

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HỒ CHÍ MINH  
KHOA CÔNG NGHỆ THÔNG TIN



HCMUTE

BÁO CÁO ĐỒ ÁN CUỐI KỲ  
**ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU BÁN HÀNG  
CHO CÔNG TY THIẾT BỊ PHẦN CỨNG**

Môn học: KHO DỮ LIỆU  
Mã lớp học phần: DAWH430784\_23\_2\_01  
GVHD: ThS. Nguyễn Văn Thành

Nhóm sinh viên thực hiện: Nhóm 7

|                      |          |
|----------------------|----------|
| Nguyễn Anh Tuấn      | 21133113 |
| Nguyễn Hoàn Thảo     | 21133081 |
| Nguyễn Thành Trung   | 21133090 |
| Tăng Huỳnh Minh Tiến | 21133088 |

TP. Hồ Chí Minh, 13 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   | Nguyễn Anh Tuân      | 21133113 | 100%           |
| 2   | Nguyễn Hoàn Thảo     | 21133081 | 100%           |
| 3   | Nguyễn Thành Trung   | 21133090 | 100%           |
| 4   | Tăng Huỳnh Minh Tiến | 21133088 | 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 cảm ơn sâu sắc đến **thầy Nguyễn Văn Thành**, thầy đã tận tình giảng dạy và hỗ trợ trong suốt quá trình tìm hiểu và học tập môn **Khoa dữ liệu**. Trong quá trình học tập, thầy đã nhiệt huyết trong từng lời giải đáp, những kiến thức, kỹ năng, tài liệu quý giá mà thầy đã truyền đạt học không chỉ là nền tảng cho quá trình thực hiện báo cáo cuối kì mà còn là kiến thức nền tảng cho nhiều phần sau này. Rất nhiều công sức và nỗ lực đã được bỏ ra, nhưng do chưa có kinh nghiệm trong việc xây dựng đề tài và những hạn chế về kiến thức, bài báo cáo này không tránh khỏi sai sót. Chúng em luôn sẵn sàng tiếp thu những ý kiến từ thầy để cải thiện bài báo cáo cũng như đề tài lần này. Một lần nữa, nhóm em xin bày tỏ lòng biết ơn đến thầy. Xin kính chúc thầy luôn dồi dào sức khỏe, đạt được nhiều thành công trong công việc.

## MỤC LỤC

|                                                                                        |    |
|----------------------------------------------------------------------------------------|----|
| CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI .....                                                    | 1  |
| 1.1. Lý do chọn đề tài.....                                                            | 1  |
| 1.2. Tổng quan về tập dữ liệu .....                                                    | 1  |
| 1.2.1. Nguồn dữ liệu .....                                                             | 1  |
| 1.2.2. Mô tả chi tiết tập dữ liệu .....                                                | 2  |
| a) Tập hardwareStore.csv .....                                                         | 2  |
| b) Bốn bảng dữ liệu (order_items,orders,customers,employees )trong<br>OT Database..... | 2  |
| 1.2.3. Import file csv vào 1 Table Store trong Database để dễ dàng sử dụng             | 4  |
| 1.3. Giới thiệu các công cụ được sử dụng.....                                          | 4  |
| CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP) .....                       | 5  |
| 2.1. Xác định các Business Process và bảng Fact.....                                   | 5  |
| 2.1.1. Business Process: Sales Analysis.....                                           | 5  |
| 2.1.2. Business Process: Inventory Management.....                                     | 5  |
| 2.2. Xây dựng các bảng Dimension .....                                                 | 6  |
| 2.2.1. DimDate .....                                                                   | 6  |
| 2.2.2. DimProduct .....                                                                | 6  |
| 2.2.3. DimLocation .....                                                               | 6  |
| 2.2.4. DimWareHouse.....                                                               | 6  |
| 2.2.5. DimCustomer.....                                                                | 6  |
| 2.2.6. DimEmployee .....                                                               | 7  |
| 2.2.7. DimCategory .....                                                               | 7  |
| 2.3. Star Schema (Lược đồ hình sao).....                                               | 7  |
| CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS).....                                         | 8  |
| 3.1. Import Dữ liệu vào các bảng dimension: .....                                      | 8  |
| 3.1.1.Date Dimension .....                                                             | 10 |
| a) Load từ nguồn vào bảng stgDate .....                                                | 10 |
| b) Load dữ liệu từ stgDate vào DimDate .....                                           | 14 |
| 3.1.2.Product Dimension.....                                                           | 16 |
| a) Load từ nguồn vào bảng stgProduct .....                                             | 16 |
| b) Load dữ liệu từ stgProduct vào DimProduct .....                                     | 19 |
| 3.3.3.Location Dimension.....                                                          | 22 |
| a) Load từ nguồn vào bảng stgLocation .....                                            | 22 |
| b) Load dữ liệu từ stgLocation vào bảng DimLocation.....                               | 25 |
| 3.3.4.Customer Dimension .....                                                         | 29 |
| a) Load dữ liệu từ nguồn vào stgCustomer.....                                          | 29 |
| b) Load dữ liệu từ stgCustomer vào bảng DimCustomer .....                              | 31 |
| 3.3.5.Employee Dimension.....                                                          | 34 |
| a) Load dữ liệu từ nguồn vào stgEmployee .....                                         | 34 |
| b) Load dữ liệu từ stgEmployee vào bảng DimEmployee .....                              | 37 |
| 3.3.6.Warehouse Dimension .....                                                        | 41 |
| a) Load dữ liệu từ nguồn vào stgWarehouse .....                                        | 41 |
| b) Load dữ liệu từ stgWarehouse vào DimWarehouse .....                                 | 45 |

|                                                                                             |     |
|---------------------------------------------------------------------------------------------|-----|
| 3.3.7.Category Dimension .....                                                              | 49  |
| a) Load dữ liệu từ nguồn vào stgCategory .....                                              | 49  |
| b) Load dữ liệu từ stgCategory vào DimCategory.....                                         | 53  |
| 3.2. Import dữ liệu vào bảng các bảng fact.....                                             | 57  |
| 3.2.1. Fact Sales .....                                                                     | 57  |
| a) Load dữ liệu từ nguồn vào stgSales.....                                                  | 57  |
| b) Load dữ liệu vào Sales_fact.....                                                         | 61  |
| 3.2.2. Fact Product Inventory.....                                                          | 75  |
| a) Load dữ liệu từ nguồn vào stgInventory.....                                              | 75  |
| b) Load dữ liệu vào Inventory_fact .....                                                    | 78  |
| CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS).....                                                     | 91  |
| 4.1. Quá trình xây dựng mô hình .....                                                       | 91  |
| 4.1.1. Tạo Data Source.....                                                                 | 91  |
| 4.1.2. Tạo Data Source View .....                                                           | 93  |
| 4.2. Quá trình xây dựng khối Cube.....                                                      | 96  |
| 4.3. Cấu hình Hierarchy .....                                                               | 98  |
| 4.3.1. Tạo Hierarchy cho Dim Date .....                                                     | 98  |
| 4.3.2. Tạo Hierarchy cho Dim Location .....                                                 | 100 |
| 4.3.3. Tạo Hierarchy cho Dim Category .....                                                 | 101 |
| 4.3.4. Tạo Hierarchy cho Dim Product .....                                                  | 103 |
| 4.3.5. Tạo Hierarchy cho Dim Employee .....                                                 | 105 |
| 4.3.6. Tạo Hierarchy cho Dim WareHouse .....                                                | 106 |
| 4.3.7. Tạo Hierarchy cho Dim Customer .....                                                 | 107 |
| 4.4. Thực hiện phân tích dữ liệu .....                                                      | 109 |
| 4.4.1. Câu hỏi: Số lượng các loại sản phẩm bán ra trong tháng/ quý/ năm.                    | 109 |
| 4.4.1.1. Sử dụng công cụ SSAS .....                                                         | 109 |
| 4.4.1.2. Sử dụng Pivot Table trong Excel .....                                              | 110 |
| 4.4.1.3. Sử dụng Power BI.....                                                              | 110 |
| 4.4.2. Câu hỏi: Cho biết doanh số các sản phẩm trong tháng/ quý/ năm...111                  | 111 |
| 4.4.2.1. Sử dụng công cụ SSAS .....                                                         | 111 |
| 4.4.2.2. Sử dụng Pivot Table trong Excel .....                                              | 112 |
| 4.4.2.3. Sử dụng Power BI.....                                                              | 112 |
| 4.4.3. Câu hỏi: Cho biết doanh số bán hàng của từng nhân viên trong tháng/ quý/ năm.....114 | 114 |
| 4.4.3.1. Sử dụng công cụ SSAS .....                                                         | 114 |
| 4.4.3.2. Sử dụng Pivot Table trong Excel .....                                              | 116 |
| 4.4.3.3. Sử dụng Power BI.....                                                              | 116 |
| 4.4.4. Câu hỏi: Chi tiêu của khách hàng theo thời gian .....                                | 118 |
| 4.4.4.1. Sử dụng công cụ SSAS .....                                                         | 118 |
| 4.4.4.2. Sử dụng Pivot Table trong Excel .....                                              | 119 |
| 4.4.4.3. Sử dụng Power BI.....                                                              | 120 |
| 4.4.5. Câu hỏi: Quản lý số lượng sản phẩm phân bổ trong các kho. ....121                    | 121 |
| 4.4.5.1. Sử dụng công cụ SSAS .....                                                         | 121 |

|                                                |     |
|------------------------------------------------|-----|
| 4.4.5.2. Sử dụng Pivot Table trong Excel ..... | 122 |
| 4.4.5.3. Sử dụng Power BI.....                 | 122 |
| CHƯƠNG 5: KẾT LUẬN .....                       | 124 |
| 5.1. Kết quả đạt được .....                    | 124 |
| 5.2. Những hạn chế .....                       | 124 |
| 5.3. Tài liệu tham khảo .....                  | 124 |

# PHÂN CÔNG NHIỆM VỤ

| Nhiệm vụ                                                                           | Minh Tiến | Hoàn Thảo | Thành Trung | Anh Tuấn |
|------------------------------------------------------------------------------------|-----------|-----------|-------------|----------|
| Tìm kiếm tập dữ liệu                                                               |           |           |             | x        |
| <b>Mô hình hóa dữ liệu vào workbook</b>                                            | x         | x         | x           | x        |
| + DimDate                                                                          |           |           |             |          |
| + DimWareHouse                                                                     | x         |           |             |          |
| + DimCustomer                                                                      |           |           |             |          |
| + DimLocation                                                                      |           |           |             | x        |
| + InventoryFact                                                                    |           |           |             |          |
| + SalesFact                                                                        |           |           | x           |          |
| + DimEmployee                                                                      |           |           |             |          |
| + DimProduct                                                                       |           | x         |             |          |
| + DimCategory                                                                      |           |           |             |          |
| <b>Thực hiện ETL đưa dữ liệu vào các bảng dim và fact dùng SSIS</b>                | x         | x         | x           | x        |
| + Tạo và đưa dữ liệu vào các dim stages và các bảng dim                            |           |           |             | x        |
| + Tạo và đưa dữ liệu vào FactSales stage và FactSales                              |           |           | x           |          |
| + Tạo và đưa dữ liệu vào InventoryFact stage và InventoryFact                      | x         |           |             |          |
| Nhập dữ liệu vào SSAS, tạo data source, data source view và cube                   |           | x         |             |          |
| <b>Tạo các phân cấp hierarchy cho các bảng dim và các attribute relationships.</b> | x         | x         | x           | x        |
| + Hierarchy cho DimDate                                                            | x         |           |             |          |
| + Hierarchy cho                                                                    |           |           | x           |          |

|                                                 |  |   |  |   |
|-------------------------------------------------|--|---|--|---|
| DimLocation,<br>DimProduct                      |  |   |  |   |
| + Hierarchy cho<br>DimCategory,<br>DimEmployee  |  | x |  |   |
| + Hierarchy cho<br>DimWareHouse,<br>DimCustomer |  |   |  | x |

| Nhiệm vụ                                                                                                                                   | Minh Tiến | Hoàn Thảo | Thành Trung | Anh Tuấn |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----------|-----------|-------------|----------|
| <b>Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Sales Analysis</b>                                                        | x         | x         | x           | x        |
| + Số lượng các loại sản phẩm bán ra trong tháng/ quý/ năm.                                                                                 | x         |           |             |          |
| + Cho biết doanh số các sản phẩm trong tháng/ quý/ năm.                                                                                    |           |           | x           |          |
| + Câu hỏi: Cho biết doanh số bán hàng của từng nhân viên trong tháng/ quý/ năm.                                                            |           | x         |             |          |
| + Câu hỏi: Chi tiêu của khách hàng theo thời gian.                                                                                         |           |           |             | x        |
| <b>Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Inventory management: quản lý số lượng sản phẩm phân bổ trong các kho</b> | x         | x         | x           |          |
| + Trả lời câu hỏi dùng SSAS                                                                                                                |           | x         |             |          |
| + Trả lời câu hỏi dùng Power BI Desktop                                                                                                    | x         |           |             |          |
| + Trả lời câu hỏi dùng Excel Pivot                                                                                                         |           |           | x           |          |

# CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

## 1.1. Lý do chọn đề tài

- Trong thời đại công nghệ 4.0 ngày nay, Hầu hết mọi nơi, vào bất kỳ thời điểm nào, đều bao gồm sự hiện diện và hoạt động của các công nghệ. Sự tăng cao về nhu cầu sử dụng các thiết bị công nghệ làm cho việc cung cấp, bảo trì và sửa chữa trở nên vô cùng quan trọng. Để đảm bảo rằng các quá trình từ sản xuất, vận chuyển, lưu trữ đến phân phối sản phẩm diễn ra một cách mượt mà và tự động, một hệ thống quản lý toàn diện là không thể thiếu. Thường thường, dữ liệu được thu thập và lưu trữ ở mỗi giai đoạn khác nhau, và việc kết hợp chúng một cách cẩn thận có thể giúp doanh nghiệp ra những quyết định chính xác và hiệu quả hơn. Vì lý do này, nhóm của chúng em đã quyết định nghiên cứu và thực hiện đề tài " Xây dựng kho dữ liệu cho chuỗi cửa hàng bán lẻ phần cứng". Mục tiêu của chúng em là tạo ra một Kho dữ liệu linh hoạt và dễ quản lý, giúp công ty nắm bắt thông tin một cách hiệu quả và từ đó, đưa ra các quyết định chiến lược và có ý nghĩa.

## 1.2. Tổng quan về tập dữ liệu

### 1.2.1. Nguồn dữ liệu

- Nhóm sử dụng tập dữ liệu hardwareStore.csv lấy từ kaggle kết hợp với bốn bảng dữ liệu (order\_items,orders,customers,employees )trong OT Database được lấy từ trang web Oracle Tutorial.
- Đường dẫn tập csv [hardwareStore.csv](#) (Đã thêm cột date để phục vụ cho việc tạo data warehouse).
- Đường dẫn tải tập dữ liệu (Nhóm chỉ lấy 4 bảng order\_items, orders, customers, employees) : [OT Oracle Sample Database](#)

## 1.2.2. Mô tả chi tiết tập dữ liệu

### a) Tập hardwareStore.csv

Gồm 23 cột và 408104 dòng

|                        |                         |
|------------------------|-------------------------|
| Tên cột                | Mô tả                   |
| Date                   | Ngày kiểm kê            |
| CATEGORY_ID            | ID của loại sản phẩm    |
| CATEGORY_NAME          | Tên loại sản phẩm       |
| PRODUCT_ID             | ID của sản phẩm         |
| PRODUCT_NAME           | Tên sản phẩm            |
| DESCRIPTION            | Mô tả sản phẩm          |
| DESCRIPTION - Detail 1 | Mô tả chi tiết sản phẩm |
| DESCRIPTION - Detail 2 | Mô tả chi tiết sản phẩm |
| DESCRIPTION - Detail 3 | Mô tả chi tiết sản phẩm |
| DESCRIPTION - Detail 4 | Mô tả chi tiết sản phẩm |
| STANDARD_COST          | Giá sản xuất sản phẩm   |
| LIST_PRICE             | Giá niêm yết            |
| COUNTRY_ID             | ID của quốc gia         |
| REGION_ID              | ID của vùng             |
| LOCATION_ID            | ID của khu vực          |
| WAREHOUSE_ID           | ID của kho              |
| QUANTITY               | Số lượng sản phẩm       |
| WAREHOUSE_NAME         | Tên kho                 |
| ADDRESS                | Địa chỉ kho             |
| POSTAL_CODE            | Mã bưu điện             |
| CITY                   | Thành phố               |
| STATE                  | Bang                    |
| COUNTRY_NAME           | Tên đất nước            |

### b) Bốn bảng dữ liệu (order\_items,orders,customers,employees )trong OT Database

Bảng orders (105 dòng, 5 cột)

| Biến        | Mô tả                                                         |
|-------------|---------------------------------------------------------------|
| ORDER_ID    | Mã đơn hàng (Khóa chính)                                      |
| CUSTOMER_ID | Mã khách hàng (Khóa ngoại trỏ đến CUSTOMERS(CUSTOMER_ID))     |
| STATUS      | Tình trạng đơn hàng                                           |
| SALESMAN_ID | Mã người bán hàng (Khóa ngoại trỏ đến EMPLOYEES(EMPLOYEE_ID)) |

|            |                   |
|------------|-------------------|
| ORDER_DATE | Ngày tạo đơn hàng |
|------------|-------------------|

Bảng order\_items (665 dòng 5 cột)

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

Bảng Customers (319 dòng 5 cột)

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

Bảng employees ( 107 dòng, 8 cột)

| Biến        | Mô tả                                  |
|-------------|----------------------------------------|
| EMPLOYEE_ID | Mã nhân viên (Khóa chính)              |
| FIRST_NAME  | Họ nhân viên                           |
| LAST_NAME   | Tên nhân viên                          |
| EMAIL       | Email của nhân viên                    |
| PHONE       | Số điện thoại của nhân viên            |
| HIRE_DATE   | Ngày tuyển dụng làm việc của nhân viên |

|            |                                |
|------------|--------------------------------|
| MANAGER_ID | Mã người quản lý của nhân viên |
| JOB_TITLE  | Vị trí công việc của nhân viên |

### 1.2.3. Import file csv vào 1 Table Store trong Database để dễ dàng sử dụng

Để có thể dễ dàng kiểm soát các kiểu dữ liệu khi đưa vào Datawarehouse thì nhóm em quyết định đưa dữ liệu từ file CSV vào table Store để dễ dàng kiểm soát

### **1.3. Giới thiệu các công cụ được sử dụng**

- Visual Studio 2019/2022 tích hợp SSIS (SQL Server Integration Services) và SSAS (SQL Server Analysis Services).
  - SQL Server Management Studio 2019
  - Power BI.

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

## 2.1. Xác định các Business Process và bảng Fact

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

|   | A                     | B              | C               | D                        | E                                 | F         | G       | H        | I        | J    | K        | L        | M    | N    | O     | P     | Q     | R     | S     |
|---|-----------------------|----------------|-----------------|--------------------------|-----------------------------------|-----------|---------|----------|----------|------|----------|----------|------|------|-------|-------|-------|-------|-------|
| 1 | Instructions!         |                |                 |                          |                                   | Warehouse | Country | Category | Products | Date | Employee | Customer | Dim8 | Dim9 | Dim10 | Dim11 | Dim12 | Dim13 | Dim14 |
| 2 | Business Process Name | Fact Table     | Fact Grain Type | Granularity              | Facts                             |           |         |          |          |      |          |          |      |      |       |       |       |       |       |
| 3 | Sales Reporting       | sales_fact     | Transaction     | one row per order detail | Quantity, Unit Price, Total Price | x         | x       | x        | x        | x    | x        |          |      |      |       |       |       |       |       |
| 4 | Inventory Analysis    | Inventory_fact | Periodic        | one row per product      | Quantity, List_Price, Profit      | x         | x       | x        | x        | x    |          |          |      |      |       |       |       |       |       |
| 5 |                       |                |                 |                          |                                   |           |         |          |          |      |          |          |      |      |       |       |       |       |       |

### 2.1.1. Business Process: Sales Analysis

- Các câu hỏi cụ thể được đặt ra:
  - + Số lượng các loại sản phẩm bán ra trong tháng/ quý/ năm.
  - + Cho biết doanh số các sản phẩm trong tháng/ quý/ năm.
  - + Cho biết doanh số bán hàng của từng nhân viên trong tháng/ quý/ năm.
  - + Câu hỏi: Chi tiêu của khách hàng theo thời gian.

#### - Bảng SalesFact

| Instructions! | Dimension / Fact Table | Attribute / Fact Name | Description                |
|---------------|------------------------|-----------------------|----------------------------|
|               | SalesFact              | ProductKey            | Foreignkey to DimProduct   |
|               | SalesFact              | CategoryKey           | Foreignkey to DimCategory  |
|               | SalesFact              | OrderDateKey          | Foreignkey to DimOrderDate |
|               | SalesFact              | CustomerKey           | Foreignkey to DimCustomer  |
|               | SalesFact              | EmployeeKey           | Foreignkey to DimEmployee  |
|               | SalesFact              | Quantity              | Quantity of each product   |
|               | SalesFact              | Unit_Price            | Unit price of each product |
|               | SalesFact              | Total_Price           | Total sales of store       |

### 2.1.2. Business Process: Inventory Management

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

#### - Bảng InventoryFact

| Instructions! | Dimension / Fact Table | Attribute / Fact Name | Description                           |
|---------------|------------------------|-----------------------|---------------------------------------|
|               | InventoryFact          | ProductKey            | Foreign to DimProduct                 |
|               | InventoryFact          | CategoryKey           | Foreign to DimCategory                |
|               | InventoryFact          | DateKey               | Foreign to DimDate                    |
|               | InventoryFact          | LocationKey           | Foreign to DimLocation                |
|               | InventoryFact          | WareHouseKey          | Foreign to DimWareHouse               |
|               | InventoryFact          | Quantity              | Quantity of each product in warehouse |
|               | InventoryFact          | List_Price            | List price of product in warehouse    |
|               | InventoryFact          | Profit                | Profit of each product in warehouse   |

## 2.2. Xây dựng các bảng Dimension

### 2.2.1. DimDate

| Instructions!          | Attribute / Fact Name | Description | Alternate Names | Sample Values |
|------------------------|-----------------------|-------------|-----------------|---------------|
| Dimension / Fact Table |                       |             |                 |               |
| Date                   | DayOfWeek             | DayOfWeek   |                 | 5             |
| Date                   | DayName               | DayName     |                 | Friday        |
| Date                   | DayOfMonth            | DayOfMonth  |                 | 1             |
| Date                   | DayOfYear             | DayOfYear   |                 | 1             |
| Date                   | WeekOfYear            | WeekOfYear  |                 | 1             |
| Date                   | MonthName             | MonthName   |                 | January       |
| Date                   | MonthOfYear           | MonthOfYear |                 | 1             |
| Date                   | Quarter               | Quarter     |                 | 1             |
| Date                   | Year                  | Year        |                 | 1999          |
| Date                   | IsWeekDay             | IsWeekDay   |                 | Y             |

### 2.2.2. DimProduct

| Instructions!          | Attribute / Fact Name | Description                  | Alternate Names | Sample Values                 |
|------------------------|-----------------------|------------------------------|-----------------|-------------------------------|
| Dimension / Fact Table |                       |                              |                 |                               |
| Products               | ProductID             | ID of products               |                 | 1                             |
| Products               | ProductName           | Name of products             |                 | Intel i6                      |
| Products               | Description           | Description product          |                 | Speed:3.6GHz,Cores:6,TDP:140W |
| Products               | Description_Detail_1  | Description product          |                 | Speed:3.6GHz                  |
| Products               | Description_Detail_2  | Description product          |                 | Cores:6                       |
| Products               | Description_Detail_3  | Description product          |                 | TDP:140W                      |
| Products               | Description_Detail_4  | Description product          |                 | Max RAM:32GB                  |
| Products               | Standard_Cost         | Costs to produce the product |                 | 535.47                        |

### 2.2.3. DimLocation

| Instructions!          | Attribute / Fact Name | Description      | Alternate Names | Sample Values |
|------------------------|-----------------------|------------------|-----------------|---------------|
| Dimension / Fact Table |                       |                  |                 |               |
| Location               | CountryID             | ID of country    |                 | AU            |
| Location               | RegionID              | ID of region     |                 | 3             |
| Location               | LocationID            | ID of location   |                 | 12            |
| Location               | CountryName           | Name of country  |                 | Mexico        |
| Location               | State                 | State of country |                 | Nuevo León    |
| Location               | City                  | City of country  |                 | Monterrey     |

### 2.2.4. DimWareHouse

| Instructions!          | Attribute / Fact Name | Description          | Alternate Names | Sample Values  |
|------------------------|-----------------------|----------------------|-----------------|----------------|
| Dimension / Fact Table |                       |                      |                 |                |
| WareHouse              | WarehouseID           | ID of warehouse      |                 | 9              |
| WareHouse              | Address               | Address of warehouse |                 | 1298 Vileparle |
| WareHouse              | WarehouseName         | Name of warehouse    |                 | Bombay         |

### 2.2.5. DimCustomer

| Instructions!          | Attribute / Fact Name | Description         | Alternate Names | Sample Values |
|------------------------|-----------------------|---------------------|-----------------|---------------|
| Dimension / Fact Table |                       |                     |                 |               |
| Customer               | CustomerID            | ID of customer      |                 | 2             |
| Customer               | Address               | Address of customer |                 | Yen Lang      |
| Customer               | CustomerName          | Name of customer    |                 | Kay Tran      |
| Customer               | Website               | Website             |                 | Bombay.com    |
| Customer               | Credit_Limit          | Credit              |                 | 1000          |

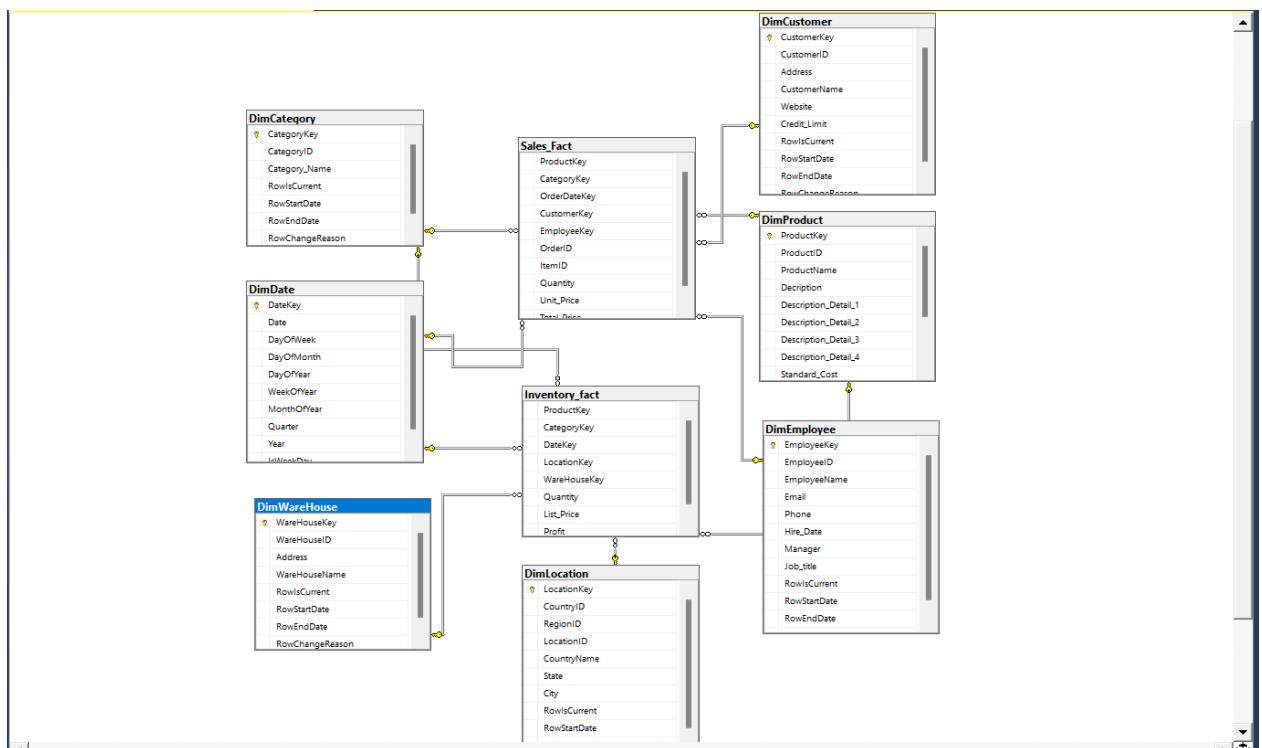
## 2.2.6. DimEmployee

| Instructions!          |          | Attribute / Fact Name | Description           | Alternate Names | Sample Values |
|------------------------|----------|-----------------------|-----------------------|-----------------|---------------|
| Dimension / Fact Table | Employee | EmployeeID            | ID of employee        |                 | 3             |
| Employee               |          | EmployeeName          | Name of employee      |                 | Alba          |
| Employee               |          | Email                 | Email of employee     |                 | abc@gmail.com |
| Employee               |          | Phone                 | Phone of employee     |                 | 0234478988    |
| Employee               |          | Hire Date             | Date of employee      |                 | 05/09/2013    |
| Employee               |          | Manager               | Manager of employee   |                 | Zala          |
| Employee               |          | Job_title             | Job_title of employee |                 | Staff         |

## 2.2.7. DimCategory

| Instructions!          |          | Attribute / Fact Name | Description      | Alternate Names | Sample Values |
|------------------------|----------|-----------------------|------------------|-----------------|---------------|
| Dimension / Fact Table | Category | CategoryID            | ID of category   |                 | 1             |
| Category               |          | Category_Name         | Name of category |                 | Storage       |

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

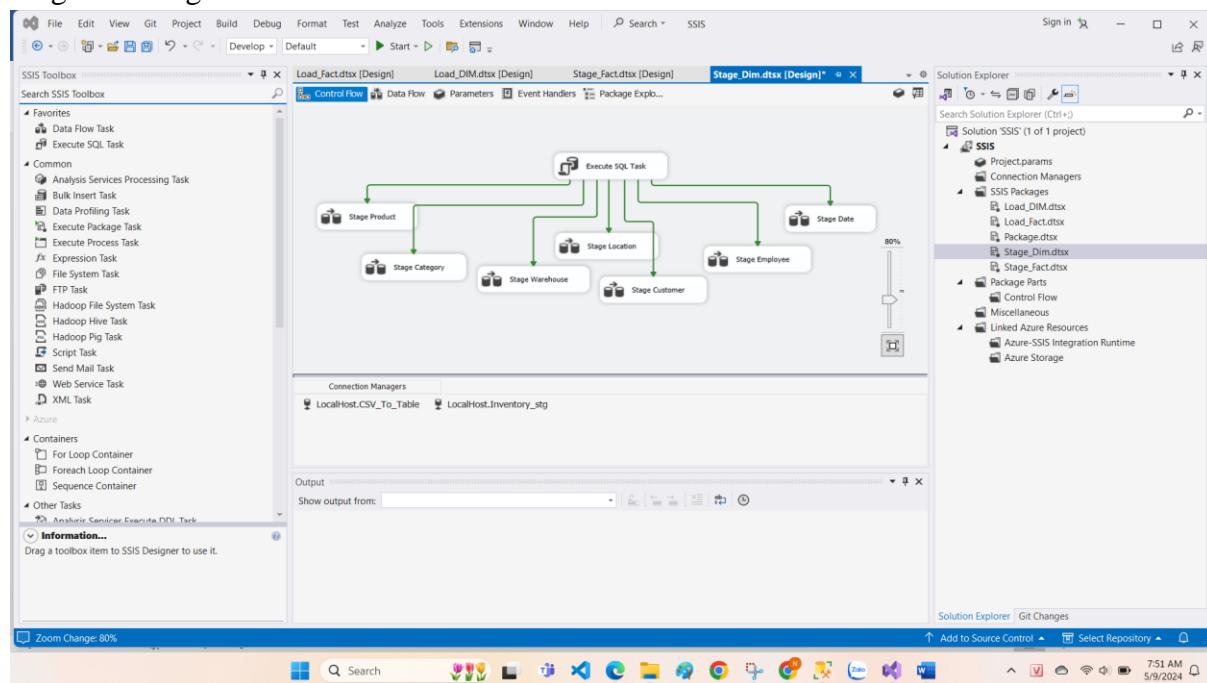


# CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

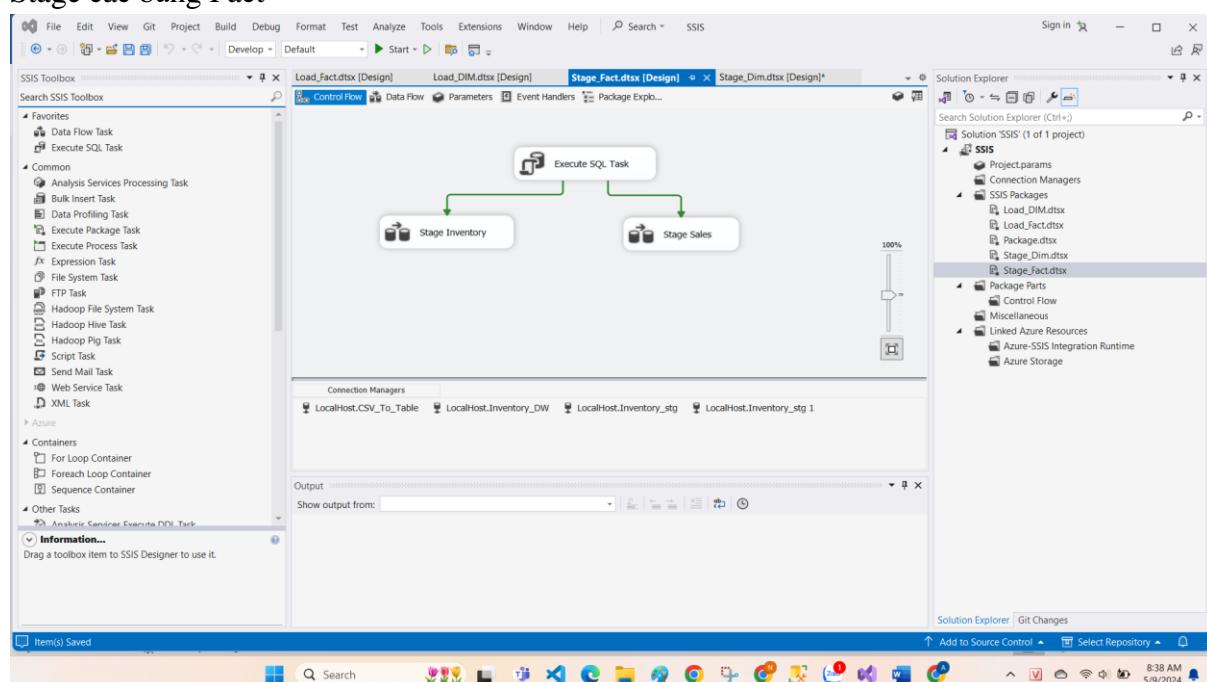
## 3.1. Import Dữ liệu vào các bảng dimension:

Ta phải Load từ SRC vào Stage

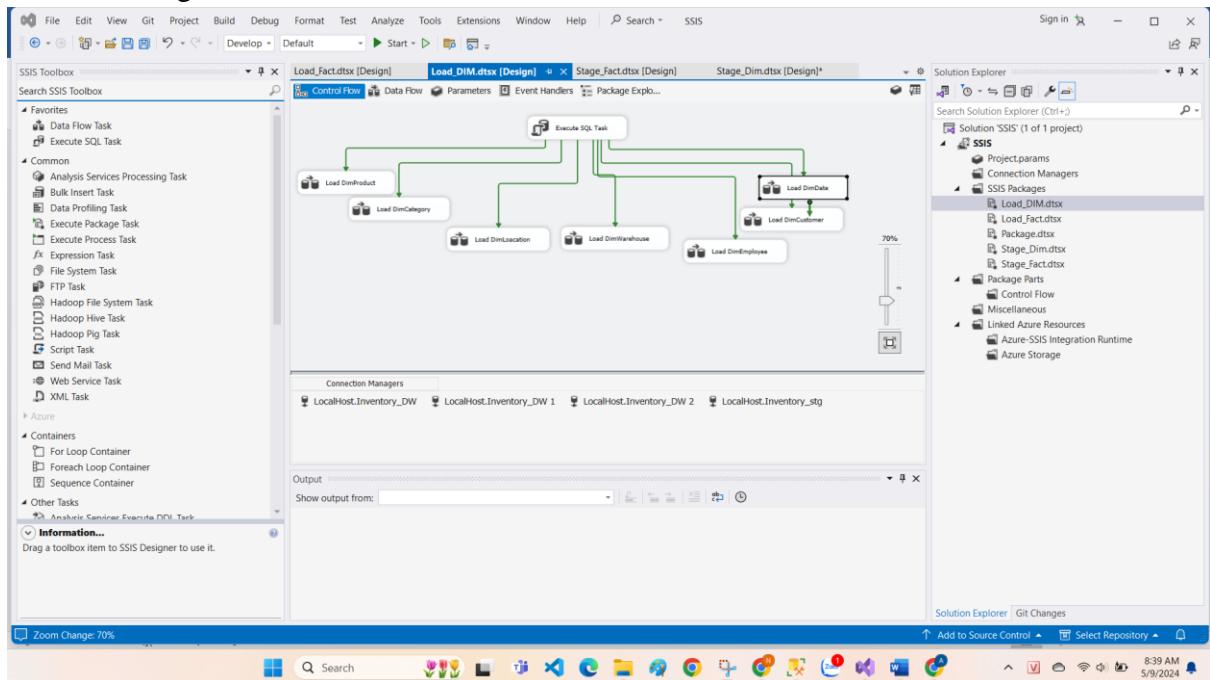
Stage các bảng Dim



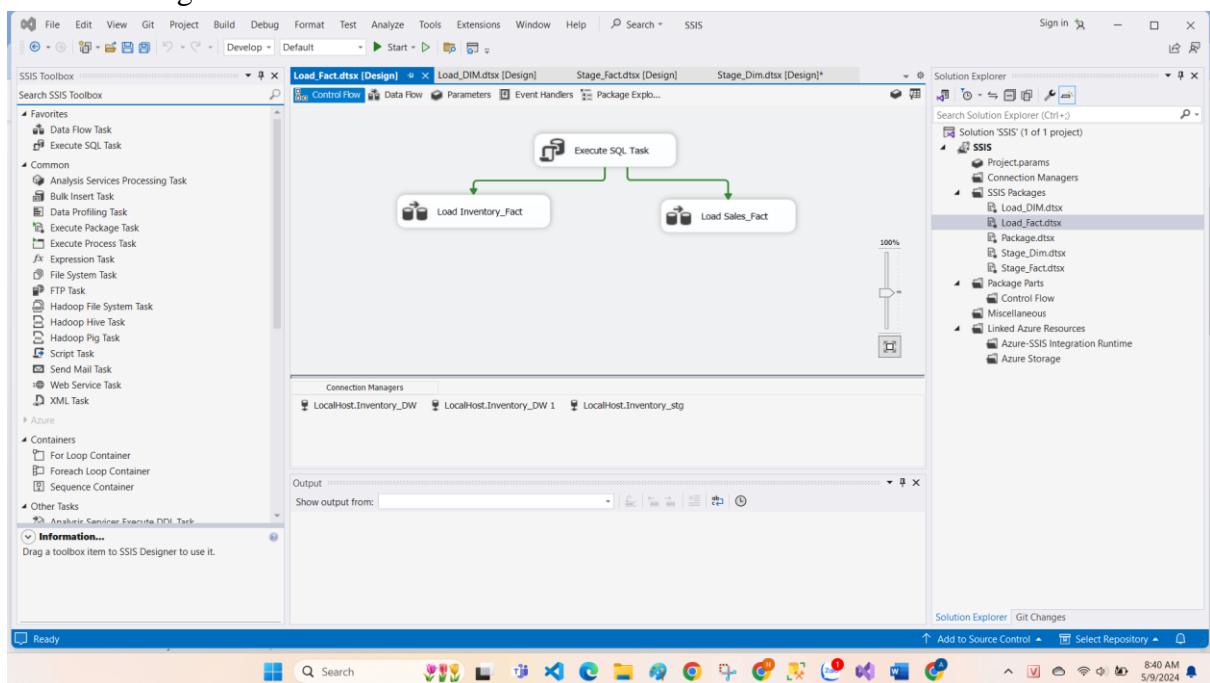
Stage các bảng Fact



Sau đó Load từ Stage vào các bảng Dim và bảng Fact  
 Load các bảng Dim

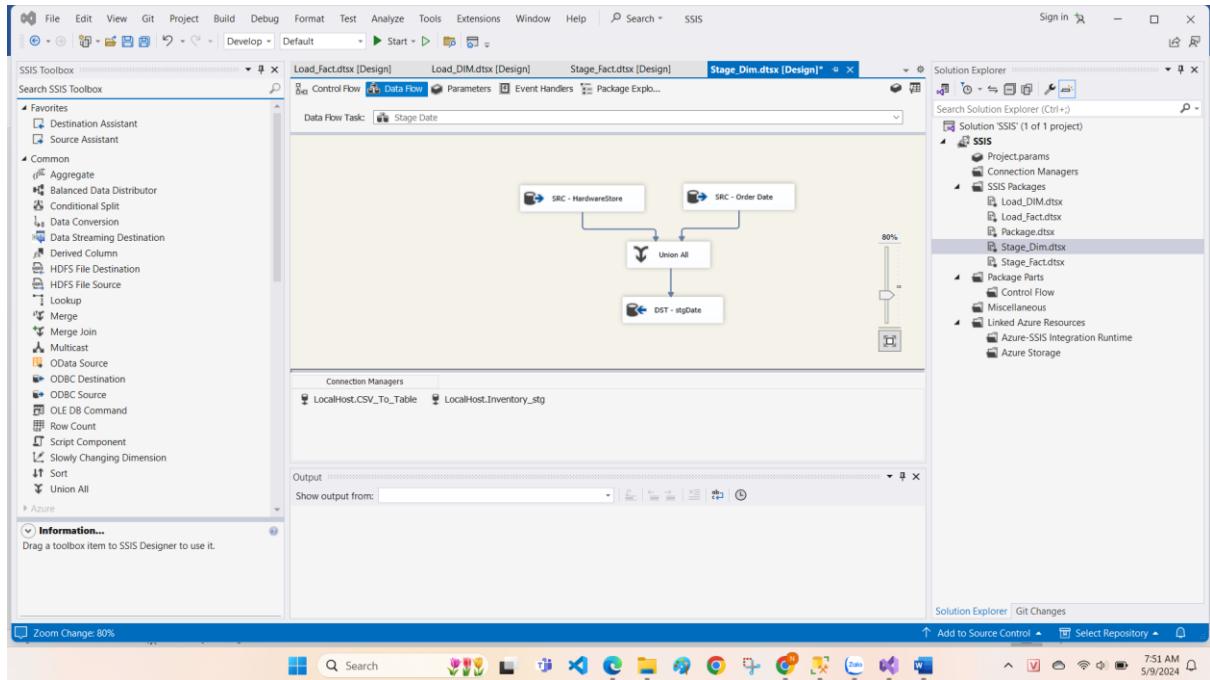


Load các bảng Fact

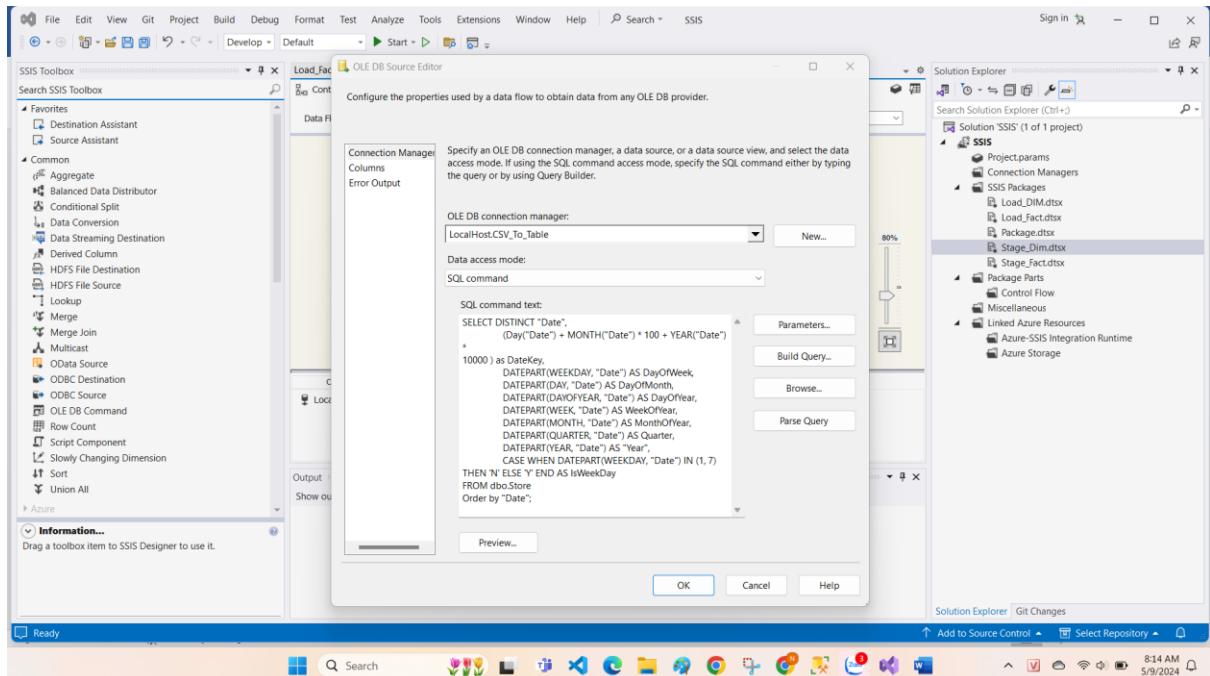


### 3.1.1.Date Dimension

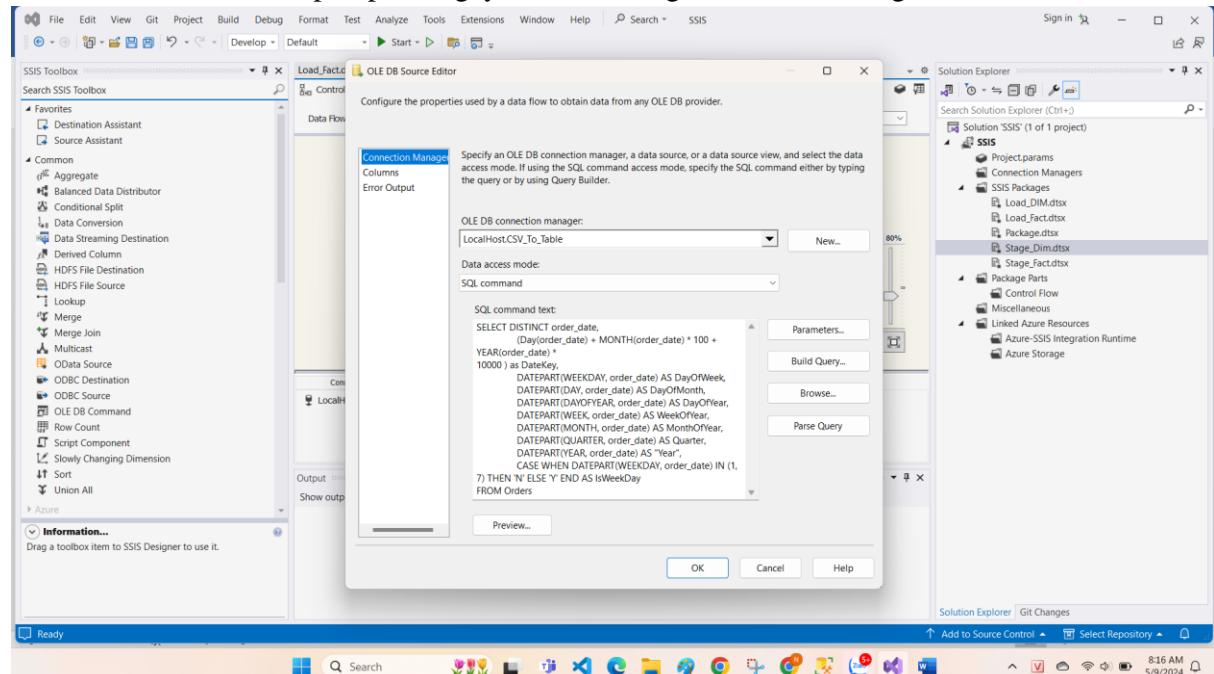
#### a) Load từ nguồn vào bảng stgDate



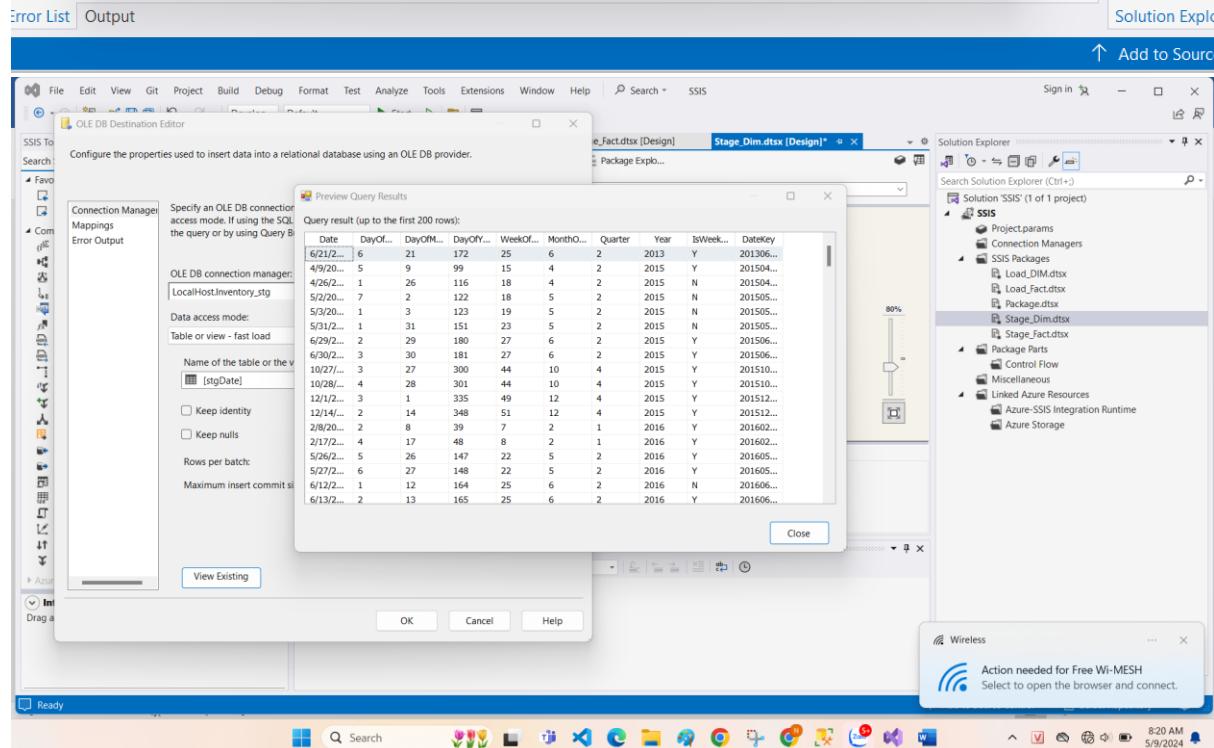
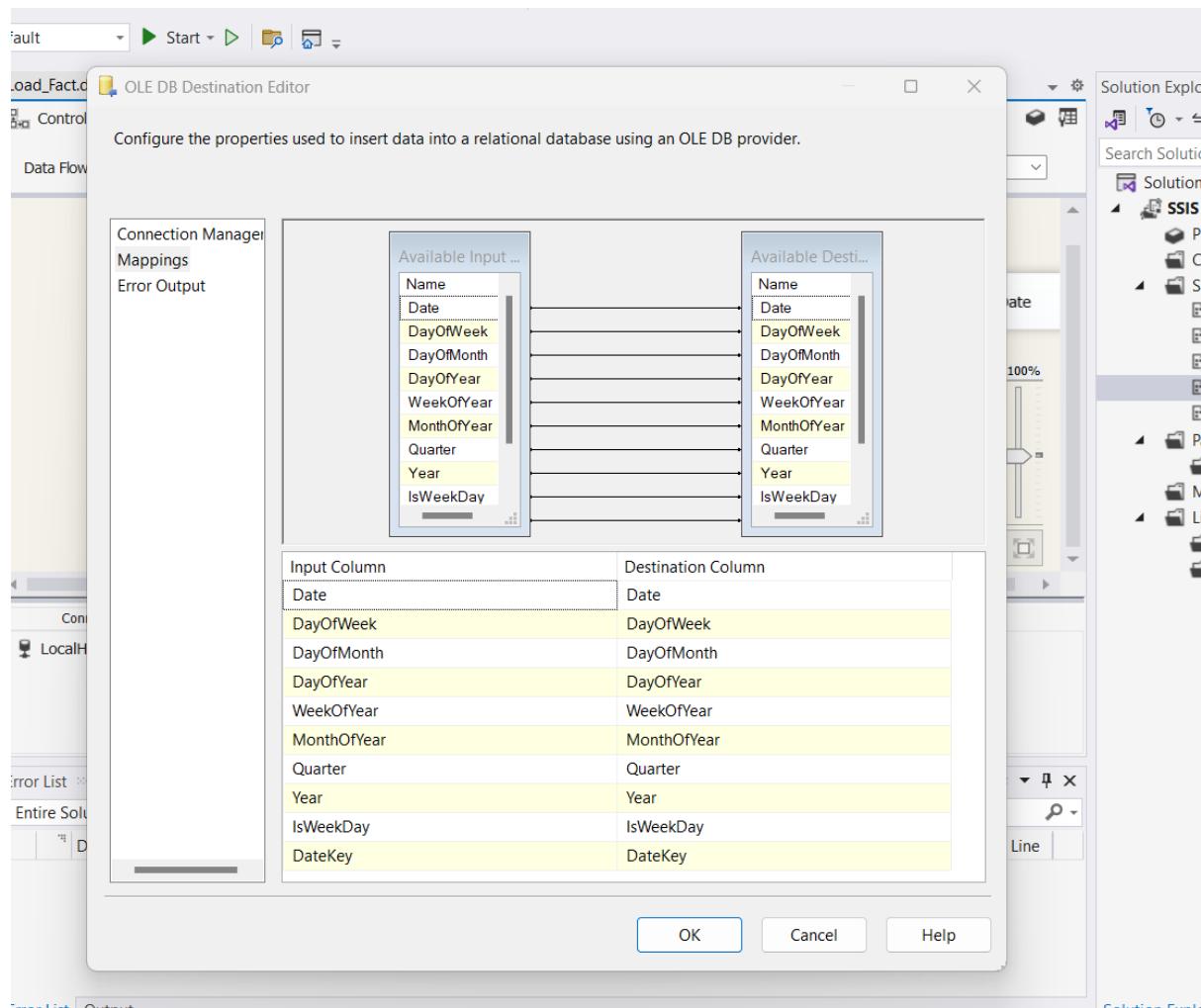
**\_SRC -HardwareStore** là tập hợp ngày lấy từ bảng Store dùng để thống kê các sản phẩm trong kho



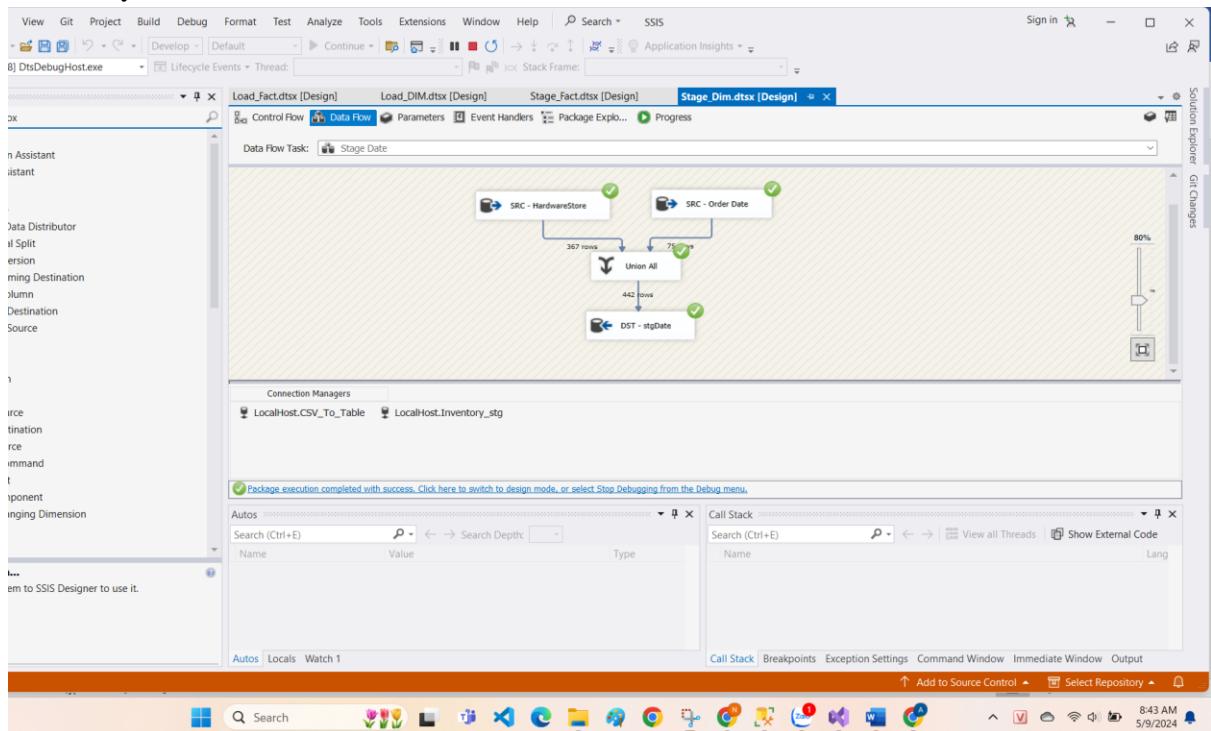
## SRC – Order Date là tập hợp các ngày có thực hiện giao dịch mua hàng



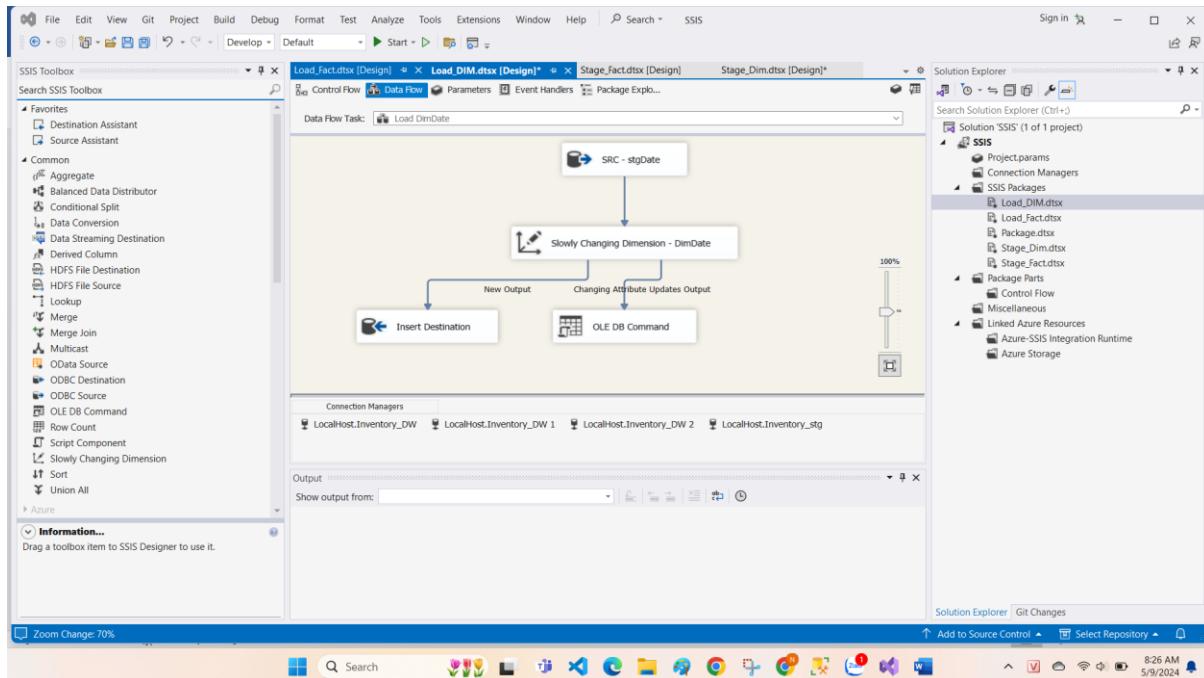
Ta sẽ hội 2 nguồn dữ liệu này lại với nhau để có thể tạo ra stgDate chứa thời gian thông kê kho và thời gian giao dịch



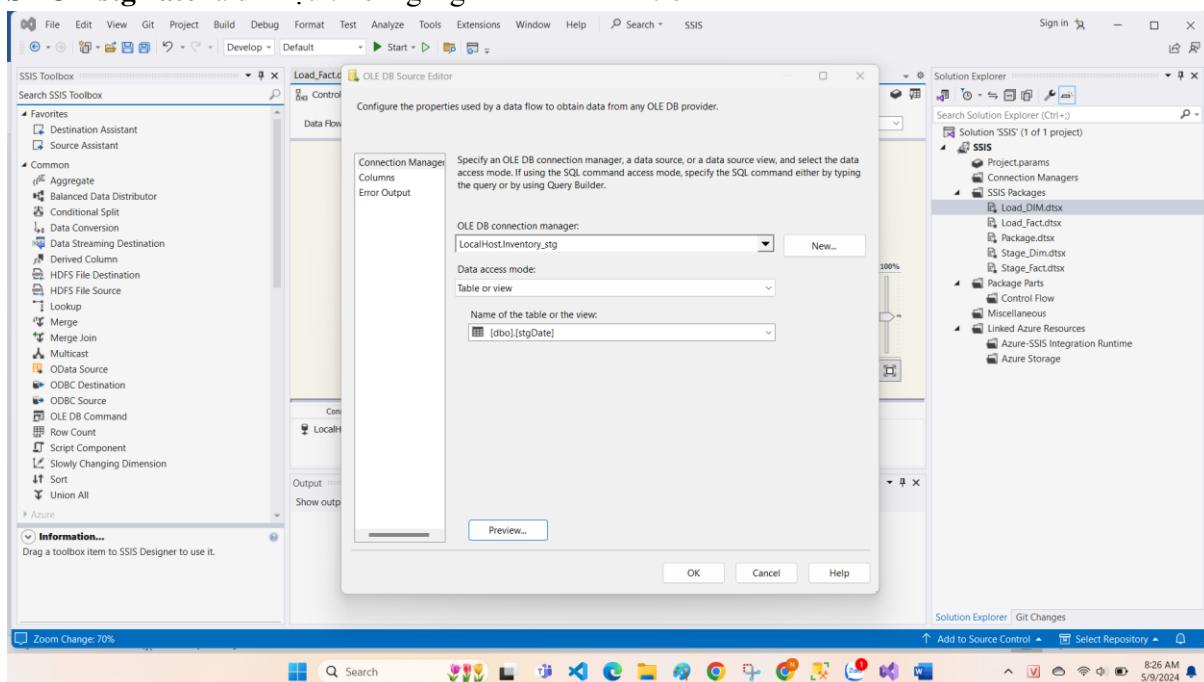
## Khởi chạy



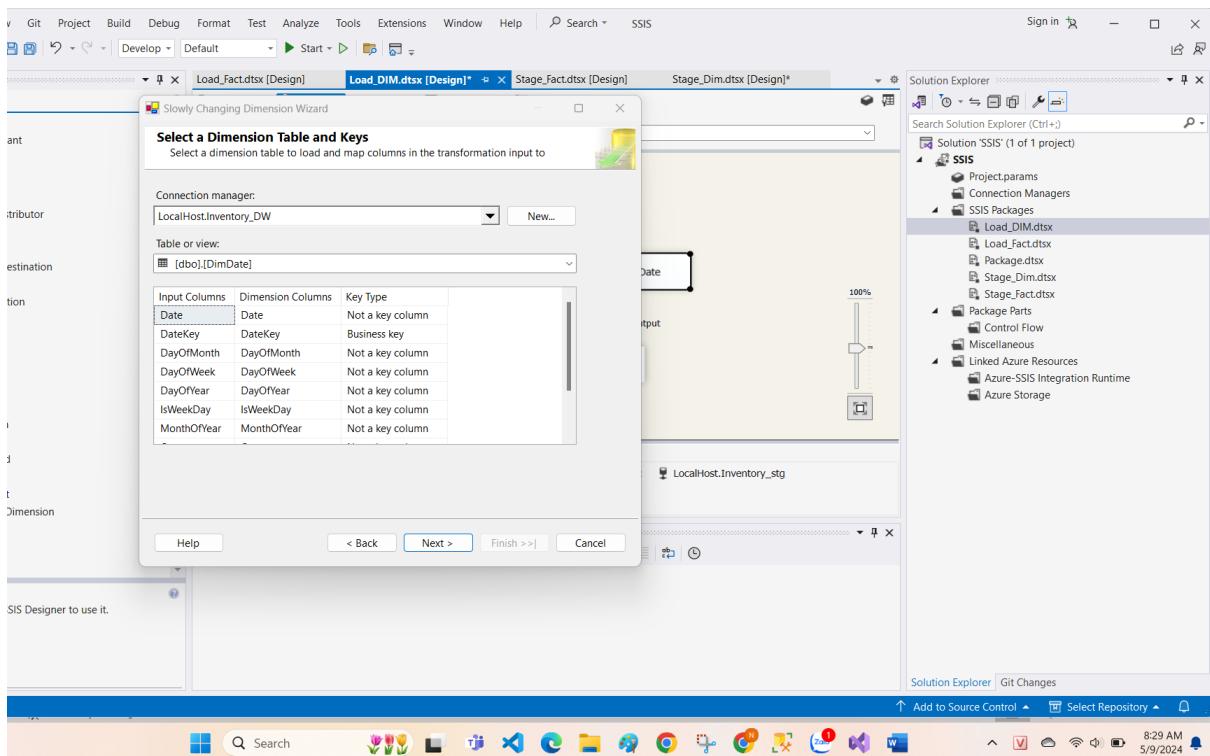
## b) Load dữ liệu từ stgDate vào DimDate



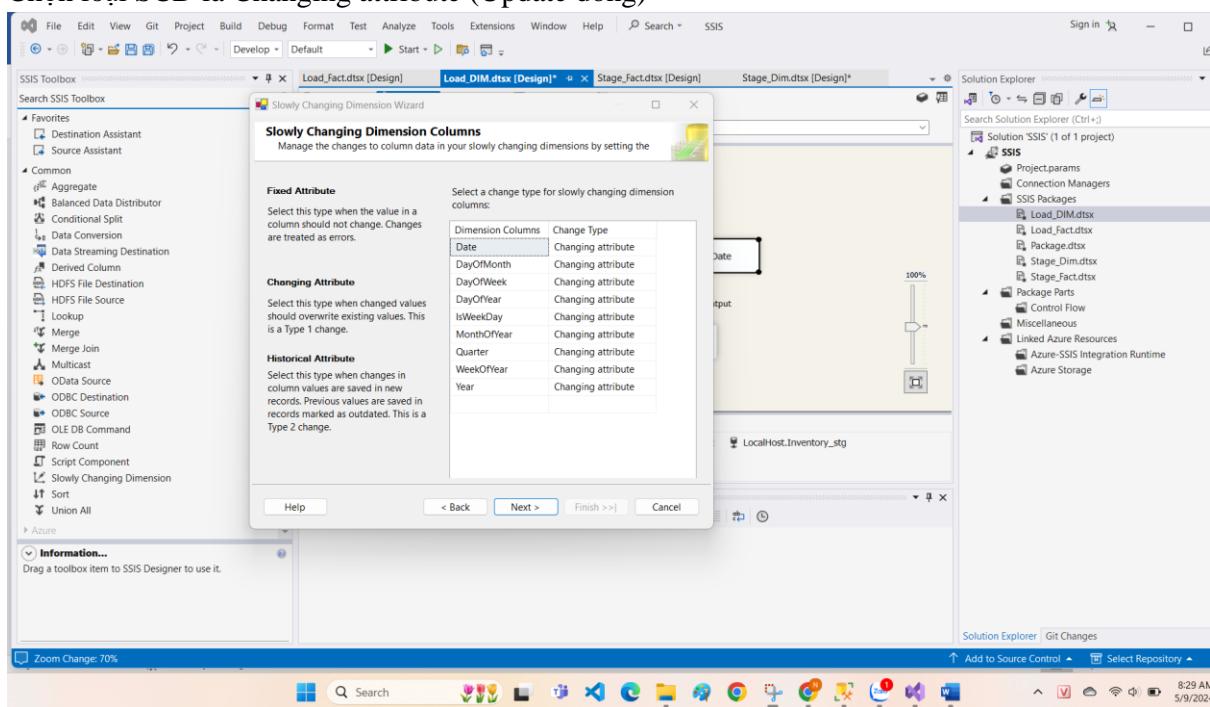
**Src – stgDate** là dữ liệu từ bảng stgDate đã load ở trên



