

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



BÁO CÁO MÔN HỌC KHO DỮ LIỆU VÀ OLAP

Topic:

**PHÂN TÍCH VÀ ĐÁNH GIÁ CÁC TRẬN ĐẤU OLYMPIC TỪ NĂM
1986 – 2022**

Giảng viên hướng dẫn: **ThS. Đỗ Thị Minh Phụng**

Sinh viên thực hiện: **Lê Gia Bảo – 20521101**

Ho Chi Minh City, May 2024

Table of Contents

CHƯƠNG 1. PHÂN TÍCH VÀ MÔ TẢ DỮ LIỆU	4
1.1 GIỚI THIỆU VỀ DATASET.....	4
1.1.1 Dataset.....	4
1.1.2 Mô tả số dòng, số cột, thời gian thu thập.....	4
1.1.3 Link dataset.....	4
1.1.4 Mô tả chi tiết các thuộc tính dữ liệu gốc.....	4
1.1.5 Kho dữ liệu đã xử lý	6
1.1.6 Hướng chủ đề.....	6
1.2 XÂY DỰNG KHO DỮ LIỆU.....	6
1.2.1 Lược đồ hình sao.....	6
1.2.2 Các bảng Dimension.....	7
1.3 CÁC CÂU TRUY VẤN.....	9
CHƯƠNG 2: QUÁ TRÌNH TRÍCH XUẤT DỮ LIỆU, BIẾN ĐỔI VÀ NẠP DỮ LIỆU VÀO KHO DỮ LIỆU (QUÁ TRÌNH SSIS).	9
2.1 CHUẨN BỊ CÔNG CỤ VÀ DATA WAREHOUSE.	9
2.2 THỰC HIỆN QUÁ TRÌNH SSIS	10
Tạo project “Integration Service Project” mới.....	10
Tạo database	11
Thiết lập kết nối.....	12
Quá trình import dữ liệu từ file CSV với SQL server	14
Quá trình gộp dữ liệu 4 bảng dữ liệu	19
Quá trình làm sạch dữ liệu.....	37
Quá trình tạo các bảng Dimension.....	45
Quá trình tạo bảng FACT	71
Tạo Execute SQL Task.....	80
CHƯƠNG 3. PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS).	86
3. 1 QUÁ TRÌNH SSAS TRONG VISUAL STUDIO 2022.....	86
3.1.1 Tạo project trong visual studio	86
3.1.2 Xác định nguồn dữ liệu – Tạo mới Data Source.	87

3.1.3	Tạo mới Data Source Views.....	92
3.1.4	Tạo Cubes	95
3.1.5	Tạo các bảng Dimension	98
3. 2	THỰC HIỆN CÁC CÂU TRUY VẤN.....	104
CHƯƠNG 4. QUÁ TRÌNH TẠO REPORT BẰNG CÔNG CỤ POWER.....		135
5.1	NHẬP DỮ LIỆU NGUỒN	135
THỰC HIỆN CÁC CÂU TRUY VẤN.....		137
CHƯƠNG 5 KHAI THÁC DỮ LIỆU – DATA MINING		142
5.1	HƯỚNG PHÂN TÍCH.....	142
5.2	THỰC HIỆN	143
5.2.1	Decision Tree.....	143
5.2.2	Random Forest.....	145

CHƯƠNG 1. PHÂN TÍCH VÀ MÔ TẢ DỮ LIỆU

1.1 GIỚI THIỆU VỀ DATASET.

1.1.1 Dataset

- Tên dataset: Olympic Summer & Winter Games, 1986-2022
- Datasets này dùng để phân tích dữ liệu về hơn 19.000 huy chương, 150.000 kết quả, 74.000 vận động viên, 20.000 tiểu sử và 52 nước tổ chức Thế vận hội Olympic mùa hè và mùa đông.

1.1.2 Mô tả số dòng, số cột, thời gian thu thập

- Bộ dữ liệu gồm 159021 dòng dữ liệu và 41 cột thuộc tính bao gồm: Bảng olympic_athletes gồm 75904 dòng dữ liệu và 7 cột thuộc tính Bảng olympic_host gồm 52 dòng dữ liệu và 7 cột thuộc tính.
- Bảng olympic_medals gồm 21697 dòng dữ liệu và 12 cột thuộc tính. Bảng olympic_results gồm 162804 dòng dữ liệu và 15 cột thuộc tính.
- Dữ liệu được thu thập từ năm 1986 đến 2022

1.1.3 Link dataset

Link dataset: <https://www.kaggle.com/datasets/piterfm/olympic-games-medals-19862018>

1.1.4 Mô tả chi tiết các thuộc tính dữ liệu gốc

STT	Tên thuộc tính	Ý nghĩa	Kiểu dữ liệu
1	athlete_url	Link giới thiệu vận động viên	Url
2	athlete_full_name	Tên đầy đủ vận động viên	String
3	first_game	Tên giải đấu đầu tiên tham gia	String
4	athlete_year_birthday	Năm sinh vận động viên	Integer
5	athlete_medals	Số huy chương đạt được theo từng loại	String
6	games_participants	Tổng số lần tham gia giải đấu	Integer
7	bio	Tiểu sử vận động viên	String

Mô tả chi tiết bảng olympic_athletes

STT	Tên thuộc tính	Ý nghĩa	Kiểu dữ liệu
1	game_slug	Mã giải đấu	String
2	game_end_date	Thời gian kết thúc	String
3	game_start_date	Thời gian bắt đầu	String
4	game_location	Địa điểm tổ chức	String
5	game_name	Tên giải đấu	String
6	game_season	Mùa giải đấu	String
7	game_year	Năm tổ chức giải đấu	Integer

Mô tả chi tiết bảng olympic_host

STT	Tên thuộc tính	Ý nghĩa	Kiểu dữ liệu
1	discipline_title	Thể loại môn thi đấu	String
2	slug_game	Mã giải đấu	String
3	event_title	Nội dung của thể loại môn thi đấu	String
4	event_gender	Giới tính được tham gia	String
5	medal_type	Loại huân chương đạt được	String
6	participant_type	Tham gia theo cá nhân hoặc theo đội	String
7	participant_title	Tên của đội tham gia	String
8	athlete_url	Link giới thiệu vận động viên	Url
9	athlete_full_name	Tên vận động viên	String
10	country_name	Tên quốc gia	String
11	country_code	Mã quốc gia	String
12	country_3_letter_code	Mã quốc gia gồm 3 ký tự	String

Mô tả chi tiết bảng olympic_medals

STT	Tên thuộc tính	Ý nghĩa	Kiểu dữ liệu

1	discipline_title	Thể loại môn thi đấu	String
2	event_title	Nội dung thể loại môn thi đấu	String
3	slug_game	Mã giải đấu	String
4	participant_type	Tham gia theo cá nhân hoặc theo đội	String
5	medal_type	Loại huân chương đạt được	String
6	athlete	Link giới thiệu vận động viên (team từ 2 người)	Url
7	rank_equal	Đồng vị trí hay không	String
8	rank_position	Vị trí	Integer
9	country_name	Tên quốc gia	String
10	country_code	Mã quốc gia	String
11	country_3_letter_code	Mã quốc gia gồm 3 ký tự	String
12	athlete_url	Link giới thiệu vận động viên	Url
13	athlete_full_name	Tên đầy đủ vận động viên	String
14	value_unit	Giá trị của trận đấu	String
15	value_type	Loại giá trị	String

Mô tả chi tiết bảng olympic_results

1.1.5 Kho dữ liệu đã xử lý

- Sau khi lọc dữ liệu ta được 257782 dòng và 20 thuộc tính để sử dụng cho việc phân tích đề tài

1.1.6 Hướng chủ đề

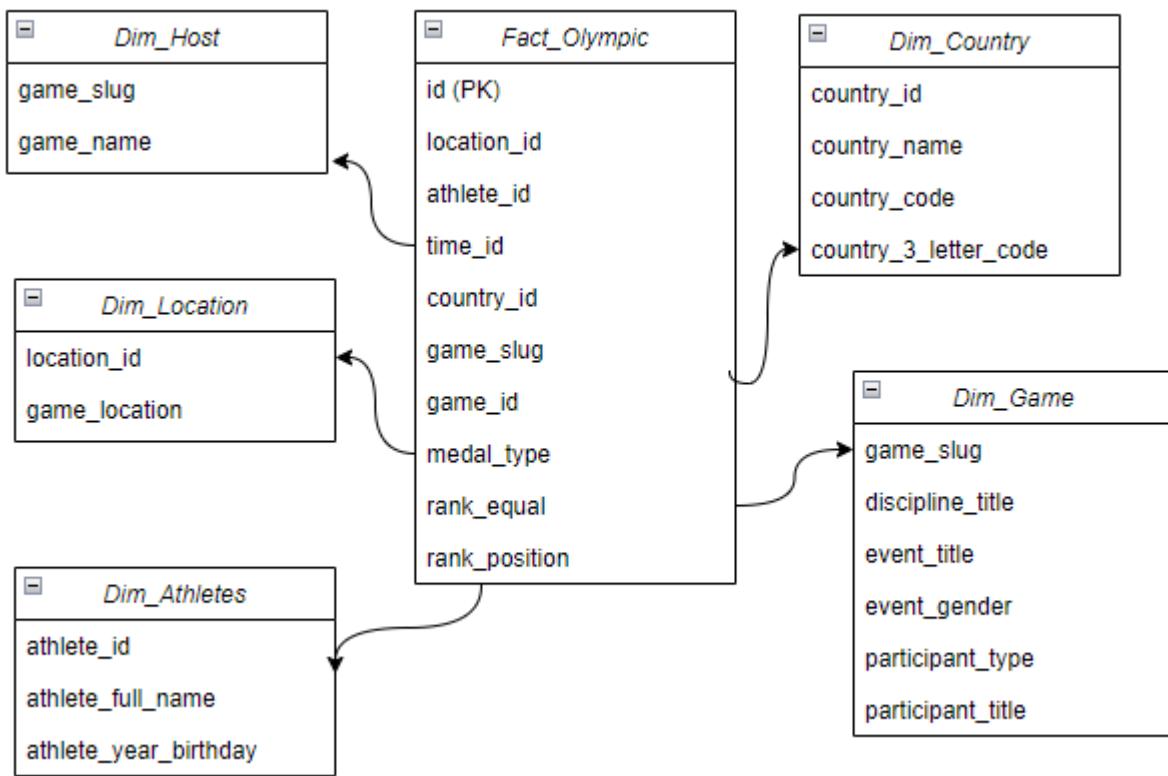
- Hướng chủ đề của nhóm sẽ là phân tích kết quả và đánh giá về các trận đấu Olympic Games.

1.2 XÂY DỰNG KHO DỮ LIỆU

1.2.1 Lược đồ hình sao

- Lược đồ hình sao chứa 7 bảng. Trong đó bảng FACT là bảng sự kiện chứa các khoá ngoại và các độ đo. Dim_Host, Dim_Country, Dim_Time, Dim_Location, Dim_Game và Dim_Athlete là các bảng chiều chứa thuộc tính

khoá chính và các thuộc tính khác của chiều.



1.2.2 Các bảng Dimension

A Bảng DIM_Host

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
🔑	game_slug	String	Mã giải đấu
	game_name	String	Tên giải đấu

B Bảng DIM_Country

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
🔑	country_id	Integer	Mã quốc gia
	country_name	String	Tên quốc gia
	country_code	String	Mã quốc gia

	country_code_3_letter_code	String	Mã quốc gia (3 ký tự)
--	----------------------------	--------	-----------------------

C Bảng DIM_Time

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
🔑	time_id	Integer	Mã thời gian
	game_end_date	String	Thời gian kết thúc giải đấu
	game_start_date	String	Thời gian bắt đầu giải đấu
	game_season	String	Mùa tổ chức giải đấu
	game_year	Integer	Năm tổ chức giải đấu

D Bảng DIM_Location

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
🔑	location_id	Integer	Mã địa điểm tổ chức
	game_location	String	Tên địa điểm tổ chức

E Bảng DIM_Game

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
🔑	game_id	Integer	Mã trận đấu
	discipline_title	String	Tên bộ môn thi đấu
	event_title	String	Tên nội dung thi đấu
	event_gender	String	Giới tính vdv tham gia thi đấu
	participant_type	String	Hình thức thi đấu theo cá nhân hoặc đội
	participant_title	String	Tên đội thi (nếu thi đấu theo đội)

F Bảng DIM_Athlete

	Tên thuộc tính	Kiểu dữ liệu	Mô tả

	athlete_id	Integer	Mã vận động viên
	athlete_full_name	String	Tên đầy đủ vận động viên
	athlete_year_birth	Integer	Năm sinh vận động viên

G Bảng kiện FACT

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
	FACT_id	Integer	Mã thứ tự
	country_id	Integer	Mã quốc gia
	athlete_id	Integer	Mã vận động viên
	time_id	Integer	Mã thời gian
	location_id	Integer	Mã địa điểm
	game_id	Integer	Mã bộ môn thi đấu
	game_slug	Integer	Mã giải đấu
	medal_type	String	Loại huy chương
	rank_equal	String	Đồng vị trí
	rank_position	Integer	Vị trí

1.3 CÁC CÂU TRUY VẤN.

CHƯƠNG 2: QUÁ TRÌNH TRÍCH XUẤT DỮ LIỆU, BIẾN ĐỔI VÀ NẠP DỮ LIỆU VÀO KHO DỮ LIỆU (QUÁ TRÌNH SSIS).

2.1 CHUẨN BỊ CÔNG CỤ VÀ DATA WAREHOUSE.

- Tải và cài đặt công cụ SQL Server 2019 , SQL Server Management Studio
- Cài đặt Visual Studio
- Cài đặt các Extension:
 - SQL Analysis Services
 - SQL Server Integration Services Projects



SQL Server Integration Services Projects

Microsoft | 1.661.828 installs | ★★★★★ (204) | Free

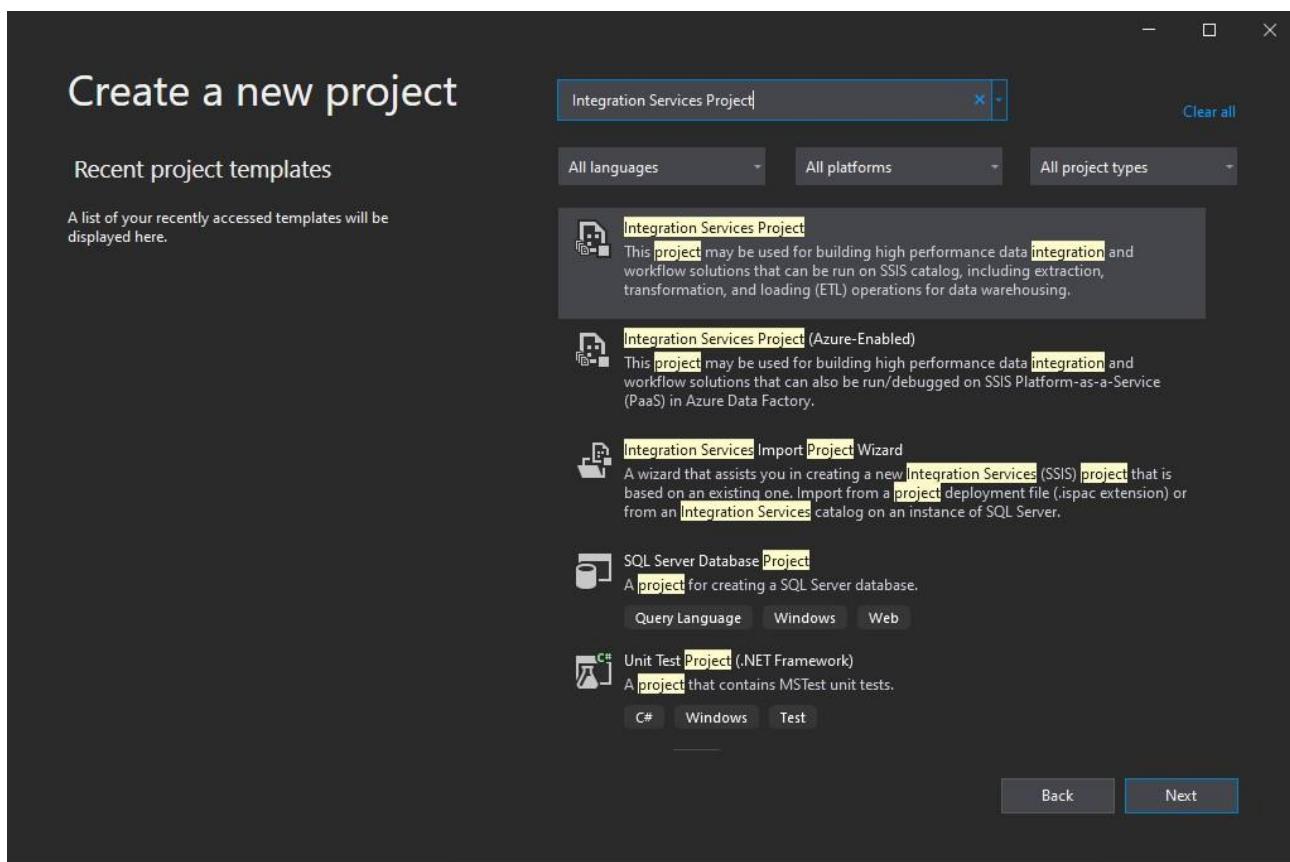
This project may be used for building high performance data integration and workflow solutions, including extraction, transformation, and loading (ETL) operations for data warehousing.

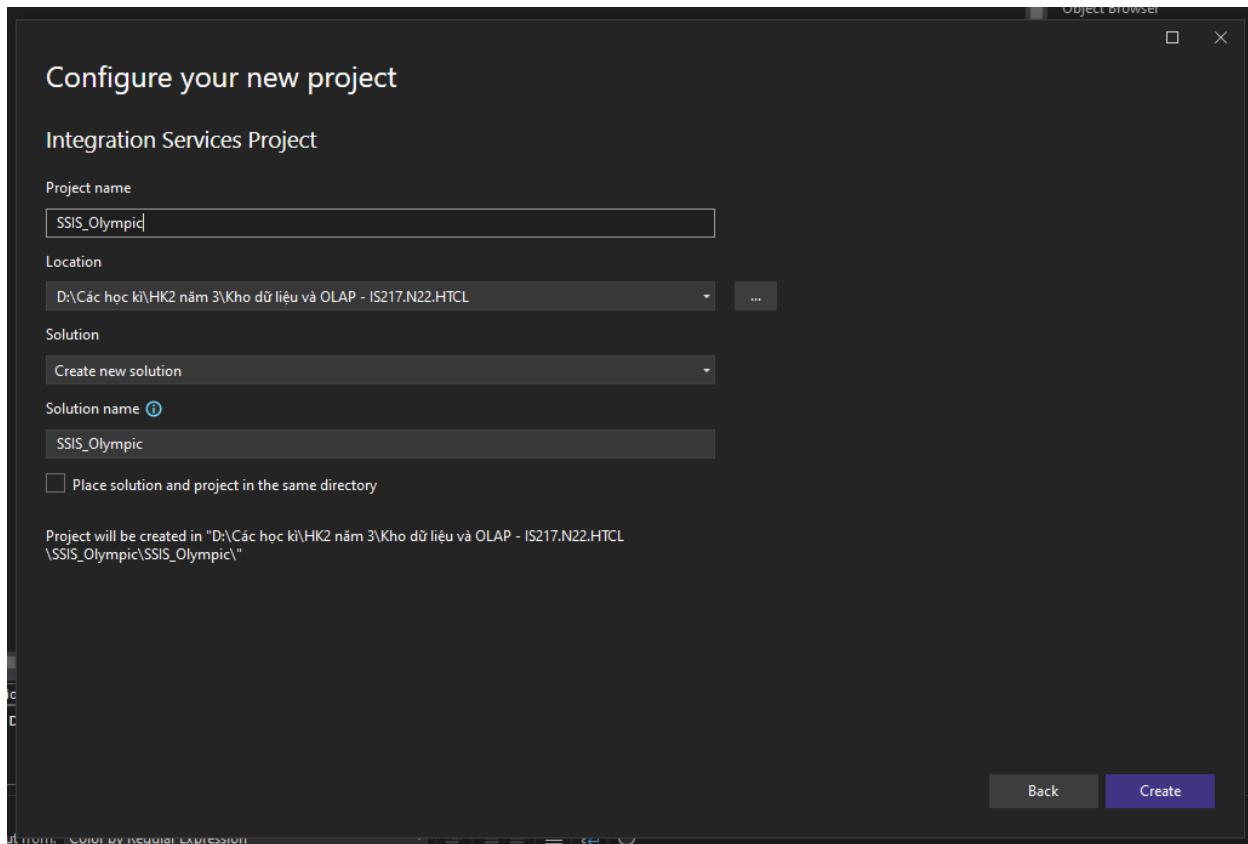
[Download](#)

2.2 THỰC HIỆN QUÁ TRÌNH SSIS

Tạo project “Integration Service Project” mới.

- Ở tab Create a new project , chọn Integration Services Project.





Tạo database

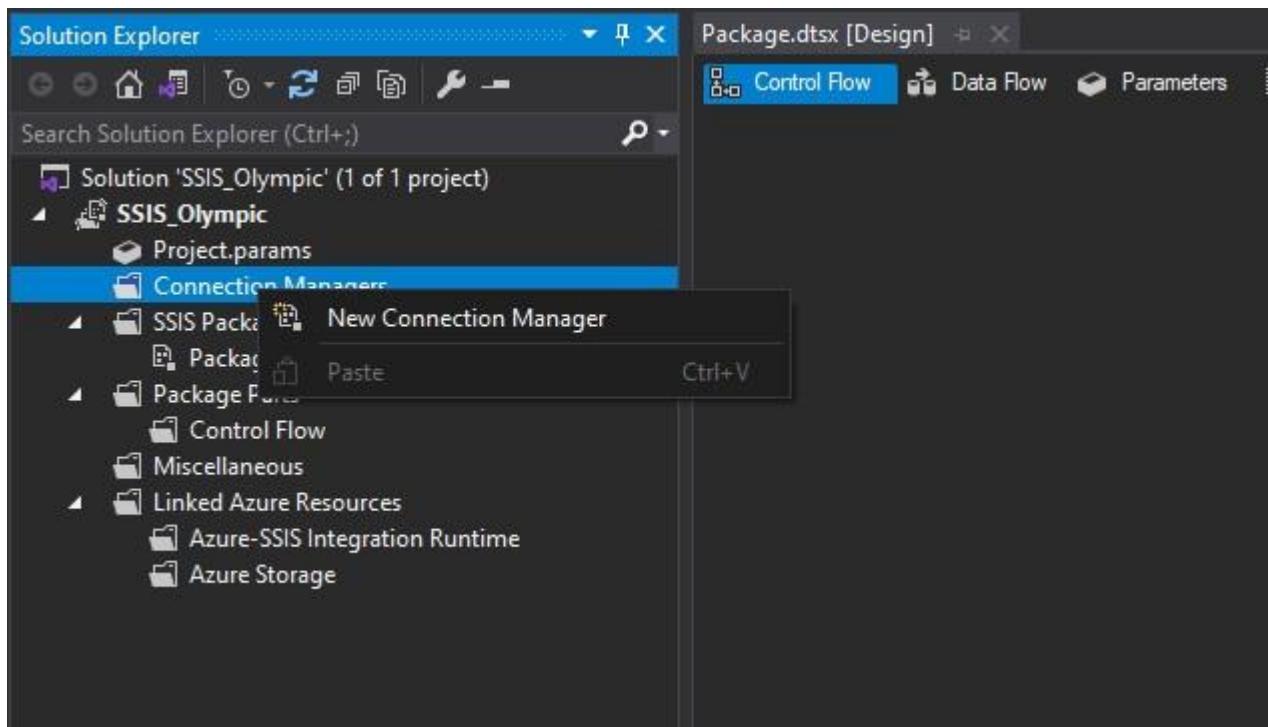
- Tạo 2 database
 - OriginalData: lưu những dữ liệu gốc được import từ CSV, bao gồm các:
 - Data_Original: được import thẳng từ file csv.
 - Data_Clean: chứa dữ liệu sử dụng sau khi làm sạch
 - Data_NULL: chứa dữ liệu Null trong quá trình làm sạch
 - Olympic_DWH: Data Warehouse

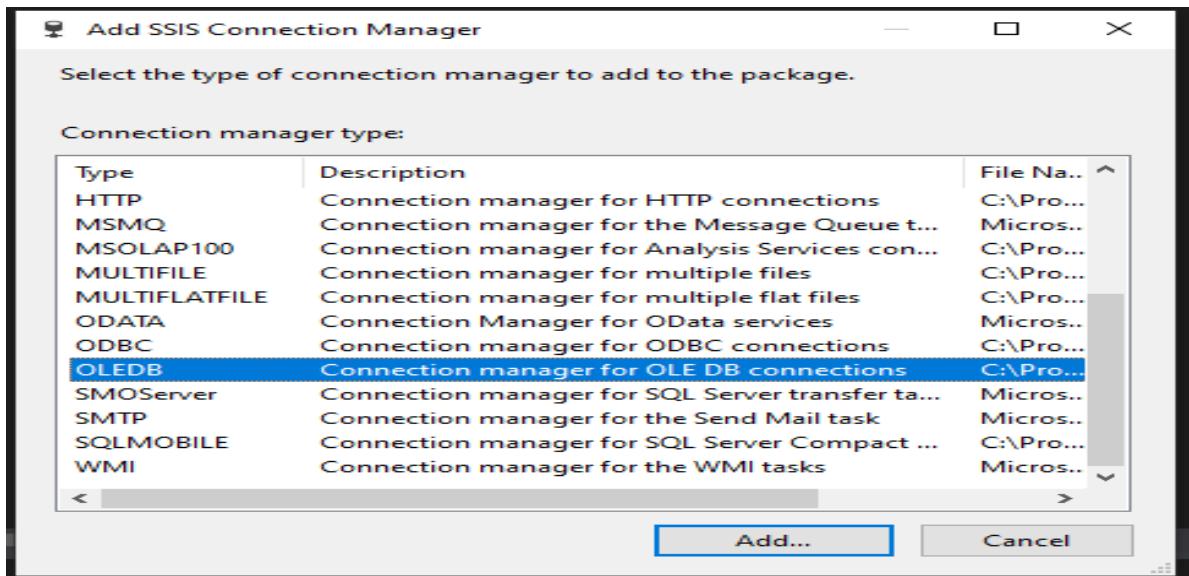
The screenshot shows the SSMS interface. The Object Explorer on the left lists the databases in the 'ZERA\SQLEXPRESS' instance: System Databases, Database Snapshots, Olympic_DW, OriginalData, Security, Server Objects, Replication, PolyBase, Management, and XEvent Profiler. The 'master' database is selected. In the center, a query window titled 'SQLQueryForDatabase...' shows two database creation statements:

```
CREATE DATABASE OriginalData  
CREATE DATABASE Olympic_DW
```

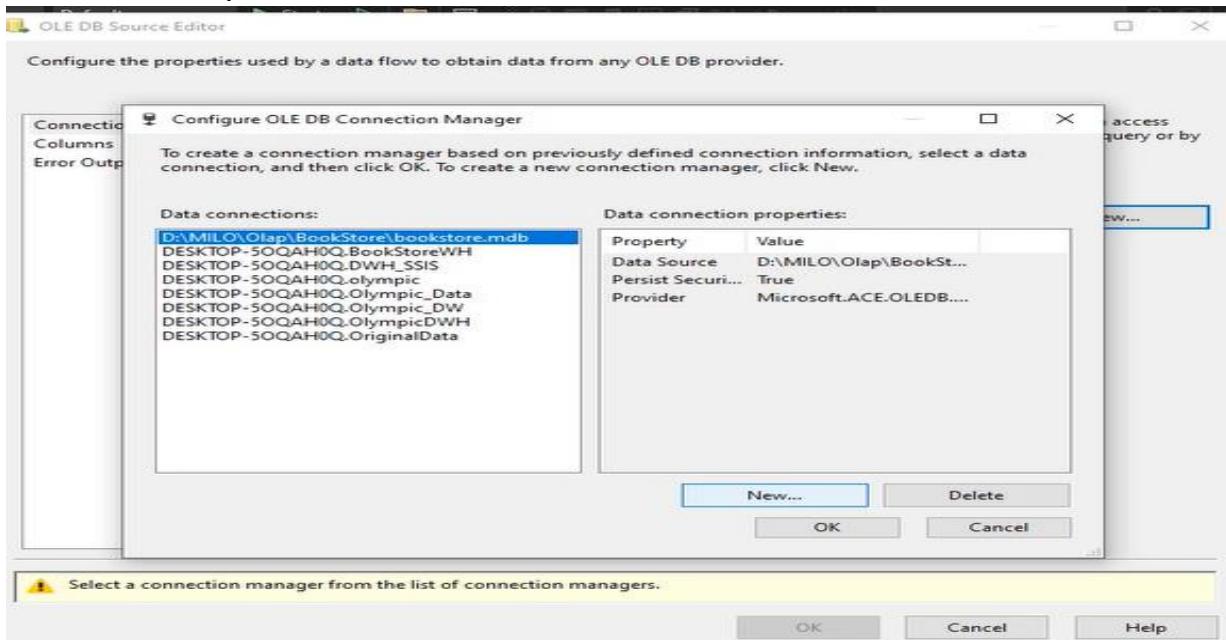
Thiết Lập kết nối

- Tại cửa sổ Solution Explorer, click chuột phải chọn New Connection Manager, chọn kết nối với OLDEDB

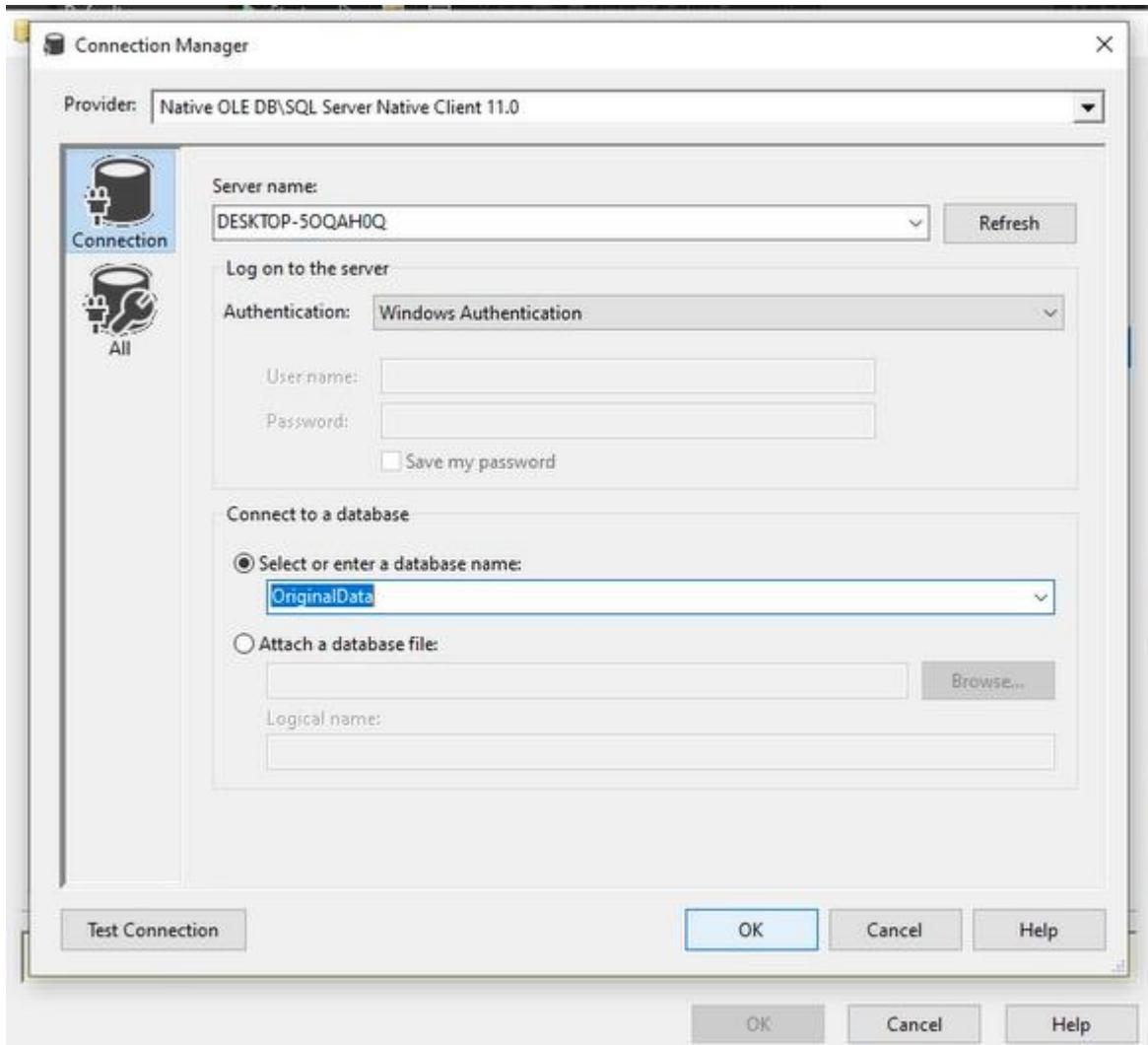




- Chọn new để tiến hành thêm connection.



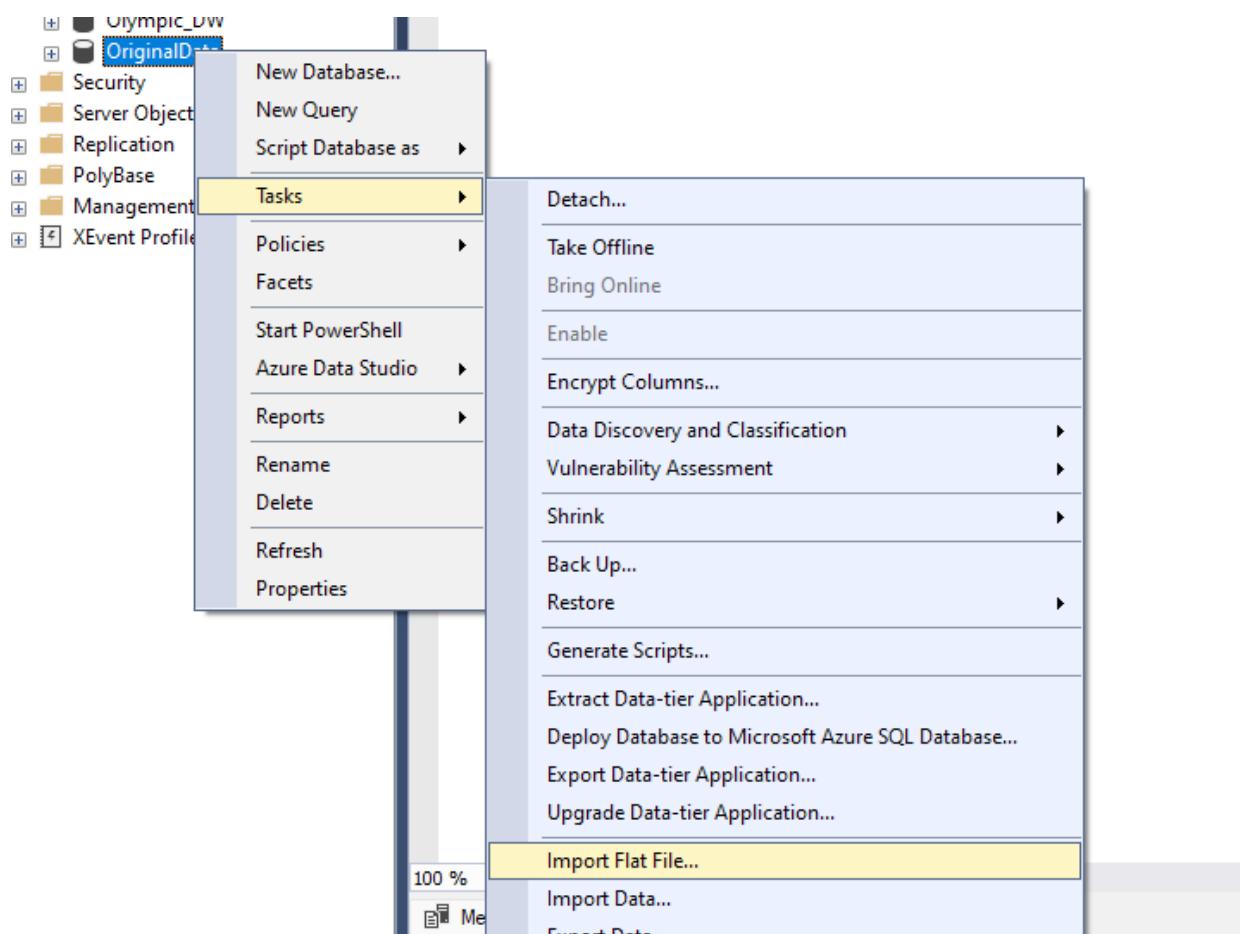
- Điền tên Server SQL vào hàng Server name, sau đó chọn Database “OriginalData” ở mục Select or enter a database name. Tiếp theo nhấn Enter để hoàn tất.



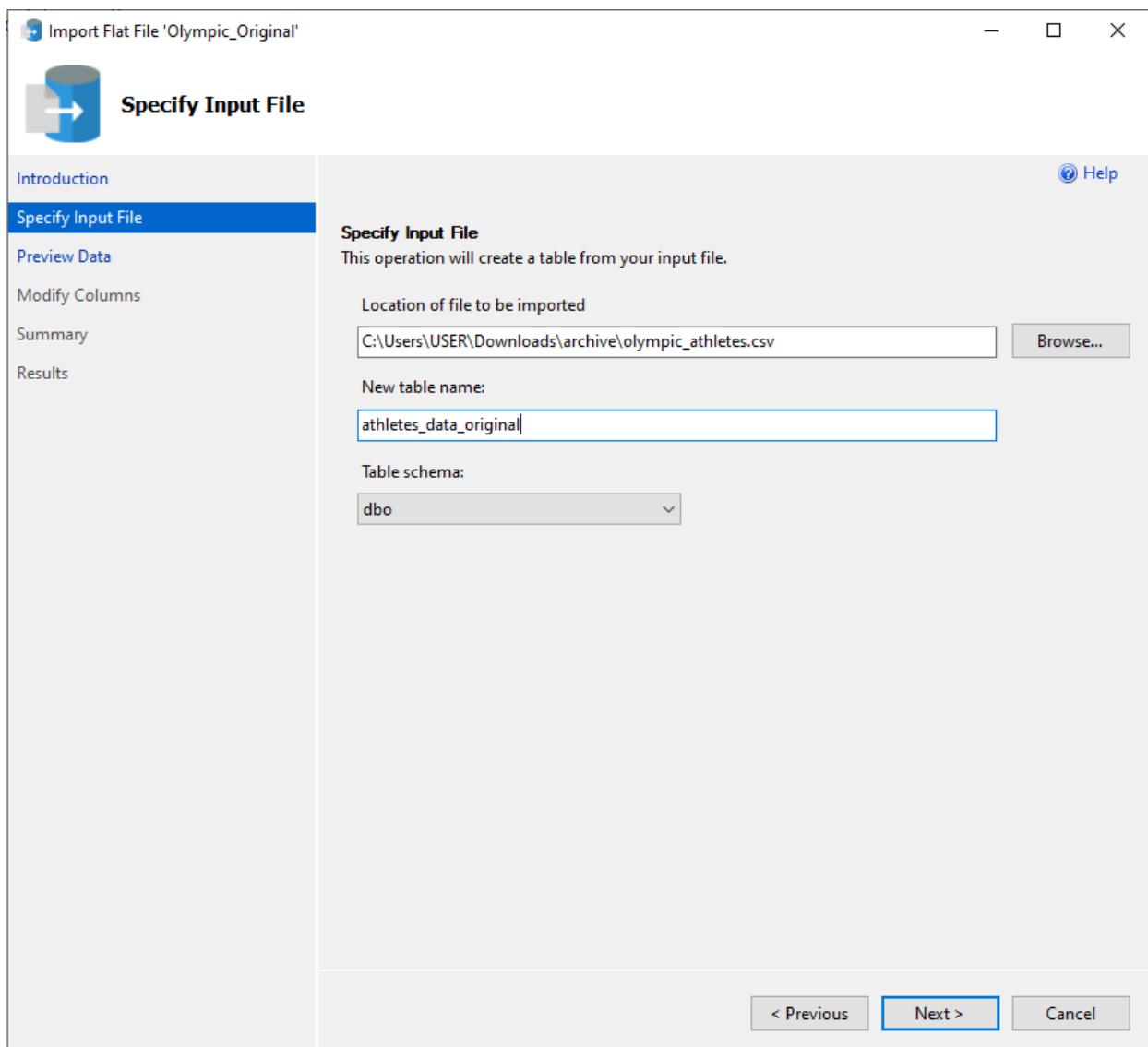
- Ta cũng thực hiện tương tự để connect đến với database “Olympic_DWH”

Quá trình import dữ liệu từ file CSV với SQL server

- Trong database, ta click phải chuột vào “OriginalData” → task → import Flat file để lấy dữ liệu từ file CSV.



- Tại phần Import Flat File ‘OriginalData’, chọn Specific Input File sau đó chọn vị trí file muốn đưa dữ liệu. Ấn Next để qua bước tiếp theo



- Chọn Finish để hoàn thành Import Flat File:

Import Flat File 'Olympic_Original'

Preview Data

Introduction Specify Input File Preview Data Modify Columns Summary Results

Help

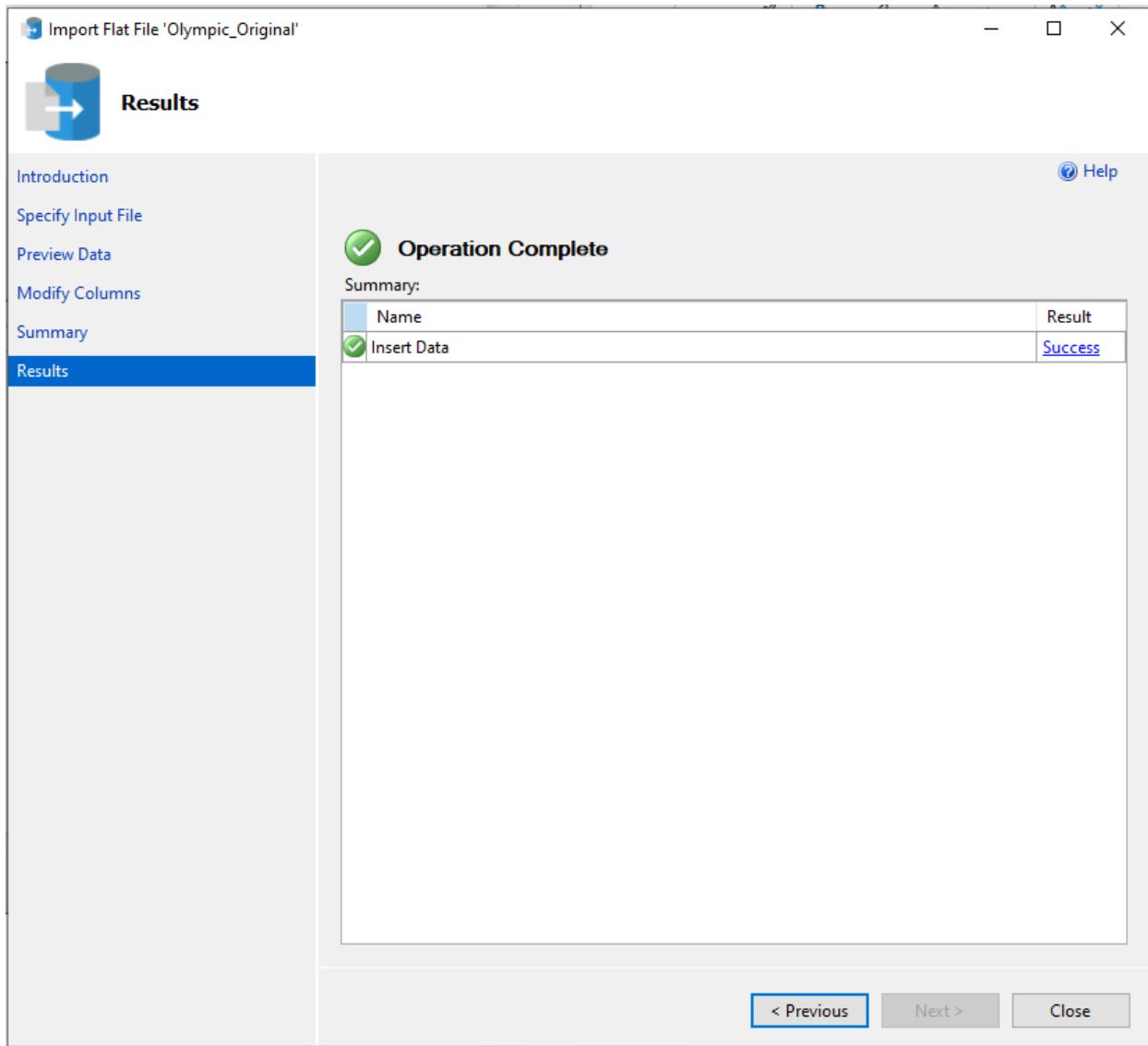
Preview Data

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

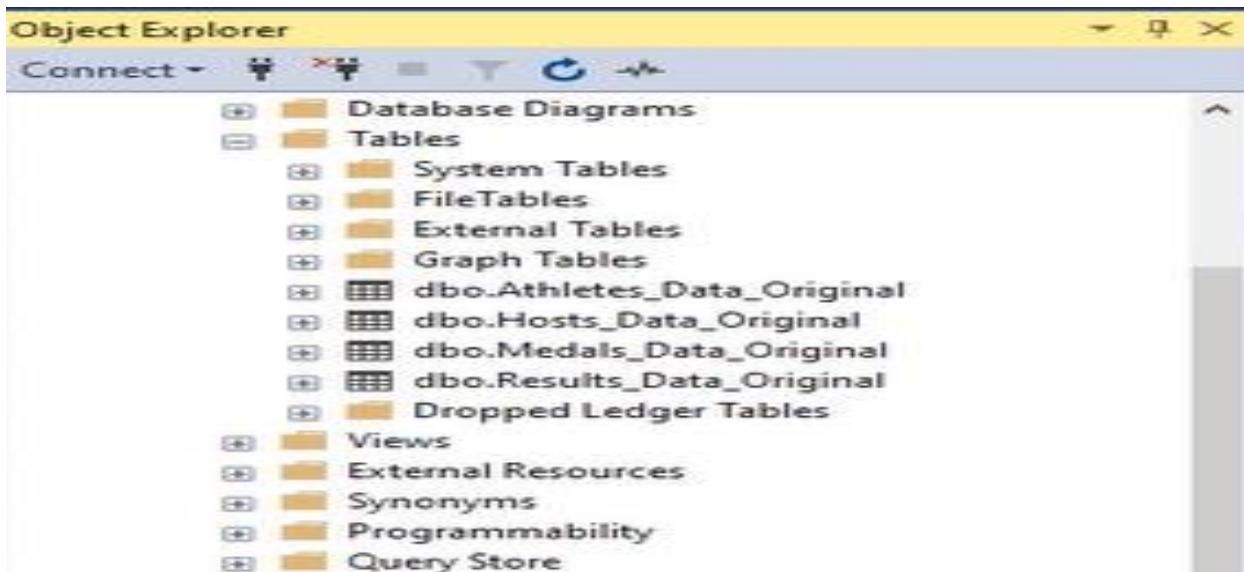
athlete_url	athlete_full_name	games_participat	first_game	athlete_year_birth	athlete
https://olympic.org/athletes/cooper-wood	Cooper WOOD...	1	Beijing 2022	2000	
https://olympic.org/athletes/felix-elofsson	Felix ELOFSSON	2	PyeongChang ...	1995	
https://olympic.org/athletes/dylan-walczyk	Dylan WALCZYK	1	Beijing 2022	1993	
https://olympic.org/athletes/olli-penttala	Olli PENTTALA	1	Beijing 2022	1995	
https://olympic.org/athletes/dmitriy-reikhe	Dmitriy REIKHE...	1	Beijing 2022	1989	
https://olympic.org/athletes/matt-graham	Matt GRAHAM	3	Sochi 2014	1994	1S
https://olympic.org/athletes/ikuma-horishi	Ikuma HORISHI...	2	PyeongChang ...	1997	1B
https://olympic.org/athletes/daichi-hara	Daichi HARA	2	PyeongChang ...	1997	1B
https://olympic.org/athletes/laurent-dumais	Laurent DUMAIS	1	Beijing 2022	1996	
https://olympic.org/athletes/james-mathes	James MATHES...	2	PyeongChang ...	1995	
https://olympic.org/athletes/pavel-kolmakov	Pavel KOLMAKOV	3	Sochi 2014	1996	
https://olympic.org/athletes/kosuke-kobashi	Kosuke KOBASHI	1	Beijing 2022	1994	
https://olympic.org/athletes/brodie-summers	Brodie SUMMERS	3	Sochi 2014	1993	
https://olympic.org/athletes/severi-vierela	Severi VIERELA	1	Beijing 2022	2001	
https://olympic.org/athletes/marco-tade	Marco TADE	1	Beijing 2022	1995	
https://olympic.org/athletes/william-feneley	William FENELEY	1	Beijing 2022	1999	
https://olympic.org/athletes/mikael-kingsbury	Mikael KINGSBURY	3	Sochi 2014	1992	1G2S

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

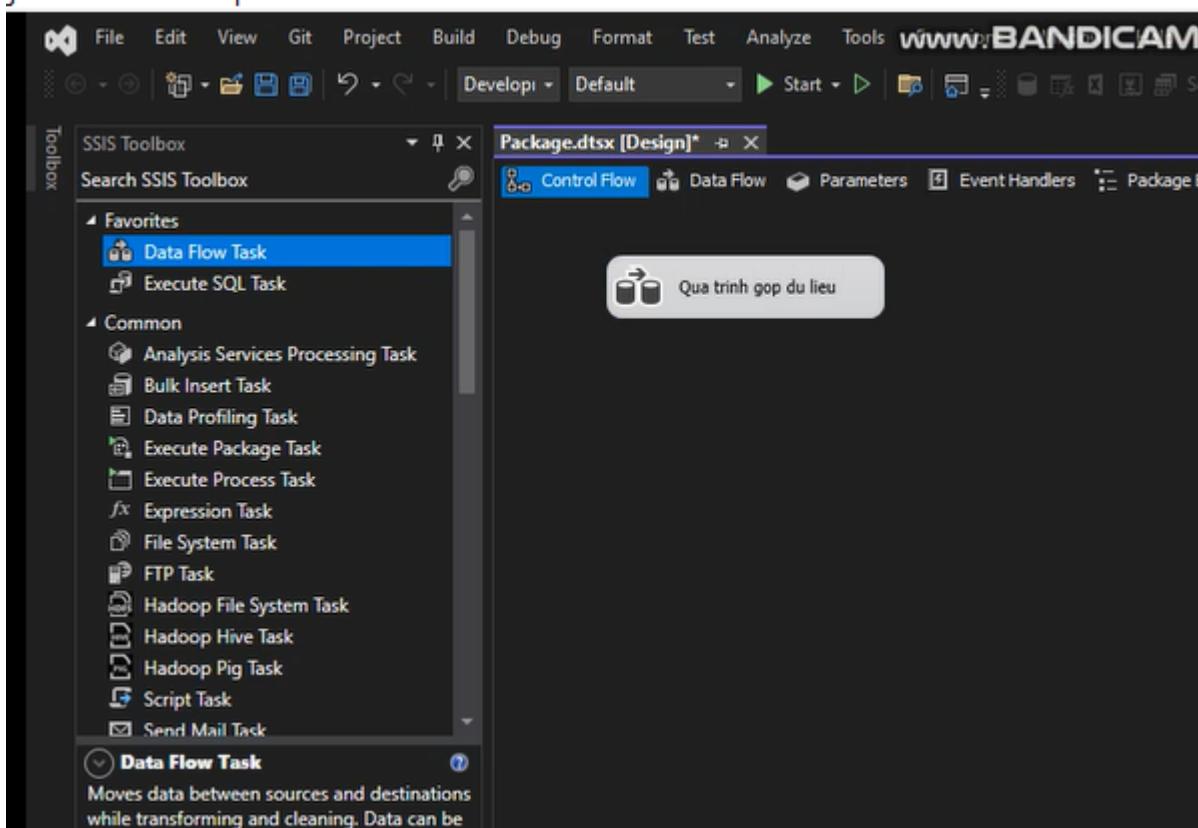
< Previous Next > Cancel



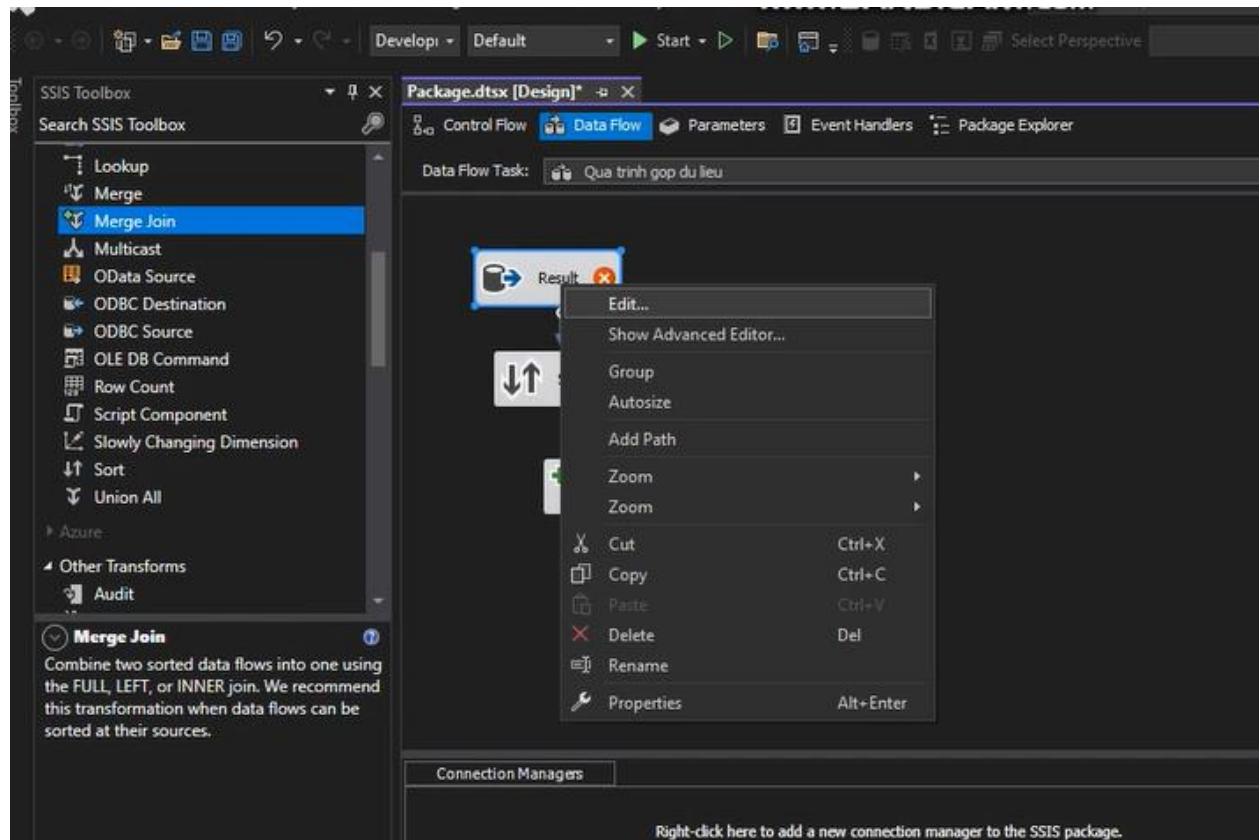
- Ta cũng thực hiện tương tự cho các file: Host, Medal, Result
- Sau khi thực hiện xong, sẽ có được kết quả như hình sau

