

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

-----OOBOOKOO-----



BÁO CÁO ĐỒ ÁN MÔN HỌC
KHO DỮ LIỆU VÀ OLAP

ĐỀ TÀI
**XÂY DỰNG KHO DỮ LIỆU PHÂN TÍCH XU HƯỚNG
PHIM ẢNH TRÊN NỀN TẢNG TOÀN CẦU TMDB**

Giảng viên: Đỗ Thị Kim Phụng

Lớp: IS217.O22.HTCL

Thành viên: Nhóm 1

Ngô Đức Hoàng Hiệp – 21520846

Phạm Thị Trâm Anh – 21520146

Thành phố Hồ Chí Minh, tháng 5 năm 2024

MỤC LỤC

PHẦN 1: GIỚI THIỆU KHO DỮ LIỆU.....	5
1.1. Lý do chọn đề tài	5
1.2. Giới thiệu bộ dữ liệu	5
1.2.1. Tên cơ sở dữ liệu	5
1.2.2. Nguồn cung cấp	5
1.2.3. Thời gian thu thập.....	5
1.2.4. Cấu trúc cơ sở dữ liệu.....	5
1.2.5. Các thuộc tính phân tích	6
1.3. Mô hình kho dữ liệu	7
1.3.1. Lược đồ kho dữ liệu	7
1.3.2. Mô tả các bảng trong kho dữ liệu	7
1.4. Các câu truy vấn.....	9
PHẦN 2: QUÁ TRÌNH XÂY DỰNG KHO DỮ LIỆU (SSIS)	11
2.1. Chuẩn bị các công cụ	11
2.2. Chuẩn bị cơ sở dữ liệu	12
2.3. Xử lý dữ liệu.....	12
2.4. Tạo mới project SSIS	13
2.5. Tạo bảng Dim và bảng Fact	14
2.5.1. Bảng Dim_release_date	18
2.5.2. Bảng Dim_adult.....	27
2.5.3. Bảng Dim_runtime.....	31
2.5.4. Bảng Dim_language	35

2.5.5. Bảng Dim_movies.....	39
2.5.6. Bảng Fact	47
2.5.7. Chạy dự án SSIS.....	71
2.5.8. Kiểm tra dữ liệu các bảng	75
2.6. Lược đồ cơ sở dữ liệu sau khi hoàn thành	78
PHẦN 3: THỰC HIỆN PHÂN TÍCH TRÊN KHO DỮ LIỆU (SSAS).....	79
3.1. Tạo project SSAS mới.....	79
3.2. Xác định dữ liệu nguồn (Data Sources)	81
3.3. Xác định khung nhìn dữ liệu nguồn (Data source view)	85
3.4. Tạo chiều thời gian.....	88
3.5. Tạo cube	93
3.6. Điều chỉnh các thuộc tính của cube	99
3.6.1. Tạo hierarchy cho chiều thời gian	99
3.6.2. Đổi tên một số thuộc tính.....	104
3.7. Chạy dự án SSAS	105
3.8. Thực hiện 15 câu truy vấn (MDX, Manual và Pivot Excel).....	106
PHẦN 4: QUÁ TRÌNH TẠO BÁO CÁO DỮ LIỆU(SSRS).....	129
4.1. Tạo báo cáo với Reporting Service trên Visual Studio.....	129
4.1.1. Cấu hình Report Server Configuration Manager.....	129
4.1.2. Khởi tạo và cấu hình project Visual Studio.....	131
4.1.3. Báo cáo 1	136
4.1.4. Báo cáo 2	140
4.1.5. Báo cáo 3	142
4.1.6. Deploy các báo cáo	143

4.2. Tạo báo cáo với PowerBI.....	145
4.2.1. Cài đặt PowerBI	145
4.2.2. Tạo project SSRS trên Power BI	145
4.2.3. Báo cáo 1	147
4.2.4. Báo cáo 2	150
4.2.5. Báo cáo 3	151
4.2.6. Deploy các báo cáo	153

PHẦN 1: GIỚI THIỆU KHO DỮ LIỆU

1.1. Lý do chọn đề tài

TMDB (The Movie Database) không chỉ là một nền tảng phổ biến mà còn được coi là một nguồn thông tin đáng tin cậy trong cộng đồng yêu thích điện ảnh. Với hàng triệu bộ phim cùng với đánh giá và nhận xét từ người dùng trên khắp thế giới, TMDB trở thành một kho tài nguyên vô cùng phong phú, giúp chúng ta tiếp cận và phân tích xu hướng phim một cách sâu sắc.

Bằng việc khai thác dữ liệu từ TMDB, chúng ta có thể đánh giá được sự phổ biến của các dòng phim, nhận diện những thể loại nào đang nổi lên, và cũng nhận thấy những yếu tố như đánh giá của khán giả và doanh thu phòng vé đang ảnh hưởng ra sao. Phân tích xu hướng trên TMDB không chỉ giúp các nhà làm phim, nhà sản xuất và nhà phát hành hiểu rõ hơn về nhu cầu của thị trường mà còn cung cấp cho họ cái nhìn chiến lược, từ đó họ có thể ra quyết định đúng đắn khi sản xuất và phát hành các bộ phim, thu hút sự chú ý của khán giả.

1.2. Giới thiệu bộ dữ liệu

1.2.1. Tên cơ sở dữ liệu

Cơ sở dữ liệu về phim ảnh năm 2024 (The movies database).

1.2.2. Nguồn cung cấp

Dữ liệu được lấy từ nền tảng Kaggle, cơ sở dữ liệu được tác giả trích xuất từ trang website The Movie Database (TMDB).

1.2.3. Thời gian thu thập

Từ 10/2023 đến 3/2024.

1.2.4. Cấu trúc cơ sở dữ liệu

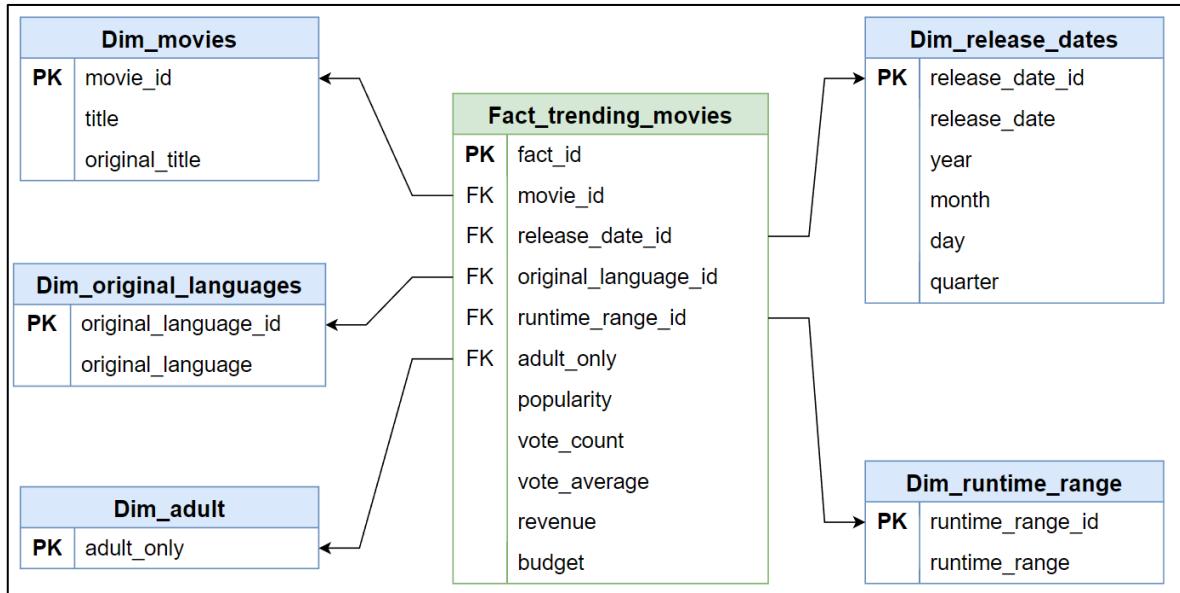
- Số cột: 23
- Số dòng: 1 005 152

1.2.5. Các thuộc tính phân tích

STT	Tên thuộc tính	Kiểu dữ liệu	Mô tả
1	id	int	Khóa chính của bộ phim
2	title	nvarchar(255)	Tiêu đề phim
3	vote_average	float	Điểm bình chọn trung bình
4	vote_count	int	Tổng bình chọn
5	status	nvarchar(50)	Trạng thái của bộ phim gồm: Released, In Production, Post Production, Planned, Canceled
6	release_date	date	Ngày công chiếu.
7	revenue	int	Tổng doanh thu(đơn vị \$)
8	runtime	int	Thời lượng của phim(đơn vị phút)
9	adult	bool	Thể hiện bộ phim chỉ phù hợp cho người xem thành niên. Chứa hai giá trị True hoặc False.
10	backdrop_path	nvarchar(255)	Đường dẫn của hình nền phim.
11	budget	int	Chi phí sản xuất phim(đơn vị \$)
12	homepage	nvarchar(255)	Đường dẫn đến trang chủ của phim
13	imdb_id	nvarchar(20)	ID của phim trên nền tảng IMDB.
14	original_title	nvarchar(255)	Tên gốc của bộ phim
15	original_language	Nvarchar(255)	Ngôn ngữ gốc của bộ phim
16	overview	nvarchar(255)	Mô tả ngắn gọn của bộ phim
17	popularity	float	Điểm phổ biến của bộ phim.
18	poster_path	nvarchar(255)	Đường dẫn đến ảnh áp phích của phim
19	tagline	nvarchar(255)	Câu khẩu hiệu của bộ phim
20	genres	nvarchar(50)	Thể loại của bộ phim
21	production_companies	nvarchar(50)	Công ty tham gia sản xuất bộ phim
22	production_countries	nvarchar(50)	Quốc gia tham gia sản xuất bộ phim
23	spoken_languages	nvarchar(50)	Ngôn ngữ được nói trong phim

1.3. Mô hình kho dữ liệu

1.3.1. Lược đồ kho dữ liệu



1.3.2. Mô tả các bảng trong kho dữ liệu

1.3.2.1. Bảng Fact_trending_movies

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
Khóa chính	fact_id	int	Mã fact
Khóa ngoại	movie_id	int	Mã bộ phim
Khóa ngoại	release_date	datetime	Ngày công chiếu
Khóa ngoại	original_language_code	varchar(2)	Mã ngôn ngữ gốc (đầu tiên) được sử dụng trong phim.
Khóa ngoại	runtime_range_id	int	Mã thời lượng phim
Khóa ngoại	adult_only	bool	Mã chỉ thị rằng phim chỉ dành cho người lớn (18+)
	popularity	float	Điểm đánh giá độ thịnh hành của bộ phim
	vote_count	int	Tổng số lượng đánh giá của người xem dành cho bộ phim

	vote_average	float	Điểm đánh giá trung bình của người xem danh cho bộ phim
	revenue	float	Tổng doanh thu của bộ phim.
	budget	float	Tổng chi phí sản xuất cho bộ phim.

1.3.2.2. Bảng Dim_movies

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
Khóa chính	movie_id	int	Mã bộ phim
	title	nvarchar(1000)	Tên bộ phim
	original_title	nvarchar(1000)	Tên ban đầu của phim

1.3.2.3. Bảng Dim_release_dates

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
Khóa chính	release_date_id	int	Khóa chính
	release_date	datetime	Ngày tháng năm công chiếu
	year	int	Năm công chiếu
	month	int	Tháng công chiếu
	day	int	Ngày công chiếu
	quarter	int	Quý công chiếu

1.3.2.4. Bảng Dim_original_languages

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
Khóa chính	original_language_id	int	Khóa chính tăng tự động.
	Original_language	varchar(2)	Mã ký hiệu ngôn ngữ, gồm 2 ký tự

1.3.2.5. Bảng Dim_adult

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
Khóa chính	adult	bool	Chỉ thị chỉ dành cho người lớn: <ul style="list-style-type: none"> • True: chỉ cho phép người xem trên 18 tuổi.

			<ul style="list-style-type: none"> • False: không giới hạn độ tuổi.
--	--	--	--

1.3.2.6. Bảng Dim_runtime_range

	Tên thuộc tính	Kiểu dữ liệu	Mô tả
Khóa chính	runtime_id	int	Mã dung lượng phim
	runtime_range	nvarchar(20)	Khoảng thời lượng: <ul style="list-style-type: none"> • under 1h: dưới 1h • 1h - 2h: từ 1h đến 2h • 2h - 3h: từ 2h đến 3h • over 3h: trên 3h

1.4. Các câu truy vấn

Câu 1: Tổng doanh thu và ngân sách của các phim theo các mốc thời lượng chiếu.

Câu 2: Liệt kê 5 bộ phim có doanh thu cao nhất, sắp xếp theo thứ tự giảm dần.

Câu 3: Thống kê 10 tên phim có ngân sách cao nhất được công chiếu vào năm 2023 sắp xếp giảm dần theo độ thịnh hành.

Câu 4: Số lượng phim được công bố vào các quý của năm 2023.

Câu 5: Thống kê tên phim, số lượng bình chọn và điểm bình chọn trung bình của các bộ phim có ngân sách trên 1.5 tỷ đô la.

Câu 6: Top 3 bộ phim có ngân sách cao nhất theo 3 ngôn ngữ Anh, Việt và Trung.

Câu 7: Thống kê 2 phim phổ biến nhất có tiêu đề bắt đầu bằng ký tự “H” ở từng năm từ năm 2020 đến 2024.

Câu 8: Số lượng phim, số lượng bình chọn và điểm bình chọn trung bình theo các mốc thời lượng có ngôn ngữ gốc là Anh, Trung, Việt.

Câu 9: Tổng doanh thu của các phim được công bố từ năm 2020 đến năm 2024, thực hiện drilldown từ năm sang quý.

Câu 10: Số lượt bình chọn và điểm thịnh hành của các bộ phim có ngôn ngữ gốc là tiếng Việt, tiếng Anh, có thời lượng dưới 1h công chiếu từ năm 2020 đến 2023.

Câu 11: Thống kê số lượng và phần trăm các bộ phim có chi phí sản xuất lớn hơn doanh thu (gấp tình trạng lỗ) trong 5 năm gần nhất.

Câu 12: Thống kê số lượng phim và tổng số lượt bình chọn theo từng quý từ năm 2021 đến 2024 của hai ngôn ngữ là tiếng Anh và tiếng Việt.

Câu 13: Thống kê số lượng phim và tổng doanh thu của theo từng năm từ 2021 đến 2024 và độ tuổi người xem cho phép(trên 18 tuổi hay không).

Câu 14: Top 5 các phim có lượt bình chọn thấp nhất được công chiếu trong khoảng từ năm 2021 đến 2024, có doanh thu và ngân sách lớn hơn 0.

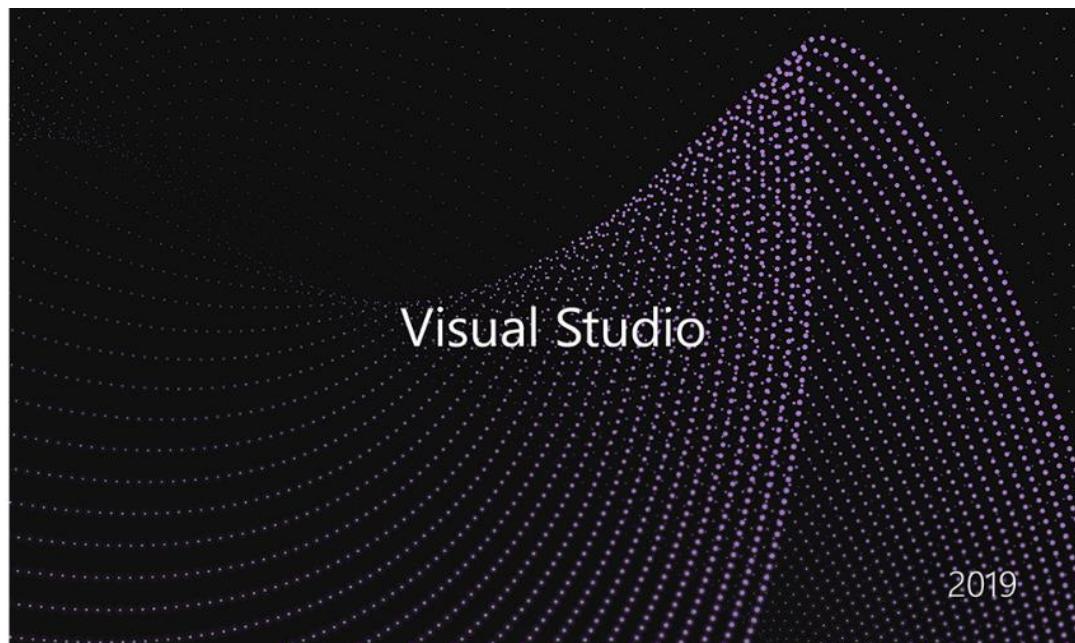
Câu 15: Thống kê tổng số lượt bình chọn, điểm thịnh hành, tổng ngân sách của các năm có số lượng phim công chiếu lớn 15000, sắp xếp theo thứ tự giảm dần theo tổng ngân sách.

PHẦN 2: QUÁ TRÌNH XÂY DỰNG KHO DỮ LIỆU (SSIS)

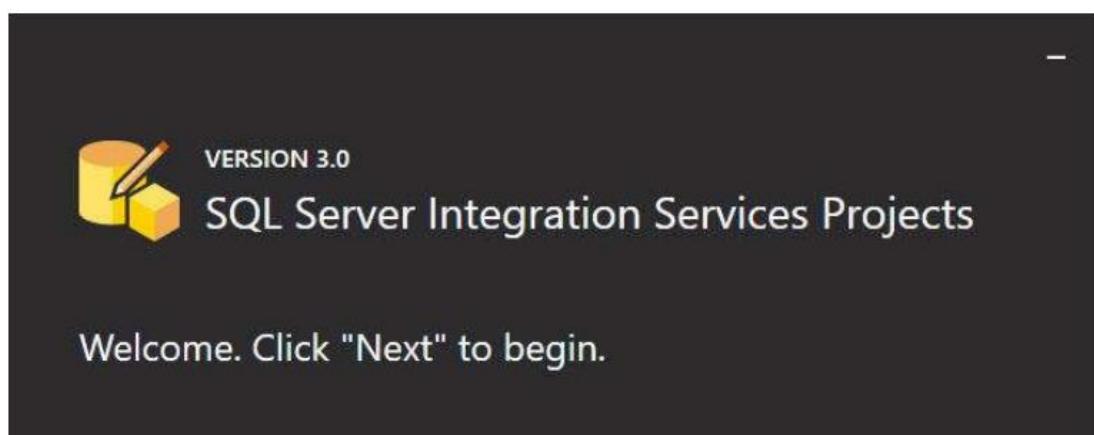
2.1. Chuẩn bị các công cụ

Để thực hiện được quá trình SSIS ta cần chuẩn bị và cài đặt các công cụ sau:

- + Visual Studio Community 2019

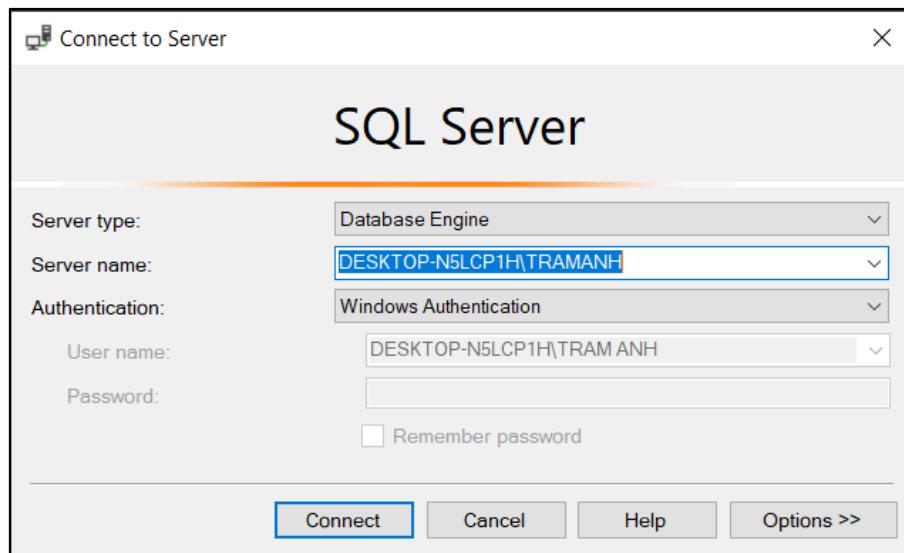


- + SQL Server Integration Services Project



2.2. Chuẩn bị cơ sở dữ liệu

Bước 1: Mở SQL Server 2019 và kết nối với server bằng tài khoản user của window (Windows Authentication).



Bước 2: Khởi tạo một cơ sở dữ liệu có tên MoviesTMDb, đây là nơi lưu các bảng Dim và bảng Fact cùng dữ liệu của các bảng đó.

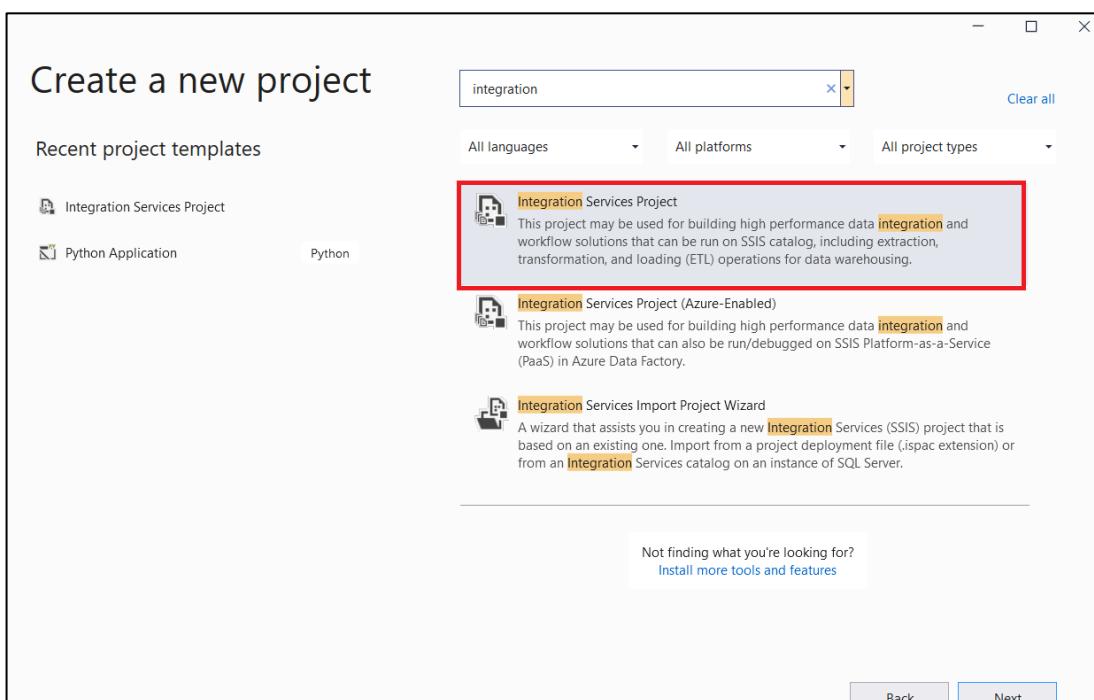
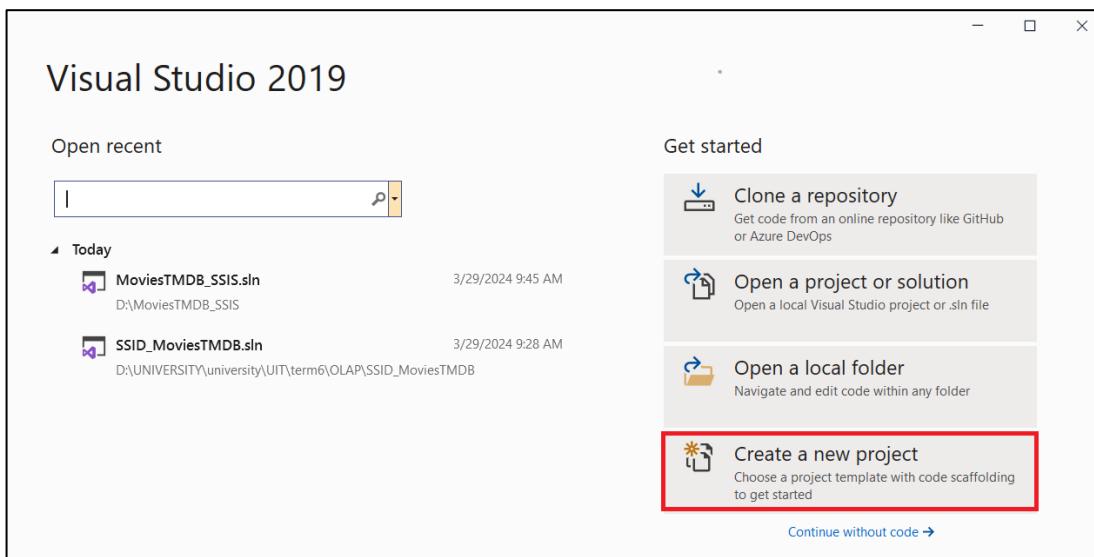
2.3. Xử lý dữ liệu

Cột runtime (dung lượng bộ phim) trong file .csv gốc được biến đổi từ giá trị liên tục đến theo dạng khoảng (bin). Bằng cách sử dụng Python với cú pháp if...else..., chúng tôi chia thành 4 đoạn: under 1h, 1h – 2h, 2h – 3h và over 3h.

runtime_range
2h - 3h
1h - 2h
2h - 3h

2.4. Tạo mới project SSIS

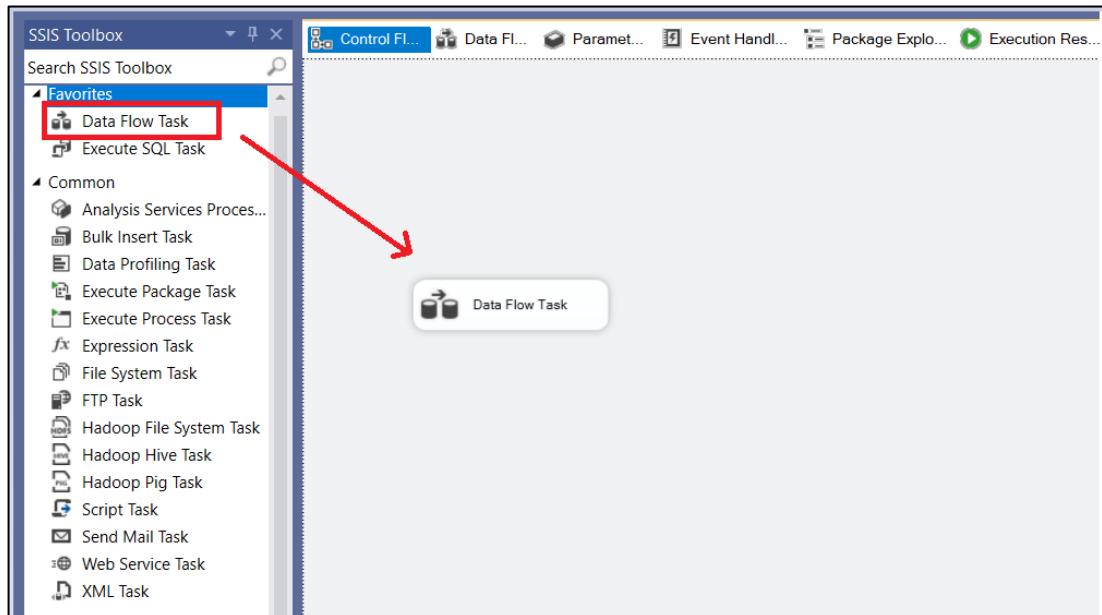
Bước 1: Mở Visual Studio 2019 và chọn “Create a new project”, chọn Integration Services Project và chọn Next.



