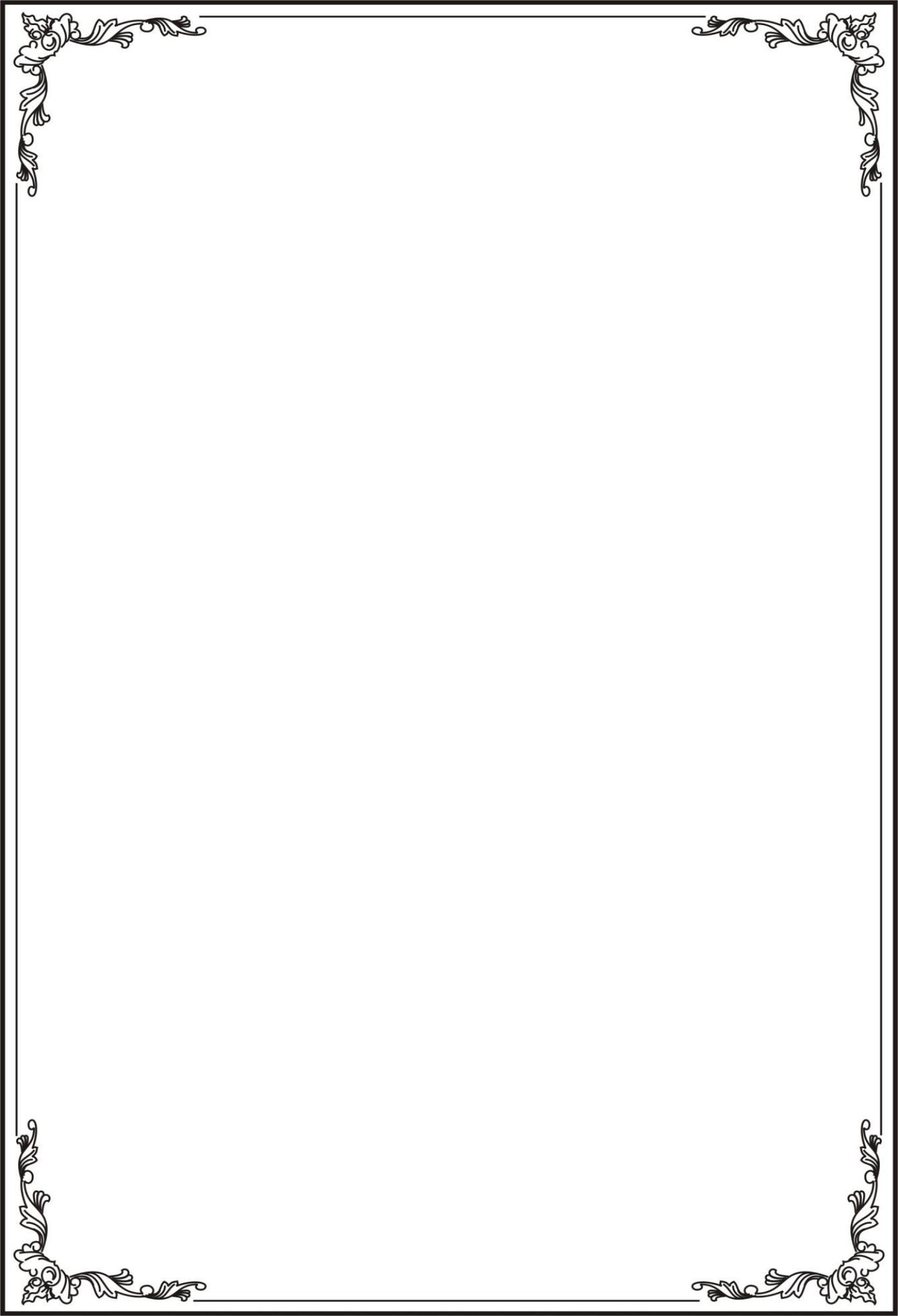
|  |  |  |
| --- | --- | --- |
| **Người hướng dẫn:** | **Th.s Nguyễn Sỹ Thiệu** | |
|  |  | |
| **Sinh viên thực hiện:** | **Nguyễn Như Chiến - 7133112076** | |
|  | |

**BÁO CÁO**

**HÀ NỘI, 2024**

2



**QUẢN TRỊ DỮ LIỆU LỚN VỚI APACHE SPARK**



**HỌC VIỆN CHÍNH SÁCH VÀ PHÁT TRIỂN**

**KHOA KINH TẾ SỐ**

-----















-----

Mục lục

**[I. Bài toán](#_Toc195436470)** [1](#_Toc195436470)

***[Mục tiêu](#_Toc195436471)*** [1](#_Toc195436471)

**[II. Cơ sở dữ liệu cho bài toán](#_Toc195436472)** [1](#_Toc195436472)

[2.1. Mô hình kiến trúc tổng thể 2](#_Toc195436473)

[2.2. Nhóm bảng sự kiện (Fact Tables) 2](#_Toc195436474)

[2.3. Nhóm bảng chiều (Dimension Tables) 3](#_Toc195436475)

[2.4. Nhóm bảng phụ trợ / trung gian (Staging & Lookup) 5](#_Toc195436476)

[2.5. Mô hình liên kết giữa các bảng 5](#_Toc195436477)

**[III. Xây dựng luồng tích hợp dữ liệu (ETL) bằng SQL và SSIS](#_Toc195436478)** [6](#_Toc195436478)

[Bước 1: Dọn sạch dữ liệu cũ – Execute SQL Task 6](#_Toc195436479)

[Bước 2: Trích xuất dữ liệu từ nguồn – Get Data from Sources 8](#_Toc195436480)

[Bước 3: Biến đổi dữ liệu – Transform Data 18](#_Toc195436481)

**[Giải thích SCD Type 1](#_Toc195436482)** [20](#_Toc195436482)

**[Xử lý thay đổi với SCD Type 2](#_Toc195436483)** [22](#_Toc195436483)

[Bước 4: Nạp dữ liệu vào Data Warehouse – Load to DW 24](#_Toc195436484)

**[IV. Sử dụng CSDL tạo báo cáo](#_Toc195436485)** [25](#_Toc195436485)

**[IV. Phân tích báo cáo](#_Toc195436486)** [30](#_Toc195436486)

# **I. Bài toán**

Trong bối cảnh thị trường ngày càng cạnh tranh, các doanh nghiệp cần không chỉ bán được sản phẩm mà còn phải hiểu rõ hành vi tiêu dùng, sở thích và giá trị của từng nhóm khách hàng. Đồng thời, việc nắm bắt hiệu suất bán hàng theo khu vực địa lý, sản phẩm và từng nhân viên cũng là yếu tố then chốt để ra quyết định đúng đắn, tối ưu hóa hoạt động kinh doanh.

Bài toán này mô phỏng một hệ thống phân tích dữ liệu kinh doanh dựa trên bộ dữ liệu mẫu AdventureWorks 2022 – một cơ sở dữ liệu do Microsoft phát hành, tái hiện hoạt động bán hàng, sản xuất và phân phối của một công ty kinh doanh. Dữ liệu phản ánh một hệ thống kinh doanh thực tế, gồm nhiều đối tượng: khách hàng, nhân viên, sản phẩm, đơn hàng, vùng lãnh thổ, v.v.

## ***Mục tiêu***

Bài toán hướng tới việc xây dựng một hệ thống phân tích dữ liệu hoàn chỉnh, đóng vai trò như một trợ lý ra quyết định cho doanh nghiệp. Cụ thể:

* Hiểu rõ hành vi và phân khúc khách hàng: Ai đang mua hàng? Mua bao nhiêu? Mua loại sản phẩm gì? Bao lâu mua một lần?
* Đánh giá hiệu quả sản phẩm và khu vực bán hàng: Sản phẩm nào bán chạy? Khu vực nào tạo ra doanh thu cao? Có sự khác biệt theo thời gian hay không?
* Phân tích hiệu suất đội ngũ kinh doanh: Nhân viên nào có doanh số tốt? Ai đang phụ trách khu vực nào? Hiệu quả bán hàng theo từng người ra sao?

Các bước triển khai

Để thực hiện mục tiêu trên, nhóm đã thiết kế quy trình gồm 4 giai đoạn chính:

1. Trích xuất dữ liệu từ hệ quản trị cơ sở dữ liệu SQL Server – nơi lưu trữ toàn bộ dữ liệu giao dịch, sản phẩm, khách hàng,...
2. Xử lý và biến đổi dữ liệu bằng công cụ SSIS (SQL Server Integration Services), bao gồm làm sạch, chuẩn hóa, và xây dựng các bảng trung gian (staging) để dễ dàng kiểm soát chất lượng dữ liệu.
3. Xây dựng mô hình Data Warehouse theo kiến trúc Snowflake Schema, đảm bảo tính linh hoạt khi phân tích đa chiều.
4. Trực quan hóa dữ liệu và phân tích chuyên sâu trên Power BI, tạo ra các dashboard hỗ trợ ra quyết định trực tiếp từ dữ liệu.

# **II. Cơ sở dữ liệu cho bài toán**

Hệ thống phân tích dữ liệu trong dự án được xây dựng từ cơ sở dữ liệu mẫu AdventureWorks 2022 – một cơ sở dữ liệu do Microsoft phát triển, mô phỏng quy trình hoạt động của một công ty sản xuất và phân phối sản phẩm. AdventureWorks 2022 cung cấp tập dữ liệu phong phú, phản ánh các hoạt động kinh doanh như: quản lý đơn hàng, khách hàng, sản phẩm, nhân viên và khu vực bán hàng.

## 2.1. Mô hình kiến trúc tổng thể

Kiến trúc tổng thể của hệ thống phân tích được thiết kế theo hướng hiện đại, bao gồm các thành phần sau:

* Nguồn dữ liệu: AdventureWorks 2022 (SQL Server)
* ETL (Extract - Transform - Load): thực hiện bằng SQL Server Integration Services (SSIS)
* Lưu trữ dữ liệu: theo mô hình Data Warehouse, sử dụng kiến trúc Snowflake Schema
* Công cụ trực quan hóa và phân tích: Power BI

Hệ thống dữ liệu trong Data Warehouse được tổ chức thành 3 nhóm bảng chính: bảng sự kiện (Fact Tables), bảng chiều (Dimension Tables) và bảng trung gian hỗ trợ (Staging/Lookup Tables).

## 2.2. Nhóm bảng sự kiện (Fact Tables)

Các bảng này lưu trữ các sự kiện giao dịch như đơn hàng và chi tiết đơn hàng – là trung tâm của các phân tích kinh doanh.

🔹Sales.SalesOrderHeader

Lưu thông tin tổng quát của từng đơn đặt hàng: khách hàng, nhân viên bán hàng, khu vực, và các chỉ số doanh thu.

|  |  |
| --- | --- |
| Cột | Mô tả |
| SalesOrderID | Mã đơn hàng |
| CustomerID | Mã khách hàng |
| CustomerAccountNumber | Mã tài khoản KH |
| SalesPersonID | Mã nhân viên bán hàng |
| TerritoryID | Mã khu vực bán hàng |
| TerritoryName | Tên khu vực |
| SubTotal | Tổng tiền hàng |
| TaxAmt | Thuế |
| Freight | Phí vận chuyển |
| TotalDue | Tổng thanh toán |
| OrderDate | Ngày đặt hàng |
| DueDate | Ngày đến hạn giao |

🔹 Sales.SalesOrderDetail

Chứa thông tin chi tiết về từng mặt hàng trong mỗi đơn hàng.

|  |  |
| --- | --- |
| Cột | Mô tả |
| SalesOrderID | Mã đơn hàng |
| SalesOrderDetailID | Mã dòng hàng |
| ProductID | Mã sản phẩm |
| OrderQty | Số lượng đặt |
| UnitPrice | Đơn giá |
| LineTotal | Thành tiền |

## 2.3. Nhóm bảng chiều (Dimension Tables)

Các bảng chiều cung cấp ngữ cảnh và đặc điểm cho các sự kiện, giúp phân tích theo các khía cạnh như sản phẩm, khách hàng, nhân viên, khu vực,...

