

**TRƯỜNG ĐẠI HỌC KINH TẾ
KHOA THỐNG KÊ – TIN HỌC**



**BÁO CÁO CUỐI KỲ
MÔN: THỰC HÀNH THIẾT KẾ KHO VÀ PHÂN TÍCH
DỮ LIỆU KINH DOANH**

Lớp : **46K21.2**

Giảng viên hướng dẫn : **Đặng Quốc Viện**

Thành viên nhóm : Hoàng Dũng

Văn Phú Đạt

Đào Thị Hương Giang

Lê Thị Kiều Linh

Lưu Vĩnh Khôi

Trần Thị Anh Thư

Lê Thị Hoàng Phụng

Đà Nẵng 12/2023

MỤC LỤC

| | |
|--|-----------|
| I. GIỚI THIỆU VỀ DỮ LIỆU..... | 2 |
| 1. <i>Orders</i> - Đơn đặt hàng | 2 |
| 2. <i>Region</i> - Khu vực | 2 |
| 3. <i>Product</i> - Sản phẩm..... | 2 |
| 4. <i>Supplier</i> - Nhà cung cấp | 3 |
| II. THIẾT KẾ CƠ SỞ DỮ LIỆU | 3 |
| 1. Thiết kế cơ sở dữ liệu Staging | 3 |
| 2. Thiết kế Datawarehouse (Kho dữ liệu) | 4 |
| III. XÂY DỰNG TIỀN TRÌNH ETL | 5 |
| 1. Tiến trình xây dựng..... | 5 |
| 2. Module tổng hợp..... | 6 |
| 2.1 <i>ETL cho các bảng ở Staging</i> | 7 |
| 2.2 <i>ETL cho các bảng Dim</i> | 9 |
| 2.3 <i>ETL cho bảng Fact_Sales</i> | 15 |
| 3. ETL cho bảng Fact_Sales..... | 19 |
| 4. Thực hiện truy vấn dữ liệu đã được đổ vào SQL | 22 |
| 4.1 <i>Staging_Region</i> | 22 |
| 4.2 <i>Staging_Product</i> | 23 |
| 4.3 <i>Staging_Supplier</i> | 23 |
| 4.4 <i>Staging_Orders</i> | 24 |
| 4.5. <i>Staging_Fact_Sales</i> | 24 |
| 4.6 <i>Dim_Product</i> | 25 |
| 4.7 <i>Dim_Region</i> | 25 |
| 4.8 <i>Dim_Supplier</i> | 26 |
| 4.9 <i>Dim_Time</i> | 26 |
| 4.10 <i>Fact_Sales</i> | 26 |
| IV. OLAP..... | 28 |
| 1. Data Source..... | 28 |
| 2. Data source view | 29 |
| 3. Data Cube | 30 |
| 4. Dimensions | 31 |
| 5. Dữ liệu Olap trên SQL | 33 |
| 6. Phân tích OLAP | 34 |
| 7. Pivot Table..... | 35 |
| V. DASHBOARD..... | 38 |
| 1. Phân tích biểu đồ..... | 38 |
| 2. Kết luận | 41 |
| 3. Giải pháp..... | 42 |

DANH MỤC HÌNH ẢNH

| | |
|---|----|
| Hình 1. Dữ liệu của đơn đặt hàng rượu tại khu vực IOWA..... | 2 |
| Hình 2. Dữ liệu khu vực..... | 2 |
| Hình 3. Dữ liệu về sản phẩm được cung cấp tại IOWA..... | 3 |
| Hình 4. Dữ liệu của các nhà cung cấp tại khu vực IOWA..... | 3 |
| Hình 5. Thiết kế cơ sở dữ liệu của staging..... | 4 |
| Hình 6. Lược đồ hình sao của kho dữ liệu đích..... | 5 |
| Hình 7. Module ETL tổng hợp..... | 6 |
| Hình 8. Quá trình ETL cho các bảng ở staging..... | 7 |
| Hình 9. Module thực hiện xóa dữ liệu ở Staging..... | 8 |
| Hình 10. Data flow lấy dữ liệu cho các bảng staging..... | 8 |
| Hình 11. Lấy dữ liệu từ nguồn excel..... | 9 |
| Hình 12. OLE DB Source lấy dữ liệu cho bảng Staging_Region..... | 9 |
| Hình 13. Quá trình ETL cho các bảng Dim..... | 10 |
| Hình 14. ETL cho bảng Dim_Product..... | 10 |
| Hình 15. Aggregate cho bảng Dim_Product..... | 11 |
| Hình 16. Select a Dimension Table and Keys cho bảng Dim_Product..... | 11 |
| Hình 17. Slowly changing Dimension Columns cho bảng Dim_Product..... | 12 |
| Hình 18. Quá trình ETL cho bảng Dim_Time..... | 12 |
| Hình 19. Aggregate cho bảng Dim_Time..... | 13 |
| Hình 20. Data Conversion Transformation cho bảng Dim_Time..... | 13 |
| Hình 21. Derived Column cho bảng Dim_Time..... | 14 |
| Hình 22. . Scrip Transformation cho bảng Dim_Time..... | 14 |
| Hình 23. Select a Dimension Table and Keys cho bảng Dim_Time..... | 15 |
| Hình 24. Quá trình ETL cho bảng Fact_Sales..... | 15 |
| Hình 25. Sắp xếp các cột được chọn cho bảng Fact_Sales..... | 16 |
| Hình 26. Lookup cho bảng Fact_Sales..... | 16 |
| Hình 27. Lookup cho bảng Fact_Sales..... | 17 |
| Hình 28. Lookup cho bảng Fact_Sales..... | 17 |
| Hình 29. Lookup cho bảng Fact_Sales..... | 18 |
| Hình 30. Thực hiện đổ dữ liệu vào bảng Fact_Sales..... | 18 |
| Hình 31. ETL cho bảng Fact_Sales..... | 19 |
| Hình 32. Thực hiện đổ dữ liệu vào bảng Staging_Fact_Sales..... | 20 |
| Hình 33. Lookup cho bảng Fact_Sales..... | 20 |
| Hình 34. Conditional Split Transformation Editor vào bảng Staging_Fact_Sales..... | 21 |
| Hình 35. Thực hiện lệnh update..... | 22 |
| Hình 36. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Region..... | 22 |
| Hình 37. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Product..... | 23 |

| | |
|---|----|
| Hình 38. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Supplier | 23 |
| Hình 39. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Order... | 24 |
| Hình 40. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Fact_Sales | 24 |
| Hình 41. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Dim_Product..... | 25 |
| Hình 42. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Dim_Region | 25 |
| Hình 43. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Dim_Supplier.... | 26 |
| Hình 44. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Dim_Time | 26 |
| Hình 45. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của bảng Fact_Sales | 27 |
| Hình 46. Thực hiện update dữ liệu của bảng Fact_Sales..... | 27 |
| Hình 47. Kết nối Server và chọn database | 28 |
| Hình 48. Đặt tên cho source | 28 |
| Hình 49. Chọn bảng có trong database để tạo views | 29 |
| Hình 50. Đặt tên cho Data Source Views | 29 |
| Hình 51. Chọn Measure | 30 |
| Hình 52. Đặt tên cho cube..... | 30 |
| Hình 53. Data cube sau khi tạo thành công..... | 31 |
| Hình 54. Dimensions Supplier..... | 31 |
| Hình 55. Dimensions Product..... | 32 |
| Hình 56. Dimensions Time | 32 |
| Hình 57. Dimensions Region | 33 |
| Hình 58. Deploy thành công..... | 33 |
| Hình 59. Dữ liệu Olap đã được đưa về SQL | 34 |
| Hình 60. Phân tích doanh thu theo tháng | 34 |
| Hình 61. Phân tích sản phẩm theo lợi nhuận và doanh thu..... | 35 |
| Hình 62. Biểu đồ đường thể hiện tổng doanh thu của các tháng trong năm 2021 | 36 |
| Hình 63. Biểu đồ cột thể hiện doanh thu của 12 sản phẩm trong tháng 12 | 37 |
| Hình 64. Biểu đồ Grand Total | 38 |
| Hình 65. Biểu đồ Trend Line..... | 38 |
| Hình 66. Biểu đồ cột Measurement by Product | 39 |
| Hình 67. Biểu đồ đường Sale by Supplier and Date..... | 39 |
| Hình 68. Biểu đồ cloud by Region of Sale..... | 40 |
| Hình 69. Biểu đồ tích lũy | 41 |

I. GIỚI THIỆU VỀ DỮ LIỆU

Bộ dữ liệu nói về doanh thu bán rượu của khu vực IOWA.

1. *Orders* - Đơn đặt hàng

- Gồm 7 trường và 1000 dòng dữ liệu

| OrderID | RegionID | SupplierID | ProductID | Sale | Quantity | Date |
|-----------------|----------|------------|-----------|--------|----------|------------|
| INV-33237000023 | 97 | 395 | 87510 | 23.19 | 1 | 27/07/2021 |
| INV-33524200039 | 52 | 395 | 87510 | 278.28 | 12 | 10/04/2021 |
| INV-33462200108 | 77 | 395 | 88922 | 89.28 | 2 | 27/03/2021 |
| INV-33459600008 | 57 | 395 | 87510 | 69.57 | 3 | 21/09/2021 |
| INV-33468300052 | 17 | 395 | 87510 | 69.57 | 3 | 24/05/2021 |
| INV-33501300003 | 77 | 395 | 87510 | 115.95 | 5 | 19/03/2021 |
| INV-33467100035 | 17 | 395 | 87510 | 46.38 | 2 | 07/06/2021 |
| INV-33440900085 | 52 | 395 | 89207 | 157.5 | 6 | 15/08/2021 |
| INV-33851200054 | 25 | 395 | 88922 | 89.28 | 2 | 05/04/2021 |
| INV-33178400003 | 14 | 395 | 87510 | 278.28 | 12 | 15/06/2021 |
| INV-33248200009 | 32 | 395 | 89207 | 157.5 | 6 | 07/04/2021 |
| INV-33872600048 | 25 | 395 | 87510 | 69.57 | 3 | 24/10/2021 |
| INV-33754000063 | 23 | 395 | 87510 | 23.19 | 1 | 05/07/2021 |
| INV-33510200061 | 46 | 395 | 88922 | 178.56 | 4 | 01/02/2021 |
| INV-33893700089 | 53 | 395 | 89207 | 52.5 | 2 | 22/01/2021 |
| INV-33458900085 | 57 | 395 | 87510 | 278.28 | 12 | 23/07/2021 |
| INV-33506300009 | 96 | 395 | 87510 | 278.28 | 12 | 06/09/2021 |
| INV-33180000013 | 11 | 395 | 87510 | 278.28 | 12 | 11/07/2021 |
| INV-33415800011 | 82 | 395 | 89207 | 157.5 | 6 | 23/11/2021 |
| INV-33764100058 | 48 | 395 | 87510 | 69.57 | 3 | 10/08/2021 |
| INV-33693900031 | 73 | 395 | 87510 | 46.38 | 2 | 26/09/2021 |

Hình 1. Dữ liệu của đơn đặt hàng rượu tại khu vực IOWA

2. *Region* - Khu vực

- Gồm 2 trường và 77 dòng dữ liệu

| RegionID | Region |
|----------|------------|
| 6 | BENTON |
| 7 | BLACK HAWK |
| 8 | BOONE |
| 9 | BREMER |
| 11 | BUENA VIST |
| 13 | CALHOUN |
| 14 | CARROLL |
| 15 | CASS |
| 16 | CEDAR |
| 17 | CERRO GORD |
| 18 | CHEROKEE |
| 19 | CHICKASAW |

Hình 2. Dữ liệu khu vực

3. *Product* - Sản phẩm

- Gồm 5 trường và 12 dòng dữ liệu

| ProductID | Product_Name | Bottle_volume_ml | State_bottle_retail | COG |
|-----------|-----------------|------------------|---------------------|-------|
| 18928 | Lauders | 1750 | 14.24 | 9.49 |
| 23823 | Five Star | 200 | 1.97 | 1.31 |
| 84152 | 99 Oranges | 50 | 7.74 | 5.16 |
| 84159 | 99 Apples | 750 | 11.25 | 7.5 |
| 84170 | 99 Bananas | 750 | 11.25 | 7.5 |
| 84172 | 99 Bananas Mini | 50 | 7.74 | 5.16 |
| 84187 | 99 Cinnamon | 100 | 1.43 | 0.95 |
| 84217 | 99 Grapes | 100 | 1.43 | 0.95 |
| 84404 | 99 Mangoes | 100 | 1.43 | 0.95 |
| 87510 | 1800 Silver | 750 | 23.19 | 15.46 |
| 88922 | 1800 Cristalino | 750 | 44.64 | 29.76 |
| 89207 | 1800 Reposado | 1000 | 26.25 | 17.5 |

Hình 3. Dữ liệu về sản phẩm được cung cấp tại IOWA

4. *Supplier* - Nhà cung cấp

- Gồm 2 trường và 3 dòng dữ liệu

| SupplierID | Supplier |
|------------|-----------------------|
| 55 | SAZERAC NORTH AMERICA |
| 297 | Laird & Company |
| 395 | PROXIMO |

Hình 4. Dữ liệu của các nhà cung cấp tại khu vực IOWA

II. THIẾT KẾ CƠ SỞ DỮ LIỆU

1. Thiết kế cơ sở dữ liệu Staging

Các bảng ở CSDL Staging trong quá trình ETL được sử dụng cho mục đích tiền xử lý và chuẩn bị dữ liệu trước khi đưa vào kho dữ liệu (Data Warehouse).

- **Staging Orders:** Lưu trữ và xử lý những thông tin về đơn đặt hàng được lấy từ nguồn trước khi đưa dữ liệu vào bảng Dim_Time và Fact_Sales. Tạo thêm 1 cột Profit để tiện cho việc update dữ liệu cho Fact
- **Staging Product:** Lưu trữ và xử lý những thông tin về các sản phẩm được lấy từ nguồn trước khi đưa dữ liệu vào bảng Dim_Product.
- **Staging Supplier:** Lưu trữ và xử lý những thông tin về các nhà cung cấp được lấy từ nguồn trước khi đưa dữ liệu vào bảng Dim_Supplier.

- **Staging Region:** Lưu trữ và xử lý những thông tin về các khu vực được lấy từ nguồn trước khi đưa dữ liệu vào bảng Dim_Region.
- **Staging Fact_Sales:** Bảng có cấu trúc và kiểu dữ liệu tương tự với bảng Fact_Sales, lưu trữ dữ liệu của bảng Fact_Sales được lấy từ Staging trước khi thực hiện quá trình Incremental Load từ bảng này sang bảng Fact_Sales.

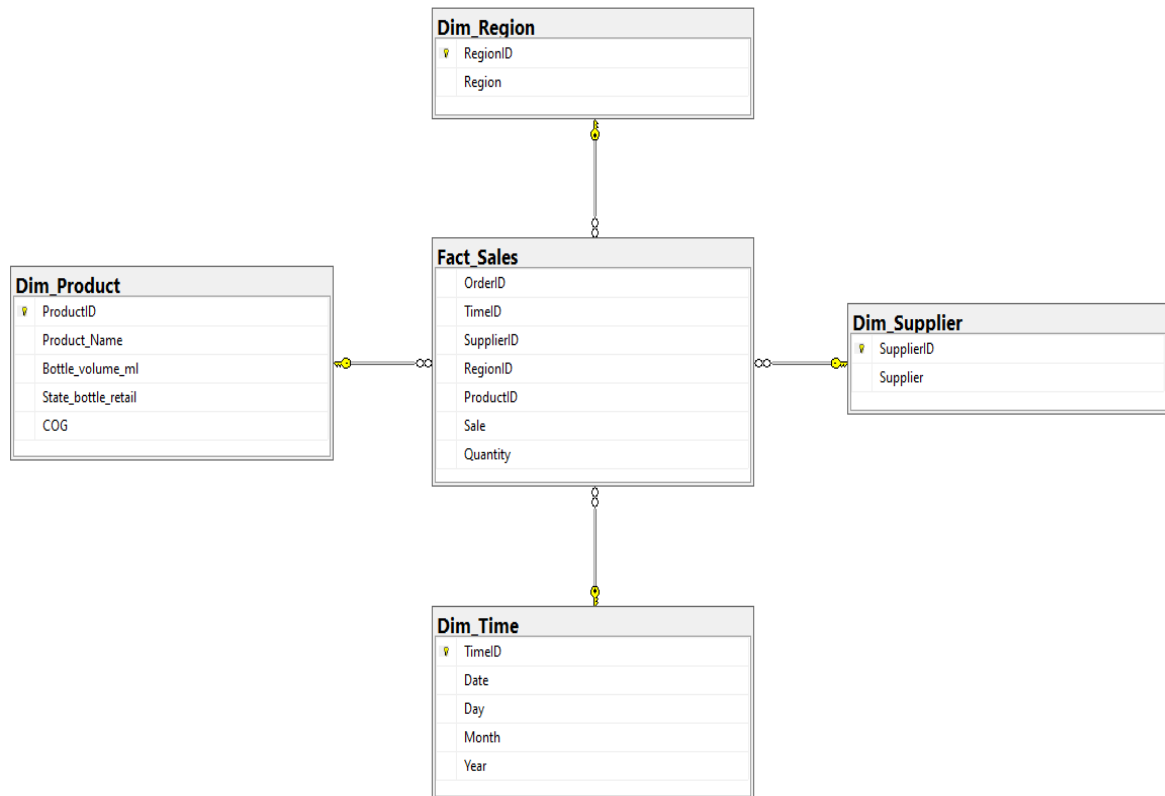
| | | | | | | | | | |
|---------------------------|---|------------|--------------|------------------|---------------------|-----------|------|----------|--|
| Staging Product | <table><tr><td>ProductID</td></tr><tr><td>Product_Name</td></tr><tr><td>Bottle_volume_ml</td></tr><tr><td>State_bottle_retail</td></tr><tr><td>COG</td></tr><tr><td></td></tr></table> | ProductID | Product_Name | Bottle_volume_ml | State_bottle_retail | COG | | | |
| ProductID | | | | | | | | | |
| Product_Name | | | | | | | | | |
| Bottle_volume_ml | | | | | | | | | |
| State_bottle_retail | | | | | | | | | |
| COG | | | | | | | | | |
| | | | | | | | | | |
| Staging Order | <table><tr><td>OrderID</td></tr><tr><td>SupplierID</td></tr><tr><td>RegionID</td></tr><tr><td>ProductID</td></tr><tr><td>Date</td></tr><tr><td>Sale</td></tr><tr><td>Quantity</td></tr><tr><td></td></tr></table> | OrderID | SupplierID | RegionID | ProductID | Date | Sale | Quantity | |
| OrderID | | | | | | | | | |
| SupplierID | | | | | | | | | |
| RegionID | | | | | | | | | |
| ProductID | | | | | | | | | |
| Date | | | | | | | | | |
| Sale | | | | | | | | | |
| Quantity | | | | | | | | | |
| | | | | | | | | | |
| Staging Region | <table><tr><td>RegionID</td></tr><tr><td>Region</td></tr><tr><td></td></tr></table> | RegionID | Region | | | | | | |
| RegionID | | | | | | | | | |
| Region | | | | | | | | | |
| | | | | | | | | | |
| Staging Supplier | <table><tr><td>SupplierID</td></tr><tr><td>Supplier</td></tr><tr><td></td></tr></table> | SupplierID | Supplier | | | | | | |
| SupplierID | | | | | | | | | |
| Supplier | | | | | | | | | |
| | | | | | | | | | |
| Staging Fact Sales | <table><tr><td>OrderID</td></tr><tr><td>TimeID</td></tr><tr><td>SupplierID</td></tr><tr><td>RegionID</td></tr><tr><td>ProductID</td></tr><tr><td>Sale</td></tr><tr><td>Quantity</td></tr><tr><td></td></tr></table> | OrderID | TimeID | SupplierID | RegionID | ProductID | Sale | Quantity | |
| OrderID | | | | | | | | | |
| TimeID | | | | | | | | | |
| SupplierID | | | | | | | | | |
| RegionID | | | | | | | | | |
| ProductID | | | | | | | | | |
| Sale | | | | | | | | | |
| Quantity | | | | | | | | | |
| | | | | | | | | | |

Hình 5. Thiết kế cơ sở dữ liệu của staging

2. Thiết kế Datawarehouse (Kho dữ liệu)

Gồm 4 bảng Dimension và 1 bảng Fact Table

- **Dim_Region:** Bảng chiều các khu vực
- **Dim_Time:** Bảng chiều thời gian
- **Dim_Supplier:** Bảng chiều các nhà cung cấp
- **Dim_Product:** Bảng chiều các sản phẩm
- **Fact_Sales:** Bảng sự kiện tổng hợp liên quan đến doanh thu của các cửa hàng tại nhiều khu vực



Hình 6. Lược đồ hình sao của kho dữ liệu đích

III. XÂY DỰNG TIỀN TRÌNH ETL

1. Tiến trình xây dựng

Tổng quan về tiến trình ETL được xây dựng trong đề tài được thực hiện thông qua 2 tiến trình như sau:

- Tiến trình Đổ Dữ liệu từ Nguồn vào Staging:
Ban đầu, dữ liệu từ nguồn được chọn để đổ vào khu vực staging. Nhằm đảm bảo tính toàn vẹn và tính nhất quán của dữ liệu, phương pháp tải toàn bộ dữ liệu (FullLoad) được thực hiện. Trước khi đổ dữ liệu vào staging, quá trình xóa dữ liệu được thực hiện trên các bảng tương ứng để làm mới không gian lưu trữ. Sau đó, dữ liệu được đổ từ nguồn vào các bảng staging tương ứng.
- Tiến trình Đổ Dữ liệu từ Staging vào Kho Dữ liệu:
Sau khi dữ liệu đã được chứa trong staging, tiến trình đổ dữ liệu từ staging vào kho dữ liệu được thực hiện. Quá trình này áp dụng kỹ thuật tải dữ liệu tăng dần Incremental Load với mục đích chỉ

đổ những dữ liệu mới và dữ liệu đã có cập nhật lại vào các bảng chiều (Dimension) và bảng số liệu (Fact). Đối với phương pháp tải dữ liệu tăng dần sử dụng kỹ thuật SCD (Slowly Changing Dimension) để xử lý những thay đổi trong dữ liệu của các bảng chiều.

