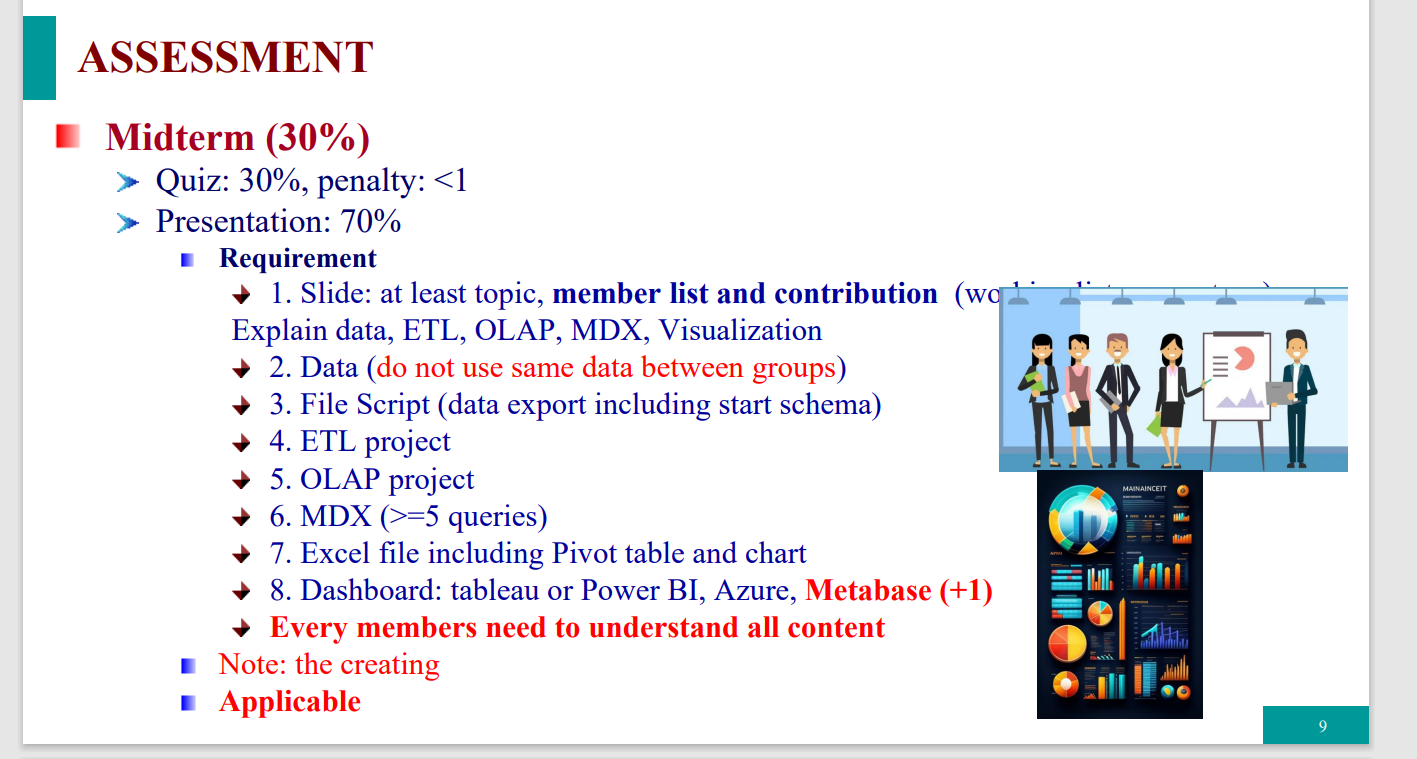
Yêu cầu bài tập



3.File Script

4.ETL project

* ETL CSV: Toàn,Tuyết 22-23/2
* ETL to schema: Hằng,Huệ 24/2

5. OLAP project: Hằng,Huy 25/2

6. Excel file including pivot table and chart: Tuyết, 27/2

7. MDX: Huy,Huệ 26/2

8. Dashboard: tableau or Power BI, Azure, Metabase(+1): Toàn 27/2

27/2 hoàn thành xong slide

28/2 Thuyết trình thử

—-Tên thành viên:

1. Trần Thị Huệ
2. Trần Gia Huy
3. Nguyễn Kim Toàn
4. Nguyễn Thị Thanh Hằng
5. Nguyễn Thị Tuyết

—--Yêu cầu slide: [Báo cáo giữa kì - Bài thuyết trình (canva.com)](https://www.canva.com/design/DAFyXwJ-Lps/_tEdLKkYJP7k8MASQeV7HA/edit?fbclid=IwAR2WKOfPQ_U7-CDqpv3kisRCATYBvA9JSmn8M7FzsGt7xlintF0nmrvQD6U)

1. Nội dung:

* Tên dự án
* Tên thành viên
* Công việc và phần trăm đóng góp
* Giải thích dữ liệu chọn (các thuộc tính): [Travel Insurance (kaggle.com)](https://www.kaggle.com/datasets/mhdzahier/travel-insurance?fbclid=IwAR2xcHRH2qNVq1xCDqetrjR-94fK2sIcCuk06WvQFcfeeUfJG783XBdWNyk)
* Nội dung được phân công từ mục 4,5,6,7,8

1. Màu sắc, animation: đẹp, hợp là được

## 

## **Explain data**

Tập dữ liệu của một công ty dịch vụ bảo hiểm du lịch bên thứ ba có trụ sở tại Singapore.

Các thuộc tính:

* **Agency**: Công ty hoặc hãng hàng không cung cấp sản phẩm bảo hiểm.
* **Agency Type**: Loại hình của công ty hoặc hãng hàng không (ví dụ: Travel Agency - Công ty du lịch, Airlines - Hãng hàng không).
* **Distribution Channel**: Kênh phân phối của sản phẩm bảo hiểm (ví dụ: Online - Trực tuyến, Offline - Ngoại tuyến).
* **Product Name: Tên của sản phẩm bảo hiểm.**
* **Claim**: Cho biết liệu có yêu cầu bồi thường không.

*Người mua bảo hiểm đã gửi yêu cầu bồi thường hay không. Nếu giá trị của mục này là "Yes", điều đó có nghĩa là người mua đã gửi yêu cầu bồi thường cho một sự cố nào đó mà họ đã gặp phải và mong muốn nhận được tiền bồi thường từ công ty bảo hiểm. Ngược lại, nếu giá trị là "No", có nghĩa là người mua không yêu cầu bồi thường và chưa gặp phải sự cố nào cần được bảo hiểm chi trả.*

* **Duration**: Thời gian du lịch (trong số ngày).
* **Destination**: Điểm đến của chuyến đi.
* **Net Sales**: Doanh số bán hàng sau khi loại bỏ chiết khấu hoặc chi phí khác.

Doanh thu (cty nhận) - Hoa hồng (chia cho Đại lý) - CP khác

* **Commission** (in value): Giá trị hoa hồng hoặc phí môi giới.
* **Gender**: Giới tính của người mua.
* **Age**: Tuổi của người mua

Dimensions:

* Agency
* Destination
* Product

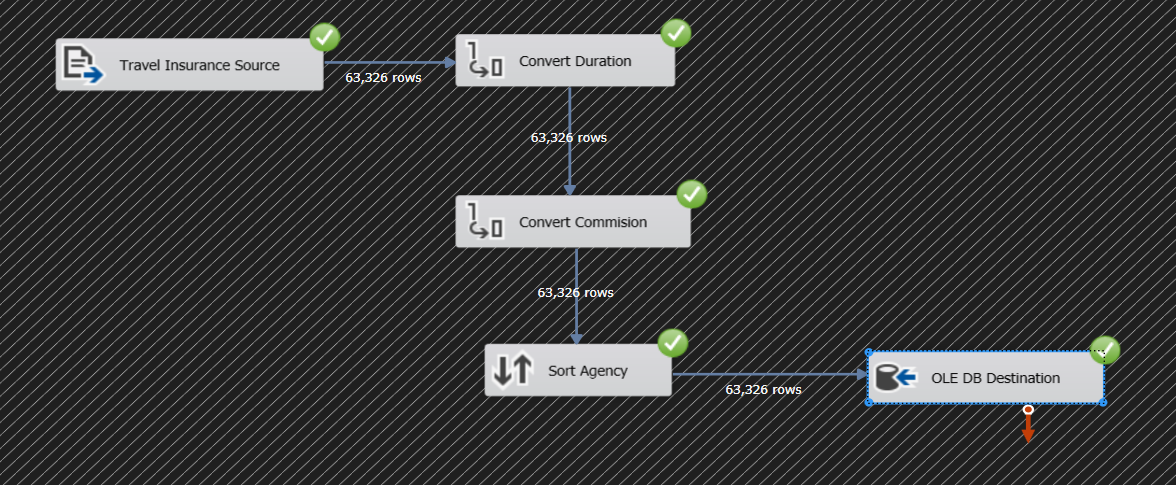
Measures:

* Commission
* Duration

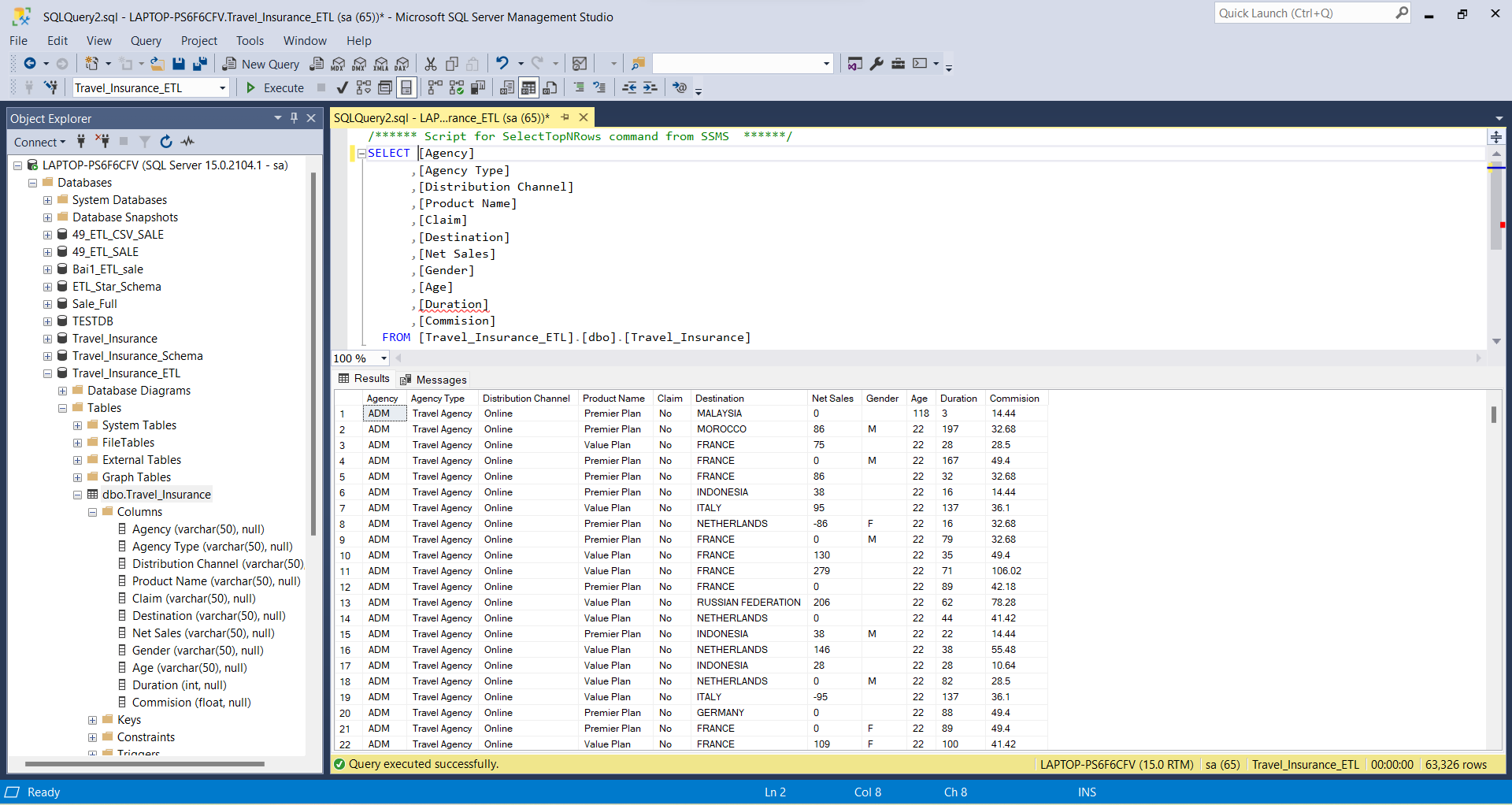
## **ETL project**

Ảnh ETL:

* Travel Insurance Source: thuộc tính(Flat file Source) import dữ liệu từ csv
* Convert (Duration/commision): thuộc tính(data convert) chuyển đổi dữ liệu
* sort: sắp xếp theo cột Agency
* OLE DB Destination: xuất dữ liệu sang SQL

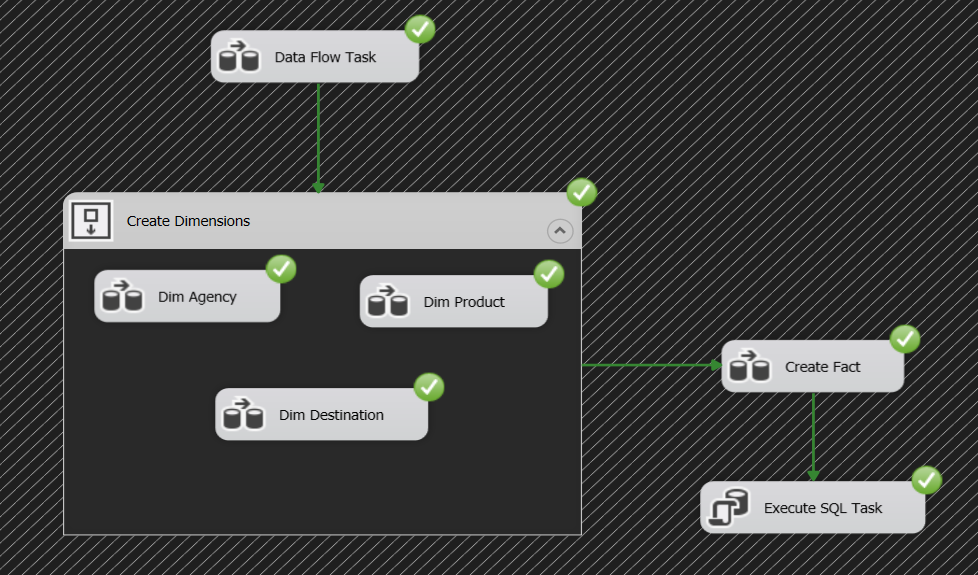


Ảnh chạy SQL:



# **ETL to Schema**

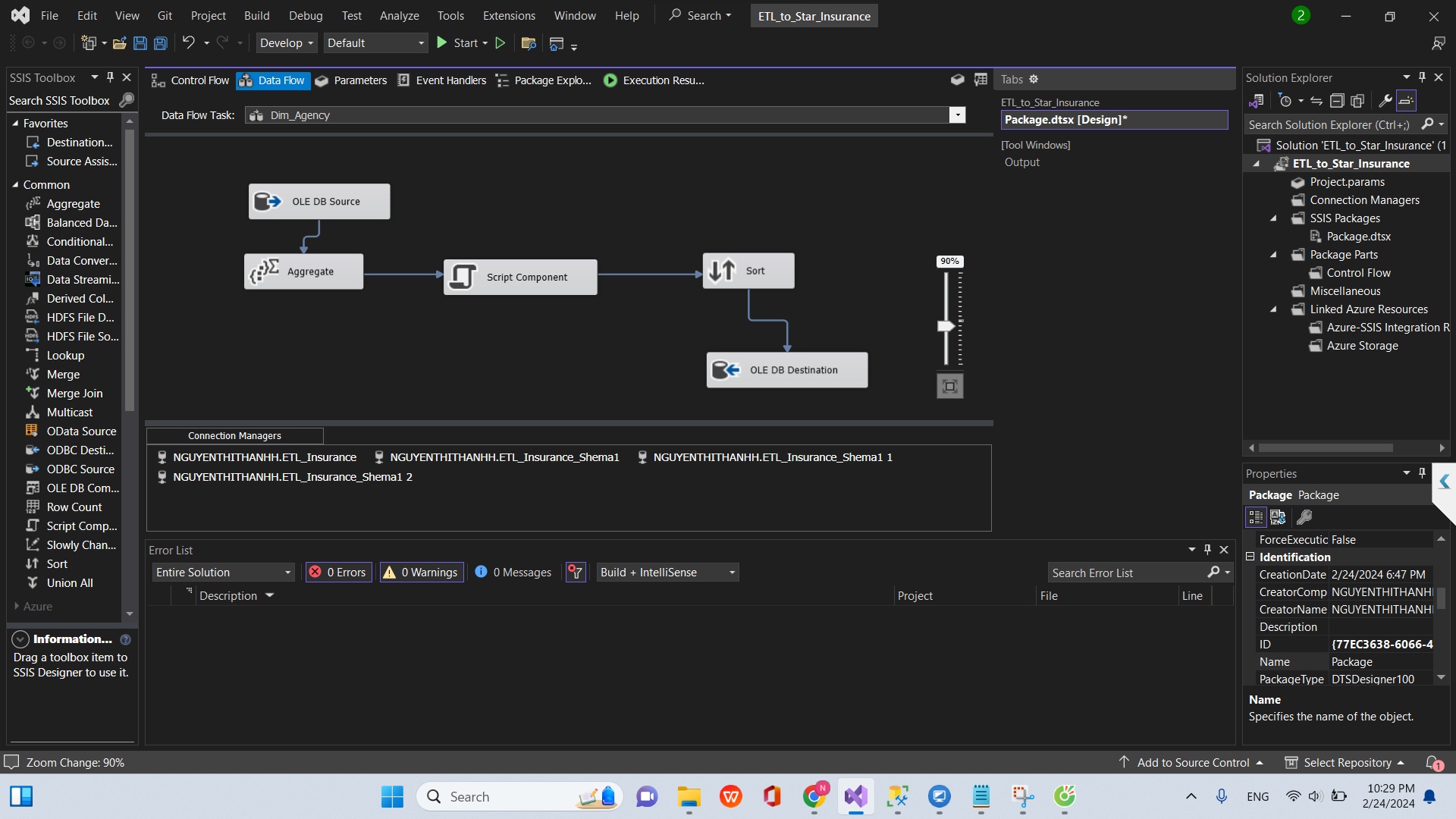
* Ảnh chạy thành công
* Data flow task: lấy dữ liệu sql từ kết quả của ETL CSV
* Create Dimensions: tạo các dimension cho ETL Schema
  + Dim Agency
  + Dim Product
  + Dim Destination
* Create Fact: bảng fact chứa khóa ngoại và measures
* Execute SQL Task: import dữ liệu sang sql

