

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

THƯ MỜI



ĐỒ ÁN

ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU VỀ LỊCH SỬ CÁC
KỲ THI VẬN HỘI OLYMPIC

MÔN HỌC: KHO DỮ LIỆU VÀ OLAP

LỚP: IS217.O22.HTCL

GVHD: Ths. Đỗ Thị Minh Phụng

SINH VIÊN THỰC HIỆN

Họ tên	MSSV
Trương Mỹ Song Dân	20520424
Phạm Thanh Nhựt	20521728

TP. HỒ CHÍ MINH, 2024

MỤC LỤC

LỜI CẢM ƠN.....	1
NHẬN XÉT CỦA GIẢNG VIÊN	2
CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN VỀ DỮ LIỆU	3
1. PHÁT BIỂU VỀ DỮ LIỆU	3
1.1. Lý do chọn đề tài.....	3
1.2. Mô tả về dữ liệu	3
1.3. Thuộc tính của kho dữ liệu.....	4
1.4. Tiền xử lí dữ liệu	13
1.5. Kho dữ liệu đã được xử lí	13
2. XÂY DỰNG KHO DỮ LIỆU	16
2.1. Lược đồ bông tuyết (snowflake schema)	16
2.2. Chi tiết các bảng.....	16
3. NỘI DUNG CÂU TRUY VẤN.....	20
CHƯƠNG 2: QUI TRÌNH TÍCH HỢP DỮ LIỆU- SQL SERVER INTERATION SERVICES (SSIS)	22
1. CẤU HÌNH CÁC THIẾT LẬP VÀ KHỞI TẠO DỰ ÁN.....	22
1.1. Cấu hình cài đặt trên Visual Studio.....	22
1.2. Khởi tạo kho dữ liệu trên SQL Server	23
1.3. Khởi tạo dự án SSIS trên Visual Studio 2022	25
2. TIẾN HÀNH THỰC HIỆN DỰ ÁN SSIS	29

2.1.	Thiết lập các kết nối	29
2.2.	Đưa dữ liệu từ tệp csv vào SQL Server	35
2.3.	Quá trình nối dữ liệu từ các bảng.....	43
2.4.	Quá trình làm sạch dữ liệu	57
2.5.	Quá trình khởi tạo các bảng Dimension.....	70
2.6.	Quá trình khởi tạo bảng FACT.....	89
2.7.	Quá trình khởi tạo các ràng buộc khoá ngoại giữa bảng Dimension và Fact .	98
2.8.	Quá trình xoá các bảng Dimension và Fact	101
2.9.	Quá trình xoá các ràng buộc khoá ngoại giữa các bảng Dimension và Fact	105
2.10.	Quá trình loại bỏ dữ liệu trùng sau khi đưa dữ liệu vào.....	109
3.	KẾT QUẢ:.....	112
4.	KIỂM TRA DỮ LIỆU	118
CHƯƠNG 3: QUI TRÌNH PHÂN TÍCH DỮ LIỆU- SQL SERVER ANALYSIS SERVICES (SSAS)		131
1.	KHỞI TẠO DỰ ÁN SSAS TRÊN VISUAL STUDIO 2022	131
2.	XÁC ĐỊNH DỮ LIỆU NGUỒN	134
3.	XÁC ĐỊNH KHUNG DỮ LIỆU NGUỒN.....	141
4.	XÁC ĐỊNH CÁC CHIỀU DỮ LIỆU	147
5.	XÁC ĐỊNH CÁC KHỐI VÀ ĐO LƯỜNG.....	177
6.	PHÂN CẤP BẢNG CHIỀU	185
7.	KHỞI CHẠY DỰ ÁN SSAS.....	188
7.1.	Đổi tên máy chủ (server name)	188
7.2.	Chạy dự án SSAS.....	190

8.	Quá trình phân tích dữ liệu bằng công cụ SSAS trên các khối cube.....	193
8.1.	Câu truy vấn 1	193
8.2.	Câu truy vấn 2	196
8.3.	Câu truy vấn 3	198
8.4.	Câu truy vấn 4	199
8.5.	Câu truy vấn 5	201
8.6.	Câu truy vấn 6	203
8.7.	Câu truy vấn 7	204
8.8.	Câu truy vấn 8	206
8.9.	Câu truy vấn 9	209
8.10.	Câu truy vấn 10	210
8.11.	Câu truy vấn 11	211
8.12.	Câu truy vấn 12	213
8.13.	Câu truy vấn 13	214
8.14.	Câu truy vấn 14	215
8.15.	Câu truy vấn 15	217
8.	Quá trình phân tích dữ liệu bằng ngôn ngữ truy vấn MDX	219
8.1.	Câu truy vấn 1	219
8.2.	Câu truy vấn 2	220
8.3.	Câu truy vấn 3	221
8.4.	Câu truy vấn 4	222
8.5.	Câu truy vấn 5	223
8.6.	Câu truy vấn 6	224

8.7.	Câu truy vấn 7	225
8.8.	Câu truy vấn 8	227
8.9.	Câu truy vấn 9	228
8.10.	Câu truy vấn 10	229
8.11.	Câu truy vấn 11	230
8.12.	Câu truy vấn 12	231
8.13.	Câu truy vấn 13	232
8.14.	Câu truy vấn 14	233
8.15.	Câu truy vấn 15	234
9.	Quá trình phân tích dữ liệu bằng công cụ Pivot Excel	236
9.1.	Câu truy vấn 1	236
9.2.	Câu truy vấn 2	236
9.3.	Câu truy vấn 3	237
9.4.	Câu truy vấn 4	238
9.5.	Câu truy vấn 5	239
9.6.	Câu truy vấn 6	240
9.7.	Câu truy vấn 7	241
9.8.	Câu truy vấn 8	242
9.9.	Câu truy vấn 9	243
9.10.	Câu truy vấn 10	244
9.11.	Câu truy vấn 11	245
9.12.	Câu truy vấn 12	246
9.13.	Câu truy vấn 13	247

9.14. Câu truy vấn 14	248
9.15. Câu truy vấn 15	249
CHƯƠNG 4: QUÁ TRÌNH LẬP BÁO BIỂU (SSRS).....	251
1. QUÁ TRÌNH LẬP BÁO BIỂU BẰNG CÔNG CỤ MICROSOFT SSRS (Report Builder).....	251
1.1. Tạo project SSRS	251
1.2. Tạo kết nối với kho dữ liệu	255
1.3. Thực hiện lập báo biểu trên Report Builder.....	258
2. QUÁ TRÌNH LẬP BÁO BIỂU BẰNG CÔNG CỤ POWER BI.....	279
2.1. Thiết lập kết nối dữ liệu với Power BI.....	279
2.2. Lập báo biểu.....	284
CHƯƠNG 5: QUY TRÌNH KHAI THÁC DỮ LIỆU – DATA MINING	294
1. TỔNG QUAN	294
2. DỮ LIỆU	294
3. VẼ BIỂU ĐỒ.....	300
4. CHIA TẬP DỮ LIỆU	302
5. ĐÁNH GIÁ MÔ HÌNH	303
6. THUẬT TOÁN.....	303
6.1. Hồi quy tuyến tính (Linear Regression).....	303
6.2. Long Short-Term Memory (LSTM).....	305
7. KẾT QUẢ.....	309
TÀI LIỆU THAM KHẢO	310

MỤC LỤC HÌNH ẢNH

Hình 1. Bộ dữ liệu olympic_athletes.csv.....	6
Hình 2. Bộ dữ liệu olympic_hosts.csv.....	8
Hình 3 Bộ dữ liệu olympic_medals	10
Hình 4 Bộ dữ liệu olympic_results.csv	13
Hình 5 Sơ đồ bông tuyết của bộ dữ liệu Olympic	16
Hình 6 Cài đặt extension	22
Hình 7 Cài đặt SSIS Projects.....	23
Hình 8 Kết nối SQL Server	24
Hình 9 Tạo các cơ sở dữ liệu mới.....	25
Hình 10 Tạo project SSIS mới.....	26
Hình 11 Chọn project.....	27
Hình 12 Đặt tên và chọn thư mục cho Project.....	28
Hình 13 Màn hình chính của SSIS Project.....	29
Hình 14 Chọn Connection Manager.....	30
Hình 15 Tạo mới Connection Manager	30
Hình 16 Chọn loại Connection Manager là OLEDB	31
Hình 17 Tạo kết nối mới.....	32
Hình 18 Chọn nguồn dữ liệu	33
Hình 19 Kiểm tra kết nối	34
Hình 20 Chọn Data Connection	34
Hình 21 Tạo kết nối với Raw Data thành công	35
Hình 22 Tạo kết nối với DataWarehouse thành công	35
Hình 23 Thêm dữ liệu vào SQL Server	36

Hình 24 Chọn nguồn dữ liệu	37
Hình 25 Xem trước dữ liệu.....	38
Hình 26 Điều chỉnh các cột	39
Hình 27 Tổng quan quá trình thêm dữ liệu	40
Hình 28 Thêm dữ liệu thành công	41
Hình 29 Thêm dữ liệu Athletes thành công	42
Hình 30 Thêm dữ liệu Hosts thành công	42
Hình 31 Thêm dữ liệu Medals thành công	43
Hình 32 Thêm dữ liệu Results thành công	43
Hình 33 Tạo data flow task.....	44
Hình 34 Tạo data Source	44
Hình 35 Chọn nguồn dữ liệu	45
Hình 36 Nối Result với Sort	46
Hình 37 Lựa chọn thứ tự sort các thuộc tính	46
Hình 38 Data source Athletes	47
Hình 39 Sort Athletes	47
Hình 40 Lựa chọn thứ tự để sắp xếp các thuộc tính	48
Hình 41 Data source Hosts	49
Hình 42 Nối Host với Sort.....	49
Hình 43 Lựa chọn các thuộc tính để sắp xếp.....	49
Hình 44 Merge 2 nguồn dữ liệu.....	50
Hình 45 Chọn Merge Join Left input.....	50
Hình 46 Nối các thuộc tính của 2 nguồn dữ liệu	51
Hình 47 Sắp xếp lại sau khi nối.....	52
Hình 48 Lựa chọn thứ tự để sắp xếp các thuộc tính	52
Hình 49 Thực hiện nối với nguồn dữ liệu Hosts	53
Hình 50 Lựa chọn các thuộc tính để sắp xếp.....	54
Hình 51 Thực hiện sắp xếp sau khi nối	55

Hình 52 Lựa chọn các thuộc tính để sắp xếp.....	55
Hình 53 Đưa dữ liệu vào Raw_Data	56
Hình 54 Tạo mới bảng Raw_Data.....	56
Hình 55 Nối các cột trong bảng.....	57
Hình 56 Nối data flow mới với data flow nối dữ liệu	58
Hình 57 Tạo nguồn dữ liệu Raw_Data	58
Hình 58 Tạo kết nối với bảng Raw_Data	59
Hình 59 Chọn các cột để sử dụng.....	60
Hình 60 Tạo công cụ Conditional Split	61
Hình 61 Tạo các hàm NULL_Output và NotNULL_Output	62
Hình 62 Chọn Null_Output	63
Hình 63 Tạo đích đến Data_NULL	63
Hình 64 Tạo bảng Data_NULL	64
Hình 65 Nối các cột thuộc tính.....	65
Hình 66 Sắp xếp dữ liệu sau khi xử lí	66
Hình 67 Lựa chọn các thuộc tính để sắp xếp.....	67
Hình 68 Tạo đích đến Data_Cleaned.....	68
Hình 69 Tạo bảng Data_Cleaned.....	69
Hình 70 Nối các cột trong bảng.....	70
Hình 71 Tạo Sequence Container “Create Dim Table”	71

MỤC LỤC BẢNG BIỂU

Bảng 1. Bảng các thuộc tính của bộ dữ liệu Olympic_Athletes.....	5
Bảng 2. Bảng các thuộc tính của bộ dữ liệu olympic_host	6
Bảng 3 Bảng các thuộc tính của bộ dữ liệu Olympic_medals.....	8
Bảng 4 Bảng các thuộc tính của bộ dữ liệu Olympic_results.....	10
Bảng 5 Bảng các thuộc tính sau khi xử lí dữ liệu.....	13
Bảng 6 Bảng thuộc tính của Fact_Olympic.....	16
Bảng 7 Bảng thuộc tính của Dim_Athlete.....	17
Bảng 8 Bảng thuộc tính của Dim_Game	17
Bảng 9 Bảng thuộc tính của Dim_Result	18
Bảng 10 Bảng thuộc tính của Dim_Time	18
Bảng 11 Bảng thuộc tính của Dim_Country	19
Bảng 12 Bảng thuộc tính của Dim_Competition	19

LỜI CẢM ƠN

Chúng em muốn gửi lời biết ơn đến đội ngũ giảng viên của trường đại học Công Nghệ Thông Tin, nhất là với các thầy cô của khoa Hệ Thông Tin đã cung cấp cho chúng em nhiều kiến thức cần thiết cùng những kỹ năng cũng như hỗ trợ cơ sở vật chất để thực hiện đồ án. Đặc biệt, chúng em muốn gửi lời biết ơn đến ThS. *Đỗ Thị Minh Phụng* đã tận tình hướng dẫn, chia sẻ kiến thức cùng kinh nghiệm, góp ý và chỉnh sửa nhằm giúp đỡ chúng em kết thúc kỳ học được nhanh chóng.

Xuất phát từ mục đích học tập và xây dựng kho dữ liệu, các phương pháp xây dựng kho dữ liệu và phân tích dữ liệu trên kho dữ liệu về lịch sử các kỳ thi vận hội Olympic. Dựa trên những kiến thức được Cô cung cấp trên trường kết hợp với việc tự tìm hiểu những công cụ và kiến thức mới, nhóm đã cố gắng thực hiện đồ án một cách tốt nhất.

Qua thời gian thực hiện đồ án này, mỗi thành viên trong nhóm đã vận dụng những kỹ năng cần thiết và phối hợp với nhau để tìm hiểu kiến thức mới. Vì vậy, nhóm đã cố gắng hết mình để sử dụng các kiến thức có sẵn nhằm hoàn thiện đồ án này một cách tốt nhất. Dù vậy, trong quá trình thực hiện đồ án cũng khó mà tránh khỏi việc gặp phải một số lỗi. Vì vậy, chúng em mong có được sự góp ý của thầy cô để chúng em tiếp tục hoàn thiện kiến thức và chỉnh sửa, dùng nó phục vụ vào những dự án khác trong tương lai.

Chúng em xin trân trọng cảm ơn!

Nhóm sinh viên thực hiện

NHẬN XÉT CỦA GIẢNG VIÊN

CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN VỀ DỮ LIỆU

1. PHÁT BIỂU VỀ DỮ LIỆU

1.1. Lý do chọn đề tài

Thể vận hội thể thao Olympic là một trong những sự kiện thể thao nổi tiếng nhất với quy mô trên toàn cầu. Sự kiện Thể vận hội có nhiều cuộc thi thể thao mùa đông và mùa hè với hàng nghìn vận động viên tham gia tranh tài đến từ hơn 200 quốc gia trên thế giới. Ủy ban Olympic quốc tế (IOC) tổ chức Thể vận hội Olympic mùa đông và Thể vận hội Olympic mùa hè cứ bốn năm một lần, và cứ cách nhau hai năm sẽ có một kỳ Thể vận hội được diễn ra. Thể vận hội gần nhất là Thể vận hội mùa đông vừa diễn ra tại thủ đô Bắc Kinh, Trung Quốc vào năm 2022. Thể vận hội mùa hè được tổ chức tại Tokyo, Nhật Bản vào năm 2020, Paris, Pháp vào năm 2024 và Los Angeles, Mỹ vào năm 2028.

Lần đầu tiên ngọn đuốc Olympic 2008 đến Việt Nam vào cuối tháng 04/2008 và là qua Thành phố Hồ Chí Minh. Đó chính là vị thế và danh dự của Thành phố Hồ Chí Minh đối với cả nước và trên trường quốc tế.

Với tinh thần thể thao cao thượng và trong sáng, nhân dân Việt Nam nói chung và nhân dân Thành phố Hồ Chí Minh nói riêng hết lòng ủng hộ và cổ vũ cho Olympic 2008, sự kiện thể thao quan trọng nhất của nhân dân thế giới được tổ chức tại Bắc Kinh. Đội tuyển quốc gia của Việt Nam cũng tham gia nhiều môn thi đấu tại Olympic Bắc Kinh 2008 lần này như một sự hưởng ứng tinh thần thượng võ của Đại hội thể thao lớn nhất hành tinh – Thể vận hội Olympic.

Chính vì vậy mà nhóm em đã chọn đề tài “**Xây dựng kho dữ liệu về lịch sử các kỳ Thể vận hội Olympics**” để tìm hiểu rõ hơn cũng như muốn có thêm nhiều kiến thức về thời đại.

1.2. Mô tả về dữ liệu

- Bộ dữ liệu Olympics Historical Dataset: gồm các bộ dữ liệu của sự kiện Thể vận hội

- Olympic cập nhật (từ năm 1896 đến 2022) cho bất kỳ người đam mê thể thao/dữ liệu nào sử dụng để trực quan hóa và tạo một số thông tin hiểu biết về bộ dữ liệu sự kiện thế vận hội Olympic.
- Ngày cập nhật gần nhất: 12/04/2022 (cập nhật thêm dữ liệu của Olympic Beijing 2022)
- Thông tin cung cấp từ kho dữ liệu:
 - Đây là một bộ dữ liệu lịch sử về Thế vận hội Olympic, bao gồm tất cả các nội dung thi từ Athens 1896 đến Bắc Kinh 2022. Bộ dữ liệu bao gồm kết quả, huy chương, vận động viên và chủ nhà. Dữ liệu được tạo ra từ Thế vận hội.
 - Bộ dữ liệu cung cấp:
 - Hơn 21.000 nội dung thi
 - 162.000 kết quả
 - 74.000 vận động viên
 - 20.000 tiểu sử
 - 53 quốc gia đăng cai của Thế vận hội Olympic mùa hè và mùa đông.
- Đơn vị cung cấp nguồn dữ liệu: Dữ liệu được thu thập từ website olympics.com có dữ liệu Olympic mới nhất được đặt từ Thế vận hội Mùa hè Athene năm 1896 đến Thế vận hội mùa đông Bắc Kinh 2022.
- Đường dẫn nguồn dữ liệu: [tại đây](#).

1.3. Thuộc tính của kho dữ liệu

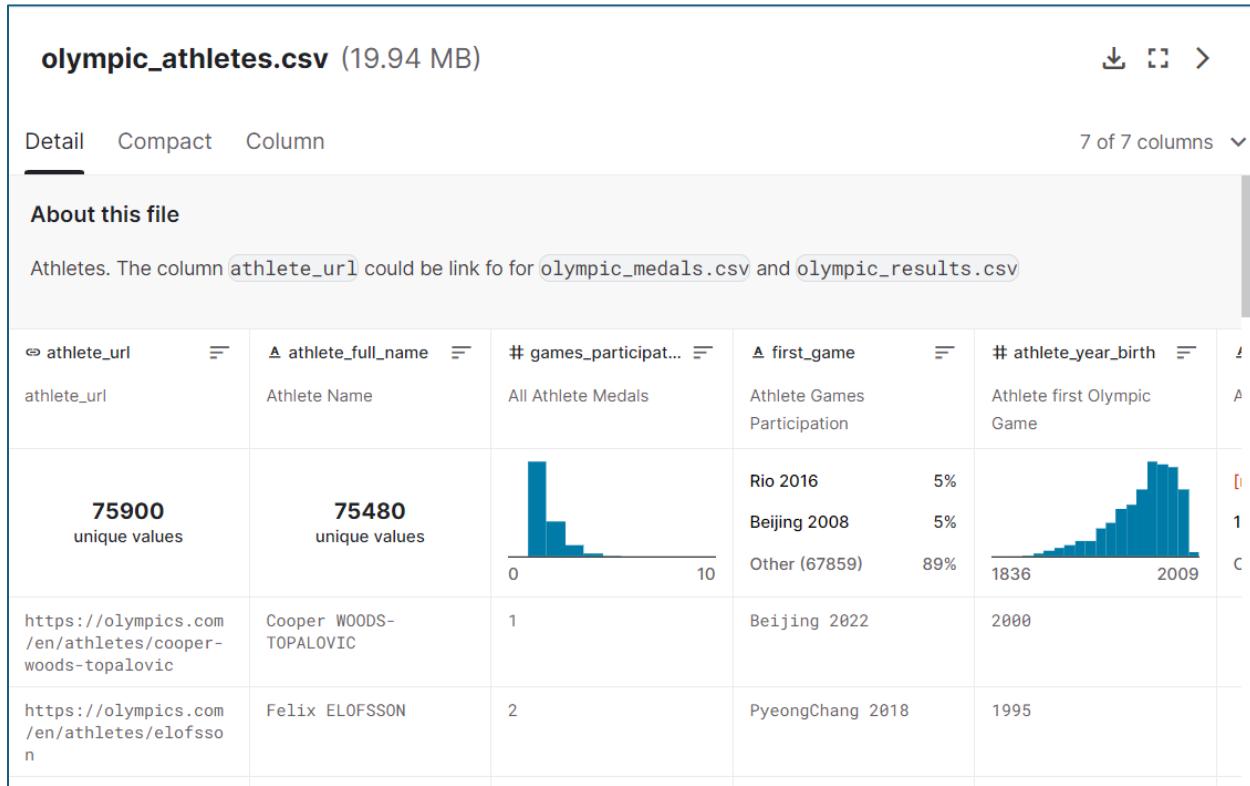
Các thuộc tính của bộ dữ liệu:

1.3.1. Bộ dữ liệu *olympic_athletes*

Gồm 75,903 dòng dữ liệu và 7 thuộc tính.

Bảng 1. Bảng các thuộc tính của bộ dữ liệu Olympic_Athletes

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Athlete_url	url	Đường dẫn thông tin vận động viên
2	Athlete_full_name	Text	Họ tên đầy đủ của vận động viên
3	Competitions_participants	Number	Số lần tham dự các kỳ thi vận hội
4	First_competitions	Text	Tên kì thi vận hội đầu tiên tham gia
5	Athlete_year_birth	Number	Năm sinh của vận động viên
6	Athlete_medals	Number	Số huy chương đạt được
7	bio	Text	Thông tin tiểu sử của vận động viên



Hình 1. Bộ dữ liệu *olympic_athletes.csv*

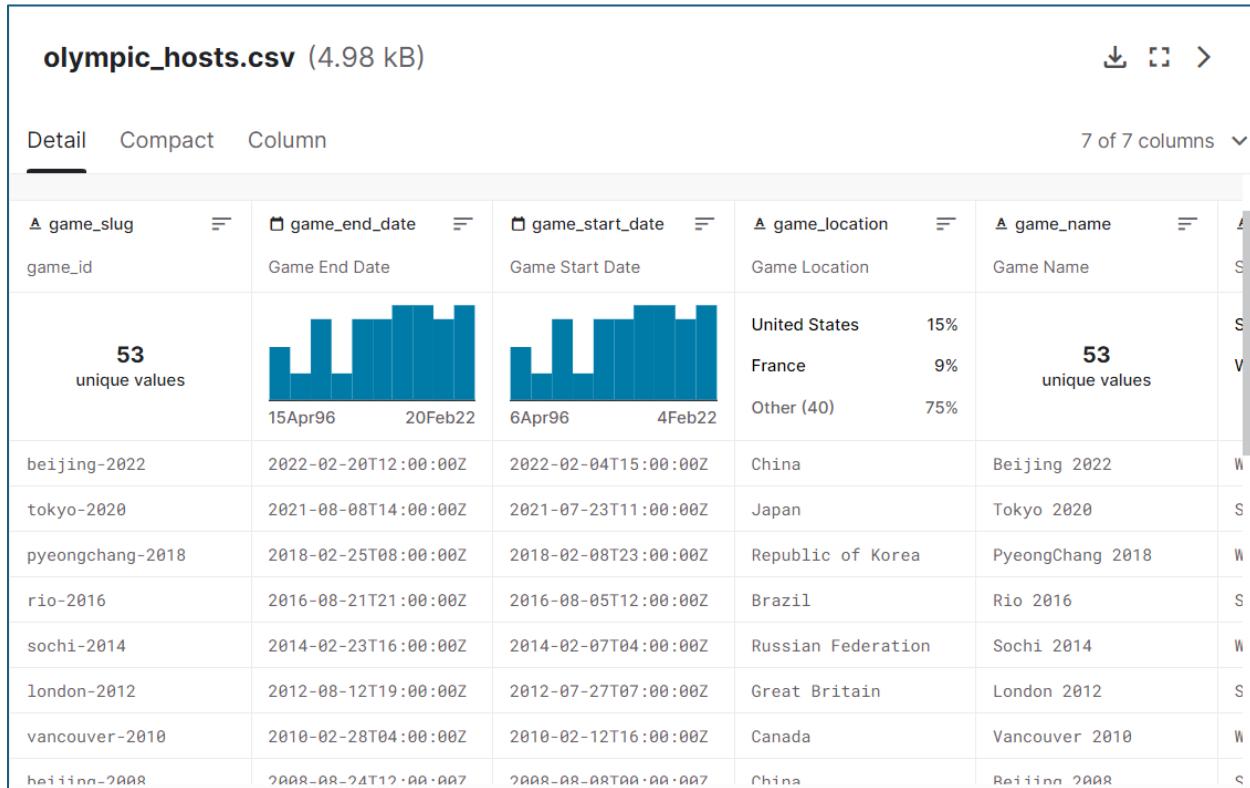
1.3.2. Bộ dữ liệu *olympic_hosts*

Gồm 53 dòng dữ liệu và 7 thuộc tính.

Bảng 2. Bảng các thuộc tính của bộ dữ liệu *olympic_host*

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Competition_slug	Text	Mã thẻ vận hội
2	Competition_end_date	Datetime	Thời gian bế mạc thẻ vận hội
3	Competition_start_date	Datetime	Thời gian khai mạc thẻ vận hội
4	Competition_location	Text	Địa điểm tổ chức

5	Competition_name	Text	Tên thể vận hội Định dạng: địa điểm tổ chức + “” + Năm tổ chức
6	Competition_season	Text	Mùa thể vận hội Gồm 2 mùa: Hè và đông (tương ứng Summer và Winter)
7	Competition_year	Text	Năm tổ chức thể vận hội



Hình 2. Bộ dữ liệu olympic_hosts.csv

1.3.3. Bộ dữ liệu olympic_medals

Gồm 21,697 dòng dữ liệu và 12 thuộc tính.

Bảng 3 Bảng các thuộc tính của bộ dữ liệu Olympic_medals

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Discipline_title	Text	Thể loại bộ môn thi đấu
2	Slug_competition	Text	Mã thể vận hội
3	Event_tittle	Text	Nội dung thi đấu của thể loại bộ môn thi đấu

4	Event_gender	Text	<p>Giới tính dành cho bộ môn thi đấu.</p> <p>Gồm 4 loại: Nam, Nữ, Nam + Nữ, Mở rộng (tương ứng: Men, Woman, Mixed, Open)</p>
5	Medal_type	Text	<p>Loại huy chương đạt được.</p> <p>Gồm 3 loại: Vàng, Bạc, Đồng (tương ứng: GOLD, SILVER, BRONZE)</p>
6	Participant_type	Text	<p>Hình thức tham gia của nội dung thi đấu.</p> <p>Gồm 2 loại: cá nhân và tiếp sức (tương ứng: Athlete và CompetitionTeam)</p>
7	Participant_title	Text	Tên đội tham gia
8	Athlete_url	Text	Đường dẫn thông tin vận động viên
9	Athlete_full_name	Text	Tên vận động viên

10	Country_name	Text	Tên quốc gia đại diện
11	Country_code	Text	Mã quốc gia đại diện
12	Country_3_letter_code	Text	Mã quốc gia gồm 3 kí tự

olympic_medals.csv (3.11 MB) ↓ ☰ >

Detail Compact Column 12 of 12 columns ▾

About this file

Medals. The column `slug_game` is the link for `olympic_metadata_host.csv`.

▲ discipline_title	Discipline	▲ slug_game	game_id	▲ event_title	Event	▲ event_gender	Gender	▲ medal_type	Medal type	▲
Athletics Swimming Other (16854)	14% 8% 78%	tokyo-2020 rio-2016 Other (19446)	5% 5% 90%	1436 unique values	Men Women Other (1442)	64% 29% 7%	BRONZE GOLD Other (7059)	35% 33% 33%	A G	
Curling		beijing-2022		Mixed Doubles		Mixed	GOLD		G	
Curling		beijing-2022		Mixed Doubles		Mixed	GOLD		G	
Curling		beijing-2022		Mixed Doubles		Mixed	SILVER		G	

Hình 3 Bộ dữ liệu olympic_medals

1.3.4. Bộ dữ liệu olympic_results

Gồm 162,688 dòng dữ liệu và 15 thuộc tính.

Bảng 4 Bảng các thuộc tính của bộ dữ liệu Olympic_results

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
-----	----------------	--------------	---------

1	Discipline_tittle	Text	Thể loại bộ môn thi đấu
2	Event_tittle	Text	Nội dung thi đấu của thể loại bộ môn thi đấu
3	Slug_competition	Text	Mã thể vận hội
4	Participant_type	Text	Hình thức tham gia của nội dung thi đấu Gồm 2 loại: cá nhân và tiếp sức (tương ứng: Athlete và CompetitionTeam)
5	Medal_type	Text	Loại huy chương đạt được. Gồm 3 loại: Vàng, Bạc, Đồng (tương ứng: GOLD, SILVER, BRONZE)
6	Athlete	Array	Đường dẫn thông tin vận động viên
7	Rank_equal	Text	Có xảy ra đồng hạng hay không
8	Rank_position	Text	Thứ hạng của nội dung

9	Country_name	Text	Tên quốc gia đại diện
10	Country_code	Text	Mã quốc gia đại diện
11	Country_3_letter_code	Text	Mã quốc gia gồm 3 ký tự
12	Athlete_url	Text	Đường dẫn thông tin vận động viên
13	Athlete_full_name	Text	Tên vận động viên
14	Value_unit	Text	Đơn vị giá trị tính điểm
15	Value_type	Text	Loại giá trị tính điểm

olympic_results.csv (24.04 MB)

Detail Compact Column 15 of 15 columns

About this file

Results. The column `slug_game` is the link for `olympic_metadata_host.csv`.

discipline_title	event_title	slug_game	participant_type	medal_type						
Discipline	Event	game_id	Participant Type	Medal Type						
Athletics	12%	individual mixed	2%	rio-2016	5%	Athlete	87%	[null]	88%	[null]
Swimming	9%	individual road race...	2%	london-2012	5%	GameTeam	13%	BRONZE	4%	[null]
Other (128841)	79%	Other (157429)	97%	Other (147601)	91%			Other (13177)	8%	C
Curling		Mixed Doubles		beijing-2022		GameTeam		GOLD		M

Hình 4 Bộ dữ liệu `olympic_results.csv`

1.4. Tiền xử lý dữ liệu

1.5. Kho dữ liệu đã được xử lý

Sau khi thực hiện việc tiền xử lý dữ liệu, ta thu được **129.932 dòng dữ liệu** và **20 thuộc tính** để sử dụng cho việc phân tích

Bảng 5 Bảng các thuộc tính sau khi xử lý dữ liệu

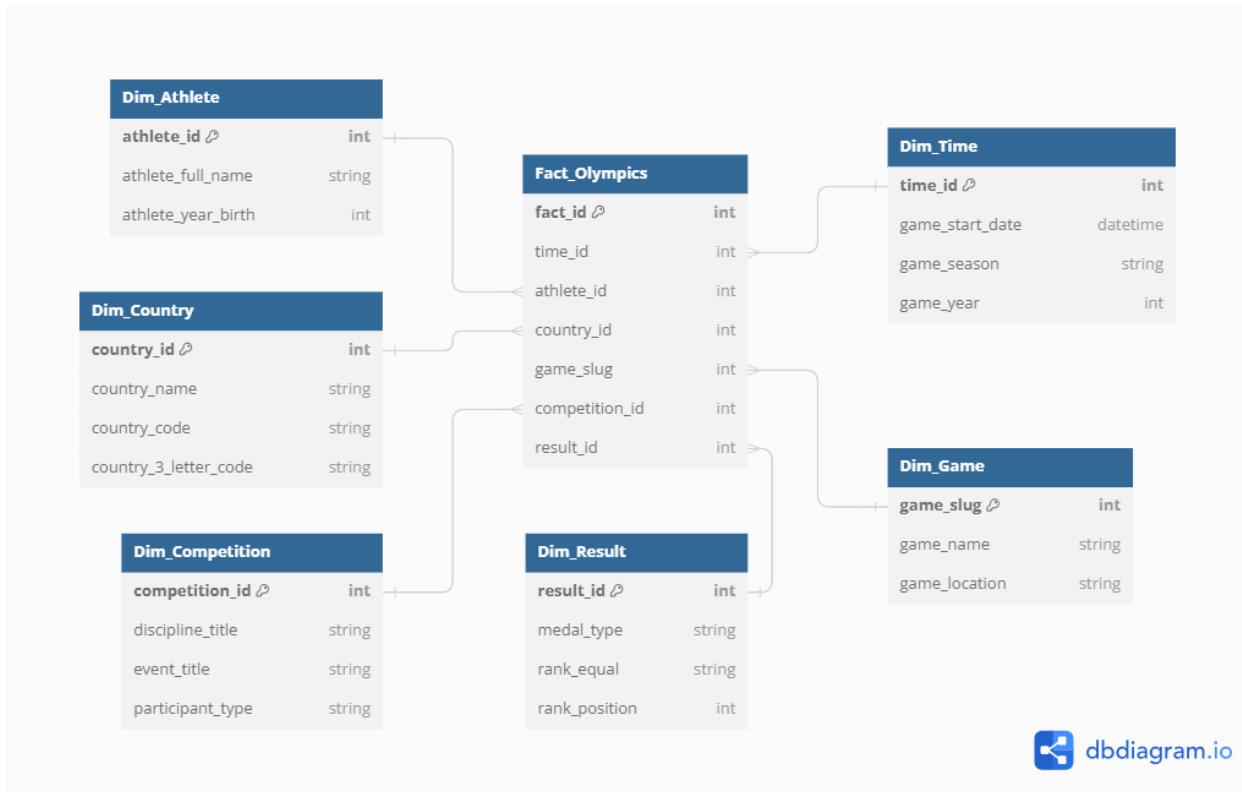
STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	slug_game	Text	Mã thể vận hội
2	game_start_date	DateTime	Ngày khai mạc thể vận hội
3	game_end_date	DateTime	Ngày bế mạc thể vận hội

4	game_location	Text	Địa điểm tổ chức
5	game_name	Text	Tên thể vận hội
6	game_season	Text	Mùa thể vận hội
7	game_year	Text	Năm tổ chức thể vận hội
8	discipline_title	Text	Thể loại bộ môn thi đấu
9	event_title	Text	Nội dung thi đấu của thể loại bộ môn thi đấu
10	participant_type	Text	Hình thức tham gia của nội dung thi đấu. Gồm 2 loại: cá nhân & tiếp sức (giá trị: Athlete & CompetitionTeam)
11	metal_type	Text	Loại huy chương đạt được. Gồm 3 loại: Vàng, Bạc, Đồng (giá trị: GOLD, SILVER, BRONZE)
12	country_name	Text	Tên quốc gia đại diện

13	country_code	Text	Mã quốc gia đại diện
14	country_3_letter_code	Text	Mã quốc gia gồm 3 ký tự
15	rank_equal	Text	Có xảy ra đồng hạng hay không?
16	rank_position	Text	Thứ hạng của nội dung
17	athlete	Array	Đường dẫn thông tin vận động viên (team 2 người trở lên)
18	athlete_url	url	Đường dẫn thông tin vận động viên
19	athlete_full_name	Text	Họ tên đầy đủ của vận động viên
20	athlete_year_birth	Number	Năm sinh của vận động viên

2. XÂY DỰNG KHO DỮ LIỆU

2.1. Lược đồ bông tuyết (snowflake schema)



Hình 5 Sơ đồ bông tuyết của bộ dữ liệu Olympic

2.2. Chi tiết các bảng

2.2.1. Bảng Fact_Olympic

Bảng 6 Bảng thuộc tính của Fact_Olympic

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Fact_id (Primary Key)	int	Mã của bảng sự kiện
3	Time_id	int	Mã thời gian
4	Athlete_id	int	Mã vận động viên

5	Country_id	int	Mã quốc gia đại diện
6	Game_slug	int	Mã thể vận hội
7	Competition_id	int	Mã môn thi
8	Result_id	int	Mã kết quả

2.2.2. *Bảng Dim_Athlete*

Bảng 7 Bảng thuộc tính của Dim_Athlete

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Athlete_id (Primary Key)	int	Mã vận động viên
2	Athlete_full_name	string	Tên đầy đủ của vận động viên
3	Athlete_year_birth	int	Năm sinh của vận động viên

2.2.3. *Bảng Dim_Game*

Bảng 8 Bảng thuộc tính của Dim_Game

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Game_slug (Primary Key)	string	Mã thể vận hội
2	Game_name	string	Tên vận hội
3	Game_location	string	Địa điểm tổ chức thể vận hội

2.2.4. *Bảng Dim_Result*

Bảng 9 Bảng thuộc tính của Dim_Result

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	result_id (Primary Key)	int	Mã kết quả
2	Medal_type	String	Loại huy chương đạt được. Gồm 3 loại: Vàng, Bạc, Đồng (giá trị: GOLD, SILVER, BRONZE)
	Rank_equal	int	Có xảy ra đồng hạng hay không?
	Rank_position	int	Thứ hạng của nội dung

2.2.5. *Bảng Dim_Time*

Bảng 10 Bảng thuộc tính của Dim_Time

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Time_id (Primary Key)	int	Mã thời gian
2	Game_start_date	Datetime	Thời gian bắt đầu
3	Game_season	String	Mùa của thời gian thi đấu
4	Game_year	Int	Năm thi đấu

2.2.6. *Bảng Dim_Country*

Bảng 11 Bảng thuộc tính của Dim_Country

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Country_id	Int	Mã quốc gia
2	Country_name	String	Tên quốc gia
3	Country_code	String	Mã code của quốc gia
4	Country_3_letter_code	String	Mã code 3 chữ của quốc gia

2.2.7. *Bảng Dim_Competition*

Bảng 12 Bảng thuộc tính của Dim_Competition

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Competition_id	Int	Mã cuộc thi
2	Discipline_title	String	Thể loại bộ môn thi đấu
3	Event_tittle	String	Nội dung thi đấu của thể loại bộ môn thi đấu
4	Participant_type	String	Hình thức tham gia của nội dung thi đấu.

		Gồm 2 loại: cá nhân & tiếp sức (giá trị: Athlete & CompetitionTeam)
--	--	---

3. NỘI DUNG CÂU TRUY VẤN

- **Câu 1:** Liệt kê top 10 vận động viên đạt huy chương vàng nhất, sắp xếp theo thứ tự giảm dần.
- **Câu 2:** Liệt kê tất cả các vận động viên tham gia Olympic từ 3 lần trở lên, sắp theo theo thứ tự tăng dần
- **Câu 3:** Liệt kê 10 môn thi đấu có số lượng huy chương đồng nhiều nhất trong kỳ thế vận hội Tokyo 1964.
- **Câu 4:** Vận động viên France dành được nhiều huy chương bạc nhất
- **Câu 5:** Quốc gia dành được nhiều huy vàng nhất từ năm 1980 đến năm 2000
- **Câu 6:** Thống kê số lượng vận động viên của các quốc gia tham gia thi đấu môn Ski Jumping trong thế vận hội Sochi 2014.
- **Câu 7:** Quốc gia đạt số lượng huy chương bạc nhiều nhất thế vận hội London 2012.
- **Câu 8:** Liệt kê top 5 nội dung thi có số lượng vận động viên tham gia nhiều nhất trong kỳ thế vận hội Tokyo 2020.
- **Câu 9:** Thống kê số lượng vận động viên tham gia các nội dung thi đấu qua các kỳ Olympics đến nay của đội tuyển Việt Nam.
- **Câu 10:** Liệt kê số lượng chi tiết từng huy chương của Pháp theo từng kỳ Olympics.
- **Câu 11:** Với mỗi quốc gia đưa ra 3 môn thi đấu có nhiều vận động viên tham gia nhiều nhất từ trước đến nay
- **Câu 12:** Các vận động viên bị trùng hạng tư của nước United States of America
- **Câu 13:** Tìm vận động viên giành được nhiều huy chương vàng nhất Olympic Beijing 2008

- **Câu 14:** Kì Olympic với nhiều vận động viên tham gia nhất
- **Câu 15:** Top 10 các môn thi có số lượng vận động viên tham gia nhiều nhất với địa điểm tổ chức Olympic nằm ở quốc gia Great Britain

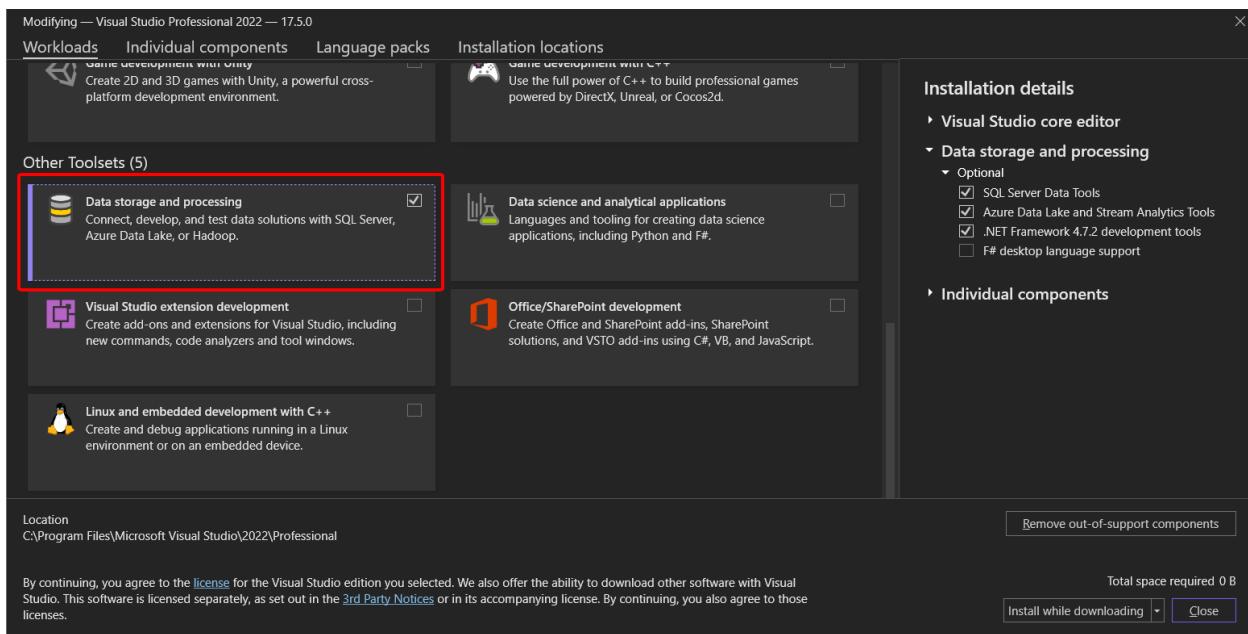
CHƯƠNG 2: QUI TRÌNH TÍCH HỢP DỮ LIỆU- SQL SERVER INTERACTION SERVICES (SSIS)

1. CẤU HÌNH CÁC THIẾT LẬP VÀ KHỞI TẠO DỰ ÁN

1.1. Cấu hình cài đặt trên Visual Studio

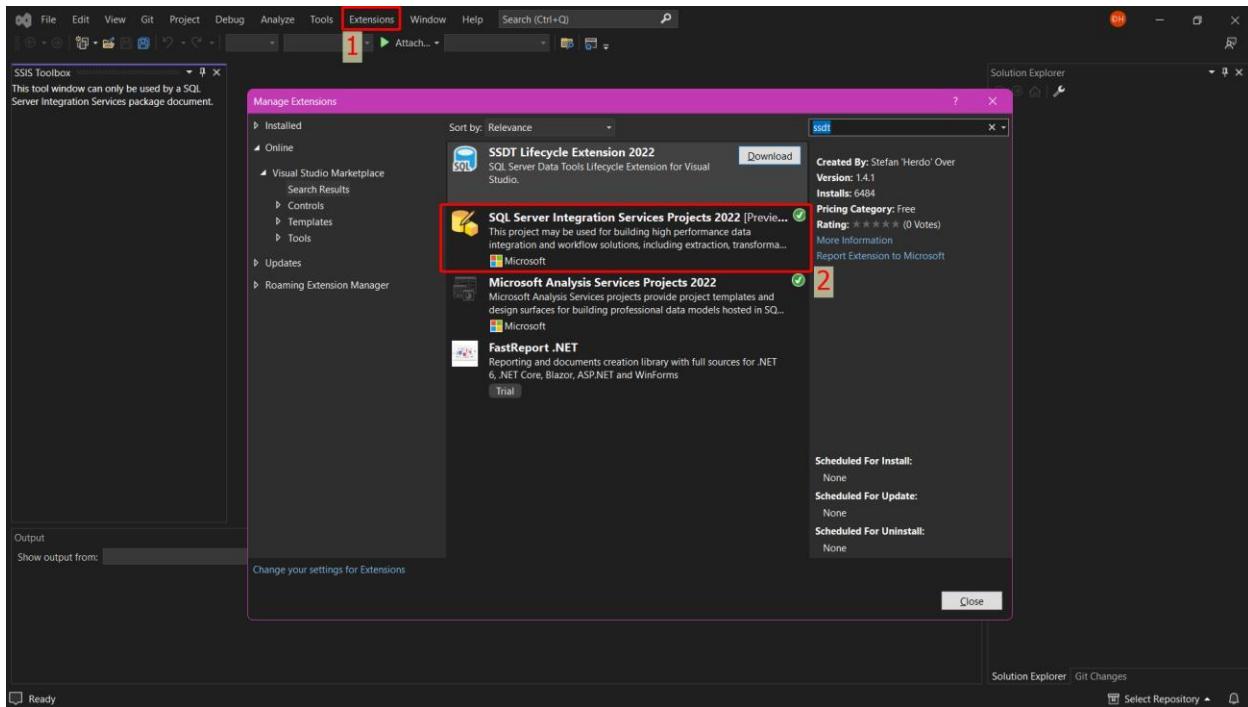
Để bắt đầu qui trình SSIS, ta cần cài đặt các công cụ như sau:

- SQL Server 2019.
- Visual Studio 2022.
- Gói extension **Data Storage and Processing** trong Visual Studio.



Hình 6 Cài đặt extension

- Công cụ SQL Server Data Tools: Truy cập **Visual Studio 2022 → Extensions → Manage Extensions → SSIS Projects 2022**

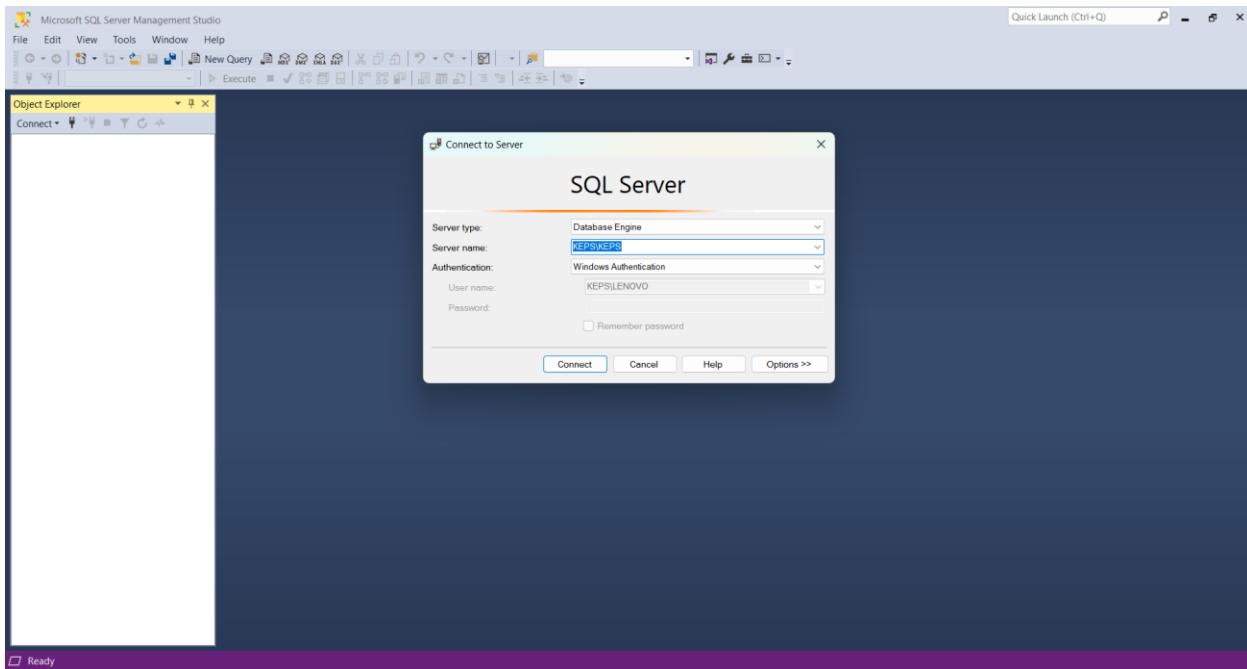


Hình 7 Cài đặt SSIS Projects

1.2. Khởi tạo kho dữ liệu trên SQL Server

Các bước khởi tạo kho dữ liệu bằng SQL Server 2019:

- Bước 1: Truy cập **SQL Server Management Studio** và kết nối đến Server.



Hình 8 Kết nối SQL Server

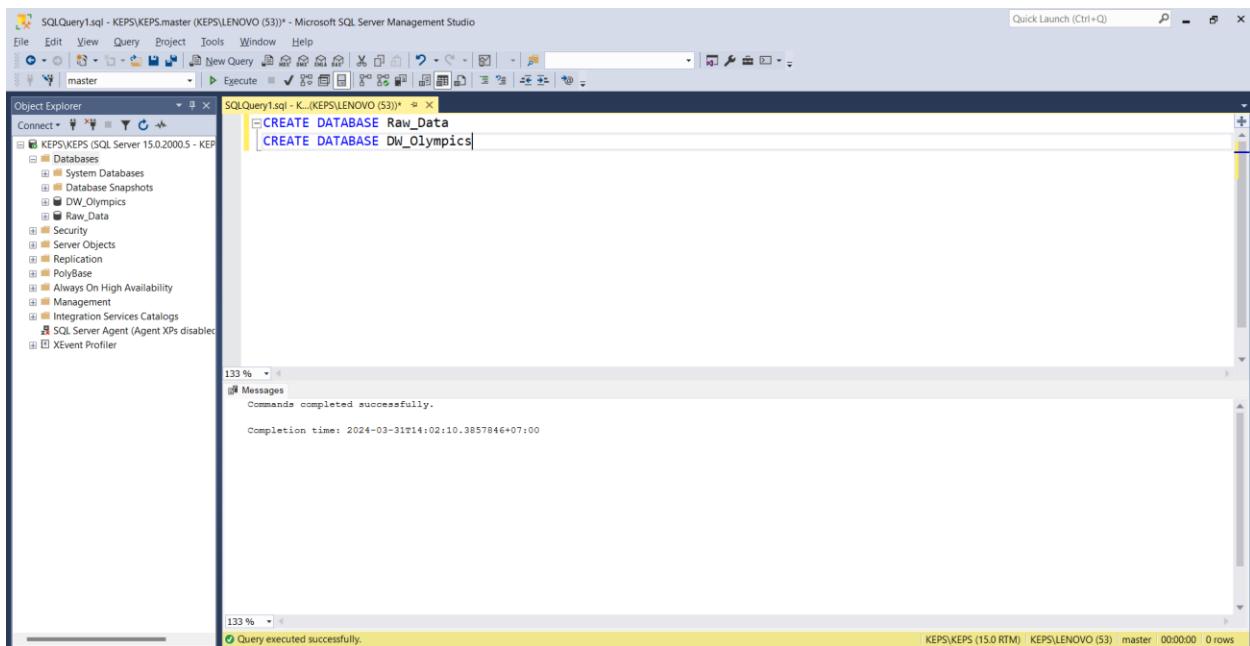
- Bước 2: Khởi tạo và cấu hình cho kho dữ liệu

 - + *Tạo hai cơ sở dữ liệu mới:*

 - Một cơ sở dữ liệu lưu trữ thông tin tập dataset chứa dữ liệu gốc được import trực tiếp từ các file csv, dữ liệu sử dụng sau khi làm sạch và các dữ liệu rỗng.
 - Một cơ sở dữ liệu lưu trữ thông tin của kho dữ liệu bao gồm các bảng Dimensions và Fact sau khi thực hiện quá trình ETL.

 - + *Các bước thực hiện*

 - Tại màn hình chính, chọn **New Query** để mở màn hình truy vấn mới.
 - Tạo 2 cơ sở dữ liệu mới có tên là RawData và DW_Olympics bằng cú pháp: CREATE DATABASE RawData; CREATE DATABASE DW_Olympics.
 - Sau đó, cửa sổ **Object Explorer** sẽ xuất hiện 2 cơ sở dữ liệu vừa tạo.

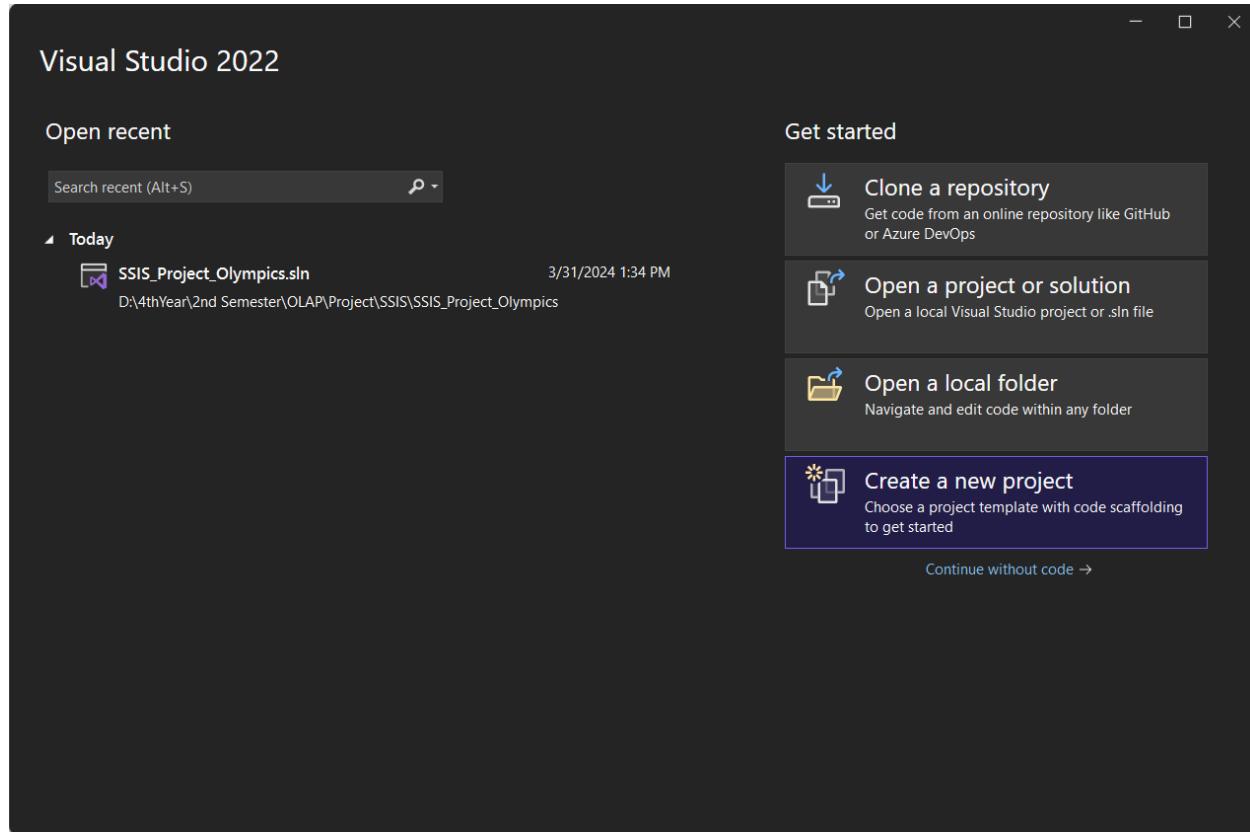


Hình 9 Tạo các cơ sở dữ liệu mới

1.3. Khởi tạo dự án SSIS trên Visual Studio 2022

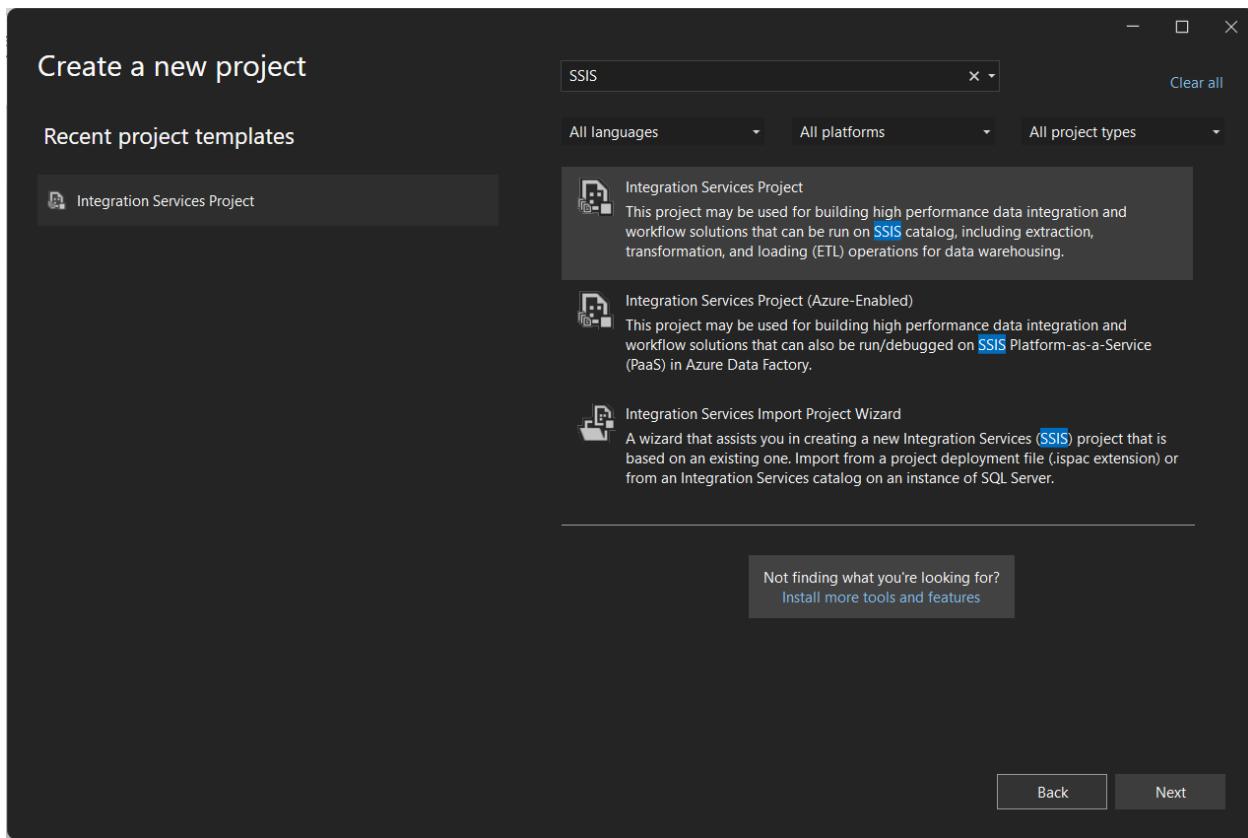
Các bước khởi tạo dự án SSIS trên Visual Studio 2022:

- Bước 1: Truy cập **Visual Studio 2022** → Chọn **Create a new project**

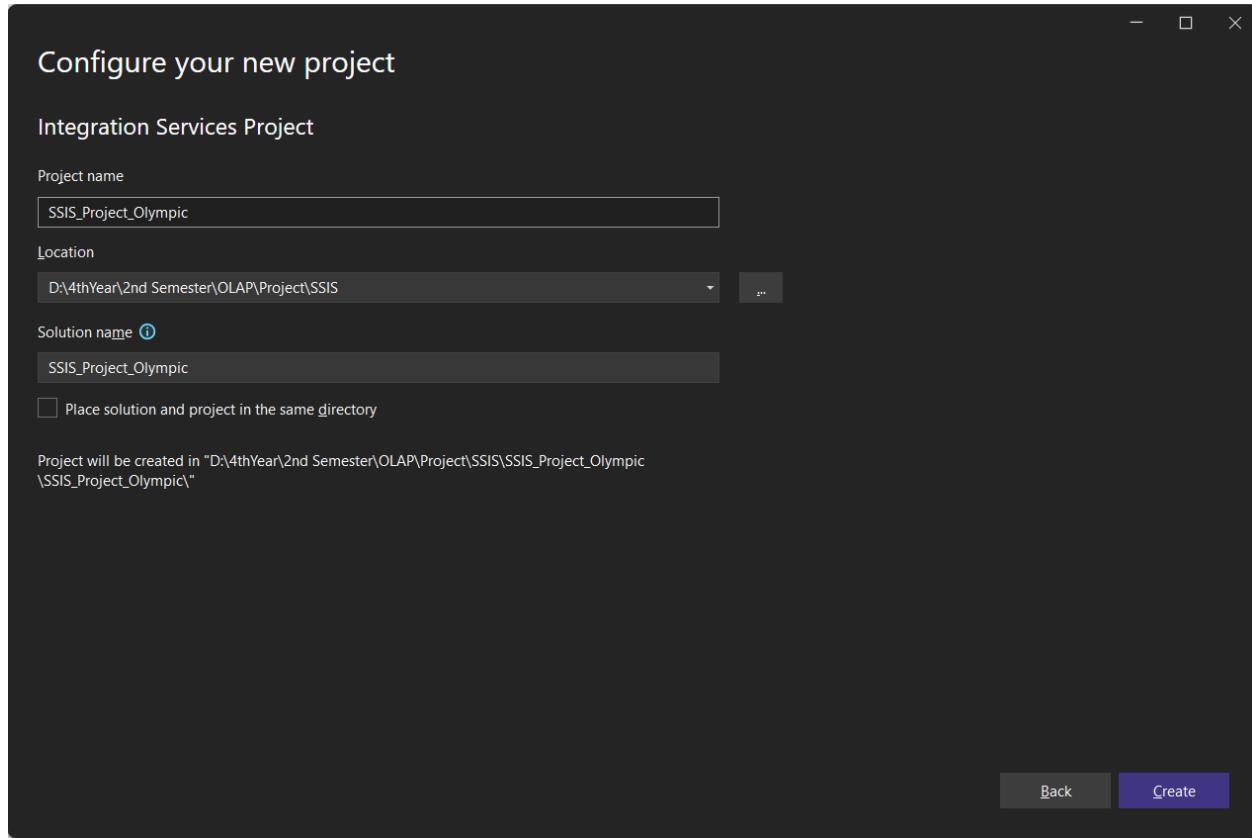


Hình 10 Tạo project SSIS mới

- Bước 2: Tìm kiếm từ khoá SSIS và chọn **Integration Services Project** → Đổi tên project thành **SSIS_Project_Olympics** → Chọn **Create**

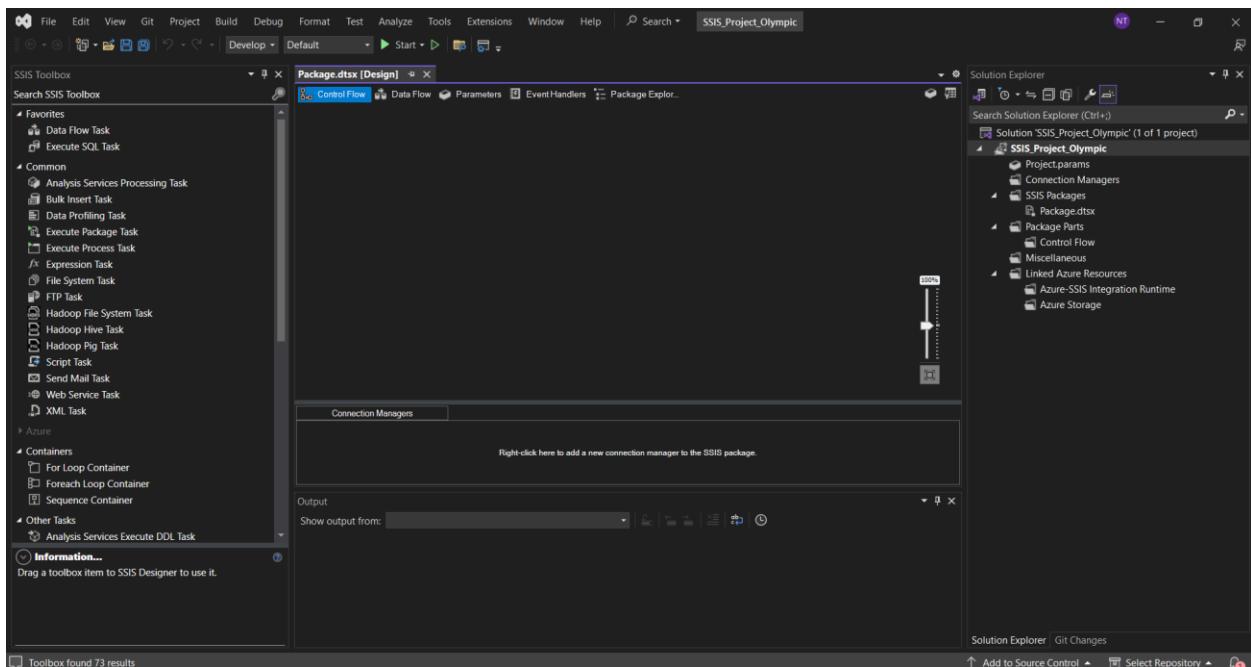


Hình 11 Chọn project



Hình 12 Đặt tên và chọn thư mục cho Project

- Bước 3: Giao diện của dự án SSIS xuất hiện, sau đó ta sử dụng các tính năng bên cửa sổ **SSIS Toolbox**.



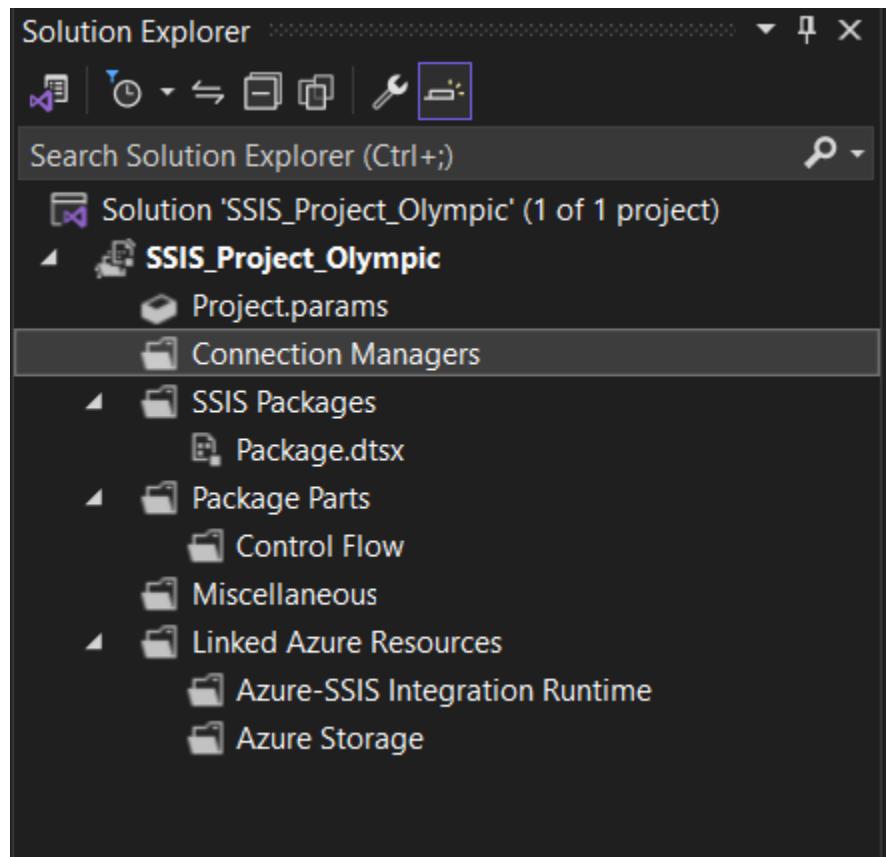
Hình 13 Màn hình chính của SSIS Project

2. TIẾN HÀNH THỰC HIỆN DỰ ÁN SSIS

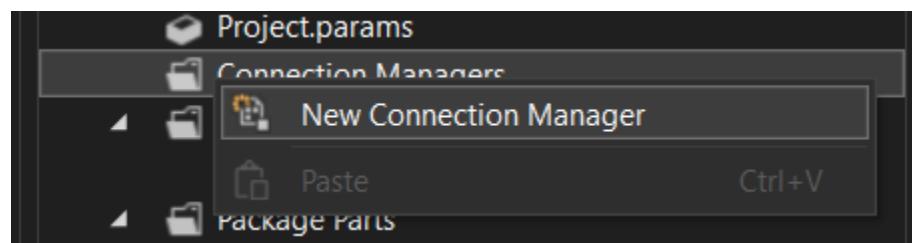
2.1. Thiết lập các kết nối

Để thiết lập kết nối cho dự án SSIS ta thực hiện các bước sau:

- Bước 1: Tại cửa sổ **Solution Explorer**, nhấp chuột phải vào **Connection Managers** và chọn **New Connection Manager**.

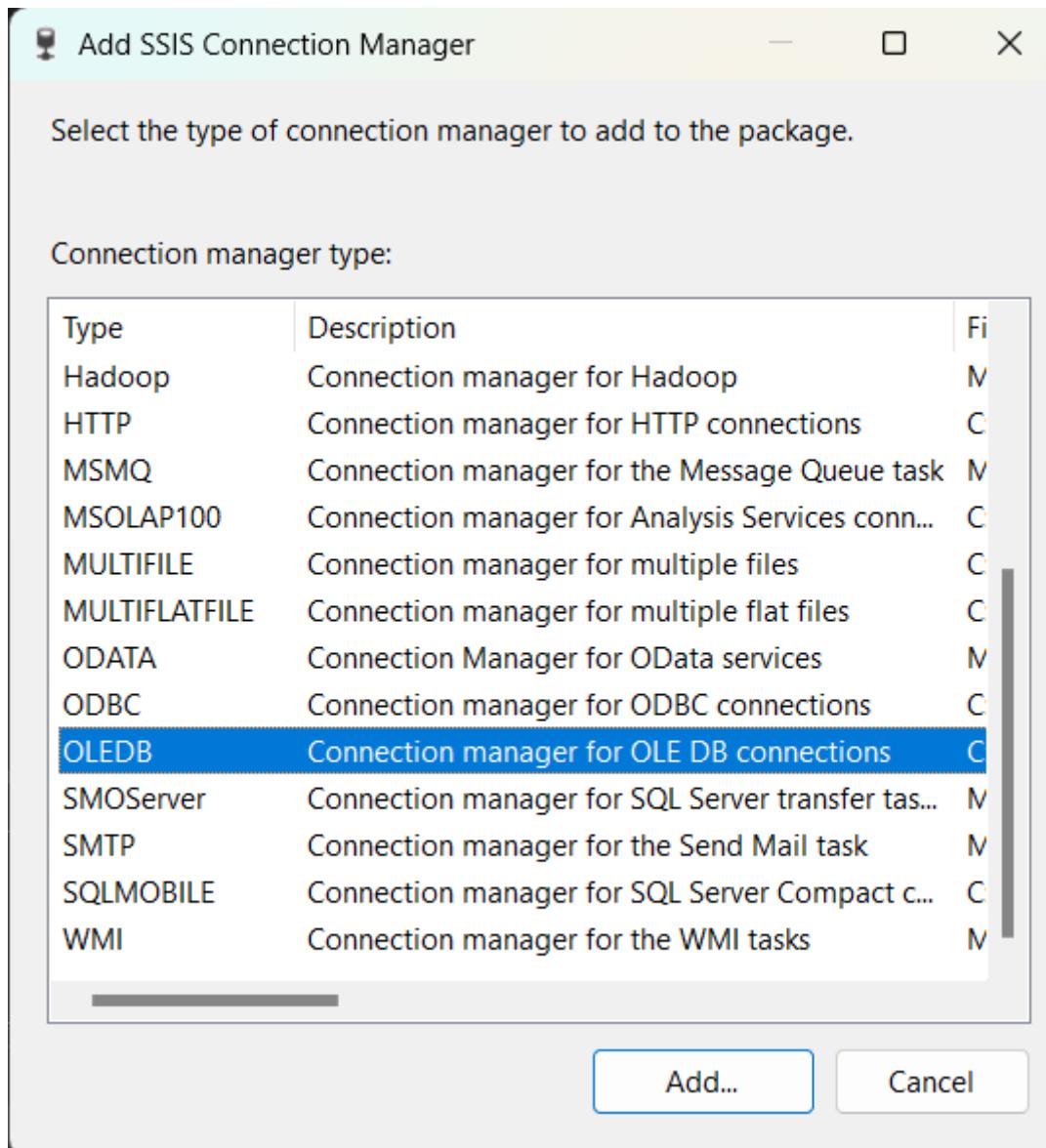


Hình 14 Chọn Connection Manager



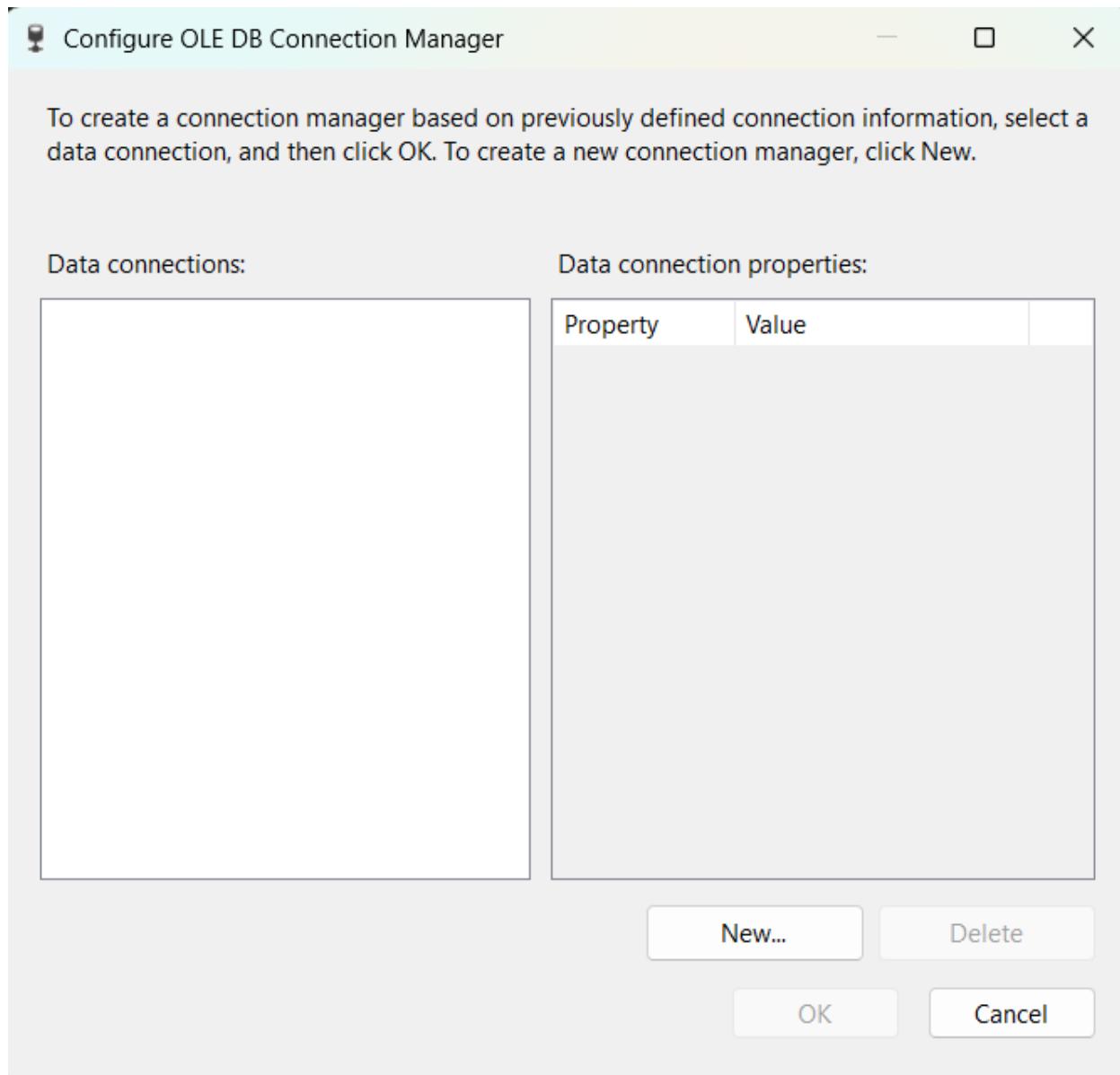
Hình 15 Tạo mới Connection Manager

- Bước 2: Chọn loại kết nối là OLEDB.



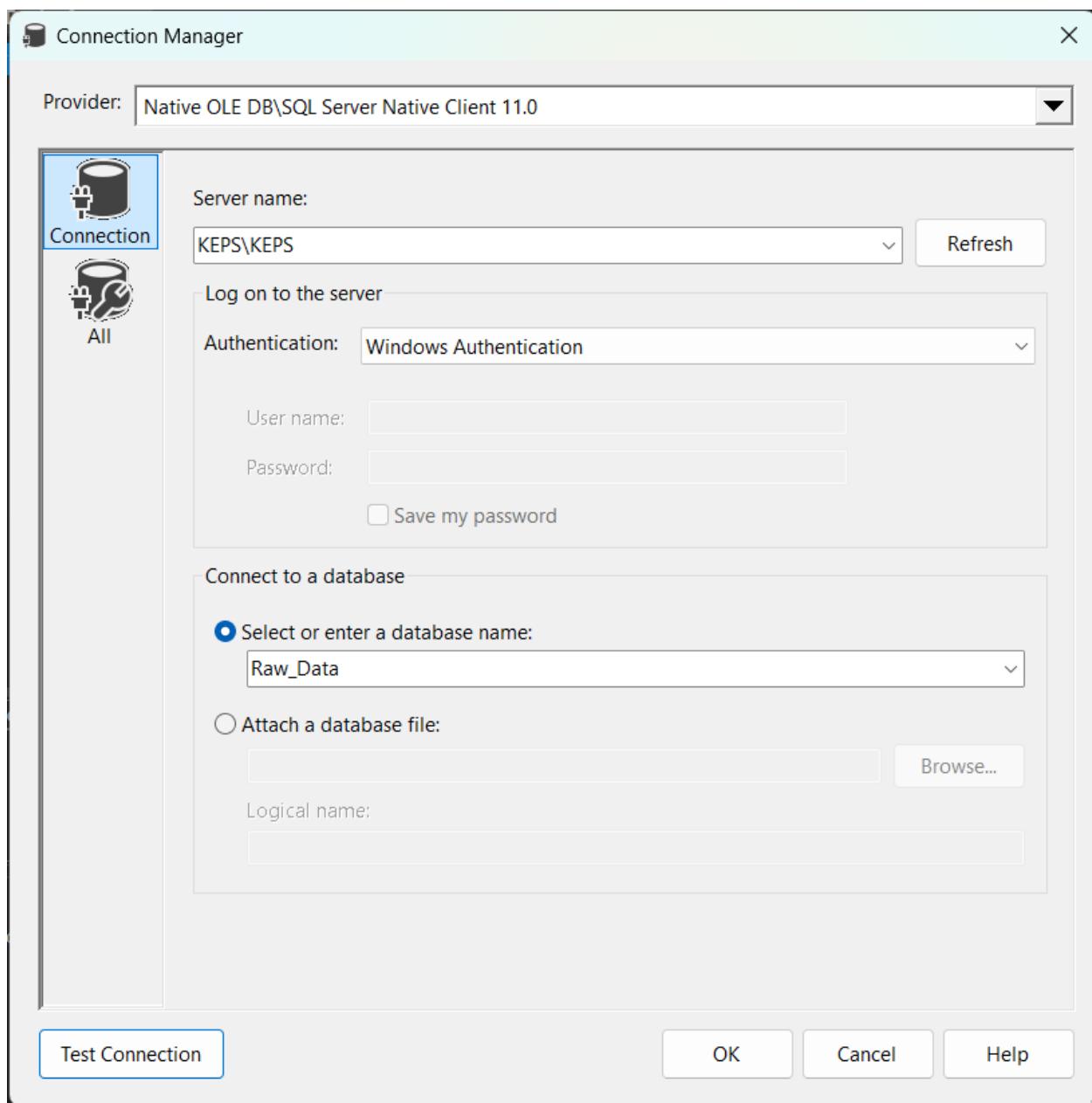
Hình 16 Chọn loại Connection Manager là OLEDB

- Bước 3: Tại cửa sổ **Configure OLE DB Connection Manager**, chúng ta không tìm thấy các cơ sở dữ liệu vừa tạo, vì vậy cần chọn **New** để tạo kết nối mới.

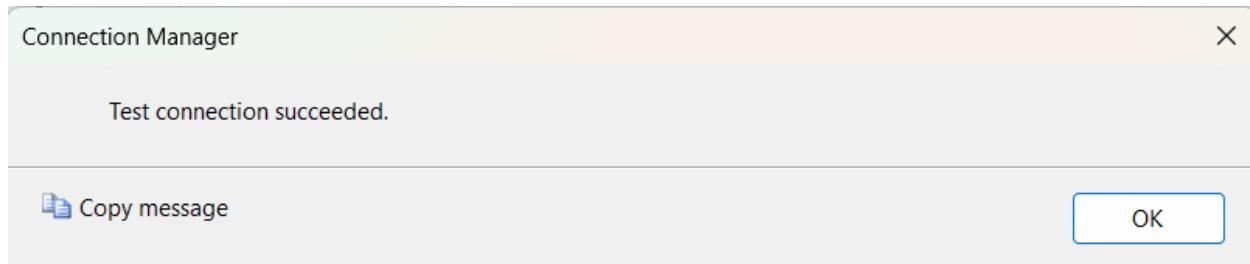


Hình 17 Tạo kết nối mới

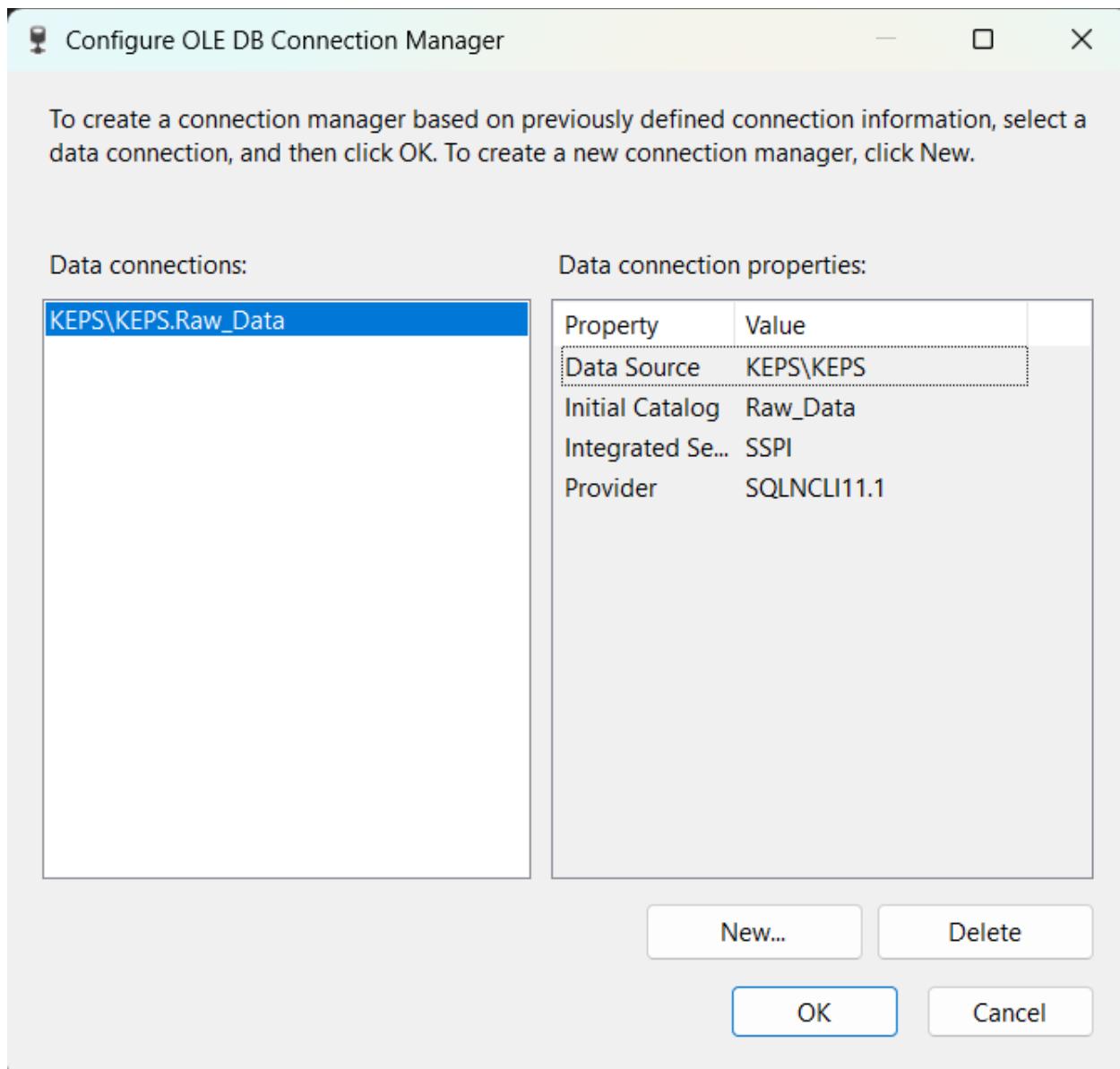
- Bước 4: Chọn Provider **SQL Server Native Client 11.0**, điền thông tin tên Server, chọn Database cần tạo kết nối, chọn **Test Connection** để kiểm tra kết nối. Sau khi kiểm tra thành công thì chọn **OK** để tạo kết nối.



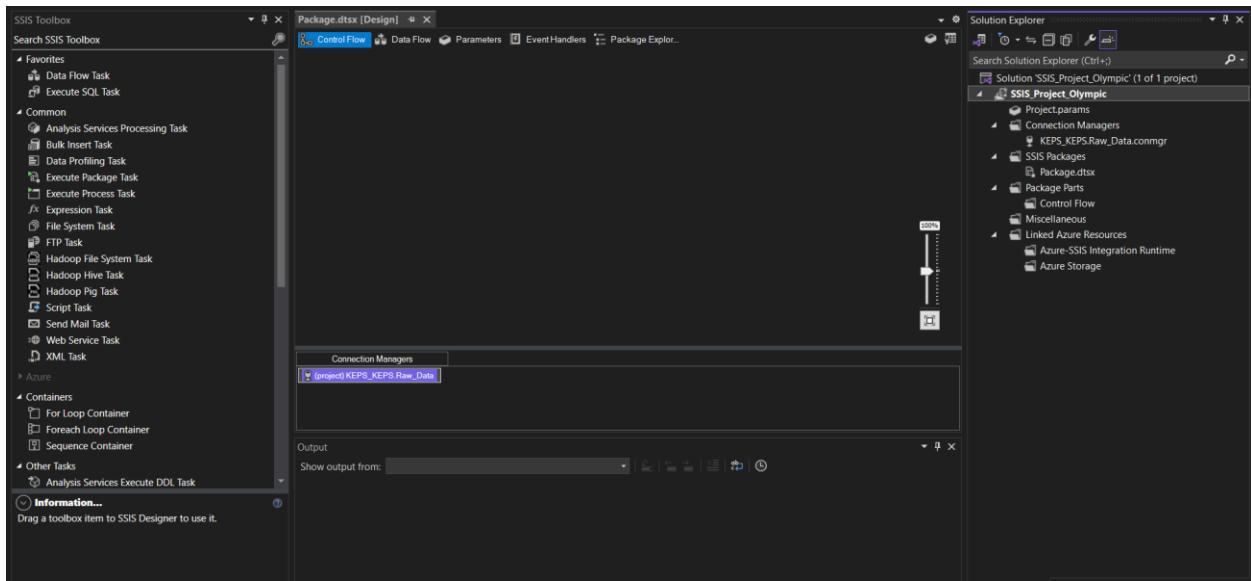
Hình 18 Chọn nguồn dữ liệu



Hình 19 Kiểm tra kết nối

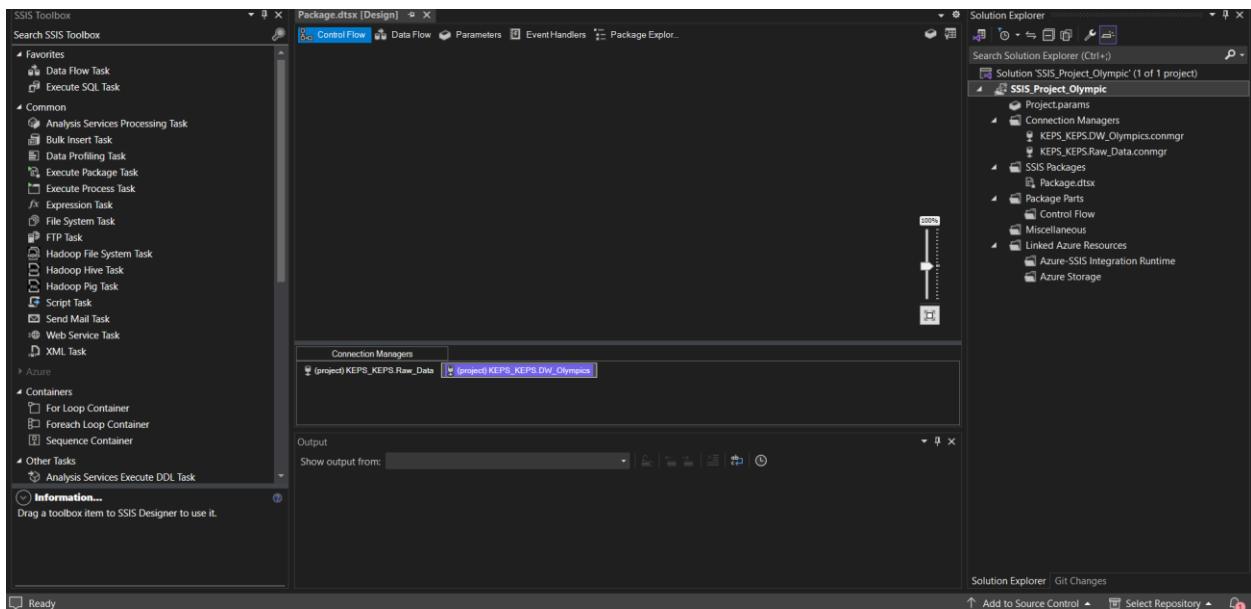


Hình 20 Chọn Data Connection



Hình 21 Tạo kết nối với Raw Data thành công

Tương tự, chúng ta thực hiện kết nối với cơ sở dữ liệu DW_Olympic

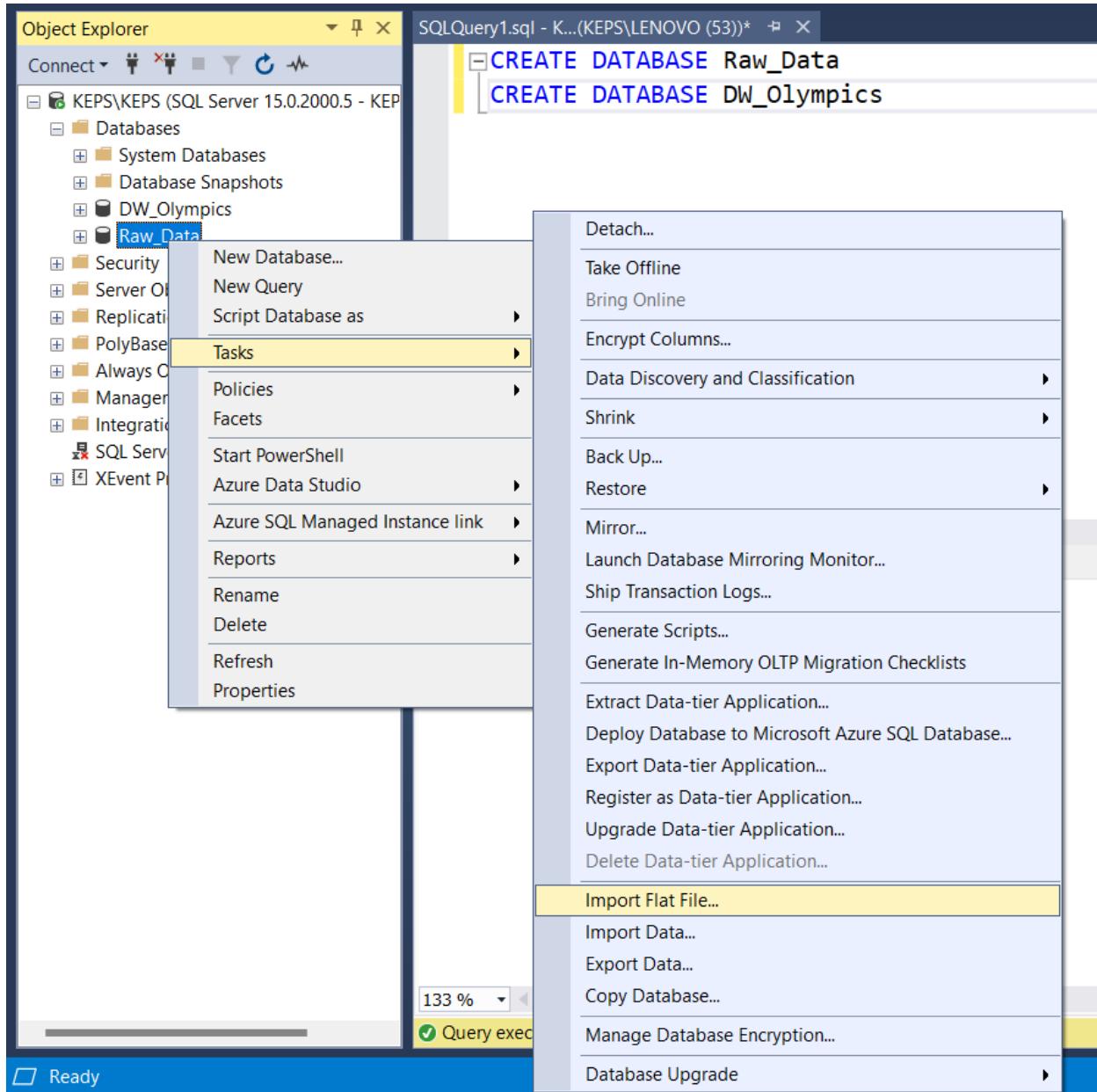


Hình 22 Tạo kết nối với DataWarehouse thành công

2.2. Đưa dữ liệu từ tệp csv vào SQL Server

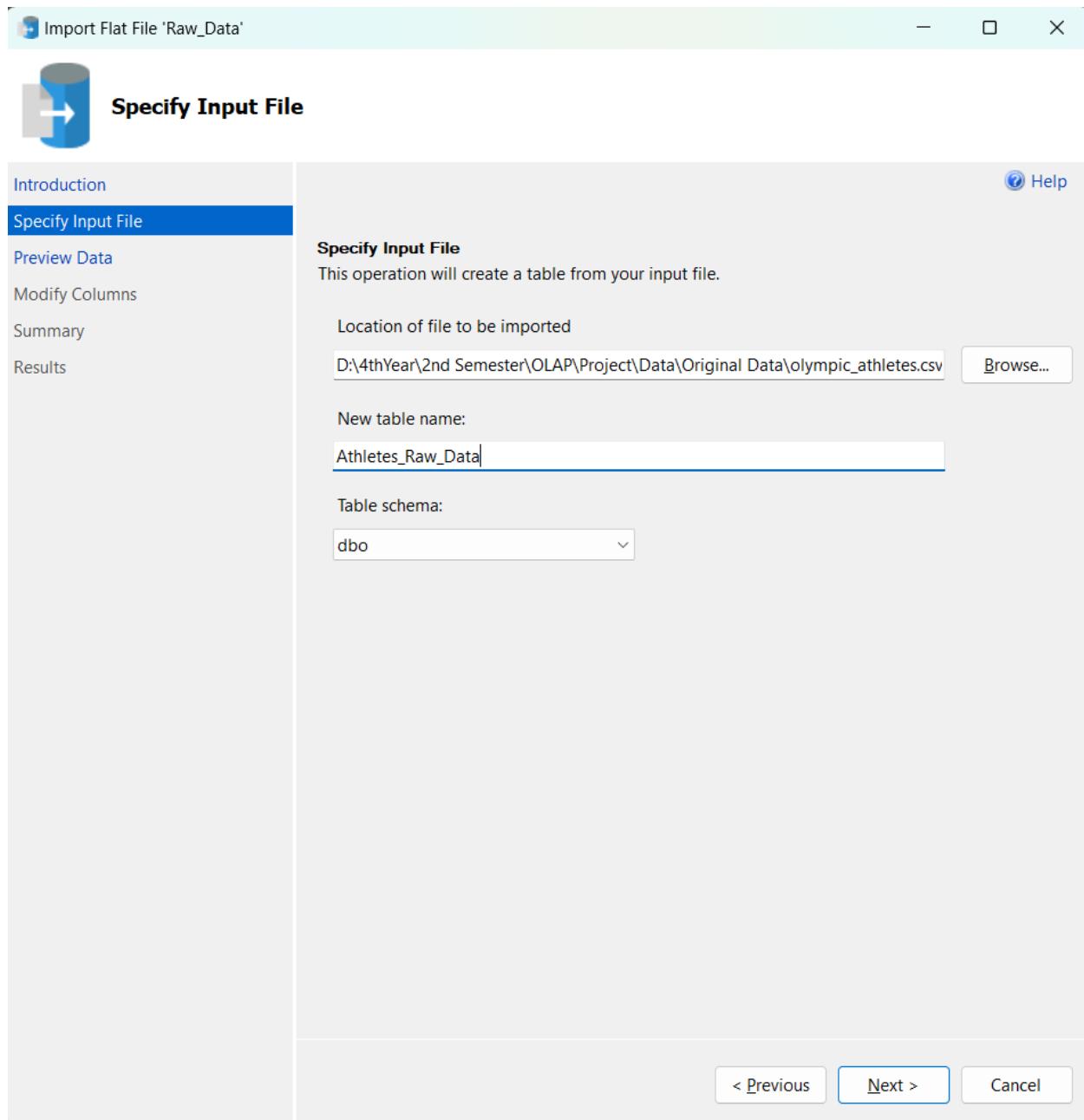
Để tiến hành đưa dữ liệu từ tệp csv vào hệ quản trị cơ sở dữ liệu SQL Server, ta thực hiện các bước sau:

- Bước 1: Tại cửa sổ **Object Explorer**, nhấp chuột phải tại Database **RawData** ở thư mục Databases, chọn **Tasks → Import Flat File**.



Hình 23 Thêm dữ liệu vào SQL Server

- Bước 2: Tại cửa sổ **Import Flat File ‘RawData’**, tại tab **Specify Input File**, chọn đường dẫn tệp có đuôi .csv, thực hiện đổi tên bảng thành ‘Athletes_Raw_Data’, sau đó chọn **Next**.



Hình 24 Chọn nguồn dữ liệu

- Bước 3: Tại tab **Preview Data**, kiểm tra dữ liệu vừa đưa vào, chọn vào ô *Use Rich Data Type Detection*, sau đó chọn **Next**.