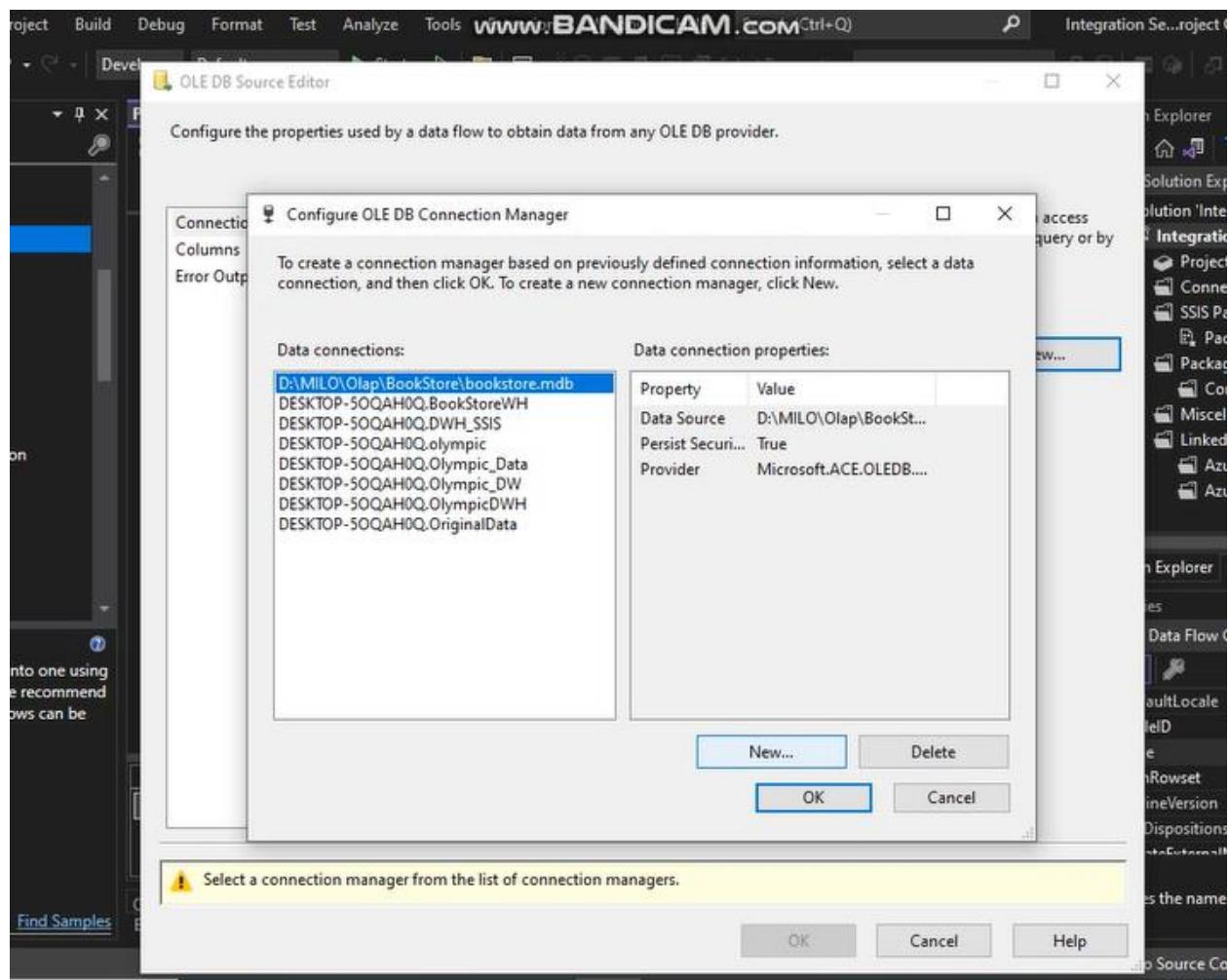


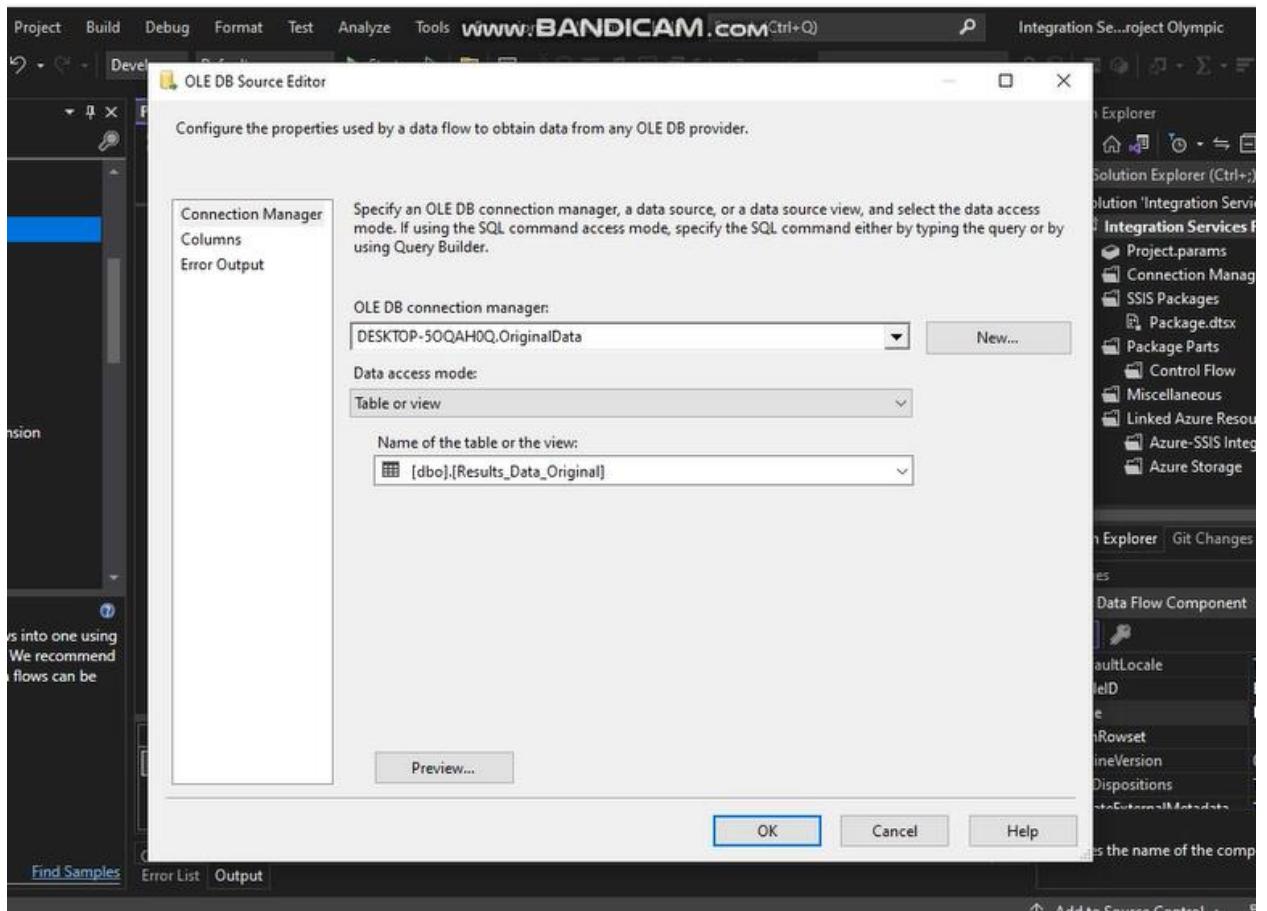
Quá trình gộp dữ liệu 4 bảng dữ liệu

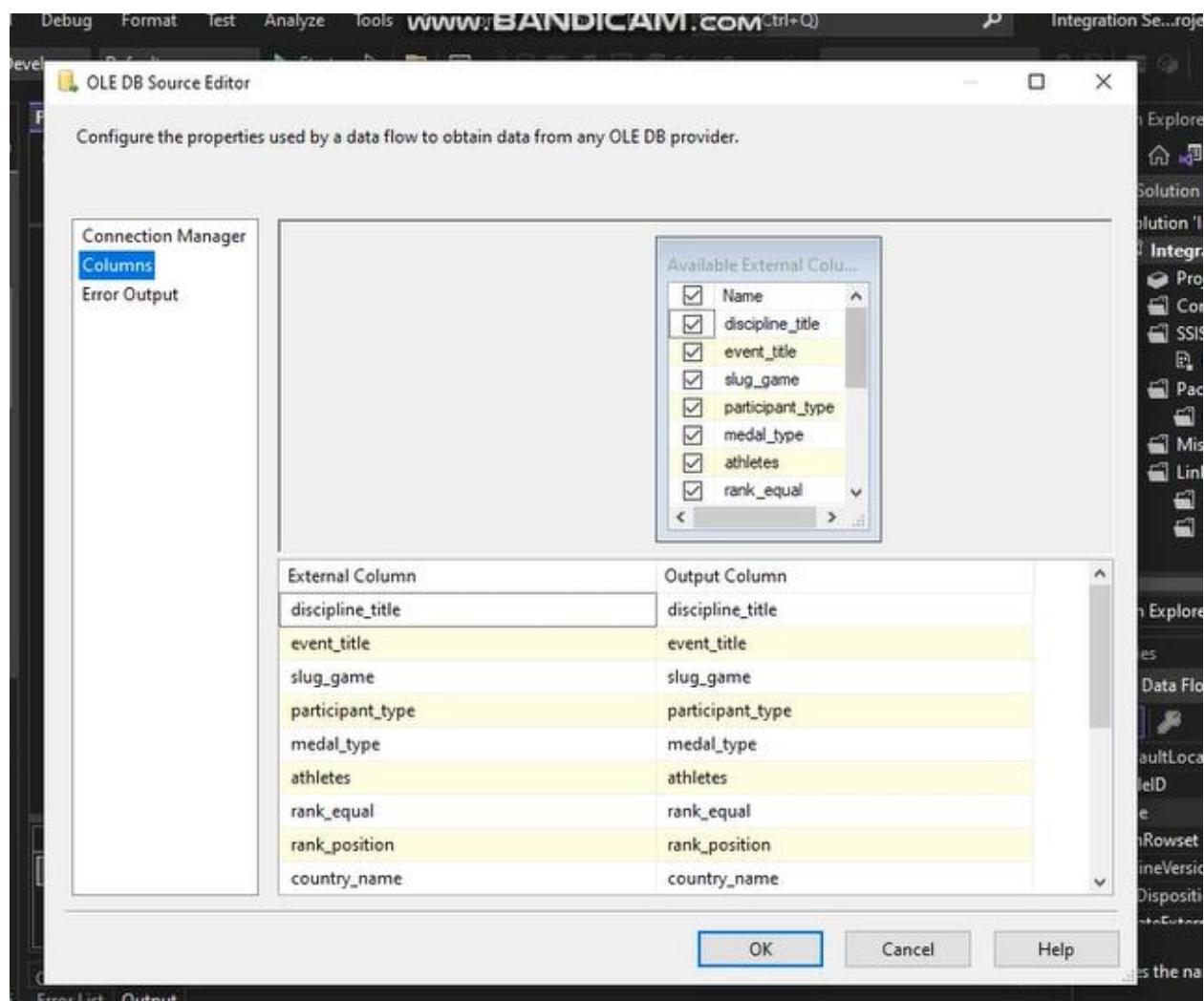


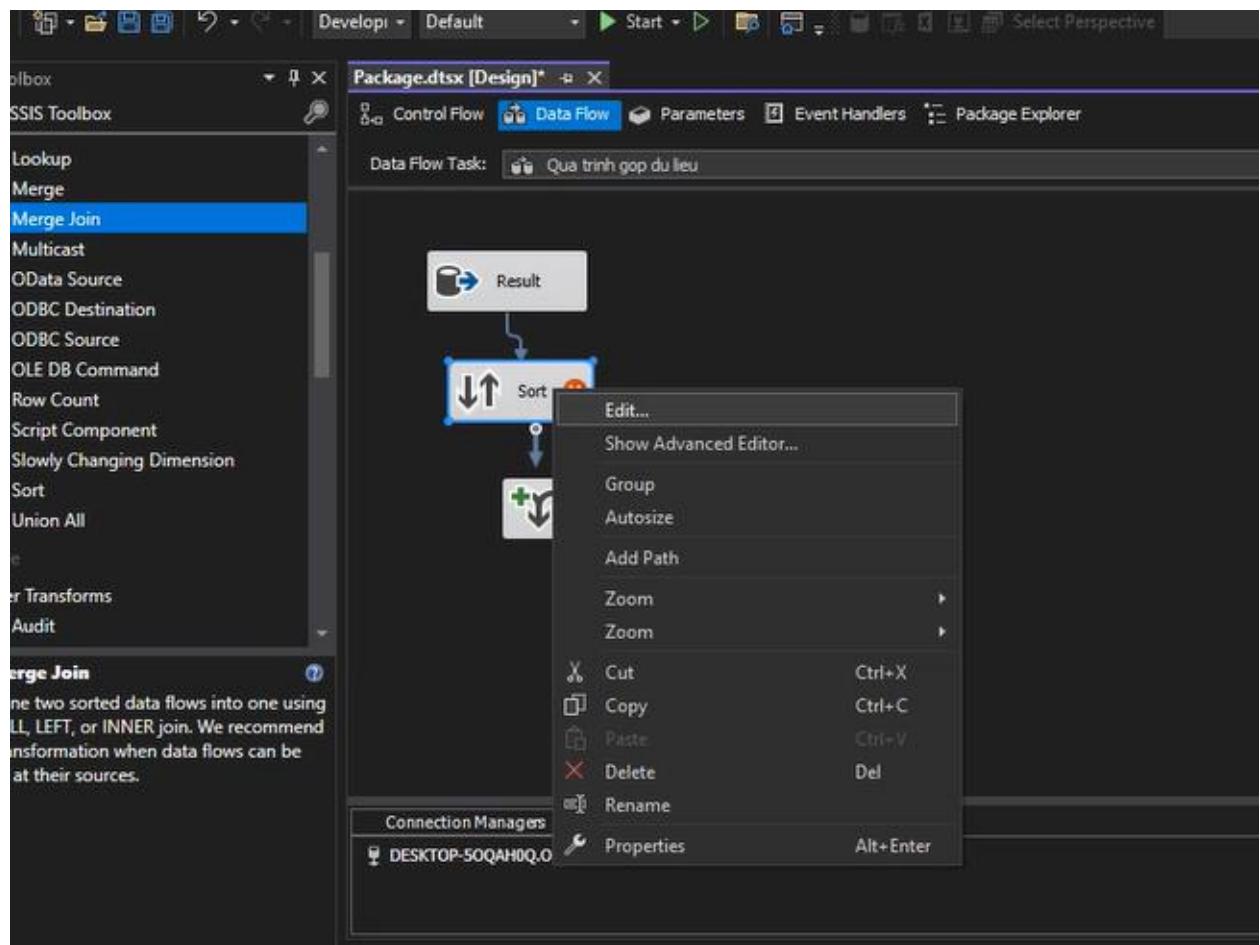
- Sau khi tạo được Data Flow Task “Quá trình gộp dữ liệu”, ta click đúp chuột để chuyển sang tab Data Flow và bắt đầu tiến hành merge dữ liệu.













Sort Transformation Editor

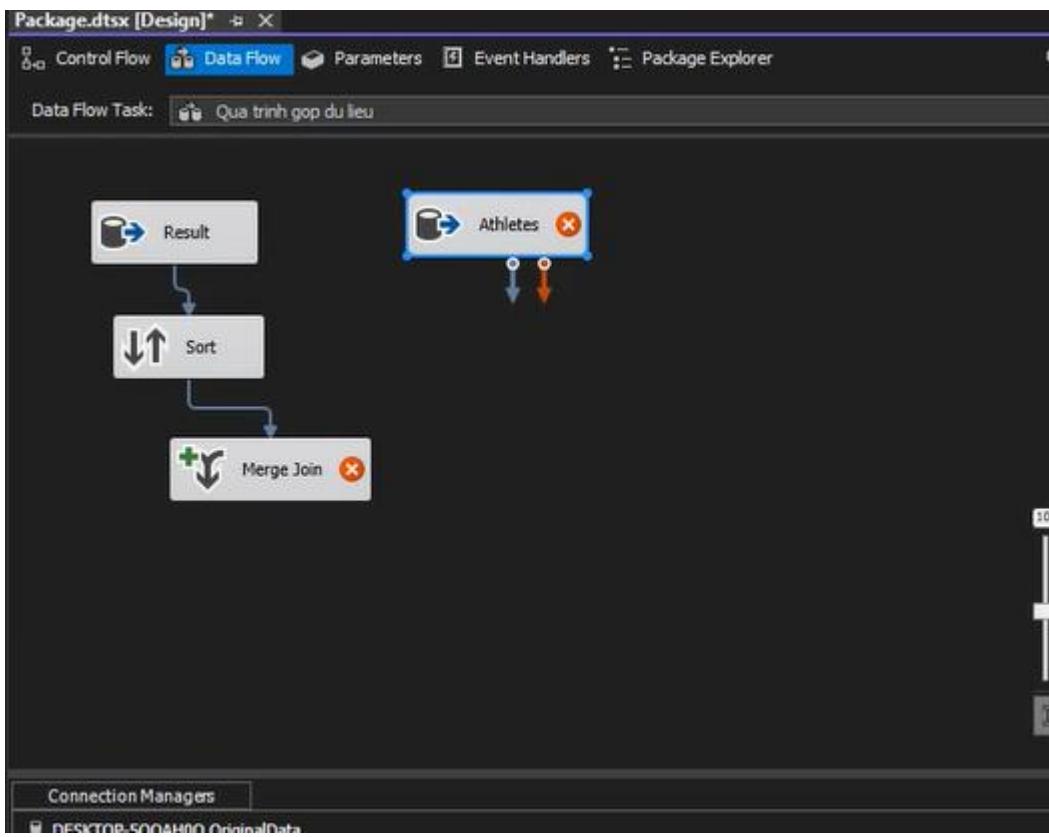
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

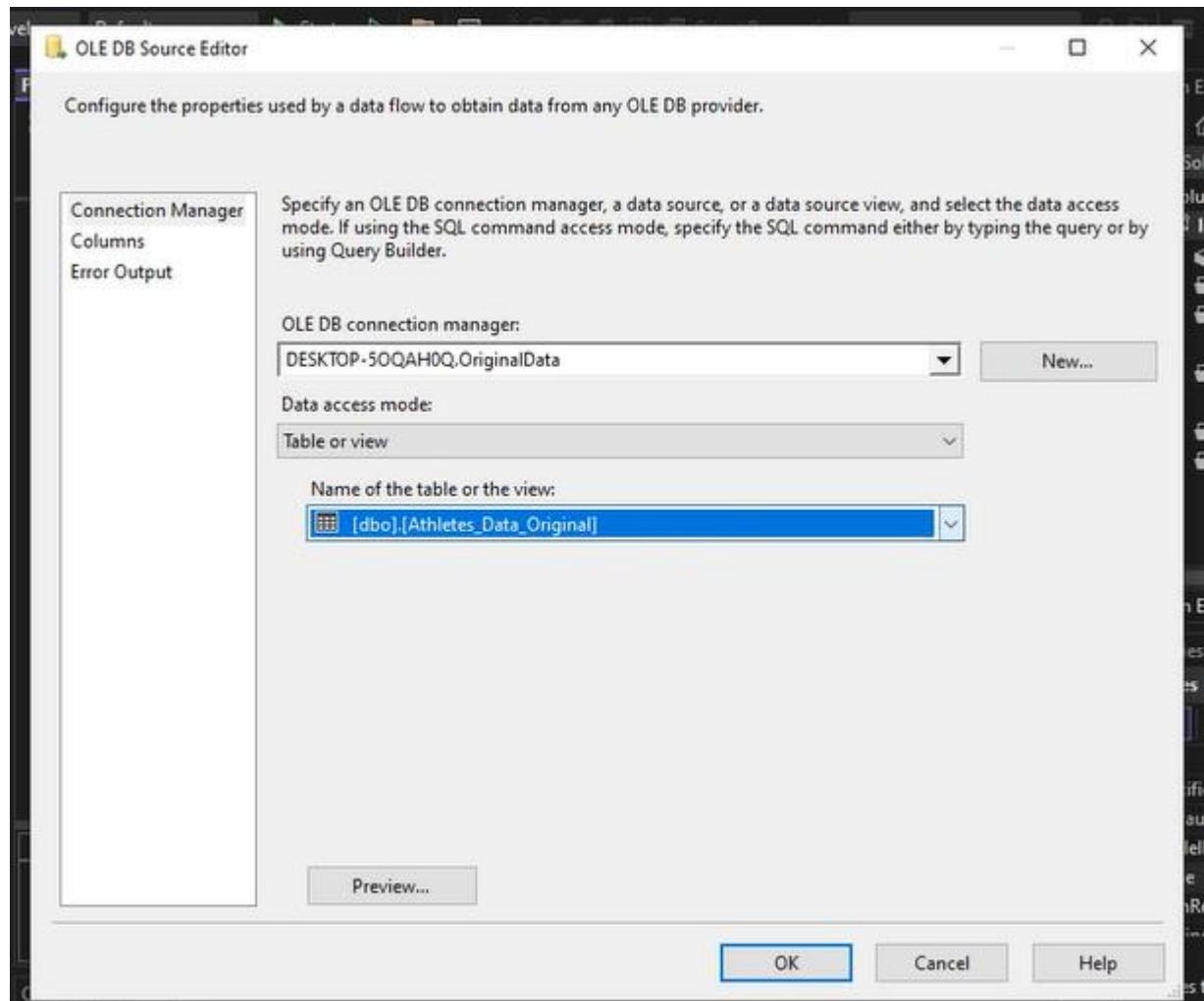
Input Column	Output Alias	Source Component	Re	Sort Type	Sort Order	Con
athlete_url	athlete_url			ascending	1	
athlete_full_name	athlete_full_name			ascending	2	

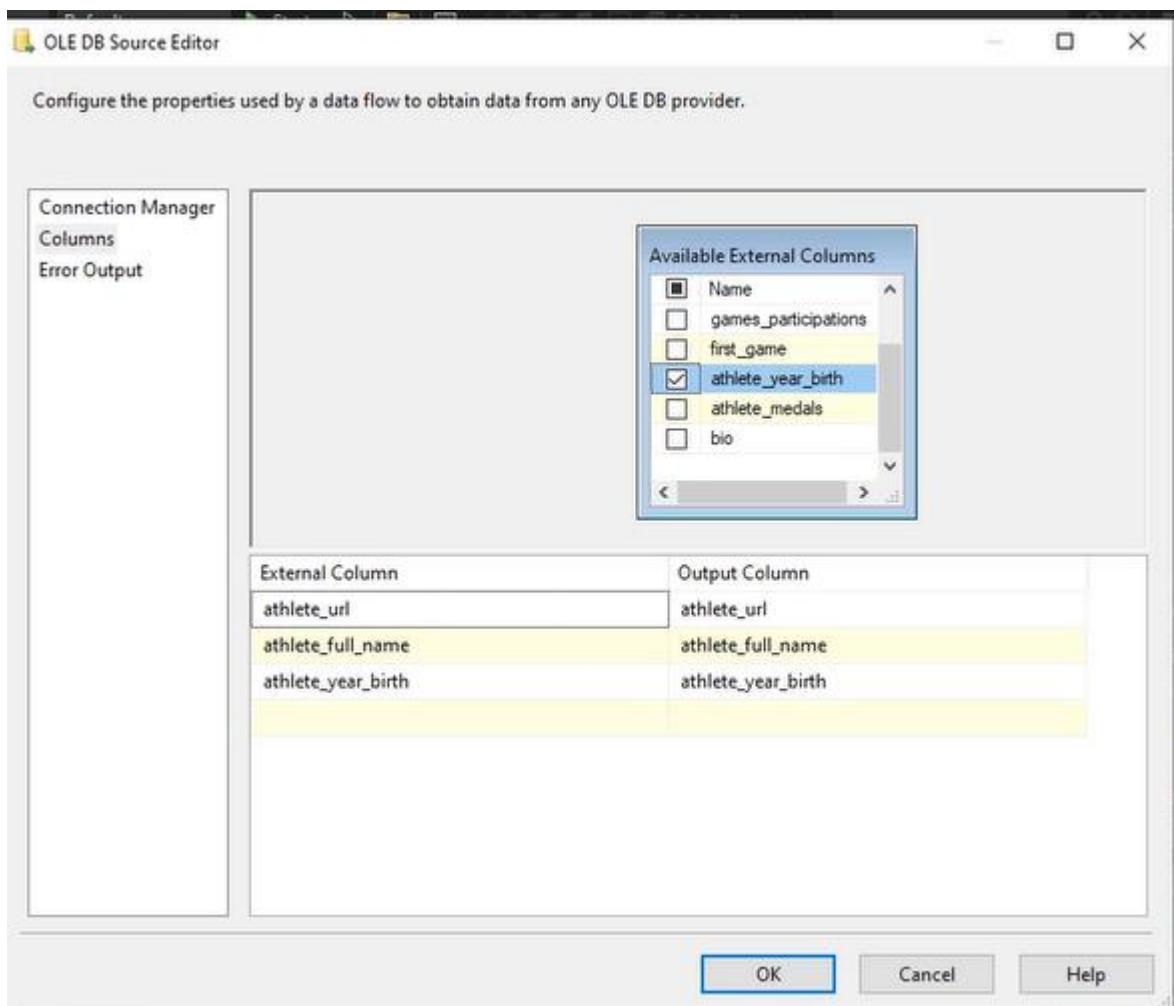
Remove rows with duplicate sort values.

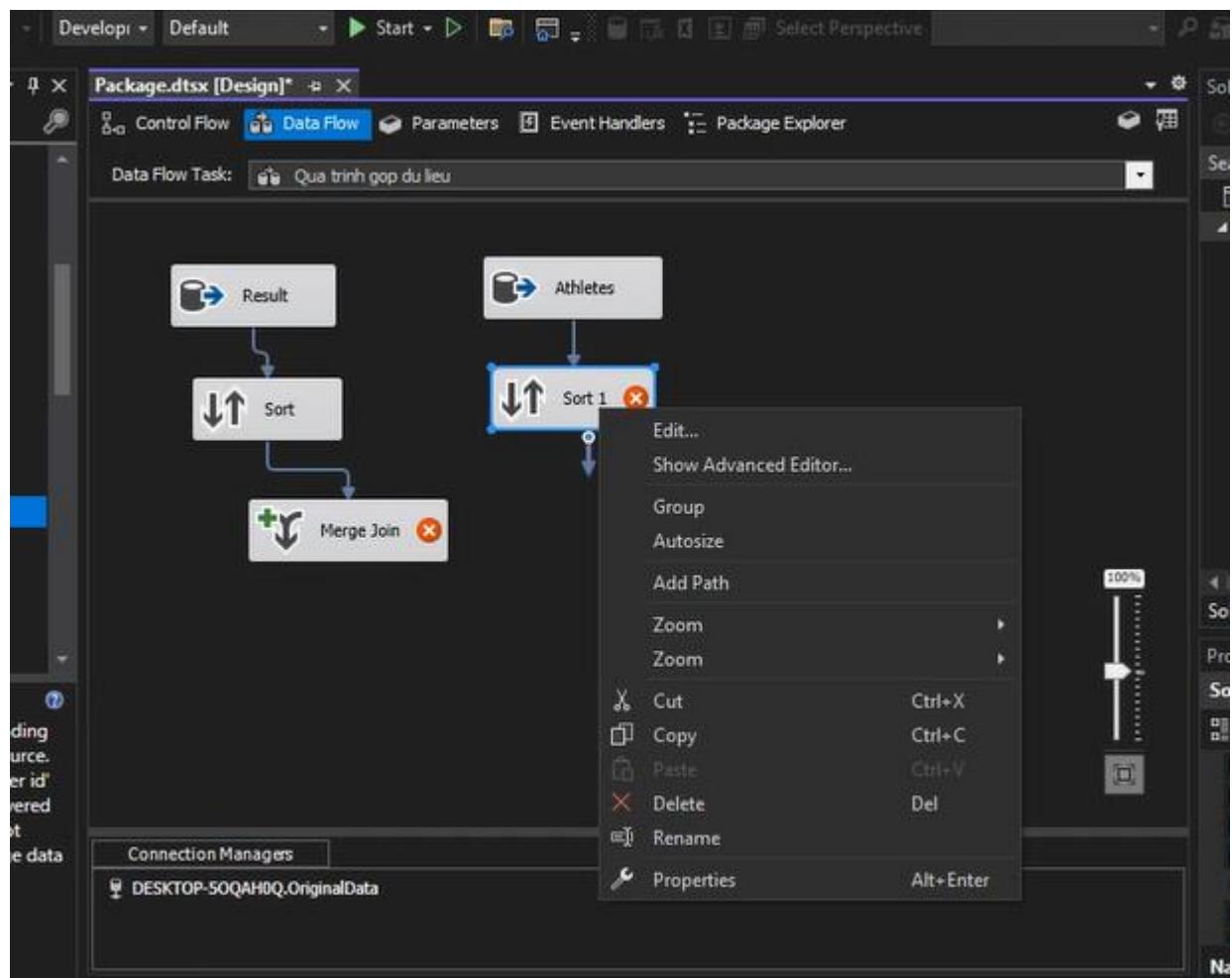
Available Input Columns	
<input type="checkbox"/>	Name
<input type="checkbox"/>	country_code
<input type="checkbox"/>	country_3_lett...
<input checked="" type="checkbox"/>	athlete_url
<input checked="" type="checkbox"/>	athlete_full_na...

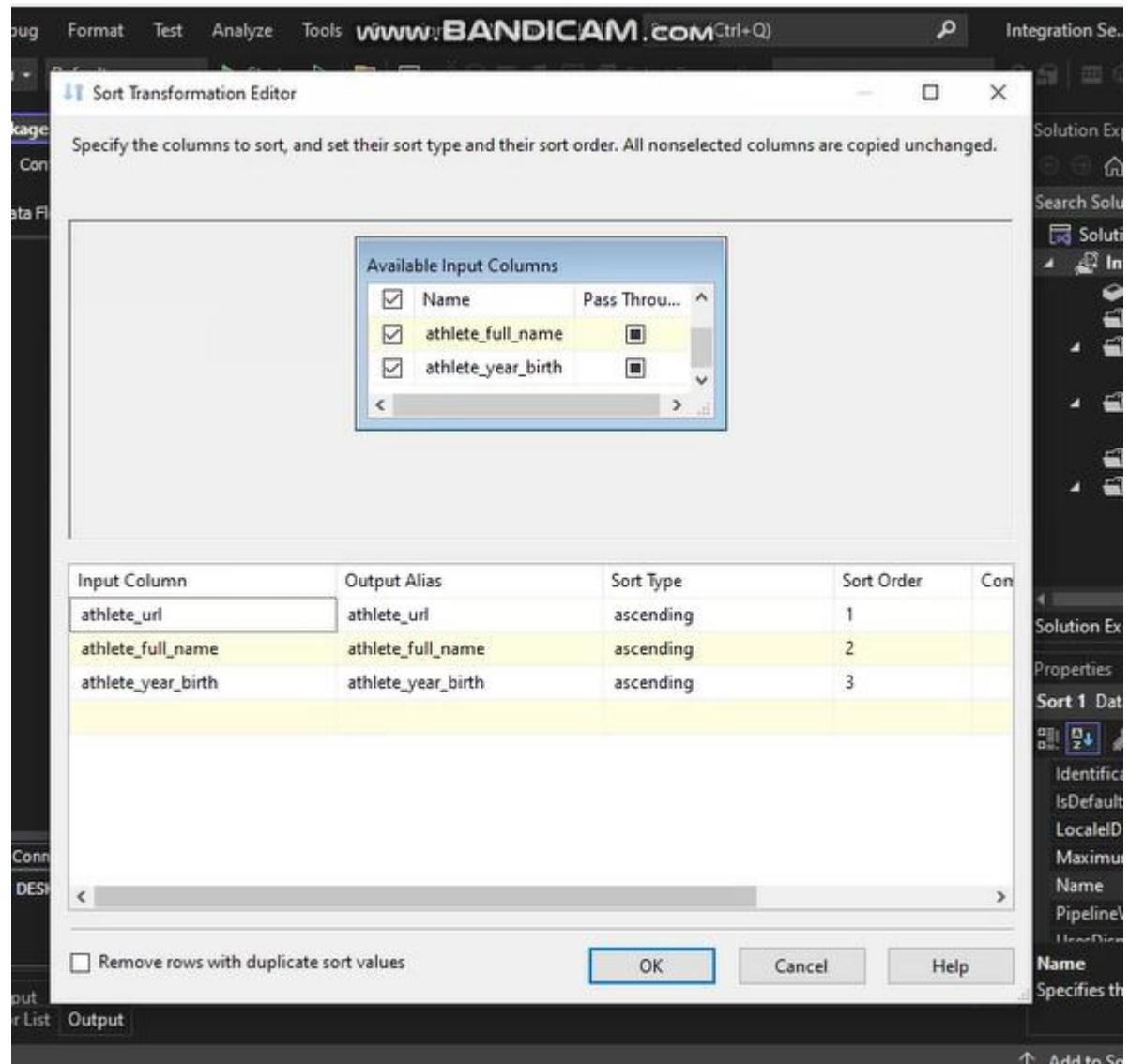
Column name: athlete_full_name
Data type: Unicode string [DT_WSTR]
Length: 100
Scale: 0
Precision: 0











Merge Join Transformation Editor

Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.

Join type: Inner join Swap Inputs

Sort

Name	Order	Join Key
discipline_title	0	<input type="checkbox"/>
event_title	0	<input type="checkbox"/>
slug_game	0	<input type="checkbox"/>
participant_type	0	<input type="checkbox"/>
medal_type	0	<input type="checkbox"/>
athletes	0	<input type="checkbox"/>
rank_equal	0	<input type="checkbox"/>
rank_position	0	<input type="checkbox"/>
country_name	0	<input type="checkbox"/>
country_code	0	<input type="checkbox"/>
country_3_lett...	0	<input type="checkbox"/>
athlete_url	1	<input checked="" type="checkbox"/>

Sort 1

Name	Order	Join Key
athlete_url	1	<input type="checkbox"/>
athlete_full_name	2	<input checked="" type="checkbox"/>
athlete_year_birth	3	<input type="checkbox"/>

Input **Input Column** **Output Alias**

Sort	discipline_title	discipline_title
Sort	event_title	event_title
Sort	slug_game	slug_game
Sort	participant_type	participant_type
Sort	medal_type	medal_type
Sort	athletes	athletes

OK Cancel Help

Package.dtsx [Design] * Data Flow Parameters Event Handlers Package Explorer

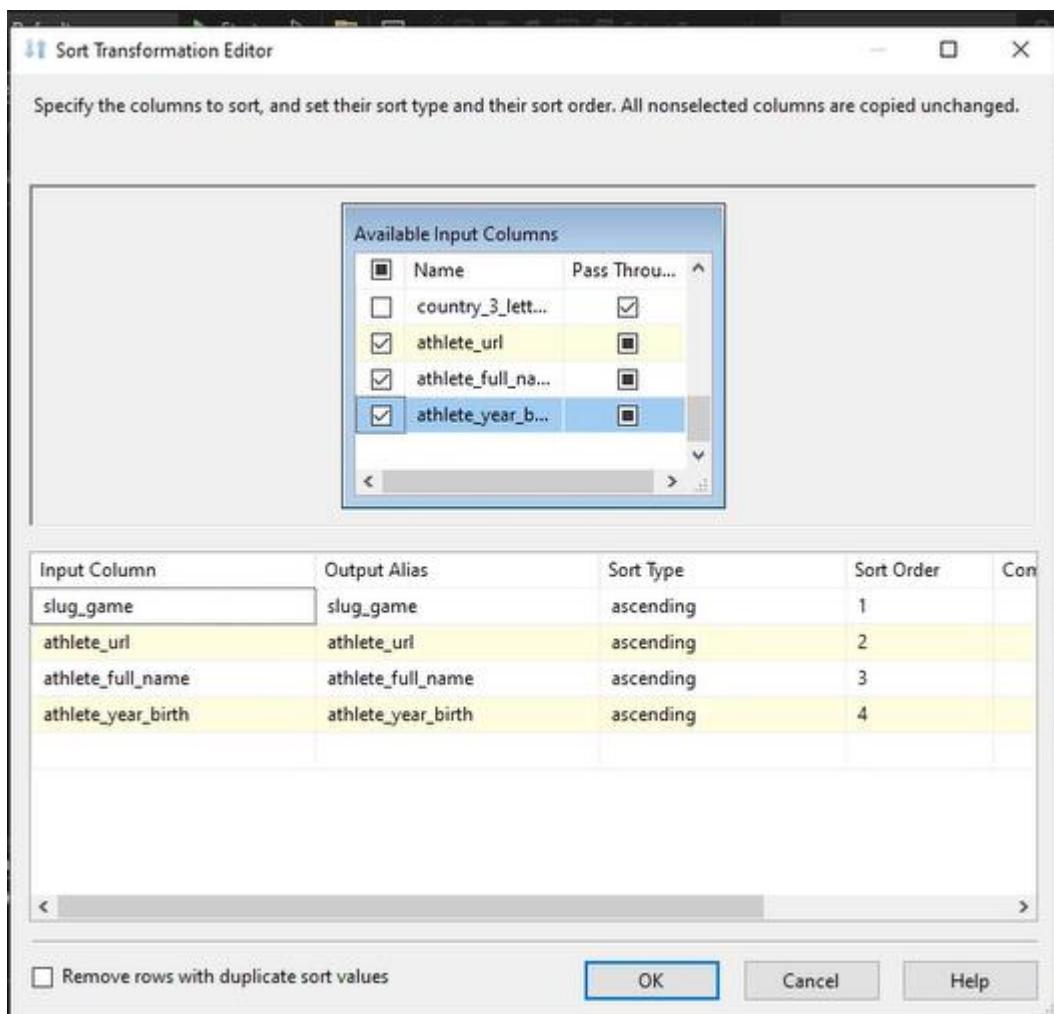
Data Flow Task: Quá trình gộp dữ liệu

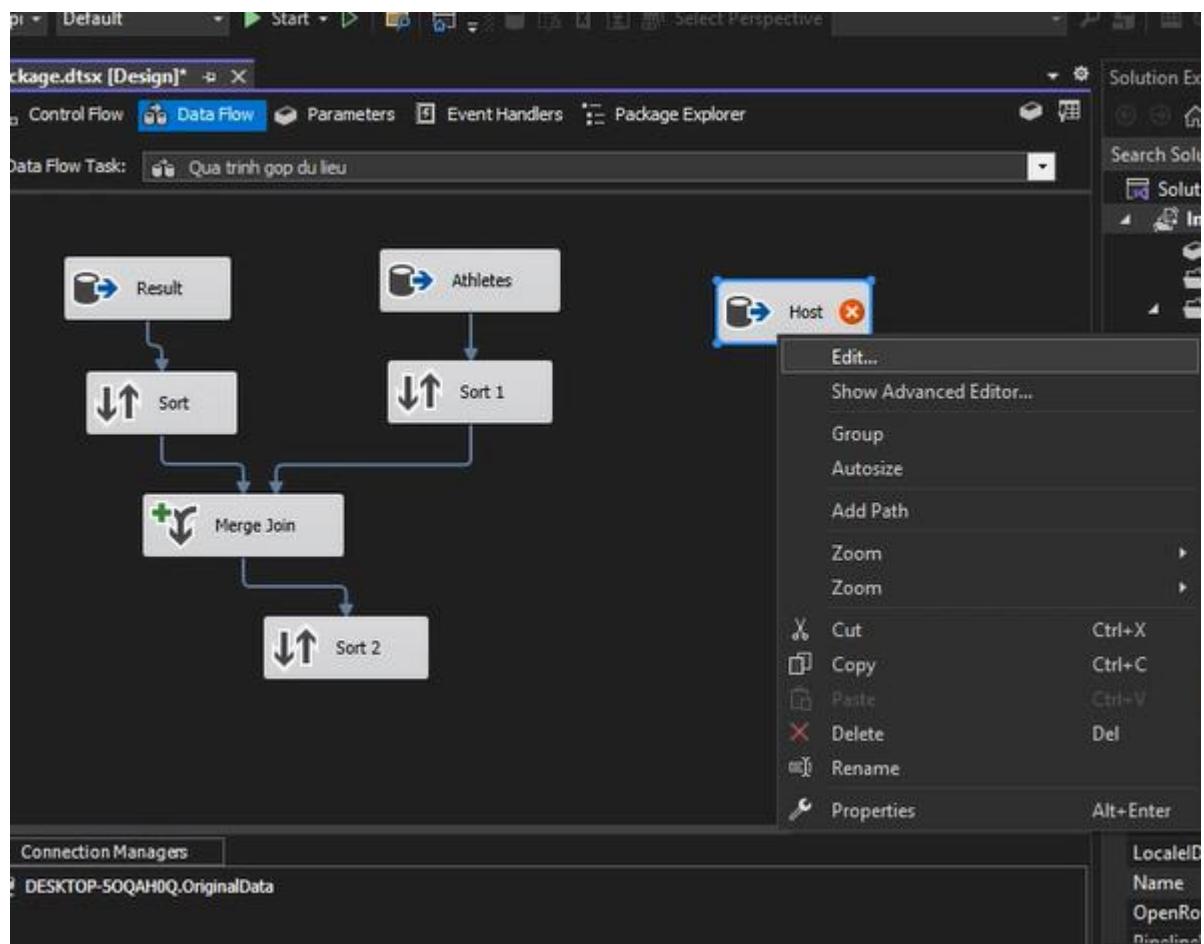
```

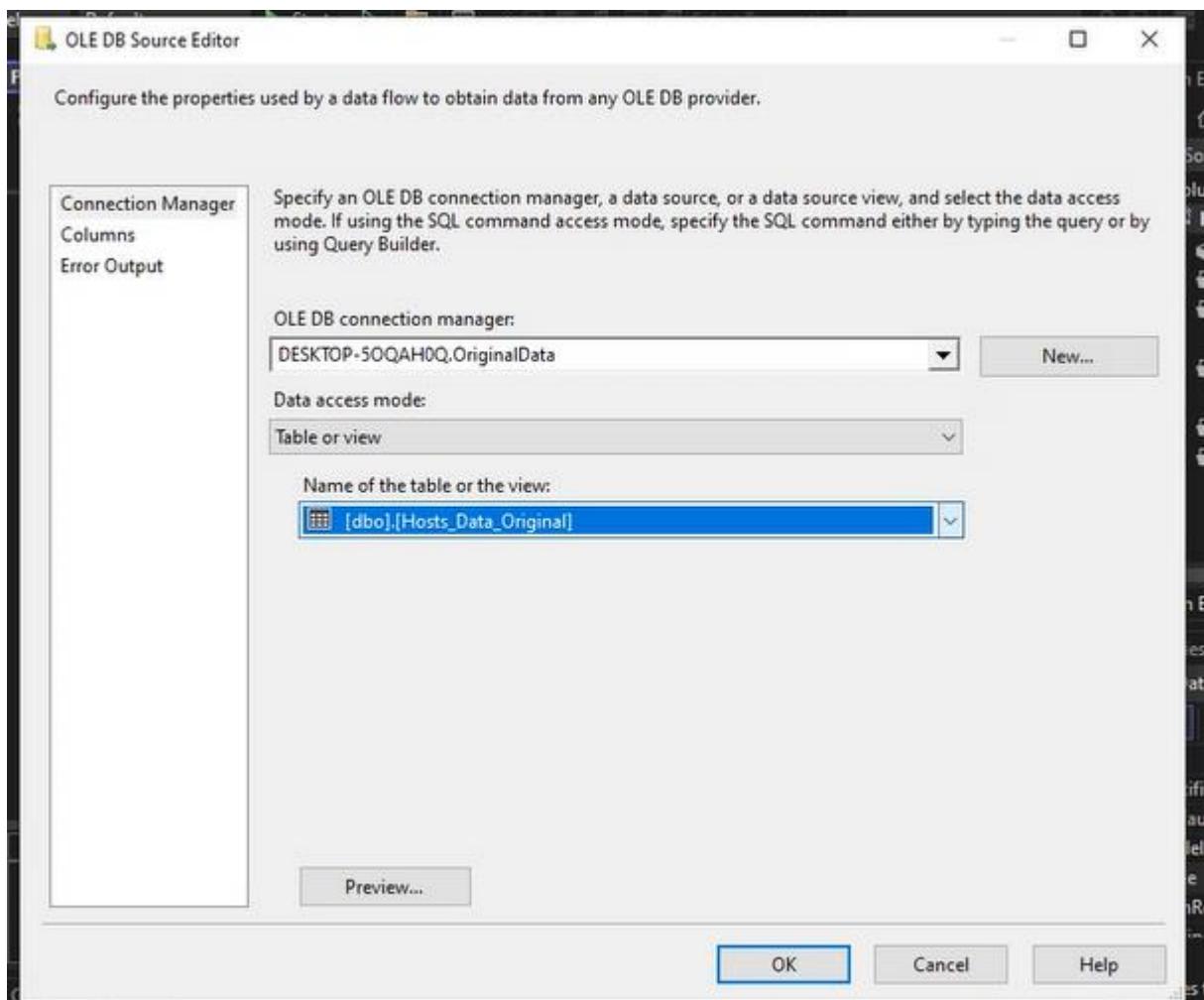
graph TD
    Result[Result] --> Sort[Sort]
    Sort --> MergeJoin[Merge Join]
    Athletes[Athletes] --> Sort1[Sort 1]
    Sort1 --> MergeJoin
    MergeJoin --> Sort2[Sort 2]
    Sort2 --> End(( ))
  
```

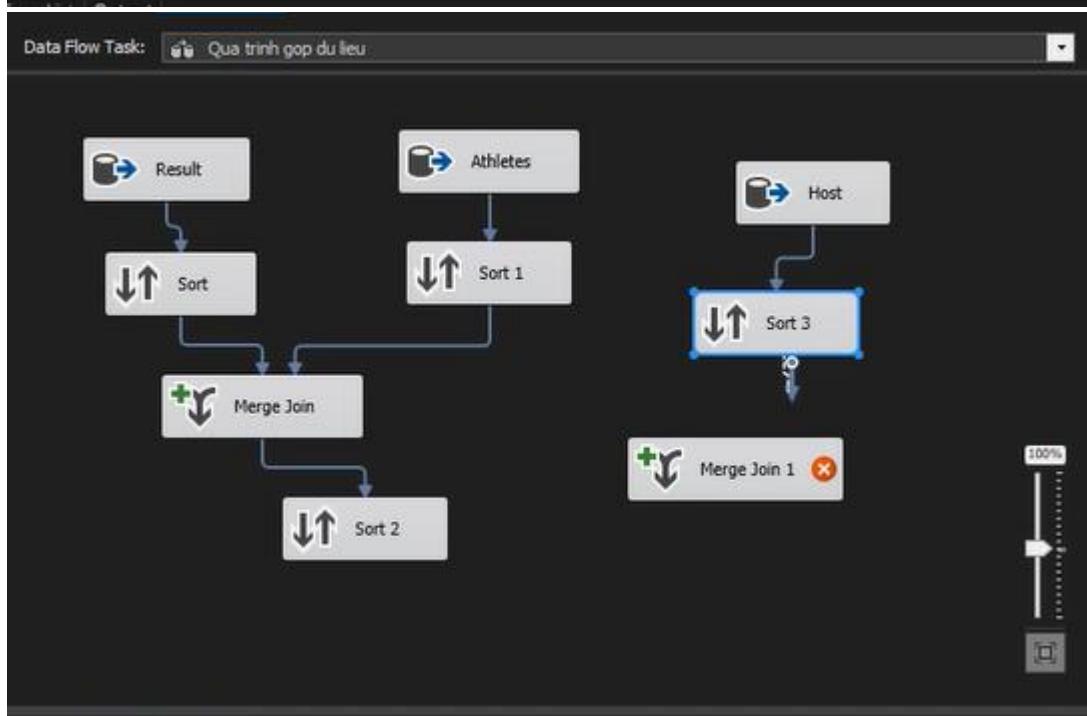
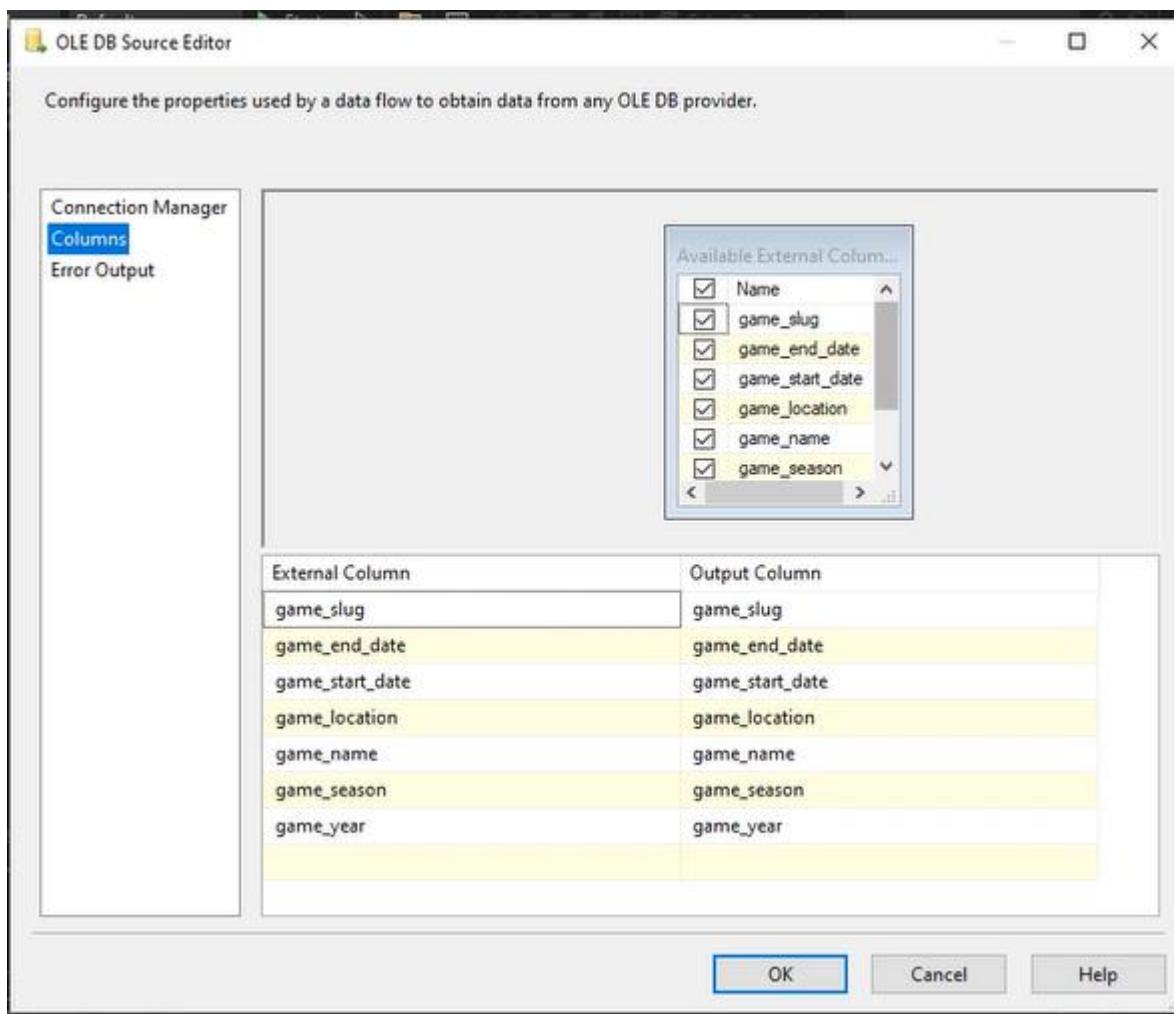
Connection Managers

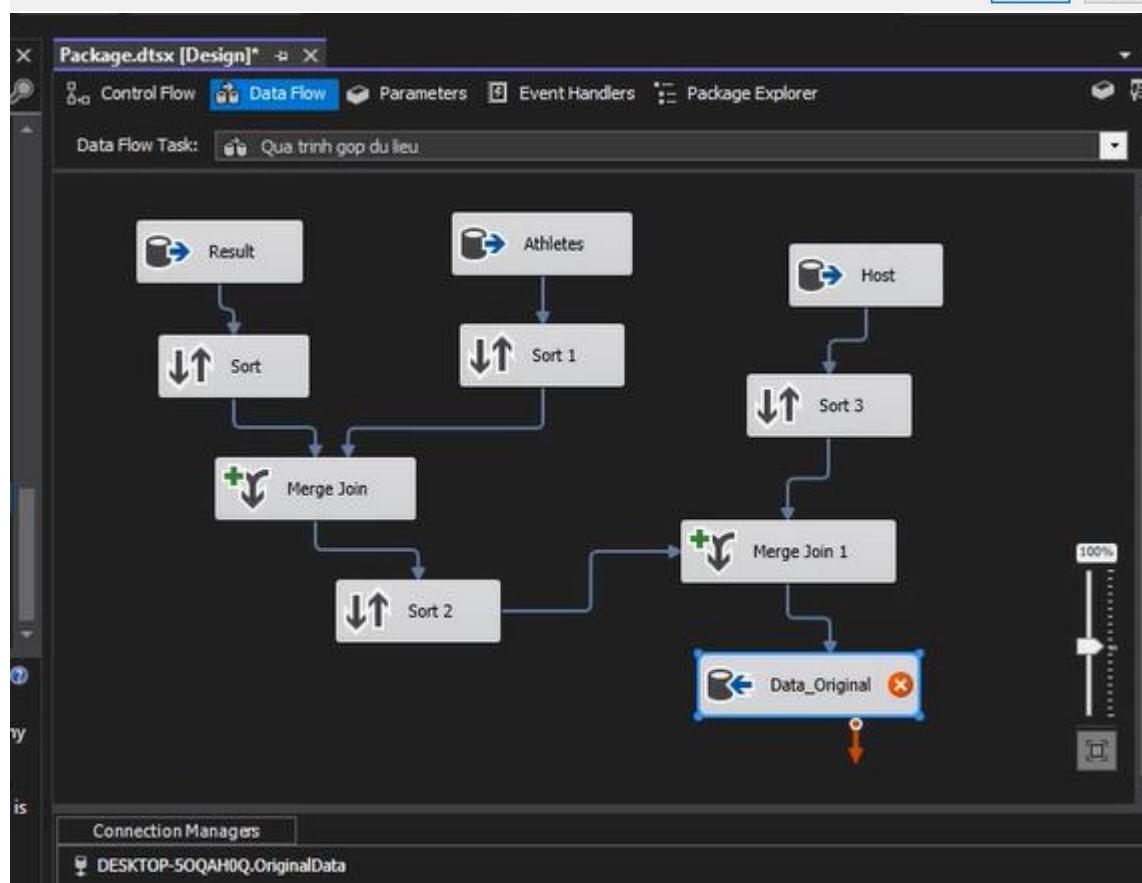
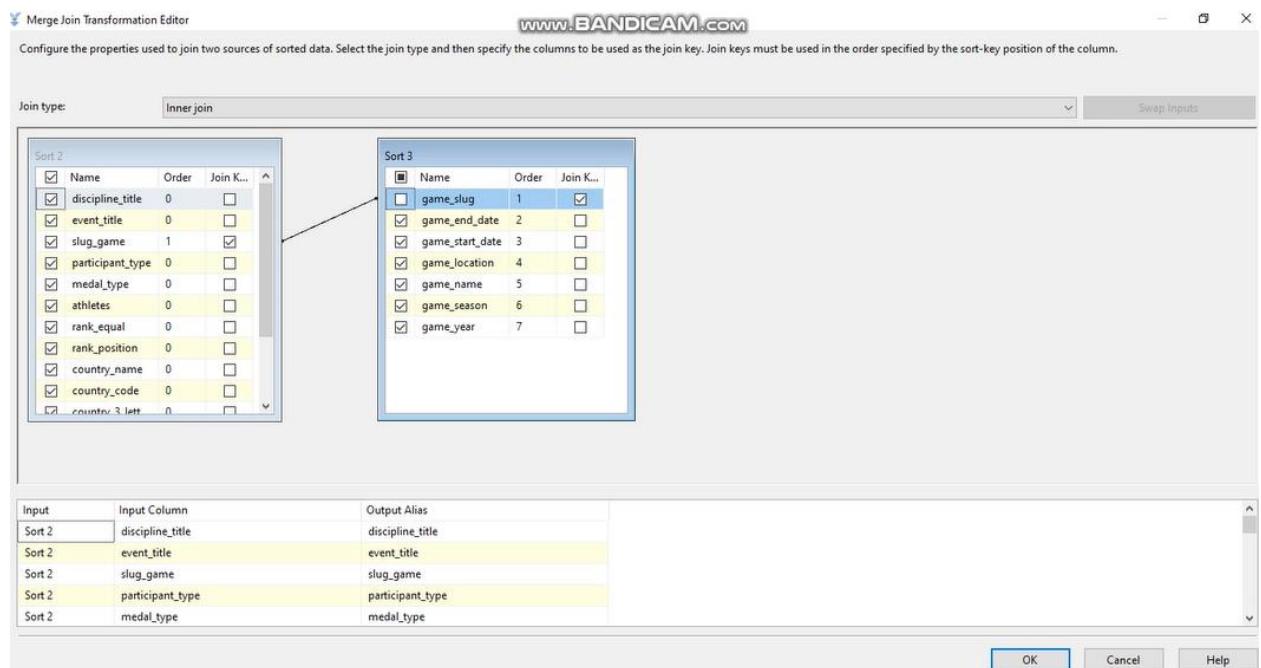
- DESKTOP-50QAH0Q.OriginalData

