1. **Giải thích dữ liệu**

Tên của bộ dữ liệu : Weather Conditions in World War Two.

Link của bộ dữ liệu: <https://www.kaggle.com/smid80/weatherww2/data>

* 1. Giải thích loại bỏ dữ liệu thừa:

|  |  |  |
| --- | --- | --- |
| Tên thuộc tính ban đầu | Ý nghĩa | Lí do loại bỏ |
| WindGustSpd | Tốc độ gió cực đại (km/h) | Trùng lặp ý nghĩa và giá trị với thuộc tính SPD, ở đây SPQ là tốc độ gió cực đại với đơn vị đo là Hải lí |
| Snowfall | Độ dày tuyết (mm) | Trùng lặp ý nghĩa và giá trị với thuộc tính SNF, ở đây SNF là độ dày của tuyết rơi với đơn vị đo là inch |
| MeanTemp | Nhiệt độ trung bình (oC) | Do đã có maxtemp (nhiệt độ lớn nhất) và mintemp (nhiệt độ nhỏ nhất) nên không cần thiết thuộc tính meantemp |
| YR | Năm quan sát | Trùng lặp với thuộc tính Data(ngày tháng năm) quan sát |
| MO | Tháng quan sát |
| DA | Ngày quan sát |
| MAX | Nhiệt độ lớn nhất (oF) | Trùng lặp ý nghĩa với Maxtemp và mintemp, ở đây maxtemp và mintemp là đo theo độ C |
| MIN | Nhiệt độ nhỏ nhất (oF) |
| MEA | Nhiệt độ trung bình (oF) |
| FT | Frozen Ground Top (depth in inches) | Không có dữ liệu. |
| FB | Frozen Ground Base (depth in inches) |
| FTI | Frozen Ground Thickness (thickness in inches) |
| ITH | Ice Thickness on Water (inches and tenths) |
| SD3 | Snow depth at 0030 GMT includes ice pellets in inches and tenths |
| RHX | 24-hour maximum relative humidity, as a whole percent |
| RHN | 24-hour minimum relative humidity, as a whole percent |
| RVG | River guage in feet and tenths |
| WTE | Water equivalent of snow and ice on ground in inches and hundredths |
| TSHDSBRSGF | Là viết tắt của: Thunder; Sleet; Hail; Dust or Sand; Smoke or Haze; Blowing Snow; Rain; Snow; Glaze; Fog; 0 = No, 1 = Yes | Trùng lặp ý nghĩa và giá trị với thuộc tính PoorWeather |
| SND | Snow depth (includes ice pellets) recorded at 1200 GMT except 0000 GMT in Far East Asian Area in inches and tenths | Do chỉ có một giá trị duy nhất là 0, không có ý nghĩa để sử dụng |
| PRCP | Lượng mưa (inch) và lấy đến 2 số thập phân | Trùng ý nghĩa và giá trị với thuộc tính Precip, ở đây nghĩa là lượng mưa theo mm |

* 1. Giải thích thuộc tính bảng fact:

|  |  |  |
| --- | --- | --- |
| Tên thuộc tính | Ý nghĩa | Ghi chú |
| FactID | Khóa chính |  |
| StationID | Khóa ngoại |  |
| DateID | Khóa ngoại |  |
| Precip | Lượng mưa (mm) |  |
| MaxCelsius | Nhiệt độ tối đa (độ C) | Đổi tên từ thuộc tính MaxTemp |
| MinCelsius | Nhiệt độ tối thiểu (độ C) | Đổi tên từ thuộc tính MinTemp |
| PoorWeatherID | Khóa ngoại |  |
| MaximumWindySpeedKnots | Tốc độ gió cực đại (hải lý) | Đổi tên từ thuộc tính SPD |
| SnowDepthInch | Độ dày của tuyết (inch) | Đổi tên từ thuộc tính SNF |
| WindyTime | Thời điểm gió thổi | Đổi tên từ thuộc tính PGT |
| WindyDriection | Hướng gió thổi | Đổi tên từ thuộc tính DR |

* 1. Giải thích thuộc tính bảng Date:

|  |  |
| --- | --- |
| Tên thuộc tính | Ý nghĩa |
| DateID | Khóa chính |
| Day | Ngày quan sát |
| Mon | Tháng quan sát |
| Year | Năm quan sát |

* 1. Giải thích thuộc tính bảng Country

|  |  |
| --- | --- |
| Tên thuộc tính | Ý nghĩa |
| CountryID | Khóa chính |
| CountryName | Tên của quốc gia |

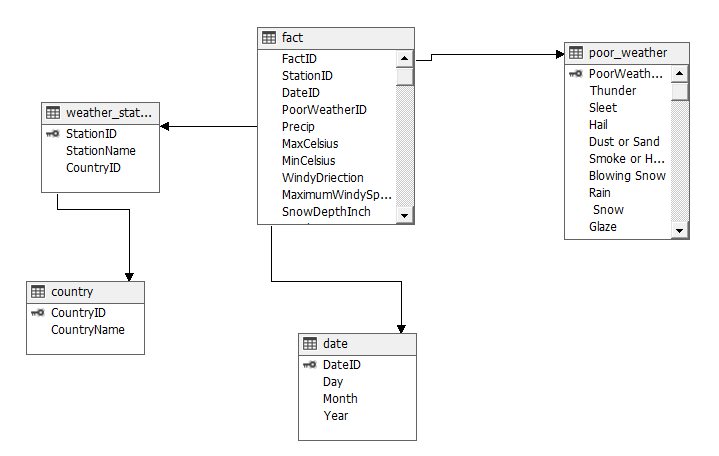
* 1. Giải thích thuộc tính bảng WeatherStation

|  |  |
| --- | --- |
| Tên thuộc tính | Ý nghĩa |
| StationID | Khóa chính |
| StationName | Tên của trạm đo thời tiết |
| CountryID | Khóa ngoại |

* 1. Giải thích bảng PoorWeather

|  |  |  |
| --- | --- | --- |
| Tên thuộc tính | Ý nghĩa | Ghi chú |
| PoorWeatherID | Khóa chính |  |
| Thunder | Sấm sét | Có 3 giá trị thể hiện  1 : có thời tiết tương ứng xảy ra  0 : không có thời tiết ứng xảy ra  -1 : không xác định |
| Sleet | Mưa kèm tuyết rơi |
| Hail | Mưa đá |
| DustOrSand | Bão cát |
| SmokeOrHaze | Sương mù dày đặc |
| BlowingSnow | Tuyết thổi mạnh |
| Rain | Mưa |
| Snow | Tuyết nhẹ |
| Glaze | Hiện tượng đóng băng |
| Fog | Sương mù |

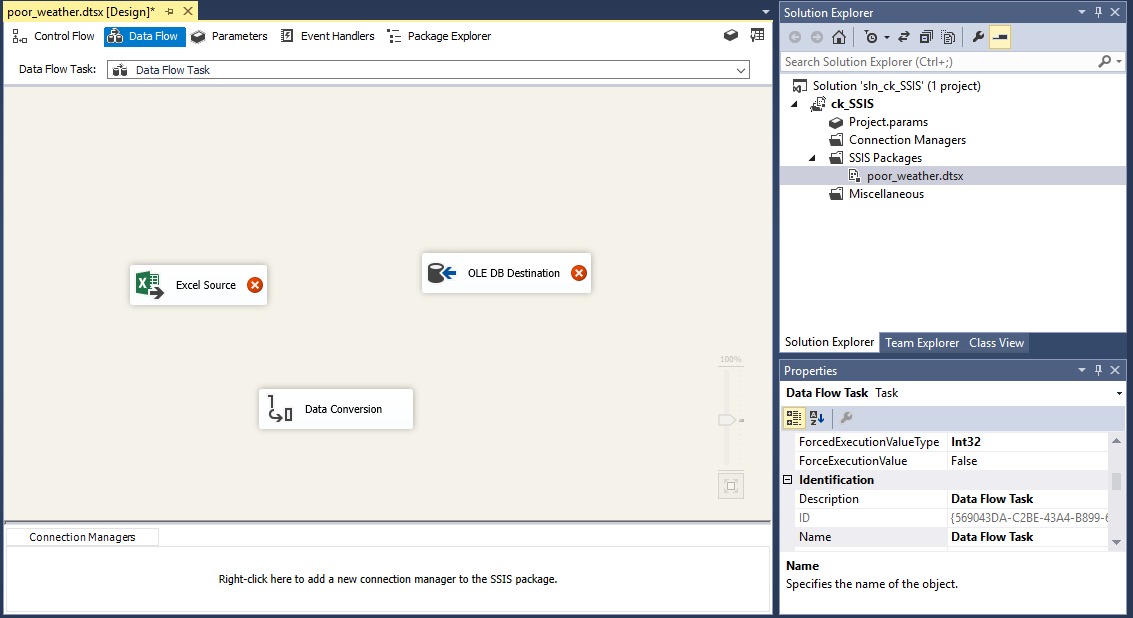
* 1. Thiết kế lược đồ dữ liệu



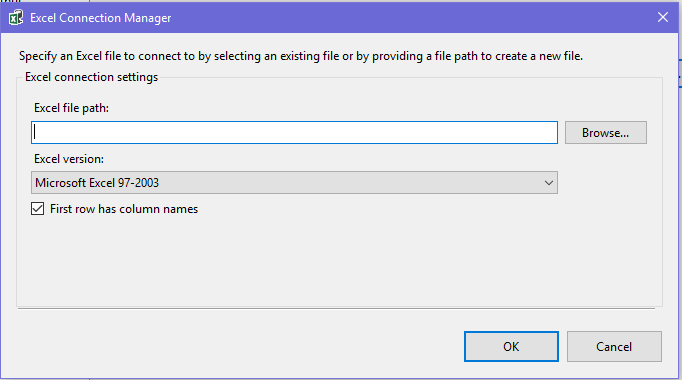
**2. Quá trình SSIS**

**2.1 import dữ liệu bảng poor weather**

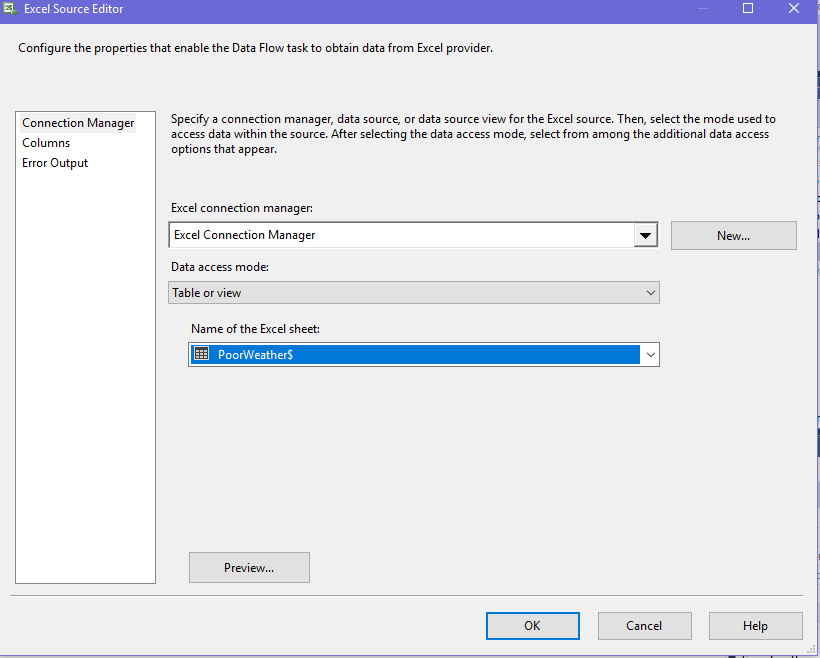
**- tạo mới package poor weather**



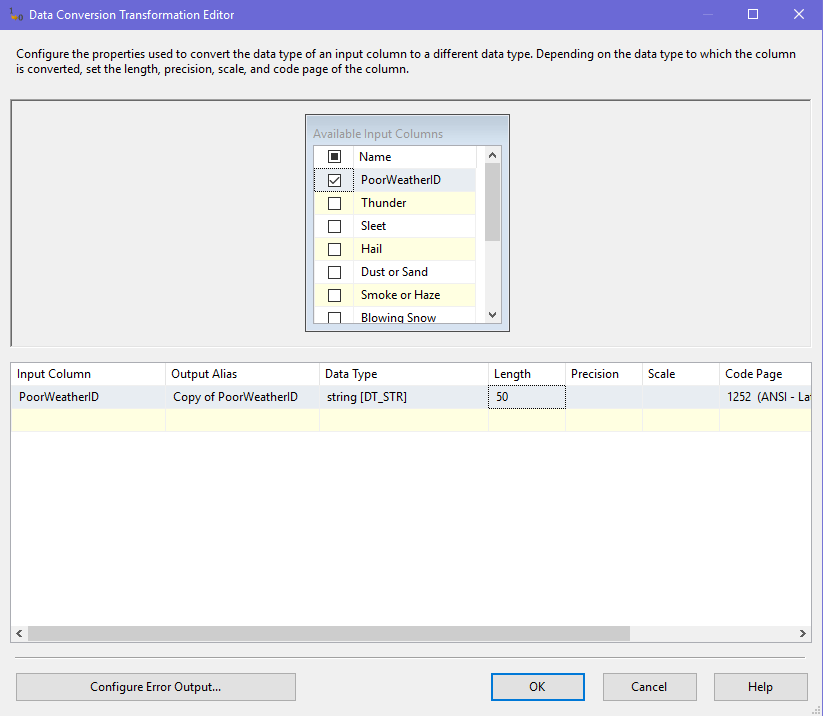
* **Tiến hành chọn file excel nguồn**



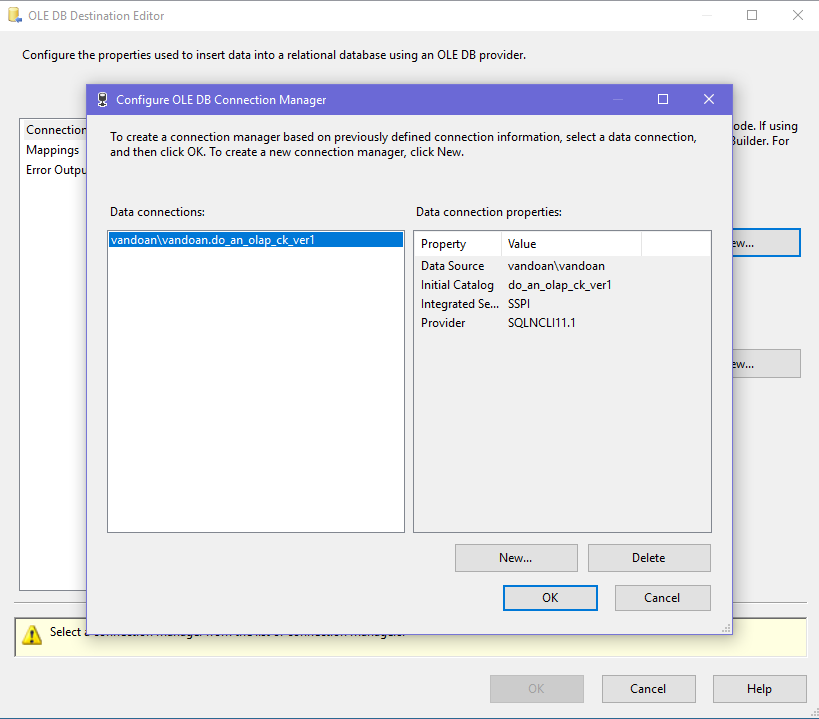
* **Chọn sheet dữ liệu phù hợp**



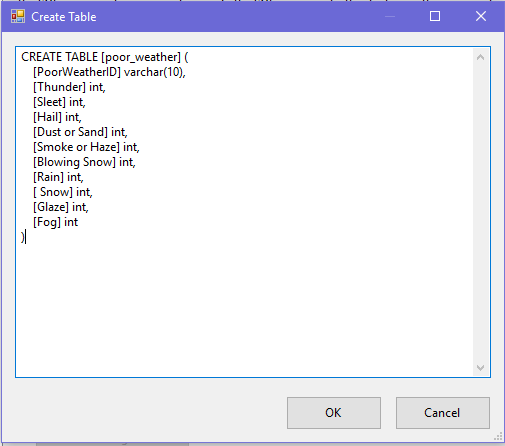
* **Tùy chỉnh lại data conversion để chuyển đổi dữ liệu string trước khi import**