Import Flat File 'Raw_Data'

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_participated	first_game	athlete_year_born	athlete_medals	bio
https://olympicdb.com/api/v1/athletes/1	Cooper WOODWARD	1	Beijing 2022	2000		
https://olympicdb.com/api/v1/athletes/2	Felix ELOFSEN	2	PyeongChang 2018	1995		
https://olympicdb.com/api/v1/athletes/3	Dylan WALCOTT	1	Beijing 2022	1993		
https://olympicdb.com/api/v1/athletes/4	Olli PENTTALA	1	Beijing 2022	1995		
https://olympicdb.com/api/v1/athletes/5	Dmitriy REINHOLD	1	Beijing 2022	1989		
https://olympicdb.com/api/v1/athletes/6	Matt GRAHAM	3	Sochi 2014	1994	1S	
https://olympicdb.com/api/v1/athletes/7	Ikuma HORIZUMI	2	PyeongChang 2018	1997	1B	
https://olympicdb.com/api/v1/athletes/8	Daichi HARA	2	PyeongChang 2018	1997	1B	
https://olympicdb.com/api/v1/athletes/9	Laurent DUPOUY	1	Beijing 2022	1996		
https://olympicdb.com/api/v1/athletes/10	James MATTHEWS	2	PyeongChang 2018	1995		
https://olympicdb.com/api/v1/athletes/11	Pavel KOLMACHEV	3	Sochi 2014	1996		
https://olympicdb.com/api/v1/athletes/12	Kosuke SUGAWARA	1	Beijing 2022	1994		
https://olympicdb.com/api/v1/athletes/13	Brodie SUMMERS	3	Sochi 2014	1993		
https://olympicdb.com/api/v1/athletes/14	Severi VIERI	1	Beijing 2022	2001		
https://olympicdb.com/api/v1/athletes/15	Marco TADE	1	Beijing 2022	1995		
https://olympicdb.com/api/v1/athletes/16	William FENSTERWALD	1	Beijing 2022	1999		
https://olympicdb.com/api/v1/athletes/17	Mikael KINNUNEN	3	Sochi 2014	1992	1G2S	In training
https://olympicdb.com/api/v1/athletes/18	So MATSUDA	1	Beijing 2022	1999		
https://olympicdb.com/api/v1/athletes/19	Walter WALLACE	2	PyeongChang 2018	2000	1G	
https://olympicdb.com/api/v1/athletes/20	Bradley WILLIAMS	3	Sochi 2014	1992		

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

< Previous Next > Cancel

Hình 25 Xem trước dữ liệu

- Bước 4: Tại tab **Modify Columns**, kiểm tra các kiểu dữ liệu và thay đổi nếu có, nếu không thì bỏ qua và chọn **Next**.

Modify Columns

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

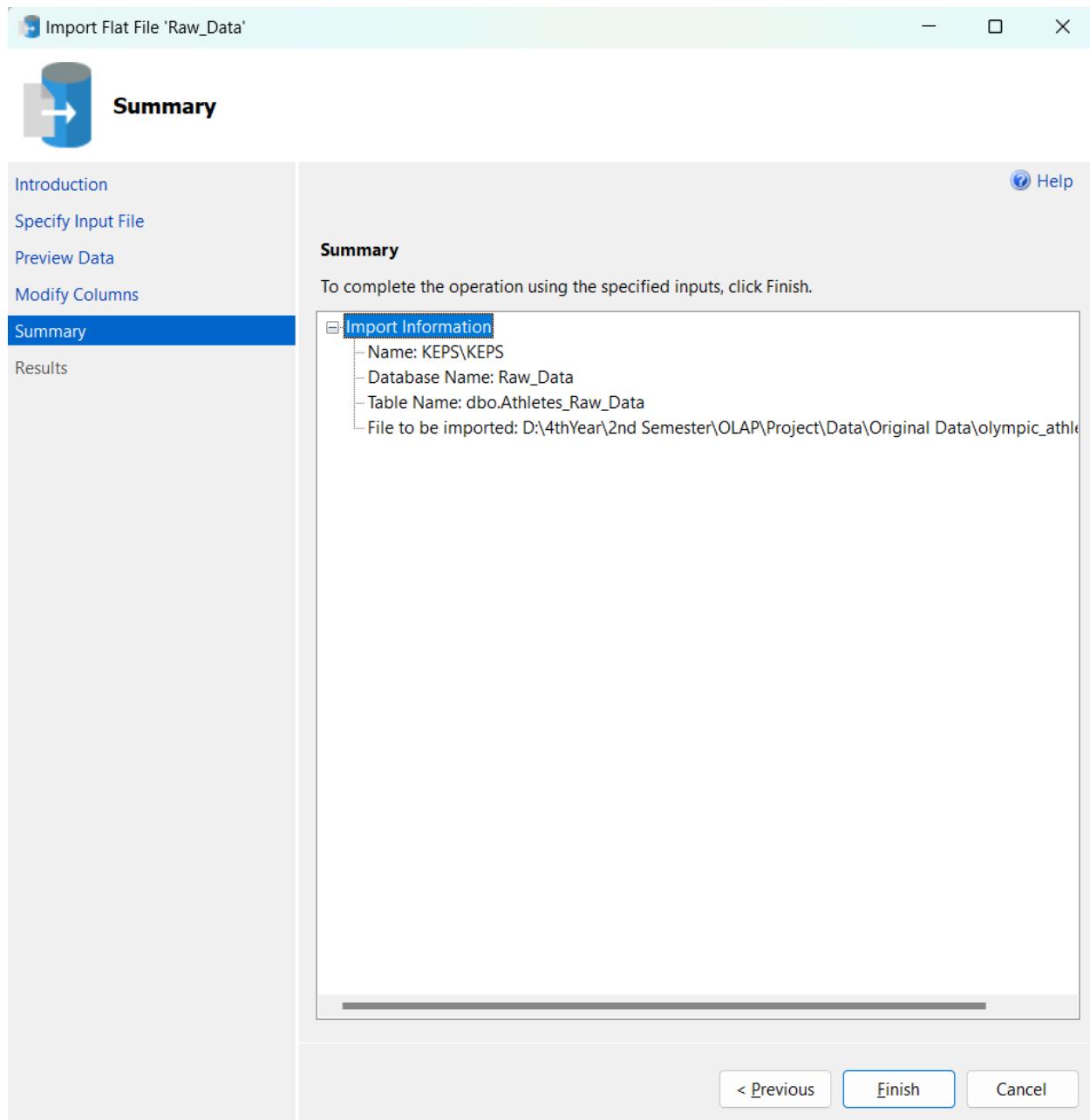
Column Name	Data Type	Primary Key	Allow Nulls
athlete_url	nvarchar(100)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
athlete_full_name	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
games_participations	tinyint	<input type="checkbox"/>	<input checked="" type="checkbox"/>
first_game	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
athlete_year_birth	smallint	<input type="checkbox"/>	<input checked="" type="checkbox"/>
athlete_medals	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
bio	nvarchar(max)	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Row granularity of error reporting (performance impact with smaller ranges)

< Previous Cancel

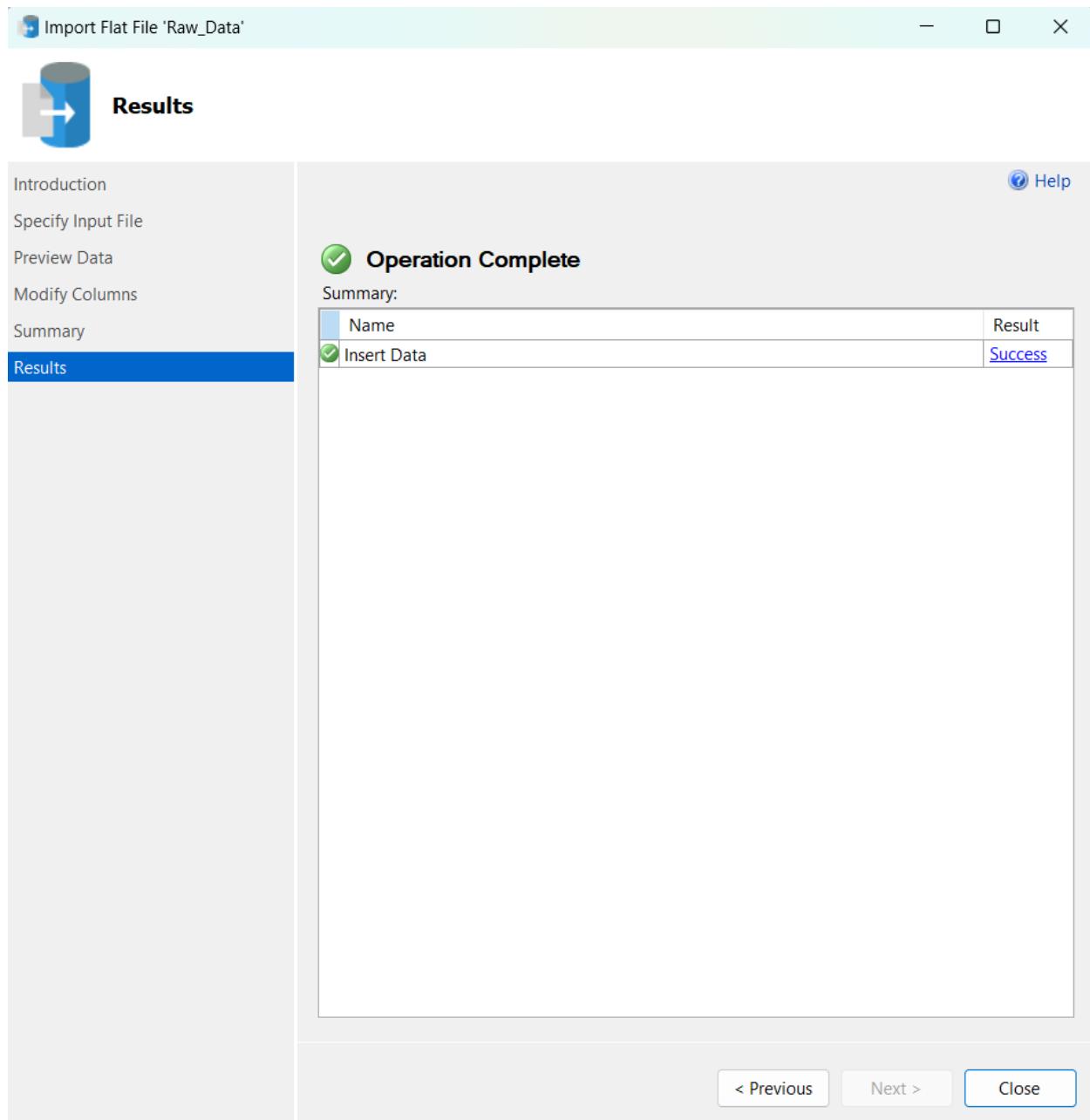
Hình 26 Điều chỉnh các cột

- Bước 5: Chọn **Finish** để bắt đầu thực hiện việc đưa dữ liệu vào SQL Server.



Hình 27 Tổng quan quá trình thêm dữ liệu

- Bước 6: Nếu hiển thị ra giao diện bên dưới thì việc đưa dữ liệu vào SQL Server **thành công**, ngược lại thì chúng ta chọn vào **Error** ở cột **Result** để kiểm tra lỗi và tiến hành sửa lỗi.



Hình 28 Thêm dữ liệu thành công

Thực hiện việc đưa dữ liệu từ các tệp phẳng còn lại vào SQL Server tương tự như các bước trên. Ta được kết quả truy vấn của cơ sở dữ liệu RawData như các hình bên dưới

The screenshot shows the Object Explorer on the left with the connection to KEPS\KEPS (SQL Server 15.0.2000.5 - KEPS). The SQL Query window contains the following code:

```
SELECT * FROM [dbo].Athletes_Raw_Data
```

The results grid displays 53 rows of athlete data, including columns like athlete_id, athlete_full_name, games_participations, first_name, athlete_year_birth, athlete_medals, and bio. The bio column contains URLs to Wikipedia pages for each athlete.

At the bottom of the results grid, it says "Query executed successfully." and shows the status bar with "KEPS\KEPS (15.0 RTM) | KEPS\LENOVO (53) | Raw_Data | 00:00:00 | 75,904 rows".

Hình 29 Thêm dữ liệu Athletes thành công

The screenshot shows the Object Explorer on the left with the connection to KEPS\KEPS (SQL Server 15.0.2000.5 - KEPS). The SQL Query window contains the following code:

```
SELECT * FROM [dbo].Athletes_Raw_Data
SELECT * FROM [dbo].Hosts_Raw_Data
```

The results grid displays 53 rows of host data, including columns like game_slug, game_end_date, game_start_date, game_location, game_name, game_season, and game_year. The game_name column contains the names of the host cities and years.

At the bottom of the results grid, it says "Query executed successfully." and shows the status bar with "KEPS\KEPS (15.0 RTM) | KEPS\LENOVO (53) | Raw_Data | 00:00:00 | 53 rows".

Hình 30 Thêm dữ liệu Hosts thành công

The screenshot shows the Object Explorer on the left with the database 'KEPS\KEPS' selected. The SQL Query window on the right contains the following T-SQL code:

```

SELECT * FROM [dbo].Athletes_Raw_Data
SELECT * FROM [dbo].Hosts_Raw_Data
SELECT * FROM [dbo].Medals_Raw_Data

```

The results grid displays 21,697 rows of data from the Medals Raw Data table, listing various athletes, their disciplines, events, genders, medal types, and countries.

Hình 31 Thêm dữ liệu Medals thành công

The screenshot shows the Object Explorer on the left with the database 'KEPS\KEPS' selected. The SQL Query window on the right contains the following T-SQL code:

```

SELECT * FROM [dbo].Athletes_Raw_Data
SELECT * FROM [dbo].Hosts_Raw_Data
SELECT * FROM [dbo].Medals_Raw_Data
SELECT * FROM [dbo].Results_Raw_Data

```

The results grid displays 162,804 rows of data from the Results Raw Data table, listing athletes, their disciplines, events, and results.

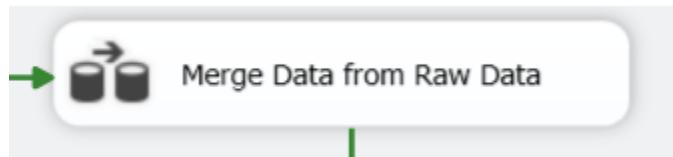
Hình 32 Thêm dữ liệu Results thành công

2.3. Quá trình nối dữ liệu từ các bảng

Sau khi đưa dữ liệu từ tệp csv vào cơ sở dữ liệu, ta cần thực hiện việc nối (merge) các dữ liệu từ các bảng một cách tự động, quá trình thực hiện như sau:

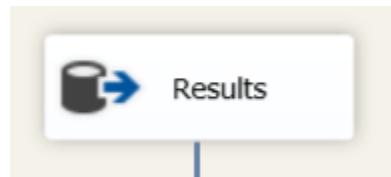
- Bước 1: Trong project SSIS của Visual Studio, tại cửa sổ **Control Flow**, ta tạo một **Data Flow Task** mới bằng cách kéo từ **SSIS Toolbox** vào cửa sổ làm việc, sau đó

nhấn chuột phải tại Task đó và chọn **Rename** để đổi tên thành quá trình nối dữ liệu *Merge Data from Raw Data*.



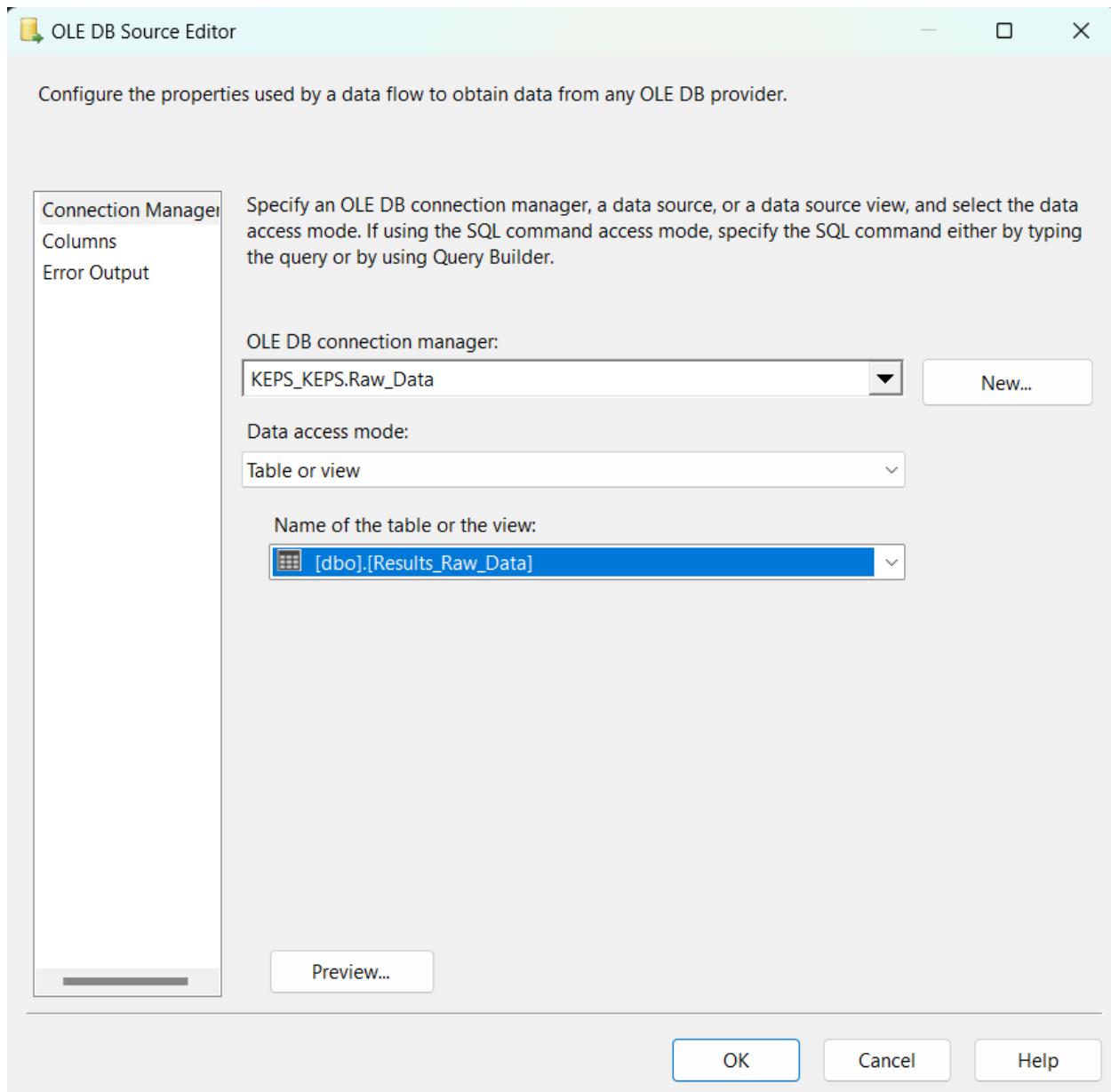
Hình 33 Tạo data flow task

- Bước 2: Chuyển qua cửa sổ **Data Flow**, chọn **Data Flow Task** “*Merge Data from Raw Data*”. Tại cửa sổ SSIS Toolbox, chọn nguồn dữ liệu là **OLE DB Source** và kéo vào cửa sổ làm việc, sau đó chọn Rename và đổi tên thành “*Results*”.



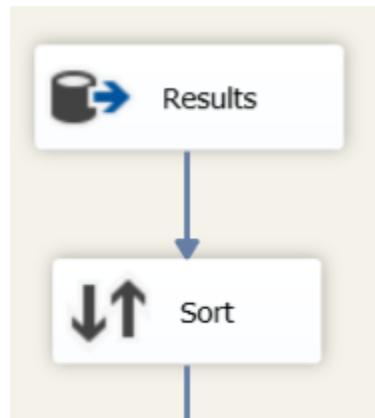
Hình 34 Tạo data Source

- Bước 3: Nhấp đúp chuột vào nguồn dữ liệu là “*Results*”, tạo một **Connection Manager** mới (nếu đã tạo ở trên thì bỏ qua), chọn bảng nguồn cần lấy, ta có thể xem trước bằng cách chọn **Preview**, nếu cần loại bỏ cột thì chọn tab **Columns** và bỏ chọn các cột không cần thiết, chọn **OK** để hoàn tất.



Hình 35 Chọn nguồn dữ liệu

- Bước 4: Tiến hành sắp xếp các dữ liệu, chọn sự kiện **Sort** tại Toolbox, nối **Results** vào **Sort**



Hình 36 Nối Result với Sort

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 Thr...	
<input checked="" type="checkbox"/> discipline_title	<input type="button" value="..."/>	
<input checked="" type="checkbox"/> event_title	<input type="button" value="..."/>	
<input checked="" type="checkbox"/> slug_game	<input type="button" value="..."/>	

Input Column	Output Alias	Sort Type	Sort Order	Con
discipline_title	discipline_title	ascending	4	
event_title	event_title	ascending	5	
slug_game	slug_game	ascending	3	
participant_type	participant_type	ascending	7	
medal_type	medal_type	ascending	6	
rank_equal	rank_equal	ascending	8	
rank_position	rank_position	ascending	9	
country_name	country_name	ascending	10	
country_code	country_code	ascending	11	
country_3_letter_code	country_3_letter_code	ascending	12	
athlete_url	athlete_url	ascending	1	
athlete_full_name	athlete_full_name	ascending	2	

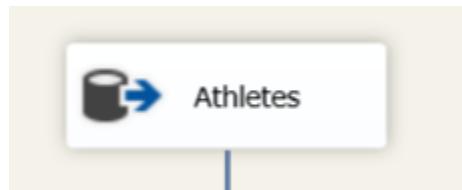
Remove rows with duplicate sort values

OK **Cancel** **Help**

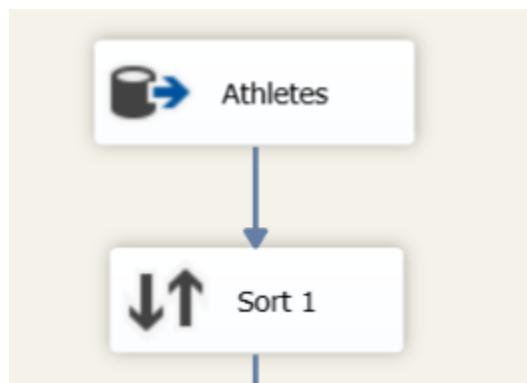
Hình 37 Lựa chọn thứ tự sort các thuộc tính

→ Thực hiện các bước trên tương tự với các dữ liệu **Athletes**, **Hosts**.

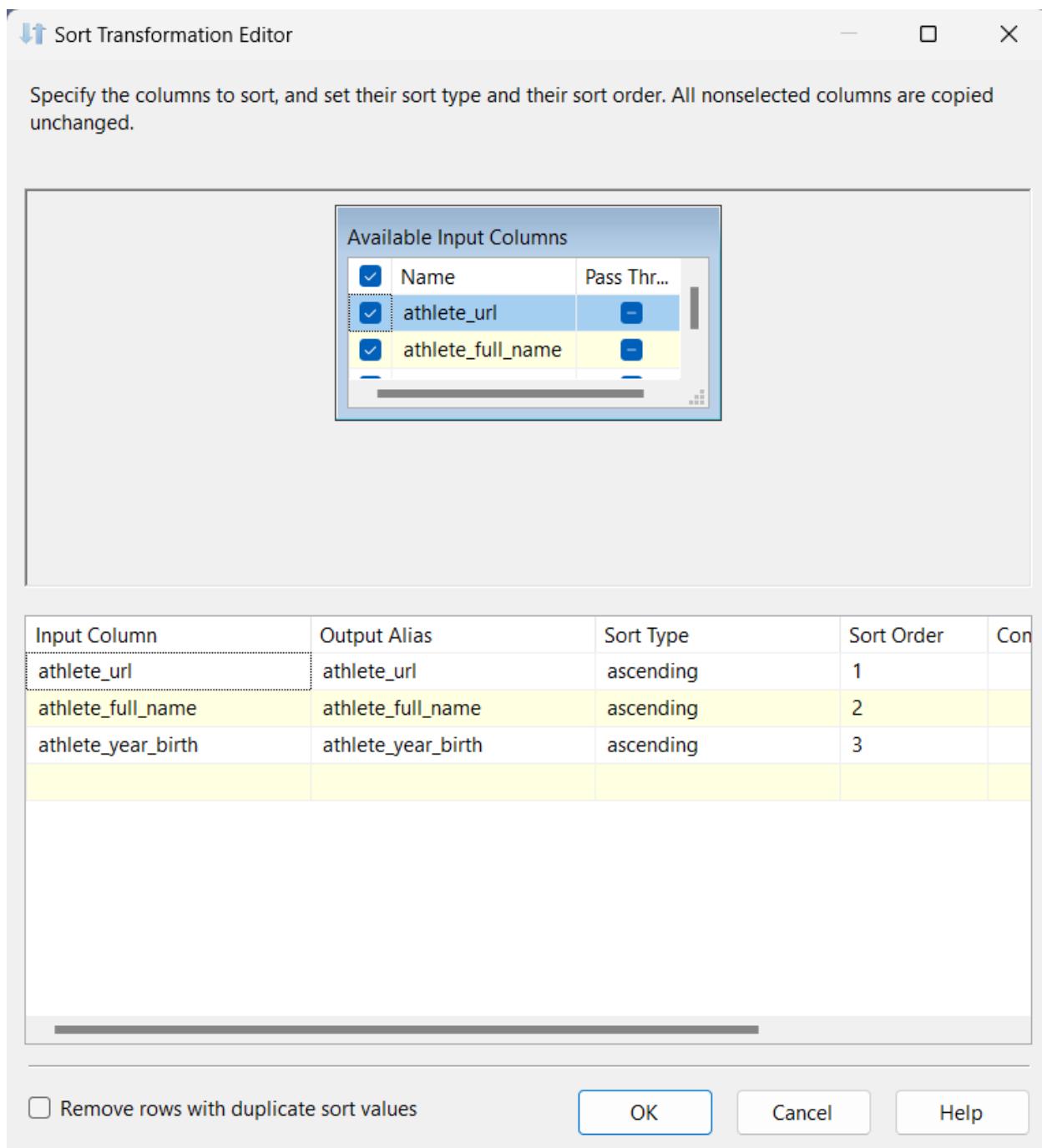
Athletes:



Hình 38 Data source Athletes

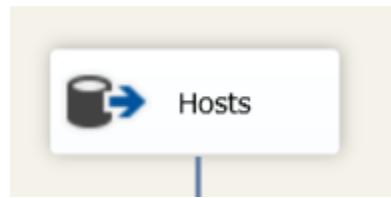


Hình 39 Sort Athletes



Hình 40 Lựa chọn thứ tự để sắp xếp các thuộc tính

Hosts:



Hình 41 Data source Hosts



Hình 42 Nối Host với Sort

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		
<input checked="" type="checkbox"/>	Name	Pass Thr...
<input checked="" type="checkbox"/>	game_slug	—
<input checked="" type="checkbox"/>	game_start_date	—
<input checked="" type="checkbox"/>	game_location	—
<input checked="" type="checkbox"/>	game_name	—
<input checked="" type="checkbox"/>	game_season	—

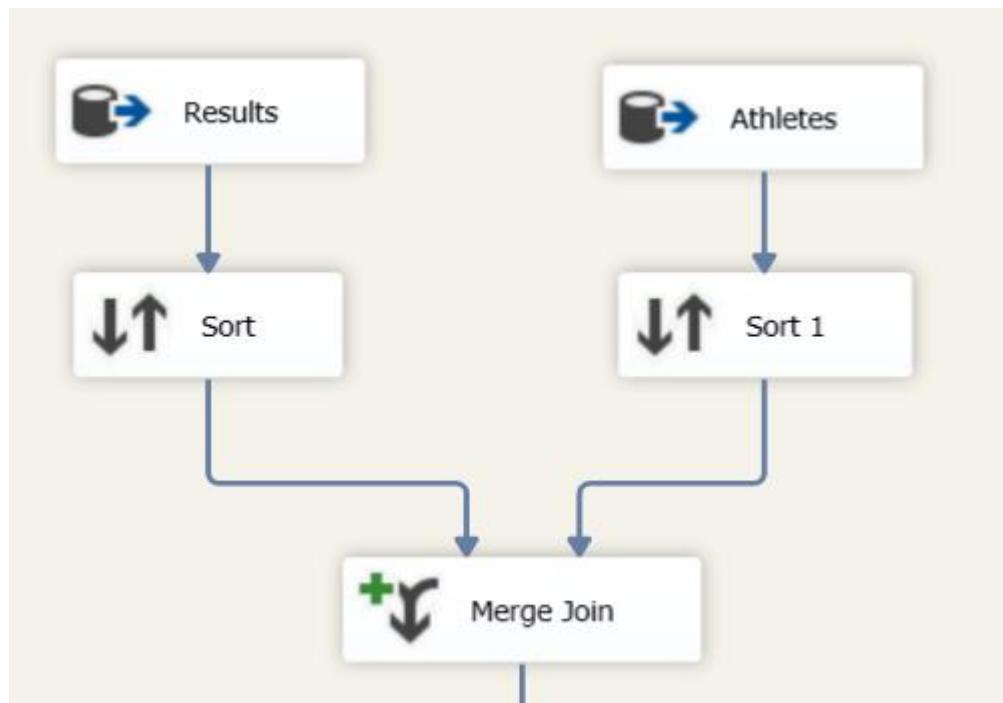
Input Column	Output Alias	Sort Type	Sort Order	Con
game_slug	game_slug	ascending	1	
game_start_date	game_start_date	ascending	2	
game_location	game_location	ascending	3	
game_name	game_name	ascending	4	
game_season	game_season	ascending	5	
game_year	game_year	ascending	6	

Remove rows with duplicate sort values

OK Cancel Help

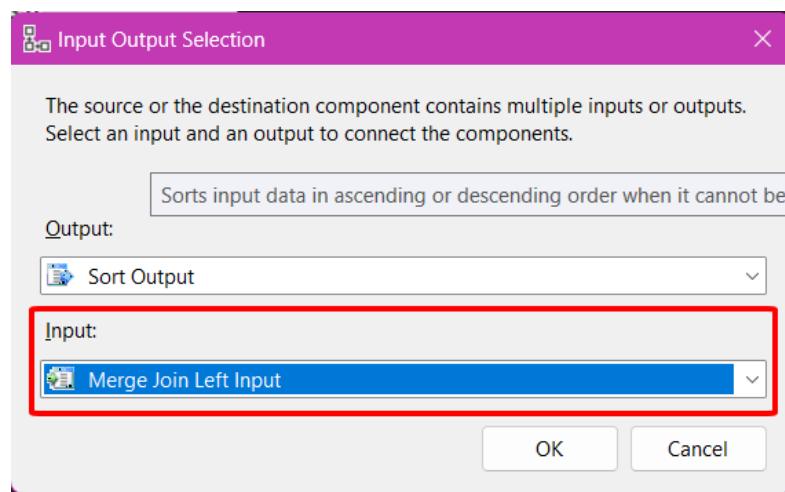
Hình 43 Lựa chọn các thuộc tính để sắp xếp

- Bước 6: Thực hiện kết nối dữ liệu của các Data source bằng cách sử dụng **Merge Join**



Hình 44 Merge 2 nguồn dữ liệu

Khi nối vào Merge Join sẽ hiển thị ra hộp thoại bên dưới, ta cần chọn kiểu đầu vào là **Merge Join Left** hay **Merge Join Right**, ở đây ta chọn **Merge Join Left** và chọn **OK**



Hình 45 Chọn Merge Join Left input

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 a

Join type: Inner join

The screenshot shows the 'Merge Join Transformation Editor' interface. At the top, it says 'Merge Join Transformation Editor' and 'Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used a'. Below this, the 'Join type:' is set to 'Inner join'. There are two 'Sort' boxes on the left and right sides, each containing a table with columns 'Name', 'Ord...', and 'Join...'. Arrows point from the 'Sort' boxes to the 'Input Column' column in the main mapping grid below. The 'Sort' boxes contain the following data:

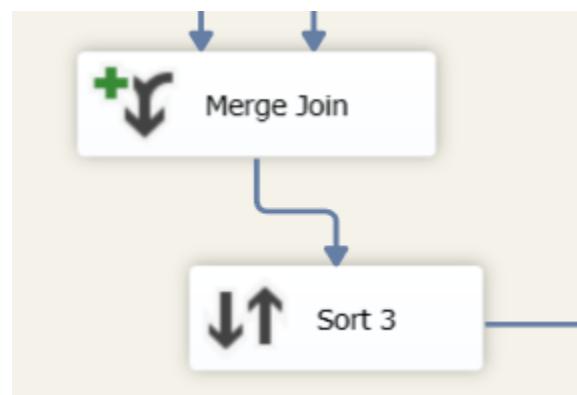
Name	Ord...	Join...
discipline_title	4	<input type="checkbox"/>
event_title	5	<input type="checkbox"/>
slug_game	3	<input type="checkbox"/>
participant_type	7	<input type="checkbox"/>
medal_type	6	<input type="checkbox"/>
athletes	0	<input type="checkbox"/>

Name	Ord...	Join...
athlete_url	1	<input checked="" type="checkbox"/>
athlete_full_name	2	<input checked="" type="checkbox"/>

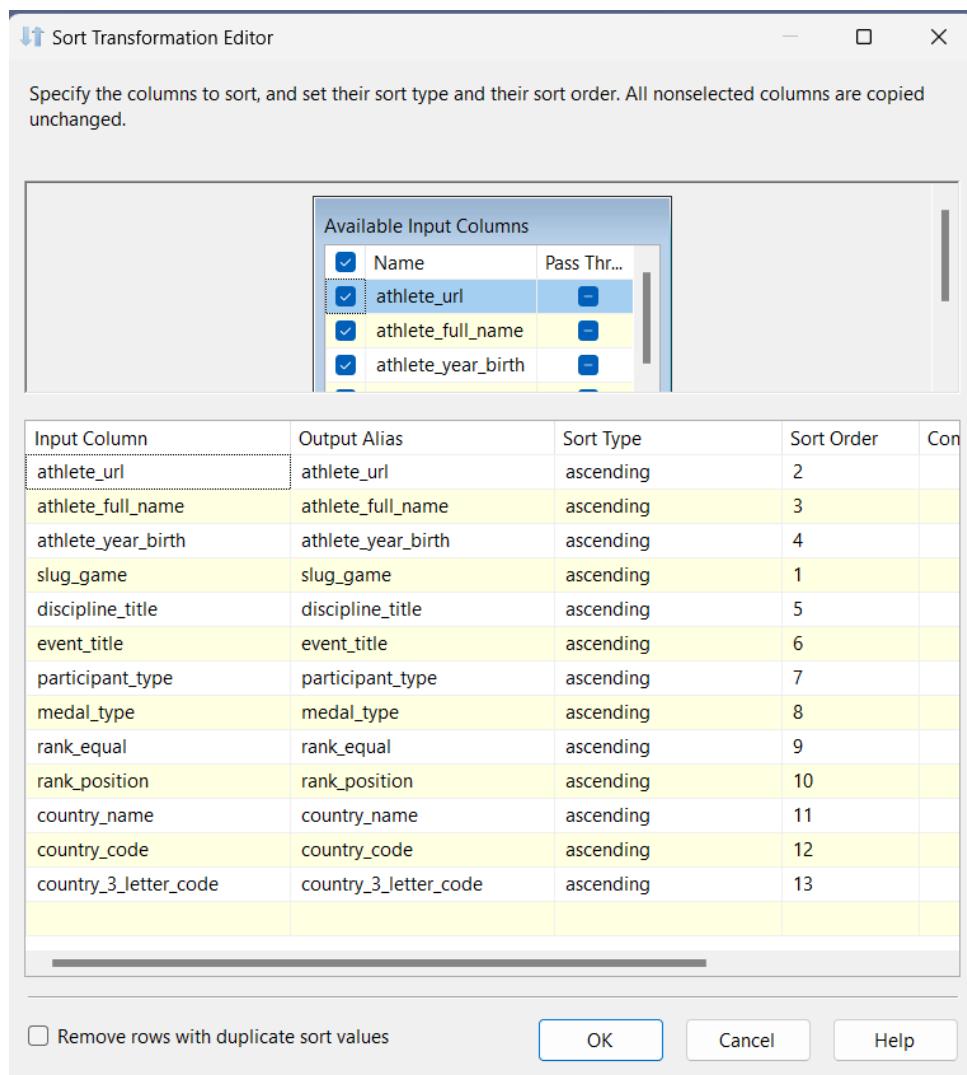
The main mapping grid below has three columns: 'Input', 'Input Column', and 'Output Alias'. It lists 14 rows, each mapping a source column to an output alias. The rows are:

Input	Input Column	Output Alias
Sort	athlete_url	athlete_url
Sort	athlete_full_name	athlete_full_name
Sort 1	athlete_year_birth	athlete_year_birth
Sort	slug_game	slug_game
Sort	discipline_title	discipline_title
Sort	event_title	event_title
Sort	participant_type	participant_type
Sort	medal_type	medal_type
Sort	rank_equal	rank_equal
Sort	rank_position	rank_position
Sort	country_name	country_name
Sort	country_code	country_code
Sort	country_3_letter_code	country_3_letter_code

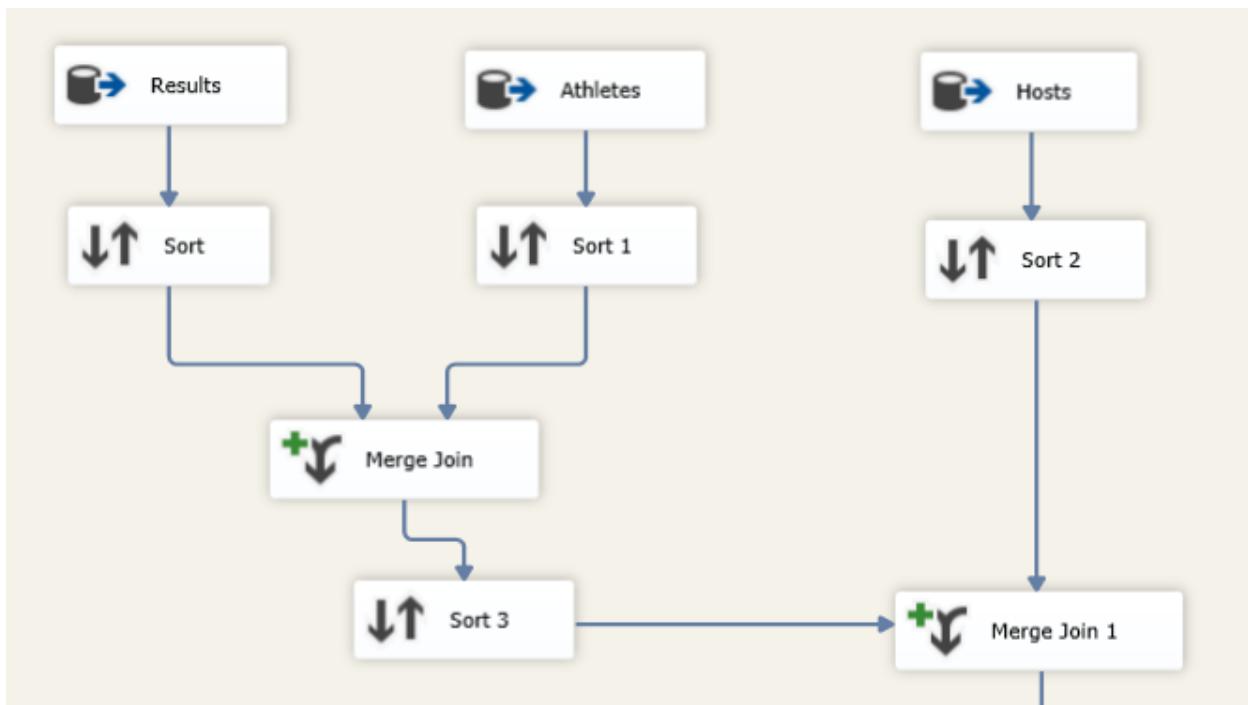
Hình 46 Nối các thuộc tính của 2 nguồn dữ liệu



Hình 47 Sắp xếp lại sau khi nối



Hình 48 Lựa chọn thứ tự để sắp xếp các thuộc tính



Hình 49 Thực hiện nối với nguồn dữ liệu Hosts

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.

Join type: Inner join

The screenshot shows the 'Merge Join Transformation Editor' interface. It consists of two 'Sort' tables and a 'Mapping' table below them.

Sort 3 (Left):

Name	Ord...	Join...
athlete_url	2	<input type="checkbox"/>
athlete_full_name	3	<input type="checkbox"/>
athlete_year_birth	4	<input type="checkbox"/>
slug_game	1	<input checked="" type="checkbox"/>
discipline_title	5	<input type="checkbox"/>
event_title	6	<input type="checkbox"/>

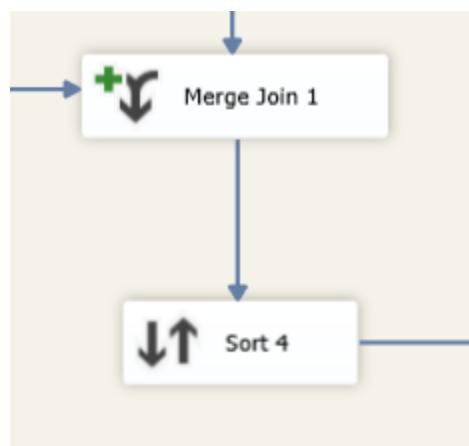
Sort 2 (Right):

Name	Ord...	Join...
game_slug	1	<input checked="" type="checkbox"/>
game_start_date	2	<input type="checkbox"/>
game_location	3	<input type="checkbox"/>
game_name	4	<input type="checkbox"/>
game_season	5	<input type="checkbox"/>

Mapping Table:

Input	Input Column	Output Alias
Sort 2	game_slug	game_slug
Sort 3	athlete_full_name	athlete_full_name
Sort 3	athlete_year_birth	athlete_year_birth
Sort 3	discipline_title	discipline_title
Sort 3	event_title	event_title
Sort 3	participant_type	participant_type
Sort 3	medal_type	medal_type
Sort 3	rank_equal	rank_equal
Sort 3	rank_position	rank_position
Sort 3	country_name	country_name
Sort 3	country_code	country_code
Sort 3	country_3_letter_code	country_3_letter_code
Sort 2	game_start_date	game_start_date
Sort 2	game_location	game_location
Sort 2	game_name	game_name
Sort 2	game_season	game_season
Sort 2	game_year	game_year

Hình 50 Lựa chọn các thuộc tính để sắp xếp



Hình 51 Thực hiện sắp xếp sau khi nối

Sort Transformation Editor

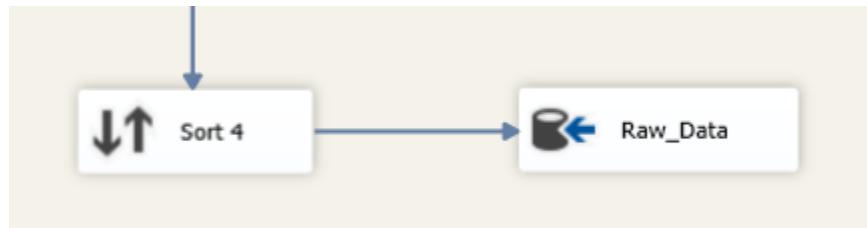
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Available Input Columns				
	Name	Pass Thr...		
<input checked="" type="checkbox"/>	game_slug	<input type="button" value="..."/>		
<input checked="" type="checkbox"/>	athlete_full_name	<input type="button" value="..."/>		
<input checked="" type="checkbox"/>	athlete_year_birth	<input type="button" value="..."/>		

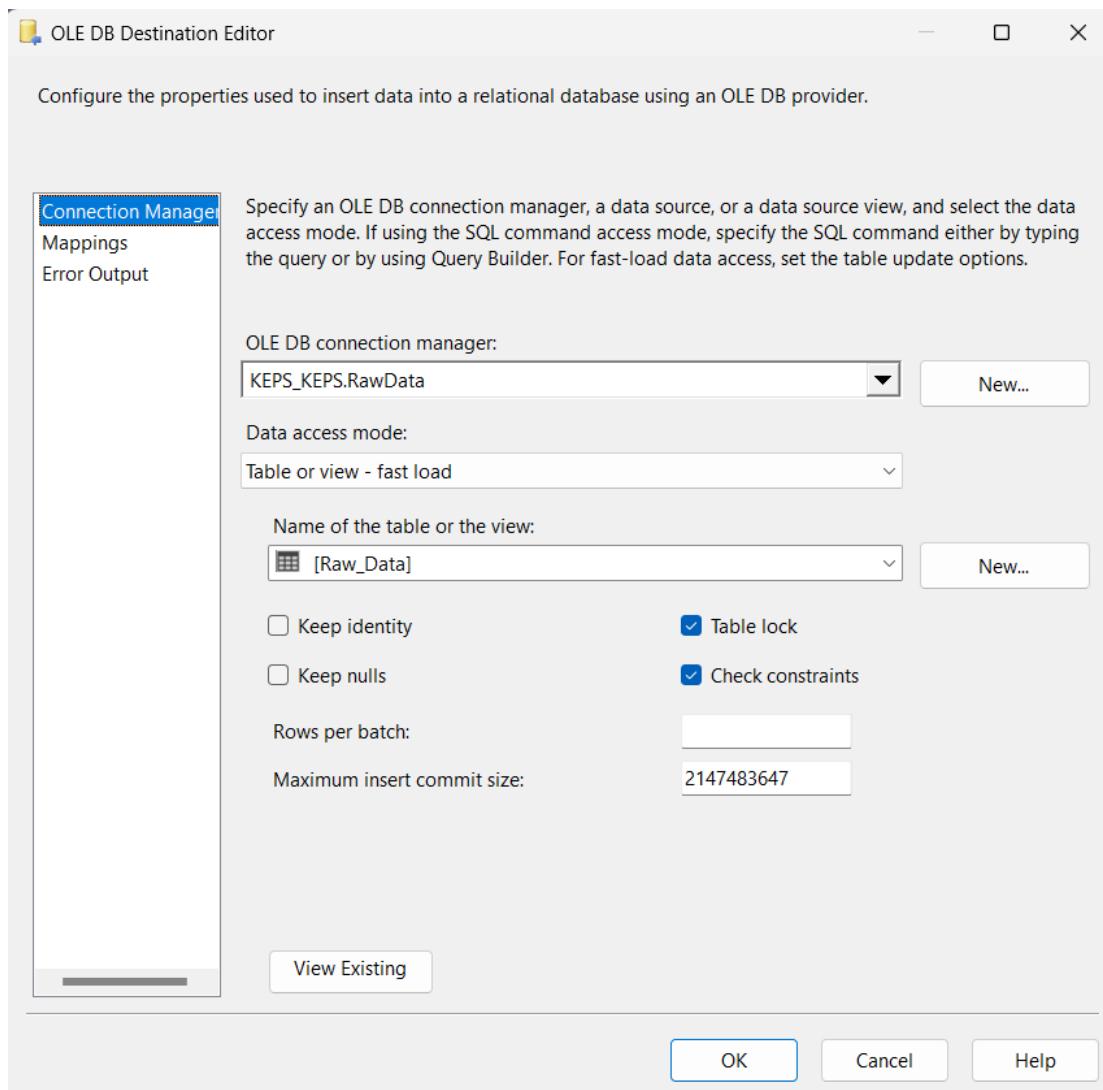
Input Column	Output Alias	Sort Type	Sort Order	Comparison Flag
game_slug	game_slug	ascending	1	
athlete_full_name	athlete_full_name	ascending	2	
athlete_year_birth	athlete_year_birth	ascending	3	
discipline_title	discipline_title	ascending	4	
event_title	event_title	ascending	5	
participant_type	participant_type	ascending	6	
medal_type	medal_type	ascending	7	
rank_equal	rank_equal	ascending	8	
rank_position	rank_position	ascending	9	
country_name	country_name	ascending	10	
country_code	country_code	ascending	11	
country_3_letter_code	country_3_letter_code	ascending	12	
game_start_date	game_start_date	ascending	13	
game_location	game_location	ascending	14	
game_name	game_name	ascending	15	
game_season	game_season	ascending	16	
game_year	game_year	ascending	17	

Hình 52 Lựa chọn các thuộc tính để sắp xếp

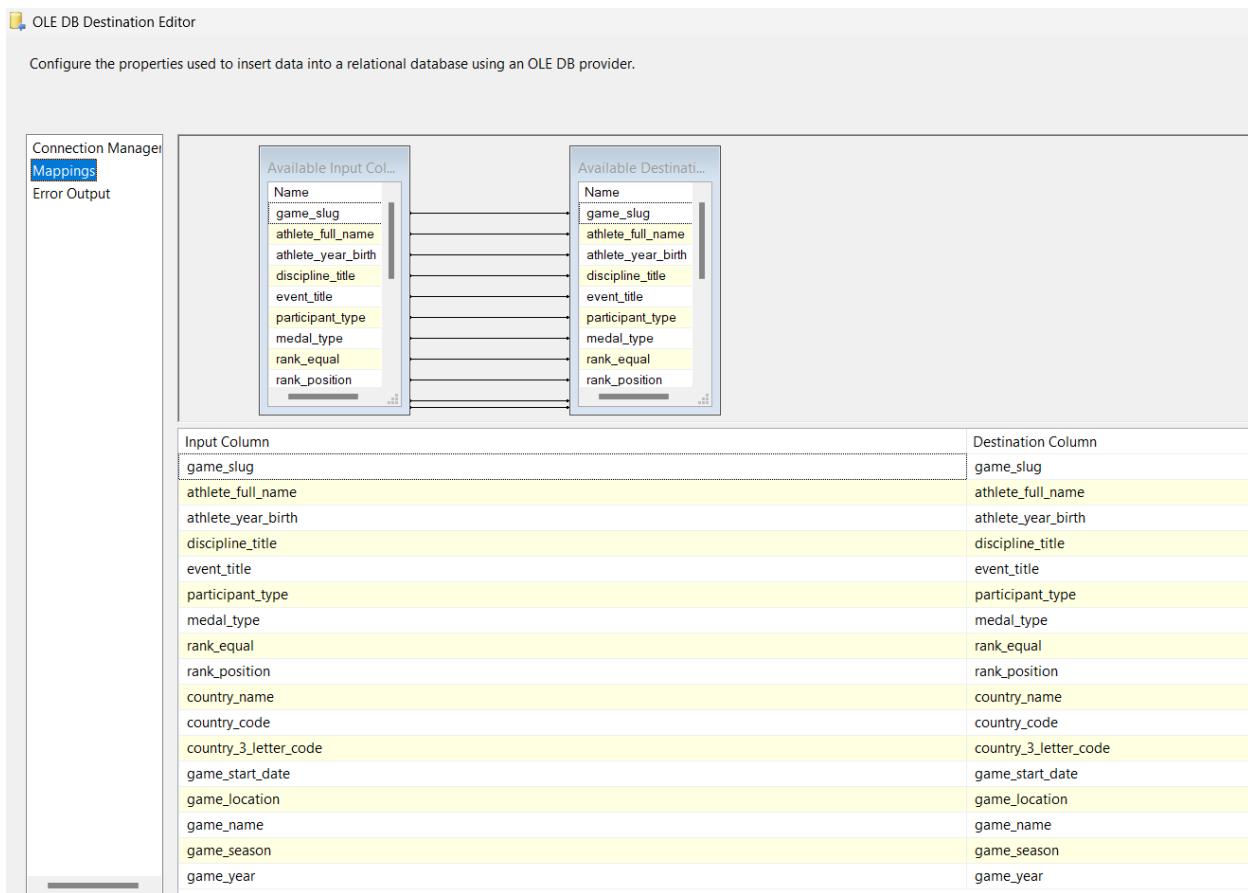
- Bước 7: Chọn bảng dữ liệu sẽ đổ dữ liệu vào, ở đây ta sẽ chọn Destination là “**OLE DB Destination**” với kho dữ liệu “**Raw_Data**”.



Hình 53 Đưa dữ liệu vào Raw_Data



Hình 54 Tạo mới bảng Raw_Data

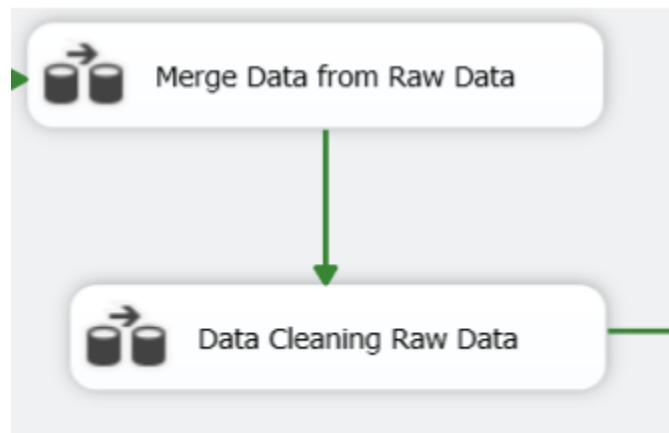


Hình 55 Nối các cột trong bảng

2.4. Quá trình làm sạch dữ liệu

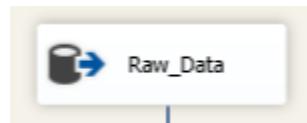
Để tiến hành việc làm sạch dữ liệu trong dự án SSIS, ta thực hiện các bước sau:

- Bước 1: Tại cửa sổ Control Flow, thực hiện việc kéo thả công cụ Data Flow Task tại SSIS Toolbox, sau đó đổi tên thành Data Cleaning Raw Data. Nối quy trình nối dữ liệu *Merge Data From RawData* vào *Data Cleaning Raw Data*.



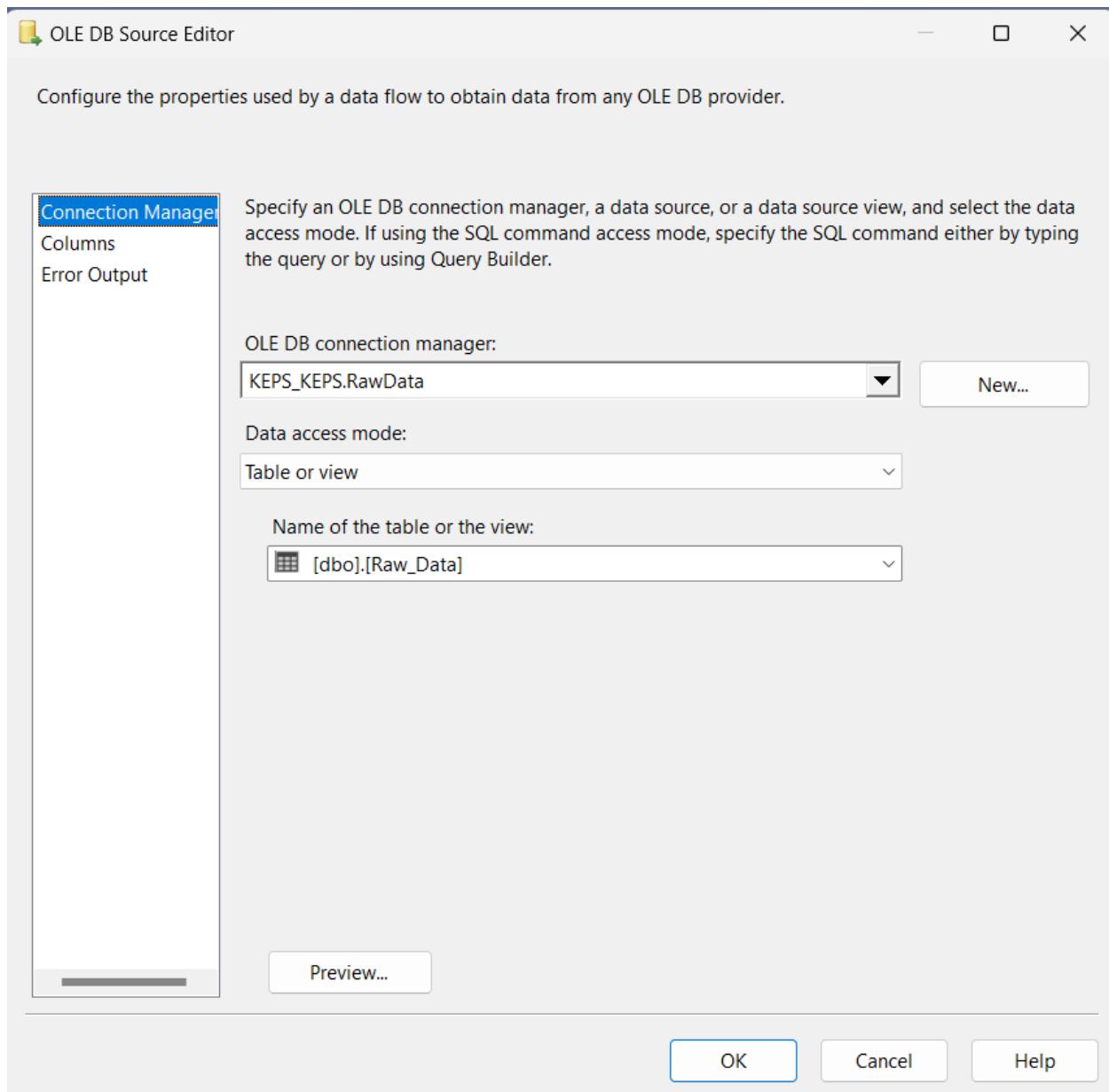
Hình 56 Nối data flow mới với data flow nối dữ liệu

- Bước 2: Chọn quy trình *Data Cleaning DB_RawData*, tại cửa sổ **Data Flow** tạo một nguồn dữ liệu mới **OLE DB Source** ở SSIS Toolbox, sau đó đổi tên thành *Raw_Data*



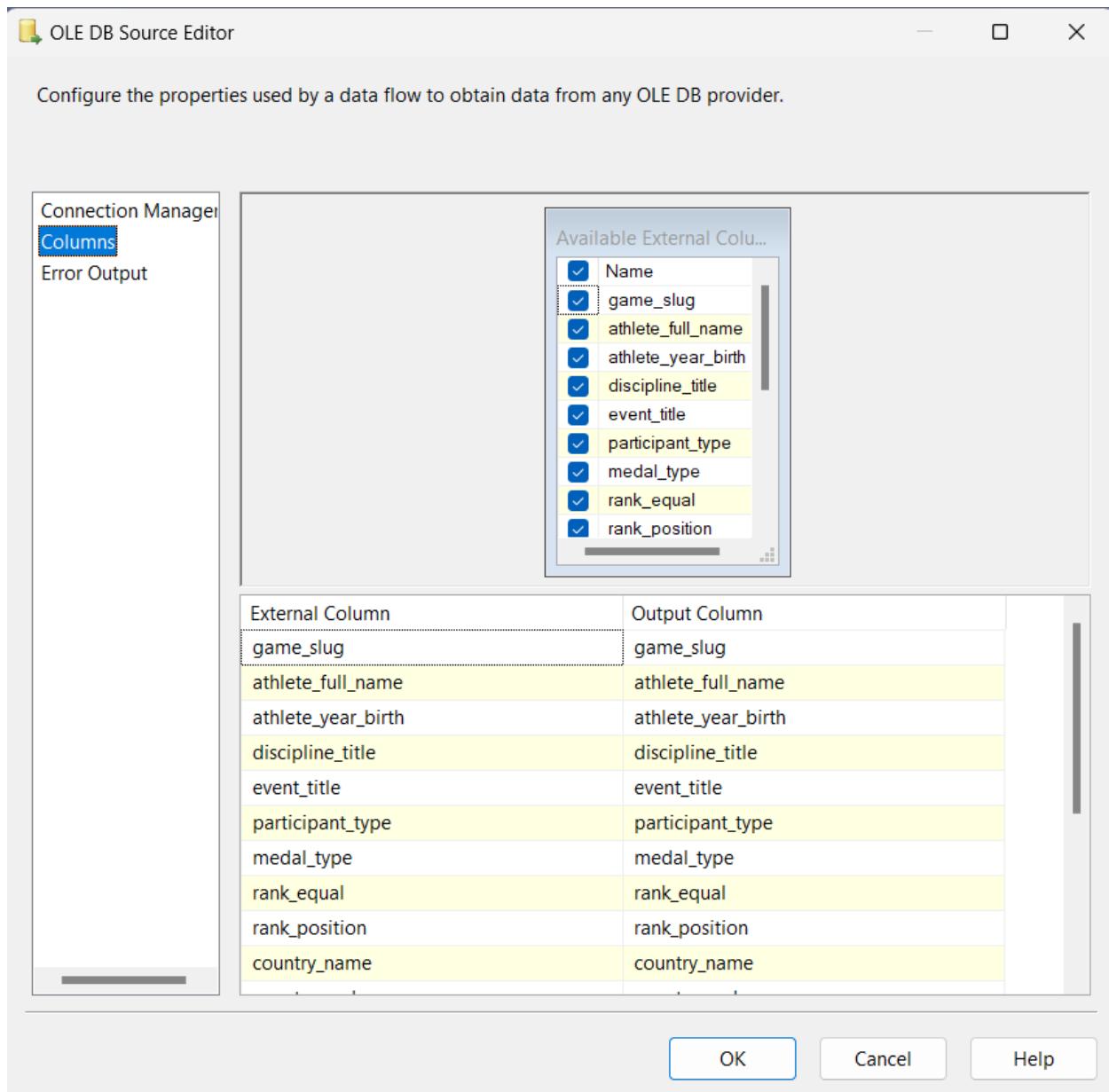
Hình 57 Tạo nguồn dữ liệu *Raw_Data*

- Bước 3: Chọn **Edit** tại *Raw_Data*, tại mục Connection Manager, chọn OLE DB connection manager là **RawData** và chọn bảng **Raw_Data** và chọn **OK** để hoàn tất



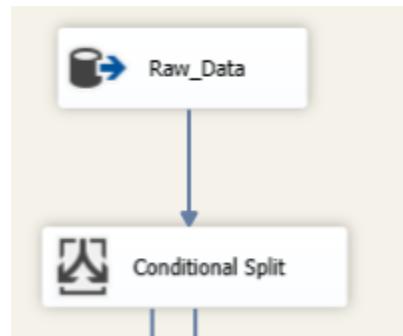
Hình 58 Tạo kết nối với bảng Raw_Data

- Bước 4: Chuyển sang mục **Columns**, chọn tất cả các cột và chọn **OK**



Hình 59 Chọn các cột để sử dụng

- Bước 5: Thực hiện việc **loại bỏ các dữ liệu rỗng (NULL)** bằng công cụ Conditional Split, kéo thả công cụ **Conditional Split** tại SSIS Toolbox.



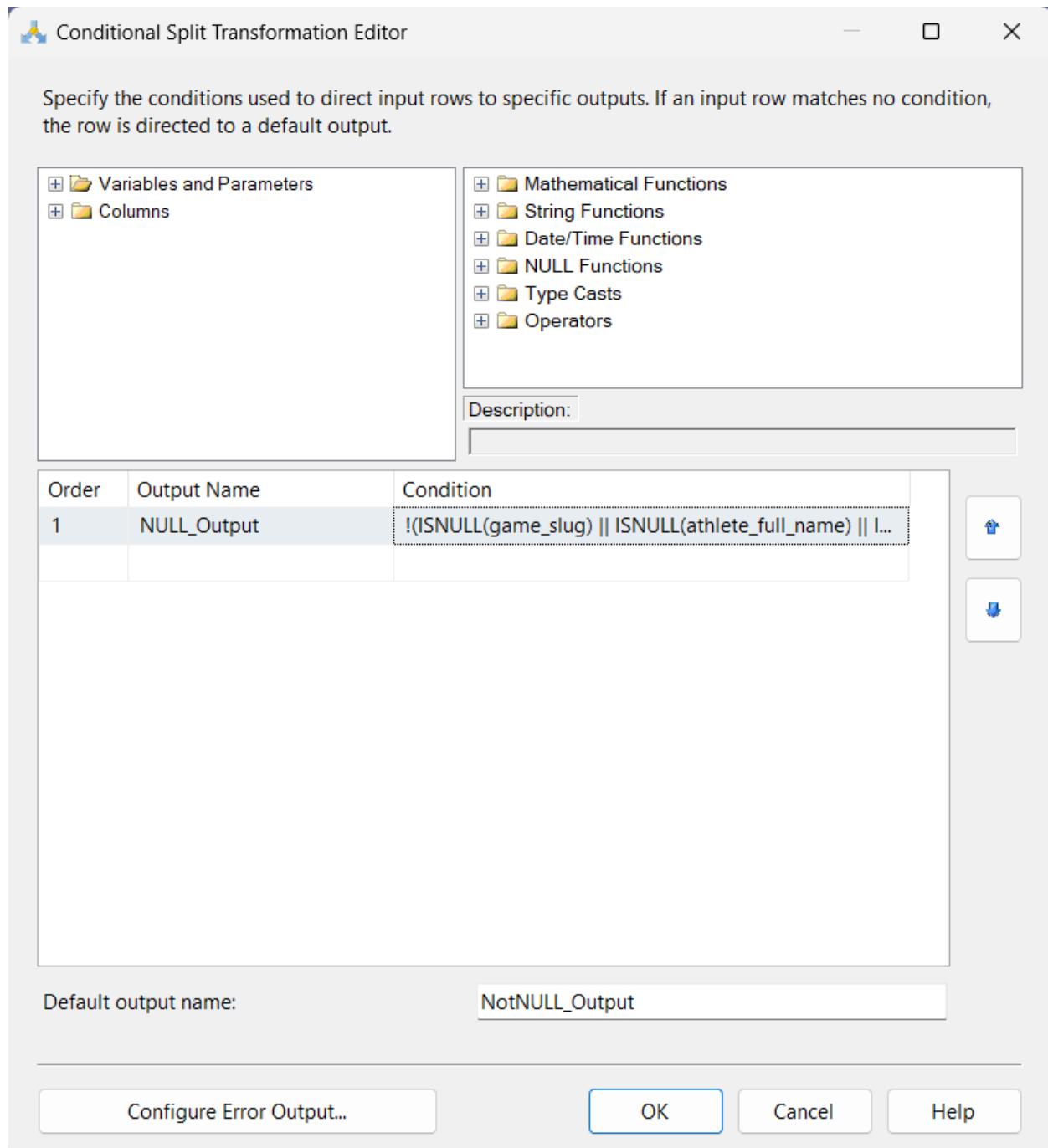
Hình 60 Tạo công cụ Conditional Split

- Bước 6: Chọn Edit trong Conditional Split, lọc các dữ liệu rỗng bằng hàm **ISNULL**.
Tạo một output có tên là *Null_Output* và lọc với điều kiện:

```

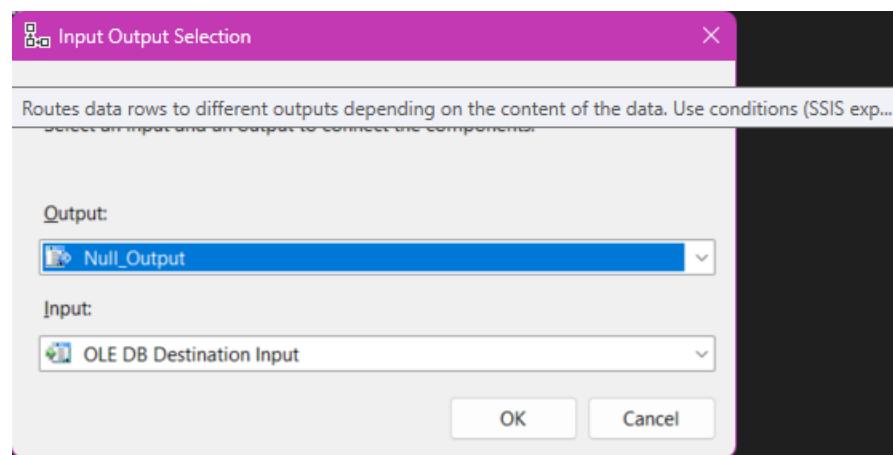
!(ISNULL(game_slug) // ISNULL(athlete_full_name) // ISNULL(athlete_year_birth) //
ISNULL(discipline_title) // ISNULL(event_title) // ISNULL(participant_type) //
ISNULL(medal_type) // ISNULL(rank_equal) // ISNULL(rank_position) //
ISNULL(country_name) // ISNULL(country_code) // ISNULL(country_3_letter_code) //
ISNULL(game_start_date) // ISNULL(game_location) // ISNULL(game_name) //
ISNULL(game_season) // ISNULL(game_year))
  
```

bằng cách kéo thả các hàm và các cột tương ứng. Sau đó đặt tên cho **Default output name** là *NOTNULL_Output* tương ứng với dữ liệu đầu ra không chứa các thuộc tính rỗng. Khi hoàn tất, chọn **OK**.

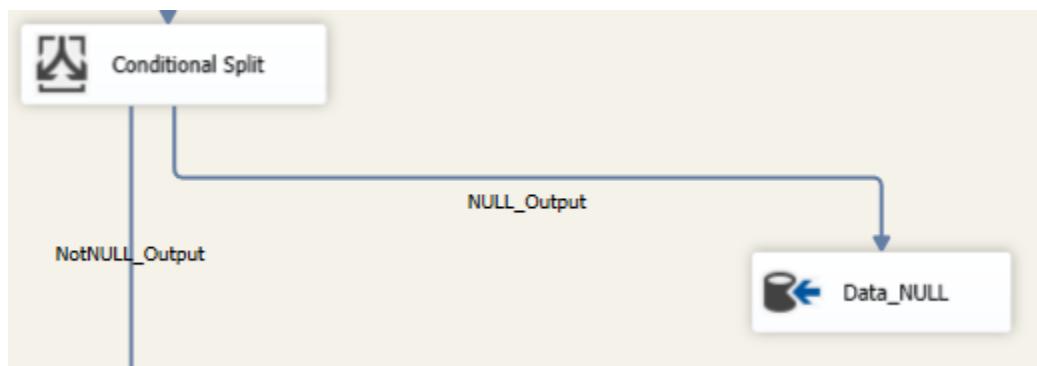


Hình 61 Tạo các hàm NULL_Output và NotNULL_Output

- Bước 7: Chọn dữ liệu đưa vào là OLE DB Destination, kéo thả công cụ vào cửa sổ Data Flow, sau đó đổi tên thành *Data_NULL* với **Output_Name** là *Null_Output*.

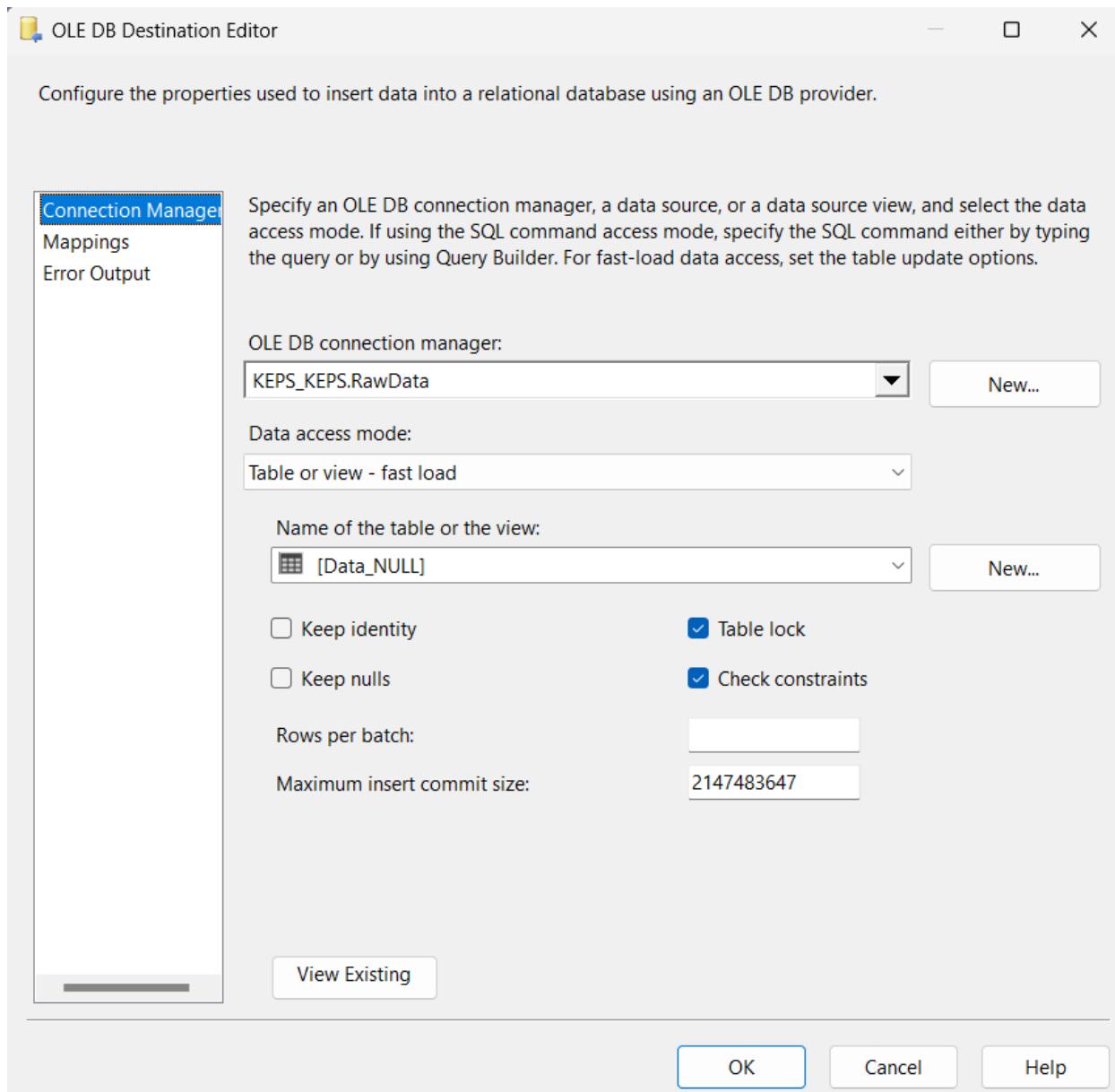


Hình 62 Chọn Null_Output

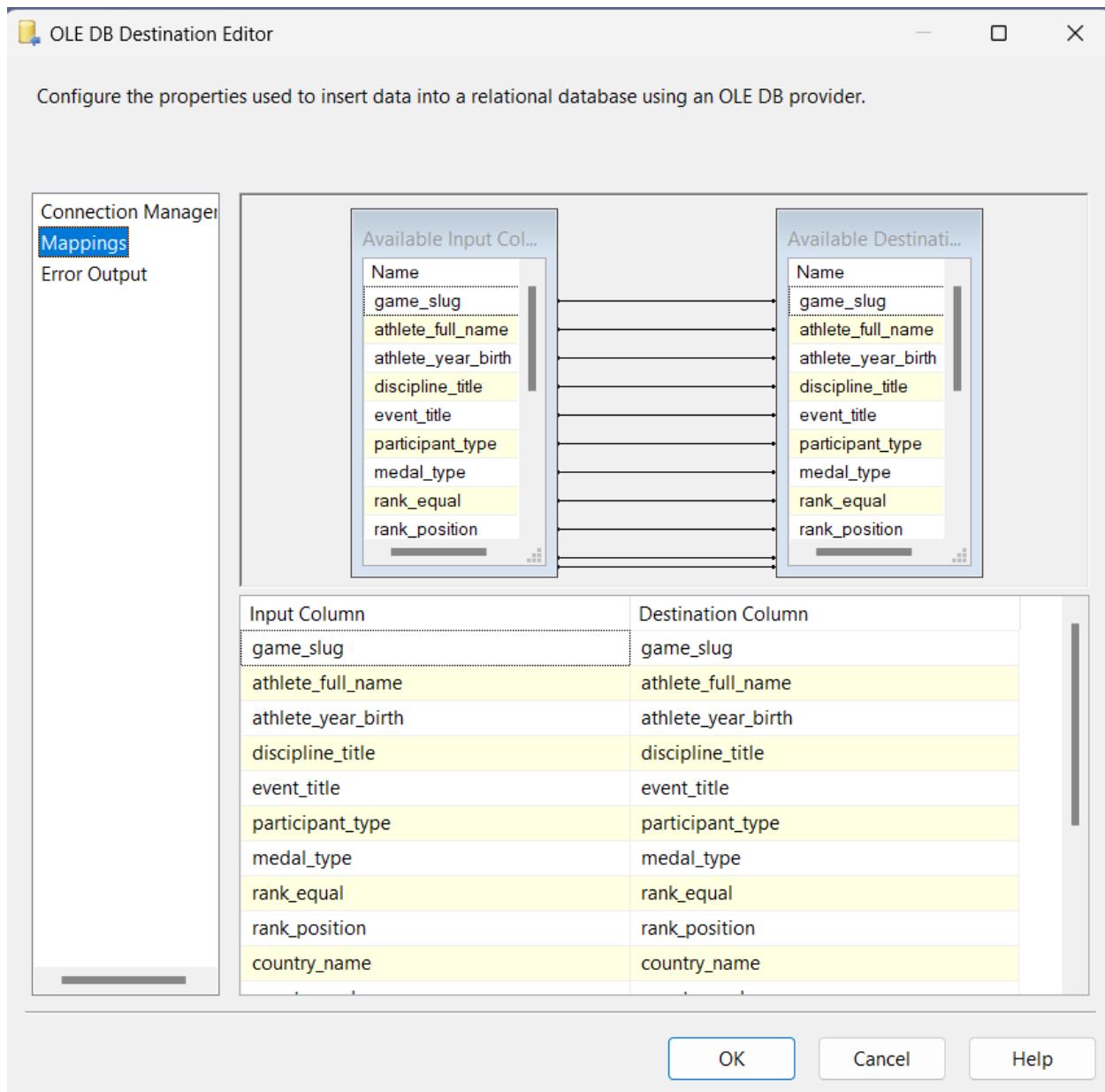


Hình 63 Tạo đích đến Data_NULL

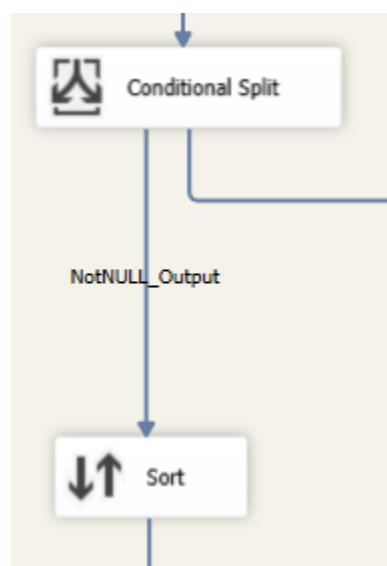
- Bước 8: Thiết lập cho OLE DB Destination *Data_NULL*. Đầu tiên, chọn New để tạo một bảng mới có tên là *DATA_NULL* và chọn **OK**. Sau đó, sử dụng các thiết lập cho Connection Manager và Mappings



Hình 64 Tạo bảng Data_NULL

*Hình 65 Nối các cột thuộc tính*

- Bước 9: Thực hiện việc sắp xếp lại dữ liệu bằng công cụ Sort để tránh dẫn đến trùng lặp dữ liệu



Hình 66 Sắp xếp dữ liệu sau khi xử lý

Sort Transformation Editor

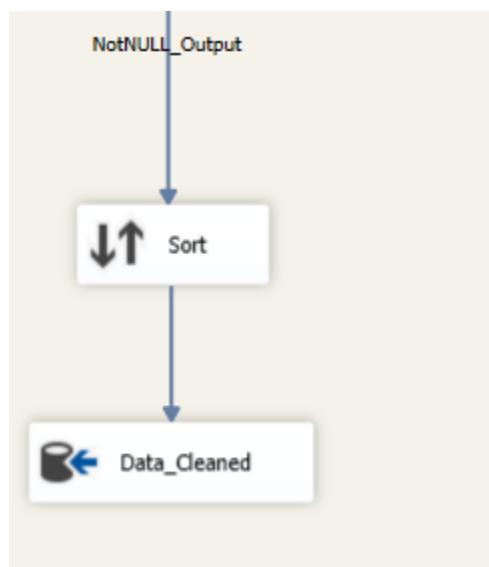
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Available Input Columns		
	Name	Pass Thr...
<input checked="" type="checkbox"/>	game_slug	<input type="button" value="▼"/>
<input checked="" type="checkbox"/>	athlete_full_name	<input type="button" value="▼"/>
<input checked="" type="checkbox"/>	athlete_year_birth	<input type="button" value="▼"/>
<input checked="" type="checkbox"/>	discipline_title	<input type="button" value="▼"/>
<input checked="" type="checkbox"/>	event_title	<input type="button" value="▼"/>
<input checked="" type="checkbox"/>	participant_type	<input type="button" value="▼"/>
<input type="button" value="..."/>		

Input Column	Output Alias	Sort Type	Sort Order	Comparison Flags
game_slug	game_slug	ascending	1	
athlete_full_name	athlete_full_name	ascending	2	
athlete_year_birth	athlete_year_birth	ascending	3	
discipline_title	discipline_title	ascending	4	
event_title	event_title	ascending	5	
participant_type	participant_type	ascending	6	
medal_type	medal_type	ascending	7	
rank_equal	rank_equal	ascending	8	
rank_position	rank_position	ascending	9	
country_name	country_name	ascending	10	
country_code	country_code	ascending	11	
country_3_letter_code	country_3_letter_code	ascending	12	
game_start_date	game_start_date	ascending	13	
game_location	game_location	ascending	14	
game_name	game_name	ascending	15	
game_season	game_season	ascending	16	
game_year	game_year	ascending	17	

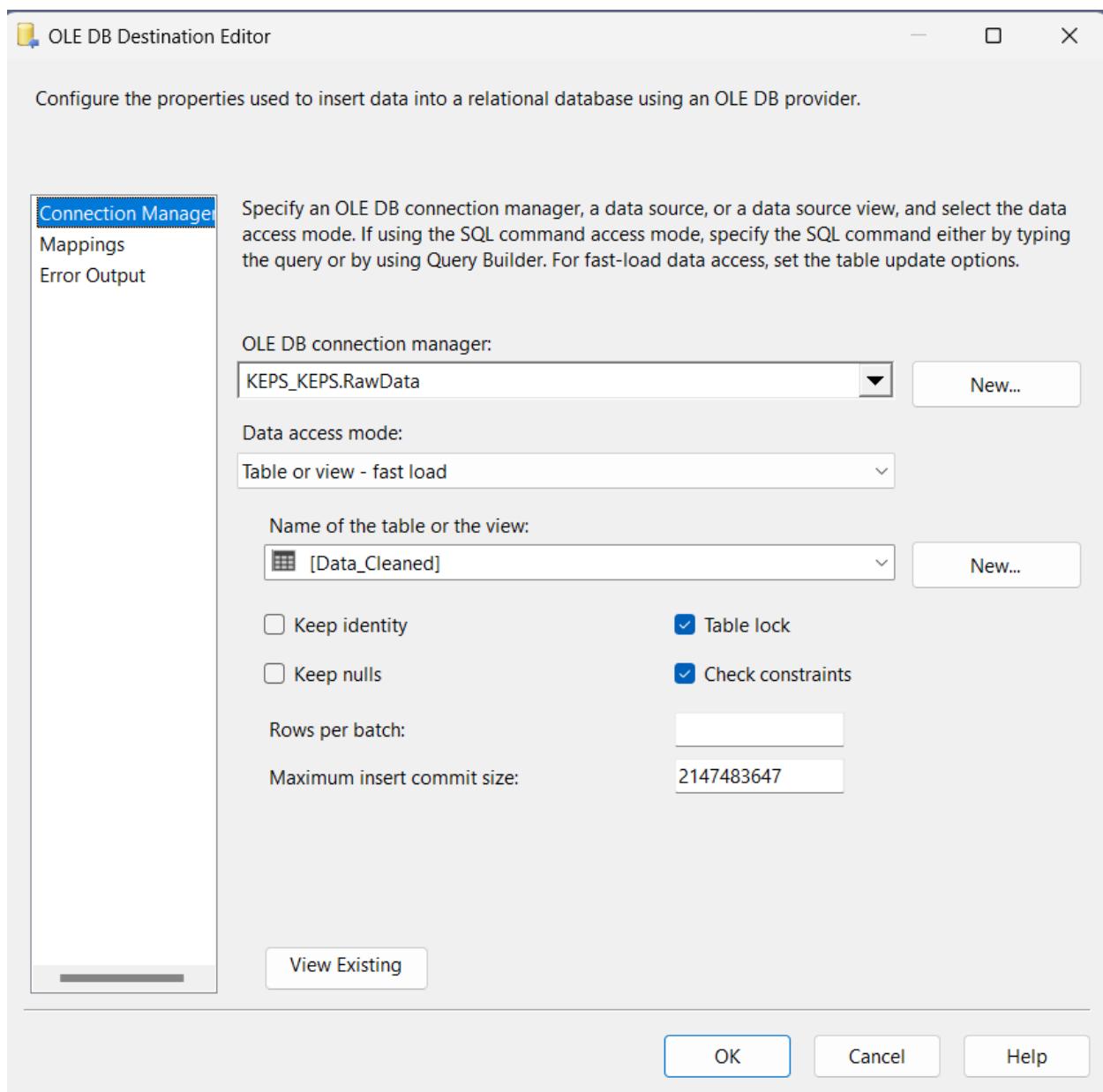
Hình 67 Lựa chọn các thuộc tính để sắp xếp

- Bước 10: Chọn dữ liệu đưa vào là OLE DB Destination, kéo thả công cụ vào cửa sổ Data Flow, sau đó đổi tên thành *Data_Cleaned* với **Output_Name** là *NOTNULL_Output* trên hành động Sort

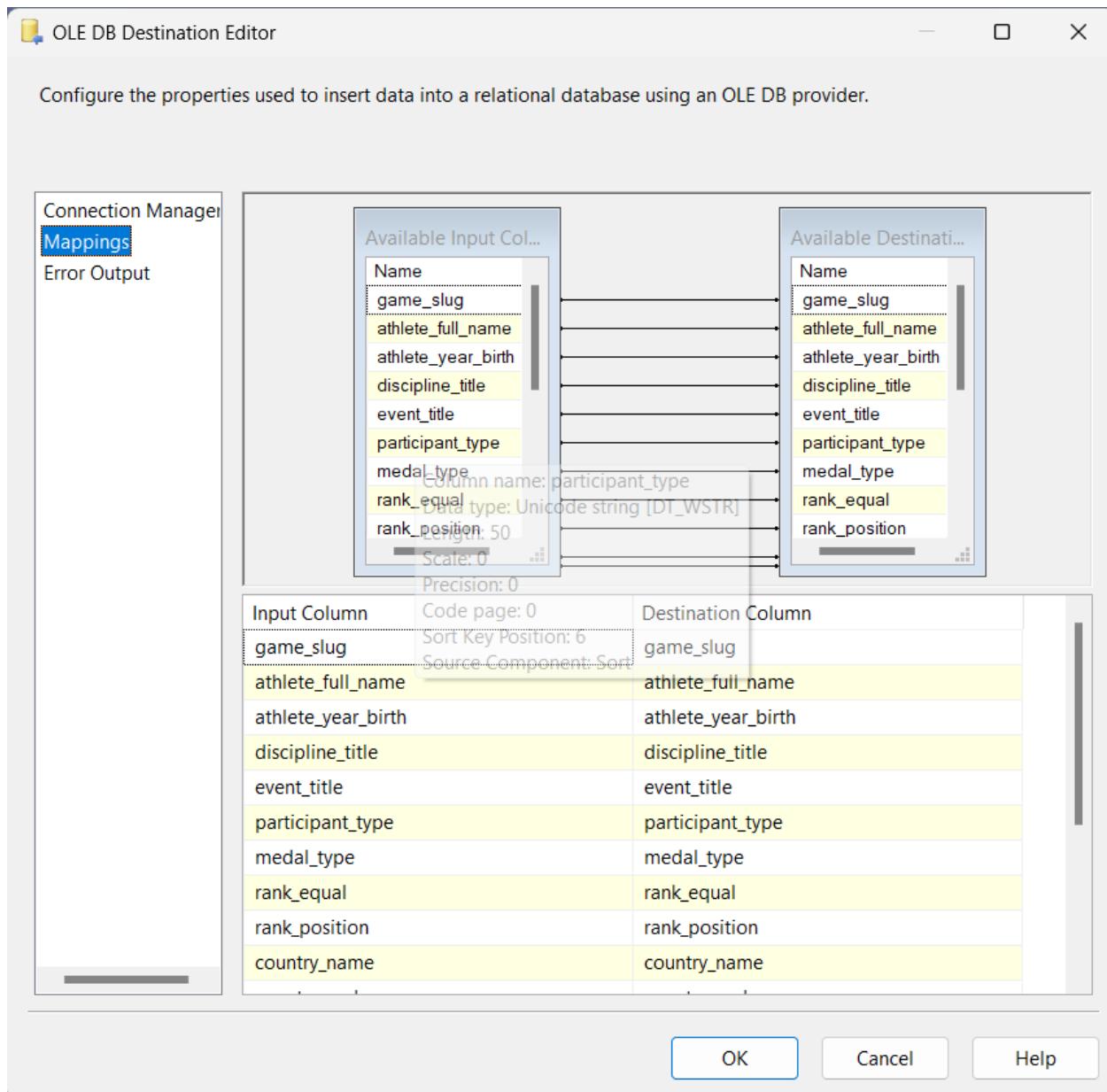


Hình 68 Tạo đích đến Data_Cleaned

- Bước 11: Thiết lập Data_Cleaned



Hình 69 Tạo bảng Data_Cleaned

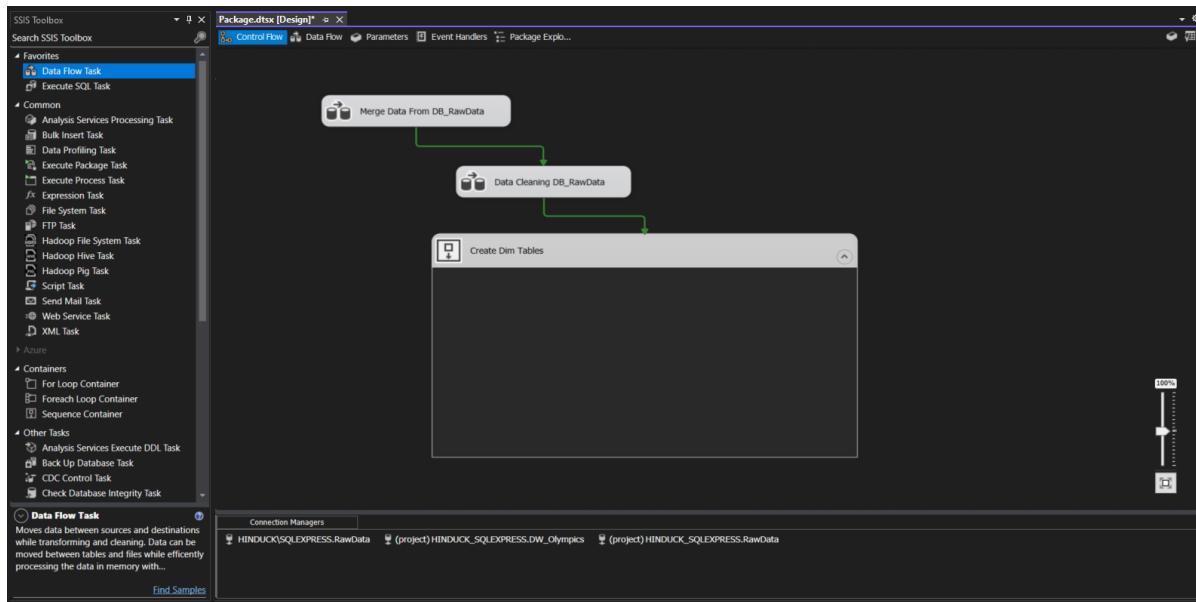


Hình 70 Nối các cột trong bảng

2.5. Quá trình khởi tạo các bảng Dimension

Sau quá trình xử lý dữ liệu đầu vào, ta thực hiện việc tạo các bảng Dimension dựa trên lược đồ hình sao.

Để xây dựng một quy trình việc thực hiện việc tạo các bảng Dimension, trước hết ta cần tạo một Sequence Container và đổi tên thành Create Dim Tables từ cửa sổ SSIS Toolbox.



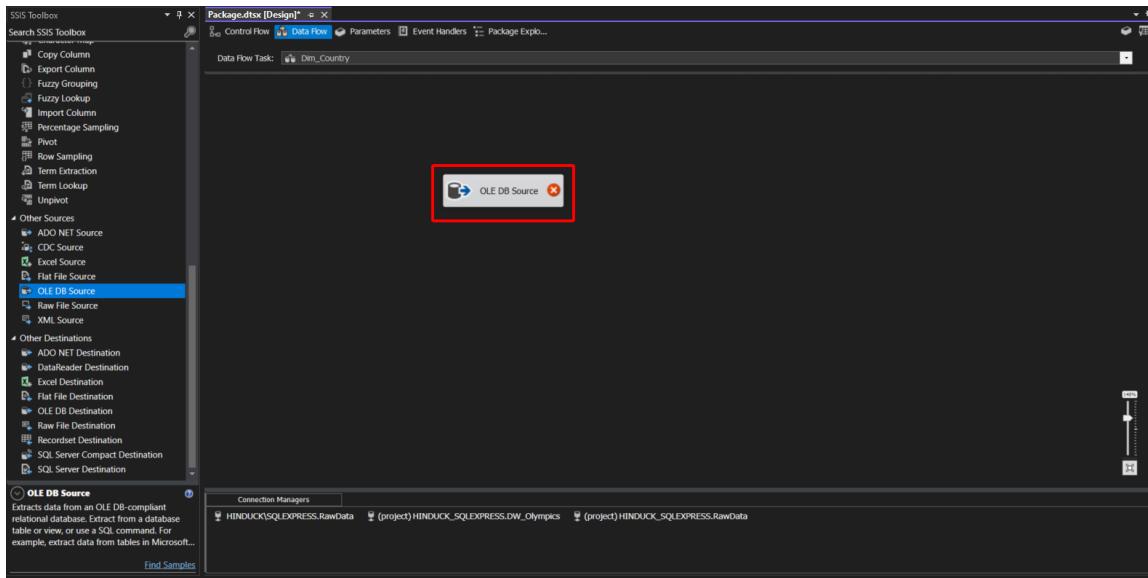
Hình 71 Tạo Sequence Container “Create Dim Table”

Để khởi tạo bảng *Dim_Country* cho tác vụ khởi tạo các bảng Dimension trong dự án SSIS, ta thực hiện các bước sau:

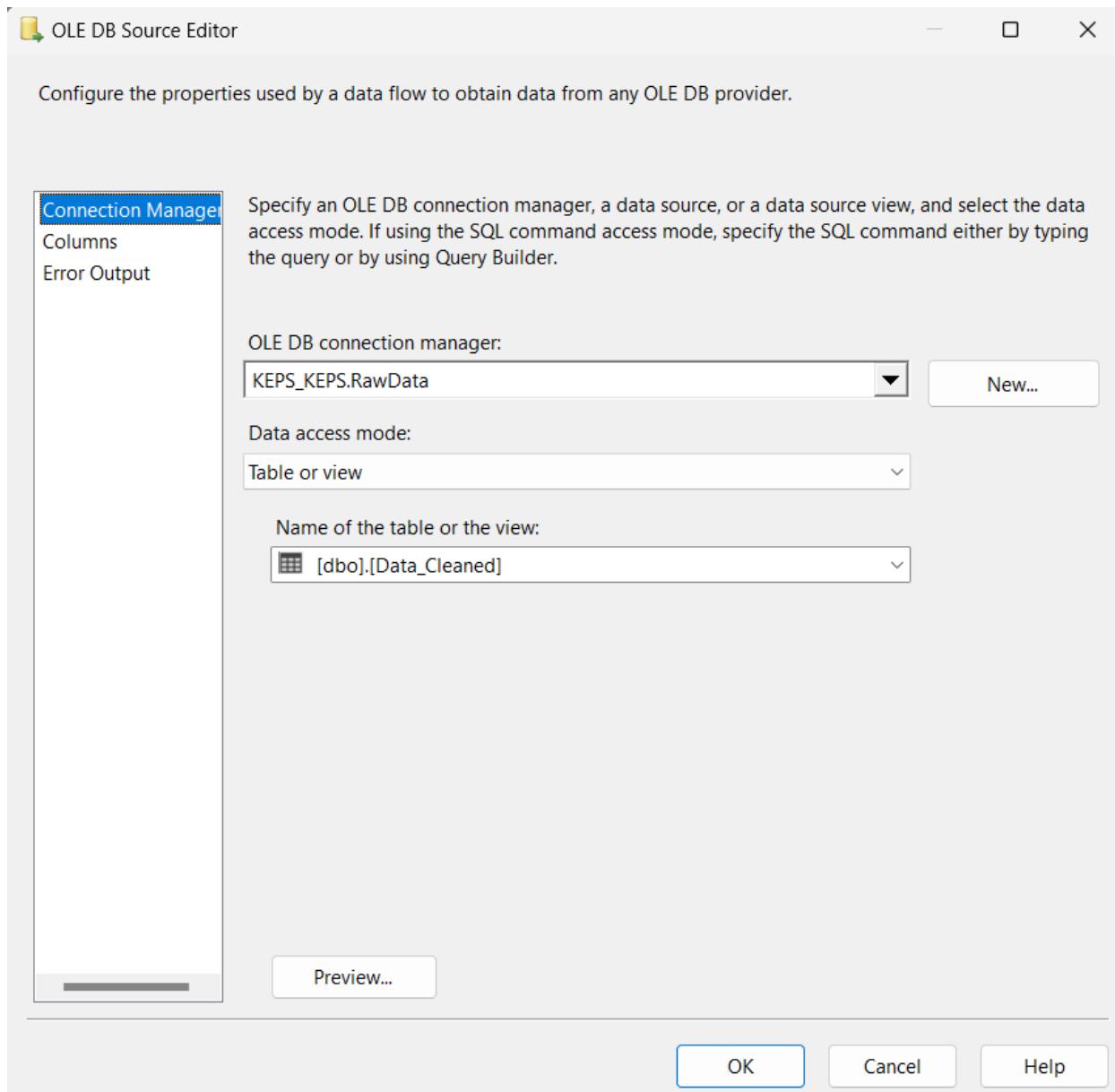
- **Bước 1:** Tại cửa sổ **Control Flow**, thực hiện việc kéo thả công cụ **Data Flow Task** tại SSIS Toolbox vào tác vụ *Create Dim Tables*, sau đó đổi tên thành *Dim_Country*.



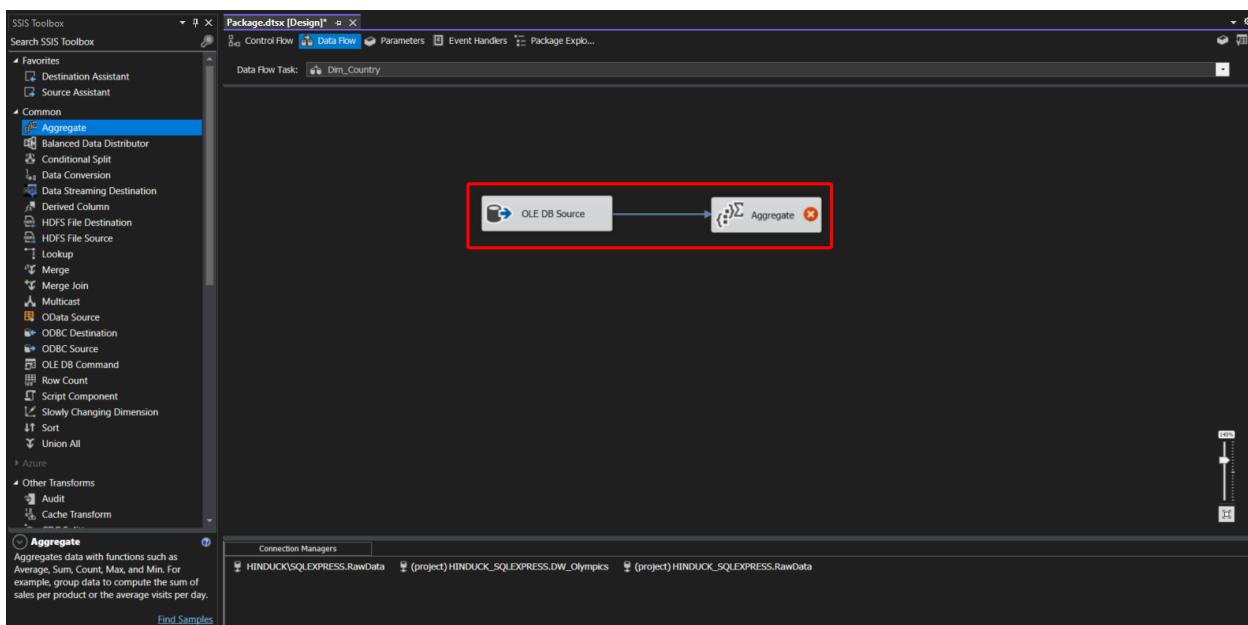
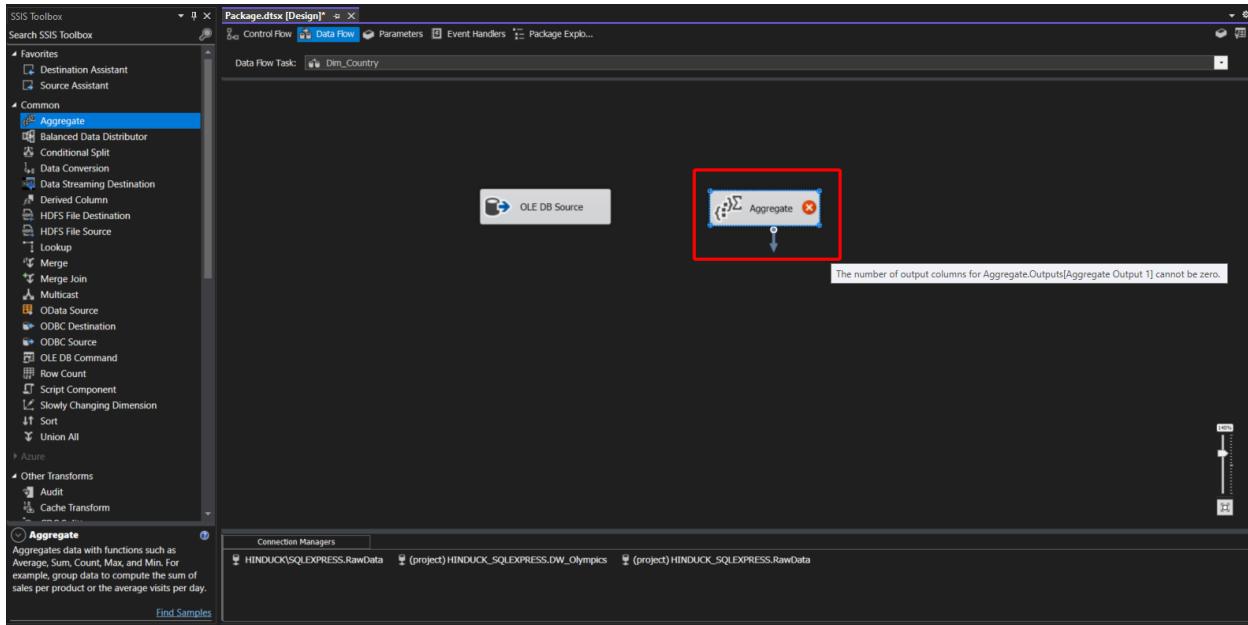
- **Bước 2:** Chọn quy trình *Dim_Country*, tại cửa sổ **Data Flow** tạo một nguồn dữ liệu mới **OLE DB Source** ở SSIS Toolbox.



