

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**MÔN HỌC**

**KHO DỮ LIỆU VÀ OLAP-IS217.022.HTCL**

**ĐỀ TÀI :**

**PHÂN TÍCH Ô NHIỄM KHÔNG KHÍ Ở HÀN QUỐC**

**GVHD: Đỗ Thị Minh Phụng**

**NHÓM: 19**

**THÀNH VIÊN:**

**Nguyễn Dương Chí Tâm - 21520439**

**Hoàng Văn Tiến Nhật - 19521951**

**TP HCM, Ngày 19 Tháng 5 Năm 2024**

# Mục lục

## Contents

I.	Giới thiệu .....	7
1.	Tổng quan đề tài.....	7
1.1.	Lý do chọn đề tài.....	7
1.2.	Nội dung dữ liệu .....	8
1.3.	Mô tả dữ liệu .....	8
1.3.1.	Mô tả các cột thuộc tính trong Dataset: .....	8
2.	Xây dựng kho dữ liệu .....	9
2.1.	Lược đồ hình sao.....	9
2.2.	Các bảng chiều .....	9
2.3.	Bảng Fact .....	10
II.	Tích hợp dữ liệu vào kho (SSIS) .....	11
1.	Tạo project và thực hiện kết nối.....	11
1.1.	Tạo project .....	11
1.2.	Tạo cơ sở dữ liệu và thiết lập kết nối .....	13
1.2.1.	Tạo cơ sở dữ liệu: .....	13
1.2.2.	Thiết lập kết nối đến cơ sở dữ liệu : .....	14
2.	Chuẩn bị dữ liệu gốc, import dữ liệu gốc.....	18
3.	Quá trình làm sạch dữ liệu .....	25
4.	Quá trình tạo các bảng Dimension.....	37
4.1.	Dim_Location .....	37
4.2.	Dim_Date.....	46
5.	Quá trình tạo bảng Fact .....	56
6.	Tạo các ràng buộc khóa ngoại .....	70
7.	Thực thi .....	73
8.	Dữ liệu sau khi hoàn thành .....	73
9.	Diagram đồ án sau khi hoàn thành .....	76
III.	Phân tích dữ liệu trong kho (SSAS).....	76
1.	Quá trình thực hiện SSAS .....	76

1.1.	<b>Quá trình thực hiện trên công cụ Visual Data Studio</b>	76
1.2.	<b>Điều chỉnh các Dimation</b>	92
2.	<b>Quá trình phân tích dữ liệu bằng công cụ SSAS trên các khối CUBE</b>	101
2.1.	<b>Cho biết tổng lượng PM25 ở thành phố Pado-Ri theo từng năm</b>	101
2.2.	<b>Cho biết vùng có chỉ số 03 cao nhất trong tháng 6 năm 2017</b>	103
2.3.	<b>Cho biết thành phố nào có tổng lượng Pm10 trên 50000</b>	105
2.4.	<b>Cho biết top 5 quận có chỉ số So2 cao nhất trong từng năm</b>	107
2.5.	<b>Cho biết ngày nào có tổng lượng NO2 cao nhất</b>	110
2.6.	<b>Cho biết tổng lượng CO ở vùng Deajeon theo từng năm</b>	112
2.7.	<b>Đưa ra 3 ngày có lượng khí Pm25 ít nhất tại Paju</b>	114
2.8.	<b>Cho biết 3 thành phố có lượng khí Pm10 nhiều nhất trong năm 2017</b>	115
2.9.	<b>Với từng vùng liệt kê lượng khí So2 theo từng năm, trừ Gwangju và Gyeonggi</b>	117
2.10.	<b>Cho biết thành phố có tổng lượng o3 cao nhất và thành phố có tổng lượng o3 thấp nhất</b>	119
2.11.	<b>Với mỗi quận đưa ra 3 ngày có tổng lượng CO cao nhất</b>	120
2.12.	<b>Liệt kê những thành phố có tổng lượng o3 từ 10000 đến 20000</b>	122
2.13.	<b>Đưa ra 7 ngày có lượng Pm10 cao nhất tại quận Jeonbuk</b>	124
2.14.	<b>Liệt kê tổng lượng So2 của tất cả các vùng theo từng năm</b>	125
2.15.	<b>Liệt kê những năm có tổng lượng Pm25 từ 20000-200000</b>	126
3.	<b>Quá trình phân tích dữ liệu bằng công cụ PIVOT EXCEL</b>	127
3.1.	<b>Cho biết tổng lượng PM25 ở thành phố Pado-Ri theo từng năm</b>	127
3.2.	<b>Cho biết vùng có chỉ số 03 cao nhất trong tháng 6 năm 2017</b>	128
3.3.	<b>Cho biết thành phố nào có tổng lượng Pm10 trên 50000</b>	128
3.4.	<b>Cho biết top 5 quận có chỉ số So2 cao nhất trong từng năm</b>	129
3.5.	<b>Cho biết ngày nào có tổng lượng NO2 cao nhất</b>	129
3.6.	<b>Cho biết tổng lượng CO ở vùng Deajeon theo từng năm</b>	130
3.7.	<b>Đưa ra 3 ngày có lượng khí Pm25 ít nhất tại Paju</b>	130
3.8.	<b>Cho biết 3 thành phố có lượng khí Pm10 nhiều nhất trong năm 2017</b>	130
3.9.	<b>Với từng vùng liệt kê lượng khí So2 theo từng năm, trừ Gwangju và Gyeonggi</b>	131

3.10. Cho biết thành phố có tổng lượng o3 cao nhất và thành phố có tổng lượng o3 thấp nhất .....	131
3.11. Với mỗi quận đưa ra 3 ngày có tổng lượng CO cao nhất.....	131
3.12. Liệt kê những thành phố có tổng lượng o3 từ 10000 đến 20000.....	131
3.13. Đưa ra 7 ngày có lượng Pm10 cao nhất tại quận Jeonbuk .....	132
3.14. Liệt kê tổng lượng So2 của tất cả các vùng theo từng năm. ....	132
3.15. Liệt kê những năm có tổng lượng Pm25 từ 20000-200000.....	133
<b>4. Quá trình phân tích dữ liệu bằng ngôn ngữ MDX .....</b>	<b>133</b>
4.1. Cho biết tổng lượng PM25 ở thành phố Pado-Ri theo từng năm.....	133
4.2. Cho biết vùng có chỉ số 03 cao nhất trong tháng 6 năm 2017.....	134
4.3. Cho biết thành phố nào có tổng lượng Pm10 trên 50000. ....	135
4.4. Cho biết top 5 quận có chỉ số So2 cao nhất trong từng năm.....	136
4.5. Cho biết ngày nào có tổng lượng NO2 cao nhất .....	137
4.6. Cho biết tổng lượng CO ở vùng Deajeon theo từng năm. ....	137
4.7. Đưa ra 3 ngày có lượng khí Pm25 ít nhất tại Paju .....	138
4.8. Cho biết 3 thành phố có lượng khí Pm10 nhiều nhất trong năm 2017. ....	139
4.9. Với từng vùng liệt kê lượng khí So2 theo từng năm, trừ Gwangju và Gyeonggi.	
140	
4.10. Cho biết thành phố có tổng lượng o3 cao nhất và thành phố có tổng lượng o3 thấp nhất. ....	141
4.11. Với mỗi quận đưa ra 3 ngày có tổng lượng CO cao nhất.....	142
4.12. Liệt kê những thành phố có tổng lượng o3 từ 10000 đến 20000.....	143
4.13. Đưa ra 7 ngày có lượng Pm10 cao nhất tại quận Jeonbuk .....	144
4.14. Liệt kê tổng lượng So2 của tất cả các vùng theo từng năm. ....	145
4.15. Liệt kê những năm có tổng lượng Pm25 từ 20000-200000.....	146
<b>IV. Quá trình lập báo biểu (SSRS) .....</b>	<b>147</b>
<b>1. Quá trình lập báo biểu bằng công cụ Report builder.....</b>	<b>147</b>
1.1. Tạo project SSRS .....	147
1.2. Tạo kết nối kho dữ liệu .....	148
1.3. Thực hiện lập báo biểu trên Report Builder.....	152
1.3.1. Báo biểu 1: (Report Grouping) đưa ra số lượng ô nhiễm SO2 trong từng năm của từng vùng .....	152

<b>1.3.2. Báo biểu 2: (Report Grouping +Total +Matrix )</b>	<b>Báo cáo về tổng lượng ô nhiễm NO2 tại Paju theo ngày , tháng ,năm.</b>	163
<b>1.3.3. Báo biểu 3:(Report Grouping +Total)</b>	<b>Báo cáo với từng thành phố đưa ra tổng lượng PM10 từ 2014-2018.....</b>	169
<b>2. Quá trình lập báo biểu bằng công cụ Power BI .....</b>		175
<b>2.1. Kết nối Power BI với database.....</b>		175
<b>2.2. Báo cáo 1: đưa ra số lượng ô nhiễm SO2 trong từng năm của từng vùng .....</b>		178
<b>2.3. Báo cáo 2 : Báo cáo về tổng lượng ô nhiễm NO2 tại Paju theo tháng, năm. .</b>		180
<b>2.4. Báo cáo 3 : Báo cáo với từng thành phố đưa ra tổng lượng PM10 từ 2014-2018.</b>		181
<b>V. Quá trình Datamining.....</b>		183
<b>1. Chọn lại kho dữ liệu Bike_Buyer .....</b>		183
<b>1.1. Tổng quan kho dữ liệu .....</b>		183
<b>1.2. Mô tả dữ liệu .....</b>		183
<b>2. Quá trình thực hiện Data Mining với thuật toán cây quyết định (Microsoft decision trees).....</b>		183
<b>2.1. Đưa dữ liệu vào kho (như SSAS).....</b>		183
<b>2.2. Thiết lập cấu trúc datamining.....</b>		183
<b>2.3. Thực hiện Deploy project data minning .....</b>		195
<b>2.4. Phân tích mining bằng cây quyết định và đưa ra tập luật .....</b>		199
<b>2.5. So sánh với các thuật toán khác.....</b>		204
<b>VI. Tài liệu tham khảo.....</b>		209

## **Lời cảm ơn**

Trong chặng đường của cuộc sống, có lẽ ai cũng đã từng trải qua các cung bậc cảm xúc khác nhau, có thành công cũng có thất bại, dấu thế đó cũng là thành quả của cá nhân cũng như tập thể. Và đăng sau đó chính là sự giúp đỡ từ mọi người xung quanh. Hôm nay, để có thể hoàn thành được đồ án môn học này, nhóm chúng em rất biết ơn quý thầy cô đã hỗ trợ tận tình, cung cấp cho chúng em kiến thức cũng như lời khuyên sâu sắc.

Lời đầu tiên, chúng em xin chân thành cảm ơn đến toàn thể giảng viên trường Đại học Công nghệ Thông tin - Đại học Quốc gia TP.HCM cũng như nhà trường đã cùng tri thức, tâm huyết đã truyền đạt cho chúng em trong suốt thời gian thực hiện đồ án môn học.

Đặc biệt hơn hết, nhóm chúng em xin gửi lời cảm ơn chân thành đến cô Đỗ Thị Kim Phụng – Giảng viên môn Kho dữ liệu và Olab đã trực tiếp giảng dạy, truyền đạt kiến thức, kinh nghiệm, hướng dẫn chúng em một cách cụ thể, tận tâm giúp chúng em hoàn thành tốt đồ án môn học của mình. Chúc cô luôn dồi dào sức khỏe, luôn luôn nhiệt huyết với nghề để có thể tiếp tục giảng dạy, dùi dắt cho thế hệ sinh viên tiếp theo.

Ngoài ra, chúng em cũng gửi lời cảm ơn tập thể lớp IS217.022.HTCL nói chung cũng như các bạn thành viên trong nhóm nói riêng đã cùng nhau học tập, cùng nhau nghiên cứu, học hỏi để thực hiện đồ án một cách tốt nhất.

Cuối cùng, nhóm chúng em đã hoàn thành đồ án mang tên "Phân tích dữ liệu ô nhiễm không khí ở Hàn Quốc". Mặc dù đã vận dụng tối đa những gì đã học nhưng vẫn khó tránh khỏi sai sót, vì vậy nhóm chúng em rất mong nhận được sự góp ý từ phía quý thầy cô để tiếp tục hoàn thiện dự án một cách tốt nhất. Qua đó, rút ra bài học và tích lũy thêm kinh nghiệm làm hành trang cho tương lai.

Chúng em xin hết và xin chân thành cảm ơn cô và các bạn!

Trân trọng cảm ơn.

TP. Hồ Chí Minh, ngày 19 tháng 5 năm 2023

## I. Giới thiệu

### 1. Tổng quan đề tài

#### 1.1. Lý do chọn đề tài

Trong bối cảnh của sự phát triển vượt bậc công nghệ, lượng dữ liệu được sinh ra và lưu trữ ngày càng nhiều. Nhận thức về tiềm năng và giá trị của dữ liệu đã khiến việc phân tích dữ liệu trở thành một xu hướng quan trọng, đặc biệt là trong lĩnh vực quản lý. Cùng có nhu cầu về phân tích quản lý, chúng em tìm kiếm tệp dữ liệu phù hợp với yêu cầu của môn học, đồng thời đảm bảo được nguồn gốc dữ liệu minh bạch có ý nghĩa trong thực tế.

Đặc biệt trong thời kỳ công nghiệp phát triển như hiện nay, tình hình ô nhiễm đã và đang trở nên ngày một trầm trọng. Do đó, việc phân tích dữ liệu sẽ là một đóng góp quan trọng trong việc phân tích và đưa ra phương án giải quyết hợp lý để giảm bớt lượng ô nhiễm trong môi trường sống.

Dữ liệu cũng đóng vai trò quan trọng trong việc tạo ra sự nhận thức cộng đồng về vấn đề ô nhiễm môi trường. Bằng cách trình bày dữ liệu một cách trực quan và dễ hiểu, chúng ta có thể giúp tăng cường ý thức và động viên mọi người tham gia vào các hoạt động bảo vệ môi trường và giảm thiểu tác động tiêu cực đến sức khỏe và cuộc sống của chúng ta.

Data Analytics cung cấp các công cụ, phần mềm, thuật toán và nền tảng giúp chúng ta phân tích các số liệu về các chỉ số ô nhiễm, vị trí ô nhiễm. Nhờ vào việc này, chúng em có thể tạo ra các biểu đồ và bản đồ về tình trạng ô nhiễm hiện tại, từ đó có cơ sở để đề xuất các giải pháp và hành động cụ thể để giải quyết vấn đề một cách chính xác và hiệu quả.

## 1.2. Nội dung dữ liệu

Bộ dữ liệu chứa dữ liệu chất lượng không khí ở cấp độ hàng ngày của các trạm khác nhau trên nhiều thành phố ở Hàn Quốc.

- Bộ dữ liệu gồm 34530 dòng và 13 cột thuộc tính.
- Link dataset: <https://www.kaggle.com/datasets/calebreigada/south-korean-pollution>

## 1.3. Mô tả dữ liệu

Mô hình kho dữ liệu đề xuất theo hướng chủ đề: Quản lý

Mô tả chi tiết bảng Fact, các bảng Dimension trong kho dữ liệu đề xuất: STT, Tên cột, Kiểu dữ liệu, Ý nghĩa :

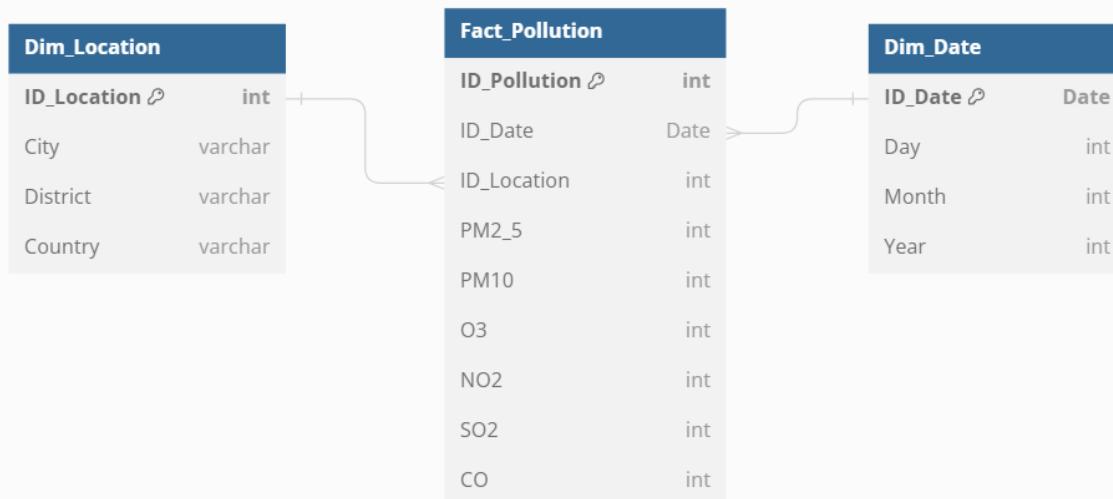
### 1.3.1. Mô tả các cột thuộc tính trong Dataset:

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	ID	int	Số thứ tự
2	Date	Date	Ngày đo kết quả
3	PM 2.5	int	Bụi mịn 2.5 micrometer trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ ).
4	PM 10	int	Bụi mịn 10 micrometer trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )
5	NO2	int	Nitric Dioxide( $\text{NO}_2$ ) trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )

6	O3	int	Ozone (O3) trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )
7	SO2	int	Sulfur Dioxide(SO2) trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )
8	CO	int	Carbon Monoxide trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )
9	Lat	float	Vĩ độ
10	Long	float	Kinh độ
11	City	Varchar	Tên thành phố
12	District	Varchar	Huyện
13	Country	Varchar	Quốc gia

## 2. Xây dựng kho dữ liệu

### 2.1. Lược đồ hình sao



### 2.2. Các bảng chiều

Dim\_Location :

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
1	ID_Location	int	Primary key	ID của vị trí
2	City	Varchar		Thành phố
3	District	Varchar		Huyện
4	Country	Varchar		Quốc gia

Dim\_Date :

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
1	ID_Date	Date	Primary key	Chi tiết ngày
2	Day	Int		Ngày
3	Month	Int		Tháng
4	Year	int		Năm

### 2.3. Bảng Fact

Fact\_Pollution :

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Ý nghĩa
1	ID_Polluton	Int	Primary key	ID của bảng
2	ID_Date	Date	Foreign key	Ngày
3	ID_Location	Int	Foreign key	ID vị trí thành phố
4	PM2_5	Int		Bụi mịn 2.5 micrometer trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ ).
5	PM10	Int		Bụi mịn 10 micrometer trên mỗi mét

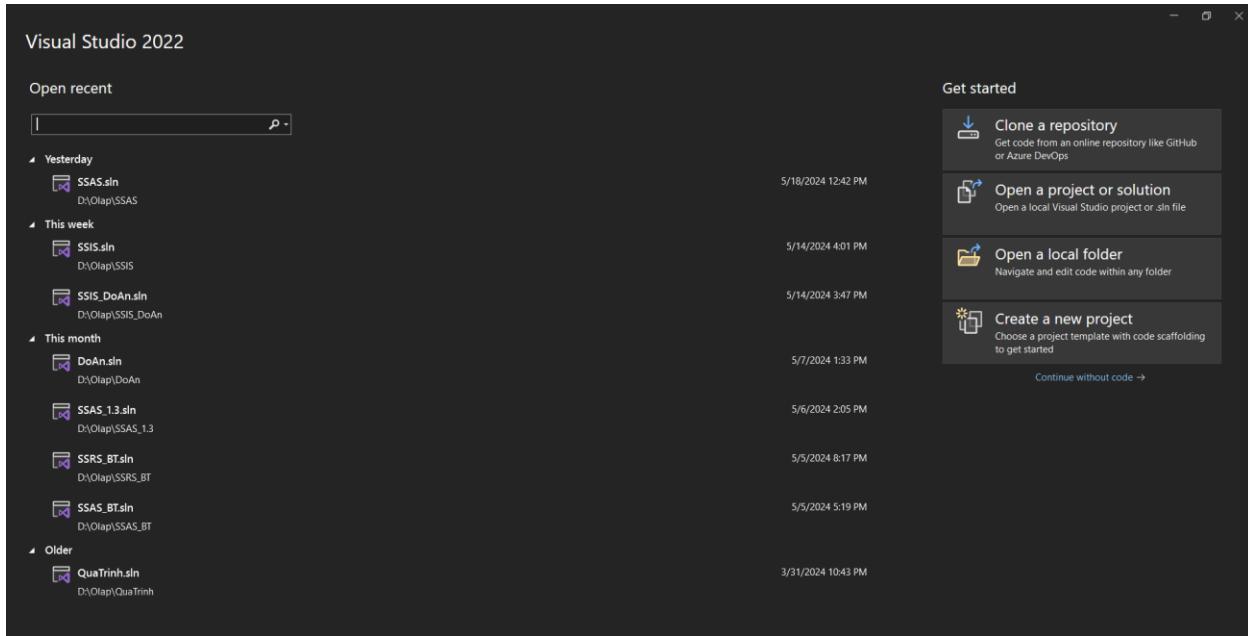
				khối của không khí ( $\mu\text{g}/\text{m}^3$ )
6	O3	Int		Ozone (O3) trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )
7	NO2	Int		Nitric Dioxide(NO2) trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )
8	SO2	Int		Sulfur Dioxide(SO2) trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )
9	CO	int		Carbon Monoxide trên mỗi mét khối của không khí ( $\mu\text{g}/\text{m}^3$ )

## II. Tích hợp dữ liệu vào kho (SSIS)

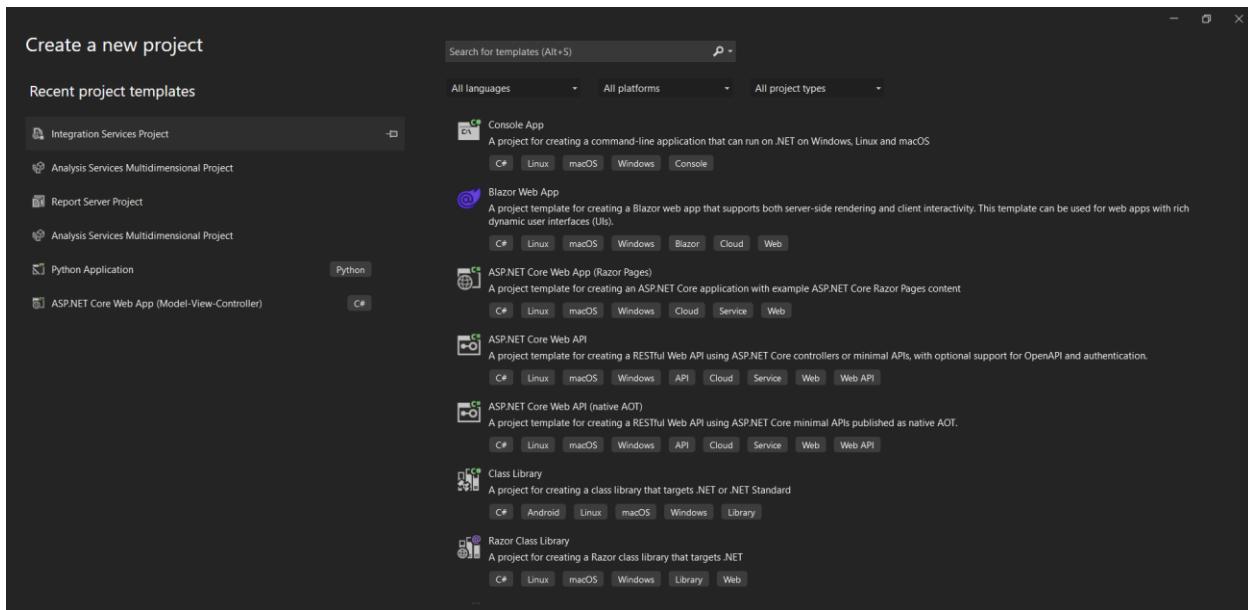
### 1. Tạo project và thực hiện kết nối

#### 1.1. Tạo project

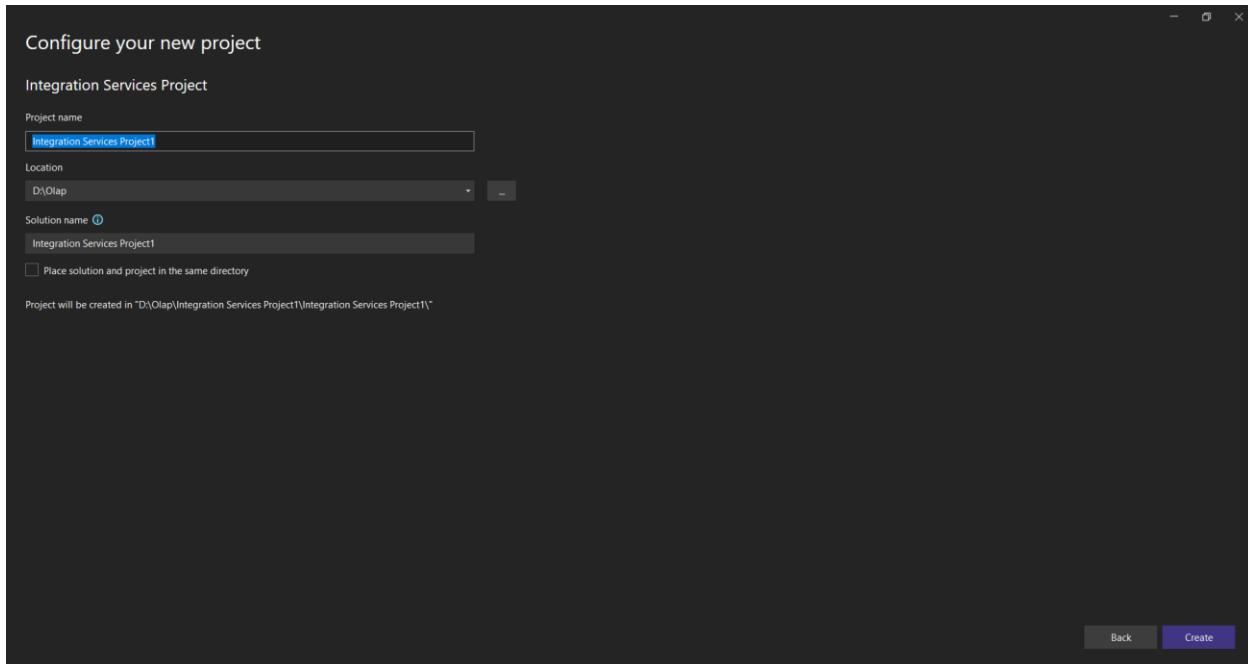
**Bước 1:** Mở Visual Studio (2022) và tạo mới dự án



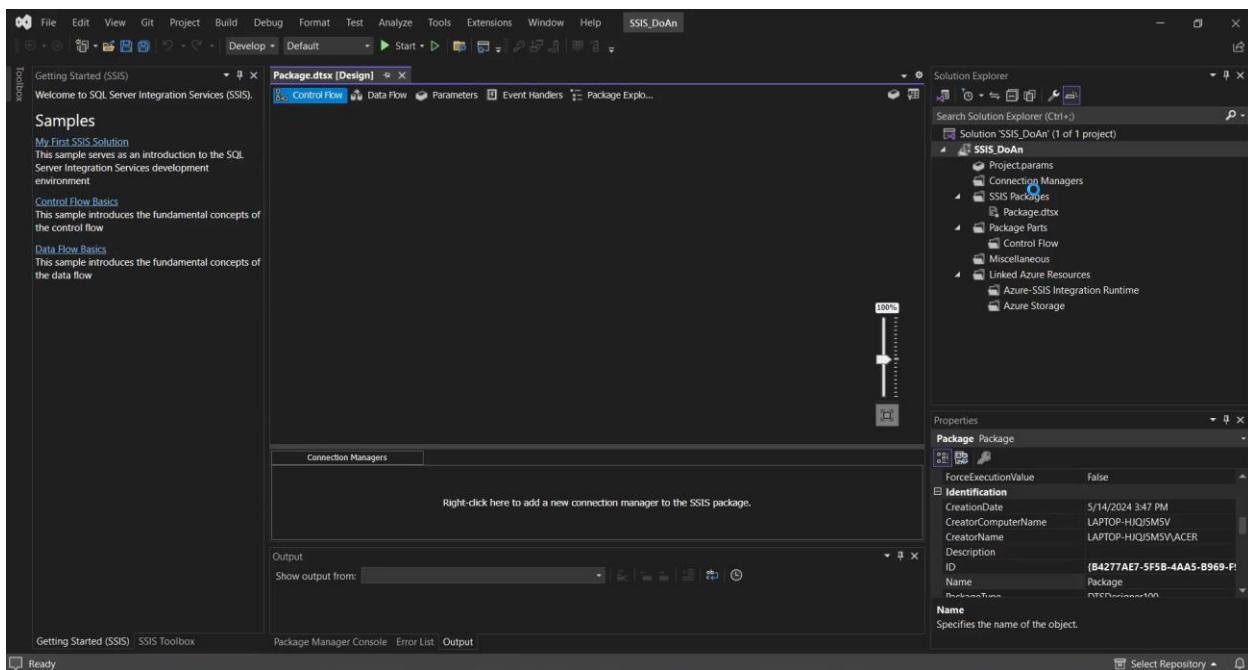
## Bước 2: Để tạo dự án SSIS ta sử dụng công cụ “Integration Services Project”.



## Bước 3: Điền các thông tin như tên dự án , địa điểm lưu dự án và tạo dự án.



**Bước 4:** Sau khi tạo hoàn thành ta có kết quả.



## 1.2. Tạo cơ sở dữ liệu và thiết lập kết nối

### 1.2.1. Tạo cơ sở dữ liệu:

Ta sử dụng công cụ MSSMS 18 để tạo 2 database cho dự án

The screenshot shows the SSMS interface. In the Object Explorer, there is a connection to 'LAPTOP-HJQ5M5V\SQLEXPRESS (SQL Server 16.0.1000.6 - LAP...)'. Under the 'Databases' node, several databases are listed: System Databases, Database Snapshots, buffet, and CakeShop. In the center, a query window titled 'SQLQuery1.sql - L..JQ5M5V\ACER (74)\*' displays the following T-SQL code:

```
create database DW_AirPollution  
create database OriginalData
```

Chi tiết database :

OriginalData: là nơi lưu trữ dữ liệu gốc, có chứa các bảng :

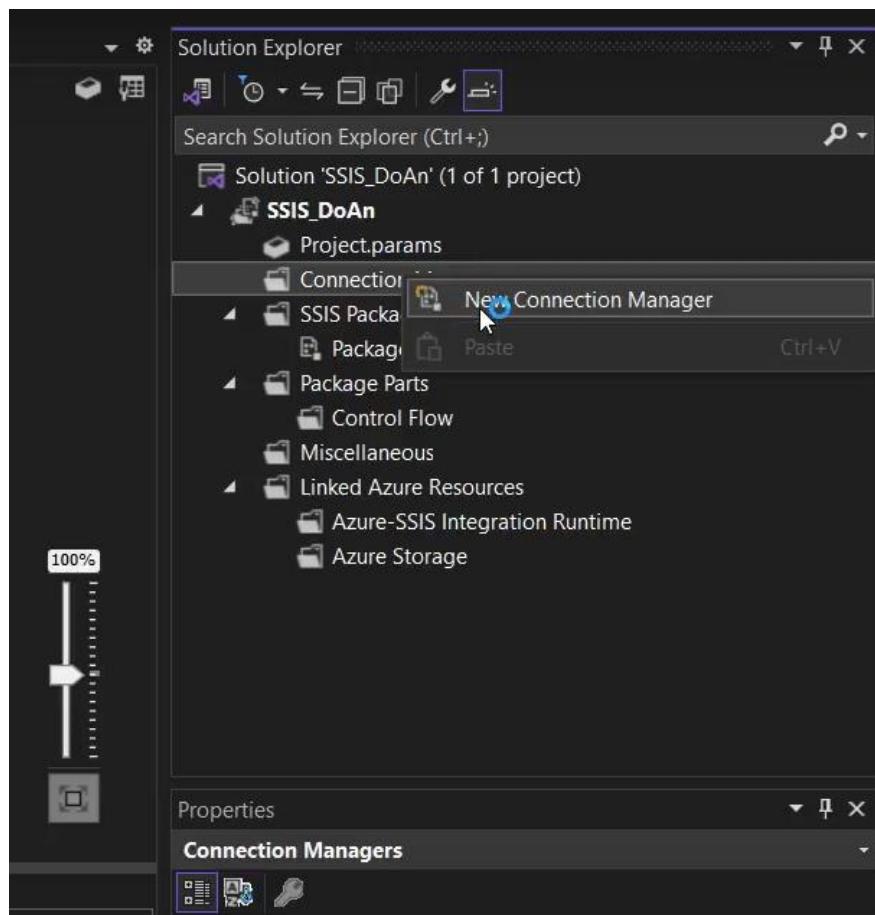
- Clear\_Data: chứa các dữ liệu sau khi được làm sạch.
- Null\_Data: chứa các dữ liệu Null trong quá trình làm sạch.
- Original: Dữ liệu data gốc được import thẳng từ file CSV.

DW\_AirPollution: là nơi lưu trữ dữ liệu của các bảng Dim , Fact :

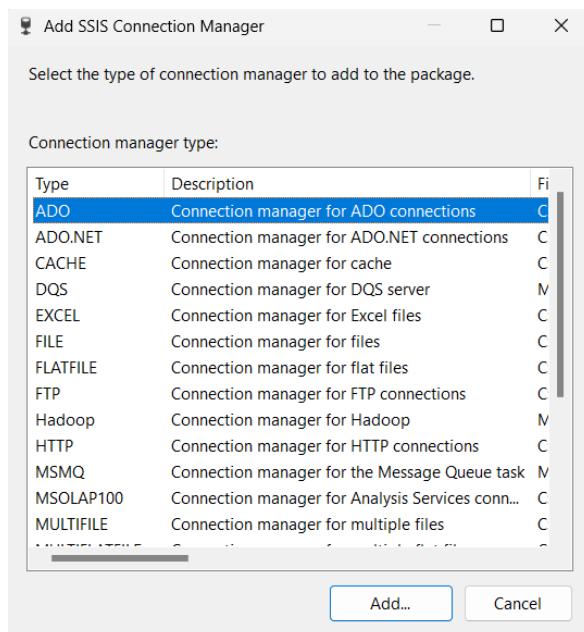
- Fact\_Pollution:
- Dim\_Location:
- Dim\_Date :

### **1.2.2. Thiết lập kết nối đến cơ sở dữ liệu :**

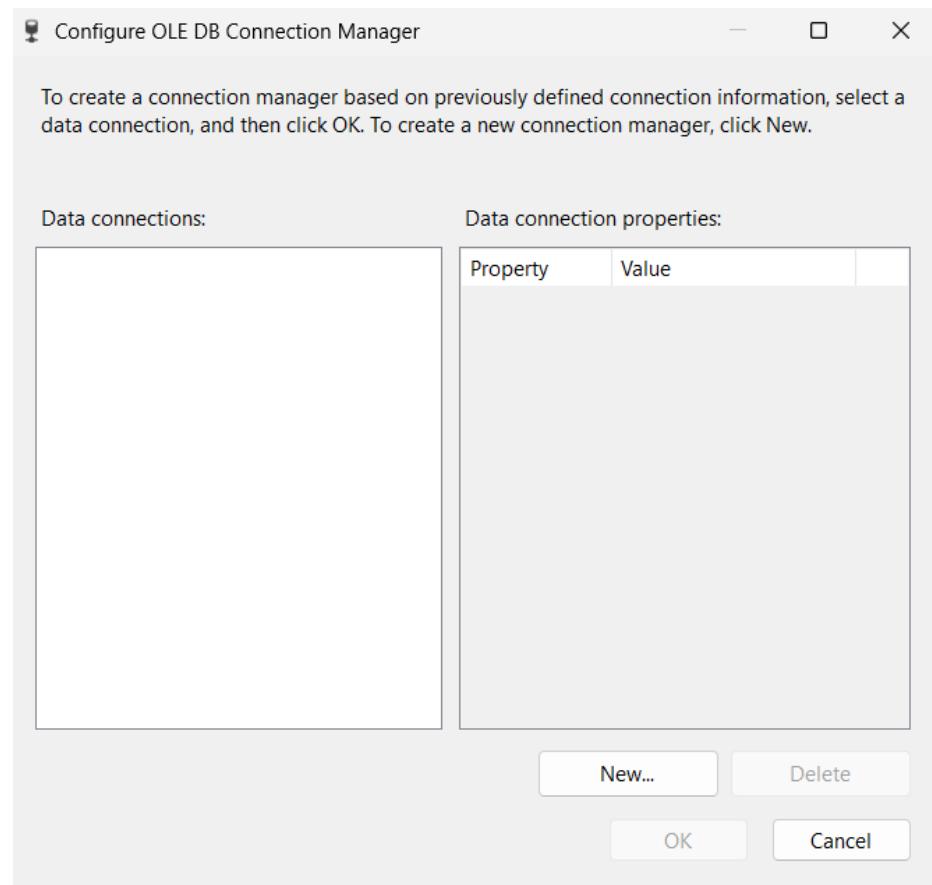
Bước 1 : Ta thực hiện tạo “New Connection Manager” ở trong dự án SSIS



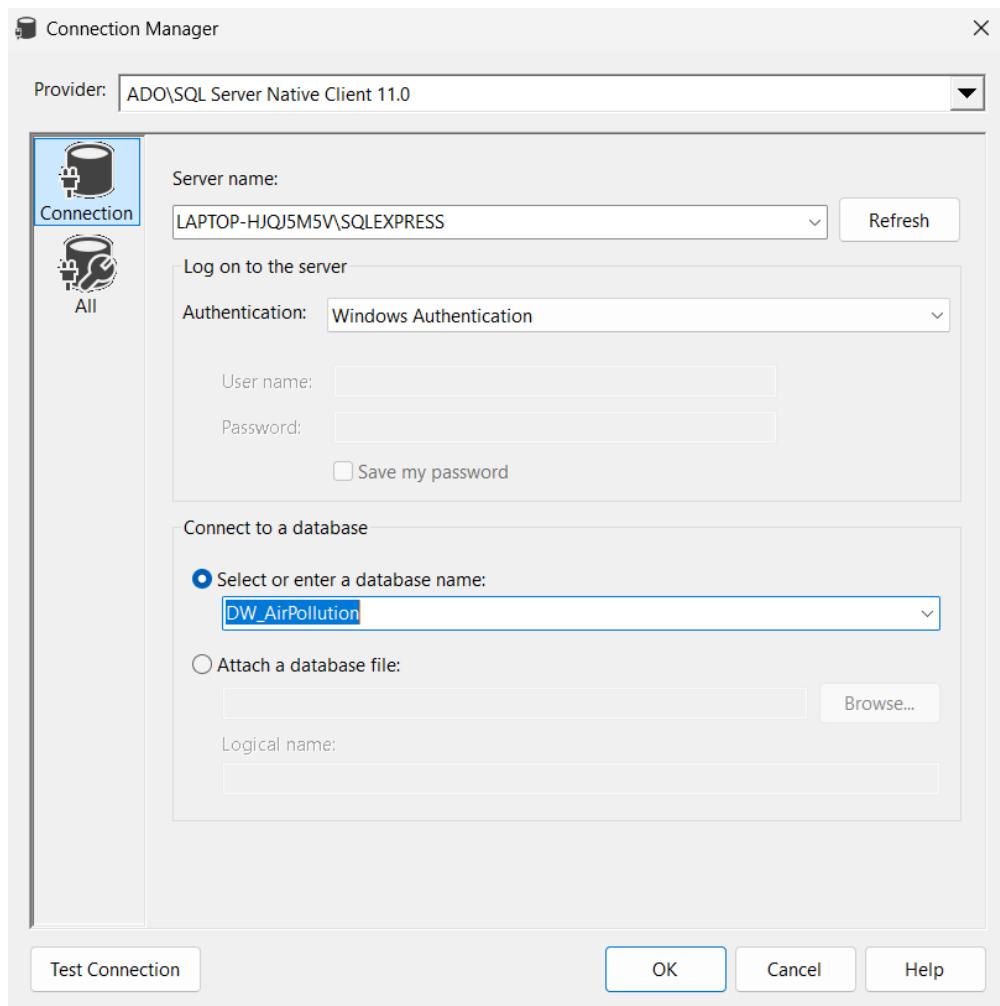
Bước 2: Ta chọn kết nối bằng kiểu "ADO"



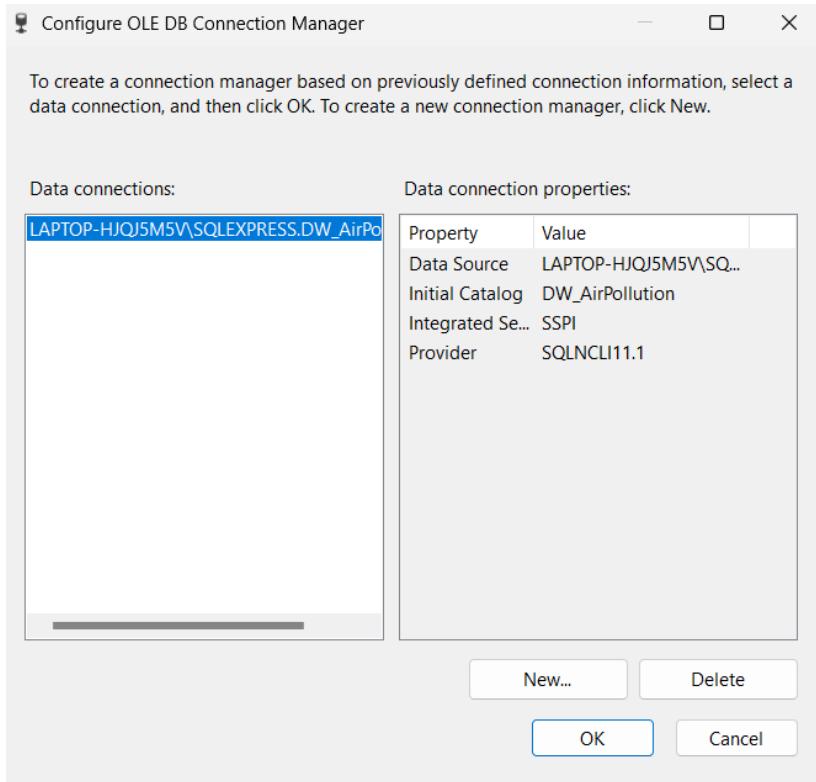
Bước 3: Ta thực hiện việc tạo mới kết nối đến database bằng cách chọn "New"



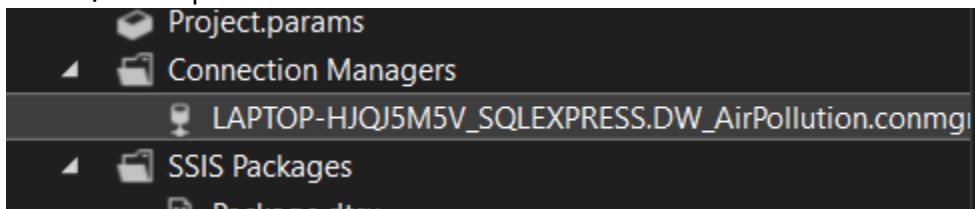
Bước 4 : Ta chọn đến database cần kết nối



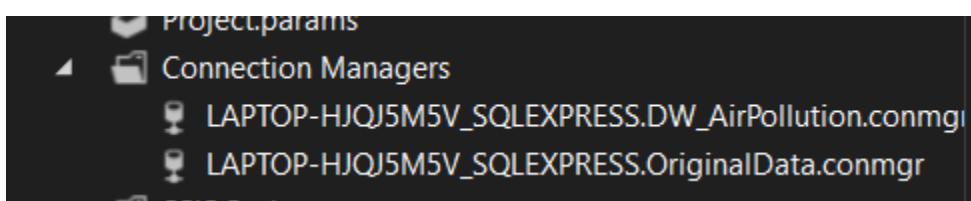
Bước 5 : Sau khi hoàn tất ta chọn ok để thực hiện kết nối



Ta được kết quả như sau :

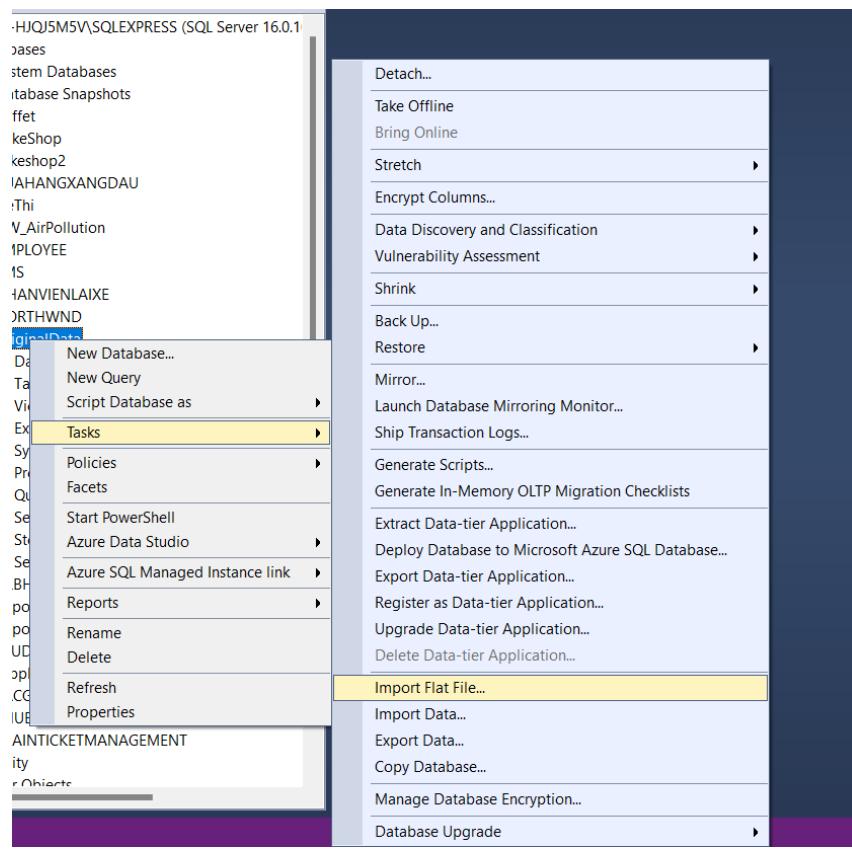


Ta thực hiện tương tự các bước trên để kết nối đến database OriginalData, Ta được kết quả :

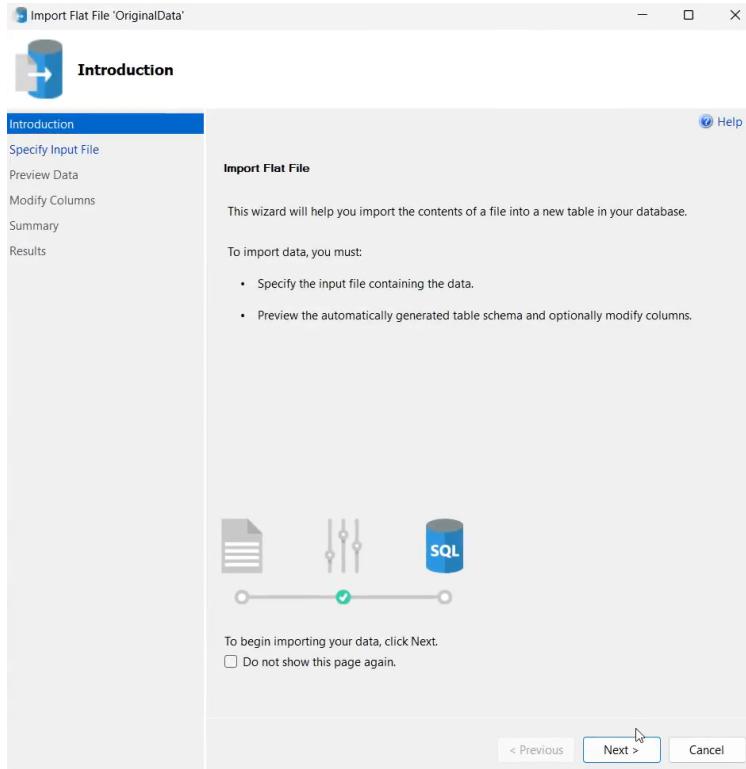


## 2. Chuẩn bị dữ liệu gốc, import dữ liệu gốc

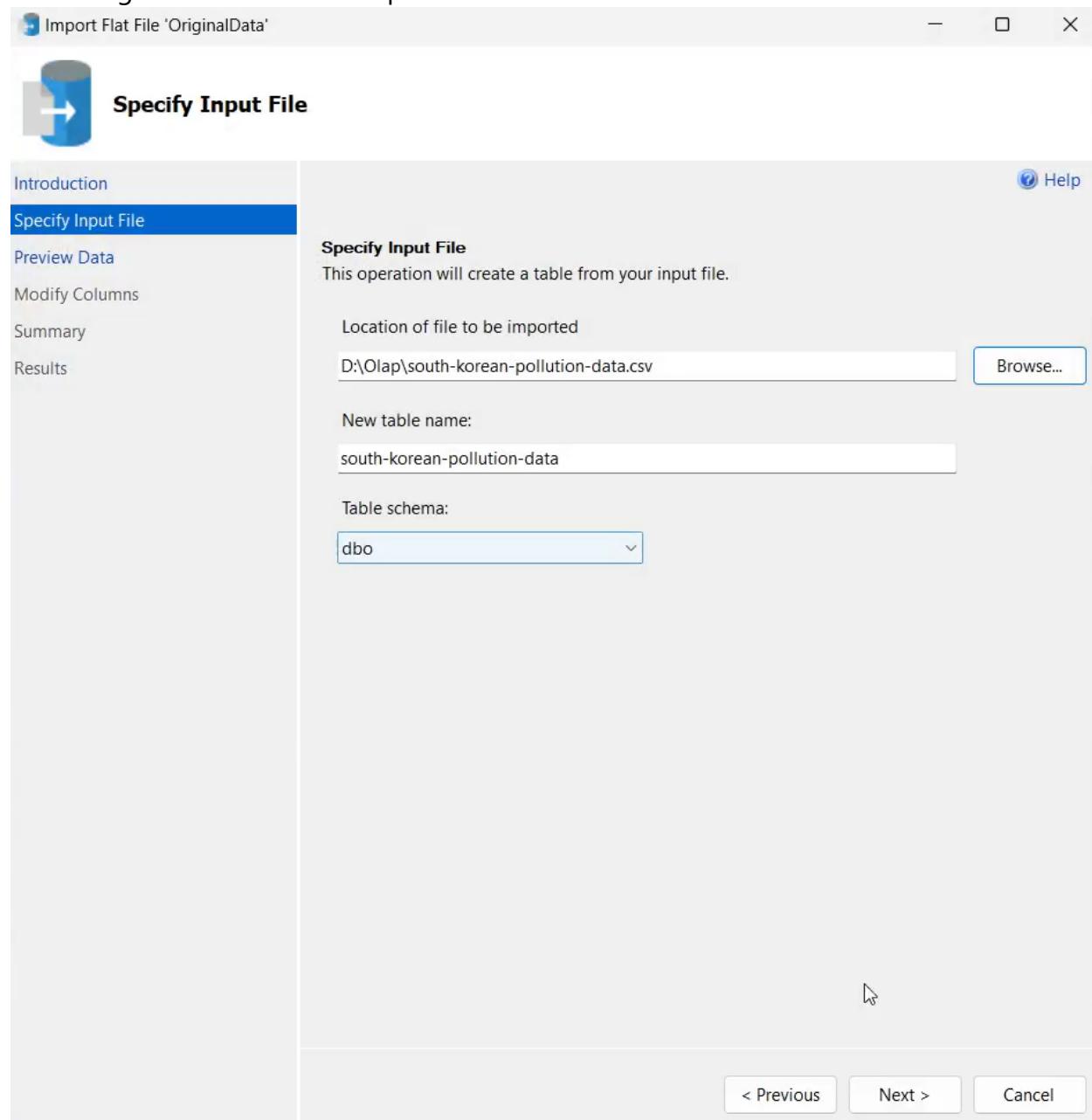
Bước 1 : Ta thực hiện việc đổ dữ liệu gốc của dự án ta vào database OriginalData bằng cách import Flat File



Trang để import file hiện ra:



Bước 2 : Ta chọn file CSV dữ liệu mà ta muốn import vào database , ở đây sẽ import file data gốc là "south-korean-pollution-data.csv"



Bước 3 : Ta xem và kiểm tra dữ liệu đã được nhập nếu ok thì ấn "Next".

Import Flat File 'OriginalData'

## Preview Data

Introduction  
Specify Input File  
**Preview Data**  
Modify Columns  
Summary  
Results

Help

**Preview Data**

This operation analyzed the input file structure to generate the preview below for up to the first 50 rows.

column1	date	pm25	pm10	o3	no2	so2
0	2022/2/1	112	31	35	2	1
1	2022/2/2	92	21	35	2	1
2	2022/2/3	60	20	35	1	1
3	2022/2/4	51	27	33	1	1
4	2022/2/5	57	24	27	2	1
5	2022/2/6	51	23	33	3	1
6	2022/2/7	61	21	34	3	1
7	2022/2/8	58	21	34	2	1
8	2022/2/9	56	22	36	2	1
9	2022/2/10	58	45	44	3	2
10	2022/2/11	107	51	39	3	1
11	2022/2/12	129	35	37	3	1
12	2022/2/13	85	28	38	4	1
13	2022/2/14	74	34	34	1	0
14	2022/2/15	88	24	33	1	2
15	2022/2/16	64	25	30	3	1
16	2022/2/17	66	21	34	3	1
..	.....	..	..	..	..	..

Column names changed due to invalid characters, duplication, etc. Column names can be edited in Modify Columns page.

Use Rich Data Type Detection - may provide a closer type fit. However, cells with anomalous values may be dropped.

< Previous Next  Cancel

Bước 4 : Ta thực hiện sửa datatype cho các cột thuộc tính cho chính xác .

Import Flat File 'OriginalData'

## Modify Columns

Introduction Specify Input File Preview Data Modify Columns

Help

Summary Results

**Modify Columns**

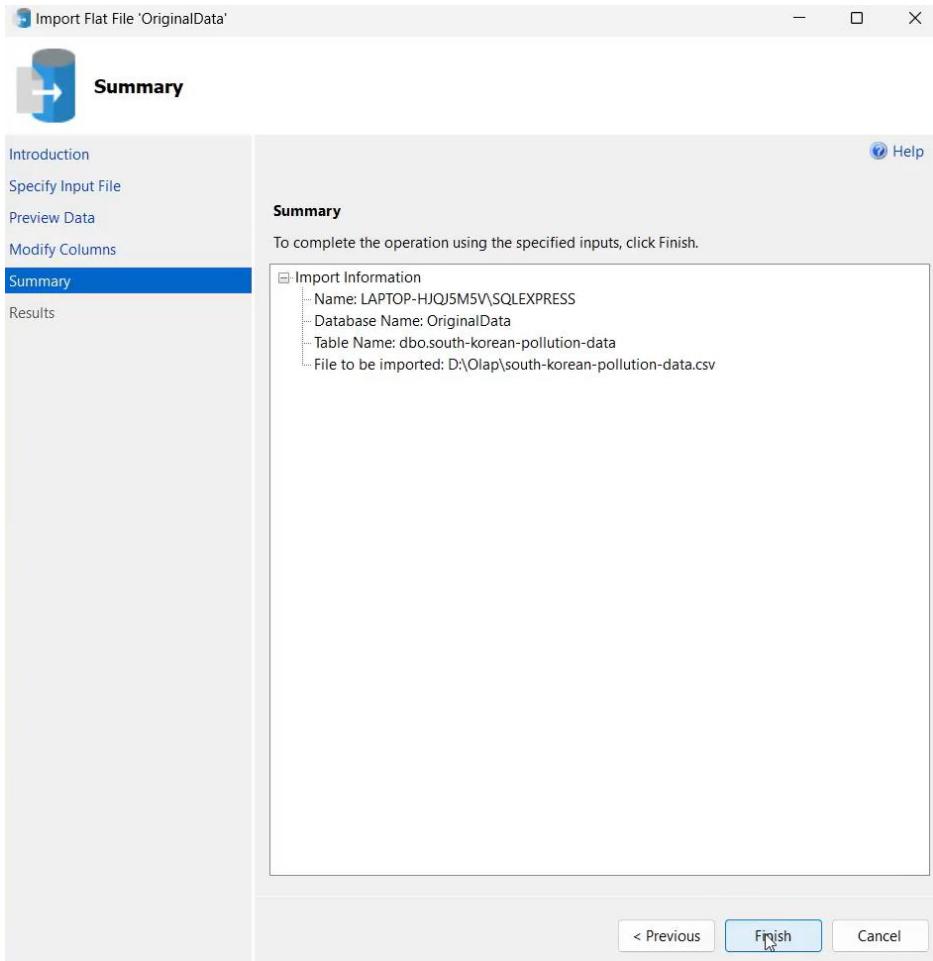
This operation generated the following table schema. Please verify if schema is accurate, and if not, please make any changes.

Column Name	Data Type	Primary Key	Allow Nulls
ID	int	<input checked="" type="checkbox"/>	<input type="checkbox"/>
date	date	<input type="checkbox"/>	<input type="checkbox"/>
pm25	int	<input type="checkbox"/>	<input type="checkbox"/>
pm10	int	<input type="checkbox"/>	<input type="checkbox"/>
o3	int	<input type="checkbox"/>	<input type="checkbox"/>
no2	int	<input type="checkbox"/>	<input type="checkbox"/>
so2	int	<input type="checkbox"/>	<input type="checkbox"/>
co	int	<input type="checkbox"/>	<input type="checkbox"/>
Lat	float	<input type="checkbox"/>	<input type="checkbox"/>
Long	float	<input type="checkbox"/>	<input type="checkbox"/>
City	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
District	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
Country	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>

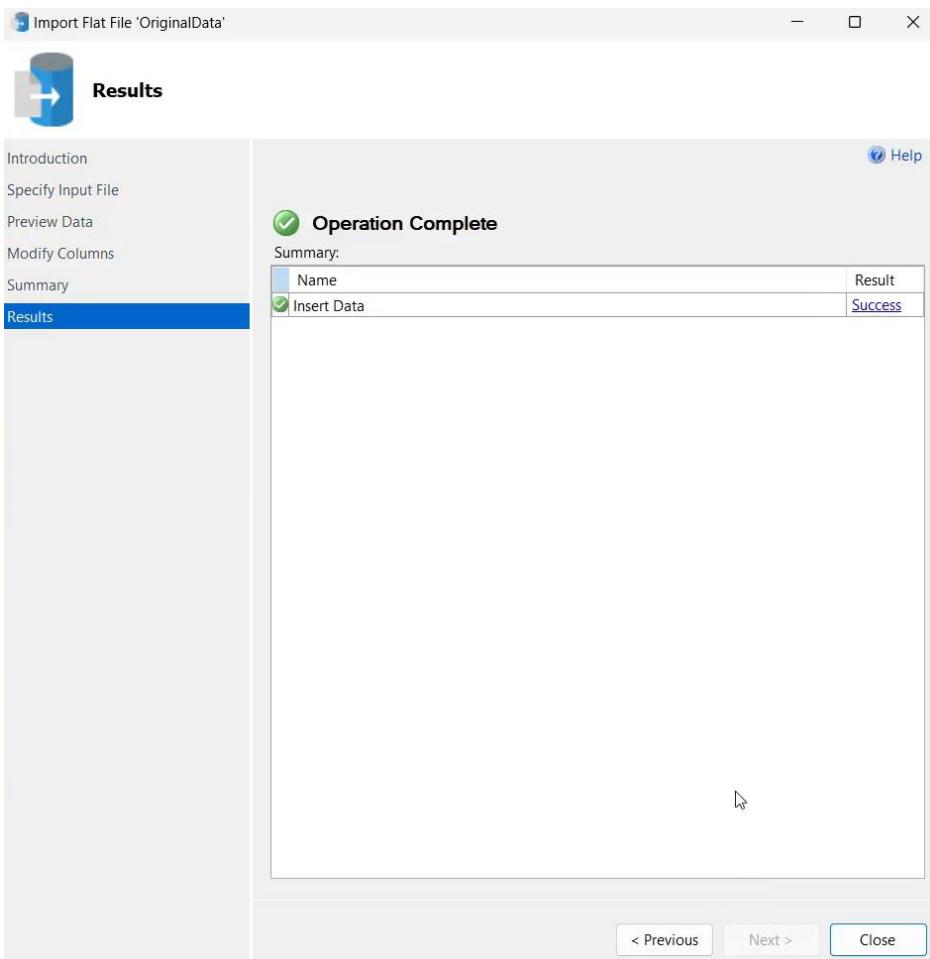
Row granularity of error reporting (performance impact with smaller ranges) No Range

< Previous Next > Cancel

Bước 5 : Ta ấn Finish để thực hiện việc import



Ta được kết quả như sau :



Như vậy dữ liệu đã được thêm thành công . Ta có thể kiểm tra bằng cách truy vấn trong SQLServer :

SQLQuery1.sql - L...JQJ5M5V\ACER (62)\*

```
Select * from [dbo].[south-korean-pollution-data]
```

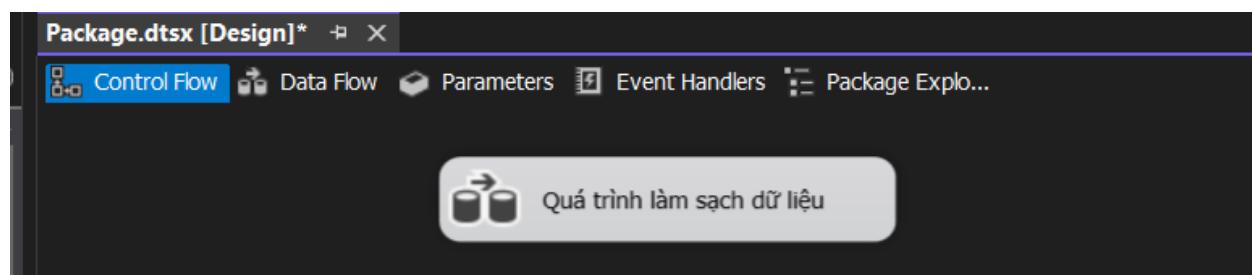
131 %

ID	date	pm25	pm10	o3	no2	so2	co	Lat	Long	City	District	Country	
1	0	2022-02-01	112	31	35	2	1	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
2	1	2022-02-02	92	21	35	2	1	0	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
3	2	2022-02-03	60	20	35	1	1	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
4	3	2022-02-04	51	27	33	1	1	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
5	4	2022-02-05	57	24	27	2	1	5	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
6	5	2022-02-06	51	23	33	3	1	5	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
7	6	2022-02-07	61	21	34	3	1	5	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
8	7	2022-02-08	58	21	34	2	1	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
9	8	2022-02-09	56	22	36	2	1	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
10	9	2022-02-10	58	45	44	3	2	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
11	10	2022-02-11	107	51	39	3	1	5	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
12	11	2022-02-12	129	35	37	3	1	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
13	12	2022-02-13	85	28	38	4	1	5	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
14	13	2022-02-14	74	34	34	1	0	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
15	14	2022-02-15	88	24	33	1	2	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
16	15	2022-02-16	64	25	30	3	1	5	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
17	16	2022-02-17	66	21	34	3	1	3	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
18	17	2022-02-18	57	29	35	3	1	3	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
19	18	2022-02-19	69	30	35	1	0	3	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
20	19	2022-02-20	69	0	0	0	0	0	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
21	20	2022-01-03	59	29	25	2	3	5	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
22	21	2022-01-04	70	25	26	3	1	3	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
23	22	2022-01-05	66	21	27	3	1	3	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
24	23	2022-01-06	59	33	28	3	1	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
25	24	2022-01-07	75	38	23	5	1	4	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
26	25	2022-01-08	97	60	27	7	1	6	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
27	26	2022-01-09	155	60	32	4	1	6	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea
28	27	2022-01-10	154	33	26	4	2	5	38.2089004516602	127.949501037598	Bangsan-Myeon	Gangwon	South Korea

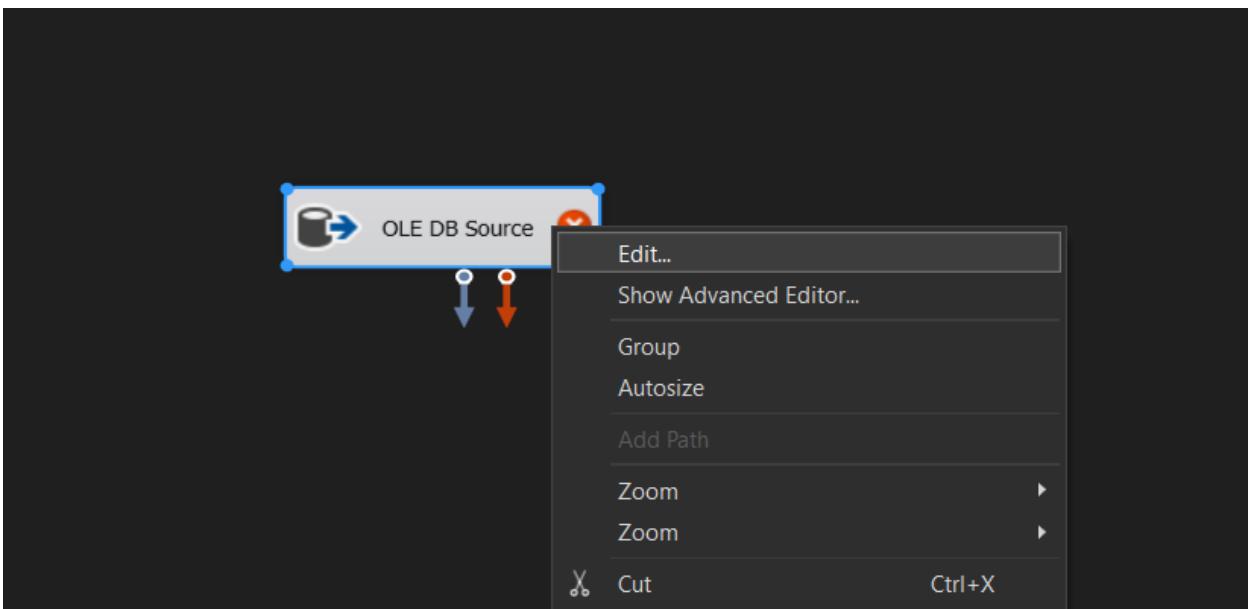
Query executed successfully.

### 3. Quá trình làm sạch dữ liệu

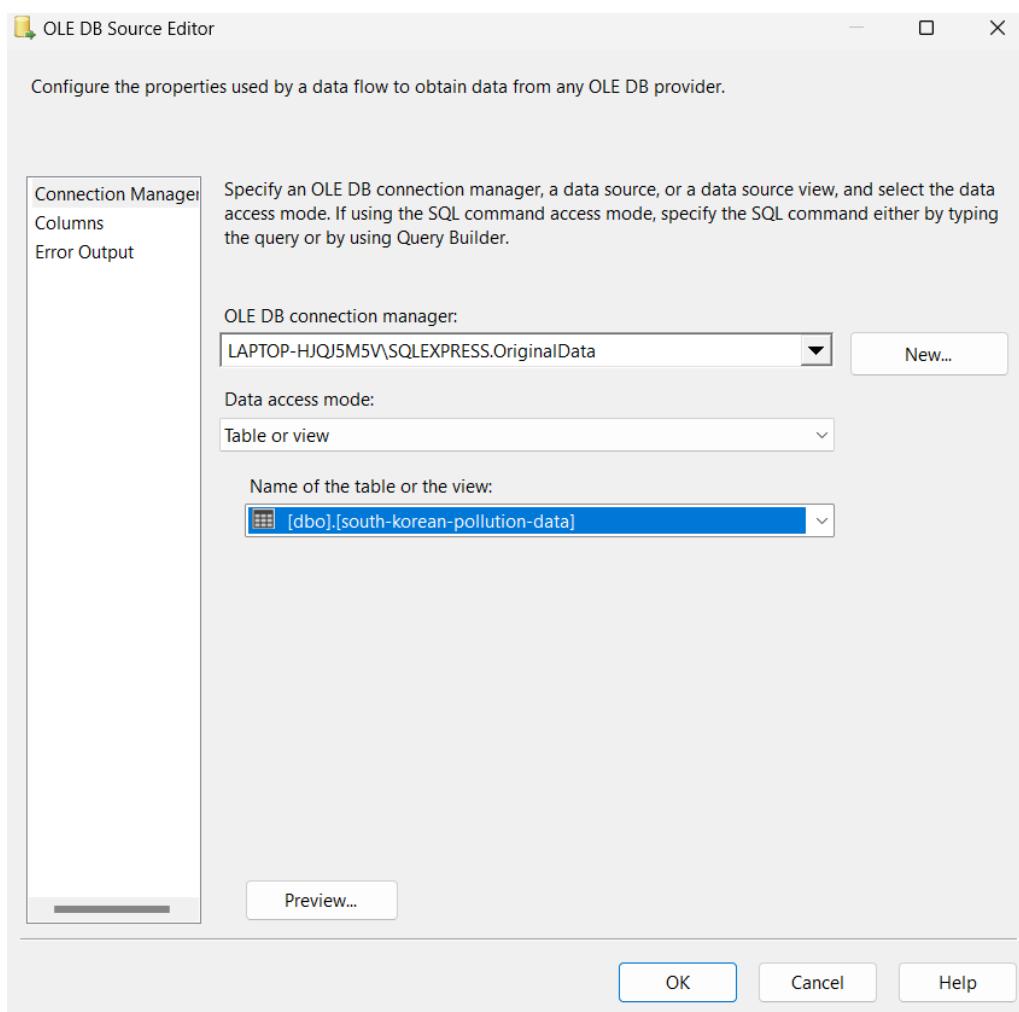
Bước 1: Ở công cụ SSIS của Visual Studio . Ta thực hiện việc tạo mới 1 "Data Flow Task" có tên là "Quá trình làm sạch dữ liệu " .



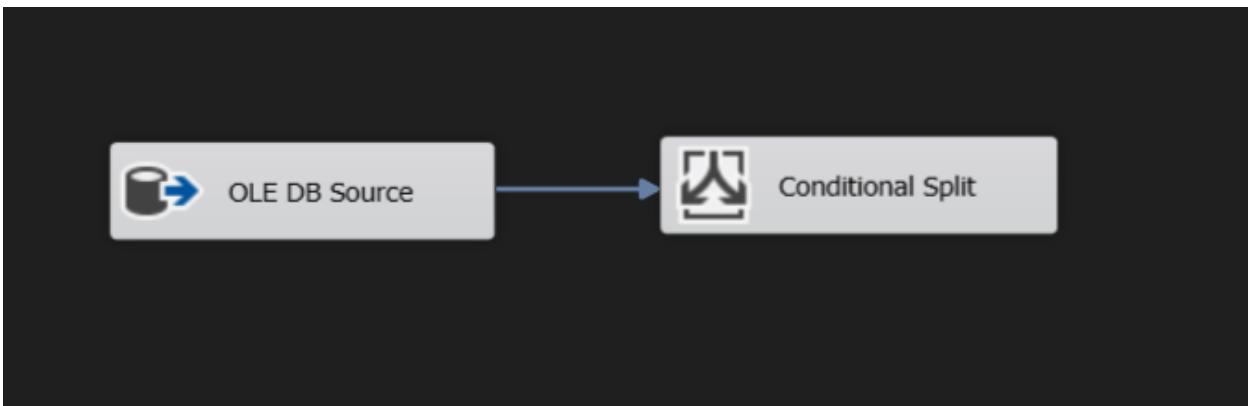
Bước 2 : Ta thực hiện chỉnh sửa "Data Flow Task" bằng cách ấn vào Edit . Ta thực hiện thêm "OLE DB Source" và cài đặt.



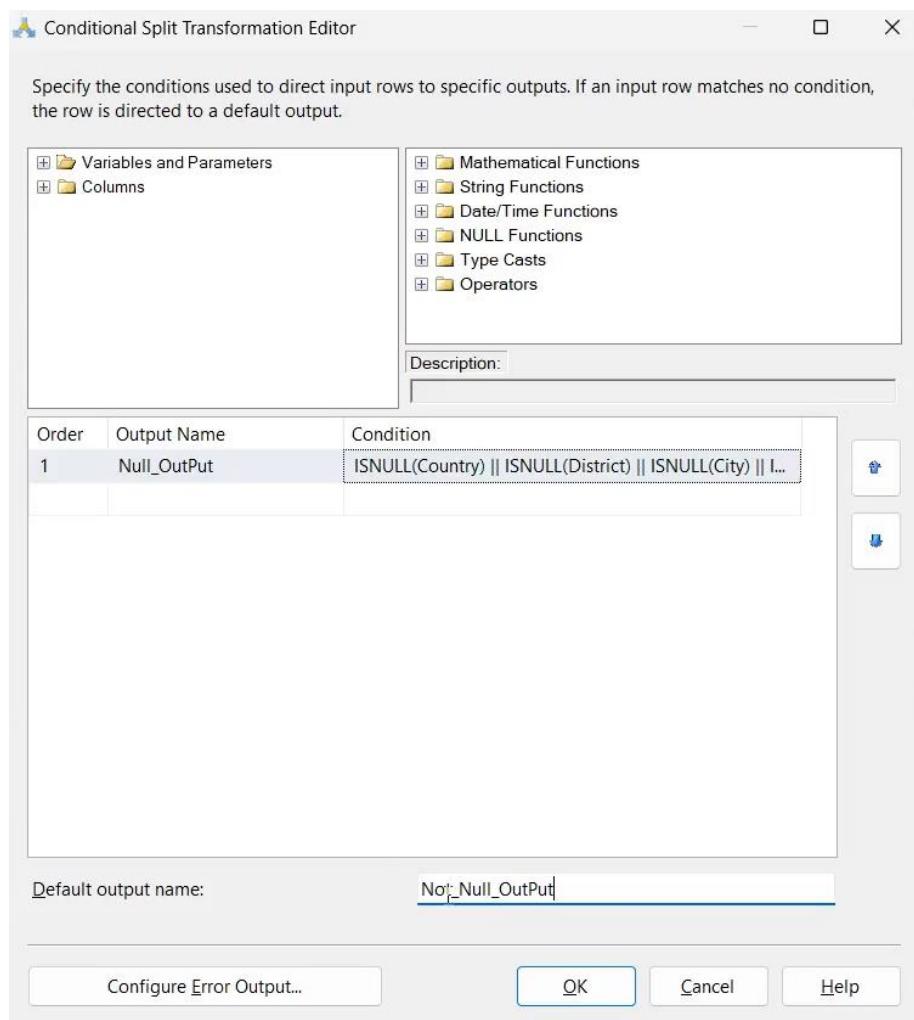
Ta thực hiện việc kết nối đến database OriginalData và chọn bảng dữ liệu gốc.



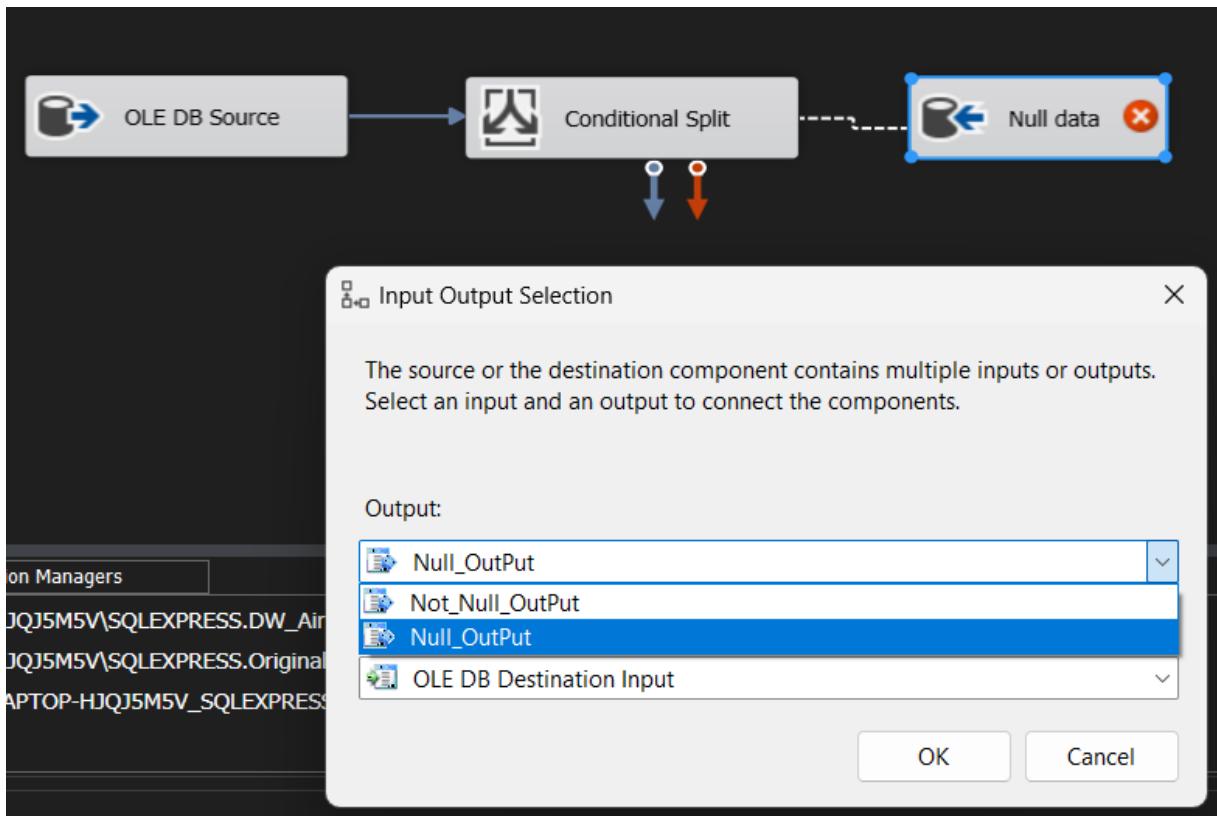
Bước 3 : Ta thêm “Condition Split” để tạo điều kiện lọc các hàng chứa giá trị Null trong dữ liệu gốc .



