**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**□&□**

**MÔN: KHO DỮ LIỆU**

Ảnh có chứa văn bản, mẫu họa

Mô tả được tạo tự động

**BÁO CÁO ĐỒ ÁN**

**ĐỀ TÀI**

**XÂY DỰNG KHO DỮ LIỆU**

**CHO CỬA HÀNG BÁN LẺ XE ĐẠP**

**GVHD: ThS. Nguyễn Văn Thành**

**Mã lớp:**

**Nhóm thực hiện:**

1. Nguyễn Đức Toàn - 22110437
2. Phạm Quốc Đức – 22110310
3. Đào Thanh Hiếu - 22110321
4. Trương Hồng Hạnh - 22110318

TP. Hồ Chí Minh, tháng 05 năm 2025

**DANH SÁCH THÀNH VIÊN THAM GIA**

**THỰC HIỆN ĐỀ TÀI VÀ VIẾT BÁO CÁO**

**Môn: Kho dữ liệu -** *HỌC KÌ II – NĂM HỌC 2024 – 2025*

|  |  |  |  |
| --- | --- | --- | --- |
| **5STT** | **HỌ VÀ TÊN** | **MSSV** | **TỶ LỆ ĐÓNG GÓP** |
| 1 | Nguyễn Đức Toàn | 22110437 | 100% |
| 2 | Đào Thanh Hiếu | 22110321 | 100% |
| 4 | Phạm Quốc Đức | 22110310 | 100% |
| 3 | Trương Hồng Hạnh | 22110318 | 100% |

**Nhận xét của giảng viên:**

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

*Ngày … tháng 05 năm 2024*

*Giảng viên chấm điểm*

*Ths. Nguyễn Văn Thành*

**LỜI CẢM ƠN**

Chúng em xin gửi lời biết ơn sâu sắc đến Thạc sĩ Nguyễn Văn Thành, người đã dành thời gian và công sức để truyền đạt những kiến thức quý báu cho chúng em trong suốt thời gian qua.

Tham gia môn học Kho Dữ liệu do Thầy dạy, chúng em đã trải qua những trải nghiệm học tập đầy ý nghĩa và bổ ích. Những kiến thức mà Thầy chia sẻ không chỉ là những con số và dữ liệu, mà còn là những cảm xúc, sự đam mê và sự tận tụy từ Thầy. Điều này đã giúp chúng em không chỉ hiểu sâu hơn về lĩnh vực này mà còn phát triển kỹ năng và lòng nhiệt thành trong học tập.

Mặc dù chúng em đã cố gắng hết sức, nhưng chúng em nhận thức rõ rằng công trình của mình vẫn còn nhiều điểm chưa hoàn hảo. Do đó, chúng em rất trân trọng và mong nhận được sự đánh giá và góp ý từ Thầy để bài đồ án của chúng em trở nên hoàn thiện hơn.

Một lần nữa, chúng em xin chân thành cảm ơn lòng tận tụy và sự hỗ trợ từ Thầy.

Trân trọng!

**PHÂN CÔNG NHIỆM VỤ**

| **Nhiệm vụ** | Đức Toàn | Thanh Hiếu | Quốc Đức | Hồng Hạnh |
| --- | --- | --- | --- | --- |
| Tìm kiếm tập dữ liệu | x | x | x | x |
| **Mô hình hóa dữ liệu vào workbook** | **x** | **x** | **x** | **x** |
| + DimStaff  + DimDate |  |  |  | x |
| + DimStore  + OrderFact | x |  |  |  |
| + DimProduct  + FactSales |  | **x** |  |  |
| + DimLocation  + DimCustomer |  |  | **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 SaleFact | x |  |  |  |
| + Tạo và đưa dữ liệu vào order Fact |  |  | x |  |
| **Data Cube Design with SSAS** |  | x |  | **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 DimDate, DimCustomer, DimStore |  | x |  |  |
| + Hierarchy cho DimLocation, DimStaff, DimProduct |  |  |  | x |
| Tạo Bike Store Dw Cube |  | x |  | x |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nhiệm vụ** | Đức Toàn | Thanh Hiếu | Quốc Đức | Hồng Hạnh |
| **Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Sales Reporting** | x |  |  | x |
| + Tổng doanh thu của chuỗi cửa hàng? Doanh thu của cửa hàng ở từng khu vực? |  |  |  | x |
| + Tổng số lượng xe đạp bán ra? Số lượng bán ra của mỗi khu vực? | x |  |  |  |
| + Tổng số tiền của từng khu vực theo mỗi tháng/ quý/ năm? |  | x |  | x |
| + Top những sản phẩm có giá trị bán ra cao nhất? |  |  |  | x |
| + Top những nhân viên xuất sắc nhất? (tính theo tổng giá trị sản phẩm bán ra) | x |  |  |  |
| **Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Order and delivery** |  | x | x |  |
| + Tổng số đơn hàng của chuỗi cửa hàng? Tổng Số đơn hàng của mỗi cửa hàng? |  | x |  |  |
| + Tổng thời gian giao hàng của chuỗi chuỗi cửa hàng và của mỗi cửa hàng? |  |  | x |  |
| + Số lượng đơn hàng theo từng tháng/ quý/ năm của từng cửa hàng và của mỗi cửa hàng? |  |  | x |  |
| + Tổng giá trị sản phẩm bán ra của chuỗi chuỗi cửa hàng và của mỗi mỗi cửa hàng theo từng ngày/ tháng/ quý/ năm là bao nhiêu? |  | x |  |  |
| **Trả lời câu hỏi dùng SSAS** | **x** |  |  | **x** |
| **Trả lời câu hỏi dùng Power BI Desktop** |  | **x** | **x** |  |

**MỤC LỤC**

[CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI 4](#_Toc1127755243)

[1.1. Lý do chọn đề tài 4](#_Toc491700679)

[1.2. Tổng quan về tập dữ liệu 4](#_Toc496238297)

[1.2.1. Nguồn dữ liệu 4](#_Toc1598056378)

[1.2.2. Mô tả chi tiết tập dữ liệu 5](#_Toc2103055664)

[1.2.2.1. Bảng customers 6](#_Toc623198875)

[1.2.2.2. Bảng stores 6](#_Toc825745706)

[1.2.2.3. Bảng staffs 7](#_Toc1268753828)

[1.2.2.4. Bảng orders 7](#_Toc1580456133)

[1.2.2.5. Bảng brands 8](#_Toc500997588)

[1.2.2.6. Bảng categories 9](#_Toc2130565276)

[1.2.2.7. Bảng products 9](#_Toc1268263654)

[1.2.2.8. Bảng stock 9](#_Toc1932036303)

[1.2.2.9. Bảng order\_items 9](#_Toc1781401881)

[1.3. Giới thiệu các công cụ được sử dụng 10](#_Toc1951891140)

[CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP) 11](#_Toc474198408)

[2.1. Xác định các Business Process và bảng Fact 11](#_Toc540829239)

[2.1.1. Business Process: Sales report 11](#_Toc1566225769)

[2.1.2. Business Process: Order and delivery 12](#_Toc1849438304)

[2.2. Xây dựng các bảng Dimension 12](#_Toc831268650)

[2.2.1. DimDate 12](#_Toc1751076815)

[2.2.2. DimLocation 12](#_Toc1232131217)

[2.2.3. DimCustomers 13](#_Toc244556098)

[2.2.4. DimProducts 13](#_Toc2118122375)

[2.2.5. DimStaffs 13](#_Toc446340306)

[2.2.6. DimStores 13](#_Toc481143971)

[CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS) 15](#_Toc1310789642)

[3.1. Tạo ETL project và solution 15](#_Toc1554262168)

[3.2. Thiết lập các Connection 16](#_Toc493175468)

[3.3. Import Dữ liệu từ file csv vào database bike\_store\_origin 22](#_Toc1660696903)

[3.4. Import Dữ liệu vào các bảng trong bike\_store\_stage: 27](#_Toc487888662)

[3.4.1. Bảng date stage (date\_stg) 28](#_Toc852177988)

[3.4.2. Bảng store stage (store\_stg) 34](#_Toc987002473)

[3.4.3. Bảng order item stage (order\_item\_stg) 34](#_Toc408569843)

[3.4.4.Bảng stock stage (stock\_stg) 35](#_Toc1328131842)

[3.4.5. Bảng staffstage (staff\_stg) 36](#_Toc966841576)

[3.4.6. Bảng order stage (order\_stg) 38](#_Toc1845094630)

[3.4.7. Bảng location stage (location\_stg) 39](#_Toc841583730)

[3.4.8. Bảng product stage (product\_stg) 41](#_Toc1170450968)

[3.4.9. Bảng customer stage (customer\_stg) 45](#_Toc819212656)

[3.4.10. Tạo Execute SQL Task để Truncate các bảng trong stage 48](#_Toc989448475)

[3.5. Import dữ liệu từ stage vào các bảng dimension 51](#_Toc1694640604)

[3.5.1. Bảng Date Dimension (DimDate) 51](#_Toc1896504914)

[3.5.2. Bảng Location Dimension (DimLocation) 54](#_Toc1605800824)

[3.5.3. Bảng Store Dimension (DimStore) 55](#_Toc1267457119)

[3.5.4. Bảng Product Dimension (DimProduct) 58](#_Toc1261504440)

[3.5.5. Bảng Customer Dimension (DimCustomer) 59](#_Toc2146559896)

[3.5.6. Bảng Staff Dimension (DimStaff) 59](#_Toc1654386119)

[3.6. Import dữ liệu từ stage vào các bảng fact 59](#_Toc1796724324)

[3.6.1. Bảng Sale Fact (sale\_fact) 59](#_Toc1778153001)

[3.6.2. Bảng Order Fact (order\_fact) 63](#_Toc1035056155)

[CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS & POWER BI) 68](#_Toc1571256385)

[4.1. Quá trình xây dựng mô hình 68](#_Toc499655536)

[4.1.1. Tạo Data Source 69](#_Toc1049533782)

[4.1.2. Tạo Data Source View 72](#_Toc83597616)

[4.2. Quá trình xây dựng khối Cube 73](#_Toc570826974)

[4.3. Cấu hình Hierarchy 78](#_Toc1079500869)

[4.3.1. Tạo Hierarchy cho Dim Date 78](#_Toc1229255427)

[4.3.2. Tạo Hierarchy cho Dim Location 79](#_Toc1303277449)

[4.3.4. Tạo Hierarchy cho Dim Product 81](#_Toc1998369590)

[4.3.5. Tạo Hierarchy cho Dim Staff 82](#_Toc2100943697)

[4.4. Thực hiện phân tích dữ liệu 84](#_Toc1634591505)

[4.4.1. Báo cáo doanh số (Sales reporting) 84](#_Toc1968612111)

[CHƯƠNG 5: KẾT LUẬN 92](#_Toc183200248)

[5.1. Kết quả đạt được 92](#_Toc1702501985)

[5.2. Những hạn chế 92](#_Toc878749267)

[5.3. Tài liệu tham khảo 93](#_Toc893890050)

# CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

## Lý do chọn đề tài

Trong thời đại công nghiệp hóa, hiện hiện đại hóa như ngày nay, việc xây dựng một kho dữ liệu cung cấp nền tảng thông tin rõ ràng cho các hoạt động kinh doanh, từ việc quản lý sản phẩm, khách hàng, quản lý giao hàng và báo cáo doanh thu là một điều cần thiết. Việc tận dụng sức mạnh của dữ liệu và công nghệ thông tin để cải thiện quy trình làm việc và đưa ra các quyết định một cách chuẩn xác. Chính vì những lý do đó, nhóm chúng em chọn đề tài *“Xây dựng Kho dữ liệu cho chuỗi cửa hàng bán lẻ xe đạp”* với mong muốn taọ nên một Kho dữ liệu giúp cho chuỗi cửa hàng bán lẻ xe đạp có thể dễ dàng quản lý, thao tác và từ đó đưa ra những quyết định đúng đắn.

## Tổng quan về tập dữ liệu

### Nguồn dữ liệu

Để xây dựng kho dữ liệu cho cửa hàng bán lẻ xe đạp, nhóm đã sử dụng tập dữ liệu Bike Store Relational Database, được lấy từ trang web Kaggle.com. Kaggle là một nền tảng trực tuyến nổi tiếng cung cấp các tập dữ liệu chất lượng từ nhiều lĩnh vực khác nhau, giúp cho việc nghiên cứu và phân tích dữ liệu trở nên dễ dàng hơn. Tập Bike Store Relational Database là một nguồn thông tin phong phú về hoạt động kinh doanh của các cửa hàng bán lẻ xe đạp, bao gồm thông tin về sản phẩm, đơn hàng, khách hàng, và các giao dịch liên quan.

Đường dẫn tải tập dữ liệu: [Bike Store Relational Database](https://www.kaggle.com/datasets/dillonmyrick/bike-store-sample-database?select=stores.csv)

### Mô tả chi tiết tập dữ liệu

Tập dữ liệu ‘Bike Store Relational Database’ bao gồm thông tin về hoạt động kinh doanh của các cửa hàng bán lẻ xe đạp. Dữ liệu bao gồm các 9 bảng với mô tả như sau:

| **Tên bảng** | **Mô tả** | **Số dòng** |
| --- | --- | --- |
| BRANDS | lưu trữ thông tin về các thương hiệu xe đạp | 9 dòng |
| CATEGORIES | lưu trữ thông tin về các danh mục sản phẩm | 7 dòng |
| CUSTOMERS | lưu trữ thông tin về các khách hàng | 1445 dòng |
| ORDER\_ITEMS | lưu trữ thông tin về các mục trong đơn hàng | 4722 dòng |
| ORDERS | lưu trữ thông tin về các đơn hàng | 1615 dòng |
| PRODUCTS | lưu trữ thông tin về các sản phẩm | 321 dòng |
| STAFFS | lưu trữ thông tin về nhân viên | 10 dòng |
| STOCKS | lưu trữ thông tin về tồn kho của sản phẩm | 939 dòng |
| STORES | lưu trữ thông tin về các cửa hàng bán lẻ xe đạp | 3 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. |
| fist\_name | Tên khách hàng |
| last\_name | Họ và tên đệm của khách hàng |
| phone | Số điện thoại khách hàng |
| email | Địa chỉ email khách hàng |
| street | Tên đường |
| city | Thành phố |
| state | Bang đang sống |
| zip\_code | Mã code vận chuyển |

#### Bảng stores

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| store\_id | Mã cửa hàng (Khóa chính) |
| store\_name | Tên cửa hàng |
| phone | Số điện thoại liên hệ của cửa hàng |
| email | Email liên hệ của cửa hàng |
| street | Tên đường |
| city | Tênthành phố |
| state | Tên bang |
| zip\_code | Mã code vận chuyển |

#### Bảng staffs

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| staff\_id | Mã nhân viên (Khóa chính) |
| first\_name | Họ nhân viên |
| last\_name | Tên nhân viên |
| email | Email của nhân viên |
| phone | Số điện thoại của nhân viên |
| active | đánh dấu nhân viên còn làm việc hay không |
| store\_id | Mã cửa hàng (Khóa ngoại trỏ đến stores(store\_id) |
| manager\_id | Mã id của quản lý (khóa ngoại trỏ đến staffs(staffs\_id) |

#### 

#### Bảng orders

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| order\_id | Mã đặt hàng (Khóa chính) |
| customer\_id | Mã khách hàng (Khóa ngoại trỏ đến customers(customer\_id)) |
| order\_status | Trạng thái đặt hàng |
| order\_date | Ngày đặt hàng |
| require\_date | Ngày dự kiến giao hàng |
| shipped\_date | Ngày vận chuyển |
| store\_id | Mã cửa hàng (Khóa ngoại trỏ đến store(store\_id)) |
| staff\_id | Mã nhân viên (Khóa ngoại trỏ đến staffs(staff\_id)) |

#### Bảng brands

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

#### Bảng categories

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

#### Bảng products

|  |  |
| --- | --- |
| **Biến** | **Mô tả** |
| product\_id | Mã sản phẩm (Khóa chính) |
| product\_name | Tên sản phẩm |
| brand\_id | Mã thương hiệu (Khóa ngoại) |
| category\_id | Mã danh mục (Khóa ngoại) |
| model\_year | Năm sản xuất |
| list\_price | Giá sản phẩm |

#### Bảng stock

| **Biến** | **Mô tả** |
| --- | --- |
| order\_id | Mã đặt hàng (Khoái ngoại trỏ đến orders(order\_id)) |
| item\_id | Mã mặt hàng (Khóa chính) |
| product\_id | Mã sản phẩm (Khóa ngoại trỏ trỏ đến bảng products(product\_id)) |
| quantity | Số lượng |
| list\_price | Bảng giá |
| discount | Khuyến mãi |

#### 

#### Bảng order\_items

| **Biến** | **Mô tả** |
| --- | --- |
| order\_id | Mã đặt hàng (Khoái ngoại trỏ đến orders(order\_id)) |
| item\_id | Mã mặt hàng (Khóa chính) |
| product\_id | Mã sản phẩm (Khóa ngoại trỏ trỏ đến bảng products(product\_id)) |
| quantity | Số lượng |
| list\_price | Bảng giá |
| discount | Khuyến mãi |

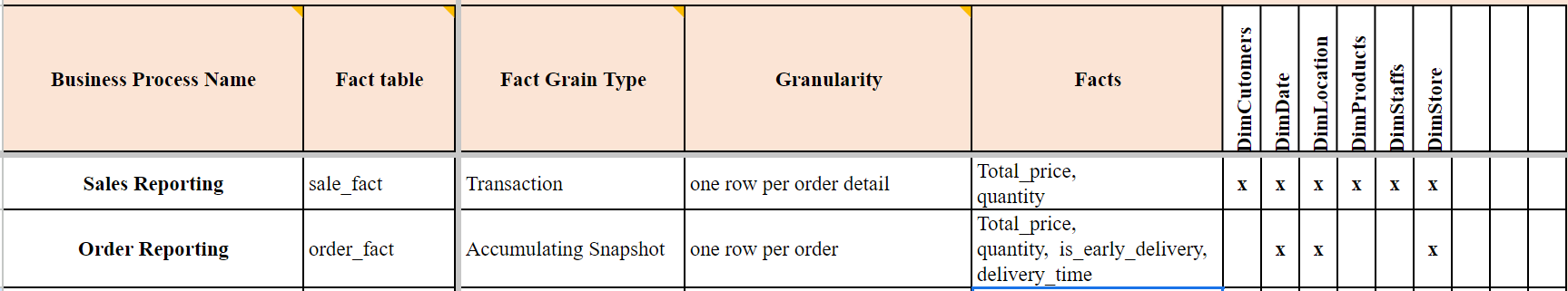
## Giới thiệu các công cụ được sử dụng

* 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.



**Business requirement**

Công ty bán lẻ xe đạp muốn tạo các data mart từ CSDL với các mục đích sau:

**Báo cáo doanh số (sales reporting):** người quản lý muốn theo dõi doanh số bán hàng theo thương hiệu (brands), danh mục sản phẩm (categories), khách hàng (customers), đơn hàng (orders), nhân viên (staffs), sản phẩm (product) và cửa hàng (store) để có thể biết được sản phẩm, thương hiệu, danh mục sản phẩm nào nào bán chạy nhất, nhân viên nào đặt nhiều hóa (order) đơn nhất, đối tượng khách hàng nào mua nhiều nhất, thời gian nào bán hàng được nhất, cửa hàng nào kinh doanh tốt nhất.

**Quản lý đơn hàng (Order and delivery):** người quản lý muốn phân tích xem những khách hàng thường mua đơn hàng lớn, tỉ lệ đơn giao muộn, các cửa hàng, các địa chỉ có tỉ lệ đơn giao muộn cao để có các chiến lược phân khúc khách hàng, chiến lược phân bổ giao hàng hợp lý.

Business requirement có thể được tổng hợp bởi Enterprise Bus Matrix sau:

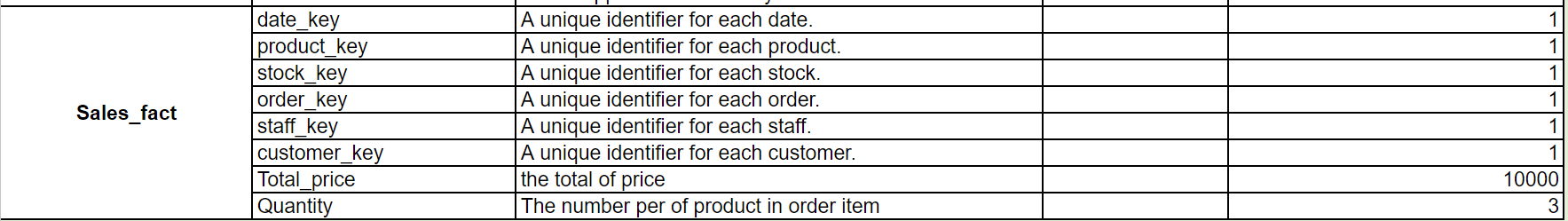
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dimension Bus. Process** | **customers** | **date** | **products** | **staffs** | **stores** | **location** |
| **Báo cáo doanh số** | x | x | x | x | x | x |
| **Quản lý đơn hàng** |  | x |  |  | x | x |

### 2.1.1. Business Process: Sales report

Các câu hỏi cụ thể được đặt ra:

* Tổng doanh thu của chuỗi cửa hàng là bao nhiêu? Doanh thu của cửa hàng ở từng khu vực?
* Tổng số lượng xe đạp bán ra là bao nhiêu? Số lượng bán ra của mỗi khu vực?
* Tổng số tiền của từng khu vực theo mỗi quý?
* Top những sản phẩm có giá trị bán ra cao nhất?
* Top những nhân viên xuất sắc nhất? (tính theo tổng giá trị sản phẩm bán ra)

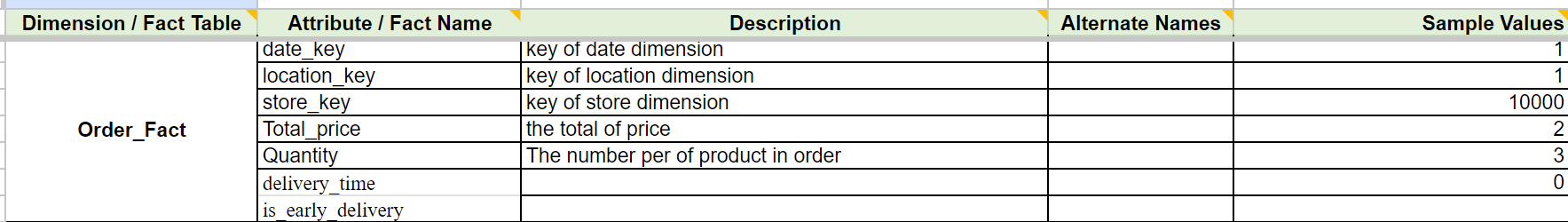
Bảng **Sales\_fact**



### 2.1.2. Business Process: Order and delivery

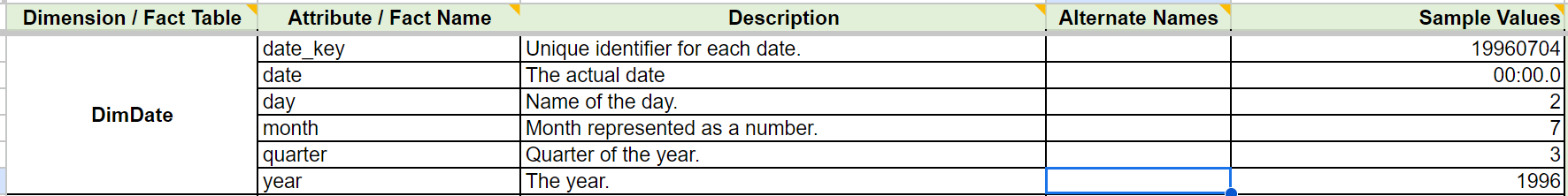
Các câu hỏi cụ thể được đặt ra:

* Tổng số đơn hàng của chuỗi cửa hàng là bao nhiêu? Số lượng đơn hàng của mỗi cửa hàng?
* Tổng thời gian giao hàng của chuỗi chuỗi cửa hàng và của mỗi cửa hàng?
* Số lượng đơn hàng theo từng quý của từng cửa hàng và của mỗi cửa hàng?
* Tổng giá trị sản phẩm bán ra của chuỗi chuỗi cửa hàng và của mỗi mỗi cửa hàng theo từng tháng là bao nhiêu?