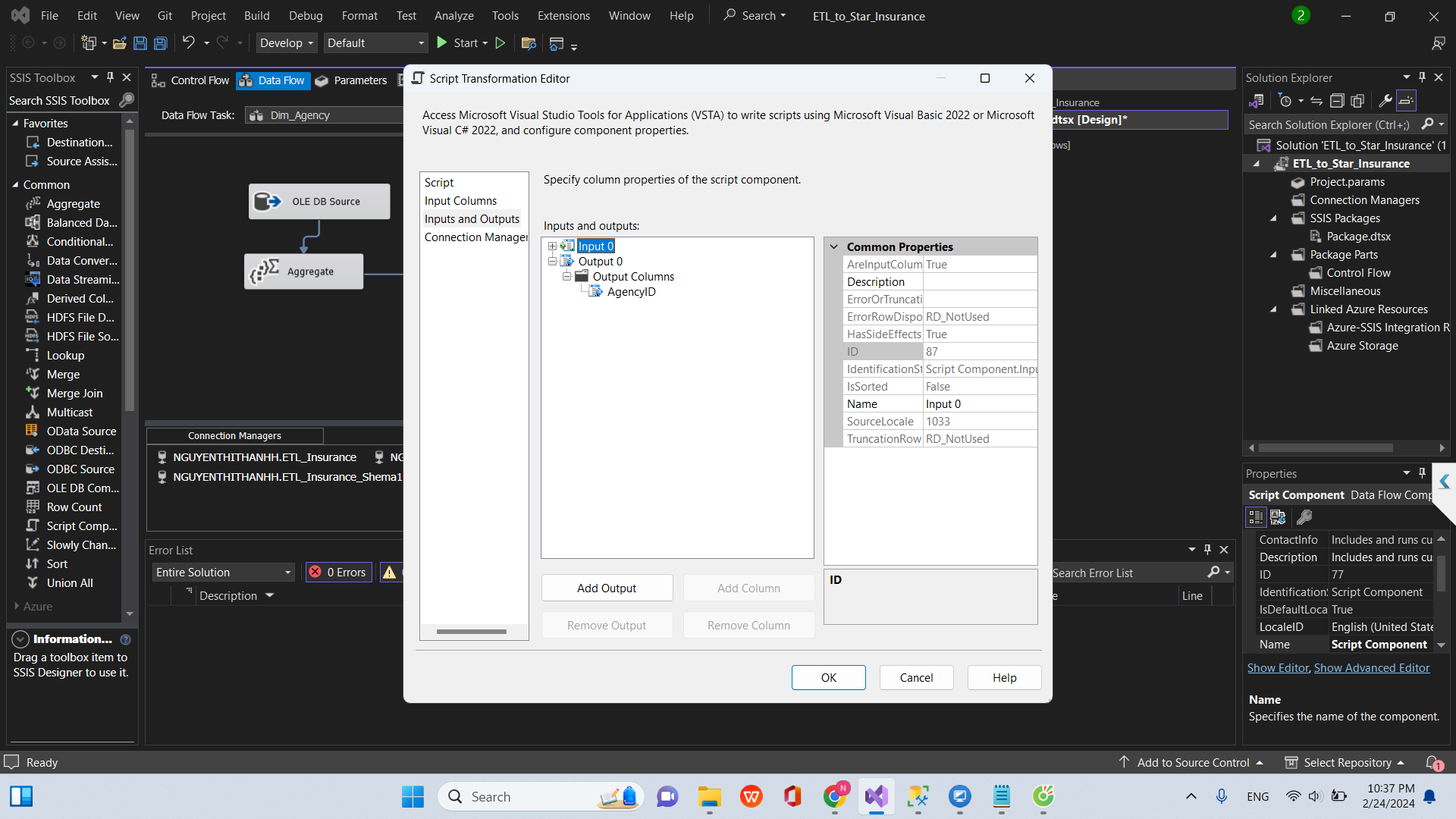


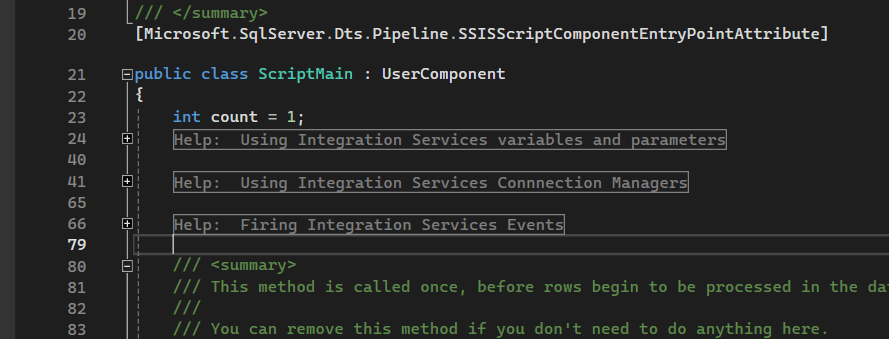
* Ảnh bảng chiều Agency

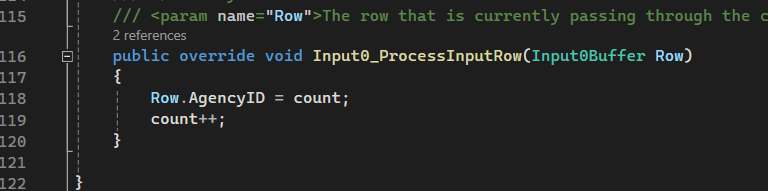
Dim Agency: (Tương tự với Dim Product, Dim Destination)

* OLE DB Source(Object Linking and Embedding, Database): lấy dữ liệu gốc sql từ kết quả của ETL CSV
* Aggregate: tích hợp dữ liệu để có thể thực hiện các phép tính tổng hợp như tổng, trung bình, tối đa, tối thiểu, v.v., trên dữ liệu được nhóm lại theo một hoặc nhiều cột.
* Script Component: viết mã script tự động bằng ngôn ngữ C# hoặc Visual Basic.NET để xử lí dữ liệu
* Sort: sắp xếp dữ liệu theo agency
* OLE DB Destination: xuất dữ liệu bảng chiều agency sang sql