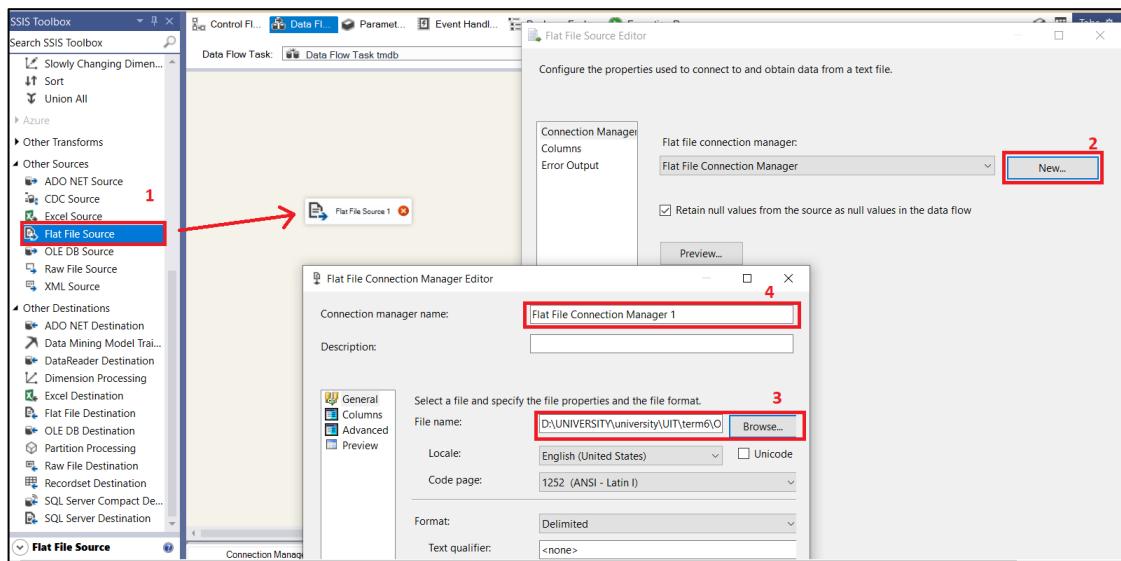
Bước 2: Đặt tên và thiết lập đường dẫn cho Project.

2.5. Tạo bảng Dim và bảng Fact

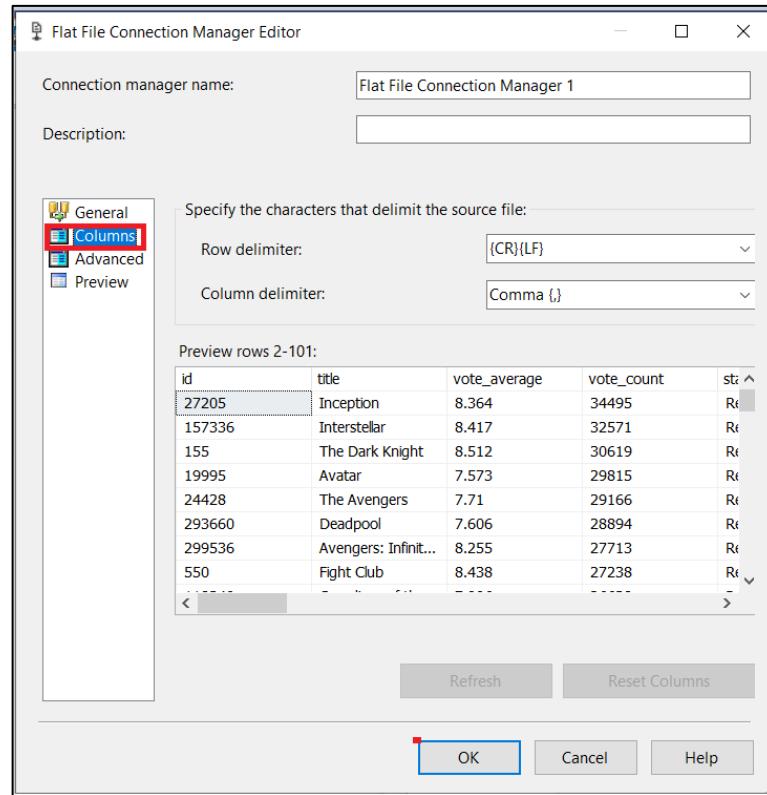
Giai đoạn 1: Load dữ liệu gốc từ file .csv vào Data Flow



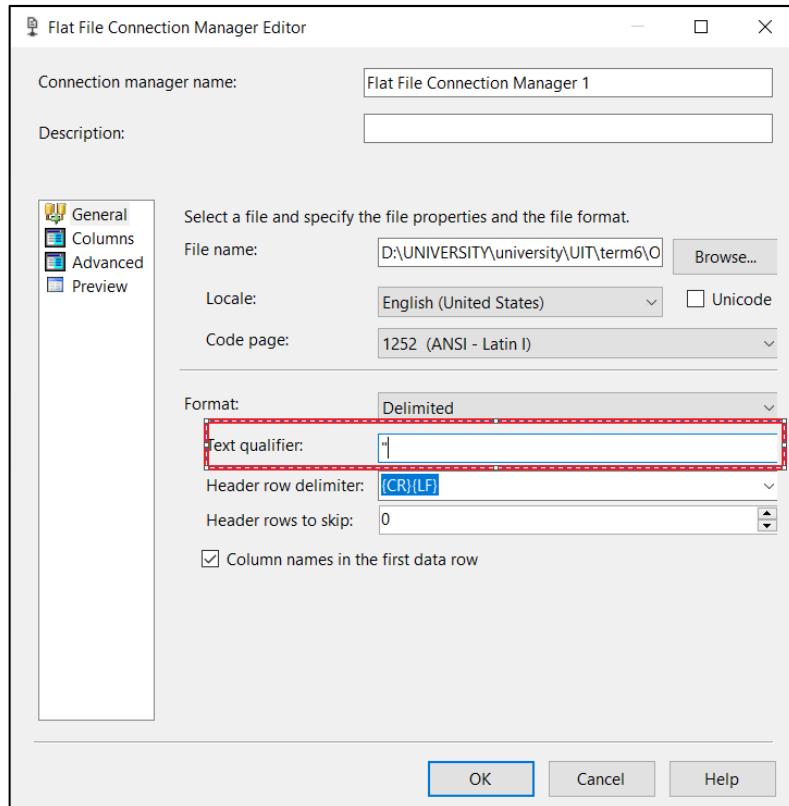
Bước 1: Trong Data Flow, tạo một đối tượng Flat File Source để lấy dữ liệu gốc từ file .csv. Chọn New để tạo một Flat File Connection Manager. Lấy đường dẫn tới file .csv và đặt tên cho Connection manager.



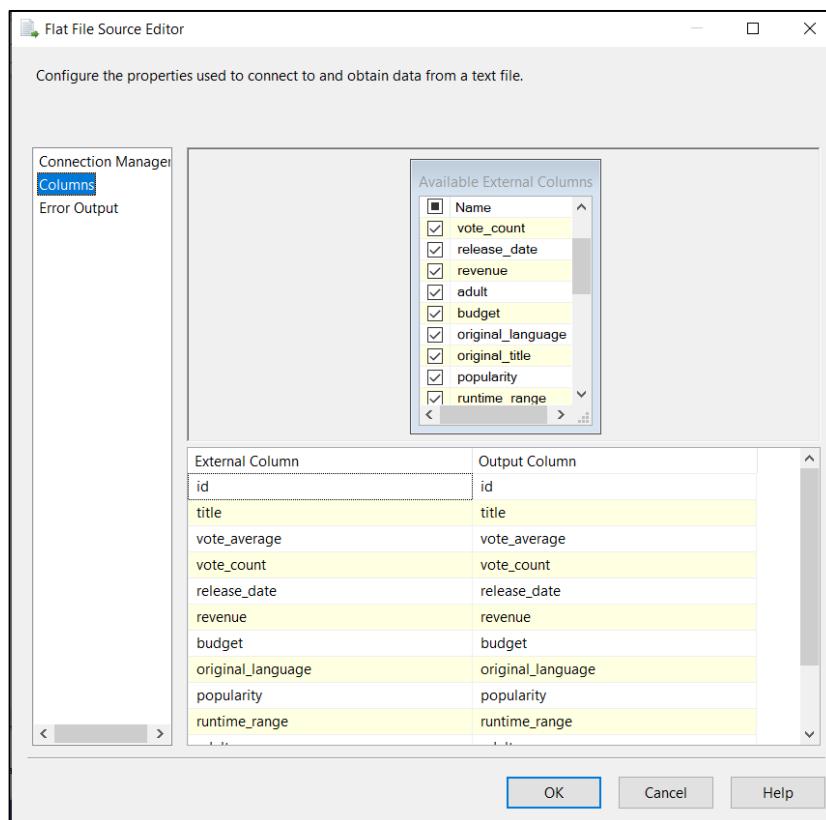
Bước 2: Xem lại các cột dữ liệu trong file dữ liệu đã được tải lên ở menu Columns.



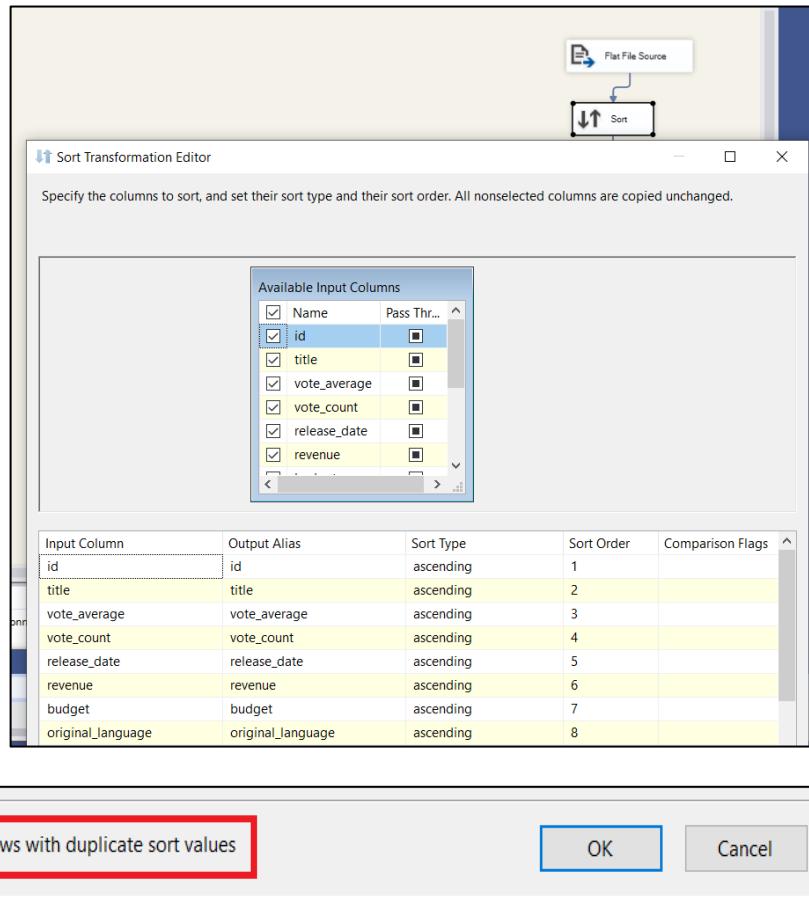
Lưu ý: Nếu giá trị chưa được tách đúng cột thì chúng ta phải điều chỉnh ở Text qualifier.



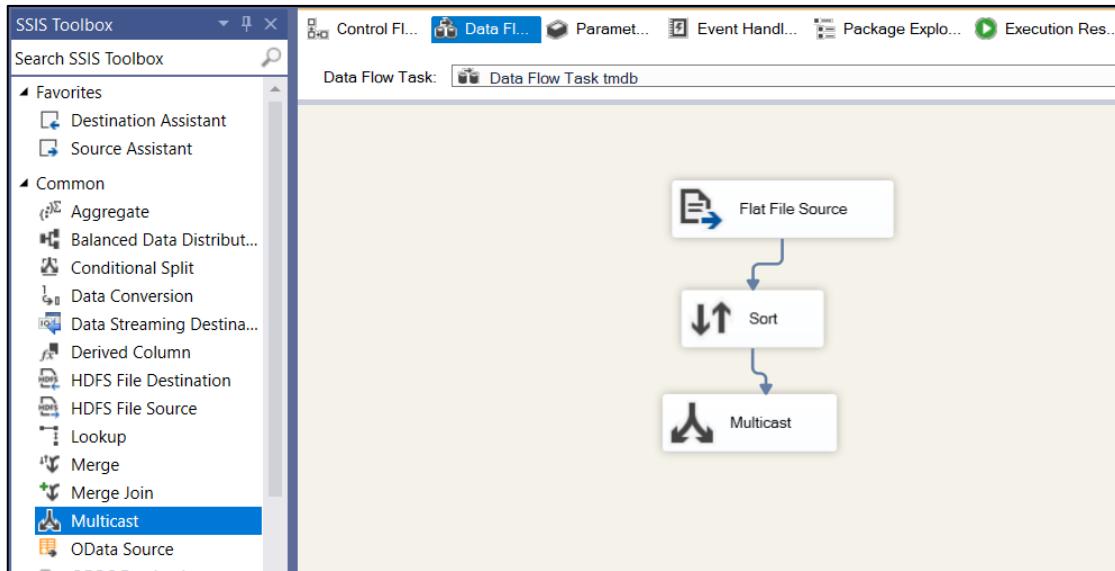
Bước 3: Lựa chọn các cột cần thiết và ấn OK



Bước 4: Tạo Sort để xóa các dòng giống nhau. Chọn tất cả các trường dữ liệu.



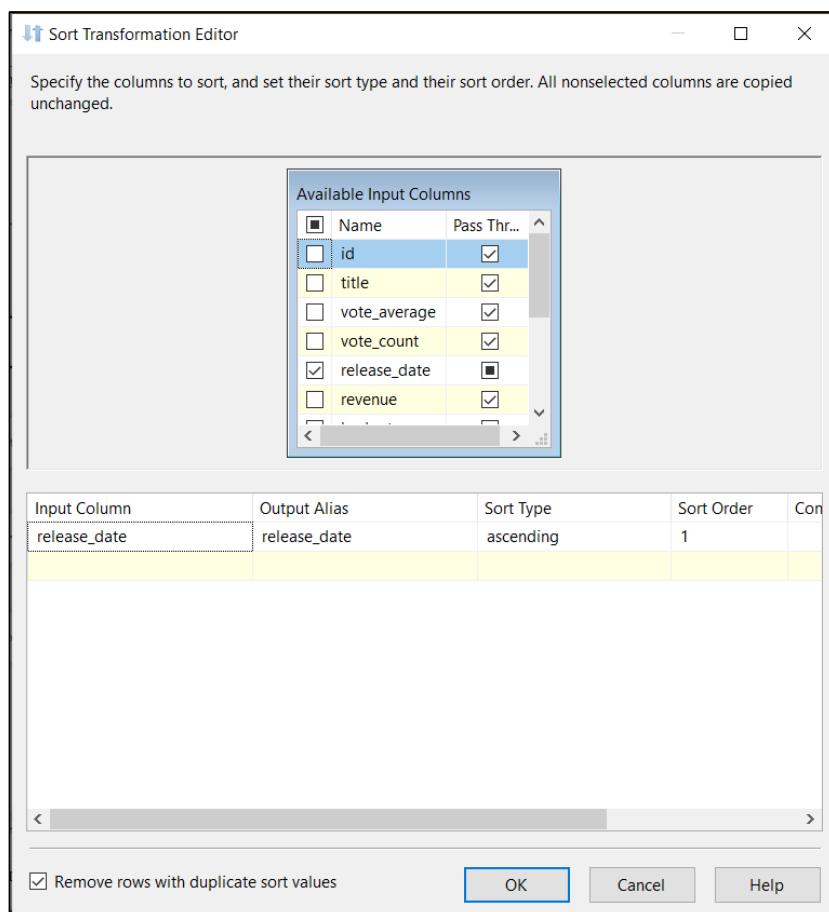
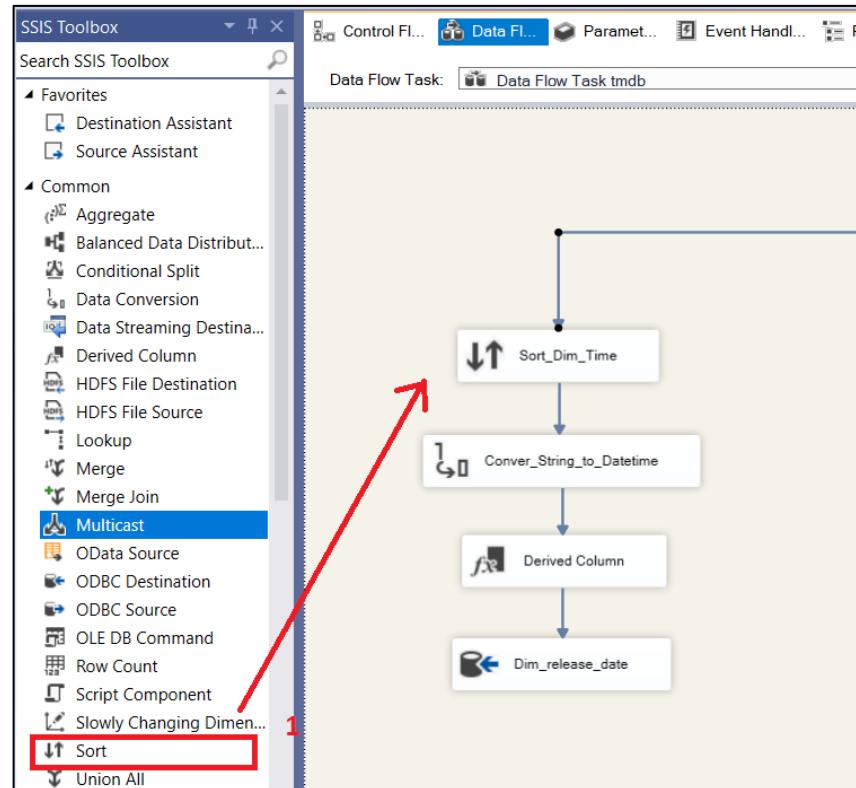
Bước 5: Tạo Multicast để phân tán dữ liệu từ Flat File Source đến các Dimension. Tiến hành kết nối Flat File Source và Multicast.



