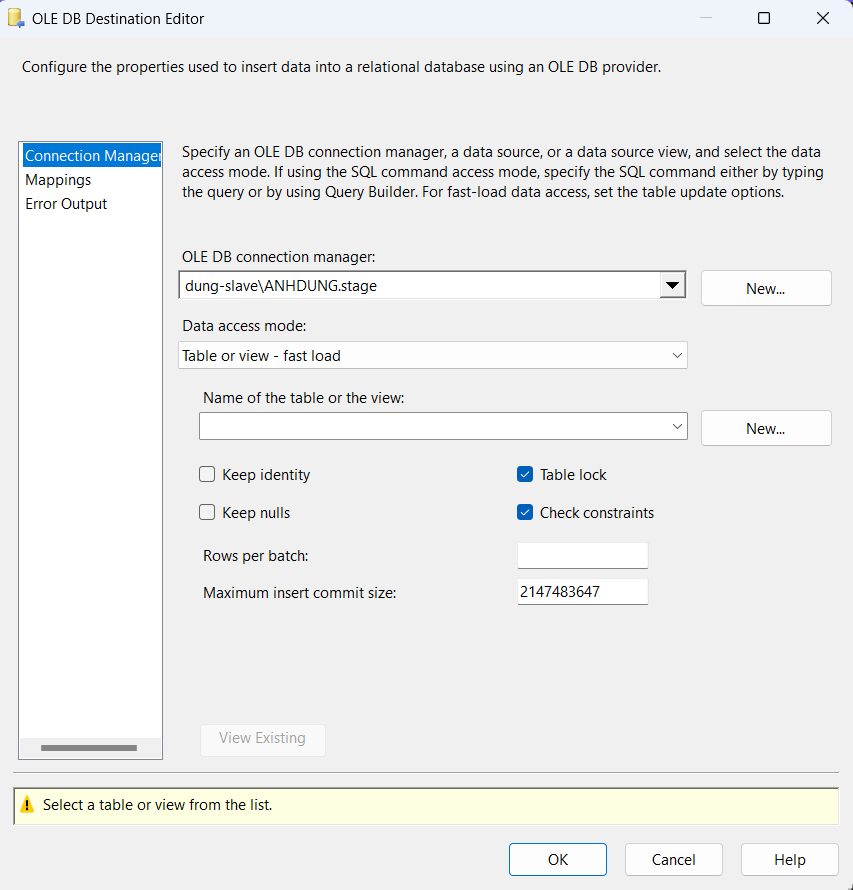
****

Tạo một Destination Assistant có tên DST - Stage stgProducts. Kết nối SRC -

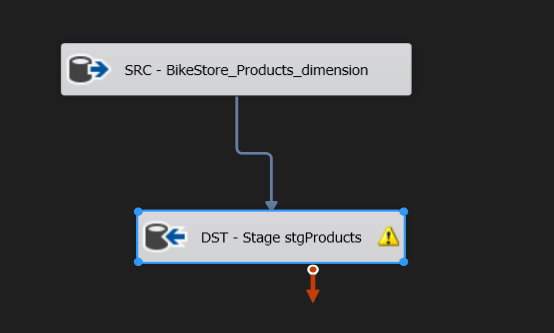
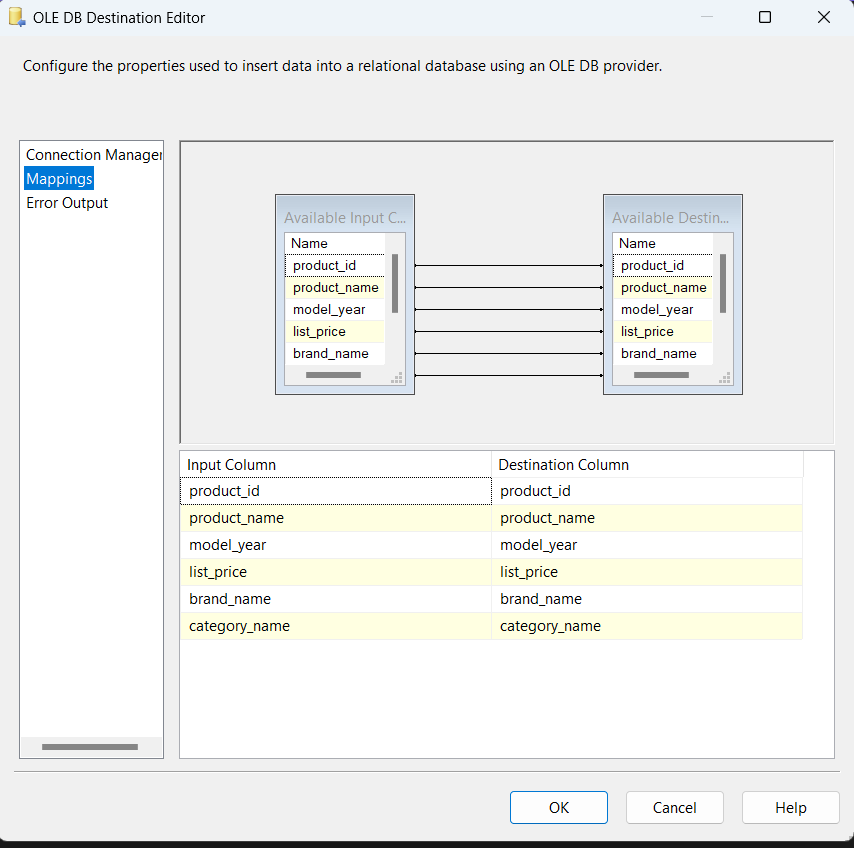
Northwind Products với DST - Stage stgNorthwindProducts. Double click vào DST - Stage

stgProducts để cấu hình. Click New…để tạo bảng. Đặt tên nó là

stgProducts. Ta có một số màn hình kết quả sau:

****

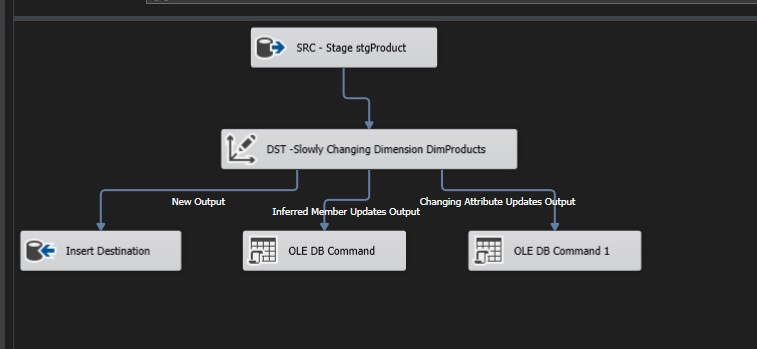
****

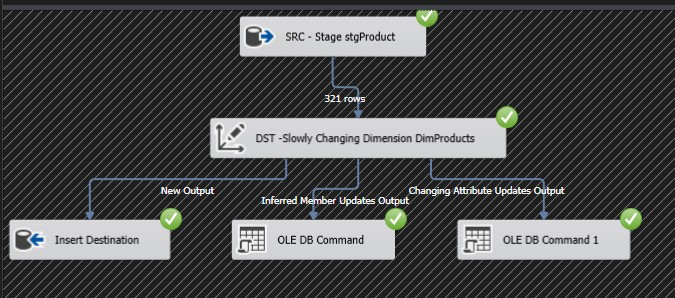
****

##### **Load dữ liệu từ Stage vào Dimension**

**- Double-click vào DF - Load to DimProduct task để mở data flow design**

**Surface và thực hiện các bước cấu hình tương tự mục 3.3.**

**Kết quả thu được**

****

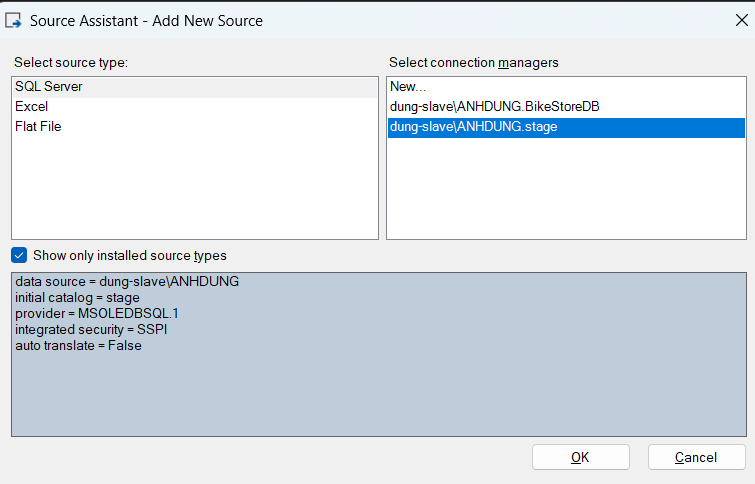
1. **DimStore**

##### **Load dữ liệu từ Source vào Stage**

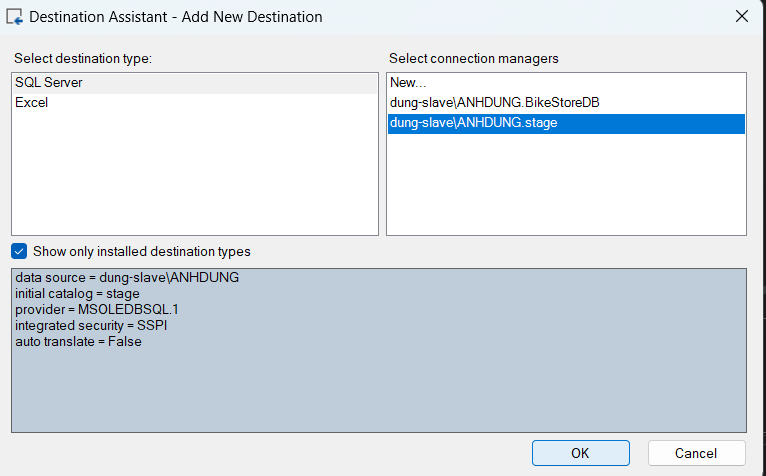
- Double-click **DF -  Stage Stores** task để mở nó trong **Data Flow design surface:**

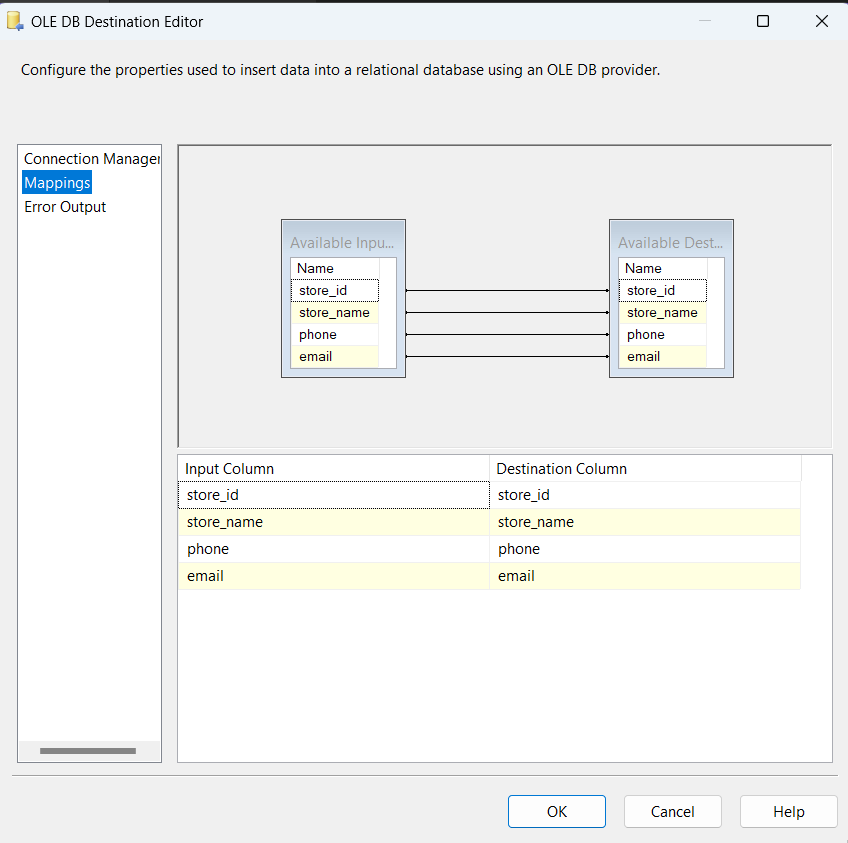
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

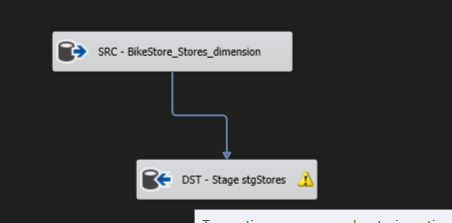


+ Destination Assistant:





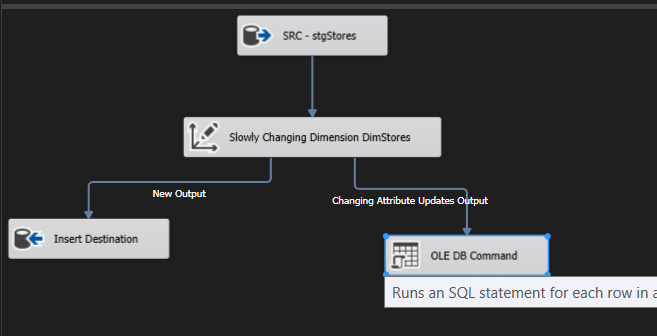
+ Sau khi cấu hình xong ta được:



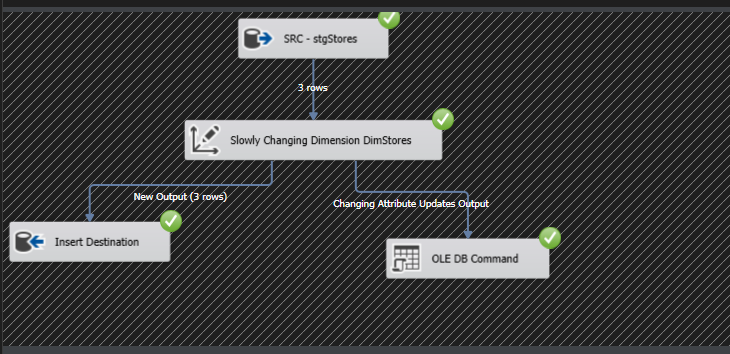
##### **Load dữ liệu từ Stage vào Dimension**

- Double-click vào **DF - Load to DimProduct** task để mở data flow design

Surface và thực hiện các bước cấu hình tương tự mục 3.3.

z

Kết quả:



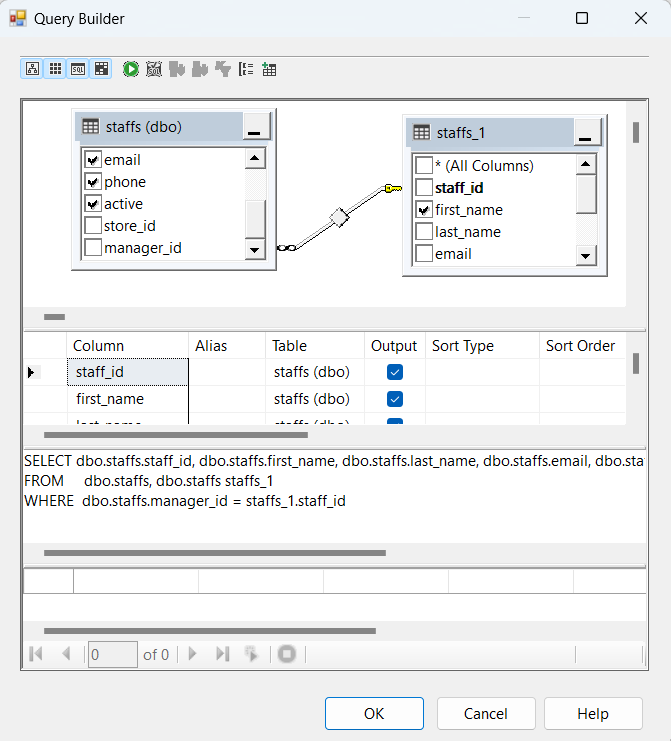
1. **Staffs**

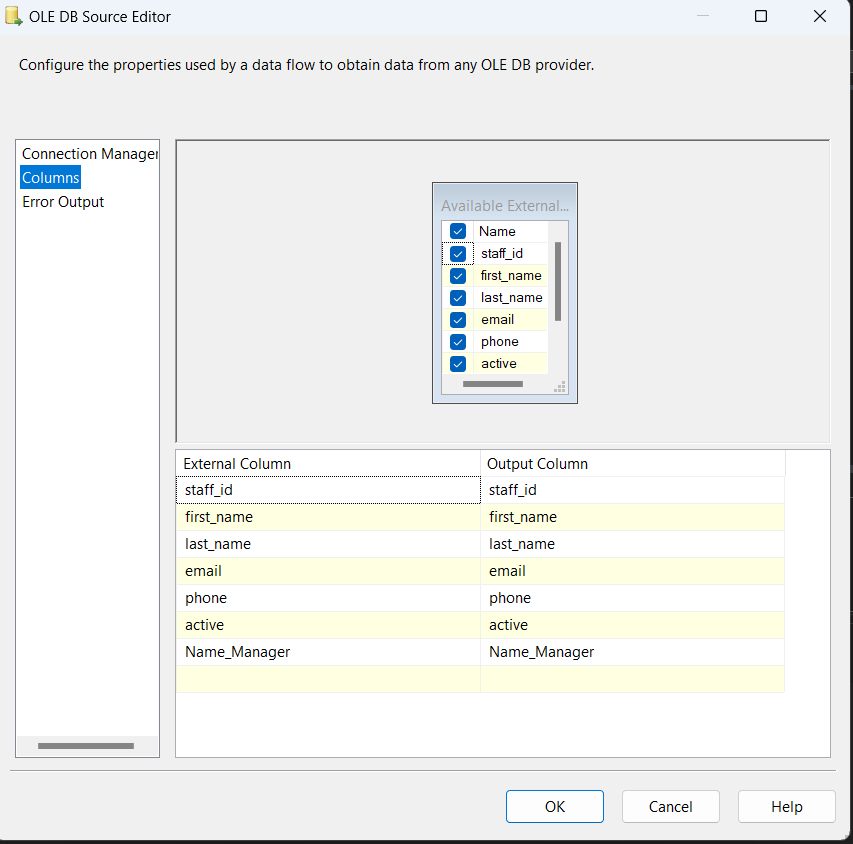
##### **Load dữ liệu từ Source vào Stage**

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

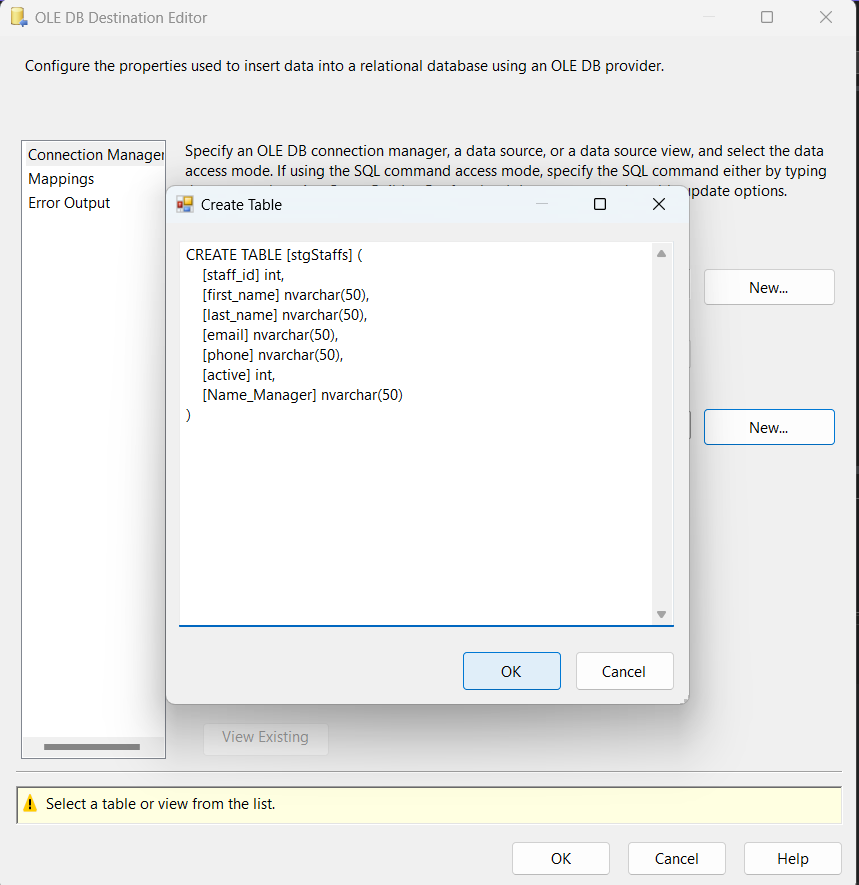
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