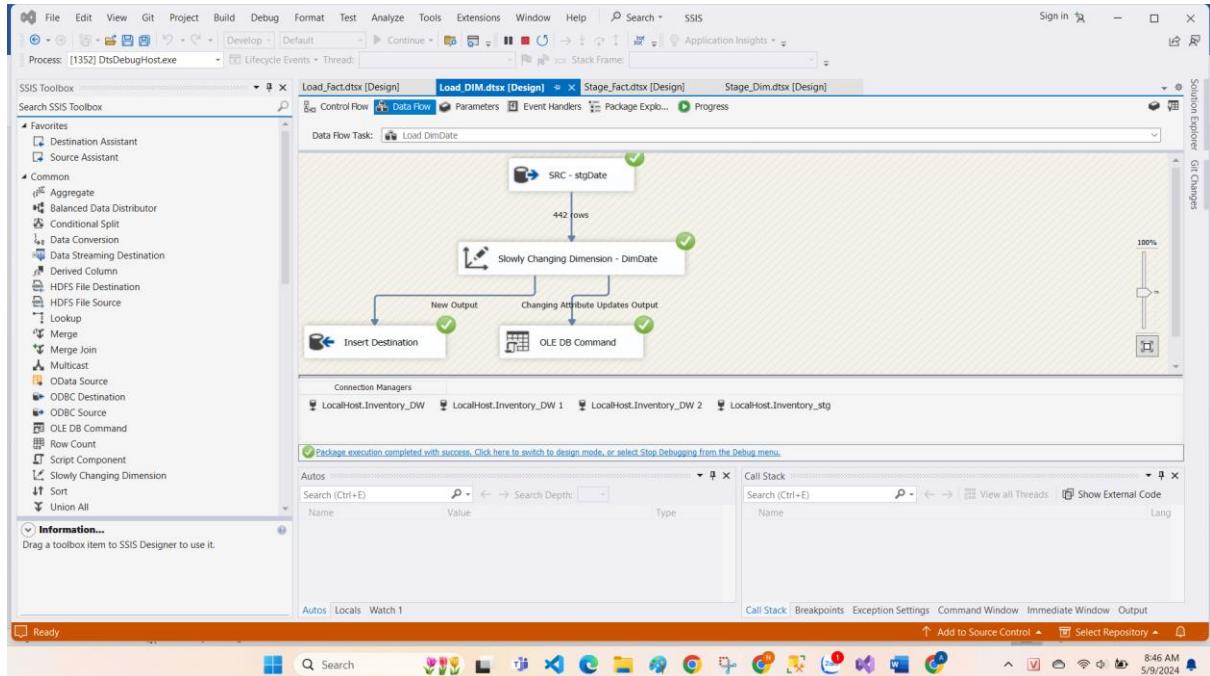
**Slowly Changing Dimension – DimDate** load dữ liệu từ stgDate vào DimDate



## Chọn loại SCD là Changing attribute (Update dòng)

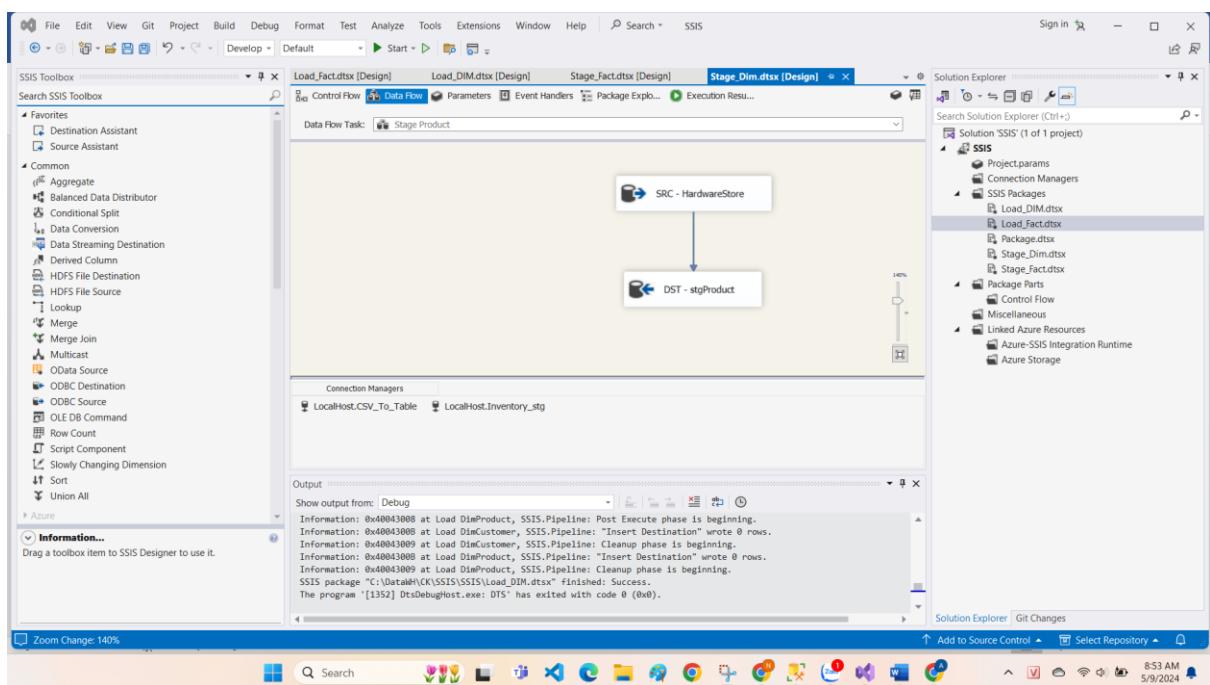


## Khởi chạy

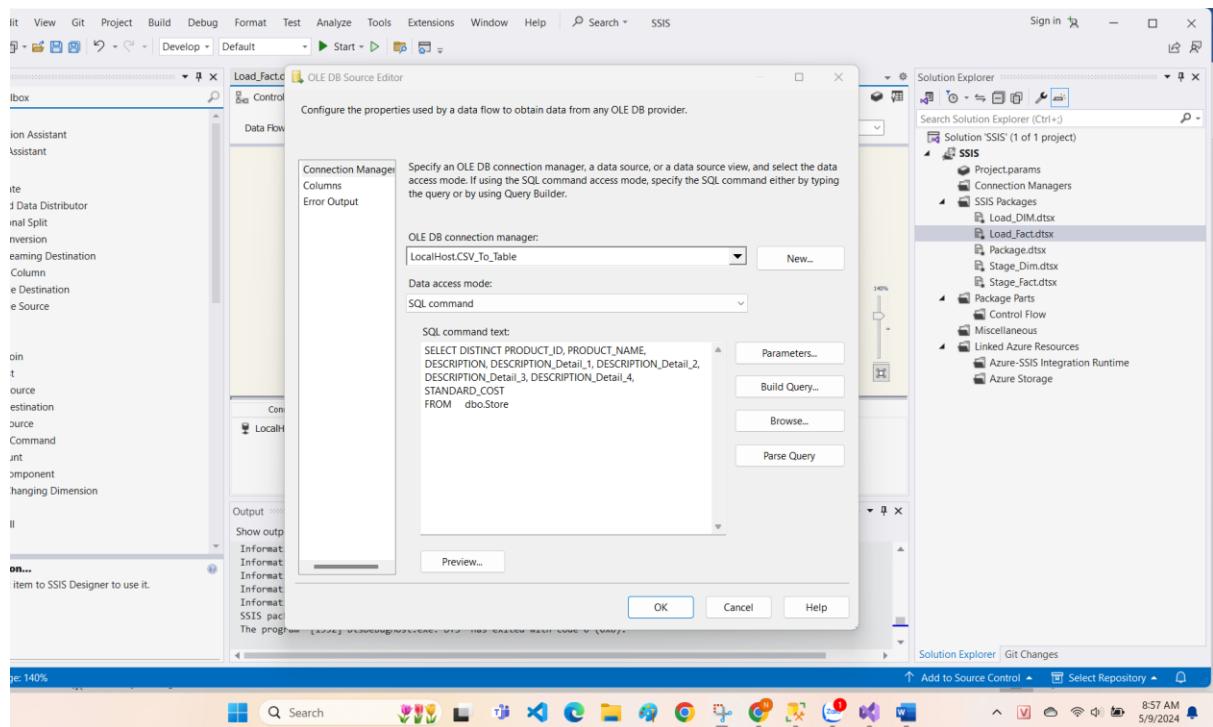


### 3.1.2. Product Dimension

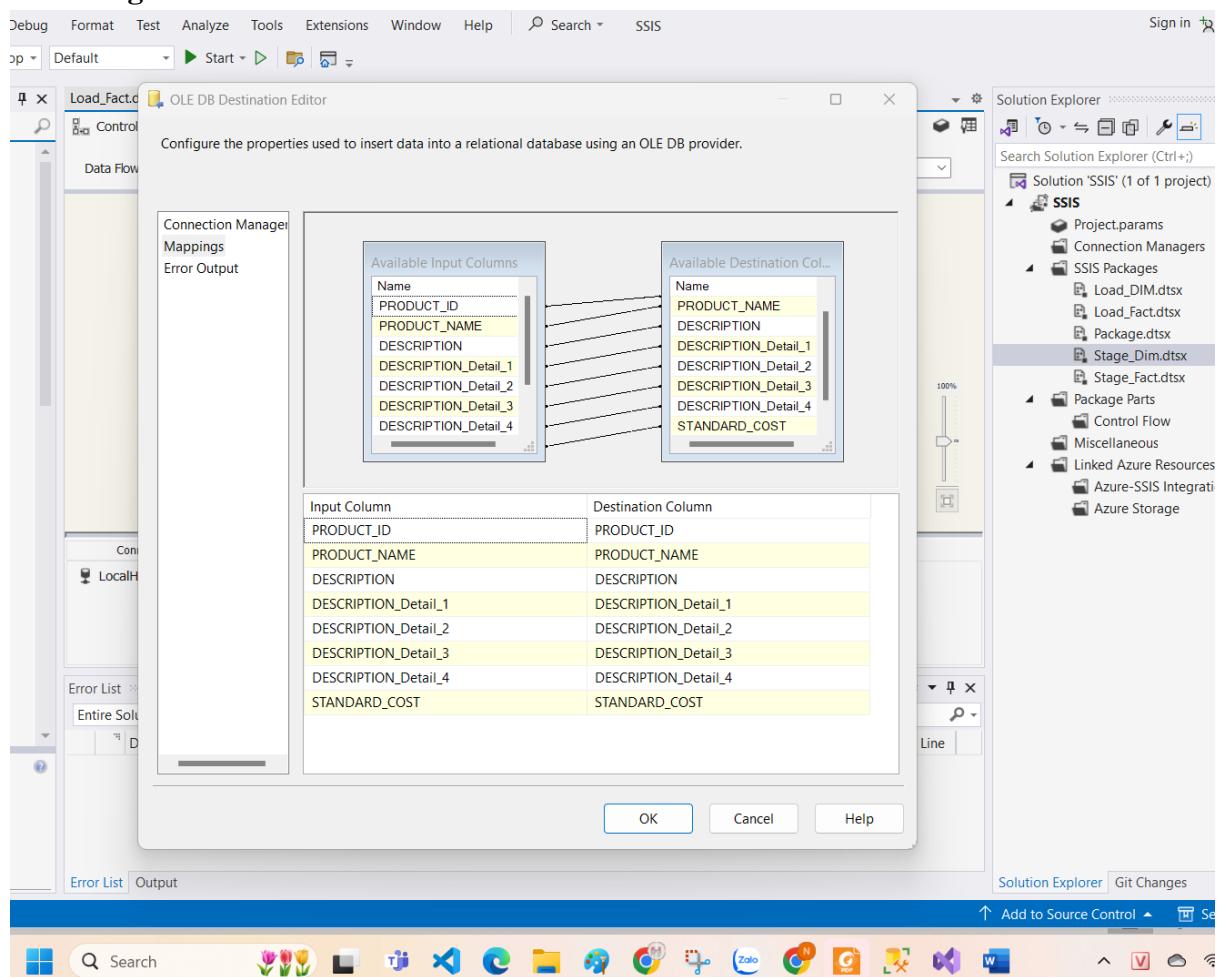
#### a) Load từ nguồn vào bảng stgProduct

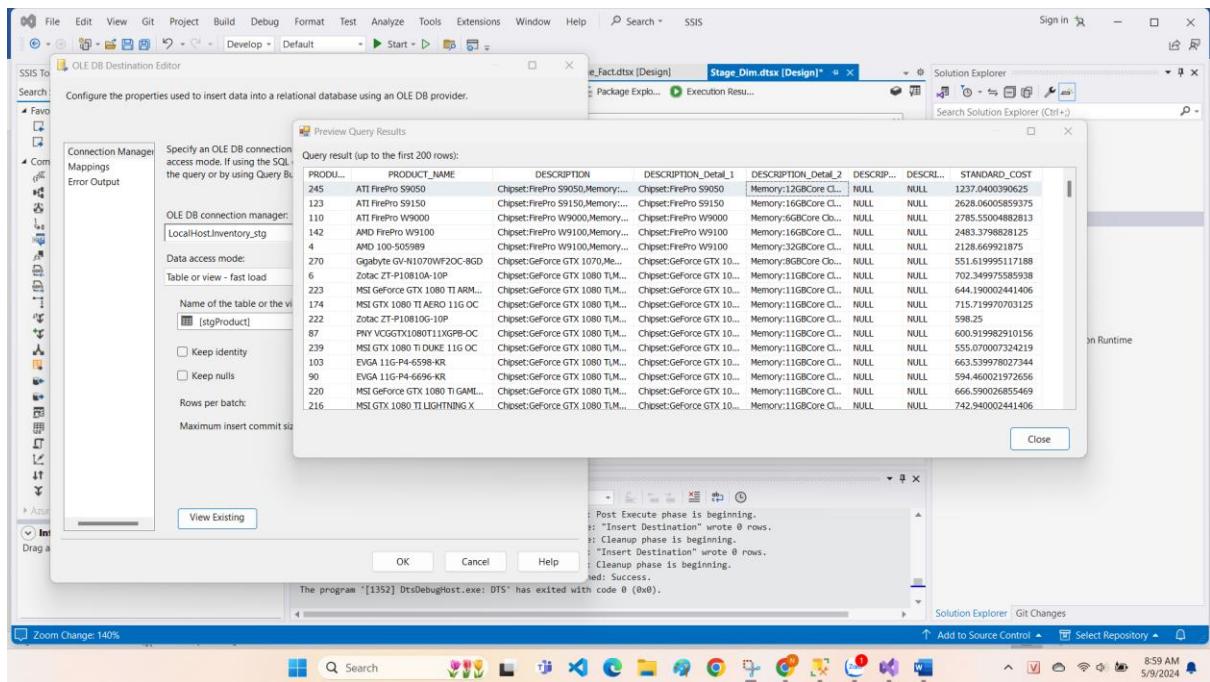


**Src – HardwareStore** chứa dữ liệu các Product từ bảng Store

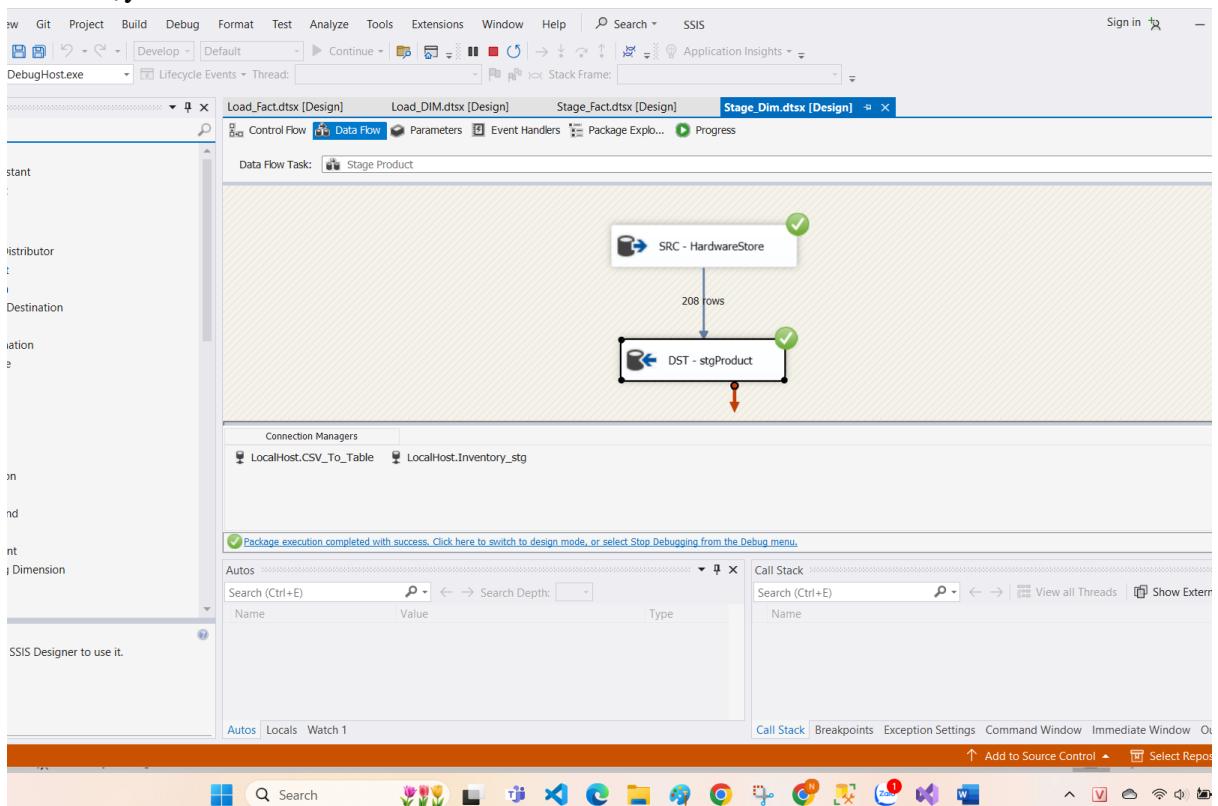


## DST – stgProduct

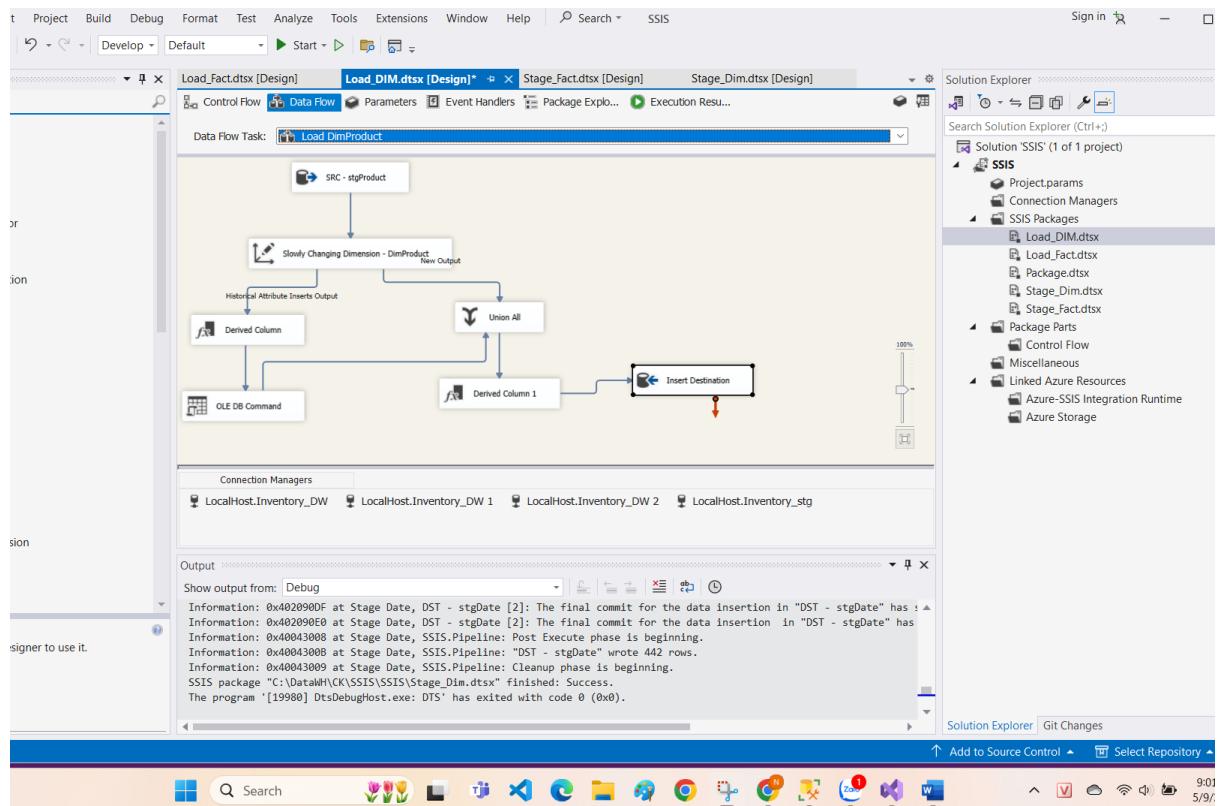




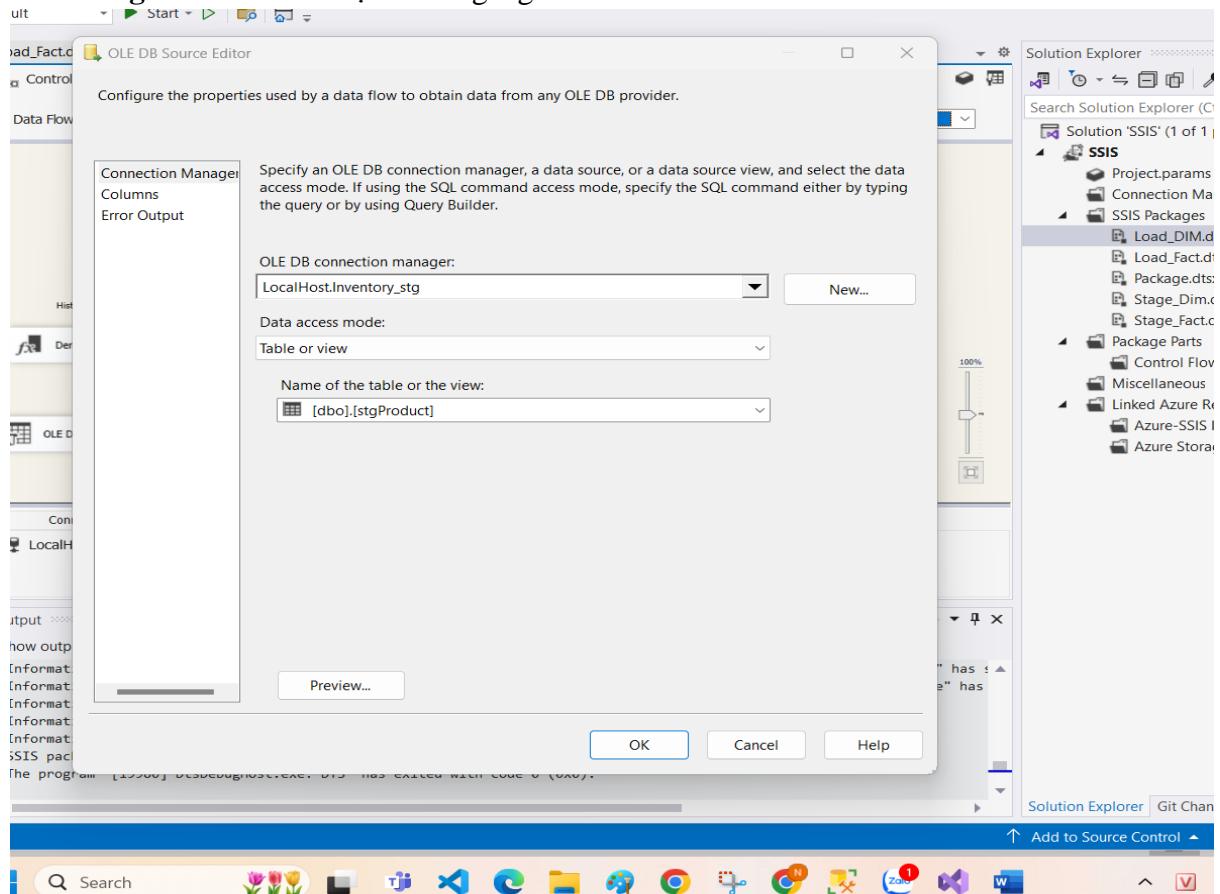
## Khởi chạy



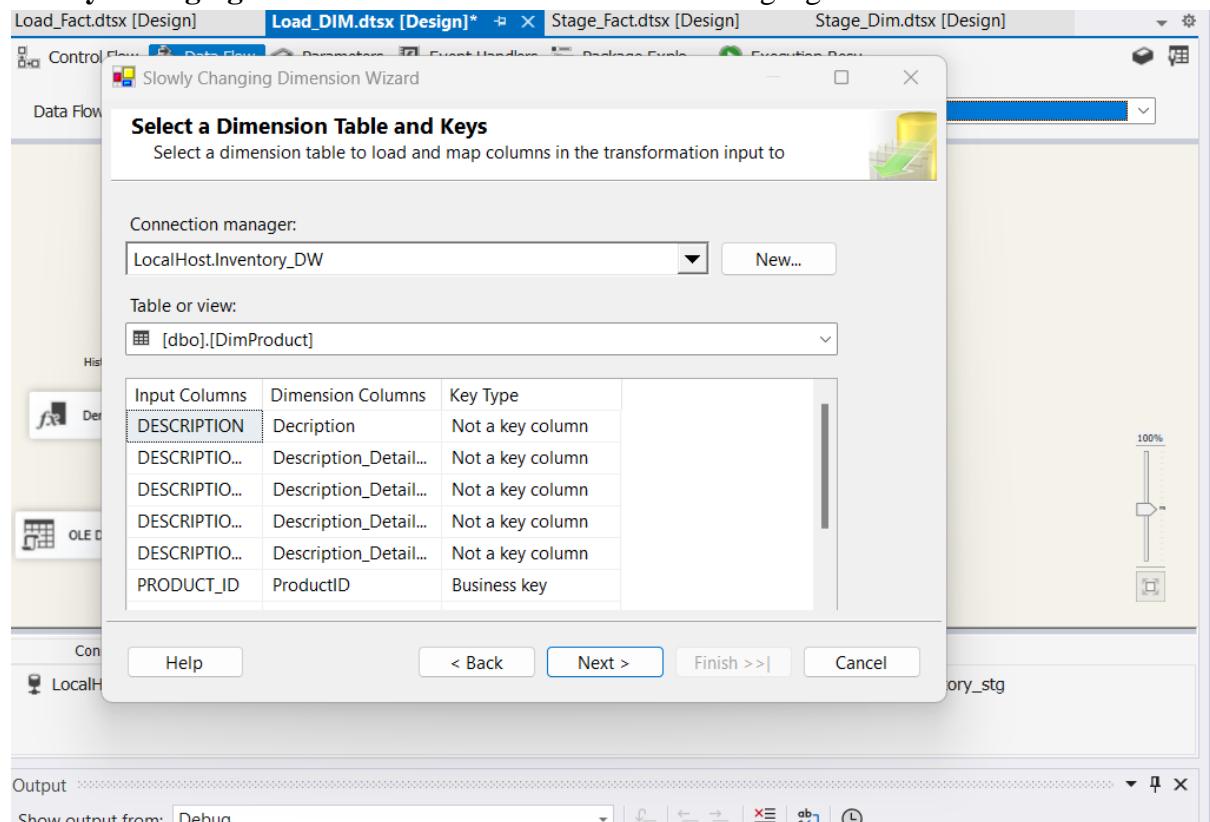
## b) Load dữ liệu từ stgProduct vào DimProduct



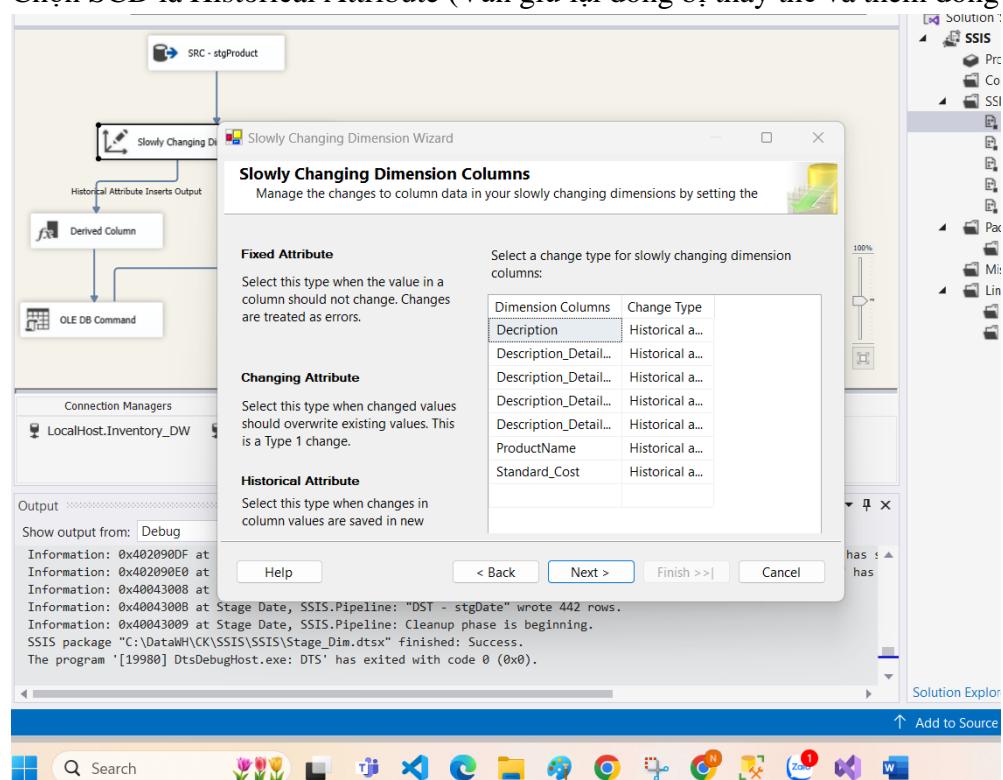
**Src – stgProduct** là dữ liệu từ bảng stgProduct



## Slowly Changing Dimension – DimProduct Load từ bảng stgProduct vào DimProduct



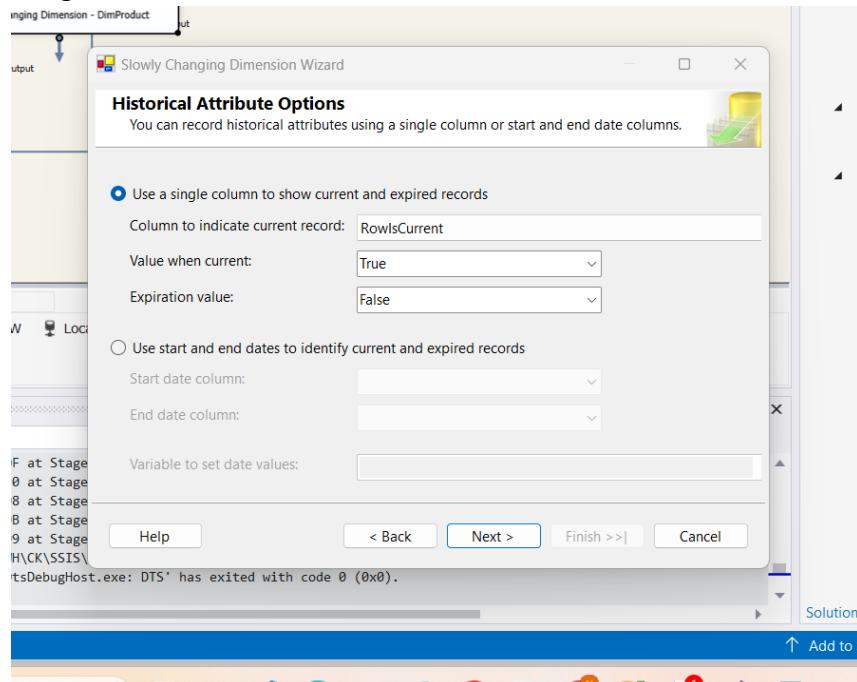
Chọn SCD là Historical Attribute (Vẫn giữ lại dòng bị thay thế và thêm dòng mới)



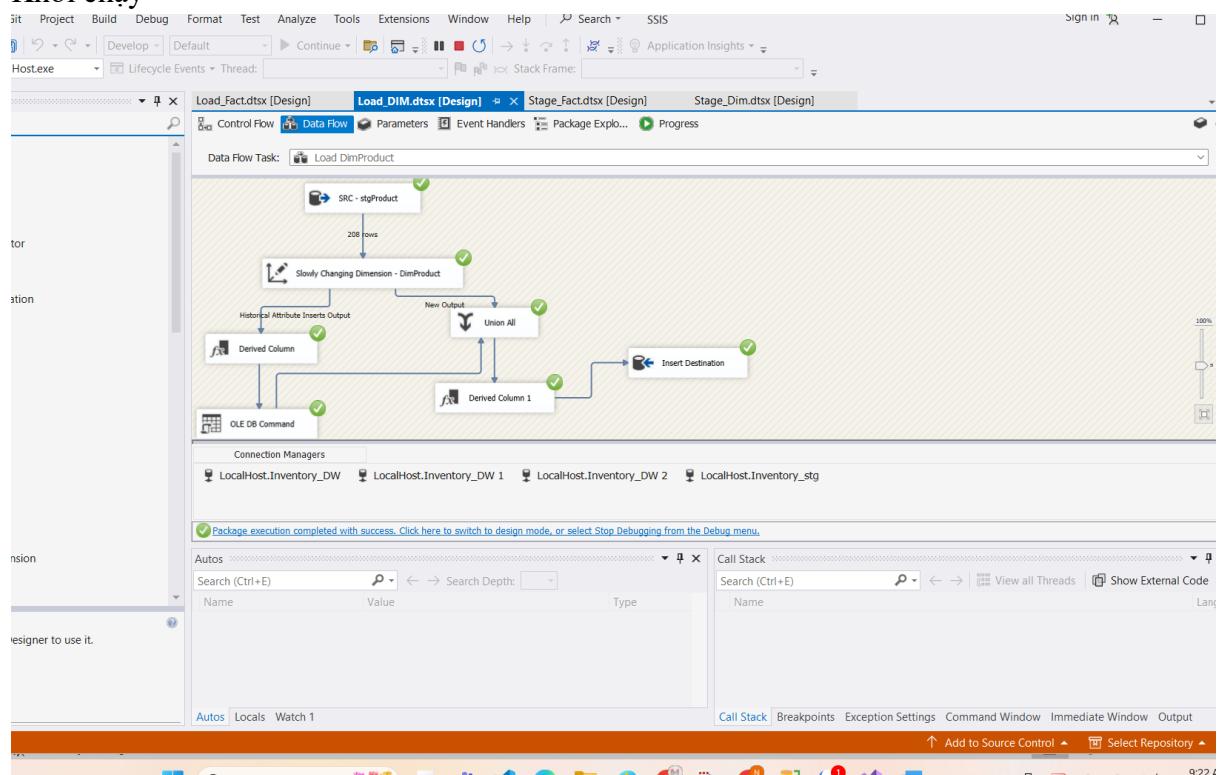
Khi một dòng trong bảng DimProduct cột RowIsCurrent sẽ có giá trị là 1

Ví dụ 1 khách hàng hiện tại đang có hạng tiêu dùng là bạc thì rowiscurrent là 1

Sau đó được nâng hạng lên vàng thì sẽ thêm 1 dòng mới với mã khách hàng đó và hạng tiêu dùng là vàng với rowiscurrent là 1 và sẽ sửa rowiscurrent của dòng cũ thành 0 mà không xóa dòng cũ

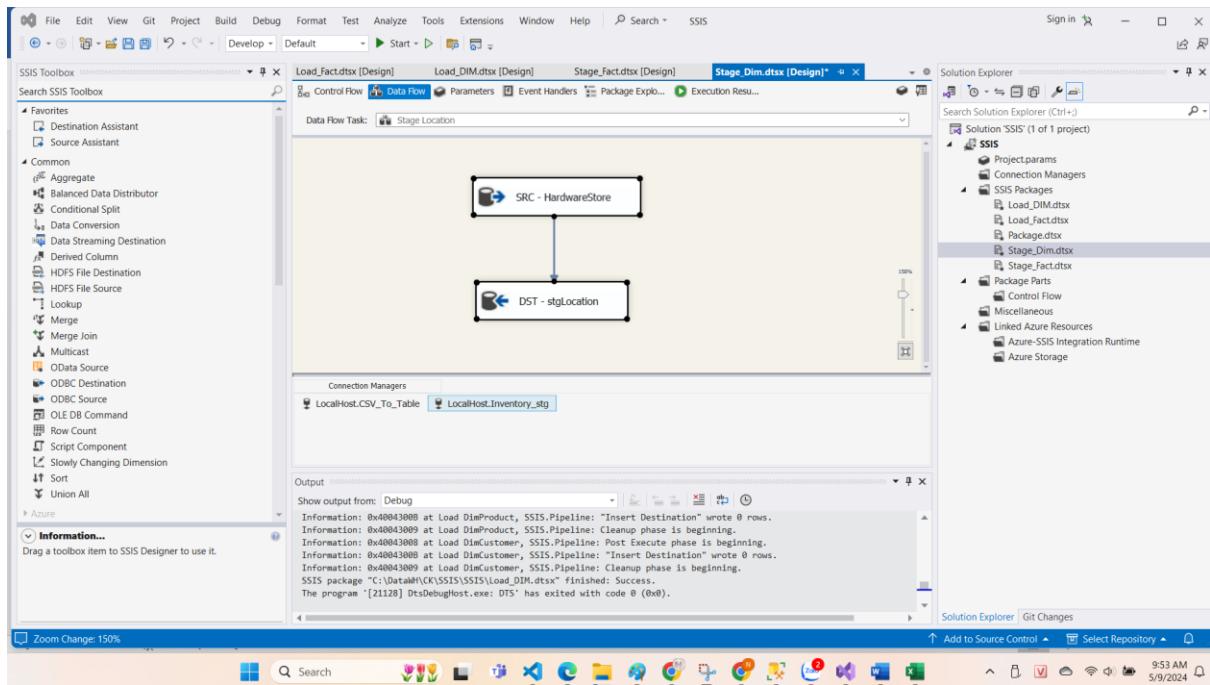


## Khởi chạy

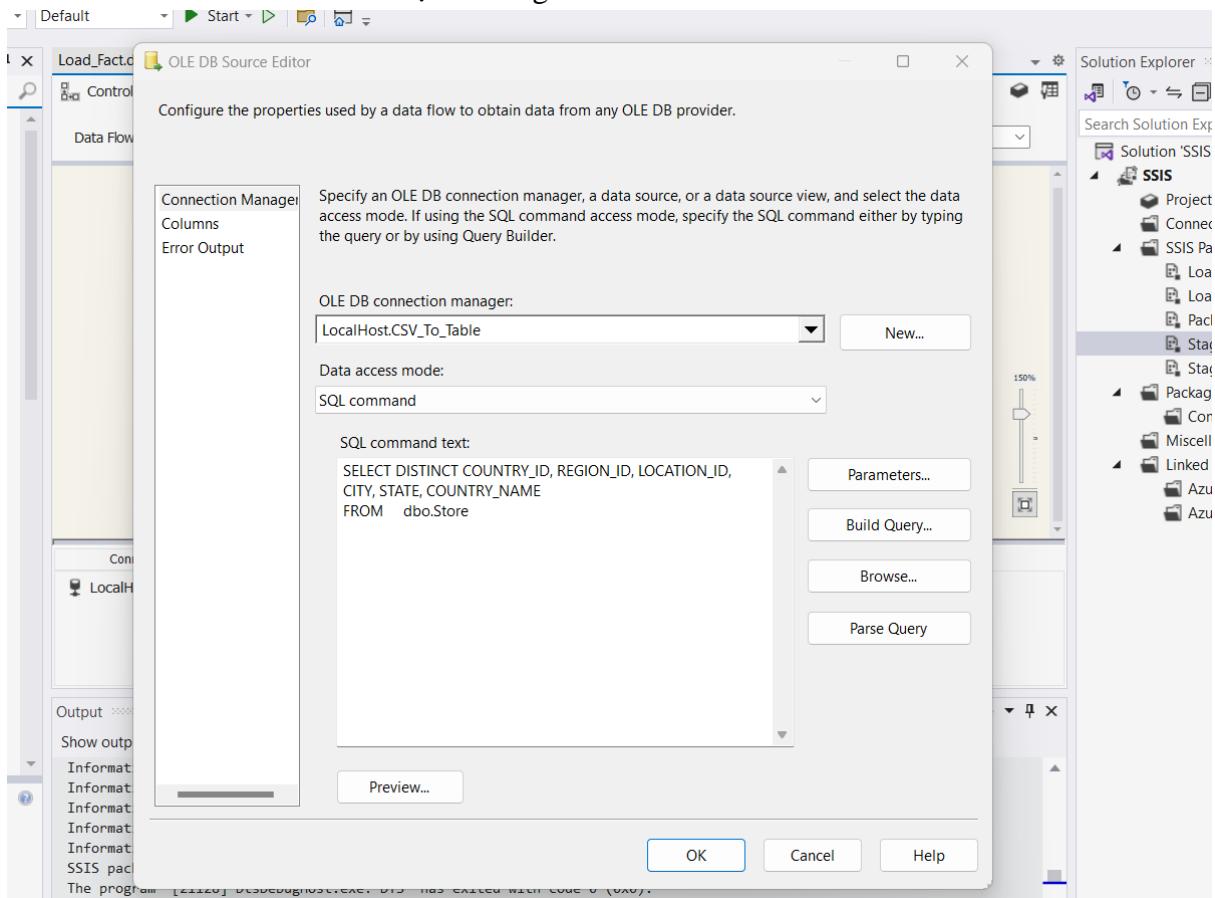


### 3.3.3.Location Dimension

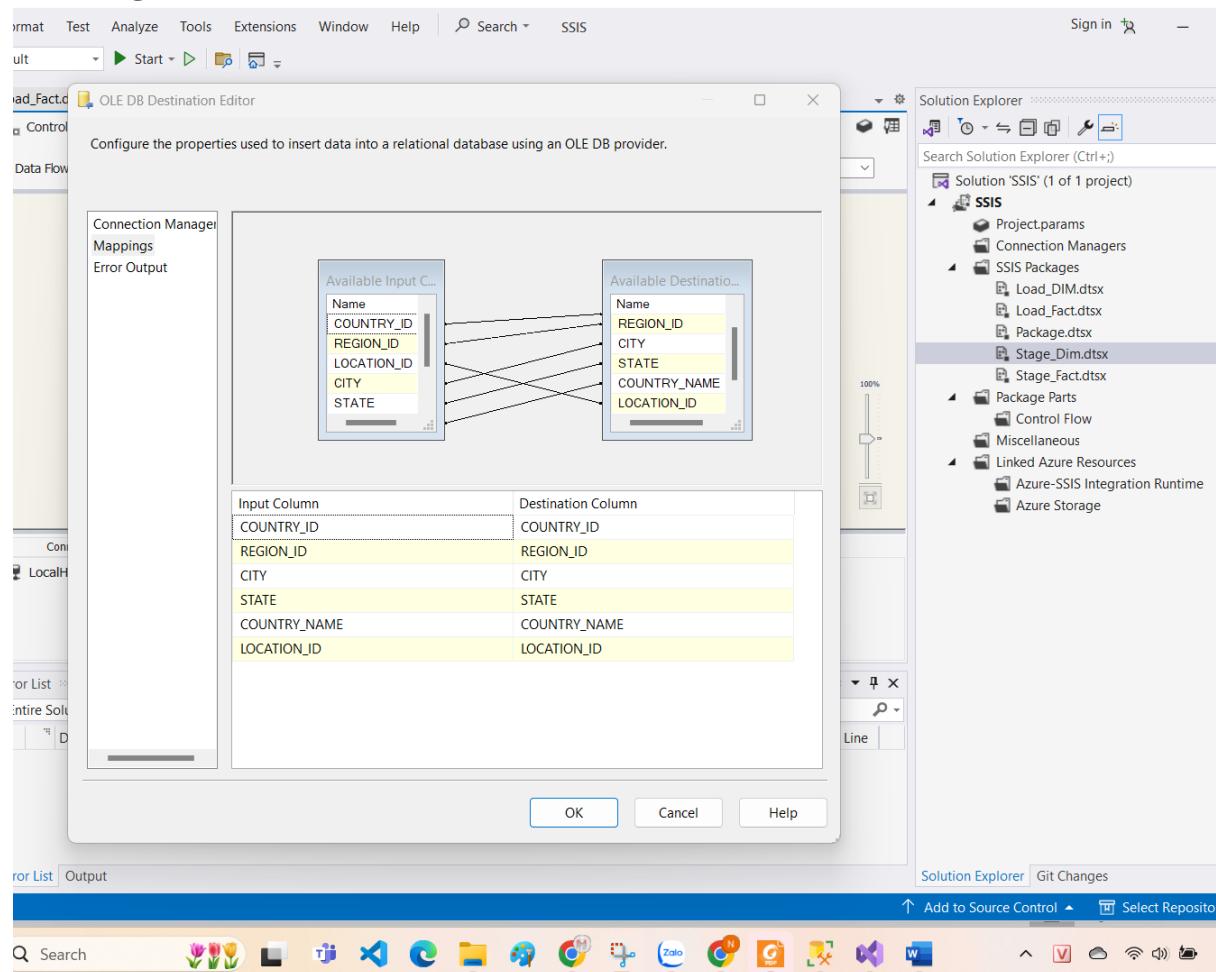
#### a) Load từ nguồn vào bảng stgLocation

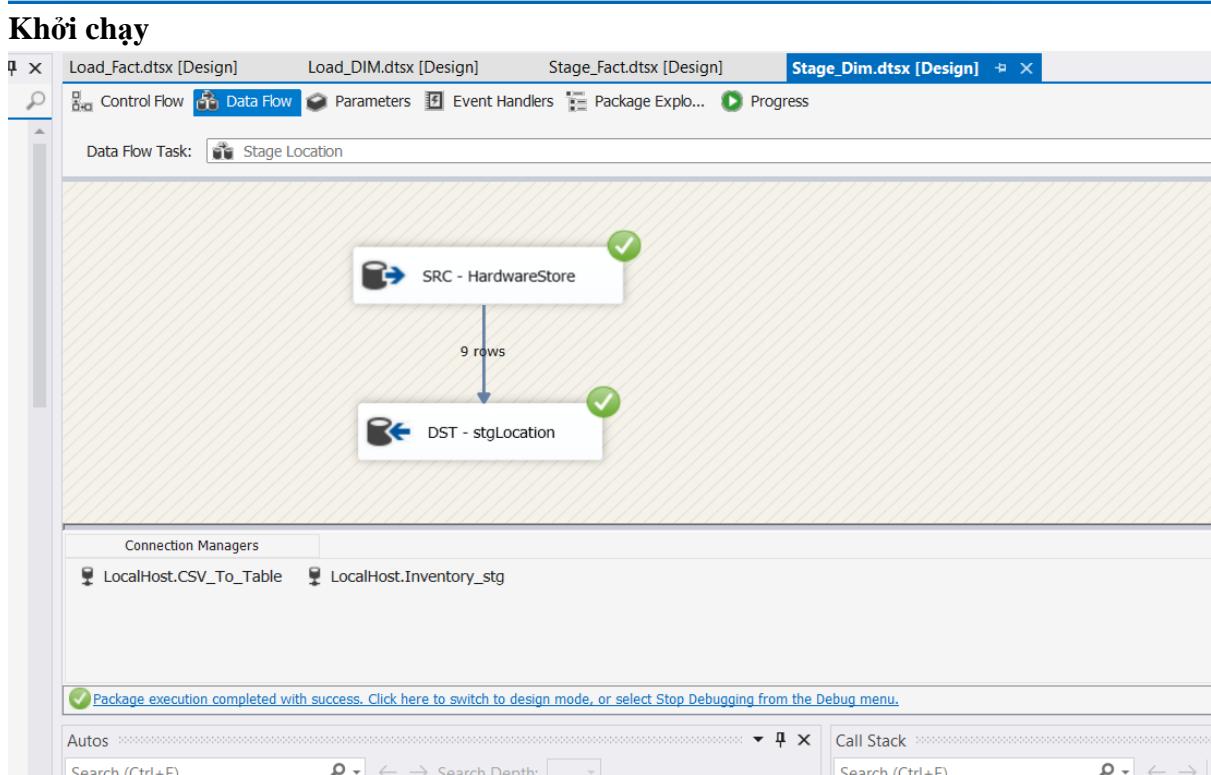
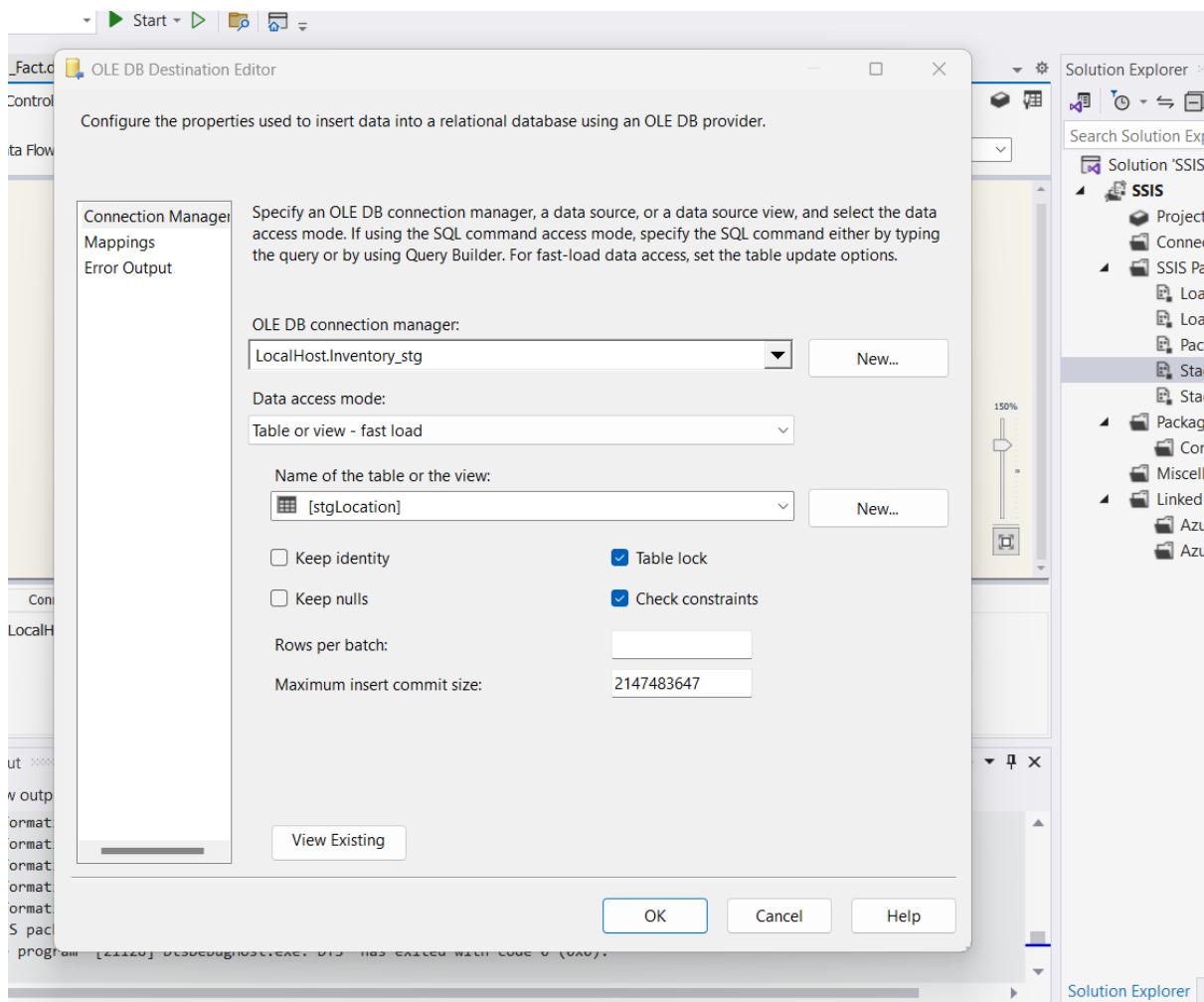


**SRC – HardwareStore** là dữ liệu từ bảng Store

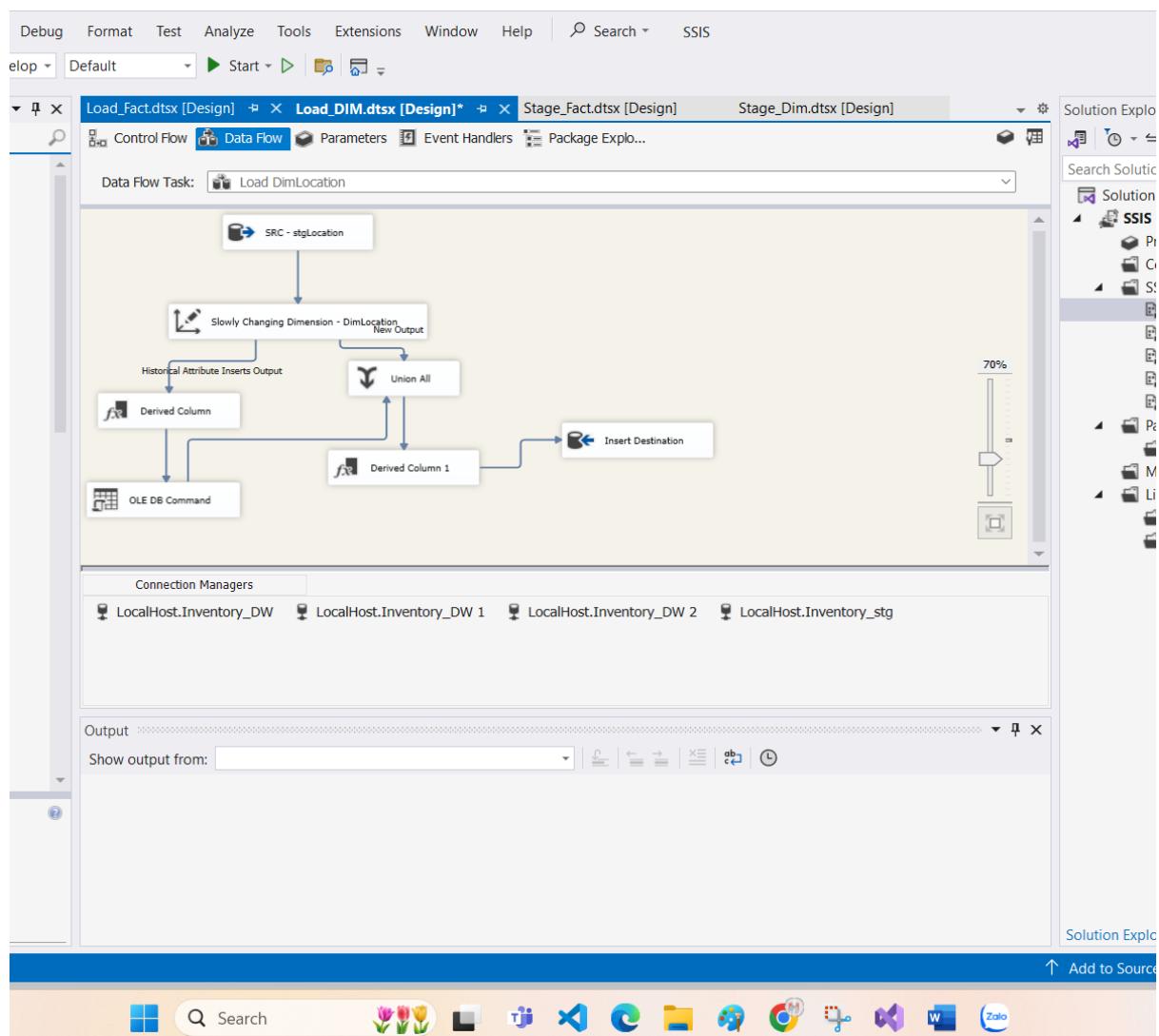


## DST – stgLocation

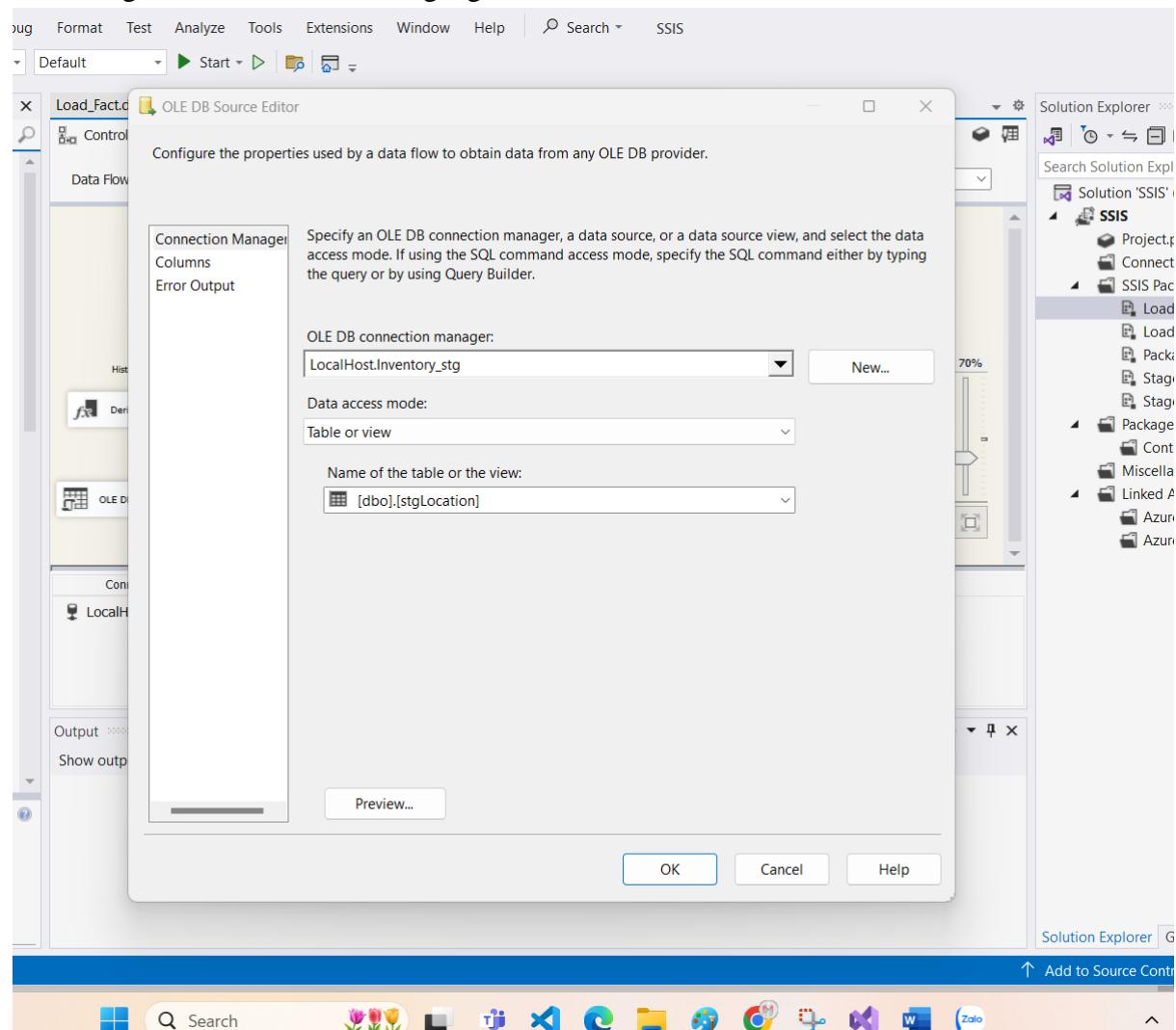




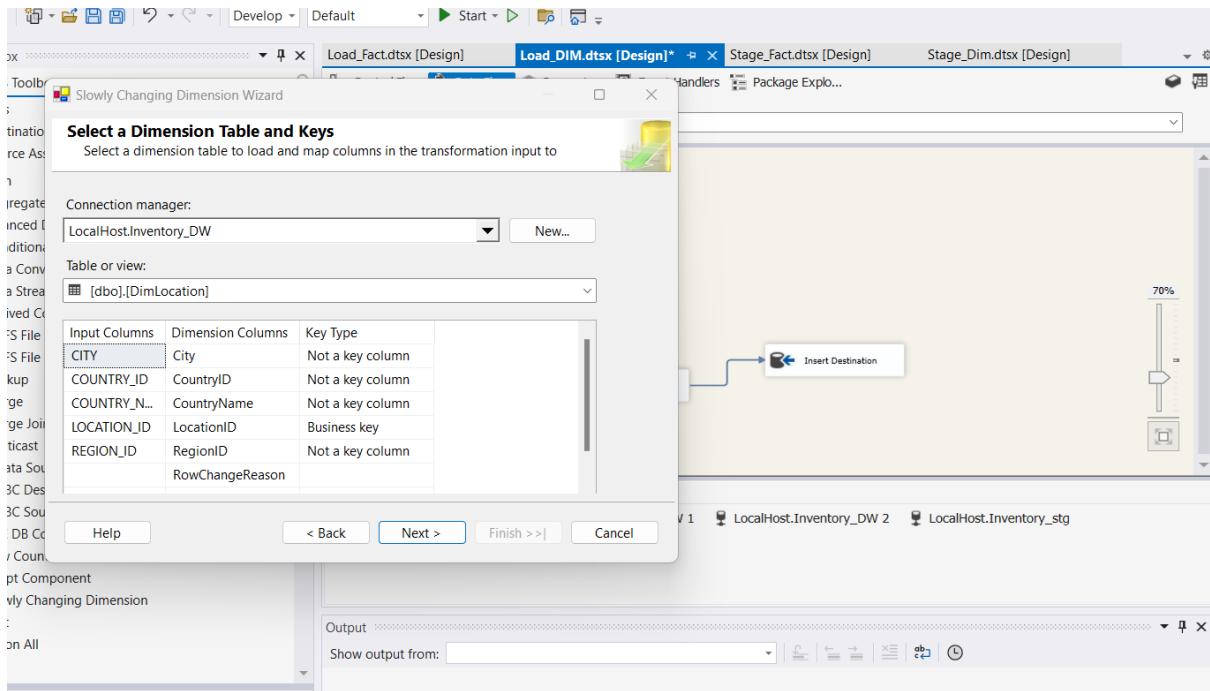
## b) Load dữ liệu từ stgLocation vào bảng DimLocation



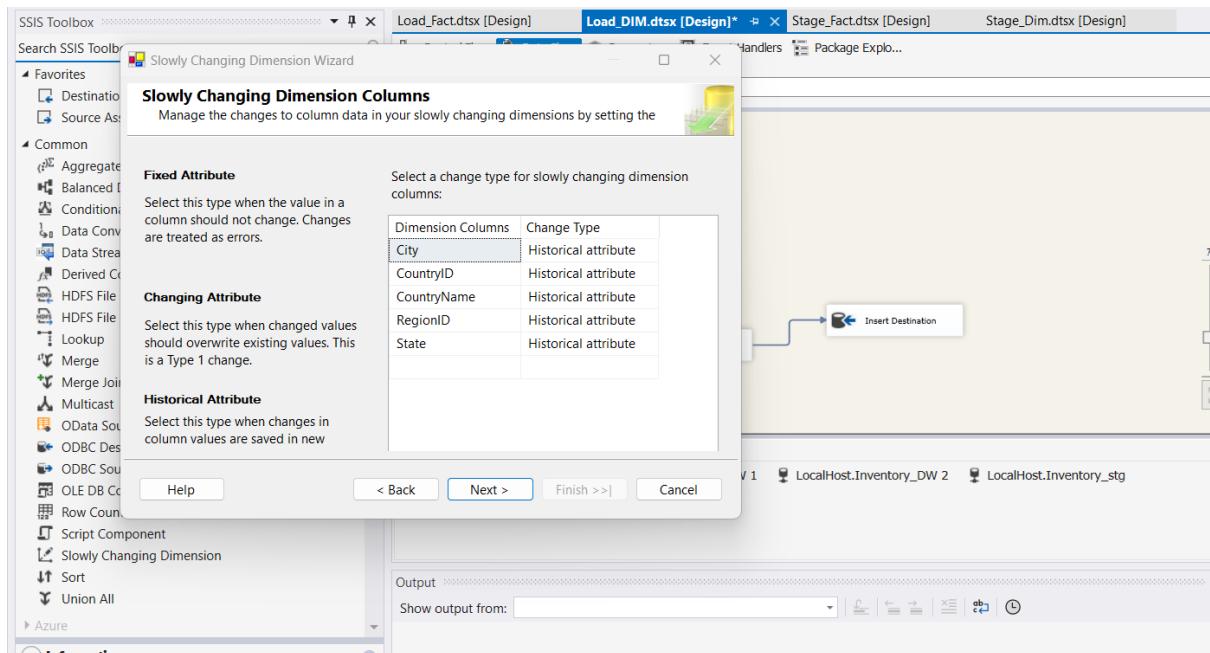
SRC – stgLocation là dữ liệu bảng stgLocation