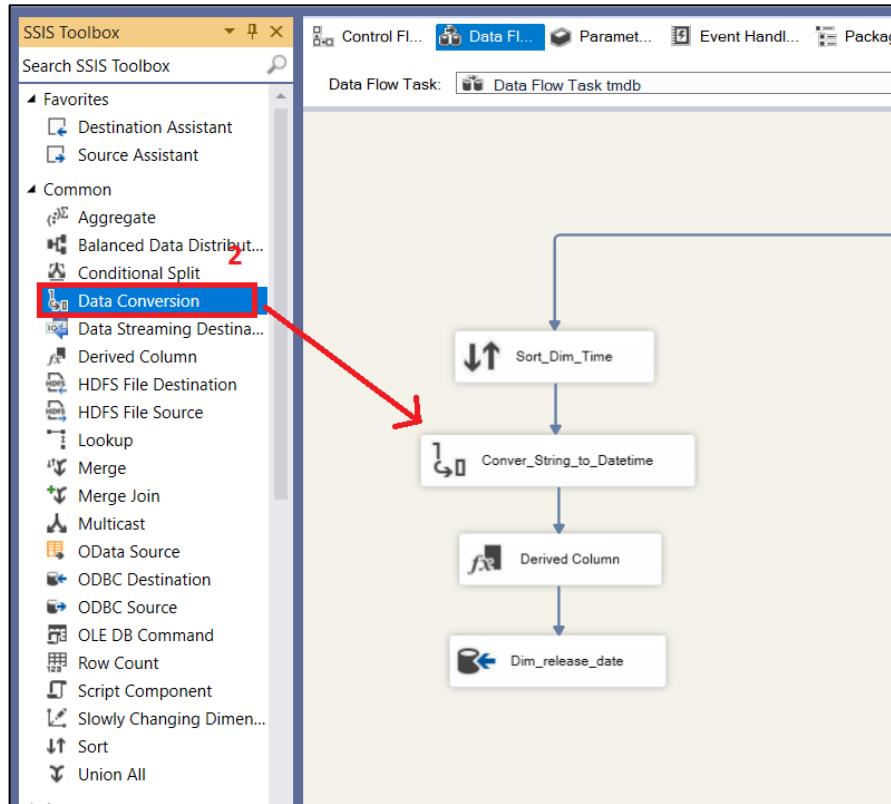
Giai đoạn 2: Tạo bảng Dim và bảng Fact

2.5.1. Bảng Dim_release_date

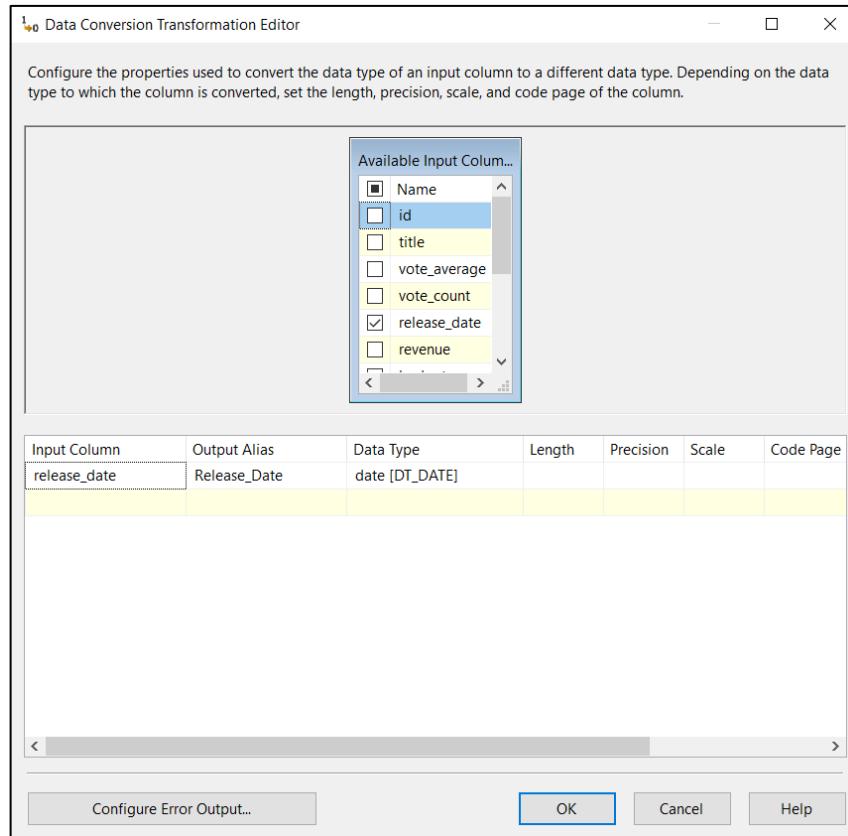
Bước 1: Tạo mới một Sort có tên là Sort_Dime_Time để lấy ra các cột dữ liệu cần thiết cho Dim_Time. Nhấn chuột phải và nhấn Edit để chọn Start_Time làm cột dữ liệu cho Sort_Dime_Time. Tick chọn Remove rows with duplicate sort values xóa các dòng dữ liệu trùng nhau, sau đó chọn OK



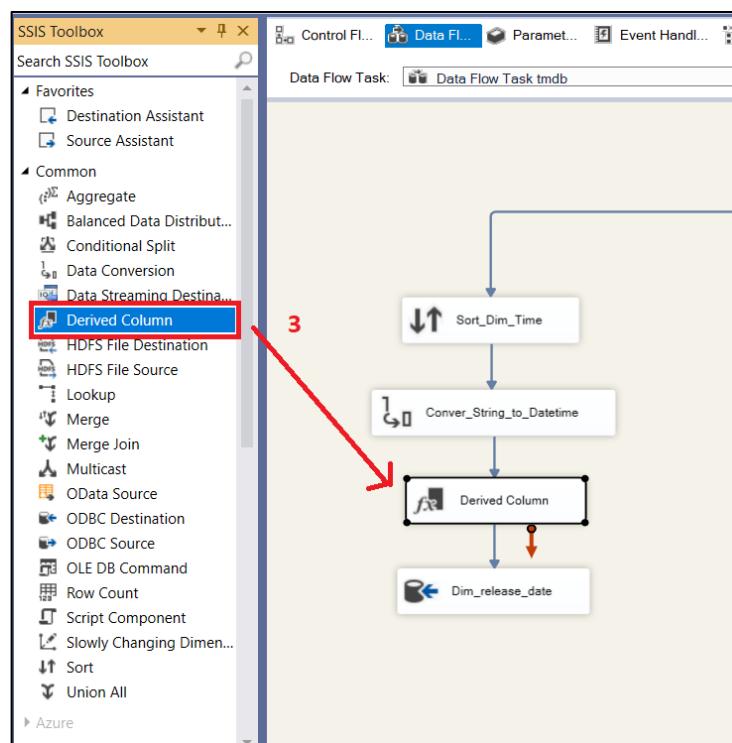
Bước 2: Kiểu dữ liệu DateTime khi lấy từ dữ liệu gốc sẽ được mặc định là kiểu string. Ta dùng Data Conversion để chuyển đổi kiểu string về dạng DateTime. Đặt tên cho Data Conversion này là Dim_Time Change type string to datetime.



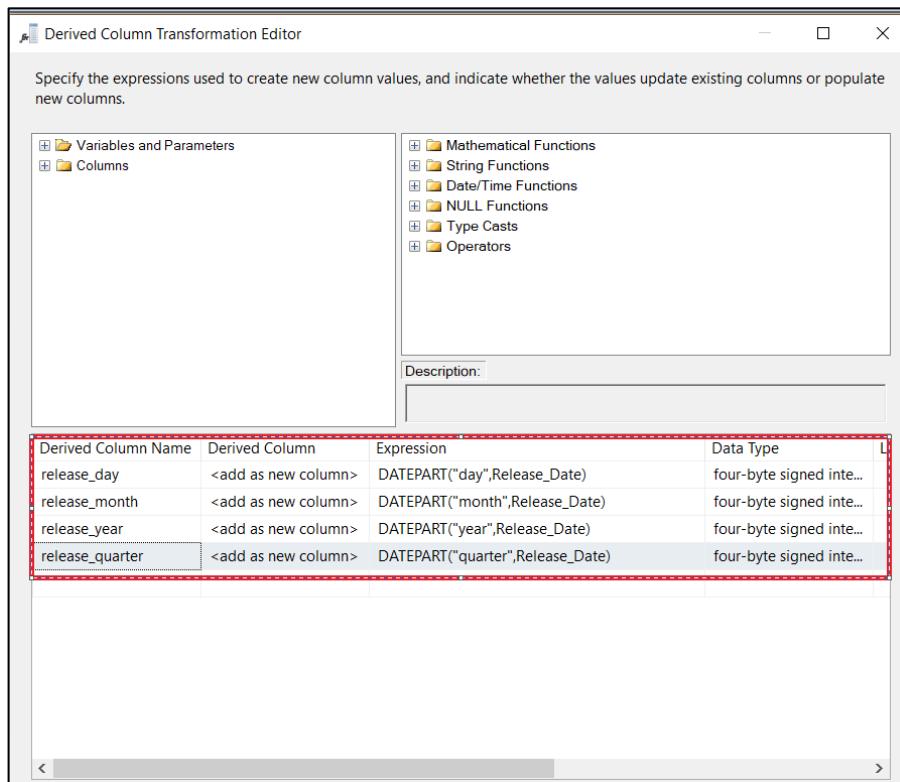
Click chuột phải vào Data conversion này và chọn Edit, click chọn cột Start_Time, ở cột Data Type ta thấy mặc định kiểu dữ liệu là string, ta chọn lại kiểu dữ liệu cho cột này date[DT_DATE]. Đặt lại alias và ấn OK.



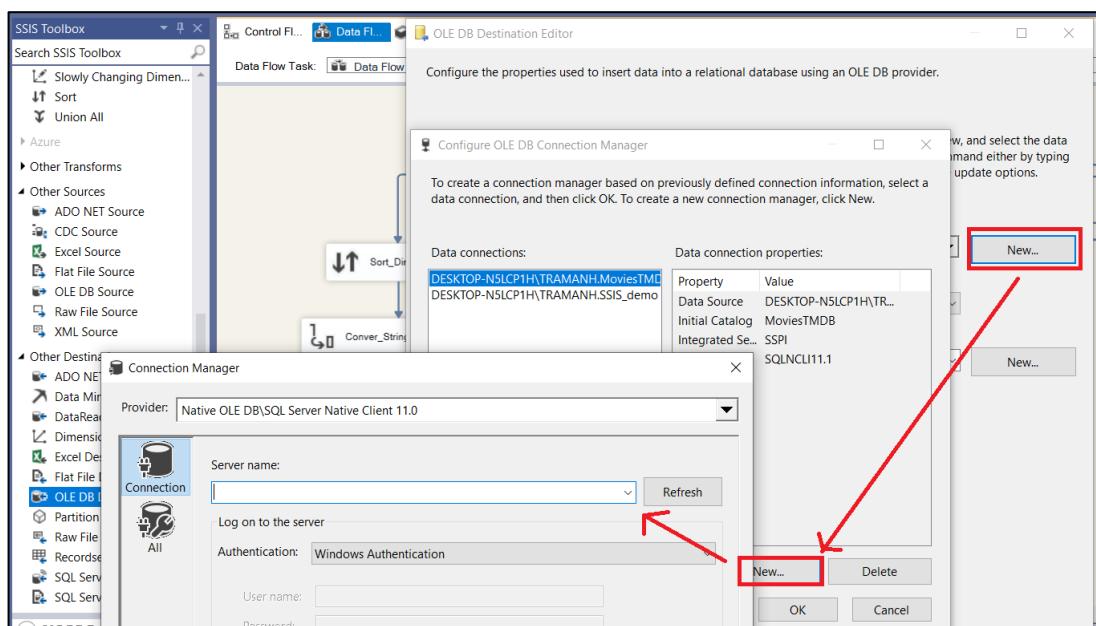
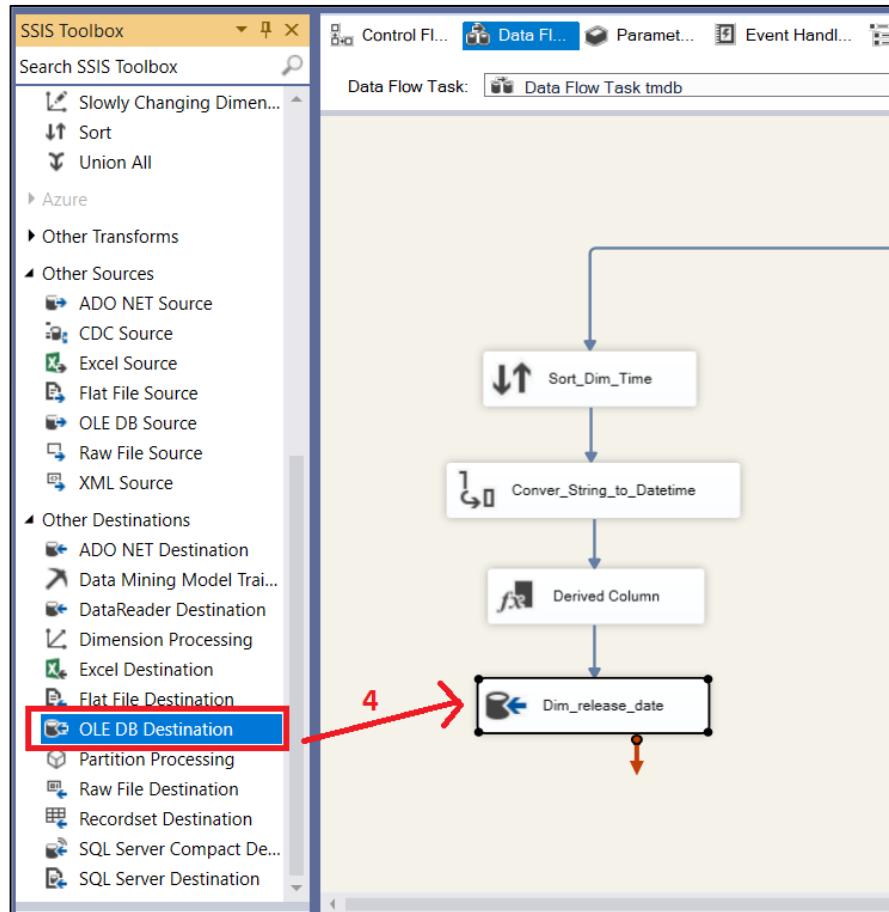
Bước 3: Thêm thành phần Derived Column và chọn Edit để chia cột dữ liệu.



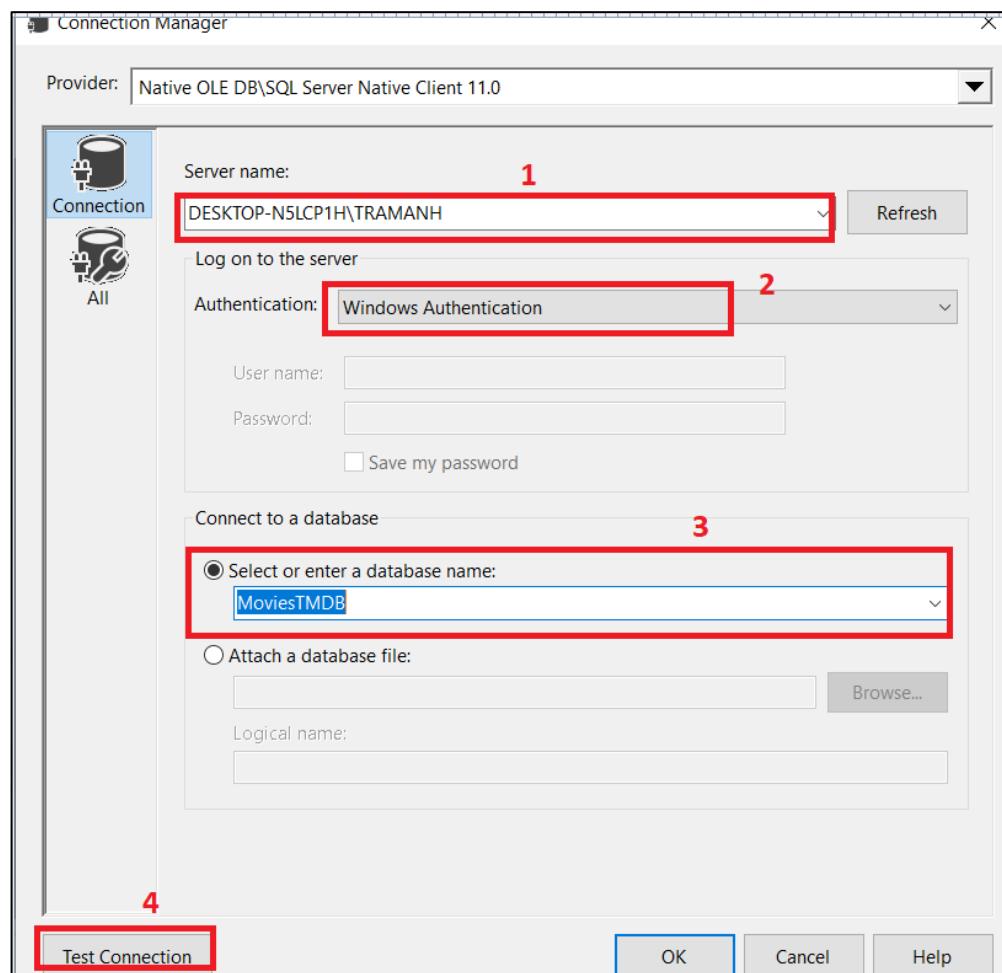
Chia dữ liệu từ cột Release_Date có kiểu dữ liệu dd/MM/yyyy thành các cột release_day, release_month, release_year, release_quarter. Chọn phương thức DATEPART(, cột dữ liệu).

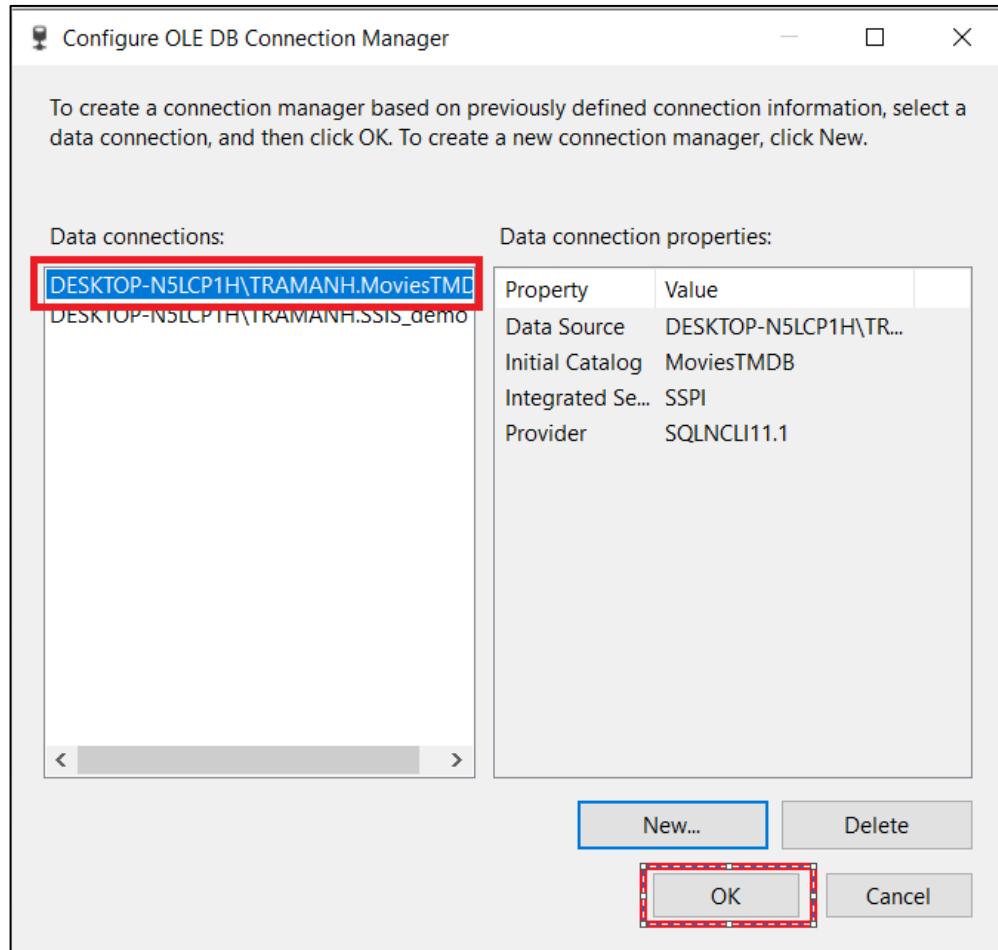


Bước 6: Tạo Dim_release_date từ một OLE DB Destination. Double click vào OLE DB Destination này để tạo một connection mới đến MS SQL Server. Tiếp tục chọn New... để tạo một connection mới.



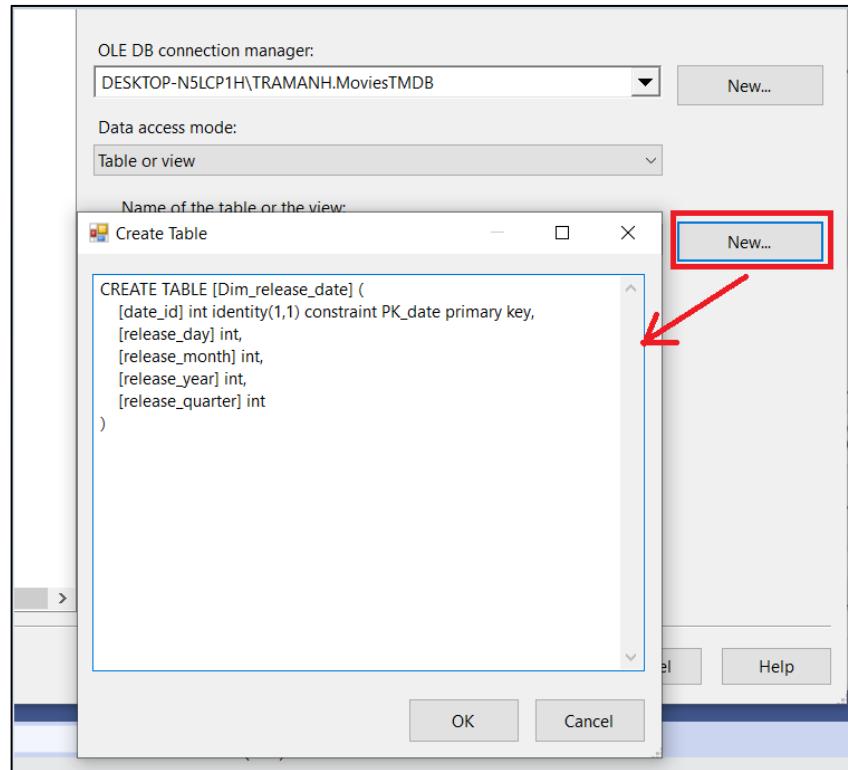
Chọn tên server name trùng với server name MS SQL Server để ta có thể kết nối đến datawarehouse MoviesTMDB vừa tạo. Kết nối đến server bằng tài khoản window mặc định (Windows Authentication) Nhấn Test Connection để kiểm tra kết nối



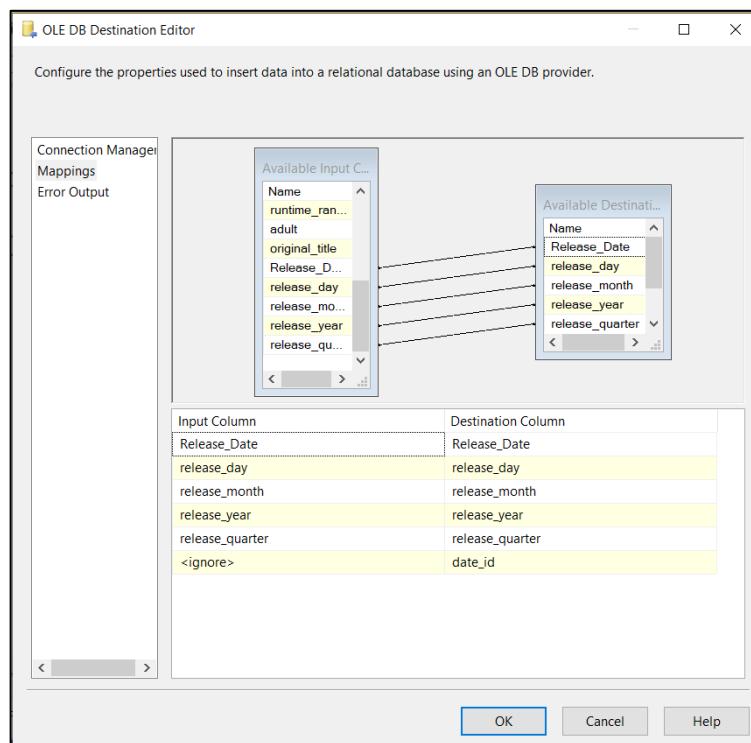


Chọn New.. để tạo mới bảng. Nội dung câu lệnh SQL tạo bảng Dim_release_date như sau:

```
CREATE TABLE [Dim_release_date] (
    [date_id] int identity(1,1) constraint PK_date primary key,
    [release_date] datetime,
    [release_day] int,
    [release_month] int,
    [release_year] int,
    [release_quarter] int
)
```

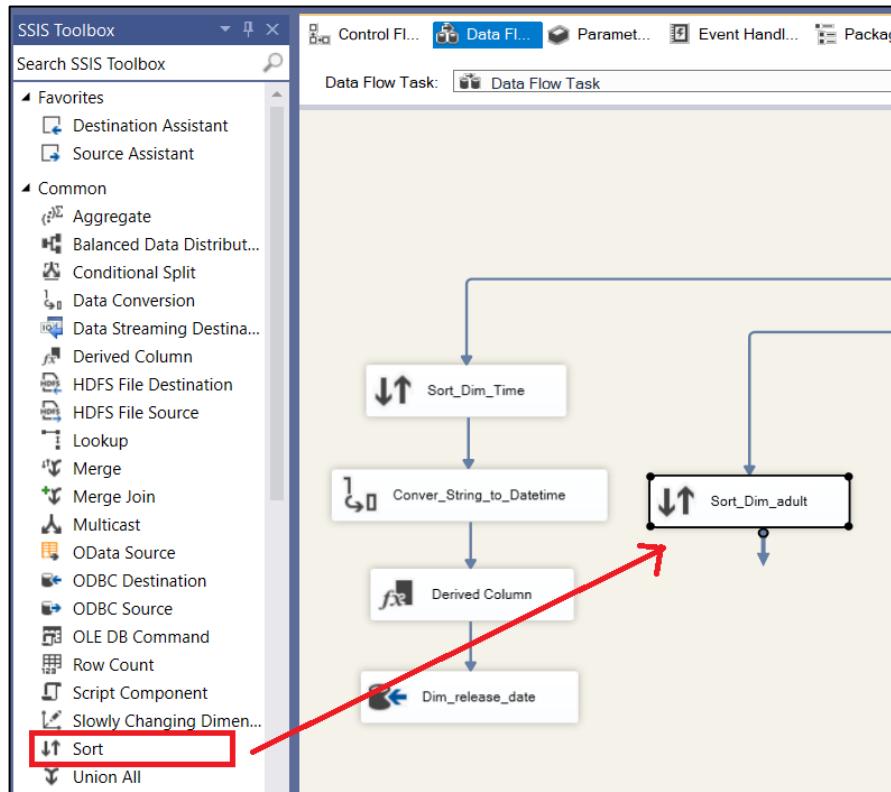


Tiếp đến ta cần chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu.
Chọn OK để hoàn tất thiết lập.