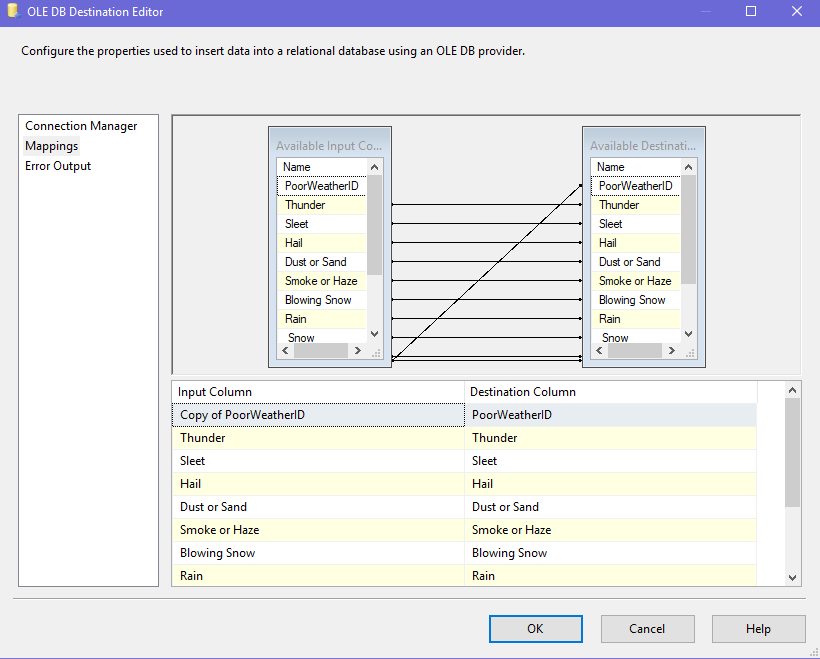
* **Tạo connection tới database**



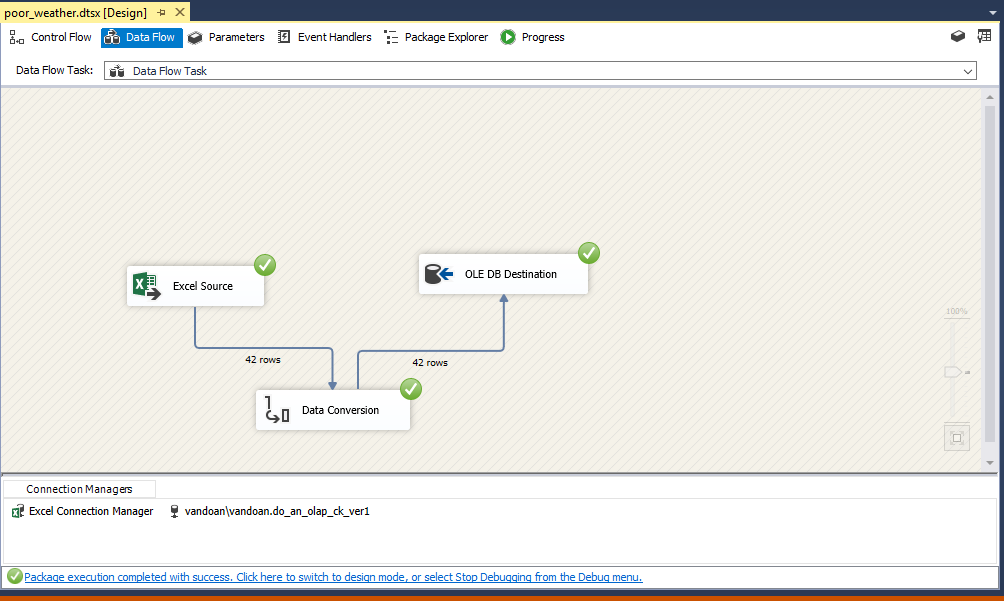
* **Tạo bảng dữ liệu**



* **Mapping lại cột dữ liệu**

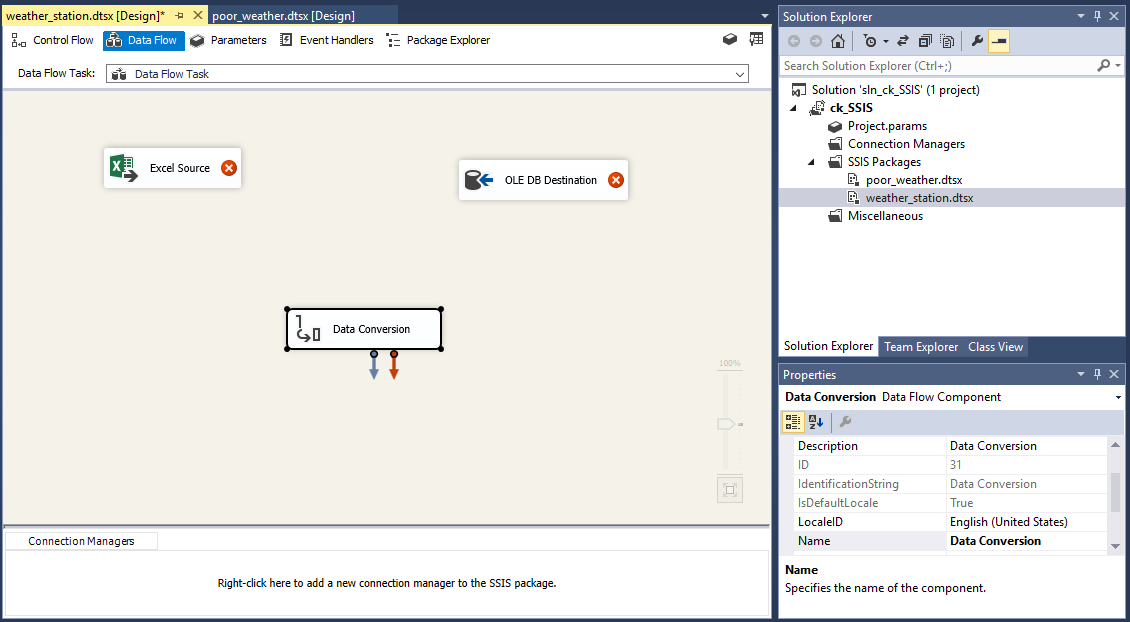


* **Khởi chạy để import dữ liệu**

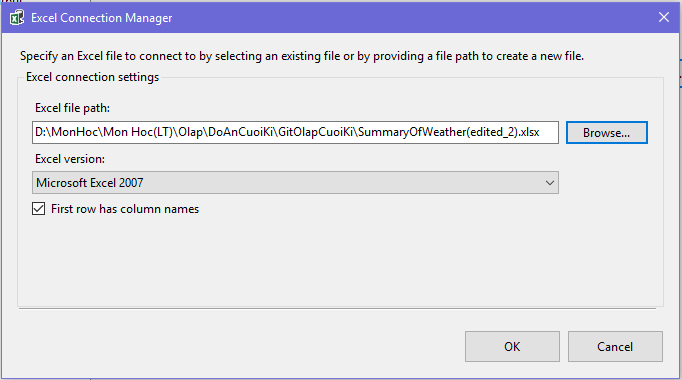


**2.2 import dữ liệu bảng weather station.**

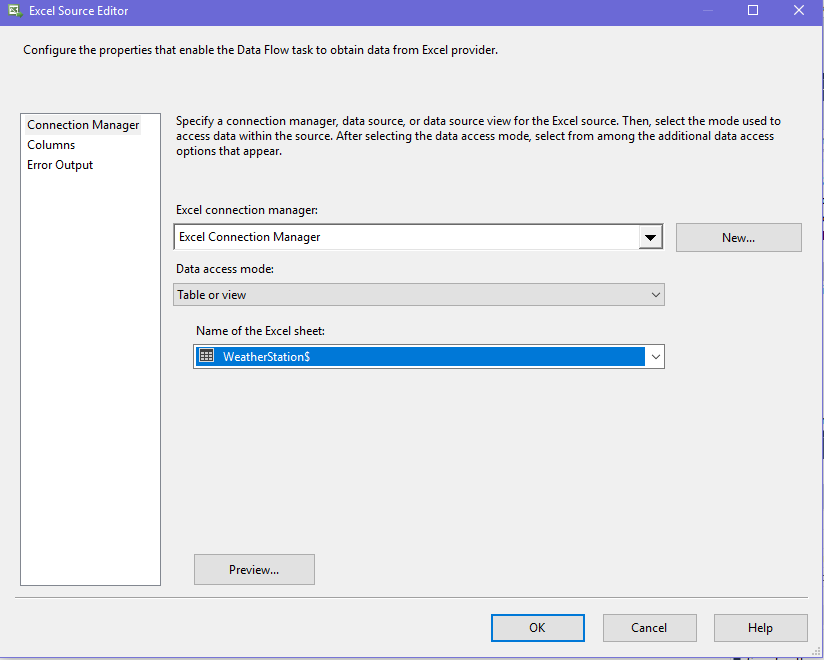
* **Khởi tạo package weather station.**



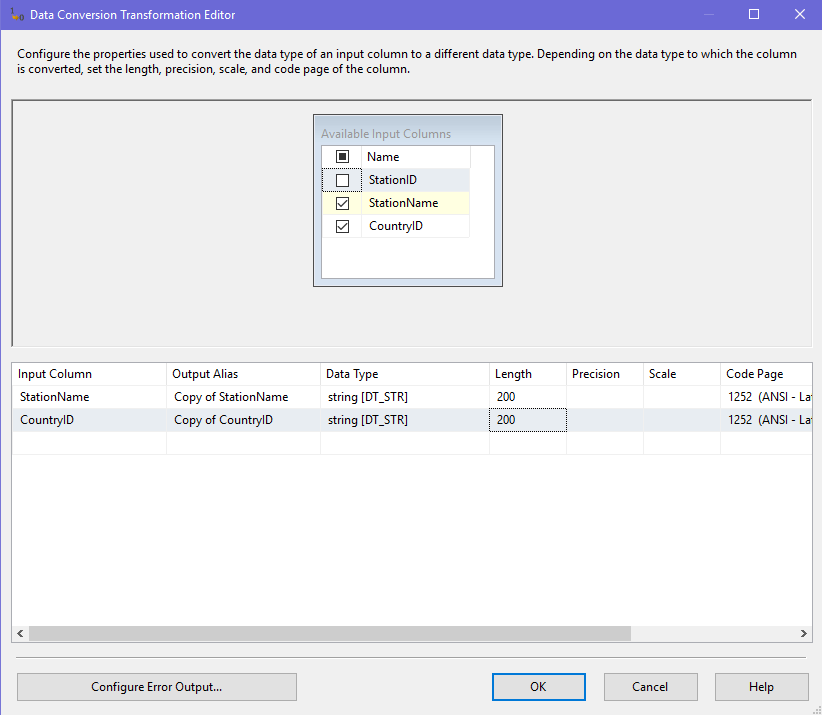
**- chọn dữ liệu để import**



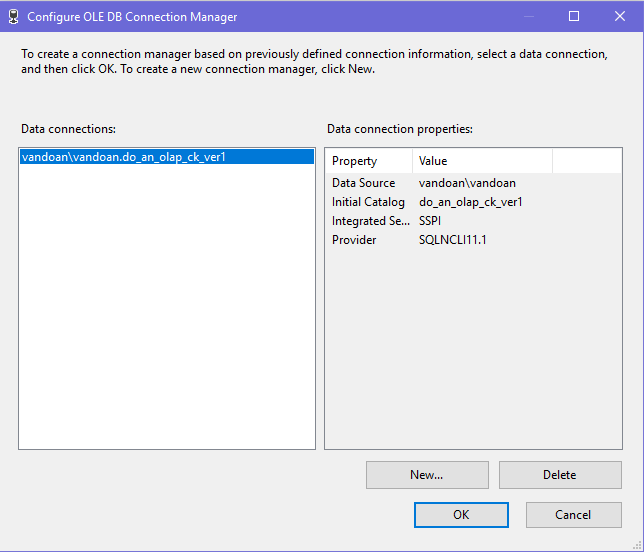
**- chọn sheet để import dữ liệu**



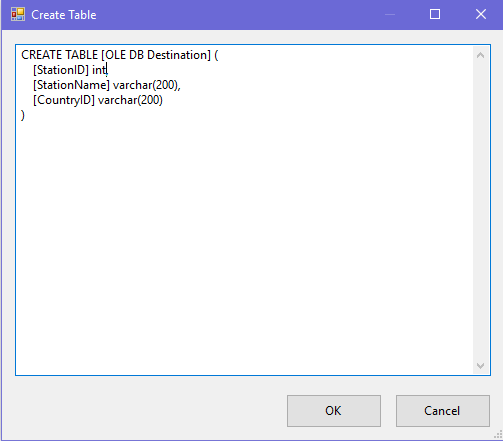
**- dùng data conversion để chuyển đổi dữ liệu string trước khi import**



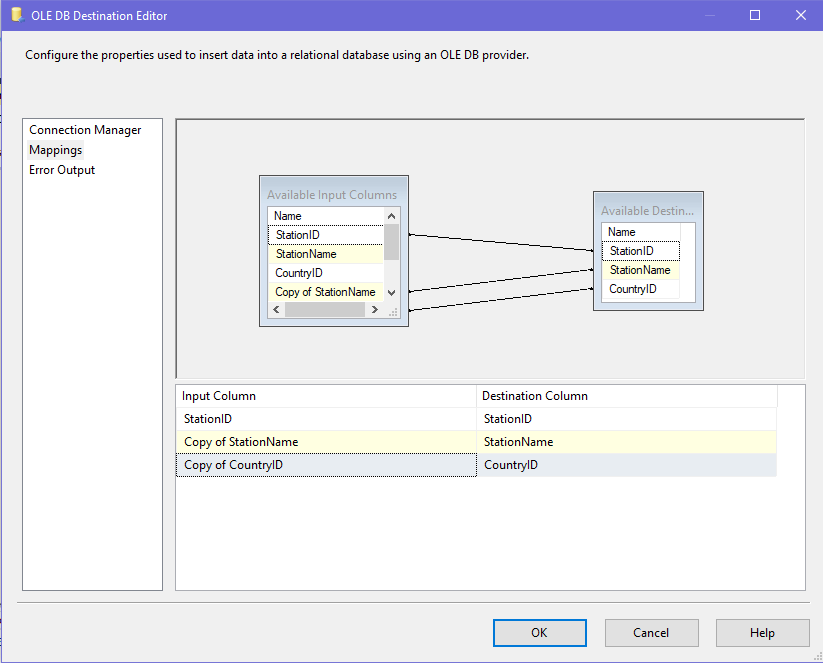
**- chọn kết nối tới database**



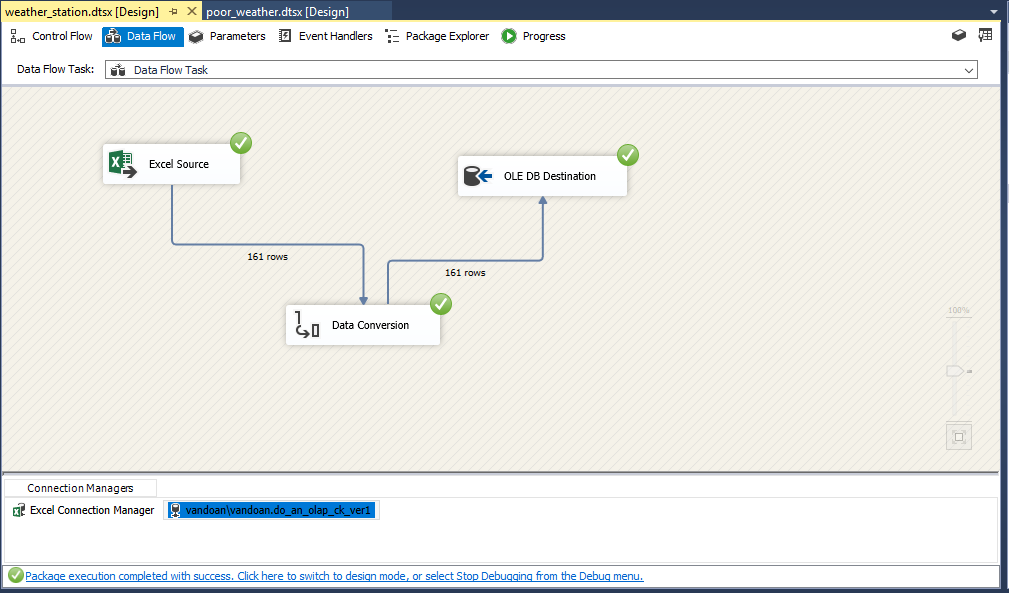
- tạo bảng dữ liệu



- mapping lại dữ liệu

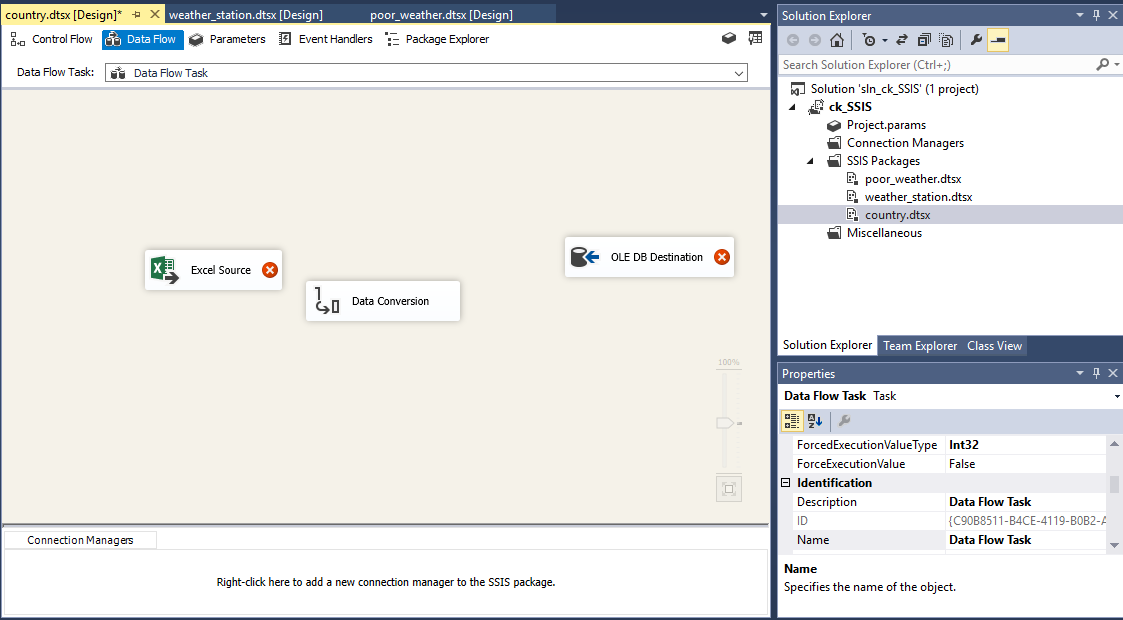


- tiến hành import dữ liệu

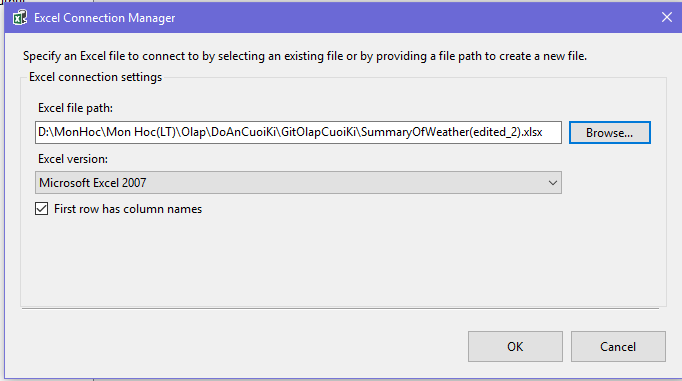


**2.3 import dữ liệu bảng country.**

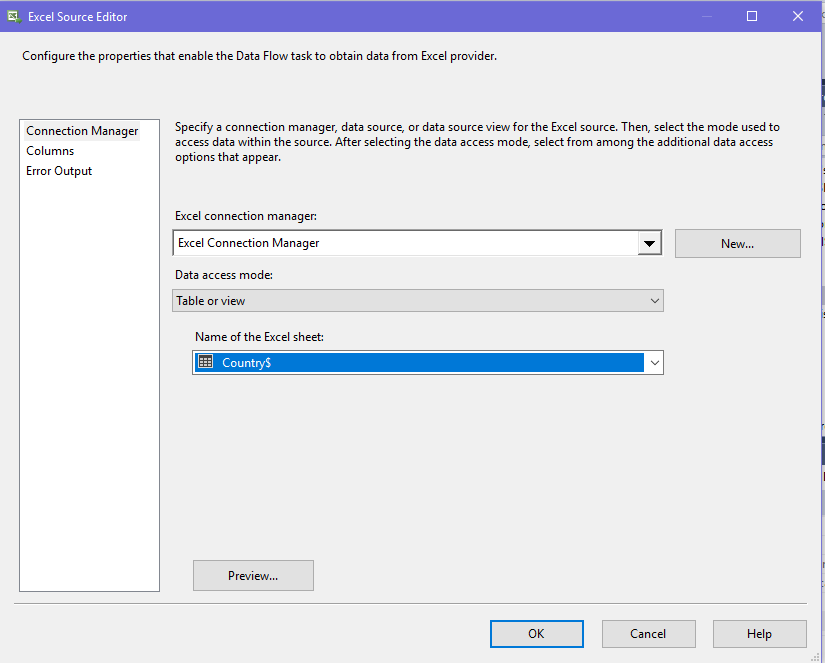
**- khởi tạo package contry**



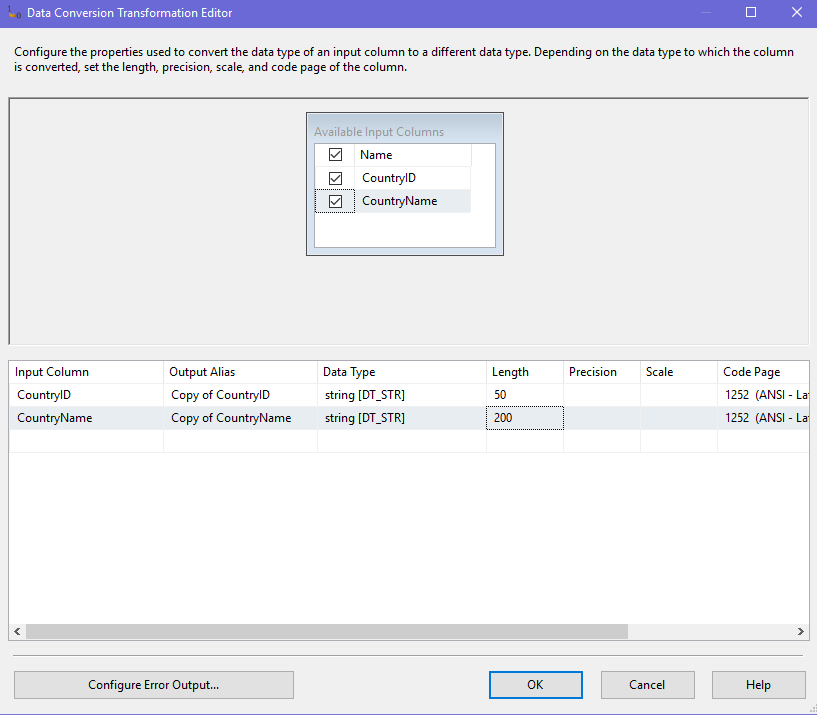
**- tiến hành chọn dữ liệu nguồn**



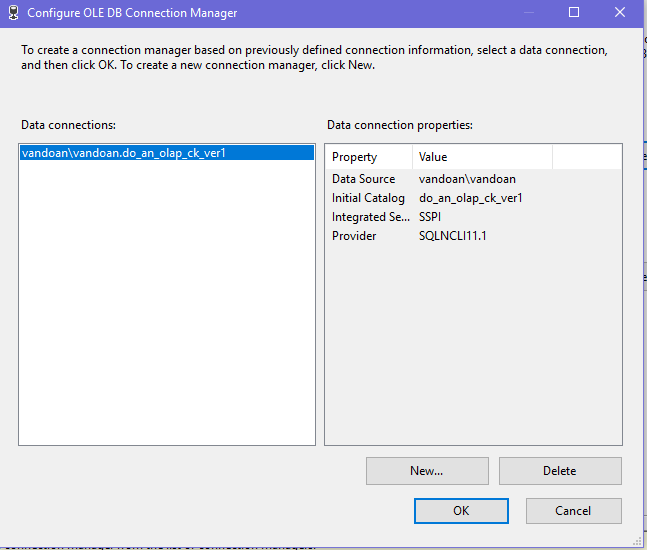
**- tiến hành chọn sheet phù hợp**



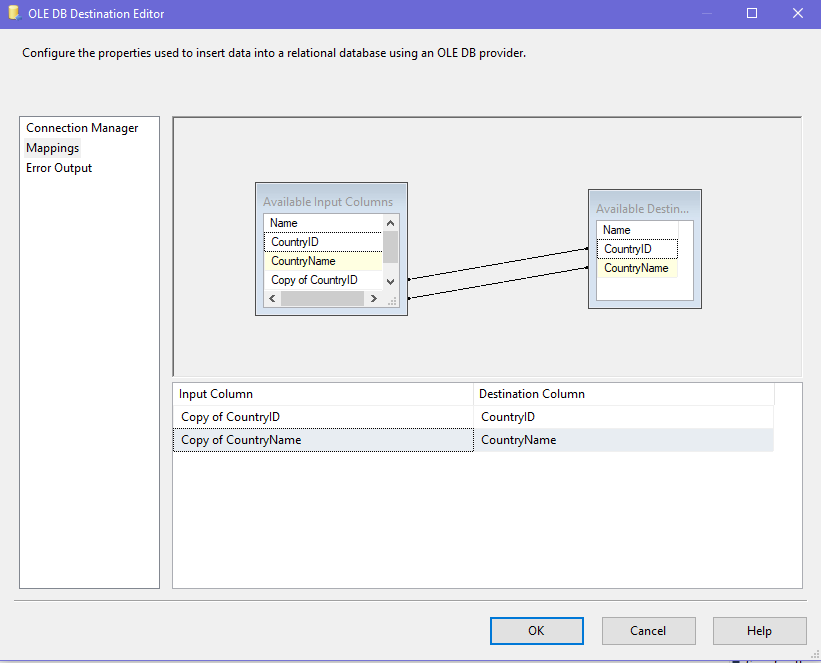
**- sử dụng data conversion để chuyển đổi dữ liệu kiểu string**