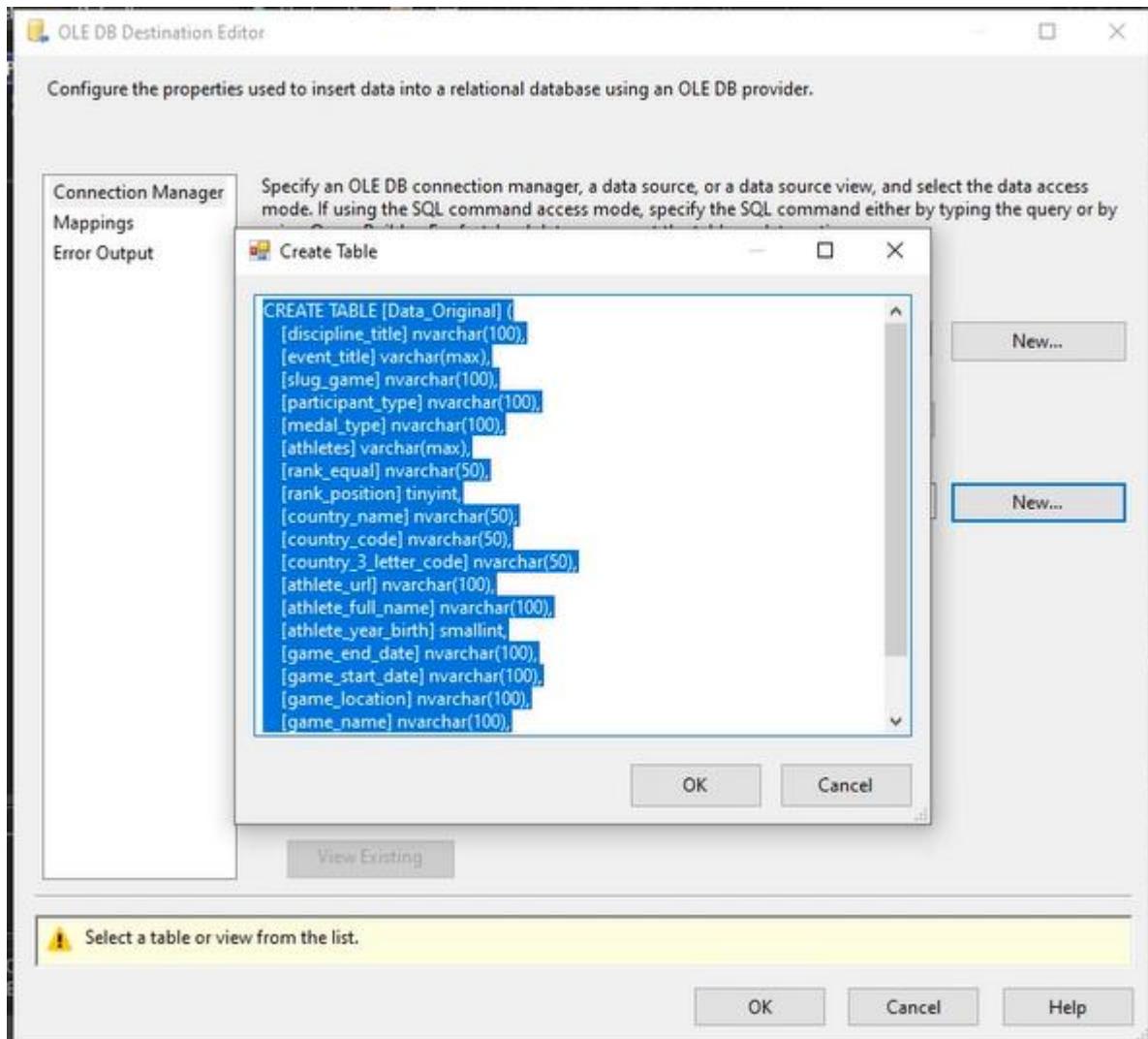






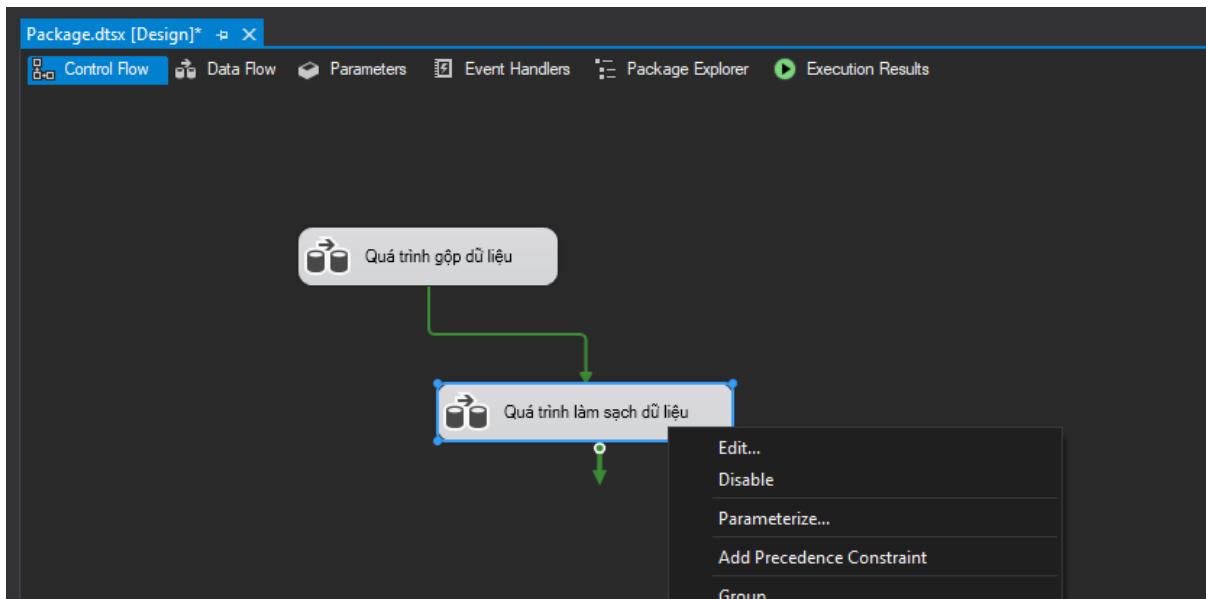




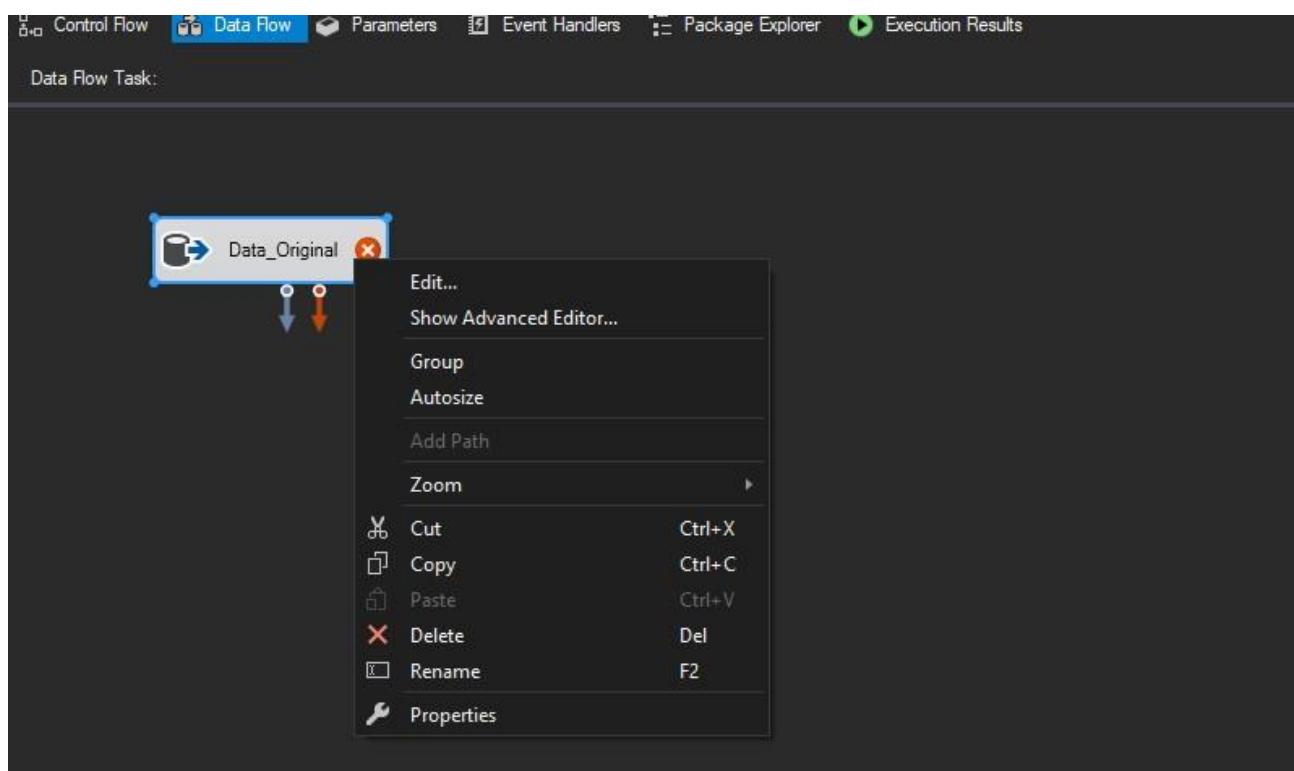


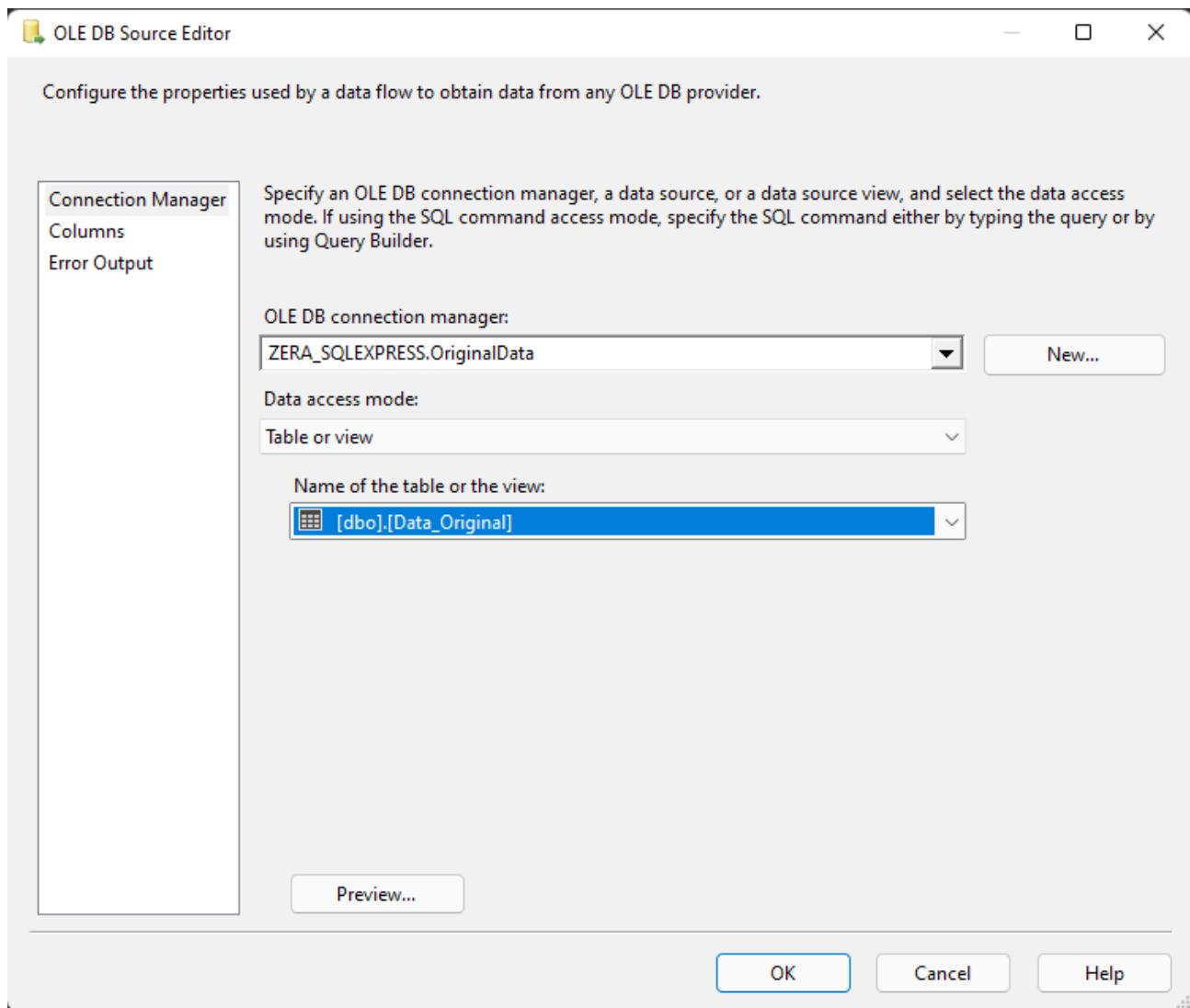
Quá trình làm sạch dữ liệu.

- Tạo mới một Data Flow Task và đặt tên là “Lam sach dữ liệu”. Trong Task này ta sẽ tiến hành làm sạch dữ liệu, loại bỏ các thuộc tính không cần thiết, những dữ liệu không xác định.

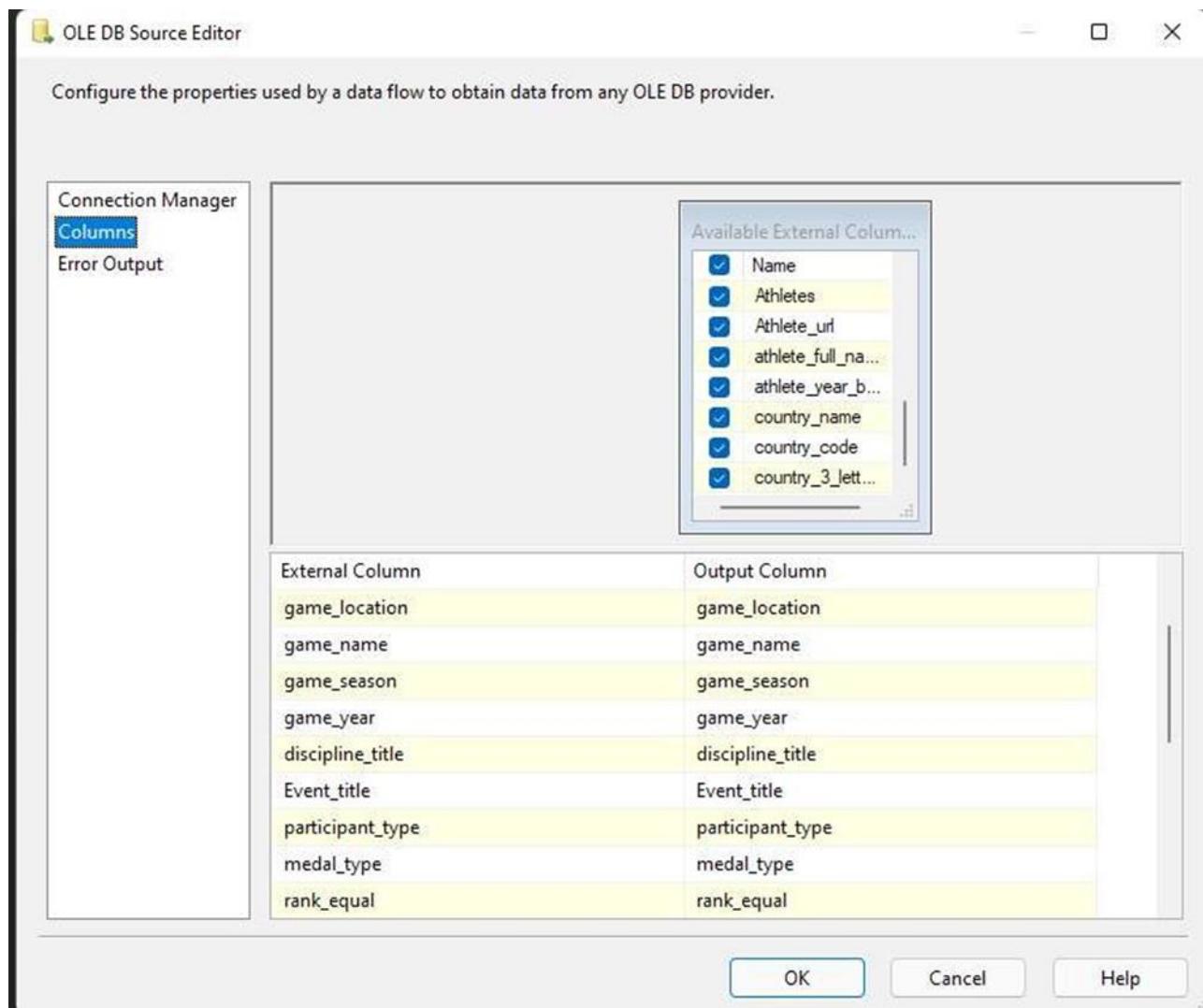


- Dữ liệu mà ta sử dụng để làm sạch chính là bảng Data_Original mà ta đã tiến hành gộp ở bước trên.

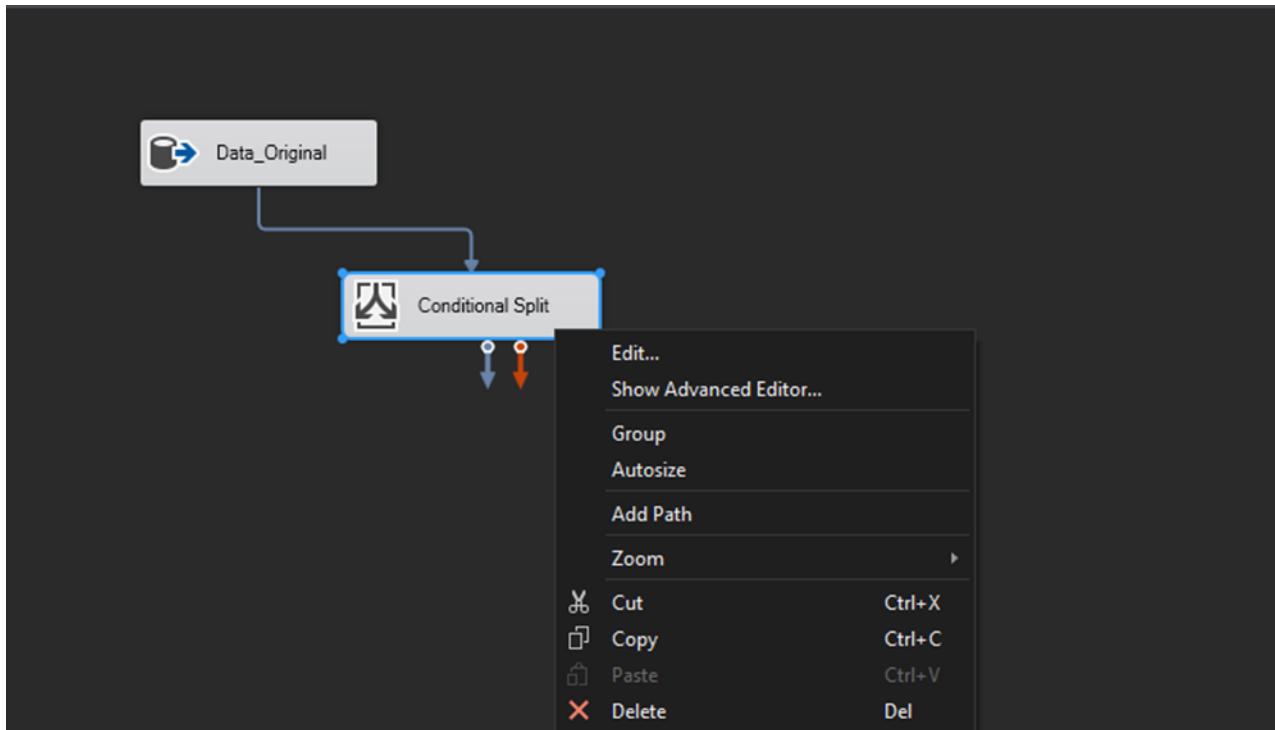




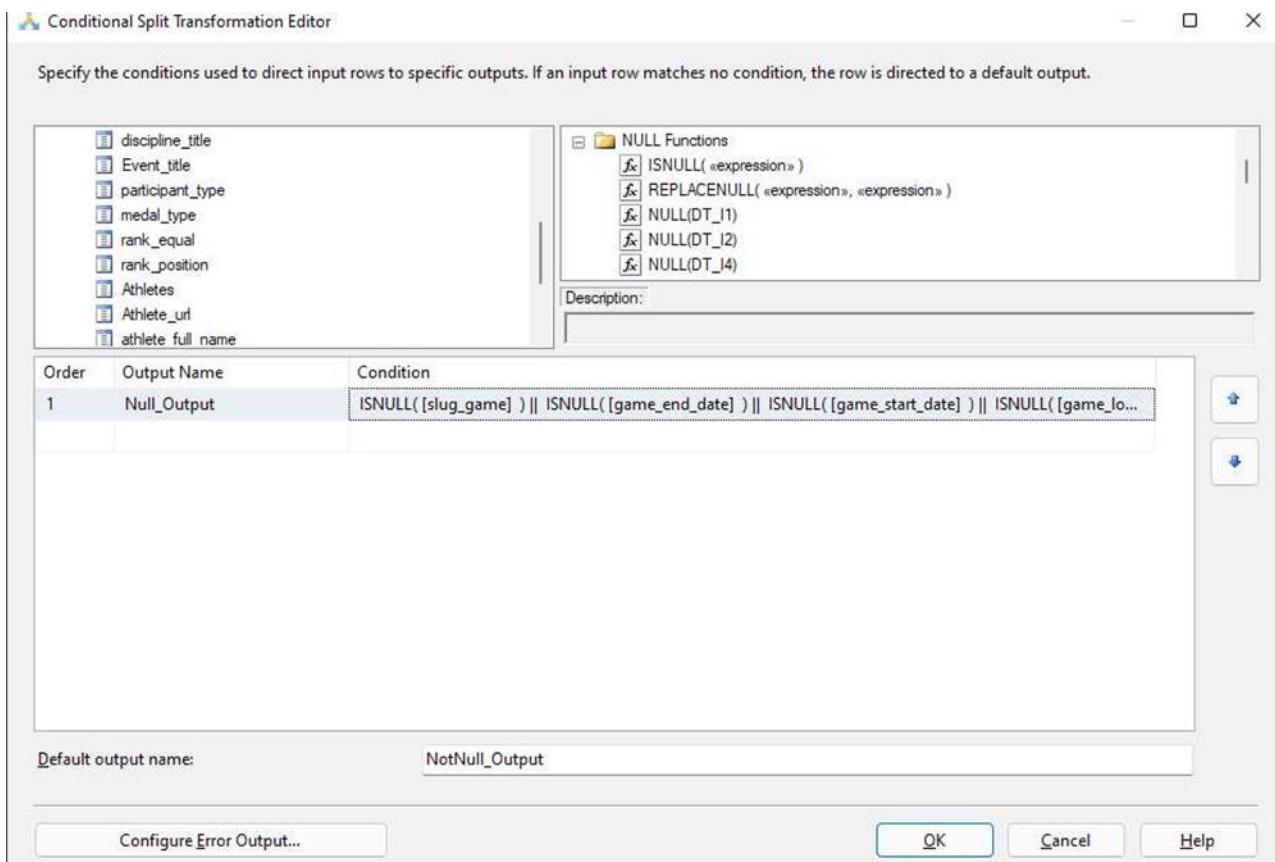
- Trong “OLE DB Source Editor” ở phần Columns chọn tất cả các cột rồi chọn OK.



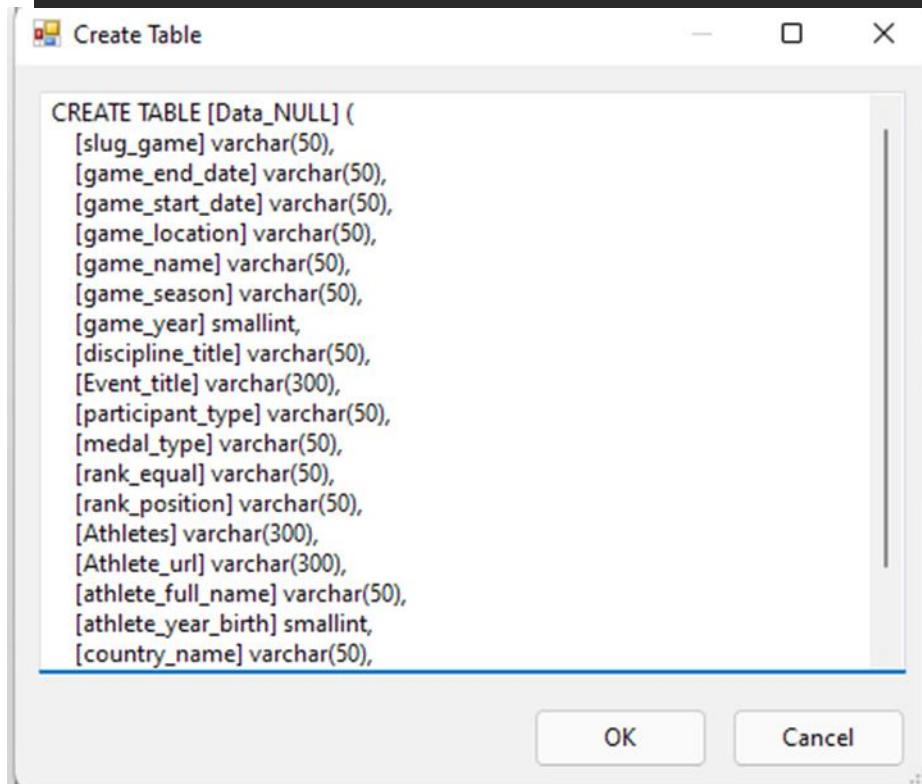
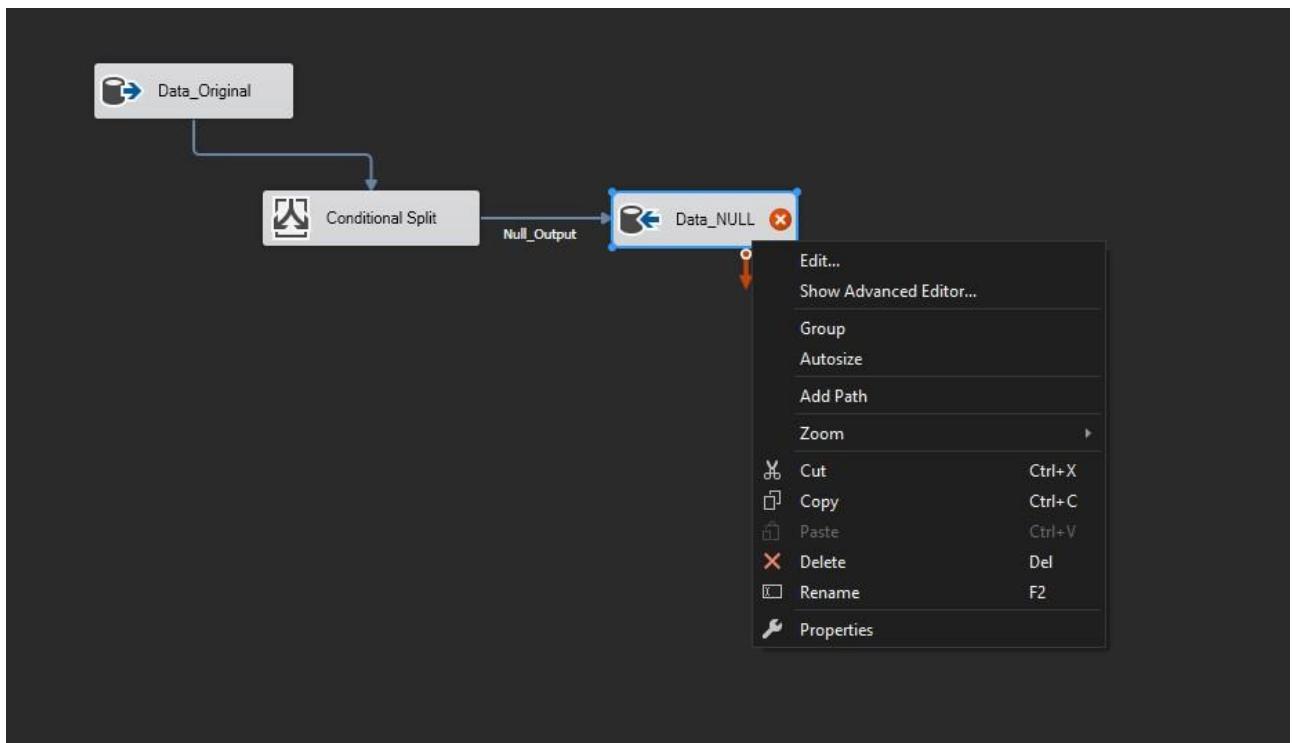
- Tạo và cấu hình “Condition Split” để lọc dữ liệu NULL
- Kéo thả công cụ “Condition Split” ở “SSIS Toolbox”. Click vào và chọn Edit.

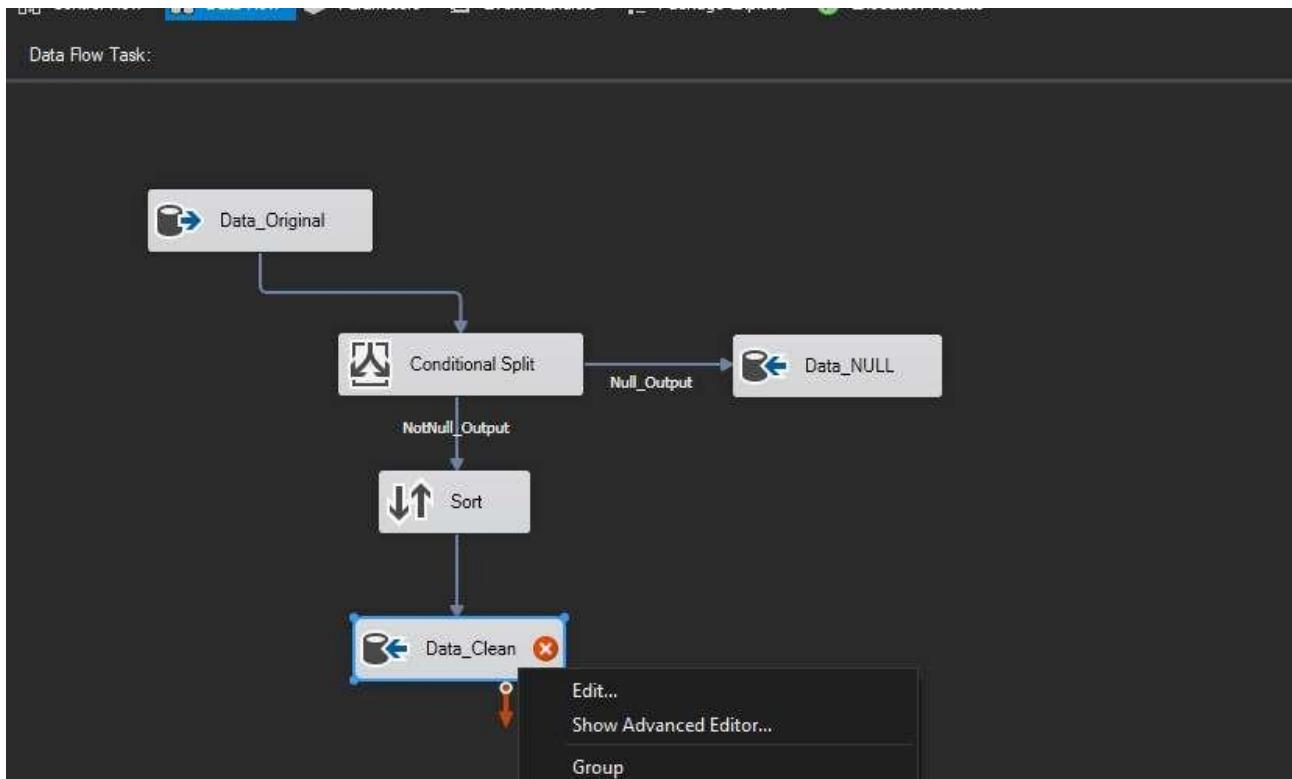


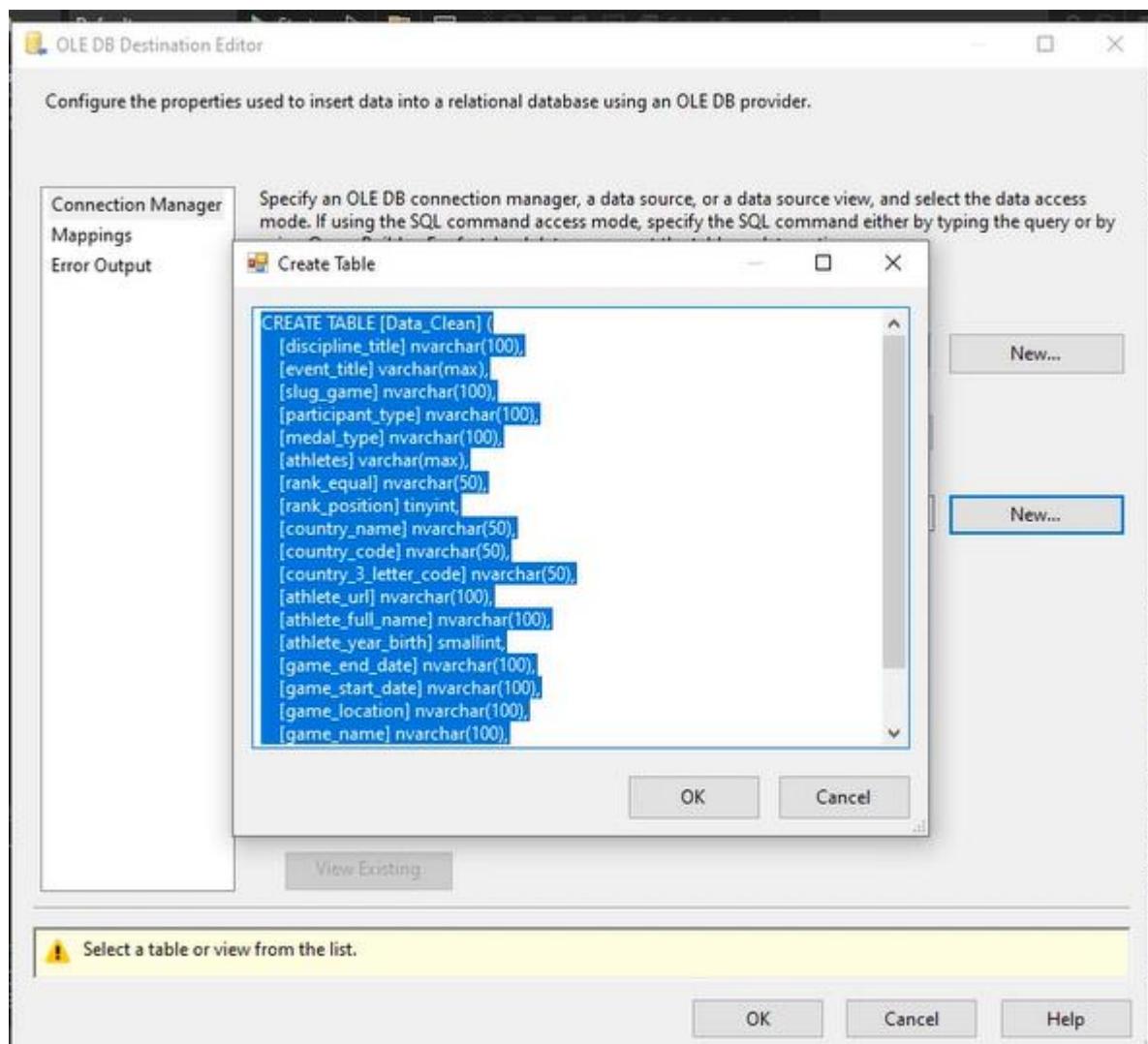
- Sử dụng hàm ISNULL để lọc NULL



- Dữ liệu Null và Dữ liệu không Null sẽ được lưu vào hai bảng riêng

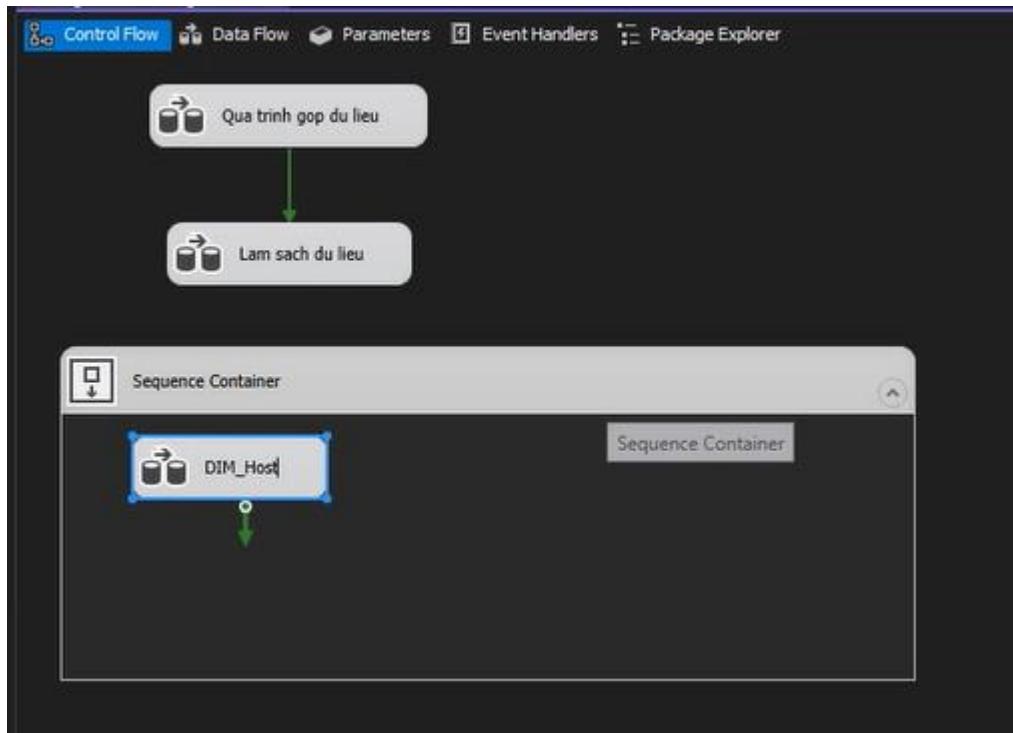


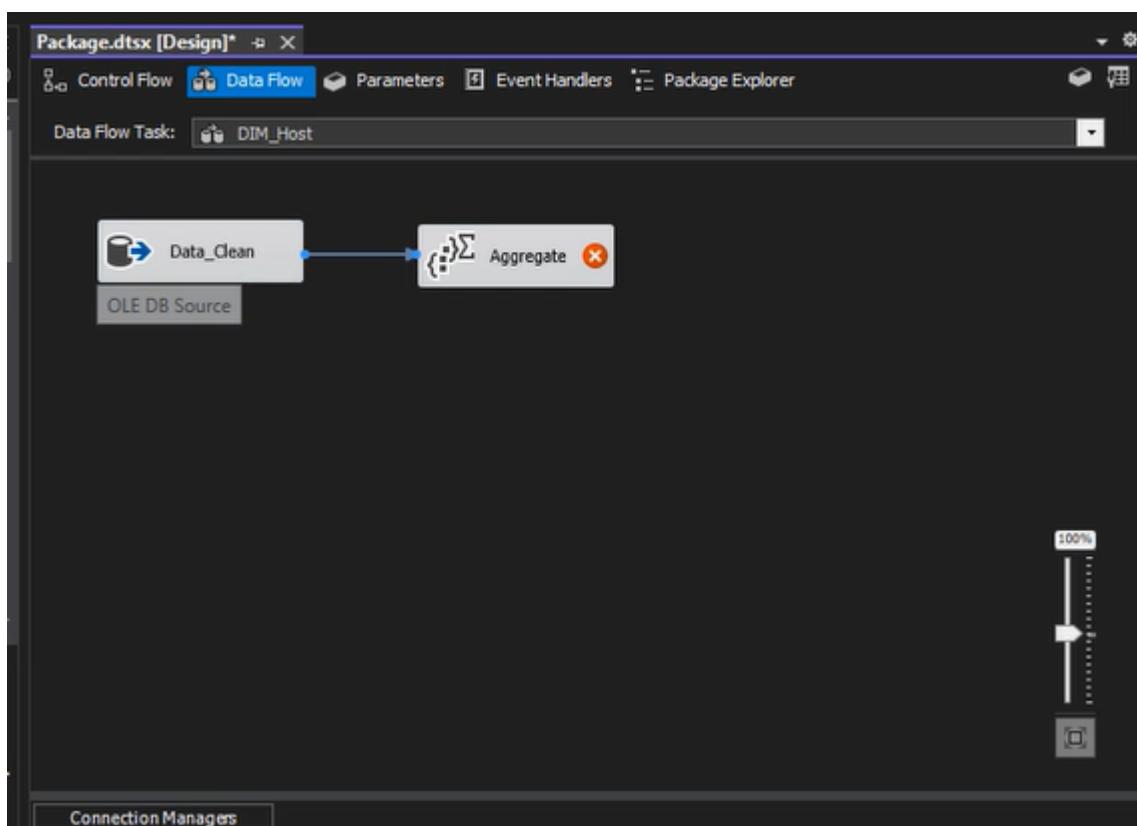
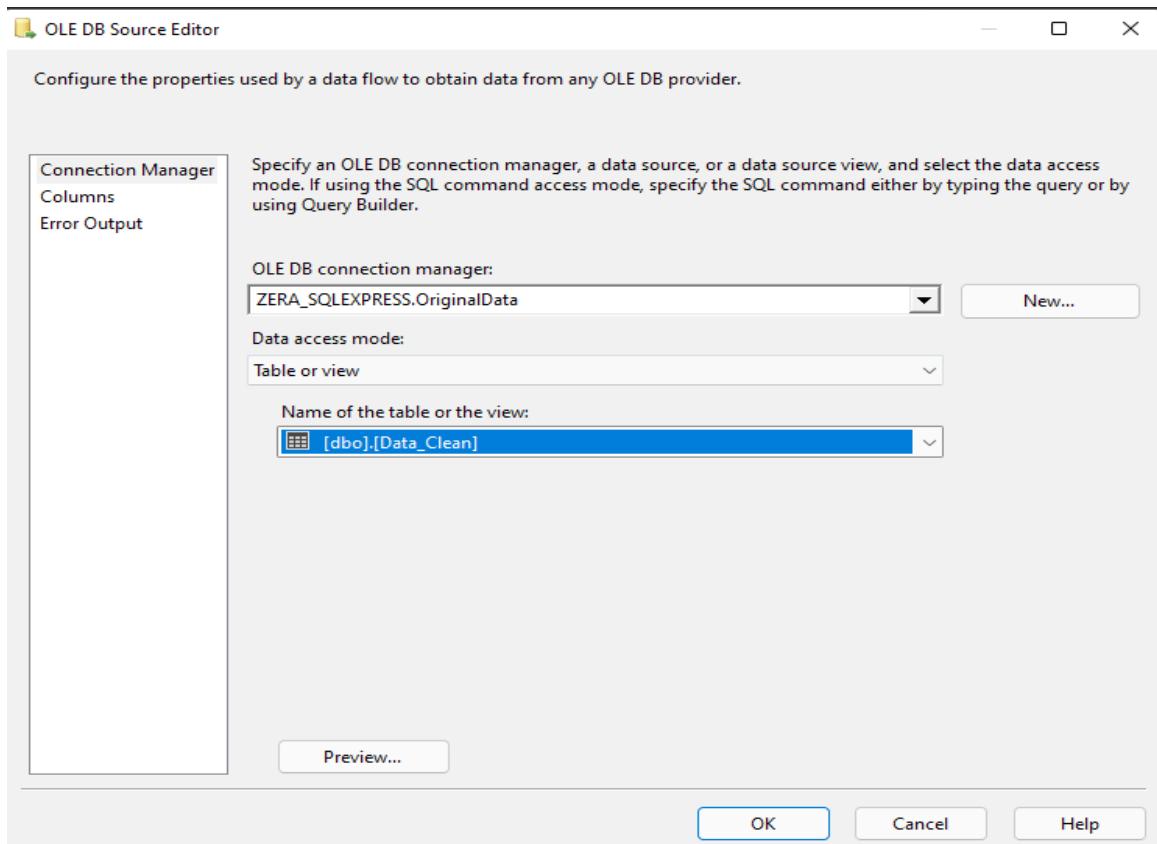




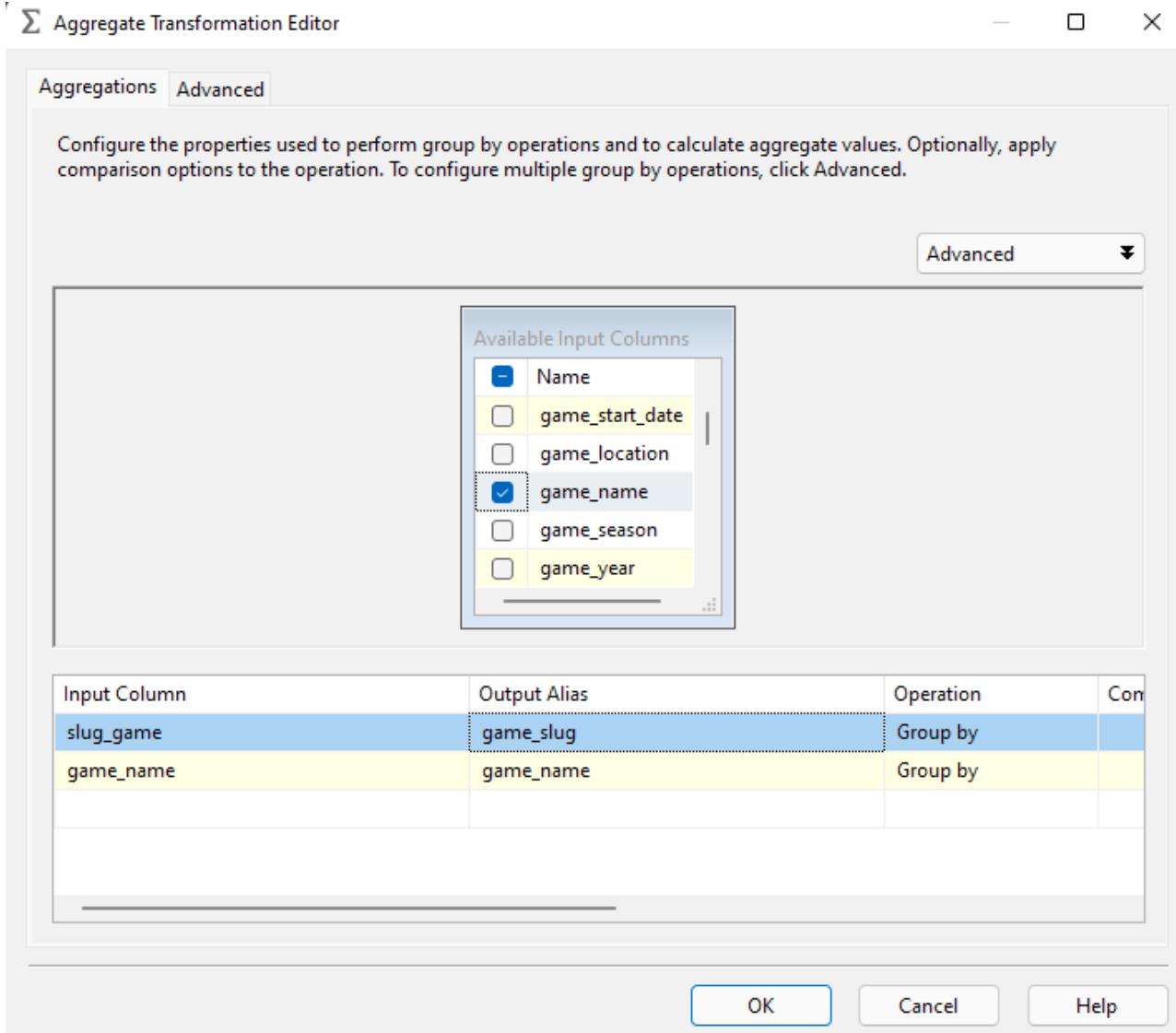
Quá trình tạo các bảng Dimension.