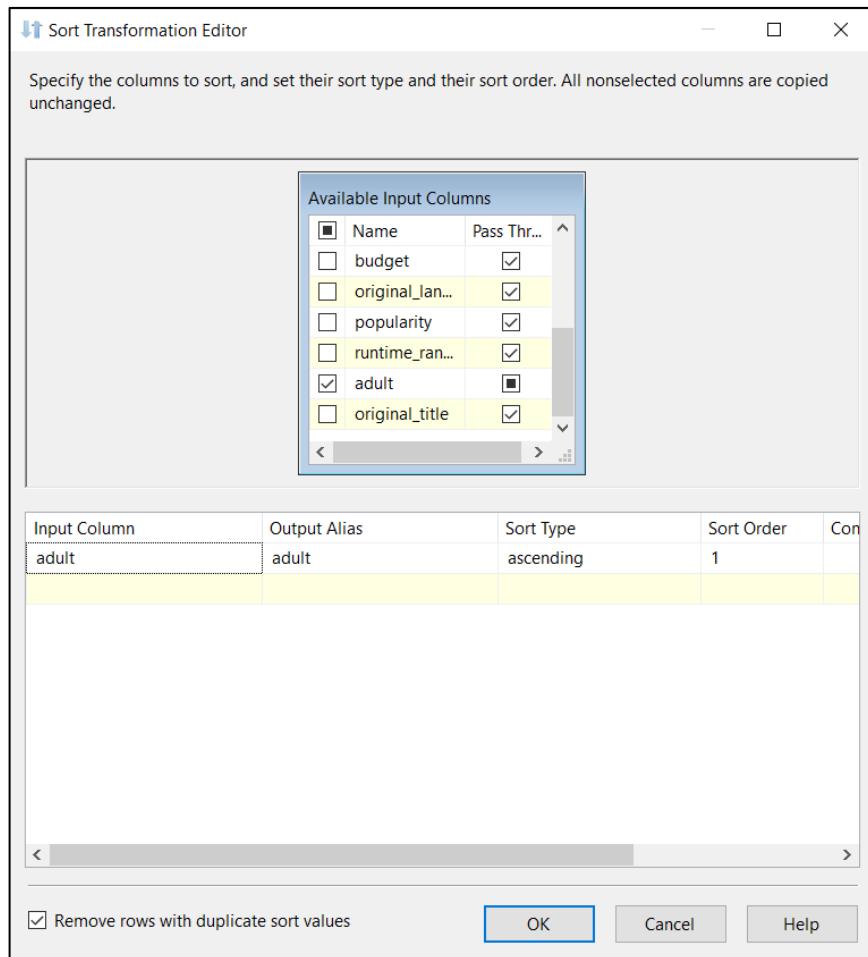


2.5.2. Bảng Dim_adult

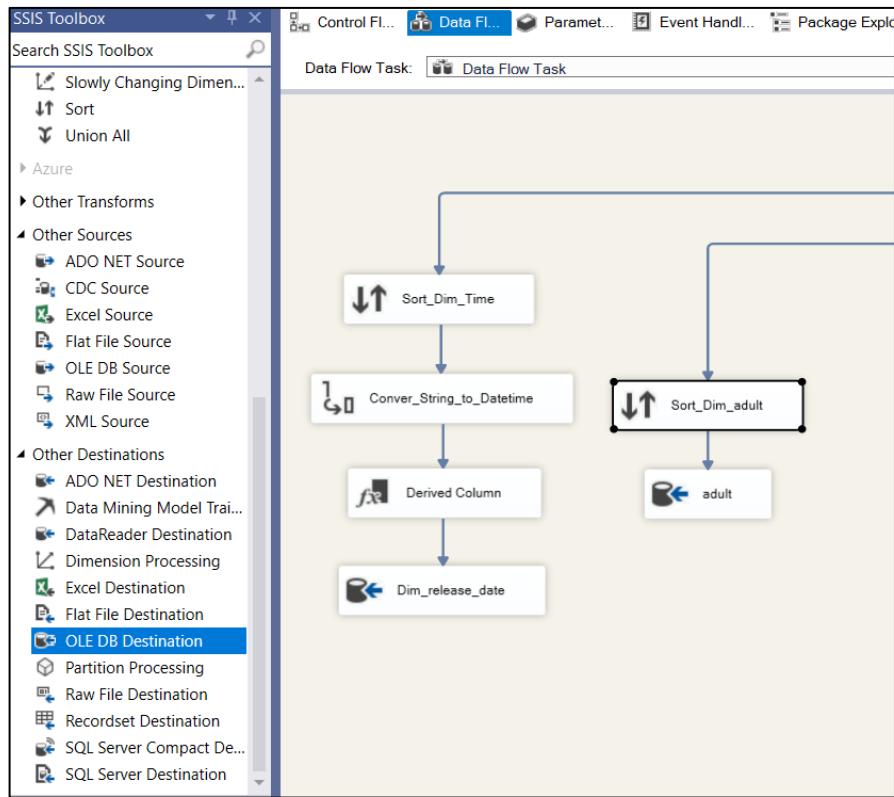
Bước 1: Chọn một Sort để tạo ra Sort_Dim_adult cho Dim_adult



Bước 2: Click chuột phải vào Sort_Dim_adult, chọn Edit: chọn cột adult làm các cột để để dữ liệu vào Sort_Dim_adult. Tick chọn Remove rows with duplicate sort values xóa đi các dòng dữ liệu trùng nhau và sau đó chọn OK.



Bước 3: Tạo mới một OLE DB Destination để đổ dữ liệu gốc sau khi đã được xử lý.

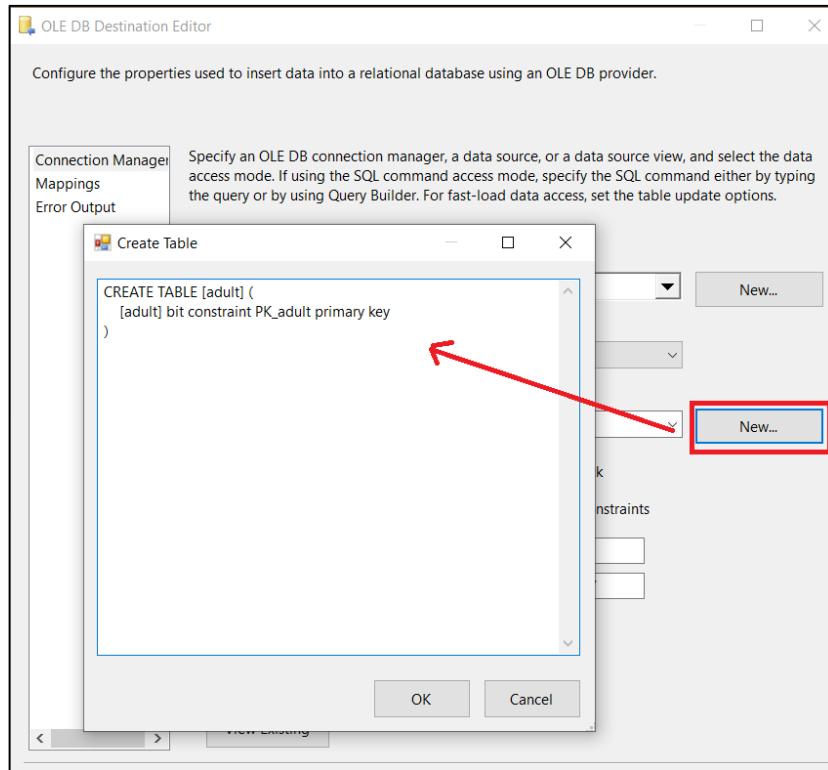


Bước 4: Connection đến kho dữ liệu đã được tạo khi tạo Dim_release_date, vì vậy ta chỉ cần chọn New... để tạo bảng Dim_adult.

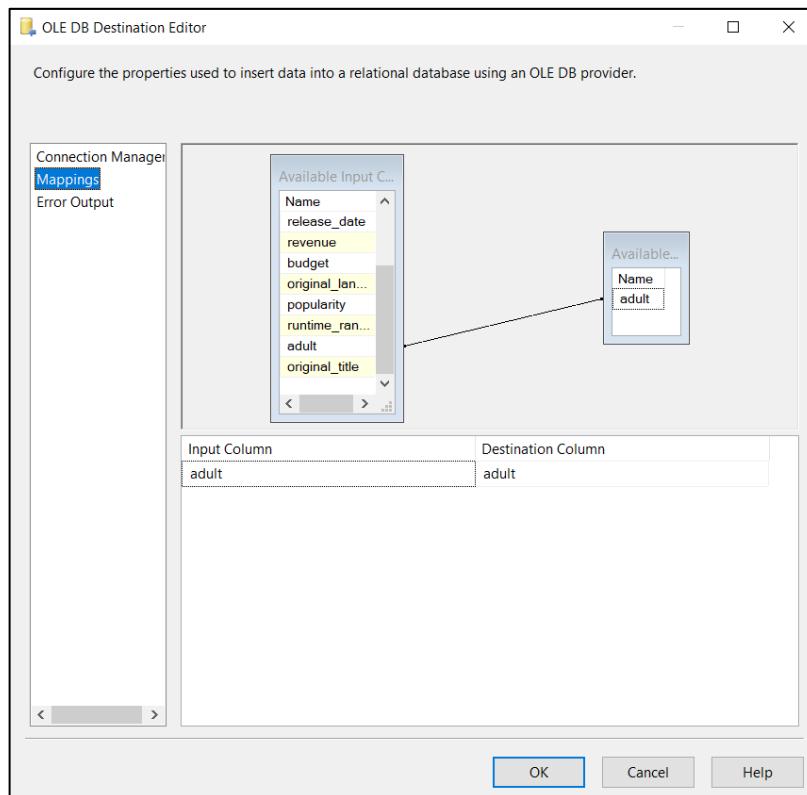
CREATE TABLE [dim_adult] (

[adult] bit constraint PK_adult primary key

)

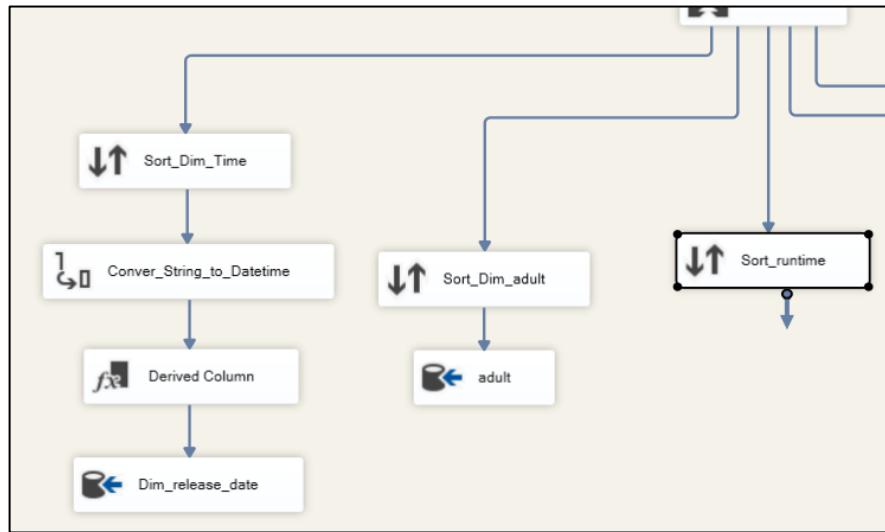


Bước 6: Tiếp đến ta cần chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu. Chọn OK để hoàn tất thiết lập



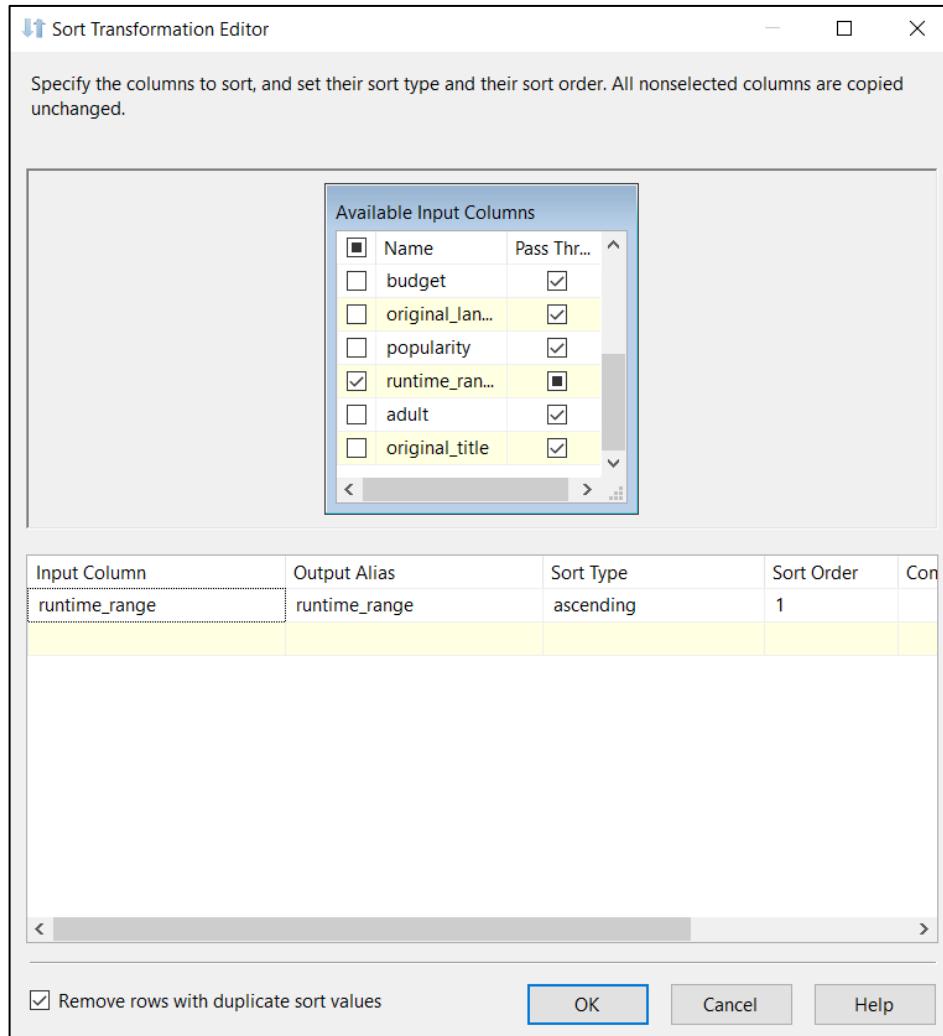
2.5.3. Bảng Dim_runtime

Bước 1: Chọn một Sort để tạo ra Sort_Dim_runtime cho Dim_runtime.

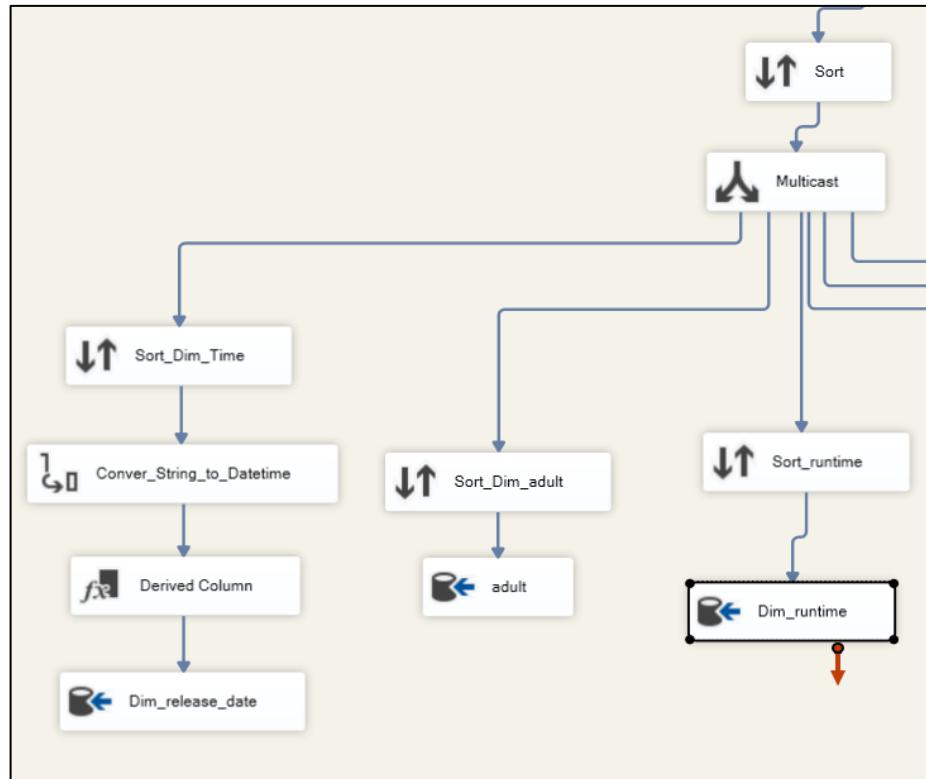


Bước 2: Click chuột phải vào Sort_Dim_runtime, chọn Edit: chọn runtime_range làm các cột để đổ dữ liệu vào Sort_Dim_runtime.

Tick chọn Remove rows with duplicate sort values xóa đi các dòng dữ liệu trùng nhau và sau đó chọn OK.



Bước 3: Tạo mới một OLE DB Destination để đổ dữ liệu gốc sau khi đã được xử lý vào trong bảng Dim_runtime.



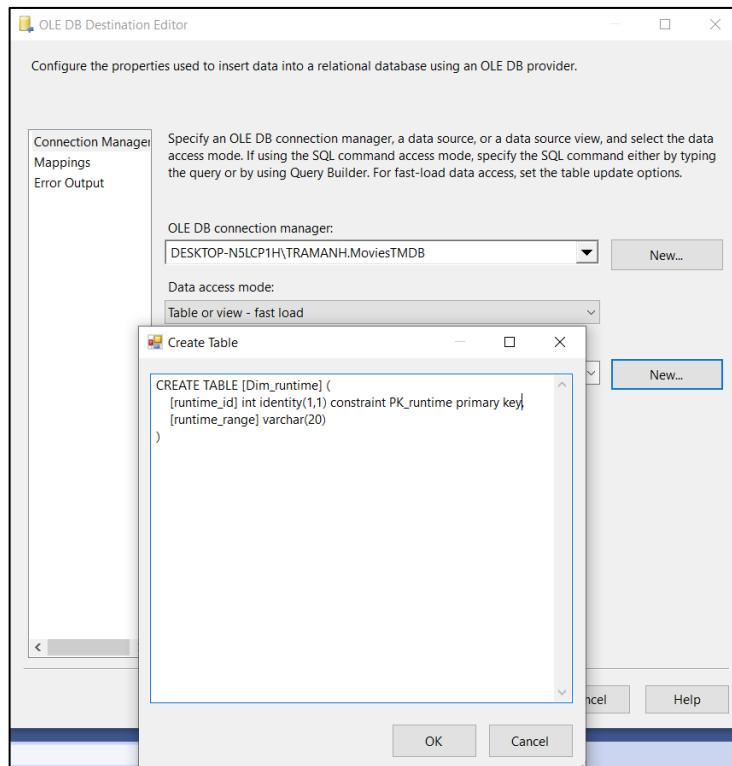
Bước 4: Connection đến kho dữ liệu đã được tạo khi tạo Dim_release_date, vì vậy ta chỉ cần chọn New... để tạo bảng Dim_runtime.

```
CREATE TABLE [Dim_runtime] (
```

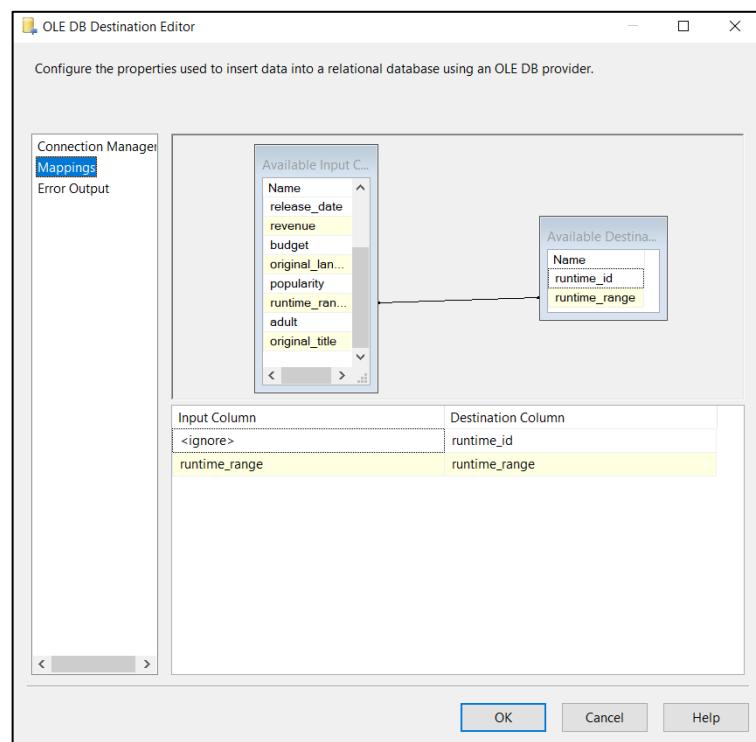
```
    [runtime_id] int identity(1,1) constraint PK_runtime primary key,
```

```
    [runtime_range] varchar(20)
```

```
)
```

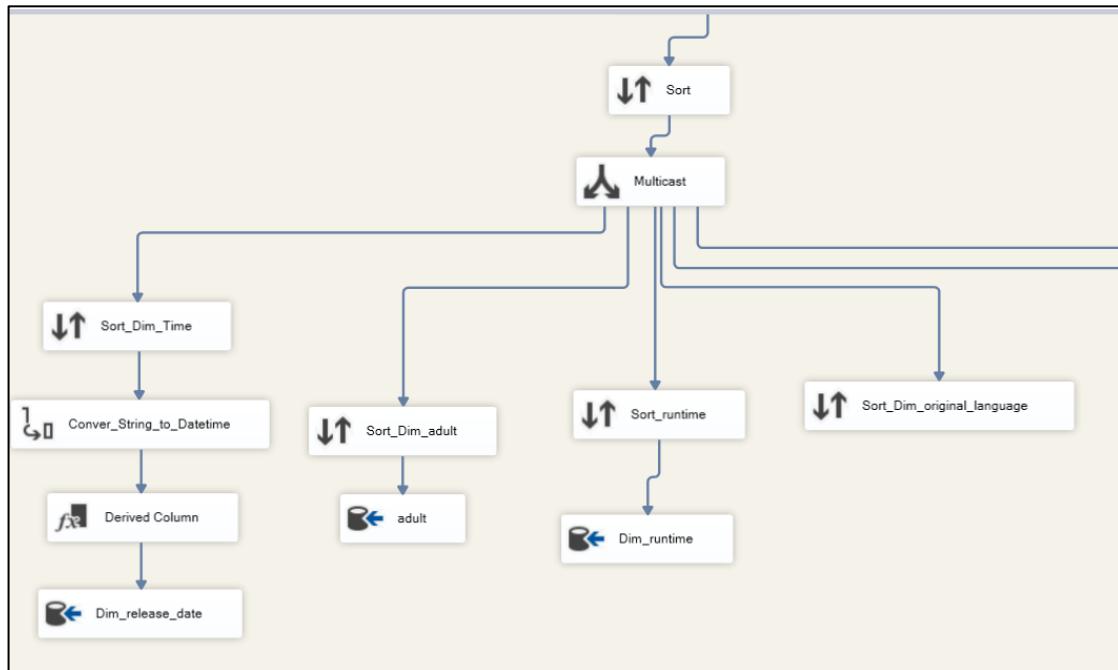


Bước 5: Tiếp đến ta cần chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu. Chọn OK để hoàn tất thiết lập.