- Đối với bảng số liệu (Fact), tạo ra một bảng tạm là Stag_Fact_Sales để so sánh với bảng đích Fact_Sales. Quá trình so sánh này cho phép cập nhật những dữ liệu mới và cập nhật lại những dữ liệu có sự thay đổi từ bảng tạm sang bảng đích.

2. Module tổng hợp



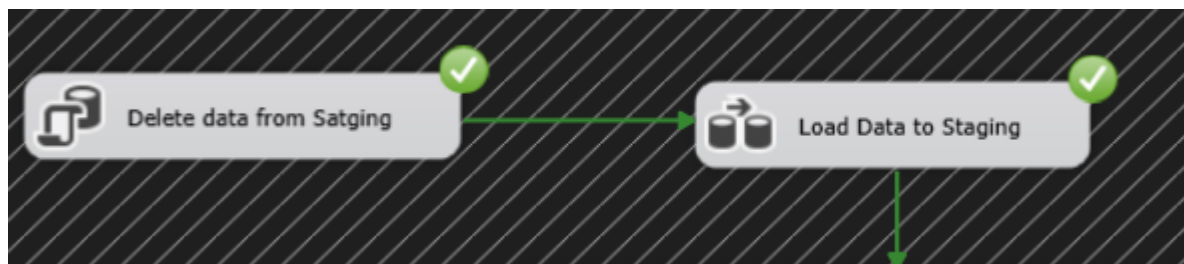
Hình 7. Module ETL tổng hợp

- **Delete data from Staging:** Là module dùng để xóa dữ liệu các bảng ở cơ sở dữ liệu Staging trước khi thực hiện đổ dữ liệu từ Nguồn vào Staging.
- **Load data to Staging:** Là module tiến hành đổ dữ liệu từ Nguồn vào Staging.

- ***Incremental Data Load to Dimension (Dim_Product, Dim_Region, Dim_Supplier, Dim_Time):*** Là module tiến hành lấy dữ liệu từ các bảng ở Staging để đổ vào các bảng Dim theo phương pháp Incremental Data Load.
- ***Load data to Staging_Fact_Sales:*** Là module thực hiện thông qua công cụ Lookup của SSIS bằng cách tìm khóa ngoại để có dữ liệu trùng khớp với các bảng Dim, sau đó dữ liệu được đổ vào bảng tạm của bảng Fact là Staging_Fact_Sales.
- ***Incremental Data Load to Fact_Sales:*** Là module thực hiện việc so sánh dữ liệu ở bảng Staging_Fact_Sales so với bảng Fact_Sales để cập nhật những dữ liệu mới và cập nhật lại những dữ liệu có thay đổi vào bảng Fact_Sales cho mỗi lần ETL.

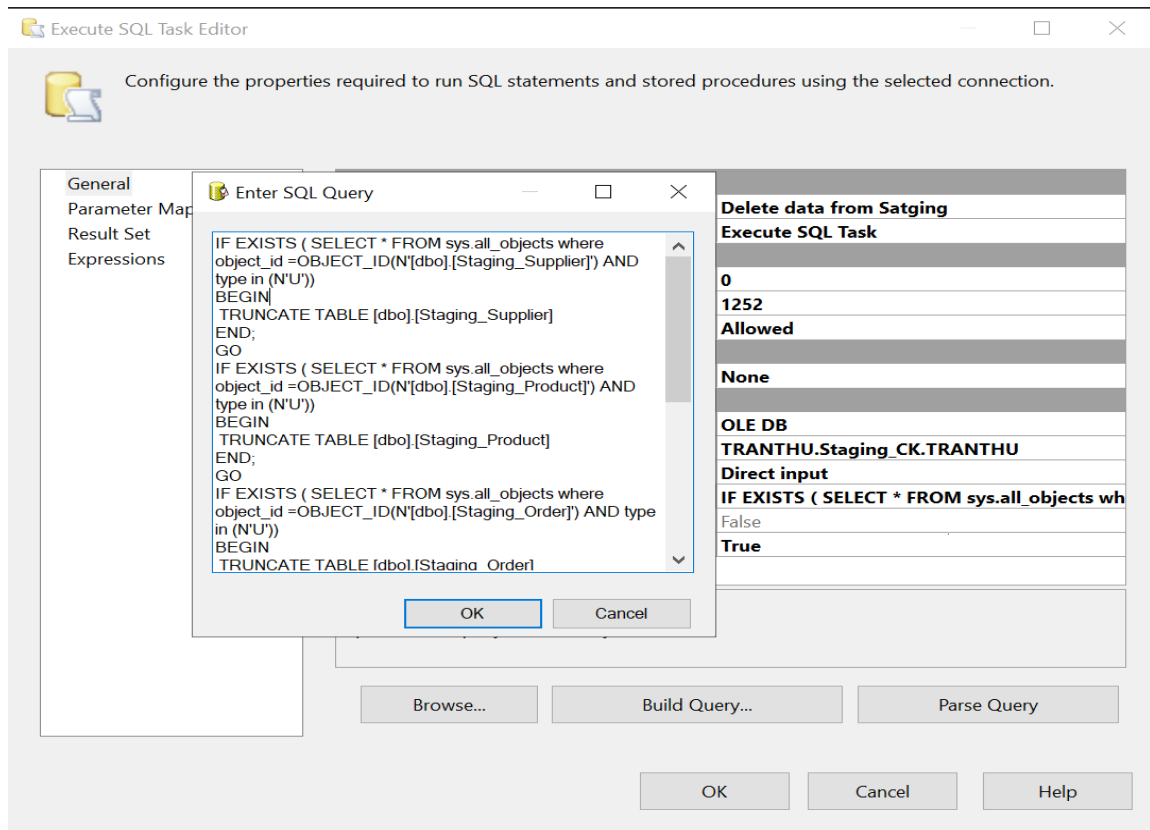
2.1 ETL cho các bảng ở Staging

Trong giai đoạn đầu của quá trình ETL, thực hiện lấy dữ liệu từ Nguồn cho các bảng ở Staging bằng phương pháp tải toàn bộ dữ liệu (FullLoad), với phương pháp tải toàn bộ dữ liệu từ Nguồn vào Staging cho mỗi lần chạy ETL cần xóa tất cả dữ liệu ở các bảng Staging trước khi đổ dữ liệu từ Nguồn vào



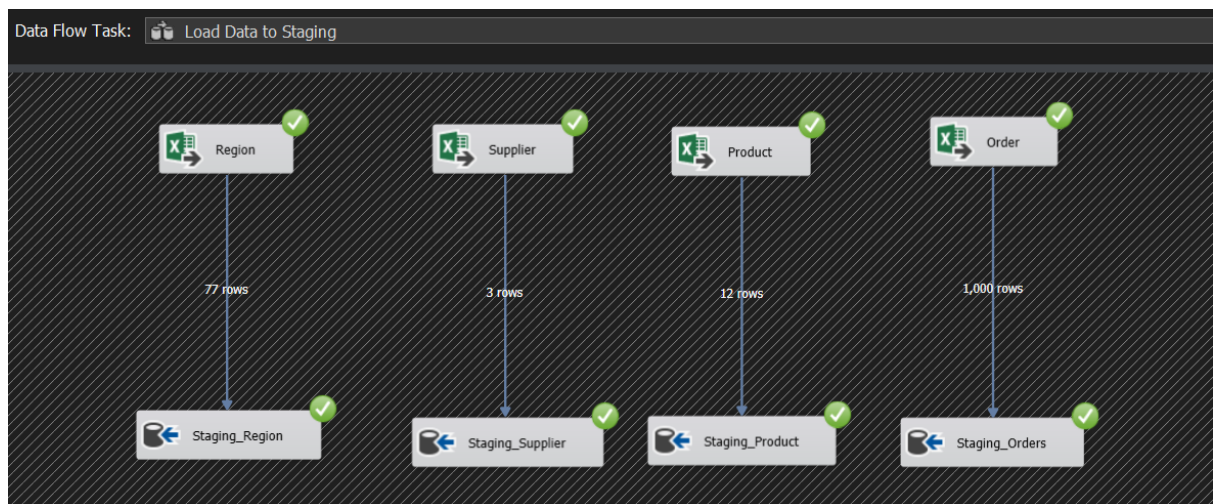
Hình 8. Quá trình ETL cho các bảng ở staging

- Đầu tiên Module Delete data from Staging sẽ thực hiện xóa dữ liệu ở các bảng Staging trước khi thực hiện quá trình đổ dữ liệu nguồn vào Staging



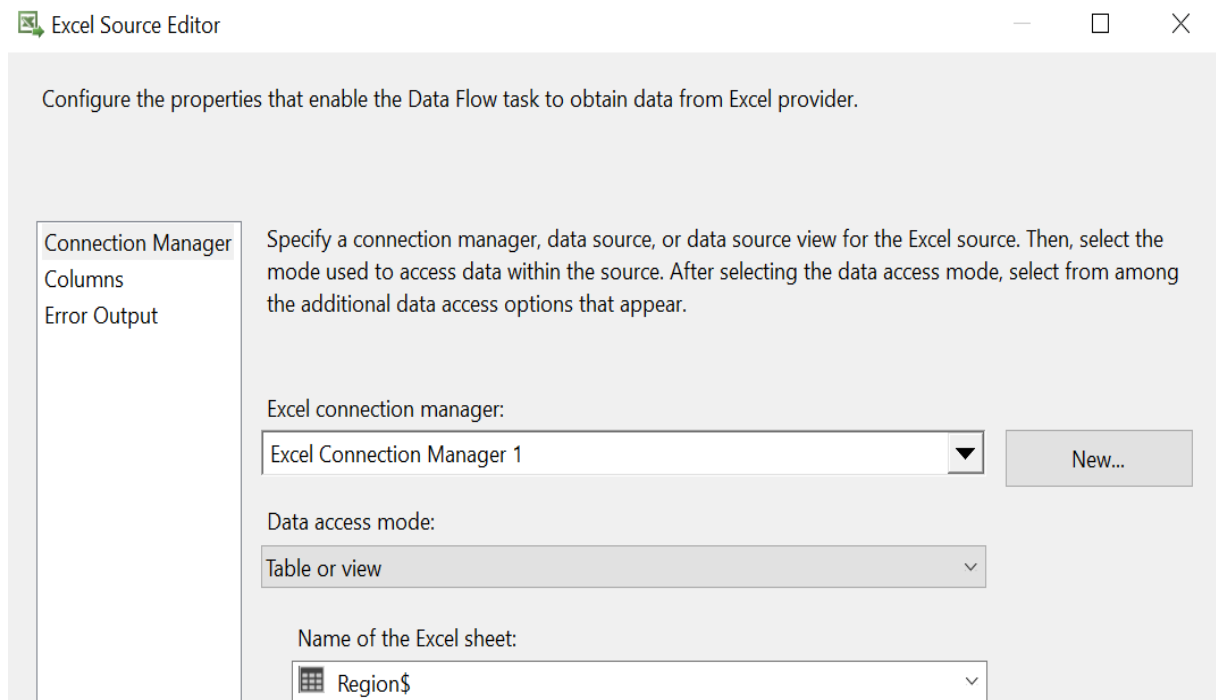
Hình 9. Module thực hiện xóa dữ liệu ở Staging

- Sau khi xóa dữ liệu ở Staging, module Load data to Staging sẽ thực hiện đổ dữ liệu từ nguồn vào các bảng Staging.



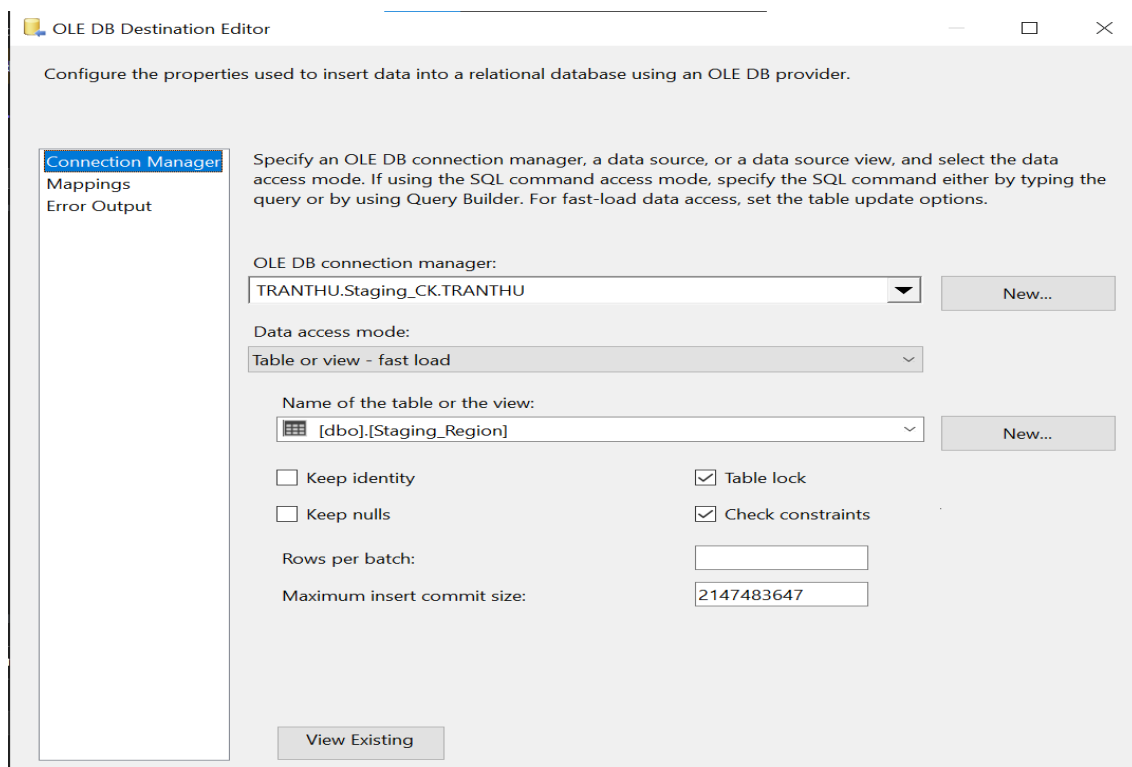
Hình 10. Data flow lấy dữ liệu cho các bảng staging

- Thực hiện lấy dữ liệu Region từ nguồn Excel



Hình 11. Lấy dữ liệu từ nguồn excel

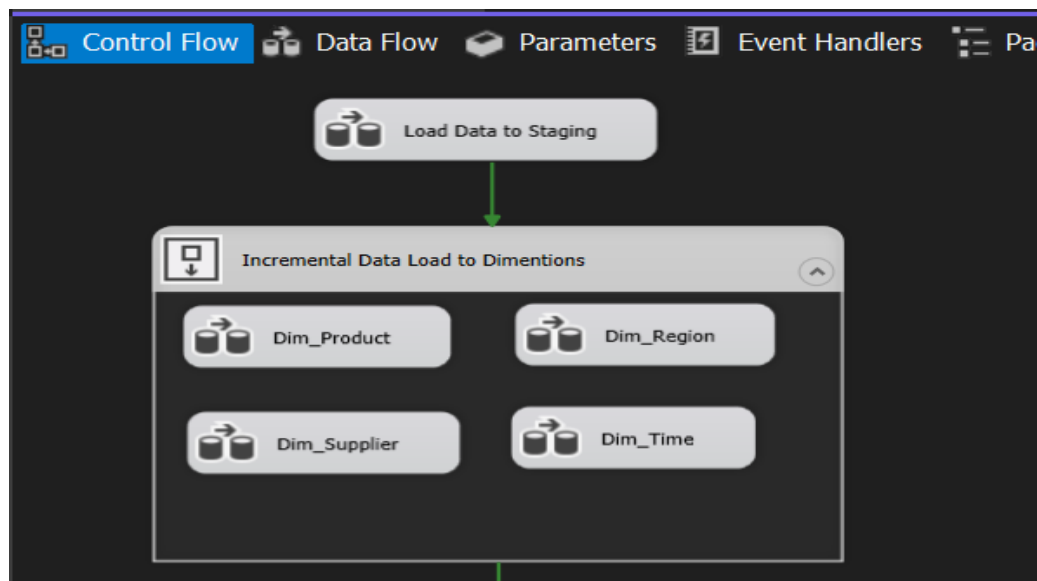
- Sau đó đổ trực tiếp vào Staging_Region đã tạo sẵn trước đó. Các Staging Supplier, Product thực hiện tương tự



Hình 12. OLE DB Source lấy dữ liệu cho bảng Staging_Region

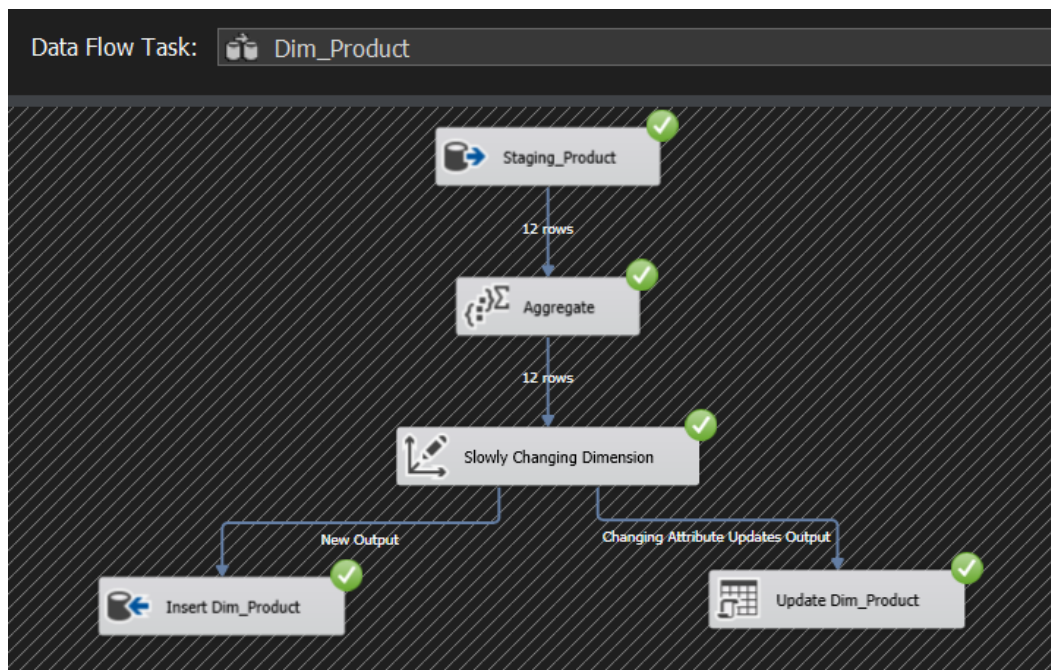
2.2 ETL cho các bảng Dim

Sau khi dữ liệu đã được đổ vào Staging, thực hiện ETL lấy dữ liệu từ các bảng Staging để đổ vào các bảng Dim theo phương pháp tải dữ liệu tăng dần (Incremental Data Load).



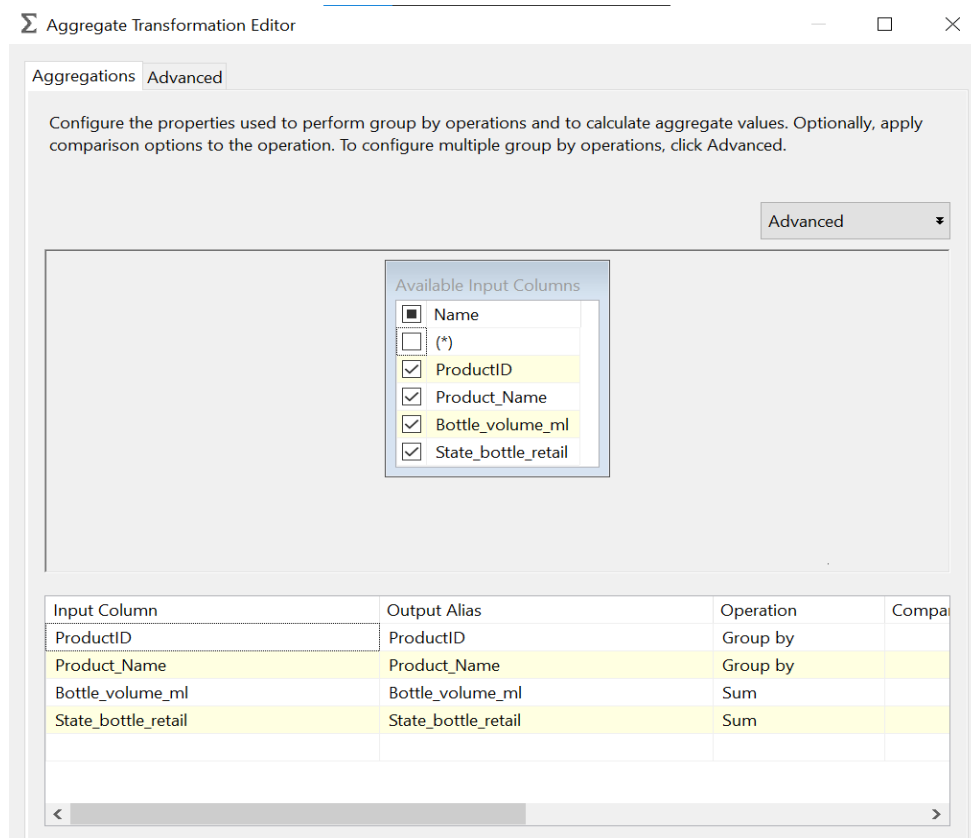
Hình 13. Quá trình ETL cho các bảng Dim

- Trong quá trình ETL từ Staging vào các bảng Dim sử dụng phương pháp Slowly Changing Dimension (SCD) cho việc Incremental Load.