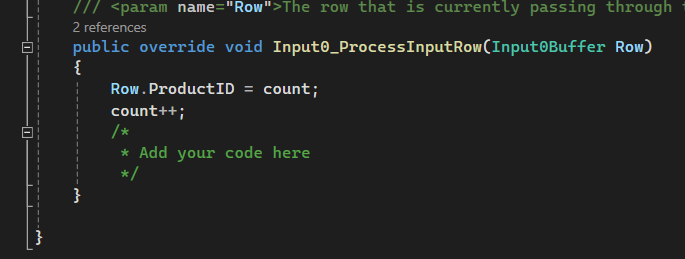


* Ảnh bảng chiều Product

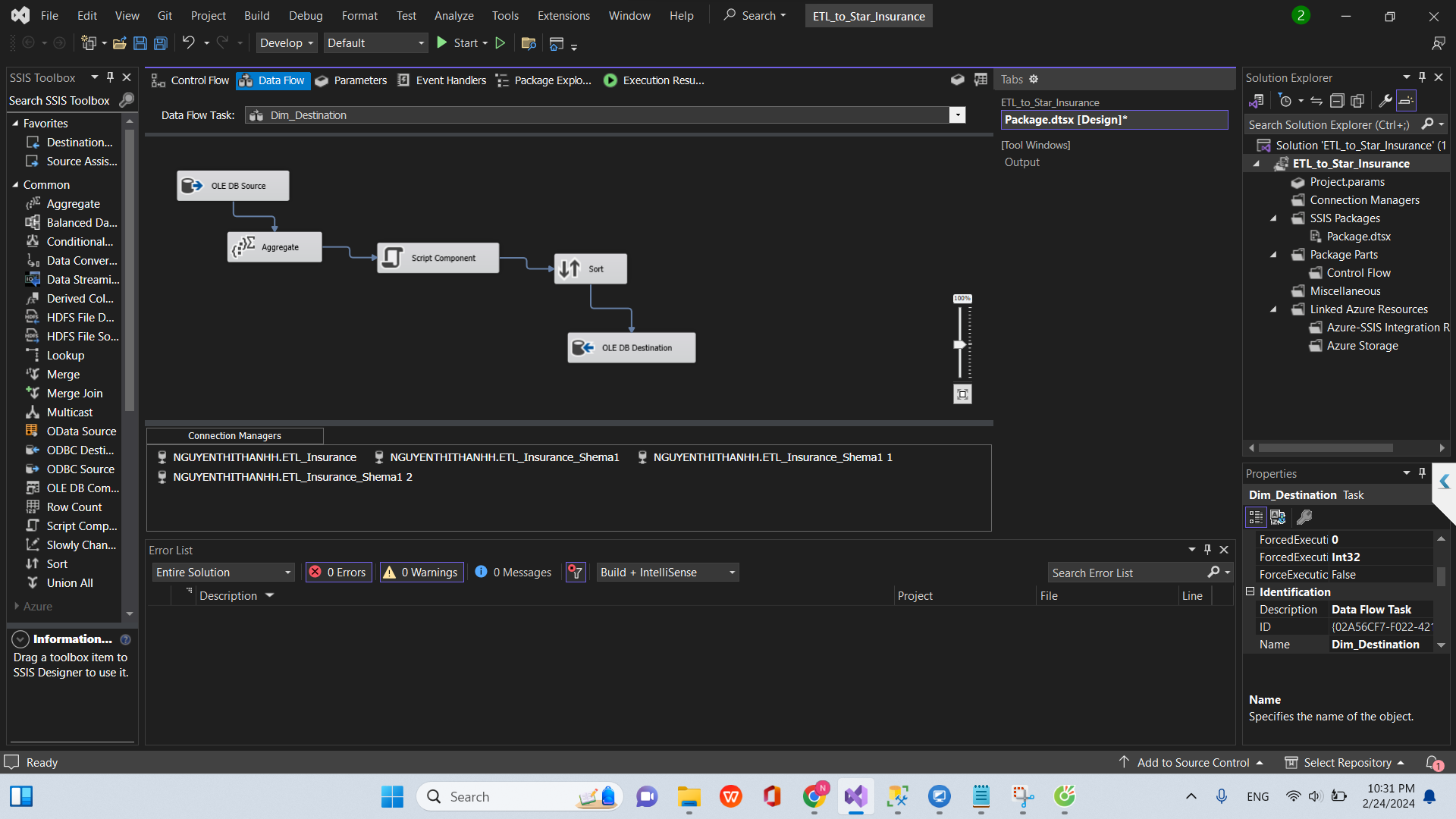


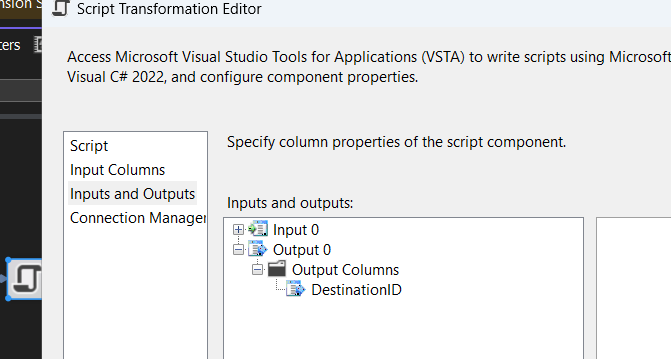


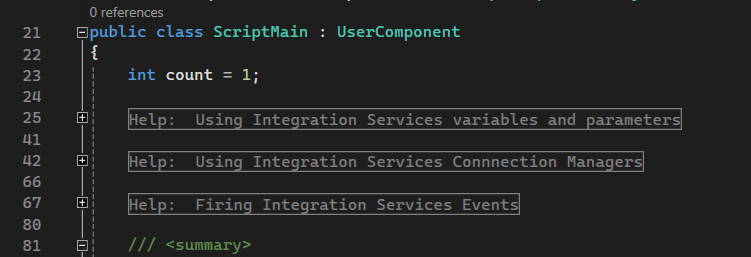


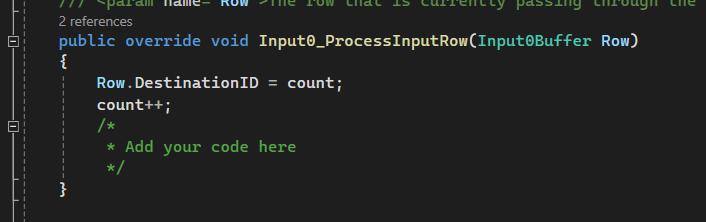


* Ảnh bảng chiều Destination





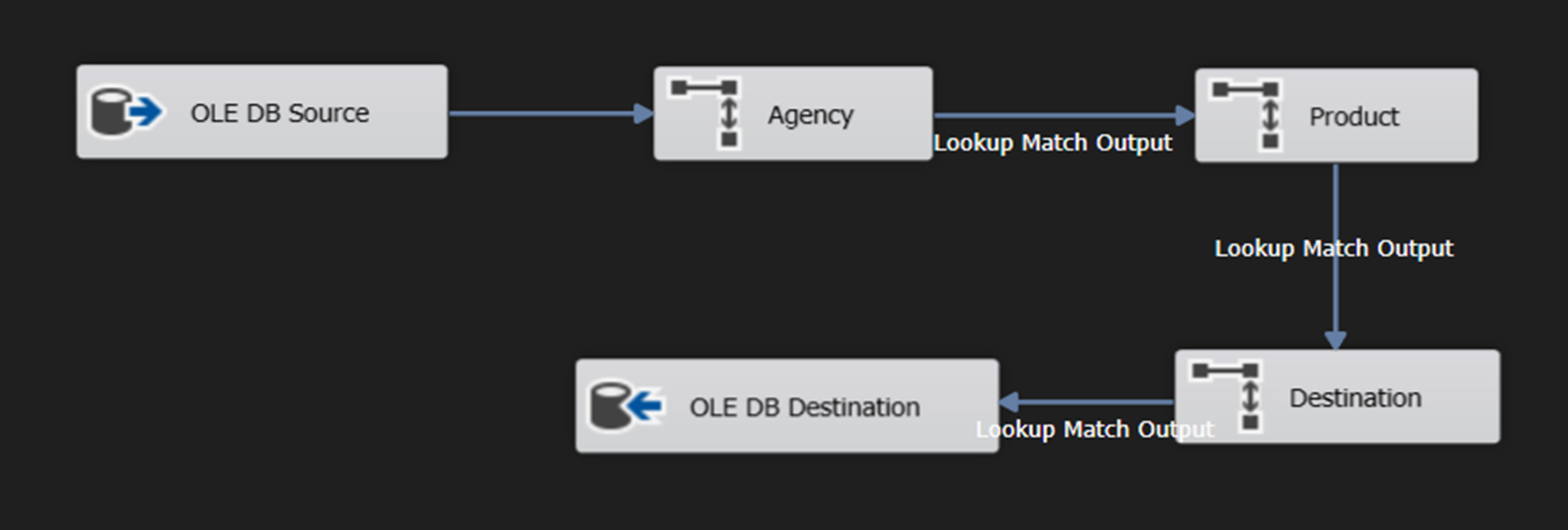


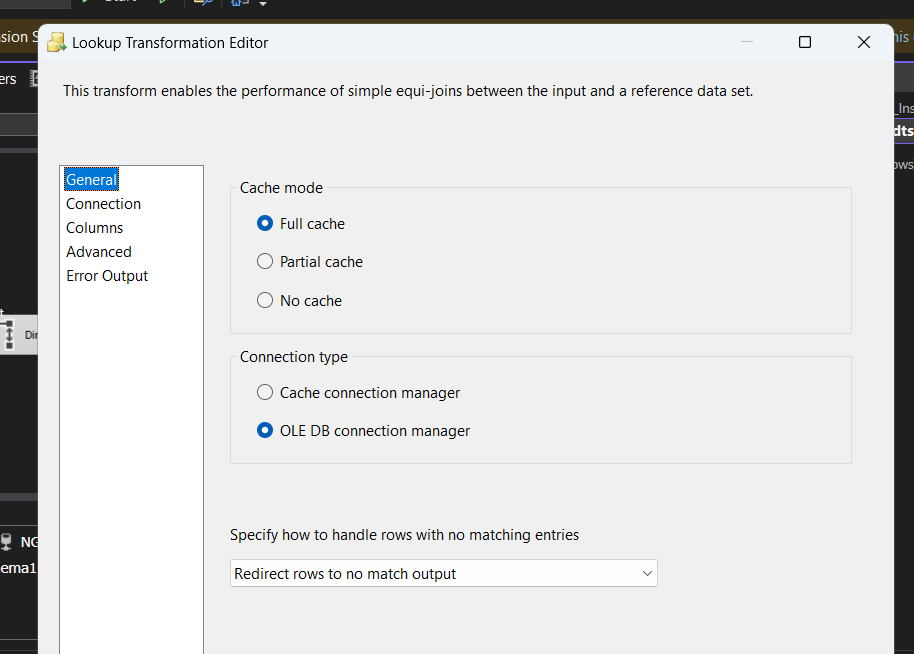


* Ảnh bảng Fact

Lookup Component (Thành phần Lookup):

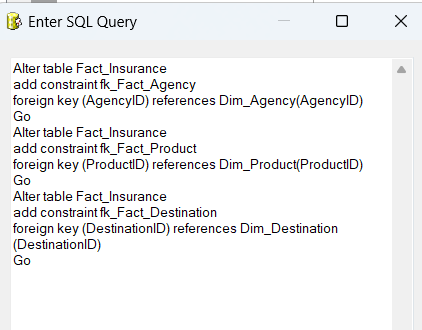
* Chức năng chính:
  + Tra cứu (Lookup): So sánh giữa giá trị của một cột trong nguồn dữ liệu đầu vào với một bảng hoặc tập hợp dữ liệu khác để tìm giá trị tương ứng.
  + Kết hợp (Join): Kết hợp thông tin từ nguồn dữ liệu đầu vào với thông tin từ bảng hoặc tập hợp dữ liệu khác dựa trên một điều kiện
  + Lookup Match Output: nối và kiểm tra dữ liệu có trùng hợp hay không.