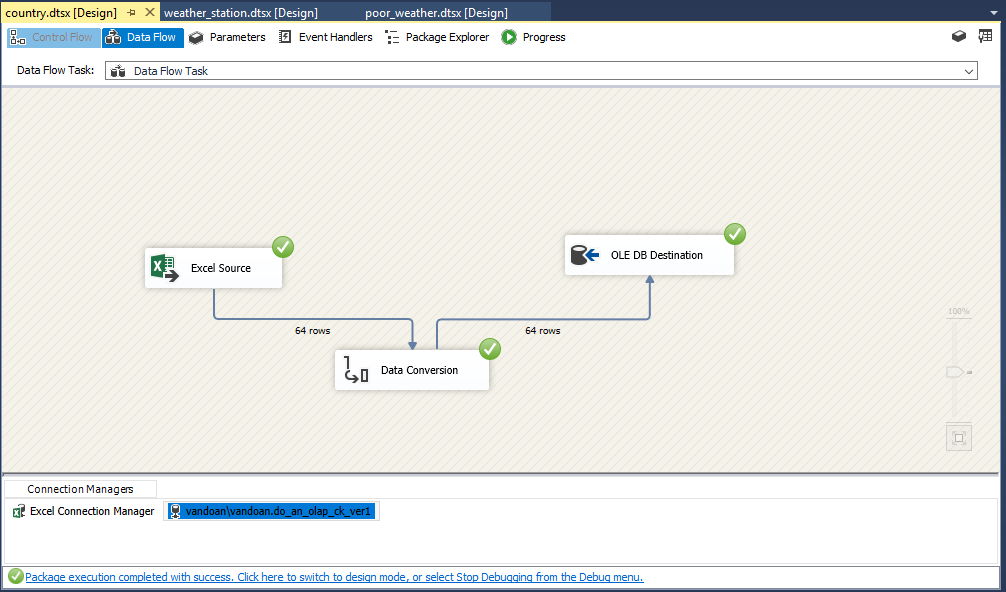
**- chọn nguồn database cần import**



**- tiến hành mapping dữ liệu**

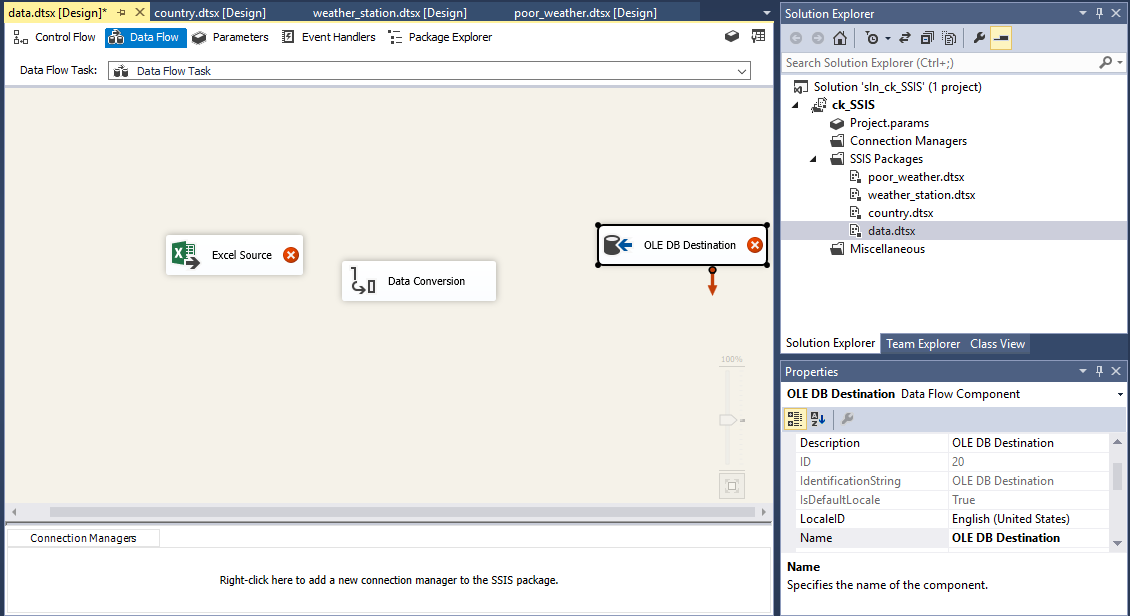


**- tiến hành import dữ liệu**

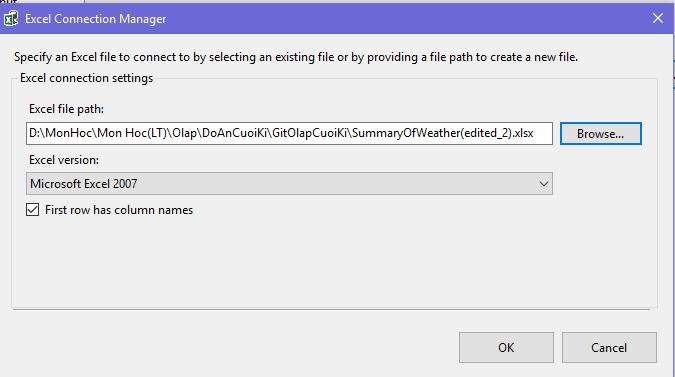


**2.4 import dữ liệu bảng date.**

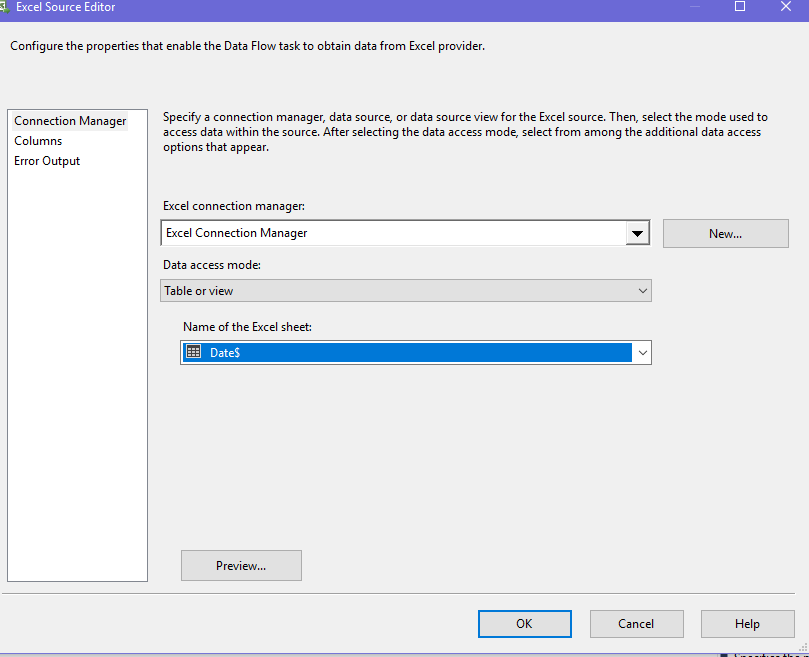
**- khởi tạo package date**



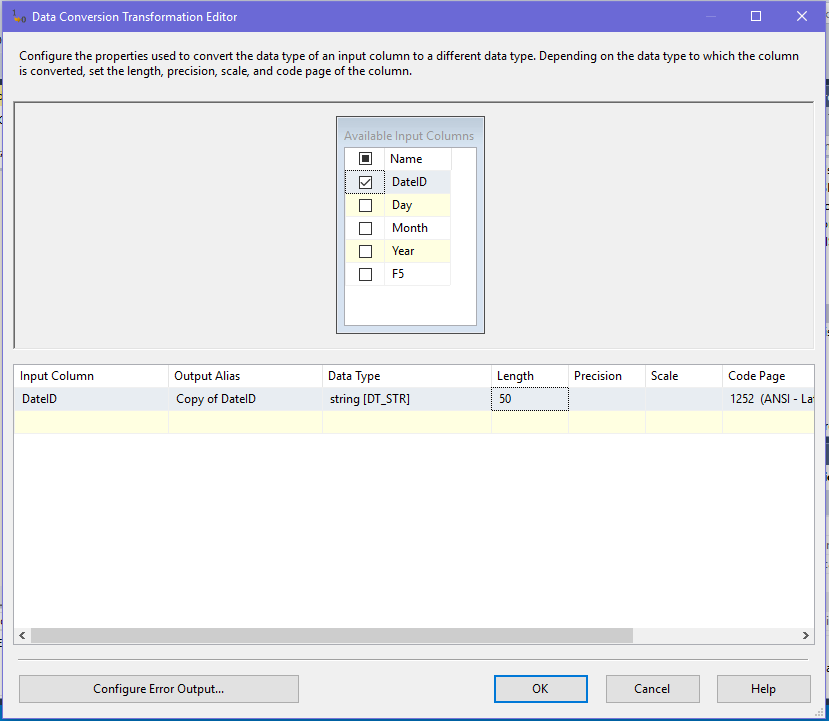
**- chọn nguồn dữ liệu cần import**



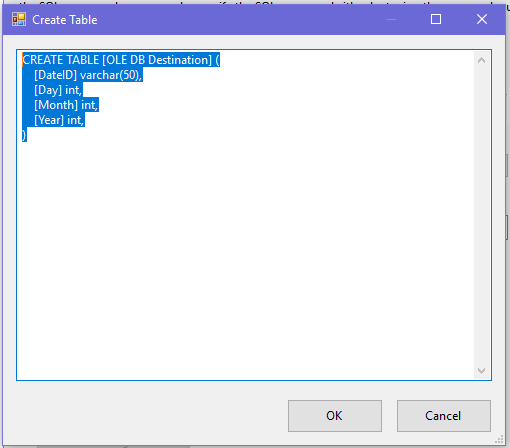
**- chọn sheet dữ liệu phù hợp**



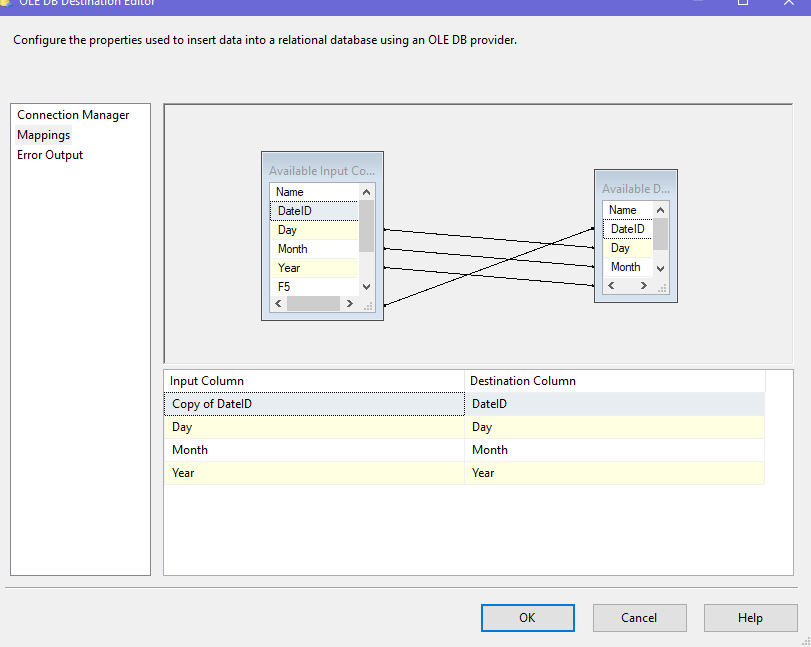
**- dùng data conversion để chuyển đổi dữ liệu kiểu string**



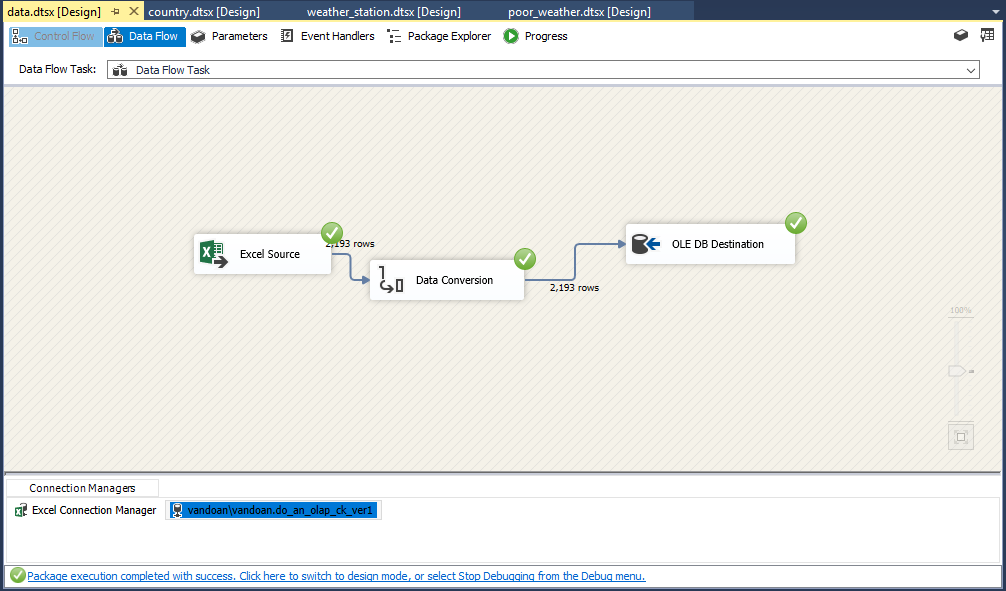
**- tạo bảng date**



**- tiến hành mapping dữ liệu**

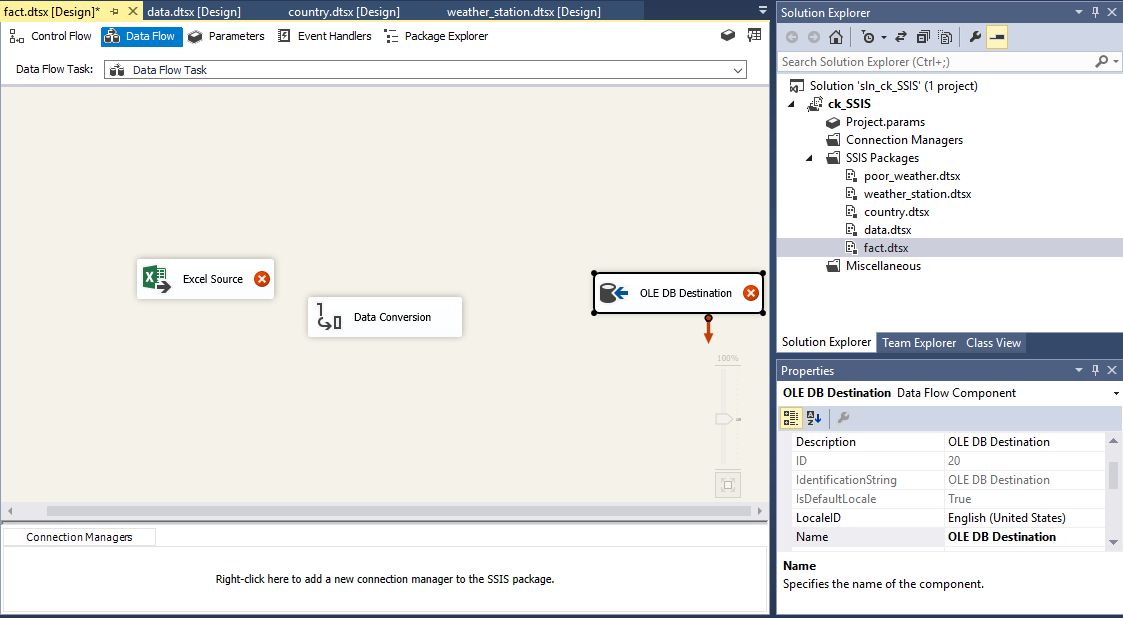


**- khởi chạy import dữ liệu**

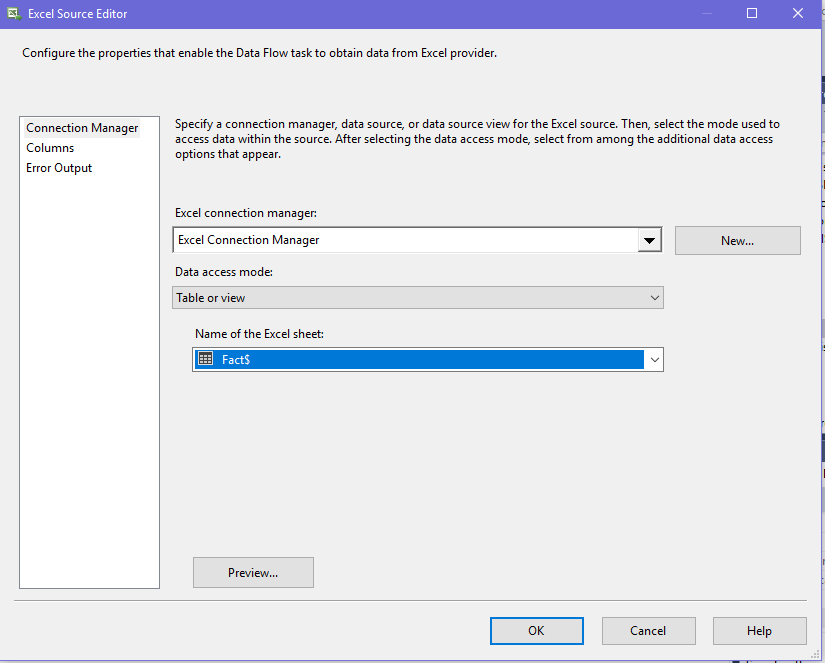


**2.5 import dữ liệu bảng fact.**

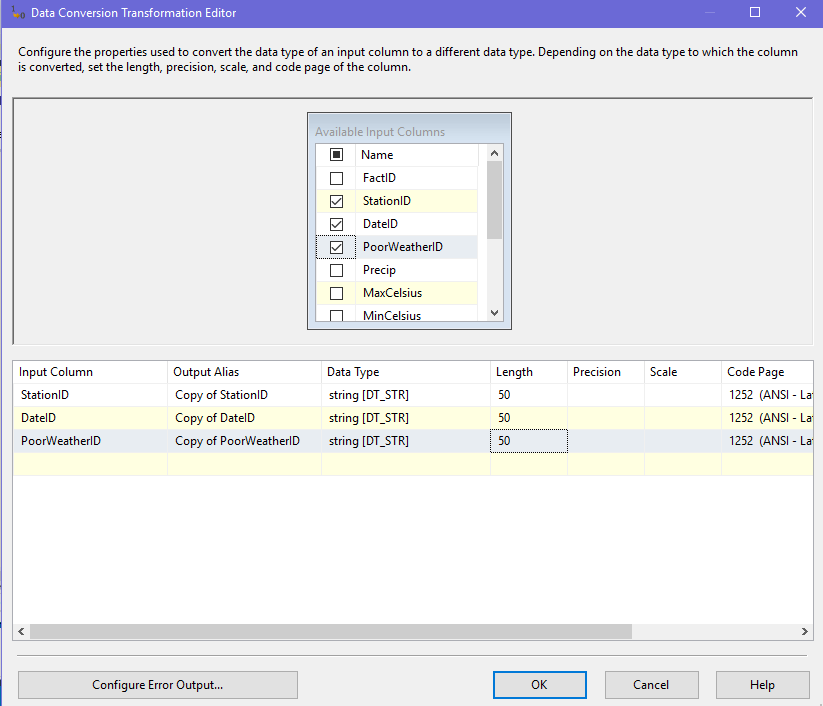
**- khởi tạo package fact**



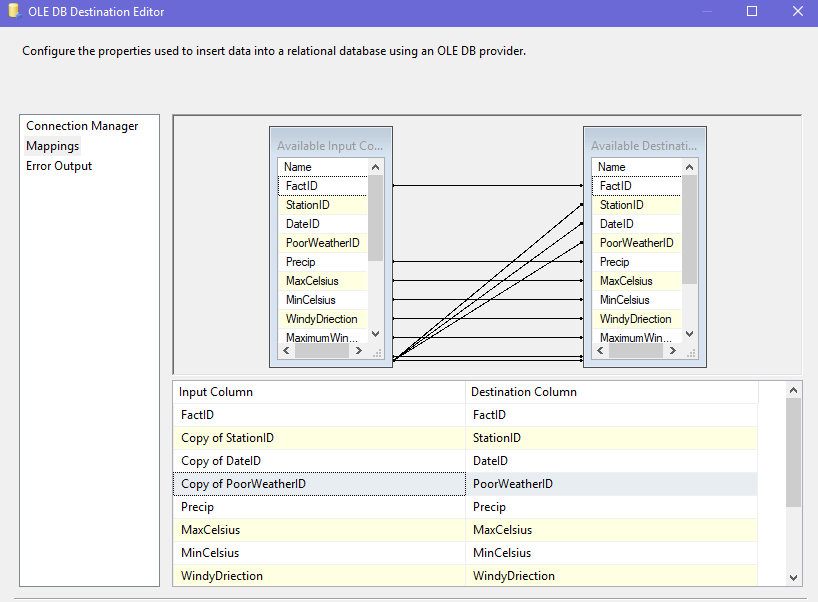
* **Chọn file excel nguồn và chọn sheet phù hợp**