🔹DimProducts  
 (Nguồn:Production.Product)  
 Thông tin chi tiết về sản phẩm.

|  |  |
| --- | --- |
| Cột | Mô tả |
| ProductID | Mã sản phẩm |
| Name | Tên sản phẩm |
| ProductNumber | Mã số sản phẩm |
| Color, StandardCost, ListPrice | Thuộc tính sản phẩm |
| ProductSubcategoryID | Nhóm phụ |

🔹DimProductSubcategory&DimProductCategory

(Nguồn: Production.ProductSubcategory, Production.ProductCategory)

Phân loại sản phẩm thành từng nhóm và danh mục

🔹 DimCustomers(Nguồn: Sales.Customer, Person.Person)

Thông tin khách hàng bao gồm tài khoản và chi tiết cá nhân.

|  |  |
| --- | --- |
| Cột | Mô tả |
| CustomerID | Mã khách hàng |
| AccountNumber | Mã tài khoản |
| PersonID | Tham chiếu đến bảng Person |
| TerritoryID | Mã khu vực khách hàng |
| FirstName, LastName, Title | Họ tên và danh xưng |

🔹 DimSalesPersons

(Nguồn: HumanResources.Employee, Sales.SalesPerson)

Thông tin nhân viên bán hàng.

|  |  |
| --- | --- |
| Cột | Mô tả |
| BusinessEntityID | Mã nhân viên |
| JobTitle | Chức vụ |
| HireDate | Ngày tuyển |
| Gender | Giới tính |

🔹 DimSalesTerritories

(Nguồn: Sales.SalesTerritory)

Thông tin về khu vực bán hàng.

|  |  |
| --- | --- |
| Cột | Mô tả |
| TerritoryID | Mã khu vực |
| Name | Tên khu vực |
| CountryRegionCode | Mã quốc gia |
| Group | Nhóm khu vực |

🔹 DimDate Bảng ngày chuẩn hóa giúp phân tích theo thời gian: năm, quý, tháng,...

(Tạo từ dữ liệu OrderDate, DueDate...)

|  |  |
| --- | --- |
| Cột | Mô tả |
| DateKey | Khóa ngày |
| FullDate | Ngày đầy đủ |
| Year, Quarter, Month, Weekday | Các thành phần thời gian |

## 2.4. Nhóm bảng phụ trợ / trung gian (Staging & Lookup)

Được sử dụng trong giai đoạn ETL để xử lý, làm sạch và chuẩn hóa dữ liệu trước khi nạp vào kho dữ liệu.

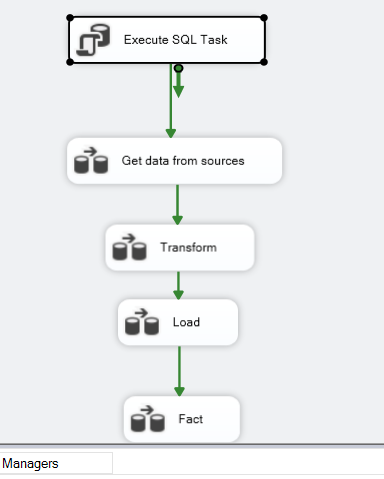
* STG\_SalesOrderHeader *(Nguồn: Sales.SalesOrderHeader)*
* STG\_SalesOrderDetail *(Nguồn: Sales.SalesOrderDetail)*
* STG\_DimDate *(Tạo từ các cột ngày)*
* STG\_Customer, STG\_Product, STG\_Employee (Nếu cần staging thêm từ các bảng nguồn)

## 2.5. Mô hình liên kết giữa các bảng

|  |  |  |  |
| --- | --- | --- | --- |
| **Bảng A** | **Khóa chính/ngoại** | **Bảng B** | **Mô tả quan hệ** |
| SalesOrderHeader | SalesOrderID | SalesOrderDetail | 1 đơn hàng có nhiều dòng chi tiết |
| SalesOrderHeader | CustomerID | Customer | Mỗi đơn hàng thuộc về một khách hàng |
| SalesOrderHeader | SalesPersonID | Employee | Mỗi đơn hàng do một nhân viên phụ trách |
| SalesOrderHeader | TerritoryID | SalesTerritory | Mỗi đơn hàng thuộc một khu vực địa lý |
| SalesOrderDetail | ProductID | Product | Mỗi dòng chi tiết ứng với một sản phẩm |
| Product | ProductSubcategoryID | ProductSubcategory | Mỗi sản phẩm thuộc một nhóm phụ |
| ProductSubcategory | ProductCategoryID | ProductCategory | Mỗi nhóm phụ thuộc về một nhóm chính |
| SalesOrderHeader.OrderDate | Ngày đặt hàng | DimDate.DateKey | Kết nối với bảng thời gian chuẩn hóa để phân tích theo ngày |

# **III. Xây dựng luồng tích hợp dữ liệu (ETL) bằng SQL và SSIS**

Tổng quan luồng ETL:

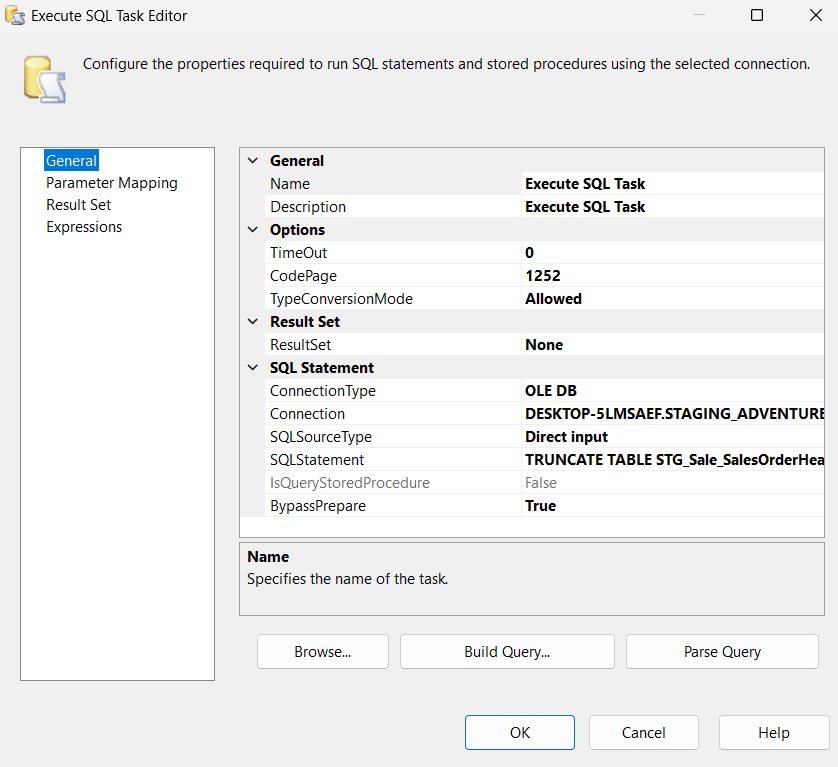


Để thiết kế hệ thống phân tích dữ liệu kinh doanh cho doanh nghiệp, nhóm đã xây dựng một quy trình tích hợp dữ liệu (ETL) hoàn chỉnh bằng công cụ SQL Server Integration Services (SSIS). Quy trình này được tổ chức thành các bước rõ ràng, đảm bảo luồng dữ liệu nhất quán, sạch và tối ưu hóa cho việc phân tích. Cụ thể:

## **Bước 1: Dọn sạch dữ liệu cũ – Execute SQL Task**

Trước khi bắt đầu nạp dữ liệu mới, **Execute SQL Task** được sử dụng để thực thi các câu lệnh TRUNCATE TABLE trên các bảng **Staging** nhằm:

* Xóa toàn bộ dữ liệu cũ trước khi load.
* Tránh lỗi trùng lặp khi chạy ETL nhiều lần.
* Đảm bảo dữ liệu staging nhẹ, phục vụ hiệu suất xử lý sau này.



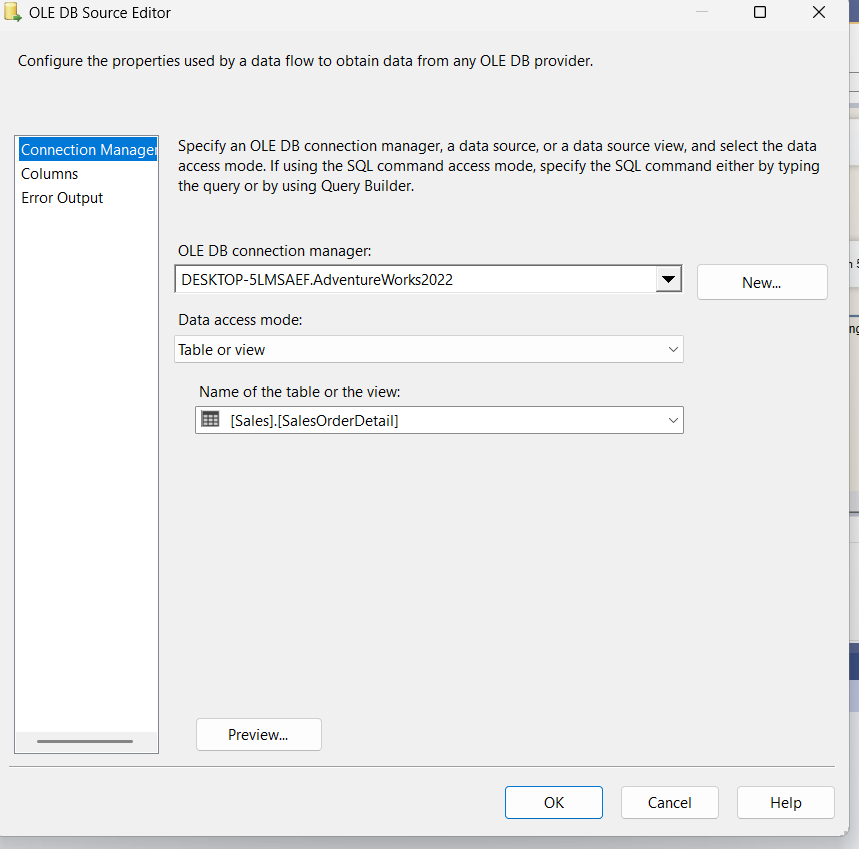
TRUNCATE TABLE STG\_SalesOrderDetail;

TRUNCATE TABLE STG\_SalesOrderHeader;

TRUNCATE TABLE STG\_DimDate;

-- ... các bảng dimension khác

## **Bước 2: Trích xuất dữ liệu từ nguồn – Get Data from Sources**

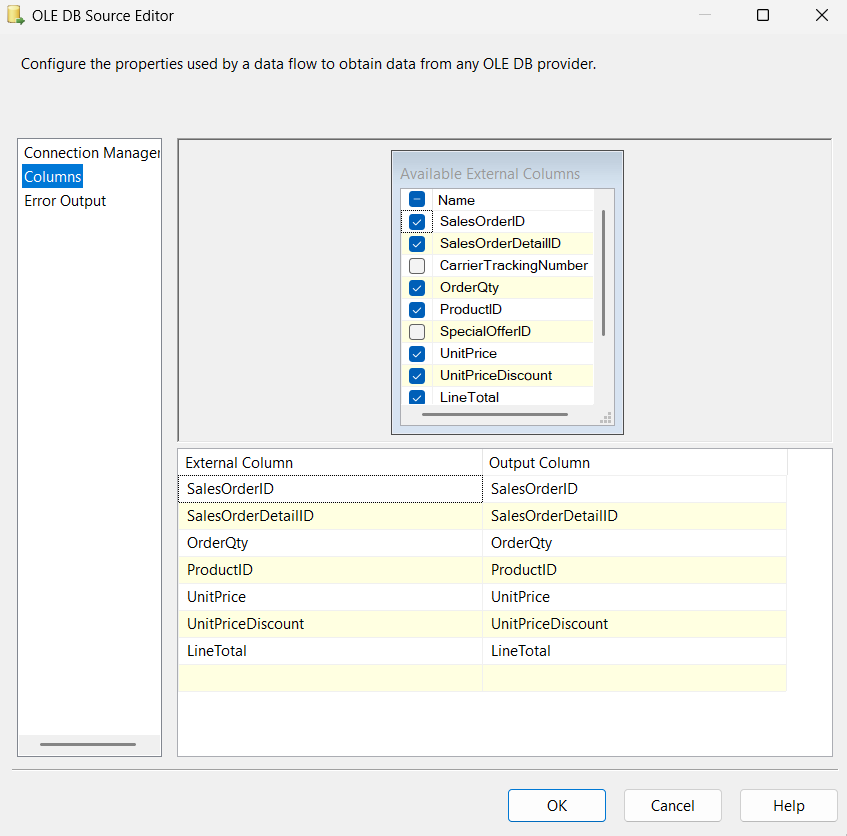


Dữ liệu được lấy từ hệ thống cơ sở dữ liệu **AdventureWorks 2022**, với các bảng gốc phục vụ từng khía cạnh của phân tích kinh doanh. Dưới đây là danh sách bảng nguồn và trường dữ liệu chính được lựa chọn:

|  |  |
| --- | --- |
| **Bảng nguồn** | **Trường dữ liệu chính** |
| Sales.SalesOrderDetail | SalesOrderID, SalesOrderDetailID, ProductID, OrderQty, UnitPrice, LineTotal |
| Sales.SalesOrderHeader | SalesOrderID, CustomerID, SalesPersonID, SubTotal, TaxAmt, OrderDate, Freight |
| Production.Product | ProductID, Name, ProductNumber, Color, Size, Weight, StandardCost, ListPrice |
| Purchasing.ProductVendor | ProductID, VendorID, StandardPrice, MinOrderQty, MaxOrderQty |
| Purchasing.Vendor | VendorID, Name (VendorName), CreditRating |
| Sales.SalesTerritory | TerritoryID, Name, CountryRegionCode, [Group], SalesYTD, SalesLastYear |
| Sales.SalesPerson | SalesPersonID, SalesYTD, SalesLastYear, TerritoryID |
| HumanResources.Employee | BusinessEntityID, BirthDate, Gender, JobTitle |
| Person.Person | BusinessEntityID, FirstName, LastName (→ FullName) |
| Person.PersonPhone | BusinessEntityID, PhoneNumber |
| Person.EmailAddress | BusinessEntityID, EmailAddress |
| Person.Address | AddressID, AddressLine1, City, PostalCode, ModifiedDate |