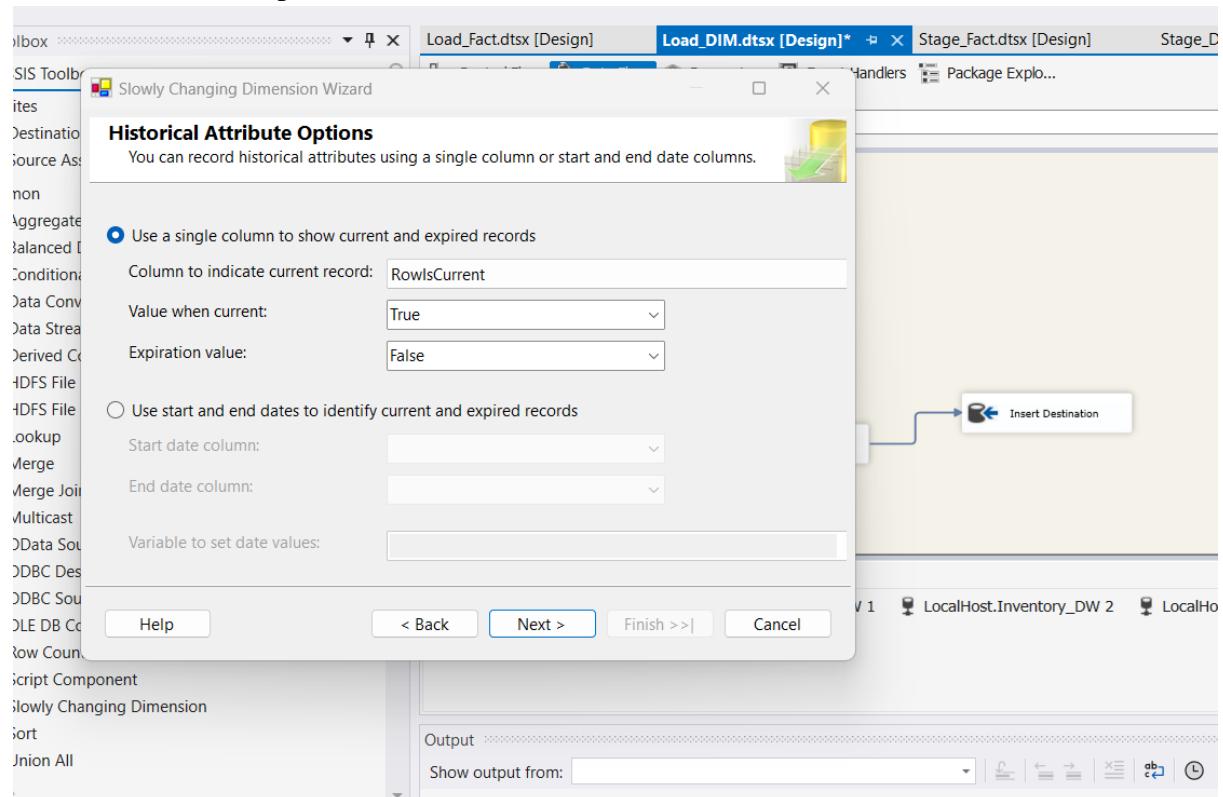
Slowly Changing Dimension – DimLocation



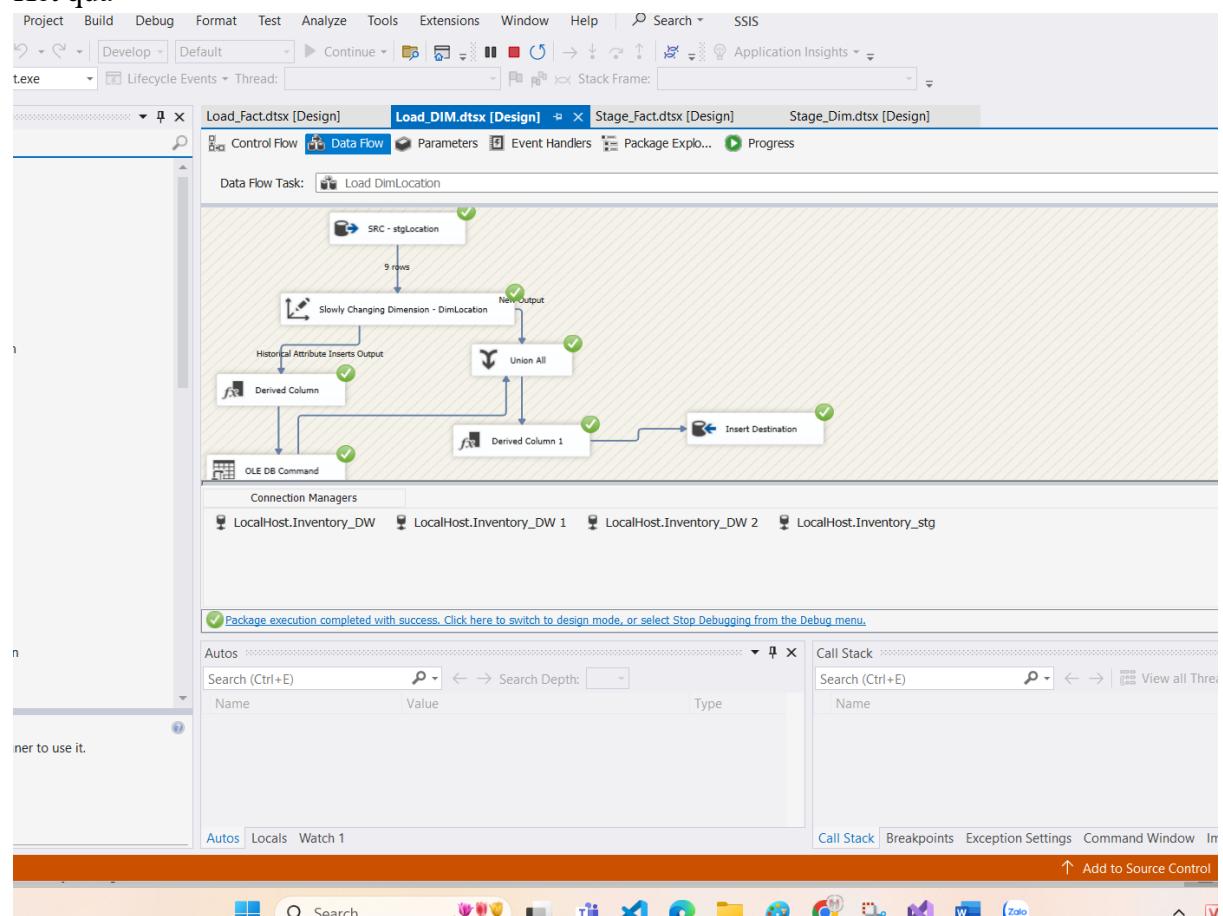
## SCD chọn Historical Attribute



## Historical attribute options chon RowIsCurrent

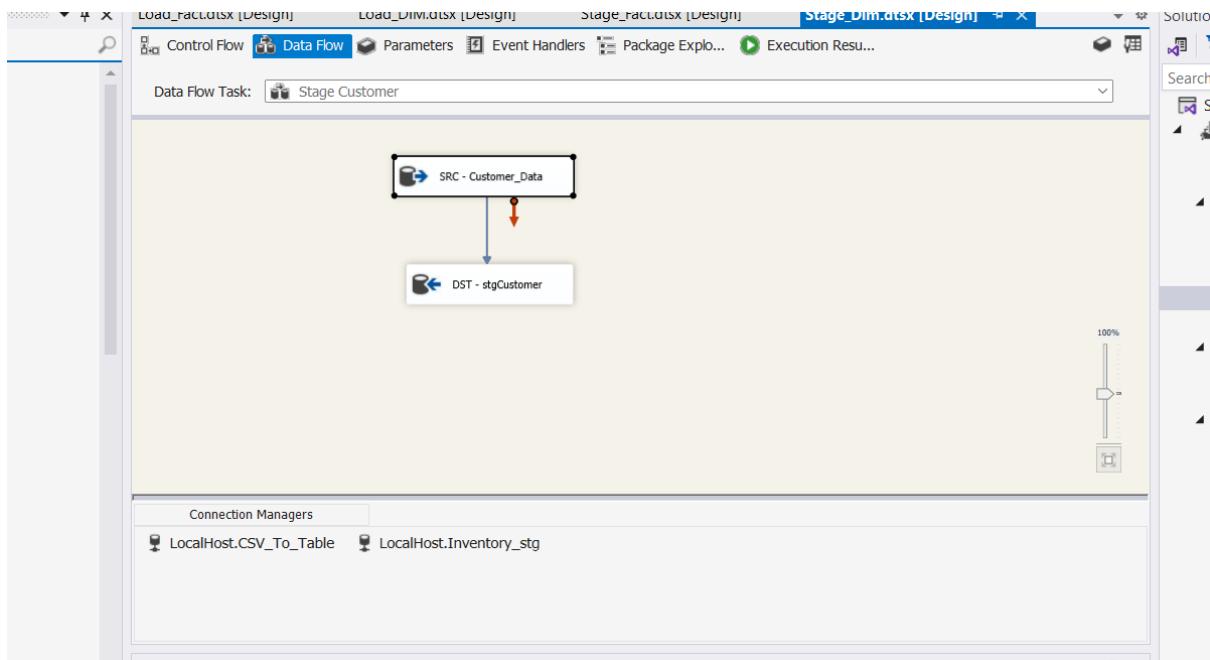


## Kết quả

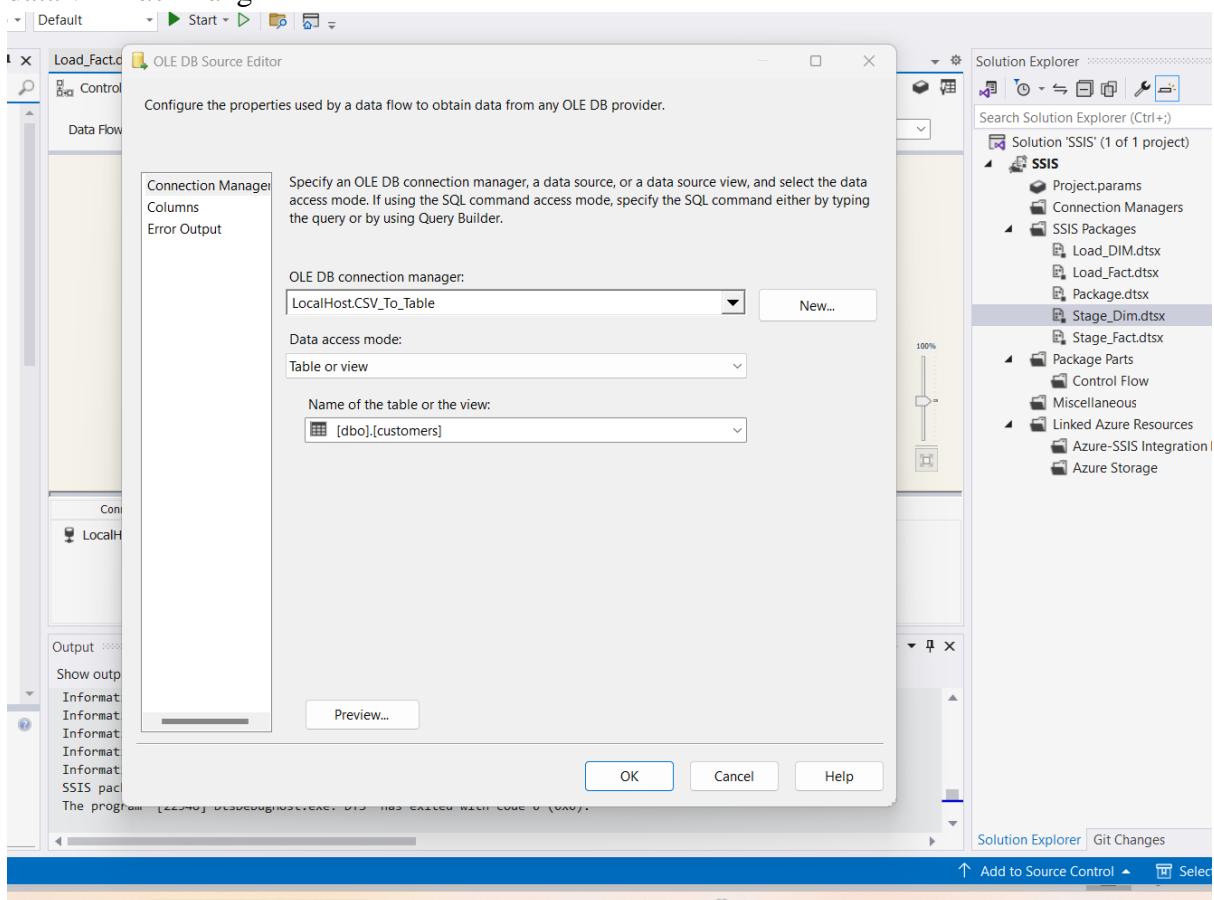


### 3.3.4.Customer Dimension

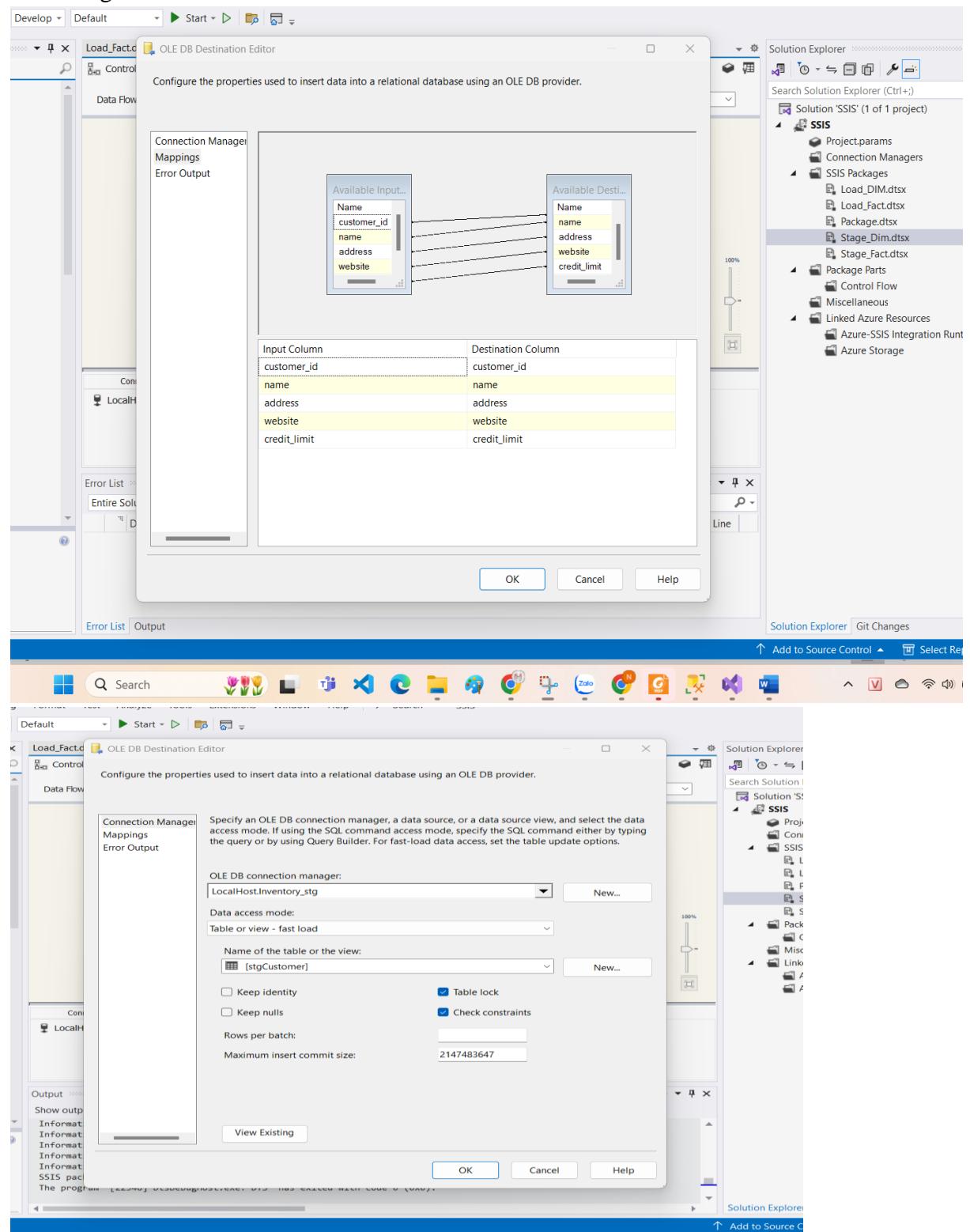
#### a) Load dữ liệu từ nguồn vào stgCustomer



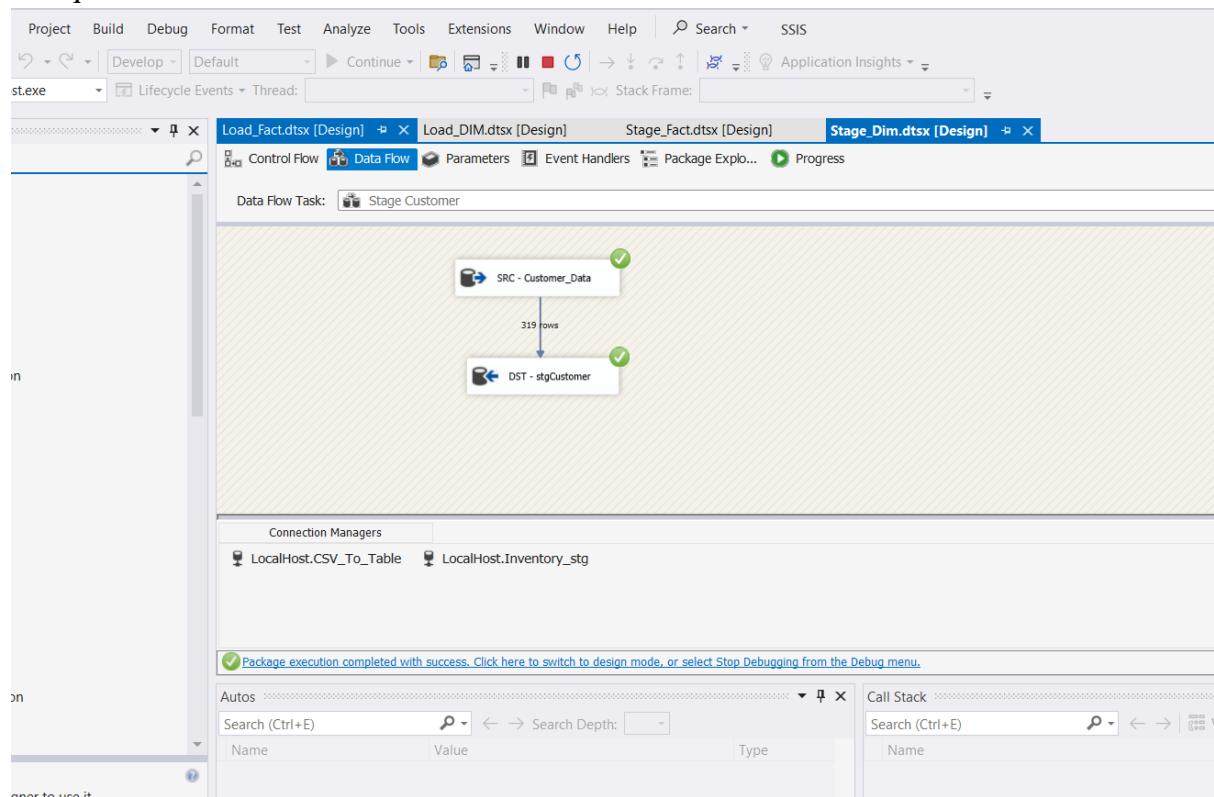
**Src – Customer\_Data** là dữ liệu từ bảng customer trong database CSV\_To\_Table chứa data về khách hàng



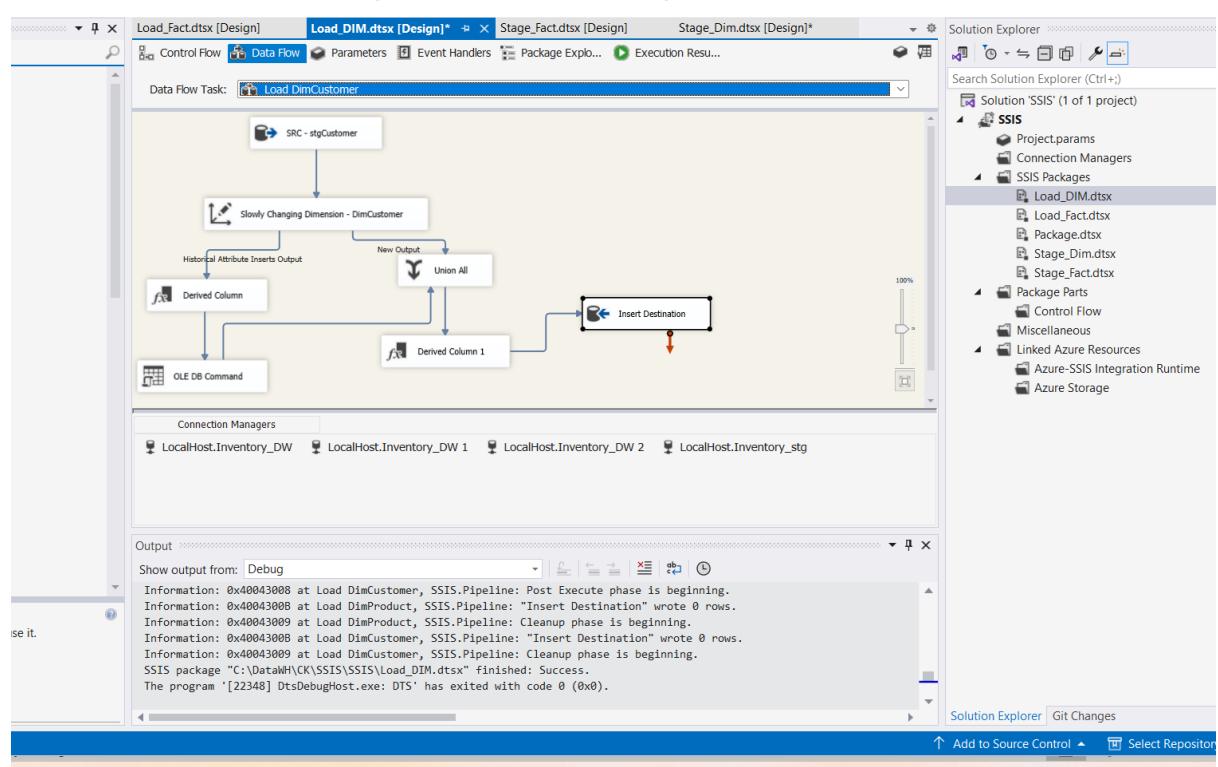
## DST – stgCustomer



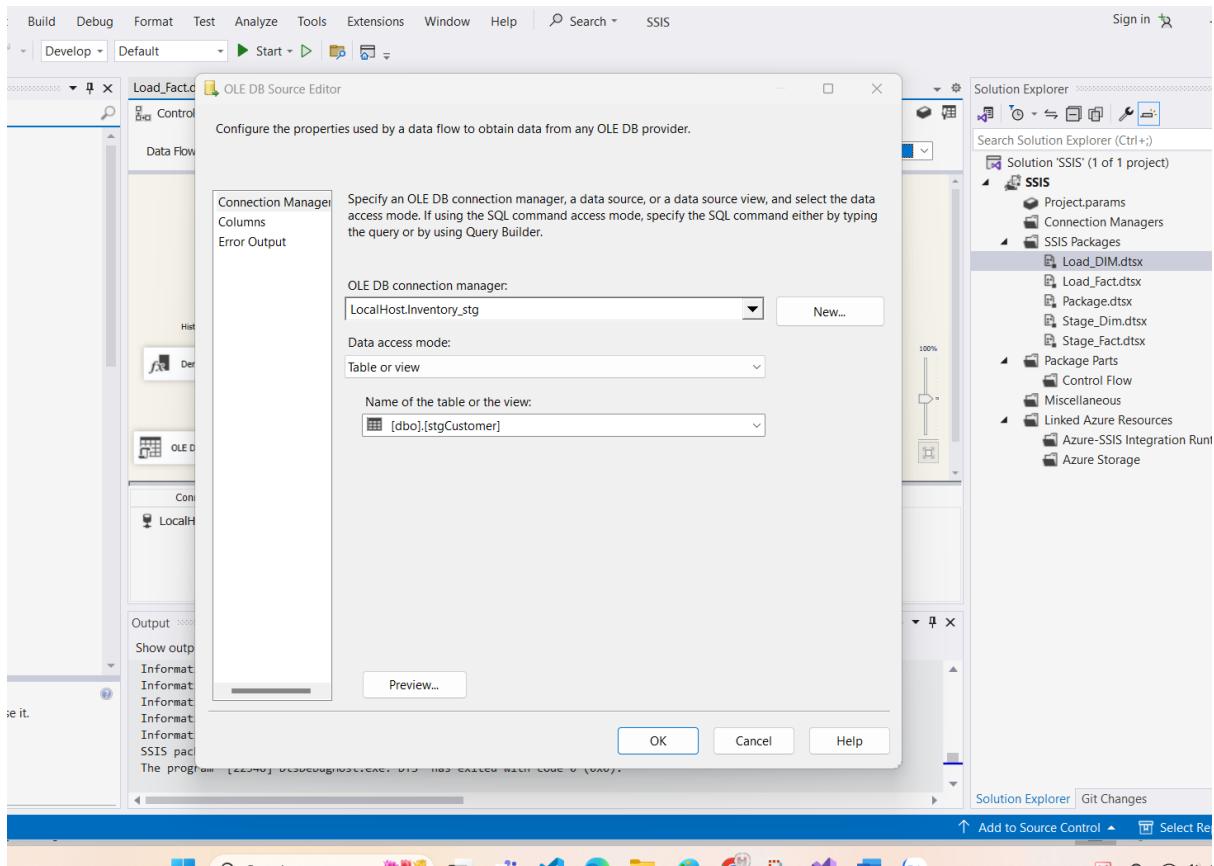
## Kết quả



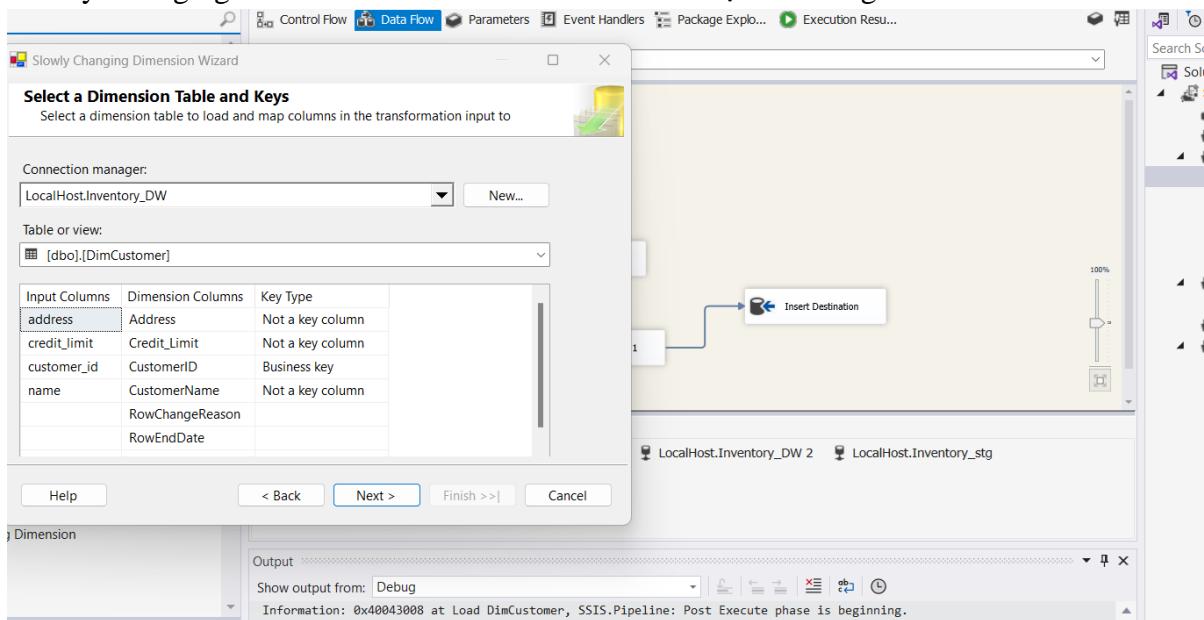
## b) Load dữ liệu từ stgCustomer vào bảng DimCustomer



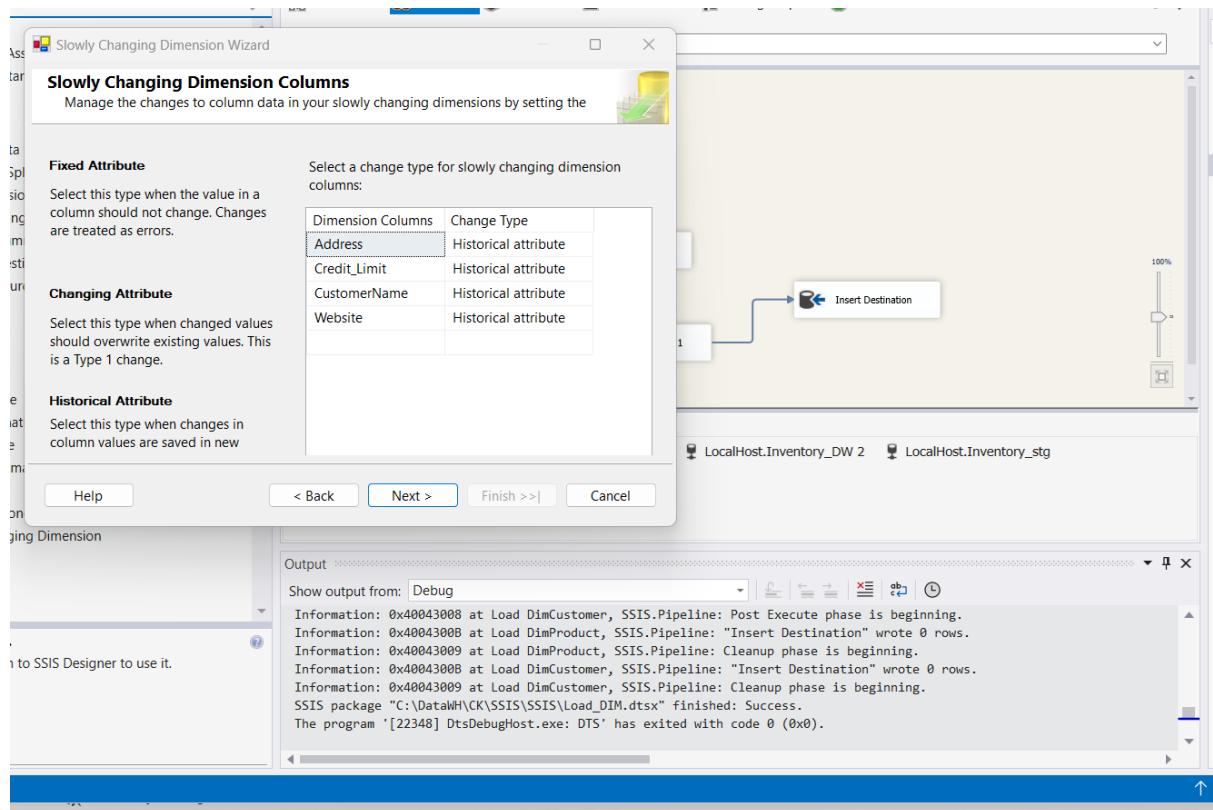
SRC - stgCustomer chứa dữ liệu bảng stgCustomer



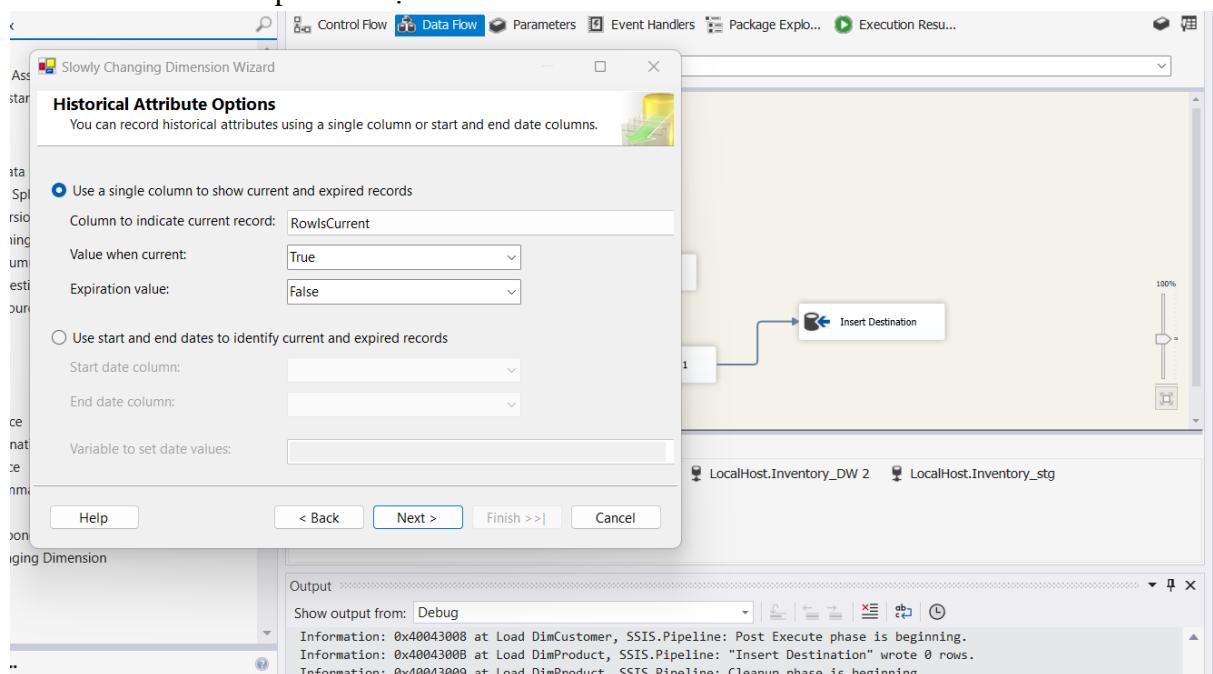
## Slowly Changing Dimension – DimCustomer load dữ liệu vào bảng DimCustomer



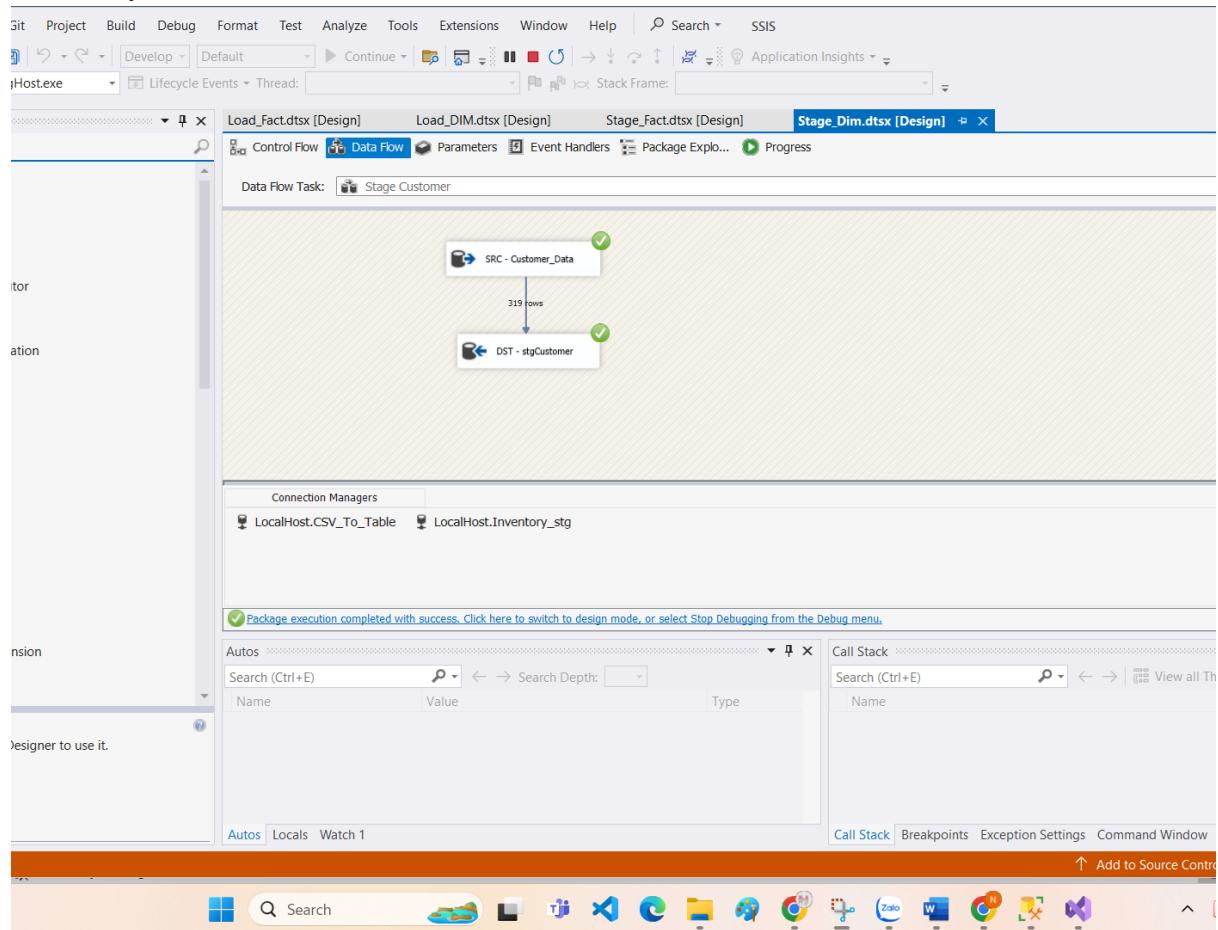
## SCD chọn Historical attribute



## Historical attribute options chọn RowIsCurrent

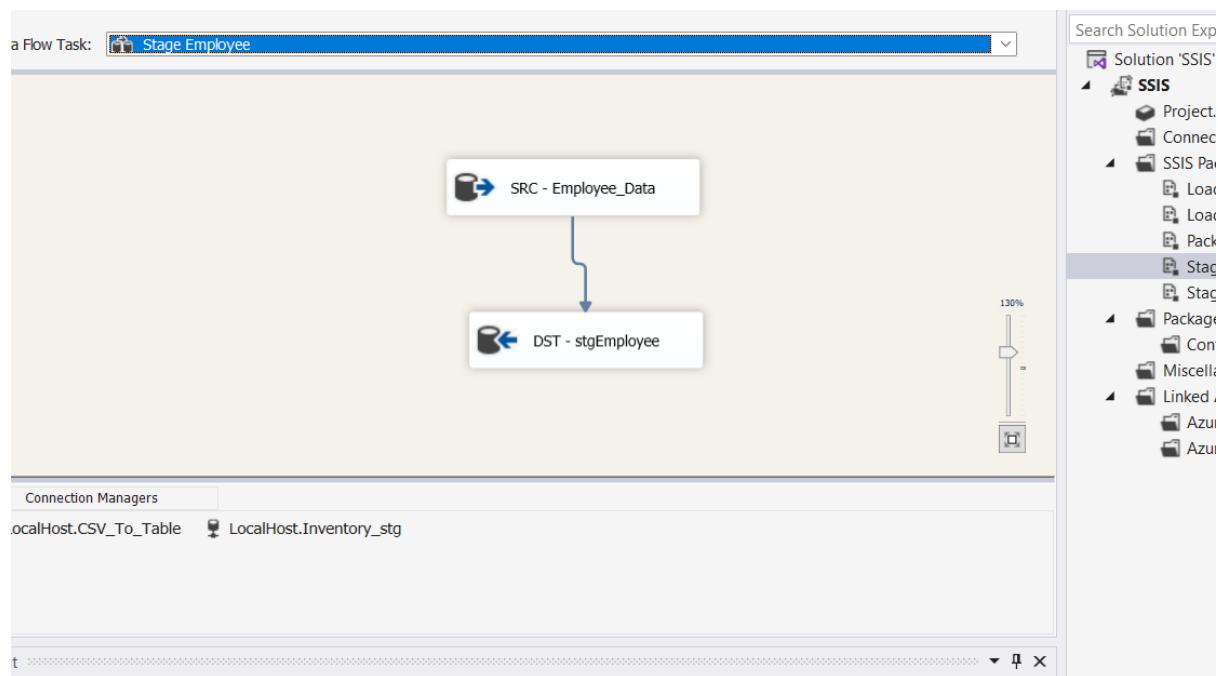


## Khởi chạy

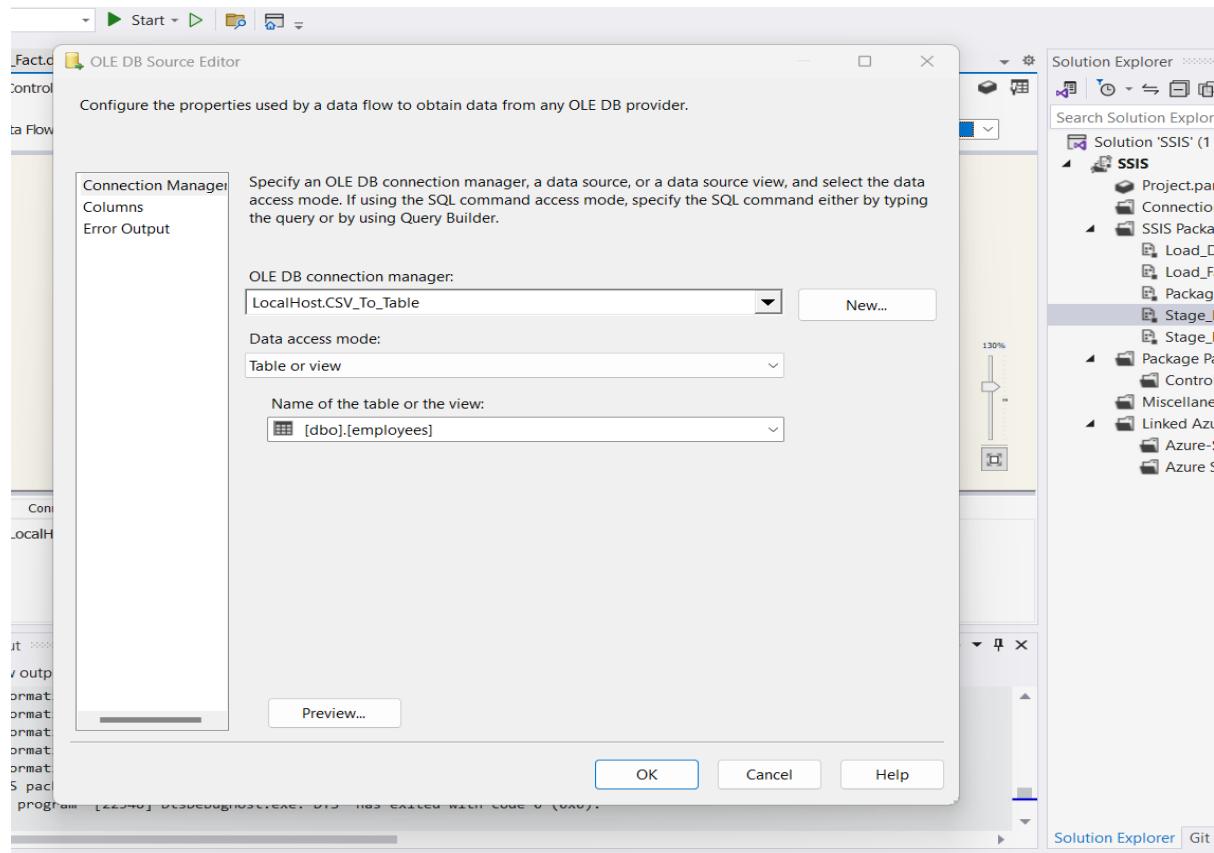


### 3.3.5. Employee Dimension

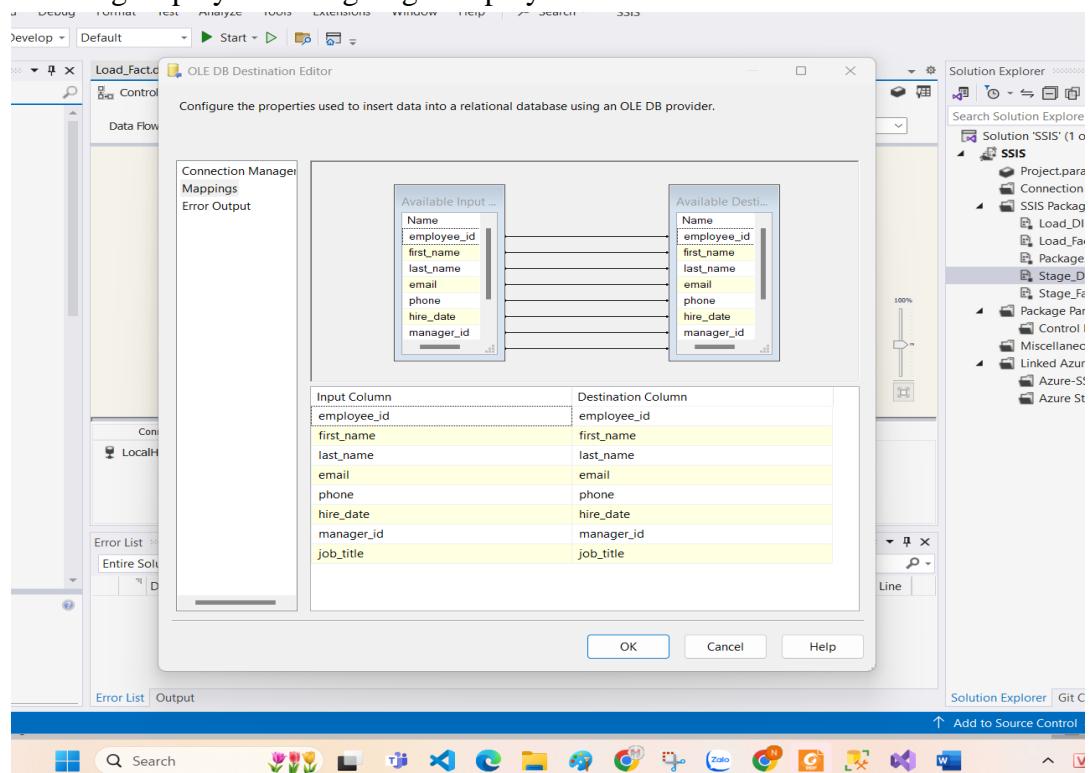
#### a) Load dữ liệu từ nguồn vào stgEmployee

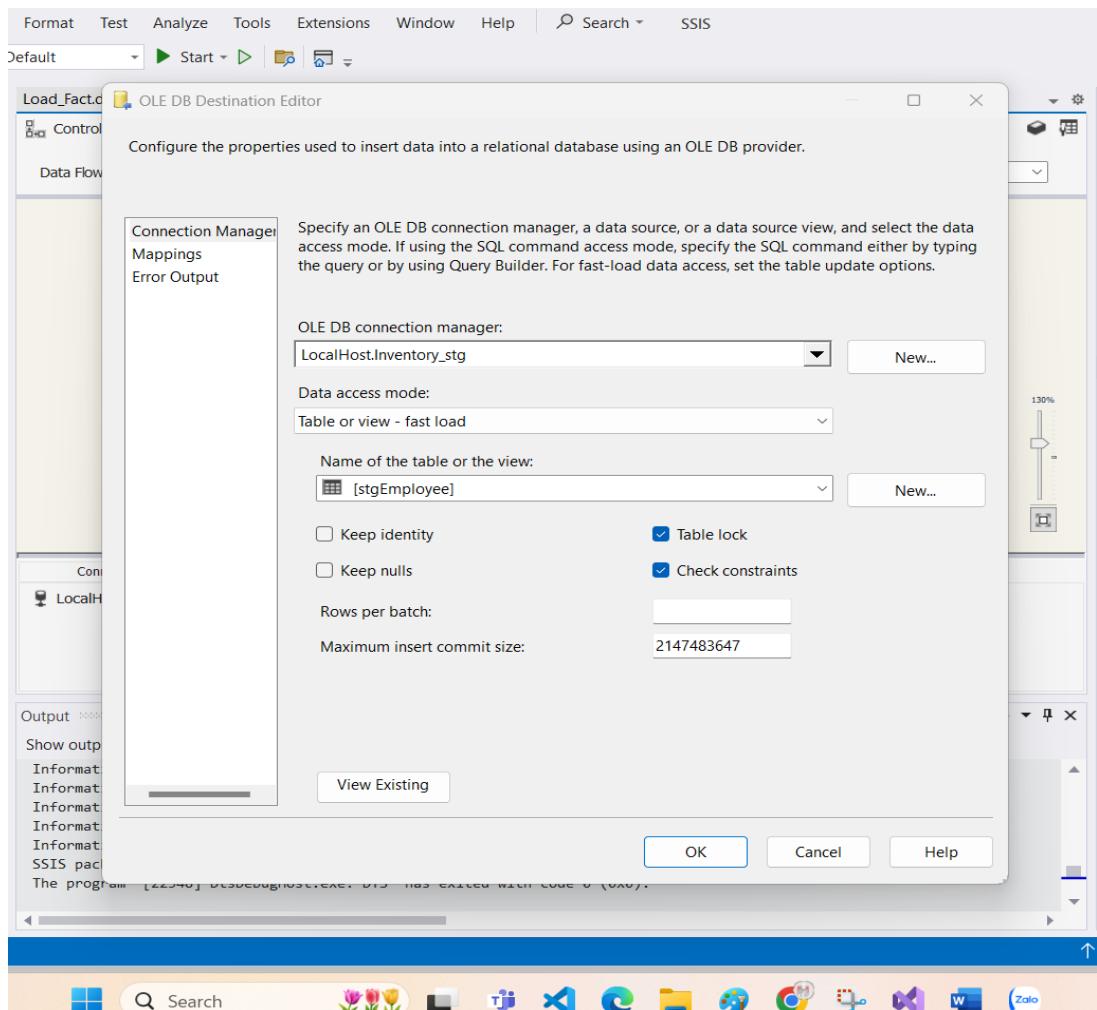


**SRC – Employee\_Data** là dữ liệu từ bảng Employee trong database CSV\_To\_Table chứa data về nhân viên

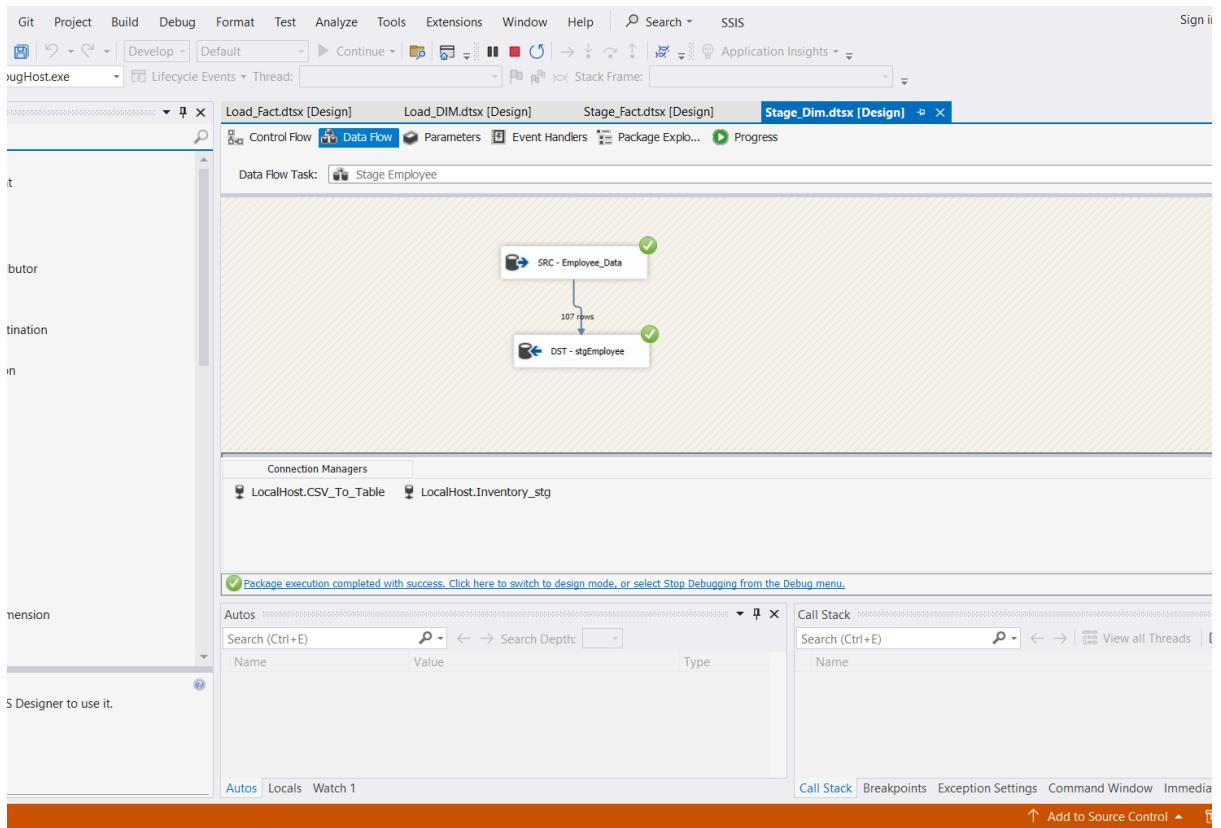


**DST -stgEmployee** là bảng stage Employee

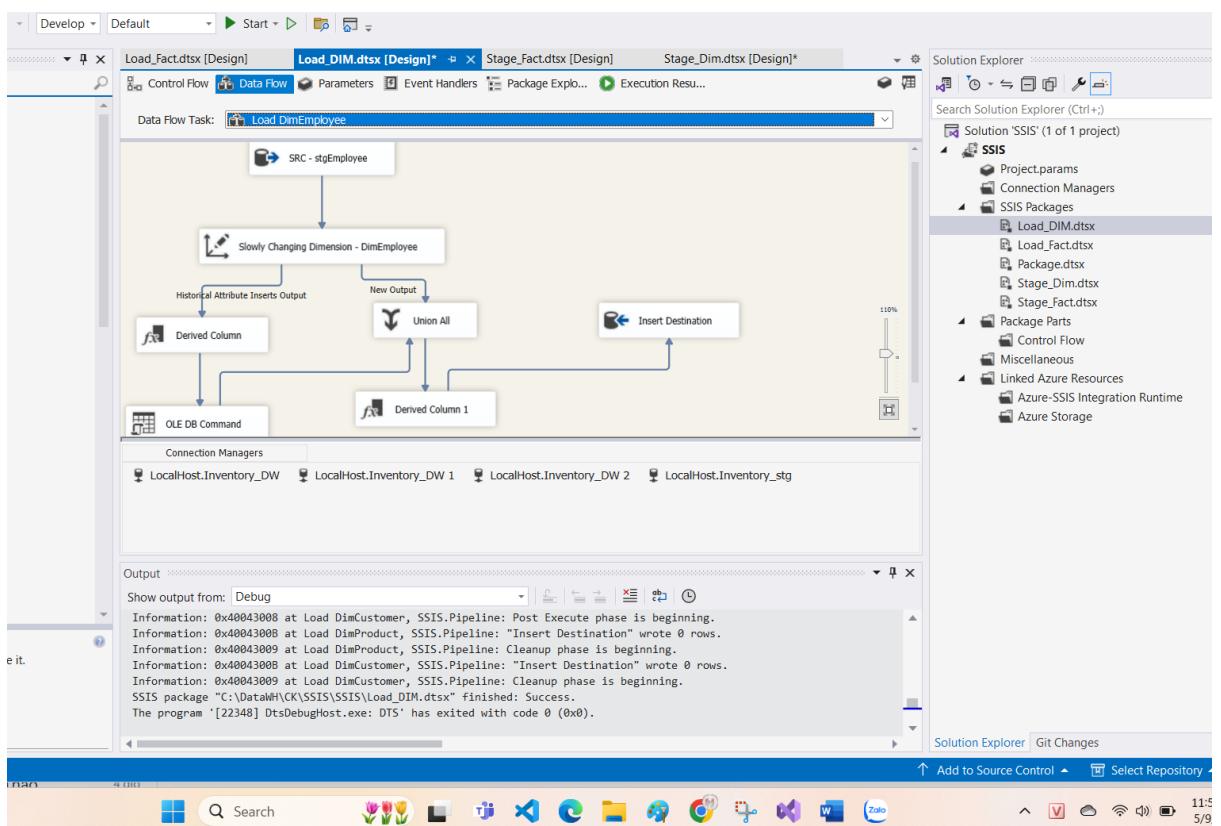




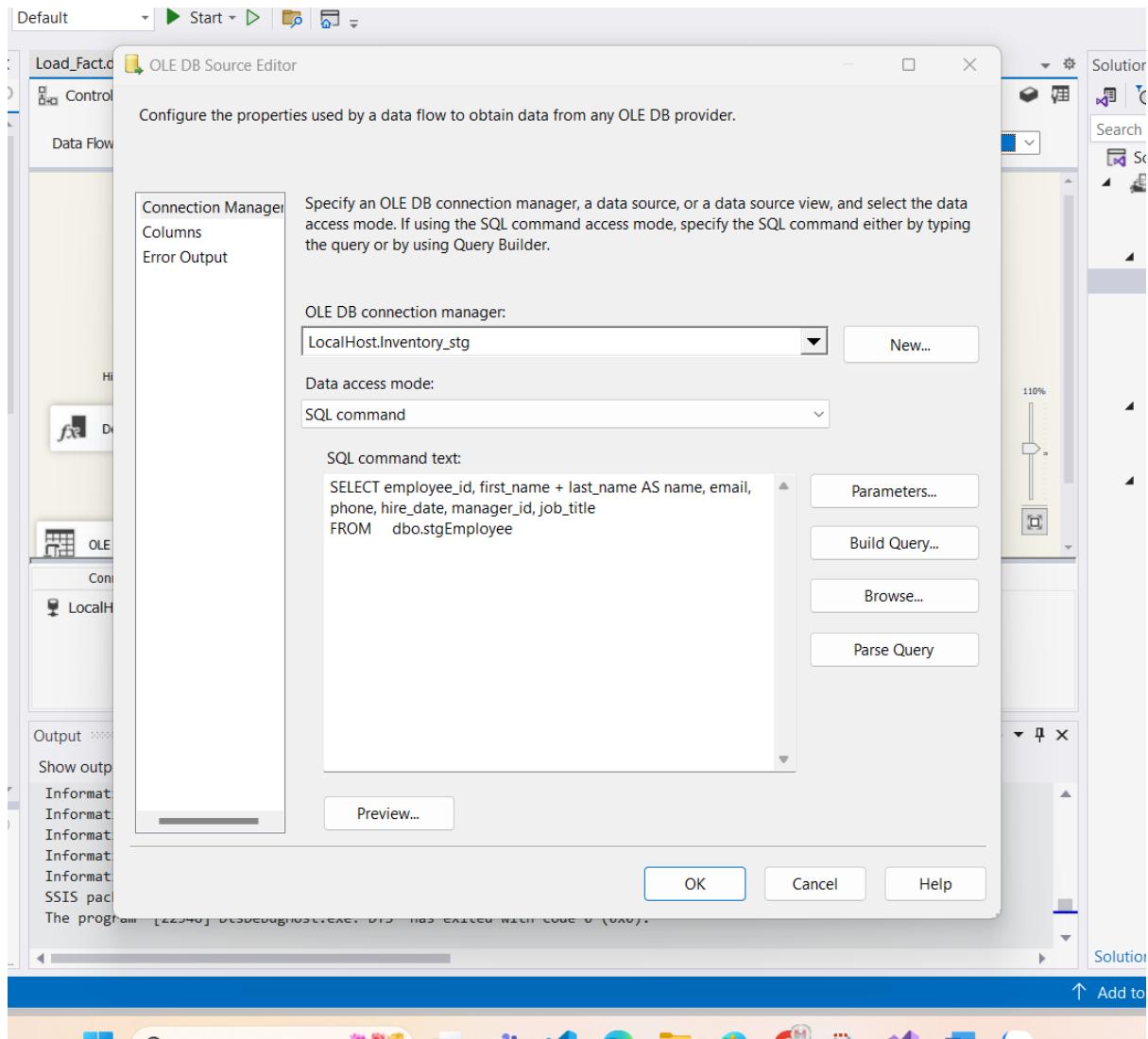
Khởi chạy



### b) Load dữ liệu từ stgEmployee vào bảng DimEmployee



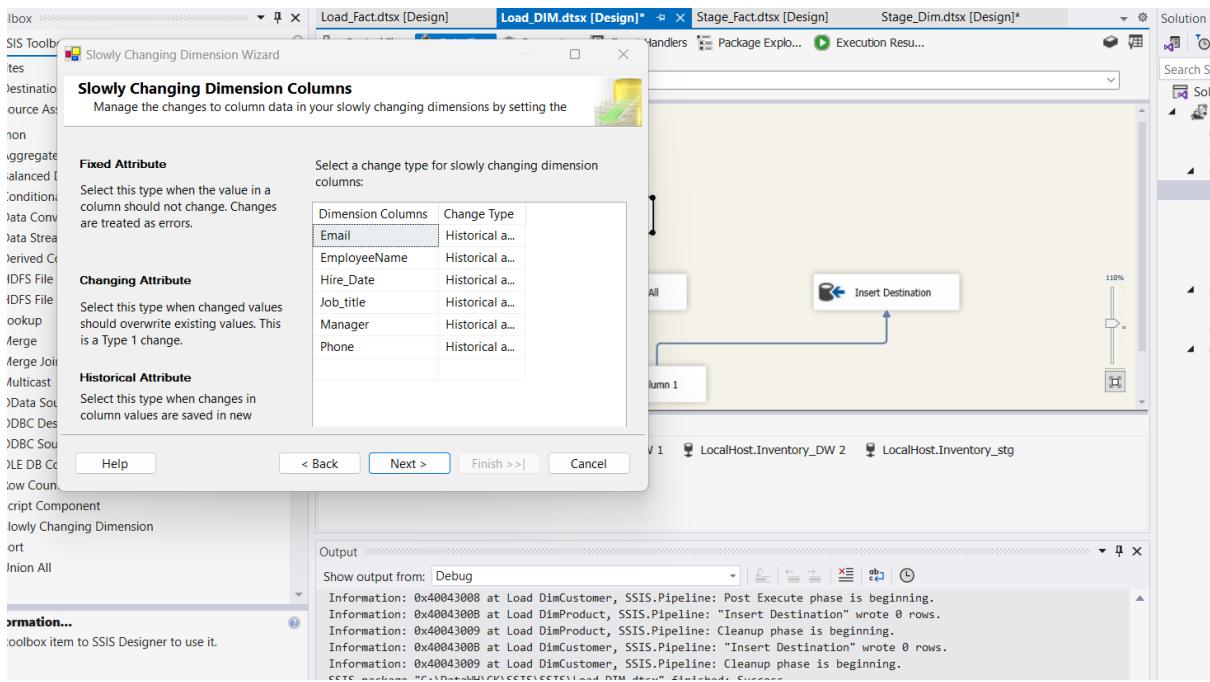
SRC – stgEmployee chứa dữ liệu bảng Employee



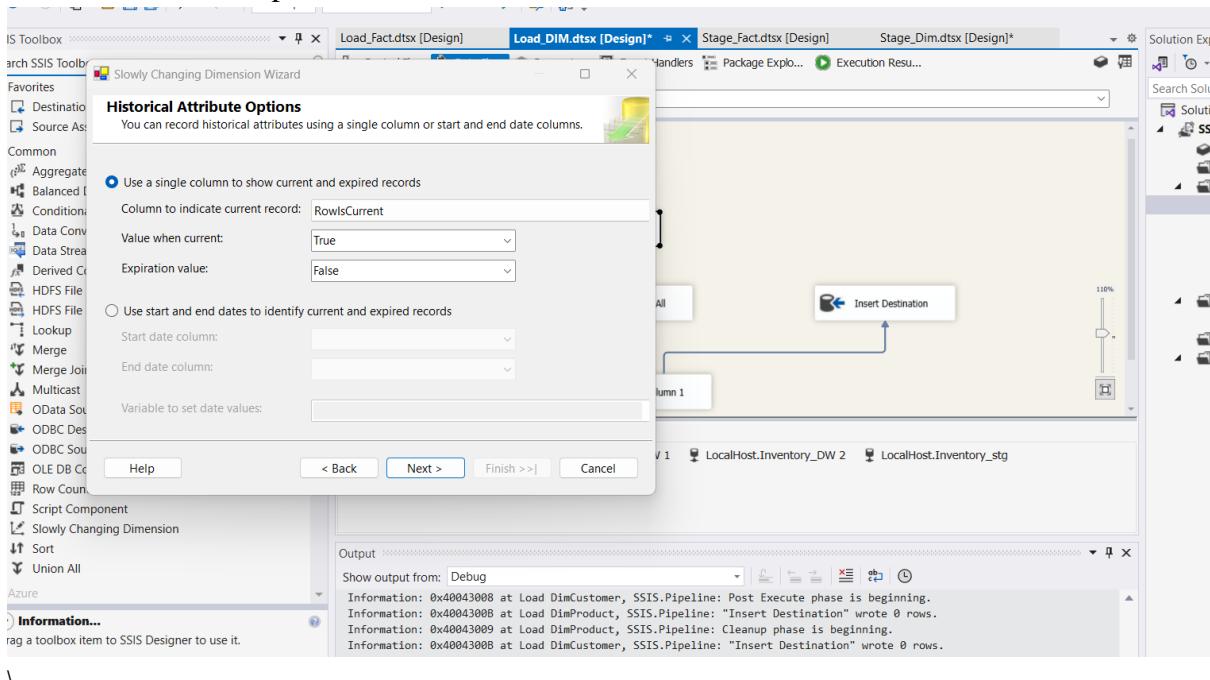
## Slowly Changing Dimension – DimEmployee

The program '[22540] D:\desdebughost\SSIS\Slowly Changing Dimension - DimEmployee.dtsx' has exited with code 0 (0x0).