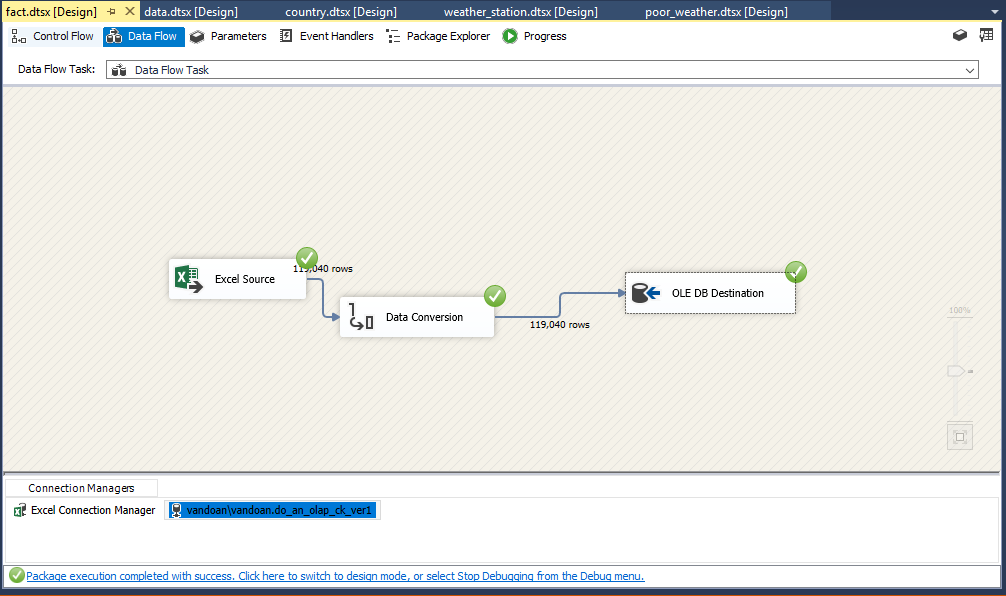
* **Dùng data conversion để chuyển đổi dữ liệu string**



* **Tiến hành mapping dữ liệu**

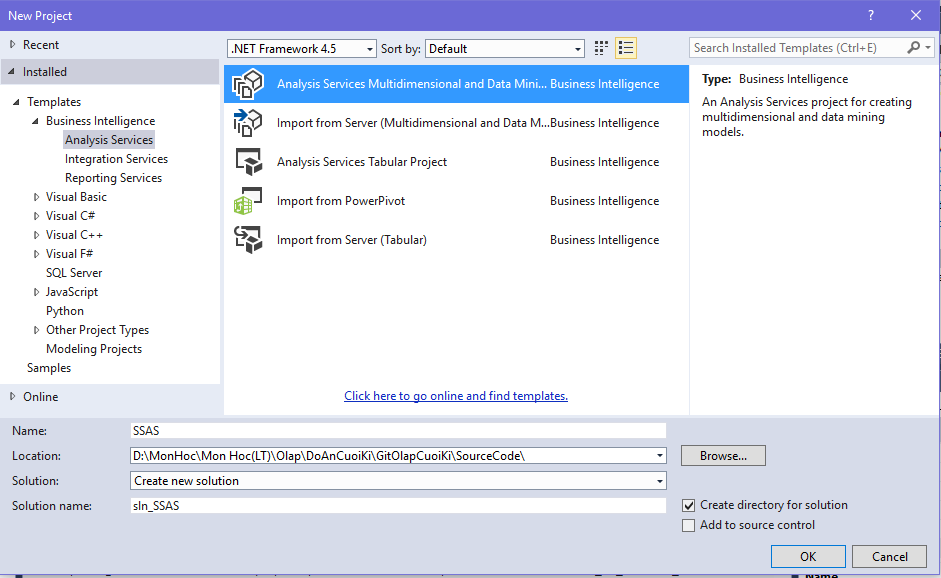


* **Tiến hành import dữ liệu**

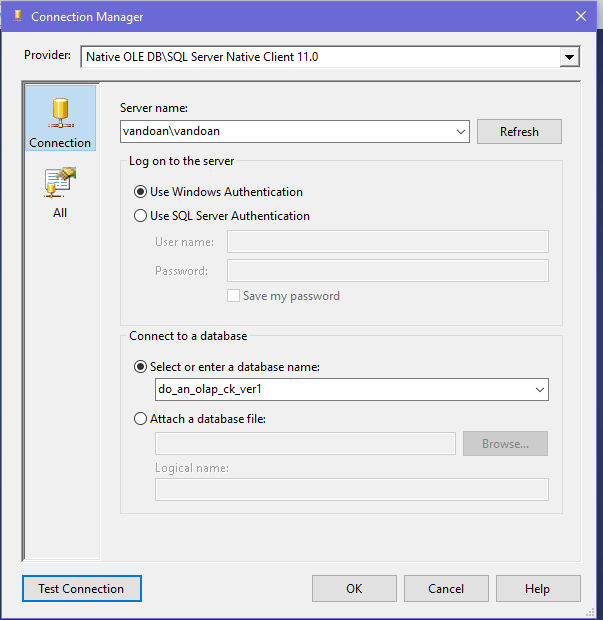


**3. quá trình SSAS**

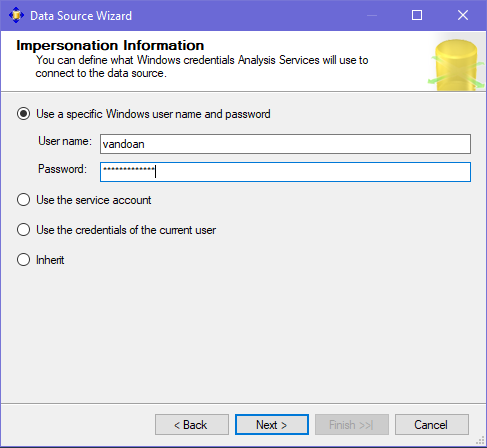
**- tạo mới project SSAS**



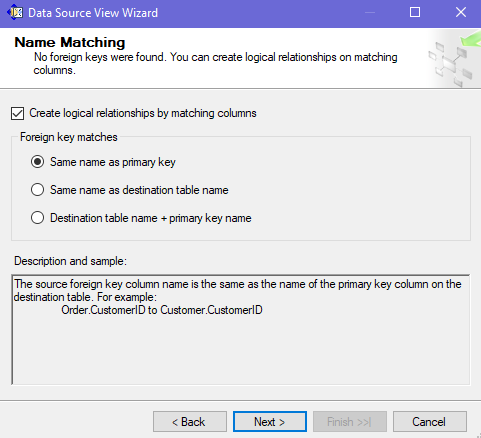
* **Tạo mới data source, chọn database sau khi đã import qua quá trình SSIS**



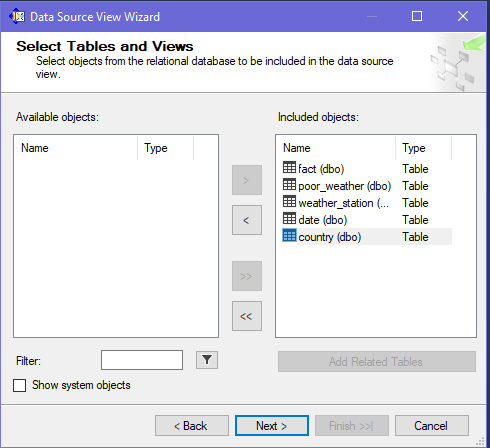
* **Chọn tài khoản windows để được cấp quyền truy cập database**



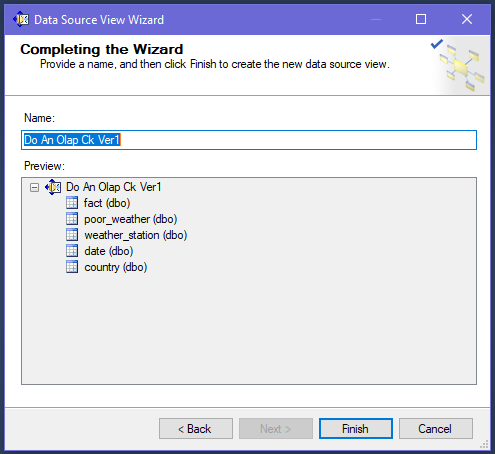
* **Tạo mới data source view**



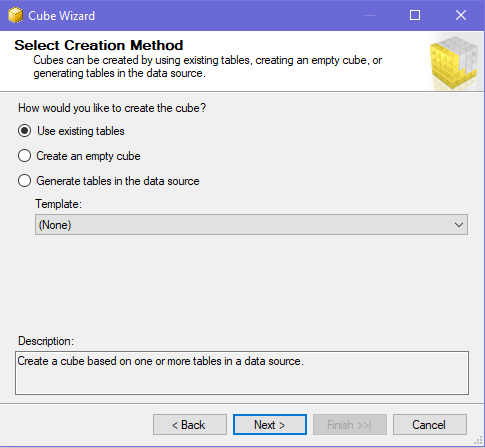
* **Tiến hành chọn table**



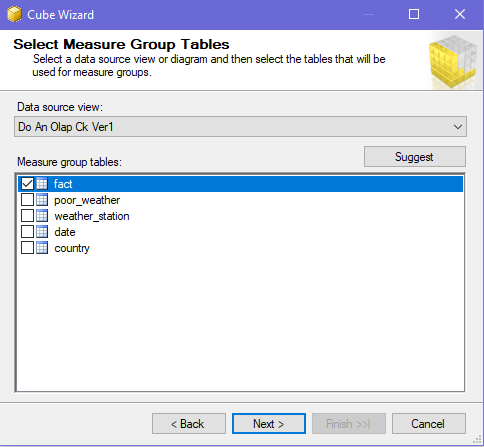
* **Kiểm tra và xác nhận hoàn tất tạo data source view**



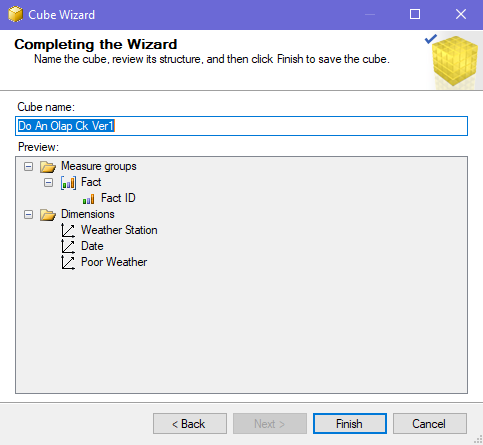
* **Tạo mới cube**



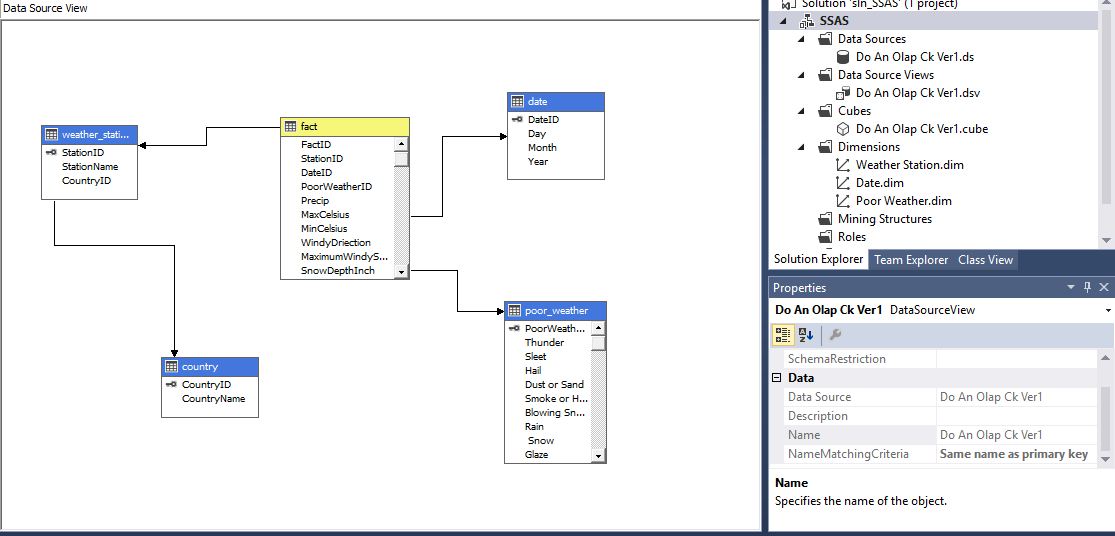
* **Chọn bảng để áp dụng độ đo**



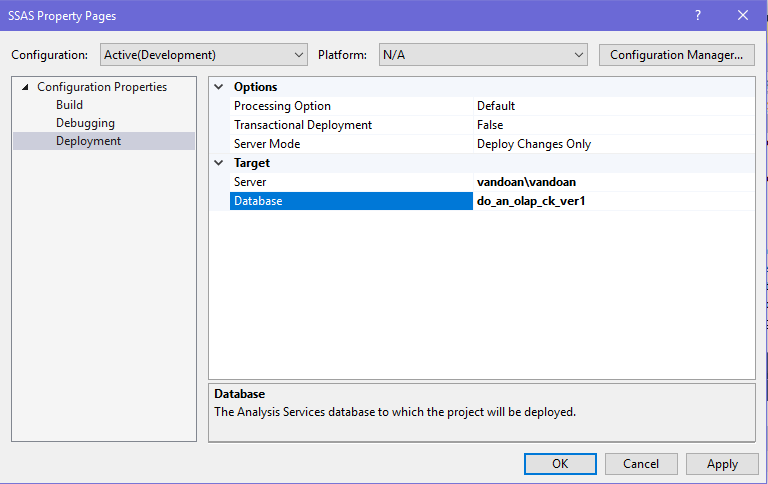
* **Chọn tên cube và xác nhận hoàn tất**



* **Kết quả sau khi tạo mới cube, đồng thời các dimesion tương ứng cũng được tạo ra**



* **Cấu hình server và database**



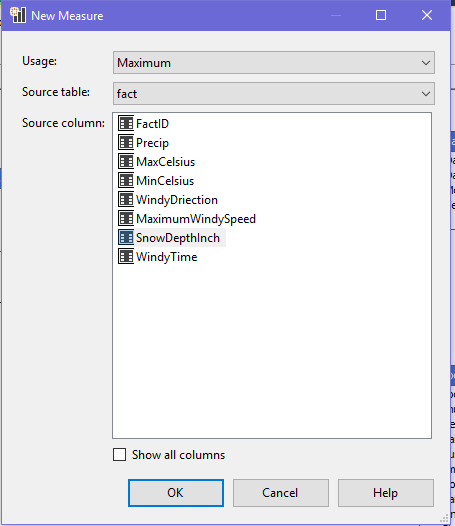
**4. Xử lí truy vấn với SSAS trên visual và SQL Server**

**4.1 Danh sách truy vấn**

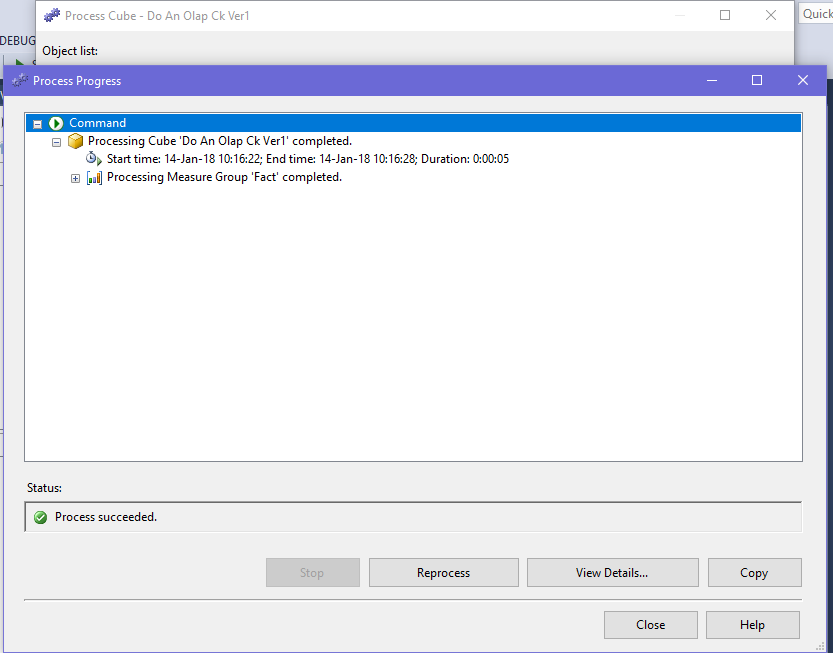
**a. Độ dày của tuyết sâu nhất trong tất cả các năm ở từng quốc gia.**

**- Trên visual studio**

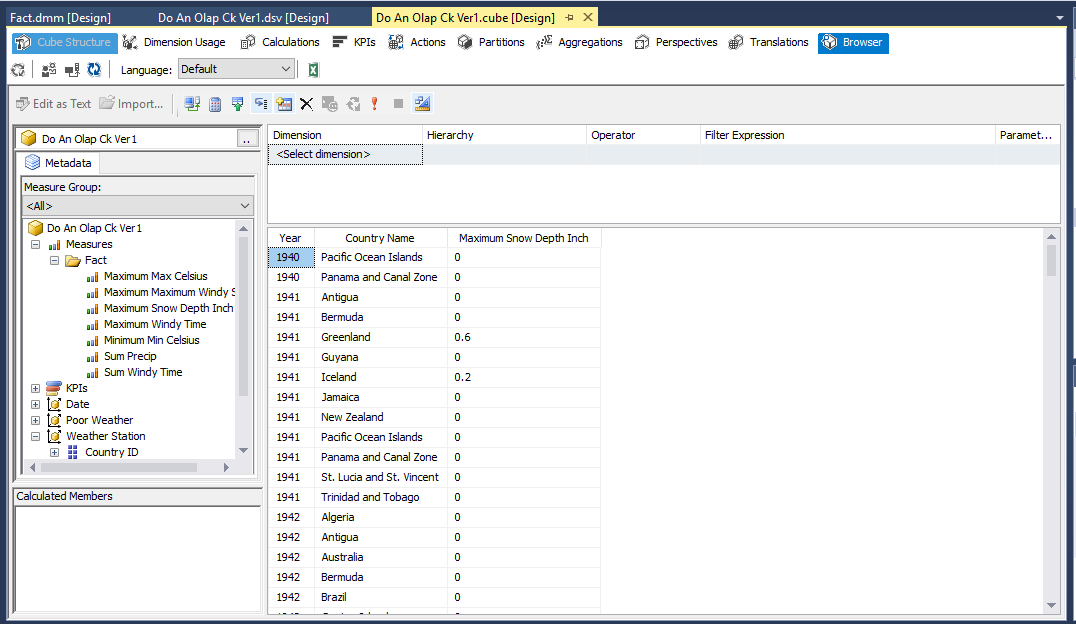
**+ Tạo mới measures Maximum Snow Depth Inch**