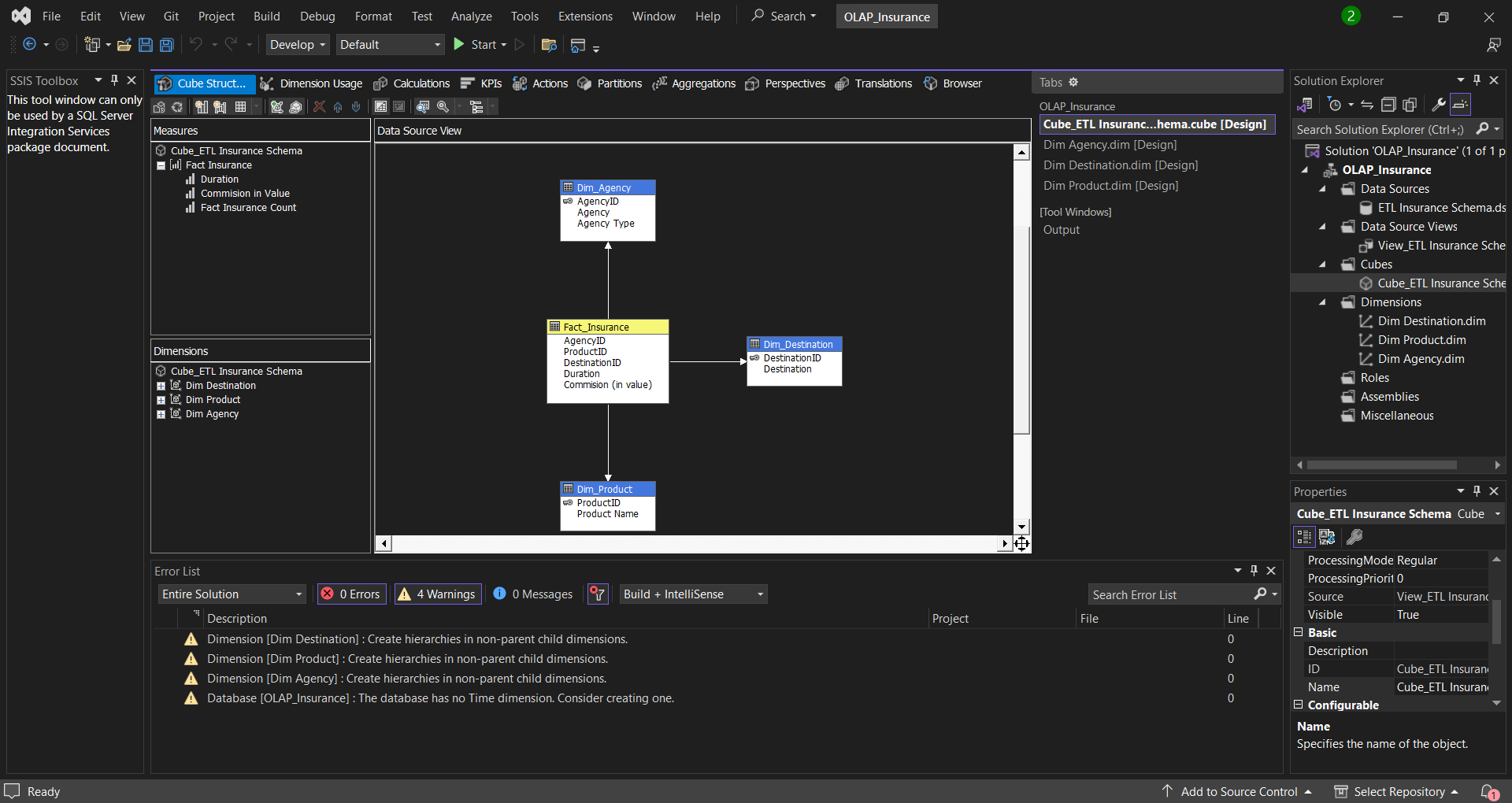
Execute



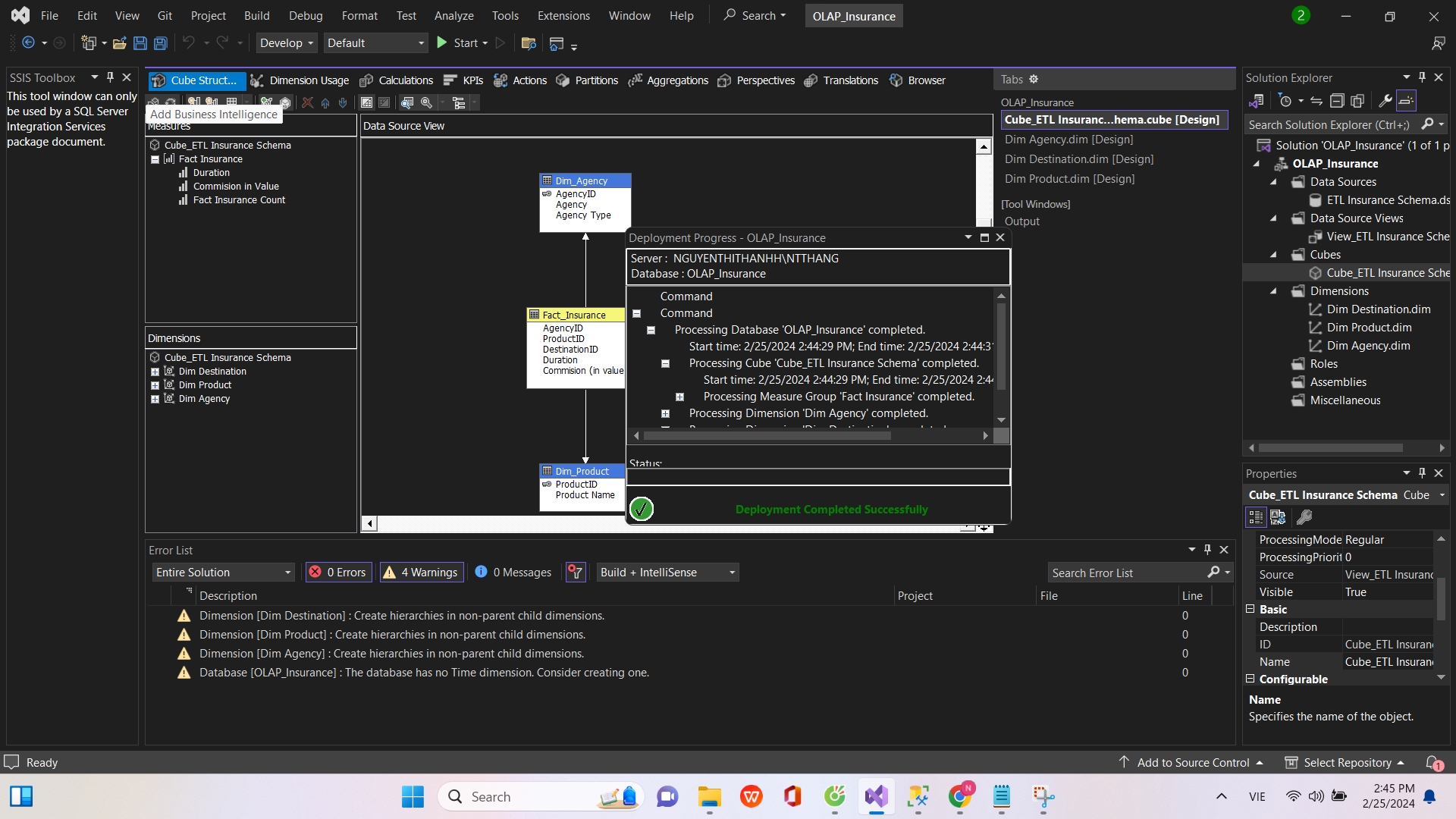


## **OLAP**

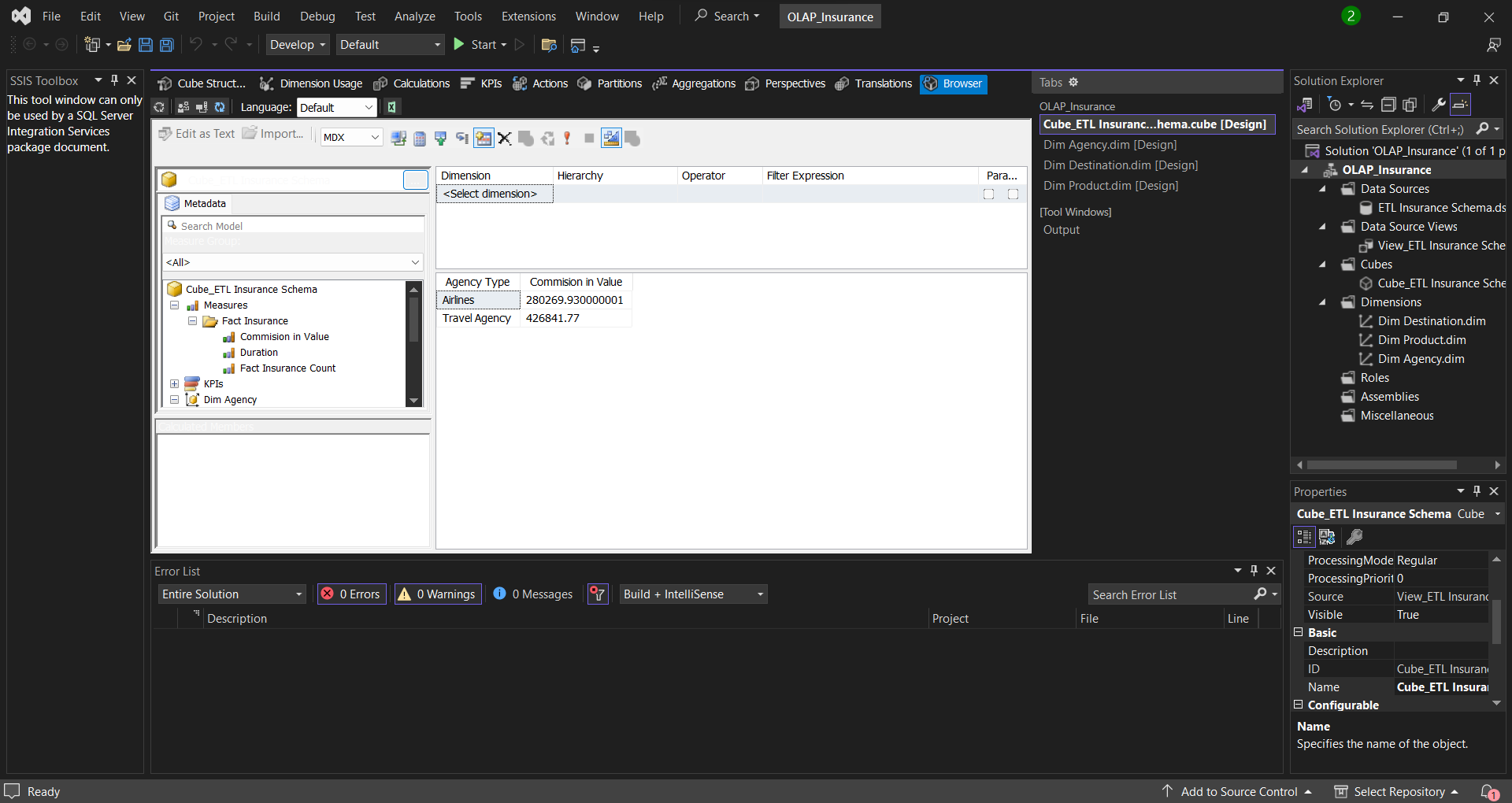
Tạo Cube