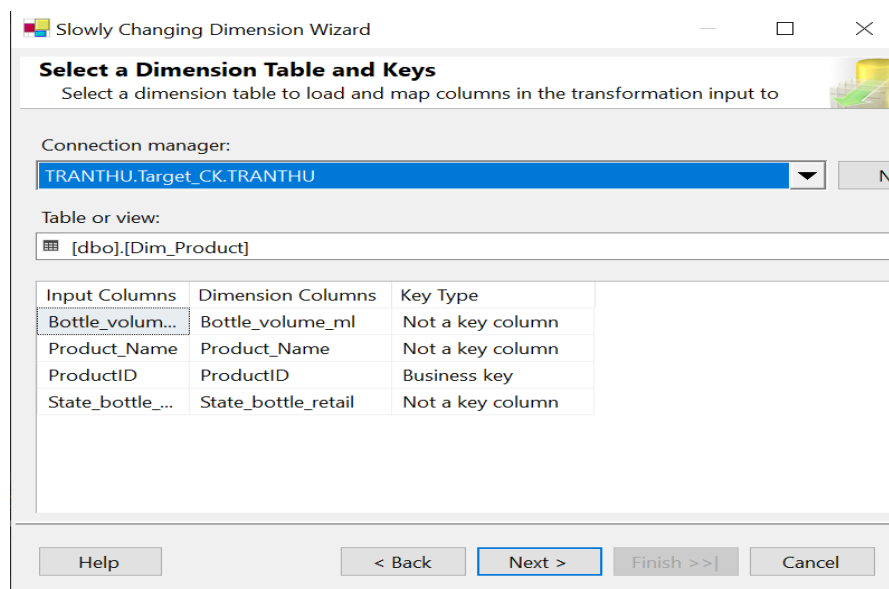
Hình 14. ETL cho bảng Dim_Product

- Sử dụng Aggregate để chọn ra và gộp nhóm những cột được sử dụng trong tiến trình đồ dữ liệu vào các bảng Dim.

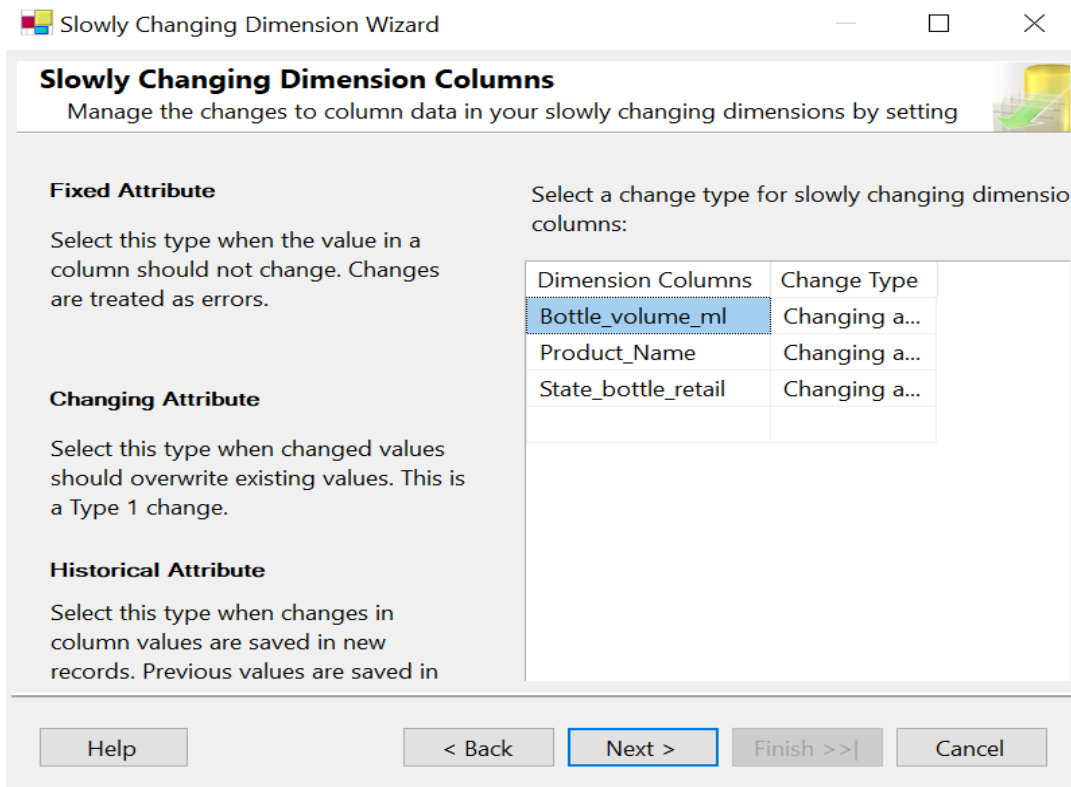


Hình 15. Aggregate cho bảng Dim_Product

- Sử dụng Slowly Changing Dimension cho quá trình Incremental Data Load từ Staging vào các bảng Dim.

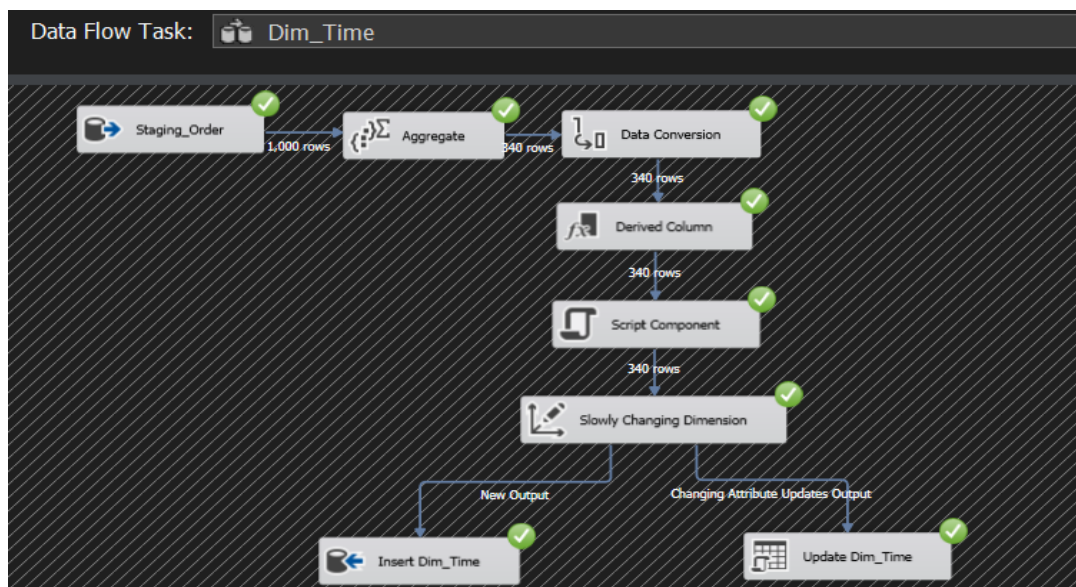


Hình 16. Select a Dimension Table and Keys cho bảng Dim_Product.



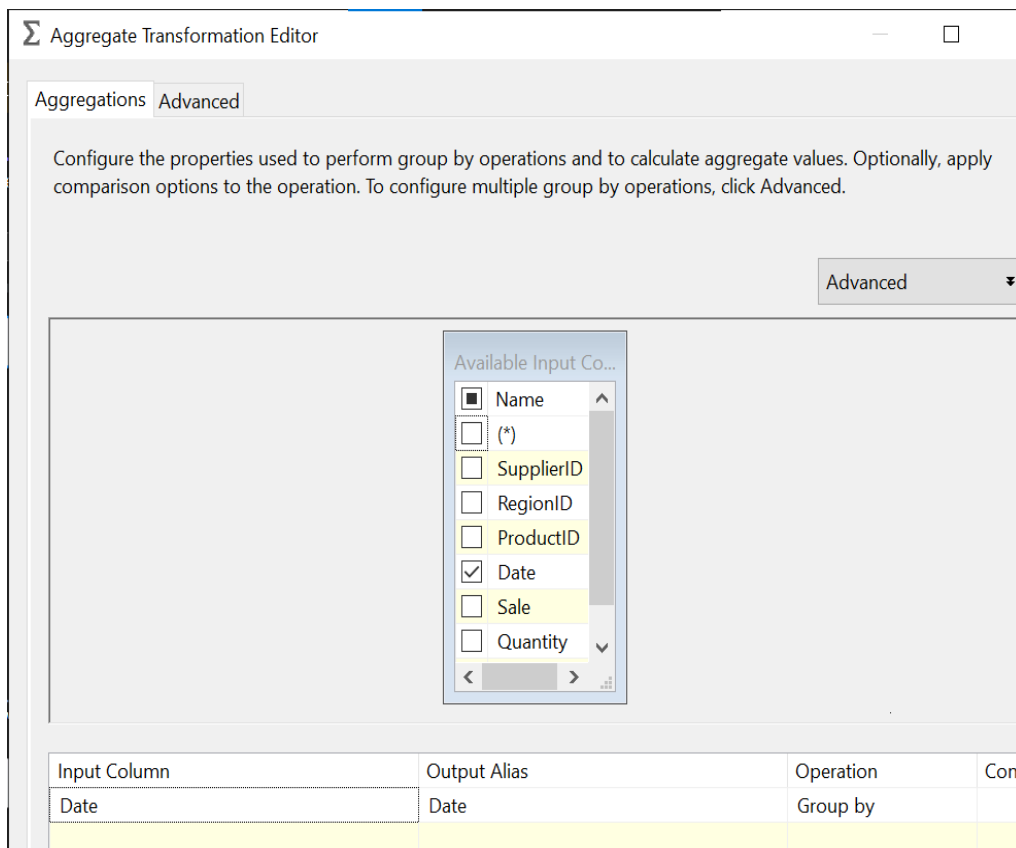
Hình 17. Slowly changing Dimension Columns cho bảng Dim_Product

- Các Dimension Supplier và Product thực hiện tương tự. Còn về Dimension Time. Đầu tiên lấy dữ liệu từ Staging_Order và chọn ra cột Date để tiến hành thiết kế Dim_Time



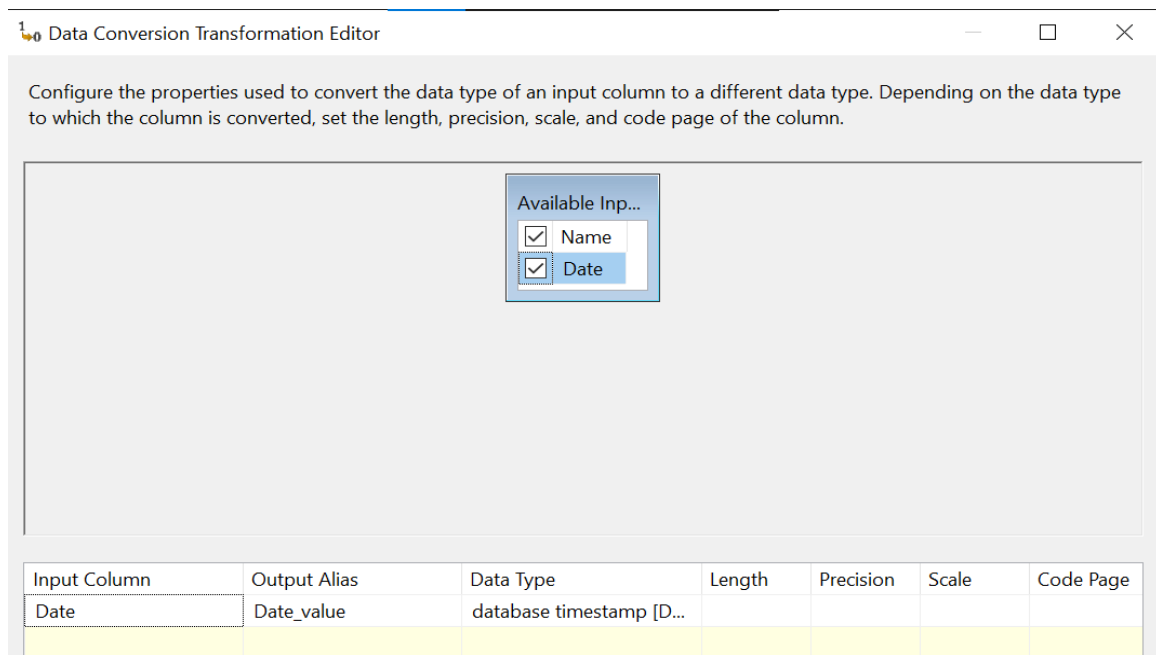
Hình 18. Quá trình ETL cho bảng Dim_Time

- Sử dụng Aggregate để chọn ra và gộp nhóm Date



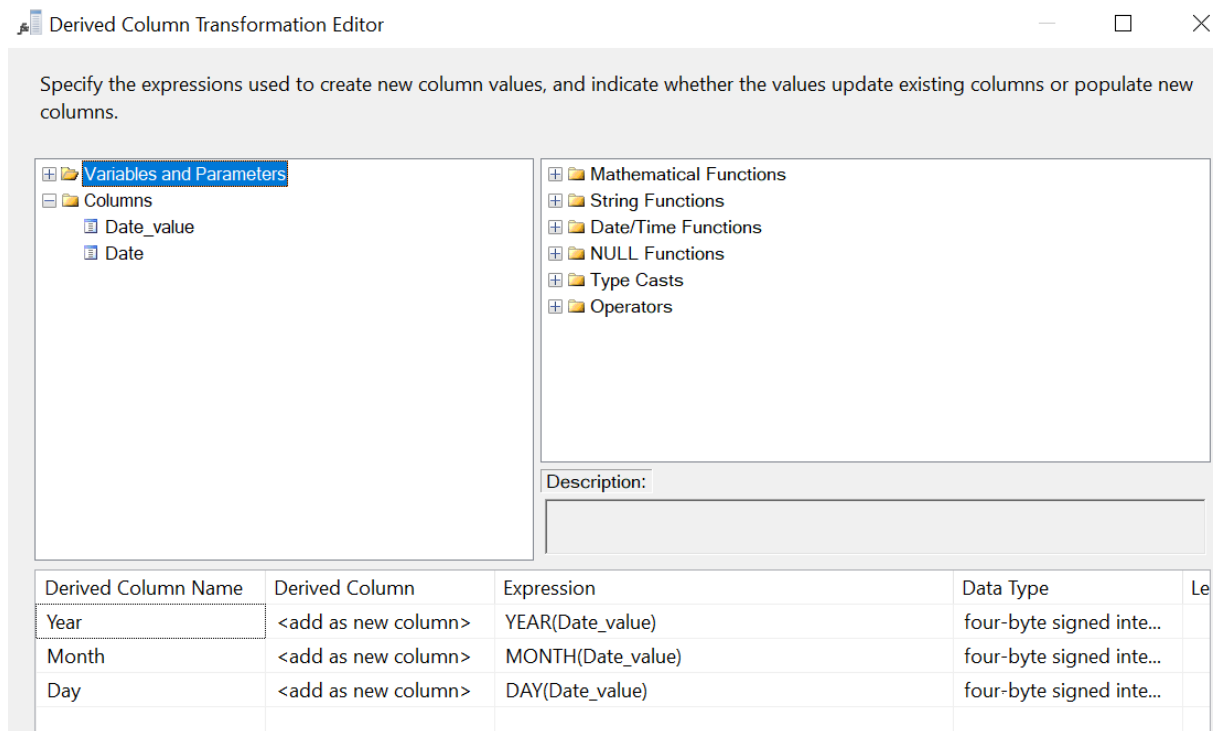
Hình 19. Aggregate cho bảng Dim_Time

- Sử dụng Data Conversion Transformation để thực hiện chuyển đổi kiểu dữ liệu sang dạng database timestamp



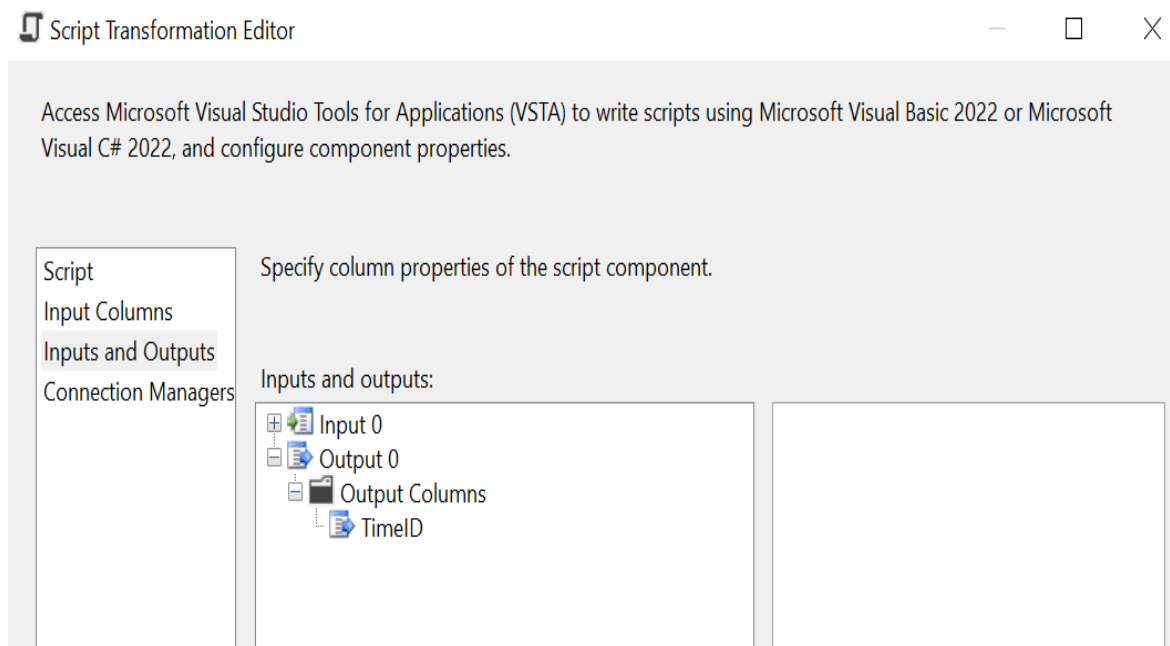
Hình 20. Data Conversion Transformation cho bảng Dim_Time

- Tiếp theo dùng Derived Column để lấy ra các cột Year, Month, Day từ cột Date vừa chuyển đổi kiểu dữ liệu



Hình 21. Derived Column cho bảng Dim_Time

- Sử dụng Script Transformation để tạo TimeID là khóa chính cho Dim_Time



Hình 22. . Scrip Transformation cho bảng Dim_Time

- Sử dụng Slowly Changing Dimension

Slowly Changing Dimension Wizard

Select a Dimension Table and Keys

Select a dimension table to load and map columns in the transformation input to

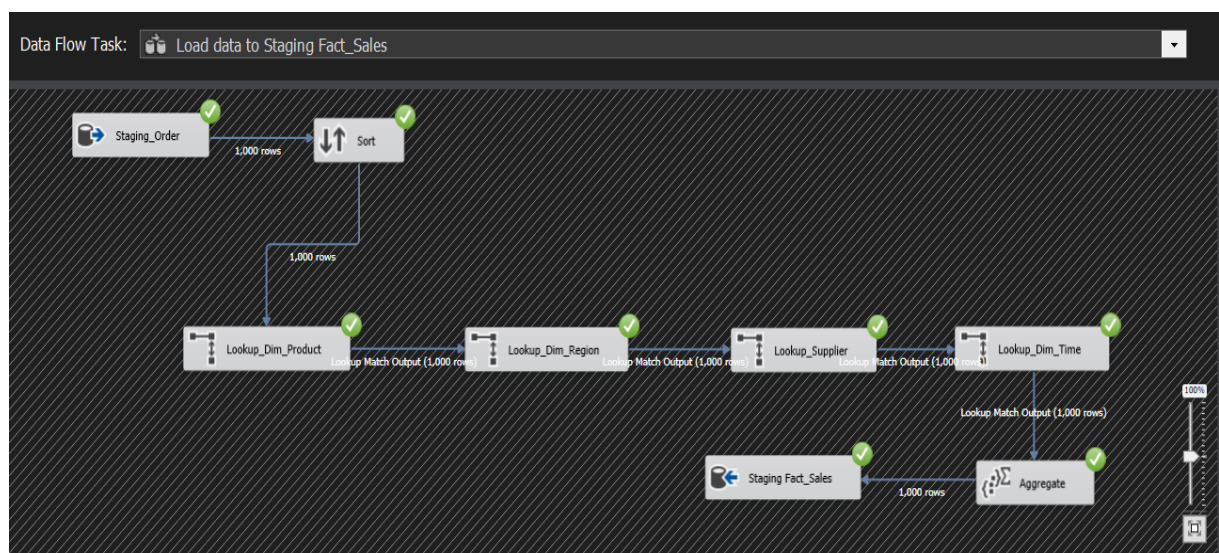
Connection manager:
 TRANTHU.Target_CK.TRANTHU