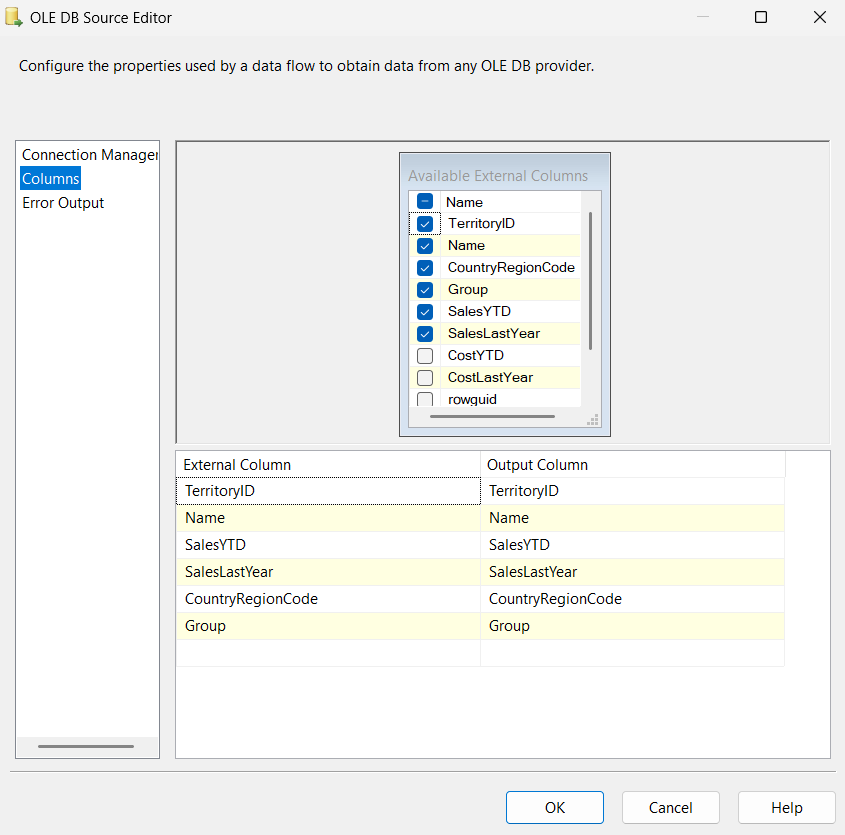
Bảng OrderDetails (Chi tiết đơn hàng), lấy các trường:

* SalesOrderID
* SalesOrderDetailID
* ProductID
* OrderQty
* UnitPrice
* UnitPriceDiscount
* LineTotal



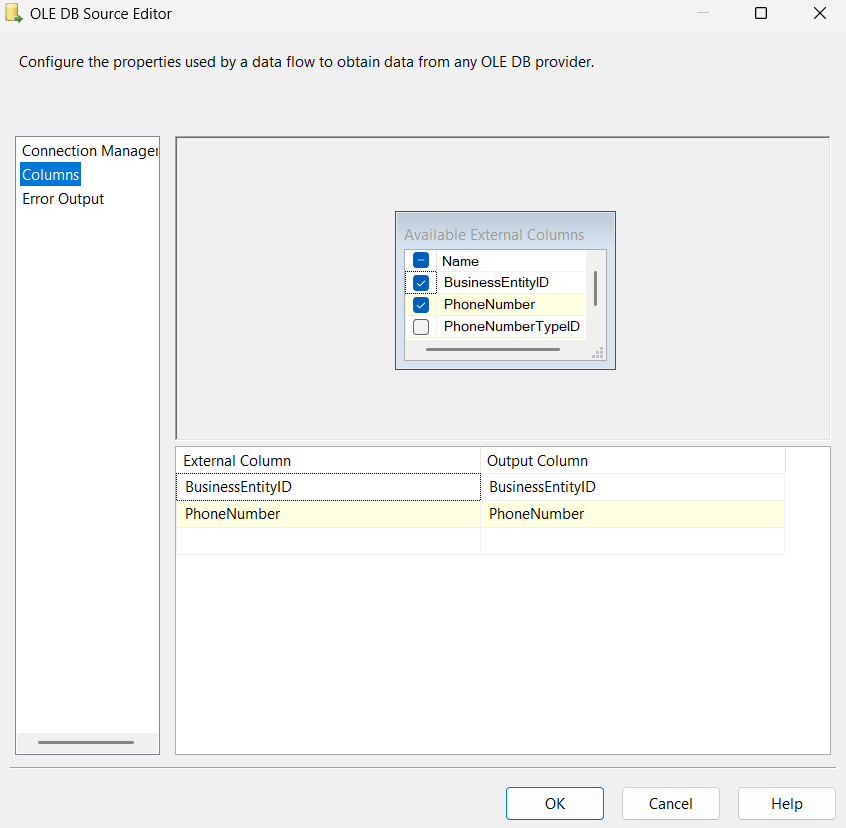
Bảng SalesTerritories (Khu vực bán hàng), lấy các trường:

* TerritoryID
* TerritoryName (Name)
* CountryRegionCode
* Group
* SalesYTD
* SalesLastYear



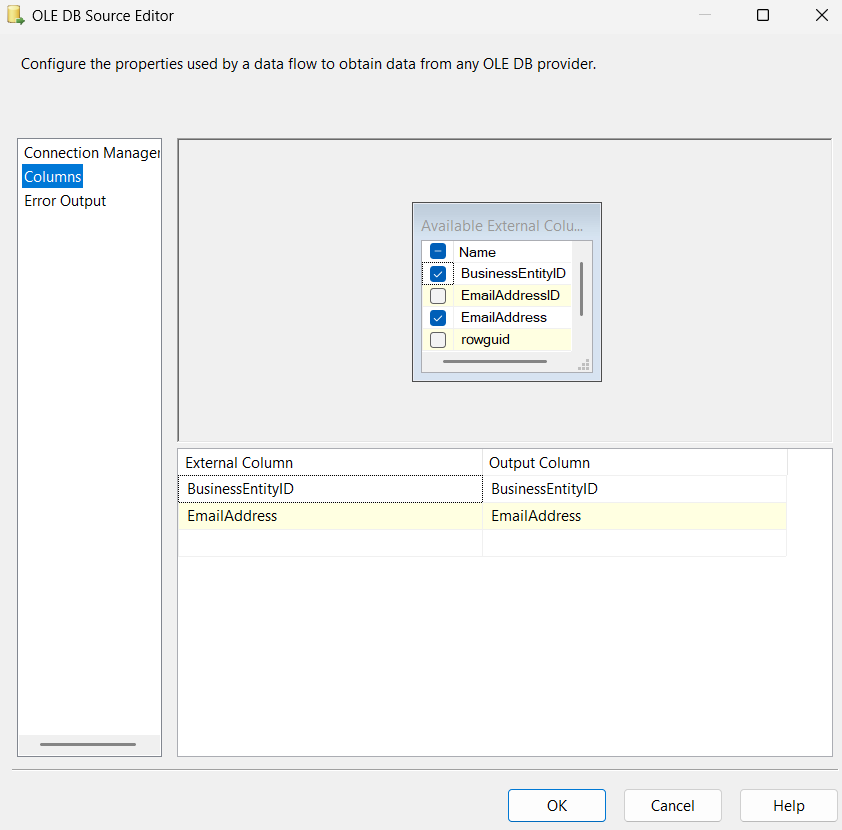
Bảng Person.PersonPhone, lấy trường dữ liệu:

* BusinessEntidyID
* Phonenumber



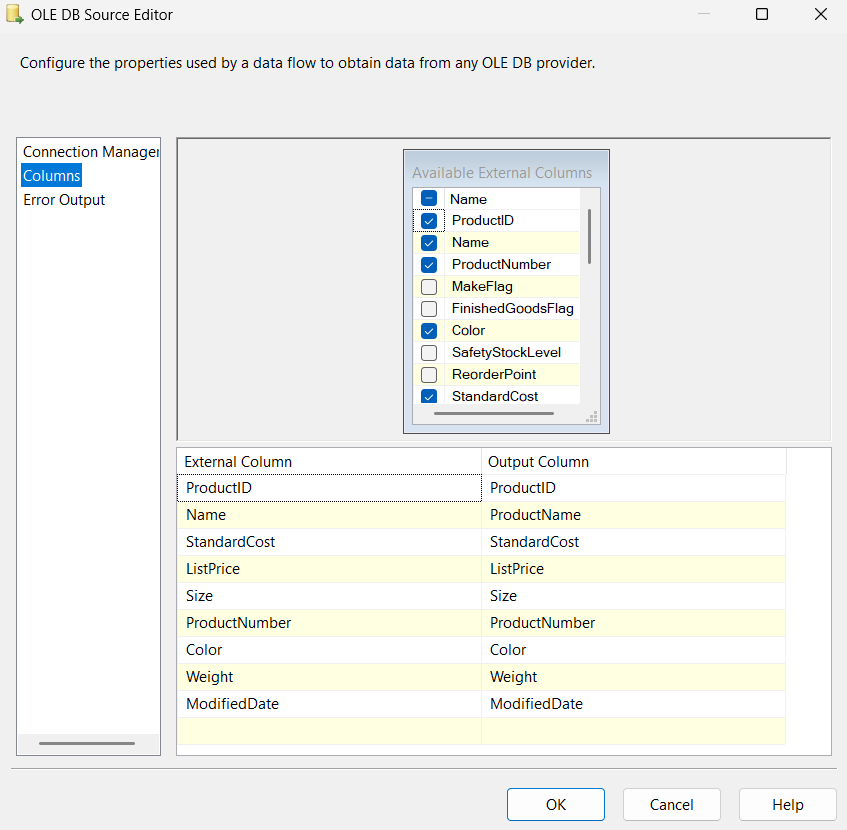
Bảng Person.EmailAddress, lấy trường dữ liệu:

* BusinessEntidyID
* Emailaddress



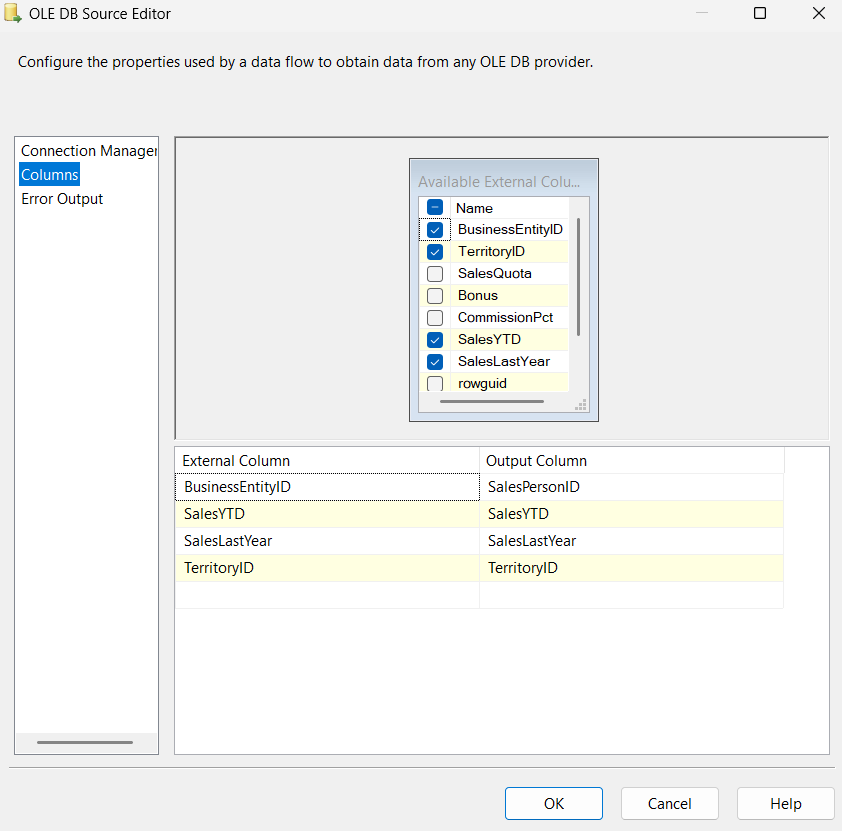
Bảng Production.Product, lấy trường dữ liệu:

* ProductID
* ProductName (Name)
* ProductNumber
* StandardCost
* ListPrice
* Color
* Size
* Weight
* ModifiedDate



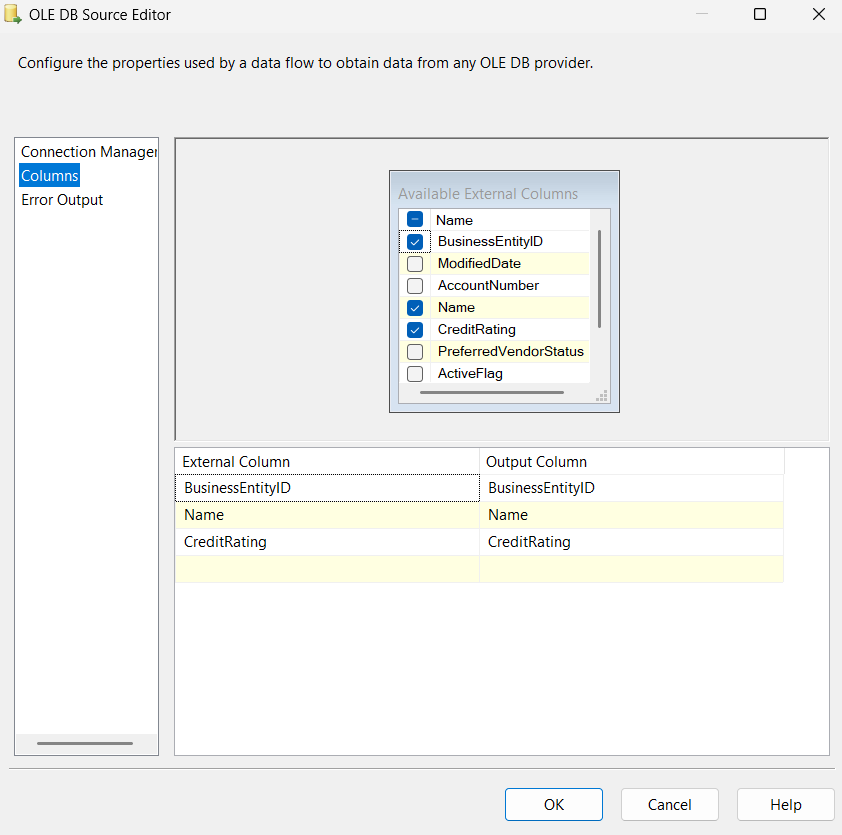
Bảng Sales.SalesPerson, lấy trường dữ liệu:

* SalesPersonID (BusinessEntityID)
* SalesYTD
* TerritoryID
* SalesLastYear



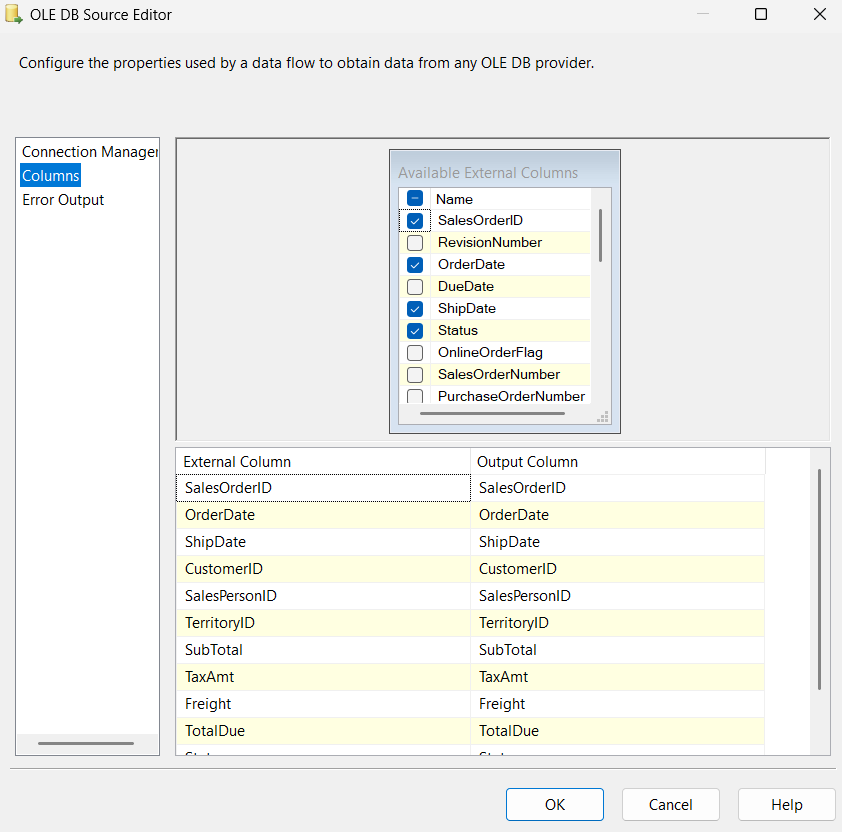
Bảng Purchasing.Vendor, lấy trường dữ liệu:

* VendorName (Name)
* Name
* CreditRating



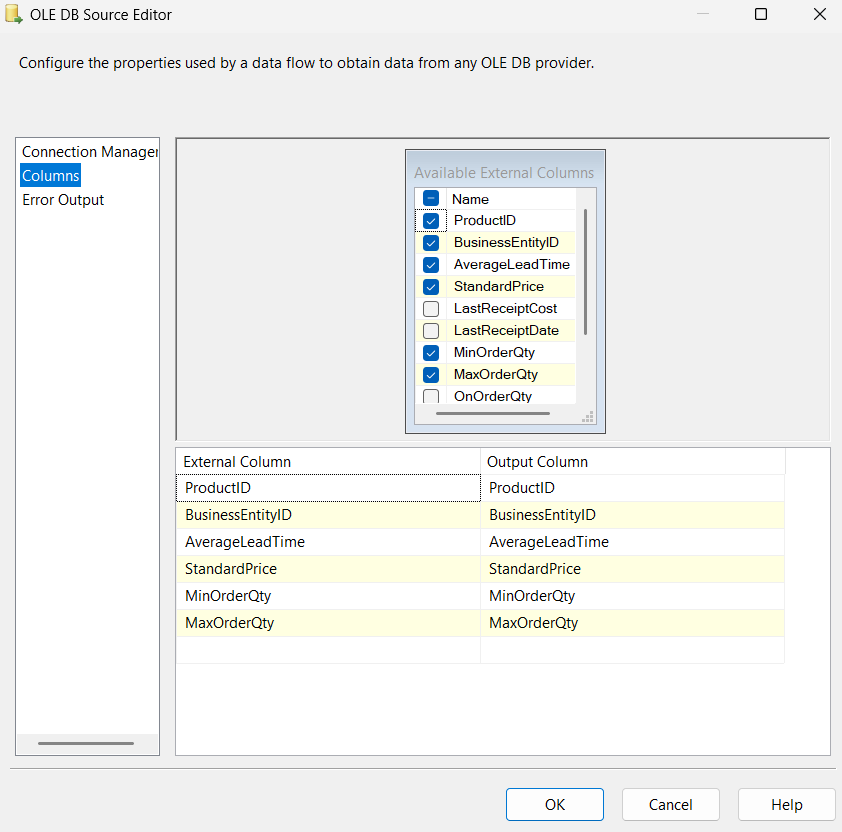
Bảng Sales.SalesOrderHeader, lấy trường dữ liệu:

* SalesOrderID
* CustomerID
* SalesPersonID
* TerritoryID
* SubTotal
* TaxAmt
* TotalDue
* OrderDate
* DueDate
* Freight
* Status



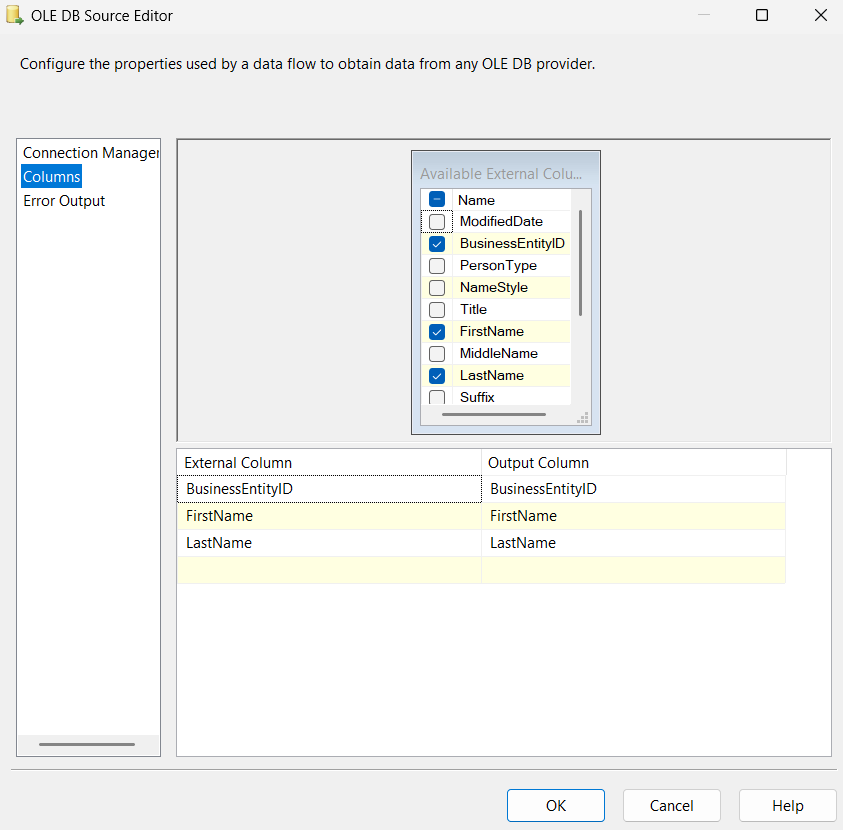
Bảng Purchasing.ProductVentor, lấy trường dữ liệu:

* ProductID
* VendorID (BusinessEntityID)
* StandardPrice
* LastReceiptCost
* AverageLeadTime
* MinOrderQty
* MaxOrderQty



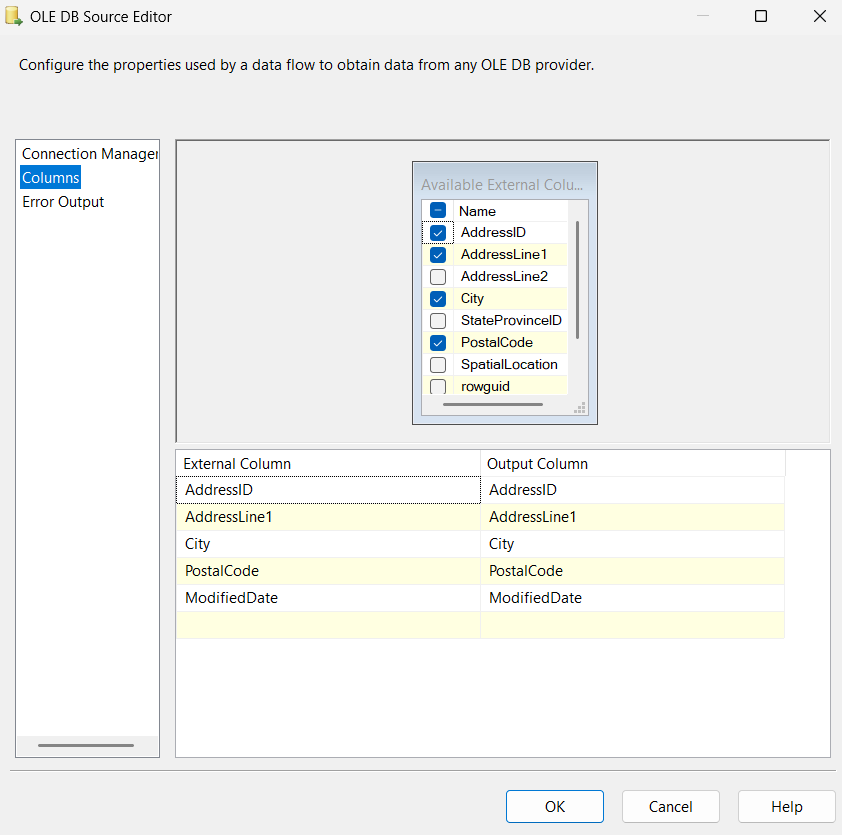
Bảng Person.Person, lấy trường dữ liệu:

* BusinessEntidyID
* Full Name (Bao gồm First Name + Last Name)