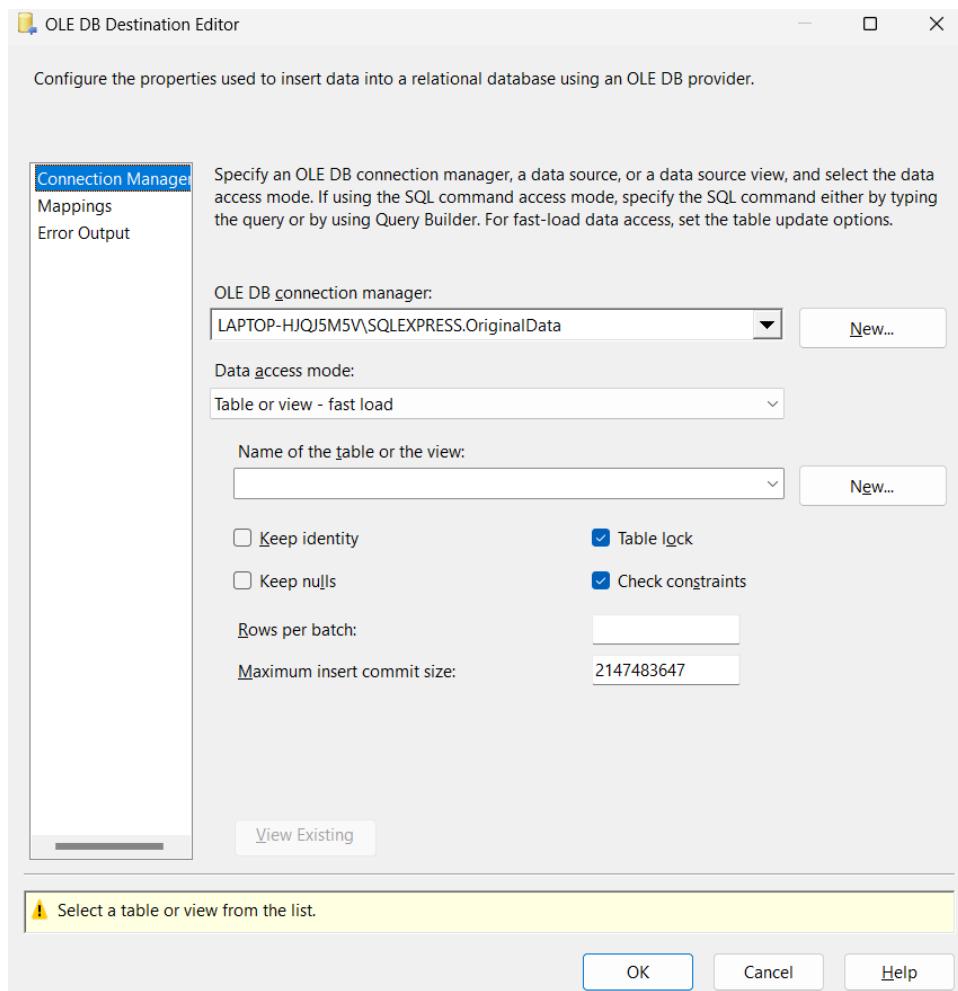
Ta cấu hình trong “ Conditional Split” để có thể lọc được các giá trị null trong dữ liệu, ta sử dụng hàm ISNULL() để lọc các giá trị null , các giá trị null sẽ được đưa ra Null\_OutPut và output mặc định là Not\_Null\_OutPut



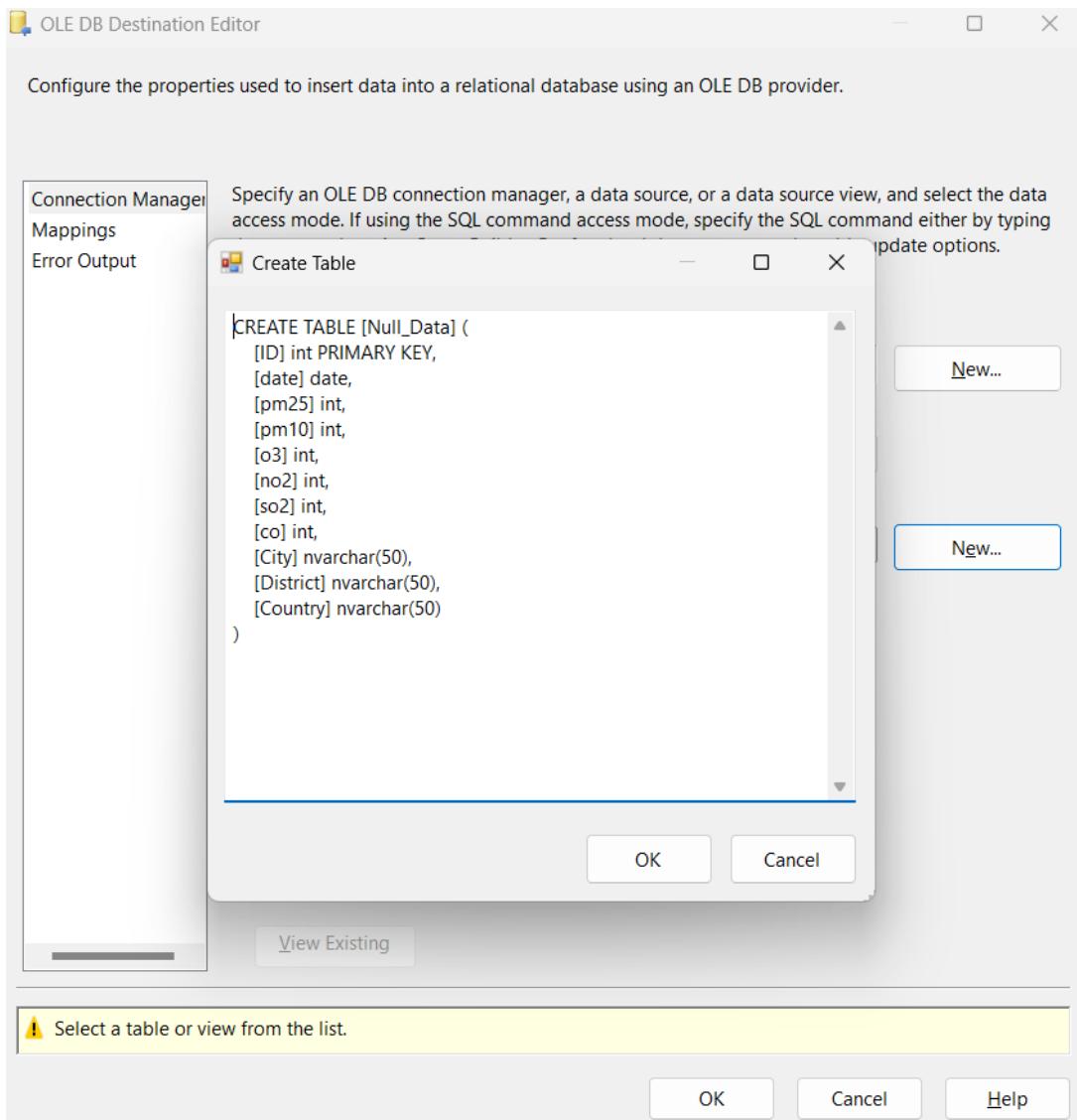
Bước 4 : Ta tạo “OLE DB Destination” với tên là “Null data ” để lưu các giá trị Null\_OutPut



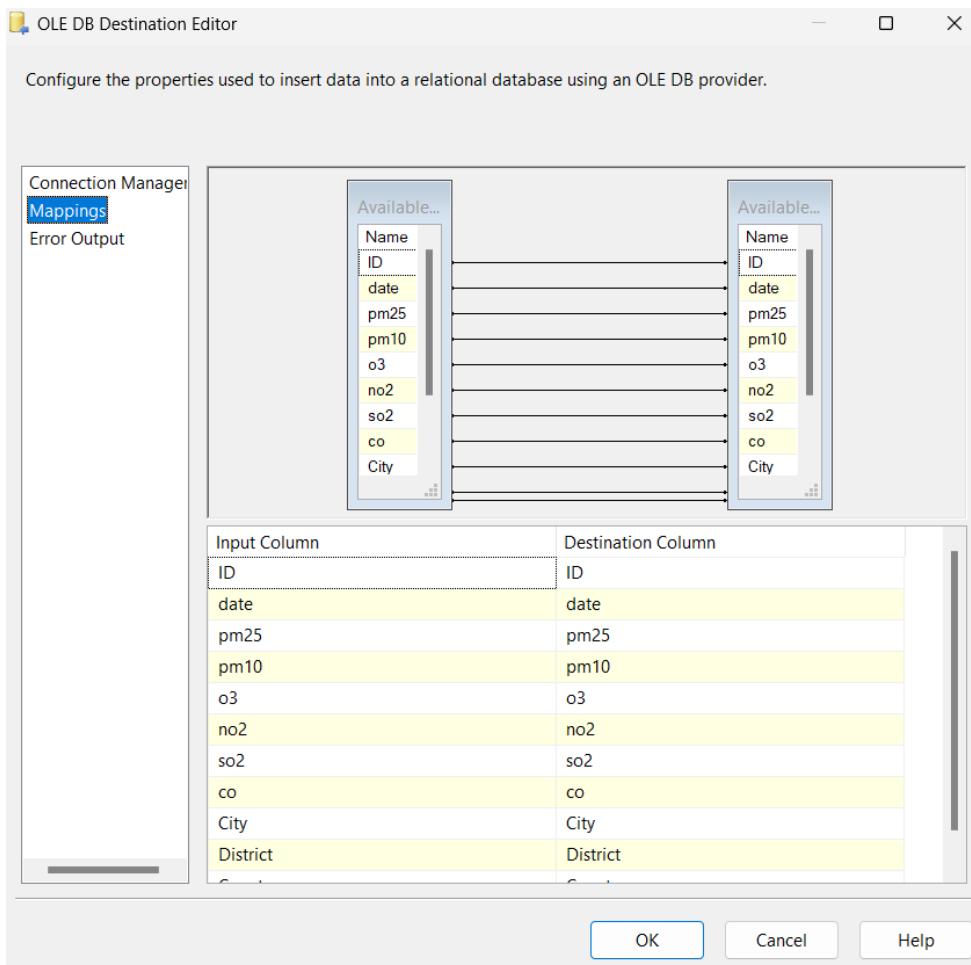
Ta cấu hình “OLE DB Destination” bằng cách chọn database là OriginalData và thực hiện việc tạo thêm một bảng tên là “Null\_data” để lưu giá trị null



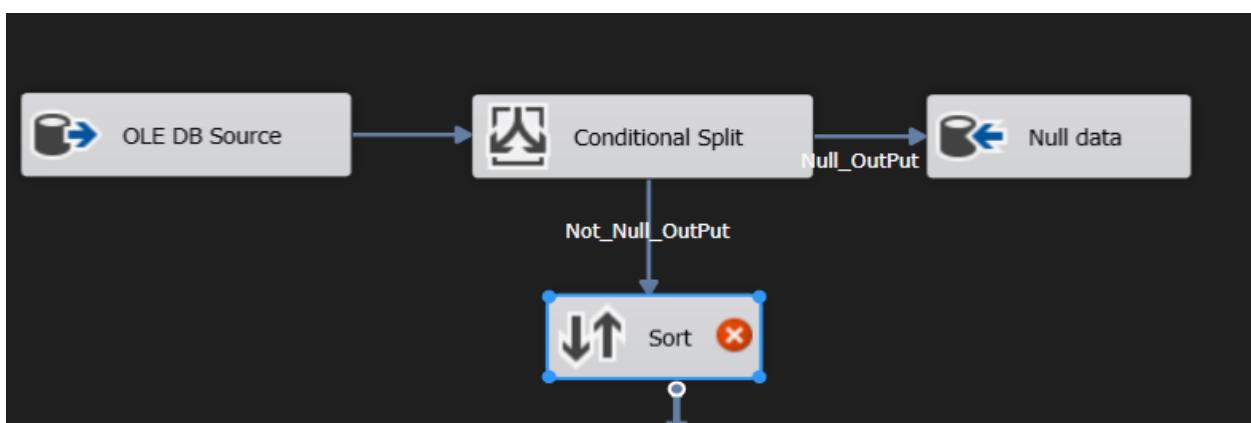
Ta thực hiện tạo bảng bằng cách chọn “new” và nhập lệnh SQL



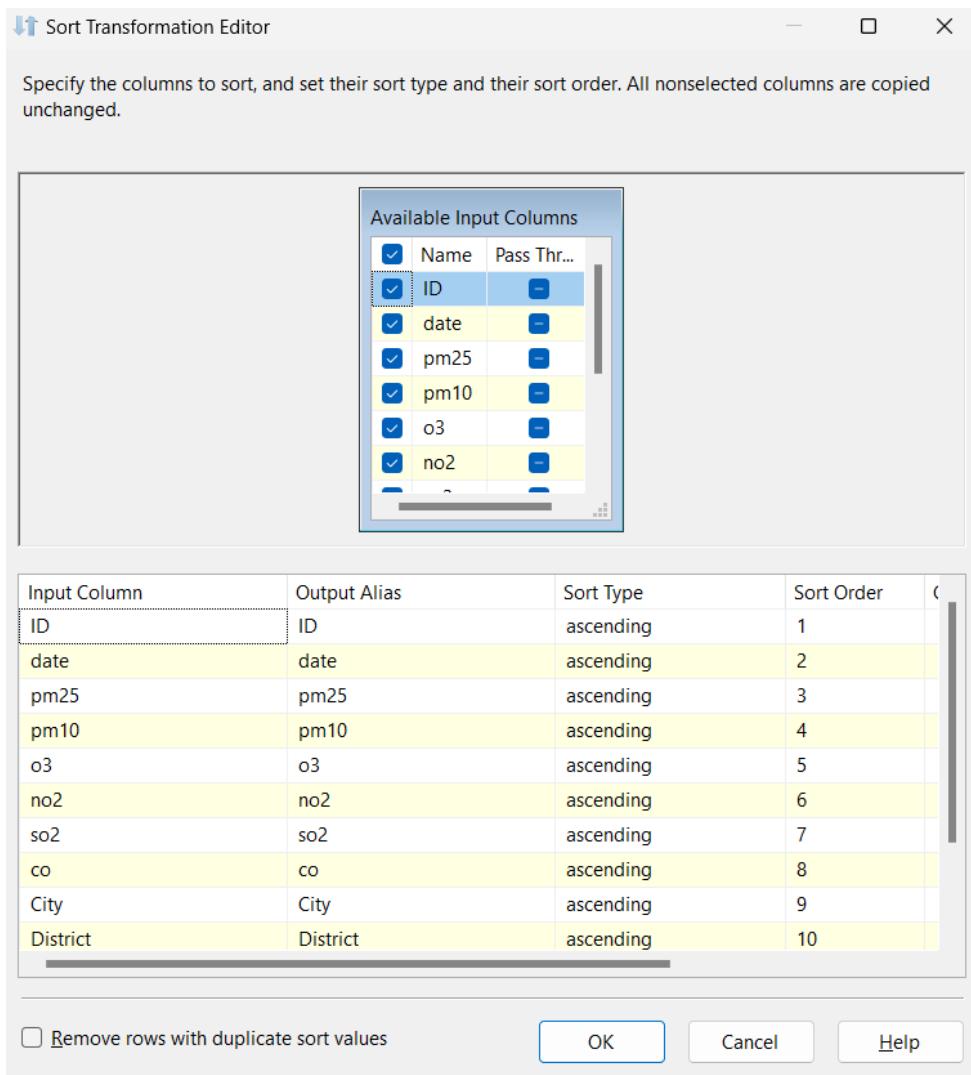
Tiếp theo ta kiểm tra mapping các thuộc tính và chọn "ok" nếu hoàn tất.



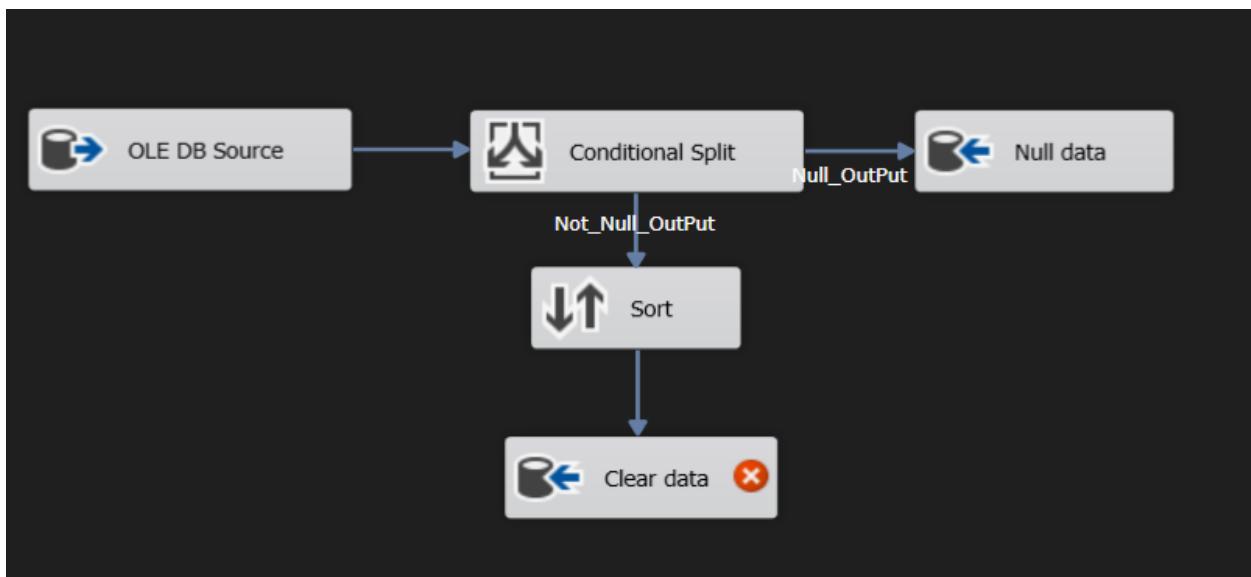
Bước 5: Tiếp theo ta tạo “Sort” với đầu vào là NotNull\_OutPut để lọc dữ liệu sạch .



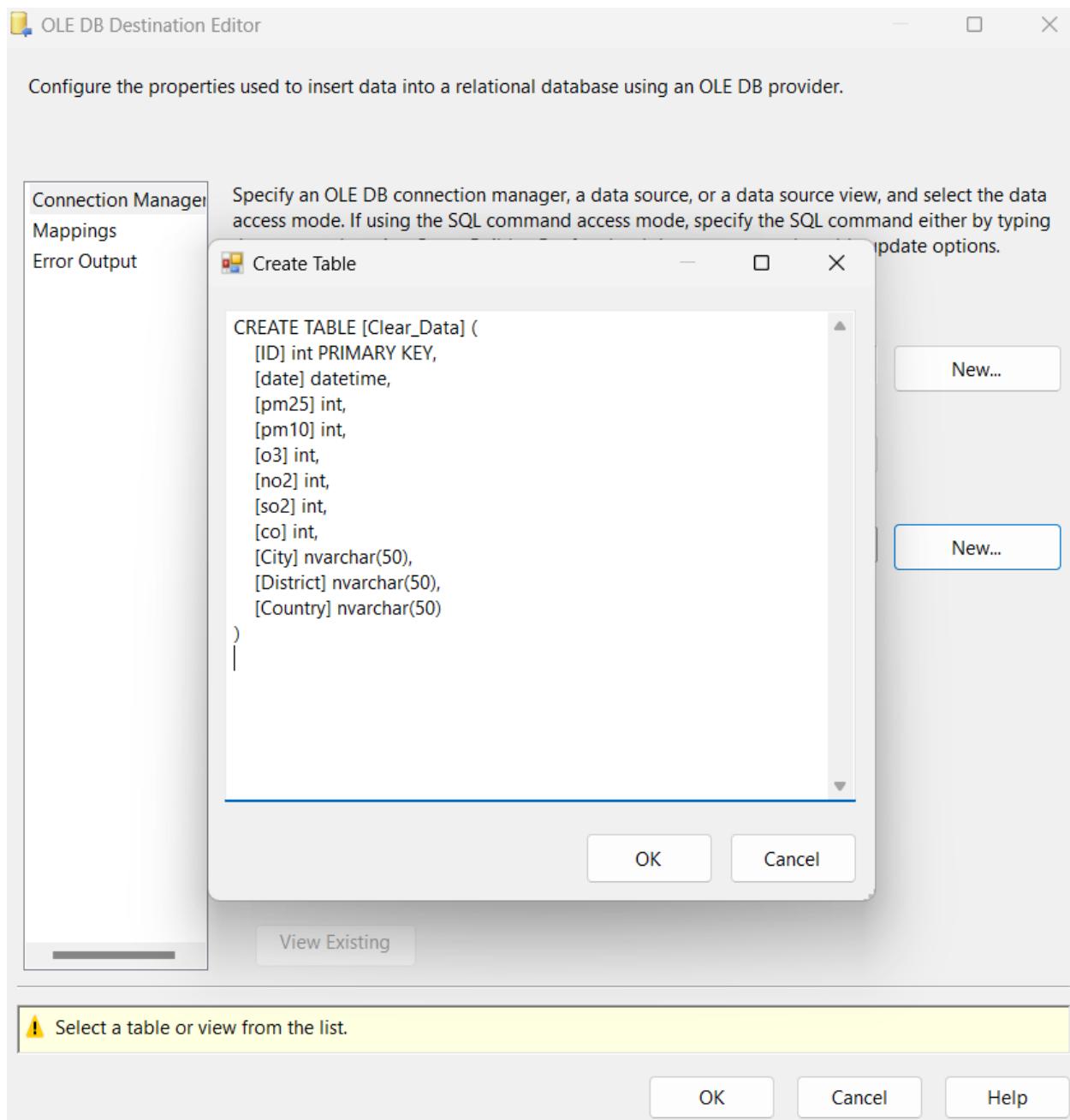
Ta thực hiện cấu hình cho “Sort” , ta chọn tất cả các thuộc tính và chọn “ok” để hoàn tất.



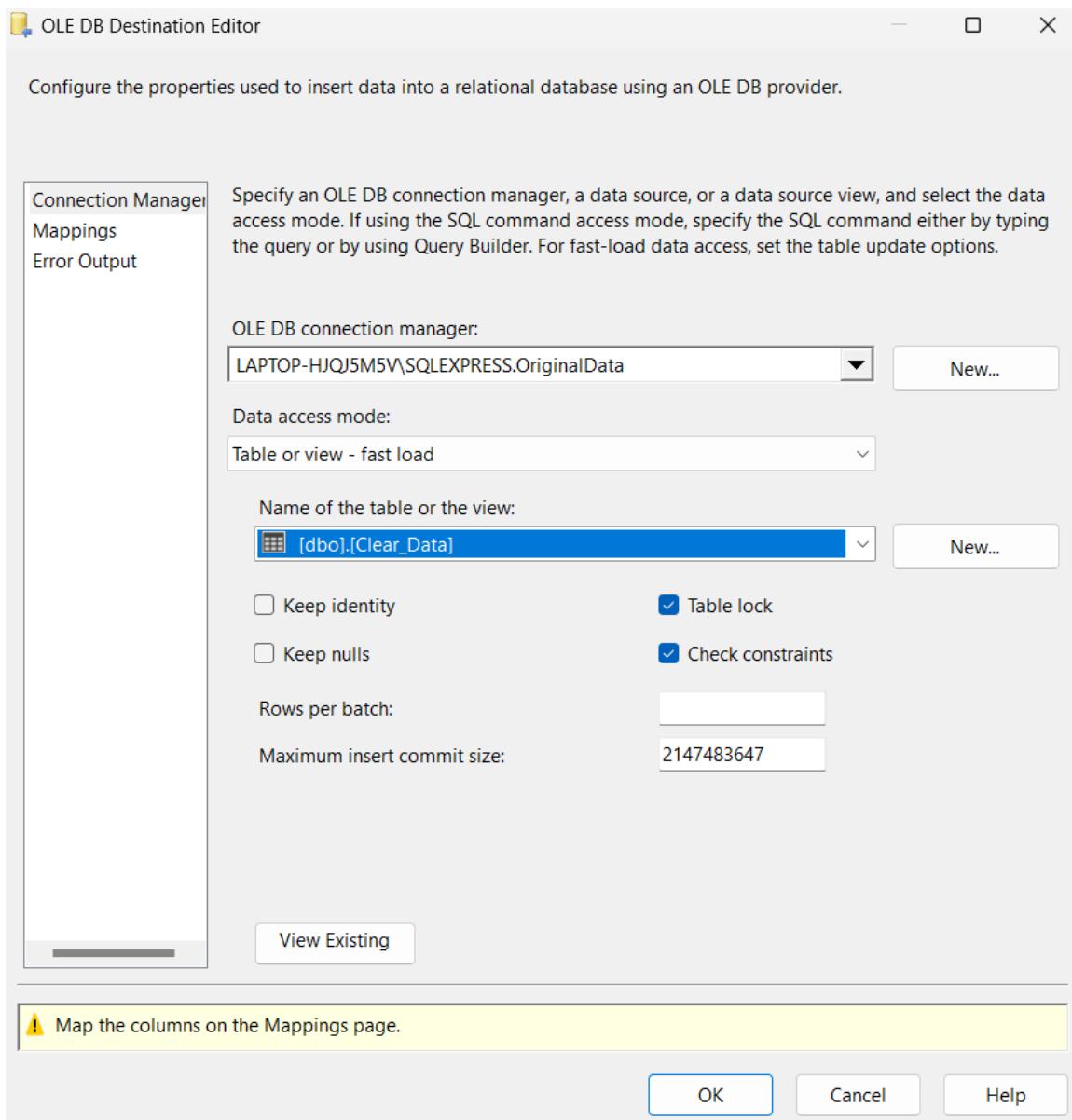
Bước 6 : Ta tạo thêm “OLE DB Destination” với tên là “Clear data” để lưu các giá trị Not\_Null\_OutPut.

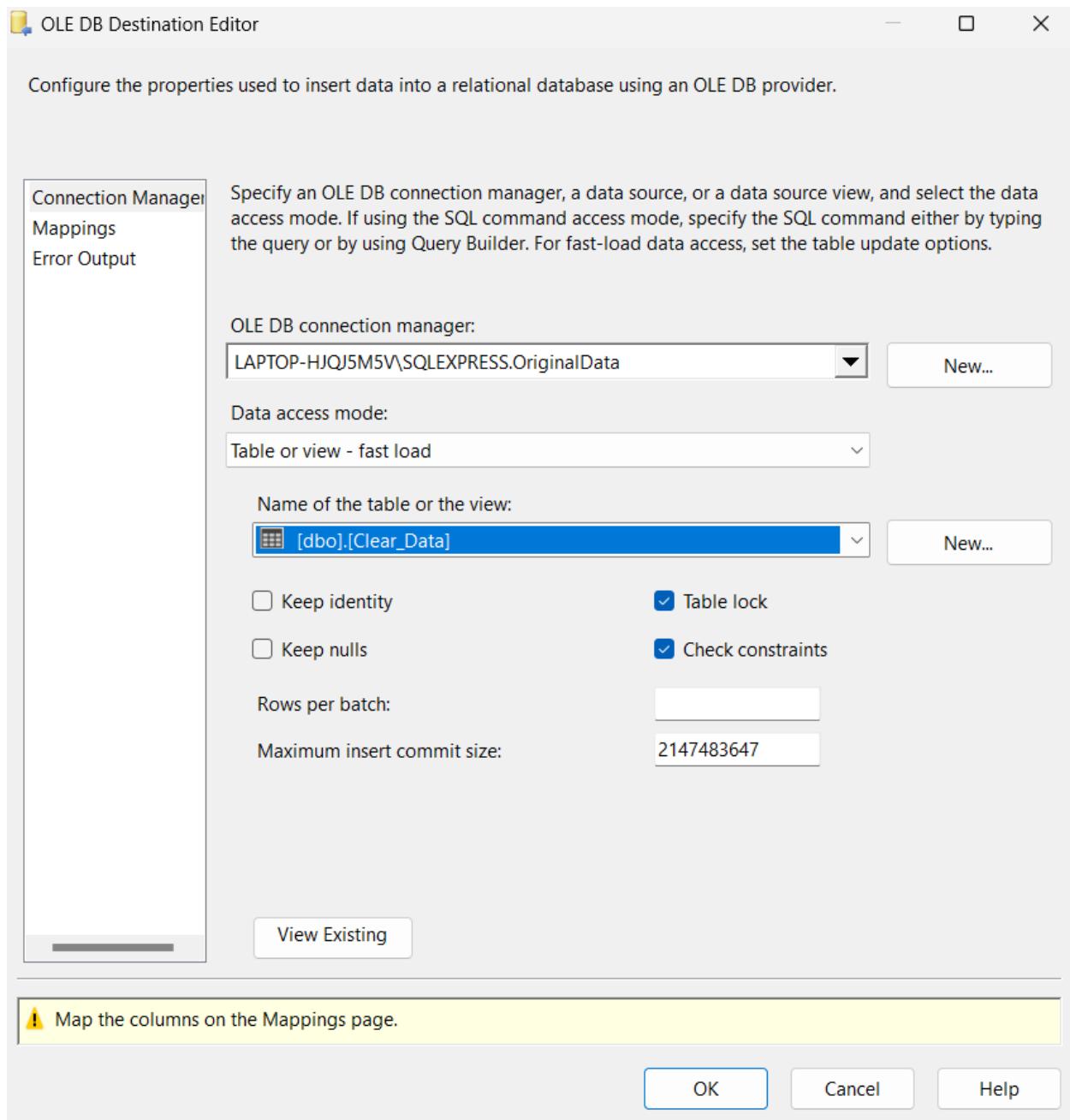


Ta thực hiện cấu hình tương tự như bên “Null data” , ta tạo bảng tên là “Clear\_data” để lưu giá trị sạch .

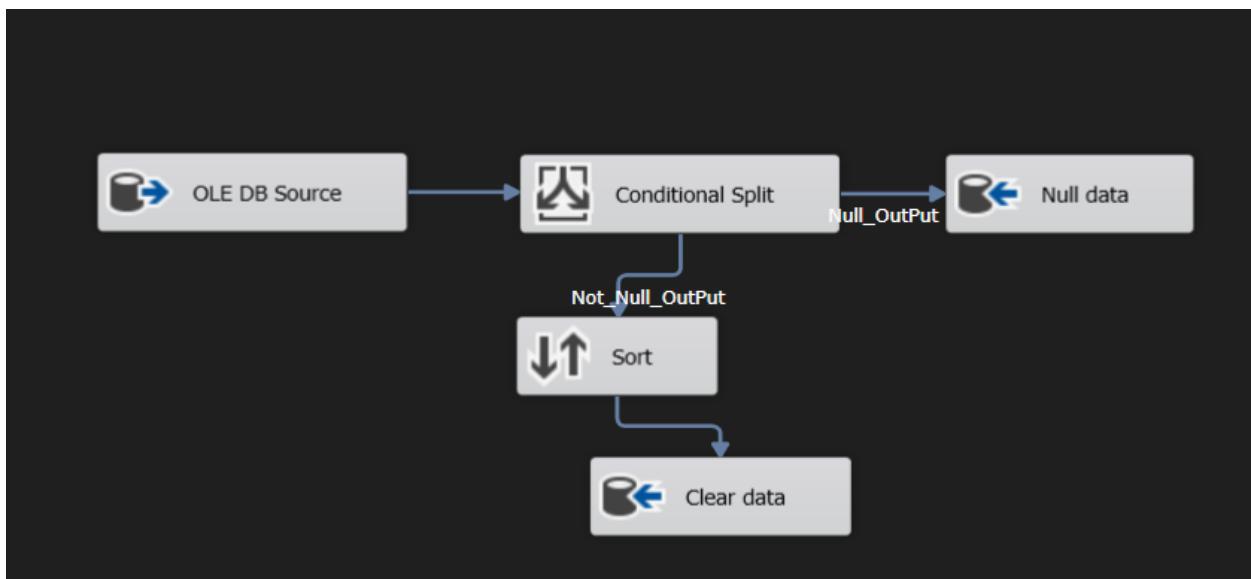


Sau khi tạo bảng ta được kết quả .



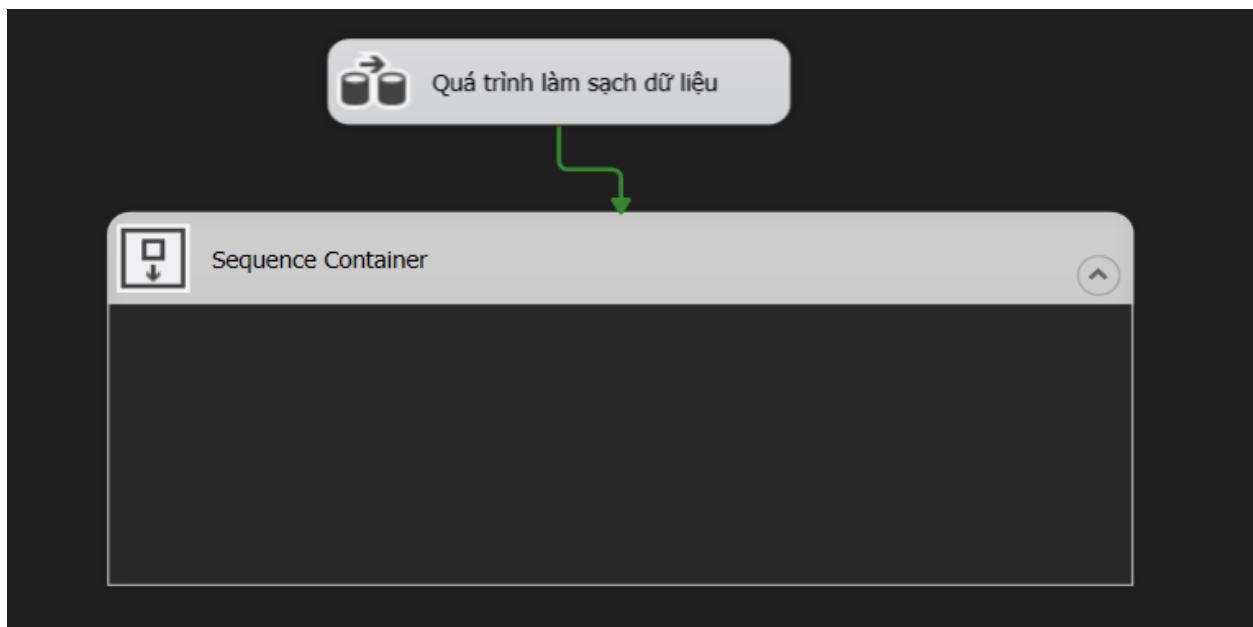


Data flow của quá trình làm sạch dữ liệu sau khi hoàn thành:



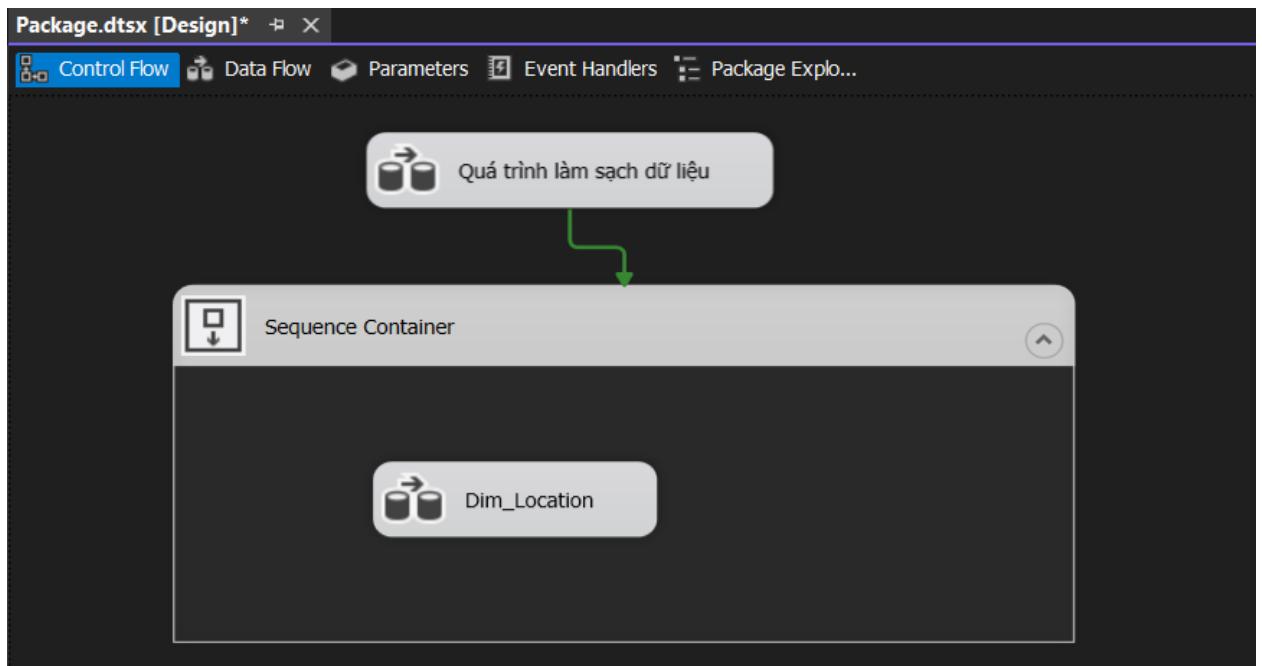
#### 4. Quá trình tạo các bảng Dimension

Trước tiên ta cần tạo “Sequence Container” để chứa các “Data Flow” là Dim\_Location và Dim\_Date .

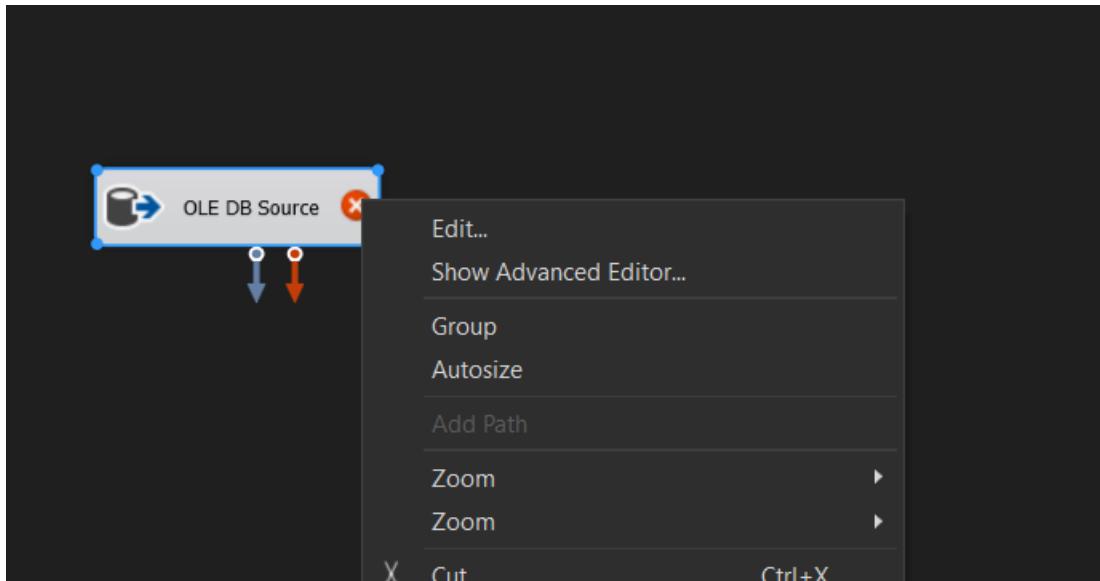


##### 4.1. Dim\_Location

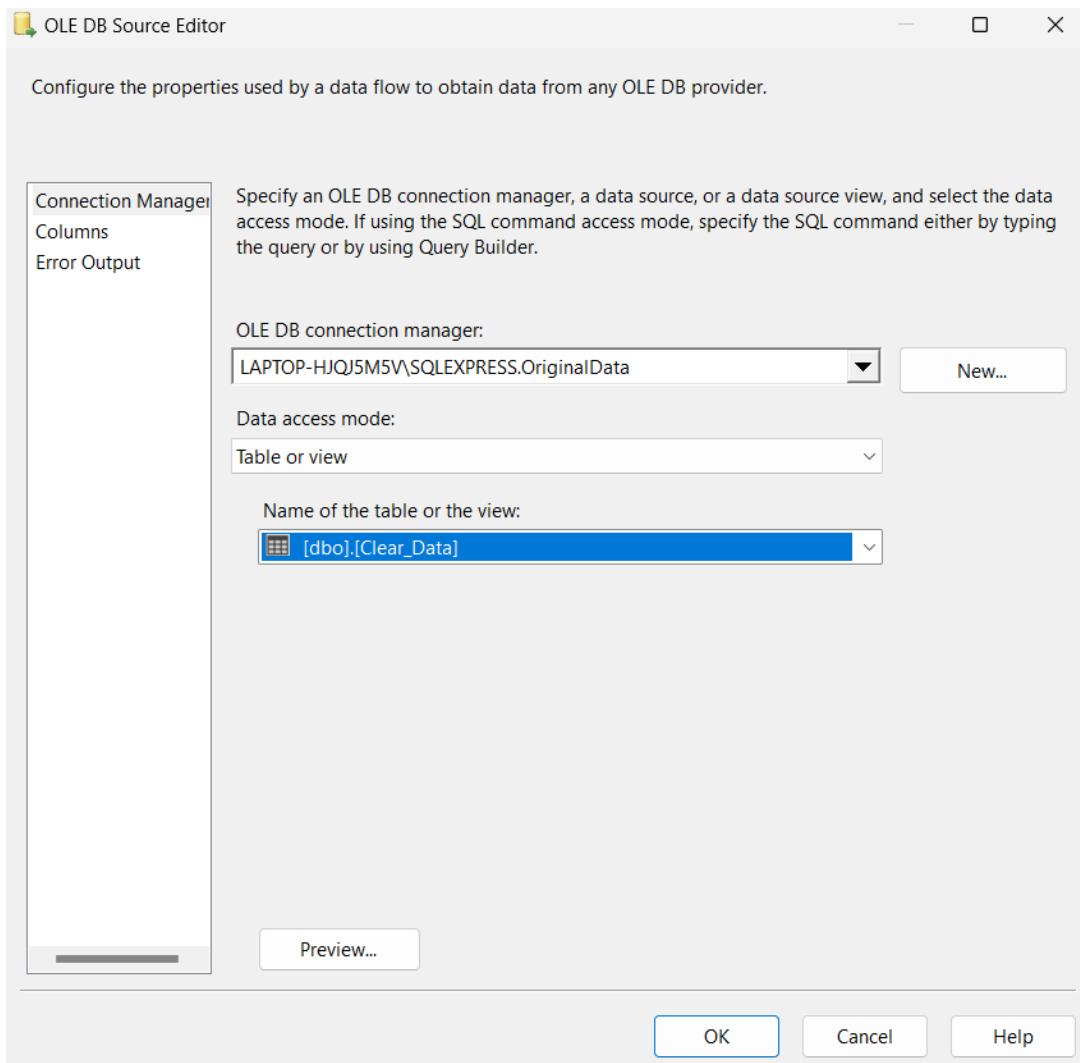
Bước 1 : Ta thực hiện tạo 1 “Data flow” tên là “Dim\_Location” và thực hiện việc chỉnh sửa .



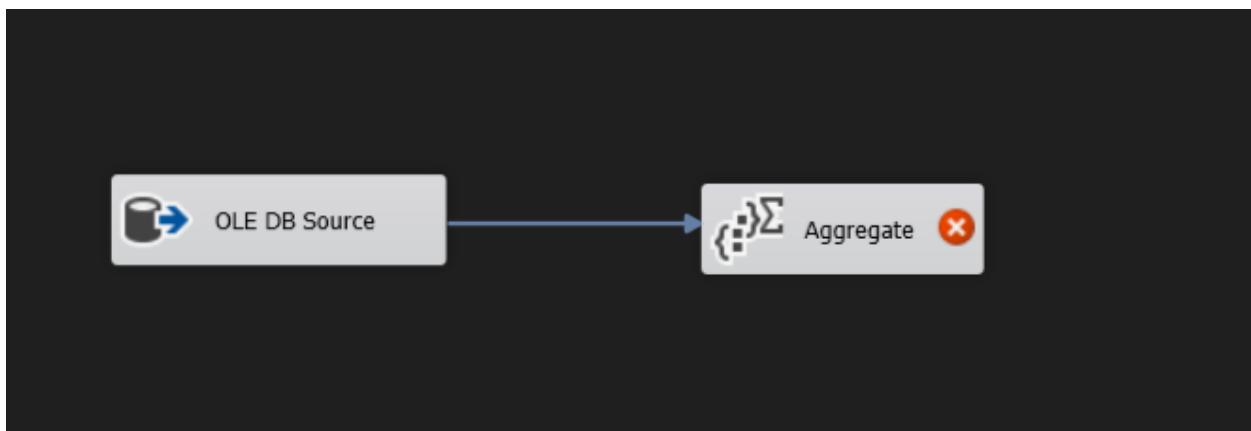
Bước 2 : Ta thực hiện chỉnh sửa “Data Flow Task” bằng cách ấn vào Edit . Ta thực hiện thêm “OLE DB Source” và cài đặt.



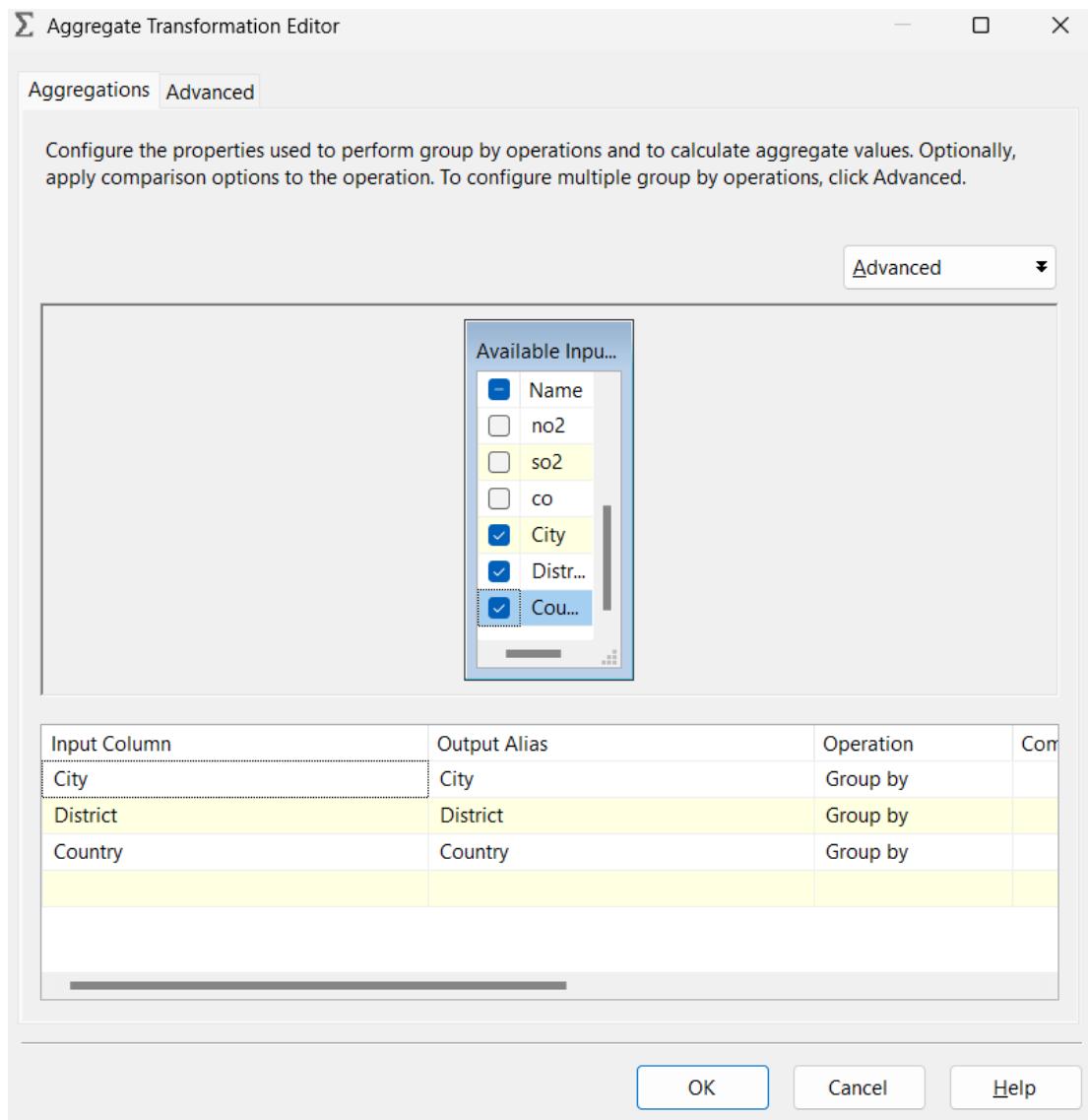
Ta thực hiện việc kết nối đến database OriginalData và chọn bảng dữ liệu “Clear\_data” để làm nguồn dữ liệu.



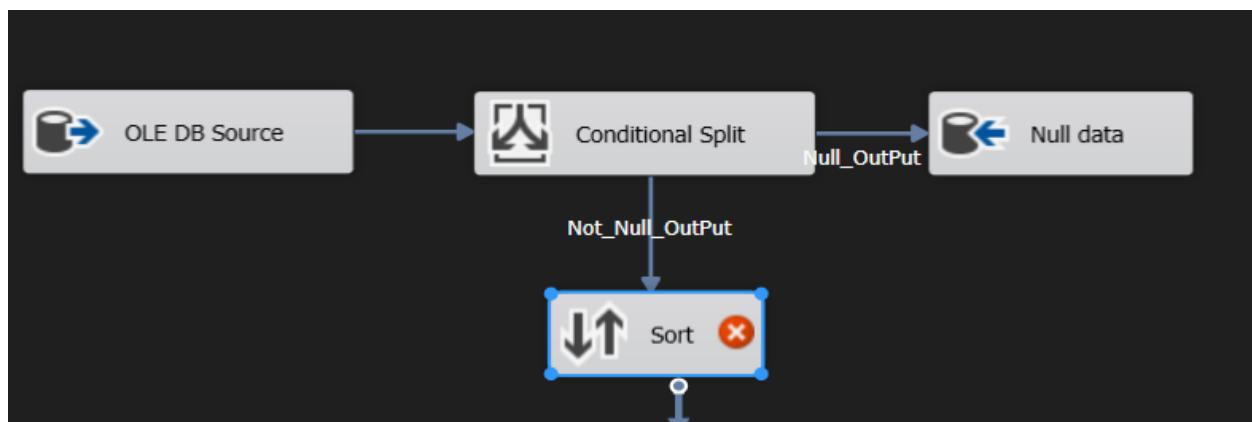
Bước 3: Ta thực hiện việc tạo "Aggregate"



Ta chọn các thuộc tính cho bảng Dim\_Location .



Bước 4: Tiếp theo ta tạo “Sort” và chọn Edit để chỉnh sửa.



Ta thực hiện cấu hình cho “Sort”, ta chọn tất cả các thuộc tính và chọn “ok” để hoàn tất.

Sort Transformation Editor

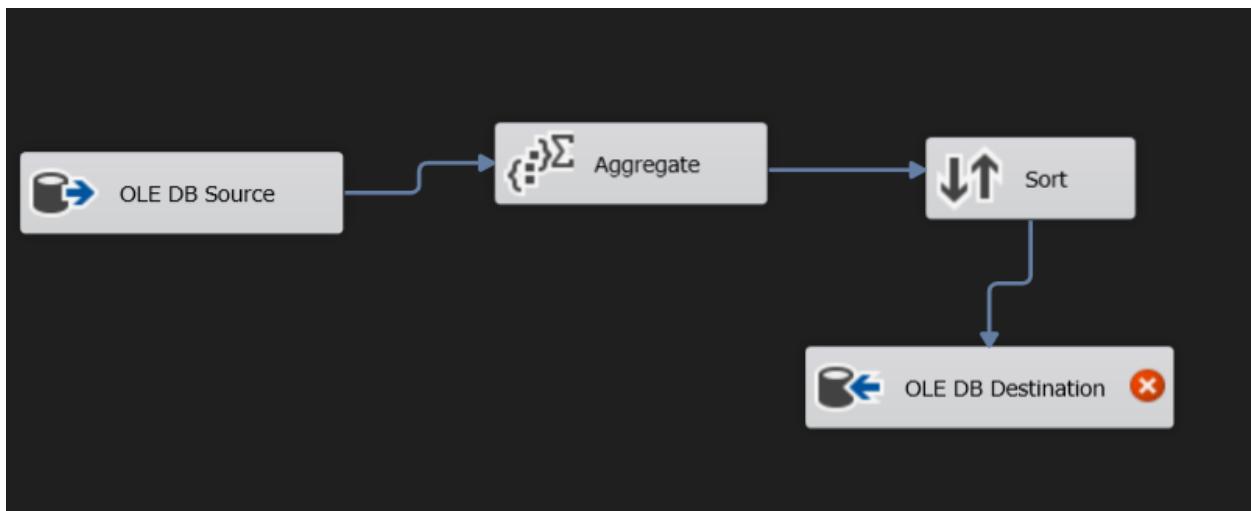
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Input Column	Output Alias	Sort Type	Sort Order	Condition
City	City	ascending	1	
District	District	ascending	2	
Country	Country	ascending	3	

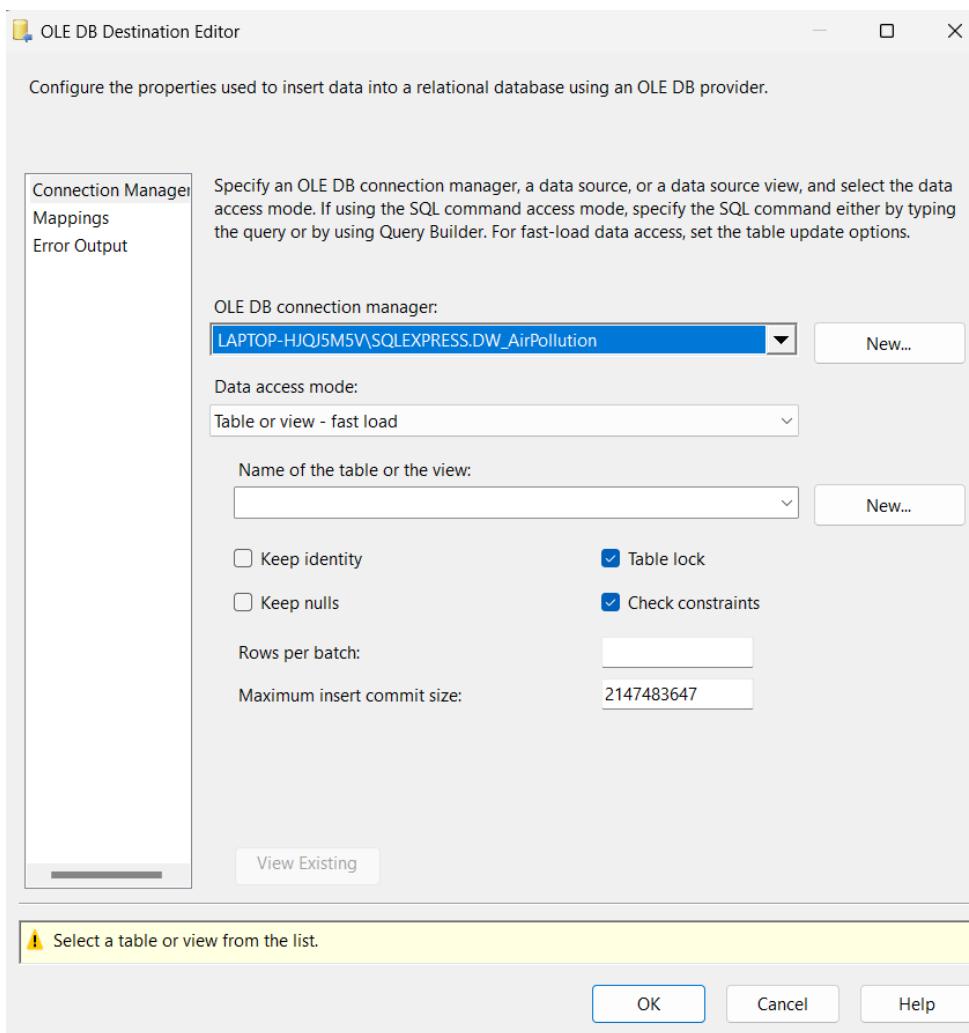
Remove rows with duplicate sort values

OK Cancel Help

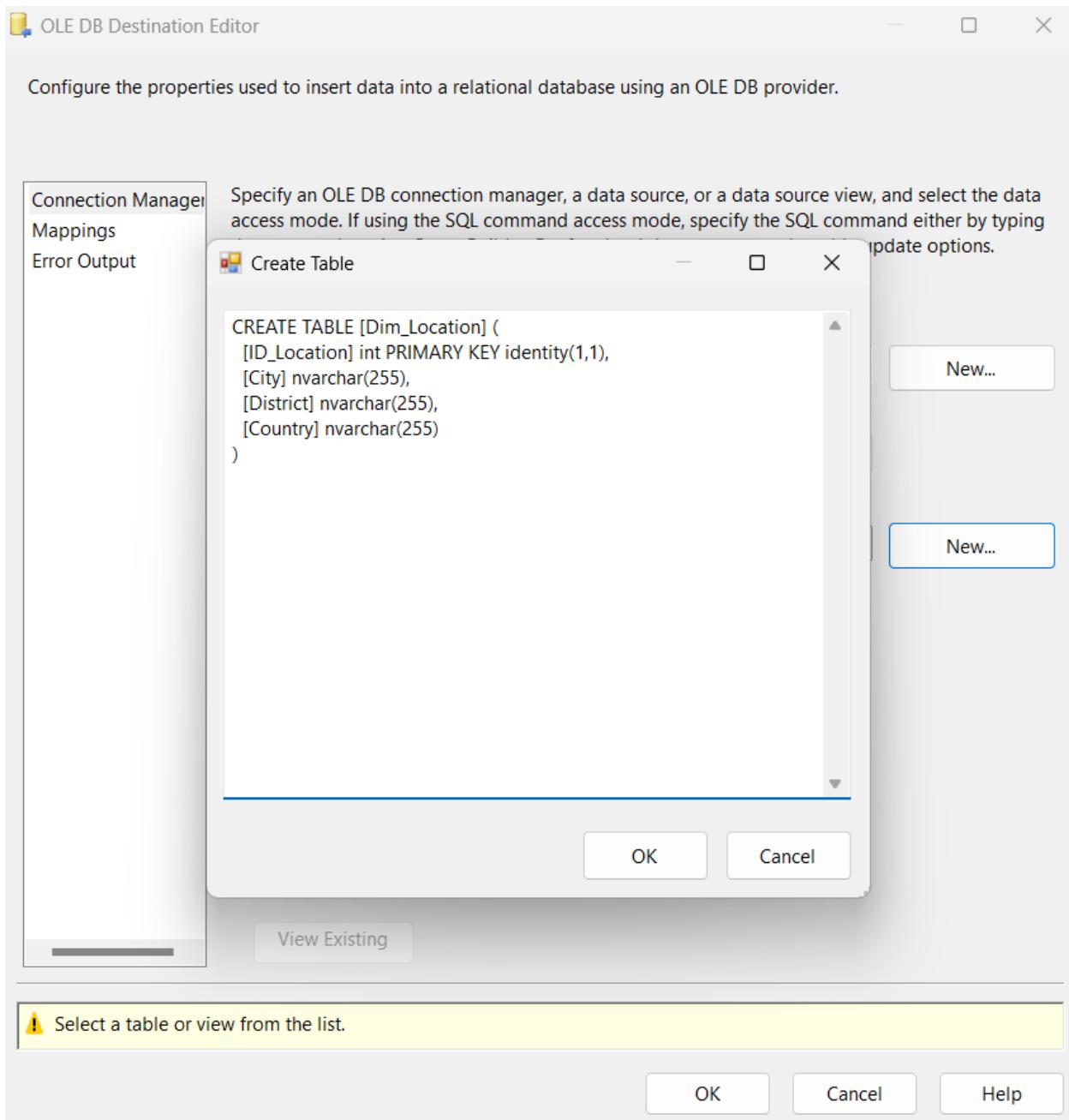
Bước 5: Cuối cùng ta tạo “OLE DB Destination” .



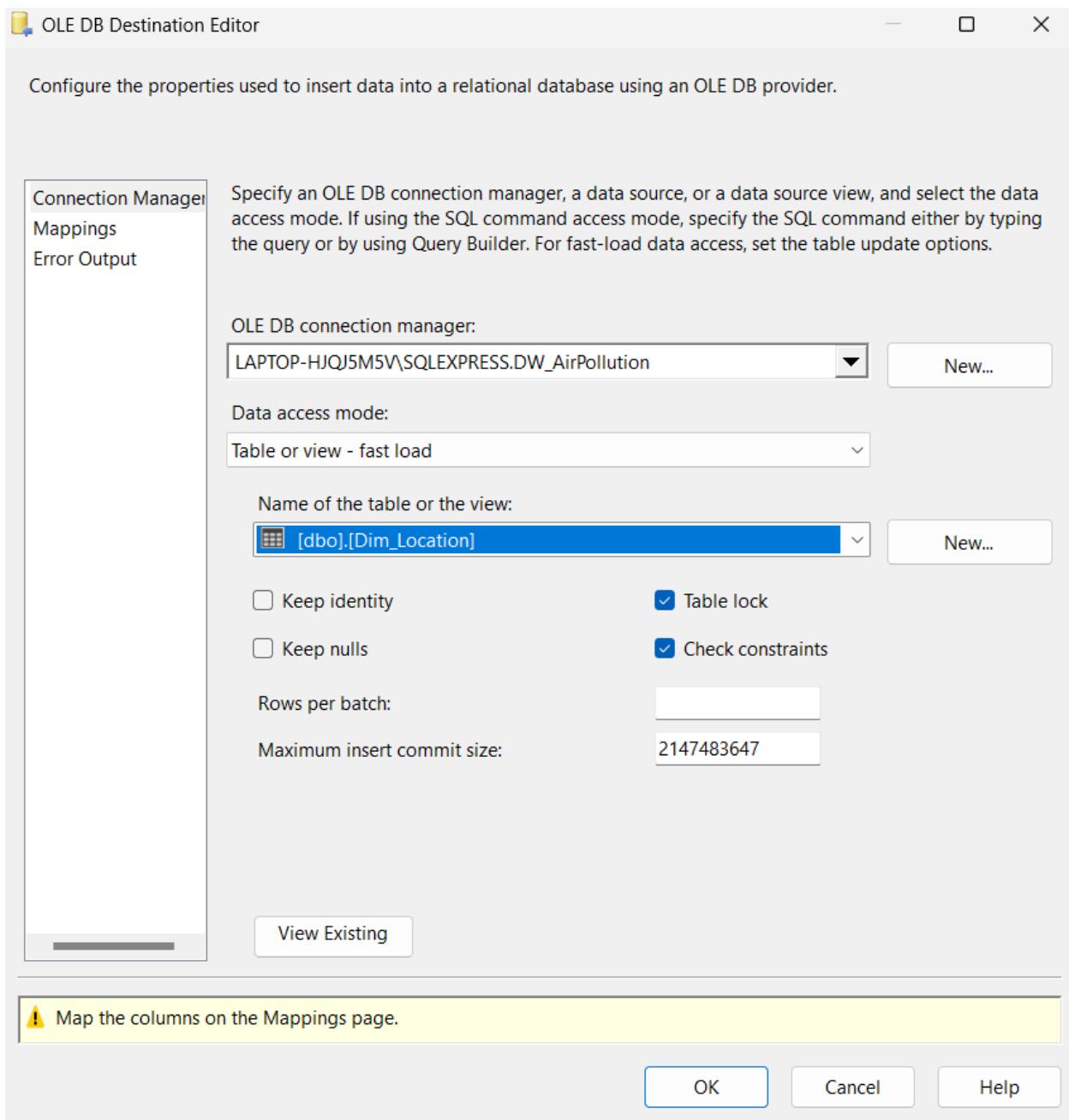
Ta thực hiện việc kết nối đến database “DW\_AirPollution” .



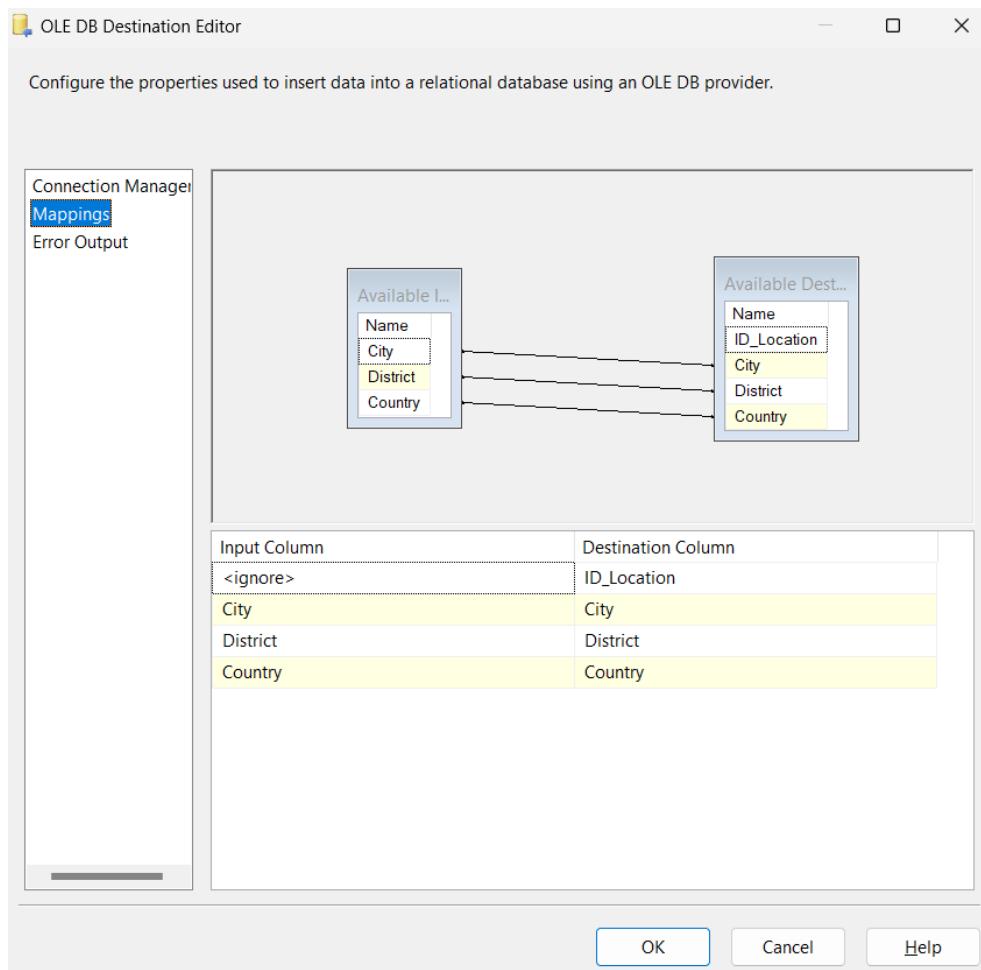
Tiếp đến ta thực hiện việc tạo mới table "Dim\_Location" và chạy lệnh SQL để tạo.



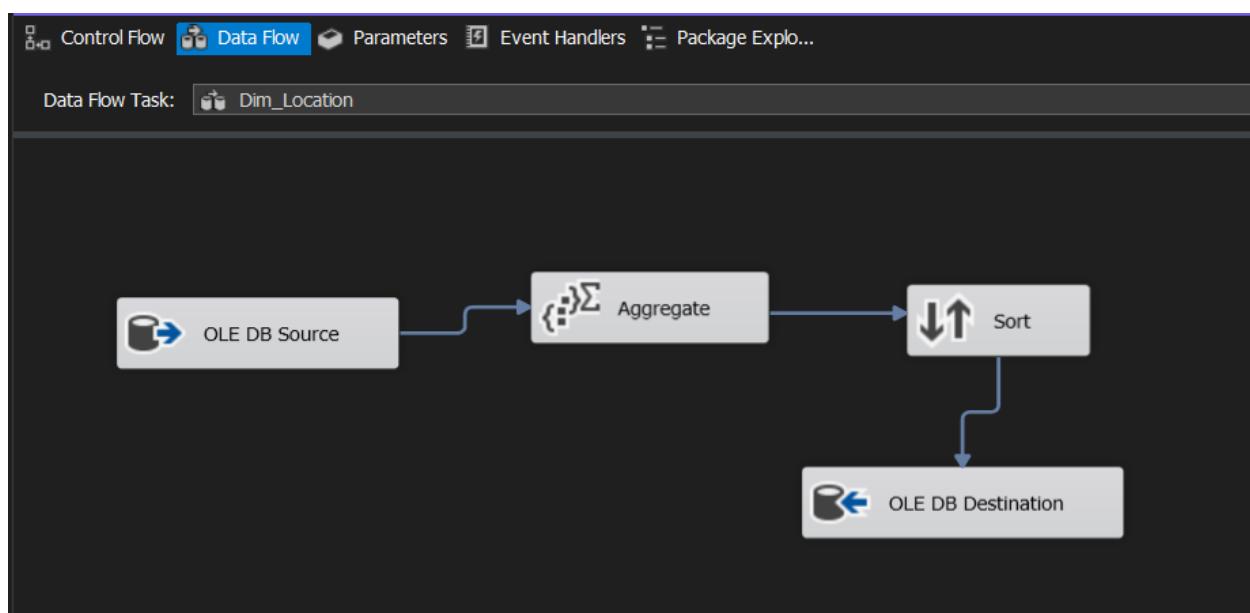
Sau khi tạo hoàn tất , ta chọn "Dim\_Location"



Cuối cùng ta thực hiện kiểm tra mapping và chọn "ok" để hoàn tất .

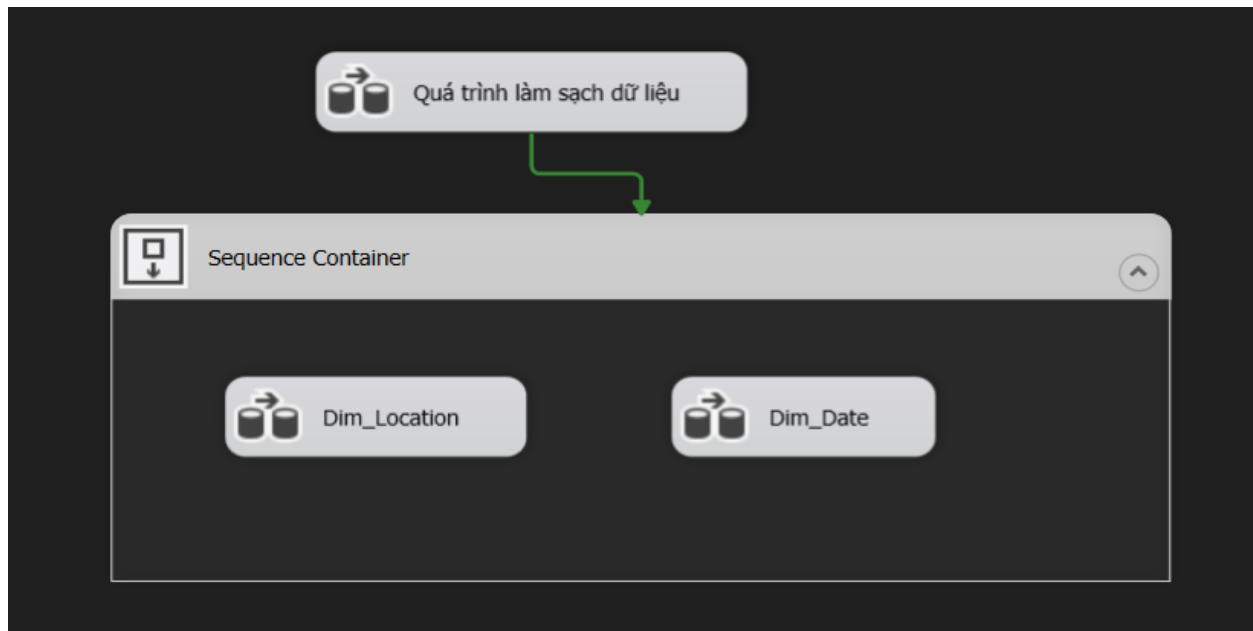


Data Flow của quá trình tạo bảng Dim\_Location sau khi hoàn thành.

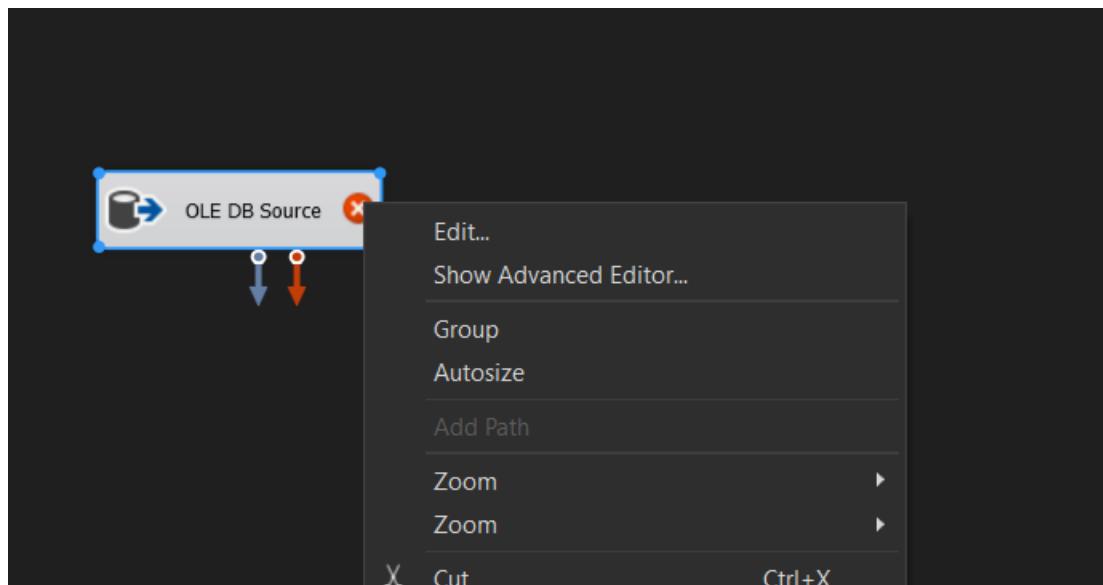


#### 4.2. Dim\_Date

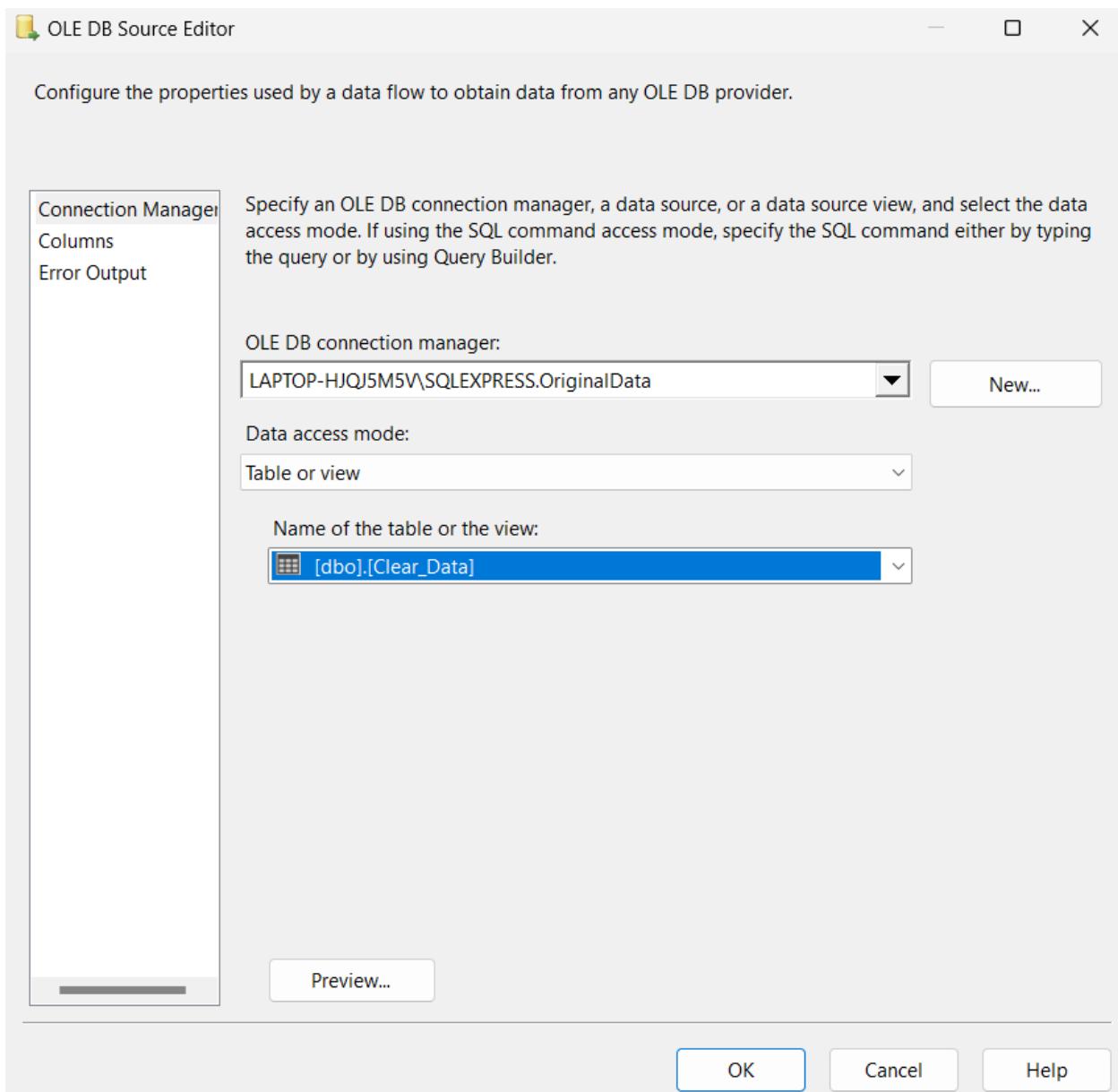
Bước 1 : Ta thực hiện tạo 1 “Data flow Task” tên là “Dim\_Date” và thực hiện việc chỉnh sửa .



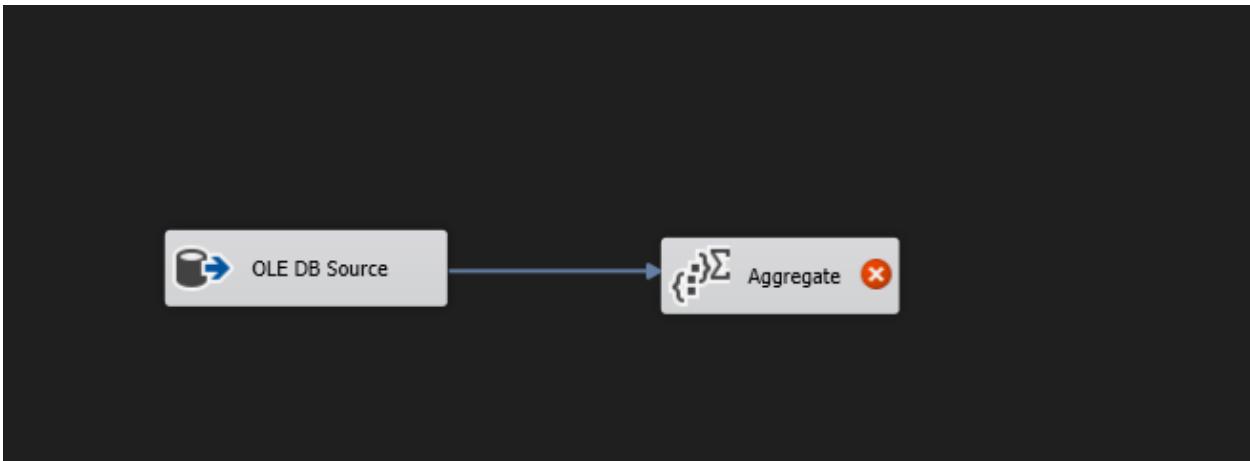
Bước 2 : Ta thực hiện chỉnh sửa “Data Flow Task” bằng cách ấn vào Edit . Ta thực hiện thêm “OLE DB Source” và cài đặt.