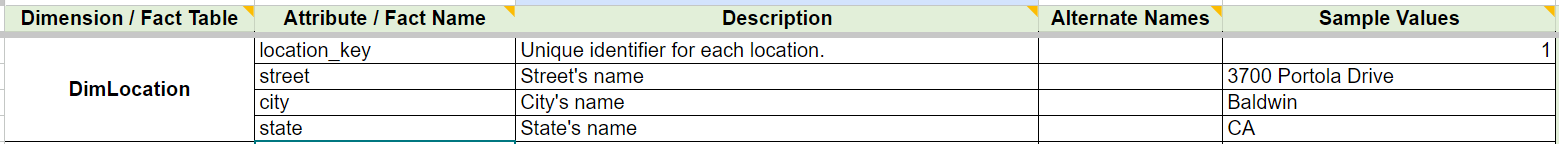


## 2.2. Xây dựng các bảng Dimension

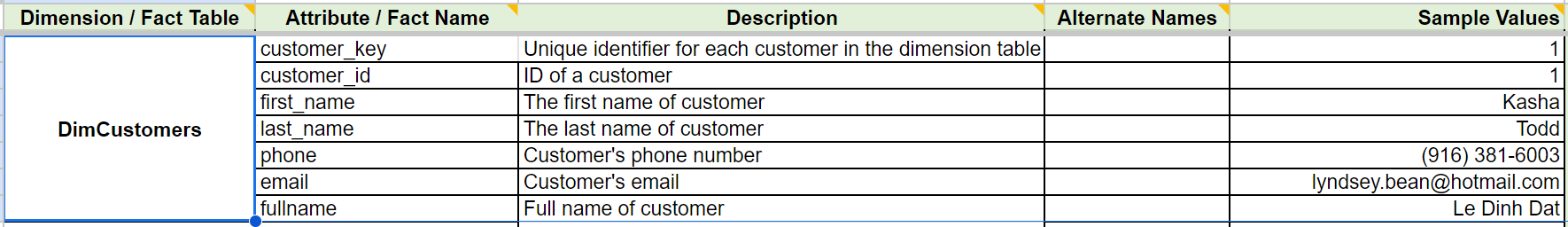
### 2.2.1. DimDate



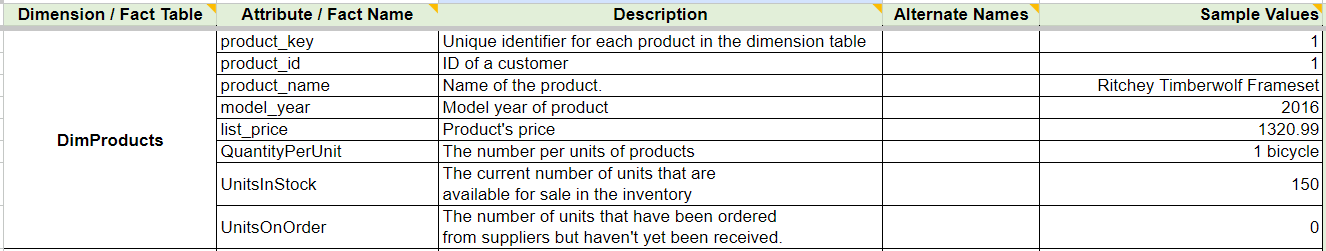
### 2.2.2. DimLocation



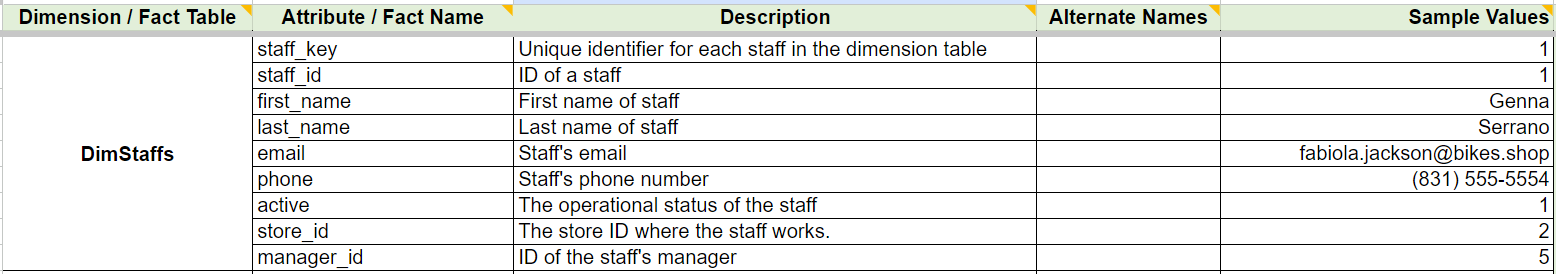
### 2.2.3. DimCustomers



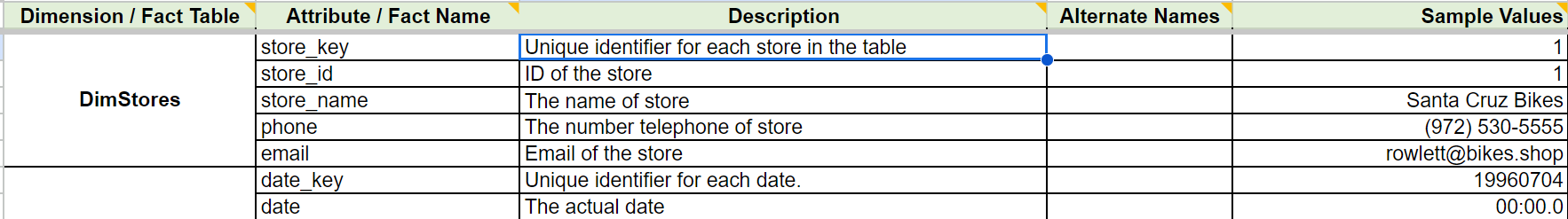
### 2.2.4. DimProducts



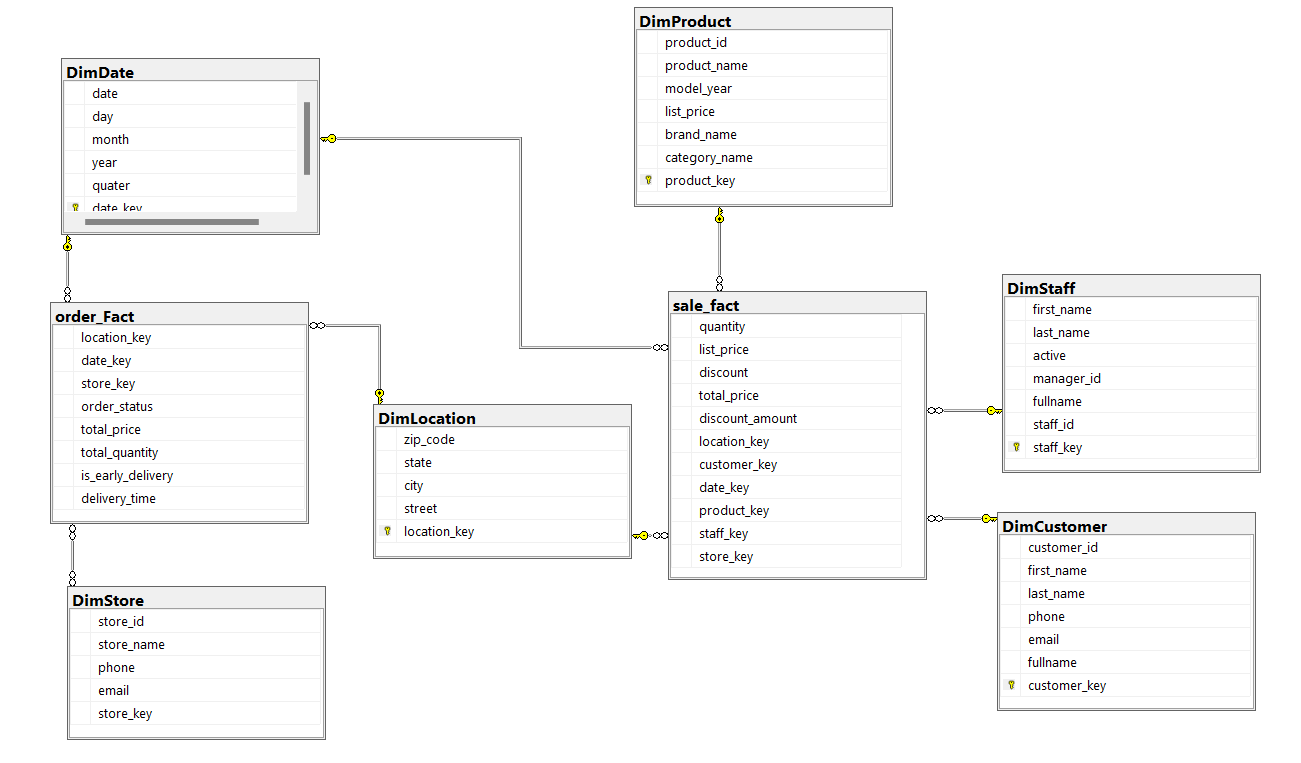
### 2.2.5. DimStaffs



### 2.2.6. DimStores



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



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

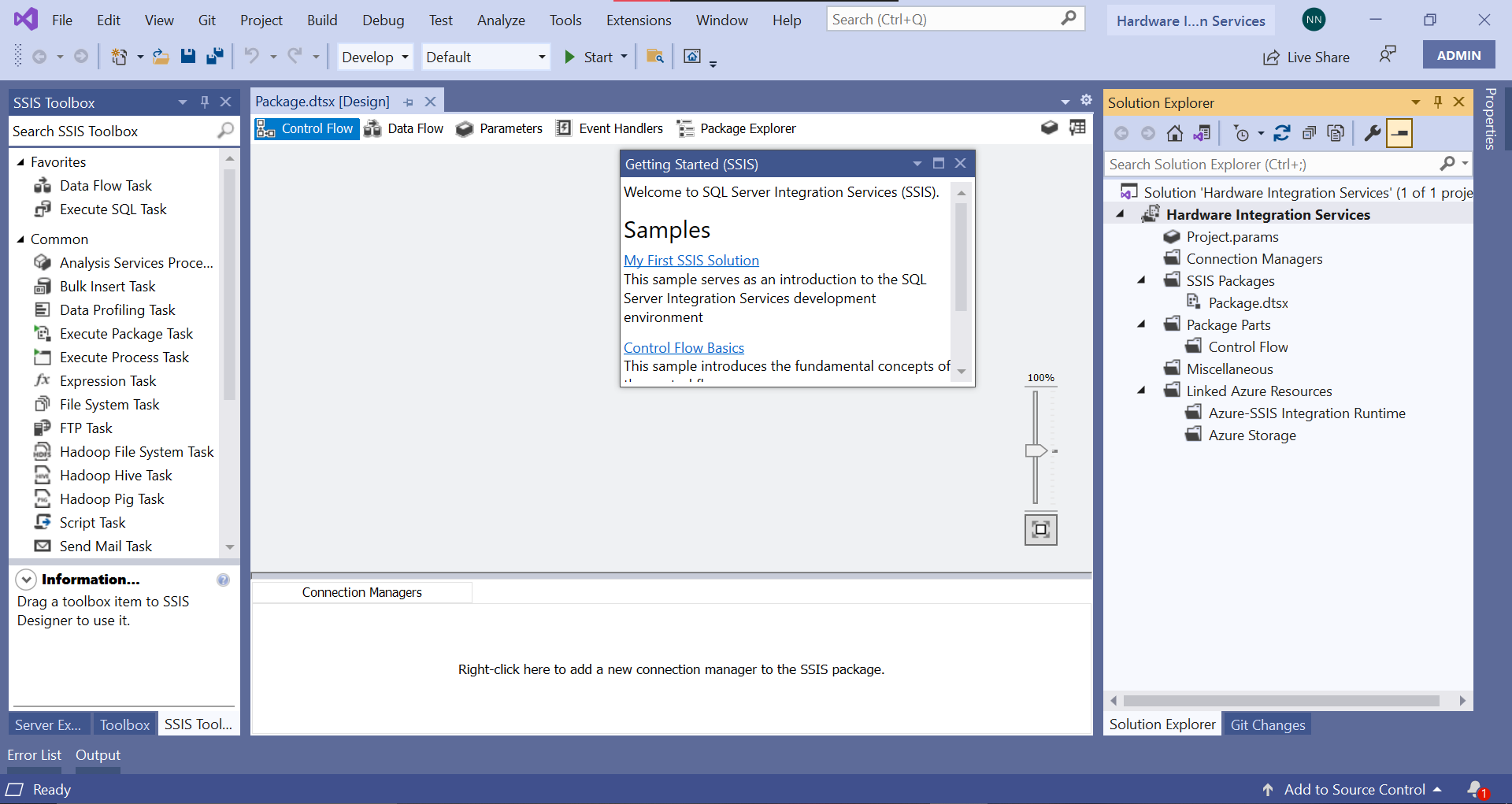
## 3.1. Tạo ETL project và solution

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

a. Chọn **Integration Services Project**

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

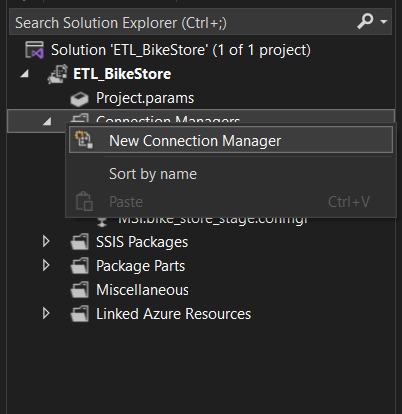
1. Click **OK**



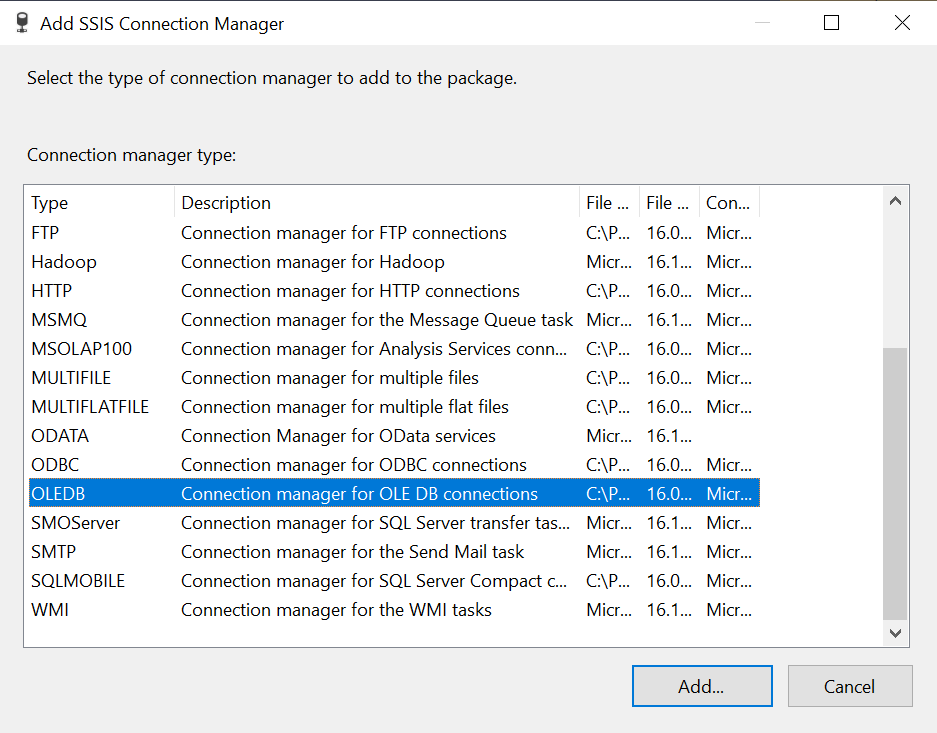
## 3.2. Thiết lập các Connection

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

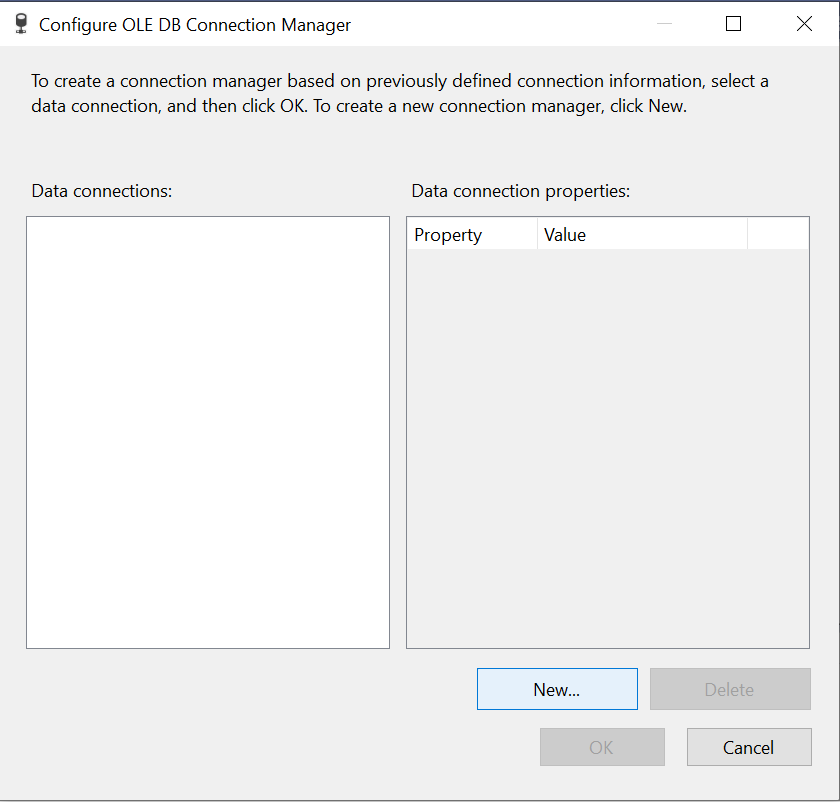
**Connection Manager**.



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



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



Đặt tên cho các thông tin như bên dưới:

Provider: **Microsoft OLE DB Provider for SQL Server**

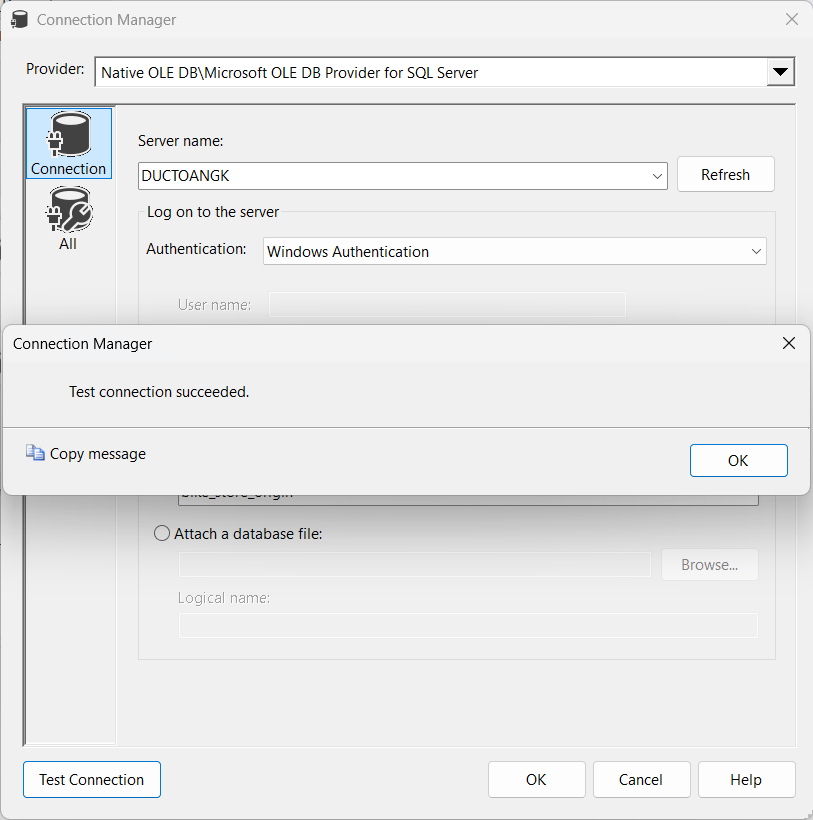
Server Name: **DUCTOANGK**

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

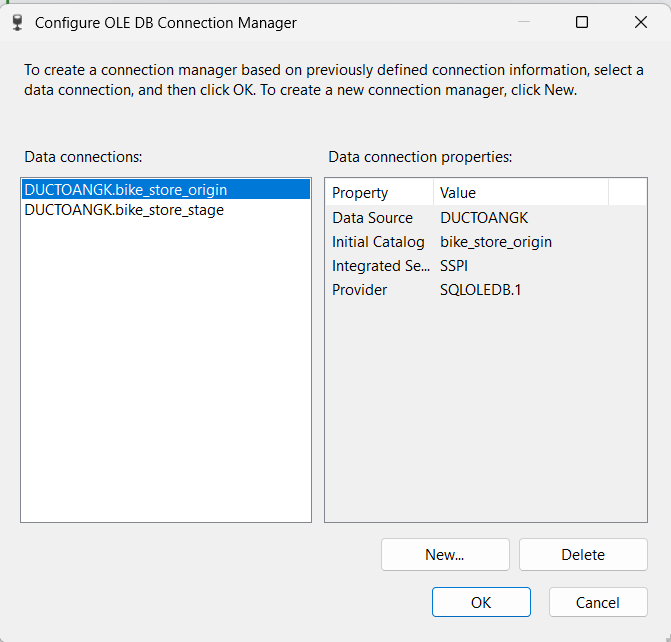
Connect to a database: **bike\_store\_origin**



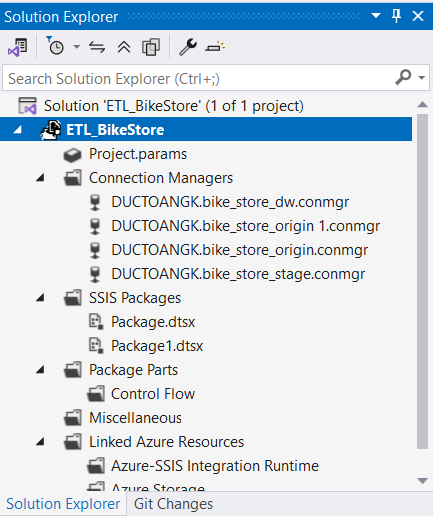
Click **Test Connection** để kiểm tra connection hoạt động, sau đó click **OK** để lưu connection.



1. Trở lại màn hình **Configure OLE DB Connection Manager**, ta được **MSI.bike\_store\_dw** vừa được tạo:

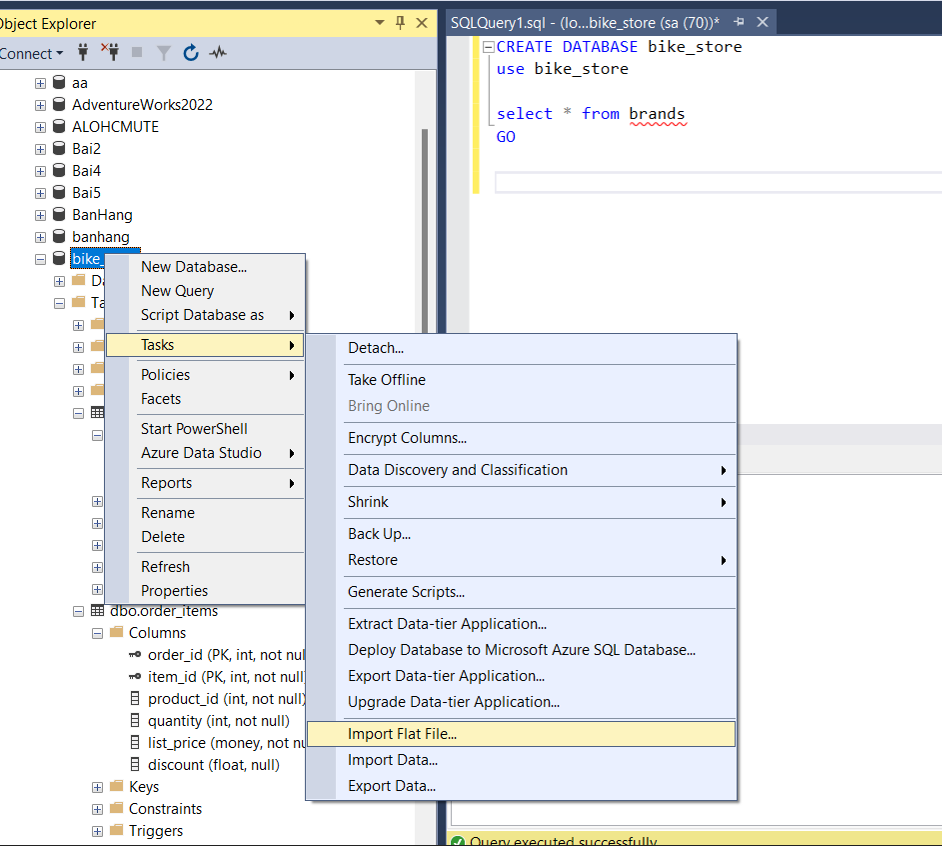


1. Chọn và click **OK** để thêm connection này vào project.
2. Lặp lại các bước trên (bước 1->5) cho bike\_store\_stage vaf bike\_store\_dw.

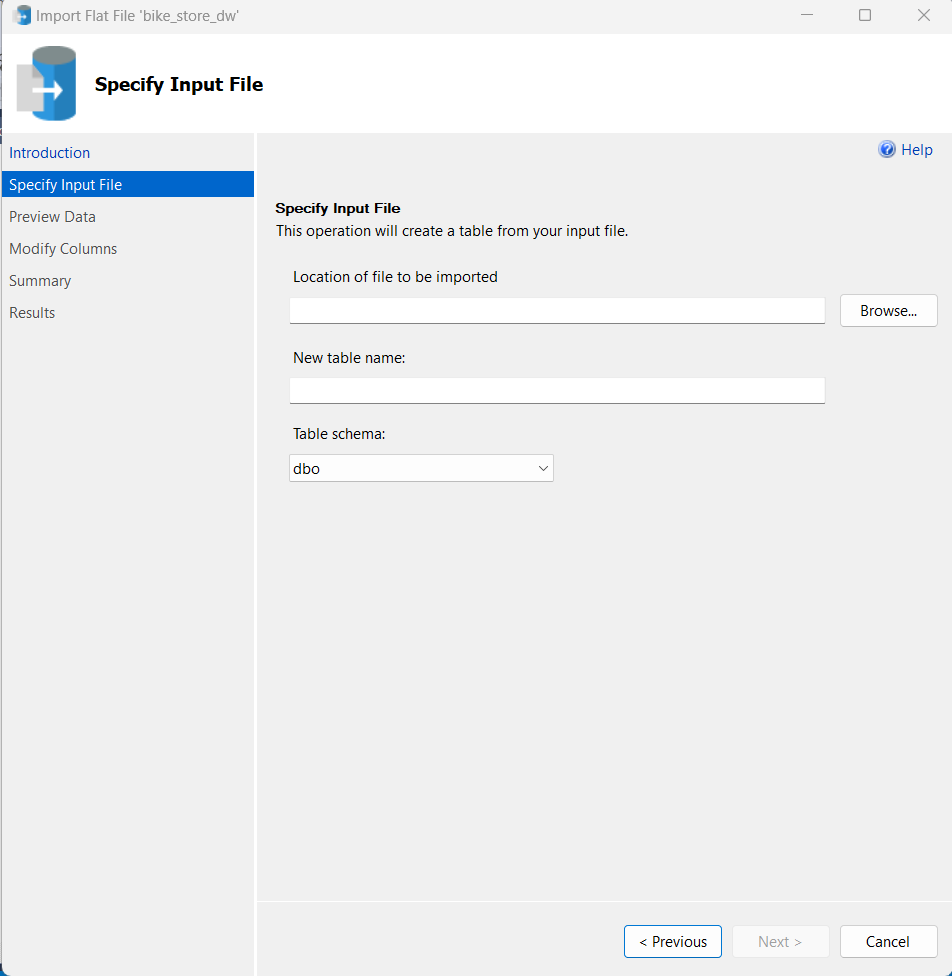


## 3.3. Import Dữ liệu từ file csv vào database bike\_store\_origin

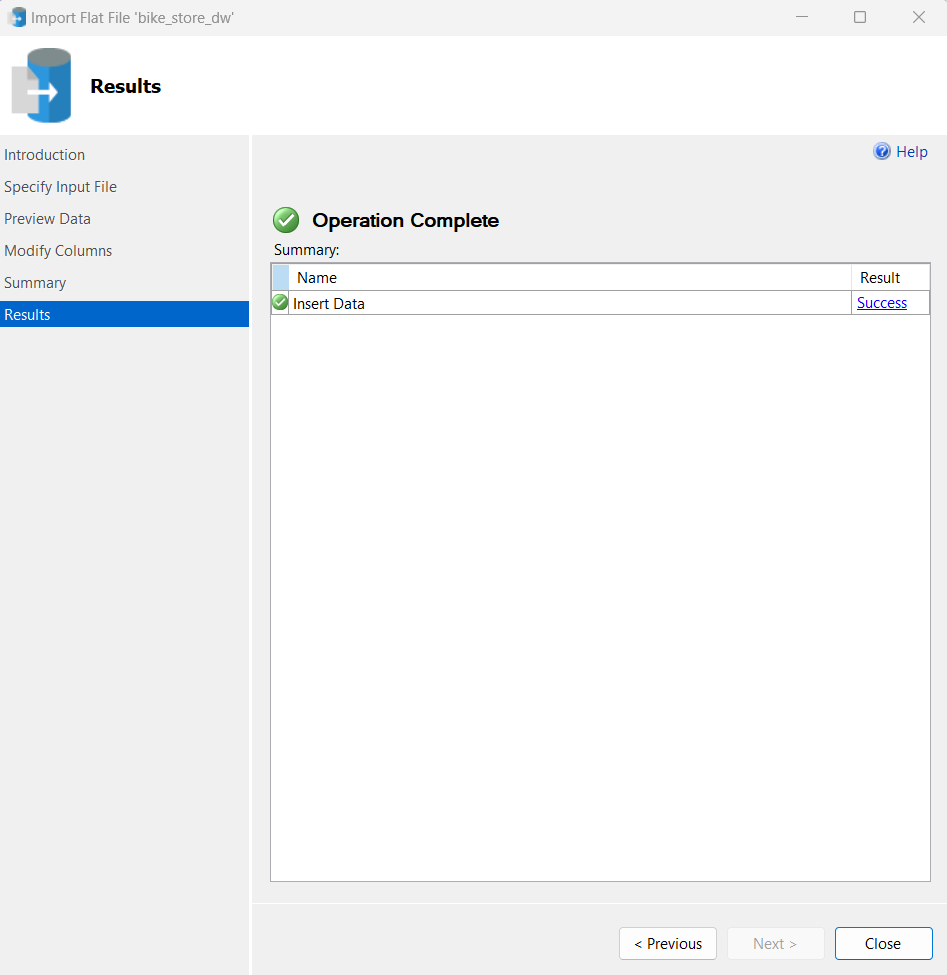
1. Click chuột phải vào bike\_store\_prigin trong phần task chọn import flat file



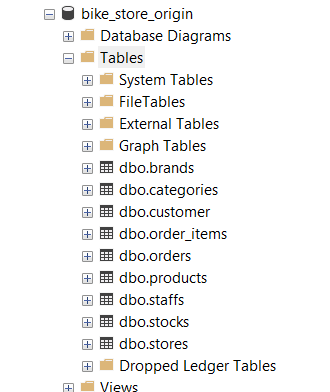
1. Chọn đường dẫn đến file csv, đặt tên cho bảng và bấm **Next**

****

1. Thêm kiểu dữ liệu, đặt khóa chính và đặt thuộc tính Allow Nulls cho các cột. Chọn **Next** để tiếp
2. Nhấn **Finish** để hoàn thành việc import file csv vào sql server.

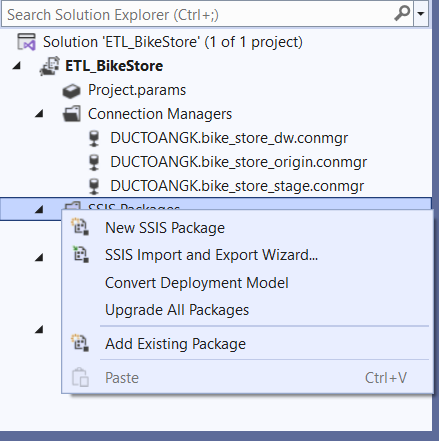


1. Thực hiện lại các bước để import các bảng còn lại bào bike\_store\_origin.



## 3.4. Import Dữ liệu vào các bảng trong bike\_store\_stage:

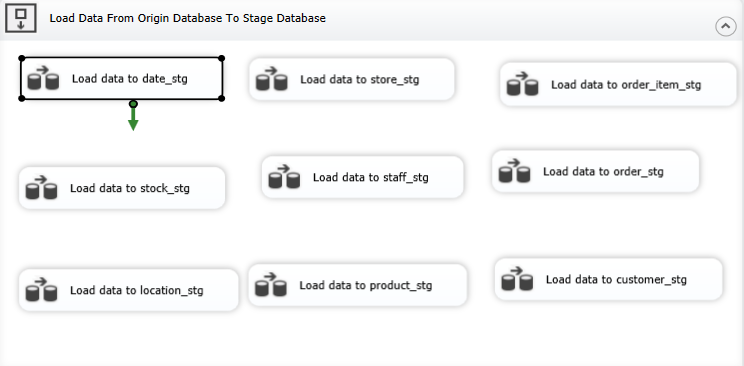
Tạo 1 package mới để thực hiện load dữ liệu:



Tạo một sequences container để chứa các data flow đẩy dữ liệu từ bike\_store\_origin sang bike\_store\_stage

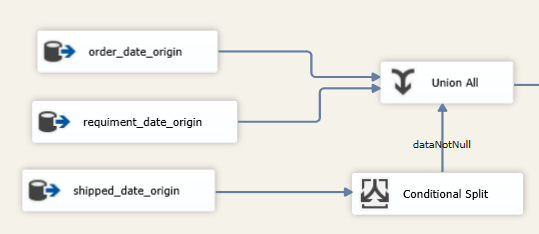
### 3.4.1. Bảng date stage (date\_stg)

Tạo một data flow đặt tên là “Load To date\_stg” đặt bên trong sequence container vừa tạo:

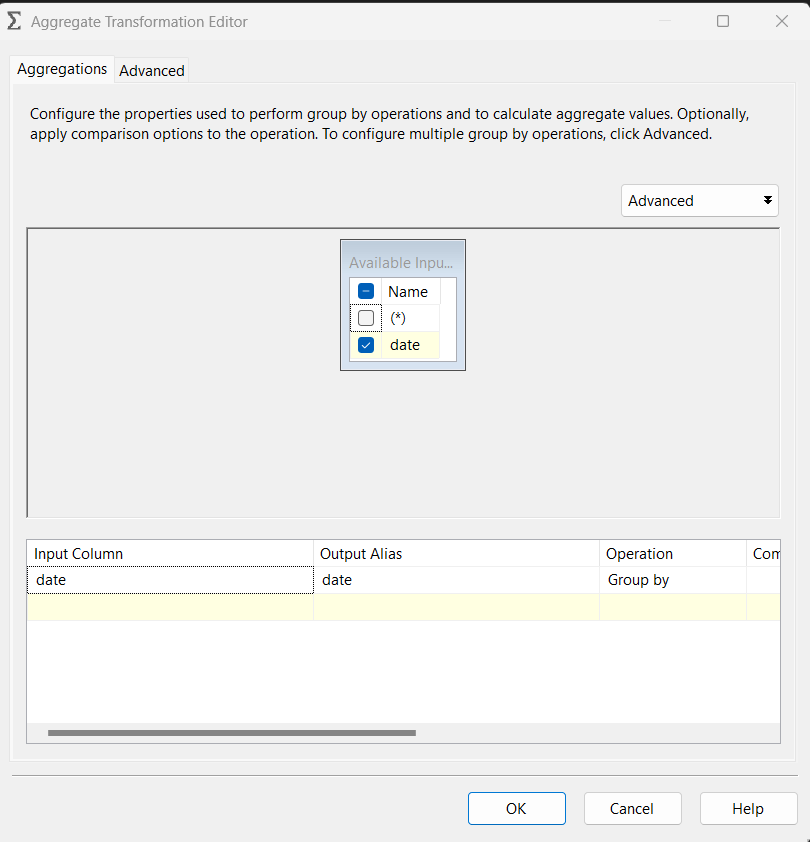


Thiết lập cho data flow:

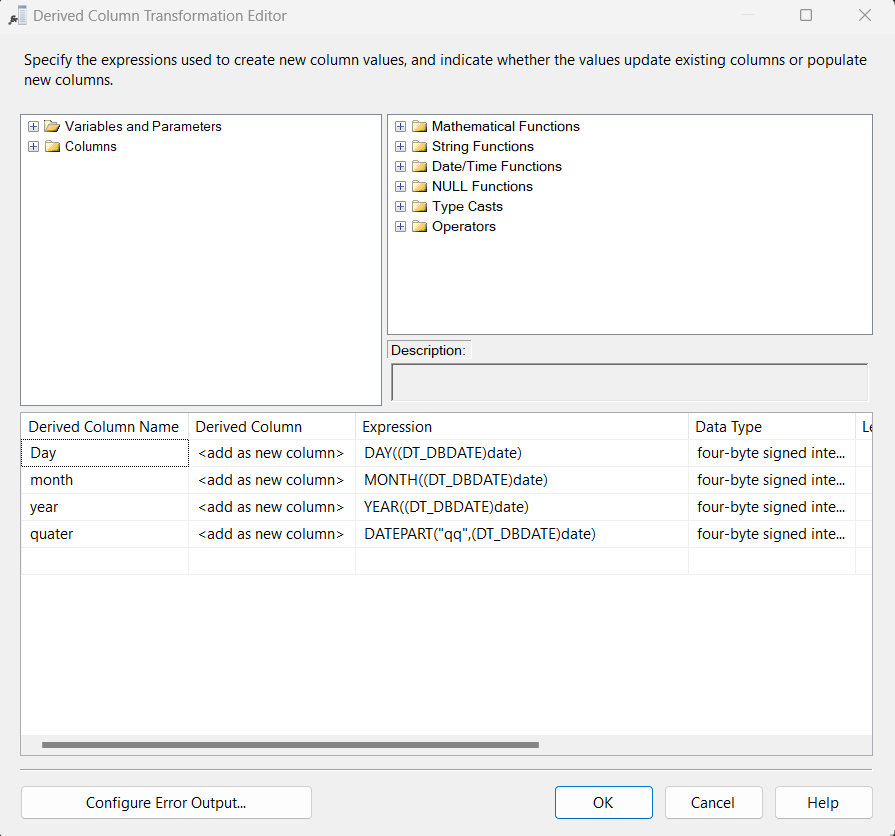
* Từ SSIS toolbox lấy ra 3 Source Assistant lần lượt chứa dữ liệu của cột order\_date, requiment\_date và shipped\_date.
* Vì trong cột shipped\_date có dữ liệu null nên cần sử dụng conditional Split để chia dữ liệu thành tập null và tập không null, sau đó dùng Union All để nối tất cả dữ liệu với nhau, tạo thành tập dữ liệu chứa tất cả các ngày xuất hiện trong database.