Bảng Person.Address, lấy trường dữ liệu:

* Address ID
* Address (AddressLine1)
* City
* PostalCode
* ModifiedDate



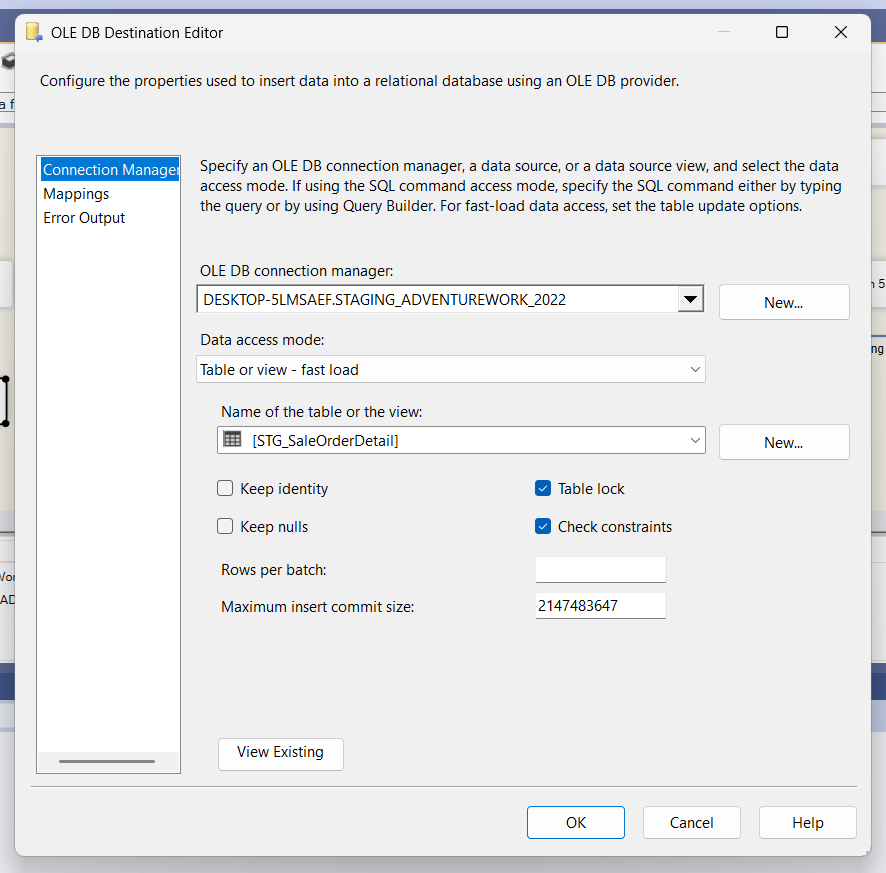
Bảng HumanResources.Employee, lấy trường dữ liệu:

* BirthDate
* Gender
* Marital Status
* JobTitle



Dữ liệu sau khi được trích xuất sẽ được đưa vào **các bảng STAGING**, với tên chuẩn hóa như:

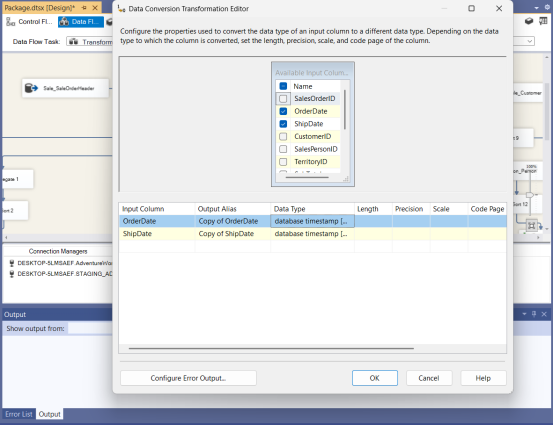
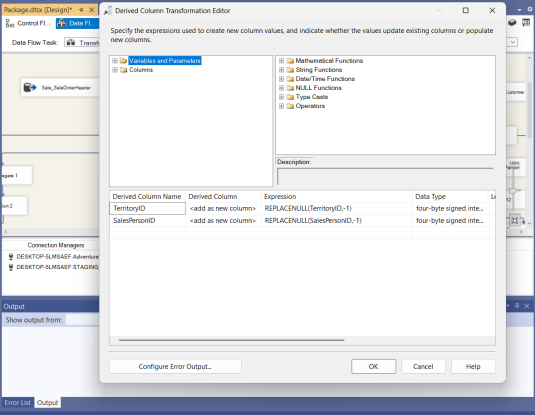
* STG\_SalesOrderDetail, STG\_SalesOrderHeader, STG\_Product
* STG\_Vendor, STG\_ProductVendor, STG\_SalesTerritory, STG\_SalesPerson
* STG\_Employee, STG\_Person, STG\_PersonPhone, STG\_EmailAddress, STG\_Address



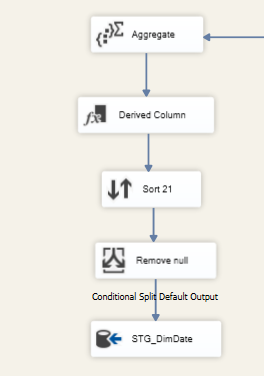
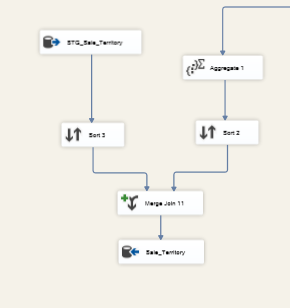
## **Bước 3: Biến đổi dữ liệu – Transform Data**

Trong giai đoạn này, dữ liệu từ các bảng **Staging** được xử lý và chuẩn hóa để sẵn sàng đưa vào mô hình Data Warehouse. Bao gồm các thao tác:

**Xử lý dữ liệu đã lấy ở vùng nhớ tạm, làm sạch dữ liệu**

**Xử lý dữ liệu dữ liệu theo từng luồng, đối chiếu bảng Fact với bảng Dim**

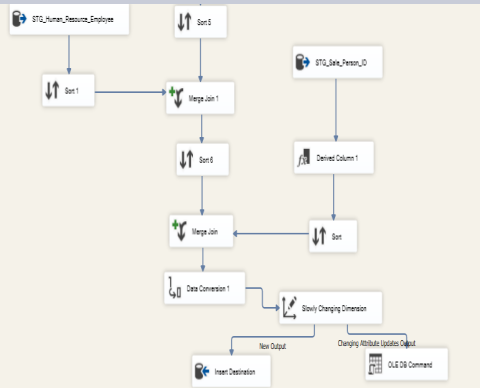
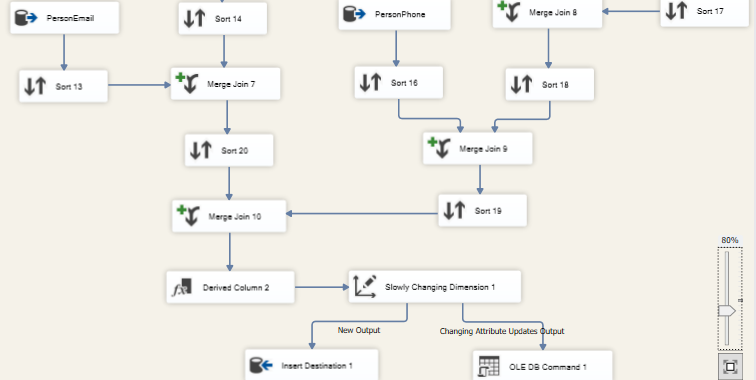
 

#### **Đối chiếu giữa Fact và Dimension:**

|  |  |
| --- | --- |
| **Bảng Fact (chính)** | **Bảng Dimension (liên quan)** |
| SalesOrderOverview | Customers, SalesPersons, SalesTerritories |
| OrderDetails | Products |
| VendorProductInfo | Products, Vendors (kết hợp từ ProductVendor + Vendor) |

Mỗi dòng dữ liệu trong bảng Fact được ánh xạ đầy đủ đến các thông tin chiều tương ứng (customer, product, territory...).

Trong quá trình này, nhóm đã sử dụng **Slowly Changing Dimension Type 1 (SCD Type 1)** bằng SSIS để xử lý cập nhật thông tin thay đổi ở các bảng chiều như khách hàng, nhân viên bán hàng, sản phẩm…

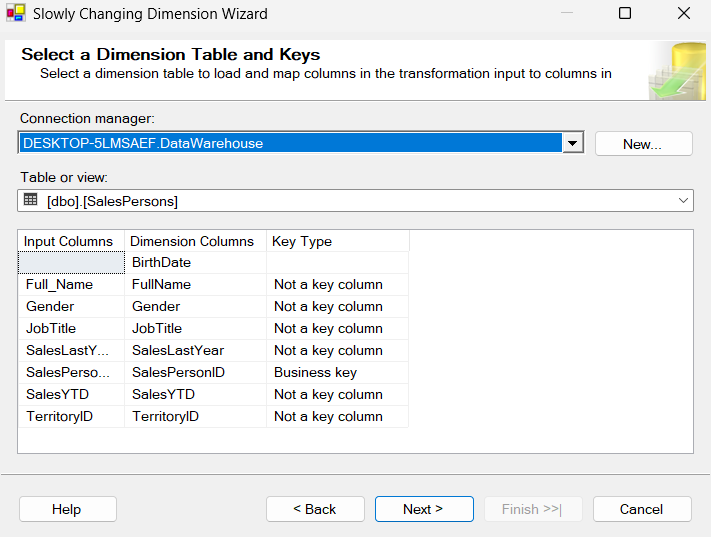
 

### **Giải thích SCD Type 1**

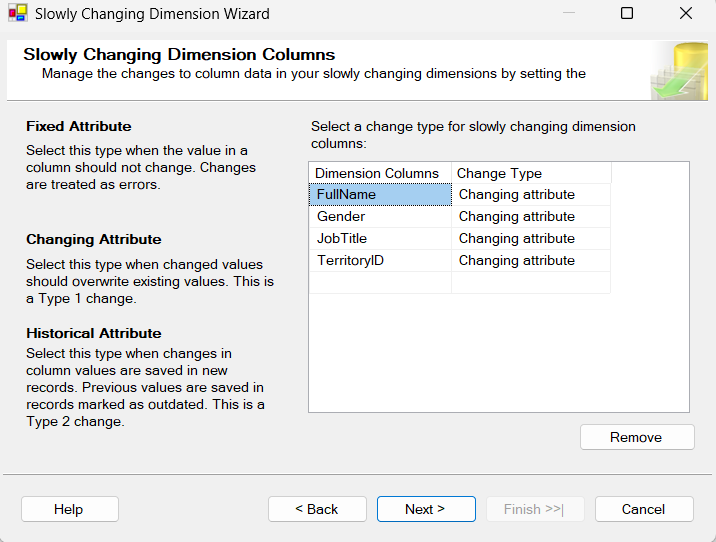
Với SCD Type 1, khi thông tin của một chiều (ví dụ: tên khách hàng hoặc chức vụ nhân viên) bị thay đổi, hệ thống sẽ **cập nhật đè trực tiếp thông tin mới lên dòng dữ liệu cũ** trong bảng Dimension.