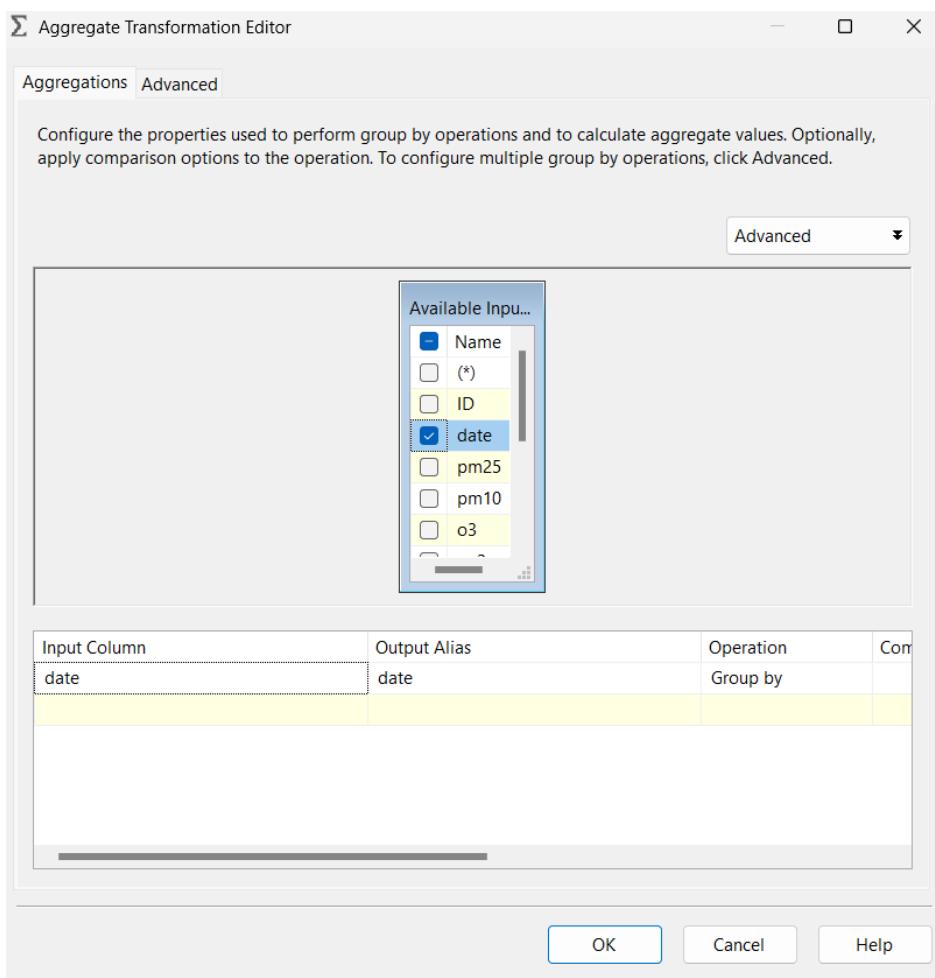
Ta thực hiện việc kết nối đến database OriginalData và chọn bảng dữ liệu “Clear\_data” để làm nguồn dữ liệu.



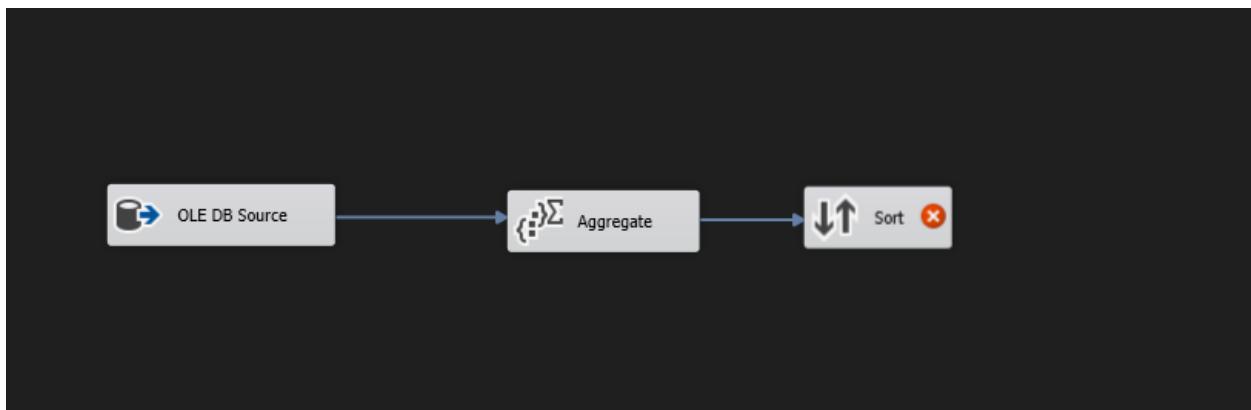
Bước 3: Ta thực hiện việc tạo “Aggregate”



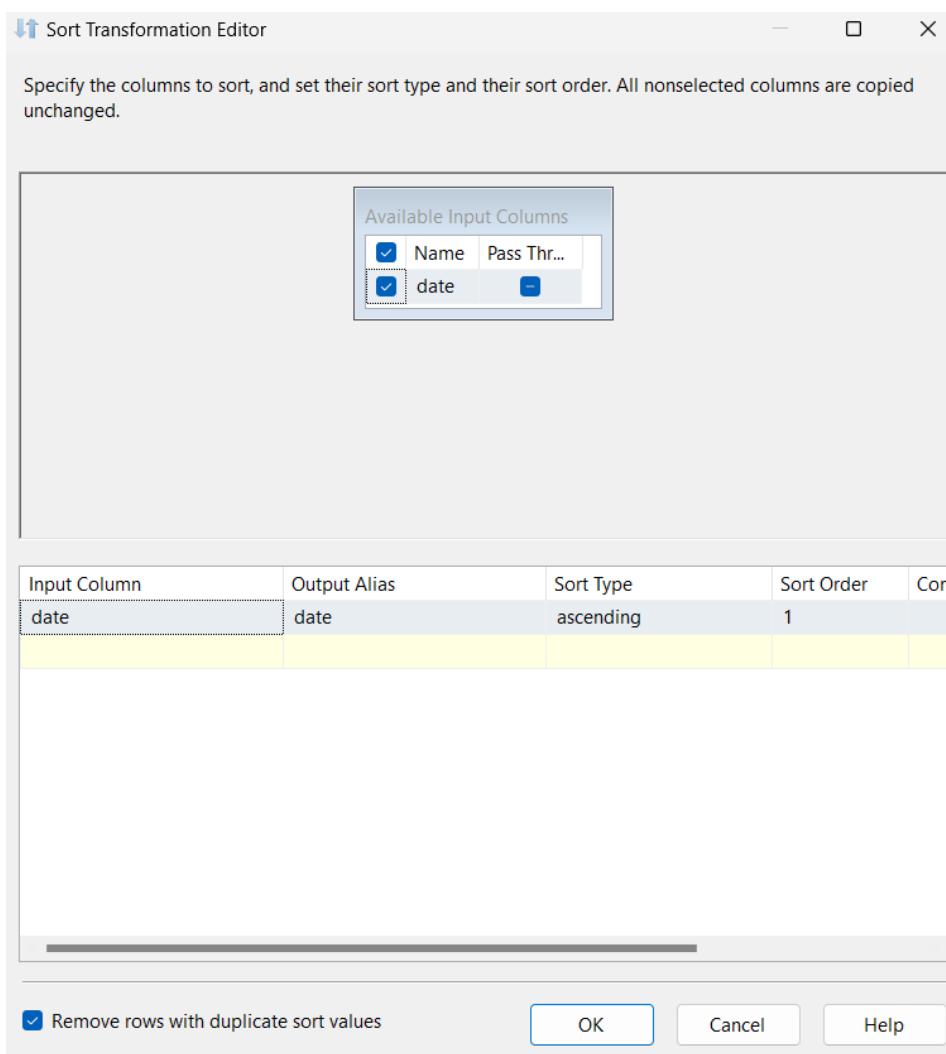
Ta chọn các thuộc tính cho "Dim\_Date"



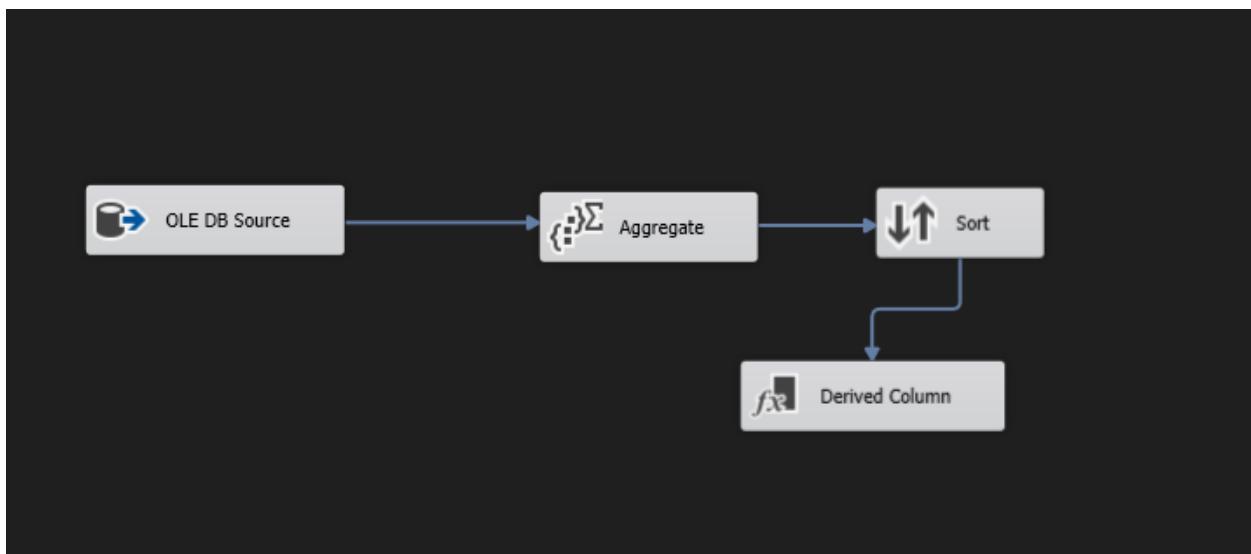
Bước 4: Ta tạo mới "Sort "



Ta thực hiện cấu hình cho "Sort" , ta chọn tất cả các thuộc tính và chọn "ok" để hoàn tất.



Bước 5 : Ta tạo “Derived Column” để có thể chia dữ liệu cho cột “Day”, “Month”, “Year”



Ta sử dụng hàm “DAY”, “MONTH”, “YEAR” để có thể lấy được dữ liệu ngày , tháng ,năm từ “date”.

Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Derived Column Name	Derived Column	Expression	Data Type	Length
Day	<add as new column>	DAY( [date] )	four-byte signed integer	
Month	<add as new column>	MONTH( [date] )	four-byte signed integer	
Year	<add as new column>	YEAR( [date] )	four-byte signed integer	

Variables and Parameters

Columns

Date/Time Functions

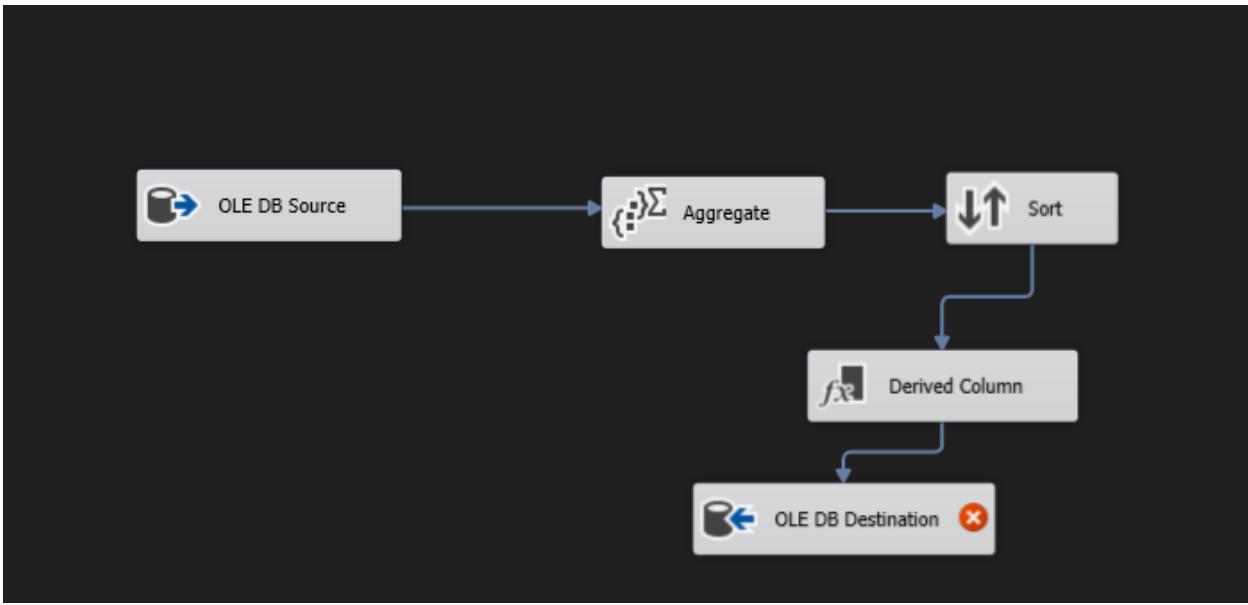
- DATEADD( «datepart», «number», «date» )
- DATEDIFF( «datepart», «startdate», «enddate» )
- DATEPART( «datepart», «date» )
- DAY( «date» )
- GETDATE()
- GETUTCDATE()
- MONTH( «date» )
- YEAR( «date» )

NULL Functions

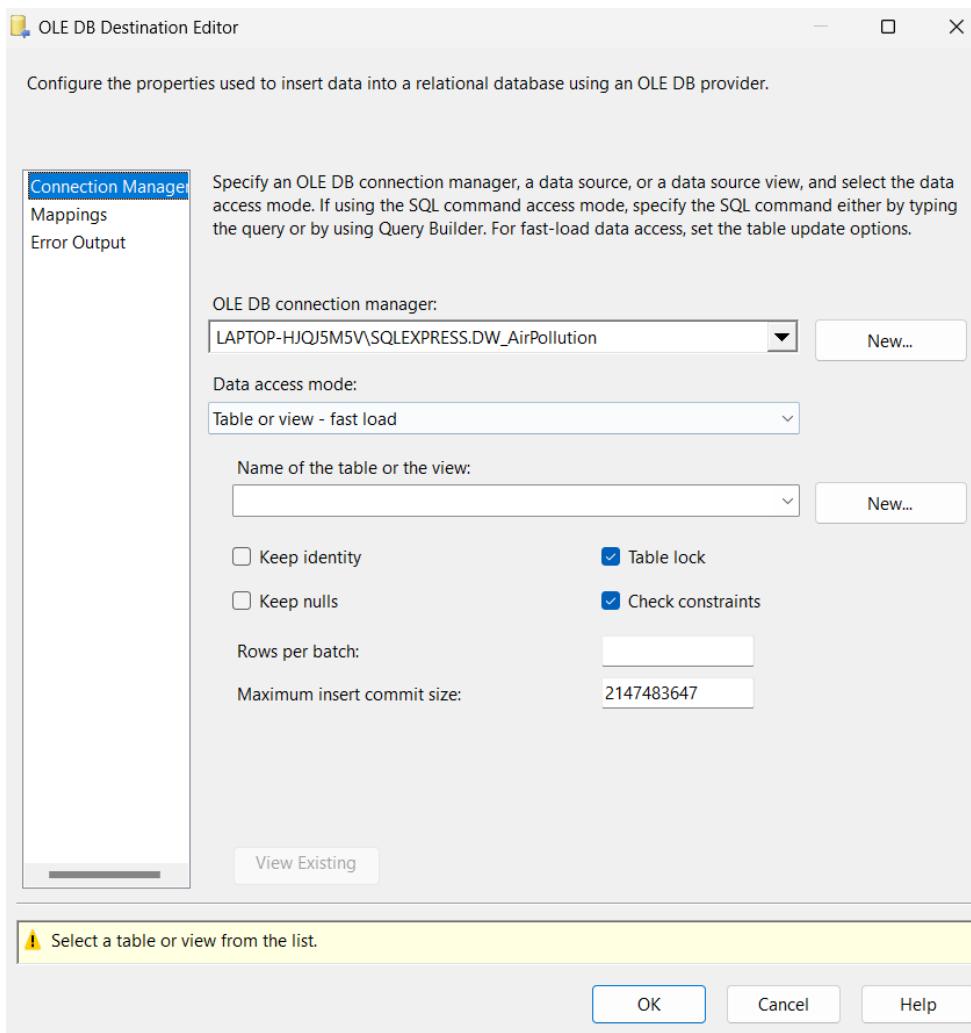
Description:

Configure Error Output... OK Cancel Help

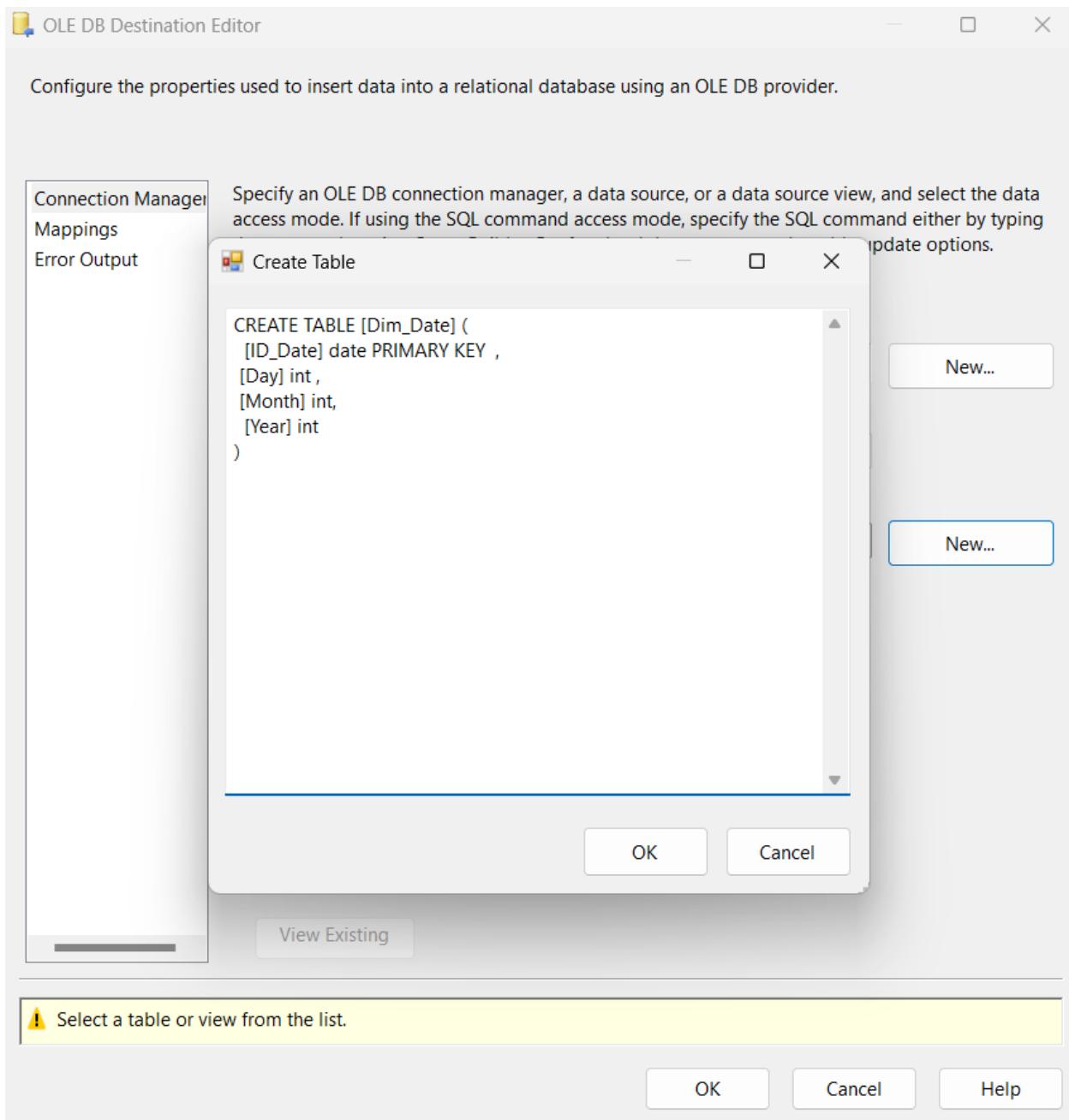
Bước 6: Cuối cùng ta tạo “OLE DB Destination” .



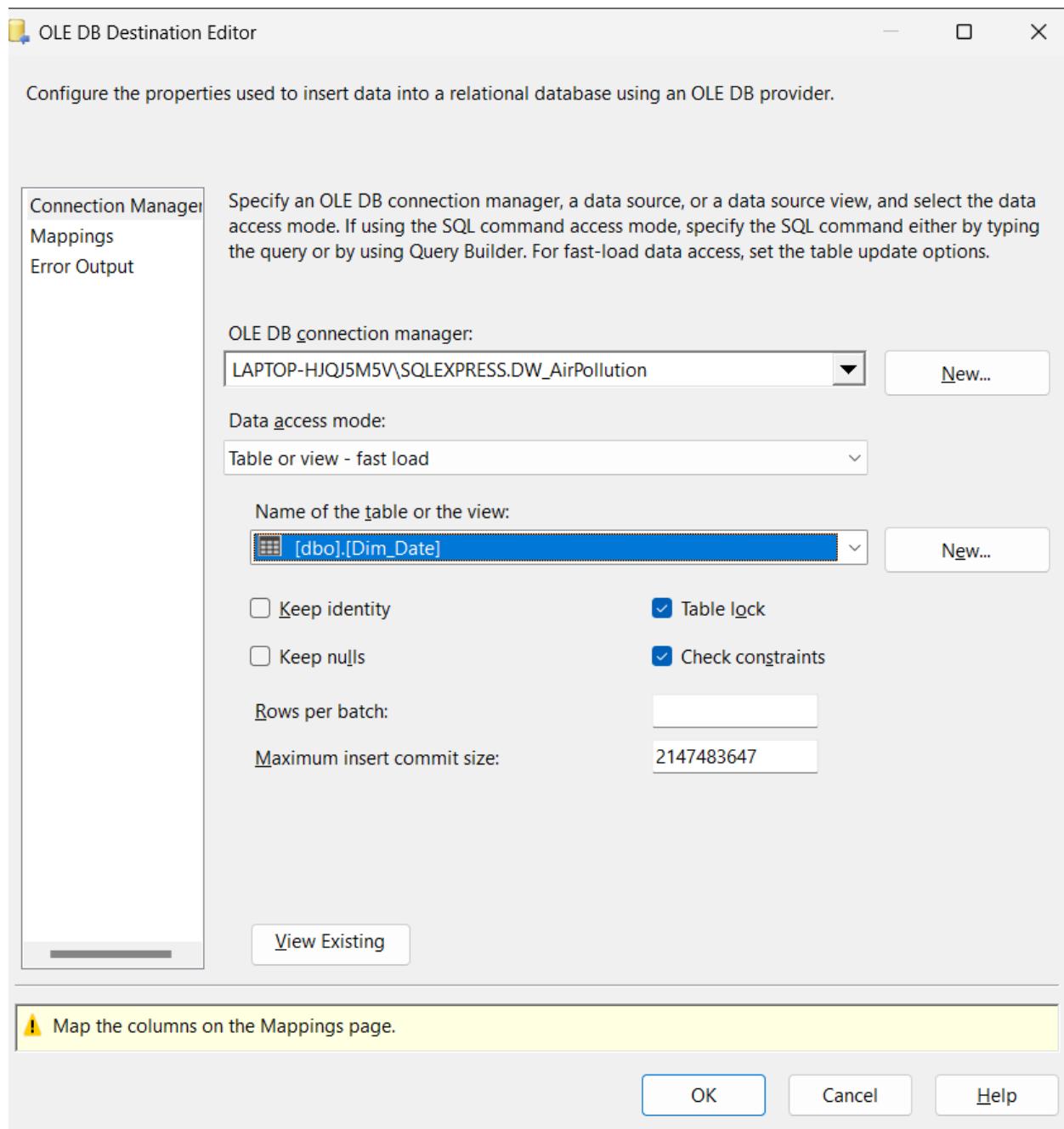
Ta thực hiện việc kết nối đến database “DW\_AirPollution” .



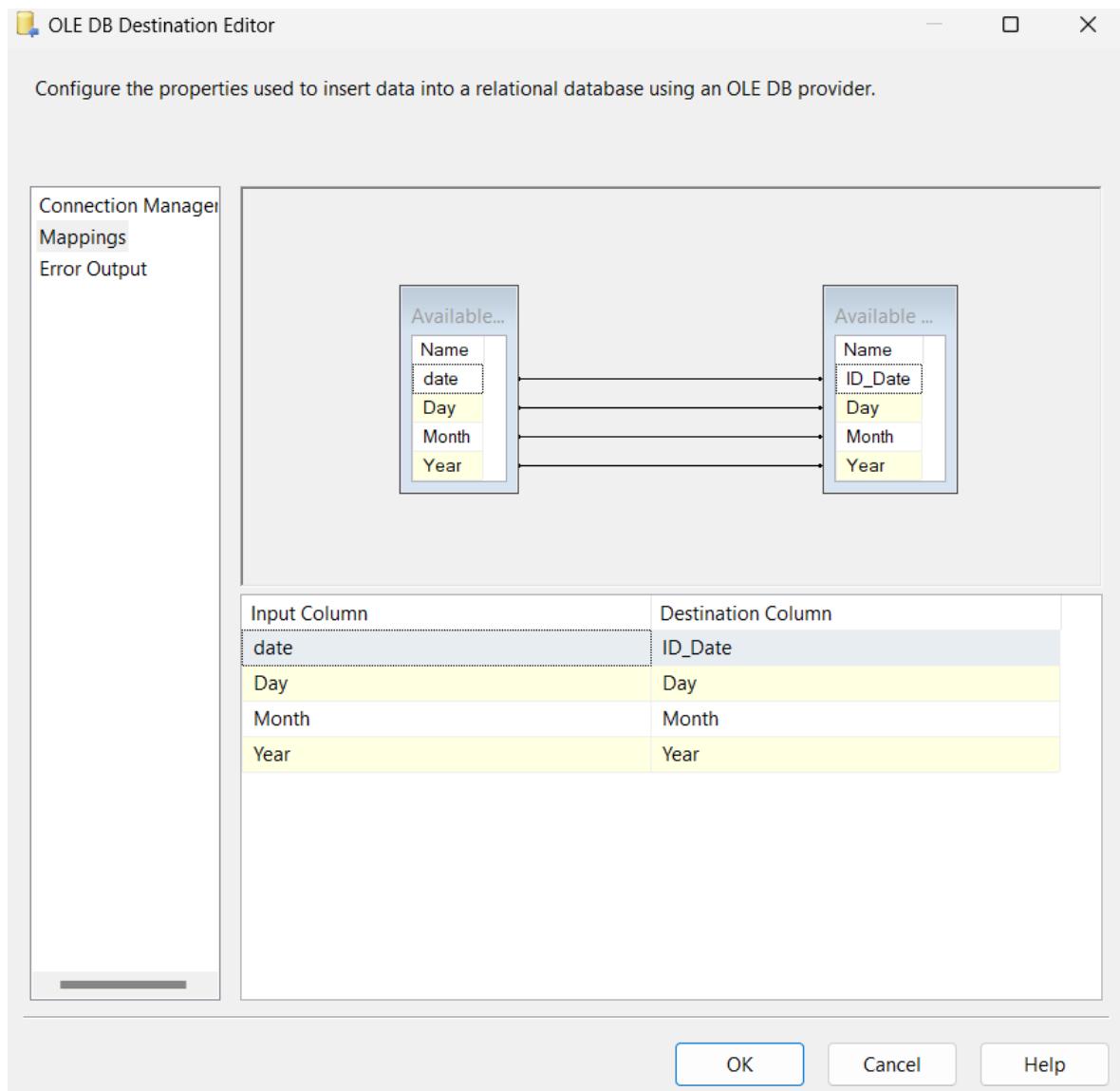
Tiếp đến ta thực hiện việc tạo mới table "Dim\_Date" và chạy lệnh SQL để tạo.



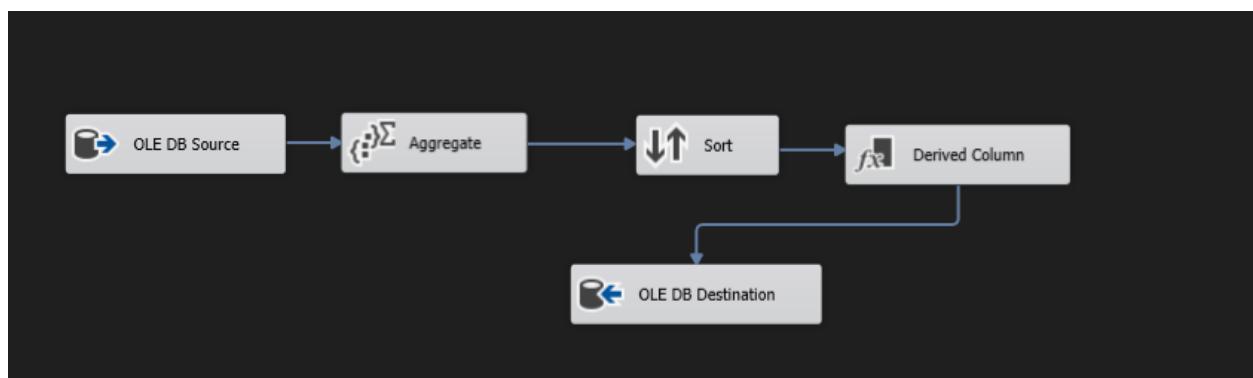
Sau khi tạo hoàn tất , ta chọn "Dim\_Date"



Cuối cùng ta thực hiện kiểm tra mapping và chọn "ok" để hoàn tất .

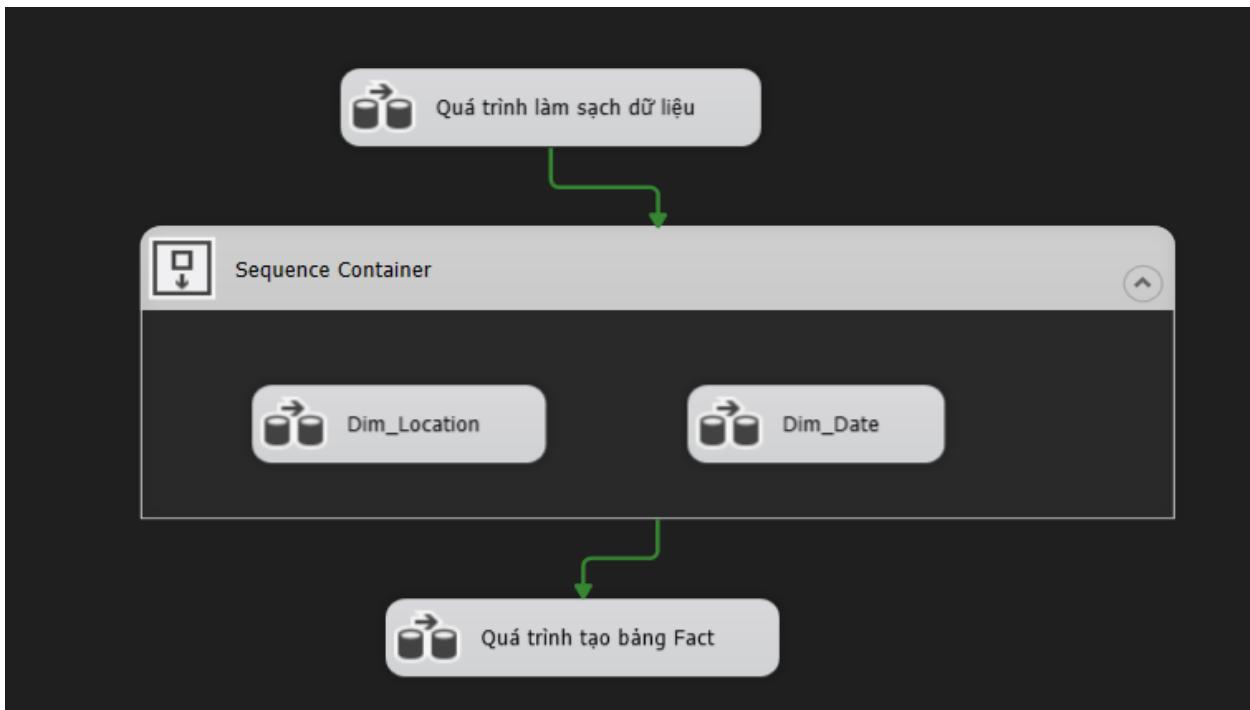


Data Flow của quá trình tạo bảng Dim\_Date sau khi hoàn thành.

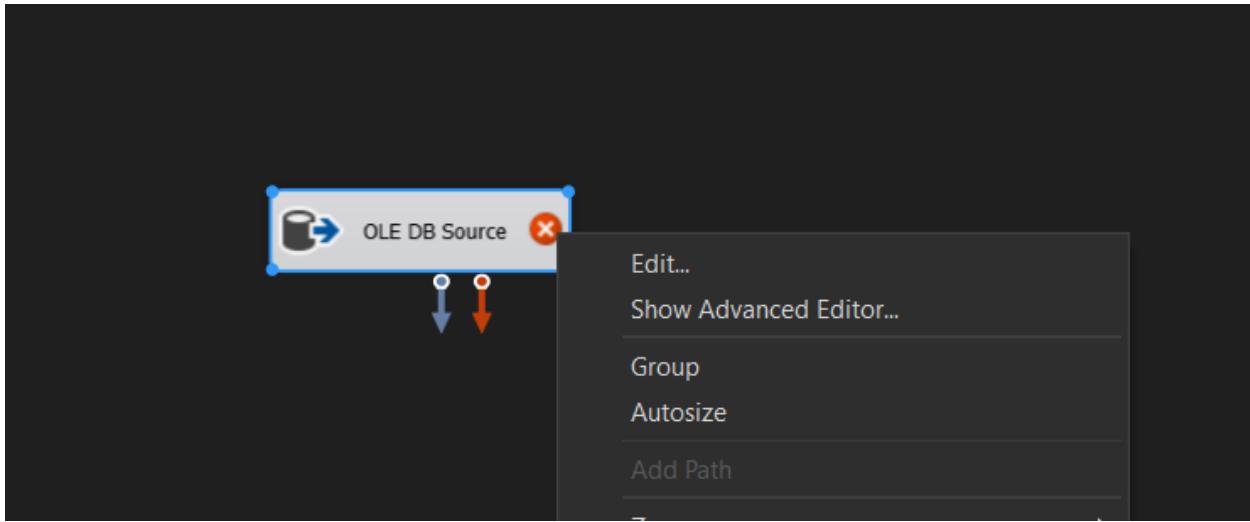


## 5. Quá trình tạo bảng Fact

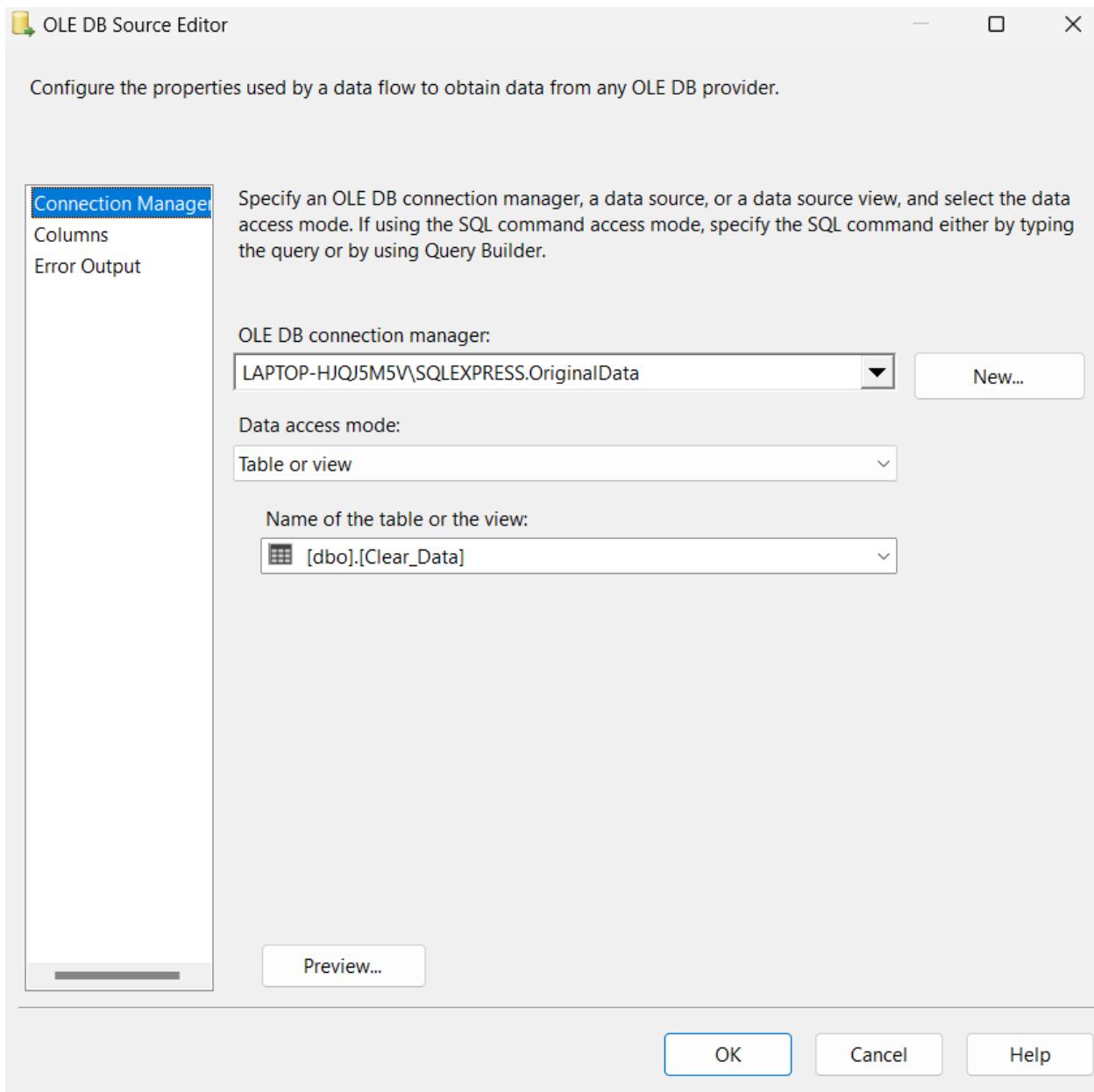
Bước 1: Ta thực hiện việc tạo mới “Data Flow Task” Tên là “Quá trình tạo bảng Fact”.



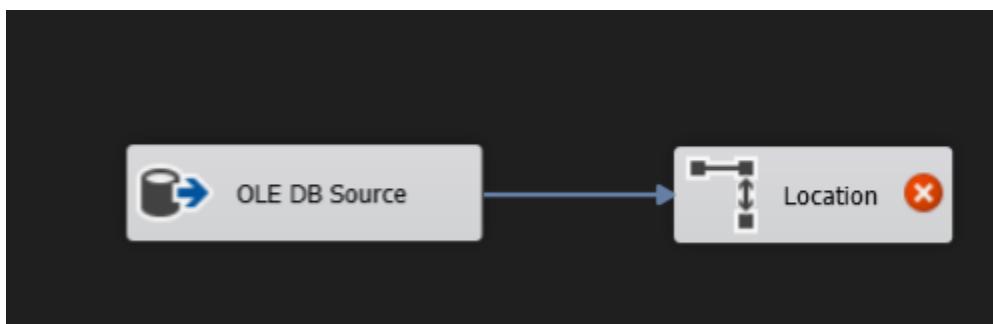
Bước 2 : Ta thực hiện chỉnh sửa “Data Flow Task” bằng cách ấn vào Edit . Ta thực hiện thêm “OLE DB Source” và cài đặt.



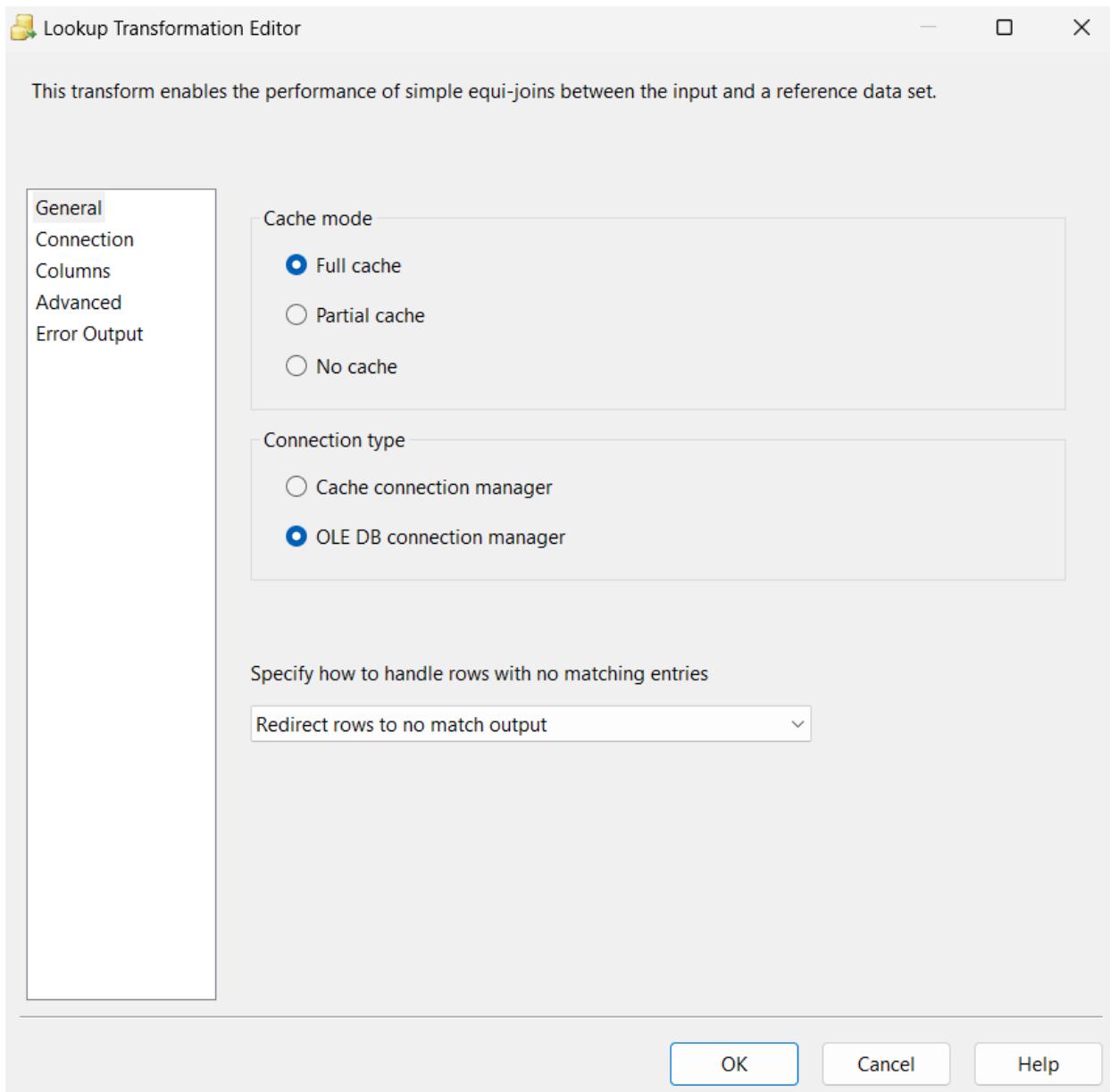
Ta thực hiện việc kết nối đến database OriginalData và chọn bảng dữ liệu “Clear\_data” để làm nguồn dữ liệu.



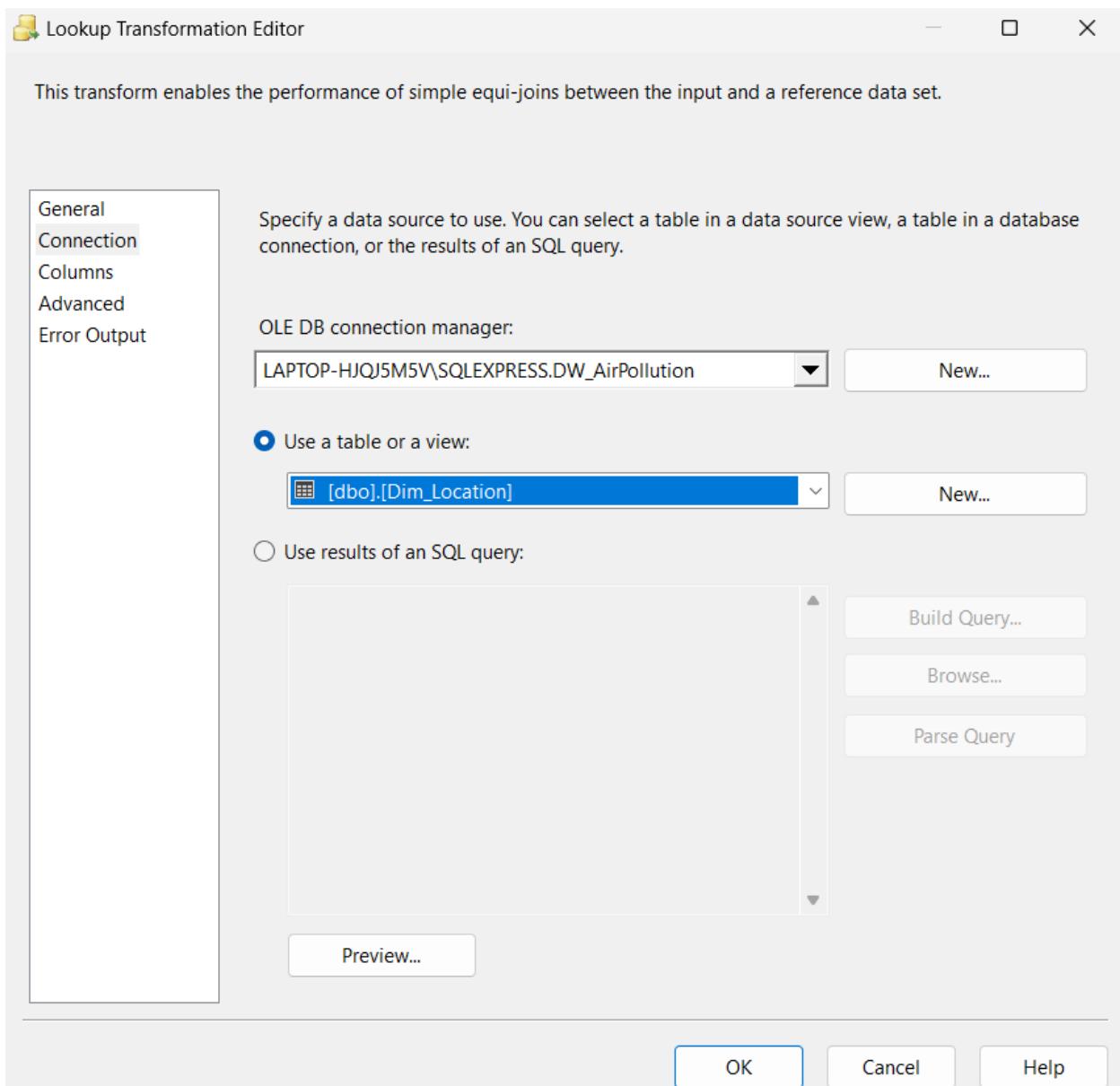
Bước 3 : Ta tạo mới “Look up” với tên là “Location” cho các thuộc tính từ bảng “Dim\_Location” ở database DW\_AirPollution .



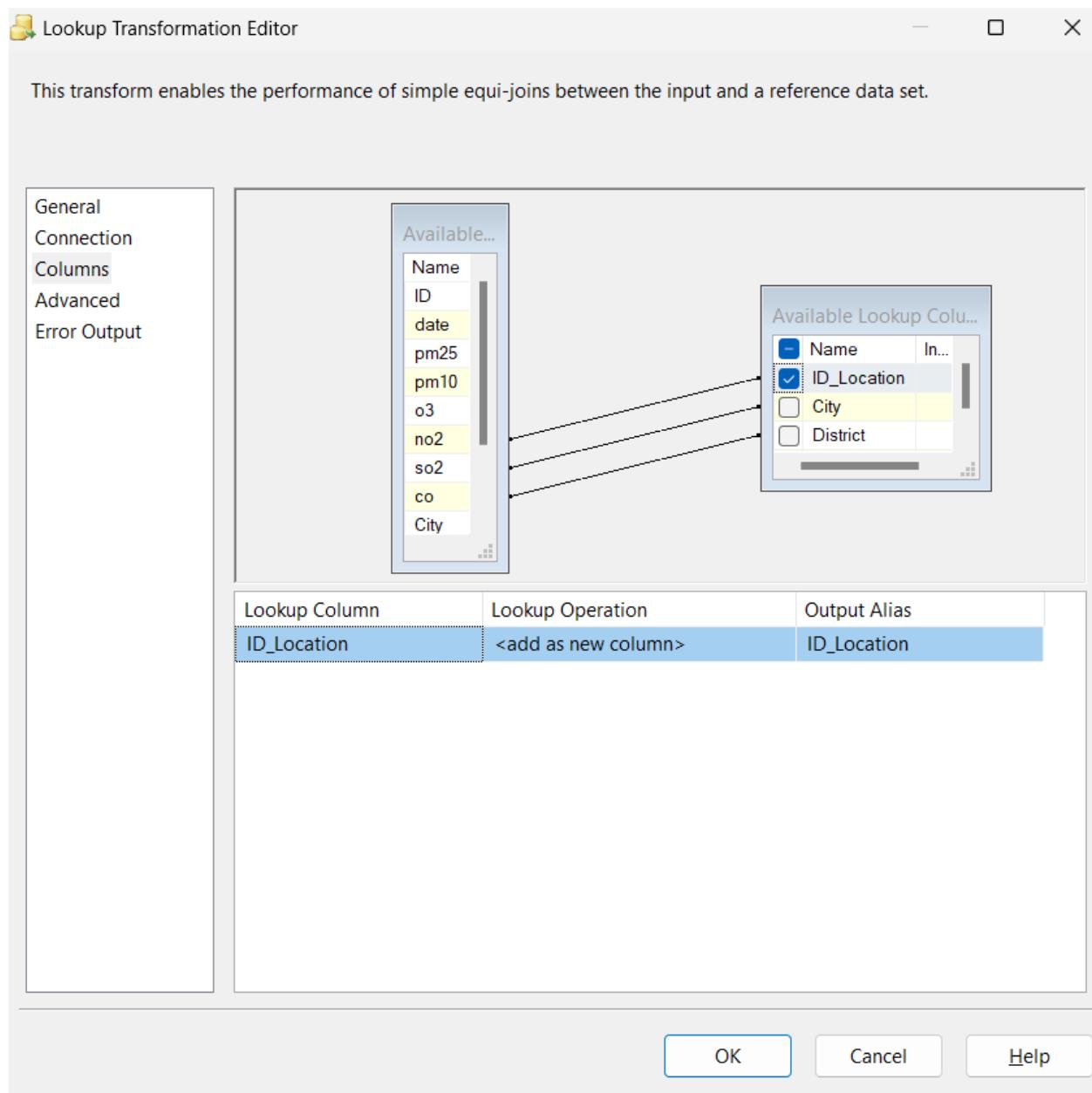
Ta thực hiện cấu hình "Look up" bằng cách chọn "Edit". Ta chọn "Redirect rows to no match output" .



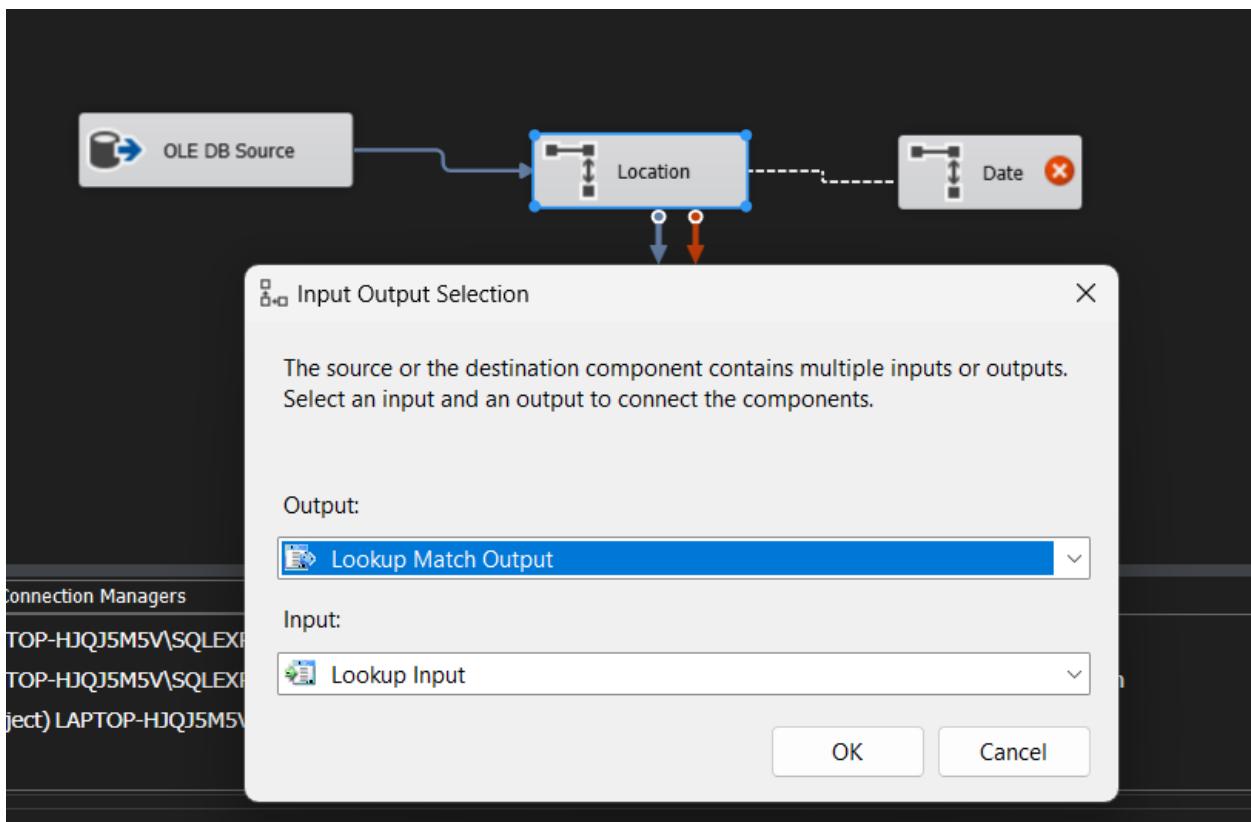
Tiếp theo ta chọn bảng "Dim\_Location"



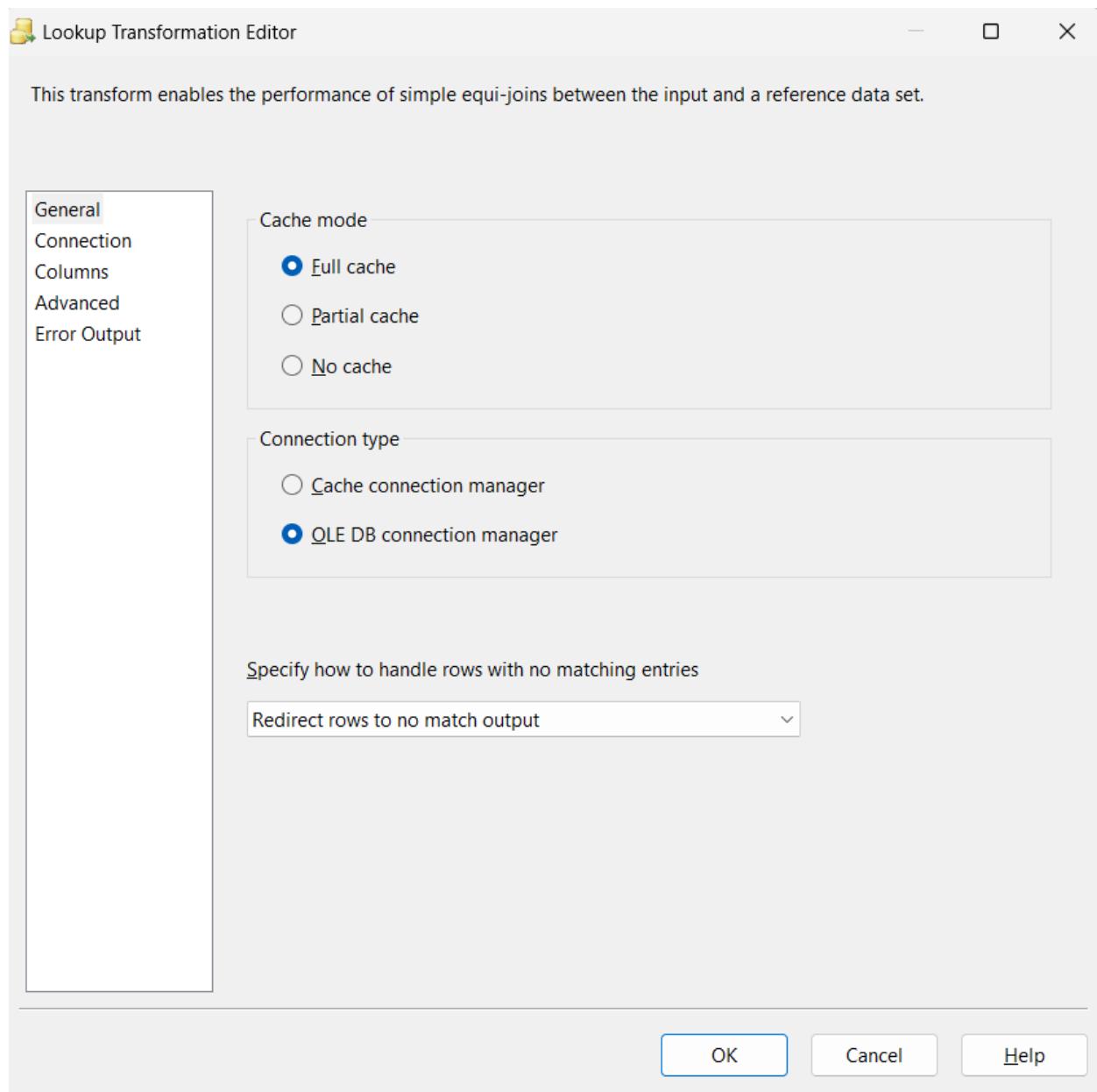
Ta thực hiện kéo thả với các hàng tương ứng và chọn ok.



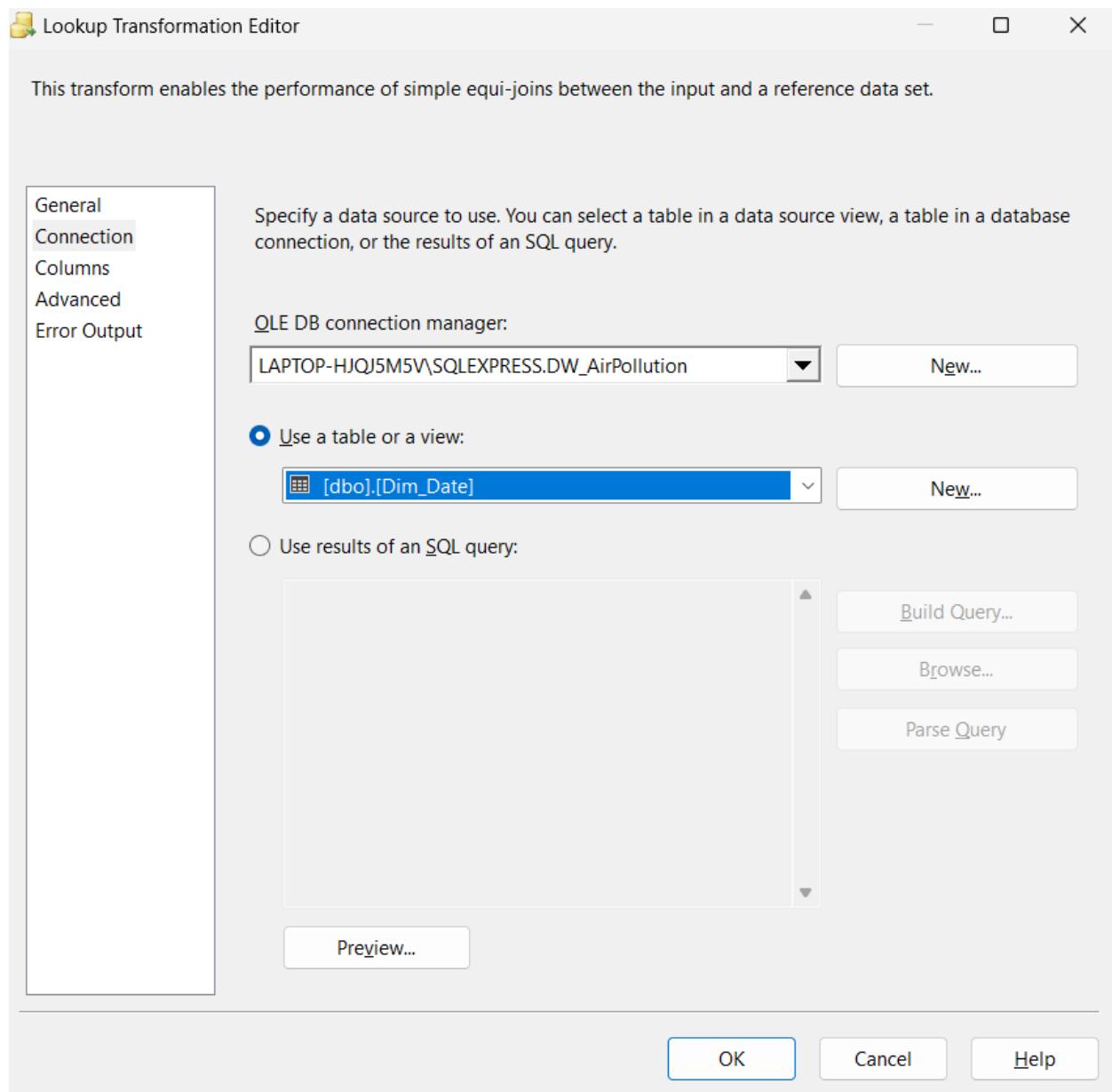
Bước 4 : Tương tự ta tạo mới “Look up” với tên là “Date” cho các thuộc tính từ bảng “Dim\_Date” ở database DW\_AirPollution . Ta chọn “Look up Match Output”.



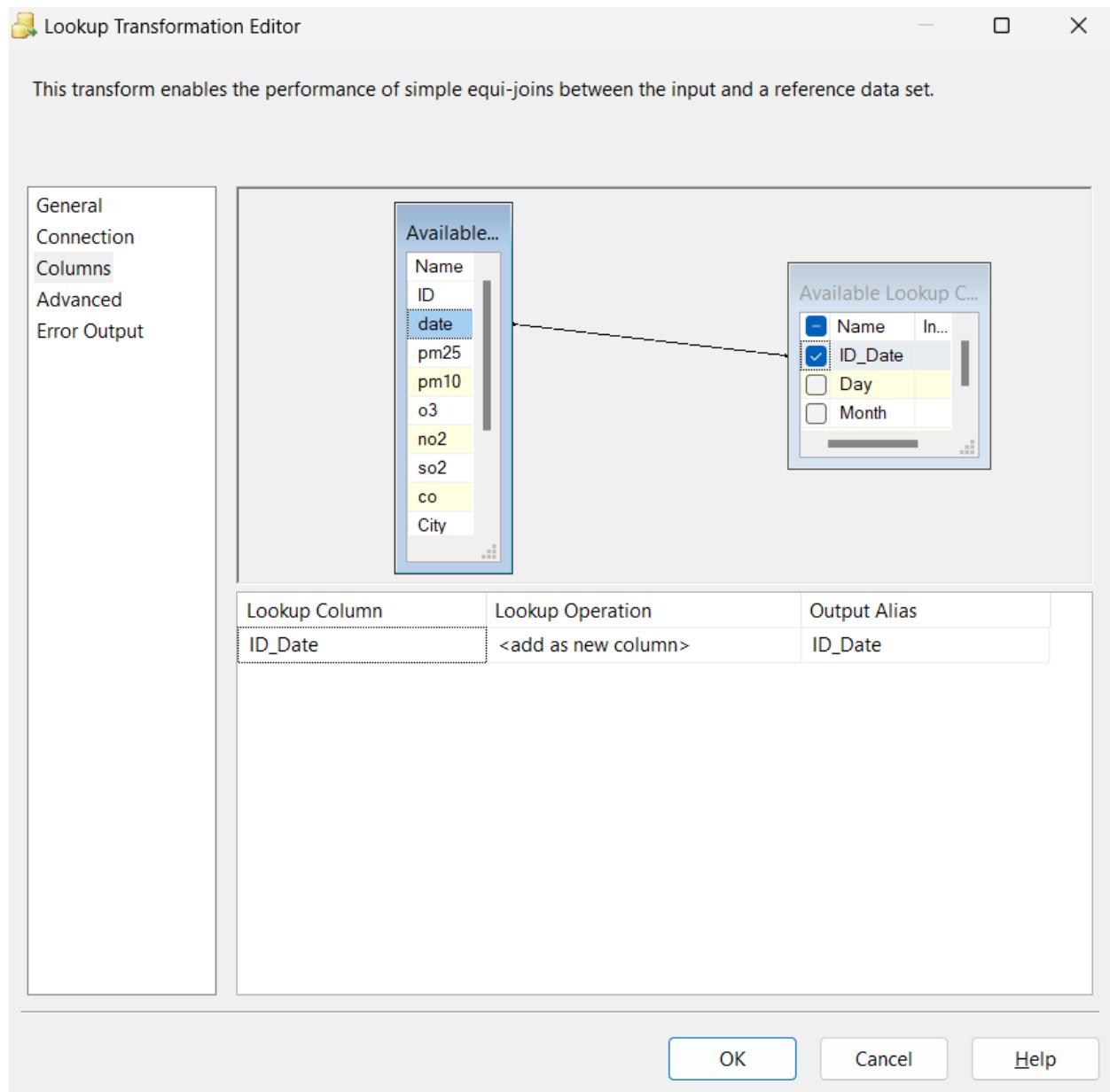
Ta thực hiện cấu hình “Look up” bằng cách chọn “Edit”. Ta chọn “Redirect rows to no match output” .



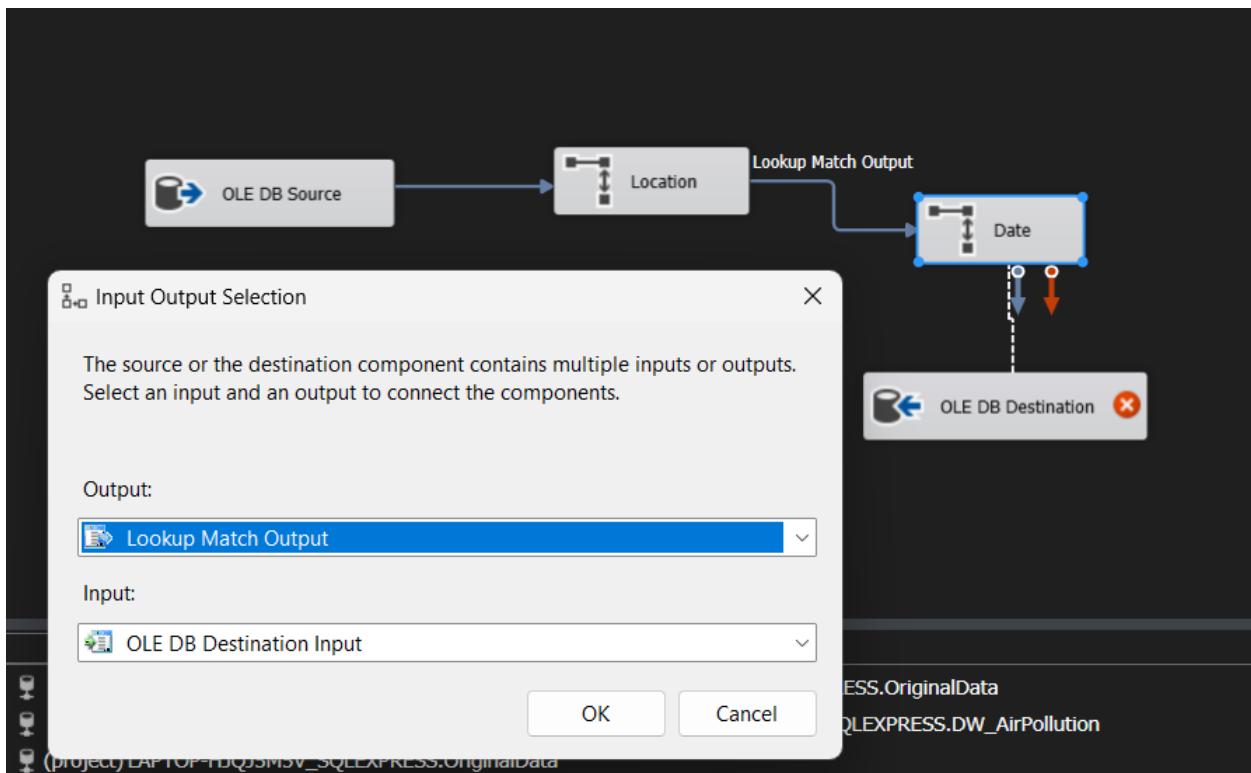
Tiếp theo ta chọn bảng "Dim\_Date"



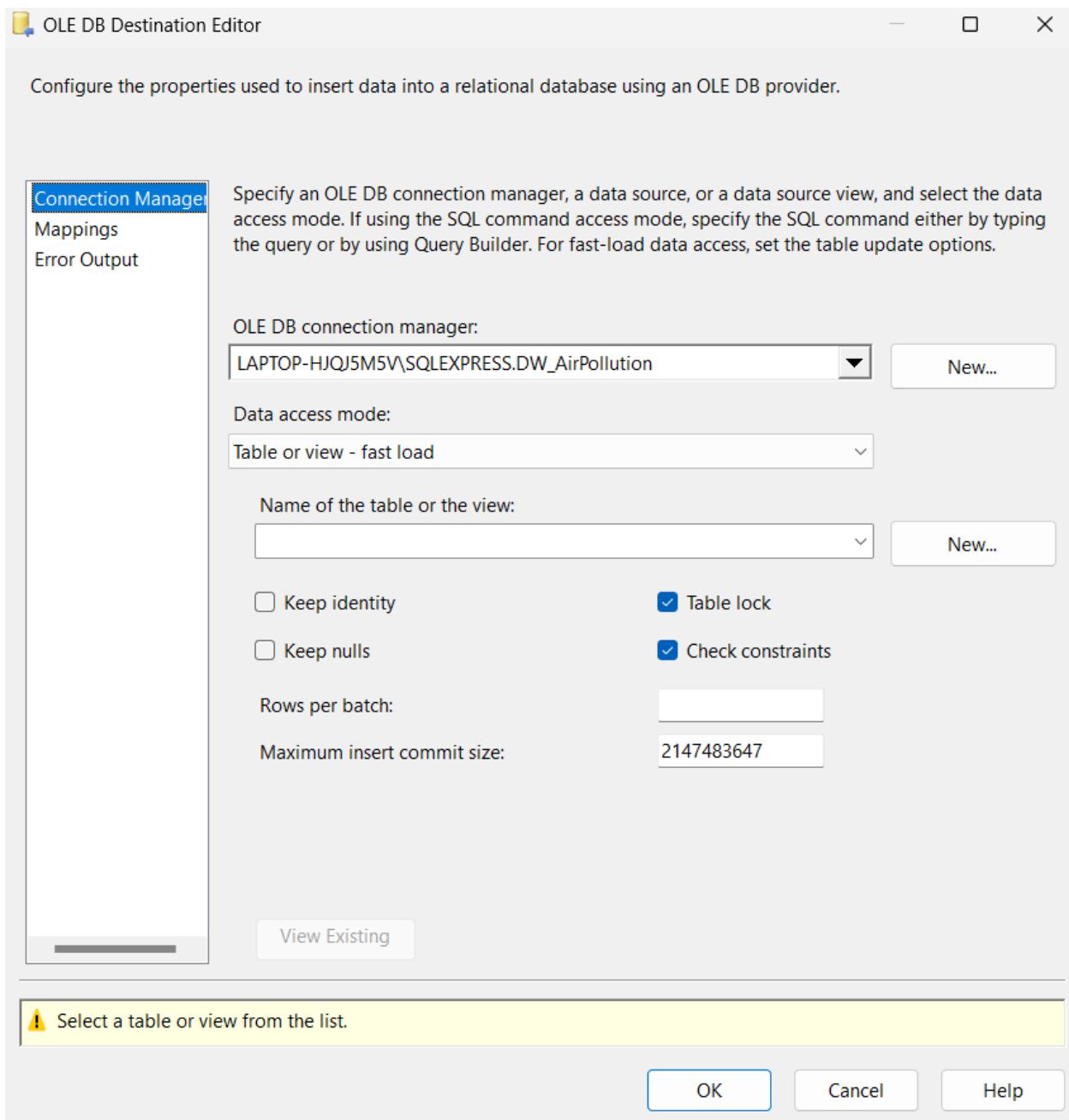
Ta nối "date" với "ID\_date" và chọn ok.



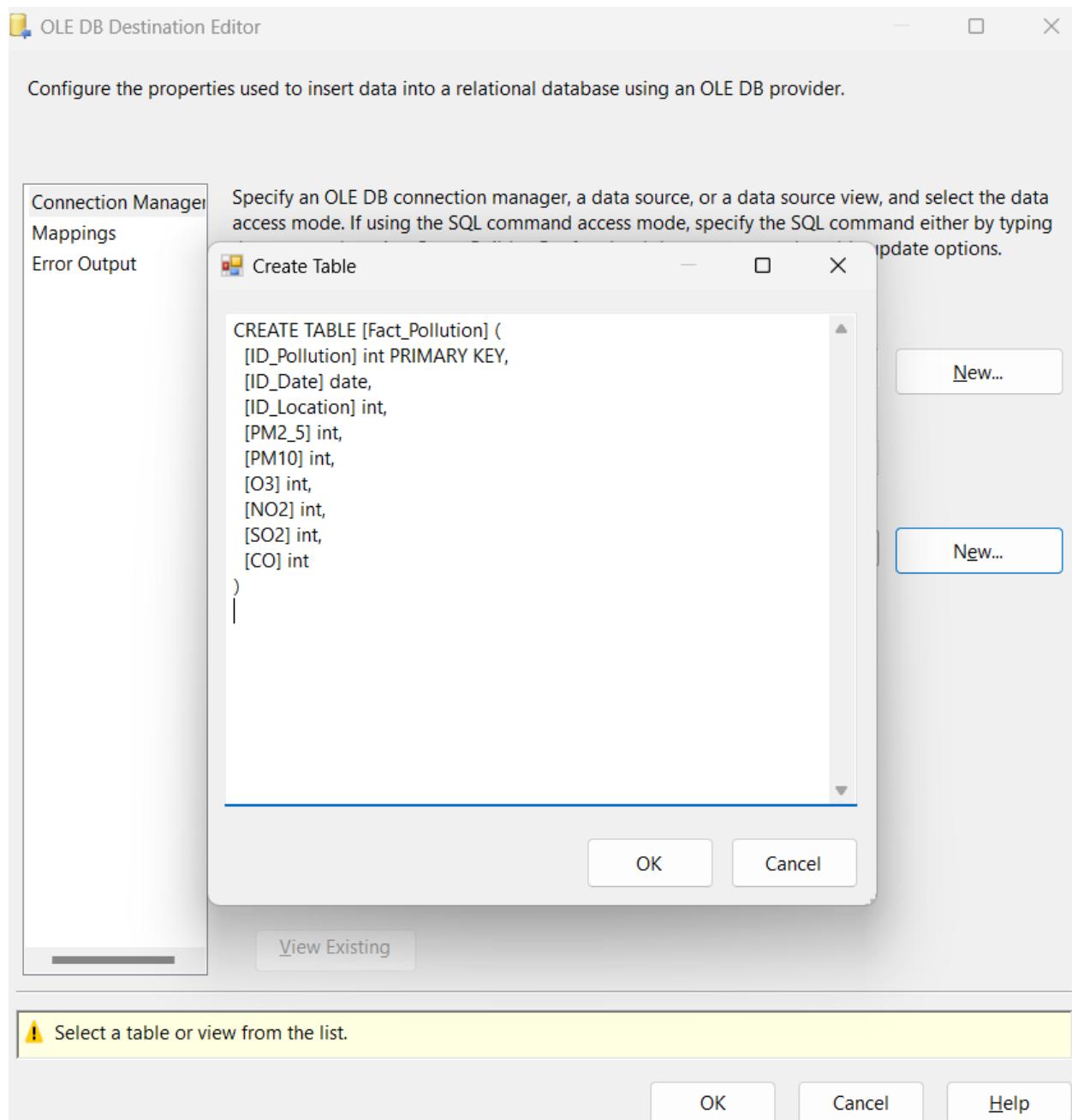
Bước 5 : Ta thực hiện việc tạo “OLE DB Destination” và “Lookup Match OutPut” .



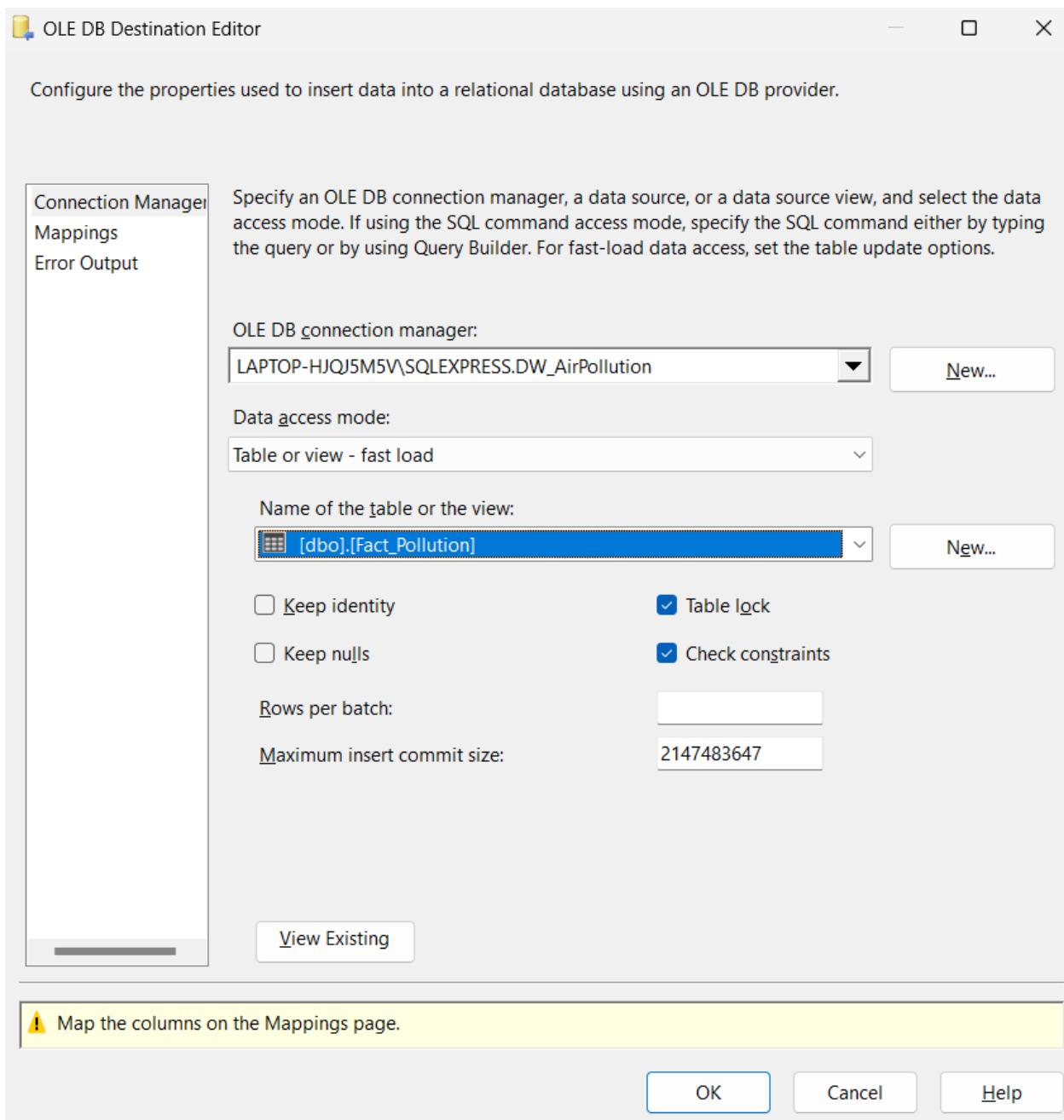
thực hiện việc kết nối đến database “DW\_AirPollution” .