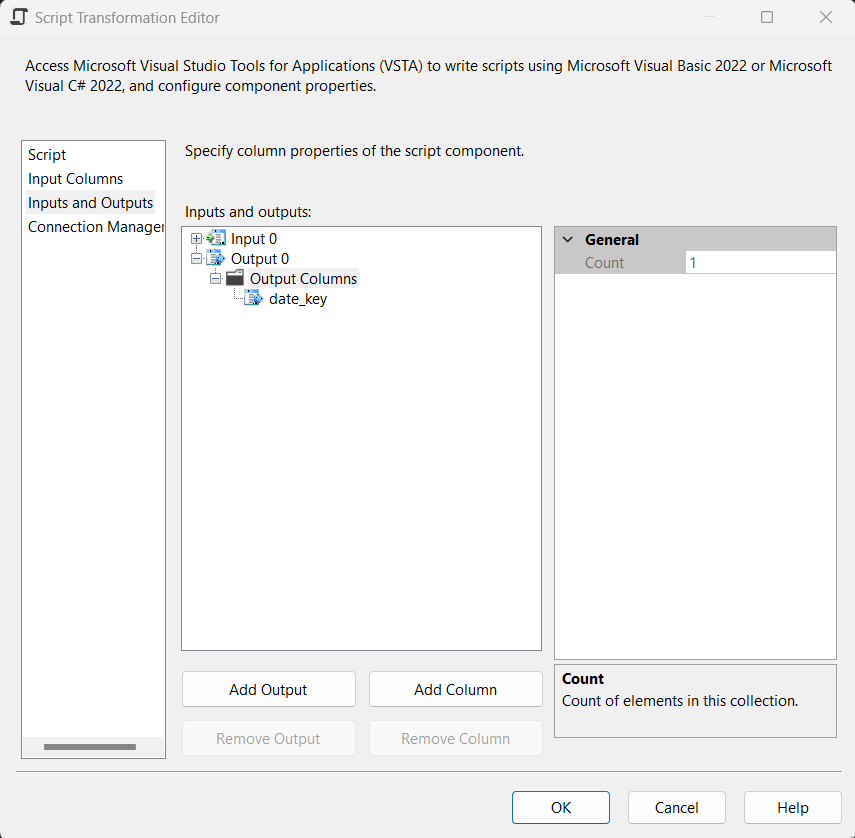
* Sử dụng aggregate để group by cột date nhằm xóa đi những dữ liệu bị trùng.



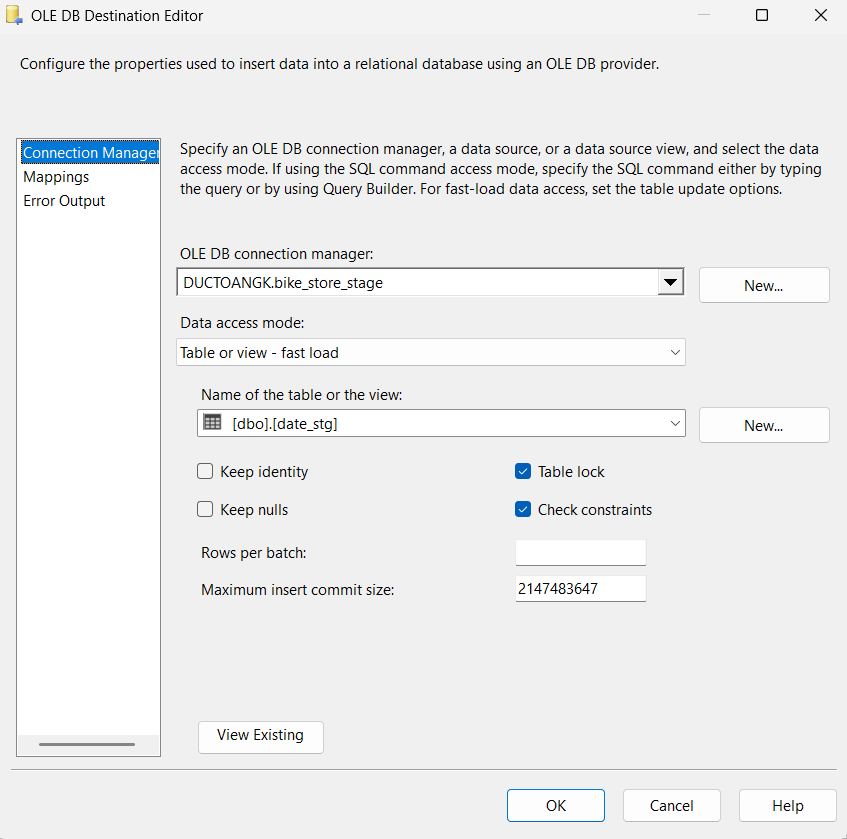
* Sử dụng delivery column nhằm thêm các cột day, month, quater, year bằng cách sử dụng các câu lệnh để tách dữ liệu từ cột date.



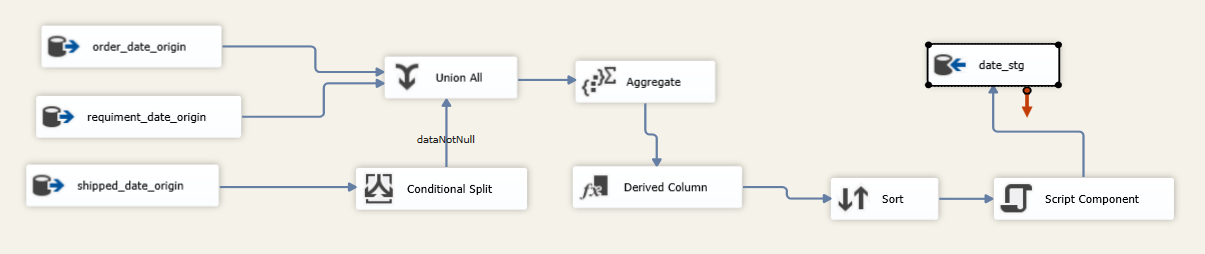
Dùng Script Component để tạo cột key tự động tăng cho bảng date\_stg.

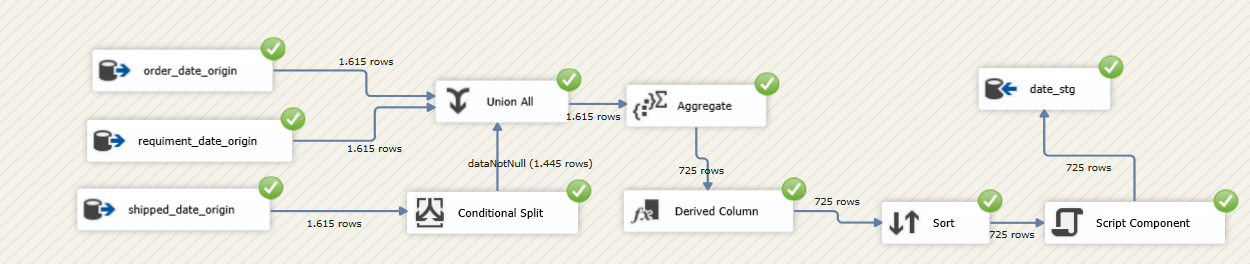


* Lấy Destination Assistant để lưu dữ liệu bảng date\_stg.



Toàn bộ quá trình đẩy dữ liệu từ bảng nguồn sang bảng date\_stg:



Thực thi:  


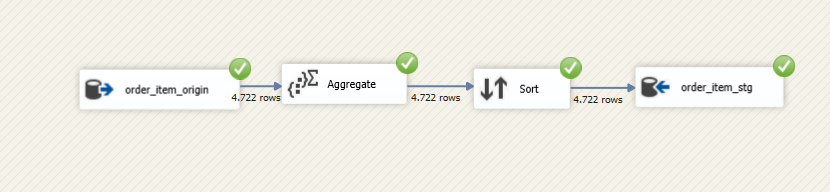
Thực hiện tương tự để đưa dữ liệu từ các bảng còn lại của bike\_store\_origin sang bike\_store\_stage.

**3.4.2. Bảng store stage (store\_stg)**

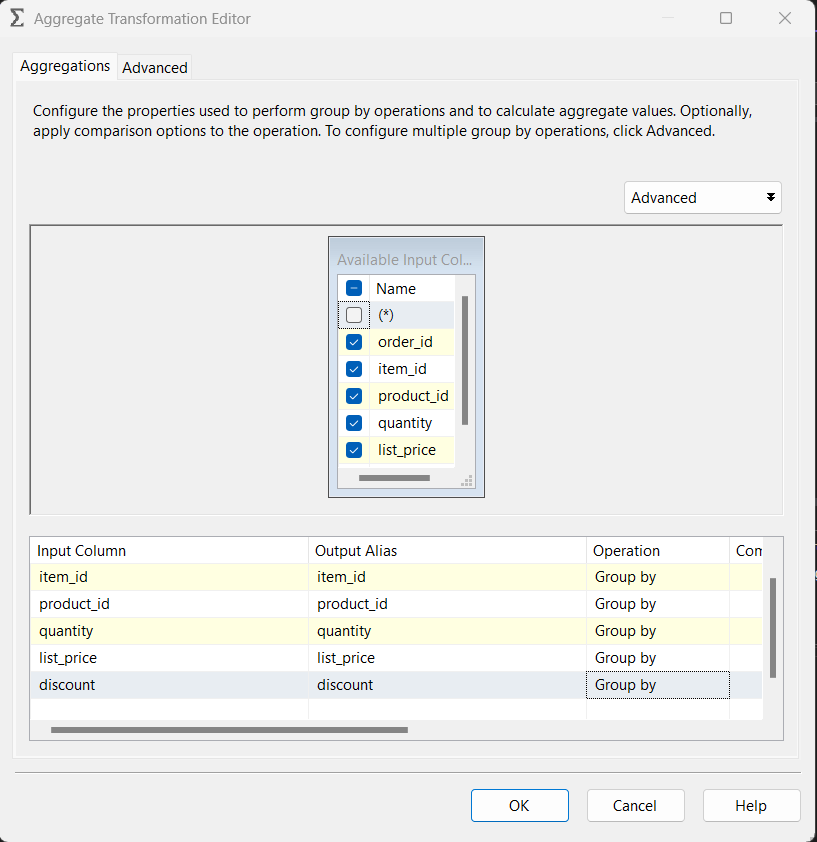
Sử dụng Sort để sắp xếp lại dữ liệu

#### 

### 3.4.3. Bảng order item stage (order\_item\_stg)

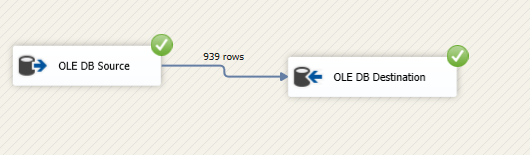


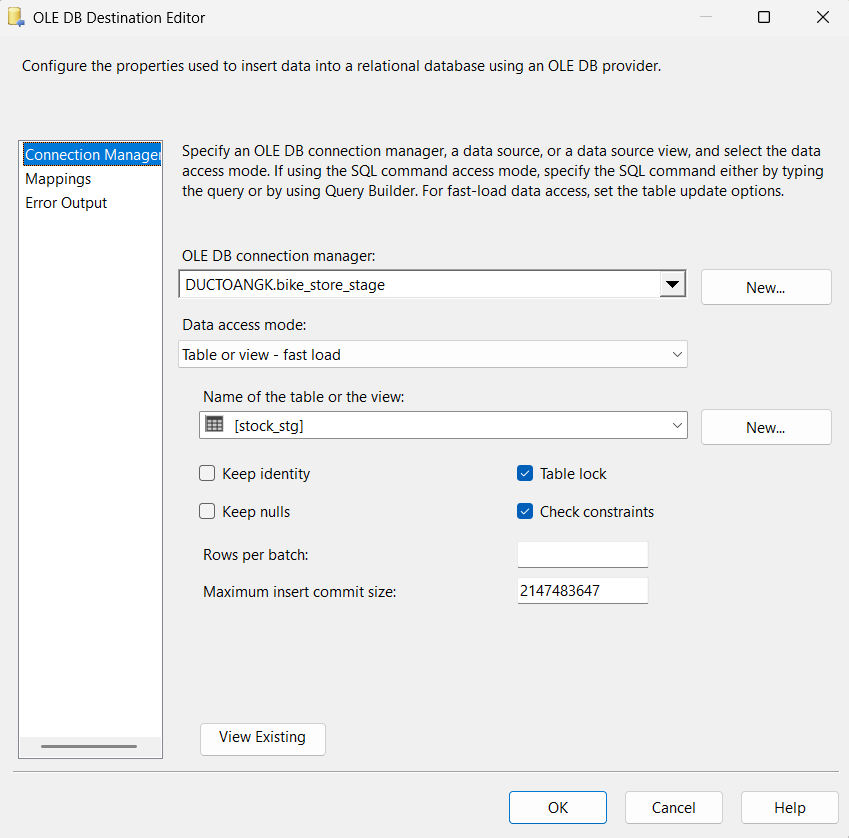
Sử dụng aggregate để group by tất cả cột, những hàng dữ liệu trùng sẽ bị loại bỏ sau bước này.



### 3.4.4.Bảng stock stage (stock\_stg)

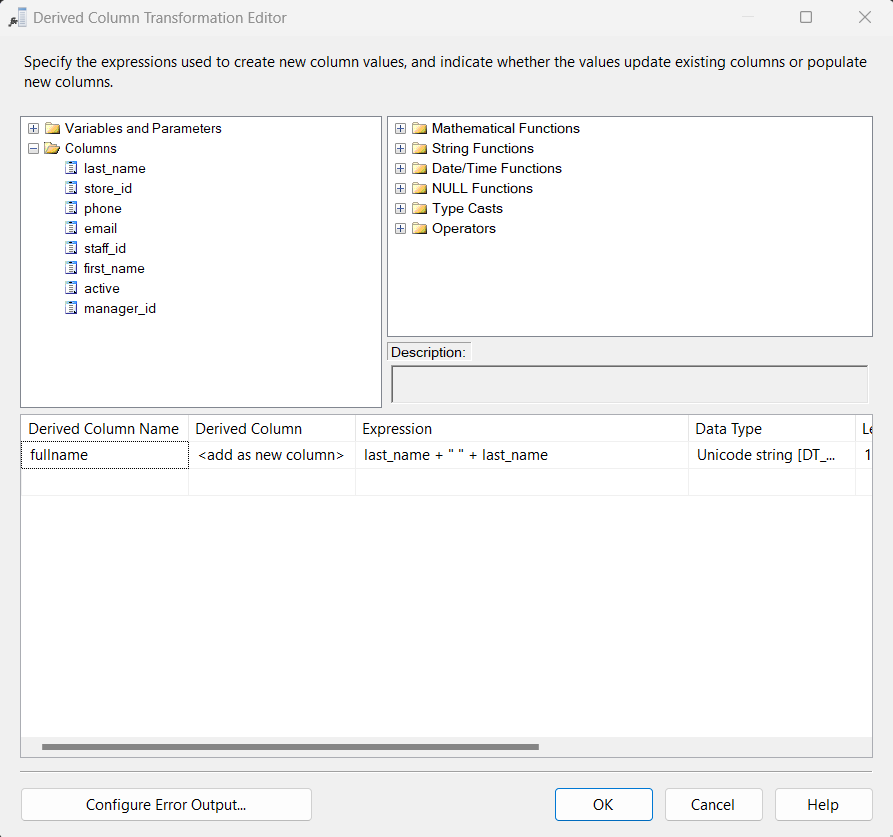
Đưa thẳng dữ liệu từ bảng stock trong dữ liệu nguồn vào bảng stock\_stg.





### 3.4.5. Bảng staffstage (staff\_stg)

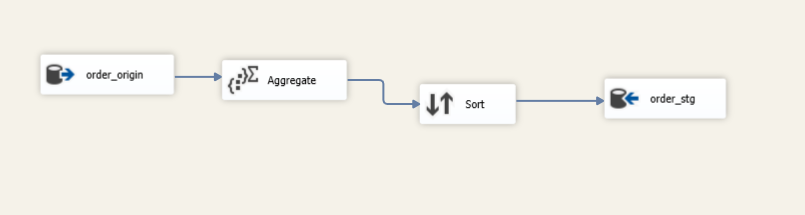
Sử dụng derived column để tạo thêm cột fullname cho bảng staff\_stg.



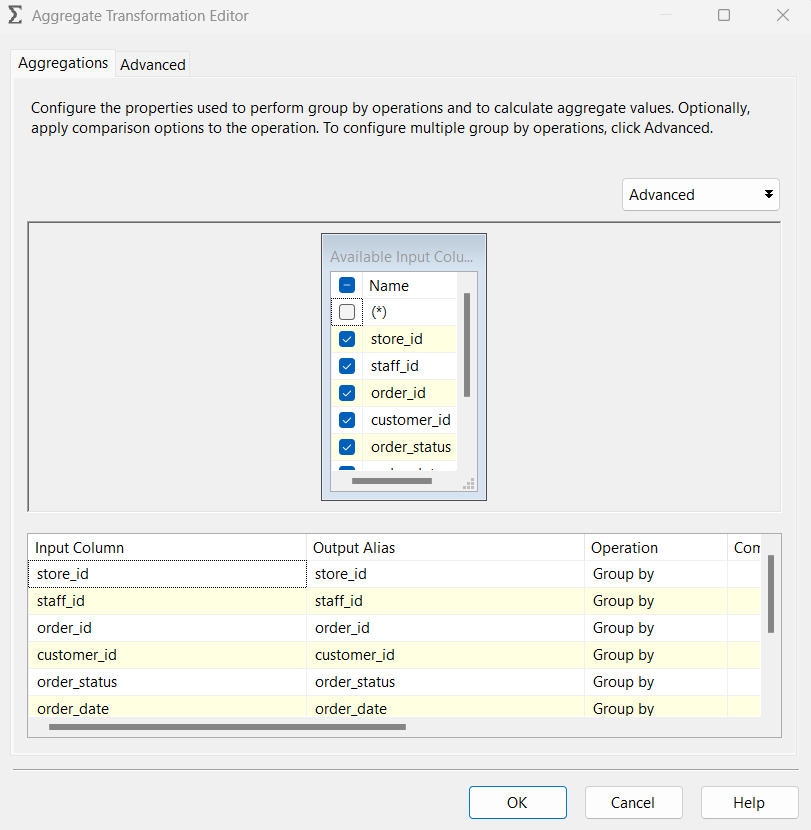
Thực hiện sắp xếp và đưa dữ liệu vào stage:



### 3.4.6. Bảng order stage (order\_stg)

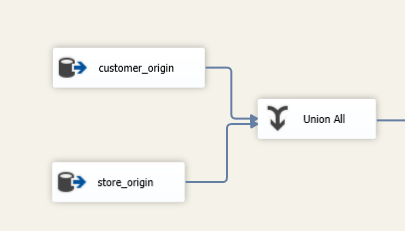


Sử dụng Aggregate group by tất cả các cột để loại bỏ tất cả các dòng dữ liệu trùng lặp

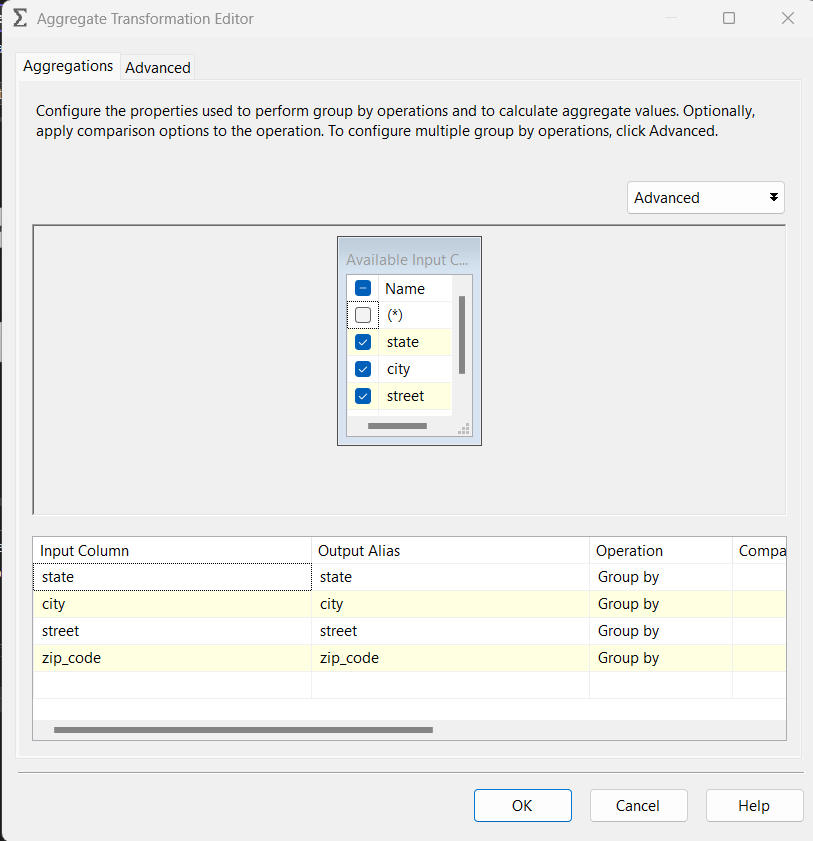


### 3.4.7. Bảng location stage (location\_stg)

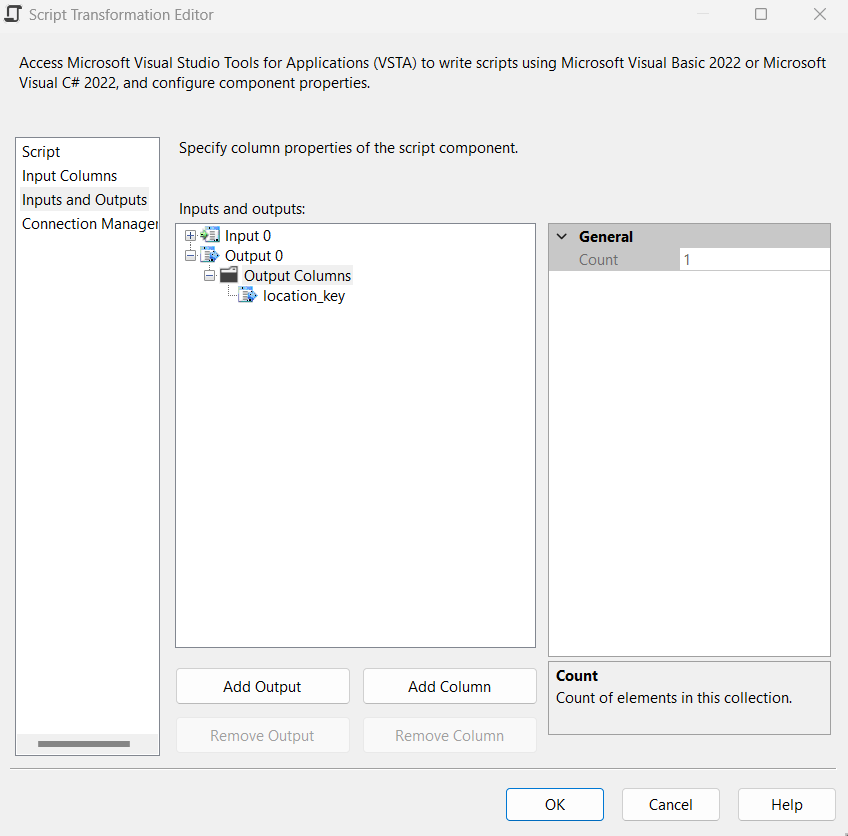
Lấy 2 Source Assistant để chứa dữ liệu địa chỉ của khách hàng và địa chỉ cửa hàng, thực hiện Union All để nối dữ liệu lại, ta có dữ liệu mới chứa tất cả địa chỉ trong database.



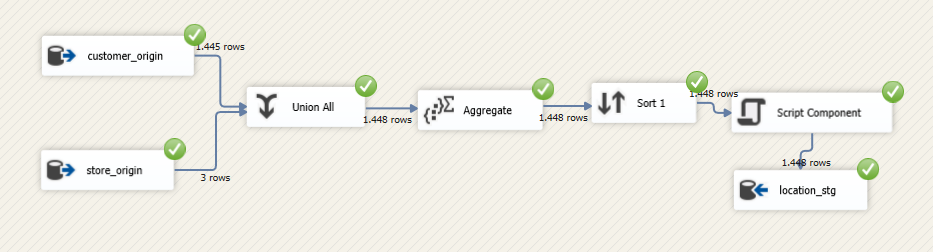
Sau đó dùng aggregate để group by tất cả cột nhằm xóa các dòng dữ liệu bị trùng lặp.



Sử dụng Script component để tạo cột location\_key có giá trị tăng tự động.

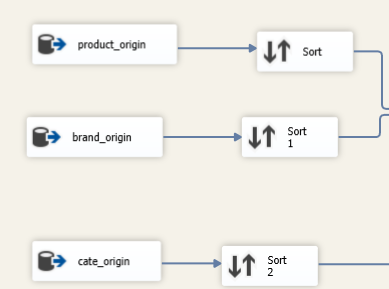


Thực thi:



### 3.4.8. Bảng product stage (product\_stg)

Lấy 3 Source Assistant từ SSIS Toolbox để chứa dữ liệu từ bảng products, brands, categorys.



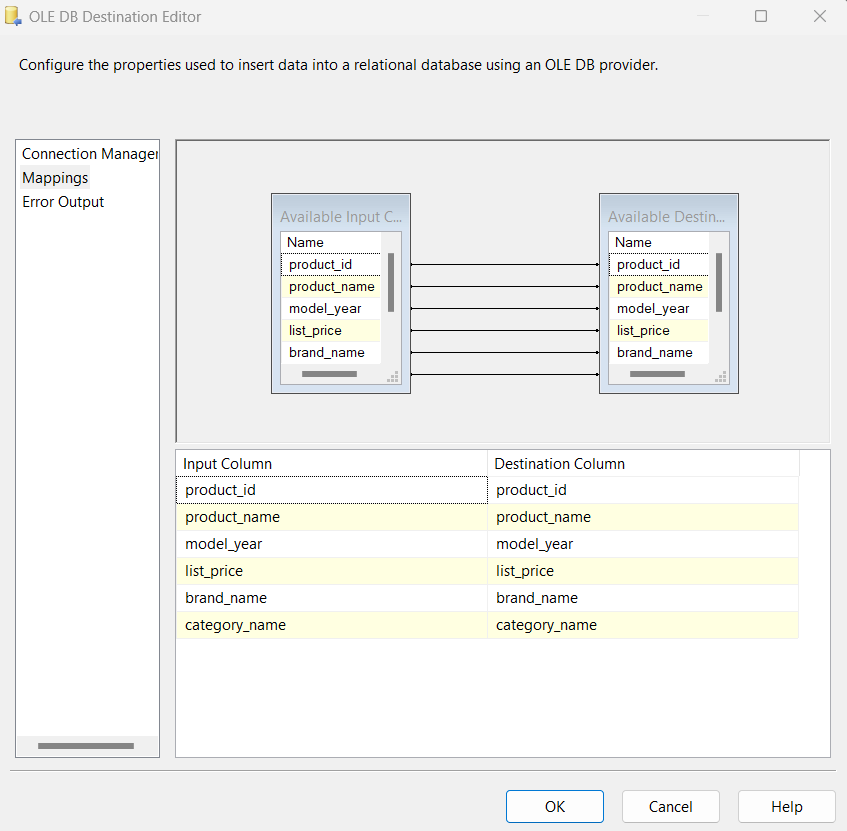
Sử dụng Merge join để lấy tên brand và tên category đưa vào bảng products.