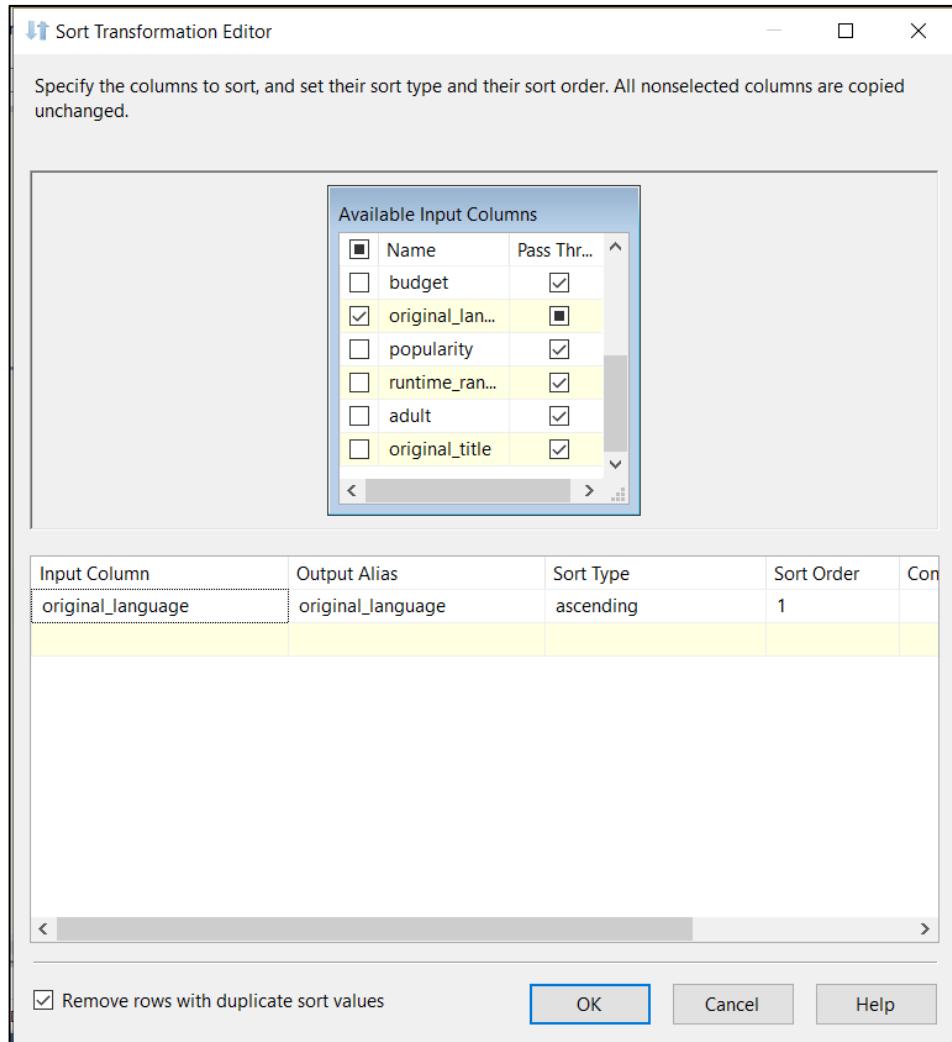
2.5.4. Bảng Dim_language

Bước 1: Chọn một Sort để tạo ra Sort_Dim_original_language cho Dim_original_language.

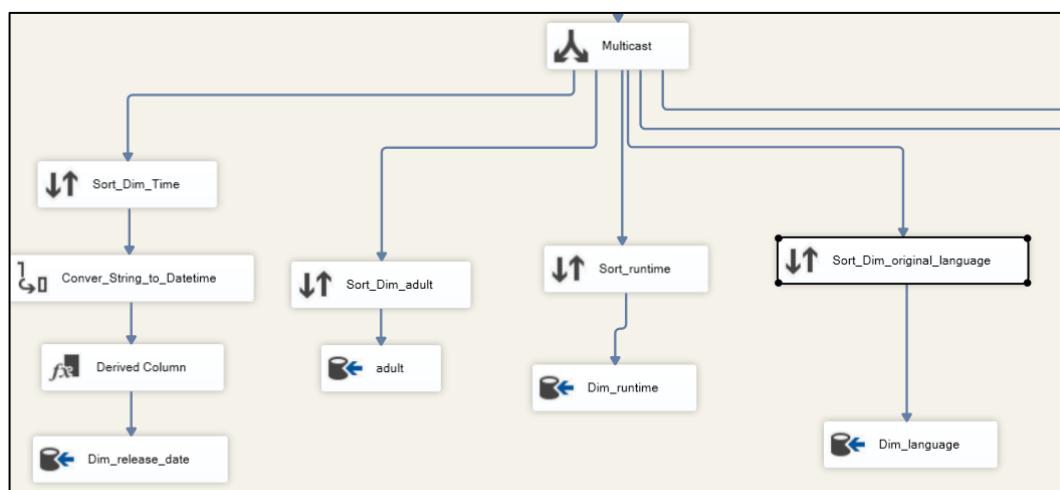


Bước 2: Click chuột phải vào Sort_Dim_original_language, chọn Edit: chọn original_language làm các cột để đổ dữ liệu vào Sort_Dim_original_language.

Tick chọn Remove rows with duplicate sort values xóa đi các dòng dữ liệu trùng nhau và sau đó chọn OK.



Bước 3: Tạo mới một OLE DB Destination để đổ dữ liệu gốc sau khi đã được xử lý vào trong bảng Dim_Airport kho dữ liệu US_ACCIDENTS_WH.



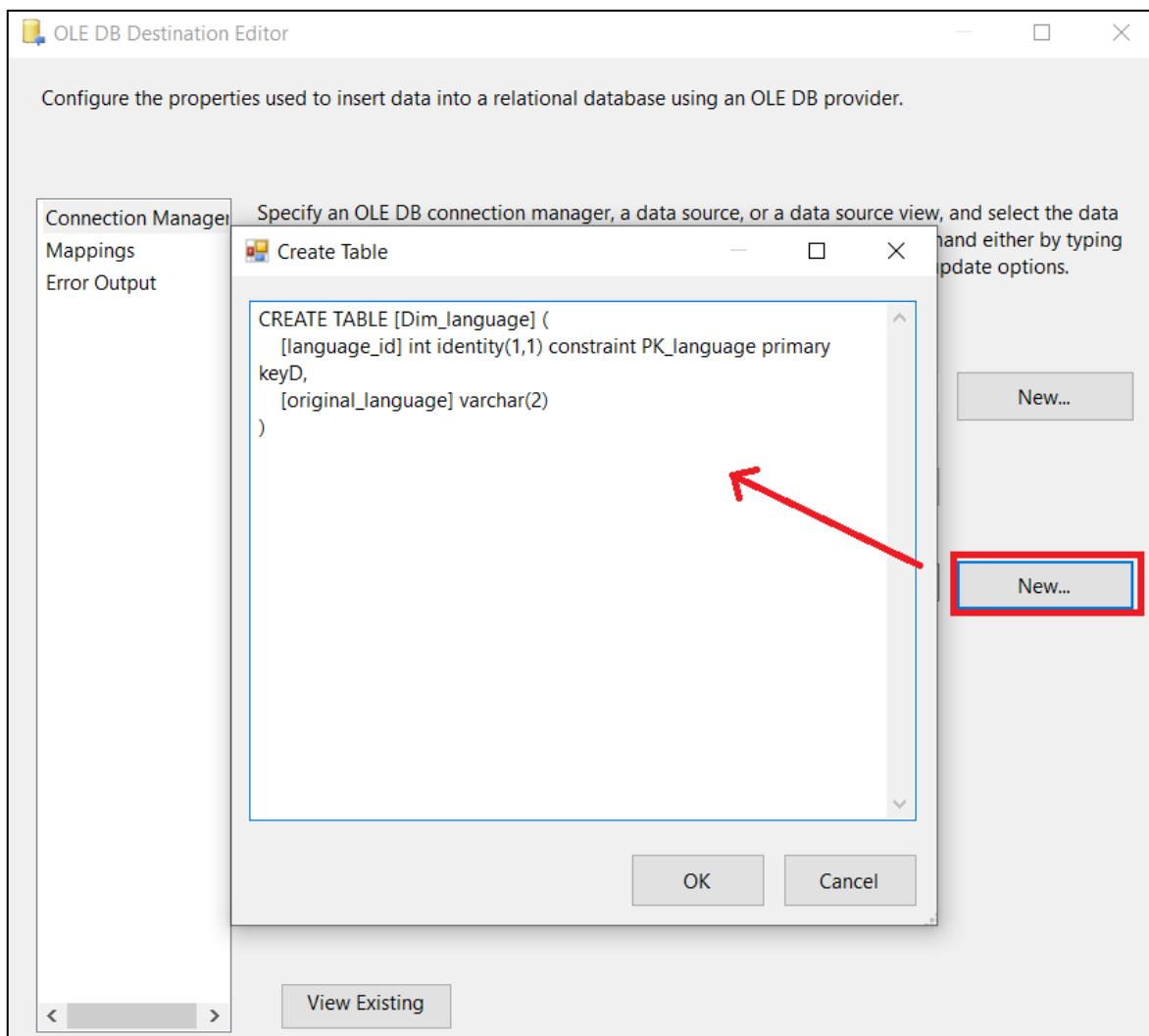
Bước 4: Connection đến kho dữ liệu đã được tạo khi tạo Dim_release_date, vì vậy ta chỉ cần chọn New... để tạo bảng Dim_original_language.

```
CREATE TABLE [Dim_language] (
```

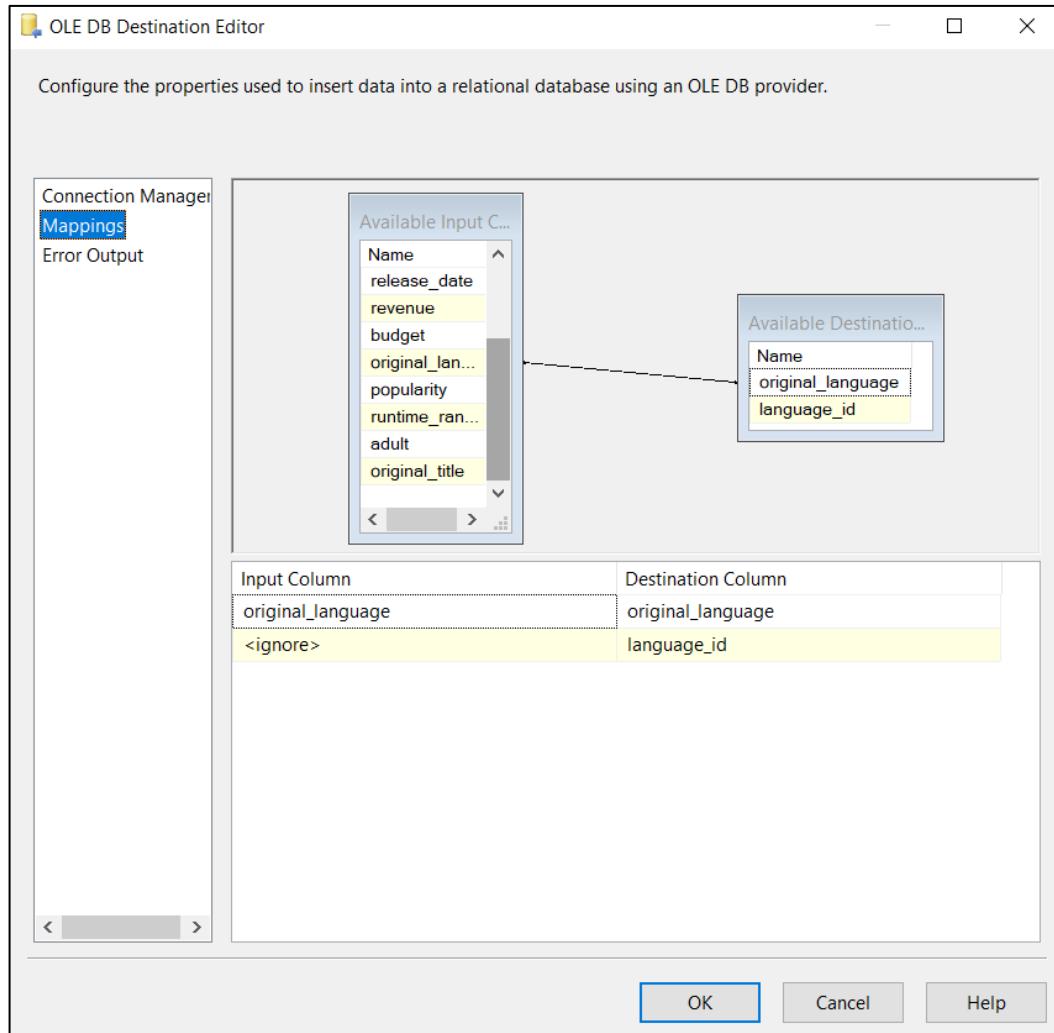
```
    [language_id] int identity(1,1) constraint PK_language primary keyD,
```

```
    [original_language] varchar(2)
```

```
)
```



Bước 5: Tiếp đến ta cần chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu. Chọn OK để hoàn tất thiết lập.



*** Lưu ý:**

Trong file .csv, bộ dữ liệu MoviesTMDB có 2 vấn đề:

- TH1: Các dòng giống nhau tất cả các trường dữ liệu
- TH2: Các dòng cùng chung ID phim, tên phim nhưng trường dữ liệu khác nhau.

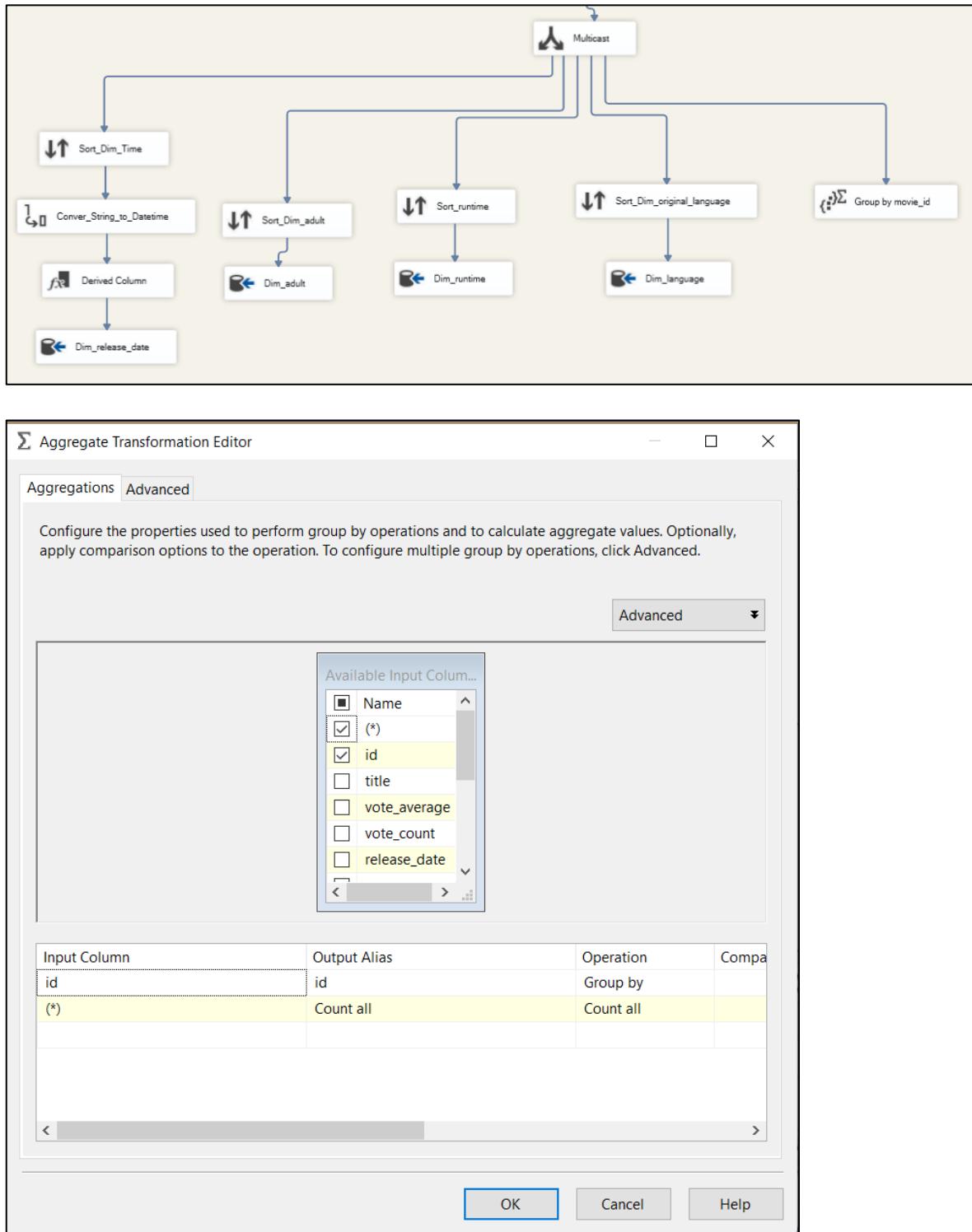
Do đó, chúng ta cần thao tác:

- Xóa tất cả các dòng giống nhau trong TH1.
- Xóa và giữ lại duy nhất 1 dòng trong TH2.

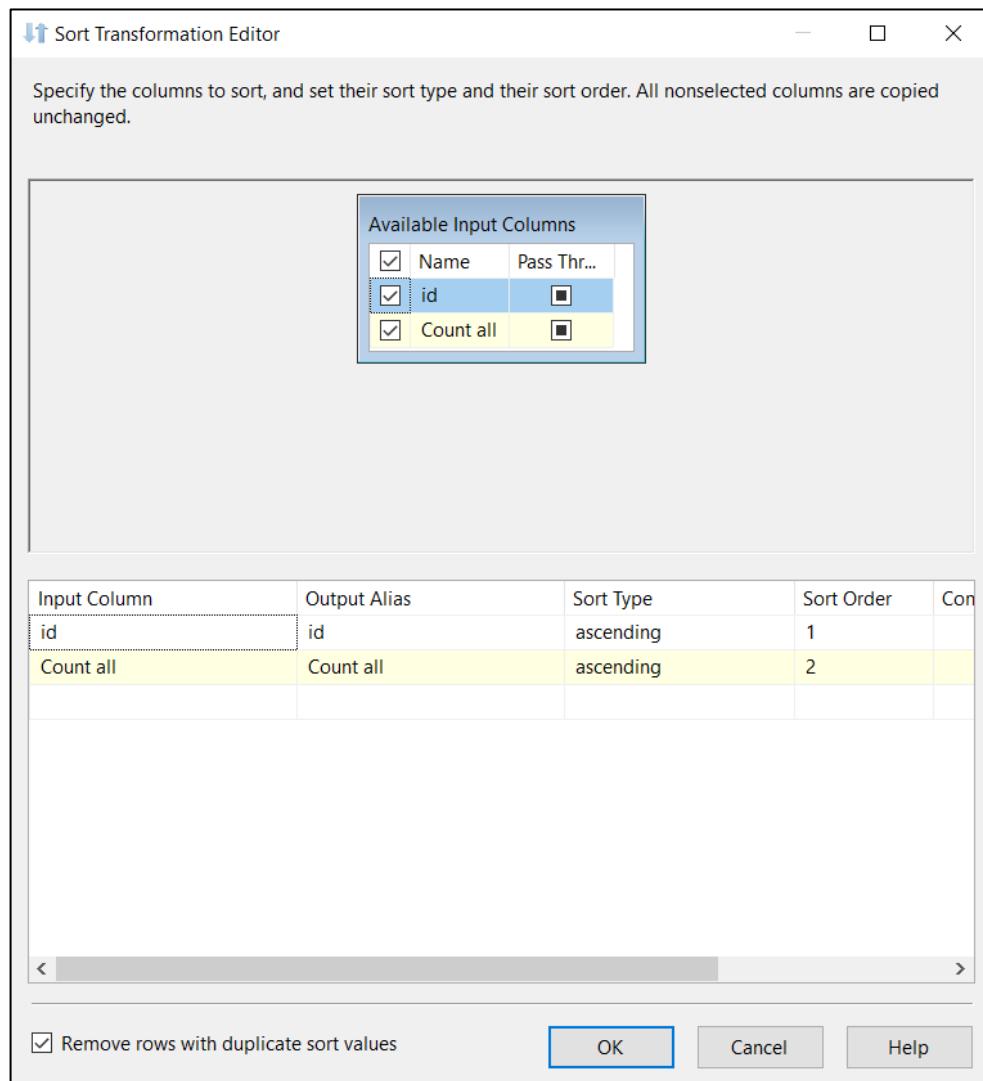
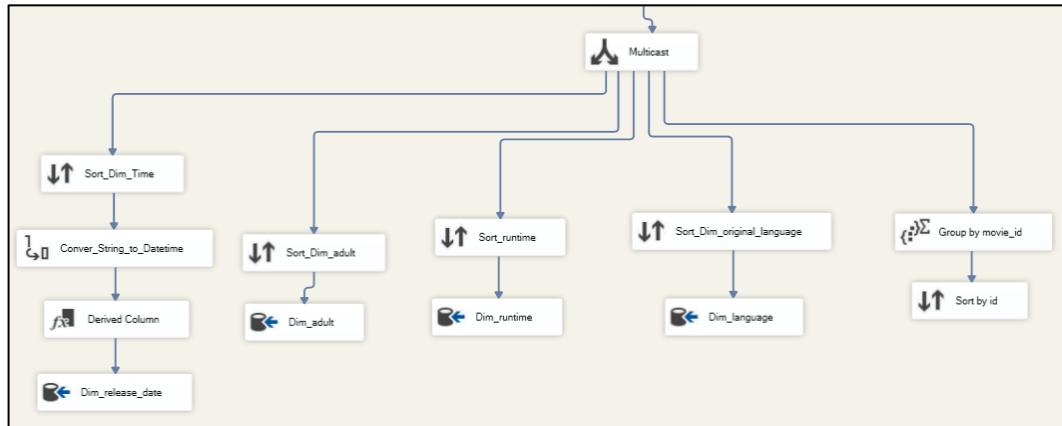
Trong quá trình đổ bảng Dim_movies và Fact_raw, chúng ta **thực hiện xử lý dữ liệu trước** rồi sử dụng Multicast để đổ vào bảng Dim_movies và Fact_raw.

2.5.5. Bảng Dim_movies

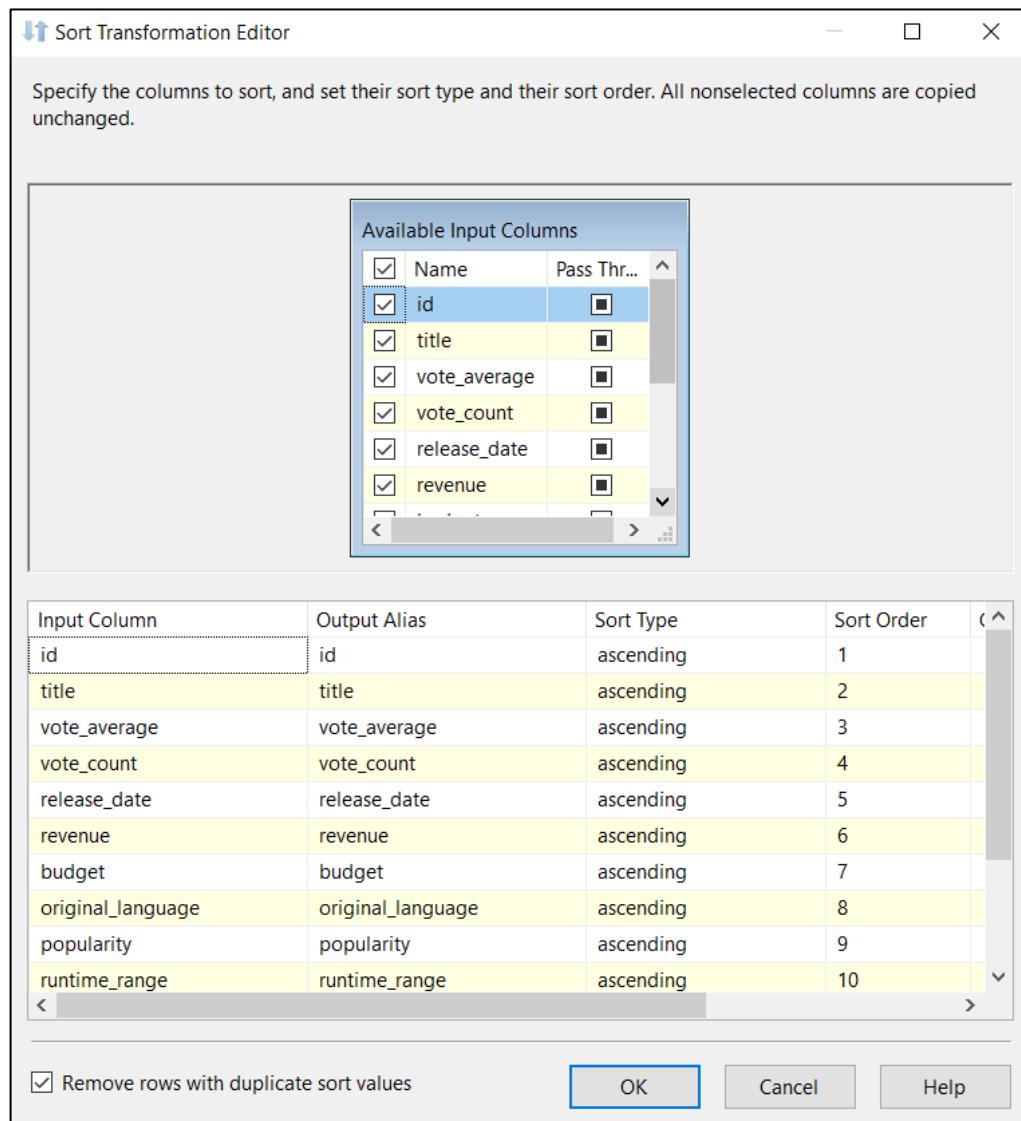
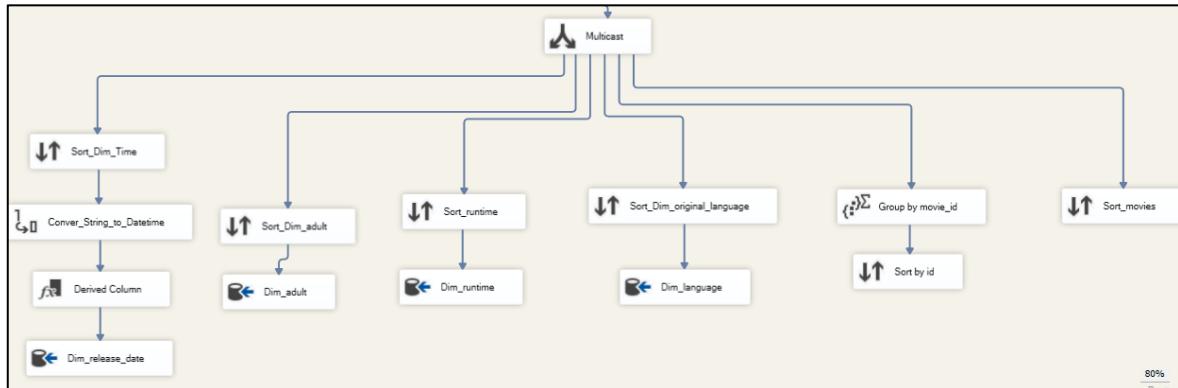
Bước 1: Chọn một Aggregate để thực hiện Group by theo ID. Thêm cột Count all đếm số dòng cùng ID và ấn OK.