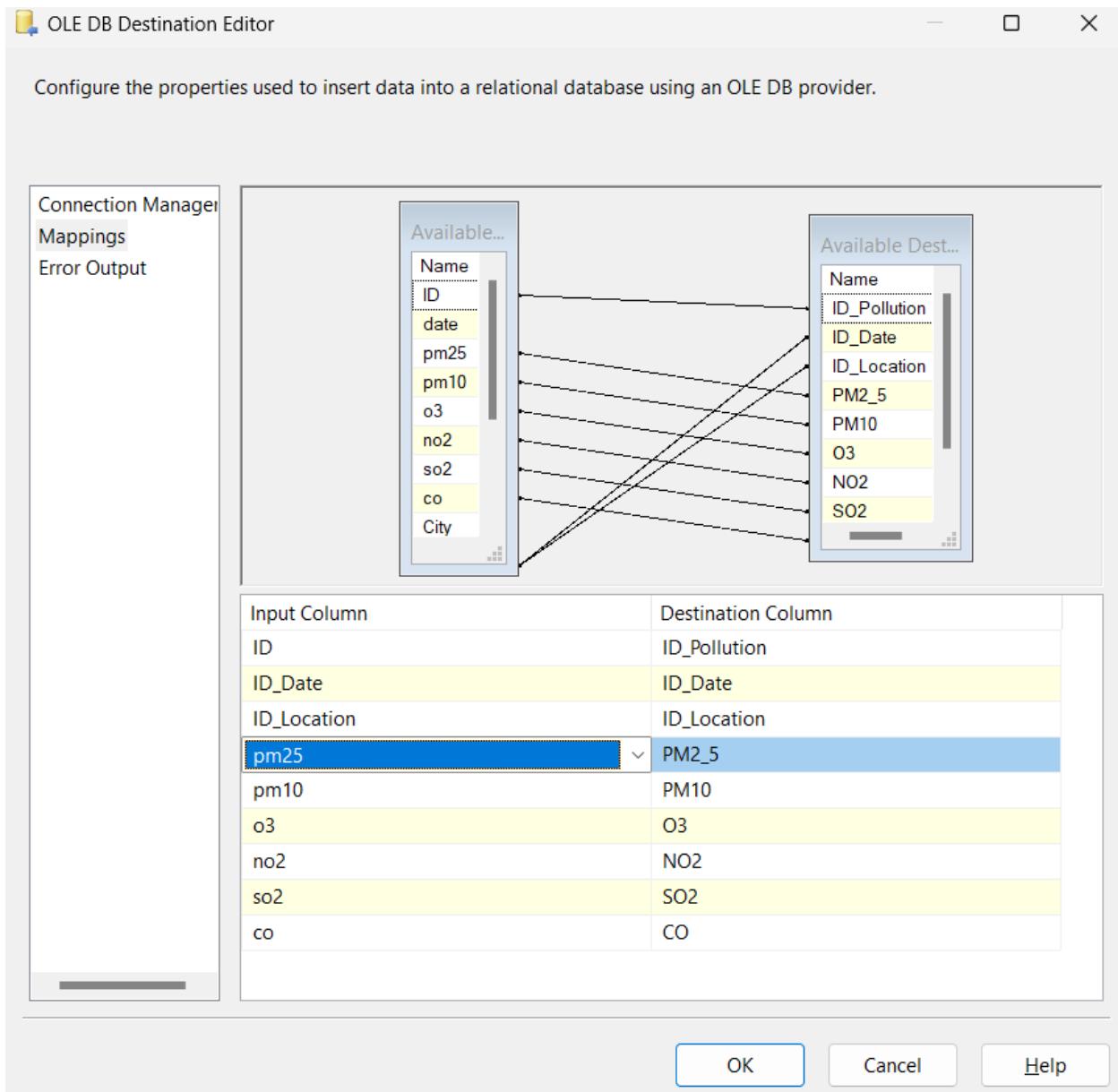
Tiếp đến ta thực hiện việc tạo mới table “Fact\_Pollution” và chạy lệnh SQL để tạo.



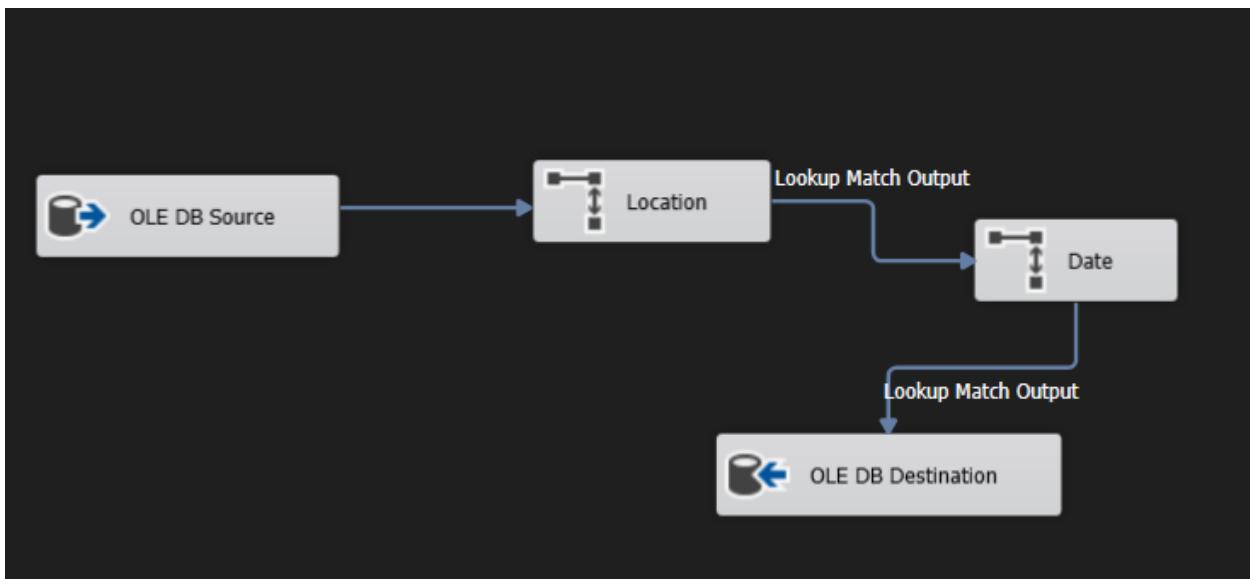
Sau khi tạo hoàn tất , ta chọn "Fact\_Pollution"



Cuối cùng ta thực hiện kiểm tra mapping và chọn "ok" để hoàn tất .

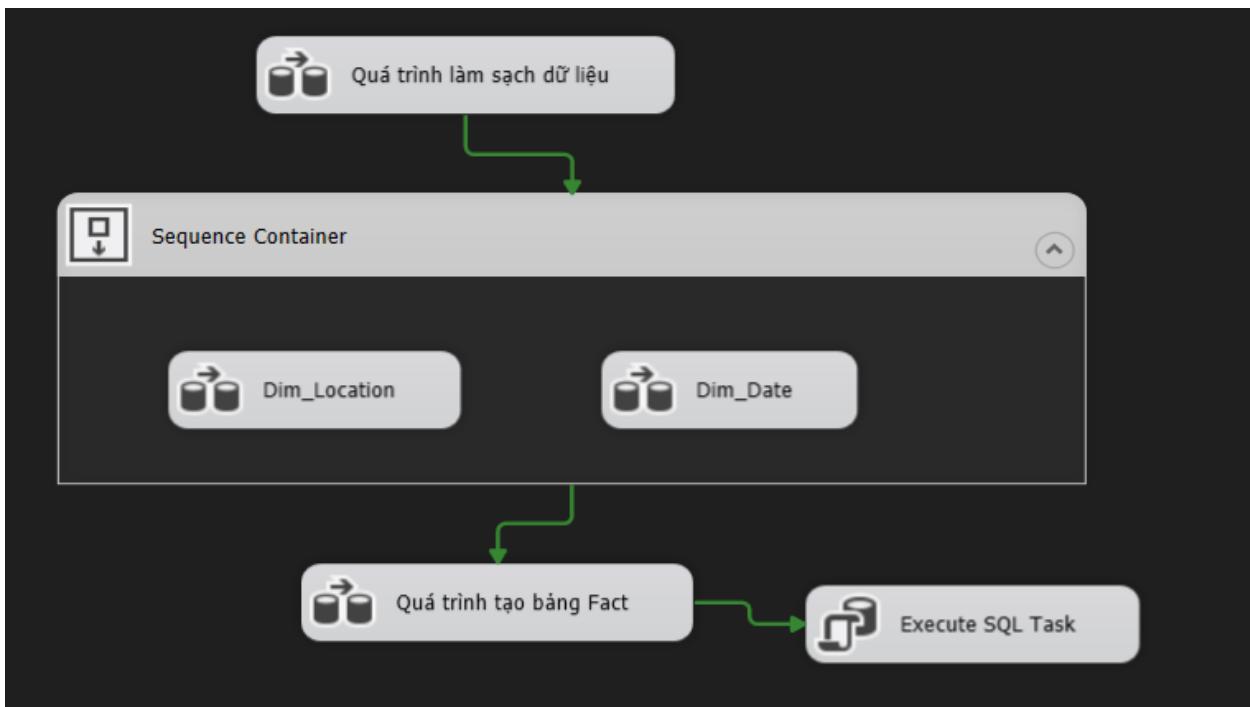


Data Flow của quá trình tạo bảng Fact\_Pollution sau khi hoàn thành.

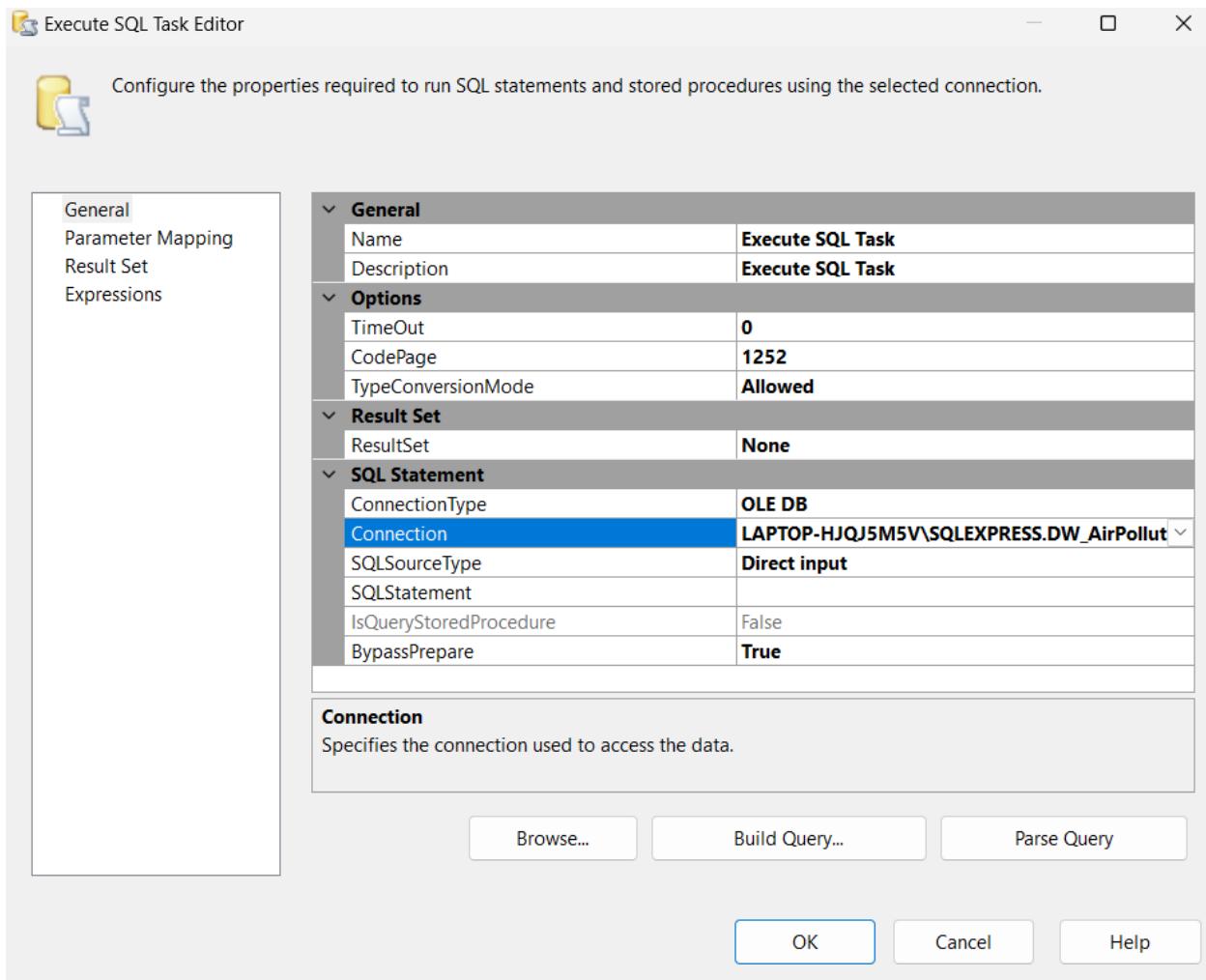


## 6. Tạo các ràng buộc khóa ngoại

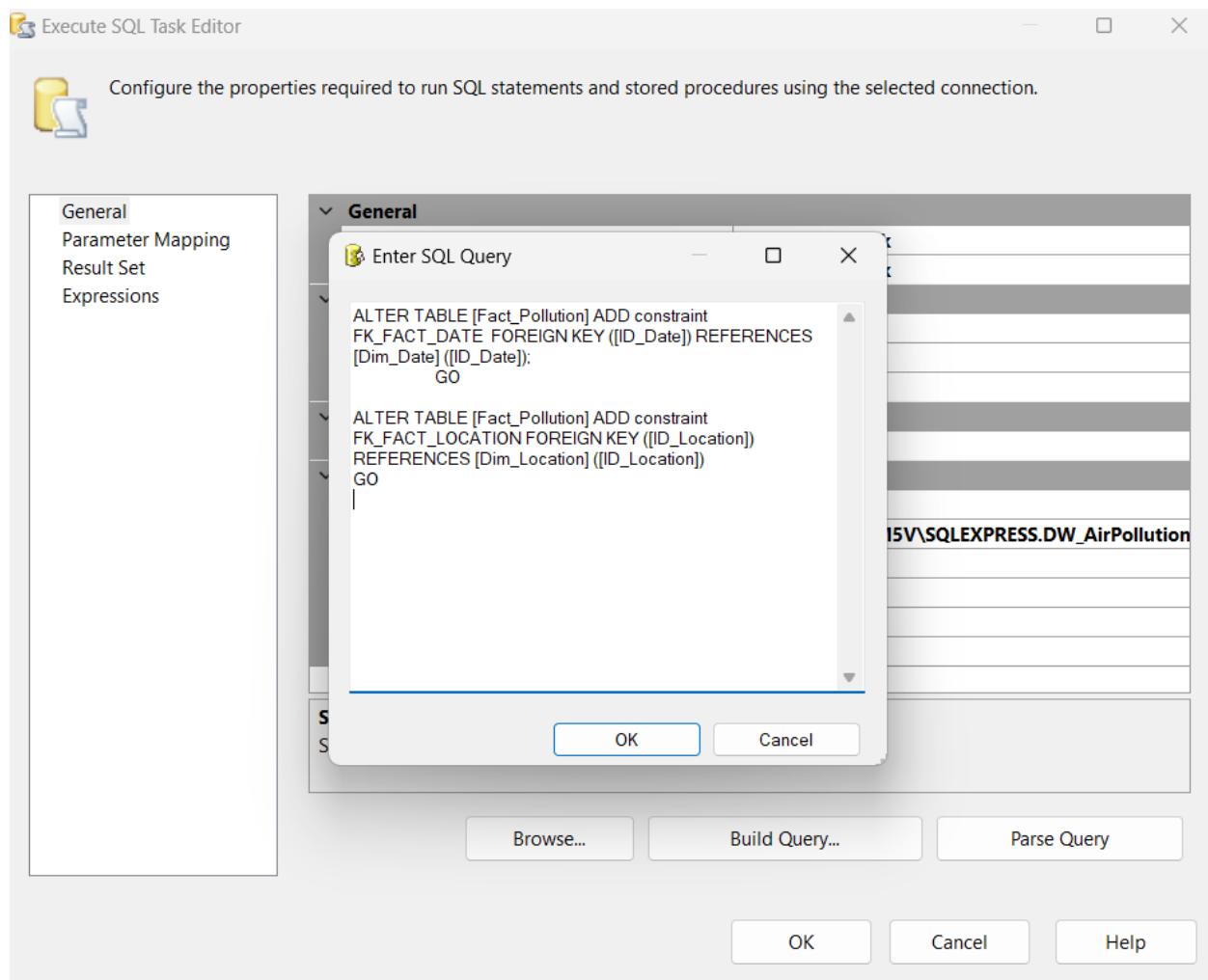
Ta tạo mới “Excute SQL Task” .



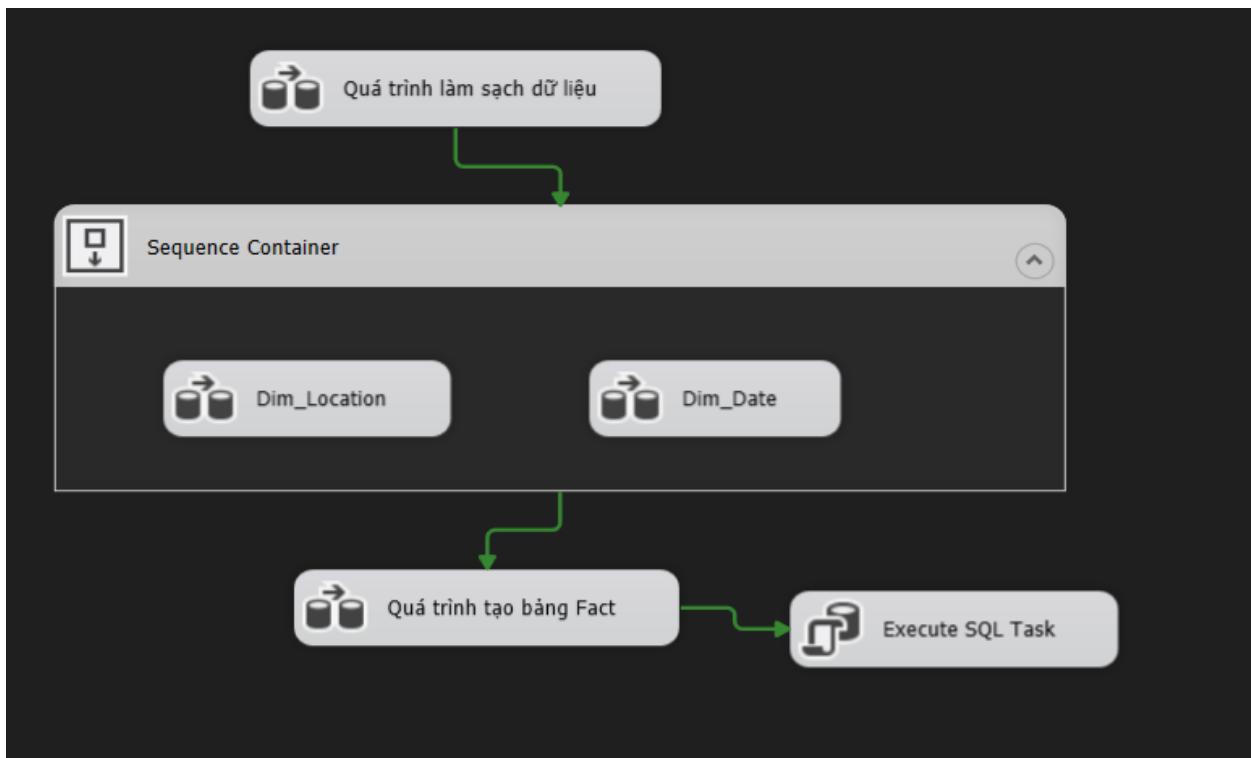
Ta thực hiện việc chỉnh sửa .Ở phần “connection” ta chọn kết nối đến “DW\_AirPollution”.



Ở phần "SQLStatement" ta thêm các lệnh ràng buộc khóa ngoại cho database và chọn Ok để hoàn tất .

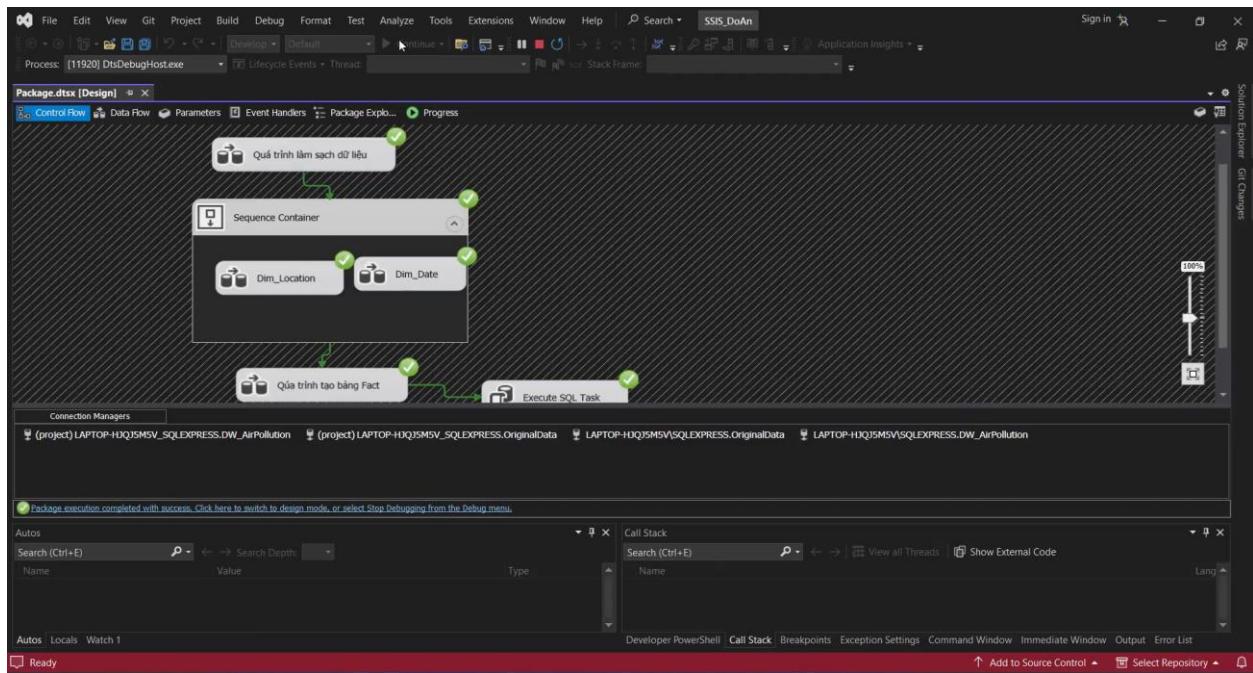


Data Flow của quá trình tạo ràng buộc khóa ngoại sau khi hoàn thành.



## 7. Thực thi

Ta thực hiện chạy ta được kết quả như sau , thì đã là triển khai thành công :



## 8. Dữ liệu sau khi hoàn thành

Dữ liệu ở bảng "Dim\_Date":

The screenshot shows the Microsoft SQL Server Management Studio interface. On the left, the Object Explorer pane displays the database structure, including the DW\_AirPollution database and its tables. In the center, a query window titled 'SQLQuery1.sql' contains the SQL command: 'select \* from [dbo].[Dim\_Date]'. The results pane on the right shows a table with 21 rows of data, each representing a date from January 1, 2014, to January 20, 2014. The columns are ID\_Date, Day, Month, and Year. A message at the bottom of the results pane indicates that the query was executed successfully.

ID_Date	Day	Month	Year
2013-12-31	31	12	2013
2014-01-01	1	1	2014
2014-01-02	2	1	2014
2014-01-03	3	1	2014
2014-01-04	4	1	2014
2014-01-05	5	1	2014
2014-01-06	6	1	2014
2014-01-07	7	1	2014
2014-01-08	8	1	2014
2014-01-09	9	1	2014
2014-01-10	10	1	2014
2014-01-11	11	1	2014
2014-01-12	12	1	2014
2014-01-13	13	1	2014
2014-01-14	14	1	2014
2014-01-15	15	1	2014
2014-01-16	16	1	2014
2014-01-17	17	1	2014
2014-01-18	18	1	2014
2014-01-19	19	1	2014
2014-01-20	20	1	2014

Dữ liệu ở bảng "Dim\_Location":

SQLQuery1.sql - LAPTOP-HJQJ5M5V\SQLEXPRESS.DW\_AirPollution (LAPTOP-HJQJ5M5V\ACER (74)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

DW\_AirPollution Execute

Object Explorer

SQLQuery1.sql - L...JQJ5M5V\ACER (74)\*

```
select * from [dbo].[Dim_Location]
```

ID_Location	City	District	Country
1	Bangsan-Myeon	Gangwon	South Korea
2	Cheongnim-Dong	Gyeongbuk	South Korea
3	Geumchon-Dong	Paju-Si	Gyeonggi
4	Gwanak-Gu	Seoul	South Korea
5	Hwasun-Eup	Jeonnam	South Korea
6	Hyeksin-Dong	Jeonbuk	South Korea
7	Jeongnim-Dong	Seo-Gu	Daejeon
8	Jungang-Way	Chuncheon-Si	Gangwon
9	Juwol-Dong	Nam-Gu	Gwangju
10	Nowon-Gu	Seoul	South Korea
11	Onui-Dong	Gangwon	South Korea
12	Pado-Ri	Chungnam	South Korea
13	Paju	Gyeonggi	South Korea
14	Saemangeum	Jeonbuk	South Korea
15	Saesol-Dong	Gyeonggi	South Korea
16	Sangjusi	Gyeongbuk	South Korea
17	Seolseong-Myeon	Gyeonggi	South Korea
18	Seonggeo-eup	Chungnam	South Korea
19	Soi-Myeon	Chungbuk	South Korea
20	Taeha-Ri	Gyeongbuk	South Korea

Query executed successfully.

Dữ liệu ở bảng "Fact\_Pollution":

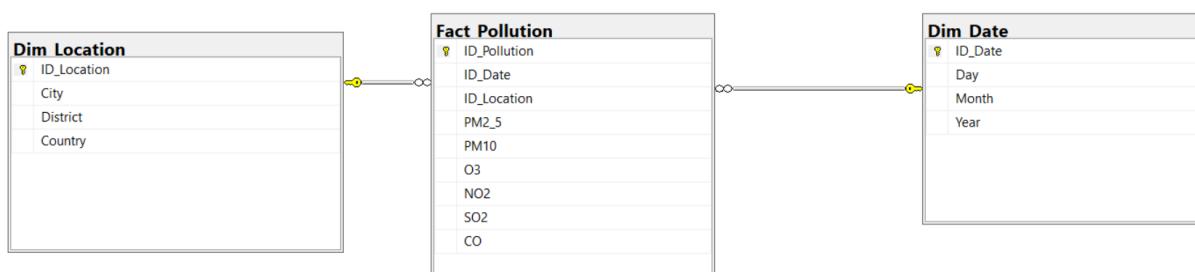
```

SQLQuery1.sql - LAPTOP-HJQJ5M5V\SQLEXPRESS.DW_AirPollution (LAPTOP-HJQJ5M5V\ACER (74)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
DW_AirPollution Execute description, ...
Object Explorer
Connect Databases System Databases Database Snapshots buffet CakeShop cakeshop2 CUAHANGXANGDAU DeThi DW_AirPollution Database Diagrams Tables System Tables FileTables External Tables Graph Tables dbo.Dim_Date dbo.Dim_Location dbo.Fact_Pollution Views External Resources Synonyms Programmability Query Store Service Broker Storage Security EMPLOYEE EMS NHANVIENLAIXE NORTHWND OriginalData QLBH ReportServer ReportServerTempDB
SQLQuery1.sql - L...JQJ5M5V\ACER (74)*
select * from [dbo].[Dim_Date]
select * from [dbo].[Dim_Location]
select * from [dbo].[Fact_Pollution]
Results Messages
ID_Pollution ID_Date ID_Location PM2_5 PM10 O3 NO2 SO2 CO
1 0 2022-02-01 1 112 31 35 2 1 4
2 1 2022-02-02 1 92 21 35 2 1 0
3 2 2022-02-03 1 60 20 35 1 1 4
4 3 2022-02-04 1 51 27 33 1 1 4
5 4 2022-02-05 1 57 24 27 2 1 5
6 5 2022-02-06 1 51 23 33 3 1 5
7 6 2022-02-07 1 61 21 34 3 1 5
8 7 2022-02-08 1 58 21 34 2 1 4
9 8 2022-02-09 1 56 22 36 2 1 4
10 9 2022-02-10 1 58 45 44 3 2 4
11 10 2022-02-11 1 107 51 39 3 1 5
12 11 2022-02-12 1 129 35 37 3 1 4
13 12 2022-02-13 1 85 28 38 4 1 5
14 13 2022-02-14 1 74 34 34 1 0 4
15 14 2022-02-15 1 88 24 33 1 2 4
16 15 2022-02-16 1 64 25 30 3 1 5
17 16 2022-02-17 1 66 21 34 3 1 3
18 17 2022-02-18 1 57 29 35 3 1 3
19 18 2022-02-19 1 69 30 35 1 0 3
20 19 2022-02-20 1 69 0 0 0 0 0
21 20 2022-01-03 1 59 29 25 2 3 5

```

Query executed successfully.

## 9. Diagram đồ án sau khi hoàn thành

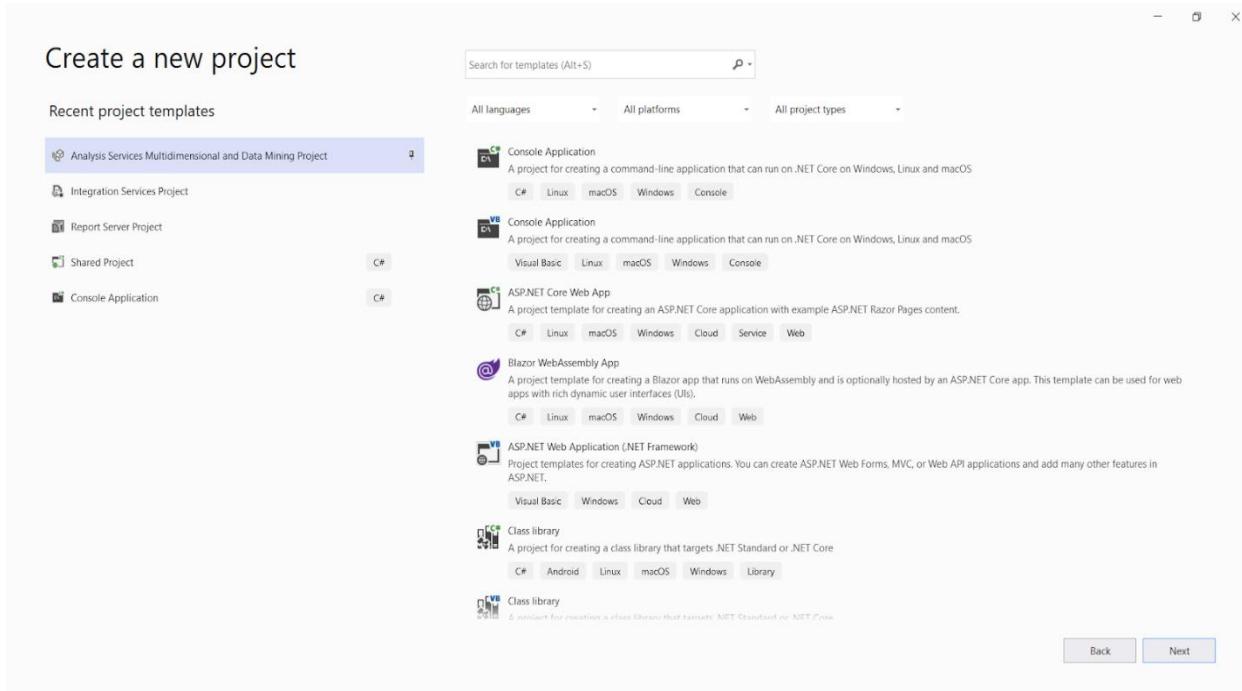


## III. Phân tích dữ liệu trong kho (SSAS)

### 1. Quá trình thực hiện SSAS

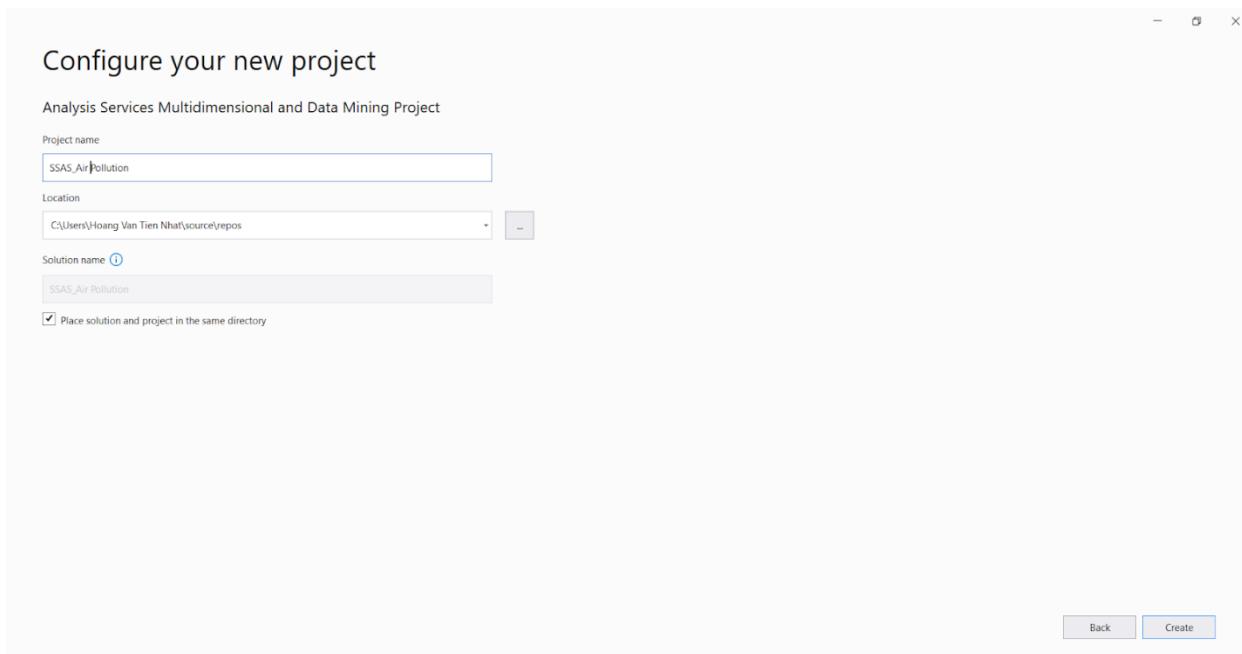
#### 1.1. Quá trình thực hiện trên công cụ Visual Data Studio

## Bước 1: Tạo project SSAS trong Visual Data Studio 2019

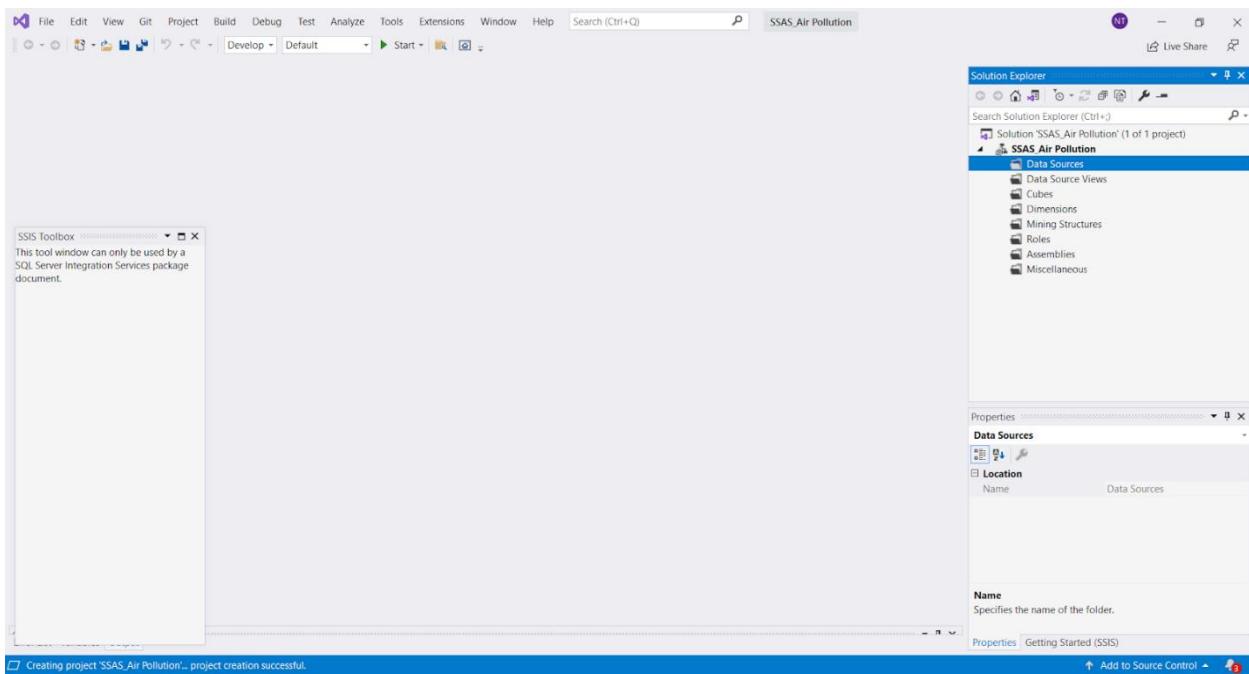


## Bước 2: Chọn **Analysis Services Multidimensional and Data Mining**

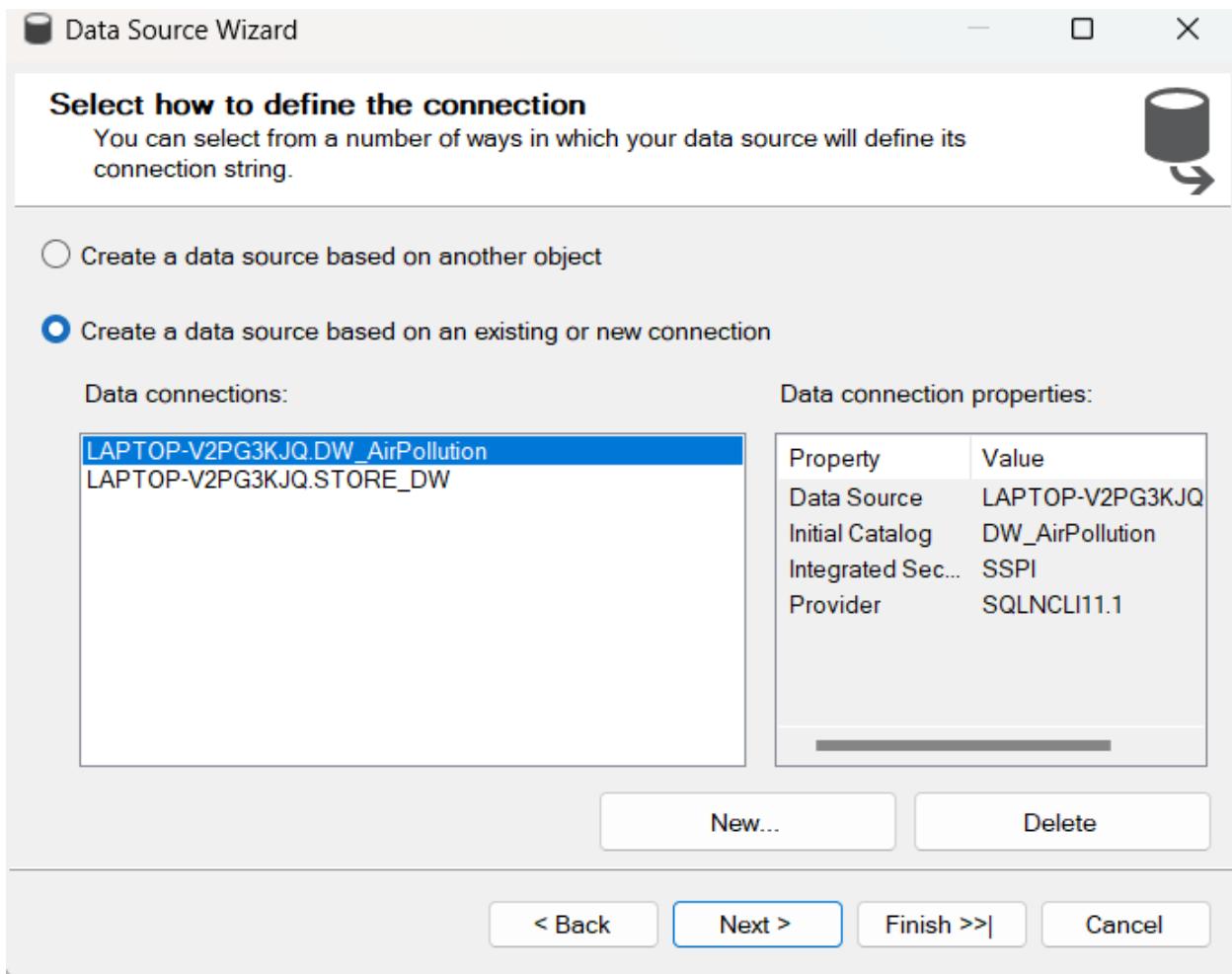
**Project** và đặt tên .



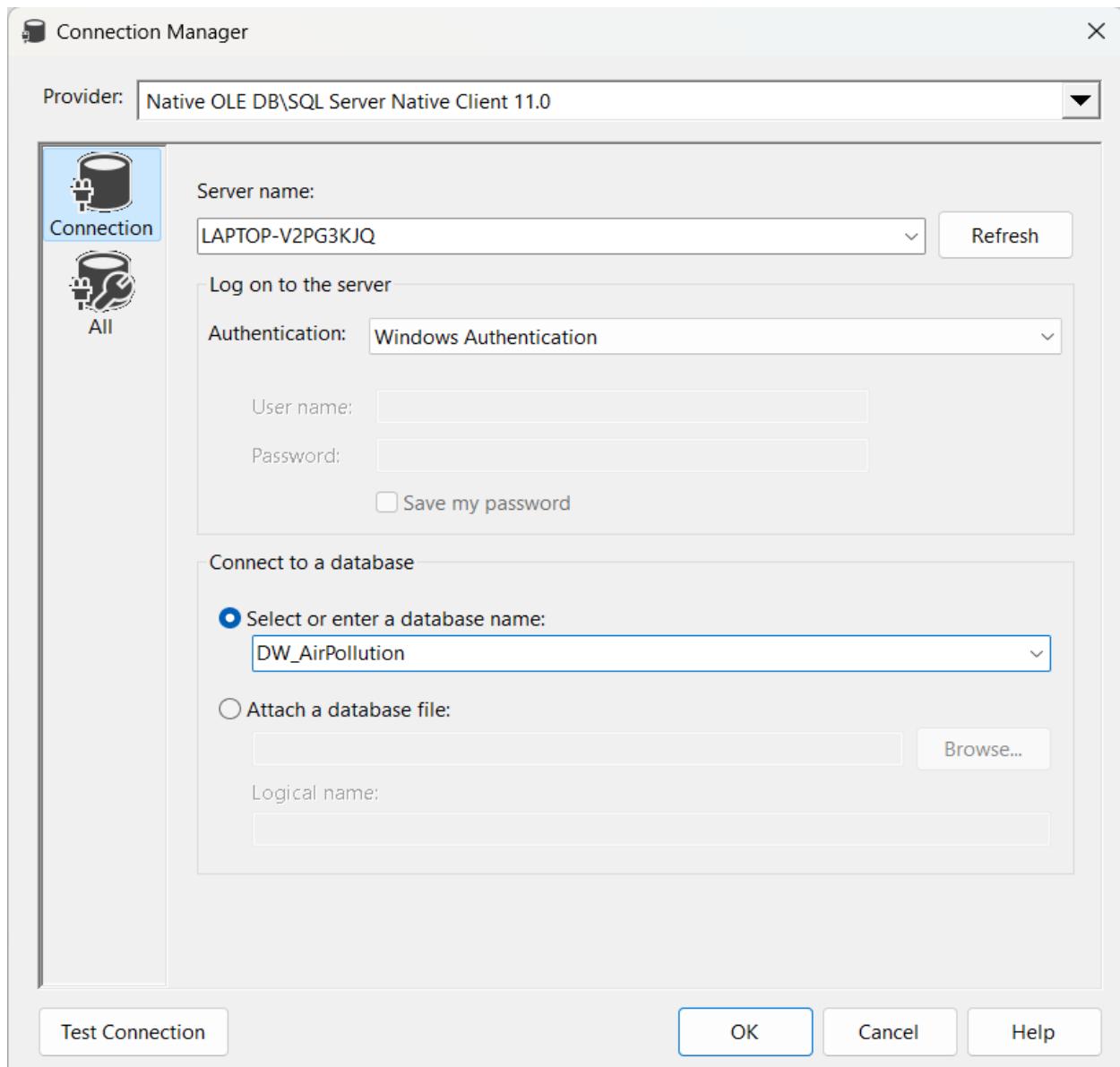
## Bước 3: Chọn **Data Source** để chọn Database đã tạo ở quá trình SSIS.



**Bước 4:** Click chọn **Create a data source based on an existing or new connection.**

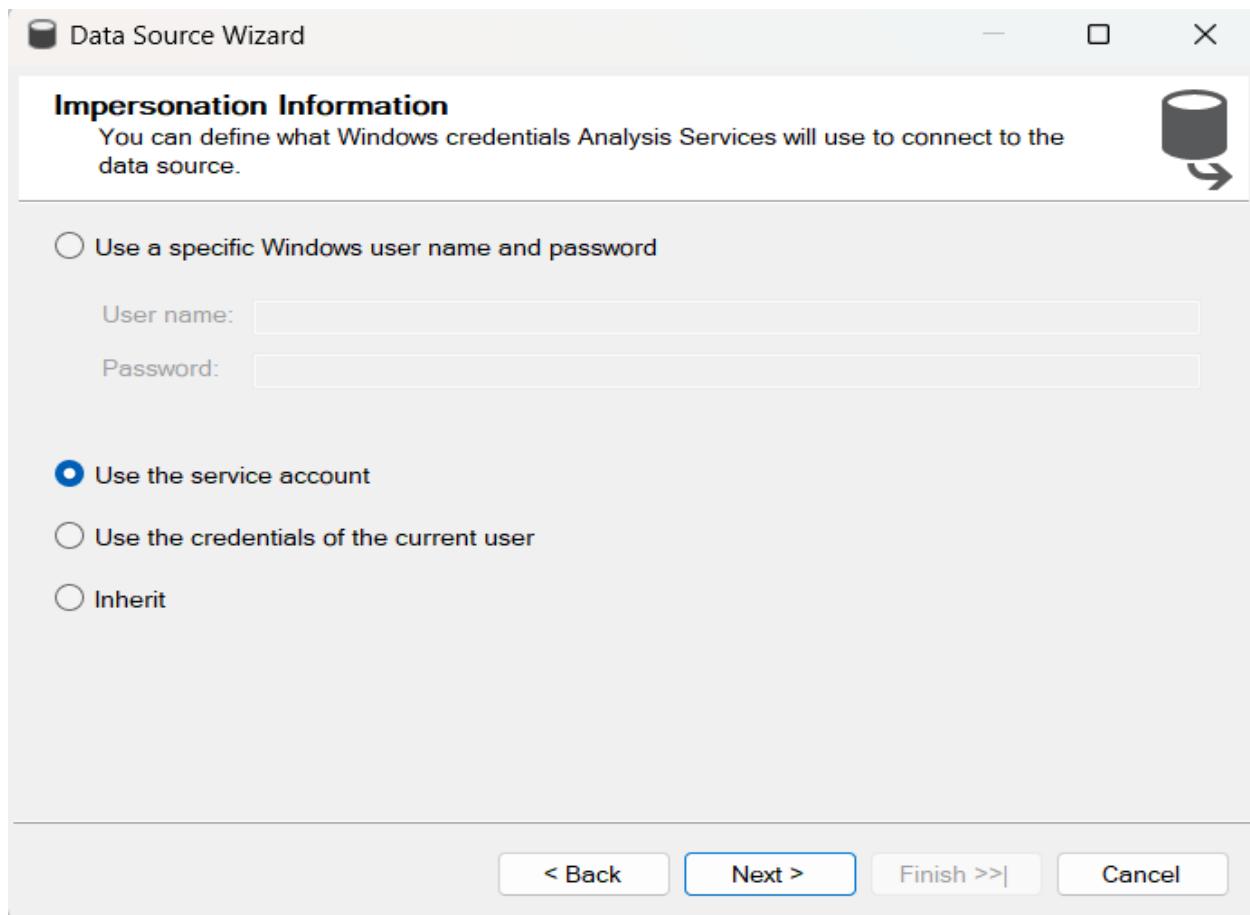


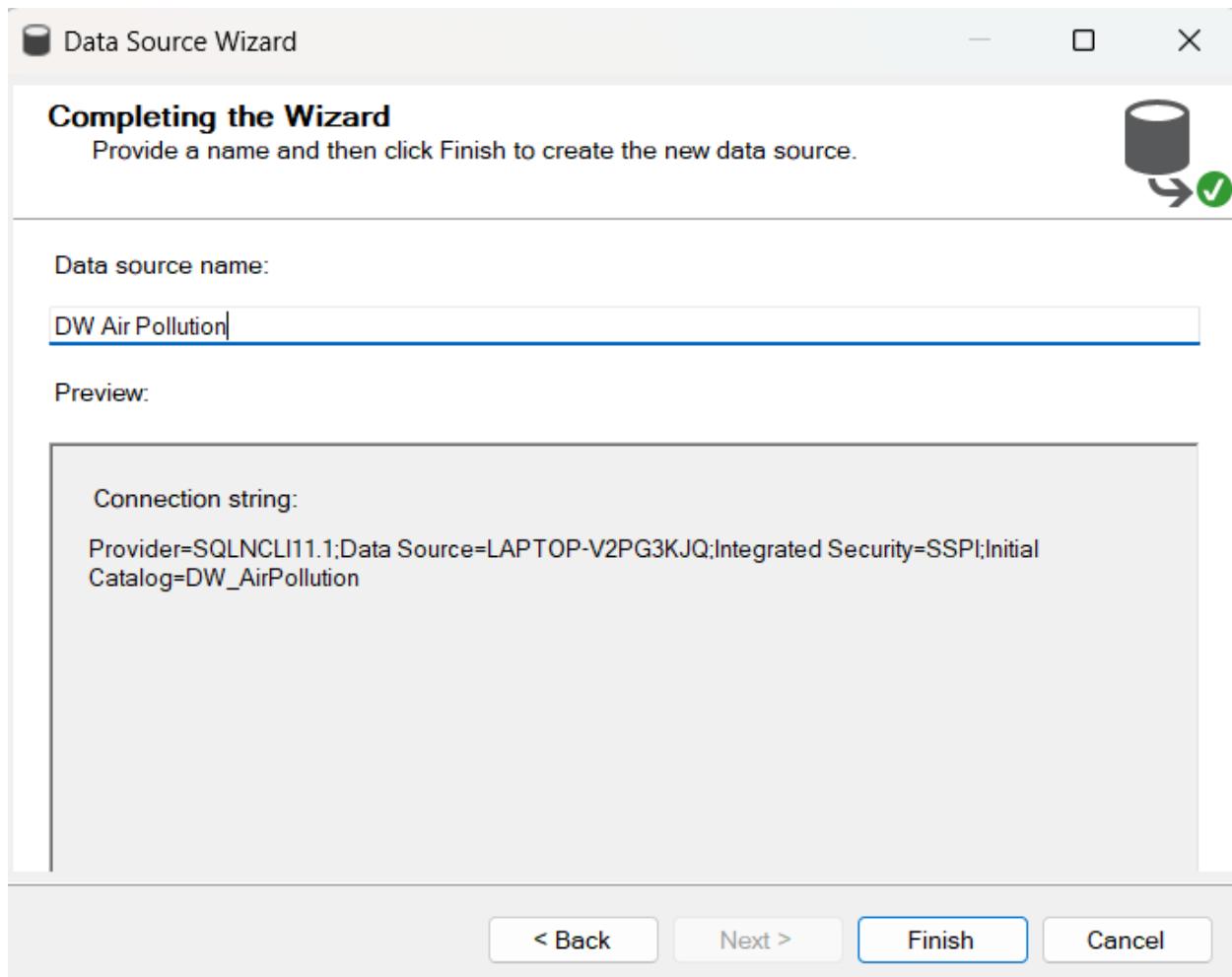
**Bước 5:** Click **New** để tạo **Data Connections** mới. Chọn **server name** là tên database trong MSSQ và chọn tên database cần thực hiện quá trình SSAS.



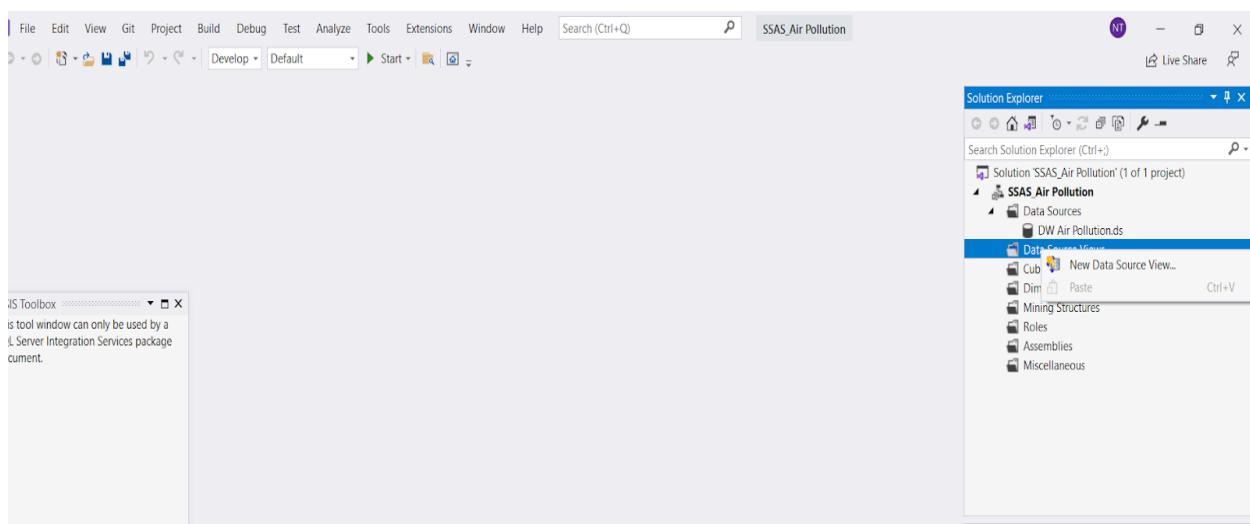
**Bước 6:** Click **OK** và chọn **Next**.

**Bước 7:** Chọn **Use the service account**. Click **Next** và **Finish**.

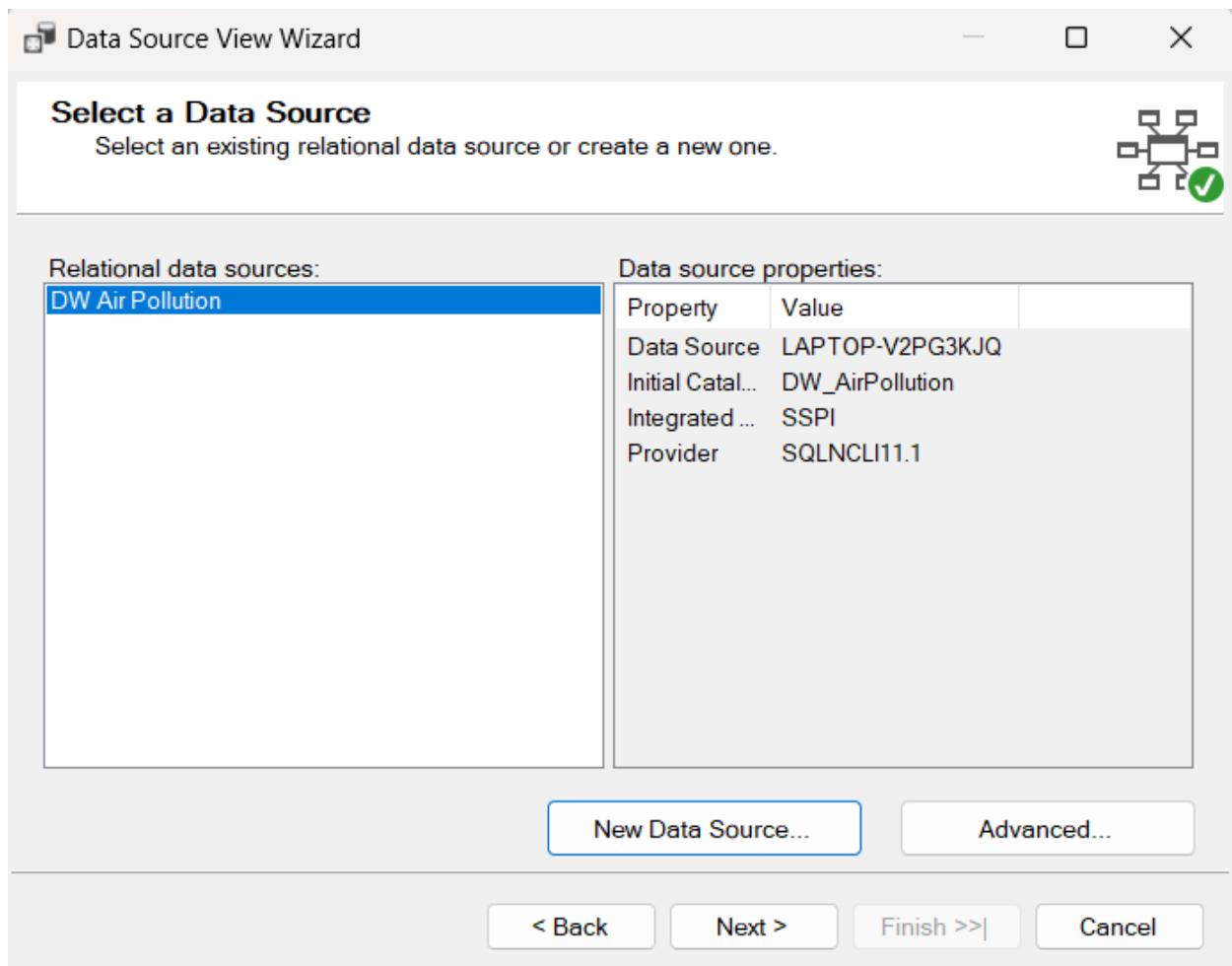




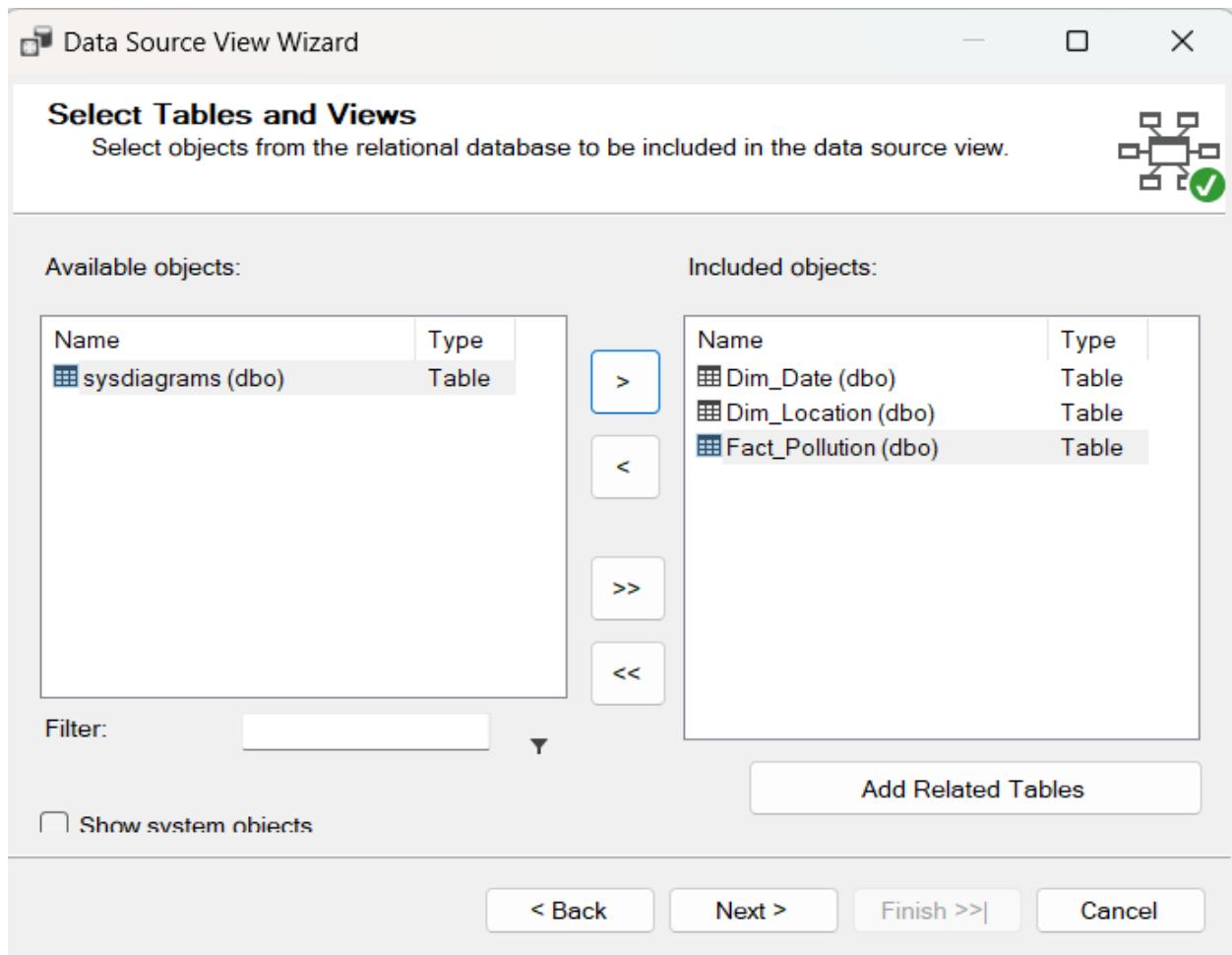
### Bước 8: Tại phần Data Source Views chọn New Data Source View



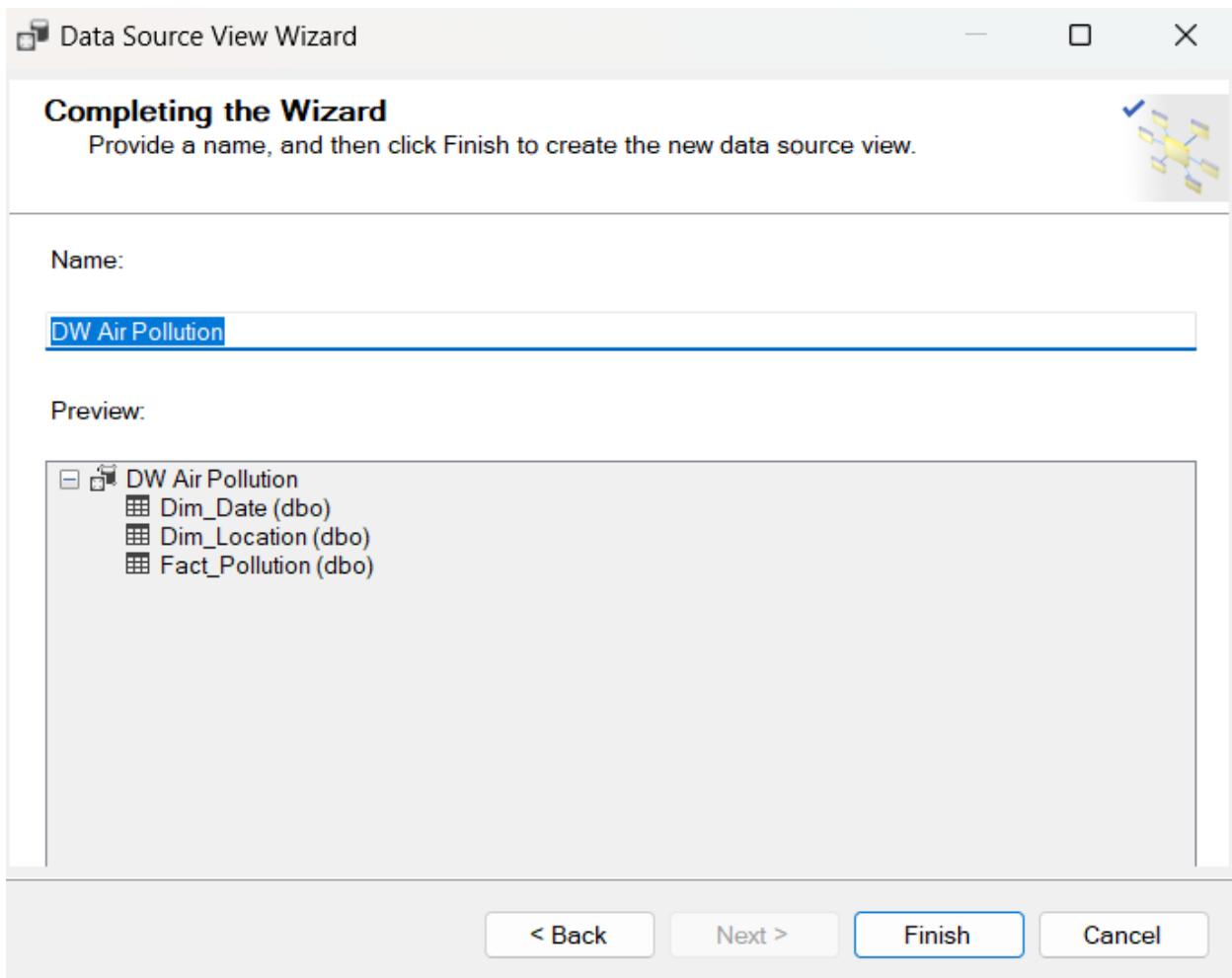
### Bước 9: Click Next.



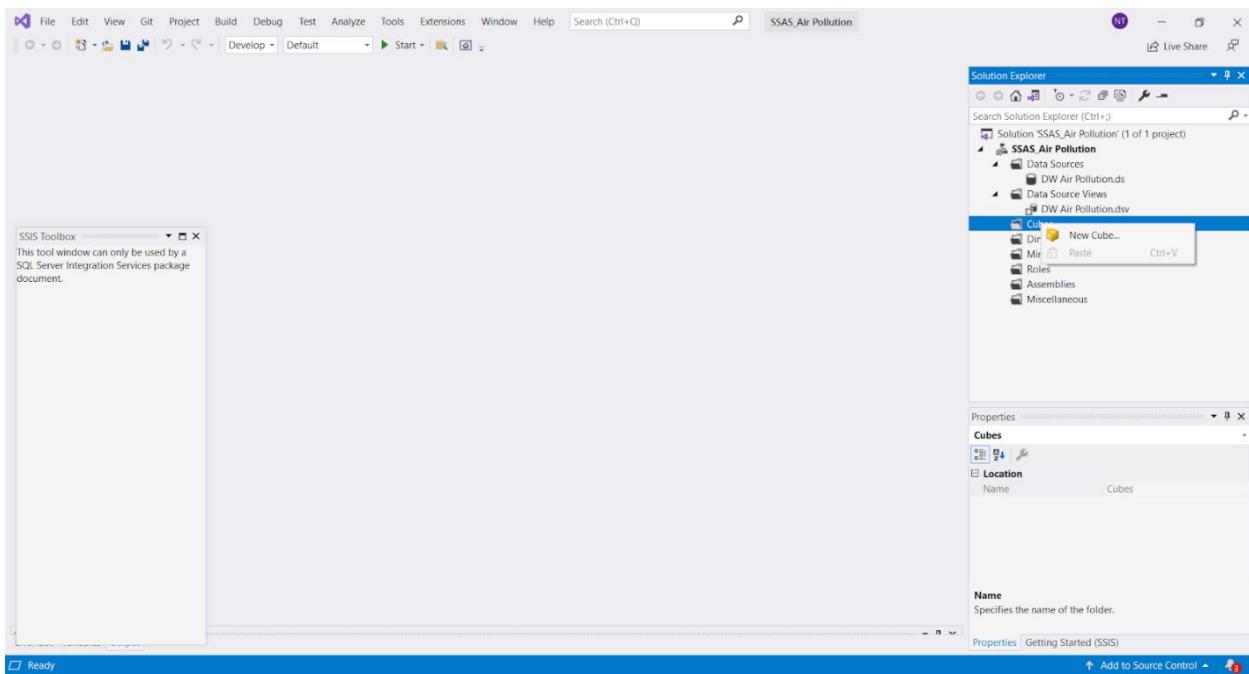
**Bước 10:** Chọn Bảng **Fact** và click **Add Related Tables** để chọn các bảng có quan hệ với bảng **Fact**. Cuối cùng chọn **Next**.



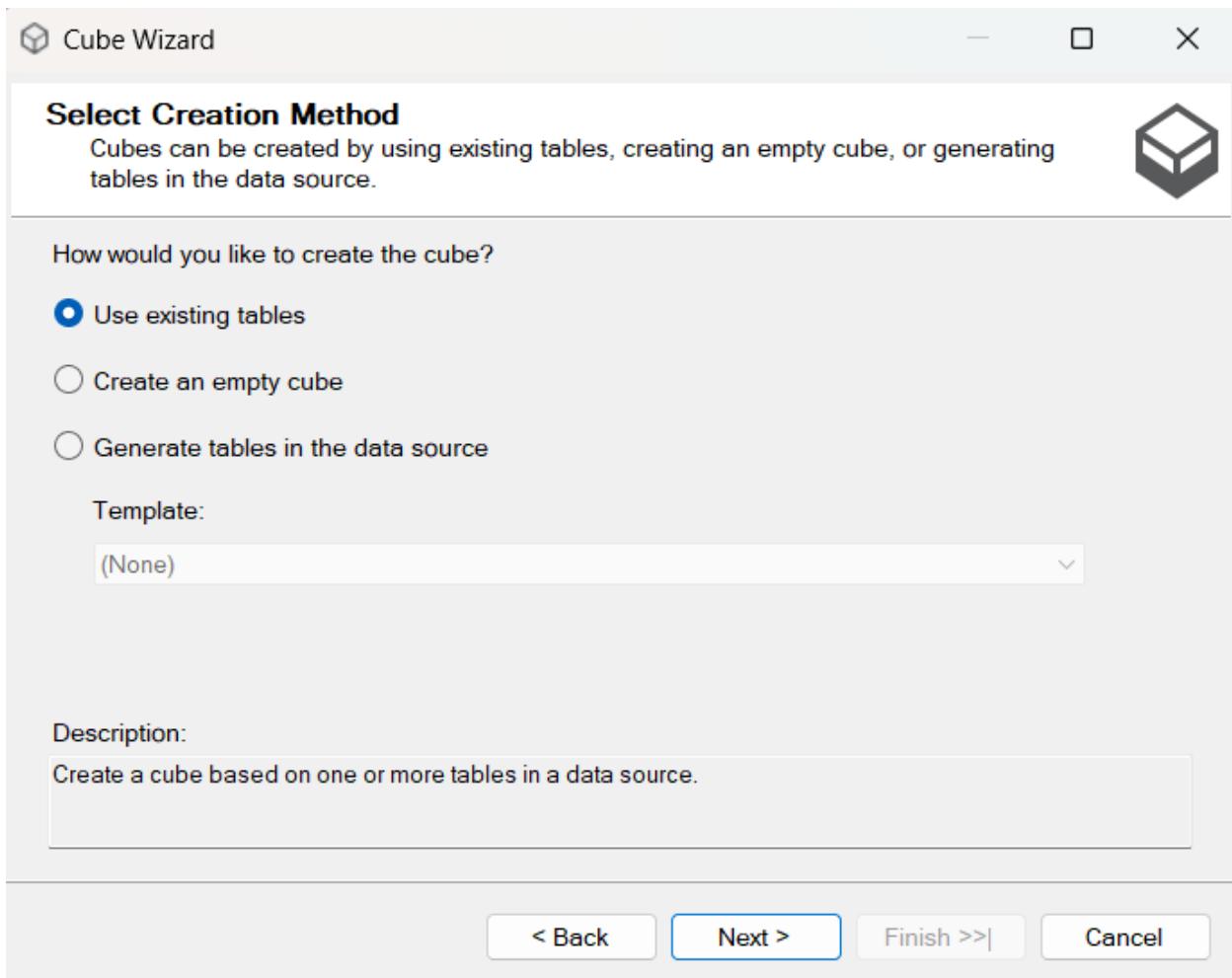
Sau khi tạo xong click chọn Next và Finish



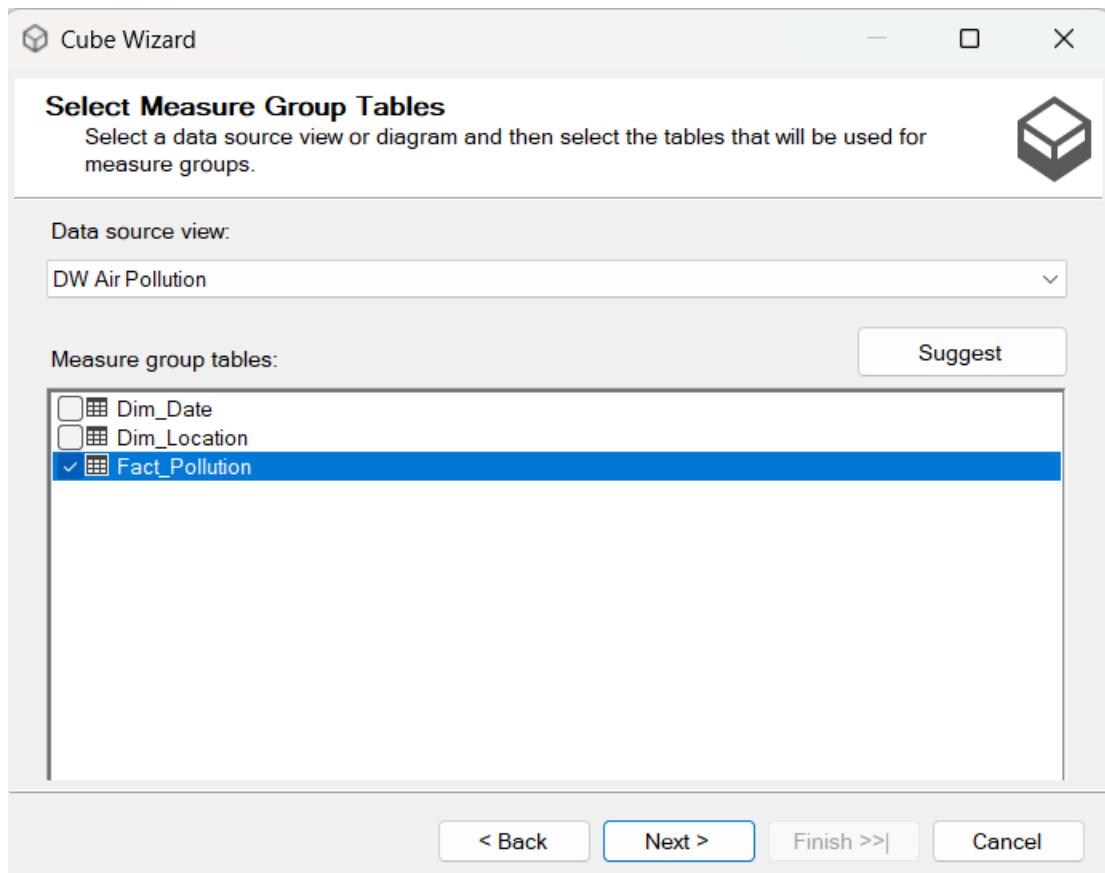
### Bước 11: Tạo Cube. Chọn New Cube



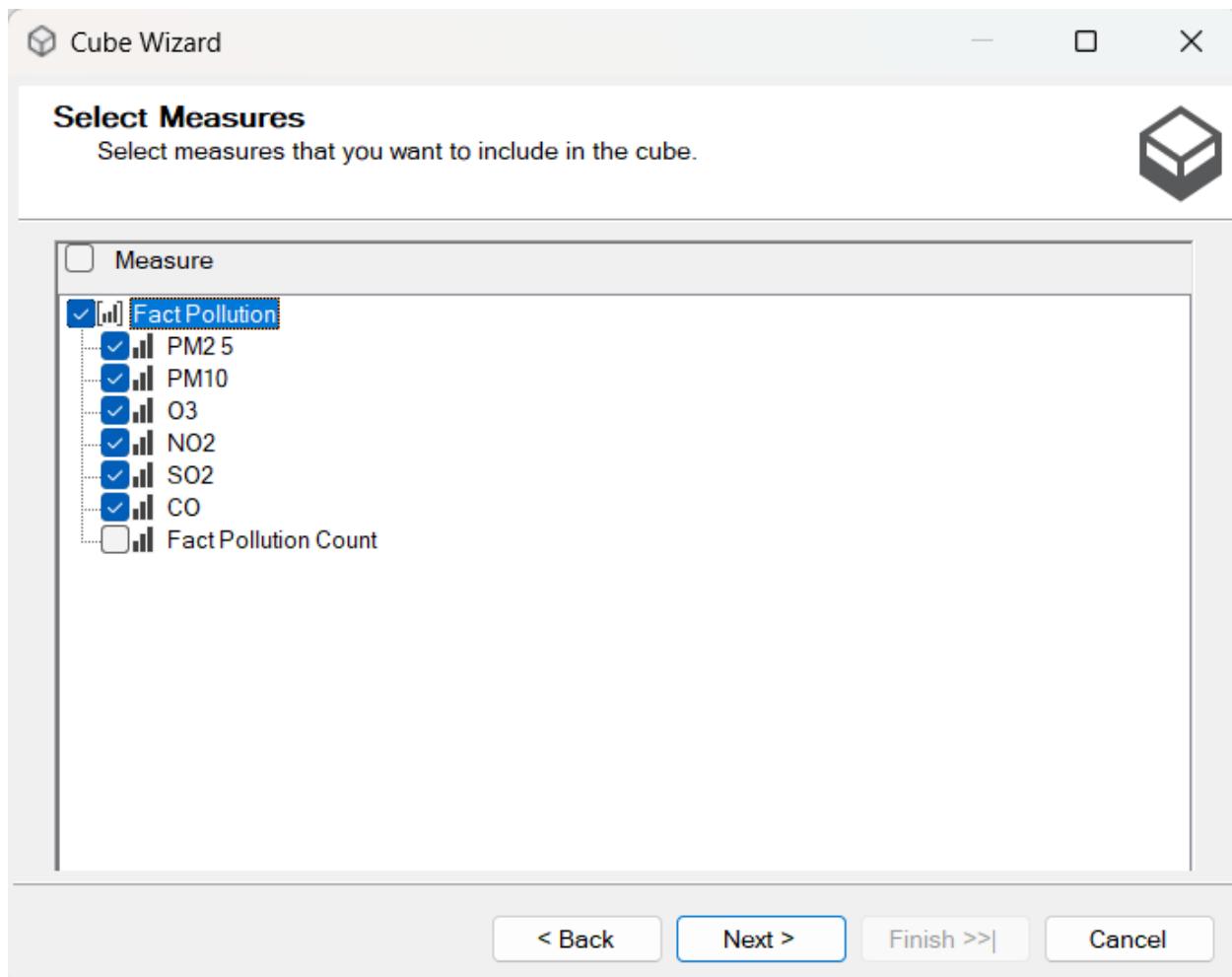
**Bước 12:** Chọn **Use existing tables** và chọn **Next**.