Đưa dữ liệu bảng product mới được tích hợp thêm tên brand và category vào bảng product\_stg trong database stage.

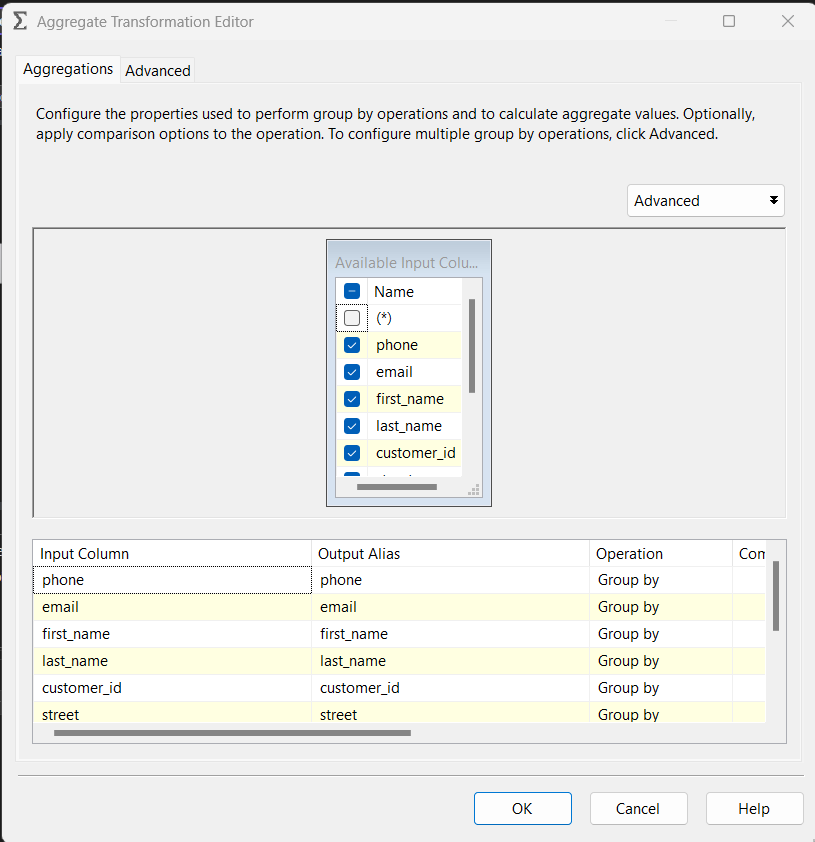


Các cột trong bảng product\_stg.

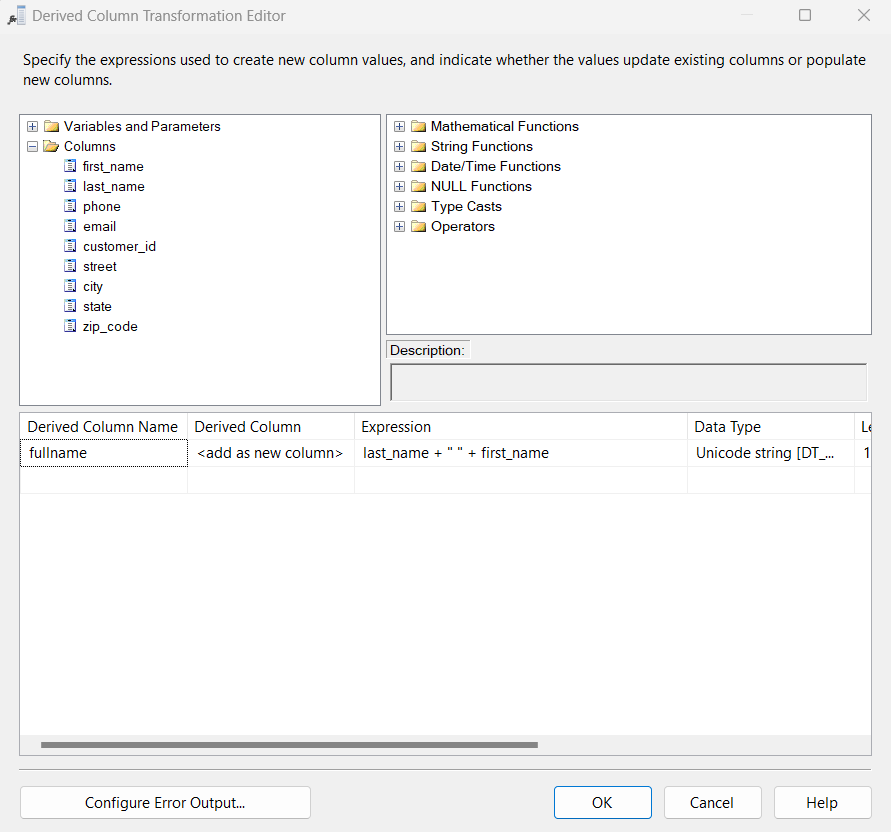


### 3.4.9. Bảng customer stage (customer\_stg)

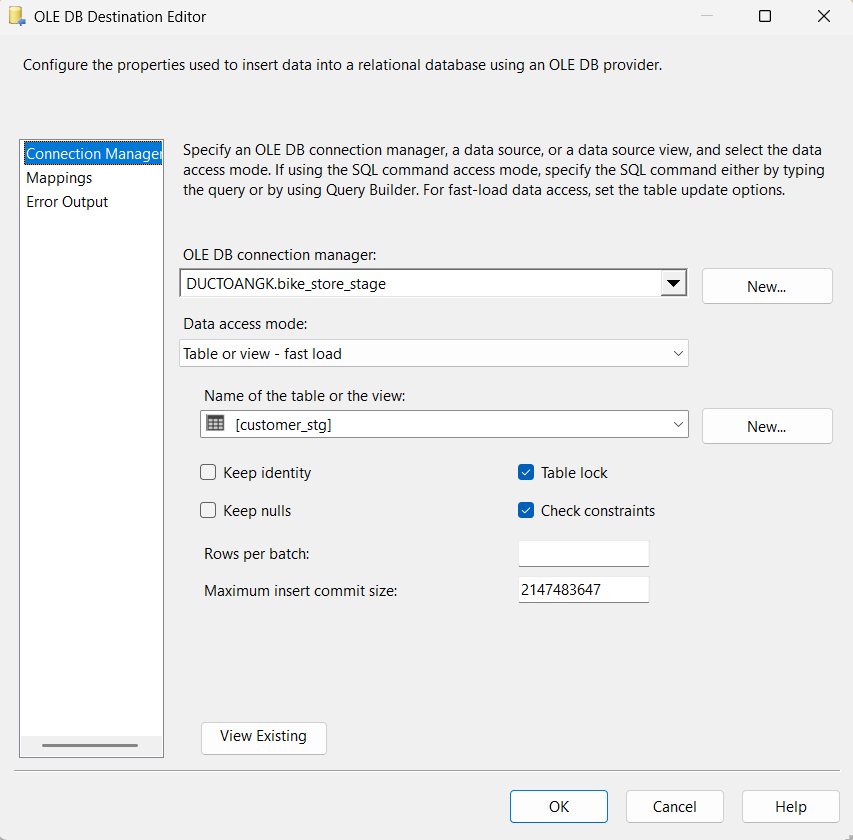
Dùng Aggregate để group by tất cả các cột, các dòng dữ liệu trùng nhau sẽ được xóa.



Sử dụng derived column để tạo thêm cột fullname cho bảng customer.

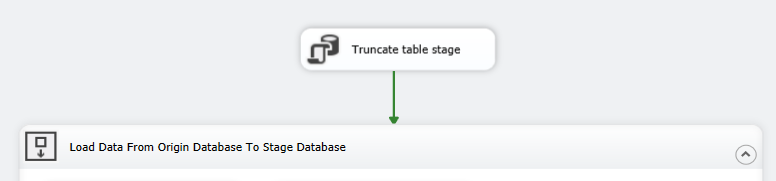


Tạo destination Assistant để đưa dữ liệu vào bảng customer\_stg.

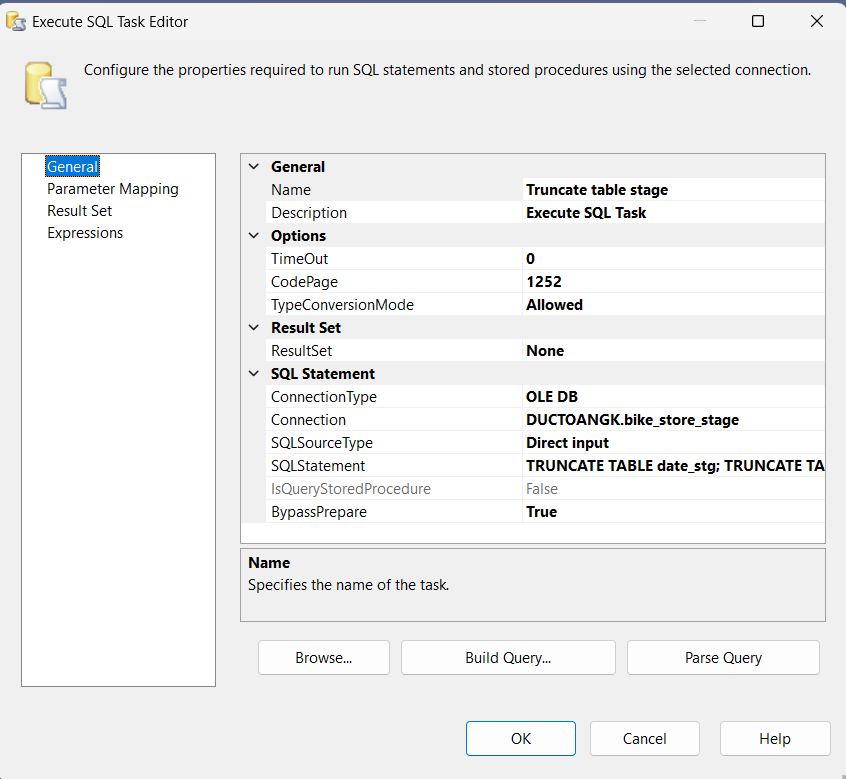


### 3.4.10. Tạo Execute SQL Task để Truncate các bảng trong stage

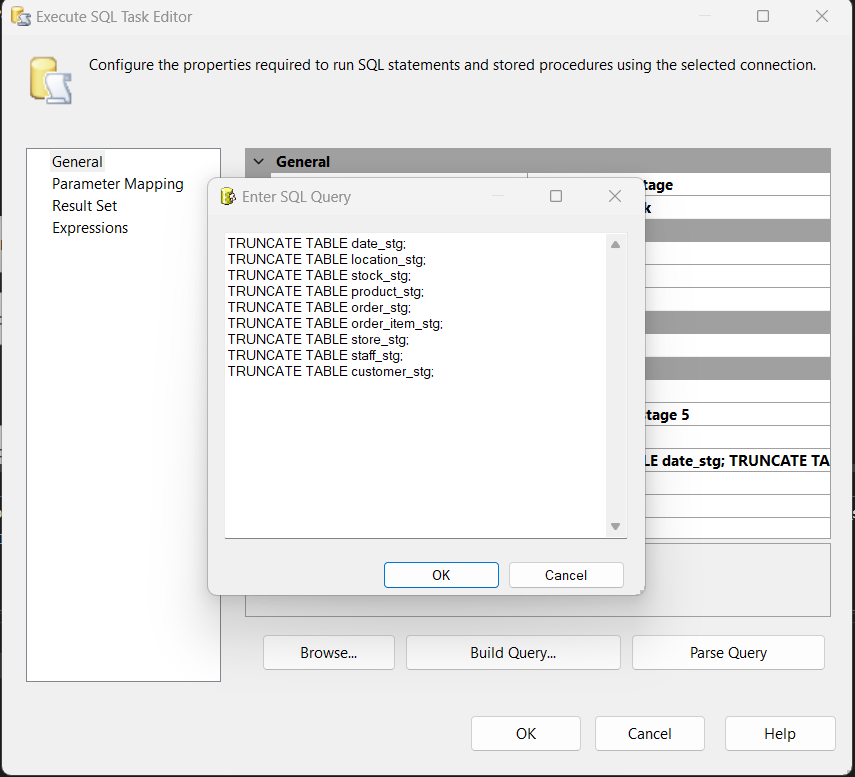
Lấy 1 Execute SQL Task từ SSIS Toolbox và đổi tên thành “Truncate table stage” .



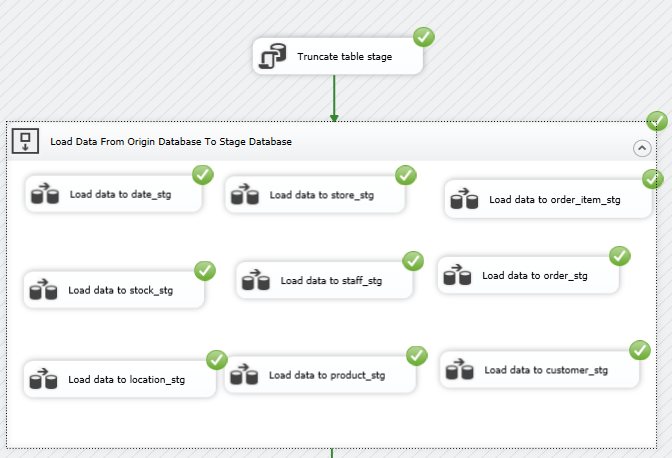
Cấu hình để connect tới bike\_store\_stage.



Trong phần SQL Statement, viết đoạn lệnh sau để truncate tất cả các bảng trong stage.

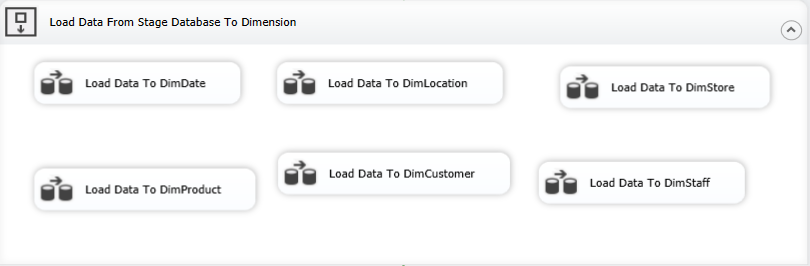


Thực thi phần đưa dữ liệu từ database nguồn sang stage.



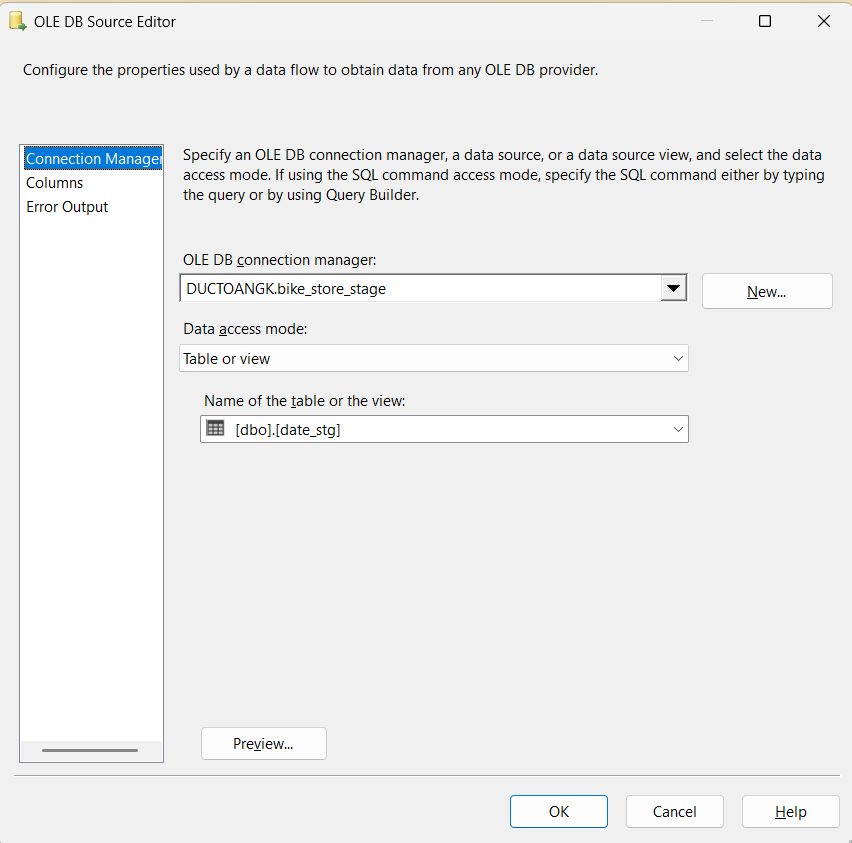
## 3.5. Import dữ liệu từ stage vào các bảng dimension

Tạo một sequence container để chứa các data flow và sửa tên thành “Load Data From Stage Database To Dimension”

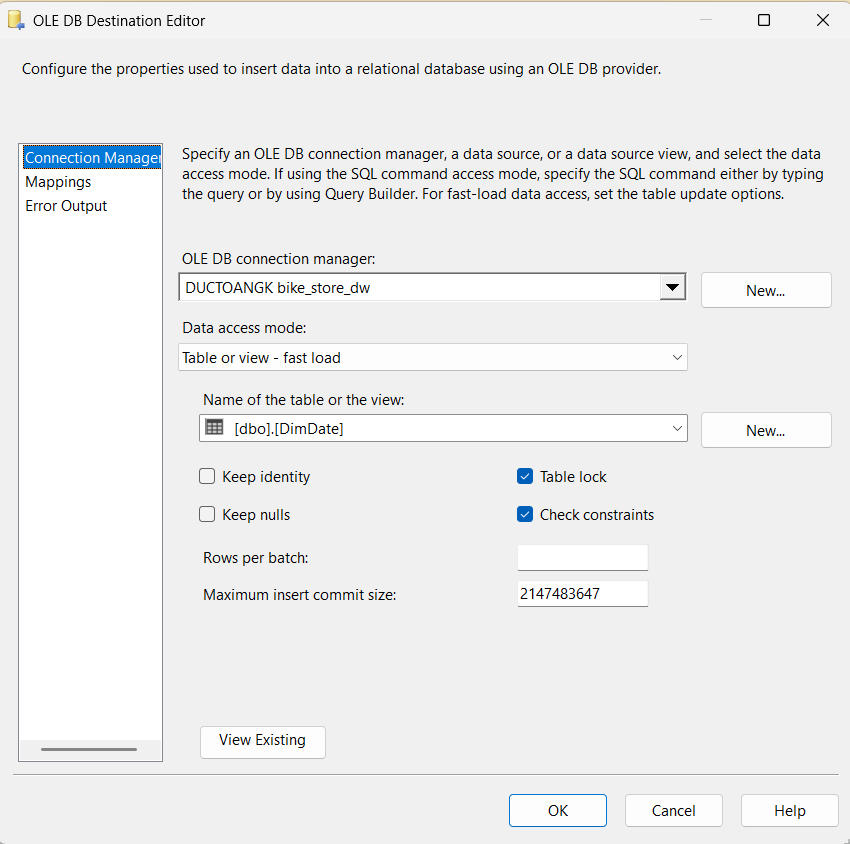


### 3.5.1. Bảng Date Dimension (DimDate)

Sử dụng 1 Source Assistant để lấy dữ liệu từ bảng date\_stage

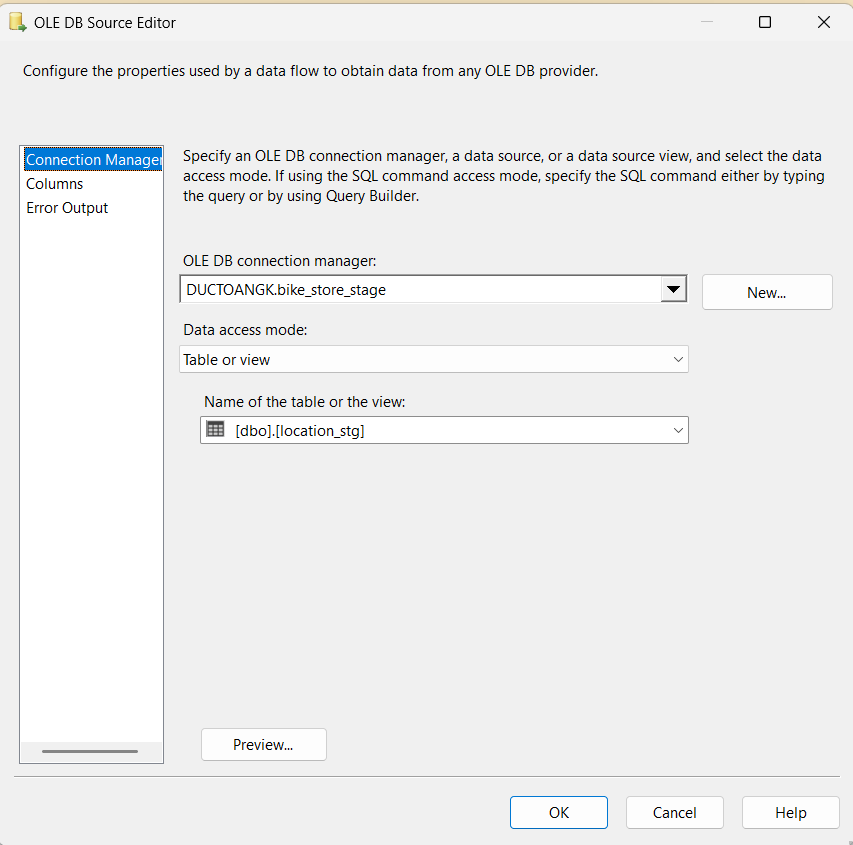


Sử dụng Destination Assistant để lấy dữ liệu lưu vào bảng DimDate.

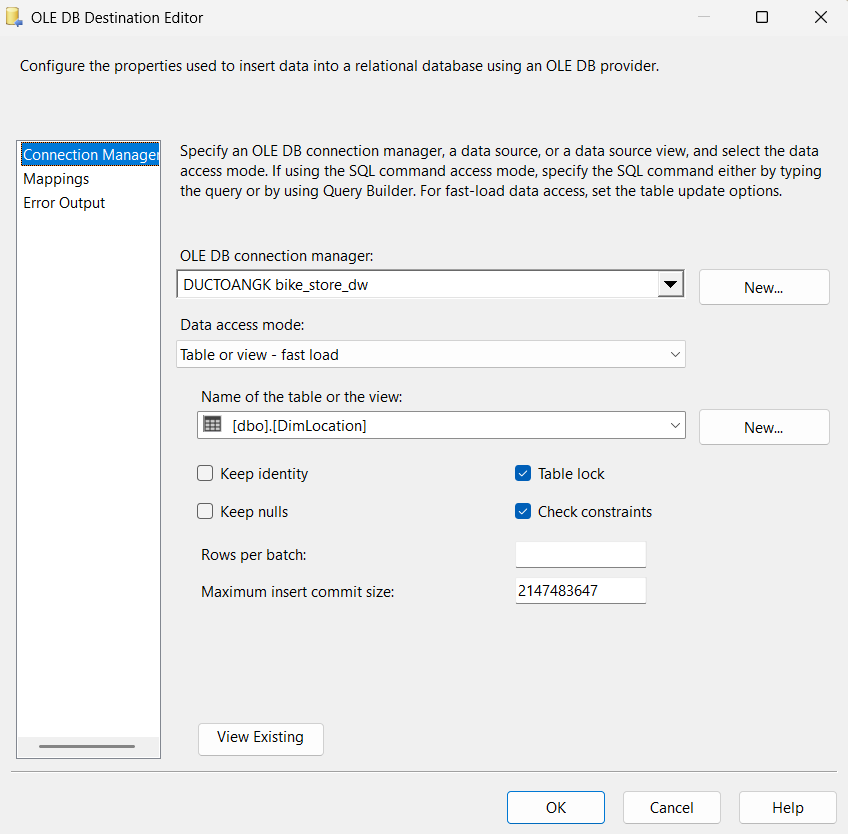


### 3.5.2. Bảng Location Dimension (DimLocation)

Sử dụng 1 Source Assistant để lấy dữ liệu từ bảng location\_stage.

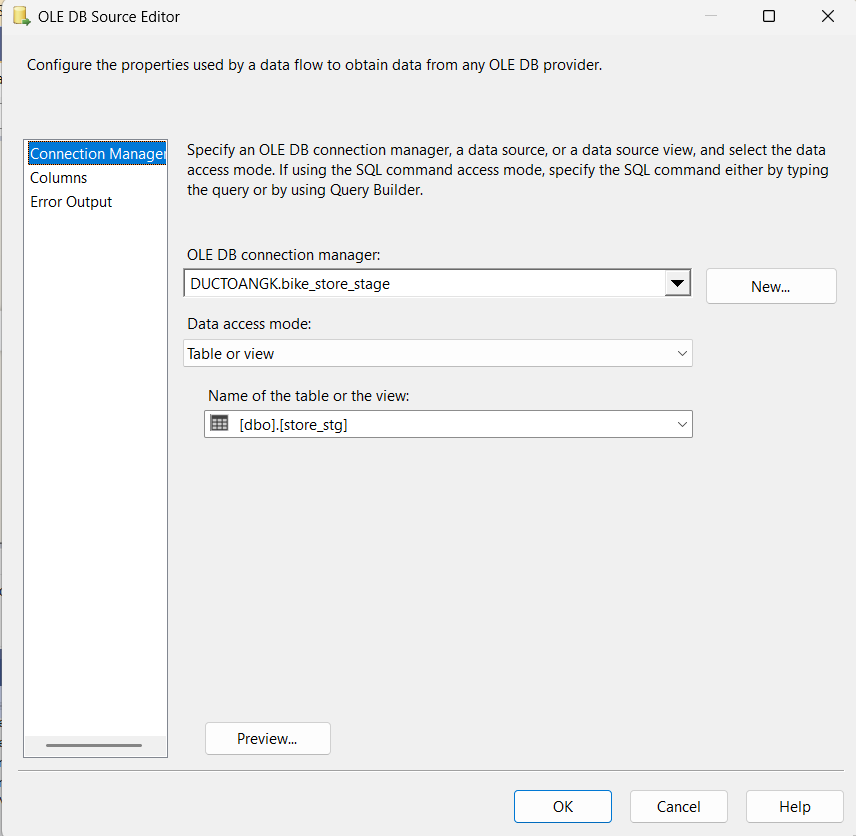


Sử dụng Destination Assistant để lấy dữ liệu lưu vào bảng DimLocation.



### 3.5.3. Bảng Store Dimension (DimStore)

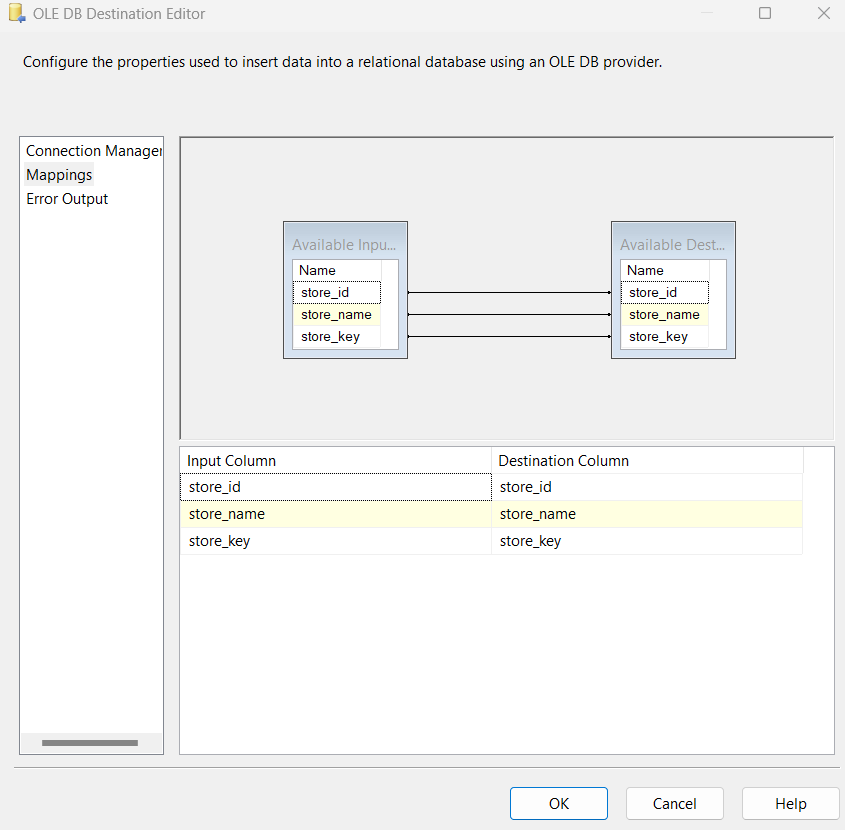
Sử dụng 1 Source Assistant để lấy dữ liệu từ bảng store\_stage.



Sử dụng Script Component để thêm cột store\_key tăng tự động.

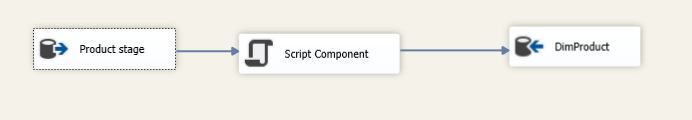


Sử dụng Destination Assistant để lấy dữ liệu sau khi đã tích hợp thêm cột store\_key lưu vào bảng DimStore.

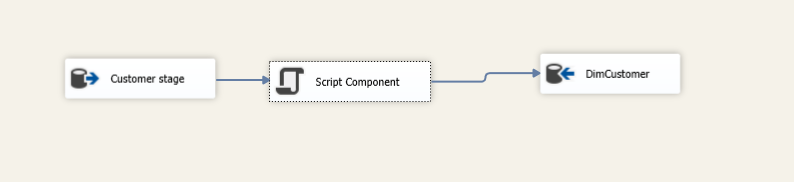


### 3.5.4. Bảng Product Dimension (DimProduct)

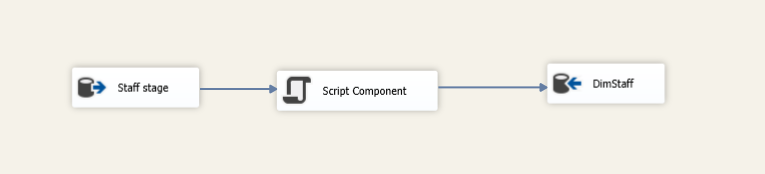
Sử dụng Script Component để thêm cột product\_key tương tự bảng DimStore.