Table or view:
 [dbo].[Dim_Time]

| Input Columns | Dimension Columns | Key Type |
|---------------|-------------------|------------------|
| Date | Date | Not a key column |
| Day | Day | Not a key column |
| Month | Month | Not a key column |
| TimeID | TimeID | Business key |
| Year | Year | Not a key column |

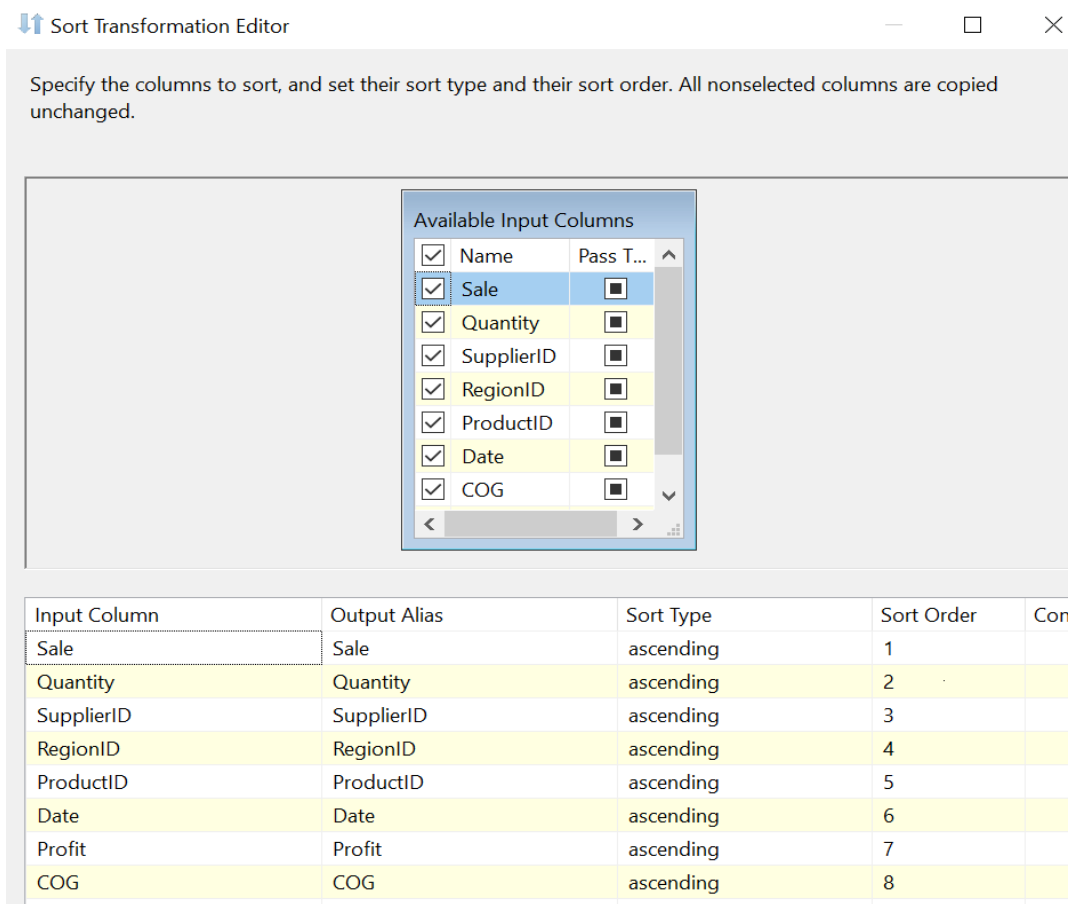
Hình 23. Select a Dimension Table and Keys cho bảng Dim_Time

2.3 ETL cho bảng Fact_Sales



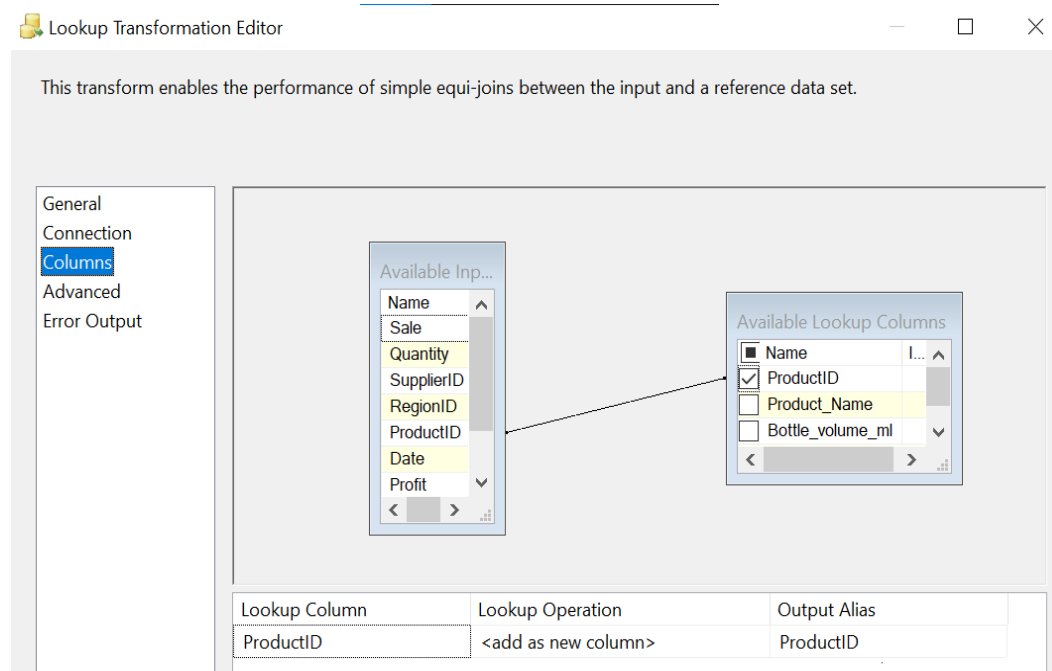
Hình 24. Quá trình ETL cho bảng Fact_Sales

- Sort để sắp xếp các cột được chọn

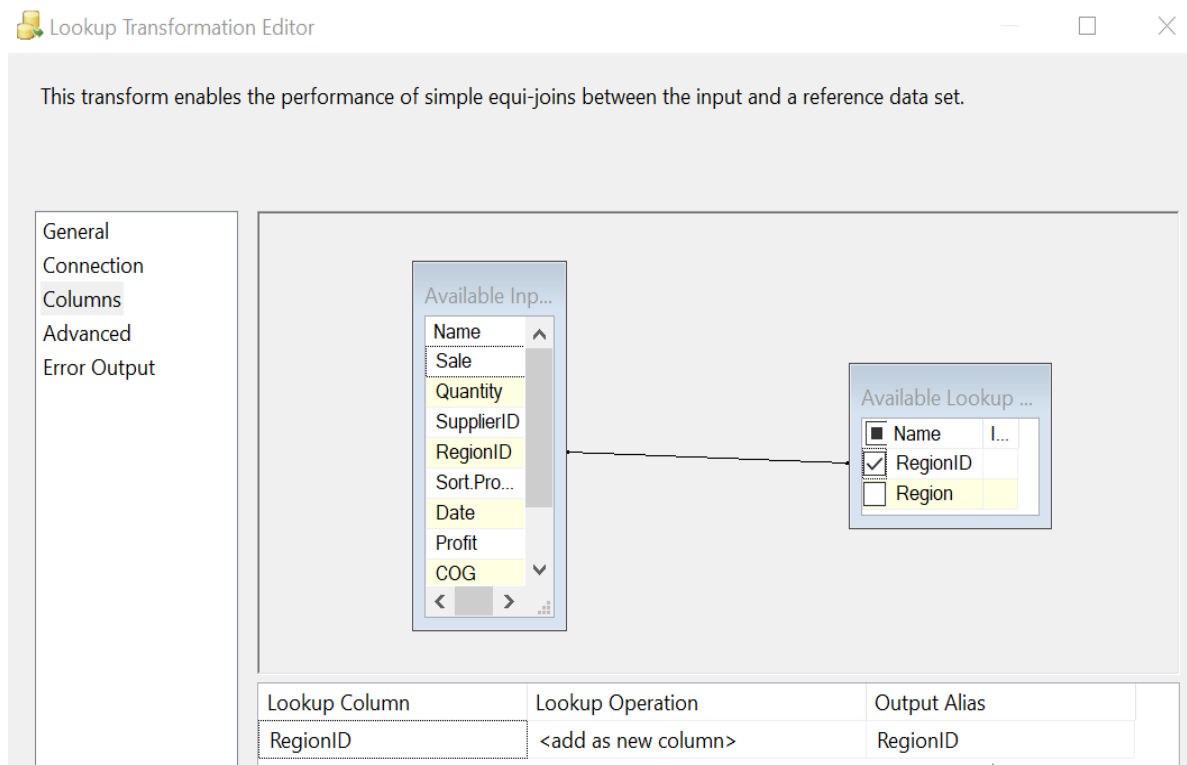


Hình 25. Sắp xếp các cột được chọn cho bảng Fact_Sales

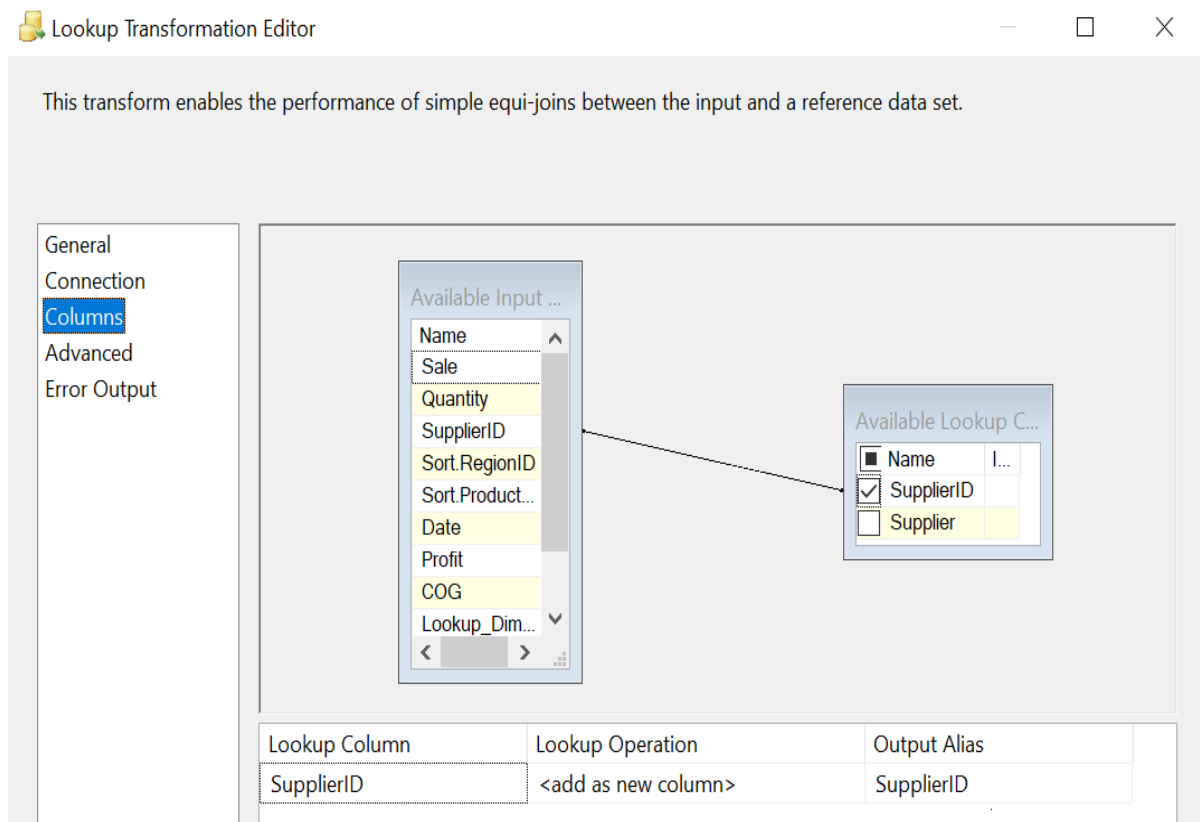
- Sử dụng Look up thực hiện tham chiếu giữa Dimension và Fact Table



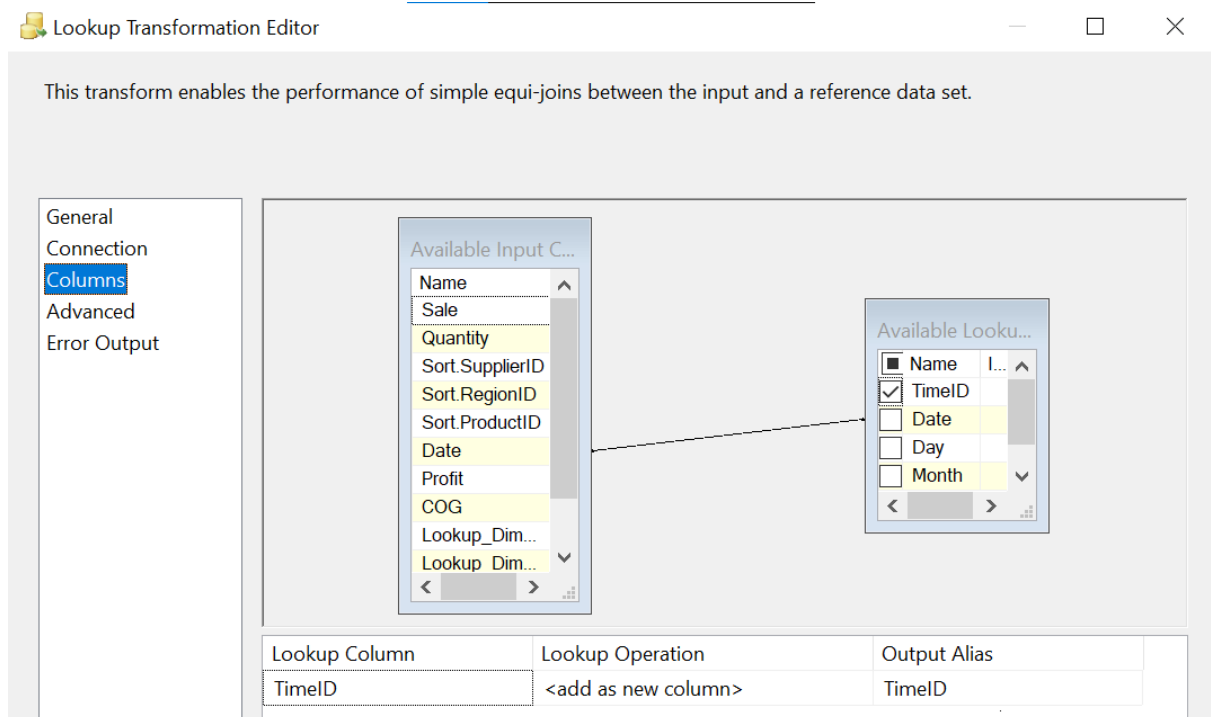
Hình 26. Lookup cho bảng Fact_Sales



Hình 27. Lookup cho bảng Fact_Sales

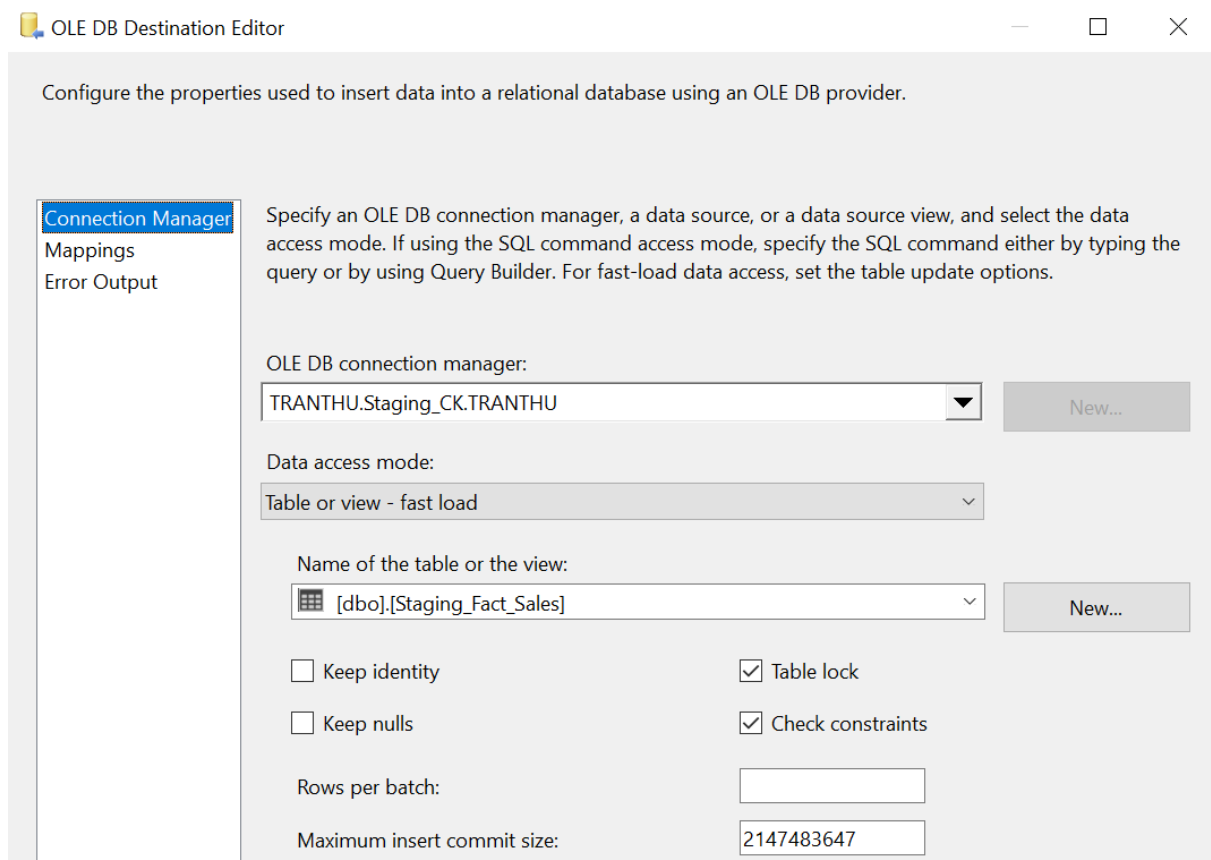


Hình 28. Lookup cho bảng Fact_Sales



Hình 29. Lookup cho bảng Fact_Sales

- Thực hiện đổ dữ liệu vào bảng tạm thời Staging_Fact_Sales



Hình 30. Thực hiện đổ dữ liệu vào bảng Fact_Sales

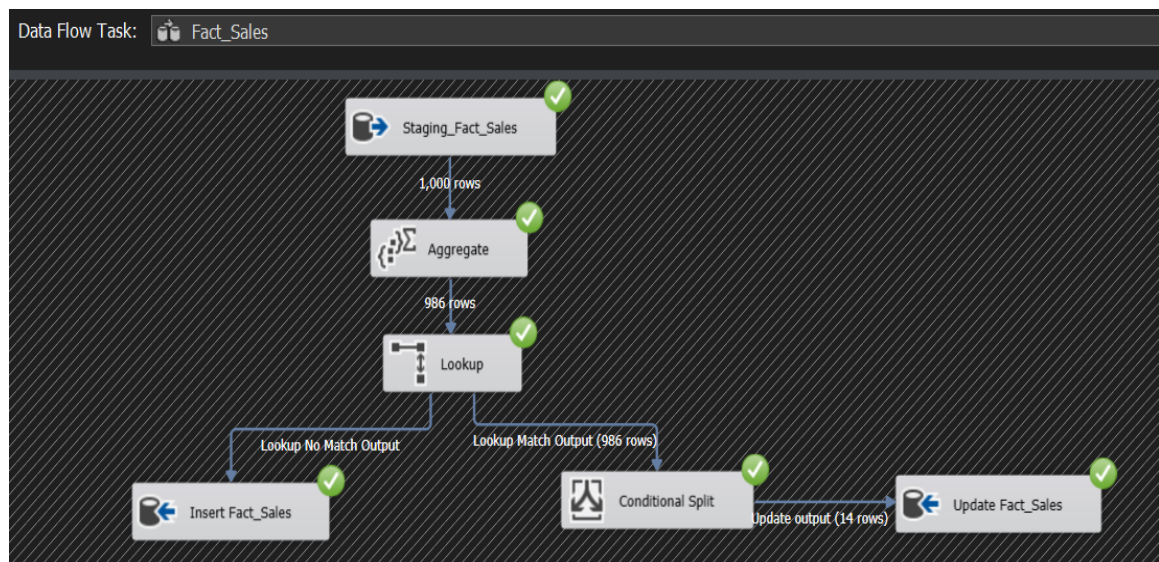
3. ETL cho bảng Fact_Sales

Tổng quan về quá trình ETL cho bảng Fact_Sales đảm bảo việc Incremental Load từ bảng Stag_Fact_Sales vào bảng Fact_Sales. Giai đoạn này được thực hiện để kiểm tra sự tương thích giữa dữ liệu trong bảng tạm của bảng Fact (Stag_Fact_Sales) và dữ liệu trong bảng Fact (Fact_Sales). Mục tiêu chính của giai đoạn này là cập nhật những dữ liệu mới và cập nhật lại những dữ liệu có thay đổi từ bảng tạm Stag_Fact_Sales vào bảng Fact_Sales cho mỗi lần chạy ETL.