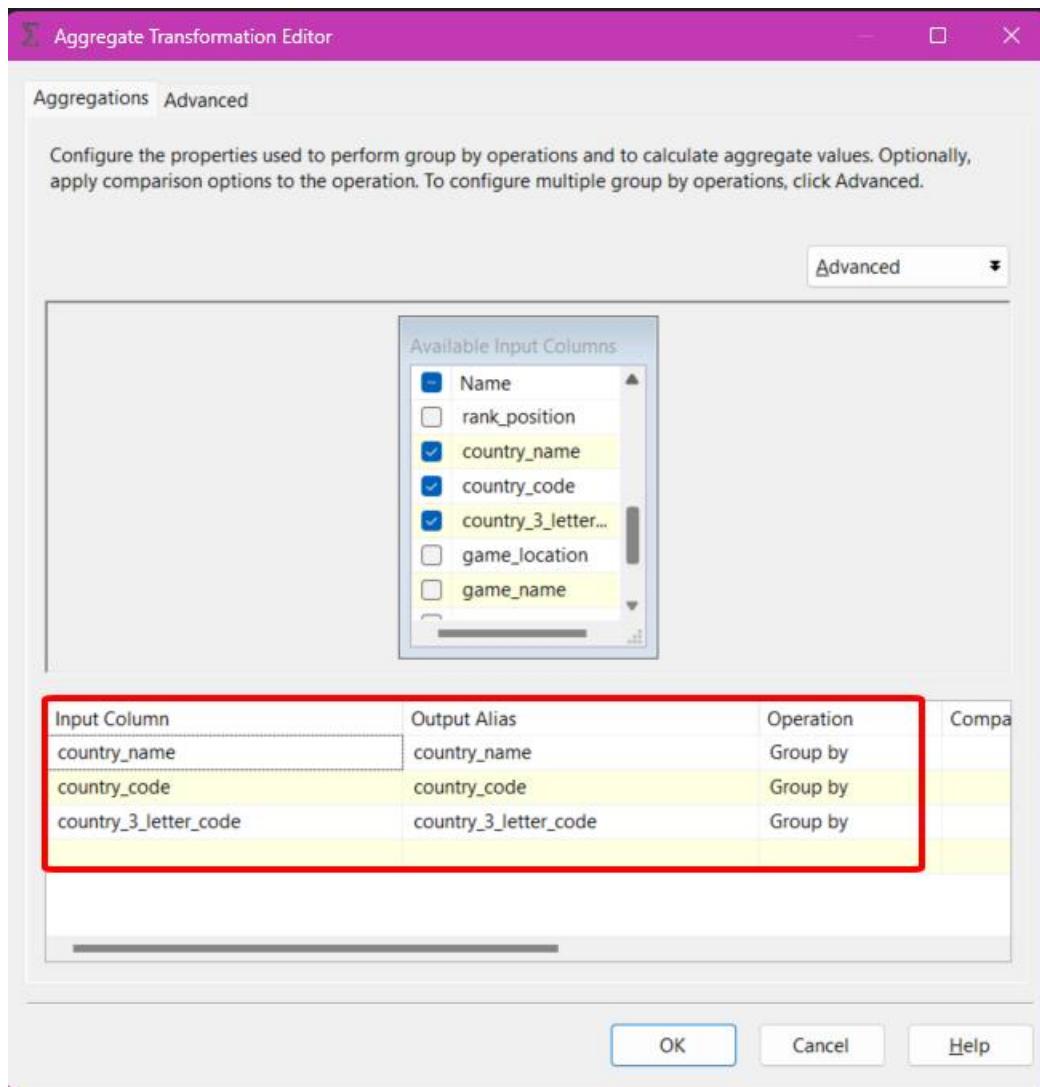
- **Bước 3:** Chọn **Edit** tại OLE DB Source để thiết lập cấu hình, tại mục Connection Manager, chọn OLE DB connection manager là **RawData** và chọn bảng **Data_Cleaned** và chọn **OK** để hoàn tất.



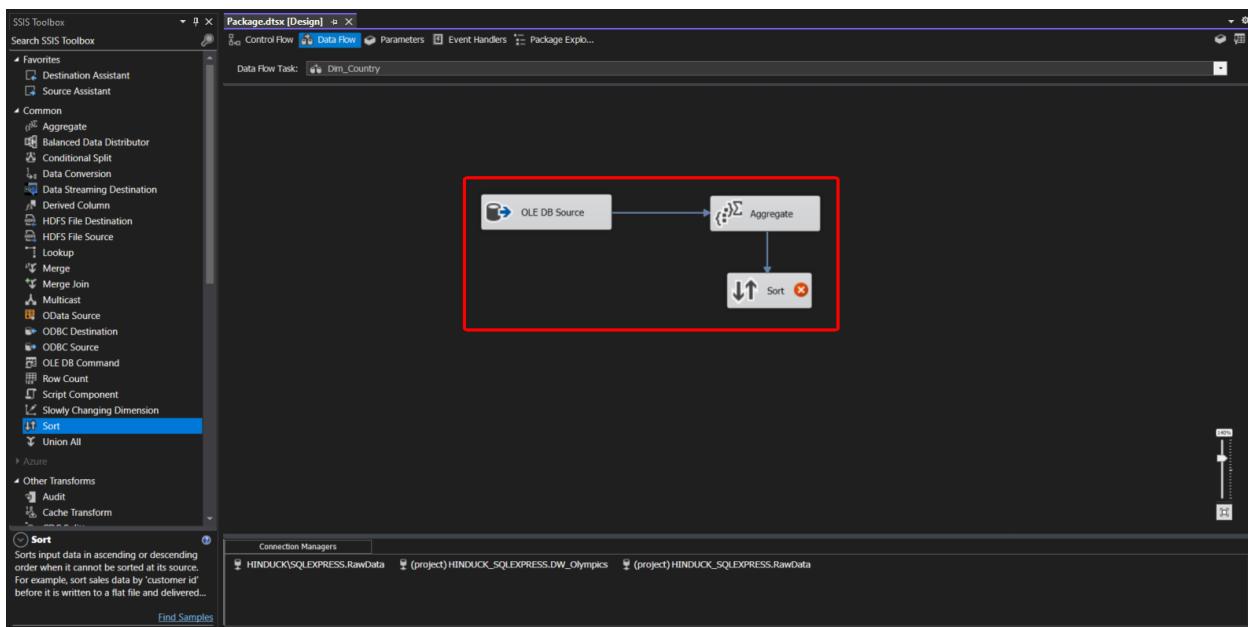
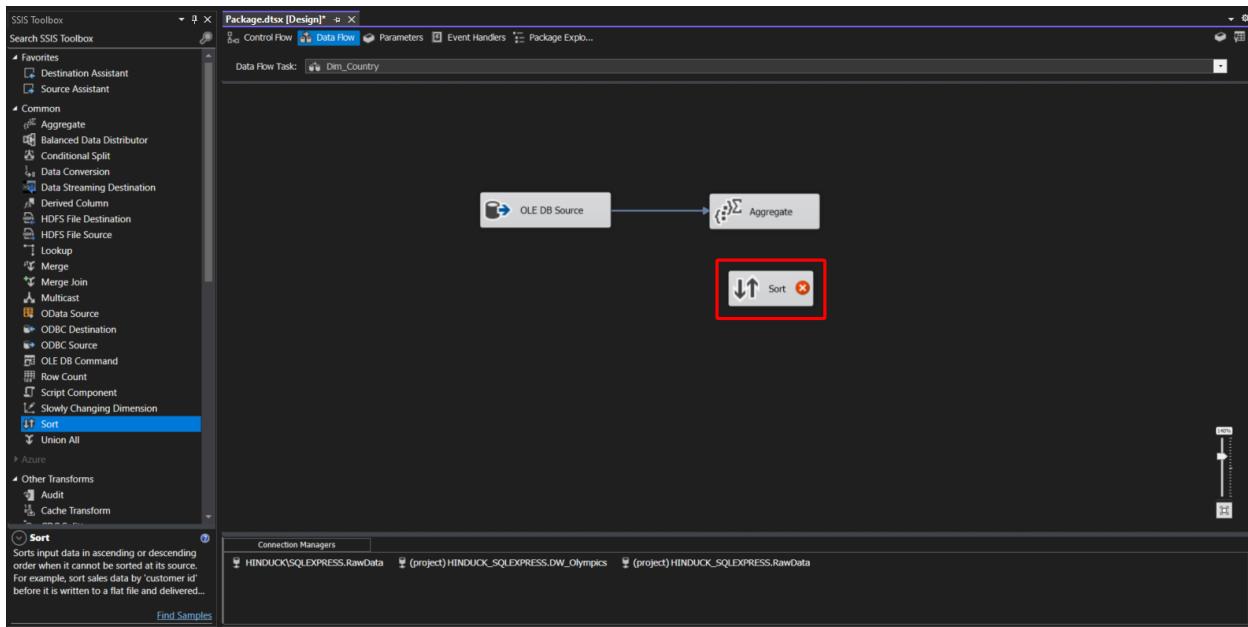
- Bước 4: Ta cần tổng hợp dữ liệu trong bảng *Dim_Country*, kéo thả để sử dụng công cụ **Aggregate** trong SSIS Toolbox.



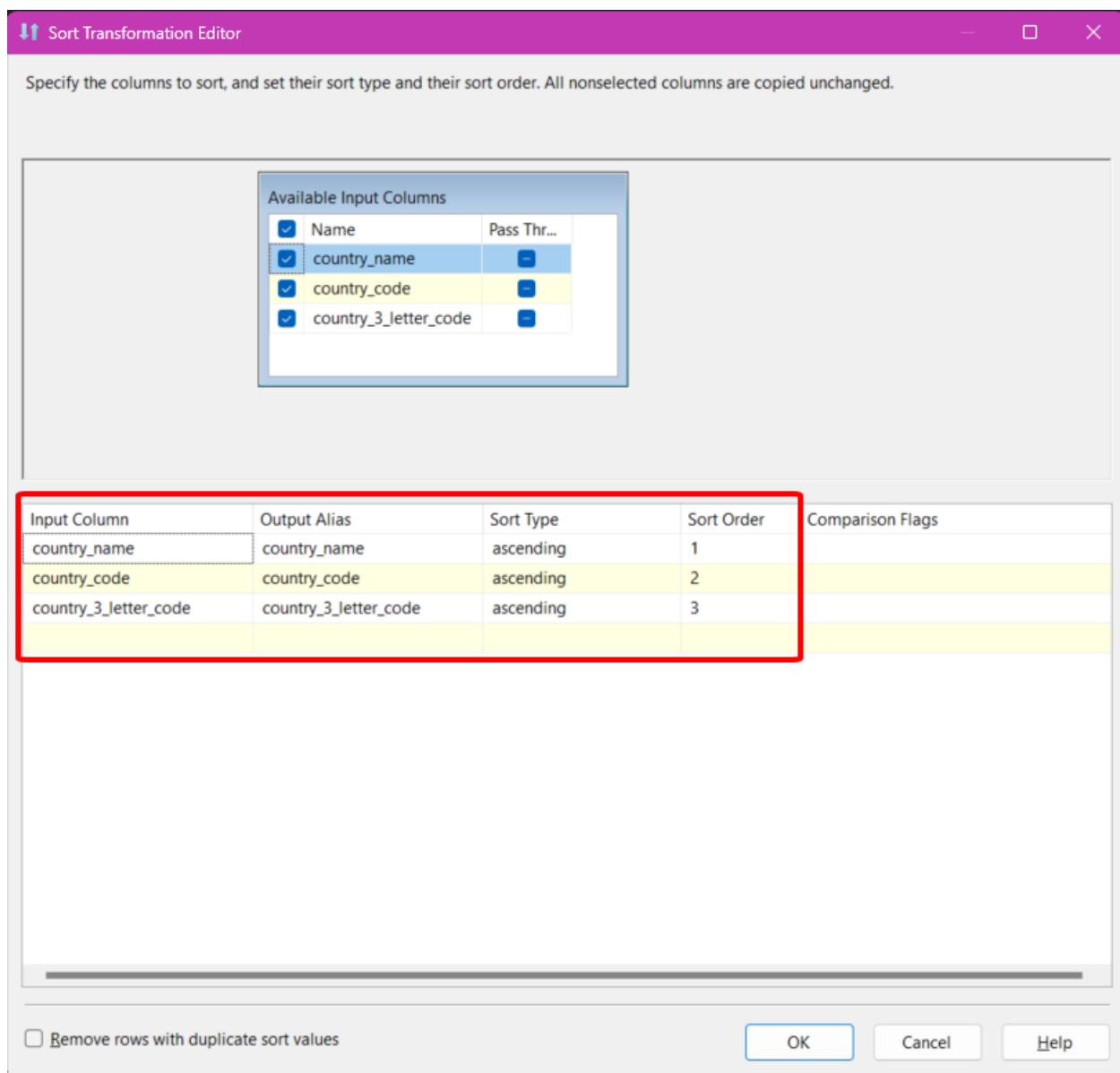
- Bước 5:** Chọn **Edit** tại **Aggregate**, chọn các cột cần sử dụng để tạo bảng *Dim_Country*, ở đây ta chọn cột *country_name*, *country_code* và *country_3_letter_code*. Chọn **OK** để hoàn tất.



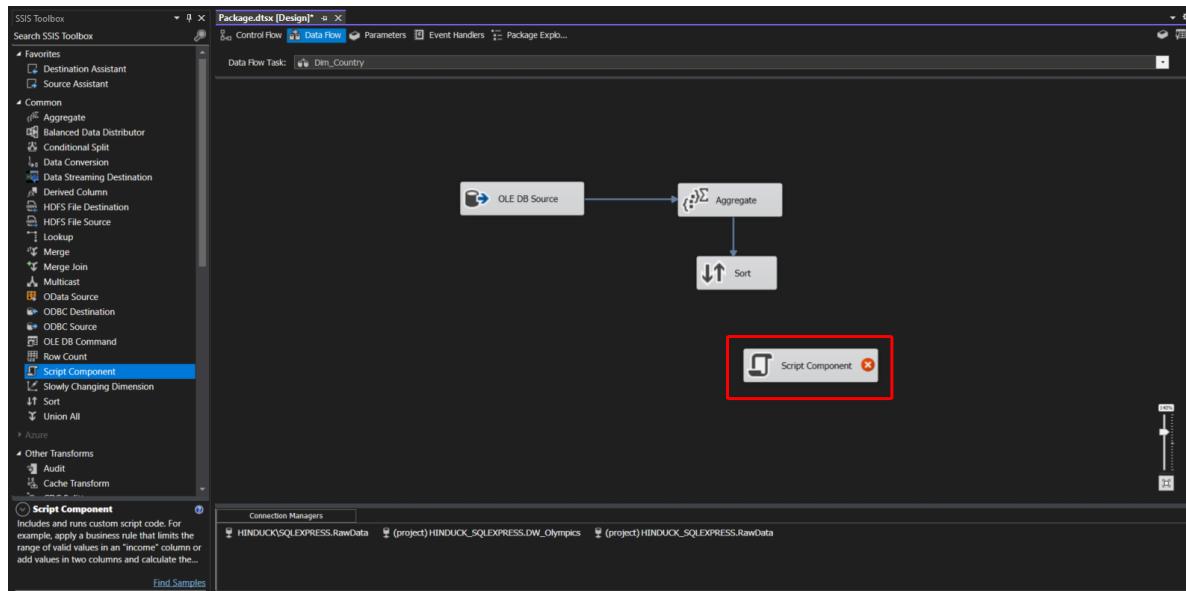
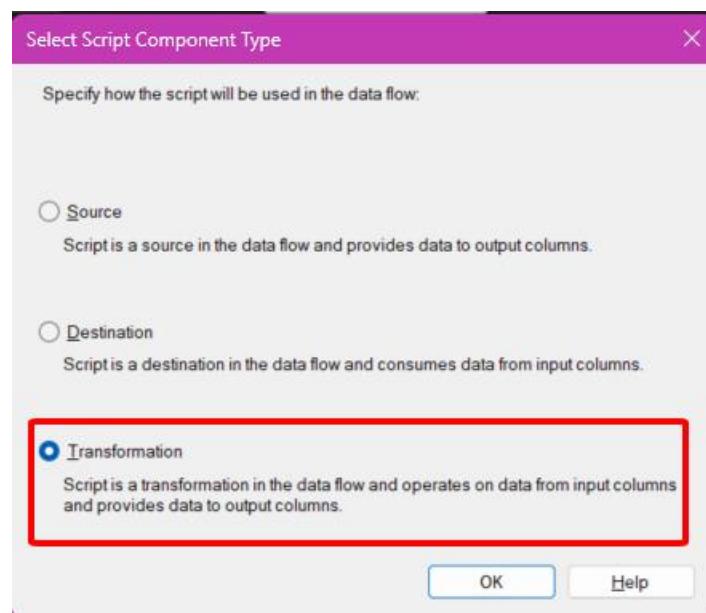
- Bước 6: Tiến hành sắp xếp các dữ liệu, chọn sự kiện **Sort** tại Toolbox, nối từ **Aggregate** vào **Sort**.

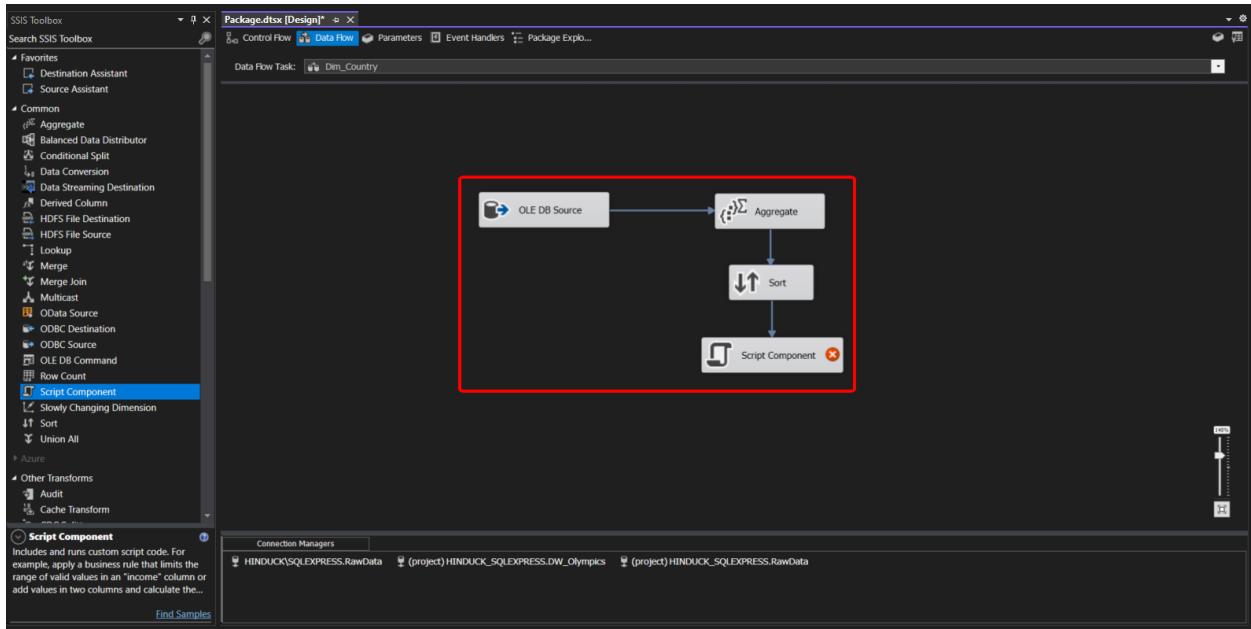


- **Bước 7:** Chọn **Edit** tại **Sort**, chọn các cột cần sử dụng để sắp xếp. Chọn **OK** để hoàn tất.



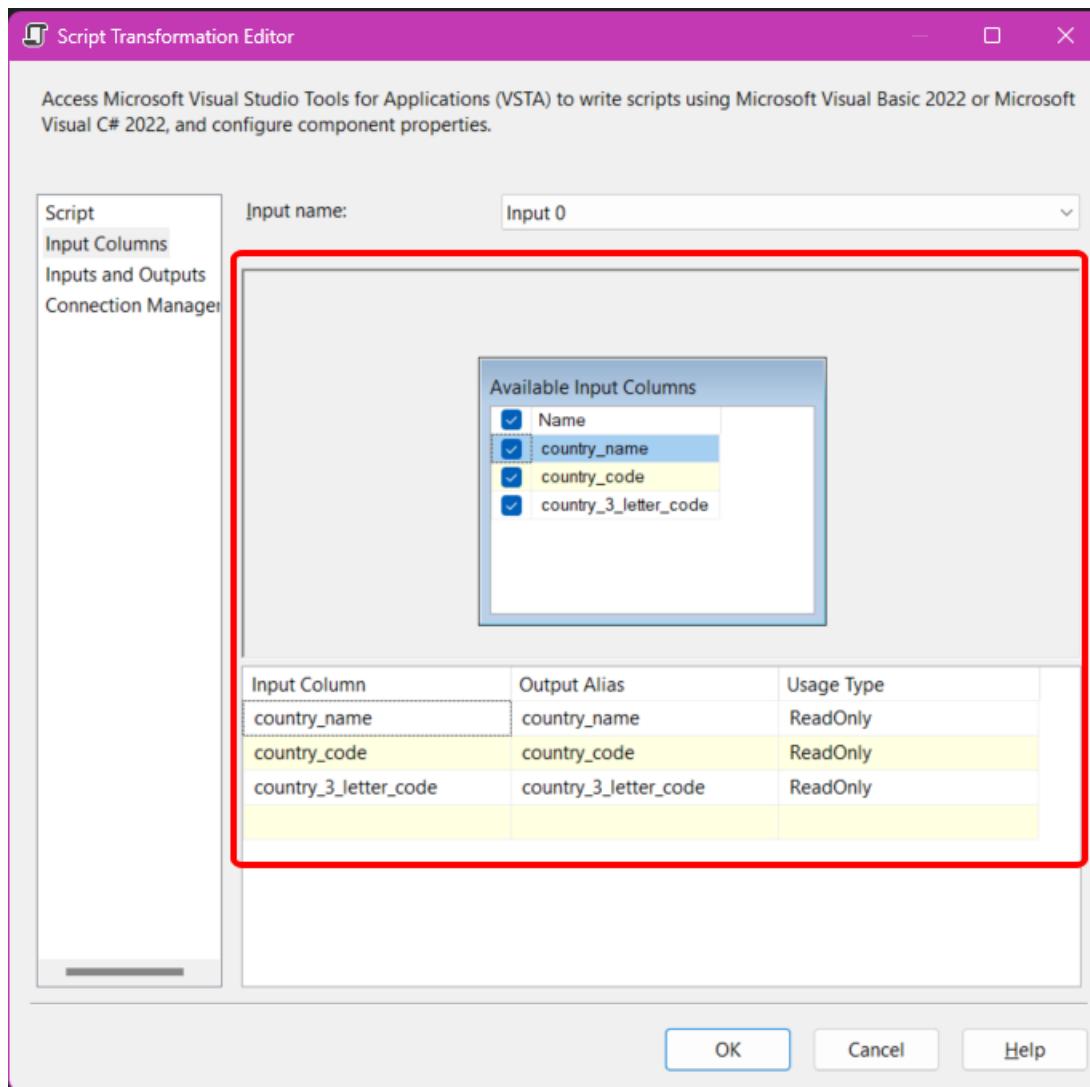
- **Bước 8:** Thêm một tác vụ **Script Component** vào quá trình tạo bảng *Dim_Country*. Kéo thả để sử dụng tác vụ **Script Component**, sau đó chọn loại Transformation Script Component và nối từ các tác vụ trên vào **Script Component** đã chọn và thực hiện thiết lập.



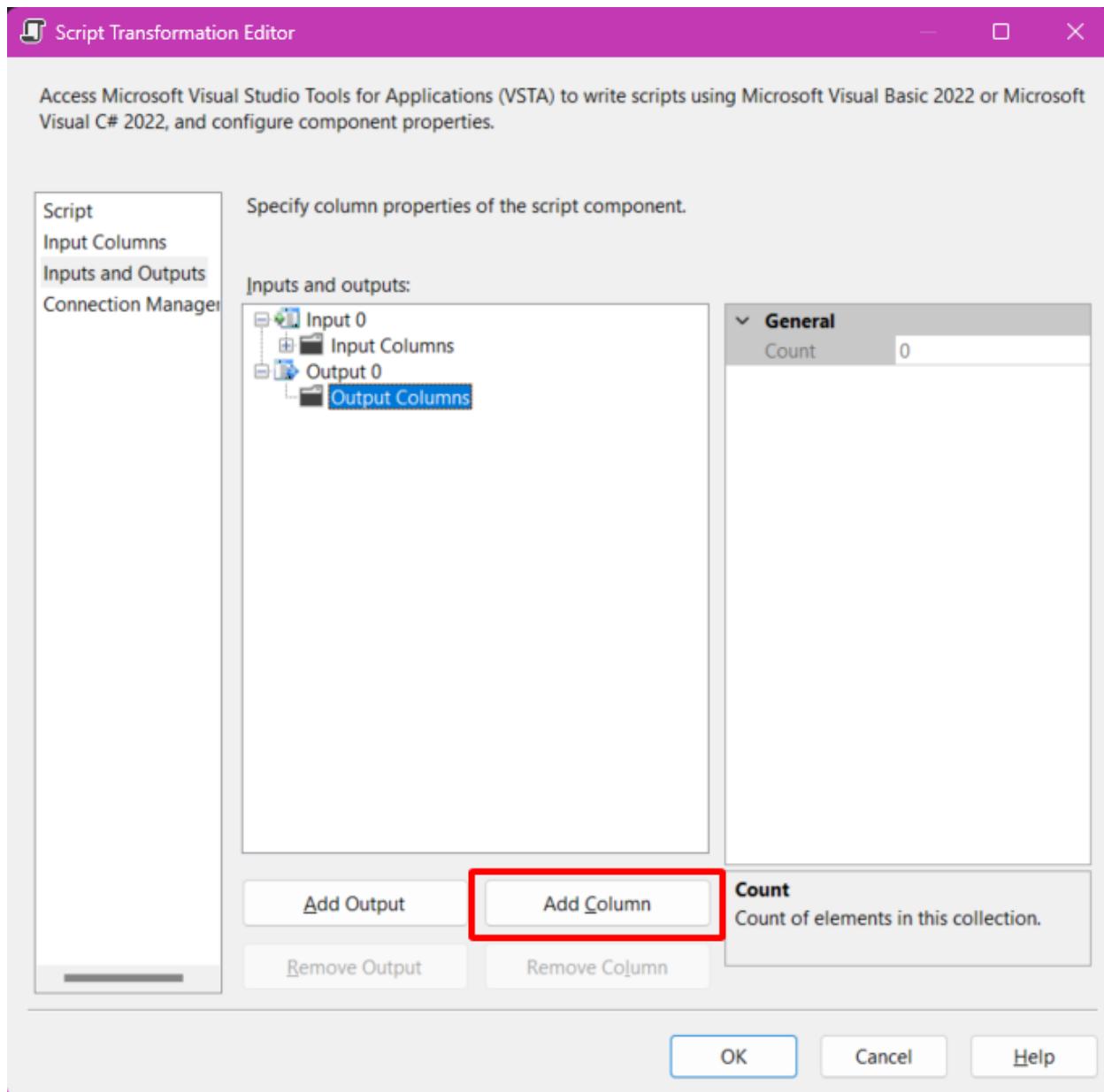


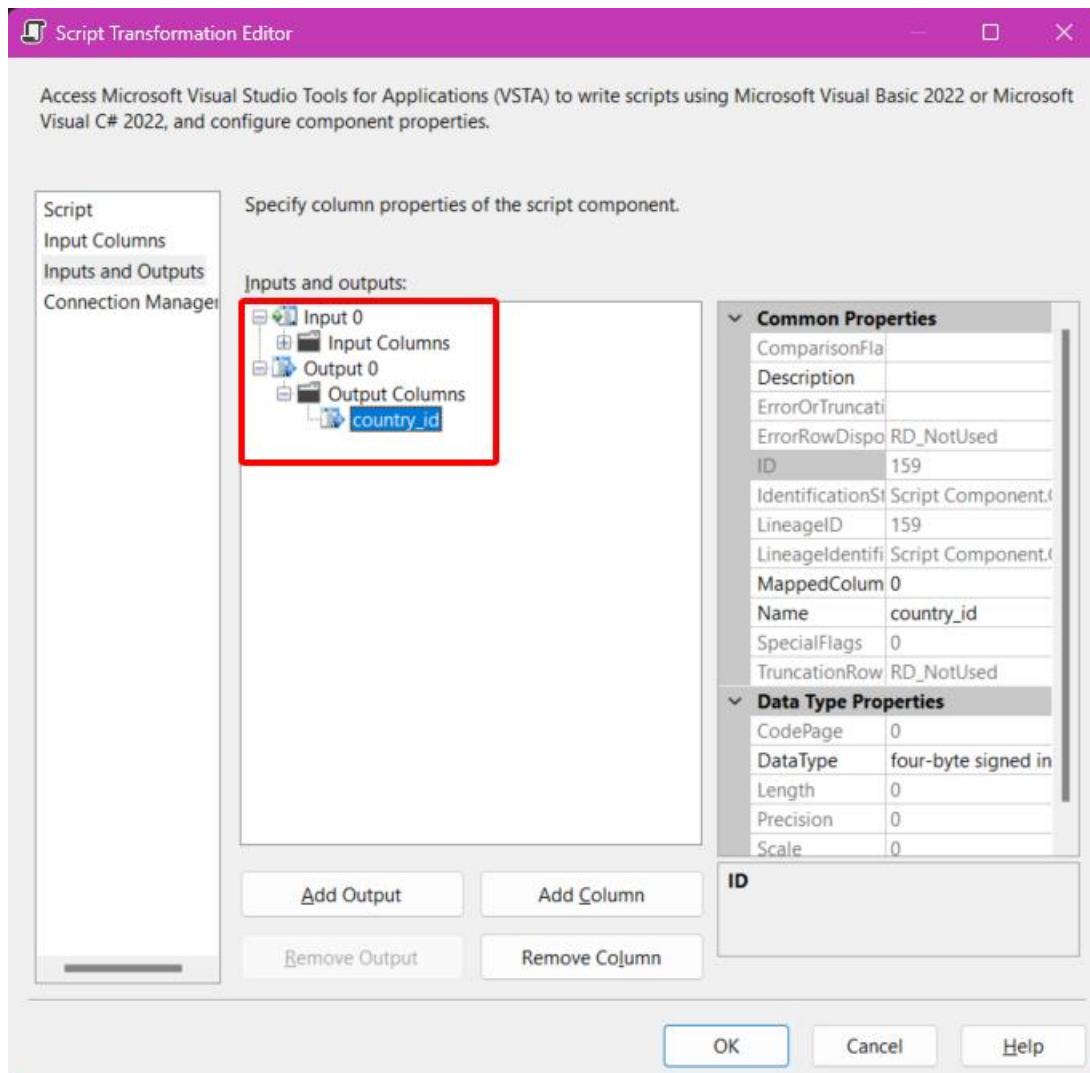
- Bước 9: Chọn **Edit** tại *Script Component*:

- Trong mục **Input Columns** chọn các cột thành phần cần sử dụng, ở đây ta chọn 3 cột *country_name*, *country_code* và *country_3_letter_code*.

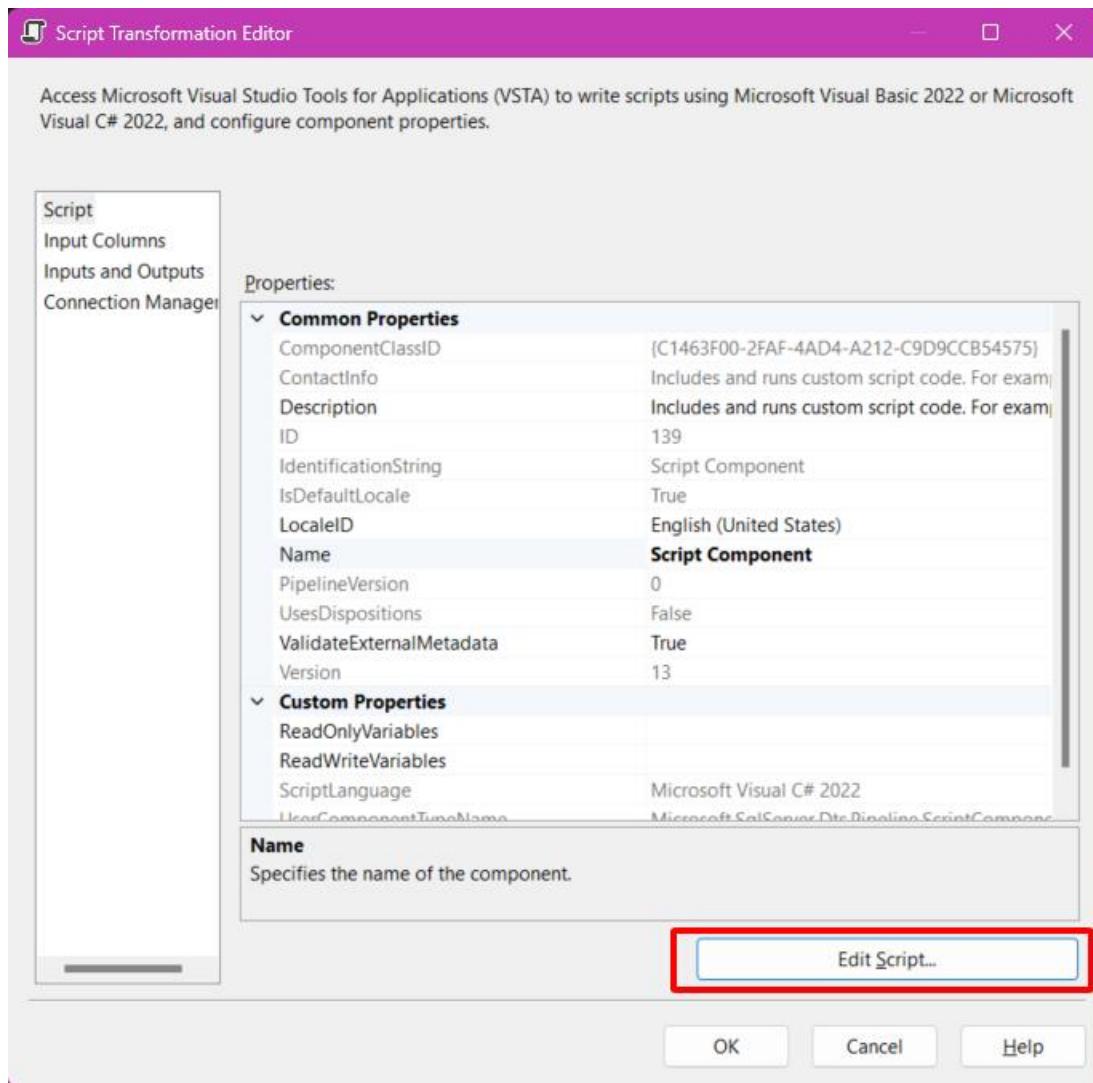


- Trong mục **Inputs và Outputs**, chọn **Add Column** tại Output 0 và đặt tên là *country_id*.





- Trong mục Script, chọn **Edit Script** để mở một cửa sổ chỉnh sửa đoạn mã, sau đó thêm các câu lệnh. Đoạn mã này dùng để tự động tăng giá trị *country_id* vừa thêm ở trên.



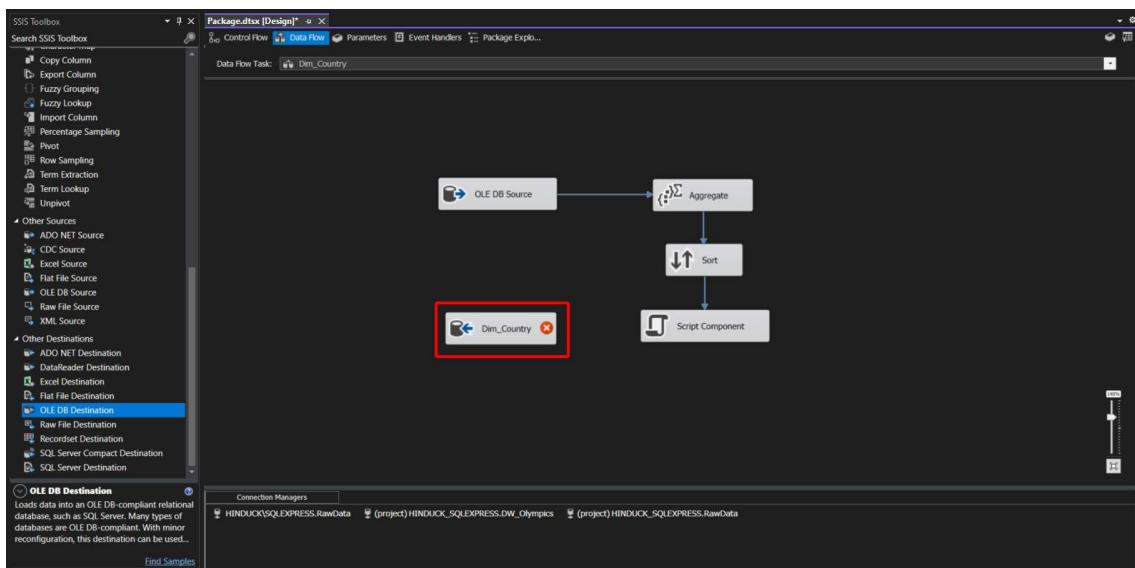
```
public class ScriptMain : UserComponent
{
    int count = 1;
    Help: Using Integration Services variables and parameters
```

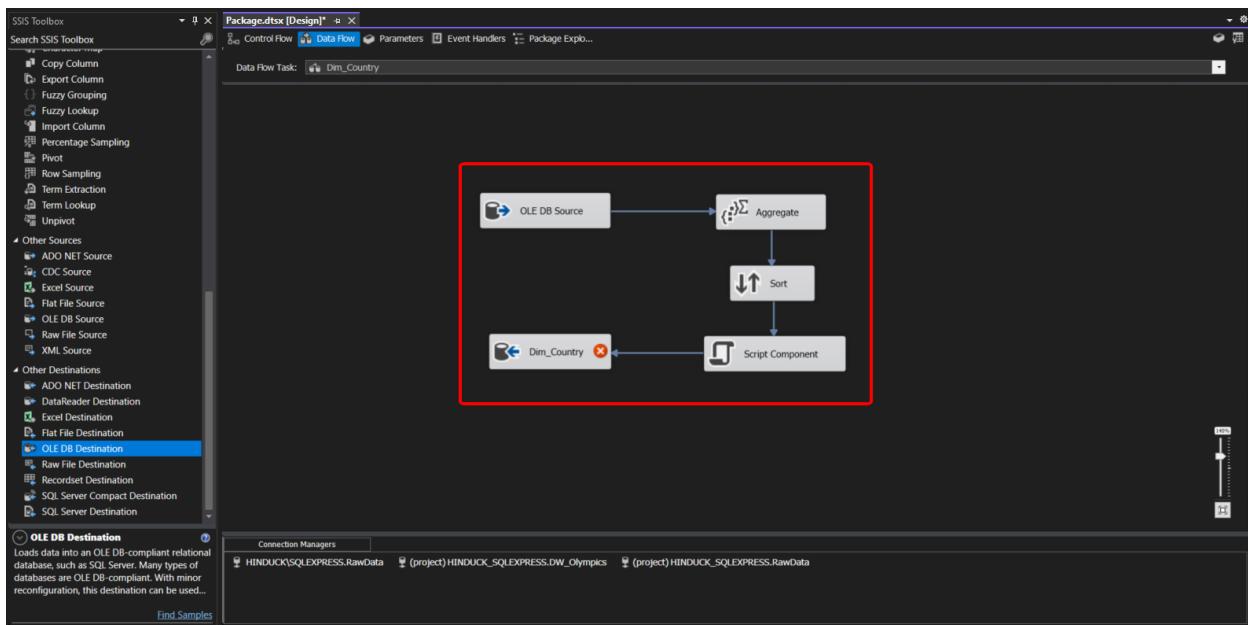
```

/// <summary>
/// This method is called once for every row that passes through the component from Input0.
///
/// Example of reading a value from a column in the 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</param>
2 references
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    Row.countryid = count;
    count++;
}

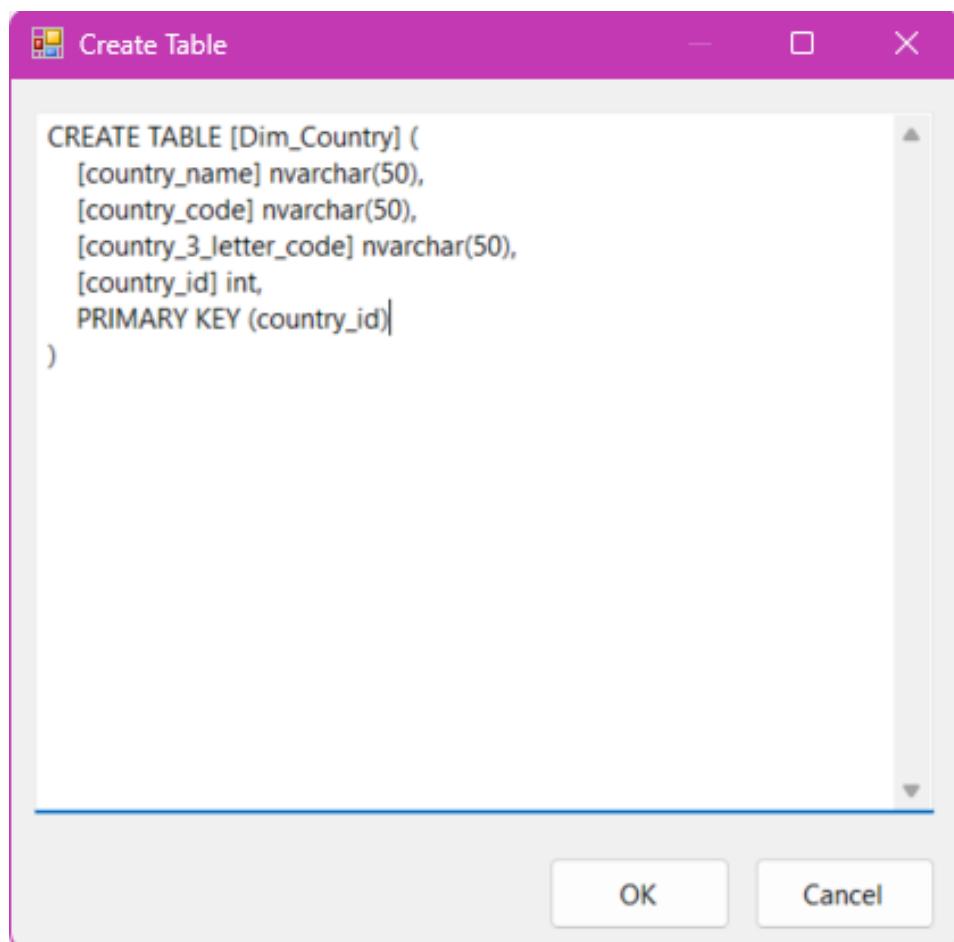
```

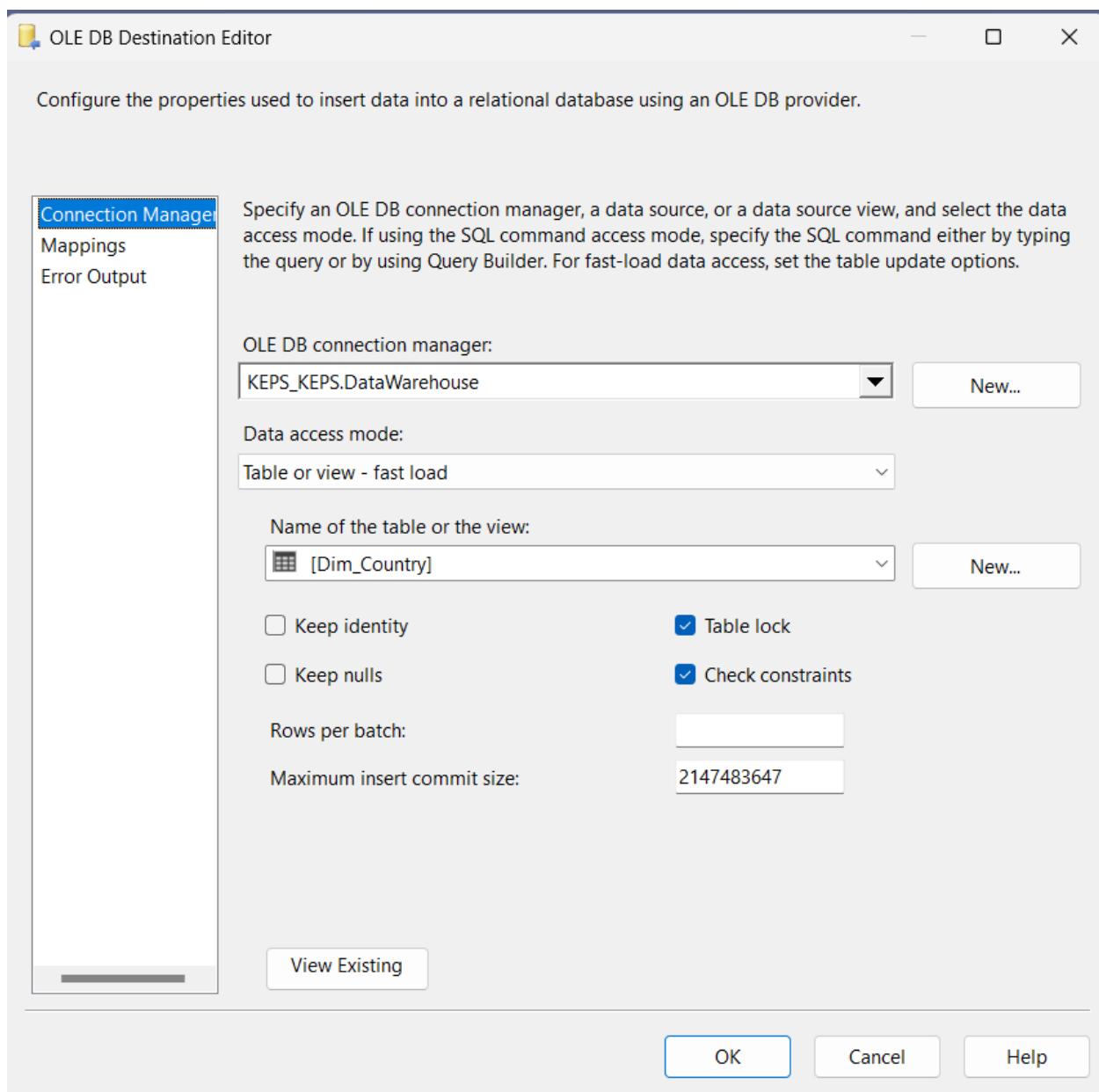
- Chọn **OK** để hoàn tất.
- Bước 10: Chọn bảng dữ liệu trong kho dữ liệu làm nguồn dữ liệu đến, sử dụng công cụ **OLE DB Destination**, được nối vào từ các tác vụ trên và đổi tên thành *Dim_Country*.

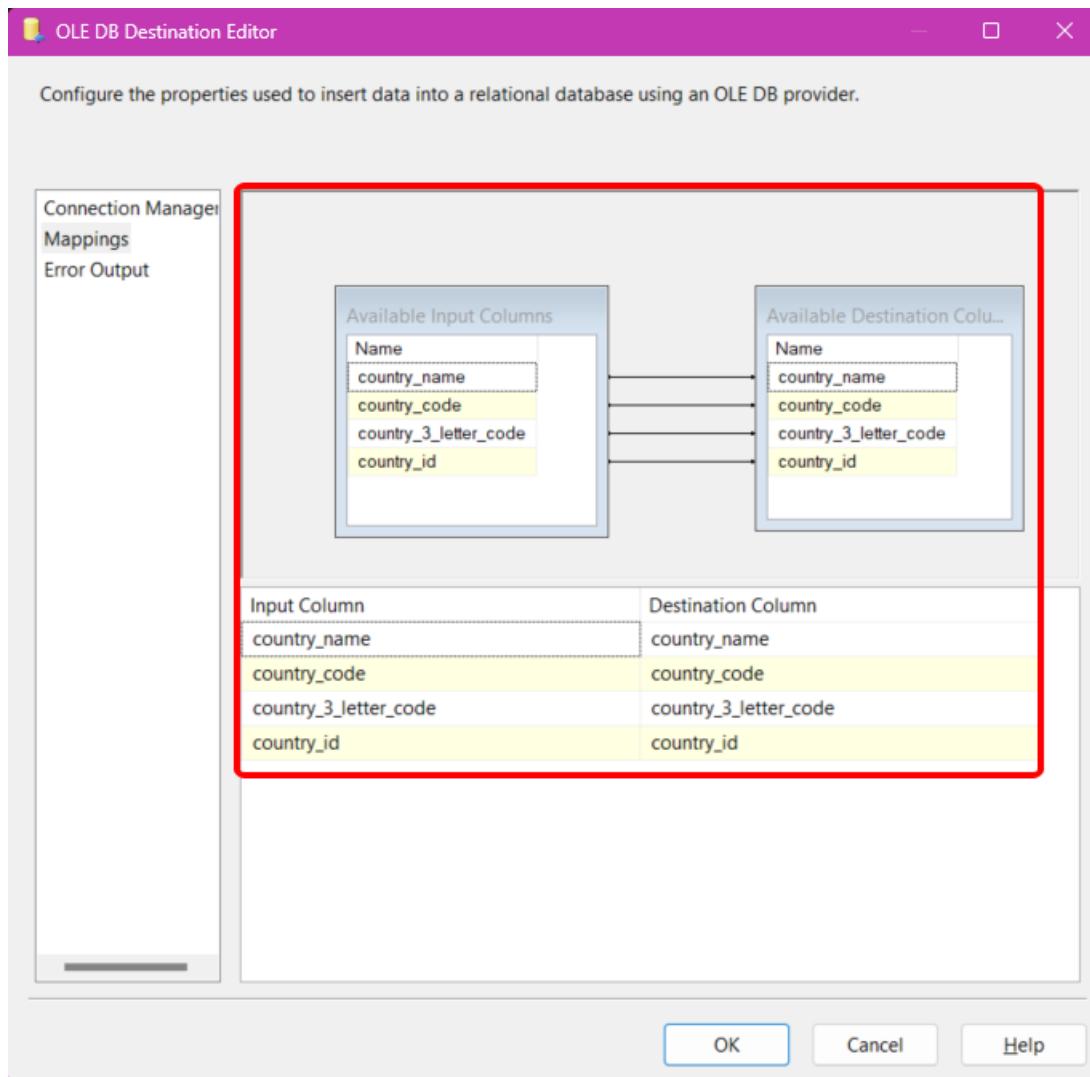




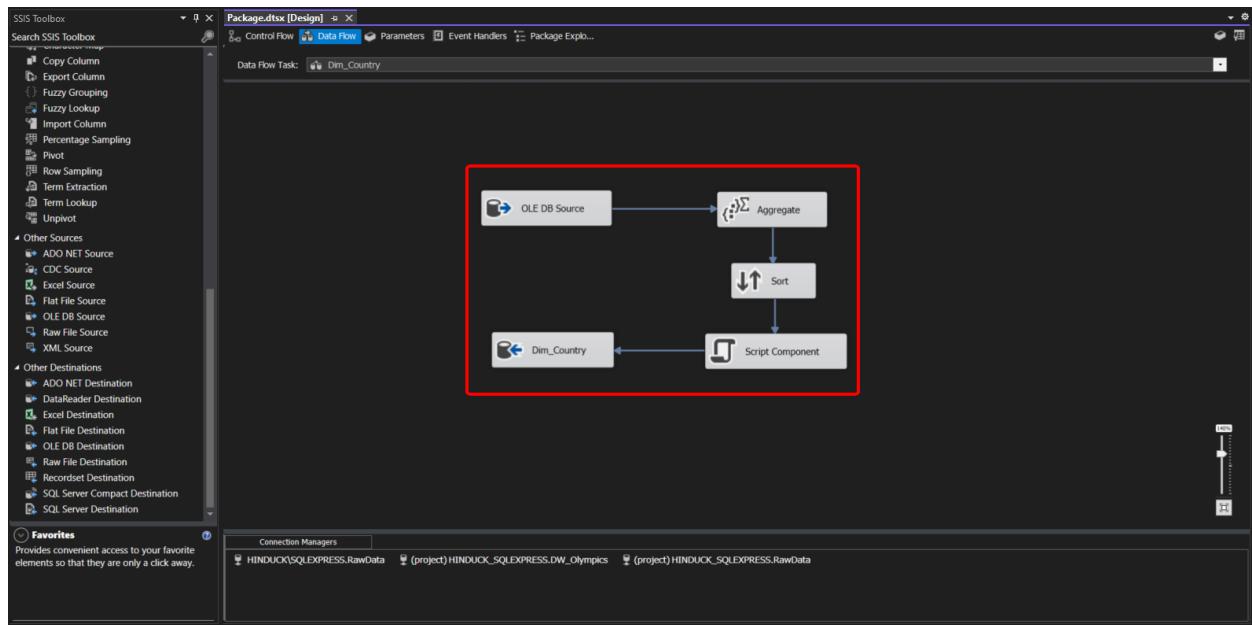
- Bước 11: Chọn **Edit** tại *Dim_Country*, chọn New để tạo bảng mới có tên *Dim_Country* trên kho dữ liệu *DataWarehouse* với thuộc tính khóa chính là *country_id* và thiết lập Connection Manager và Mappings trên bảng *Dim_Country*.







2.

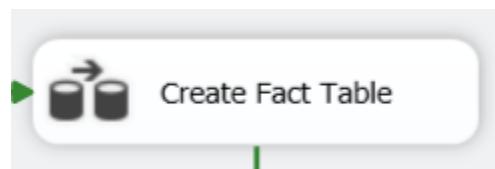


→ Thực hiện tương tự với các bảng Dimension khác

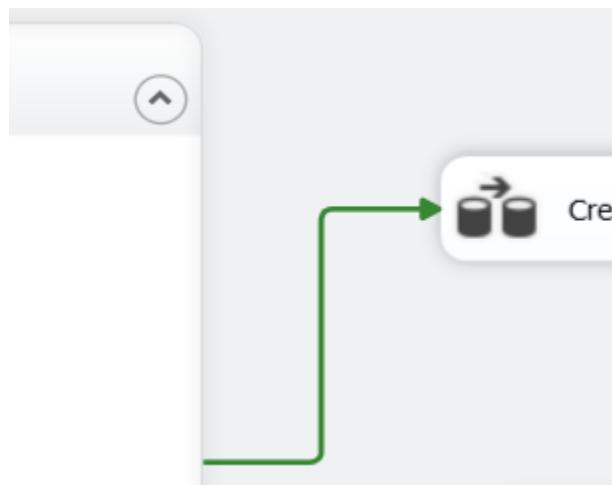
2.6. Quá trình khởi tạo bảng FACT

Để thực hiện quá trình khởi tạo bảng Fact, ta thực hiện các bước sau:

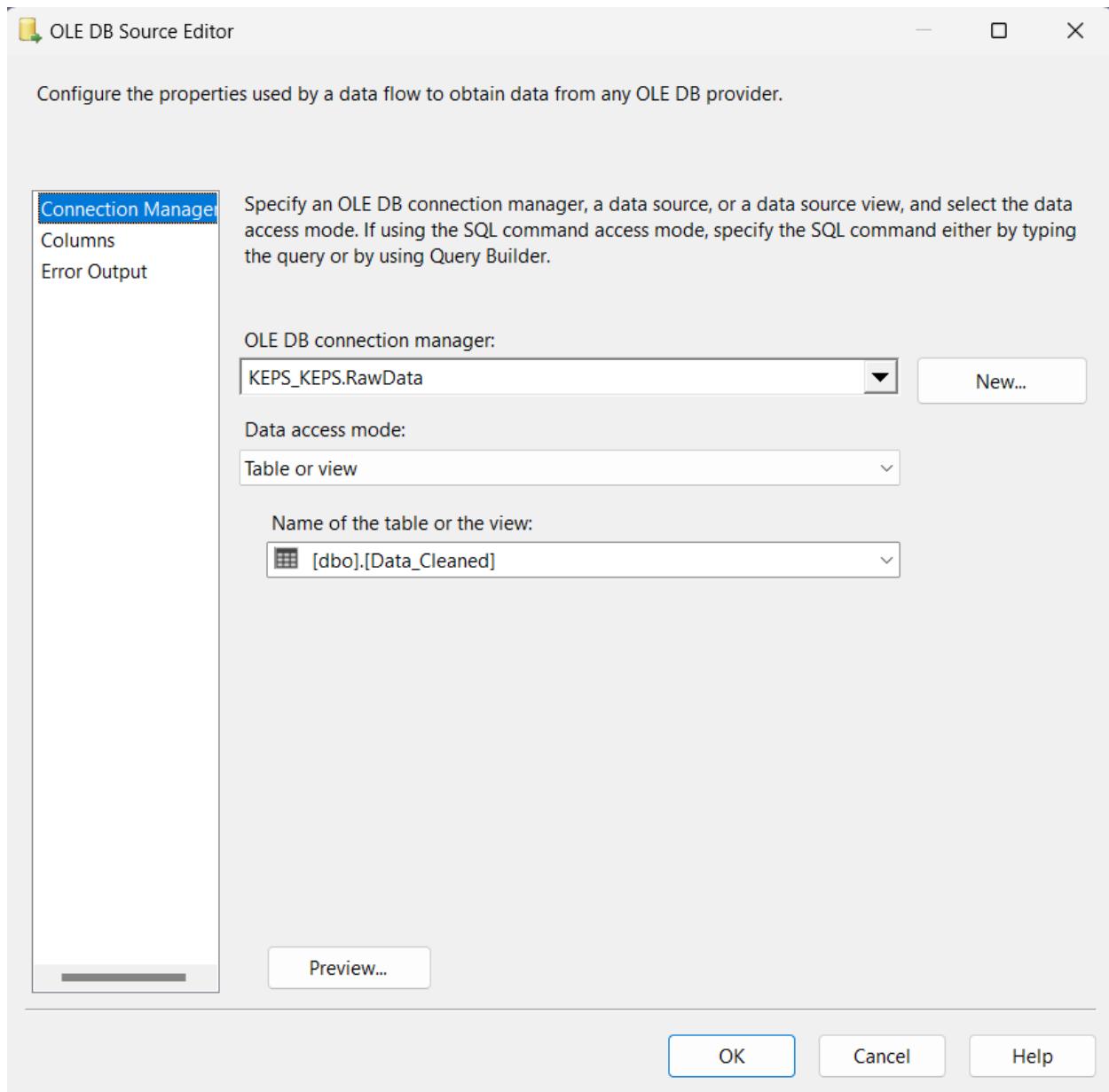
- Bước 1: Tại cửa sổ Control Flow, thực hiện việc kéo thả công cụ Data Flow Task tại SSIS Toolbox vào cửa sổ, sau đó đổi tên quá trình thành Create Fact Table



- Bước 2: Nối các tác vụ trước vào quá trình tạo bảng Fact.

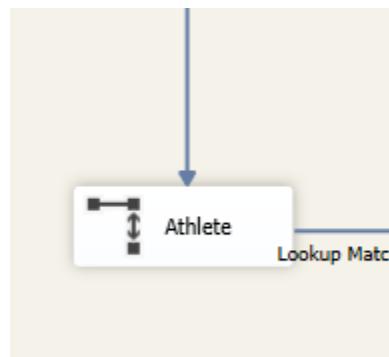


- Bước 3: Chọn **Edit** tại OLE DB Source để thiết lập cấu hình, tại mục Connection Manager, chọn OLE DB connection manager là **RawData** và chọn bảng **Data_Cleaned** và chọn **OK** để hoàn tất.

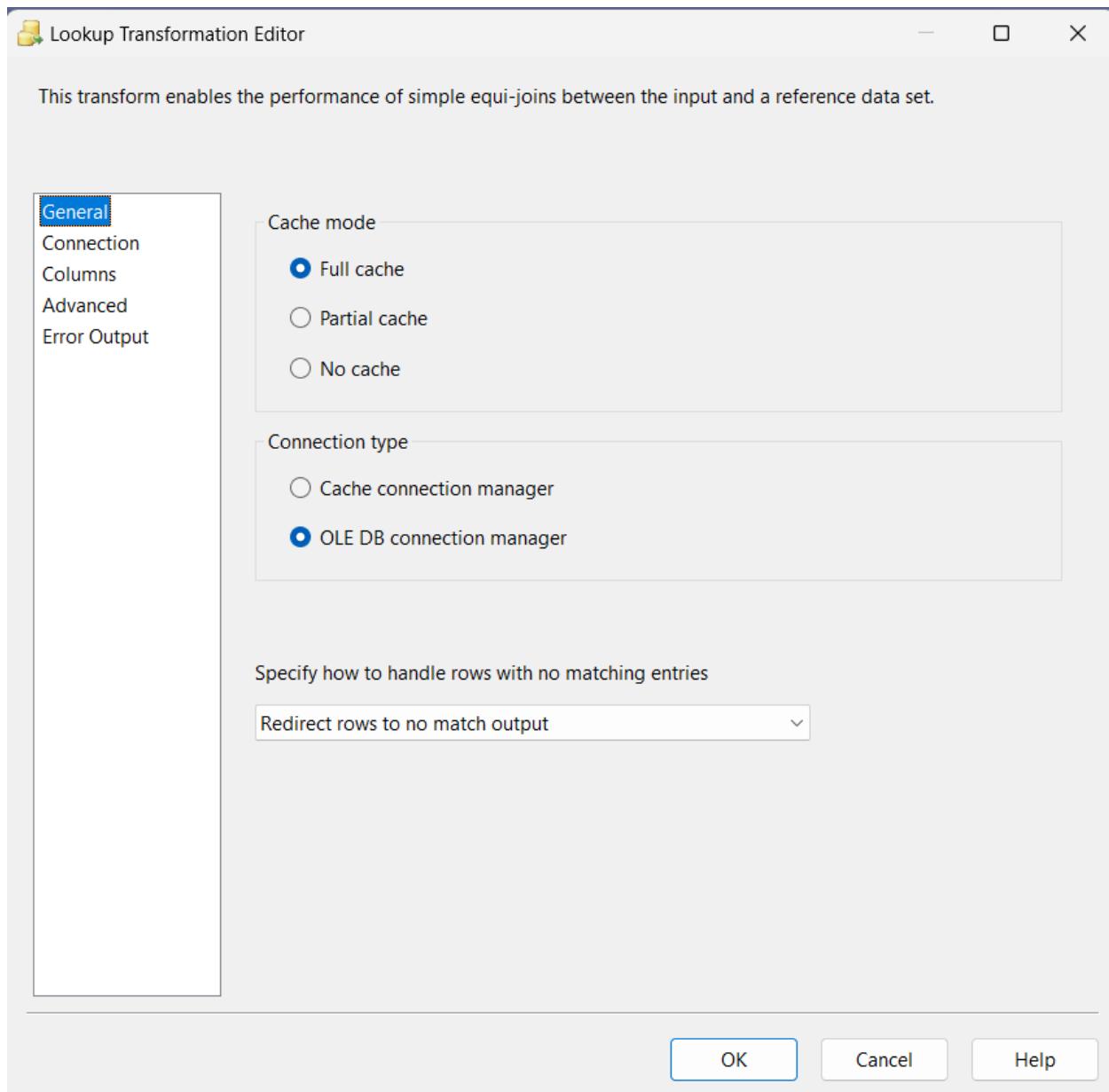


- Bước 4: Ta cần lấy dữ liệu trong các bảng Dimension, kéo thả để sử dụng công cụ **Lookup** trong SSIS Toolbox. Sử dụng công cụ này cho từng bảng Dimension.

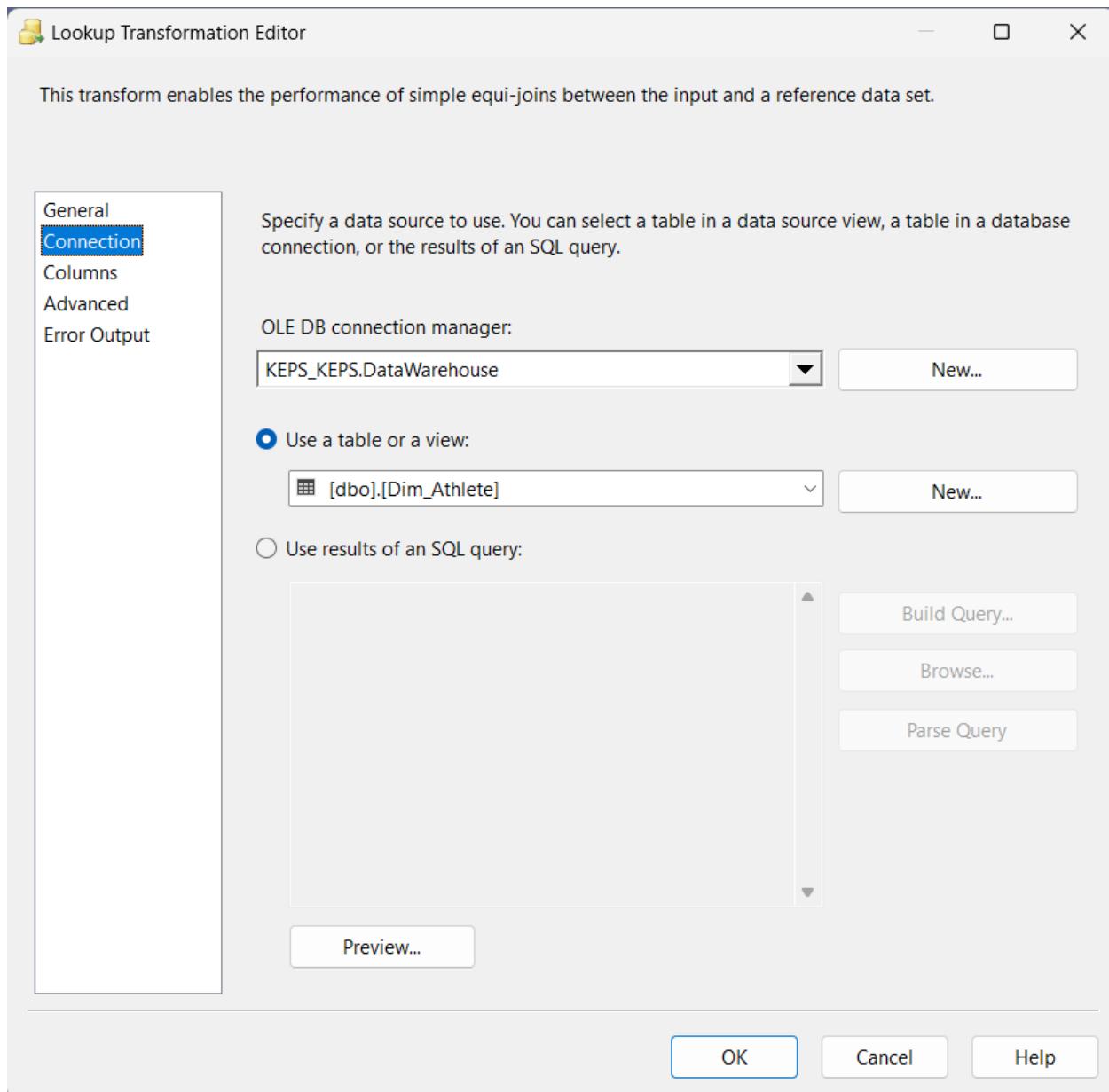
Dim_Athlete:



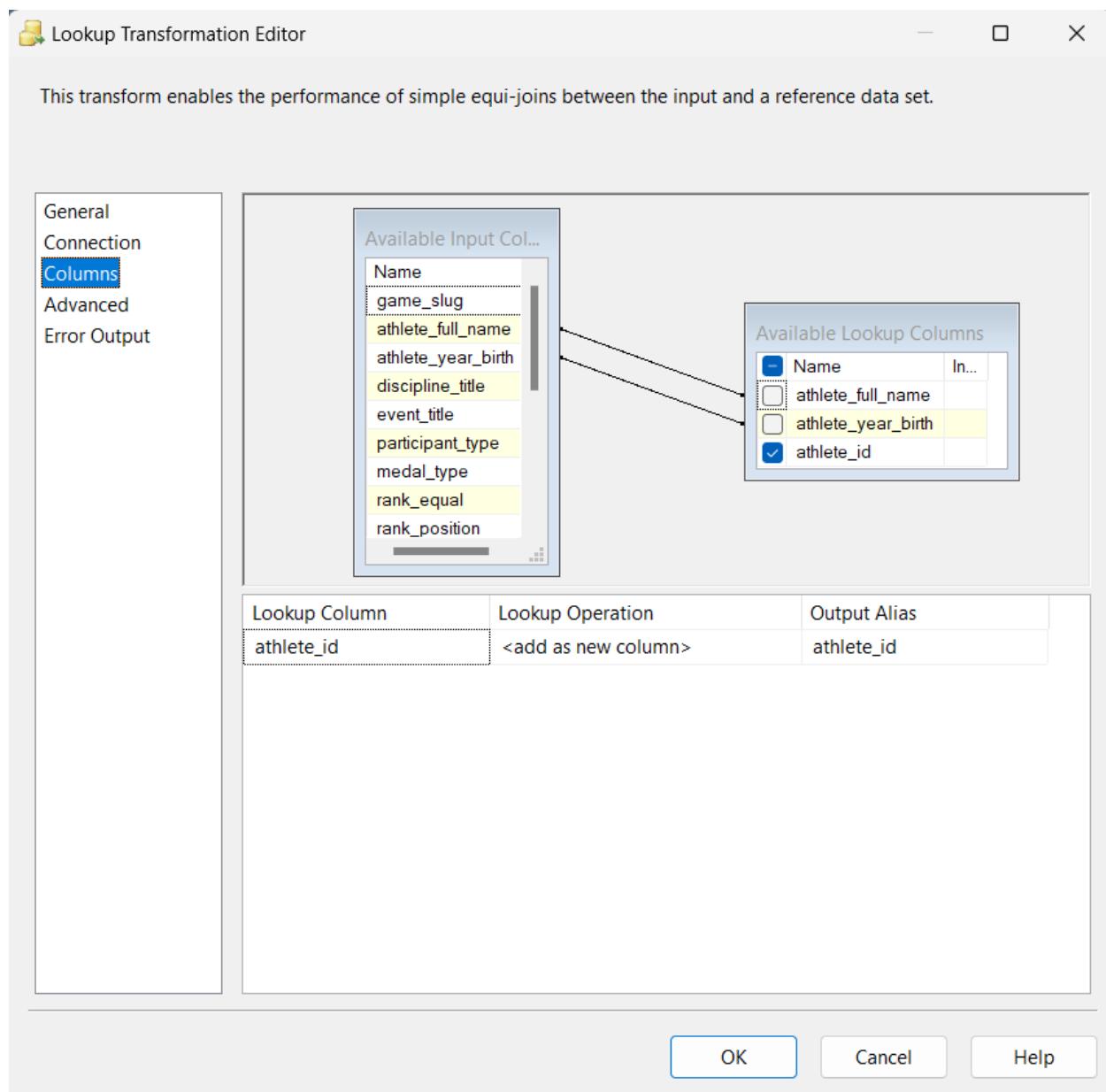
Chọn **Edit** tại *Lookup*, chọn các cột cần sử dụng để tạo bảng *Fact* trên Dim_Host. Tại General, ta chọn **Cache Mode** là *Full Cache*, **Connection Type** là *OLE DB Connection Manager*, **Specify how to handle rows with no matching entires** là *Redirect rows to no match output*



Tại **Connection**, ta chọn kết nối đến kho dữ liệu (database) *DataWarehouse* và tạo mới trên kho dữ liệu bảng *Dim_Athlete*

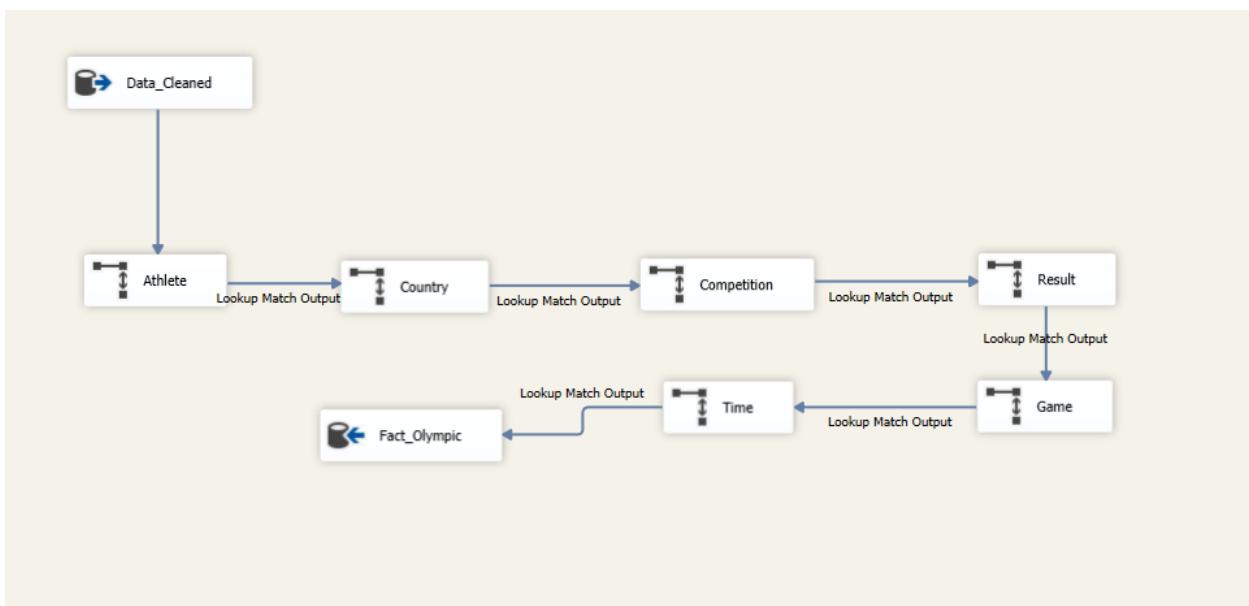
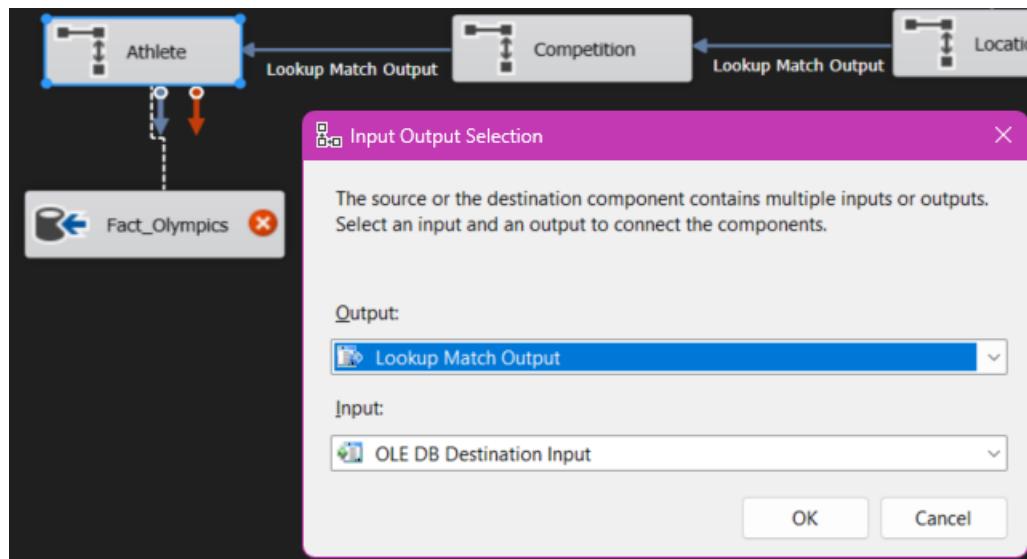


Tại **Columns**, ta kéo cột cần lookup sang kho dữ liệu từ Available Input Columns sang Available Lookup Columns, chọn *athlete_id* là **Lookup Column**. Sau đó chọn **OK** để hoàn tất.

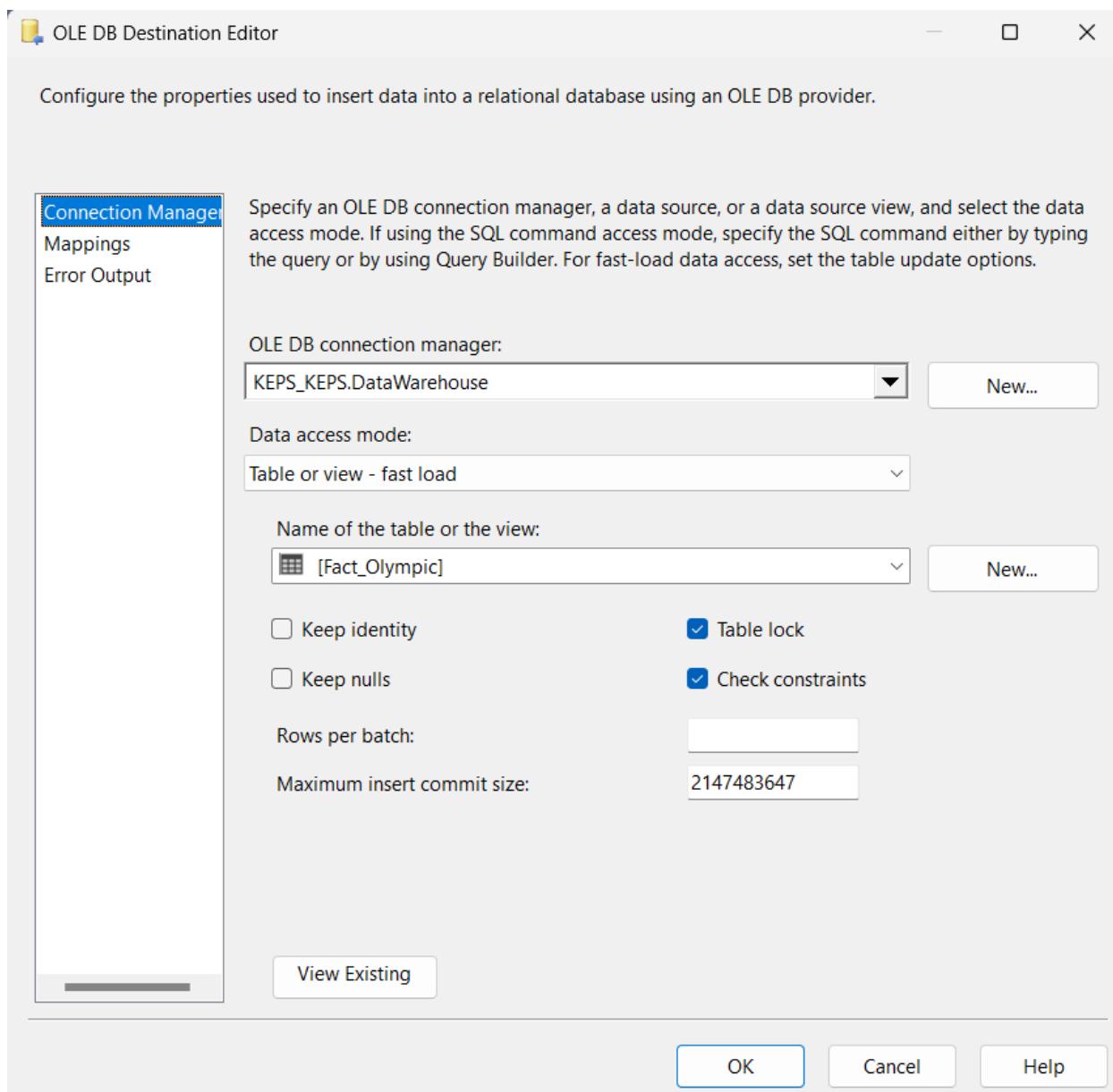


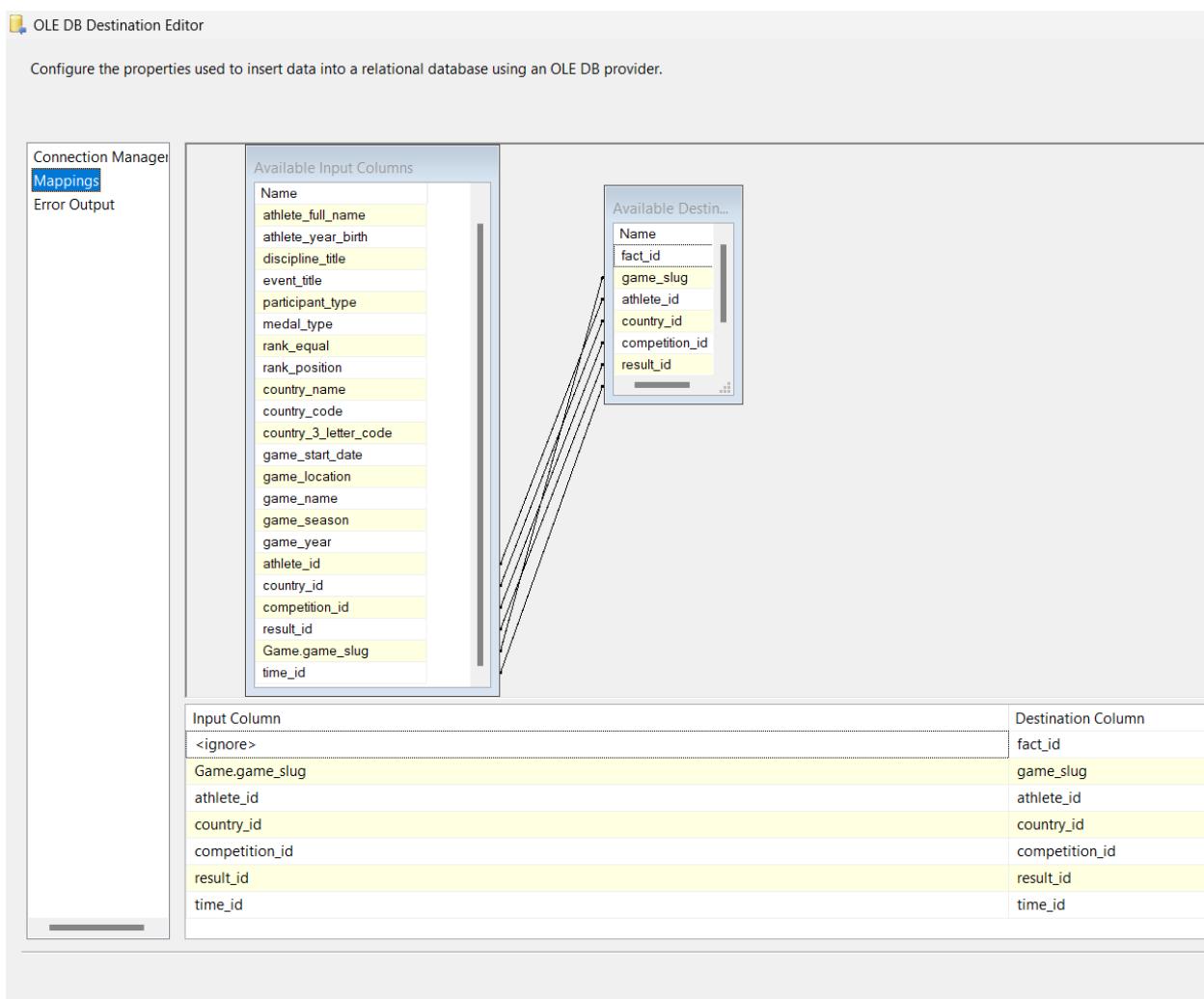
Thực hiện tương tự với các Dimension còn lại

- Bước 5: Chọn bảng dữ liệu trong kho dữ liệu làm nguồn dữ liệu đến, sử dụng công cụ **OLE DB Destination**, được nối vào từ các tác vụ trên và đổi tên thành *Fact_Olympic*



- Bước 6: Chọn **Edit** tại *Fact_Olympic*, chọn **New** để tạo bảng mới có tên *Fact_Olympic* với thuộc tính khóa chính là ***fact_id*** và thiết lập Connection Manager và Mappings trên bảng *Fact_Olympic*.

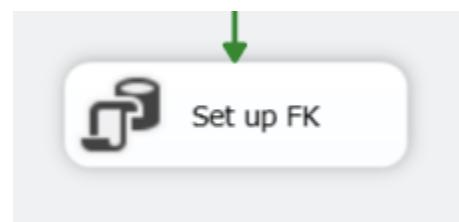




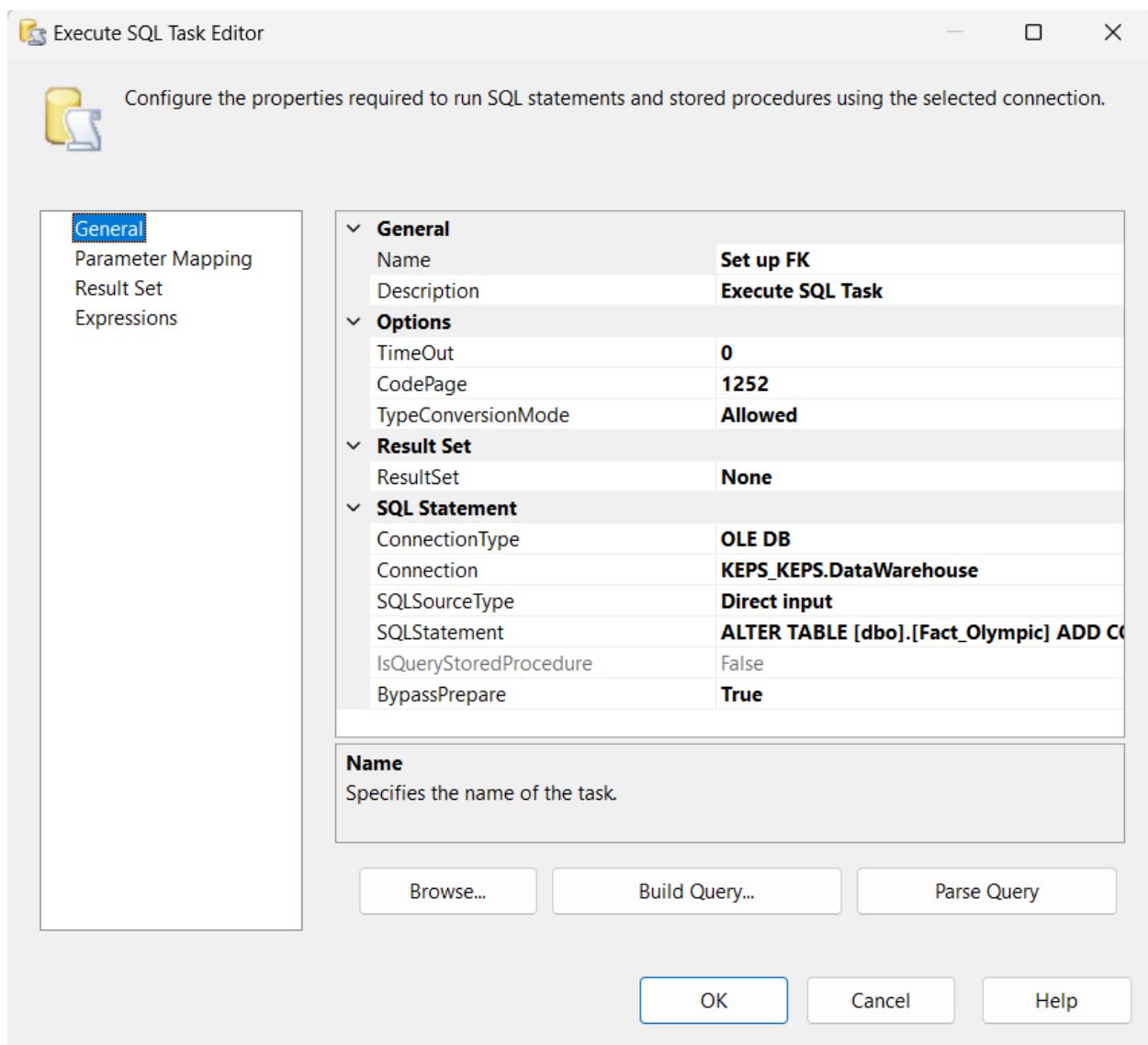
2.7. Quá trình khởi tạo các ràng buộc khoá ngoại giữa bảng Dimension và Fact

Để khởi tạo các ràng buộc khoá ngoại giữa các bảng Dimension và bảng Fact, ta thực hiện các bước sau:

- Bước 1: Tại cửa sổ **Control Flow**, thực hiện việc kéo thả công cụ **Execute SQL Task** tại SSIS Toolbox và nối vào tác vụ *Create Fact Tables*, sau đó đổi tên thành *Setup FK*.



- Bước 2: Chọn **Edit** tại quy trình *Setup FK Constraints* để thiết lập cấu hình, sẽ xuất hiện cửa sổ **Execute SQL Task Editor**, trong **SQL Statement** tại mục **Connection**, chọn kho dữ liệu *DataWarehouse*



- Bước 3: Thiết lập **SQLStatement**, nhập câu lệnh để tạo các ràng buộc khoá ngoại như sau:

```
#thêm FK
```

```
ALTER TABLE [dbo].[Fact_Olympic] ADD CONSTRAINT FK_FACT_ATHLETE
FOREIGN KEY ([athlete_id])
REFERENCES [dbo].Dim_Athlete ([athlete_id]);
```

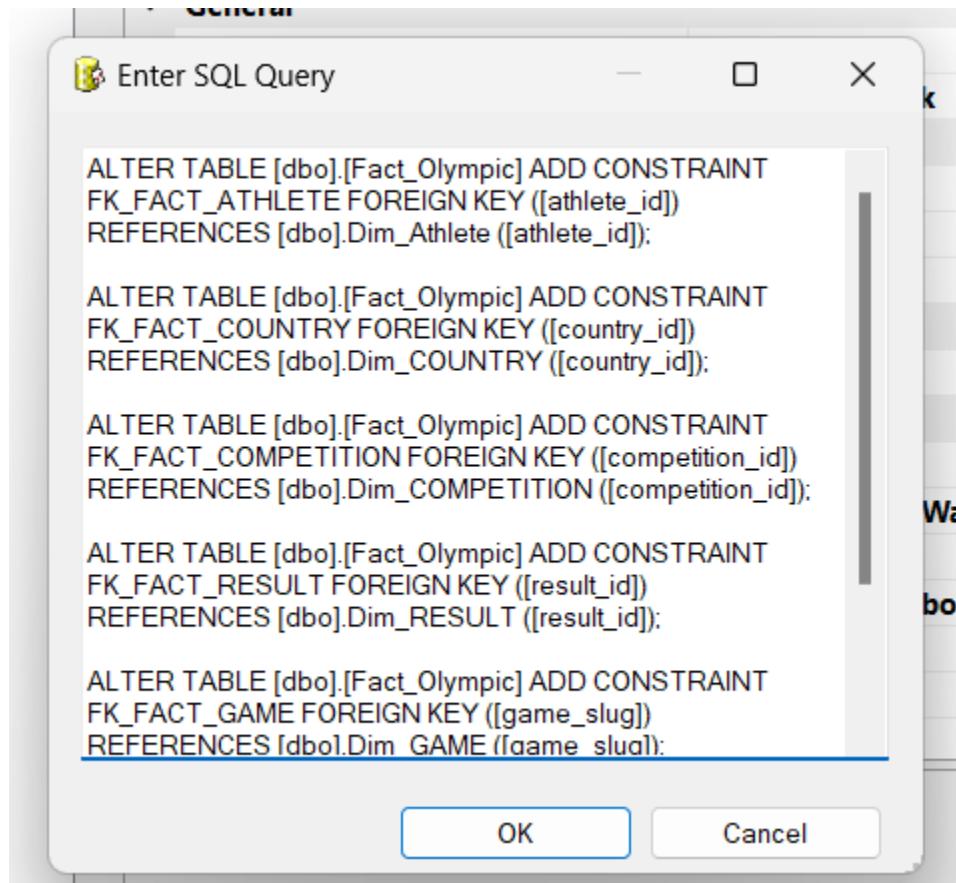
```
ALTER TABLE [dbo].[Fact_Olympic] ADD CONSTRAINT FK_FACT_COUNTRY
FOREIGN KEY ([country_id])
REFERENCES [dbo].Dim_COUNTRY ([country_id]);
```

```
ALTER TABLE [dbo].[Fact_Olympic] ADD CONSTRAINT
FK_FACT_COMPETITION FOREIGN KEY ([competition_id])
REFERENCES [dbo].Dim_COMPETITION ([competition_id]);
```

```
ALTER TABLE [dbo].[Fact_Olympic] ADD CONSTRAINT FK_FACT_RESULT
FOREIGN KEY ([result_id])
REFERENCES [dbo].Dim_RESULT ([result_id]);
```

```
ALTER TABLE [dbo].[Fact_Olympic] ADD CONSTRAINT FK_FACT_GAME
FOREIGN KEY ([game_slug])
REFERENCES [dbo].Dim_GAME ([game_slug]);
```

```
ALTER TABLE [dbo].[Fact_Olympic] ADD CONSTRAINT FK_FACT_TIME
FOREIGN KEY ([time_id])
REFERENCES [dbo].Dim_TIME ([time_id]);
```



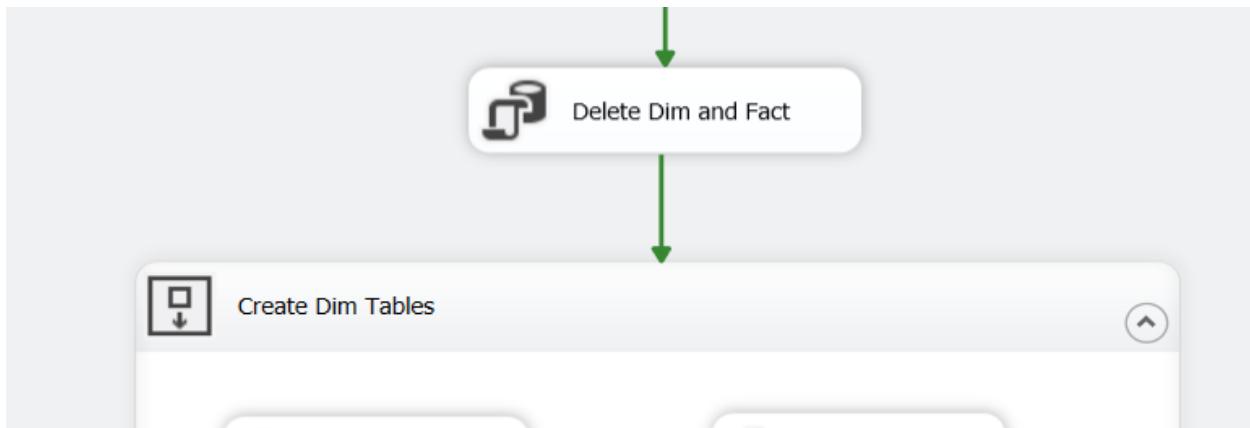
- Bước 4: Chọn **OK** để hoàn tất.

2.8. Quá trình xoá các bảng Dimension và Fact

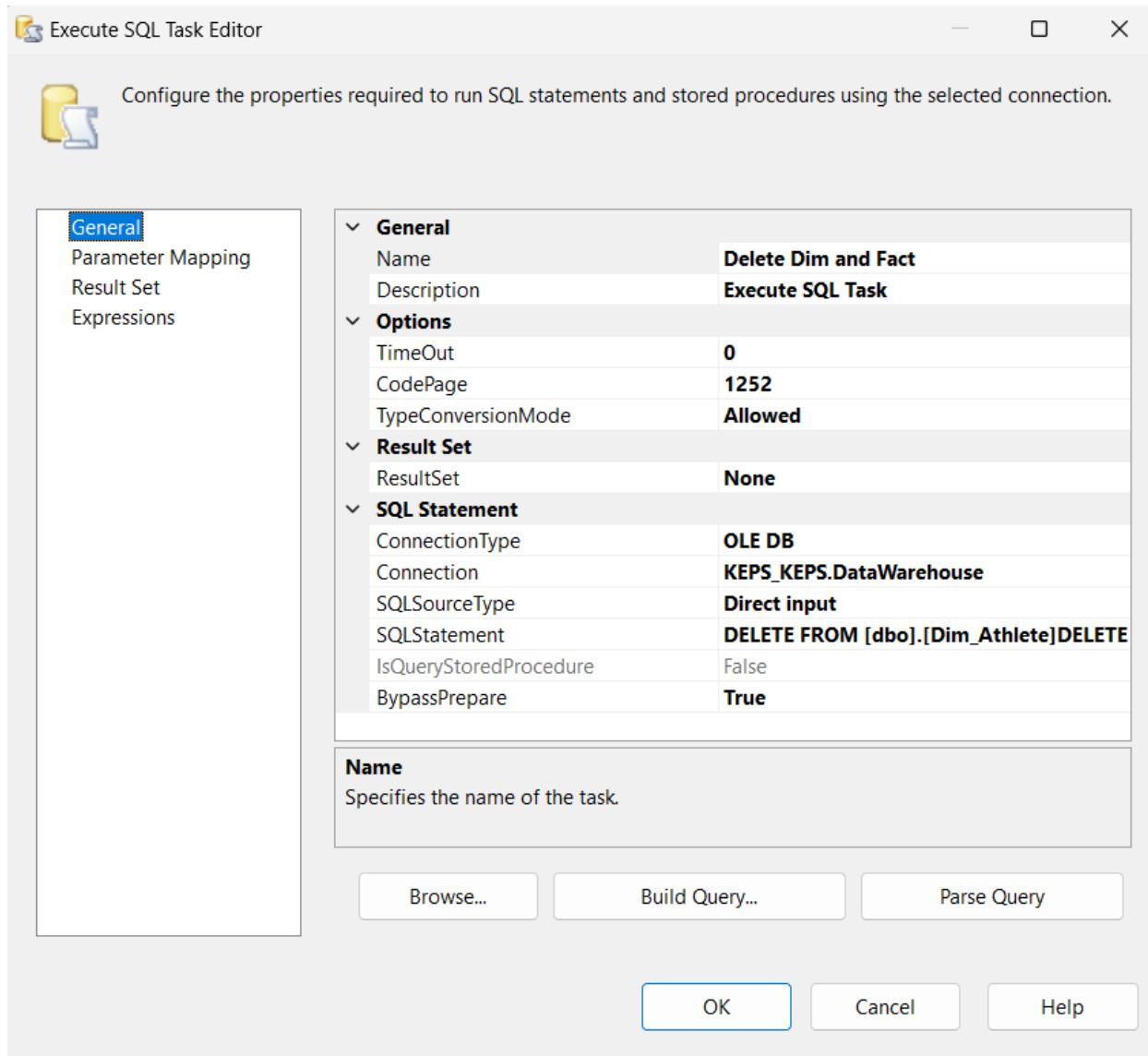
Sau mỗi lần thực thi việc đưa dữ liệu vào các bảng Dimension và Fact sẽ bị ghi đè dữ liệu. Do đó, ta cần loại bỏ các bảng Dimension và Fact cũ ra khỏi kho dữ liệu. Để loại bỏ dữ

liệu trùng sau mỗi lần thực thi việc đưa dữ liệu vào những lần tiếp theo, ta thực hiện các bước sau:

- Bước 1: Tại cửa sổ **Control Flow**, thực hiện việc kéo thả công cụ **Execute SQL Task** tại SSIS Toolbox, sau đó đổi tên thành *Delete from Dimension and Fact tables*.



- Bước 2: Chọn **Edit** tại quy trình *Delete from Dimension and Fact tables* để thiết lập cấu hình, sẽ xuất hiện cửa sổ **Execute SQL Task Editor**, trong **SQL Statement** tại mục **Connection**, chọn kho dữ liệu *DataWarehouse*.