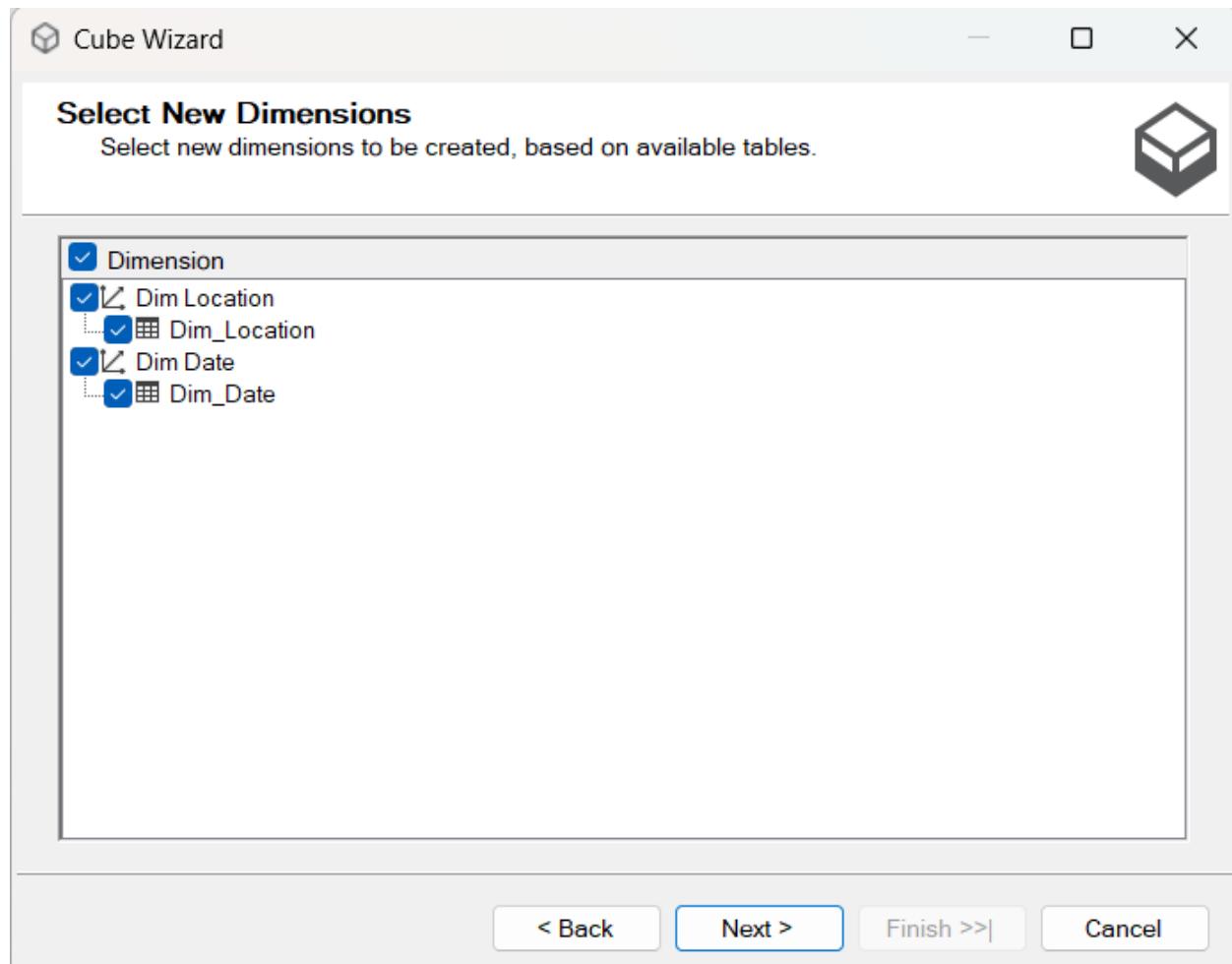
**Bước 13:** Chọn bảng **Fact** và click **Next**.

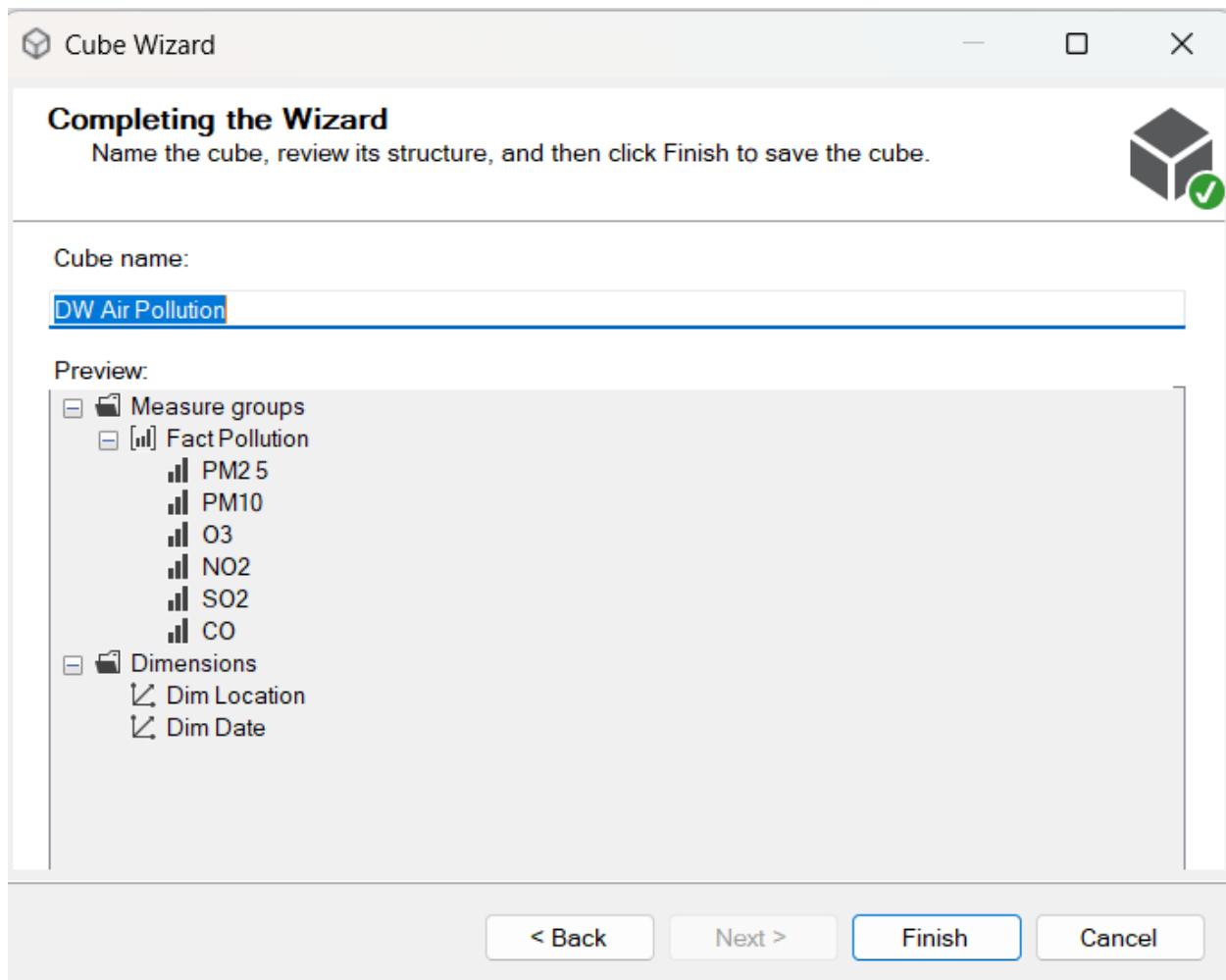


**Bước 14:** Tại bảng **Fact**, bỏ cột **Fact Pollution Count** và click **Next**.

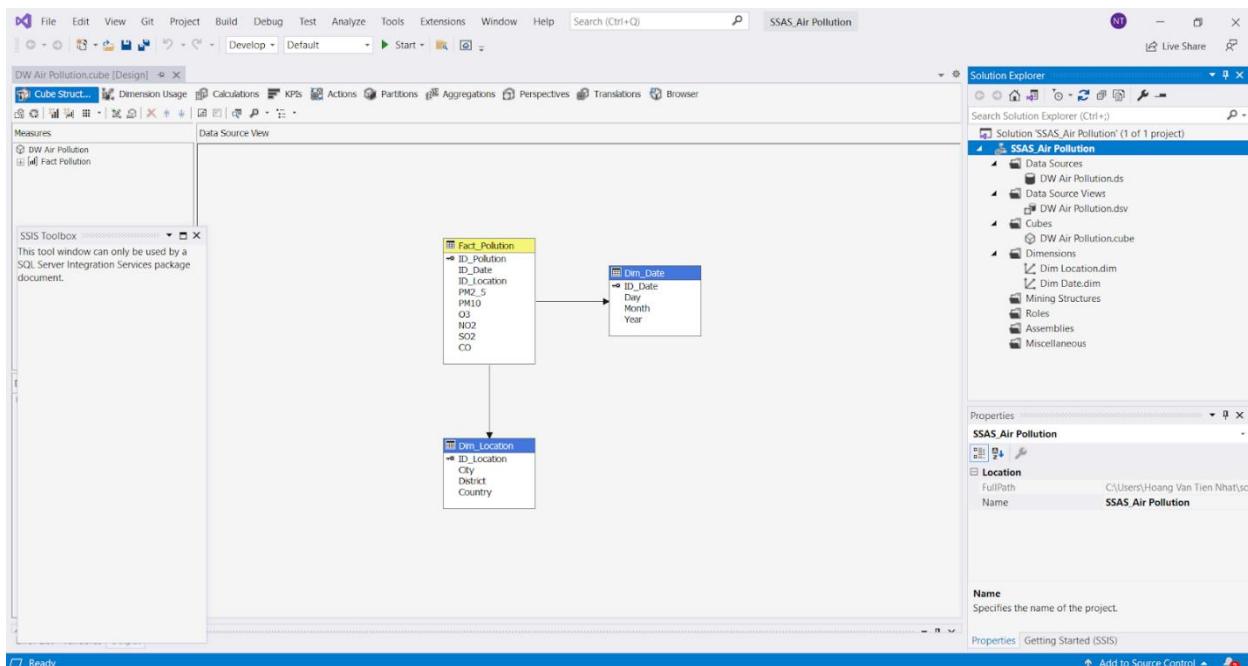


**Bước 15:** Click chọn **Next** sau khi xem xét hết các Dimension, nhấn Finish sau khi hoàn thành



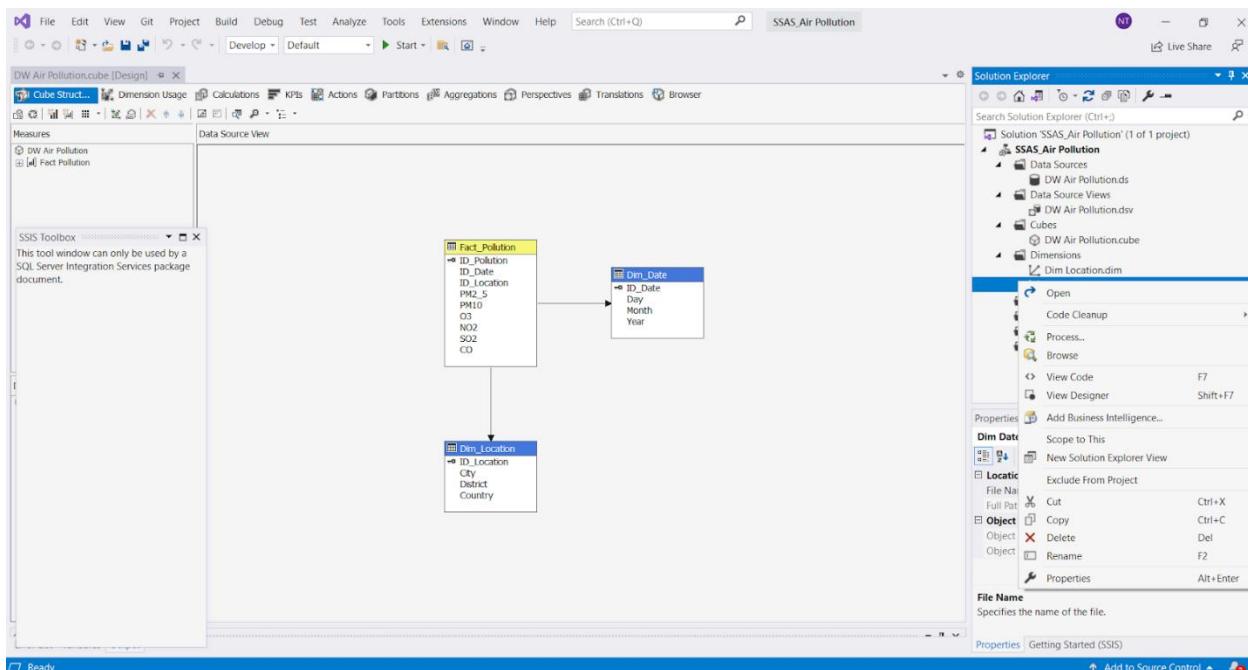


**Bước 16:** Sau khi hoàn thành, màn hình sẽ đưa ra lược đồ Data Warehouse của database.

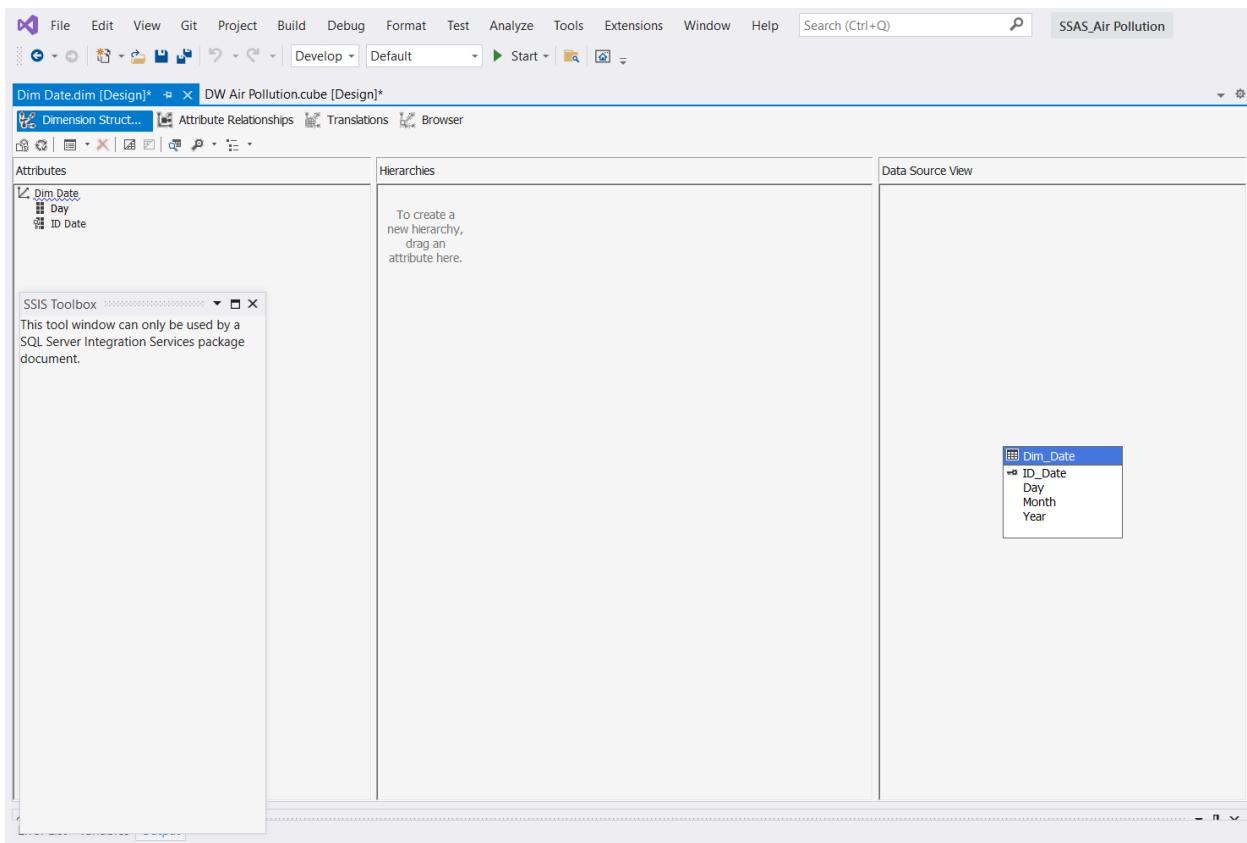


## 1.2. Điều chỉnh các Dimension

**Bước 1:** Tại Dim Date, click chọn Open để mở Dim Date.



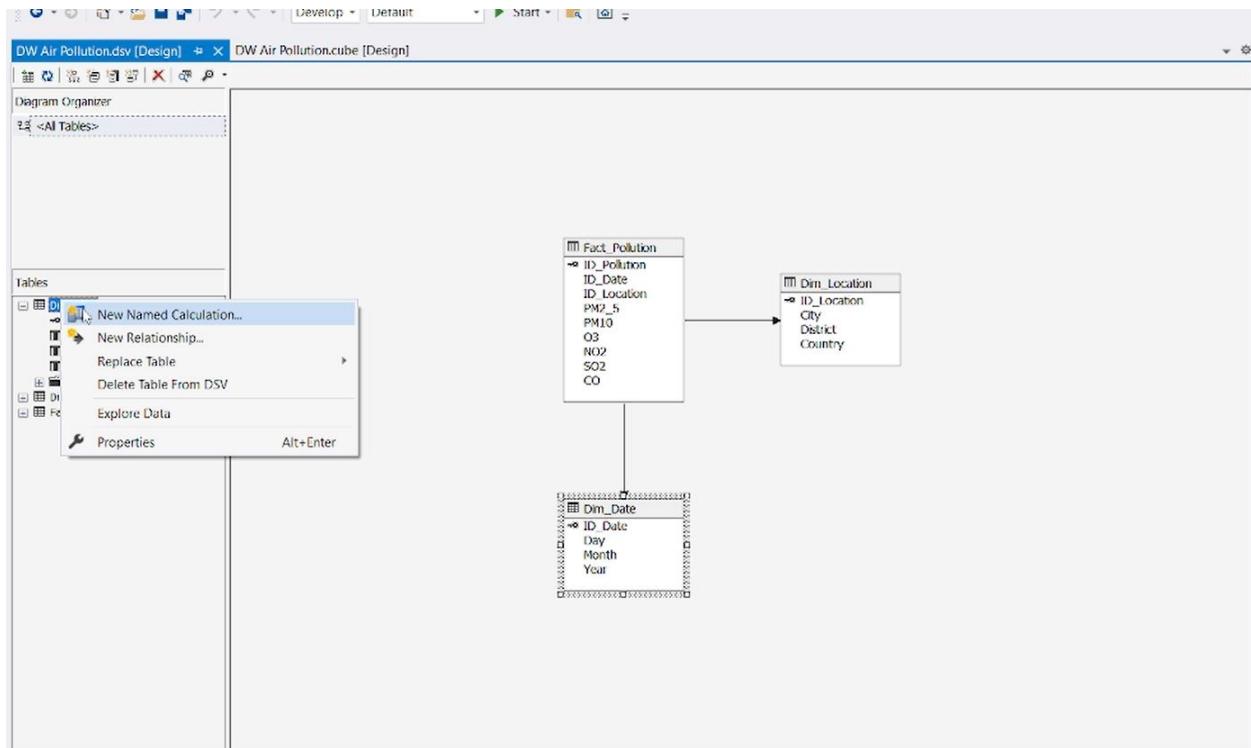
**Bước 2:** Kéo thả các thuộc tính của Dim Date vào **Attributes**.



**Bước 3:** Kéo thả các thuộc tính vừa để vào trong **Attributes** vào **Hierachies** để tạo cây phân cấp.

**Tại sao lại tạo mới “Day, Month” :** Ta sẽ ví dụ khi phân cấp Day, Month, Year trong bảng chiều Dim\_Date , bởi vì thuộc tính dữ liệu không bị trùng lặp, nếu dữ liệu của cột hiện tại có thuộc tính con và bị trùng lặp , thì sẽ không thể phân cấp rõ ràng được . Nên ta sẽ tạo thêm 1 column mới với dữ liệu không lặp hoặc dùng ID có sẵn để Unique thuộc tính bị lặp.

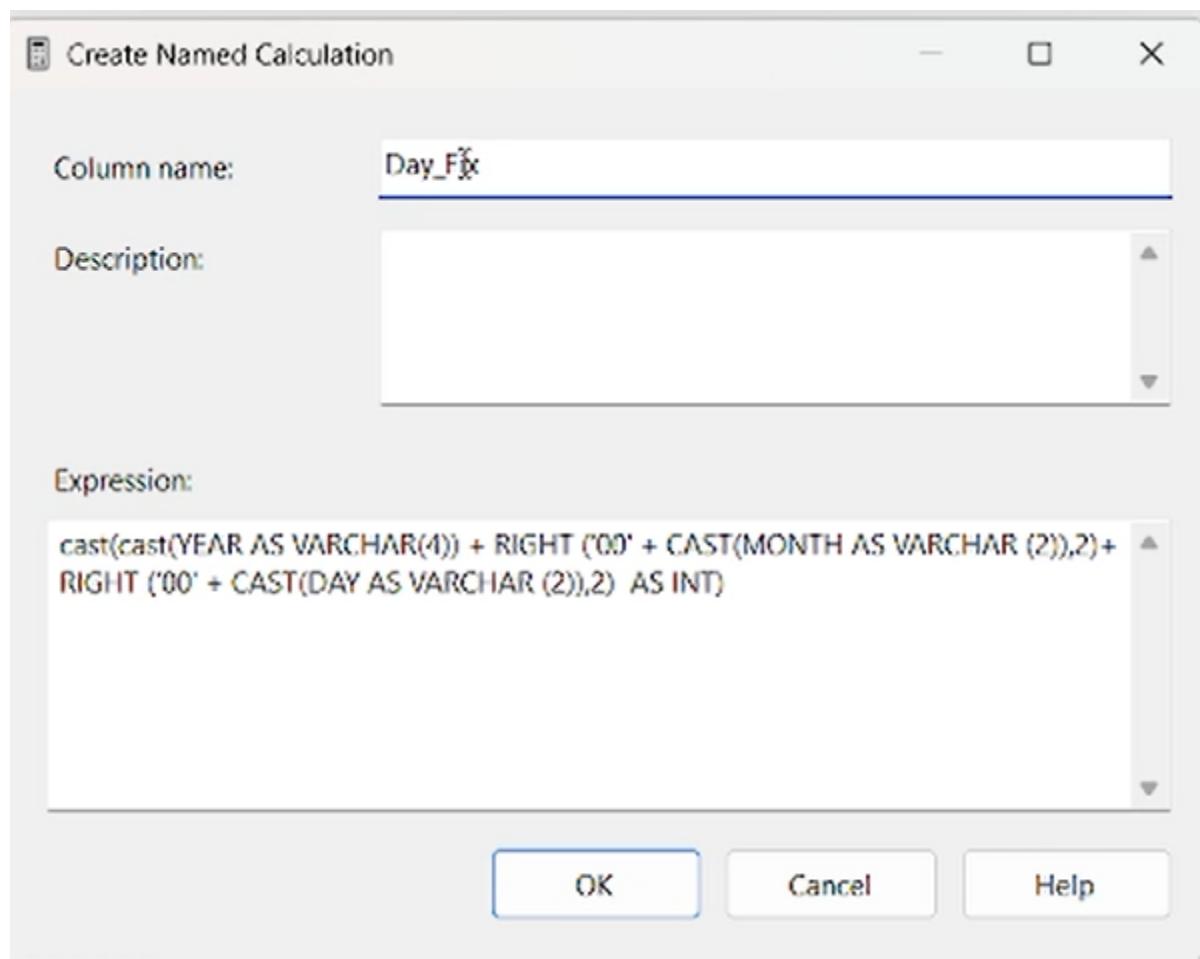
**Bước 4:** Vào Data Source View file, tìm bảng chiều cần thêm column, nhấn chuột phải và chọn Option **New Named Calculation**.



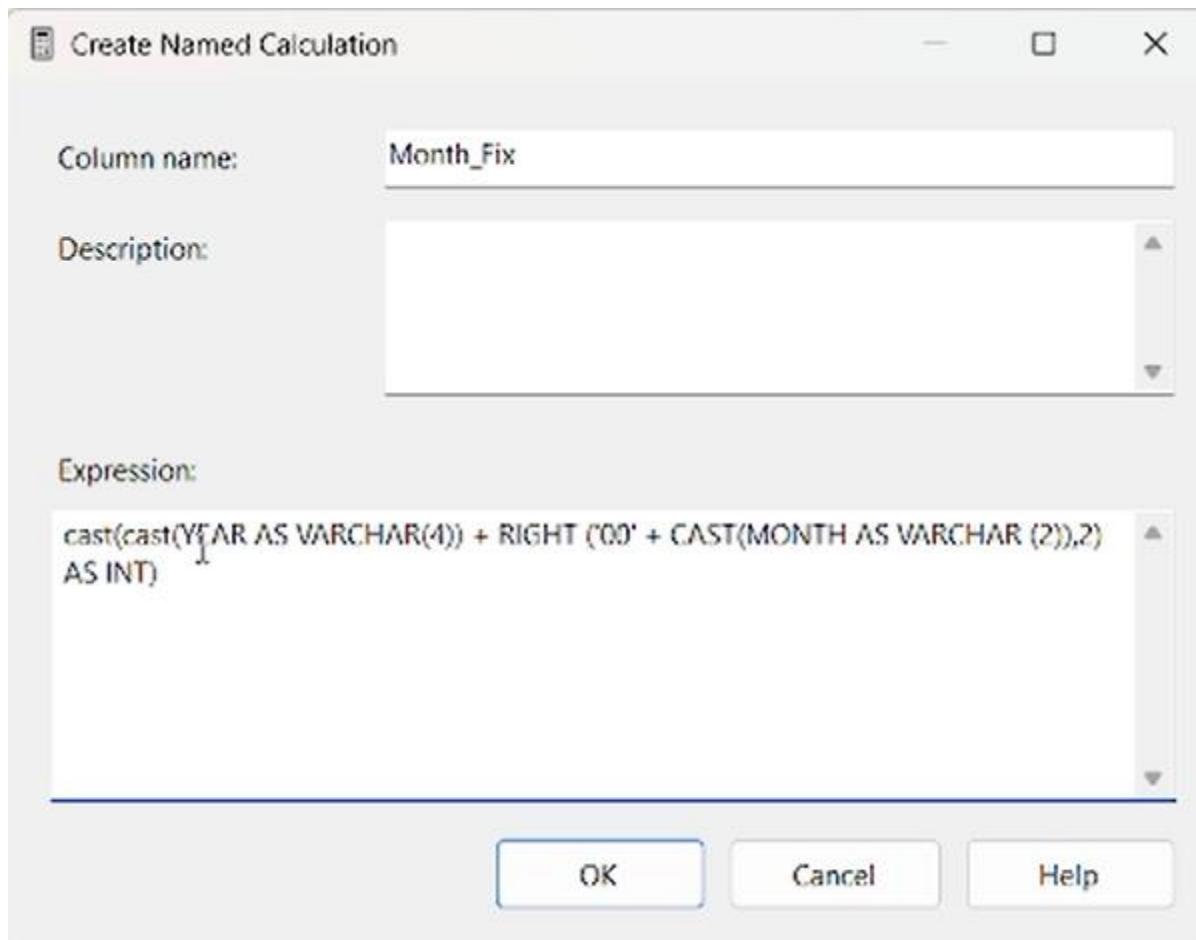
**Bước 5 :** Nhập tên Column , mô tả ( có thể không nhập) và expression. SQL expression ở đây là biểu thức tạo ra cột mới với định nghĩa ,

cast(cast(YEAR AS VARCHAR(4)) + RIGHT ('00' + CAST(MONTH AS VARCHAR (2)),2)+  
RIGHT ('00' + CAST(DAY AS VARCHAR (2)),2) AS INT)

Câu ở trên dùng để tạo ra một column Day\_fix mới từ những cột có sẵn (Year và Month, Day), nối chuỗi và chuyển kiểu thành INT , ví dụ



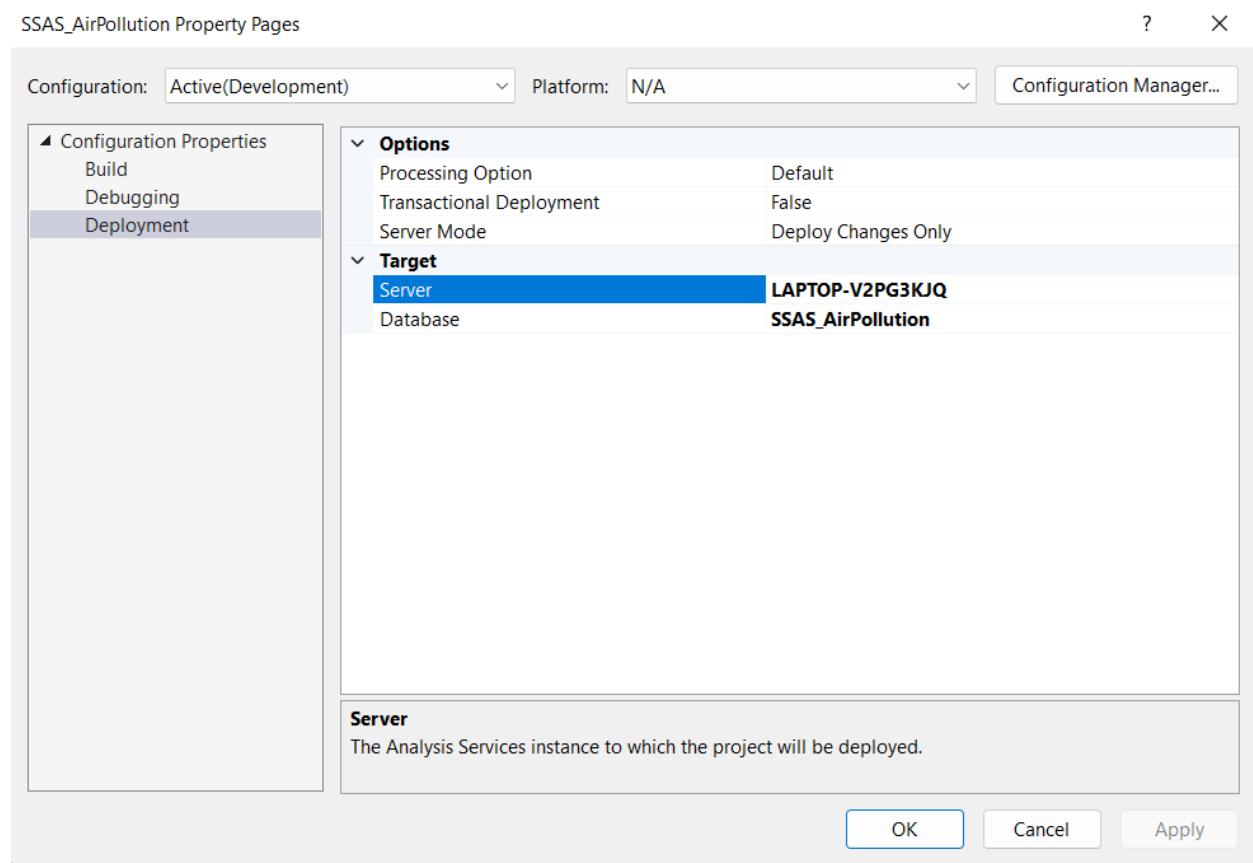
**Bước 6 :** Thực hiện tương tự cho column Month\_fix



**Bước 7 :** Vào bảng chiều Dim\_Date, kéo hai columns mới tạo vào **Attributes**, thế các thuộc tính cũ (Month và Day) bằng hai columns mới tạo (Month\_fix và Day\_fix) ở hierachies

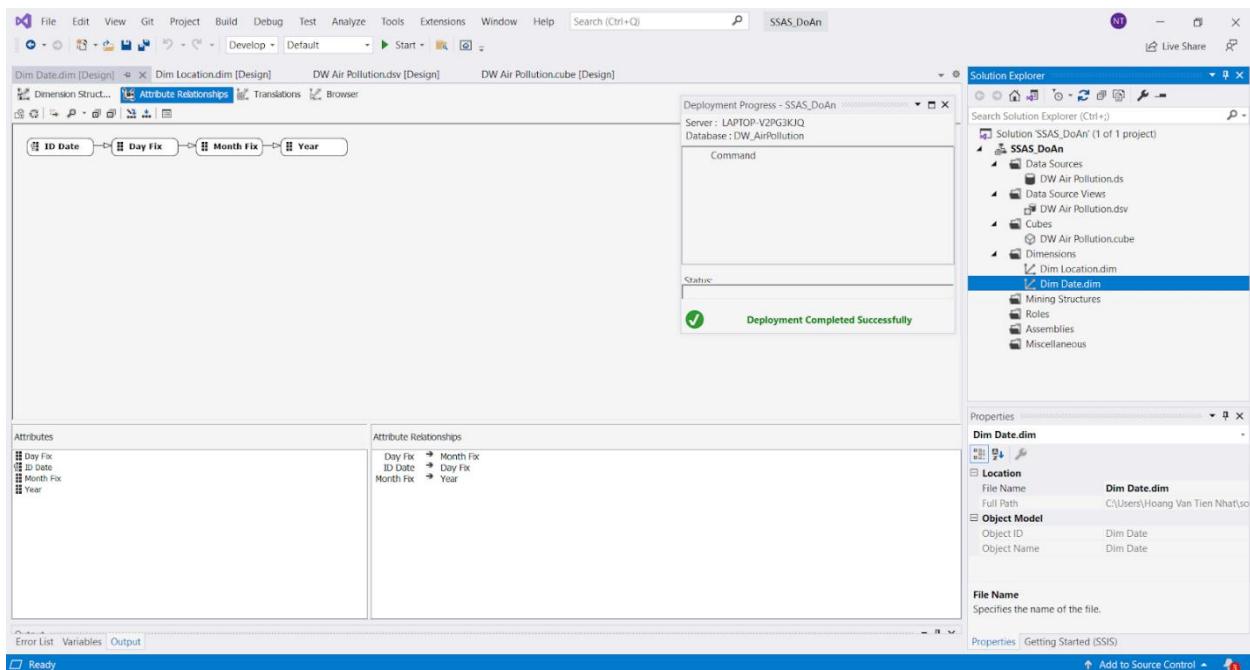
**Bước 8 :** Di chuyển qua cột **Attributes Relationship** để sắp xếp lại phân cấp

**Bước 9:** Chọn tên **Server** và **Database** cần thực hiện.

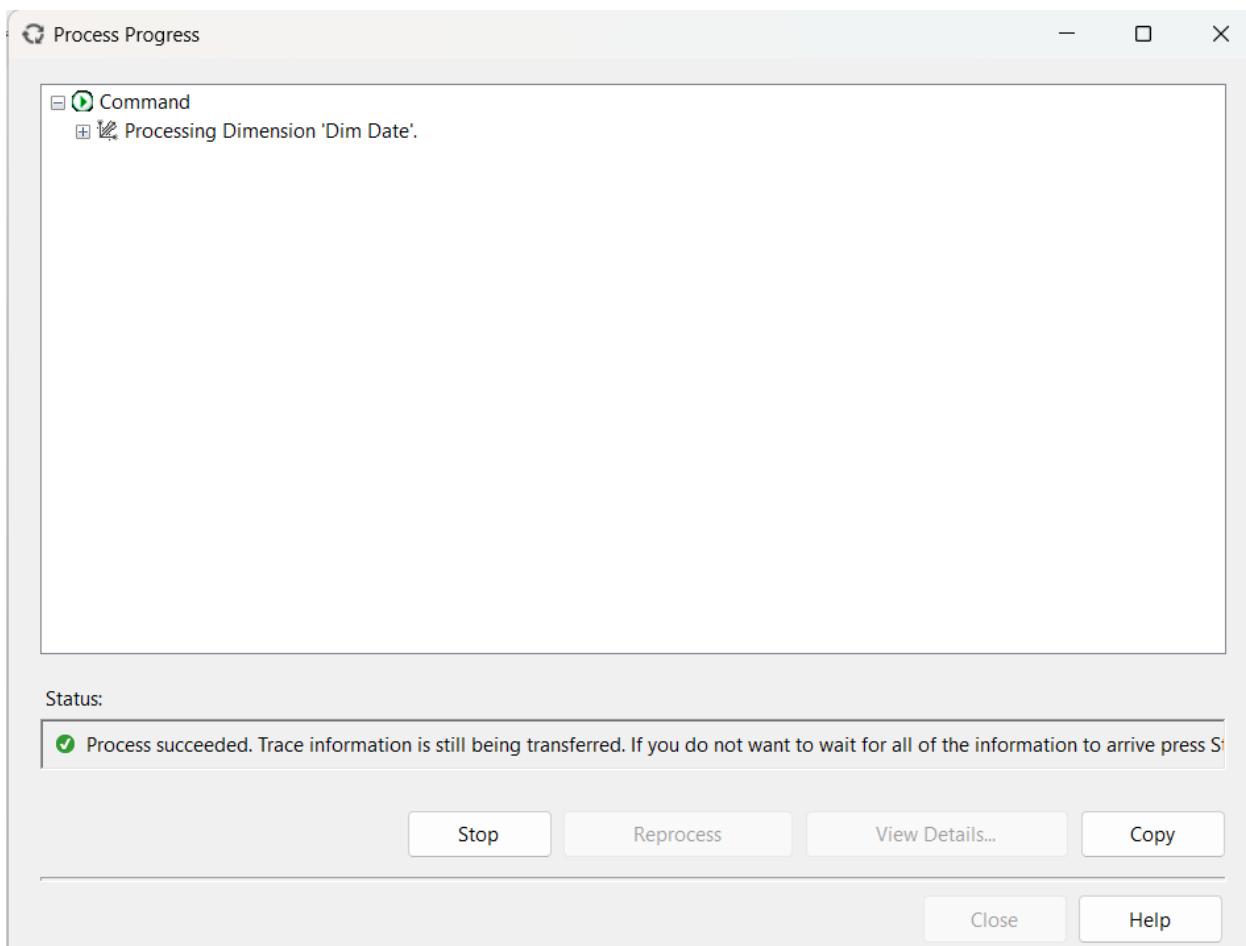


**Bước 10 :**

**Bước 10.1** Quá trình **Process** của Dim Date thực thi thành công, Click **Run** để quá trình **Process** thực thi

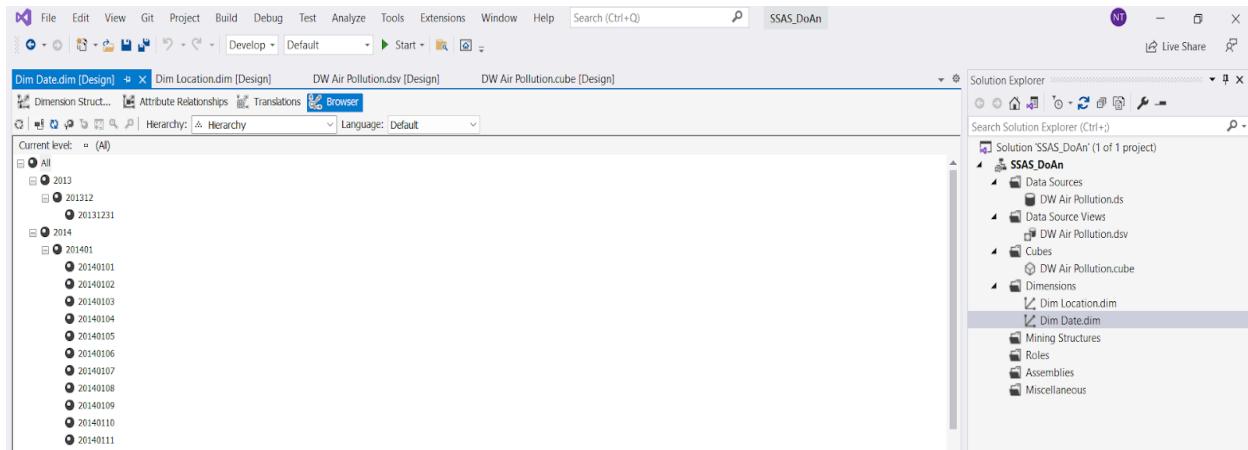


## Bước 10.2: Quá trình Process của Dim Date thực thi thành công

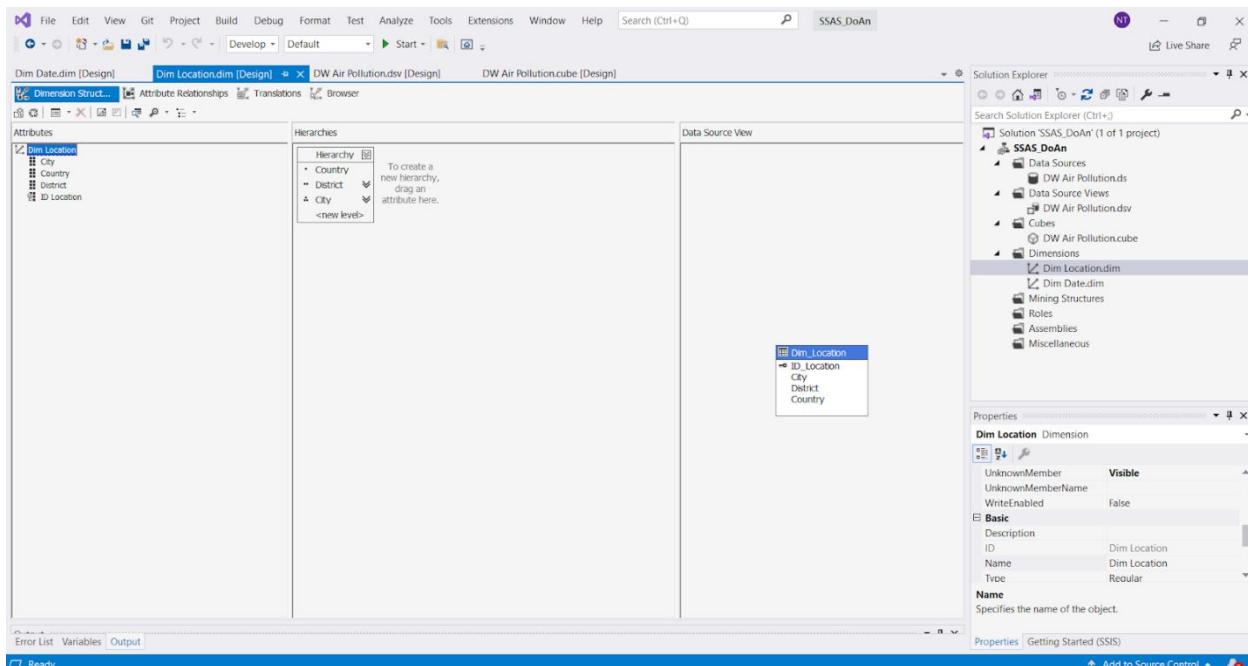


**Bước 10.3 :** Kiểm tra lại việc phân cấp của bảng Dim Date đã có xảy việc phân cấp theo thứ tự các thuộc tính.

Click vào phần **Browser** để kiểm tra và nhận thấy việc phân cấp đã thực hiện thành công.

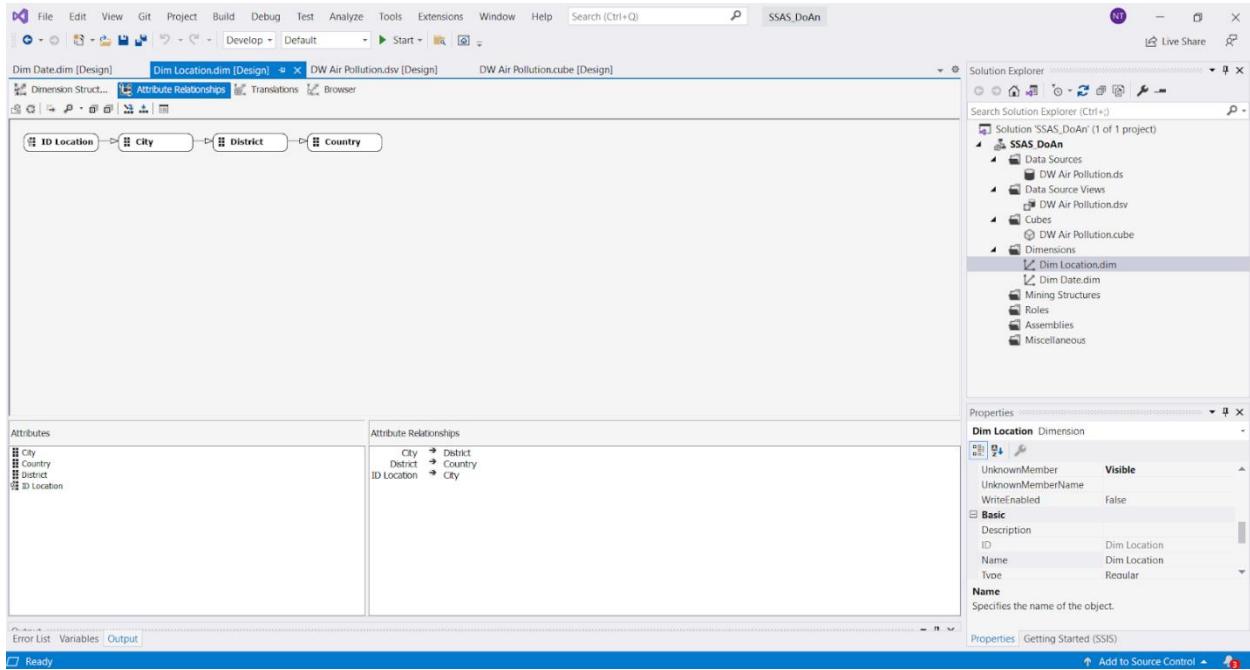


**Bước 11:** Tiếp tục thực hiện còn lại với Dim Location. Tạo **Hierarchies** cho Dim Location.

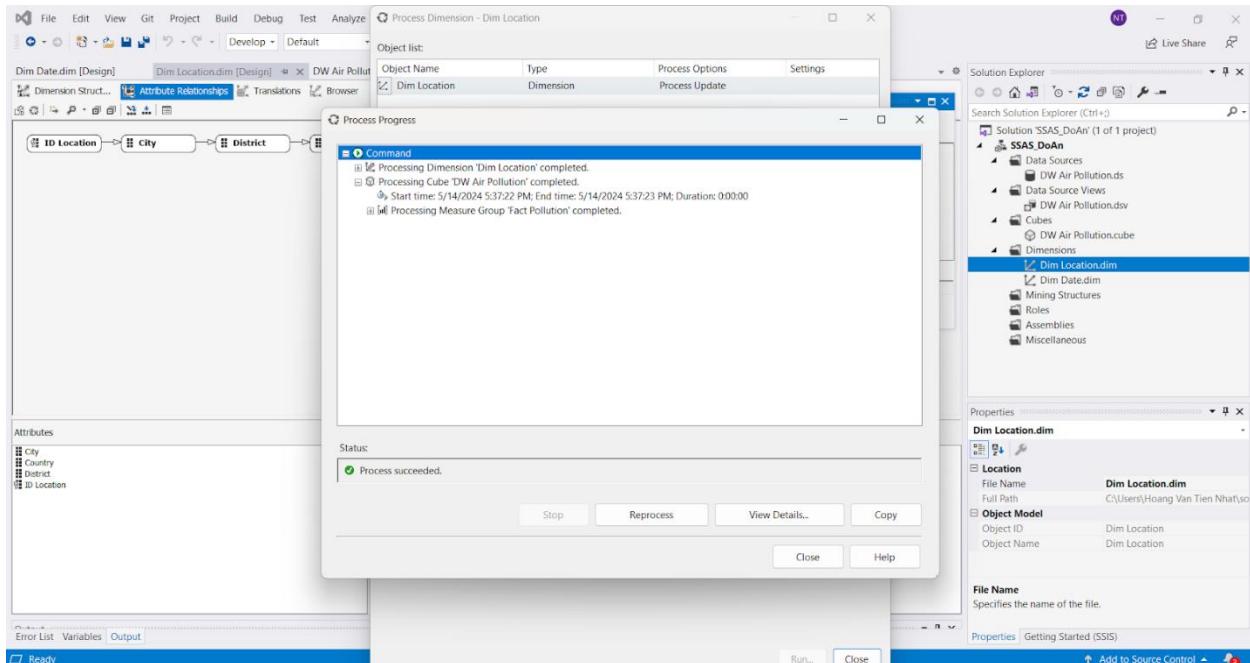


**Bước 12 :**Tại phần **Attribute Relationship**, sửa đổi một số **Attribute**

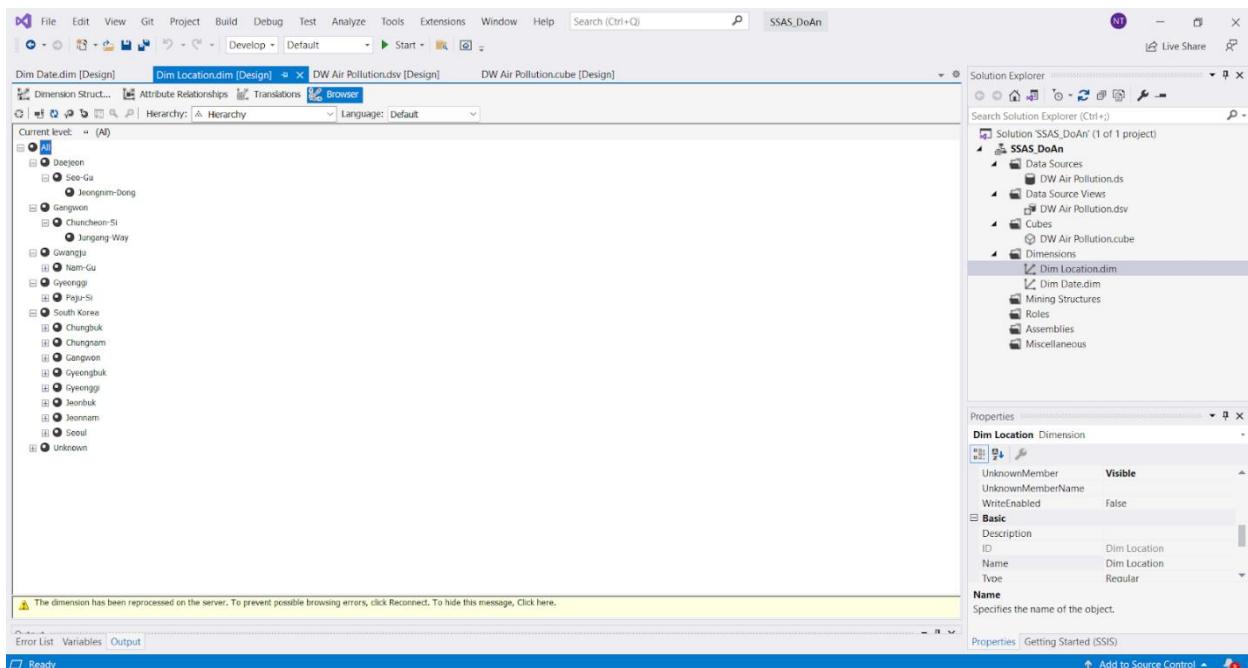
**Relationship** sau khi tạo **Hierachies** trước đó, xem lại các mối quan hệ giữa các thuộc tính sau khi sửa đổi .



**Bước 13 : Tiến hành Deploy và Process bảng Dim Location. Việc thực thi quá trình Deploy và Process thành công**



**Bước 14:** Vào **Browser** Kiểm tra việc phân cấp các thuộc tính của Dim Location đã thực hiện thành công hay không. Việc phân cấp đã thành công.



## 2. Quá trình phân tích dữ liệu bằng công cụ SSAS trên các khối CUBE

### 2.1. Cho biết tổng lượng PM25 ở thành phố Pado-Ri theo từng năm.

**Bước 1 :** Kéo thuộc tính City của bảng Dim\_Location sang cửa sổ lọc dữ liệu và cửa sổ thực thi câu truy vấn

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Location	City	Equal	<Select dimension>	

## Bước 2 : Kéo độ đo PM25 của bảng Fact\_Pollution sang cửa sổ thực thi truy vấn

The screenshot shows the SSAS cube design interface. The top navigation bar includes File, Edit, View, Git, Project, Build, Debug, Test, Analyze, Tools, Extensions, Window, Help, and a search bar. Below the navigation bar, tabs for Dim Date.dim [Design], Dim Location.dim [Design], DW Air Pollution.dsv [Design], and DW Air Pollution.cube [Design] are visible. The DW Air Pollution.cube [Design] tab is active. The main workspace displays the Fact\_Pollution table with a single row: City (Pado-Ri) and PM2 5 (89201). On the left, a tree view shows the schema: DW Air Pollution (Measure Group), KPIs, Dim Date (Hierarchy, Day Fix, ID Date, Month Fix, Year, Hierarchy), Dim Location (City, Country, District, ID Location), and Calculated Members.

## Bước 3 : Tại cửa sổ lọc trường Filter Expression chọn thành phố Pado-Ri

This screenshot shows the same SSAS cube design interface as the previous one, but with a modal dialog open over the workspace. The dialog is titled 'Filter Expression' and contains a list of locations under the 'City' hierarchy. The 'Pado-Ri' option is checked. At the bottom right of the dialog are 'OK' and 'Cancel' buttons.

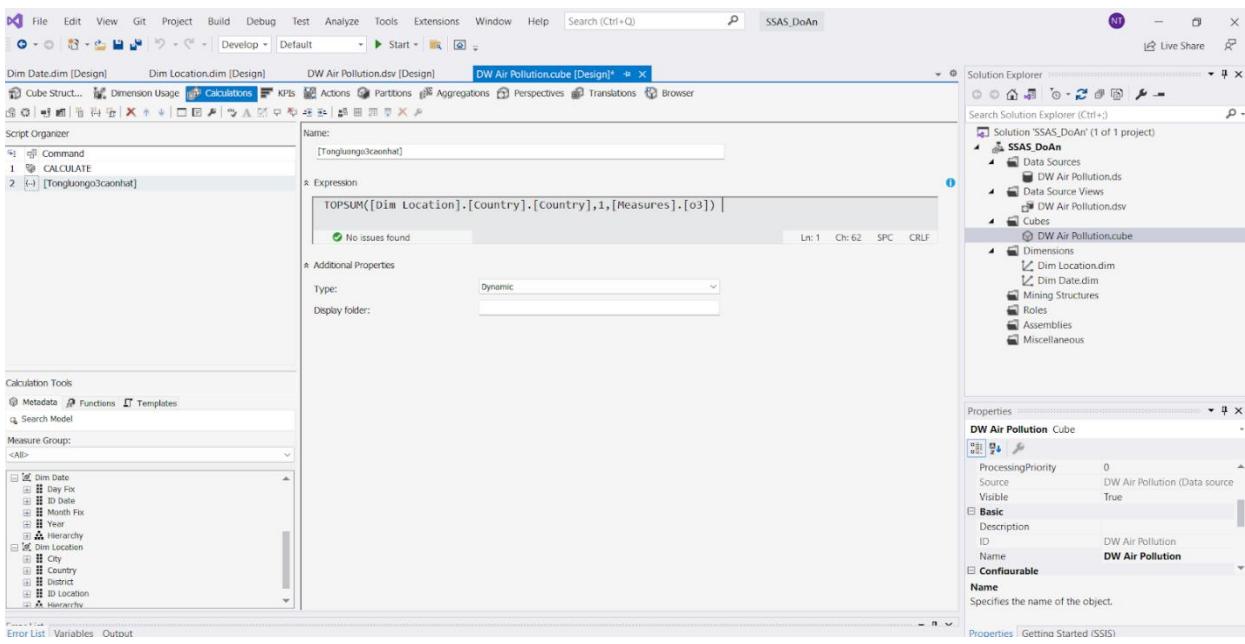
## Bước 4 : Kéo thuộc tính Year từ bảng Dim\_Date sang để lấy từng năm .

Kết quả của câu truy vấn

City	Year	PM2 5
Pado-Ri	2014	0
Pado-Ri	2015	229
Pado-Ri	2016	46
Pado-Ri	2017	0
Pado-Ri	2018	12542
Pado-Ri	2019	23537
Pado-Ri	2020	22294
Pado-Ri	2021	25907
Pado-Ri	2022	4646

## 2.2. Cho biết vùng có chỉ số O3 cao nhất trong tháng 6 năm 2017.

Bước 1: Tạo một nameset Tongluongo3caonhat dùng để lấy vùng có lượng o3 cao nhất.



Bước 2 : Kéo thuộc tính Country và độ đo 03 sang cửa sổ thực thi truy vấn.

Country	O3
Daejeon	95087
Gangwon	96594
Gwangju	89978
Gyeonggi	100431
South Korea	809415

Bước 3 : Lọc theo từng tháng và dùng mame set Tongluongo3caonhat. Và

có kết quả như hình bên dưới:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Date	Country	In	{ Tongluongo3caonhat }	
Dim Date	Hierarchy	Equal	{ 201706 }	

Country	O3
South Korea	8632

### 2.3. Cho biết thành phố nào có tổng lượng Pm10 trên 50000.

**Bước 1 :** Tạo một nameset TongLuongPm10 > 50000 dùng để lấy tổng lượng khí Pm10 lớn hơn 50000 ug/m<sup>3</sup> .

The screenshot shows the SSAS Management Studio interface. The top menu bar includes File, Edit, View, Git, Project, Build, Debug, Test, Analyze, Tools, Extensions, Window, Help, and a search bar. The title bar shows the project name "SSAS\_DoAn". The main workspace has tabs for "Dim Date.dim [Design]", "Dim Location.dim [Design]", "DW Air Pollution.dsv [Design]", and "DW Air Pollution.cube [Design]\*". The "DW Air Pollution.cube [Design]" tab is active. The left sidebar contains "Script Organizer" and "Calculation Tools" sections. The "Calculation Tools" section shows a tree view of measures: DW Air Pollution > Measures > Fact Pollution > CO, NO2, O3, PM10, PM2 5, SO2. Below this is a "Dim Date" section with Day Fix, ID Date, and TD Date. The right pane displays the creation of a calculated member. The "Name" field contains "[TongluongPm10 > 50000]". The "Expression" field contains the MDX code: `FILTER([Dim Location].[city].[City],[Measures].[PM10] > 50000)`. A status bar at the bottom right indicates "Ln: 1 Ch: 63 SPC CRLF".

Bước 2 : Kéo thuộc tính City và độ đo Pm10 sang cửa sổ thực thi truy vấn .

The screenshot shows the SSAS Management Studio interface with the same layout as the previous one. The "DW Air Pollution.cube [Design]" tab is active. The left sidebar shows the cube structure with dimensions like Dim Date, Dim Location, and measures like PM10, PM2 5, SO2. The right pane displays the results of an MDX query. The query results table has columns "City" and "PM10". The data is as follows:

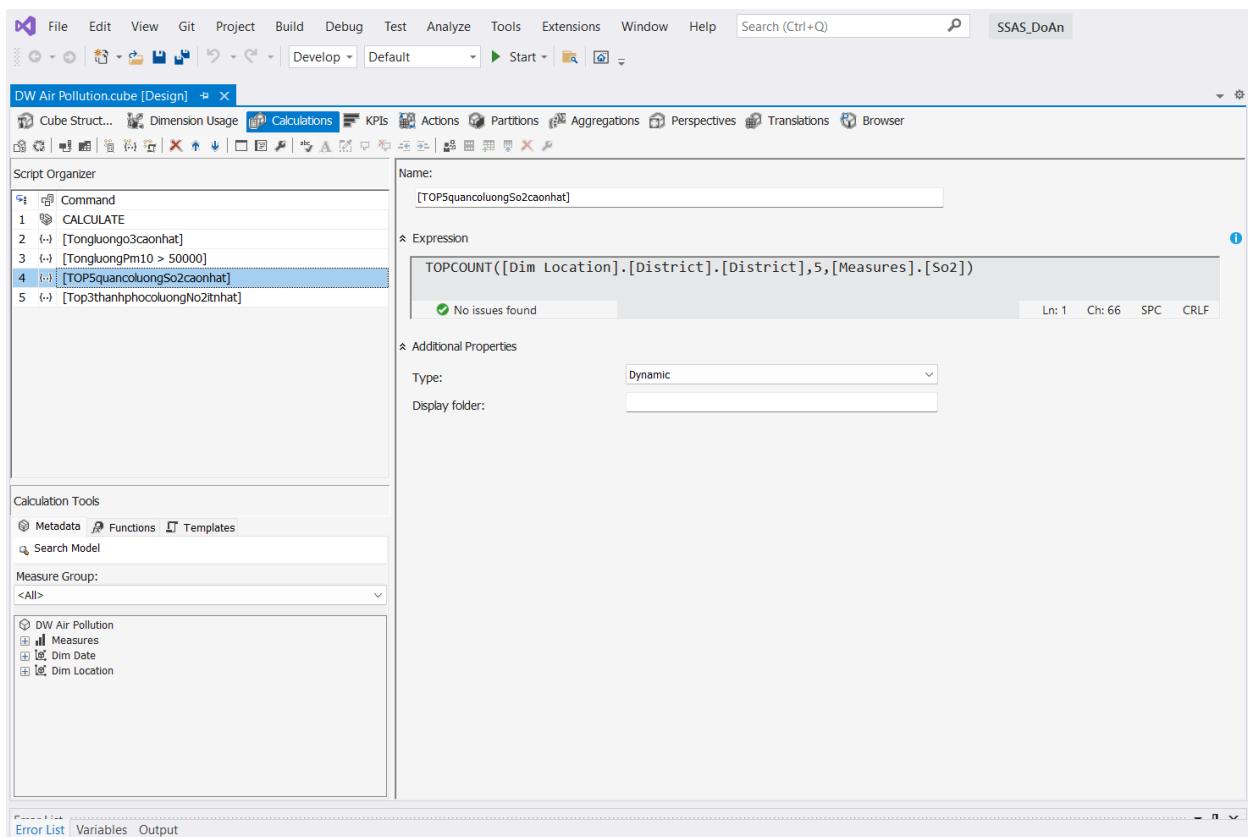
City	PM10
Bangsan-Myeon	76339
Cheongnim-Dong	29642
Geumchon-Dong	117789
Gwanak-Gu	107644
Hwasun-Eup	27019
Hyeoksin-Dong	20154
Jeongnim-Dong	91570
Jungang-Way	99428
Juwol-Dong	97010
Nowon-Gu	99531
Onui-Dong	20057
Pado-Ri	102347
Paju	58016
Saemangeum	497
Saesol-Dong	18595
Sangjusi	44672
Seolseong-Myeon	119361
Seonggeo-eup	51858
Sol-Myeon	18372
Taeha-Ri	273

Bước 3 : Lọc theo từng thành phố trong thuộc tính phân cấp và dùng nameset **TongluongPm10 > 50000**. Vf sẽ có kết quả như hình bên dưới.

City	PM10
Bangsan-Myeon	76339
Geumchon-D...	117789
Gwanak-Gu	107644
Jeongnim-Dong	91570
Jungang-Way	99428
Juwol-Dong	97010
Nowon-Gu	99531
Pado-Ri	102347
Paju	58016
Seolseong-My...	119361
Seonggeoeup	51858

#### 2.4. Cho biết top 5 quận có chỉ số So2 cao nhất trong từng năm.

Bước 1: Tạo một nameset **Top5quancoluongSo2caonhat** dùng để lấy quận có lượng So2 thấp nhất.



Bước 2 : Kéo thuộc tính District và độ đo So2 sang cửa sổ thực thi truy vấn .

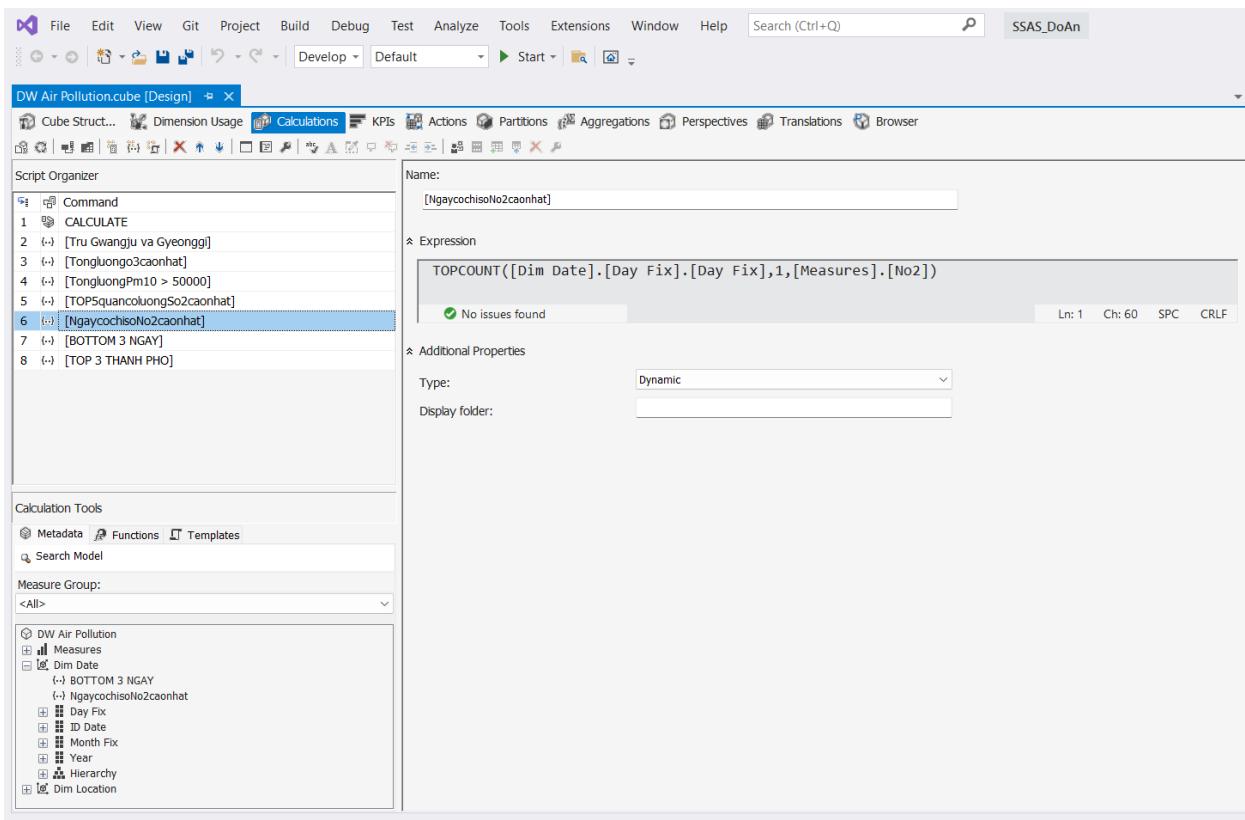
District	SO2
Chuncheon-Si	10625
Chungbuk	1148
Chungnam	12708
Gangwon	4988
Gyeongbuk	12368
Gyeonggi	10320
Jeonbuk	1387
Jeonnam	2899
Nam-Gu	10375
Paju-Si	13216
Seo-Gu	7893
Seoul	34761

Bước 3 : Lọc theo từng năm và dùng name set **Top5quancoluongSo2caonhat** . Và có kết quả như hình bên dưới:

The screenshot shows the Microsoft Analysis Services (SSAS) cube editor interface. The top menu bar includes File, Edit, View, Git, Project, Build, Debug, Test, Analyze, Tools, Extensions, Window, Help, and a search bar. The title bar says "DW Air Pollution.cube [Design]". The main area has tabs for Cube Struct..., Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser. A toolbar below the tabs includes icons for Edit as Text, Import..., MDX, and various navigation and search tools. On the left, a tree view shows the cube structure with nodes for DW Air Pollution, Metadata, Measure Group, and various dimensions like Dim Location, Dim Date, and KPIs. The right side contains a table with columns for Dimension, Hierarchy, Operator, Filter Expression, and Parameters. The table shows data for Dim Location (District) and Dim Date (Year). Below the table is a grid showing SO2 values for different districts. At the bottom, there are tabs for Error List, Variables, and Output.

## 2.5. Cho biết ngày nào có tổng lượng NO<sub>2</sub> cao nhất .

Bước 1: Tạo một nameset **NgaycochisoNo2caonhat** dùng để lấy ngày có lượng SO<sub>2</sub> cao nhất.



Bước 2 : Kéo thuộc tính Day Fix và độ đo No2 sang cửa sổ thực thi truy vấn .

ID Date	SO2
2013-12-31	78
2014-01-01	60
2014-01-02	61
2014-01-03	51
2014-01-04	49
2014-01-05	69
2014-01-06	78
2014-01-07	67
2014-01-08	32
2014-01-09	59
2014-01-10	66
2014-01-11	90
2014-01-12	43
2014-01-13	57
2014-01-14	77
2014-01-15	91
2014-01-16	80
2014-01-17	60
2014-01-18	61
2014-01-19	59
2014-01-20	85
2014-01-21	71
2014-01-22	62

Bước 3 : Lọc theo từng ngày và dùng name set **NgaycochisoNo2caonhat**. Và có kết quả như hình bên dưới:

DW Air Pollution.cube [Design]

Dimension Hierarchy Operator Filter Expression Parameters

Dim Date	Day Fix	In	NgaycochisoNo2caonhat	<input type="checkbox"/> <input type="checkbox"/>
----------	---------	----	-----------------------	---

Calculated Members

## 2.6. Cho biết tổng lượng CO ở vùng Deajeon theo từng năm.

**Bước 1 :** Kéo thuộc tính Country của bảng Dim\_Location sang cửa sổ lọc dữ liệu và cửa sổ thực thi câu truy vấn

DW Air Pollution.cube [Design]

Dimension Hierarchy Operator Filter Expression Parameters

Dim Location	Country	Equal		<input type="checkbox"/> <input type="checkbox"/>
--------------	---------	-------	--	---

Calculated Members

**Bước 2 :** Kéo độ đo CO của bảng Fact\_Pollution sang cửa sổ thực thi truy vấn

DW Air Pollution.cube [Design] SSAS\_DoAn

Cube Structure | Dimension Usage | Calculations | KPIs | Actions | Partitions | Aggregations | Perspectives | Translations | Browser

Language: Default

Filter Expression:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Location	Country	Equal	<Select dimension>	<input type="checkbox"/> <input type="checkbox"/>

Table Data:

Country	CO
Daejeon	13850
Gangwon	13869
Gwangju	15229
Gyeonggi	17468
South ...	96641

### Bước 3 : Tại cửa sổ lọc trường Filter Expression chọn vùng Deajeon

DW Air Pollution.cube [Design] SSAS\_DoAn

Cube Structure | Dimension Usage | Calculations | KPIs | Actions | Partitions | Aggregations | Perspectives | Translations | Browser

Language: Default

Filter Expression:

Deajeon

OK Cancel

Bước 4 : Kéo thuộc tính Month fix từ bảng Dim Date sang để lấy từng năm, ta được kết quả.

```

SELECT
    [Dim Location].[Country].<Select dimension>
    ON COLUMNS
    ,[Measures].[CO]
    ON ROWS
FROM [DW Air Pollution]
WHERE [Dim Location].[Country] = 'Daejeon'
GROUP BY [Dim Location].[Year]

```

Country	Year	CO
Daejeon	2013	7
Daejeon	2014	2012
Daejeon	2015	1950
Daejeon	2016	1775
Daejeon	2017	1283
Daejeon	2018	1574
Daejeon	2019	1865
Daejeon	2020	1346
Daejeon	2021	1718
Daejeon	2022	320

## 2.7. Đưa ra 3 ngày có lượng khí Pm25 ít nhất tại Paju .

Bước 1 : Xây dựng name set [BOTTOM 3 NGAY]

```

[BOTTOM 3 THANH]
{BottomCount(
    NonEmpty(
        [Dim Date].[Month Fix].[Month Fix],
        [Measures].[PM2 5]
    ),
    3,
    [Measures].[PM2 5]
)}

```

Bước 2 : Deploy Nameset sau đó qua Browser kéo thả những thuộc tính cần thiết cho câu truy vấn ta được kết quả.

## 2.8. Cho biết 3 thành phố có lượng khí Pm10 nhiều nhất trong năm 2017.

Bước 1 : Tạo Name Set [TOP 3 THANH PHO] như hình :

Bước 2 : Kéo thả thuộc tính City trong bảng Dim Location, Year trong bảng Dim Date và độ đo Pm10.

DW Air Pollution.cube [Design] SSAS\_DoAn

Dimension Hierarchy Operator Filter Expression Parameters

<Select dimension>

City	Year	PM10
Bangsan...	2014	12959
Bangsan...	2015	10648
Bangsan...	2016	10299
Bangsan...	2017	8744
Bangsan...	2018	7991
Bangsan...	2019	8524
Bangsan...	2020	7355
Bangsan...	2021	8387
Bangsan...	2022	1432
Cheongni...	2019	941
Cheongni...	2020	12963
Cheongni...	2021	13771
Cheongni...	2022	1967
Geumcho...	2013	0
Geumcho...	2014	16502
Geumcho...	2015	15497
Geumcho...	2016	15555
Geumcho...	2017	13316
Geumcho...	2018	13499
Geumcho...	2019	14486
Geumcho...	2020	13061
Geumcho...	2021	13792
Geumcho...	2022	2081
Gwanak-Gu	2013	0
Gwanak-Gu	2014	14991

Bước 3: Kéo thả Name Set [TOP 3 THANH PHO] sang trường lọc và lọc theo năm 2017 ta được kết quả như hình :

DW Air Pollution.cube [Design] SSAS\_DoAn

Dimension Hierarchy Operator Filter Expression Parameters

Dim Date Year Equal { 2017 }

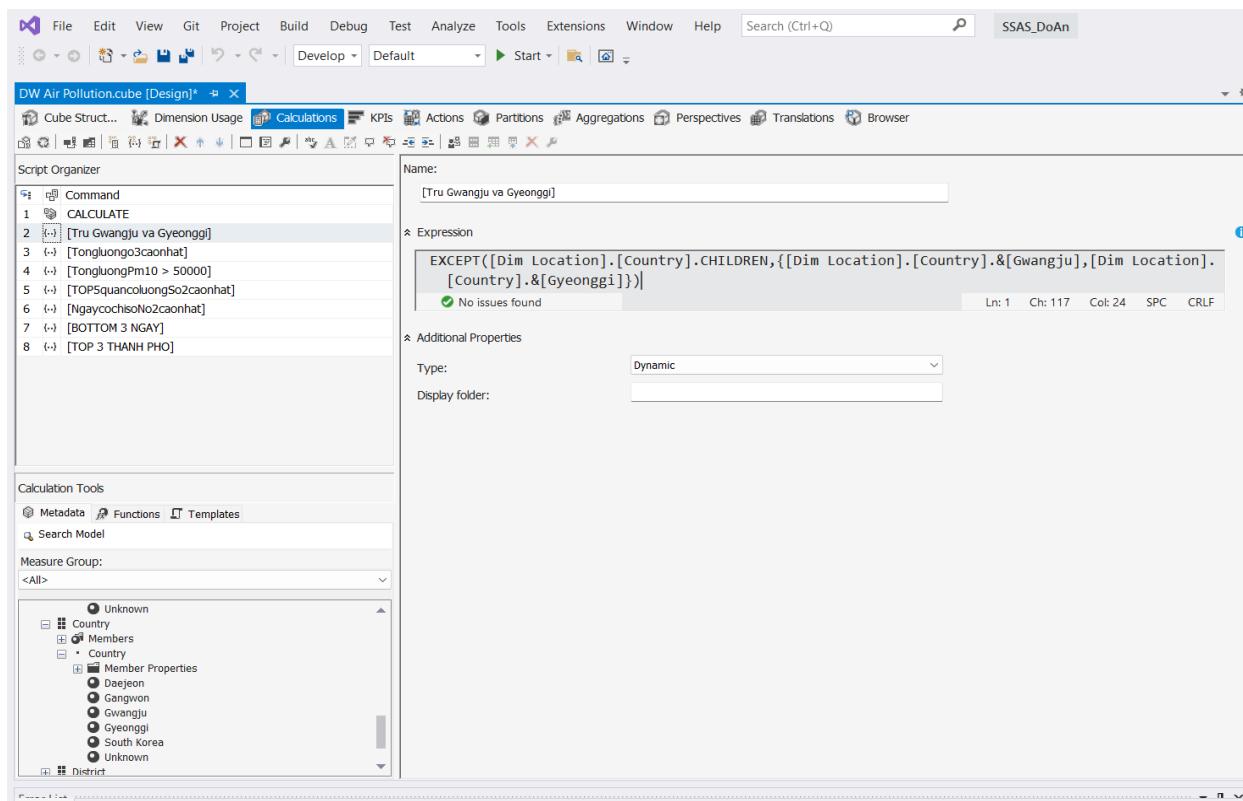
Dim Location City In TOP 3 THANH PHO

<Select dimension>

City	Year	PM10
Geumchon-Dong	2017	13316
Jungang-Way	2017	12422
Seolseong-Myeon	2017	14601

## 2.9. Với từng vùng liệt kê lượng khí So2 theo từng năm, trừ Gwangju và Gyeonggi.

Bước 1 : Tạo Nameset [Tru Gwangju va Gyeonggi]



Bước 2 : Kéo thả thuộc tính Country trong bảng Dim Location , Year trong bảng Dim Date và độ đo S02 trong bảng Fact Pollution.

The screenshot shows the SSAS cube design interface with the cube 'DW Air Pollution.cube [Design]' selected. The left pane displays the cube's structure, including dimensions like 'Dim Date' and 'Dim Location', and measures like 'CO', 'NO2', 'O3', 'PM10', 'PM2 5', and 'SO2'. The right pane shows a table with columns 'Country', 'Year', and 'SO2'. The data includes entries for Daejeon from 2013 to 2022, Gangwon from 2013 to 2022, and Gwangju for 2013, 2014, and 2015.

Dimension	Hierarchy	Operator	Filter Expression	Parameters
<Select dimension>				
Country	Year	SO2		
Daejeon	2013	8		
Daejeon	2014	1464		
Daejeon	2015	1289		
Daejeon	2016	1037		
Daejeon	2017	787		
Daejeon	2018	751		
Daejeon	2019	858		
Daejeon	2020	508		
Daejeon	2021	1004		
Daejeon	2022	187		
Gangwon	2013	15		
Gangwon	2014	1935		
Gangwon	2015	2018		
Gangwon	2016	1436		
Gangwon	2017	1318		
Gangwon	2018	1128		
Gangwon	2019	1056		
Gangwon	2020	792		
Gangwon	2021	782		
Gangwon	2022	145		
Gwangju	2013	11		
Gwangju	2014	2151		
Gwangju	2015	1282		

Bước 3 : Kéo thả Nameset [Tru Gwangju va Gyeonggi] sang trường lọc ta được kết quả như hình:

The screenshot shows the SSAS cube design interface with the same cube selected. The left pane shows the cube structure. The right pane shows a table with columns 'Country', 'Year', and 'SO2'. A filter is applied to the 'Dim Location' dimension, setting the 'Country' hierarchy to 'In' and the filter expression to 'Tru Gwangju va Gyeonggi'. The resulting data includes entries for Daejeon from 2016 to 2022, Gangwon from 2013 to 2022, and South Korea for 2013 to 2018.