Module Incremental Data Load to Fact_Sales được thực hiện thông qua các bước:

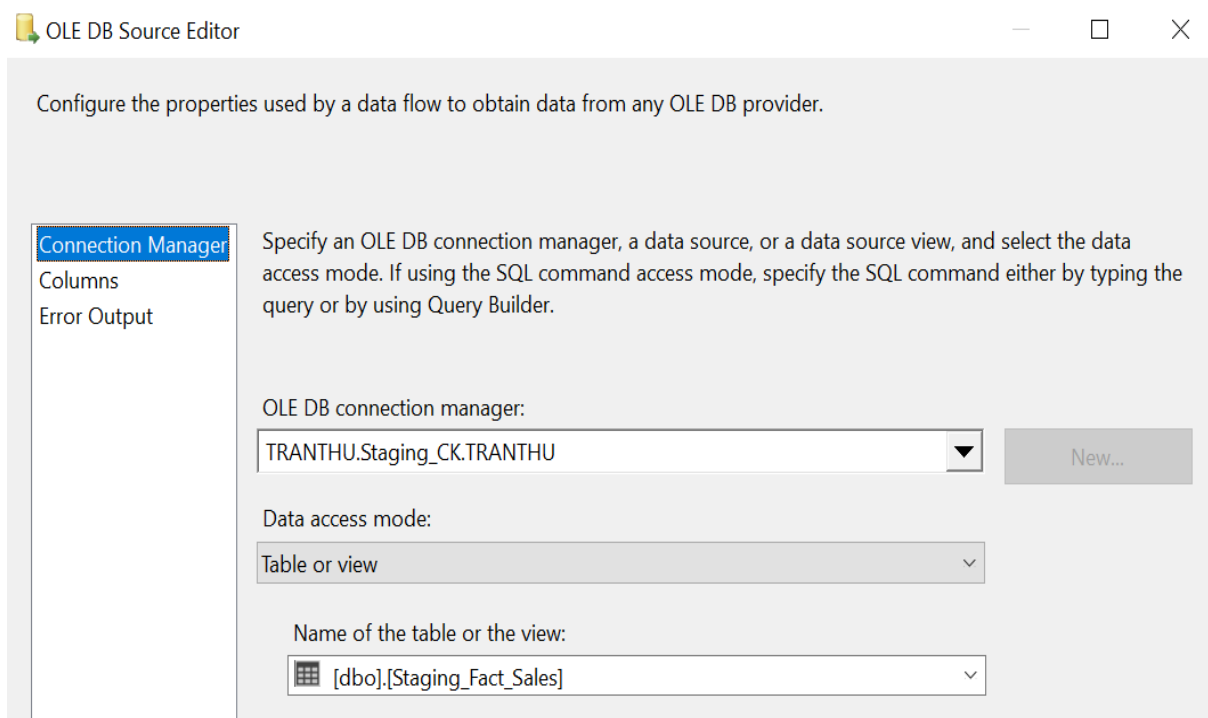
Bước 1: Module Compares Staging_Fact_Sales with Fact_Sales:

Kiểm tra và So sánh dữ liệu trong bảng Staging_Fact_Sales với bảng Fact (Fact_Sales) để tìm ra những dữ liệu mới và dữ liệu có sự thay đổi nhằm phục vụ cho phương pháp Incremental Load. Để làm điều này, chúng ta sử dụng kỹ thuật Lookup để xác định dữ liệu mới và dữ liệu có sự thay đổi thông qua việc ánh xạ các trường dữ liệu định danh giữa hai bảng.



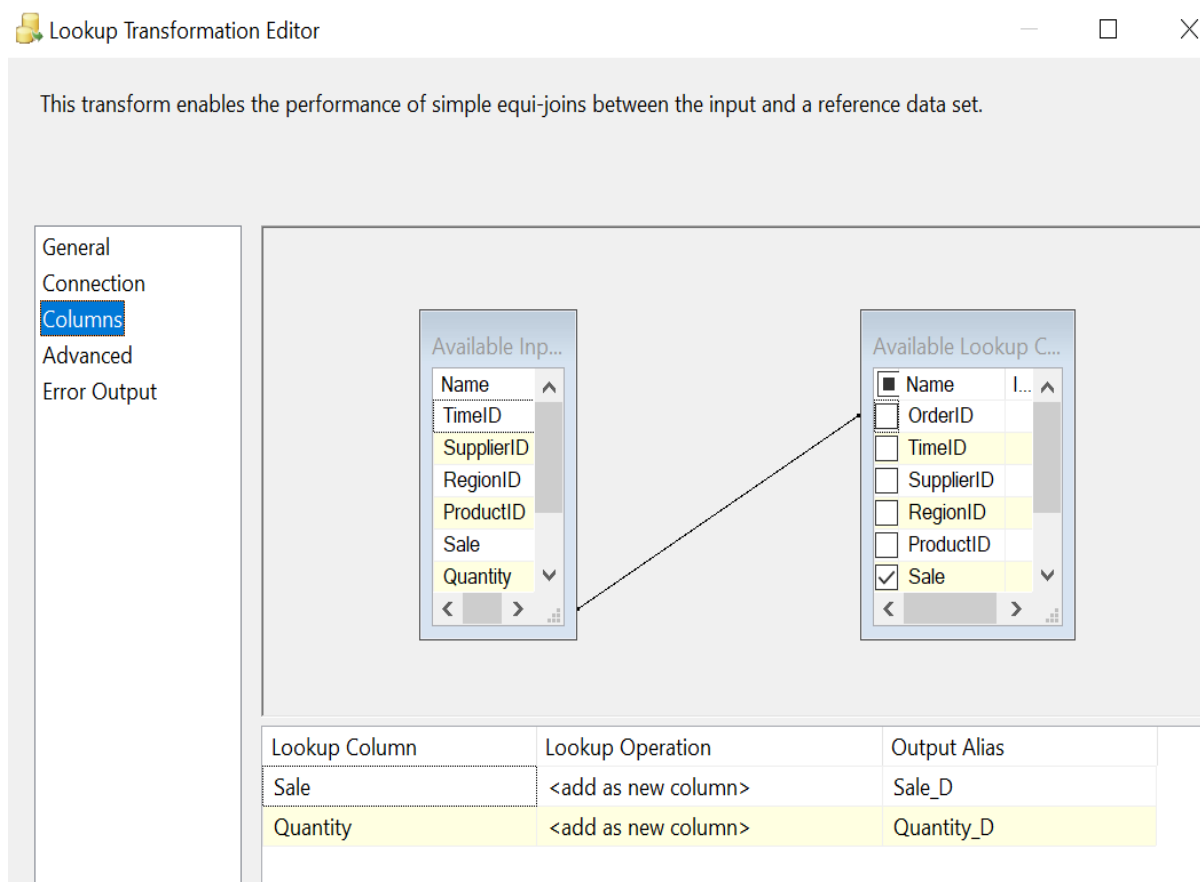
Hình 31. ETL cho bảng Fact_Sales

- Lấy dữ liệu từ bảng tạm của Fact là Staging_Fact_Sales để thực hiện so sánh



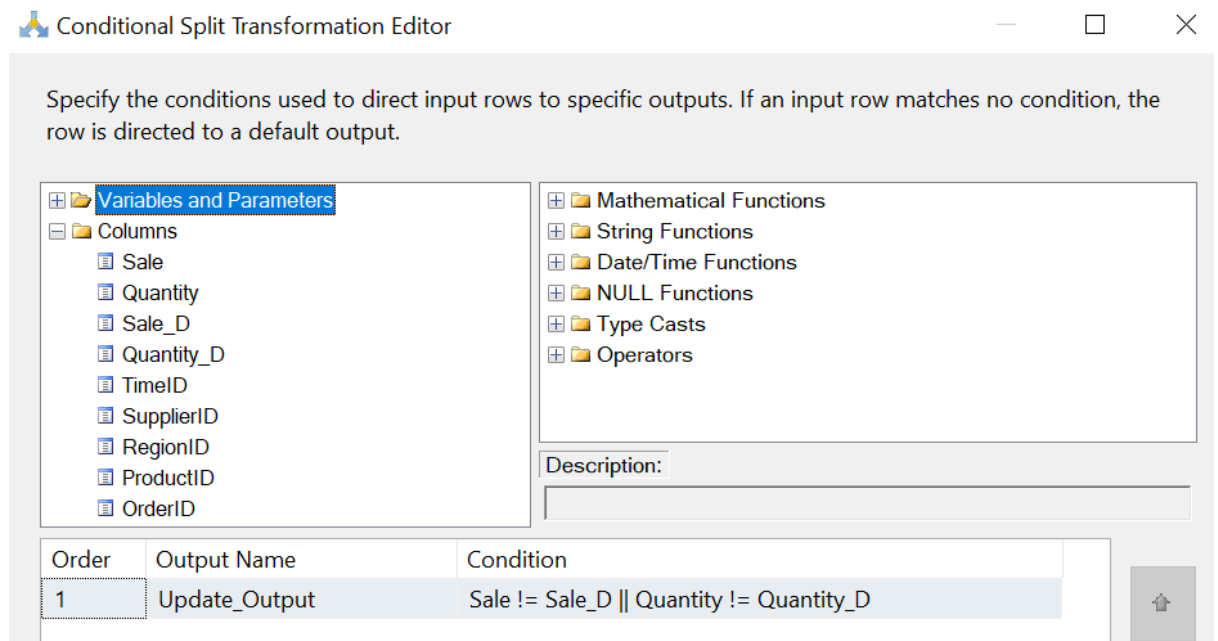
Hình 32. Thực hiện đổ dữ liệu vào bảng Staging_Fact_Sales

- Sử dụng Look up để tham chiếu bảng tạm Staging_Fact_Sales và bảng Fact_Sales và lấy ra các độ đo từ bảng Fact để so sánh



Hình 33. Lookup cho bảng Fact_Sales

- Có 2 trường hợp Look No Match Output có nghĩa là dữ liệu không khớp tức là dữ liệu chưa có trong Fact_Sales sẽ được insert trực tiếp vào Fact_Sales. Look Match Output có nghĩa là dữ liệu đã có trong Fact nên sẽ sử dụng Conditional Split để so sánh 2 bảng. Nếu các độ đo trong bảng tạm Staging_Fact_Sales khác với độ đo trong Fact_Sales thì sẽ thực hiện update dữ liệu

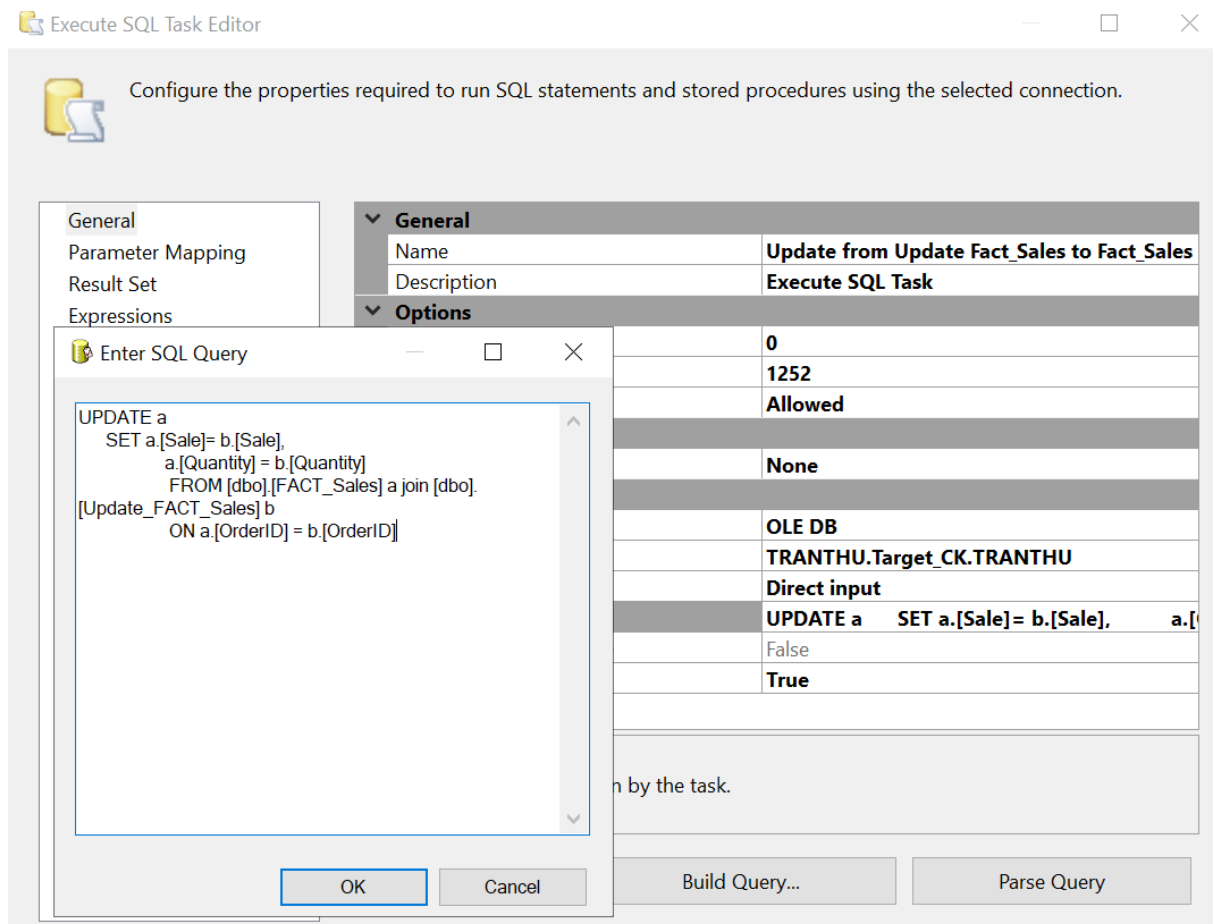


*Hình 34. Conditional Split Transformation Editor vào bảng
Staging_Fact_Sales*

Bước 2: Module Update from Update Fact_Sales to Fact_Sales thực hiện cập nhật dữ liệu từ Update Fact_Sales vào Fact_Sales:

- Sau khi thêm dữ liệu đã có sự cập nhật lại vào bảng Update Fact_Sales, chúng ta thực hiện bước cập nhật dữ liệu từ bảng Update Fact_Sales vào bảng Fact_Sales.

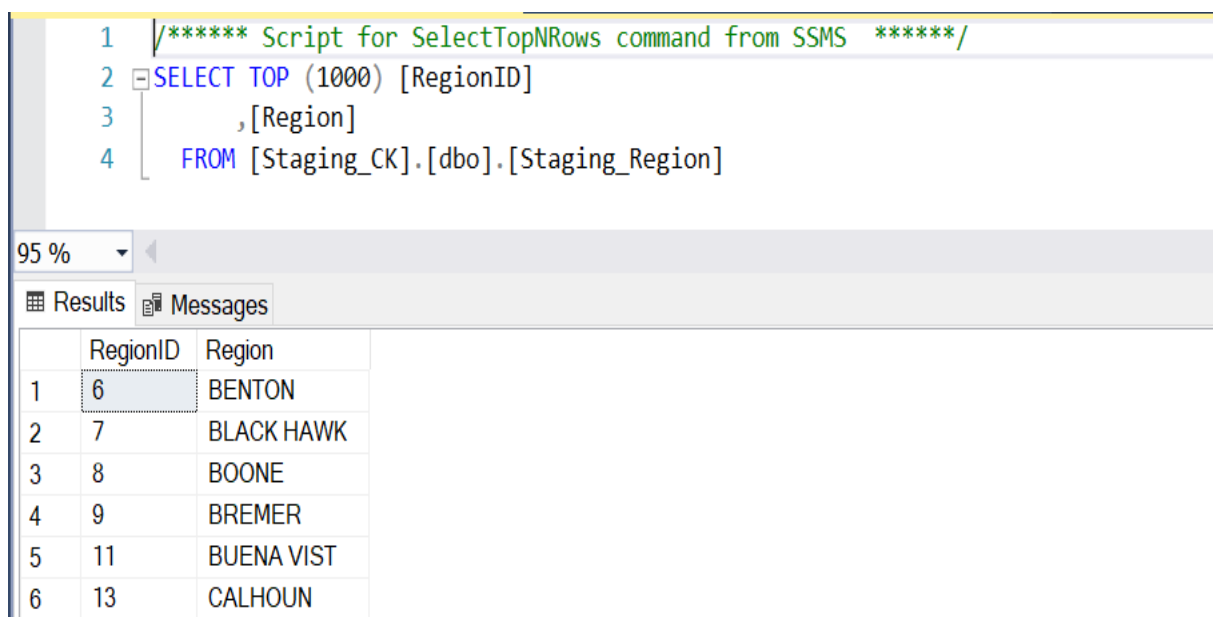
Sau bước cập nhật này, những dữ liệu đã được cập nhật sẽ được thêm vào bảng Fact_Sales.



Hình 35. Thực hiện lệnh update

4. Thực hiện truy vấn dữ liệu đã được đổ vào SQL

4.1 Staging_Region



Hình 36. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Region

4.2 Staging_Product

```

1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [ProductID]
3      , [Product_Name]
4      , [Bottle_volume_ml]
5      , [State_bottle_retail]
6  FROM [Staging_CK].[dbo].[Staging_Product]

```

95 %

Results Messages

| | ProductID | Product_Name | Bottle_volume_ml | State_bottle_retail |
|---|-----------|-----------------|------------------|---------------------|
| 1 | 18928 | Lauders | 1750 | 14.24 |
| 2 | 23823 | Five Star | 200 | 1.97 |
| 3 | 84152 | 99 Oranges | 50 | 7.74 |
| 4 | 84159 | 99 Apples | 750 | 11.25 |
| 5 | 84170 | 99 Bananas | 750 | 11.25 |
| 6 | 84172 | 99 Bananas Mini | 50 | 7.74 |
| 7 | 84187 | 99 Cinnamon | 100 | 1.43 |

Hình 37. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Product

4.3 Staging_Supplier

| | | | | |
|---|---|--|--|--|
| 1 | /***** Script for SelectTopNRows command from SSMS *****/ | | | |
| 2 | SELECT TOP (1000) [SupplierID] | | | |
| 3 | , [Supplier] | | | |
| 4 | FROM [Staging_CK].[dbo].[Staging_Supplier] | | | |

| | | |
|------|------------|-----------------------|
| 95 % | Results | Messages |
| | SupplierID | Supplier |
| 1 | 55 | SAZERAC NORTH AMERICA |
| 2 | 297 | Laird & Company |
| 3 | 395 | PROXIMO |

Hình 38. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Supplier

4.4 Staging_Orders

```
1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [OrderID]
3      , [SupplierID]
4      , [RegionID]
5      , [ProductID]
6      , [Date]
7      , [Sale]
8      , [Quantity]
9  FROM [Staging_CK].[dbo].[Staging_Order]
```

95 %

Results Messages

| | OrderID | SupplierID | RegionID | ProductID | Date | Sale | Quantity |
|---|-----------------|------------|----------|-----------|-------------------------|--------|----------|
| 1 | INV-33237000023 | 395 | 97 | 87510 | 2021-09-22 00:00:00.000 | 23.19 | 1 |
| 2 | INV-33524200039 | 395 | 52 | 87510 | 2021-05-08 00:00:00.000 | 278.28 | 12 |
| 3 | INV-33462200108 | 395 | 77 | 88922 | 2021-08-20 00:00:00.000 | 89.28 | 2 |
| 4 | INV-33459600008 | 395 | 57 | 87510 | 2021-10-26 00:00:00.000 | 69.57 | 3 |
| 5 | INV-33468300052 | 395 | 17 | 87510 | 2021-09-27 00:00:00.000 | 69.57 | 3 |
| 6 | INV-33501300003 | 395 | 77 | 87510 | 2021-06-17 00:00:00.000 | 115.95 | 5 |
| 7 | INV-33467100035 | 395 | 17 | 87510 | 2021-06-29 00:00:00.000 | 46.38 | 2 |

Hình 39. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Order

4.5. Staging_Fact_Sales

```
1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [TimeID]
3      , [SupplierID]
4      , [RegionID]
5      , [ProductID]
6      , [Sale]
7      , [Quantity]
8      , [COG]
9      , [Profit]
10 FROM [Staging_CK].[dbo].[Staging_Fact_Sales]
```

95 %

Results Messages

| | TimeID | SupplierID | RegionID | ProductID | Sale | Quantity | COG | Profit |
|---|--------|------------|----------|-----------|------|----------|------|--------|
| 1 | 65 | 297 | 30 | 23823 | 3.94 | 2 | 1.31 | 2.63 |
| 2 | 227 | 297 | 30 | 23823 | 3.94 | 2 | 1.31 | 2.63 |
| 3 | 228 | 55 | 21 | 84404 | 4.29 | 3 | 0.95 | 3.34 |
| 4 | 49 | 55 | 17 | 84217 | 5.72 | 4 | 0.95 | 4.77 |
| 5 | 225 | 55 | 30 | 84217 | 5.72 | 4 | 0.95 | 4.77 |
| 6 | 206 | 297 | 77 | 23823 | 5.91 | 3 | 1.31 | 4.6 |

Hình 40. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Staging_Fact_Sales

4.6 Dim_Product

```

1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [TimeID]
3      , [SupplierID]
4      , [RegionID]
5      , [ProductID]
6      , [Sale]
7      , [Quantity]
8      , [COG]
9      , [Profit]
10 FROM [Staging_CK].[dbo].[Staging_Fact_Sales]