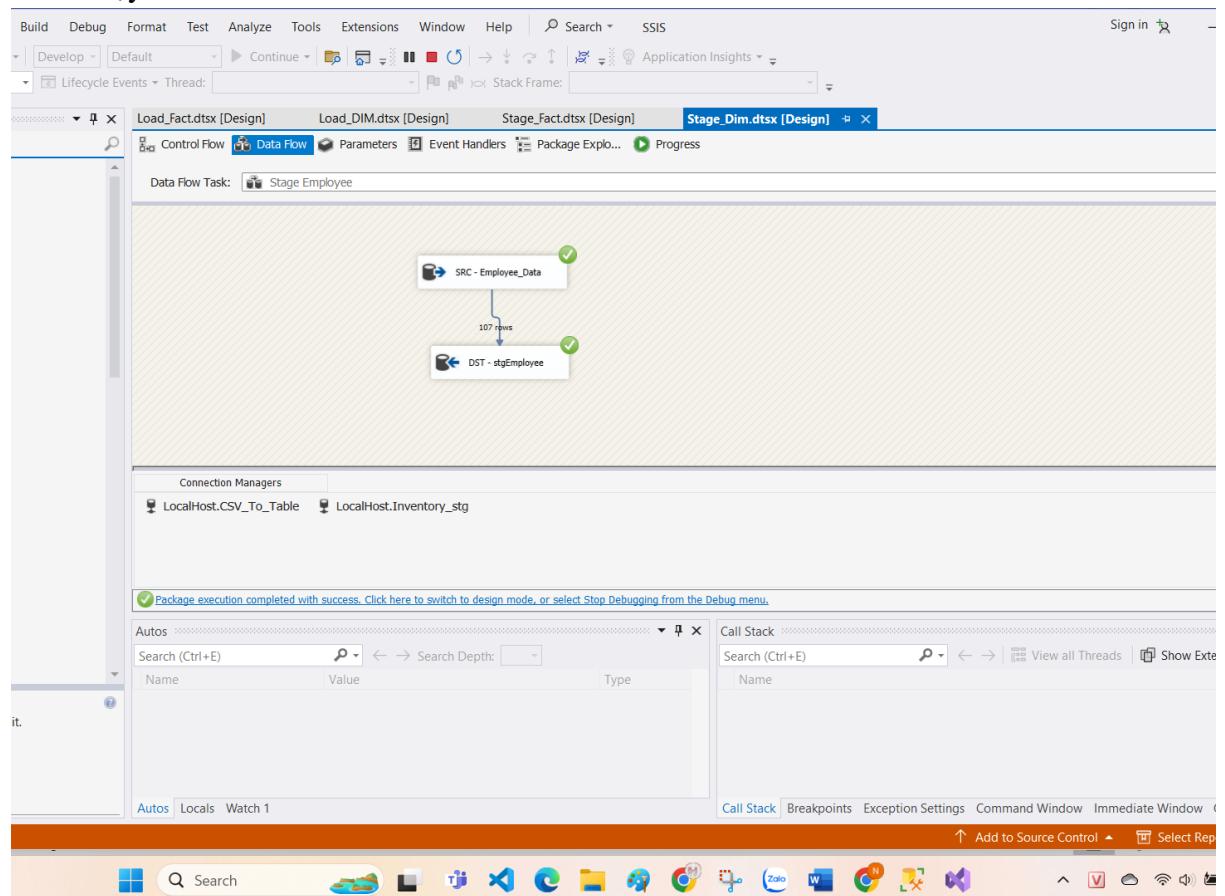
SCD chọn Historical attribute



## Historical attribute options chọn RowIsCurrent

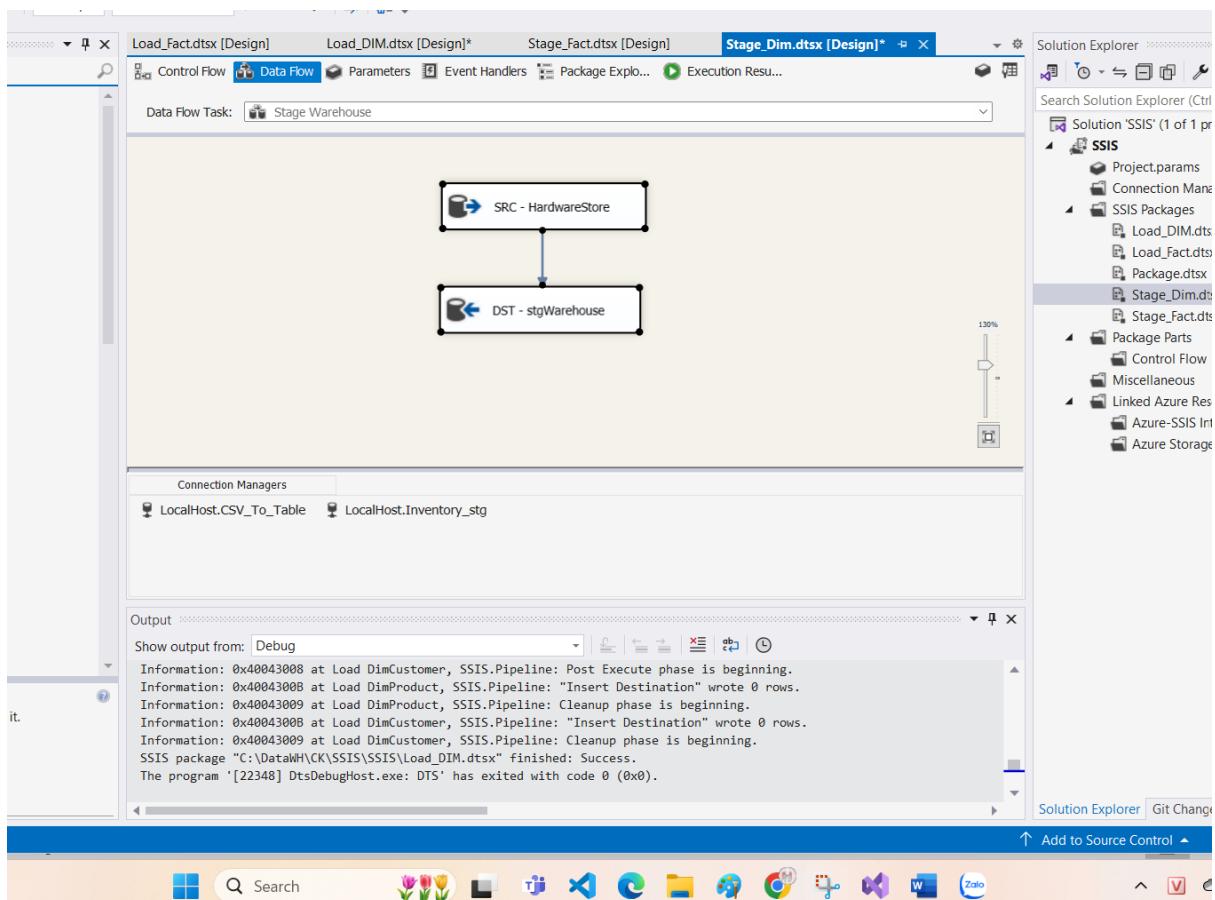


## Khởi chạy

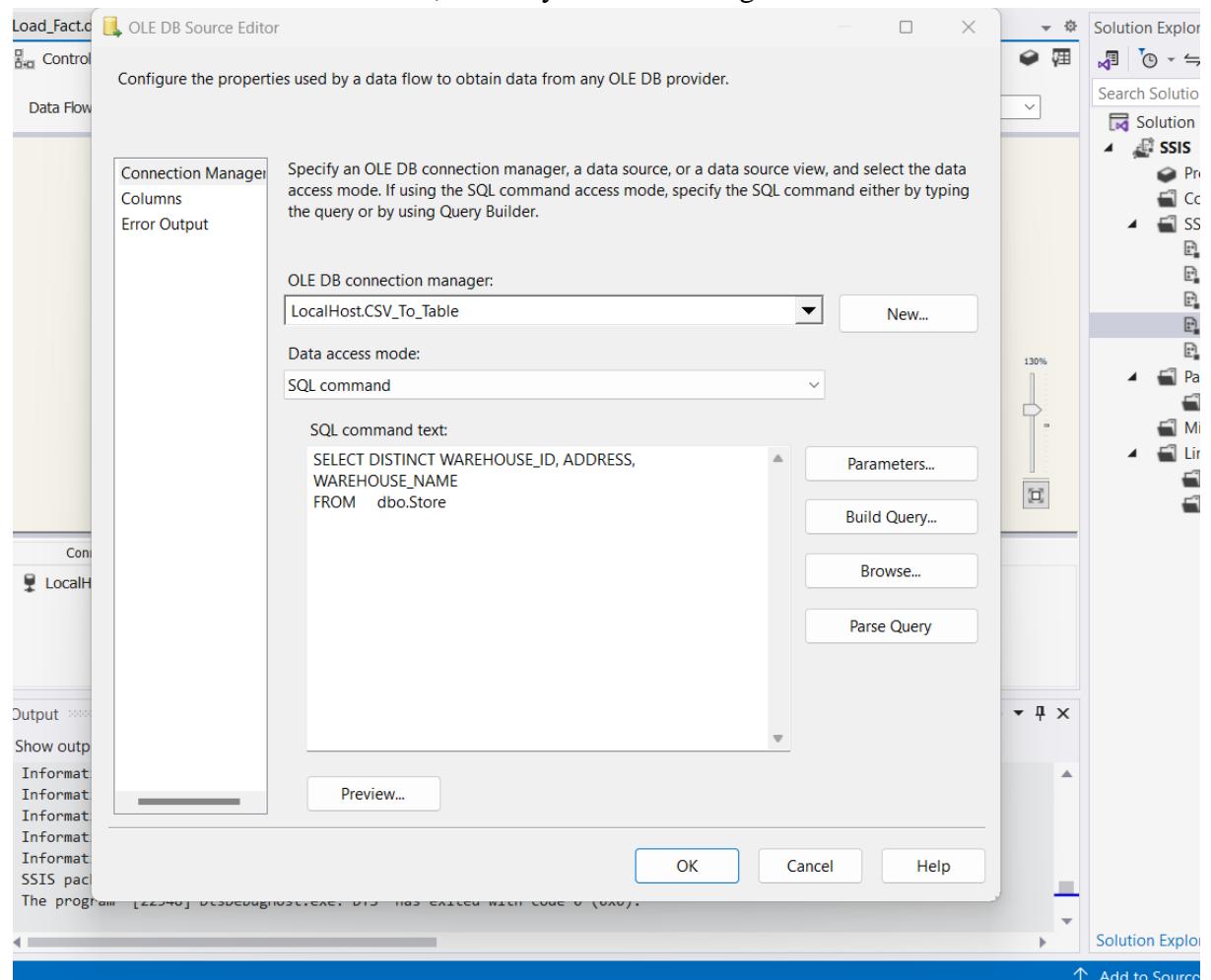


### 3.3.6. Warehouse Dimension

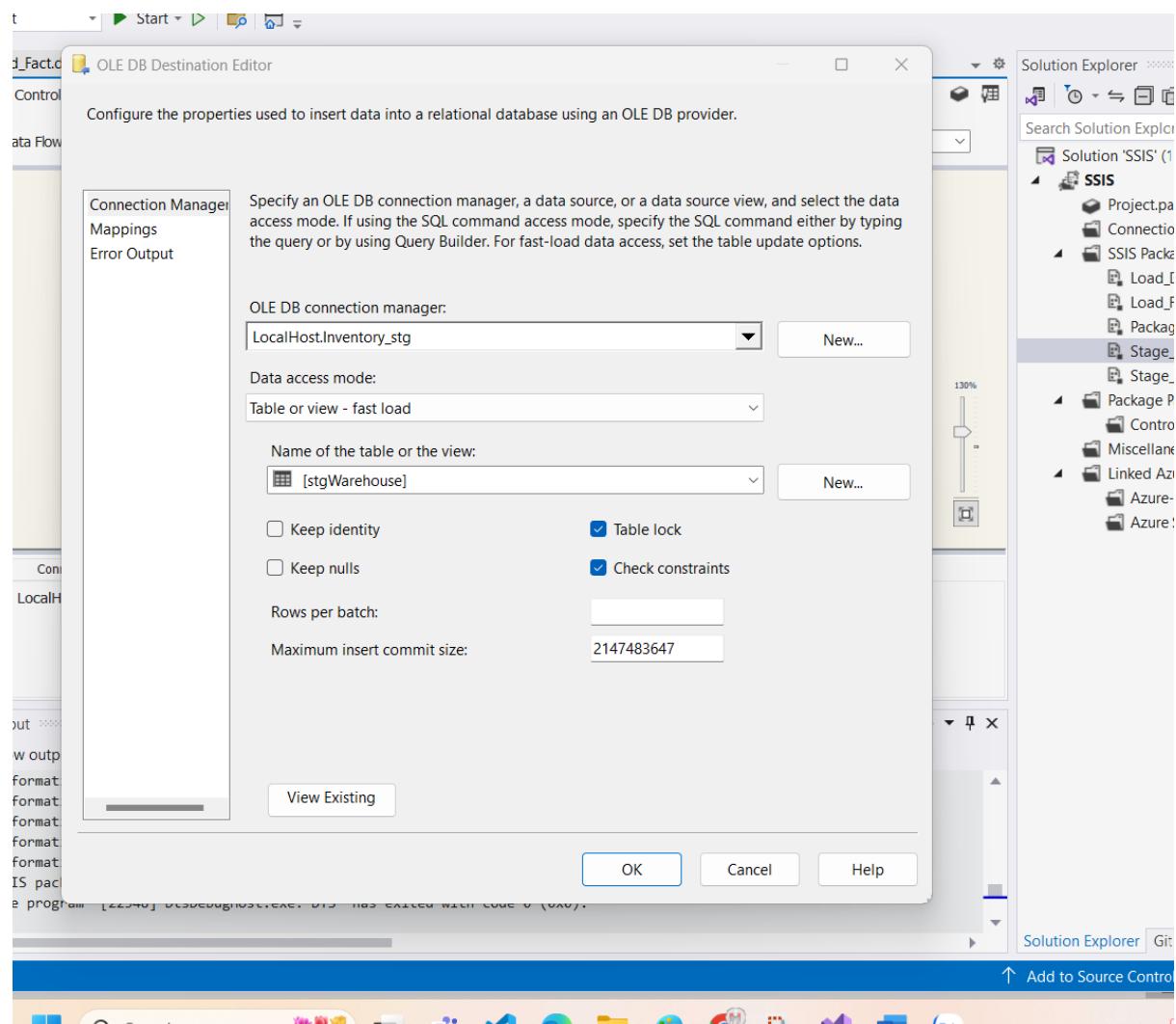
#### a) Load dữ liệu từ nguồn vào stgWarehouse

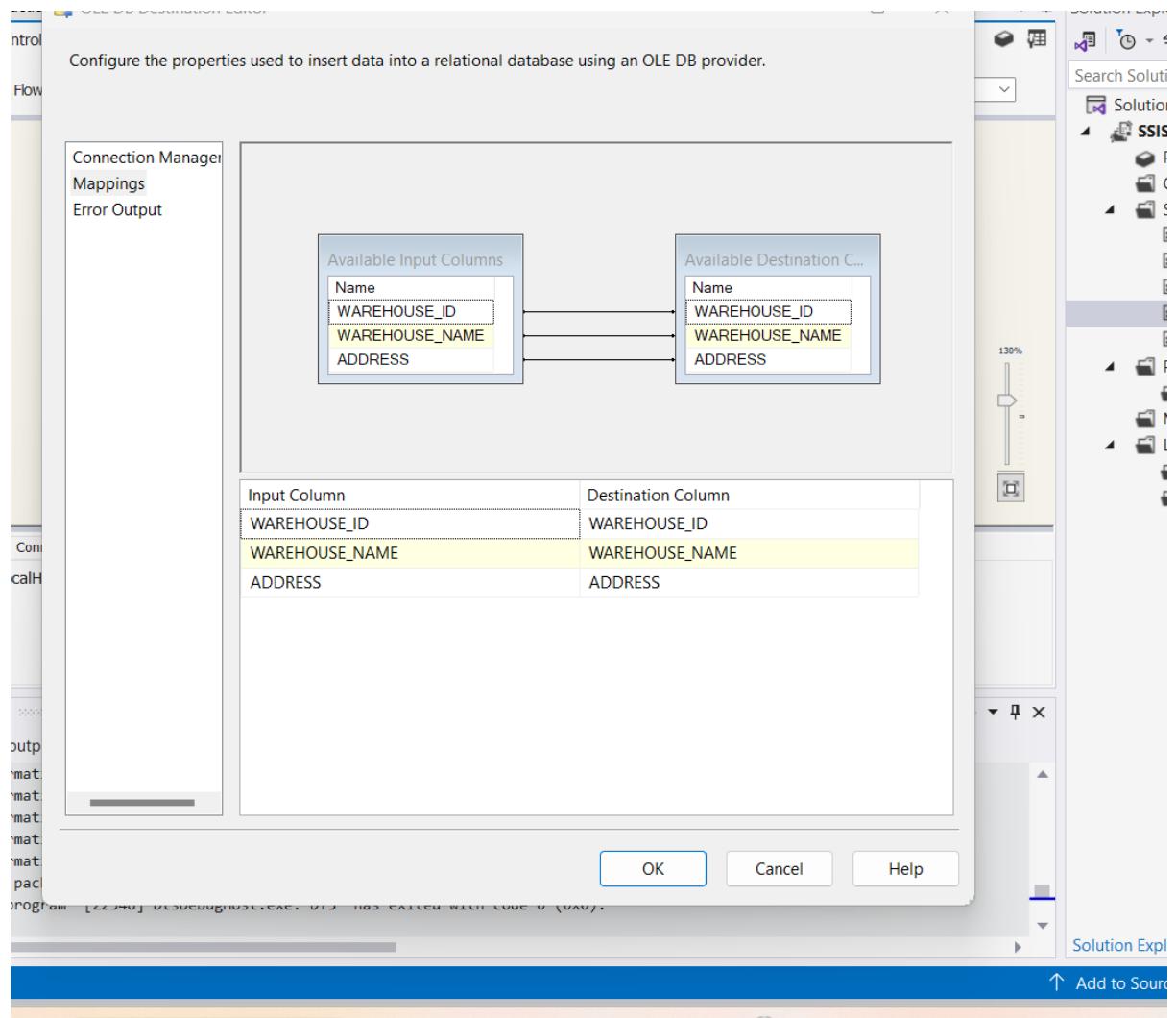


## SRC – HardwareStore chưa dữ liệu cần lấy để load vào stgWarehouse

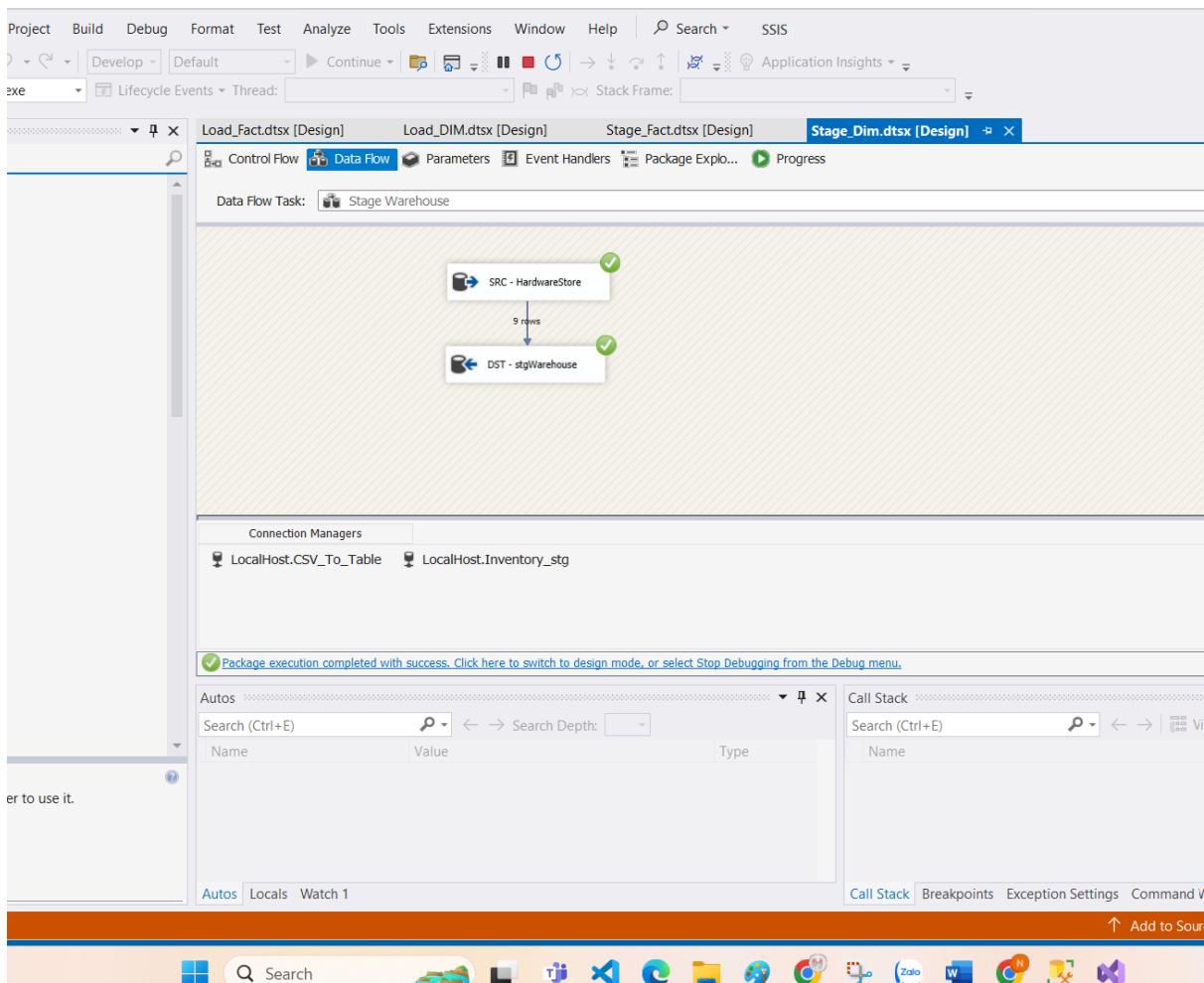


## DST – stgWarehouse

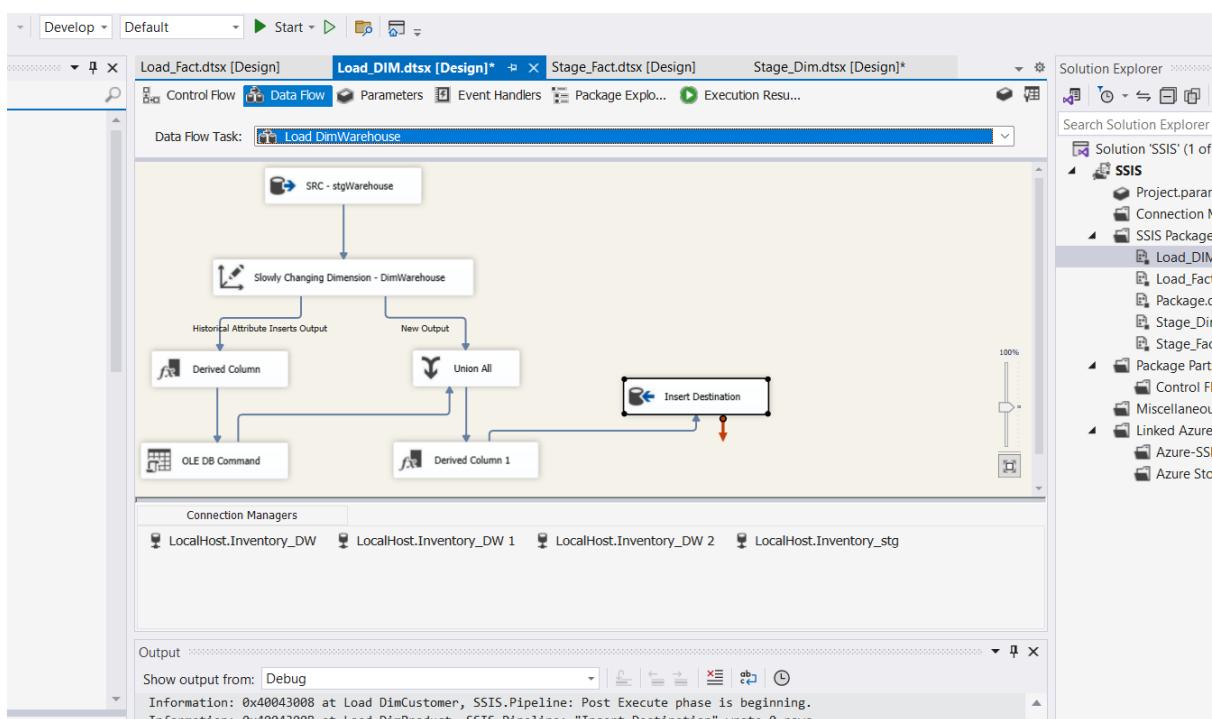




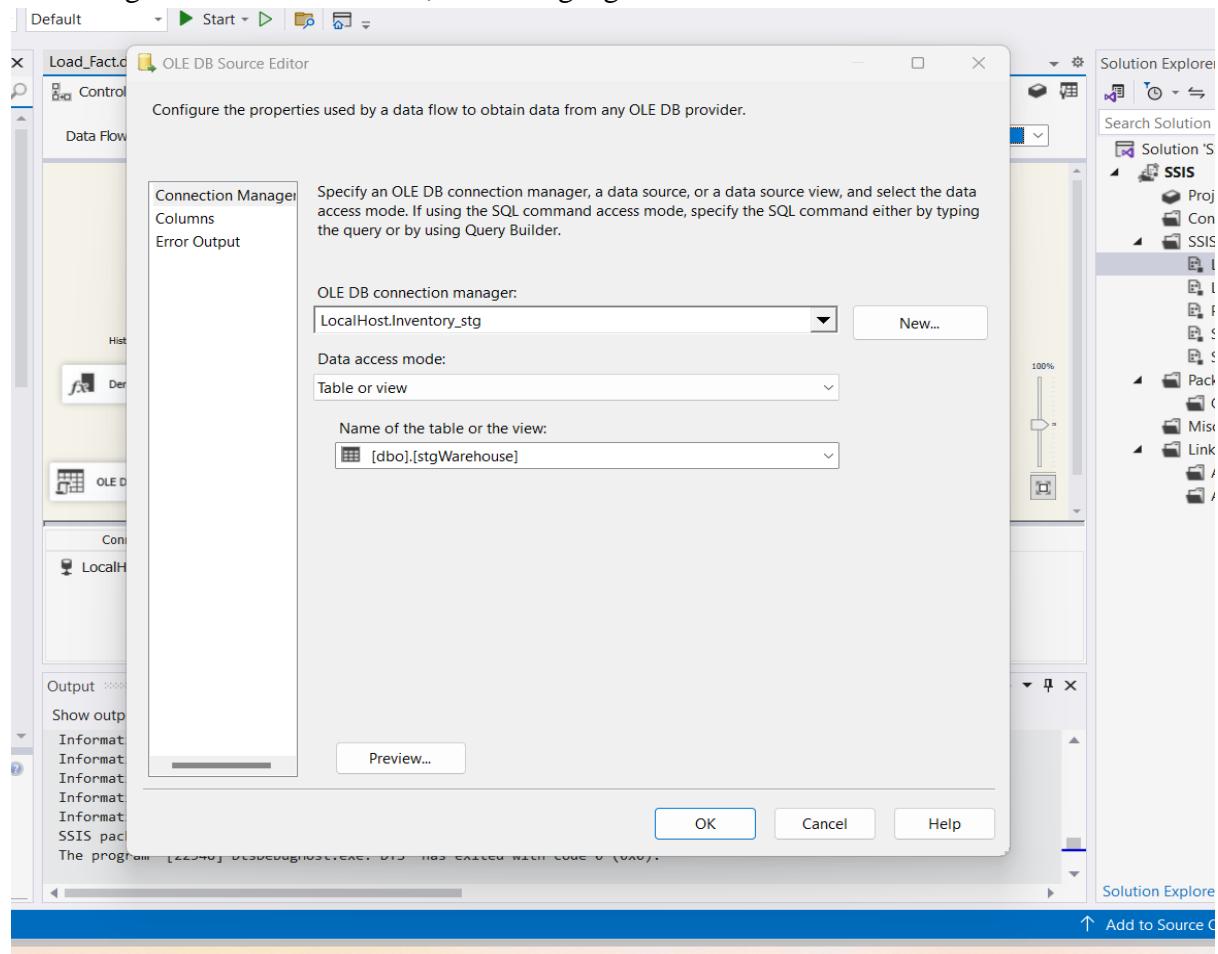
Khởi chạy



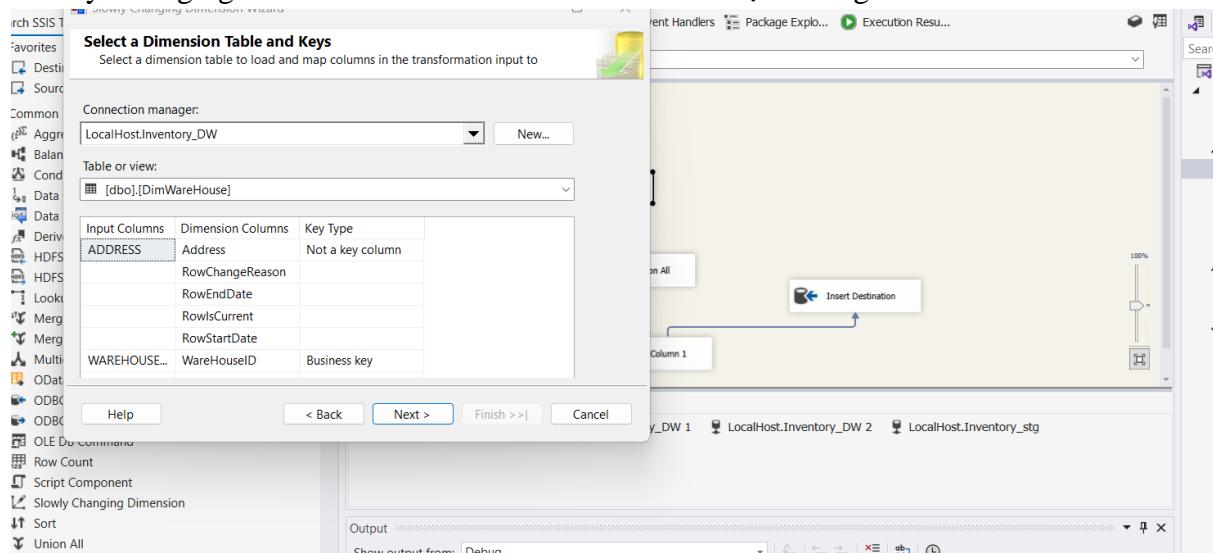
### b) Load dữ liệu từ stgWarehouse vào DimWarehouse



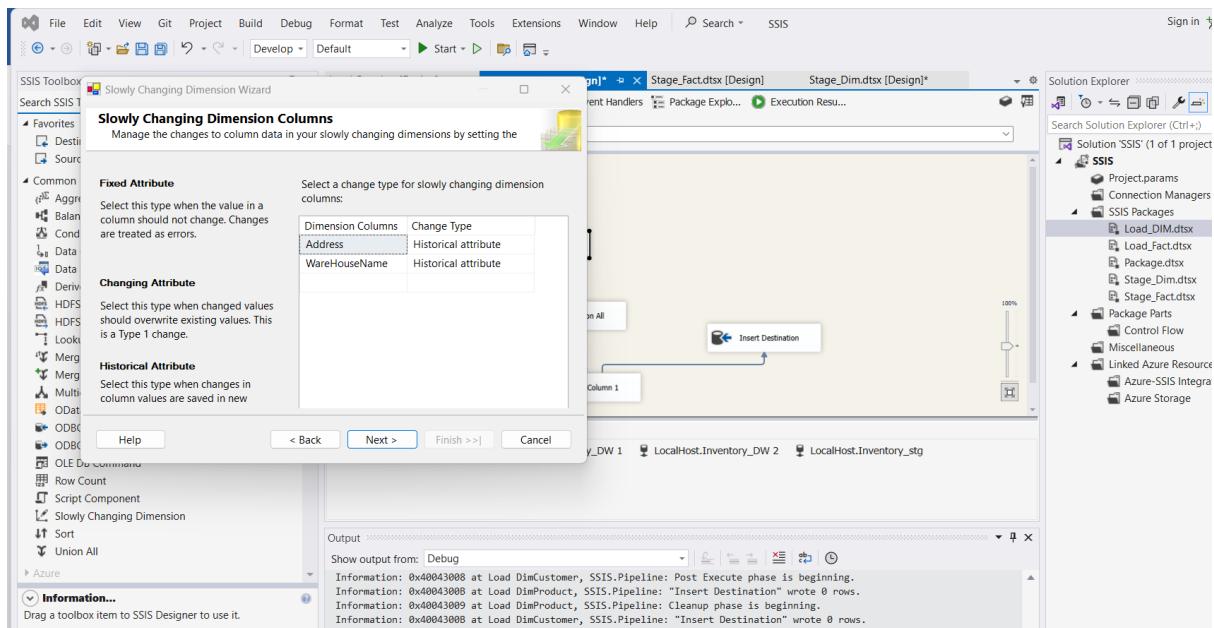
## SRC – stgWarehouse chứa dữ liệu của bảng stgWarehouse



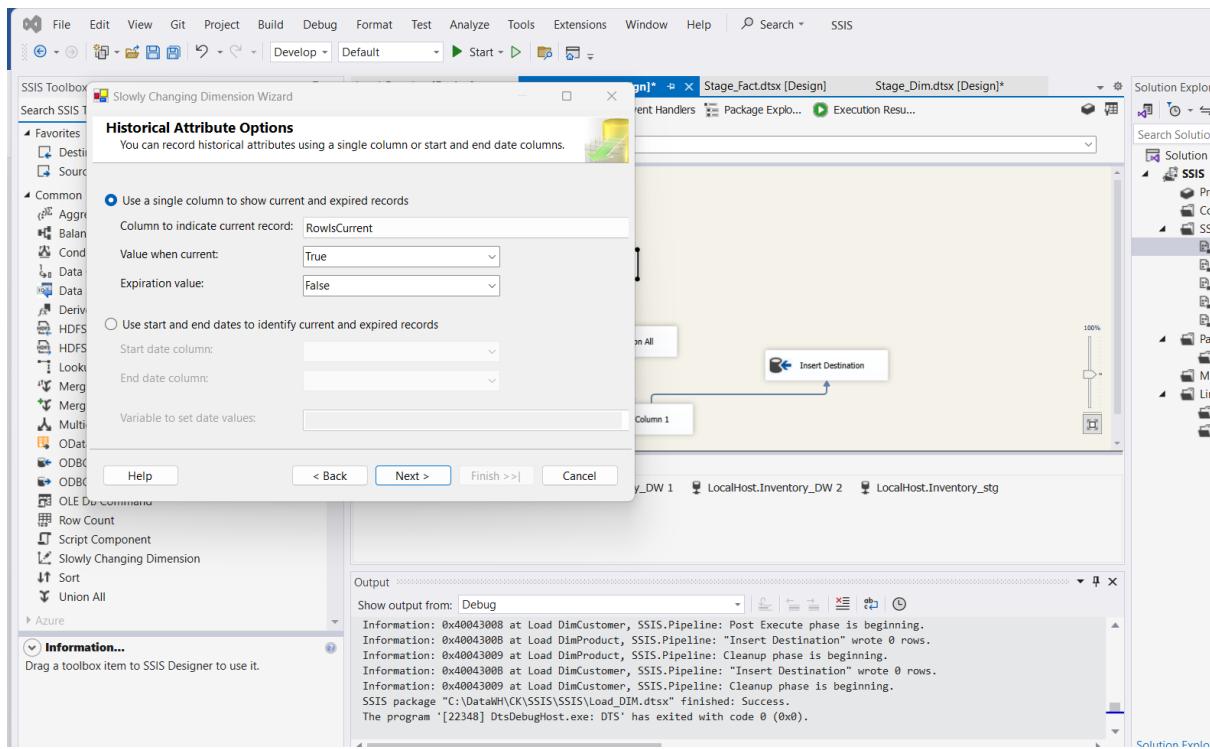
## Slowly Changing Dimension – DimWarehouse load dữ liệu từ stage vào DimWarehouse



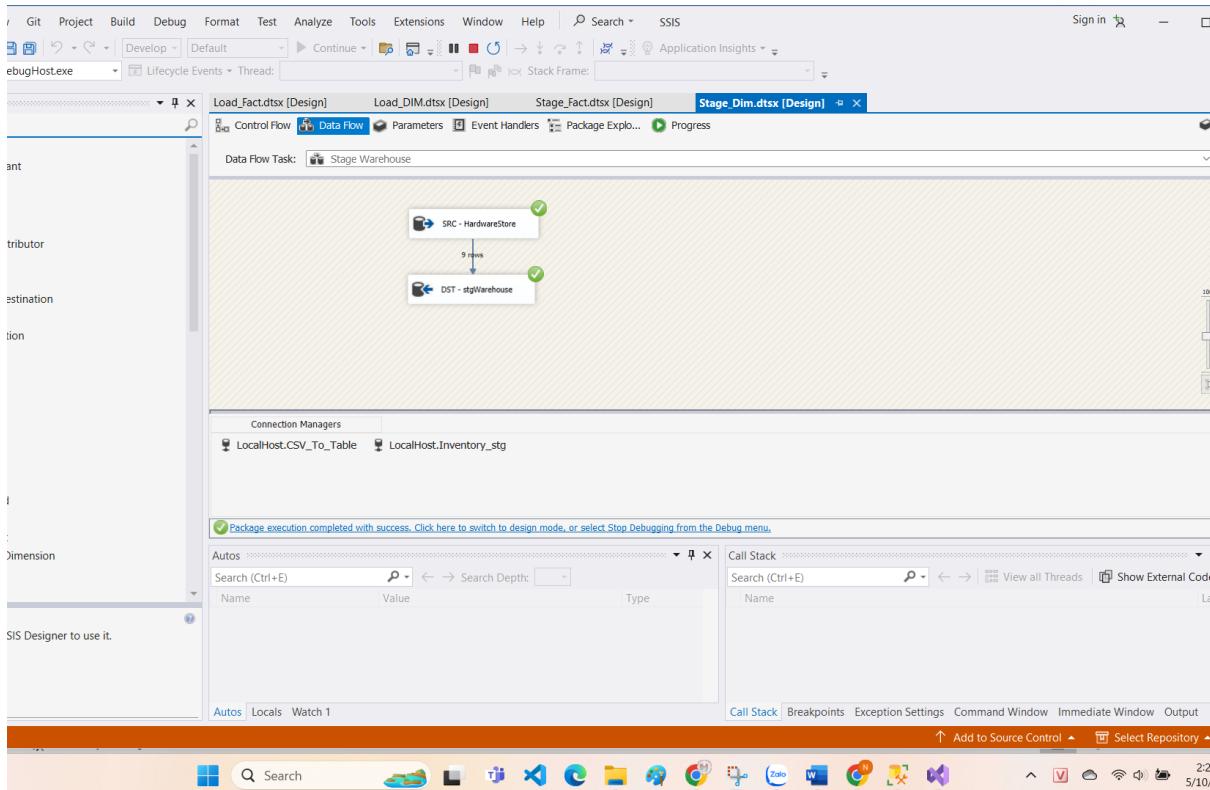
SCD chọn Historical attribute



## Historical attribute options chọn RowIsCurrent

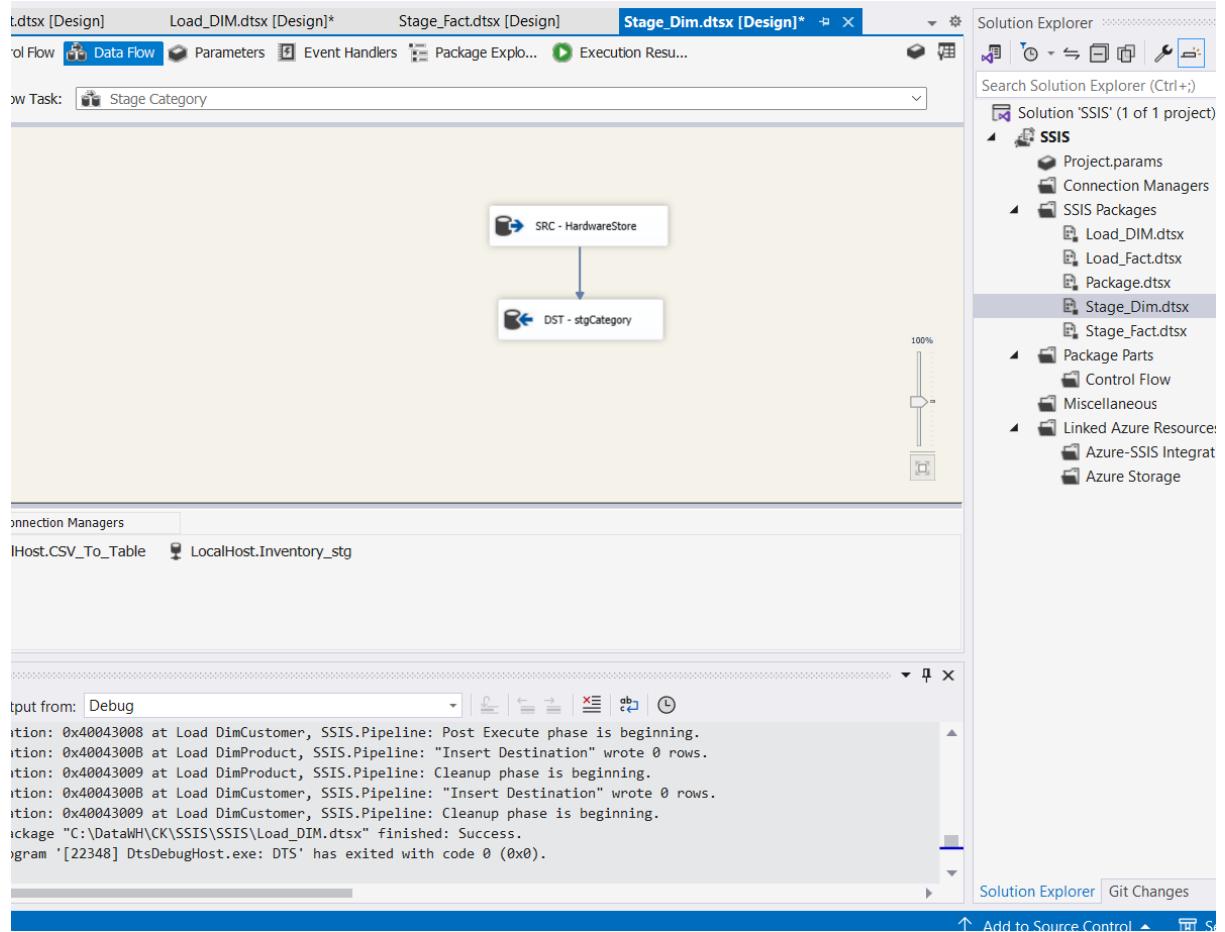


## Khởi chạy

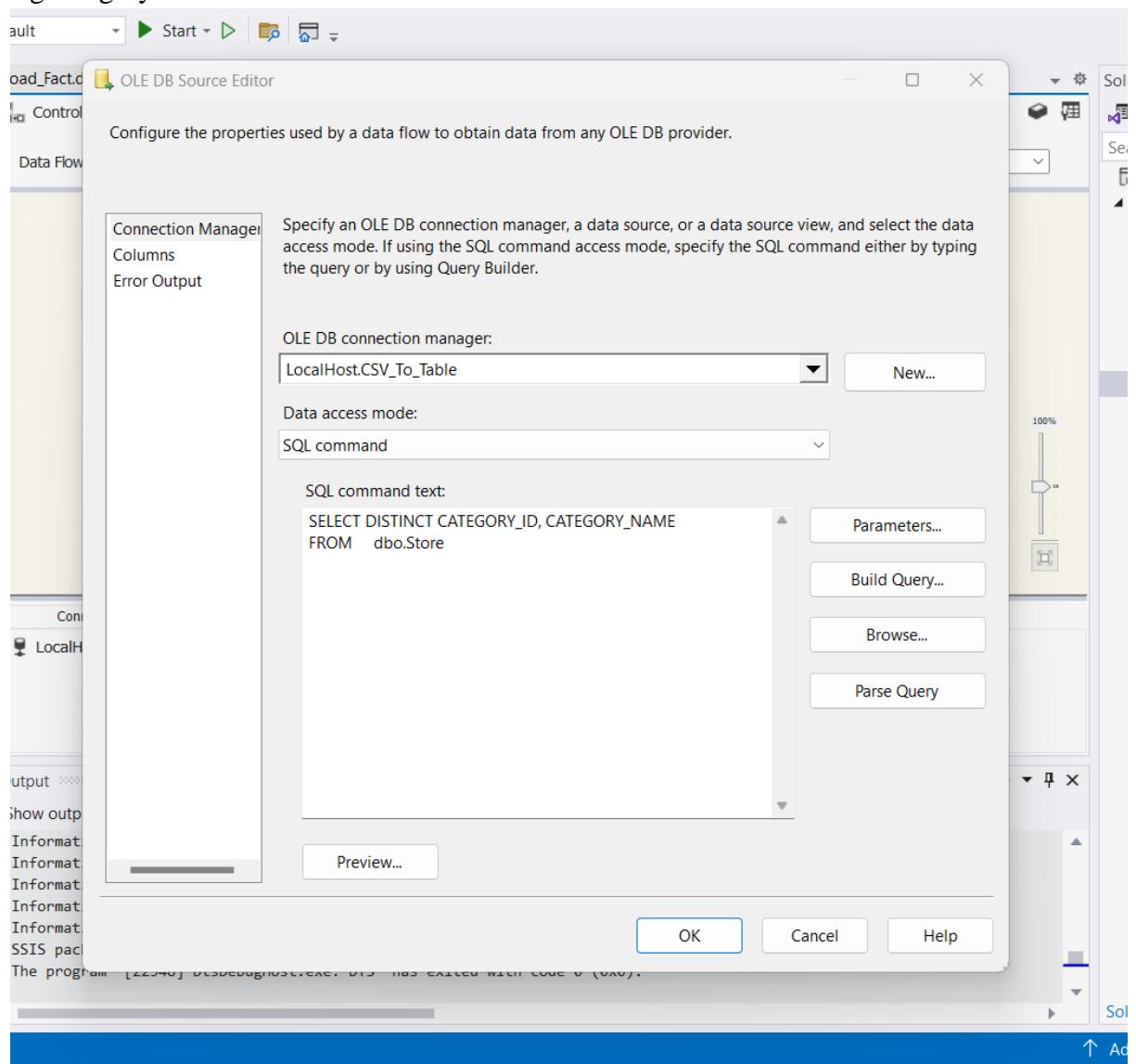


### 3.3.7. Category Dimension

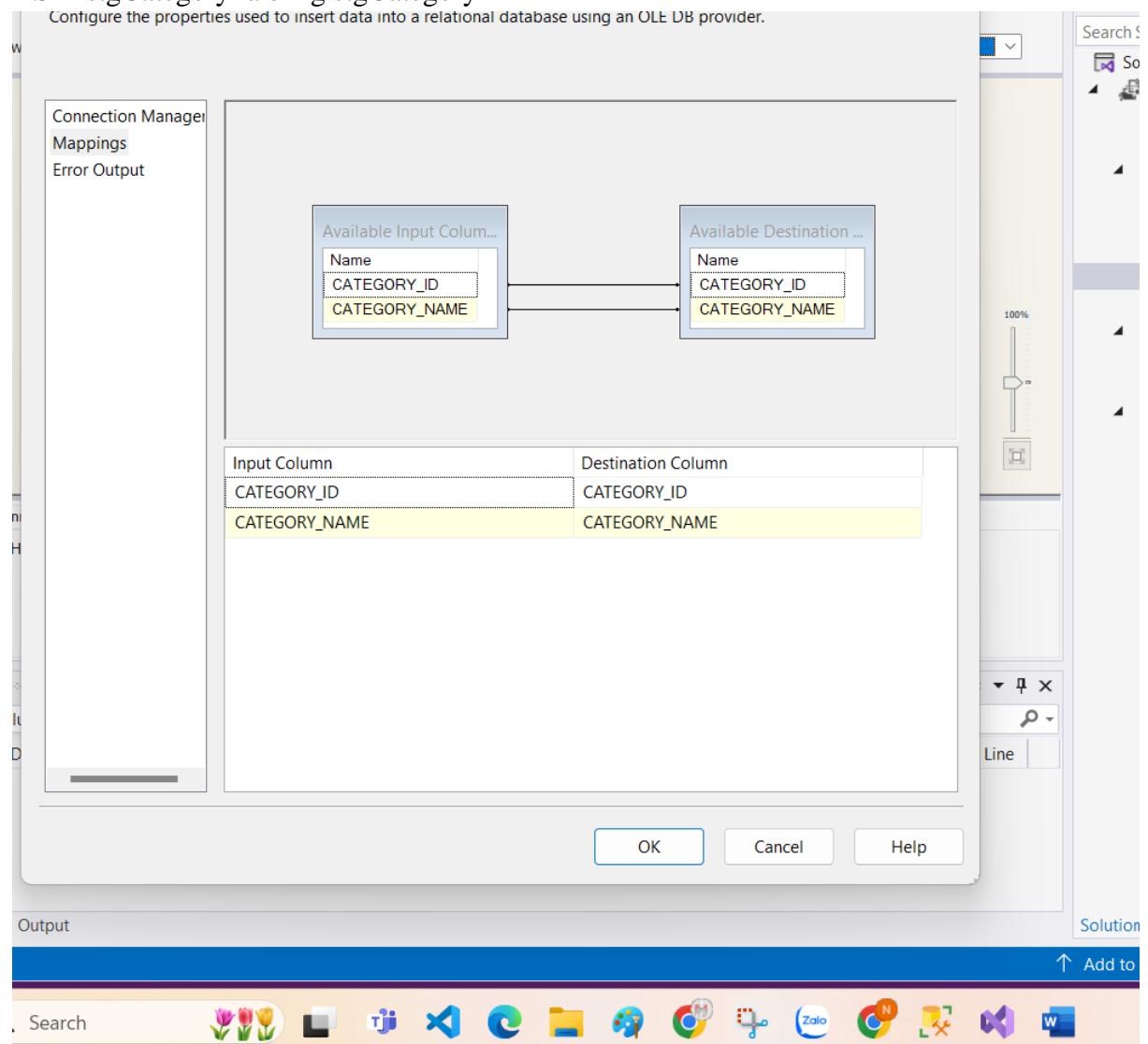
#### a) Load dữ liệu từ nguồn vào stgCategory

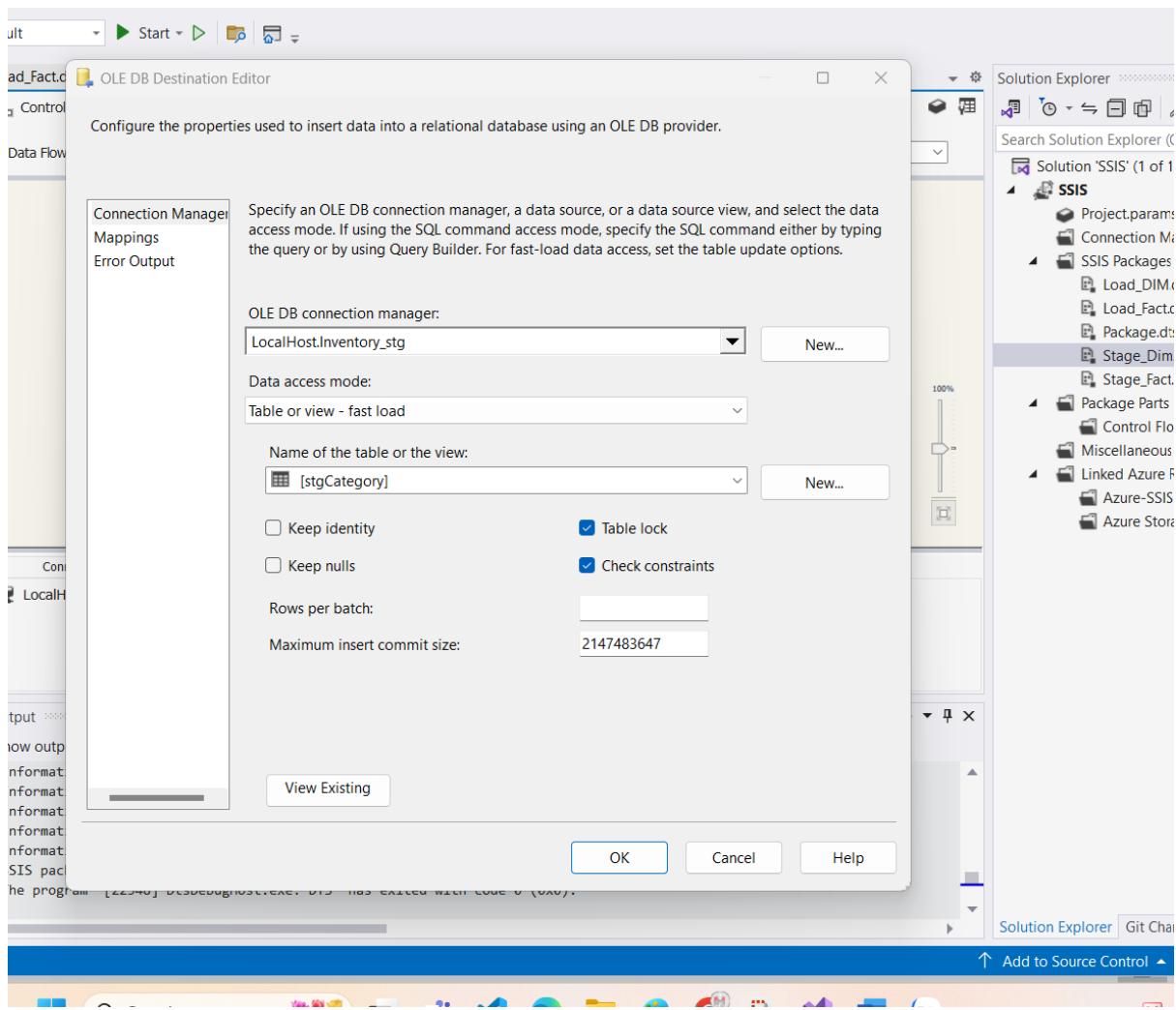


SRC – HardwareStore là dữ liệu đã được truy vấn để lấy các giá trị cần thiết cho bảng stgCategory

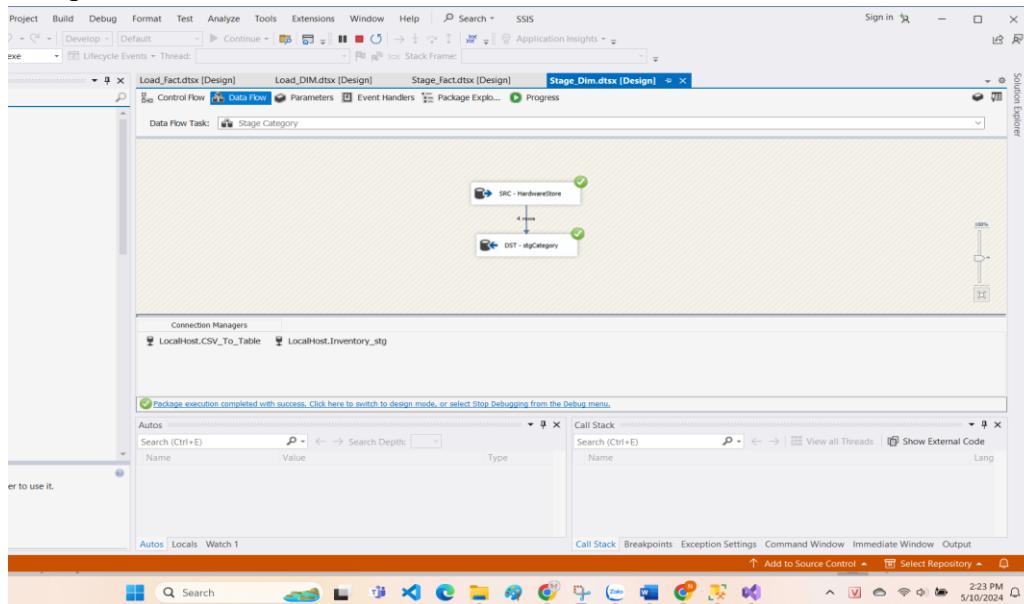


## DST -stgCategory là bảng stgCategory

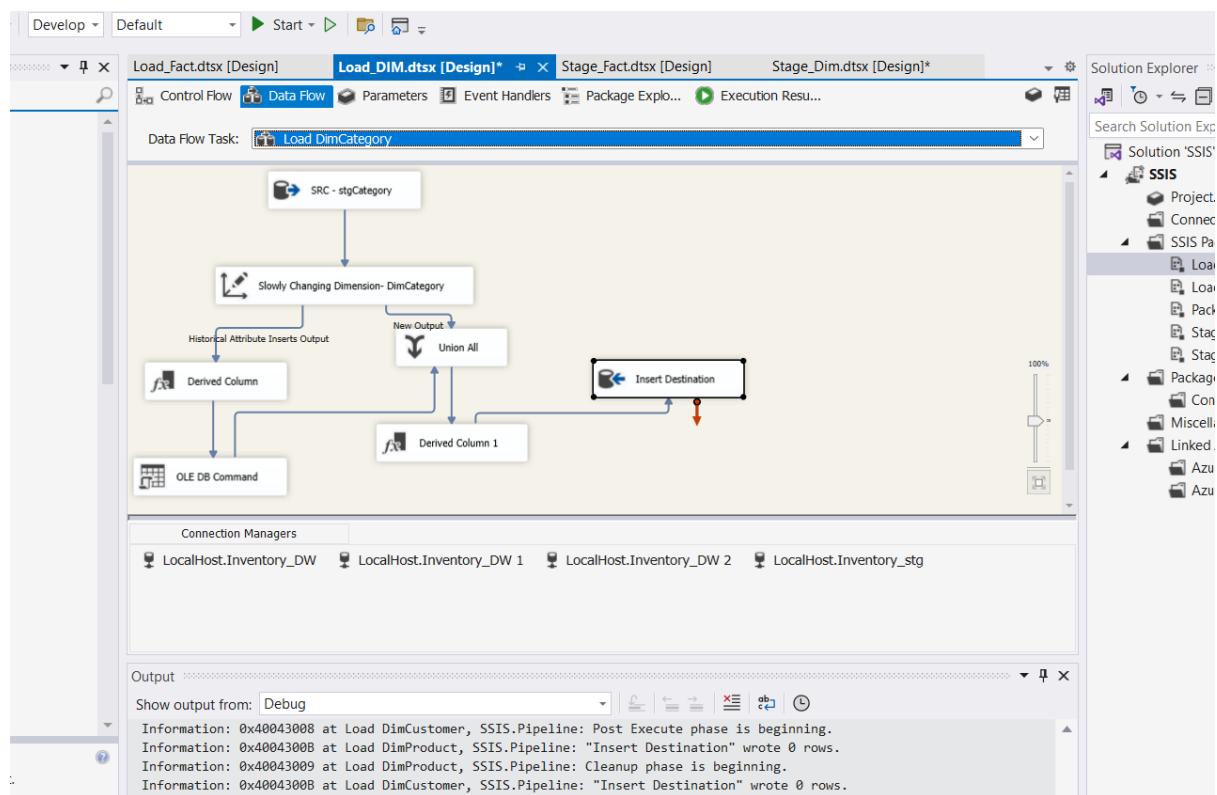




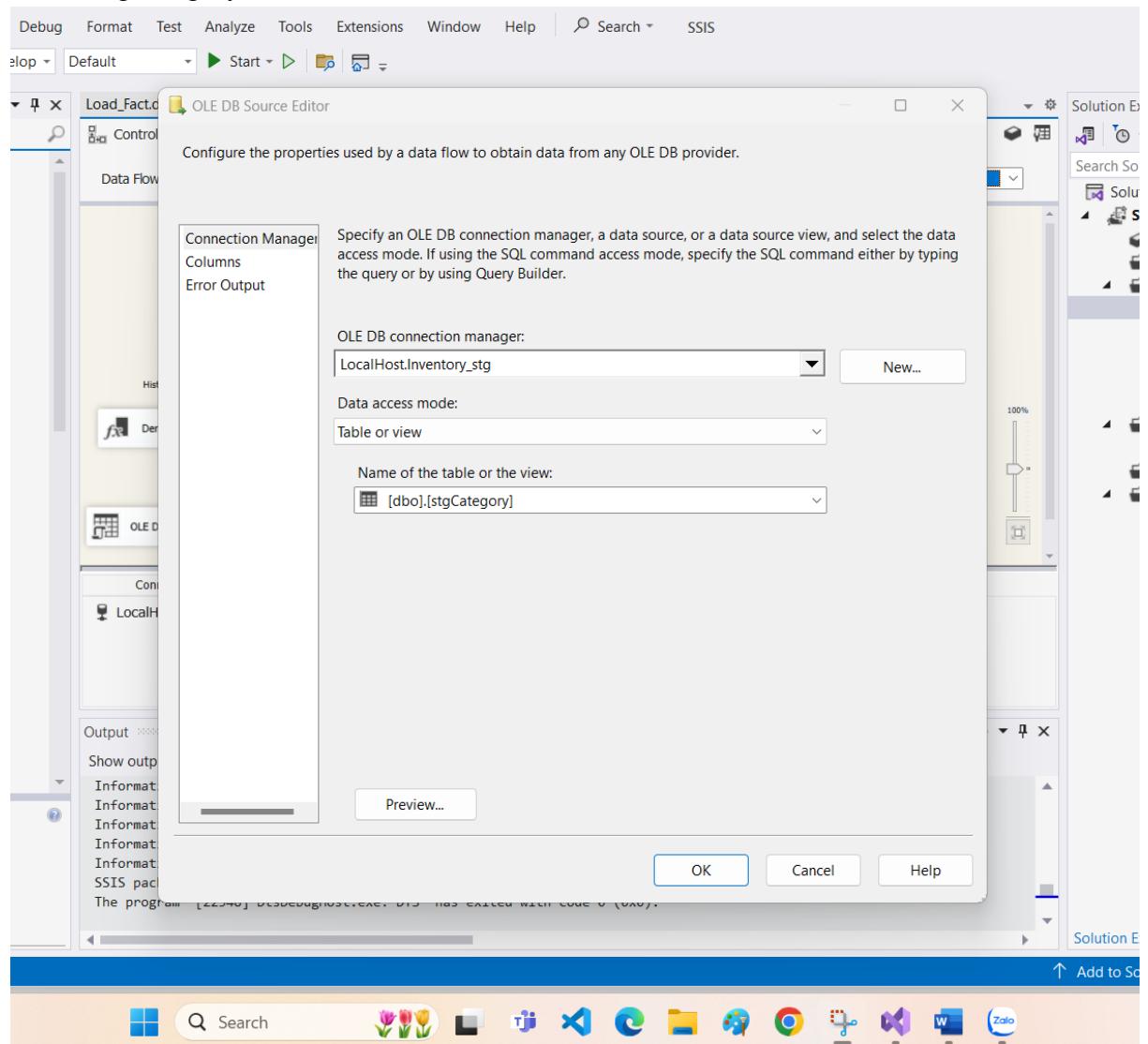
Kết quả



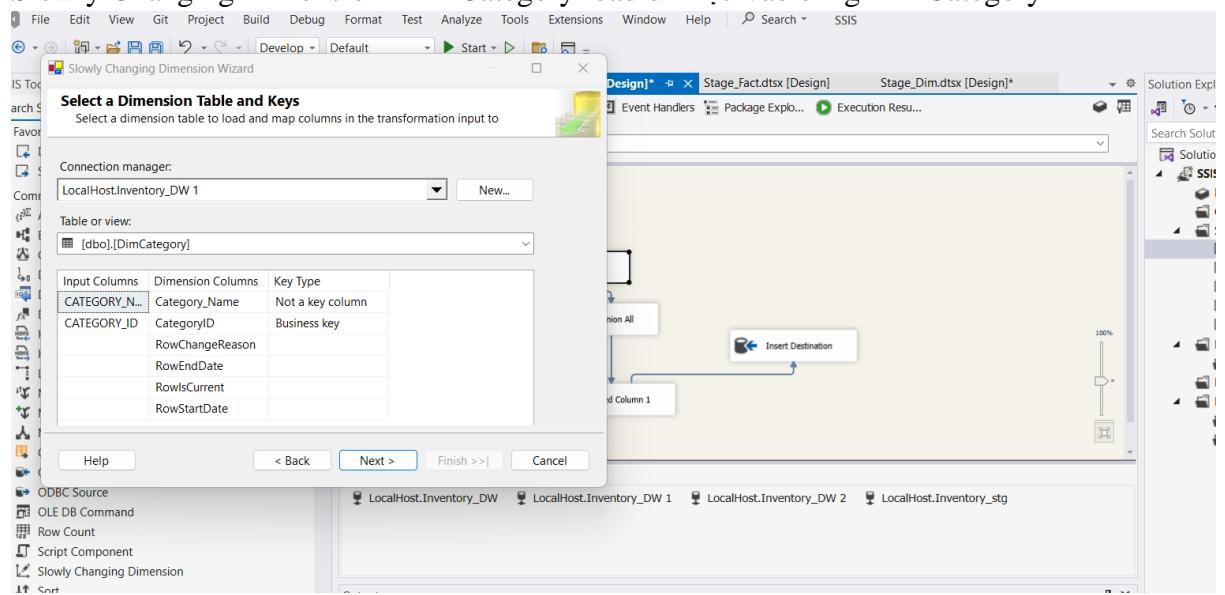
b) Load dữ liệu từ stgCategory vào DimCategory



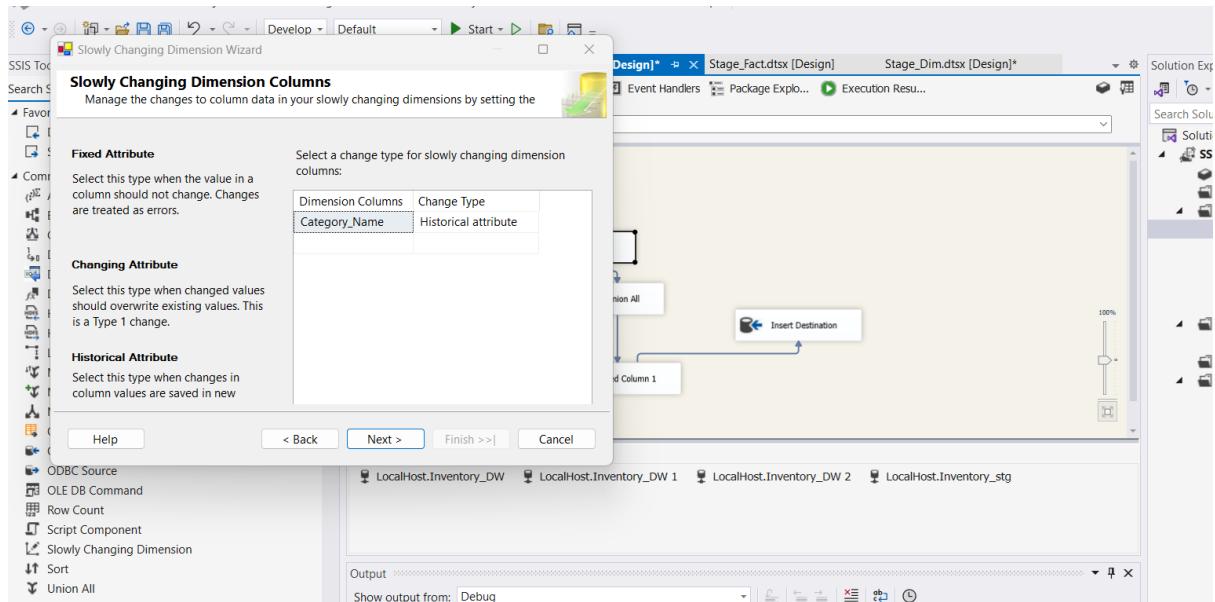
## SRC – stgCategory



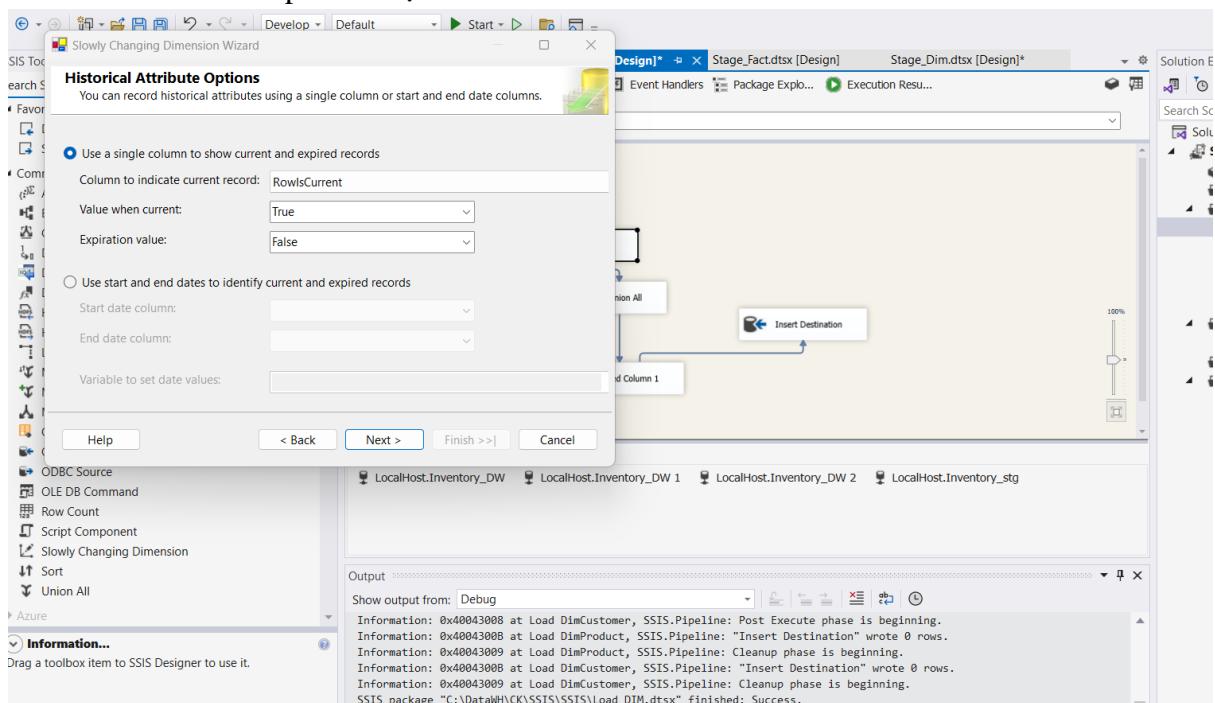
## Slowly Changing Dimension- DimCategory load dữ liệu vào bảng DimCategory