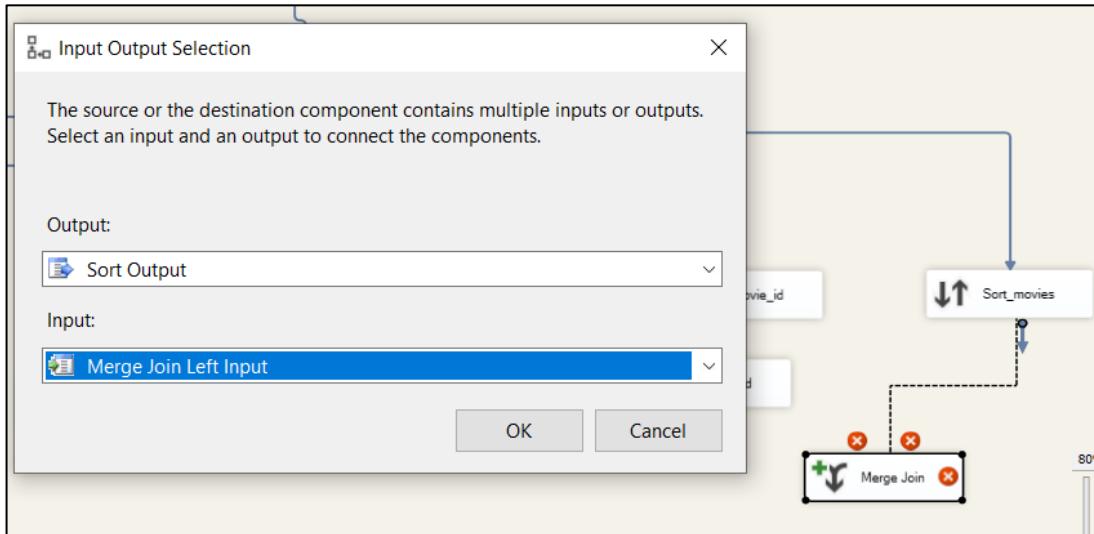
Bước 2: Chọn một Sort để tạo ra Sort by id để xóa hết các id, Count all giống nhau. Chọn Edit: chọn id và Count all. Tick chọn Remove rows with duplicate sort values xóa đi các dòng dữ liệu trùng nhau và sau đó chọn OK.



Bước 3: Chọn một Sort để tạo ra Sort_movie. Chọn Edit: chọn tất cả thuộc tính. Tick chọn Remove rows with duplicate sort values xóa đi các dòng dữ liệu trùng nhau và sau đó chọn OK.

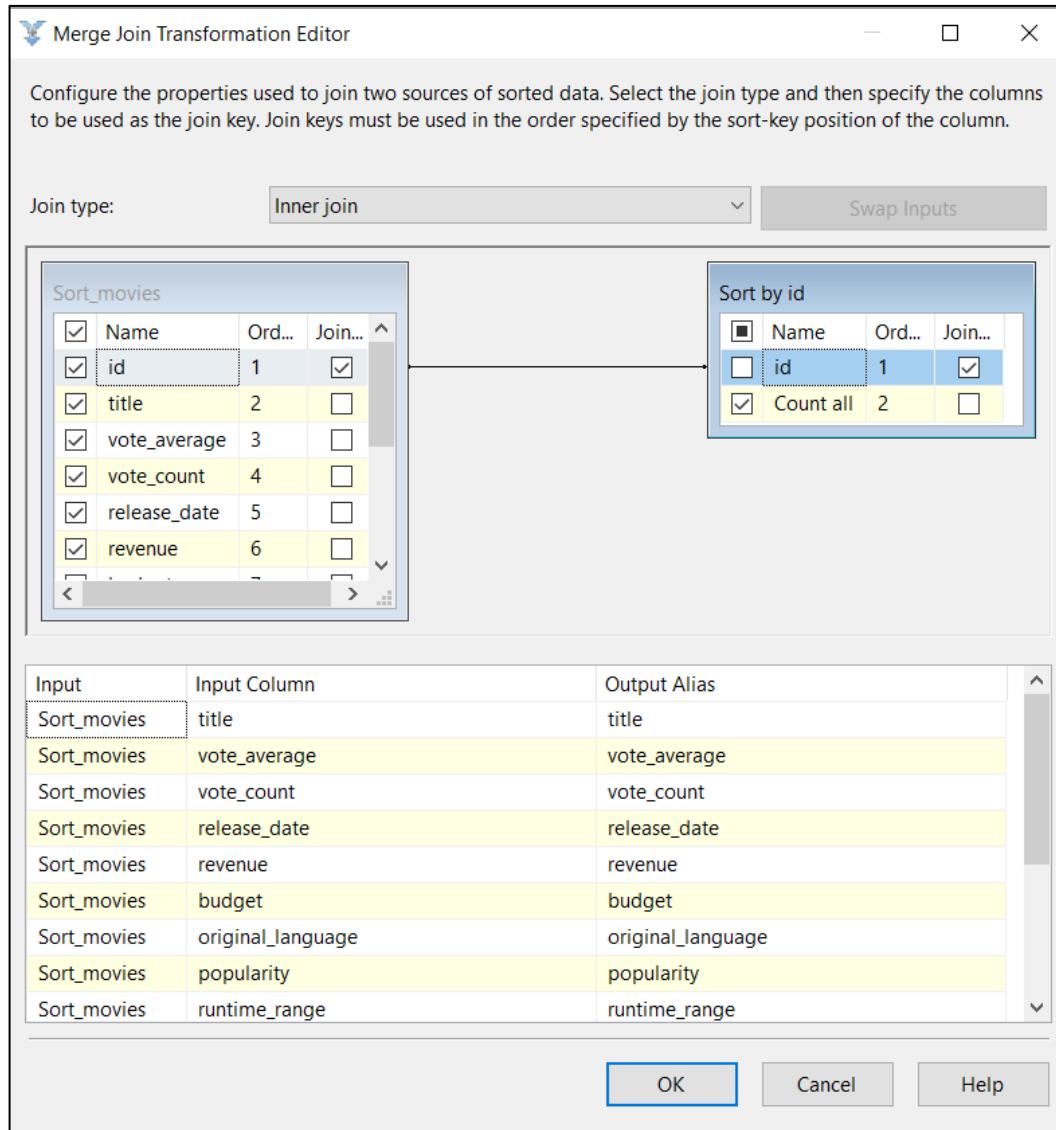


Bước 4: Tạo một Merge Join và nối với Sort_movies, tiếp theo chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong file .csv bất kể có kết quả khi thực hiện phép kết trái với cột id của Sort id hay không.

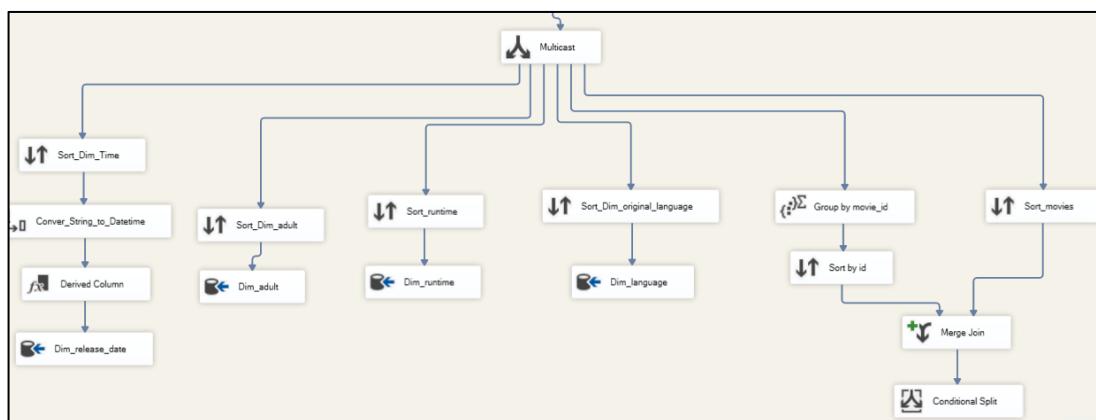


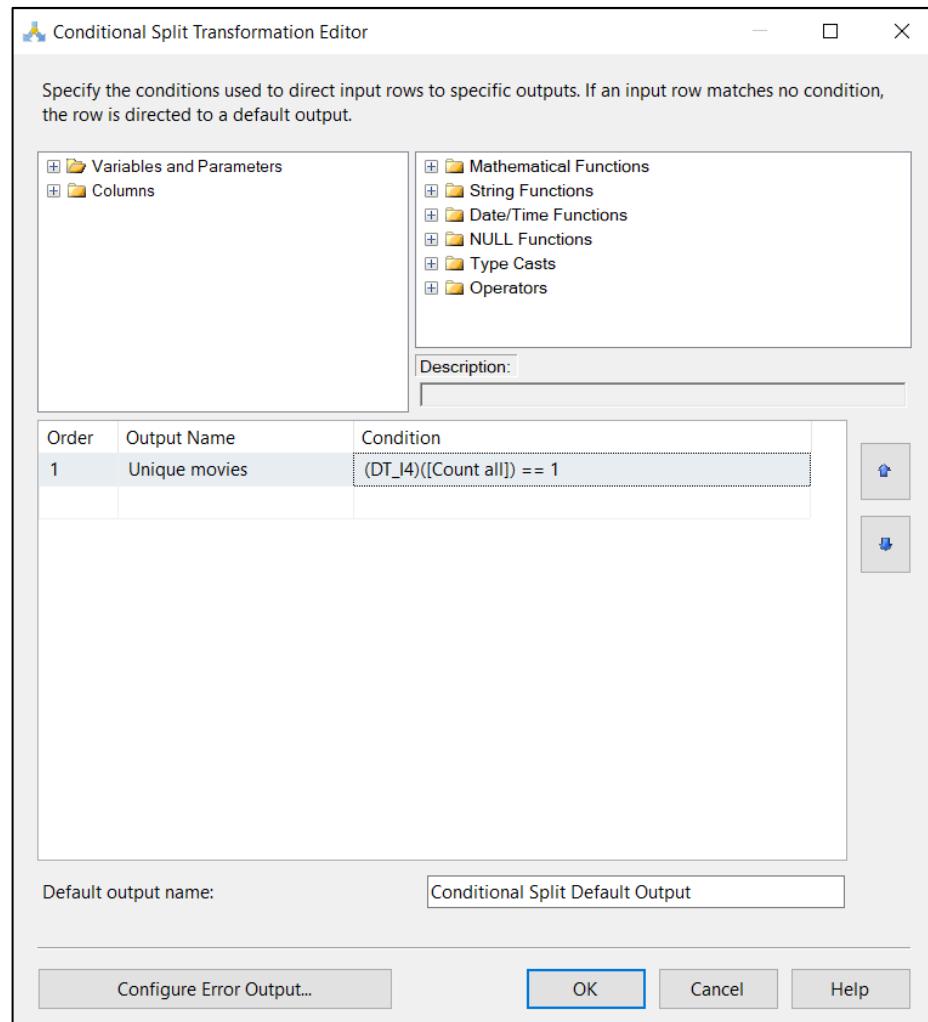
Bước 5: Sau đó, tạo kết nối giữa Sort id với Merge Join. Chuột phải vào Merge Join và nhấn Edit, một hộp thoại merge editor xuất hiện: ở đây ta tick chọn tất cả các cột của Sort_movie nhưng không lấy thuộc tính id của Sort by id.

Kết quả sau khi merge là bảng chứa hết thuộc tính của Sort_movies và thêm Count all.

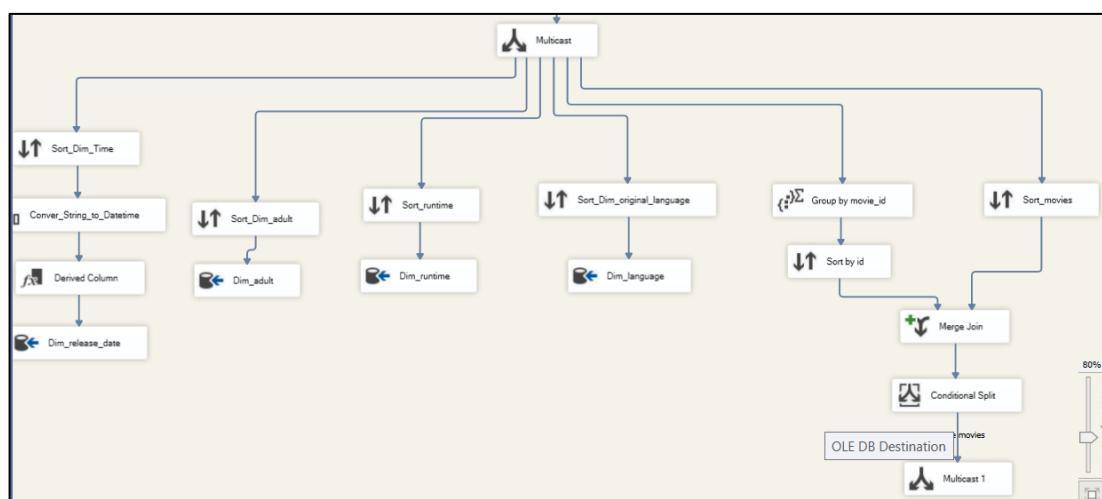


Bước 6: Chúng tôi chỉ lấy các giá trị duy nhất (unique movies) do đó, tạo một Conditional Split. Đặt điều kiện (DT_I4)([Count all]) == 1. Và ấn OK.



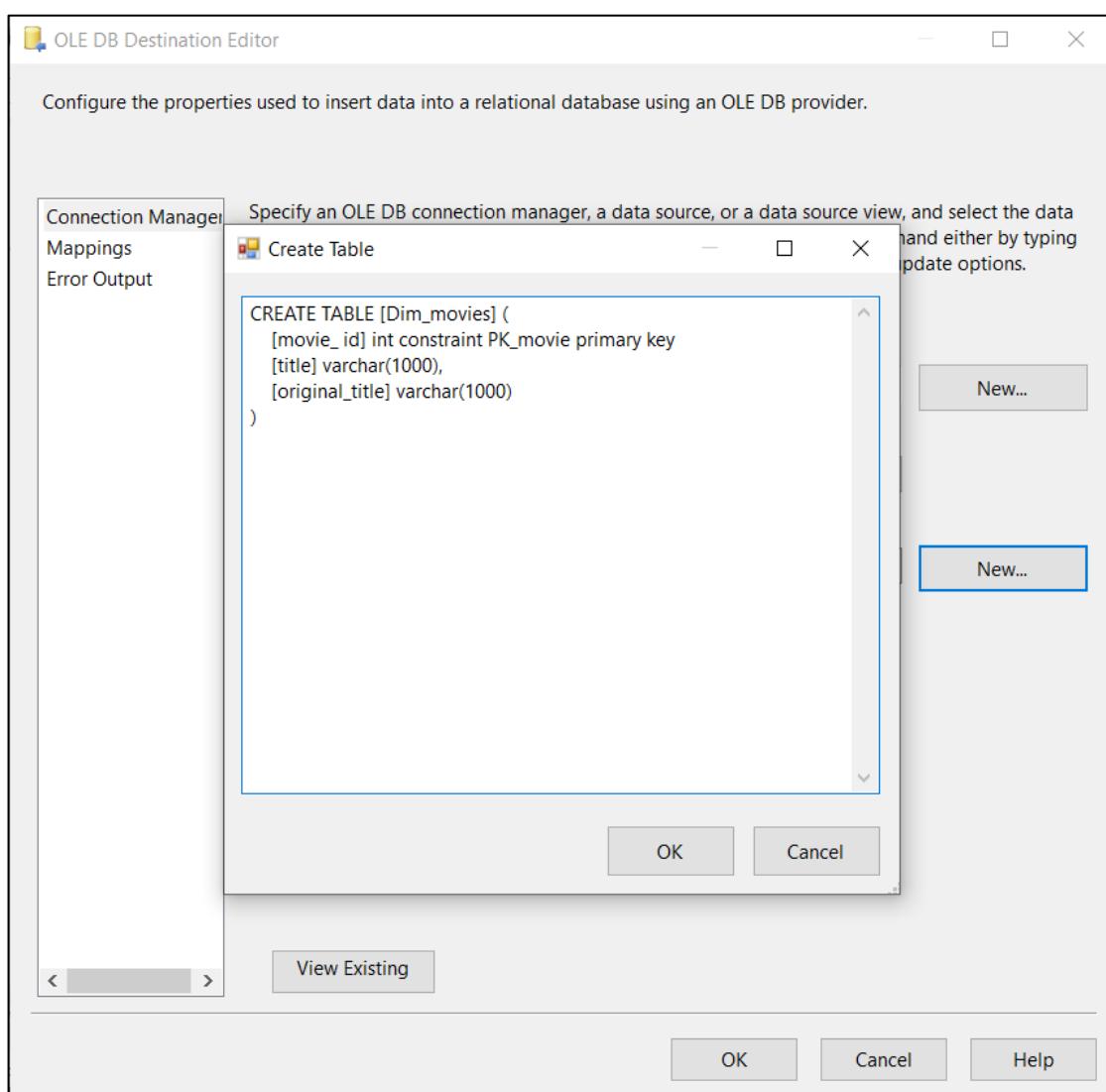


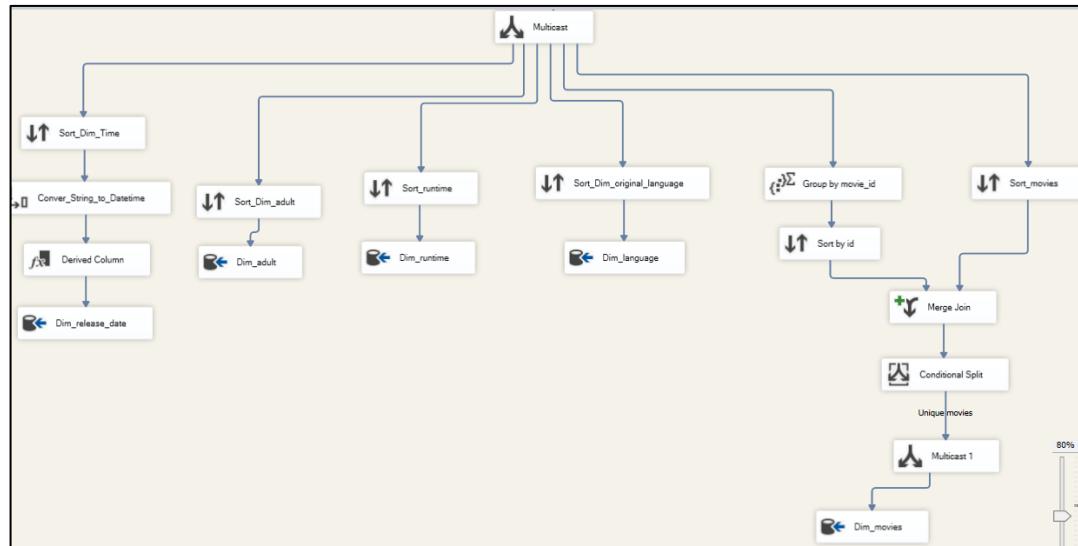
Bước 7: Tạo Multicast để phân tán dữ liệu từ Merge Join đến Dim_movies.
Tiến hành kết nối Merge Join và Multicast.



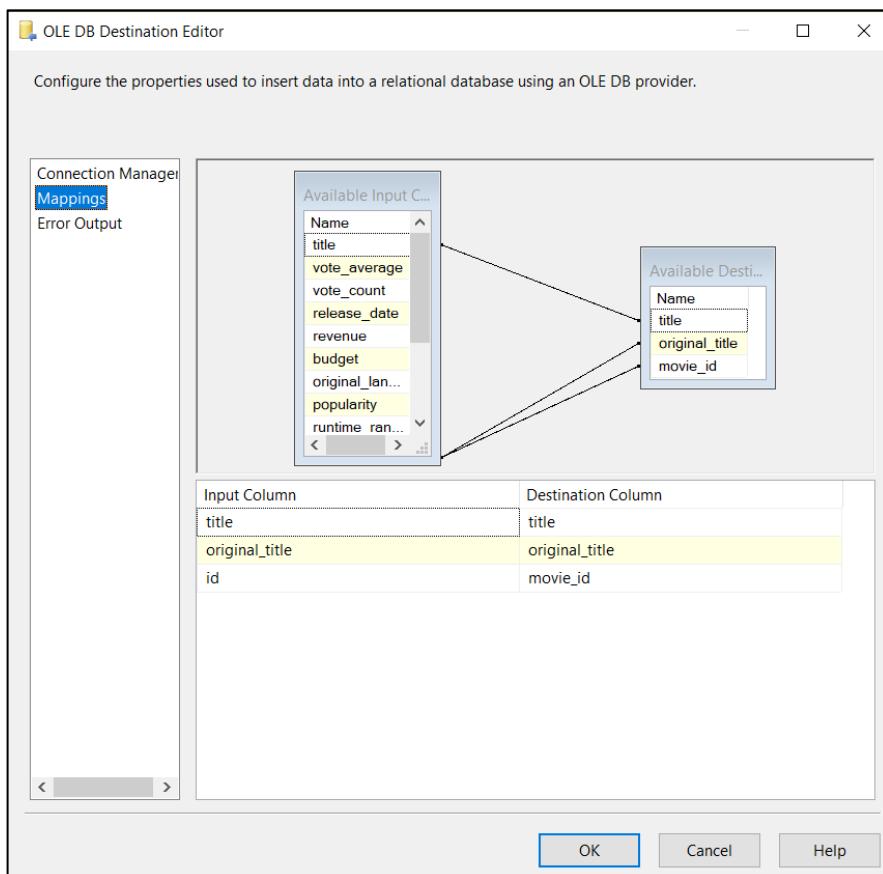
Bước 8: Connection đến kho dữ liệu đã được tạo khi tạo Dim_Time, vì vậy ta chỉ cần chọn New... để tạo bảng Dim_movies.

```
CREATE TABLE [Dim_movies] (
    [movie_id] int constraint PK_movie primary key,
    [title] varchar(1000),
    [original_title] varchar(1000)
)
```





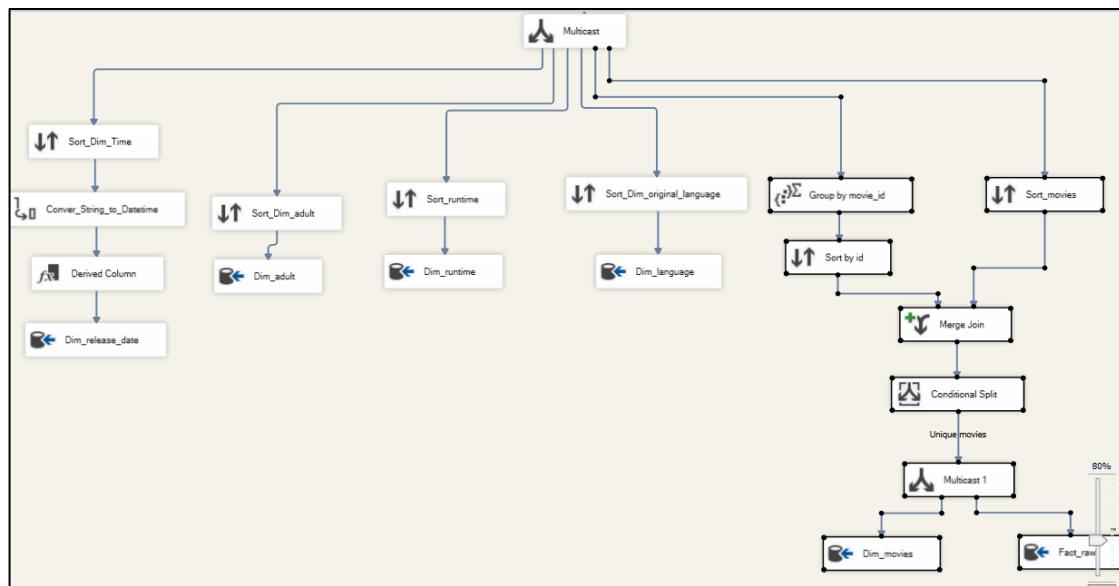
Bước 9: Tiếp đến ta cần chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu. Chọn OK để hoàn tất thiết lập.