```

95 %

Results

Messages

| | TimeID | SupplierID | RegionID | ProductID | Sale | Quantity | COG | Profit |
|---|--------|------------|----------|-----------|------|----------|------|--------|
| 1 | 65 | 297 | 30 | 23823 | 3.94 | 2 | 1.31 | 2.63 |
| 2 | 227 | 297 | 30 | 23823 | 3.94 | 2 | 1.31 | 2.63 |
| 3 | 228 | 55 | 21 | 84404 | 4.29 | 3 | 0.95 | 3.34 |
| 4 | 49 | 55 | 17 | 84217 | 5.72 | 4 | 0.95 | 4.77 |
| 5 | 225 | 55 | 30 | 84217 | 5.72 | 4 | 0.95 | 4.77 |
| 6 | 206 | 297 | 77 | 23823 | 5.91 | 3 | 1.31 | 4.6 |

Hình 41. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Dim_Product

4.7 Dim_Region

| | | |
|---|---|--|
| 1 | /***** Script for SelectTopNRows command from SSMS *****/ | |
| 2 | SELECT TOP (1000) [RegionID] | |
| 3 | , [Region] | |
| 4 | FROM [Target_CK].[dbo].[Dim_Region] | |

| | | |
|------|----------|------------|
| 95 % | Results | Messages |
| 1 | RegionID | Region |
| 1 | 6 | BENTON |
| 2 | 7 | BLACK HAWK |
| 3 | 8 | BOONE |
| 4 | 9 | BREMER |
| 5 | 11 | BUENA VIST |

Hình 42. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Dim_Region

4.8 Dim_Supplier

| | | |
|---|---------------------------------------|--|
| 1 /***** Script for SelectTopNRows command from SSMS *****/ | | |
| 2 | SELECT TOP (1000) [SupplierID] | |
| 3 | , [Supplier] | |
| 4 | FROM [Target_CK].[dbo].[Dim_Supplier] | |

| | | |
|------|------------|-----------------------|
| 95 % | Results | Messages |
| 1 | SupplierID | Supplier |
| 2 | 55 | SAZERAC NORTH AMERICA |
| 3 | 297 | Laird & Company |
| 4 | 395 | PROXIMO |

Hình 43. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Dim_Supplier

4.9 Dim_Time

```
1 /***** Script for SelectTopNRows command from SSMS *****/
2 SELECT TOP (1000) [TimeID]
3     ,[Date]
4     ,[Day]
5     ,[Month]
6     ,[Year]
7 FROM [Target_CK].[dbo].[Dim_Time]
```

95 %

Results Messages

| | TimeID | Date | Day | Month | Year |
|---|--------|-------------------------|-----|-------|------|
| 1 | 1 | 2021-03-09 00:00:00.000 | 9 | 3 | 2021 |
| 2 | 2 | 2021-11-14 00:00:00.000 | 14 | 11 | 2021 |
| 3 | 3 | 2021-02-24 00:00:00.000 | 24 | 2 | 2021 |
| 4 | 4 | 2021-10-29 00:00:00.000 | 29 | 10 | 2021 |
| 5 | 5 | 2021-03-10 00:00:00.000 | 10 | 3 | 2021 |
| 6 | 6 | 2021-11-15 00:00:00.000 | 15 | 11 | 2021 |

Hình 44. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của Dim_Time

4.10 Fact_Sales

```

1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [TimeID]
3      , [SupplierID]
4      , [RegionID]
5      , [ProductID]
6      , [Sale]
7      , [Quantity]
8      , [COG]
9      , [Profit]
10 FROM [Target_CK].[dbo].[Fact_Sales]

```

95 %

Results Messages

| | TimeID | SupplierID | RegionID | ProductID | Sale | Quantity | COG | Profit |
|---|--------|------------|----------|-----------|------|----------|------|--------|
| 1 | 65 | 297 | 30 | 23823 | 3.94 | 2 | 1.31 | 2.63 |
| 2 | 227 | 297 | 30 | 23823 | 3.94 | 2 | 1.31 | 2.63 |
| 3 | 228 | 55 | 21 | 84404 | 4.29 | 3 | 0.95 | 3.34 |
| 4 | 49 | 55 | 17 | 84217 | 5.72 | 4 | 0.95 | 4.77 |
| 5 | 225 | 55 | 30 | 84217 | 5.72 | 4 | 0.95 | 4.77 |
| 6 | 206 | 297 | 77 | 23823 | 5.91 | 3 | 1.31 | 4.6 |
| 7 | 216 | 297 | 77 | 23823 | 5.91 | 3 | 1.31 | 4.6 |

Hình 45. Thực hiện truy vấn dữ liệu đã được đổ vào SQL của bảng Fact_Sales

4.11 Update Fact_Sales

```

1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [OrderID]
3      , [TimeID]
4      , [SupplierID]
5      , [RegionID]
6      , [ProductID]
7      , [Sale]
8      , [Quantity]
9  FROM [Target_CK].[dbo].[Update Fact_Sales]

```

95 %

Results Messages

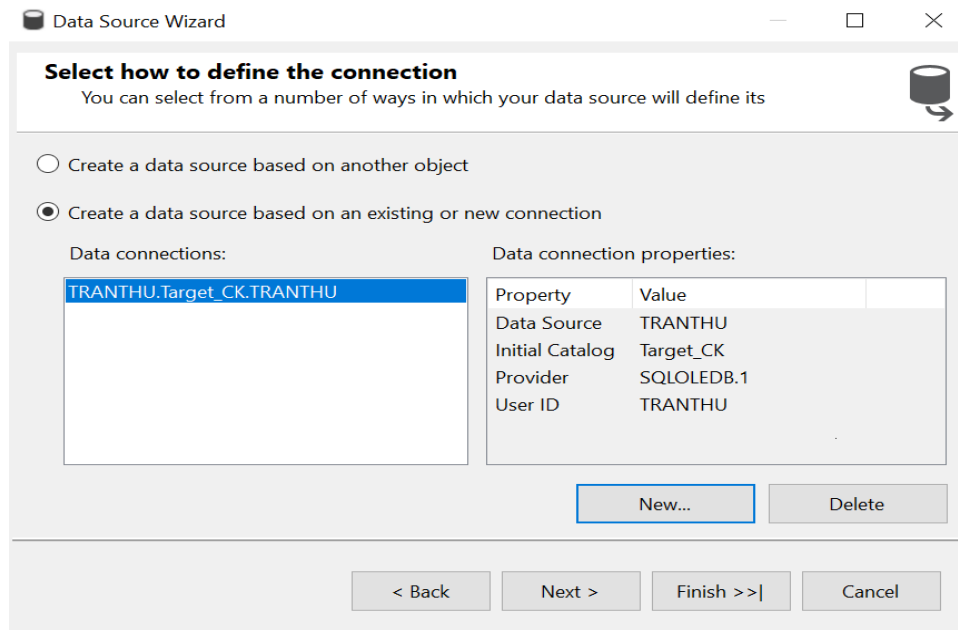
| | OrderID | TimeID | SupplierID | RegionID | ProductID | Sale | Quantity |
|---|-----------------|--------|------------|----------|-----------|------|----------|
| 1 | INV-33237000023 | 267 | 395 | 97 | 87510 | 23.2 | 2 |

Hình 46. Thực hiện update dữ liệu của bảng Fact_Sales

IV. OLAP

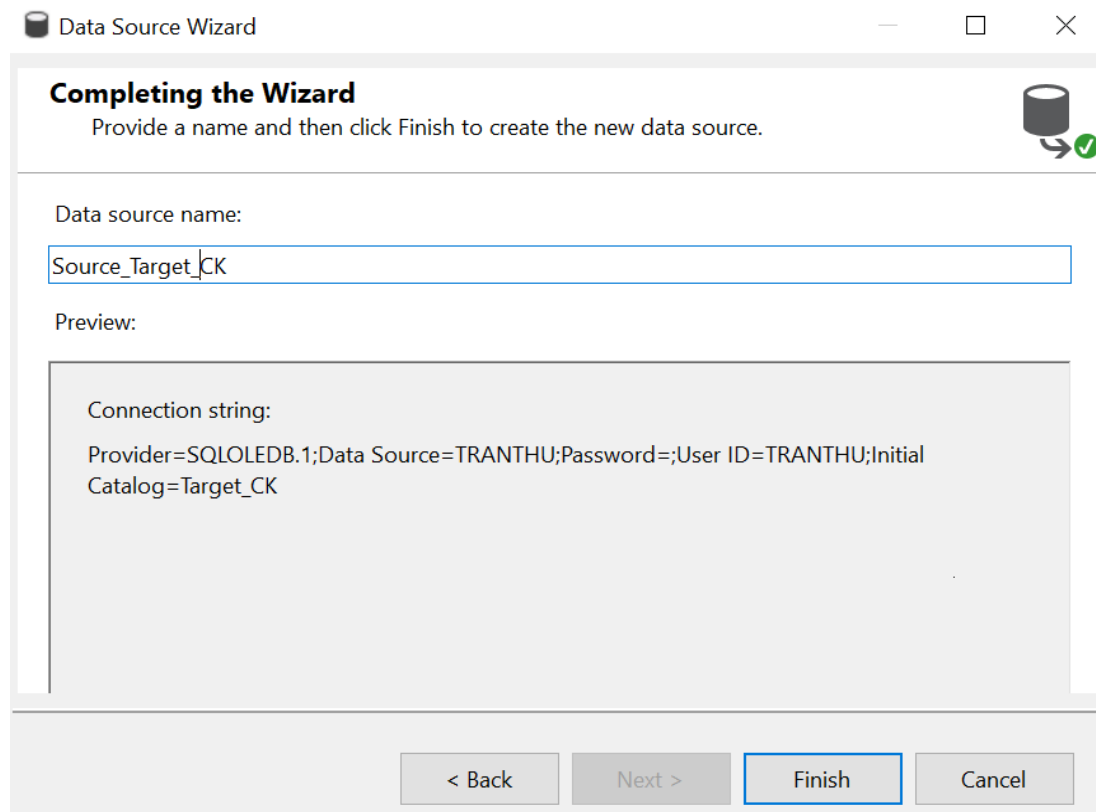
1. Data Source

- Kết nối với server và chọn database đã ETL ở trên.



Hình 47. Kết nối Server và chọn database

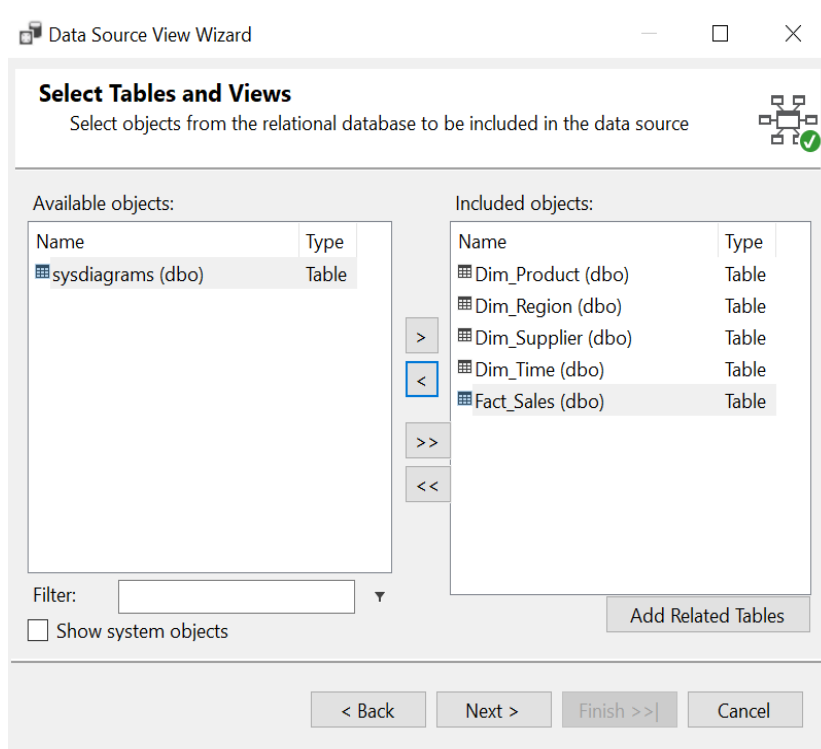
- Đặt tên cho source



Hình 48. Đặt tên cho source

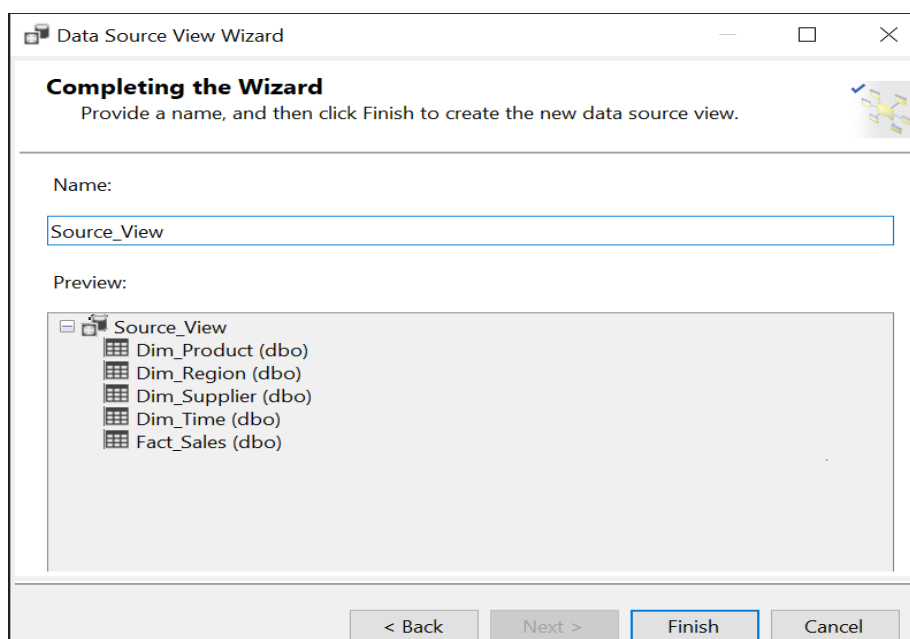
2. Data source view

- Chọn data sources
- Chọn 5 bảng Dim_Product, Dim_Region, Dim_Supplier, Dim_Time, Fact_Sales có trong database để tạo views.