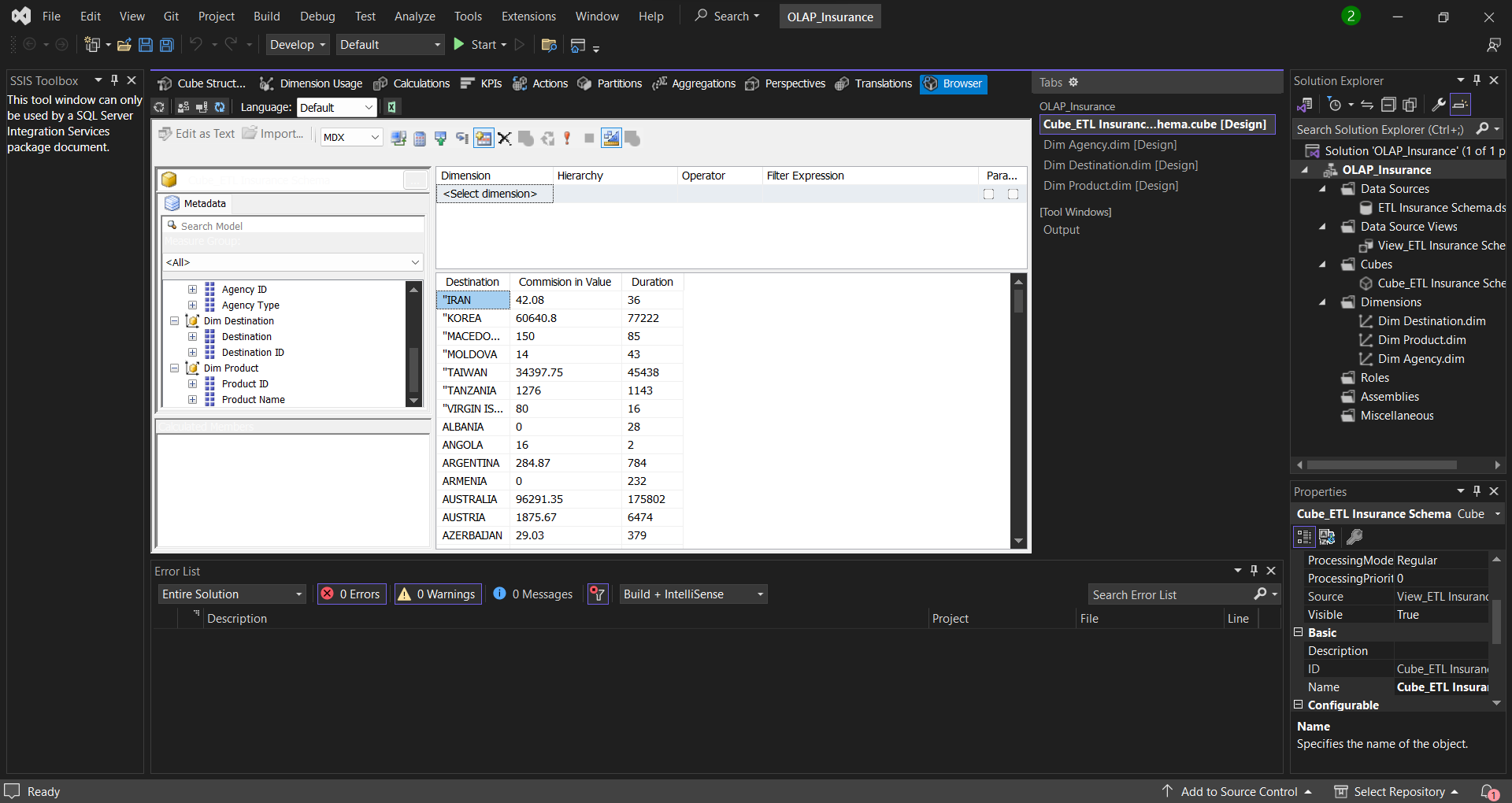


Deploy thành công



1 số ví dụ

* Hoa hồng theo loại hình đại lý bảo hiểm du lịch
* Hoa hồng và thời gian du lịch cho từng địa điểm