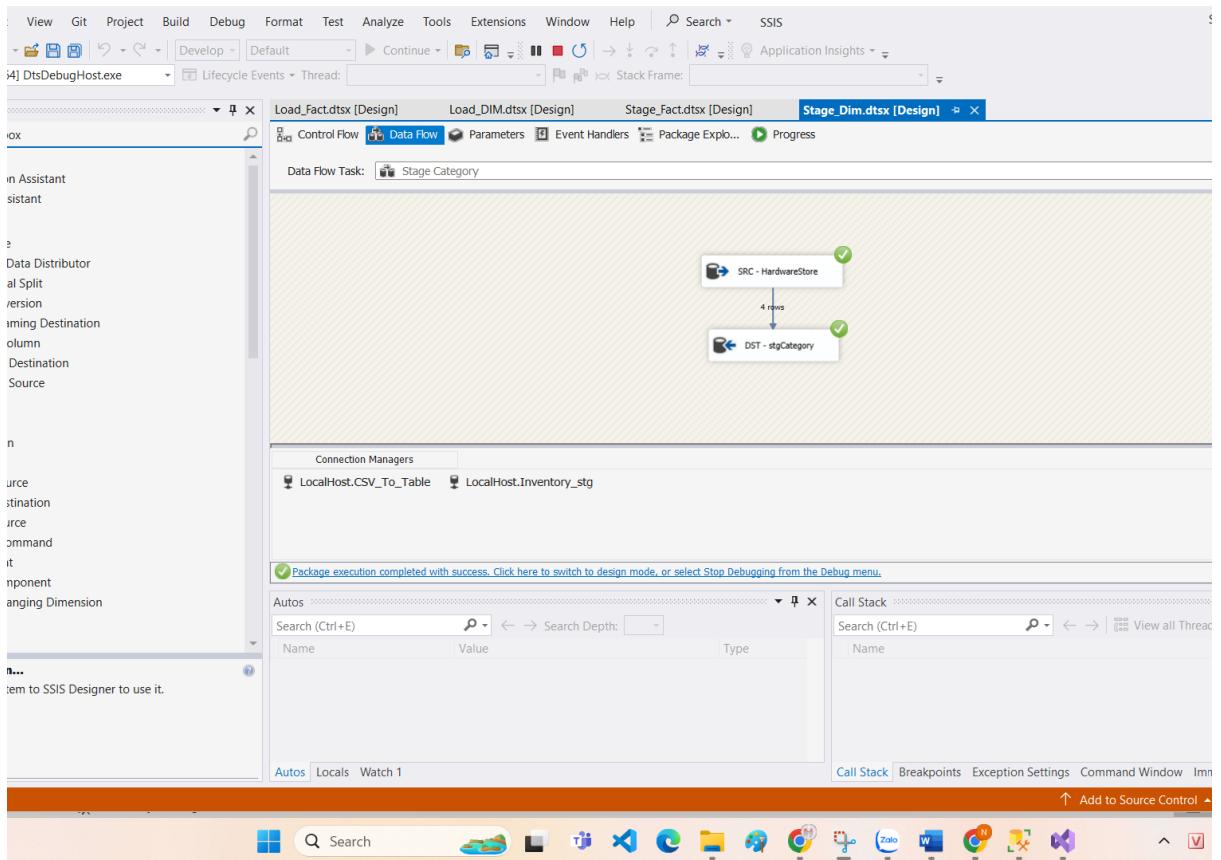
## SCD chọn historical attribute



## Historical attributes options chọn RowIsCurrent



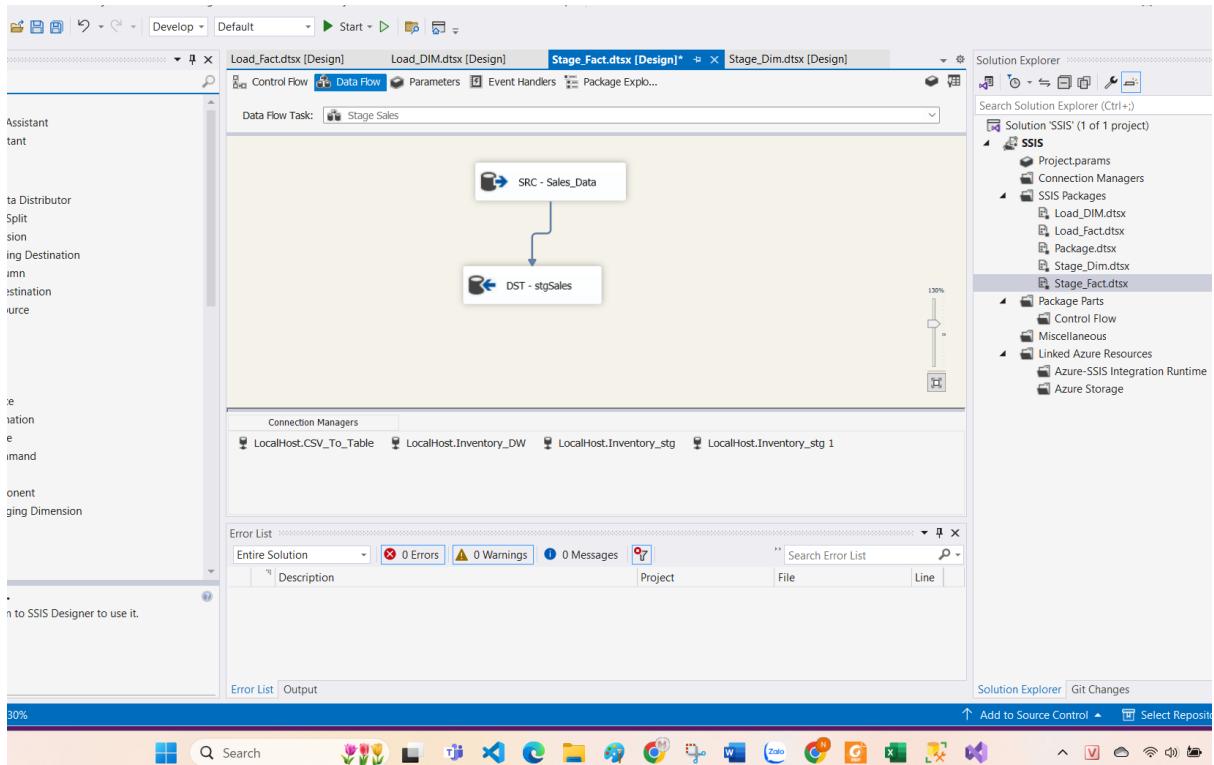
Khởi chạy



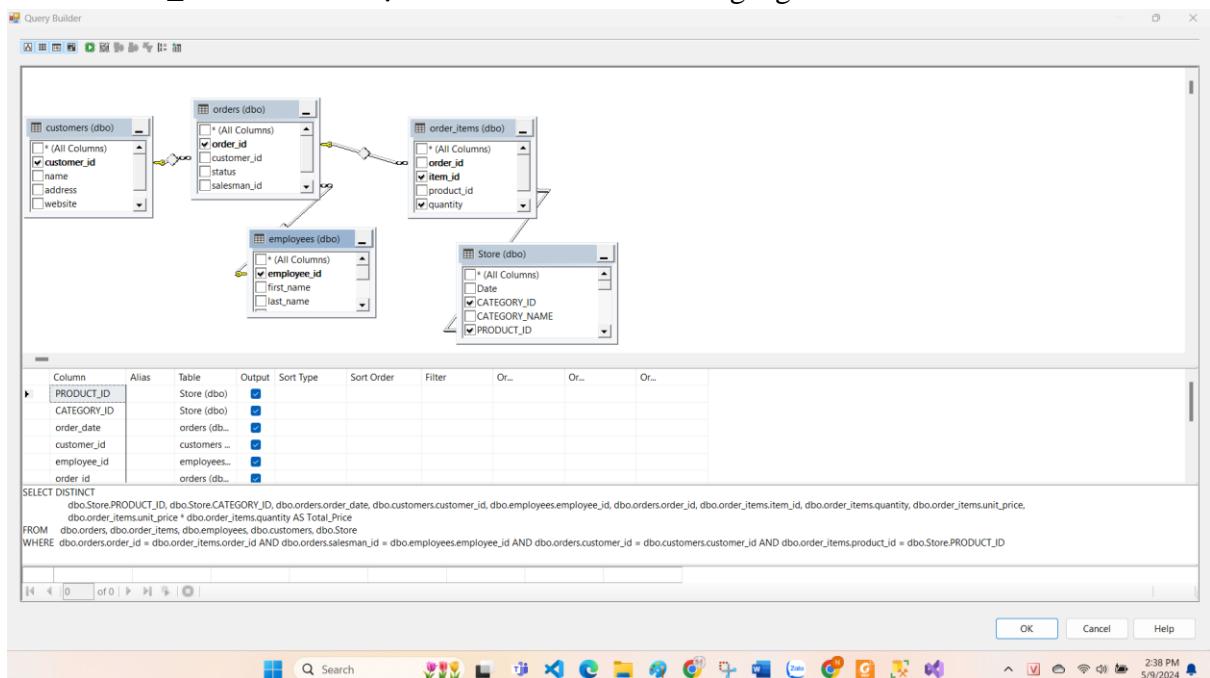
## 3.2. Import dữ liệu vào bảng các bảng fact

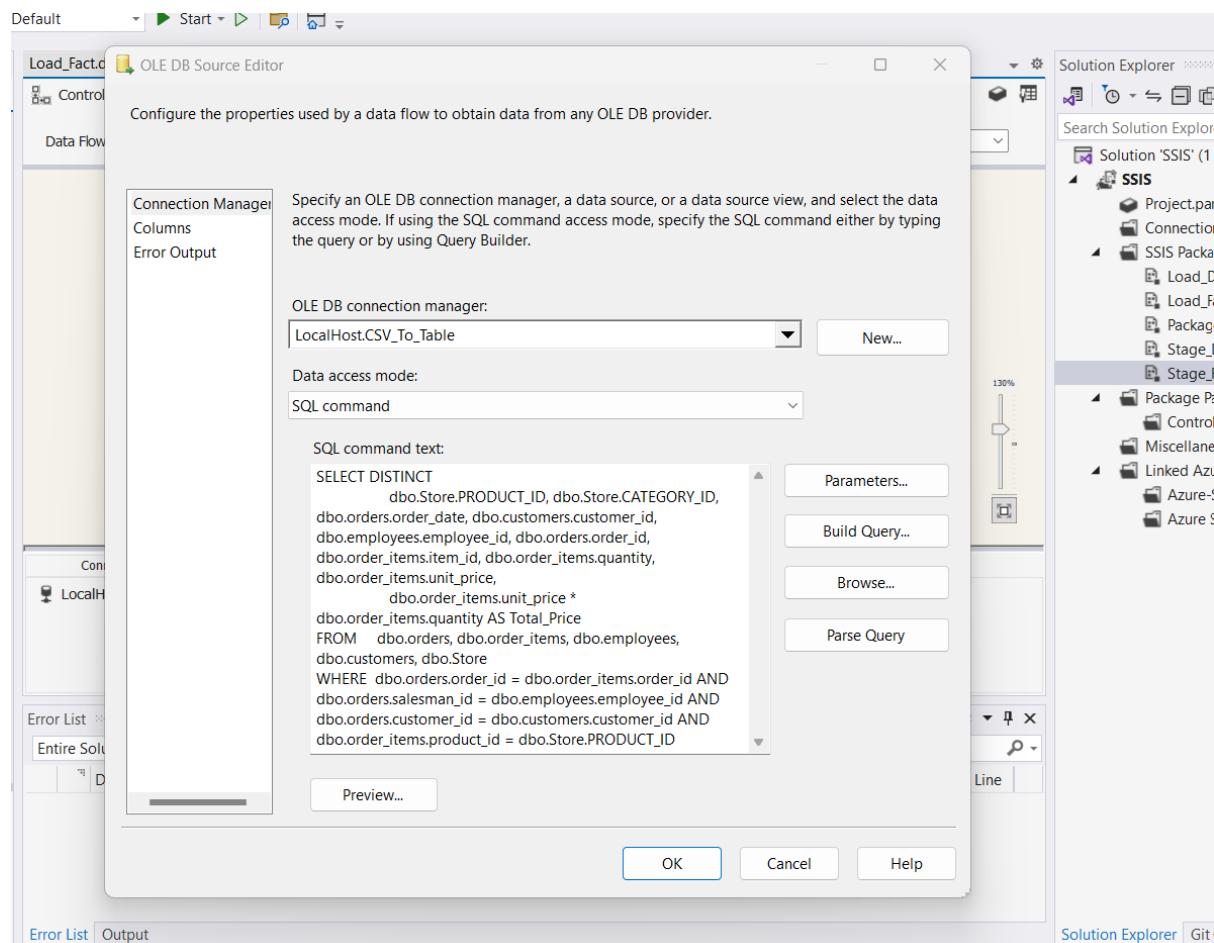
### 3.2.1. Fact Sales

#### a) Load dữ liệu từ nguồn vào stgSales

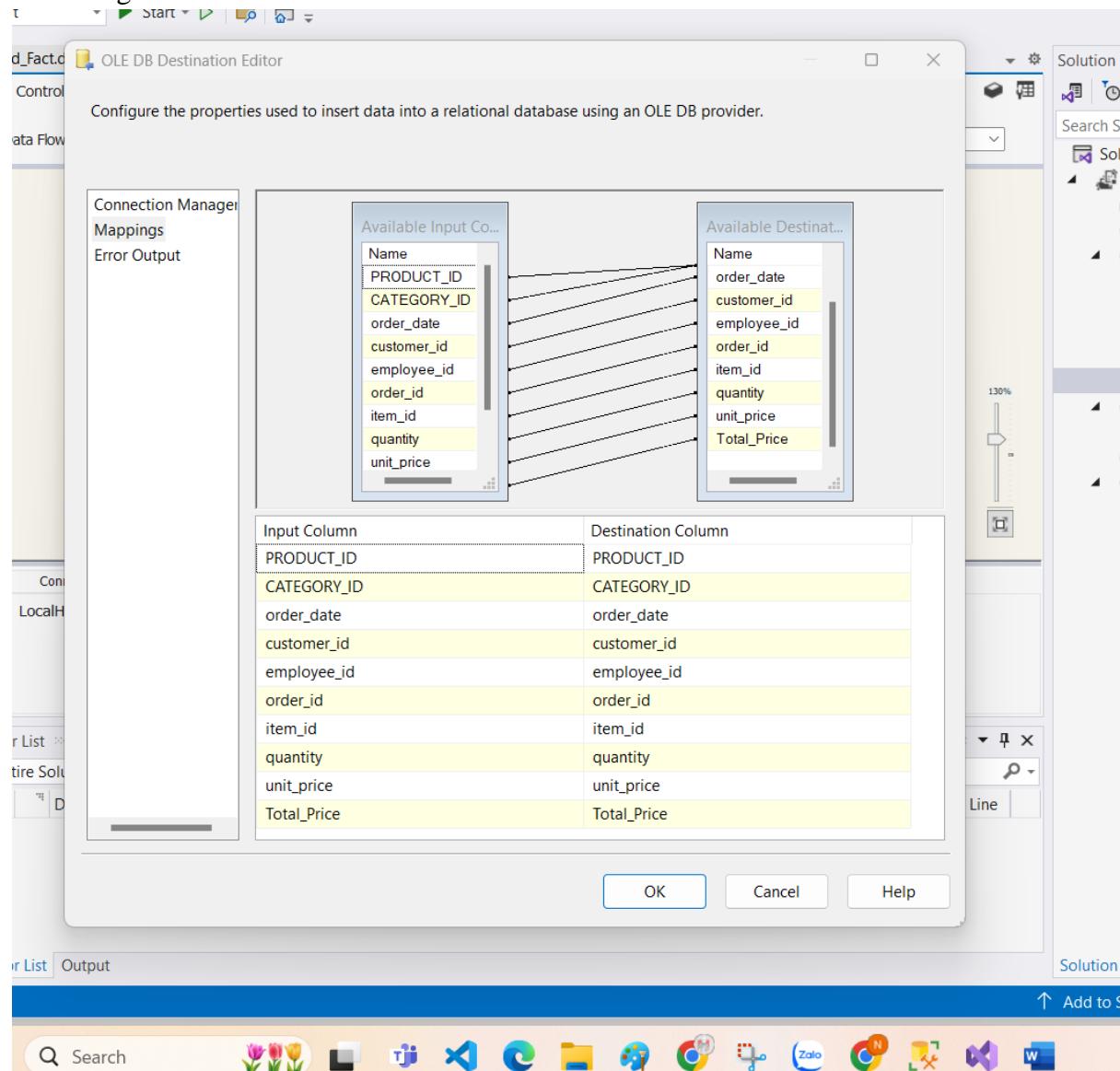


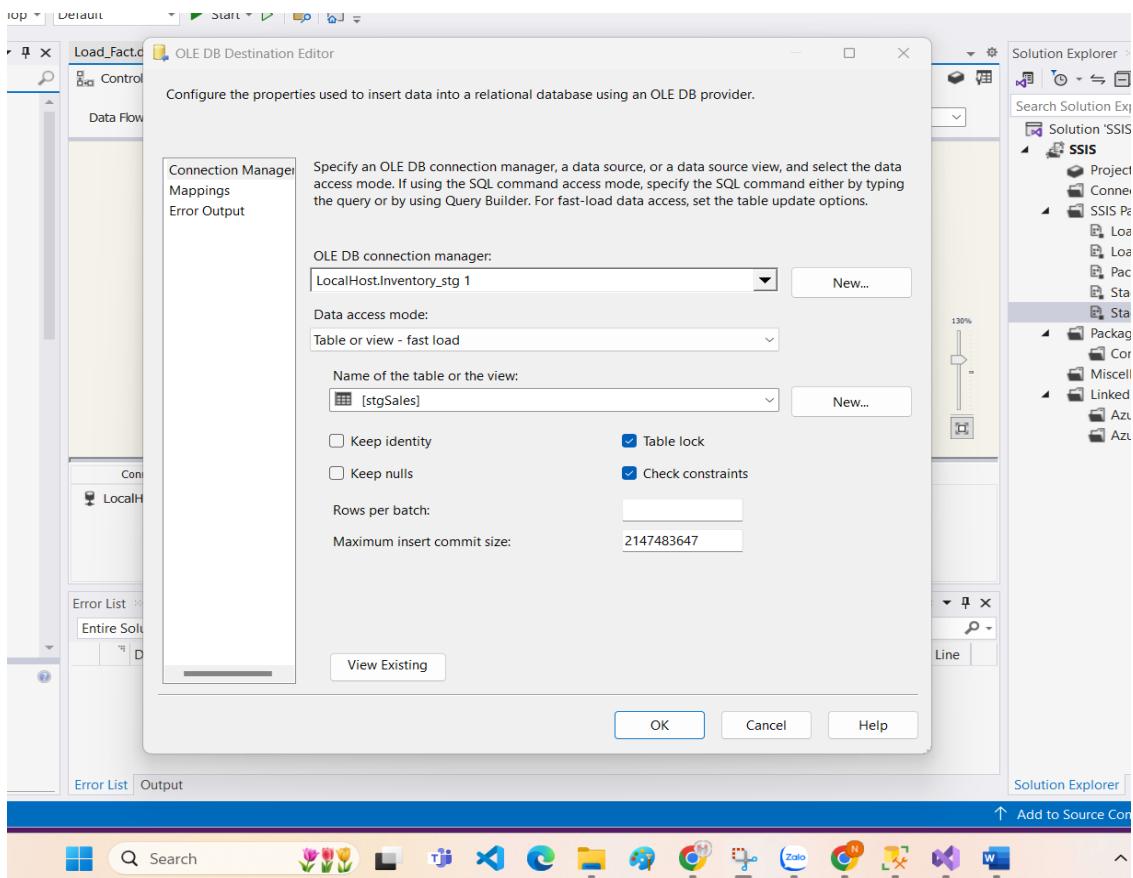
SRC – Sales\_Data chưa dữ liệu cần thiết để đưa vào bảng stgSales



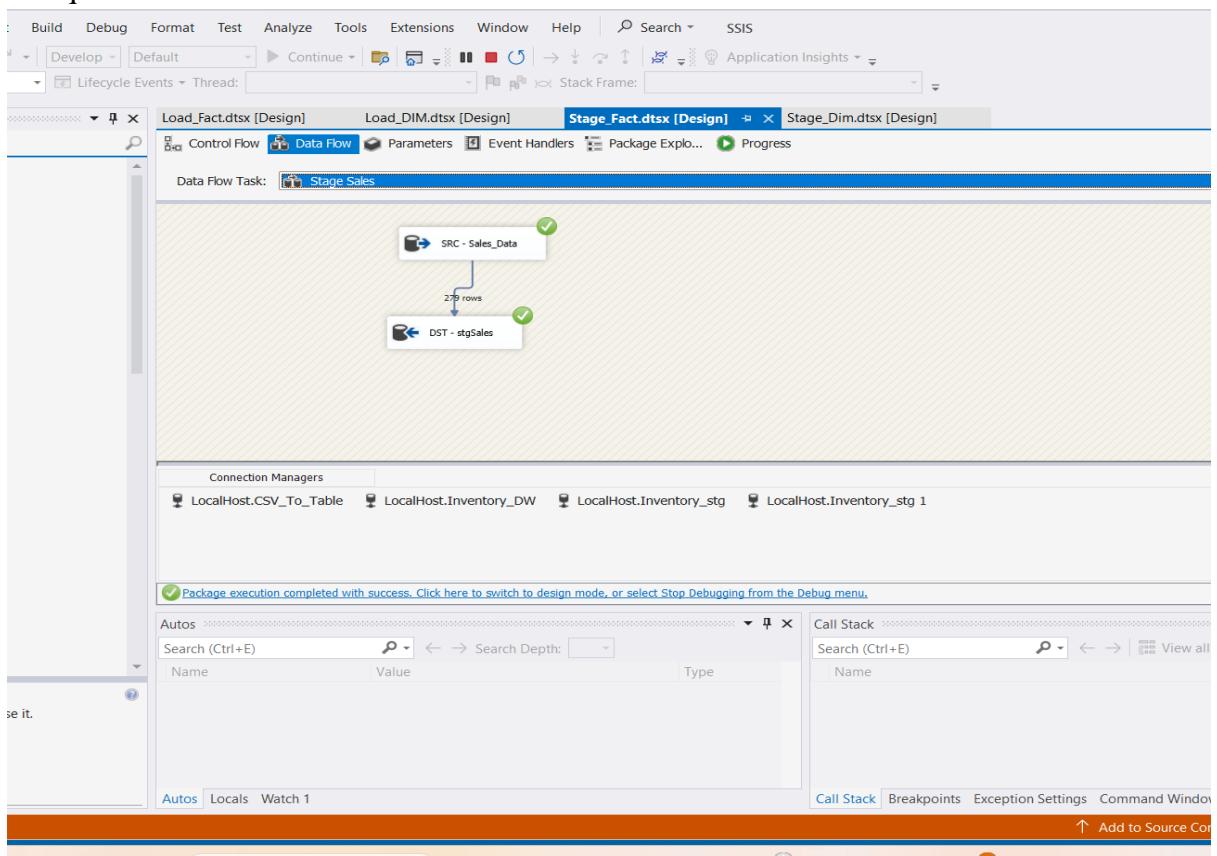


## DST – stgSales

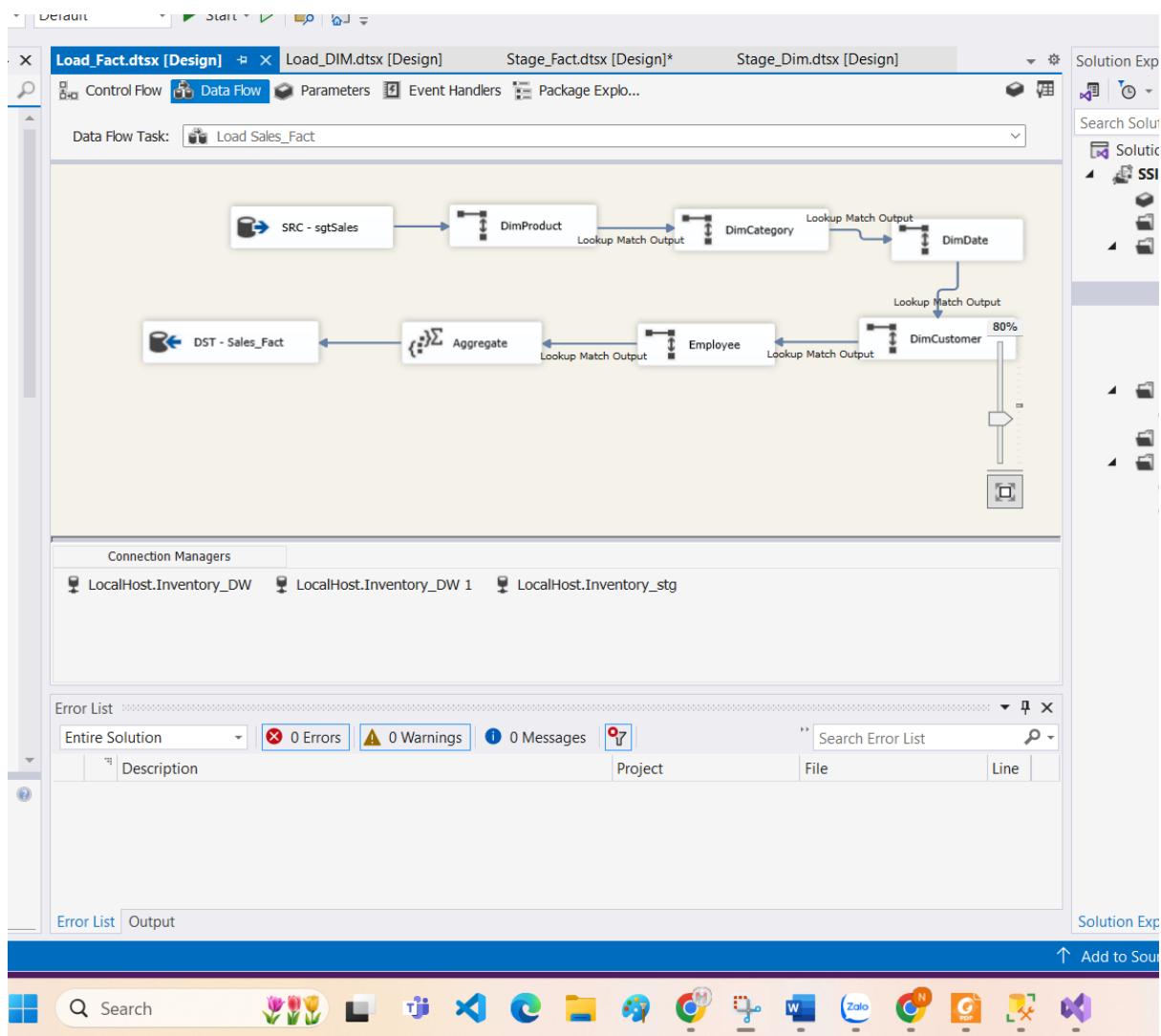




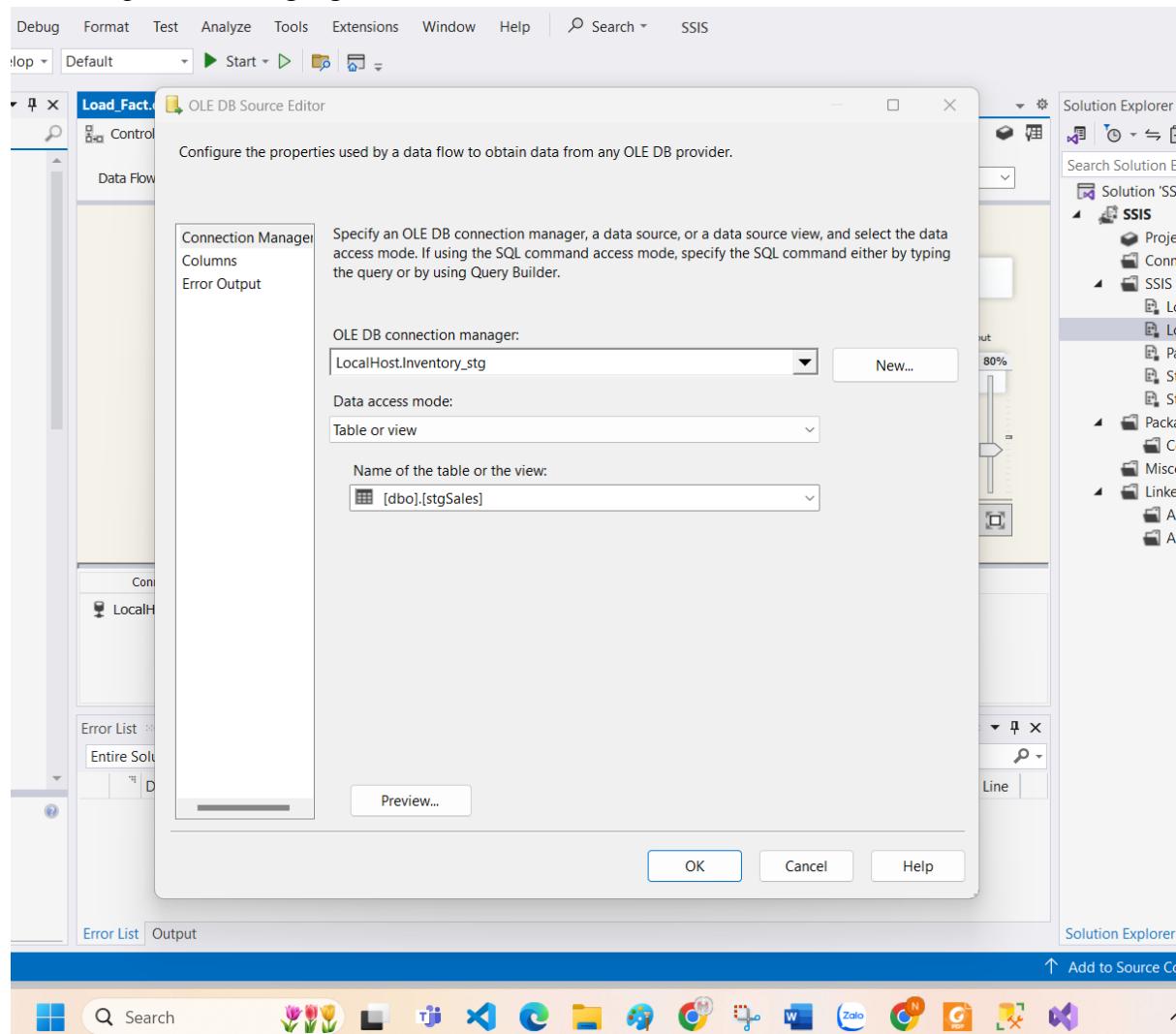
## Kết quả



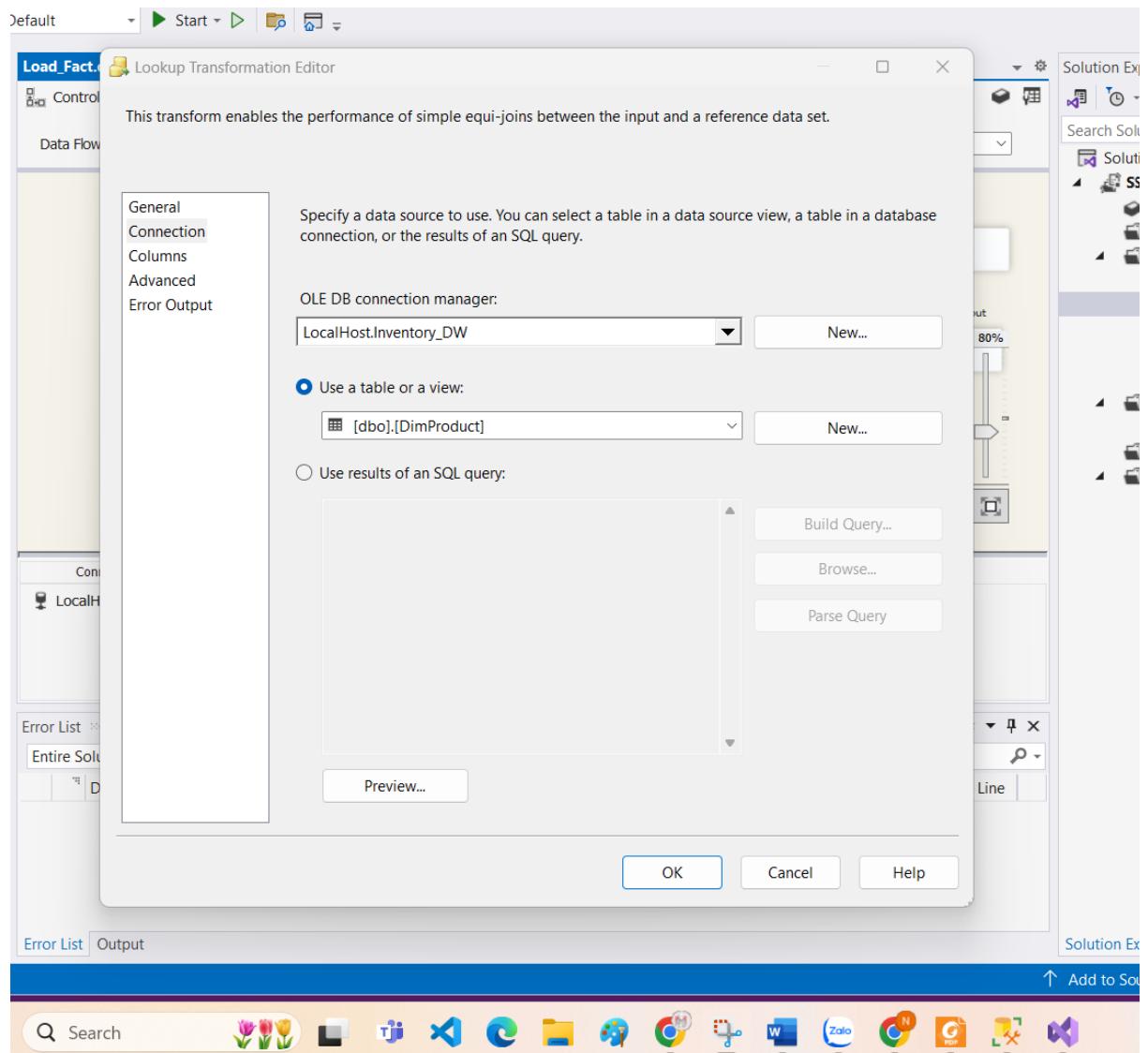
## b) Load dữ liệu vào Sales\_fact

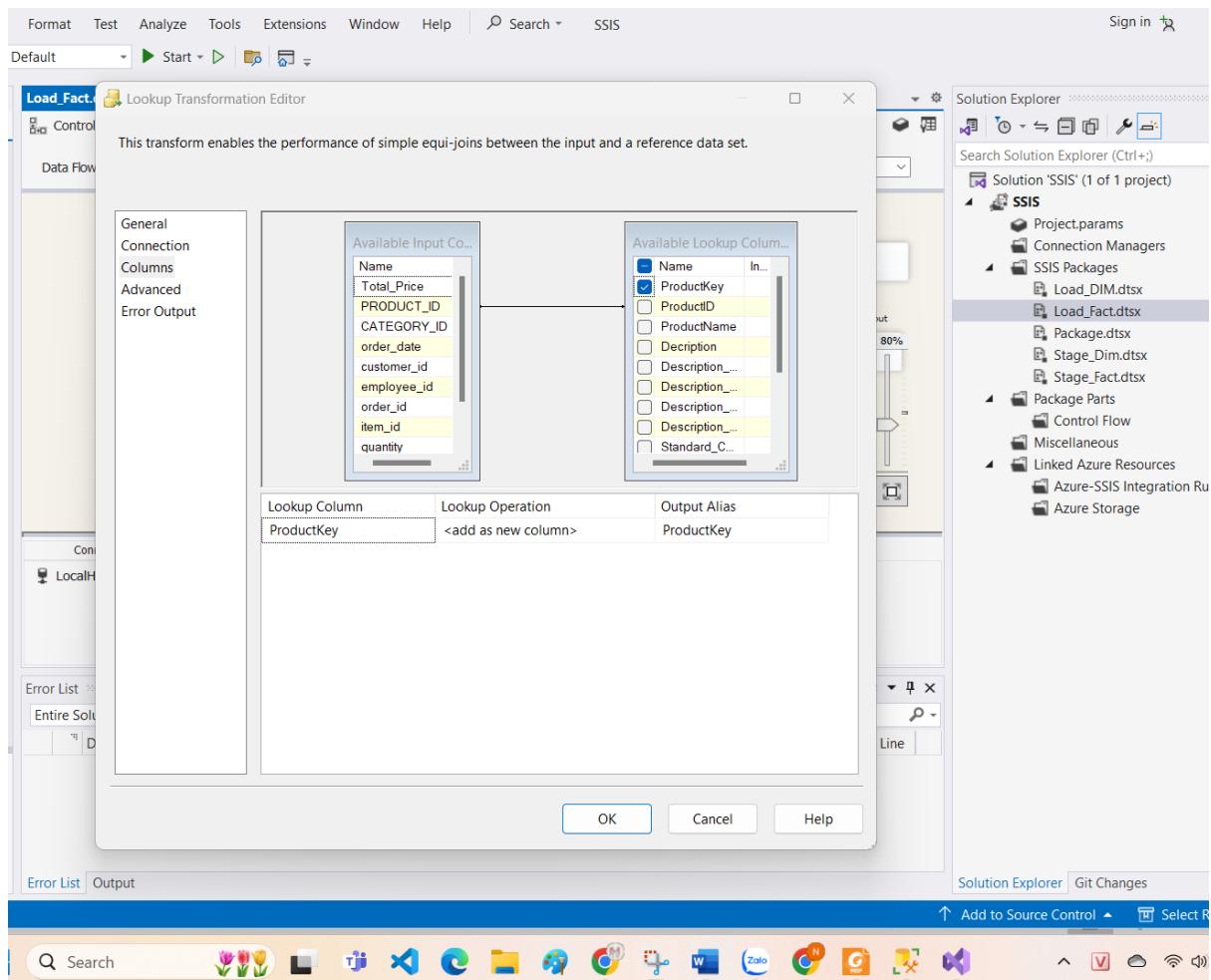


SRC – stgSales là bảng stgSales đã load từ trước

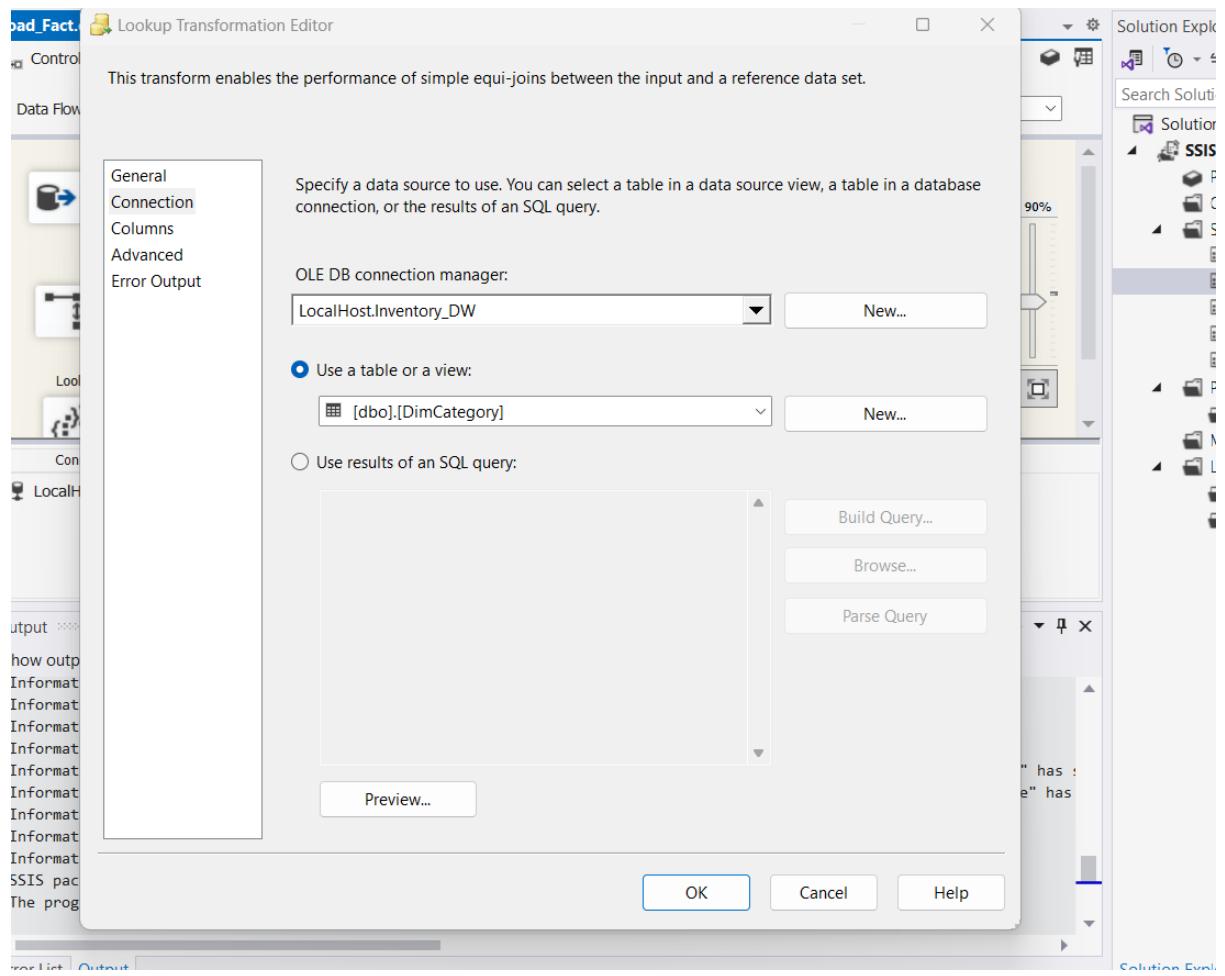


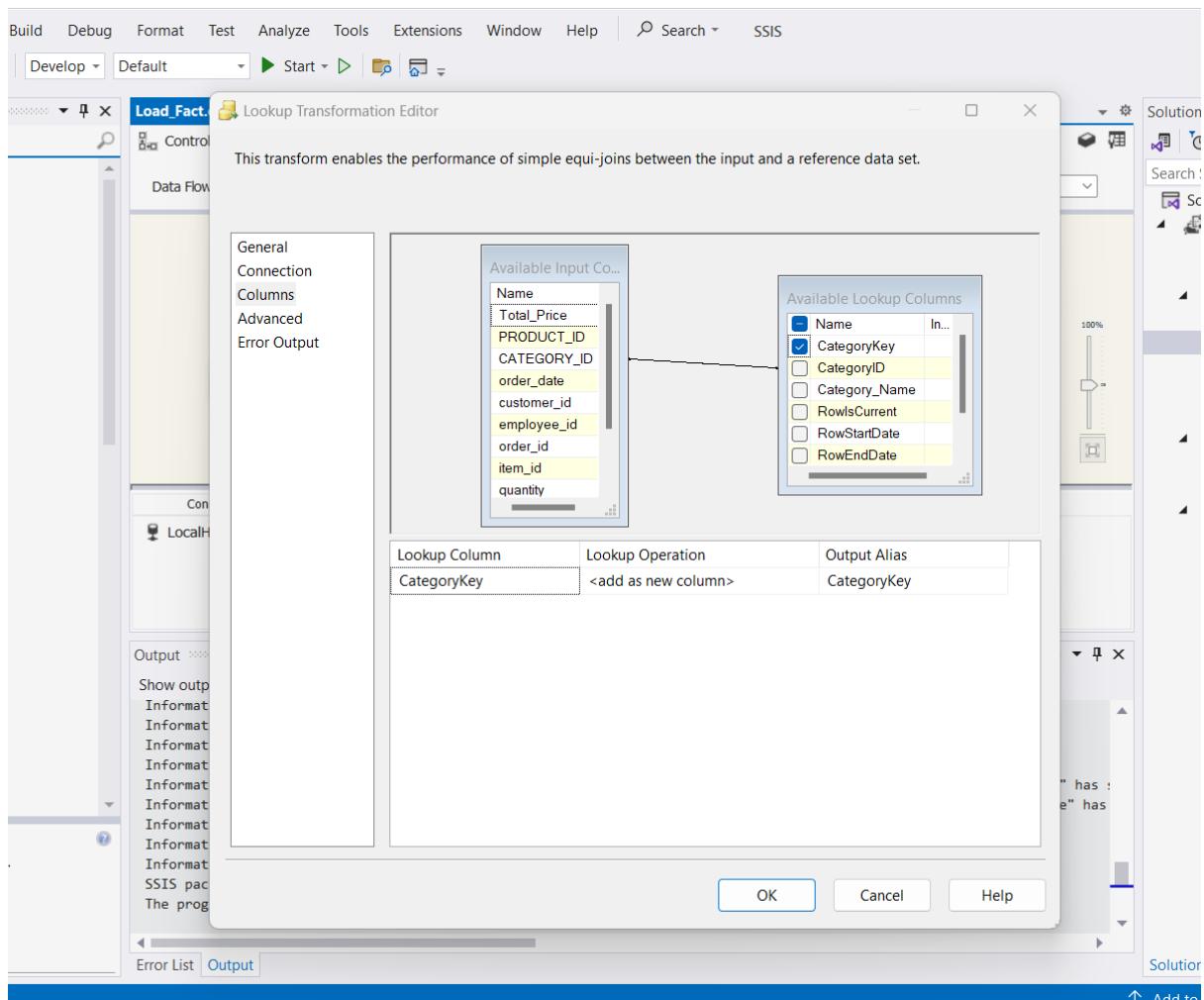
Lookup DimProduct ta sẽ map hai ProductID từ hai bảng lại với nhau và lấy ra thuộc tính ProductKey trong bảng DimProduct



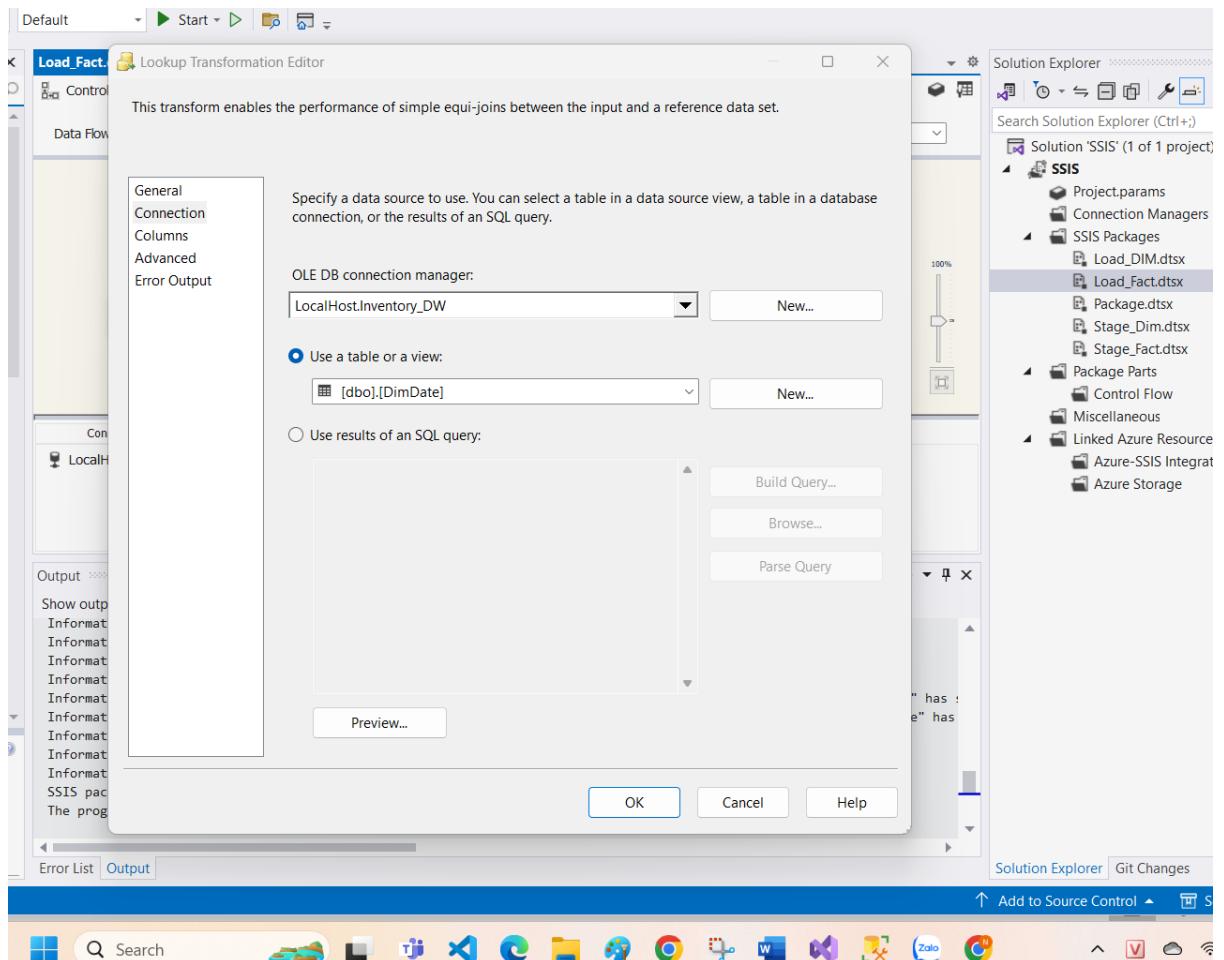


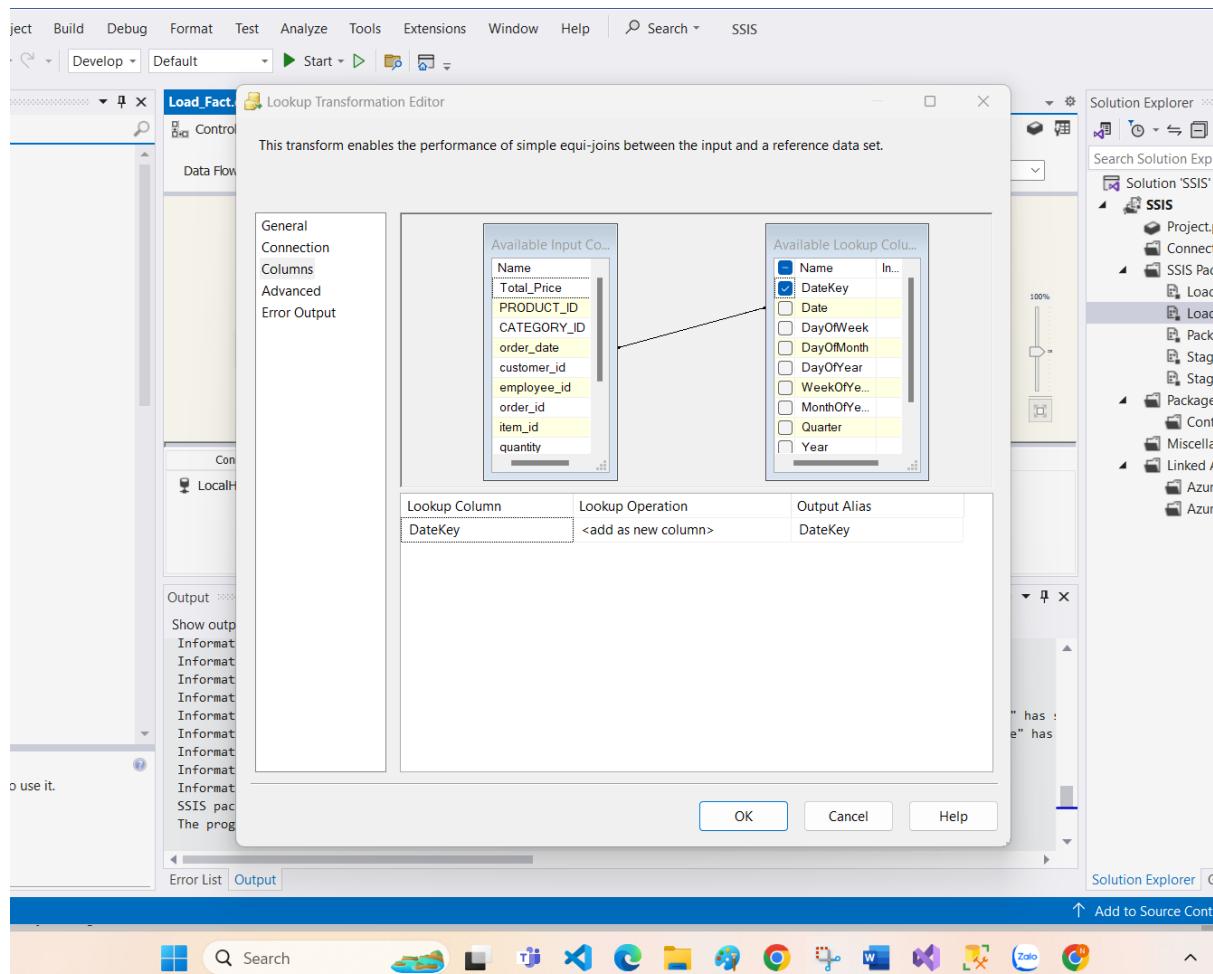
Lookup DimCategory ta sẽ map 2 thuộc tính CategoryId trong 2 bảng lại với nhau để lấy ra thuộc tính CategoryKey trong bảng DimCategory



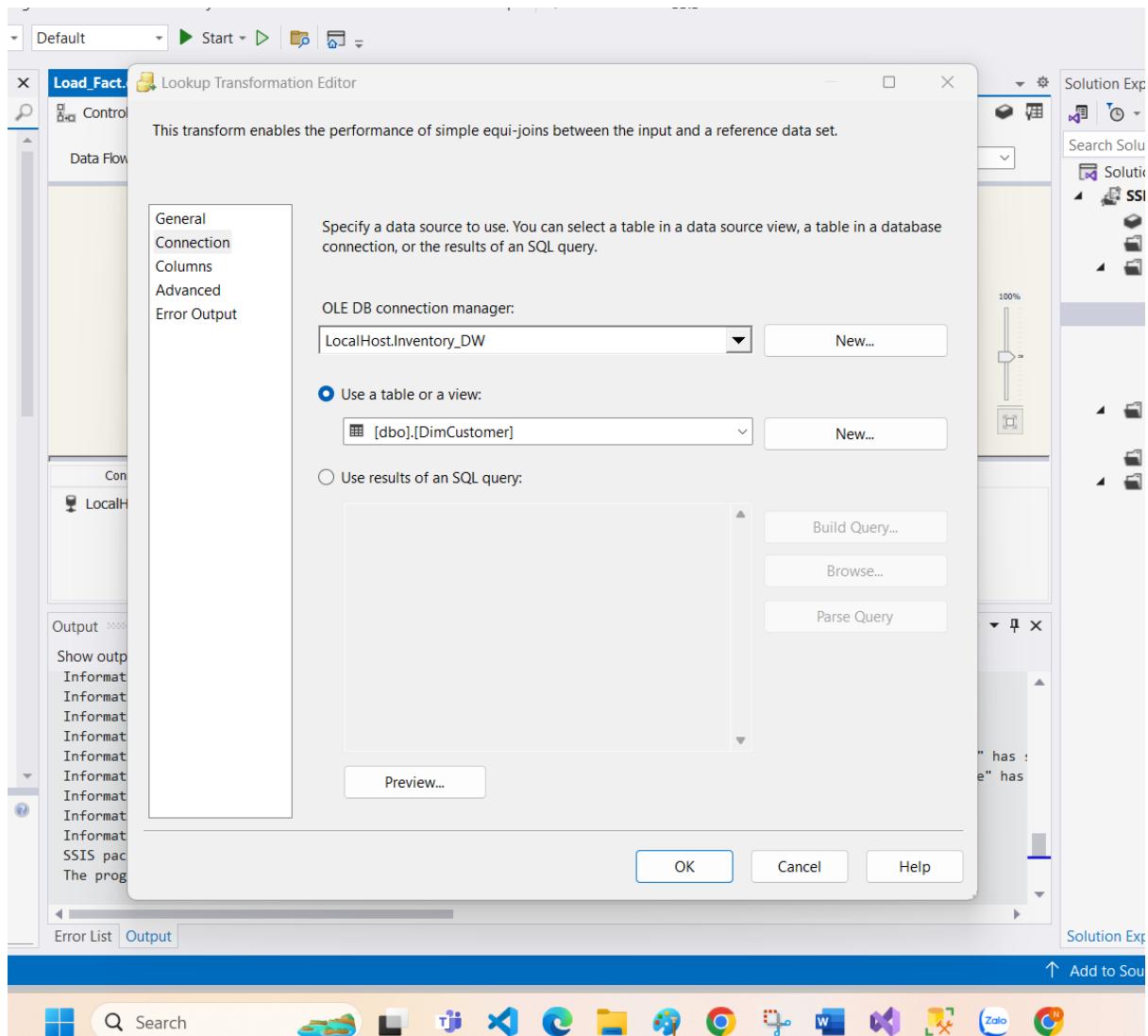


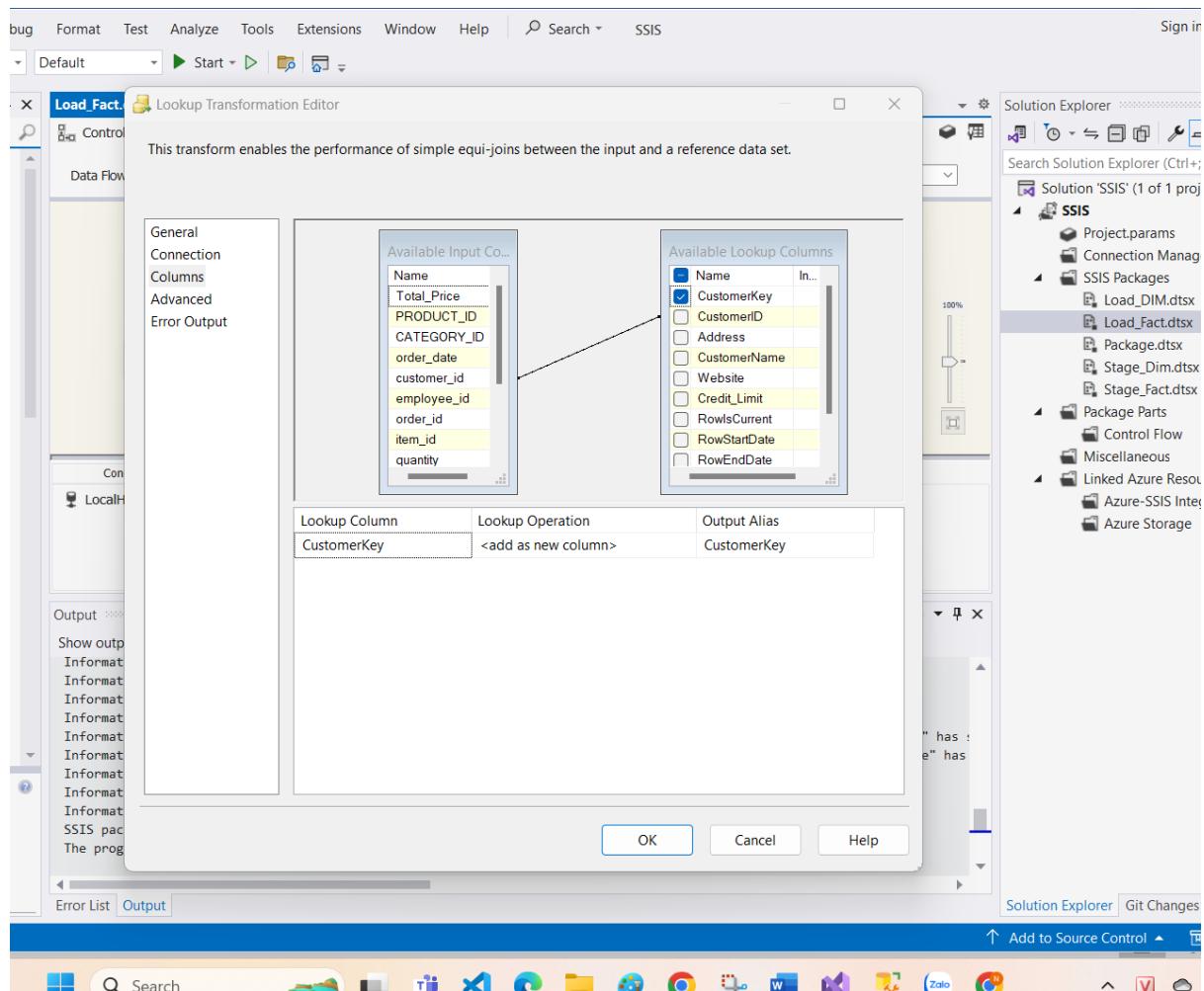
Lookup DimDate ta sẽ map cột order\_date ở bảng stgSales với cột Date trong bảng DimDate để lấy ra DateKey



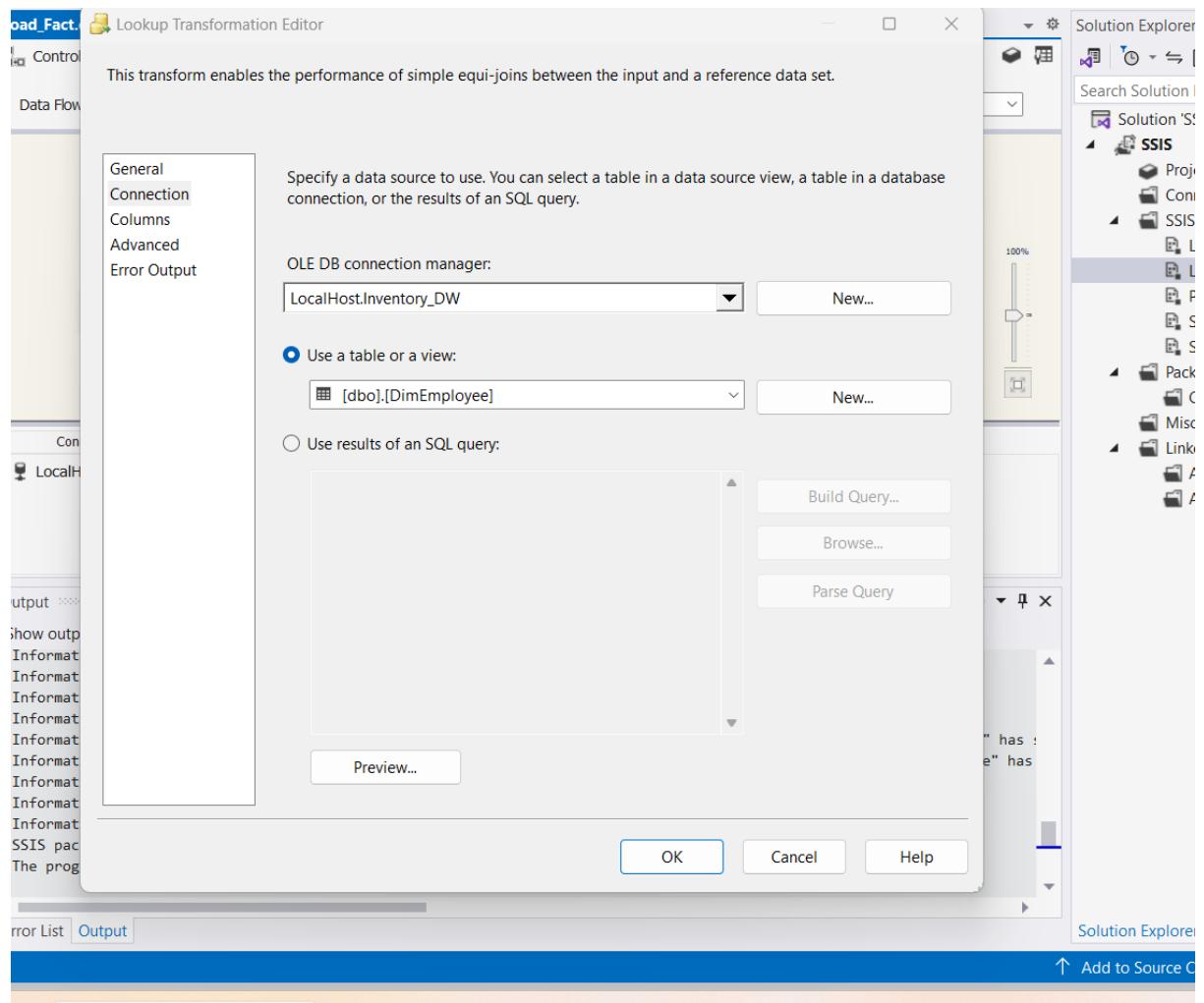


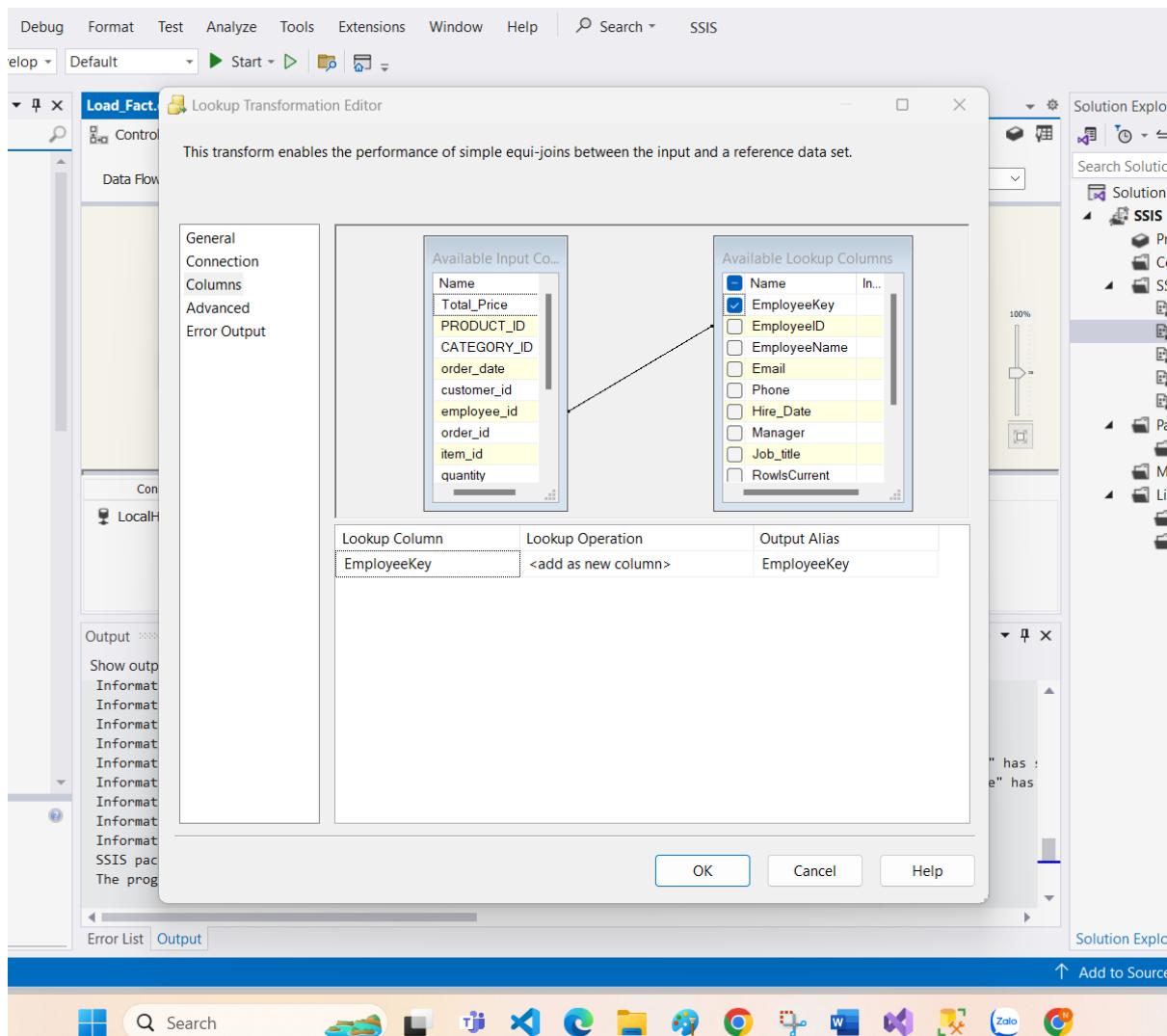
Lookup DimCustomer ta sẽ map cột customer\_id trong bảng stgSales với cột CustomerID trong bảng DimCustomer để có thể lấy ra cột CustomerKey



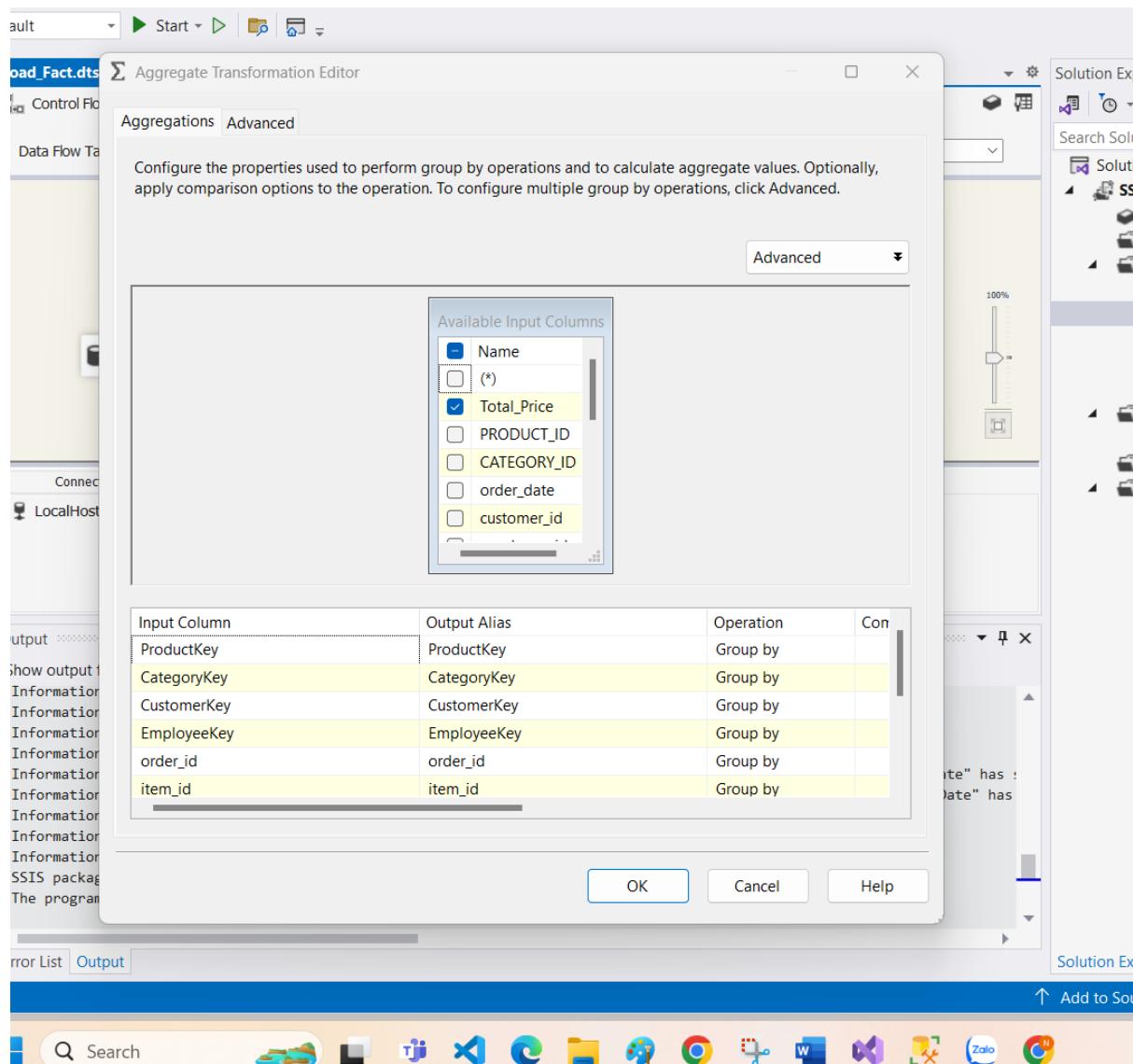


Lookup Dim Employee ta sẽ map cột employee\_id trong bảng stgSales với cột EmployeeID trong bảng DimEmployee