2.2.7.1 Tạo bảng DIM_Host





- Với DIM_Host, ta chỉ cần 2 thuộc tính : game_slug và game_name



- Cấu hình và sort dữ liệu

Sort Transformation Editor

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

Available Input Columns			
	Name	Pass Thru...	
<input checked="" type="checkbox"/>	Name		-
<input checked="" type="checkbox"/>	game_slug		-
<input checked="" type="checkbox"/>	game_name		-

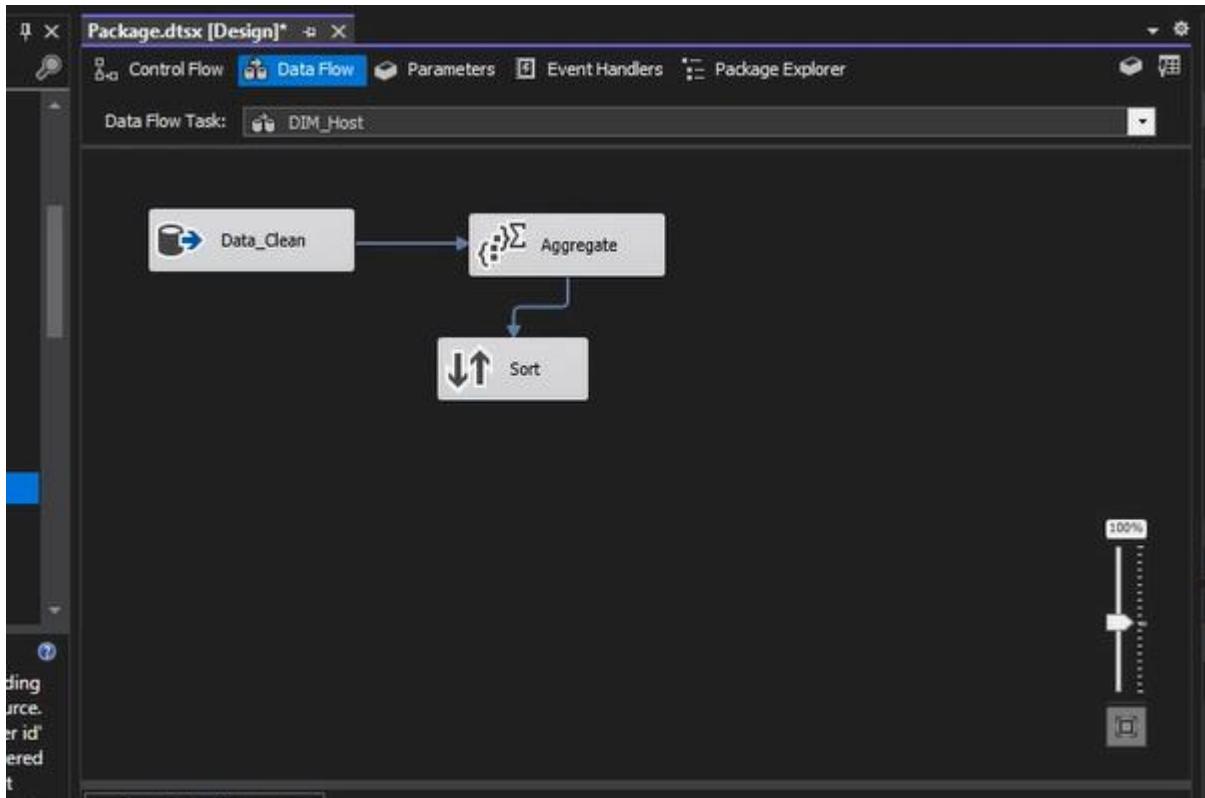
Input Column	Output Alias	Sort Type	Sort Order	Con
game_slug	game_slug	ascending	1	
game_name	game_name	ascending	2	

Remove rows with duplicate sort values

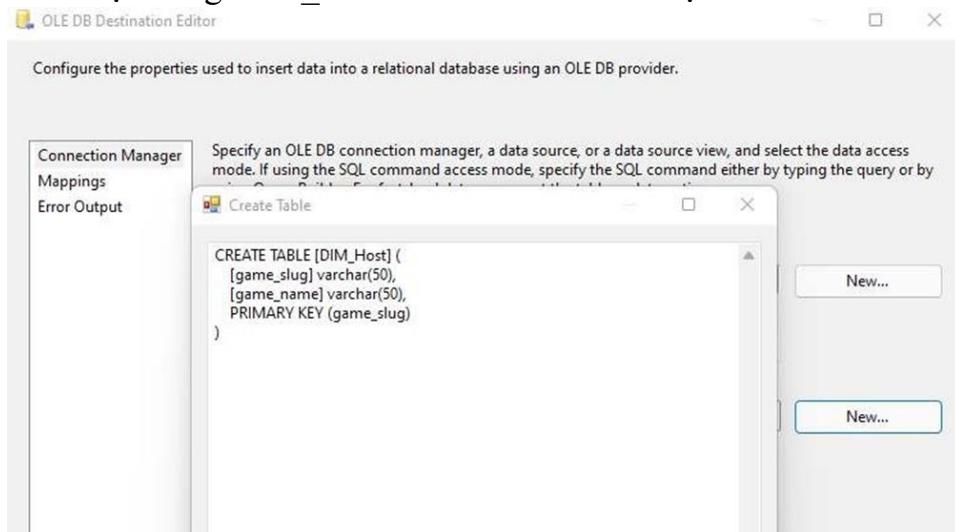
OK

Cancel

Help

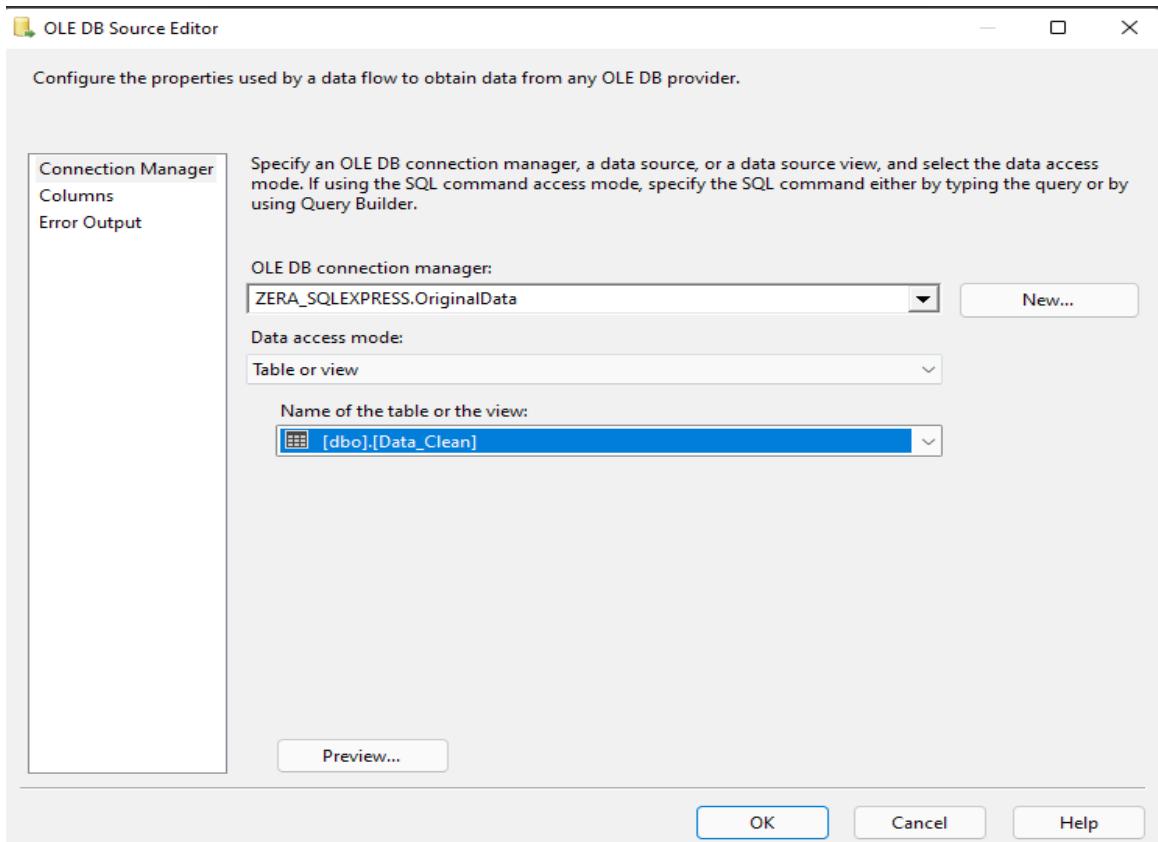
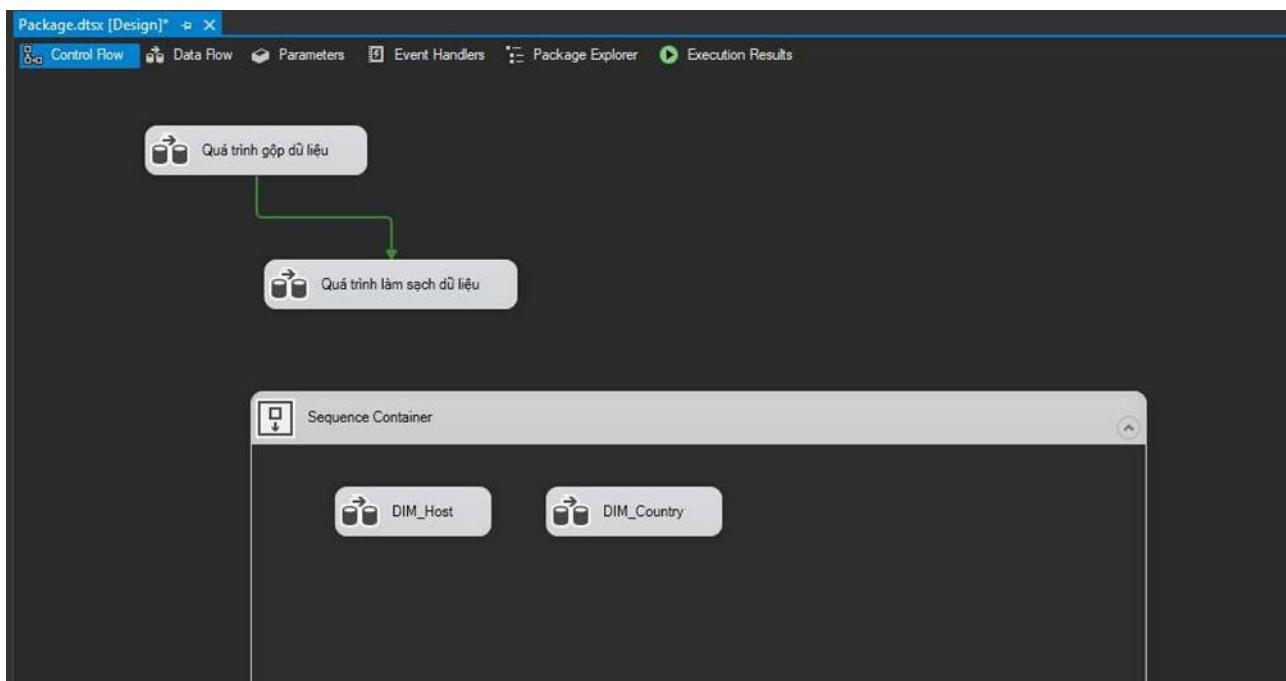


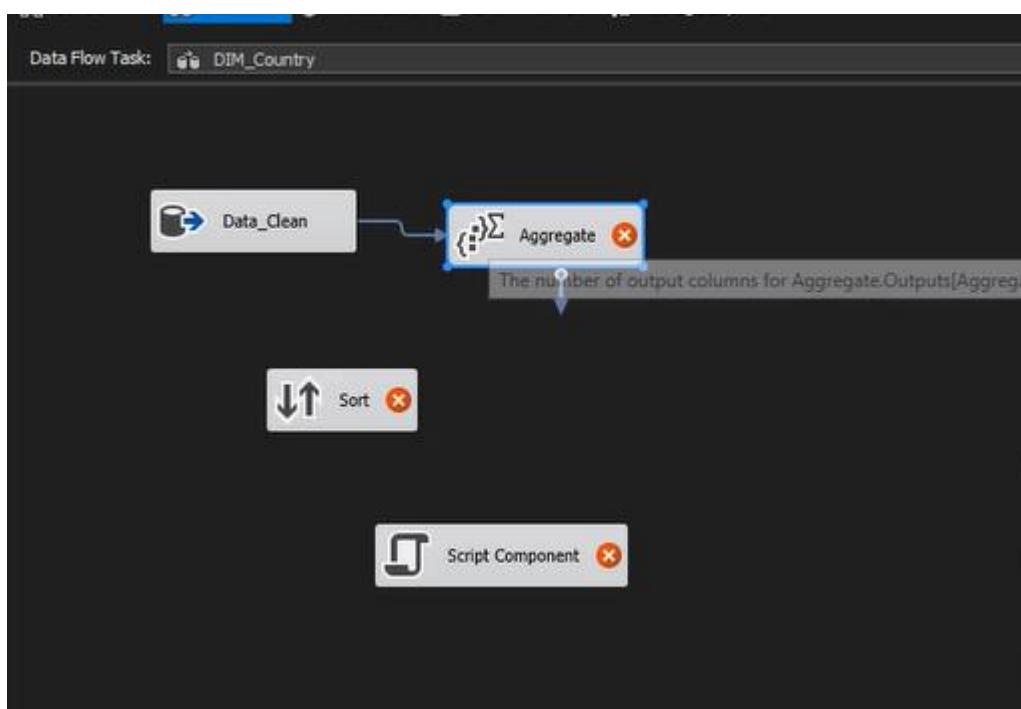
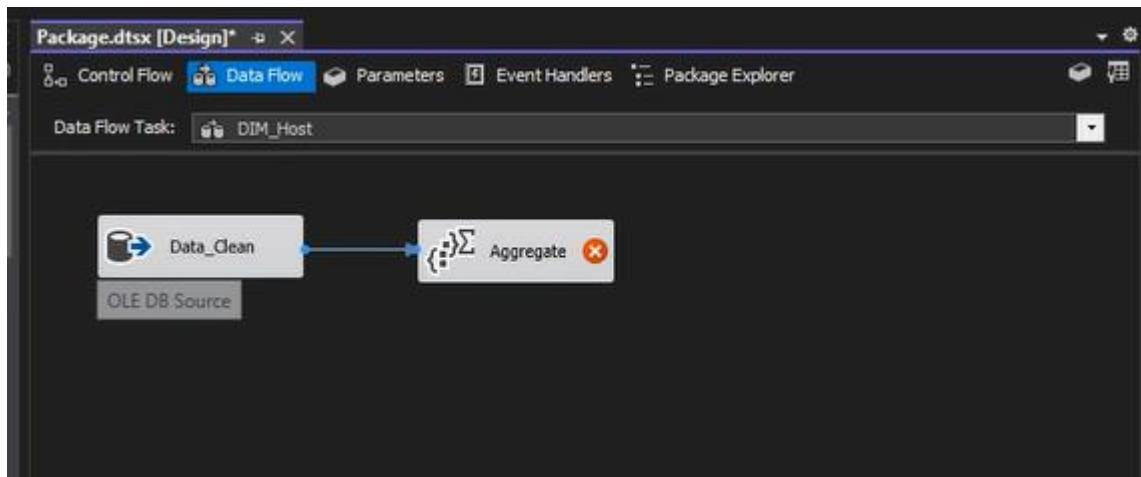
- Tạo bảng DIM_Host và lưu với kho dữ liệu

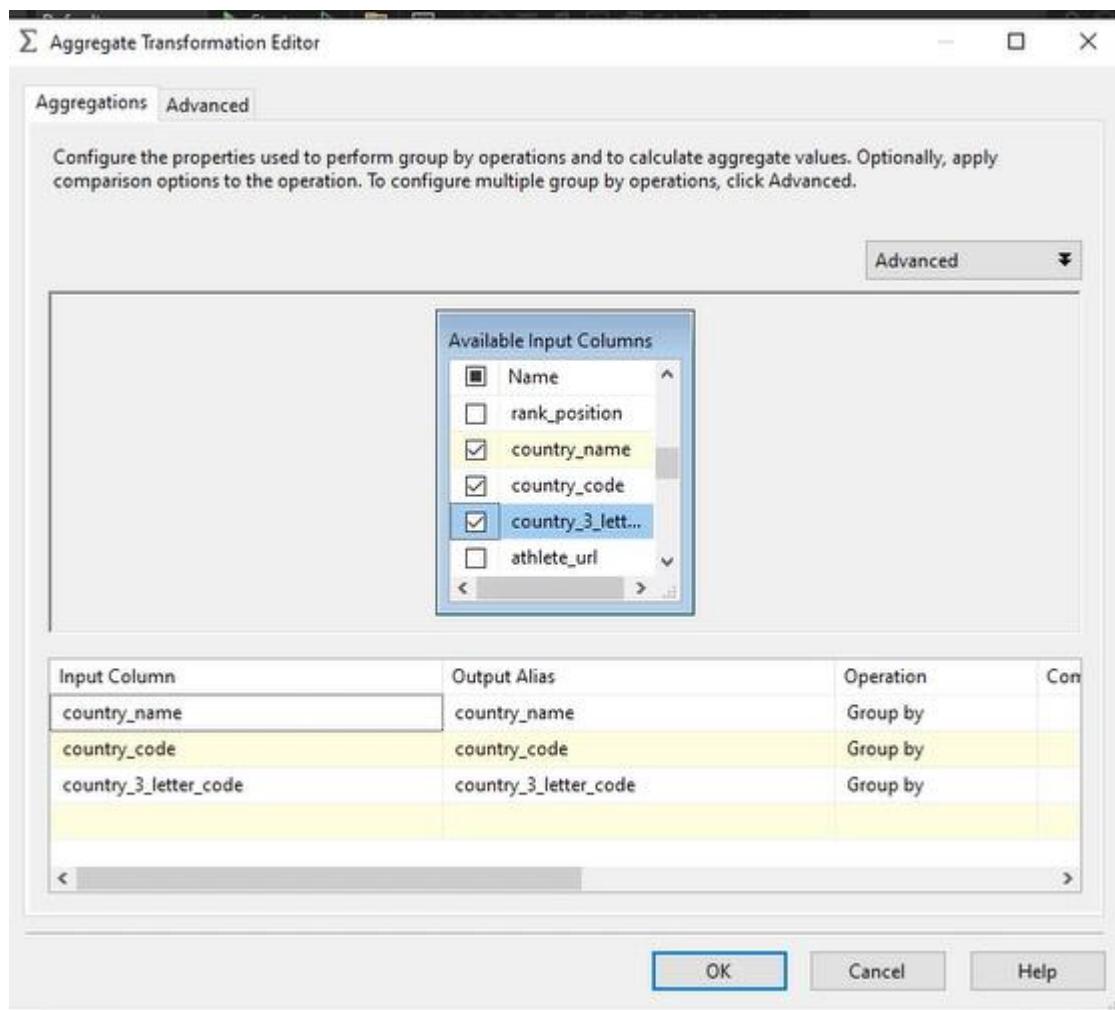


2.2.7.2 Tạo bảng DIM_Country

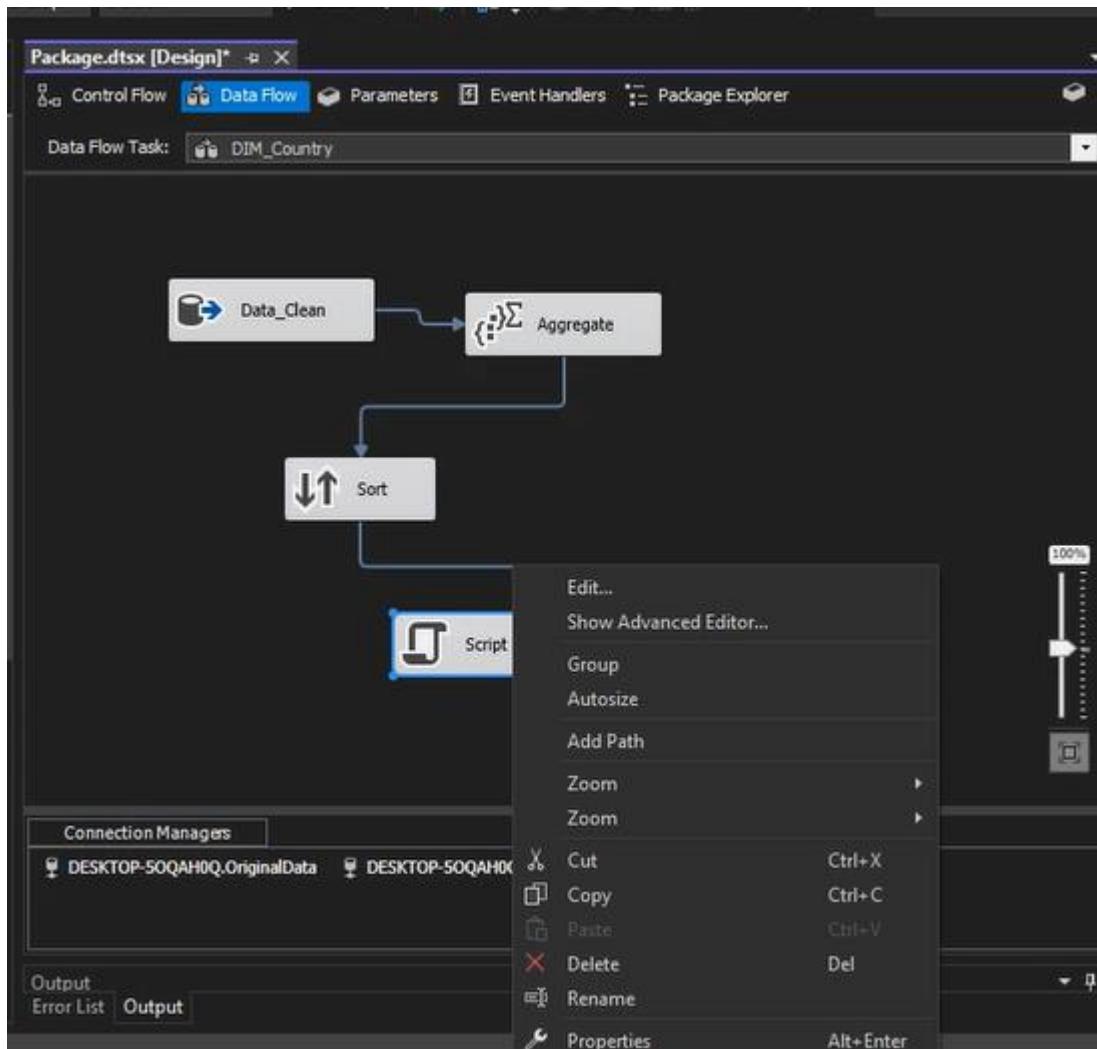
- Tương tự như trên, ta cũng sẽ tạo một Data Flow Task để tiến hành tạo với lưu DIM_Country vào kho dữ liệu.

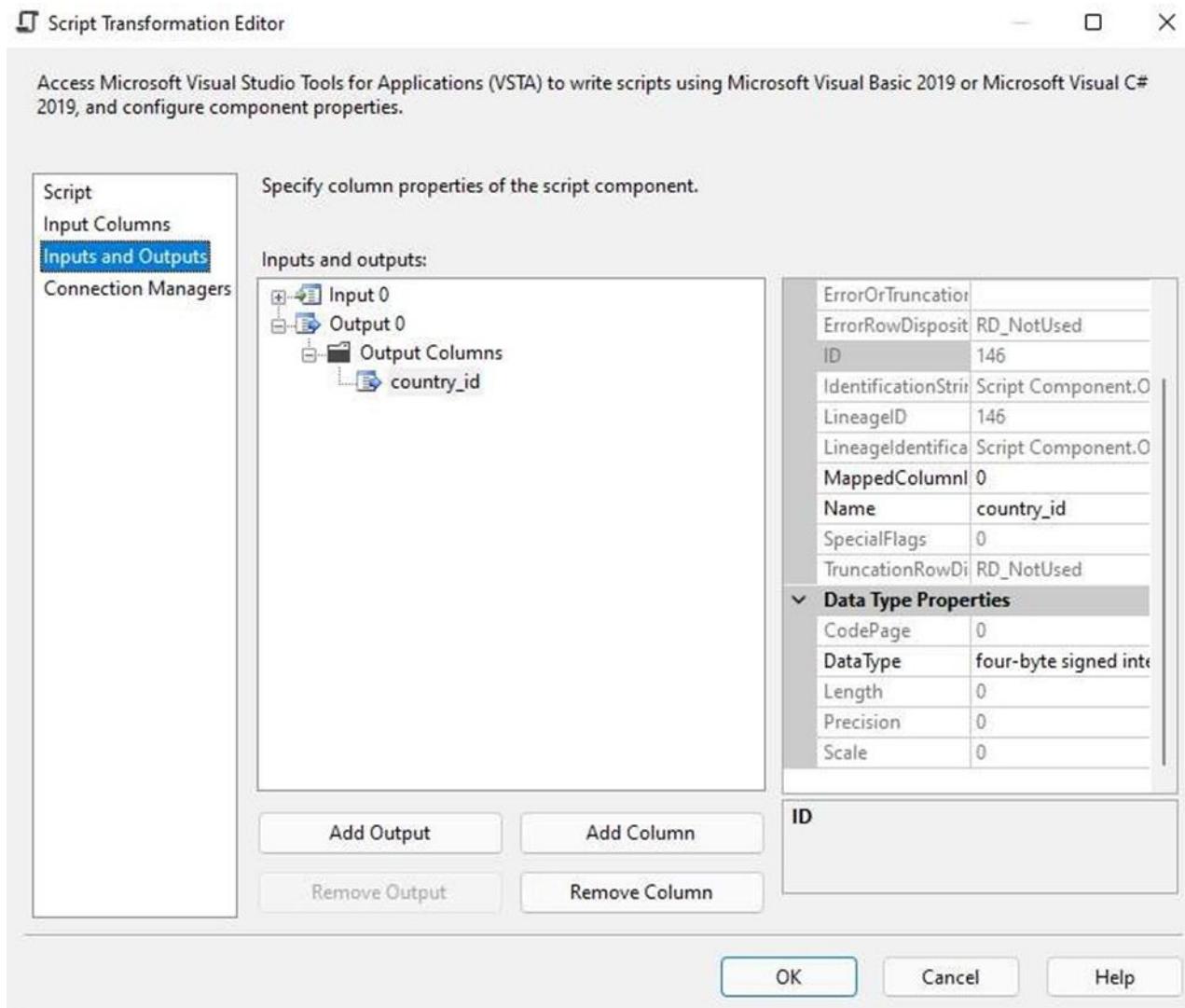






- Ở bước này mình sẽ dùng Script Transformation để tạo country_id và thiết lập cho thuộc tính này tăng tự động theo step = 1

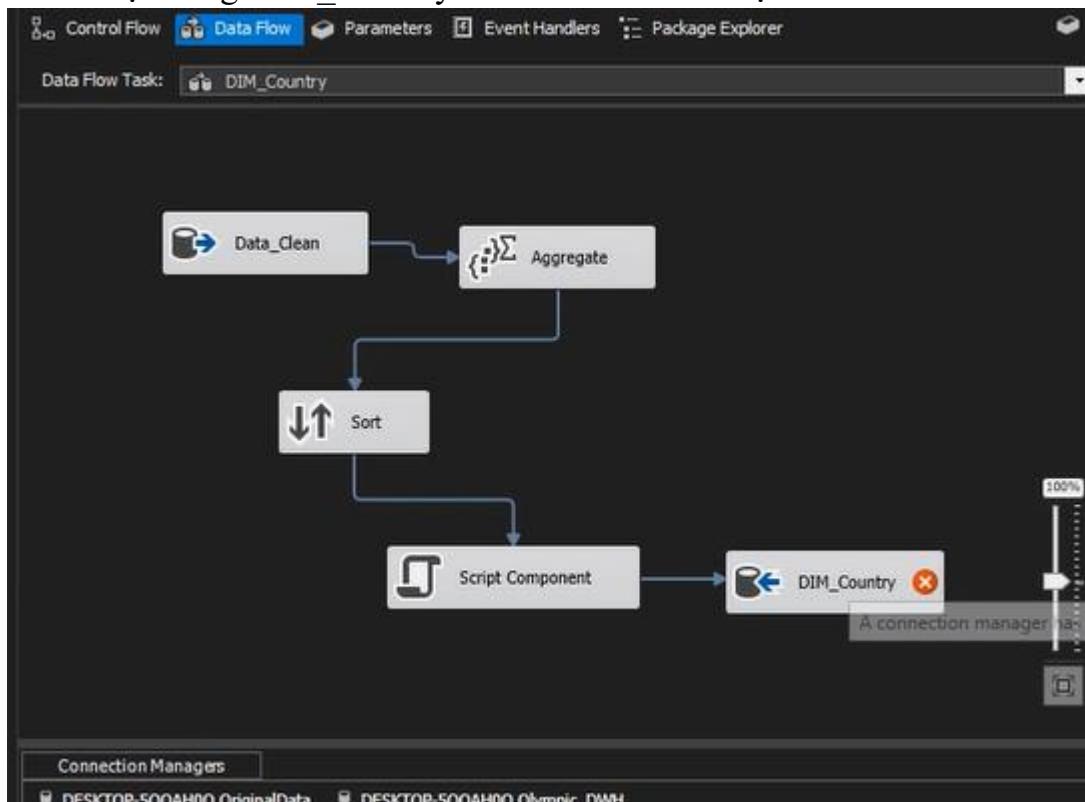


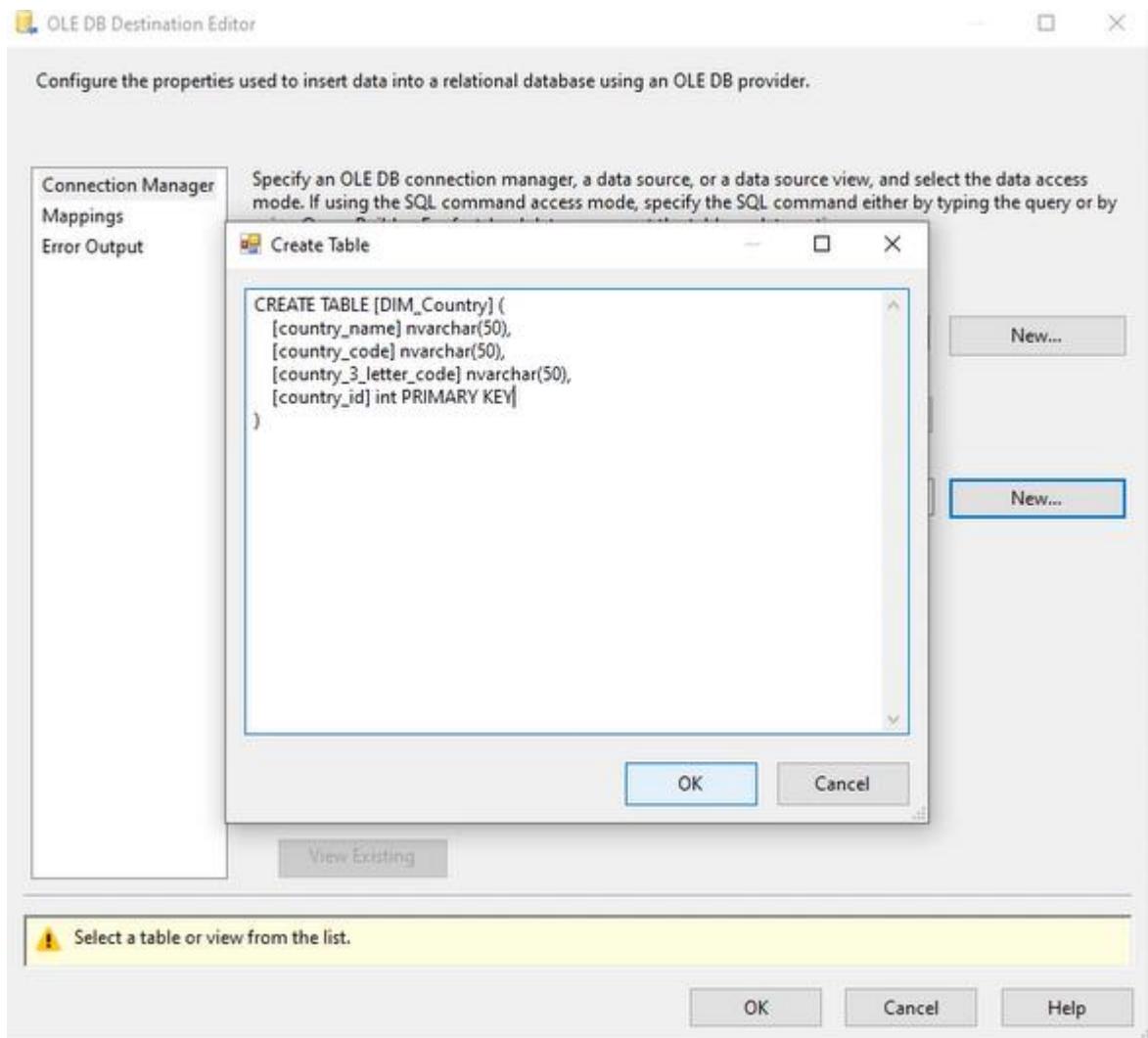


```
/// <summary>
/// </summary>
[Microsoft.SqlServer.Dts.Pipeline.SSISScriptComponentEntryPointAttribute]
public class ScriptMain : UserComponent
{
    int count = 1;
    Help: Using Integration Services variables and parameters
```

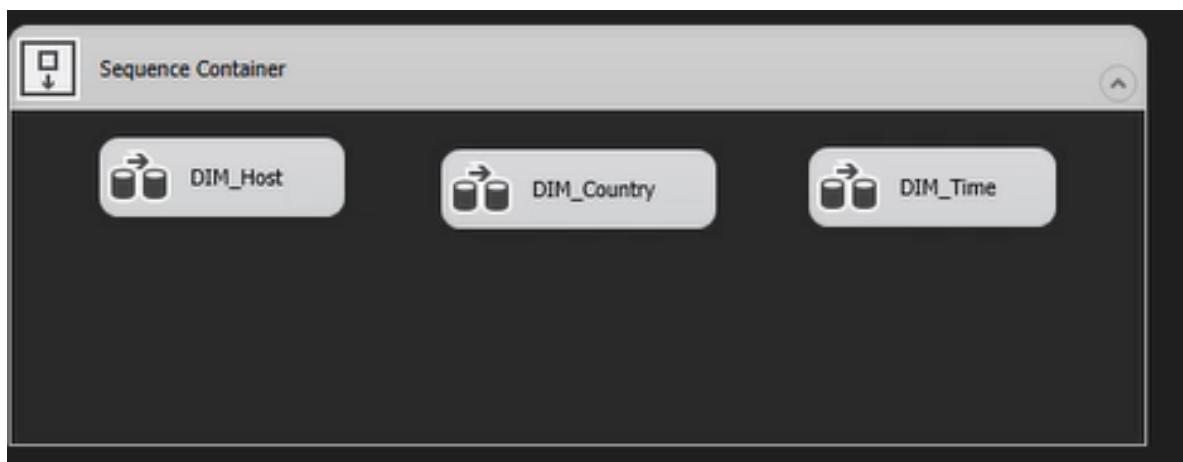
```
    /// Example of reading a value from a column in the row:  
    /// string zipCode = Row.ZipCode  
    ///  
    /// Example of writing a value to a column in the row:  
    /// Row.ZipCode = zipCode  
    /// </summary>  
    /// <param name="Row">The row that is currently passing through the component  
2 references  
public override void Input0_ProcessInputRow(Input0Buffer Row)  
{  
    Row.countryid = count;  
    count++;  
}
```

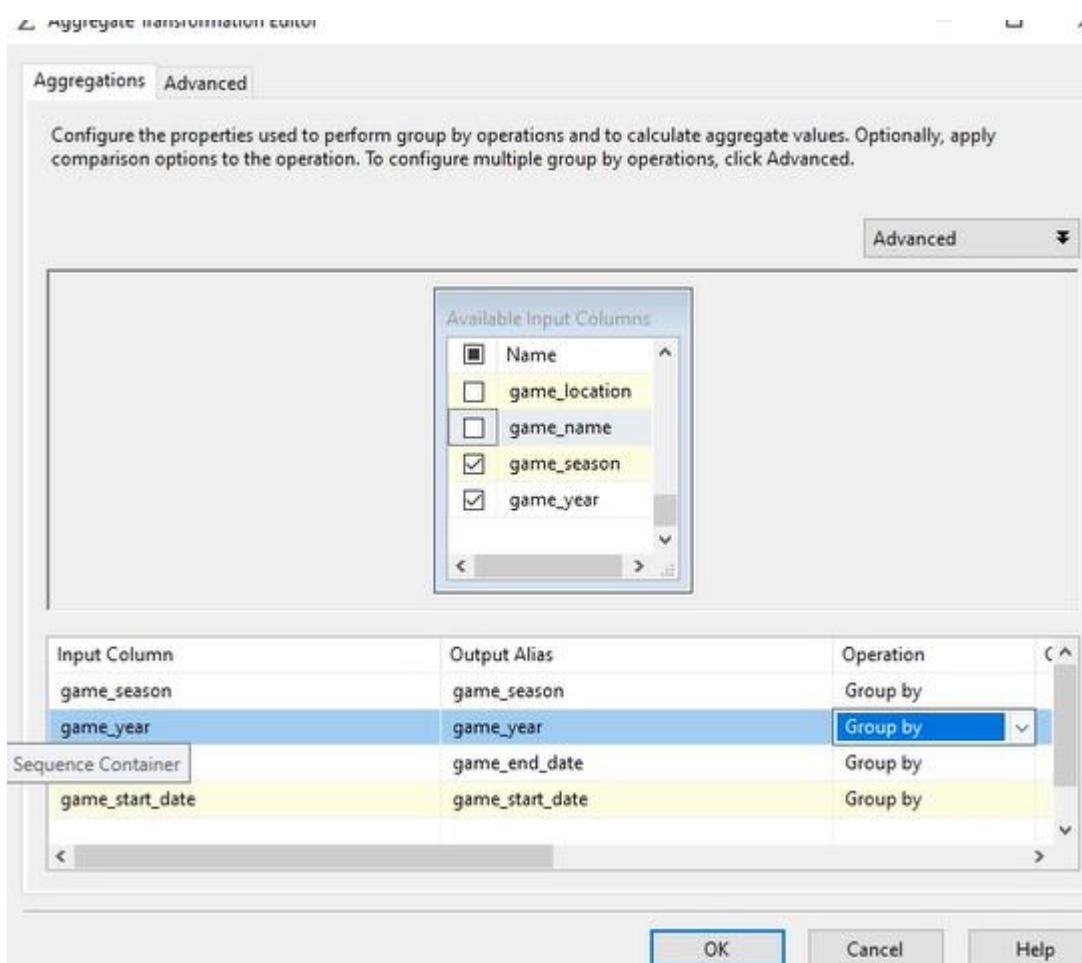
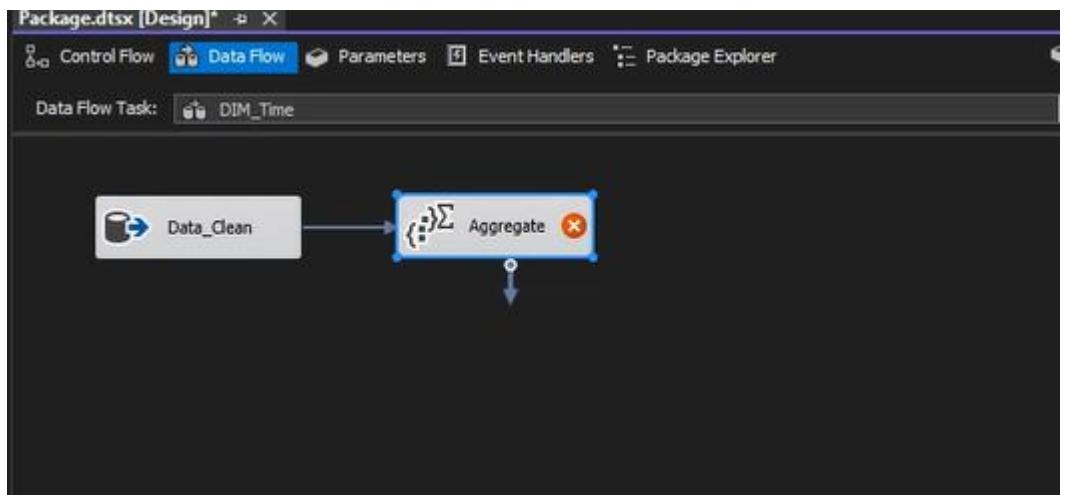
- Tạo bảng DIM_Country và lưu vào kho dữ liệu.

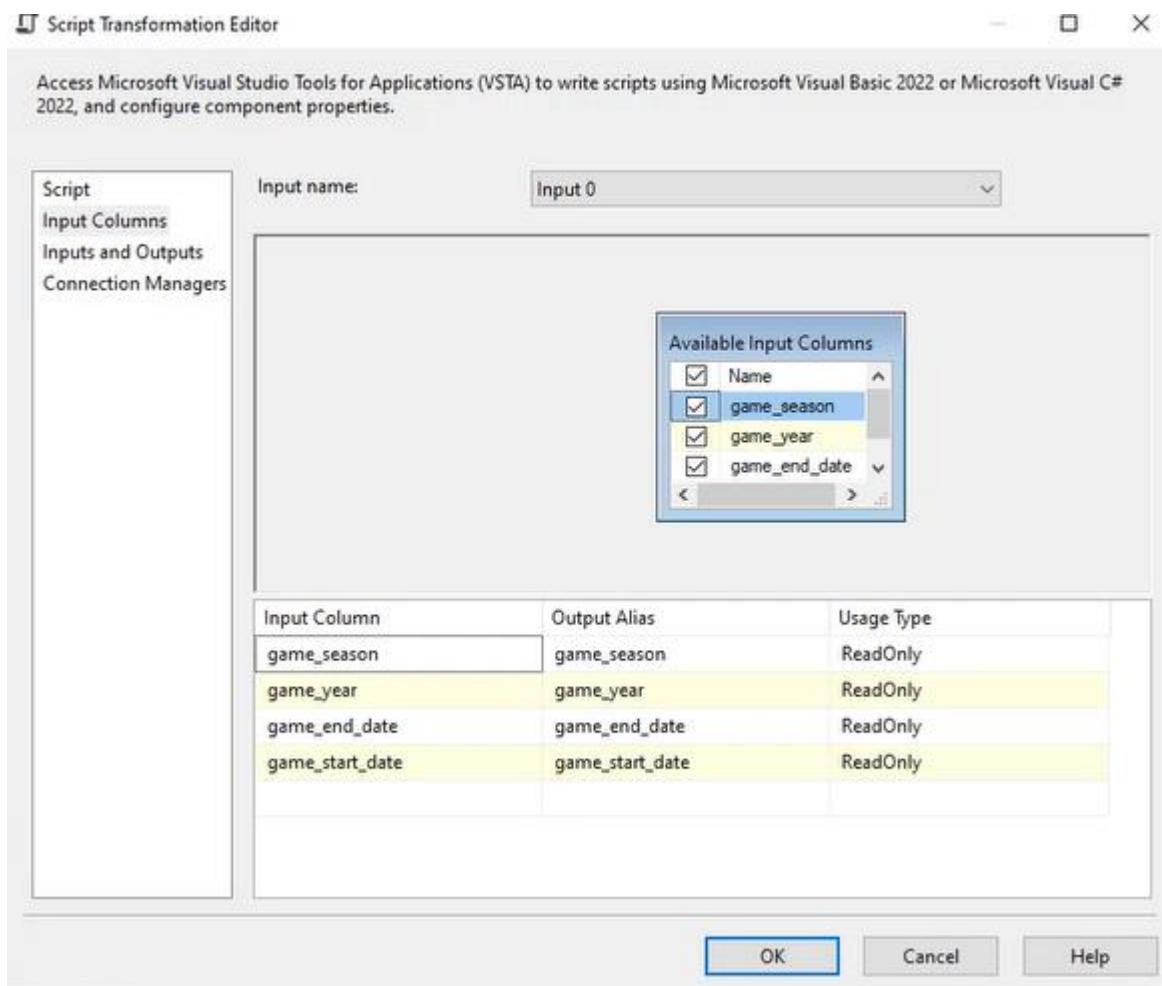
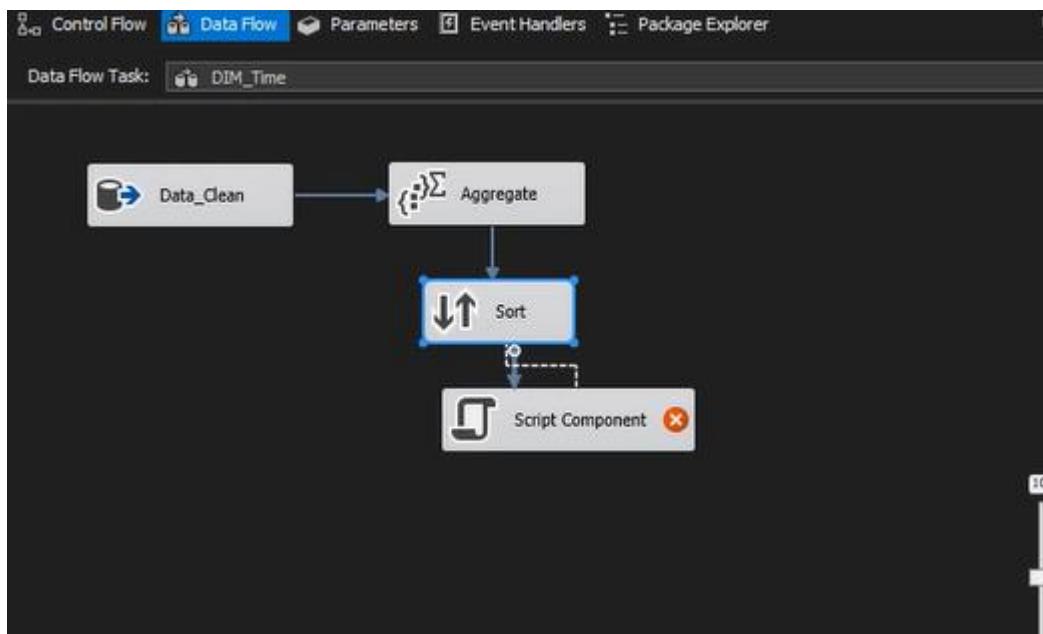


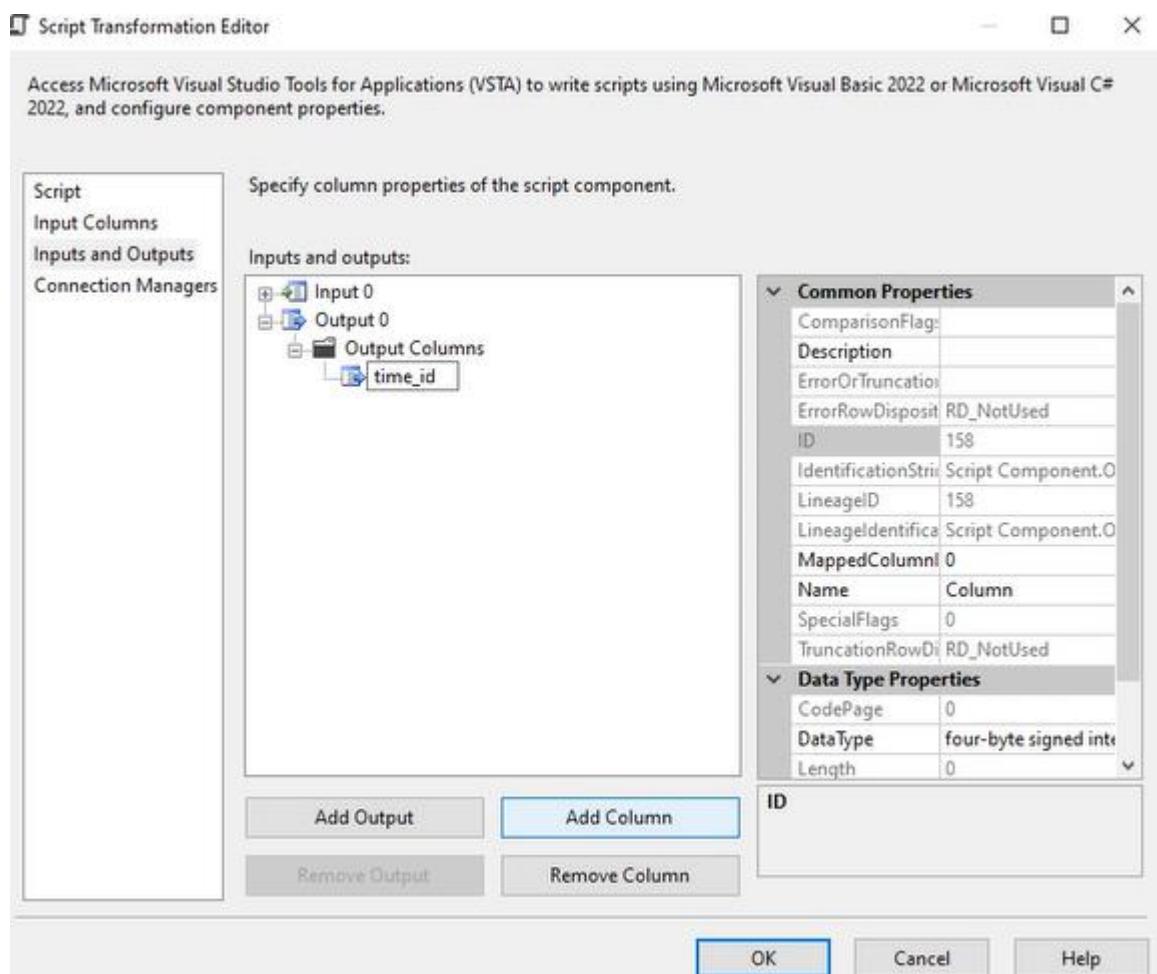


2.2.7.3 Tạo bảng DIM_Time

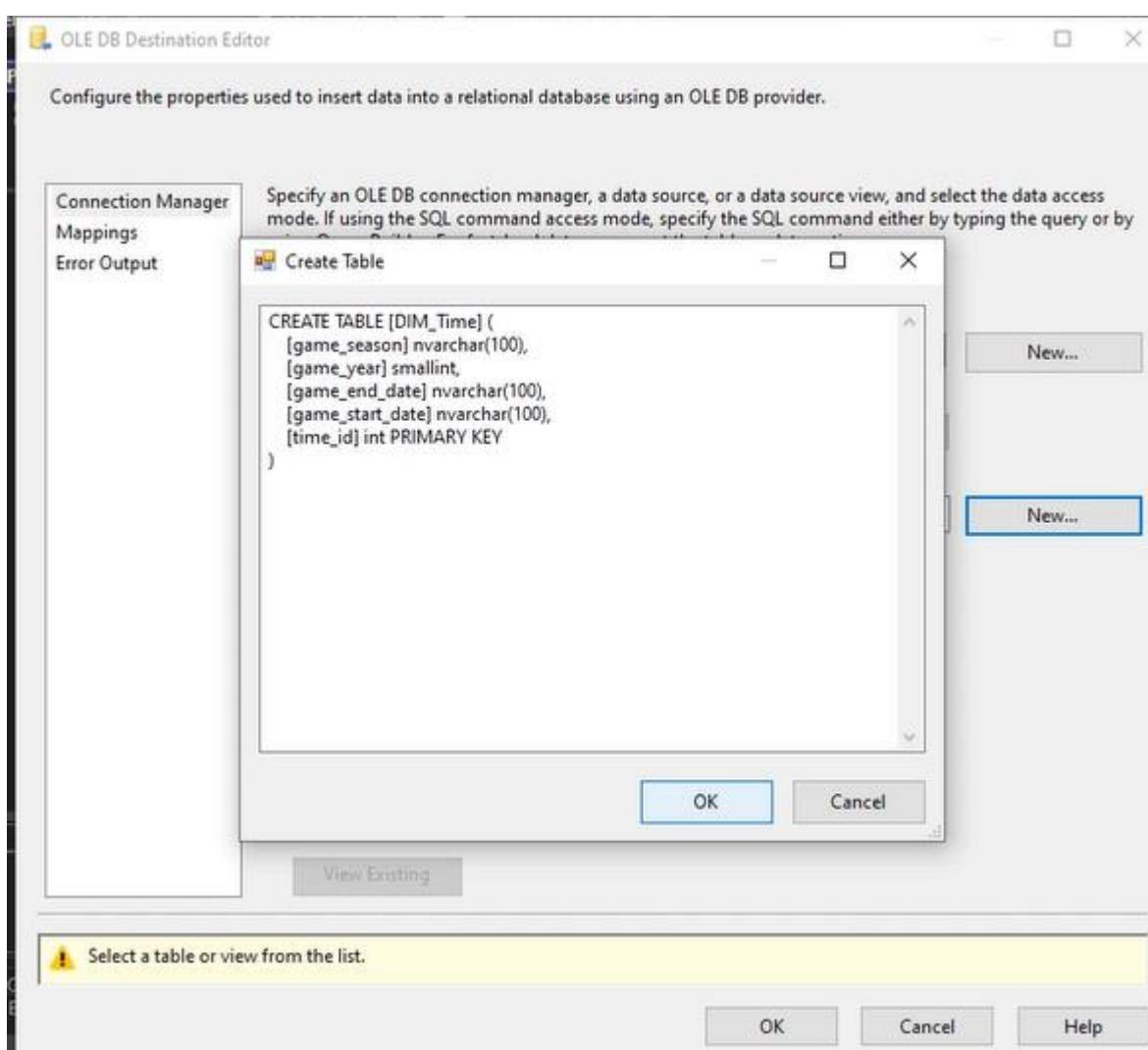
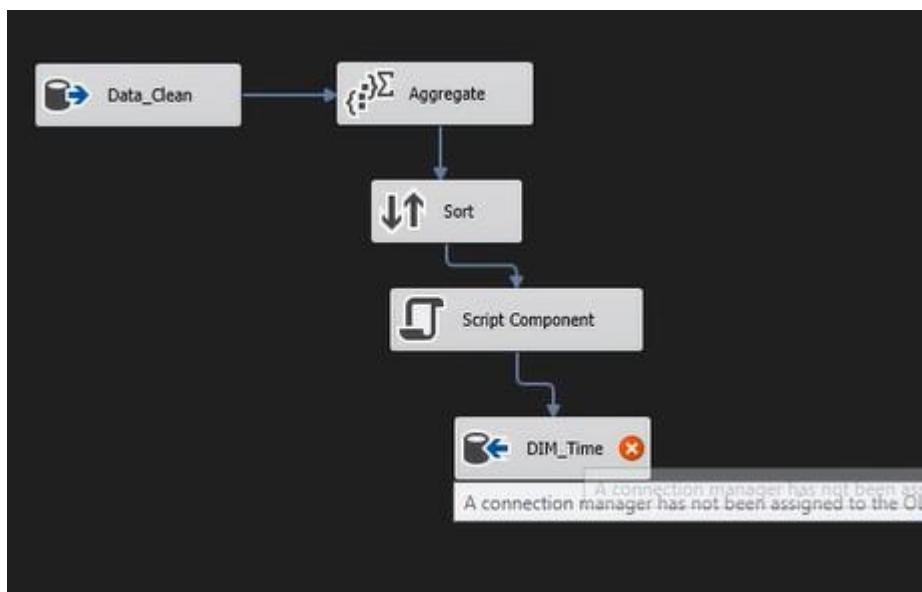




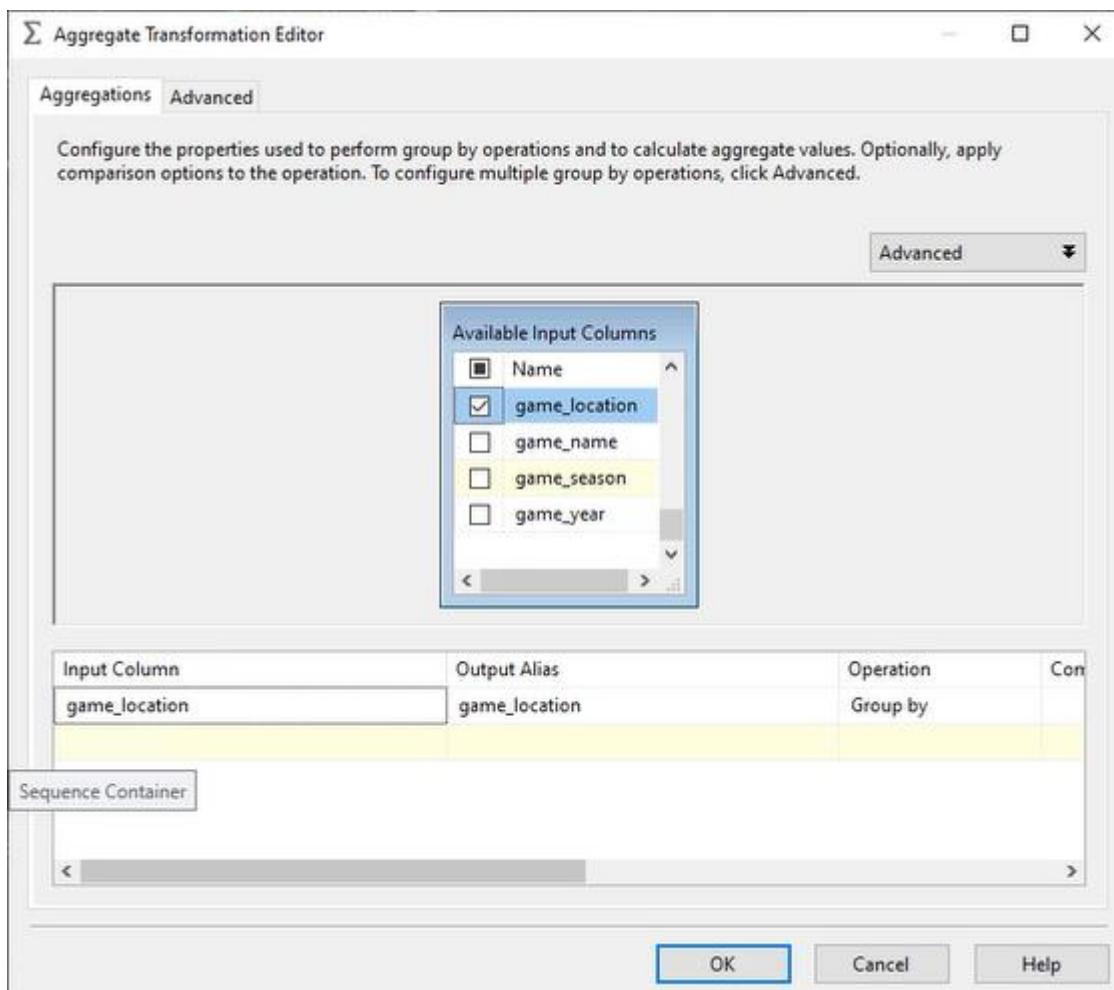


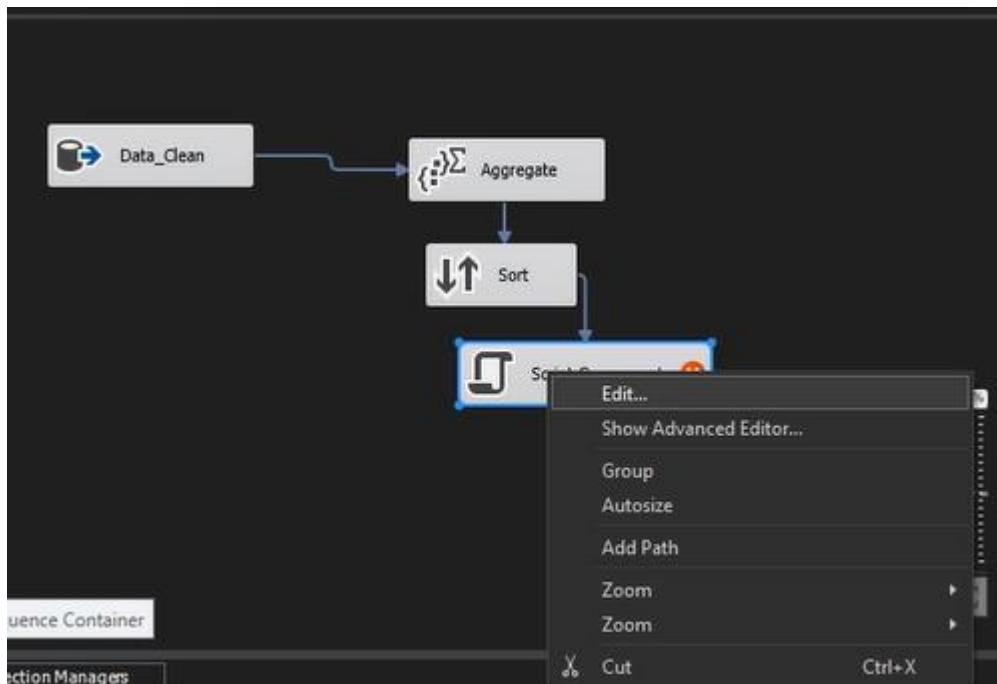
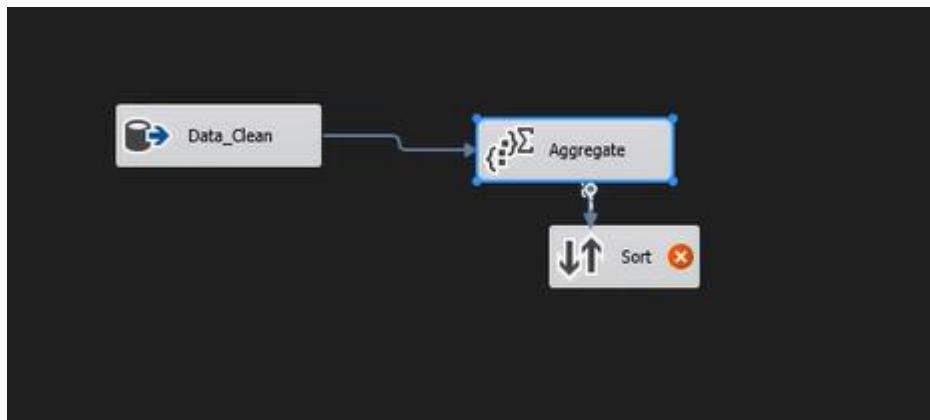