- Bước 3: Thiết lập **SQLStatement**, nhập câu lệnh để loại bỏ các bảng Dimension và Fact:

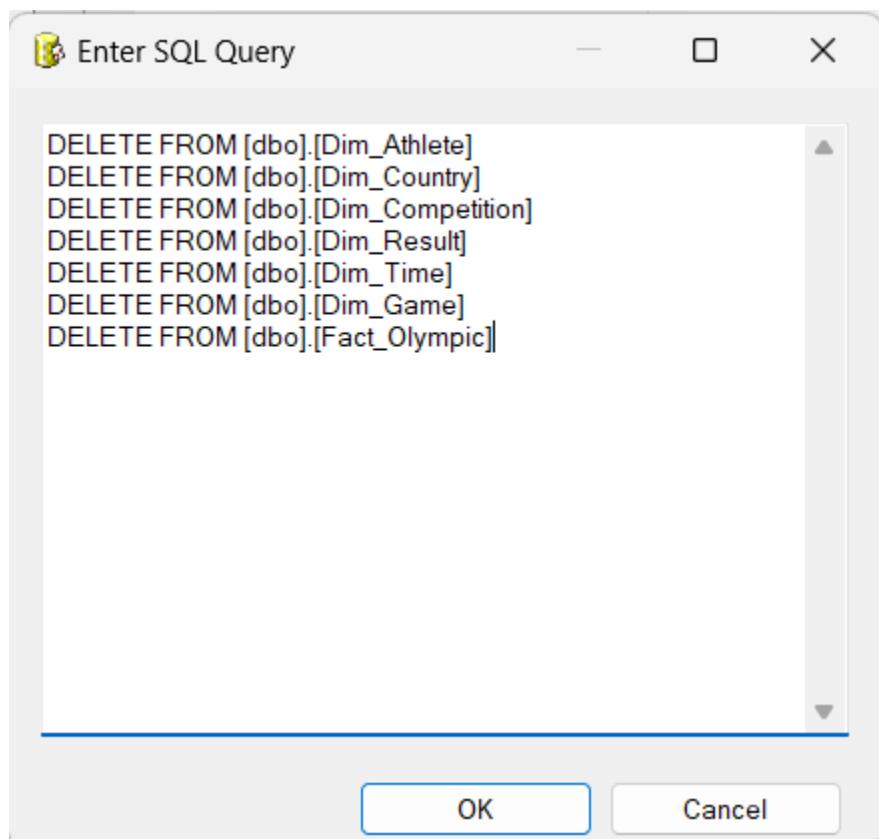
```
#Xóa Dim và Fact
```

```
DELETE FROM [dbo].[Dim_Athlete]
```

```
DELETE FROM [dbo].[Dim_Country]
```

```
DELETE FROM [dbo].[Dim_Competition]
```

```
DELETE FROM [dbo].[Dim_Result]  
DELETE FROM [dbo].[Dim_Time]  
DELETE FROM [dbo].[Dim_Game]  
DELETE FROM [dbo].[Fact_Olympic]
```



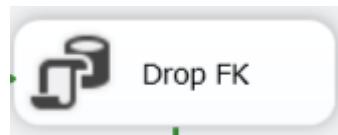
- Bước 4: Nối quá trình *Delete from Dimension and Fact tables* vào sau quá trình làm sạch dữ liệu, sau đó nối quá trình loại bỏ các bảng Dimension và Fact vào quá trình tạo bảng Dimension



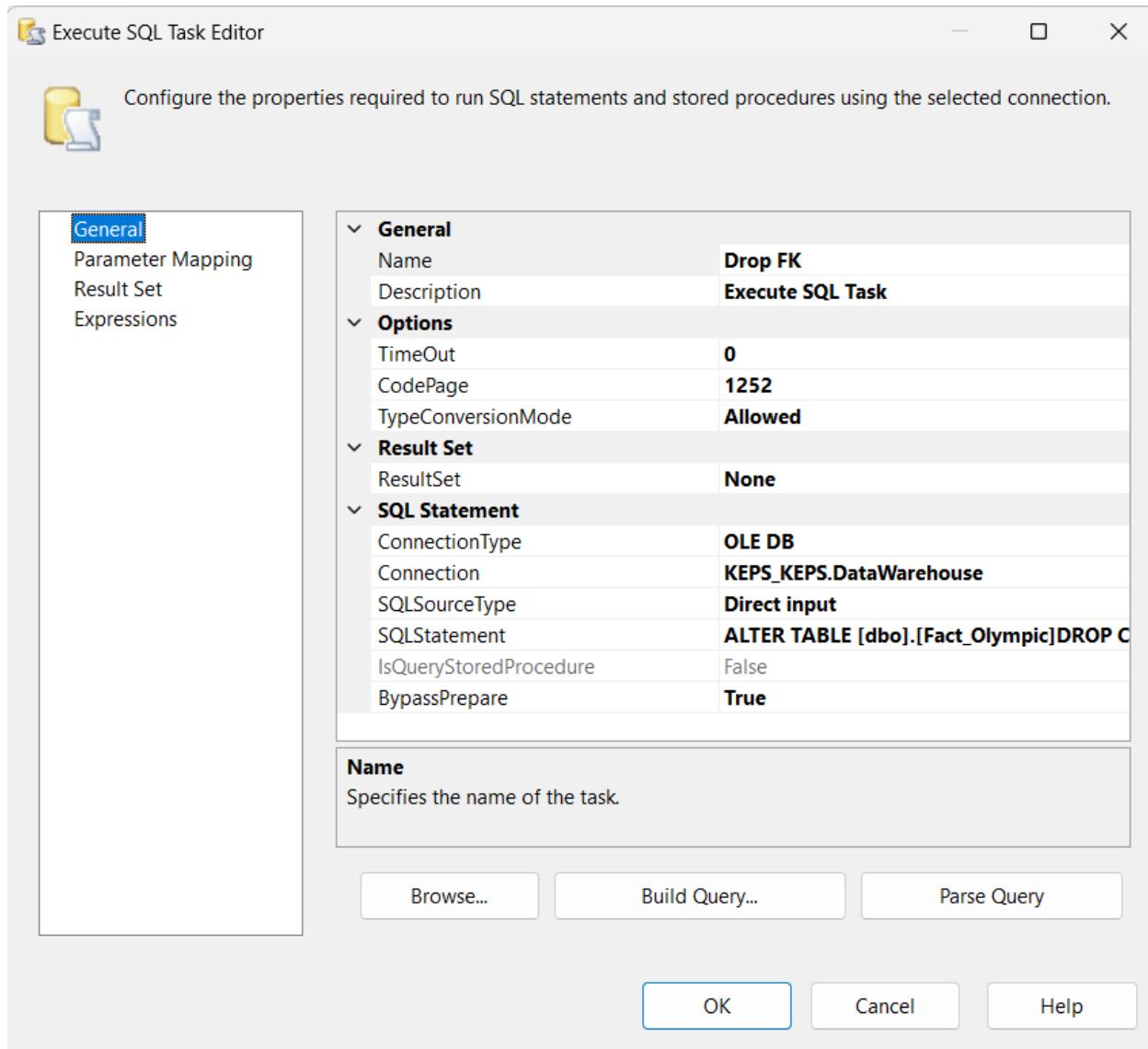
2.9. Quá trình xoá các ràng buộc khoá ngoại giữa các bảng Dimension và Fact

Trước khi thực hiện việc loại bỏ các bảng Dimension và Fact, ta cần loại bỏ các ràng buộc khoá ngoại giữa các bảng với nhau.. Để loại bỏ các ràng buộc khoá ngoại giữa các bảng Dimension và Fact, ta thực hiện các bước sau:

- Bước 1: Tại cửa sổ Control Flow, thực hiện việc kéo thả công cụ Execute SQL Task tại SSIS Toolbox, sau đó đổi tên thành Drop FK Constraints.



- Bước 2: Chọn **Edit** tại quy trình *Drop FK Constraints* để thiết lập cấu hình, sẽ xuất hiện cửa sổ **Execute SQL Task Editor**, trong **SQL Statement** tại mục **Connection**, chọn kho dữ liệu DataWarehouse



- Bước 3: Thiết lập **SQLStatement**, nhập câu lệnh để loại bỏ các ràng buộc khoá ngoại giữa các bảng Dimension và Fact:

#Xóa FK

```
ALTER TABLE [dbo].[Fact_Olympic]
```

```
DROP CONSTRAINT FK_FACT_ATHLETE
```

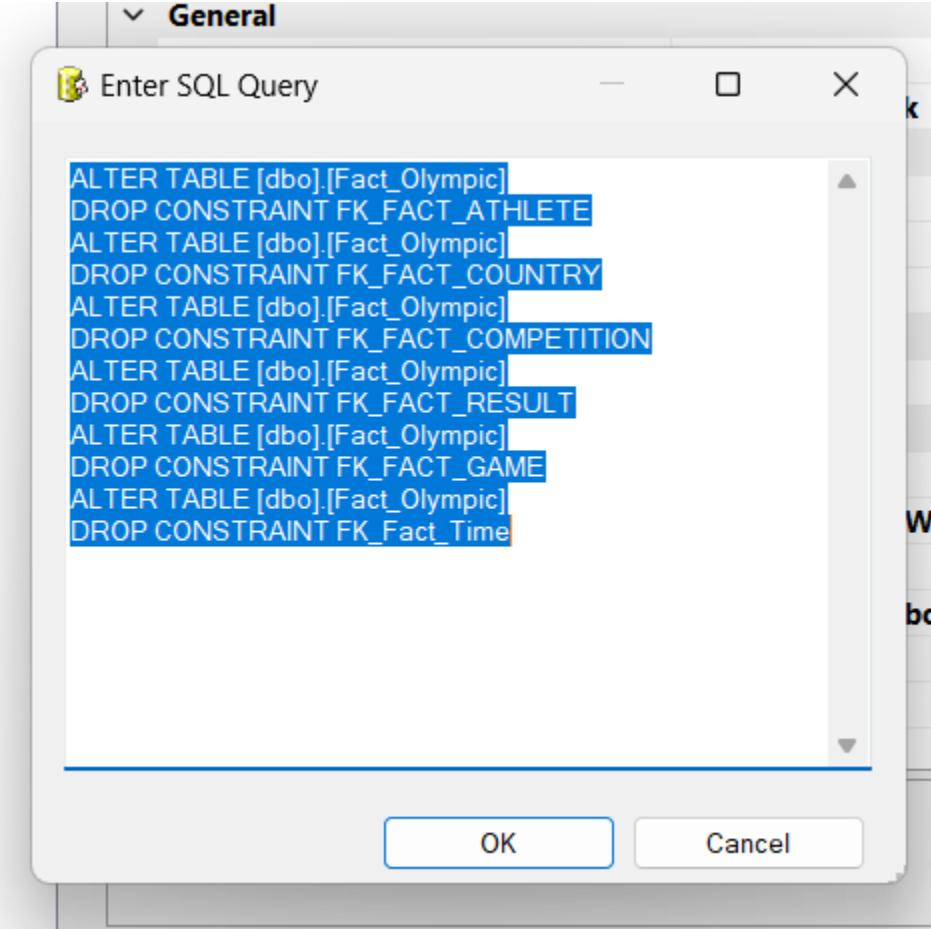
```
ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_COUNTRY

ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_COMPETITION

ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_RESULT

ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_GAME

ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_Fact_Time
```



```

ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_ATHLETE
ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_COUNTRY
ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_COMPETITION
ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_RESULT
ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_FACT_GAME
ALTER TABLE [dbo].[Fact_Olympic]
DROP CONSTRAINT FK_Fact_Time
  
```

The screenshot shows a Windows-style dialog box titled "General" with the sub-titile "Enter SQL Query". Inside the main area, there is a scrollable text box containing a series of SQL commands. These commands are used to drop various foreign key constraints from the "Fact_Olympic" table in a database. The constraints dropped include FK_FACT_ATHLETE, FK_FACT_COUNTRY, FK_FACT_COMPETITION, FK_FACT_RESULT, FK_FACT_GAME, and FK_Fact_Time.

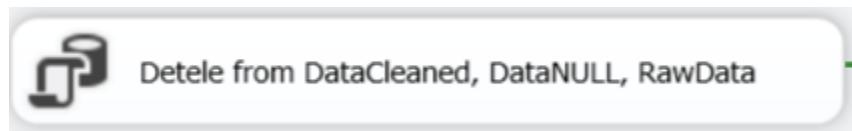
- Bước 4: Nối quá trình *Drop FK Constraints* vào sau quá trình làm sạch dữ liệu, sau đó nối quá trình loại bỏ các ràng buộc khoá ngoại trên các bảng Dimension và Fact vào trước quá trình xoá các bảng Dimension và Fact.



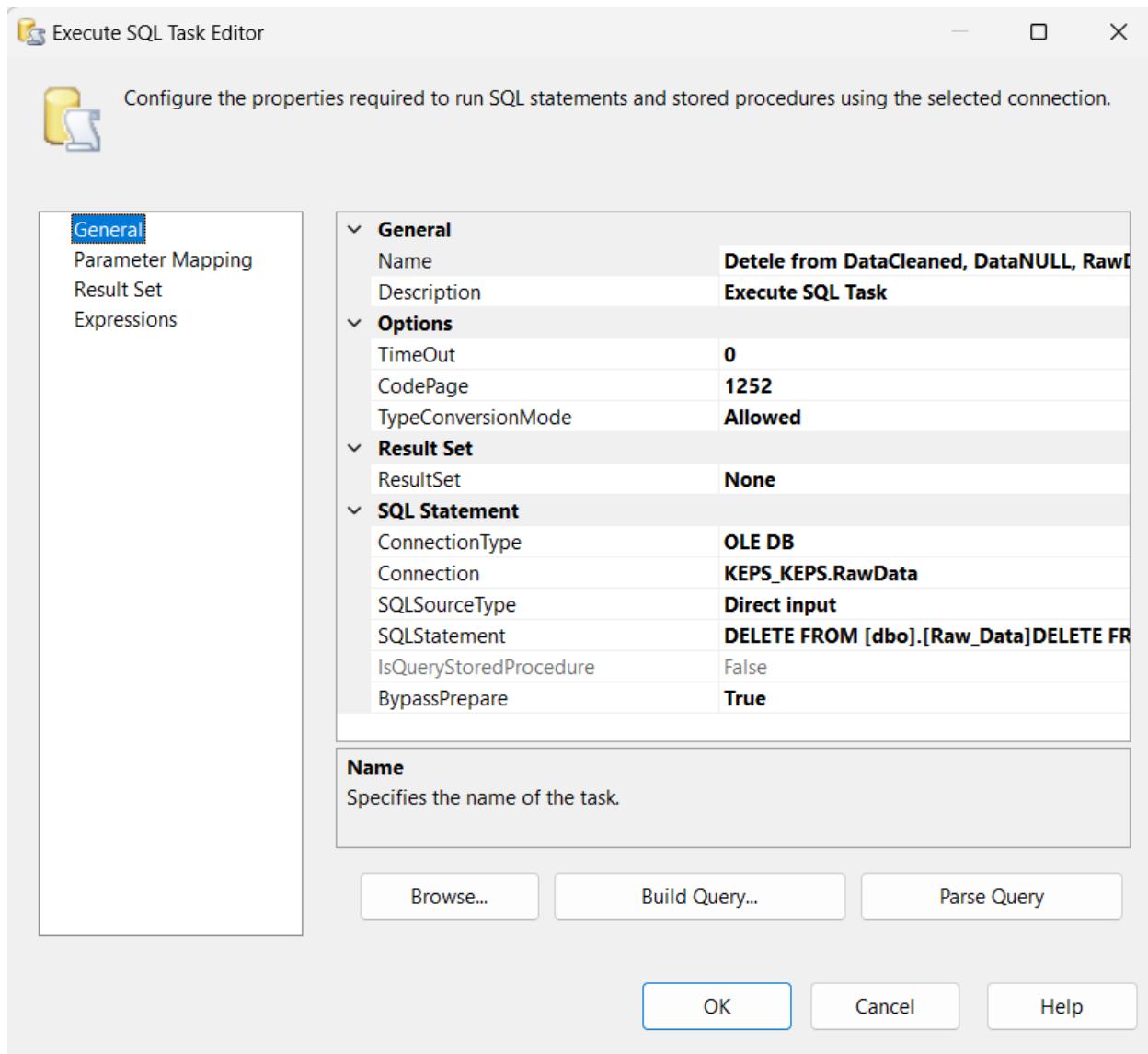
2.10. Quá trình loại bỏ dữ liệu trùng sau khi đưa dữ liệu vào

Sau mỗi lần thực thi việc đưa dữ liệu vào bảng Raw_Data, Data_Cleaned và Data_Null. Để loại bỏ dữ liệu trùng sau mỗi lần thực thi việc đưa dữ liệu vào những lần tiếp theo, ta thực hiện các bước sau:

- Bước 1: Tại cửa sổ Control Flow, thực hiện việc kéo thả công cụ Execute SQL Task tại SSIS Toolbox, sau đó đổi tên thành Delete from DataCleaned and DataNull tables.



- Bước 2: Chọn **Edit** tại quy trình *Delete from DataCleaned and DataNull tables* để thiết lập cấu hình, sẽ xuất hiện cửa sổ **Execute SQL Task Editor**, trong **SQL Statement** tại mục **Connection**, chọn kho dữ liệu *RawData*.



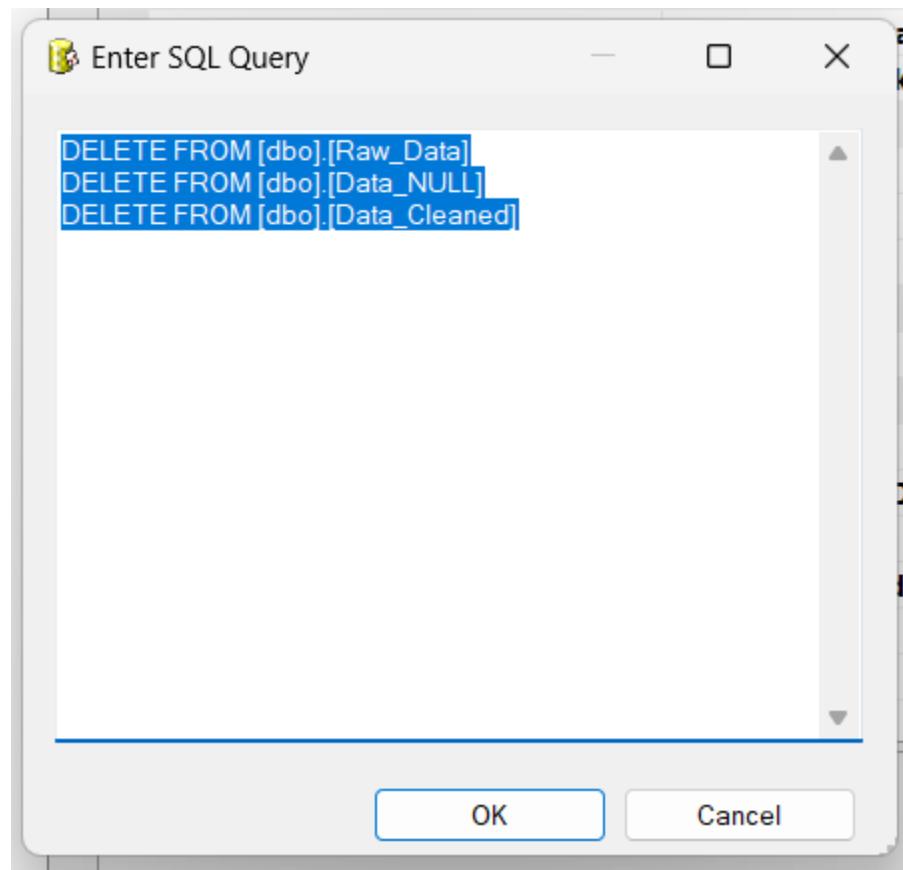
- Bước 3: Thiết lập **SQLStatement**, nhập câu lệnh để loại bỏ dữ liệu trùng lặp sau:

#Xóa data

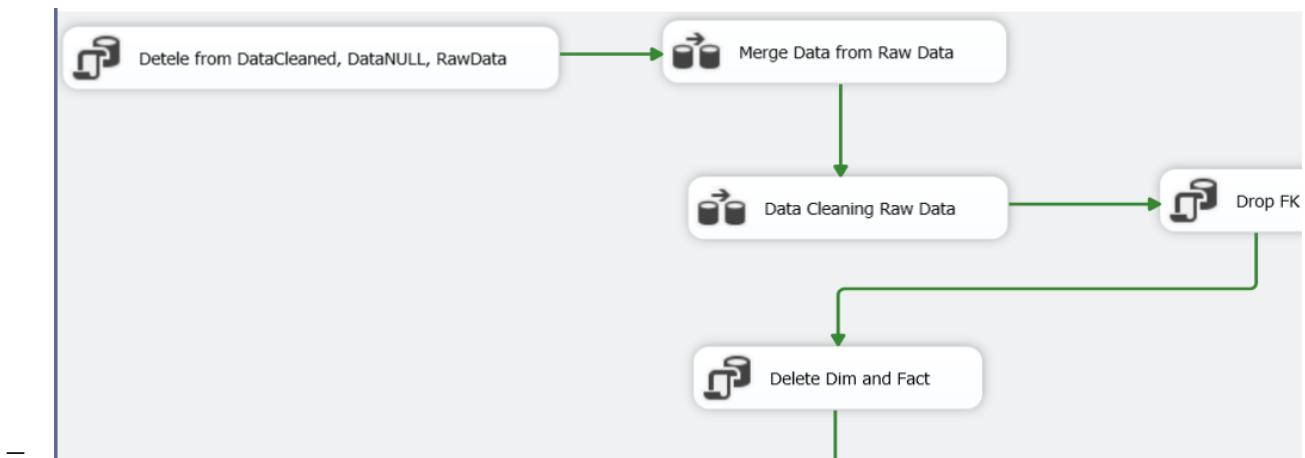
```
DELETE FROM [dbo].[Raw_Data]
```

```
DELETE FROM [dbo].[Data_NULL]
```

```
DELETE FROM [dbo].[Data_Cleaned]
```



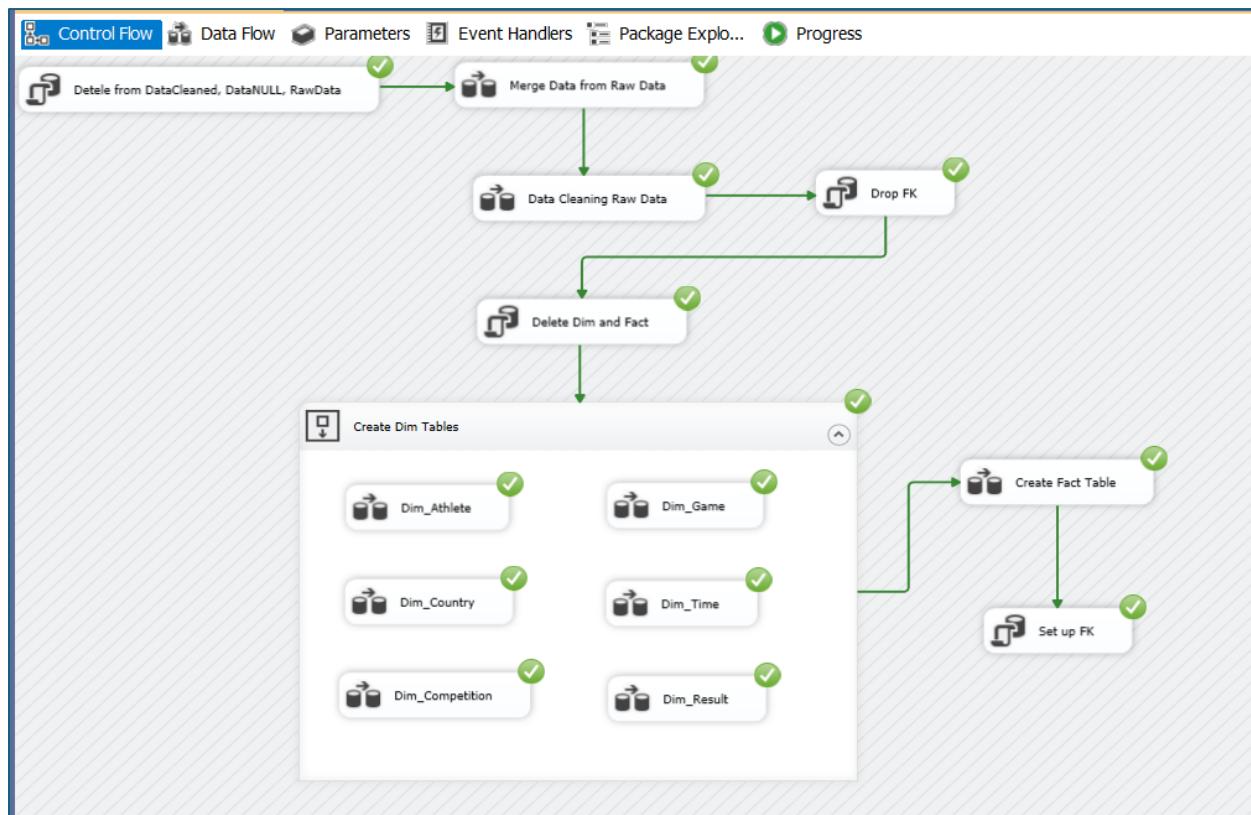
- Bước 4: Nối quá trình *Delete from DataCleaned and DataNull tables* vào sau quá trình nối dữ liệu từ các bảng, sau đó nối quá trình loại bỏ dữ liệu trùng trên vào quá trình làm sạch dữ liệu.



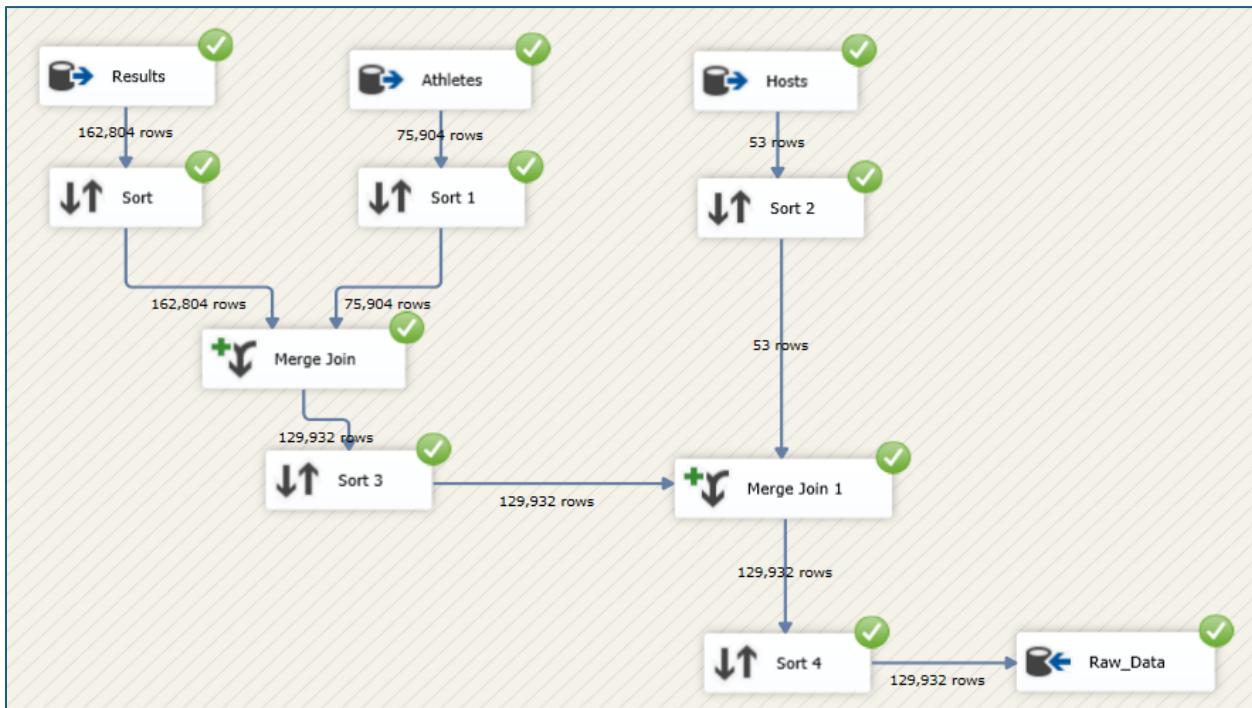
3. KẾT QUẢ:

Để thực thi dòng dữ liệu để đưa vào kho dữ liệu, ta chọn nút Start ở thanh công cụ hoặc nhấn nút F5 để bắt đầu khởi chạy dòng dữ liệu.

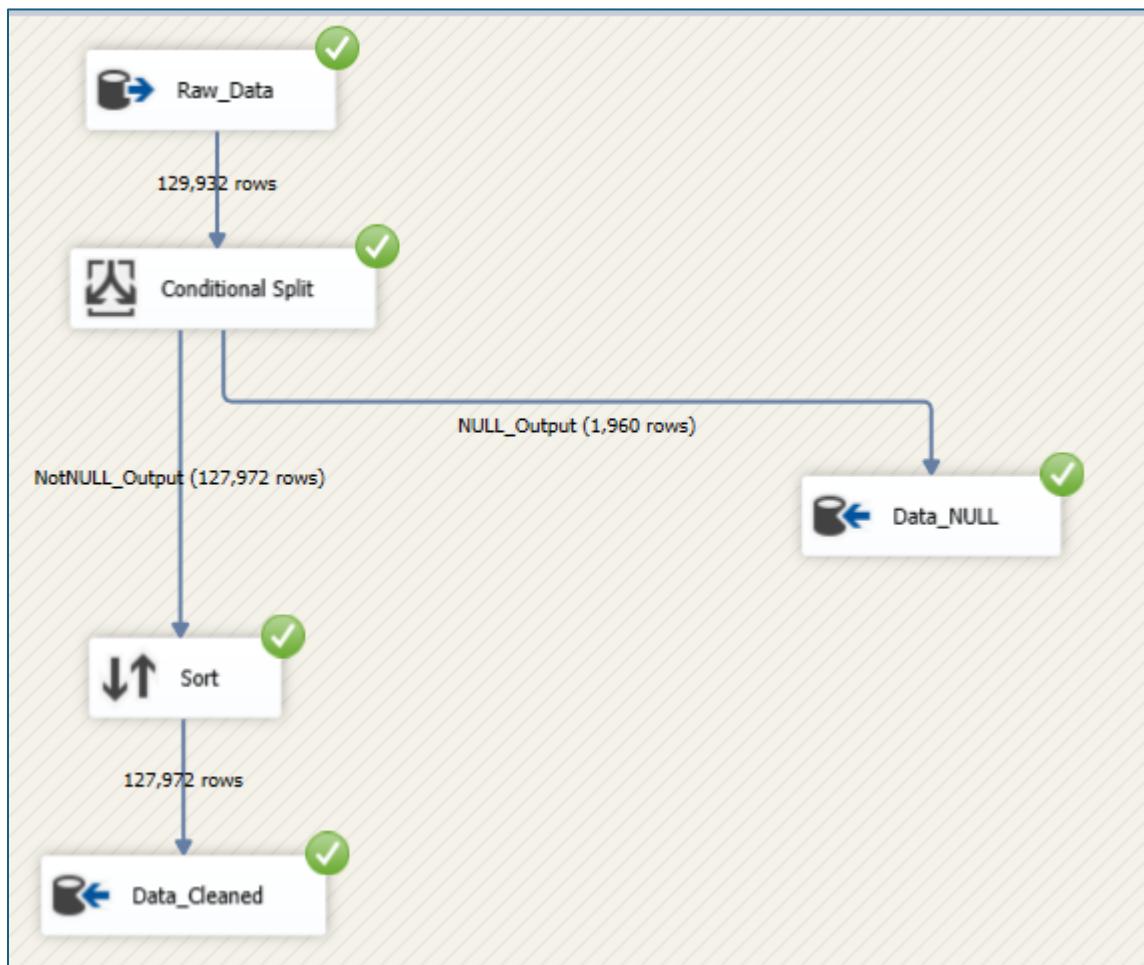
- Dòng dữ liệu chung:



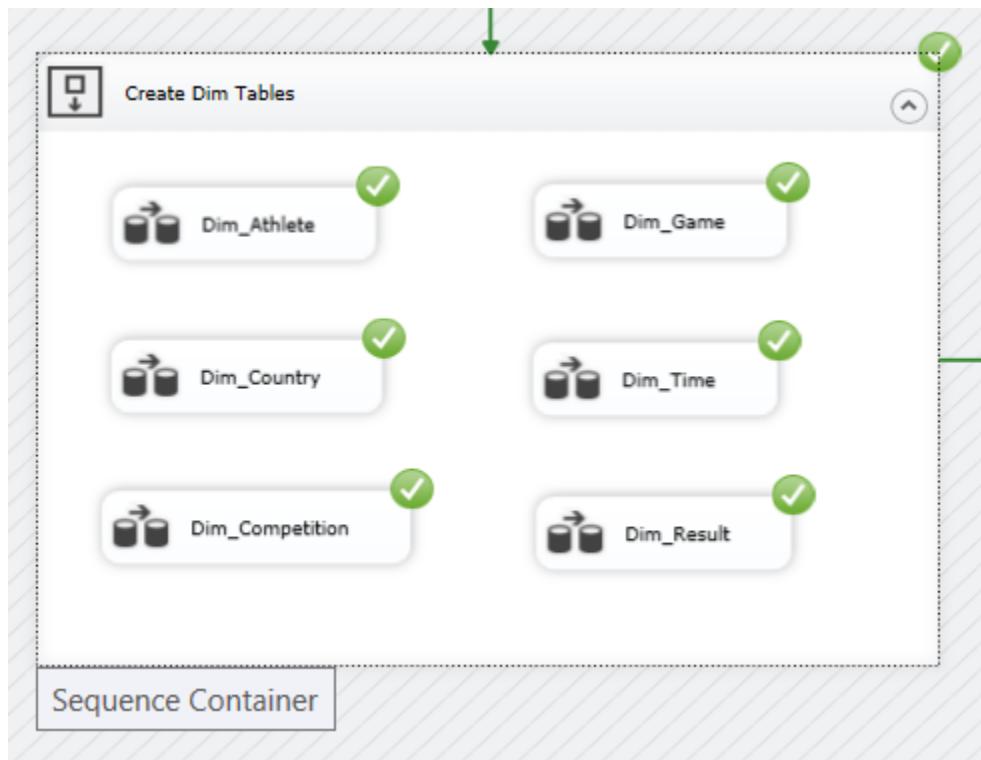
- Quá trình nối dữ liệu từ các bảng



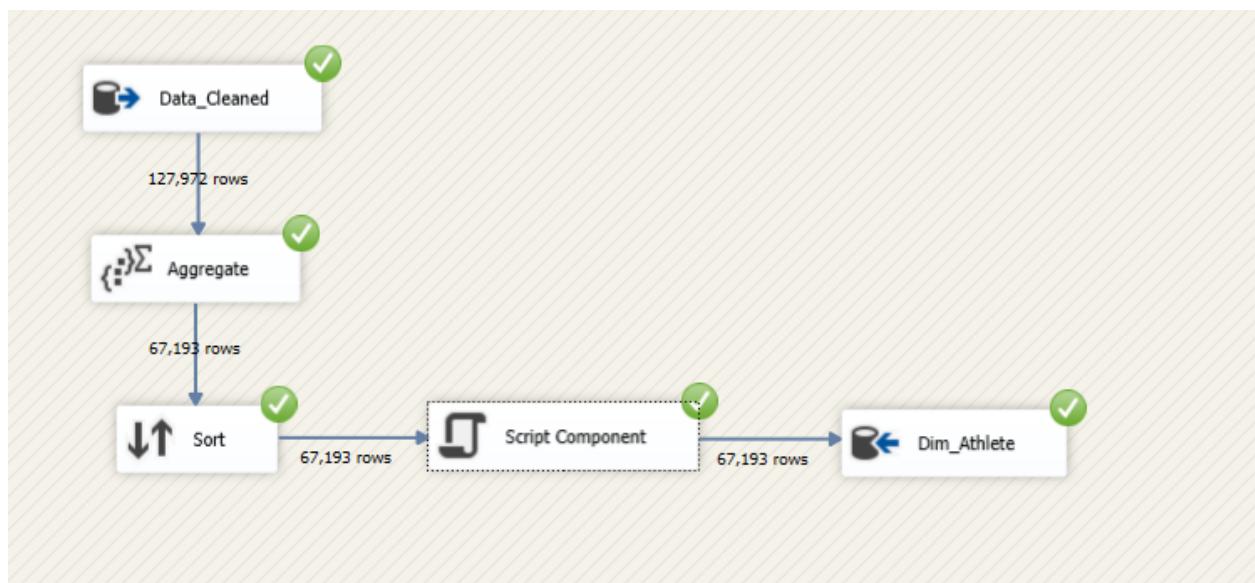
- Quá trình làm sạch dữ liệu



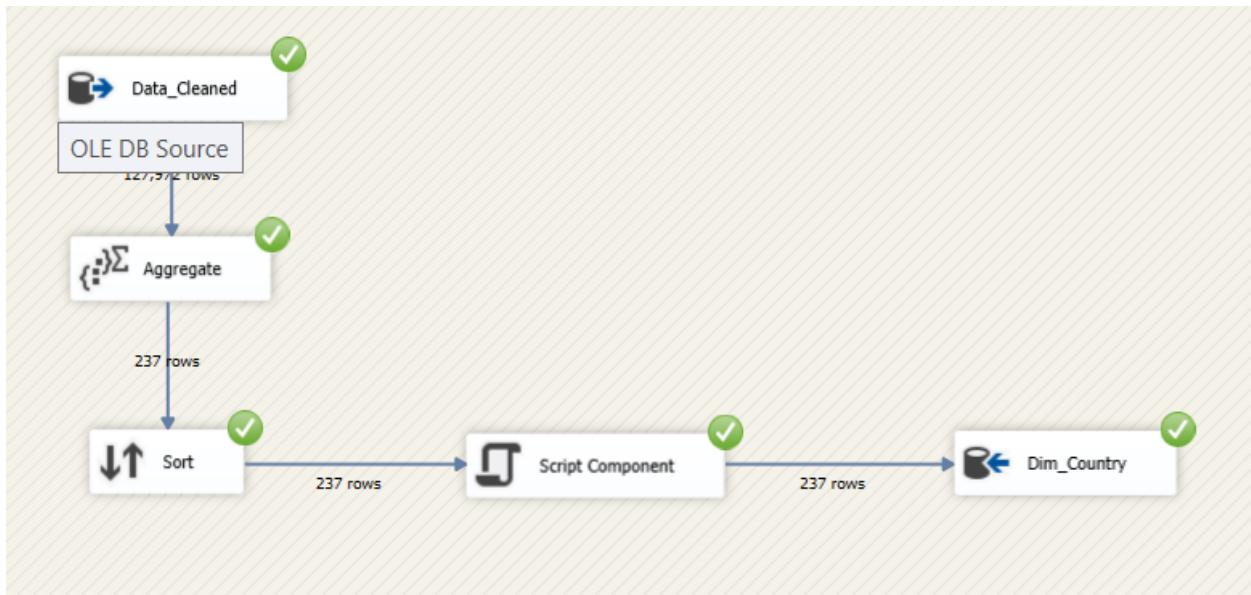
- Quá trình tạo các bảng Dim



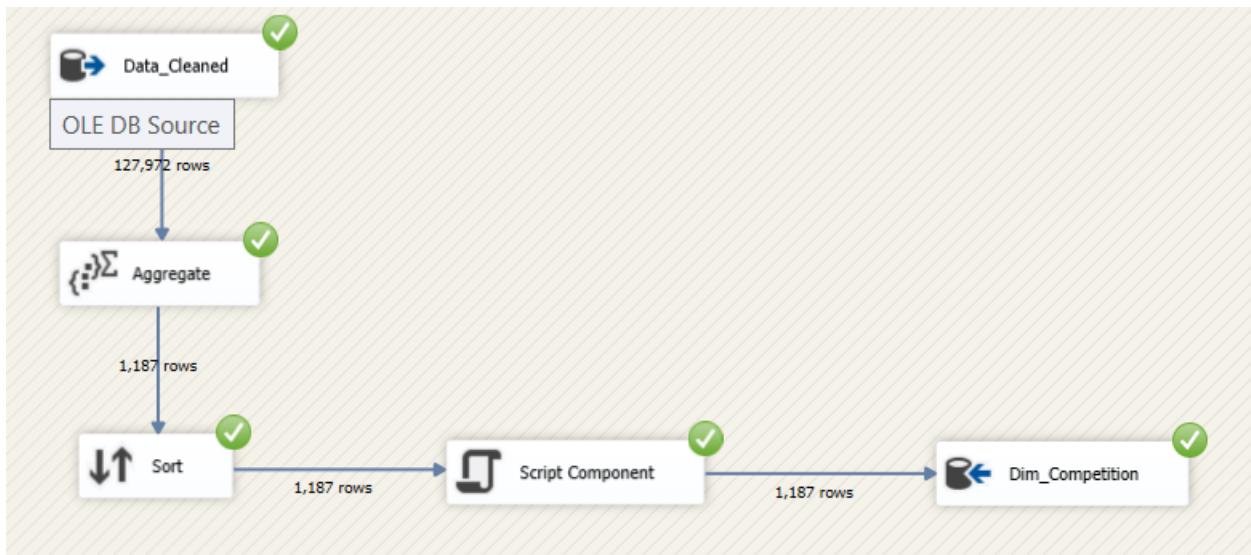
+ Dim_Athlete



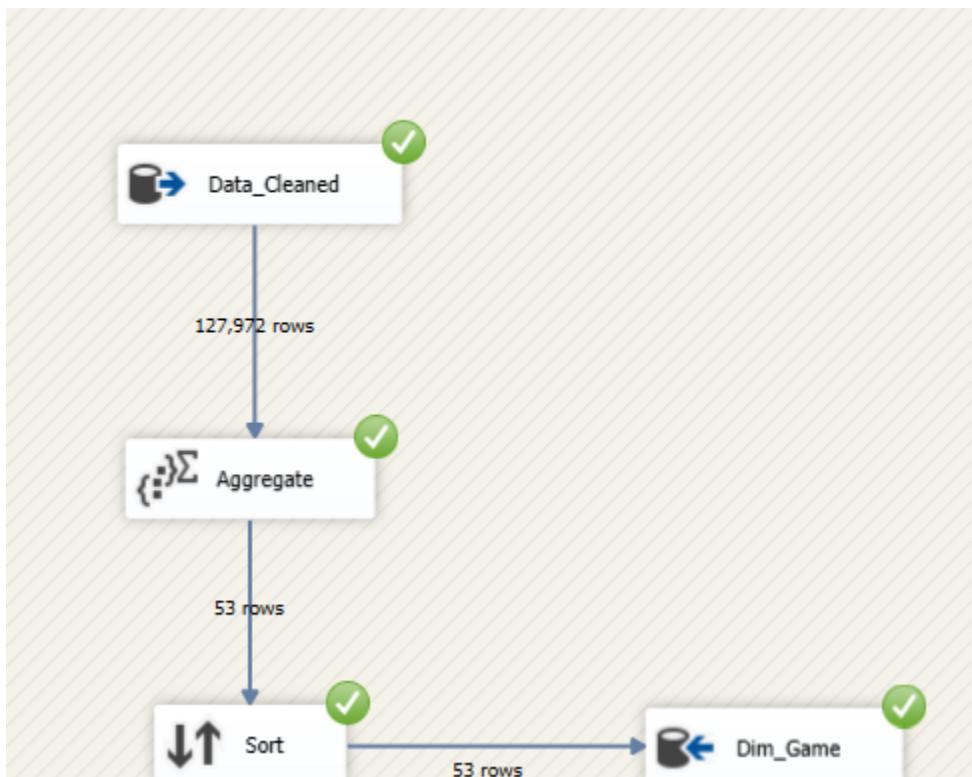
+ Dim_Country



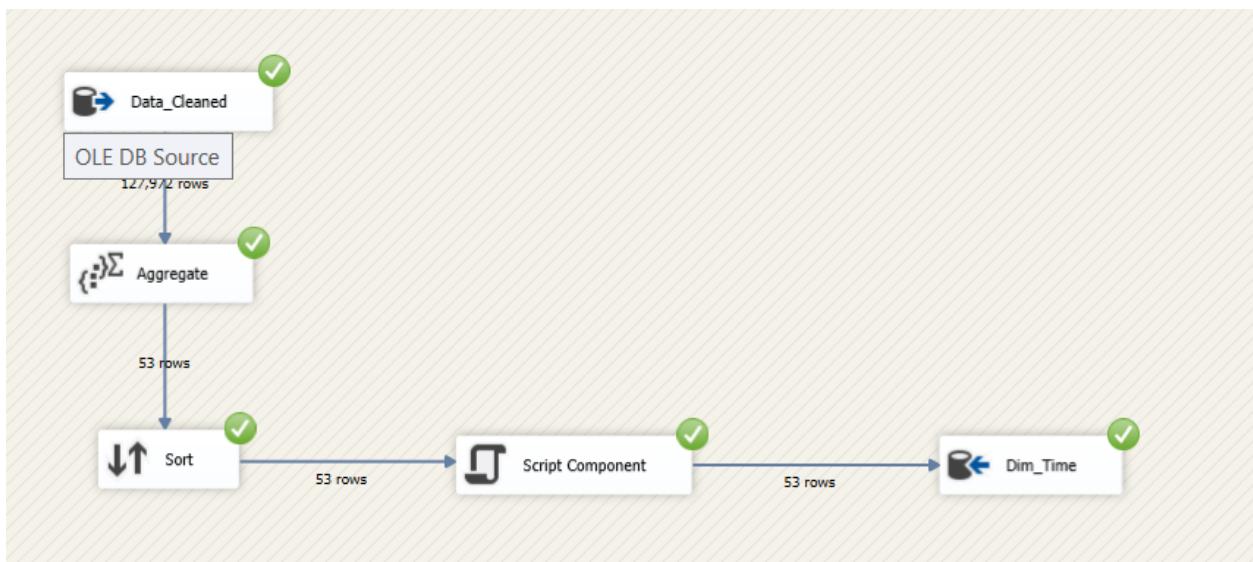
+ Dim_Country



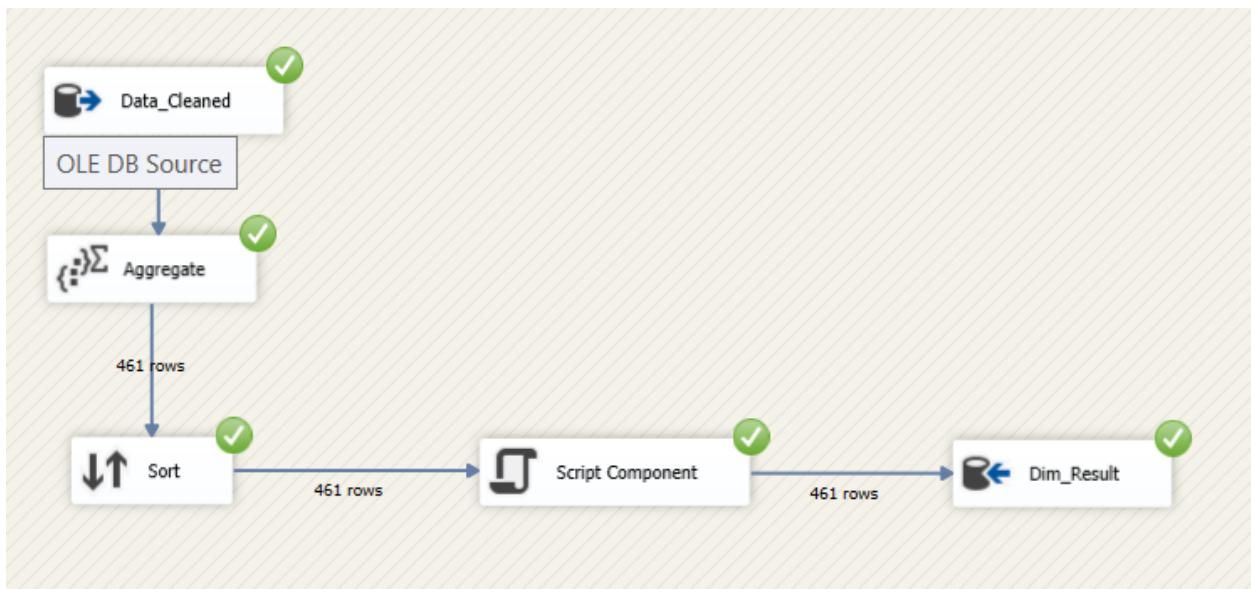
+ Dim_Game



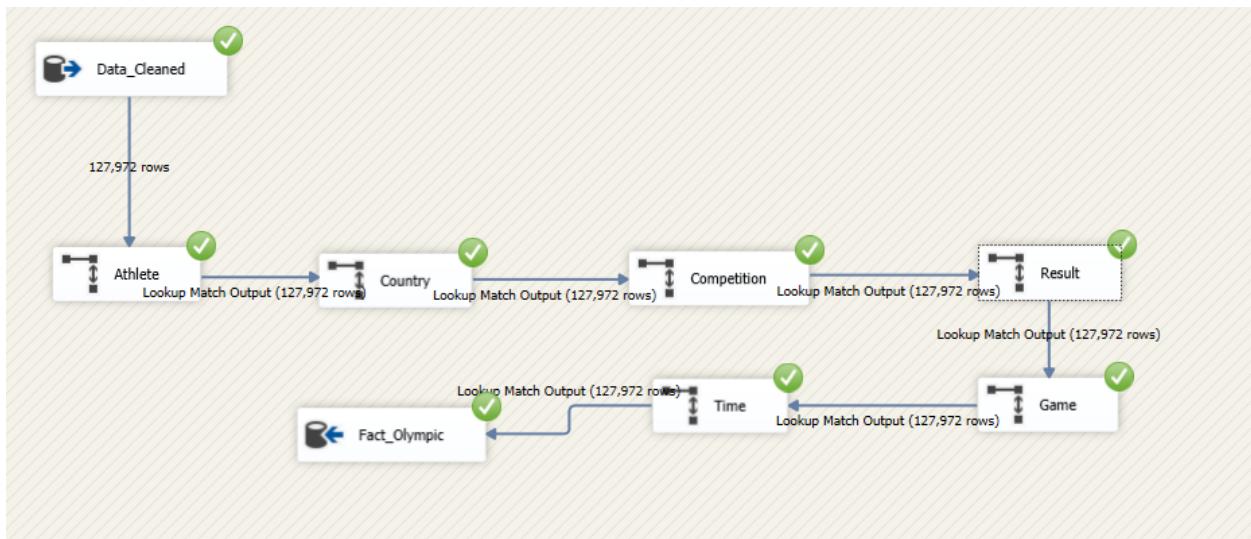
+ Dim_Time



+ Dim_Result



- Quá trình tạo bảng Fact



4. KIỂM TRA DỮ LIỆU

- Dim_Athlete

SQLQuery1.sql - K... (KEPS\LENOVO (54))*

```
SELECT * FROM Dim_Athlete
SELECT * FROM Dim_Country
SELECT * FROM Dim_Competition
SELECT * FROM Dim_Game
SELECT * FROM Dim_Time
SELECT * FROM Dim_Result
SELECT * FROM Fact_Olympic
```

100 %

Results Messages

	athlete_full_name	athlete_year_birth	athlete_id
1	. DENI	1989	1
2	. PRIYANKA	1996	2
3	. RAHUL	1996	3
4	A Baser WASIQI	1975	4
5	AJ HURT	2000	5
6	A. DARNIS	NULL	6
7	A. Germaine GOLDING	1887	7
8	A. TURNOVSKY	NULL	8
9	A.J. MILLER	1899	9
10	Aage Avaldorff MEYER	1904	10
11	Aage Ingvar ERIKSEN	1917	11
12	Aage JUSTESEN	1918	12
13	Aage MYHRVOLD	1918	13
14	Aage RASMUSSEN	1889	14
15	Aage RUBAEK-NIELSEN	1913	15
16	Aage STEEN	NULL	16

- Dim_Country

SQLQuery1.sql - K... (KEPS\LENOVO (54)) * X

```
SELECT * FROM Dim_Athlete
SELECT * FROM Dim_Country
SELECT * FROM Dim_Competition
SELECT * FROM Dim_Game
SELECT * FROM Dim_Time
SELECT * FROM Dim_Result
SELECT * FROM Fact_Olympic
```

100 %

Results Messages

	country_name	country_code	country_3_letter_code	country_id
1	Afghanistan	AF	AFG	1
2	Albania	AL	ALB	2
3	Algeria	DZ	ALG	3
4	American Samoa	AS	ASA	4
5	Andorra	AD	AND	5
6	Angola	AO	ANG	6
7	Antigua and Barbuda	AG	ANT	7
8	Argentina	AR	ARG	8
9	Armenia	AM	ARM	9
10	Aruba	AW	ARU	10
11	Australasia	NULL	ANZ	11
12	Australia	AU	AUS	12
13	Austria	AT	AUT	13
14	Azerbaijan	AZ	AZE	14
15	Bahamas	BS	BAH	15
16	Bahrain	BH	BRN	16
17	Brunei Darussalam	BN	BRU	17

- Dim_Compétition

SQLQuery1.sql - K...(KEPS\LENOVO (54))* ✎ X

```
SELECT * FROM Dim_Athlete
SELECT * FROM Dim_Country
SELECT * FROM Dim_Competition
SELECT * FROM Dim_Game
SELECT * FROM Dim_Time
SELECT * FROM Dim_Result
SELECT * FROM Fact_Olympic
```

100 %

Results Messages

	discipline_title	event_title	participant_type	competition_id
1	Alpine Skiing	alpin combined men	Athlete	1
2	Alpine Skiing	alpine combined men	Athlete	2
3	Alpine Skiing	alpine combined women	Athlete	3
4	Alpine Skiing	downhill men	Athlete	4
5	Alpine Skiing	downhill women	Athlete	5
6	Alpine Skiing	giant slalom men	Athlete	6
7	Alpine Skiing	giant slalom women	Athlete	7
8	Alpine Skiing	Ladies' Alpine Combined	Athlete	8
9	Alpine Skiing	Ladies' Downhill	Athlete	9
10	Alpine Skiing	Ladies' Giant Slalom	Athlete	10
11	Alpine Skiing	Ladies' Slalom	Athlete	11
12	Alpine Skiing	Ladies' Super-G	Athlete	12
13	Alpine Skiing	Men's Alpine Combined	Athlete	13
14	Alpine Skiing	Men's Downhill	Athlete	14
15	Alpine Skiing	Men's Giant Slalom	Athlete	15
16	Alpine Skiing	Men's Slalom	Athlete	16
17	Alpine Skiing	Men's Super G	Athlete	17

- Dim_Game

SQLQuery1.sql - K...(KEPS\LENOVO (54))*

```
SELECT * FROM Dim_Athlete
SELECT * FROM Dim_Country
SELECT * FROM Dim_Competition
SELECT * FROM Dim_Game
SELECT * FROM Dim_Time
SELECT * FROM Dim_Result
SELECT * FROM Fact_Olympic
```

100 %

Results Messages

	game_slug	game_name	game_location
1	albertville-1992	Albertville 1992	France
2	amsterdam-1928	Amsterdam 1928	Netherlands
3	antwerp-1920	Antwerp 1920	Belgium
4	athens-1896	Athens 1896	Greece
5	athens-2004	Athens 2004	Greece
6	atlanta-1996	Atlanta 1996	United States
7	barcelona-1992	Barcelona 1992	Spain
8	beijing-2008	Beijing 2008	China
9	beijing-2022	Beijing 2022	China
10	berlin-1936	Berlin 1936	Germany
11	calgary-1988	Calgary 1988	Canada
12	chamonix-1924	Chamonix 1924	France
13	cortina-d-ampezzo-1956	Cortina d'Ampezzo 1956	Italy
14	garmisch-partenkirchen-1936	Garmisch-Partenkirchen 1936	Germany
15	grenoble-1968	Grenoble 1968	France
16	helsinki-1952	Helsinki 1952	Finland
17	innbruck-1964	Innsbruck 1964	Austria

- Dim_Time

SQLQuery1.sql - K... (KEPS\LENOVO (54))*

```
SELECT * FROM Dim_Athlete
SELECT * FROM Dim_Country
SELECT * FROM Dim_Competition
SELECT * FROM Dim_Game
SELECT * FROM Dim_Time
SELECT * FROM Dim_Result
SELECT * FROM Fact_Olympic
```

100 %

Results Messages

	game_start_date	game_season	game_year	time_id
1	1896-04-06T11:38:39Z	Summer	1896	1
2	1900-05-14T08:50:39Z	Summer	1900	2
3	1904-07-01T14:00:00Z	Summer	1904	3
4	1908-04-27T08:00:00Z	Summer	1908	4
5	1912-05-05T08:00:00Z	Summer	1912	5
6	1920-04-20T07:00:00Z	Summer	1920	6
7	1924-01-25T08:00:00Z	Winter	1924	7
8	1924-05-04T07:00:00Z	Summer	1924	8
9	1928-02-11T08:00:00Z	Winter	1928	9
10	1928-05-17T07:00:00Z	Summer	1928	10
11	1932-02-04T13:00:00Z	Winter	1932	11
12	1932-07-30T16:00:00Z	Summer	1932	12
13	1936-02-06T08:00:00Z	Winter	1936	13
14	1936-08-01T07:00:00Z	Summer	1936	14
15	1948-01-30T07:00:00Z	Winter	1948	15
16	1948-07-29T07:00:00Z	Summer	1948	16
17	1952-02-14T07:00:00Z	Winter	1952	17

- Dim_Result

SQLQuery1.sql - K... (KEPS\LENOVO (54)) * X

```
SELECT * FROM Dim_Athlete
SELECT * FROM Dim_Country
SELECT * FROM Dim_Competition
SELECT * FROM Dim_Game
SELECT * FROM Dim_Time
SELECT * FROM Dim_Result
SELECT * FROM Fact_Olympic
```

100 %

Results Messages

	medal_type	rank_equal	rank_position	result_id
1	NULL	NULL	NULL	1
2	NULL	NULL	1	2
3	NULL	NULL	2	3
4	NULL	NULL	3	4
5	NULL	NULL	4	5
6	NULL	NULL	5	6
7	NULL	NULL	6	7
8	NULL	NULL	7	8
9	NULL	NULL	8	9
10	NULL	NULL	9	10
11	NULL	NULL	10	11
12	NULL	NULL	11	12
13	NULL	NULL	12	13
14	NULL	NULL	13	14
15	NULL	NULL	14	15
16	NULL	NULL	15	16

- Fact_Olympic

SQLQuery1.sql - K... (KEPS\LENOVO (54))* □ X

```
SELECT * FROM Dim_Athlete
SELECT * FROM Dim_Country
SELECT * FROM Dim_Competition
SELECT * FROM Dim_Game
SELECT * FROM Dim_Time
SELECT * FROM Dim_Result
SELECT * FROM Fact_Olympic
```

100 %

Results Messages

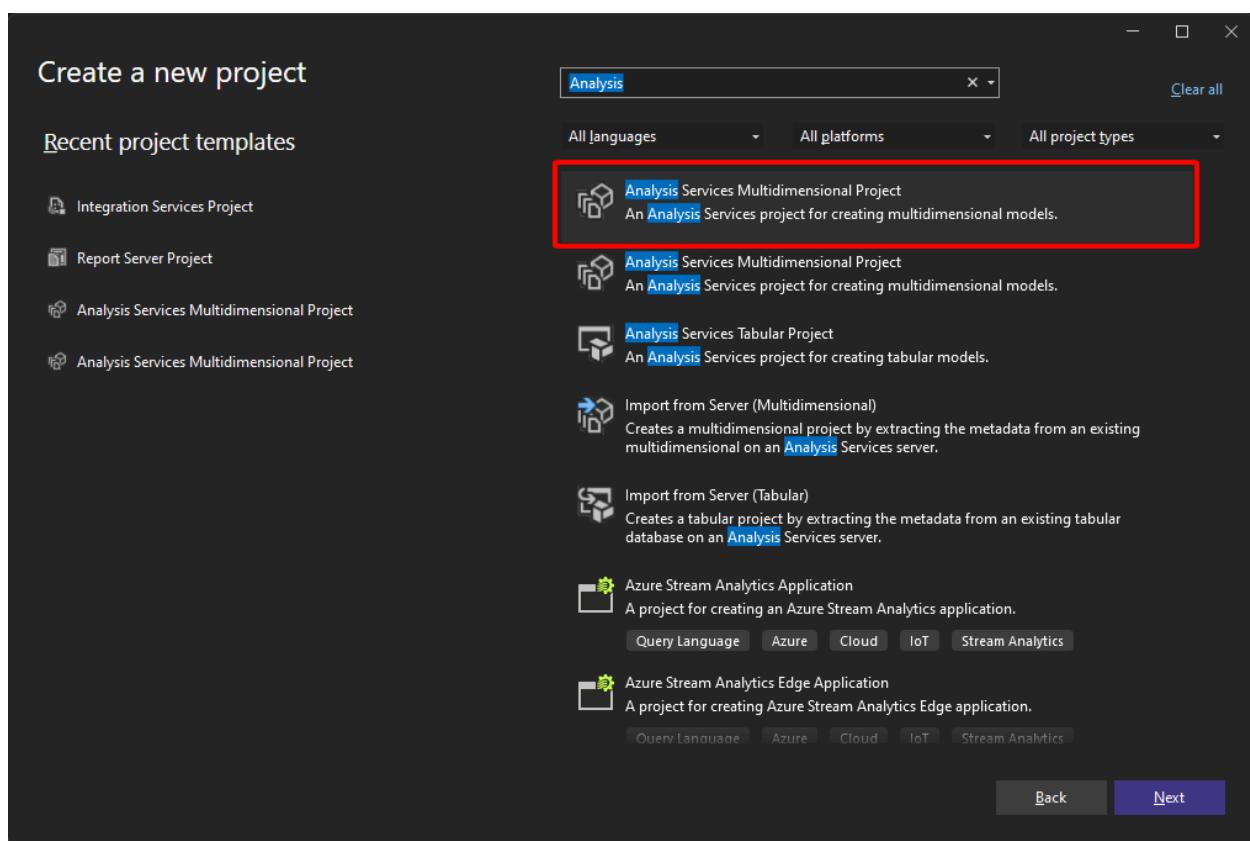
	fact_id	game_slug	athlete_id	country_id	competition_id	result_id	time_id
1	12657529	albertville-1992	230	213	343	103	37
2	12657530	albertville-1992	230	213	354	1	37
3	12657531	albertville-1992	298	198	2	1	37
4	12657532	albertville-1992	298	198	23	28	37
5	12657533	albertville-1992	329	118	2	31	37
6	12657534	albertville-1992	329	118	6	27	37
7	12657535	albertville-1992	329	118	25	1	37
8	12657536	albertville-1992	502	172	214	42	37
9	12657537	albertville-1992	502	172	217	39	37
10	12657538	albertville-1992	510	234	207	86	37
11	12657539	albertville-1992	510	234	215	85	37
12	12657540	albertville-1992	516	150	805	30	37
13	12657541	albertville-1992	516	149	807	458	37
14	12657542	albertville-1992	586	12	501	15	37
15	12657543	albertville-1992	664	74	2	1	37
16	12657544	albertville-1992	664	74	4	1	37
17	12657545	albertville-1992	702	162	211	17	37

CHƯƠNG 3: QUI TRÌNH PHÂN TÍCH DỮ LIỆU- SQL SERVER ANALYSIS SERVICES (SSAS)

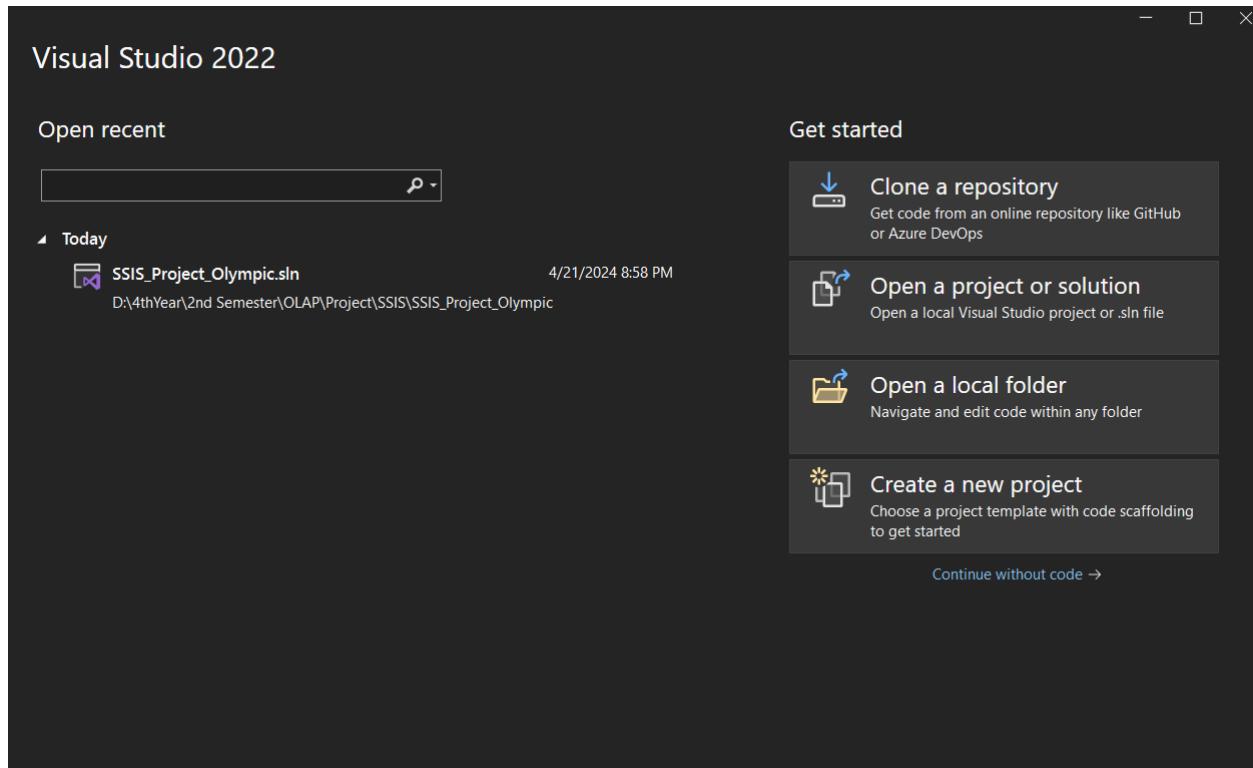
1. KHỞI TẠO DỰ ÁN SSAS TRÊN VISUAL STUDIO 2022

Các bước khởi tạo dự án SSAS trên Visual Studio 2022

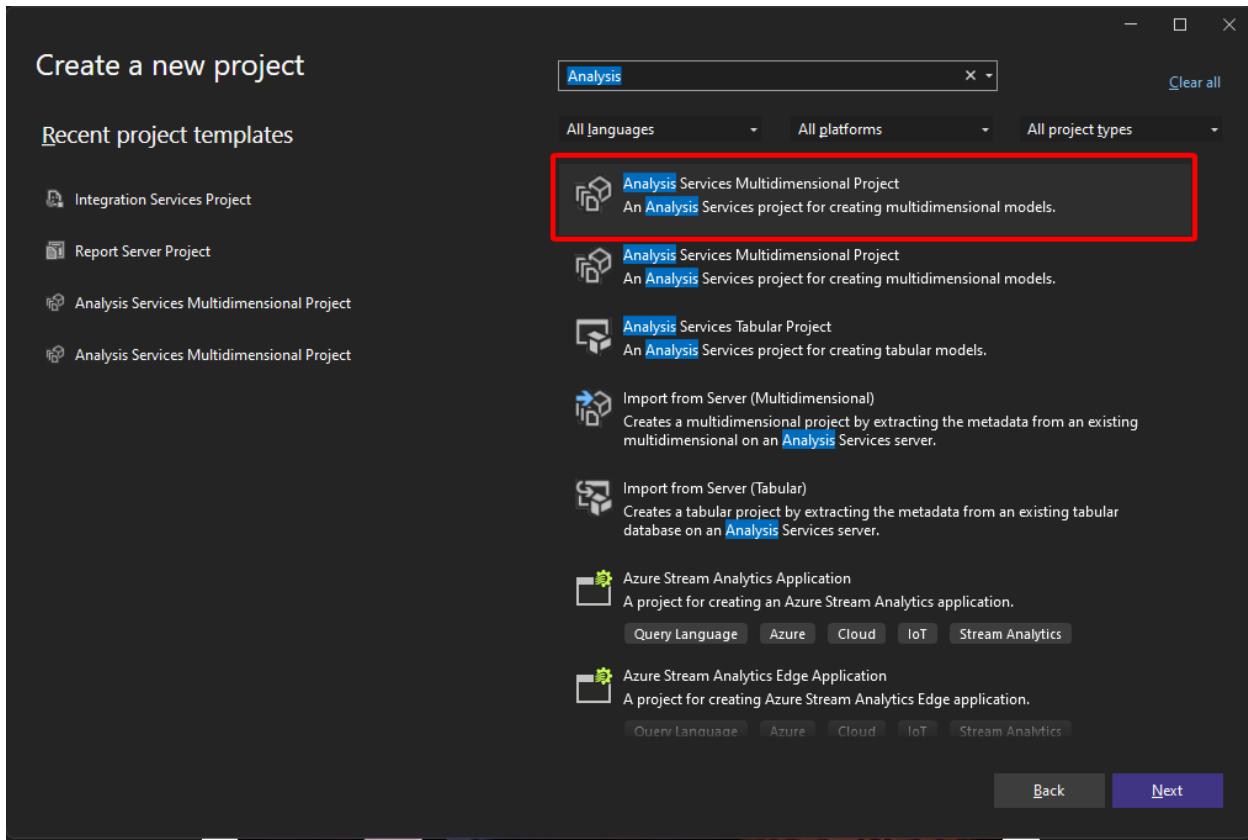
- Chuẩn bị các công cụ sau:
 - + Microsoft Analysis Services Project Extension:



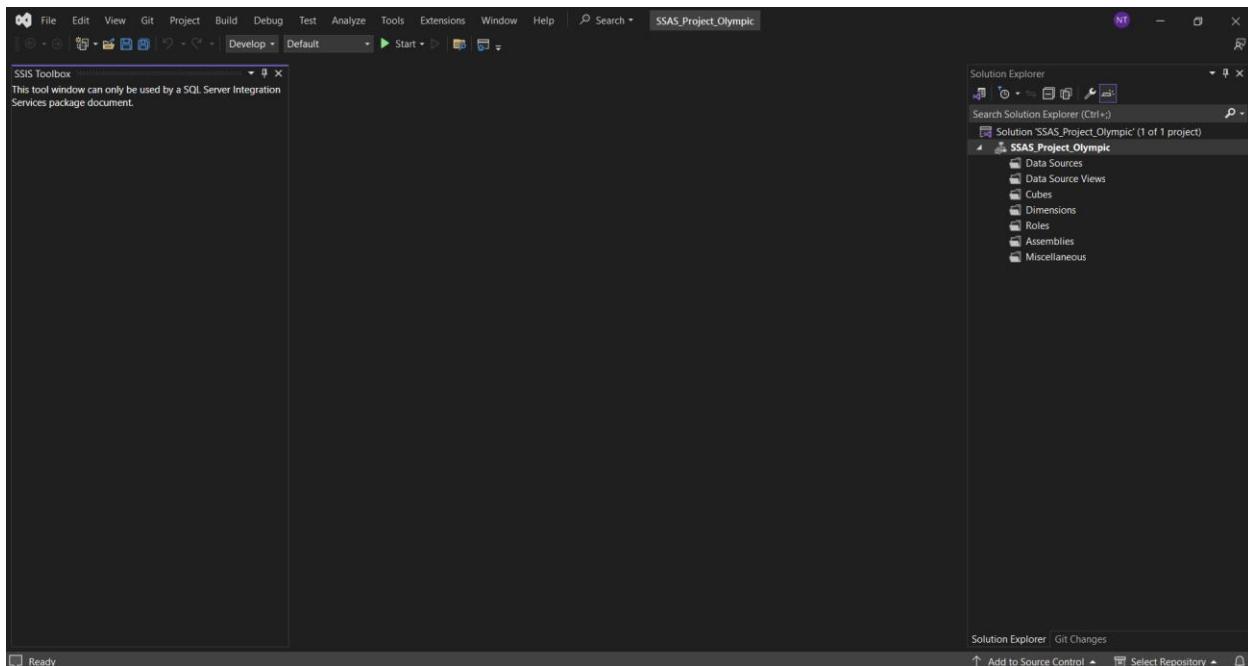
- Bước 1: Truy cập Visual Studio 2022 => Chọn Create new project



- Bước 2: Tìm kiếm từ khóa Analysis và chọn Analysis Services Multidimension Project → Đổi tên project thành *SSAS_Project_Olympic* → Chọn Create.



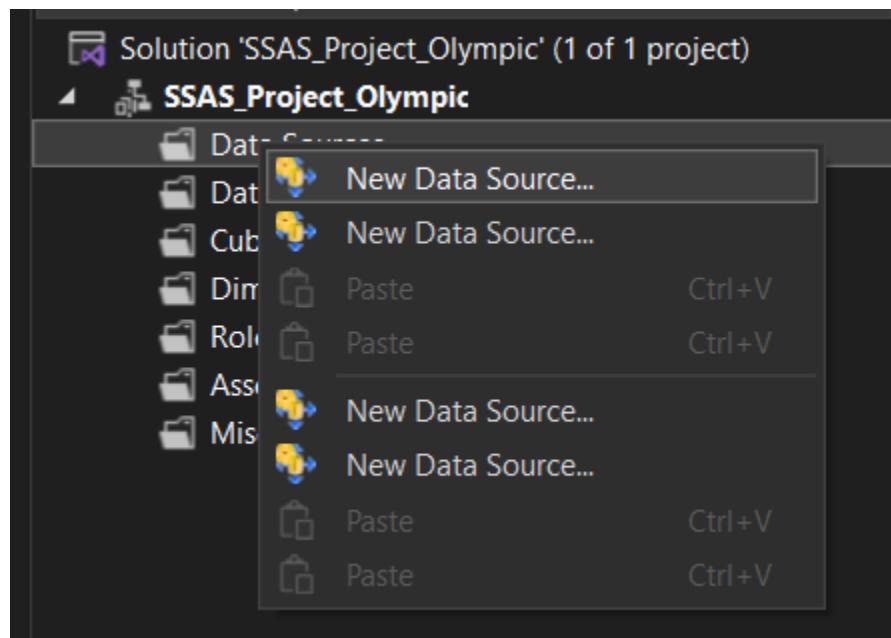
- **Bước 3: Giao diện của dự án SSAS đã xuất hiện**



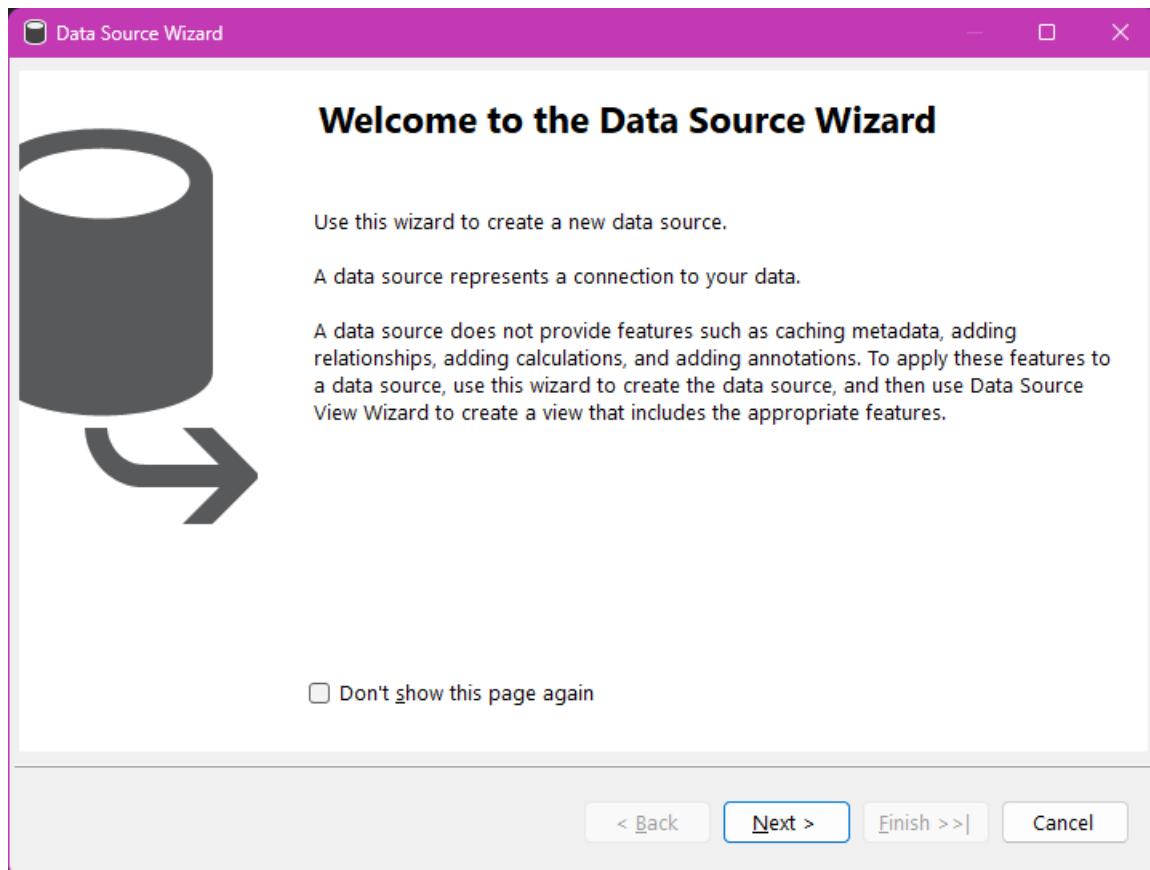
2. XÁC ĐỊNH DỮ LIỆU NGUỒN

Để xác định dữ liệu nguồn (Define Data Source) cho một dự án SSAS mới, ta thực hiện các bước sau:

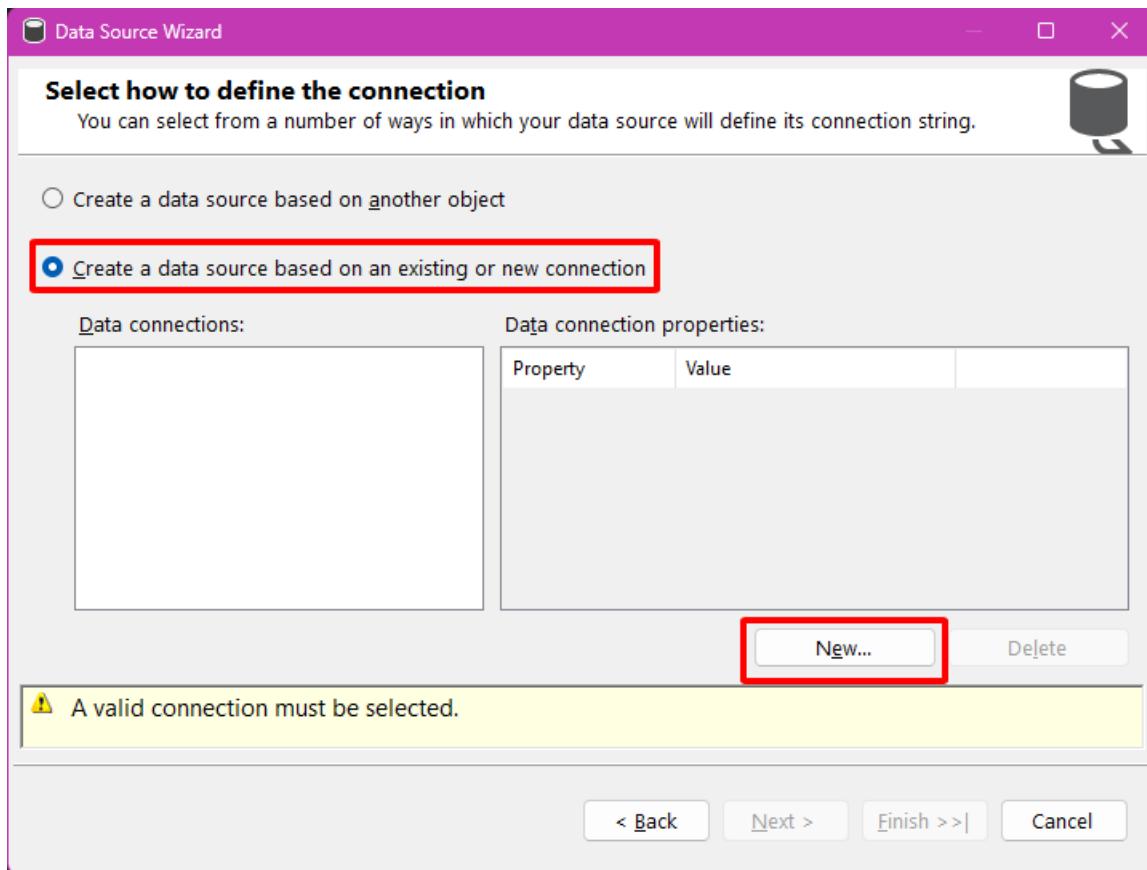
- Bước 1: Tại cửa sổ Solution Explorer, chuột phải vào Data Source => New Data Source để thêm một nguồn dữ liệu.



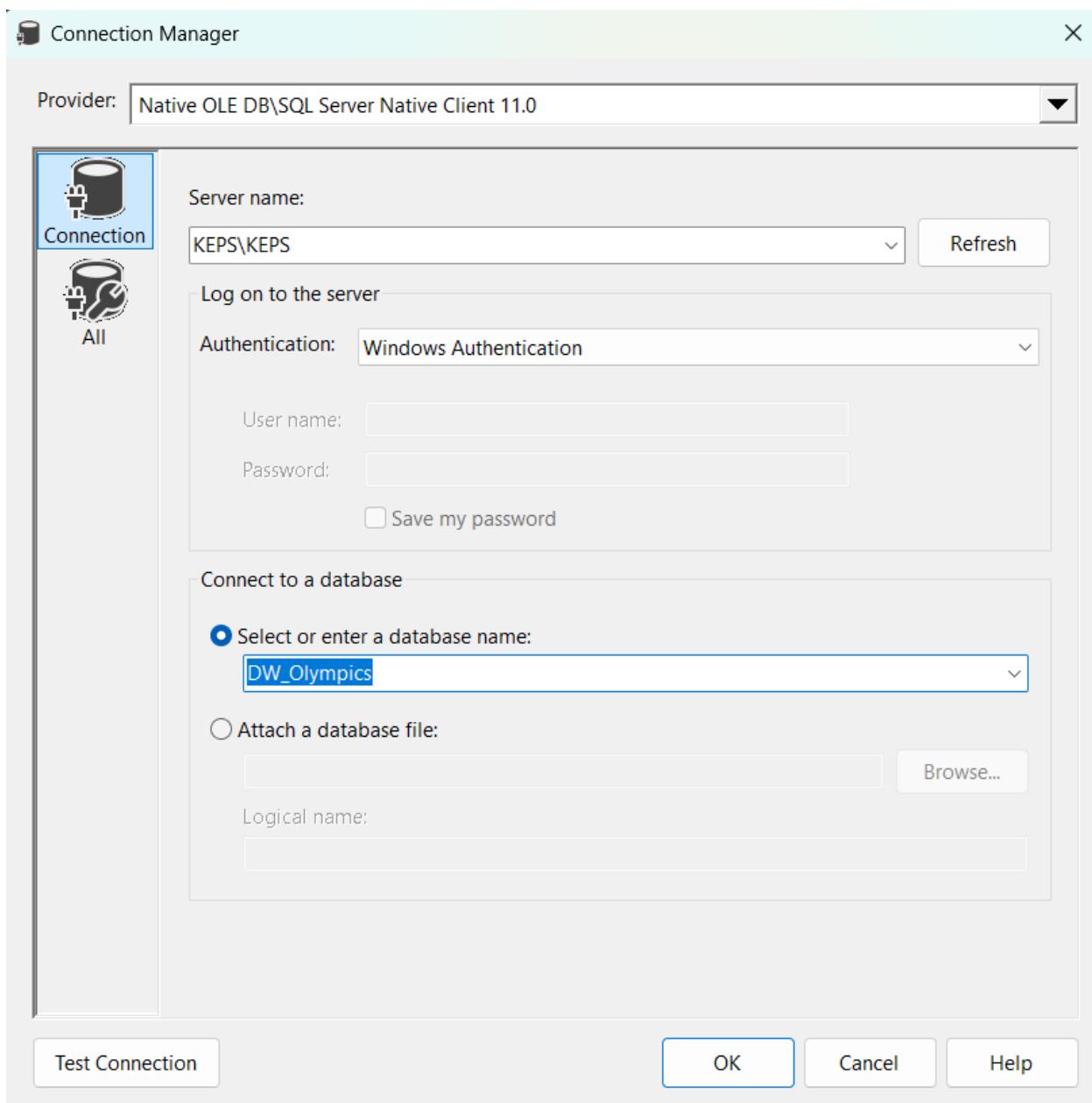
- Bước 2: Cửa sổ **Data Source Wizard** xuất hiện, nhấp **Next** để tiếp tục.

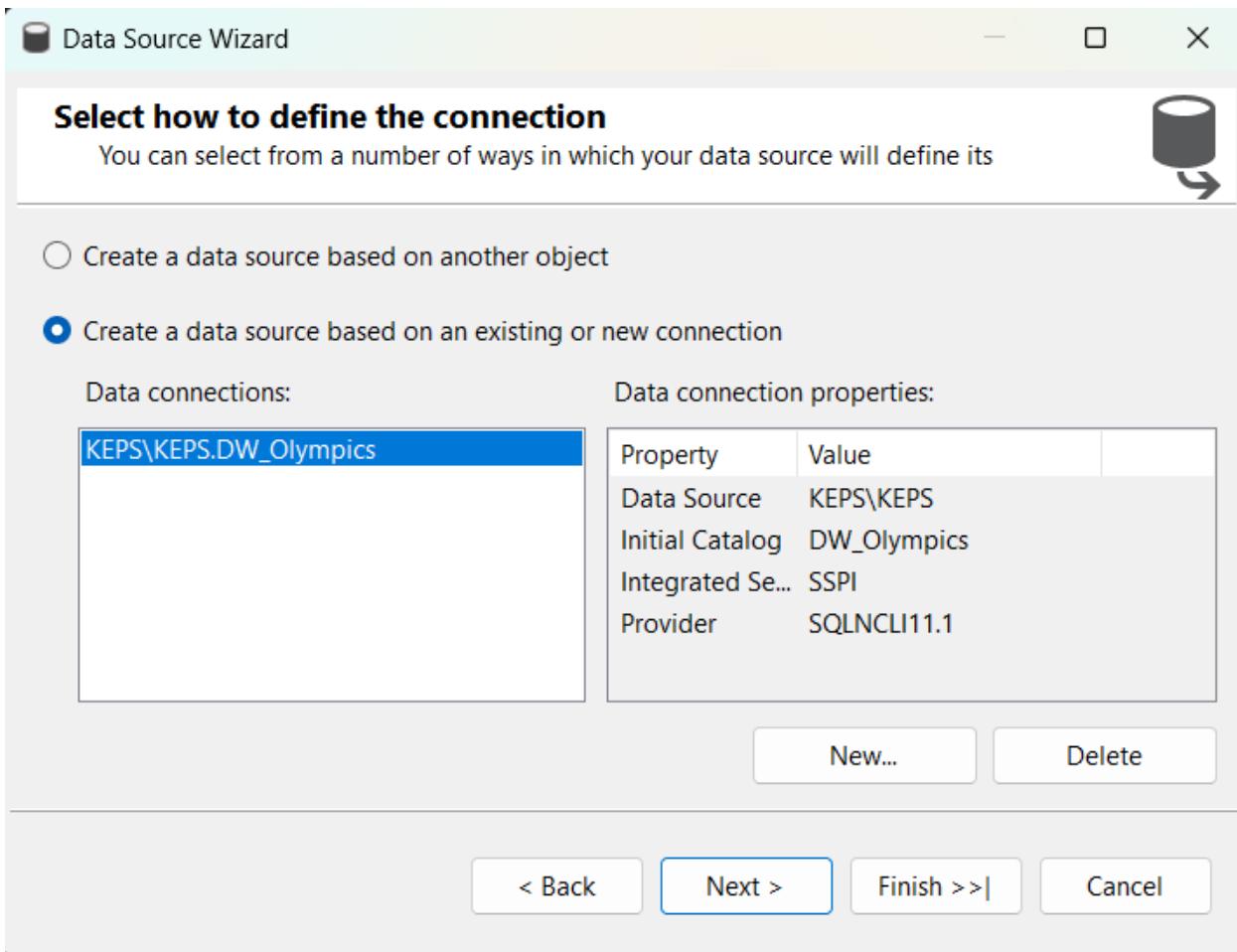


- Bước 3: Chọn nguồn dữ liệu kết nối (Data Source) cần sử dụng đến dự án SSAS, ở đây ta chọn kết nối từ dự án SSIS (xem chương 2) đã tạo. Chọn *Create a data source based on an existing or new connection* → **New** để khởi tạo một kết nối mới.

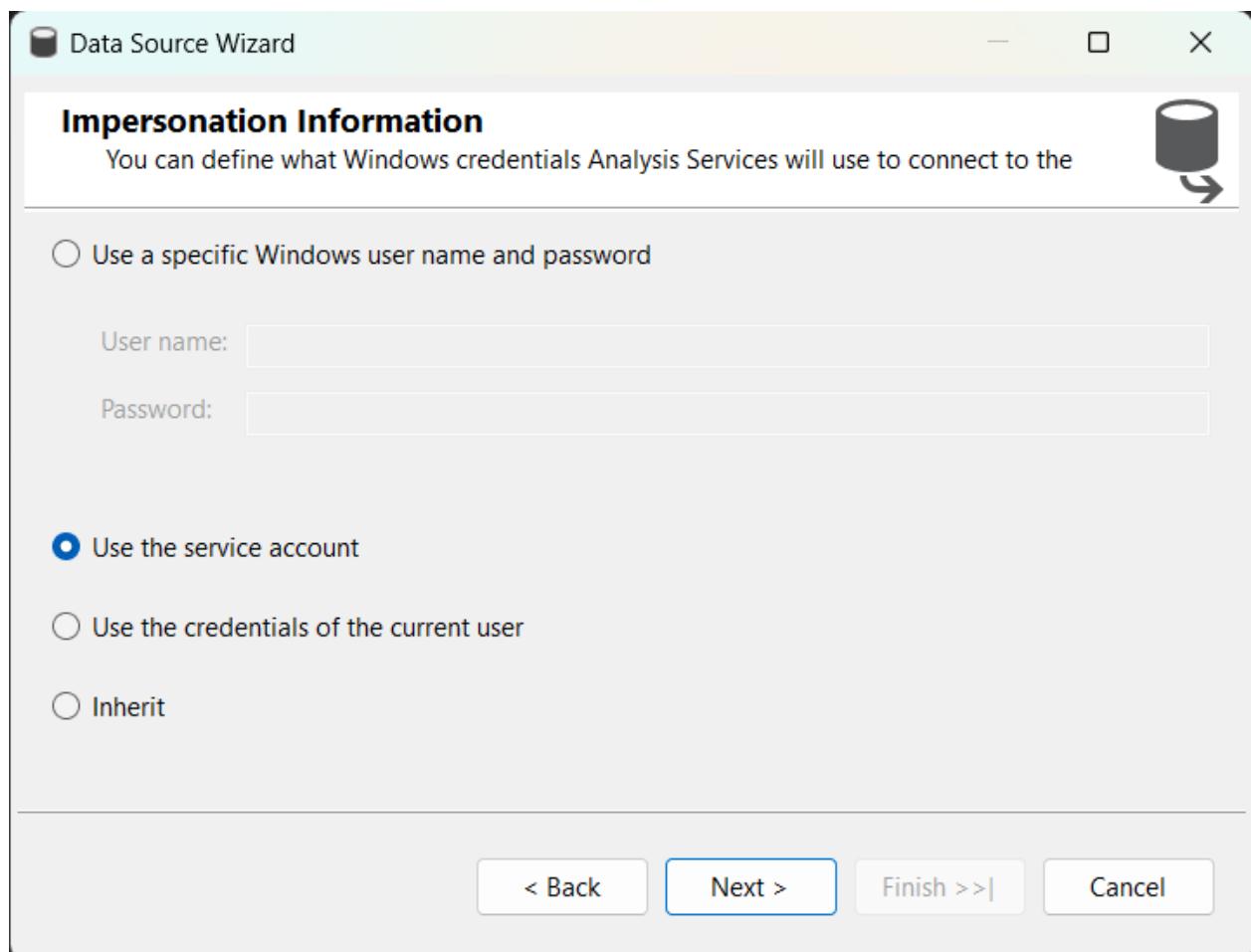


- Bước 4: Tại cửa sổ **Connection Manager**, cấu hình cho Connection từ cơ sở dữ liệu. Chọn tên Server, xác thực vào server và kết nối đến kho dữ liệu.

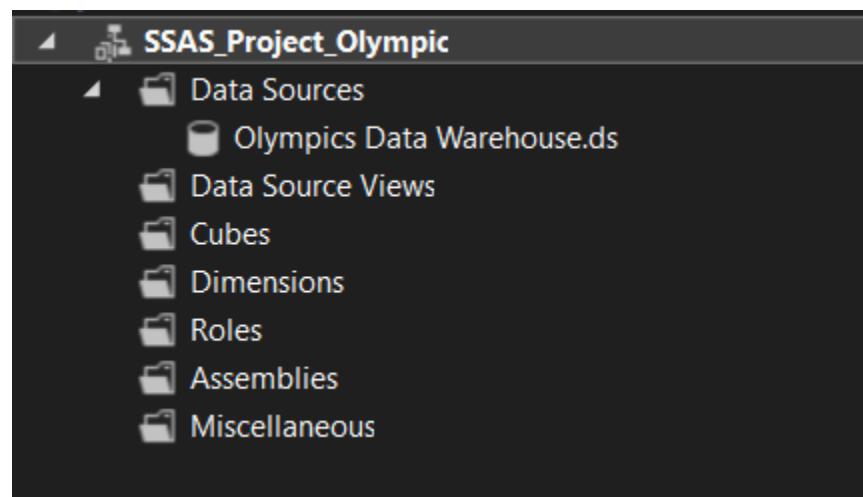
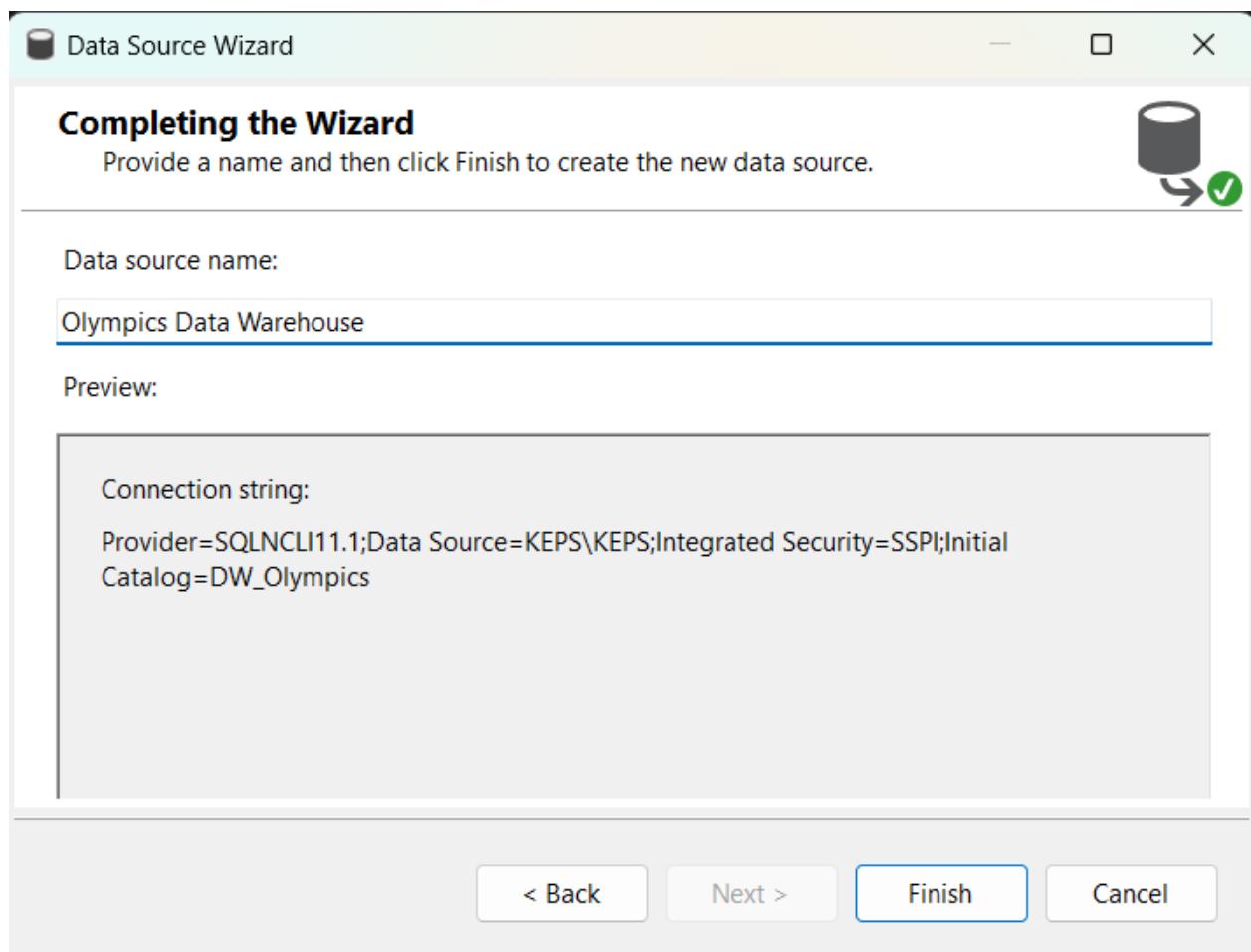




- Bước 5: Tại cửa sổ **Impersonation Information**, định nghĩa Security Credentials cho Analysis Services để kết nối tới data source. Impersonation ảnh hưởng đến tài khoản Windows dùng để kết nối đến data source khi Windows Authentication được chọn. Analysis Services thì không hỗ trợ impersonation để xử lý OLAP objects. Chọn **Use the service account** → Next.



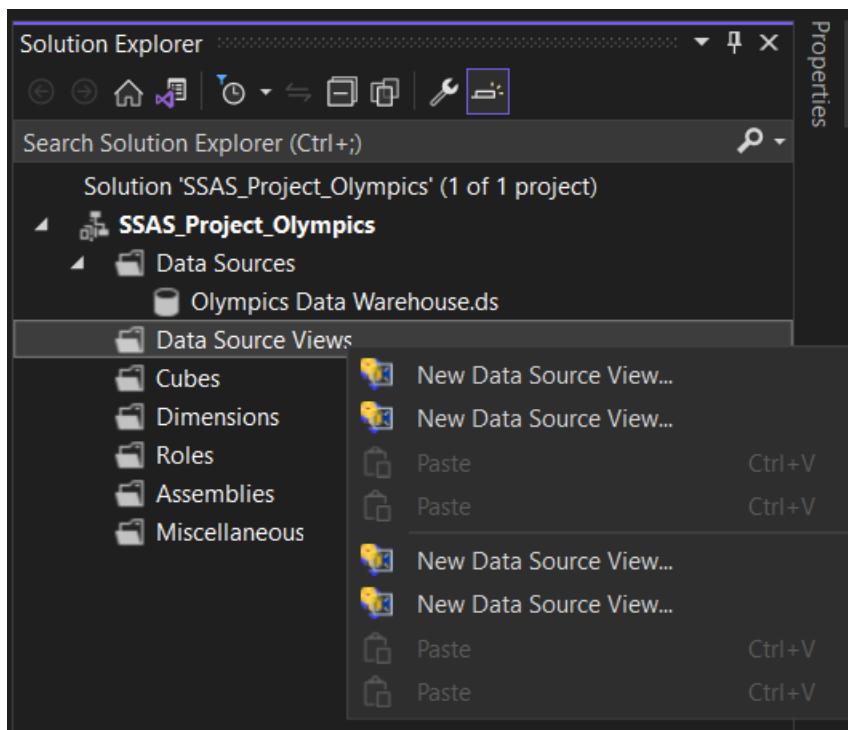
- Bước 6: Đổi tên Data Source Name là *Olympics Data Warehouse* và chọn **Finish** để hoàn tất.