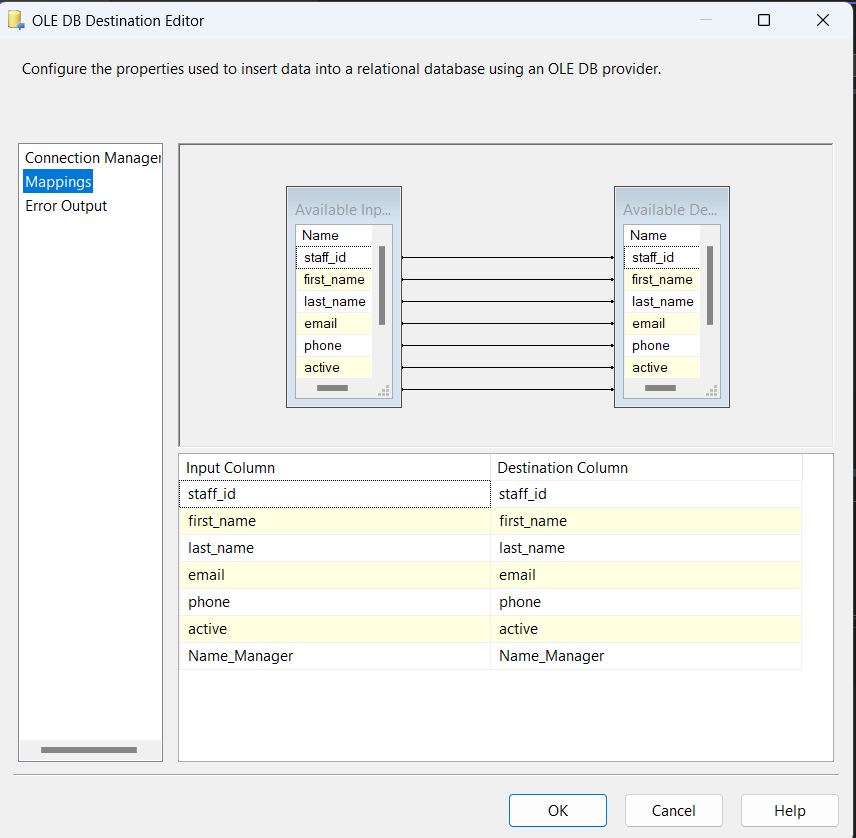
+ Source Assistant:

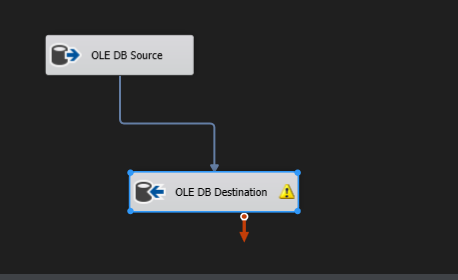




+ Destination Assistant:



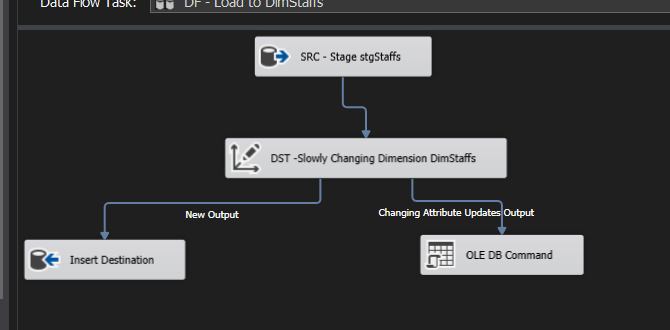




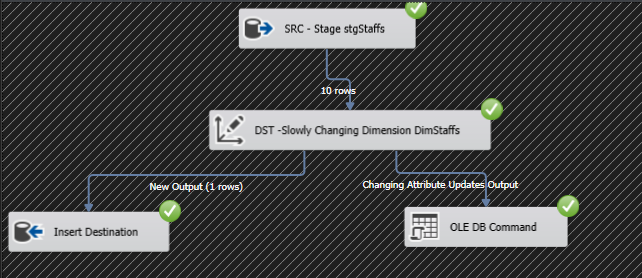
##### **Load dữ liệu từ Stage vào Dimension**

- Double-click vào **DF - Load to DimProduct** task để mở data flow design

Surface và thực hiện các bước cấu hình tương tự mục 3.3.



Kết quả:



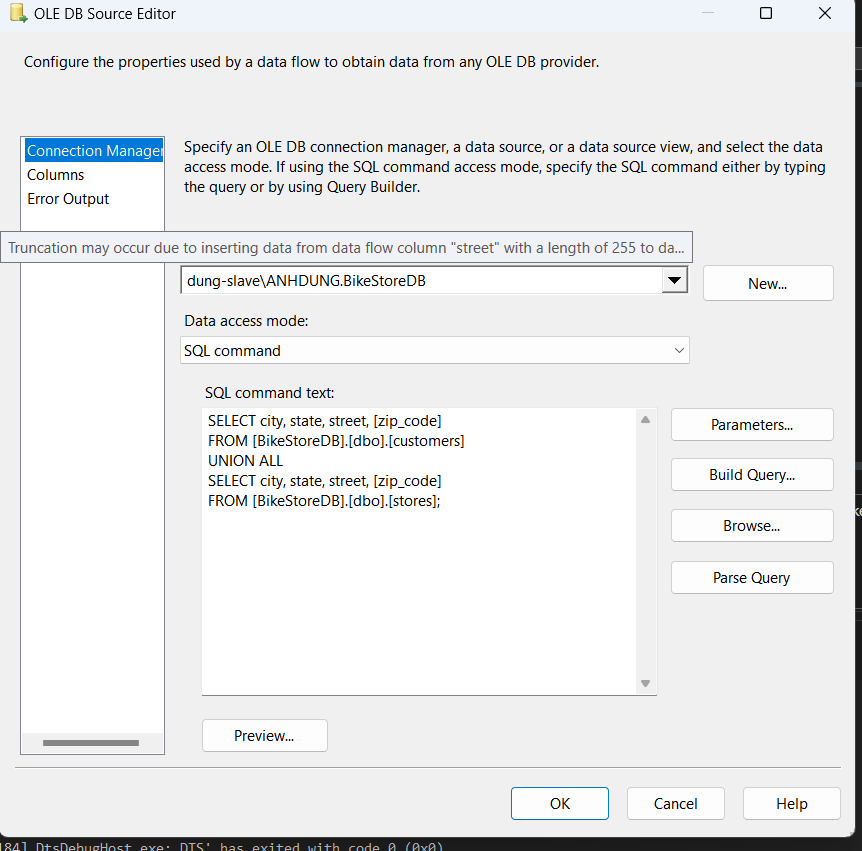
1. **Location**

##### **Load dữ liệu từ Source vào Stage**

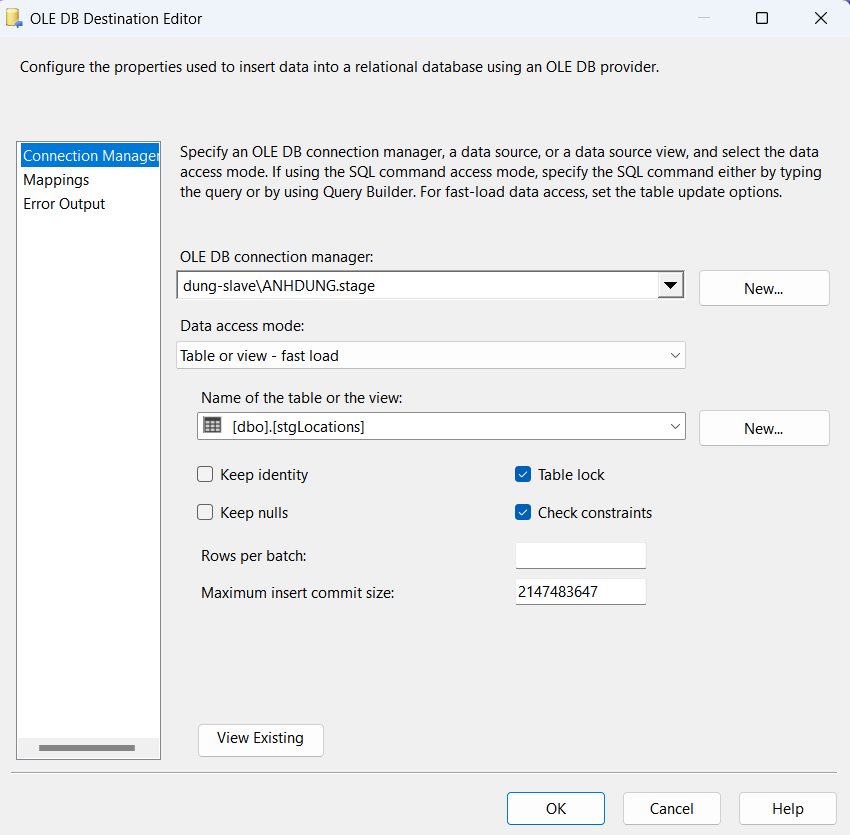
- Double-click **DF -  Stage location** task để mở nó trong **Data Flow design surface:**

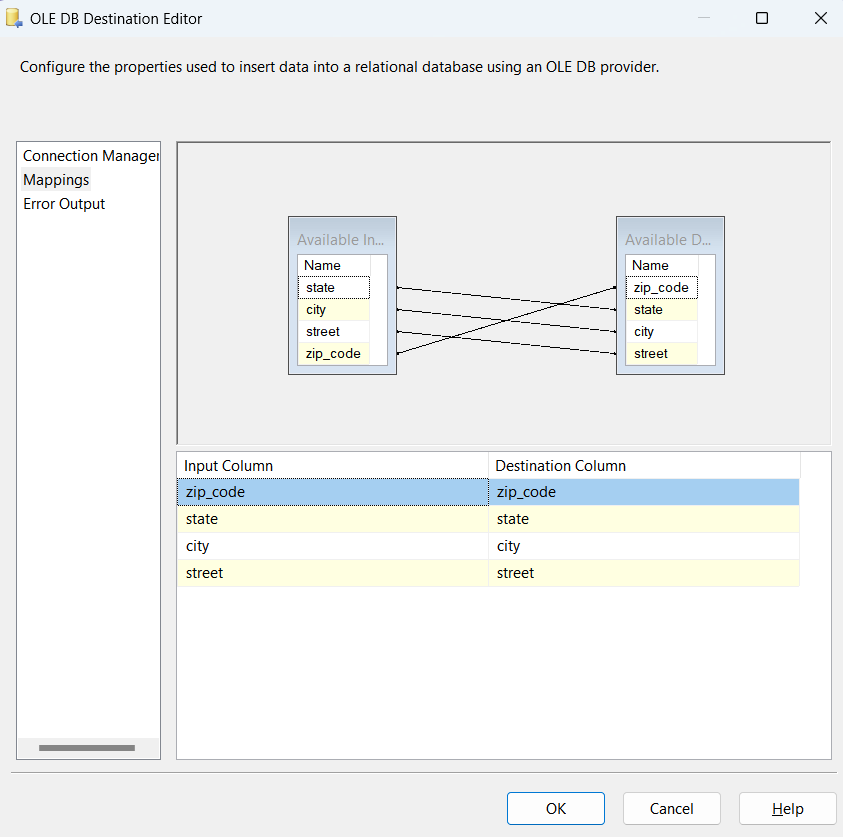
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

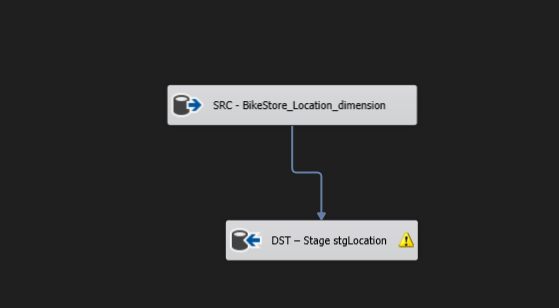


+ Destination Assistant:





Kết quả thu được:



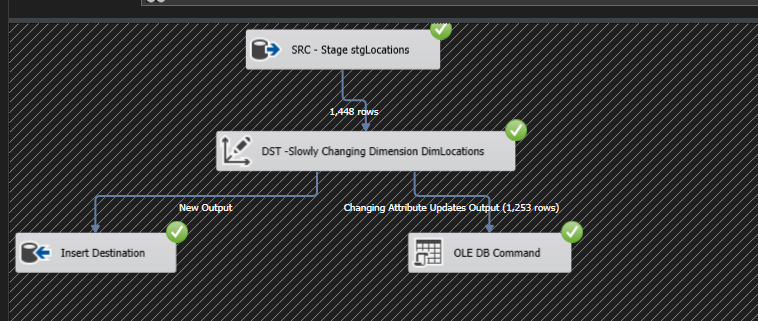
##### **Load dữ liệu từ Stage vào Dimension**

- Double-click vào **DF - Load to DimProduct** task để mở data flow design

Surface và thực hiện các bước cấu hình tương tự mục 3.3.



Kết quả:



1. **Date**

##### **Load dữ liệu từ Source vào Stage**

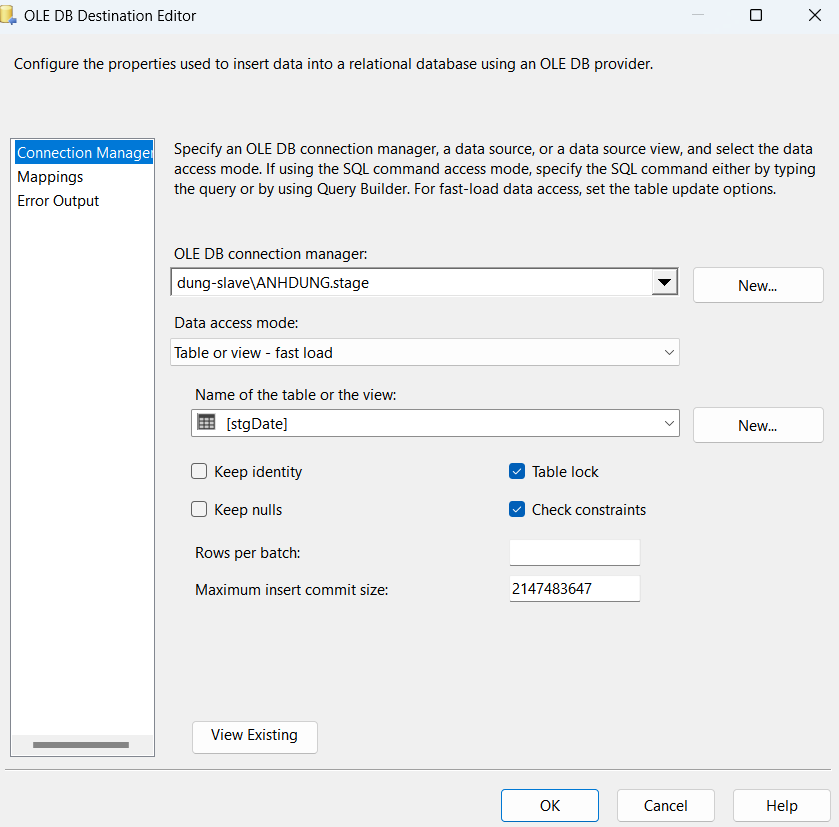
- Double-click **DF -  Stage date** task để mở nó trong **Data Flow design surface:**

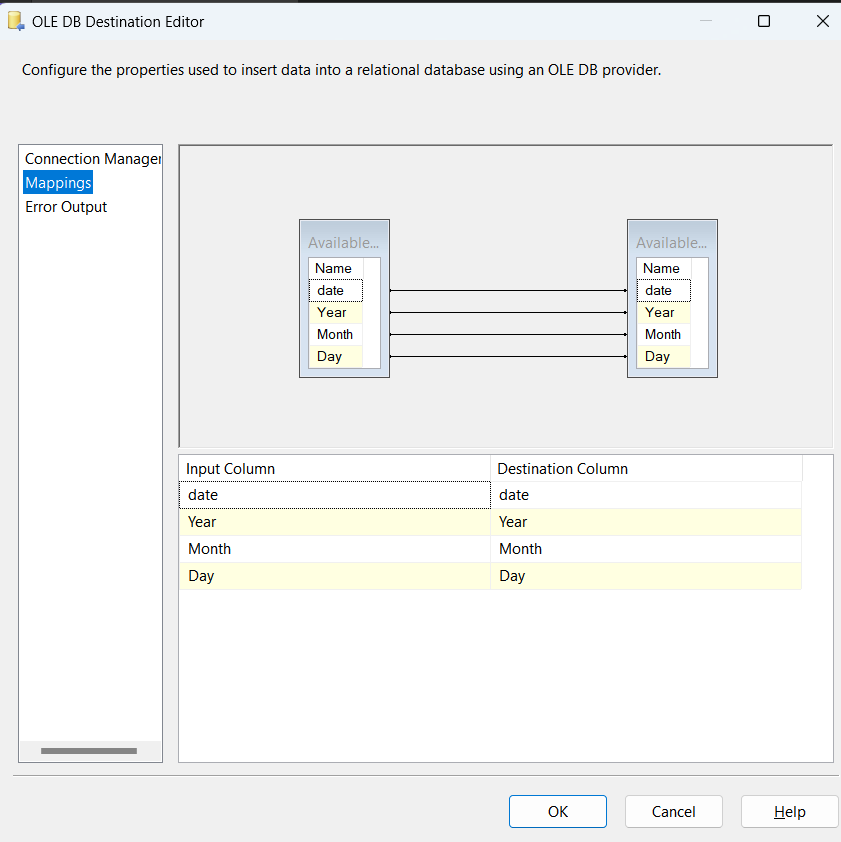
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

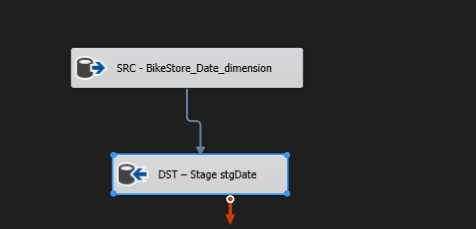


+ Destination Assistant:





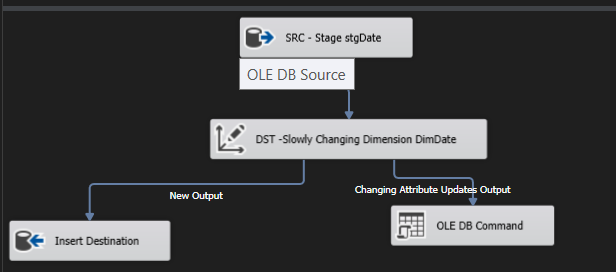
Kết quả thu được:



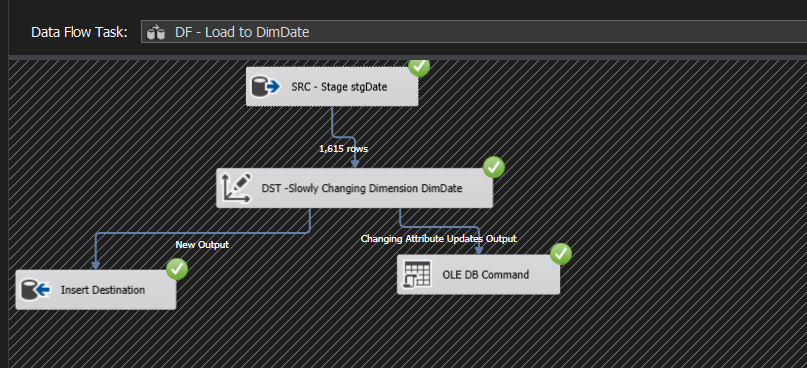
##### **Load dữ liệu từ Stage vào Dimension**

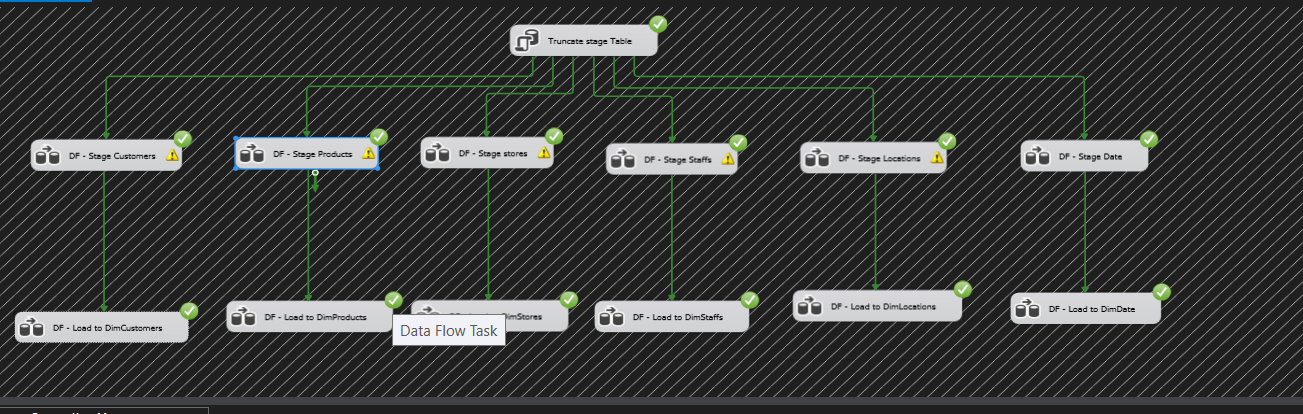
- Double-click vào **DF - Load to DimDate** task để mở data flow design

Surface và thực hiện các bước cấu hình tương tự mục 3.3



Kết quả:



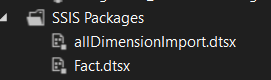


## Import dữ liệu vào bảng các bảng fact

#### 3.4.1. Fact Sales

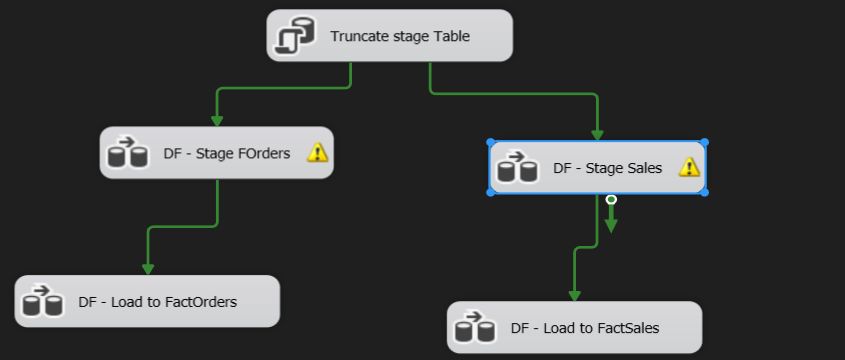
##### 3.4.1.1. Tạo Package

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

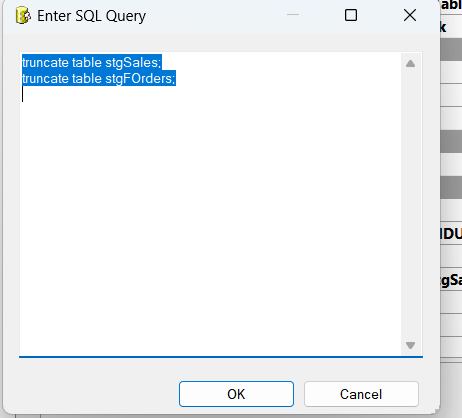


##### 3.4.1.2. Thiết lập Control Flow

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



- Thiết lập truncate table SQL task

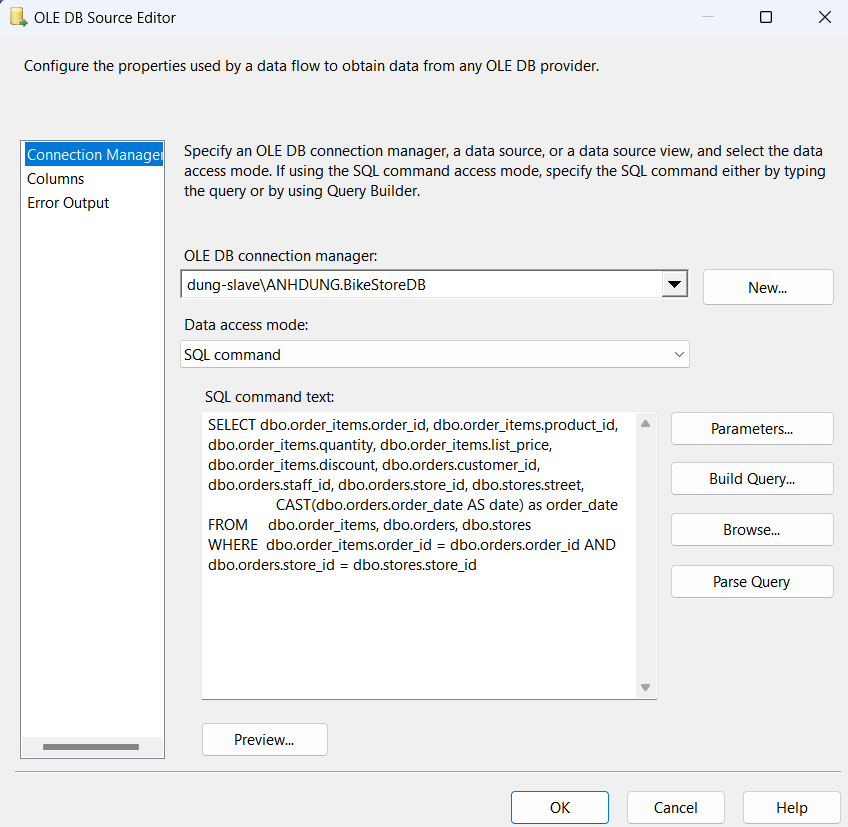


##### 3.4.1.3. Load dữ liệu từ Source vào Stage

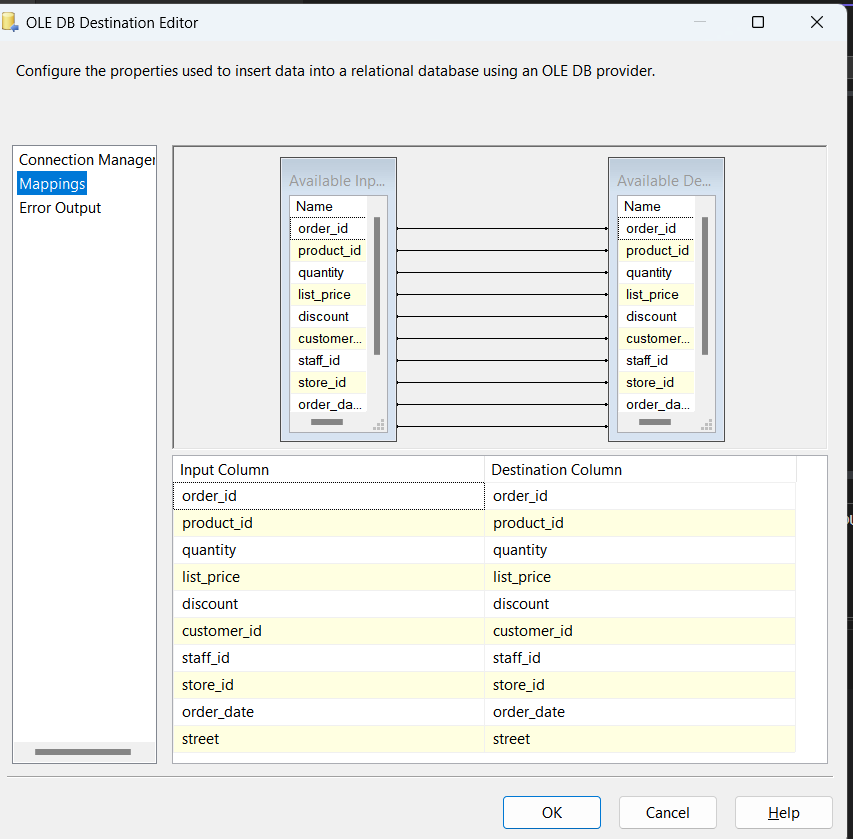
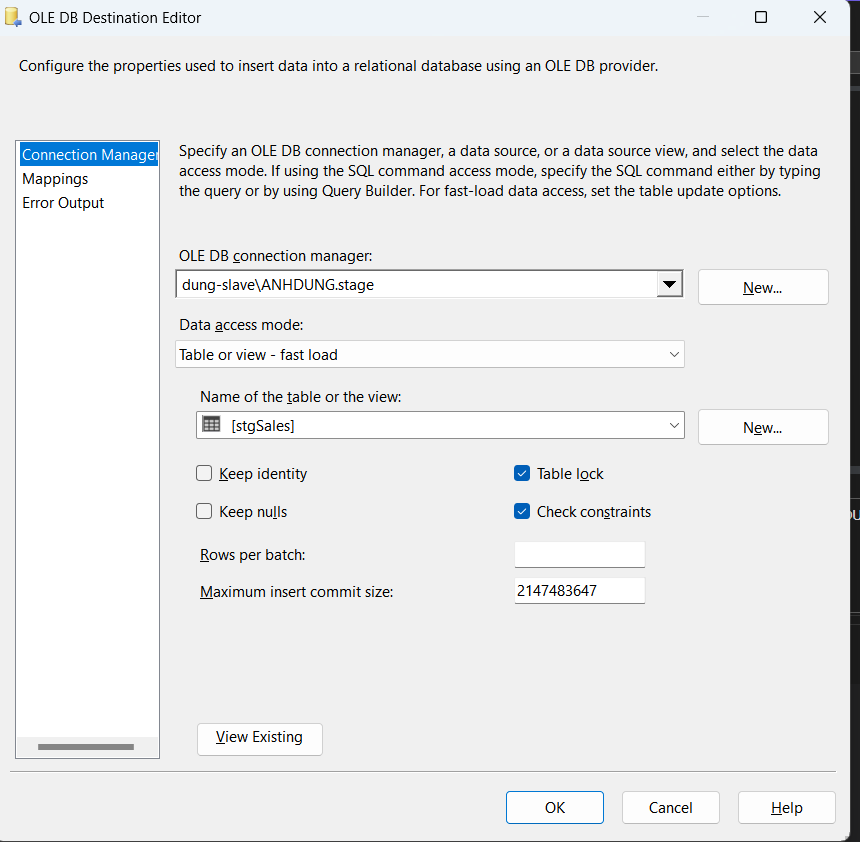
- Double-click **DF - Stage Sales** task để mở nó trong **Data Flow design surface:**

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

+ Source Assistant:



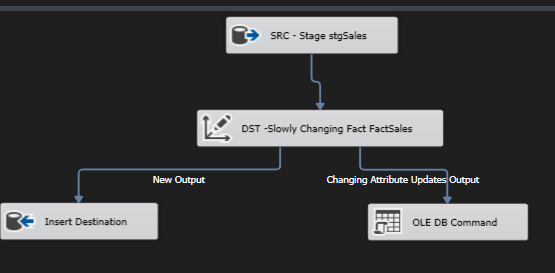
+ Destination Assist



Kết quả thu được:

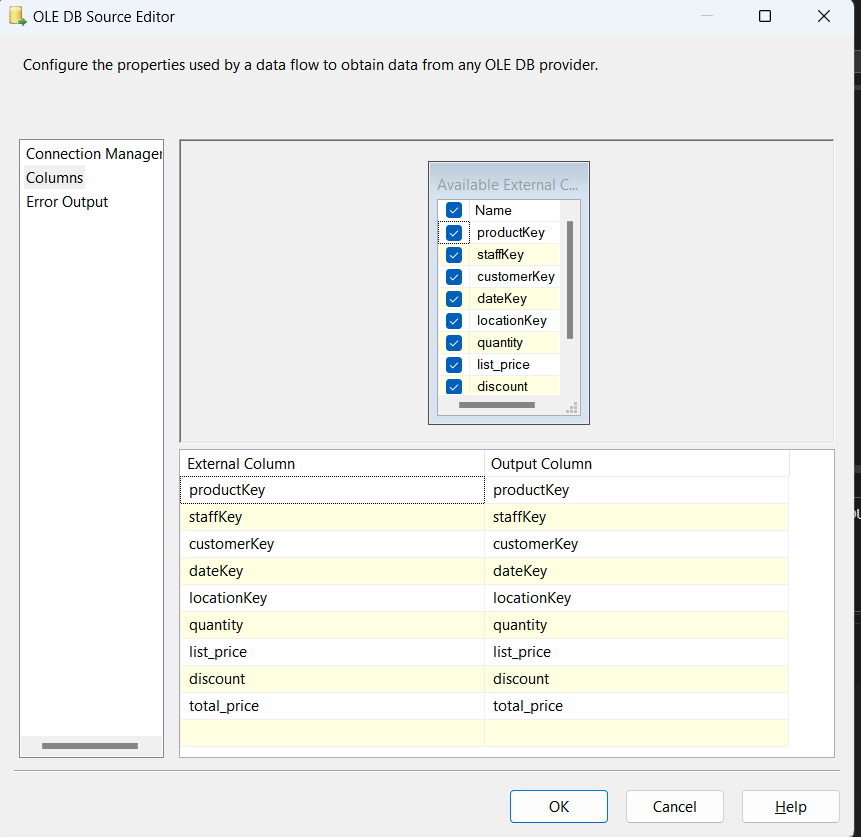
##### 3.4.1.4. Load dữ liệu từ Stage vào Fact

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

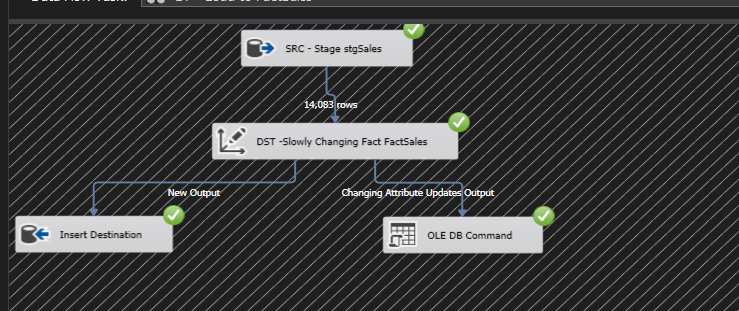


Cấu hình SRC - Stage stgSales





F5 ta thu được kết quả:



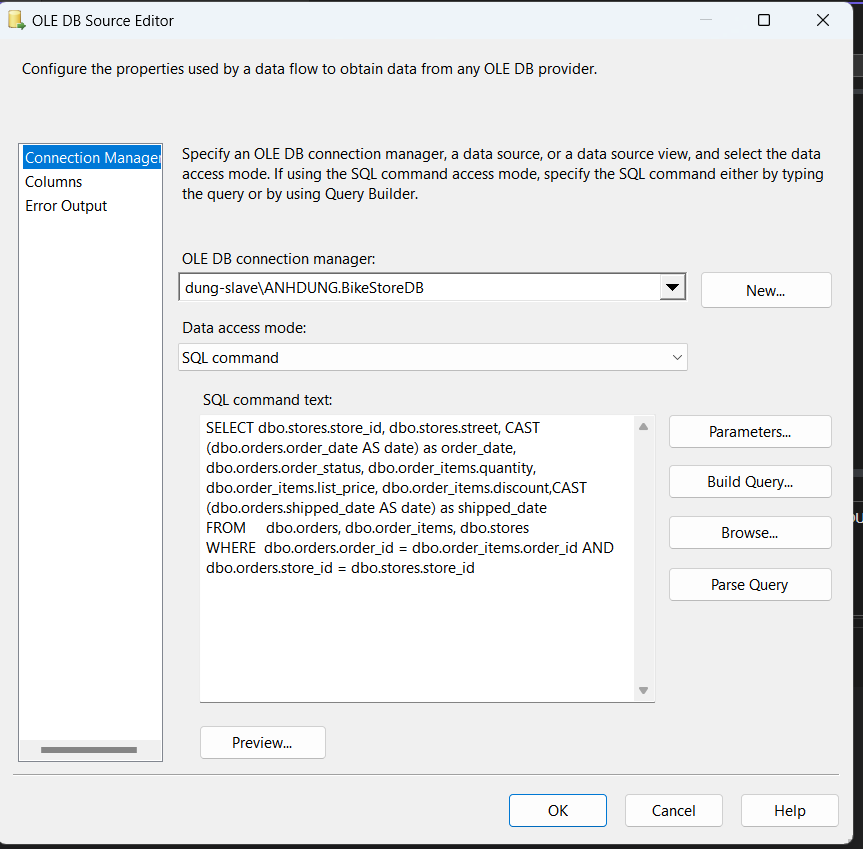
**3.4.2. Fact Orders**

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

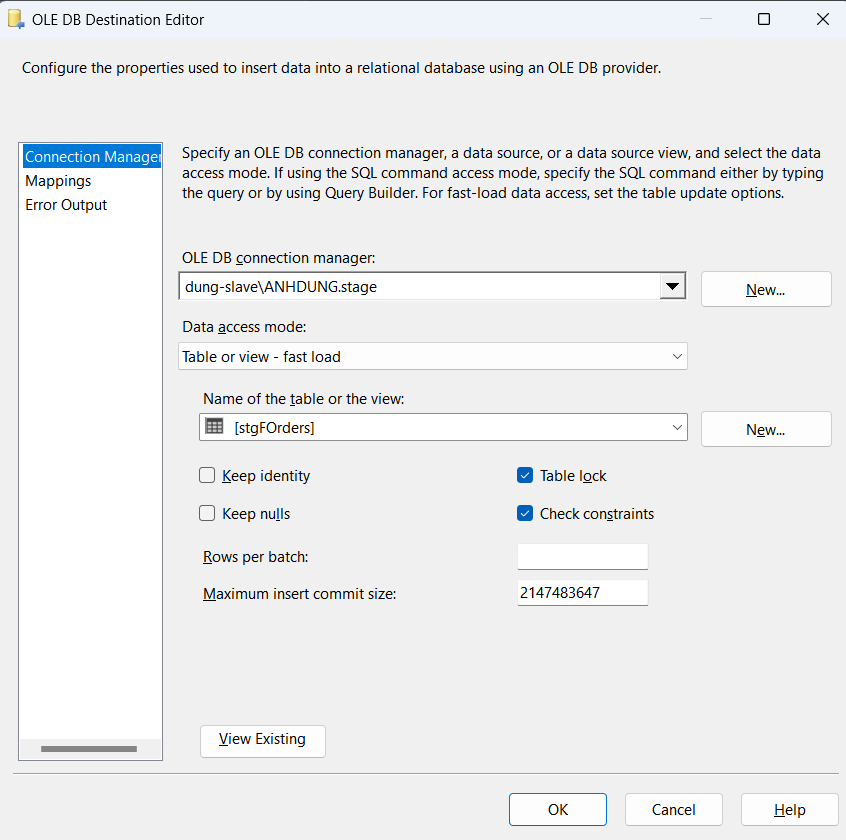
- Double-click **DF - Stage Orders** task để mở nó trong **Data Flow design surface:**

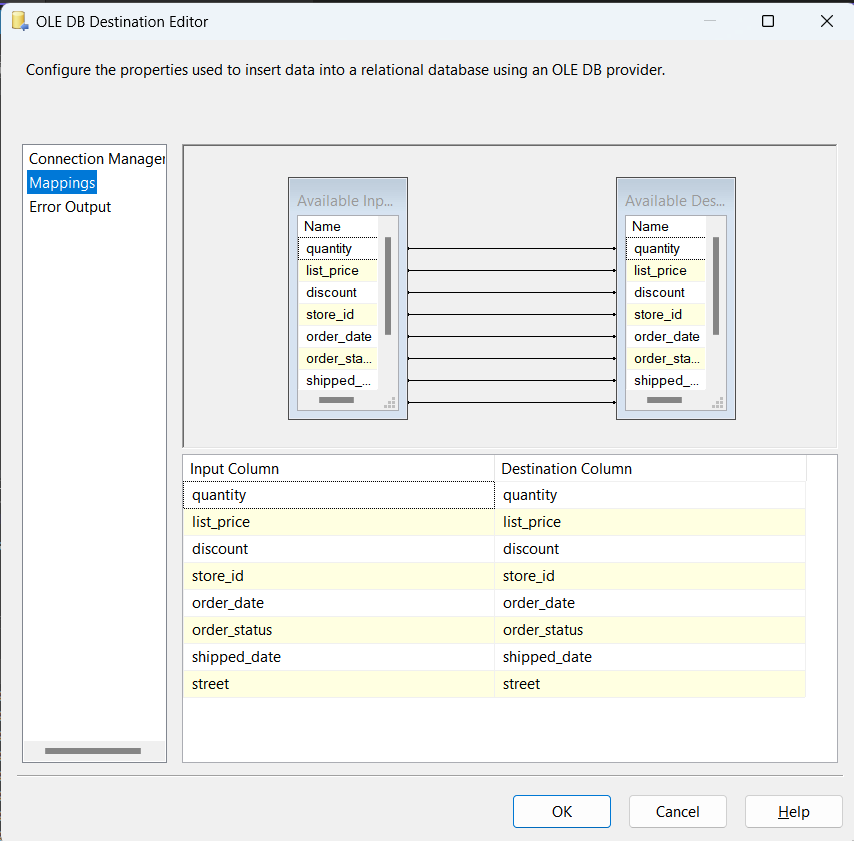
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

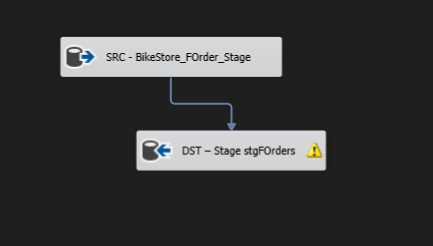


+ Destination Assist



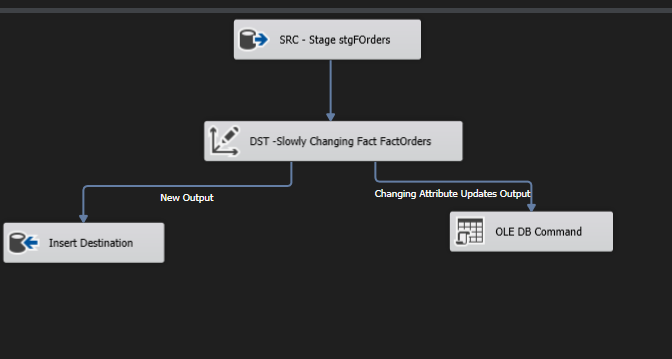


Kết quả thu được:

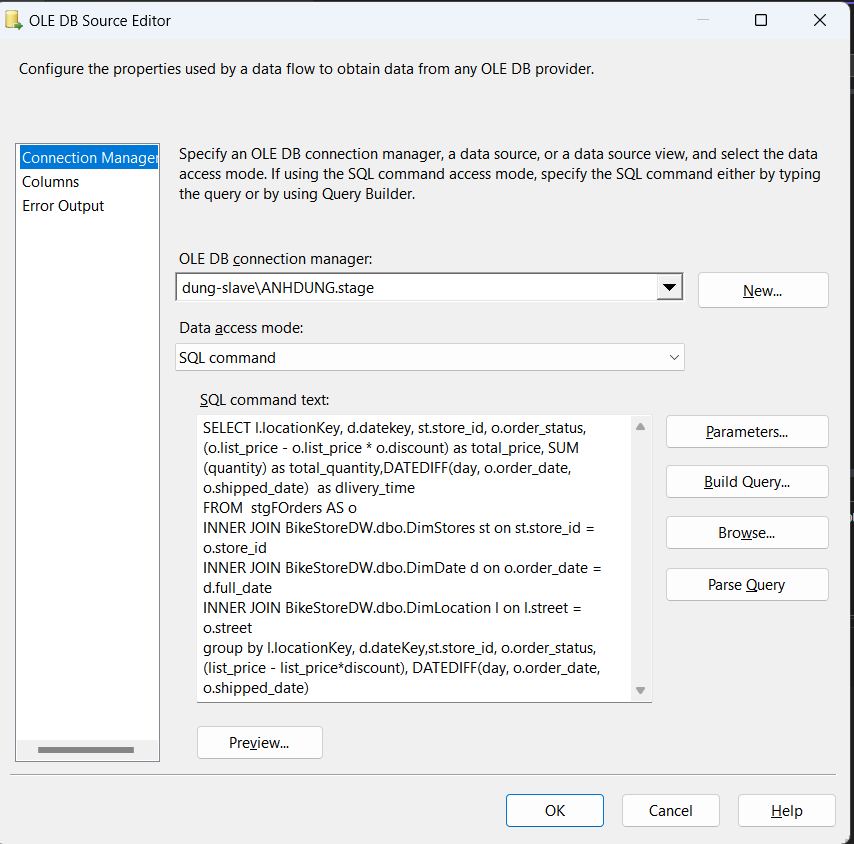


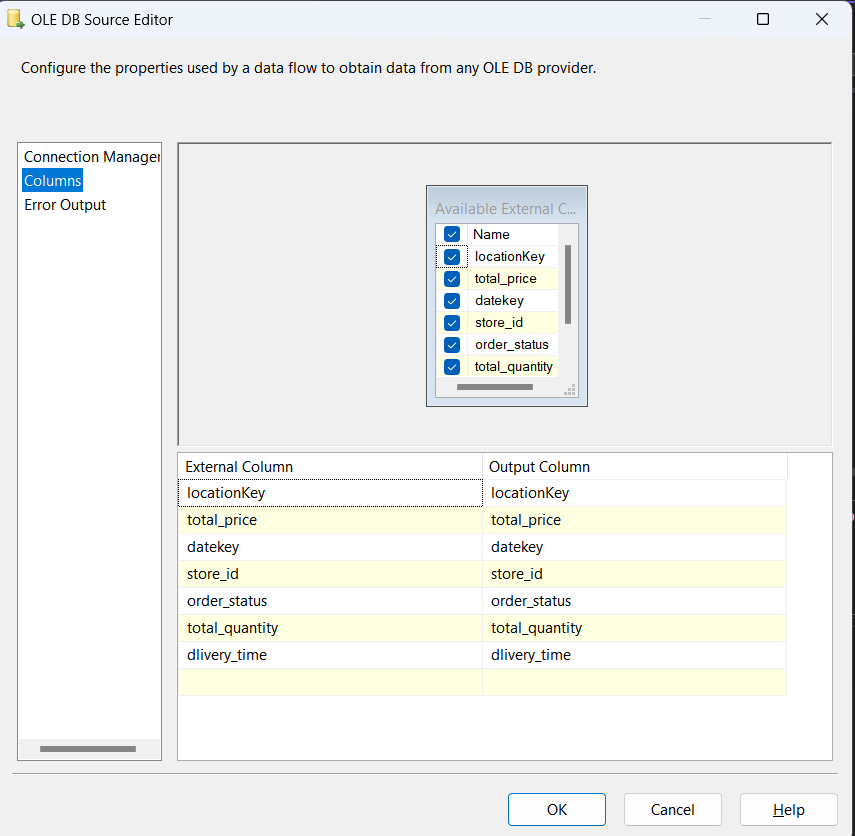
##### 3.4.2.2. Load dữ liệu từ Stage vào Fact

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

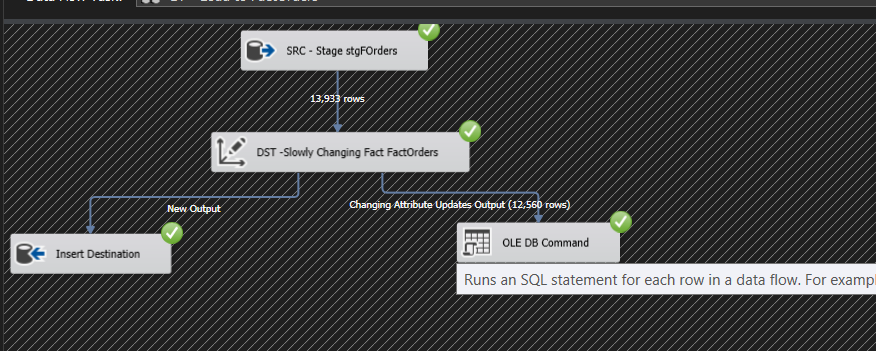


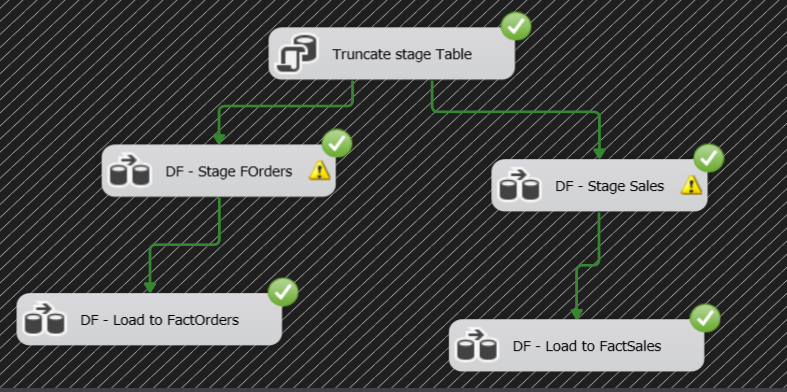
Cấu hình SRC - Stage stgFOrders:





F5 ta thu được



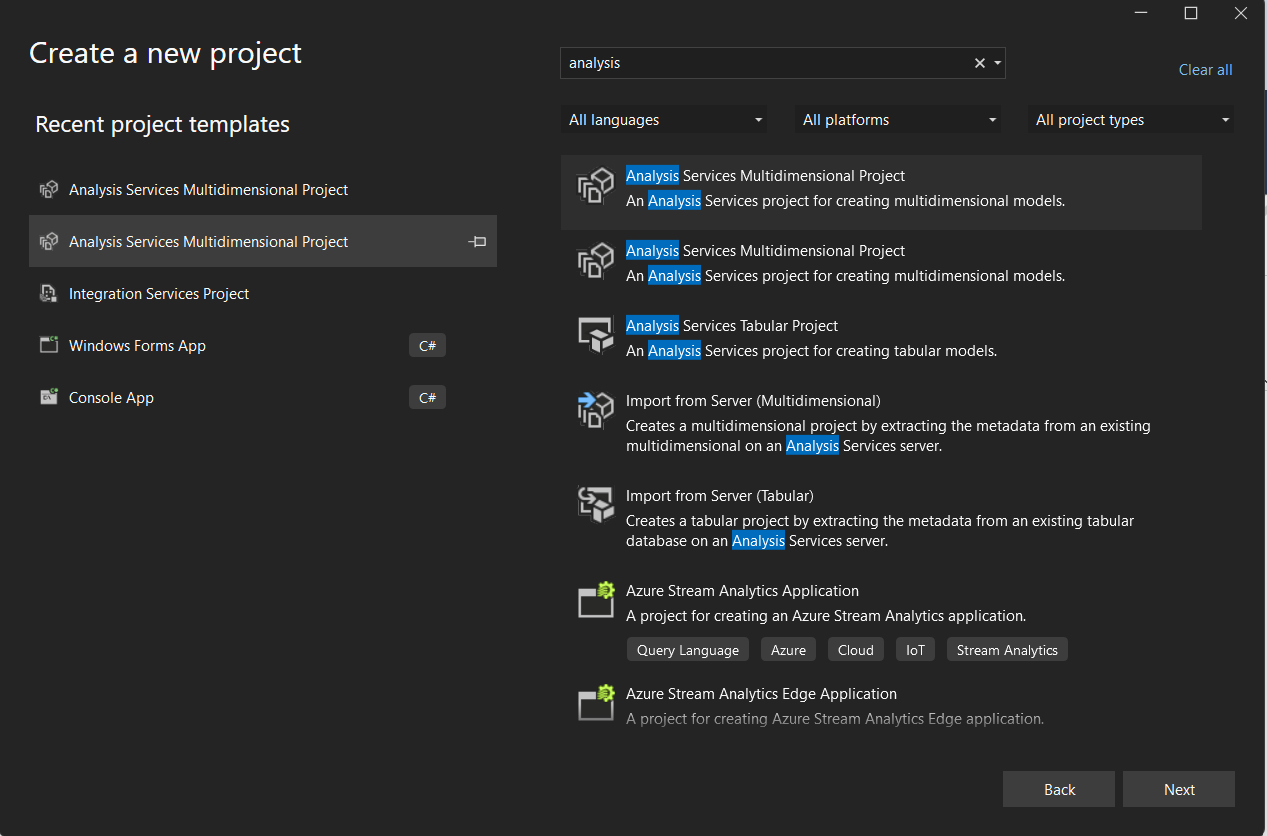


# CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)

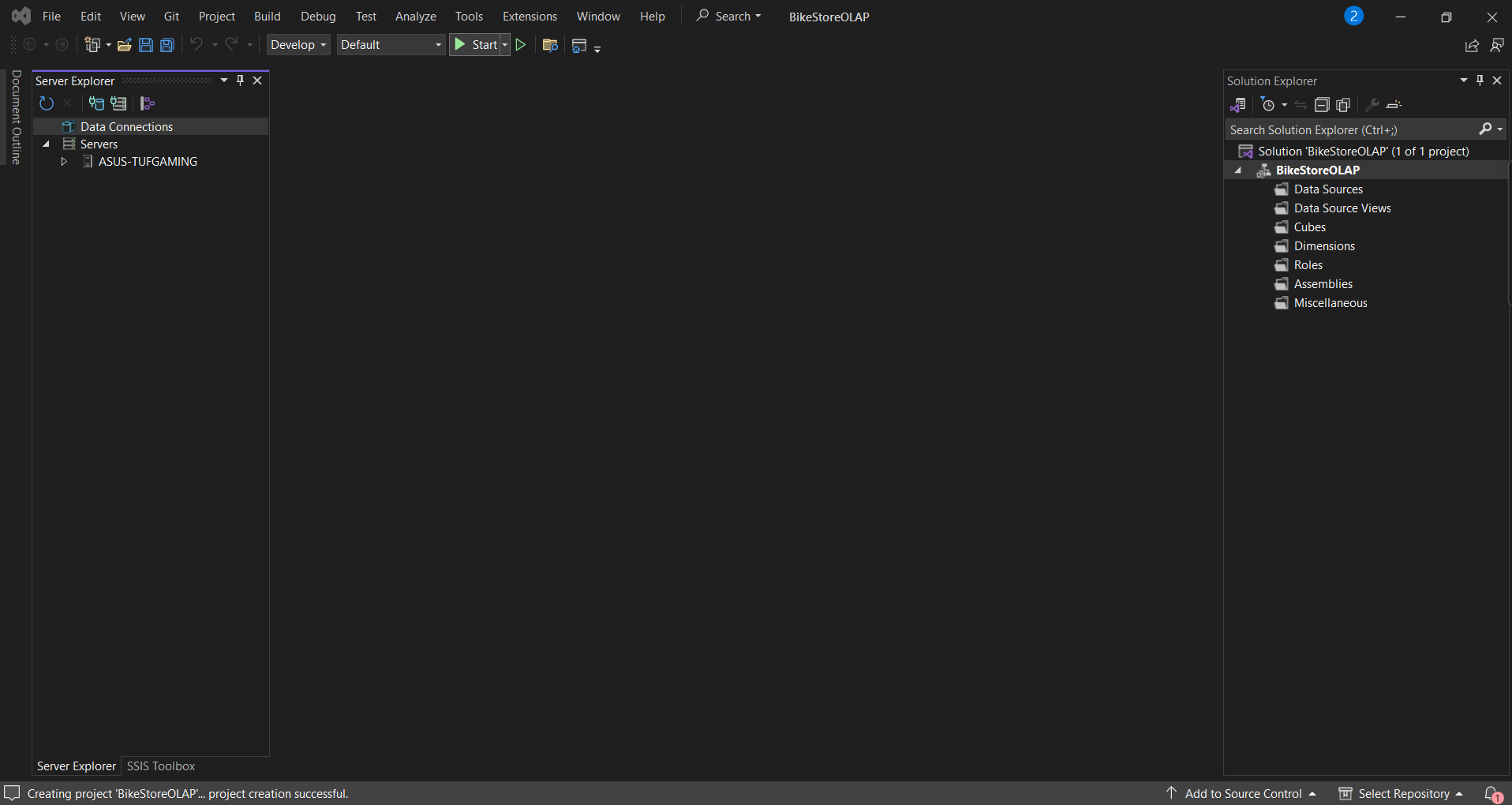
## 4.1. Quá trình xây dựng mô hình

1. **Tạo một Analysis Services Project**

Mở Visual Studio -> File -> New -> Project -> Analysis Services Multidimensional Project, đặt tên BikeStoreOLAP và click OK.

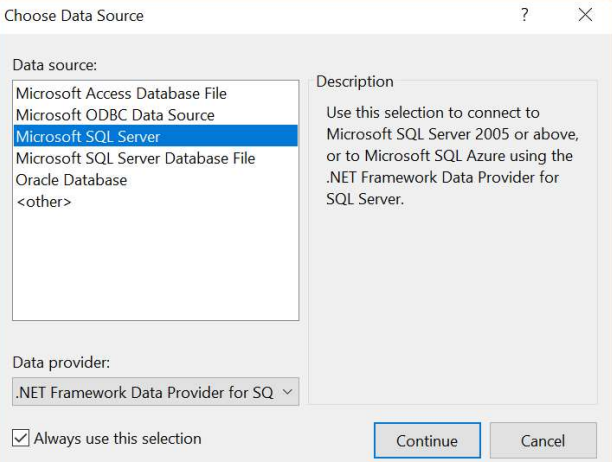


Giao diện của SSAS sau khi tạo một project mới như hình bên dưới:

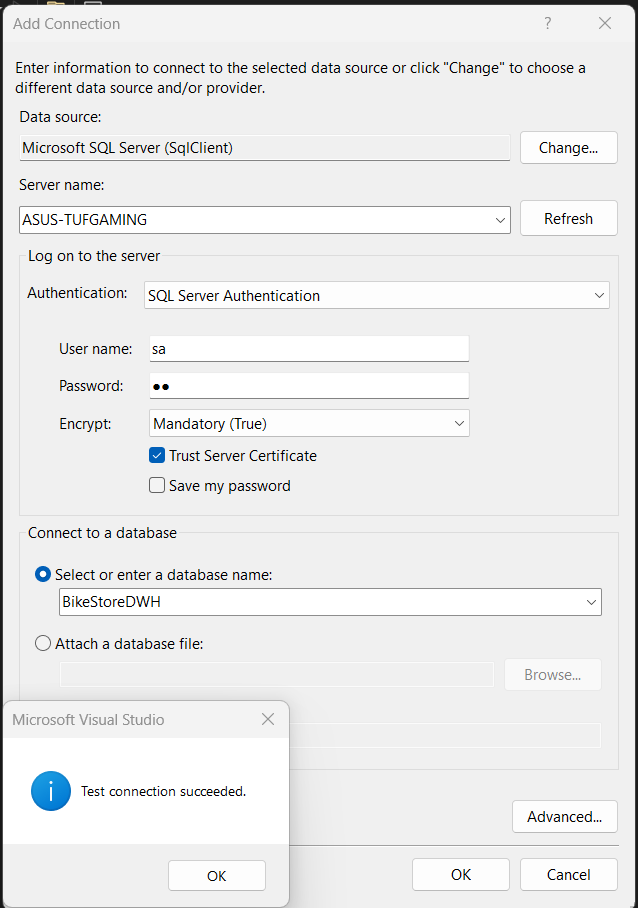


1. **Kết nối đến Analysis Services Server**

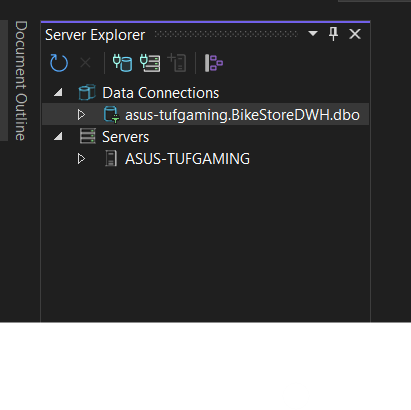
Right click vào “Data Connections” ở mục Server Explorer chọn “Add Connection …” sẽ xuất hiện dialog “Choose Data Source”, chọn “Microsoft SQL Server” và click “Continue



Xuất hiện dialog “Add Connection”, chọn “Data Source”, “Server name”, và “database name” phù hợp. Click OK.

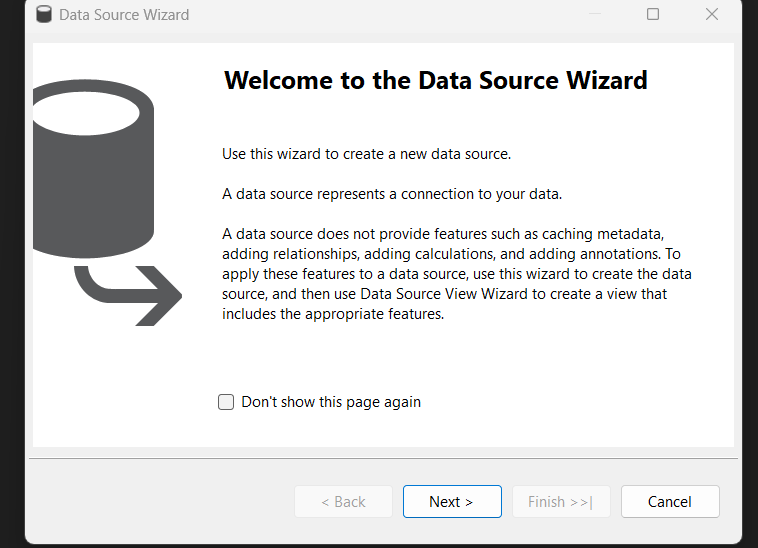


Trở lại màn hình chính, ta sẽ thấy thông tin về “Data Connections” được cập nhật và hiển thị như hình sau:

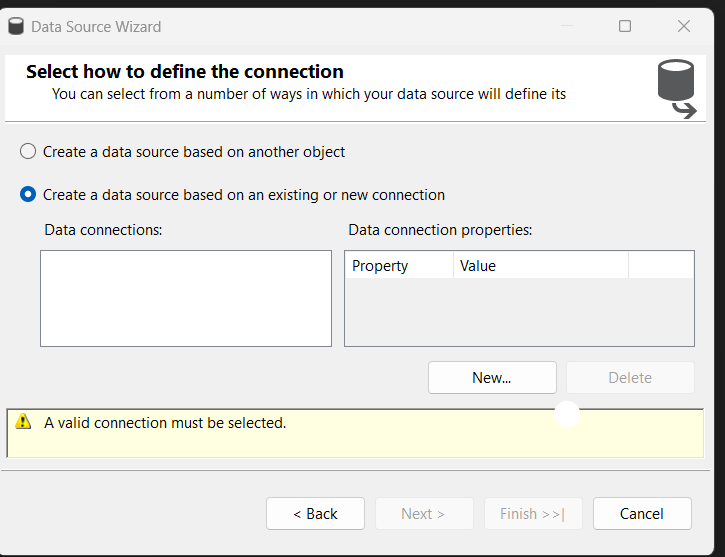


### Tạo Data Source

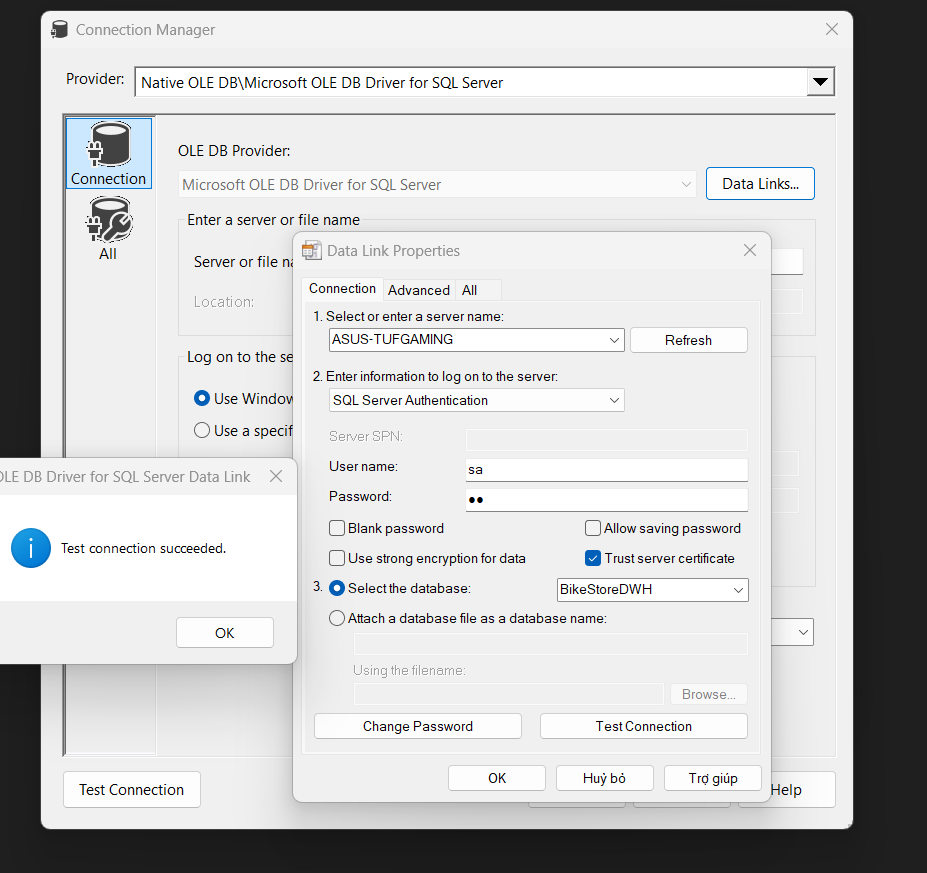
Right-click vào “Data Source” ở Solution Explorer và click New Data Source… sẽ xuất hiện Data Source Wizard dialog