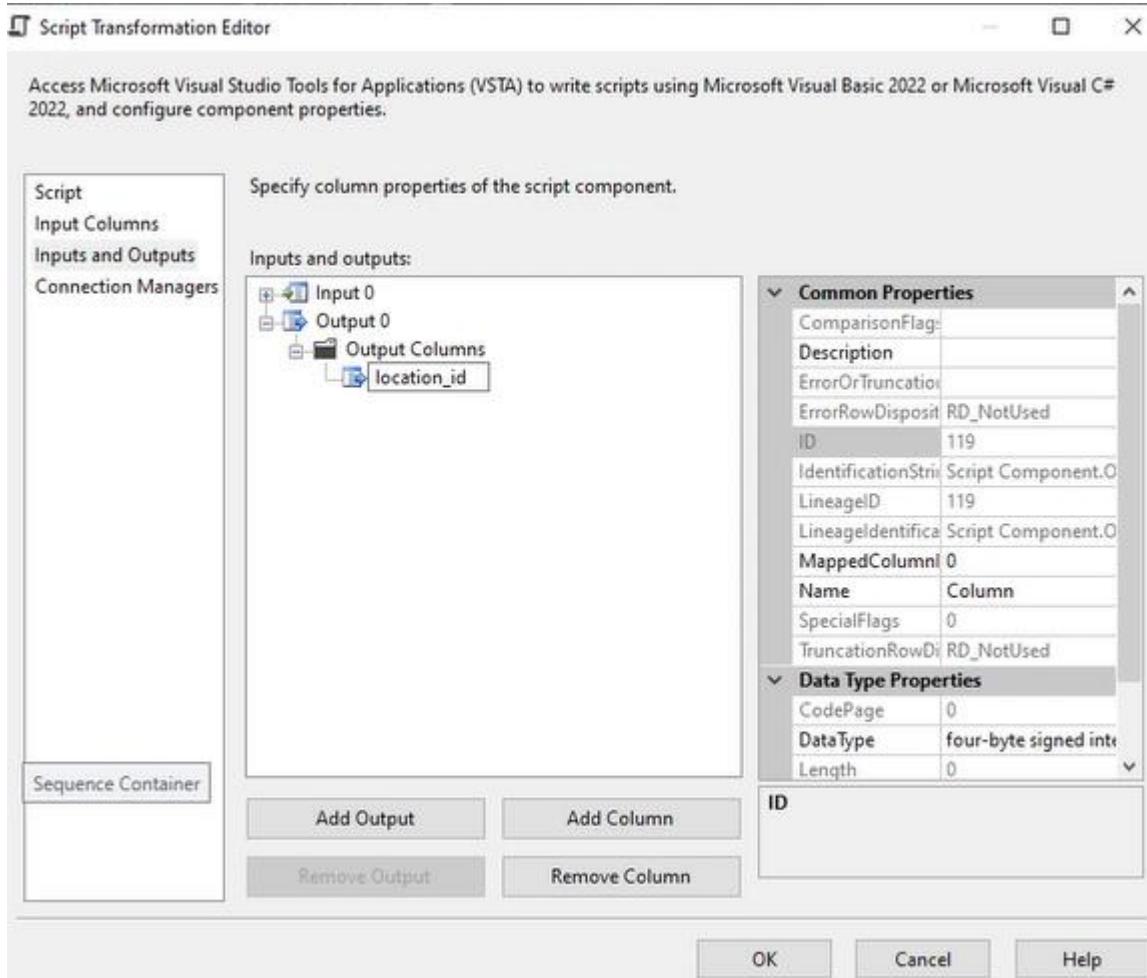
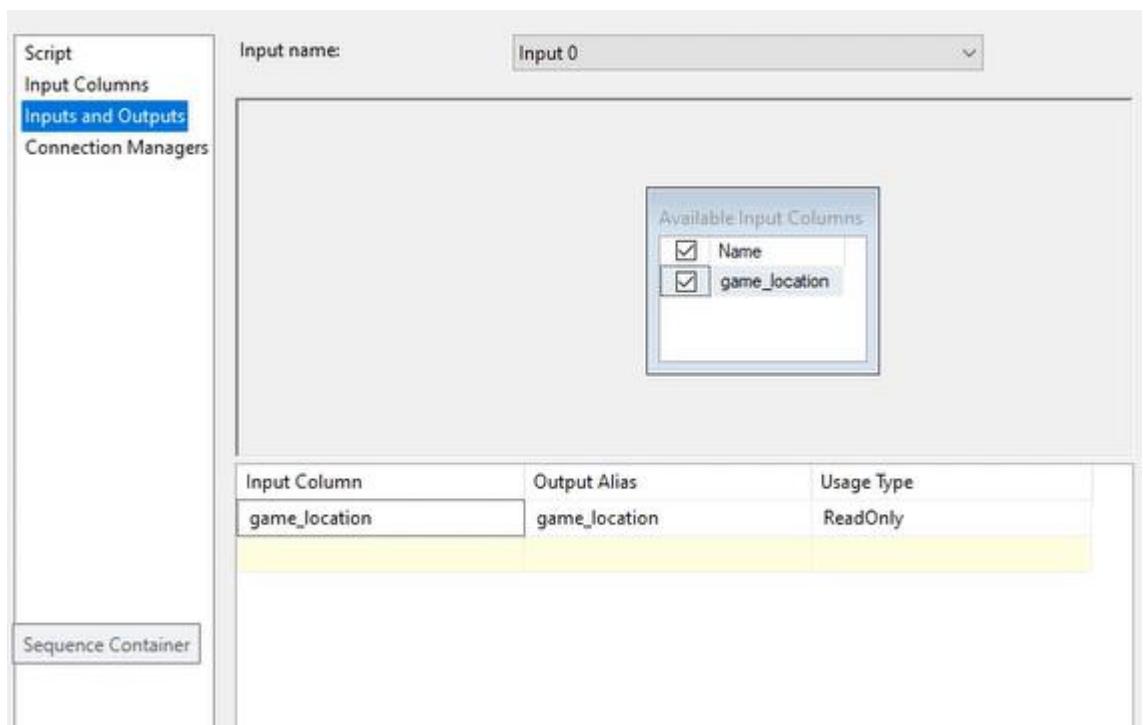
```
 3     /// </summary>
 4     /// <param name="Row">The row that is currently passing through the com
 5     int count = 1;
 6     2 references
 7     public override void Input0_ProcessInputRow(Input0Buffer Row)
 8     {
 9         /*
10          * Add your code here
11          */
12         Row.timeid = count;
13         count++;
14     }
15 }
```



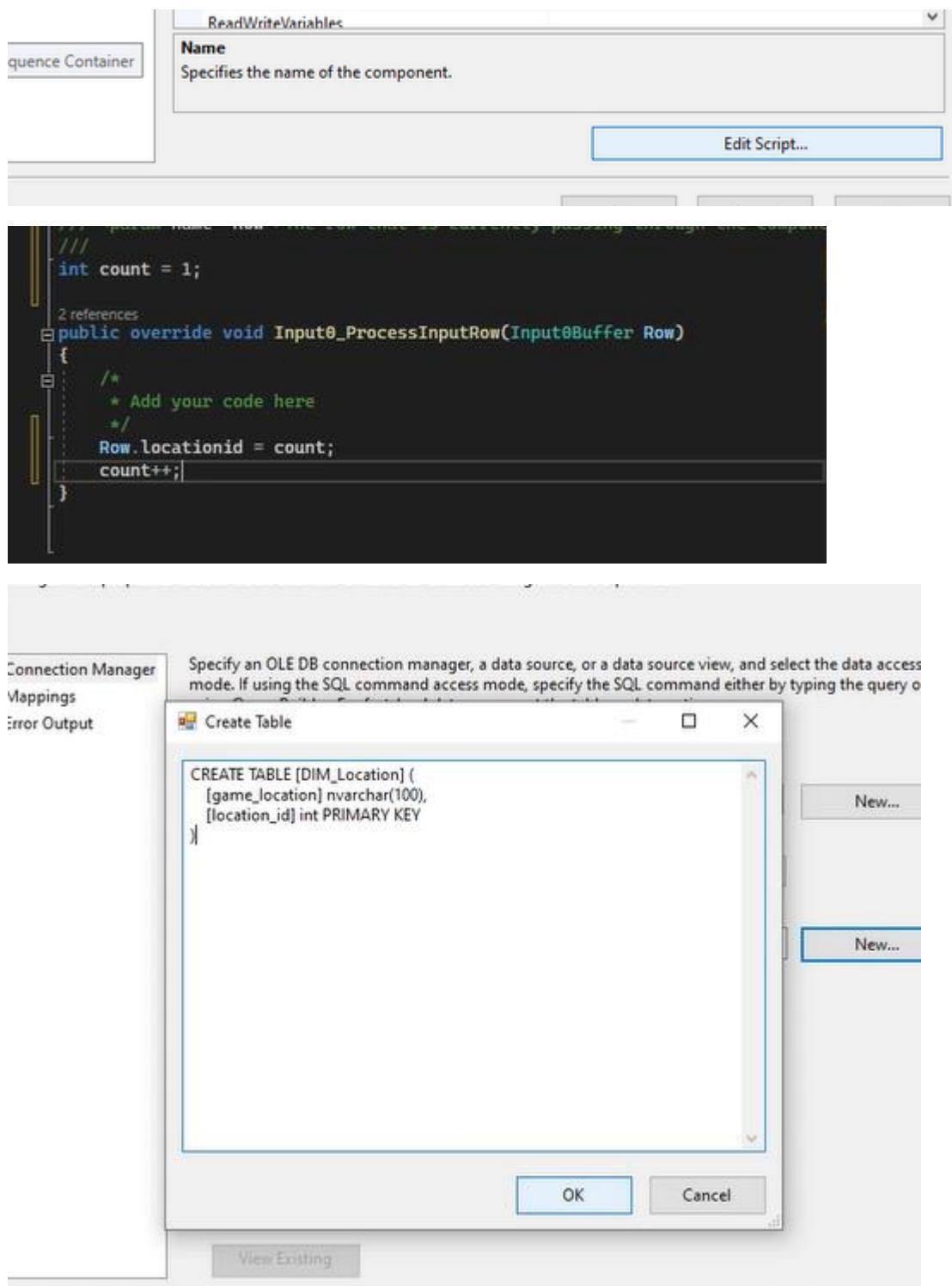
2.2.7.4 Tạo bảng DIM_Location

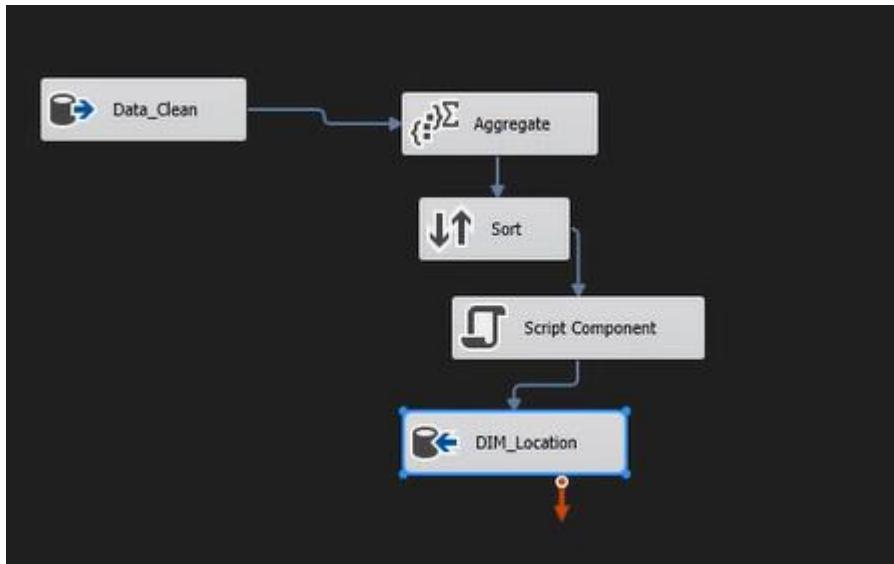




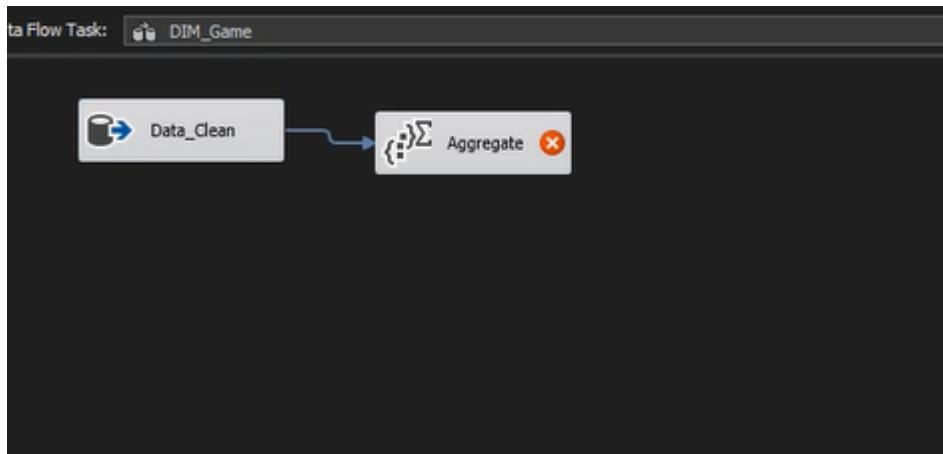


- Click vào Edit Script để thiết lập tăng location_id





2.2.7.5 Tạo bảng DIM_Game



Advanced ▾

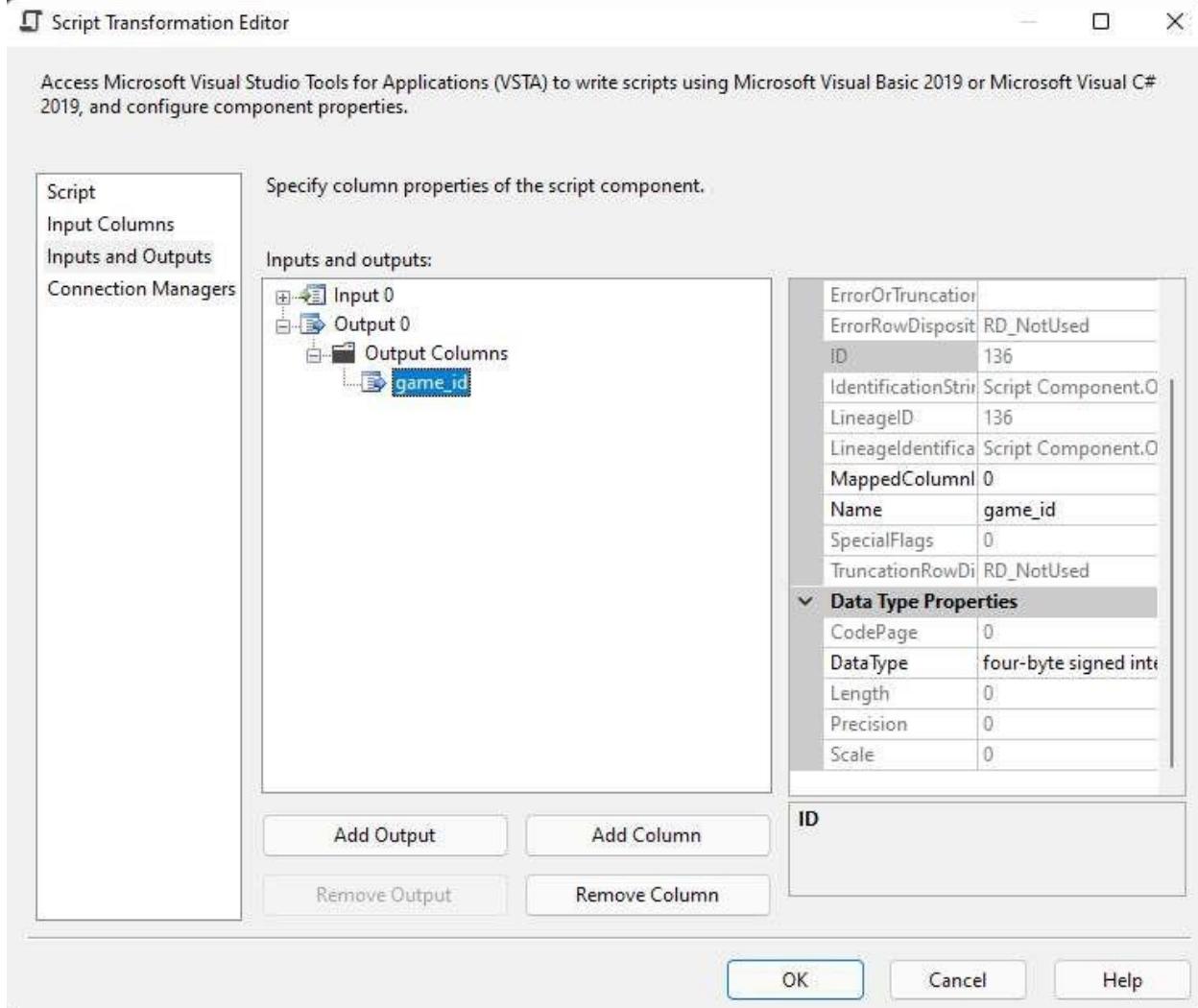
Available Input Columns

<input checked="" type="checkbox"/> Name
<input type="checkbox"/> game_location
<input type="checkbox"/> game_name
<input type="checkbox"/> game_season
<input type="checkbox"/> game_year

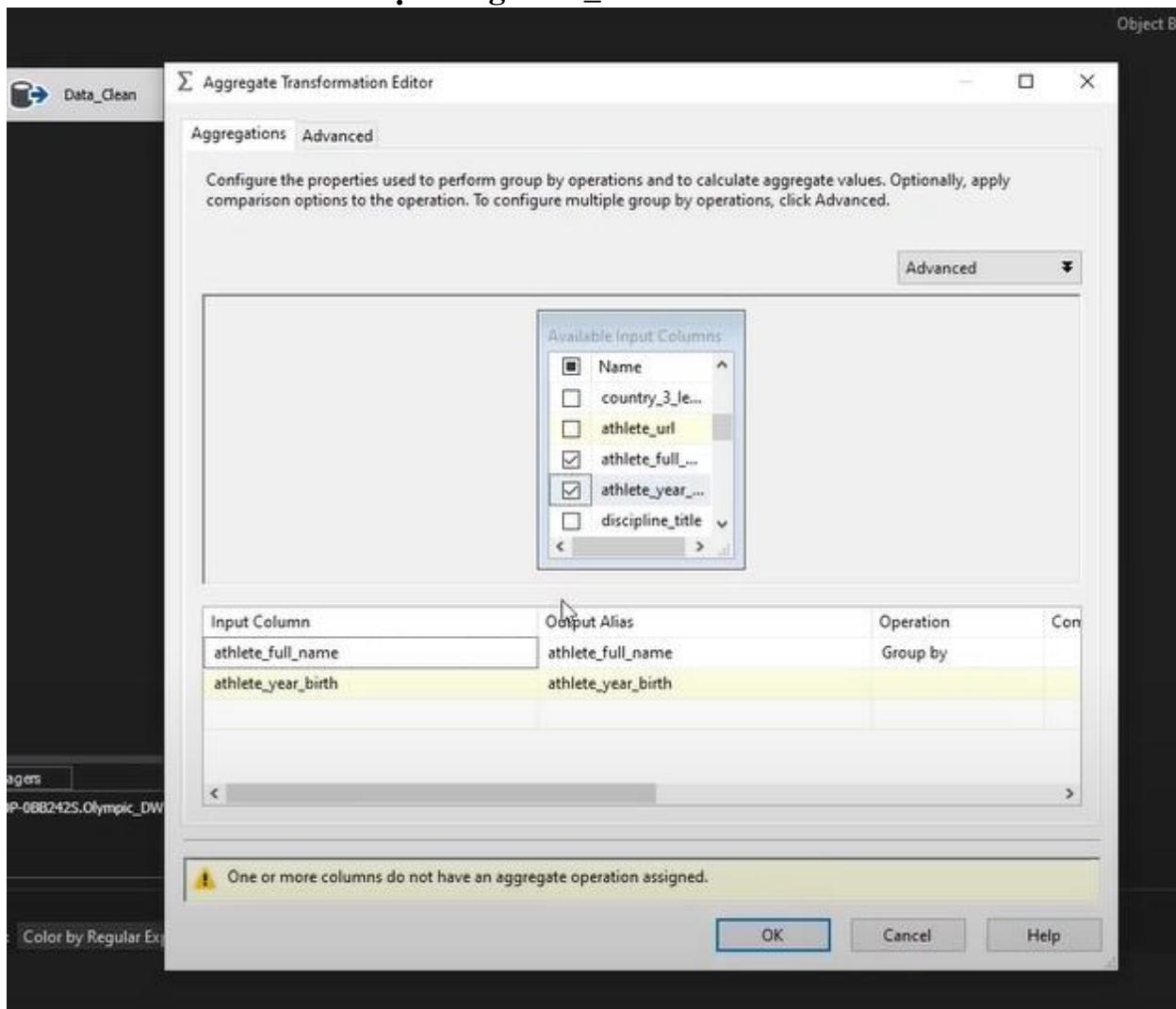
< >

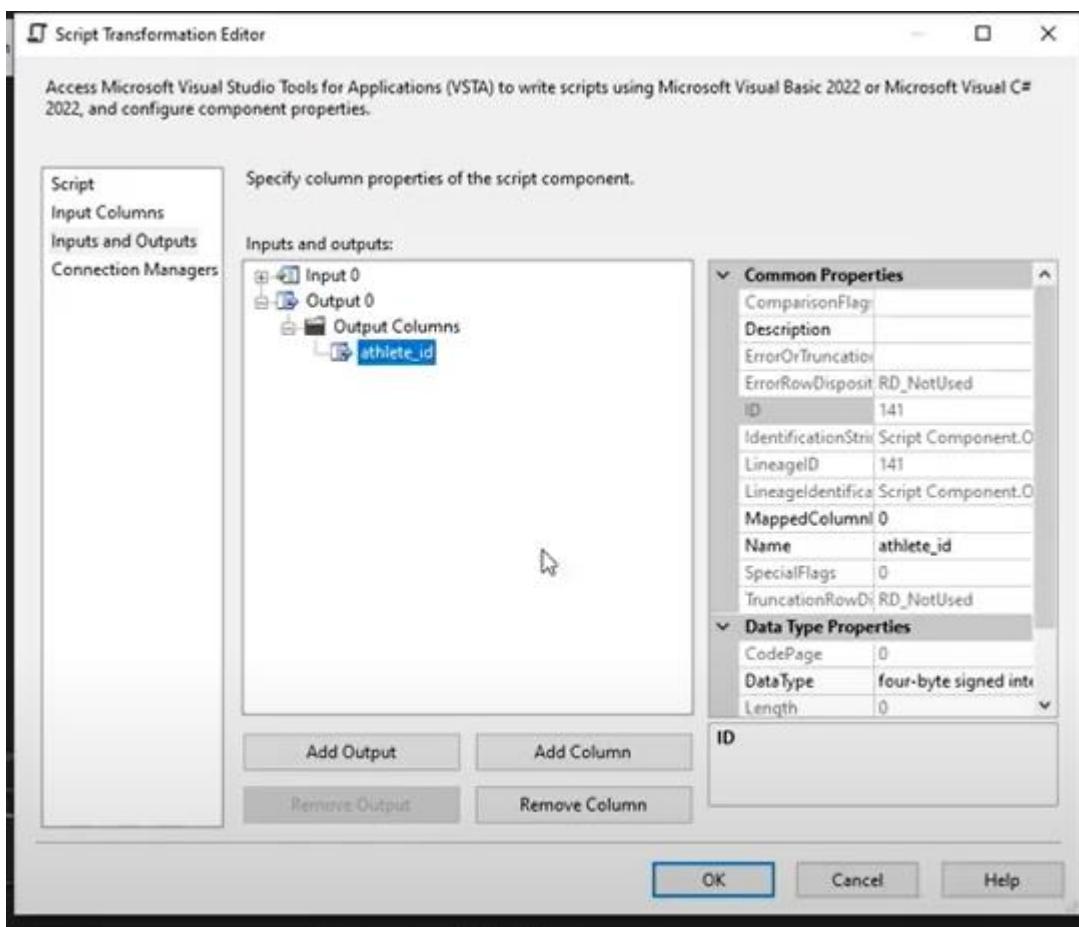
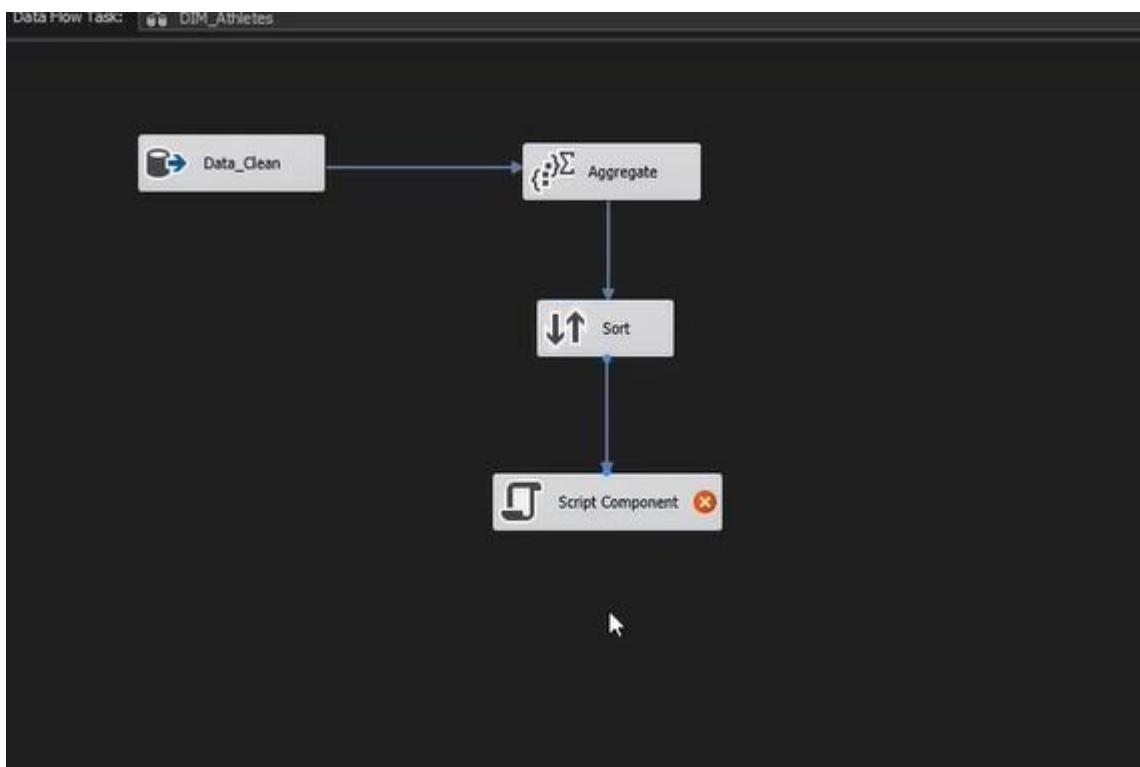
Input Column	Output Alias	Operation	Con
discipline_title	discipline_title	Group by	
event_title	event_title	Count	
participant_type	participant_type	Group by	

< >

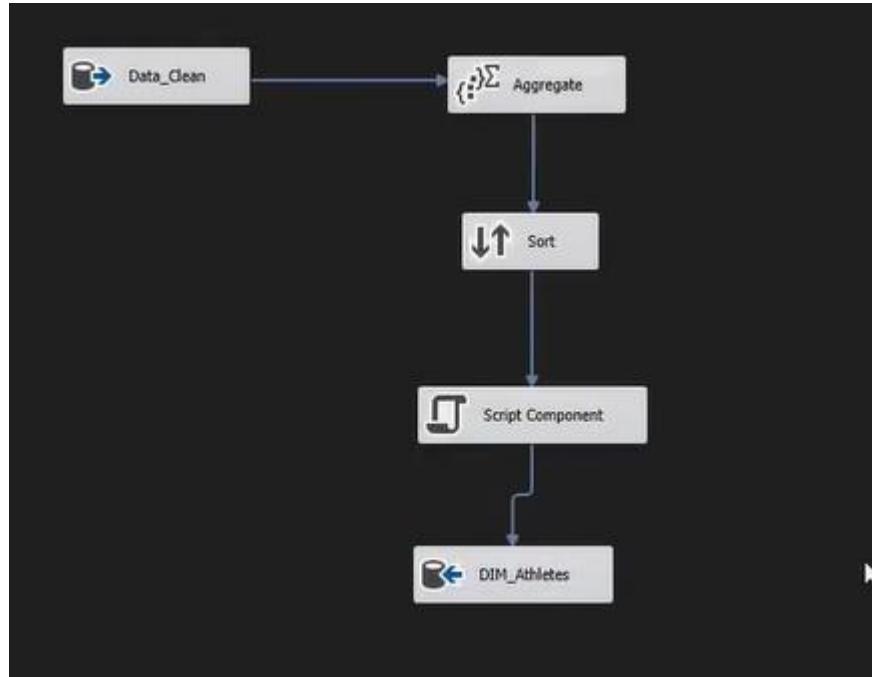


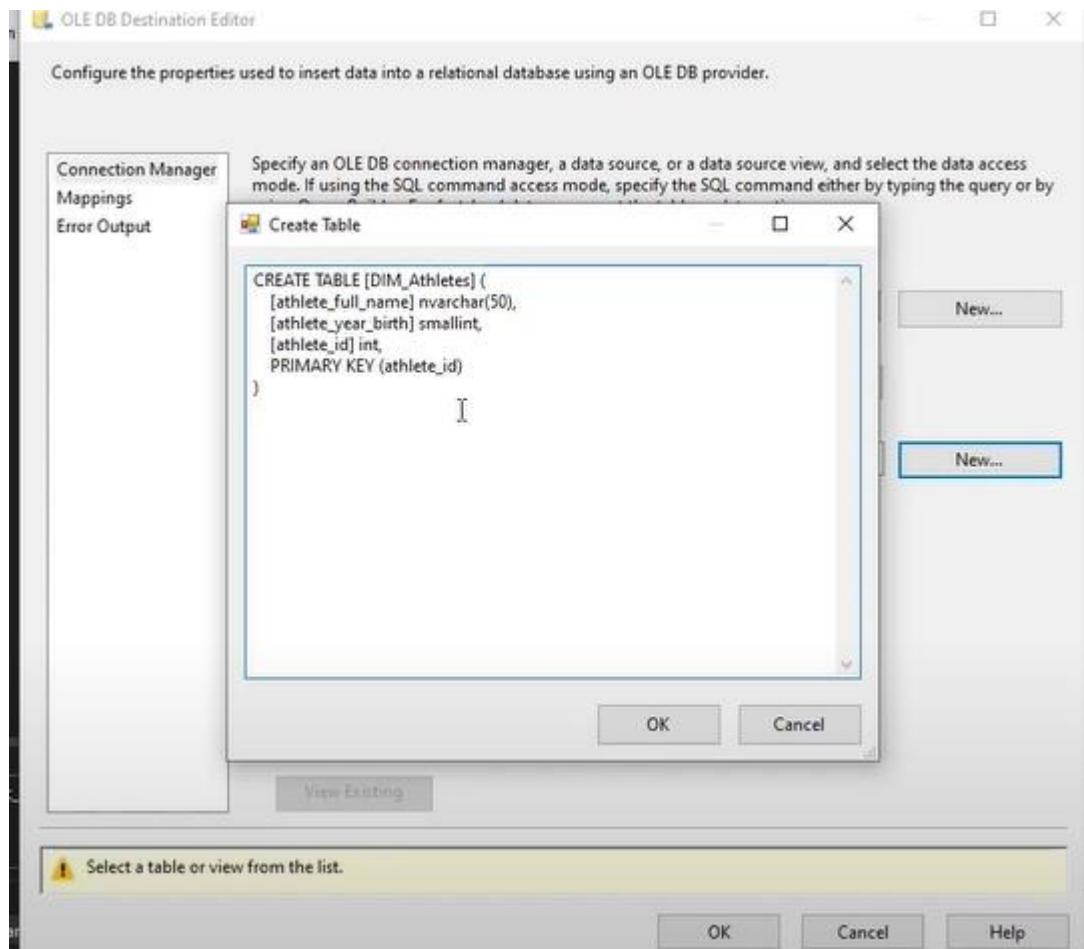
2.2.7.6 Tạo bảng DIM_Athlete





```
/// <param name="Row">The row that is currently passing through the component</param>
///
int count = 1;
2 references
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    /*
     * Add your code here
     */
    Row.athleteid = count;
    count++;
}
```



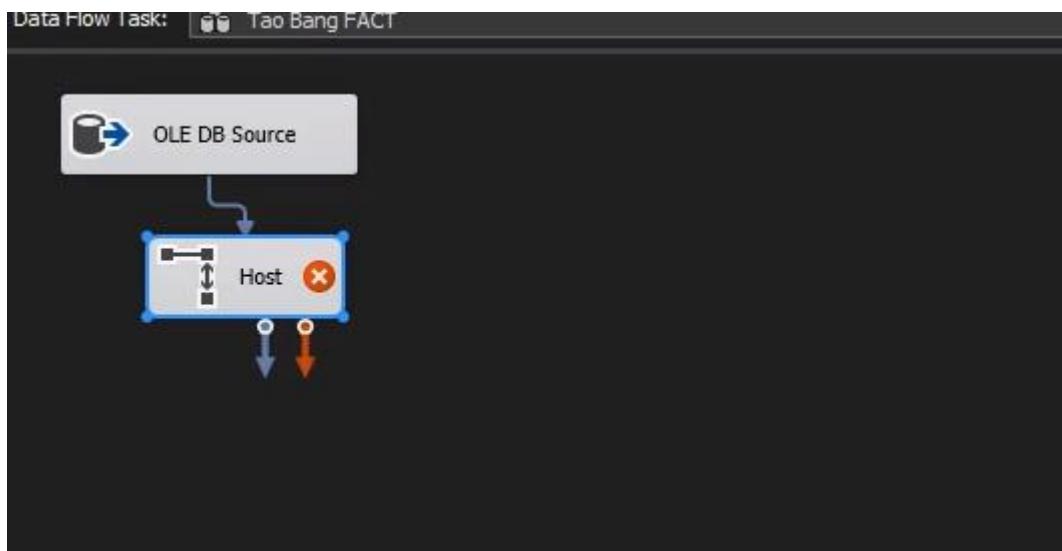


Quá trình tạo bảng FACT

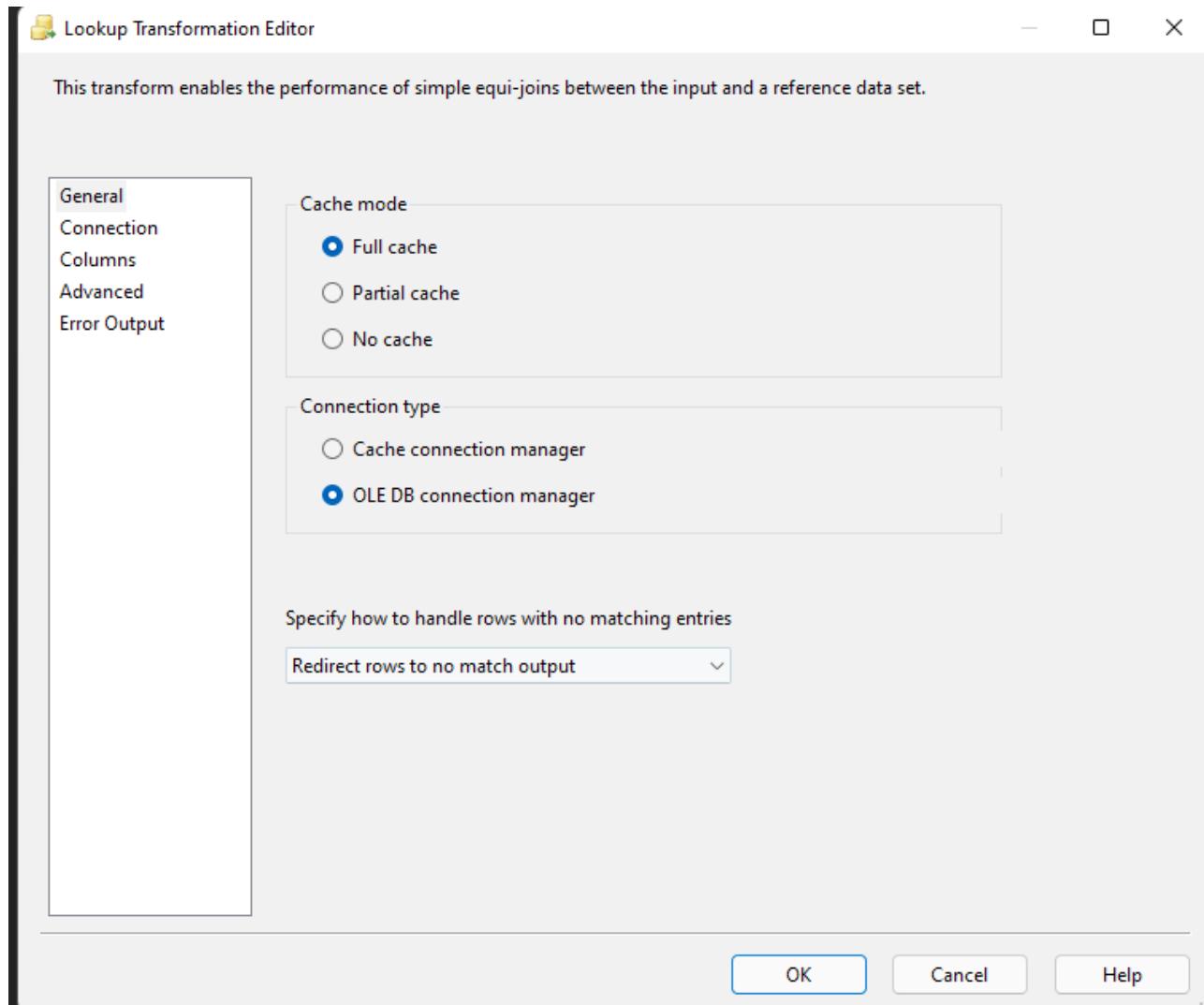
- Tạo mới một Data Flow Task và đặt trên là Tạo bảng FACT, trong này sẽ có các bước tạo bảng FACT



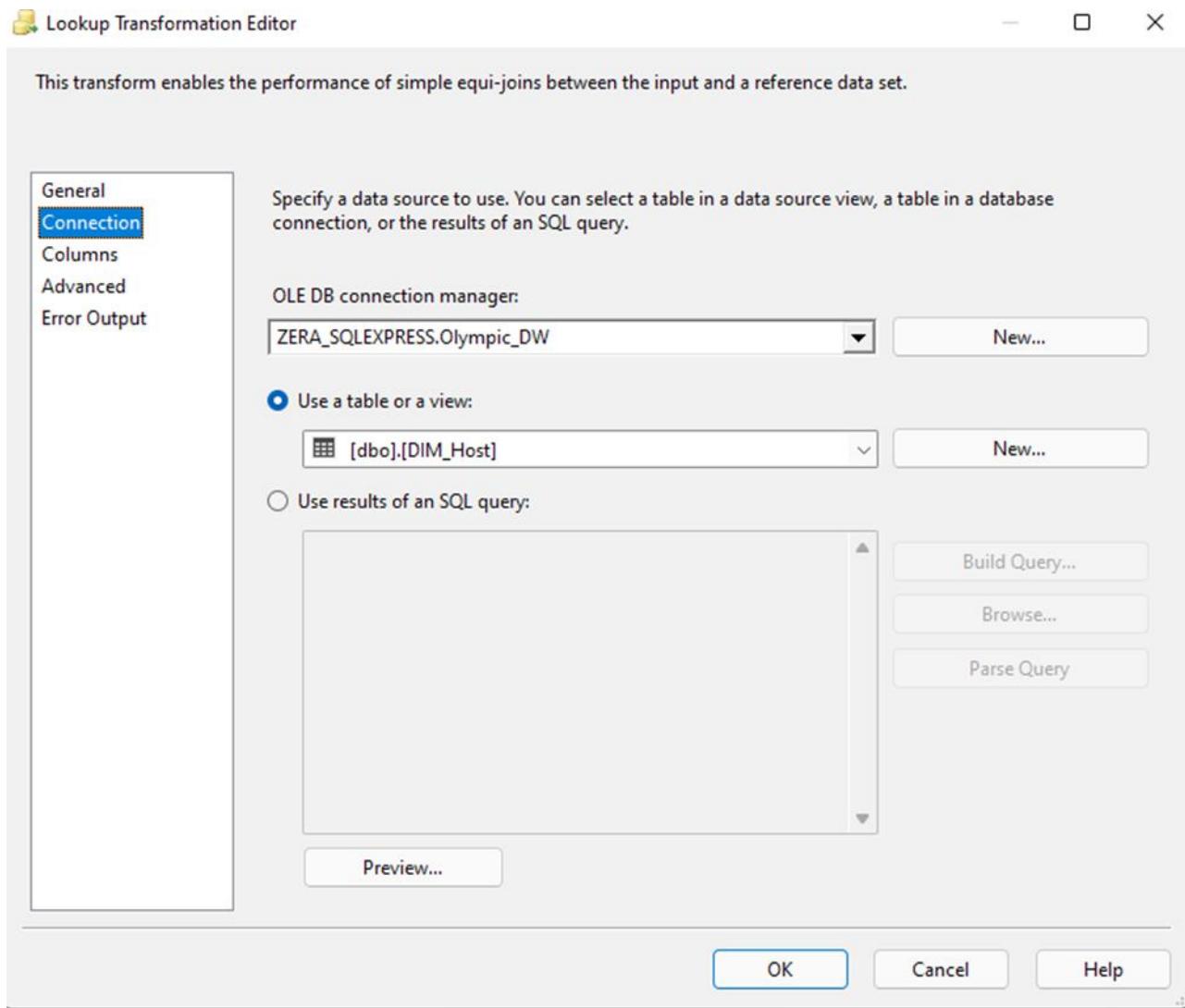
- Chọn kết nối tới database OriginalData và chọn bảng CleanData
- Sử dụng LookUp để ánh xạ dữ liệu



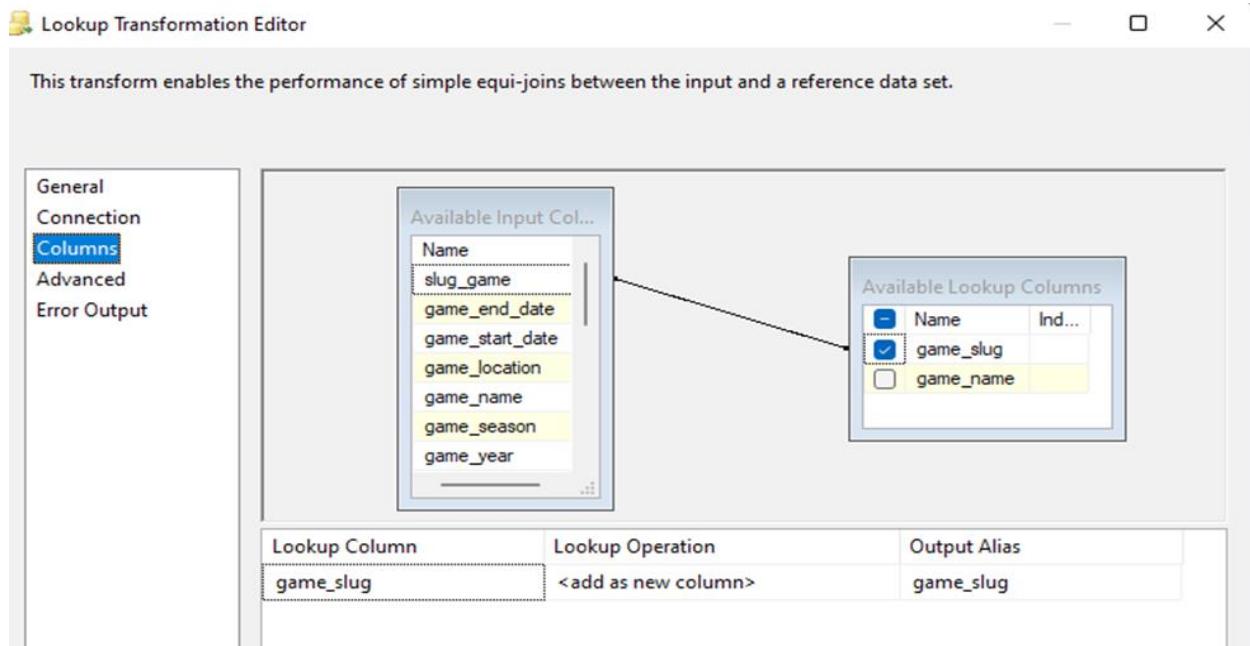
- Chọn Redirect rows to no match output.



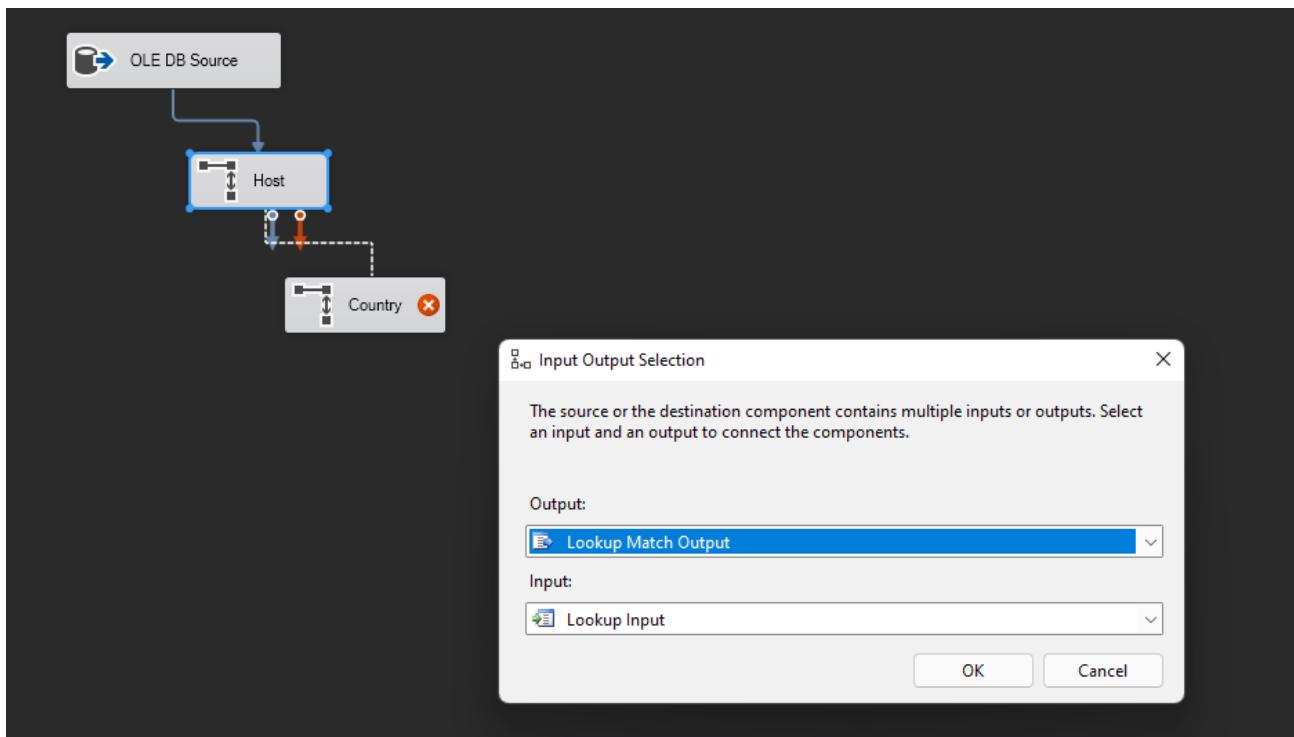
- Chọn connection tới database Olympic_DW và chọn bảng Dim_Host.



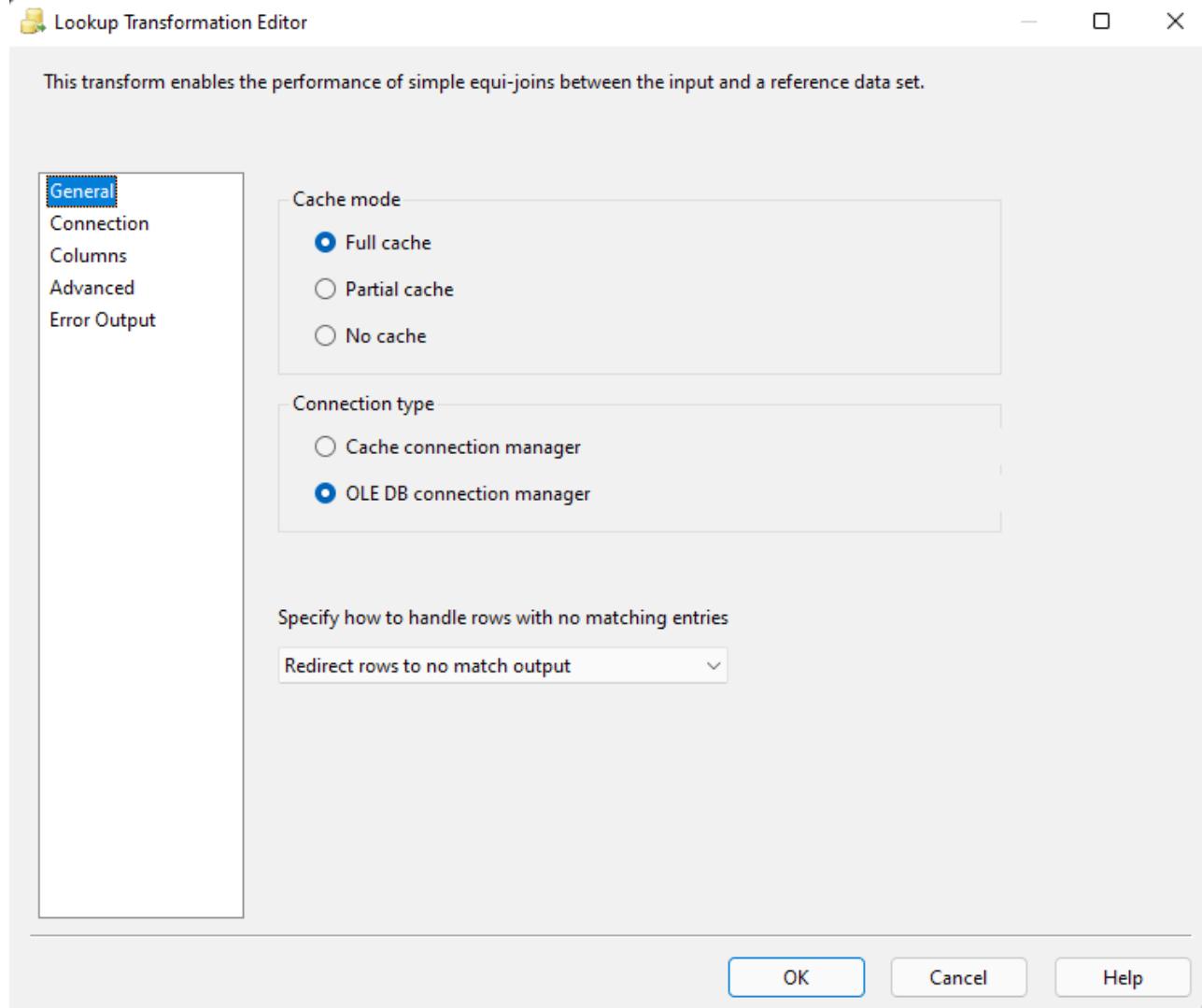
- Chọn cột look up và chọn OK.



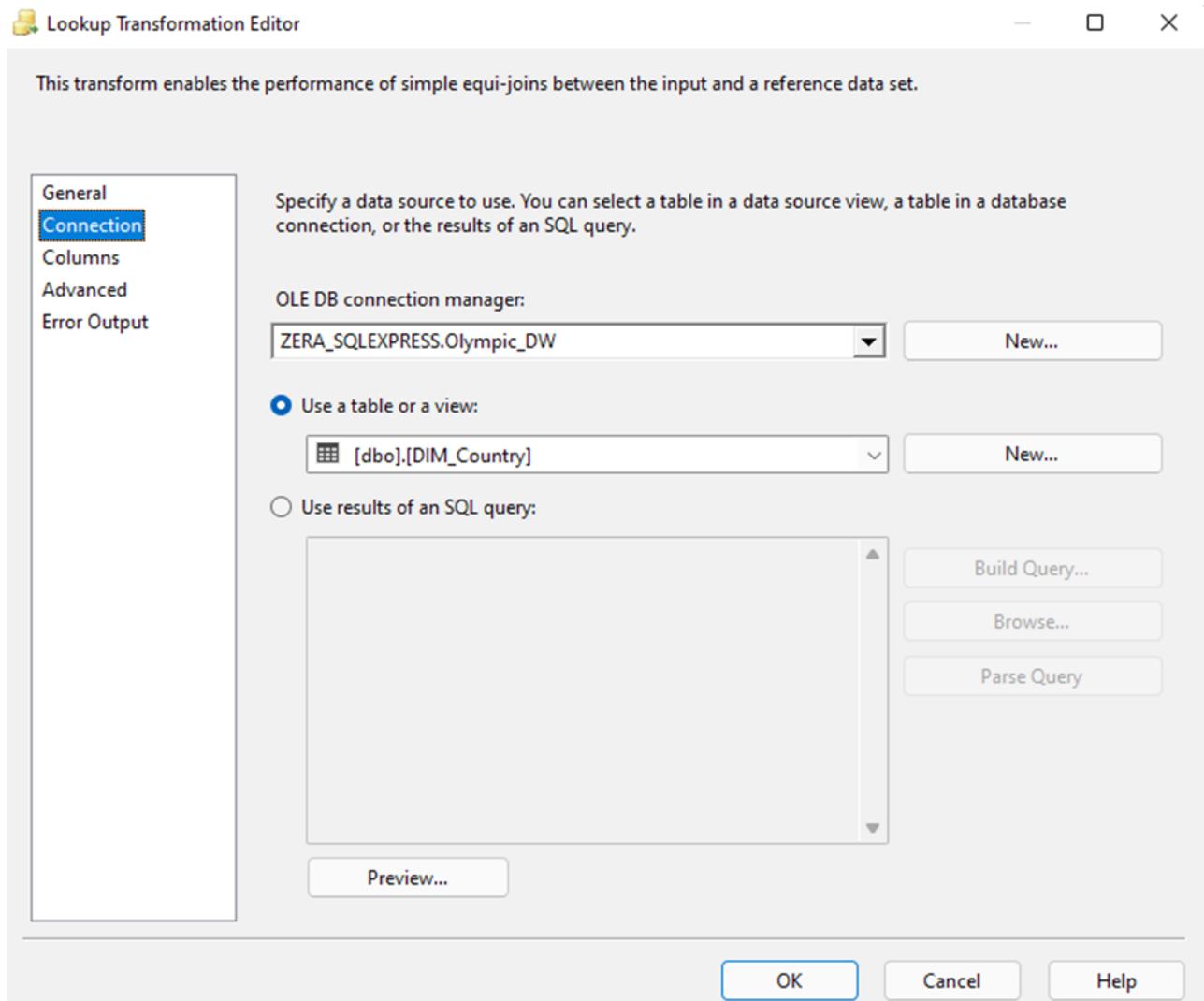
- Kéo thả và tạo mới Look up, đặt tên là Country.