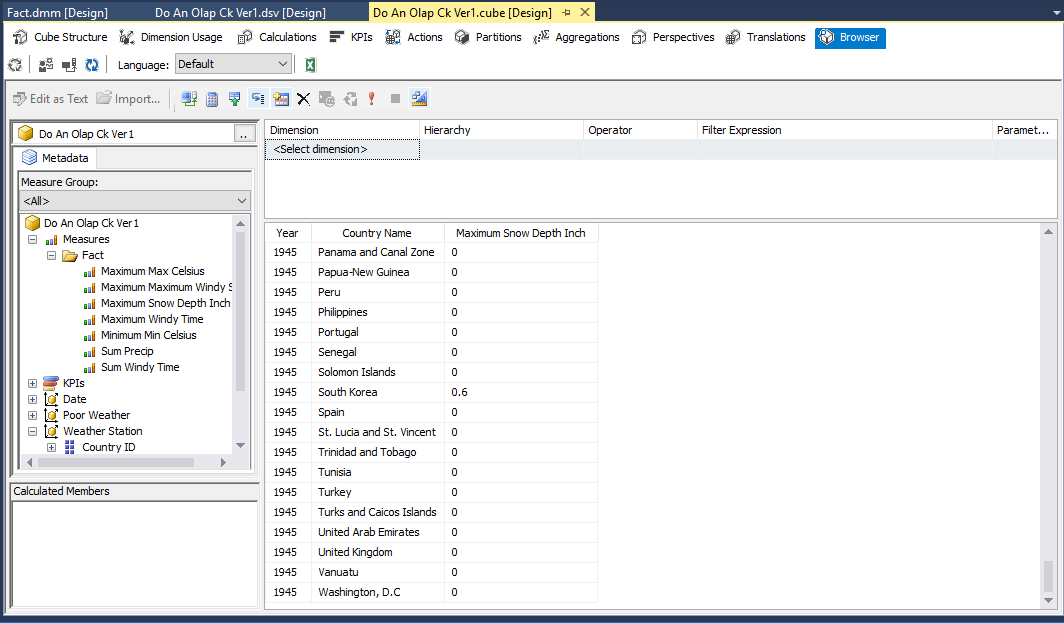
**+ Tiến hành chạy process cube để tạo measures**



**+Kết quả thực hiện thống kê tại visual trên tab brower (trang đầu tiên)**

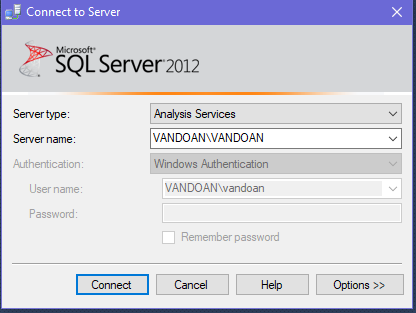


**+ Kết quả thực hiện thống kê trên visual tại tab brower(Trang cuối cùng).**

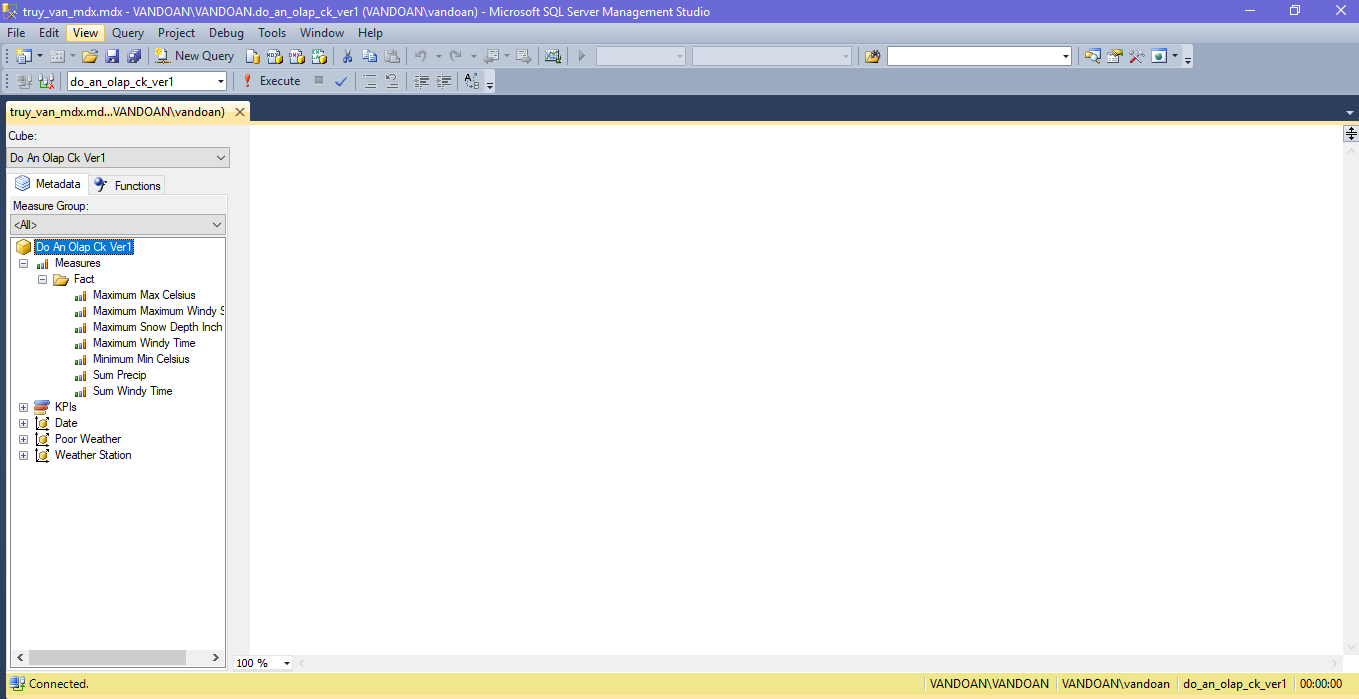


* **Thực hiện bằng truy vấn MDX trên microsoft SQL Server Analysis**

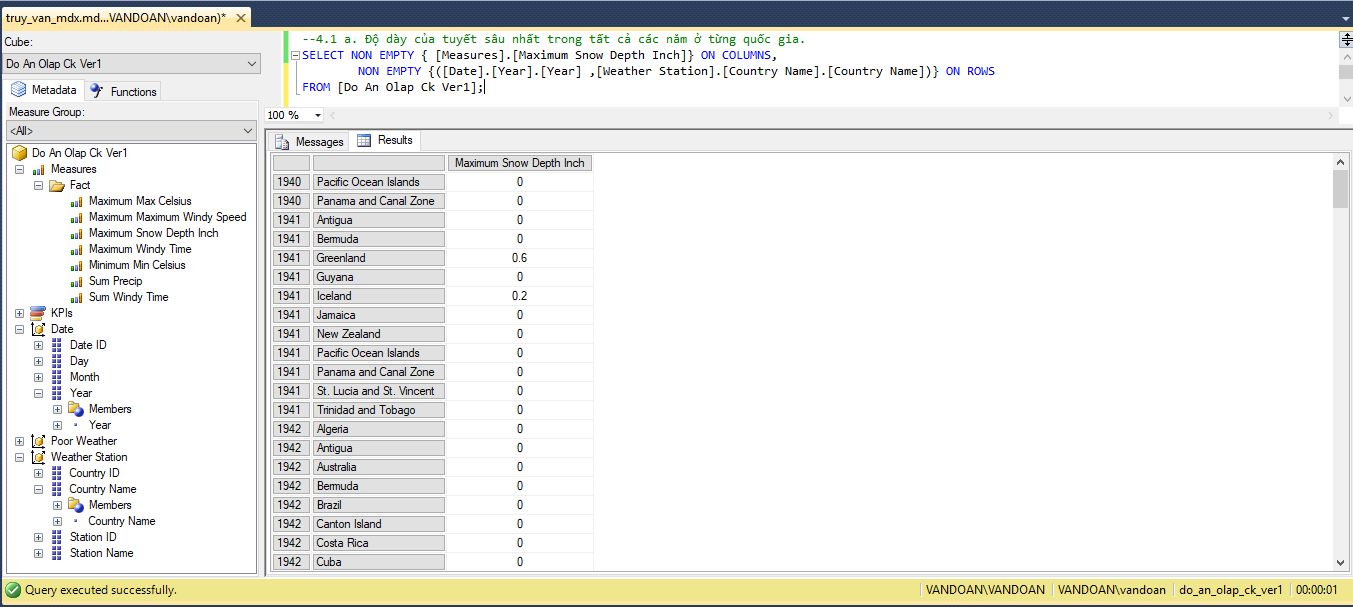
**+Khởi tạo kết nối trên SQL Server**



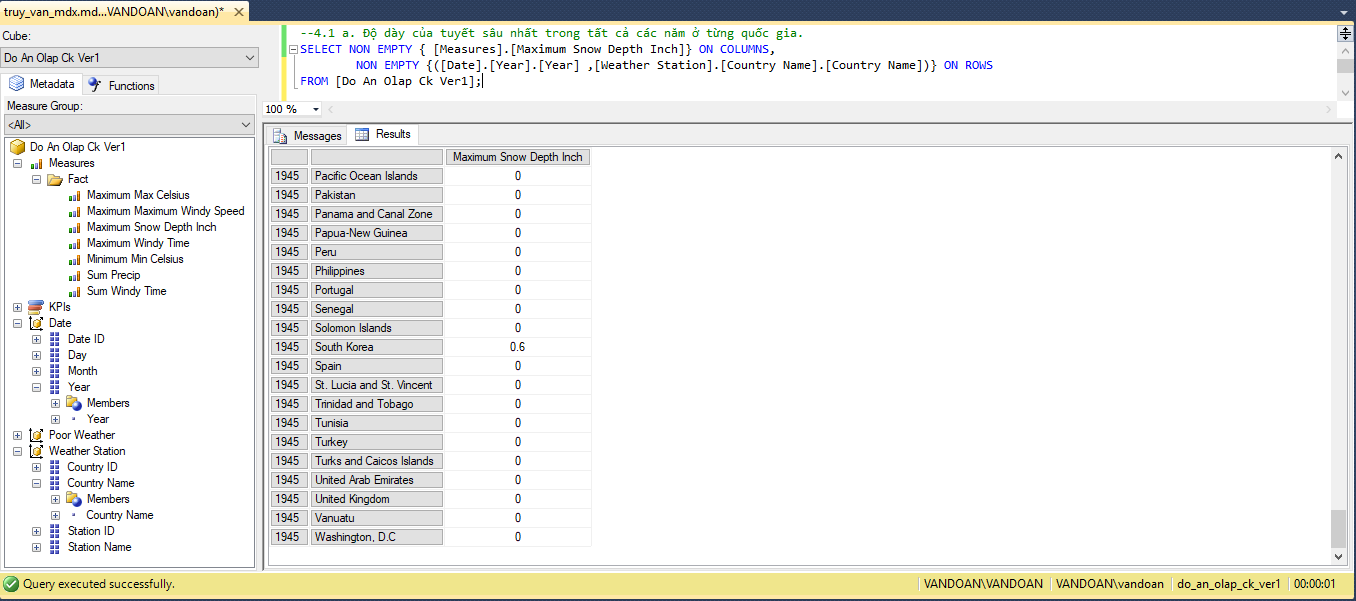
**+ Chọn database dùng để sử dụng MDX**



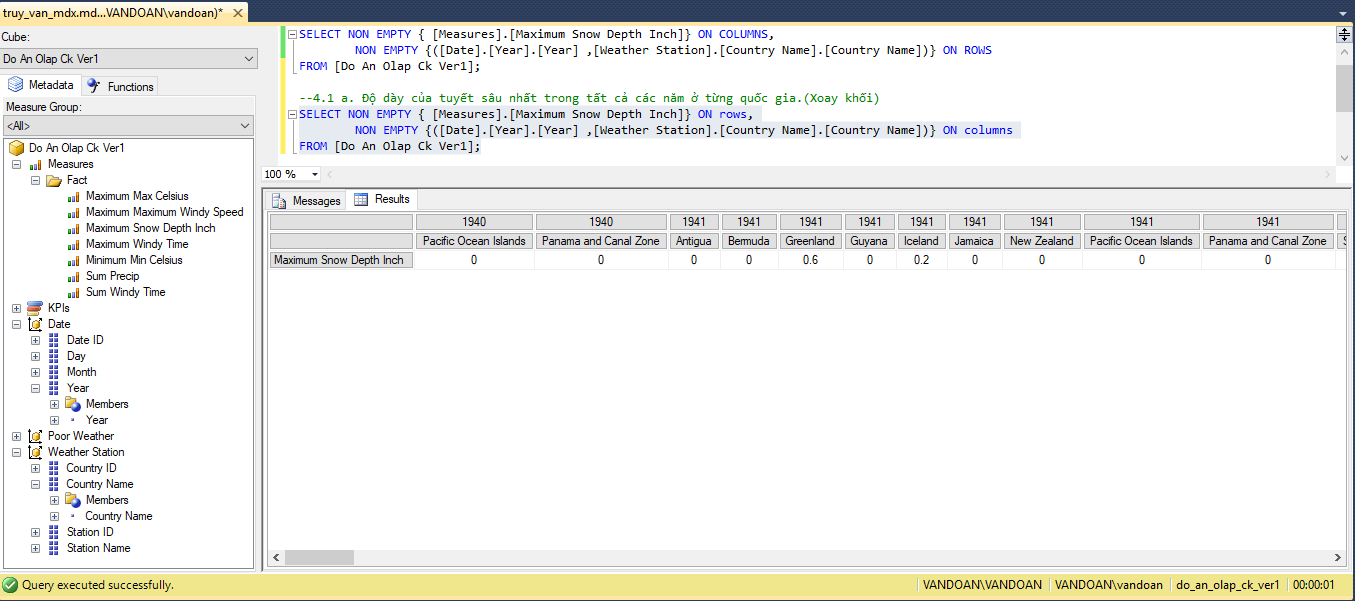
**+ Kết quả khi thực hiện truy vấn bằng ngôn ngữ MDX (trang đầu tiên)**



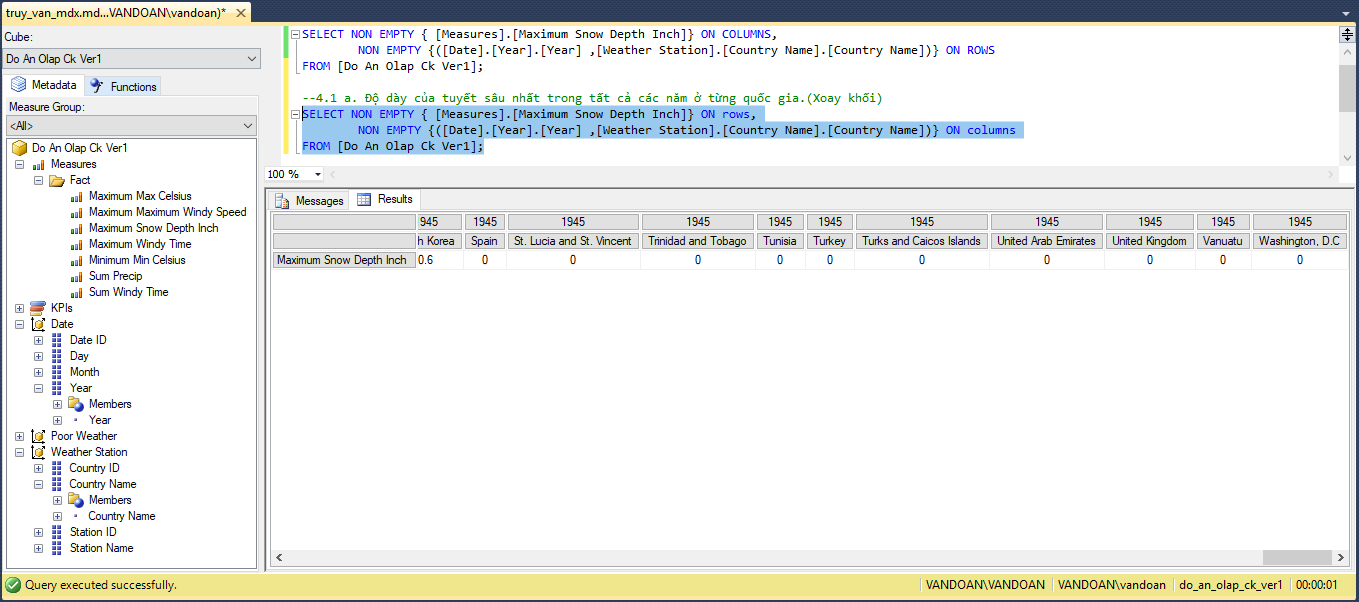
**+Kết quả truy vấn khi thực hiện bằng ngôn ngữ MDX(trang cuối cùng).**



**+ Xoay khổi dữ liệu(Pivot-Trang đầu tiên)**



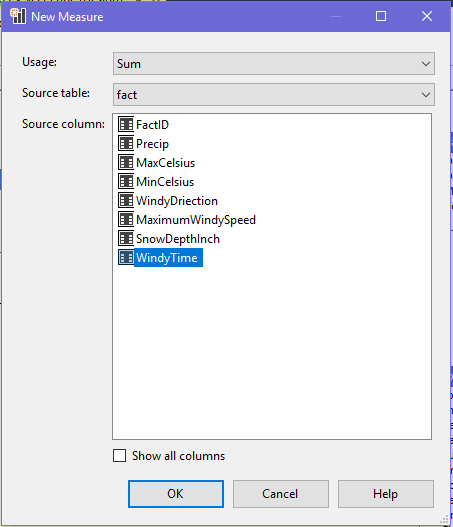
**+Xoay khối dữ liệu(Pivot-Trang cuối)**



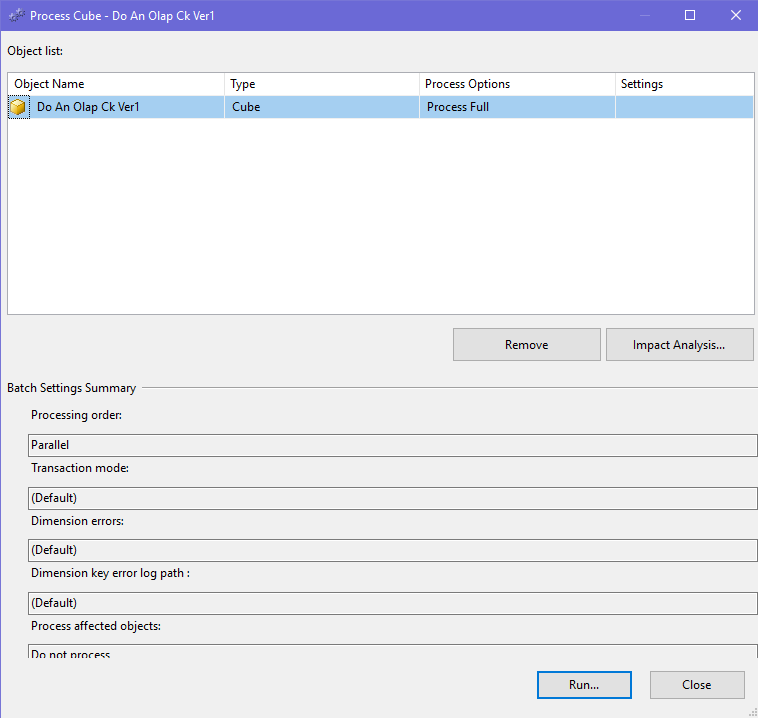
**b. Tổng thời gian gió thổi ở tất cả các quốc gia trong năm 1942.**