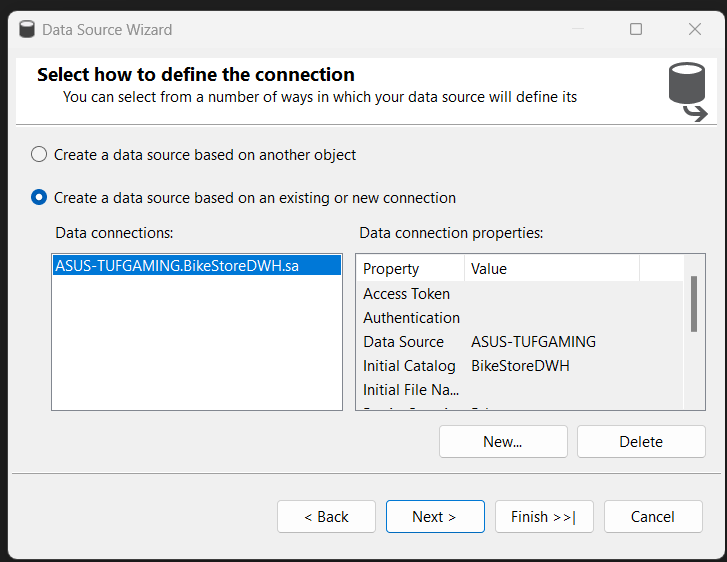


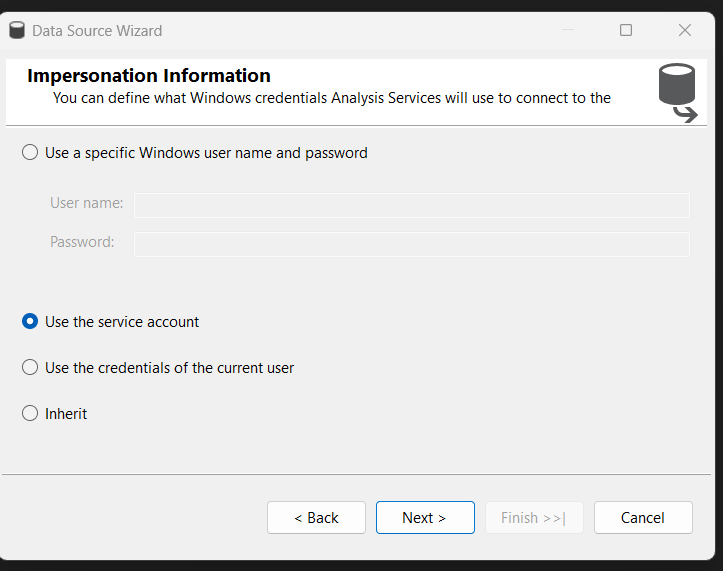
Click Next. Chọn option “Create a data source based on an existing or new connection” và click New…

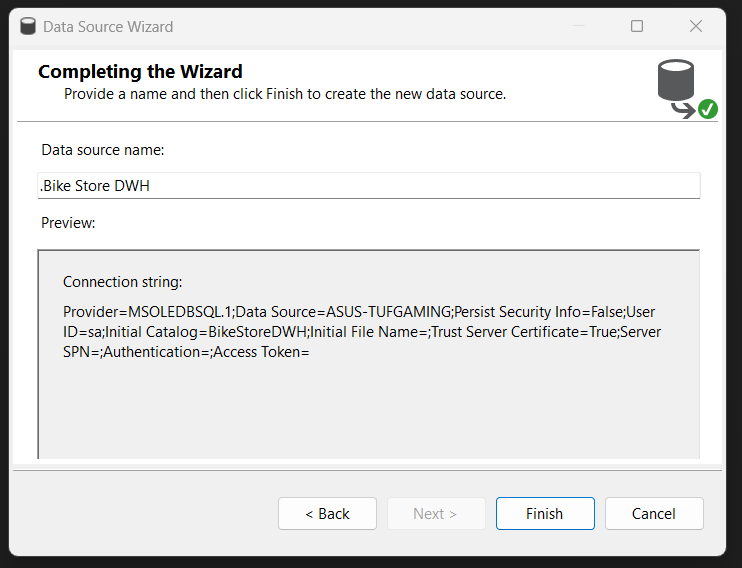


Điền thông tin Server name phù hợp và đặt tên database name và click OK

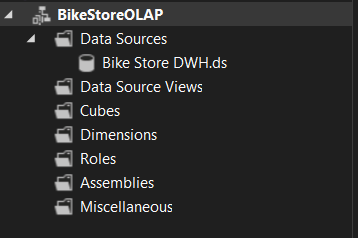








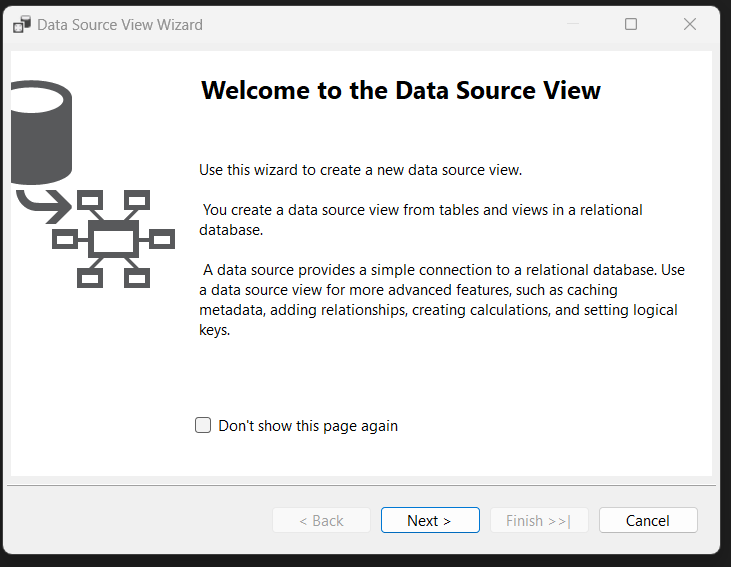
Click Finnish để hoàn thành việc kết nối. Sau khi hoàn thành bạn sẽ thấy thông tin về Data Source được hiển thị ở mục Solution Explorer như hình bên dưới.

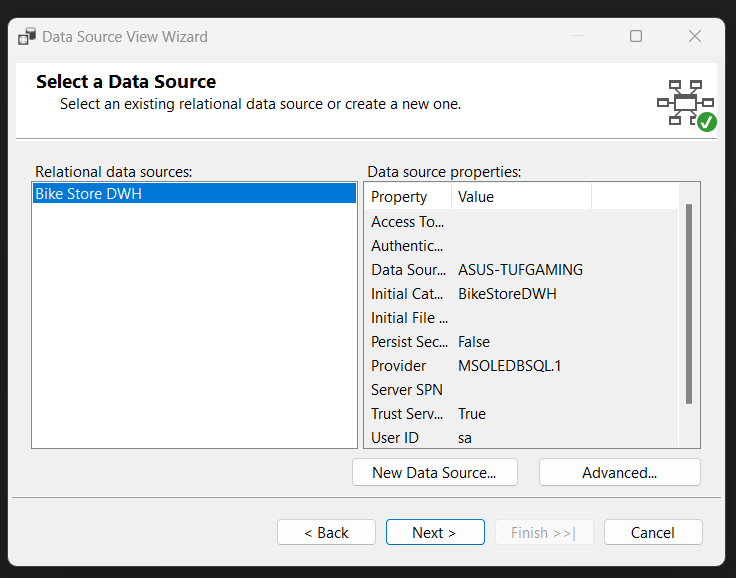


### Tạo Data Source View

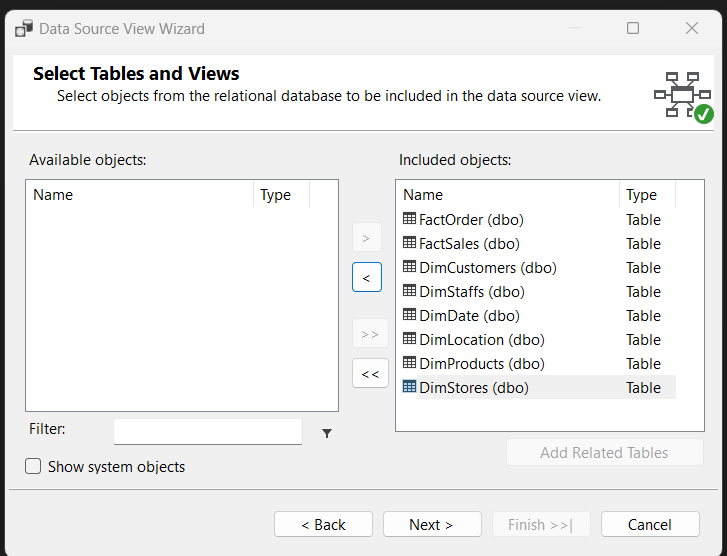
Data source view chứa logical model của CSDL (tables, keys, columns, và các constraints) sẽ được sử dụng bởi OLAP database để tạo các data cube.

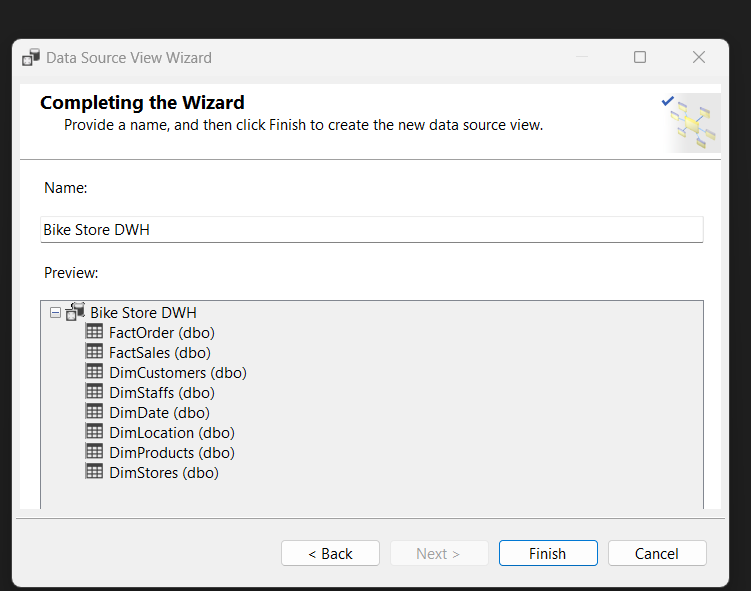
Right-click vào Data Source Views từ Solution Explorer và click New Data Source View… sẽ xuất hiện Data Source View Wizard, click Next,





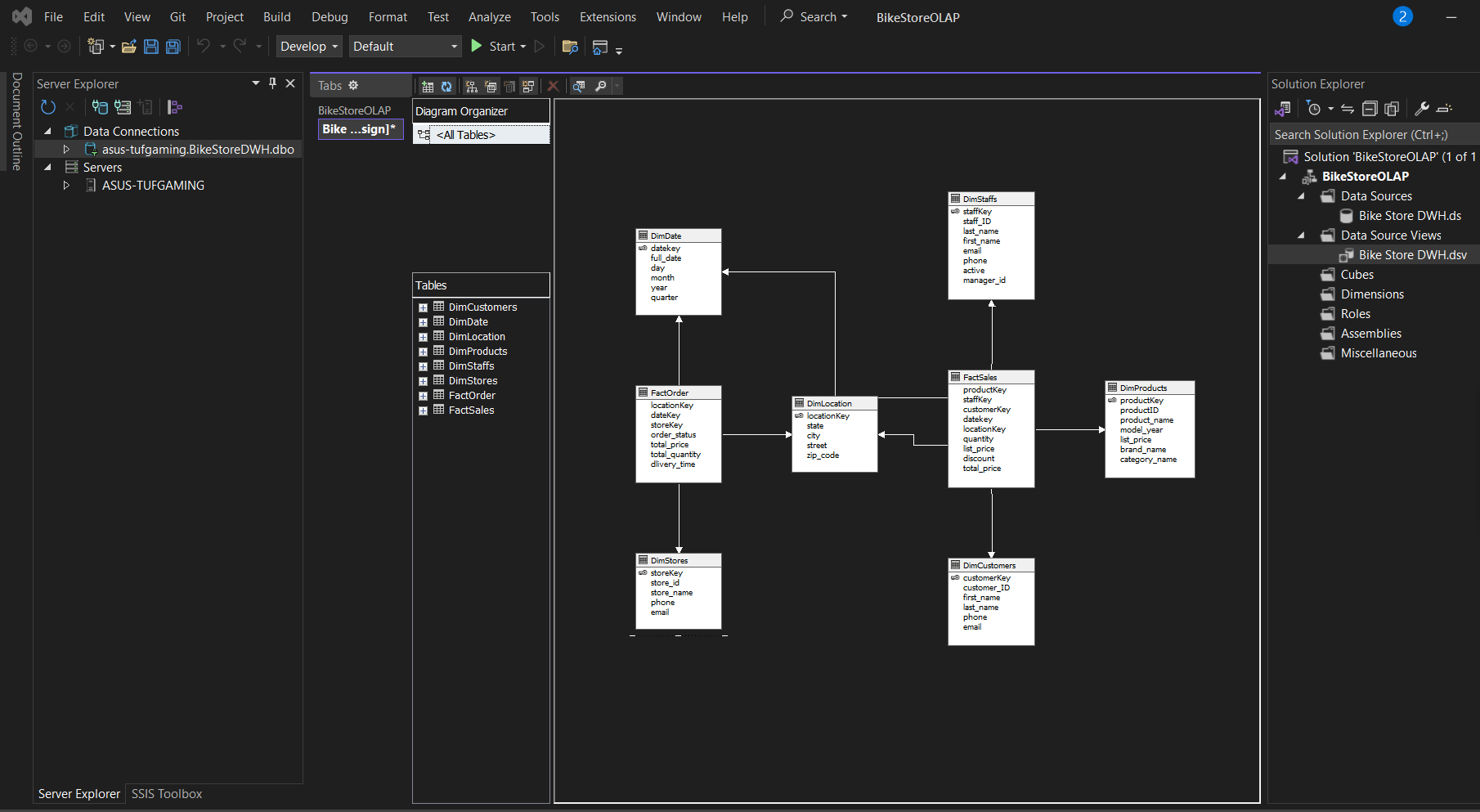
Click Next Chọn tất cả các table cho dimensional model và click Next.





Click Finnish để kết thúc.

Double-click vào Bike Store DWH Data Source View vừa mới tạo ta sẽ thấy star schema ở Design surface như sau:

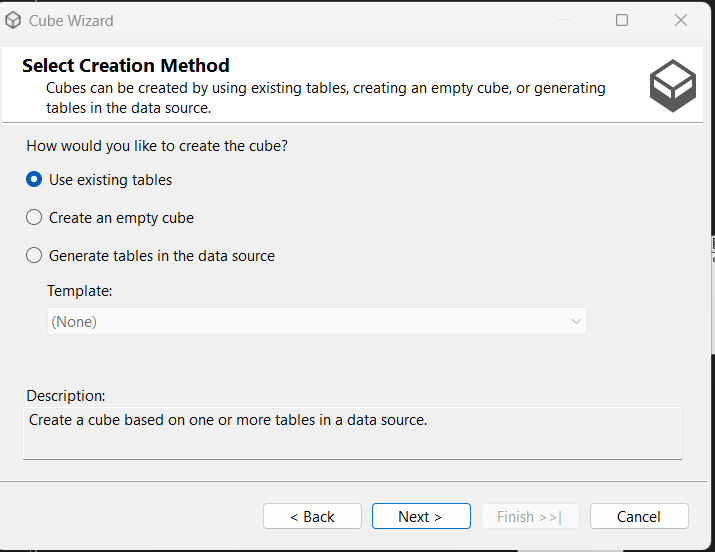


## 4.2. Quá trình xây dựng khối Cube

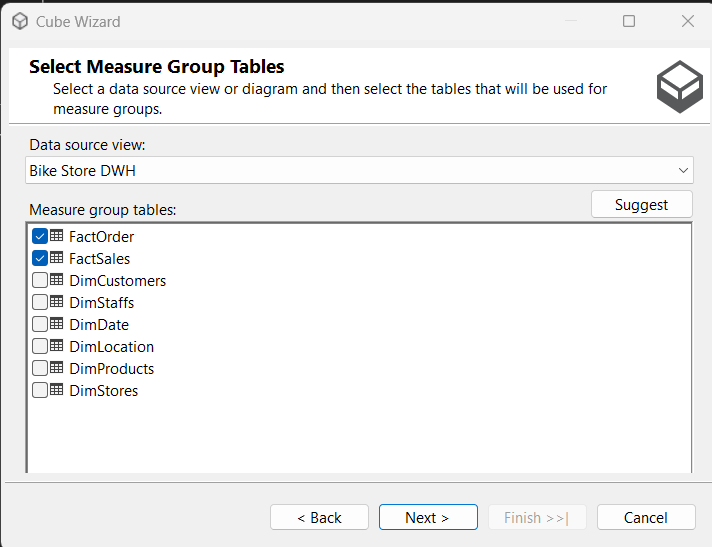
Right-click vào Cubes từ Solution Explorer và chọn New Cube… sẽ xuất hiện Cube Wizard dialog. Click Next.



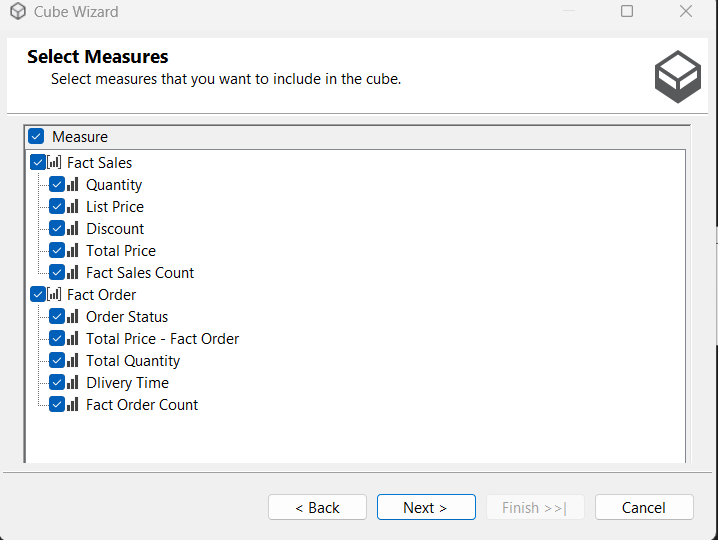
Click Next



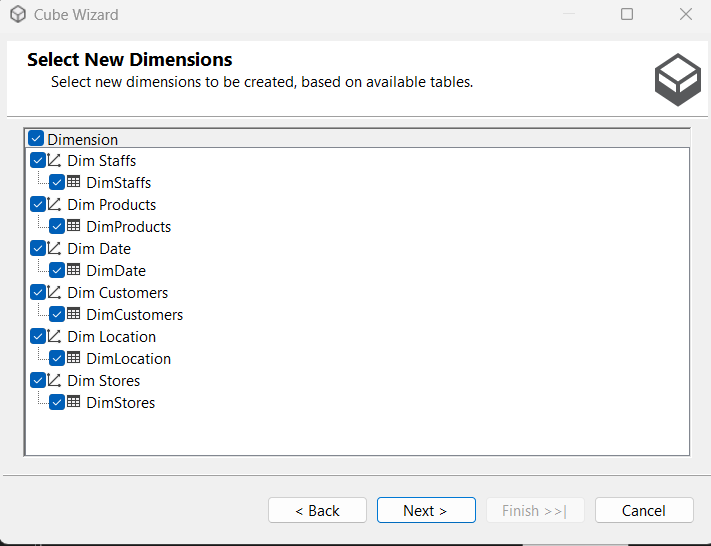
Click Next



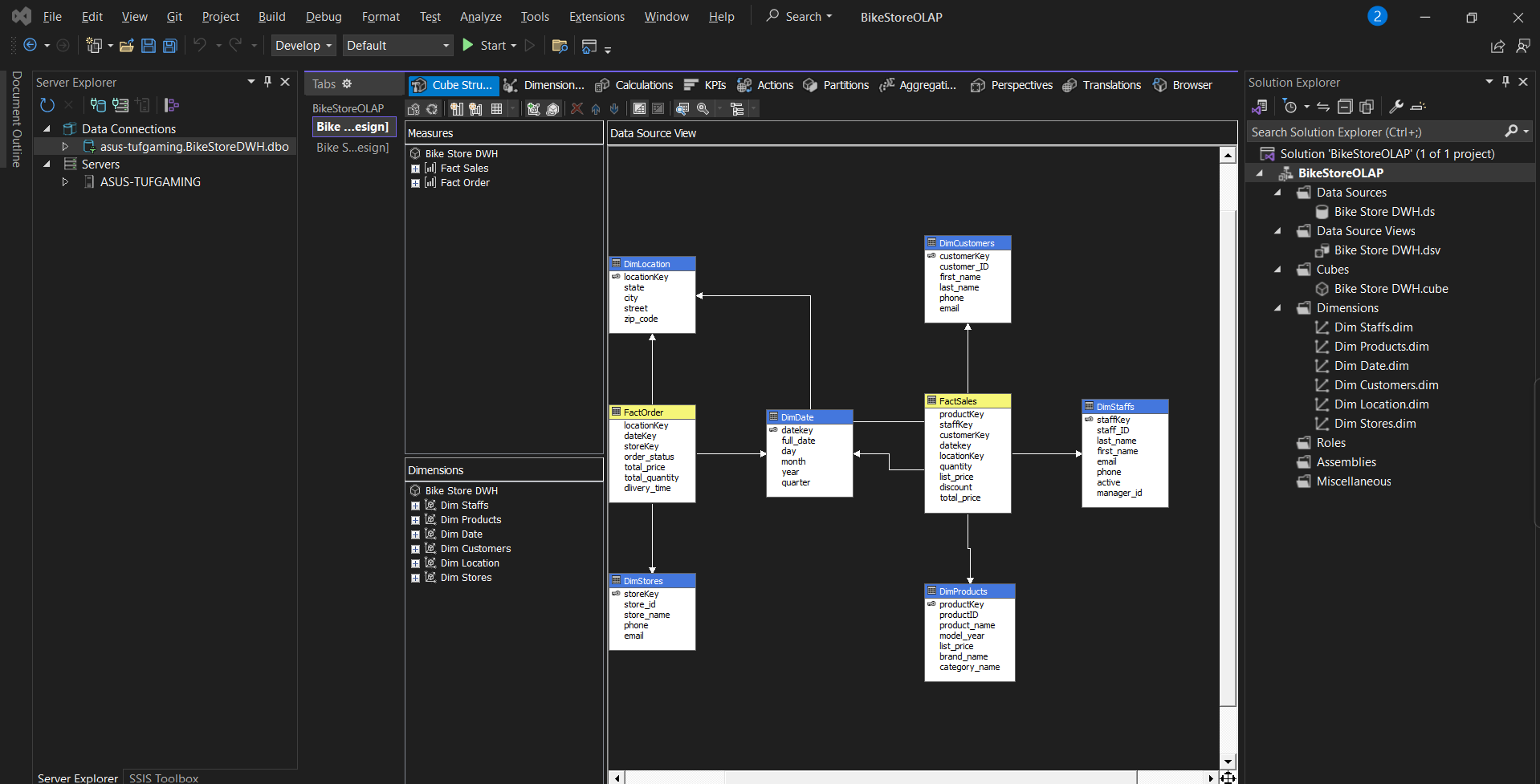
Chọn FactOrder và FactSales



Chọn (tất cả) các measures từ fact table để đưa vào cube và click Next.