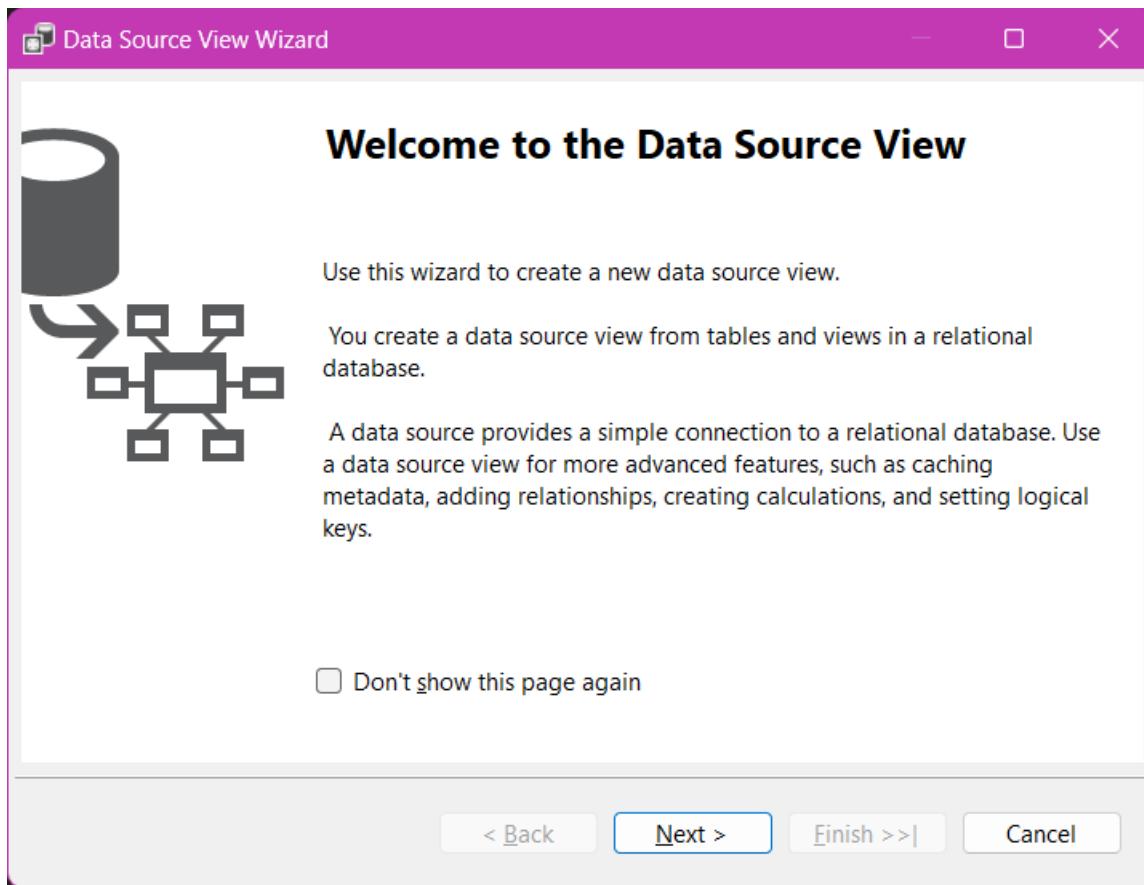
3. XÁC ĐỊNH KHUNG DỮ LIỆU NGUỒN

Để xác định khung nhìn dữ liệu nguồn (Define Data Source View) cho một dự án SSAS mới, ta thực hiện các bước sau:

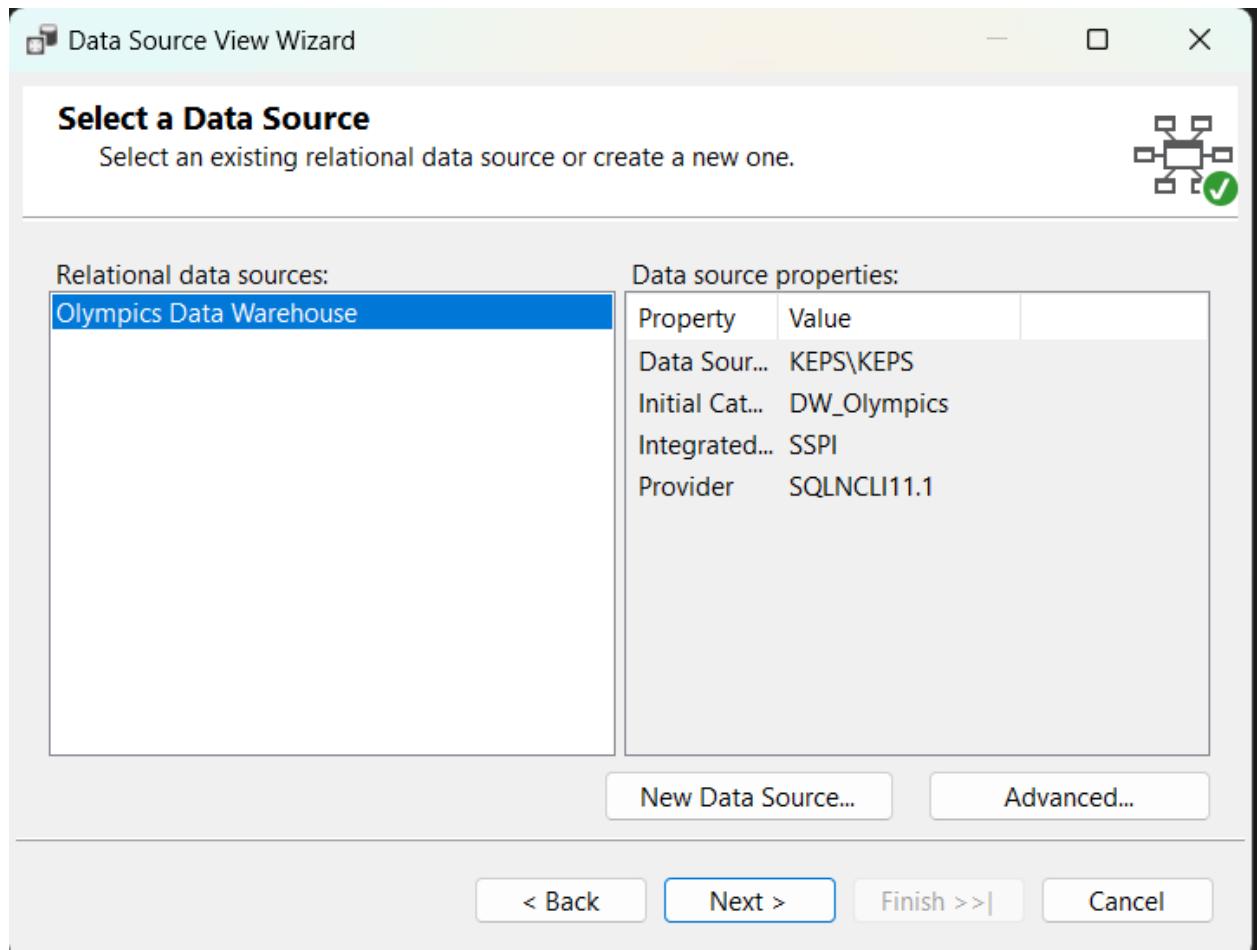
- Bước 1: Tại cửa sổ **Solution Explorer**, chuột phải vào **Data Source Views** → **New Data Source View** để thêm một nguồn dữ liệu



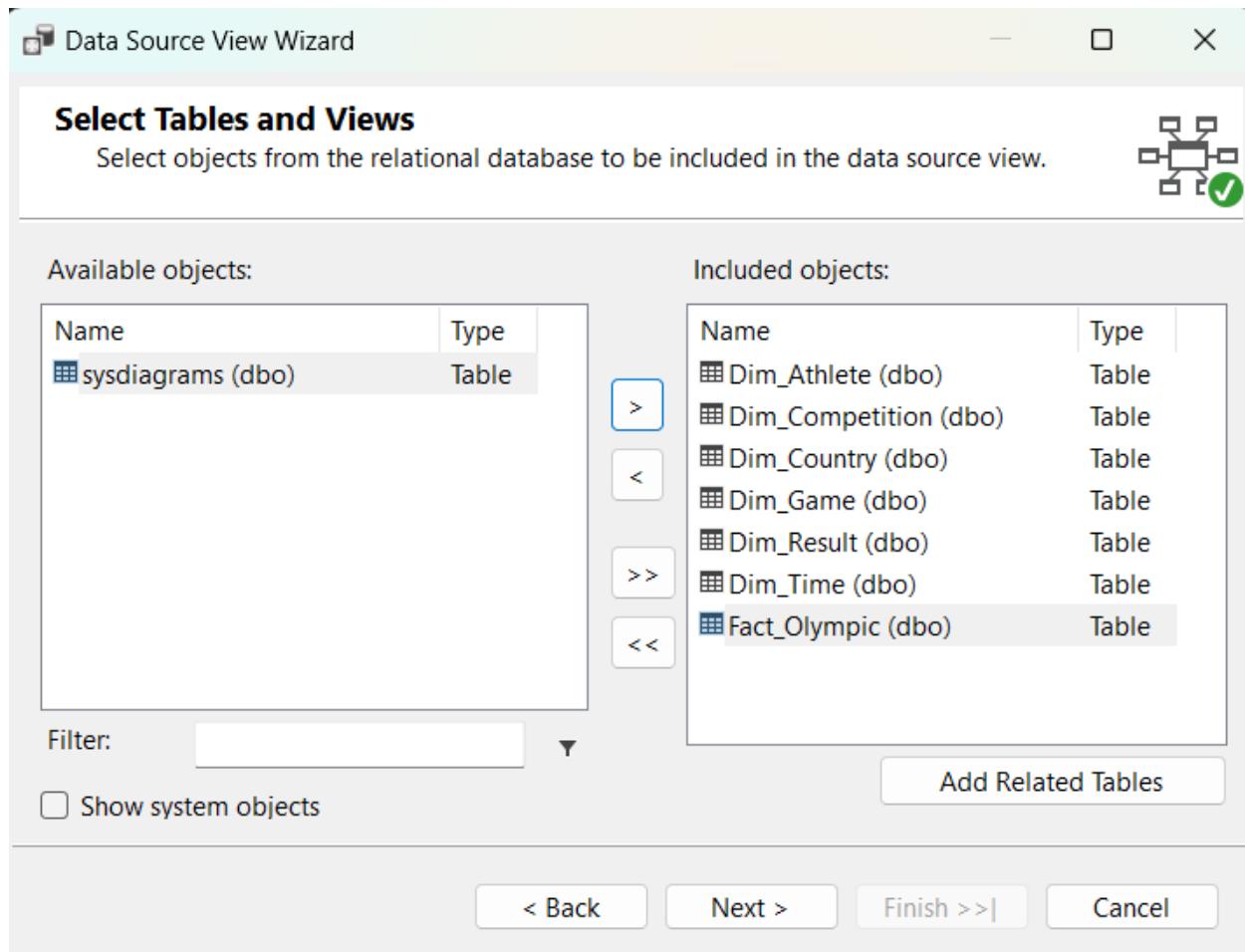
- Bước 2: Cửa sổ **Data Source View Wizard** xuất hiện, nhấn **Next** để tiếp tục.



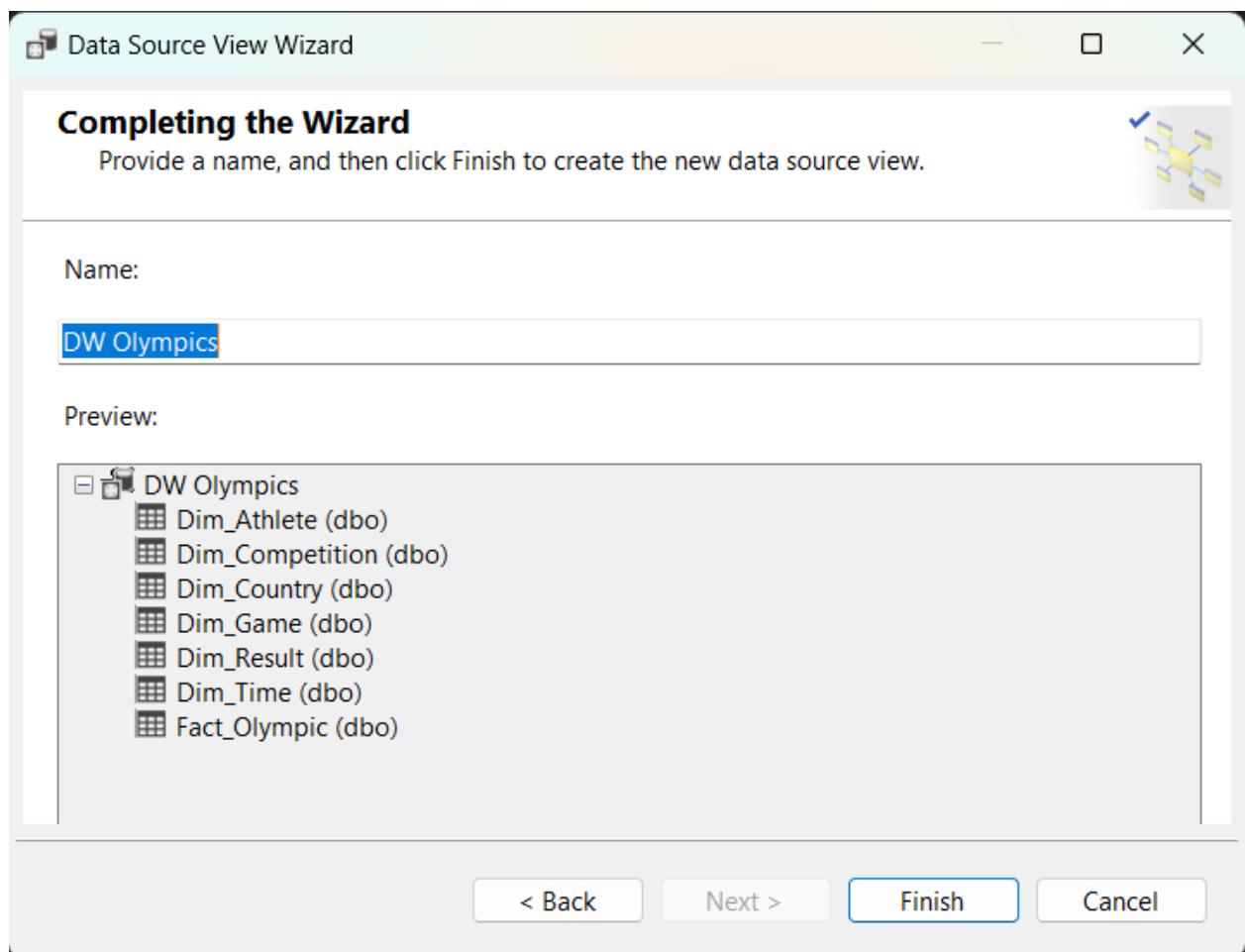
- Bước 3: Tại cửa sổ **Select Data Source**, chọn nguồn dữ liệu kết nối (Relational Data Source) cần sử dụng đến dự án SSAS, ở đây ta chọn kết nối từ dự án SSIS (xem chương 2) đã tạo trên Relational Data Sources. Chọn **Next** để khởi tạo một khung nhìn kết nối mới



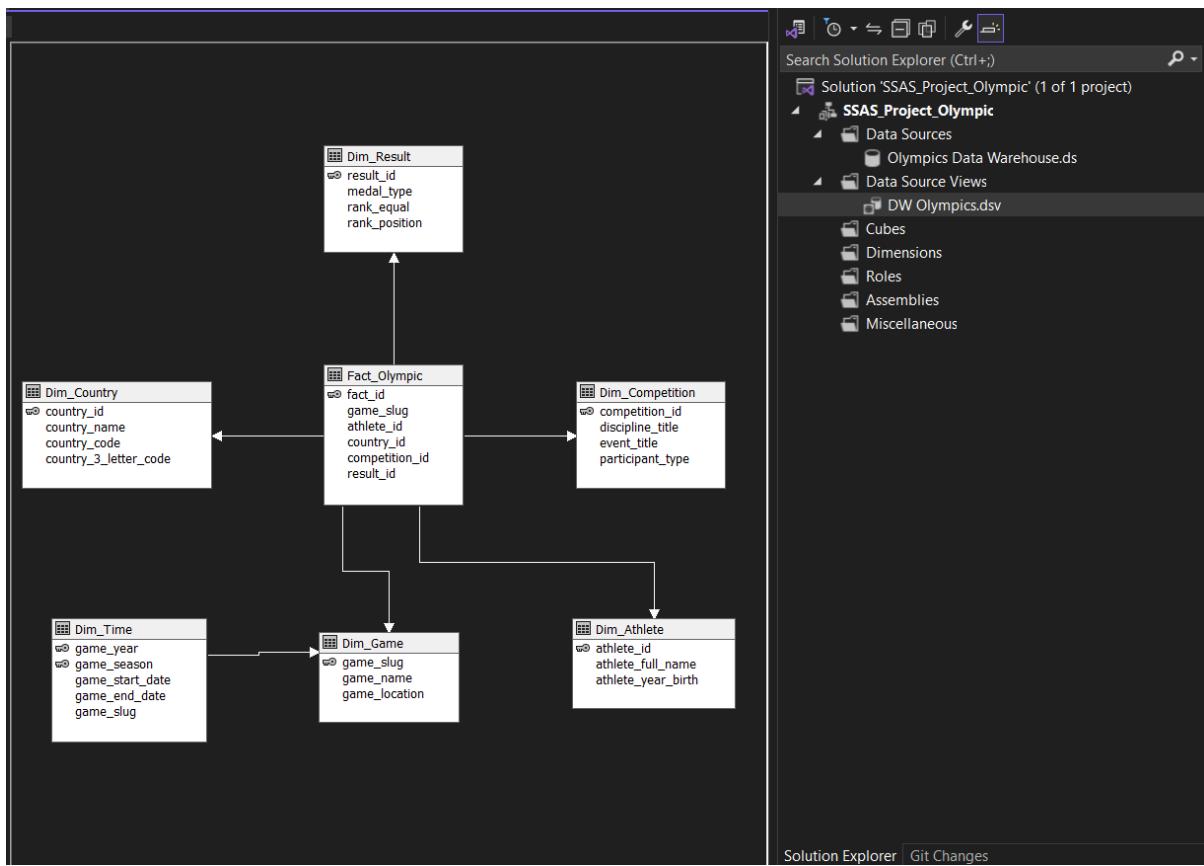
- Bước 4: Tại cửa sổ **Data Source View Wizard**, chọn các bảng DIM và FACT sử dụng trong quá trình phân tích, sau đó chọn **Next**



- Bước 5: Chọn Finish để tạo Data Source View mới.



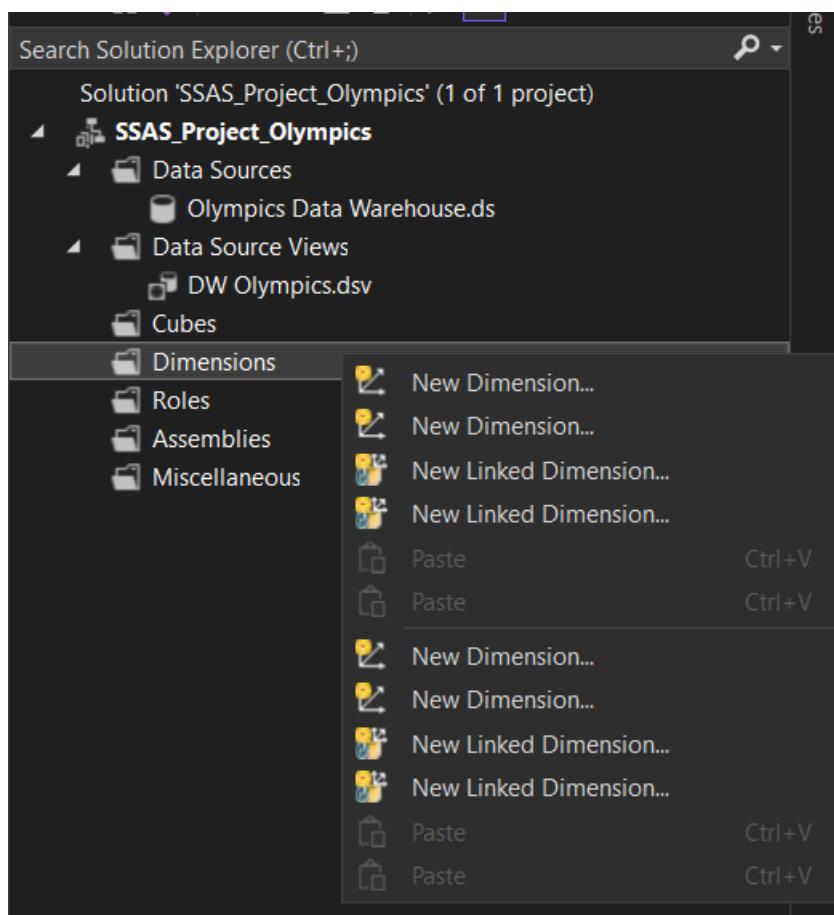
- Kết quả: Chọn tệp *DW Olympics.dsv* → **Open**.



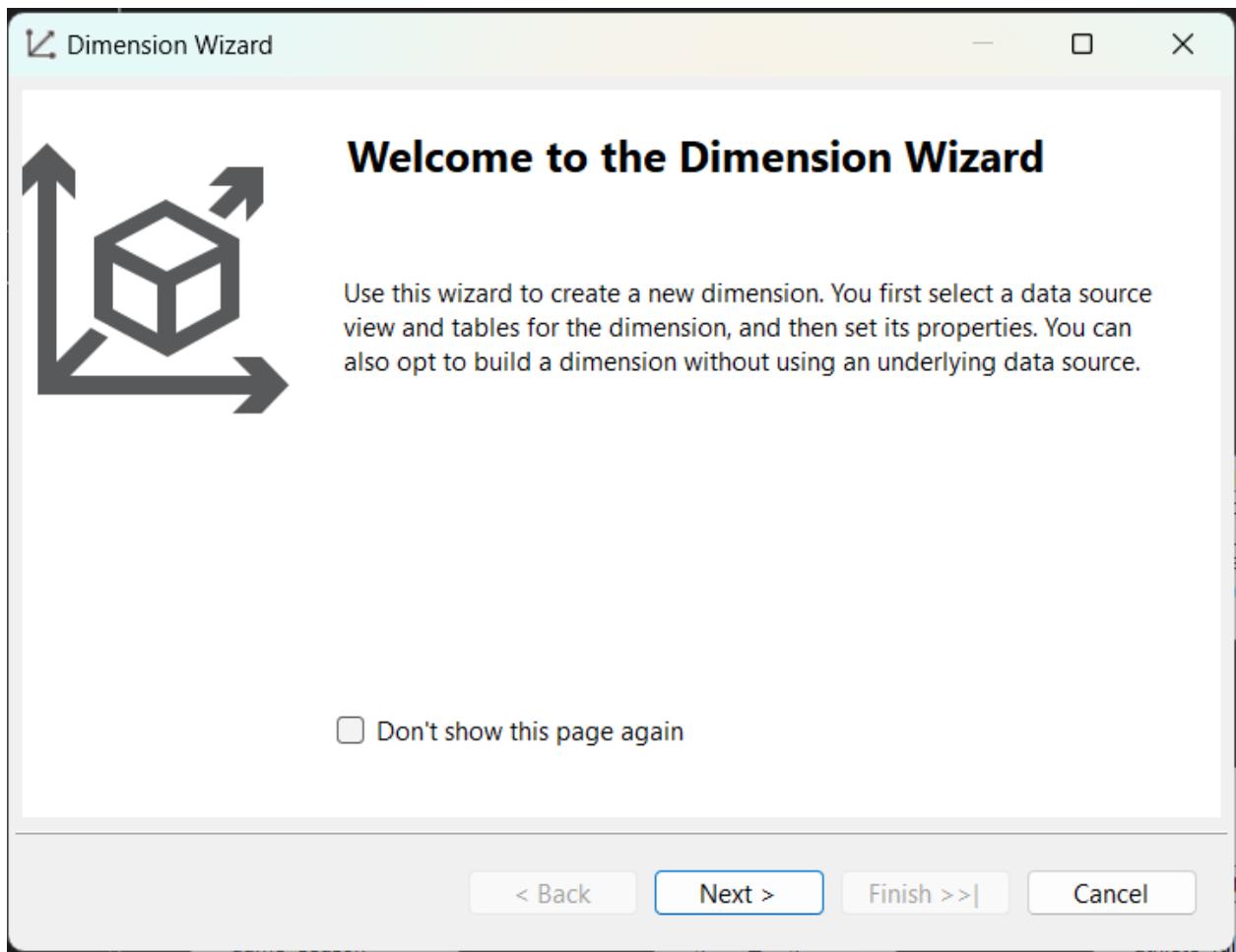
4. XÁC ĐỊNH CÁC CHIỀU DỮ LIỆU

Để xác định các chiều dữ liệu (Define a Dimension) cho một dự án SSAS mới, ta thực hiện các bước sau:

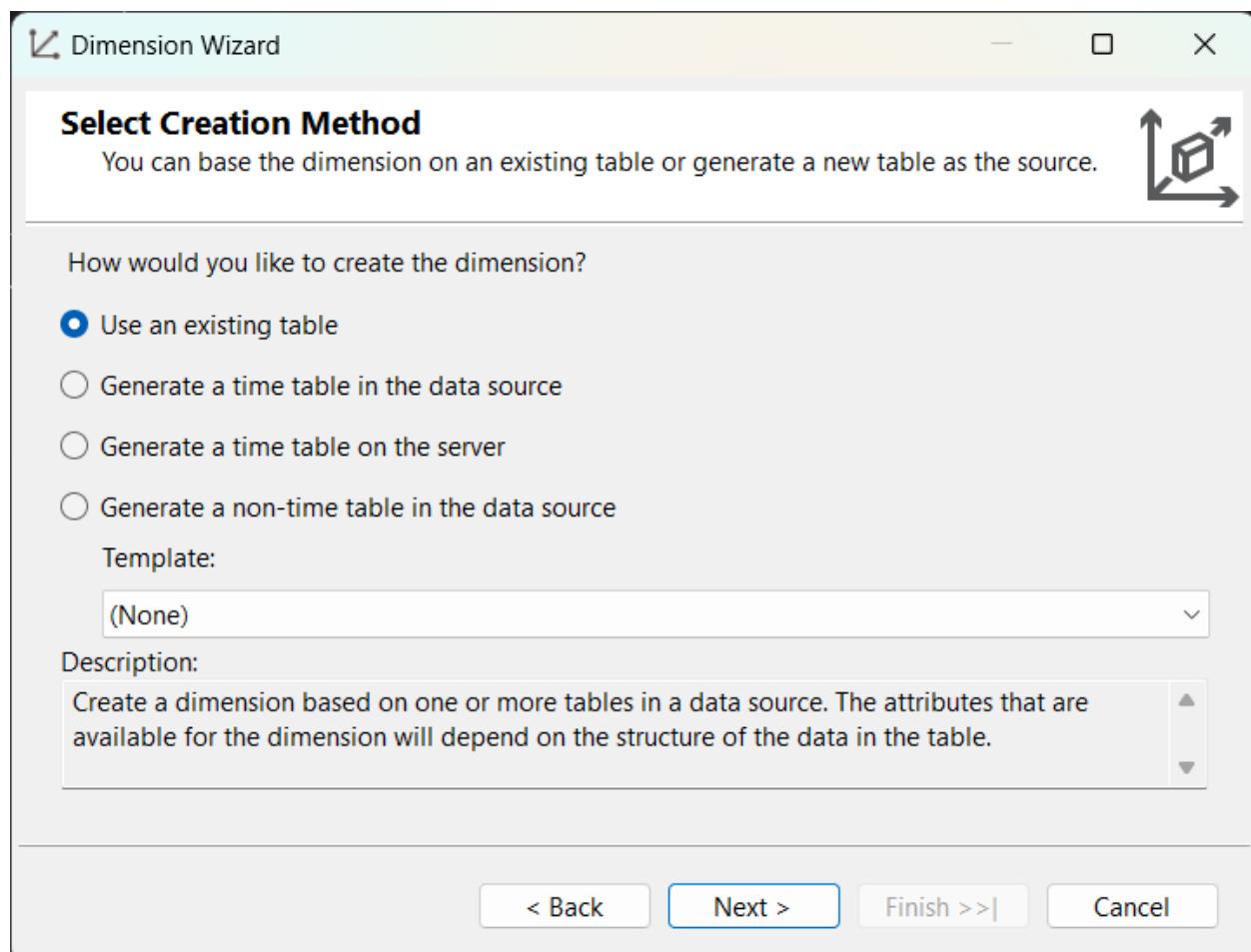
- Bước 1: Tại cửa sổ **Solution Explorer**, chuột phải vào **Dimensions** → **New Dimension** để thêm một chiều dữ liệu.



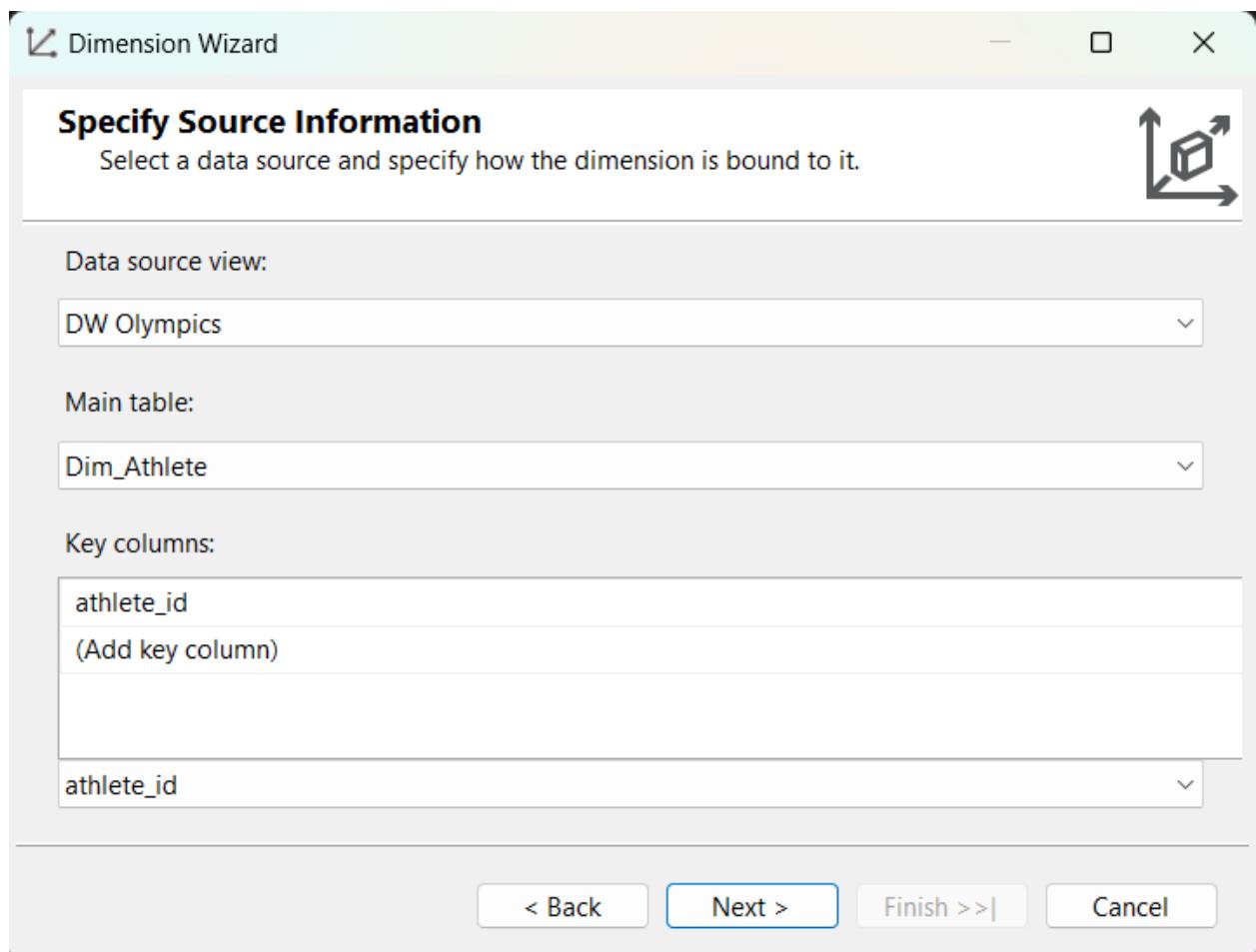
- Bước 2: Cửa sổ **Dimension Wizard** xuất hiện, nhấp **Next** để tiếp tục.



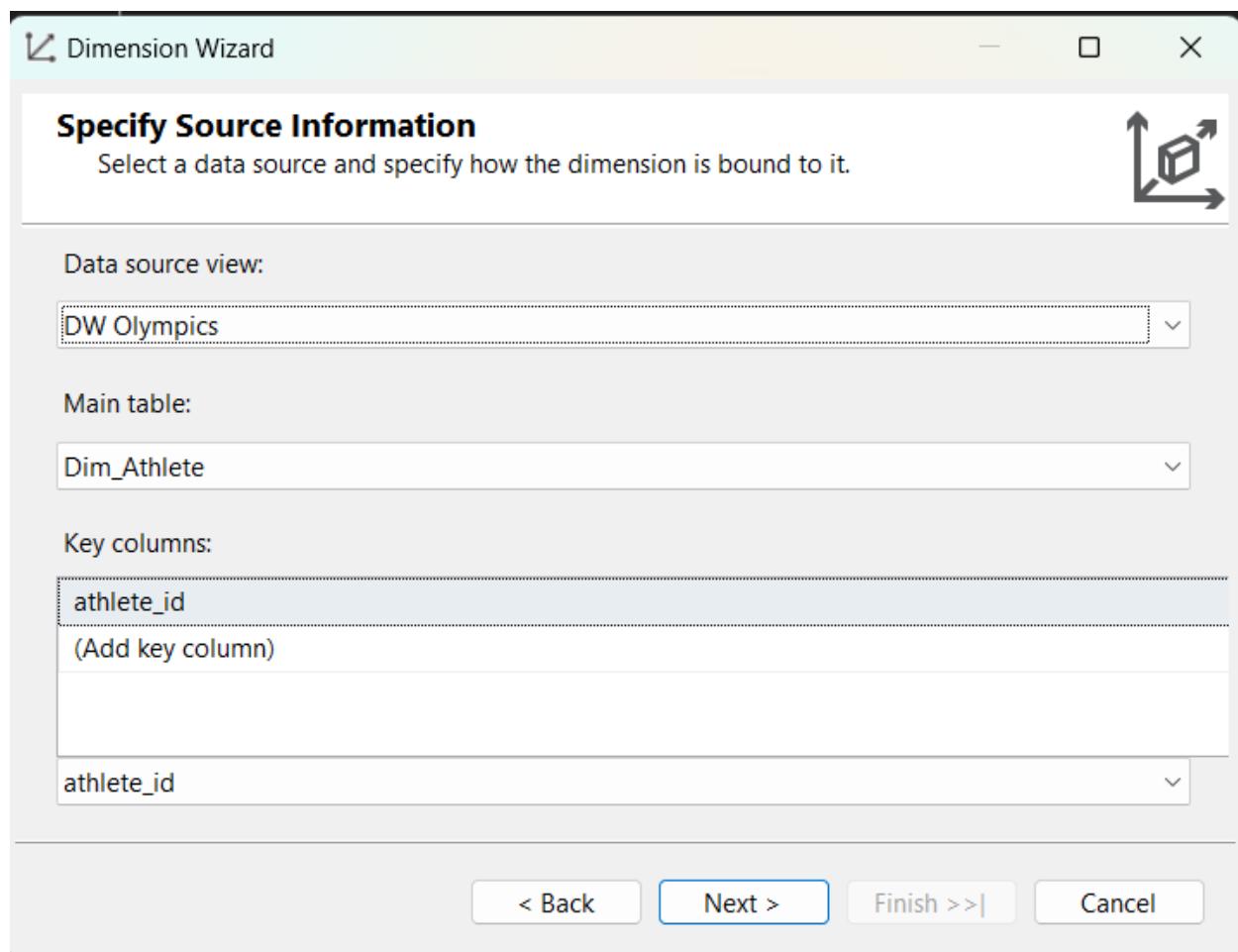
- Bước 3: Tại cửa sổ **Select Creation Method**, chọn **Use an Existing Table** → Next để chọn các bảng đã tồn tại.



- Bước 4: Tiến hành tạo các bảng chiều. Tại cửa sổ **Source Information**, chọn bảng cần tạo chiều dữ liệu → Next. Tại cửa sổ **Select Dimension Attributes**, chọn các thuộc tính cần thiết cho chiều dữ liệu đó.



+ Dim_Athlete



Dimension Wizard

Select Dimension Attributes

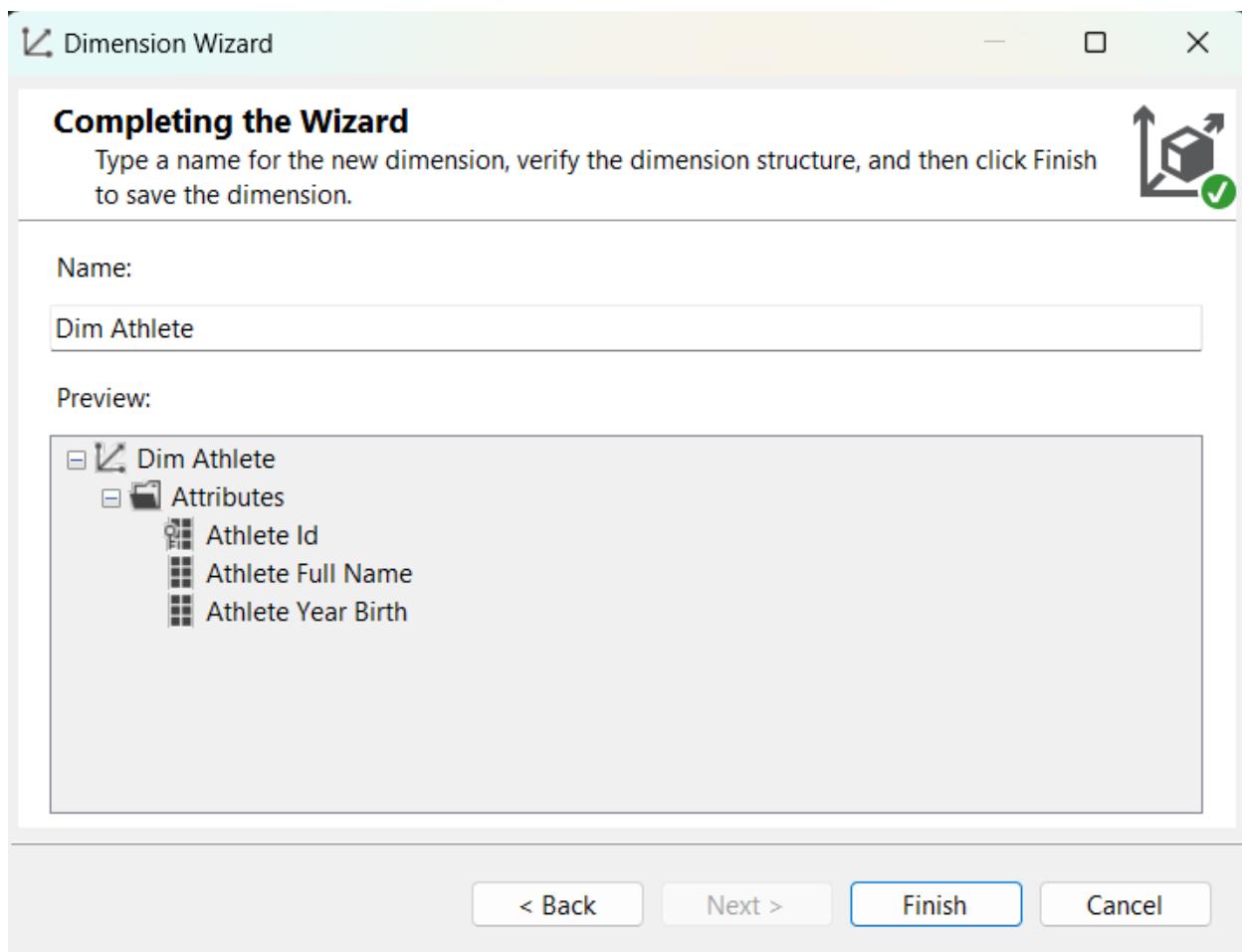
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.

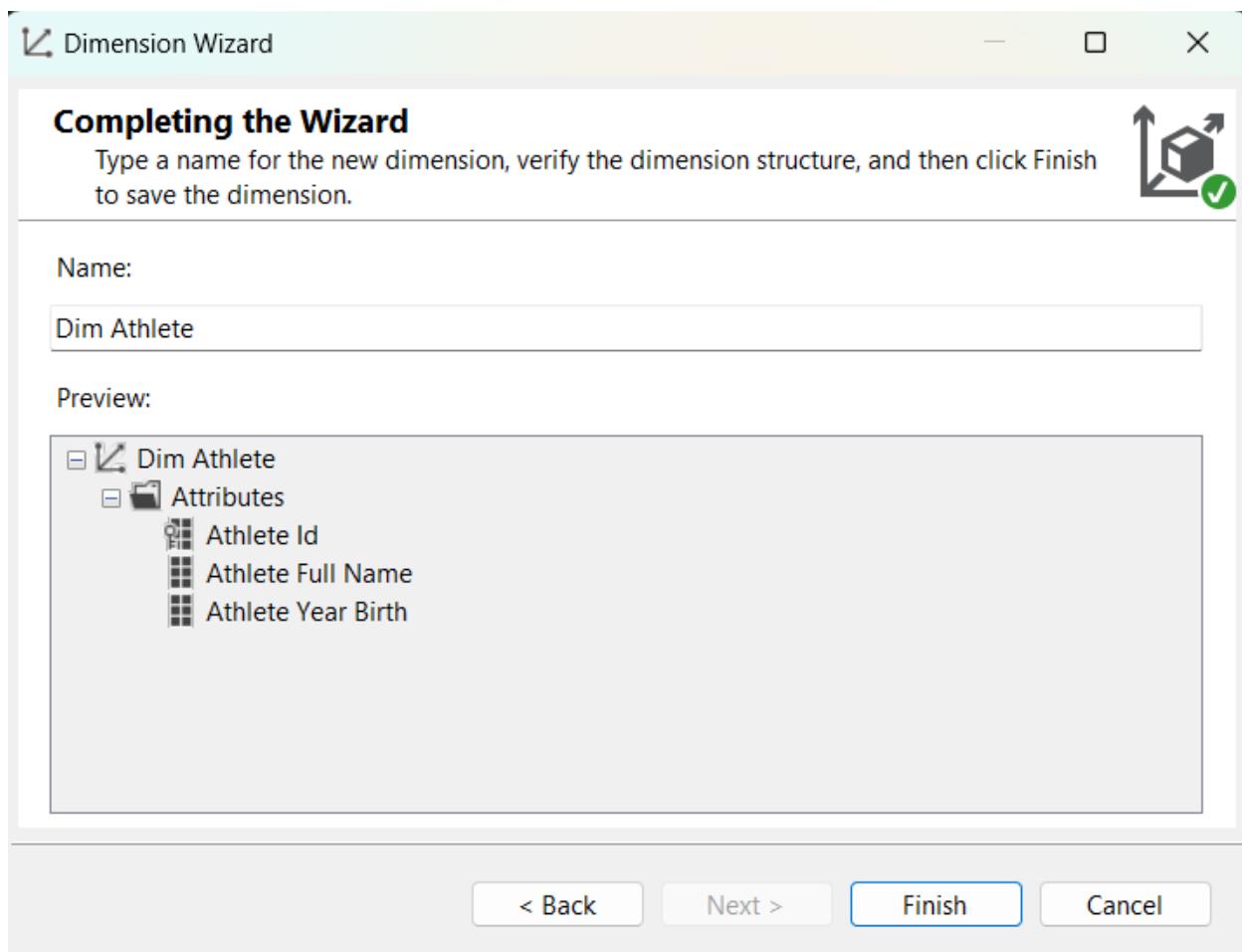


Available attributes:

Attribute Name	Enable Browsing	Attribute Type
Athlete Id	<input checked="" type="checkbox"/>	Regular
Athlete Full Name	<input checked="" type="checkbox"/>	Regular
Athlete Year Birth	<input checked="" type="checkbox"/>	Regular

< Back Next > Finish >> | Cancel





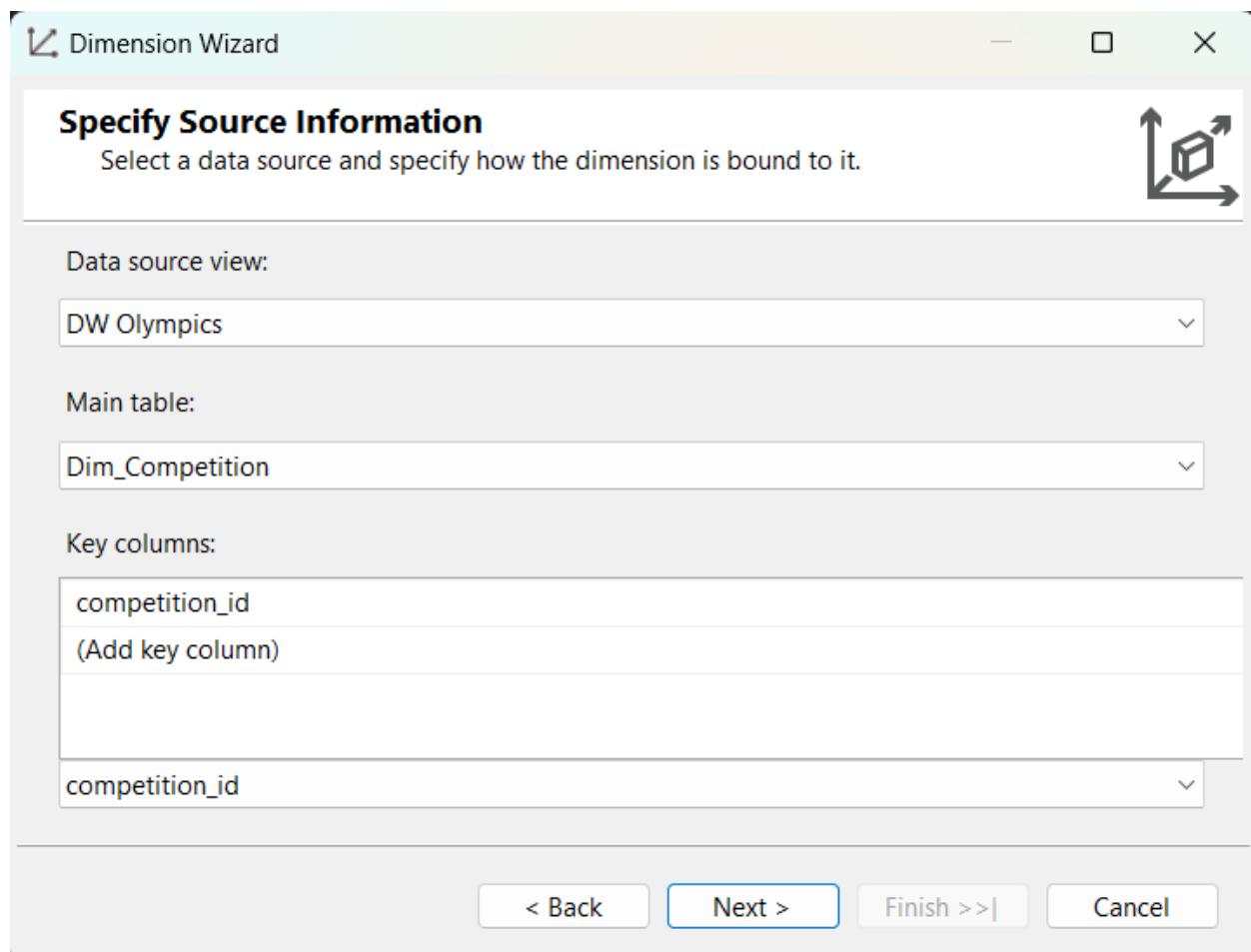
The screenshot shows the 'Dimension Structure' tool interface. The top navigation bar includes tabs for 'Dimension Struct...', 'Attribute Relationships', 'Translations', and 'Browser'. Below the tabs are standard window controls (minimize, maximize, close) and a toolbar with icons for new dimension, new hierarchy, new attribute, and search.

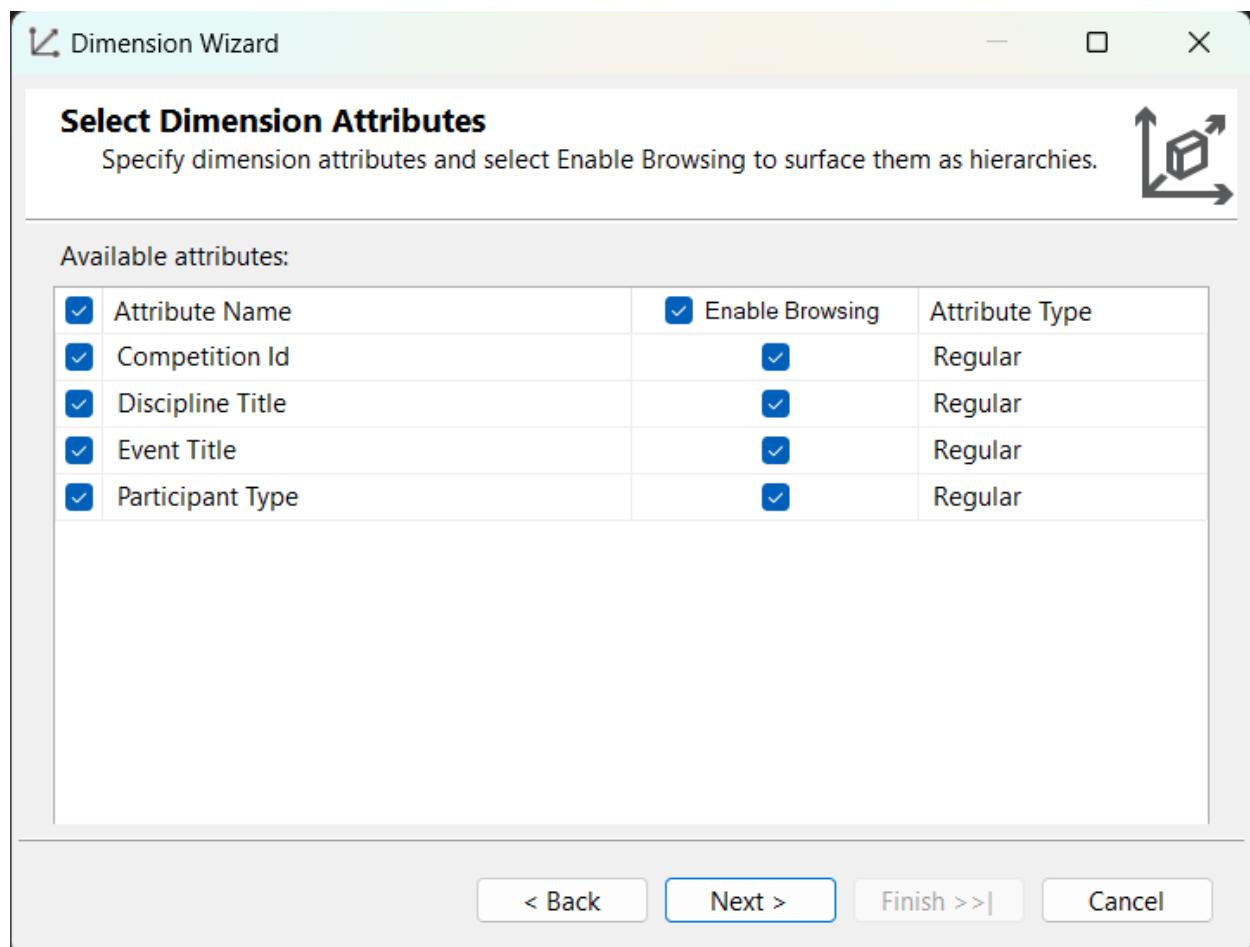
Attributes pane: Displays a tree view under 'Dim Athlete' with three attributes: 'Athlete Full Name', 'Athlete Id', and 'Athlete Year Birth'.

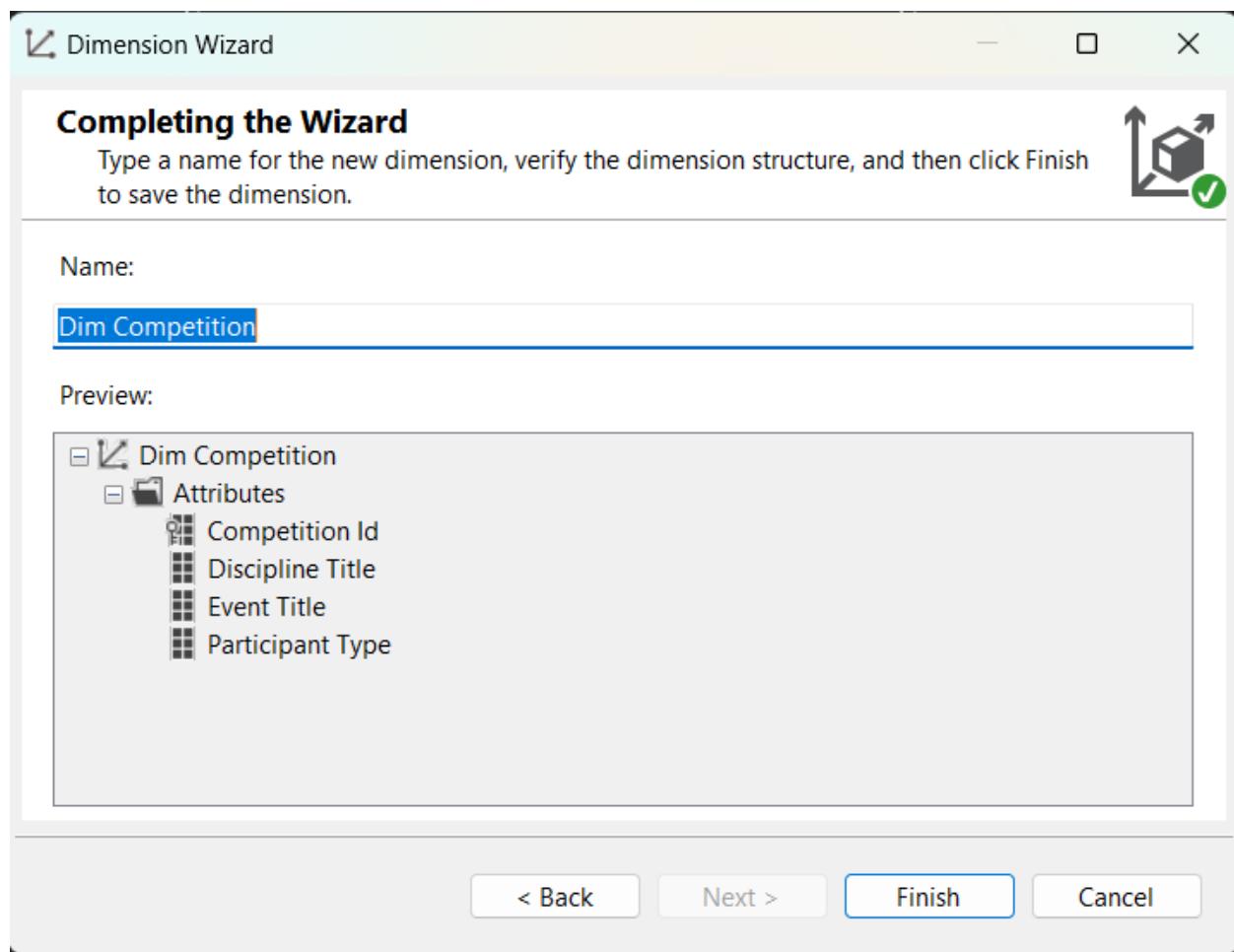
Hierarchies pane: Contains the message: 'To create a new hierarchy, drag an attribute here.'

Data Source View pane: Shows a list of attributes for 'Dim_Athlete': 'athlete_id', 'athlete_full_name', and 'athlete_year_birth'. The 'athlete_id' attribute is highlighted with a blue background.

+ Dim_Competition



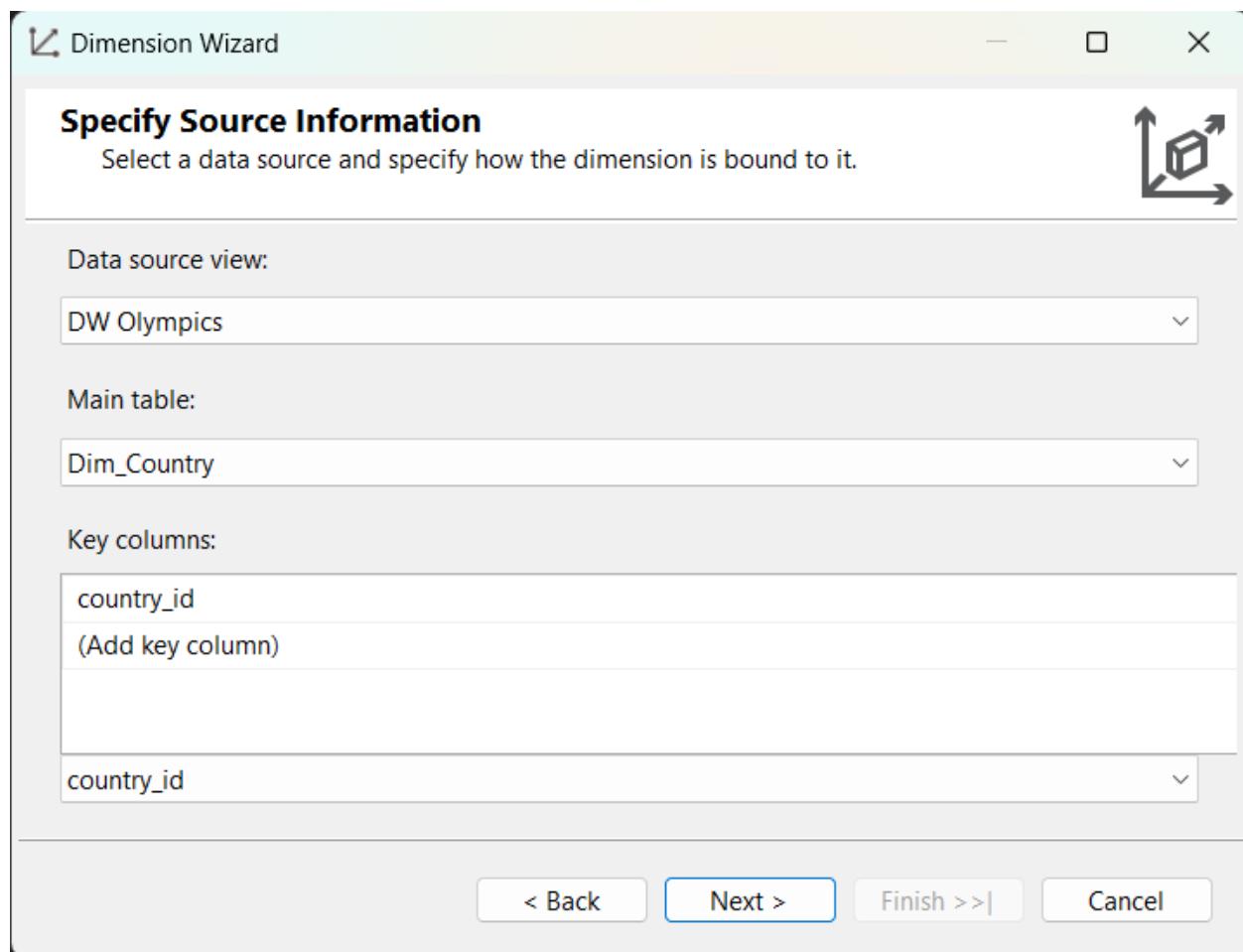




The screenshot shows the Dimension Structure tool interface with three main panes:

- Attributes** pane (left): Displays the selected dimension "Dim_Competition" with its attributes: Competition Id, Discipline Title, Event Title, and Participant Type.
- Hierarchies** pane (center): A placeholder area with the instruction: "To create a new hierarchy, drag an attribute here."
- Data Source View** pane (right): Shows the physical structure of the dimension with the table "Dim_Competition" and its columns: competition_id, discipline_title, event_title, and participant_type.

+ Dim_Country



Dimension Wizard

Select Dimension Attributes

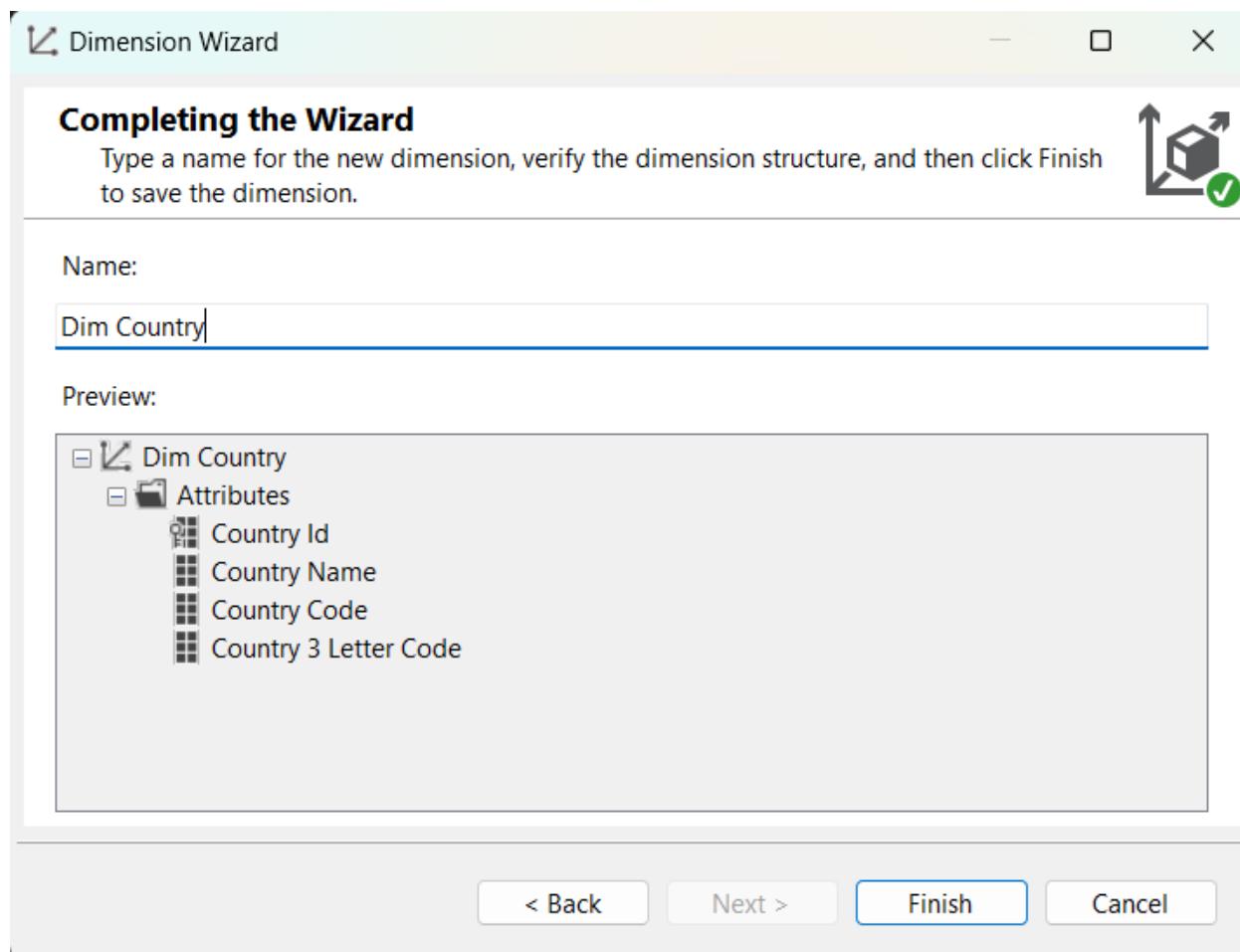
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.



Available attributes:

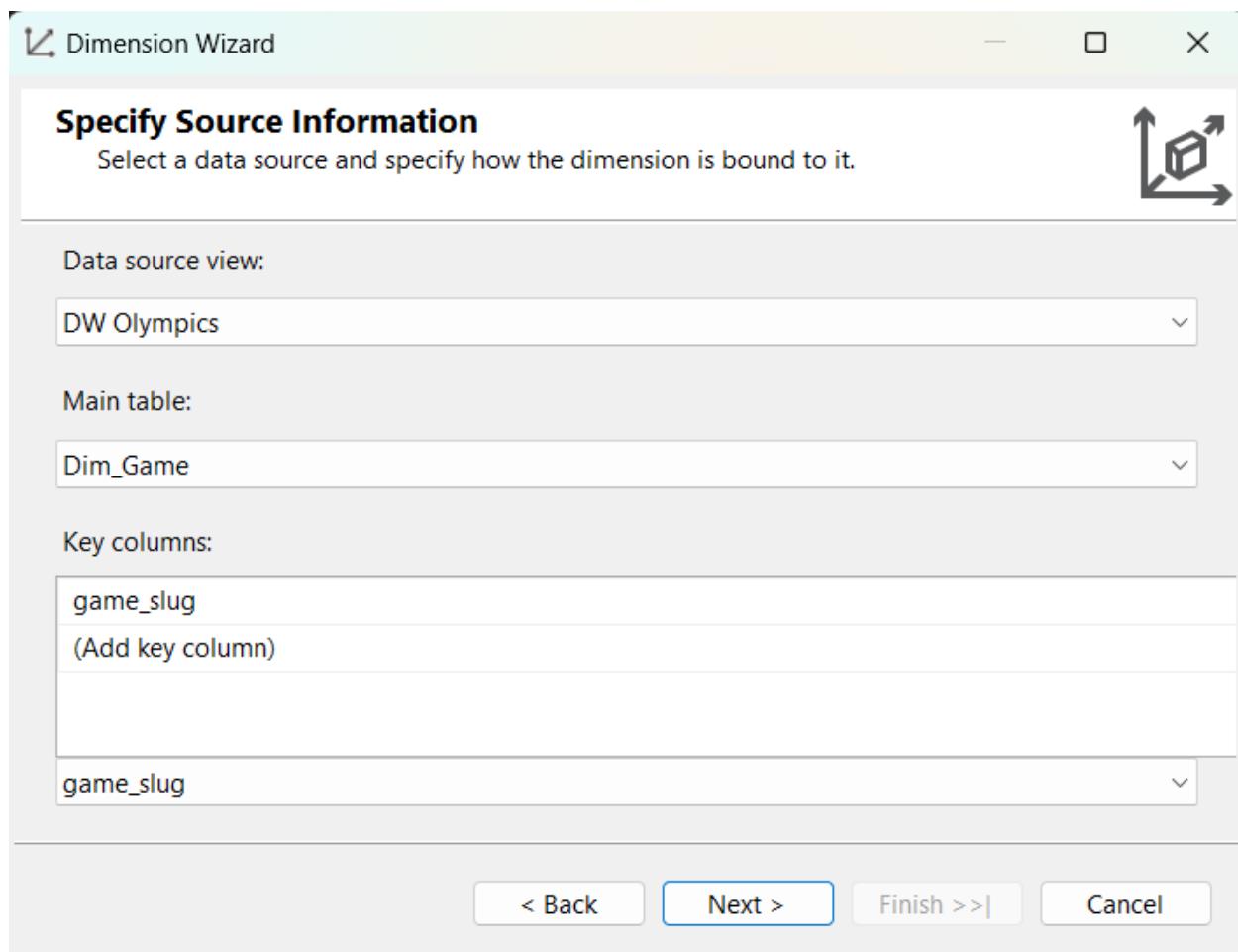
	Attribute Name	Enable Browsing	Attribute Type
<input checked="" type="checkbox"/>	Attribute Name	<input checked="" type="checkbox"/>	Regular
<input checked="" type="checkbox"/>	Country Id	<input checked="" type="checkbox"/>	Regular
<input checked="" type="checkbox"/>	Country Name	<input checked="" type="checkbox"/>	Regular
<input checked="" type="checkbox"/>	Country Code	<input checked="" type="checkbox"/>	Regular
<input checked="" type="checkbox"/>	Country 3 Letter Code	<input checked="" type="checkbox"/>	Regular

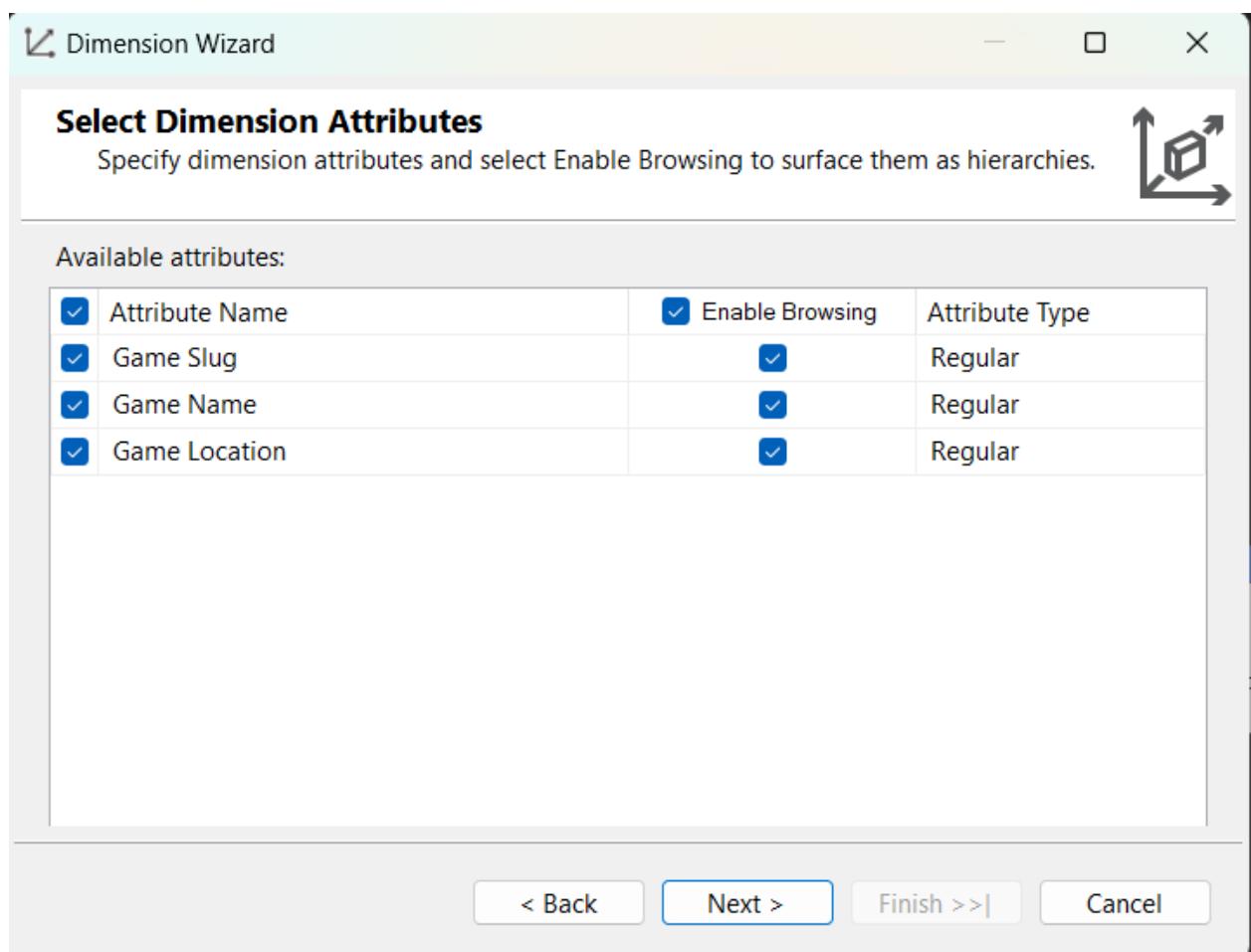
< Back Next > Finish >> Cancel

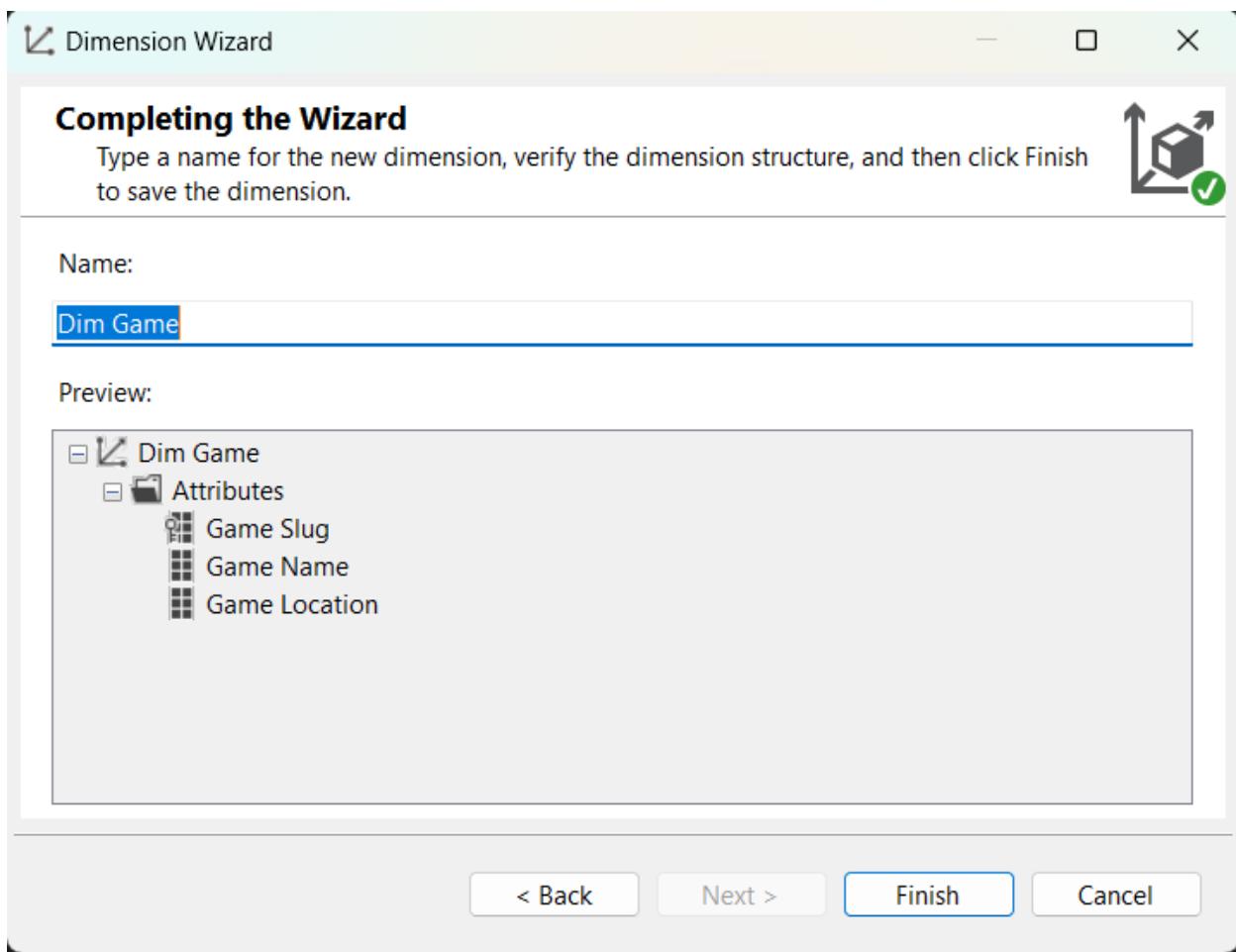


The screenshot shows the 'Dimension Structure' tool interface. The top menu bar includes 'Dimension Struct...', 'Attribute Relationships', 'Translations', and 'Browser'. The left panel, titled 'Attributes', contains a tree view with a checked node 'Dim Country' expanded to show four children: 'Country 3 Letter Code', 'Country Code', 'Country Id', and 'Country Name'. The middle panel, titled 'Hierarchies', has a placeholder message: 'To create a new hierarchy, drag an attribute here.' The right panel, titled 'Data Source View', displays a table structure for 'Dim_Country' with columns: country_id, country_name, country_code, and country_3_letter_code.

+ Dim_Game



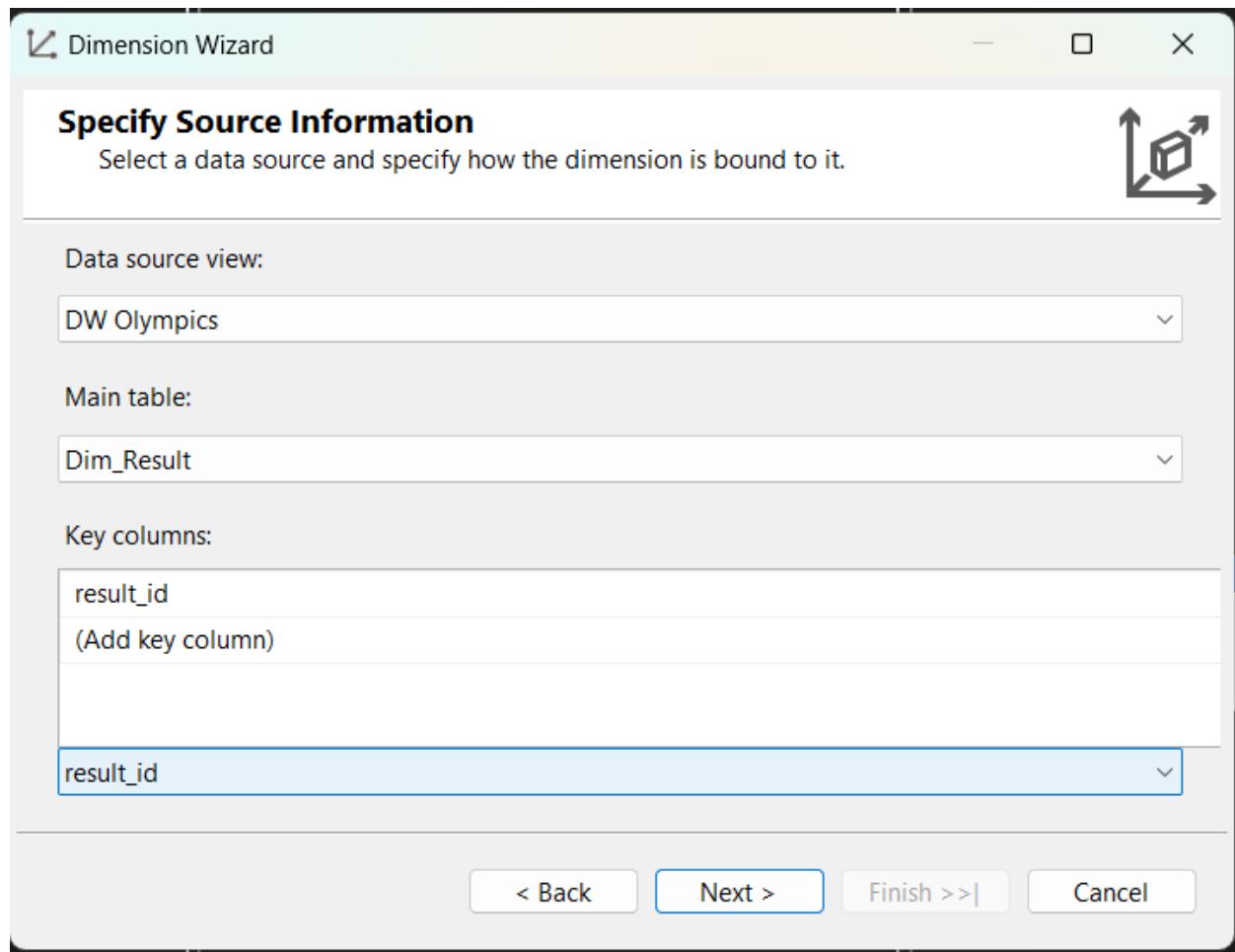


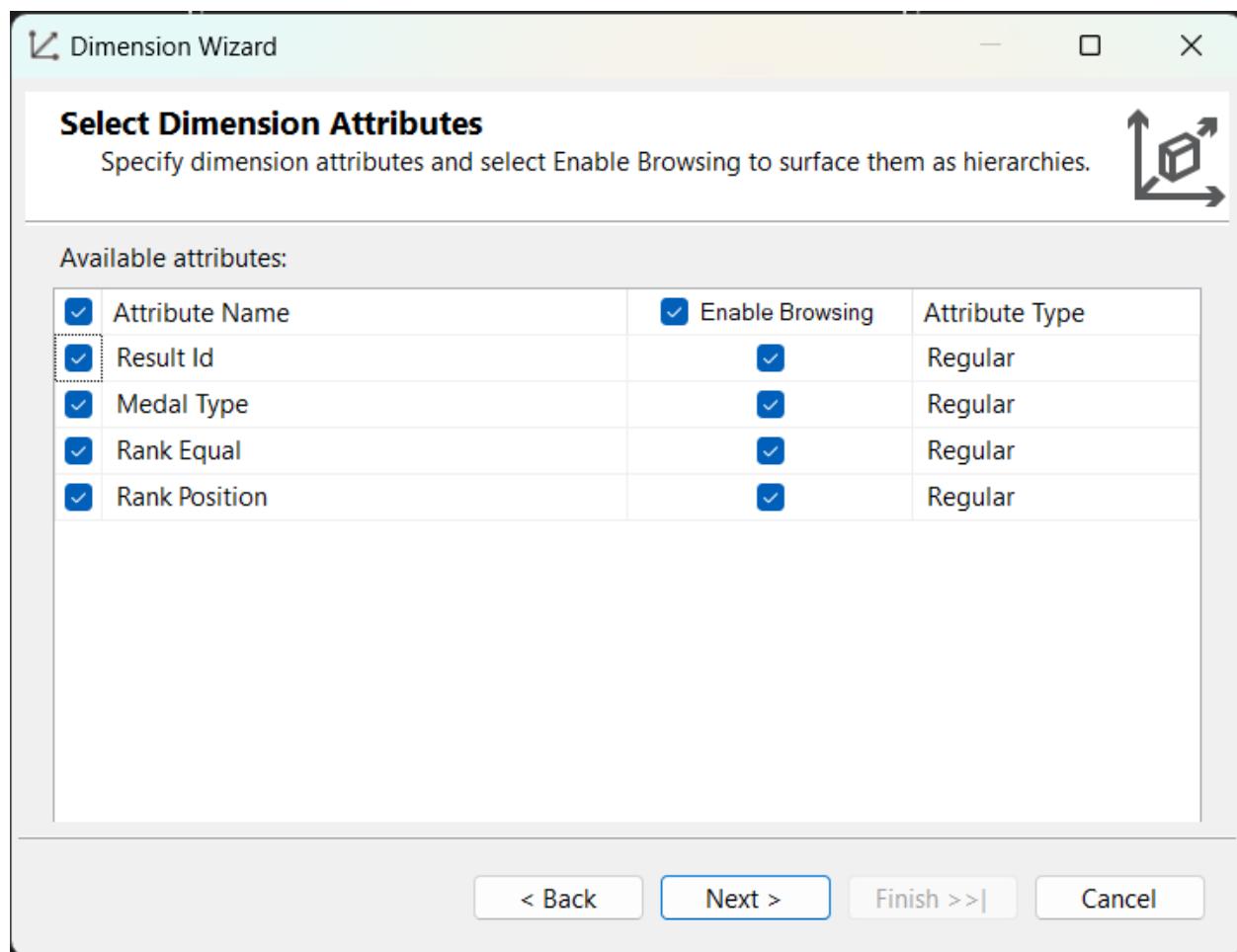


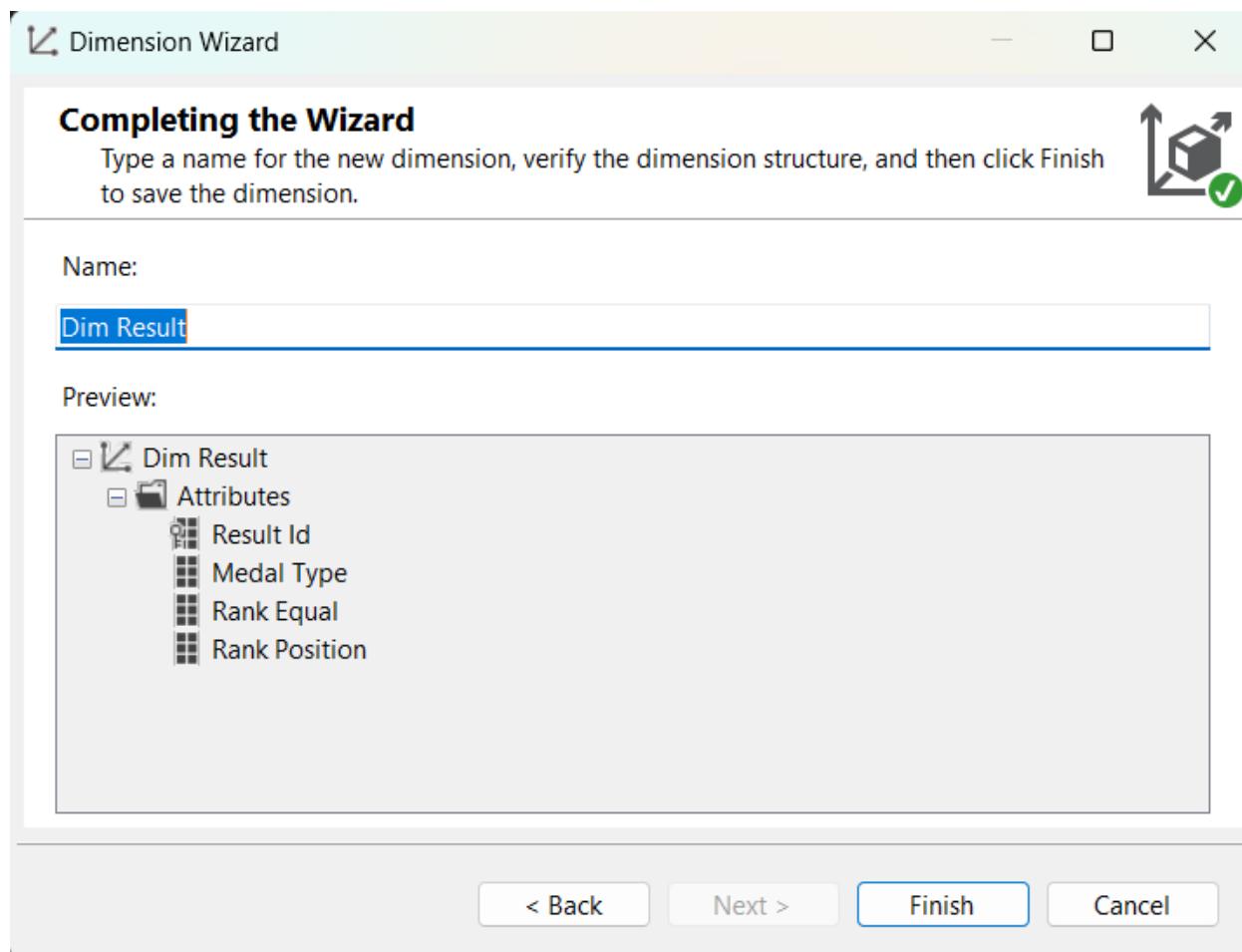
The screenshot shows a software interface for data modeling, likely a data warehouse or OLAP tool. The interface is divided into three main sections:

- Attributes:** On the left, a tree view under the heading "Dim Game" lists three attributes: "Game Location", "Game Name", and "Game Slug".
- Hierarchies:** In the center, there is a placeholder area with the text "To create a new hierarchy, drag an attribute here."
- Data Source View:** On the right, a preview window titled "Dim_Game" shows the three attributes: "game_slug", "game_name", and "game_location".

+ Dim_Result

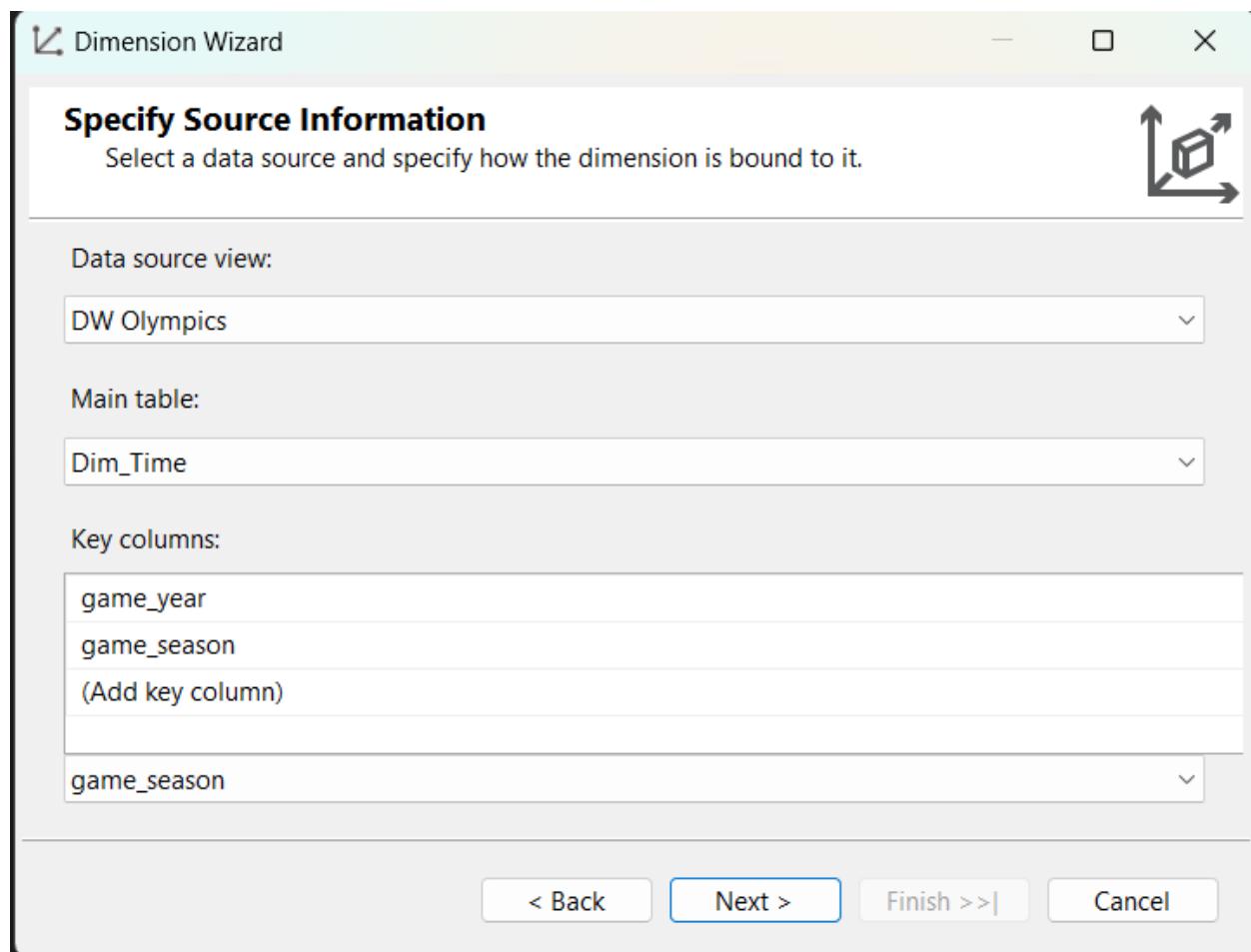


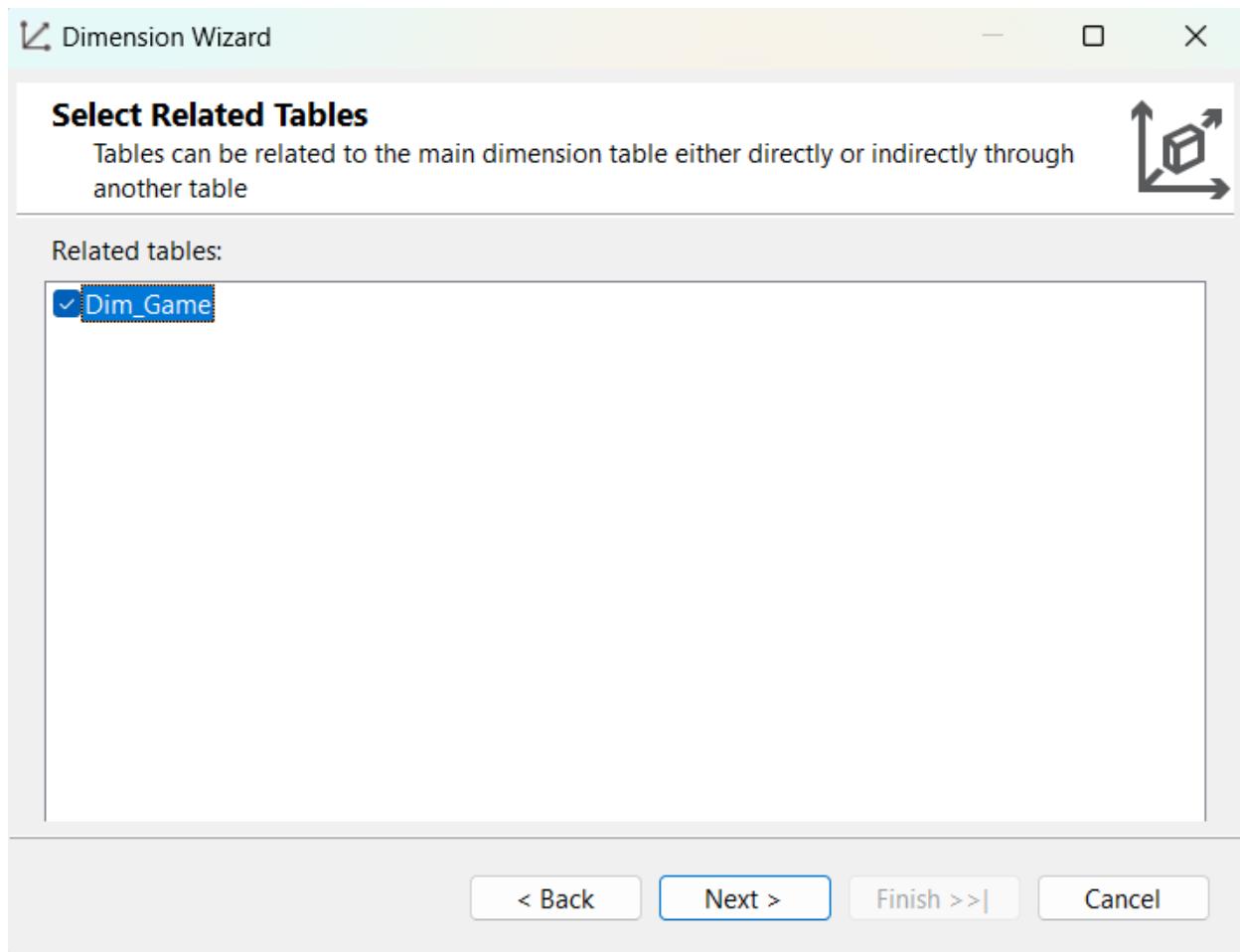




Attributes	Hierarchies	Data Source View
<ul style="list-style-type: none"><input checked="" type="checkbox"/> Dim_Result<ul style="list-style-type: none">Medal TypeRank EqualRank PositionResult Id	To create a new hierarchy, drag an attribute here.	<ul style="list-style-type: none"><input checked="" type="checkbox"/> Dim_Result<ul style="list-style-type: none">result_idmedal_typerank_equalrank_position

+ Dim_Time





Dimension Wizard

Select Dimension Attributes

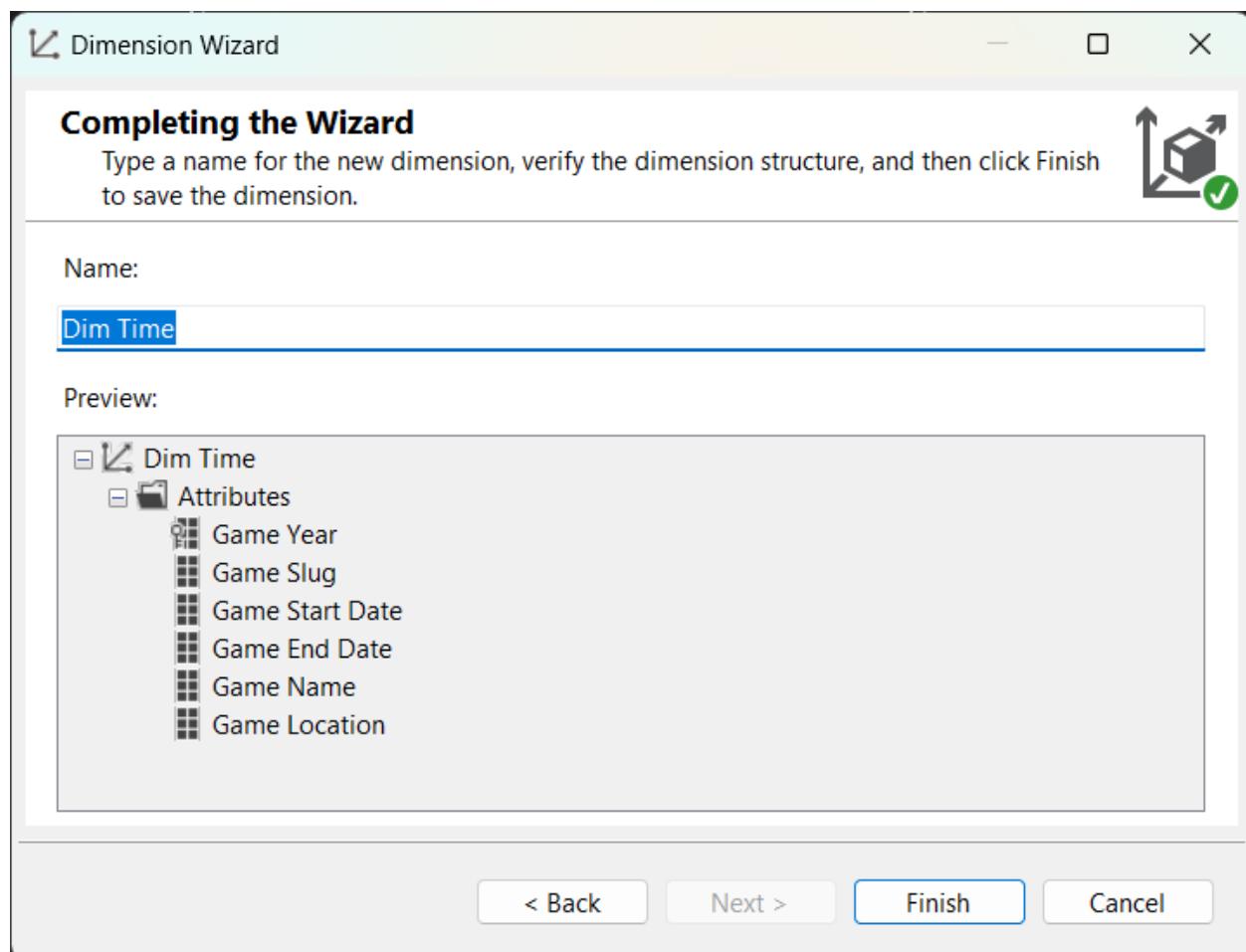
Specify dimension attributes and select Enable Browsing to surface them as hierarchies.