2.5.6. Bảng Fact

* **Lưu ý:** Vì bước xử lý dữ liệu đã được thực hiện ở bảng Dim_movies, nên chúng ta chỉ cần thực hiện thao tác đổ dữ liệu vào bảng Fact_raw thông qua Multicast.

Bước 1: Tiến hành tạo bảng Fact và đặt tên là Fact_Raw từ một OLE DB Destination.



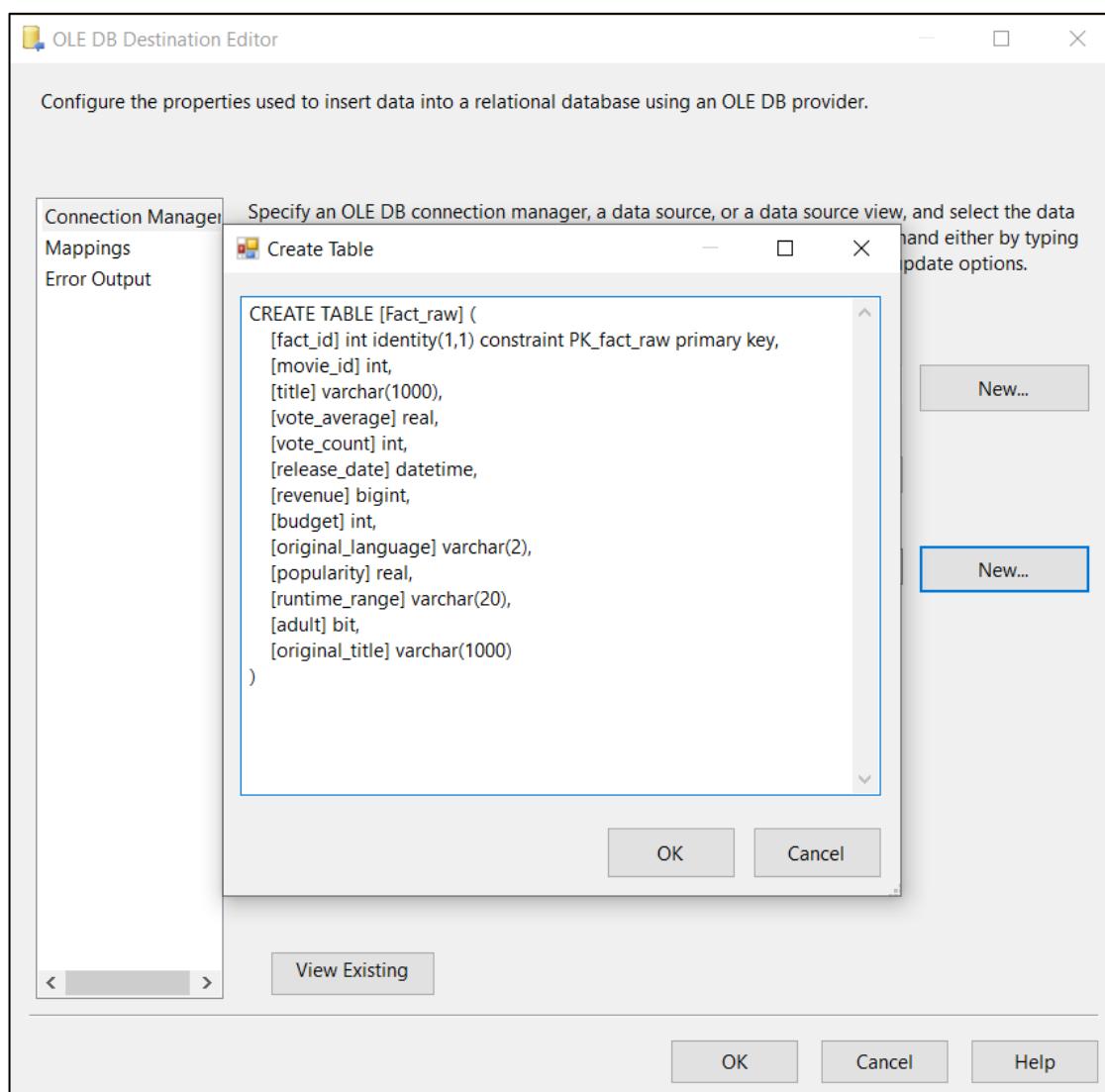
Bước 2: Click chuột phải và chọn Edit để tạo bảng Fact_Raw có các cột là tất cả các cột từ dữ liệu gốc và chứa tất cả các dòng dữ liệu.

CREATE TABLE [Fact_raw] (

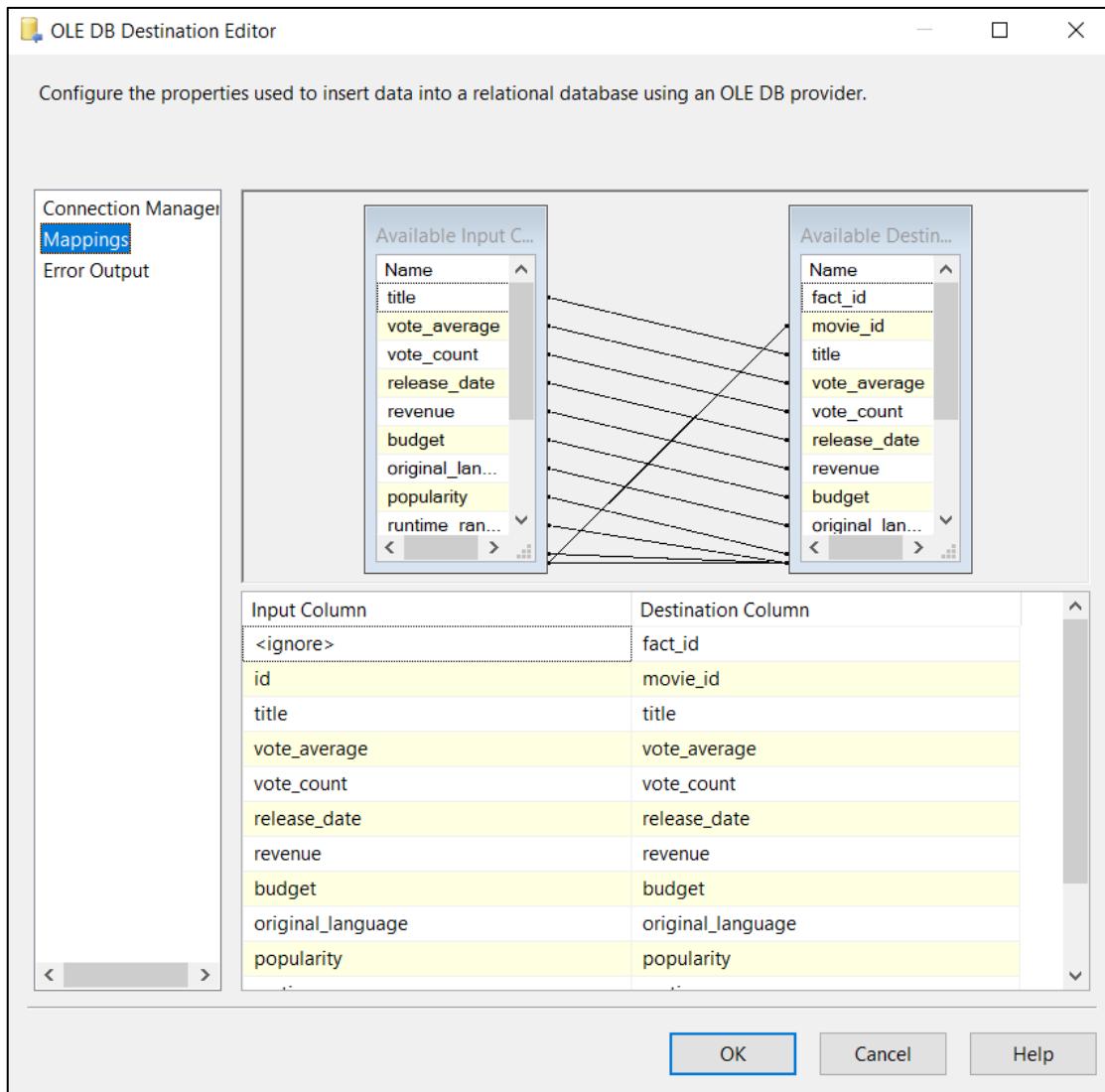
```

[fact_id] int identity(1,1) constraint PK_fact_raw primary key,
[movie_id] int,
[title] varchar(1000),
[vote_average] real,
[vote_count] int,
[release_date] datetime,
[revenue] bigint,
[budget] int,
  
```

```
[original_language] varchar(2),  
[popularity] real,  
[runtime_range] varchar(20),  
[adult] bit,  
[original_title] varchar(1000)  
)
```

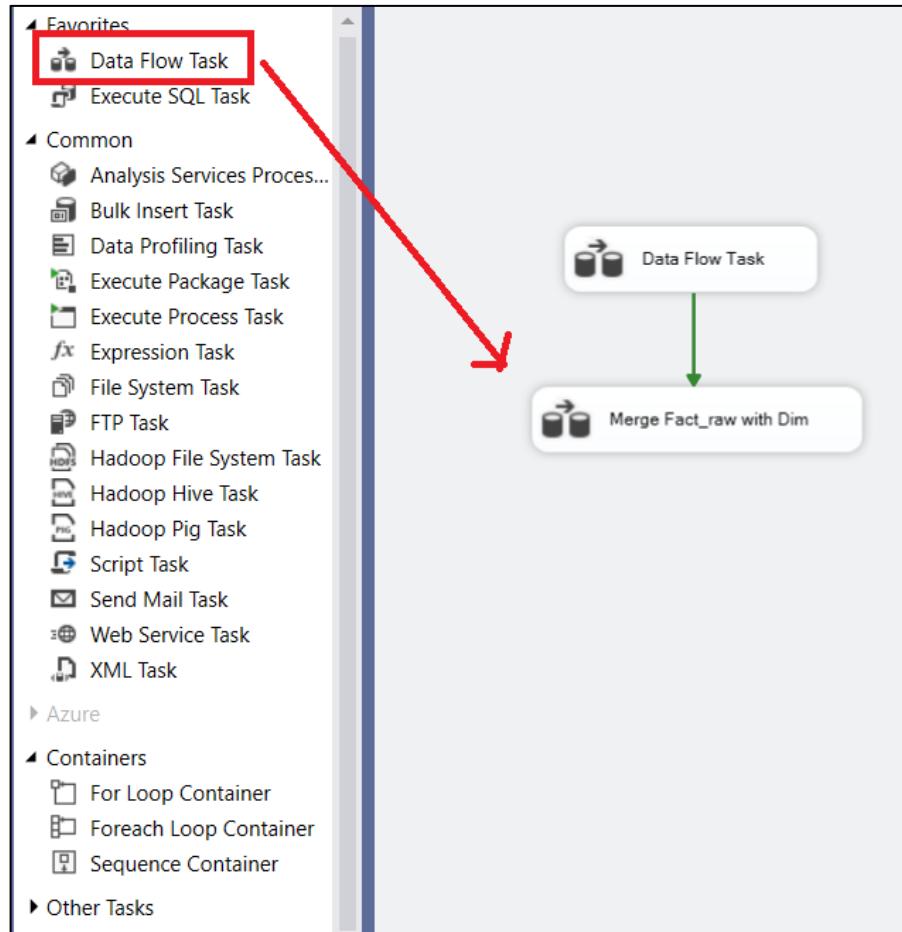


Bước 3: Tiếp đến ta cần chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu. Cuối cùng nhấn nút OK để hoàn tất quá trình tạo bảng.

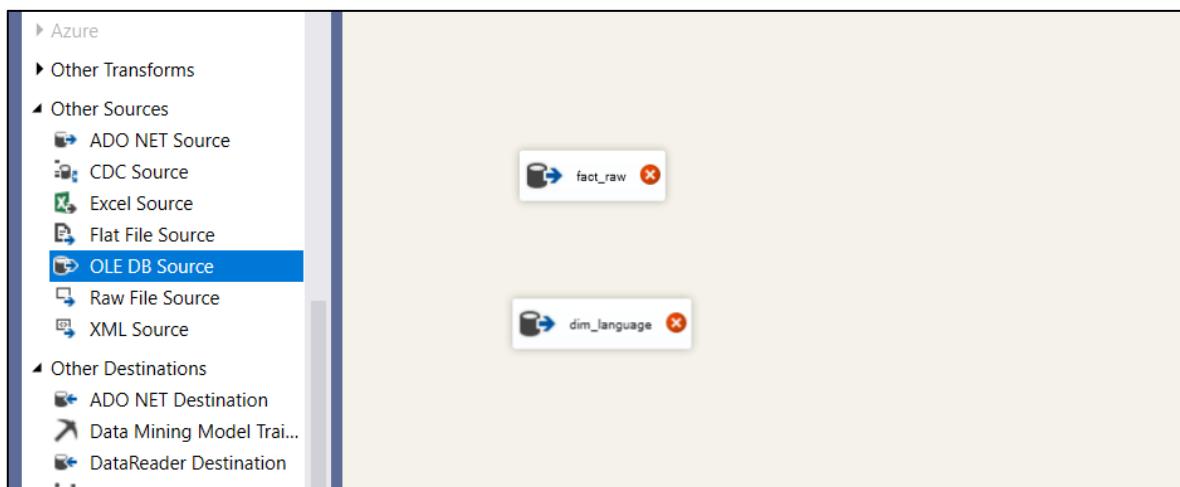


Tiếp theo đây ta sẽ thực hiện quá trình lần lượt loại bỏ các cột dữ liệu trùng của bảng Fact với các Dimension, thực hiện thêm khóa ngoại vào bảng Fact nhằm thu gọn bảng Fact, tối ưu hóa quá trình phân tích dữ liệu.

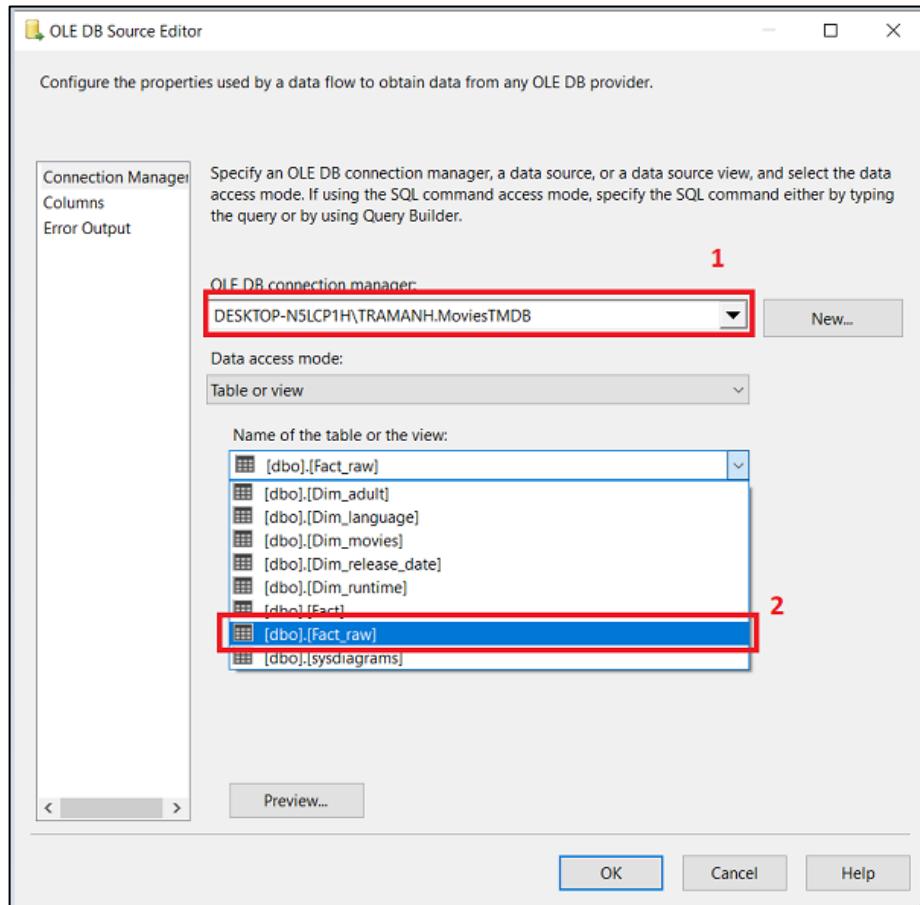
Bước 4: Tạo Data Flow Task để thực hiện quá trình Merge Fact_raw with Dim.



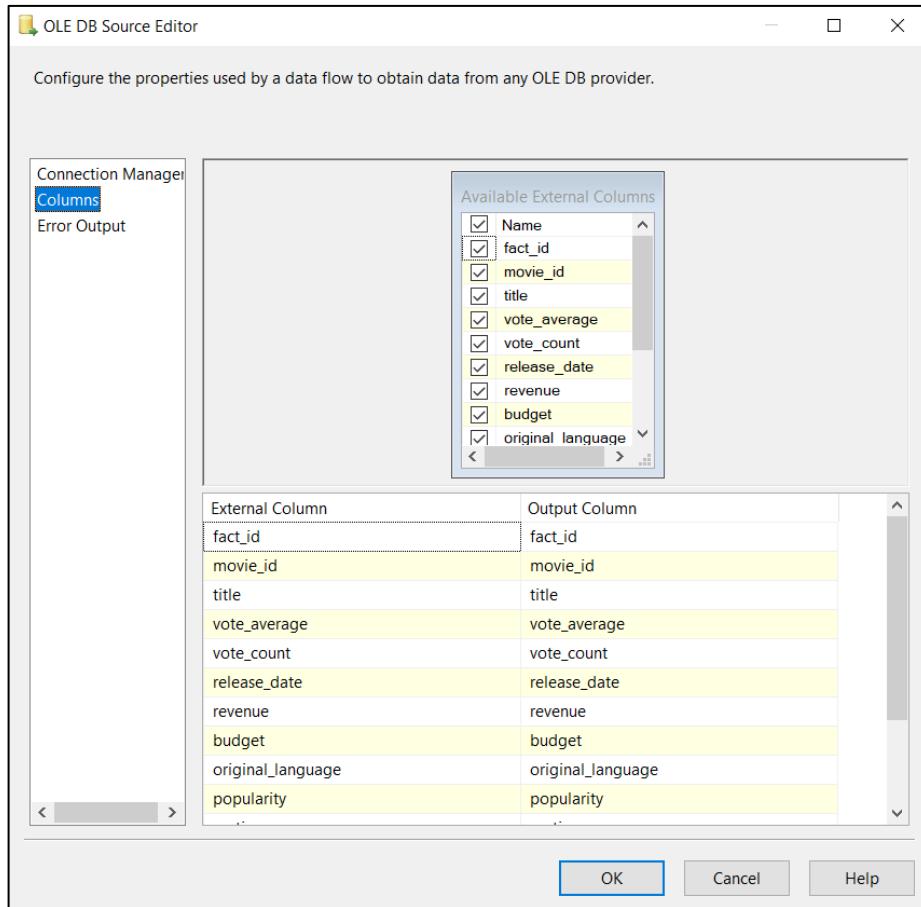
Bước 5: Click chuột phải vào Data Flow Task nói trên và chọn Edit, trong tab Data Flow ta tạo 2 OLE DB Source và đổi tên Fact_Raw và Dim_language.



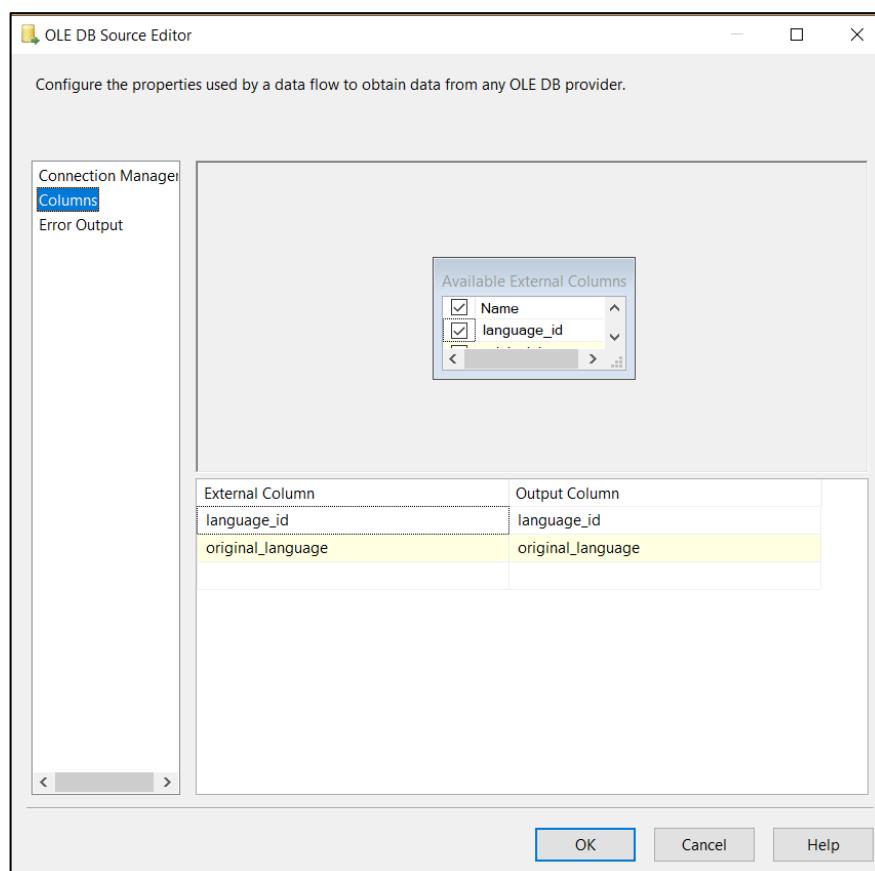
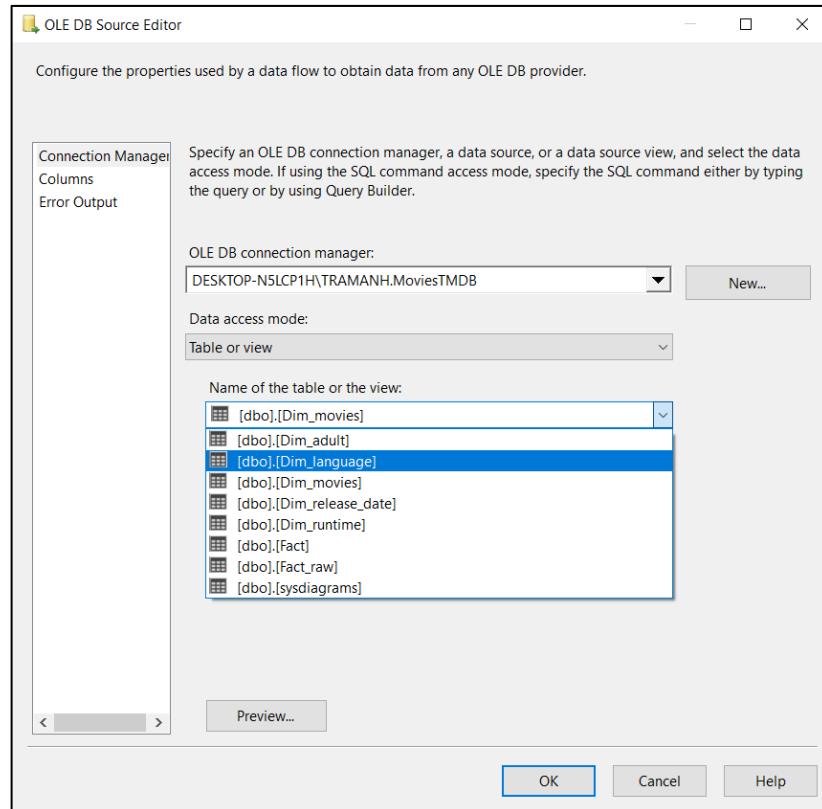
Bước 6: Click chuột phải chọn Edit, sau đó chọn bảng Fact_Raw đã tạo trước đó làm data source cho bảng Fact_Raw mới này.