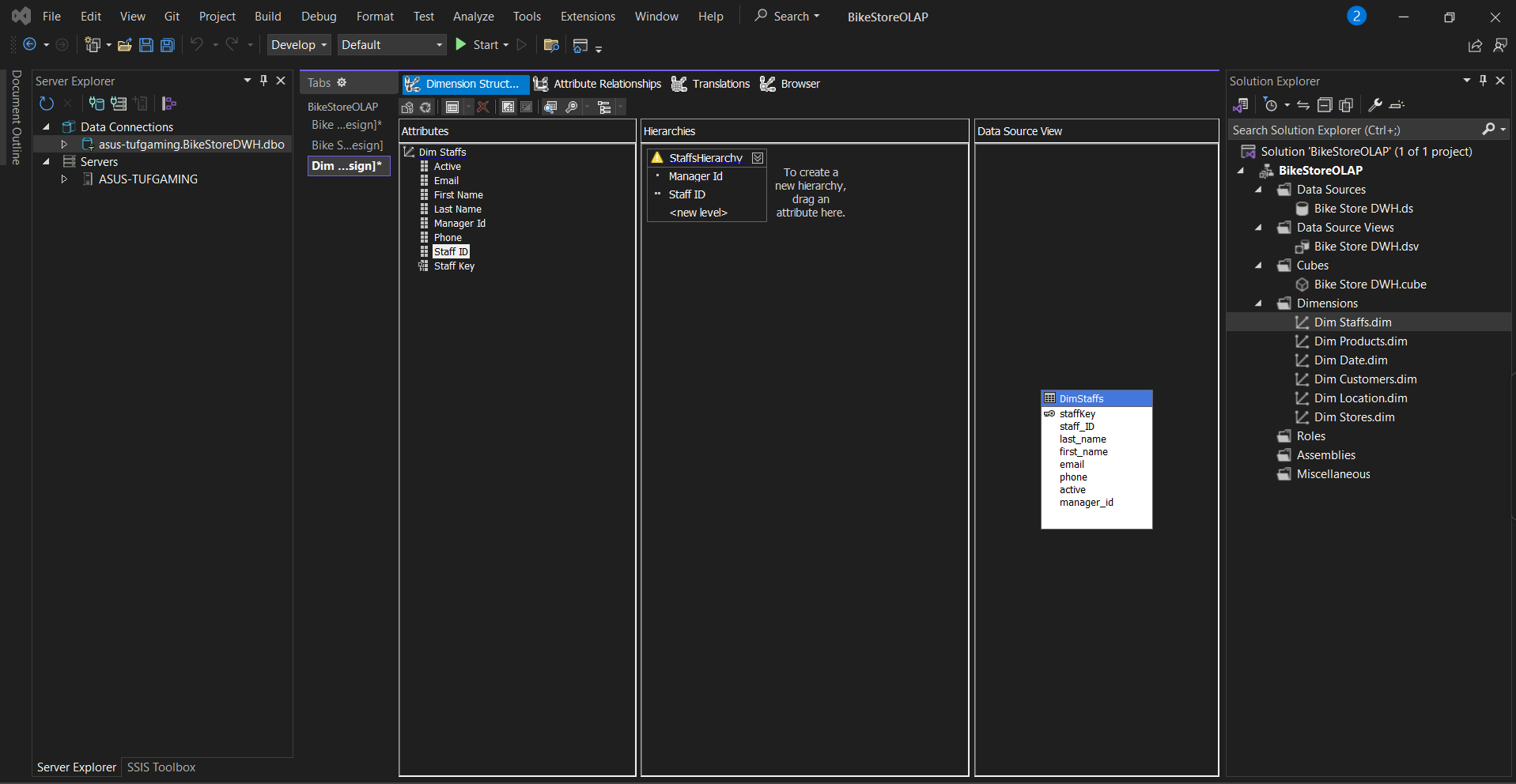
Sau khi kết thúc, màn hình của cube design sẽ như bên dưới



## 4.3. Cấu hình các dimension

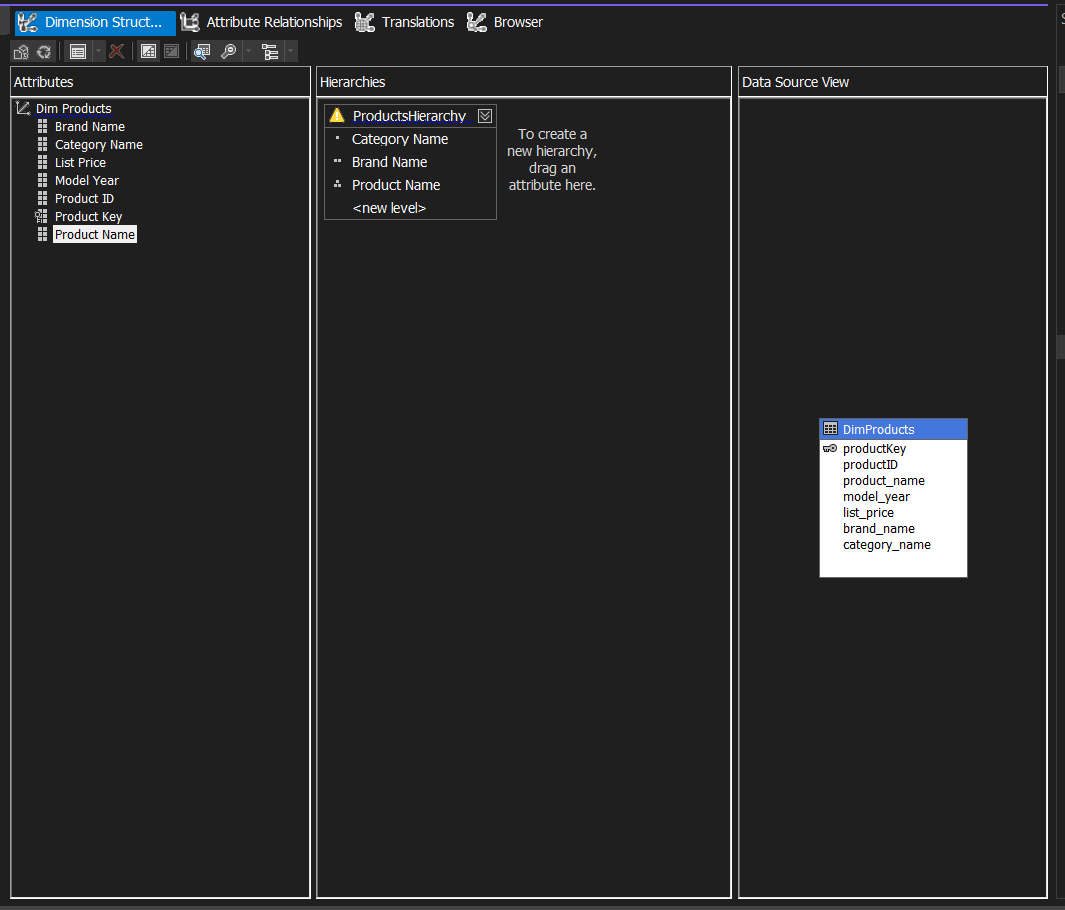
**Dim Staffs.dim**

StaffsHierarchies: Manager ID -> Staffs ID.



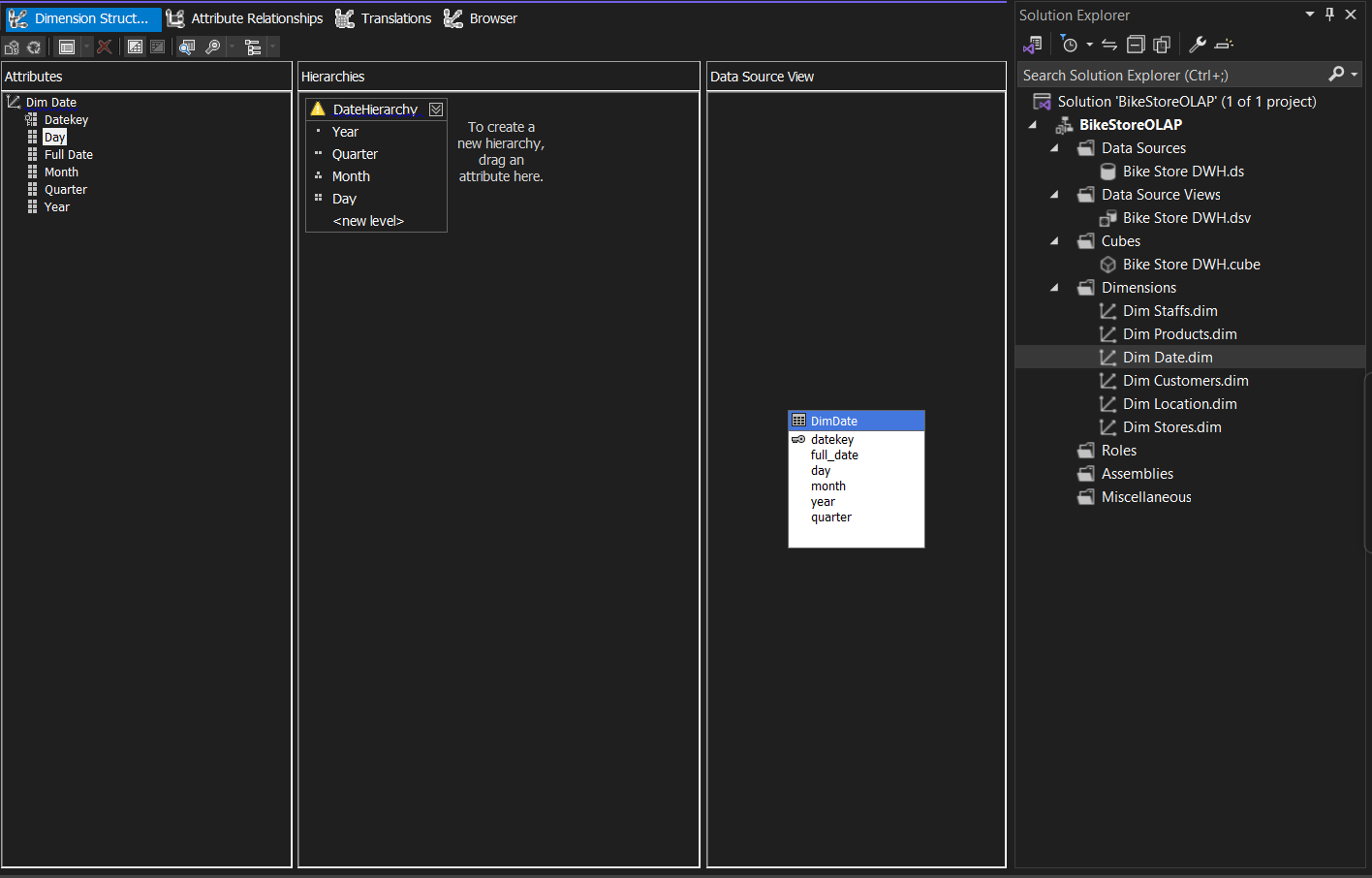
**Dim Products.dim**

ProductsHierarchies: Category Name -> Brand Name -> Product Name.

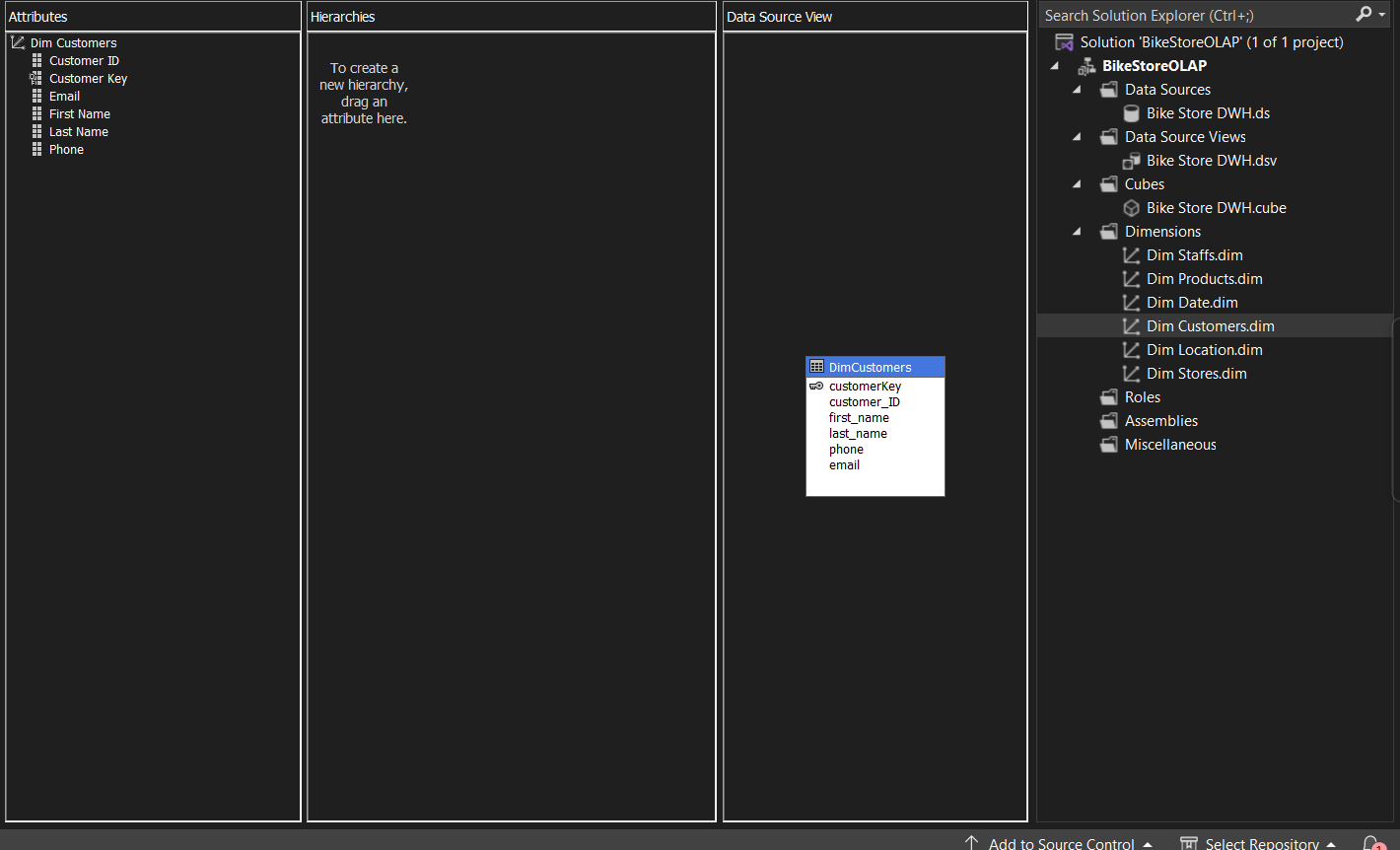


**Dim Date.dim**

Datehierarchies:Year -> quarter -> Month -> Day.

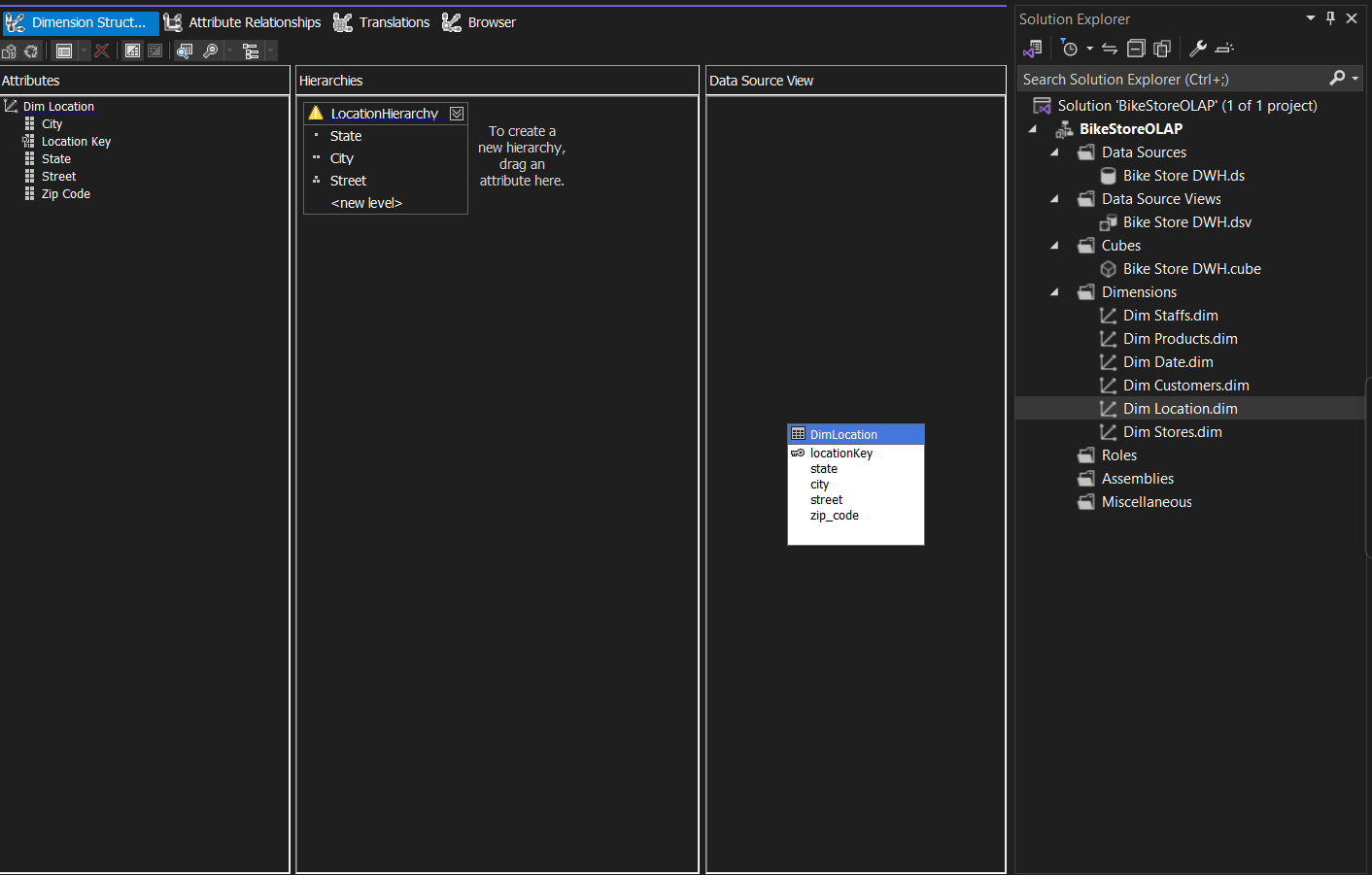


**Dim Customers**

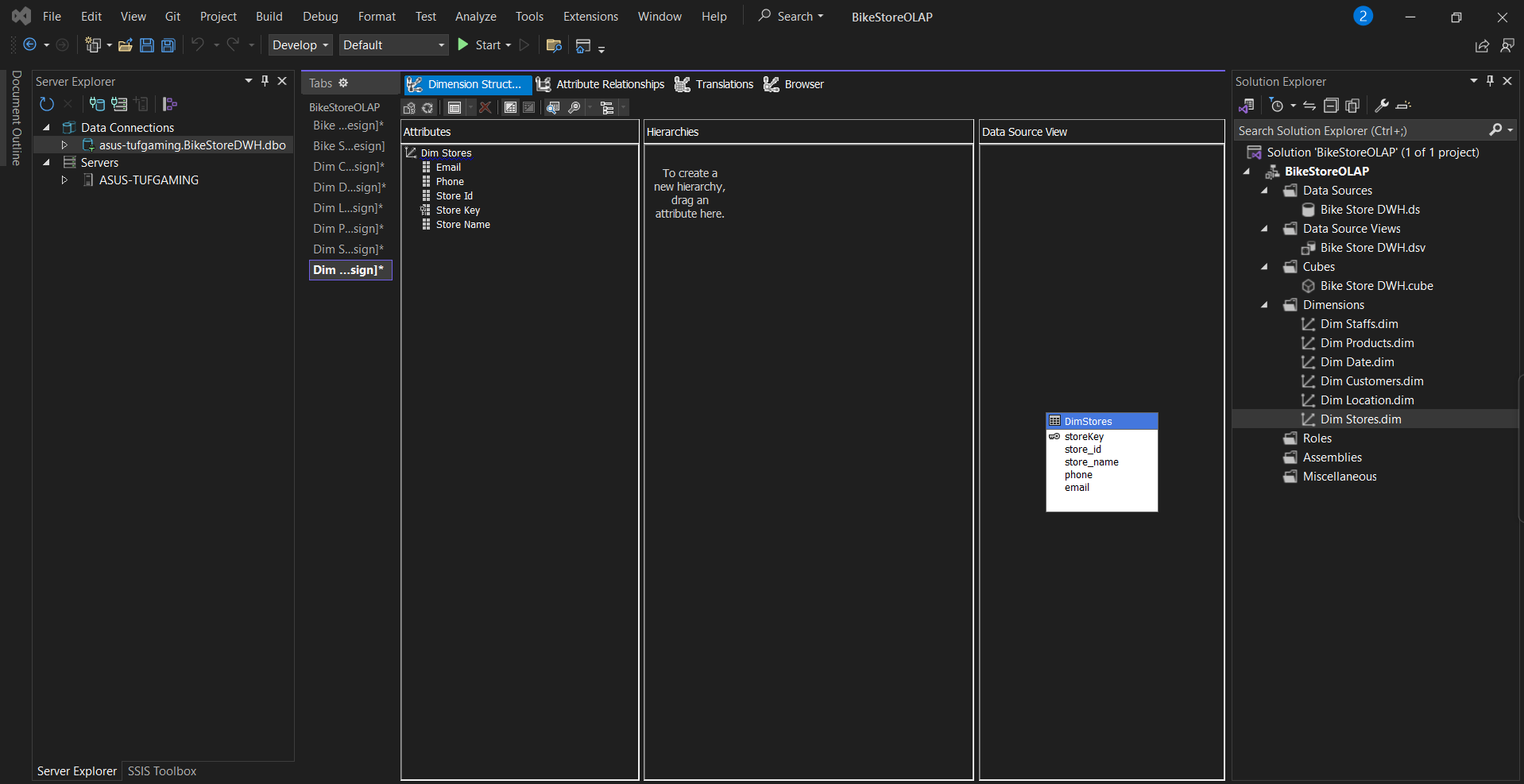


**Dim Location**

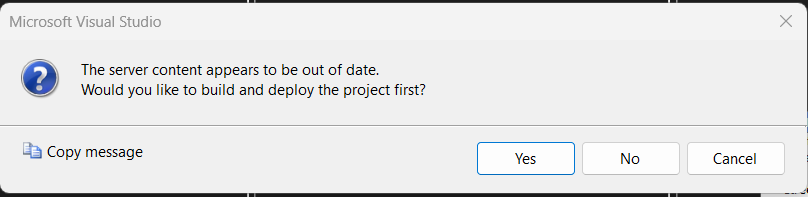
LocationHierarchies: State -> City -> Street .