## **MDX**

SELECT [measures][dimensions] [on columns|rows]

FROM []

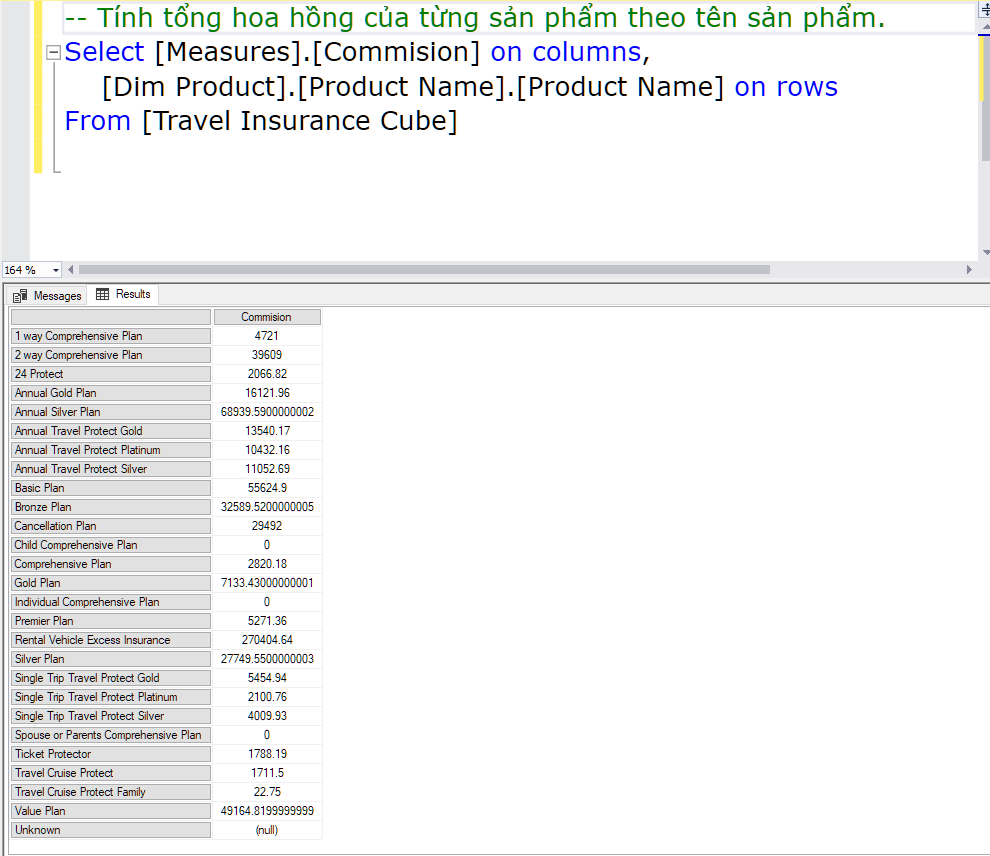
[WHERE

* Tính tổng hoa hồng theo tên sản phẩm

*Select [Measure].[Commission in value] on columns,*

*[Dim Product].[Product Name].[Product Name] on rows*

*From []*

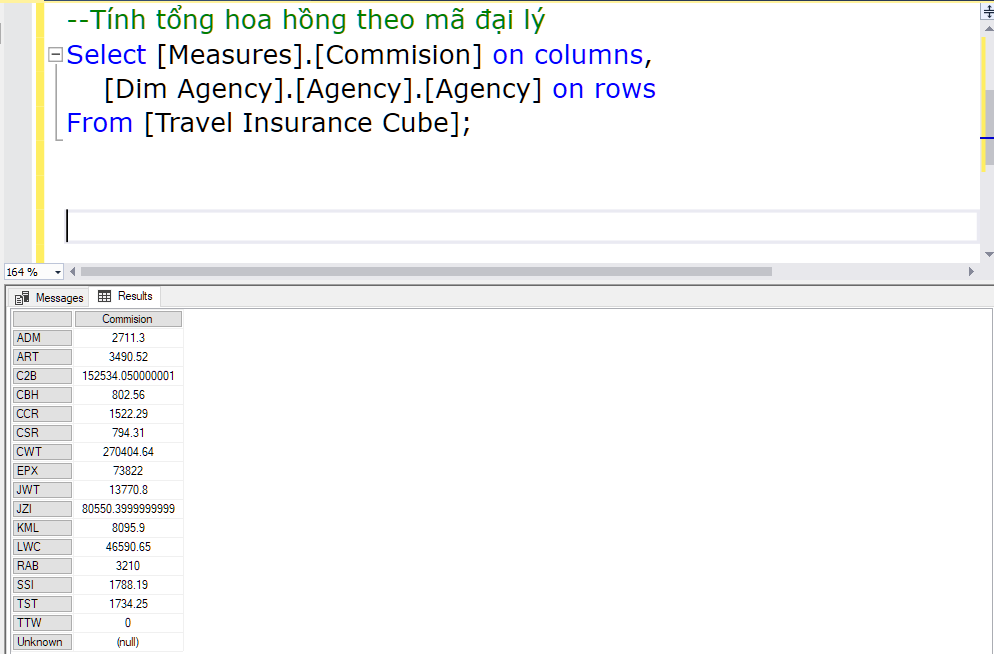


* Tính tổng hoa hồng theo mã đại lý

*Select [Measure].[Commission in value] on columns,*

*[Agency].[Agency].[Agency] on rows*

*From Cube\_ETL Insurance Schema*

**

* Tính tổng hoa hồng theo địa điểm

*Select [Measure].[Commission in value] on columns,*

*[Destination].[Destination].[Destination] on rows*

*From Cube\_ETL Insurance Schema*

* Tính thời gian trung bình của sản phẩm theo từng đại lý

*Select [Measure].[Duration] on columns,*

*[Agency].[Agency].[Agency] on rows*

*From Cube\_ETL Insurance Schema*

*Where [Product].[Product Name]*

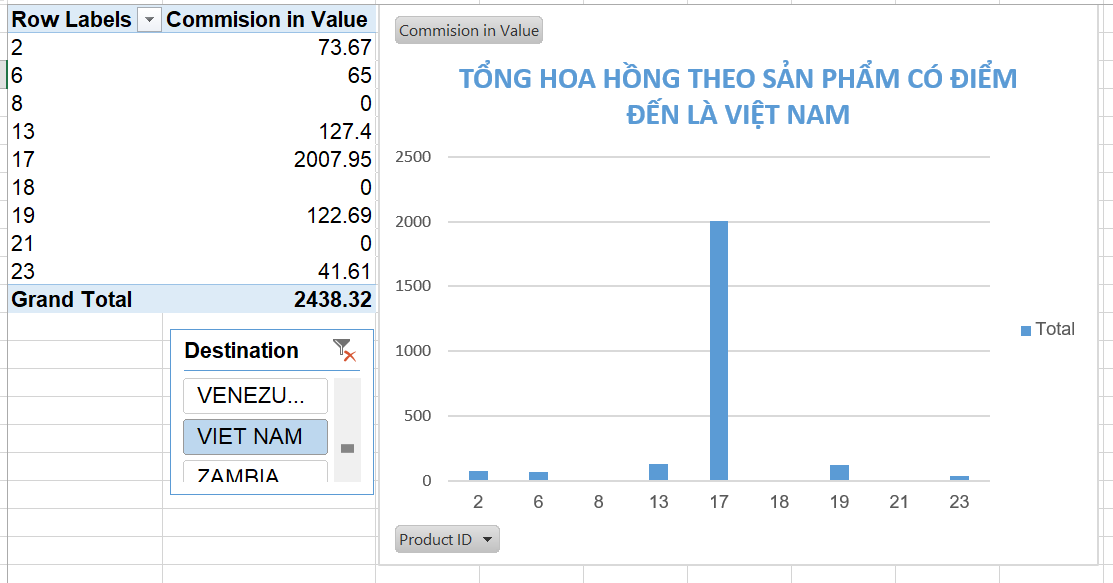
* Tính số sản phẩm mà mỗi đại lý bán được