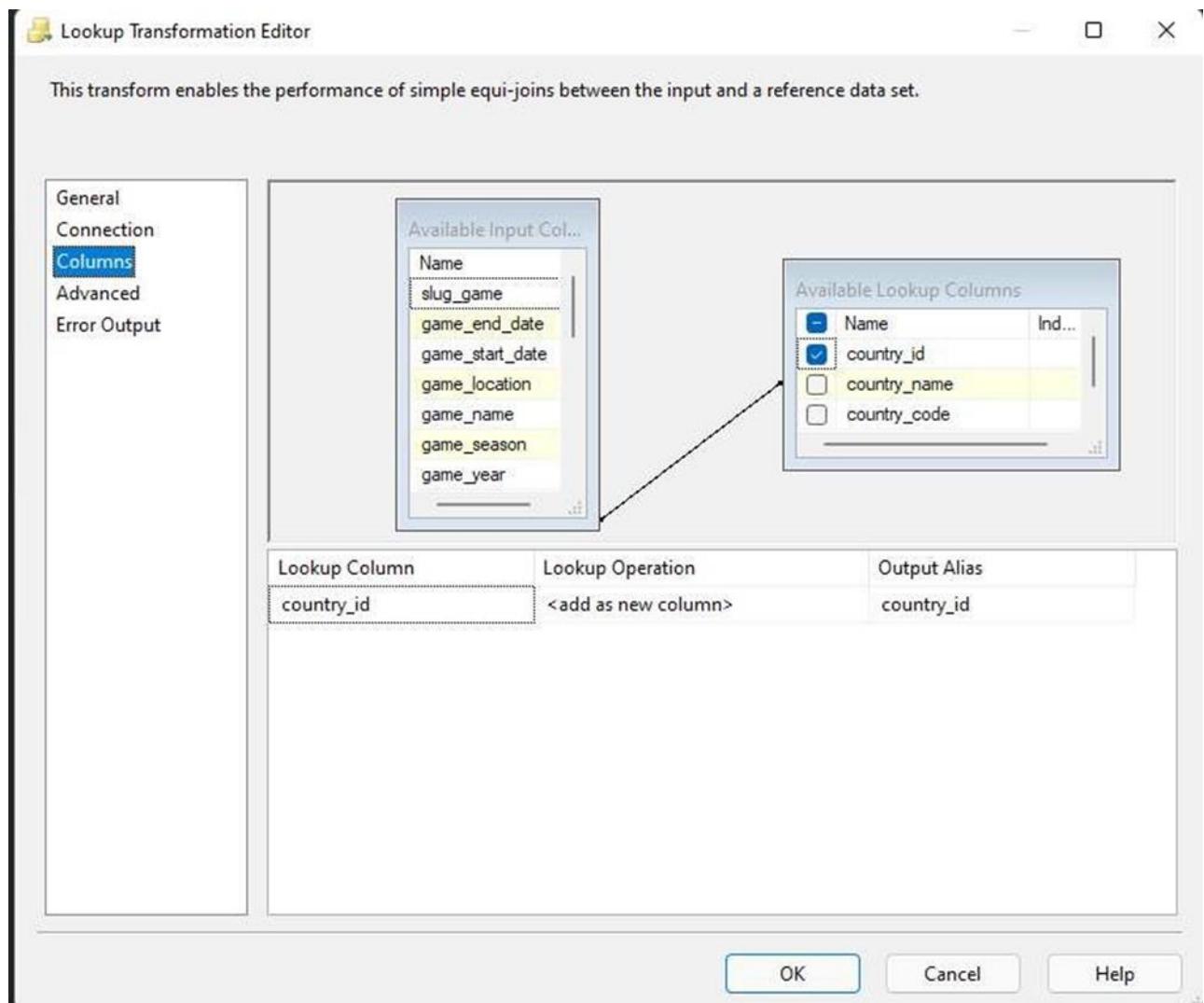
- Chọn Redirect rows to no match output.



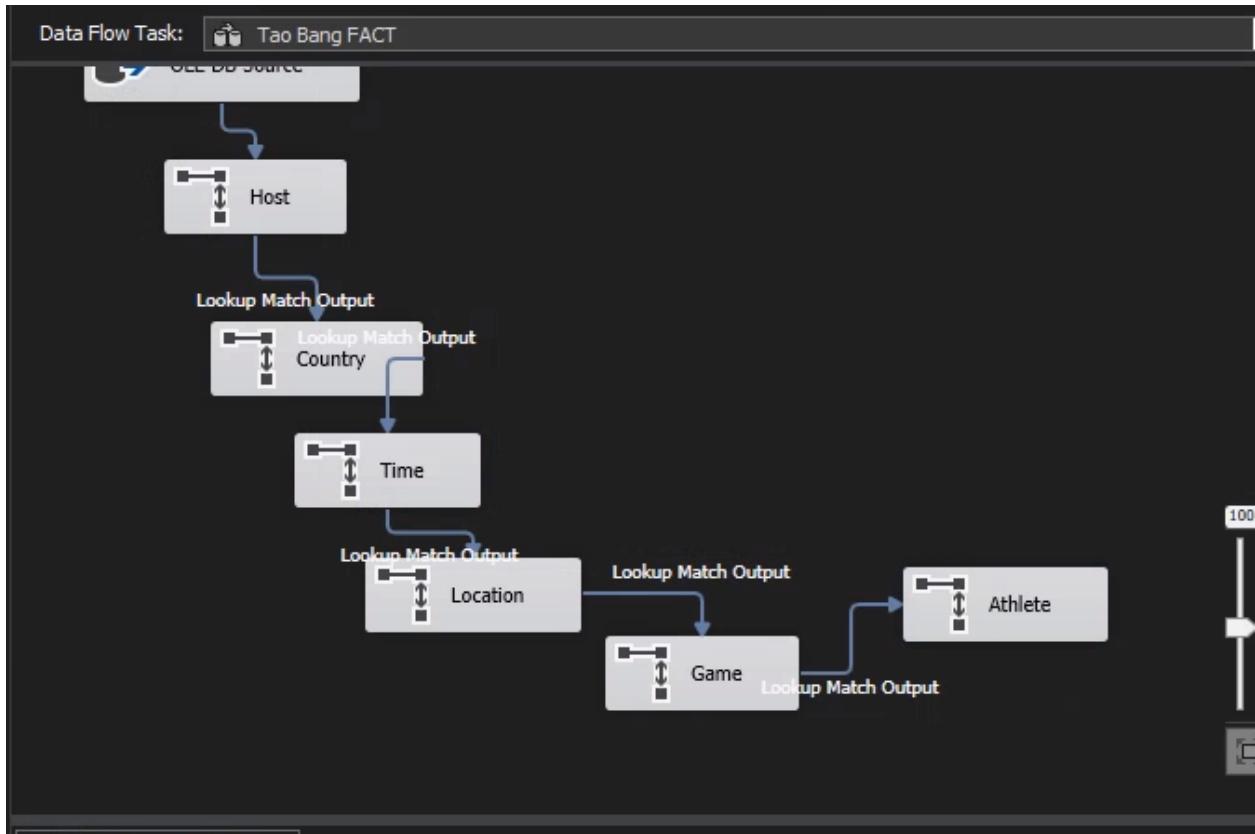
- Chọn connection tới database Olympic_DW và chọn bảng Dim_Country.



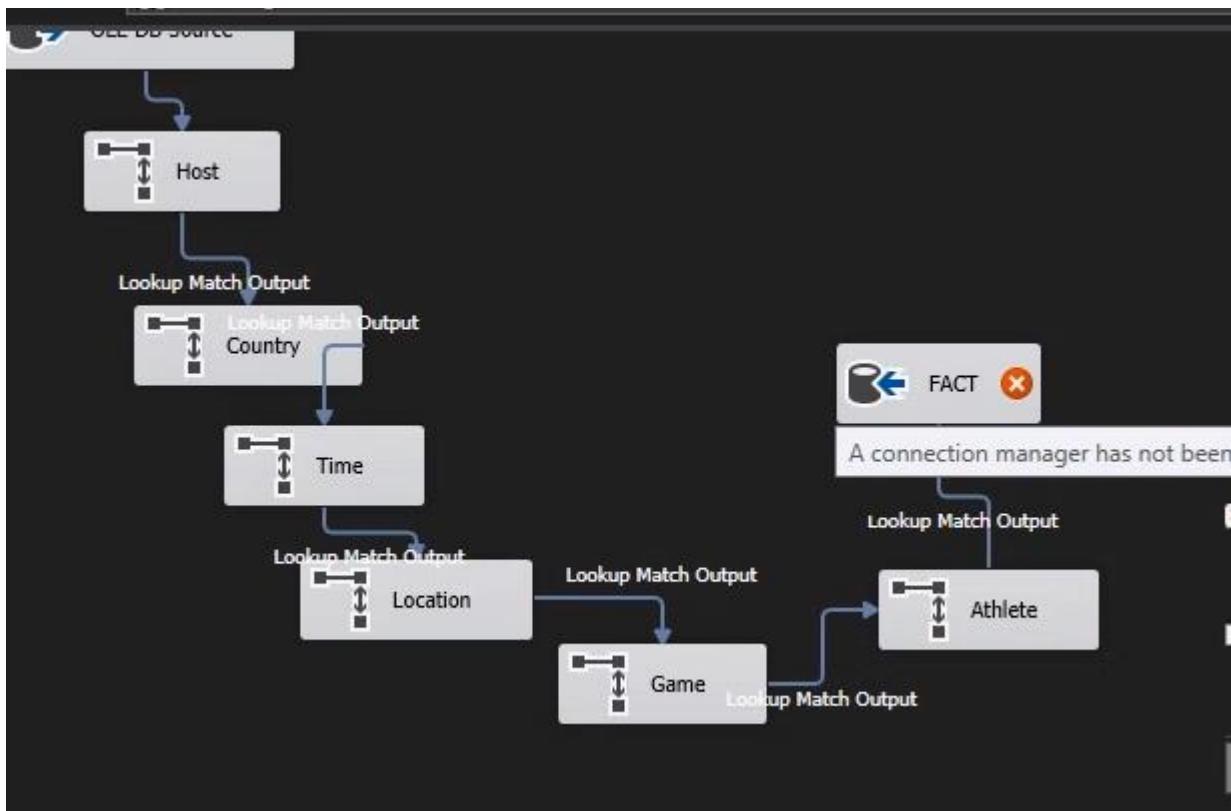
- Chọn cột look up và chọn OK



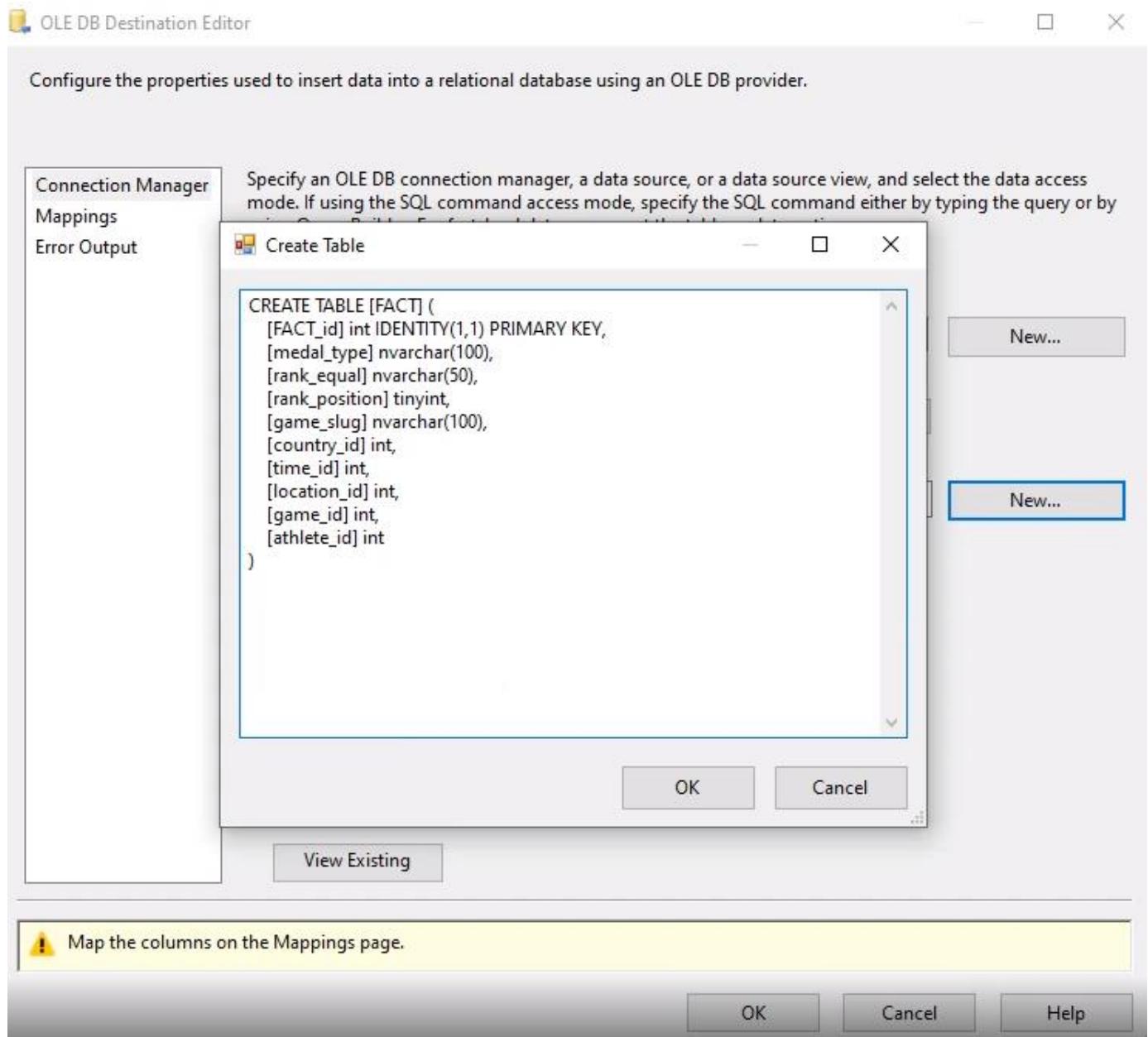
- Ta sẽ thực hiện tương tự như vậy với các bảng Dimension còn lại.
- Cuối cùng ta sẽ có được kết quả như hình sau:



- Tiếp theo ta tạo mới OLE DB Destination, đặt tên là FACT

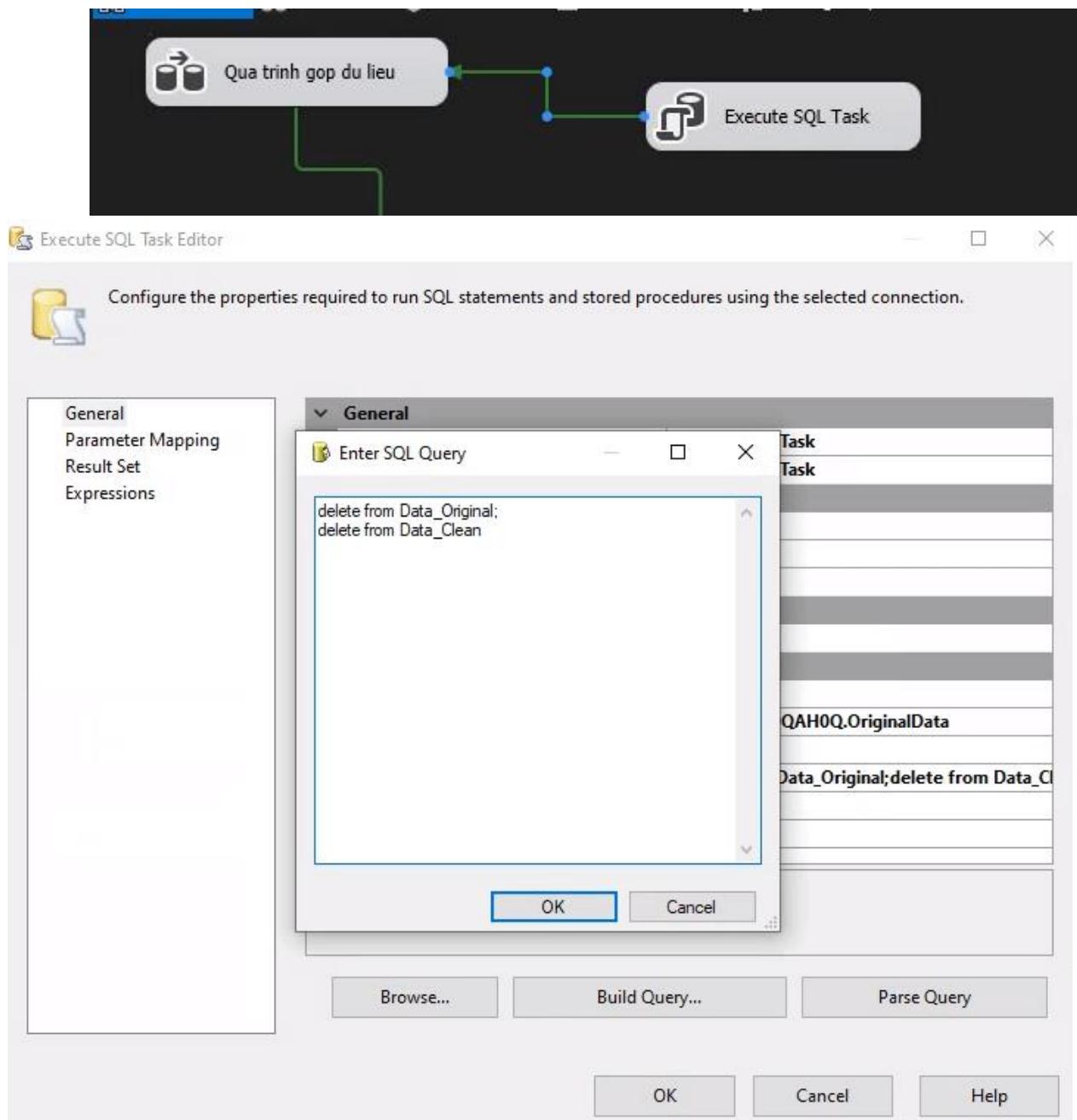


- Tiến hành Edit để tạo mới 1 bảng FACT trong kho dữ liệu

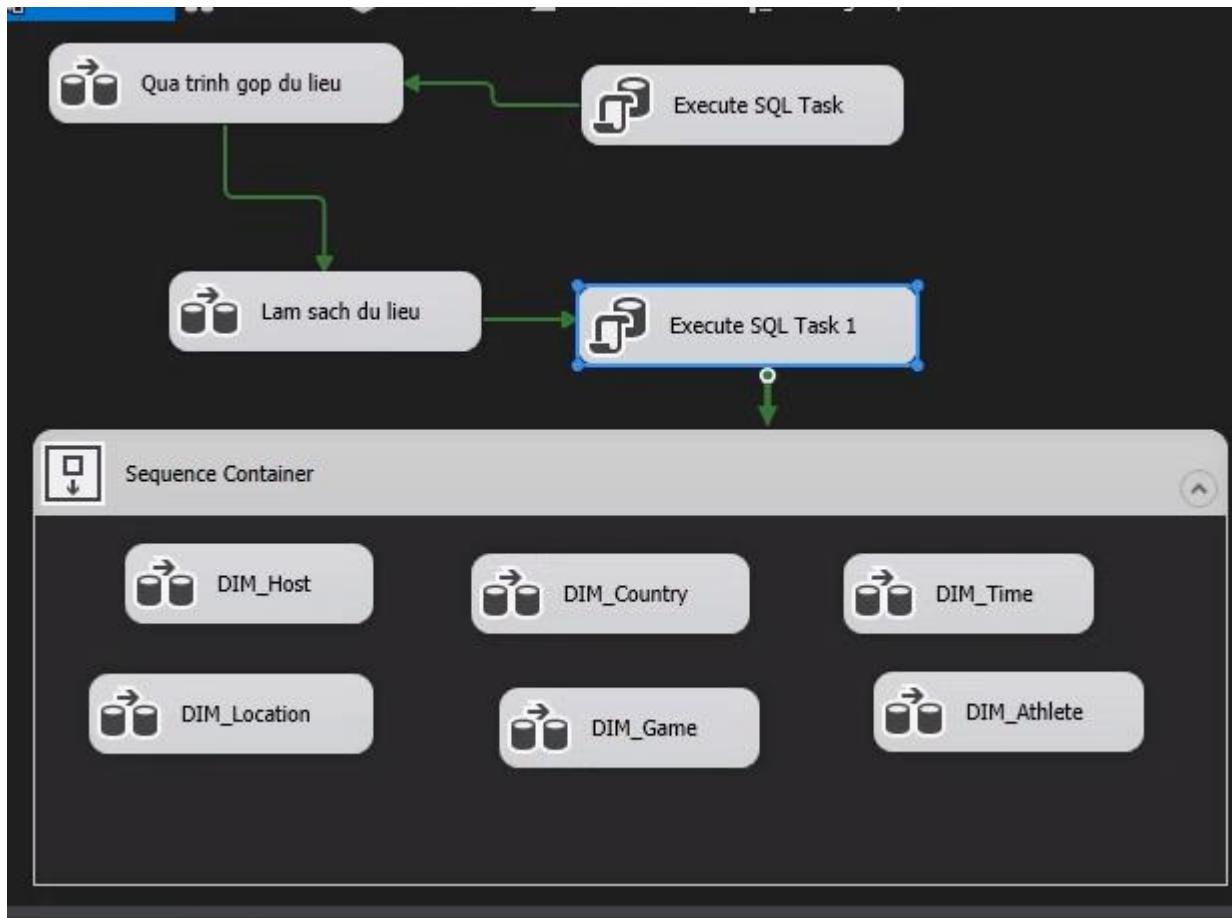


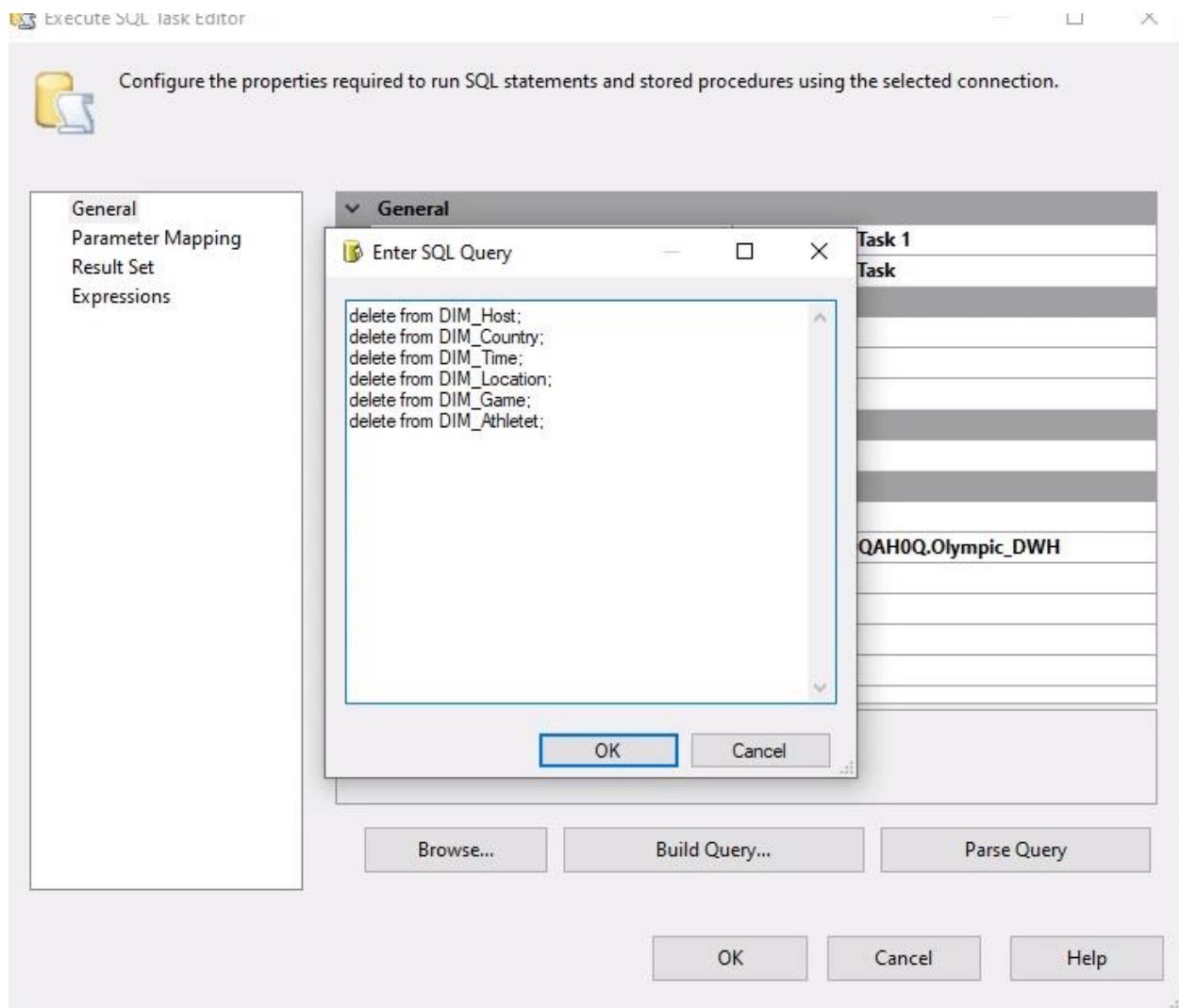
Tạo Execute SQL Task

- Xóa bảng Data_Original và Data_Clean

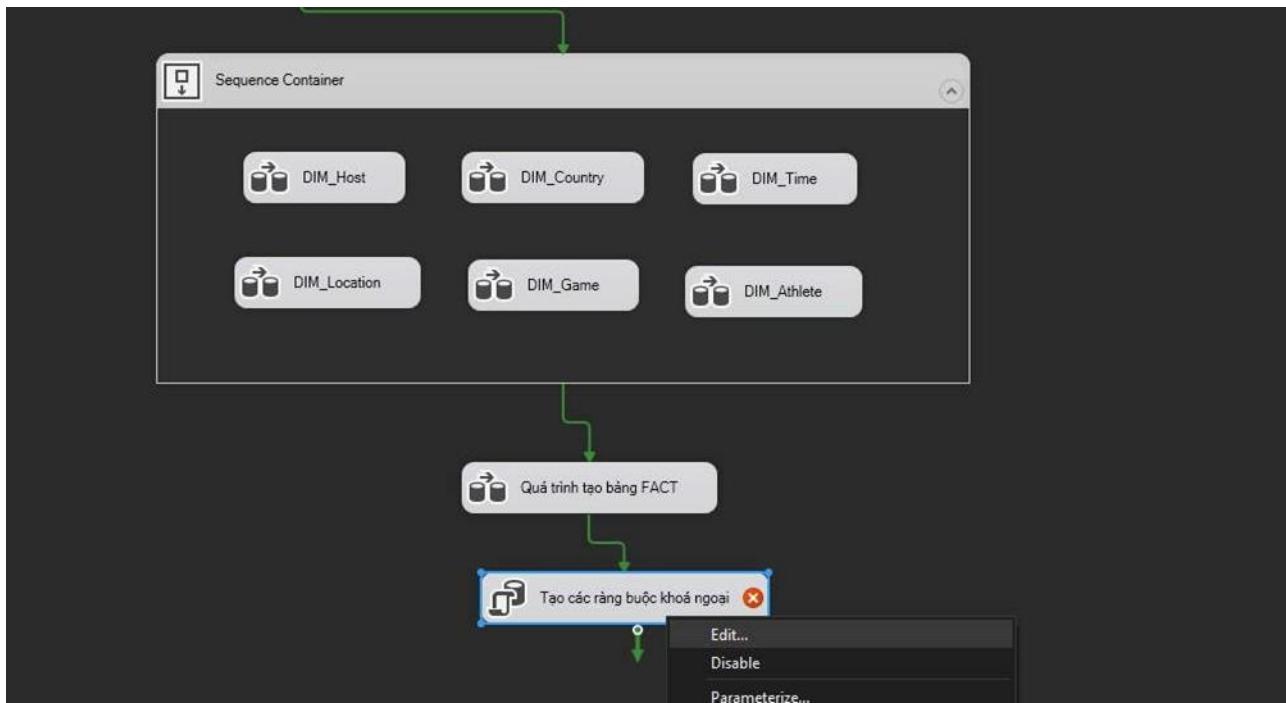


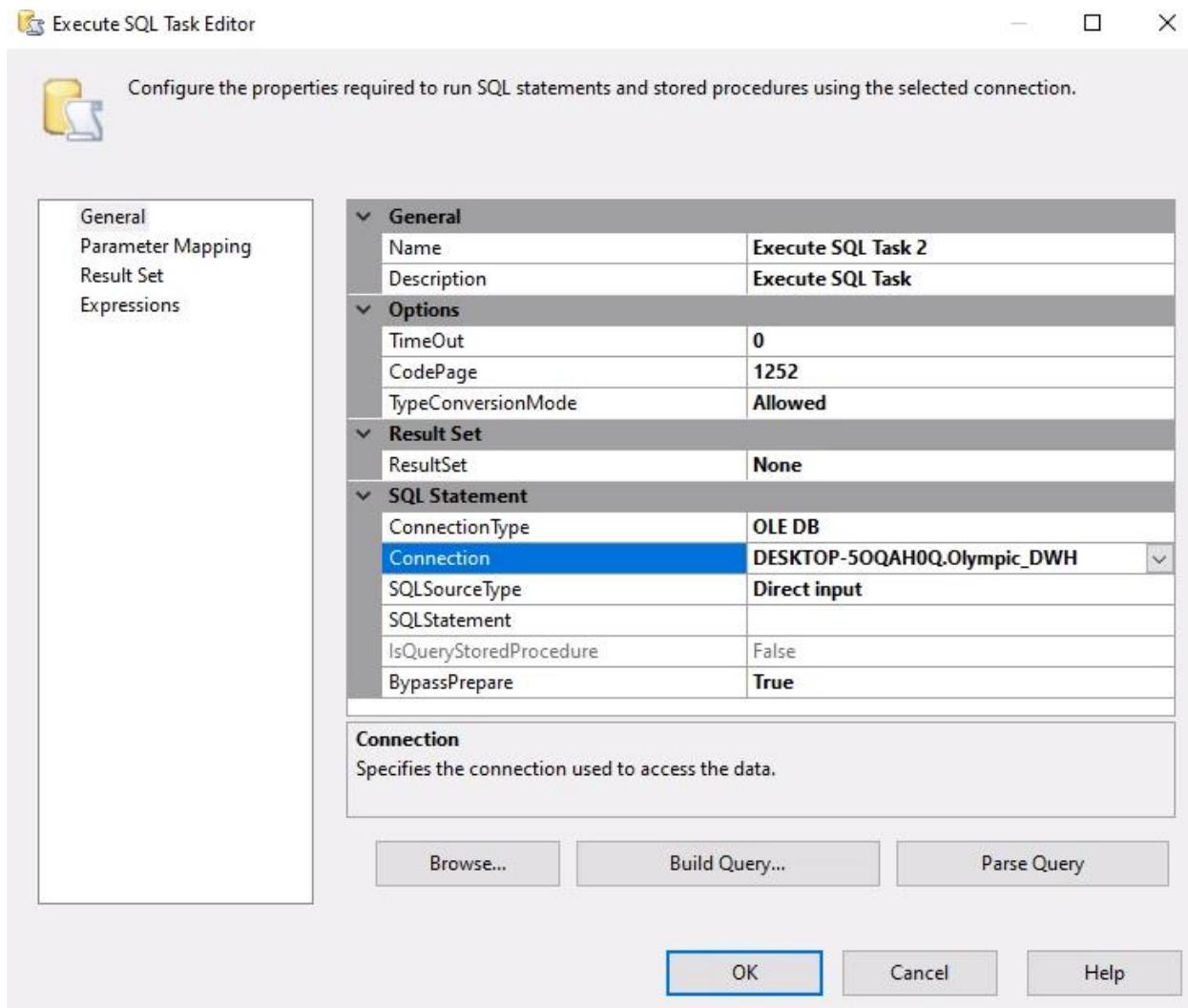
- Xóa các bảng Dimension





- Tạo ràng buộc giữa bảng FACT và các Dimension





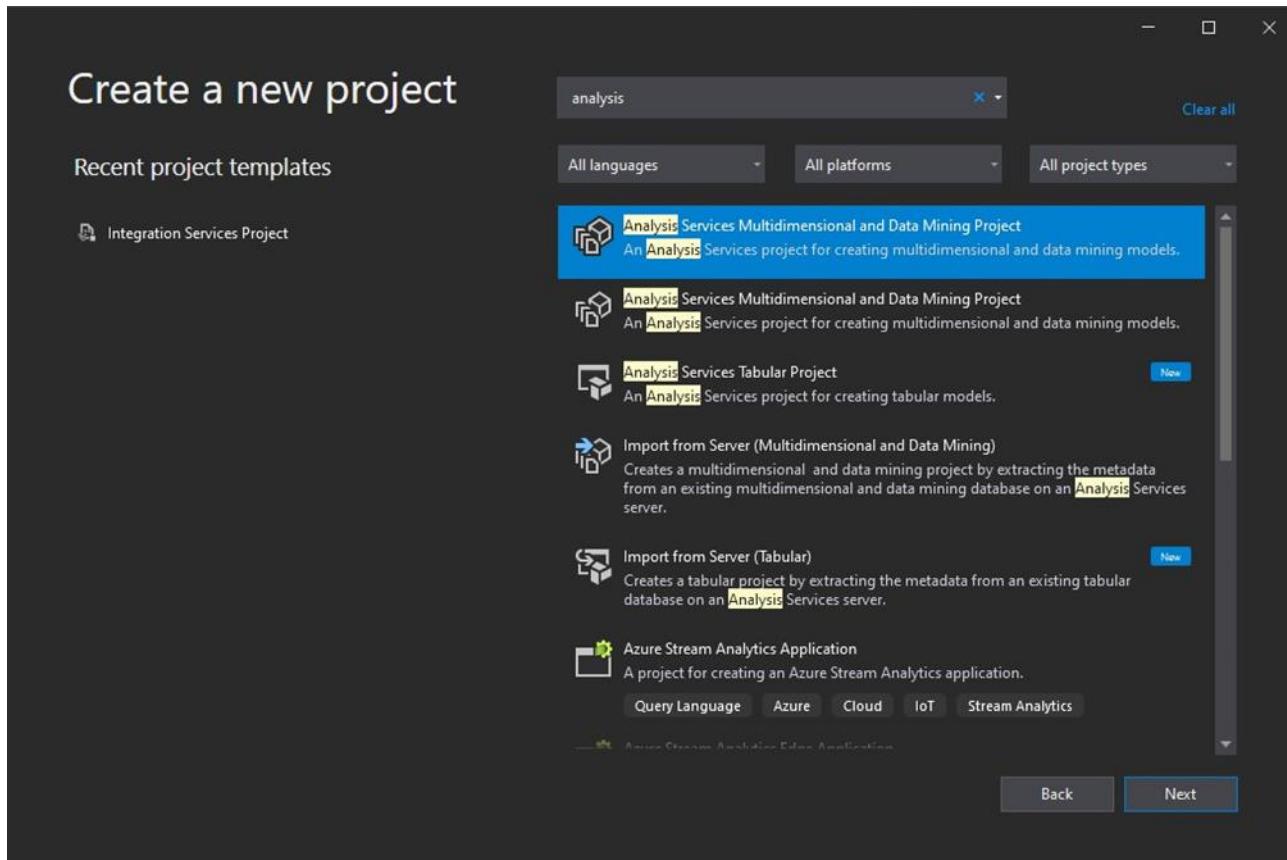
The screenshot shows the 'Enter SQL Query' dialog box. The title bar includes the Bandicam watermark. The query editor displays the following SQL code:

```
ALTER TABLE FACT ADD CONSTRAINT FK_FACT_HOST FOREIGN KEY(game_slug) REFERENCES DIM_Host(game_slug)
ALTER TABLE FACT ADD CONSTRAINT FK_FACT_COUNTRY FOREIGN KEY(country_id) REFERENCES DIM_Country(country_id)
ALTER TABLE FACT ADD CONSTRAINT FK_FACT_TIME FOREIGN KEY(time_id) REFERENCES DIM_Time(time_id)
ALTER TABLE FACT ADD CONSTRAINT FK_FACT_LOCATION FOREIGN KEY(location_id) REFERENCES DIM_Location(location_id)
ALTER TABLE FACT ADD CONSTRAINT FK_FACT_GAME FOREIGN KEY(game_id) REFERENCES DIM_Game(game_id)
ALTER TABLE FACT ADD CONSTRAINT FK_FACT_ATHLETES FOREIGN KEY(athlete_id) REFERENCES DIM_Athletes(athlete_id)
```

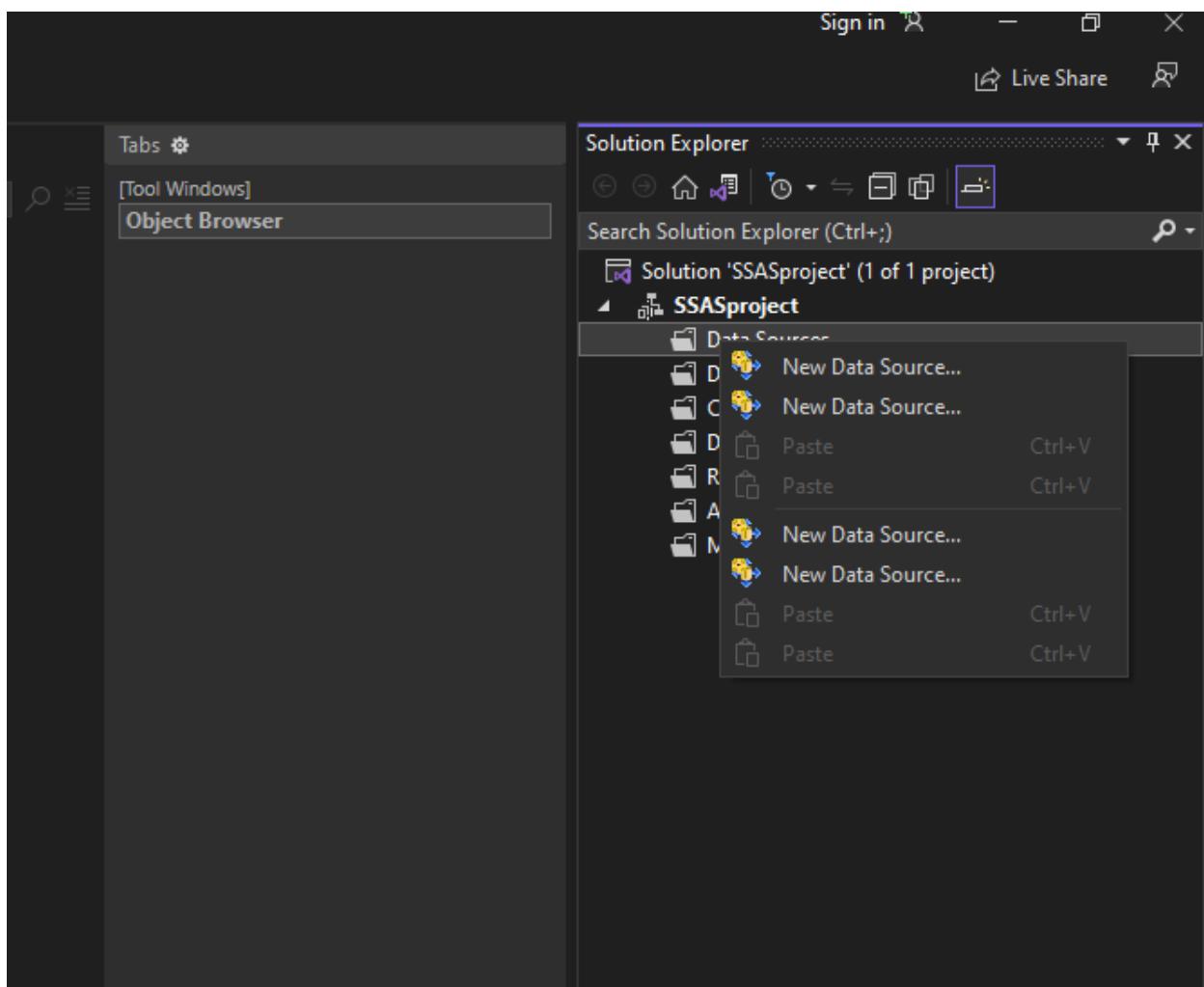
CHƯƠNG 3. PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS).

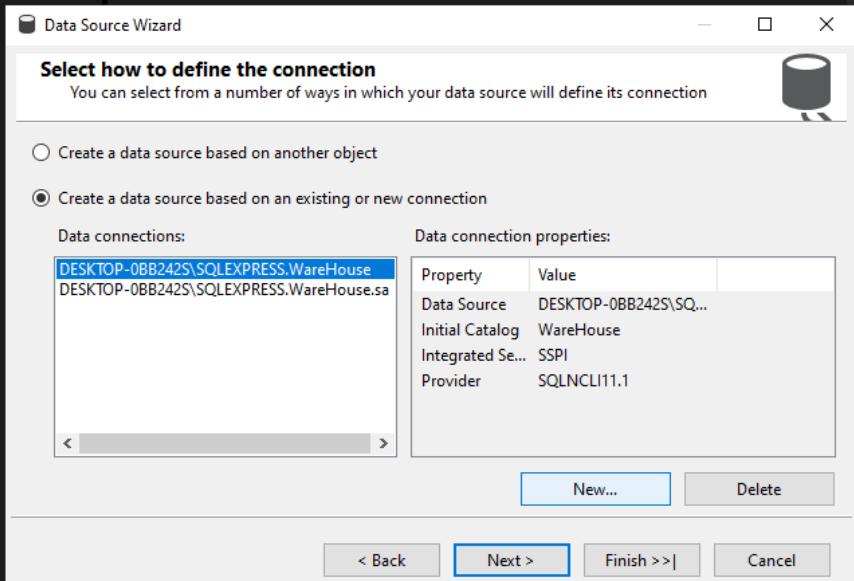
3.1 QUÁ TRÌNH SSAS TRONG VISUAL STUDIO 2022.

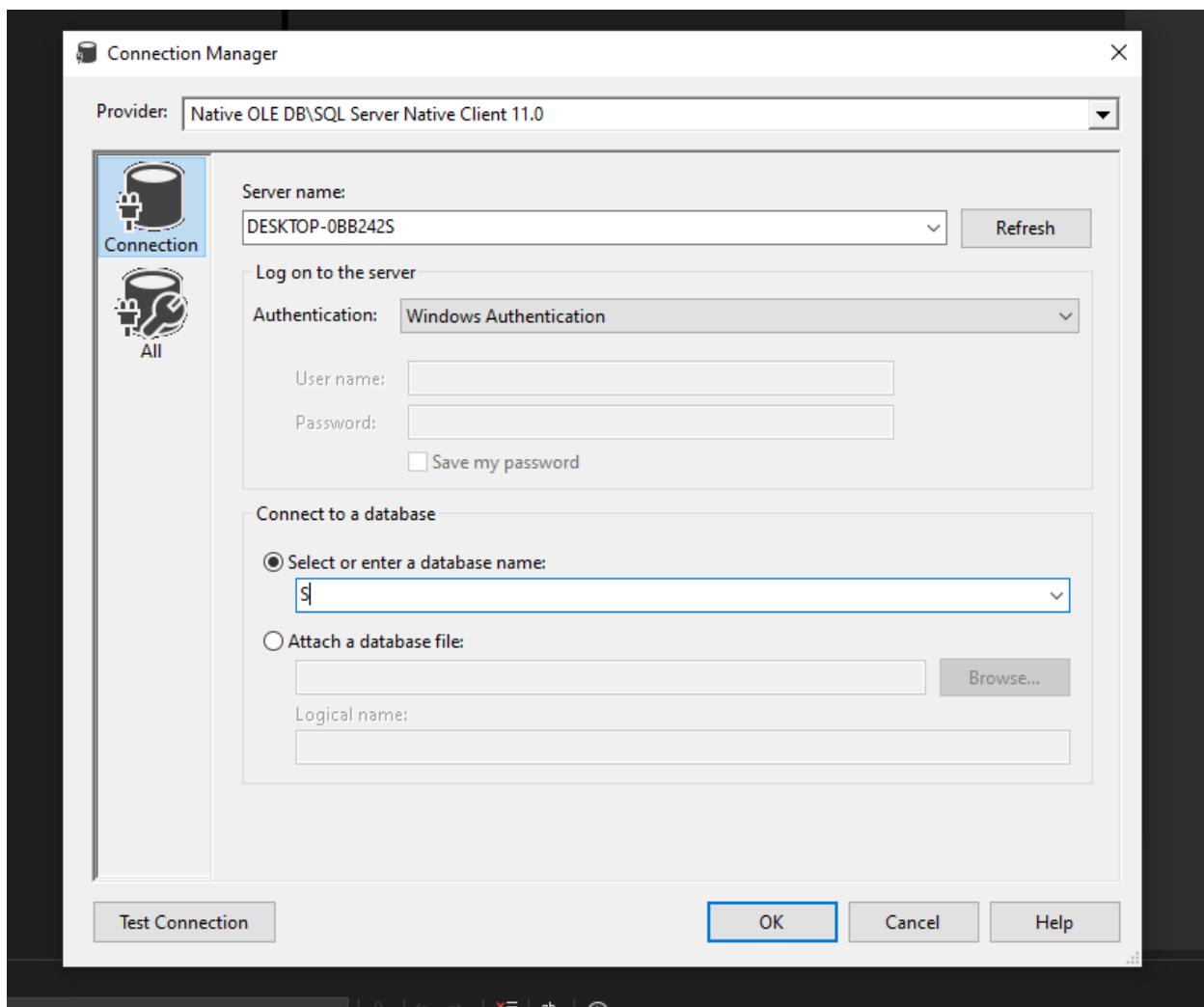
3.1.1 Tạo project trong visual studio

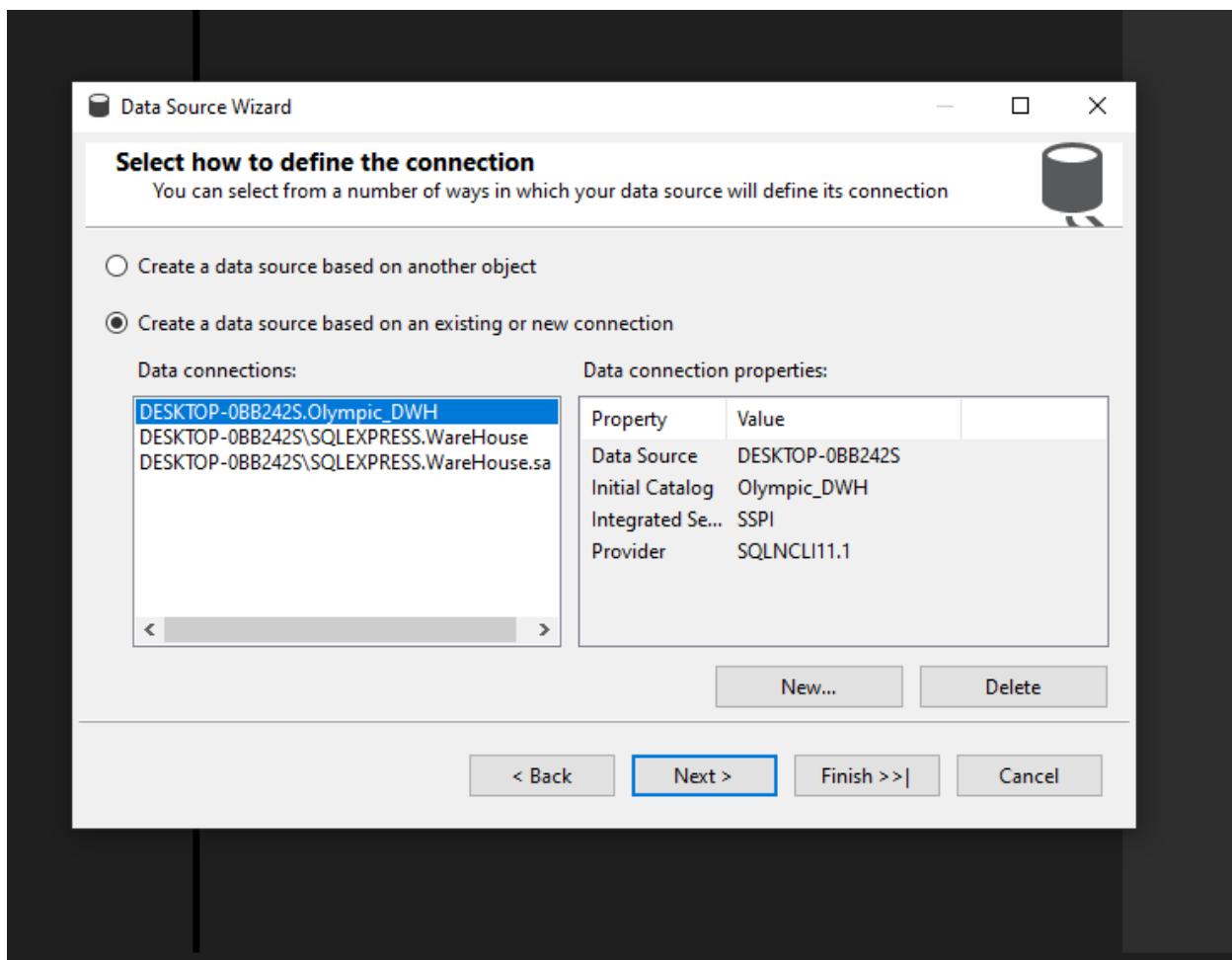


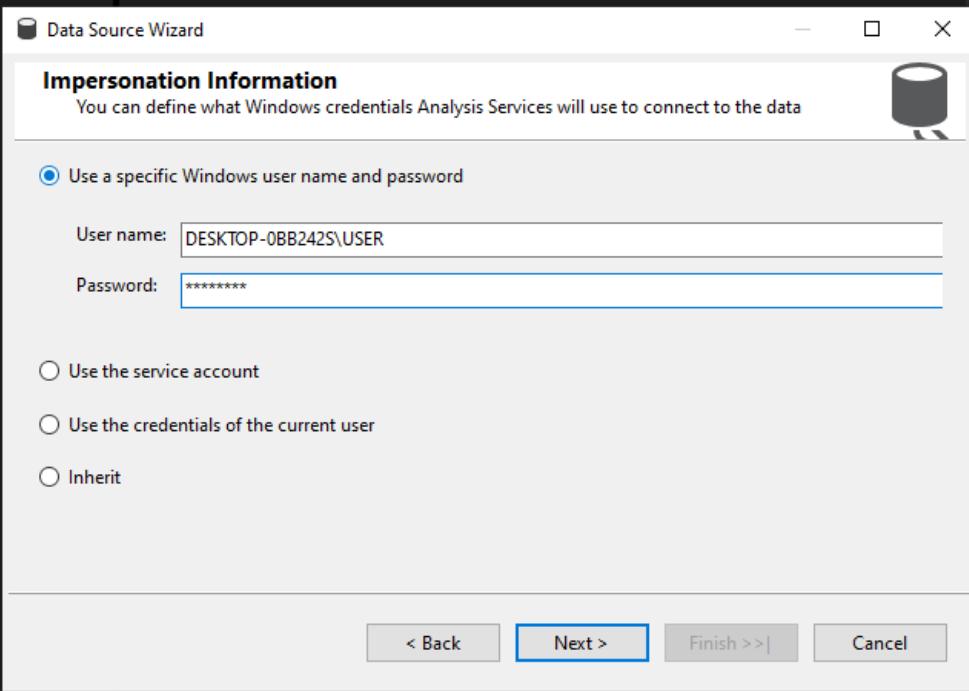
3.1.2 Xác định nguồn dữ liệu – Tạo mới Data Source.