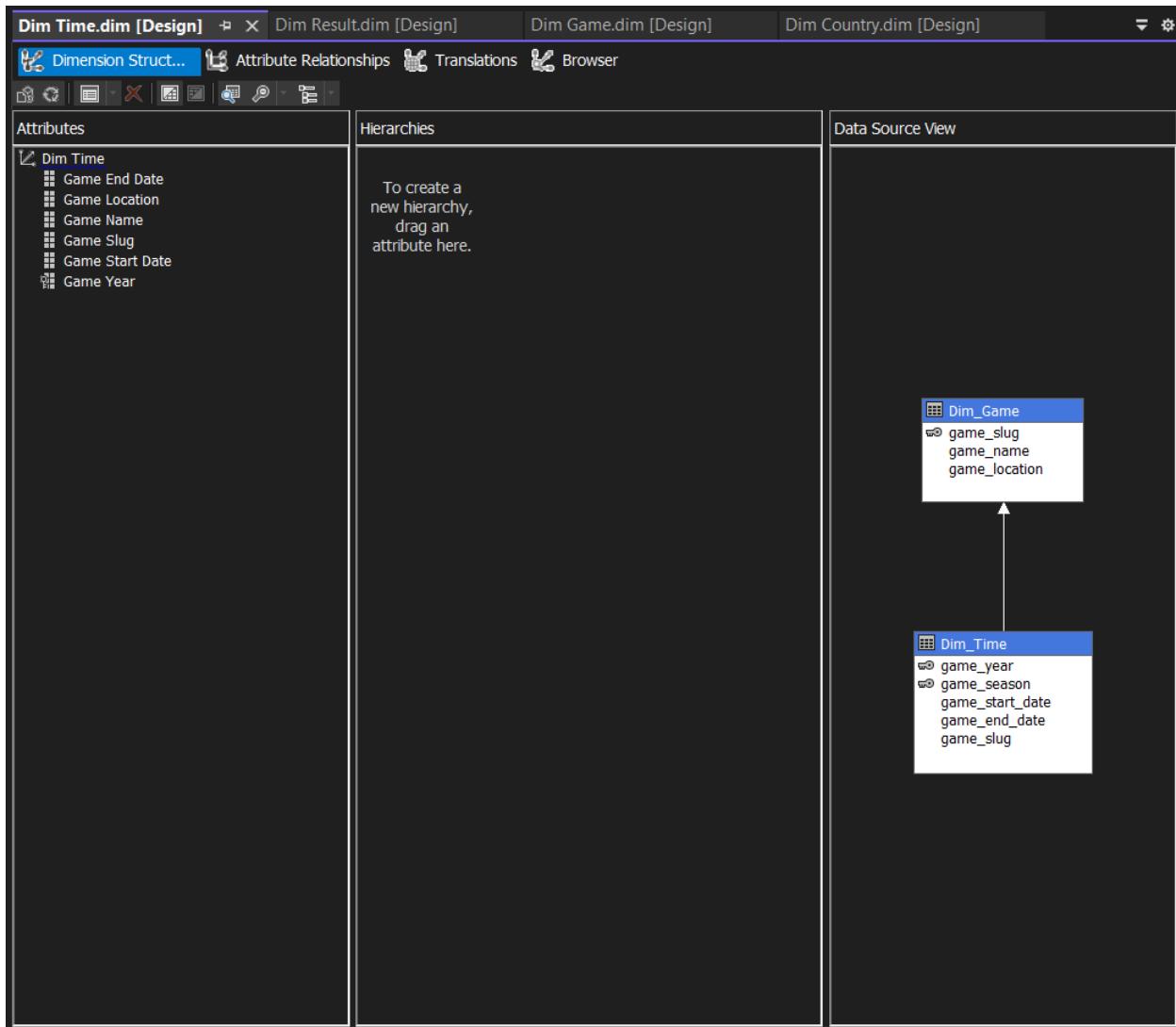


Available attributes:

Attribute Name	Enable Browsing	Attribute Type
Game Year	<input checked="" type="checkbox"/>	Regular
Game Start Date	<input checked="" type="checkbox"/>	Regular
Game End Date	<input checked="" type="checkbox"/>	Regular
Game Slug	<input checked="" type="checkbox"/>	Regular
Game Name	<input checked="" type="checkbox"/>	Regular
Game Location	<input checked="" type="checkbox"/>	Regular

< Back Next > | Finish >> Cancel

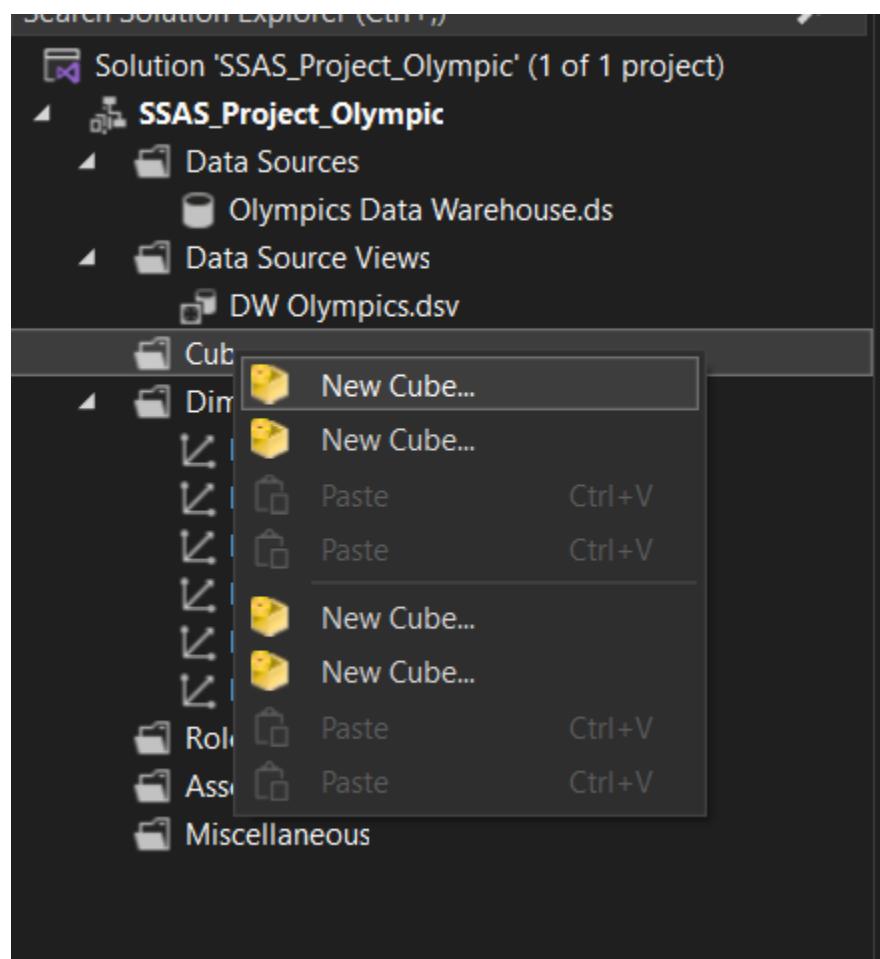




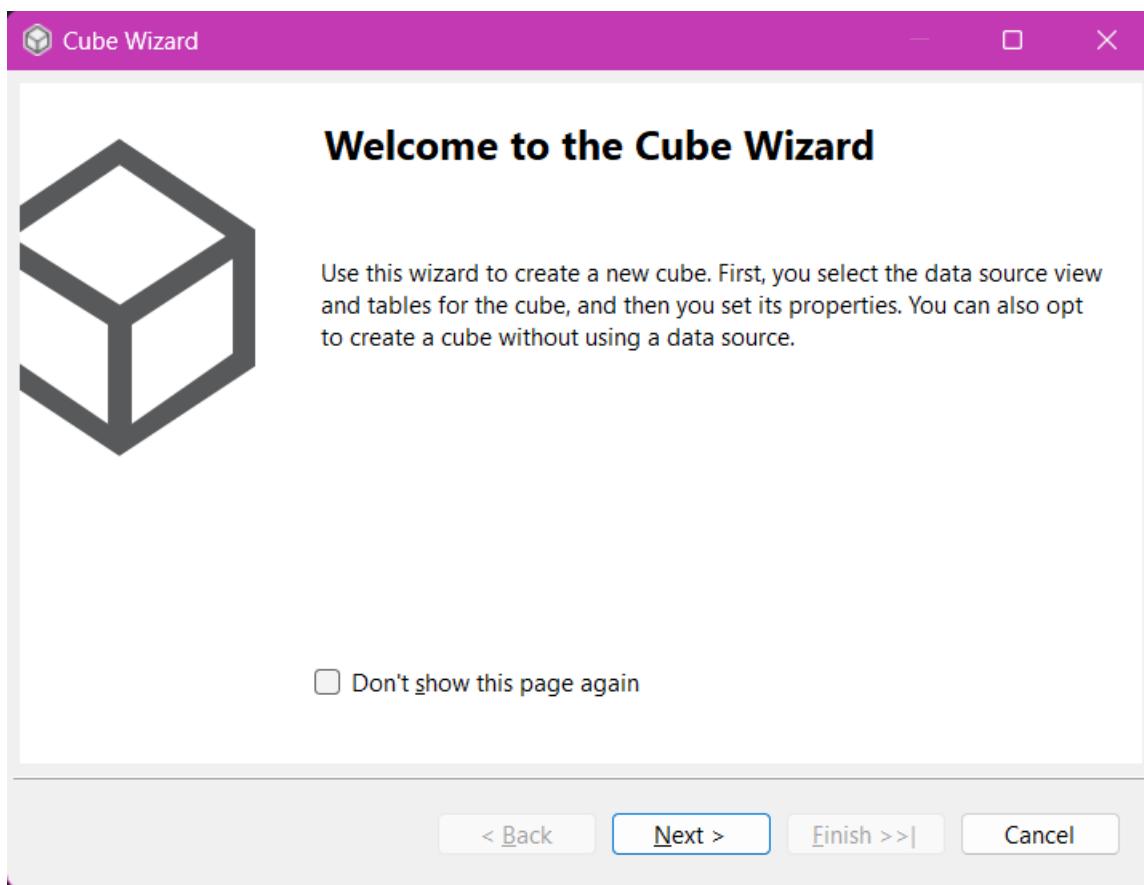
5. XÁC ĐỊNH CÁC KHỐI VÀ ĐO LƯỜNG

Để xác định khối và các độ đo lường (Define Cubes & Measures) cho một dự án SSAS mới, ta thực hiện các bước sau:

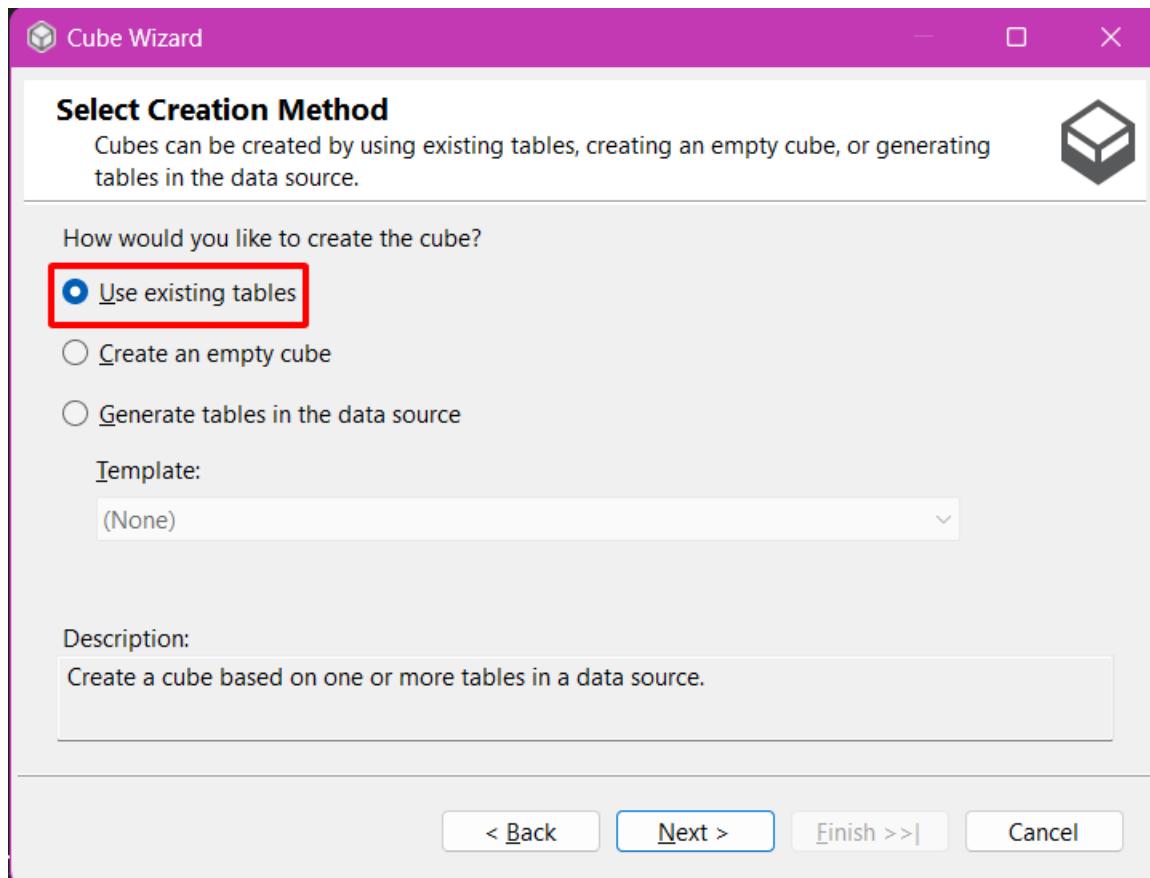
- Bước 1: Tại cửa sổ **Solution Explorer**, chuột phải vào **Cubes** → **New Cube** để thêm một khối dữ liệu mới.



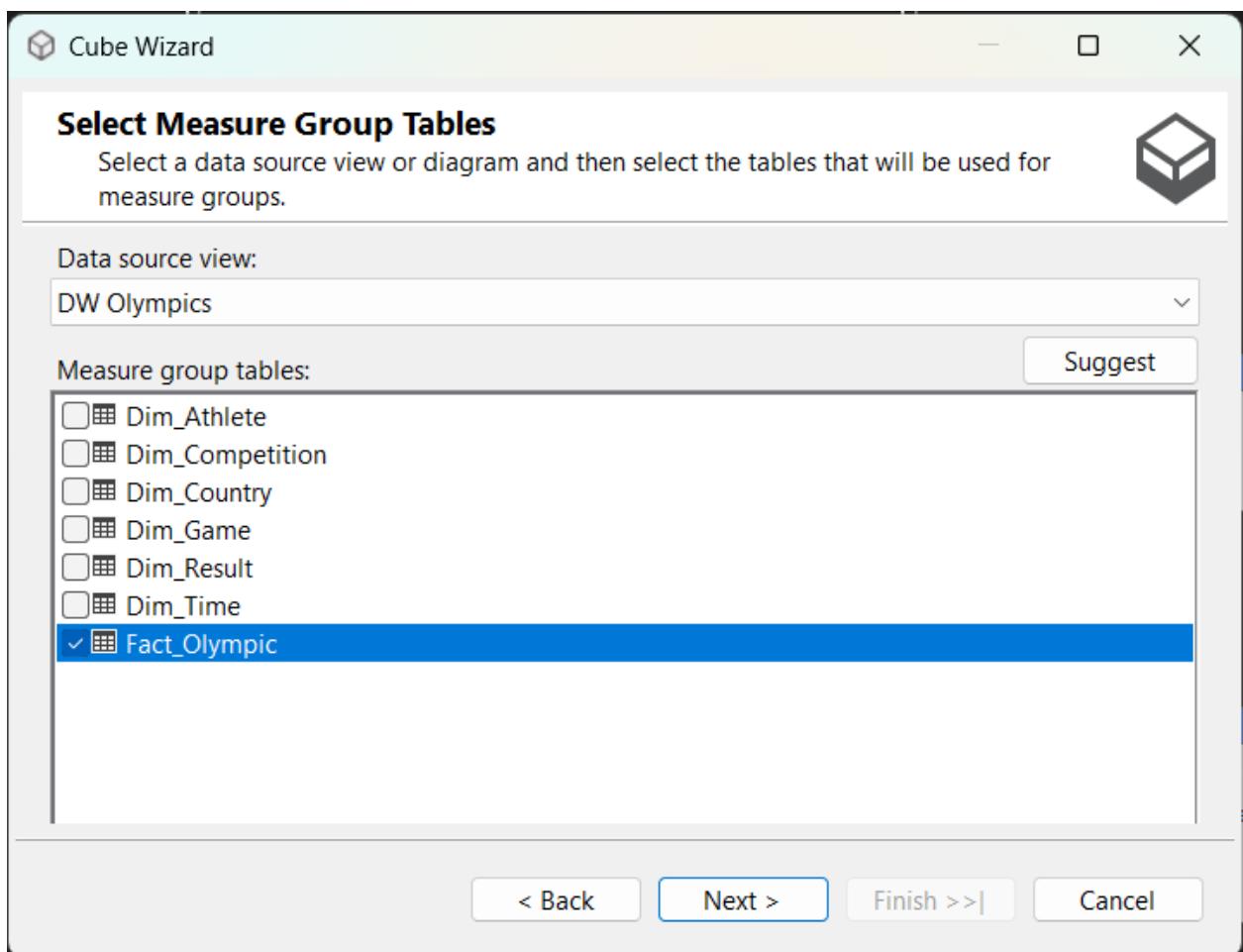
- Bước 2: Cửa sổ **Cube Wizard** xuất hiện, nhấn **Next** để tiếp tục.



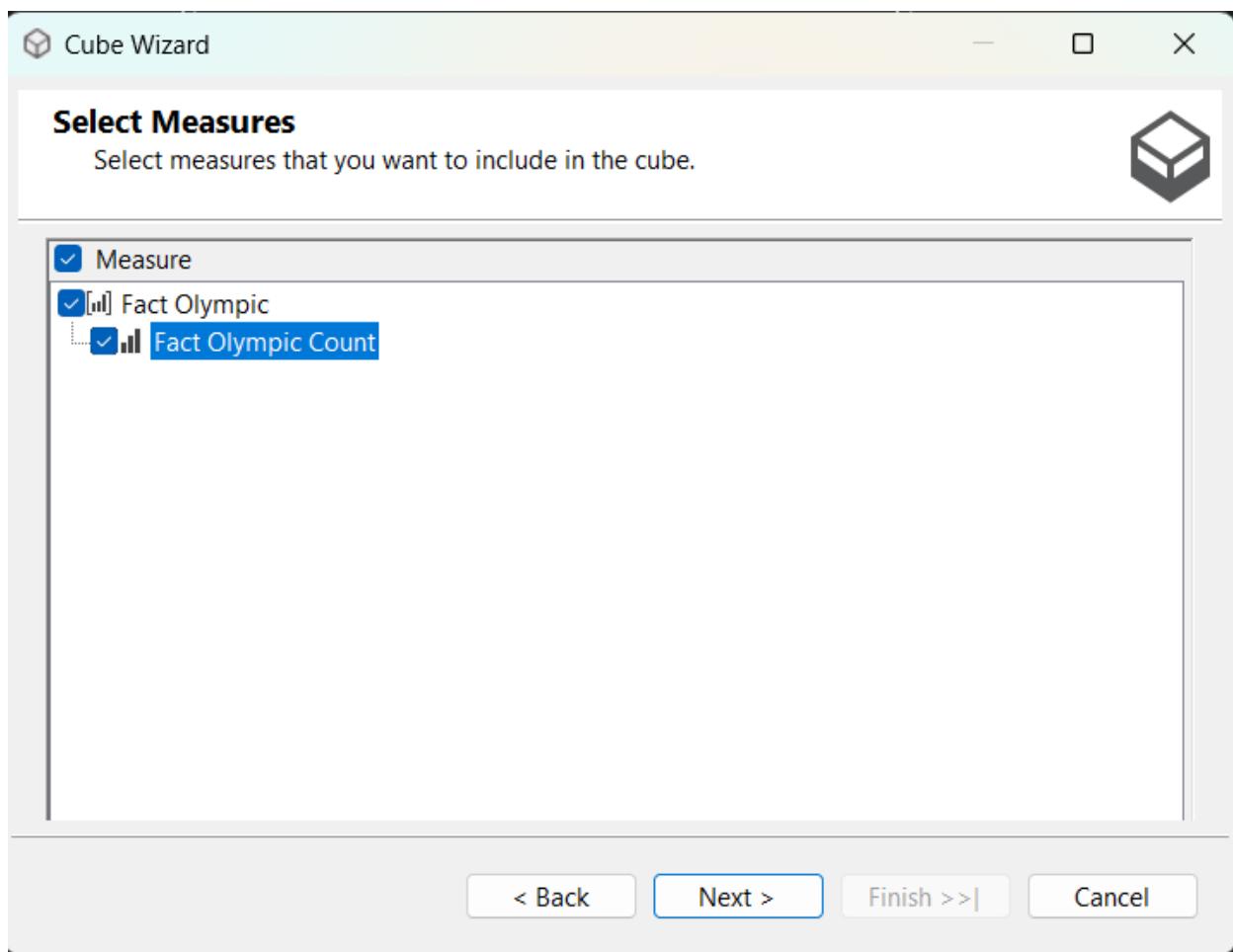
- Bước 3: Tại cửa sổ **Select Creation Method**, chọn **Use an Existing Table** → Next để chọn các bảng đã tồn tại.



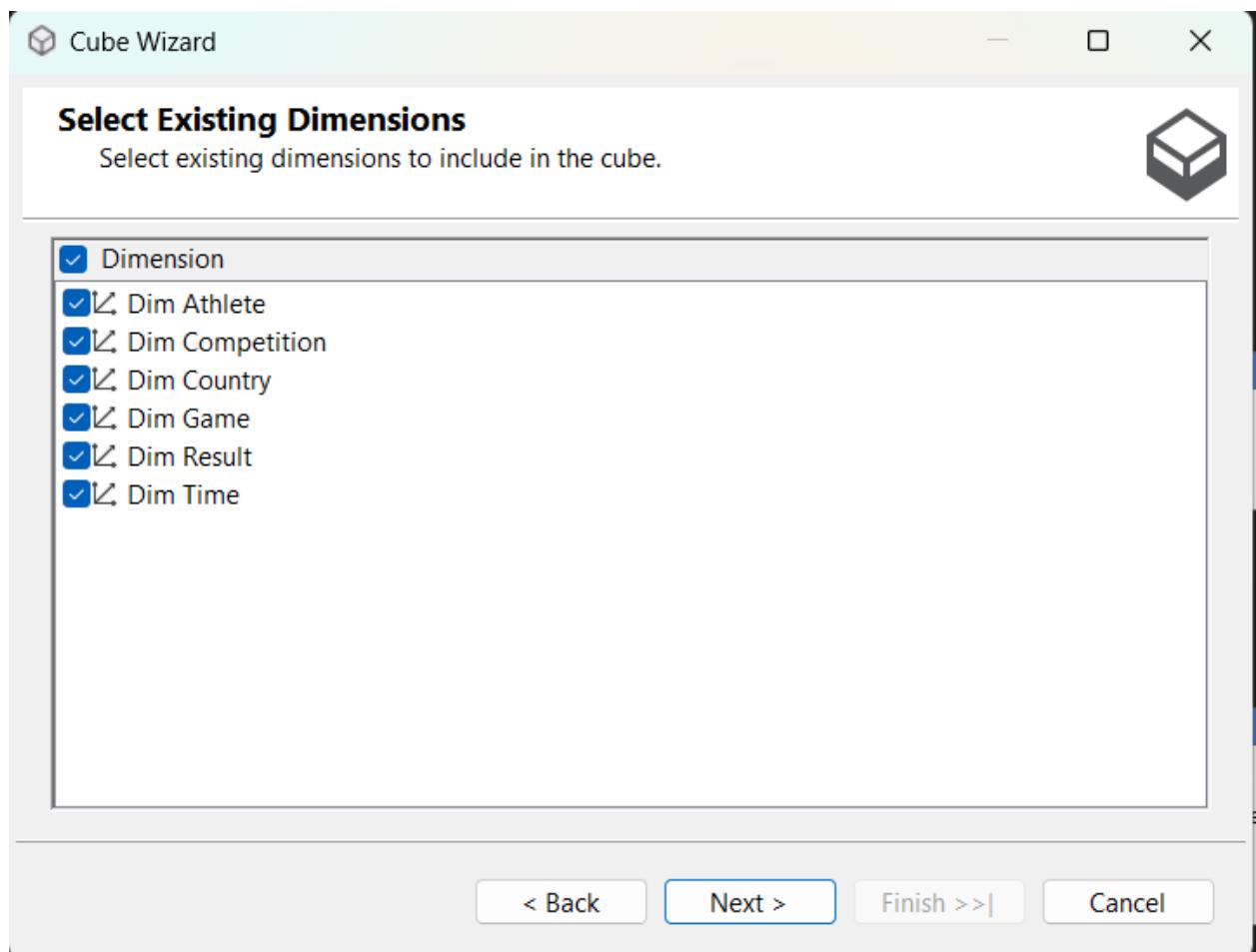
- Bước 4: Chọn bảng **Fact_Olympic** để phân chia các nhóm đo đạc (Measure Groups)



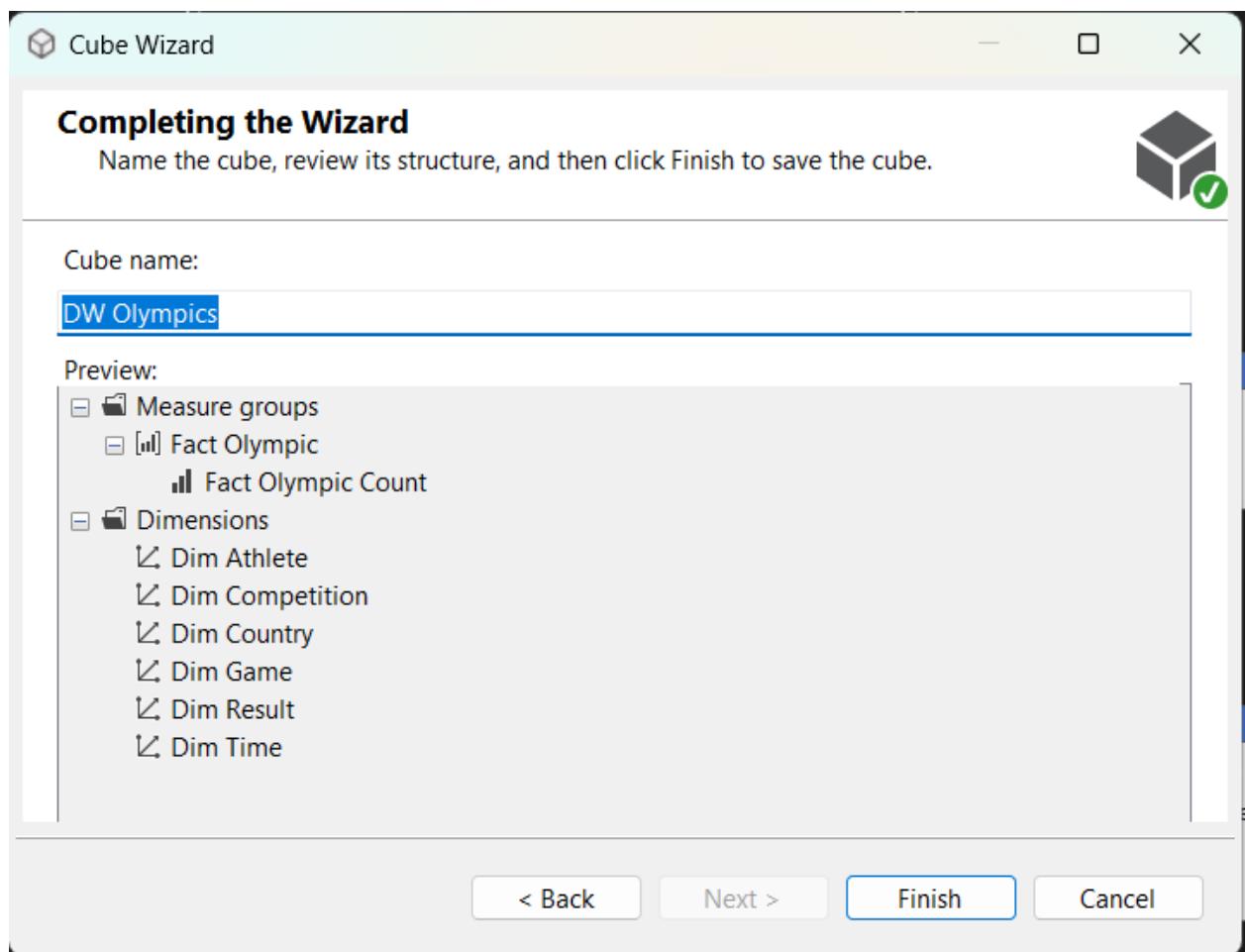
- Bước 5: Chọn các thuộc tính định lượng trong bảng **Fact_Olympic** được đề xuất.



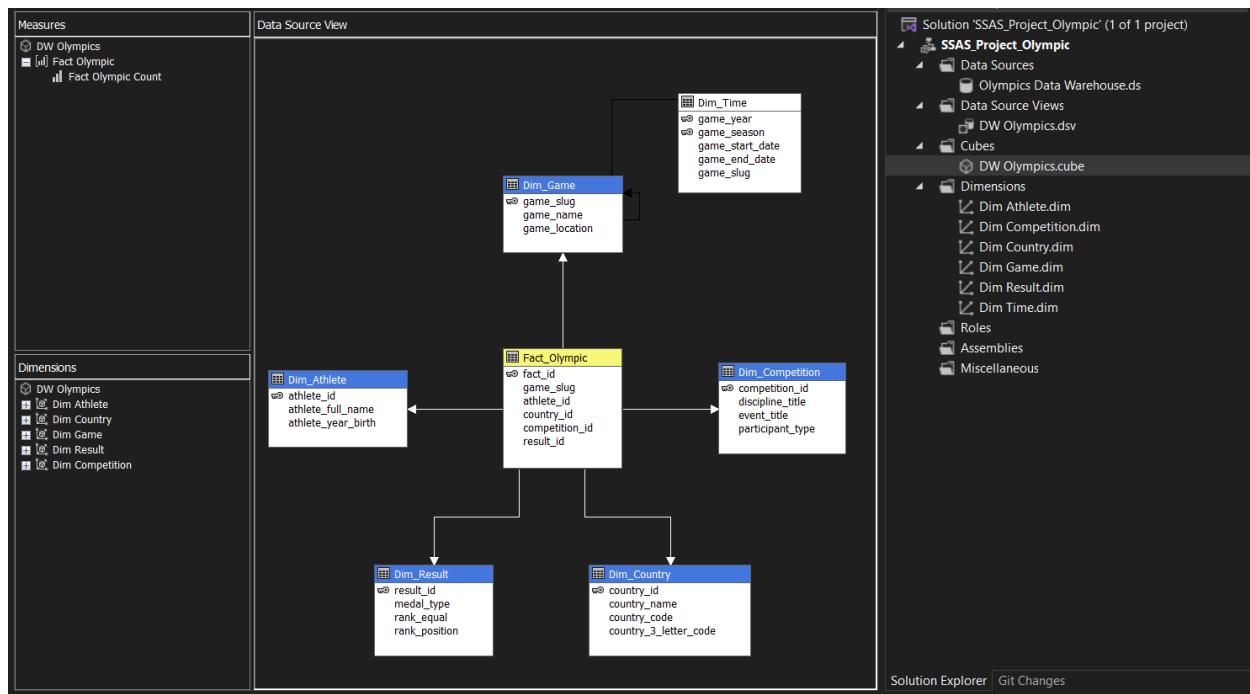
- Bước 6: Chọn danh sách các **Dimension** đã tạo (chương 3 mục 4), chọn **Next** để tiếp tục.



- Bước 7: Khởi tạo chiều dữ liệu (**Dimension mới**) trên bảng *Fact_Olympic*, chọn Next để tiếp tục.



- Bước 8: Chọn **Finish** để hoàn tất quá trình tạo Cube

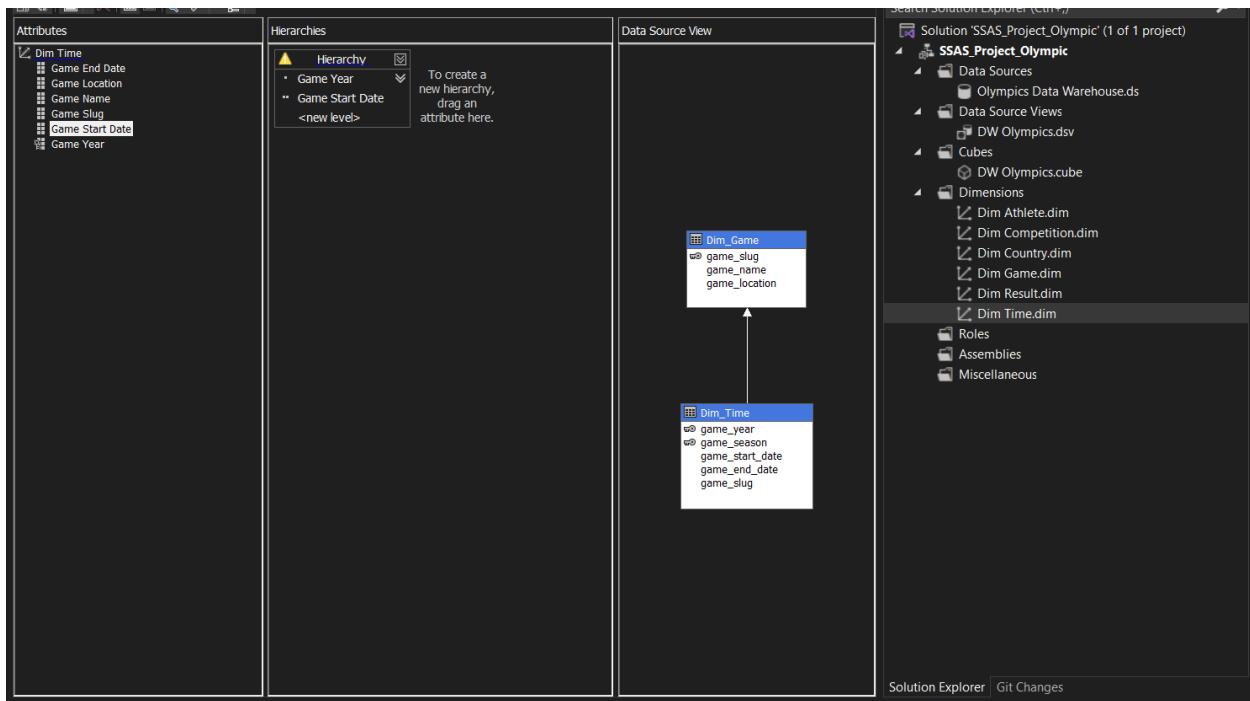


6. PHÂN CẤP BẢNG CHIỀU

Bảng Dim_Time

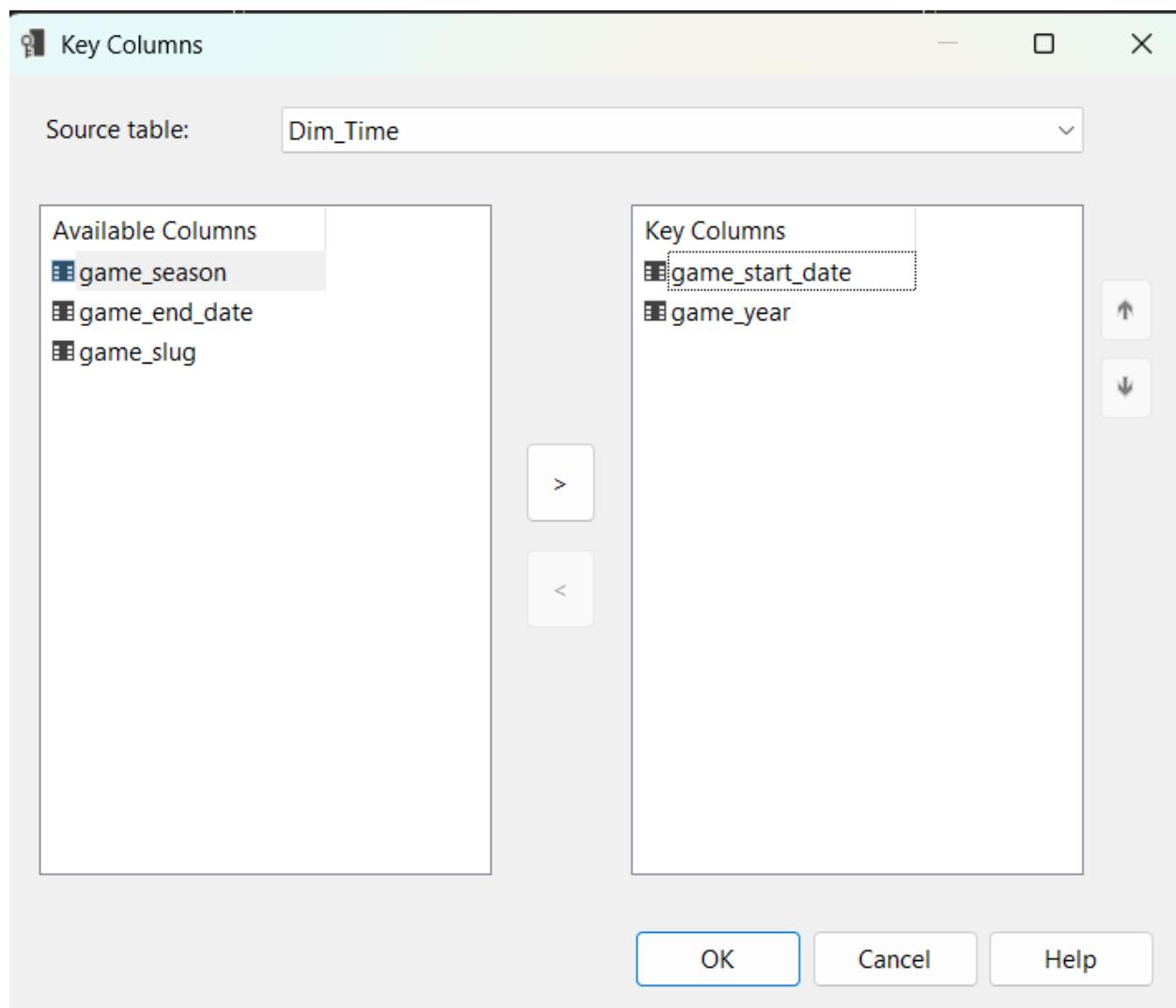
Để phân cấp trong bảng chiều *Dim_Time*, ta thực hiện các bước sau:

- Bước 1: Tại tệp *Dim_Time.dim*, kéo những thuộc tính cần phân cấp từ cửa sổ **Attributes** qua cửa sổ **Hierarchies**.

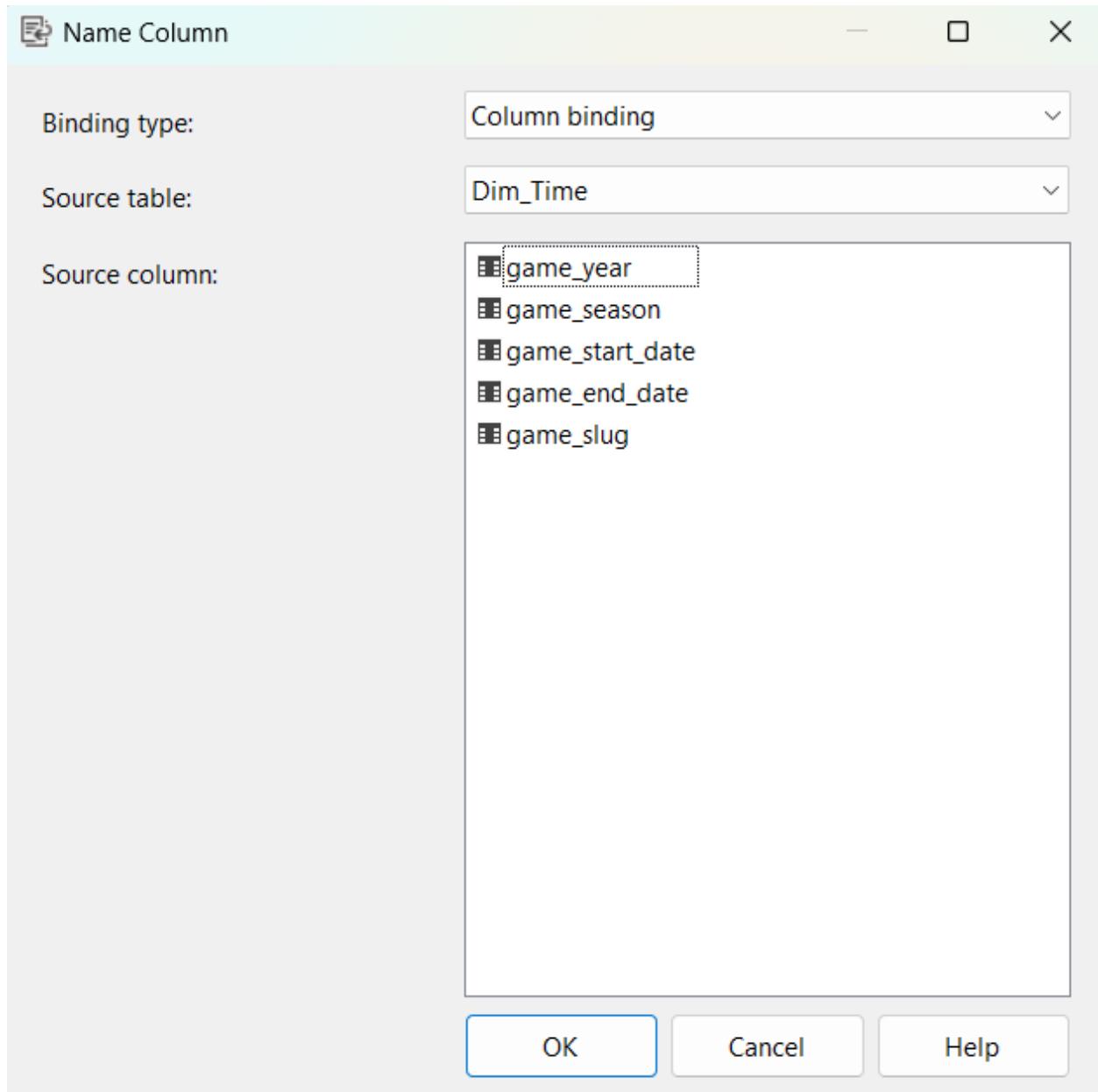


- Bước 2: Chính khóa dòng và tên dòng của thuộc tính *Game_Start_Date*
 - Vì thuộc tính *Game_Year* là thuộc tính cấp nhỏ nhất sẽ lấy khóa dòng gồm chính nó và những thuộc tính cấp cao hơn.
 - Để chỉnh sửa khoá dòng, click chuột phải vào thuộc tính *Game_Start_Date* => **Properties** => **Key Columns**.

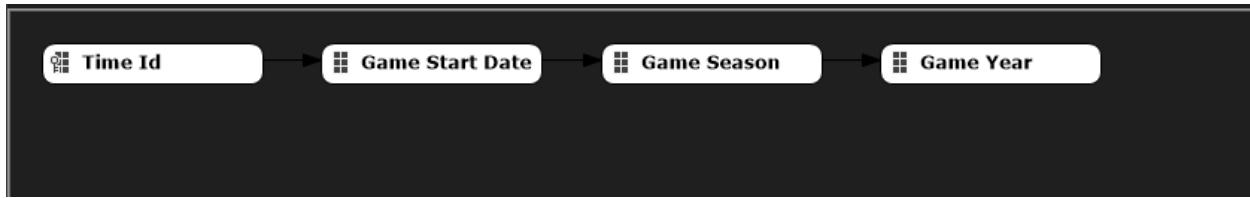
Game Start Date DimensionAttribute	
Source	
CustomRollupColumn	(none)
CustomRollupPropertiesCol	(none)
KeyColumns	Dim_Time.game_start_date
NameColumn	(none)
ValueColumn	(none)



- + Để chỉnh sửa tên dòng, click chuột phải vào thuộc tính *Game_Start_Date => Properties => Name Column.*



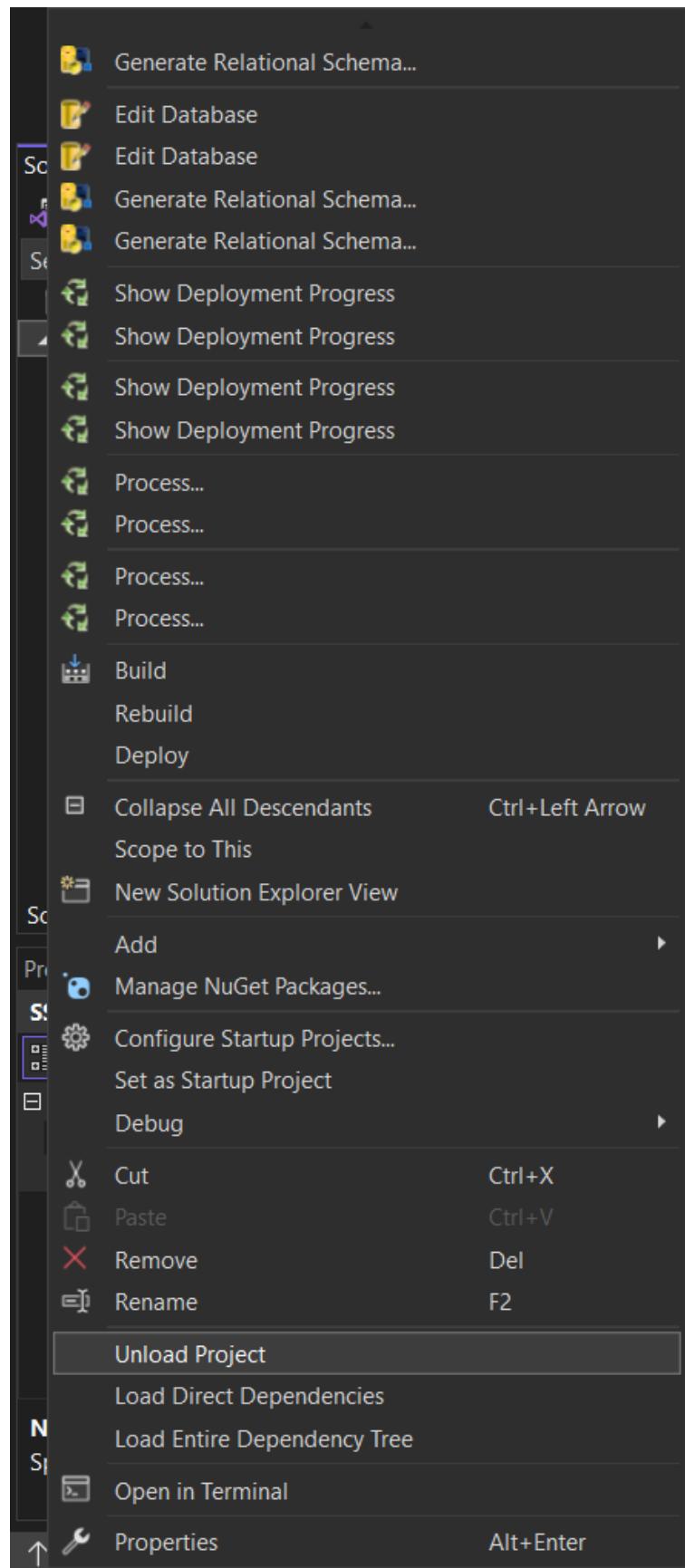
- Bước 3: Chuyển đến cửa sổ **Attribute Relationships** để liên kết các mối quan hệ phân cấp và tiến hành phân cấp:



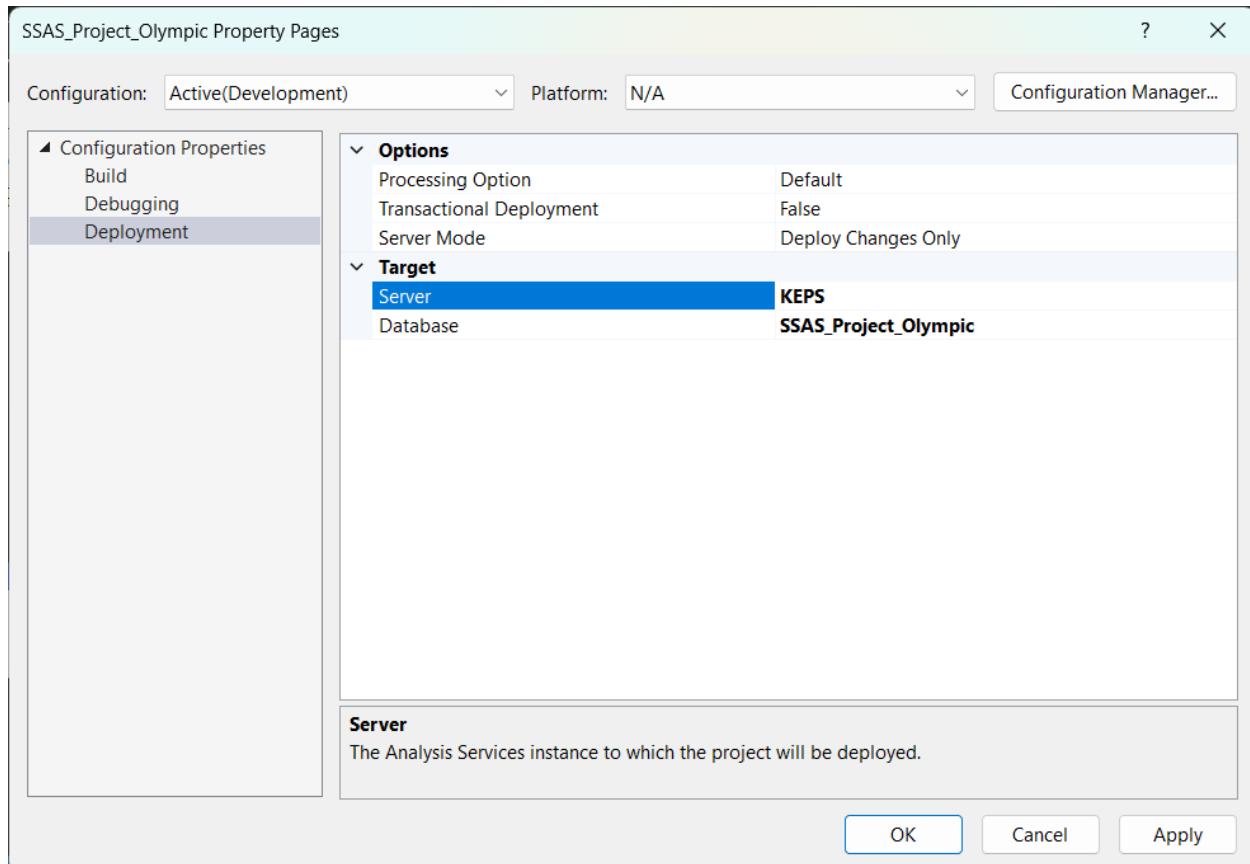
7. KHỞI CHẠY DỰ ÁN SSAS

7.1. Đổi tên máy chủ (server name)

- Bước 1: Nhấn chuột phải vào tên dự án *SSAS_Project_Olympic* → **Properties**

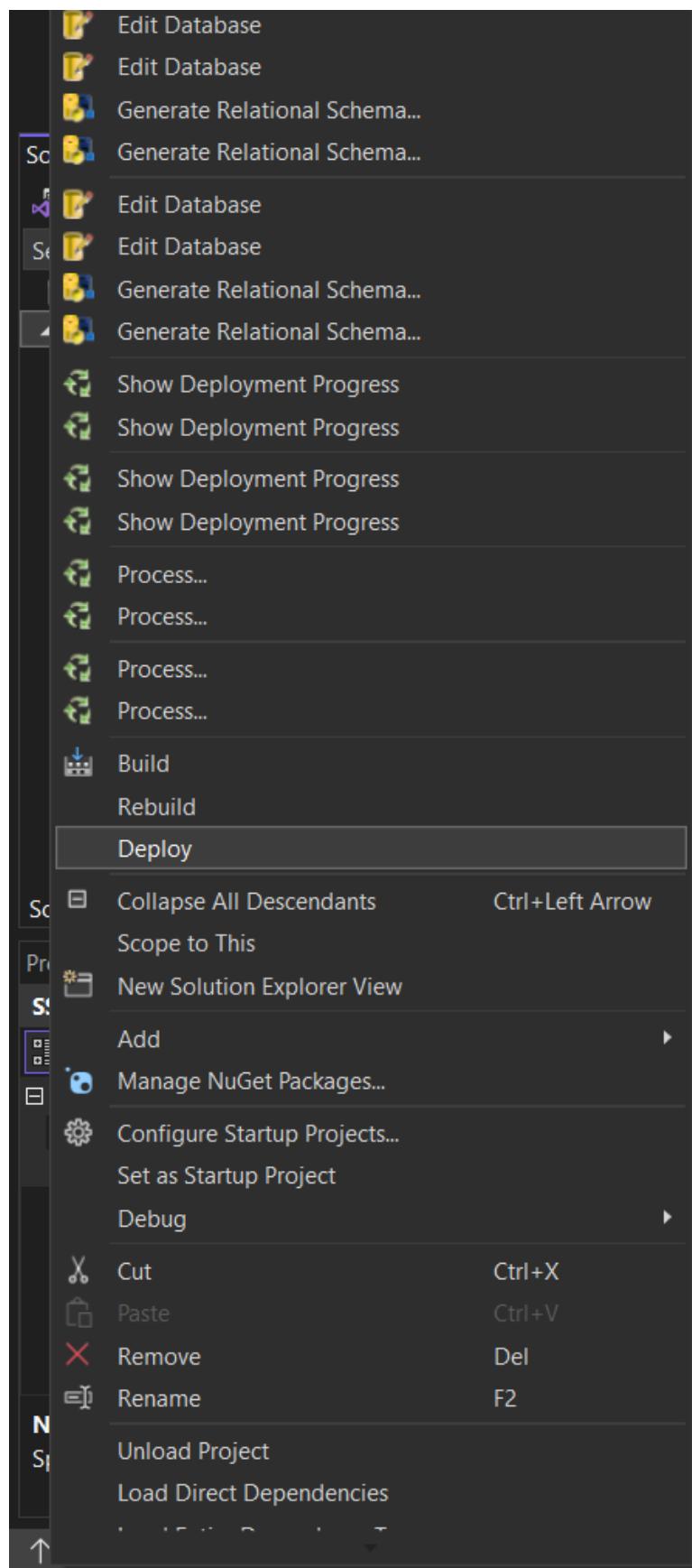


- Bước 2: Chọn tab Deployment → Đổi tên server từ localhost thành tên server

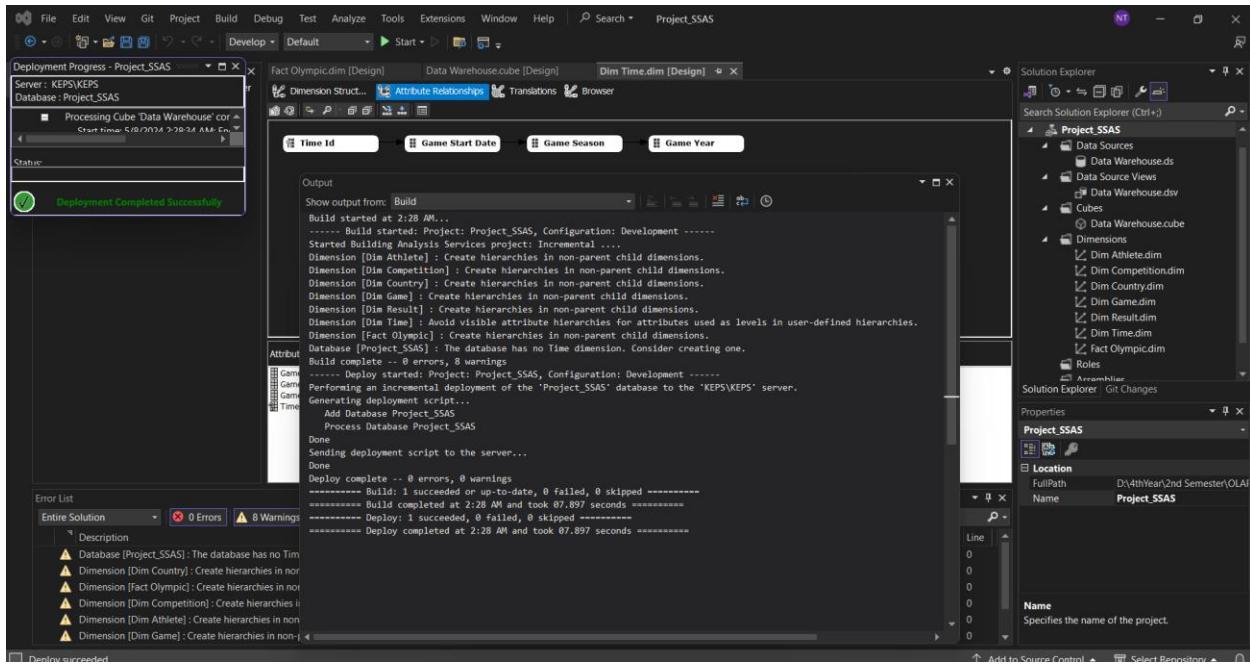


7.2. Chạy dự án SSAS

Nhấn chuột phải vào Project *SSAS_Project_Olympic* → Deploy



Kết quả:



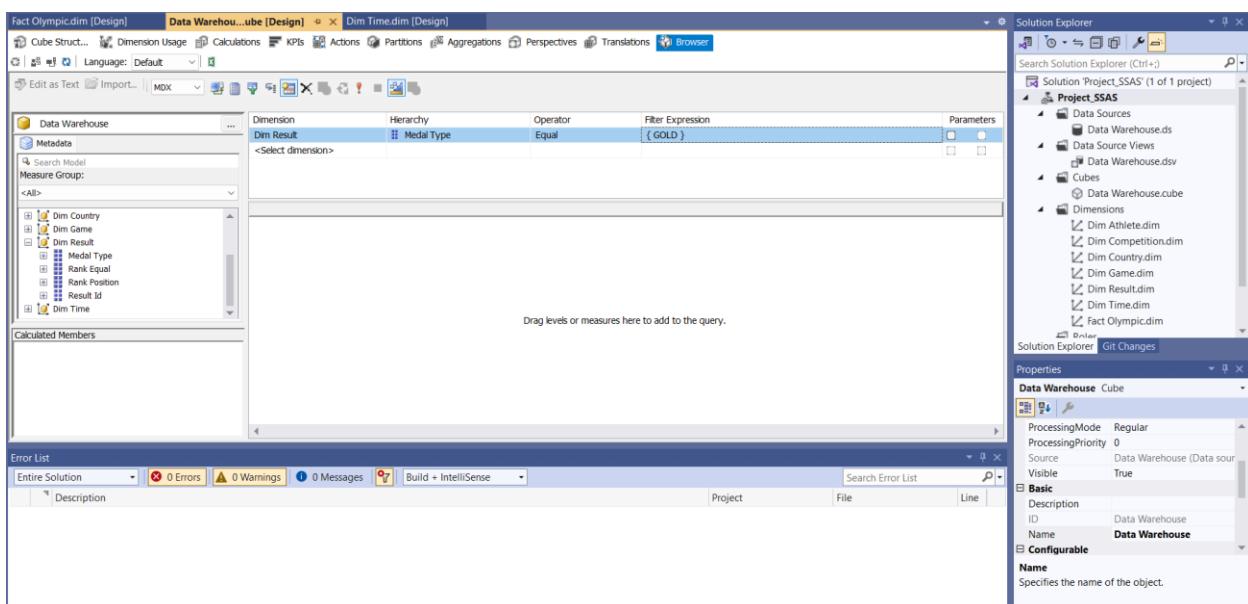
8. Quá trình phân tích dữ liệu bằng công cụ SSAS trên các khối cube

8.1. Câu truy vấn 1

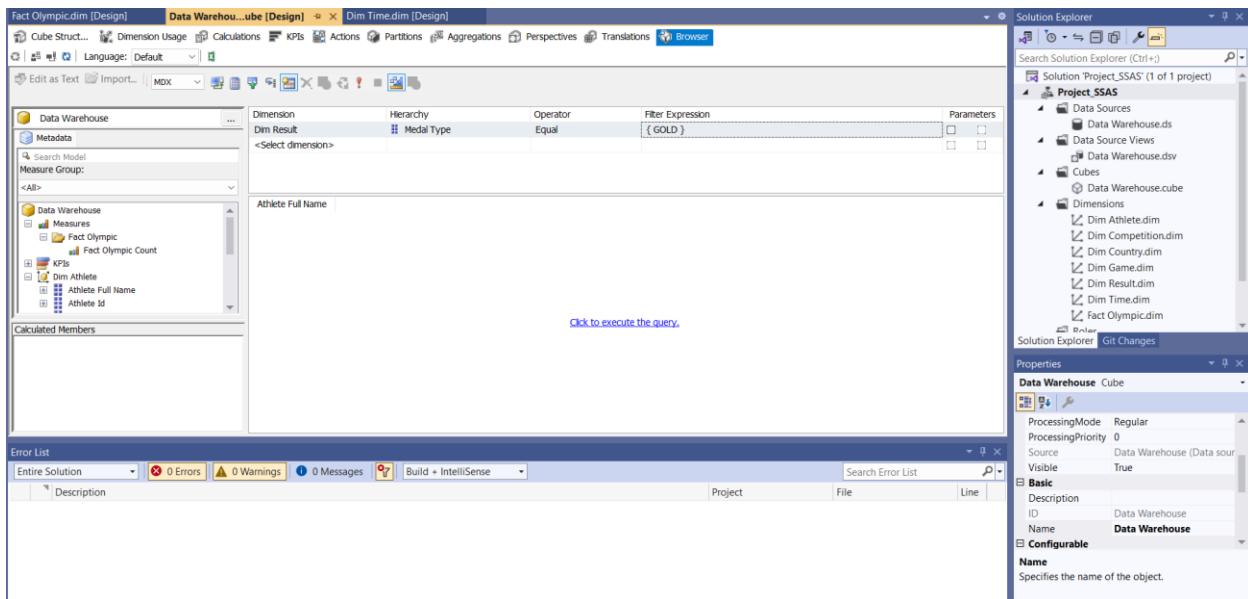
Nội dung: Liệt kê top 10 vận động viên đạt huy chương vàng nhất, sắp xếp theo thứ tự giảm dần

Thực hiện:

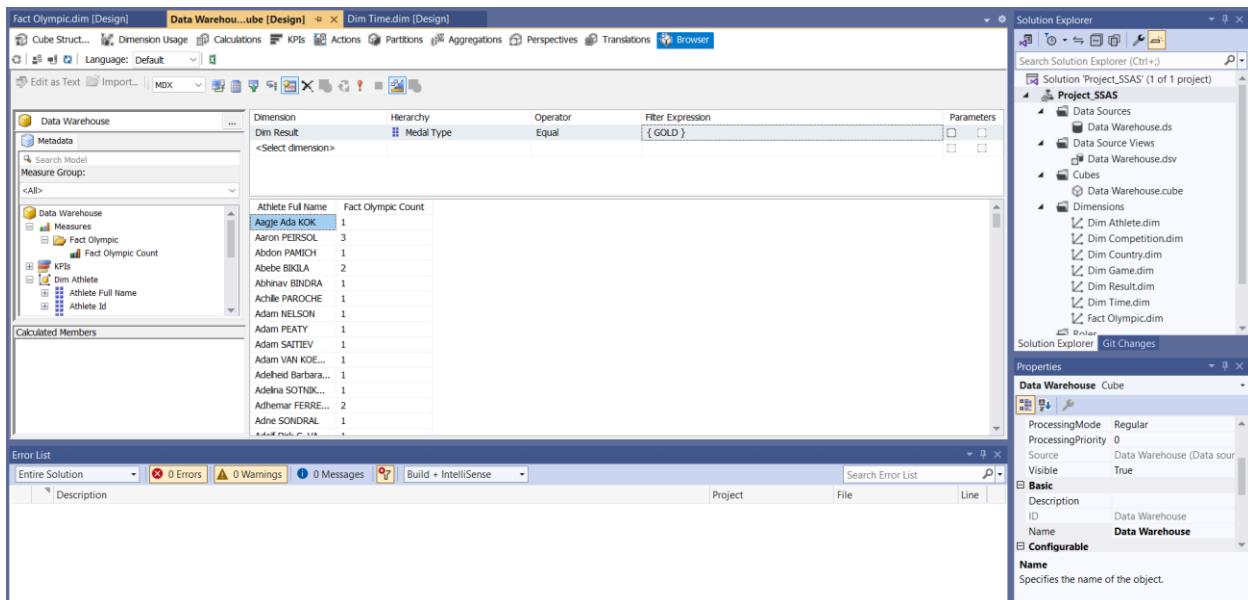
- Bước 1: Kéo thuộc tính Medal Type của bảng Dim Result sang cửa sổ lọc dữ liệu. Tại cột Filter Expression chọn GOLD.



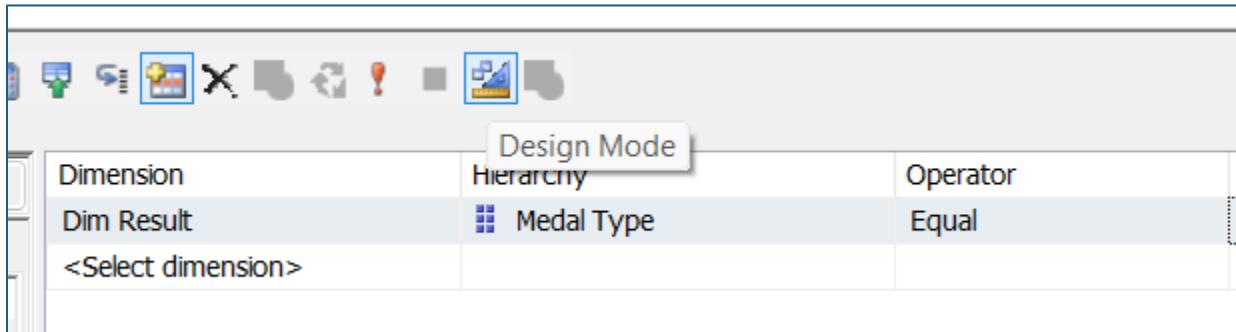
- Bước 2: Kéo thuộc tính Athlete Full Name của bảng Dim Athlete sang số thực thi câu truy vấn



- Bước 3: Kéo độ đo Fact Olympics Count của bảng Fact Olympics sang cửa sổ thực thi truy vấn



- Bước 4: Chuyển từ Design Mode sang Script Mode bằng cách nhấp chuột vào biểu tượng Design Mode như hình minh họa:



- Bước 5: Xóa bỏ dòng tô xanh như hình minh họa:

```

SELECT NON EMPTY { [Measures].[Fact Olympic Count] } ON COLUMNS, NON EMPTY { ([Dim Athlete].[Athlete Full Name].[Athlete Full Name].ALLMEMBERS) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM ( SELECT ( { [Dim Result].[Medal Type].&[GOLD] } ) ON COLUMNS FROM [Data Warehouse] ) WHERE ( [Dim Result].[Medal Type].&[GOLD] ) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
  
```

- Bước 6: Dán đoạn code sau vào dòng màu xanh vừa xóa:

TOPCOUNT([Dim Athlete].[Athlete Full Name].children**,10,[Measures].[Fact Olympic Count])**

- Bước 7: Nhấn Execute Query tại cửa sổ thực thi truy vấn ta được kết quả như hình minh họa :

```

SELECT NON EMPTY { [Measures].[Fact Olympic Count] } ON COLUMNS, NON EMPTY { TOPCOUNT([Dim Athlete].[Athlete Full Name].children,10,[Measures].[Fact Olympic Count]) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM ( SELECT ( { [Dim Result].[Medal Type].&[GOLD] } ) ON COLUMNS FROM [Data Warehouse] ) WHERE ( [Dim Result].[Medal Type].&[GOLD] ) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
  
```

Athlete Full Name	Fact Olympic Count
Michael PHELPS	13
Ray EWRY	8
Carl LEWIS	7
Björn DAEHLIE	6
Boris SHAKHIN	6
Larisa LATYNINA	6
Lidiya SKOBLIKO...	6
Nikolay ANDRIA...	6
Paavo NURMI	6
Usain BOLT	6

8.2. Câu truy vấn 2

Nội dung: Liệt kê tất cả các vận động viên tham gia Olympic từ 3 lần trở lên, sắp theo thứ tự tăng dần

Thực hiện:

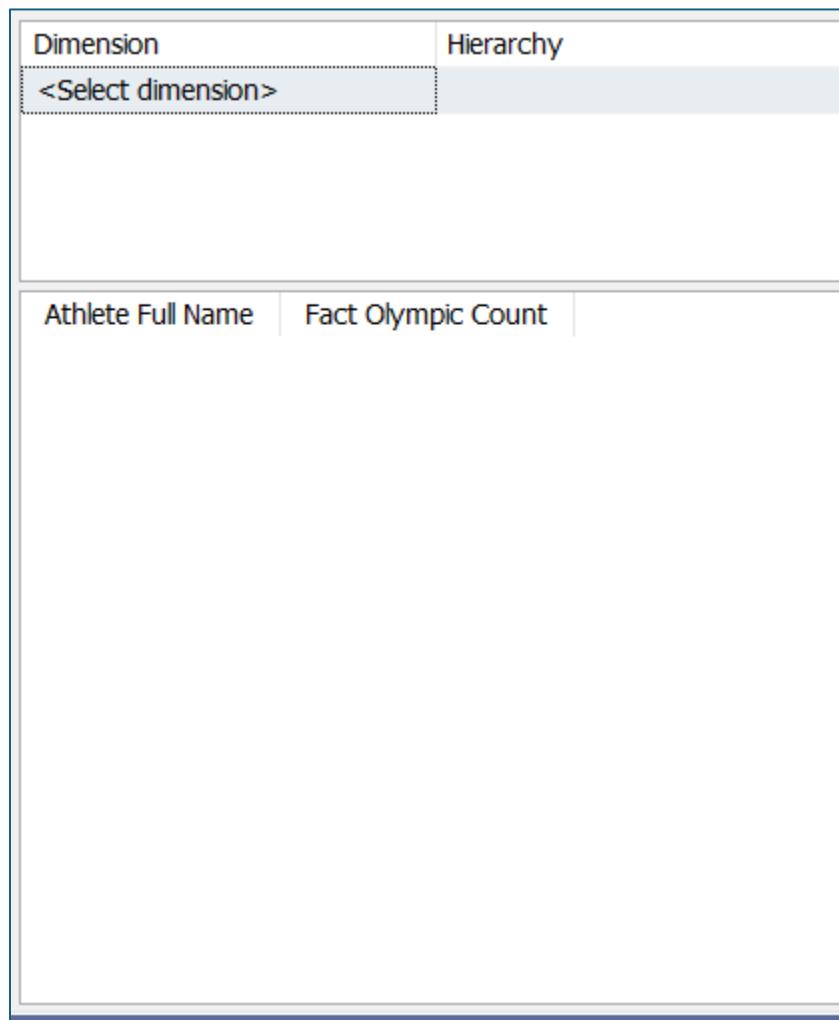
- Bước 1: Tạo một nameset [VDV THAM GIA TREN 3 OLYMPICS] dùng để lọc các động viên tham gia Olympics từ 3 lần trở lên.

The screenshot shows the Script Organizer interface. On the left, there's a sidebar titled "Script Organizer" with sections for "Command", "CALCULATE", and "[VDV THAM GIA TREN 3 OLYMPIC]" which is currently selected. Below this is the "Calculation Tools" section with tabs for "Metadata", "Functions", "Templates", and "Search Model". Under "Measure Group", it shows "<All>". On the right, the main area has a "Name:" field containing "[VDV THAM GIA TREN 3 OLYMPIC]". Below it is an "Expression" section with a code editor containing the following DAX:

```
FILTER {[Dim Athlete].[Athlete Full Name].[Athlete Full Name].MEMBERS, [Measures].[Fact Olympic Count] >=3}
```

The code editor shows a green circular icon with a checkmark and the text "No issues found". At the bottom right of the editor are buttons for "Ln: 1", "Ch: 107", "SPC", and "CRLF". A status bar at the bottom right indicates "Type: Dynamic".

- Bước 2: Kéo thuộc tính Athlete Full Name và độ đo Fact Olympics Count sang cửa sổ thực thi truy vấn.



- Bước 3: Dùng name set [VDV THAM GIA TREN 3 OLYMPICS], ta được kết quả:

The screenshot shows the MDX Editor interface with the following details:

- MDX Bar:** Edit as Text, Import..., MDX, etc.
- Query Area:**

```
SELECT
    NON EMPTY
        {[Measures].[Fact Olympic Count]} ON COLUMNS
    FROM
        [Dim Athlete]
    WHERE
        [Athlete Full Name] IN
            {<Select dimension>}

```
- Result Area:**

Athlete Full Name	Fact Olympic Count
Aage Avallorff	3
Aagje Ada KOK	3
Arne ROINE	8
Arne SALOVAARA	4
Aaron BLUNCK	3
Aaron MARCH	4
Aaron PEIRSON	5
Atos Tuomas J...	3
Abbas ZANDI	3
Abbey WETZEL	4
Abby LARSON	3
Abdalaati IGUIDER	4
Abdehafid BENCI...	4
Abdelkader CHADI	3
- Left Sidebar:**
 - Data Warehouse
 - Metadata
 - Search Model
 - Measure Group: <All>
 - Data Warehouse
 - Measures
 - Fact Olympic
 - Fact Olympic Count
 - KPIs
 - Dim Athlete
 - VDV THAM GIA TREN 3 OLYMPIC
 - Athlete Full Name
- Calculated Members:** (Empty)

8.3. Câu truy vấn 3

Nội dung: Liệt kê 10 môn thi đấu có số lượng huy chương đồng nhiều nhất trong kỳ thi vận hội Tokyo 1964.

Thực hiện:

- Bước 1: Tạo một nameset [TOP 10 MON THI DAU] dùng để lọc các môn thi đấu có nhiều huy chương đồng nhất

Script Organizer

Name: [TOP 10 MON THI DAU]

Expression

```
TAIL(ORDER([Dim Competition].[Discipline Title].CHILDREN,[Measures].[Fact Olympic Count], ASC),10)
```

No issues found

Additional Properties

Type: Dynamic

Display folder:

Calculation Tools

Measure Group: <All>

Data Warehouse

- Measures
 - Fact Olympic
 - Fact Olympic Count
- Dim Athlete
- Dim Competition

- Bước 2: Kéo thuộc tính Medal Type của bảng Dim Result sang cửa sổ lọc dữ liệu và lọc lấy các dòng dữ liệu có thuộc tính Medal Type là BRONZE.

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Result	Medal Type	Equal	{ BRONZE }	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				

Drag levels or measures here to add to the query.

Data Warehouse

Metadata

Search Model

Measure Group: <All>

Participant Type

Dim Country

Dim Game

Dim Result

Medal Type

Rank Equal

Rank Position

Result Id

Calculated Members

- Bước 3: Kéo thuộc tính Game Name của bảng Dim Game sang cửa sổ lọc dữ liệu và lọc lấy các dòng dữ liệu có thuộc tính Game Name là Tokyo 1964.

The screenshot shows a Data Warehouse interface. On the left, there's a navigation pane with 'Data Warehouse' and 'Metadata' sections, and a search bar. Below that is a tree view of measures, KPIs, and dimensions. A 'Calculated Members' section is also present. On the right, there's a query builder with four rows of filters:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Game	Game Name	Equal	{ Tokyo 1964 }	<input type="checkbox"/> <input checked="" type="checkbox"/>
Dim Result	Medal Type	Equal	{ BRONZE }	<input type="checkbox"/> <input checked="" type="checkbox"/>
Dim Competition	Discipline Title	In	TOP 10 MON THI DAU	<input type="checkbox"/> <input checked="" type="checkbox"/>

Below the filters is a table titled 'Fact Olympic Count' with one row: 'Click to execute the query.'

- Bước 4: Kéo name set TOP 10 MON THI DAU sang cửa sổ lọc dữ liệu, thuộc tính Discipline Title và độ đo Fact Olympics Count sang cửa sổ thực thi truy vấn, ta được kết quả.

The screenshot shows the same Data Warehouse interface after executing the query. The 'Fact Olympic Count' table now contains data:

Discipline Title	Fact Olympic Count
Athletics	31
Cycling Track	3
Diving	4
Fencing	4
Gymnastics ...	9
Judo	2
Shooting	5
Swimming	13
Weightlifting	7
Wrestling	14

8.4. Câu truy vấn 4

Nội dung: Vận động viên France dành được nhiều huy chương bạc nhất

Thực hiện:

- Bước 1: Tạo một nameset TOP 1 VDV dùng để lấy ra vận động viên giành nhiều huy chương nhất

Script Organizer

Name: [TOP 1 VDV]

Expression:

```
TAIL(ORDER(([Dim Athlete].[Athlete Full Name].CHILDREN,[Measures].[Fact Olympic Count],ASC),1))
```

No issues found

Type: Dynamic

Display folder:

Calculation Tools

Data Warehouse

- Metadata
- Functions
- Templates
- Search Model

Measure Group: <All>

Dim Athlete

Dim Competition

Dim Country

Dim Game

- Bước 2: Kéo thuộc tính Medal Type của bảng Dim Result sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn SILVER

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Result	Medal Type	Equal	{ SILVER }	
<Select dimension>				

Drag levels or measures here to add to the query.

- Bước 3: Kéo thuộc tính Country Name của bảng Dim Country sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn France.

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Result	Medal Type	Equal	{ SILVER }	
Dim Country	Country Name	Equal	{ France }	
<Select dimension>				

- Bước 4: Kéo name set TOP 1 VDV sang cửa sổ lọc dữ liệu, thuộc tính Athlette Full Name và độ đo Fact Olympics Count sang cửa sổ thực thi truy vấn, ta được kết quả.

Athlete Full Name	Fact Olympic Count
Alain MIMOUN	3

8.5. Câu truy vấn 5

Nội dung: Quốc gia dành được nhiều huy vàng nhất từ năm 1980 đến năm 2000

Thực hiện:

- Bước 1: Tạo một nameset TOP 1 QUOC GIA dùng để lấy ra quốc gia giành nhiều huy chương nhất

- Bước 2: Kéo thuộc tính Medal Type của bảng Dim Result sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn GOLD.

The screenshot shows the Data Warehouse Metadata interface. On the left, there's a tree view of dimensions: Dim Country, Dim Game, Dim Result, Dim Time, and others. The 'Dim Result' node is expanded. In the center, there's a query builder window with the following configuration:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Result	Medal Type	Equal	{ GOLD }	<input type="checkbox"/> <input type="checkbox"/>

A message at the bottom says "Drag levels or measures here to add to the query."

- Bước 3: Kéo thuộc tính Game Year của bảng Dim Time sang cửa sổ lọc dữ liệu, tại cột Operator chọn Range (Inclusive) và cho giá trị trong khoảng 1980 đến 2000.

The screenshot shows the Data Warehouse Metadata interface. The 'Dim Time' node is selected in the tree view. In the query builder window, the configuration is as follows:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Result	Medal Type	Equal	{ GOLD }	<input type="checkbox"/> <input type="checkbox"/>
Dim Time	Game Year	Range (Exclusive)	1980 : 2000	<input type="checkbox"/> <input type="checkbox"/>

A message at the bottom says "Drag levels or measures here to add to the query."

- Bước 4: Kéo name set TOP 1 QUOC GIA sang cửa sổ lọc dữ liệu, thuộc tính Country Name và độ đo Fact Olympics Count sang cửa sổ thực thi truy vấn, ta được kết quả

The screenshot shows the Data Warehouse interface with the following details:

- Dimension:** Dim Country
- Hierarchy:** Country Name
- Operator:** In
- Filter Expression:** TOP 1 QUOC GIA
- Parameters:** (checkboxes)
- Result:** United States ... 150

The left sidebar shows the Data Warehouse structure with measures like Fact Olympic Count and Fact Athlete.

8.6. Câu truy vấn 6

Nội dung: Thống kê số lượng vận động viên của các quốc gia tham gia thi đấu môn Ski Jumping trong thế vận hội Sochi 2014.

Thực hiện:

- Bước 1: Kéo thuộc tính Discipline Title của bảng Dim Competition sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn Ski Jumping

The screenshot shows the Data Warehouse interface with the following details:

- Dimension:** Dim Competition
- Hierarchy:** Discipline Title
- Operator:** Equal
- Filter Expression:** {SkiJumping}
- Parameters:** (checkboxes)
- Result:** Drag levels or measures here to add to the query.

The left sidebar shows the Data Warehouse structure with measures like Competition Id, Discipline Title, Event Title, Participant Type, and Dim Country.

- Bước 2: Kéo thuộc tính Game Name của bảng Dim Game sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn Sochi 2014

- Bước 3: Kéo tính Athlette Full Name và độ đo Fact Olympics Count sang cửa sổ thực thi truy vấn, ta được kết quả:

Country Name	Fact Olympic Count
Austria	10
Bulgaria	2
Canada	10
Czech Republic	8
Estonia	4
Finland	9
France	5
Germany	11
Italy	8
Japan	11
Kazakhstan	4
Norway	12
Poland	8
Republic of Korea	6
Denmark	1

8.7. Câu truy vấn 7

Nội dung: Quốc gia đạt số lượng huy chương bạc nhiều nhất thế vận hội London 2012

Thực hiện:

- Bước 1: Kéo thuộc tính Medal Type của bảng Dim Result sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn SILVER.

The screenshot shows the Data Warehouse interface with a query setup. In the 'Filter Expression' column for the 'Dim Result' dimension, the value is set to '{ SILVER }'. The 'Calculated Members' section is currently empty.

- Bước 2: Kéo thuộc tính Game Name của bảng Dim Game sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn London 2012.

The screenshot shows the Data Warehouse interface with the query updated. The 'Filter Expression' column for the 'Dim Game' dimension is now set to '{ London 2012 }'.

- Bước 3: Kéo name set TOP 1 QUOC GIA sang cửa sổ lọc dữ liệu, thuộc tính Country Name và độ đo Fact Olympics Count sang cửa sổ thực thi truy vấn, ta được kết quả

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Result	Medal Type	Equal	{ SILVER }	<input type="checkbox"/> <input type="checkbox"/>
Dim Game	Game Name	Equal	{ London 2012 }	<input type="checkbox"/> <input type="checkbox"/>
Dim Country	Country Name	In	TOP 1 QUOC GIA	<input type="checkbox"/> <input type="checkbox"/>

Country Name | Fact Olympic Count
People's Republic of China | 25

8.8. Câu truy vấn 8

Nội dung: Liệt kê top 5 nội dung thi có số lượng vận động viên tham gia nhiều nhất trong kỳ thi vận hội Tokyo 2020.

Thực hiện:

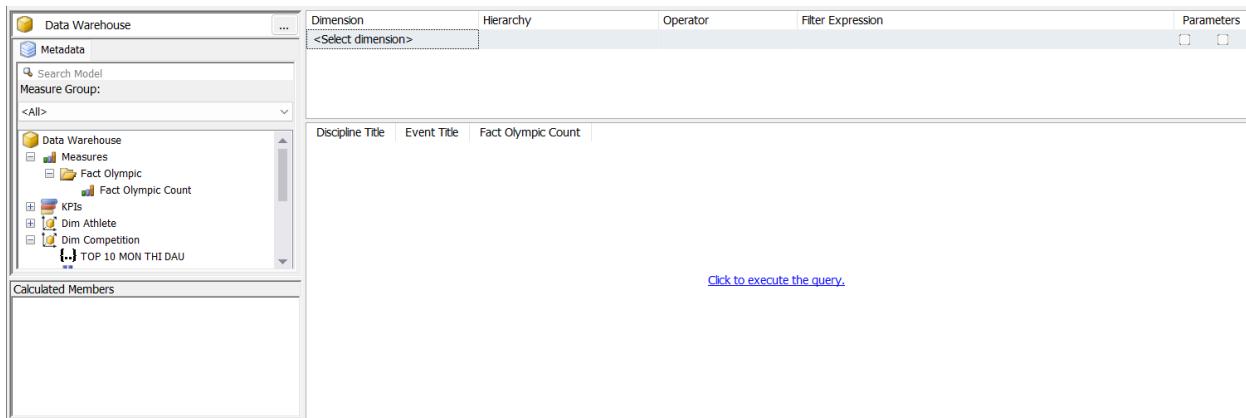
- Bước 1: Tạo name set TOP 5 NOI DUNG THI

Name: [TOP 5 NOI DUNG THI]

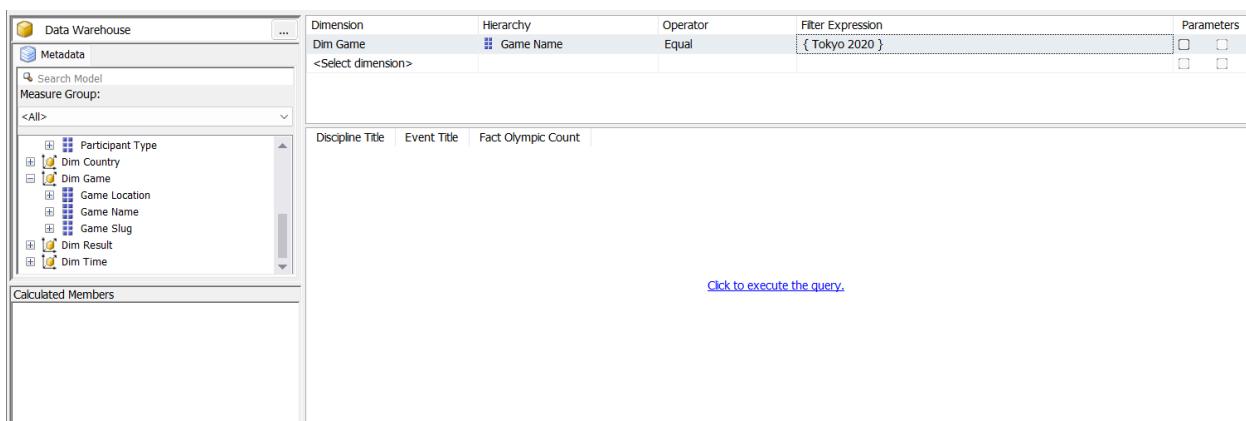
Expression: TOPCOUNT([Dim Competition].[Discipline Title].CHILDREN*[Dim Competition].[Event Title].CHILDREN,5,[Measures].[Fact Olympic Count])
No issues found

Additional Properties: Type: Dynamic

- Bước 2: Deploy và lưu nameset lại qua tab Brower để tiến hành thực hiện câu truy vấn, kéo thả những thuộc tính Discipline Title, Event Title Alias và độ đo Fact Olympics Count vào vùng truy vấn như hình minh họa.



- Bước 3: Kéo thuộc tính Game Name của bảng Dim Game sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn Tokyo 2020.



- Bước 4: Kéo thả một name set bất kì vào trường lọc để làm name set thay thế cho [TOP 5 NOI DUNG THI] vì name set này chứa nhiều thuộc tính nên không xác định được của bảng DIM nào nên không hiển thị phải dùng name set thay thế.

The screenshot shows the SSAS Data Model Designer interface. On the left, there's a tree view of the Data Warehouse, Metadata, and Measure Group sections. The main pane is titled 'Dimension Filter' and contains two rows of filters:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Game	Game Name	Equal	{ Tokyo 2020 }	<input type="checkbox"/> <input checked="" type="checkbox"/>
Dim Athlete	Athlete Full Name	In	TOP 10 VDV	<input type="checkbox"/> <input checked="" type="checkbox"/>

Below the filters, there's a section labeled '<Select dimension>' and a button 'Click to execute the query.'

- Bước 5: Tắt chế độ Design Mode để chuyển sang chế độ MDX Mode. Xác định vị trí name set thay thế (tô màu xanh). Tiến hành thay thế name set.

The screenshot shows the SSAS Data Model Designer interface. The main pane displays a complex MDX query:

```

SELECT NON EMPTY { [Measures].[Fact Olympic Count] } ON COLUMNS, NON EMPTY { {[Dim Competition].[Discipline Title].[Discipline Title].ALLMEMBERS * [Dim Competition].[Event Title].[Event Title].ALLMEMBERS ) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM ( SELECT ( { [TOP 5 NOI DUNG THI] } ) ON COLUMNS FROM ( SELECT ( { [Dim Game].[Game Name]&[Tokyo 2020] } ) ) ON COLUMNS FROM [Data Warehouse]) WHERE ( [Dim Game].[Game Name]&[Tokyo 2020] ) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
  
```

The pane includes a tree view of dimensions like KPIs, Dim Athlete, Dim Competition, and Dim Country, and a button 'Click to execute the query.'