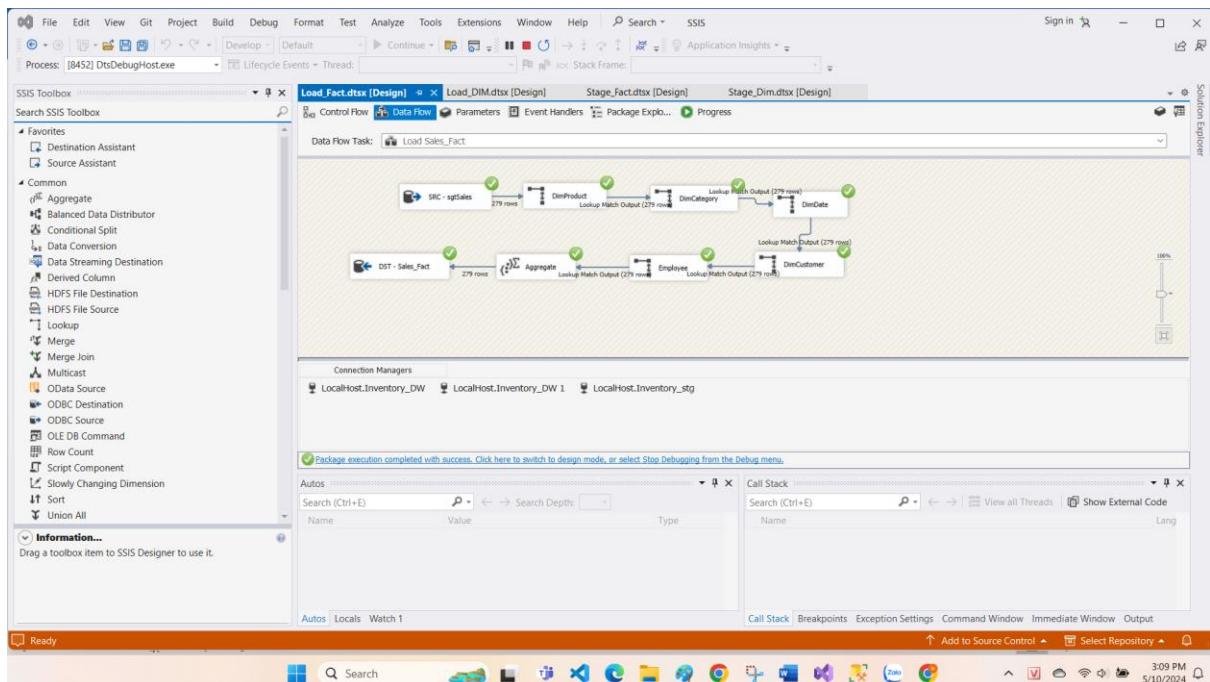




**Aggregate** Ta sẽ lấy những cột cần thiết để đưa vào bảng fact

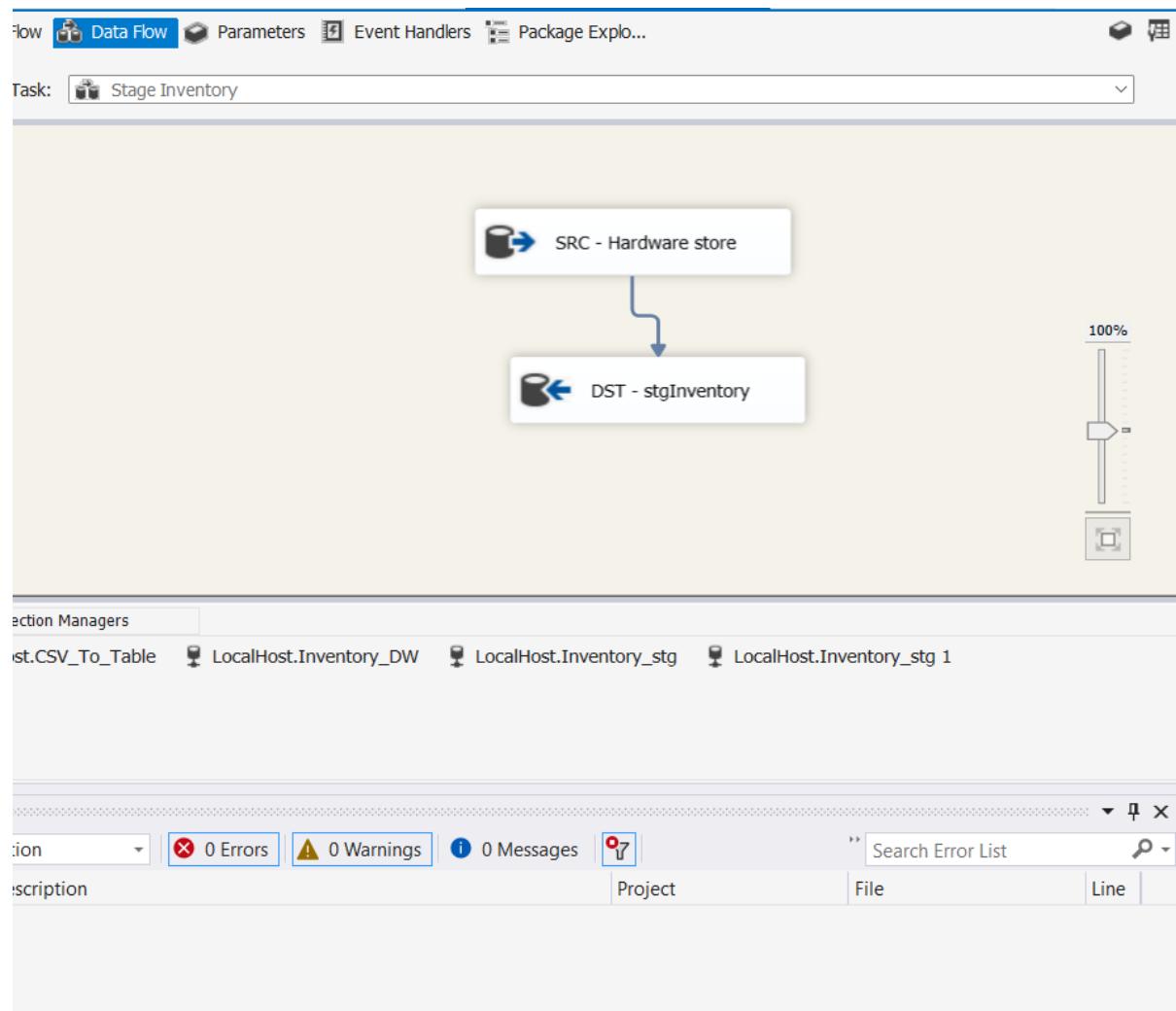


Kết quả



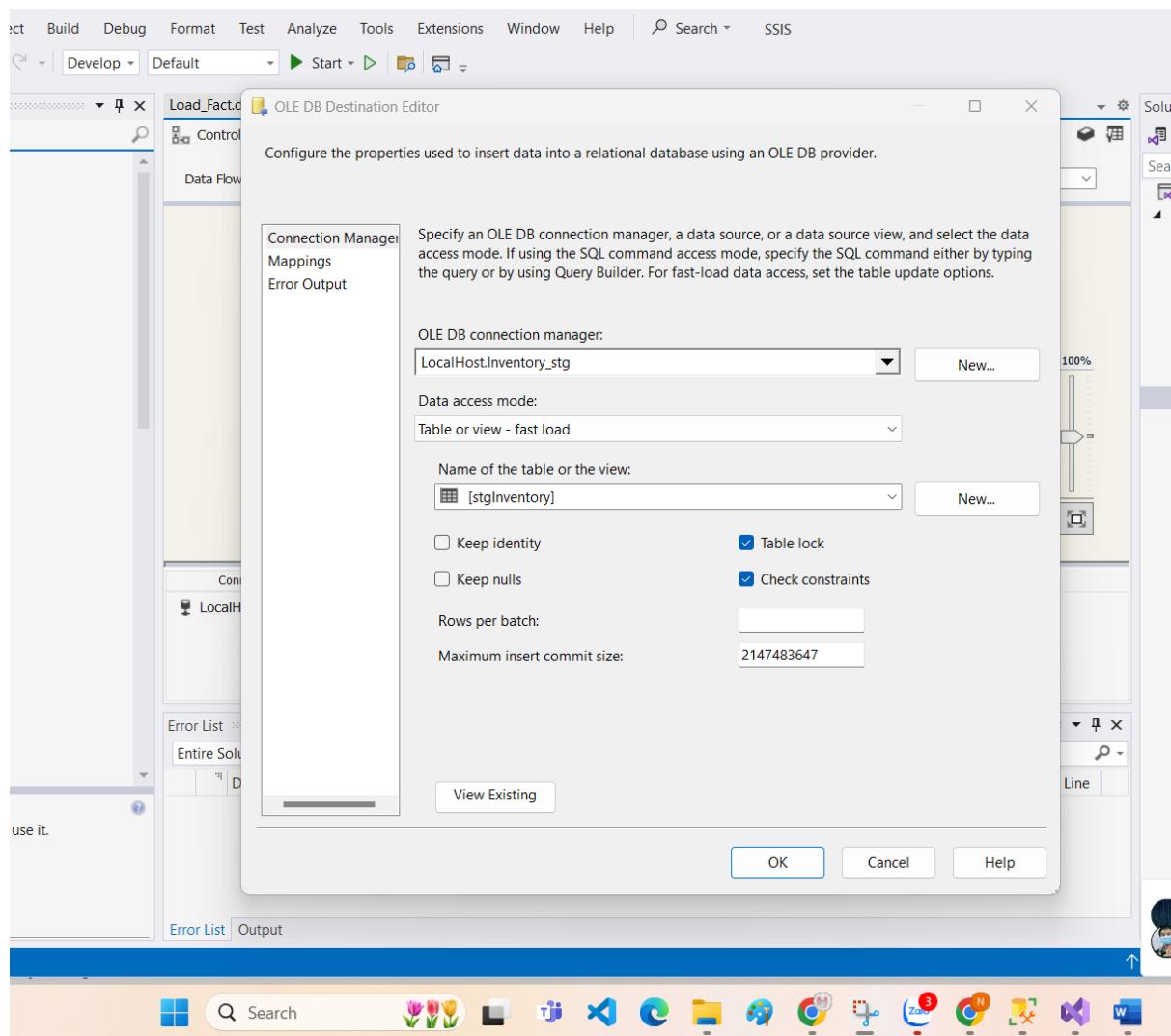
### 3.2.2. Fact Product Inventory

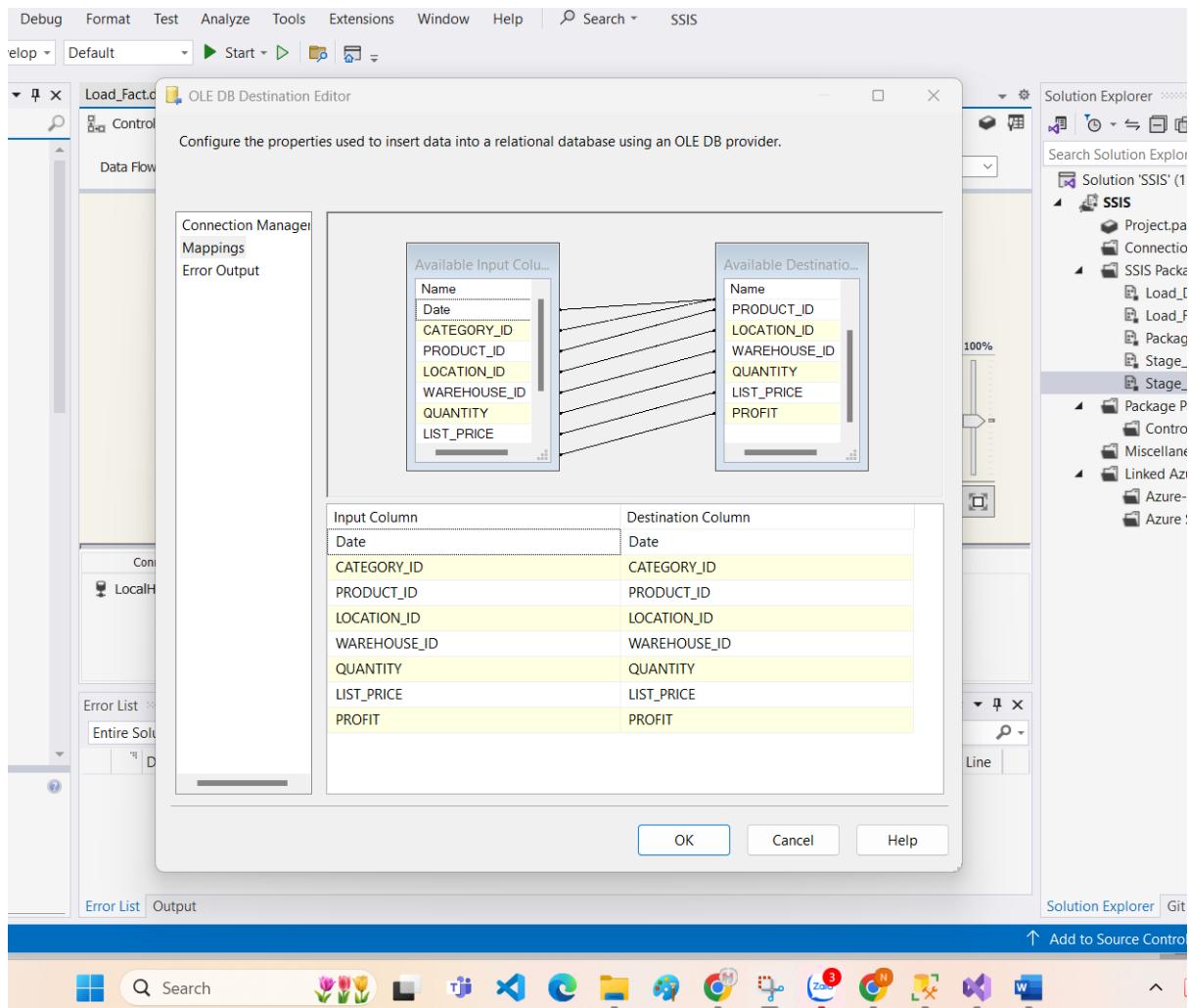
#### a) Load dữ liệu từ nguồn vào stgInventory



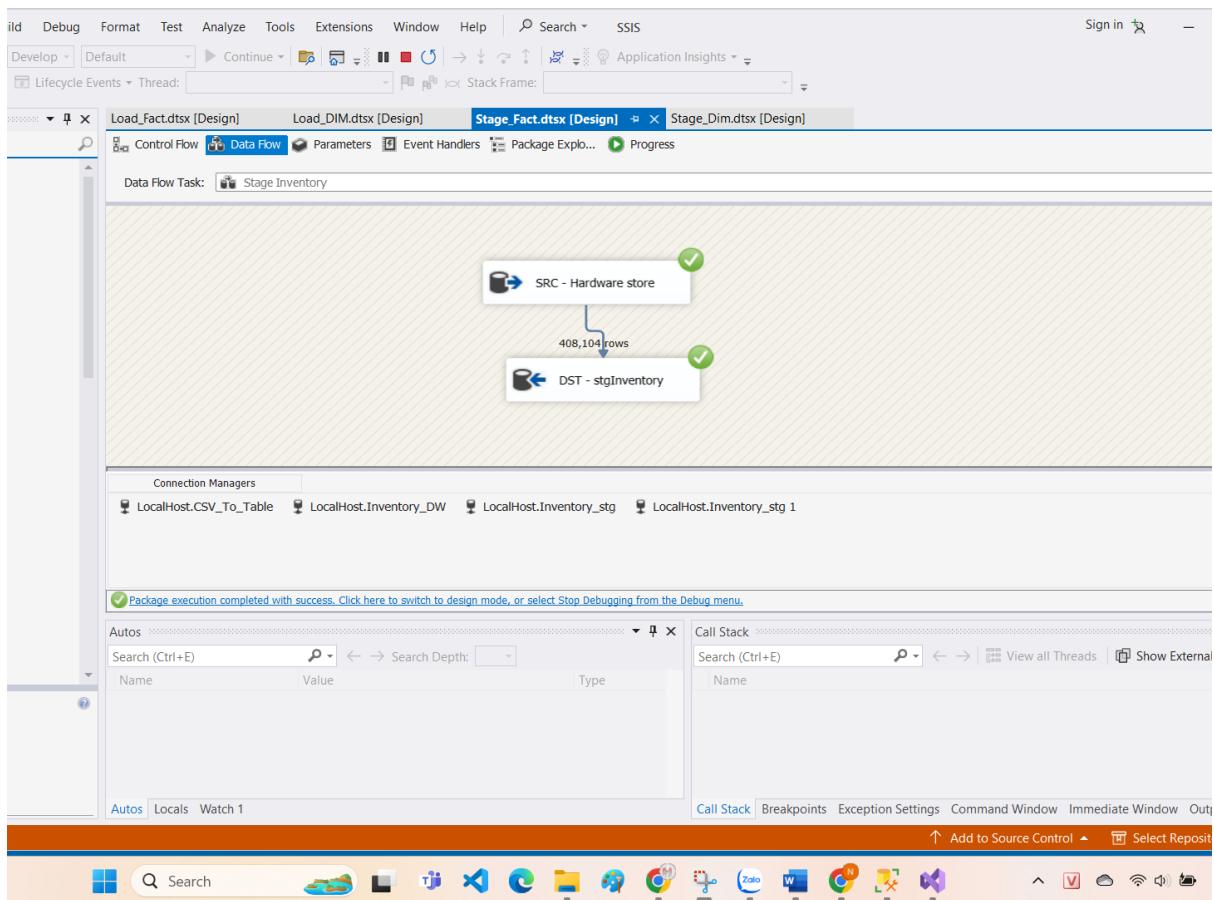
SRC – Hardware store là dữ liệu nguồn ta lấy những thuộc tính cần thiết để đưa vào bảng stgInventory

DST – stgInventory là bảng stgInventory

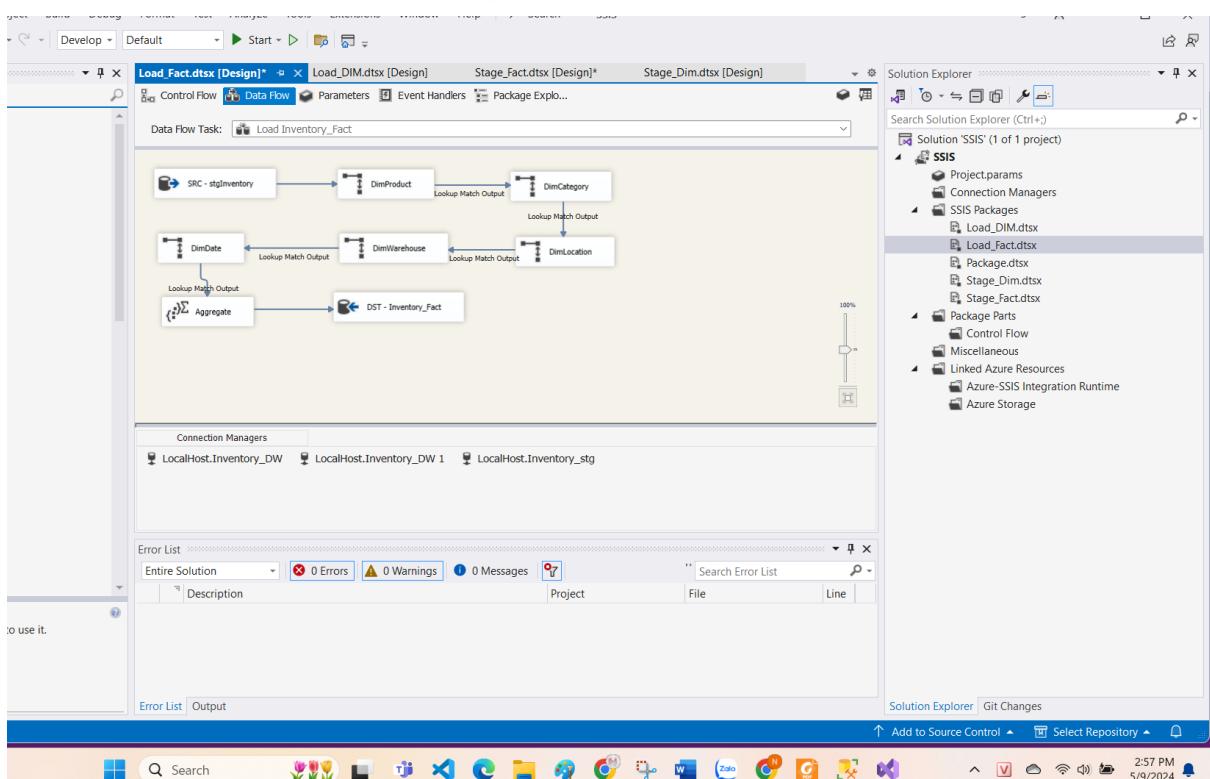




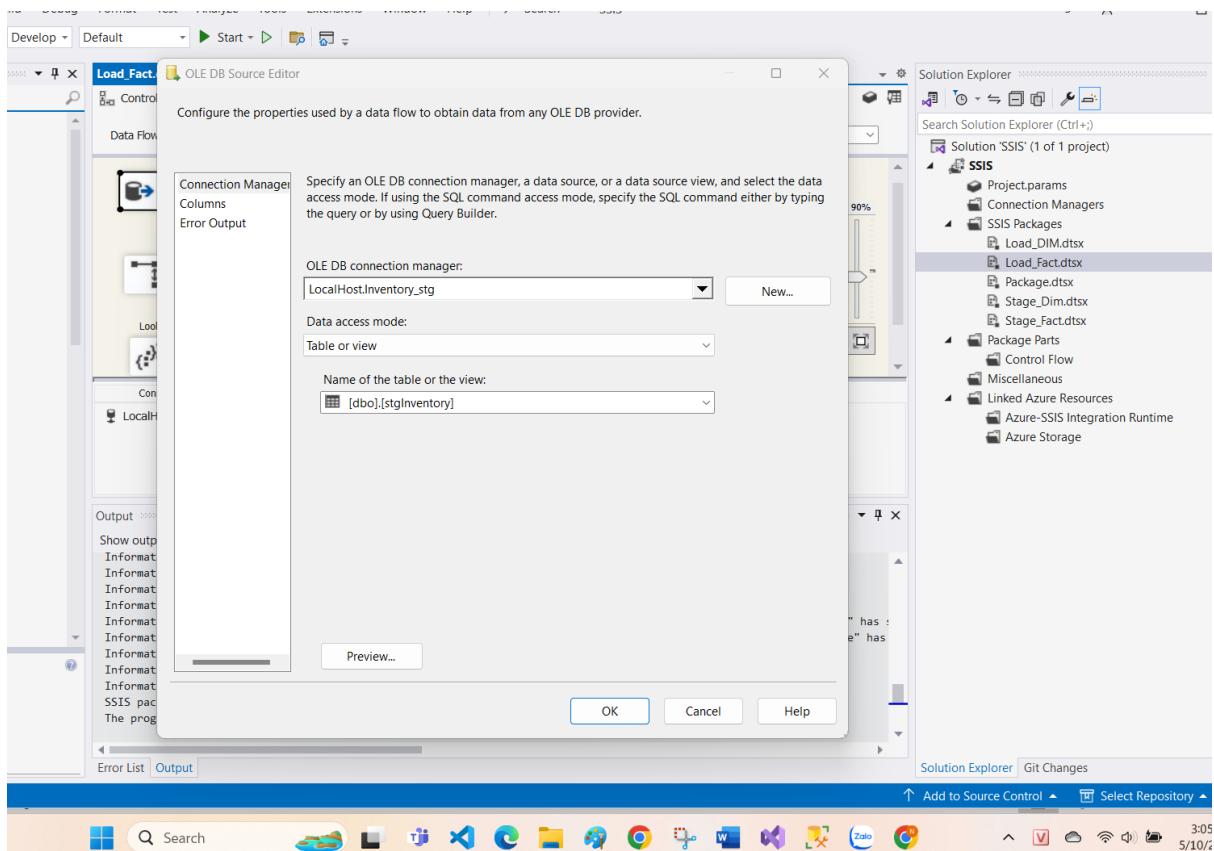
Kết quả



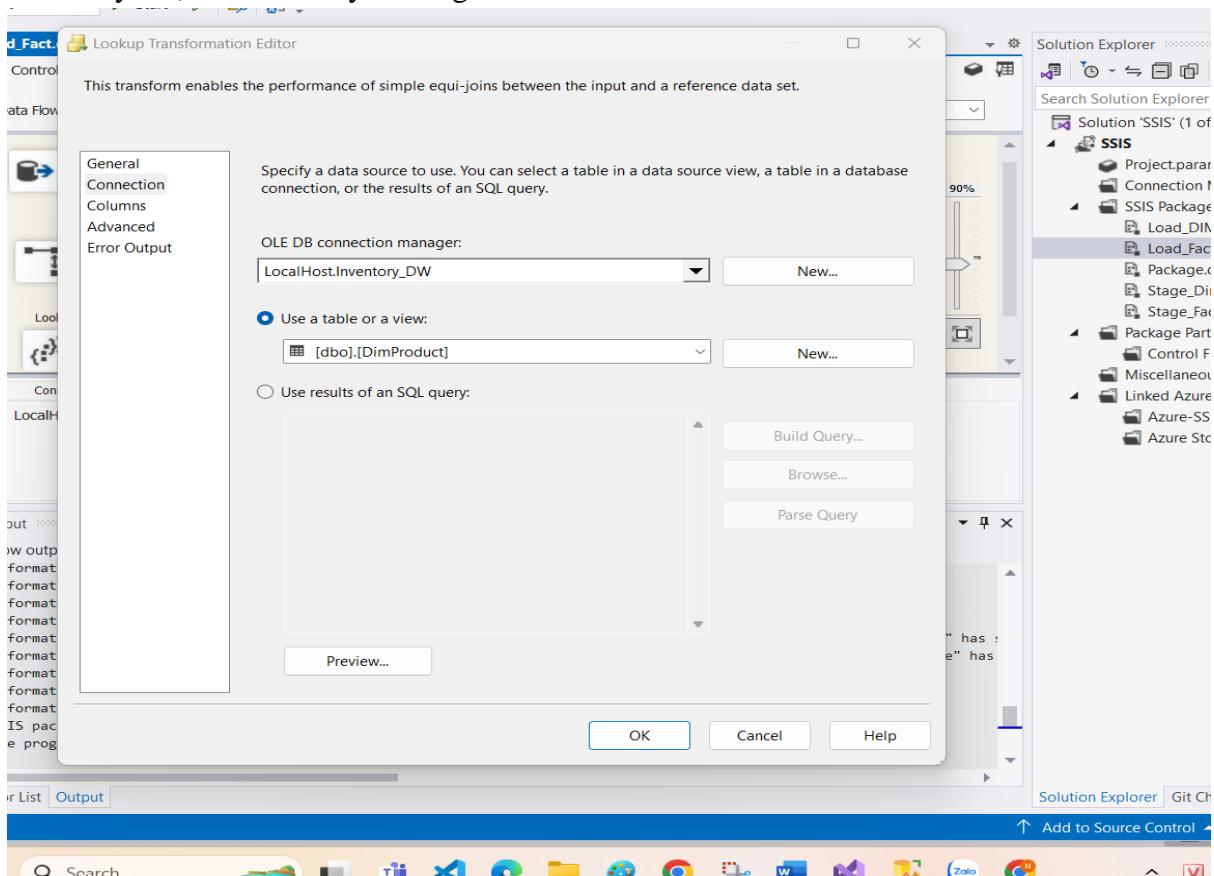
### b) Load dữ liệu vào Inventory\_fact

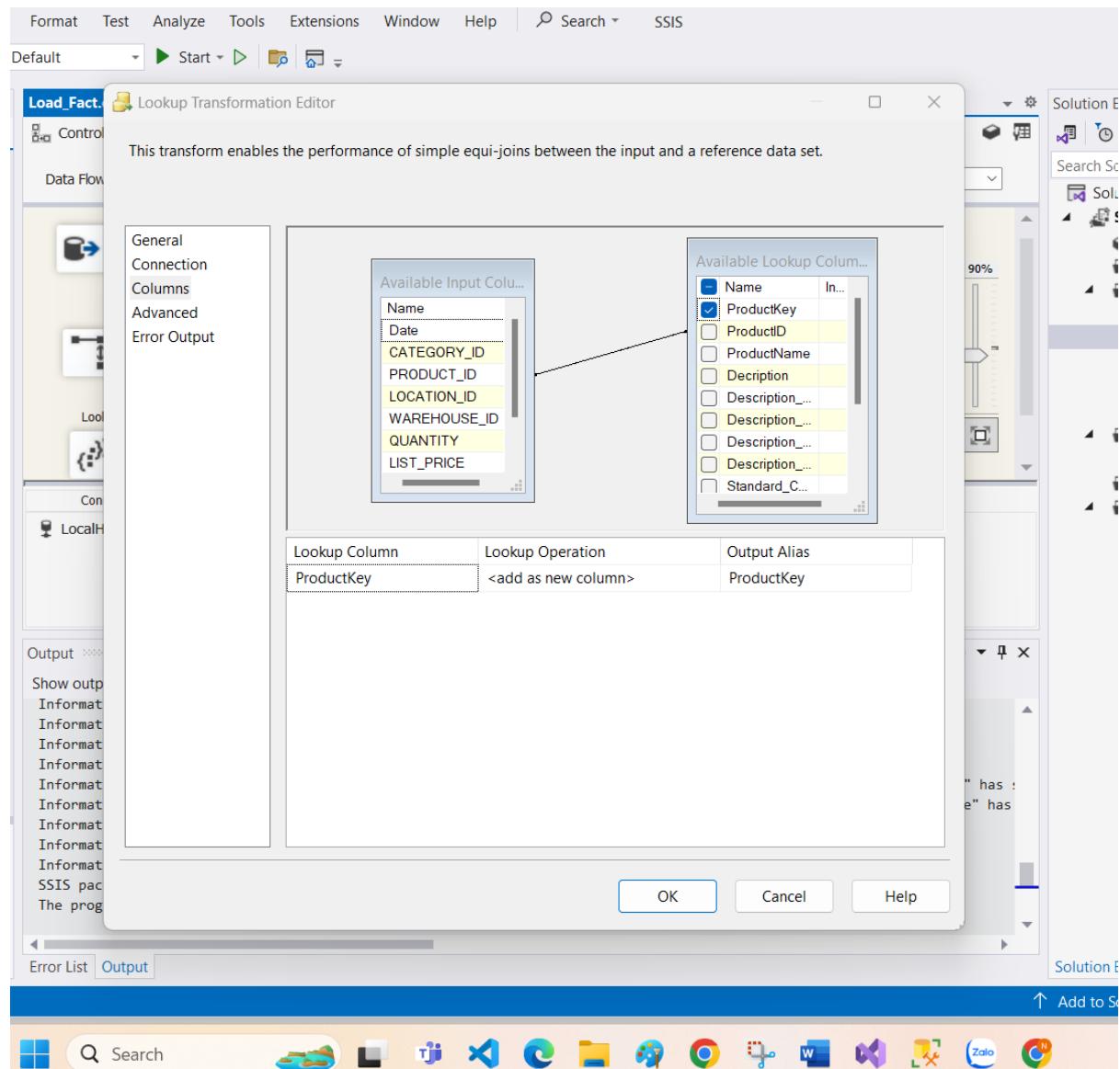


SRC – stgInventory là bảng stgInventory

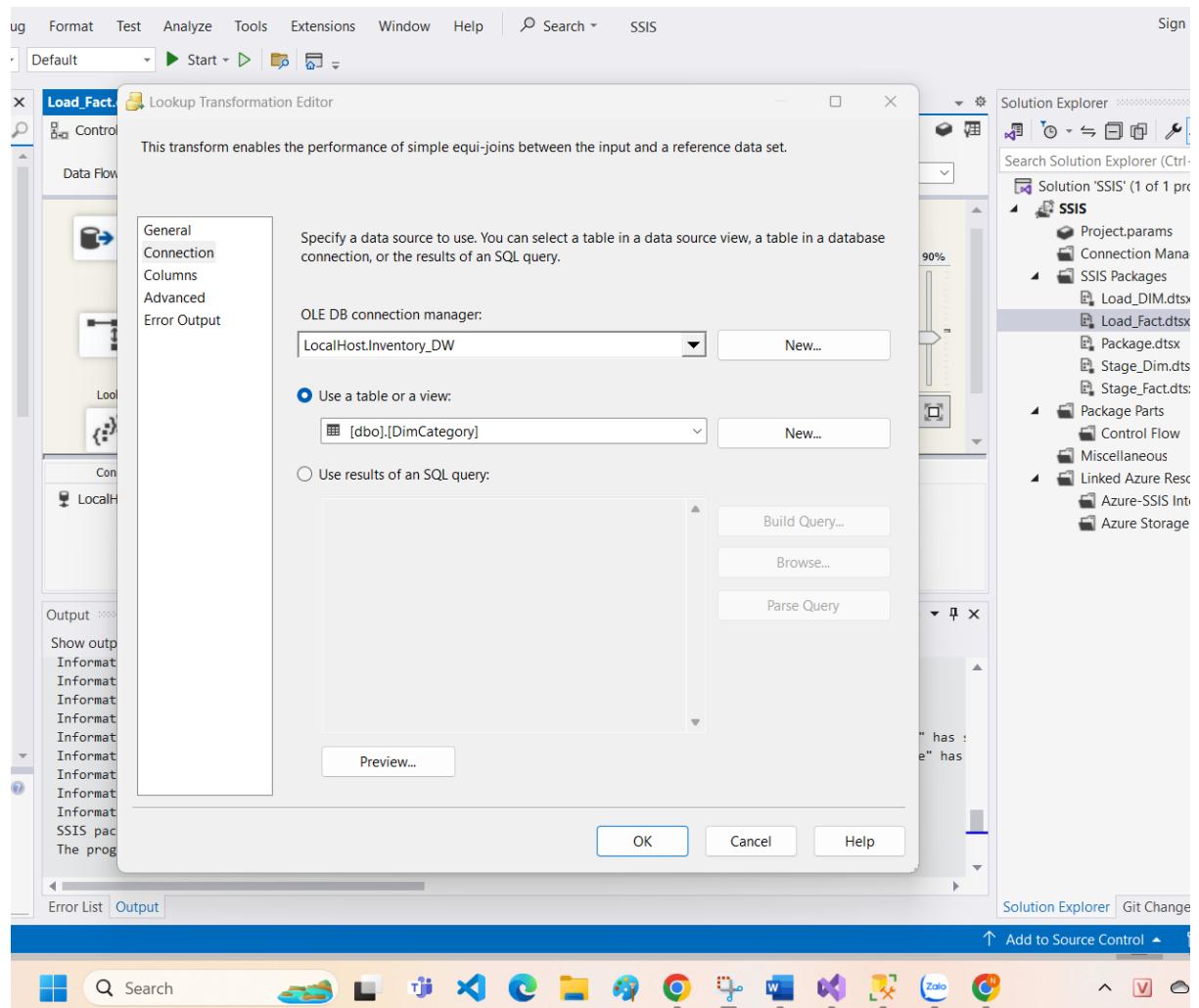


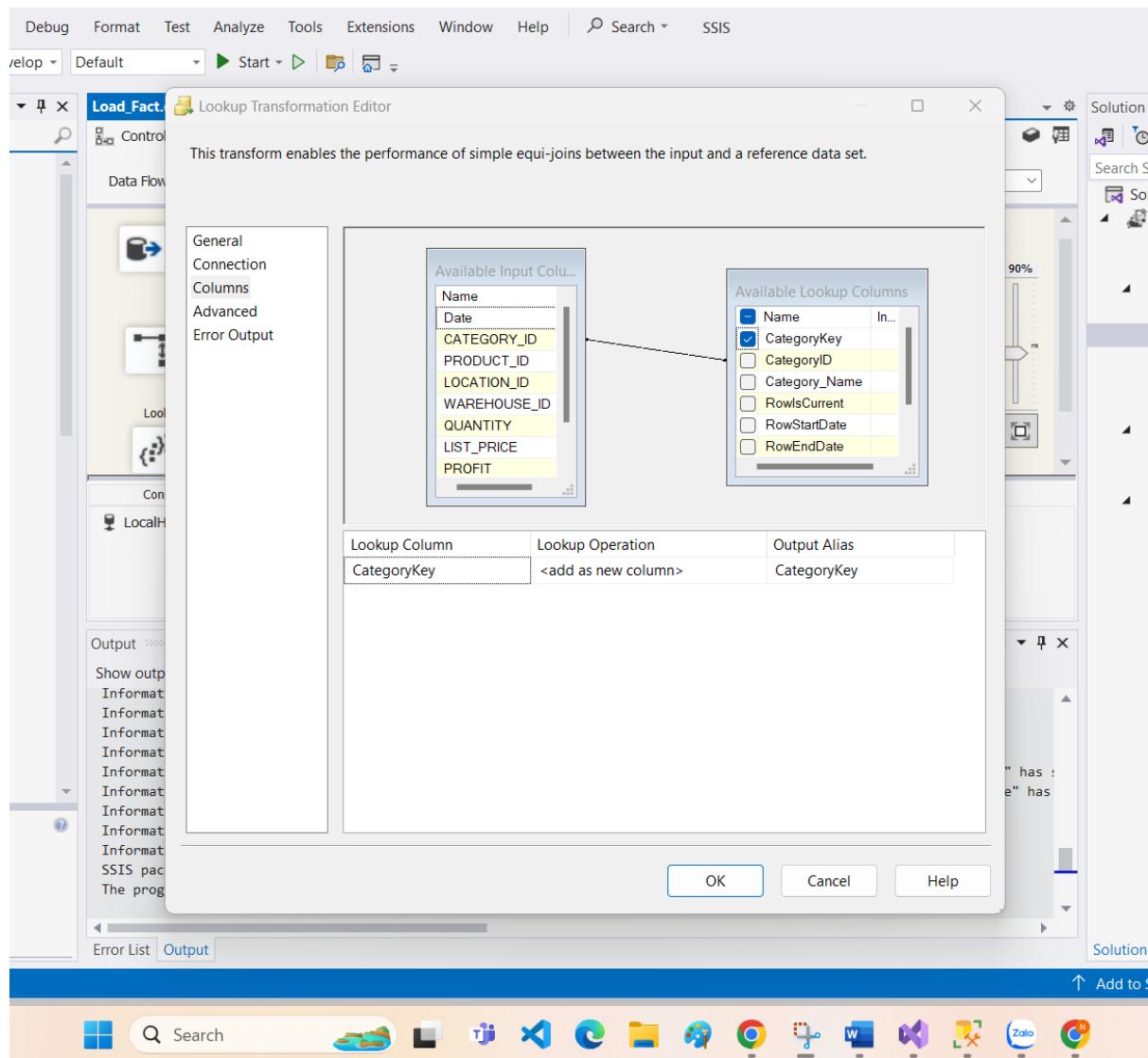
Lookup DimProduct ta sẽ map 2 cột ProductID trong 2 bảng stgInventory và DimProduct để có thể lấy được ProductKey từ bảng DimProduct



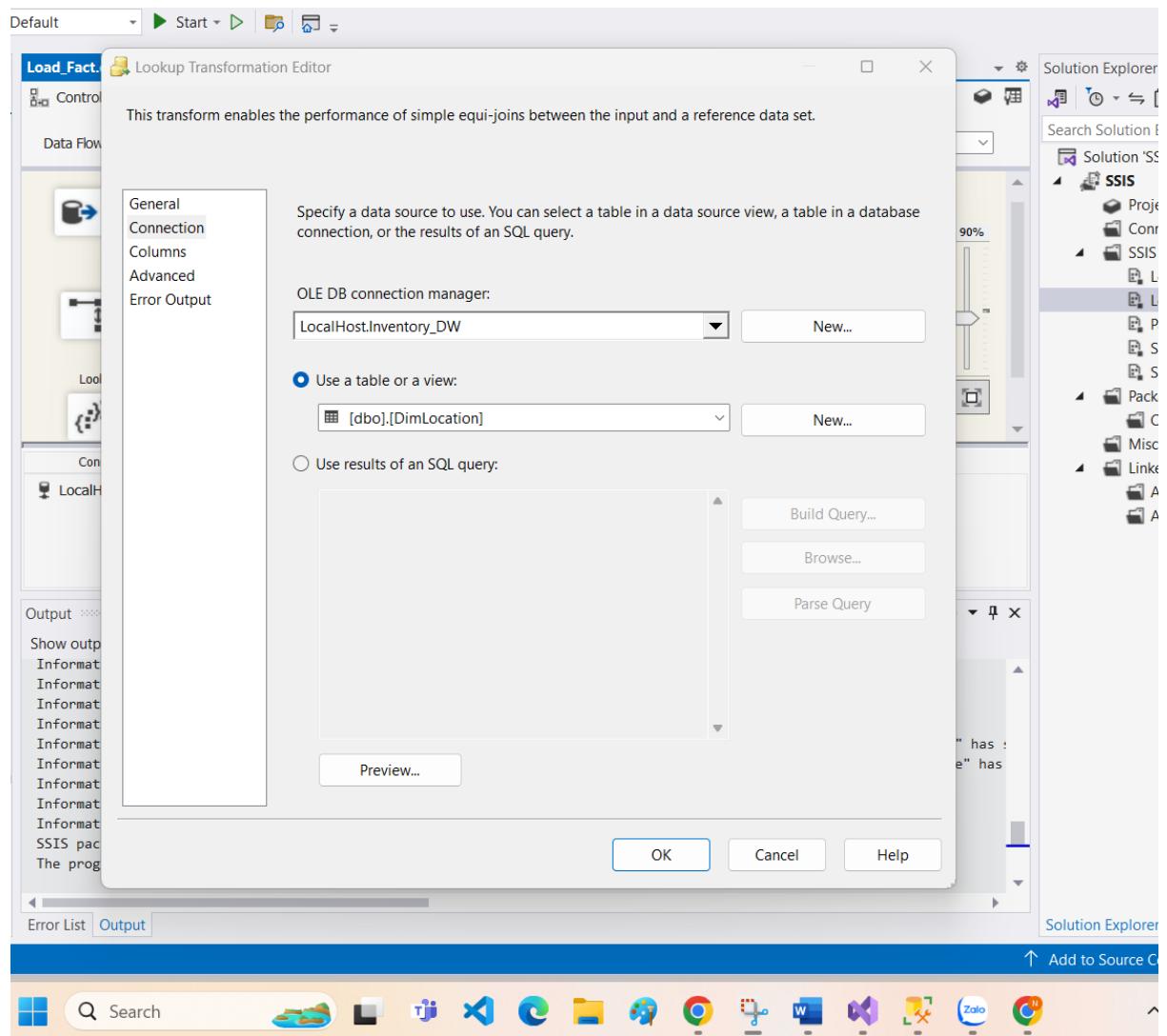


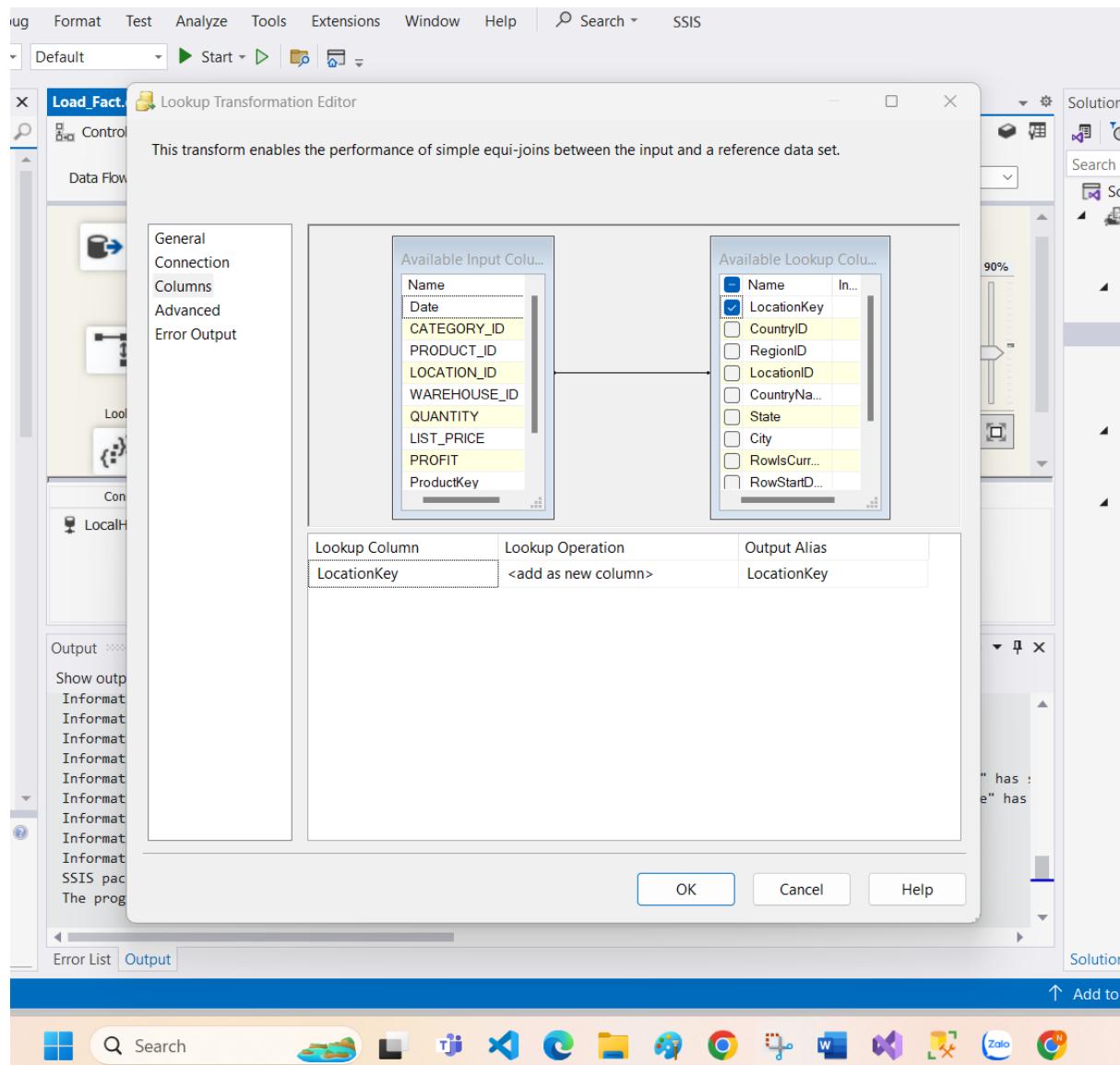
Lookup DimCategory ta sẽ map 2 cột CategoryID trong 2 bảng stgInventory và DimCategory để có thể lấy được CategoryKey trong bảng DimCategorys



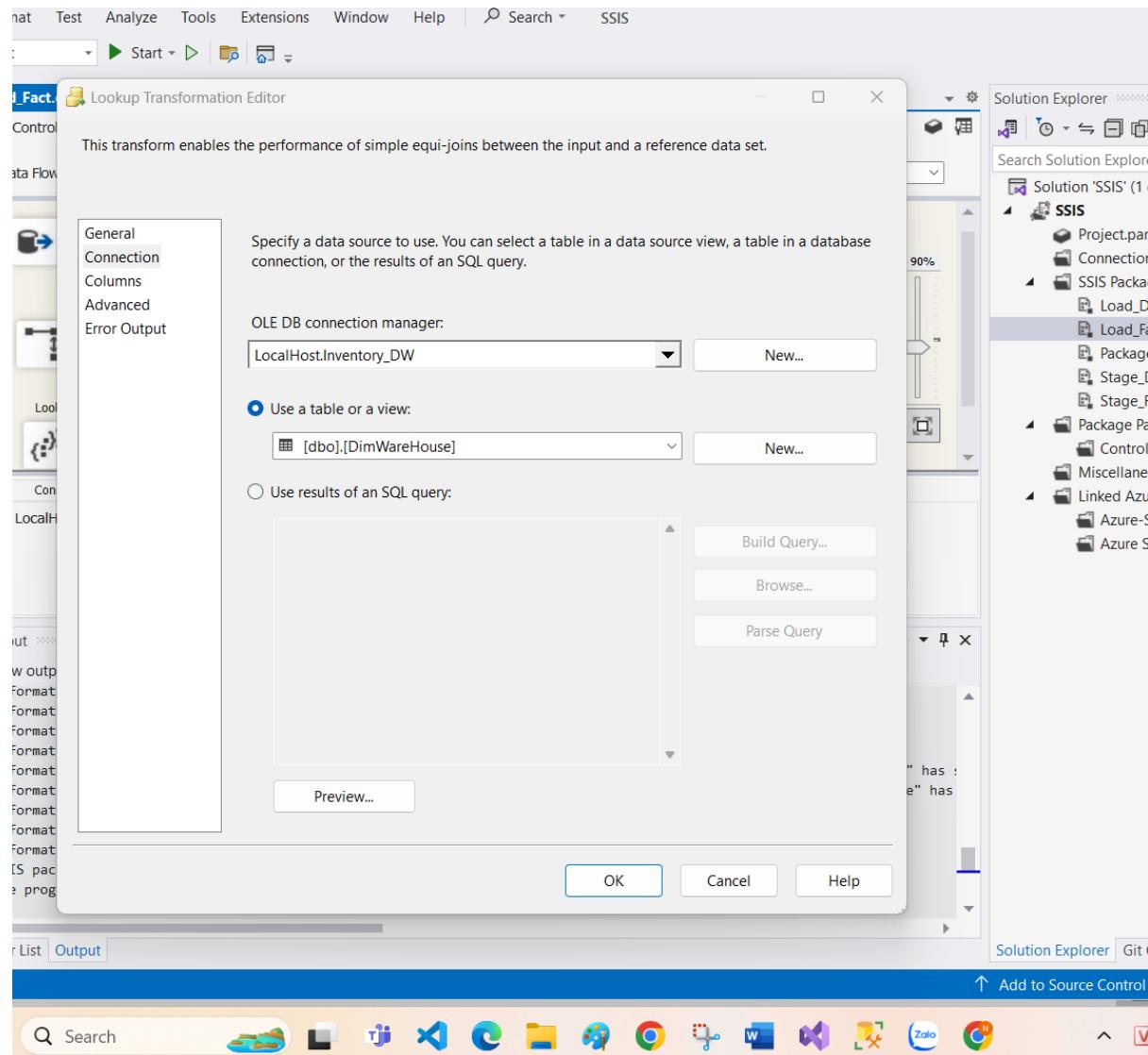


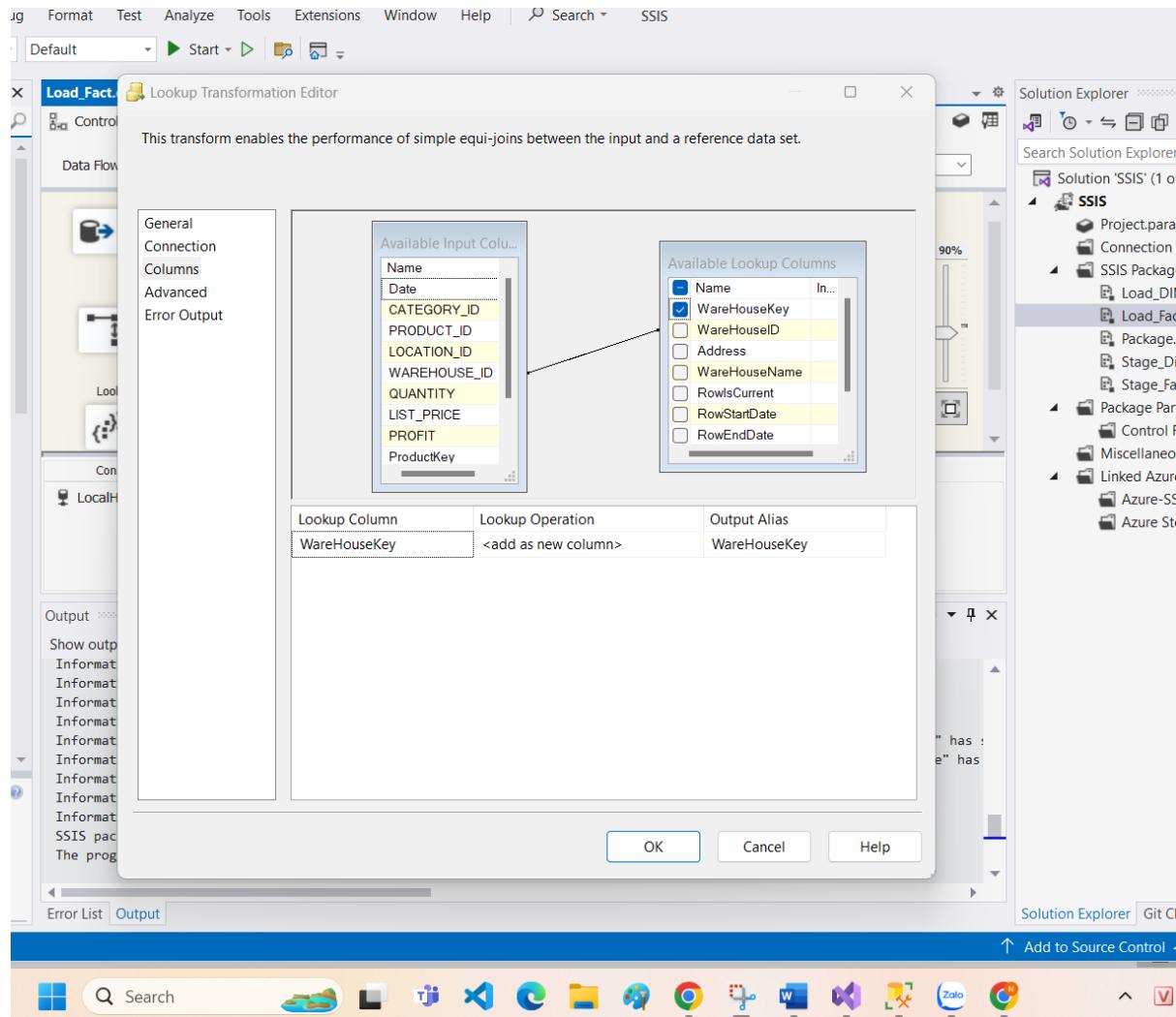
Lookup DimLocation ta sẽ map 2 cột LocationID trong 2 bảng stgInventory và DimLocation để có thể lấy được LocationKey trong bảng DimLocation



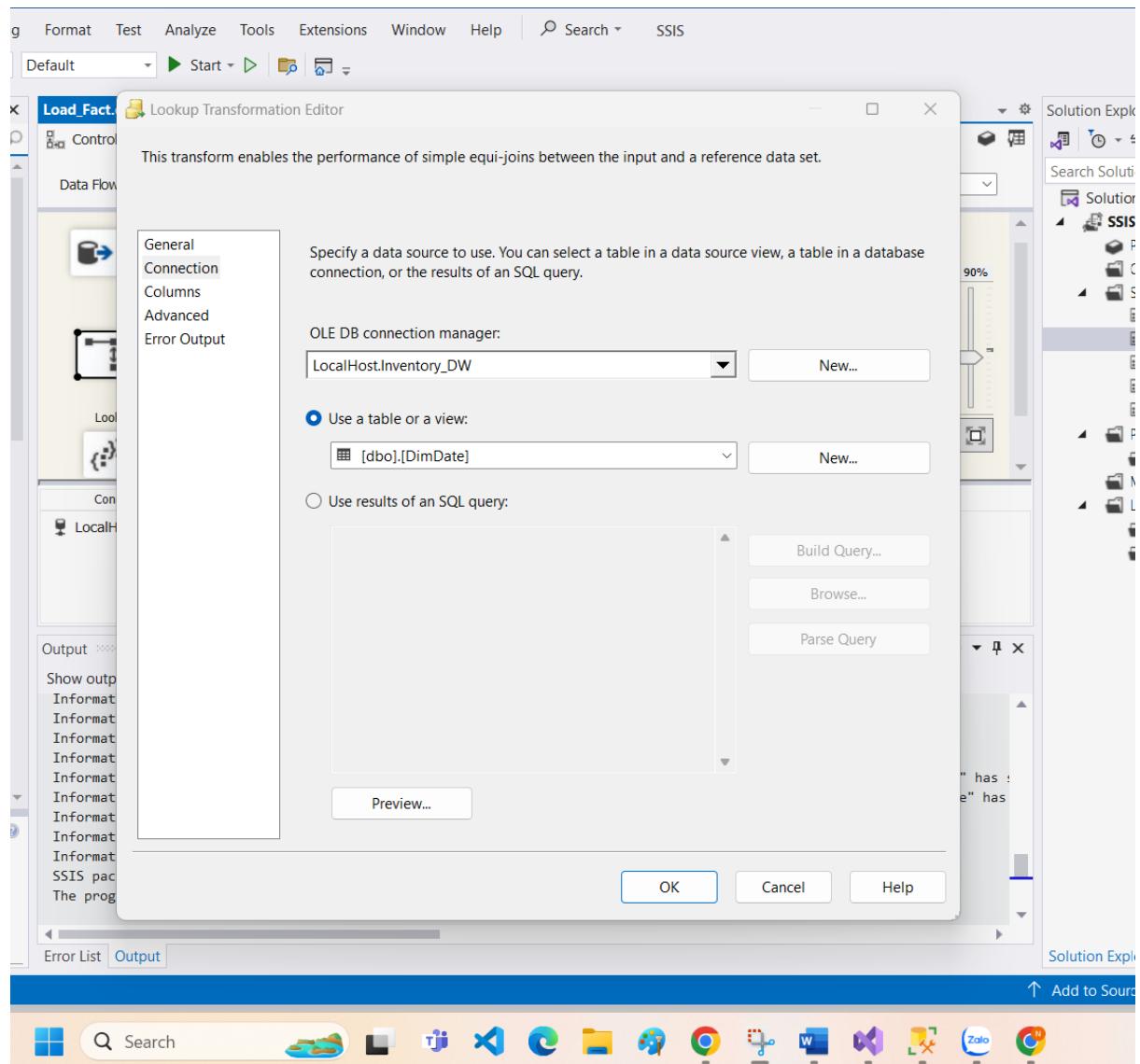


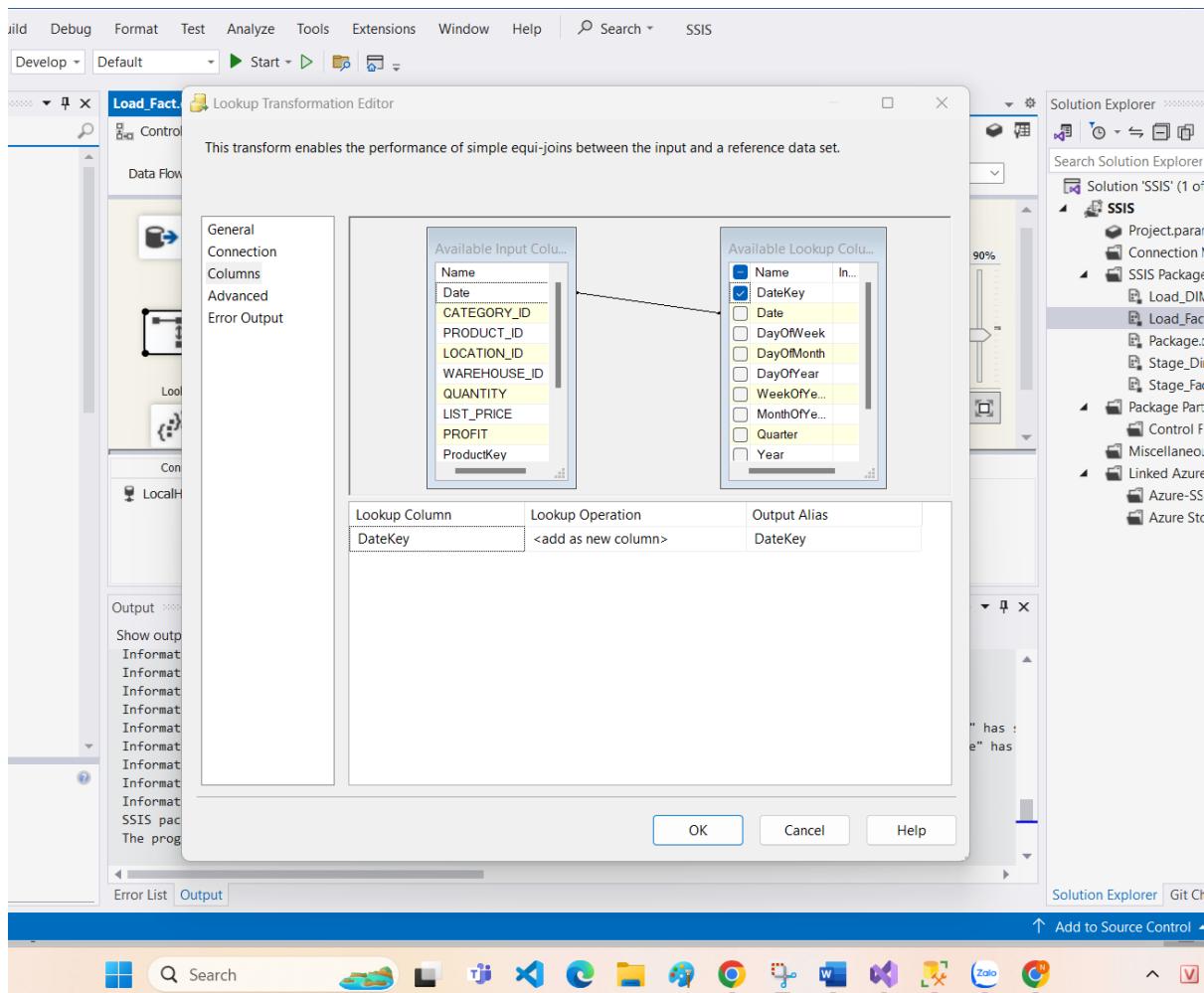
Lookup DimWarehouse ta sẽ map 2 cột WarehouseID trong 2 bảng stgInventory và DimWarehouse để lấy được WarehouseKey trong bảng DimWarehouse



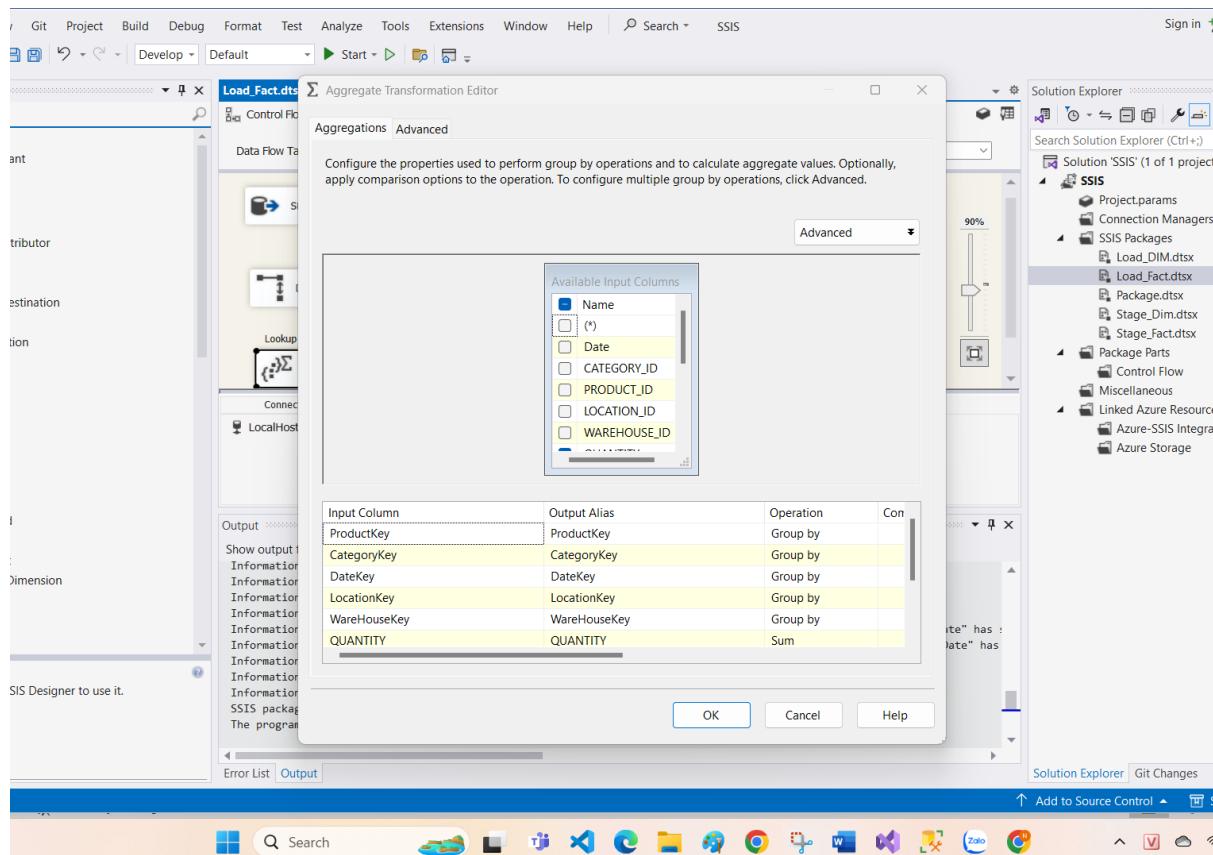


Lookup DimDate ta sẽ map 2 cột Date trong 2 bảng stgInventory và DimDate để lấy được DateKey trong bảng DimDate

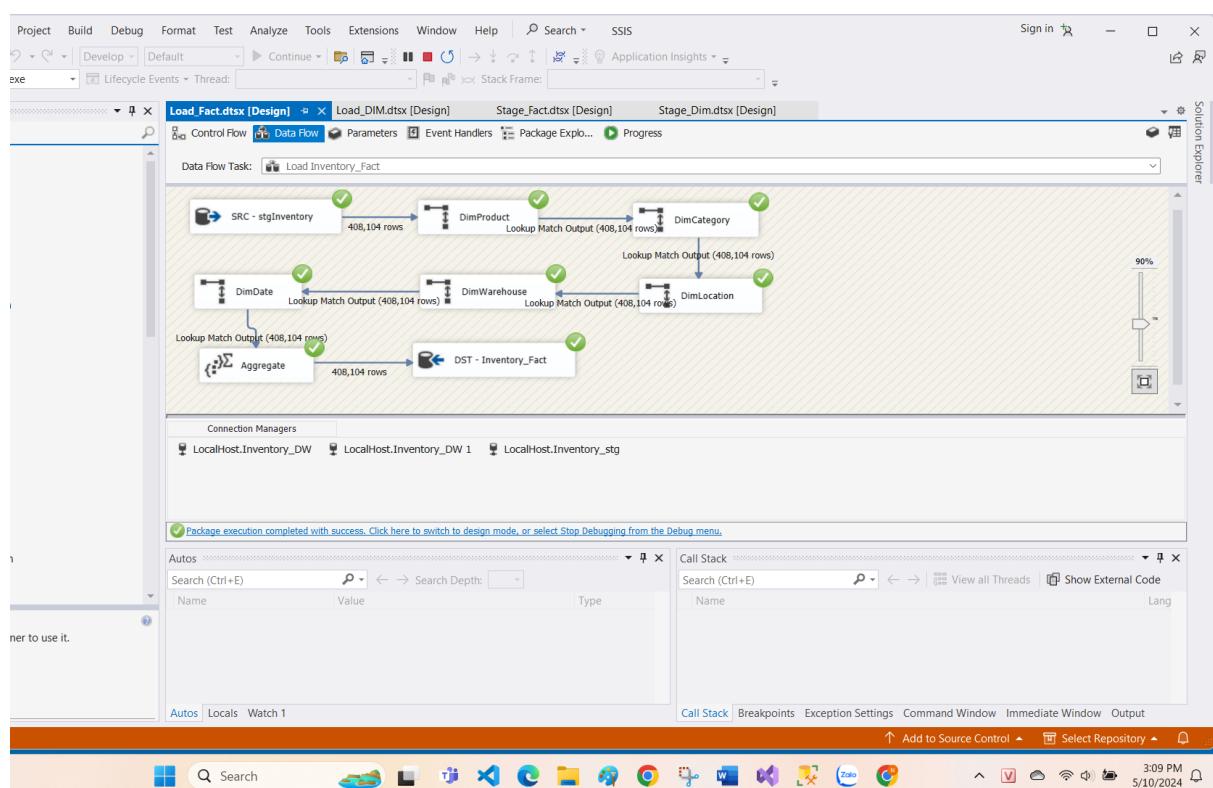




Aggregate ta sẽ chỉ lấy những cột cần thiết để đưa vào bảng Inventory\_fact  
DST- Inventory fact là bảng Inventory\_Fact