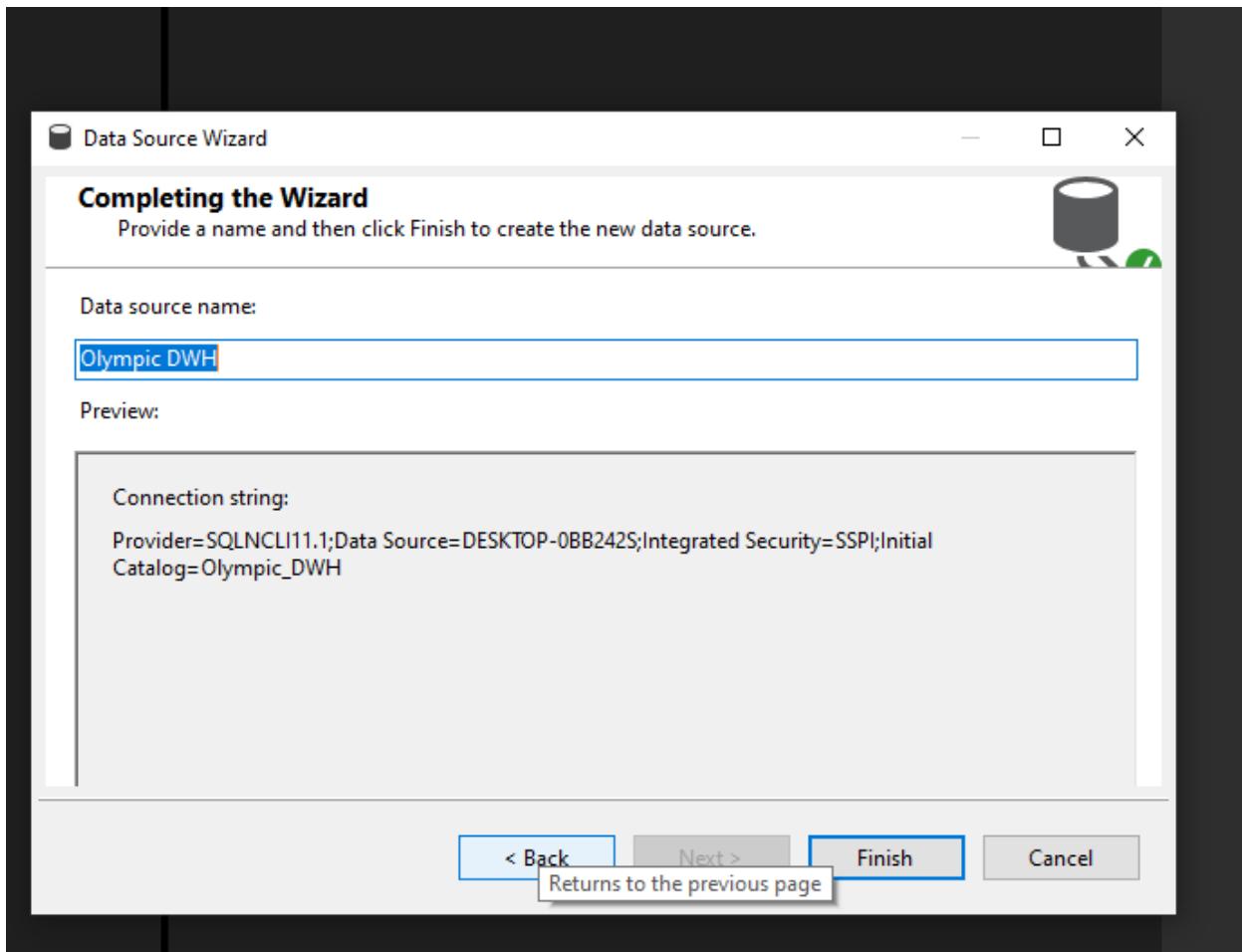




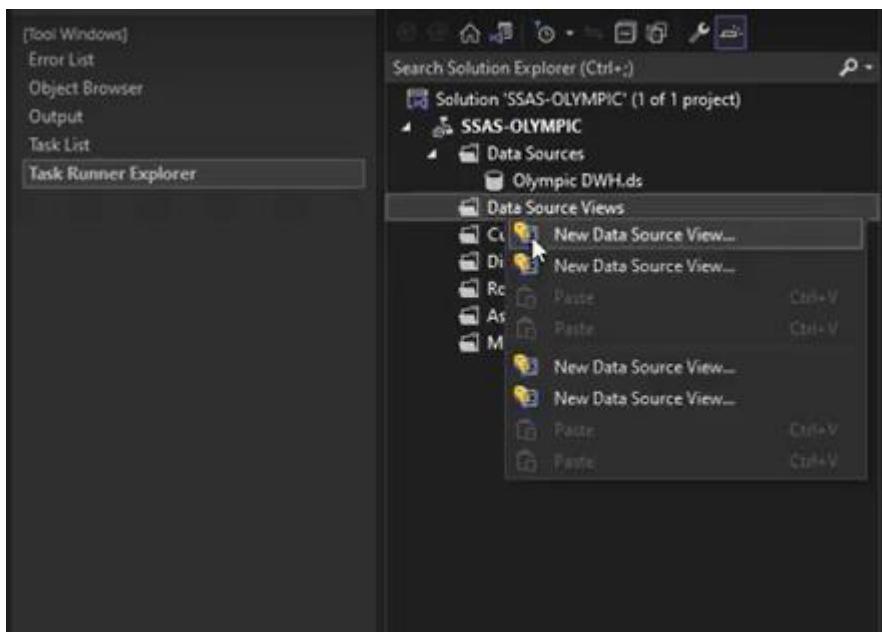


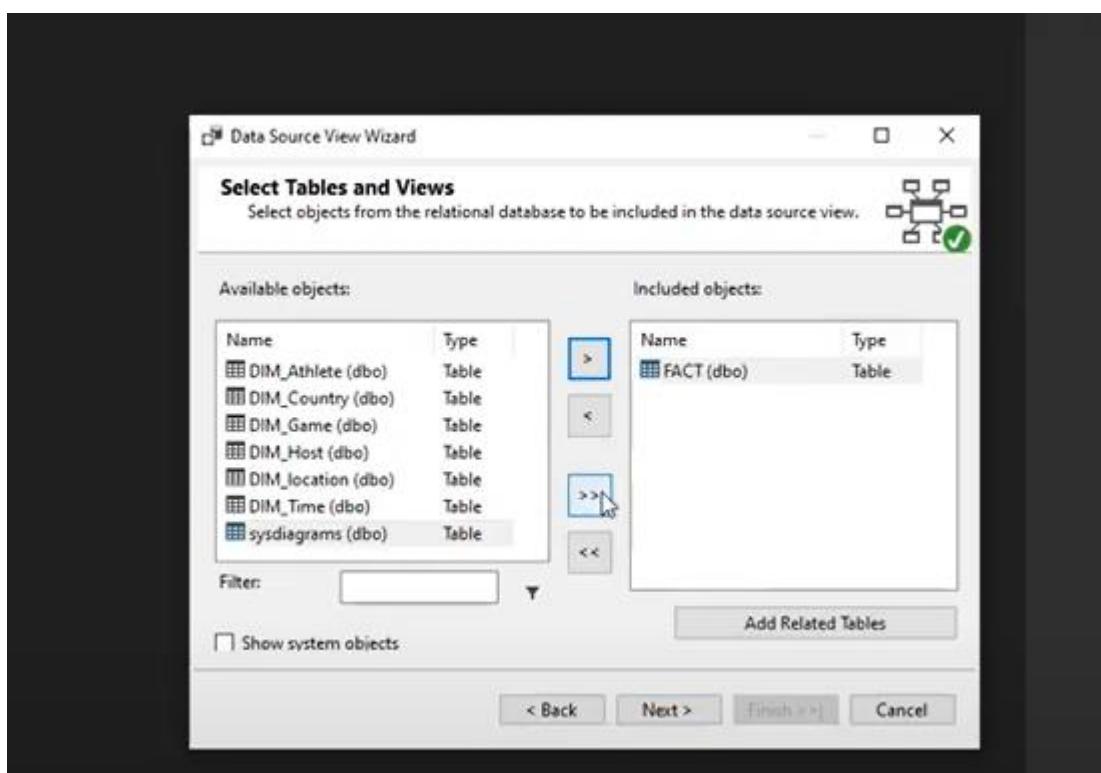
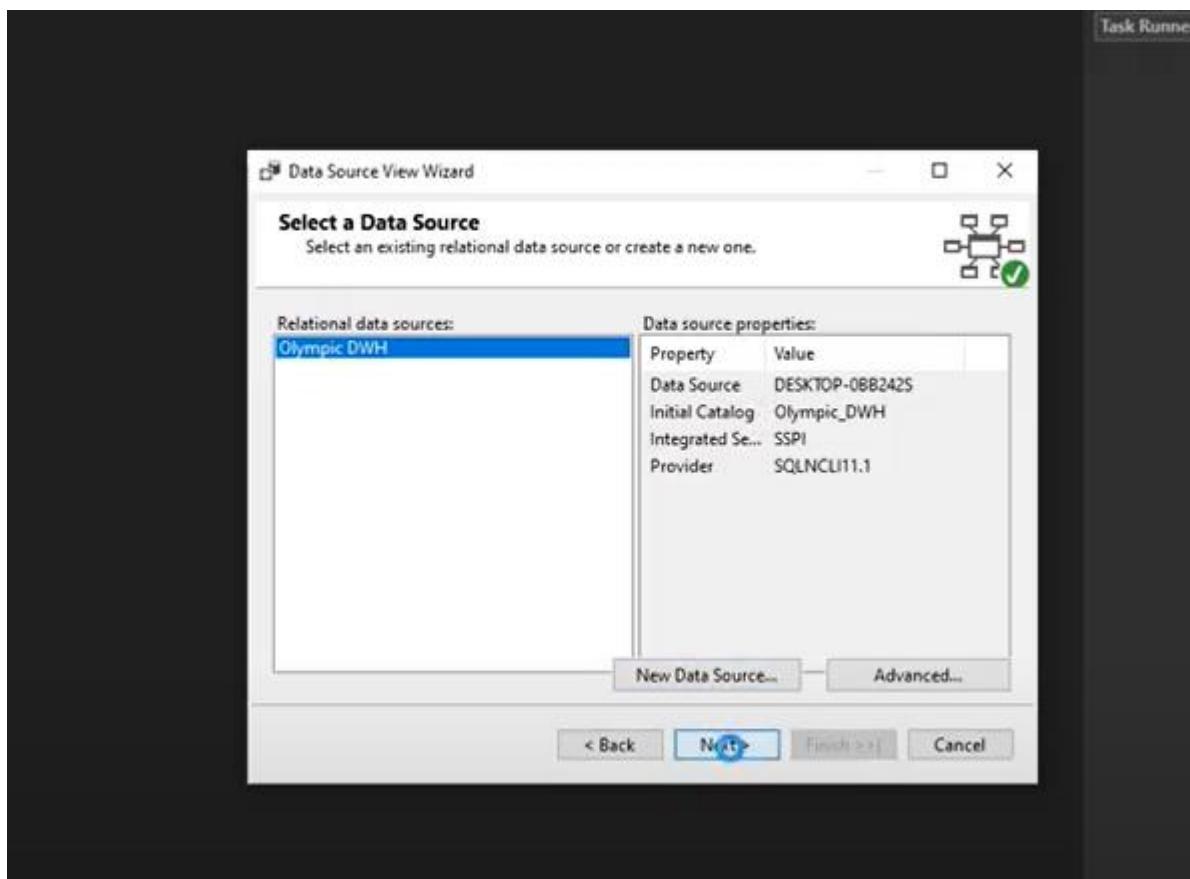


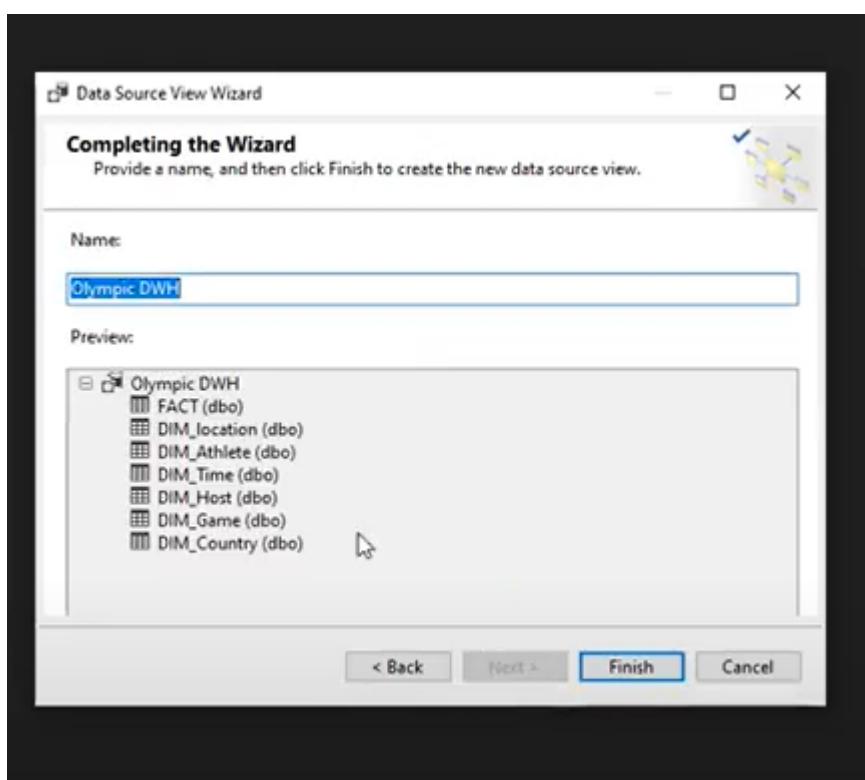
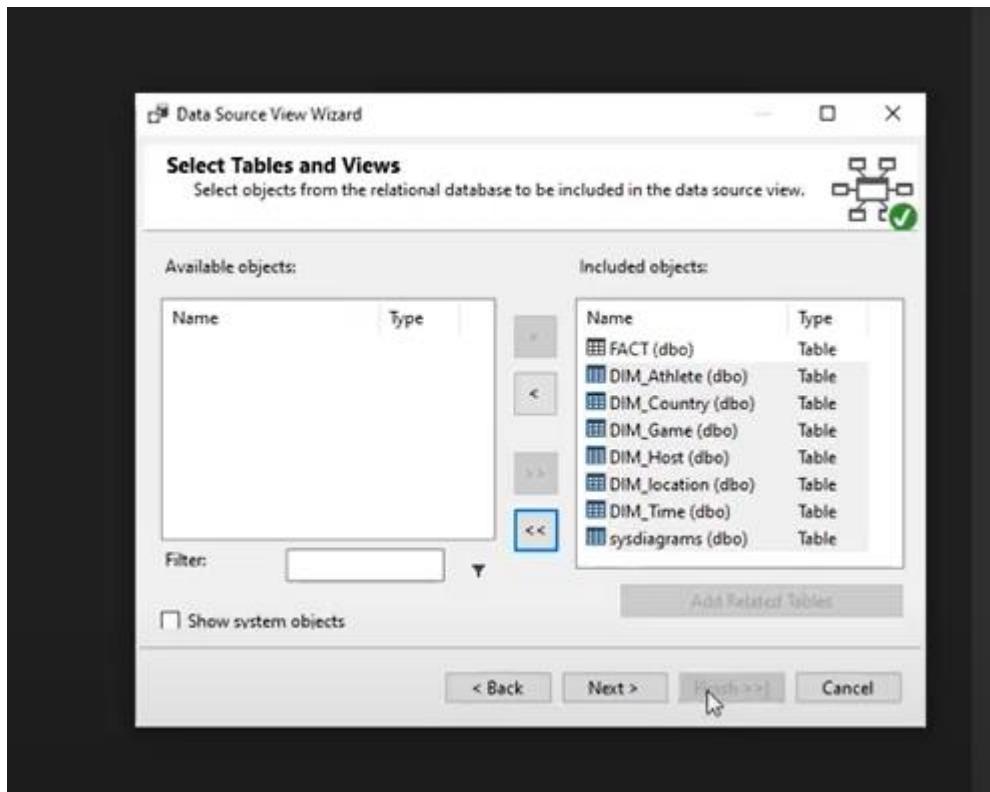
▼ | ⌂ | ← → | × | ⌂ | ⌂ | ⌂

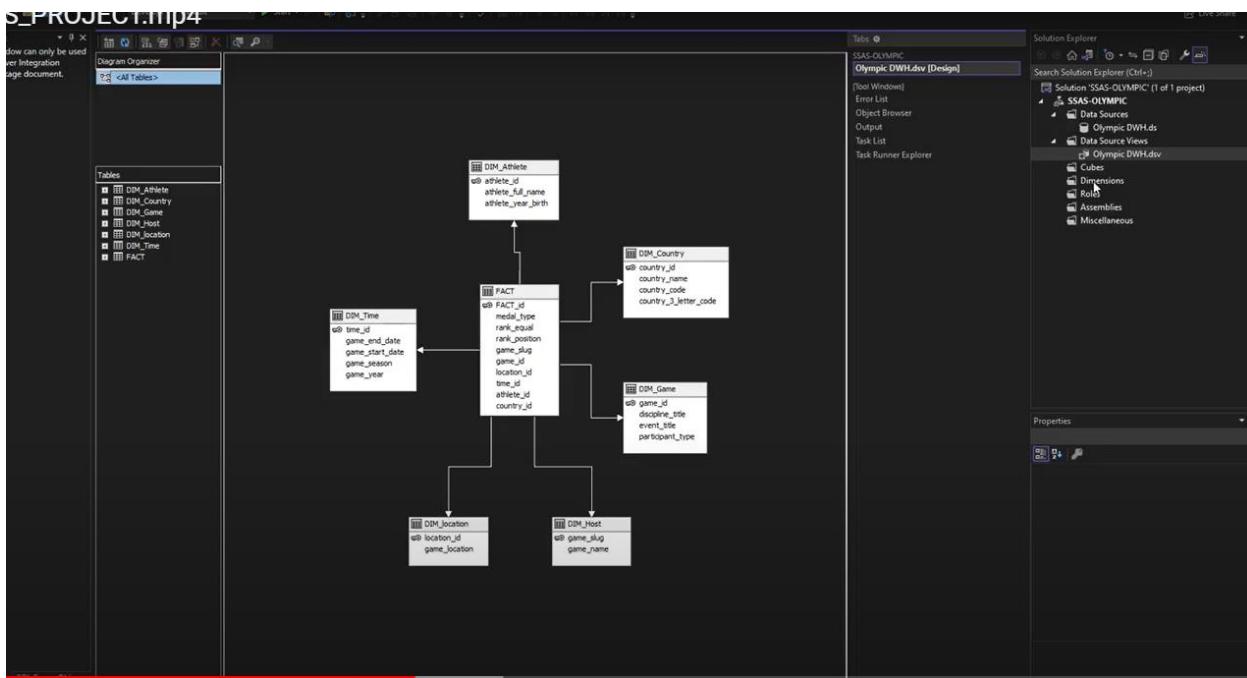


3.1.3 Tạo mới Data Source Views

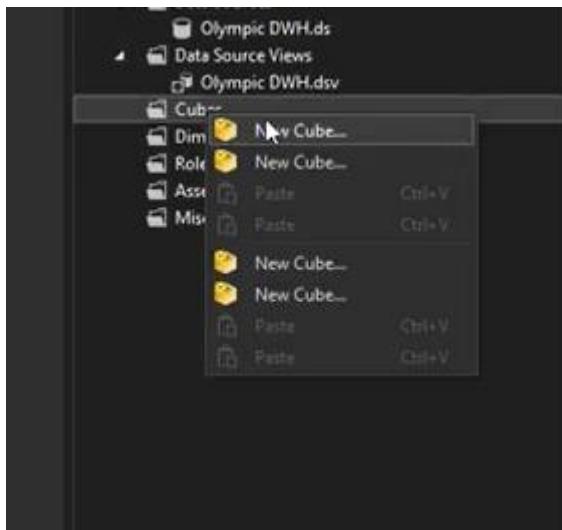


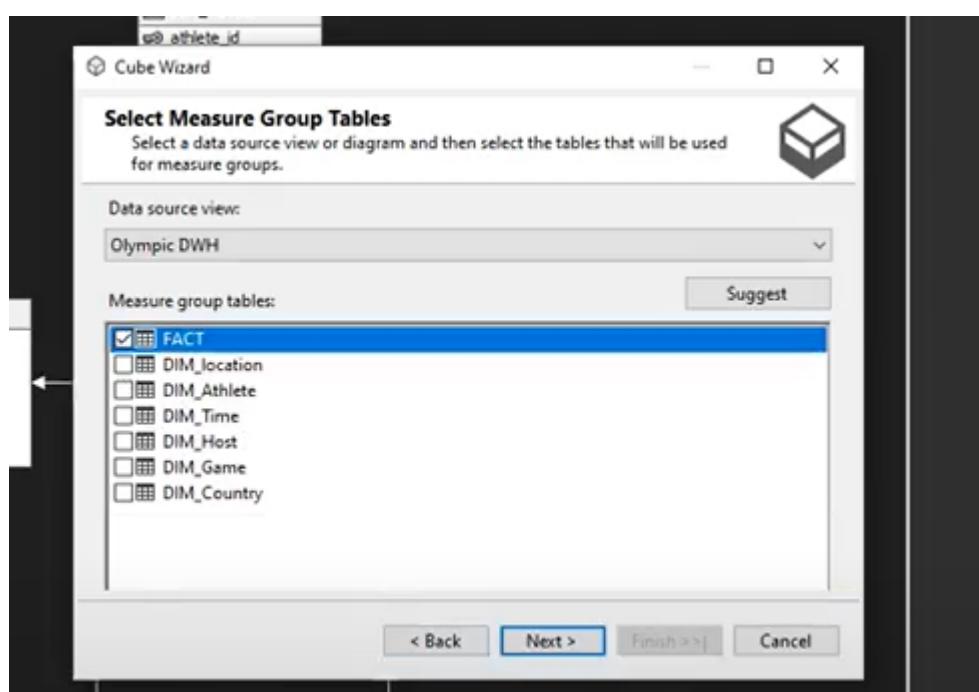
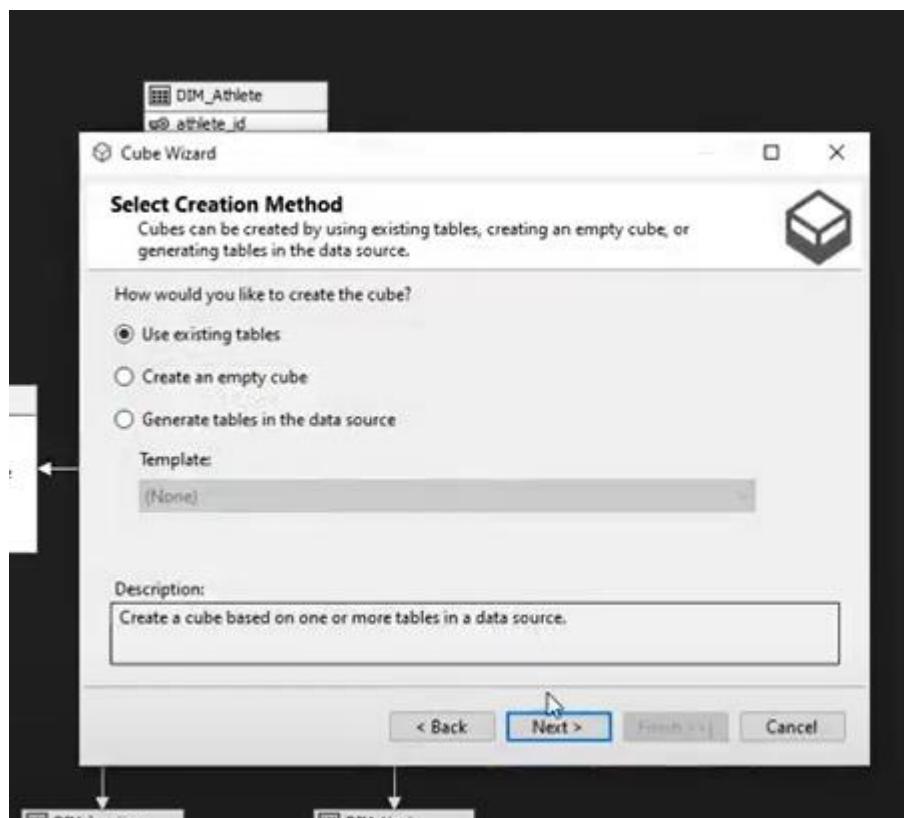


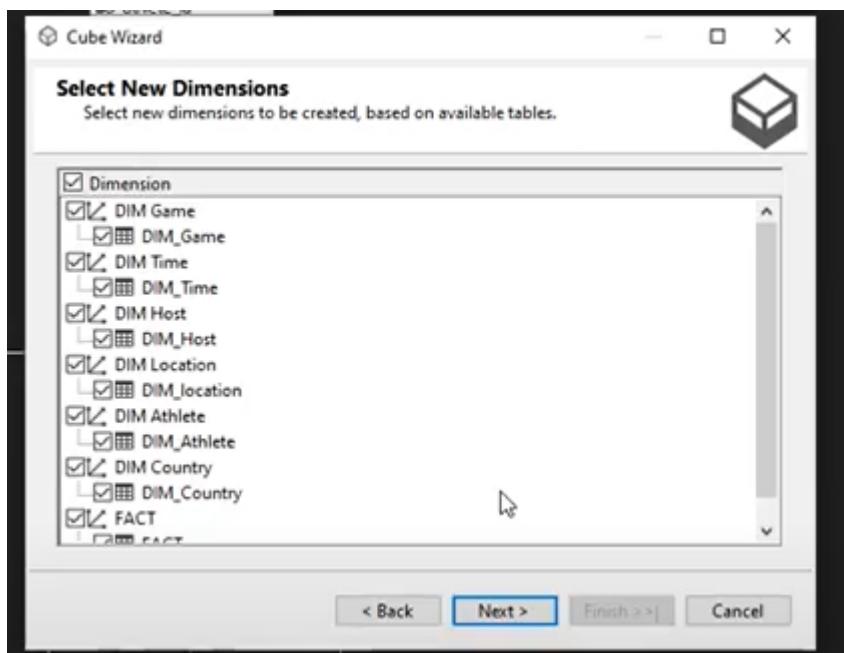
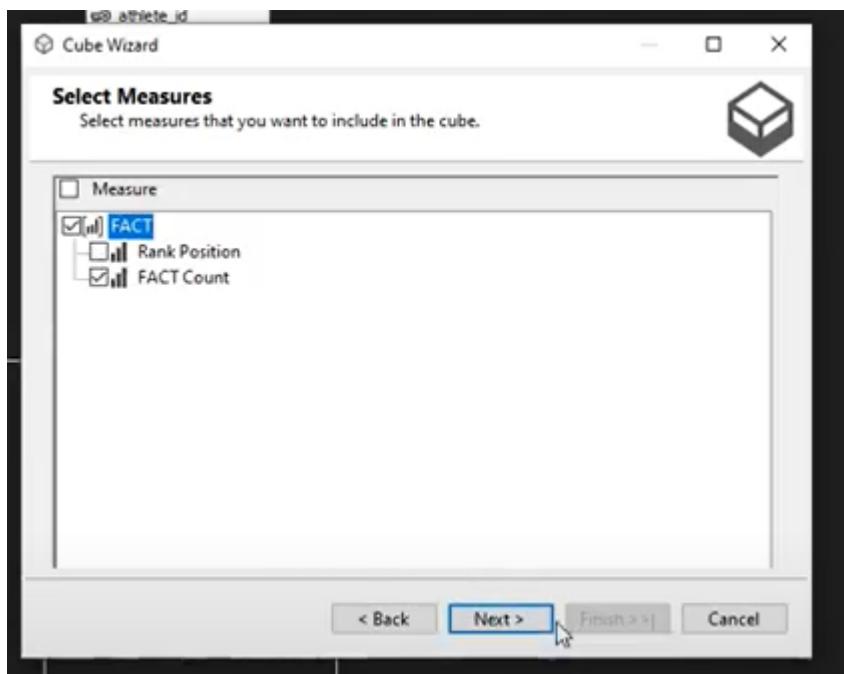


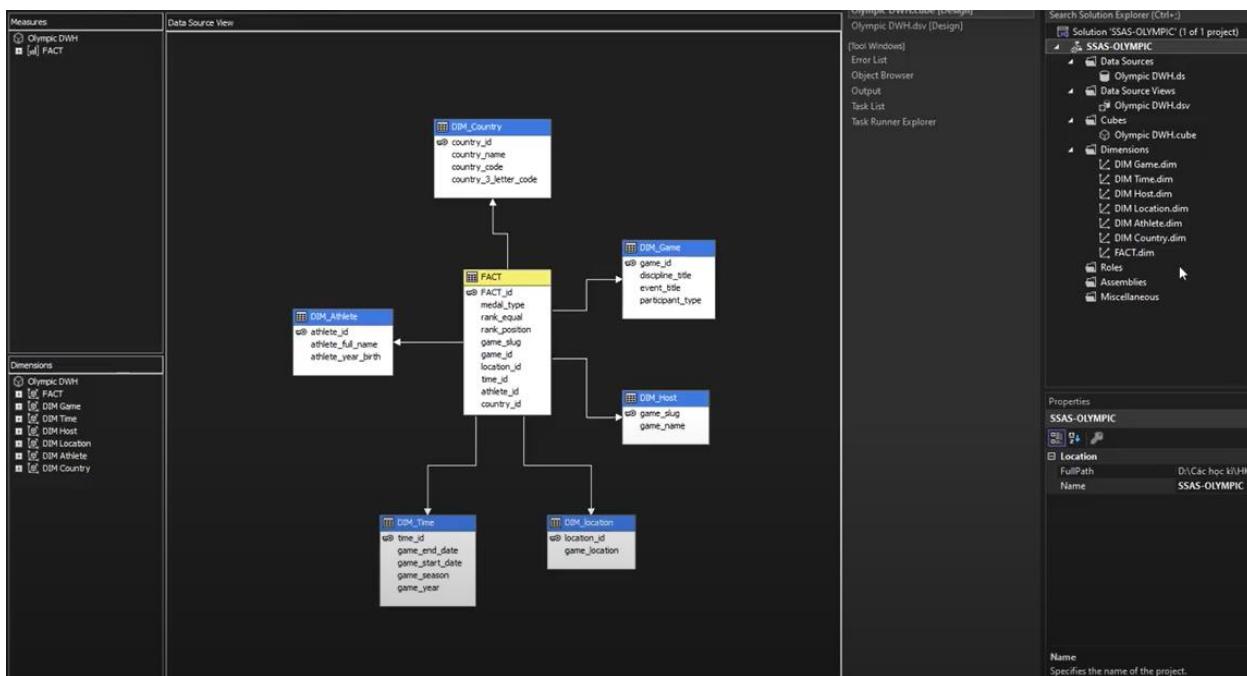


3.1.4 Tạo Cubes







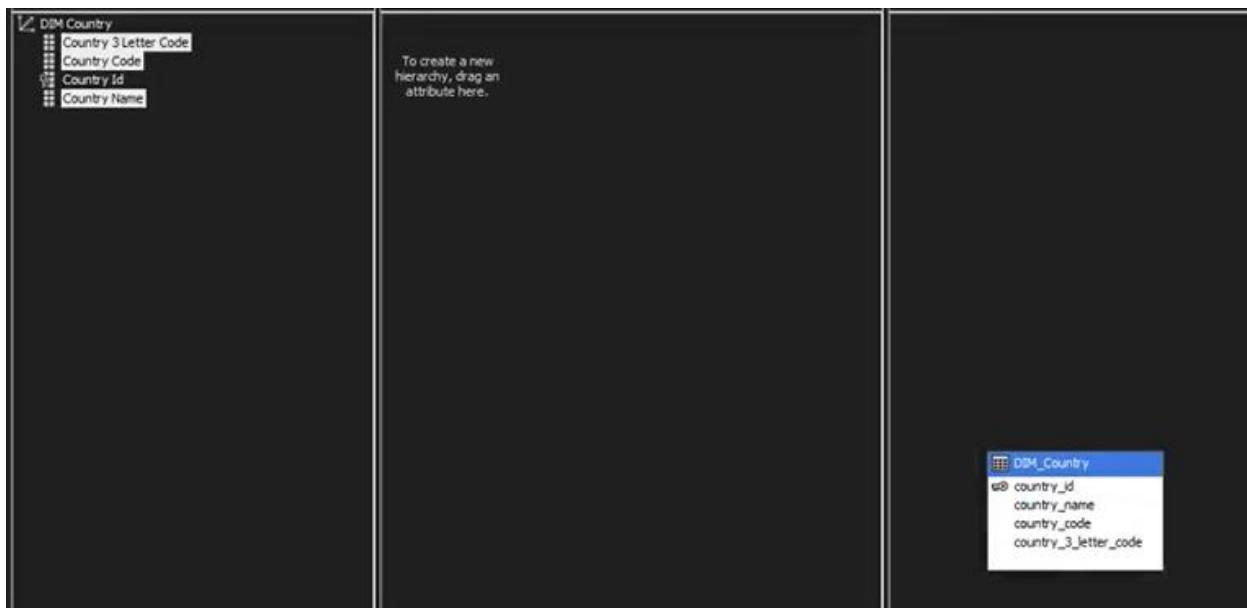
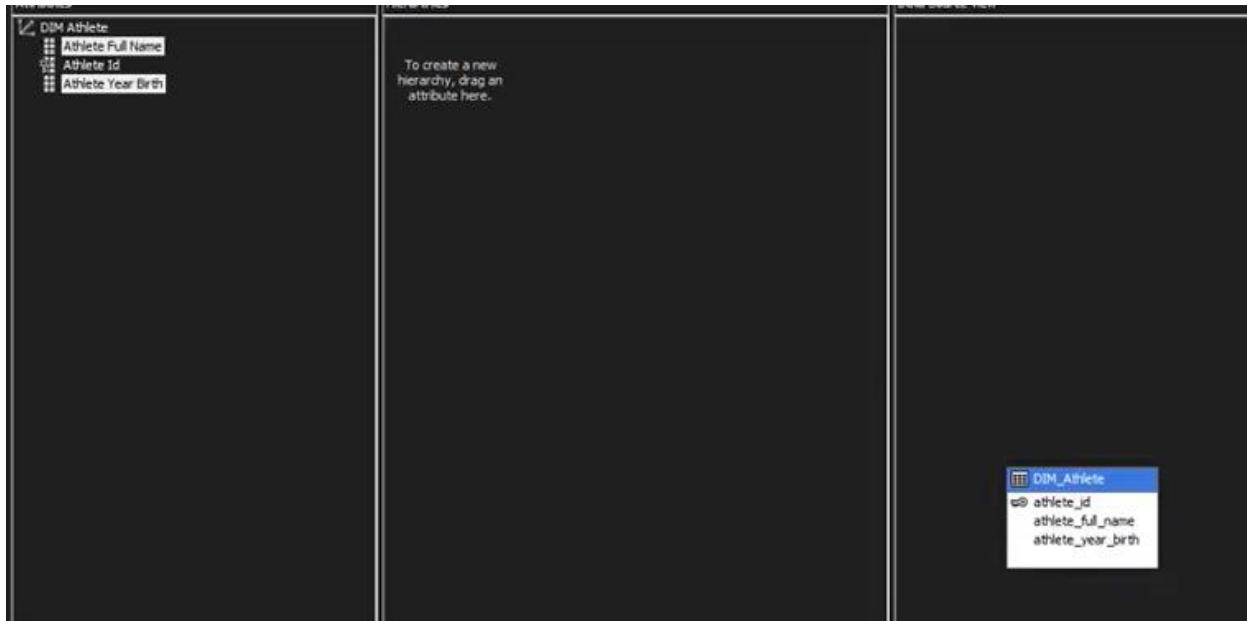


3.1.5 Tạo các bảng Dimension

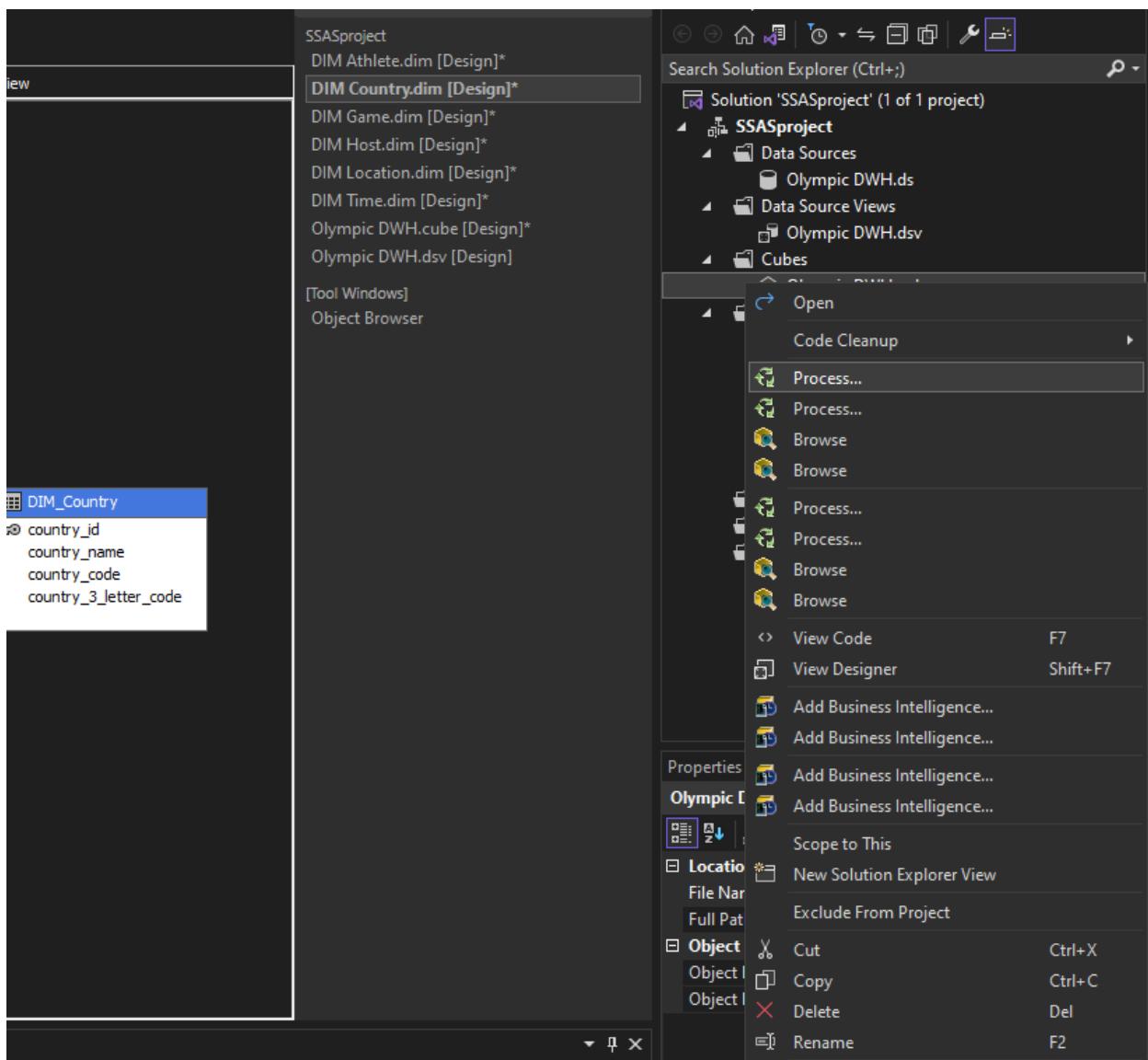
<p>DIM Time</p> <ul style="list-style-type: none">Game End DateGame SeasonGame Start DateGame YearTime Id	<p>To create a new hierarchy, drag an attribute here.</p>	<p>DIM_Time</p> <ul style="list-style-type: none">time_idgame_end_dategame_start_dategame_seasongame_year
<p>Attributes</p> <p>DIM Game</p> <ul style="list-style-type: none">Discipline TitleEvent TitleGame IdParticipant Type	<p>Hierarchies</p> <p>To create a new hierarchy, drag an attribute here.</p>	<p>Data Source View</p> <p>DIM_Game</p> <ul style="list-style-type: none">game_iddiscipline_titleevent_titleparticipant_type

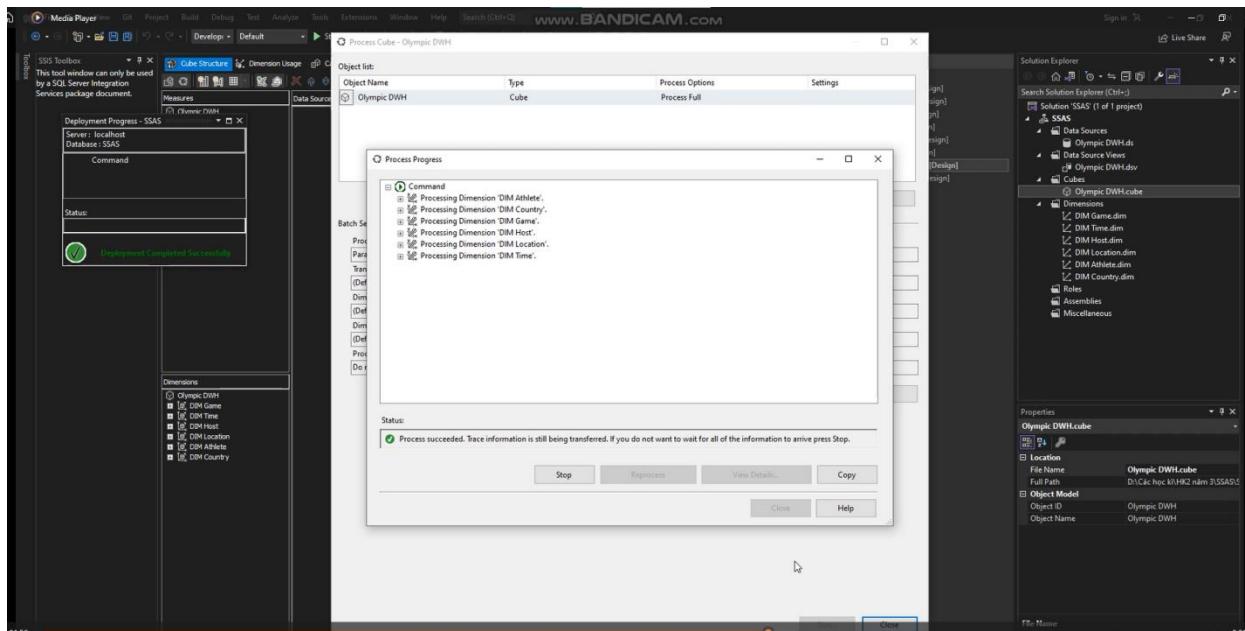
Attributes	Hierarchies	Data Source View
<ul style="list-style-type: none"><input checked="" type="checkbox"/> DIM Host<ul style="list-style-type: none">Game NameGame Slug	To create a new hierarchy, drag an attribute here.	 <p>DIM_Host</p> <ul style="list-style-type: none">game_sluggame_name

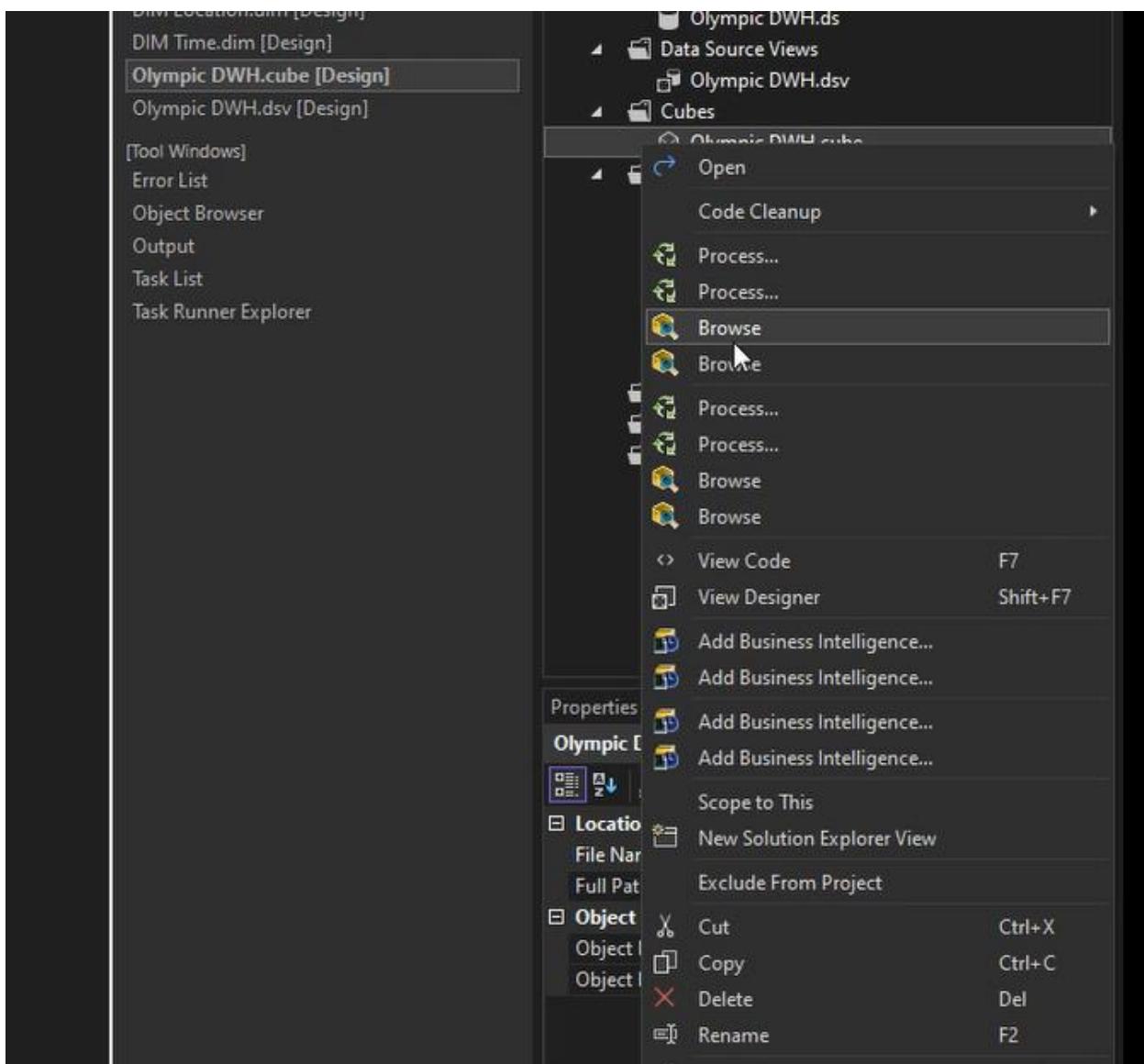
Attributes	Hierarchies	Data Source View
<ul style="list-style-type: none"><input checked="" type="checkbox"/> DIM Location<ul style="list-style-type: none">Game LocationLocation Id	To create a new hierarchy, drag an attribute here.	 <p>DIM_Location</p> <ul style="list-style-type: none">location_idgame_location



- Tiến hành process cube







3.2 THỰC HIỆN CÁC CÂU TRUY VẤN

Câu 1: Liệt kê số lần tham gia thế vận hội Olympic Games các vận động viên có năm sinh bằng 1900.

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

Edit as Text Import... MDX

Language: Default

Dimension Hierarchy Operator Filter Expression Parameters

DIM Athlete Athlete Year Birth Equal { 1990 }

<Select dimension>

Athlete Full Name	FACT Count
Aaron ROYLE	2
Abdelali DARAA	1
Abdulaziz ALSHATTI	1
Abeba AREGAWI	1
Abigail IROZURU	1
Abigail JOO	2
Adhraf KHARROUBI	1
Adam NOWICKI	1
Adam ZAMPA	9
Adel MECHAAL	1
Adlan ABDURASHIDOV	1
Adrian BLOCKI	1
Adrian DZIOLKO	1
Adrienn TOTH	1
Agata PERENC	1
Agnese KOKLACA	1
Ahmed MEJRI	1
Ahmed Mohamed AL-MERJABI	1
Ahreum NA	3
Ahti TOIVANEN	2
Aisling Margaret COONEY	1
Ajmal FAISAL	1
Akari OGATA	1
Akiko NAKAMURA	2
Alaaeldin ABOUELKASSEM	3
Alan GOGAEV	1
Albert TORRES BARCELO	1
Alberto Ezequiel MELIAN	2
Alberto Ignacio PALMETTA	1
Alberto MINO	1
Aleh YURENIA	1
Aleksander ANDRIENKO	2

```
  Select [Measures].[FACT Count] on columns,
  [DIM Athlete].[Athlete Full Name].Members on Rows
  From [Olympic DWH]
  Where [DIM Athlete].[Athlete Year Birth].&[1900];
```

100 %

Messages Results

	FACT Count
All	314
Aame Aatami PEUSSA	2
Achille SOUCHARD	1
Adolfo SCHLEGEL	1
Albert AESCHMANN	3
Alberto CAVALLERO	1
Alfred Harold HOPKINS	1
Alfredo DINALE	1
Alfredo PORZIO	1
Amaury DE LA MOUSSAYE	1
André LAUSEIG	1
Anton ZWERINA	1
Antoni CEJZIK	2
Antonin SVOBODA	2
Antonio MARTINS GUIMARAES	4
Armas Rikkarti KINNUNEN	1
Arthur PORRITT	3
Arthur REINMANN	2
Arturs MOTTMILLERS	3
Arvo HAAVISTO	2
Asanosuke MATSUI	1
Beatrix LOUGHREN	2
Bertil FASTEN	2
Bino BINI	1
Bohumil STINNY	1
Brutus Kerr HAMILTON	3
Carlos ABARCA GONZALEZ	1
Cecil Charles MCMASTER	3
Charles JEWTRAW	3
Charles PADDOCK	4
Christopher HUDSON	1
Clarence PINKSTON	5
Constant BUCHER	3

The screenshot shows a Microsoft Power BI interface. On the left, a PivotTable Fields pane displays various dimensions and measures. Under 'Dimensions' (J), 'Athlete Year Birth' is selected. Under 'Measures' (I), 'FACT Count' is selected. The main area shows a PivotTable with 'Athlete Year Birth' in the Rows section and 'FACT Count' in the Values section. The data grid contains approximately 38 rows of athlete names and their birth years.

Athlete Year Birth	FACT Count
1990	1
Abdelali DARAA	1
Abdullah ALSHAITI	1
Ahmed ARAFAWI	1
Abigail IROZURO	1
Abigael JOCH	2
Achraf KHARROUBI	1
Adam NOWICKI	1
Adam ZAMPÀ	9
Adel MECHAAL	1
Adlan ABDURASHIDOV	1
Adrian BLOCKI	1
Adrian DZIOLKO	1
Adrienn TOTH	1
Agata PERENC	1
Agnieszka KOKKACA	1
Ahmed Moustafa AL-MERJABI	1
Ahmet NAL	3
Ahti TOIVANEN	2
Aistling Margaret COONEY	1
Ajmal FAISAL	1
Akari OGATA	1
Akihiko NAKAMURA	2
Alaaeldin ABOUELKASSEM	3
Alan GOGAEV	1
Albert TORRES BARCELÓ	1
Alberto Ezequiel MELIAN	2
Alberto Ignacio PALMETTA	1
Alberto MINO	1
Ali YILMAZ	1
AlexANDRE ANDRIENKO	2
Aleksandr MENKOV	1
Aleksandra JARMOLINSKA	2
Aleksander YERGIN	1

Câu 2: Liệt kê số lần tham gia thế vận hội Olympic Games mùa hè của các vận động viên trong năm 2016

Default

MDX

Dimension Hierarchy Operator Filter Expression Parameters

DIM Time Game Season Equal { Summer }

DIM Time Game Year Equal { 2016 }

<Select dimension>

Athlete Full Name	FACT Count
Aaron COOK	1
Aaron ROYLE	1
Ababel YESHANEH	1
Abadi HADIS	1
Abbas QALI	1
Abbey WEITZEL	2
Abbos RAKHMO...	1
Abdalaati IGUIDER	1
Abdel Aziz MEHE...	1
Abdelaziz MOHA...	1
Abdelhafid BENC...	1
Abdelkader CHADI	1
Abdelkebir OUA...	1
Abdelkhaled ELB...	1
Abdellatif Moha...	1
Abdelmajid EL HI...	1
Abderrahmane B...	1
Abderrahmane ...	1
Abdi Hakin ULAD	1
Abdi NAGEEYE	1
Abdoul Khadre M...	1
Abdoulrazak ISS...	1
Abdul OMAR	1
Abdulaziz ALSHA...	1
Abdulkadir ABDU...	1
Abdullah ALRAS...	1
Abdullah Hel BAKI	1
Abdullo TANGRIEV	1
Abdulrahman AL ...	1
Abeku Gyekye J...	1
Abhinav BINDRA	1
Abigail JOHNSTON	1
Abigel JOO	1
Abraham Naibei ...	1
Abraham NIYON...	1
Achraf KHARRO	1

MDXQuery1.mdx - TOP-0BB242\\$USER* SQLQuery1.sql - DE_0BB242\\$USER (6)

Cube: Olympic DWH

Metadata Functions

Search Model

Measure Group: <All>

Olympic DWH

- Measure
- Functions
- Search Model
- Measure Group: <All>
- Olympic DWH
- Measures
- Dimensions
- DIM Athlete
- DIM Country
- DIM Game
- DIM Host
- DIM Location
- DIM Time

```

Select [Measures].[FACT Count] on columns,
FILTER([DIM Athlete].[Athlete Full Name].Members, [Measures].[FACT Count] > 0
From [Olympic DWH]
Where CrossJoin([DIM Time].[Game Season].&[Summer],[DIM Time].
[Game Year].&[2016]));

```

214 %

	FACT Count
All	6366
Aaron COOK	1
Aaron ROYLE	1
Ababel YESHANEH	1
Abadi HADIS	1
Abbas GALI	1
Abbey WEITZEL	2
Abbey RAKHMONOV	1
Abdalati GUIDER	1
Abdel Aziz MEHELBA	1
Abdelaziz MOHAMED AHMED	1
Abdelhafid BENCHABLA	1
Abdelkader CHADI	1
Abdelkebir OUADDAR	1
Abdelkhalik ELBANNA	1
Abdelatif Mohamed Ahmed MOHAMED	1
Abdelmajid EL HISOUF	1
Abderahmane BENAMADI	1
Abderahmane MANSOURI	1
Abdi Hakim ULAD	1
Abdi NAGEYEY	1
Abdoul Khadre Mbaye NIANE	1
Abdourazak ISSOUFOU ALFAGA	1
Abdu OMAR	1
Abdulaziz ALSHATTI	1

PivotTable	Active Field	Group
A1	Game Season	
1 Game Season	Summer	
2 Game Year	2016	
4 Row Labels	FACT Count	
5 Aaron COOK	1	
6 Aaron ROYLE	1	
7 Ababel YESHANEH	1	
8 Abadi HADIS	1	
9 Abbas QALI	1	
10 Abbey WEITZEIL	2	
11 Abbos RAKHMONOV	1	
12 Abdalaati IGUIDER	1	
13 Abdel Aziz MEHELBA	1	
14 Abdelaziz MOHAMED AHMED	1	
15 Abdelhafid BENCHABLA	1	
16 Abdelkader CHADI	1	
17 Abdelkebir OUADDAR	1	
18 Abdelkhalek ELBANNA	1	
19 Abdellatif Mohamed Ahmed MOHAMED	1	
20 Abdelmajid EL HISSOUF	1	
21 Abderrahmane BENAMADI	1	
22 Abderrahmane MANSOURI	1	
23 Abdi Hakin ULAD	1	
24 Abdi NAGEEYE	1	
25 Abdoul Khadre Mbaye NIANE	1	
26 Abdoulrazak ISSOUFOU ALFAGA	1	
27 Abdul OMAR	1	
28 Abdulaziz ALSHATTI	1	
29 Abdulkadir ABDULLAYEV	1	
30 Abdullah ALRASHIDI	1	
31 Abdullah Hel BAKI	1	
32 Abdullo TANGRIEV	1	
33 Abdulrahman AL FAIHAN	1	
34 Abeku Gyekye JACKSON	1	
35 Abhinav BINDRA	1	
36 Abigail JOHNSTON	1	
37 Abigel JOO	1	

Câu 3: Tổng số vận động viên đã tham gia vào thế vận hội Olympic Games 2020 tại Nhật Bản

Screenshot of the SSAS Cube Structure tool interface.

The top navigation bar includes: Cube Structure, Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser.

The left sidebar shows the "Metadata" node expanded, listing various dimensions and measures:

- Olympic DWH
- Measures
- FACT
- Rank Position
- KPIs
- DIM Athlete
- Athlete Full Name
- Athlete Id
- Athlete Year Birth
- DIM Country
- Country 3 Letter Code
- Country Code
- Country Id
- Country Name
- DIM Game
- Athletes
- Discipline Title
- Event Title
- Game Id
- Medal Type
- Participant Type
- DIM Host
- Game Name
- Game Slug
- DIM Location
- Game Location
- Location Id
- DIM Time
- Game End Date
- Game Season
- Game Start Date
- Game Year
- Time Id

The main pane displays a query builder with the following filter settings:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
DIM Time	Game Year	Equal	{ 2020 }	
DIM Location	Game Location	Equal	{ Japan }	

Below the filters, a result grid shows:

Game Location	FACT Count
Japan	5484

The right side of the interface shows the "Tabs" pane with the following tabs listed:

- SSAS
- DIM Athlete.dim [Design]
- DIM Country.dim [Design]
- DIM Game.dim [Design]
- DIM Host.dim [Design]
- DIM Location.dim [Design]
- DIM Time.dim [Design]
- Olympic DWH.dim [Design]
- Olympic DWH.csv [Design]

Other tabs include: [Tool Windows], Error List, Object Browser, Output, Task List, and Task Runner Explorer.

Screenshot of the SQL Server Management Studio (SSMS) interface showing a query window.

The title bar reads: SQLQuery1.sql - DE...0BB242S\USER (61)

The query window contains the following T-SQL code:

```
Select [Measures].[FACT Count] on columns,
[DIM Location].[Game Location].&[Japan] on Rows
From [Olympic DWH]
Where ([DIM Time].[Game Year].&[2020]);
```

The results pane at the bottom shows the output of the query:

FACT Count
Japan 5484

A	B	C	D
1	YEAR	2020	
2			
3	Row Labels	FACT Count	
4	Japan	5484	
5	Grand Total	5484	
6			
7			
8			
9			
10			

Câu 4: Truy vấn 10 quốc gia có số lượng vận động viên tham gia nhiều nhất, giảm dần theo số lượng tham gia.

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists various database objects like 'sa', 'sa', 'Cubes', 'Dimensions', 'Mining Structures', 'Roles', 'Assemblies', and 'Management'. The MDX Query Editor in the center contains the following query:

```

Select [Measures].[FACT Count] on columns,
    HEAD(
        ORDER(
            [DIM Country].[Country Name].Members,
            [Measures].[FACT Count],
            desc),
        10
    ) on Rows
From [Olympic DWH];
  
```

The Results pane on the right displays the top 10 countries with their FACT Count:

Country	FACT Count
All	123310
United States of America	9037
France	5130
Italy	5026
Great Britain	4652
Germany	4581
Canada	4457
Japan	4391
Sweden	4313
Finland	3389

The screenshot shows a Microsoft Excel spreadsheet with a table titled "FACT Count". The table has two columns: "Row Labels" (Column A) and "FACT Count" (Column B). The data includes 11 countries and their respective counts, plus a Grand Total. The table is styled with alternating row colors.

	A	B	C	D	E
1	Row Labels	FACT Count			
2	United States of America	9037			
3	France	5130			
4	Italy	5026			
5	Great Britain	4652			
6	Germany	4581			
7	Canada	4457			
8	Japan	4391			
9	Sweden	4313			
10	Finland	3389			
11	Australia	3159			
12	Grand Total	48135			
13					
14					
15					
16					
17					

Câu 5: Liệt kê tổng số vận động viên tham gia môn 1500 mét xe lăn (nam)

The screenshot shows the Microsoft Analysis Services (SSAS) Dimension Usage tool interface. On the left, there is a tree view of the 'Olympic DWH' cube structure, including dimensions like 'DIM Game', 'DIM Athlete', 'DIM Country', etc., and measures like 'FACT Count'. In the center, a configuration pane displays a filter for the 'DIM Game' dimension. The filter is set to 'Event Title' with the operator 'Equal' and the value '{1,500 metres Wheelchair Men}'. Below this, a preview table shows a single row: 'Event Title' is '1,500 metres Wheelchair Men' and 'FACT Count' is '15112'.