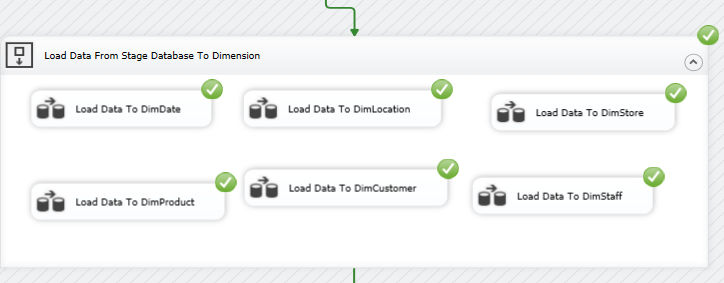
### 3.5.5. Bảng Customer Dimension (DimCustomer)



### 3.5.6. Bảng Staff Dimension (DimStaff)



Thực thi load dữ liệu từ stage vào các bảng dim của datawarehouse

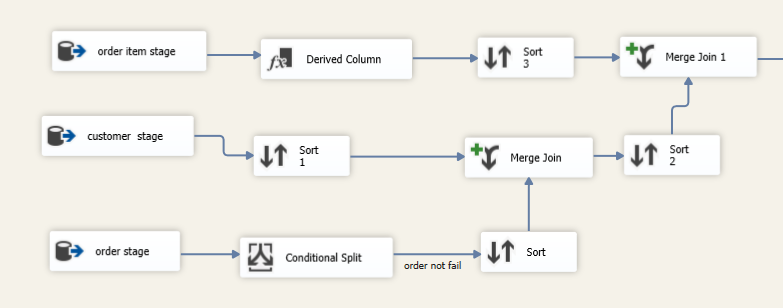


## 3.6. Import dữ liệu từ stage vào các bảng fact

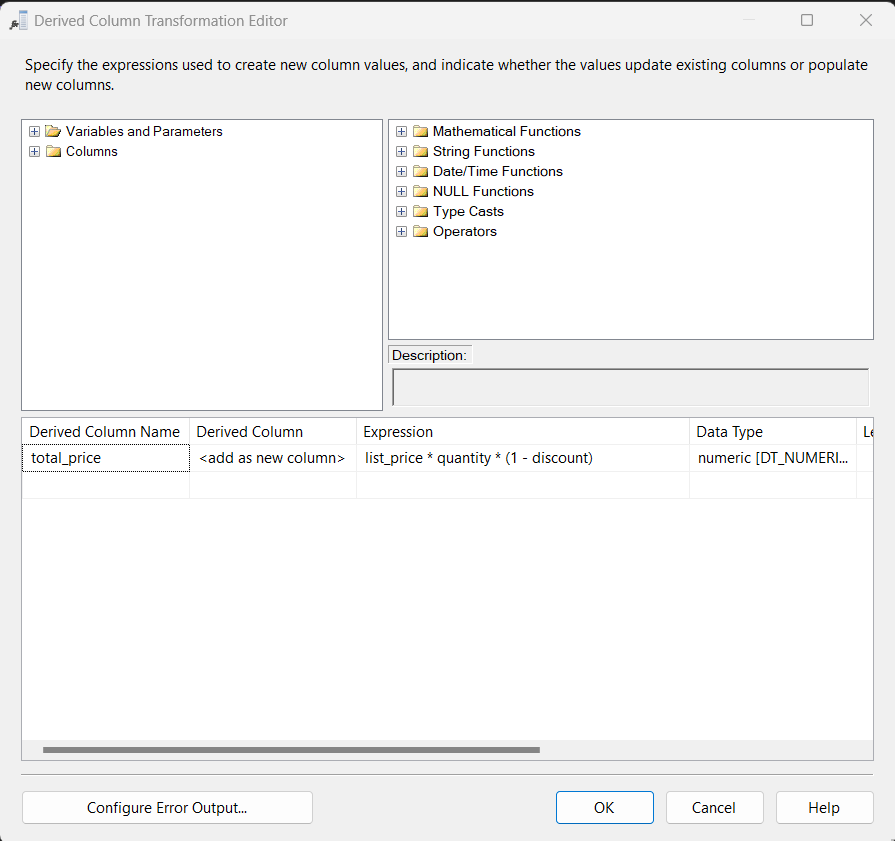
Tạo một sequence container để chứa các data flow và sửa tên thành “Load data From stage to Fact”

### 3.6.1. Bảng Sale Fact (sale\_fact)

Lấy dữ liệu từ các bảng order\_stg, order-item, customer\_stg, Join các bảng với nhanh bằng Merge Join.

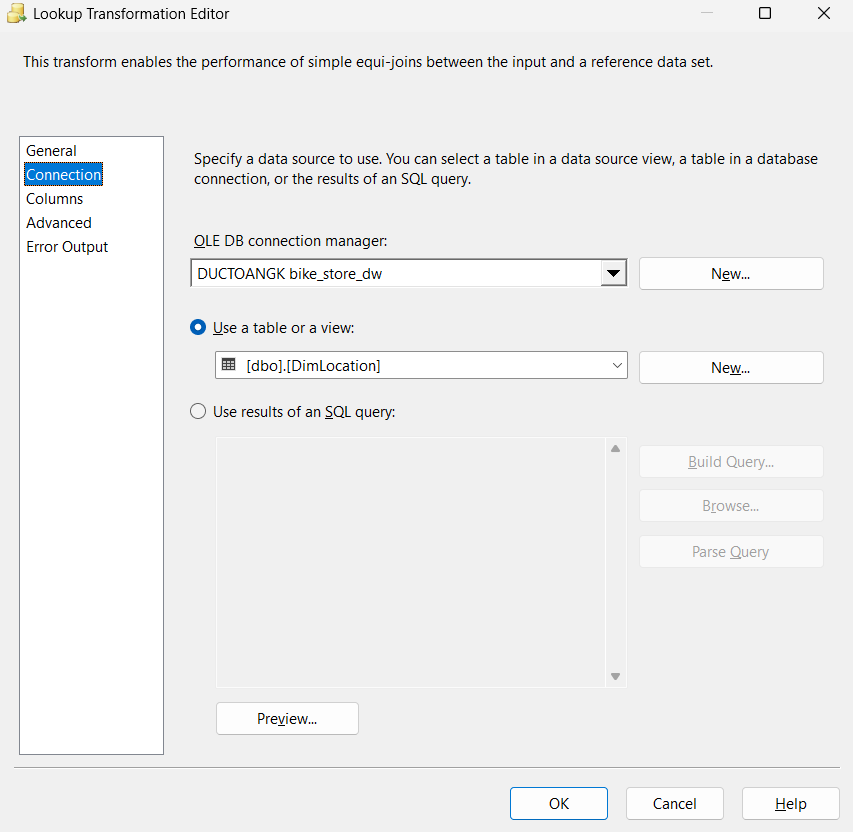


Derived column được dùng để thêm cột total\_price cho mỗi chi tiết đơn hàng.

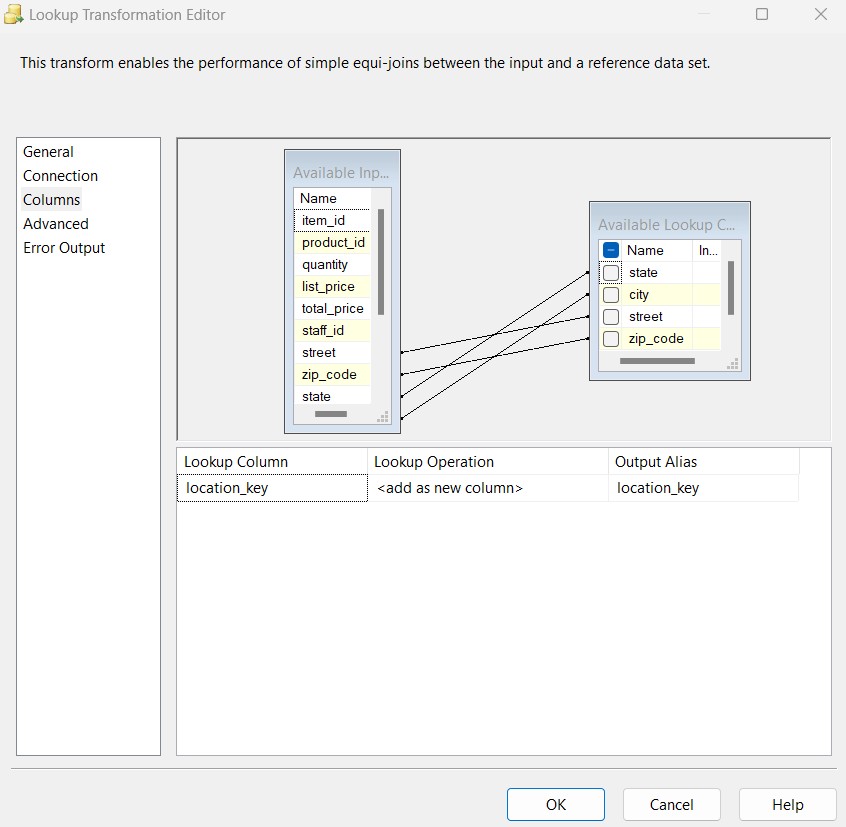


Sử dụng Lookup để ánh xạ các key từ các bảng Dim vào bảng fact

Thiết lập connect tới bảng dim trong datawarehouse.

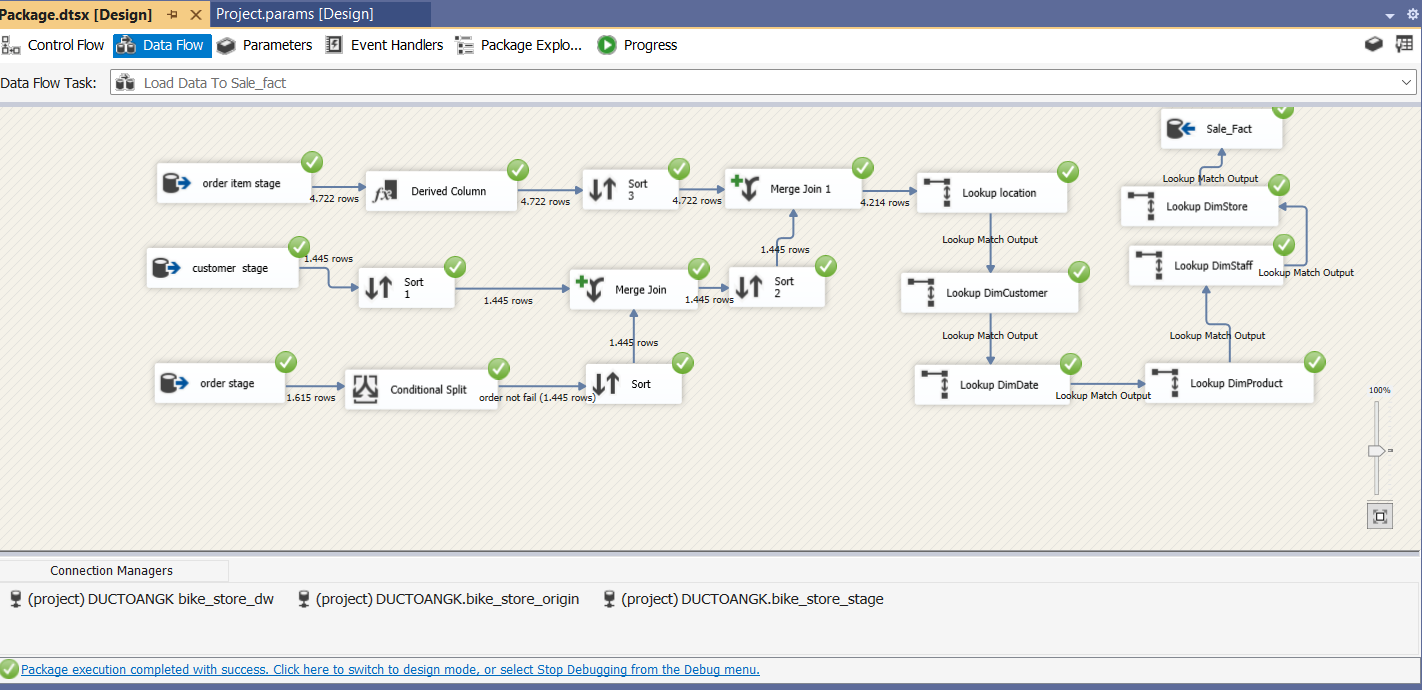


Ánh xạ các cột vào bảng Dim để lấy ra key.



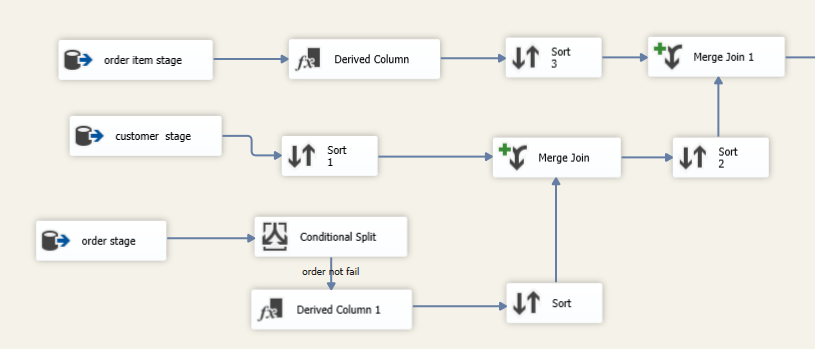
Làm tường tự cho các bảng DimCustomer, DimDate, DimProduct, DimStaff, DimStore.

Thực thi:

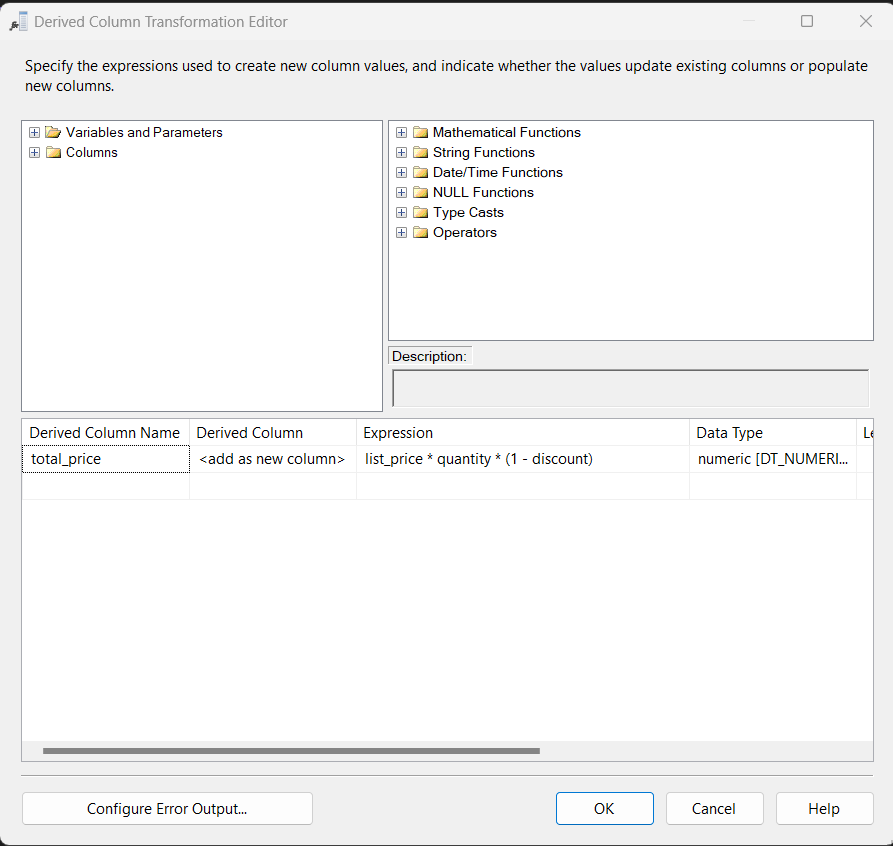


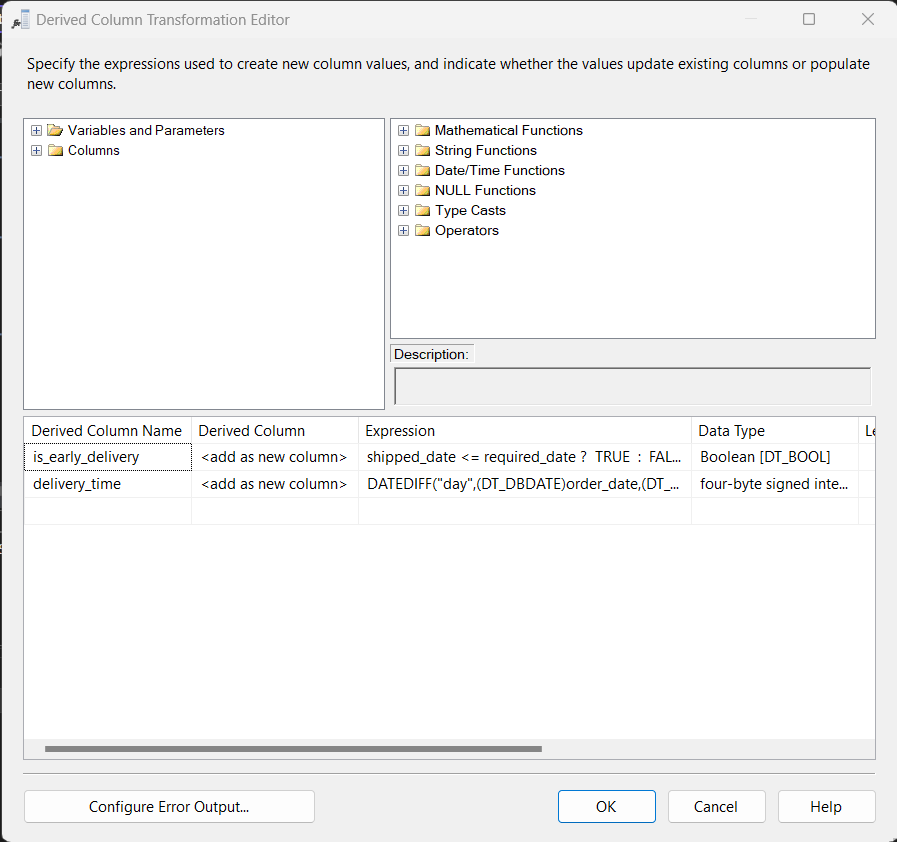
### 3.6.2. Bảng Order Fact (order\_fact)

Lấy dữ liệu từ các bảng order\_stg, order-item, customer\_stg, Join các bảng với nhanh bằng Merge Join.

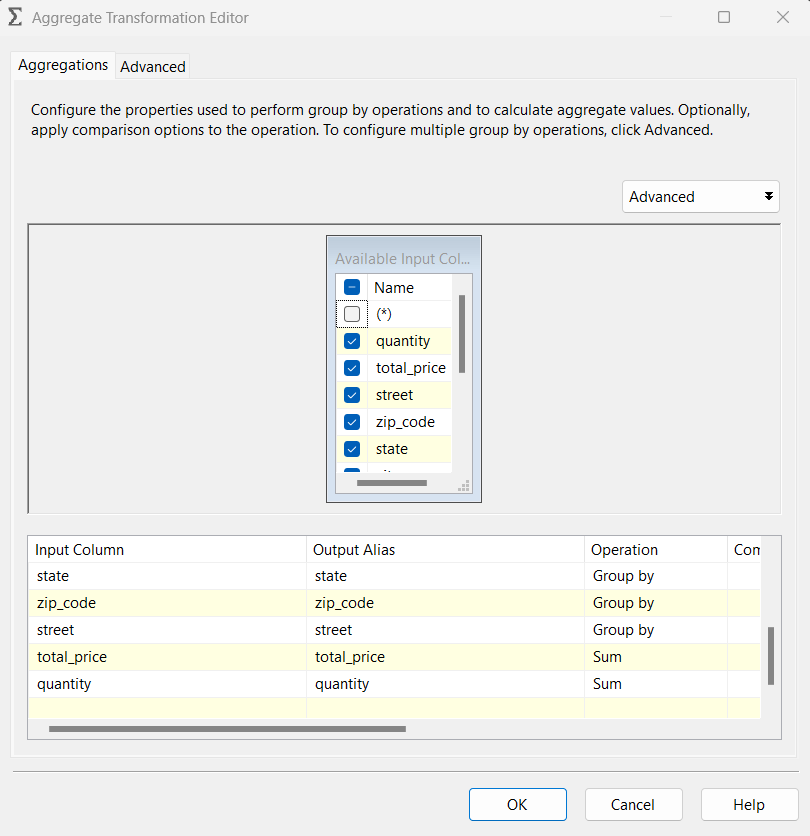


Sử dụng Derived column để tạo thêm cột total\_price, delivery\_time, is\_early\_delivery.

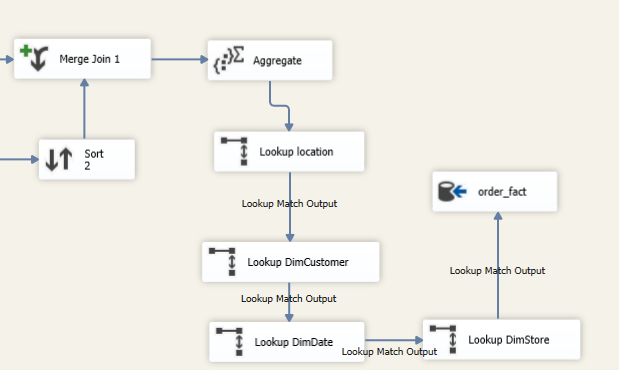




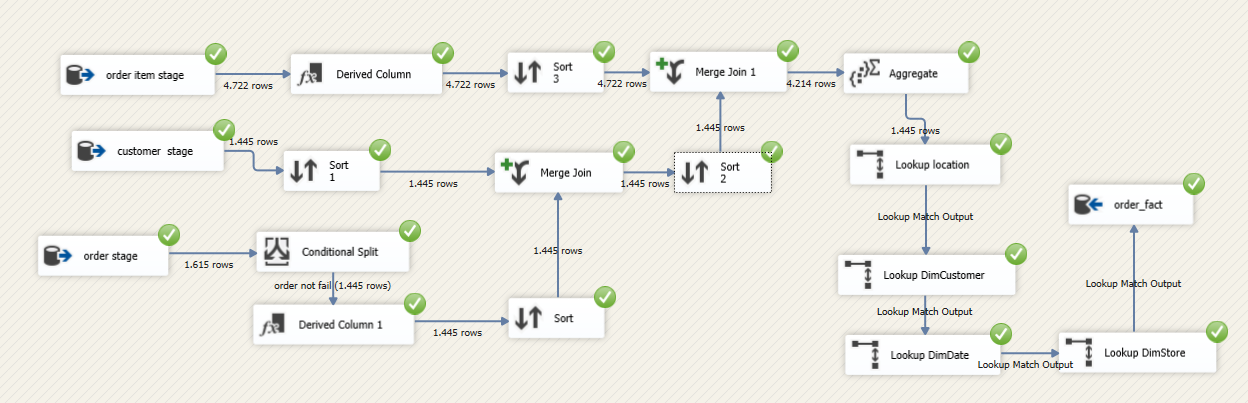
Dùng aggregate để nhóm các chi tiết đơn có cùng mã đơn đồng thời tính tổng giá trị đơn và tổng số lượng hàng.



Sử dụng hàm Lookup tương tự phần sale\_fact để lấy các key và đưa dữ liệu bảng ra bảng order fact.



Thực thi:

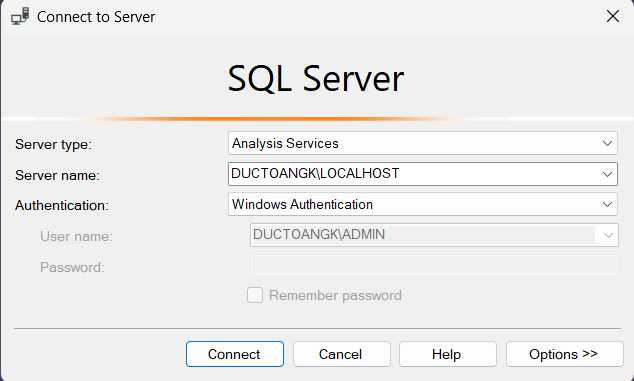


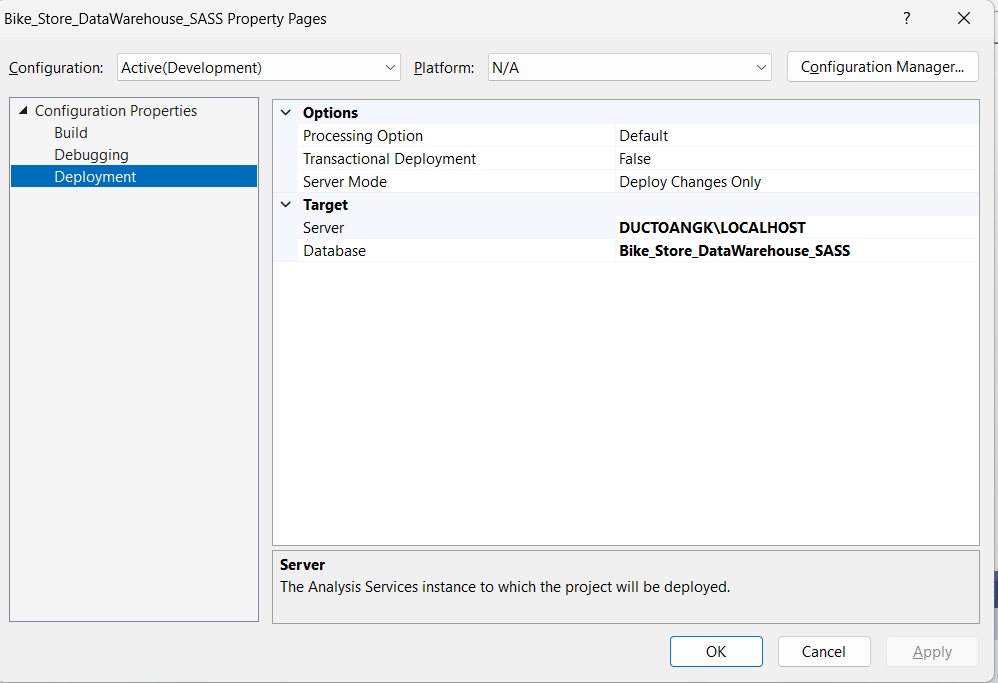
# CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS & POWER BI)

## 

## 4.1. Quá trình xây dựng mô hình

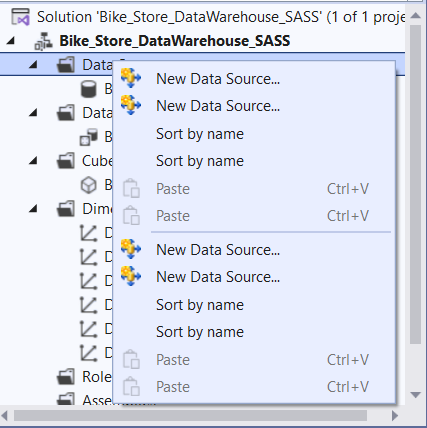
Thiết lập lại Server cho hệ thống: Nhấp chuột phải vào tên Solution → Deployment → Target → Đổi tên Server đến Analysis Services.

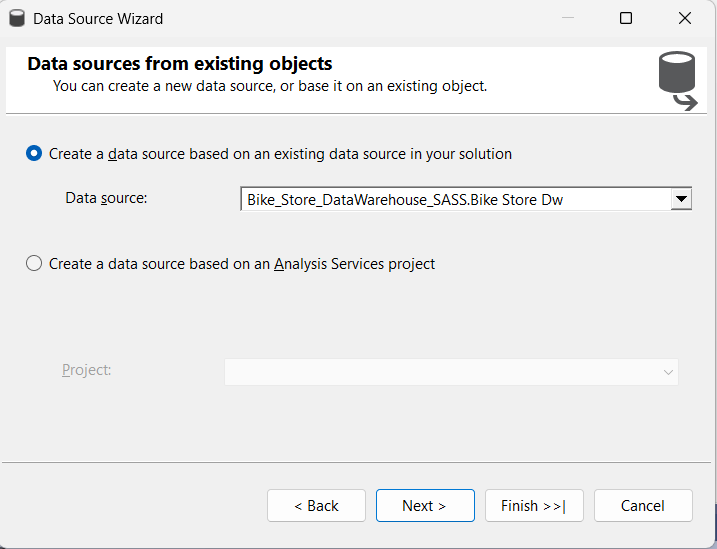


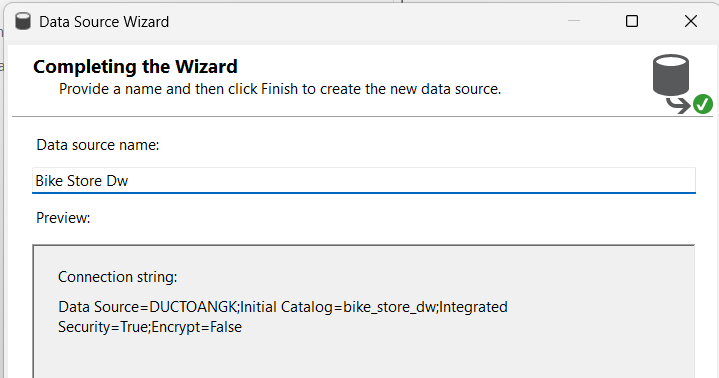


### 4.1.1. Tạo Data Source

* Nhấp chuột phải vào Data Sources → New Data Resource.. → New (Data Connection) → Điền Server name, database name → Ok
* Chọn data connection vừa tạo → Next → Use this service account → Đặt tên → Finish

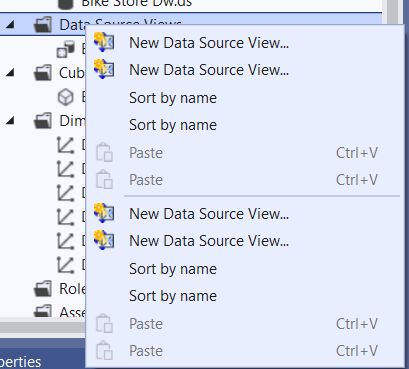




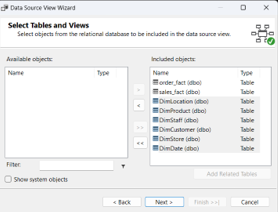


### 4.1.2. Tạo Data Source View

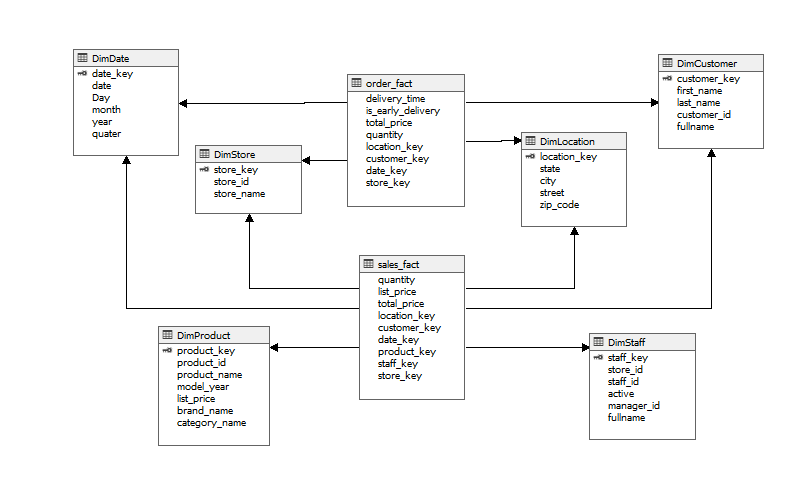
* Nhấp chuột phải vào Data Sources View→ New Data Resource View.. → Chọn data sources vừa tạo ở trên → Next.