Hình 49. Chọn bảng có trong database để tạo views

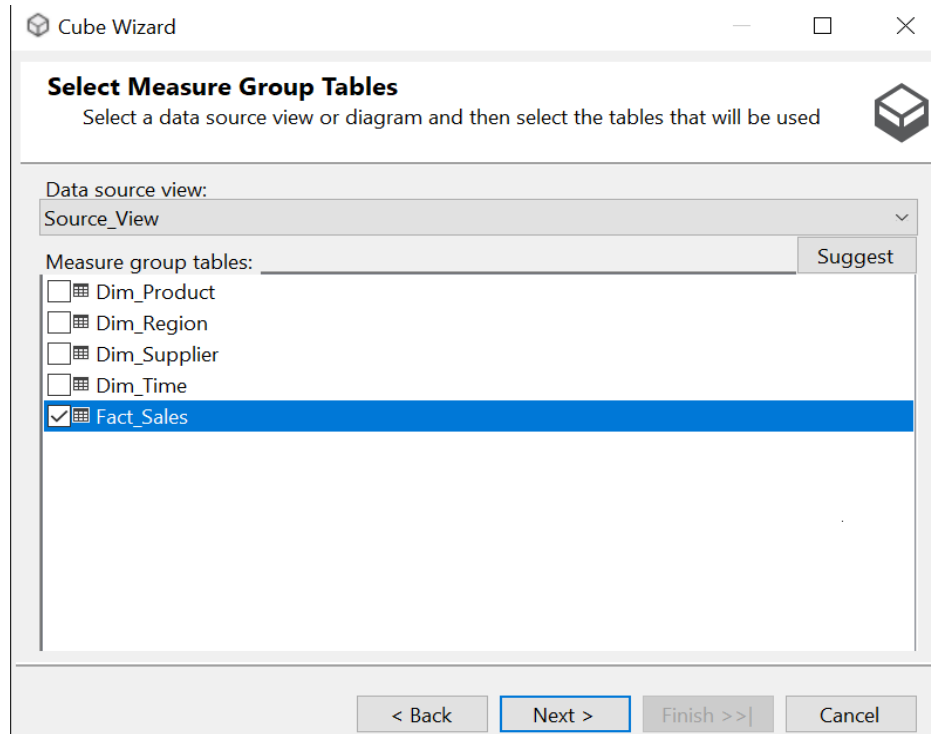
- Đặt tên chọn Data source View



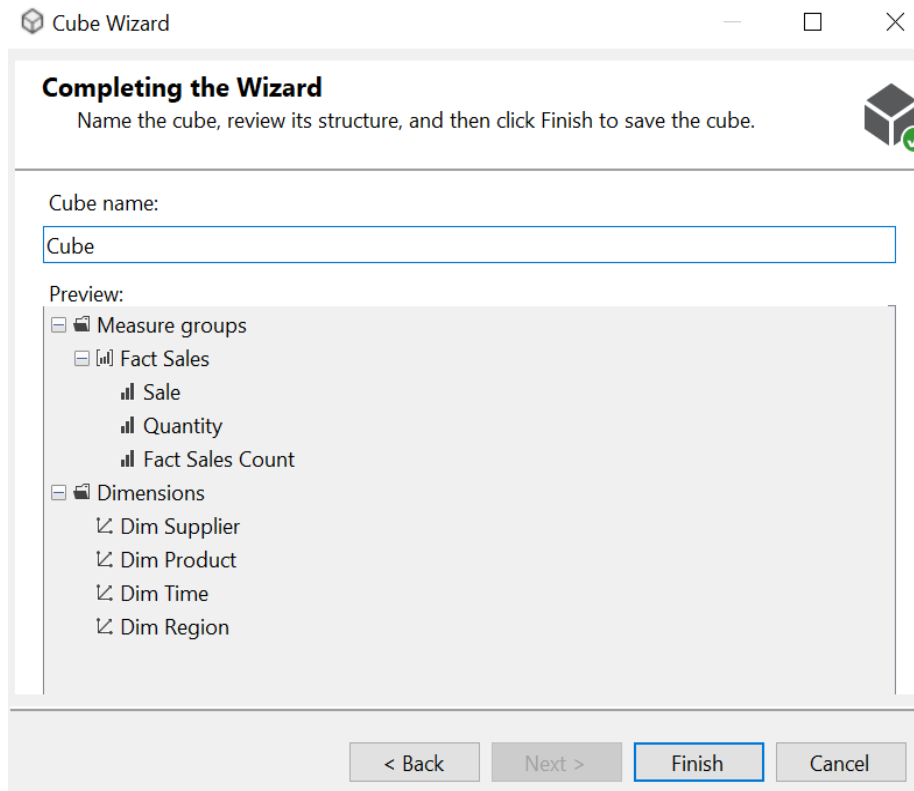
Hình 50. Đặt tên cho Data Source Views

3. Data Cube

- Chọn Measure trong bảng Fact_Sales và Dimension

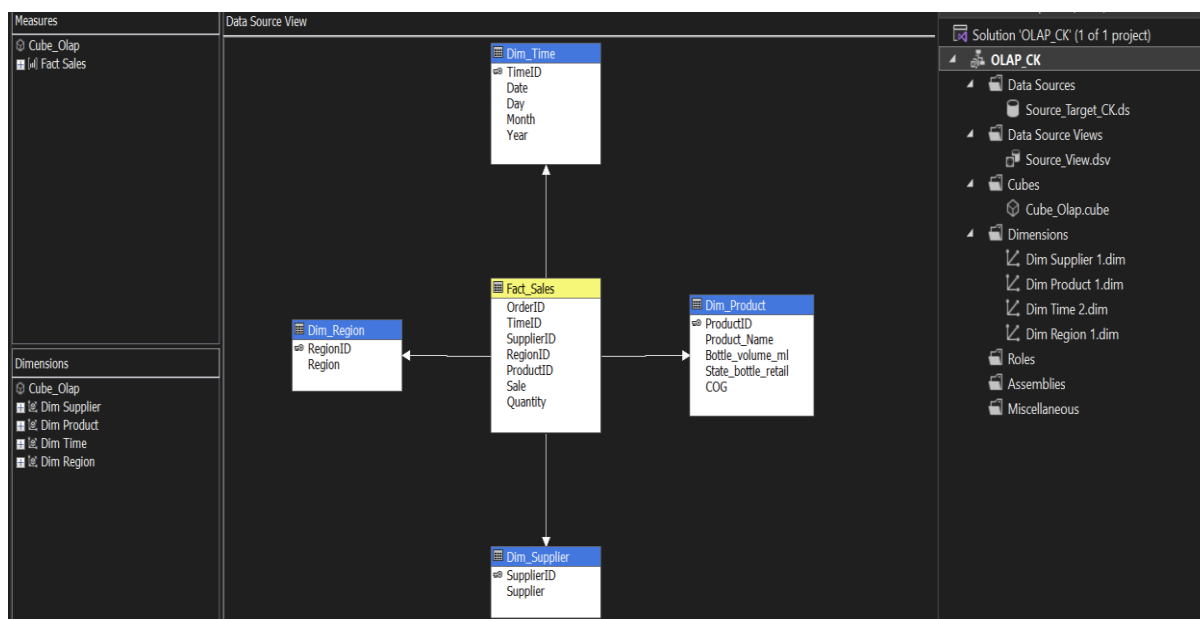


Hình 51. Chọn Measure



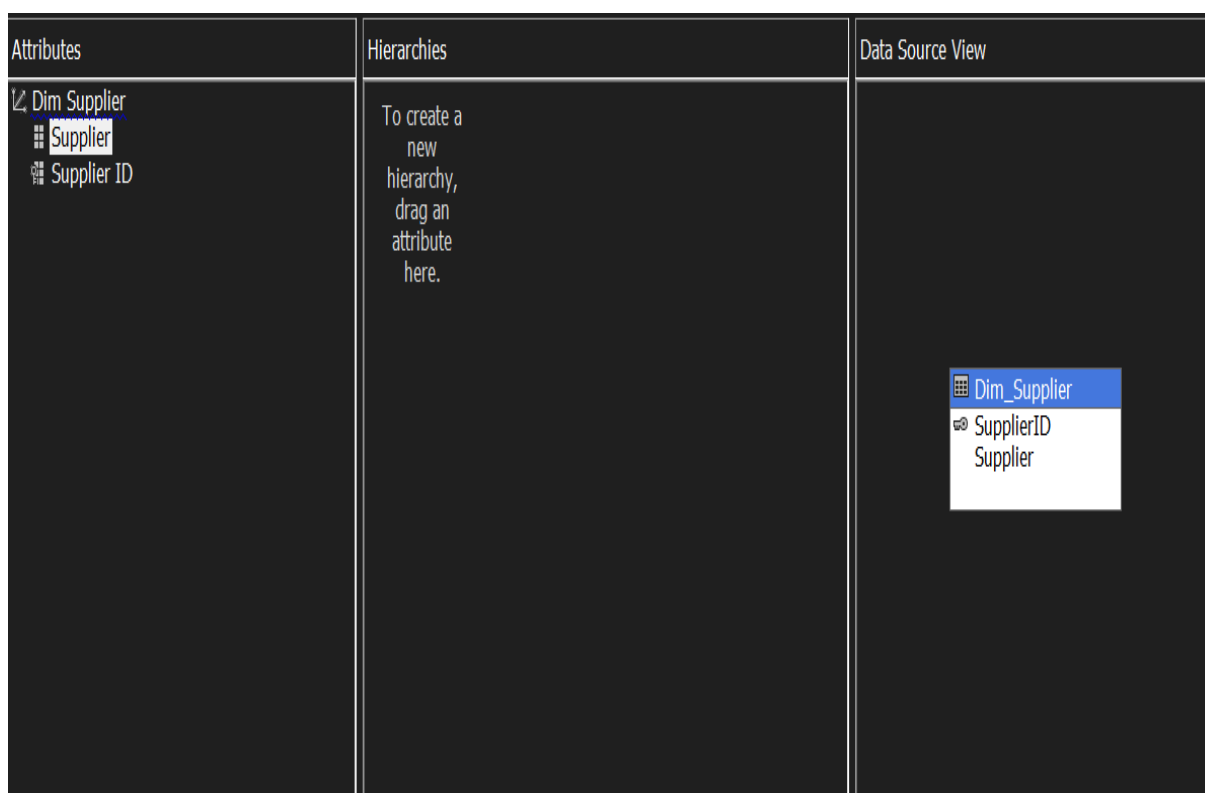
Hình 52. Đặt tên cho cube

- Data Cube sau khi tạo thành công

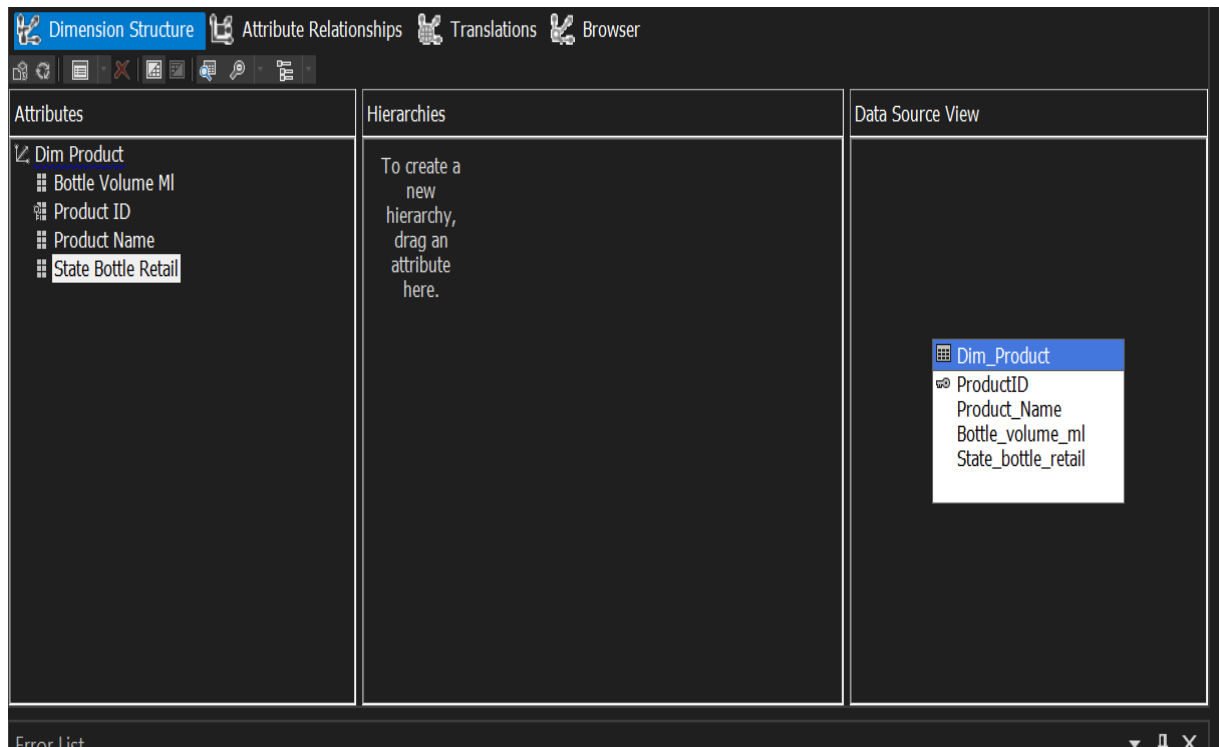


Hình 53. Data cube sau khi tạo thành công

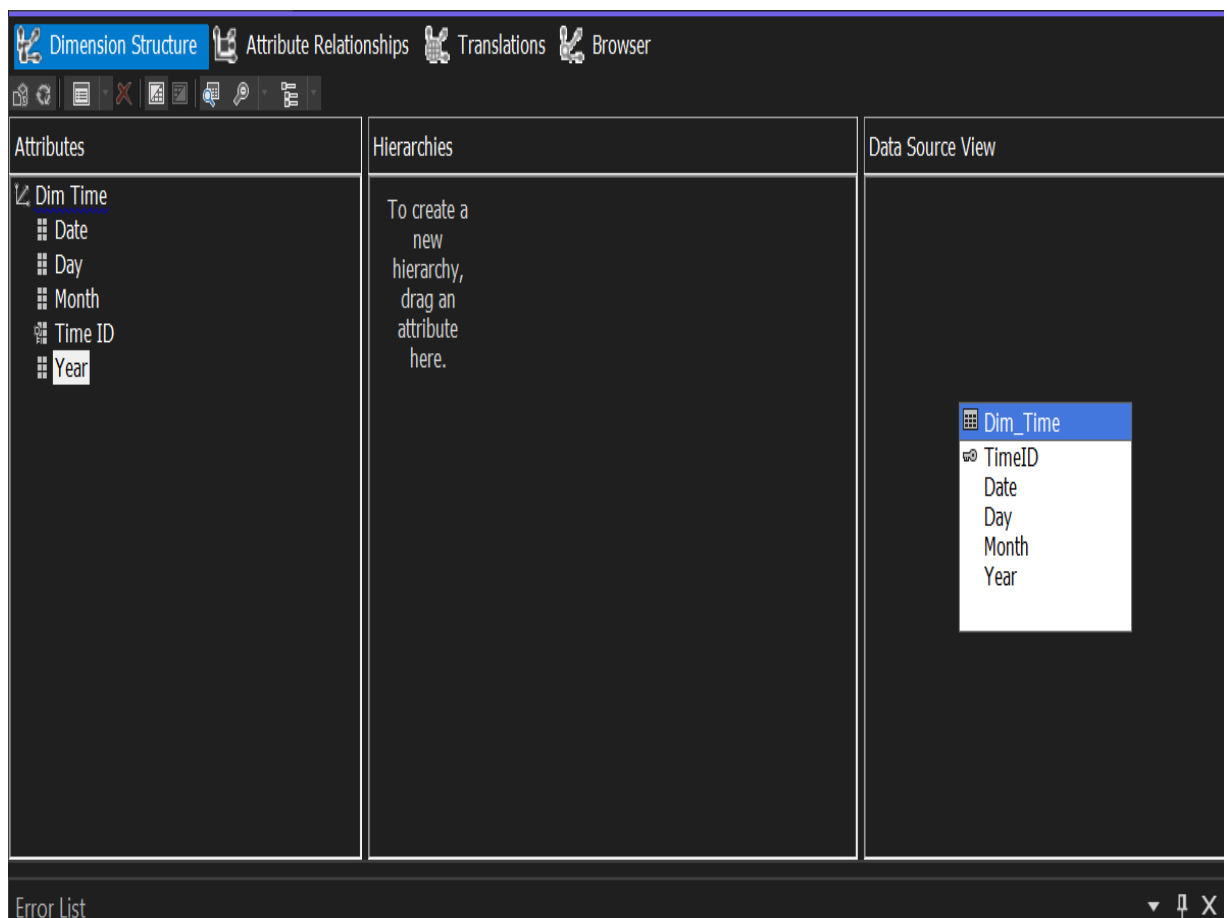
4. Dimensions



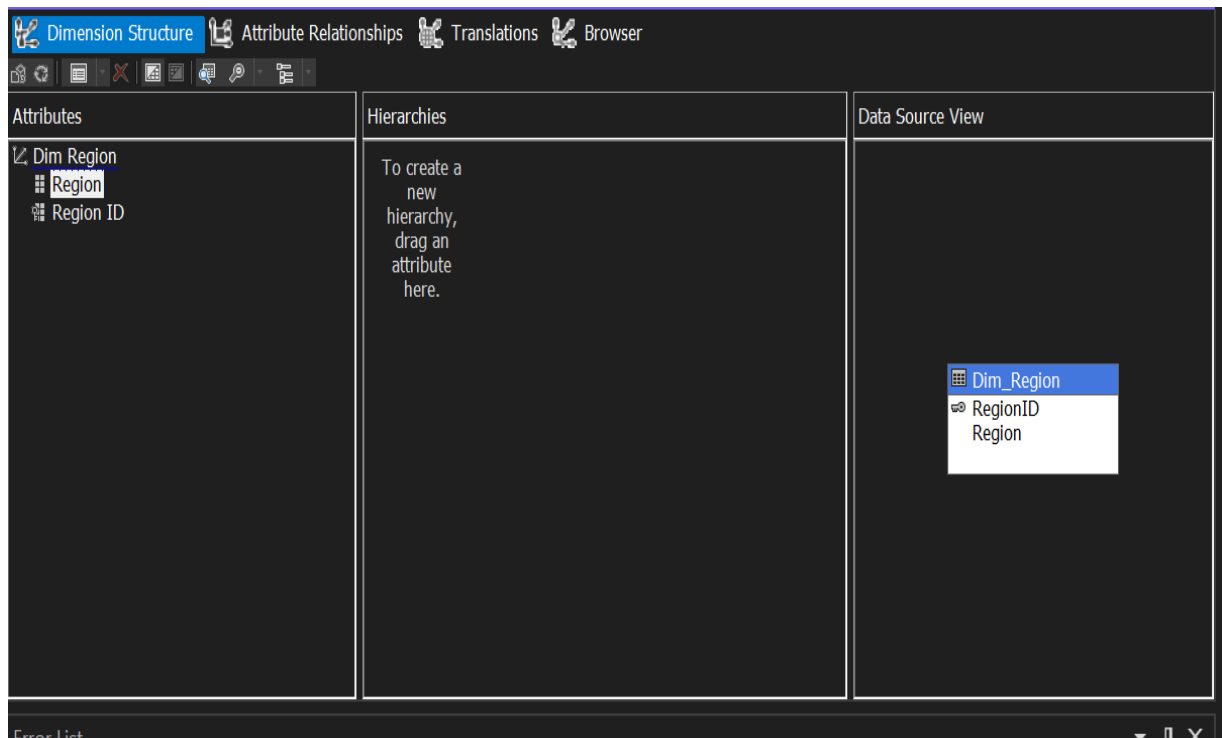
Hình 54. Dimensions Supplier



Hình 55. Dimensions Product

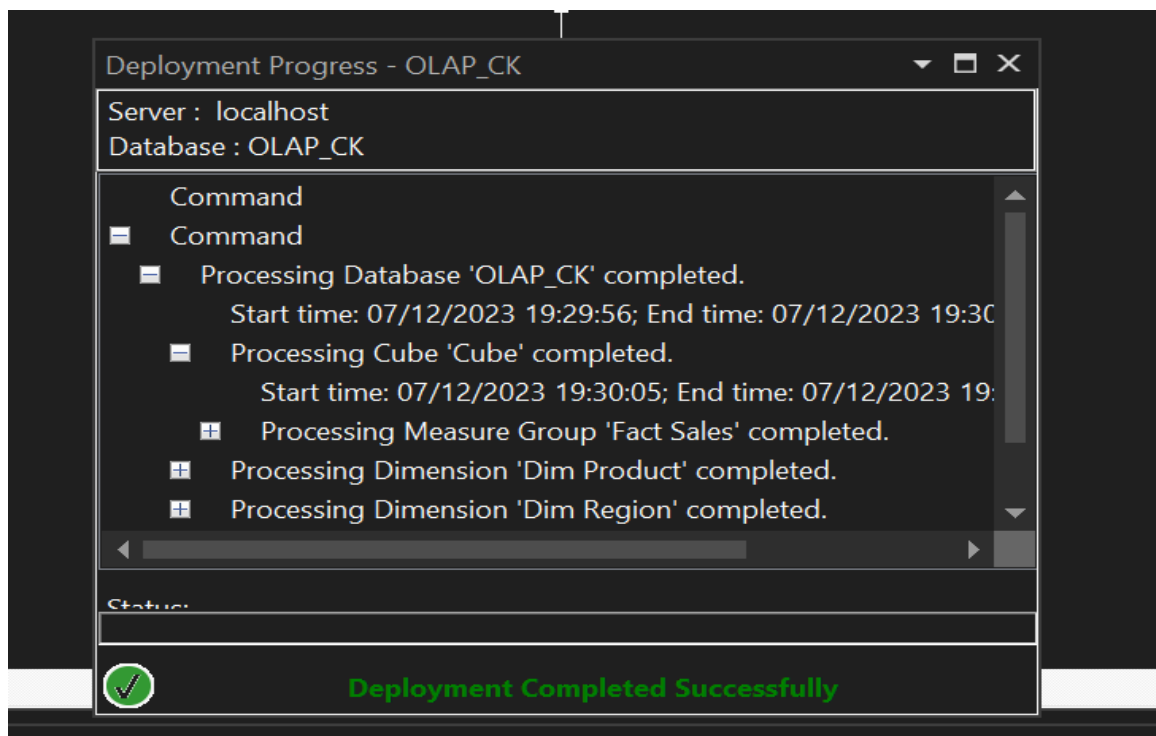


Hình 56. Dimensions Time



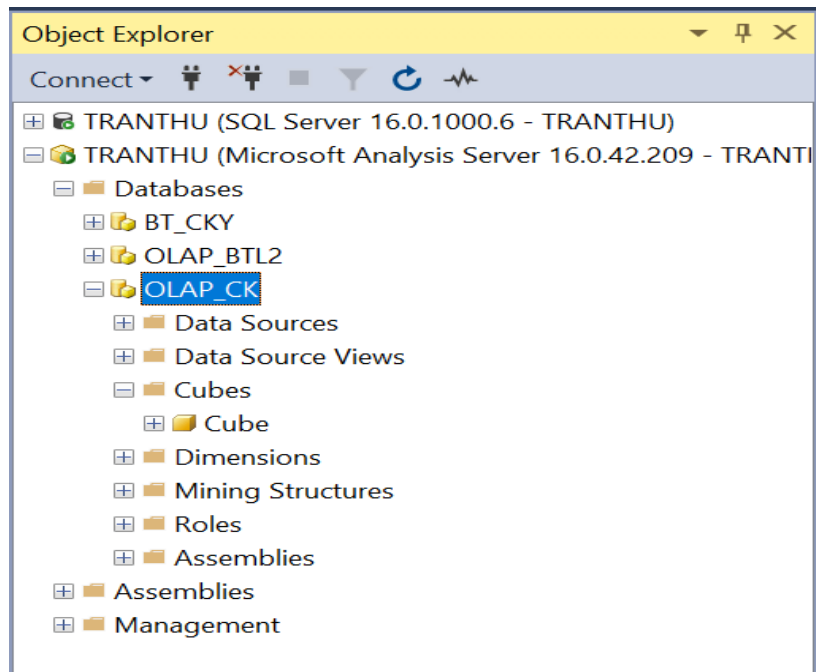
Hình 57. Dimensions Region

- Deploy thành công



Hình 58. Deploy thành công

5. Dữ liệu Olap trên SQL



Hình 59. Dữ liệu Olap đã được đưa về SQL

6. Phân tích OLAP

- Sử dụng Dim Time để phân tích doanh thu theo tháng.

The screenshot shows the SQL Server Data Tools (SSDT) interface. The 'Cube_Olap' cube is selected in the 'Metadata' pane. The 'Measures' list includes 'Fact Sales Count', 'Price', 'Quantity', and 'Sale'. The 'Dimensions' list includes 'Dim Product' and 'Dim Region'. The 'Calculated Members' pane is empty. The 'MDX' query editor shows the following query:

```

SELECT [Dim Time].[Month] ON COLUMNS, [Fact Sales].[Fact Sales Count] ON ROWS
FROM [Cube_Olap]

```

The query result is displayed in a table with the following data:

| Month | Sale |
|-------|-----------|
| 1 | 185465.62 |
| 10 | 96064.93 |
| 11 | 139939.89 |
| 12 | 412092.12 |
| 2 | 141634.19 |
| 3 | 5765.15 |
| 4 | 188709.46 |
| 5 | 278801.97 |
| 6 | 141295.45 |
| 7 | 141967.09 |
| 8 | 52651.75 |
| 9 | 5023.38 |

Hình 60. Phân tích doanh thu theo tháng

- Sử dụng Dim Product để phân tích các tổng số lượng sản phẩm bán được theo tổng doanh thu

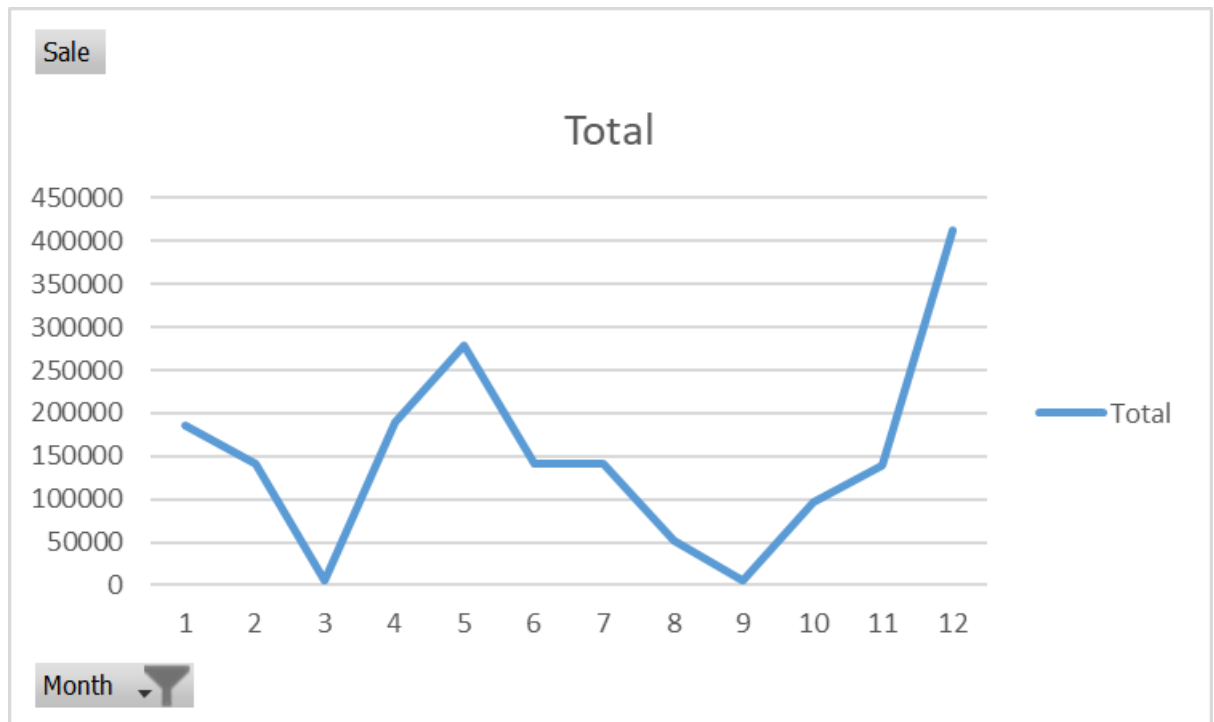
| | | | |
|----------------------------|-----------|-------------|-------------------|
| Language: Default | | | |
| Edit as Text Import... MDX | | | |
| Cube_Olap | | | |
| Metadata | | | |
| Search Model | | | |
| Measure Group: | | | |
| <All> | | | |
| Cube_Olap | | | |
| Measures | | | |
| Fact Sales | | | |
| Fact Sales Count | | | |
| Price | | | |
| Quantity | | | |
| Sale | | | |
| KPIs | | | |
| Dim Product | | | |
| Dim Region | | | |
| Calculated Members | | | |
| Dimension | Hierarchy | Operator | Filter Expression |
| <Select dimension> | | | |
| Product Name | Quantity | Price | Sale |
| 1800 Cristalino | 118 | 44.64 | 5267.52 |
| 1800 Reposa... | 115 | 26.25 | 3018.75 |
| 1800 Silver | 788 | 23.16058... | 18250.54 |
| 99 Apples | 225 | 2613.878... | 588122.75 |
| 99 Bananas | 399 | 1478.897... | 590080.25 |
| 99 Bananas ... | 1030 | 7.739999... | 7972.19999... |
| 99 Cinnamon | 1028 | 1.43 | 1470.04 |
| 99 Grapes | 2185 | 42.57905... | 93035.24 |
| 99 Mangoes | 2727 | 1.43 | 3899.61 |
| 99 Oranges | 256 | 7.74 | 1981.44 |
| Five Star | 2941 | 78.65531... | 231325.27 |
| Lauders | 1345 | 182.1467... | 244987.39 |