**- Trên visual studio**

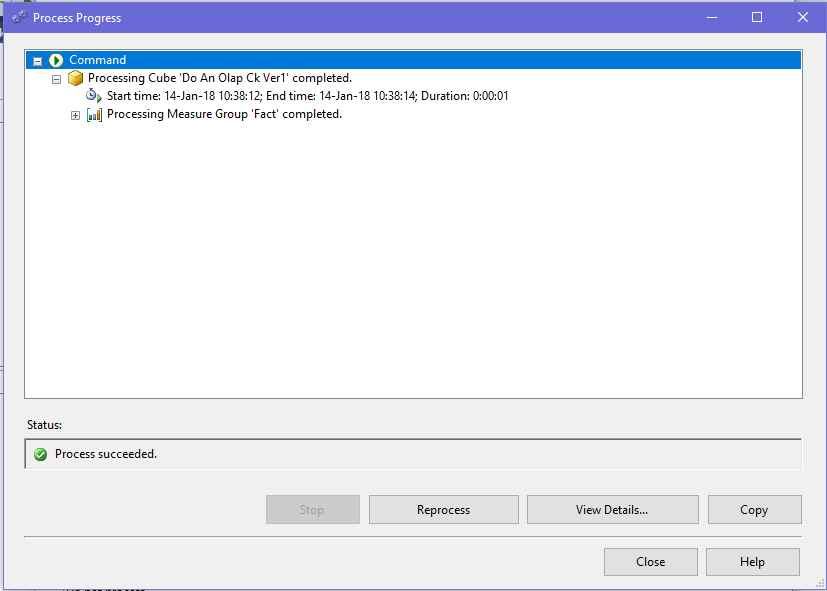
**+ Khởi tạo measures Sum Windy Time**



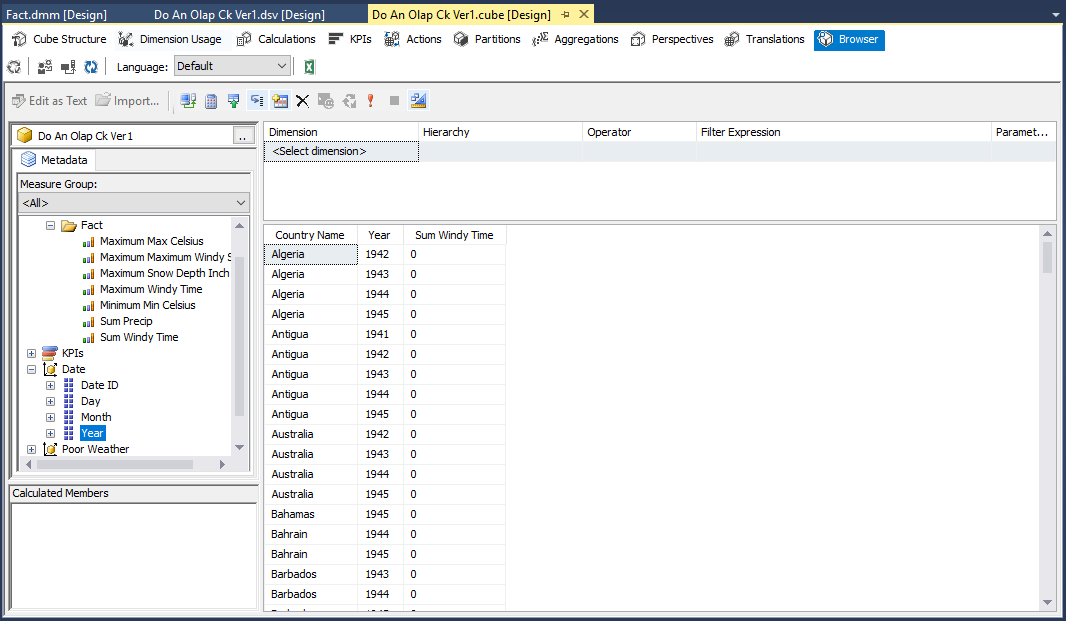
**+ Tiến hành khởi chạy để tạo measures**



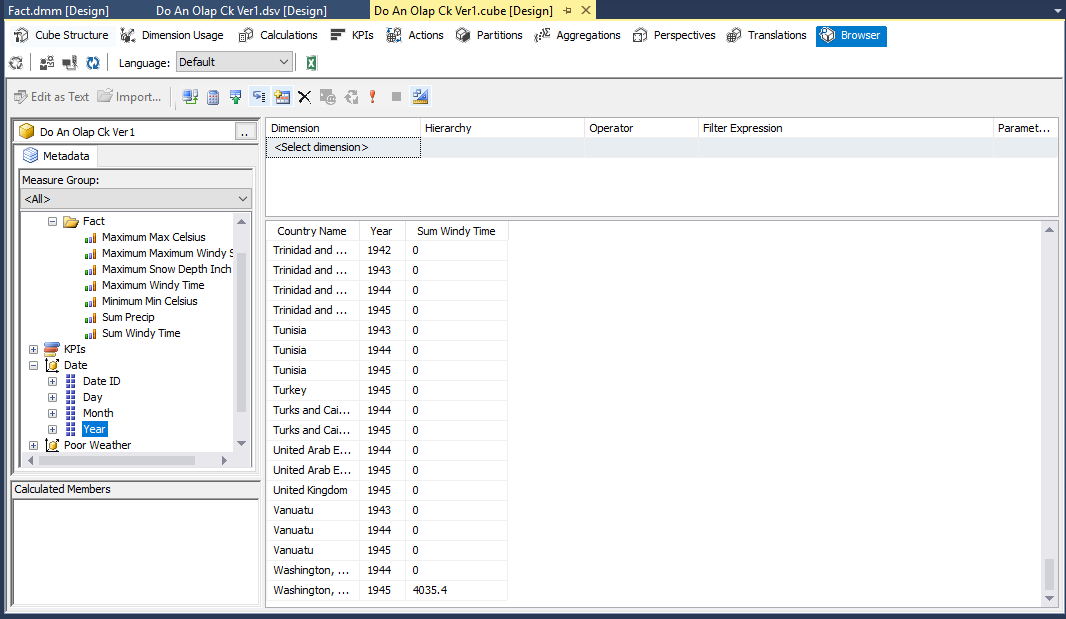
**+ Kết quả thông báo thành công.**



**+ Kết quả thực hiện truy vấn trên visual studio(Trang đầu tiên)**

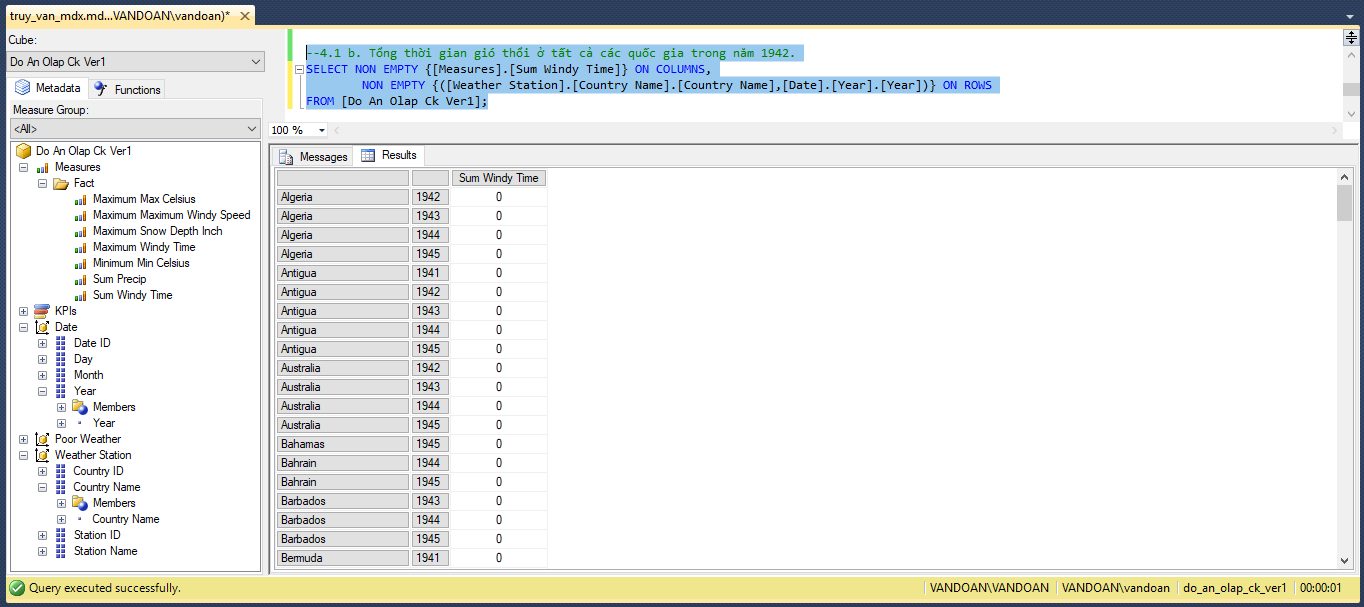


**+Kết quả truy vấn khi thực hiện trên visual studio(Trang cuối cùng)**

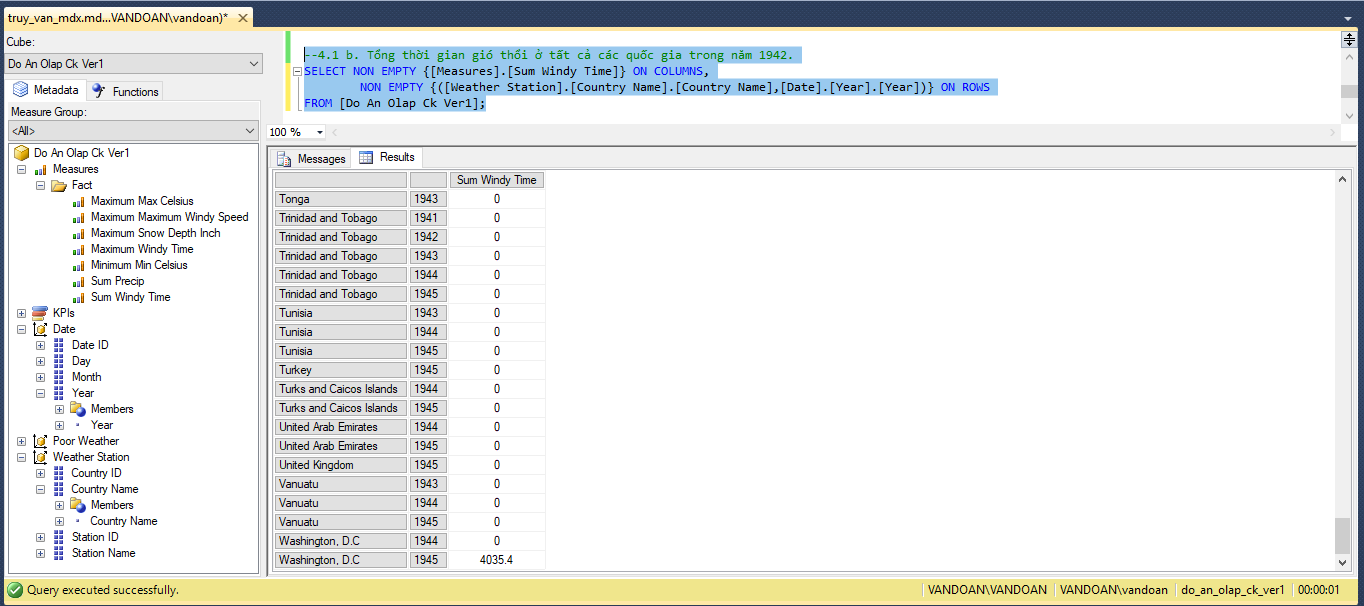


* **Thực hiện truy vấn trên SQL Server bằng ngôn ngữ MDX.**

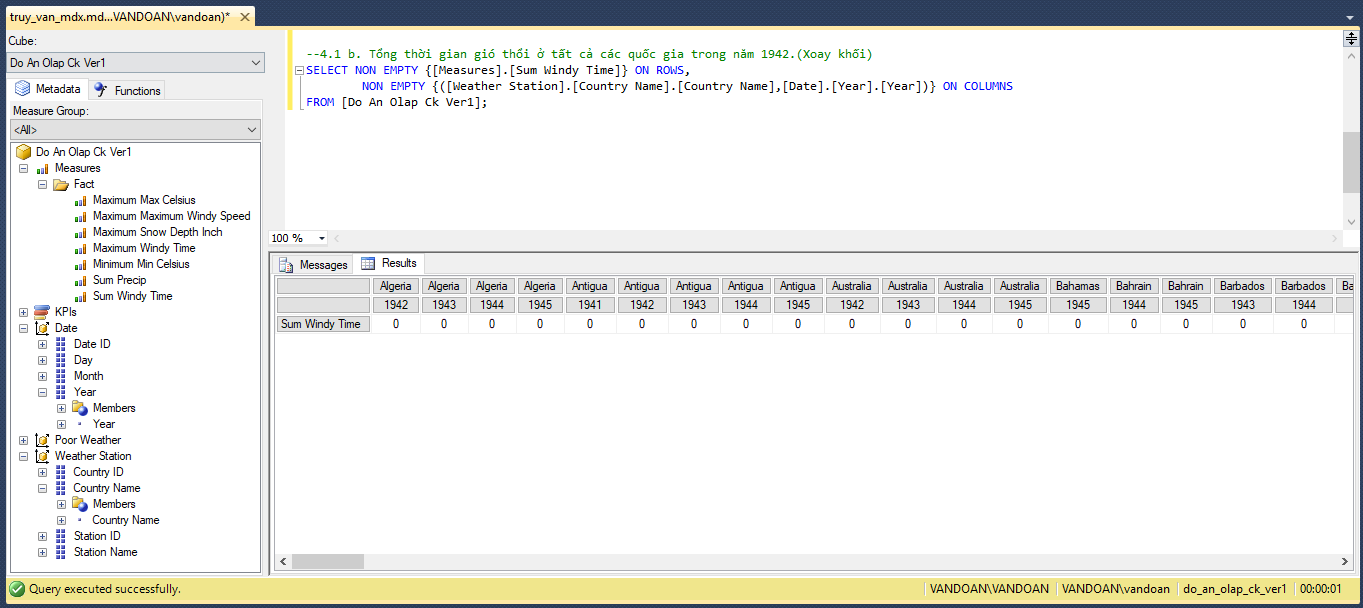
**+ Kết quả khi thực hiện truy vấn bằng ngôn ngữ MDX(trang đầu tiên)**



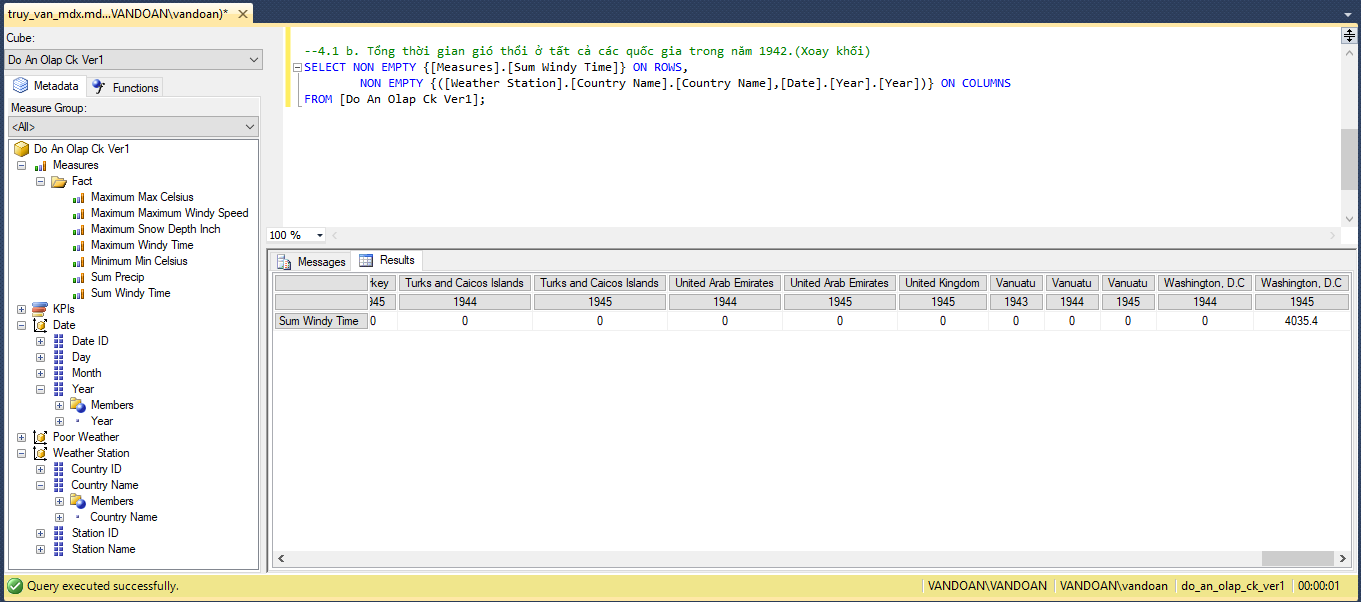
**+Kết quả khi thực hiện truy vấn bằng ngôn ngữ MDX(trang cuối cùng).**



**+Xoay khối dữ liệu(Pivot-Trang đầu)**



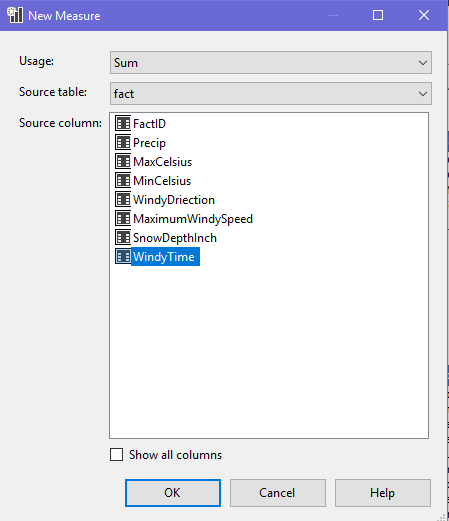
**+ Xoay khối dữ liệu(Pivot-Trang cuối)**



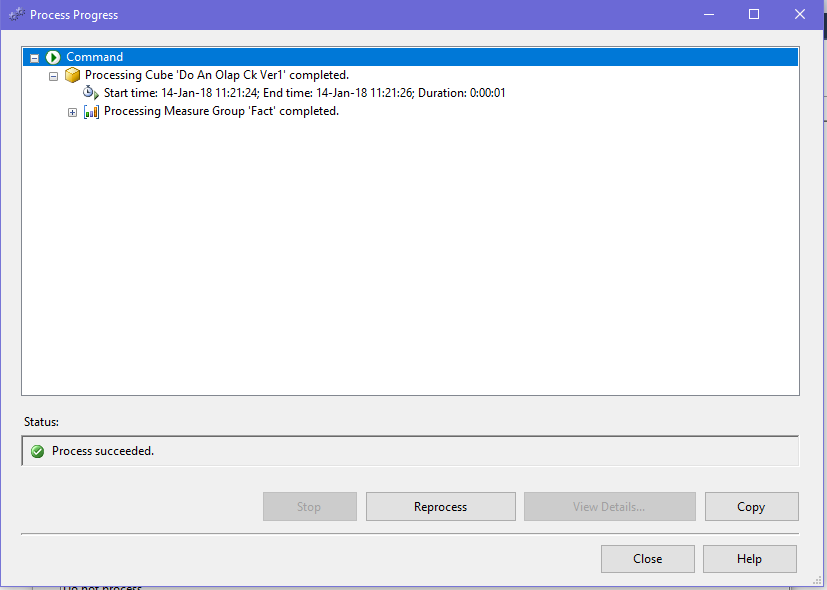
**c. Tổng thời gian gió thổi theo từng trạm đo trong từng tháng với thời tiết có sấm sét.**