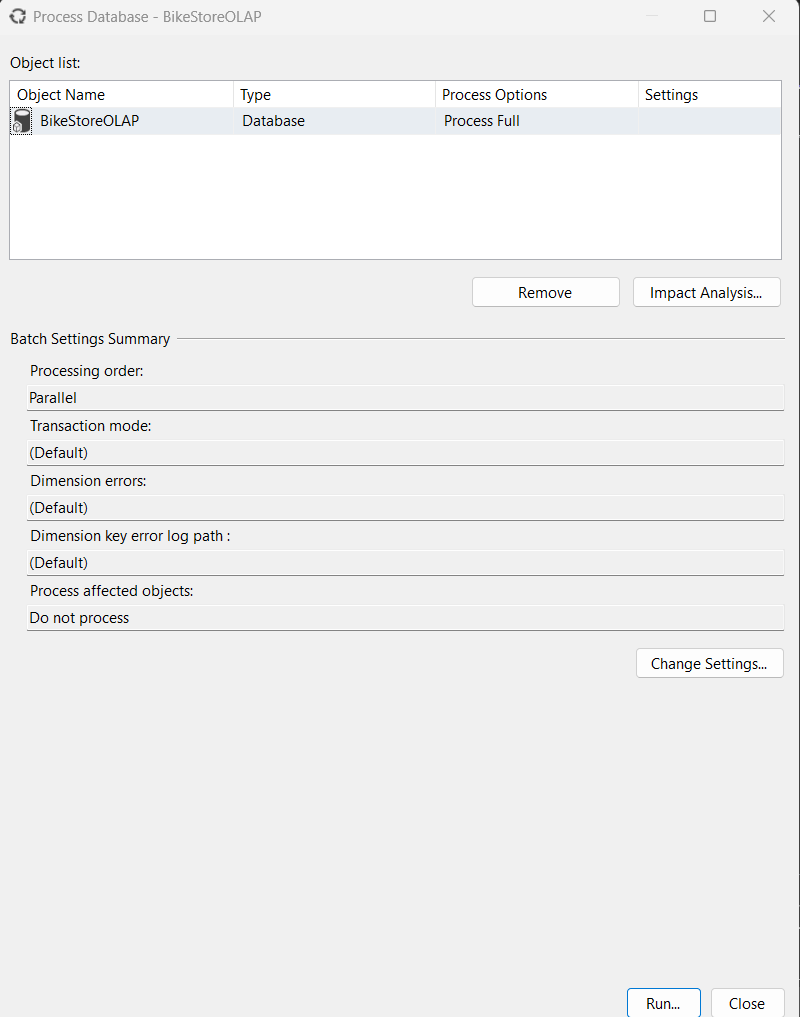
**Dim Stores:**



Sau khi cấu hình xong Ta Right-Click vào BikeStoreOLAP và ấn Process



Ấn Yes để tiếp Tục.



Click Run để thực thi. Nếu thành công, Process Progress dialog sau sẽ xuất hiện.



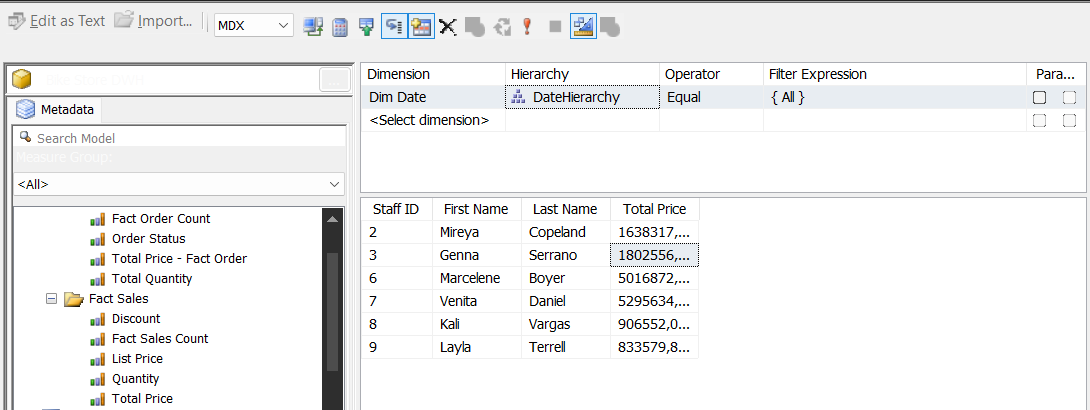
Thành Công.

## 4.4. Thực hiện phân tích dữ liệu

### Câu hỏi: Cho Biết xếp hạng KPI của Nhân viên trong Tháng/ Quý/ Năm.

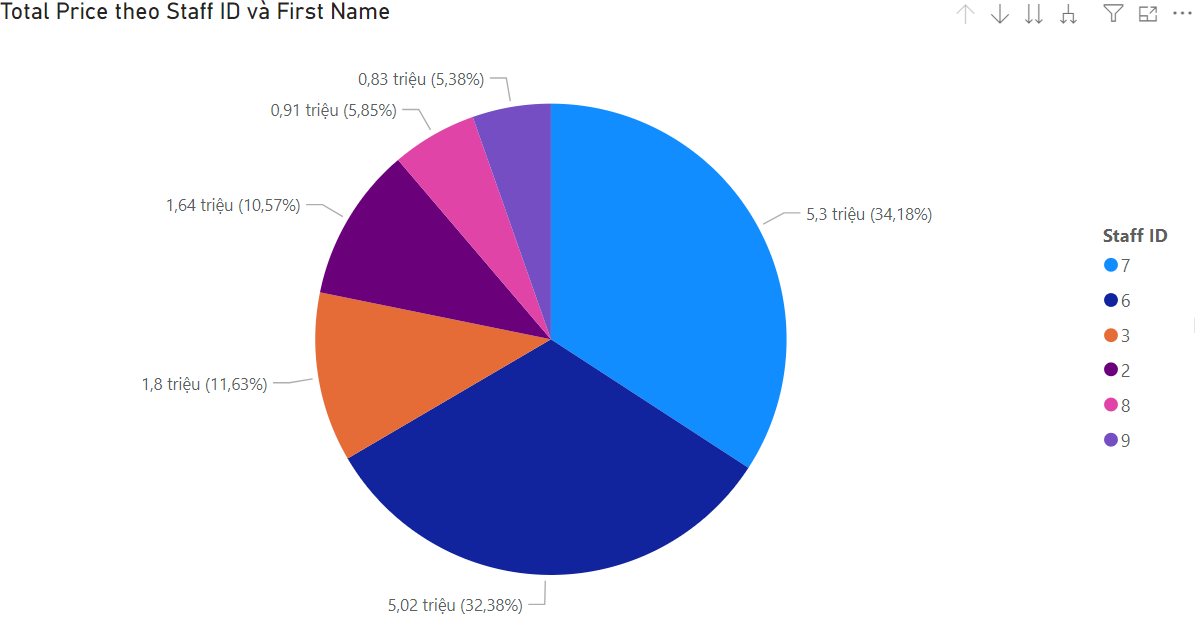
**Sử dụng công cụ SSAS**

Hiển thị KPI của Nhân viên bán hàng từ trước đến nay



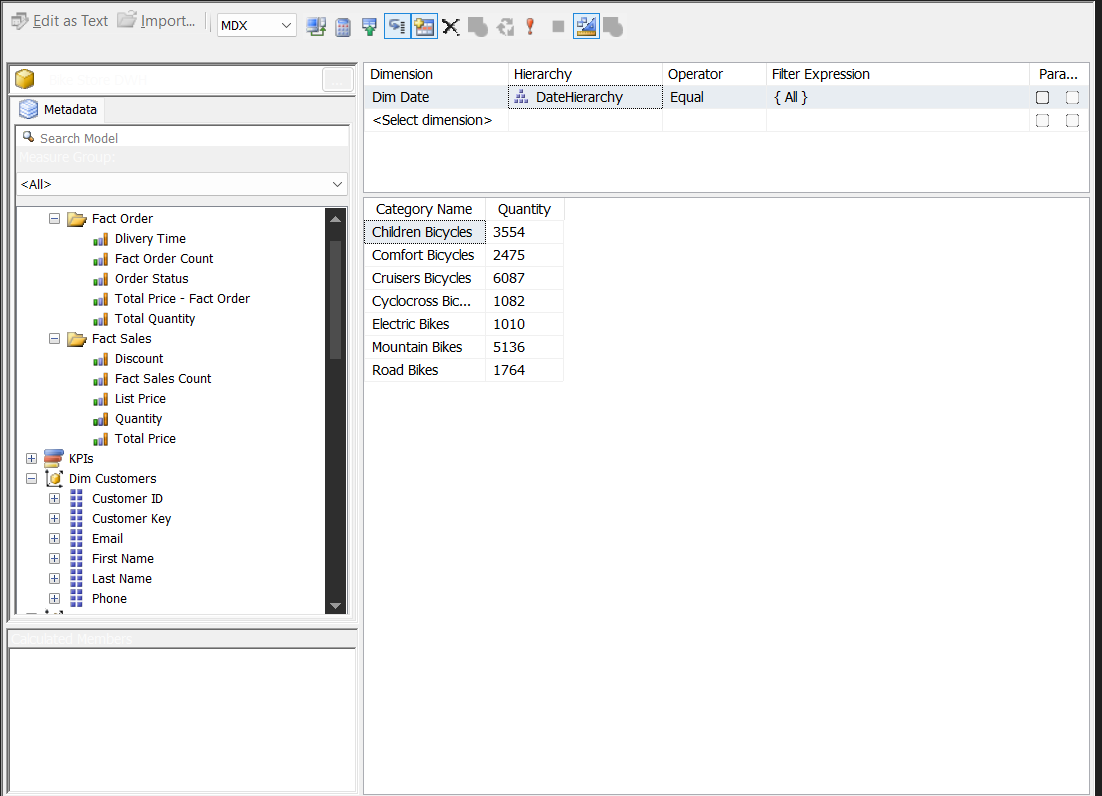
Có 7 nhân viên bán hàng. Ta thấy Doanh Số bán hàng cao nhất thuộc về nhân viên có Staff ID là 7, bằng cách này ta có thể Biết được nhân viên nào xuất sắc nhất để tiến hành khen thưởng hoặc cân nhắc thăng chức vụ.

**Sử dụng Power BI**



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

**Sử dụng công cụ SSAS**



**Sử dụng Power BI**

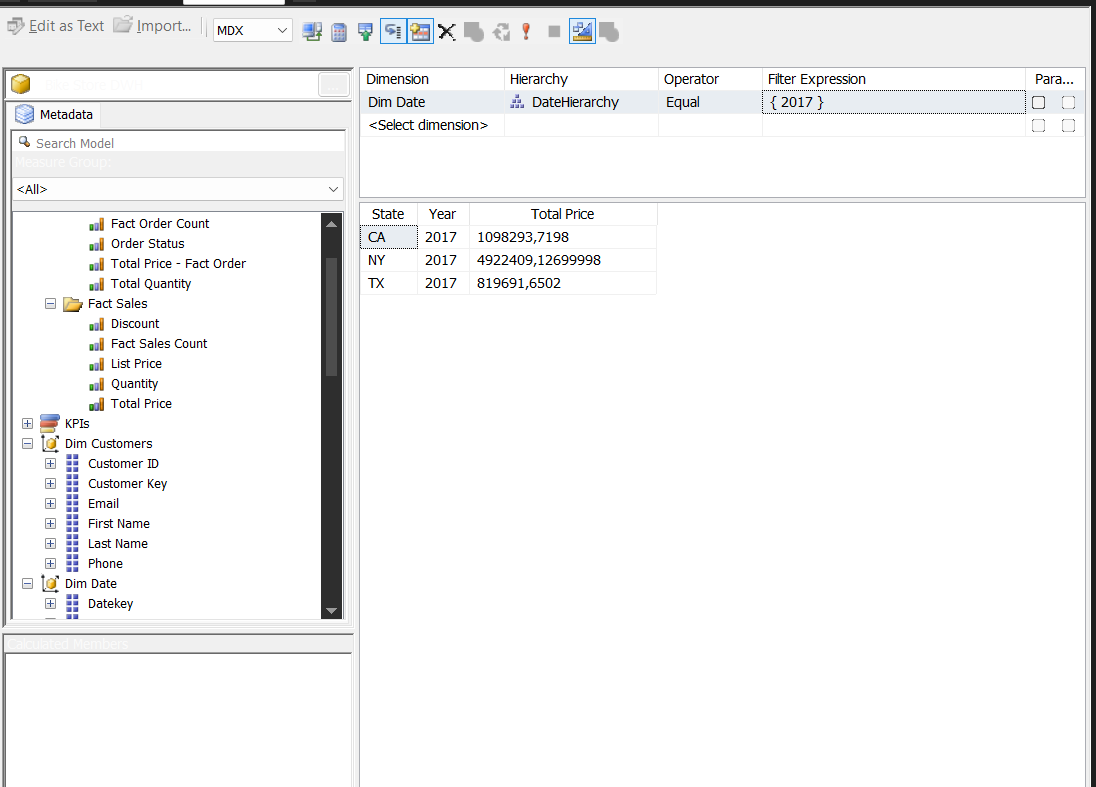
Biểu đồ danh mục các sản phẩm được tiêu thụ nhiều nhất

Ta thấy danh mục sản phẩm Cruisers Bicycles được tiêu thụ nhiều nhất với 6087 sản phẩm.

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

**Sử dụng công cụ SSAS**

Số liệu thể hiện tổng số tiền mua sắm từ Bang tính từ trước cho đến nay trên Nước Mỹ.



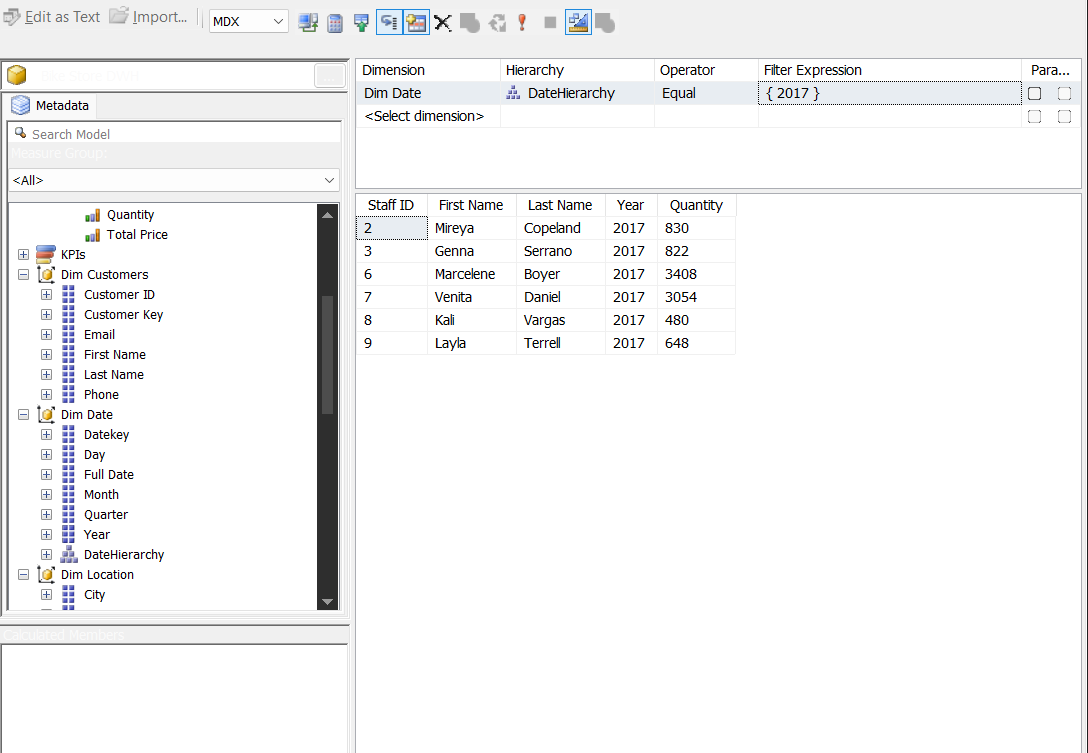
**Sử dụng Power BI**



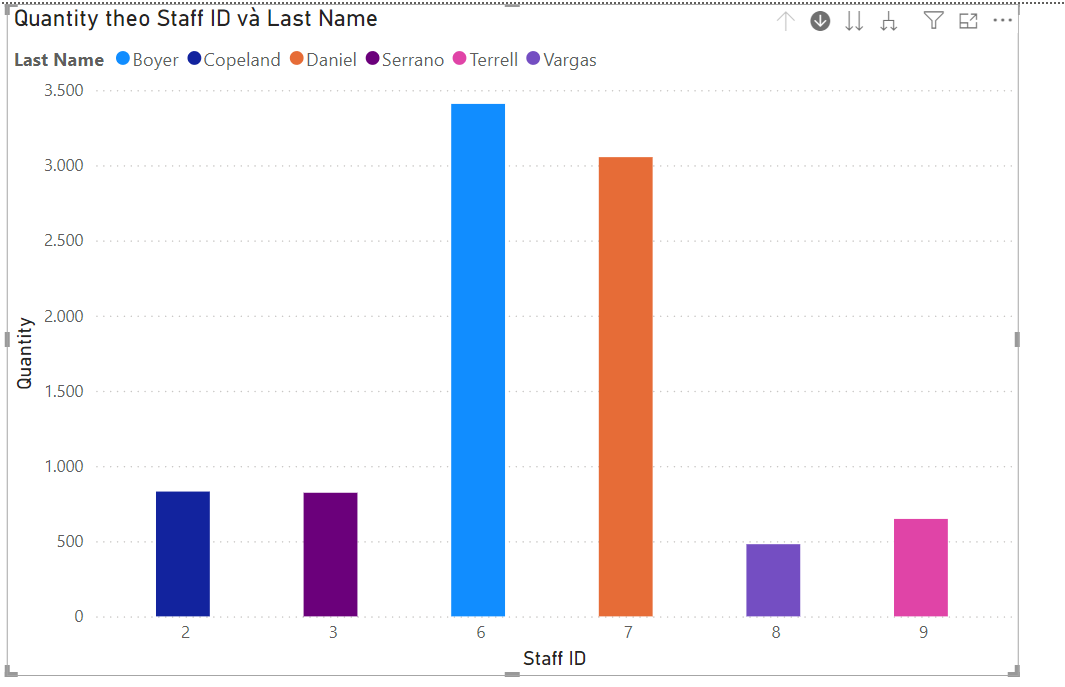
### Câu hỏi: Cho biết top những nhân viên bán được nhiều sản phẩm nhất Trong Tháng/ Quý/ Năm.

**Sử dụng công cụ SSAS**

Số liệu biểu diễn số lượng sản phẩm nhân viên bán được trong năm 2017.