* Đẩy order\_fact và sale\_fact sang bảng included objects → Add related tables → Next → Finish

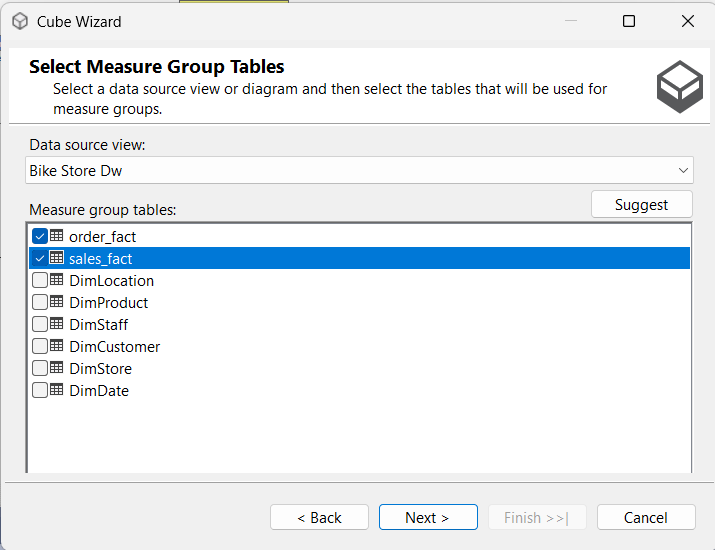


Ấn đúp vào data sources view vừa tạo để xem diagram

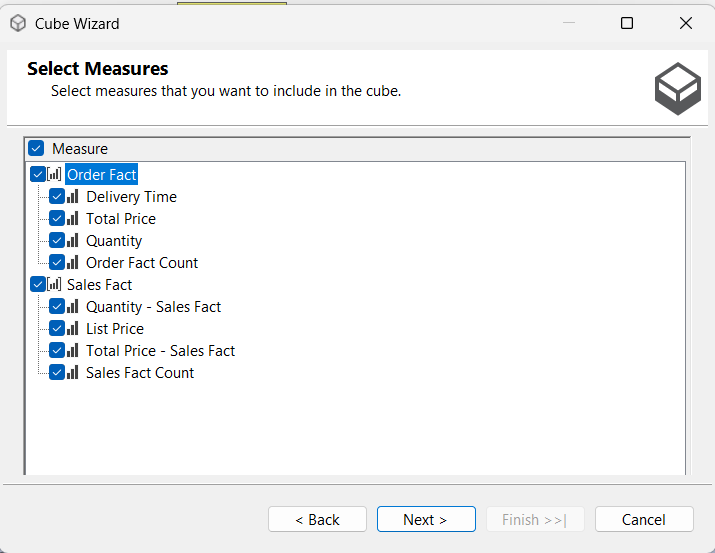


## 4.2. Quá trình xây dựng khối Cube

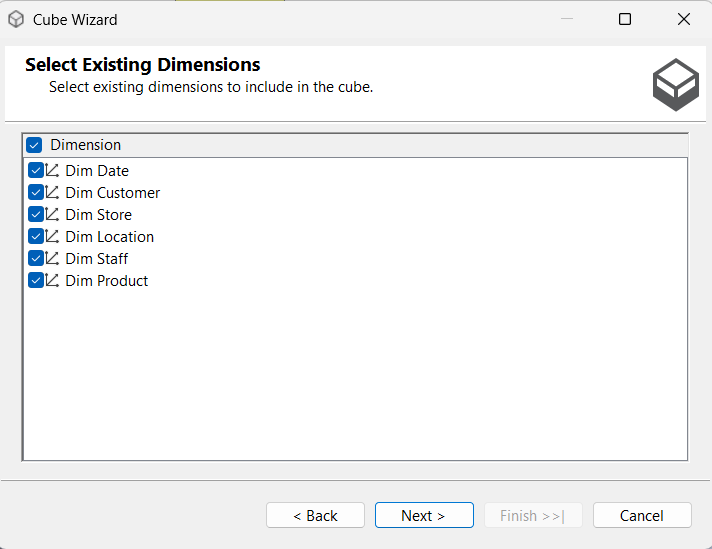
* Nhấp chuột phải vào **Cubes** ở khung Solution Explorer → New Cube → Use existing tables → Next → Chọn **order\_fact** và **sales\_fact** → Next



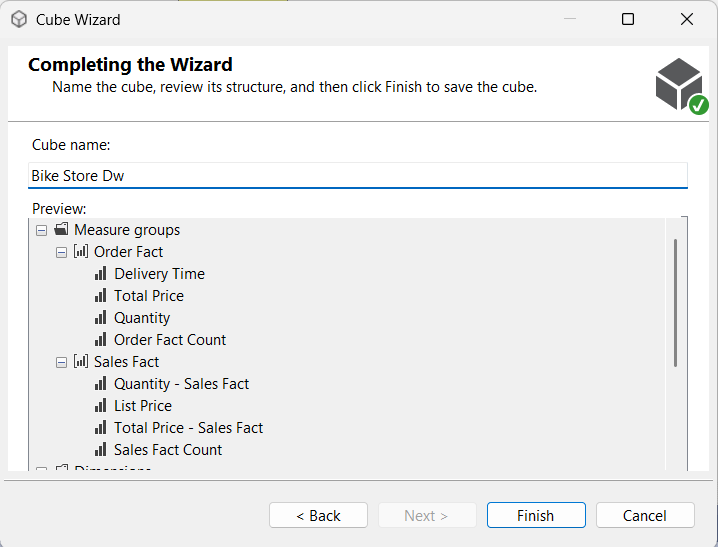
* Chọn những Measure cần thiết.



* Khởi tạo các bảng Dimension

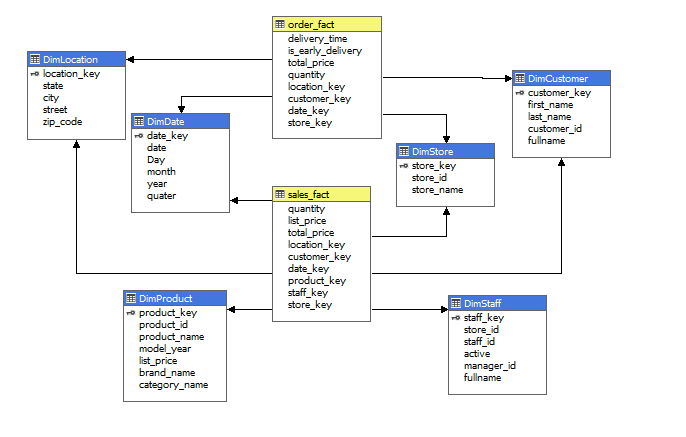


Đặt tên cho Cube là Bike Store Dw.



Ấn Finish để hoàn tất việc khởi tạo.

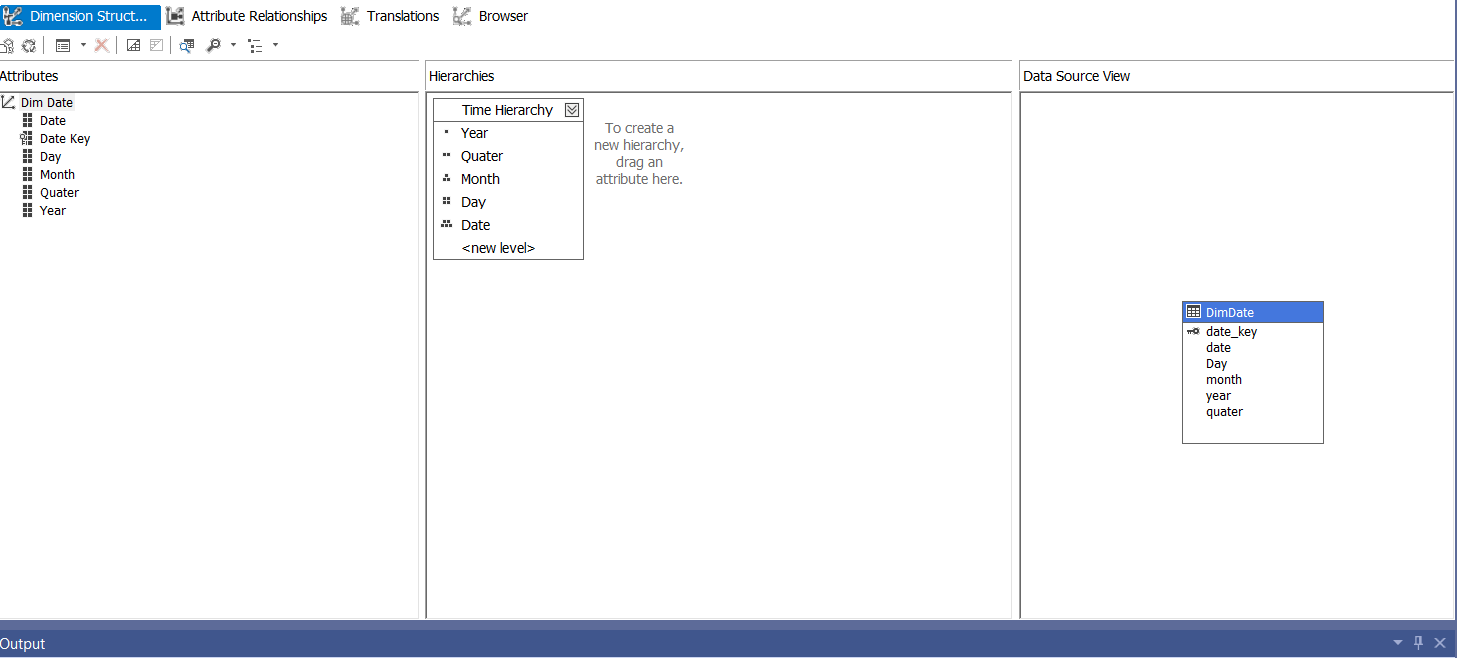
* Các bảng Fact, Dimension và quan hệ giữa chúng sẽ được thể hiện trong Diagram



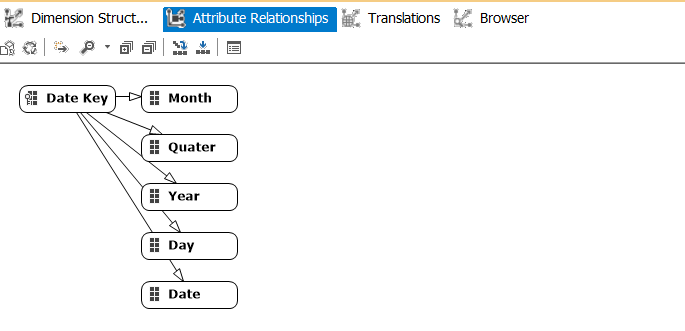
## 4.3. Cấu hình Hierarchy

#### 4.3.1. Tạo Hierarchy cho Dim Date

* Kéo thả các thuộc tính Date, Day, Month, Quater, Year từ bảng DimDate trong panel Data Source View sang panel Attributes.
* Sau đó lần lượt kéo thả các thuộc tính từ pannel Attribute này sang panel Hierachy. Chú ý thứ tự Year→Quater→Month→Day→Date.
* Nhấn chuột phải cào bảng Hierachy → Rename →Đặt lại tên bảng là Time Hierachy

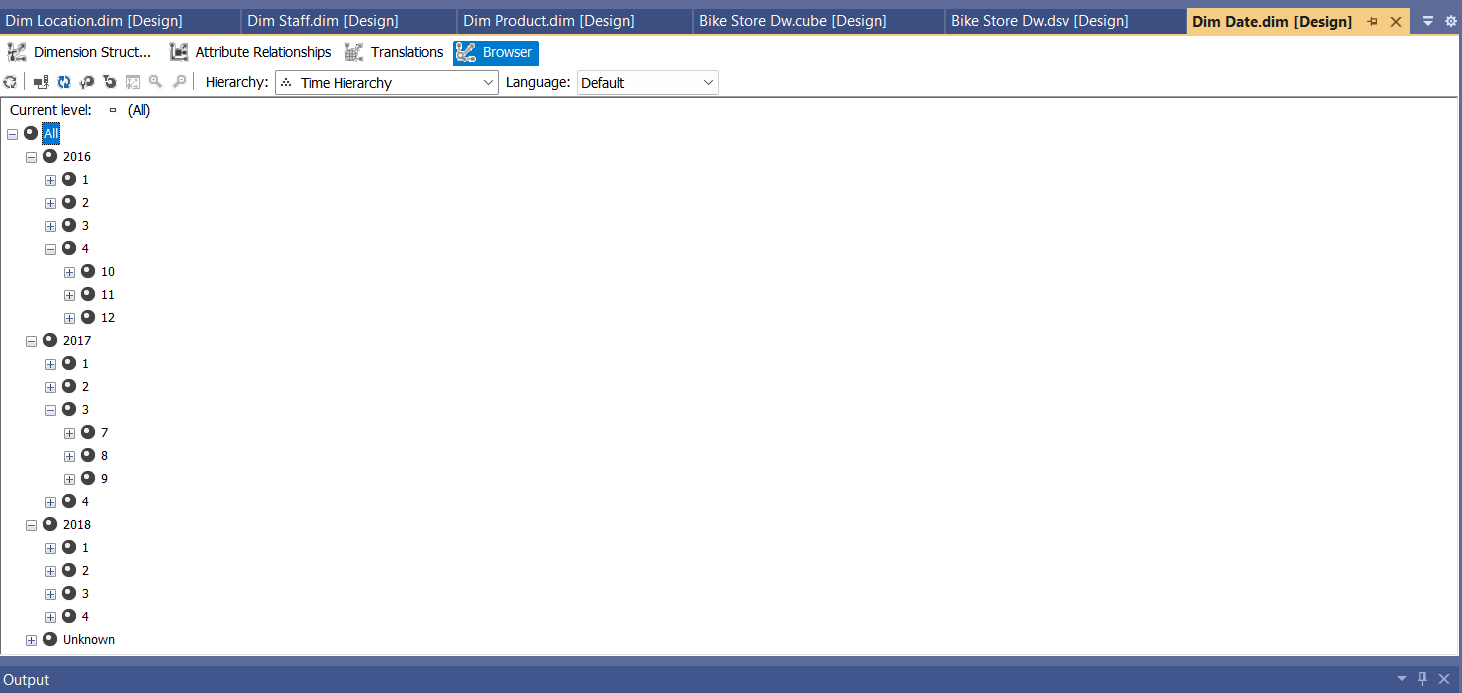


* Tạo các Attribute Relationship



Nhấn chuột phải vào DimDate.dim → Process (để load dữ liệu cho bảng Dim Date)

Sau đó, chuyển đến tab Browser, sẽ xem được` thời gian được phân cấp thep năm(2016, 2017, 2018) → quý(1, 2, 3, 4) → tháng(1, 2, …12) → ngày(1, 2,...31) →date

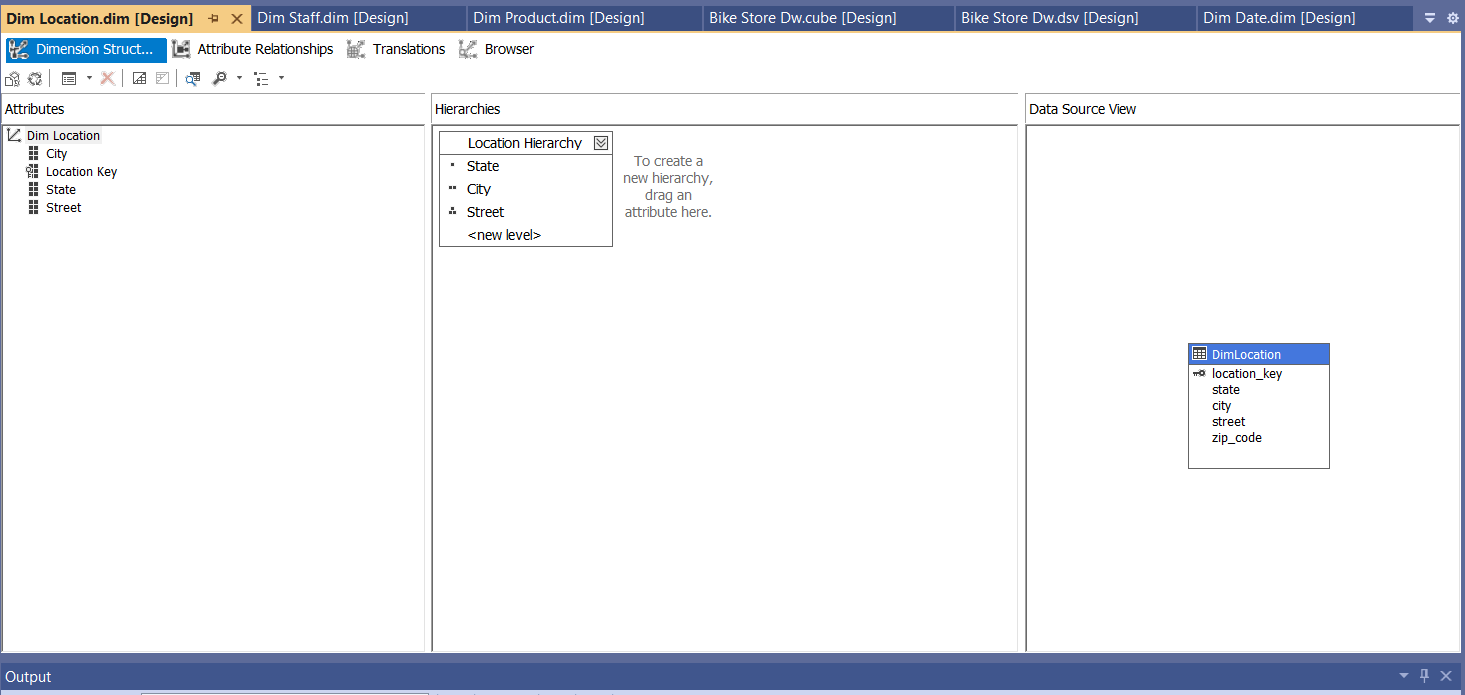


#### 4.3.2. Tạo Hierarchy cho Dim Location

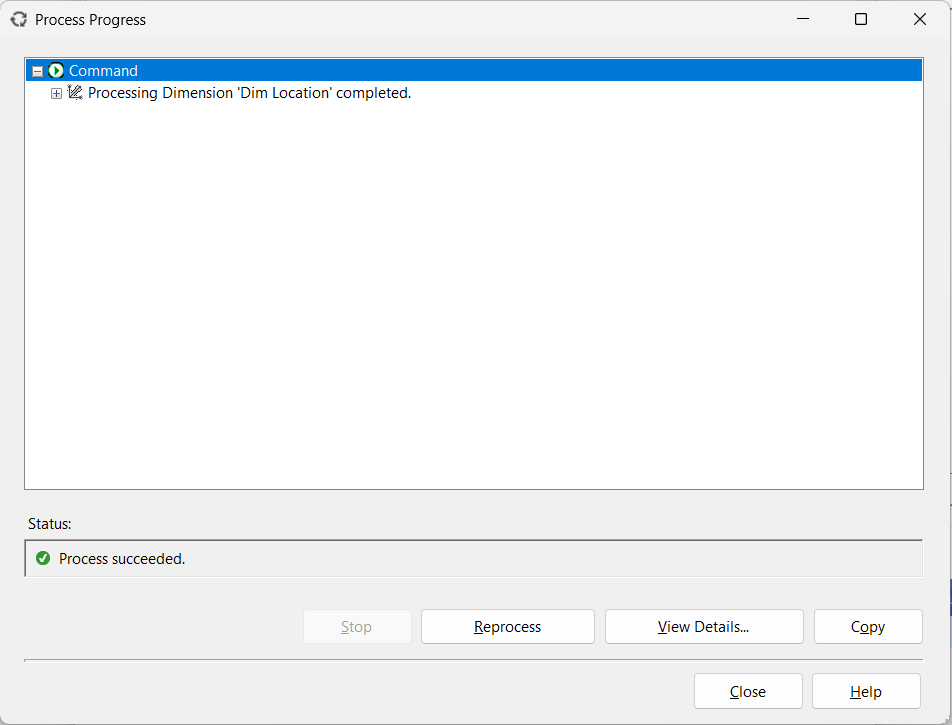
* Tương tự như cách tạo Hierachy cho Dim Date, ta tạo hierachy cho Dim Location

Thứ tự State→City→Street.

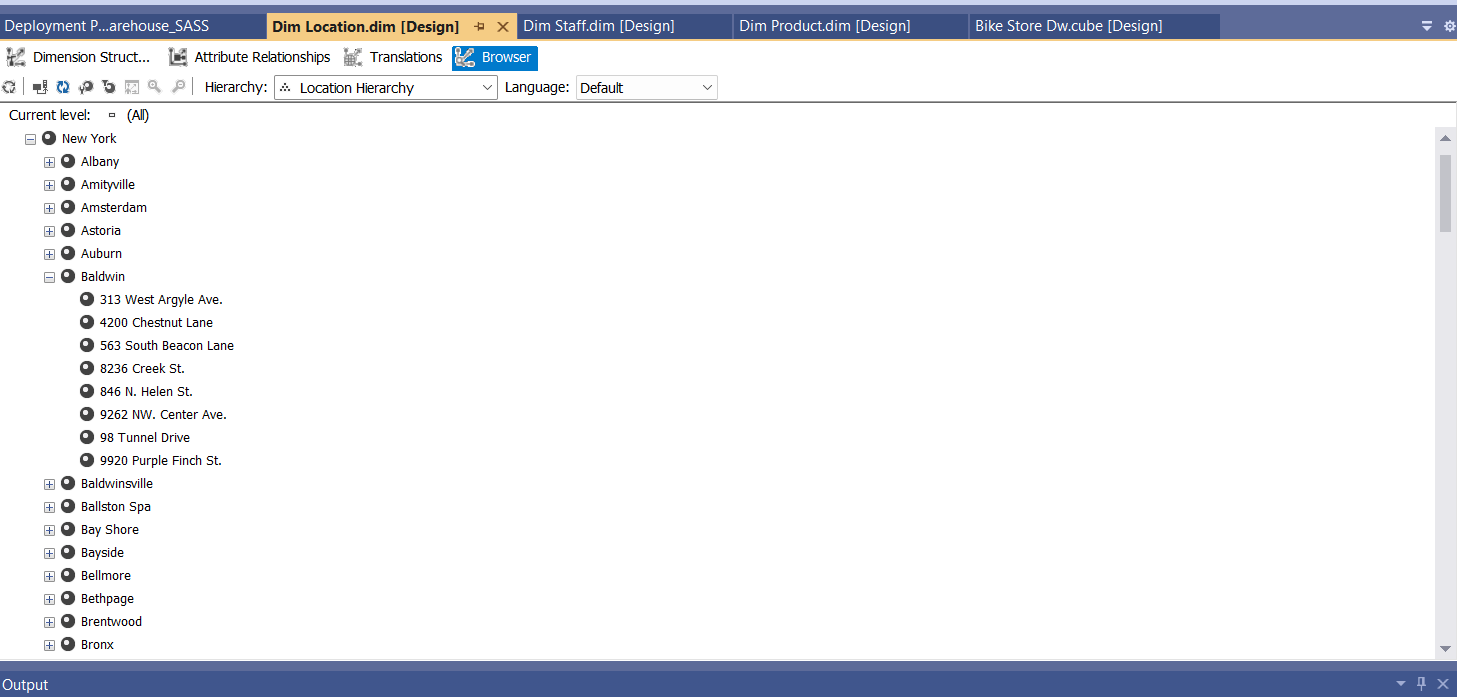
Đặt lại tên bảng là Location Hierachy



* Nhấn chuột phải vào Dim Location.dim → Process (để load dữ liệu cho bảng Dim Location)



* Chuyển đến tab Browers. Vị trí được sắp xếp theo Bang→ Thành phố → Đường.`



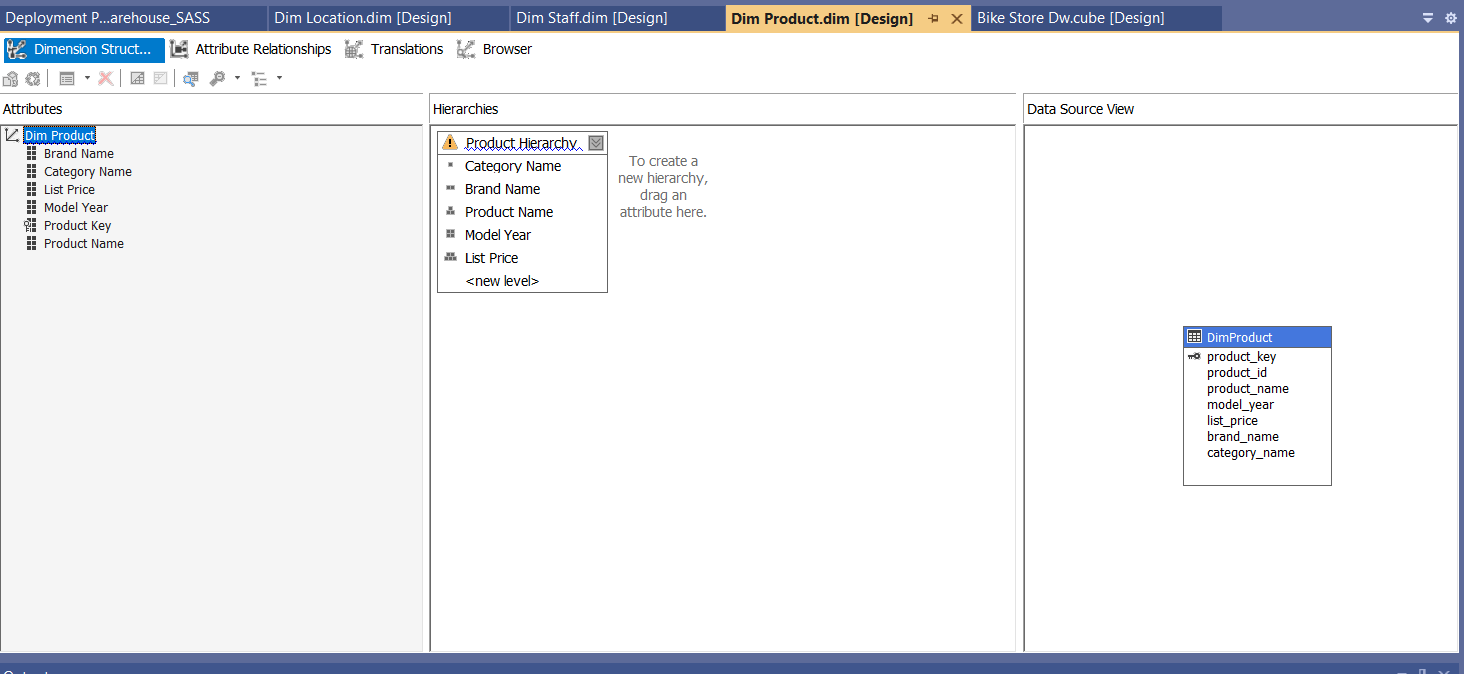
#### 

#### 4.3.4. Tạo Hierarchy cho Dim Product

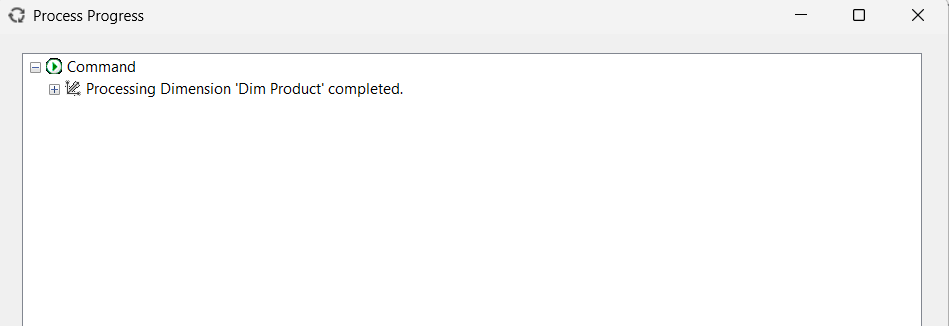
* Tương tự, tạo hierachy cho Dim Product

Thứ tự Category Name→Brand Name→Product Name → Model year →List price.