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Location	Country	In	Tru Gwangju va Gyeonggi	
<Select dimension>				
Country	Year	SO2		
Daejeon	2016	1037		
Daejeon	2017	787		
Daejeon	2018	751		
Daejeon	2019	858		
Daejeon	2020	508		
Daejeon	2021	1004		
Daejeon	2022	187		
Gangwon	2013	15		
Gangwon	2014	1935		
Gangwon	2015	2018		
Gangwon	2016	1436		
Gangwon	2017	1318		
Gangwon	2018	1128		
Gangwon	2019	1056		
Gangwon	2020	792		
Gangwon	2021	782		
Gangwon	2022	145		
South Korea	2013	30		
South Korea	2014	7806		
South Korea	2015	7984		
South Korea	2016	7734		
South Korea	2017	5920		
South Korea	2018	9907		

## 2.10. Cho biết thành phố có tổng lượng o3 cao nhất và thành phố có tổng lượng o3 thấp nhất.

Bước 1 : Tạo Nameset [TopthanhpovaBottomthanphotongluongO3]

The screenshot shows the SSAS\_DoS interface. In the Script Organizer, a new item named [TopthanhpovaBottomthanphotongluongO3] is being created under the CALCULATE section. The expression is defined as:

```
UNION(TOPCOUNT(NONEMPTY([Dim Location].[City].[City].MEMBERS, [Measures].[.03])), 1, [Measures].[.03]), BOTTOMCOUNT(NONEMPTY([Dim Location].[City].[City].MEMBERS, [Measures].[.03])), 1, [Measures].[.03])
```

No issues were found. The Type is set to Dynamic. The Calculation Tools pane shows DW Air Pollution selected. The Measure Group pane shows <All> selected.

Bước 2 : Kéo thả thuộc tính City trong bảng Dim Location và độ đo o3 và kéo thả nameset [TopthanhpovaBottomthanphotongluongO3] ta được kết quả như hình :

## 2.11. Với mỗi quận đưa ra 3 ngày có tổng lượng CO cao nhất.

Ta dùng hàm Generate và CurrentMember :

Generate(

[Dim Location].[District].Children,

TopCount (

{[Dim Location].[District].CurrentMember \* [Dim Date].[Day Fix].

Children}

,3, [Measures].[CO]

)

)

**Bước 1:** Copy hàm trên vào bộ nhớ tạm. Qua browser kéo thả những thuộc tính cần thiết cho câu truy vấn như hình sau đây:

District	Day Fix	CO
Chuncheon-Si	20131231	9
Chuncheon-Si	20140101	9
Chuncheon-Si	20140102	13
Chuncheon-Si	20140103	15
Chuncheon-Si	20140104	12
Chuncheon-Si	20140105	14
Chuncheon-Si	20140106	16
Chuncheon-Si	20140107	12
Chuncheon-Si	20140108	3
Chuncheon-Si	20140109	7
Chuncheon-Si	20140110	9
Chuncheon-Si	20140111	11
Chuncheon-Si	20140112	6
Chuncheon-Si	20140113	10
Chuncheon-Si	20140114	11
Chuncheon-Si	20140115	13
Chuncheon-Si	20140116	13

**Bước 2 :** Chuyển từ Design Mode sang Script Mode bằng cách nhấp chuột vàobiểu tượng Design Mode như hình minh họa :

**Bước 3 :** Xóa bỏ dòng tô xanh như hình minh họa :

```

SELECT NON EMPTY { [Measures].[CO] } ON COLUMNS, NON EMPTY { {[Dim Location].[District].[District].ALLMEMBERS * [Dim Date].[Day Fix].[Day Fix].ALLMEMBERS } } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [DW Air Pollution] CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
  
```

#### Bước 4 : Dán đoạn code trên vào dòng màu xanh vừa xóa

The screenshot shows the SSAS Management Studio environment. On the left, the cube browser displays the 'DW Air Pollution' cube with its dimensions (Dim Location, Dim Date, Fact Pollution), measures (CO, NO2, O3, PM10, PM2.5, SO2), and KPIs. In the center, the query editor contains the following MDX code:

```

SELECT NON EMPTY { [Measures].[CO] } ON COLUMNS, NON EMPTY { Generate(
    [Dim Location].[District].Children,
    TopCount(
        {[Dim Location].[District].CurrentMember * [Dim Date].[Day Fix].Children},
        3, [Measures].[CO]
    )
) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [DW Air Pollution] CELL PROPERTIES VALUE, BACK_COLOR,
FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
  
```

Below the code, there is a table preview with columns District, Day Fix, and CO. A link 'Click to execute the query.' is present at the bottom.

#### Bước 5 : Nhấn Execute Query tại cửa sổ thực thi truy vấn ta được kết quả như hình minh họa :

The screenshot shows the SSAS Management Studio environment with the 'DW Air Pollution.cube [Design]' tab selected. The cube browser on the left is identical to the previous screenshot. The main area displays the executed MDX query results in a table:

District	Day Fix	CO
Chuncheon-Si	20180101	46
Chuncheon-Si	20171231	41
Chuncheon-Si	20171230	30
Chungbuk	20220109	10
Chungbuk	20210121	9
Chungbuk	20201226	8
Chungnam	20190113	24
Chungnam	20190301	22
Chungnam	20190303	21
Gangwon	20210305	144
Gangwon	20200511	27
Gangwon	20190713	22
Gyeongbuk	20191210	27
Gyeongbuk	20191228	26

#### 2.12. Liệt kê những thành phố có tổng lượng o3 từ 10000 đến 20000.

Bước 1 : Tạo nameset tổng lượng o3 từ 10000-20000

The screenshot shows the SSAS Script Editor interface. In the top navigation bar, the 'Calculations' tab is selected. On the left, the 'Script Organizer' pane lists several calculated members, with the current one being edited highlighted. The main pane contains the following details:

- Name:** [Tong luong o3 10000-20000]
- Expression:**

```
FILTER([Dim Location].[City].CHILDREN,[Measures].[o3]>=10000 AND [Measures].[o3]<=20000)
```

No issues found
- Additional Properties:**
  - Type: Dynamic
  - Display folder:

Bước 2 : Chuyển sang chế độ Design mode -> Browser, kéo thả những thuộc tính cần thiết cho câu truy vấn .

The screenshot shows the SSAS Design mode interface. The top navigation bar has the 'Browser' tab selected. The left sidebar shows the cube structure, including dimensions like 'DW Air Pollution' and measures like 'Fact Pollution'. The main pane displays the results of an MDX query:

City	O3
Bangsan-Myeon	102743
Cheongnim-D...	24523
Geumchon-Do...	100431
Gwanak-Gu...	87195
Hwasun-Eup	32551
Hyeksan-Dong	20340
Jeongnim-Dong	95087
Jungang-Way	96594
Juwol-Dong	89978
Nowon-Gu	90824
Onul-Dong	15552
Pado-Ri	112312
Paju	59219
Saemangeum	532
Saesol-Dong	20473
SangJus	60163
Seokeong-My...	106030

Bước 3 : Kéo nameset [Tong luong o3 10000-20000] sang trường lọc ta được kết quả như hình bên dưới :

## 2.13. Đưa ra 7 ngày có lượng Pm10 cao nhất tại quận Jeonbuk .

Bước 1 : Xây dựng Name set [TOP 7 NGAY]

```

Name: [TOP 7 NGAY]

Expression:
{TopCount(
    NonEmpty(
        [Dim Date].[Day Fix].Children
        ,[Measures].[Pm10]
    )
    ,7
    ,[Measures].[Pm10]
)}
No issues found
Ln: 8 Ch: 11 SPC CRLF

```

Bước 2 : Deploy Nameset sau đó qua Browser kéo thả những thuộc tính cần thiết cho câu truy vấn trên ta được kết quả như hình :

DW Air Pollution.cube [Design] ...

Cube Struct... Dimension Usage Calculations KPIs Actions Partitions Aggregations Perspectives Translations Browser

Language: Default

Edit as Text Import... MDX

Dimension Hierarchy Operator Filter Expression Parameters

Dim Location	District	Equal	{ Jeonbuk }	<input type="checkbox"/> <input type="checkbox"/>
Dim Date	Day Fix	In	TOP 7 NGAY	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>

Day Fix PM10

20210328	456
20210329	277
20210506	198
20210507	344
20220211	133
20220212	127
20220213	118

Calculated Members

Measure Group: <All>

Dim Date

- [.] BOTTOM 3 THANG
- [.] NgaycochitohNo2caonhat
- [.] TOP 7 NGAY

Day Fix ID Date Month Fix Year Hierarchy Dim Location

## 2.14. Liệt kê tổng lượng So2 của tất cả các vùng theo từng năm.

Bước 1 : Kéo thả thuộc tính Year trong bảng chiều Dim Date , kéo thả thuộc tính Country trong bảng chiều Dim Location và kéo độ đo So2 sang cửa sổ thực thi truy vấn.

DW Air Pollution.cube [Design] ...

Cube Struct... Dimension Usage Calculations KPIs Actions Partitions Aggregations Perspectives Translations Browser

Language: Default

Edit as Text Import... MDX

Dimension Hierarchy Operator Filter Expression Parameters

<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>
--------------------	--	--	--	---

Year Country SO2

2013	Daejeon	8
2013	Gangwon	15
2013	Gwangju	11
2013	Gyeonggi	14
2013	South Korea	30
2014	Daejeon	1464
2014	Gangwon	1935
2014	Gwangju	2151
2014	Gyeonggi	2421
2014	South Korea	7806
2015	Daejeon	1289
2015	Gangwon	2018
2015	Gwangju	1282
2015	Gyeonggi	2178
2015	South Korea	7984
2016	Daejeon	1037
2016	Gangwon	1436

Calculated Members

Measure Group: <All>

Dim Date

- [.] BOTTOM 3 THANG
- [.] NgaycochitohNo2caonhat
- [.] TOP 7 NGAY

Day Fix ID Date Month Fix

Bước 2 : Ta được kết quả câu truy vấn .

Year	Country	SO2
2013	Daejeon	8
2013	Gangwon	15
2013	Gwangju	11
2013	Gyeonggi	14
2013	South Korea	30
2014	Daejeon	1464
2014	Gangwon	1935
2014	Gwangju	2151
2014	Gyeonggi	2421
2014	South Korea	7806
2015	Daejeon	1289
2015	Gangwon	2018
2015	Gwangju	1282
2015	Gyeonggi	2178
2015	South Korea	7984
2016	Daejeon	1037
2016	Gangwon	1436
2016	Gwangju	1254
2016	Gyeonggi	1902
2016	South Korea	7734
2017	Daejeon	787
2017	Gangwon	1318
2017	Gwangju	1111

## 2.15. Liệt kê những năm có tổng lượng Pm25 từ 20000-200000.

Bước 1 : Tạo nameset tổng lượng Pm25 từ 20000-200000

The screenshot shows the Script Organizer interface. On the left, there is a list of existing scripts and calculations. On the right, a new calculation is being defined:

- Name:** [Tong luong Pm25 từ 20000-200000]
- Expression:**

```
Filter([Dim Date].[Year].Children, [Measures].[Pm2_5] >=20000 AND [Measures].[Pm2_5] <=200000)
```

No issues found
- Additional Properties:**
  - Type: Dynamic
  - Display folder: (empty)

Bước 2 : Chuyển sang chế độ Design mode -> Browser, kéo thả những thuộc tính cần thiết cho câu truy vấn .

The screenshot shows the SSAS Dimension browser interface. At the top, there are tabs for Dimension, Hierarchy, Operator, Filter Expression, and Parameters. Below this is a table with two columns: Year and PM2.5. The table data is as follows:

Year	PM2.5
2013	0
2014	16880
2015	101214
2016	151735
2017	145761
2018	248045
2019	325609
2020	360923
2021	407292
2022	80387

Bước 3 : Kéo nameset [Tong luong Pm25 20000-200000] sang trường lọc ta được kết quả như hình bên dưới :

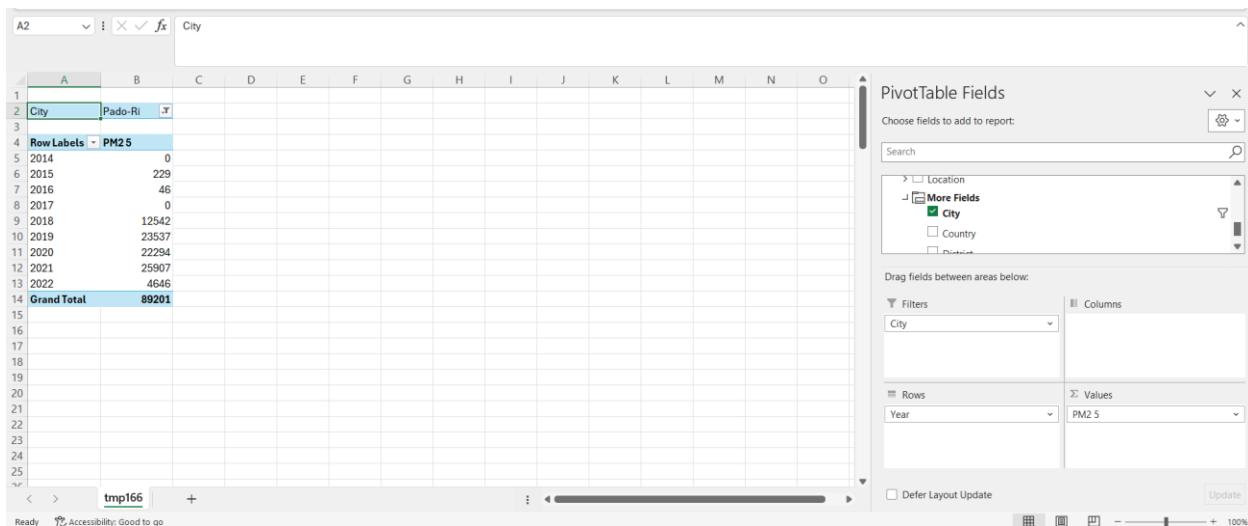
The screenshot shows the SSAS Dimension browser interface with a filter applied. The filter settings are: Dimension: Dim Date, Hierarchy: Year, Operator: In, Filter Expression: Tong luong Pm25 20000-200000. The resulting table data is as follows:

Year	PM2.5
2015	101214
2016	151735
2017	145761
2022	80387

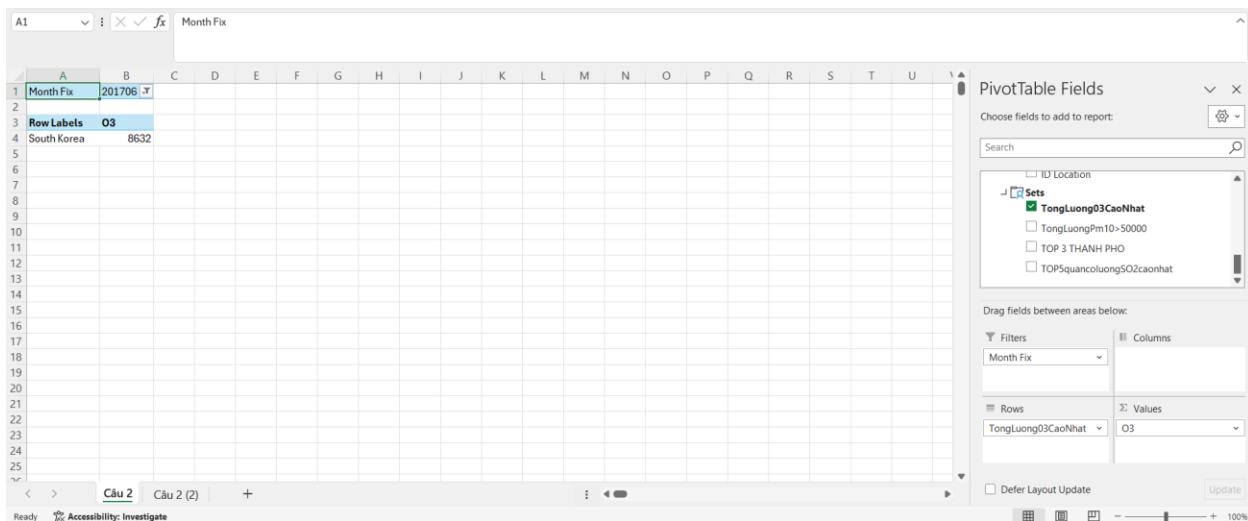
### 3. Quá trình phân tích dữ liệu bằng công cụ PIVOT EXCEL

#### 3.1. Cho biết tổng lượng PM25 ở thành phố Pado-Ri theo từng năm.

Ta thực hiện kéo thả các thuộc tính tương ứng vào "City" sẽ vào "Filter" , "Year" sẽ vào "Rows" và "PM25" là Values.



### 3.2. Cho biết vùng có chỉ số 03 cao nhất trong tháng 6 năm 2017.



### 3.3. Cho biết thành phố nào có tổng lượng Pm10 trên 50000.

### 3.4. Cho biết top 5 quận có chỉ số SO<sub>2</sub> cao nhất trong từng năm.

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable Fields pane open on the right side. The data table on the left contains the following information:

	District	SO2
1	Chuncheon-Si	10625
2	Chungnam	12708
3	Gyeongbuk	12368
4	Paju-Si	13216
5	Seoul	34761
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		

The PivotTable Fields pane lists the following fields:

- ID Location
- Sets
  - TongLuong03CaoNhat
  - TongLuongpm10>50000
  - TOP 3 THANH PHO
  - TOP5quancoluongSO2caonhat** (selected)

Below the pane, there are sections for Filters, Rows, and Values, with the value "SO2" selected for the Values section. The status bar at the bottom indicates "Ready" and "Accessibility: Good to go".

### 3.5. Cho biết ngày nào có tổng lượng NO<sub>2</sub> cao nhất .

A	B	C	D	E	F
1	Day	NO2			
2	20210120	578			
3					
4					
5					
6					
7					
8					
9					

### 3.6. Cho biết tổng lượng CO ở vùng Deajeon theo từng năm.

Screenshot of a Microsoft Excel PivotTable setup. The PivotTable Fields pane shows 'Fact Pollution' with 'CO' selected. The Rows section is set to 'Year' and the Values section is set to 'CO'. The main table shows data from 2013 to 2022, with a Grand Total of 13850 for Daejeon.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Country	Daejeon										
2												
3	Year	CO										
4	2013	7										
5	2014	2012										
6	2015	1950										
7	2016	1775										
8	2017	1283										
9	2018	1574										
10	2019	1865										
11	2020	1346										
12	2021	1718										
13	2022	320										
14	Grand Total	13850										
15												
16												

### 3.7. Đưa ra 3 ngày có lượng khí Pm25 ít nhất tại Paju .

Screenshot of a Microsoft Excel PivotTable setup. The PivotTable Fields pane shows 'Fact Pollution' with 'PM2.5' selected. The Rows section is set to 'BOTTOM 3 NGAY' and the Values section is set to 'PM2.5'. The main table shows data for Paju on three specific days in 2021.

	A	B	C	D	E	F	G
1	City	Paju					
2							
3	Day	PM2.5					
4	20210823	0					
5	20211210	0					
6	20211221	0					
7							
8							
9							
10							
11							
12							

### 3.8. Cho biết 3 thành phố có lượng khí Pm10 nhiều nhất trong năm 2017.

Screenshot of a Microsoft Excel PivotTable setup. The PivotTable Fields pane shows 'Fact Pollution' with 'PM10' selected. The Rows section is set to 'TOP 3 THANH PHO' and the Values section is set to 'PM10'. The main table shows data for three cities in 2017.

	A	B	C	D	E	F	G
1	Year	2017					
2							
3	City	PM10					
4	Geumchon-Dong	13316					
5	Jungang-Way	12422					
6	Seolseong-Myeon	14601					
7							
8							
9							
10							
11							
12							
13							

### 3.9. Với từng vùng liệt kê lượng khí So2 theo từng năm, trừ Gwangju và Gyeonggi.

Country	SO2
Deseon	8
2013	1464
2014	1297
2015	1037
2016	787
2017	751
2018	858
2019	568
2020	1004
2021	187
Gangwon	15
2013	1926
2014	2018
2015	1436
2016	1318
2017	1339
2018	1056
2019	792
2020	782
2021	145
South Korea	30
2013	7806
2014	7984
2015	7734
2016	6900
2017	9907
2018	10960
2019	13735
2020	14400
2021	2109
2022	

### 3.10. Cho biết thành phố có tổng lượng o3 cao nhất và thành phố có tổng lượng o3 thấp nhất

City	O3
Pado-Ri	112312
Saemangeum	532

### 3.11. Với mỗi quận đưa ra 3 ngày có tổng lượng CO cao nhất.

### 3.12. Liệt kê những thành phố có tổng lượng o3 từ 10000 đến 20000.

A1    City

	A	B	C	D	E	F
1	City	03				
2	Onui-Dong	15552				
3						
4						
5						
6						
7						
8						
9						

Câu 12 Câu 2 Câu 3 Câu 4 Câu 5 Câu 6 Câu 7 Câu t ... + : ← →

Ready Accessibility: Good to go

### 3.13. Đưa ra 7 ngày có lượng Pm10 cao nhất tại quận Jeonbuk .

A3    City

	A	B	C	D	E	F	G	H	I	J
1	District	Jeonbuk								
2	City	PM10								
3	20210328	456								
4	20210329	277								
5	20210506	198								
6	20210507	344								
7	20220211	133								
8	20220212	127								
9	20220213	118								
10										
11										
12										
13										
14										
15										
16										

Câu 7 Câu 8 Câu 9 Câu 10 Câu 11 Câu 12 Câu 13 Câu 14 + : ← →

### 3.14. Liệt kê tổng lượng SO2 của tất cả các vùng theo từng năm.

A screenshot of an Excel PivotTable titled 'SO2'. The table has columns for 'City' (Daejeon, Gangwon, Gwangju, Gyeonggi, South Korea), 'Year' (2013-2022), and 'SO2' concentration. The 'Grand Total' row shows values like 7893 for 2013 and 122688 for 2022. The PivotTable Fields pane on the right shows 'Country' in the Columns area and 'Year' in the Rows area.

	A	B	C	D	E	F	G	H	I	J	K
1	SO2	Column Labels									
2	City	Daejeon	Gangwon	Gwangju	Gyeonggi	South Korea	Grand Total				
3	2013		8	15	11	14	30	78			
4	2014		1464	1935	2151	2421	7806	15777			
5	2015		1289	2018	1282	2178	7984	14751			
6	2016		1037	1436	1254	1902	7734	13363			
7	2017		787	1318	1111	1555	5920	10691			
8	2018		751	1128	1137	1507	9907	14430			
9	2019		858	1056	1203	1391	10960	15468			
10	2020		508	792	1031	1104	13735	17170			
11	2021		1004	782	1059	1000	14400	18245			
12	2022		187	145	136	144	2103	2715			
13	Grand Total		7893	10625	10375	13216	80579	122688			
14											
15											
16											
17											

### 3.15. Liệt kê những năm có tổng lượng Pm25 từ 20000-200000.

A screenshot of an Excel PivotTable titled 'City'. The table has columns for 'City' (Daejeon) and 'PM2 5'. The PivotTable Fields pane on the right shows a set named 'Tong luong Pm25 tu 20000-200000' selected in the Sets section. The rows show years 2015, 2016, 2017, and 2022.

	A	B	C	D	E	F	G	H	I	J	K
1	City	PM2 5									
2	2015	101214									
3	2016	151735									
4	2017	145761									
5	2022	80387									

## 4. Quá trình phân tích dữ liệu bằng ngôn ngữ MDX

### 4.1. Cho biết tổng lượng PM25 ở thành phố Pado-Ri theo từng năm.

- Câu lệnh truy vấn bằng ngôn ngữ MDX.

```

SELECT [Measures].[Pm10] ON COLUMNS,
NON EMPTY {[Dim Location].[City].[City].MEMBERS}
HAVING [Measures].[Pm10] > 50000 ON ROWS
FROM [DW Air Pollution]

```

- Kết quả câu truy vấn :

The screenshot shows the MDXQuery1.mdx - L\_...ng Van Tien Nhat window. On the left, there's a navigation pane with a tree view of the cube structure under 'Cube: DW Air Pollution'. The tree includes 'DW Air Pollution', 'Measures', 'KPIs', 'Dim Date' (with 'BOTTOM 3 NGAY' and 'NgaycochisoNo2caenhat' children), 'Day Fix', 'ID Date', 'Month Fix', 'Year' (with 'All', '2013', '2014', '2015', '2016', '2017', '2018', '2019', '2020', '2021', '2022', and 'Unknown' children), and 'Member Properties'. On the right, the main area contains an MDX query and its results.

```
-- Câu 1. Cho biết tổng lượng PM2.5 ở thành phố Pado-Ri theo từng năm.
SELECT
NON EMPTY [Measures].[PM2.5] ON COLUMNS,
[Dim Date].[Year].MEMBERS * [Dim Location].[City].&[Pado-Ri] ON ROWS
FROM [DW Air Pollution]
```

The results table shows the total PM2.5 concentration for each year from 2013 to 2022 at the Pado-Ri location. The data is as follows:

Year	Location	PM2.5
2013	Pado-Ri	(null)
2014	Pado-Ri	0
2015	Pado-Ri	229
2016	Pado-Ri	46
2017	Pado-Ri	0
2018	Pado-Ri	12542
2019	Pado-Ri	23537
2020	Pado-Ri	22294
2021	Pado-Ri	25907
2022	Pado-Ri	4646
Unknown	Pado-Ri	(null)

#### 4.2. Cho biết vùng có chỉ số 03 cao nhất trong tháng 6 năm 2017.

- Câu lệnh truy vấn bằng ngôn ngữ MDX

SELECT NON EMPTY {[Measures].[O3]} ON COLUMNS,

NON EMPTY {TOPCOUNT([Dim Location].[Country].[Country], 1, [Measures].[O3]) \* [Dim Date].[Month Fix].&[201706]} ON ROWS

FROM [DW Air Pollution]

- Kết quả câu truy vấn :

The screenshot shows a SQL query editor window in SSMS. The code is written in T-SQL and includes four comments (Câu 1 to Câu 4) and several SELECT statements. The results pane shows a single row: South Korea, 201706, 8632.

```

--Câu 1. Cho biết tổng lượng Pm2.5 ở thành phố Pado-Ri theo từng năm.
SELECT
    NON EMPTY [Measures].[Pm2.5] ON COLUMNS,
    [Dim Date].[Year].>[Year] * [Dim Location].[City].>[Pado-Ri] ON ROWS
FROM [DW Air Pollution]

--Câu 2. Cho biết vùng có chỉ số O3 cao nhất vào tháng 6 năm 2017.
SELECT
    NON EMPTY {[Measures].[O3]} ON COLUMNS,
    NON EMPTY {TOPCOUNT([Dim Location].[Country],1,[Measures].[O3])} * [Dim Date].[Month Fix]>[201706] ON ROWS
FROM [DW Air Pollution]

--Câu 3. Cho biết thành phố nào có tổng lượng Pm10 trên 50000.
SELECT [Measures].[Pm10] ON COLUMNS,
    NON EMPTY {[Dim Location].[City]} .MEMBERS
HAVING [Measures].[Pm10] > 50000 ON ROWS
FROM [DW Air Pollution]

--Câu 4.

```

	Messages	Results
	O3	
South Korea	201706	8632

### 4.3. Cho biết thành phố nào có tổng lượng Pm10 trên 50000.

- Câu lệnh truy vấn :

```

SELECT [Measures].[Pm10] ON COLUMNS,
NON EMPTY {[Dim Location].[City]} .MEMBERS
HAVING [Measures].[Pm10] > 50000 ON ROWS
FROM [DW Air Pollution]

```

- Kết quả truy vấn :

```
--Câu 3. Cho biết thành phố nào có tổng lượng Pm10 trên 50000.
SELECT [Measures].[Pm10] ON COLUMNS,
NON EMPTY {[Dim Location].[City].[City].MEMBERS}
HAVING [Measures].[Pm10] > 50000 ON ROWS
FROM [DW Air Pollution]
--Câu 4.
```

100 %

	PM10
Bangsan-Myeon	76339
Geumchon-Dong	117789
Gwanak-Gu	107644
Jeongnim-Dong	91570
Jungang-Way	99428
Juwol-Dong	97010
Nowon-Gu	99531
Pado-Ri	102347
Paju	58016
Seolseong-Myeon	119361
Seonggeoep	51858

#### 4.4. Cho biết top 5 quận có chỉ số So2 cao nhất trong từng năm.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

```
SELECT
NON EMPTY {[Measures].[So2]} ON COLUMNS,
{topcount([Dim Location].[District].[District], 5, [Measures].[So2])} ON ROWS
FROM [DW Air Pollution]
```

- Kết quả câu truy vấn :

```
--Câu 4. Cho biết top 5 quận có chỉ số SO2 cao nhất trong tất cả các năm.
SELECT
NON EMPTY {[Measures].[SO2]} ON COLUMNS,
{topcount([Dim Location].[District].[District], 5, [Measures].[SO2])} ON ROWS
FROM [DW Air Pollution]
```

100 %

	Messages	Results
		SO2
Seoul		34761
Paju-Si		13216
Chungnam		12708
Gyeongbuk		12368
Chuncheon-Si		10625

#### 4.5. Cho biết ngày nào có tổng lượng NO2 cao nhất .

- Câu lệnh truy vấn bằng ngôn ngữ MDX

```
SELECT [Measures].[NO2] ON COLUMNS,
topcount([Dim Date].[Day Fix].[Day Fix],1,[Measures].[NO2]) ON ROWS
FROM [DW Air Pollution]
```

- Kết quả câu truy vấn :

```
--Câu 5. Cho biết ngày nào có tổng lượng NO2 cao nhất.
SELECT [Measures].[NO2] ON COLUMNS,
topcount([Dim Date].[Day Fix].[Day Fix],1,[Measures].[NO2]) ON ROWS
FROM [DW Air Pollution]
```

100 %

	Messages	Results
		NO2
20210120		578

#### 4.6. Cho biết tổng lượng CO ở vùng Deajeon theo từng năm.

- Câu lệnh truy vấn bằng ngôn ngữ MDX

```

SELECT NON EMPTY [Measures].[CO] ON COLUMNS,
NON EMPTY{[Dim Location].[Country].&[Daejeon] * [Dim Date].[Year].[Year] } ON
ROWS
FROM [DW Air Pollution]

```

- Kết quả câu truy vấn :

--Câu 6. Cho biết tổng lượng CO ở vùng Deajeon theo từng năm.

```

SELECT
    NON EMPTY [Measures].[CO] ON COLUMNS,
    NON EMPTY{[Dim Location].[Country].&[Daejeon] * [Dim Date].[Year].[Year] }ON ROWS
    FROM [DW Air Pollution]

```

		CO
Daejeon	2013	7
Daejeon	2014	2012
Daejeon	2015	1950
Daejeon	2016	1775
Daejeon	2017	1283
Daejeon	2018	1574
Daejeon	2019	1865
Daejeon	2020	1346
Daejeon	2021	1718
Daejeon	2022	320

#### 4.7. Đưa ra 3 ngày có lượng khí Pm25 ít nhất tại Paju .

- Câu lệnh truy vấn bằng ngôn ngữ MDX

```

SELECT
    [Measures].[PM2 5] ON COLUMNS,

```

```

[Dim Location].[City].&[Paju] * BottomCount(
NonEmpty([Dim Date].[Month Fix].[Month Fix],[Measures].[PM2 5]),3,[Measures].[PM2
5])

ON ROWS

FROM [DW Air Pollution]

WHERE [Dim Date].[Year].&[2018]

```

- Kết quả câu truy vấn :

The screenshot shows the SSMS interface with two tabs: 'Messages' and 'Results'. The 'Messages' tab contains the MDX code for both Câu 7 and Câu 8. The 'Results' tab displays the output of the first query, which is a table showing the top three months in 2018 for Paju with PM2.5 values.

```

--Câu 7. Đưa ra 3 tháng có tổng lượng Pm25 ít nhất trong năm 2018 tại thành phố Paju.
SELECT
    [Measures].[PM2 5] ON COLUMNS,
    [Dim Location].[City].&[Paju] * BottomCount(
        NonEmpty([Dim Date].[Month Fix].[Month Fix],[Measures].[PM2 5]),3,[Measures].[PM2
5])
    ON ROWS
FROM [DW Air Pollution]
WHERE [Dim Date].[Year].&[2018]
--Câu 8. Cho biết 3 thành phố có lượng khí Pm10 nhiều nhất trong năm 2017.
SELECT
    NON EMPTY {[Measures].[Pm10]} ON COLUMNS,
    NON EMPTY {TOPCOUNT([Dim Location].[City].CHILDREN,3,[Measures].[Pm10])} ON ROWS
FROM [DW Air Pollution]
WHERE ([Dim Date].[Year].&[2017])

```

		PM2 5
Paju	201809	1225
Paju	201808	1257
Paju	201807	1245

#### 4.8. Cho biết 3 thành phố có lượng khí Pm10 nhiều nhất trong năm 2017.

- Câu lệnh truy vấn bằng ngôn ngữ MDX

```

SELECT NON EMPTY {[Measures].[Pm10]} ON COLUMNS,
NON EMPTY {TOPCOUNT([Dim Location].[City].CHILDREN,3,[Measures].[Pm10])} ON ROWS
FROM [DW Air Pollution]

```

WHERE ([Dim Date].[Year].&[2017])

- Kết quả câu truy vấn :

```
--Câu 8. Cho biết 3 thành phố có lượng khí Pm10 nhiều nhất trong năm 2017.  
SELECT  
NON EMPTY {[Measures].[Pm10]} ON COLUMNS,  
NON EMPTY {TOPCOUNT([Dim Location].[City].CHILDREN,3,[Measures].[Pm10])} ON ROWS  
FROM [DW Air Pollution]  
WHERE ([Dim Date].[Year].&[2017])
```

City	Pm10
Seolseong-Myeon	14601
Geumchon-Dong	13316
Jungang-Way	12422

#### 4.9. Với từng vùng liệt kê lượng khí SO2 theo từng năm, trừ Gwangju và Gyeonggi.

- Câu lệnh truy vấn bằng ngôn ngữ MDX

SELECT [Measures].[SO2] ON COLUMNS,

NON EMPTY {EXCEPT([Dim Location].[Country].CHILDREN,{[Dim Location].[Country].&[Gwangju],[Dim Location].[Country].&[Gyeonggi]}) \* [Dim Date].[Year].[Year]} ON ROWS

FROM [DW Air Pollution]

- Kết quả câu truy vấn :

```

--Câu 9.Với từng vùng liệt kê lượng khí SO2 theo từng năm, trừ Gwangju và Gyeonggi.
SELECT [Measures].[SO2] ON COLUMNS,
NON EMPTY {EXCEPT([Dim Location].[Country].CHILDREN, {[Dim Location].[Country].&[Gwangju], [Dim Location].[Country].&[Gyeonggi]})} ON ROWS
FROM [DW Air Pollution]

```

100 %

		SO2
Daejeon	2013	8
Daejeon	2014	1464
Daejeon	2015	1289
Daejeon	2016	1037
Daejeon	2017	787
Daejeon	2018	751
Daejeon	2019	858
Daejeon	2020	508
Daejeon	2021	1004
Daejeon	2022	187
Gangwon	2013	15
Gangwon	2014	1935
Gangwon	2015	2018
Gangwon	2016	1436
Gangwon	2017	1318
Gangwon	2018	1128
	...	...

LAPTOP-V2PG3KJO LAPTOP-V2PG3KJO\Hoang ... DW AirPollution 00:00:01

#### 4.10. Cho biết thành phố có tổng lượng o3 cao nhất và thành phố có tổng lượng o3 thấp nhất.

- Câu lệnh truy vấn bằng ngôn ngữ MDX :

WITH

SET [TopCity] AS

TOPCOUNT(NONEMPTY([Dim Location].[City].[City].MEMBERS, [Measures].[O3]), 1, [Measures].[O3])

SET [BottomCity] AS

BOTTOMCOUNT(NONEMPTY([Dim Location].[City].[City].MEMBERS, [Measures].[O3]), 1, [Measures].[O3])

SELECT

{[Measures].[O3]} ON COLUMNS,

UNION([TopCity], [BottomCity]) ON ROWS

FROM [DW Air Pollution]

- Kết quả câu truy vấn :

```
--Câu 10.Cho biết thành phố có tổng lượng o3 cao nhất và thành phố có tổng lượng o3 thấp nhất.
WITH
    SET [TopCity] AS
        TOPCOUNT(NONEMPTY([Dim Location].[City].[City].MEMBERS, [Measures].[O3]), 1, [Measures].[O3])
    SET [BottomCity] AS
        BOTTOMCOUNT(NONEMPTY([Dim Location].[City].[City].MEMBERS, [Measures].[O3]), 1, [Measures].[O3])
SELECT
    {[Measures].[O3]} ON COLUMNS,
    UNION([TopCity], [BottomCity]) ON ROWS
FROM [DW Air Pollution]
```

100 % ▶

Messages	Results	
Pado-Ri	O3 112312	Saemangeum 532

#### 4.11. Với mỗi quận đưa ra 3 ngày có tổng lượng CO cao nhất.

- Câu lệnh truy vấn bằng ngôn ngữ MDX

SELECT

NON EMPTY [Measures].[CO] ON COLUMNS,

NON EMPTY

GENERATE(

[Dim Location].[District].Children,

TOPCOUNT(

[Dim Location].[District].CurrentMember \* [Dim Date].[Day Fix].Children,

3,

[Measures].[CO]

```

)
) ON ROWS

FROM [DW Air Pollution]

```

- Kết quả câu truy vấn :

--Câu 11. Với mỗi quận đưa ra 3 ngày có tổng lượng CO cao nhất.

```

SELECT
NON EMPTY [Measures].[CO] ON COLUMNS,
NON EMPTY
GENERATE(
    [Dim Location].[District].Children,
    TOPCOUNT(
        [Dim Location].[District].CurrentMember * [Dim Date].[Day Fix].Children,
        3,
        [Measures].[CO]
    )
) ON ROWS
FROM [DW Air Pollution]

```

		CO
Chuncheon-Si	20180101	46
Chuncheon-Si	20171231	41
Chuncheon-Si	20171230	30
Chungbuk	20220109	10
Chungbuk	20210121	9
Chungbuk	20201226	8
Chungnam	20190113	24
Chungnam	20190301	22
Chungnam	20190303	21
Gangwon	20210305	144
Gangwon	20200511	27
Gangwon	20190713	22
Gyeongbuk	20191210	27
Gyeongbuk	20191228	26
Gyeongbuk	20200104	24
Gyeonggi	20190113	26

#### 4.12. Liệt kê những thành phố có tổng lượng o3 từ 10000 đến 20000.

- Câu lệnh truy vấn bằng ngôn ngữ MDX :

```

SELECT [Measures].[o3] ON COLUMNS,
NON EMPTY
FILTER(
    {[Dim Location].[City].[City].MEMBERS},
    [Measures].[o3] >= 10000 AND [Measures].[o3] <= 20000
) ON ROWS

```

FROM [DW Air Pollution]

- Kết quả câu truy vấn :

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. In the top pane, there is an MDX query:

```
--Câu 12. Liệt kê những thành phố có tổng lượng o3 từ 10000 đến 20000.
SELECT [Measures].[o3] ON COLUMNS,
NON EMPTY
FILTER(
    {[Dim Location].[City].[City].MEMBERS},
    [Measures].[o3] >= 10000 AND [Measures].[o3] <= 20000
) ON ROWS
FROM [DW Air Pollution]
```

In the bottom pane, under the "Results" tab, the output is displayed in a grid:

	O3
Onui-Dong	15552

#### 4.13. Đưa ra 7 ngày có lượng Pm10 cao nhất tại quận Jeonbuk .

- Câu lệnh truy vấn bằng ngôn ngữ MDX :

SELECT

[Measures].[Pm10] ON COLUMNS,

TopCount(

NonEmpty(

[Dim Date].[Day Fix].Children,

([Dim Location].[District].&[Jeonbuk], [Measures].[Pm10])

),

7,

```

[Measures].[Pm10]

) ON ROWS

FROM [DW Air Pollution]

WHERE [Dim Location].[District].&[Jeonbuk]

```

- Kết quả câu truy vấn :

```

--Câu 13. Đưa ra 7 ngày có lượng Pm10 cao nhất tại quận Jeonbuk.

SELECT
    [Measures].[Pm10] ON COLUMNS,
    TopCount(
        NonEmpty(
            [Dim Date].[Day Fix].Children,
            ([Dim Location].[District].&[Jeonbuk], [Measures].[Pm10])
        ),
        7,
        [Measures].[Pm10]
    ) ON ROWS
FROM [DW Air Pollution]
WHERE [Dim Location].[District].&[Jeonbuk]

```

	PM10
20210328	456
20210507	344
20210329	277
20210506	198
20220211	133
20220212	127
20220213	118

#### 4.14. Liệt kê tổng lượng So2 của tất cả các vùng theo từng năm.

- Câu lệnh truy vấn bằng ngôn ngữ MDX :

```

SELECT [Measures].[So2] ON COLUMNS,

```

NON EMPTY

[Dim Location].[Country].Children \* [Dim Date].[Year].[Year].Members ON ROWS  
FROM [DW Air Pollution]

- Kết quả câu truy vấn :

```
--Câu 14. Liệt kê tổng lượng SO2 của tất cả các vùng theo từng năm
SELECT [Measures].[SO2] ON COLUMNS,
NON EMPTY
    [Dim Location].[Country].Children * [Dim Date].[Year].[Year].Members ON ROWS
FROM [DW Air Pollution]
```

Region	Year	SO2
Daejeon	2013	8
Daejeon	2014	1464
Daejeon	2015	1289
Daejeon	2016	1037
Daejeon	2017	787
Daejeon	2018	751
Daejeon	2019	858
Daejeon	2020	508
Daejeon	2021	1004
Daejeon	2022	187
Gangwon	2013	15
Gangwon	2014	1935
Gangwon	2015	2018
Gangwon	2016	1436
Gangwon	2017	1318
Gangwon	2018	1128

#### 4.15. Liệt kê những năm có tổng lượng Pm2.5 từ 20000-200000.

- Câu lệnh truy vấn bằng ngôn ngữ MDX :

SELECT [Measures].[Pm2.5] ON COLUMNS,  
NON EMPTY  
FILTER(

```

{[Dim Date].[Year].[Year].MEMBERS},  

[Measures].[Pm2 5] >= 20000 AND [Measures].[Pm2 5] <= 200000  

) ON ROWS  

FROM [DW Air Pollution]

```

- Kết quả câu truy vấn :

```
--Câu 15. Liệt kê những năm có tổng lượng Pm25 từ 20000-100000.  

SELECT [Measures].[Pm2 5] ON COLUMNS,  

NON EMPTY [ ]  

FILTER(  

{[Dim Date].[Year].[Year].MEMBERS},  

[Measures].[Pm2 5] >= 20000 AND [Measures].[Pm2 5] <= 200000  

) ON ROWS  

FROM [DW Air Pollution]
```

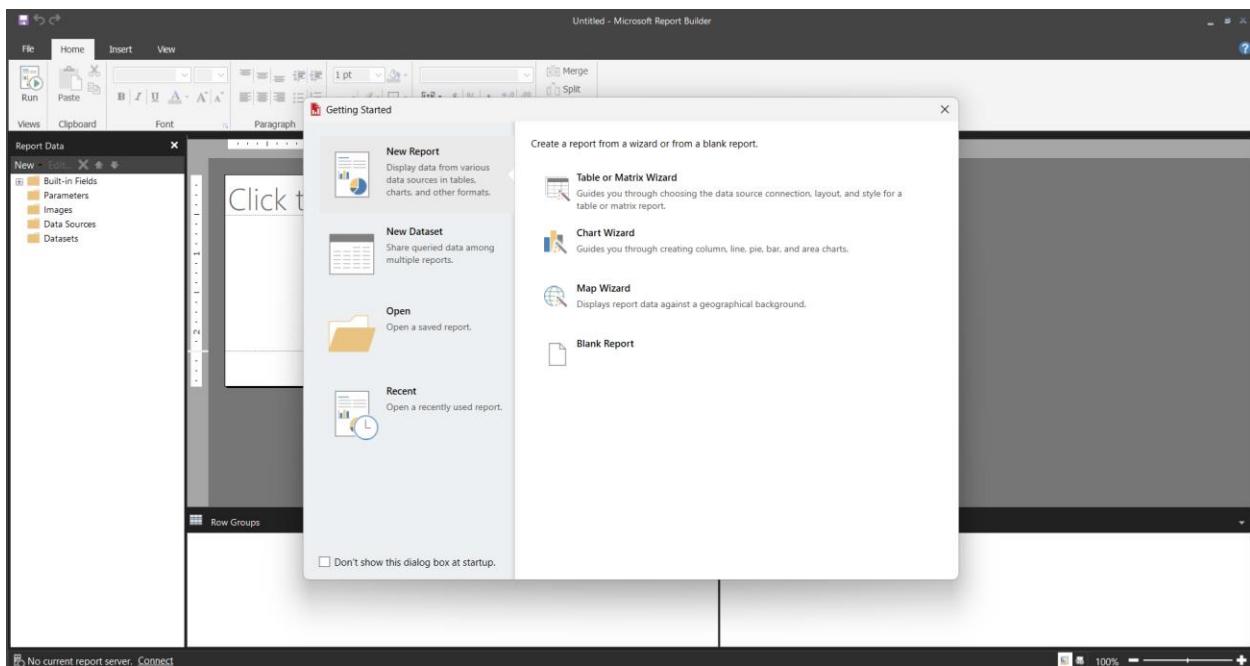
	PM2 5
2015	101214
2016	151735
2017	145761
2022	80387

#### IV. Quá trình lập báo biểu (SSRS)

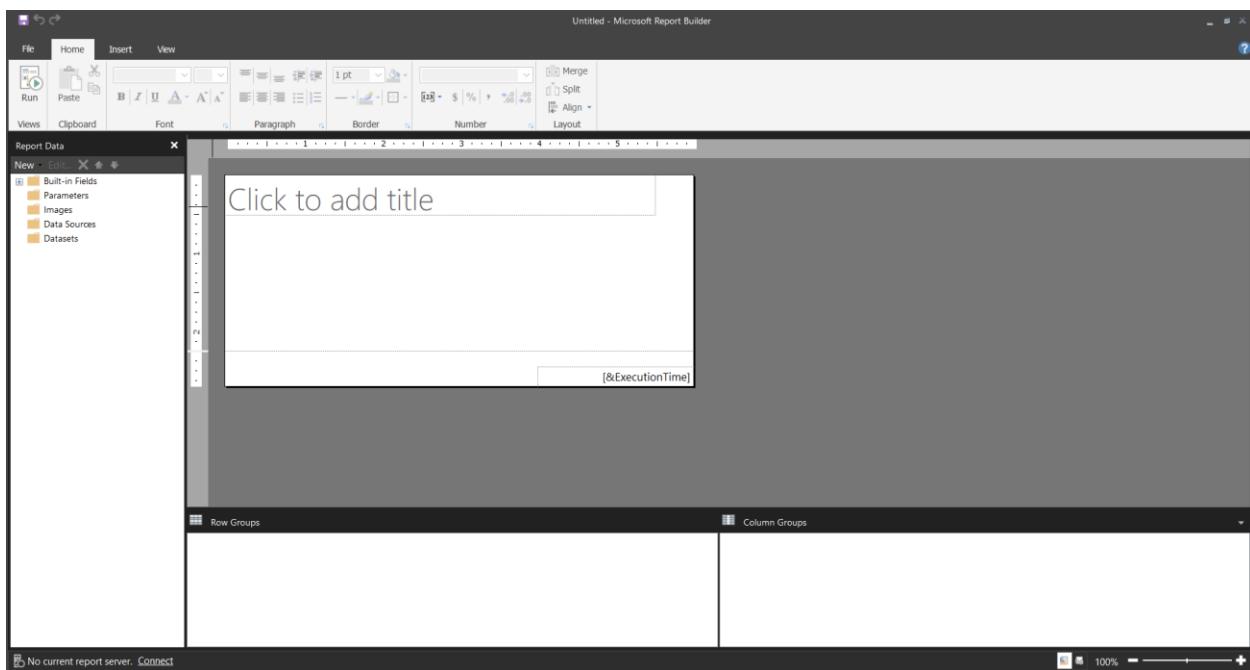
##### 1. Quá trình lập báo biểu bằng công cụ Report builder

###### 1.1. Tạo project SSRS

Ta bật công cụ Report Builder , và chọn Blank Report để tạo mới 1 Report.

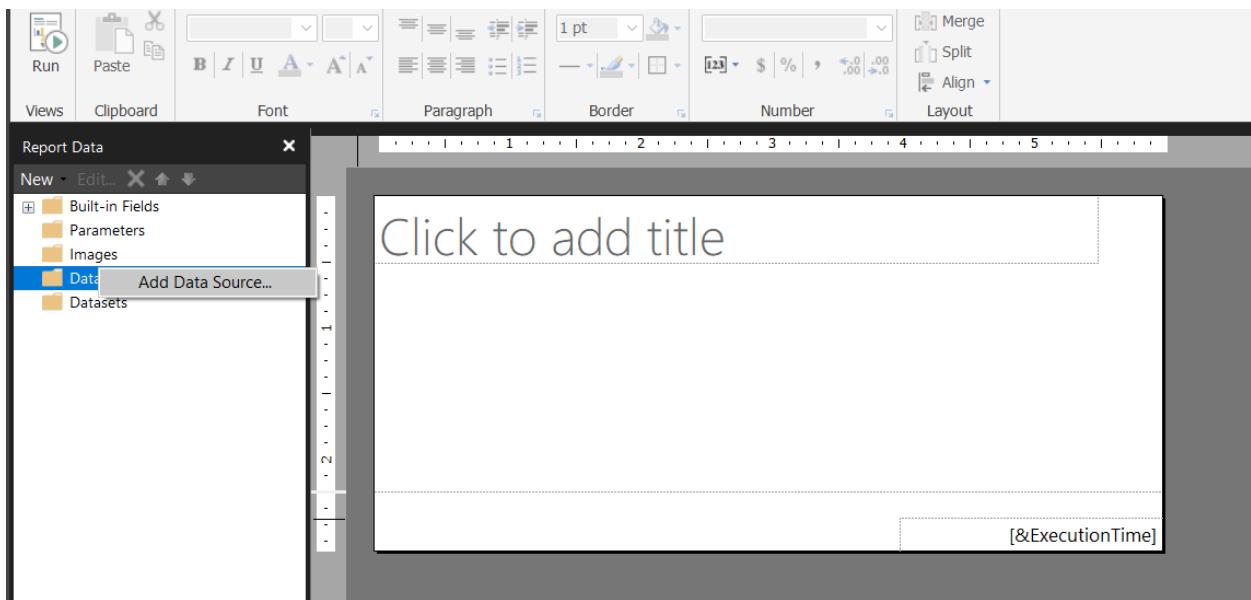


Sau khi hoàn thành

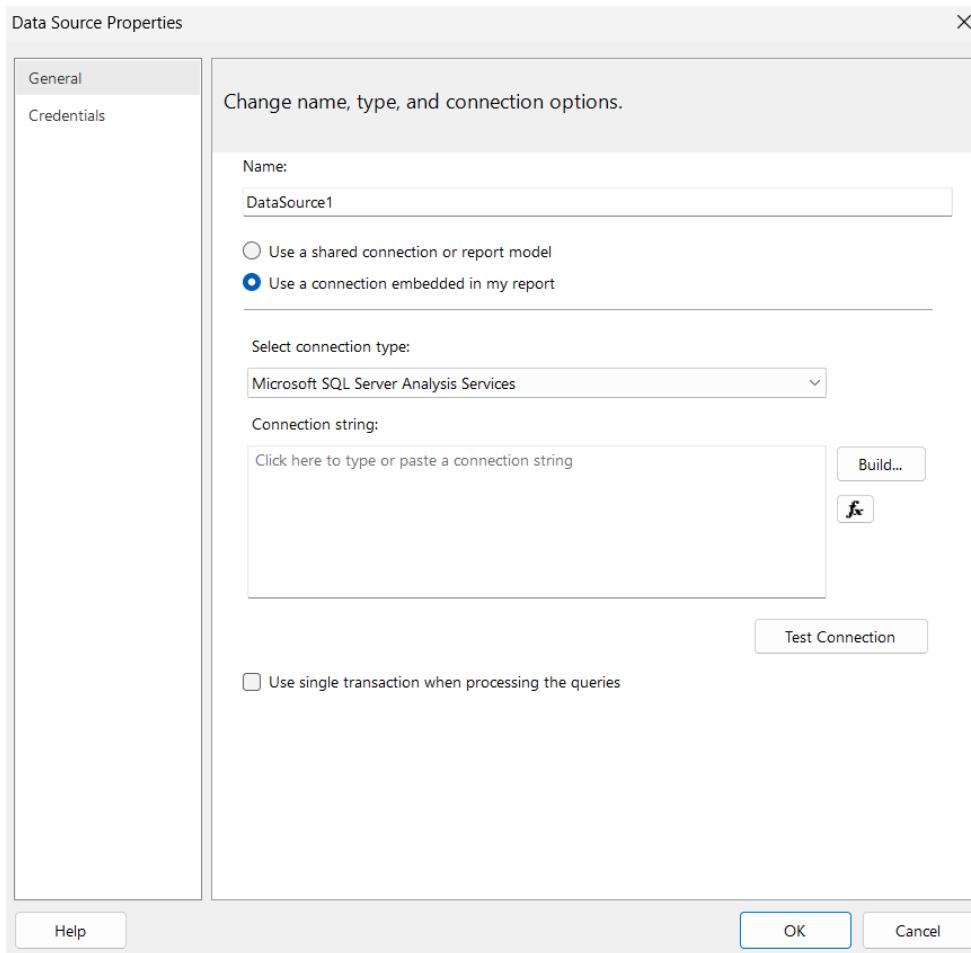


## 1.2. Tạo kết nối kho dữ liệu

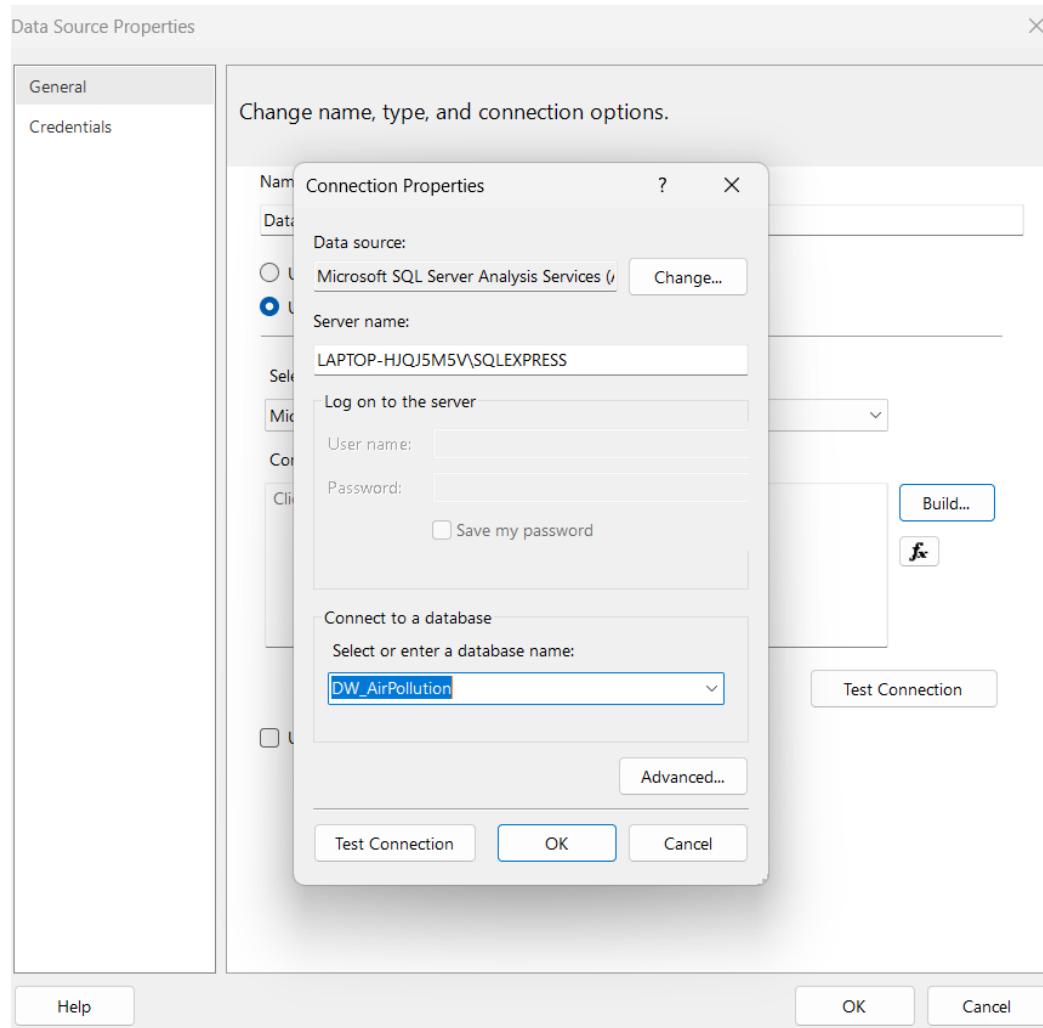
Bước 1: Ta thực hiện việc tạo mới 1 "Data Source"



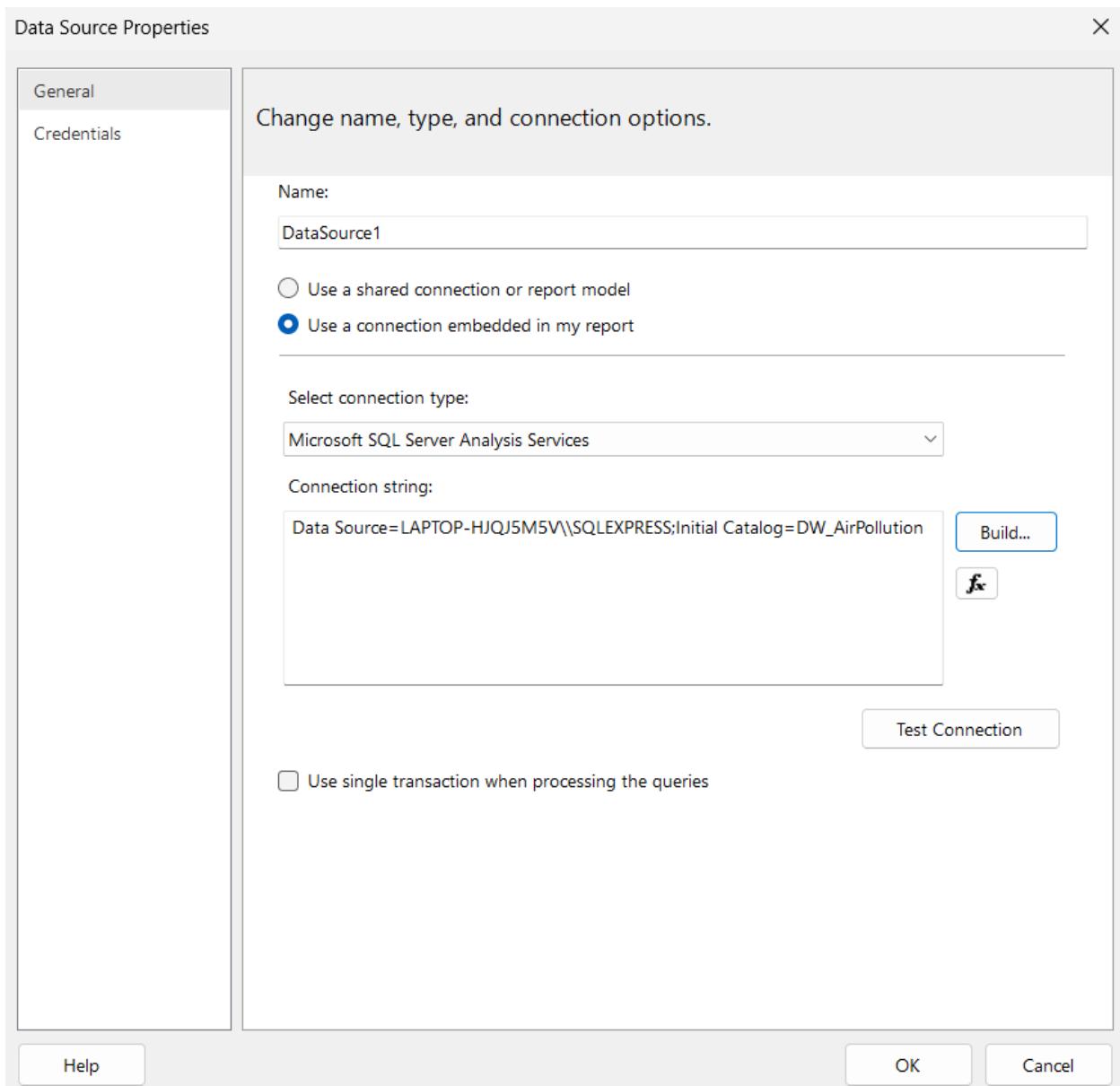
Bước 2: Ta chọn lại “ Use a connection embedded in my report ” và chọn connection “Microsoft SQL Server Analysis Services”.



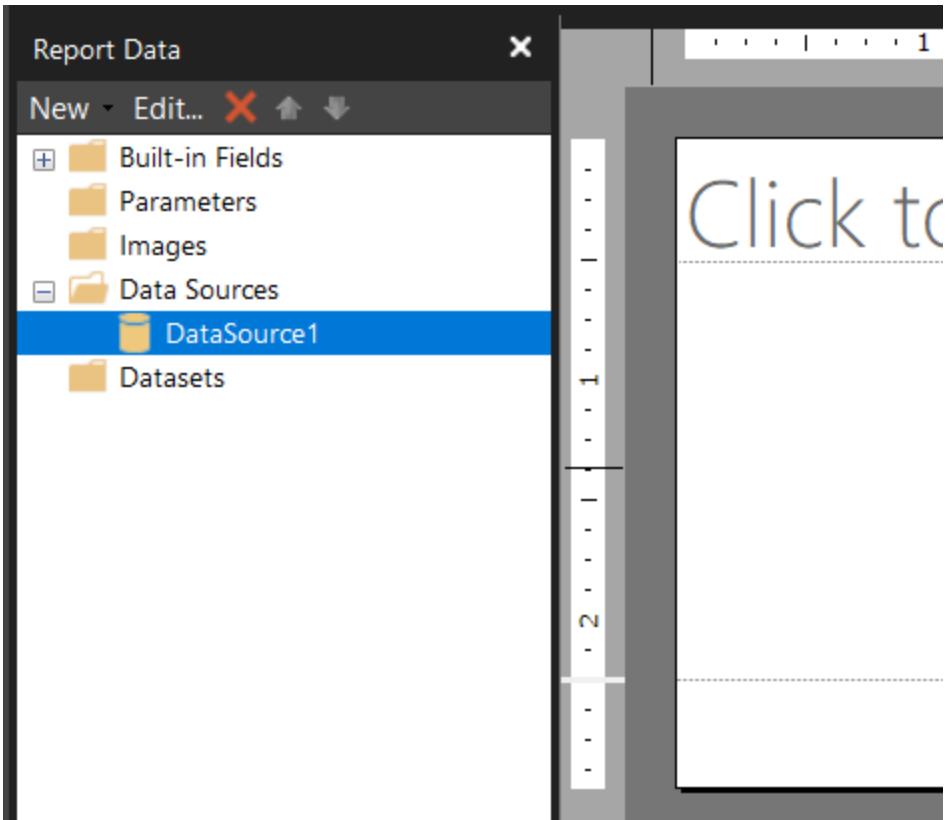
Bước 3 : để tạo mới 1 connect chọn “Build” và nhập các thông tin database cần thiết .



Bước 4 : Ta kiểm tra lại thông tin và chọn OK để hoàn tất .



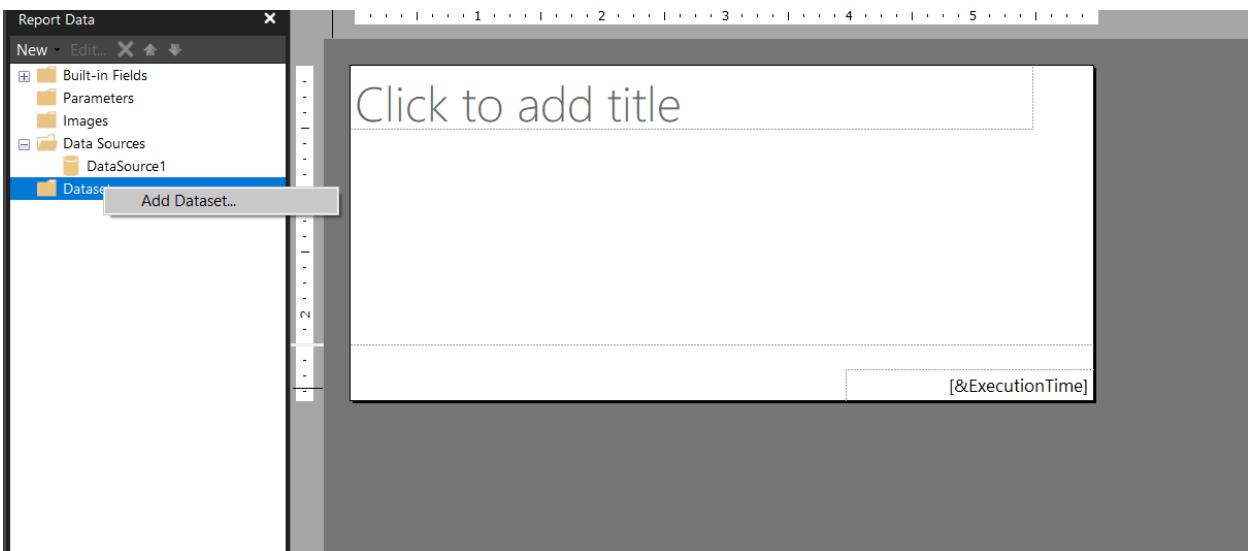
Kết quả



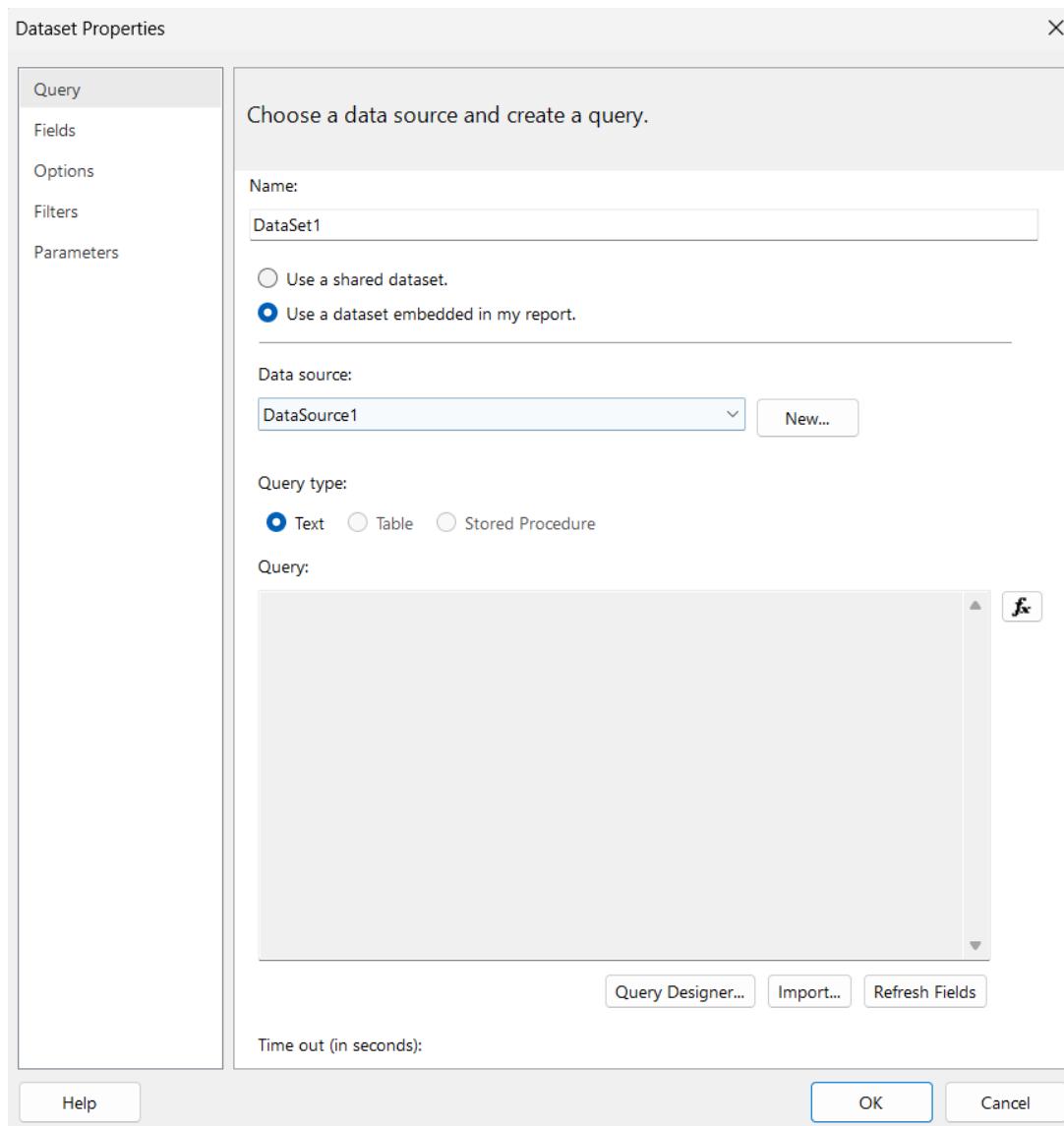
### 1.3. Thực hiện lập báo biểu trên Report Builder

#### 1.3.1. Báo biểu 1: (Report Grouping) đưa ra số lượng ô nhiễm SO2 trong từng năm của từng vùng

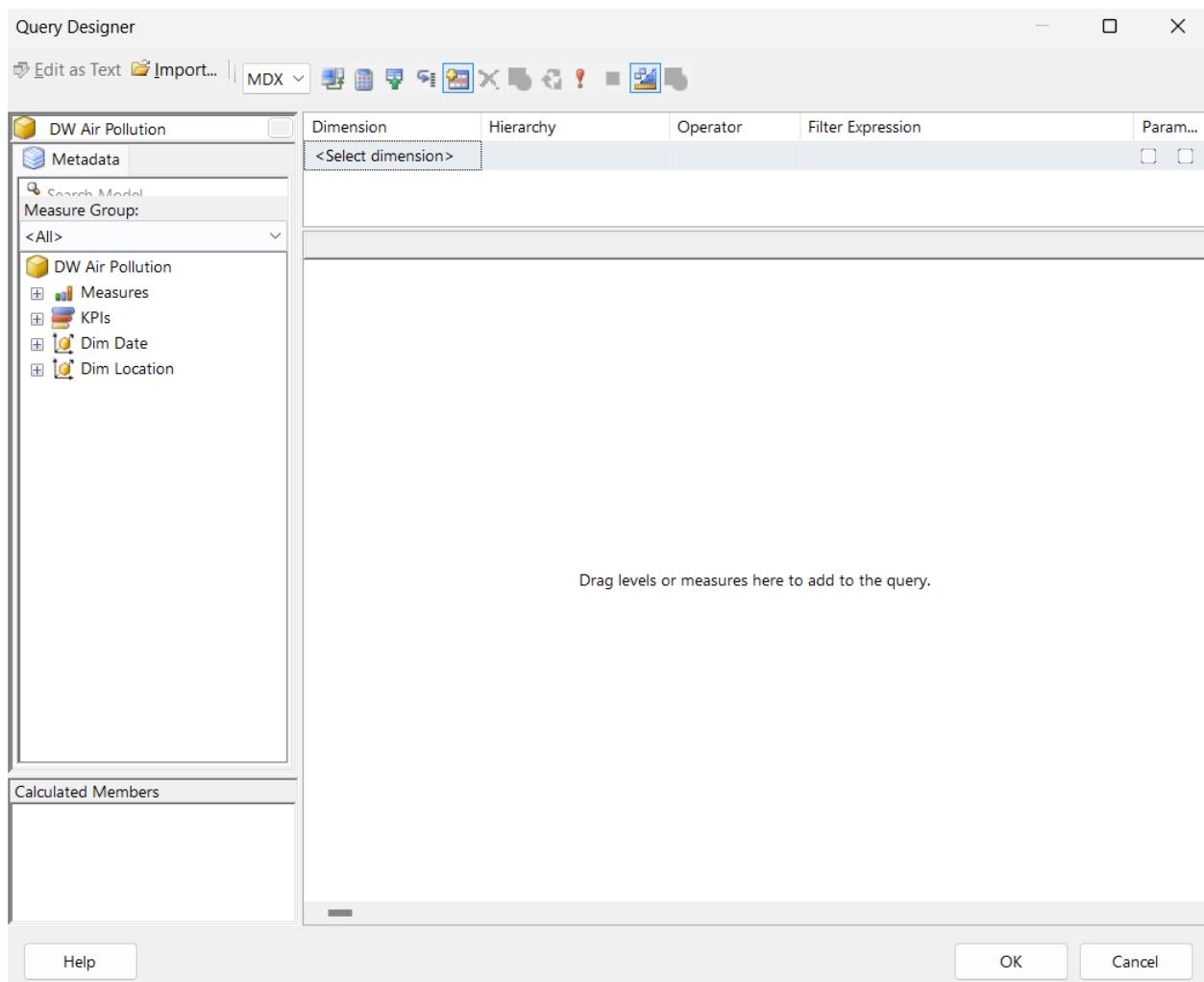
Bước 1 : Ta thêm dataset cần thiết cho báo biểu.



Bước 2 : Ta chọn lại " Use a connection embedded in my report " và chọn DataSource vừa mới tạo.



Bước 3 : Ta chọn hộp thoại “Query Designer”.



Bước 4 : Ta chọn những thuộc tính cần thiết cho báo biểu như hình minh họa sau đó nhấn OK.

Query Designer

Edit as Text Import... MDX

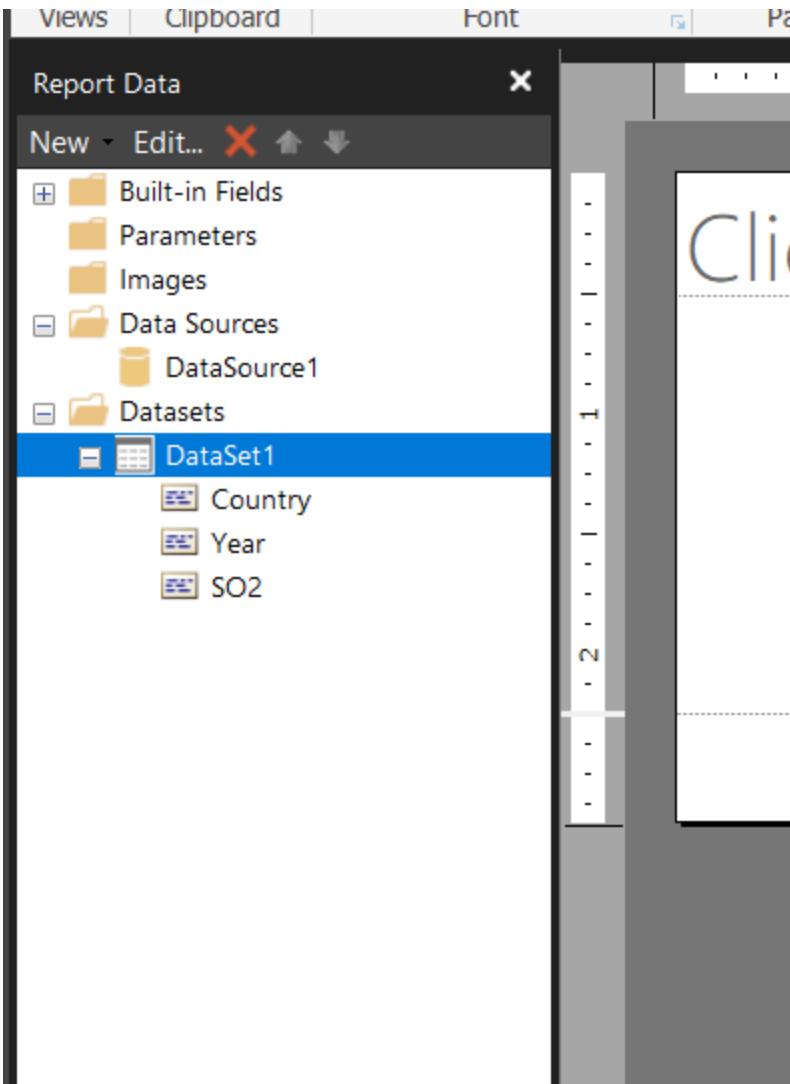
Dimension Hierarchy Operator Filter Expression Param...