**- Trên visual studio**

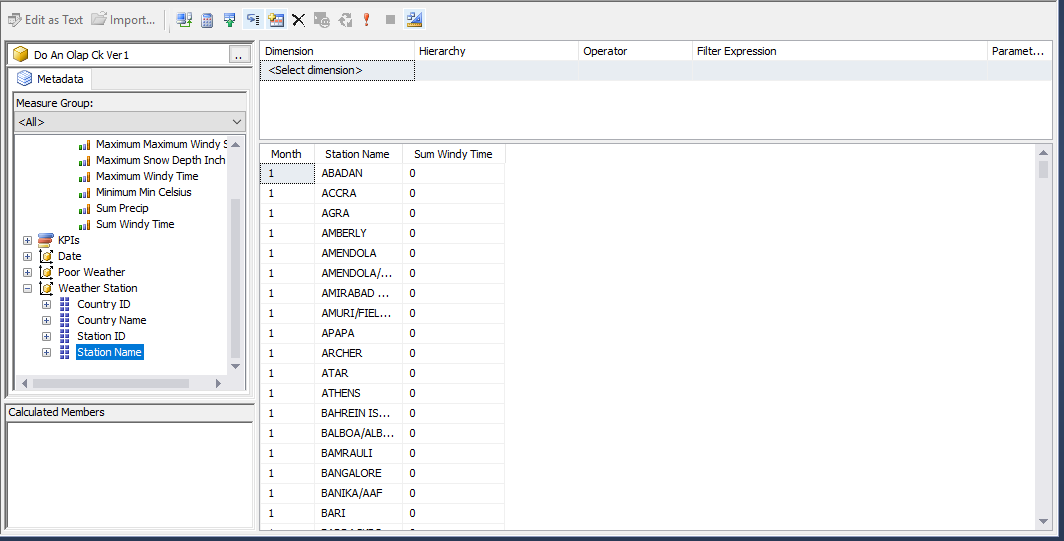
**+ Khởi tạo measures Sum Windy Time**



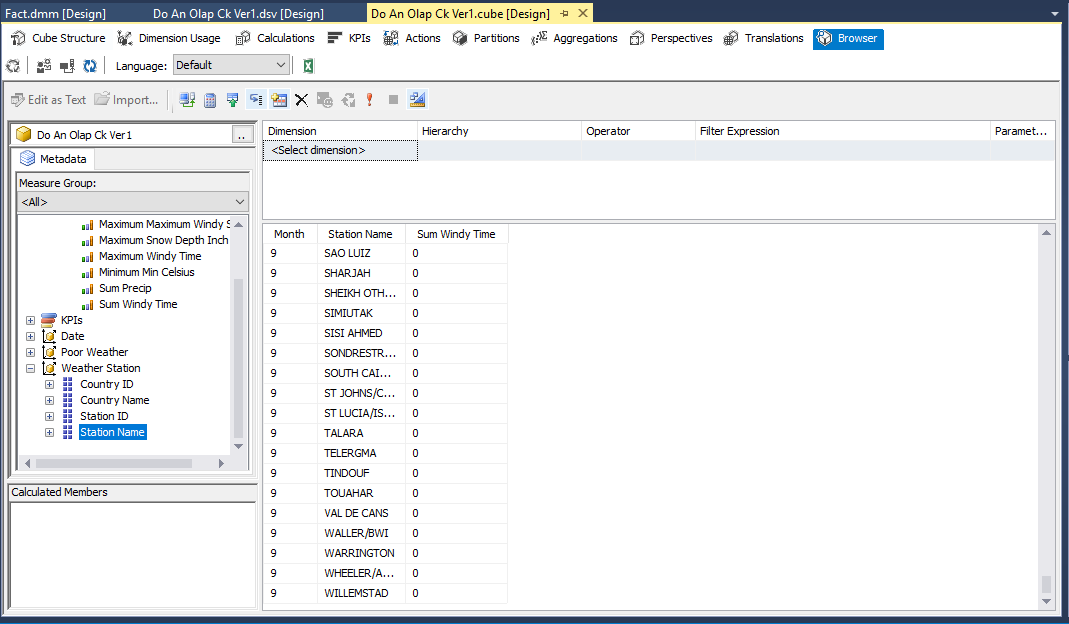
**+Khởi chạy process trên cube**



**+Kết quả khi thực hiện truy vấn trên visual(Trang đầu).**

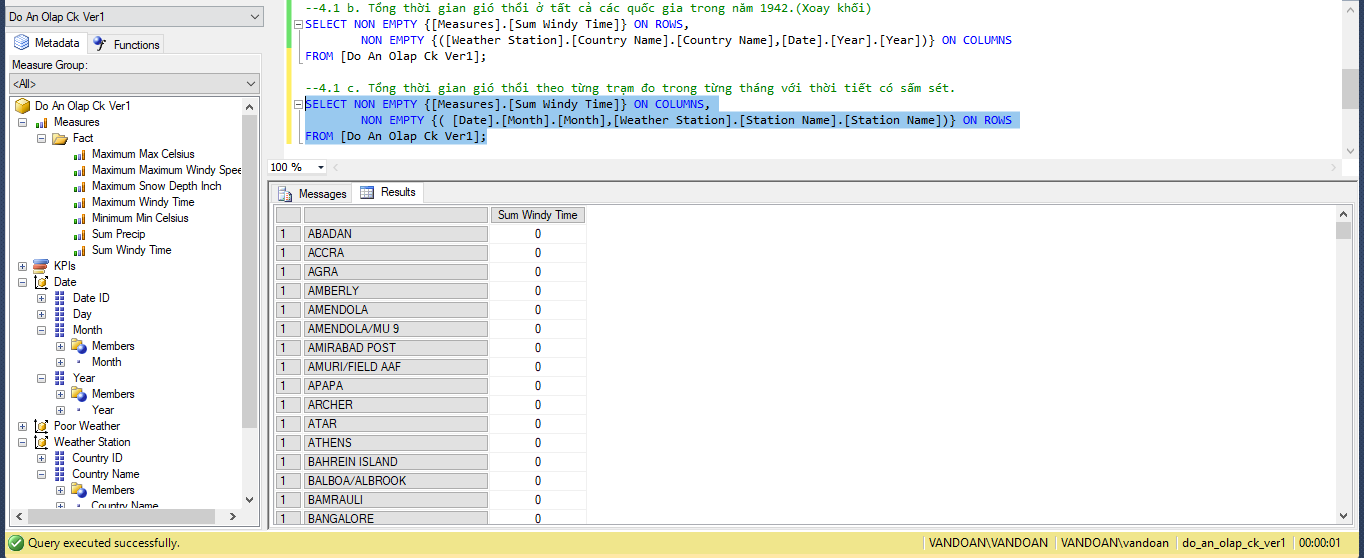


**+Kết quả truy vấn khi thực hiện trên visual(Trang cuối)**

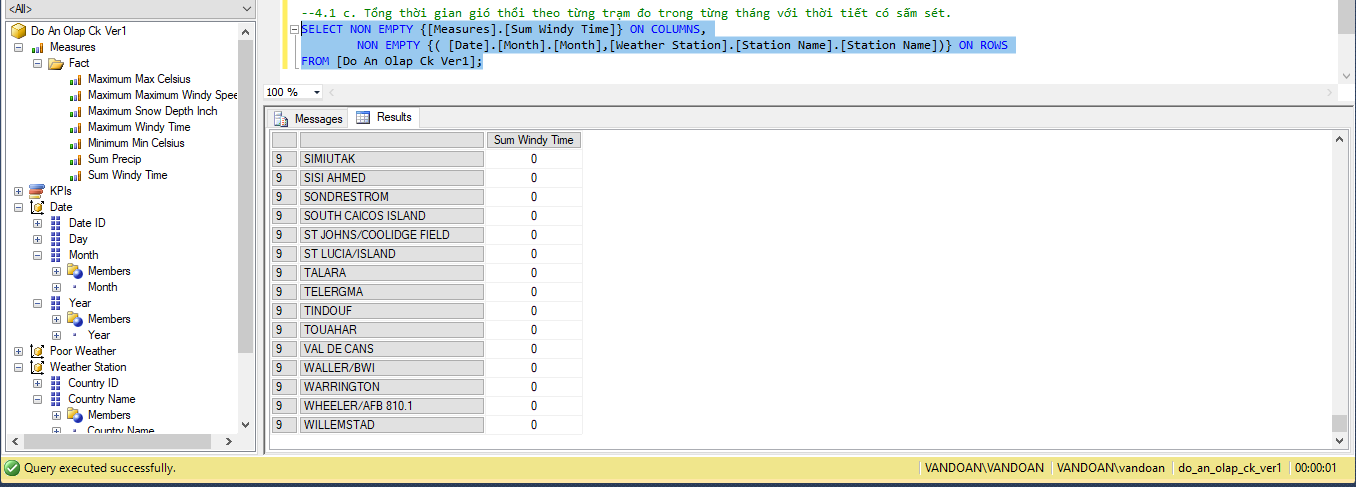


* **Truy vấn MDX trên SQL Server**

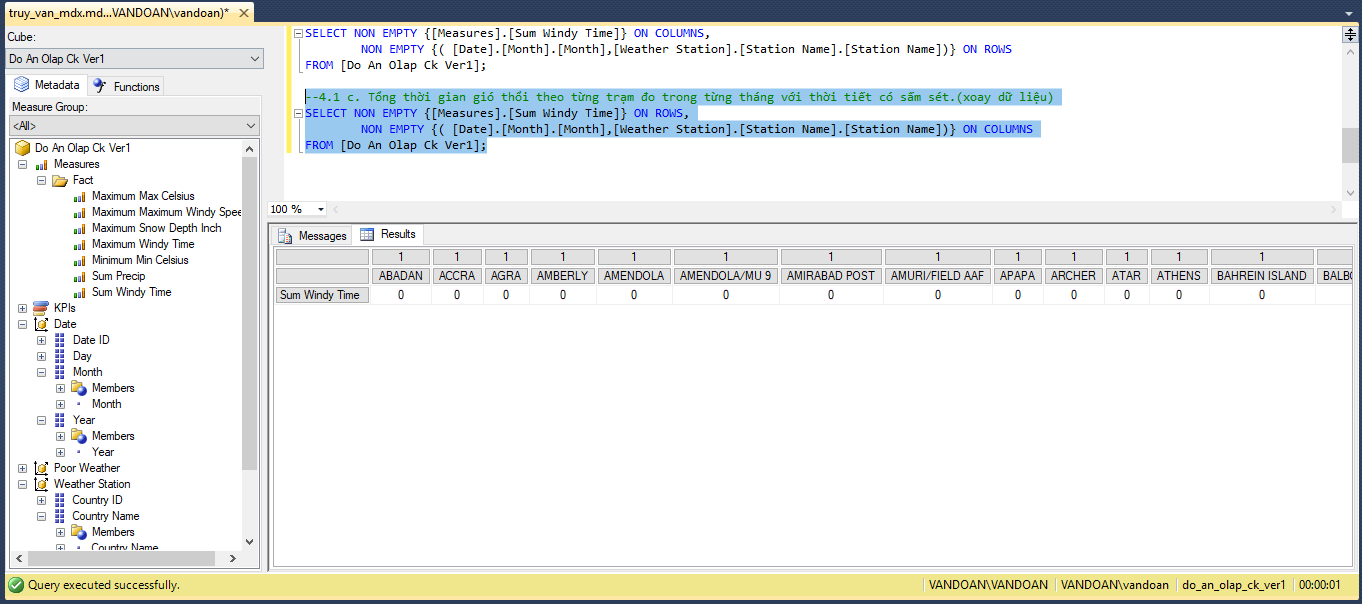
**+Kết quả truy vấn khi thực hiện bằng MDX(trang đầu)**



**+Kết quả truy vấn khi thực hiện bằng MDX(trang cuối)**



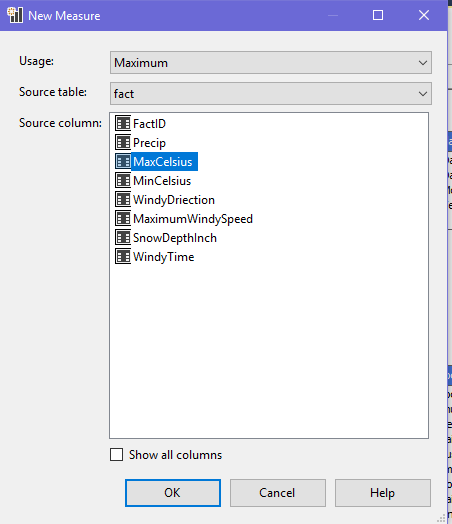
**+Xoay dữ liệu(pivot-trang)**



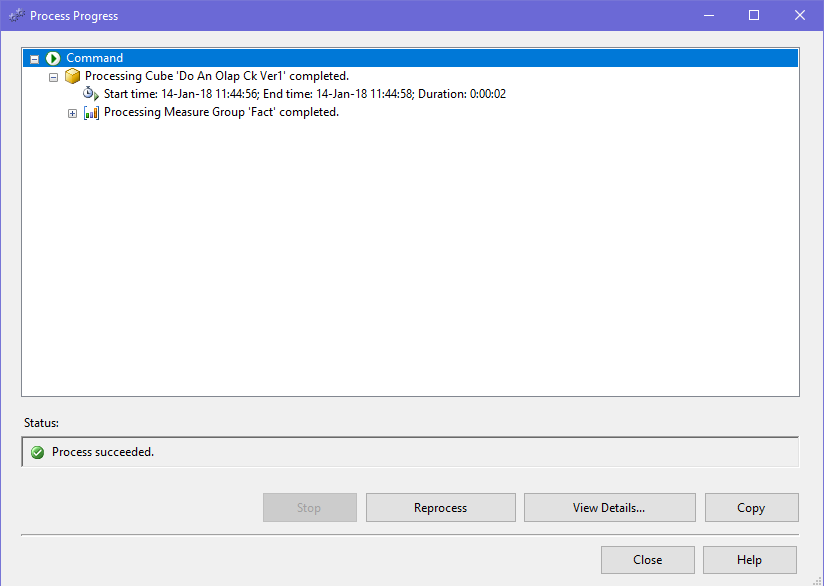
**d. Top 5 quốc gia có nhiệt độ cao nhất kèm theo thời tiết smoke and haze.**

**- Trên visual**

**+ Tạo measures Maximum Max Celcius**

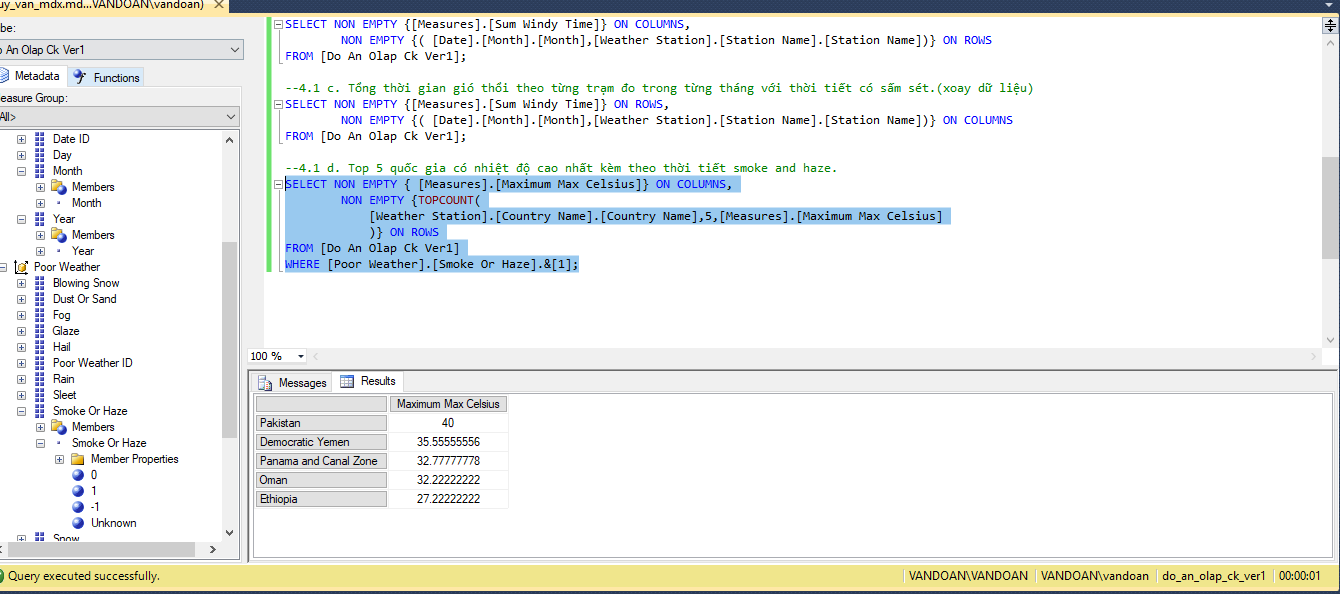


**+Khởi chạy process trên cube**

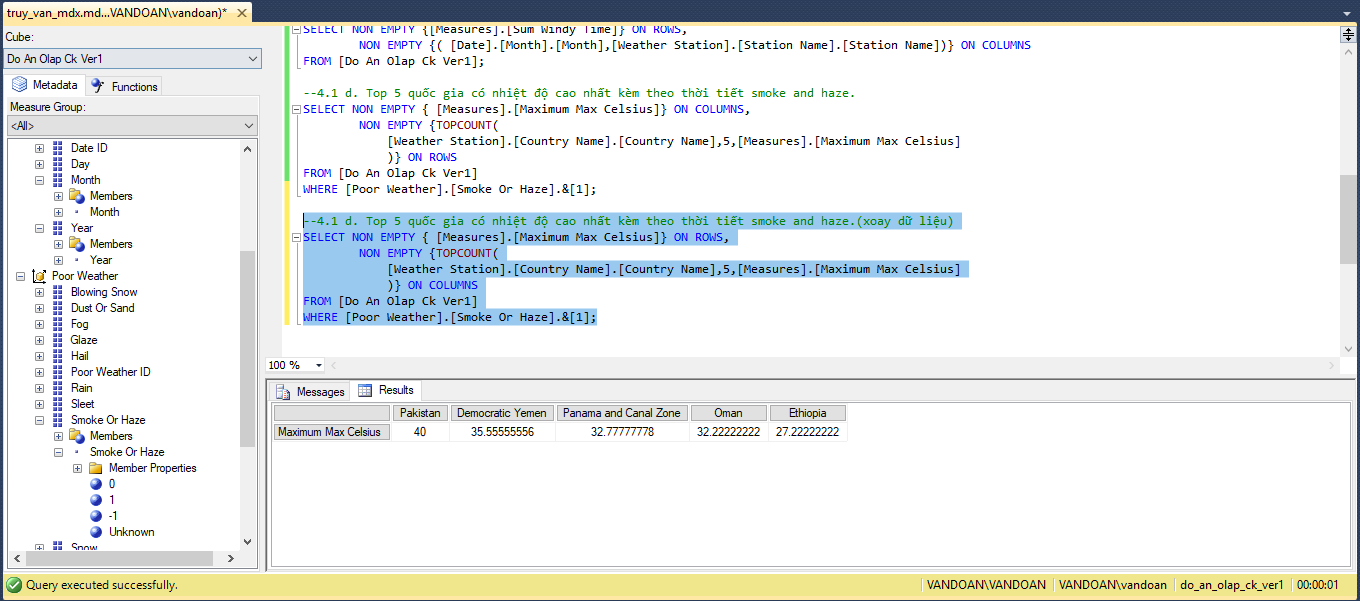


**Vì truy vấn này sử dụng hàm top count nên ta sẽ thực hiện lấy kết quả bằng ngôn ngữ truy vấn MDX**

* **Truy vấn bằng ngôn ngữ MDX**



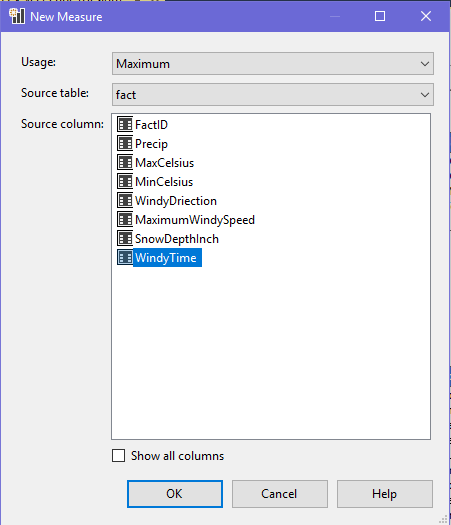
**+ Xoay khối dữ liệu(Pivot)**



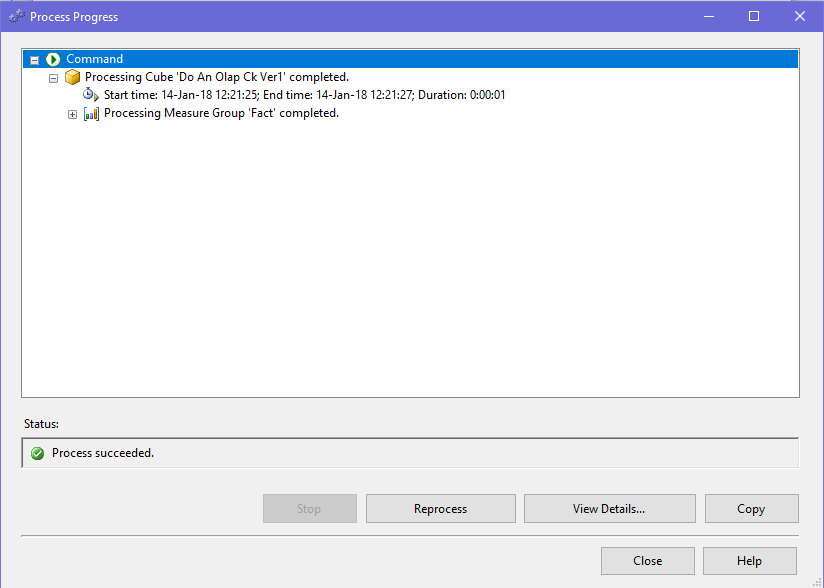
**e.Tốc độ gió thổi và thời gian gió thổi lớn nhất của quốc gia Democratic Yemen trong tất cả các tháng, thời gian tăng dần.**