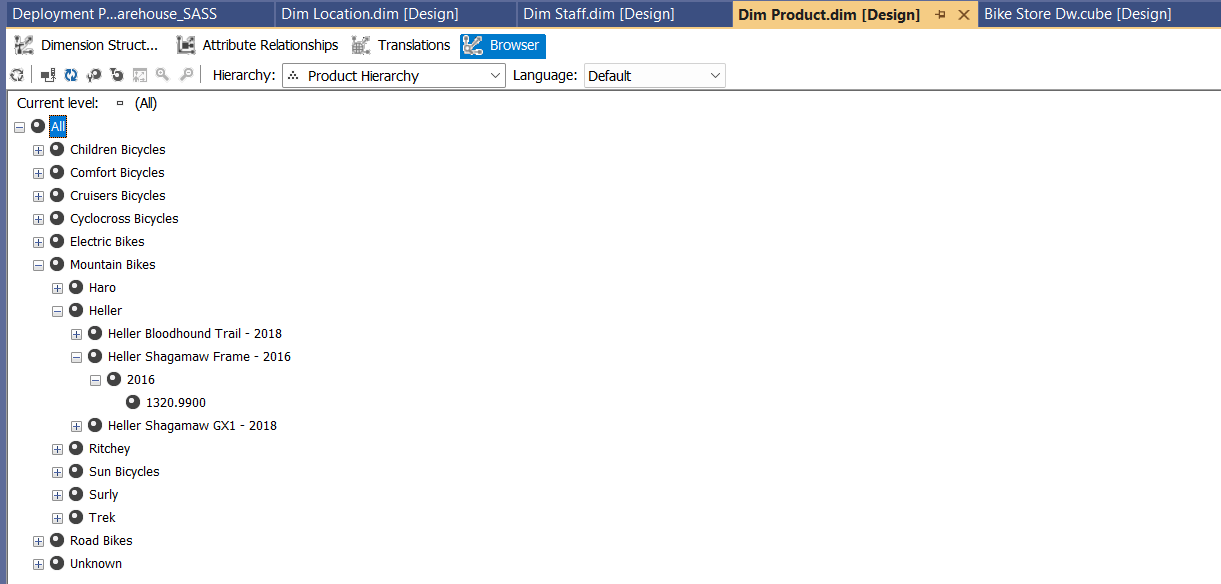
Đặt lại tên bảng là Product Hierachy



* Nhấn chuột phải vào Dim Product.dim → Process (để load dữ liệu cho bảng Dim Product)



* Chuyển qua tab Browser sẽ thấy các sản phẩm được phân loại 8 loại xe đạp chính → Tên các nhãn hàng cho mỗi loại → Tên xe → Năm sản xuất → Giá.`

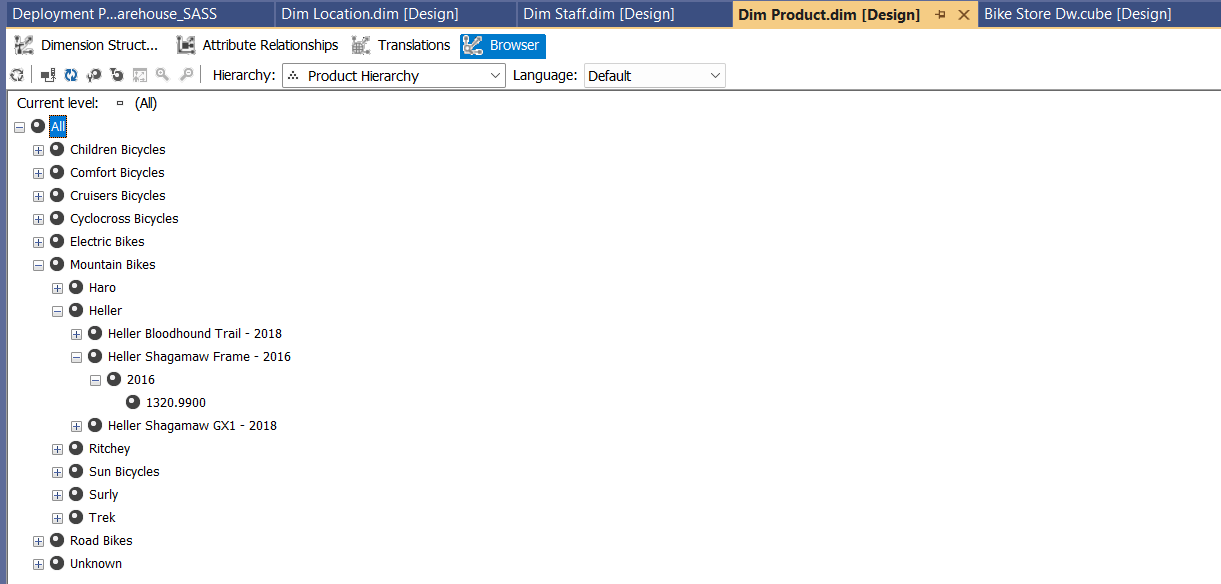


#### 4.3.5. Tạo Hierarchy cho Dim Staff

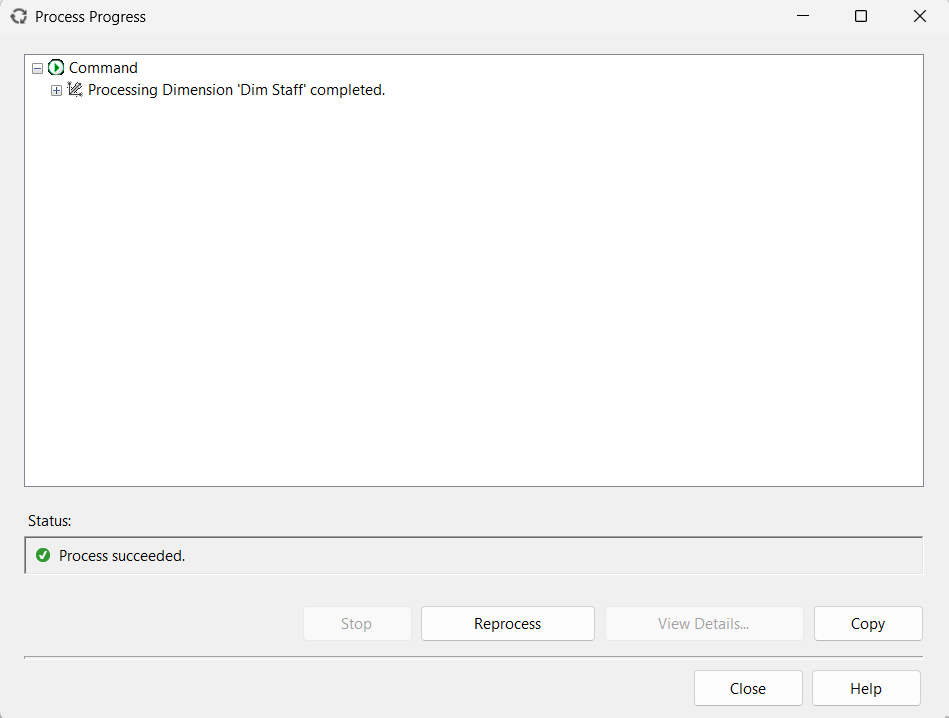
* Tương tự, ta tạo hierarchy cho Dim Staff

Thứ tự StoreId→Manager Id→Fullname→Active.

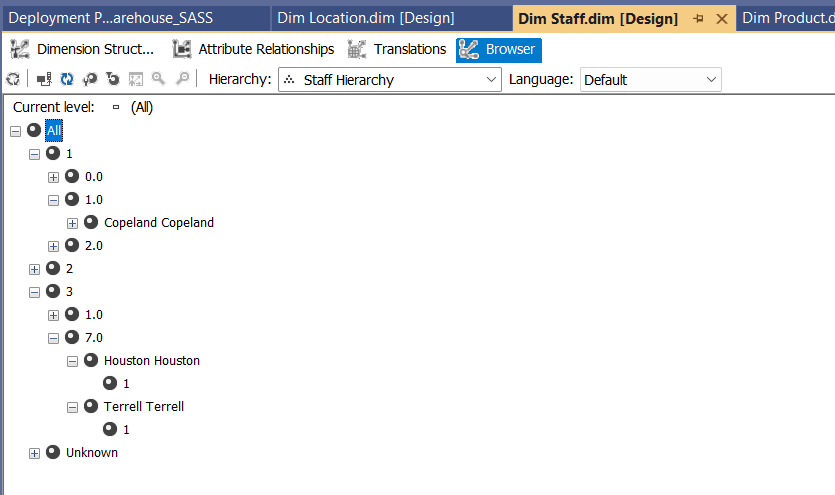
Đặt lại tên bảng là Staff Hierarchy



* Nhấn chuột phải vào Dim Staff.dim → Process (để load dữ liệu cho bảng Dim Staff)



* Chuyển đến tab Browser để xem danh sách nhân viên theo cửa hàng → Người quản lý → tên nhân v`iên → Trạng thái hoạt động.



## 4.4. Thực hiện phân tích dữ liệu

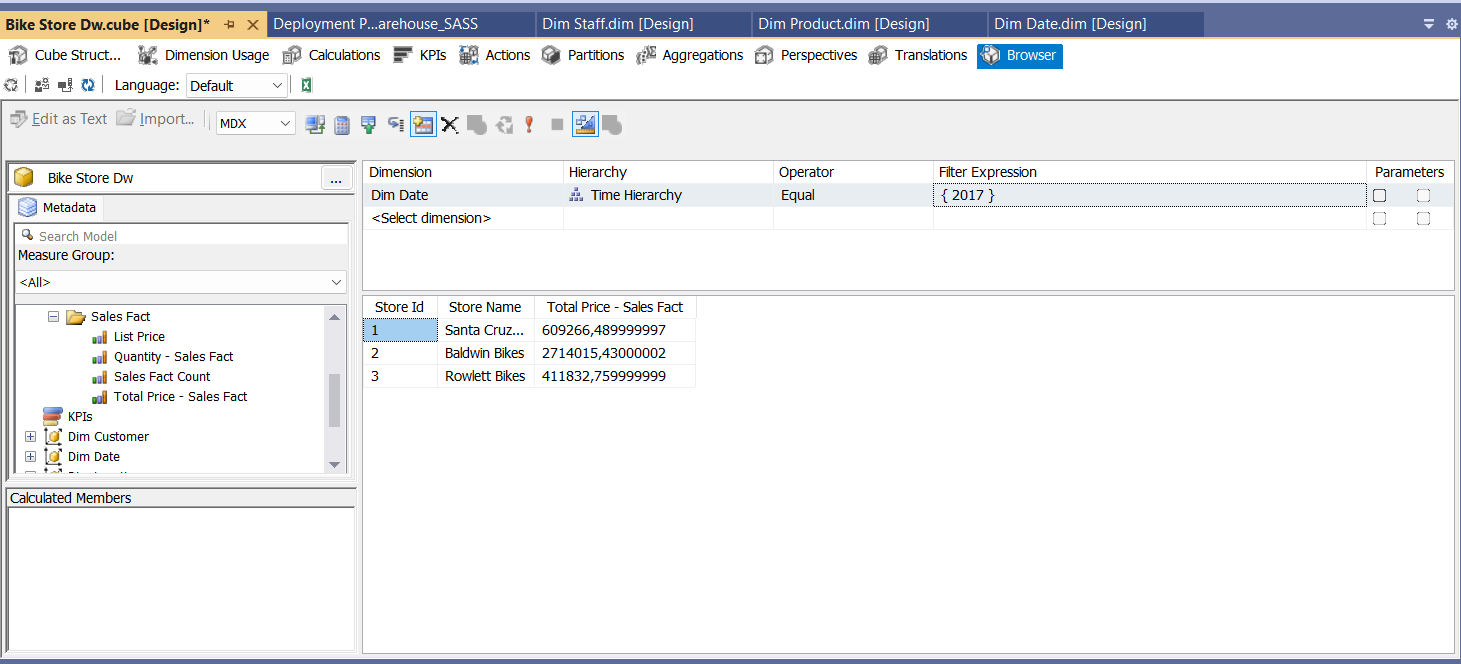
### 4.4.1. Báo cáo doanh số (Sales reporting)

Trả lời cho các câu hỏi:

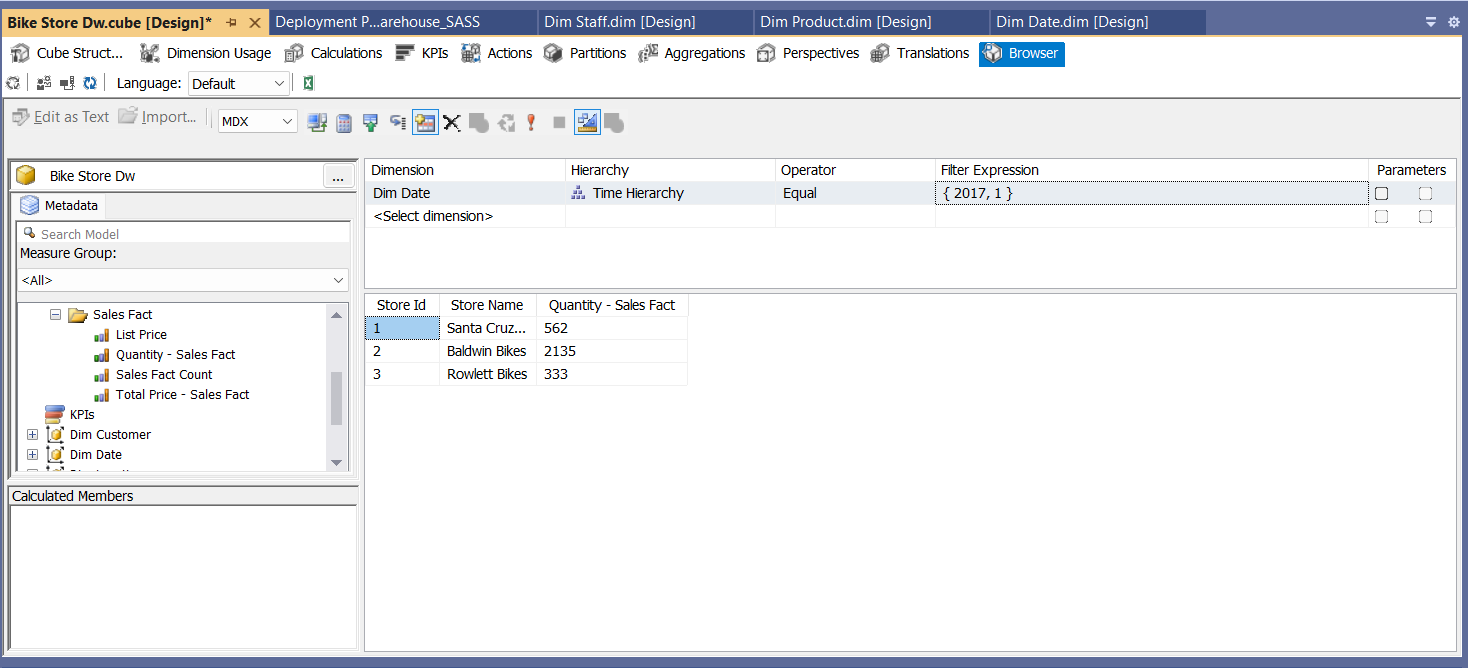
* Tổng doanh thu của chuỗi cửa hàng là bao nhiêu? Doanh thu của cửa hàng ở từng khu vực?
* Tổng số lượng xe đạp bán ra là bao nhiêu? Số lượng bán ra của mỗi khu vực?
* Tổng số tiền của từng khu vực theo mỗi quý?
* Top những sản phẩm có giá trị bán ra cao nhất?
* Top những nhân viên xuất sắc nhất? (tính theo tổng giá trị sản phẩm bán ra)

**4.4.1.1. Sử dụng công cụ SSAS**

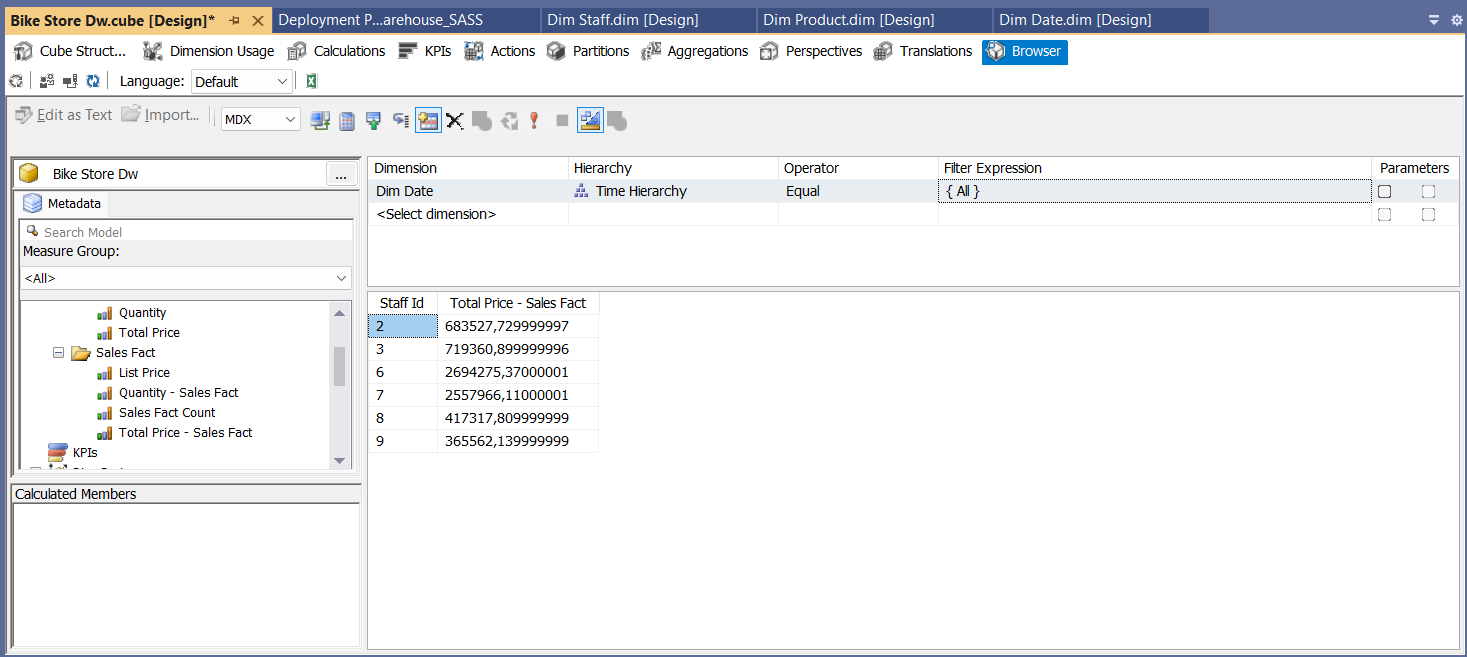
Doanh thu của 3 cửa hàng trong năm 2017.

****

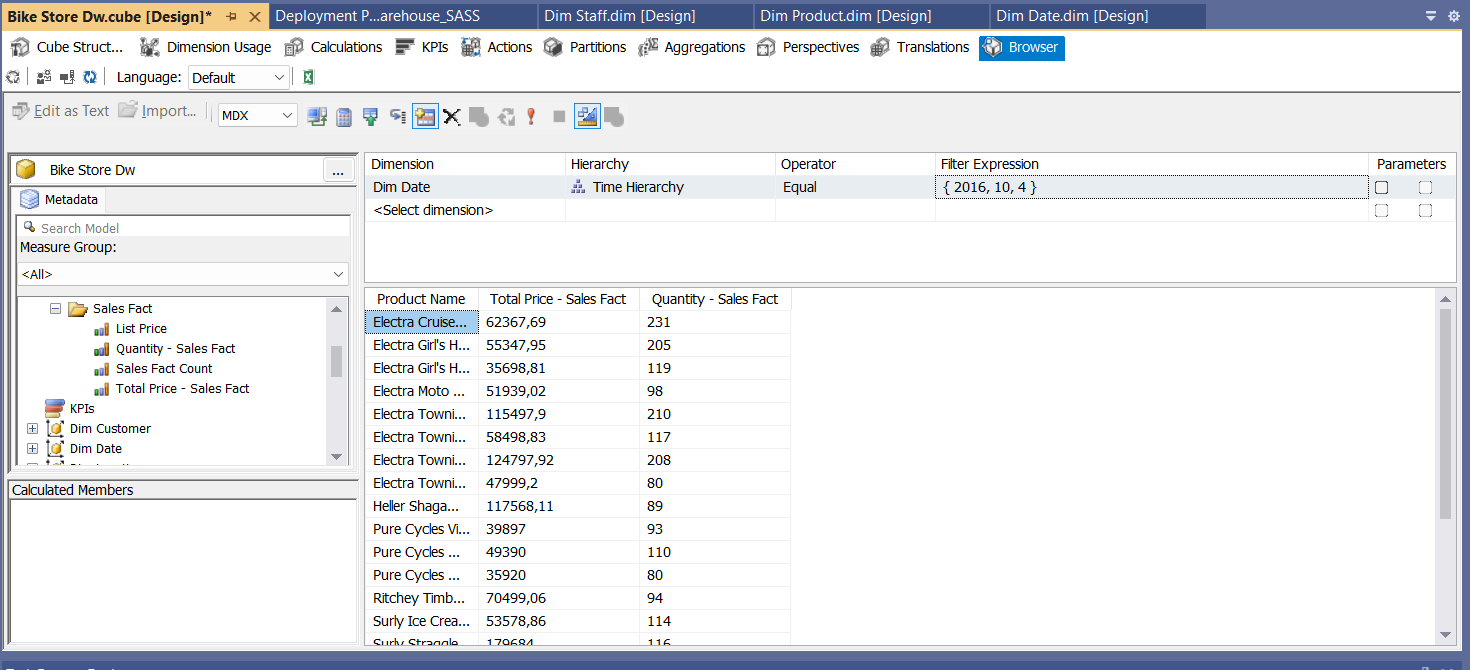
Tổng số lượng xe đạp bán ra trong quý 1 của năm 2017 ở 3 cửa hàng



Hiển thị doanh số của 6 nhân viên bán hàng từ trước đến nay.

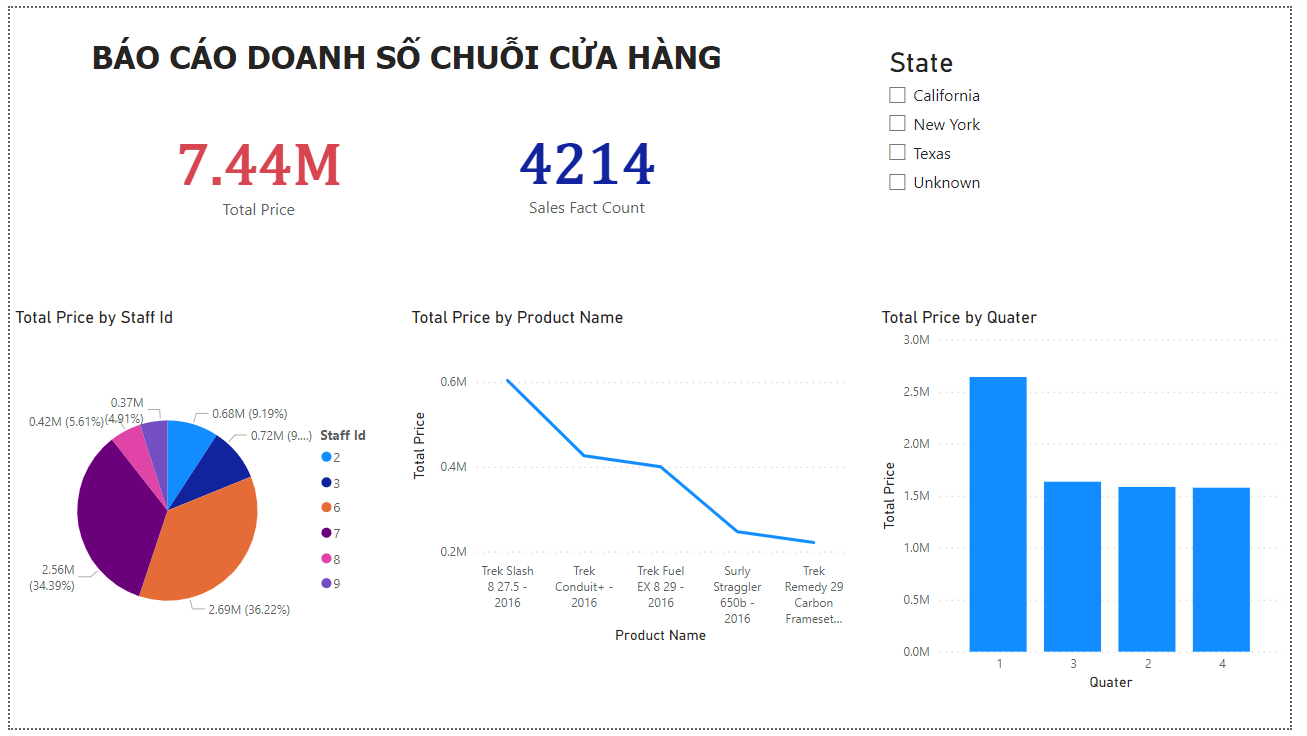


Tổng số lượng sản phẩm bán ra, tổng doanh số trên từng sản phẩm trong tháng 10 năm 2016

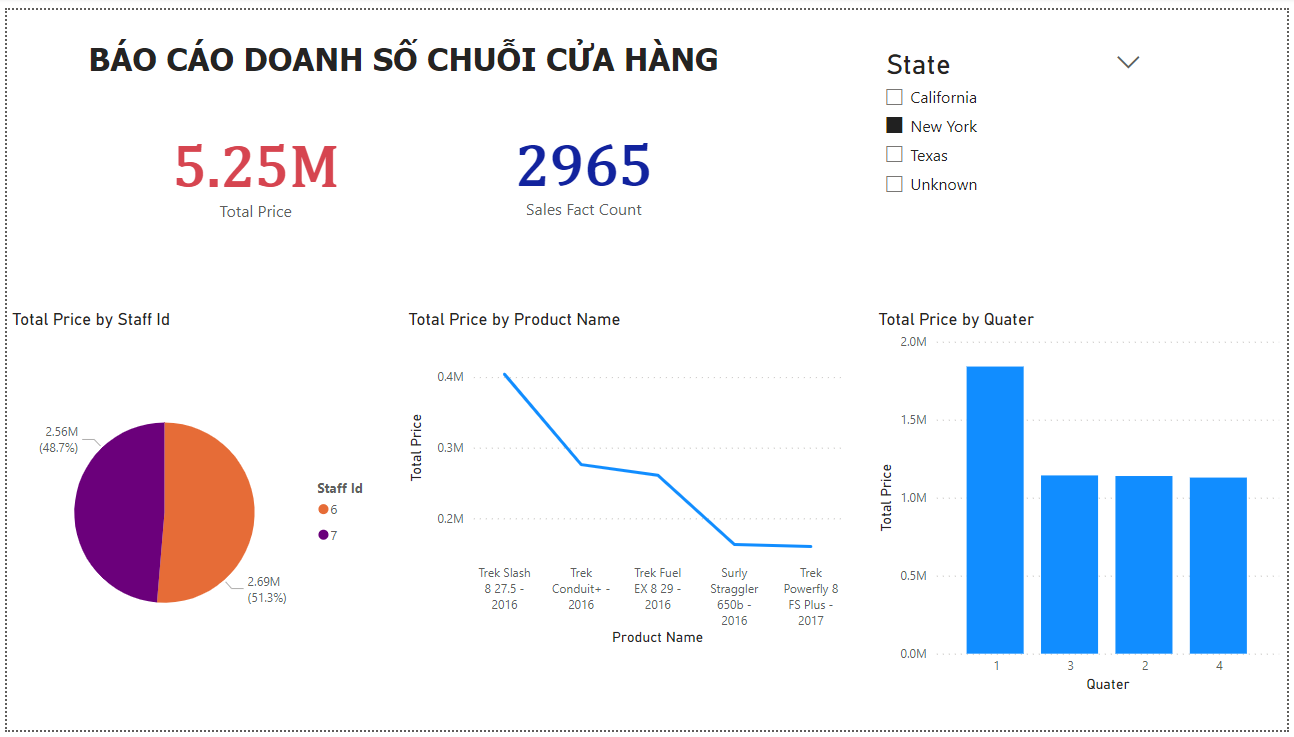


**4.4.1.2. Sử dụng Power BI**

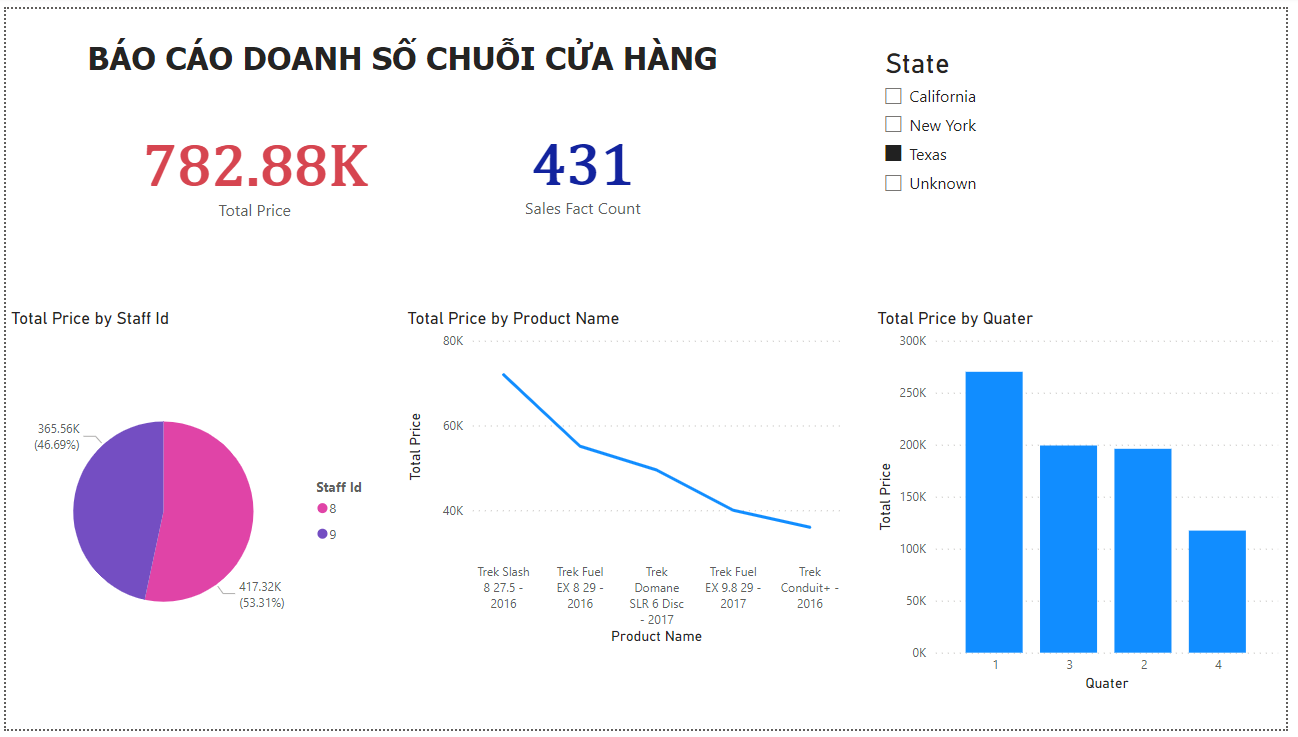
Nhìn vào biểu đồ, ta thấy được tổng doanh thu của chuỗi cửa hàng là 7.44M với số lượng đơn hàng đã bán ra là 4214 đơn hàng.