Nhóm đã thiết kế luồng xử lý SCD này trong SSIS như các hình ảnh bạn cung cấp:

* Dữ liệu sau khi được merge từ nhiều bảng (Email, Phone, Name, JobTitle...) → đưa vào module **Slowly Changing Dimension**
* Với các thay đổi "Changing Attribute" → được cập nhật trực tiếp (overwrite)
* Output chia làm 2 nhánh:
  + New Output: thêm bản ghi mới



* + Changing Attribute Updates Output: cập nhật dữ liệu thay đổi

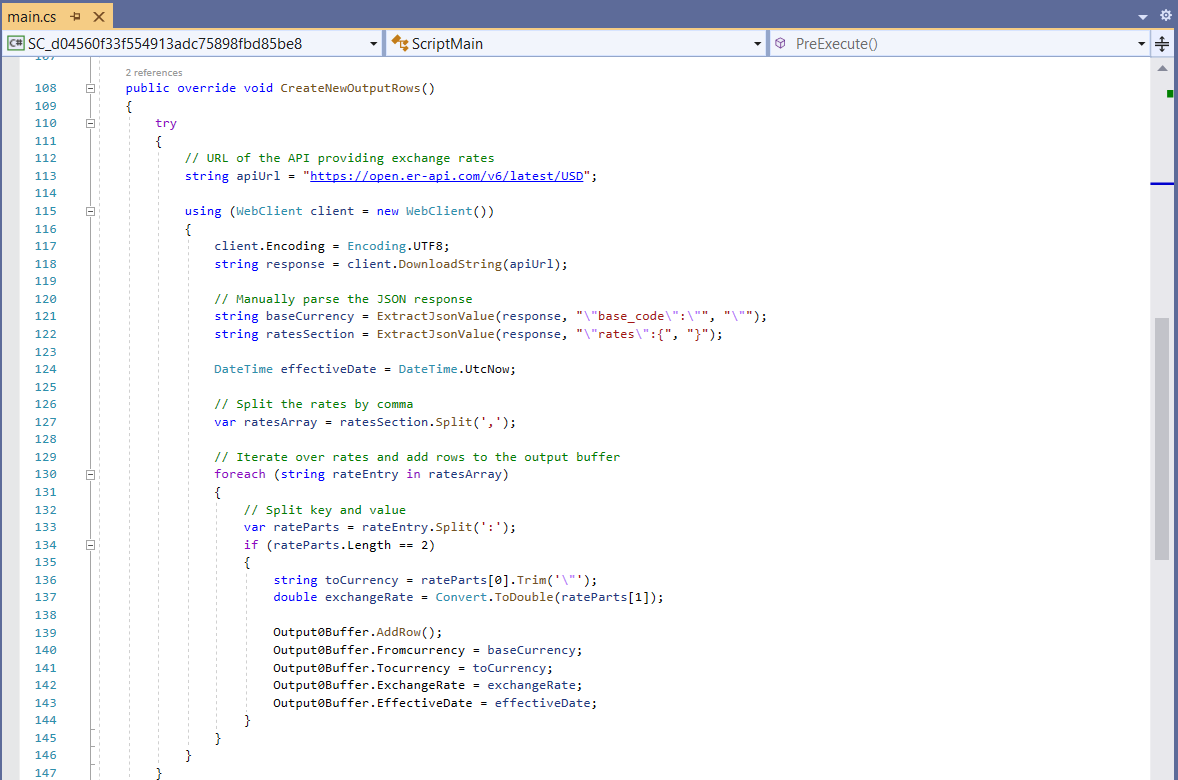


Ví dụ: Nếu nhân viên A từ "Sales Rep" → "Sales Manager", dòng dữ liệu trong Dimension sẽ được cập nhật đè chức vụ mới.

\* Ngoài việc xử lý và chuẩn hóa dữ liệu từ các bảng Staging, nhóm còn tích hợp API tỷ giá hối đoái vào quy trình ETL bằng cách sử dụng ngôn ngữ C# trong Script Task của SSIS.

#### Kết hợp tỷ giá USD → VND qua API

Để quy đổi doanh thu từ USD sang VND, nhóm đã sử dụng Script Task để gọi API từ exchangeratesapi.io và đưa dữ liệu vào bảng **Currency\_Rate**. Mã lệnh C# ví dụ như sau:



#### **Kết quả:**

* Tỷ giá được lưu vào biến trong SSIS.
* Sau đó, chèn vào bảng Currency Rate trong Data Warehouse.
* Giá trị đơn hàng bằng USD sẽ được chuyển đổi sang **VND**, phục vụ các báo cáo tài chính nội địa.

### **Xử lý thay đổi với SCD Type 2**

SCD Type 2 cho phép lưu lại lịch sử thay đổi của các thuộc tính trong bảng Dimension. Thay vì ghi đè (overwrite) như SCD Type 1, SCD Type 2 sẽ tạo bản ghi mới mỗi khi phát hiện có sự thay đổi ở các thuộc tính theo dõi (Changing Attributes).

Bước 1: Chọn những trường muốn thay đổi theo thời gian

A screenshot of a computer

Description automatically generated

Bước 2: Tìm kiếm những dữ liệu nào sẽ thay đổi theo thời gian, thì sẽ chọn “Historical Attribute”, còn cột nào không thay đổi chọn “Fixed Attribute”

A screenshot of a computer

Description automatically generated

Bước 3: Nhấn OK

A screenshot of a computer

Description automatically generated

## **Bước 4: Nạp dữ liệu vào Data Warehouse – Load to DW**

Dữ liệu sau khi xử lý sẽ được nạp vào hệ thống Data Warehouse theo cấu trúc mô hình sao:

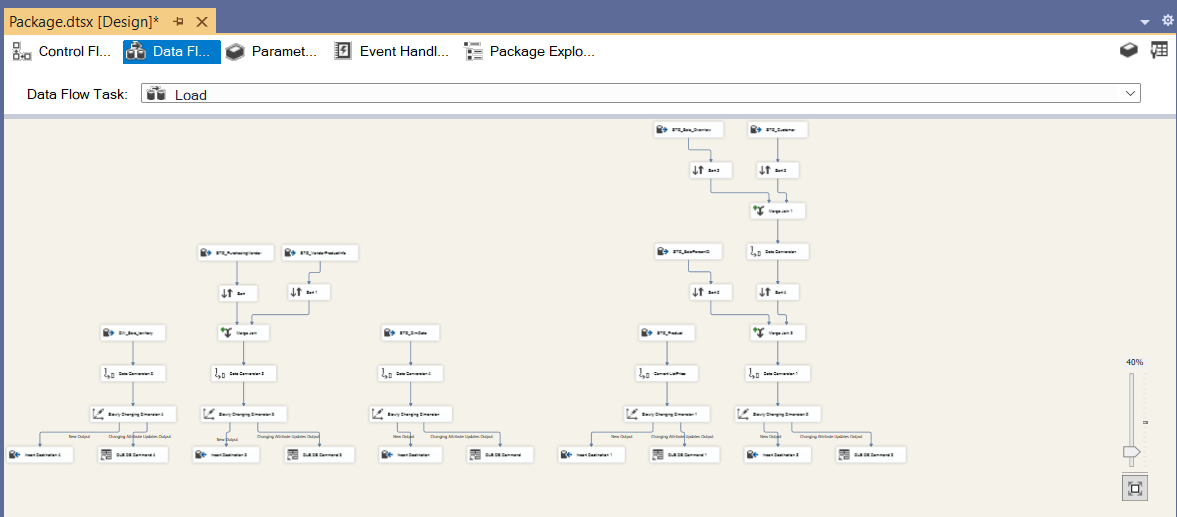
#### **Bảng Dimension:**

* Customers
* Products
* SalesPersons
* SalesTerritories
* VendorProductInfo
* STG\_DimDate (thời gian)
* Currency\_Rate (tỷ giá)

#### **Bảng Fact:**

* SalesOrderOverview (đơn hàng tổng)
* OrderDetails (chi tiết từng dòng sản phẩm)
* VendorProductInfo (thông tin giá cung ứng sản phẩm)
* Mỗi bảng Fact có các khóa ngoại liên kết với các bảng Dimension để phục vụ phân tích đa chiều như:
* Hiệu suất bán hàng theo vùng, theo nhân viên
* Chi tiết từng đơn hàng và dòng sản phẩm
* So sánh giá nhập – giá bán theo nhà cung cấp

**Kết quả**: Hệ thống Data Warehouse hoàn chỉnh đã được xây dựng từ các nguồn (staging), giúp trực quan hóa và phân tích dữ liệu trên Power BI một cách hiệu quả, hỗ trợ ra quyết định kinh doanh chính xác và kịp thời.



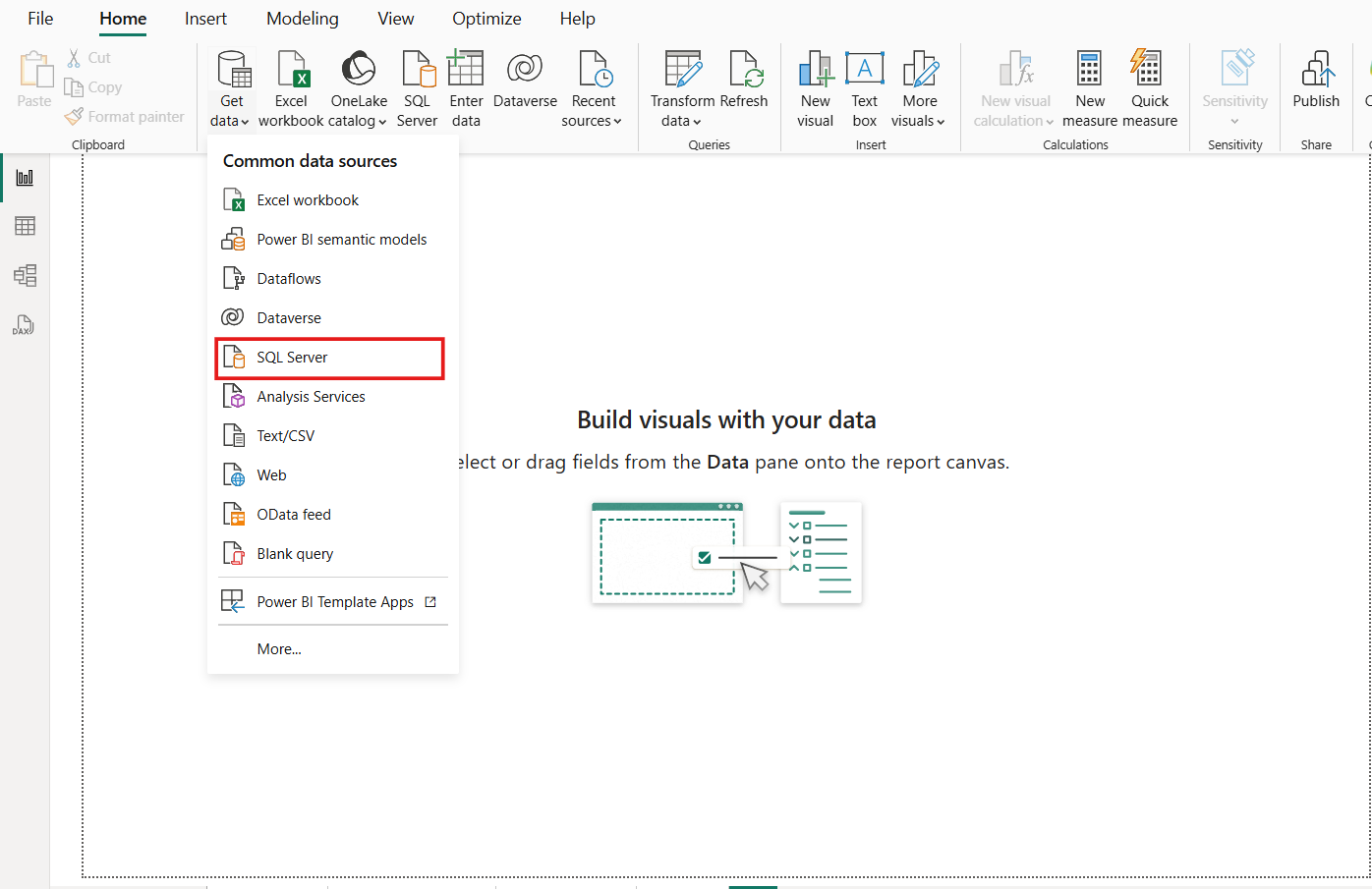
# **IV. Sử dụng CSDL tạo báo cáo**

1.Sau khi hệ thống Data Warehouse hoàn tất:

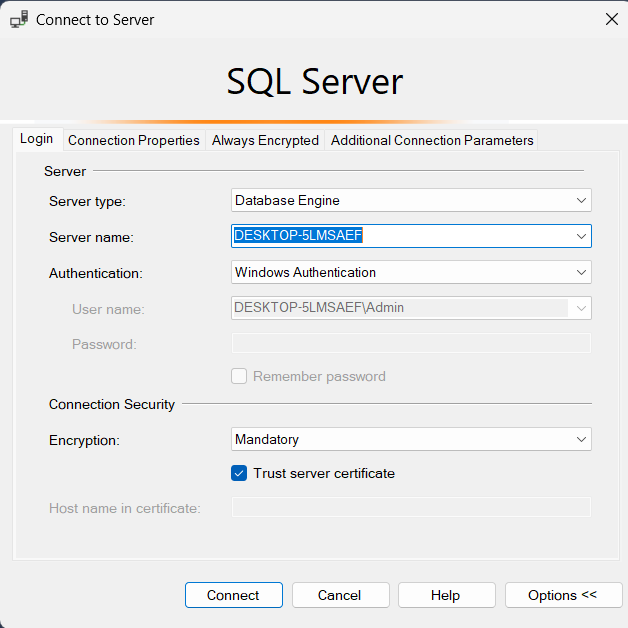
* Power BI được kết nối trực tiếp đến các bảng Fact và Dim.
* Các bảng được thiết lập mối quan hệ chuẩn hóa (1:N) để dễ dàng khai thác thông tin.
* Dashboard được xây dựng với các biểu đồ, bảng, slicer cho phép:
  + Phân tích doanh số theo sản phẩm, khách hàng, khu vực
  + So sánh hiệu suất nhân viên bán hàng
  + Theo dõi doanh số theo thời gian (năm, quý, tháng)