**- trên visual**

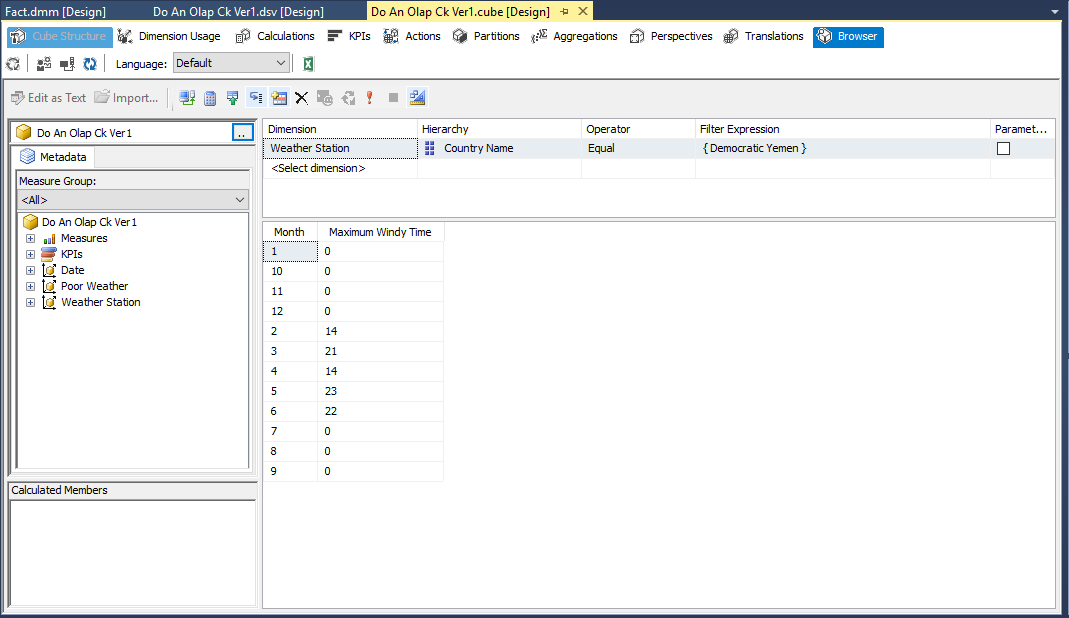
**+Khởi tạo measures**



**+Khởi chạy process trên cube**

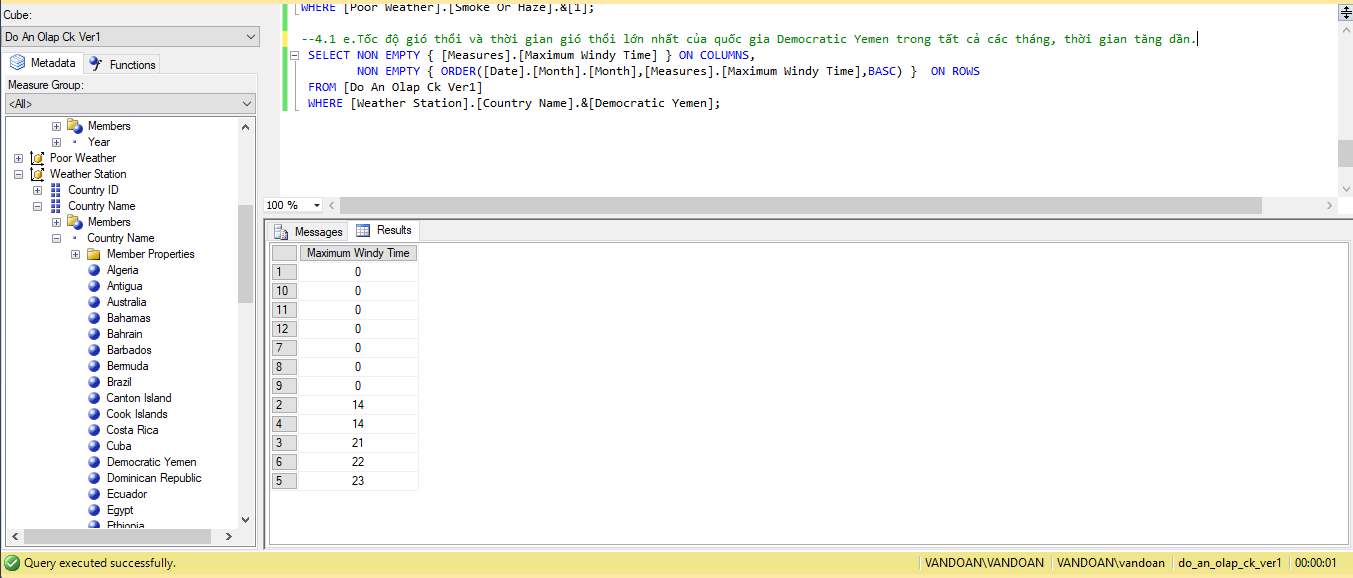


**+Kết quả thực hiện truy vấn trên visual**

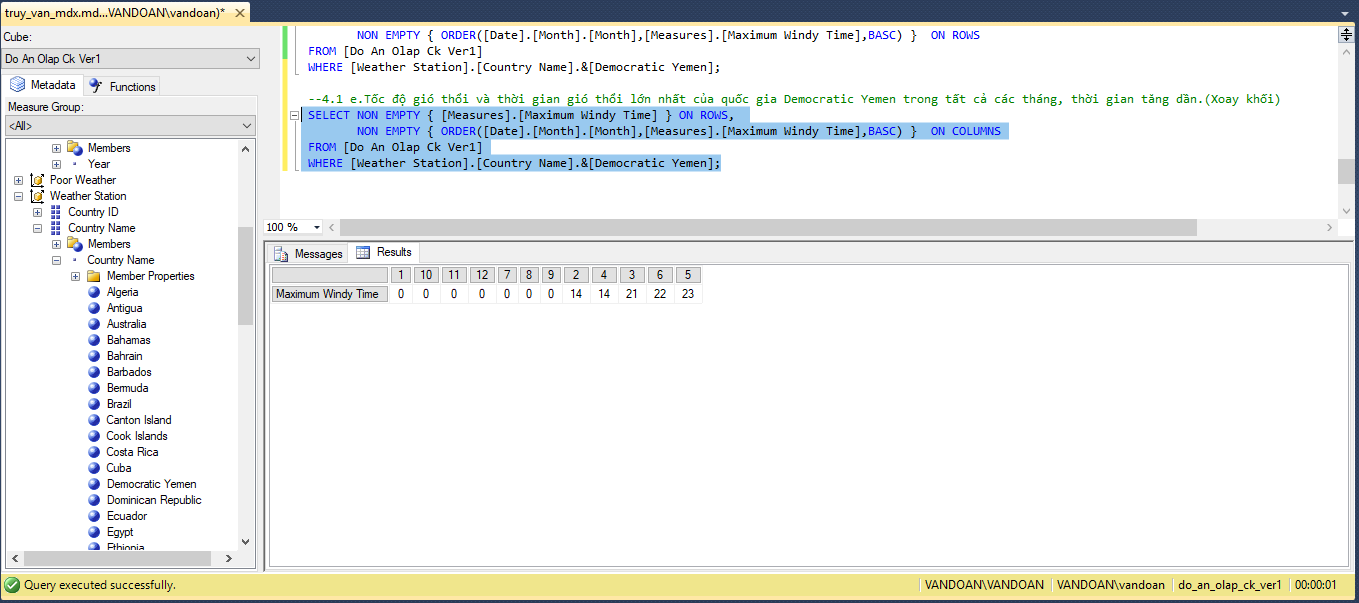


* **Thực hiện truy vấn trên SQL Server bằng MDX**

**+Kết quả truy vấn khi thực hiện bằng ngôn ngữ MDX**



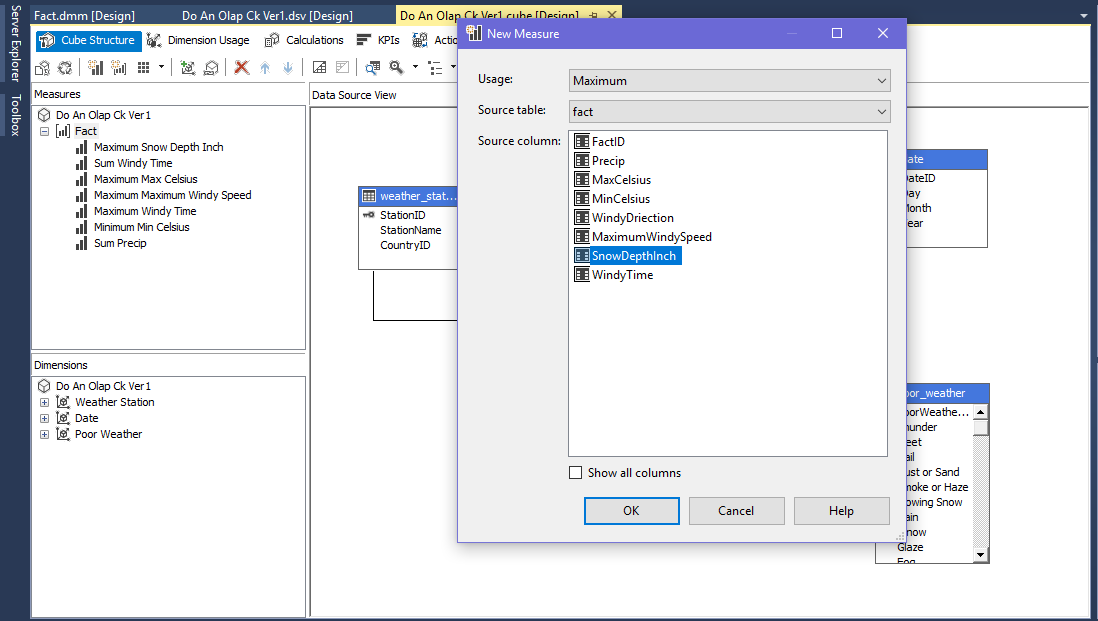
**+Xoay khối dữ liệu**



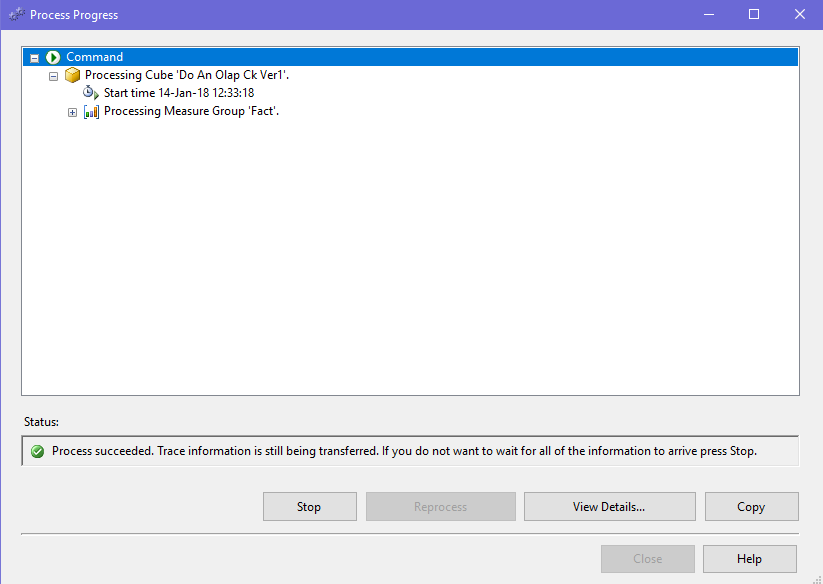
**f.Độ sâu lớn nhất của tuyết kèm theo điều kiện thời tiết có sấm set tại trạm đo simiutak trong từng tháng của năm.**

**- Thực hiện trên visual**

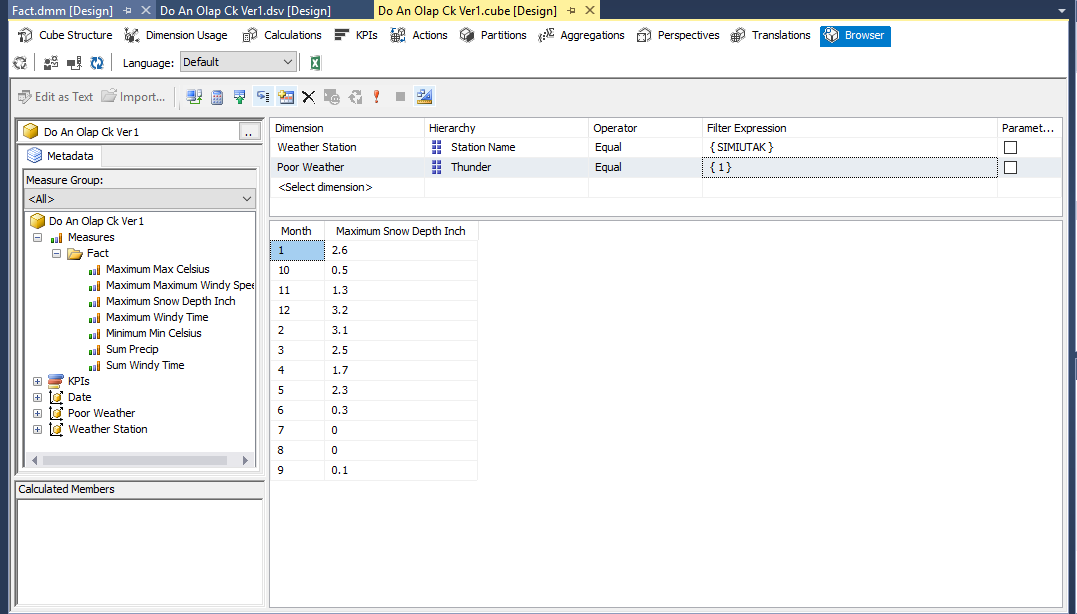
**+ Khởi tạo measure Maximum Snow Depth Inch**



**+Run process để tạo mesures**

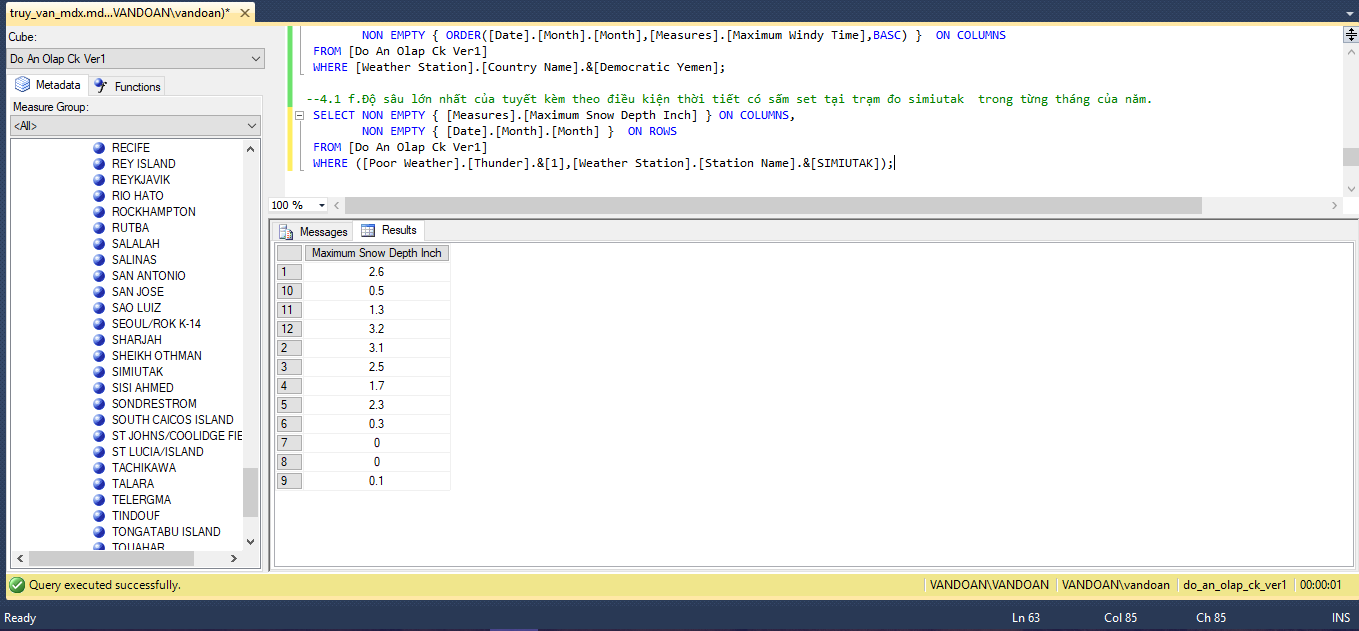


**+Kết quả khi thực hiện truy vấn trên visual**

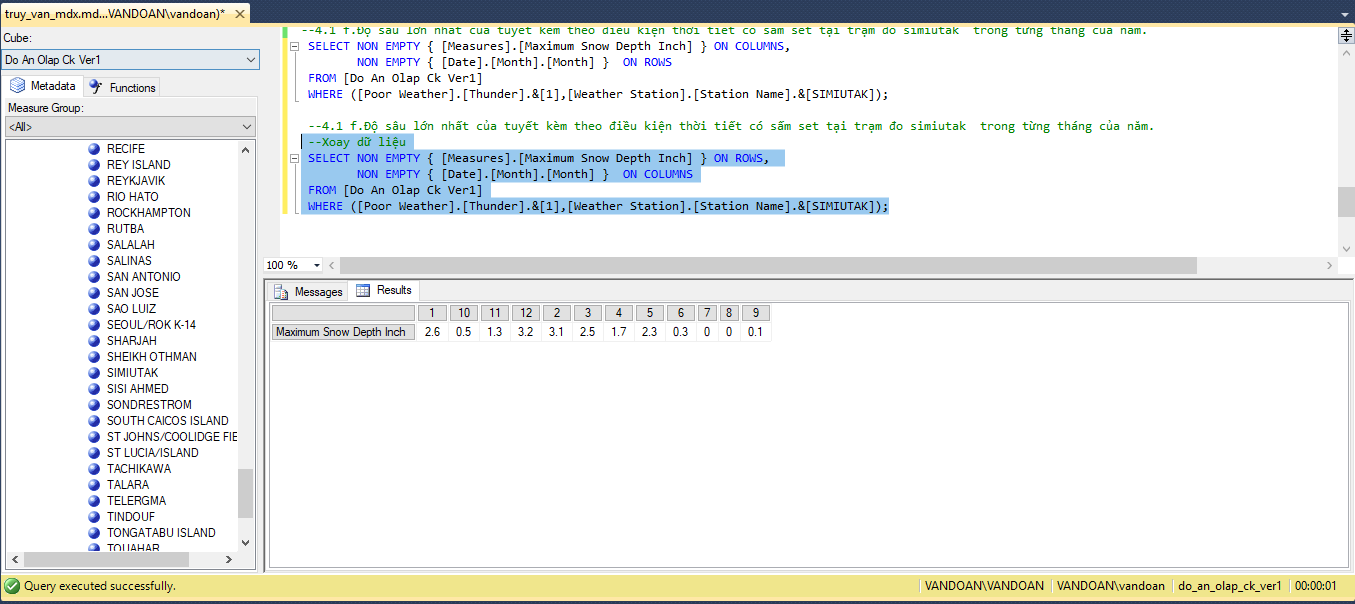


* **Truy vấn bằng MDX trên SQL Server**

**+Kết quả thực hiện truy vấn**

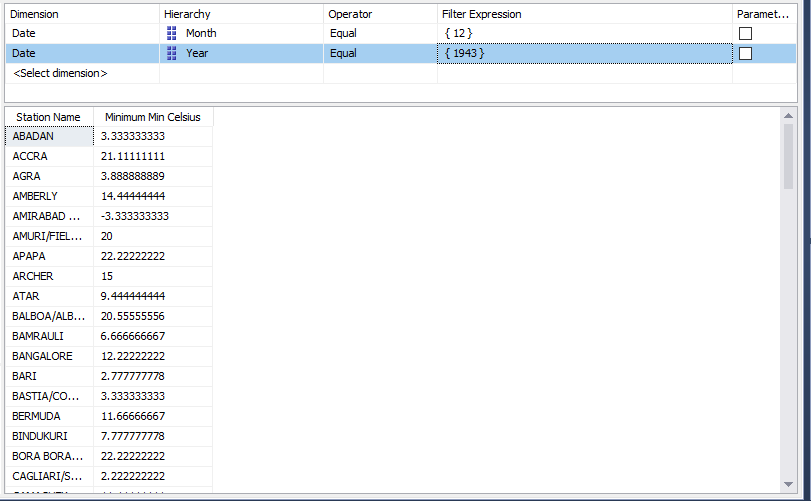


**+Xoay dữ liệu(Pivot)**



**g. dice : Nhiệt độ thấp nhất từng đo được tại các trạm đo trong tháng 12 năm**

**1943**



**h. dice: Tổng lượng mưa của tất cả các quốc gia trong năm 1944 có kèm theo sấm sét**