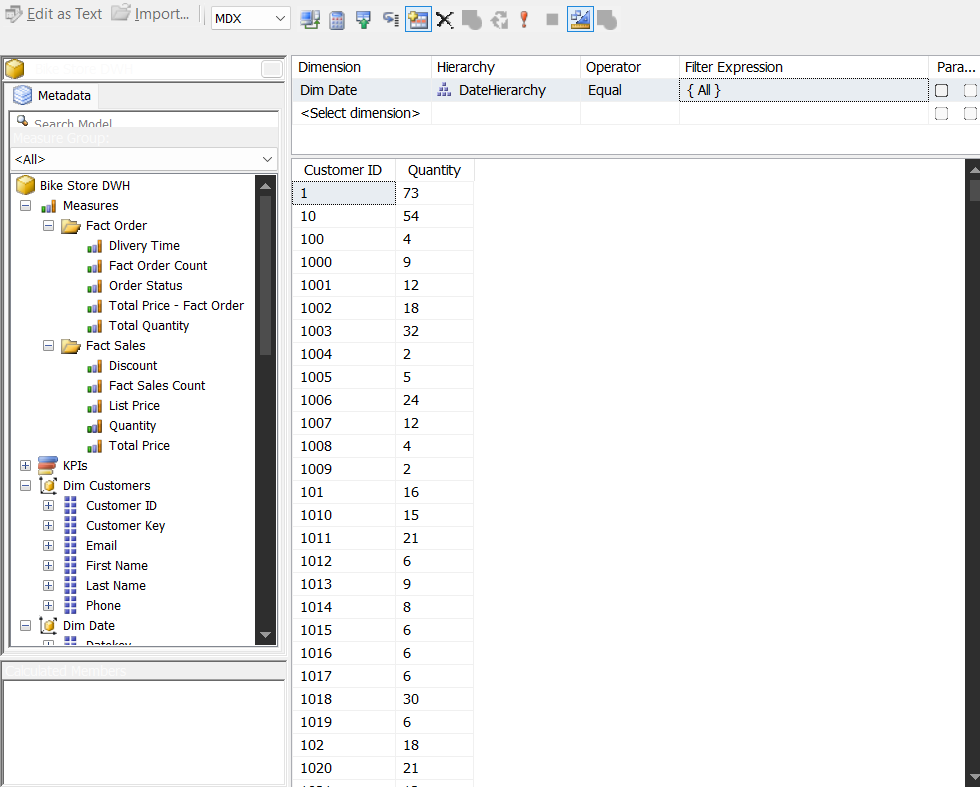


**Sử dụng Power BI**



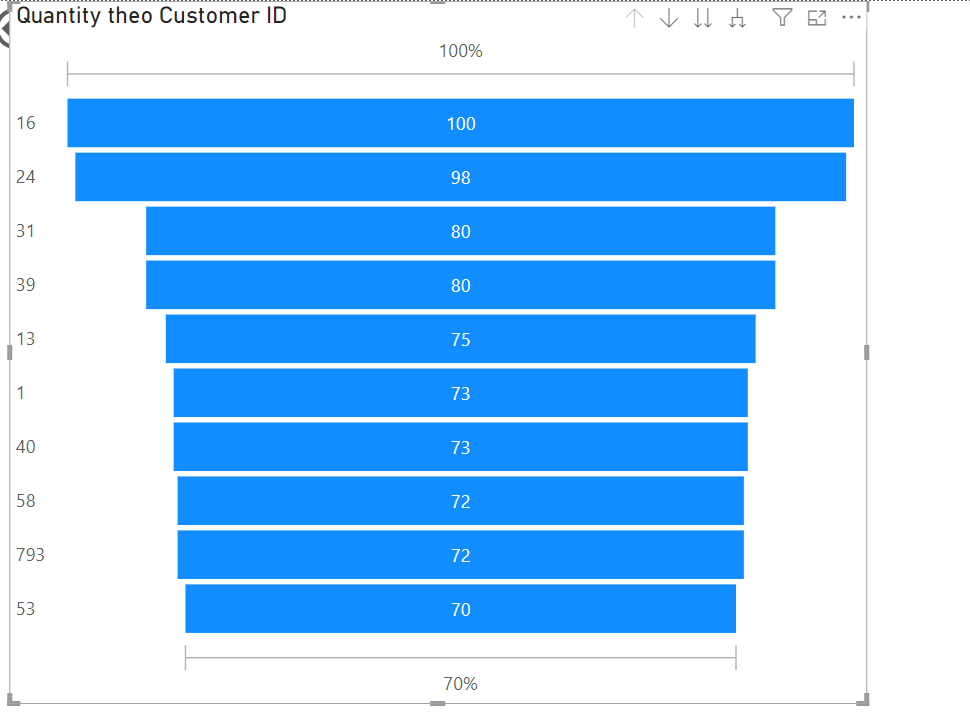
### Câu Hỏi: Chọn ra top 10 khách hàng mua nhiều sản phẩm nhất.

**Sử dụng SSAS**

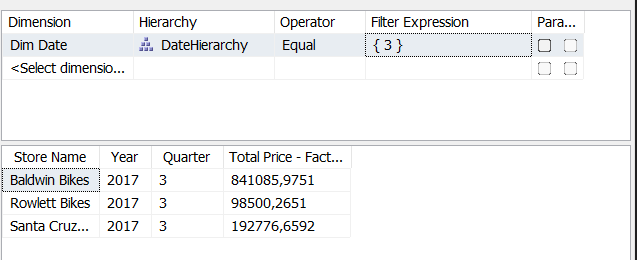


**Sử dụng Power BI**

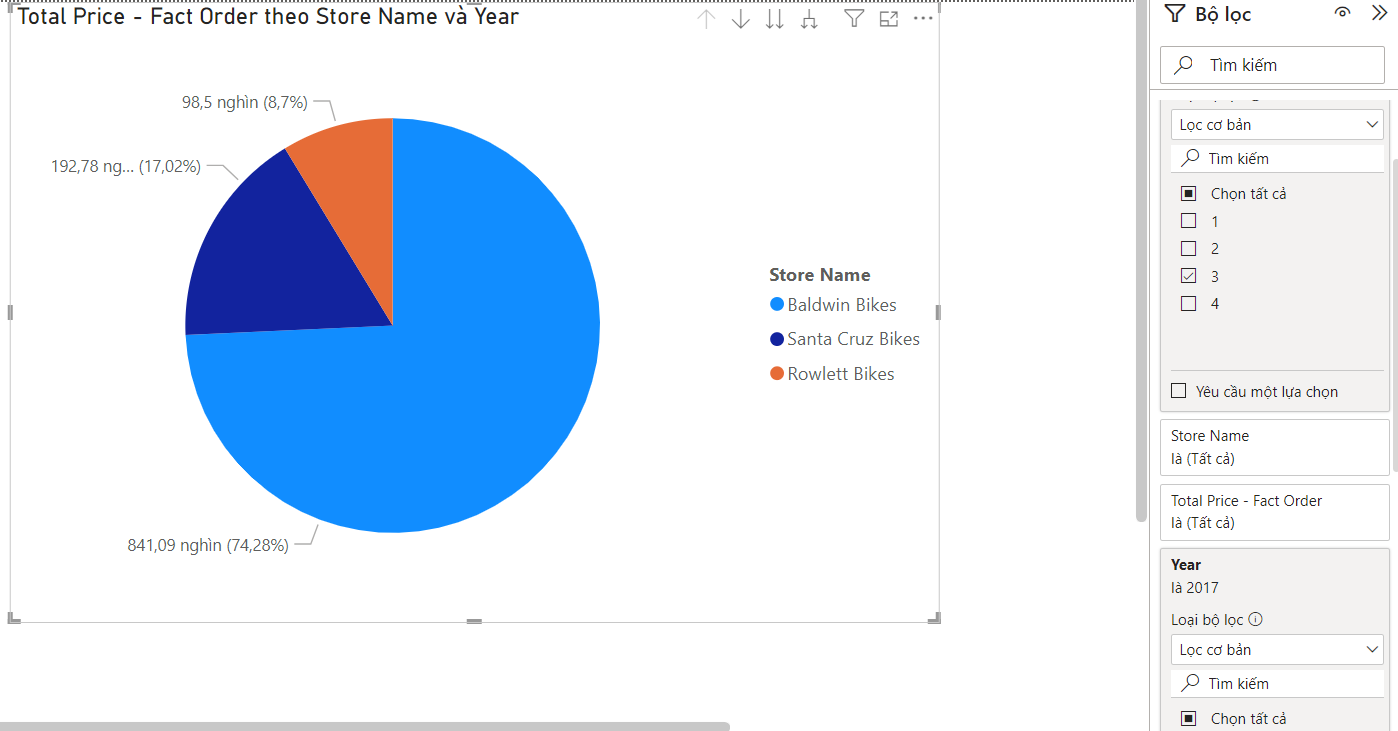
Biểu đồ Thể hiện 10 khách hàng có lượng tiêu thụ nhiều nhất



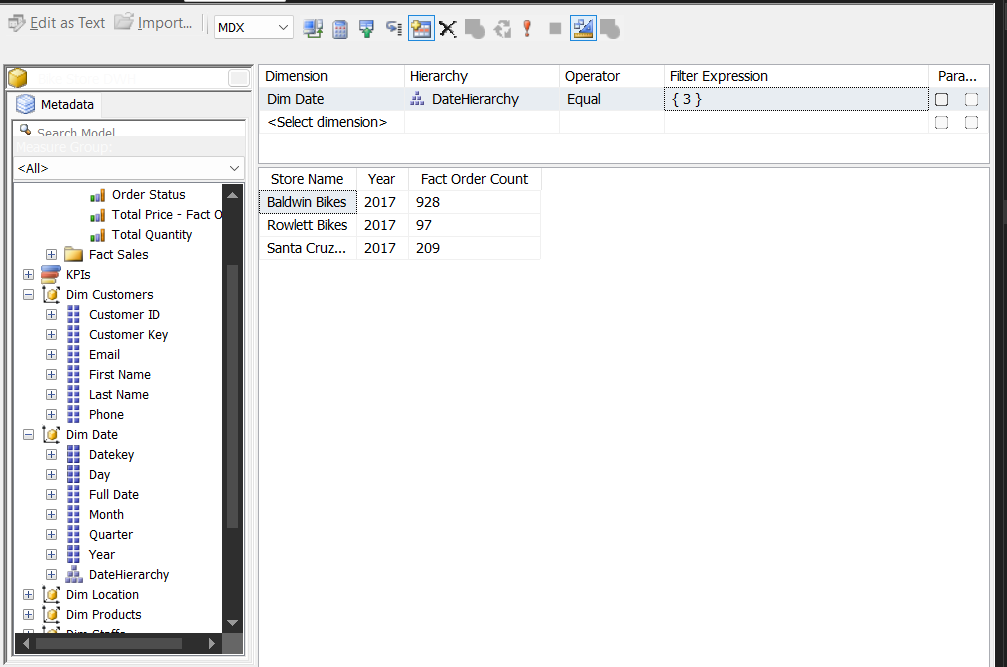
### Câu hỏi: Cho biết Tổng tiền tất cả đơn hàng của từng của hàng trong tháng/ quý/ năm.

**Sử dụng công cụ SSAS** 

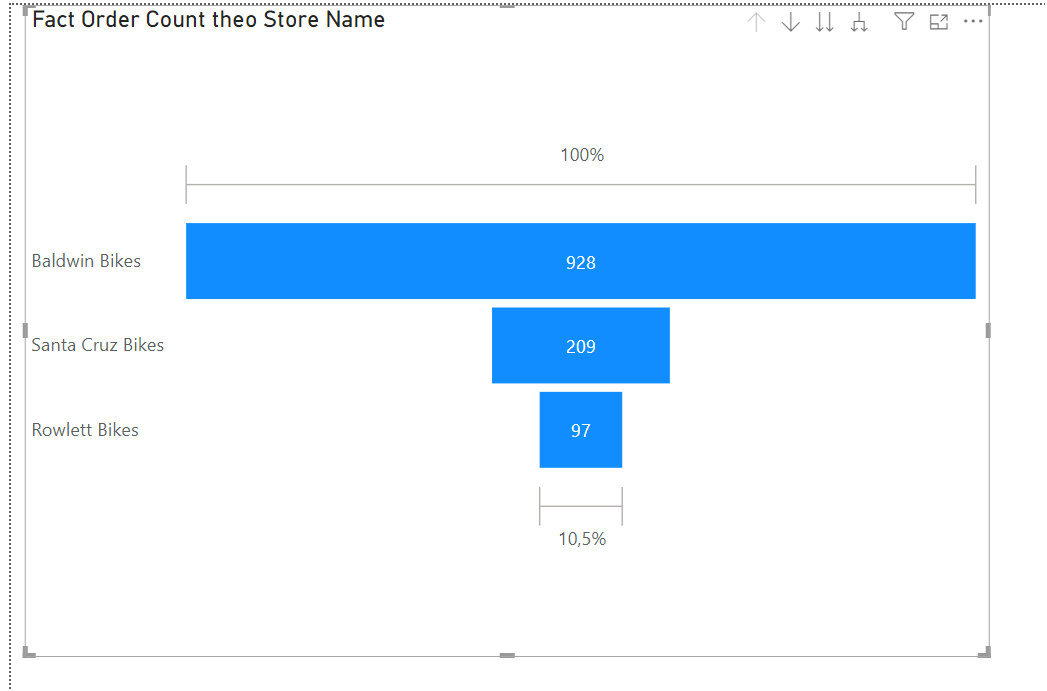
**Sử dụng Power BI**



### Câu hỏi: Cho biết Cửa Hàng có nhiều đơn hàng nhất trong tháng/ quý/ năm.

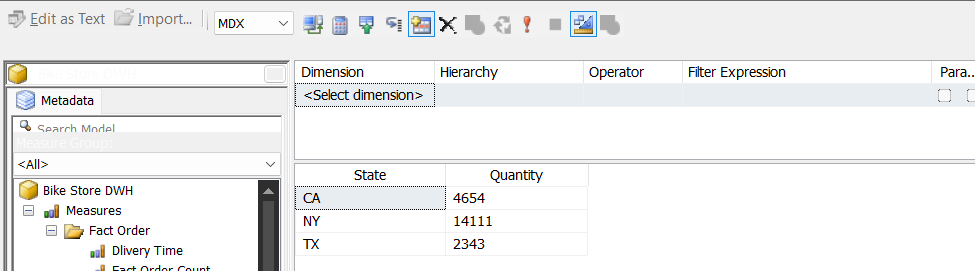
**Sử dụng công cụ SSAS** 

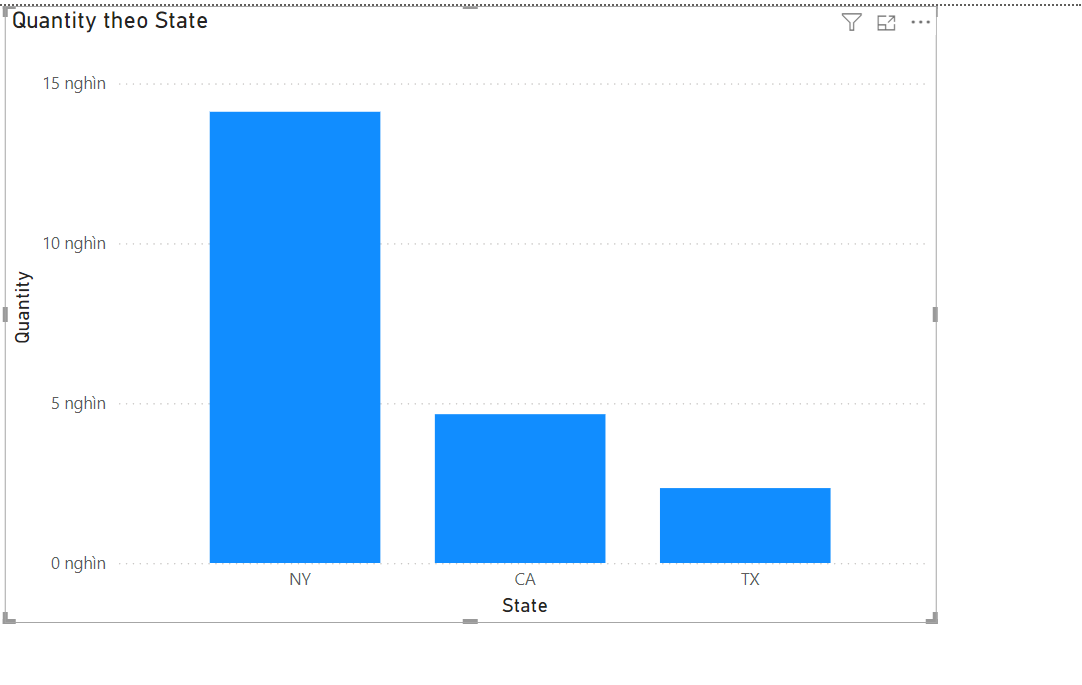
**Sử dụng Power BI**



### Câu Hỏi: Cho biết State có số lượng Sản phẩm tiêu thụ nhiều nhất.

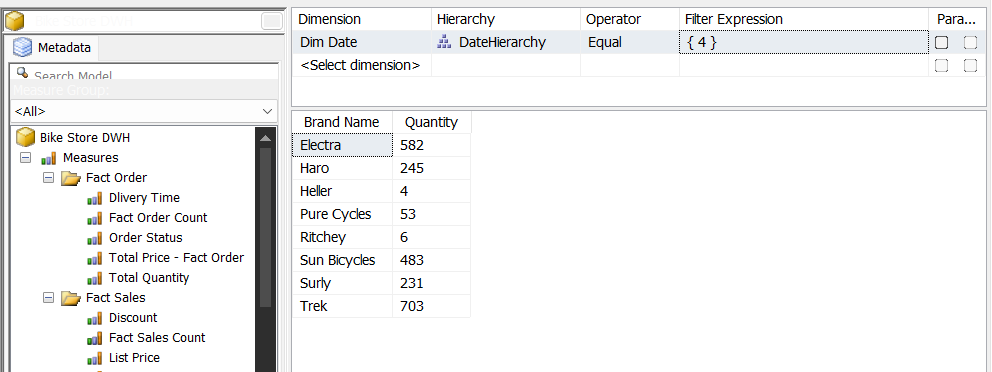
**Sử dụng công cụ SSAS**

 **Sử dụng Power BI**

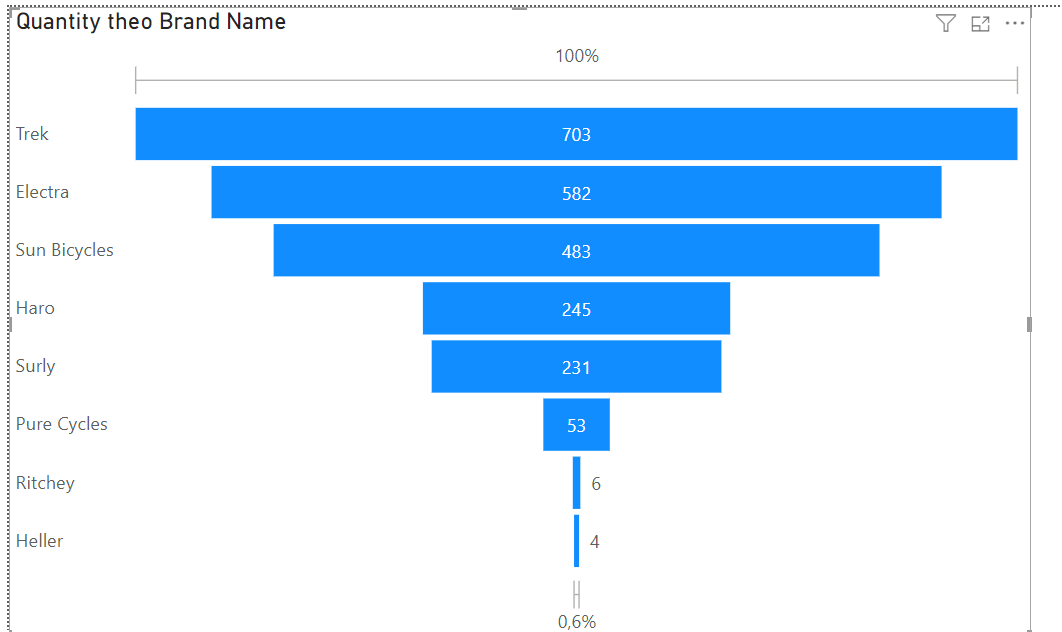


### Câu hỏi: Cho biết Nhãn Hàng được mua nhất trong tháng/ quý/ năm.

**Sử dụng công cụ SSAS**



**Sử dụng Power BI**



# CHƯƠNG 5: KẾT LUẬN

## 5.1. Kết quả đạt được

* Hiểu sâu hơn về kiến thức kho dữ liệu
* Sử dụng thành thạo hơn các công cụ SISS, SSAS, Power BI
* Nâng cao sử lý thông tin và phân tích dữ liệu

## 5.2. Những hạn chế

Chưa khai thác hết được tập dữ liệu

Chưa đạt đến mức nâng cao trong tổng thể quá trình xây dựng nâng cao thành kho dữ liệu hoàn chỉnh.

## 5.3. Tài liệu tham khảo

[Bike Store Relational Database | SQL (kaggle.com)](https://www.kaggle.com/datasets/dillonmyrick/bike-store-sample-database)

[SQL Server Sample Database (sqlservertutorial.net)](https://www.sqlservertutorial.net/getting-started/sql-server-sample-database/)

[Hướng dẫn SSAS: SSAS Cube là gì, Archikiến trúc & các loại (guru99.com)](https://www.guru99.com/vi/ssas-tutorial.html)

[Kết nối với Cơ sở SQL Server Analysis Services dữ liệu (Nhập) - Hỗ trợ của Microsoft](https://support.microsoft.com/vi-vn/office/k%E1%BA%BFt-n%E1%BB%91i-v%E1%BB%9Bi-c%C6%A1-s%E1%BB%9F-sql-server-analysis-services-d%E1%BB%AF-li%E1%BB%87u-nh%E1%BA%ADp-b6dd7f39-bea5-4e98-9aa1-39fc7b24424b)

[(477) Power BI nâng cao | Tạo Dashboard quản lý bán hàng (Kèm file thực hành) - YouTube](https://www.youtube.com/watch?v=k9dmVlPqGDM)

<https://www.youtube.com/watch?v=Vz8ttUSKpYo&t=2083s&pp=ygUEc3Nhcw%3D%3D>