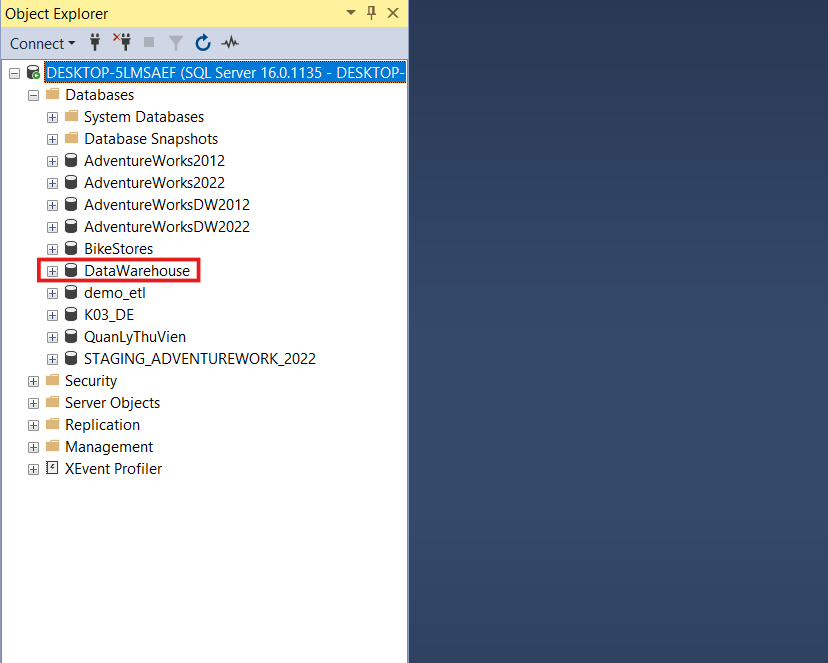
2. Kết nối với DataWarehouse để đưa dữ liệu vào báo cáo

Bước 1: Mở PowerBI => Get data, chọn SQL Server



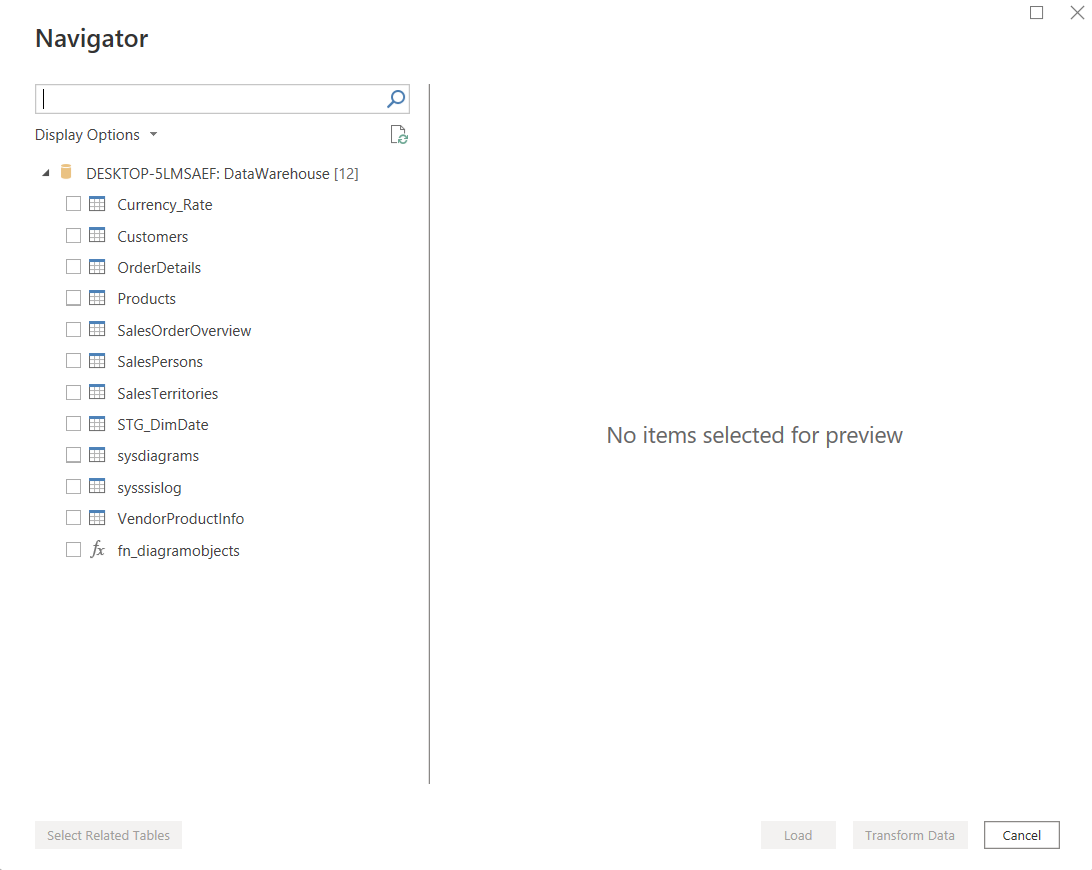
Bước 2: Điền cổng kết nối tới server và database

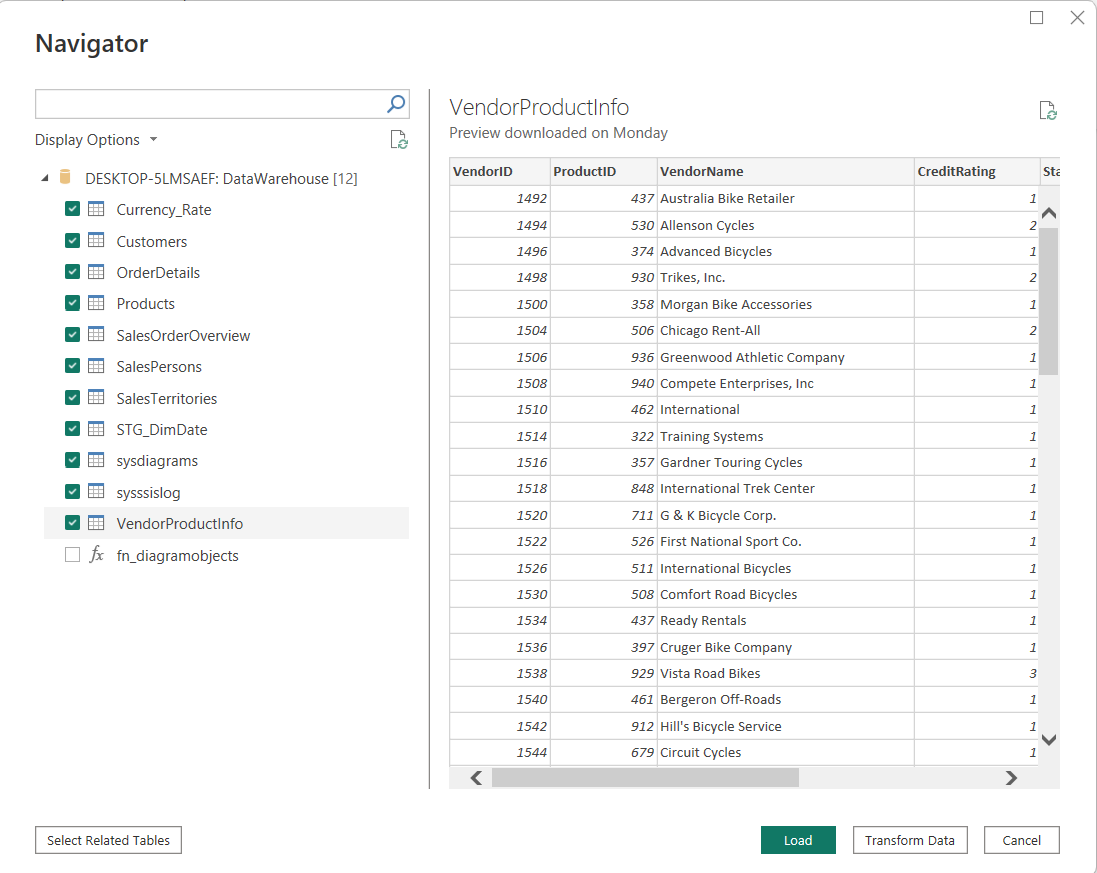




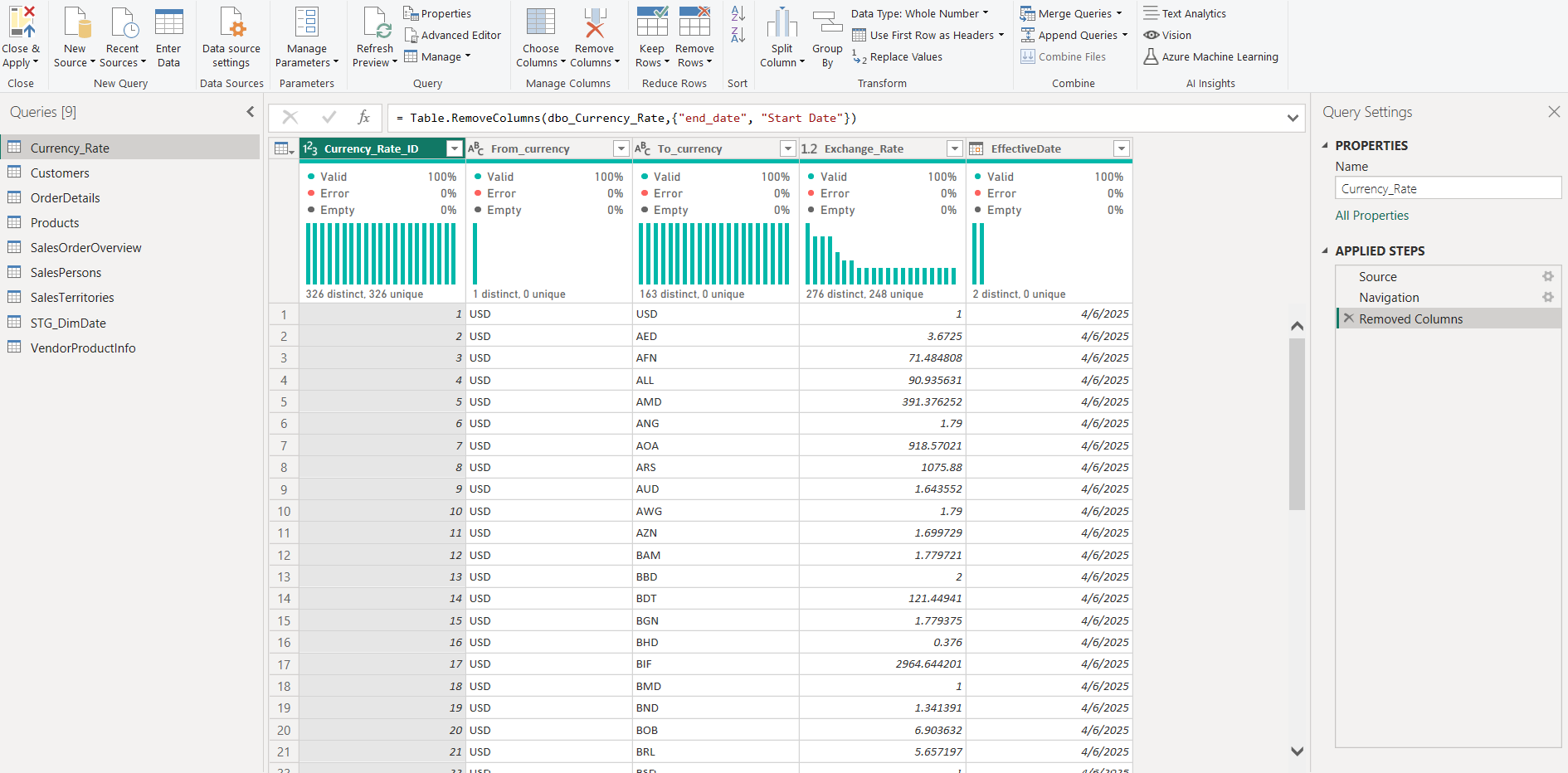


Bước 3: Sau khi kết nối xong, lấy những bảng cần lấy nhằm phục vụ cho việc biểu diễn dữ liệu trên powerBI





Bước 4: Xử lý dữ liệu trong power query trước khi đưa vào phân tích



Bước 5: Tạo kết nối giữa các bảng trong powerBI (mô hình Snowflake Schema)



# **IV. Phân tích báo cáo**



Tình hình tổng thể của công ty bao gồm:

- Tổng doanh thu

- Doanh thu theo thời gian (năm, tháng, quý)

- Phân bổ doanh thu theo vùng

- Xu hướng đơn hàng theo năm

- Biểu đồ địa lý theo lãnh thổ

Chỉ số đầu tiên được thể hiện là tổng doanh thu của toàn bộ hệ thống. Con số này phản ánh quy mô hoạt động và mức độ phát triển của doanh nghiệp trong giai đoạn báo cáo. Đây là chỉ số nền tảng cho tất cả các phân tích sâu hơn phía sau. Nếu doanh thu tăng trưởng tốt so với kỳ trước hoặc so với cùng kỳ năm trước, có thể đánh giá doanh nghiệp đang vận hành hiệu quả.