- Bước 6: Chính sửa name set [TOP 10 VDV] thành name set [TOP 5 NOI DUNG THI] và click to execute the query ta được kết quả như hình minh họa:

The screenshot shows the SSAS Data Warehouse interface. On the left, there's a navigation pane with 'Data Warehouse' and 'Metadata' sections. Under 'Measure Group', it lists 'All'. In the center, there's a large text area with a complex SQL query:

```
SELECT NON EMPTY { [Measures].[Fact Olympic Count] } ON COLUMNS, NON EMPTY { ([Dim Competition].[Discipline Title].[Discipline Title].ALLMEMBERS * [Dim Competition].[Event Title].[Event Title].ALLMEMBERS ) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM ( SELECT ( {[TOP 5 NỘI DUNG THI]} ) ON COLUMNS FROM ( SELECT ( {[Dim Game].[Game Name].&[Tokyo 2020]} ) ON COLUMNS FROM [Data Warehouse$] ) WHERE ([Dim Game].[Game Name].&[Tokyo 2020]) ) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

Below the query is a results grid titled 'Fact Olympic Count' with columns 'Discipline Title', 'Event Title', and 'Fact Olympic Count'. The data is as follows:

Discipline Title	Event Title	Fact Olympic Count
Athletics	Men's Mar...	103
Athletics	Women's...	85
Cycling Road	Men's Ro...	125
Equestrian	Eventing ...	69
Equestrian	Jumping I...	87

8.9. Câu truy vấn 9

Nội dung: Thống kê số lượng vận động viên tham gia các nội dung thi đấu qua các kỳ Olympics đến nay của đội tuyển Việt Nam.

Thực hiện:

- Bước 1: Kéo thuộc tính Country Name của bảng Dim Country sang cửa sổ lọc dữ liệu, tại cột Filter Expression chọn Vietnam.

The screenshot shows the SSAS Data Warehouse interface. On the left, there's a navigation pane with 'Data Warehouse' and 'Metadata' sections. Under 'Measure Group', it lists 'All'. In the center, there's a 'Filter Expression' configuration window. It shows a table with columns 'Dimension', 'Hierarchy', 'Operator', 'Filter Expression', and 'Parameters'. One row is selected for 'Dim Country' with 'Country Name' as the hierarchy, 'Equal' as the operator, and 'Vietnam' as the filter expression. Below this table is a message: 'Drag levels or measures here to add to the query.'

- Bước 2: Kéo thuộc tính Game Name, Discipline Title, Event Title Alias và độ đo Fact Olympics Count sang cửa sổ thực thi truy vấn, ta được kết quả

Game Name	Athlete Full Name	Discipline Title	Fact Olympic Count
Athens 2004	Huu Viet NGUYEN	Swimming	1
Athens 2004	Kien Quoc DOAN	Table Tennis	1
Athens 2004	Manh Tuong NG...	Shooting	1
Athens 2004	Quoc Huan NGU...	Taekwondo	1
Athens 2004	Thi Thiet NGUYEN	Weightlifting	1
Athens 2004	Van Hung NGUY...	Taekwondo	1
Atlanta 1996	Phoung Trinh C...	Judo	1
Barcelona 1...	Quoc Cuong NG...	Shooting	1
Beijing 2008	Anh Tuan HOANG	Weightlifting	1
Beijing 2008	Huu Viet NGUYEN	Swimming	1
Beijing 2008	Kien Quoc DOAN	Table Tennis	1
Beijing 2008	Manh Tuong NG...	Shooting	2
Beijing 2008	Thi Hoai Thi NG...	Taekwondo	1
Beijing 2008	Thi Ngan Thuon...	Gymnastics ...	5
Beijing 2008	Thi Nguen Tran T...	Taekwondo	1

8.10. Câu truy vấn 10

Nội dung: Liệt kê số lượng huy chương của Pháp theo từng kỳ Olympics.

Thực hiện:

- Bước 1: Kéo thuộc tính Game name và độ đo Fact Olympic Count vào cửa sổ truy vấn

Game Name	Fact Olympic Count
Click to execute the query.	

- Bước 2: Kéo thuộc tính Country name vào cửa sổ lọc dữ liệu, Filter “France”.

Thực thi truy vấn

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Country	Country Name	Equal	{ France }	
<Select dimension>				
	Game Name	Fact Olympic Count		
	Mexico City 1968	95		
	Montreal 1976	90		
	Moscow 1980	60		
	Munich 1972	86		
	Nagano 1998	106		
	Oslo 1952	33		
	Paris 1900	196		
	Paris 1924	174		
	PyeongChang 2018	149		
	Rio 2016	195		
	Rome 1960	92		
	Salt Lake City 2002	113		
	Sapporo 1972	53		
	Sarajevo 1984	46		

8.11. Câu truy vấn 11

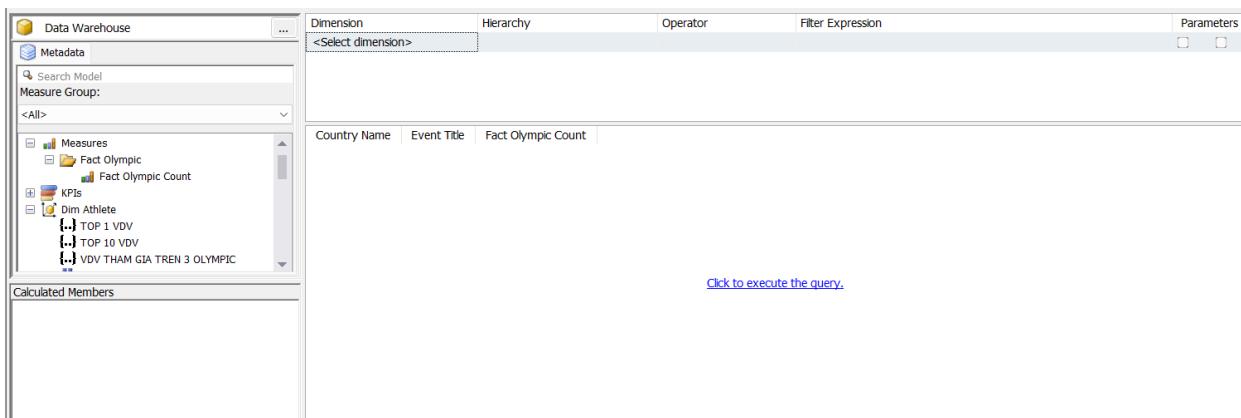
Nội dung: Với mỗi quốc gia đưa ra 3 môn thi đấu có nhiều vận động viên tham gia nhiều nhất từ trước đến nay

Thực hiện:

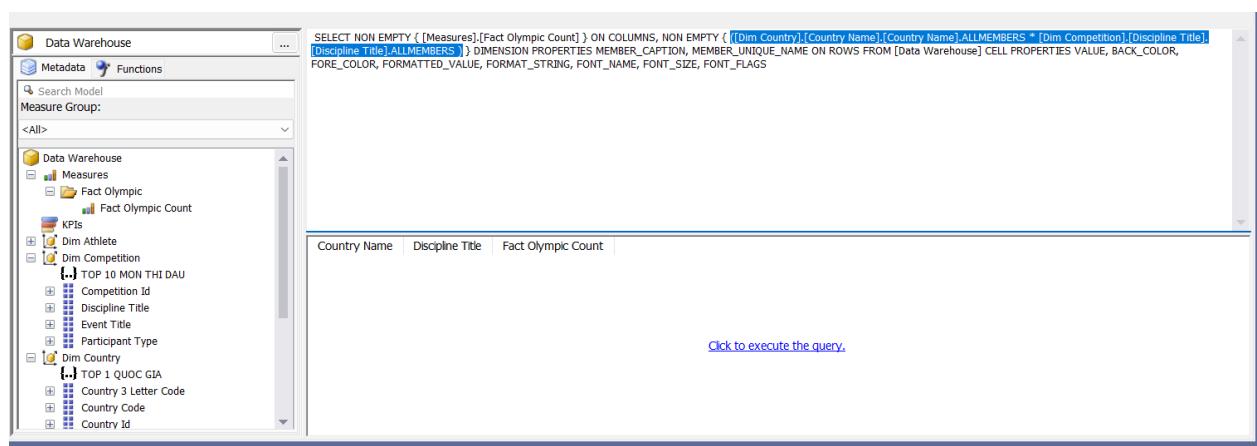
- Bước 1: Dùng hàm GENERATE và TOPCOUNT

```
Generate([Dim Country].[Country Name].Children,
TopCount( {[Dim Country].[Country Name].CurrentMember * [Dim Competition].[Discipline Title].Children},3,[Measures].[Fact Olympic Count]))
```

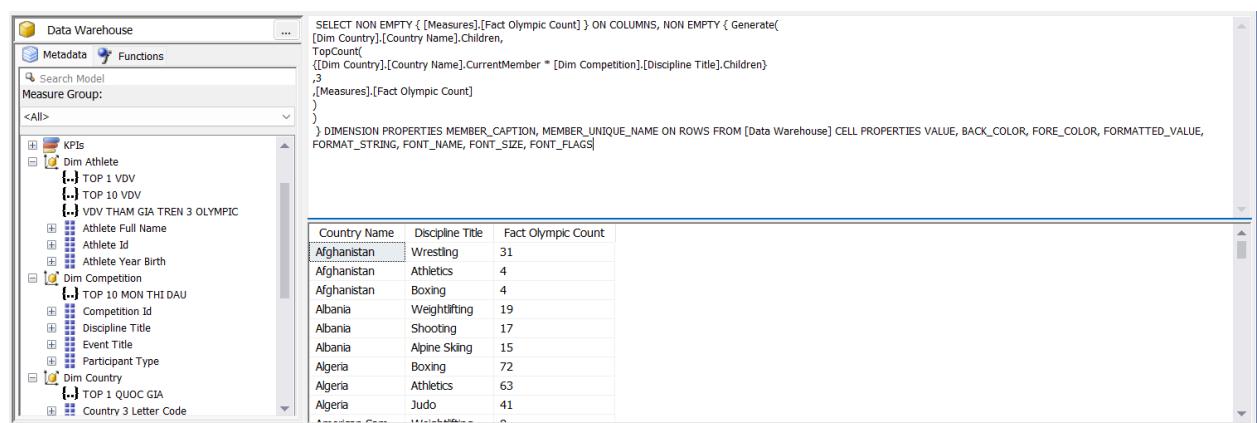
- Bước 2: Kéo thuộc tính qua cửa sổ lọc dữ liệu của công cụ BI kéo thả các thuộc tính cần thiết Country Name, Discipline Title và độ đo Fact Olympics Count câu truy vấn như hình sau đây:



- Bước 3: Chuyển từ Design Mode sang Script Mode bằng cách nhấn chuột vào biểu tượng Design Mode. Xóa bỏ dòng tô xanh như hình minh họa



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



- Bước 5: Thực hiện truy vấn

The screenshot shows a Data Warehouse interface. On the left, there's a navigation pane with sections like 'Data Warehouse', 'Metadata', 'Functions', 'Search Model', 'Measure Group', and 'All'. Under 'Dim Athlete', there are items like 'TOP 1 VDV', 'TOP 10 VDV', and 'VDV THAM GIA TREN 3 OLYMPIC'. Under 'Dim Competition', there are items like 'TOP 10 MON THI DAU', 'Competition Id', 'Discipline Title', 'Event Title', and 'Participant Type'. Under 'Dim Country', there are items like 'TOP 1 QUOC GIA' and 'Country 3 Letter Code'. The main area displays a query result grid with three columns: 'Country Name', 'Discipline Title', and 'Fact Olympic Count'. The data in the grid is as follows:

Country Name	Discipline Title	Fact Olympic Count
Afghanistan	Wrestling	31
Afghanistan	Athletics	4
Afghanistan	Boxing	4
Albania	Weightlifting	19
Albania	Shooting	17
Albania	Alpine Skiing	15
Algeria	Boxing	72
Algeria	Athletics	63
Algeria	Judo	41

The query editor pane at the top contains the following SQL-like code:

```

SELECT NON EMPTY { [Measures].[Fact Olympic Count] } ON COLUMNS, NON EMPTY { [Dim Country].[Country Name].Children, TopCount([Dim Country].[Country Name], CurrentMember * [Dim Competition].[Discipline Title].Children), 3, [Measures].[Fact Olympic Count] }
) DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [Data Warehouse] CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS

```

8.12. Câu truy vấn 12

Nội dung: Các vận động viên bị trùng hạng tư của nước United States of America

Thực thi:

- Bước 1: Kéo thuộc tính Athlete Full Name và độ đo Fact Olympic sang cửa sổ thực thi truy vấn.

Athlete Full Name	Fact Olympic Count
John Smith	4

- Bước 2: Lọc các thuộc tính phân cấp theo hình

Click to execute the query.

- Bước 3: Thực thi câu lệnh

Athlete Full Name	Fact Olympic Count
Andrea MEAD ...	1
Anne Kring KUR...	1
Arthur SEWARD	1
Colin MORIKAWA	1
Egon R. ERICKS...	1
Frank Otto HAU...	1
Janice SMITH	1
Jeanne ASHWOO...	1
Jim THORPE	1
John Richard W...	1
Inzank I YODOM	1

8.13. Câu truy vấn 13

Nội dung: Tìm vận động viên giành được nhiều huy chương vàng nhất Olympic Beijing 2008

Thực thi:

- Bước 1: Kéo các thuộc tính vào cửa sổ truy vấn và các thuộc tính phân cấp như hình

Dimension Hierarchy Operator Filter Expression Parameters

Dim Result	Medal Type	Equal	{ GOLD }	<input type="checkbox"/> <input type="checkbox"/>
Dim Game	Game Name	Equal	{ Beijing 2008 }	<input type="checkbox"/> <input type="checkbox"/>
Dim Athlete	Athlete Full Name	In	TOP 1 VDV	<input type="checkbox"/> <input type="checkbox"/>

<Select dimension>

Athlete Full Name Fact Olympic Count

Click to execute the query.

- Bước 2: Thực thi truy vấn

Athlete Full Name	Fact Olympic Count
Michael PHELPS	5

8.14. Câu truy vấn 14

Nội dung: Kì Olympic với nhiều vận động viên tham gia nhất

Thực thi:

- Bước 1: Tạo name set “Top 1 Game”

- Bước 2: Kéo các thuộc tính Game name và độ đo Fact Olympic Count sang cửa sổ truy vấn

- Bước 3: Kéo name set Top 1 Game sang cửa sổ lọc dữ liệu

- Bước 4: Thực thi truy vấn

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Athlete	Athlete Full Name	Equal		
Dim Game	Game Name	In	TOP 1 GAME	
<Select dimension>				

Game Name	Fact Olympic Count
Rio 2016	6303

8.15. Câu truy vấn 15

Nội dung: Top 10 các môn thi có số lượng vận động viên tham gia nhiều nhất với địa điểm tổ chức Olympic nằm ở quốc gia Great Britain

Thực thi:

- Bước 1: sử dụng hàm TopCount để thực thi truy vấn

```
TopCount({[Dim Competition].[Discipline
Title].Children},10,[Measures].[Fact Olympic Count])
```

- Bước 2: kéo thuộc tính Game location vào cửa sổ lọc dữ liệu, Filter “Great Britain”

The screenshot shows the Analysis Services Management Studio interface. On the left, there's a tree view of the Data Warehouse, Metadata, and Measure Group sections. In the main area, there's a table for defining dimensions. The 'Dimension' column is set to 'Dim Game', 'Hierarchy' to 'Game Location', 'Operator' to 'Equal', and 'Filter Expression' to '{ Great Britain }'. Below this table is a section labeled 'Fact Olympic Count' with the text 'Click to execute the query.' At the bottom of the main area, there's a 'Calculated Members' section which is currently empty.

- Bước 3: kéo thuộc tính Discipline Title và độ đo Fact Olympic Count vào cửa sổ truy vấn

The screenshot shows the same interface as the previous one, but with a change. The 'Discipline Title' item has been selected from the Dimension list and is now highlighted in blue. The rest of the dimension definition remains the same: 'Dimension' is 'Dim Game', 'Hierarchy' is 'Game Location', 'Operator' is 'Equal', and 'Filter Expression' is '{ Great Britain }'. The 'Calculated Members' section is still empty. The 'Fact Olympic Count' section still contains the text 'Click to execute the query.'

- Bước 4: Chuyển sang cửa sổ Design mode và thay đổi đoạn code được tô xanh

The screenshot shows the Analysis Services Management Studio in Design mode. The top navigation bar includes 'Cube Struct...', 'Dimension Usage', 'Calculations', 'KPIs', 'Actions', 'Partitions', 'Aggregations', 'Perspectives', 'Translations', and 'Browser'. The main area has tabs for 'Edit as Text', 'Import...', 'MDX', and 'Browser'. The 'MDX' tab is active. The interface is similar to the previous ones, with a tree view on the left and a query editor on the right. The query editor contains the following MDX code:

```

SELECT NON EMPTY { [Measures].[Fact Olympic Count] } ON COLUMNS, NON EMPTY { {[Dim Competition].[Discipline Title].[Discipline Title].ALLMEMBERS} } ON ROWS FROM ( SELECT { {[Dim Game].[Game Location].&[Great Britain]} } ON COLUMNS FROM [Data Warehouse] WHERE {[Game Location].&[Great Britain]} ) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS

```

- Bước 5: Dán đoạn code ở bước 1, thực thi truy vấn

Discipline Title	Fact Olympic Count
Athletics	1953
Swimming	1186
Shooting	917
Gymnastics ...	833
Wrestling	543
Boxing	401
Judo	344
Cycling Road	334
Weightlifting	325
Fencing	251

8. Quá trình phân tích dữ liệu bằng ngôn ngữ truy vấn MDX

8.1. Câu truy vấn 1

Nội dung: Liệt kê top 10 vận động viên đạt huy chương vàng nhất, sắp xếp theo thứ tự giảm dần.

Thực thi:

```
WITH MEMBER [Measures].[Số lượng huy chương vàng]
AS [Measures].[Fact Olympic Count]

SELECT [Measures].[Số lượng huy chương vàng] ON COLUMNS ,
NON EMPTY TOPCOUNT([Dim Athlete].[Athlete Full
Name].children,10,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]

WHERE [Dim Result].[Medal Type].[GOLD]
```

Kết quả:

```

--Câu 1: Liệt kê top 10 vận động viên đạt huy chương vàng nhất, sắp xếp theo thứ tự giảm dần.
WITH MEMBER [Measures].[Số lượng huy chương vàng] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng huy chương vàng] ON COLUMNS ,
NON EMPTY TOPCOUNT([Dim Athlete].[Athlete Full Name].children,10,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE [Dim Result].[Medal Type].[GOLD]

--Câu 2: Liệt kê tất cả các vận động viên tham gia Olympic từ 3 lần trở lên, sắp theo theo thứ tự tăng dần
WITH MEMBER [Measures].[Số lần tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lần tham gia] ON COLUMNS ,
NON EMPTY [Dim Athlete].[Athlete Full Name].CHILDREN
HAVING [Measures].[Số lần tham gia] >= 3 ON ROWS
FROM [Data Warehouse]

--Câu 3: Liệt kê 10 môn thi đấu có số lượng huy chương đồng nhiều nhất trong kỳ thi vận hội Tokyo 2020.
SELECT {[Measures].[Fact Olympic Count]} ON COLUMNS,
{TopCount(
{[Dim Competition].[Discipline Title].CHILDREN }
,10
,[Measures].[Fact Olympic Count]
)} ON ROWS

```

Số lượng huy chương vàng	
Michael PHELPS	13
Ray EWRY	8
Carl LEWIS	7
Björn DAEHLIE	6
Boris SHAKHIN	6
Larisa LATYNINA	6
Lidiya SKOBLIKOVA	6
Nikolay ANDRIANOV	6
Paavo NURMI	6
Usain BOLT	6

8.2. Câu truy vấn 2

Nội dung: Liệt kê tất cả các vận động viên tham gia Olympic từ 3 lần trở lên, sắp theo theo thứ tự tăng dần

Thực thi:

```
WITH MEMBER [Measures].[Số lần tham gia] AS [Measures].[Fact Olympic Count]
```

```

SELECT [Measures].[Số lần tham gia] ON COLUMNS ,
NON EMPTY [Dim Athlete].[Athlete Full Name].CHILDREN
HAVING [Measures].[Số lần tham gia] >= 3 ON ROWS
FROM [Data Warehouse]

```

Kết quả:

```

--Câu 2: Liệt kê tất cả các vận động viên tham gia Olympic từ 3 lần trở lên, sắp theo thứ tự tăng dần
WITH MEMBER [Measures].[Số lần tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lần tham gia] ON COLUMNS,
NON EMPTY [Dim Athlete].[Athlete Full Name].CHILDREN
HAVING [Measures].[Số lần tham gia] >= 3 ON ROWS
FROM [Data Warehouse]

--Câu 3: Liệt kê 10 môn thi đấu có số lượng huy chương đồng nhiều nhất trong kỳ thi vận hội Tokyo 2020.
SELECT {[Measures].[Fact Olympic Count]} ON COLUMNS,
{TopCount(
{[Dim Competition].[Discipline Title].CHILDREN }
,10
,[Measures].[Fact Olympic Count]
)} ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Result].[Medal Type]&[BRONZE]} * {[Dim Game].[Game Name]&[Tokyo 1964]}

--Câu 4: Vận động viên France dành được nhiều huy chương bạc nhất
WITH MEMBER [Measures].[Số lượng huy chương bạc] AS [Measures].[Fact Olympic Count]
SELECT {[Measures].[Fact Olympic Count]} ON COLUMNS,
NON EMPTY [Dim Athlete].[Athlete Full Name].CHILDREN
HAVING [Measures].[Số lượng huy chương bạc] >= 3 ON ROWS
FROM [Data Warehouse]

```

Athlete Full Name	Số lần tham gia
Aage Avaldorff MEYER	3
Aagje Ada KOK	3
Arme ROINE	8
Arme SALOVARA	4
Aaron BLUNCK	3
Aaron MARCH	4
Aaron PEIRSON	5
Aatos Tuomas JASKARI	3
Abbas ZANDI	3
Abbey WEITZEL	4
Abby LARSON	3
Abdalati IGUIDER	4
Abdelhafid BENCHABLA	4
Abdelkader CHADI	3
Abdelouahab FERGUENE	3
Abdihakem ABDIRAHMAN	4

8.3. Câu truy vấn 3

Nội dung: Liệt kê 10 môn thi đấu có số lượng huy chương đồng nhiều nhất trong kỳ thi vận hội Tokyo 2020.

Thực thi:

```

SELECT {[Measures].[Fact Olympic Count]} ON COLUMNS,
{TopCount(
{[Dim Competition].[Discipline Title].CHILDREN }
,10
,[Measures].[Fact Olympic Count]
)} ON ROWS
FROM [Data Warehouse]

```

WHERE {[Dim Result].[Medal Type].&[BRONZE]} * {[Dim Game].[Game Name].&[Tokyo 1964]})

Kết quả:

```

FROM [Data Warehouse]
--Câu 3: Liệt kê 10 môn thi đấu có số lượng huy chương đồng nhiều nhất trong kỳ thi vận hội Tokyo 2020.
SELECT {[Measures].[Fact Olympic Count]} ON COLUMNS,
{TopCount(
{[Dim Competition].[Discipline Title].CHILDREN }
,10
,[Measures].[Fact Olympic Count]
)} ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Result].[Medal Type].&[BRONZE]} * {[Dim Game].[Game Name].&[Tokyo 1964]})

--Câu 4: Vận động viên France dành được nhiều huy chương bạc nhất
WITH MEMBER [Measures].[Số lượng huy chương bạc] AS[Measures].[Fact Olympic Count]
SELECT
{[Measures].[Số lượng huy chương bạc]} ON COLUMNS,
NON EMPTY TopCount([Dim Athlete].[Athlete Full Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Country].[Country Name].&[France]} * {[Dim Result].[Medal Type].&[SILVER]})


```

Sport	Fact Olympic Count
Athletics	31
Wrestling	14
Swimming	13
Gymnastics Artistic	9
Weightlifting	7
Shooting	5
Diving	4
Fencing	4
Cycling Track	3
Boxing	2

8.4. Câu truy vấn 4

Nội dung: Vận động viên France dành được nhiều huy chương bạc nhất

Thực thi:

WITH MEMBER [Measures].[Số lượng huy chương bạc]

AS[Measures].[Fact Olympic Count]

SELECT

{[Measures].[Số lượng huy chương bạc]} ON COLUMNS,

NON EMPTY TopCount([Dim Athlete].[Athlete Full

Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS

FROM [Data Warehouse]

```
WHERE {[Dim Country].[Country Name]&[France]} * {[Dim Result].[Medal Type]&[SILVER]}
```

Kết quả:

```
--Câu 4: Vận động viên France dành được nhiều huy chương bạc nhất
WITH MEMBER [Measures].[Số lượng huy chương bạc] AS [Measures].[Fact Olympic Count]
SELECT
{[Measures].[Số lượng huy chương bạc]} ON COLUMNS,
NON EMPTY TopCount([Dim Athlete].[Athlete Full Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Country].[Country Name]&[France]} * {[Dim Result].[Medal Type]&[SILVER]}

--Câu 5: Quốc gia dành được nhiều huy vàng nhất từ năm 1980 đến năm 2000
WITH MEMBER [Measures].[Số lượng huy chương vàng] AS [Measures].[Fact Olympic Count]
SELECT
{[Measures].[Số lượng huy chương vàng]} ON COLUMNS,
TOPCOUNT([Dim Country].[Country Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Time].[Game Year].[1980] : [Dim Time].[Game Year].[2000]} * {[Dim Result].[Medal Type]&[GOLD]}

--Câu 6: Thống kê số lượng vận động viên của các quốc gia tham gia thi đấu môn Ski Jumping trong thế vận hội
WITH MEMBER [Measures].[Số lượng vận động viên] AS [Measures].[Fact Olympic Count]
SELECT
{[Measures].[Số lượng vận động viên]} ON COLUMNS,
NON EMPTY TopCount([Dim Country].[Country Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
```

Messages	Results
Alain MIMOUN	Số lượng huy chương bạc 3

8.5. Câu truy vấn 5

Nội dung: Quốc gia dành được nhiều huy vàng nhất từ năm 1980 đến năm 2000

Thực thi:

```
WITH MEMBER [Measures].[Số lượng huy chương vàng] AS
[Measures].[Fact Olympic Count]
```

SELECT

```
{[Measures].[Số lượng huy chương vàng]} ON COLUMNS,
```

```

TOPCOUNT([Dim Country].[Country Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE ([Dim Time].[Game Year].[1980] : [Dim Time].[Game Year].[2000}) * {[Dim Result].[Medal Type]&[GOLD]}

```

Kết quả:

The screenshot shows the Microsoft Analysis Services Management Studio. On the left, the 'Cube' dropdown is set to 'Data Warehouse'. The 'Measure Group' dropdown shows '<All>'. The 'Data Warehouse' tree view includes categories like Data Warehouse, Measures, KPIs, Dim Athlete, Dim Competition, Dim Game, Dim Result, and Dim Time. The main pane displays a T-SQL query:

```

WHERE ([Dim Country].[Country Name]&[France]) * {[Dim Result].[Medal Type]&[SILVER]}
--Câu 5: Quốc gia dành được nhiều huy vàng nhất từ năm 1980 đến năm 2000
WITH MEMBER [Measures].[Số lượng huy chương vàng] AS [Measures].[Fact Olympic Count]
SELECT {[Measures].[Số lượng huy chương vàng]} ON COLUMNS,
TOPCOUNT([Dim Country].[Country Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE ([Dim Time].[Game Year].[1980] : [Dim Time].[Game Year].[2000]) * {[Dim Result].[Medal Type]&[GOLD]}

```

The results pane at the bottom shows a table with one row:

Messages	Results	
	Số lượng huy chương vàng	
	United States of America	177

8.6. Câu truy vấn 6

Nội dung: Thông kê số lượng vận động viên của các quốc gia tham gia thi đấu môn Ski Jumping trong thế vận hội Sochi 2014.

Thực thi:

```

WITH MEMBER [Measures].[Số lượng vận động viên] AS
[Measures].[Fact Olympic Count]

```

SELECT

```
{[Measures].[Số lượng vận động viên]} ON COLUMNS,
NON EMPTY {[Dim Country].[Country Name].Children } ON ROWS
FROM [Data Warehouse]
WHERE ({[Dim Competition].[Discipline Title].&[Ski Jumping]} *
{[Dim Game].[Game Name].&[Sochi 2014]})
```

Kết quả:

Nước	Số lượng vận động viên
Austria	10
Bulgaria	2
Canada	10
Czech Republic	8
Estonia	4
Finland	9
France	5
Germany	11
Italy	8
Japan	11
Kazakhstan	4
Norway	12
Poland	8
Republic of Korea	6
Romania	2
Russian Federation	8

8.7. Câu truy vấn 7

Nội dung: Quốc gia đạt số lượng huy chương bạc nhiều nhất thế vận hội London 2012.

Thực thi:

```

WITH MEMBER [Measures].[Số lượng huy chương bạc] AS
[Measures].[Fact Olympic Count]

SELECT
{[Measures].[Số lượng huy chương bạc]} ON COLUMNS,
NON EMPTY TOPCOUNT([Dim Country].[Country
Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE ({[Dim Game].[Game Name].&[London 2012]} * {[Dim
Result].[Medal Type].&[SILVER]})

```

Kết quả:

The screenshot shows the Microsoft Analysis Services Management Studio environment. On the left, the 'Cube' dropdown is set to 'Data Warehouse'. Below it, the 'Measure Group' dropdown shows 'Data Warehouse' selected. The main area contains a T-SQL query:

```

FROM [Data Warehouse]
WHERE ({[Dim Competition].[Discipline Title].&[Ski Jumping]} * {[Dim Game].[Game Name].&[Sochi 2014]})

--Câu 7: Quốc gia đạt số lượng huy chương bạc nhiều nhất thế vận hội London 2012.
WITH MEMBER [Measures].[Số lượng huy chương bạc] AS [Measures].[Fact Olympic Count]
SELECT
{[Measures].[Số lượng huy chương bạc]} ON COLUMNS,
NON EMPTY TOPCOUNT([Dim Country].[Country Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE ({[Dim Game].[Game Name].&[London 2012]} * {[Dim Result].[Medal Type].&[SILVER]})

--Câu 8: Liệt kê top 5 nội dung thi có số lượng vận động viên tham gia nhiều nhất trong kỳ thế vận hội Tokyo 2020.
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS,
NON EMPTY TOPCOUNT({[Dim Competition].[Discipline Title].children * [Dim Competition].[Event Title].children},5,[Measures].[Số lượng VĐV tham gia]) ON ROWS
FROM [Data Warehouse]
WHERE ({[Dim Game].[Game Name].&[Tokyo 2020]})

--Câu 9: Thống kê số lượng vận động viên tham gia các nội dung thi đấu qua các kỳ Olympics đến nay của đội Trung Quốc.
WITH MEMBER [Measures].[Số lượng huy chương bạc] AS [Measures].[Fact Olympic Count]
SELECT

```

The results pane at the bottom shows a single row for 'People's Republic of China' with a value of 25.

8.8. Câu truy vấn 8

Nội dung: Liệt kê top 5 nội dung thi có số lượng vận động viên tham gia nhiều nhất trong kỳ thế vận hội Tokyo 2020.

Thực thi:

```
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS  
[Measures].[Fact Olympic Count]  
  
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,  
  
NON EMPTY TOPCOUNT( {[Dim Competition].[Discipline  
Title].children * [Dim Competition].[Event Title].children}, 5,  
[Measures].[Số lượng VĐV tham gia]) ON ROWS  
  
FROM [Data Warehouse]  
  
WHERE ([Dim Game].[Game Name].&[Tokyo 2020])
```

Kết quả:

```

FROM [Data warehouse]
WHERE {[Dim Competition].[Discipline Title].&[Ski Jumping]} * {[Dim Game].[Game Name].&[Sochi 2014]}

--Câu 7: Quốc gia đạt số lượng huy chương bạc nhiều nhất thế vận hội London 2012.
WITH MEMBER [Measures].[Số lượng huy chương bạc] AS [Measures].[Fact Olympic Count]
SELECT
    {[Measures].[Số lượng huy chương bạc]} ON COLUMNS,
    NON EMPTY TOPCOUNT([Dim Country].[Country Name].children,1,[Measures].[Fact Olympic Count]) ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Game].[Game Name].&[London 2012]} * {[Dim Result].[Medal Type].&[SILVER]}

--Câu 8: Liệt kê top 5 nội dung thi có số lượng vận động viên tham gia nhiều nhất trong kỳ thế vận hội Tokyo 2020.
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,
    NON EMPTY TOPCOUNT( {[Dim Competition].[Discipline Title].children} * [Dim Competition].[Event Title].children) ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Game].[Game Name].&[Tokyo 2020]}

--Câu 9: Thống kê số lượng vận động viên tham gia các nội dung thi đấu qua các kỳ Olympics đến nay của đội tuyển Việt Nam.
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT
    {[Measures].[Số lượng VĐV tham gia]} ON COLUMNS,
    NON EMPTY {[Dim Game].[Game Name].children} ON ROWS
FROM [Data Warehouse]

```

Messages Results

		Số lượng VĐV tham gia
Cycling Road	Men's Road Race	125
Athletics	Men's Marathon	103
Equestrian	Jumping Individual	87
Athletics	Women's Marathon	85
Equestrian	Eventing Individual	69

8.9. Câu truy vấn 9

Nội dung: Thống kê số lượng vận động viên tham gia các nội dung thi đấu qua các kỳ Olympics đến nay của đội tuyển Việt Nam.

Thực thi:

```

WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS
[Measures].[Fact Olympic Count]

SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,
NON EMPTY ({[Dim Game].[Game Name].CHILDREN * [Dim
Competition].[Discipline Title].CHILDREN * [Dim
Competition].[Event Title].children}) ON ROWS

FROM [Data Warehouse]

WHERE {[Dim Country].[Country Name].&[Vietnam]}

```

Kết quả:

The screenshot shows the SSAS Management Studio environment. On the left, the 'Cube' dropdown is set to 'Data Warehouse'. The 'Measure Group' dropdown shows '<All>'. The 'Search Model' button is visible. The main area contains T-SQL code for three queries:

```

NON EMPTY TOPCOUNT( {[Dim Competition].[Discipline Title].children * [Dim Competition].[Event Title].children}, 10)
FROM [Data Warehouse]
WHERE ({[Dim Game].[Game Name]}.&[Tokyo 2020])

--Câu 9: Thống kê số lượng vận động viên tham gia các nội dung thi đấu qua các kỳ Olympics đến nay của đội
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS,
NON EMPTY {[Dim Game].[Game Name].CHILDREN * [Dim Competition].[Discipline Title].CHILDREN * [Dim Competition].children}
FROM [Data Warehouse]
WHERE {[Dim Country].[Country Name]}.&[Vietnam])

--Câu 10: Liệt kê số lượng chi tiết từng huy chương của Pháp theo từng kỳ Olympics.
WITH MEMBER [Measures].[Số huy chương vàng] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số huy chương vàng] ON COLUMNS,
NON EMPTY {[Dim Game].[Game Name].Children} ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Country].[Country Name]}.&[France]

--Câu 11: Với mỗi quốc gia đưa ra 3 môn thi đấu có nhiều vận động viên tham gia nhiều nhất từ trước đến nay
SELECT {[Measures].[Fact Olympic Count]} ON COLUMNS,
NON EMPTY {[Dim Game].Children} ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Country].[Country Name]}.&[France]

```

The results pane shows a table titled 'Messages' with data from the third query:

			Số lượng VĐV tham gia
Athens 2004	Shooting	10m air pistol 60 shots men	1
Athens 2004	Swimming	100m breaststroke men	1
Athens 2004	Table Tennis	Singles men	1
Athens 2004	Taekwondo	80 kg men	1
Athens 2004	Taekwondo	Athens 2004 Taekwondo - 58 kg men	1
Athens 2004	Weightlifting	63kg women	1
Atlanta 1996	Judo	48kg extra lightweight women	1
Barcelona 1992	Shooting	25m rapid fire pistol 60 shots men	1
Beijing 2008	Badminton	Singles men	1
Beijing 2008	Gymnastics Artistic	balance beam women	1
Beijing 2008	Gymnastics Artistic	floor exercises women	1
Beijing 2008	Gymnastics Artistic	individual all-round women	1
Beijing 2008	Gymnastics Artistic	uneven bars women	1
Beijing 2008	Gymnastics Artistic	vault women	1
Beijing 2008	Shooting	10m air pistol 60 shots men	1
Beijing 2008	Shooting	50m pistol 60 shots men	1

8.10. Câu truy vấn 10

Nội dung: Liệt kê số lượng huy chương của Pháp theo từng kỳ Olympics.

Thực thi:

```
WITH MEMBER [Measures].[Số huy chương vàng] AS [Measures].[Fact Olympic Count]
```

```
SELECT [Measures].[Số huy chương vàng] ON COLUMNS,
```

```
NON EMPTY {[Dim Game].[Game Name].Children} ON ROWS
```

```
FROM [Data Warehouse]
```

```
WHERE {[Dim Country].[Country Name]}.&[France]
```

Kết quả:

The screenshot shows the SSMS interface with the cube editor on the left and the results grid on the right.

Cube: Data Warehouse

Measure Group: <All>

Results Grid:

	Số huy chương vàng
Albertville 1992	116
Amsterdam 1928	104
Antwerp 1920	126
Athens 1896	28
Athens 2004	187
Atlanta 1996	173
Barcelona 1992	188
Beijing 2008	203
Beijing 2022	121
Berlin 1936	131
Calgary 1988	83
Chamonix 1924	33
Cortina d'Ampezzo 1956	42
Garmisch-Partenkirchen 1936	8
Grenoble 1968	77
Helsinki 1952	189

8.11. Câu truy vấn 11

Nội dung: Với mỗi quốc gia đưa ra 3 môn thi đấu có nhiều vận động viên tham gia nhiều nhất từ trước đến nay

Thực thi:

```

SELECT {[Measures].[Fact Olympic Count]} ON COLUMNS,
{GENERATE(
[Dim Country].[Country Name].CHILDREN,
TopCount(
{[Dim Country].[Country Name].CURRENTMEMBER * [Dim Competition].[Discipline Title].Children},3,[Measures].[Fact Olympic Count])} ON ROWS
FROM [Data Warehouse]

```

Kết quả:

The screenshot shows the Microsoft SQL Server Management Studio interface. On the left, there is a tree view of the cube structure under 'Data Warehouse' with various dimensions like Dim Athlete, Dim Competition, Dim Country, etc., and measures like Measures, KPIs, and Fact Olympic Count. The main area contains two T-SQL queries labeled 'Câu 11' and 'Câu 12'. The results grid on the right displays data from the Fact Olympic Count table, showing countries, sports, and their counts.

```

WITH MEMBER [Measures].[Số huy chương vàng] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số huy chương vàng] ON COLUMNS,
NON EMPTY {[Dim Game].[Game Name].Children} ON ROWS
FROM [Data Warehouse]
WHERE ([Dim Country].[Country Name].&[France])

--Câu 11: Với mỗi quốc gia đưa ra 3 môn thi đấu có nhiều vận động viên tham gia nhiều nhất từ trước đến nay
SELECT {[Measures].[Fact Olympic Count]} ON COLUMNS,
{GENERATE(
[Dim Country].[Country Name].CHILDREN,
TopCount(
{[Dim Country].[Country Name].CURRENTMEMBER * [Dim Competition].[Discipline Title].Children},3,[Measures].[Fact Olympic Count])
) } ON ROWS
FROM [Data Warehouse]

--Câu 12: Các vận động viên bị trùng hạng tư của nước United States of America
WITH MEMBER [Measures].[Số lần trùng hạng 4] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lần trùng hạng 4] ON COLUMNS,
NON EMPTY {[Dim Athlete].[Athlete Full Name].Children} ON ROWS
FROM [Data Warehouse]
WHERE ([Dim Country].[Country Name].&[United States of America], [Dim Result].[Rank Equal].&[True], [Dim Result].[Rank Position].&[4] )

```

Messages	Results	
	Fact Olympic Count	
Afghanistan	Wrestling	31
Afghanistan	Athletics	4
Afghanistan	Boxing	4
Albania	Weightlifting	19
Albania	Shooting	17
Albania	Alpine Skiing	15
Algeria	Boxing	72
Algeria	Athletics	63
Algeria	Judo	41
American Samoa	Weightlifting	8
American Samoa	Boxing	5
American Samoa	Swimming	4
Andorra	Alpine Skiing	118
Andorra	Cross Country Skiing	16
Andorra	Shooting	10
Angola	Swimming	34

8.12. Câu truy vấn 12

Nội dung: Các vận động viên bị trùng hạng tư của nước United States of America

Thực thi:

```
WITH MEMBER [Measures].[Số lần trùng hạng 4] AS [Measures].[Fact Olympic Count]
```

```
SELECT [Measures].[Số lần trùng hạng 4] ON COLUMNS,
NON EMPTY {[Dim Athlete].[Athlete Full Name].Children} ON ROWS
FROM [Data Warehouse]
```

```
WHERE ([Dim Country].[Country Name].&[United States of America],
[Dim Result].[Rank Equal].&[True], [Dim Result].[Rank Position].&[4] )
```

Kết quả:

The screenshot shows the SSMS interface with a query window containing T-SQL code and a results window displaying the output of the query.

```

[Dim Country].[Country Name].CHILDREN,
TopCount(
{[Dim Country].[Country Name].CURRENTMEMBER * [Dim Competition].[Discipline Title].Children},3,[Measures].
FROM [Data Warehouse]

--Câu 12: Các vận động viên bị trùng hạng tư của nước United States of America
WITH MEMBER [Measures].[Số lần trung hạng 4] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lần trung hạng 4] ON COLUMNS,
NON EMPTY {[Dim Athlete].[Athlete Full Name].Children } ON ROWS
FROM [Data Warehouse]
WHERE ([Dim Country].[Country Name].&[United States of America], [Dim Result].[Rank Equal].&[True], [Dim Result].[Medal Type].&[GOLD])

--Câu 13: Tìm vận động viên giành được nhiều huy chương vàng nhất Olympic Beijing 2008
WITH MEMBER [Measures].[Số HCV] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số HCV] ON COLUMNS ,
NON EMPTY TOPCOUNT( {[Dim Athlete].[Athlete Full Name].children}, 1, [Measures].[Số HCV]) ON ROWS
FROM [Data Warehouse]
WHERE ({[Dim Result].[Medal Type].&[GOLD], [Dim Game].[Game Name].&[Beijing 2008]})

--Câu 14: Khi Olympic với nhiều vận động viên tham gia nhất
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS
FROM [Data Warehouse]

```

Messages	Results
Số lần trung hạng 4	
Andrea MEAD LAWRENCE	1
Anne Kidig KURSINSKI	1
Arthur SEWARD	1
Collin MORIKAWA	1
Egon R. ERICKSON	1
Frank Otto HAUBOLD	1
Janice SMITH	1
Jeanne ASHWORTH	1
Jim THORPE	1
John Richard WILSON	1
Joseph LYDON	1
Kc LIGHTFOOT	1
Kenneth JEWETT	1
Leo GOEHRING	1
Linda Joan METHENY	1
Myron Willis RODERICK	1

8.13. Câu truy vấn 13

Nội dung: Tìm vận động viên giành được nhiều huy chương vàng nhất Olympic Beijing 2008

Thực thi:

```
WITH MEMBER [Measures].[Số HCV] AS [Measures].[Fact Olympic Count]
```

```
SELECT [Measures].[Số HCV] ON COLUMNS ,
```

```
NON EMPTY TOPCOUNT( {[Dim Athlete].[Athlete Full Name].children}, 1, [Measures].[Số HCV]) ON ROWS
```

```
FROM [Data Warehouse]
```

WHERE ([Dim Result].[Medal Type].&[GOLD], [Dim Game].[Game Name].&[Beijing 2008])

Kết quả:

The screenshot shows the Microsoft Analysis Services (SSMS) interface. On the left, the 'Cube' pane displays the 'Data Warehouse' cube with various measure groups and dimensions like Dim Athlete, Dim Competition, Dim Game, etc. In the center, the 'Script' pane contains T-SQL DAX queries for three specific questions. The first query finds the athlete with the most gold medals in Beijing 2008. The second query finds the game with the most athletes. The third query finds the top 10 sports with the most participants. On the right, the 'Results' pane shows the output for the first query, which is Michael Phelps with 5 gold medals.

```

WITH MEMBER [Measures].[Số lần trúng hạng 4] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lần trúng hạng 4] ON COLUMNS,
NON EMPTY {[Dim Athlete].[Athlete Full Name].Children} ON ROWS
FROM [Data Warehouse]
WHERE ([Dim Country].[Country Name].&[United States of America], [Dim Result].[Rank Equal].&[True], [Dim Result].[Medal Type].&[GOLD], [Dim Game].[Game Name].&[Beijing 2008])

--Câu 13: Tìm vận động viên giành được nhiều huy chương vàng nhất Olympic Beijing 2008
WITH MEMBER [Measures].[Số HCV] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số HCV] ON COLUMNS ,
NON EMPTY TOPCOUNT( {[Dim Athlete].[Athlete Full Name].children}, 1, [Measures].[Số HCV]) ON ROWS
FROM [Data Warehouse]
WHERE ([Dim Result].[Medal Type].&[GOLD], [Dim Game].[Game Name].&[Beijing 2008])

--Câu 14: Kì Olympic với nhiều vận động viên tham gia nhất
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,
NON EMPTY TOPCOUNT( {[Dim Game].[Game Name].children}, 1, [Measures].[Số lượng VĐV tham gia]) ON ROWS
FROM [Data Warehouse]

--Câu 15: Top 10 các môn thi có số lượng vận động viên tham gia nhiều nhất với địa điểm tổ chức Olympic nà
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,
NON EMPTY TOPCOUNT( {[Dim Game].[Game Name].children}, 1, [Measures].[Số lượng VĐV tham gia]) ON ROWS
FROM [Data Warehouse]
  
```

Messages	Results
Số HCV	Michael PHELPS 5

8.14. Câu truy vấn 14

Nội dung: Kì Olympic với nhiều vận động viên tham gia nhất

Thực thi:

```
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS
[Measures].[Fact Olympic Count]
```

```
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,
NON EMPTY TOPCOUNT( {[Dim Game].[Game Name].children}, 1,
[Measures].[Số lượng VĐV tham gia]) ON ROWS
FROM [Data Warehouse]
```

Kết quả:

The screenshot shows the SSMS interface with the 'Data Warehouse' cube selected. The query editor contains T-SQL code for three queries (Câu 13, Câu 14, Câu 15) related to Olympic data. The results pane shows the output for Rio 2016, which includes a table with two rows: 'Rio 2016' and '6303'. The table has two columns: 'Số lượng VĐV tham gia' (Number of athletes) and '6303'.

```

FROM [Data Warehouse]
WHERE ([Dim Country].[Country Name].&[United States of America], [Dim Result].[Rank Equal].&[True], [Dim Result].[Medal Type].&[GOLD], [Dim Game].[Game Name].&[Beijing 2008])

--Câu 13: Tìm vận động viên giành được nhiều huy chương vàng nhất Olympic Beijing 2008
WITH MEMBER [Measures].[Số HCV] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số HCV] ON COLUMNS,
NON EMPTY TOPCOUNT( {[Dim Athlete].[Athlete Full Name].children}, 1, [Measures].[Số HCV]) ON ROWS
FROM [Data Warehouse]
WHERE {[Dim Result].[Medal Type].&[GOLD], [Dim Game].[Game Name].&[Beijing 2008]}

--Câu 14: Ki Olympic với nhiều vận động viên tham gia nhất
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS,
NON EMPTY TOPCOUNT( {[Dim Game].[Game Name].children}, 1, [Measures].[Số lượng VĐV tham gia]) ON ROWS
FROM [Data Warehouse]

--Câu 15: Top 10 các môn thi có số lượng vận động viên tham gia nhiều nhất với địa điểm tổ chức Olympic nằm ở quốc gia Great Britain
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS,
NON EMPTY TOPCOUNT( {[Dim Competition].[Discipline Title].children}, 10, [Measures].[Số lượng VĐV tham gia])
FROM [Data Warehouse]

```

8.15. Câu truy vấn 15

Nội dung: Top 10 các môn thi có số lượng vận động viên tham gia nhiều nhất với địa điểm tổ chức Olympic nằm ở quốc gia Great Britain

Thực thi:

```

WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS
[Measures].[Fact Olympic Count]

SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,
NON EMPTY TOPCOUNT( {[Dim Competition].[Discipline Title].children}, 10, [Measures].[Số lượng VĐV tham gia]) ON ROWS
FROM [Data Warehouse]

WHERE ( [Dim Game].[Game Location].&[Great Britain] )

```

Kết quả:

Cube: Data Warehouse

Measure Group: <All>

Measure: [Measures].[Số lượng VĐV tham gia]

```
--Câu 14: Ki Olympic với nhiều vận động viên tham gia nhất
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,
NON EMPTY TOPCOUNT( {[Dim Game].[Game Name].children}, 1, [Measures].[Số lượng VĐV tham gia]) ON ROWS
FROM [Data Warehouse]
```

```
--Câu 15: Top 10 các môn thi có số lượng vận động viên tham gia nhiều nhất với địa điểm tổ chức Olympic là Anh Quốc
WITH MEMBER [Measures].[Số lượng VĐV tham gia] AS [Measures].[Fact Olympic Count]
SELECT [Measures].[Số lượng VĐV tham gia] ON COLUMNS ,
NON EMPTY TOPCOUNT( {[Dim Competition].[Discipline Title].children}, 10, [Measures].[Số lượng VĐV tham gia])
FROM [Data Warehouse]
WHERE ( [Dim Game].[Game Location].&[Great Britain] )
```

Results

Môn thi	Số lượng VĐV tham gia
Athletics	1953
Swimming	1186
Shooting	917
Gymnastics Artistic	833
Wrestling	543
Boxing	401
Judo	344
Cycling Road	334
Weightlifting	325
Fencing	251

9. Quá trình phân tích dữ liệu bằng công cụ Pivot Excel

9.1. Câu truy vấn 1

Nội dung:

Liệt kê top 10 vận động viên đạt huy chương vàng nhất, sắp xếp theo thứ tự giảm dần.

Kết quả:

	MedalType	GOLD
1	MedalType	GOLD
2	Row Labels	Fact Olympic Count
3	Boris SHAKHIN	6
4	Carl LEWIS	7
5	Larisa LATYNINA	6
6	Lidiya SKOBILKOVA	6
7	Michael PHELPS	13
8	Nikolay ANDRIANOV	6
9	Paavo NURMI	6
10	Ray EWRY	8
11	Usain BOLT	6
12	Vera CASLAVSKA	6
13		

9.2. Câu truy vấn 2

Nội dung:

Liệt kê tất cả các vận động viên tham gia Olympic từ 3 lần trở lên, sắp xếp theo thứ tự tăng dần

Kết quả:

	Fact Olympic Count
1	Row Labels
2	Aage Avaldorff MEYER
3	Angie Ada KOK
4	Aarne ROINE
5	Aarne SALOVARA
6	Aaron BLUNCK
7	Aaron MARCH
8	Aaron PEIRSON
9	Aatos Tuomas JASKARI
10	Abbas ZANDI
11	Abibey WEITZEL
12	Abibey LARSON
13	Abdallaati GUIADER
14	Abdelhafid BENCHABLA
15	Abdelkader CHADI
16	Abdelouahab FERGUENE
17	Abdihamem ABDIRAHMAN
18	Abdon PAMICH
19	Abdoulalye DJIBA
20	Abdoulalye KOTE
21	Abdul Latif Mohd AL-BULUSHI
22	Abdullah ALRASHIDI
23	Abdullah YILMAZ
24	Abibe Bikila
25	Abel ANTON RODRIGO
26	Ahbinav BINDRA
27	Abraham Israel GROSSFELD

9.3. Câu truy vấn 3

Nội dung:

Liệt kê 10 môn thi đấu có số lượng huy chương đồng nhiều nhất trong kỳ thế vận hội Tokyo 2020.

Kết quả:

The screenshot shows a WPS Office spreadsheet application window. The main area displays a PivotTable with the following data:

	Medal Type	Game Name	Fact Olympic Count
1	BRONZE	Tokyo 2020	
2			
3			
4	Row Labels	Fact Olympic Count	
5	Athletics		1
6	Judo		1
7	Shooting		1
8	Swimming		3
9	Taekwondo		1
10	Wrestling		1
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			

The PivotTable Fields pane on the right side of the screen shows the following structure:

- Fields chosen for report:
 - Fact Olympic
 - Fact Olympic Count (selected)
- Sets:
 - Dim Athlete
 - Athlete Full Name
 - Athlete Id
 - Athlete Year Birth
- Filters:
 - Medal Type
 - Game Name
- Rows:
 - TOP 10 MON THI DAU
- Values:
 - Fact Olympic Count

9.4. Câu truy vấn 4

Nội dung:

Vận động viên France dành được nhiều huy chương bạc nhất

Kết quả:

The screenshot shows a Microsoft Excel-like application window titled "OLAP_PivotExcel.xlsx". The ribbon menu includes File, Home, WPS PDF, Insert, Page Layout, Formulas, Data, Review, View, Automate, Help, PivotTable Analyze, and Design. The main area displays a PivotTable with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	MedalType	SILVER														
2	CountryName	France														
3																
4	Row Labels	Fact Olympic Count														
5	Alain MIMOUN		3													
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																
27																

The PivotTable Fields pane on the right shows the following fields:

- Fact Olympic (selected)
- Fact Olympic Count
- Dim Athlete
 - Athlete Full Name
 - Athlete Id
 - Athlete Year Birth

9.5. Câu truy vấn 5

Nội dung:

Quốc gia dành được nhiều huy vàng nhất từ năm 1980 đến năm 2000

Kết quả:

The screenshot shows a Microsoft Excel spreadsheet titled "OLAP_PivotExcel.xlsx". The PivotTable Fields pane on the right indicates that the "Fact Olympic" field is selected under the "Fact Olympic Count" category. The main table area displays a single row with the following data:

	A	B
1	MedalType	GOLD
2	Game Year	(Multiple Items)
3		
4	Row Labels	Fact Olympic Count
5	United States of America	150
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		

9.6. Câu truy vấn 6

Nội dung:

Thống kê số lượng vận động viên của các quốc gia tham gia thi đấu môn Ski Jumping trong thế vận hội Sochi 2014.

Kết quả:

The screenshot shows a Microsoft Excel-like application window titled "OLAP_PivotExcel.xlsx". The PivotTable Fields pane on the right indicates the data source is "Fact Olympic" with "Fact Olympic Count" selected. The PivotTable itself shows medal counts for 23 countries, with Germany having the highest count of 11. The total medal count is 145.

Country Name	Fact Olympic Count
Austria	10
Bulgaria	2
Canada	10
Czech Republic	8
Estonia	4
Finland	9
France	5
Germany	11
Italy	8
Japan	11
Kazakhstan	4
Norway	12
Poland	8
Republic of Korea	6
Romania	2
Russian Federation	8
Slovenia	11
Switzerland	5
United States of America	11
Grand Total	145

9.7. Câu truy vấn 7

Nội dung:

Quốc gia đạt số lượng huy chương bạc nhiều nhất thế vận hội London 2012.

Kết quả:

	MediaType	Game Name
1	SILVER	London 2012
2		
3		
4	Row Labels	Fact Olympic Count
5	People's Republic of China	25
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		

9.8. Câu truy vấn 8

Nội dung:

Liệt kê top 5 nội dung thi có số lượng vận động viên tham gia nhiều nhất trong kỳ thế vận hội Tokyo 2020.

Kết quả:

The screenshot shows a Microsoft Excel-like application window titled "OLAP_PivotExcel.xlsx". The main area displays a PivotTable with the following data:

	Game Name	Tokyo 2020
Row Labels	Fact Olympic Count	
Athletics		
Men's Marathon	103	
Women's Marathon	85	
Cycling Road		
Men's Road Race	125	
Equestrian		
Eventing Individual	69	
Jumping Individual	87	

To the right of the table, the "PivotTable Fields" pane is open, showing the fields used in the report:

- Fact Olympic (selected)
- Fact Olympic Count
- Dim Athlete
 - Athlete Full Name
 - Athlete Id
 - Athlete Year Birth
- Sets

A status bar at the bottom indicates "Screenshot copied to clipboard and saved".

9.9. Câu truy vấn 9

Nội dung:

Thống kê số lượng vận động viên tham gia các nội dung thi đấu qua các kỳ Olympics đến nay của đội tuyển Việt Nam.

Kết quả:

The screenshot shows a Microsoft Excel spreadsheet titled "OLAP_PivotExcel.xlsx". The PivotTable Fields pane on the right side is open, showing the structure of the data:

- Rows:** Fact Olympic Count (selected), Fact Olympic, Sets, Dim Athlete (Athlete Full Name, Athlete Id, Athlete Year Birth).
- Columns:** None.
- Filters:** Country Name (selected), Game Name, Discipline Title, Event Title.
- Values:** Fact Olympic Count.

The main data grid displays the following data:

Country Name	Fact Olympic Count
Vietnam	1
Athens 2004	
Shooting	1
10m air pistol 60 shots men	1
Swimming	1
100m breaststroke men	1
Table Tennis	1
Singles men	1
Taekwondo	1
80 kg men	1
Athens 2004 Taekwondo - 56 kg men	1
Weightlifting	1
63kg women	1
Atlanta 1996	1
Judo	1
48kg extralightweight women	1
Barcelona 1992	1
Shooting	1
25m rapid fire pistol 60 shots men	1
Beijing 2008	1
Badminton	1
Singles men	1
Gymnastics Artistic	1
balance beam women	1
floor exercises women	1

9.10. Câu truy vấn 10

Nội dung:

Liệt kê số lượng huy chương của Pháp theo từng kỳ Olympics.

Kết quả:

The screenshot shows a Microsoft Excel-like application window titled "OLAP_PivotExcel.xlsx". The main area displays a PivotTable with data from 1928 to 2022. The columns are labeled "Country Name" and "Fact Olympic Count". The data shows that France has the highest count at 131. The PivotTable Fields pane on the right side shows the fields used in the PivotTable, including "Fact Olympic" and "Fact Olympic Count" under "Values", and "Country Name" under "Columns". The "Rows" section shows "Game Name". The status bar at the bottom indicates "1052 PM 5/13/2024".

Country Name	Fact Olympic Count
Albertville 1992	116
Amsterdam 1928	104
Antwerp 1920	126
Athens 1896	28
Athens 2004	187
Atlanta 1996	173
Barcelona 1992	188
Beijing 2008	203
Beijing 2022	121
Berlin 1936	131
Calgary 1988	83
Chamonix 1924	33
Cortina d'Ampezzo 1956	42
Garmisch-Partenkirchen 1936	8
Grenoble 1968	77
Helsinki 1952	189
Innsbruck 1964	43
Innsbruck 1976	43
Lake Placid 1932	6
Lake Placid 1980	36
Lillehammer 1994	105
London 1908	100
London 1948	148
London 2012	190

9.11. Câu truy vấn 11

Nội dung:

Với mỗi quốc gia đưa ra 3 môn thi đấu có nhiều vận động viên tham gia nhiều nhất từ trước đến nay

Kết quả:

The screenshot shows a WPS Office spreadsheet application with a PivotTable Fields pane open on the right side. The main grid displays a list of countries and their Olympic medal counts. The PivotTable Fields pane shows the structure of the data, with 'Fact Olympic' as the primary category and 'Fact Olympic Count' selected. The 'Rows' section contains 'TOP 3 MON THI DAU T...' and the 'Values' section contains 'Fact Olympic Count'. The status bar at the bottom indicates the date and time as 5/13/2024 10:52 PM.

	Fact Olympic Count
1 Row Labels	
2 Afghanistan	
3 Athletics	4
4 Boxing	4
5 Wrestling	31
6 Albania	
7 Alpine Skiing	15
8 Shooting	17
9 Weightlifting	19
10 Algeria	
11 Athletics	63
12 Boxing	72
13 Judo	41
14 American Samoa	
15 Boxing	5
16 Swimming	4
17 Weightlifting	8
18 Andorra	
19 Alpine Skiing	118
20 Cross Country Skiing	16
21 Shooting	10
22 Angola	
23 Boxing	8
24 Judo	13
25 Swimming	34
26 Antigua and Barbuda	
27 Cycling Track	9

9.12. Câu truy vấn 12

Nội dung:

Các vận động viên bị trùng hạng tư của nước United States of America

Kết quả:

The screenshot shows a WPS Office spreadsheet application with a PivotTable Fields pane open on the right side. The data grid contains a list of athletes from the United States of America, ordered by their rank position. The PivotTable Fields pane shows the following structure:

- Fact Olympic Count** (selected under Values)
- Dim Athlete** (under Fields)
 - Athlete Full Name** (selected under Columns)
 - Athlete Id**
 - Athlete Year Birth**
- Filters** (under Fields)
 - Rank Position
 - Rank Equal
 - Country Name
- Rows** (under Fields)
 - Athlete Full Name

9.13. Câu truy vấn 13

Nội dung:

Tìm vận động viên giành được nhiều huy chương vàng nhất Olympic Beijing 2008

Kết quả:

The screenshot shows a WPS Office spreadsheet application window titled "OLAP_Pivot Excel.xlsx". The main area displays a PivotTable with the following data:

	Medal Type	Game Name	Fact Olympic Count
1	GOLD	Beijing 2008	
2			
3			
4	Row Labels	Fact Olympic Count	
5	Michael PHELPS		5

To the right of the data grid is the "PivotTable Fields" pane, which lists the fields used in the PivotTable:

- Fact Olympic (selected)
- Fact Olympic Count
- Dim Athlete
 - Athlete Full Name
 - Athlete Id
 - Athlete Year Birth

The "Filters" section contains "Medal Type" and "Game Name". The "Columns" section contains "Fact Olympic Count". The "Rows" section contains "TOP 1 VDV". The "Values" section contains "Fact Olympic Count".

9.14. Câu truy vấn 14

Nội dung:

Kì Olympic với nhiều vận động viên tham gia nhất

Kết quả:

The screenshot shows a WPS Office spreadsheet application window titled "OLAP_PivotExcel.xlsx". The main area displays a PivotTable with the following data:

Row Labels	Fact Olympic Count
Rio 2016	6303

The "PivotTable Fields" pane on the right side of the interface lists the following fields:

- Fact Olympic (selected)
- Fact Olympic Count
- Dim Athlete
 - Athlete Full Name
 - Athlete Id
 - Athlete Year Birth

The "Filters" section shows "TOP 1 GAME" under the "Rows" category and "Fact Olympic Count" under the "Values" category.

9.15. Câu truy vấn 15

Nội dung:

Top 10 các môn thi có số lượng vận động viên tham gia nhiều nhất với địa điểm tổ chức Olympic nằm ở quốc gia Great Britain

Kết quả:

The screenshot shows a WPS Office spreadsheet application with a PivotTable Fields pane open on the right side. The main data grid displays the following information:

Game Location	Fact Olympic Count
Great Britain	334
Athletics	1953
Boxing	401
Cycling Road	334
Fencing	251
Gymnastics Artistic	833
Judo	344
Shooting	917
Swimming	1186
Weightlifting	325
Wrestling	543

The PivotTable Fields pane shows the following fields:

- Fact Olympic (selected)
- Fact Olympic Count
- Dim Athlete
 - Athlete Full Name
 - Athlete Id
 - Athlete Year Birth

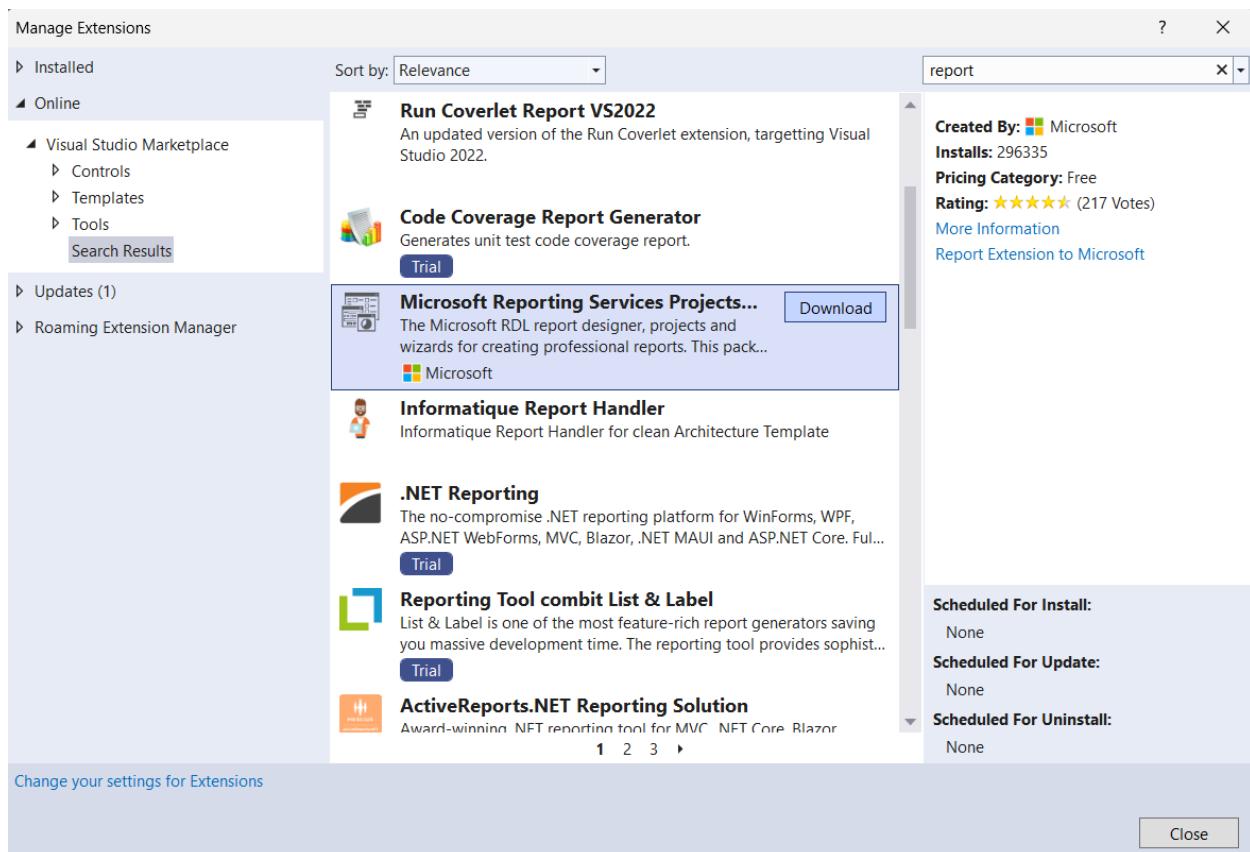
The Filters section shows "Game Location" set to "TOP 10 MON THI DAU". The Values section shows "Fact Olympic Count".

CHƯƠNG 4: QUÁ TRÌNH LẬP BÁO BIỂU (SSRS)

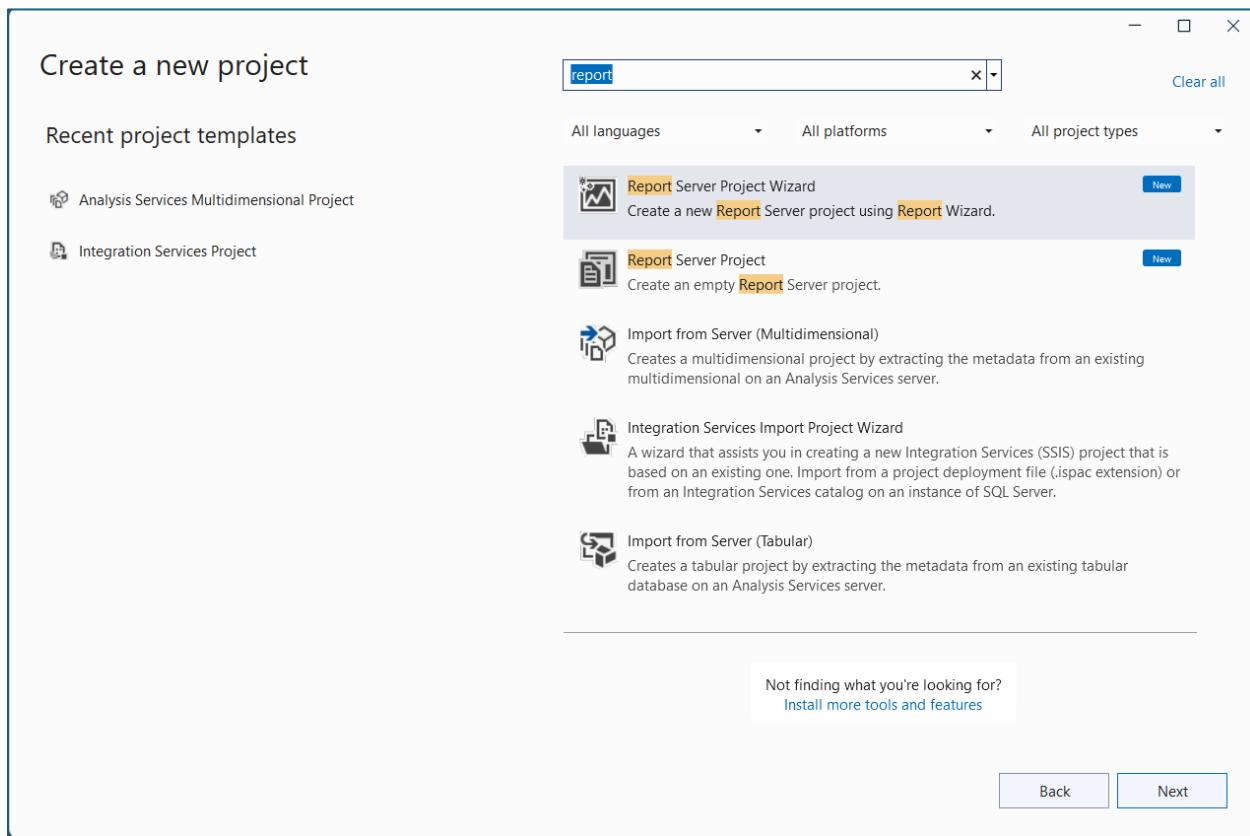
1. QUÁ TRÌNH LẬP BÁO BIỂU BẰNG CÔNG CỤ MICROSOFT SSRS (Report Builder)

1.1. Tạo project SSRS

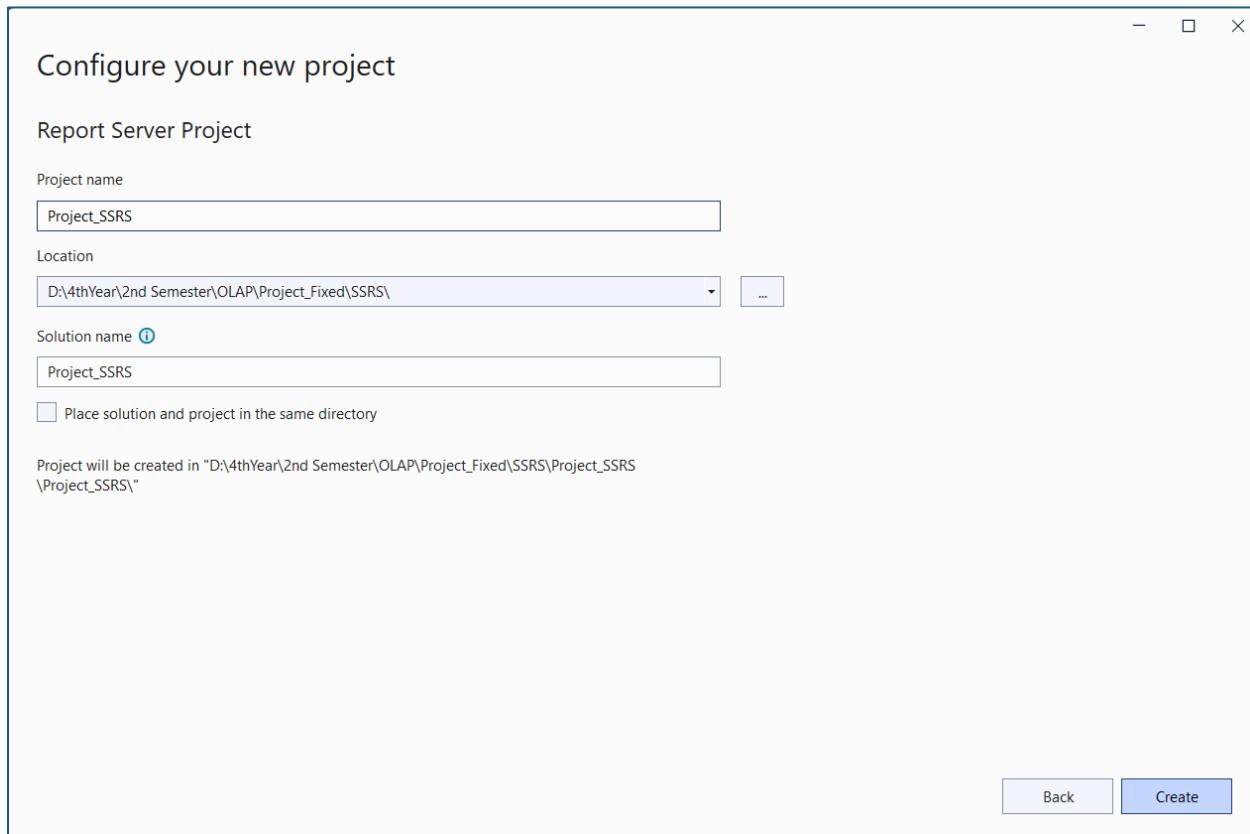
- Bước 1: Tải extension mang tên : Microsoft Reporting Services Projects



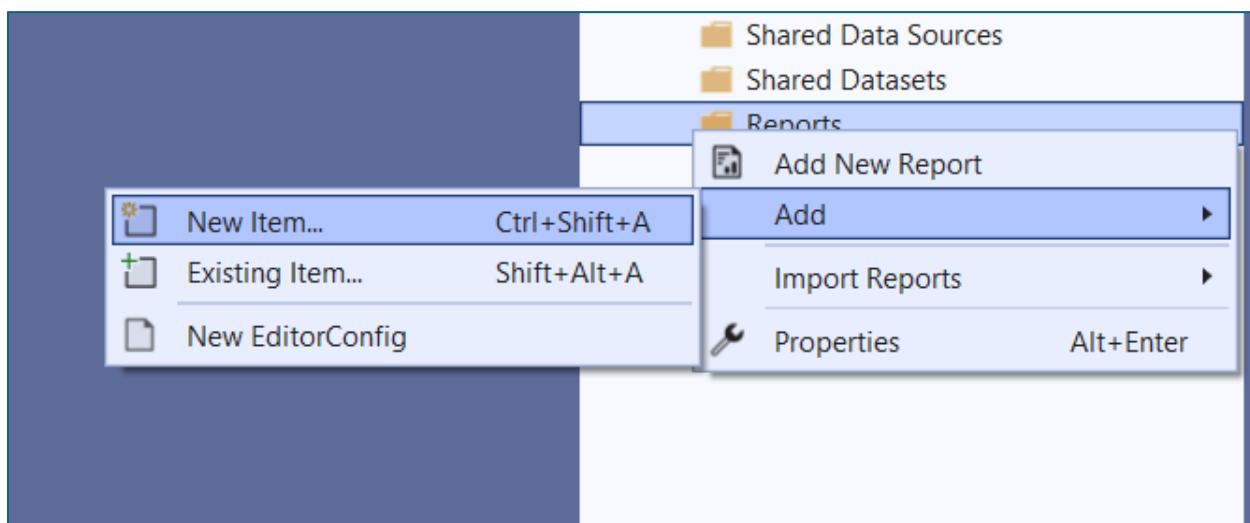
- Bước 2: Mở **Visual Studio** → **File** → **New** → **Projects** hộp thoại như hình dưới đây sẽ hiện ra:



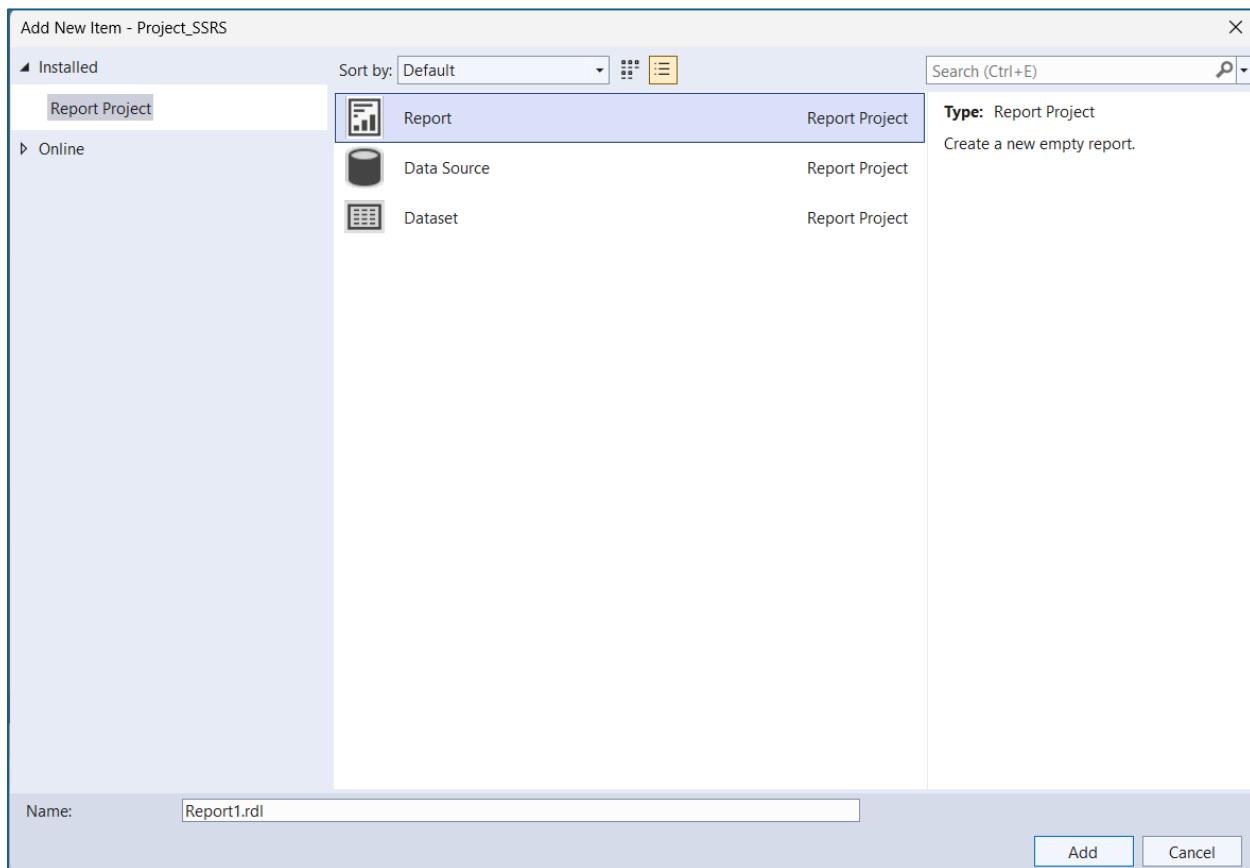
- Bước 3: chọn Report Server Project → Đặt tên project → Dẫn tới đường dẫn cần lưu → Nhấn OK.



- Bước 4: Tại cửa sổ bên phải như hình minh họa, nhấn chuột phải vào phần **Reports**
→ **Add** → **Add new item**



- Bước 5: Chọn Report → đặt tên → nhấn Add

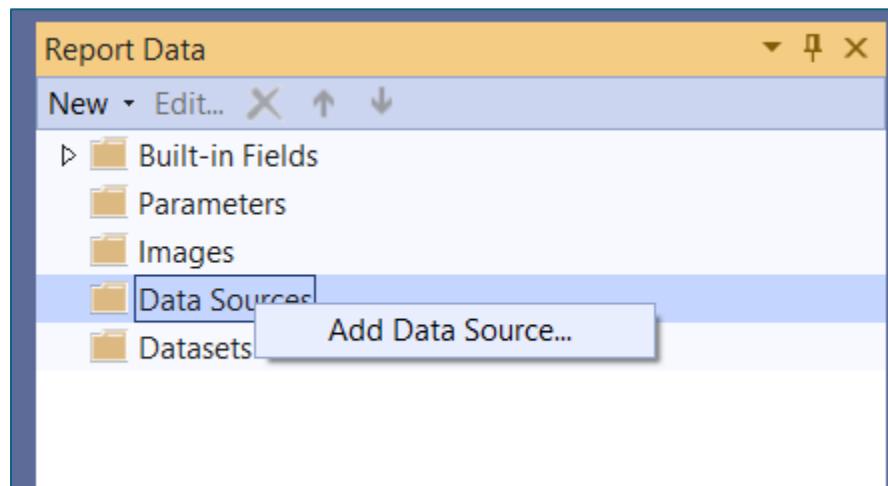


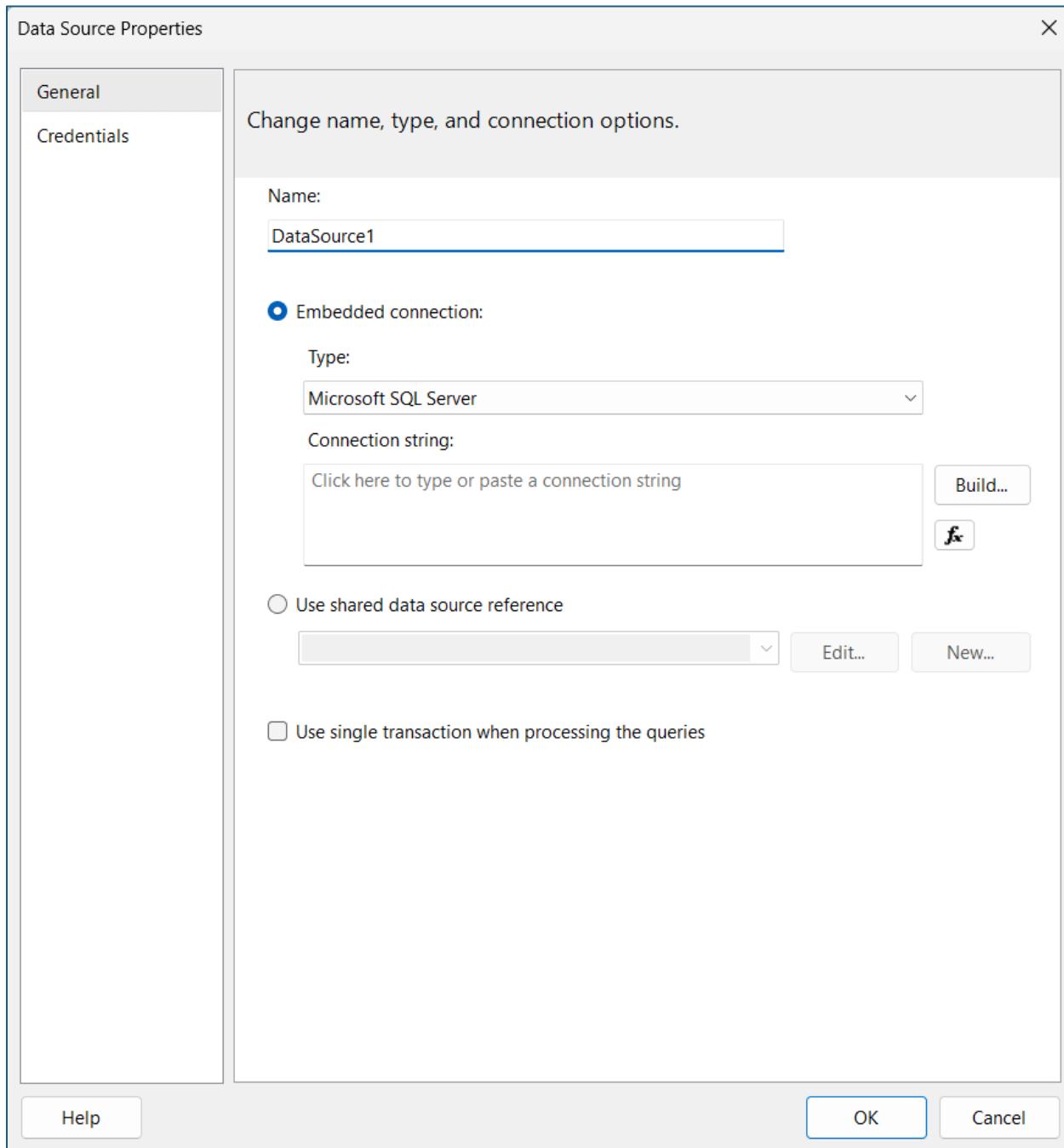
- Bước 6: Kết quả:
 - + Tab Design: Để thiết kế Report
 - + Tab Preview: Để show kết quả Report



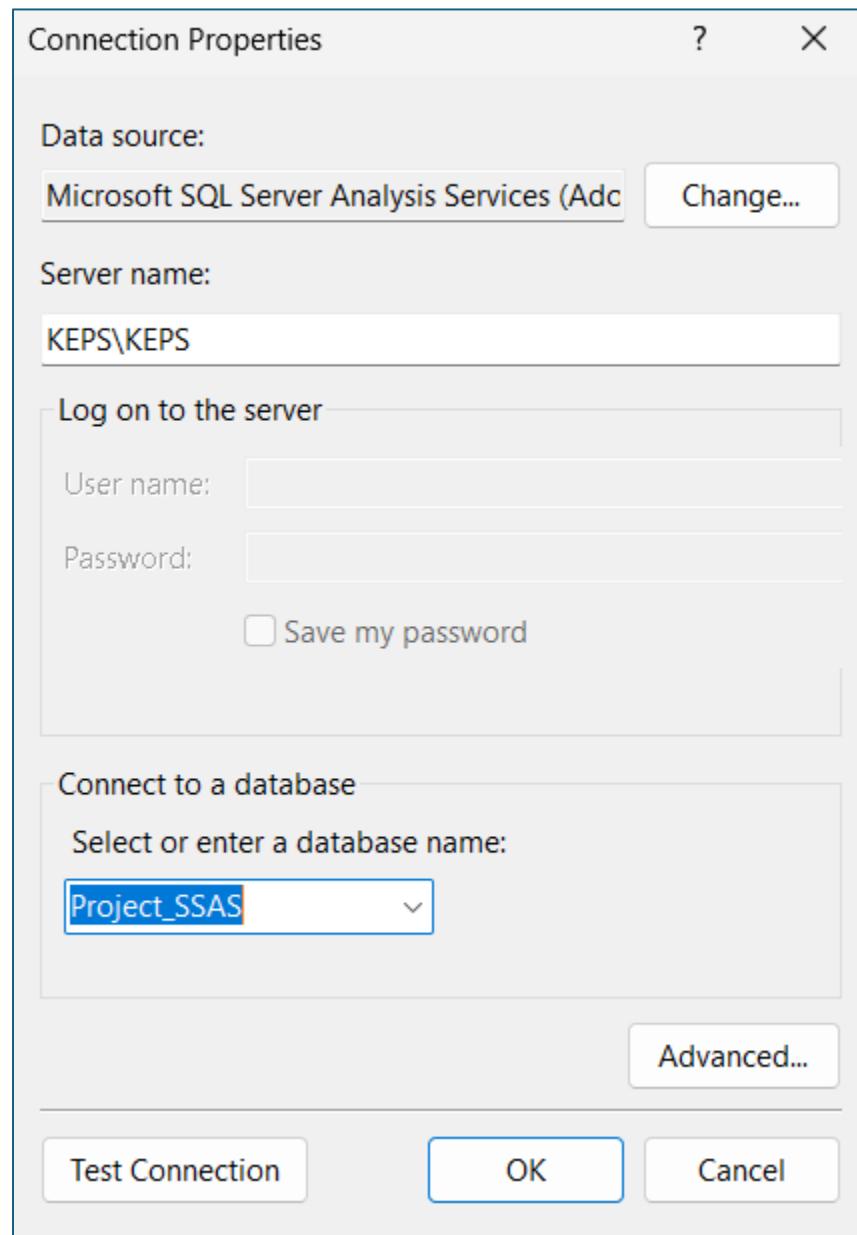
1.2. Tạo kết nối với kho dữ liệu

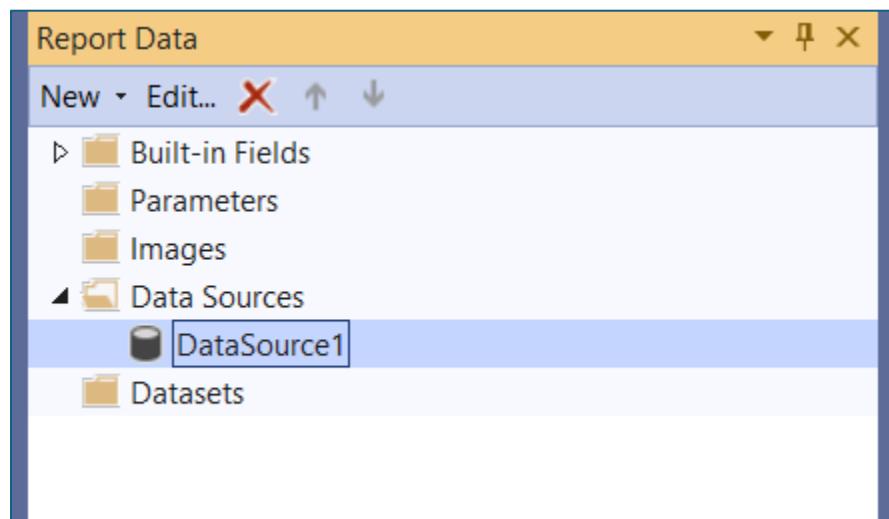
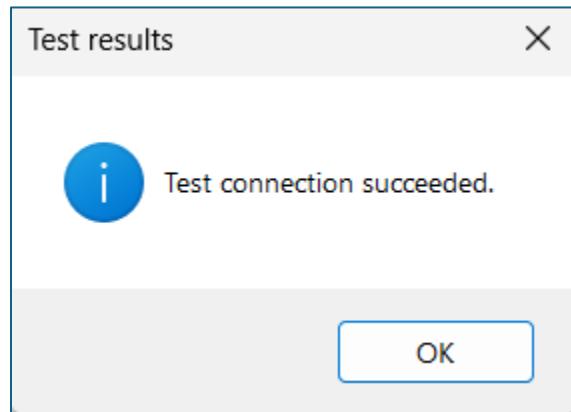
- Bước 1: Tại thanh Report Data → Chọn Data Source → chọn add data source như hình minh họa:





- Bước 2: Trường Type chọn Microsoft SQL Server Analysis Service → Chọn Type và cấu hình như hình minh họa (tùy tên) → rồi nhấn Test Connection. Nếu thành công, nhấn OK → nhấn OK

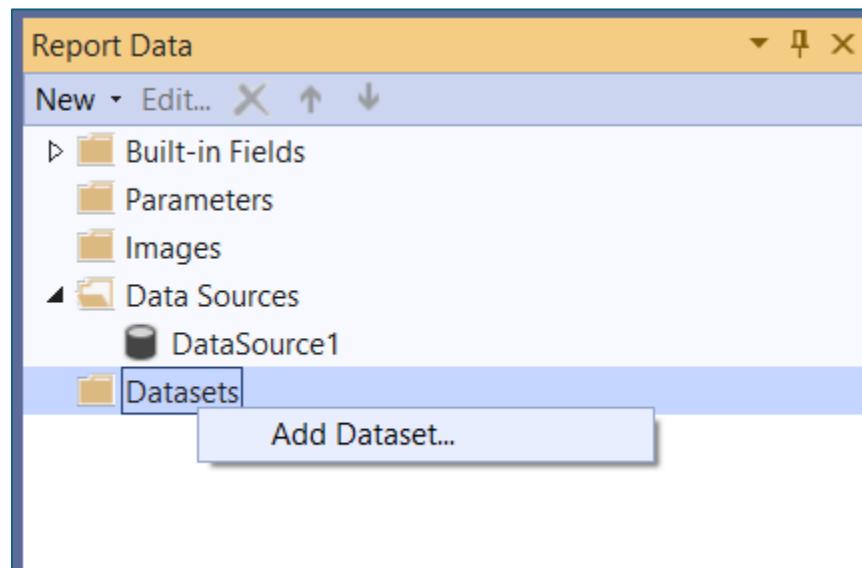




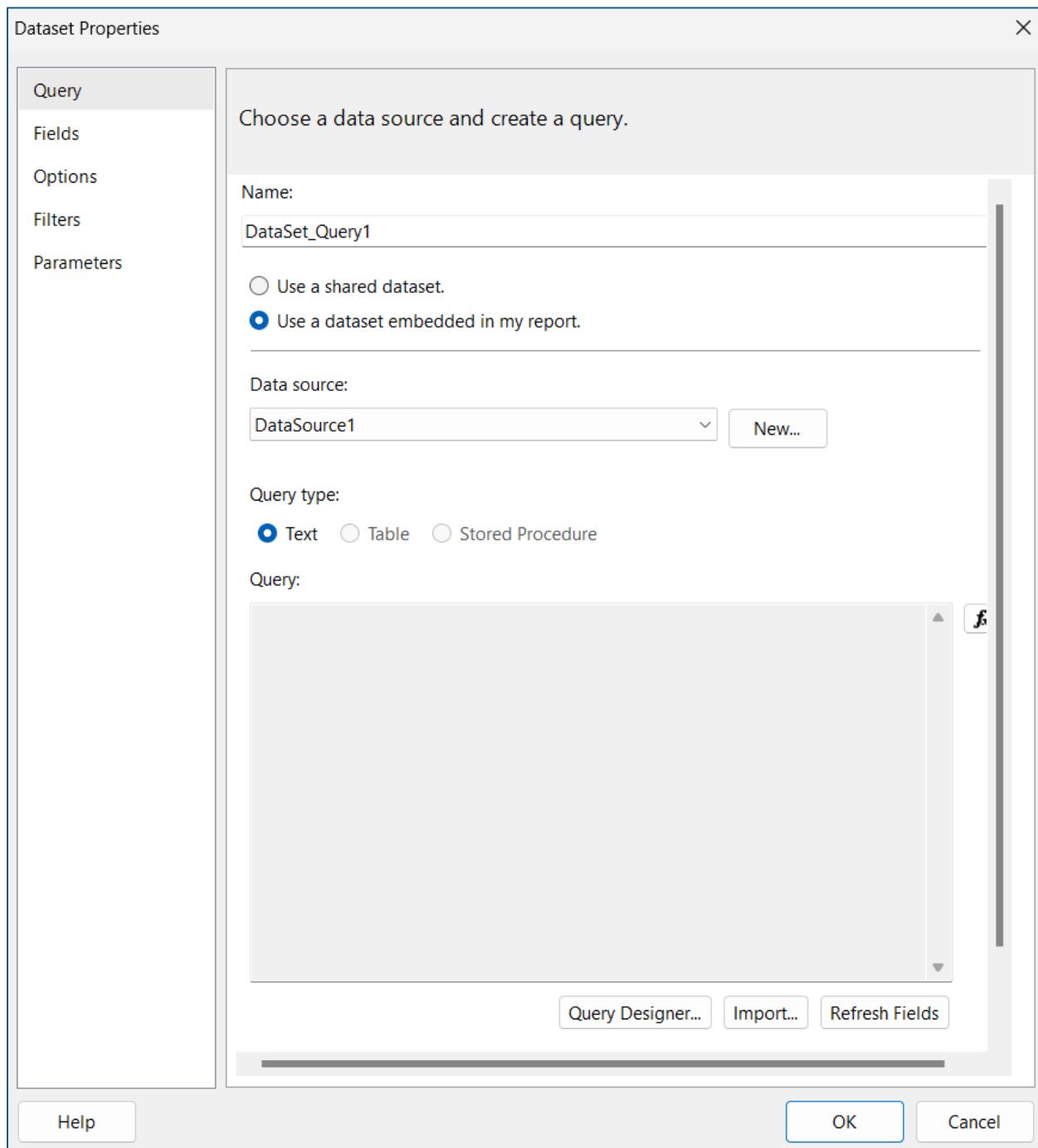
1.3. Thực hiện lập báo biểu trên Report Builder

1.3.1. Báo biểu 1: Thống kê thành tích của đội tuyển Mỹ qua các kì Olympic(2000-2020)

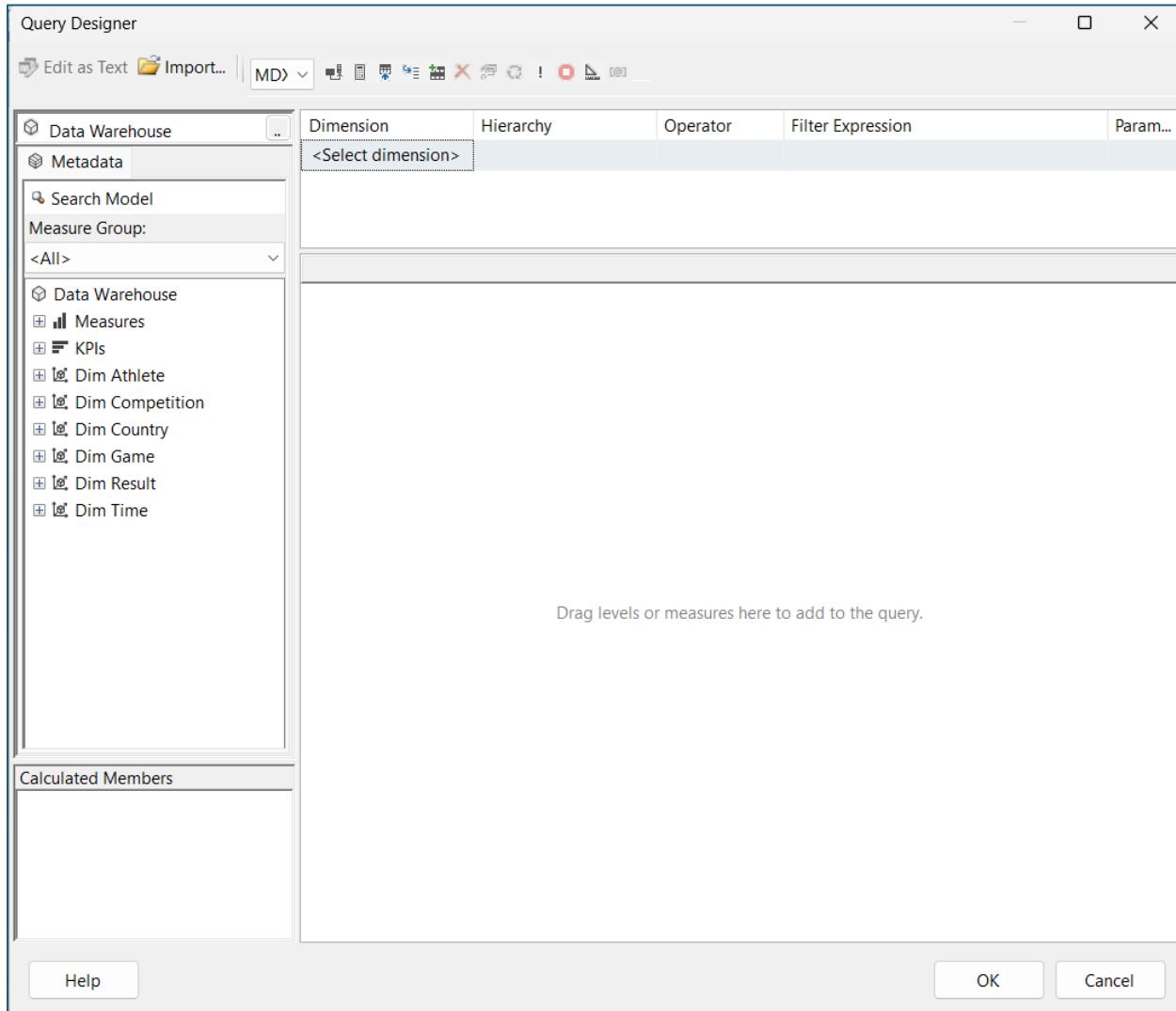
- Bước 1: Thêm datasets cần thiết cho báo biểu 1 đặt tên là datasets_query1, bằng cách nhấn chuột phải vào **Datasets** → **Add new dataset...**



- Bước 2: Trong tab Datasets Properties → Trường Query. Tại trường **Name** đặt tên là Dataset_Query1 chọn chế độ **Use a dataset embedded in my report**. Trường Data Source chọn data source tại phần tạo kết nối **data source**.



- Bước 3: Chọn **Query Designer** hộp thoại sẽ hiện ra:



- Bước 4: Chọn những thuộc tính cần thiết cho báo biểu như hình họa giống như chọn trong quá trình SSAS bằng công cụ BI sau đó nhấn OK.

Query Designer

MD

Dimension	Hierarchy	Operator	Filter Expression	Param
Dim Country	Country Name	Equal	{ United States of America }	<input type="checkbox"/>
<Select dimension>				
Game Year	Game Season	Medal Type	Fact Olympic Count	
2000	Summer	BRONZE	19	
2000	Summer	GOLD	22	
2000	Summer	SILVER	17	
2002	Winter	BRONZE	9	
2002	Winter	GOLD	8	
2002	Winter	SILVER	9	
2004	Summer	BRONZE	16	
2004	Summer	GOLD	23	
2004	Summer	SILVER	32	
2006	Winter	BRONZE	4	
2006	Winter	GOLD	9	
2006	Winter	SILVER	7	
2008	Summer	BRONZE	24	
2008	Summer	GOLD	22	
2008	Summer	SILVER	29	
2010	Winter	BRONZE	10	
2010	Winter	GOLD	8	
2010	Winter	SILVER	10	
2012	Summer	BRONZE	16	
2012	Summer	GOLD	30	
2012	Summer	SILVER	21	
2014	Winter	BRONZE	7	

Calculated Members

Help OK Cancel

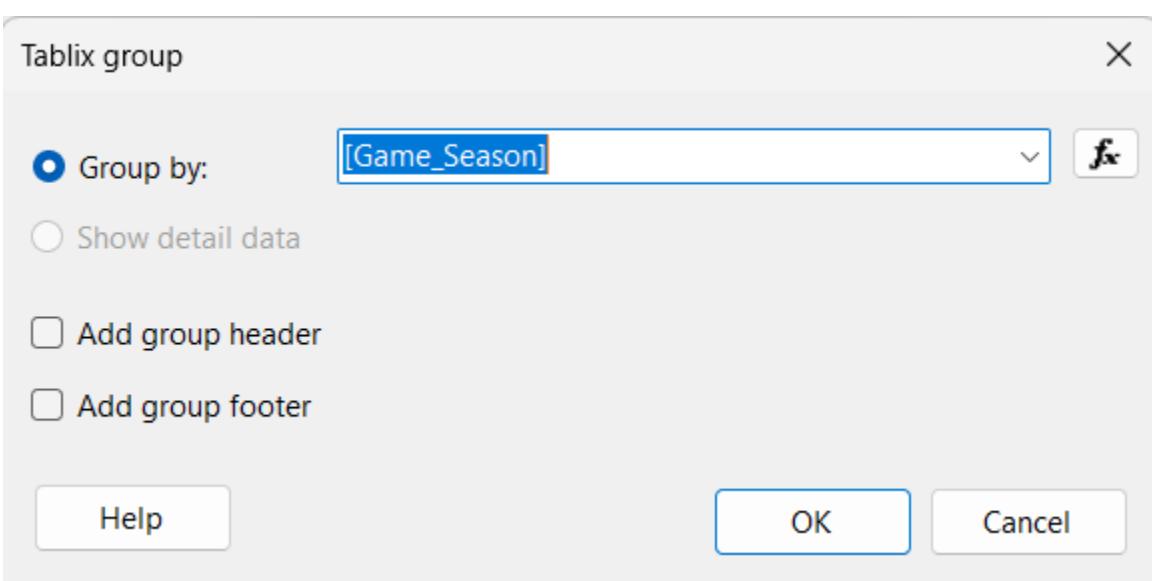
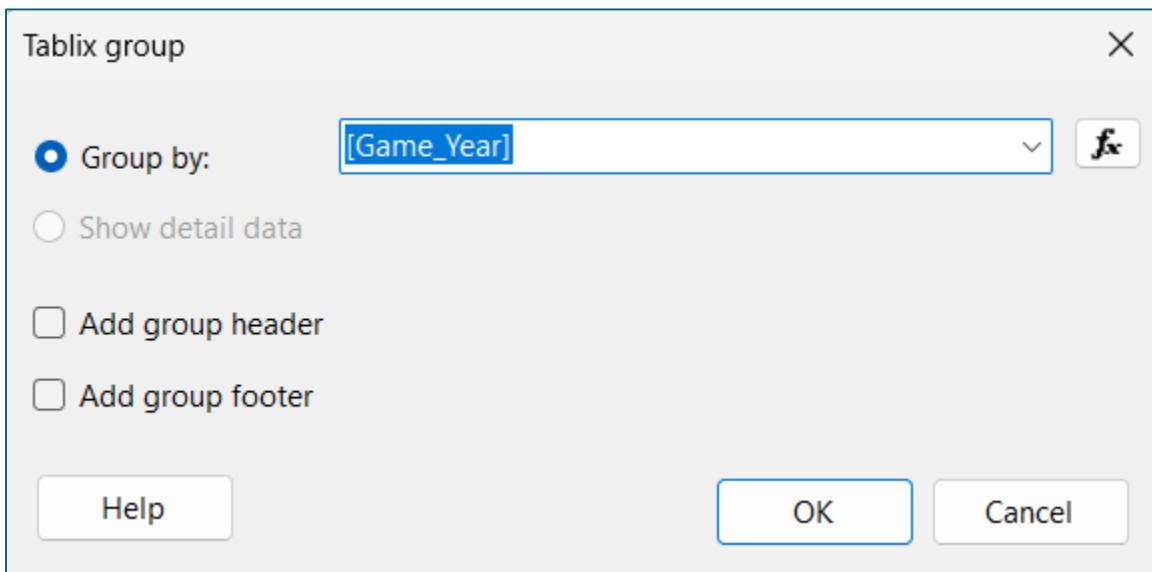
- Bước 5: Qua trường **Toolbox** kéo table qua vùng thiết kế (Design), kéo những thuộc tính vào bảng table như hình minh họa:

The screenshot shows the Microsoft Visual Studio interface with the SSRS (Report Designer) open. On the left, the Solution Explorer displays a project named 'Project_SSRS' with a single item 'DataSource1'. Below it, the Datasets folder contains 'DataSet_Query1' which includes four tables: 'Game_Year', 'Game_Season', 'Medal_Type', and 'Fact_Olympic_Count'. The main workspace shows a report titled 'THỐNG KÊ THÀNH TÍCH CỦA ĐỘI TUYỀN MỸ (2000-2020)'. Inside the report, there is a table with four columns: 'Năm thi đấu', 'Mùa thi đấu', 'Loại huy chương', and 'Số lượng'. The table rows are bound to the fields: [Game_Year], [Game_Season], [Medal_Type], and [Fact_Olympic_Count]. The Properties pane on the right shows the 'Body' border is set to 'Black, Black, Black, Black' with a border width of 1pt. The Error List pane at the bottom indicates 0 errors, 0 warnings, and 0 messages.

- Bước 6: Tùy chọn Grouping bằng các bước sau: Nhấn chuột phải vào **Detail**
→ **Add Group → Parent Group**

The screenshot shows the Microsoft Visual Studio interface with the SSRS report designer open. The 'Row Groups' pane on the left shows a single group named '(Details)'. A context menu is open over this group, with the 'Add Group' option highlighted. The submenu includes 'Parent Group...', 'Child Group...', 'Adjacent Before...', and 'Adjacent After...'. The main workspace shows the report structure with the newly added group.

- Bước 7: Grouping các giá trị **Game_year**, **Game_season**



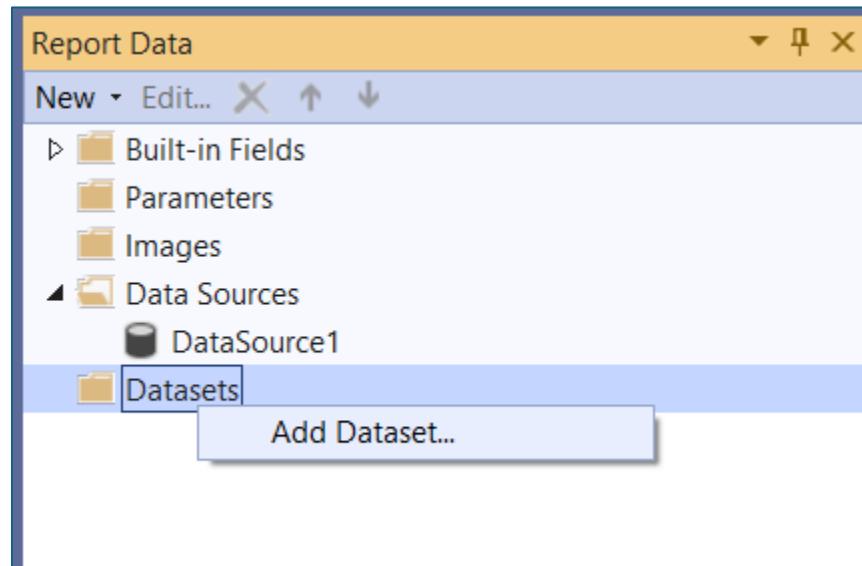
- Bước 8: In báo cáo ra file pdf bằng cách qua tab review → **save file with pdf**

THỐNG KÊ THÀNH TÍCH CỦA ĐỘI TUYỀN MỸ (2000-2020)

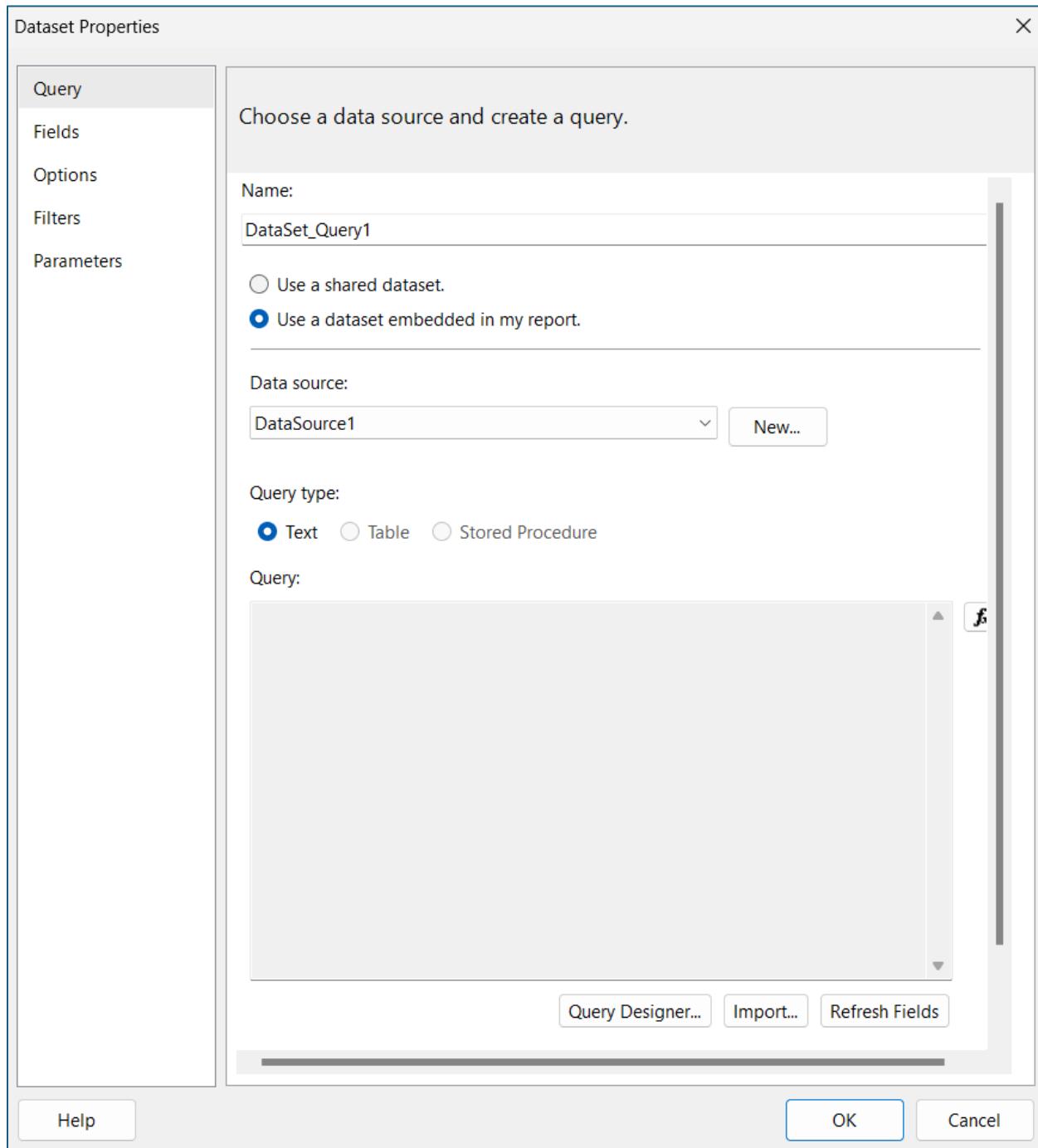
Năm thi đấu	Mùa thi đấu	Loại huy chương	Số lượng
2000	Summer	BRONZE	19
		GOLD	22
		SILVER	17
2002	Winter	BRONZE	9
		GOLD	8
		SILVER	9
2004	Summer	BRONZE	16
		GOLD	23
		SILVER	32
2006	Winter	BRONZE	4
		GOLD	9
		SILVER	7
2008	Summer	BRONZE	24
		GOLD	22
		SILVER	29
2010	Winter	BRONZE	10
		GOLD	8

1.3.2. Báo biểu 2: Thống kê top 10 vdv thành công nhất lịch sử Olympic

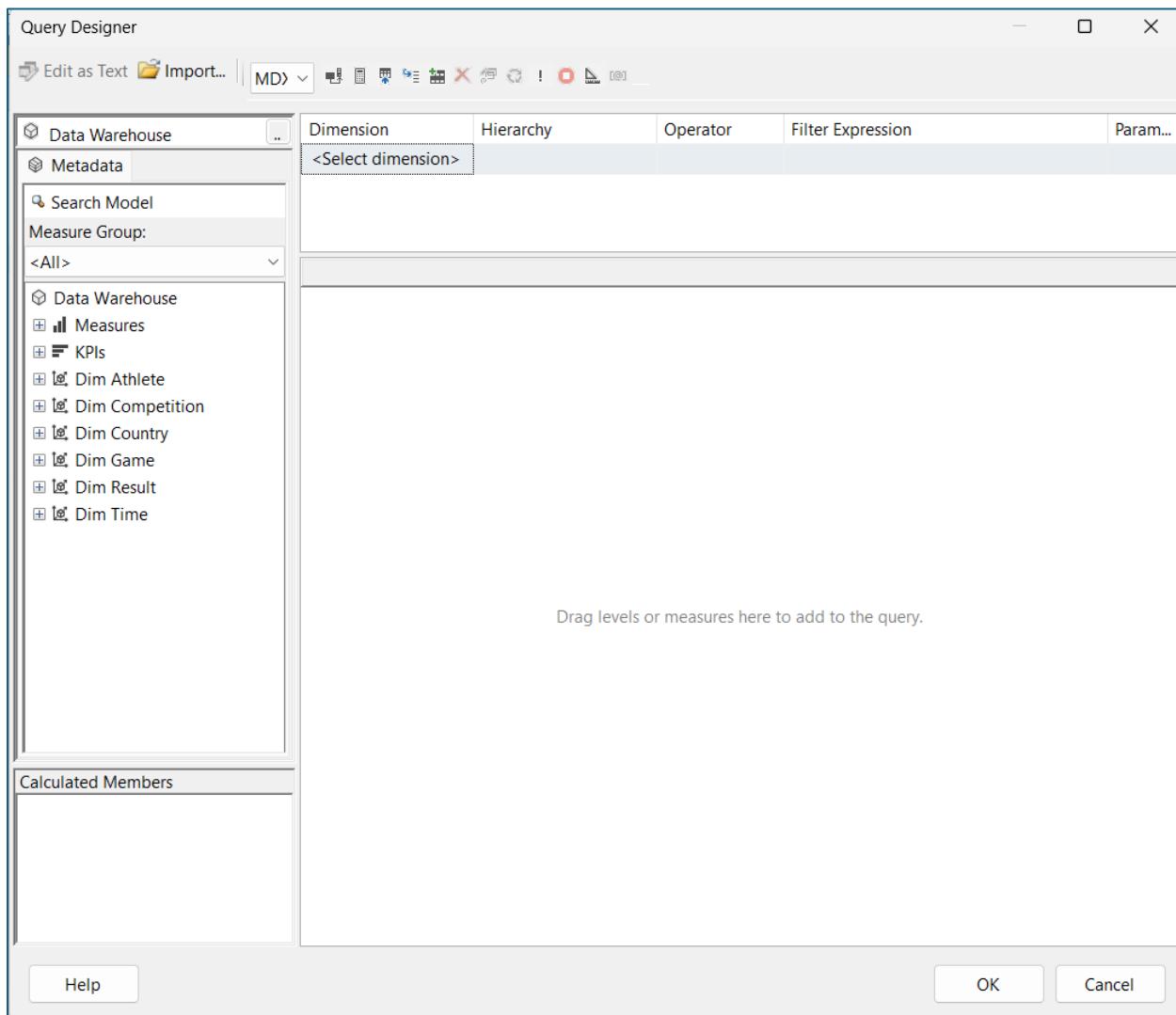
- Bước 1: Thêm datasets cần thiết cho báo biểu 1 đặt tên là datasets_query2, bằng cách nhấn chuột phải vào **Datasets** → **Add new dataset...**



- Bước 2: Trong tab Datasets Properties → Trường Query. Tại trường **Name** đặt tên là **Dataset_Query2** chọn chế độ **Use a dataset embedded in my report**. Trường Data Source chọn data source tại phần tạo kết nối **data source**.



- Bước 3: Chọn **Query Designer** hộp thoại sẽ hiện ra:



- Bước 4: Chọn những thuộc tính cần thiết cho báo biểu như hình họa giống như chọn trong quá trình SSAS bằng công cụ BI sau đó nhấn OK.

Query Designer

MD

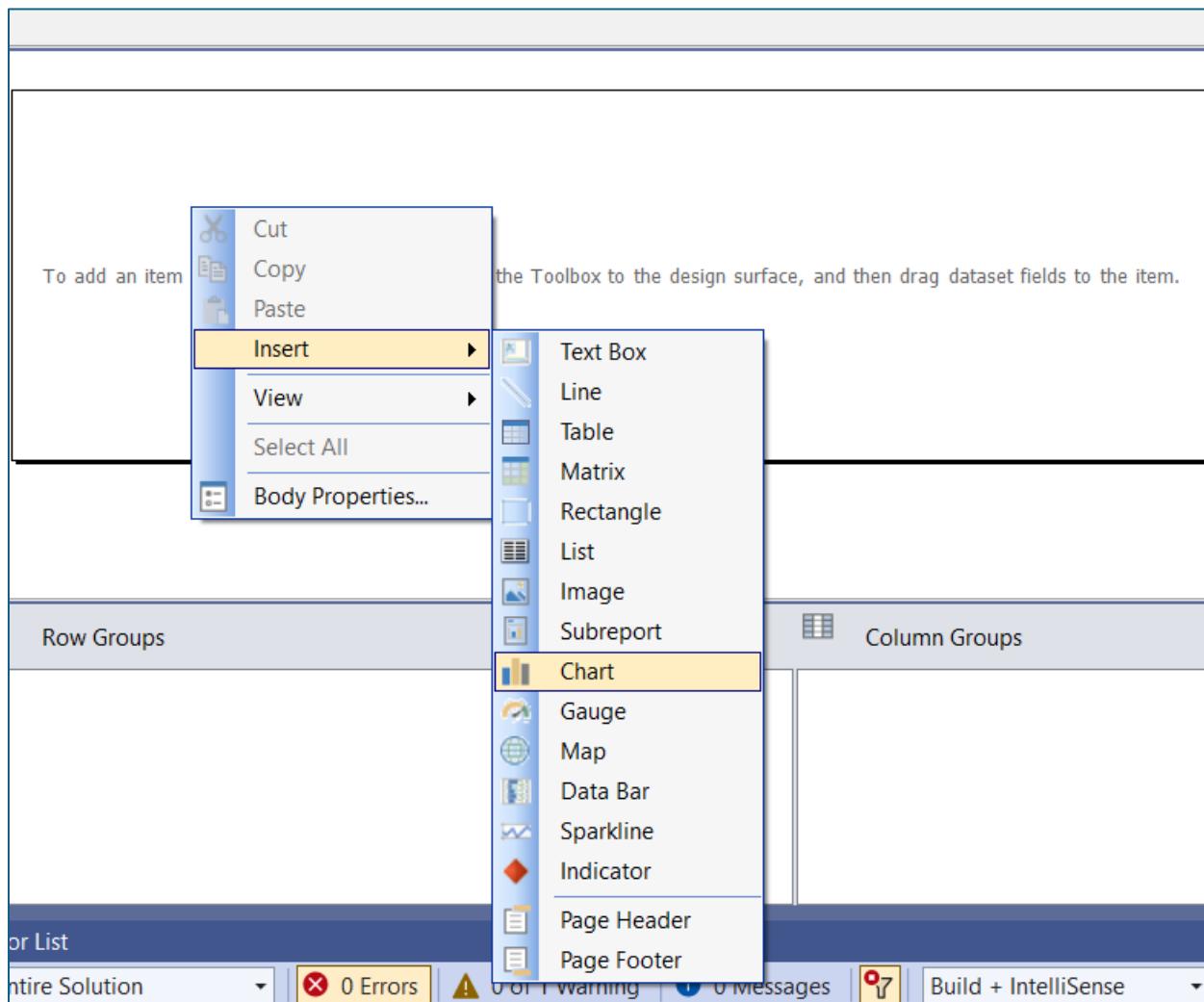
Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Result	Medal Type	Equal	{ GOLD, SILVER, BRONZE }	<input type="checkbox"/>
Dim Athlete	Athlete Full Name	In	TOP 10 VDV	
<Select dimension>				

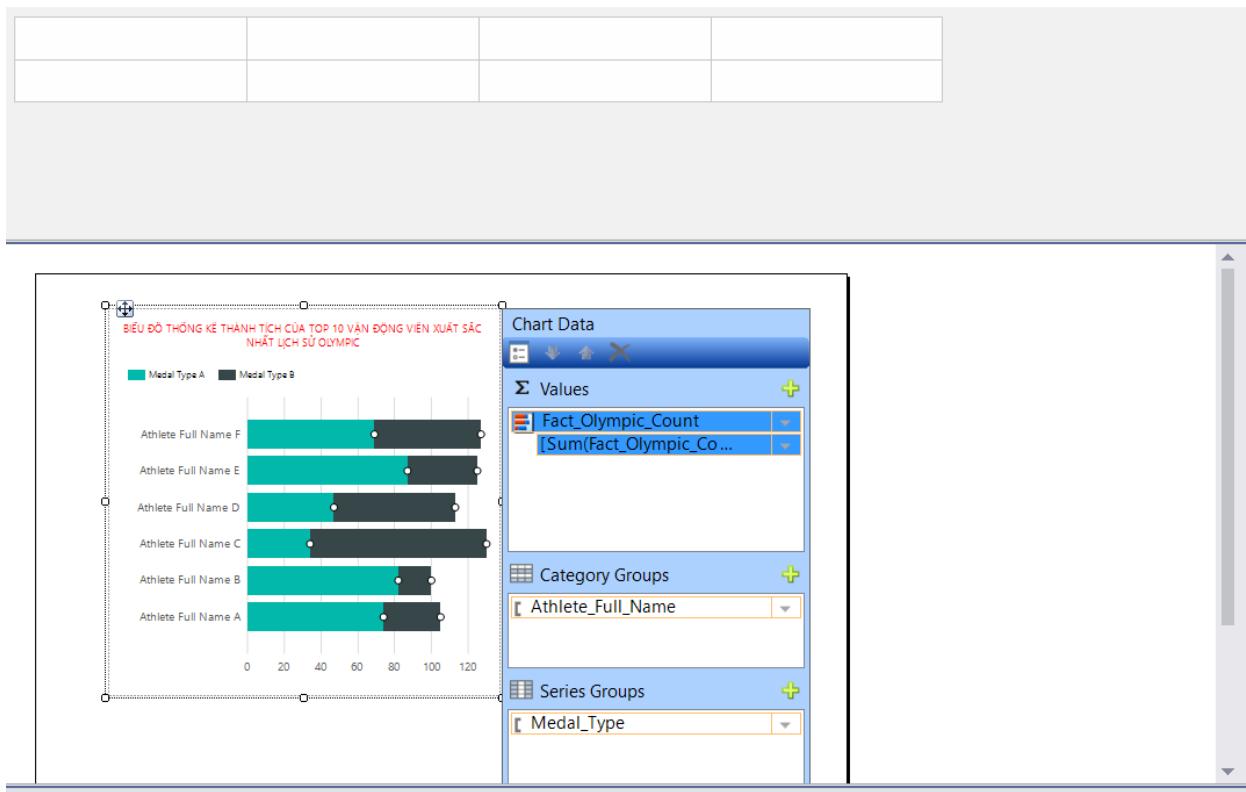
Athlete Full Name	Medal Type	Fact Olympic Count
Björn DAEHLIE	GOLD	6
Björn DAEHLIE	SILVER	3
Boris SHAKHLIN	BRONZE	2
Boris SHAKHLIN	GOLD	6
Boris SHAKHLIN	SILVER	2
Ireen WÜST	BRONZE	1
Ireen WÜST	GOLD	4
Ireen WÜST	SILVER	4
Larisa LATYNINA	BRONZE	3
Larisa LATYNINA	GOLD	6
Larisa LATYNINA	SILVER	5
Marit BJOERGEN	BRONZE	1
Marit BJOERGEN	GOLD	5
Marit BJOERGEN	SILVER	3
Michael PHELPS	BRONZE	1
Michael PHELPS	GOLD	13
Michael PHELPS	SILVER	1
Nikolay ANDRI...	BRONZE	3
Nikolay ANDRI...	GOLD	6
Nikolay ANDRI...	SILVER	3
Paavo NURMI	GOLD	6
Paavo NURMI	SILVER	3

Calculated Members

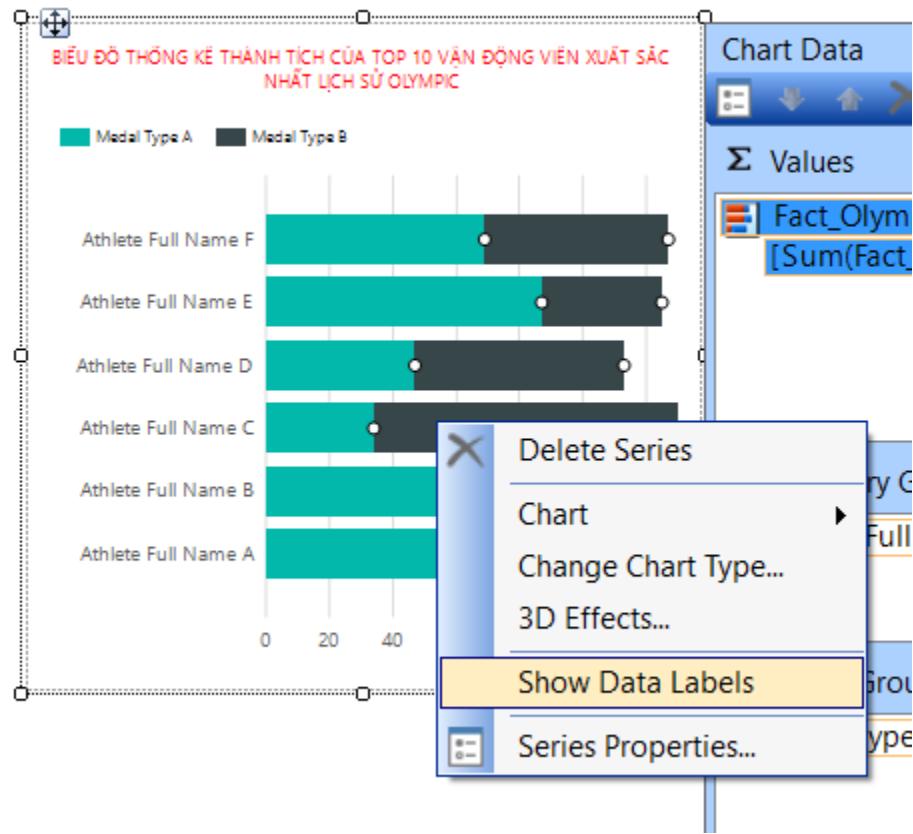
Help OK Cancel

- Bước 5: Qua trường **Toolbox** kéo table qua vùng thiết kế (Design), kéo những thuộc tính vào bảng chart như hình minh họa:

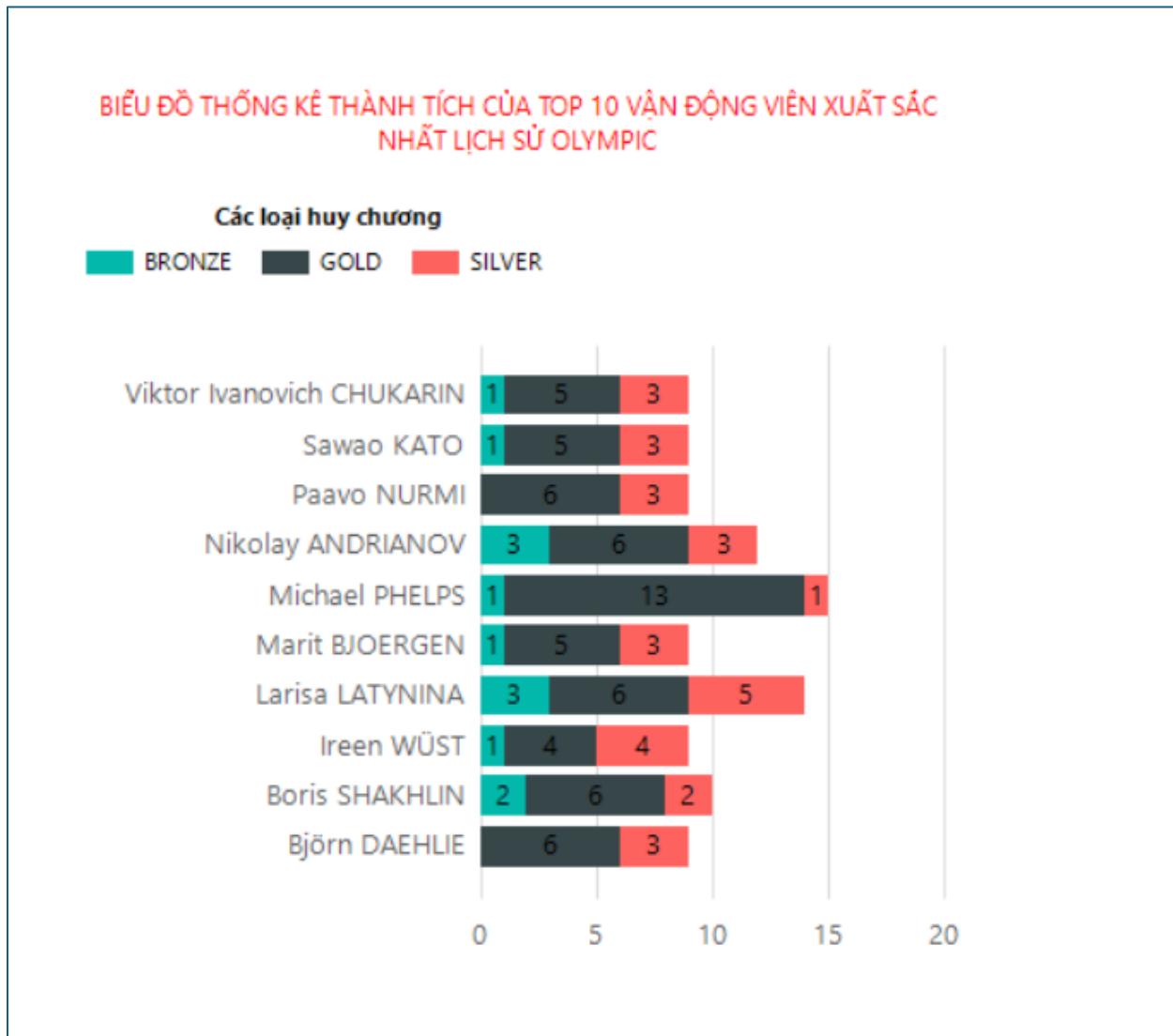




- Bước 6: Chọn **Show Data label** để thể hiện giá trị trên các cột

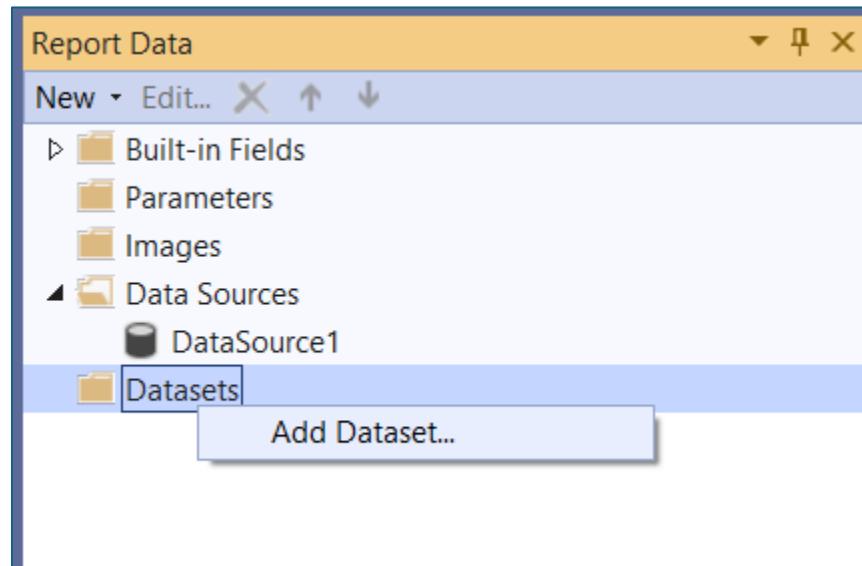


- Bước 7: In báo cáo ra file pdf bằng cách qua tab review → **save file with pdf**

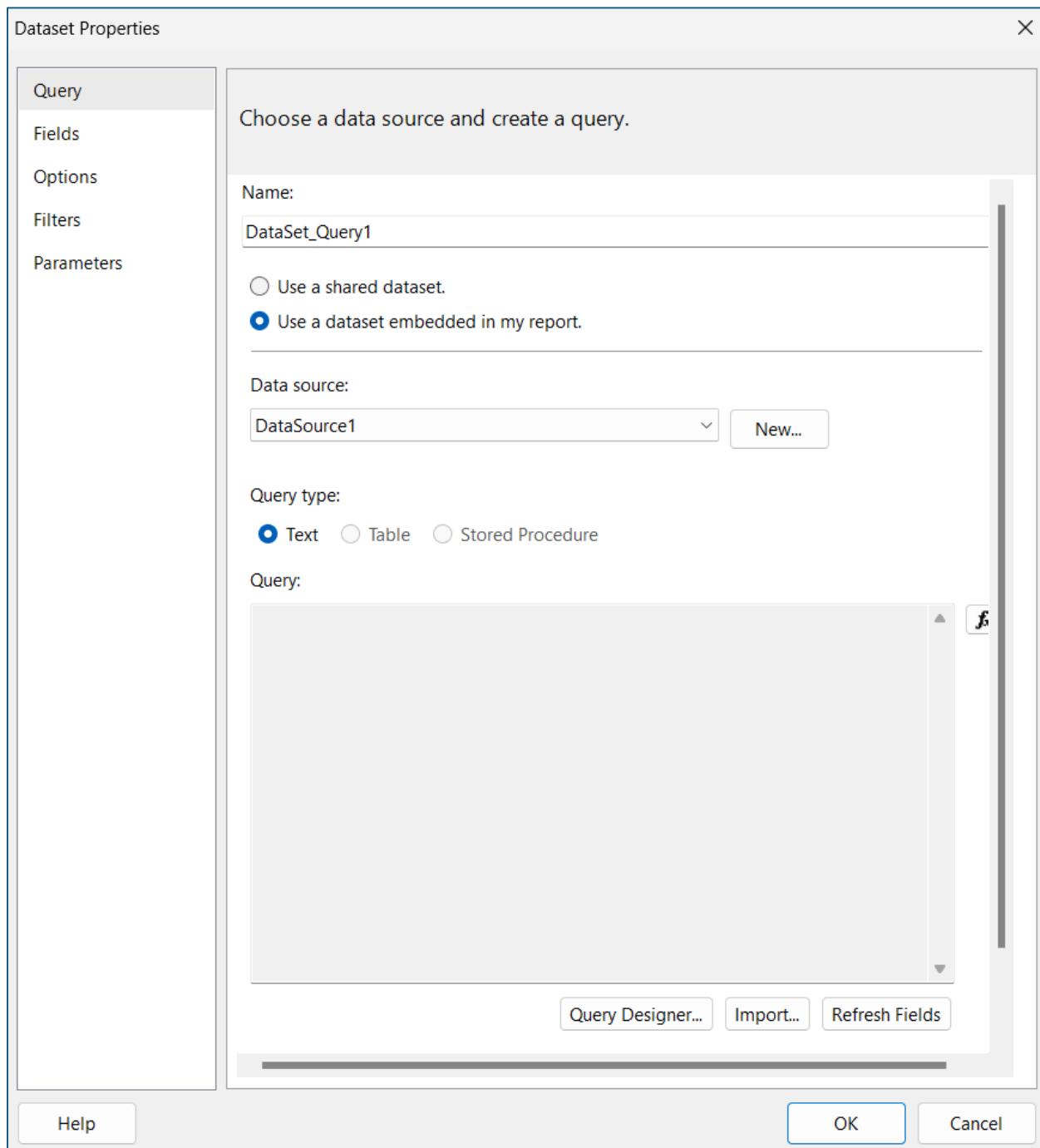


1.3.3. Báo biểu 3: Thống kê tham gia Olympic của VN

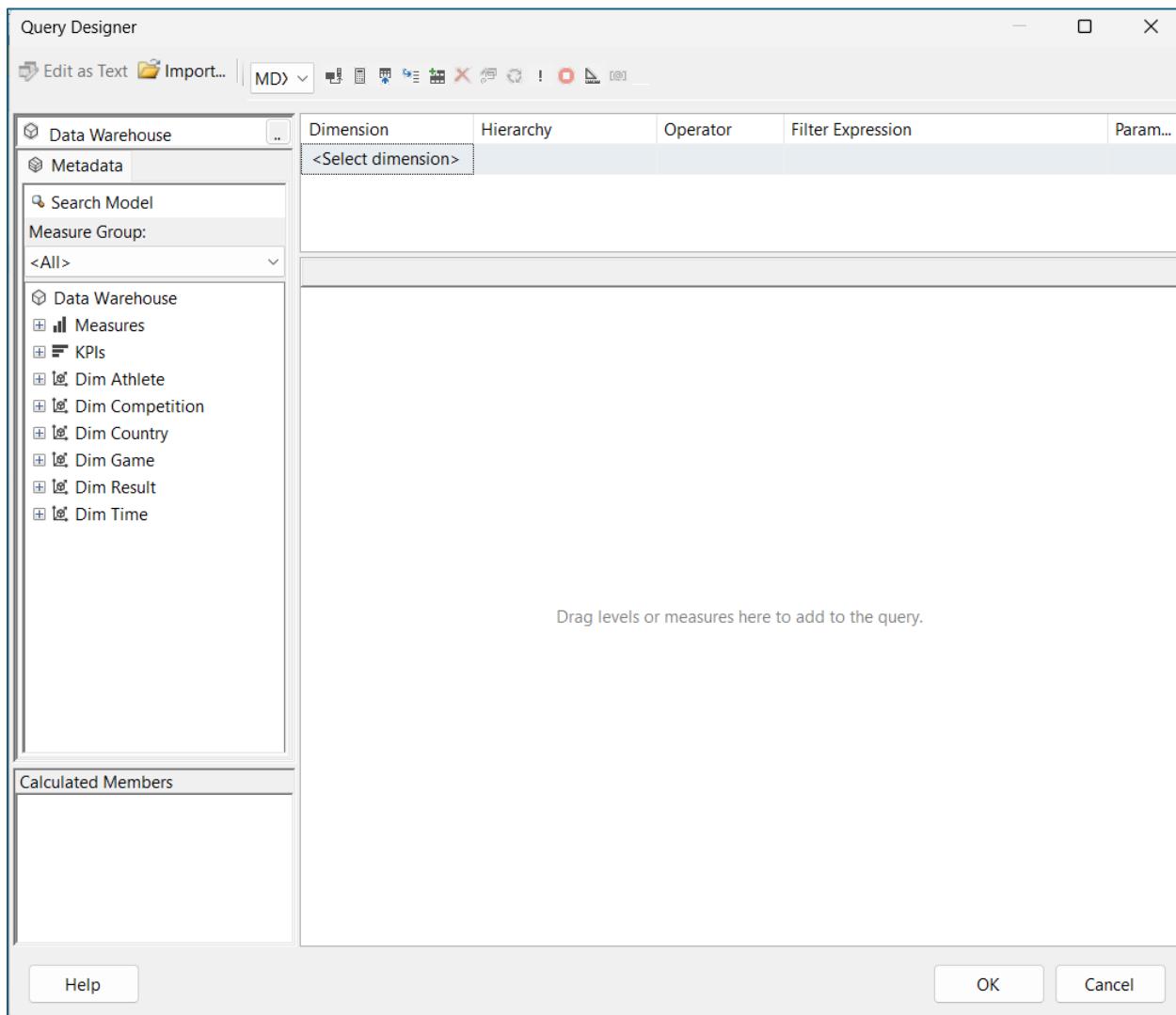
- Bước 1: Thêm datasets cần thiết cho báo biểu 1 đặt tên là datasets_query3, bằng cách nhấn chuột phải vào **Datasets** → **Add new dataset...**



- Bước 2: Trong tab Datasets Properties → Trường Query. Tại trường **Name** đặt tên là Dataset_Query3 chọn chế độ **Use a dataset embedded in my report**. Trường Data Source chọn data source tại phần tạo kết nối **data source**.



- Bước 3: Chọn **Query Designer** hộp thoại sẽ hiện ra:

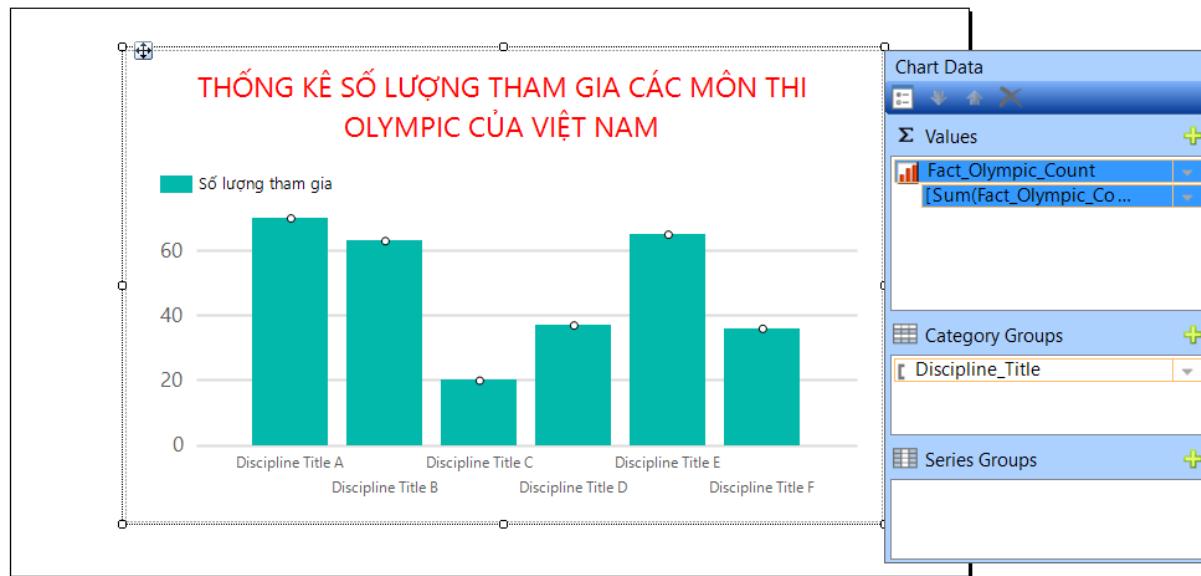


- Bước 4: Chọn những thuộc tính cần thiết cho báo biểu như hình họa giống như chọn trong quá trình SSAS bằng công cụ BI sau đó nhấn OK.

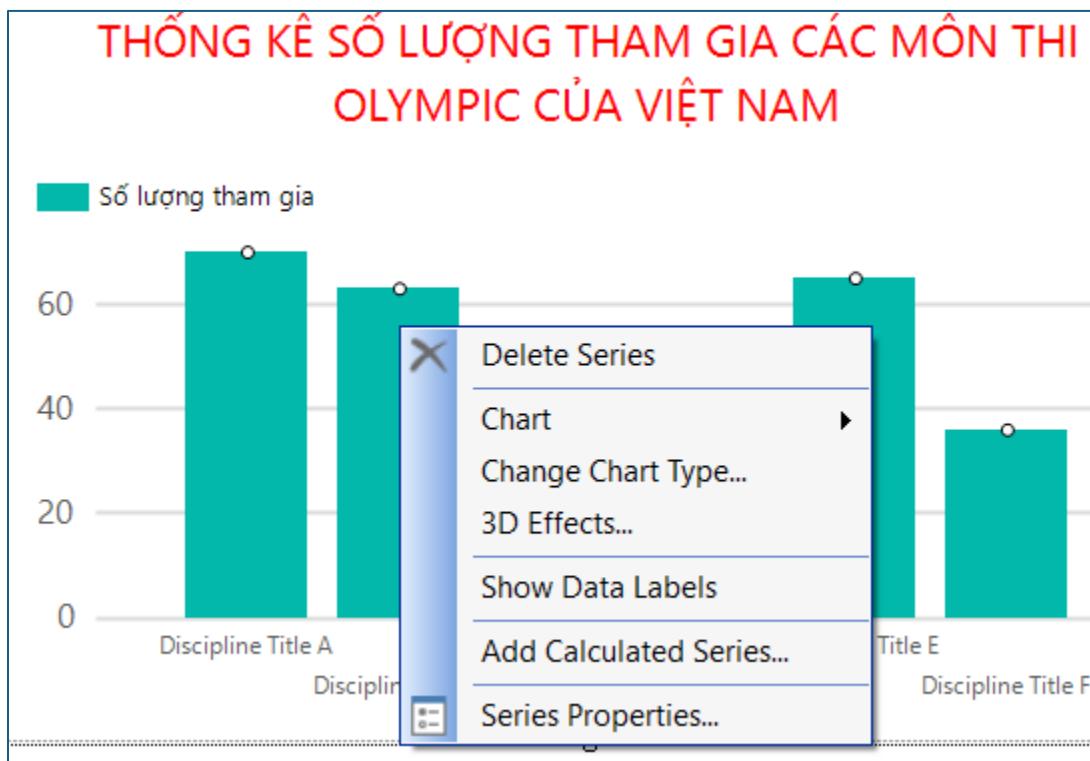
The screenshot shows the Microsoft Analysis Services Query Designer interface. On the left, there's a navigation pane with sections like Data Warehouse, Metadata, Search Model, and Measure Group. The Measure Group section is expanded, showing various dimensions and measures, with 'Fact Olympic Count' selected. In the main area, there's a table titled 'Fact Olympic Count' with columns 'Discipline Title' and 'Fact Olympic Count'. The table contains data for various sports. A filter is applied to the 'Dim Country' dimension, set to 'Country Name' with the operator 'Equal' and the value '{ Vietnam }'. The bottom right of the window has 'OK' and 'Cancel' buttons.

Discipline Title	Fact Olympic Count
Archery	2
Athletics	4
Badminton	6
Boxing	2
Cycling Road	1
Fencing	5
Gymnastics Art...	6
Judo	4
Shooting	23
Swimming	12
Table Tennis	2
Taekwondo	9
Weightlifting	10
Wrestling	8

- Bước 5: Qua trường **Toolbox** kéo table qua vùng thiết kế (Design), kéo những thuộc tính vào bảng chart như hình minh họa:

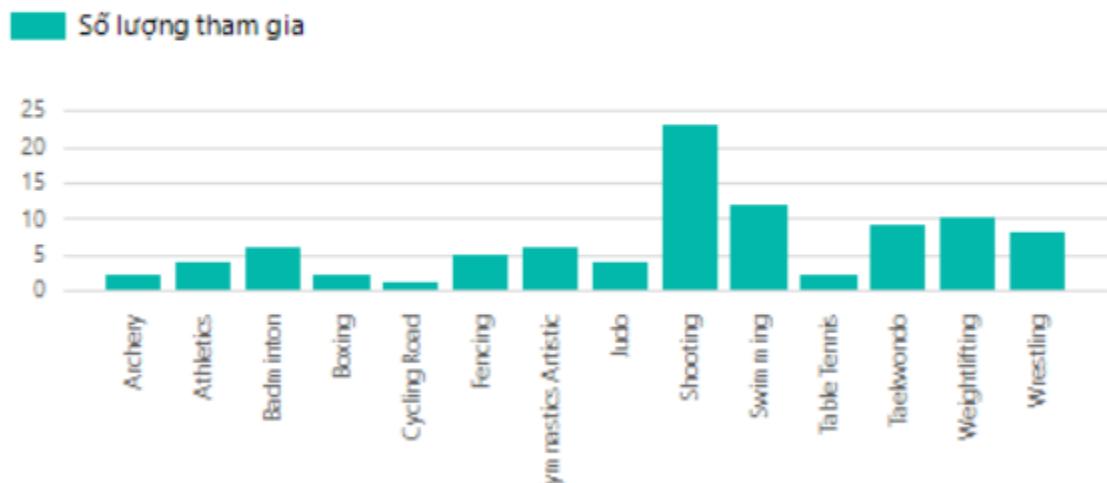


- Bước 6: Chọn **Show Data label** để thể hiện giá trị trên các cột



- Bước 7: In báo cáo ra file pdf bằng cách qua tab review → **save file with pdf**

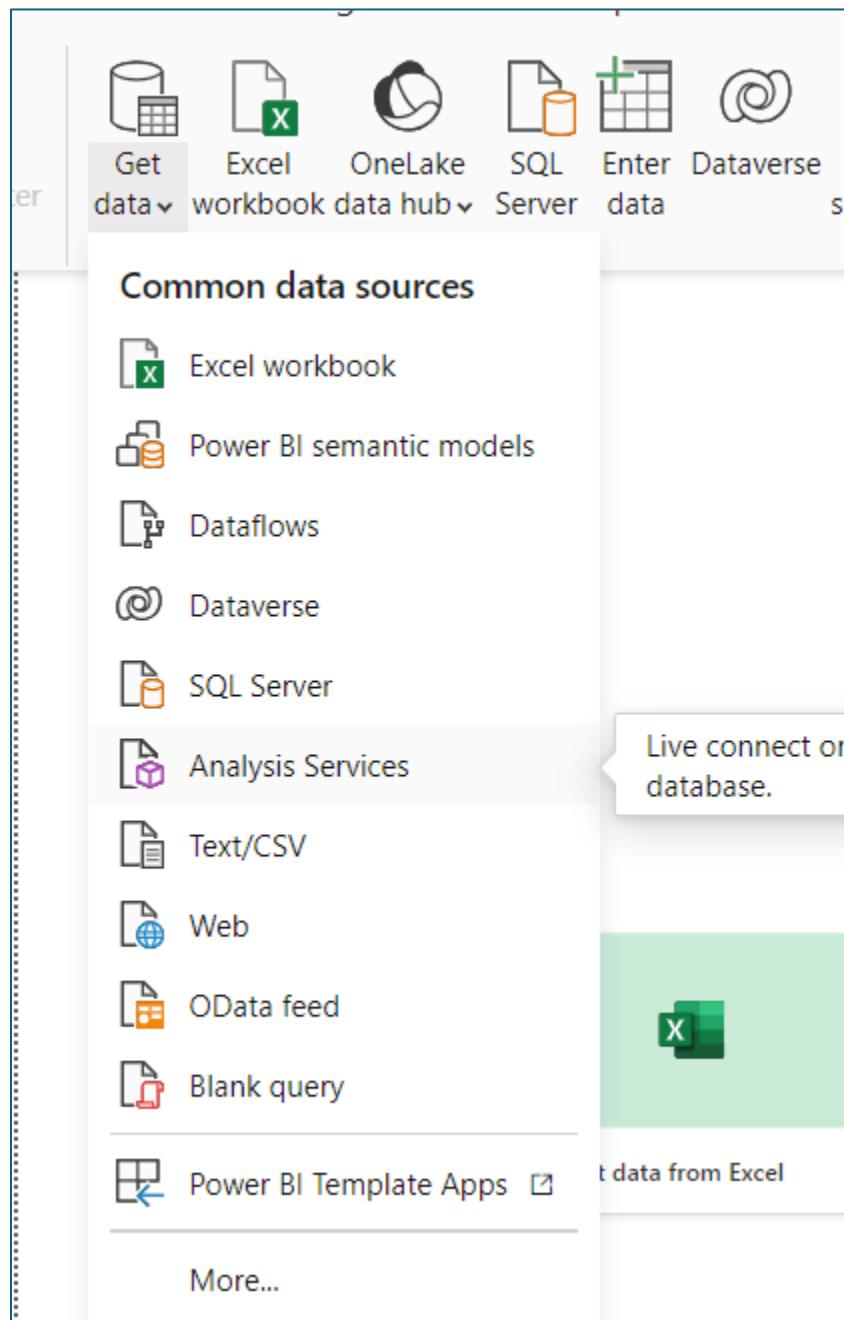
THỐNG KÊ SỐ LƯỢNG THAM GIA CÁC MÔN THI OLYMPIC CỦA VIỆT NAM



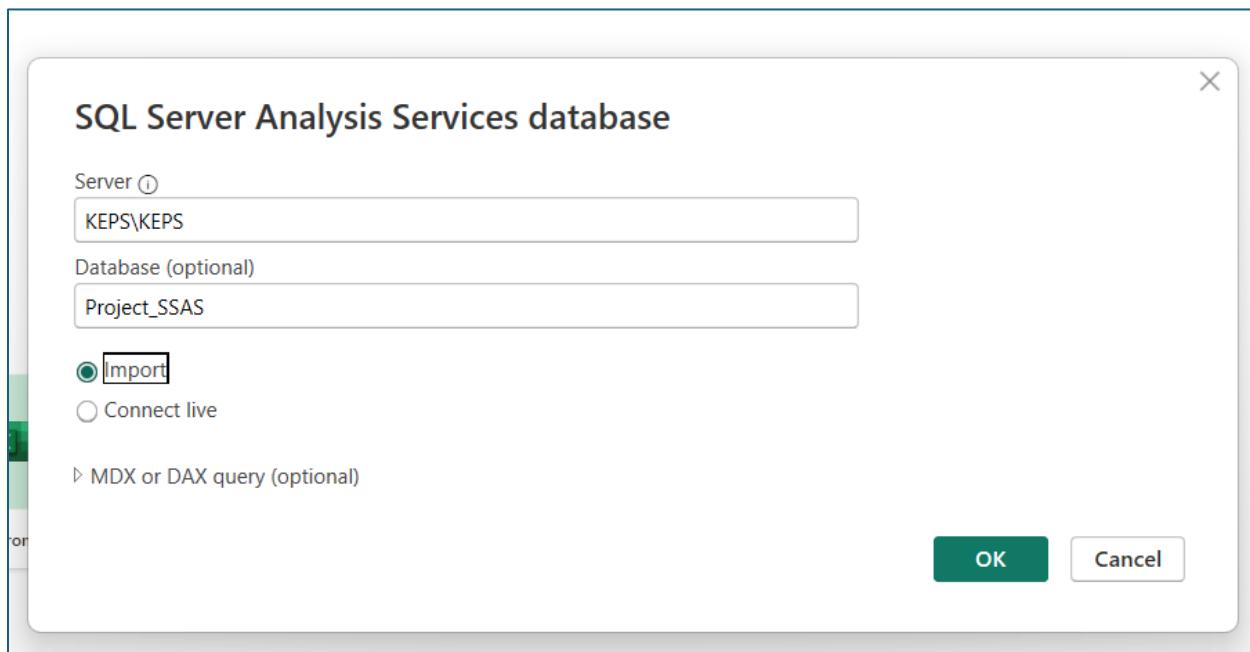
2. QUÁ TRÌNH LẬP BÁO BIỂU BẰNG CÔNG CỤ POWER BI

2.1. Thiết lập kết nối dữ liệu với Power BI

- Bước 1: tạo mới report
- Bước 2: Chọn Get Data → Analysis Services



- Bước 3:



- Bước 4: Cửa sổ lựa chọn dữ liệu để import hiện ra, chọn toàn bộ các dữ liệu của cơ sở dữ liệu SQL Analysis

Navigator

Display Options ▾

- ◀ KEPS\KEPS: Project_SSAS [1]
- ◀ Data Warehouse [1]
- ◀ Data Warehouse [7]
- ◀ Fact Olympic [1]
- Fact Olympic Count
- Dim Athlete [3]
- Dim Competition [4]
- Dim Country
- Dim Game
- Dim Result
- Dim Time

Data Warehouse

Dim Athlete.Athlete Full Name	Dim Athlete.Athlete Id	Dim Athlete.Athlete Year Birth
. DENI	1	1989
. PRIYANKA	2	1996
. RAHUL	3	1996
A Baser WASIQI	4	1975
A J HURT	5	2000
A J HURT	5	2000
A. DARNIS	6	
A. Germaine GOLDING	7	1887
A. TURNOVSKY	8	
A.J. MILLER	9	1899
Aage Avaldorff MEYER	10	1904
Aage Avaldorff MEYER	10	1904
Aage Avaldorff MEYER	10	1904
Aage Ingvar ERIKSEN	11	1917
Aage Ingvar ERIKSEN	11	1917
Aage JUSTESEN	12	1918
Aage JUSTESEN	12	1918
Aage MYHRVOLD	13	1918
Aage RASMUSSEN	14	1889

i The data in the preview has been truncated due to size limits.

Load
Transform Data
Cancel

- Bước 5: Hoàn tất nhập dữ liệu

Data »

🔍

✓ Data Warehouse ...

- Dim Athlete.Athlete Full Name
- Dim Athlete.Athlete Id
- Dim Athlete.Athlete Year Birth
- Dim Competition.Competition Id
- Dim Competition.Discipline Title
- Dim Competition.Event Title
- Dim Competition.Participant Type
- Dim Country.Country 3 Letter Code
- Dim Country.Country Code
- Dim Country.Country Id
- Dim Country.Country Name ...
- Dim Game.Game Location
- Dim Game.Game Name
- Dim Game.Game Slug
- Dim Result.Medal Type
- Dim Result.Rank Equal
- Dim Result.Rank Position
- Dim Result.Result Id
- Dim Time.Game Season
- Dim Time.Game Season.1
- Dim Time.Game Start Date

2.2. Lập báo biểu

2.2.1. Báo biểu 1: Thống kê thành tích của Michael Phelps

Tạo 3 biểu đồ:

- Thống kê các loại huân chương đạt được

The screenshot shows the 'Filters on this visual' section of a Power BI visualization. It includes a search bar at the top, followed by two filter cards: one for 'Amount' (is (All)) and one for 'Medal' (is BRONZE, GOLD, or SILVER). Below these is a placeholder 'Add data fields here'. At the bottom, there is a 'Filters on this page' section containing two filter cards: one for 'Dim Athlete.Athlete Full Name' (is Michael PHELPS) with a lock icon, and another for 'Dim Game.Game Name' (is (All)). There is also a placeholder 'Add data fields here' at the bottom of this section.

- Thống kê thành tích theo các năm

The screenshot shows the 'Filters' pane in Power BI. It is divided into two main sections: 'Filters on this visual' at the top and 'Filters on this page' at the bottom.

Filters on this visual:

- Amount: is (All)
- Medal:** is BRONZE, GOLD, or SILVER
- Year: is (All)
- Add data fields here

Filters on this page:

- Dim Athlete.Athlete Full Name:** is Michael PHELPS (with a lock icon)
- Dim Game.Game Name: is (All)

- Thông kê các nội dung đạt được nhiều huân chương

Filters on this visual

Amount
is (All)

Event Title
is (All)

Medal
is BRONZE, GOLD, or SILVER

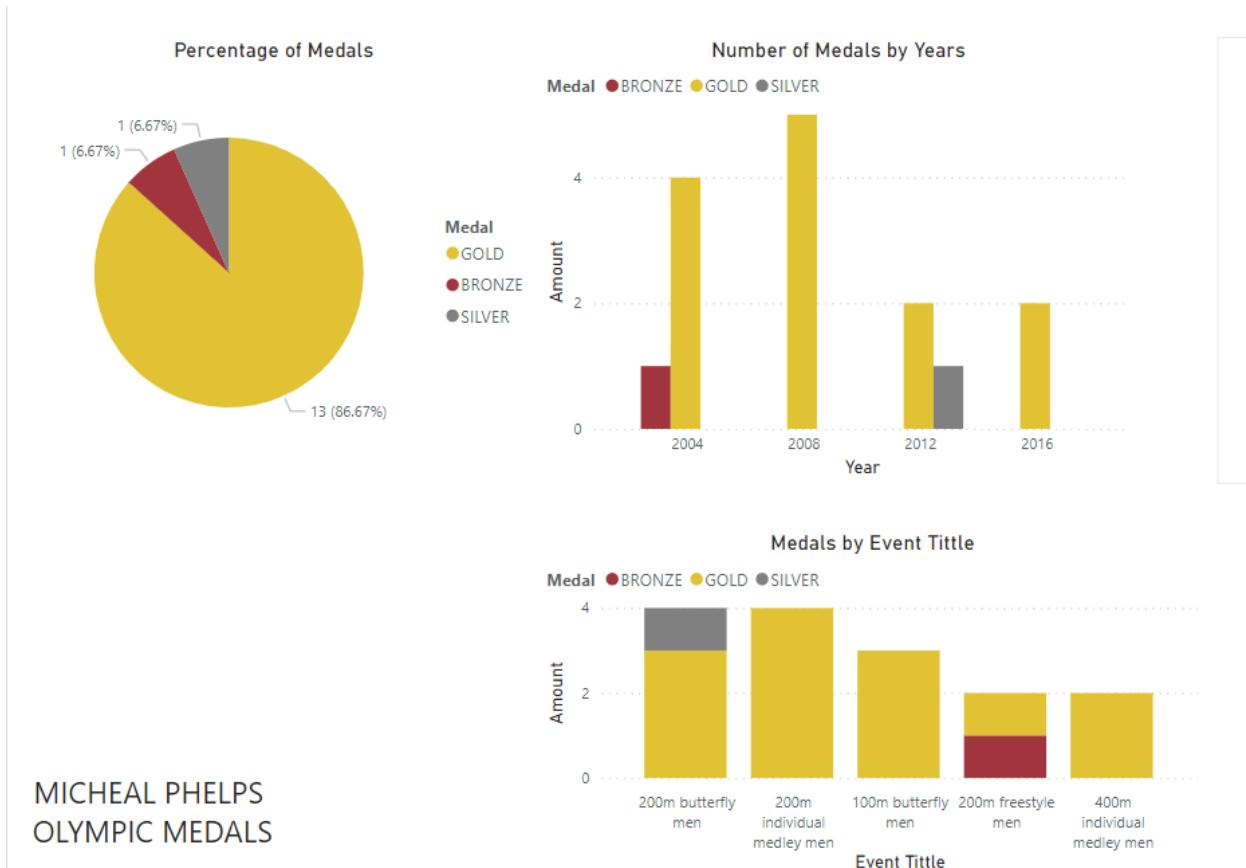
Add data fields here

Filters on this page

Dim Athlete.Athlete Full Name 
is Michael PHELPS

Dim Game.Game Name
is (All)

Kết quả:



2.2.2. Báo biểu 2: Thống kê số lượng huy chương của các quốc gia tham gia bộ môn Ski Jumping

Tạo 2 biểu đồ:

- Biểu đồ cột cho báo biểu

The screenshot shows the Power BI interface's left-hand pane, which includes the 'Filters' and 'Visualizations' sections.

Filters Section:

- Search:** A search bar at the top.
- Filters on this visual:**
 - Dim Competition....** is Ski Jumping
 - Dim Game.Game Na...** is Sochi 2014
 - Dim Country.Country ...** is (All)
 - Sum of Fact Olympic C...** is (All)
- Add data fields here**

Filters on this page:

- Add data fields here**

Filters on all pages:

- Add data fields here**

Visualizations Section:

- Build visual:** A section with three icons: a grid, a pencil, and a magnifying glass.
- Visualizations Grid:** A large grid of 24 visualization icons, including various charts (bar, line, pie, treemap), maps, and other data representation types.
- X-axis:** A dropdown menu currently set to **Dim Country.Country ...**.
- Y-axis:** A dropdown menu currently set to **Sum of Fact Olympic ...**.
- Legend:** A dropdown menu currently set to **Dim Game.Game Name**.
- Small multiples:** A section with a dropdown menu labeled **Add data fields here**.
- Tooltips:** A section with a dropdown menu labeled **Add data fields here**.

- Biểu đồ địa lý cho báo biểu

The screenshot shows the Power BI interface with the 'Filters' pane on the left and the 'Visualizations' pane on the right.

Filters Pane:

- Search:** A search bar at the top.
- Filters on this visual:**
 - Dim Country.Country ... is (All)
 - Dim Competition.Disc... is Ski Jumping
 - Sum of Fact Olympic C... is (All)
 - Dim Game.Game ... is Sochi 2014
- Filter type:** Set to "Basic filtering".
- Search:** A search bar for filtering items.
- Items:**
 - Select all
 - Albertville 1992 (116)
 - Beijing 2022 (140)
 - Calgary 1988 (113)
 - Chamonix 1924 (22)
 - Cortina d'Ampezzo ... (51)
 - Garmisch-Partenkir... (47)
- Require single selection:** A checkbox at the bottom.

Visualizations Pane:

- Build visual:** A section with three icons: a grid, a chart, and a search.
- Visualizations:** A large grid of icons representing various chart types and data visualizations.
- Location:** A section showing filters for "Dim Country.Country ...".
- Legend:** A section showing filters for "Dim Competition.Disc...".
- Add data fields here:** Buttons for "Latitude", "Longitude", and "Bubble size".

Kết quả:



2.2.3. Báo biểu 3: Thống kê thành tích ở Olympic

- Biểu đồ số lượng tham gia kì Olympic

The screenshot shows the Power BI interface with the 'Visualizations' pane open. The pane contains a grid of icons representing different types of visualizations, such as bar charts, line graphs, pie charts, and maps. Below the grid, there are sections for 'X-axis', 'Y-axis', and 'Legend', each with dropdown menus and 'Add data fields here' buttons. On the left side of the pane, there are sections for 'Filters on this visual', 'Filters on this page', and 'Filters on all pages', each with 'Add data fields here' buttons.

- Biểu đồ số huân chương đạt được

Filters

Search:

Filters on this visual

Dim Country.Country
top 10 by Count of...

Filter type ⓘ: Top N

Show items: Top 10

By value: Count of Dim Cou...

Apply filter

Dim Result.Medal...
is BRONZE, GOLD, ...

Filter type ⓘ: Basic filtering

Search:

<input checked="" type="checkbox"/> Select all		115680
<input type="checkbox"/>		3790
<input checked="" type="checkbox"/> BRONZE		4263
<input checked="" type="checkbox"/> GOLD		4186
<input checked="" type="checkbox"/> SILVER		

Visualizations

Build visual

Icon: Bar chart

Icon: Line chart

Icon: Scatter plot

Icon: Heatmap

Icon: Map

Icon: Gantt chart

Icon: Timeline

Icon: Calendar

Icon: R

Icon: Python

Icon: Power BI

Icon: Chat

Icon: Document

Icon: Trophy

Icon: Data

Icon: Flash

Icon: Globe

Icon: Network

Icon: Share

Icon: Ellipsis

Y-axis

Dim Country.Country ...

X-axis

Sum of Fact Olympic ...

Legend

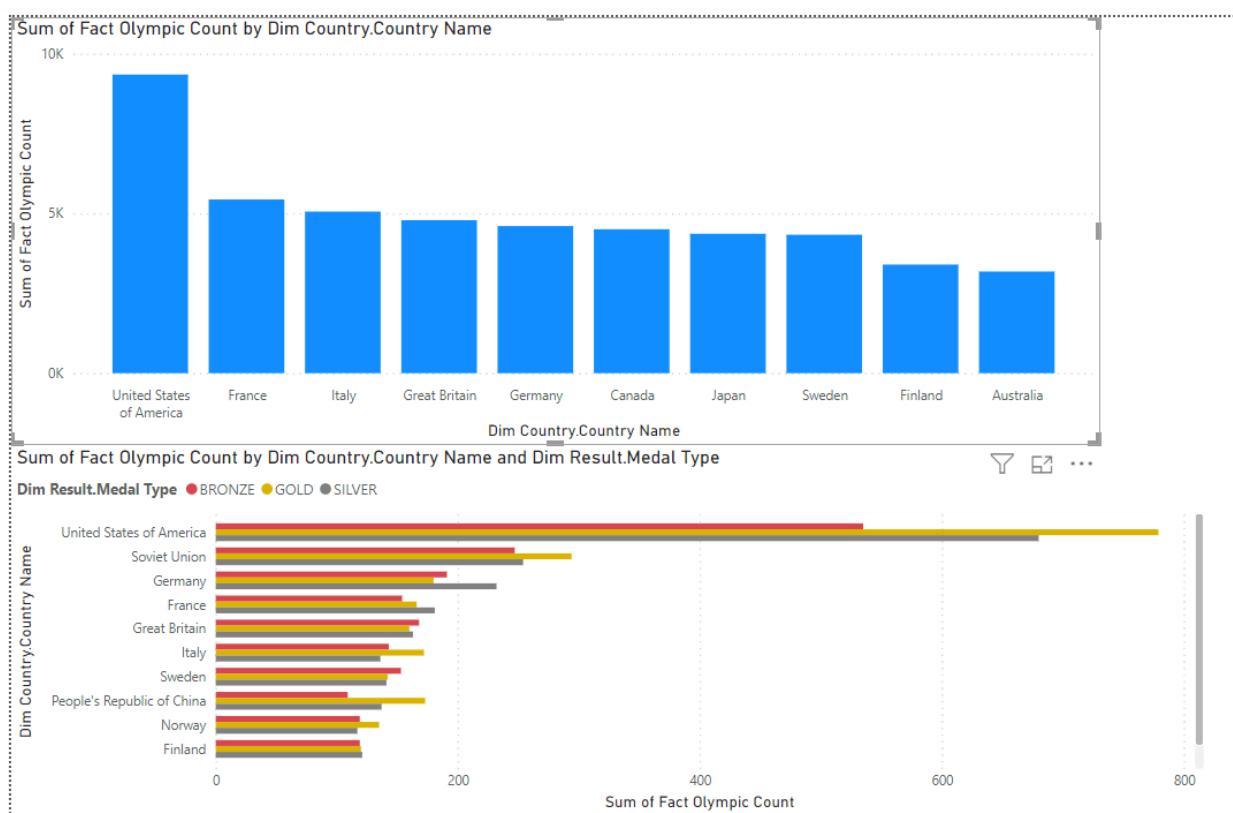
Dim Result.Medal Type

Small multiples

Add data fields here

Tooltips

Kết quả



CHƯƠNG 5: QUY TRÌNH KHAI THÁC DỮ LIỆU – DATA MINING

1. TỔNG QUAN

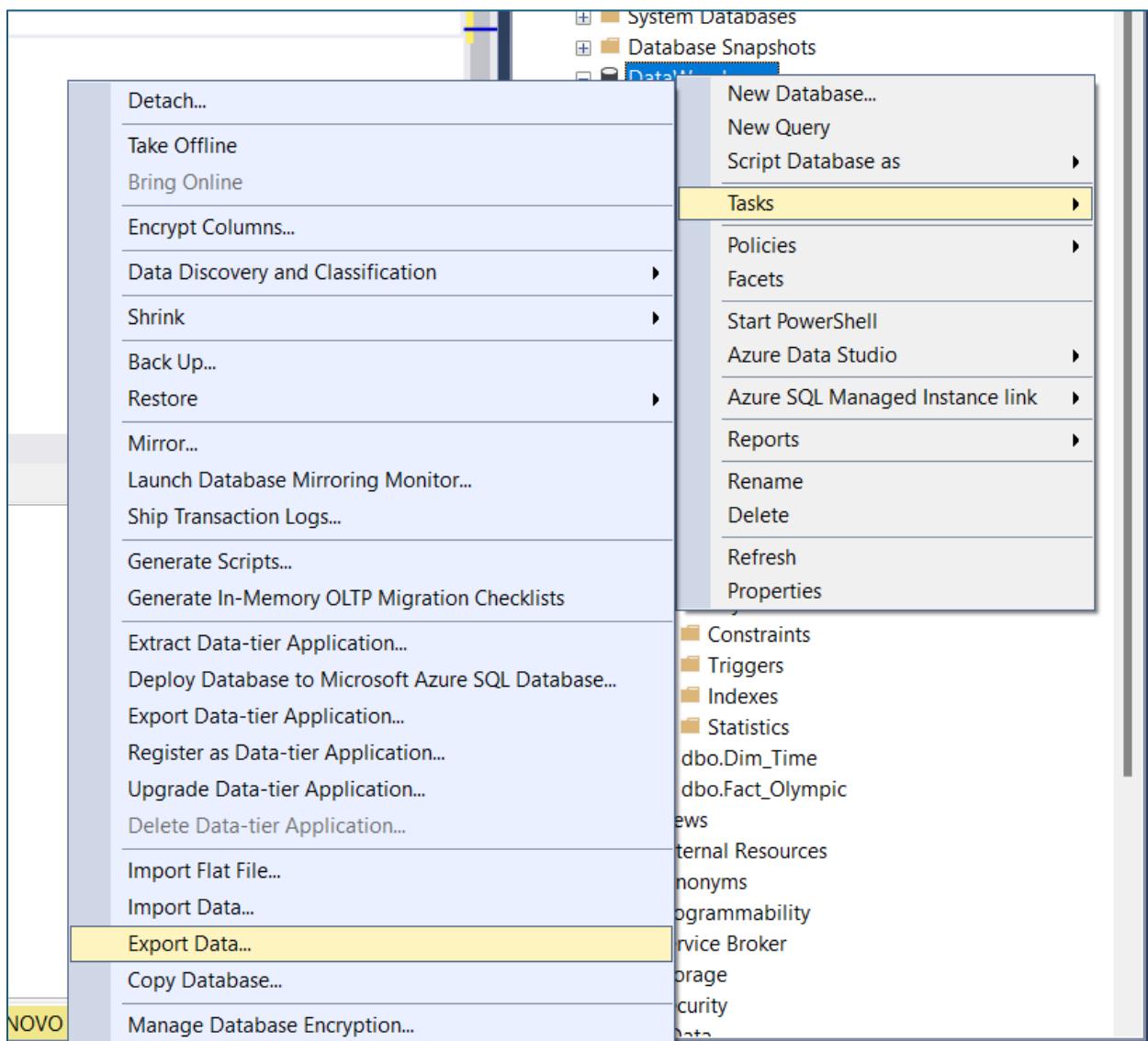
Chương này tập trung vào việc thực hiện quá trình Khai thác Dữ liệu (Data Mining) và dự báo vận động viên tham gia các kỳ Thế vận hội Olympics dựa trên năm diễn ra. Sau đó, các thuật toán Linear Regression và Long Short-Term Memory (LSTM) sẽ được đánh giá hiệu quả thông qua chỉ số RMSE (Root Mean Squared Error).

2. DỮ LIỆU

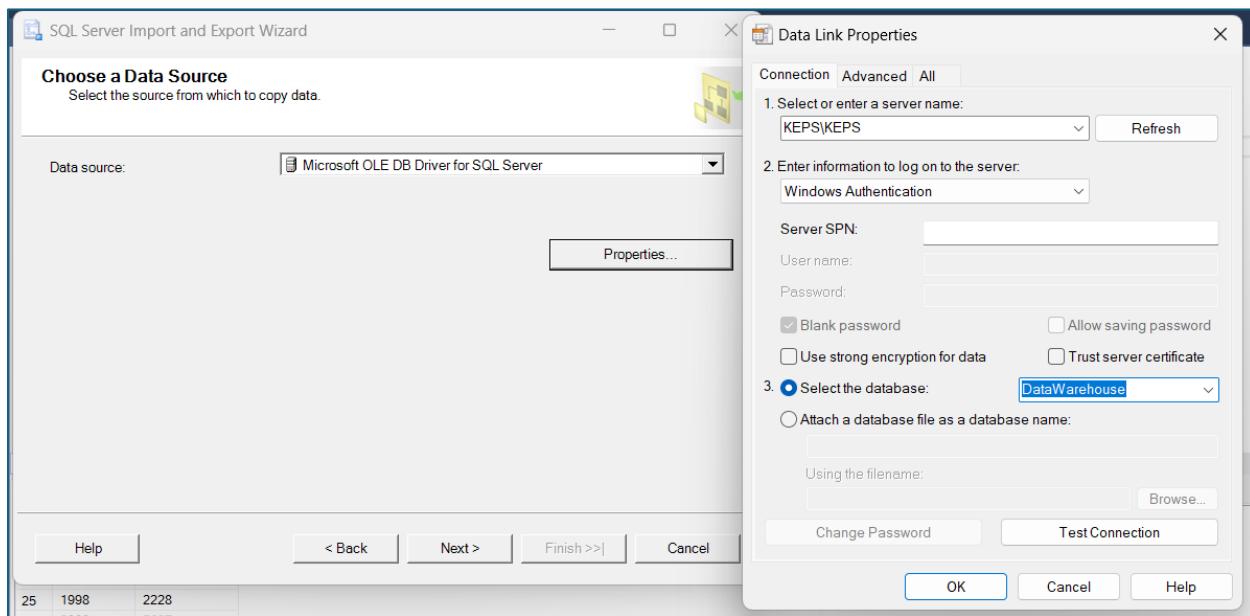
Dữ liệu được lấy bằng câu lệnh SQL sau:

```
select T1.game_year , Count(*) as 'SL VDV Tham Gia'  
from [dbo].[Dim_Time] T1, [dbo].[Fact_Olympic]  
T2,[dbo].[Dim_Competition] T3  
where T1.time_id = T2.time_id and T3.competition_id  
=T2.competition_id  
group by T1.game_year  
ORDER BY T1.game_year ASC
```

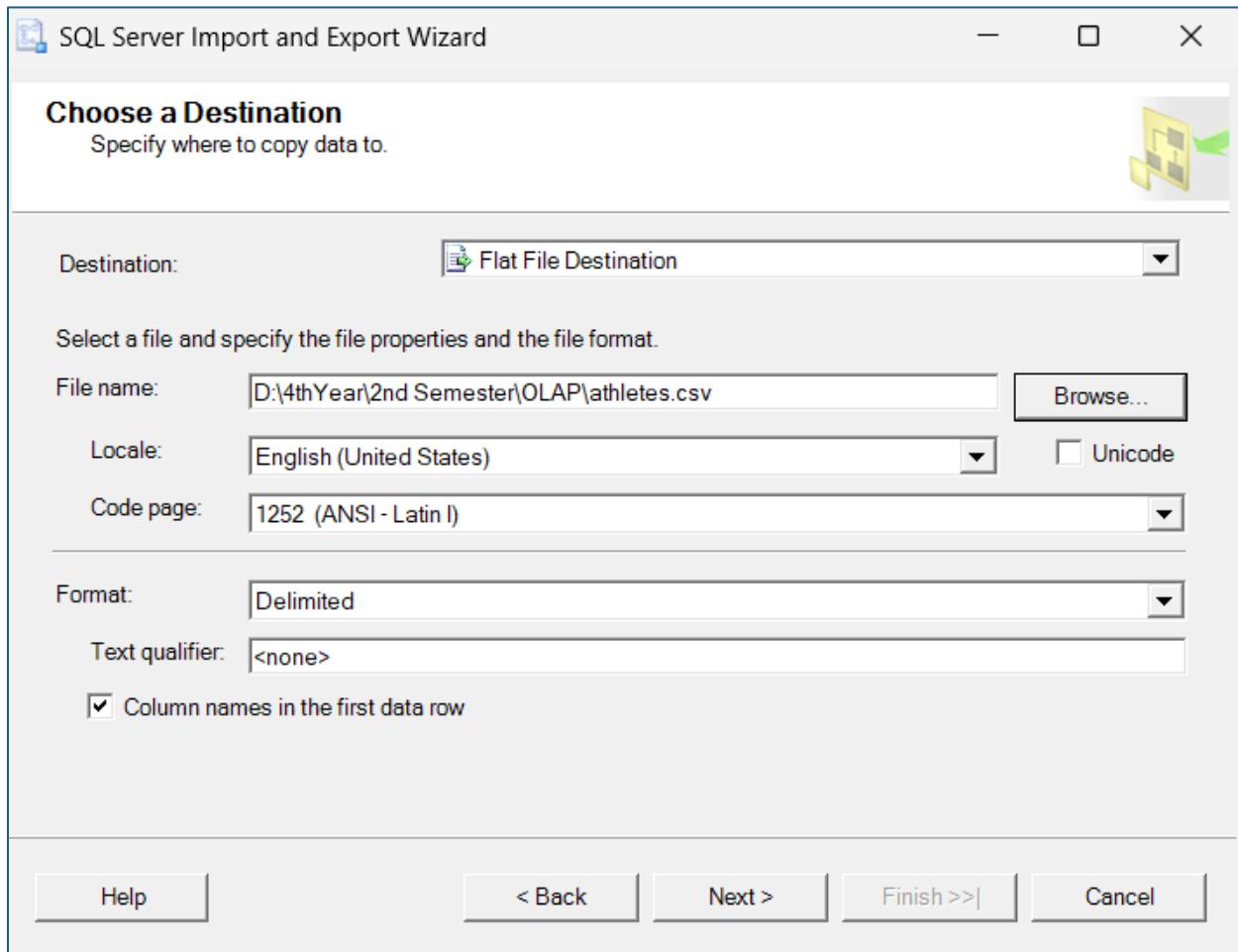
- **Bước 1:** Để có thể thực hiện quá trình khai thác dữ liệu, ta cần thực hiện xuất ra thành file CSV. Nhấn chuột phải vào database **DW_Olympics** → **Tasks > Export Data.**



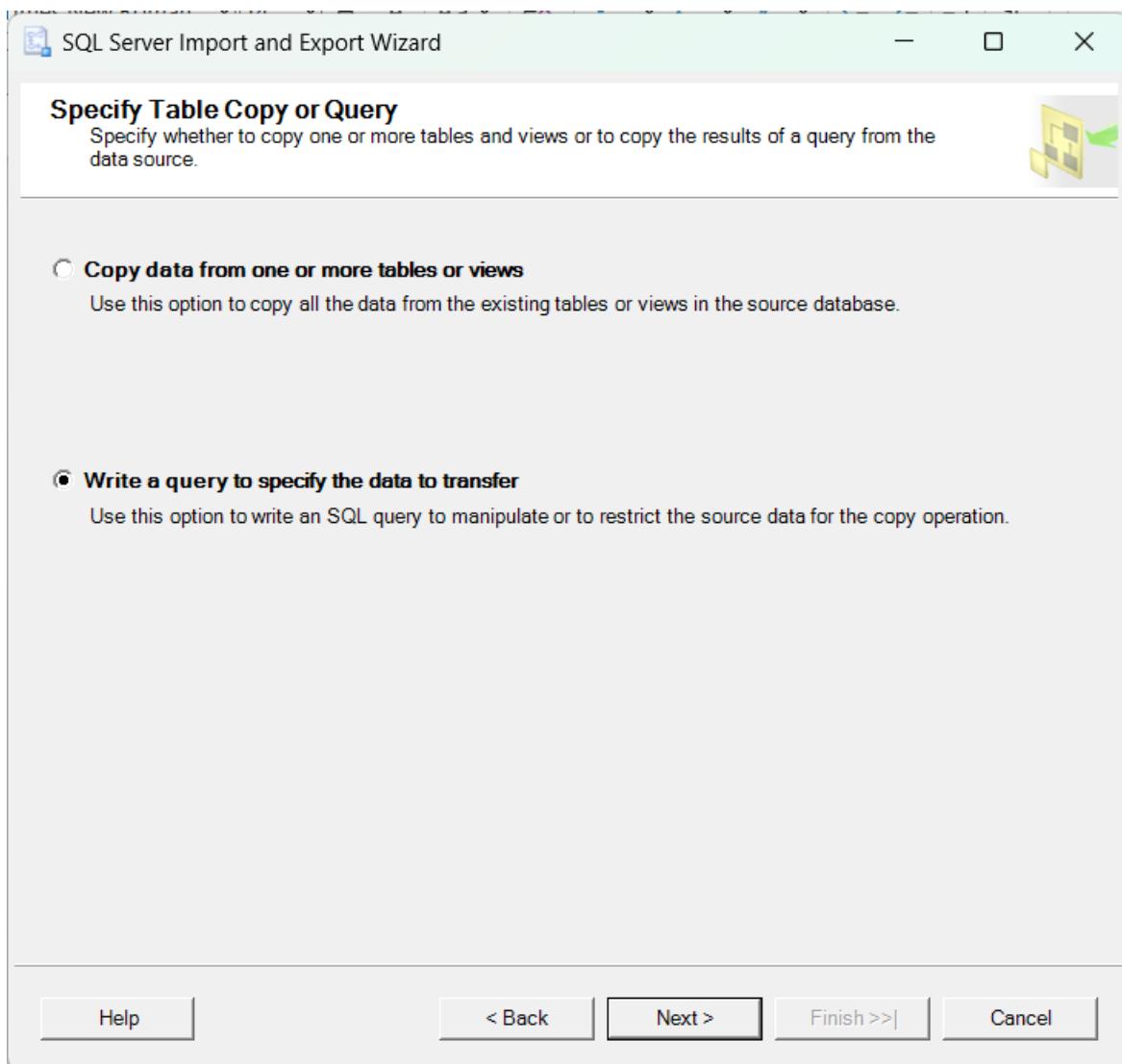
- **Bước 2:** Sau đó ta tiến hành chọn Data Source là **Microsoft OLE DB Provider for SQL Server** và chọn **Properties** và nhập các thông tin cho Connection, sau khi hoàn thành thì ấn **Next**.



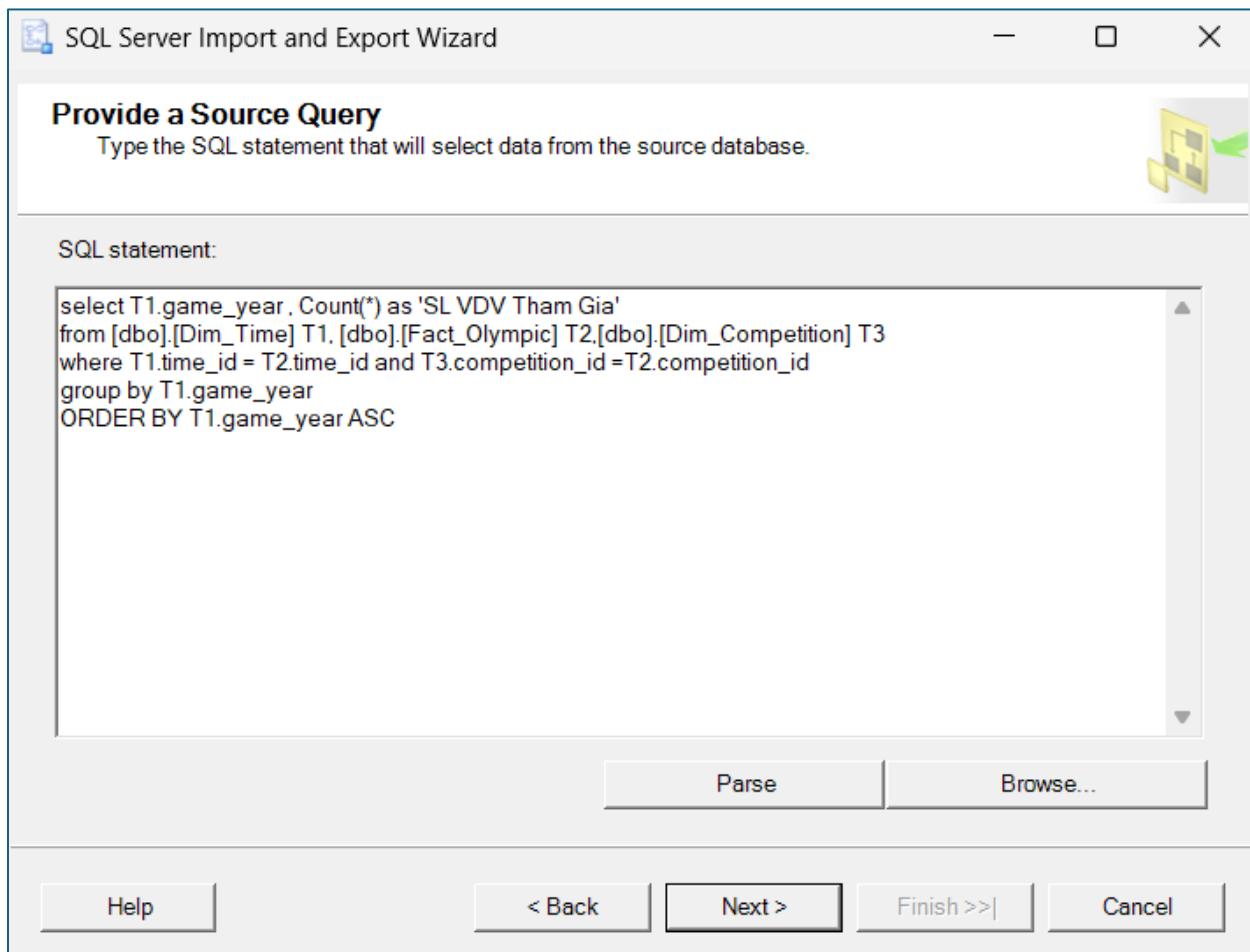
- **Bước 3:** Ké tiếp chọn **Destination**, ở đây ta chọn là **Flat File Destination** và nhấn **Browse** để nhập tên file và chọn nơi lưu trữ file csv được xuất ra và nhấn Next.



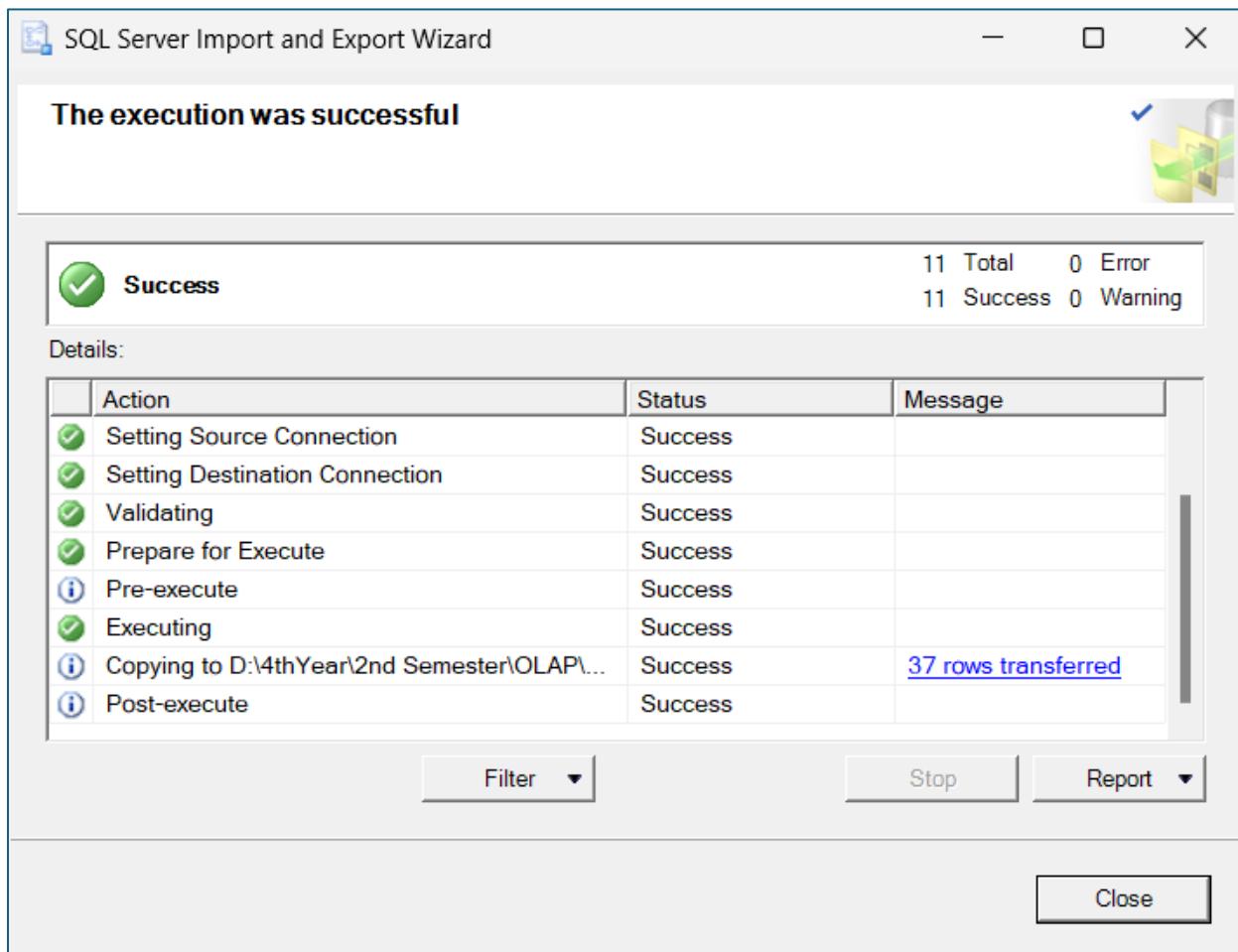
- **Bước 4:** Chọn xuất dữ liệu bằng cách viết câu truy vấn sau đó nhấp Next.



- **Bước 5:** Chép câu lệnh SQL phía trên vào và nhấn **Next** rồi sau đó **Finish** để bắt đầu xuất dữ liệu.

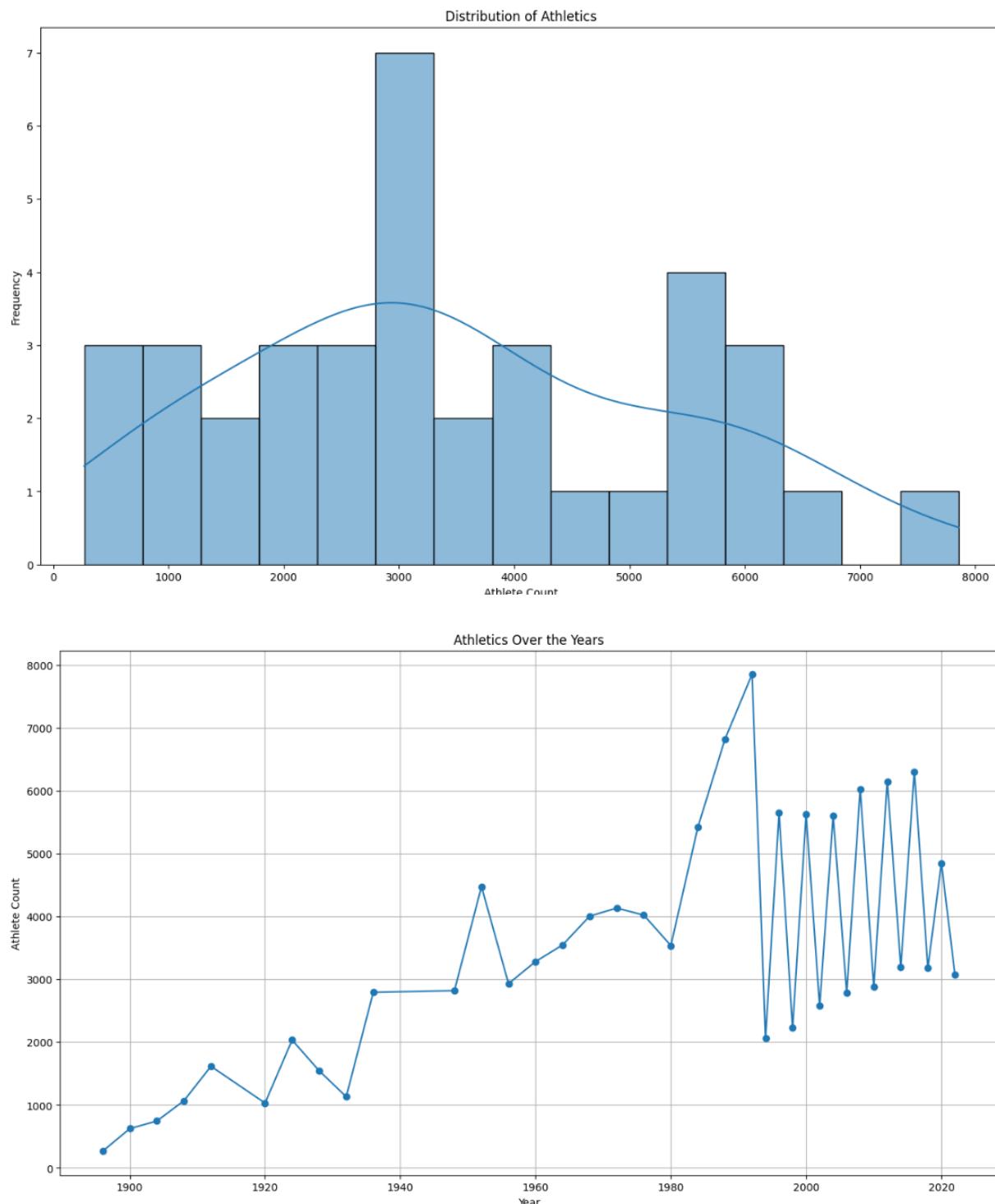


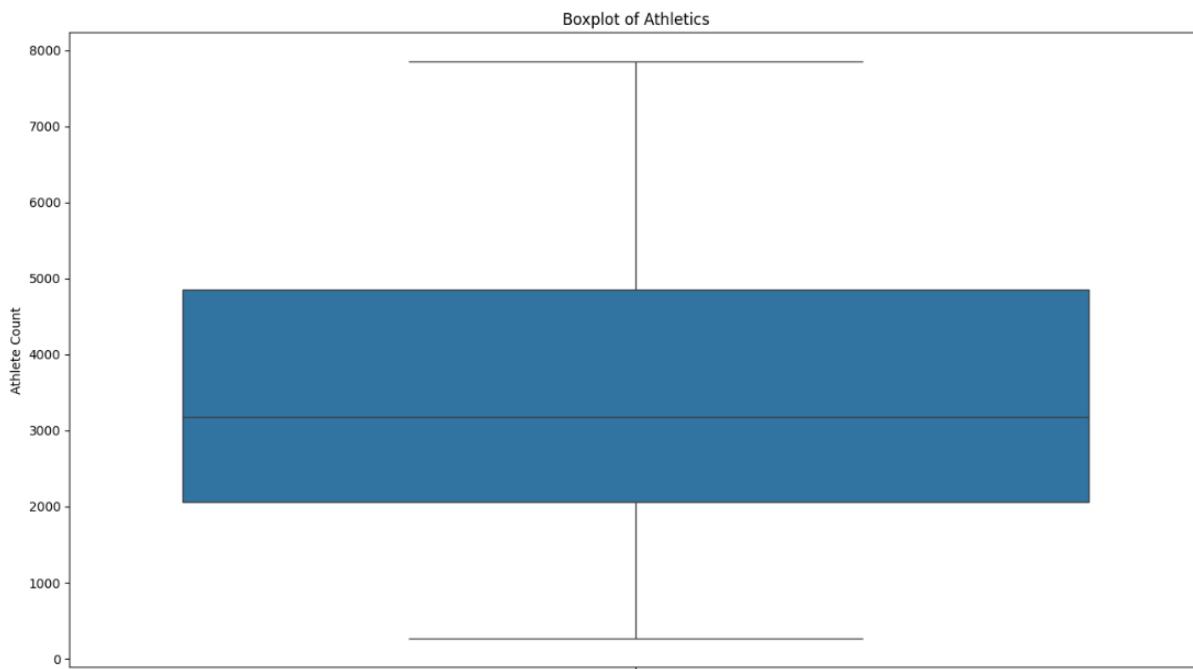
- **Bước 6:** Lúc này quy trình xuất dữ liệu sẽ được tiến hành và sẽ có kết quả thành công như hình dưới.



3. VẼ BIỂU ĐỒ

Mở Google Colaboratory và thực hiện đọc dữ liệu liệu từ file CSV ta vừa xuất ra. Ta tiến hành vẽ biểu đồ Boxplot và Histogram của dữ liệu Olympics.





❖ Nhận xét biểu đồ:

- Dữ liệu phân bổ khá đồng đều, có thể sử dụng để phân tích.
- Dữ liệu có xu hướng tăng dần về sau, do gia tăng số lượng các môn thi, tăng số lượng các quốc gia,...
- Do các kỳ Olympics tổ chức bộ môn Athletics diễn ra 4 năm/lần, ta chỉ có thể dự đoán hoặc thống kê các năm trong khoảng cách 4 năm. (bài toán này dự đoán trong 5 kỳ Olympics kế tiếp).

4. CHIA TẬP DỮ LIỆU

Trong quá trình Khai thác Dữ liệu này, cả 2 mô hình sẽ thực hiện việc chia tập dữ liệu thành hai phần: tập train và tập test theo tỉ lệ 80% cho tập train và 20% cho tập test.

Tập dữ liệu train sẽ được sử dụng để huấn luyện các mô hình dự đoán, trong khi tập dữ liệu test sẽ được sử dụng để đánh giá độ chính xác của các mô hình.

5. ĐÁNH GIÁ MÔ HÌNH

Việc đánh giá độ hiệu quả của các mô hình trong bài toán chuỗi thời gian trong phần Data mining sẽ dùng độ đo như RMSE và MAPE để so sánh 2 mô hình.

- **RMSE (Root Mean Square Error)** đo lường sai số trung bình giữa dự báo và giá trị thực tế trong đơn vị của dữ liệu. Giá trị RMSE càng nhỏ, tức là độ chính xác của mô hình càng cao
- **MAPE (Mean Absolute Percentage Error)** là một thước đo thống kê được sử dụng để đánh giá độ chính xác của mô hình dự báo. Nó tính toán sai số phần trăm trung bình giữa giá trị dự báo và giá trị thực tế.

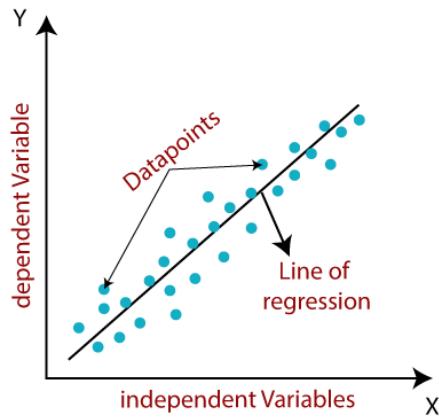
6. THUẬT TOÁN

6.1. Hồi quy tuyến tính (Linear Regression)

Linear Regression là một kỹ thuật thống kê phân tích mối quan hệ giữa hai hoặc nhiều biến và sử dụng thông tin để ước tính giá trị của các biến phụ thuộc^[1].

Thuật toán Linear Regression cho thấy mối quan hệ tuyến tính giữa biến phụ thuộc (y) và một hoặc nhiều biến độc lập (x), do đó được gọi là hồi quy tuyến tính. Vì hồi quy tuyến tính cho thấy mối quan hệ tuyến tính, có nghĩa là nó tìm cách thay đổi giá trị của biến phụ thuộc theo giá trị của biến độc lập.

Mô hình hồi quy tuyến tính cung cấp một đường thẳng dốc thể hiện mối quan hệ giữa các biến:



Công thức của hồi quy tuyến tính có thể trình bày dưới đây: [1]

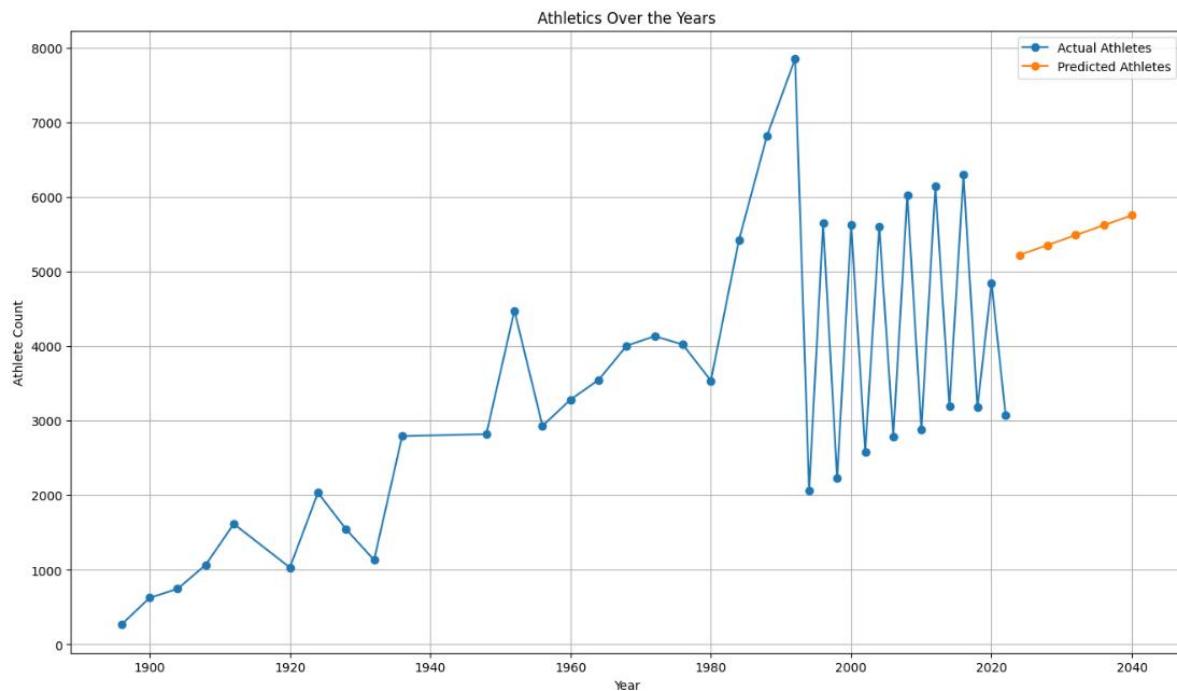
$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon$$

- y_i : biến phụ thuộc (dependent variable)
- x_{i1} : biến độc lập (independent variable)
- β_0 : hệ số chẵn (intercept coefficient)
- β_1 : hệ số hồi quy (regression coefficient)
- ε : sai số ngẫu nhiên (random error)

Hồi quy tuyến tính có thể được chia thành hai loại thuật toán:

- Simple Linear Regression: Nếu một biến độc lập duy nhất được sử dụng để dự đoán giá trị của một biến phụ thuộc số, thì thuật toán Hồi quy tuyến tính như vậy được gọi là Hồi quy tuyến tính đơn biến.
- Multiple Linear regression: Nếu nhiều hơn một biến độc lập được sử dụng để dự đoán giá trị của một biến phụ thuộc số, thì thuật toán Hồi quy tuyến tính như vậy được gọi là Hồi quy tuyến tính bội.

Tiến hành dự báo tập test và vẽ biểu đồ dự báo, ta được hình như sau:



```
test_score_rmse=np.sqrt(mean_squared_error(actual_athlete, predicted_athlete))
test_score_mape=mean_absolute_percentage_error(actual_athlete, predicted_athlete) * 100
|
print('RMSE:', test_score_rmse)
print('MAPE:', test_score_mape, '%')
```

RMSE: 1902.275645859886
MAPE: 50.75417428218185 %

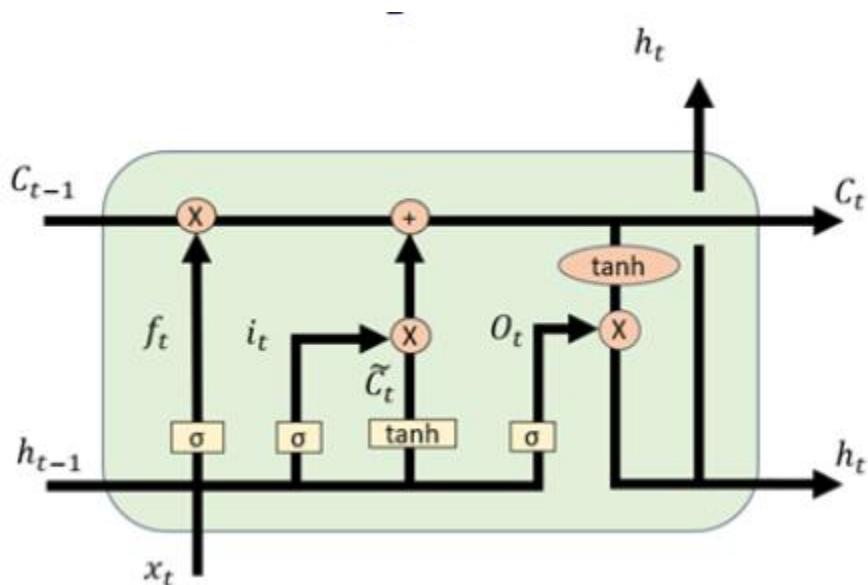
6.2. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) là một loại RNN đặc biệt có thêm tính năng ghi nhớ chuỗi dữ liệu. Việc ghi nhớ xu hướng trước đó của dữ liệu có thể thực hiện được thông qua một số công cùng với dòng bộ nhớ được tích hợp trong LSTM điển hình. [2]

- Một lớp LSTM được tạo thành từ một tập hợp các khói bộ nhớ được kết nối thường xuyên.
- Mỗi khói chứa một hoặc nhiều ô nhớ được kết nối thường xuyên thông qua ba đơn vị nhân - công đầu vào, đầu ra và công quên.

- Cung cấp các hoạt động tương tự liên tục của các hoạt động ghi, đọc và đặt lại của ô nhớ.
- Giảm thiểu nhược điểm của việc biến mất độ dốc bằng cách cho phép thông tin truyền trực tiếp hơn qua trạng thái ô nhớ.

Cấu trúc của LSTM:



Hình 6-1. Cấu trúc của LSTM

- LSTM được cấu tạo bởi các cổng, các cổng trả về các giá trị từ “0” đến “1”. Với “0” có nghĩa là cổng sẽ chặn tất cả mọi giá trị, “1” có nghĩa là cho phép mọi giá trị đi qua cổng.
- LSTM gồm các cổng: Input Gate, Forget Gate, Output Gate.

Phương trình các cổng của LSTM:

$$\text{Input gate: } i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\text{Forget gate: } f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Output gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

Trong đó:

σ : phương trình sigmoid

W_x : trọng số cho các nơ-ron cỗng (x) tương ứng

h_{t-1} : đầu ra của khối lstm trước đó (tại dấu thời gian t - 1)

x_t : đầu vào ở dấu thời gian hiện tại

b_o : độ lệch cho các cỗng tương ứng (x)

Ý nghĩa:

- Input Gate cho chúng ta biết thông tin mới nào chúng ta sẽ lưu trữ ở ô trạng thái
- Forget Gate thông báo loại bỏ thông tin khỏi ô trạng thái.
- Output Gate được sử dụng để cung cấp kích hoạt cho đầu ra cuối cùng của khối LSTM tại timestamp “t”.

Các phương trình cho ô trạng thái, ô trạng thái ứng cử viên và đầu ra cuối cùng:

Temporary cell state: $\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$

Current cell state: $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$

Final Output: $h_t = o_t * \tanh(C_t)$

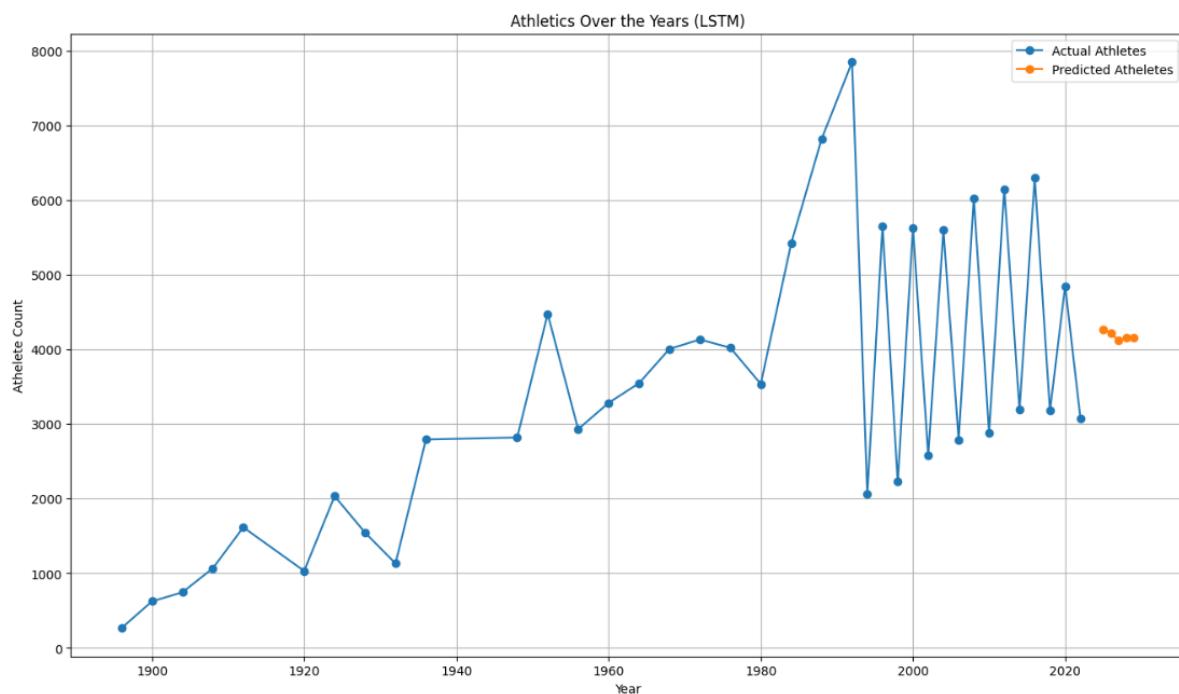
Trong đó:

C_t : ô trạng thái tại timestamp(t)

\tilde{C}_t : ứng viên cho ô trạng thái tại timestamp(t)

Ý nghĩa:

- Ở bất kỳ timestamp nào, ô trạng thái biết rằng nó cần quên những gì ở trạng thái trước đó và những gì nó cần xem xét từ timestamp hiện tại
- Cuối cùng, lọc ô trạng thái và sau đó nó được chuyển qua hàm kích hoạt để dự đoán phần nào sẽ xuất hiện dưới dạng đầu ra của đơn vị LSTM hiện tại tại timestamp t.



```
# Tính toán RMSE, MAPE, MDA trên LSTM
test_score_rmse=np.sqrt(mean_squared_error(actual_athlete, lstm_predicted_athlete))
test_score_mape=mean_absolute_percentage_error(actual_athlete, lstm_predicted_athlete) * 100

print('RMSE:', test_score_rmse)
print('MAPE:', test_score_mape, '%')
```

```
RMSE: 1267.3204242108072
MAPE: 29.13095297289214 %
```

7. KẾT QUẢ

Độ chính xác: Mô hình LSTM cho thấy mức độ chính xác tốt hơn so với mô hình Linear Regression, với sai số trung bình so sánh như bảng dưới. Điều này cho thấy mô hình LSTM có khả năng dự đoán gần đúng số lượng vận động viên tốt hơn trong trường hợp này.

LSTM có chỉ số cao hơn, do nó có khả năng mô hình hóa các mẫu chuỗi thời gian phức tạp hơn và có thể đạt được hiệu suất tốt hơn trong các tình huống phức tạp hơn.

	Linear	LSTM
RMSE	1902.275645859886	1267.3204242108072
MAPE	50.75417428218185%	29.13095297289214%

TÀI LIỆU THAM KHẢO

[1]“Linear Regression in Machine learning,” GeeksforGeeks. Available:

<https://www.geeksforgeeks.org/ml-linear-regression/>

[2]“What is LSTM? Introduction to Long Short-Term Memory | by Rebeen Hamad |

Medium.” Available: <https://medium.com/@rebeen.jaff/what-is-lstm-introduction-to-long-short-term-memory-66bd3855b9ce>