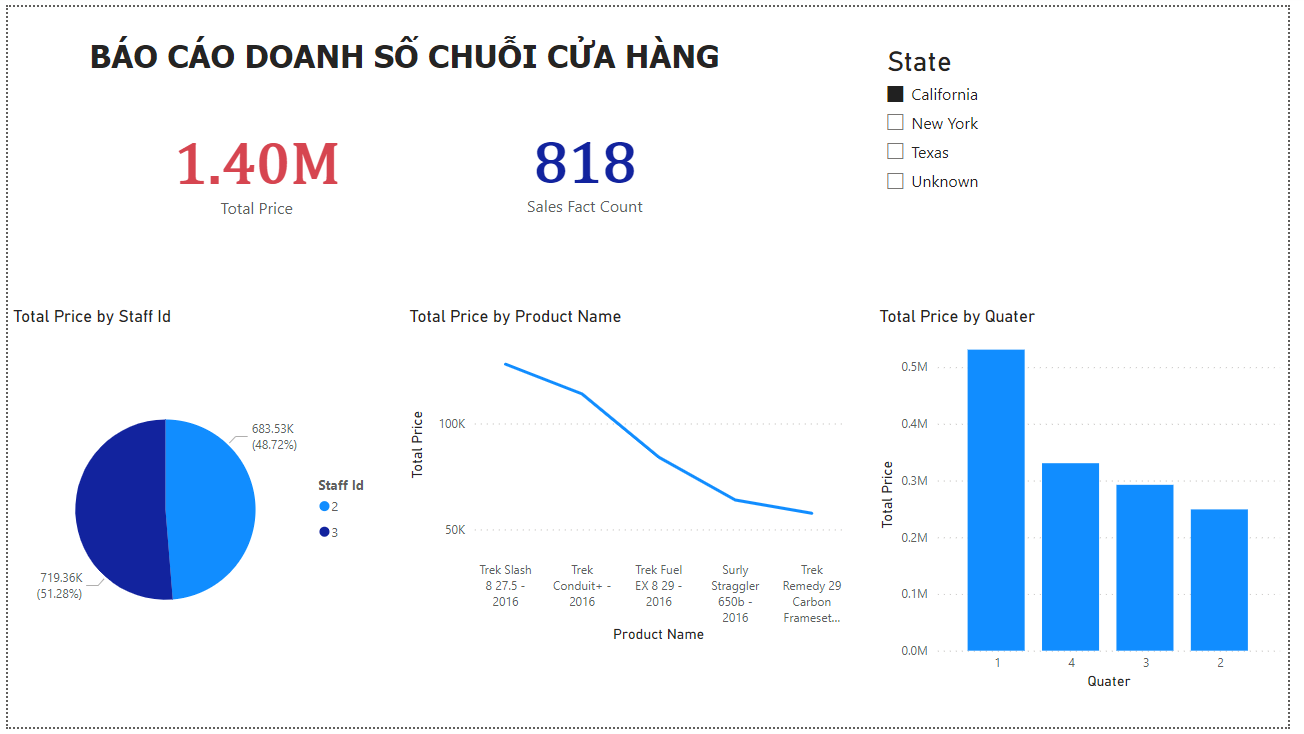
Lọc theo từng khu vực, ta sẽ thấy được mức doanh thu của doanh thu số lượng, số lượng đơn hàng, tổng giá trị các sản phẩm bán ra theo từng quý.



Bên cạnh đó, biểu đồ còn cung cung cấp cho ta biết, top những nhân viên có tổng doanh số cao nhất, đồng thời còn biết được top những sản phẩm mà khách hàng ưa chuộng và mang lại giá trị cao cho cho chuỗi cửa hàng.



Qua đó sẽ biết được nhu cầu của thị trường để chuỗi cửa hàng có thể đáp ứng được thị hiếu của khách hàng. Đồng thời thời có thể đưa những chính sách tuyên dương những nhân viên có thành tích tốt tốt cho cửa hàng.



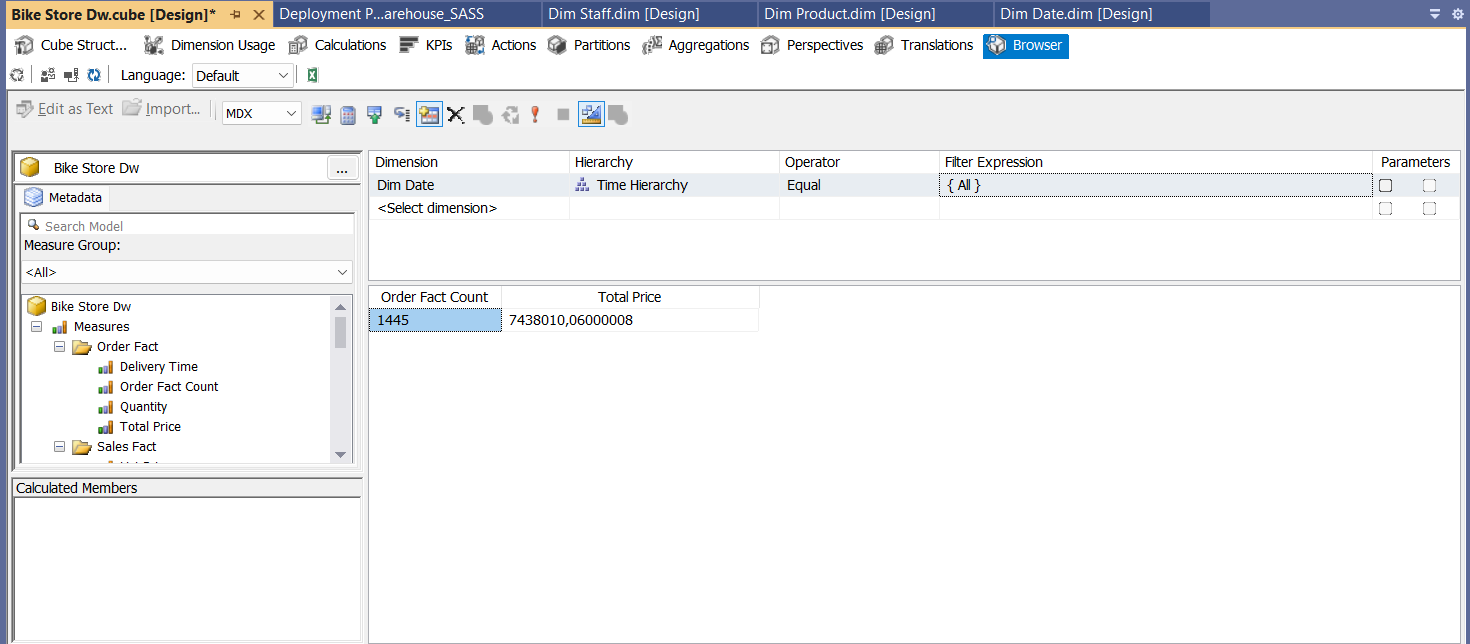
**4.4.2. Quản lý đơn hàng(Order and delivery)**

Trả lời cho các câu hỏi:

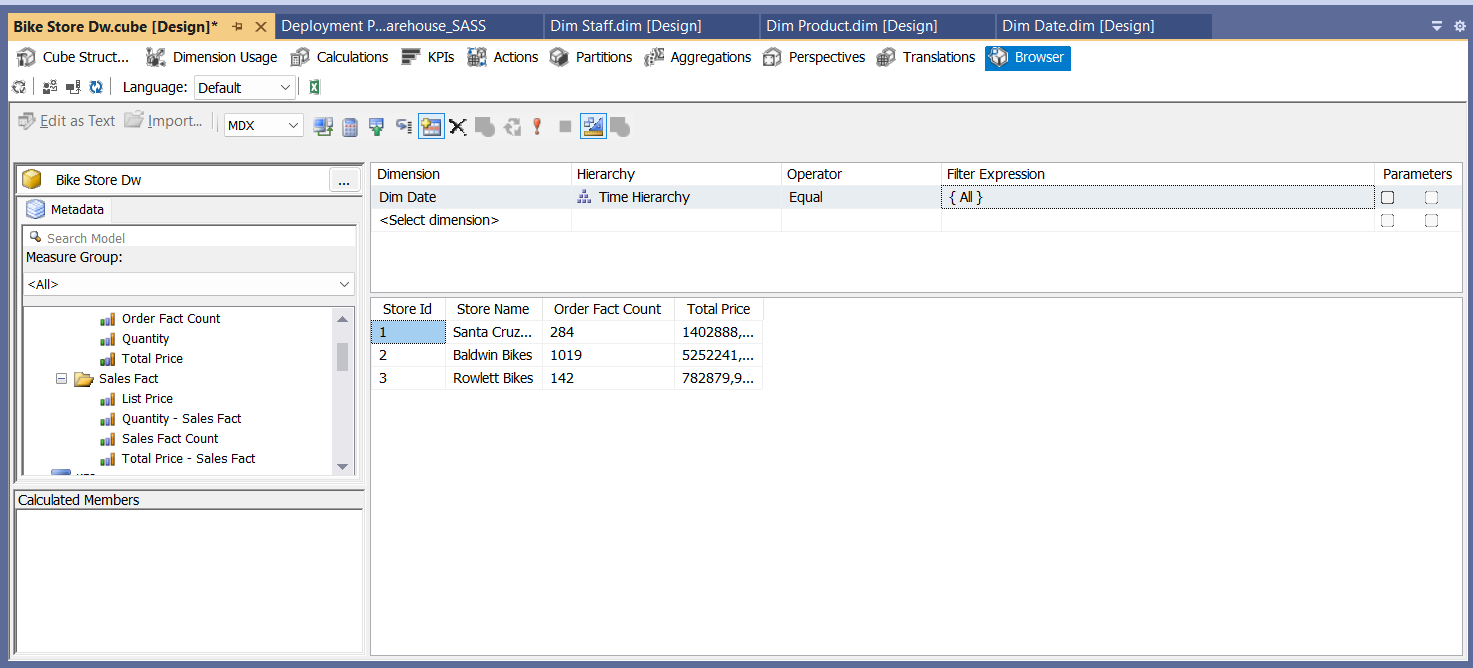
* Tổng số đơn hàng của chuỗi cửa hàng là bao nhiêu? Số lượng đơn hàng của mỗi cửa hàng?
* Tổng thời gian giao hàng của chuỗi chuỗi cửa hàng và của mỗi cửa hàng?
* Số lượng đơn hàng theo từng quý của từng cửa hàng và của mỗi cửa hàng?
* Tổng giá trị sản phẩm bán ra của chuỗi chuỗi cửa hàng và của mỗi mỗi cửa hàng theo từng tháng là bao nhiêu?

**4.4.2.1. Sử dụng SSAS**

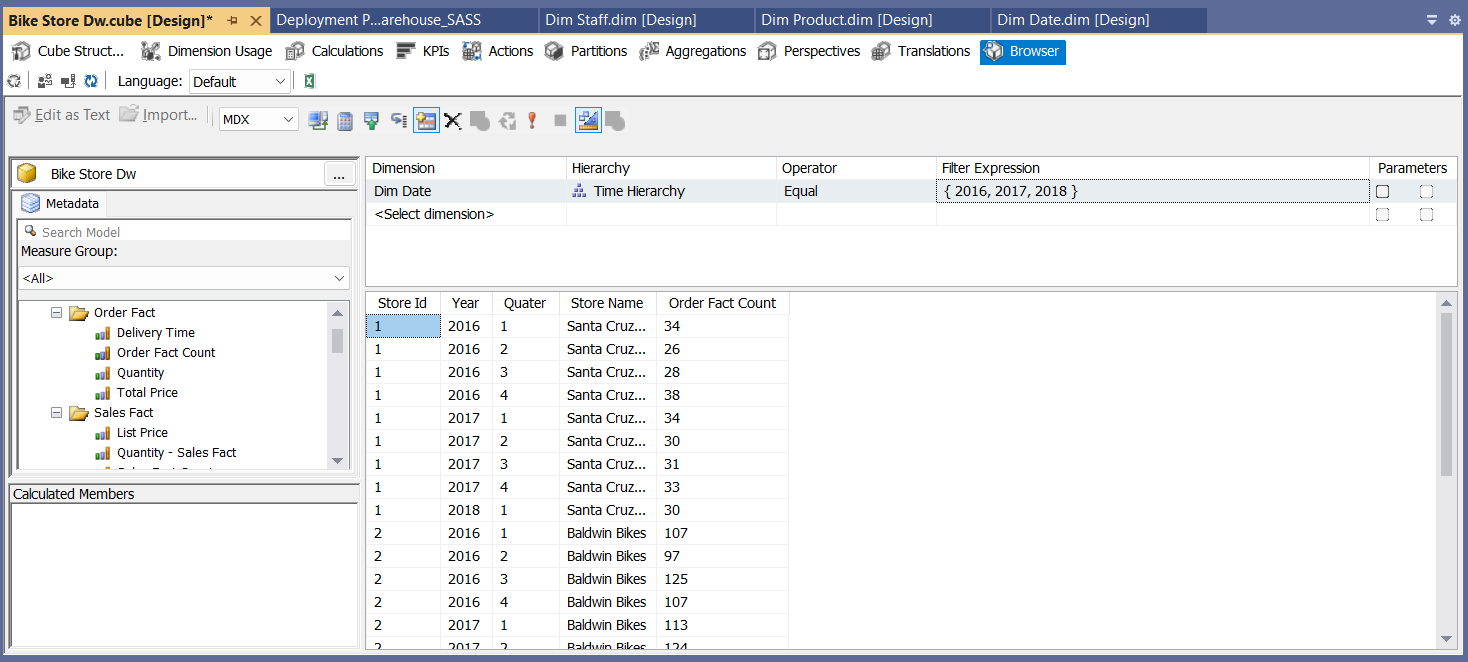
Tổng số lượng đơn hàng và tổng doanh thu của chuỗi (3 cửa hàng) từ trước đến nay.



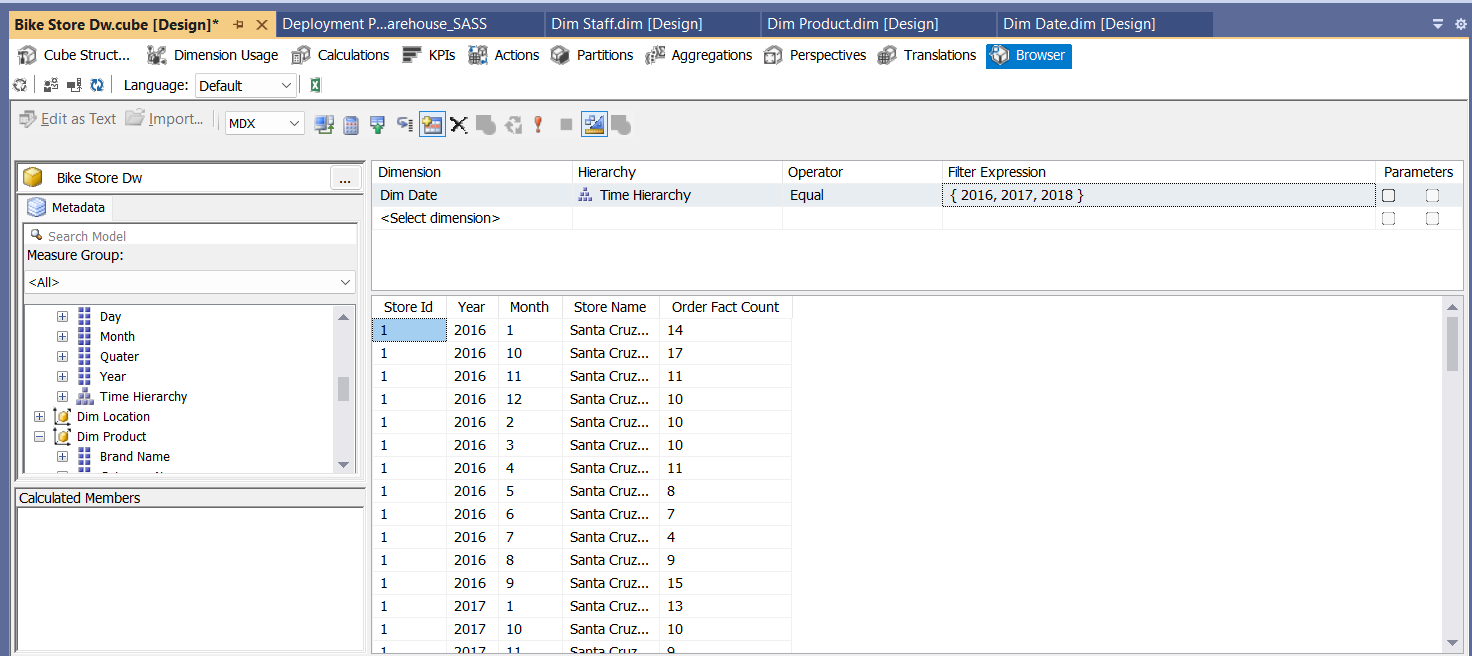
Tổng số lượng đơn hàng và doanh thu của từng cửa hàng từ trước đến nay



Tổng số lượng đơn hàng theo từng quý của mỗi cửa hàng.



Tổng giá trị sản phẩm bán ra của mỗi cửa hàng theo từng tháng.

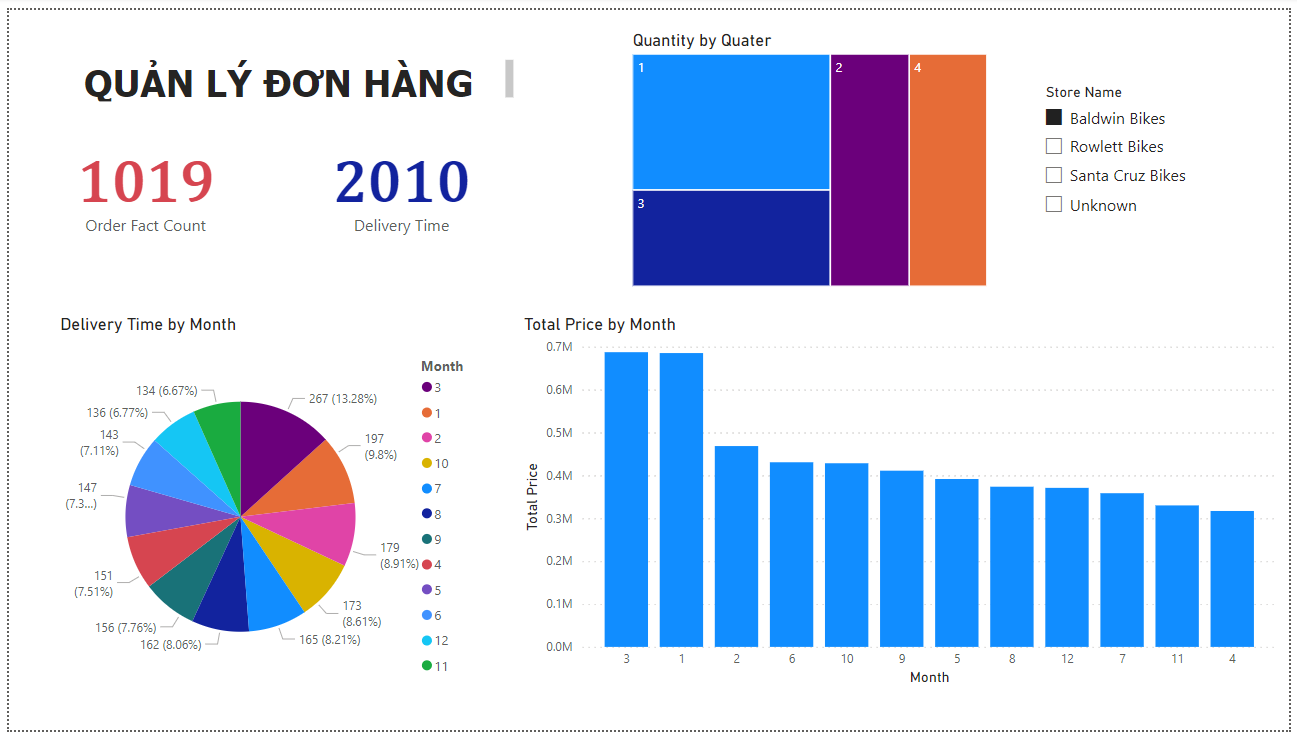


**4.4.2.2. Sử dụng Power BI**

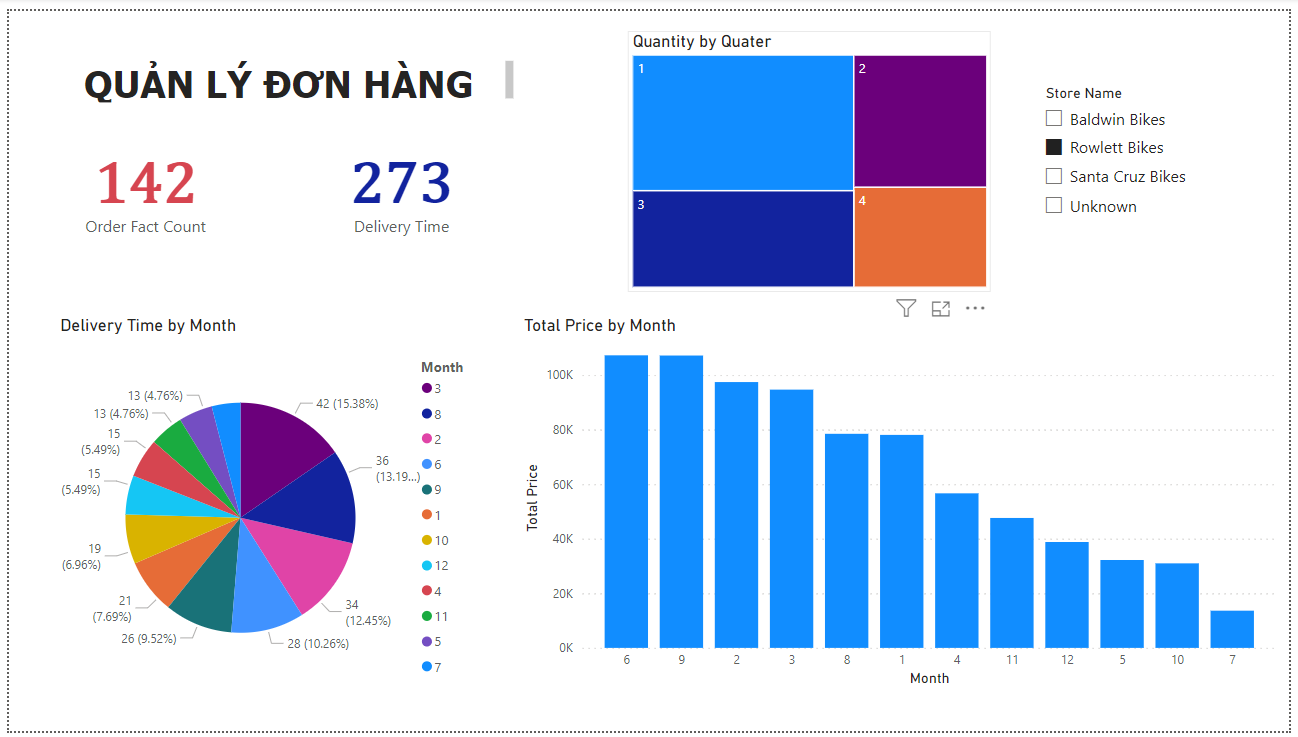
Nhìn biểu đồ ta thấy được tổng số lượng đơn hàng mà chuỗi của hàng đã bán được và và tổng thời gian đã tốn để giao hàng cho khách.



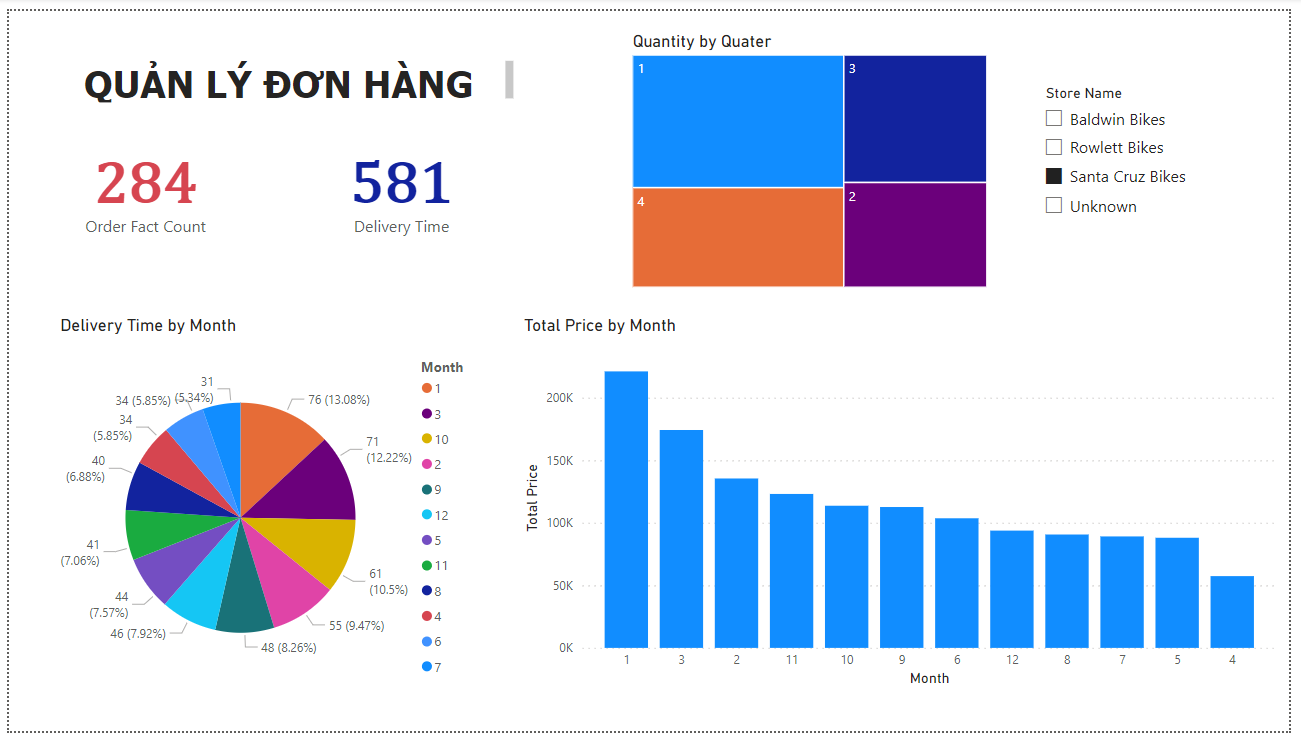
Bên cạnh đó, để thuận tiện cho việc quản lý, biểu đồ còn cung cấp số lượng đơn hàng đã bán ra theo theo từng quý.



Đồng thời tính toán giá tổng giá trị sản phẩm bán ra theo tháng của từng cửa hàng và chuỗi cửa hàng để người quản lý có thể biết rõ được doanh thu của cửa hàng thu được là bao nhiêu.



Biểu đồ còn cung cấp cho người dùng biết được thời gian giao hàng của mỗi tháng là bao nhiêu, qua đó sẽ ra những phương án phù hợp, để đơn hàng có thể đến được tay khách hàng một cách nhanh chóng nâng cao chất lượng phục vụ.



Qua biểu đồ trên, người quản lý có thể quản lý chuỗi cửa hàng một cách hiệu quả thông qua việc quản lý số đơn hàng, doanh thu và thời gian giao hàng. Đồng thời đưa ra những phương án thay đổi kịp thời nhằm nâng cao chất lượng phục vụ khách hàng.

# CHƯƠNG 5: KẾT LUẬN

## 5.1. Kết quả đạt được

Thực hiện xây dựng *‘“XÂY DỰNG KHO DỮ LIỆU CHO CỬA HÀNG BÁN LẺ XE ĐẠP”* giúp chúng em hiểu sâu hơn về kiến thức môn học Kho Dữ liệu và thành thạo hơn trong việc sử dụng các công cụ hỗ trợ quá trình ETL đưa dữ liệu vào kho dữ liệu như công cụ SSIS, SSAS, các công cụ để trực quan hóa dữ liệu như Power BI Desktop. Nâng cao kỹ năng nhìn nhận và tiếp thu, xử lý thông tin đầu vào, đặt những câu hỏi tiền đề và thảo luận hướng đi để giải quyết và trả lời câu hỏi.

## 5.2. Những hạn chế

Trong quá trình thực hiện xây dựng kho dữ liệu, nhóm chúng em không tránh khỏi gặp một số những khó khăn vì còn hạn chế về mặt kiến thức. Tập dữ liệu còn nhiều dữ kiện chưa được khai thác hết. Những thực hiện của chúng em chưa đạt đến mức nâng cao trong tổng thể quá trình xây dựng một kho dữ liệu hoàn chỉnh.

## 5.3. Tài liệu tham khảo

Tài liệu các file PDF hướng dẫn về Data Integration với SQL Server, SSIS, SSAS của GVHD Ths. Nguyễn Văn Thành.