Tiếp theo là biểu đồ thể hiện doanh thu theo từng năm, giúp quan sát được xu hướng phát triển qua thời gian. Nếu doanh thu tăng dần qua các năm, điều đó chứng minh hoạt động kinh doanh đang được duy trì ổn định hoặc có cải thiện tích cực. Ngược lại, nếu xuất hiện sự sụt giảm trong một năm cụ thể, cần xem xét nguyên nhân – có thể đến từ biến động thị trường, sản phẩm không còn phù hợp, hoặc yếu tố nội tại như thiếu hàng, marketing yếu.

Biểu đồ doanh thu theo vùng địa lý cho thấy sự phân bổ mức độ tiêu thụ của thị trường theo từng khu vực. Những khu vực mang lại doanh thu cao thường là những nơi có thị trường tiềm năng, thói quen tiêu dùng tốt hoặc mạng lưới bán hàng được triển khai hiệu quả. Ngược lại, khu vực có doanh thu thấp có thể là thị trường chưa được khai thác triệt để, hoặc đang gặp khó khăn về mặt vận hành, tiếp cận khách hàng.

Bản đồ doanh thu theo vị trí địa lý mang lại cái nhìn trực quan về mức độ tập trung của doanh số trên toàn bộ thị trường. Các khu vực có màu sắc hoặc biểu tượng lớn hơn thể hiện mức doanh thu cao. Phân tích này giúp doanh nghiệp xác định được “điểm nóng” tiêu thụ sản phẩm, từ đó có kế hoạch đầu tư, mở rộng hoặc duy trì nguồn lực phù hợp. Đồng thời, bản đồ cũng góp phần nhận diện những vùng còn yếu để đề xuất chính sách khuyến mãi, marketing hoặc điều chỉnh sản phẩm phù hợp hơn với nhu cầu địa phương.

Tổng thể, các biểu đồ trong dashboard đều tập trung phản ánh kết quả kinh doanh ở góc độ địa lý và thời gian, giúp nhà quản lý có cái nhìn tổng thể và đánh giá được hiệu quả của chiến lược bán hàng đã triển khai. Đây là cơ sở để đưa ra các quyết định quan trọng như phân bổ nguồn lực, thay đổi định hướng sản phẩm, điều chỉnh thị trường mục tiêu hay xây dựng chính sách giá phù hợp cho từng khu vực cụ thể.



Báo cáo cung cấp cái nhìn sâu về khách hàng của doanh nghiệp, thông qua các khía cạnh như số lượng khách hàng, khu vực sinh sống, độ tuổi và mức độ trung thành. Đây là những yếu tố cốt lõi giúp doanh nghiệp hiểu rõ chân dung khách hàng hiện tại, từ đó đưa ra các chiến lược tiếp thị và chăm sóc phù hợp.

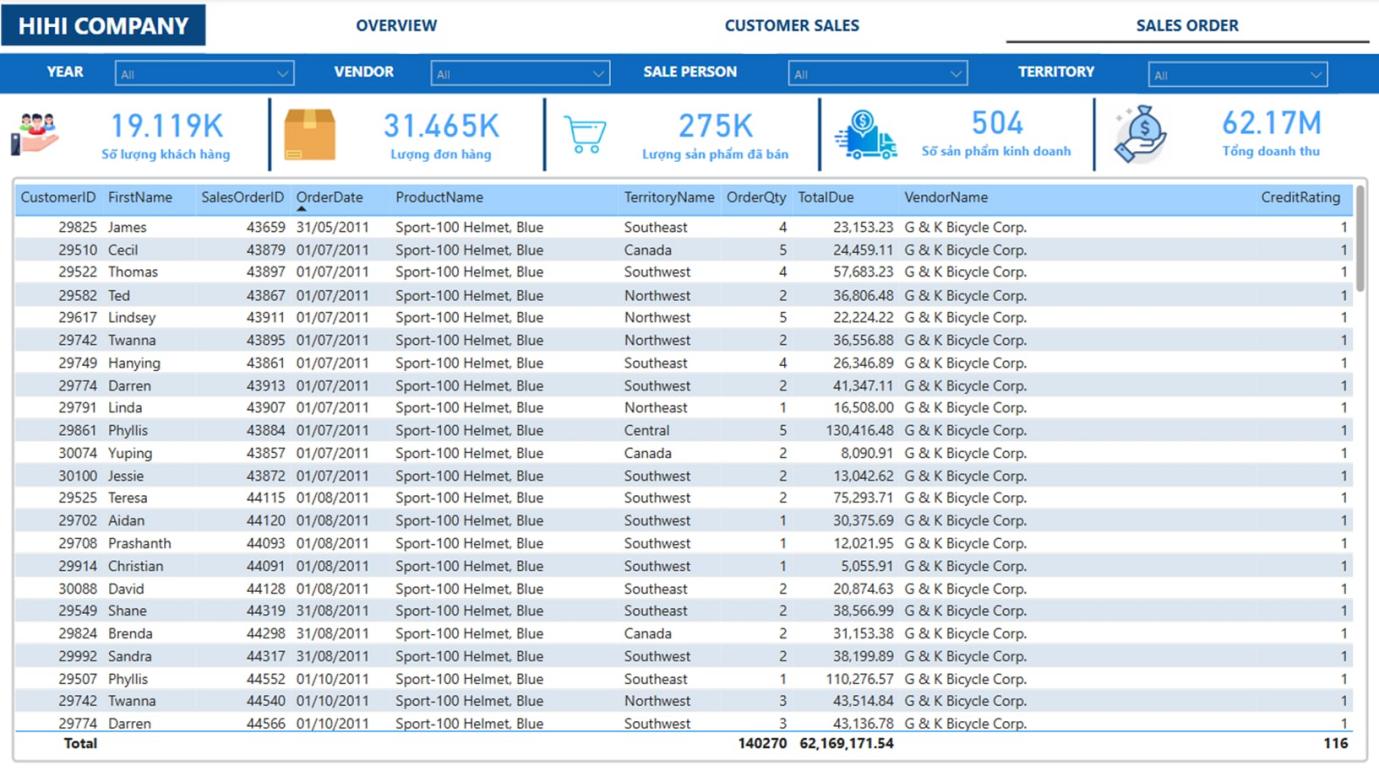
Thông tin đầu tiên được thể hiện là tổng số lượng khách hàng hiện có trong hệ thống. Con số này cho thấy quy mô tệp khách hàng doanh nghiệp đã tiếp cận được trong giai đoạn hoạt động. Nếu số lượng này tăng trưởng ổn định qua các kỳ, điều đó chứng tỏ doanh nghiệp đang thu hút tốt khách hàng mới, đồng thời giữ chân được khách hàng cũ.

Biểu đồ tiếp theo thể hiện phân bố khách hàng theo vùng địa lý. Các khu vực như miền Bắc, miền Trung, miền Nam hoặc các vùng cụ thể được thống kê số lượng khách hàng tương ứng. Từ đó, doanh nghiệp có thể đánh giá được khu vực nào đang có mật độ khách hàng cao – đây là các thị trường chính nên tiếp tục duy trì và chăm sóc tốt. Ngược lại, những khu vực có ít khách hàng có thể là nơi chưa được khai thác triệt để, hoặc sản phẩm chưa phù hợp với nhu cầu tiêu dùng tại đó.

Tiếp đến là biểu đồ phân loại khách hàng theo nhóm độ tuổi, cho thấy cơ cấu khách hàng chủ yếu nằm ở nhóm tuổi nào. Nếu phần lớn khách hàng thuộc nhóm tuổi từ 26–35 hoặc 36–45, có thể nhận định sản phẩm đang hấp dẫn đối tượng lao động trung niên, có thu nhập ổn định. Trường hợp khách hàng chủ yếu là nhóm trẻ (18–25), thì doanh nghiệp cần chú ý đến xu hướng tiêu dùng nhanh, giá rẻ và dịch vụ tiện lợi. Việc hiểu được đặc điểm tuổi tác giúp doanh nghiệp điều chỉnh nội dung truyền thông, thiết kế sản phẩm hoặc kênh tiếp cận cho phù hợp.

Một yếu tố quan trọng khác là mức độ trung thành của khách hàng, được chia thành các nhóm như: khách mới, khách quay lại, khách một lần và khách trung thành. Nếu tỷ lệ khách trung thành cao, doanh nghiệp đang có chính sách giữ chân tốt và tạo được trải nghiệm tích cực. Ngược lại, tỷ lệ khách mua một lần hoặc khách mới cao mà không có sự chuyển đổi thành khách trung thành, là dấu hiệu cảnh báo cần xem xét lại quy trình bán hàng, chăm sóc sau bán hoặc chất lượng sản phẩm.

Tổng hợp lại, dashboard phân tích khách hàng giúp doanh nghiệp nhận diện được không chỉ ai là khách hàng của mình, mà còn họ đến từ đâu, thuộc nhóm nào, và gắn bó với doanh nghiệp ra sao. Từ đó, các chiến lược như cá nhân hóa dịch vụ, phân bổ chi phí tiếp thị theo vùng hoặc triển khai các chương trình chăm sóc khách hàng thân thiết có thể được xây dựng một cách hiệu quả hơn.



Cuối cùng đó là tập trung vào phân tích hiệu suất kinh doanh của sản phẩm – một phần quan trọng trong việc đánh giá cơ cấu danh mục hàng hóa và chiến lược bán hàng của doanh nghiệp. Thông qua các biểu đồ chính như doanh thu theo loại sản phẩm, sản phẩm bán chạy nhất và xu hướng doanh thu theo thời gian, bảng điều khiển này giúp làm rõ sản phẩm nào đang mang lại hiệu quả, nhóm hàng nào cần được đầu tư thêm hoặc loại bỏ khỏi danh mục.

Phần đầu của dashboard thể hiện doanh thu theo danh mục sản phẩm (Product Category). Biểu đồ cột hoặc hình tròn minh họa tỷ trọng doanh thu của từng nhóm hàng như xe đạp (Bikes), linh kiện (Components), phụ kiện (Accessories)... Việc phân tích này giúp xác định nhóm sản phẩm chủ lực, ví dụ nếu Bikes chiếm phần lớn tổng doanh thu thì đây chính là dòng sản phẩm chính doanh nghiệp cần duy trì và phát triển. Ngược lại, nếu có nhóm hàng chiếm tỷ lệ thấp và không tăng trưởng qua thời gian, có thể cân nhắc tái cấu trúc danh mục hoặc đẩy mạnh tiếp thị.

Tiếp theo là danh sách hoặc biểu đồ thể hiện Top các sản phẩm có doanh thu cao nhất. Đây là thông tin rất thiết thực, cho phép doanh nghiệp nắm được chính xác mặt hàng nào đang bán chạy nhất, nhờ đó tập trung nguồn lực như quảng bá, nhập hàng, hoặc phân phối nhiều hơn tại các kênh phù hợp. Danh sách này cũng giúp bộ phận tồn kho kiểm soát sản phẩm bán chạy để tránh đứt hàng hoặc thiếu hụt vào mùa cao điểm.

Một biểu đồ quan trọng khác là biểu đồ thể hiện xu hướng doanh thu của sản phẩm theo thời gian – thường là biểu đồ đường (line chart). Biểu đồ này cung cấp cái nhìn về sự biến động của doanh số theo tháng hoặc quý, giúp doanh nghiệp phát hiện quy luật mua hàng theo mùa (seasonality), đánh giá hiệu quả chiến dịch bán hàng trong từng giai đoạn, cũng như dự báo được nhu cầu trong tương lai. Ví dụ, nếu thấy doanh thu nhóm Bikes tăng mạnh vào quý 2, quý 3, thì đây có thể là mùa cao điểm mua xe đạp và cần chuẩn bị hàng từ trước đó.

Tổng thể, dashboard Phân tích sản phẩm cung cấp các chỉ số cần thiết để doanh nghiệp đánh giá hiệu suất từng sản phẩm và nhóm hàng. Từ các số liệu này, có thể đưa ra những quyết định chiến lược như:

* Duy trì và phát triển nhóm sản phẩm có doanh thu ổn định, tăng trưởng cao;
* Tạm ngưng hoặc điều chỉnh những dòng sản phẩm kém hiệu quả;
* Thiết lập kế hoạch nhập hàng, quảng cáo hoặc khuyến mãi theo mùa vụ;
* Phân tích khả năng sinh lời của từng sản phẩm (nếu tích hợp thêm dữ liệu chi phí hoặc lợi nhuận).