Kết quả



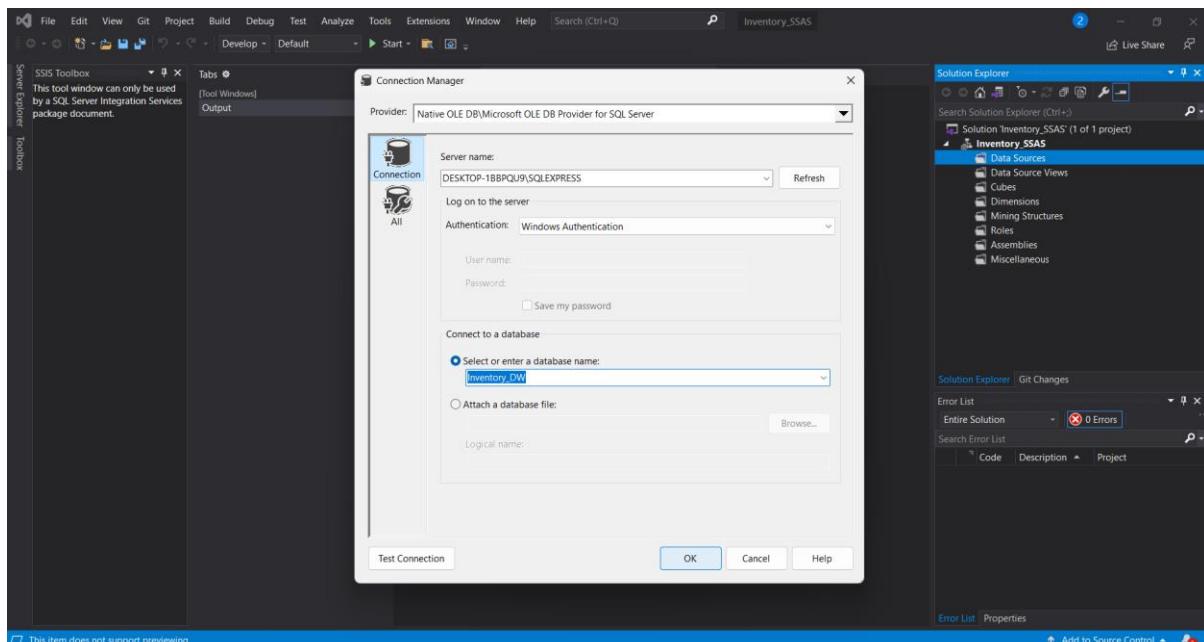


# CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)

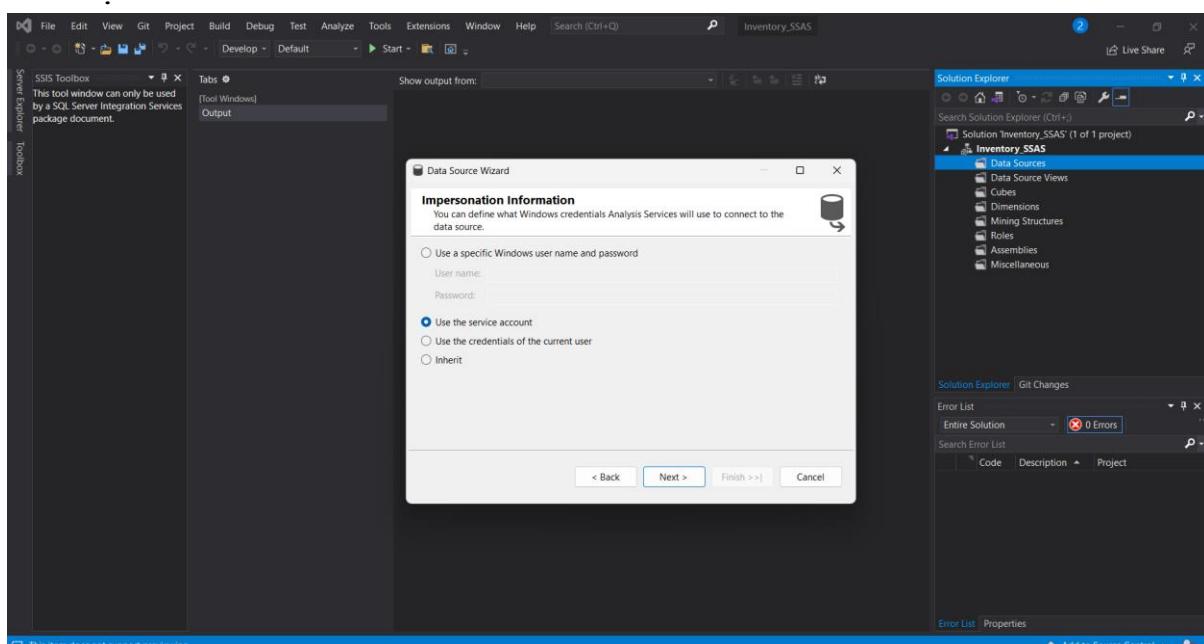
## 4.1. Quá trình xây dựng mô hình

### 4.1.1. Tạo Data Source

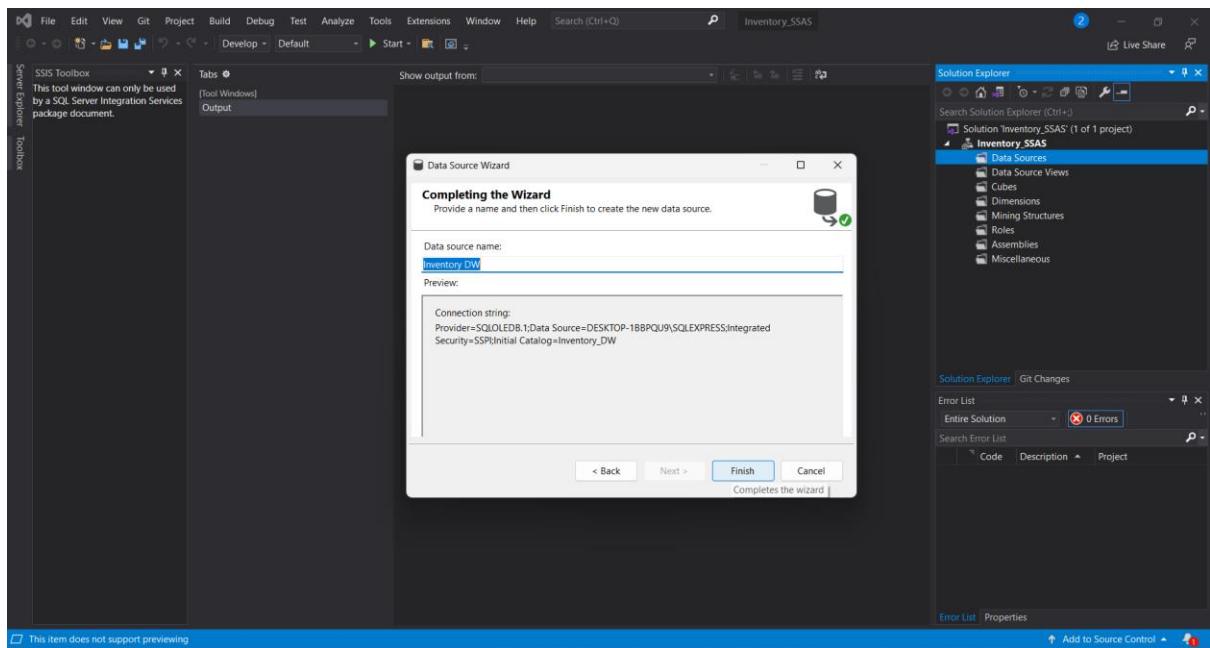
- Right-click vào Data Sources ở khung Solution Explorer. Chon New Data Source để tạo Data Source mới. Kết nối đến Database **Iventory\_DW**



- Chọn Use the service account

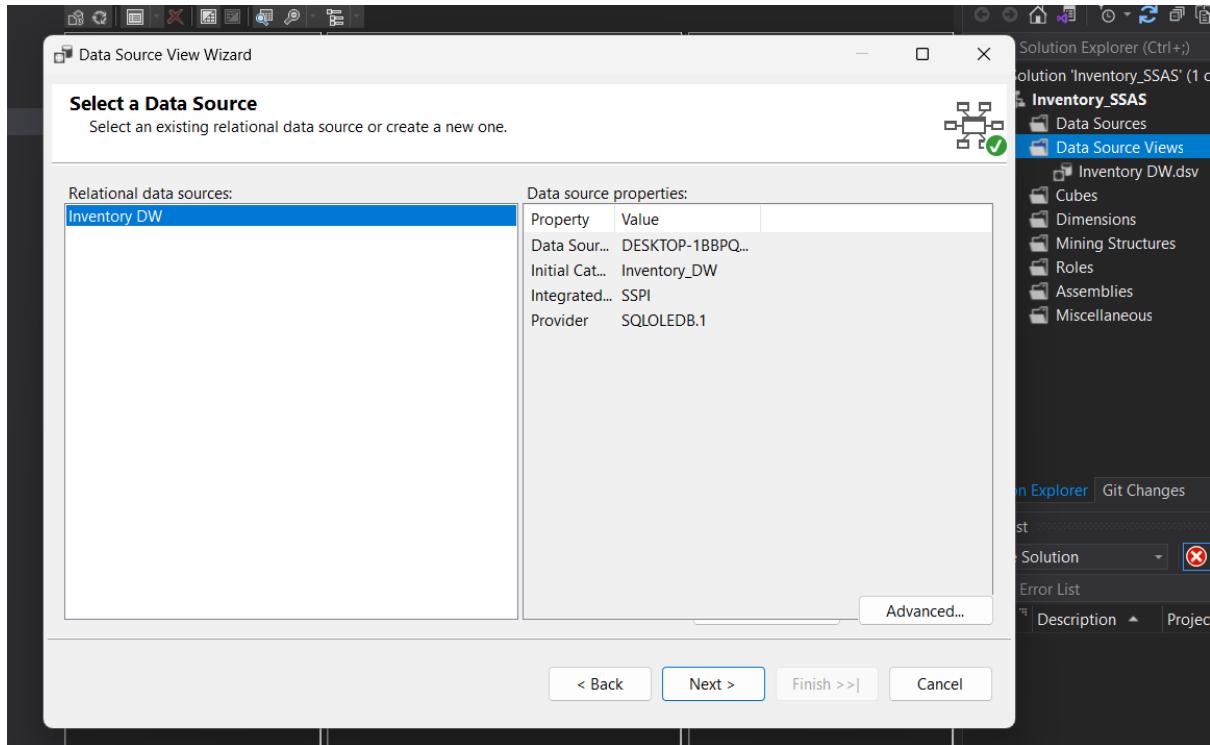


- Chọn Finish

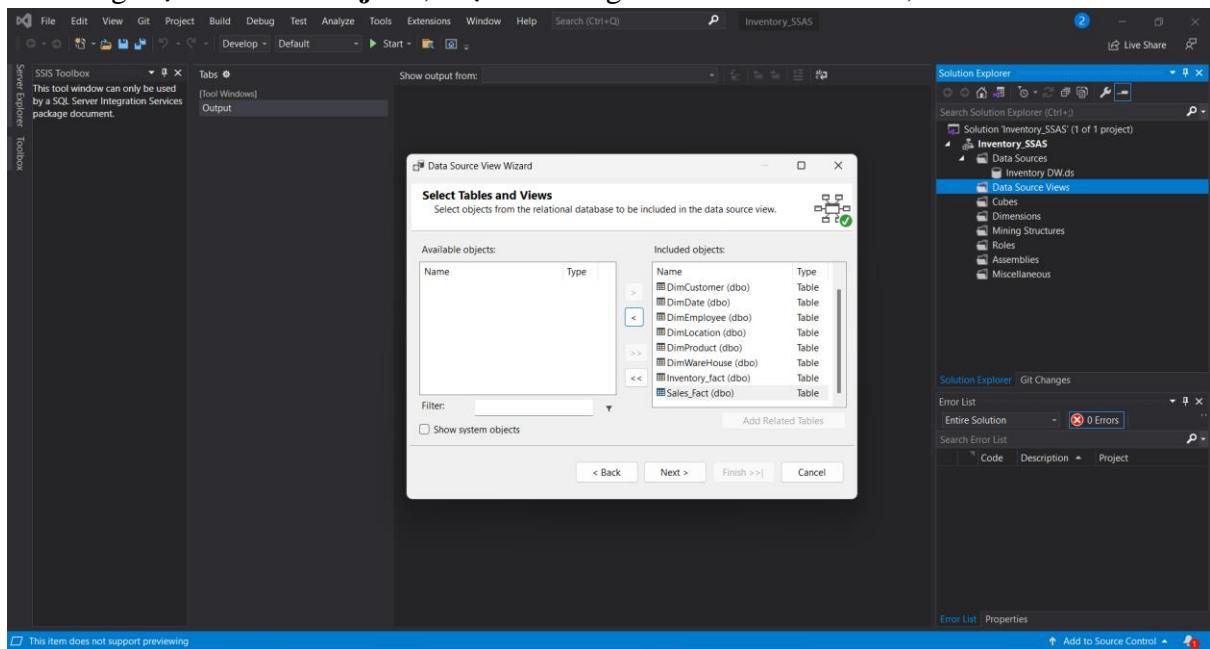


#### 4.1.2. Tạo Data Source View

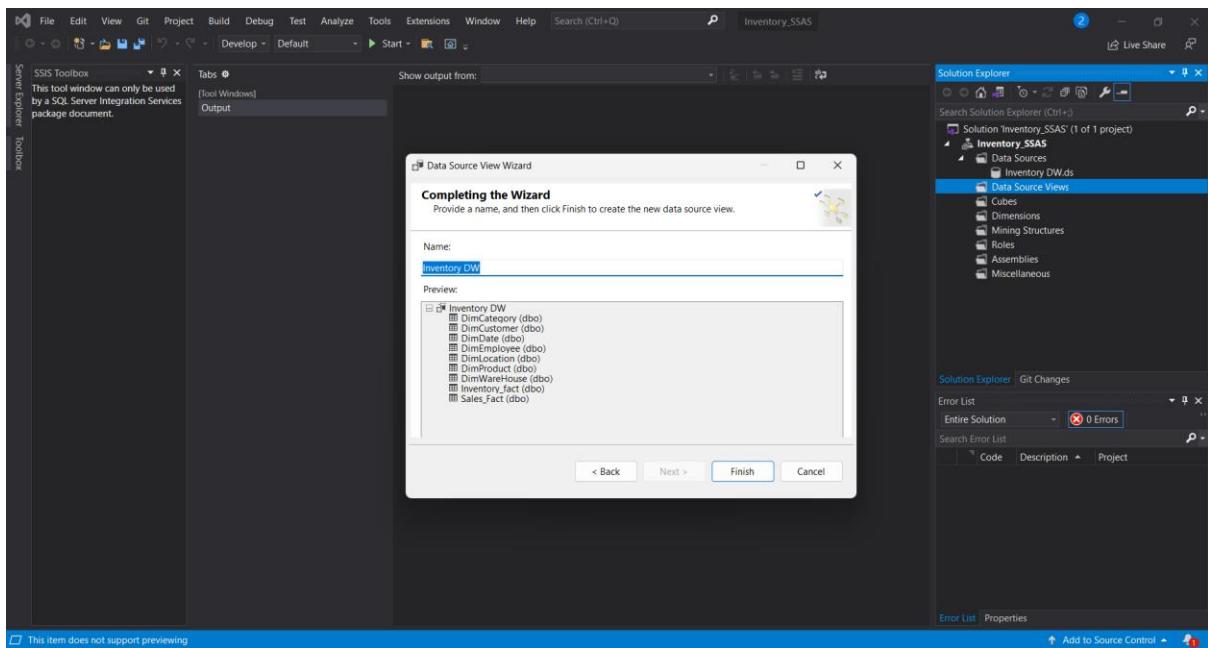
- Right-click vào Data Source Views ở khung Solution Explorer. Chon New Data Source View để tạo Data Source View mới. Chọn Data Soure là **Inventory\_DW** và chọn **Next**.



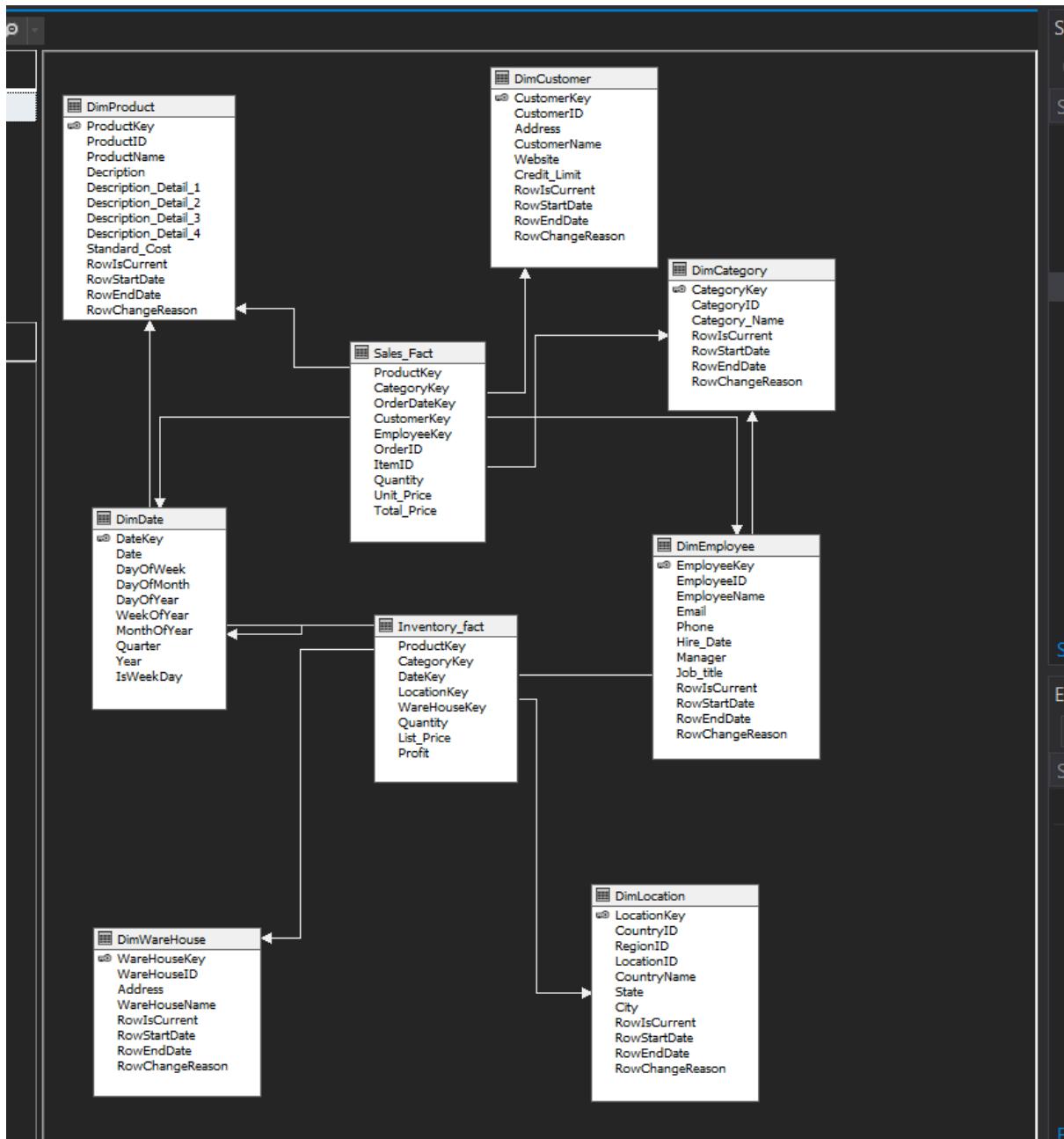
- Trong mục **Included objects**, chọn các bảng Dim và Fact cần thiết, nhấn Next



- Đặt tên cho Data Source View và nhấn Finish để hoàn tất việc tạo Data Source View

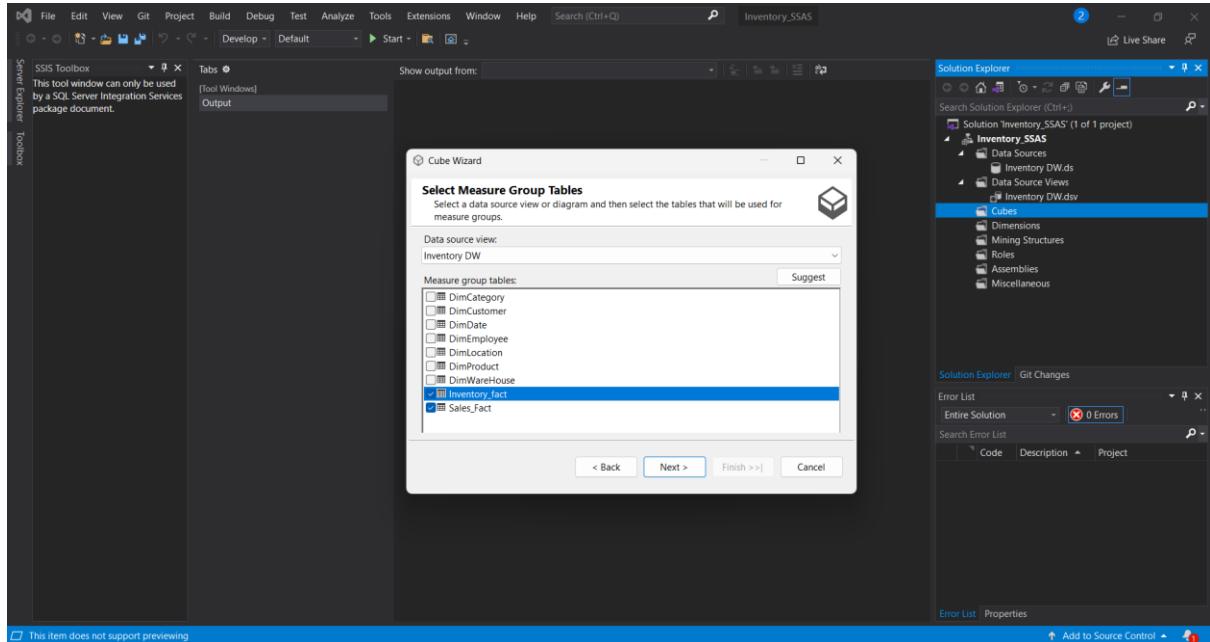


- Double-Click vào Data Source View vừa tạo để hiển thị diagram

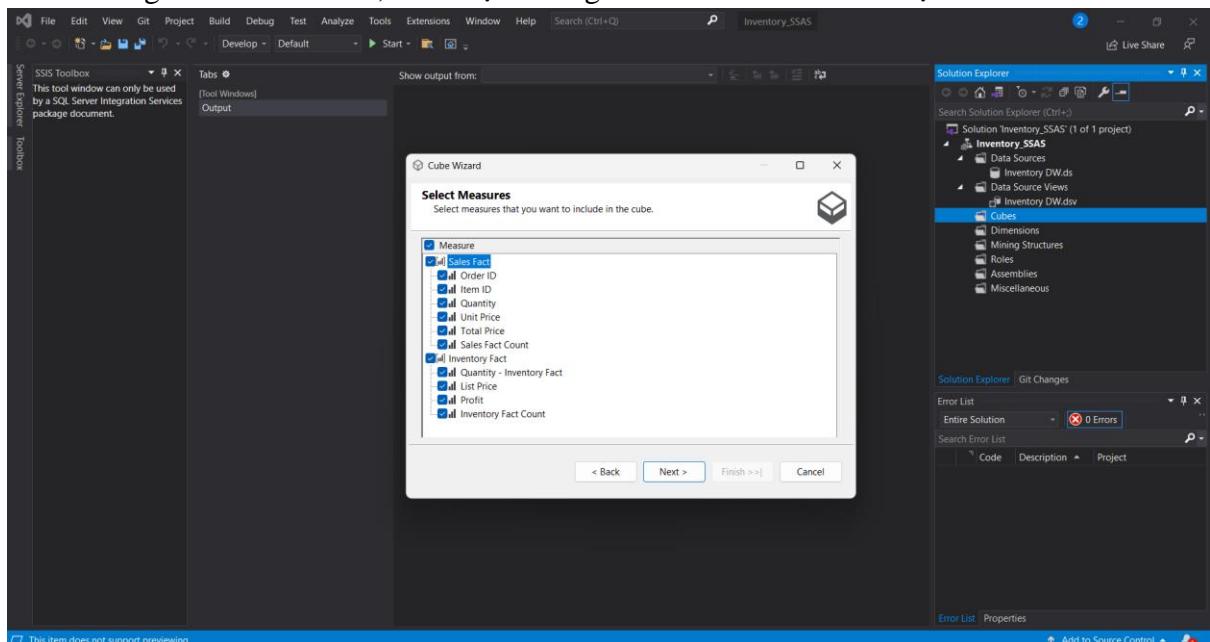


## 4.2. Quá trình xây dựng khối Cube

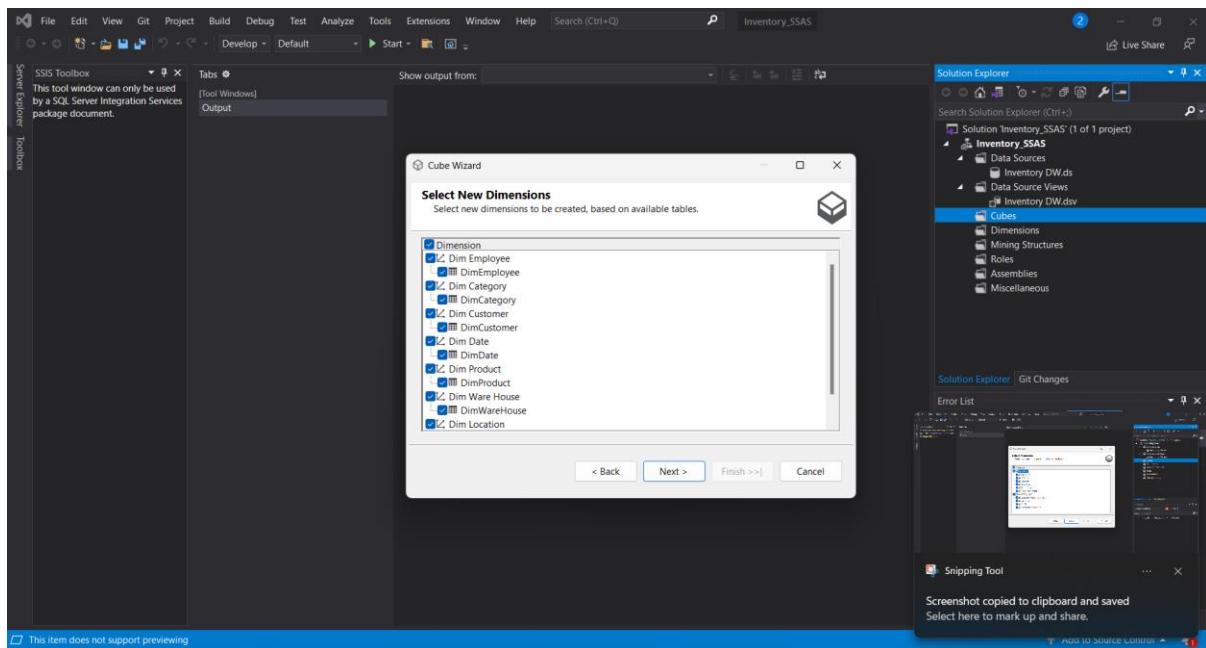
- Right-click vào **Cubes** ở khung Solution Explorer. Chon New Cube để tạo Cube mới.
- Chọn Data Source View là **Iventory DW**. Tích chọn 2 bảng **Sales\_fact** và **Inventory\_fact**, nhấn Next



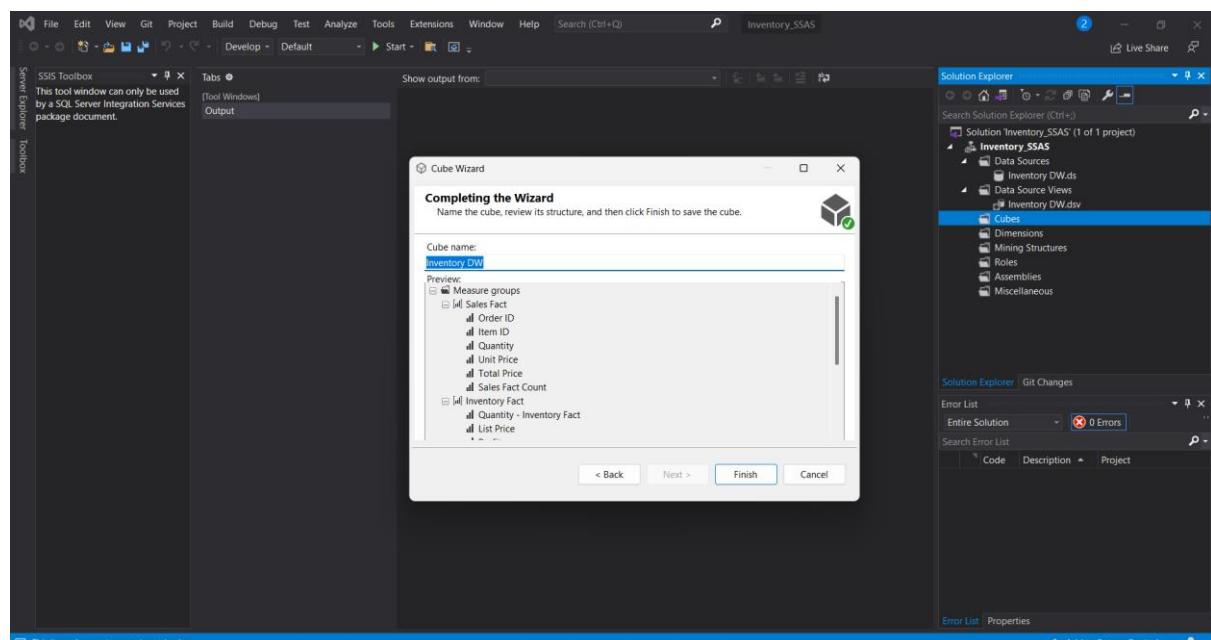
- Ở dialog Select Measures, tích chọn những measure cần thiết các mục sau



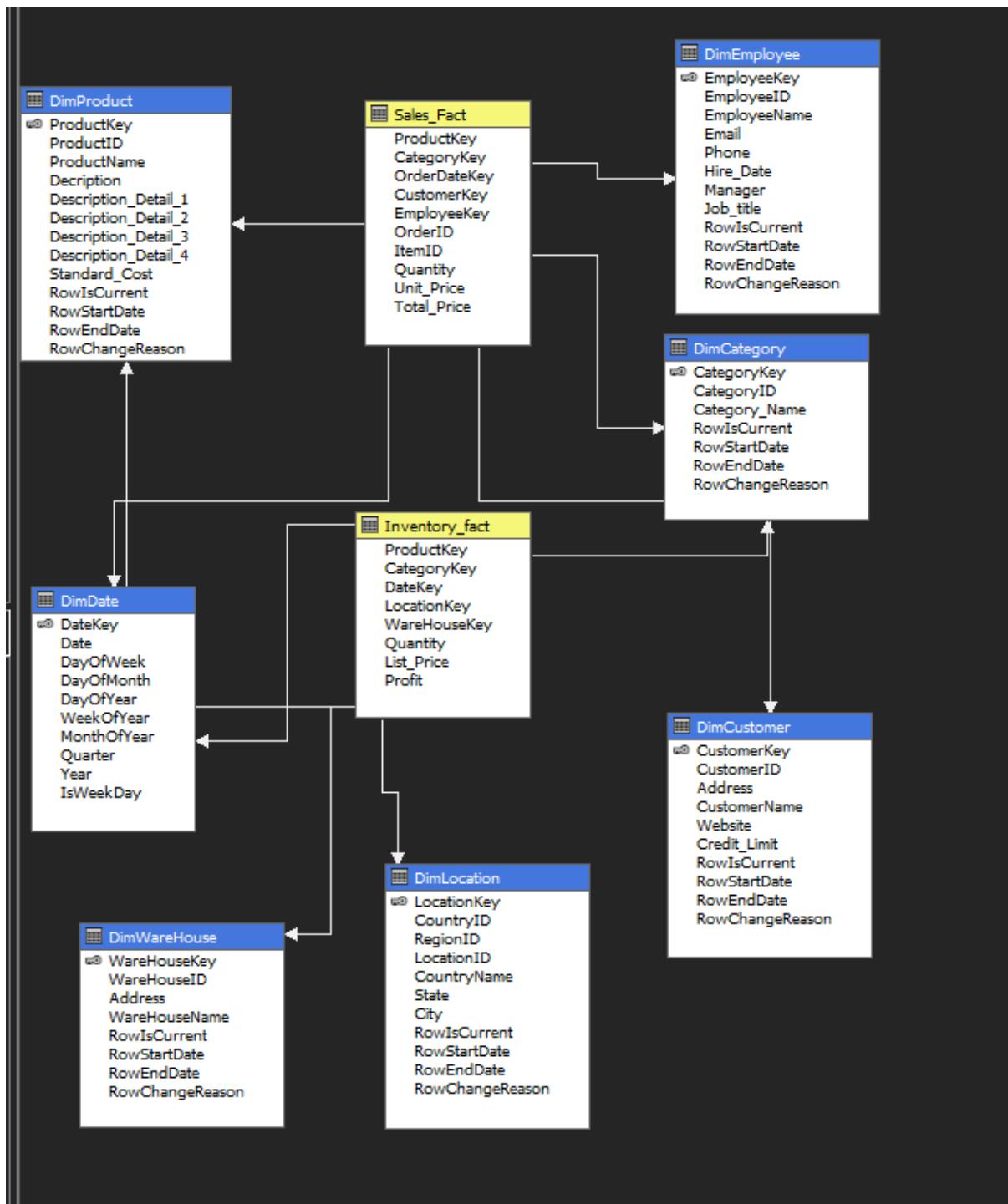
- Khởi tạo các bảng Dimension



- Đặt tên **Hardware DW Cube** để hoàn tất khởi tạo.



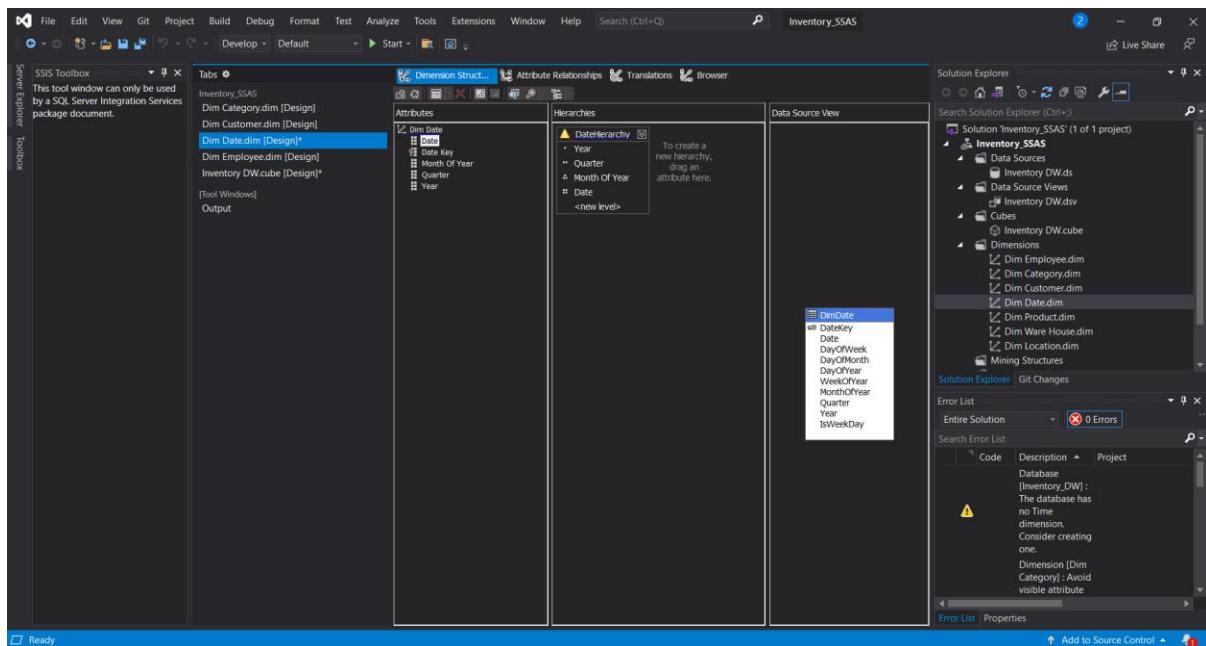
- Các bảng Fact, Dimension và quan hệ giữa chúng sẽ được thể hiện trong Diagram



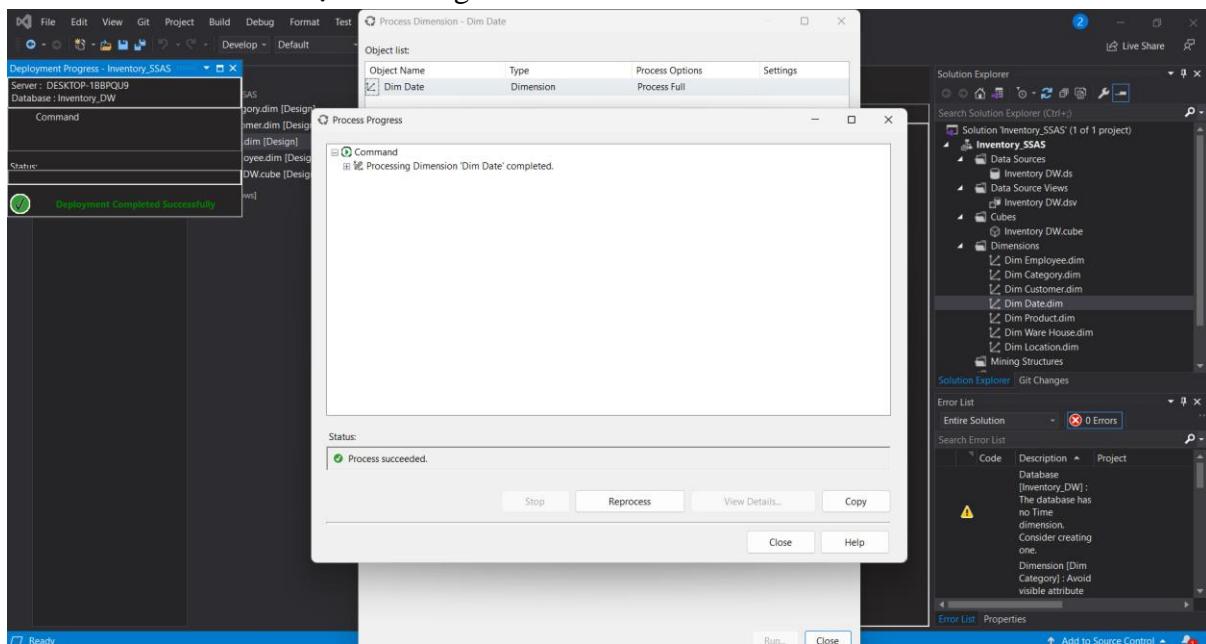
## 4.3. Cấu hình Hierarchy

### 4.3.1. Tạo Hierarchy cho Dim Date

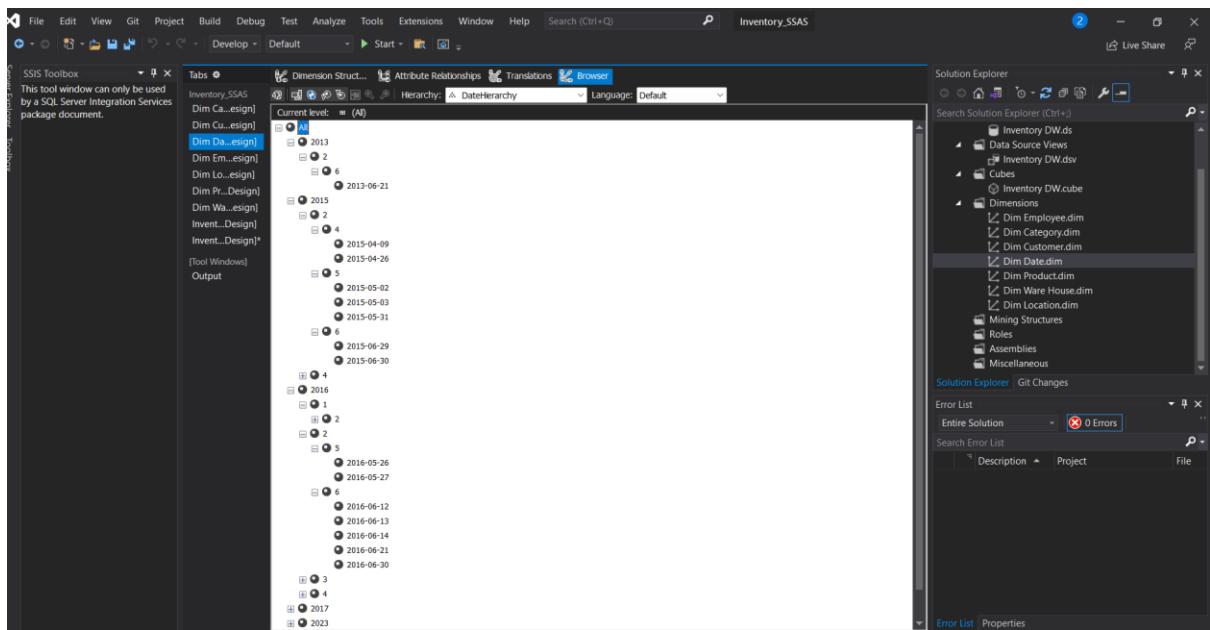
- Thêm thuộc tính vào bảng Dim: Tại table DimDate trong panel Data Source View, kéo thuộc tính cần thiết vào panel Attributes.
- Kéo thả 4 thuộc tính *Year*, *Quarter*, *Month Of Name*, *Date* từ Attributes vào panel Hierarchies để tạo ra DateHierarchy



- Process để load dữ liệu cho bảng Dim Date

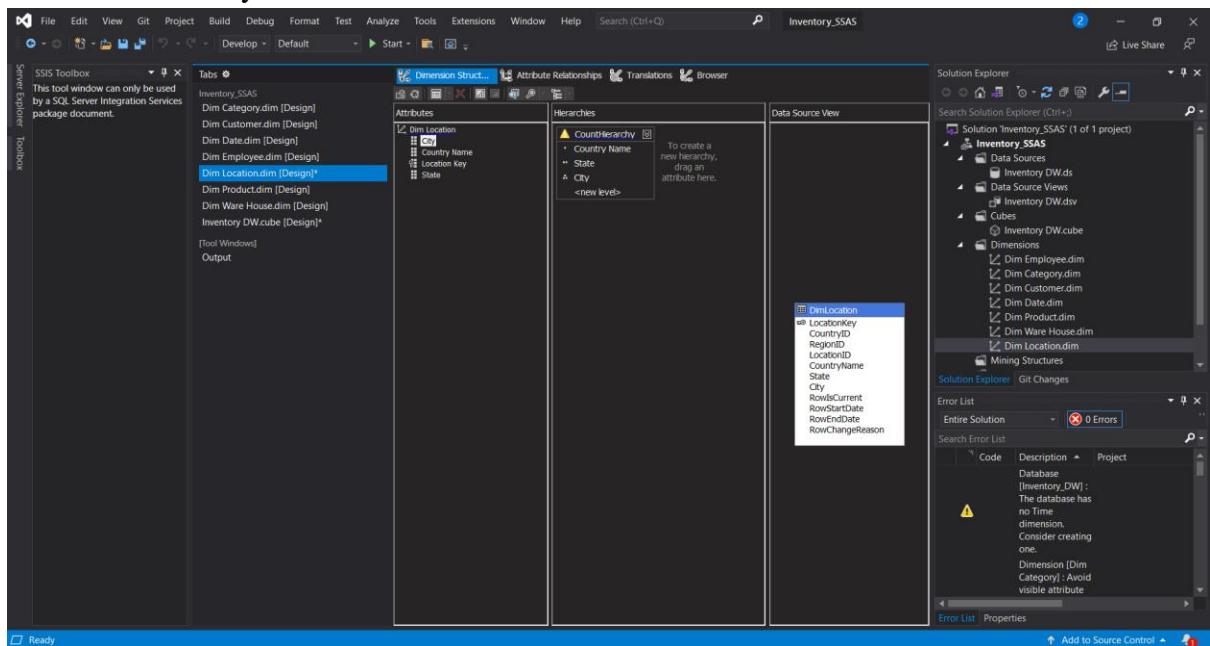


- Chuyển đến tab Browser. Ta thấy thời gian được phân cấp theo năm → quý → tháng.

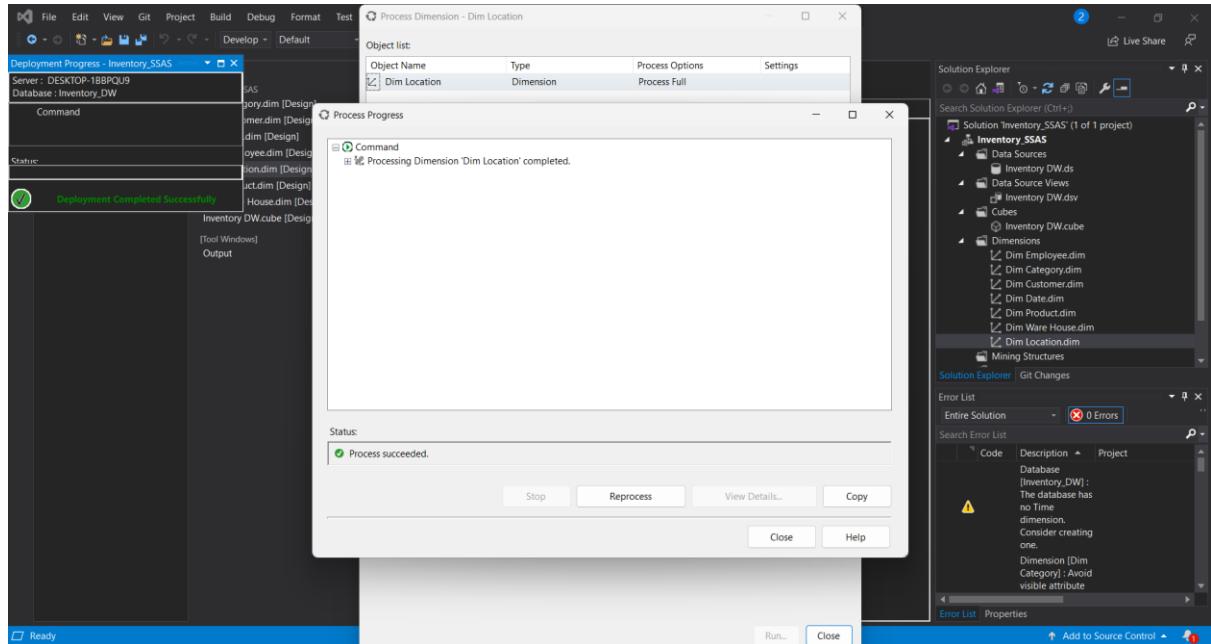


### 4.3.2. Tạo Hierarchy cho Dim Location

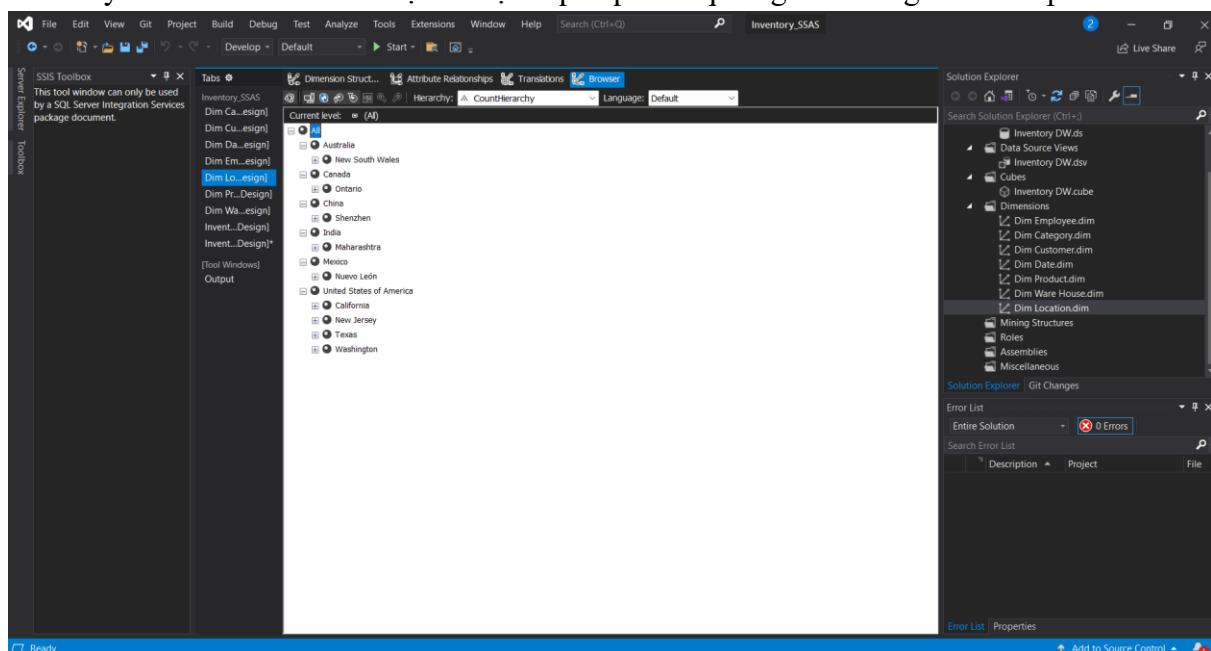
- Tại table Dim Loaction trong panel Data Source View, kéo 3 thuộc tính như hình vào panel Attributes. Sau đó kéo thả lần lượt *Country Name*, *State*, *City* vào Hierarchies để tạo CountHierarchy



### - Process bảng Dim Location

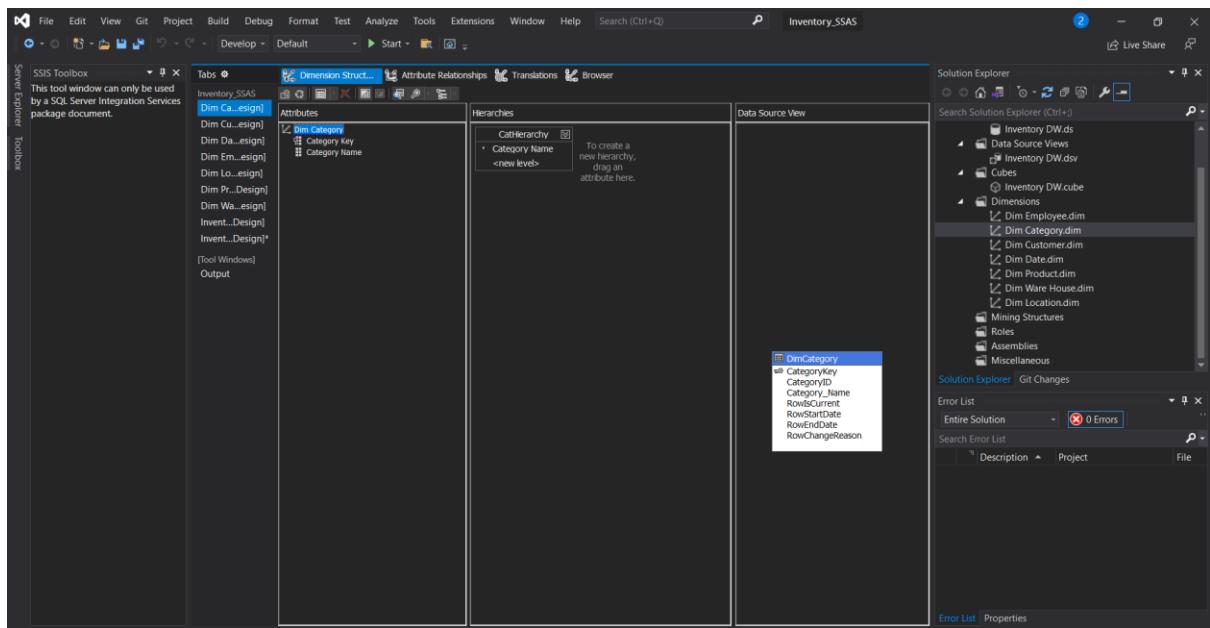


### - Chuyển đến tab Browers. Vị trí được sắp xếp theo quốc gia → bang → thành phố

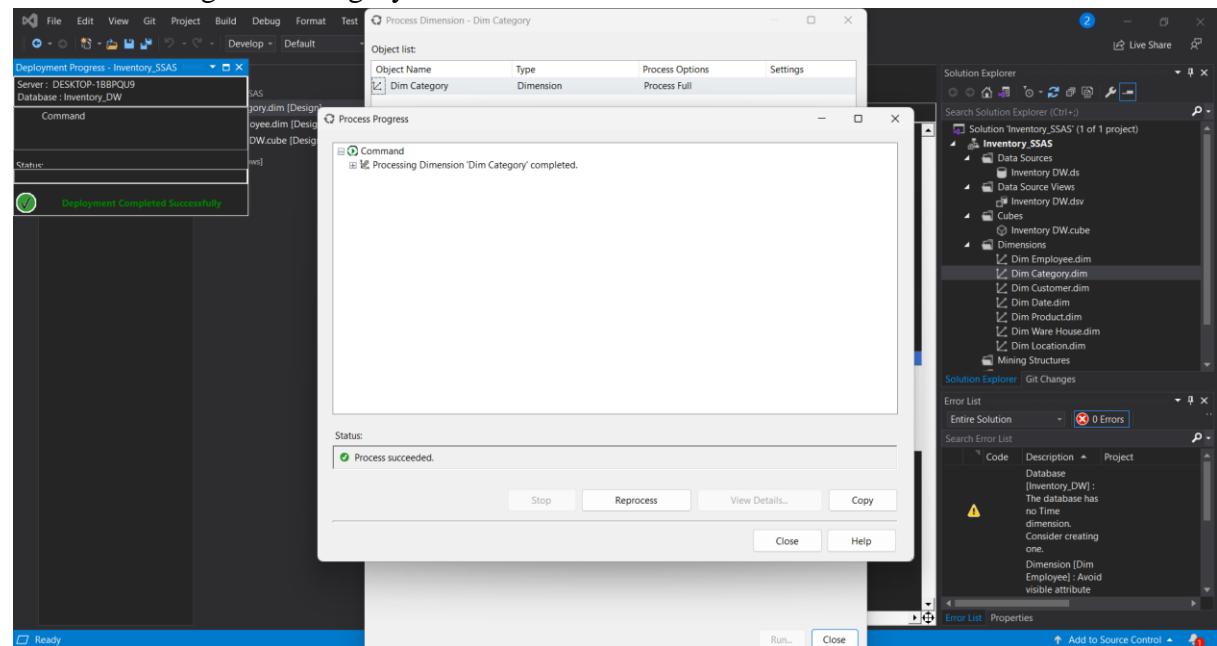


### 4.3.3. Tạo Hierarchy cho Dim Category

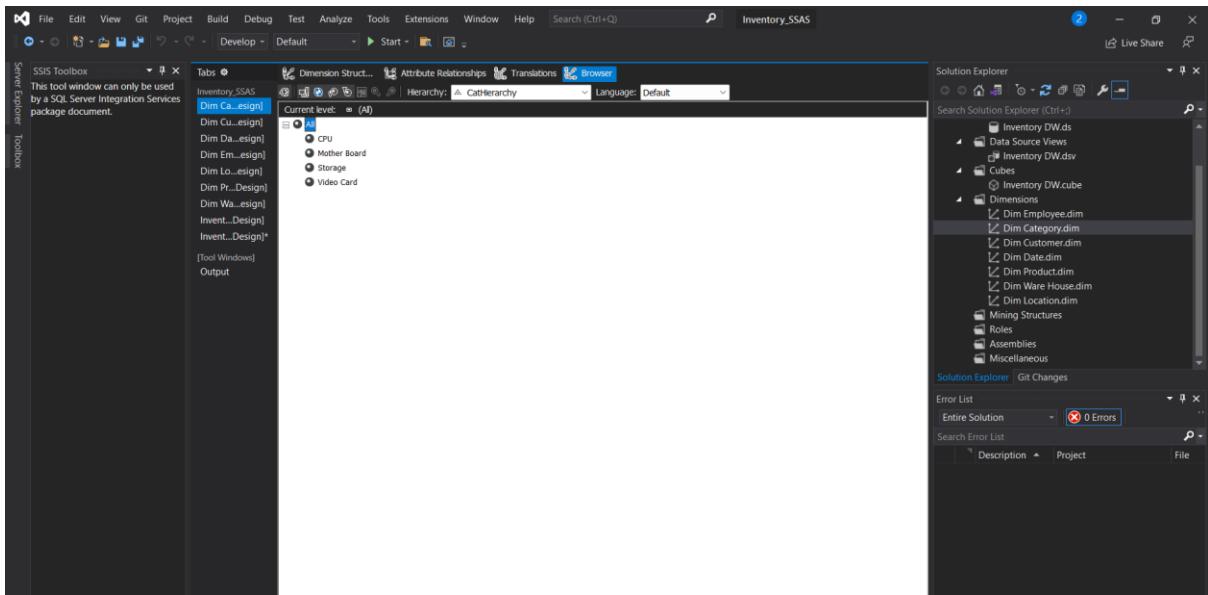
#### - Tạo CatHierachy



### - Process bảng Dim Category

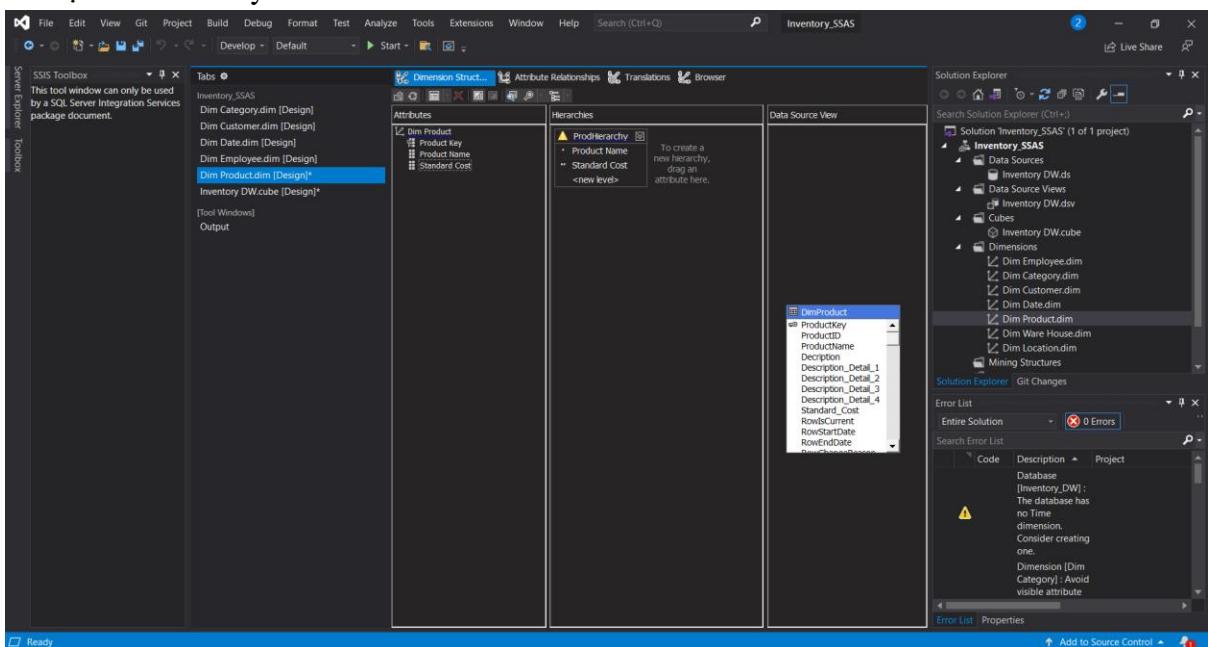


### - Chuyển qua tab Browser sẽ thấy các danh mục sản phẩm:

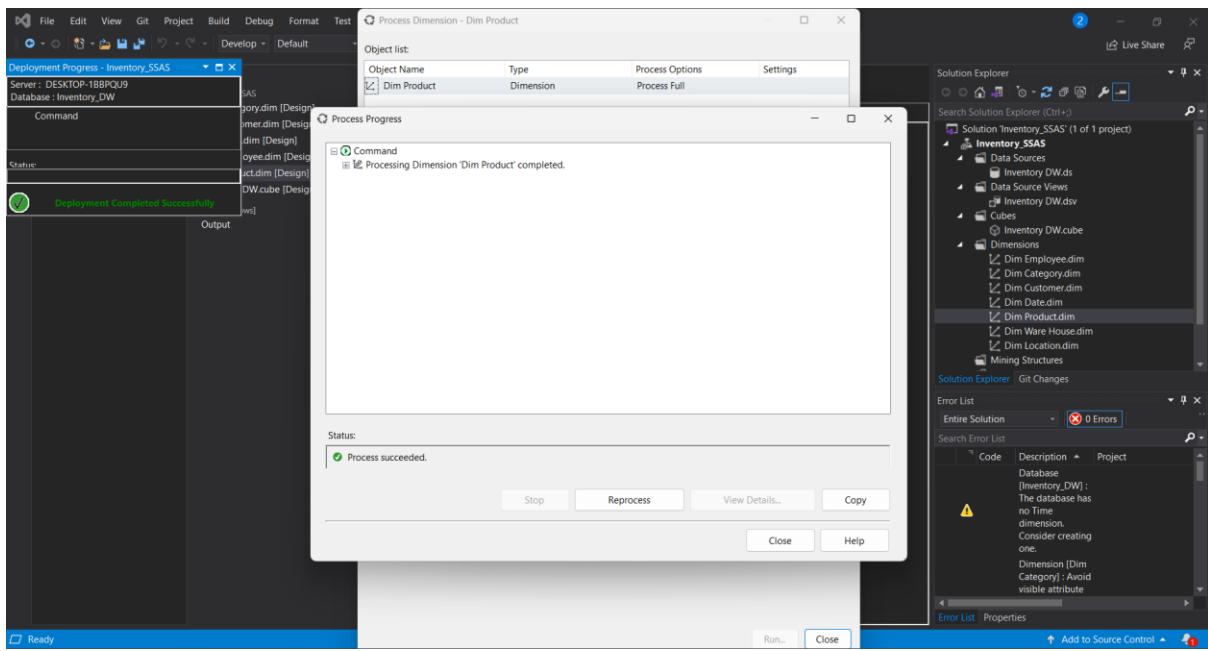


#### 4.3.4. Tạo Hierarchy cho Dim Product

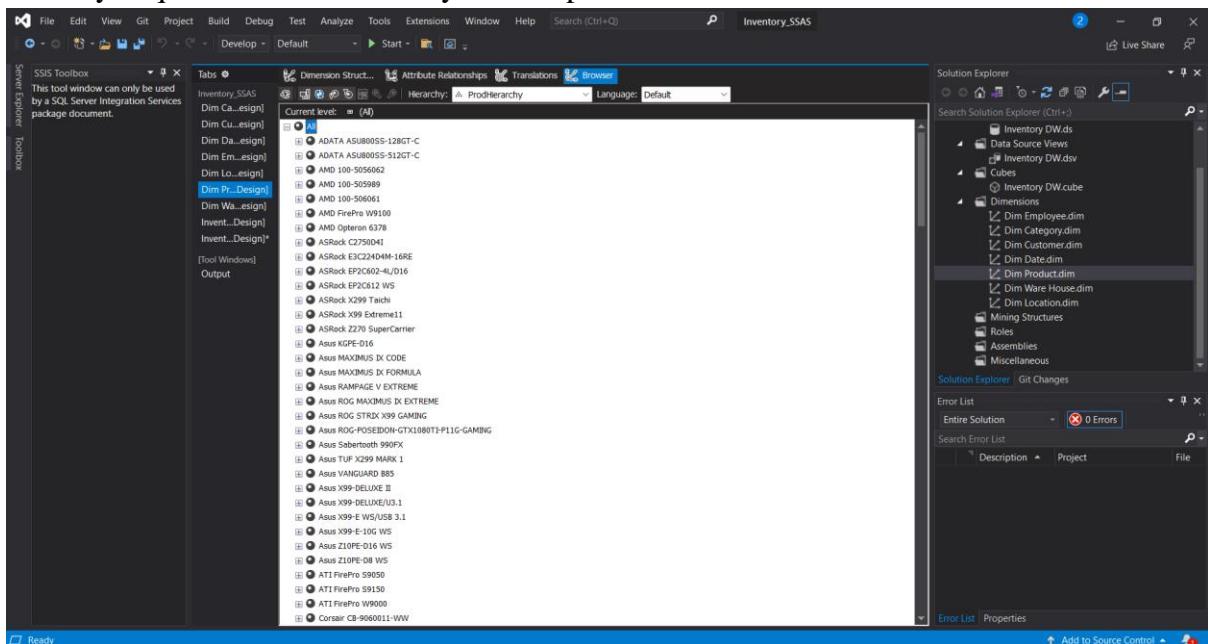
- Tạo ProHierarchy



- Process bảng Dim Product

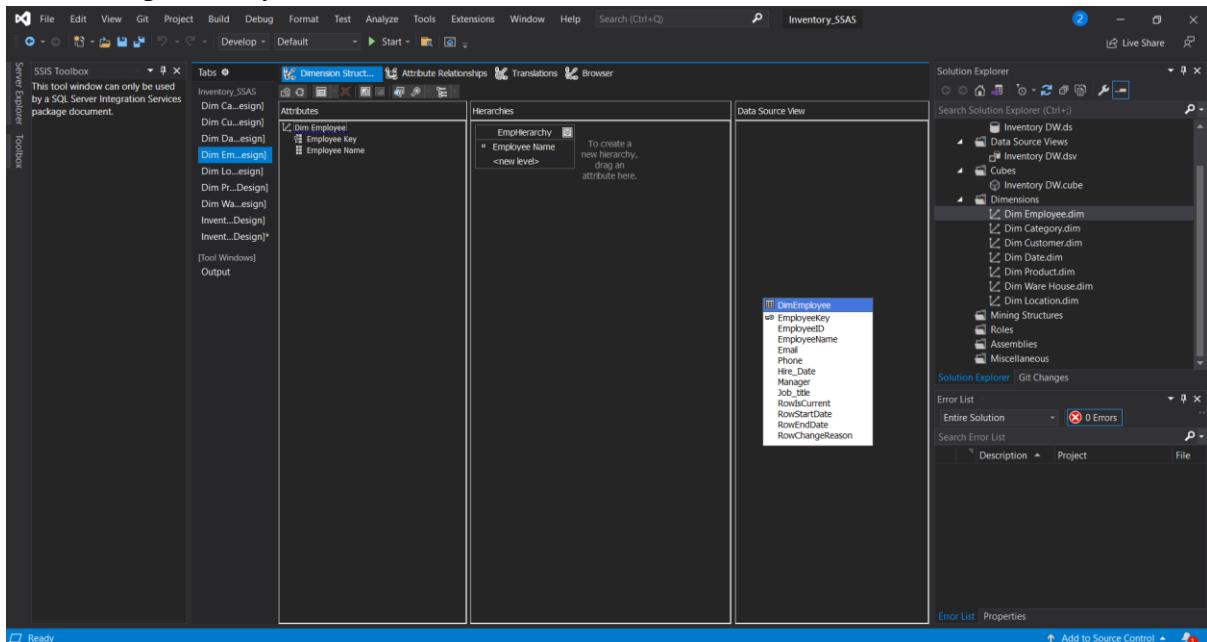


- Chuyển qua tab Browser sẽ thấy các sản phẩm:

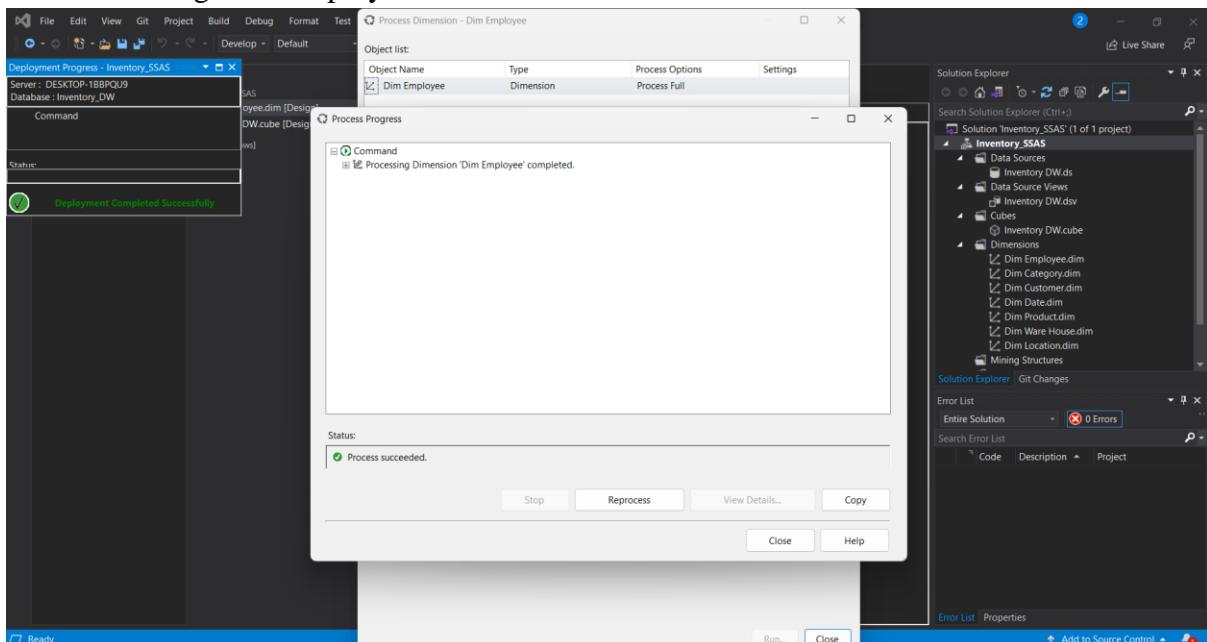


#### 4.3.5. Tạo Hierarchy cho Dim Employee

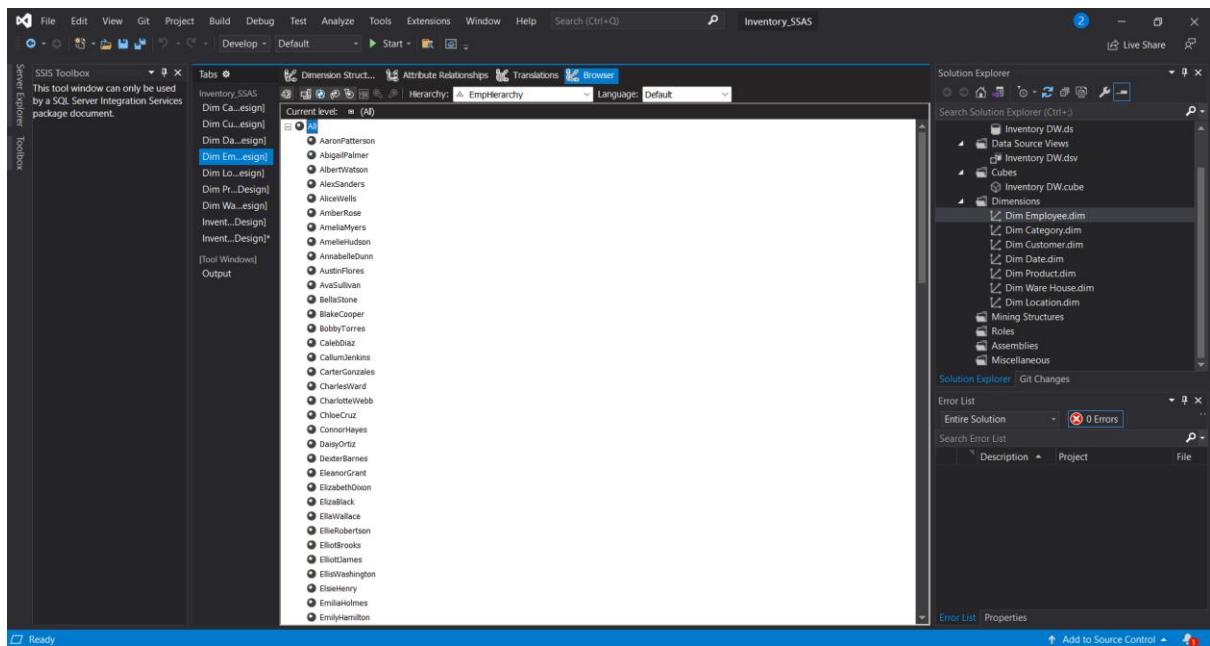
- Tạo EmpHierarchy



- Process bảng Dim Employee

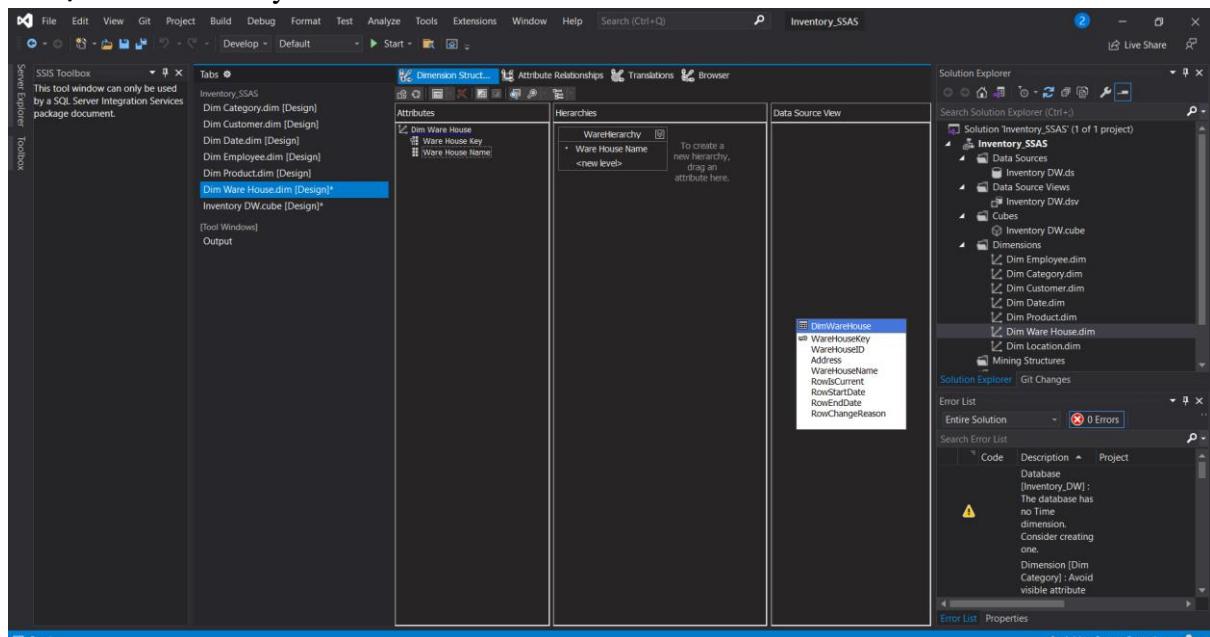


- Chuyển đến tab browser để xem danh sách nhân viên:

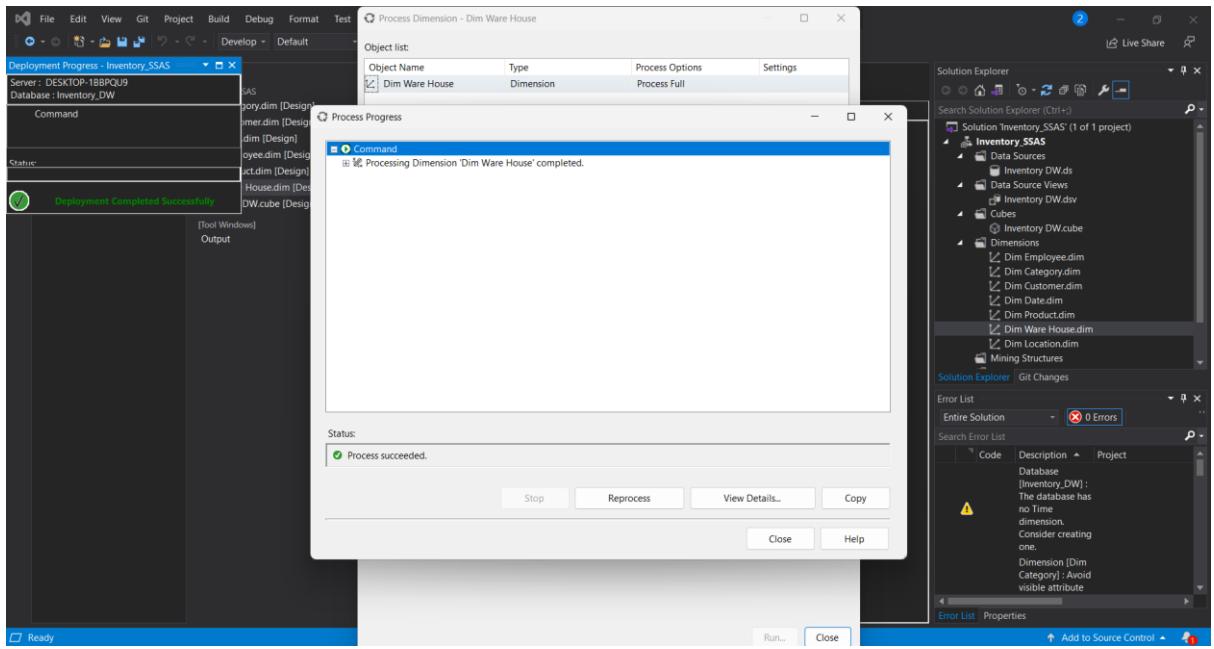


#### 4.3.6. Tạo Hierarchy cho Dim WareHouse

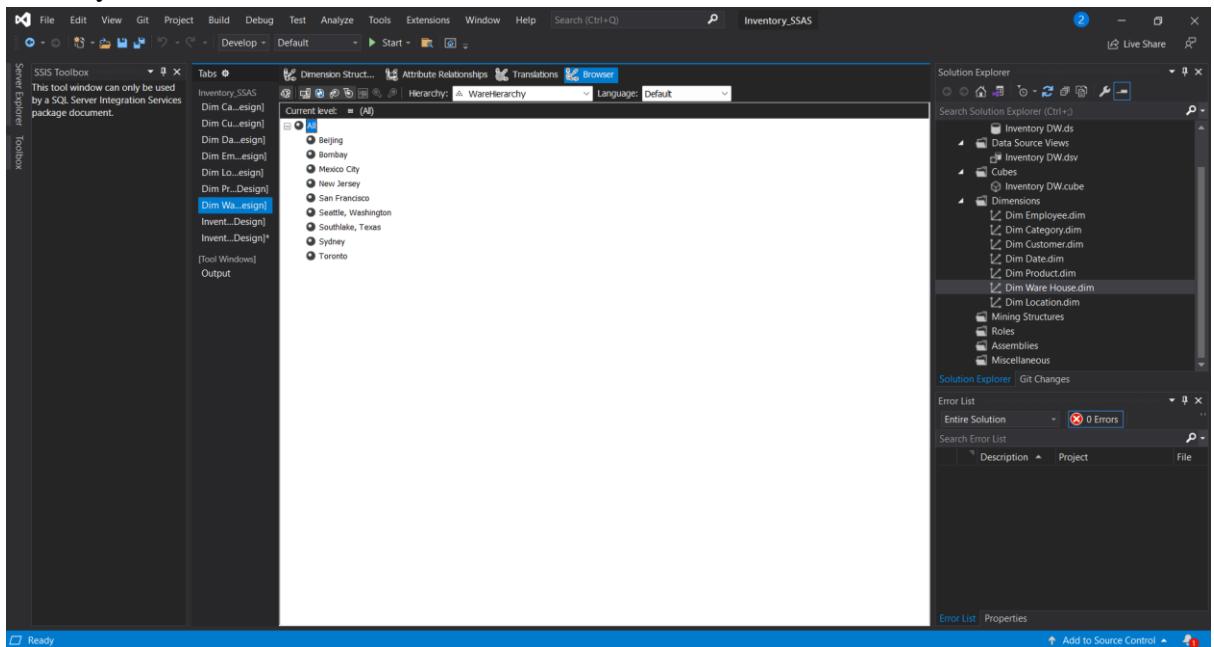
- Tạo WareHierarchy



- Process bảng Dim WareHouse

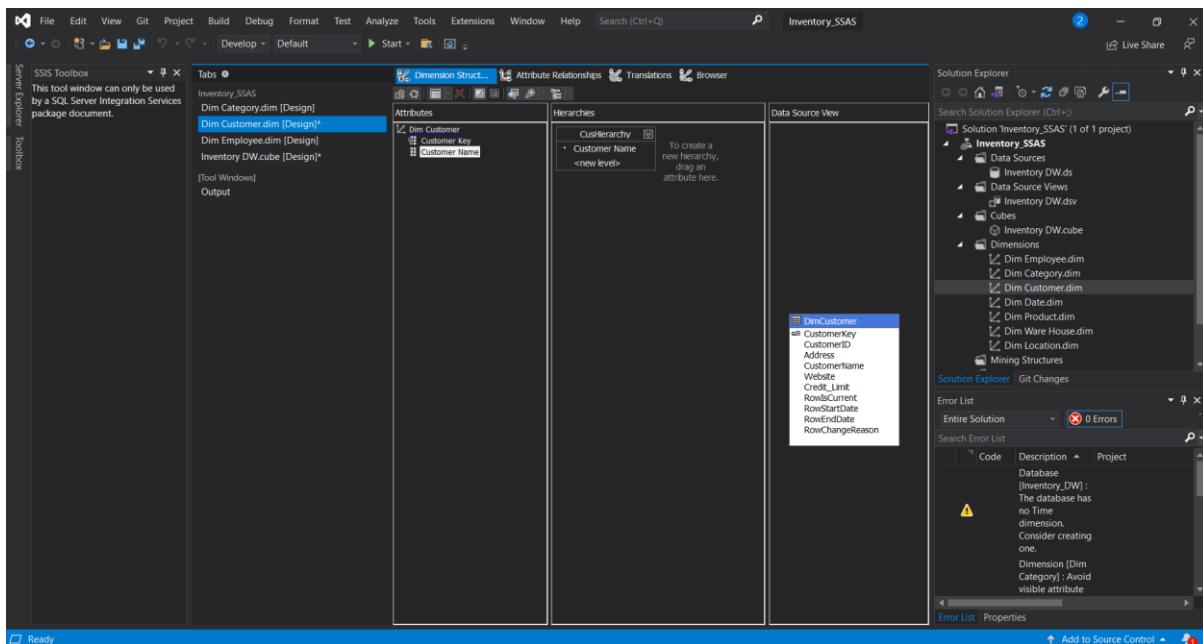


- Chuyển đến tab Brower để xem danh sách các kho:

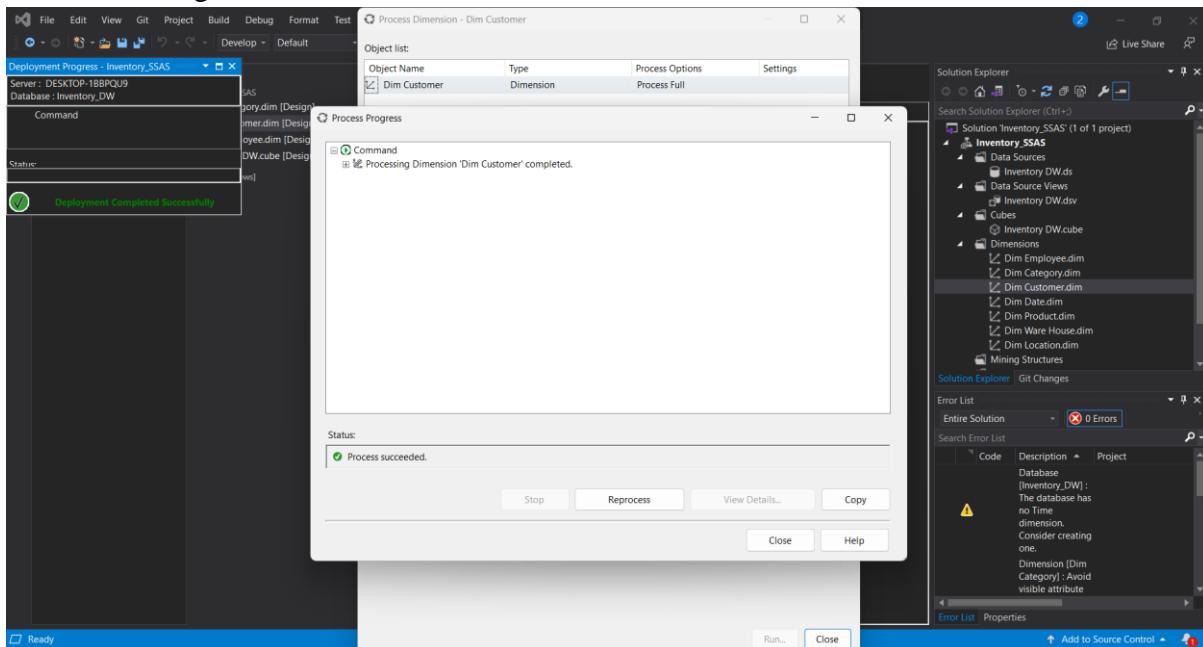


#### 4.3.7. Tạo Hierarchy cho Dim Customer

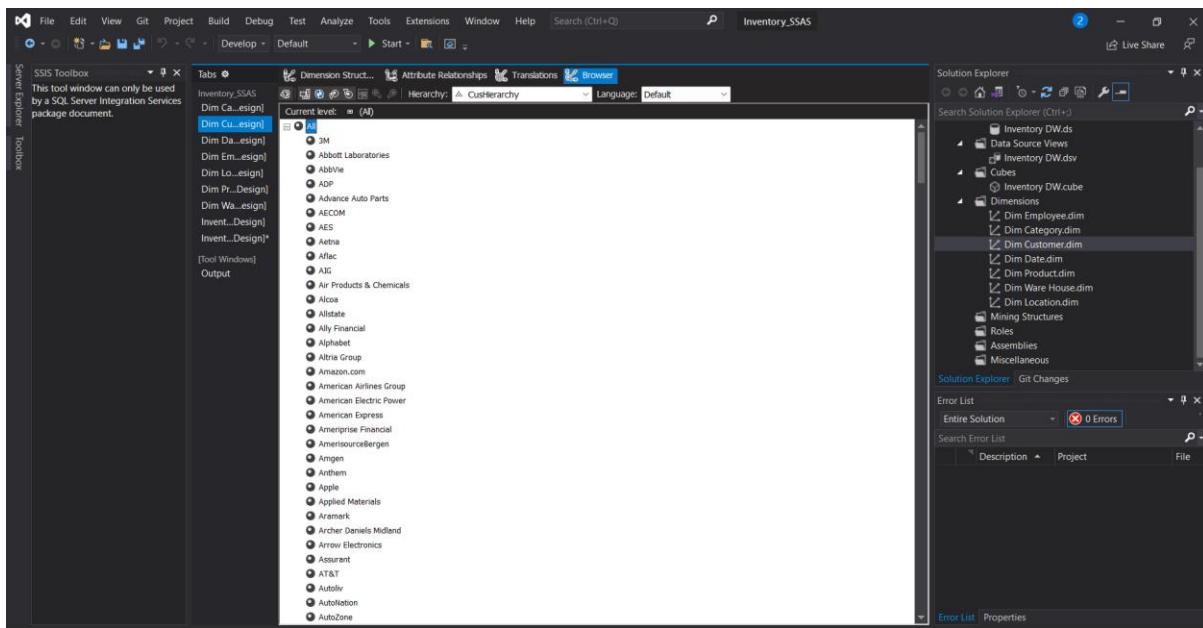
- Tạo CusHierarchy



### - Process bảng Dim Customer



### - Chuyển đến tab Brower để xem danh sách khách hàng:



## 4.4. Thực hiện phân tích dữ liệu

### 4.4.1. Câu hỏi: Số lượng các loại sản phẩm bán ra trong tháng/ quý/ năm.

#### 4.4.1.1. Sử dụng công cụ SSAS

- Hiển thị số lượng các mặt hàng, sản phẩm bán được từ trước đến nay

| Year | Category Name | Product Name                     | Quantity |
|------|---------------|----------------------------------|----------|
| 2013 | CPU           | Intel Xeon E5-2640 V3            | 141      |
| 2013 | Storage       | Corsair Dominator Platinum       | 118      |
| 2013 | Storage       | G.Skill Trident Z                | 35       |
| 2013 | Storage       | Western Digital WD500G1B0B       | 104      |
| 2015 | CPU           | Intel Core i7-3930K              | 90       |
| 2015 | CPU           | Intel Core i7-4770K              | 129      |
| 2015 | CPU           | Intel Core i7-4790K              | 87       |
| 2015 | CPU           | Intel Xeon E5-2630 V3            | 109      |
| 2015 | CPU           | Intel Xeon E5-2630 V4            | 137      |
| 2015 | CPU           | Intel Xeon E5-2640 V2            | 137      |
| 2015 | CPU           | Intel Xeon E5-2680 V3 (OEM/Trey) | 34       |
| 2015 | CPU           | Intel Xeon E5-2677W V4           | 84       |
| 2015 | CPU           | Intel Xeon E5-2690 V4            | 96       |
| 2015 | CPU           | Intel Xeon E5-2697 V2            | 138      |
| 2015 | Mother Board  | Asus K9PE-D16                    | 171      |
| 2015 | Mother Board  | Asus Sabertooth 990FX            | 106      |
| 2015 | Mother Board  | Asus TUF X299 MARK 1             | 45       |
| 2015 | Mother Board  | Asus X99-E WS/USB 3.1            | 130      |
| 2015 | Mother Board  | Asus Z10PE-D16 WS                | 75       |
| 2015 | Mother Board  | Gigabyte X299 AORUS Gaming 3     | 82       |
| 2015 | Mother Board  | MSI X99A XPOWER GAMING TITANI    | ???      |

The screenshot shows a Microsoft Excel spreadsheet titled "Power Pivot for Excel - Group7\_SSAS.xlsx". The main area displays a Power Pivot table with the following columns: Dim DateYear, Dim CategoryCategory Name, Dim ProductProductName, and MeasuresQuantity. The data includes rows for years 2013, 2014, 2015, and 2016, detailing various computer components like CPUs, Storage, and Mother Boards, along with their respective quantities.

- Nhìn vào hai hình trên ta thấy 2013 chủ yếu cửa hàng bán hai sản phẩm chính là CPU và Storage. Và từ năm 2015 trở về sau đã mở rộng thêm 2 mặt hàng gồm: Mother Board và Video Card.

#### 4.4.1.2. Sử dụng Pivot Table trong Excel

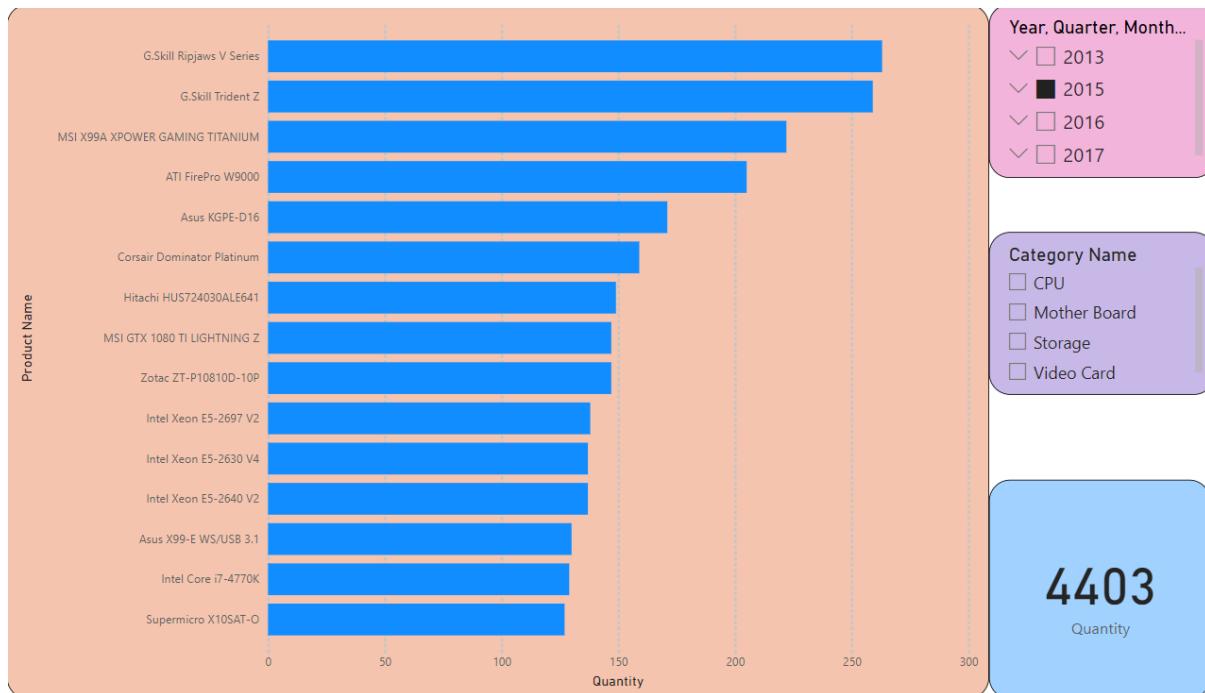
- Hiển thị theo thời gian năm/ quý/ tháng

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable set up. The PivotTable Fields pane is visible on the right side, showing fields from the 'inventory\_fact' and 'sales\_fact' tables. The PivotTable itself displays a summary of 'Quantity' by 'Year' and 'Category'. The PivotTable Fields pane shows 'DateHierarchy' under Rows and 'Quantity' under Values.

- Với hình trên thì ta cũng có thể thấy được số lượng hàng bán ra hiển thị theo năm. Nhìn vào đây có thể thấy rằng số lượng mặt hàng bán ra vào năm 2013 còn khá ít và sau đó tăng dần theo thời gian.

#### 4.4.1.3. Sử dụng Power BI

- Biểu đồ thể hiện các sản phẩm có số lượng được bán nhiều nhất theo danh mục sản phẩm và theo thời gian( Năm, quý, tháng, ngày)



- Ở biểu đồ trên, có thể thấy vào năm 2015 cửa hàng bán được tổng 4403 sản phẩm, trong đó 2 sản phẩm bán chạy nhất là *G.Skill Ripjaws V Series* và *G.Skill Trident Z*. Với mỗi sản phẩm bán được hơn 250 cái trên tổng số 4403 mặt hàng được bán ra.

#### 4.4.2. Câu hỏi: Cho biết doanh số các sản phẩm trong tháng/ quý/ năm.

##### 4.4.2.1. Sử dụng công cụ SSAS

- Doanh số của các sản phẩm

Table Import Wizard

Dimension <Select dimension>

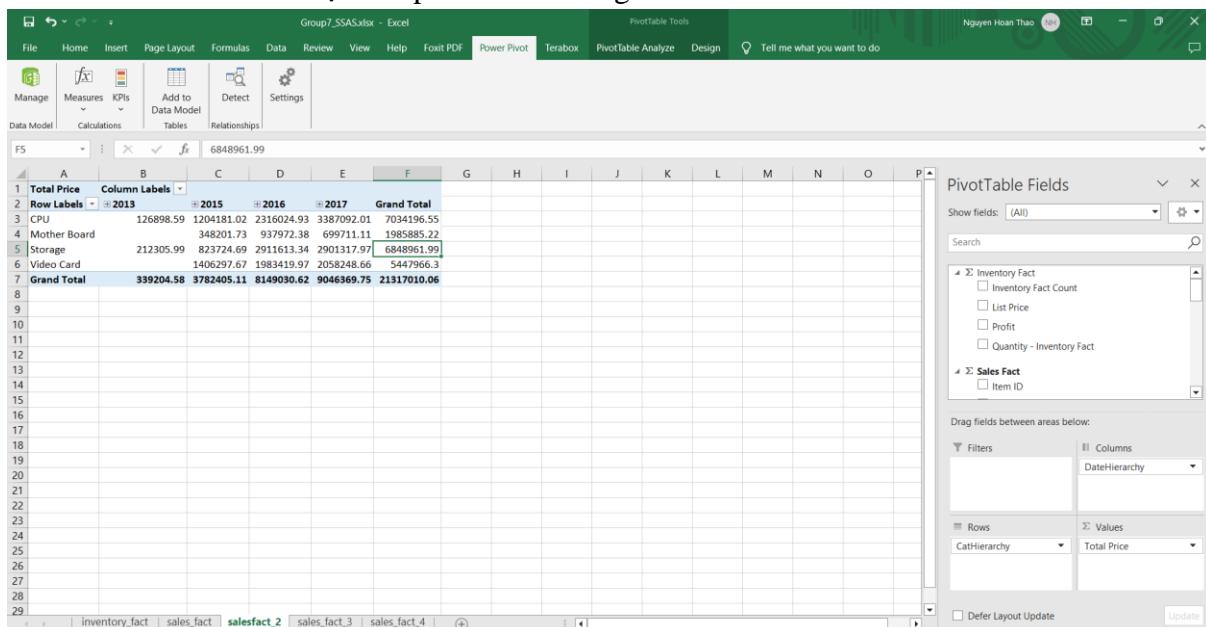
| Category Name | Product Name                     | Year | Total Price |
|---------------|----------------------------------|------|-------------|
| CPU           | AMD Opteron 6378                 | 2016 | 110816.66   |
| CPU           | Intel Core 2 Extreme QX9775      | 2016 | 28544       |
| CPU           | Intel Core 2 Extreme QX9775      | 2017 | 45492       |
| CPU           | Intel Core i7-3930K              | 2015 | 59400       |
| CPU           | Intel Core i7-3930K              | 2016 | 89100       |
| CPU           | Intel Core i7-3930K              | 2017 | 43560       |
| CPU           | Intel Core i7-3960X Extreme E... | 2017 | 152941.34   |
| CPU           | Intel Core i7-4770K              | 2015 | 103071      |
| CPU           | Intel Core i7-4790K              | 2015 | 54022.65    |
| CPU           | Intel Core i7-4790K              | 2016 | 70788.3     |
| CPU           | Intel Core i7-4790K              | 2017 | 36015.1     |
| CPU           | Intel Core i7-5930K              | 2017 | 42734.23    |
| CPU           | Intel Core i7-5960X              | 2017 | 91890.89    |
| CPU           | Intel Core i7-6950X (OEM/Tray)   | 2017 | 248838.02   |
| CPU           | Intel Core i7-7820X              | 2016 | 70590       |
| CPU           | Intel Core i7-7820X              | 2017 | 58372.5     |
| CPU           | Intel Core i9-7900X              | 2017 | 213207.93   |
| CPU           | Intel Xeon E5-1650 V3            | 2016 | 72870.81    |
| CPU           | Intel Xeon E5-1650 V3            | 2017 | 75130.37    |
| CPU           | Intel Xeon E5-1650 V4 (OEM/...   | 2017 | 63068.94    |
| CPU           | Intel Xeon E5-1660 V3 (OEM/...   | 2016 | 47939.53    |

The screenshot shows a PivotTable in Microsoft Excel. The columns are labeled: Dim Category, Category Name, Dim Product, Product Name, Dim Date, Year, and Measures, Total Price. The data includes items like G.Skill Ripjaws V Series, ATI FirePro W9000, Intel Xeon E5-2687W V4, and various storage and video cards. The total price for G.Skill Ripjaws V Series in 2016 is 825082.56.

- Hình trên cho thấy sản phẩm có doanh thu cao nhất trong các năm qua là sản phẩm *G.Skill Ripjaws V Series* với doanh thu khoảng 825082 đô la vào năm 2016.

#### 4.4.2.2. Sử dụng Pivot Table trong Excel

- Doanh số các danh mục sản phẩm theo thời gian



- Hình trên thể hiện doanh thu các mặt hàng theo năm có thể thấy được Năm 2017 là năm có tổng doanh thu cao nhất với hơn 9 triệu đô la và mặt hàng có doanh số cao nhất chính là CPU với hơn 3,3 triệu đô la.

#### 4.4.2.3. Sử dụng Power BI

- Top 10 sản phẩm có doanh số cao nhất trong năm



- Nhìn vào biểu đồ top 10 sản phẩm có doanh thu cao nhất trong năm 2016, ta có thể thấy năm 2016 có doanh thu hơn 8 triệu đô và bán được gần 11 ngàn sản phẩm. Sản phẩm *G.Skill Ripjaws V Series* là sản phẩm có doanh số cao nhất vào năm 2016 với hơn 800 nghìn đô la.

- Ngoài ra chúng ta có thể xem mặt hàng có sản phẩm bán chạy nhất theo thời gian bằng bộ lọc.



- Ta có thể thấy vào năm 2016 phân loại theo mặt hàng CPU thì sản phẩm *Intel Xeon E5 – 2695 V2* là sản phẩm có doanh số cao nhất với hơn 400 nghìn đô la thu về.

#### 4.4.3. Câu hỏi: Cho biết doanh số bán hàng của từng nhân viên trong tháng/ quý/ năm.

##### 4.4.3.1. Sử dụng công cụ SSAS

- Hiển thị doanh số bán hàng của các nhân viên theo thời gian.

Power Pivot for Excel - Group7\_SSAS.xlsx

The screenshot shows a Microsoft Excel spreadsheet titled "Power Pivot for Excel - Group7\_SSAS.xlsx". The ribbon at the top has tabs for File, Home, Design, Advanced, and Power Pivot. The Home tab is selected. The Power Pivot ribbon tab has sections for Paste Append, Paste Replace, From Database, From Data From Other Sources, Existing Connections, Refresh, and PivotTable. Below these are buttons for Data Type (Formatting), Sort and Filter, Find, and Calculations. The main area is a data grid titled "Dim EmployeeEmployee NameEmployee Name" with columns "Dim EmployeeEmployee NameEmployee Name" and "Dim DateHierarchyYear". The grid contains 21 rows of data. At the bottom, there are tabs for Query, Query1, Query2, and Query3, with Query3 selected. The status bar at the bottom shows "Record 1 of 24".

|    | Dim EmployeeEmployee NameEmployee Name | Dim DateHierarchyYear | MeasuresTotal Price | Add Column |
|----|----------------------------------------|-----------------------|---------------------|------------|
| 1  | ChloeCruz                              | 2015                  | 1718159.03          |            |
| 2  | ChloeCruz                              | 2016                  | 558540.87           |            |
| 3  | ChloeCruz                              | 2017                  | 132014.97           |            |
| 4  | DaisyOrtiz                             | 2016                  | 1269901.78          |            |
| 5  | DaisyOrtiz                             | 2017                  | 1335160.25          |            |
| 6  | EvieHarrison                           | 2016                  | 1081005.11          |            |
| 7  | EvieHarrison                           | 2017                  | 763531.28           |            |
| 8  | FlorenceFreeman                        | 2013                  | 339204.58           |            |
| 9  | FlorenceFreeman                        | 2015                  | 594289.59           |            |
| 10 | FlorenceFreeman                        | 2016                  | 753747.22           |            |
| 11 | FlorenceFreeman                        | 2017                  | 417010.25           |            |
| 12 | FreyaGomez                             | 2015                  | 692918.2            |            |
| 13 | FreyaGomez                             | 2016                  | 1861074.32          |            |
| 14 | FreyaGomez                             | 2017                  | 2689267.6           |            |
| 15 | GraceEllis                             | 2015                  | 658190.94           |            |
| 16 | GraceEllis                             | 2016                  | 72888.11            |            |
| 17 | GraceEllis                             | 2017                  | 1212427.33          |            |
| 18 | IsabelleMarshall                       | 2016                  | 748831.94           |            |
| 19 | IsabelleMarshall                       | 2017                  | 1216380.26          |            |
| 20 | LilyFisher                             | 2016                  | 559381.75           |            |
| 21 | LilyFisher                             | 2017                  | 648867.66           |            |

- Ta có thể thấy doanh số bán hàng của từng nhân viên qua các năm.

#### 4.4.3.2. Sử dụng Pivot Table trong Excel

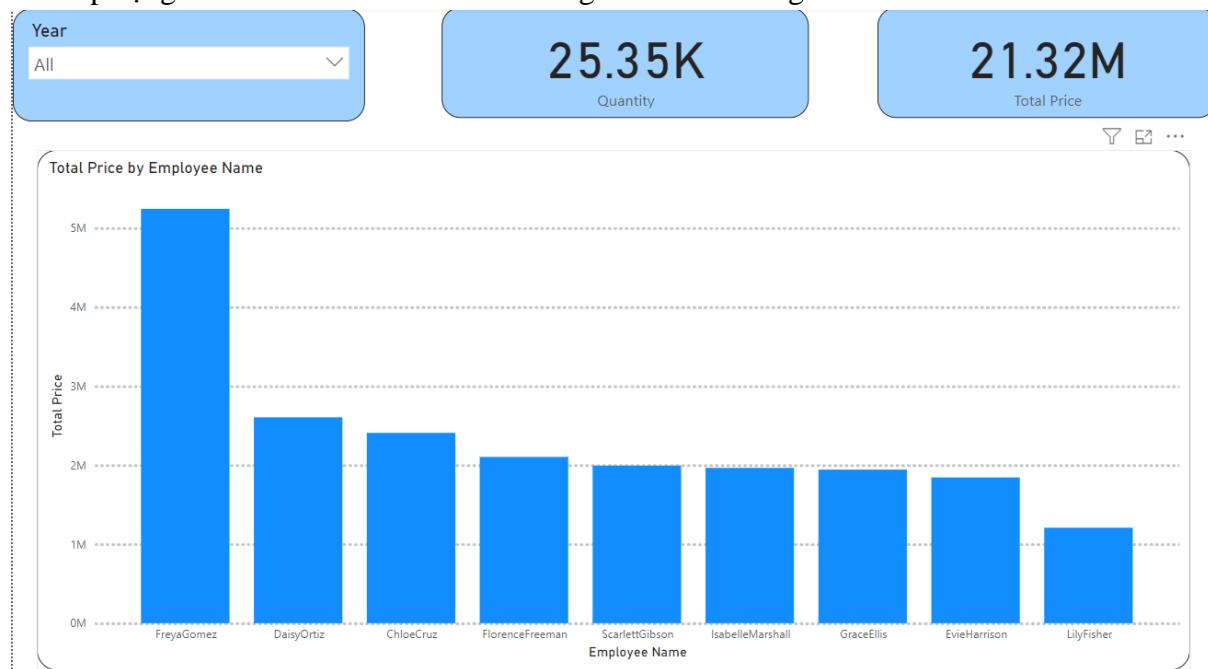
The screenshot shows a Microsoft Excel spreadsheet titled "Group7\_SSAS.xlsx - Excel". The PivotTable Fields pane on the right side lists fields from three fact tables: Inventory Fact, Sales Fact, and Salesfact\_2. The rows are grouped by Employee Name (EmpHierarchy) and the values are grouped by DateHierarchy. The PivotTable grid displays the total price for each employee across the years 2015, 2016, and 2017, with a Grand Total row.

|                    | 2015             | 2016              | 2017              | Grand Total        |
|--------------------|------------------|-------------------|-------------------|--------------------|
| ChloeCruz          | 1718159.03       | 558540.87         | 132014.97         | 2408714.87         |
| DaisyOrtiz         | 1269901.78       | 1335160.25        | 2605062.03        |                    |
| EvieHarrison       | 1081005.11       | 763531.28         | 1844536.39        |                    |
| FlorenceFreeman    | 339204.58        | 594289.59         | 753747.22         | 2104251.64         |
| FreyaGomez         | 692918.2         | 1861074.32        | 2689267.6         | 5243260.12         |
| GraceEllis         | 658190.94        | 72888.11          | 1212427.33        | 1943506.38         |
| IsabelleMarshall   | 748831.94        | 1216380.26        | 1965212.2         |                    |
| LilyFisher         | 559381.75        | 648867.66         | 1208249.41        |                    |
| ScarlettGibson     | 118847.35        | 1243659.52        | 631710.15         | 1994217.02         |
| <b>Grand Total</b> | <b>339204.58</b> | <b>3782405.11</b> | <b>8149030.62</b> | <b>9046369.75</b>  |
|                    |                  |                   |                   | <b>21317010.06</b> |

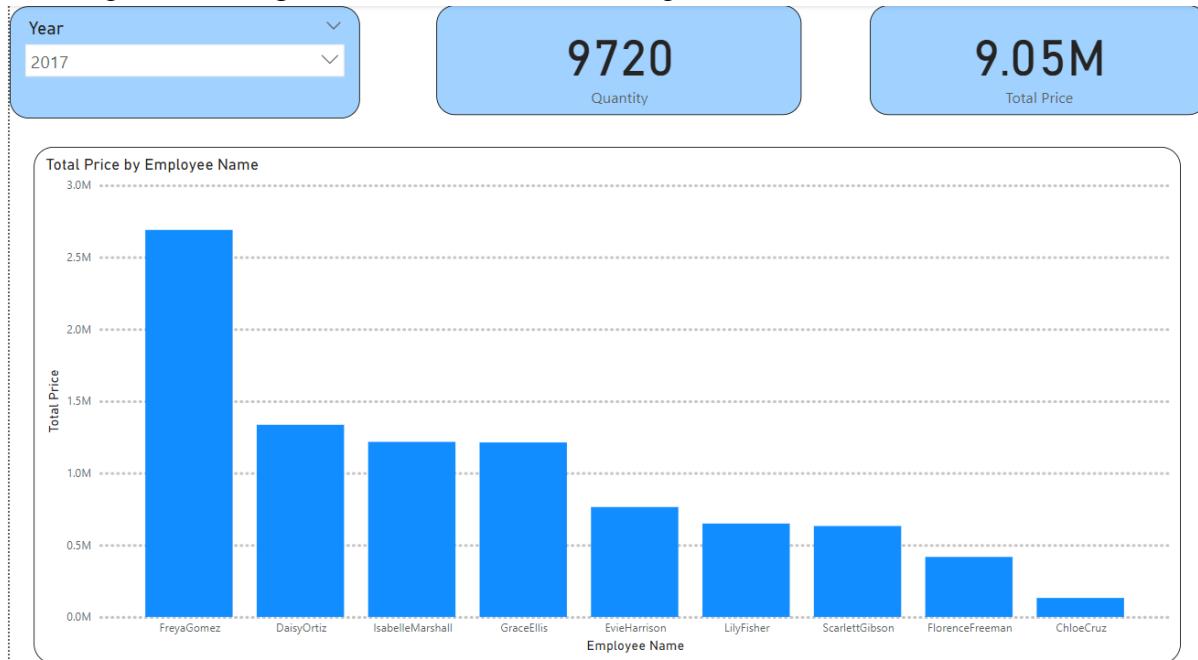
- Hình trên cho chúng ta thấy được doanh số bán hàng của các nhân viên qua từng năm, nhân thấy năm 2016 và 2017 là hai năm các nhân viên có doanh số bán hàng vượt trội hơn các năm còn lại.

#### 4.4.3.3. Sử dụng Power BI

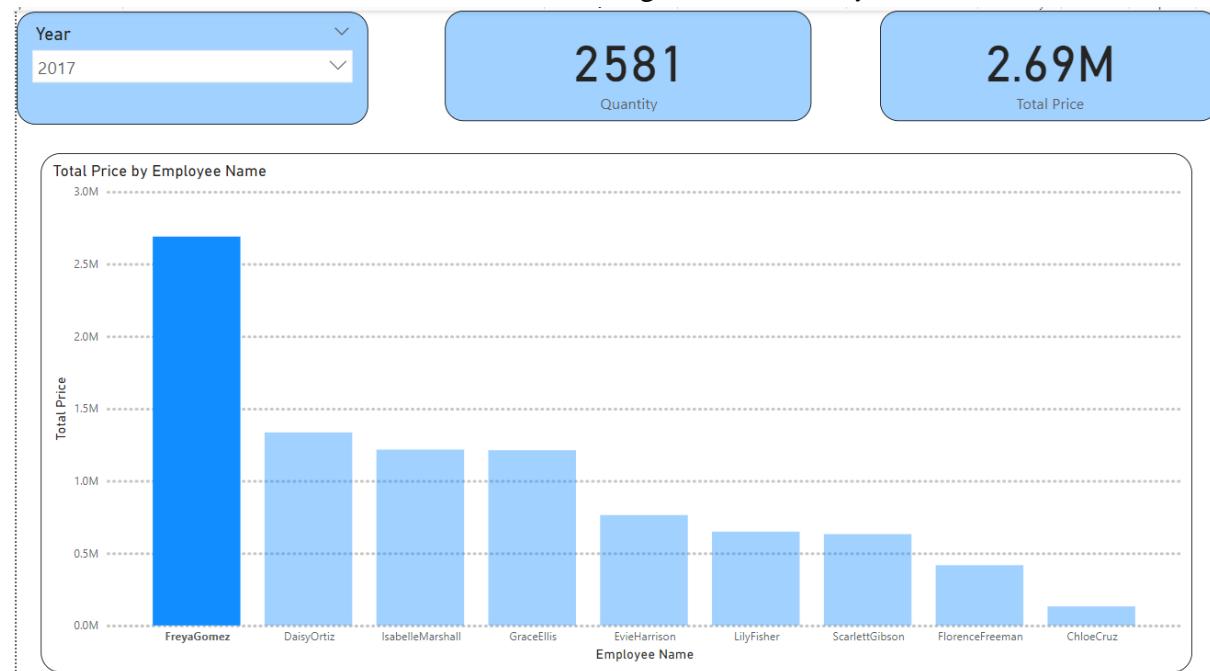
- Xếp hạng nhân viên có doanh số bán hàng theo thời gian.



- Dùng bộ lọc thời gian để xem doanh số bán hàng của các nhân viên vào năm 2017



- Vào năm 2017 nhân viên có doanh số bán hàng cao nhất là *FreyaGomez*:



- Dựa vào các số liệu thể hiện trong biểu đồ cửa hàng có thể dễ dàng xác định được các nhân viên có doanh số bán hàng cao nhất trong năm để tiến hành khen thưởng vào dịp cuối năm cũng như cân nhắc về lương thưởng cho họ, điều này sẽ giúp sự cạnh tranh tích cực để nhân viên tiếp tục phát huy và cố gắng hơn trong công việc.

## 4.4.4. Câu hỏi: Chi tiêu của khách hàng theo thời gian

### 4.4.4.1. Sử dụng công cụ SSAS

- Số liệu biểu diễn chi tiêu của khách hàng theo thời gian.

The screenshot shows the Microsoft Excel interface with the Power Pivot ribbon selected. A 'Table Import Wizard' dialog box is open, showing the 'Dimension' and 'Hierarchy' sections. The dimension selected is 'Customer Name'. The hierarchy tree shows various dimensions like 'Inventory DW', 'Measures', and 'Category Name'. Below the hierarchy, there is a table with data for 'Customer Name', 'Category Name', 'Year', and 'Total Price'. The main Excel window shows a pivot table with the following data:

| Dim Customer  | Customer Name | Category Name | Year      | Measures |
|---------------|---------------|---------------|-----------|----------|
| 1 AbbVie      | CPU           | 2016          | 99130.81  |          |
| 2 AbbVie      | CPU           | 2017          | 89609.13  |          |
| 3 AbbVie      | Mother Board  | 2017          | 104052.26 |          |
| 4 AbbVie      | Storage       | 2016          | 4784.15   |          |
| 5 AbbVie      | Storage       | 2017          | 377768.78 |          |
| 6 Alflac      | Storage       | 2016          | 399575.21 |          |
| 7 Alflac      | Storage       | 2017          | 10491.6   |          |
| 8 Alflac      | Video Card    | 2016          | 93396.42  |          |
| 9 Alcoa       | CPU           | 2015          | 240254.87 |          |
| 10 Alcoa      | CPU           | 2016          | 408416.71 |          |
| 11 Alcoa      | Mother Board  | 2015          | 51109.74  |          |
| 12 Alcoa      | Mother Board  | 2016          | 31876.22  |          |
| 13 Alcoa      | Storage       | 2015          | 103933.95 |          |
| 14 Alcoa      | Storage       | 2016          | 409414.37 |          |
| 15 Alcoa      | Video Card    | 2016          | 191018.12 |          |
| 16 AutoNation | CPU           | 2017          | 254598.66 |          |
| 17 AutoNation | Mother Board  | 2016          | 43919.04  |          |
| 18 AutoNation | Storage       | 2016          | 115980.25 |          |
| 19 AutoNation | Storage       | 2017          | 334456.71 |          |
| 20 AutoNation | Video Card    | 2016          | 99448.83  |          |
| 21 AutoNation | Video Card    | 2017          | 99748.67  |          |

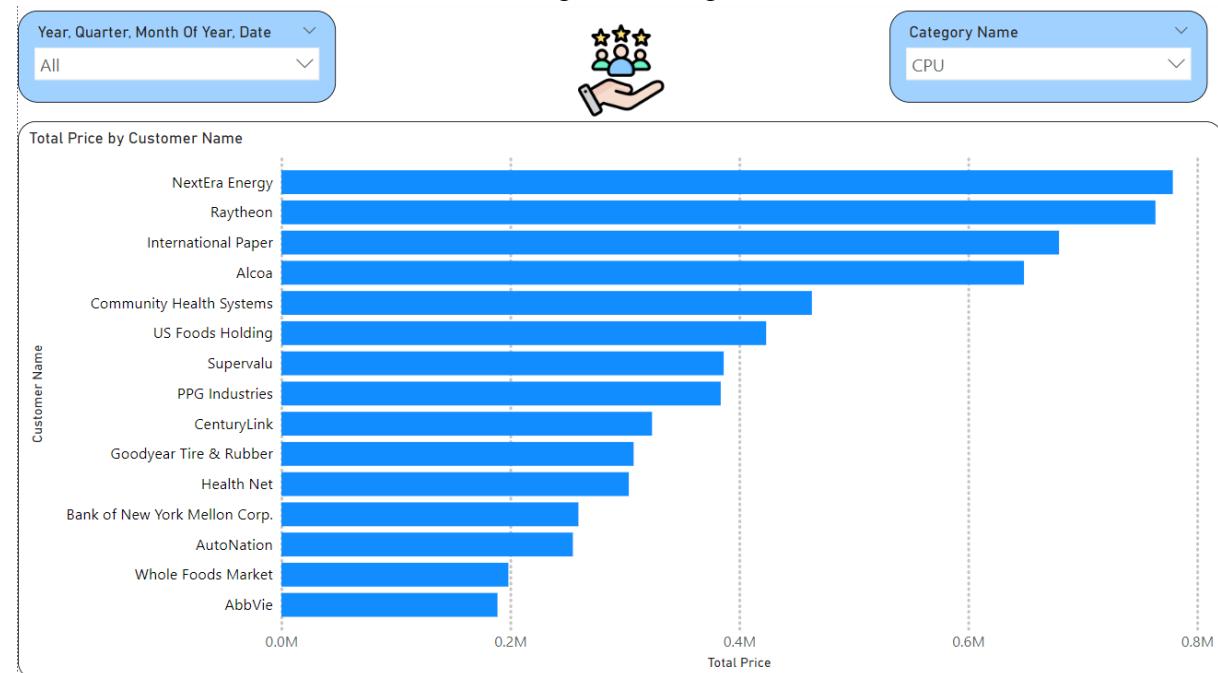
#### 4.4.4.2. Sử dụng Pivot Table trong Excel

The screenshot shows a Microsoft Excel spreadsheet titled "Group7\_SSAS.xlsx - Excel". The PivotTable Fields pane on the right side is open, displaying fields such as "Dim Category", "CatHierarchy", "Category Name", and "Dim Customer". The main table area shows sales data for companies like AbbVie, Aflac, Alcoa, AutoNation, Baker Hughes, and Bank of New York Mellon Corp. The table includes columns for "Total Price" (Column Labels) and years 2013, 2015, 2016, 2017, and Grand Total.

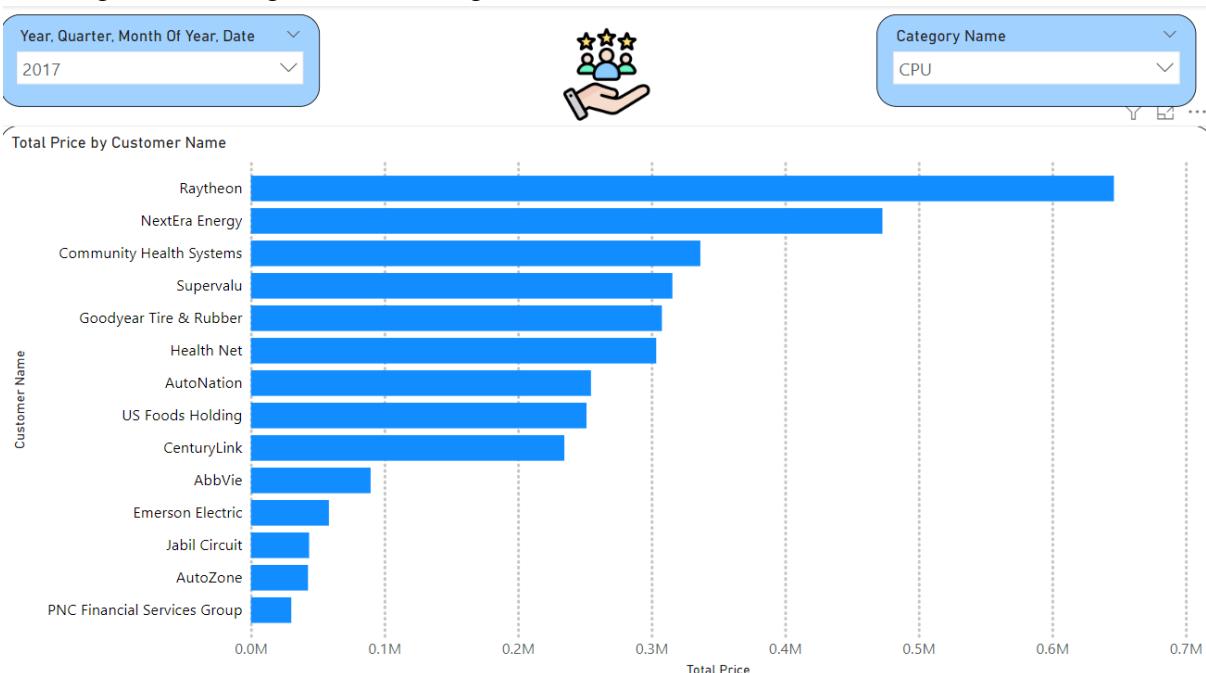
|    |                               | Column Labels | 2013      | 2015      | 2016      | 2017 | Grand Total |
|----|-------------------------------|---------------|-----------|-----------|-----------|------|-------------|
| 1  | Total Price                   |               |           |           |           |      |             |
| 2  | Row Labels                    |               | 2013      | 2015      | 2016      | 2017 | Grand Total |
| 3  | AbbVie                        |               |           |           |           |      |             |
| 4  | CPU                           |               | 99130.81  | 89609.13  | 188739.94 |      |             |
| 5  | Mother Board                  |               |           | 104052.26 | 104052.26 |      |             |
| 6  | Storage                       |               | 4784.15   | 377768.78 | 382552.93 |      |             |
| 7  | Aflac                         |               |           |           |           |      |             |
| 8  | Storage                       |               | 399575.21 | 10491.6   | 410066.81 |      |             |
| 9  | Video Card                    |               |           | 93396.42  | 93396.42  |      |             |
| 10 | Alcoa                         |               |           |           |           |      |             |
| 11 | CPU                           | 240254.87     | 408416.71 | 648671.58 |           |      |             |
| 12 | Mother Board                  | 51109.74      | 31876.22  | 82985.96  |           |      |             |
| 13 | Storage                       | 103933.95     | 409414.37 | 513348.32 |           |      |             |
| 14 | Video Card                    |               | 191018.12 | 191018.12 |           |      |             |
| 15 | AutoZone                      |               |           |           |           |      |             |
| 16 | CPU                           |               | 254598.66 | 254598.66 |           |      |             |
| 17 | Mother Board                  |               | 43919.04  | 43919.04  |           |      |             |
| 18 | Storage                       |               | 115980.25 | 334456.71 | 450436.96 |      |             |
| 19 | Video Card                    |               | 99448.83  | 99748.67  | 199197.5  |      |             |
| 20 | Baker Hughes                  |               |           |           |           |      |             |
| 21 | CPU                           |               | 42734.23  | 42734.23  |           |      |             |
| 22 | Video Card                    |               | 90497.16  | 90497.16  |           |      |             |
| 23 | Bank of New York Mellon Corp. |               |           |           |           |      |             |
| 24 | Storage                       |               | 102658.74 | 102658.74 |           |      |             |
| 25 | Video Card                    |               | 367198.64 | 367198.64 |           |      |             |
| 26 | CPU                           |               | 259421.63 | 259421.63 |           |      |             |
| 27 | Mother Board                  |               | 32899.3   | 32899.3   |           |      |             |
| 28 | Storage                       |               | 21171.21  | 21171.21  |           |      |             |
| 29 |                               |               |           |           |           |      |             |

#### 4.4.4.3. Sử dụng Power BI

- Biểu đồ thể hiện chi tiêu của khách hàng theo thời gian.



- Dùng bộ lọc thời gian và mặt hàng:



- Có thể thấy vào năm 2017 khách hàng Raytheon là khách hàng có mức chi tiêu cho mặt hàng CPU là cao nhất với khoảng 600 nghìn đô.
- Từ đây cửa hàng có thể tổ chức những chương trình tri ân khách hàng với mức chi tiêu tại cửa hàng cao nhất để kết nối thêm giữa cửa hàng và khách hàng, cũng như dựa vào tiêu chí của khách hàng để lên cái kế hoạch, chiến dịch chạy quảng cáo phù hợp với tệp khách hàng đang có.

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

### 4.4.5.1. Sử dụng công cụ SSAS

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

The screenshot shows two windows of Power Pivot for Excel. The top window is the 'Table Import Wizard' dialog, which is currently on the 'Select dimension' step. It displays a tree view of metadata from the 'Inventory DW' database, including dimensions like Country, WareHouse, Category, and Time, along with measures like Inventory Fact and Sales Fact. The bottom window shows the resulting PivotTable in the Power Pivot ribbon tab. The PivotTable has columns for Dim Location, Country Name, Country Name, Dim Ware House, Ware House Name, Dim Category, Category Name, Category Name, Dim Date, Year, and Measures. The data includes various products (CPU, Mother Board, Storage, Video Card) across locations (Australia, Canada, China) and years (2023, 2024). The total quantity for each row is listed in the 'Measures' column.

| Dim Location | Country Name | Country Name | Dim Ware House | Ware House Name | Dim Category | Category Name | Category Name | Dim Date | Year | Measures |
|--------------|--------------|--------------|----------------|-----------------|--------------|---------------|---------------|----------|------|----------|
| 1 Australia  |              | Australia    | Sydney         | CPU             |              |               |               |          | 2023 | 2791771  |
| 2 Australia  |              | Australia    | Sydney         | CPU             |              |               |               |          | 2024 | 941171   |
| 3 Australia  |              | Australia    | Sydney         | Mother Board    |              |               |               |          | 2023 | 2805687  |
| 4 Australia  |              | Australia    | Sydney         | Mother Board    |              |               |               |          | 2024 | 949431   |
| 5 Australia  |              | Australia    | Sydney         | Storage         |              |               |               |          | 2023 | 6012997  |
| 6 Australia  |              | Australia    | Sydney         | Storage         |              |               |               |          | 2024 | 2002245  |
| 7 Australia  |              | Australia    | Sydney         | Video Card      |              |               |               |          | 2023 | 2629904  |
| 8 Australia  |              | Australia    | Sydney         | Video Card      |              |               |               |          | 2024 | 859279   |
| 9 Canada     |              | Canada       | Toronto        | CPU             |              |               |               |          | 2023 | 962062   |
| 10 Canada    |              | Canada       | Toronto        | CPU             |              |               |               |          | 2024 | 322004   |
| 11 Canada    |              | Canada       | Toronto        | Mother Board    |              |               |               |          | 2023 | 1109862  |
| 12 Canada    |              | Canada       | Toronto        | Mother Board    |              |               |               |          | 2024 | 367397   |
| 13 Canada    |              | Canada       | Toronto        | Storage         |              |               |               |          | 2023 | 3154562  |
| 14 Canada    |              | Canada       | Toronto        | Storage         |              |               |               |          | 2024 | 1046972  |
| 15 Canada    |              | Canada       | Toronto        | Video Card      |              |               |               |          | 2023 | 2605688  |
| 16 Canada    |              | Canada       | Toronto        | Video Card      |              |               |               |          | 2024 | 865339   |
| 17 China     |              | China        | Beijing        | CPU             |              |               |               |          | 2023 | 2744928  |
| 18 China     |              | China        | Beijing        | CPU             |              |               |               |          | 2024 | 928649   |
| 19 China     |              | China        | Beijing        | Mother Board    |              |               |               |          | 2023 | 2172794  |
| 20 China     |              | China        | Beijing        | Mother Board    |              |               |               |          | 2024 | 724676   |
| 21 China     |              | China        | Beijing        | Storage         |              |               |               |          | 2023 | 5545849  |

#### 4.4.5.2. Sử dụng Pivot Table trong Excel

The screenshot shows a Microsoft Excel spreadsheet titled "Group7\_SSAS.xlsx". A PivotTable is being used to analyze data. The PivotTable Fields pane on the right lists fields from the "Inventory Fact" and "Sales Fact" tables, with "Quantity - Inventory Fact" checked as the value field. The main table displays product quantity by location and year. The data includes rows for Australia (Sydney, Toronto), Canada, China (Beijing), India (Bombay), and Mexico (Mexico City). The columns show Year (2023, 2024), Grand Total, and specific product categories like CPU, Mother Board, Storage, and Video Card.

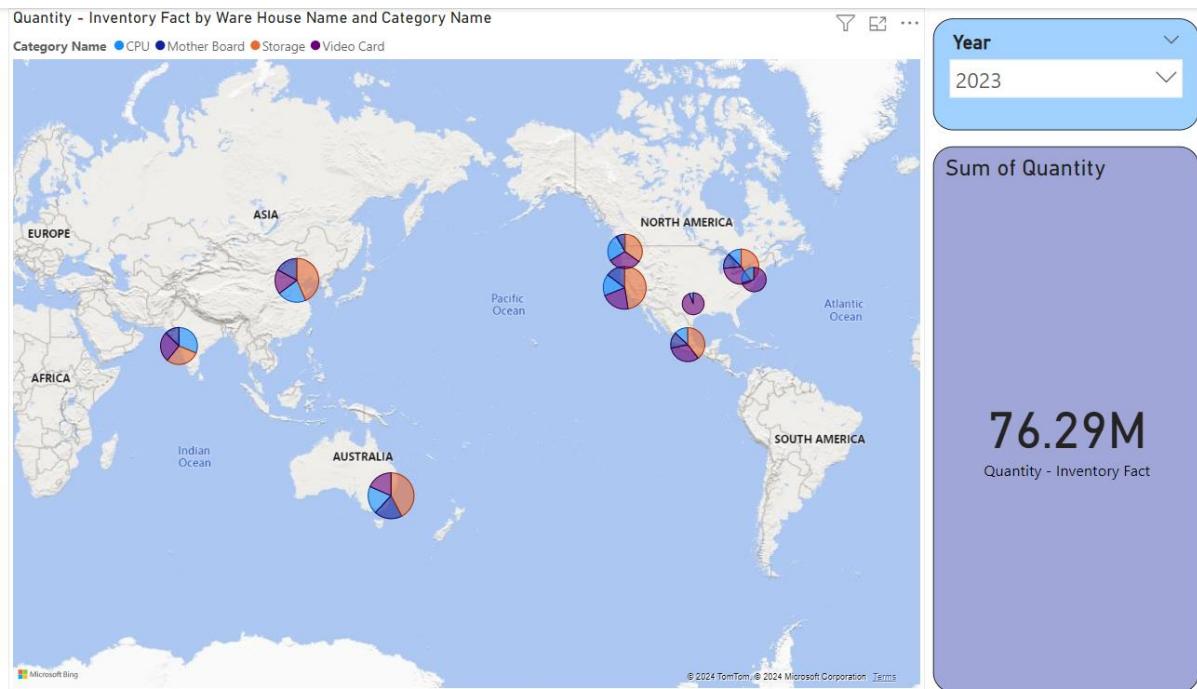
- Ta có thể thấy số lượng sản phẩm của các mặt hàng có trong các kho ở các thành phố thuộc các quốc gia như: Australia, Canada, China, India, Mexico, United States of America.

#### 4.4.5.3. Sử dụng Power BI

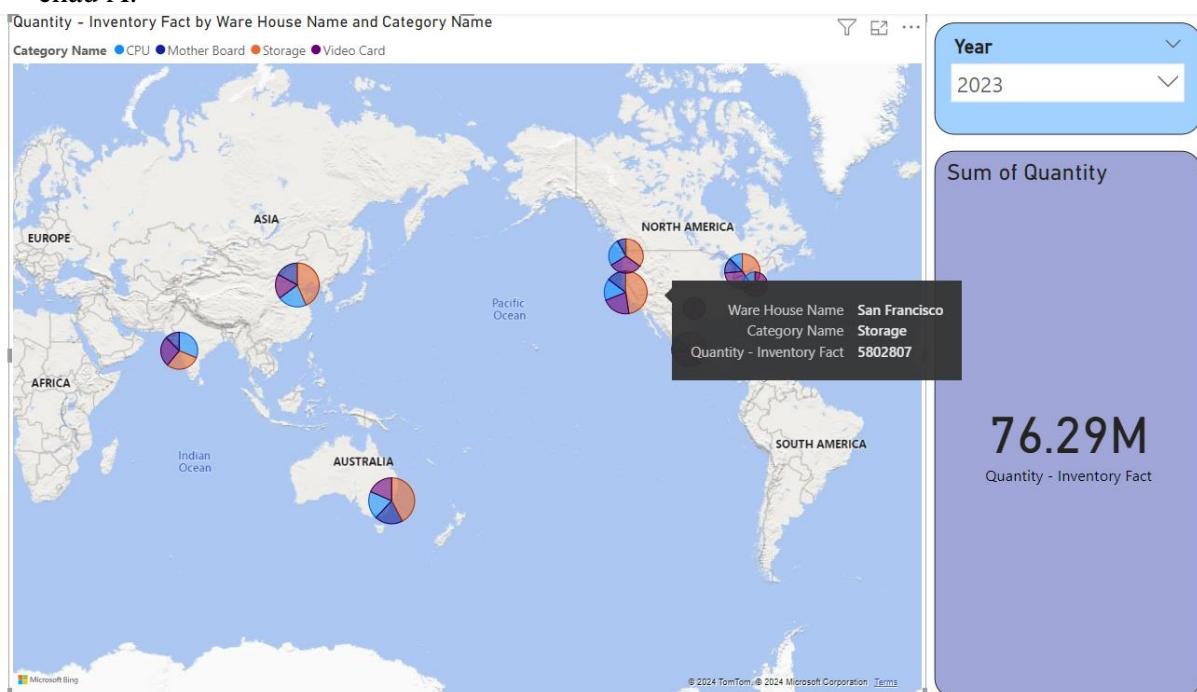
- Tạo biểu đồ thể hiện sự phân bố mặt hàng.



- Biểu đồ trên thể hiện sự phân bố của các mặt hàng ở Ấn độ vào tháng 10 quý 4 năm 2023.
- Dùng biểu đồ Map để chúng ta có cái nhìn rõ ràng hơn:



- Có thể thấy khu vực hoạt động của chuỗi cửa hàng tập trung ở hai châu lục là Châu Mỹ và châu Á.



- Ta có thể thấy số lượng mặt hàng Storage kho San Francisco vào năm 2023 là 5802807 sản phẩm.
- Dựa vào những điều này chúng ta có thể xem chi tiết sự phân bổ các mặt hàng đến các kho của các cửa hàng ở từng nước, từ đó xem xét từ nhiều yếu tố, nhu cầu... để phân bổ số lượng sản phẩm của các mặt hàng cho hợp lý.

# **CHƯƠNG 5: KẾT LUẬN**

## **5.1. Kết quả đạt được**

Thực hiện xây dựng ‘XÂY DỰNG KHO DỮ LIỆU CHO CHUỖI CỦA HÀNG BÁN LẺ PHẦN CỨNG’ giúp chúng em áp dụng thực tế được các kiến thức đã học trong quá trình học môn Kho dữ liệu, nắm vững kiến thức lý thuyết cơ bản phục vụ tốt cho việc thiết kế chương trình. Thiết kế được một kho dữ liệu tương đối hoàn chỉnh, hiểu và sử dụng được thêm được một số Cung cấp lại các kiến thức đã được học, đặc biệt là kỹ năng phân tích, giải quyết vấn đề. Nắm rõ các khái niệm cơ bản về kho dữ liệu và OLAP, các tính chất của một kho dữ liệu cần có.

Thành thạo trong việc sử dụng các công cụ hỗ trợ quá trình ETL đưa dữ liệu vào kho dữ liệu như công cụ SSIS, SSAS, các công cụ để trực quan hóa dữ liệu như Excel Pivot, Power BI Desktop.

## **5.2. Những hạn chế**

Do thời gian hạn ngắn cộng với khối lượng công việc nhiều nên trong quá trình thực hiện đồ án nhóm còn gặp phải một số vấn đề :

- + Kiến thức thực tế và kiến thức tư duy phân tích của chúng em còn hạn chế, nên tính chuyên nghiệp của đề tài chưa cao
- + Khả năng xử lý nghiệp vụ chưa linh hoạt nhạy bén.

## **5.3. Tài liệu tham khảo**

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