Select [Measures].[FACT Count] on columns,
 [DIM Game].[Event Title].&[1,500 meters Wheelchair Men] on
 Rows

From [Olympic DW];

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface with the 'Results' tab selected. The results of the executed MDX query are displayed in a table:

	FACT Count
1,500 metres Wheelchair Men	15112

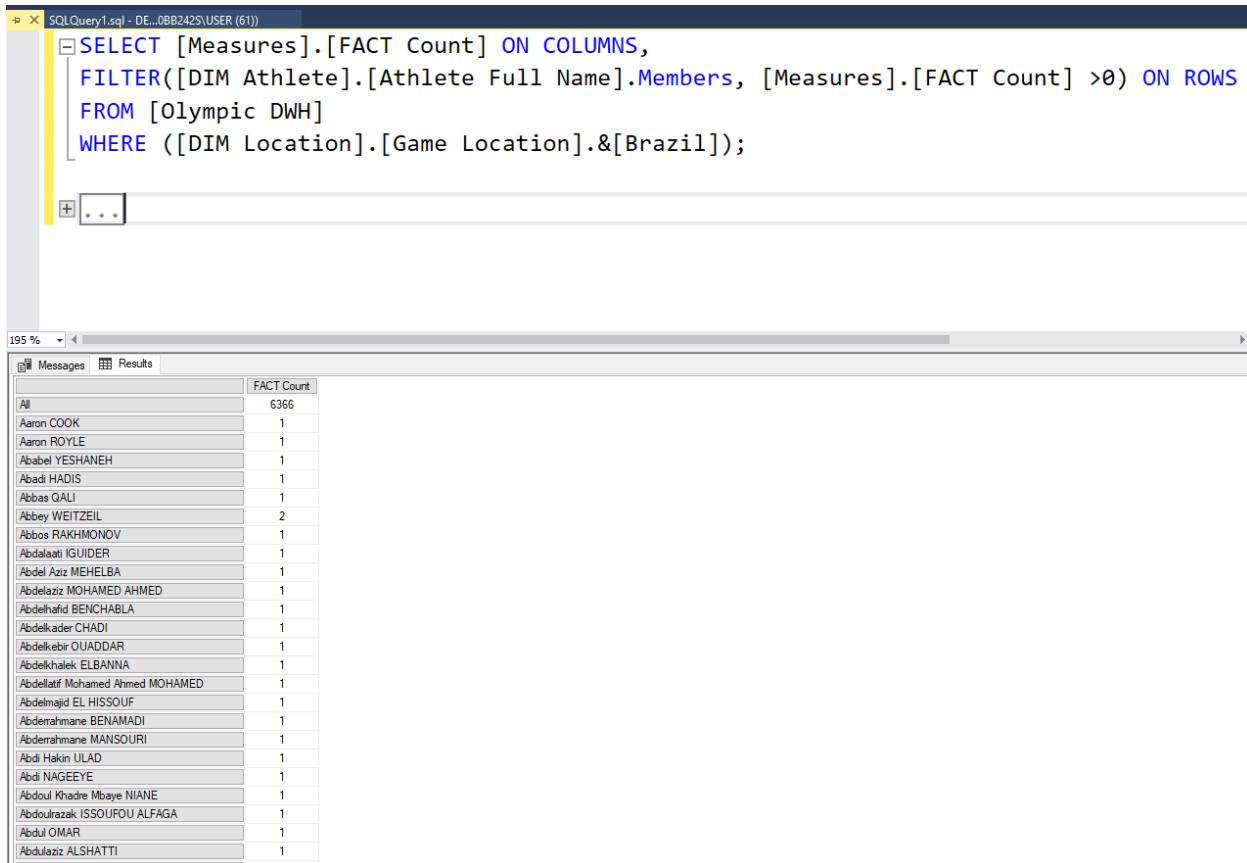
A	B	C	D
Row Labels	FACT Count		
1,500 metres Wheelchair Men	15112		
Grand Total	15112		

Câu 6: Liệt kê số lần tham gia thế vận hội Olympic Games các vận động viên tại Brazil.

The screenshot shows the Analysis Services Management Studio interface. At the top, there's a ribbon bar with various tabs like Cube Structure, Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser. The Language is set to Default. Below the ribbon is a toolbar with icons for Edit as Text, Import, MDX, and other common operations.

The main area is divided into several sections:

- Metadata:** A tree view showing the structure of the Olympic DWH cube, including Measures (FACT), KPIs, and various dimensions like DIM Athlete, DIM Country, DIM Game, DIM Host, DIM Location, and DIM Time.
- Filter:** A table where a filter is applied to the DIM Location dimension. The filter specifies a hierarchy (Game Location), operator (Equal), and filter expression ({Brazil}). There are also checkboxes for Parameters.
- Data Grid:** A large table titled "Athlete Full Name" showing the FACT Count for each athlete. The data includes rows for Aaron COOK, Aaron ROYLE, Ababel YESHANEH, Abadi HADIS, Abbas QALI, Abbey WEITZEL, Abbos RAKHMO..., Abdalaat IGUIDER, Abdel Aziz MEHE..., Abdelaziz MOHA..., Abdelhafid BENC..., Abdelkader CHADI, Abdelkebir OUA..., Abdelkhaled ELB..., Abdellatif Moha..., Abdelmajid EL HI..., Abderrahmane B..., Abderrahmane ... , Abdi Hakin ULAD, Abdi NAGEEYE, Abdoul Khadre M..., Abdoulrazaq ISS..., Abdul OMAR, Abdulaziz ALSHA..., Abdulkadir ABDU..., Abdullah ALRAS..., Abdullah Hel BAKI, Abdullo TANGRIEV, Abdulrahman AL ..., Abeku Gyekye J..., Abhinav BINDRA, Abigail JOHNSTON, Abigel JOO, Abraham Naibei ..., Abraham NIYON..., and Achraf KHARRO.



The screenshot shows a SQL Server Management Studio window with a query editor and a results grid. The query editor contains the following T-SQL code:

```
SELECT [Measures].[FACT Count] ON COLUMNS,
    FILTER([DIM Athlete].[Athlete Full Name].Members, [Measures].[FACT Count] >0) ON ROWS
FROM [Olympic DWH]
WHERE ([DIM Location].[Game Location].&[Brazil]);
```

The results grid displays a list of athletes and their fact counts. The data is as follows:

Athlete	FACT Count
All	6366
Aaron COOK	1
Aaron ROYLE	1
Ababel YESHANEH	1
Abadi HADIS	1
Abbas QALI	1
Abbey WEITZEL	2
Abbos RAKHMONOV	1
Abdalati IGUIDER	1
Abdel Aziz MEHELBA	1
Abdelaziz MOHAMED AHMED	1
Abdelhafid BENCHABLA	1
Abdelkader CHADI	1
Abdelkebir OUADDAR	1
Abdelkhalik ELBANNA	1
Abdellaif Mohamed Ahmed MOHAMED	1
Abdelmajid EL HISOUF	1
Abderahmane BENAMADI	1
Abderahmane MANSOURI	1
Abdi Hakim ULAD	1
Abdi NAGEEYE	1
Abdou Khadre Mbaye NIANE	1
Abdoulrazak ISSOUFOU ALFAGA	1
Abdul OMAR	1
Abdulaziz ALSHATTI	1

SELECT

[Measures].[FACT Count] ON COLUMNS,

FILTER ([DIM Athlete].[Athlete Full Name], [Measures].[FACT Count] >0) ON
ROWS

FROM [OLYMPIC DWH]

WHERE [DIM Location].[Game Location].&[Brazil];

A9 : X ✓ fx Abbey WEITZEIL

	A	B	C	D	E	F	G	H	I
1	Game Location	Brazil							
3	Row Labels	FACT Count							
4	Aaron COOK	1							
5	Aaron ROYLE	1							
6	Ababel YESHANEH	1							
7	Abadi HADIS	1							
8	Abbas QALI	1							
9	Abbey WEITZEIL	2							
10	Abbos RAKHMONOV	1							
11	Abdalaati IGUIDER	1							
12	Abdel Aziz MEHELBA	1							
13	Abdelaziz MOHAMED AHMED	1							
14	Abdelhafid BENCHABLA	1							
15	Abdelkader CHADI	1							
16	Abdelkebir OUADDAR	1							
17	Abdelkhalek ELBANNA	1							
18	Abdellatif Mohamed Ahmed MOHAMED	1							
19	Abdelmajid EL HISSOUF	1							
20	Abderrahmane BENAMADI	1							
21	Abderrahmane MANSOURI	1							
22	Abdi Hakim ULAD	1							
23	Abdi NAGEEYE	1							
24	Abdoul Khadre Mbaye NIANE	1							
25	Abdoulrazak ISSOUFOU ALFAGA	1							
26	Abdul OMAR	1							
27	Abdulaziz ALSHATTI	1							
28	Abdulkadir ABDULLAYEV	1							
29	Abdullah ALRASHIDI	1							
30	Abdullah Hel BAKI	1							
31	Abdullo TANGRIEV	1							
32	Abdulrahman AL FAIHAN	1							
33	Abeku Gyekye JACKSON	1							
34	Abhinav BINDRA	1							
35	Abigail JOHNSTON	1							
36	Abigel JOO	1							
37	Abraham Naibei CHEROBEN	1							
38	Abraham NIVONKIDU	1							

tmp4766

Ready Accessibility: Good to go

Câu 7: Số huy chương vàng đạt được theo từng bộ môn của các quốc gia.

MDX | Import... | MDX

Dimension **Hierarchy** **Operator** **Filter Expression** **Parameters**

FACT	Medal Type	Equal	{ GOLD }	<input type="checkbox"/>	<input type="checkbox"/>
<Select dimension>					
Country Name	Discipline Title	FACT Count			
Algeria	Athletics	4			
Algeria	Boxing	1			
Argentina	Athletics	2			
Argentina	Boxing	7			
Argentina	Judo	1			
Argentina	Taekwondo	1			
Armenia	Wrestling	2			
Australia	Archery	1			
Australia	Athletics	20			
Australia	Canoe Slalom	1			
Australia	Canoe Sprint	2			
Australia	Cycling BMX ...	1			
Australia	Cycling Road	2			
Australia	Cycling Track	6			
Australia	Diving	2			
Australia	Equestrian E...	1			
Australia	Freestyle Skiing	4			
Australia	Modern Pent...	1			
Australia	Rowing	4			
Australia	Sailing	3			
Australia	Shooting	5			
Australia	Short Track ...	1			
Australia	Skateboarding	1			
Australia	Snowboard	1			
Australia	Swimming	54			
Australia	Taekwondo	1			
Australia	Triathlon	1			
Australia	Weightlifting	1			
Austria	Alpine Skiing	38			
Austria	Athletics	1			
Austria	Canoe Sprint	1			
Austria	Cross Countr...	1			
Austria	Cycling Road	1			
Austria	Cycling Track	1			
Austria	Equestrian D...	1			
Austria	Endurance	1			

The screenshot shows the SQL Server Management Studio interface with multiple windows. On the left, the Object Explorer displays the 'Olympic DWH' cube structure, including dimensions like DIM Country, DIM Game, and FACT, and measures like FACT Count. The central pane contains an MDX query:

```

SELECT
    [Measures].[FACT Count] ON COLUMNS,
    FILTER([DIM Country].[Country Name].[Country Name].Members * [DIM Game].[Discipline Title].[Discipline Title].Members ,
    [Measures].[FACT Count]>0)ON ROWS
FROM [Olympic DWH]
WHERE [FACT].[Medal Type].&[GOLD]

```

The results pane shows the following data:

		FACT Count
Algeria	Athletics	4
Algeria	Boxing	1
Argentina	Athletics	2
Argentina	Boxing	7
Argentina	Judo	1
Argentina	Taekwondo	1
Armenia	Wrestling	2
Australia	Archery	1
Australia	Athletics	20
Australia	Canoe Slalom	1
Australia	Canoe Sprint	2
Australia	Cycling BMX Freestyle	1
Australia	Cycling Road	2
Australia	Cycling Track	6
Australia	Diving	2
Australia	Equestrian Eventing	1
Australia	Freestyle Skiing	4
Australia	Modern Pentathlon	1
Australia	Rowing	4
Australia	Sailing	3
Australia	Shooting	5

SELECT

```

[Measures].[FACT Count] ON COLUMNS,
FILTER ([DIM Country].[Country Name].[Country Name].Memebers * [Discipline Title].[Discipline Title].Memebers, [Measures].[FACT Count] >0) ON ROWS
FROM [OLYMPIC DWH]
WHERE [FACT].[Medal Type].&[GOLD]

```

	A	B	C	D	E	F	G
1	FACT Count	Column Labels	BRONZE	GOLD	SILVER	Grand Total	
2	Row Labels						
3	☒ Afghanistan						
4	Athletics		4			4	
5	Boxing		4			4	
6	Judo		2			2	
7	Shooting		1			1	
8	Taekwondo		3	2		5	
9	Wrestling		22			22	
10	☒ Albania						
11	Alpine Skiing		15			15	
12	Athletics		2			2	
13	Cycling Road		1			1	
14	Judo		3			3	
15	Shooting		17			17	
16	Swimming		11			11	
17	Weightlifting		19			19	
18	Wrestling		4			4	
19	☒ Algeria						
20	Alpine Skiing		10			10	
21	Athletics		53	2	4	3	62
22	Badminton		1				1
23	Boxing		68	5	1		74
24	Canoe Sprint		2				2
25	Cross Country Skiing		2				2
26	Cycling Road		9				9
27	Fencing		22				22
28	Gymnastics Artistic		2				2
29	Judo		39	1		1	41
30	Karate		1				1
31	Marathon Swimming		1				1
32	Rowing		8				8
33	Sailing		5				5
34	Shooting		3				3
35	Swimming		36				36
36	Table Tennis		5				5
37	Tennis		2				2
38	Weightlifting		19				19

Câu 8: Liệt kê danh sách vận động viên có thứ hạng là 2 trong giải Boxing nam ở thế vận hội Olympic Games Tokyo 2020.

The screenshot shows the Microsoft Analysis Services (SSAS) Cube Structure Editor. The top navigation bar includes options like Cube Structure, Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser. The Language dropdown is set to Default. The MDX dropdown is also set to MDX.

The left pane displays the cube's schema, including the Olympic DWH dimension, Measures, FACT, KPIs, and various dimensions like DIM Athlete, DIM Game, and DIM Time.

The main area contains an MDX query editor and a results grid. The MDX code is:

```

SELECT
    [Measures].[FACT Count] ON COLUMNS,
    NONEMPTY([DIM Athlete].[Athlete Full Name].Members) ON ROWS
FROM [Olympic DWH]
WHERE
(
    [FACT].[Rank Position].&[2],
    [DIM Game].[Discipline Title].&[Boxing],
    [DIM Time].[Game Year].&[2020]
)

```

The results grid shows the following data:

Athlete Full Name	Event Title	FACT Count
Beatriz FERREIRA	- 48kg (light-flyweight) men	1
Benjamin WHITT...	- 48kg (light-flyweight) men	1
Buse Naz CAKIR...	- 48kg (light-flyweight) men	1
Carlo PAALAM	- 48kg (light-flyweight) men	1
Duke RAGAN	- 48kg (light-flyweight) men	1
Hong GU	- 48kg (light-flyweight) men	1
Keyshawn DAVIS	- 48kg (light-flyweight) men	1
Muslim GADZHIM...	- 48kg (light-flyweight) men	1
Nesthy PETECIO	- 48kg (light-flyweight) men	1
Oleksandr KHYZ...	- 48kg (light-flyweight) men	1
Pat MCCORMACK	- 48kg (light-flyweight) men	1
Qian LI	- 48kg (light-flyweight) men	1
Richard TORREZ...	- 48kg (light-flyweight) men	1

- MDX:

SELECT

```

[Measures].[FACT Count] ON COLUMNS,
NONEMPTY([DIM Athlete].[Athlete Full Name].Members) ON ROWS
FROM [Olympic DWH]
WHERE
(
    [FACT].[Rank Position].&[2],
    [DIM Game].[Discipline Title].&[Boxing],
    [DIM Time].[Game Year].&[2020]
)

```

125\USER* - Microsoft SQL Server Management Studio

elp

MDXQuery4.mdx - TOP-0BB242S\USER* SQLQuery24.sql - D...0BB242S\USER (68) SQLQuery23.sql - D...0BB242S\USER (61) SQLQuery22.sql - D...0BB242S\USER (62)

Quick Launch (Ctrl+Q)

Cube: Olympic DWH

Measure Group: <All>

```

SELECT
    [Measures].[FACT Count] ON COLUMNS,
    NONEMPTY([DIM Athlete].[Athlete Full Name].Members) ON ROWS
FROM [Olympic DWH]
WHERE
(
    [FACT].[Rank Position].&[2],
    [DIM Game].[Discipline Title].&[Boxing],
    [DIM Time].[Game Year].&[2020]
)

```

133 % | <

	FACT Count
All	13
Beatriz FERREIRA	1
Benjamin WHITTAKER	1
Buse Naz CAKIROGLU	1
Carlo PAALEM	1
Duke RAGAN	1
Hong GU	1
Keyshawn DAVIS	1
Muslim GADZHIMAGOMEDOV	1
Neathy PETECIO	1
Oleksandr KHYZHNIAK	1
Pat MCCORMACK	1
Qian LI	1
Richard TORREZ JR	1

Câu 9: Những vận động viên Việt Nam đã từng tham gia thi đấu tại các kỳ Olympic

The screenshot shows the SSAS Management Studio interface. The left pane is the 'Metadata' browser, displaying the structure of the 'Olympic DWH' cube. It includes dimensions like DIM Athlete, DIM Country, and FACT tables, along with their respective measures and KPIs. The right pane is the 'Browser' window, which displays a query result grid. The grid has four columns: Athlete Full Name, Game Location, Discipline Title, and FACT Count. A filter is applied to the DIM Country dimension, specifically for the 'Country Name' hierarchy, with the operator set to 'Equal' and the value set to '{ Vietnam }'. The results show various athletes from Vietnam, such as Anh Tuan HOANG, Chau HUYNH, and Duc Yunn NGUYEN, along with their game locations, disciplines, and fact counts.

- MDX:

SELECT

```
[Measures].[FACT Count] ON COLUMNS,
NONEMPTY([DIM Athlete].[Athlete Full Name].Members) ON ROWS
FROM [Olympic DWH]
```

WHERE

```
( [DIM Country].[Country Name].&[Vietnam]
)
```

ibe:
lympic DWH

Metadata Functions
Search Model
Measure Group:
All >

US Virgin Islands
Uzbekistan
Vanuatu
Venezuela
Vietnam
Virgin Islands, British
Virgin Islands, US
Yemen
Yemen Arab Republ
Yugoslavia
Zambia
Zimbabwe
Unknown

DIM Game

- Discipline Title
 - Members
 - Discipline Title
- Event Title
 - Members
 - Event Title
 - Member Properties
 - 48 kg - extralightweight
 - 48 kg women
 - 48kg (light-flyweight)
 - 49 kg women
 - 52kg (flyweight) men
 - 58 kg men
 - 60 kg - extralightweight
 - 60 kg men
 - 63kg (lightweight) r
 - +100kg (heavyweight)
 - +67 kg women
 - +78kg (heavyweight)
 - +80 kg men
 - +91kg (super heavy)
 - +95kg (heavyweight)
 - +105kg men
 - +75kg women
 - 1 mile 1609m men
 - 1,500 metres Wheel
 - 1,500 metres, Handi
 - 1/2 mile 805m men
 - 1/3 mile 536m men
 - 1/4 mile 402m men

133 %

	FACT Count
All	94
Anh Tuan HOANG	1
Chau HUYNH	1
Dinh Chi NGUYEN	1
Duc Yunh NGUYEN	1
Hieu Ngan TRAN	1
Hoai Nam QUACH	2
Hoang Phi Vu NGUYEN	1
Huu Kinh NGO	1
Huu Tinh PHI	1
Huu Viet NGUYEN	2
Huy Khang PHAN	1
Huynh Chau LE	1
Ken Quoc DOAN	2
Kim Thieng NGUYEN	1
Kim Tuan THACH	1
Ie Quoc Toan TRAN	2
Manh Tuong NGUYEN	3
Minh Hien LE	1
Ngoc Anh NGUYEN	1
Ngoc Tu VAN	2
Nguyen Kim HUONG	1
Nguyen THI THANH PHUC	1

Query executed successfully.

Athlete Full Name	FACT Count
Anh Tuan HOANG	1
Chau HUYNH	1
Dinh Chi NGUYEN	1
Duc Yrnh NGUYEN	1
Hieu Ngan TRAN	1
Hoai Nam QUACH	2
Hoang Phi Vu NGUYEN	1
Huu Kinh NGO	1
Huu Tinh PHI	1
Huu Viet NGUYEN	2
Huy Khang PHAN	1
Huyhn Chau LE	1
Kien Quoc DOAN	2
Kim Thieng NGUYEN	1
Kim Tuan THACH	1
Ie Quoc Toan TRAN	2
Manh Tuong NGUYEN	3
Minh Hien LE	1
Ngoc Anh NGUYEN	1
Ngoc Tu VAN	2
Nguyen Kim HUONG	1
Nguyen THI THANH PHUC	1
Nguyen Van THUYET	1
Nguyet DO THI ANH	1
Phoung Trinh CAO NGOC	1
Phuoc HOANG	1
Quoc Cuong NGUYEN	3
Quoc Cuong TRAN	2
Quoc Huan NGUYEN	1
Quyen NGUYEN	1
Tan Tai HOANG	1
Thanh An VU	1
Thanh Ngung NGUYEN	1
Thi Anh DO	1
Thi Anh Vien NGUYEN	5
Thi Duyen HOANG	1

Câu 10: Quốc gia dành được nhiều huy vàng nhất từ năm 1990 đến năm 2000. (Loại truy vấn: Slice and dice)

- MDX:

SELECT

[Measures].[FACT Count] ON COLUMNS,

FILTER(

[DIM Country].[Country Name].[Country Name].Members * [DIM Game].[Discipline Title].[Discipline Title].Members,

[Measures].[FACT Count] > 0

) ON ROWS

FROM [Olympic DWH]

WHERE

(

```

[FACT].[Medal Type].&[GOLD],  

[DIM Time].[Game Year].[1990] : [DIM Time].[Game Year].[2000]  

)

```

MDXQuery3.mdx -...TOP-0BB2425\USER)

Cube: Olympic DWH

Metadata Functions

Search Model

Measure Group: <All>

1928
1932
1936
1948
1952
1956
1960
1964
1968
1972
1976
1980
1984
1988
1992
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Unknown

Time Id
FACT
FACT Id
Medal Type
Members
• Medal Type
Member Properties
• BRONZE
• GOLD
• SILVER
Rank Equal
Rank Position

MDXQuery2.mdx -...TOP-0BB2425\USER)* SQLQuery18.sql - D...0BB2425\USER (69) SQLQuery17.sql - D...0BB2425\USER (67) SQLQuery16.sql - D...0BB2425\USER (72)*

```

SELECT [Measures].[FACT Count] ON COLUMNS,
FILTER(
    [DIM Country].[Country Name].[Country Name].Members * [DIM Game].[Discipline Title].[Discipline Title].Members,
    [Measures].[FACT Count] > 0
) ON ROWS
FROM [Olympic DWH]
WHERE
(
    [FACT].[Medal Type].&[GOLD],
    [DIM Time].[Game Year].[1990] : [DIM Time].[Game Year].[2000]
)

```

133 % 4 Messages Results

		FACT Count
Algeria	Athletics	3
Algeria	Boxing	1
Argentina	Athletics	2
Argentina	Boxing	7
Amenia	Wrestling	1
Australia	Archery	1
Australia	Athletics	17
Australia	Canoe Sprint	1
Australia	Cycling Road	1
Australia	Cycling Track	2
Australia	Equestrian Eventing	1
Australia	Rowing	3
Australia	Shooting	3
Australia	Swimming	38
Australia	Taekwondo	1
Australia	Weightlifting	1
Austria	Alpine Skiing	23
Austria	Athletics	1
Austria	Canoe Sprint	1
Austria	Cycling Track	1
Austria	Equestrian Dressage	1
Austria	Fencing	1
Austria	Figure skating	5
Austria	Gymnastics Artistic	1

Query executed successfully.

DESKTOP-0BB2425 | DESKTOP-0BB2425\USER

Câu 11: Những vận động viên Việt Nam đã từng dành được huân chương tại các kỳ Olympic

Dimension	Hierarchy	Operator	Filter Expression	Parameters
FACT	Medal Type	Equal	{ GOLD, BRONZE, SILVER }	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
DIM Country	Country Name	Equal	{ Vietnam }	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				
Athlete Full Name	FACT Count			
Anh Tuan HOANG	1			
Hieu Ngan TRAN	1			
Xuan Vinh HOANG	2			

Country Name	Vietnam	FACT Count
Medal Type	(Multiple Items)	
Row Labels		
Anh Tuan HOANG	1	
Hieu Ngan TRAN	1	
Xuan Vinh HOANG	2	
Grand Total	4	

Câu 12: Thống kê số các vận động viên sinh năm 2000 dành được huy chương vàng

Dimension	Hierarchy	Operator	Filter Expression	Parameters
DIM Athlete	Athlete Year Birth	Equal	{ 2000 }	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
FACT	Medal Type	Equal	{ GOLD }	
<Select dimension>				

Athlete Full Name	FACT Count
Akira SONE	1
Ariarne Titmus	2
Birk RUUD	1
Chloe KIM	2
Gable Dan STEV...	1
Hannah NEISE	1
Ivet GORANOVA	1
Jade CAREY	1
Jakob INGEBRIG...	1
Kristof Kristof Milak	1
Lasha BEKAURI	1
Margaret MACNEIL	1
Mathilde GREMAUD	1
Nina DERWAEL	1
Penny OLEKSIAK	1
Qian YANG	1
Selemon BAREGA	1
Sena IRIE	1
Tom DEAN	1
Uta ABE	1
Vito DELL'AQUILA	1
Walter WALLBERG	1
Wenwen LI	1

Medal Type	GOLD	<input type="button" value="▼"/>
Athlete Year Birth	2000	<input type="button" value="▼"/>
Row Labels		FACT Count
Akira SONE	1	
Ariarne Titmus	2	
Birk RUUD	1	
Chloe KIM	2	
Gable Dan STEVESON	1	
Hannah NEISE	1	
Ivet GORANOVA	1	
Jade CAREY	1	
Jakob INGEBRIGTSEN	1	
Kristof Kristof Milak	1	
Lasha BEKAURI	1	
Margaret MACNEIL	1	
Mathilde GREMAUD	1	
Nina DERWAEL	1	
Penny OLEKSIAK	1	
Qian YANG	1	
Selemon BAREGA	1	
Sena IRIE	1	
Tom DEAN	1	
Uta ABE	1	
Vito DELL'AQUILA	1	
Walter WALLBERG	1	
Wenwen LI	1	
Grand Total	25	

Câu 13: Liệt kê số lần tham gia thế vận hội Olympic Games của các vận động viên môn boxing theo thể thức cá nhân

Dimension	Hierarchy	Operator	Filter Expression	Parameters
DIM Game	Participant Type	Equal	{ Athlete }	<input type="checkbox"/> <input type="checkbox"/>
DIM Game	Discipline Title	Equal	{ Boxing }	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>
Athlete Full Name		Participant Type	Event Title	FACT Count
Aaron POPOOLA		Athlete	- 48kg (lig...	1
Aaron PRINCE		Athlete	- 48kg (lig...	1
Abbos ATOEV		Athlete	- 48kg (lig...	2
Abdel Aziz HAMMI		Athlete	- 48kg (lig...	1
Abdel Hadi KHALLAF ALLAH		Athlete	- 48kg (lig...	2
Abdel Hamid Fouad GAD		Athlete	- 48kg (lig...	1
Abdel JEBABI		Athlete	- 48kg (lig...	1
Abdel Nabi EL SAYED MAHRAN		Athlete	- 48kg (lig...	1
Abdelalii DARAA		Athlete	- 48kg (lig...	1
Abdelamid BOUTEFNOUCHET		Athlete	- 48kg (lig...	1
Abdelaziz BOULEHIA		Athlete	- 48kg (lig...	1
Abdelaziz TOUILBINI		Athlete	- 48kg (lig...	1
Abdelhafid BENCHABLA		Athlete	- 48kg (lig...	4
Abdelhak AATAKNI		Athlete	- 48kg (lig...	1
Abdelhalim OURRADI		Athlete	- 48kg (lig...	1
Abdelhani KENZI		Athlete	- 48kg (lig...	2
Abdelhaq NADIR		Athlete	- 48kg (lig...	1
Abdelillah NHAILA		Athlete	- 48kg (lig...	1
Abdelkader CHADI		Athlete	- 48kg (lig...	3
Abdelkader WAHABI		Athlete	- 48kg (lig...	1
Abdellah TAOUANE		Athlete	- 48kg (lig...	2
Abdellatif FATHI		Athlete	- 48kg (lig...	1
Abdelmalek RAHOU		Athlete	- 48kg (lig...	1
Abdelrahman SALAH ORABI ABDELGAWWAD		Athlete	- 48kg (lig...	1

Participant Type	Athlete	
Discipline Title	Boxing	
Row Labels		FACT Count
Aaron POPOOLA		1
Aaron PRINCE		1
Abbos ATOEV		2
Abdel Aziz HAMMI		1
Abdel Hadi KHALAF ALLAH		2
Abdel Hamid Fouad GAD		1
Abdel JEBALI		1
Abdel Nabi EL SAYED MAHRAN		1
Abdelali DARAA		1
Abdelamid BOUTEFNOUCHET		1
Abdelaziz BOULEHIA		1
Abdelaziz TOUILBINI		1
Abdelhafid BENCHABLA		4
Abdelhak AATAKNI		1
Abdelhalim OURRADI		1
Abdelhani KENZI		2
Abdelhaq NADIR		1
Abdelillah NHAILA		1
Abdelkader CHADI		3
Abdelkader WAHABI		1
Abdellah TAOUANE		2
Abdellatif FATHIHI		1
Abdelmalek RAHOU		1
Abdelrahman SALAH ORABI ABDELGAWWAD		1
Abderrahim NAJM		1
Abderrazak HOUYA		1

Câu 14: Thống kê số lượng tham gia Olympic với từng nơi tổ chức của các vận động viên có quốc tịch Argentina

Dimension	Hierarchy	Operator	Filter Expression	Parameters
DIM Country	Country Name	Equal	{ Argentina }	<input type="checkbox"/> <input checked="" type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input checked="" type="checkbox"/>
Game Location	FACT Count			
Australia	65			
Australia, Sw...	22			
Austria	25			
Brazil	67			
Canada	102			
China	64			
Federal Repu...	36			
Finland	67			
France	75			
Germany	22			
Great Britain	85			
Greece	63			
Italy	60			
Japan	105			
Mexico	43			
Netherlands	7			
Norway	32			
Republic of K...	46			
Russian Fede...	21			
Spain	48			
Switzerland	14			
United States	211			
Yugoslavia	32			

Country Name	Argentina			
Row Labels	FACT Count			
Australia	65			
Australia, Sweden	22			
Austria	25			
Brazil	67			
Canada	102			
China	64			
Federal Republic of Germany	36			
Finland	67			
France	75			
Germany	22			
Great Britain	85			
Greece	63			
Italy	60			
Japan	105			
Mexico	43			
Netherlands	7			
Norway	32			
Republic of Korea	46			
Russian Federation	21			
Spain	48			
Switzerland	14			
United States	211			
Yugoslavia	32			
Grand Total	1312			

Câu 15: Liệt kê số lượng tham gia thi đấu của vận động viên Việt Nam với các bộ môn

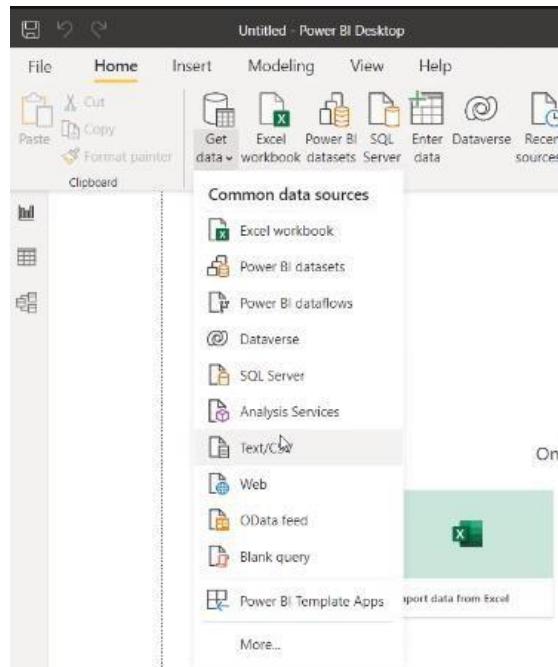
Dimension	Hierarchy	Operator	Filter Expression	Parameters
DIM Country	Country Name	Equal	{ Vietnam }	<input type="checkbox"/> <input checked="" type="checkbox"/>
<Select dimension>				
Discipline Title	FACT Count			
Archery	2			
Athletics	4			
Badminton	6			
Boxing	2			
Cycling Road	1			
Fencing	5			
Gymnastics ...	6			
Judo	4			
Shooting	23			
Swimming	12			
Table Tennis	2			
Taekwondo	9			
Weightlifting	10			
Wrestling	8			

Country Name	Vietnam	
Row Labels	FACT Count	
Archery	2	
Athletics	4	
Badminton	6	
Boxing	2	
Cycling Road	1	
Fencing	5	
Gymnastics Artistic	6	
Judo	4	
Shooting	23	
Swimming	12	
Table Tennis	2	
Taekwondo	9	
Weightlifting	10	
Wrestling	8	
Grand Total	94	

CHƯƠNG 4. QUÁ TRÌNH TẠO REPORT BẰNG CÔNG CỤ POWER

5.1 NHẬP DỮ LIỆU NGUỒN

Bước 1: Mở Power BI Desktop, tại mục **Home > Get Data > AnalysisService** để import dữ liệu đã xử lý qua quá trình SSIS và SSAS



Thực hiện import data source vào Power BI

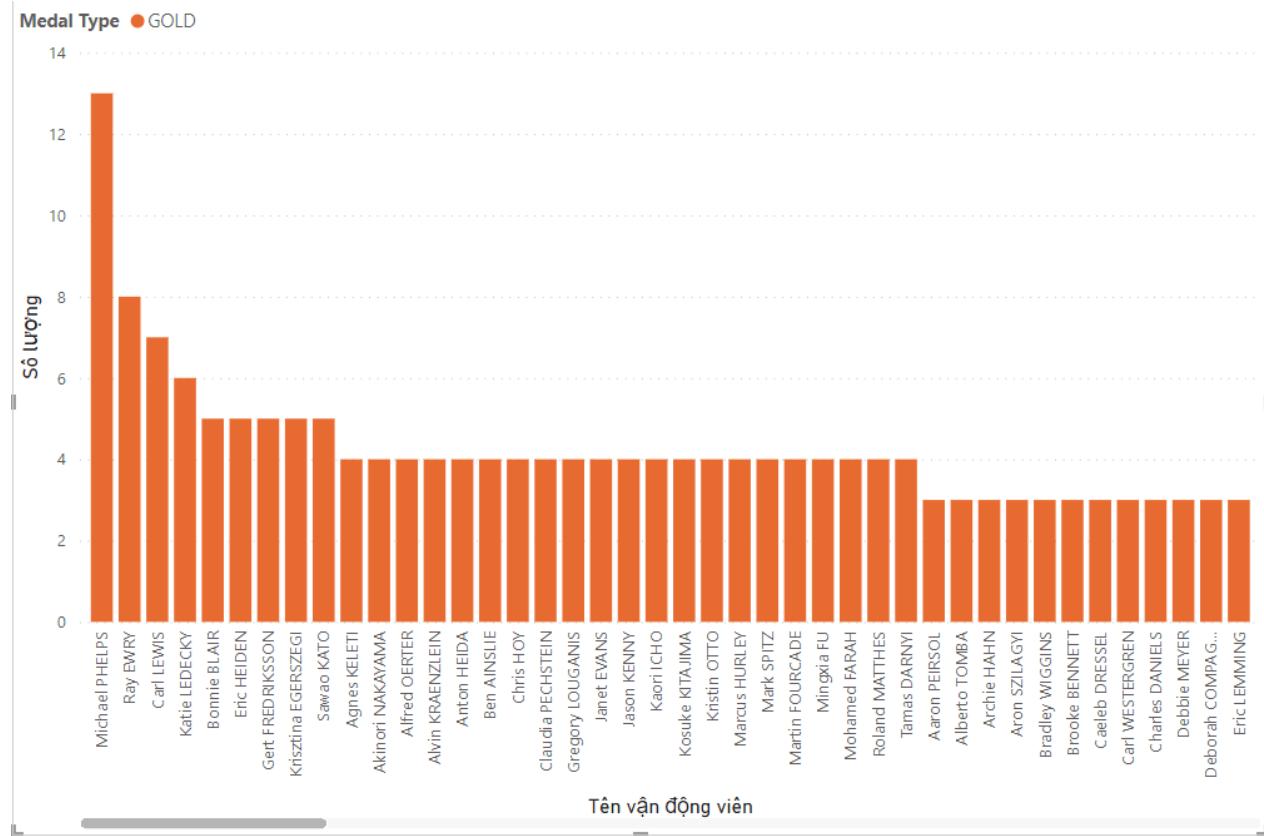
Bước 2: Tại màn hình vừa hiện lên, dán tên server của SQL Server máy mình > Nhập tên Database mình muốn sử dụng



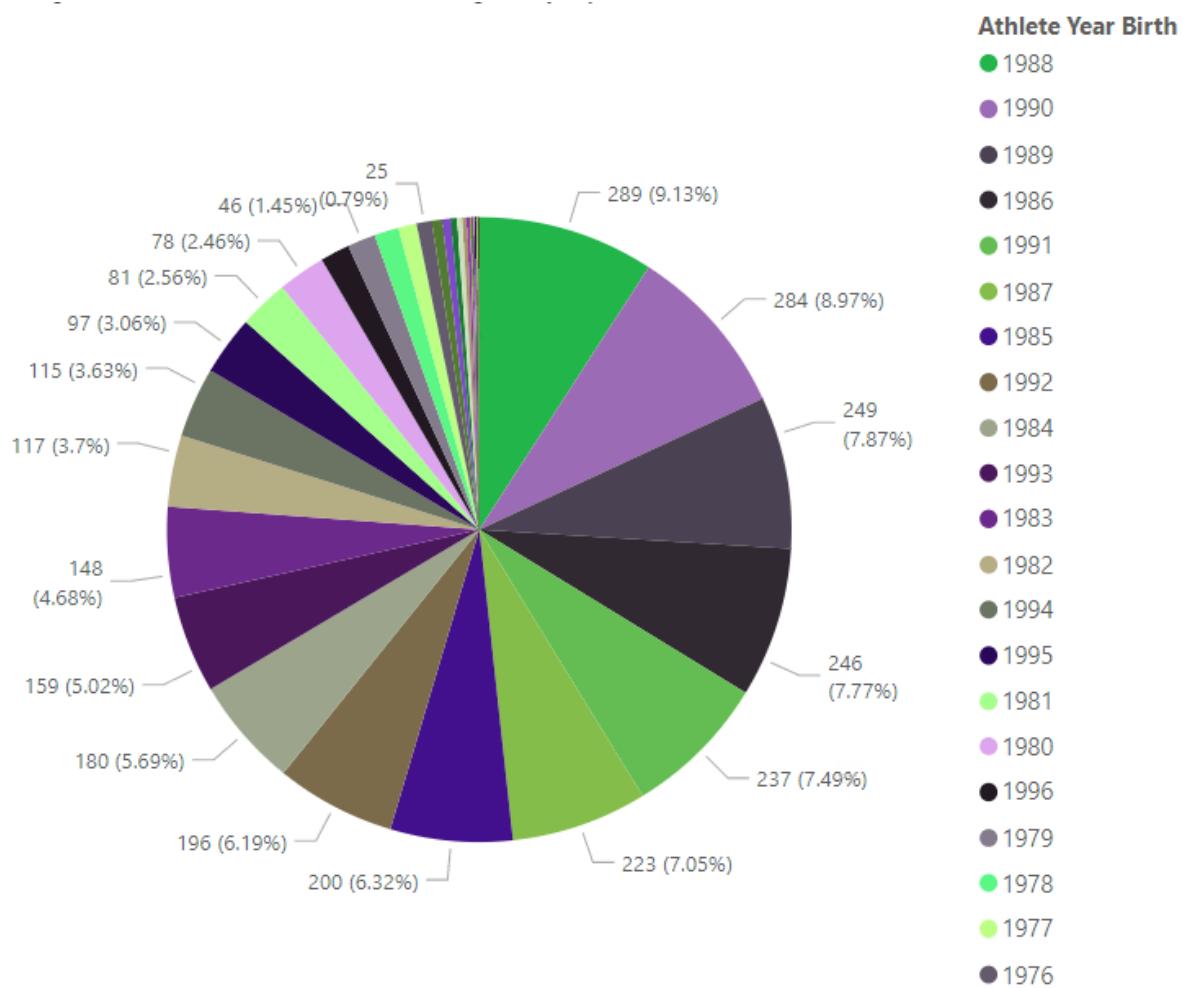
Điền thông tin đến cơ sở dữ liệu nguồn

THỰC HIỆN CÁC CÂU TRUY VẤN

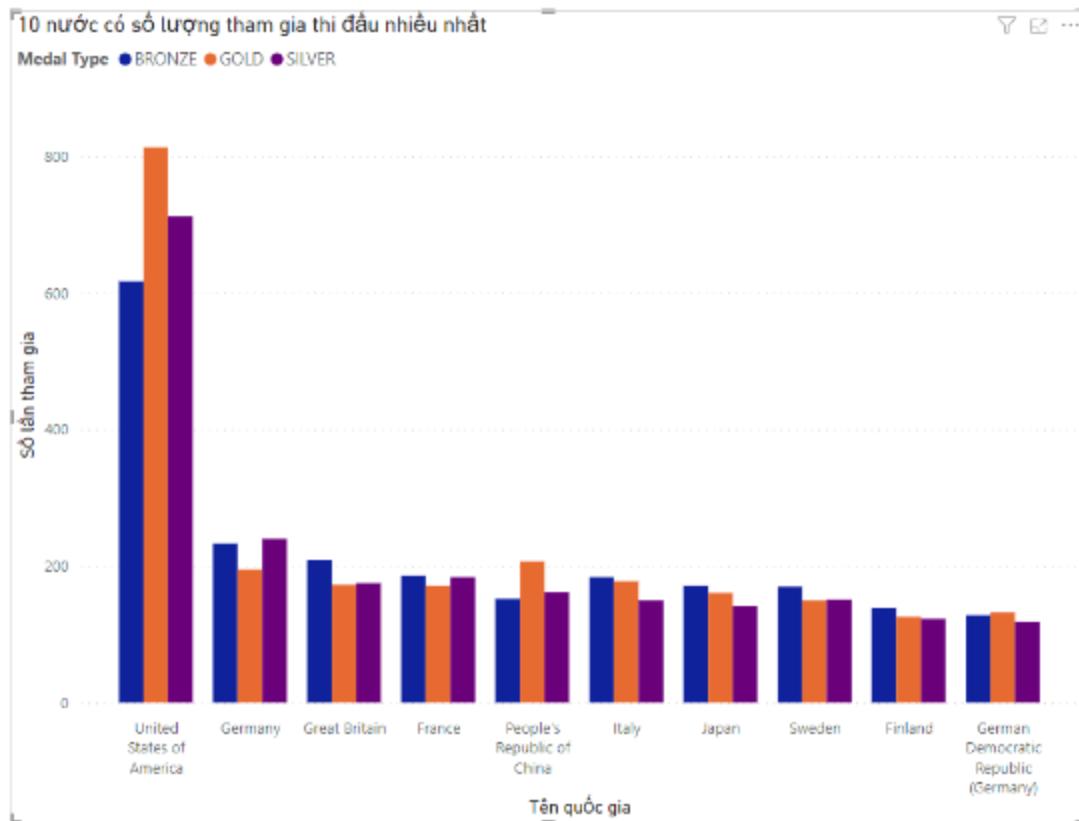
Câu 1: Vận động viên đạt huy chương vàng:



Câu 2: Thống kê độ tuổi của các vận động viên tham gia Olympic 2014:



Câu 3: 10 nước có số lượng tham gia thi đấu nhiều nhất.



Câu 4: Sử dụng Matrix: Xem số lần tham gia Olympic và huy chương đạt được của các vận động viên.

Athlete Full Name	FACT Count
Heikki SAVOLAINEN	34
BRONZE	29
GOLD	3
SILVER	1
Alfred JOCHIM	27
Takashi ONO	24
Joseph STOFFEL	23
Georges MIEZ	22
Johann (Hans) SAUTER	22
Michel MATHIOT	22
Frank Conway TURNER	21
István PELLE	21
Jean KUGELER	21
Josip PRIMOZIC	21
Kjetil Andre AAMODT	21
Lajos TOTH	21
Leon STUKELJ	21
Rafael LECUONA ASENCIO	21
Raymond DOT	21
Seiko HASHIMOTO	21
Zdenek RUZICKA	21
Emese HUNYADY	20
Ole Einar BJØRNDALEN	20
Raymond DEREVÉC	20
Total	123310

Filters Visualizations Data

Filters

- Search:
- Filters on this visual:
 - Athlete Full Name: Is (All)
 - FACT Count: Is (All)
 - Medal Type: Is (All)
 - Filter type: Basic filtering
 - Search:
 - Select all
 - BRONZE
 - GOLD
 - SILVER
 - Require single selection
 - Add data fields here
- Filters on this page:
- Filters on all pages:

Visualizations

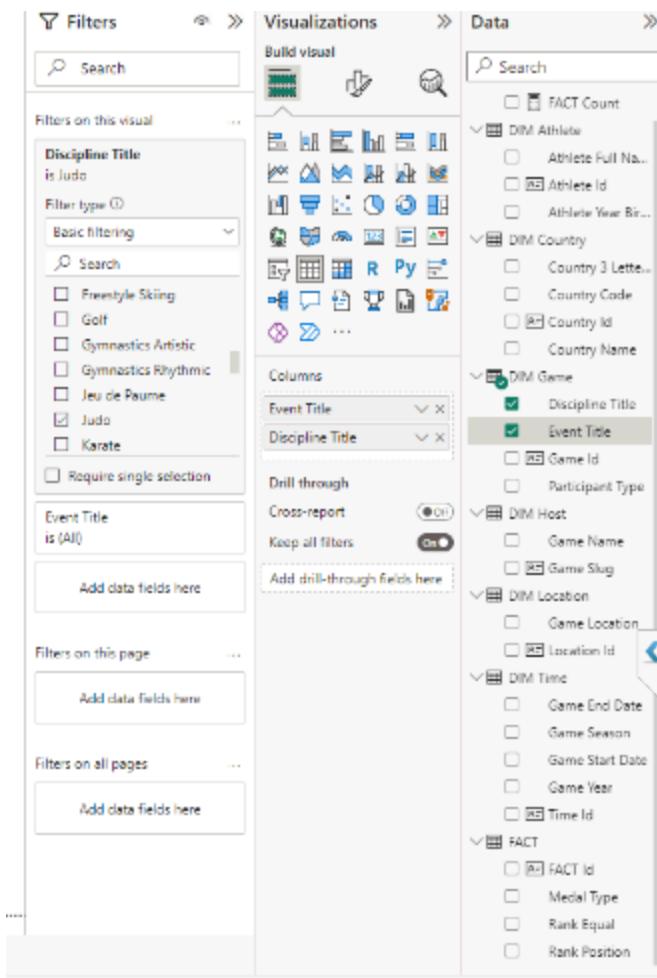
Build visual:

Data

- FACT (Measure)
 - FACT Count
- DIM Athlete
 - Athlete Full Name
 - Athlete Id
 - Athlete Year Birth
- DIM Country
 - Country 3 Letter ...
 - Country Code
 - Country Id
 - Country Name
- DIM Game
 - Discipline Title
 - Event Title
 - Game Id
 - Participant Type
- DIM Host
- DIM Location
 - Game Location
 - Location Id
- DIM Time
 - Game End Date
 - Game Season
 - Game Start Date
 - Game Year
 - Time Id
- FACT
 - FACT Id
 - Medal Type
 - Rank Equal
 - Rank Position

Câu 5: Liệt kê các hạng cân của môn Judo

Event Title	Discipline Title
- 40 kg - extralightweight women	Judo
- 48 kg women	Judo
- 60 kg -extralightweight men	Judo
- 60 kg men	Judo
- 63kg (lightweight) men	Judo
+ 100kg (heavyweight) men	Judo
+ 78kg (heavyweight) women	Judo
+ 95kg (heavyweight) men	Judo
100kg heavyweight men	Judo
48 - 52kg (half-lightweight) women	Judo
48 52kg halflightweight women	Judo
48 kg women	Judo
48kg extralightweight women	Judo
52 56kg lightweight women	Judo
52 57kg lightweight women	Judo
56 61kg halfmiddleweight women	Judo
57 - 63kg (half-middleweight) women	Judo
57 63kg halfmiddleweight women	Judo
60 - 65kg (half-lightweight) men	Judo
60 65kg halflightweight men	Judo
60 - 66kg (half-lightweight) men	Judo
60 66kg halflightweight men	Judo
60 kg men	Judo
61 66kg middleweight women	Judo
63 70kg halfmiddleweight men	Judo
63 70kg middleweight women	Judo
63kg lightweight men	Judo
65 - 71kg (lightweight) men	Judo
66 72kg halfheavyweight women	Judo
66 73kg lightweight men	Judo
68 80kg middleweight men	Judo
68kg lightweight men	Judo
70 - 78kg (half-heavyweight) women	Judo
70-79kg halfheavyweight women	Judo



CHƯƠNG 5 KHAI THÁC DỮ LIỆU – DATA MINING

5.1 HƯỚNG PHÂN TÍCH

- Chủ đề: Phân loại nhóm quốc gia dựa theo thành tích đạt được qua các kỳ Olympic.
- Thuật toán sử dụng: Decision Tree và Random Forest.
- Mô tả: Thực hiện đếm số lượng huy chương đạt được theo từng quốc gia, chia thành 3 nhóm chính: LOW, MEDIUM, HIGH.

- Tiết xử lí dữ liệu: chỉ sử dụng hai feature chính: Medal_type và Country_id. Loại bỏ những giá trị NAN ở cột medal_type, giá trị có nghĩa là tham gia thi đấu nhưng không đạt giải, do đó ta không cần đến.

5.2 THỰC HIỆN

5.2.1 Decision Tree.

- Import các thư viện cần thiết:

```

• import pandas as pd
• from sklearn.tree import DecisionTreeClassifier
• from sklearn import datasets
• from sklearn.model_selection import train_test_split
• from sklearn import metrics
• from IPython.display import Image
• import matplotlib.pyplot as plt
• from sklearn import tree

```

- Load dữ liệu:

```
• df = pd.read_csv('Fact_Olympic.csv')
```

- Tiết xử lí

```

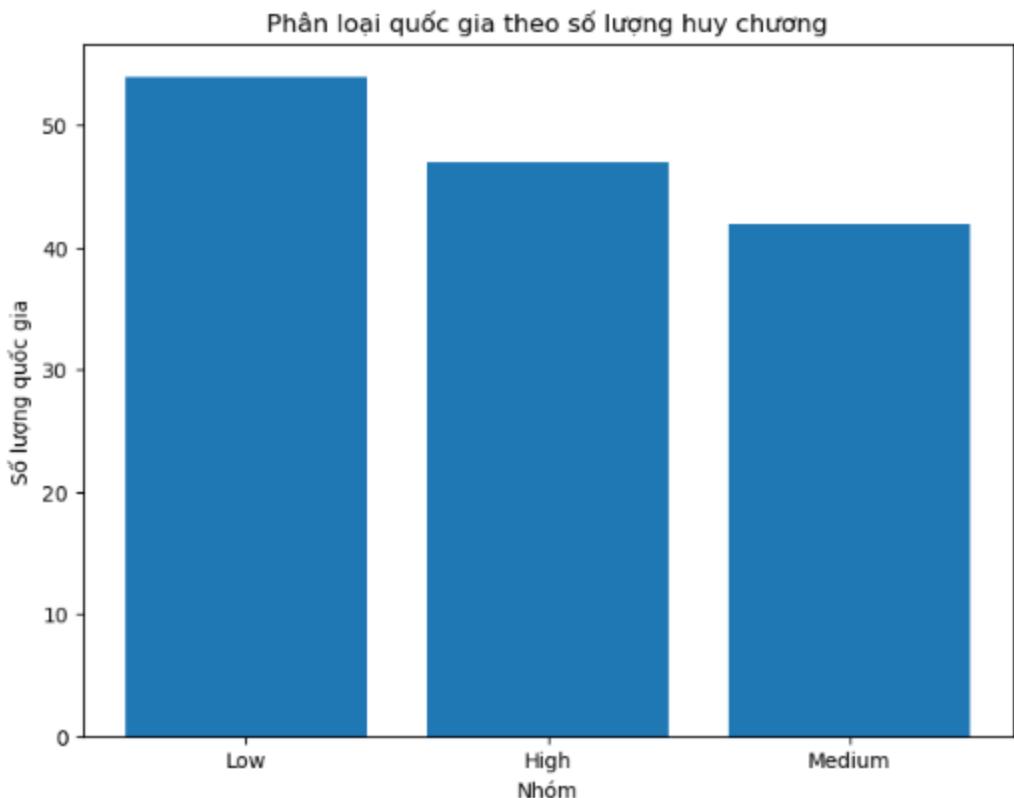
• import pandas as pd
•
• # Loại bỏ các dòng có giá trị medal_type bằng NaN
• df_clean = df.dropna(subset=['medal_type'])
•
• # Tạo dataframe mới chỉ chứa hai cột: country_id và medal_type
• df_medals = df_clean[['country_id', 'medal_type']]
•
• # Gom nhóm theo country_id và medal_type, đếm số lượng mỗi loại huy chương
• df_grouped = df_medals.groupby(['country_id',
• 'medal_type']).size().unstack(fill_value=0)
•
• # Tạo dataframe mới với hai cột: country_id và medal_count
• df_medal_count = pd.DataFrame({'country_id': df_grouped.index,
• 'medal_count': df_grouped.sum(axis=1)})
• labels = ['Low', 'Medium', 'High']
• df_medal_count['label'] = pd.qcut(df_medal_count['medal_count'],
• len(labels), labels=labels)
• # Chia dữ liệu thành input features (X) và target variable (y)
• X = df_medal_count[['medal_count']]
• y = df_medal_count['label']
•

```

```
• # Chia dữ liệu thành tập train và tập test
• X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
•
• # Tạo một đối tượng DecisionTreeClassifier
• model = DecisionTreeClassifier()
•
• # Huấn luyện mô hình trên tập train
• model.fit(X_train, y_train)
• # Đánh giá mô hình trên tập test
• accuracy = model.score(X_test, y_test)
• print("Độ chính xác của mô hình: ", accuracy)
```

→ Độ chính xác của mô hình: 1.0

```
• # Đếm số quốc gia trong mỗi nhóm
• group_counts = df_medal_count['label'].value_counts()
•
• # Tạo biểu đồ cột
• plt.figure(figsize=(8, 6))
• plt.bar(group_counts.index, group_counts.values)
•
• # Đặt tiêu đề và nhãn cho biểu đồ
• plt.title('Phân loại quốc gia theo số lượng huy chương')
• plt.xlabel('Nhóm')
• plt.ylabel('Số lượng quốc gia')
•
• # Hiển thị biểu đồ
• plt.show()
```



```

• # Lọc dữ liệu theo nhóm 'Low'
• df_low = df_medal_count[df_medal_count['label'] == 'Low']
• print("Quốc gia trong nhóm 'Low':")
• print(df_low['country_id'])

•
• # Lọc dữ liệu theo nhóm 'Medium'
• df_medium = df_medal_count[df_medal_count['label'] == 'Medium']
• print("Quốc gia trong nhóm 'Medium':")
• print(df_medium['country_id'])

•
• # Lọc dữ liệu theo nhóm 'High'
• df_high = df_medal_count[df_medal_count['label'] == 'High']
• print("Quốc gia trong nhóm 'High':")
• print(df_high['country_id'])

```

5.2.2 Random Forest.

- Import thư viện cần thiết:

```

• import pandas as pd
• from sklearn.ensemble import RandomForestClassifier
• from sklearn import datasets
• from sklearn.model_selection import train_test_split

```

```
• from sklearn import metrics  
•  
• import pydotplus  
• from IPython.display import Image  
• import matplotlib.pyplot as plt  
•
```

- Thực hiện load dữ việc và tiền xử lí tương tự như ở Decision tree.

```
• # Tạo một đối tượng Random Forest Classifier  
• model = RandomForestClassifier()  
•  
• # Huấn luyện mô hình trên tập train  
• model.fit(X_train, y_train)  
• # Đánh giá mô hình trên tập test  
• accuracy = model.score(X_test, y_test)  
• print("Độ chính xác của mô hình: ", accuracy)
```