Hình 61. Phân tích sản phẩm theo lợi nhuận và doanh thu

7. Pivot Table

- Biểu đồ thể hiện doanh thu và lợi nhuận của một số sản phẩm trong tháng 12

| Months | Sale |
|--------------------|----------------|
| 1 | 185465.62 |
| 2 | 141634.19 |
| 3 | 5765.15 |
| 4 | 188709.46 |
| 5 | 278801.97 |
| 6 | 141295.45 |
| 7 | 141967.09 |
| 8 | 52651.75 |
| 9 | 5023.38 |
| 10 | 96064.93 |
| 11 | 139939.89 |
| 12 | 412092.12 |
| Grand Total | 1789411 |

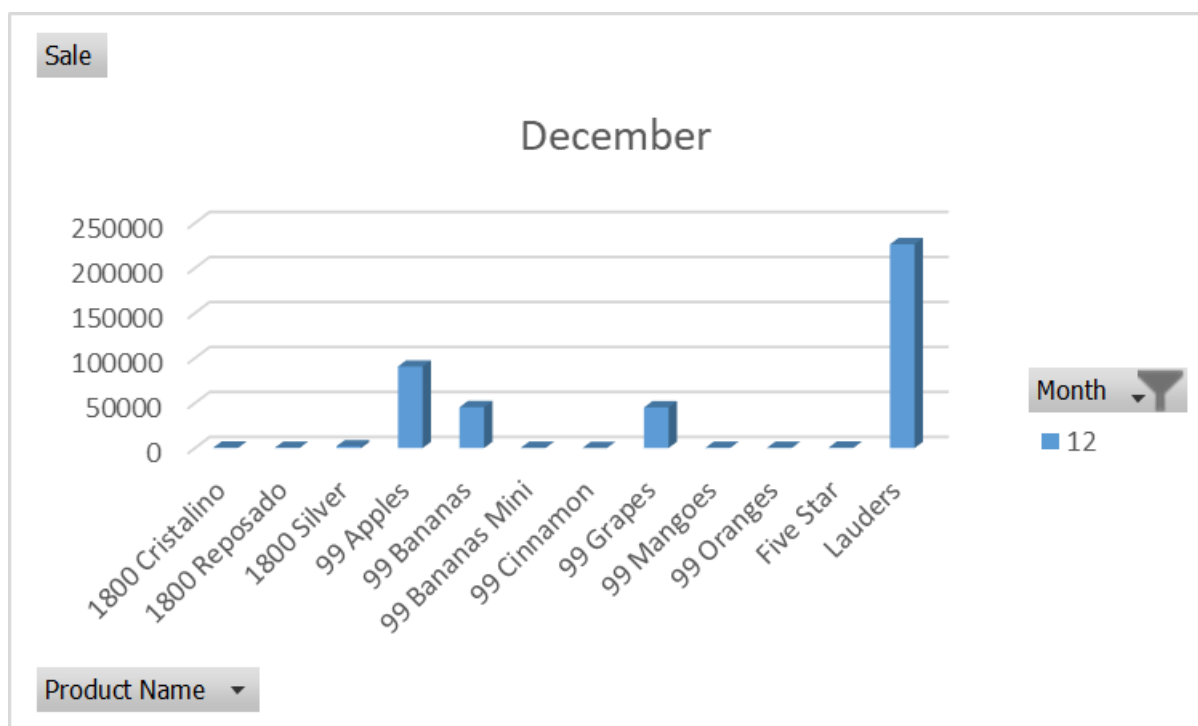


Hình 62. Biểu đồ đường thể hiện tổng doanh thu của các tháng trong năm 2021

Nhận xét: Nhìn chung doanh thu từ tháng 1 đến tháng 12 có sự biến động không đều. Đã có những tháng doanh thu giảm sâu và sau đó cũng đã có sự tăng vọt đột ngột vào tháng 12 là một dấu hiệu tích cực. Vì vậy, cần xem xét chiến lược kinh doanh và tiếp thị đã áp dụng trong từng khoảng thời gian để đạt được doanh thu tốt nhất.

- Biểu đồ thể hiện doanh thu và lợi nhuận của một số sản phẩm trong tháng 12

| Sale | Month | |
|-----------------|-----------|-------------|
| Product | 12 | Grand Total |
| 1800 Cristalino | 535.68 | 535.68 |
| 1800 Reposado | 393.75 | 393.75 |
| 1800 Silver | 1947.97 | 1947.97 |
| 99 Apples | 90496 | 90496 |
| 99 Bananas | 45428 | 45428 |
| 99 Bananas Mini | 363.78 | 363.78 |
| 99 Cinnamon | 68.64 | 68.64 |
| 99 Grapes | 45217.71 | 45217.71 |
| 99 Mangoes | 386.1 | 386.1 |
| 99 Oranges | 270.9 | 270.9 |
| Five Star | 476.74 | 476.74 |
| Lauders | 226506.85 | 226506.85 |
| Grand Total | 412092.12 | 412092.12 |



Hình 63. Biểu đồ cột thể hiện doanh thu của 12 sản phẩm trong tháng 12

Nhận xét: Vào tháng 12, có 4 sản phẩm trên tổng 12 sản phẩm bán chạy nhất trong tháng. Vì vậy cần tìm hiểu xem xét và đẩy mạnh các sản phẩm để phục vụ sở thích của khách hàng và tìm hiểu nguyên nhân các sản phẩm còn lại không được ưa chuộng để có những giải pháp tốt nhất.

- + Sản phẩm Lauder, 99 Apples, 99 Bananas có hiệu suất tốt khi cả doanh thu và lợi nhuận đều, nên duy trì.

- + Các sản phẩm có doanh số bán hàng thấp hơn so với các sản phẩm khác, nên cần xem xét để tối ưu hóa hiệu suất.

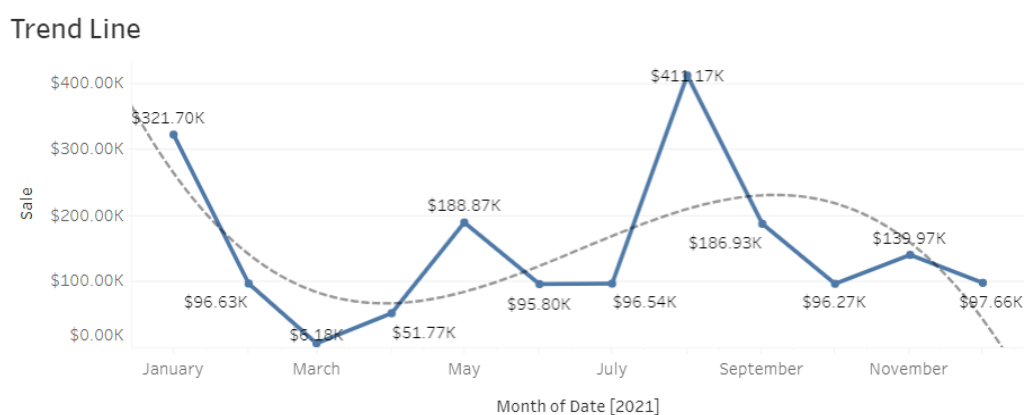
V. Dashboard

1. Phân tích biểu đồ

| Grand Total | | | |
|-----------------|-------------|--------------|-------------|
| Count Of Prod.. | Profit | Profit Ratio | Sale |
| 12 | \$1,781.64K | 99.56% | \$1,789.48K |

Hình 64. Biểu đồ Grand Total

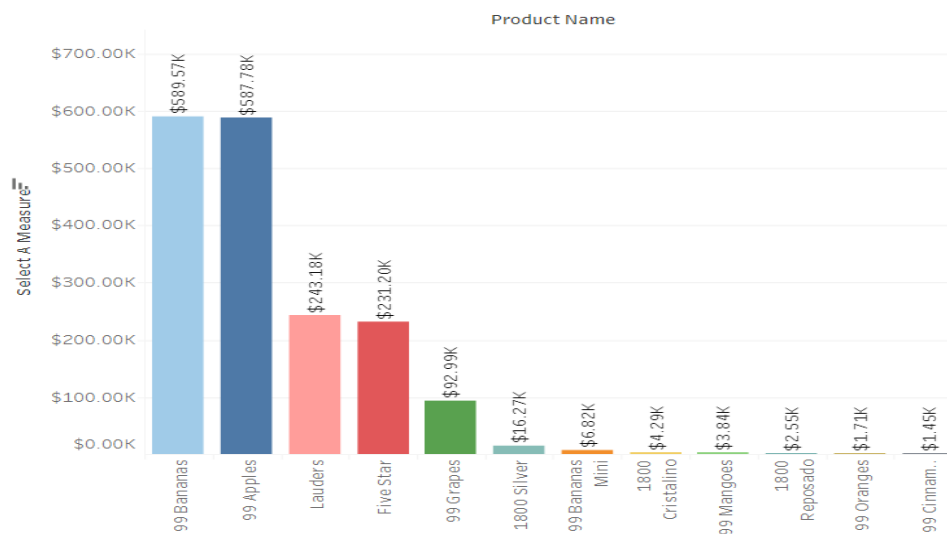
- Biểu đồ Grand Total bao gồm có 4 cột, hiển thị tổng số sản phẩm của khu vực IOWA là 12 sản phẩm và tỉ lệ phần trăm lợi nhuận là 99.56%



Hình 65. Biểu đồ Trend Line

- Trong năm doanh thu có nhiều biến động, tháng 1,5,8 đạt được doanh thu cao đặc biệt là tháng 8 đạt được doanh thu cao nhất là 411.17K và tháng 3 đạt được doanh thu thấp nhất là 6.18K.
- Biểu đồ có đường đứt thể hiện dự đoán xu hướng tăng giảm doanh thu của các tháng trong năm 2021 và đường màu xanh là doanh thu thực tế của các tháng trong năm 2021 so với đường xu hướng. Nhờ đó để biết được sự chênh lệch của đường thực tế so với đường xu hướng nó có những tác động tích cực hay là tiêu cực đối với doanh thu của khu vực IOWA.

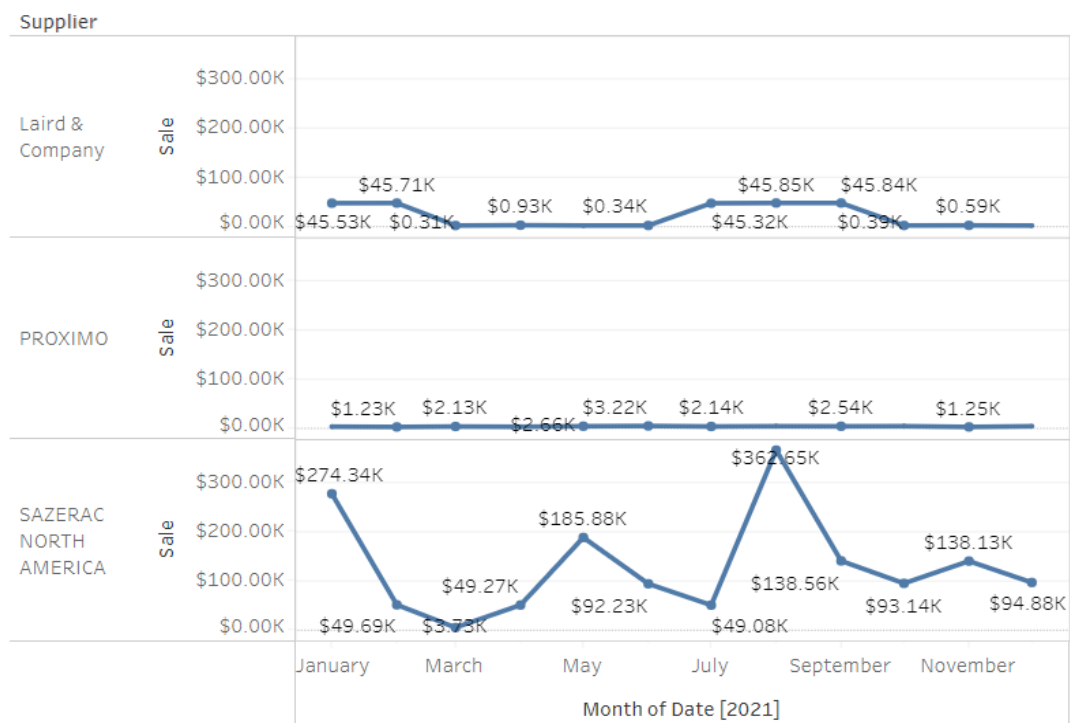
Measurement By Product



Hình 66. Biểu đồ cột Measurement by Product

Biểu đồ Measurement by product thể hiện các độ đo như quality, profit, sale của 12 sản phẩm tại khu vực IOWA.

Sale By Supplier and Date



Hình 67. Biểu đồ đường Sale by Supplier and Date

Dựa vào biểu đồ Doanh thu của các nhà cung cấp theo từng tháng trong năm 2021 cho thấy sự chênh lệch rõ rệt giữa nhà cung cấp SAZERAC NORTH AMERICA và 2 nhà cung cấp Laird & Company, nhà cung cấp PROXIMO. Xét theo từng biểu đồ ta được:

- + Laird & Company: Cho thấy sự biến động giữa các tháng không quá chênh lệch nhiều, nhưng doanh thu của nhà cung cấp này rất thấp <50k. Ngoài ra còn có sự chênh lệch lớn với nhà cung cấp SAZERAC NORTH AMERICA
- + PROXIMO: Về doanh thu của nhà cung cấp này có sự tăng và giảm nhẹ qua các tháng nhưng không đáng kể (đường thể hiện doanh thu dường như là một đường thẳng), vì doanh thu của nhà cung cấp này vẫn quá thấp và đáng báo động.
- + SAZERAC NORTH AMERICA: Về doanh thu của nhà cung cấp này có xu hướng biến động nhiều, có những tháng đạt doanh thu cao >300k và cũng có những tháng doanh thu giảm <50k, so với 2 nhà cung cấp kia thì nhà cung cấp này có doanh thu ổn định và cao hơn hẳn, nhưng vẫn cần phải có những biện pháp để giúp cân bằng doanh thu giữa các tháng.

Word Cloud By Region of Sales



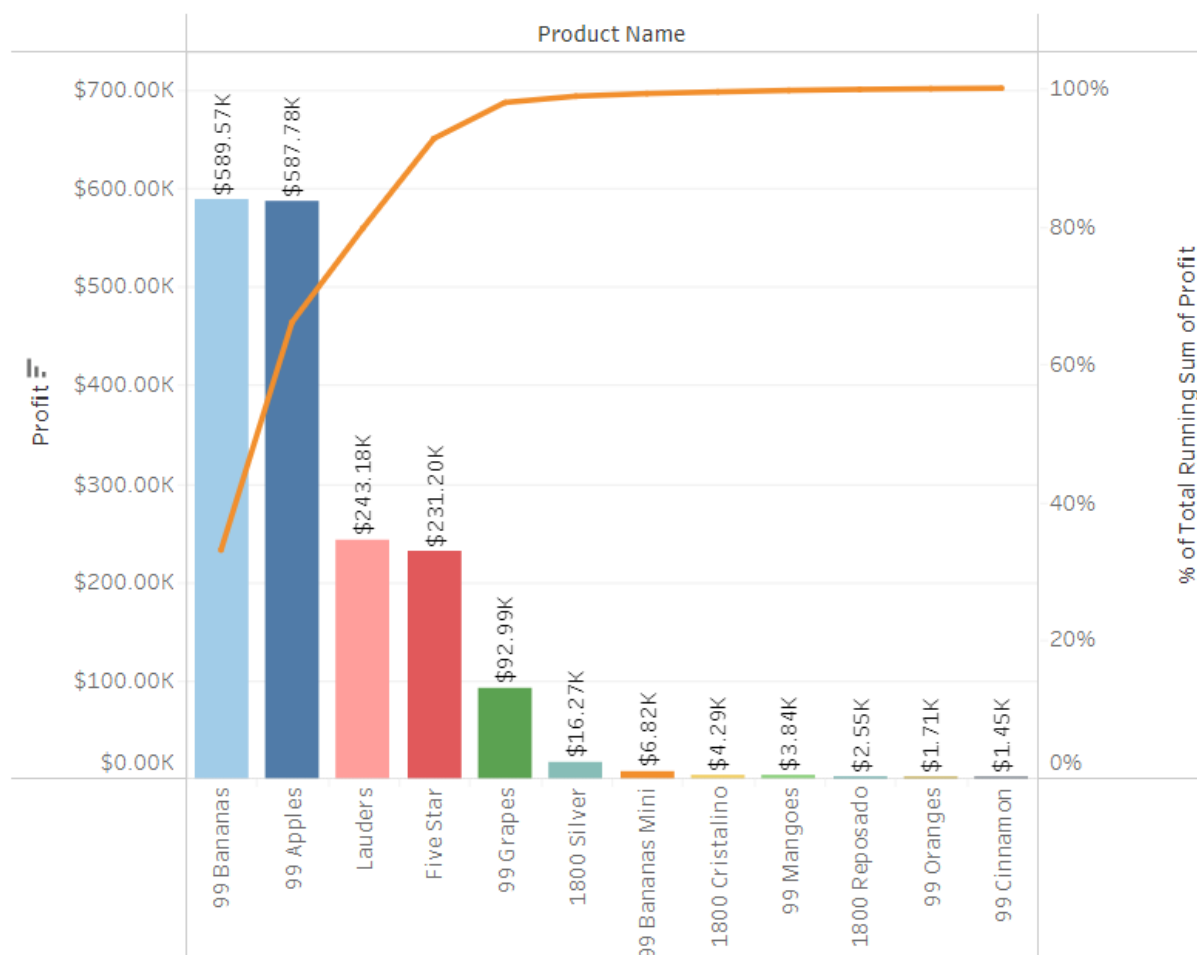
Hình 68. Biểu đồ cloud by Region of Sale

- Biểu đồ trên so sánh cho ta thấy doanh thu của các tiểu bang tại khu vực IOWA, đối với những tiểu bang mà có kích thước tên càng lớn thì doanh

thu đem lại cho khu vực càng cao và cũng ngược lại, những tiểu bang có kích thước tên càng nhỏ thì đem lại càng ít doanh thu.

- Kích thước của các tiểu bang trên biểu đồ dựa trên số đơn đặt hàng rượu tại các tiểu bang.

Pareto Chart



Hình 69. Biểu đồ tích lũy

- Biểu đồ tích lũy thể hiện cho thấy sự tích lũy về lợi nhuận của 12 sản phẩm. Các sản phẩm theo cột dọc là tích lũy riêng của từng sản phẩm và đường là tích lũy chung của tất cả các sản phẩm. Khi nhìn vào biểu đồ ta sẽ thấy có hai sản phẩm mang lại số tích lũy cao nhất là 99 Bananas và 99 Apples.

2. Kết luận

- Nhìn chung doanh thu có nhiều biến động lớn qua các tháng trong năm 2021, từ tăng vọt cho đến giảm sâu, tăng mạnh vào các tháng 5,12 và giảm mạnh vào các tháng 3, 9.
- Có sự chênh lệch lớn về doanh thu giữa các sản phẩm, hai sản phẩm là 99 Bananas và 99 Apples có tổng doanh thu và cao nhất và chênh lệch rất lớn so với doanh thu của các sản phẩm còn lại, 2 sản phẩm 99 Bananas và 99 Apples chiếm tới 66,08 % lợi nhuận của khu vực IOWA.

3. Giải pháp

- Về biến động lớn doanh thu theo các tháng IOWA nên mở rộng hoặc phát triển thêm sản phẩm hoặc dịch vụ mới để giảm thiểu tác động của sự biến động trong một khu vực cụ thể.
- Còn các tháng có xu hướng giảm sâu nên cần có những chiến lược kinh doanh để gia tăng doanh số cũng như doanh thu, ví dụ như việc đưa ra các ưu đãi, khuyến mãi để gia tăng nhu cầu mua hàng của khách hàng
- Cần xây dựng chiến lược cụ thể, tìm hiểu lại các khu vực có doanh thu thấp, tìm ra lý do khắc phục xây dựng lại chiến lược marketing với những vùng có doanh thu thấp. Đẩy mạnh hơn và khai thác tối đa lợi thế ở những khu vực mang lại doanh thu cao