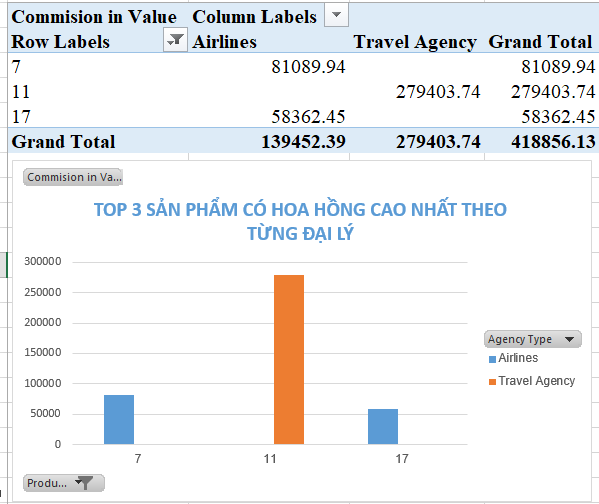
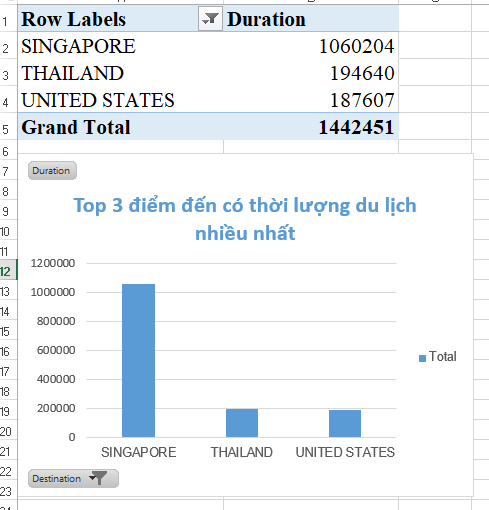
*Select [Agency].[Agency].[Agency] on columns,*

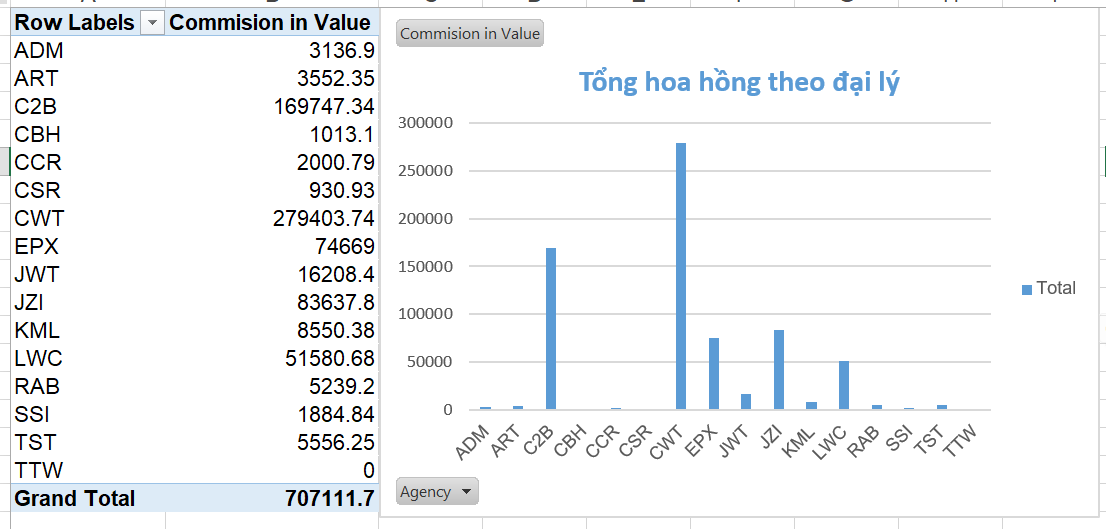
*[Product].[Product Name].[Product Name] on rows*

*From [Cube\_ETL Insurance Schema]*

* Tính số lượng mỗi sản phẩm bán được
* Tính top địa điểm khách du lịch dừng chân lại lâu nhất sắp xếp giảm dần
* Tính tổng hoa hồng của (sản phẩm cụ thể) - rental vehicle excess Insurance
* Tính thời gian trung bình của sản phẩm theo từng đại lý
* Tính tổng hoa hồng từ sản phẩm X và Y theo khoảng thời gian A đến thời gian B

7. 



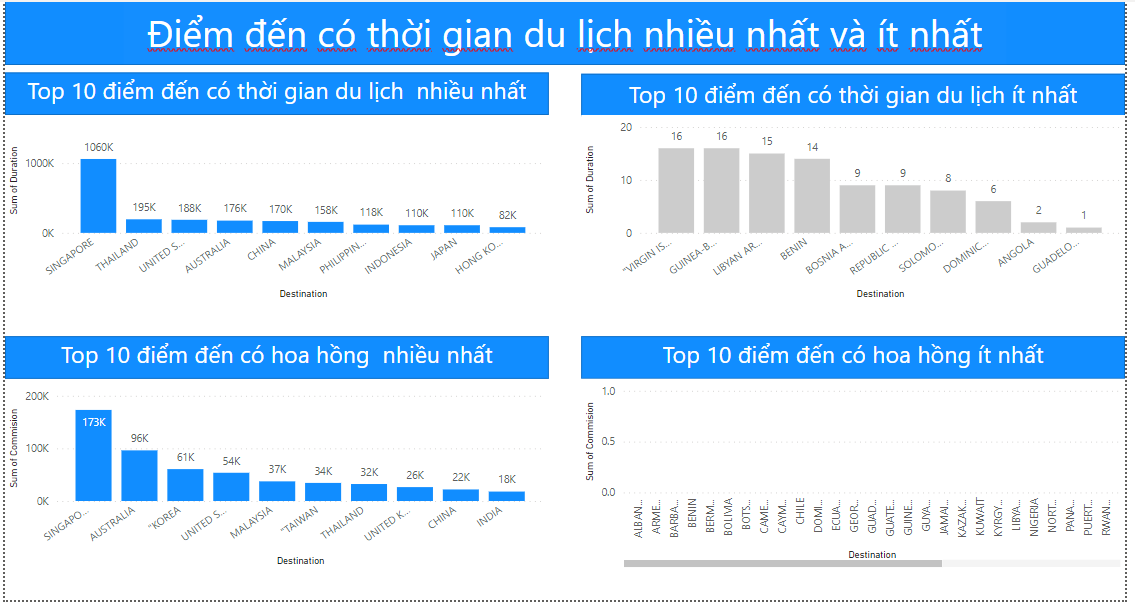


8. Dashboard



* Tổng thời gian du lịch của khách hàng: 3123K ngày
* Tổng hoa hồng thu được: 707.11K
* Biểu đồ xem xét tổng thời gian du lịch và tổng hoa hồng theo loại hình đại lý:
* Đại lý du lịch thì được sử dụng nhiều hơn so với loại hình Airlines (hãng hàng không) mặc dù số lượng hoa hồng/loại bảo hiểm khá bằng nhau (0.22 và 0.23)

→ Trong loại hình đạy lý du lịch sẽ có 1 vài chính sách hỗ trợ tốt hơn cho khách hàng ngoài hoa hồng.



* Xếp hạng 10 điểm đến có tổng thời gian du lịch nhiều nhất và ít nhất.
* Xếp hạng 10 điểm đến có hoa hồng nhiều nhất và ít nhất.

→ Nhận xét: Các quốc gia có tổng thời gian du lịch cao thì lượng hoa hồng có xu hướng cao hơn những quốc gia còn lại. Những quốc gia có tổng thời lượng du lịch thấp thì có hoa hồng thấp hơn.

→ Chứng tỏ các quốc gia mạnh về du lịch sẽ có khoản hoa hồng cao hơn các quốc gia còn lại.



* Ở loại hình đại lý Arilines tương quan giữa tổng hoa hồng và tổng thời gian du lịch theo sản phẩm không ổn định:
* Có lúc khi tổng thời lượng du lịch cho 1 sản phẩm tăng thì hoa hồng sẽ tăng theo nhưng có lúc 2 chỉ số này lại ngược lại với nhau. Ví dụ: Khi tổng thời gian du lịch của người dùng chọn sản phẩm Ticket Protector tăng đến đỉnh thì hoa hồng lại tụt xuống mức thấp nhất.
* Ở loại hình Travel Agency tương quan giữa tổng hoa hồng và tổng thời gian du lịch theo sản phẩm ổn định hơn.
* Mặc dù có một vài sản phẩm khi có tổng thời lượng du lịch tăng nhưng tổng hoa hồng lại giảm nhưng không đáng kể.

→ Nhận xét chung: Tổng hoa hồng sẽ cao hay thấp phụ thuộc chủ yếu vào điểm đến của khách hàng.