<Select dimension>

Country	Year	SO2
Daejeon	2013	8
Daejeon	2014	1464
Daejeon	2015	1289
Daejeon	2016	1037
Daejeon	2017	787
Daejeon	2018	751
Daejeon	2019	858
Daejeon	2020	508
Daejeon	2021	1004
Daejeon	2022	187
Gangwon	2013	15
Gangwon	2014	1935
Gangwon	2015	2018
Gangwon	2016	1436
Gangwon	2017	1318
Gangwon	2018	1128
Gangwon	2019	1056
Gangwon	2020	792
Gangwon	2021	782
Gangwon	2022	145
Gwangju	2013	11
Gwangju	2014	2151
Gwangju	2015	1282

Help OK Cancel

Ta được kết quả



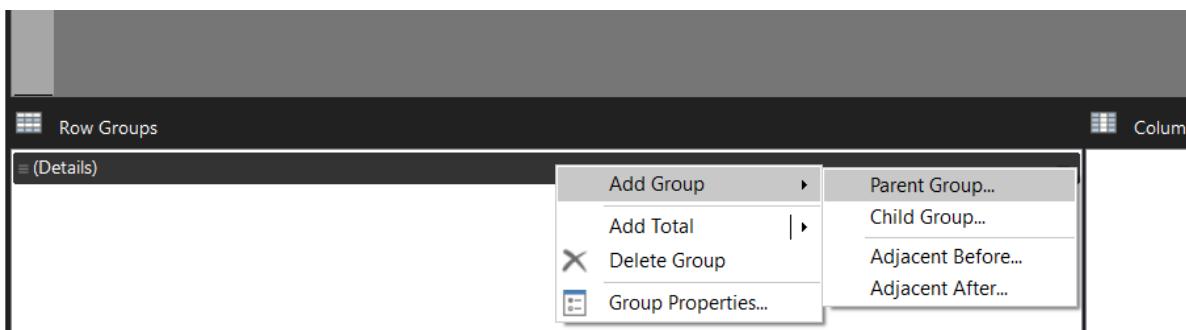
Bước 5 : Ta thực hiện tạo bảng và chỉnh sửa Report

Báo cáo tổng lượng ô nhiễm SO2 của từng vùng qua các năm

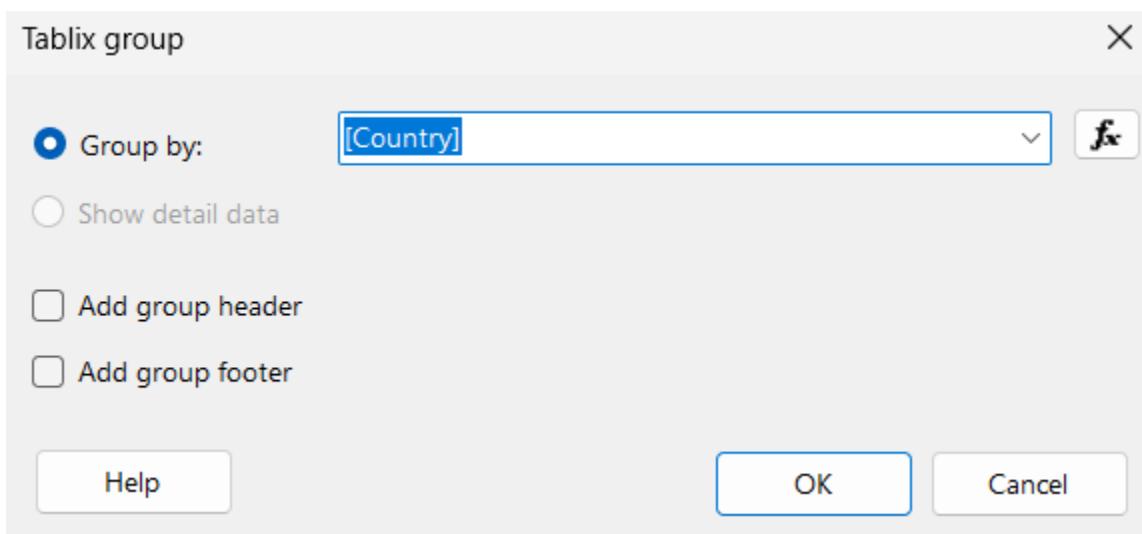
Country	Năm	SO2
[Country]	[Year]	[SO2]

[&ExecutionTime]

Bước 6 : Ta thực hiện Grouping “Country” bằng cách Add Parent Group



Ta thực hiện việc Group by : “Country”



Ta được kết quả



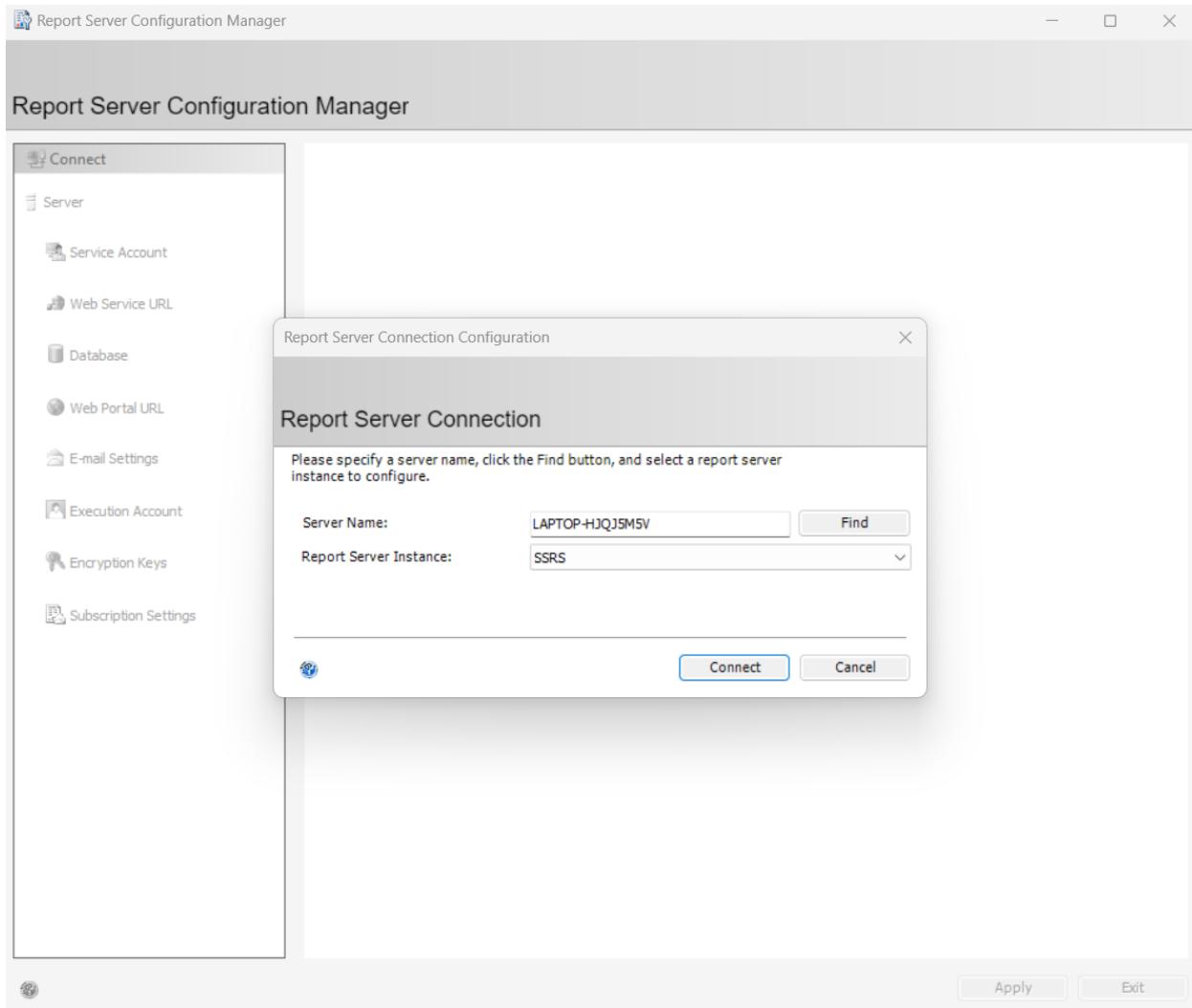
Bước 7 : Ta thực hiện việc chạy thử báo cáo bằng cách ấn “Run” , nếu thấy báo biểu hợp lý thì ta có thể xuất PDF.

#### Báo cáo tổng lượng ô nhiễm SO2 của từng vùng qua các năm

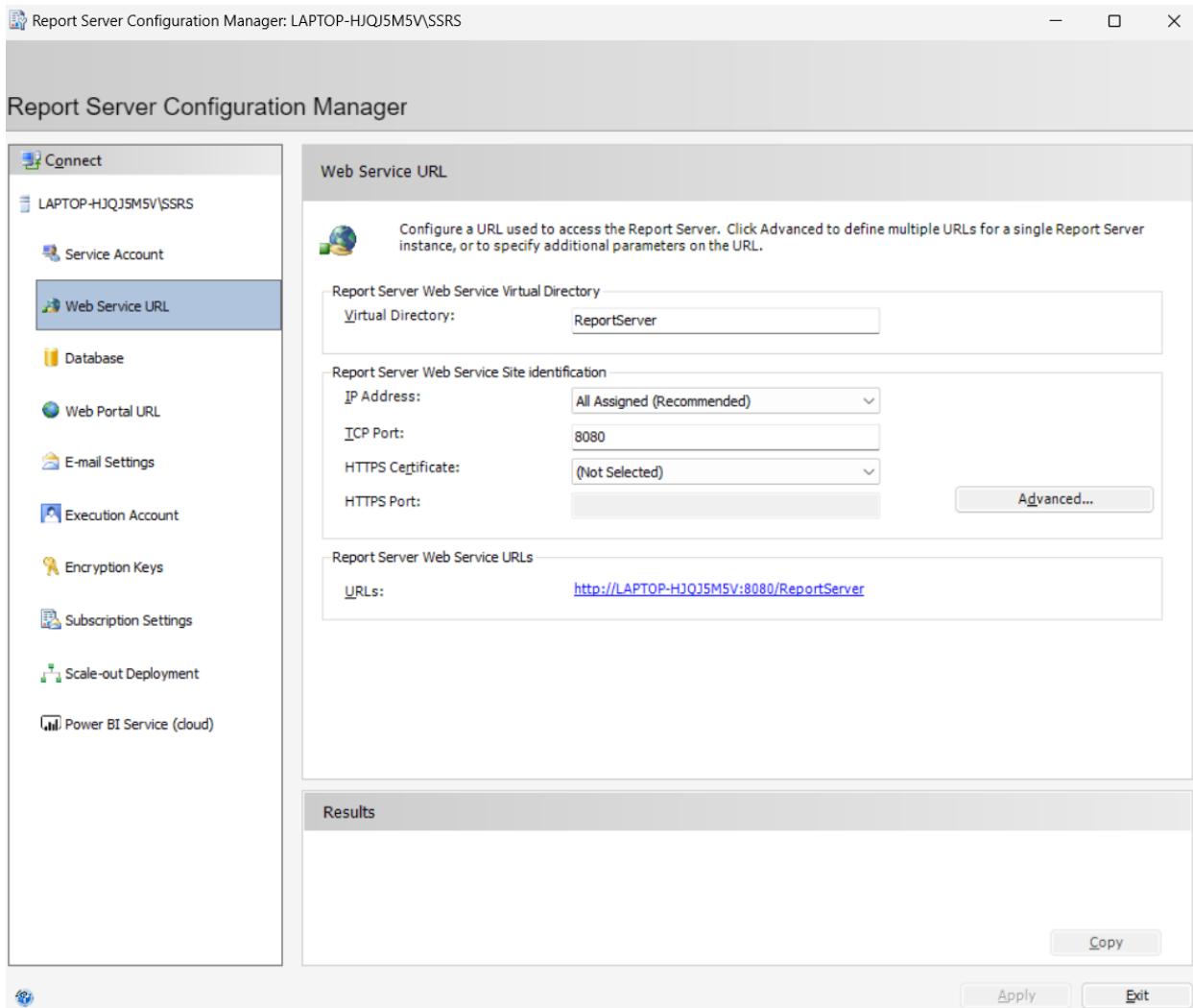
Country	Năm	SO2
Daejeon	2013	8
	2014	1464
	2015	1289
	2016	1037
	2017	787
	2018	751
	2019	858
	2020	508
	2021	1004
	2022	187
Gangwon	2013	15
	2014	1935
	2015	2018
	2016	1436
	2017	1318
	2018	1128
	2019	1056
	2020	792
	2021	782
	2022	145
Gwangju	2013	11
	2014	2151
	2015	1282
	2016	1254
	2017	1111

5/20/2024 12:17:37 AM

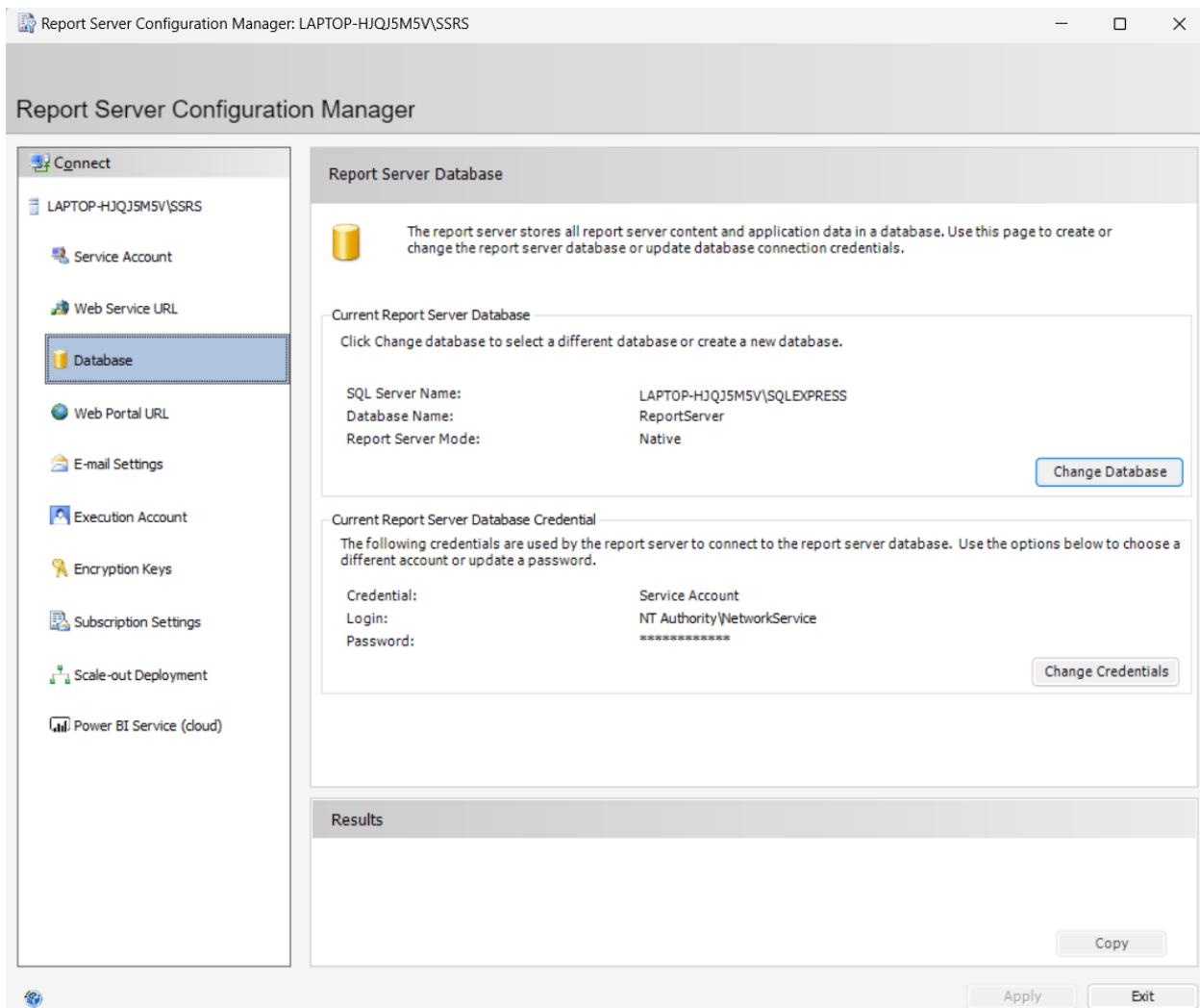
Bước 8 : Ta thực hiện việc đăng báo cáo này lên Report Server . Trước tiên ta cần cài đặt công cụ “**Report Server Configuration Manager**” . Sau đó ta khởi động công cụ và chọn Ok để kết nối.



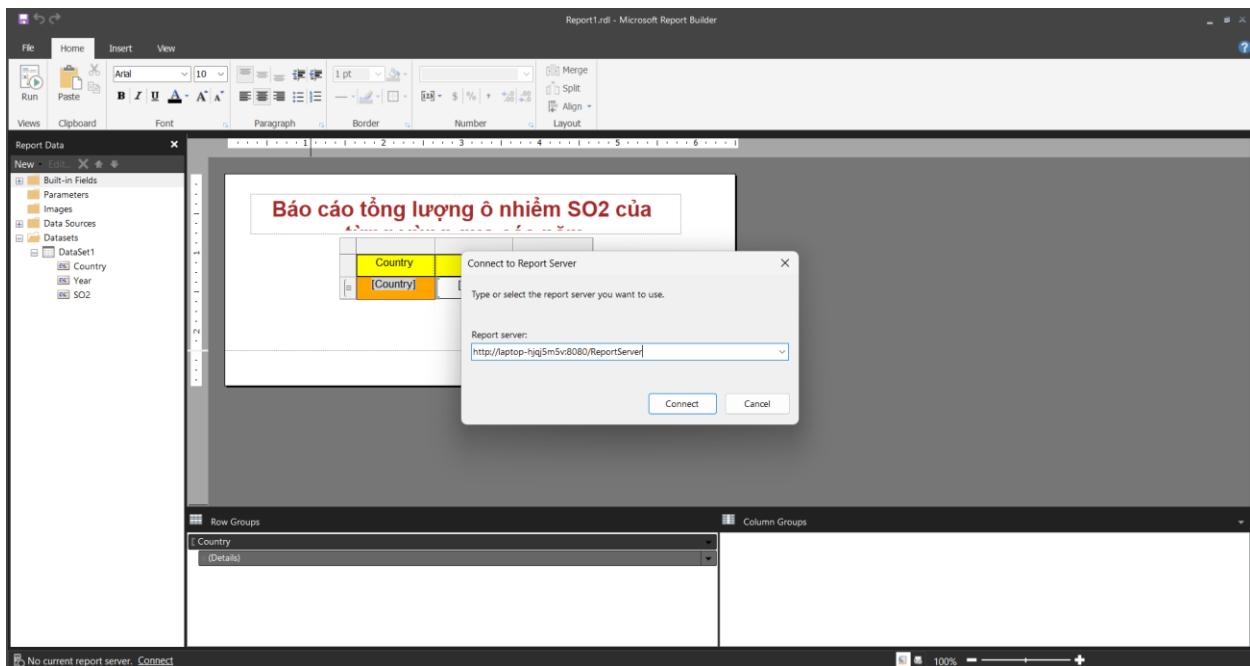
Bước 9 : Tại phần “Web Service URL” ta thực hiện cấu hình xong rồi ấn Apply.



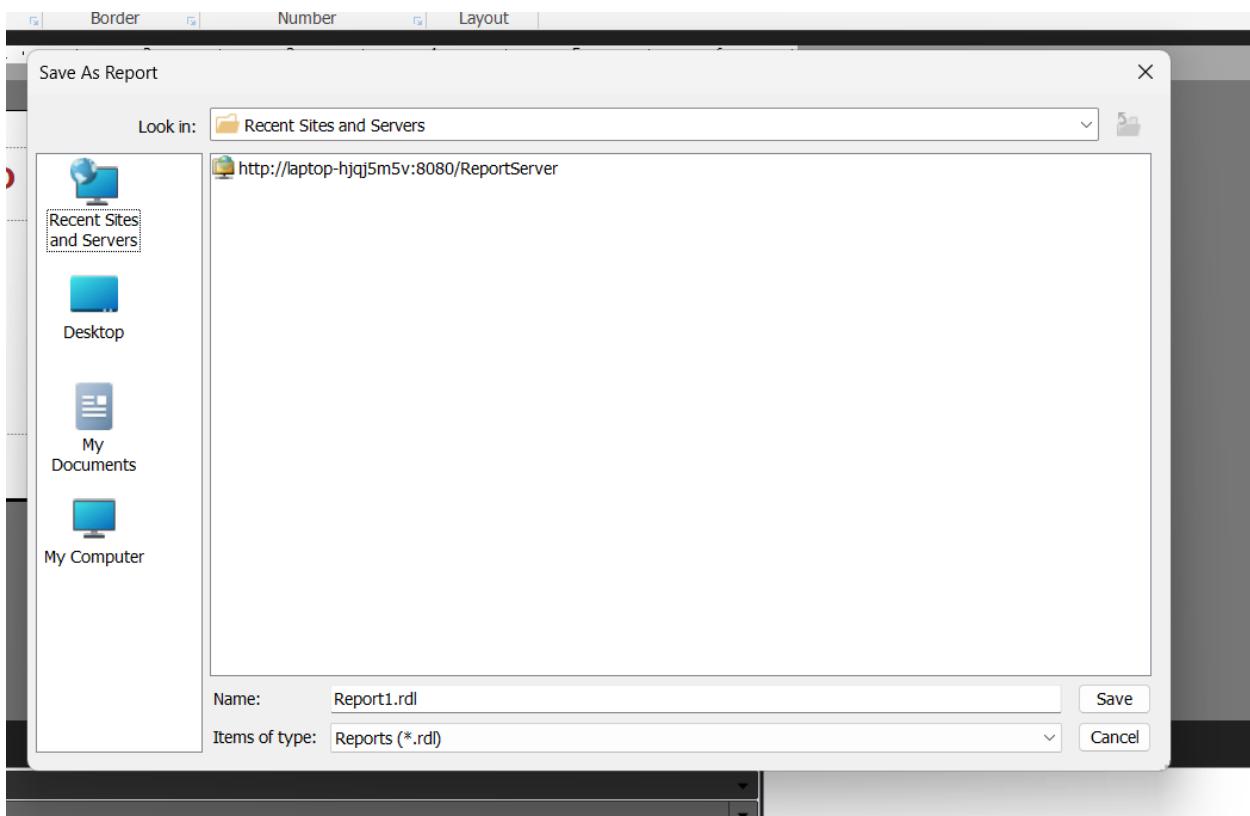
Bước 10 : Ở phần Database ta có thể tạo mới database Report bằng cách chọn “change Database”



Bước 12 : Ta thực hiện việc kết nối “Report Builder” với “Report Server”. Với link website : <http://laptop-hjqj5m5v:8080/ReportServer>



### Bước 13 : Tải lưu Report 1 vào “Report Server”



Bước 14 : Ta có thể lên trang web để kiểm tra report 1.

The screenshot shows a web browser window with the address bar containing 'Không bảo mật | laptop-hjqj5m5v:8080/ReportServer'. Below the address bar, the URL 'laptop-hjqj5m5v/ReportServer - /' is displayed. A horizontal line separates the address bar from the content area. In the content area, there is a table with three rows:

19 Tháng Năm 2024 11:08 CH	27005 Report1
19 Tháng Năm 2024 11:28 CH	67426 Report2
20 Tháng Năm 2024 1:07 SA	25695 Report3

Below the table, the text 'Version 15.0.1102.1167 de Microsoft SQL Server Reporting Services' is visible.

Bước 15 : Ta chọn Report 1 để xem .

The screenshot shows a Microsoft SQL Server Reporting Services report viewer. The title of the report is 'Báo cáo tổng lượng ô nhiễm SO2 của từng vùng qua các năm'. The report displays a table with three columns: 'Vùng' (Region), 'Năm' (Year), and 'SO2' (SO2 concentration). The data is grouped by region, with two groups shown: Daejeon and Gangwon. The table data is as follows:

Vùng	Năm	SO2
Daejeon	2013	8
	2014	1464
	2015	1289
	2016	1037
	2017	787
	2018	751
	2019	858
	2020	508
	2021	1004
	2022	187
Gangwon	2013	15
	2014	1935
	2015	2018
	2016	1436
	2017	1318
	2018	1128
	2019	1056
	2020	792
	2021	782
	2022	145

### 1.3.2. Báo biểu 2: (Report Grouping +Total +Matrix ) Báo cáo về tổng lượng ô nhiễm NO2 tại Paju theo ngày , tháng ,năm.

Bước 1 : Các bước tạo dataset tương tự như báo biểu 1. Và thực hiện chọn các thuộc tính cần thiết cho báo biểu và chọn ok.

Query Designer

Edit as Text Import... MDX

DW Air Pollution

Metadata

Search Model

Measure Group:

All >

- PM10
- PM2 5
- SO2
- KPIs
- Dim Date
  - BOTTOM 3 THANG
  - NgaycochisoNo2caon
  - Tong luong Pm25 từ :
  - Top 7 ngày
  - Day Fix
  - ID Date
  - Month Fix
  - Year
  - Date
- Dim Location
  - Tong luong O3 tu 10%
  - TongLuong03CaoNha
  - TongLuongPm10>50
  - TOP 3 THANH PHO

Calculated Members

Dimension Hierarchy Operator Filter Expression Param...

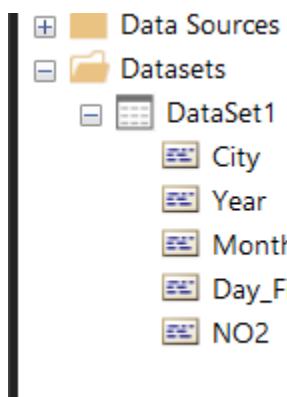
Dim Location City Equal { Paju }

<Select dimension>

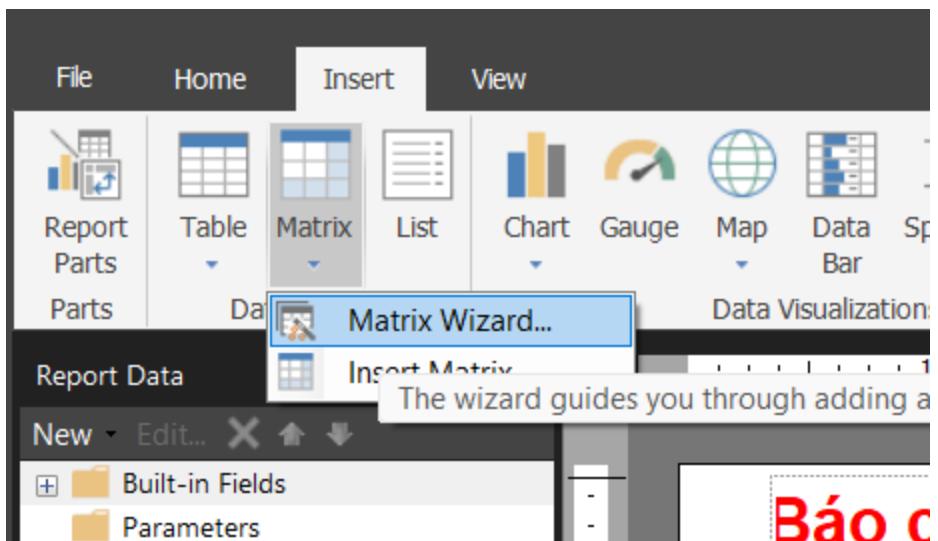
City	Day Fix	Month Fix	Year	NO2
Paju	20170...	201709	2017	4
Paju	20170...	201709	2017	4
Paju	20170...	201709	2017	0
Paju	20170...	201709	2017	4
Paju	20170...	201709	2017	3
Paju	20170...	201709	2017	2
Paju	20170...	201709	2017	4
Paju	20170...	201709	2017	12
Paju	20170...	201709	2017	7
Paju	20171...	201710	2017	3
Paju	20171...	201710	2017	2
Paju	20171...	201710	2017	2
Paju	20171...	201710	2017	5
Paju	20171...	201710	2017	2
Paju	20171...	201710	2017	2
Paju	20171...	201710	2017	4
Paju	20171...	201710	2017	3
Paju	20171...	201710	2017	8
Paju	20171...	201710	2017	2
Paju	20171...	201710	2017	3
Paju	20171...	201710	2017	3
Paju	20171...	201710	2017	2
Paju	20171...	201710	2017	3

Help OK Cancel

Ta được kết quả :



Bước 2 : Ta thực hiện tạo một Matrix vào với các tùy chọn là "Matrix Wizard"



Bước 3 : Ta kéo các thuộc tính vào trong các ô tương ứng như sau

New Table or Matrix X

Arrange fields

Arrange fields to group data in rows, columns, or both, and choose values to display. Data expands across the page in column groups and down the page in row groups. Use functions such as Sum, Avg, and Count on the fields in the Values box.

Available fields	Column groups	Row groups	Values
City Year Month_Fix Day_Fix NO2			$\Sigma$ Sum(NO2)

Help < Back Next > Cancel

Bước 4 : Ta kiểm tra và ấn next

New Table or Matrix

Choose the layout

If you choose to show subtotals and grand totals, you can place them above or below the group. Stepped reports show hierarchical structure with indented groups in the same column.

Options:

- Show subtotals and grand totals
- Blocked, subtotal below
- Blocked, subtotal above
- Stepped, subtotal above
- Expand/collapse groups

Preview

City	Year	Month Fix	Day Fix	NO2
[City]	[Year]	[Month_Fix]	[Day_Fix]	[Sum(NO2)]
		Total	Total	[Sum(NO2)]
		Total		[Sum(NO2)]
		Total		[Sum(NO2)]
Total				[Sum(NO2)]

Help < Back Next > Cancel

Bước 4 : Ta kiểm tra lần cuối và ấn OK .

City	Year	Month Fix	Day Fix	NO2
[City]	[Year]	[Month_Fix]	[Day_Fix]	[Sum(NO2)]
		Total	Total	[Sum(NO2)]
		Total		[Sum(NO2)]
		Total		[Sum(NO2)]
Total				[Sum(NO2)]

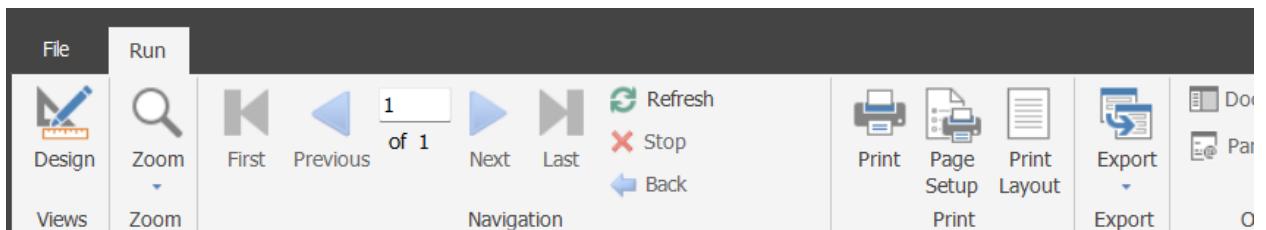
Bước 5 : Ta định dạng lại báo biểu cho hợp với báo cáo .

## Báo cáo tổng lượng ô nhiễm NO2 tại Paju

Thành phố	Năm	Tháng	Ngày	NO2
[City]	[Year]	[Month_Fix]	[Day_Fix]	[Sum(NO2)]
		Total		[Sum(NO2)]
	Total			[Sum(NO2)]
Total				[Sum(NO2)]

[&ExecutionTime]

Bước 6 : Ta ấn “Run” để kiểm tra báo biểu và xuất file PDF.



## Báo cáo tổng lượng ô nhiễm NO2 tại Paju

Thành phố	Năm	Tháng	Ngày	NO2
■ Paju	■ 2017	■ 201709	20170922	4
			20170923	4
			20170924	0
			20170925	4
			20170926	3
			20170927	2
			20170928	4
			20170929	12
			20170930	7
			Total	40
	■ 201710		Total	119
	■ 201711		Total	213
	■ 201712		Total	340
			Total	712
	■ 2018		Total	1554
	■ 2019		Total	1701
	■ 2020		Total	1460
	■ 2021		Total	1665
	■ 2022		Total	332
			Total	7424
Total				7424

5/20/2024 9:04:04 AM

Bước 7 : Ta thực hiện đẩy Report2 lên trên “Report Server” ta được kết quả

← → C (Không bảo mật | laptop-hjqj5m5v:8080/ReportServer/Pages/ReportViewer.aspx?%2fReport2&rs:Command=Render

1 sur 1 100% Rechercher | Suivant

**Báo cáo tổng lượng ô nhiễm NO2 tại PaJu**

Thành phố	Năm	Tháng	Ngày	NO2
PaJu	2017	201709	20170922	4
			20170923	4
			20170924	0
			20170925	4
			20170926	3
			20170927	2
			20170928	4
			20170929	12
			20170930	7
		Total		40
		201710	Total	119
		201711	Total	213
		201712	Total	340
			Total	712
	2018	Total		1554
	2019	Total		1701
	2020	Total		1460
	2021	Total		1665
	2022	Total		332
		Total		7424

19/5/2024

### 1.3.3. Báo biểu 3:(Report Grouping +Total) Báo cáo với từng thành phố đưa ra tổng lượng PM10 từ 2014-2018.

Bước 1 : Các bước tạo dataset tương tự như báo biểu 1. Và thực hiện chọn các thuộc tính cần thiết cho báo biểu và chọn ok.

Query Designer

Edit as Text Import... MDX

**DW Air Pollution**

Dimension Hierarchy Operator Filter Expression Param...

Dim Date Year Equal { 2014, 2015, 2017, 2018, 2016 }

<Select dimension>

City	Year	PM10
Ban...	2014	12959
Ban...	2015	10648
Ban...	2016	10299
Ban...	2017	8744
Ban...	2018	7991
Ge...	2014	16502
Ge...	2015	15497
Ge...	2016	15555
Ge...	2017	13316
Ge...	2018	13499
Gw...	2014	14991
Gw...	2015	15377
Gw...	2016	14943
Gw...	2017	11800
Gw...	2018	13261
Jeo...	2014	11403
Jeo...	2015	11169
Jeo...	2016	12014
Jeo...	2017	11018
Jeo...	2018	12266
Jun...	2014	14024
Jun...	2015	12784
Jun...	2016	14840

KPIs Dim Date Dim Location Calculated Members

Help OK Cancel

Bước 2 : Ta thực hiện chỉnh báo cáo.

**Báo cáo tổng lượng PM10 của từng thành phố từ năm 2014-2018**

City	Year	PM10
[City]	[Year]	[PM10]

[&Execution Time]

Bước 3 : Ta thực hiện việc grouping "City".

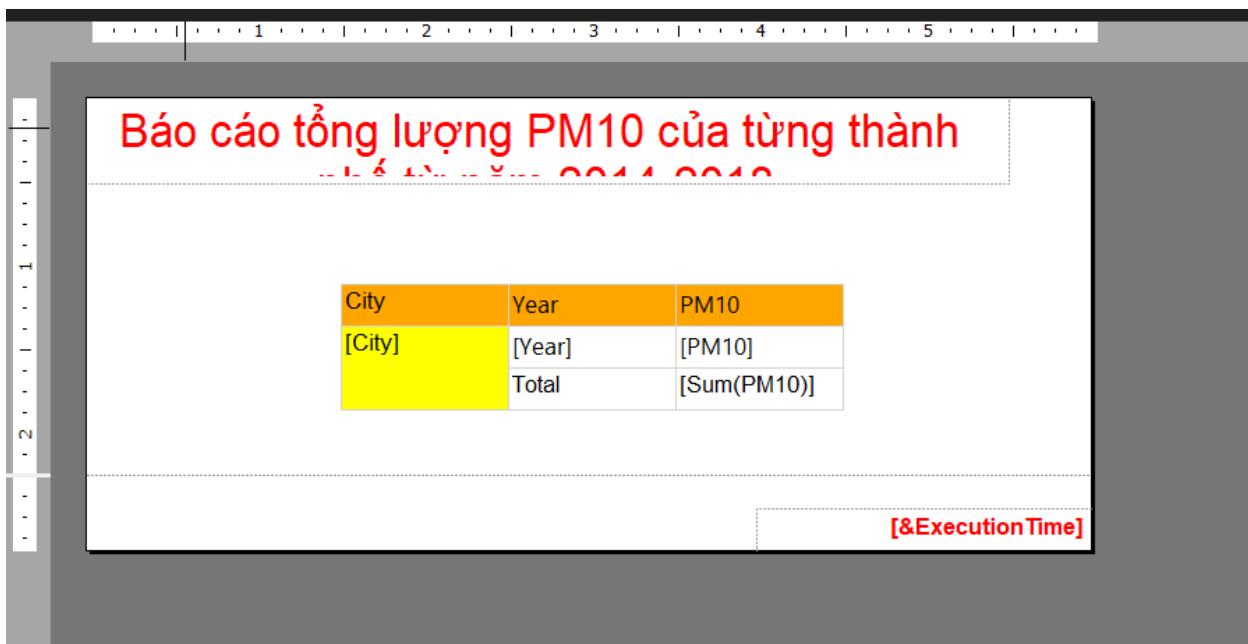
The screenshot shows the Microsoft Report Designer interface. At the top, there is a header section with the title "Báo cáo tổng lượng PM10 của từng thành phố" and a date range from "1/1/2014" to "31/12/2014". Below the header is a table with three columns: "City", "Year", and "PM10". The "City" column is highlighted with a yellow background. The table has two rows: a header row with column names and a data row with values "[City]", "[Year]", and "[PM10]". In the bottom right corner of the table, there is a red text label "[&ExecutionTime]". At the bottom of the screen, there is a "Row Groups" pane. It shows a single group named "City" under the "Row Groups" category. The "(Details)" category is also visible below it.

Bước 4 : Ta thực hiện “add total ” cho PM10.

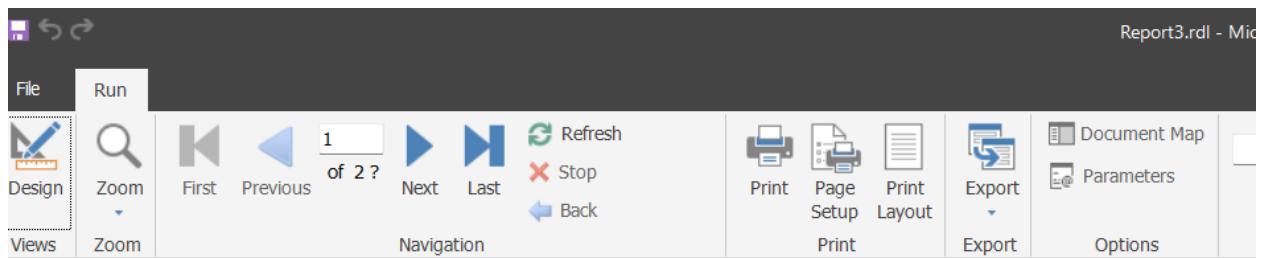
## Báo cáo tổng lượng PM10 của từng thành phố

A screenshot of a Microsoft Power BI report interface. At the top, there is a red header bar with the text "Báo cáo tổng lượng PM10 của từng thành phố". Below this is a title bar with the text "Tổng quan 2014-2019". The main area shows a table with three columns: "City", "Year", and "PM10". The first row has headers, and the second row contains data with cells labeled "[City]", "[Year]", and "[PM10]". A context menu is open over the "[PM10]" cell, showing options like "Text Box", "Cut", "Copy", "Paste", "Delete", "Select", "Expression...", "Text Box Properties...", "Tablix", "Insert Column", "Insert Row", "Delete Columns", "Delete Rows", "Add Group", "Row Group", "Add Total" (which is highlighted in grey), "Insert", and "Cell Properties...". The background of the Power BI interface is dark grey.

Ta được kết quả sau



Bước 5 : Ta thực hiện việc “Run” báo cáo và xuất file PDF.



## Báo cáo tổng lượng PM10 của từng thành phố từ năm 2014-2018

City	Year	PM10
Bangsan-Myeon	2014	12959
	2015	10648
	2016	10299
	2017	8744
	2018	7991
	Total	50641
Geumchon-Dong	2014	16502
	2015	15497
	2016	15555
	2017	13316
	2018	13499
	Total	74369
Gwanak-Gu	2014	14991
	2015	15377
	2016	14943
	2017	11800
	2018	13261
	Total	70372
Jeongnim-Dong	2014	11403
	2015	11169

Bước 6: Ta thực hiện việc kết nối "Report Builder" với "Report Server"

← → C △ Không bảo mật | laptop-hjqj5m5v:8080/ReportServer/Pages/ReportViewer.aspx?%2fReport3&rs:Command=Render

1 sur 2 ? 100% Rechercher | Suiva

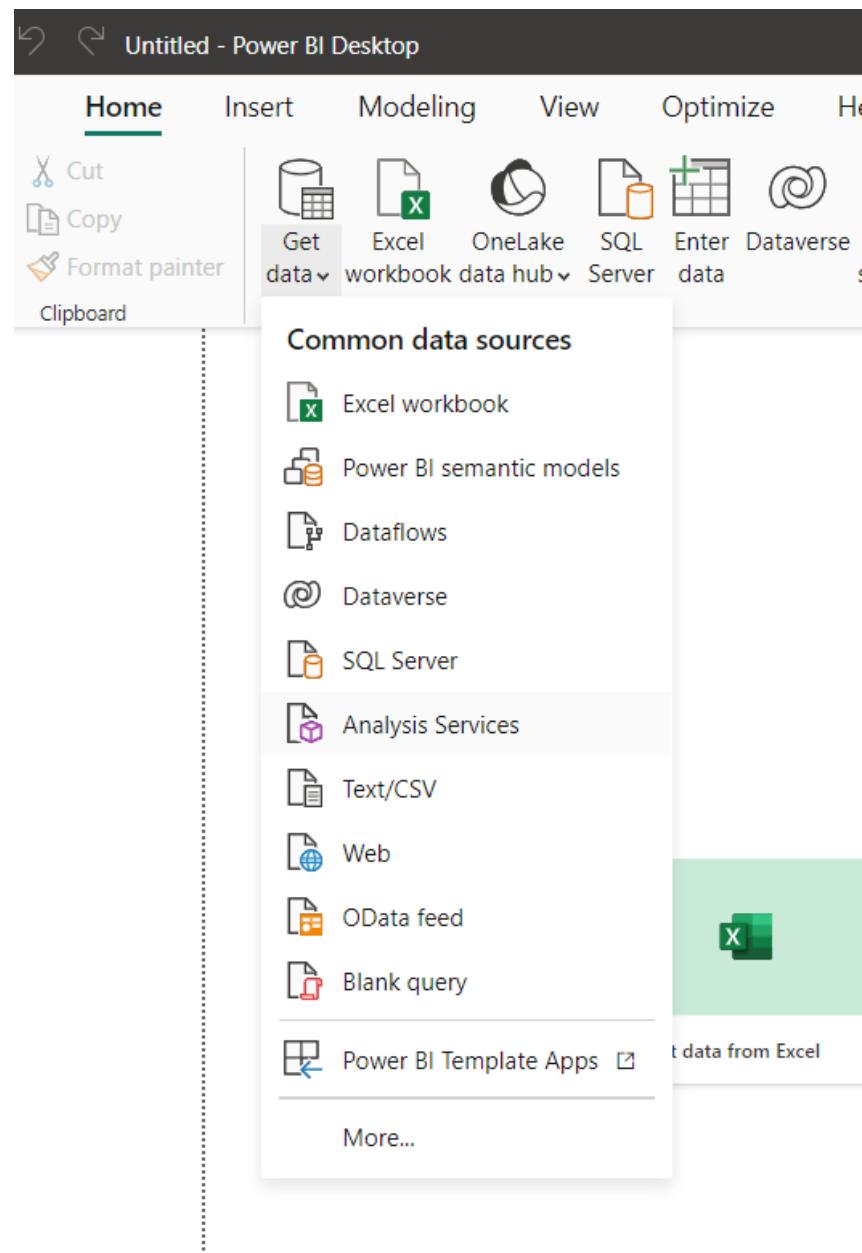
Báo cáo tổng lượng PM10 của từng thành phố từ năm 2014-2018

Thành phố	Năm	PM10
Bangsan-Myeon	2014	12959
	2015	10648
	2016	10299
	2017	8744
	2018	7991
	Total	50641
Geumchon-Dong	2014	16502
	2015	15497
	2016	15555
	2017	13316
	2018	13499
	Total	74369
Gwanak-Gu	2014	14991
	2015	15377
	2016	14943
	2017	11800
	2018	13261
	Total	70372
Jeongnim-Dong	2014	11403
	2015	11169
	2016	12014
	Total	34586

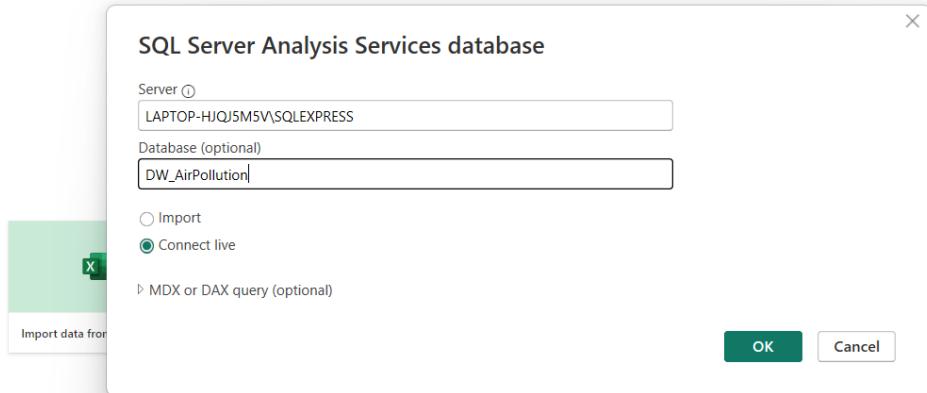
## 2. Quá trình lập báo biểu bằng công cụ Power BI

### 2.1. Kết nối Power BI với database

Bước 1 : tại trang Power BI ta chọn "Get data" sau đó là "Analysis Service"



Bước 2 : ta nhập các thông tin Server và database cần thiết .



Bước 3 : Ta chọn lấy "DW\_AirPollution" và chọn OK.

Navigator

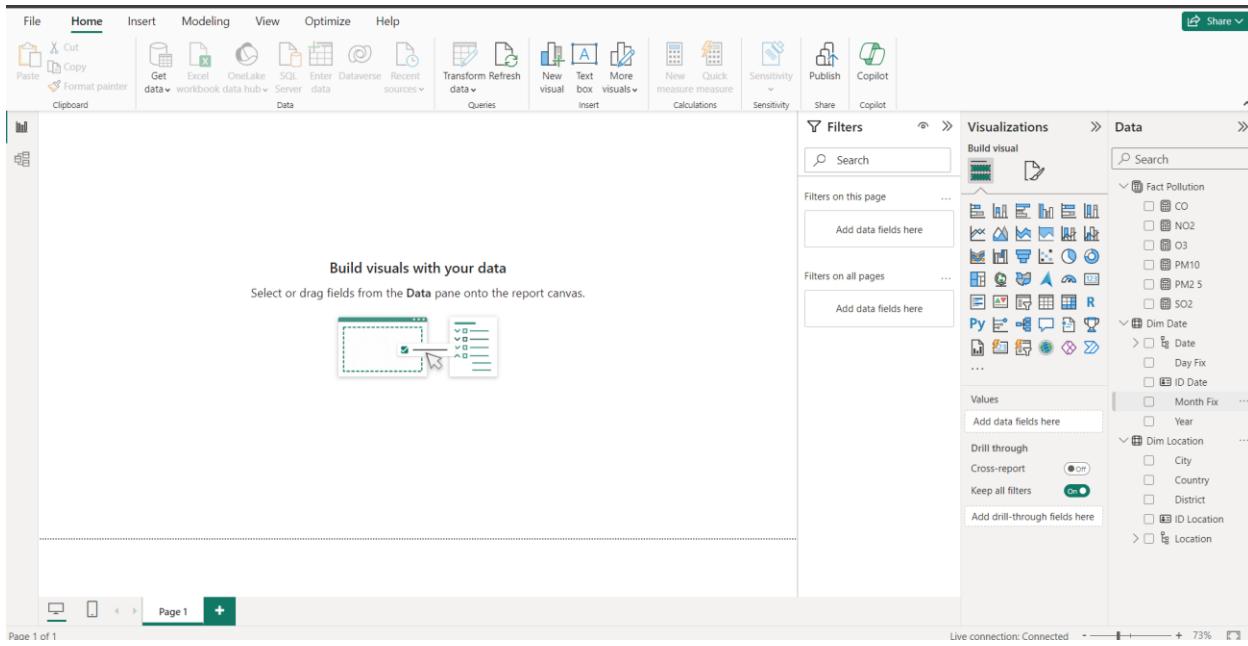
DW Air Pollution  
Last Modified: 05/20/2024 01:06:32

This model contains the following dimensions and measures  
Dim Date, Dim Location, PM2 5, PM10, O3, NO2, SO2, CO

LAPTOP-HJQJ5M5V\SQLEXPRESS: DW\_AirPollu...  
DW Air Pollution

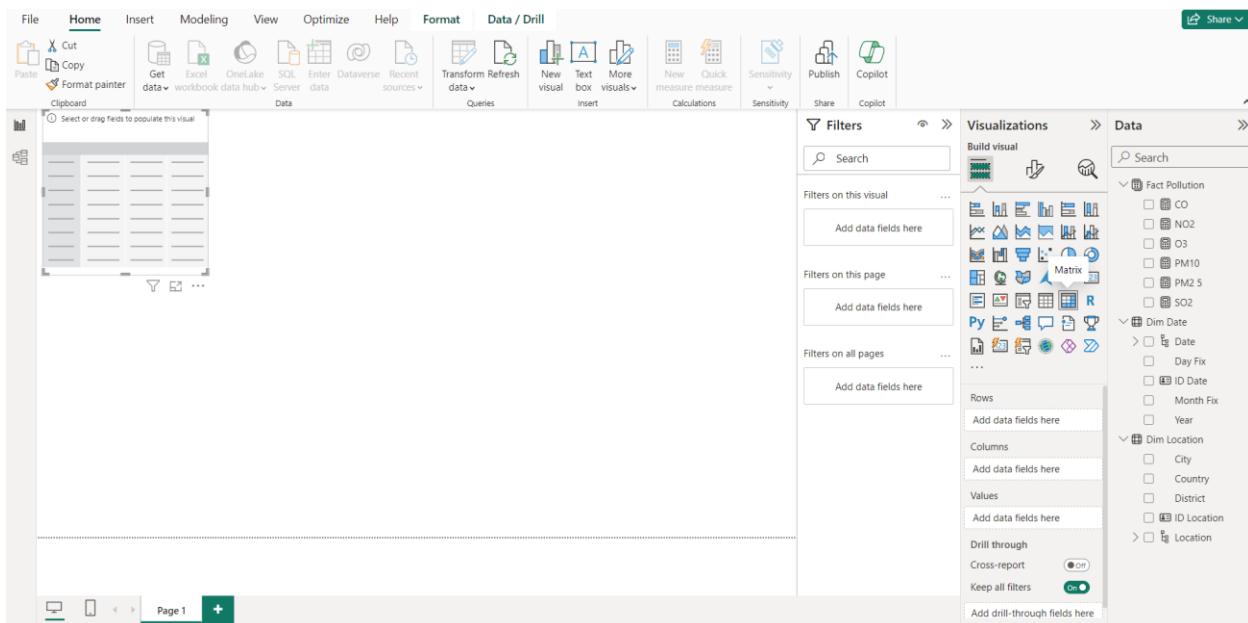
OK Cancel

Vậy ta đã kết nối thành công với database



## 2.2. Báo cáo 1: đưa ra số lượng ô nhiễm SO2 trong từng năm của từng vùng

Bước 1 : Ta chọn table Matrix để tạo báo cáo kiểu bảng .



Bước 2 : Ta thực hiện việc kéo các thuộc tính hợp lý cho bảng và chỉnh sửa thông tin bảng .

Số lượng ô nhiễm SO2 trong từng năm của từng vùng qua các năm

Country	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
Daejeon	8	1464	1289	1037	787	751	858	508	1004	187	7893
Gangwon	15	1935	2018	1436	1318	1128	1056	792	782	145	10625
Gwangju	11	2151	1282	1254	1111	1137	1203	1031	1059	136	10375
<b>Total</b>	<b>78</b>	<b>15777</b>	<b>14751</b>	<b>13363</b>	<b>10691</b>	<b>14430</b>	<b>15468</b>	<b>17170</b>	<b>18245</b>	<b>2715</b>	<b>122688</b>

Bước 3 : Ta thêm “Pie Chart” để có thể hiển thị rõ ràng hơn về dữ liệu .

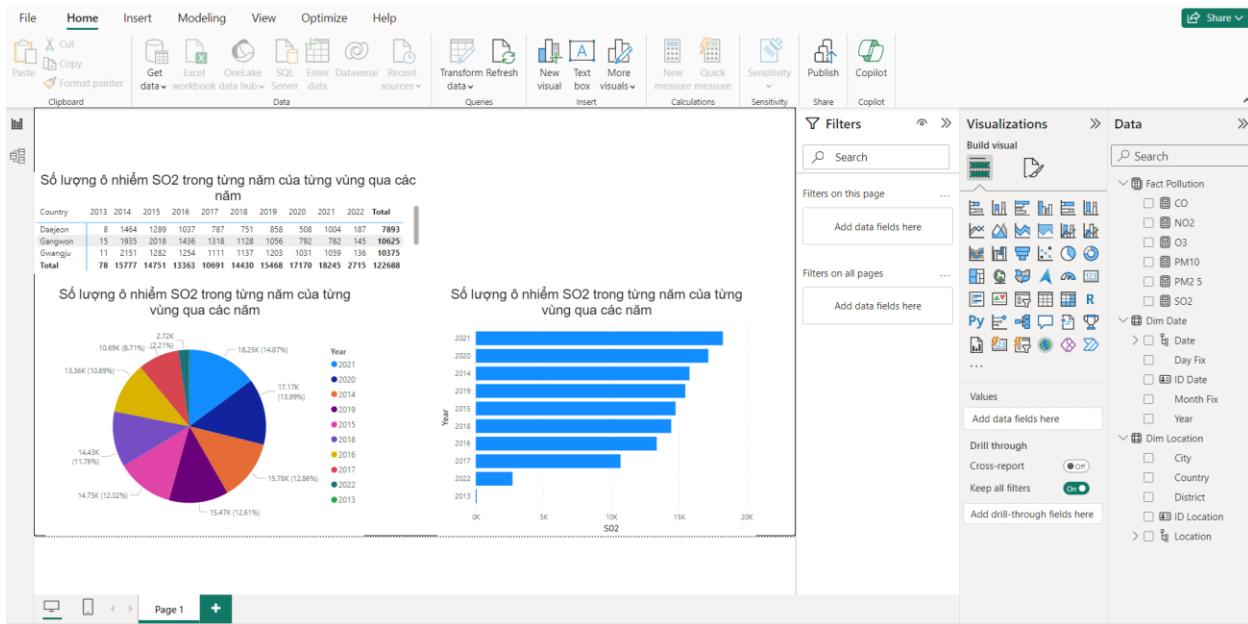
Số lượng ô nhiễm SO2 trong từng năm của từng vùng qua các năm

Year

- 2021
- 2020
- 2019
- 2018
- 2017
- 2016
- 2015
- 2014
- 2013

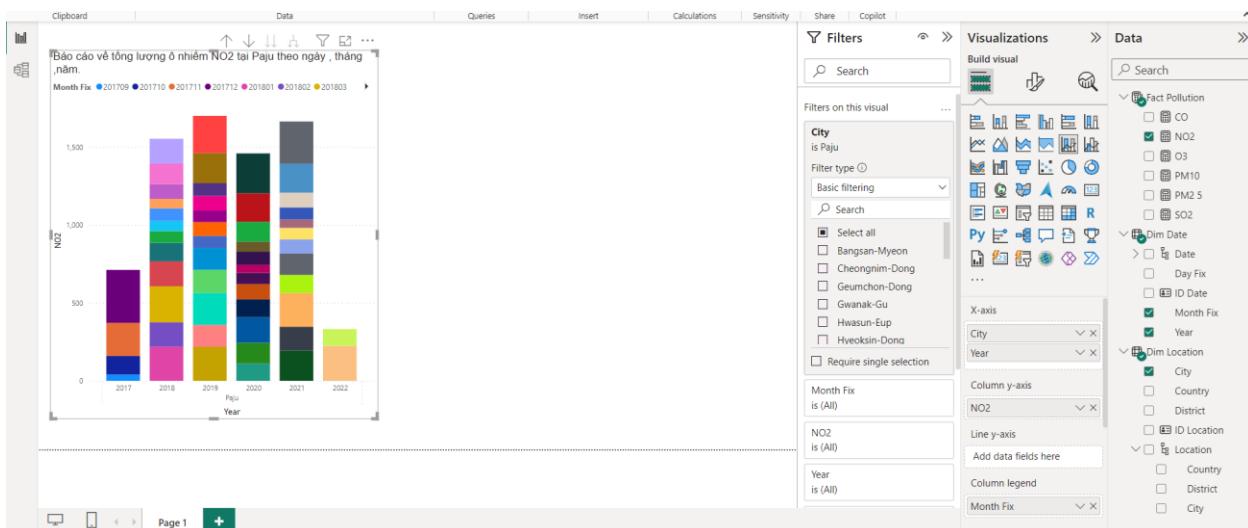
Year	Value	Percentage
2021	10.69K	(8.71%)
2020	2.72K	(2.21%)
2019	18.25K	(14.87%)
2018	17.17K	(13.99%)
2017	15.78K	(12.88%)
2016	14.75K	(12.02%)
2015	14.43K	(11.76%)
2014	13.36K	(10.89%)
2013	78	(100%)

Bước 4 : Ta thêm “Line Chart” để có thể hiển thị rõ ràng hơn về dữ liệu .

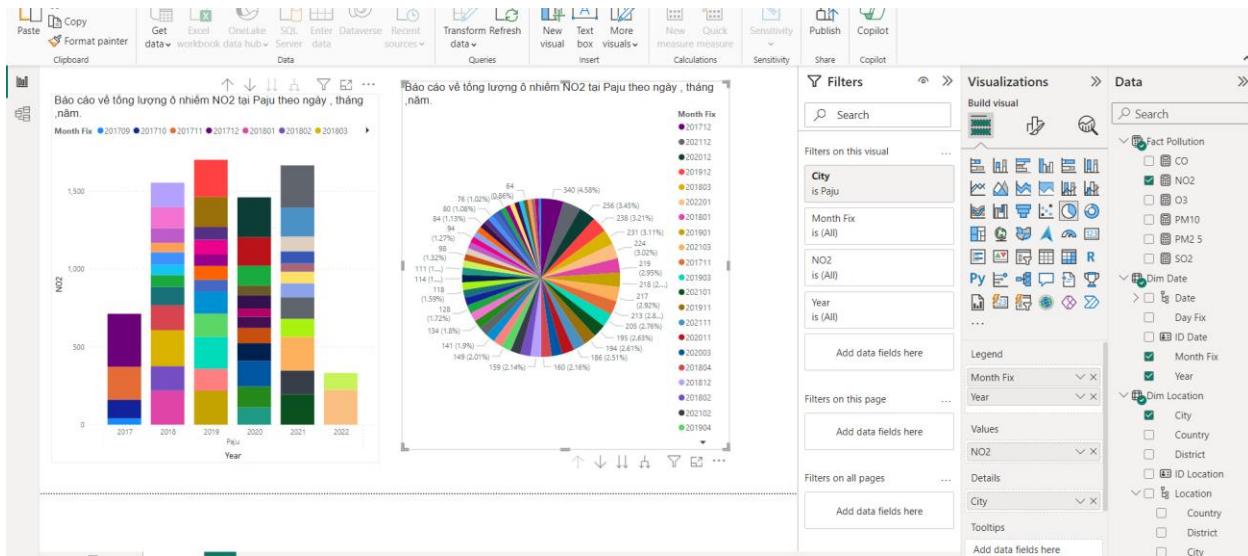


### 2.3. Báo cáo 2 : Báo cáo về tổng lượng ô nhiễm NO2 tại Paju theo tháng, năm.

Bước 1 : Ta thực hiện việc kết nối database như báo biểu 1 và tạo 1 “Line and stacked columns chart” .

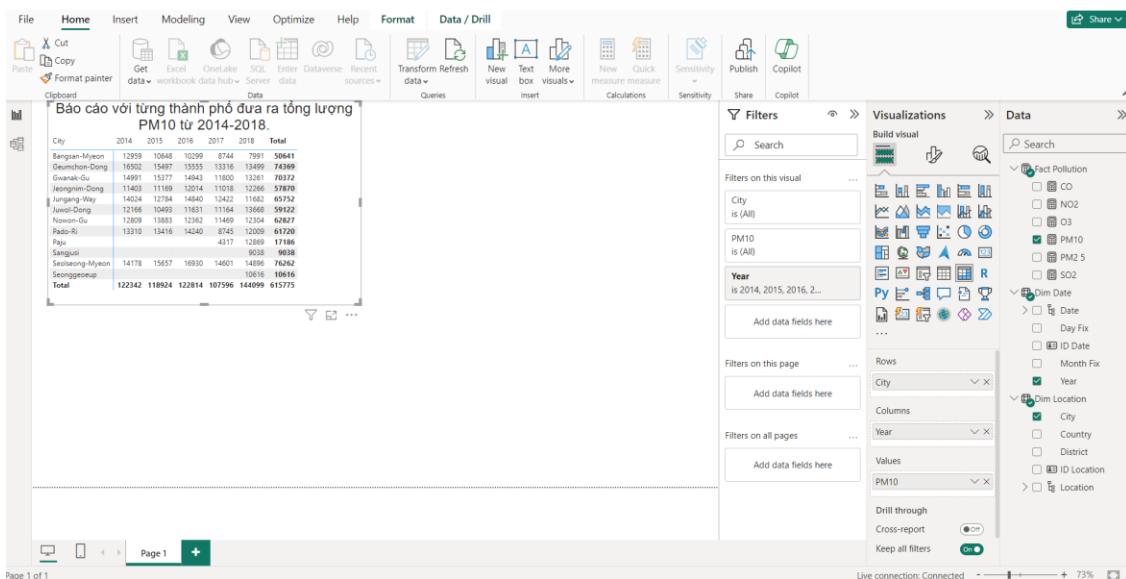


Bước 2 : Ta tạo thêm 1 “Pie chart” để hiển thị rõ hơn về thang nào chiếm lượng ô nhiễm NO2 cao nhất.

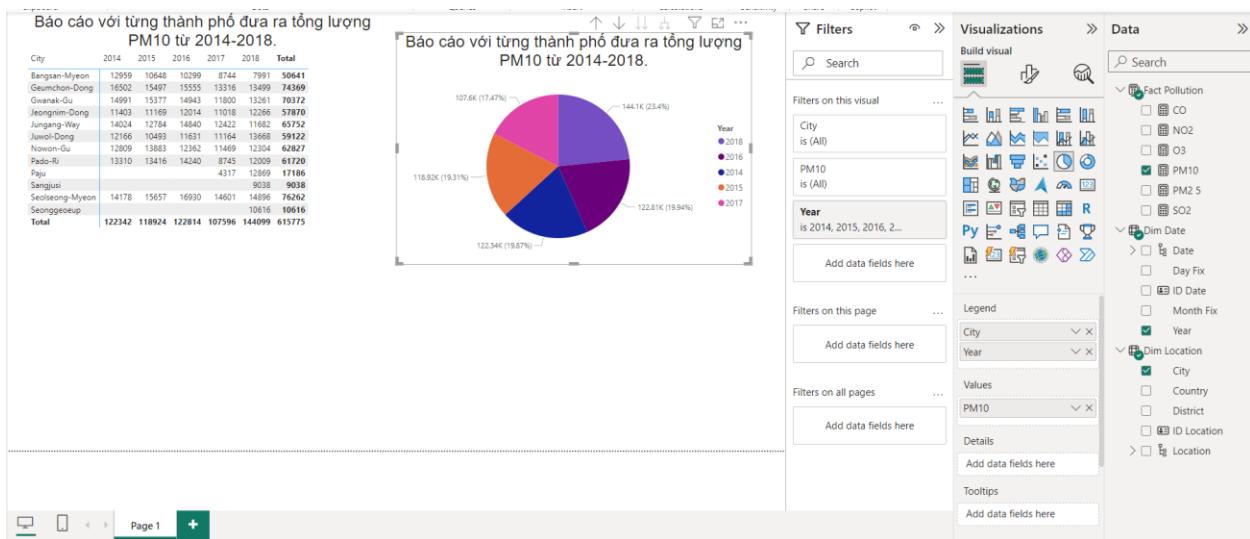


## 2.4. Báo cáo 3 : Báo cáo với từng thành phố đưa ra tổng lượng PM10 từ 2014-2018.

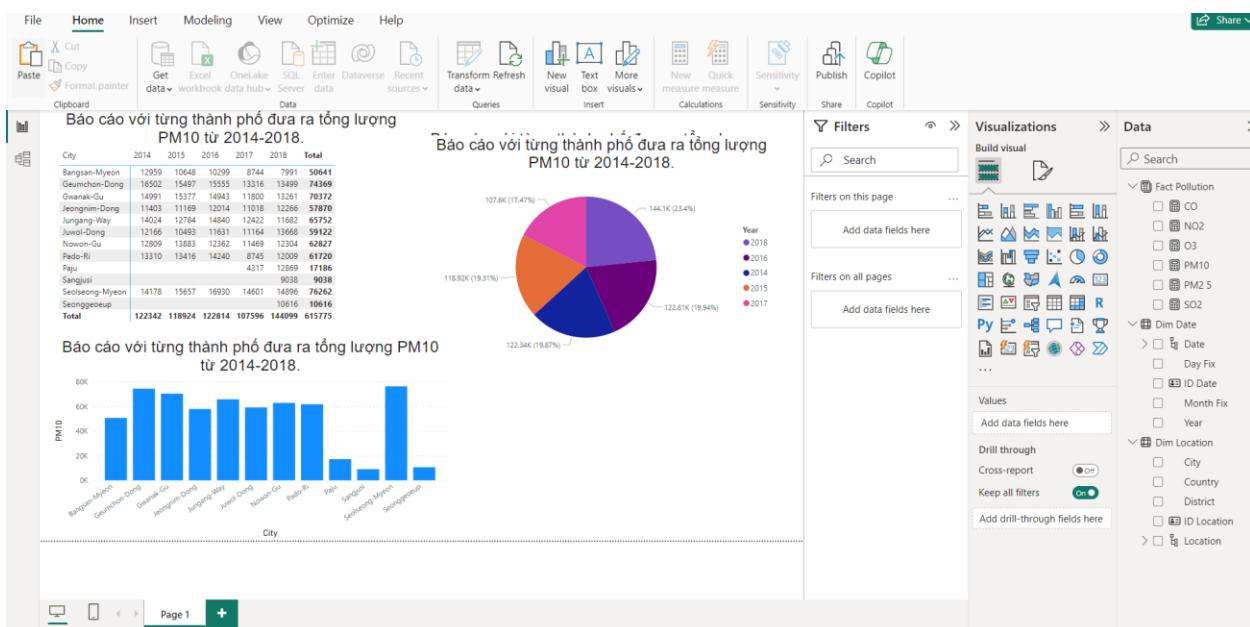
Bước 1 : Ta tạo bảng Matrix, để hiển thị chi tiết thông tin báo cáo .



Bước 2 : Ta bổ sung "Pie Chart" để có hiển thị chi tiết hơn



Bước 3 : Ta tạo lại "Stacked columns chart " để có thêm thông tin .



## V. Quá trình Datamining

### 1. Chọn lại kho dữ liệu Bike\_Buyer

#### 1.1. Tổng quan kho dữ liệu

Tập dữ liệu về người mua xe đạp có thông tin chi tiết về hơn 1000 người dùng từ các hoàn cảnh khác nhau và cho biết liệu họ có mua xe đạp hay không. Bộ dữ liệu này có thể được sử dụng cho các mô hình dự đoán bằng Thuật toán học máy. Bộ dữ liệu này bao gồm tuổi, giới tính, tình trạng hôn nhân, thu nhập hàng năm, người sở hữu nhà ở ...

Nguồn dữ liệu : <https://www.kaggle.com/datasets/heeraldedhia/bike-buyers/data>

#### 1.2. Mô tả dữ liệu

Dữ liệu bao gồm 13 cột và 1000 dòng

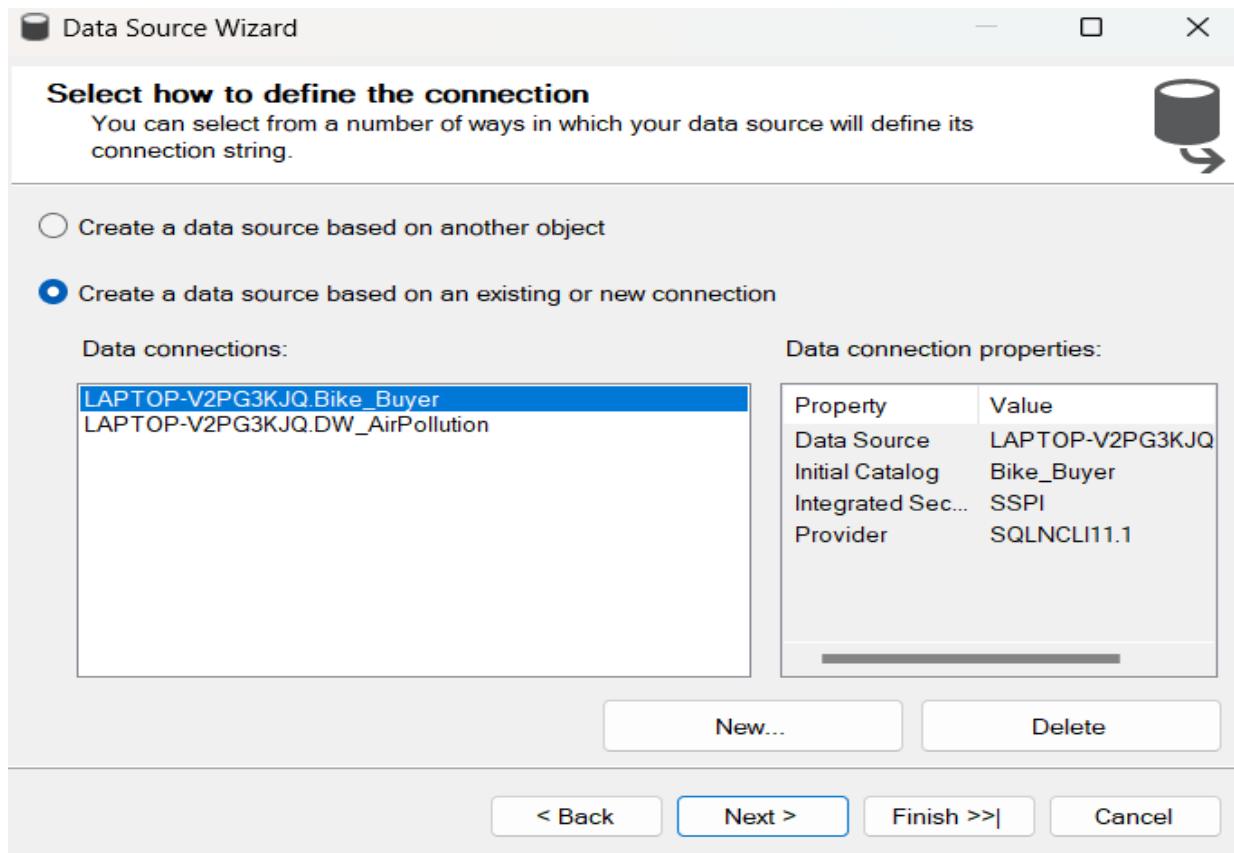
- Customer ID : mã người dùng
- Marital Status : tình trạng hôn nhân
- Gender : giới tính
- Income : thu nhập
- Children : trẻ em
- Education : giáo dục
- Occupation : nghề nghiệp
- Home Owner : chủ sở hữu nhà ở
- Cars : xe hơi
- Commute Distance : khoảng cách đi lại
- Region : vùng miền
- Age : tuổi tác
- Purchased Bike : đã mua xe đạp

### 2. Quá trình thực hiện Data Mining với thuật toán cây quyết định (Microsoft decision trees)

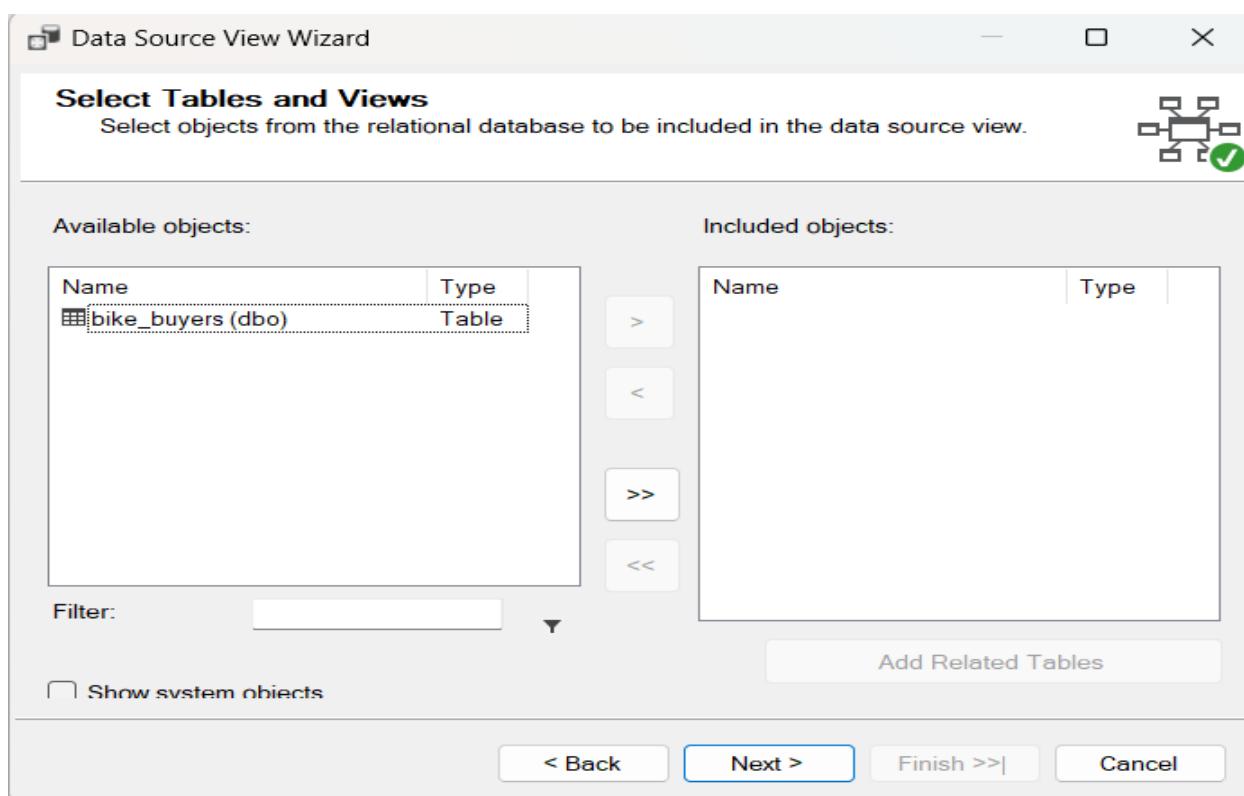
#### 2.1. Đưa dữ liệu vào kho (như SSAS)

#### 2.2. Thiết lập cấu trúc datamining

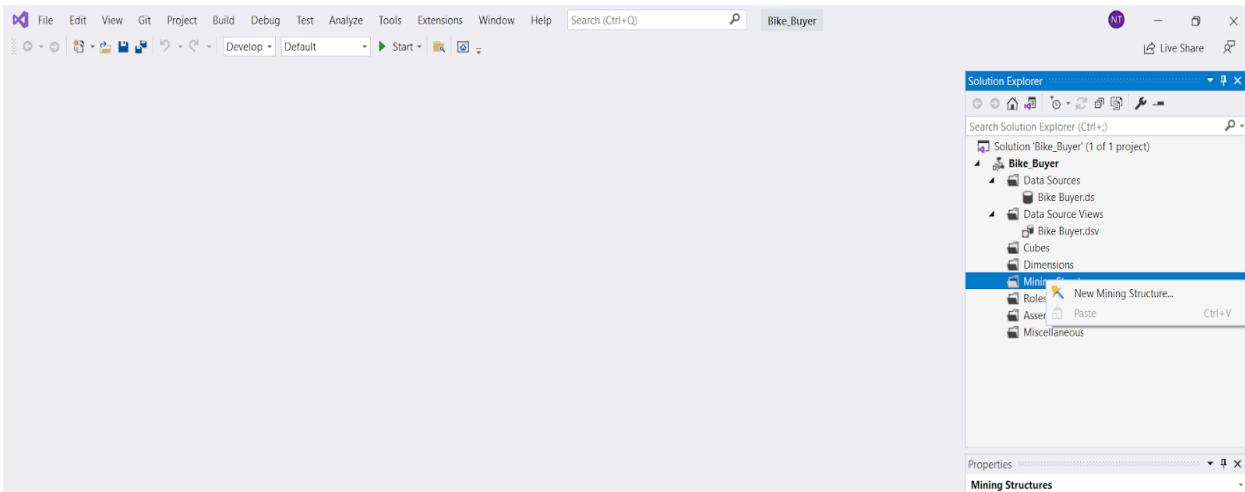
**Bước 1:** Add Datasource với database đã tạo ở phần trên



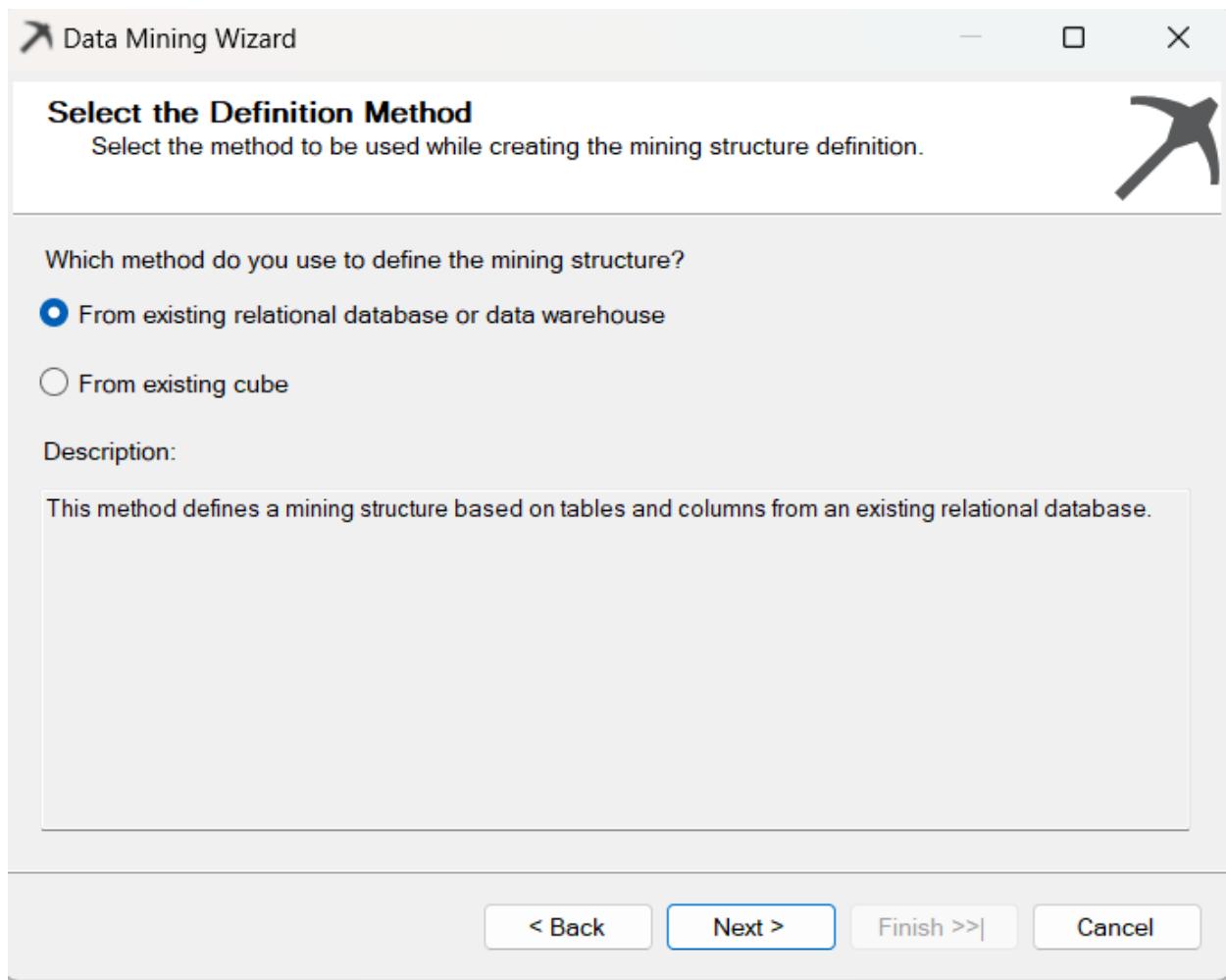
## Bước 2 : Add Datasource view với table Bike\_Buyer



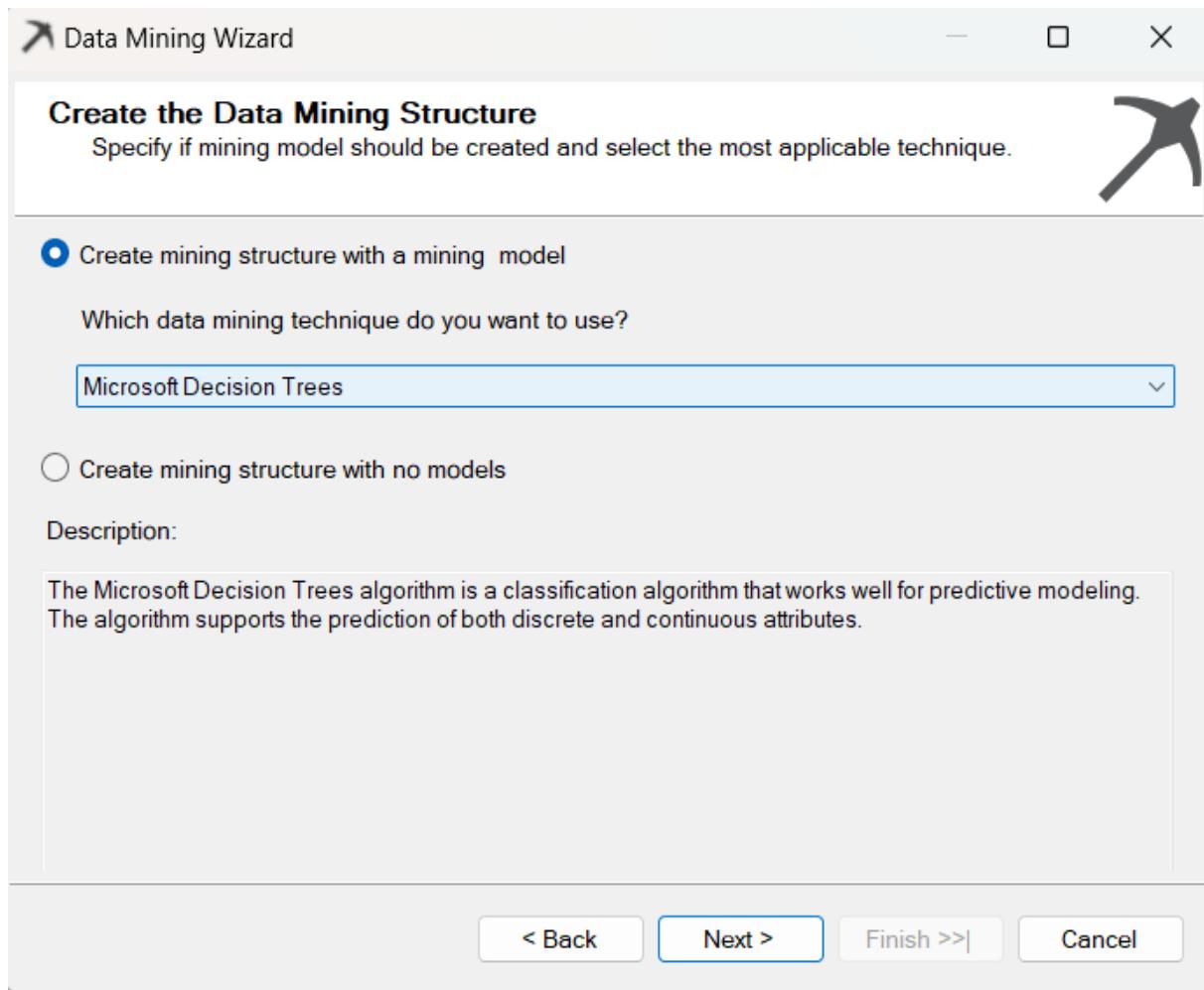
**Bước 3 :** Click chuột vào **Mining Structures** và chọn **New Mining Structure** để mở **Data Mining Wizard**.



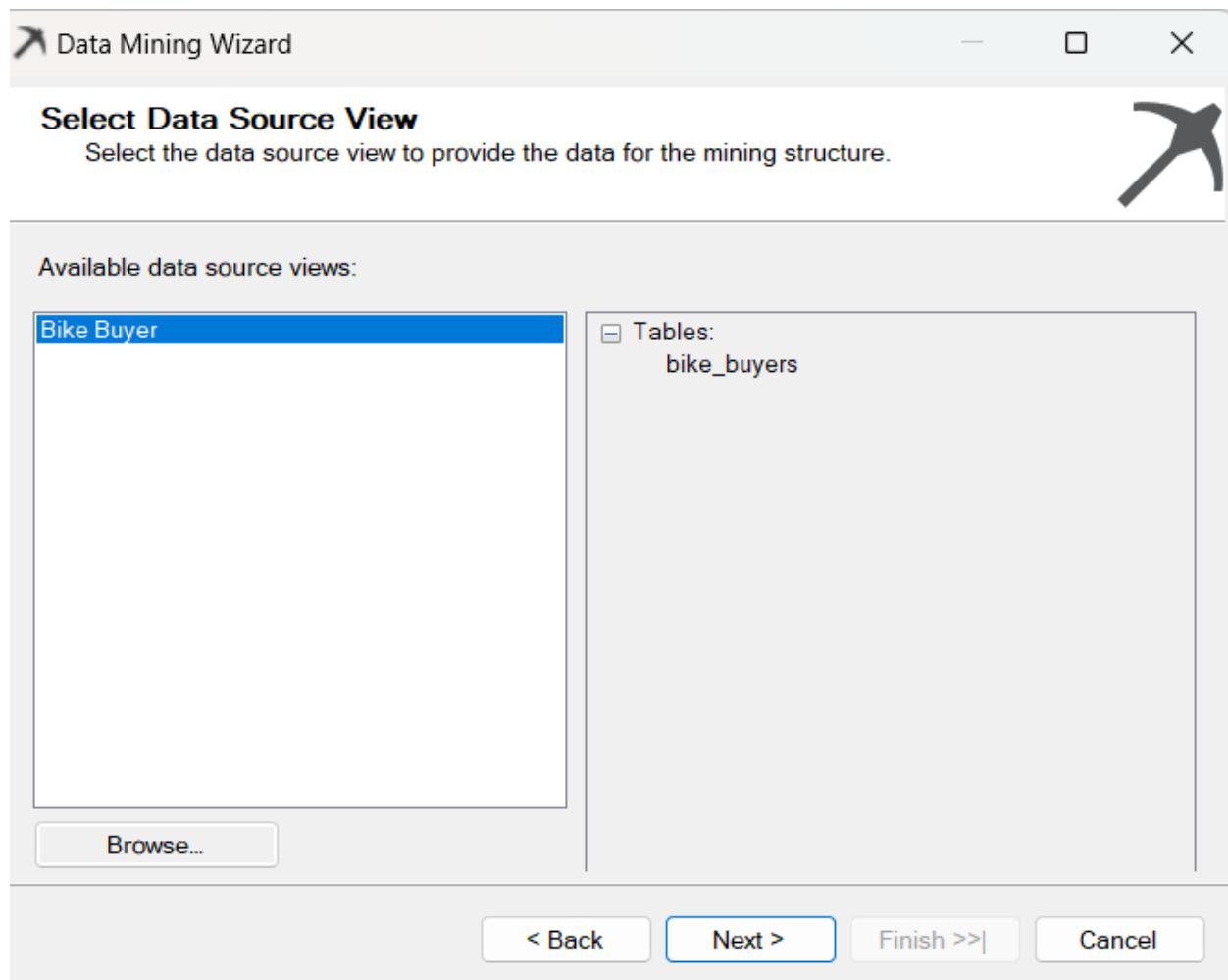
**Bước 4 :** Trên trang **Select the Definition Method** , chọn From existing relational database or data warehouse . Sau đó chọn **Next**.



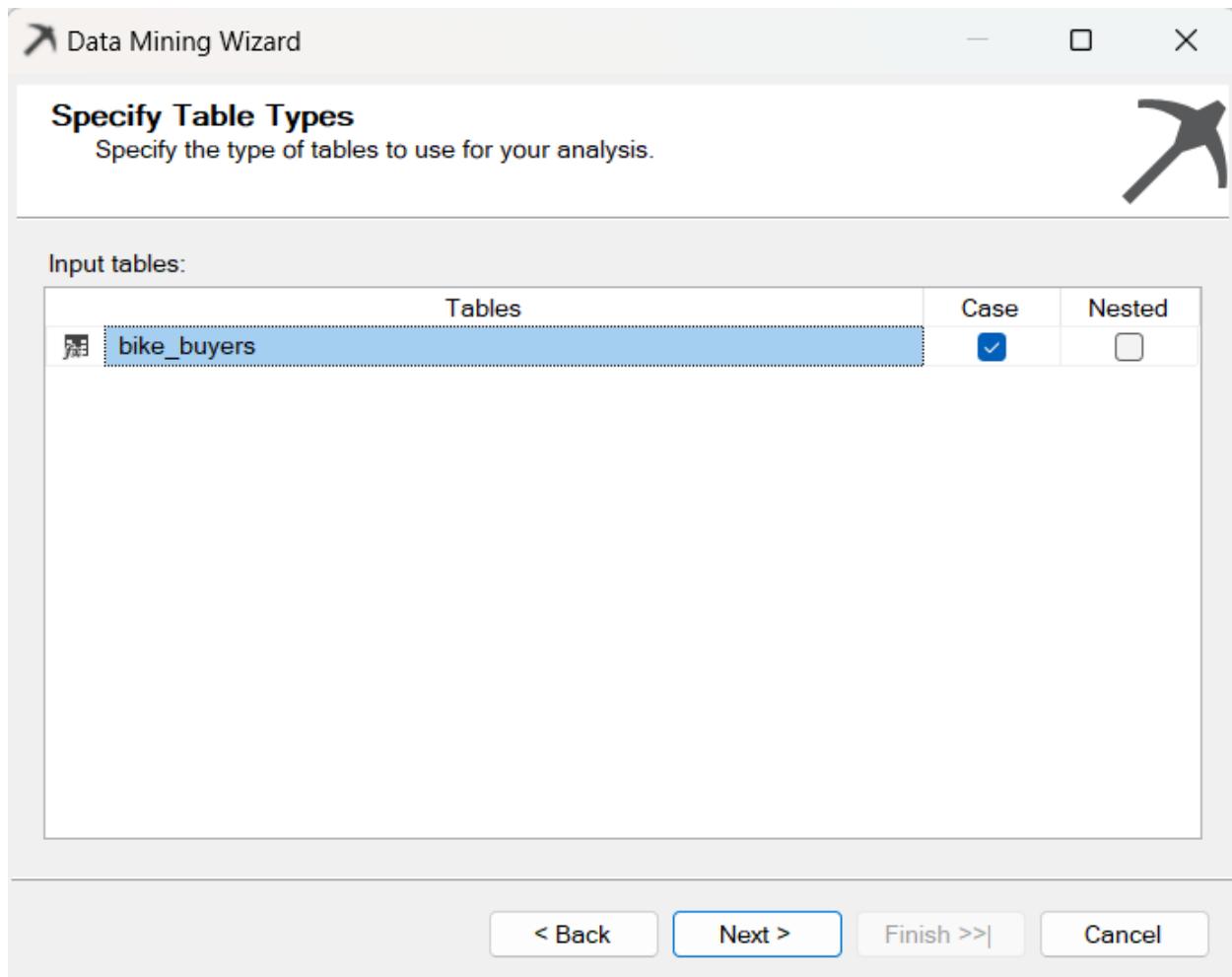
**Bước 5 :** Trên trang Create the Data Mining Structure chọn **Microsoft Decision Trees**. Sau đó, chọn Next.



**Bước 6 :** Trên trang **Select the Data Source View**, chọn Next.



**Bước 7 :** Trên trang **Specify Table Types**, chọn Next.



**Bước 8 :** Trên trang **Specify the Training Data**, cột key để mặc định , chọn một trường Purchased\_Bike để dự đoán số lượng đã mua xe đạp , chọn **Suggest**.

 Data Mining Wizard

### Specify the Training Data

Specify the columns used in your analysis.

Mining model structure:

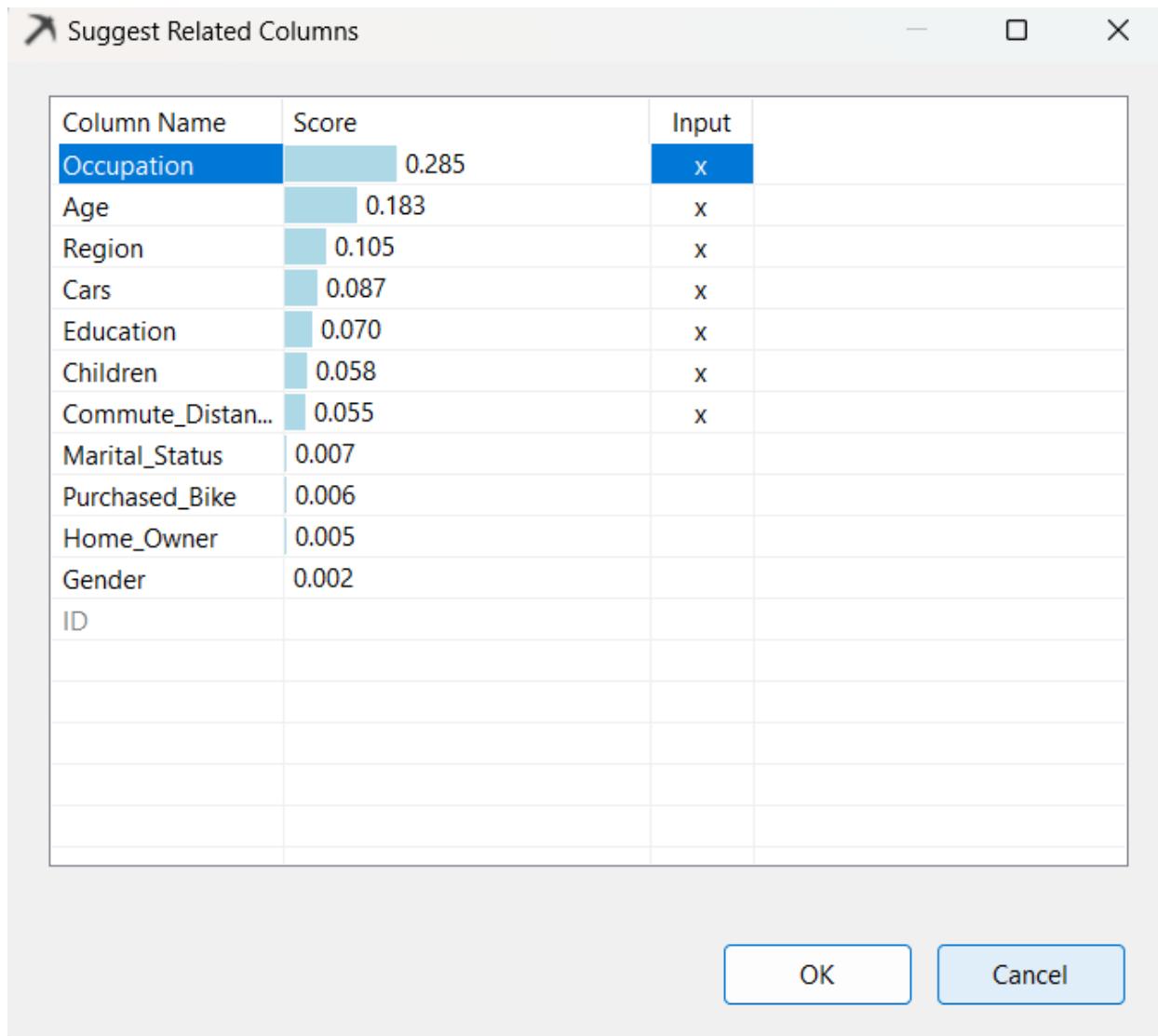
	Tables/Columns	Key	<input type="checkbox"/> Input	<input checked="" type="checkbox"/> Predi...
<input checked="" type="checkbox"/>	ID	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Income	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Marital_Status	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Occupation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Purchased_Bike	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Region	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Recommend inputs for currently selected predictable: [Suggest](#)

 No input column is defined for this mining model.

[<< Back](#) [Next >](#) [Finish >>](#) [Cancel](#)

**Bước 9 :** Suggest đưa ra các input gợi ý dựa trên Income đã chọn, ở đây ta không muốn dùng các input được gợi ý sẵn thì chọn **Cancel**.



**Bước 10 :** Chọn các cột input mong muốn , ở đây ta chọn Age, Marital\_Status, Commute\_Distance và chọn Next.

Data Mining Wizard

### Specify the Training Data

Specify the columns used in your analysis.

Mining model structure:

	Tables/Columns	Key	Input	Predi...
<input checked="" type="checkbox"/>	ID	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Income	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Marital_Status	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Occupation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Purchased_Bike	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Region	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Recommend inputs for currently selected predictable: [Suggest](#)

⚠ The mining model does not have a key column. Choose a column to use as the key.

< Back [Next >](#) [Finish >>](#) Cancel

**Bước 11:** Trên trang **Specify Column's Content and Data Type**, điều chỉnh lại kiểu dữ liệu và kiểu nội dung, có thể dùng **Detect** để xác định. Sau đó chọn **Next**.

Data Mining Wizard

### Specify Columns' Content and Data Type

Specify mining structure columns' content and data type.

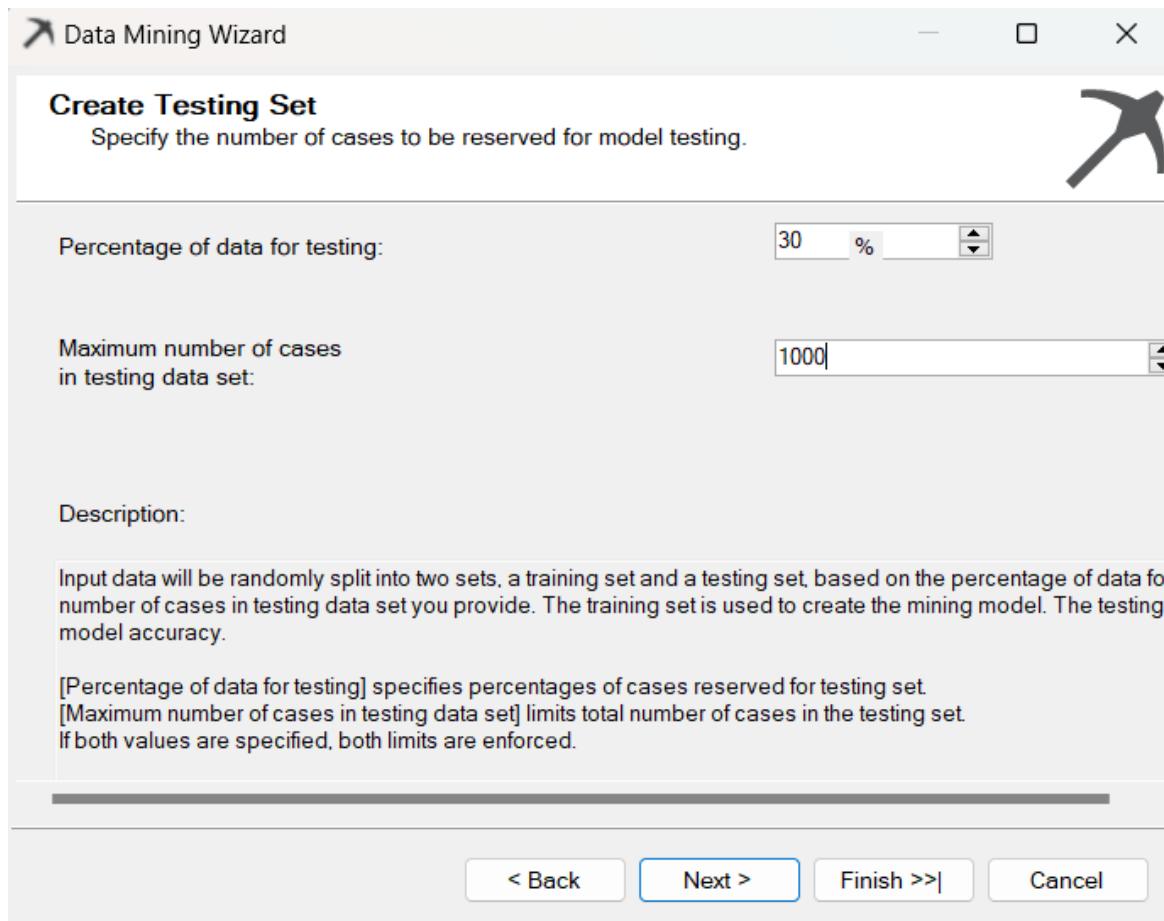
Mining model structure:

	Columns	Content Type	Data Type
Age		Continuous	Double
Commute Distance		Discrete	Text
ID		Key	Long
Marital Status		Discrete	Text
Purchased Bike		Discrete	Text

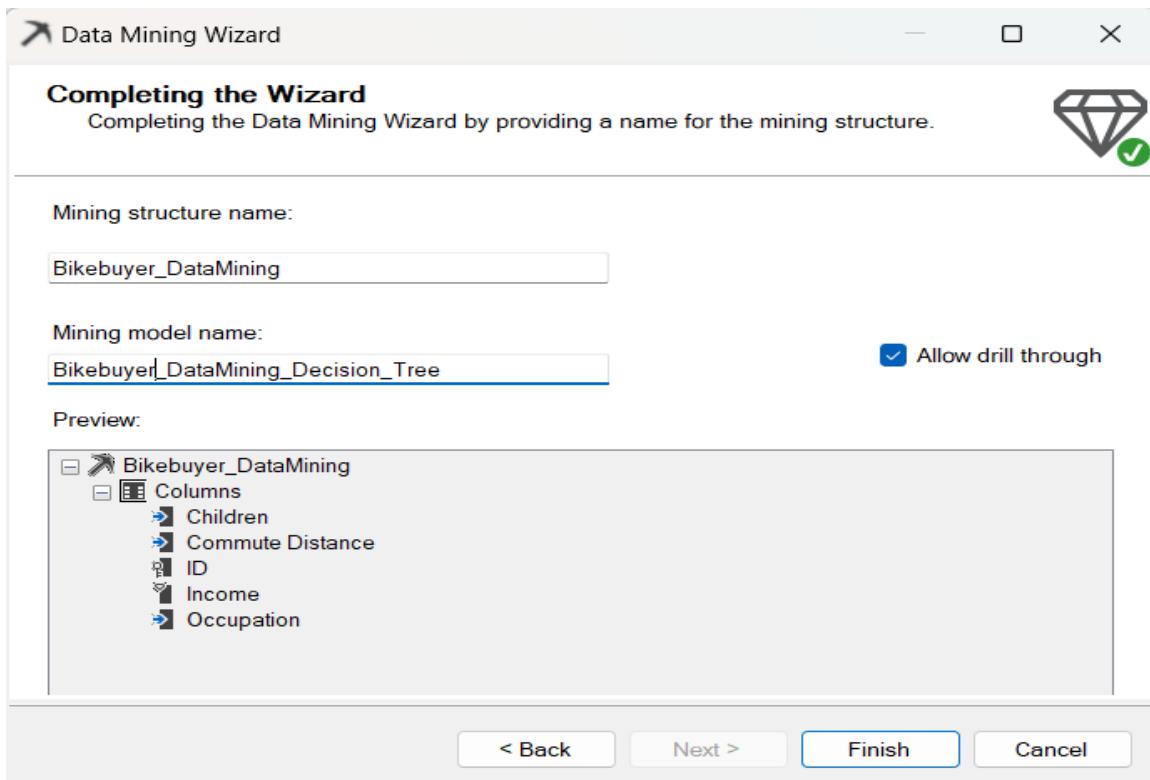
Detect continuous or discrete for numeric columns: Detect

< Back Next > Finish >> Cancel

**Bước 12 :** Trên trang **Create Testing Set**, Percentage of data for testing mặc định là 30% , Maximum number of cases in testing data để 1000. Tiếp tục chọn Next.



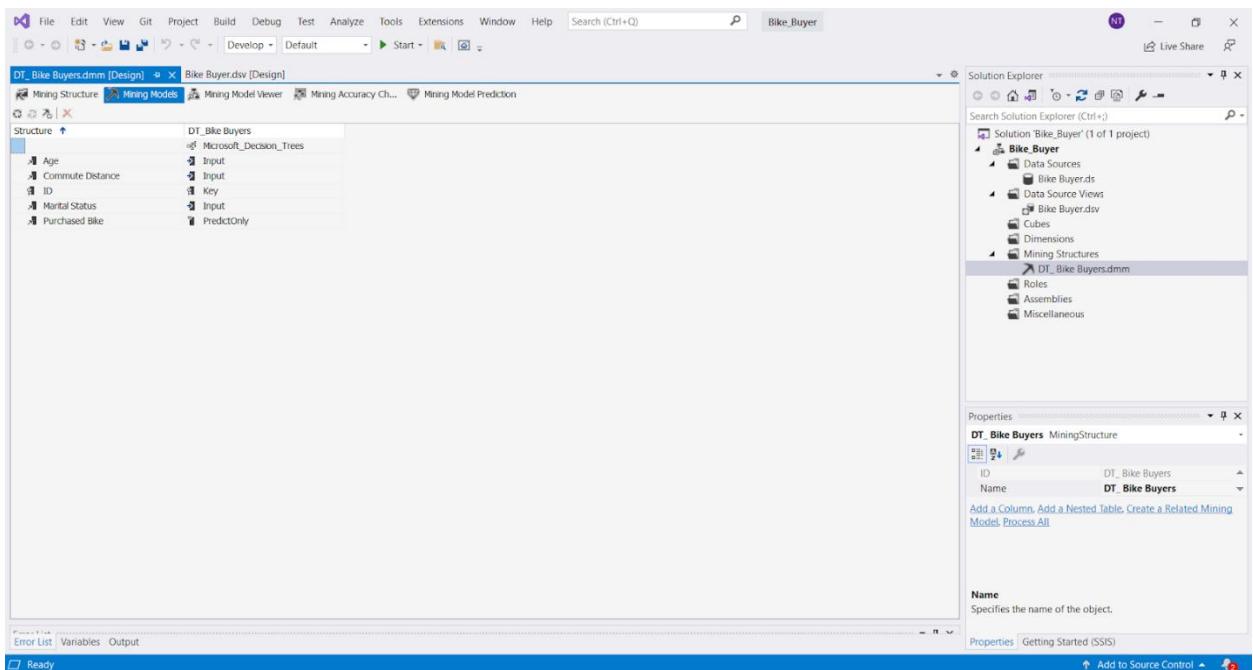
**Bước 13 :** Trên trang **Completing the Wizard**, có thể đổi tên model, click chọn vào ô **Allow drill through**. Chọn **Finish** để kết thúc quá trình thiết lập.



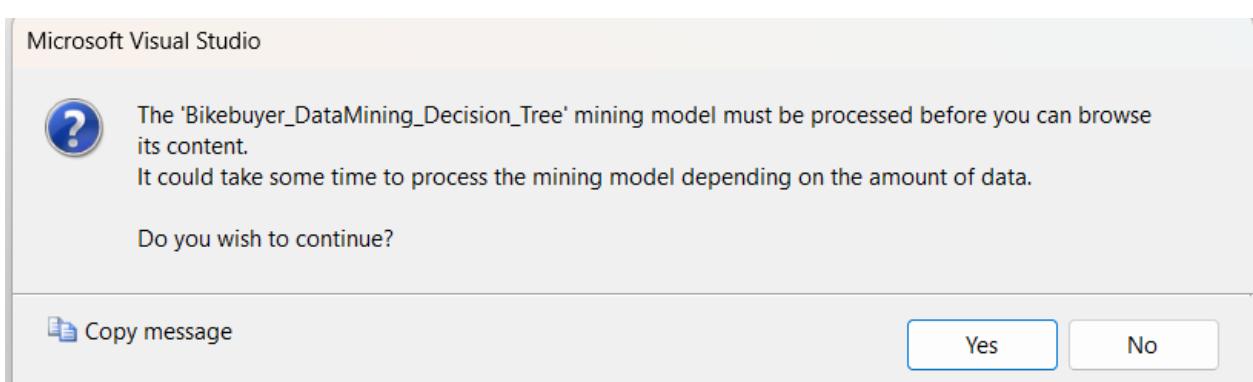
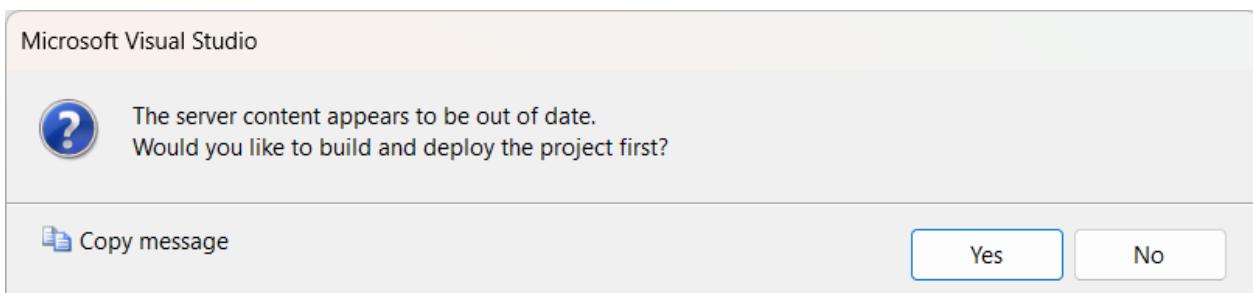
#### Bước 14 : Kết quả của việc thiết lập cấu trúc cho Data Mining.

### 2.3. Thực hiện Deploy project data minning

**Bước 1 :** Chuyển sang tab **Mining Models** để xem lại thiết lập cấu trúc data mining.



**Bước 2 :** Tiến hành deploy project , chọn Yes để tiếp tục.



**Bước 3 :** Tiếp tục chọn Run.

Process Mining Model - Bikebuyer\_DataMining\_Decision\_Tree

Object list:

Object Name	Type	Process Options	Settings
Bikebuyer_DataMining_De...	Mining Model	Process Full	

Remove Impact Analysis...

Batch Settings Summary

Processing order:  
Parallel

Transaction mode:  
(Default)

Dimension errors:  
(Default)

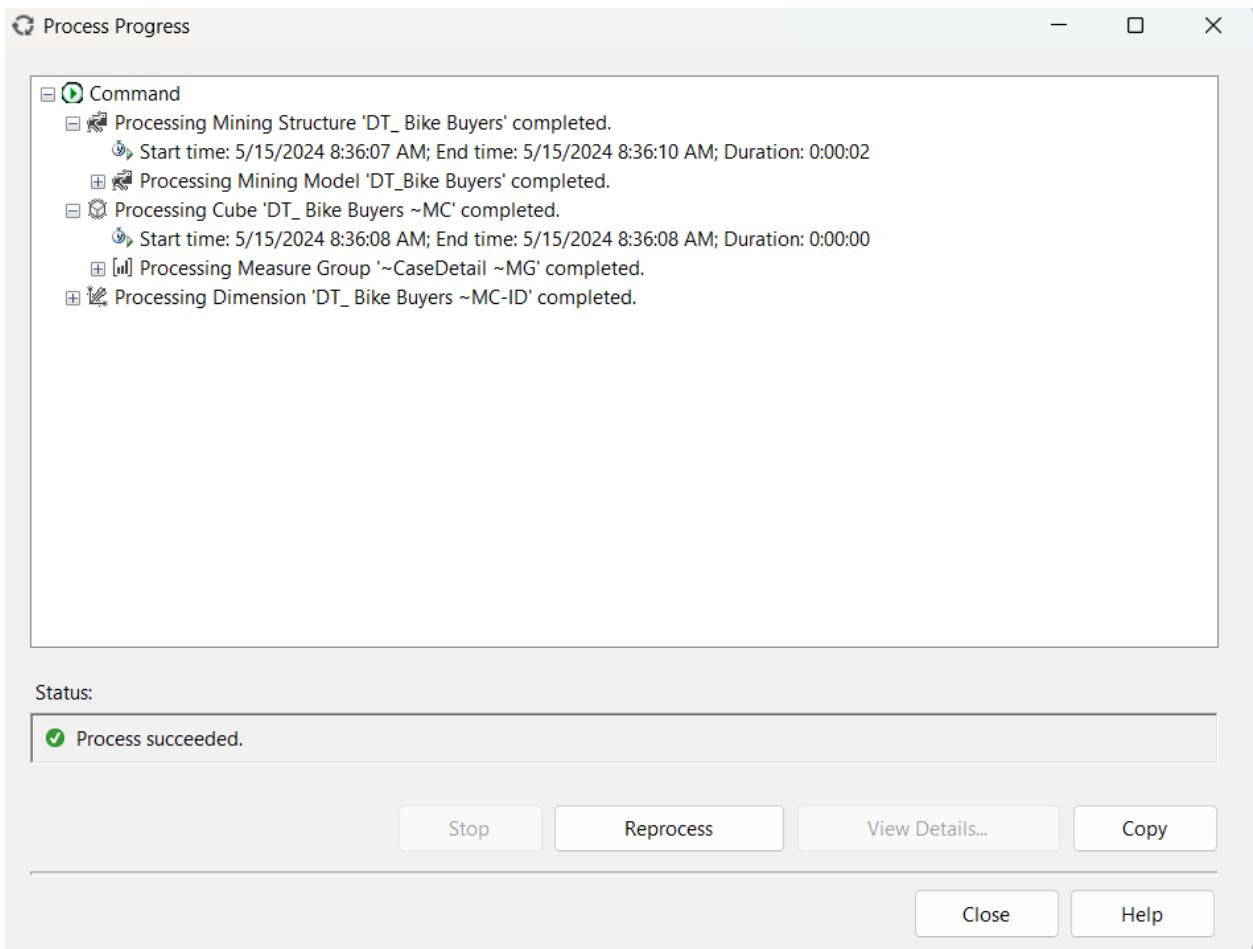
Dimension key error log path :  
(Default)

Process affected objects:  
Do not process

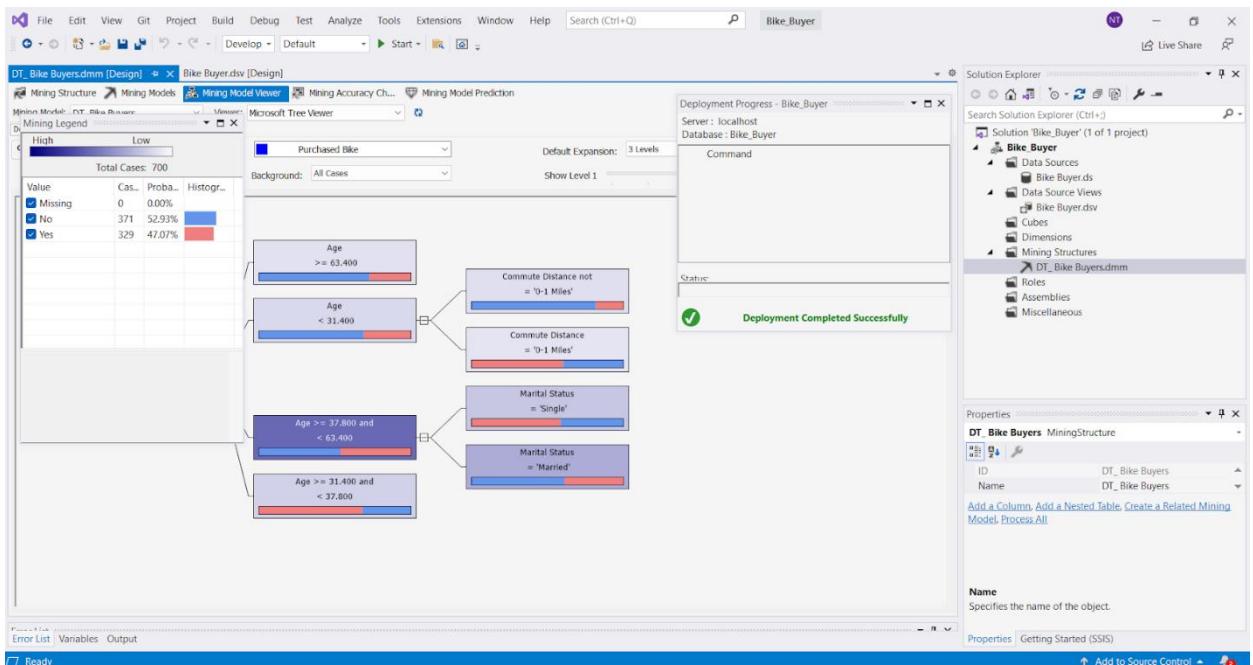
Change Settings...

Run... Close

**Bước 4 :** Sau khi tiến hành mining thành công , chọn **Close**.



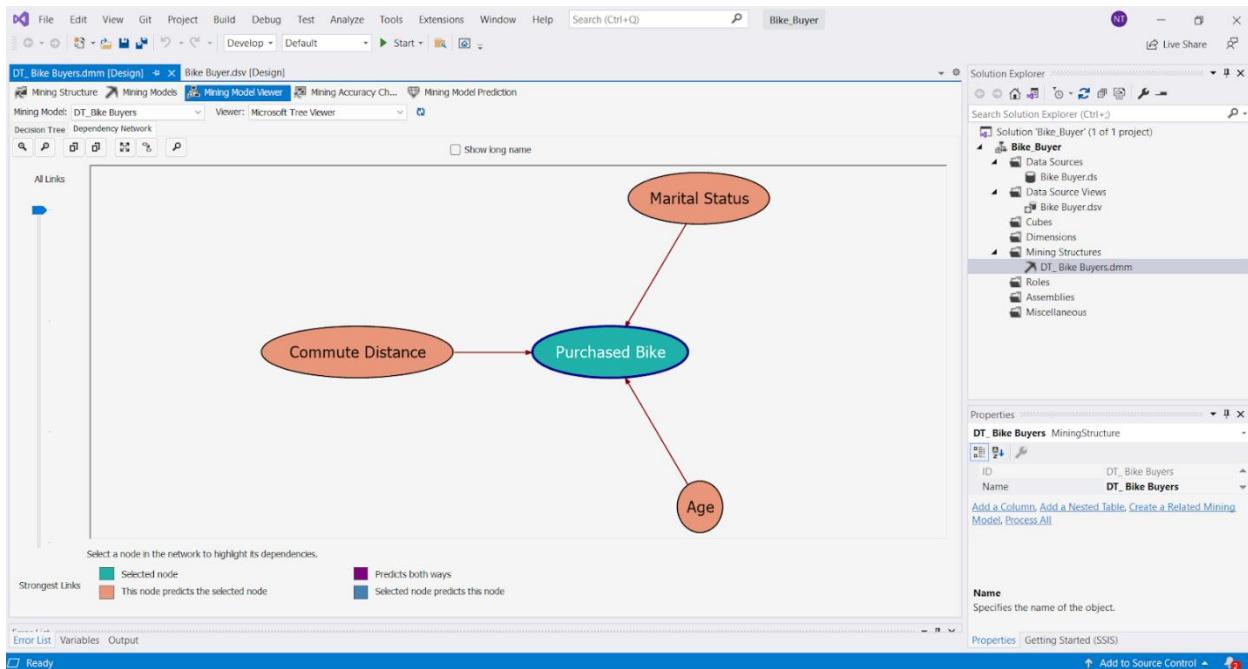
## Bước 5 : Kết quả có được mô hình cây quyết định (Decision Trees model).



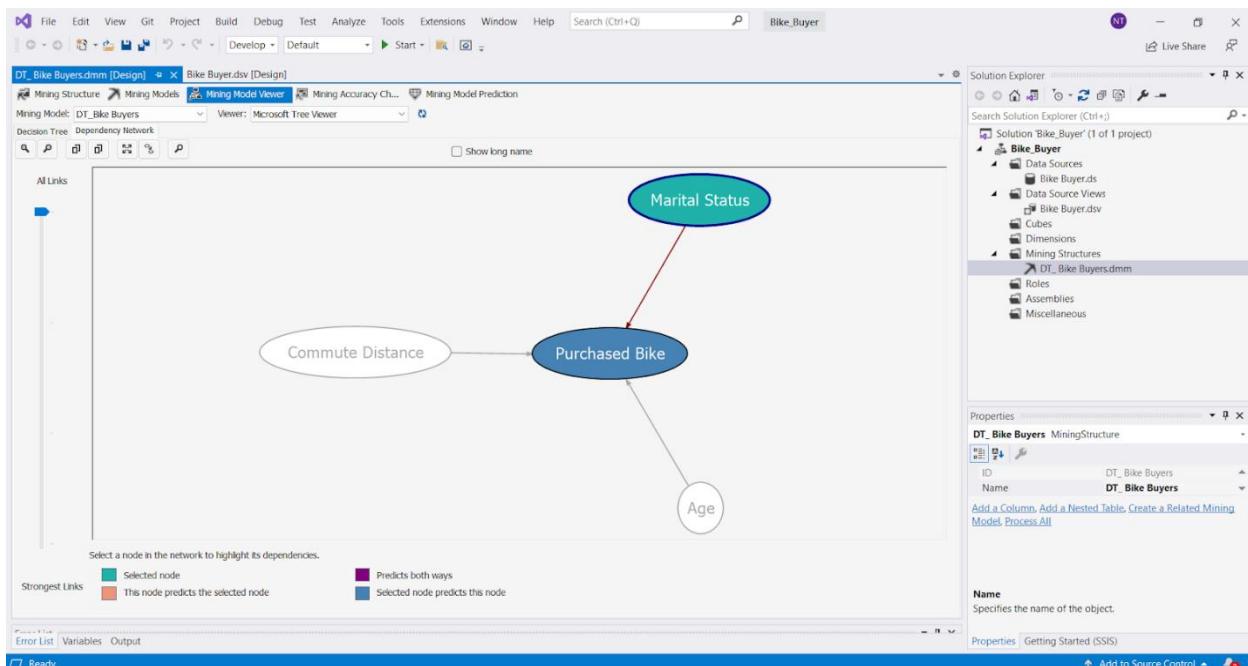
## 2.4. Phân tích mining bằng cây quyết định và đưa ra tập luật

**Bước 1 :** Sau khi đã cài đặt thành công, tiếp theo ta tiếp tục quá trình mining, vào tab **Dependency Network** để xem các cấp phụ thuộc

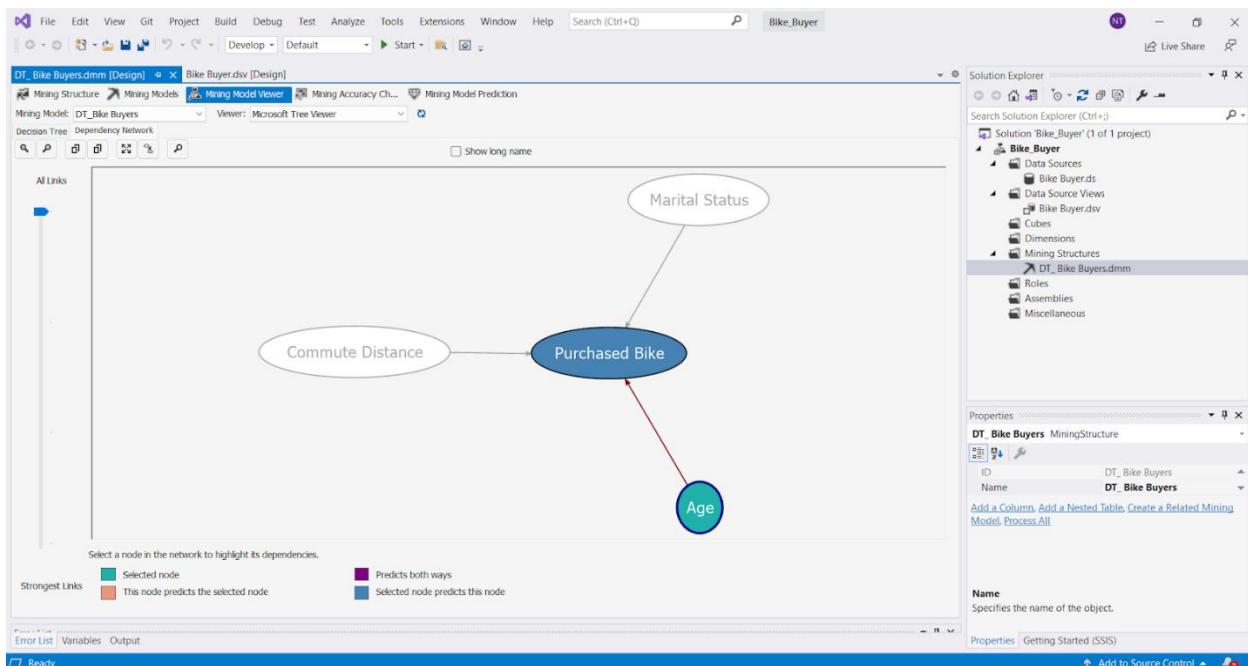
- Mức phụ thuộc cấp 3



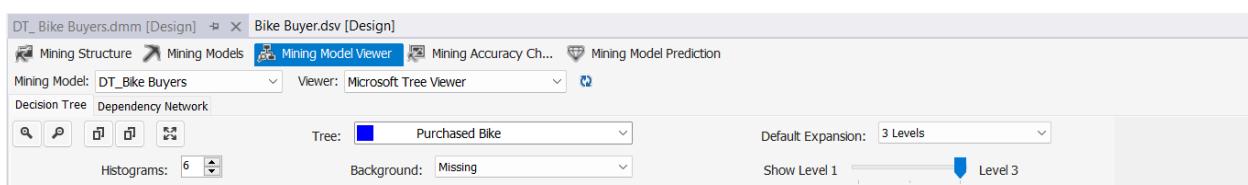
- Mức phụ thuộc cấp 2



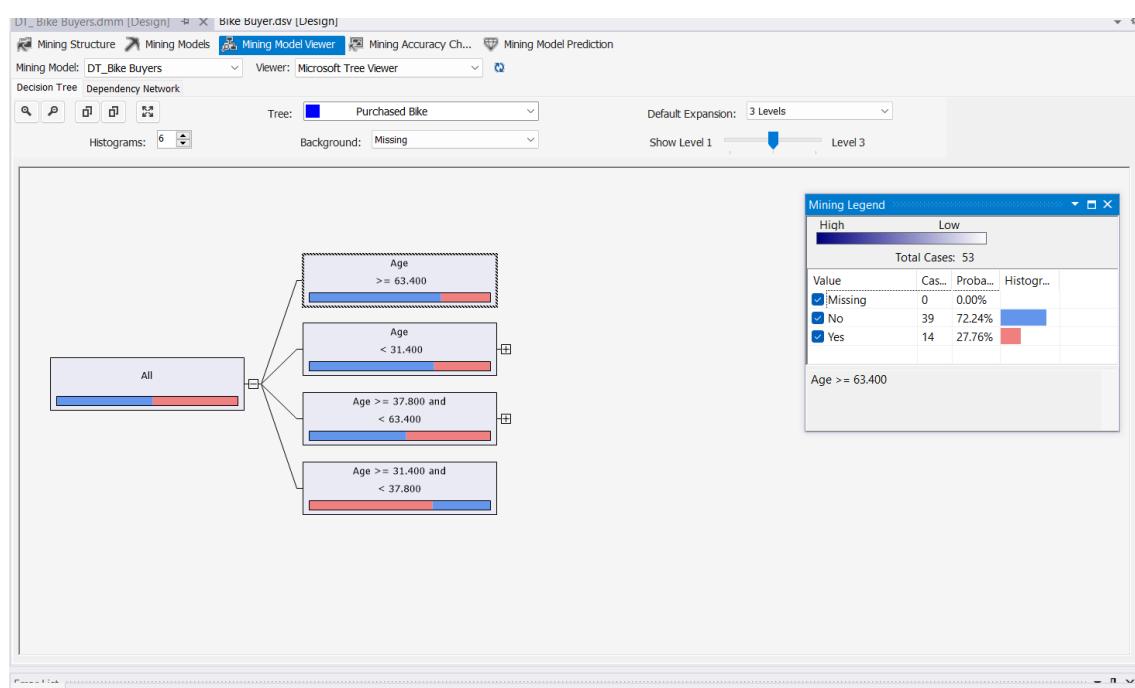
- Mức phụ thuộc cấp 1



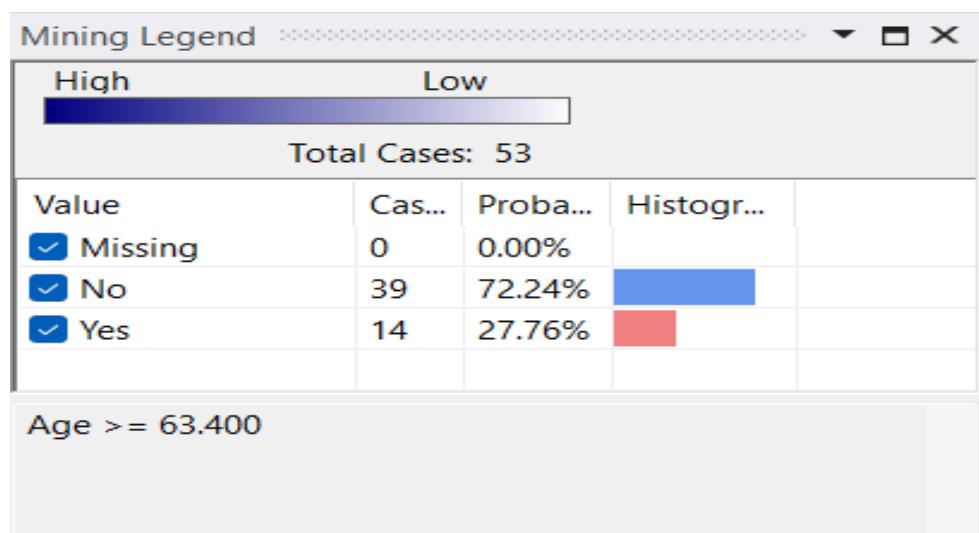
## Bước 2 : Chọn background Missing để dự đoán



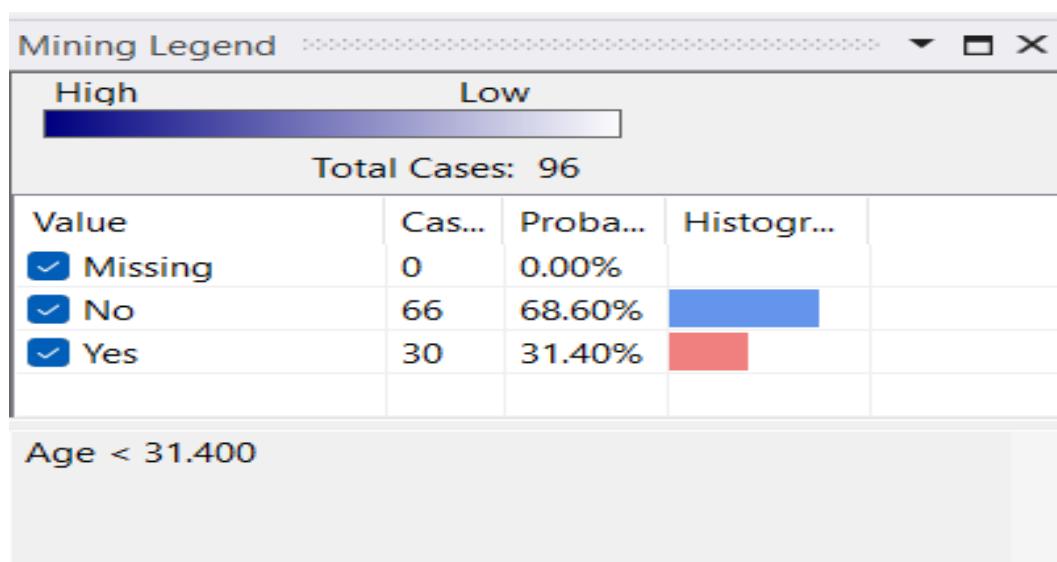
- Ở mức 2 dự đoán về tuổi ( Age )



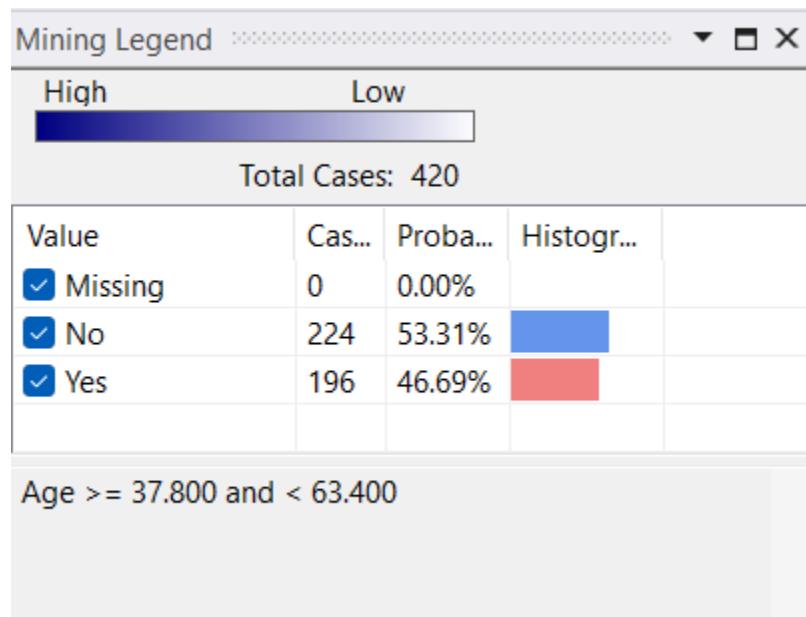
- Độ tuổi  $\geq 63,4$  đã mua xe đẹp là No chiếm 72,24%, là Yes chiếm 27.76%



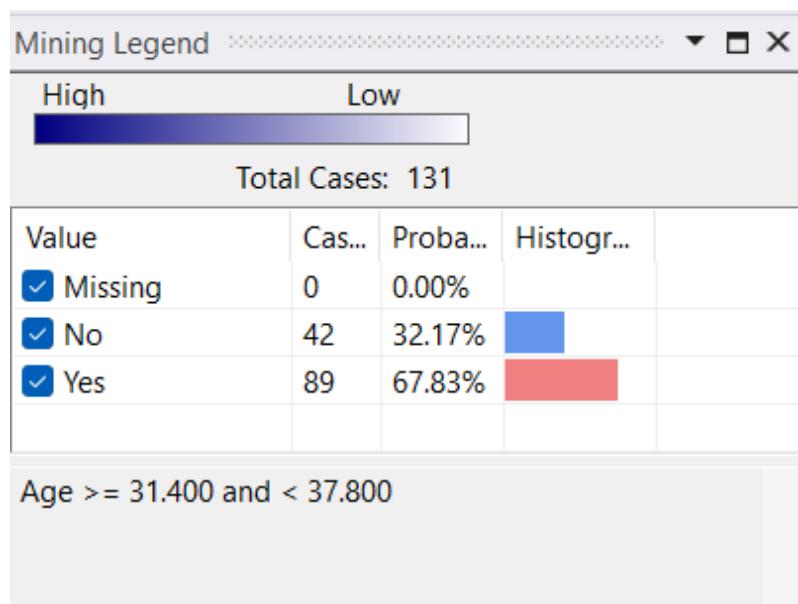
- Độ tuổi  $< 31,4$  đã mua xe đẹp là No chiếm 68,6% , là Yes chiếm 31,4%



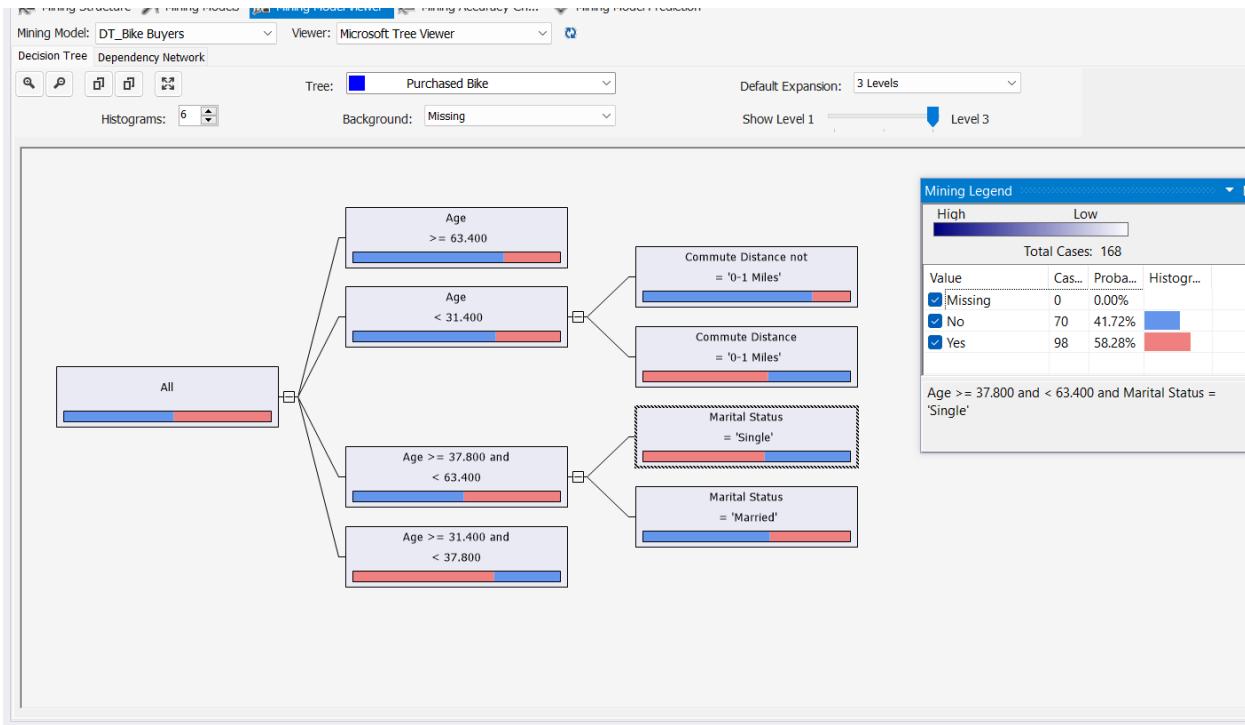
- Độ tuổi  $< 37,8$  và  $> 63,4$  đã mua xe đẹp là No chiếm 53,51% , là Yes chiếm 46,69%



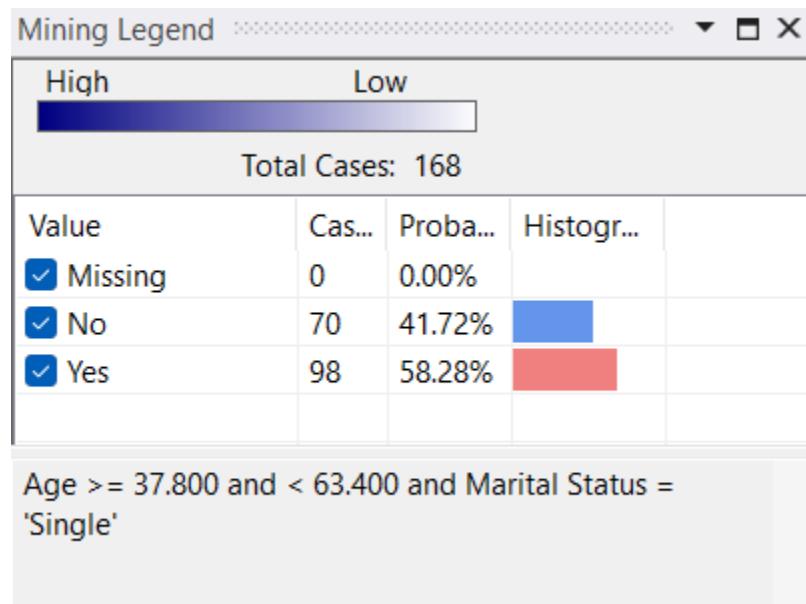
- Độ tuổi < 31,4 và > 37,8 đã mua xe đẹp là No chiếm 32,17%, là Yes chiếm 67,83%



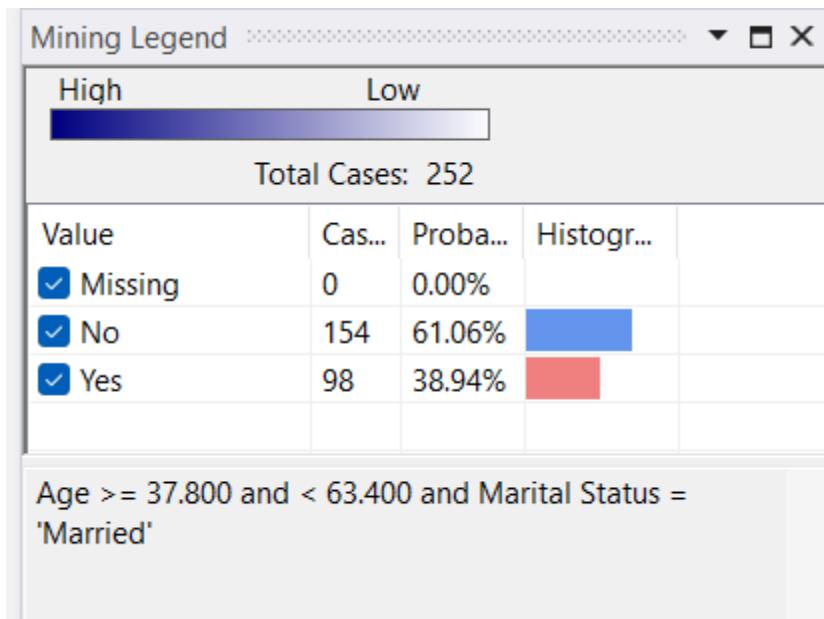
Ở mức 3 ta dự đoán về tình trạng hôn nhân (Marital Status)



- Tình trạng hôn nhân người  $\geq 37,8$  và  $< 63,4$  tuổi độc thân là No chiếm 41.72% , là Yes chiếm 58.28%



- Tình trạng hôn nhân người  $\geq 37,8$  và  $< 63,4$  tuổi đã kết hôn là No chiếm 61.06% , là Yes chiếm 38.94%

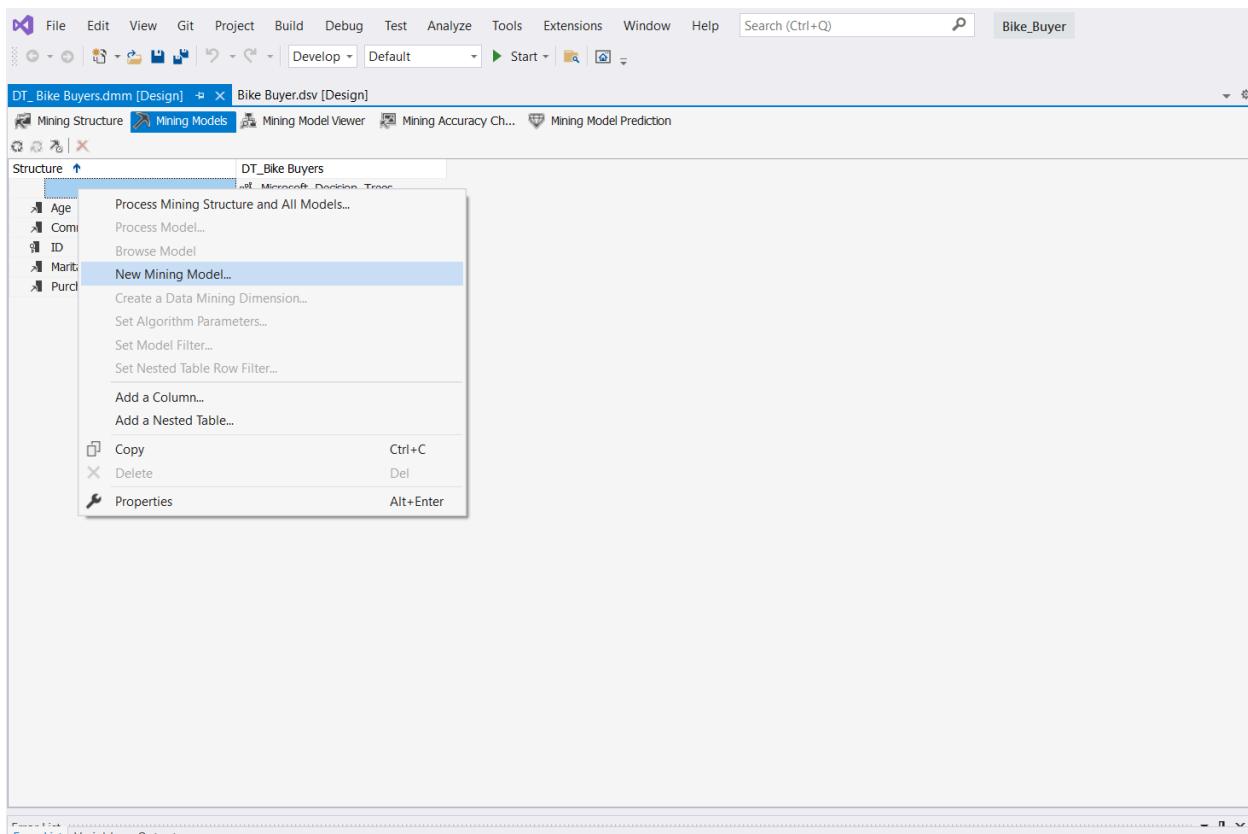


Vì phần trăm tình trạng hôn nhân của người  $\geq 37,8$  và  $< 63,3$  tuổi đã kết hôn là cao hơn nên ta chọn tình trạng hôn nhân là đã kết hôn ở mức 3.

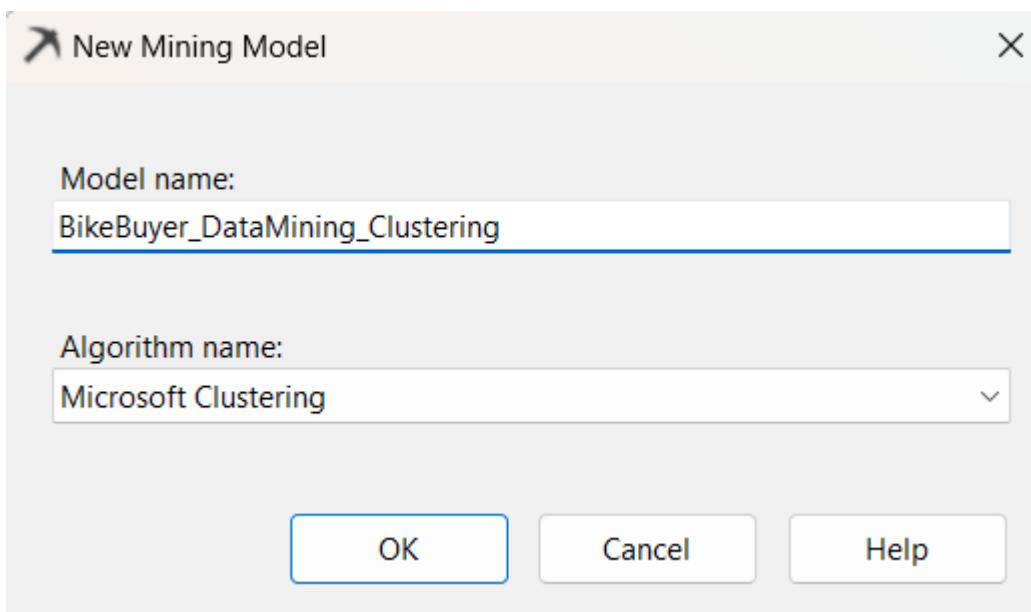
**Rút ra tập luật :** Các trường hợp người từ 38-63 tuổi mua xe đẹp , mà tình trạng hôn nhân của họ là đã kết hôn ( Marital Status = 'Married' ) thì có khả năng người mua xe đẹp và đã kết hôn là rất hạnh phúc .

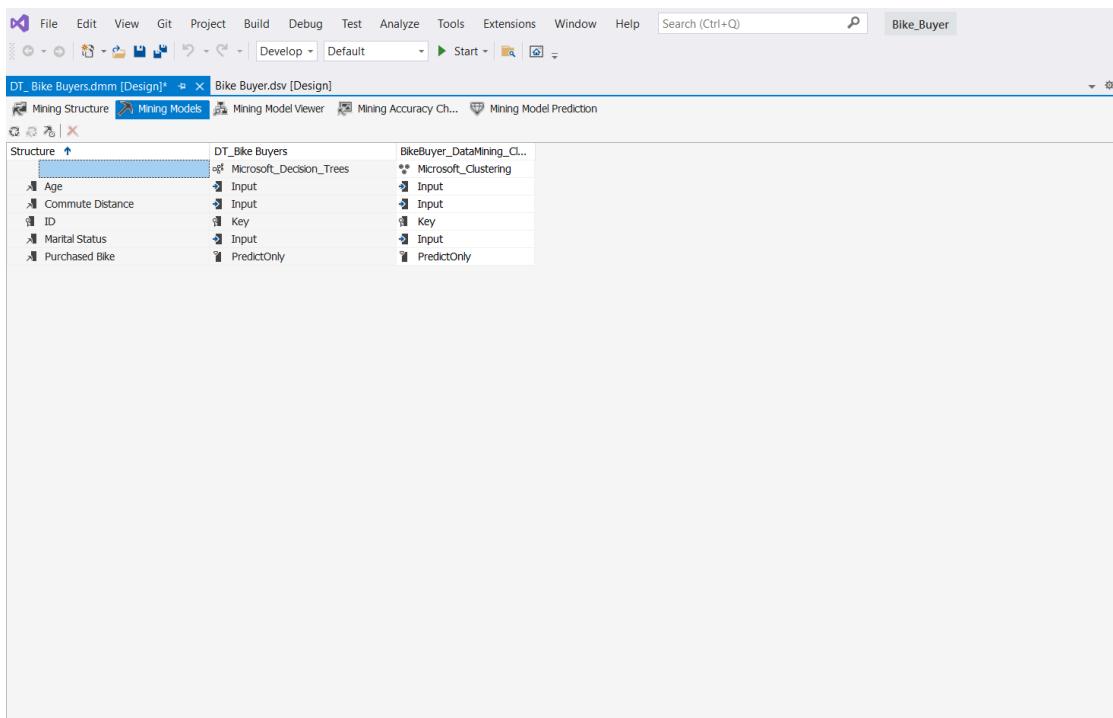
## 2.5. So sánh với các thuật toán khác

**Bước 1 :** Click chuột phải vào ô trống chọn **New Mining Model**



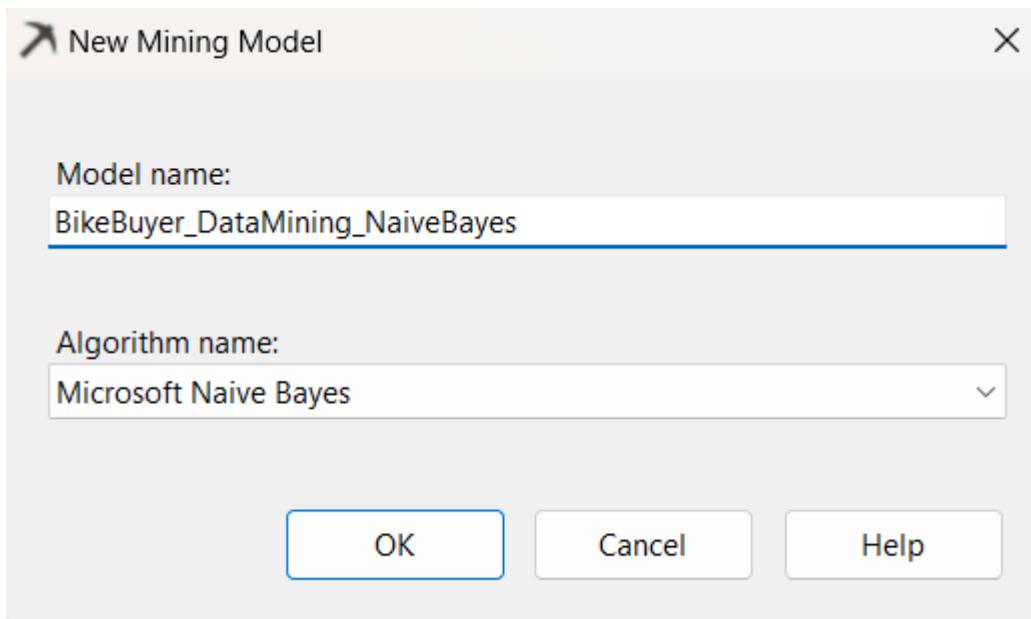
## Bước 2 : Chọn thuật toán Microsoft Clustering





### Bước 3 : Tiếp tục tạo thêm thuật toán **Mining Microsoft Naive Bayes**

- Thuật toán này không hỗ trợ mining các cột có kiểu continuous

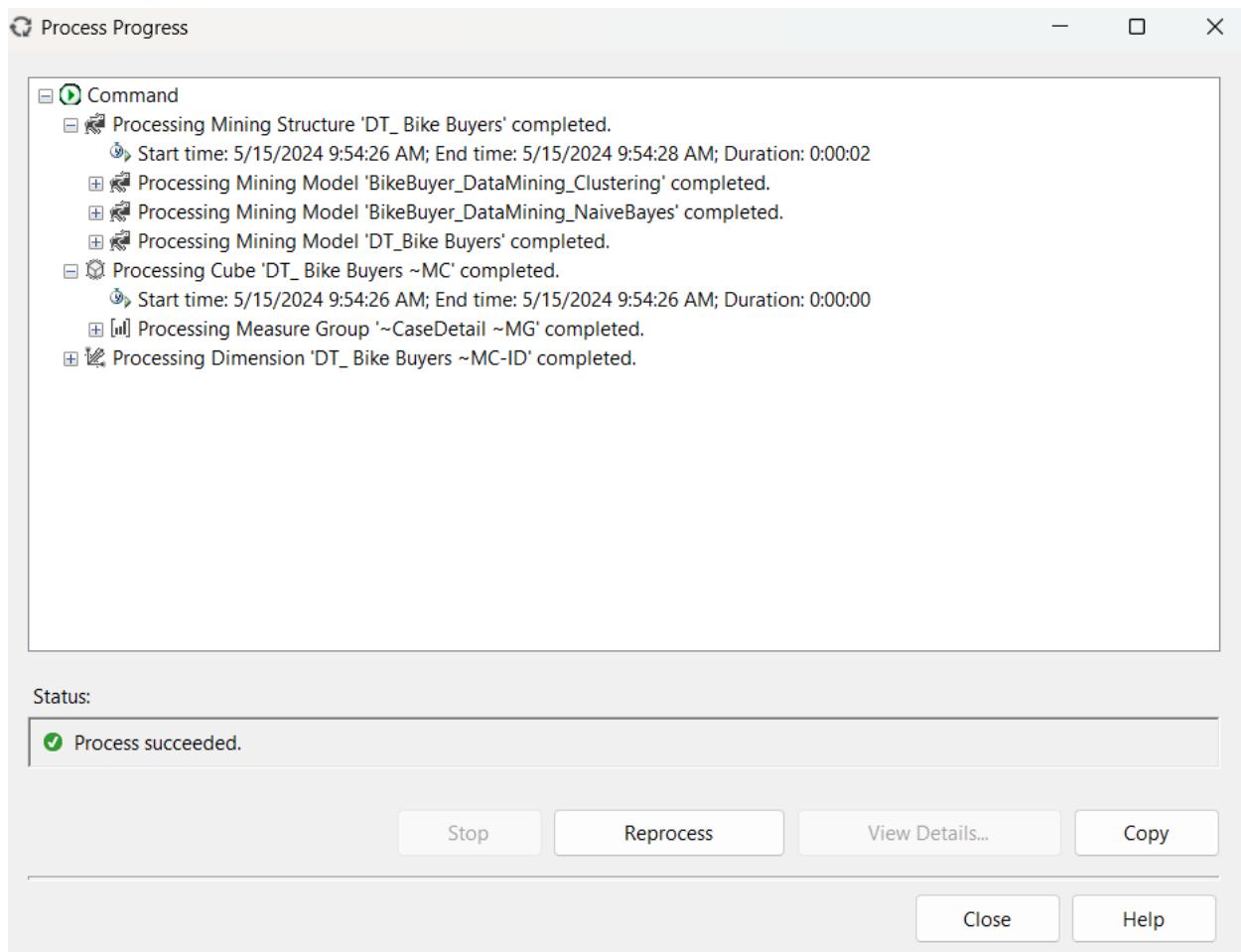


The screenshot shows the Microsoft Data Mining Model Designer window. The title bar displays "DT\_Bike Buyers.dmm [Design]" and "Bike Buyer.csv [Design]". The ribbon menu has tabs for "Mining Structure", "Mining Models", "Mining Model Viewer", "Mining Accuracy Ch...", and "Mining Model Prediction". The main area is titled "Structure" and contains a table with four columns:

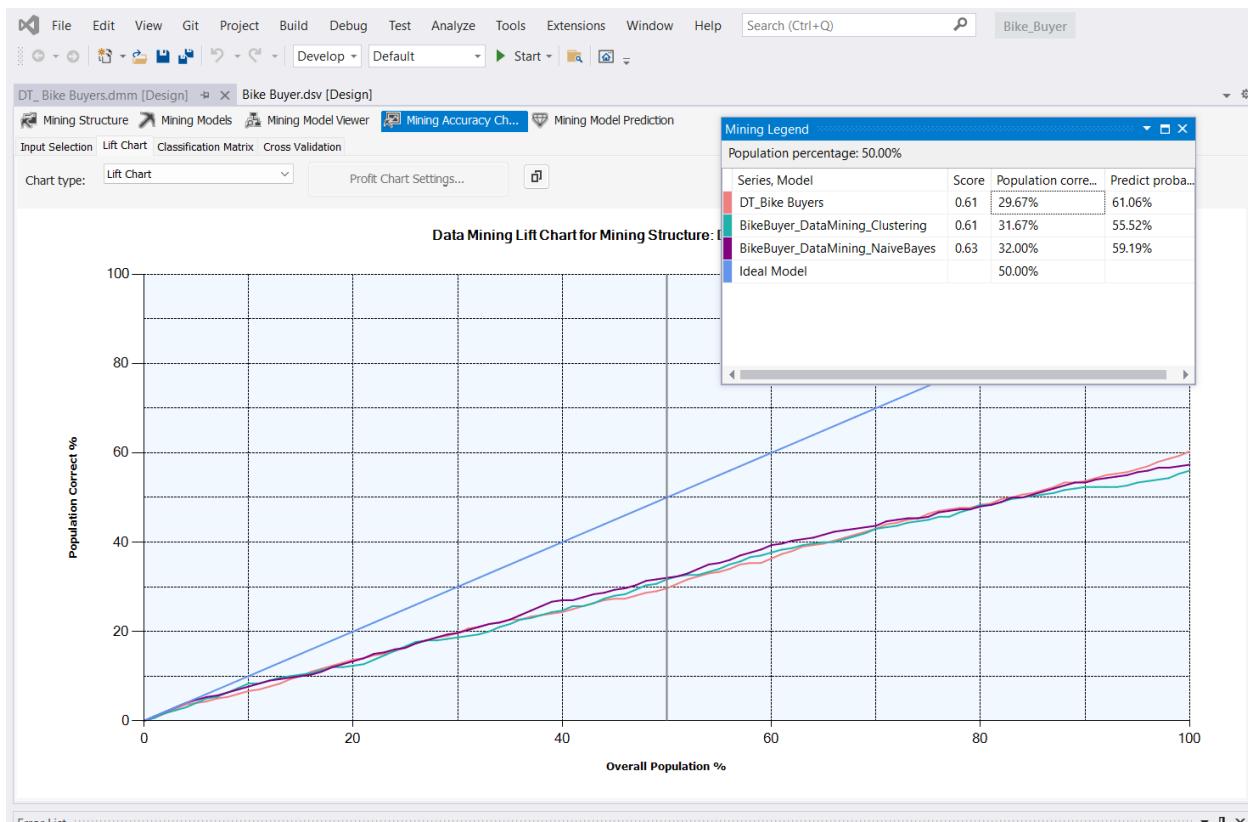
	DT_Bike Buyers	BikeBuyer_DataMining_Cl...	BikeBuyer_DataMining_N...
Age	Microsoft_Decision_Trees	Microsoft_Clustering	Microsoft_Naive_Bayes
Commute Distance	Input	Input	Ignore
ID	Input	Key	Input
Marital Status	Input	Key	Input
Purchased Bike	PredictOnly	Input	PredictOnly

At the bottom of the window, there are tabs for "Error List", "Variables", and "Output".

#### Bước 4 : Tiến hành deploy và process



- So sánh và đánh giá giữa các thuật toán bằng đồ thị Lift .
  - Để xem đồ thị Lift chart trong tab Mining Accuracy Chart chọn Lift Chart.



- Thuật toán Decision Tree có phần trăm chính xác 29.67% với số điểm 0.61
- Thuật toán Clustering có phần trăm chính xác 31.67% với số điểm 0.61
- Thuật toán NaiveBayes có phần trăm chính xác 32.00% với số điểm 0.63
- > **Kết luận** ta ưu tiên chọn thuật toán NaiveBayes.

## VI. Tài liệu tham khảo

1. **Đồ án Xây dựng kho dữ liệu Super Store .** [Link](#)
2. **Đồ án phân tích dữ liệu đại dịch COVID-19.** [Link](#)