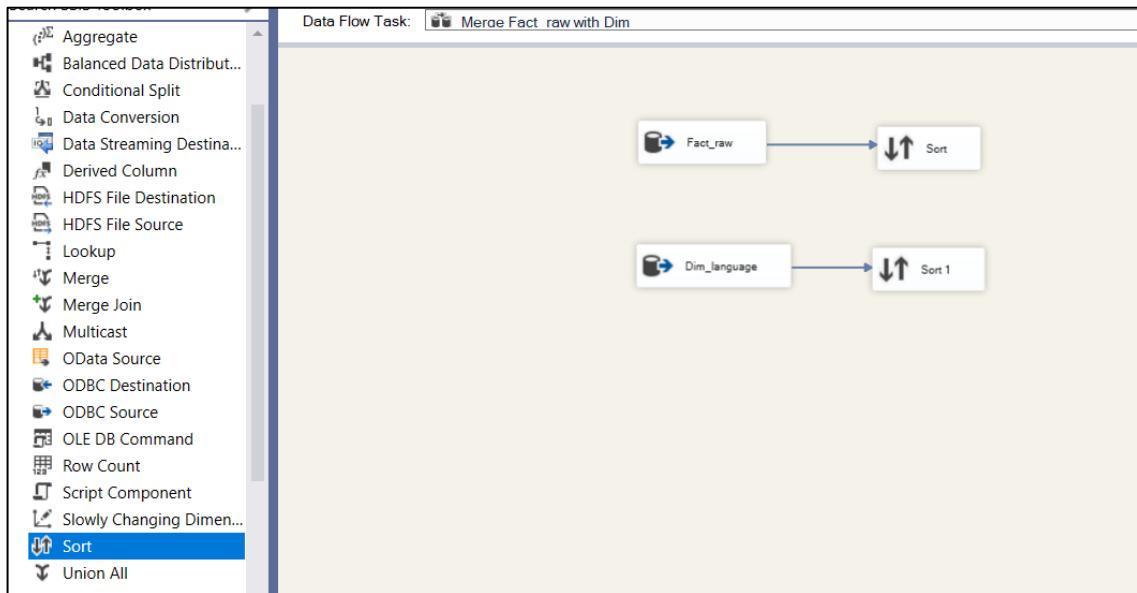
Bước 7: Chọn mục Columns để xem xét các cột được ánh xạ. Nhấn OK.



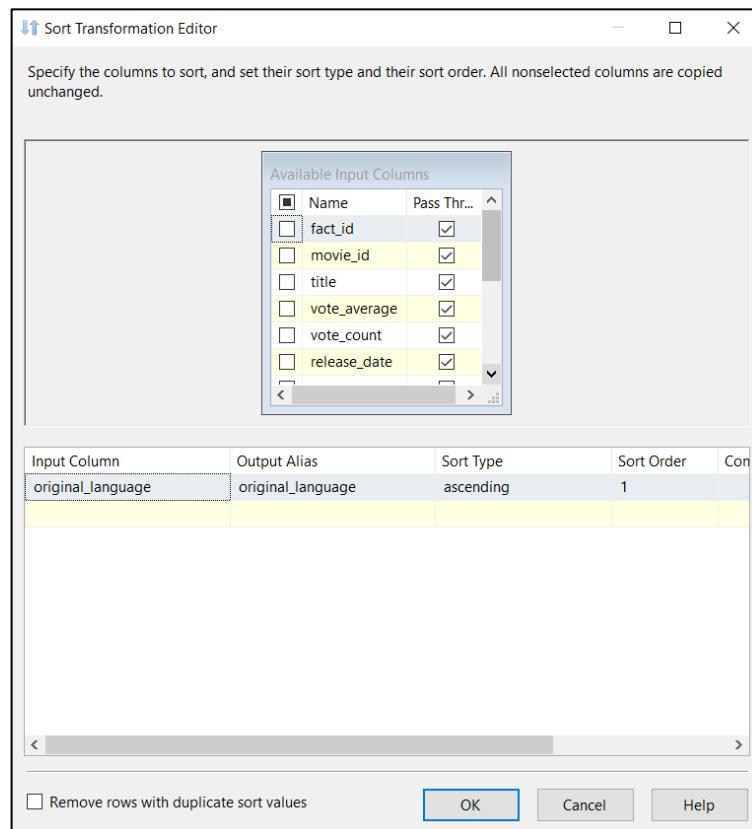
Bước 8: Tương tự thực hiện chọn ánh xạ cột cho Dim_language.



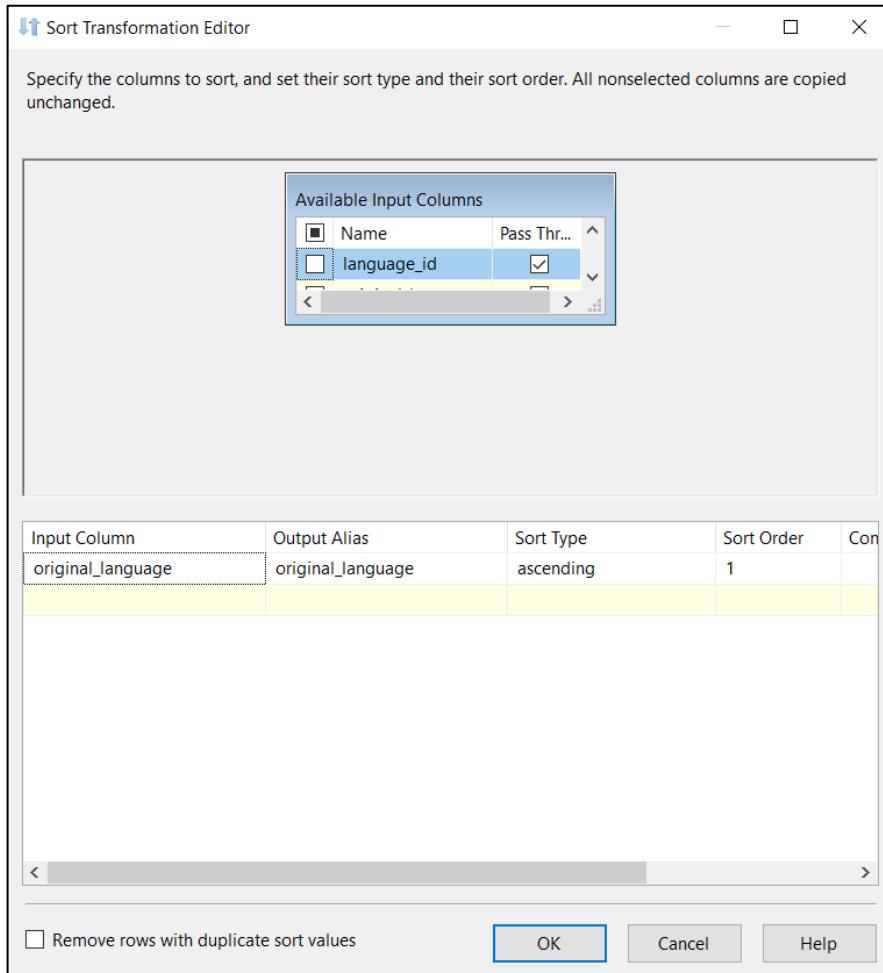
Bước 9: Tạo 2 Sort là Sort và Sort1 tương ứng với mỗi Source.



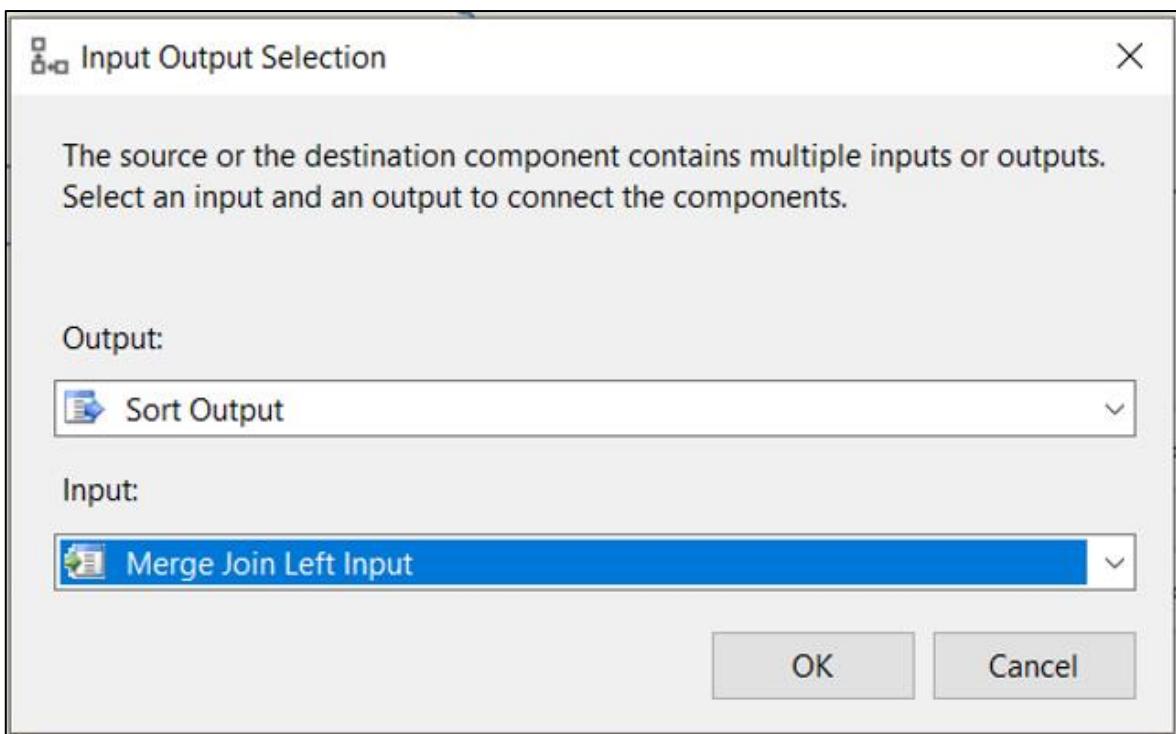
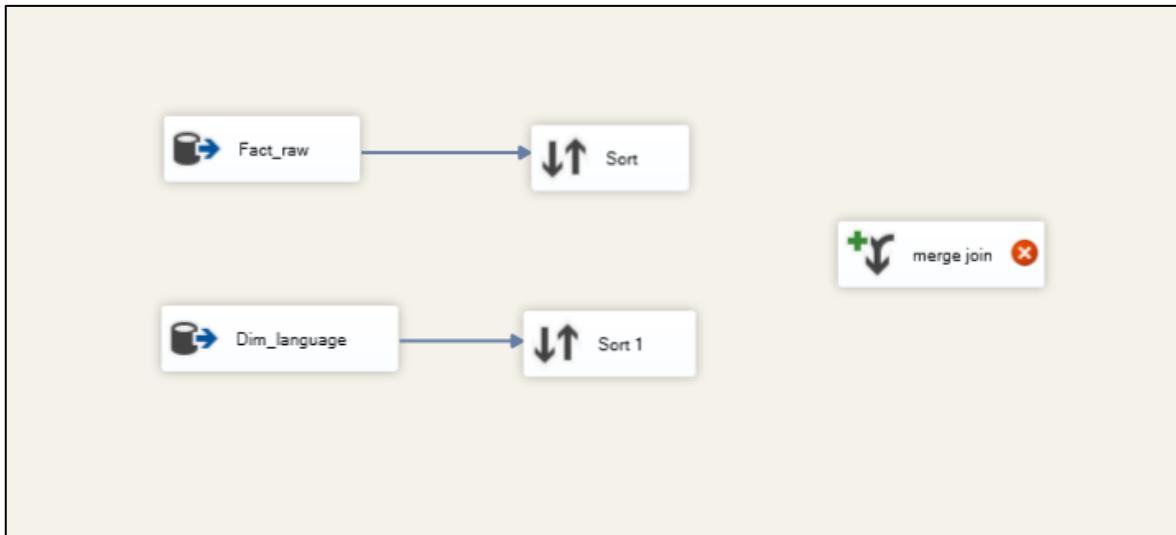
Bước 10: Ở Sort, click chuột phải chọn Edit và chọn các cột original_language theo thứ tự giống với bảng Dim_language để chuẩn bị cho quá trình merge.



Bước 11: Tương tự ta chọn các cột original_language cho Sort1.

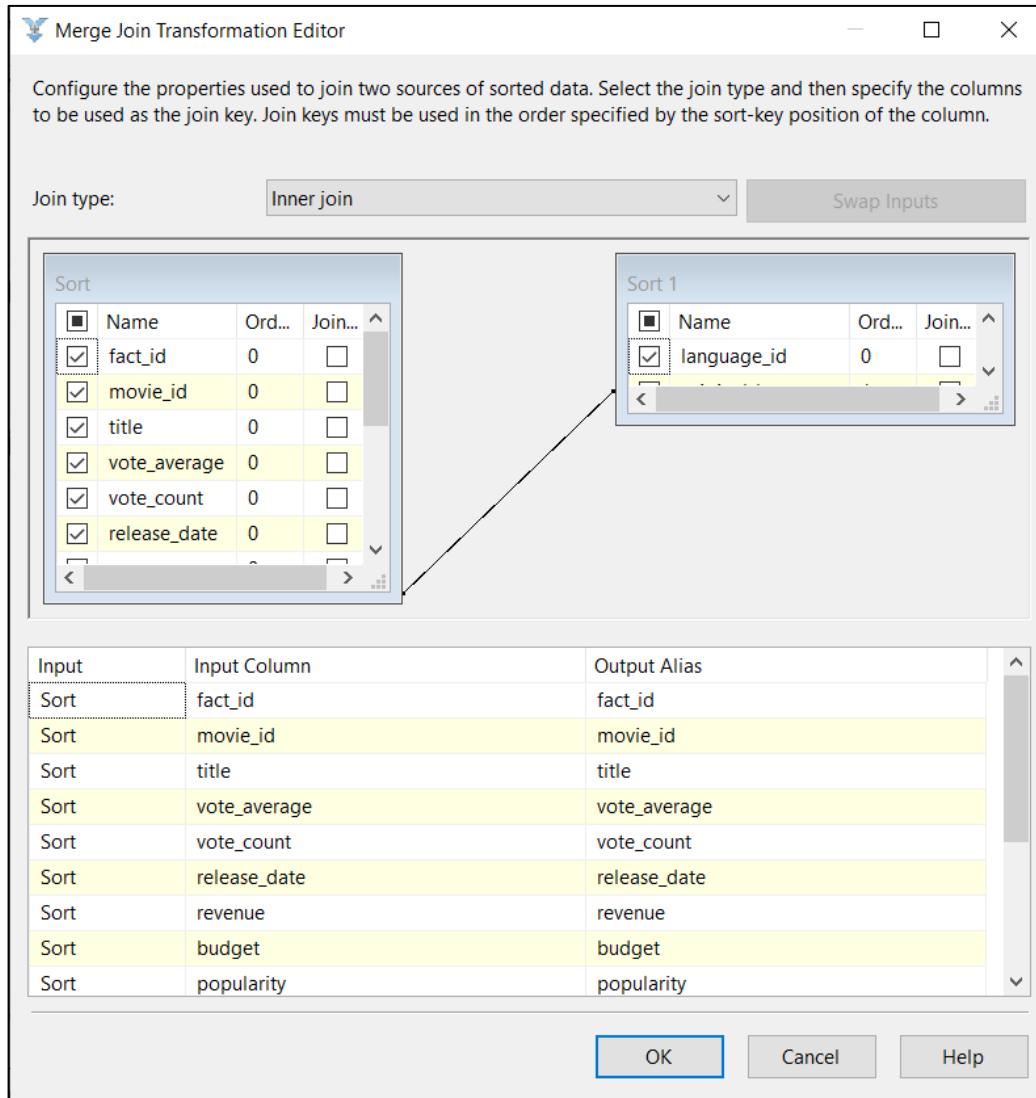


Bước 12: Tạo một Merge Join và nối với Sort, tiếp theo ta chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng Fact_Raw bất kể có kết quả khi thực hiện phép kết trái với cột ID của bảng Dim_language hay không.

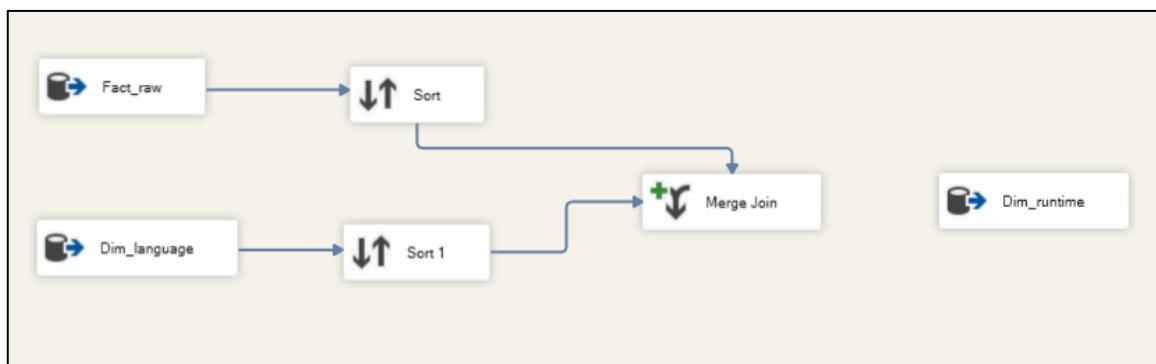


Bước 13:

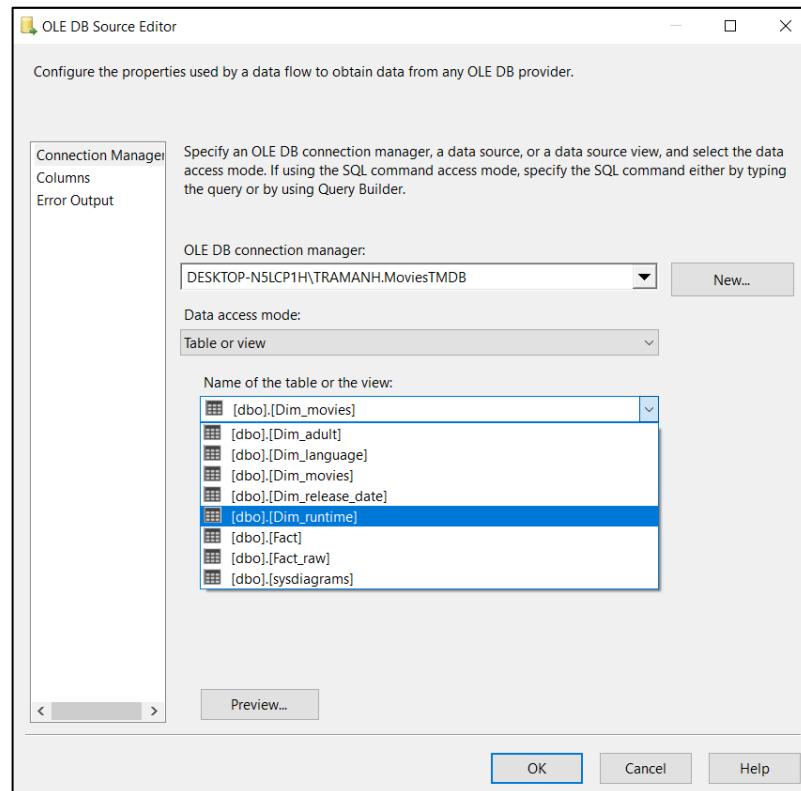
- Chuột phải vào Merge Join và nhấn Edit, một hộp thoại merge editor xuất hiện: ở đây ta tick chọn tất cả các cột của Sort nhưng không lấy original_language.
- Tiếp theo ta chọn language_id ở Sort1 để merge vào Fact_Raw
- Kết quả sau khi merge là bảng Fact_Raw không còn original_language và có thêm 1 thuộc tính mới là language_id.



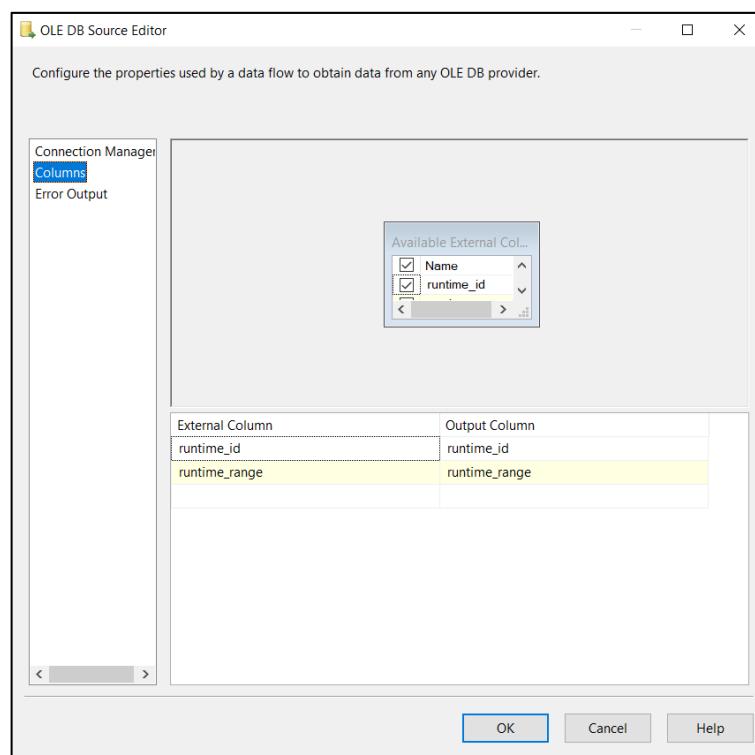
Bước 14: Click chuột phải vào Data Flow Task nói trên và chọn Edit, trong tab Data Flow ta tạo OLE DB Source và đổi thành Dim_runtime.



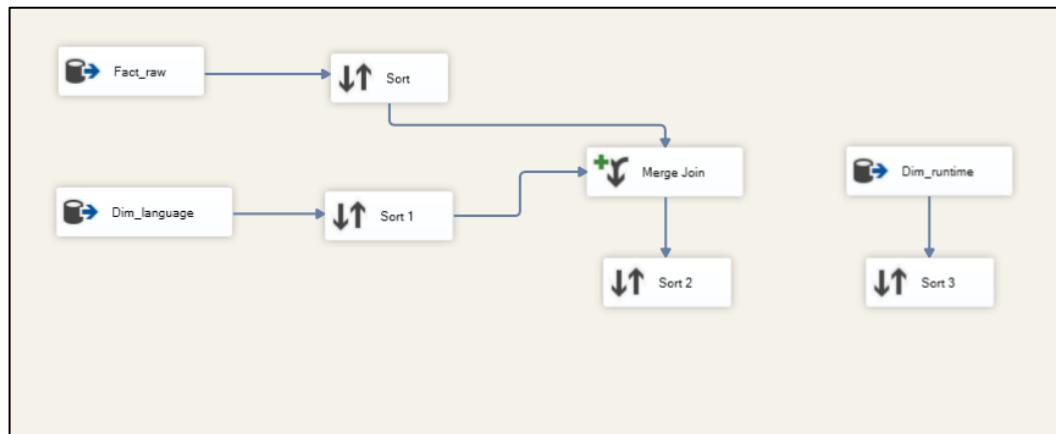
Bước 15: Thực hiện chọn ánh xạ các cột cho Dim_runtime.



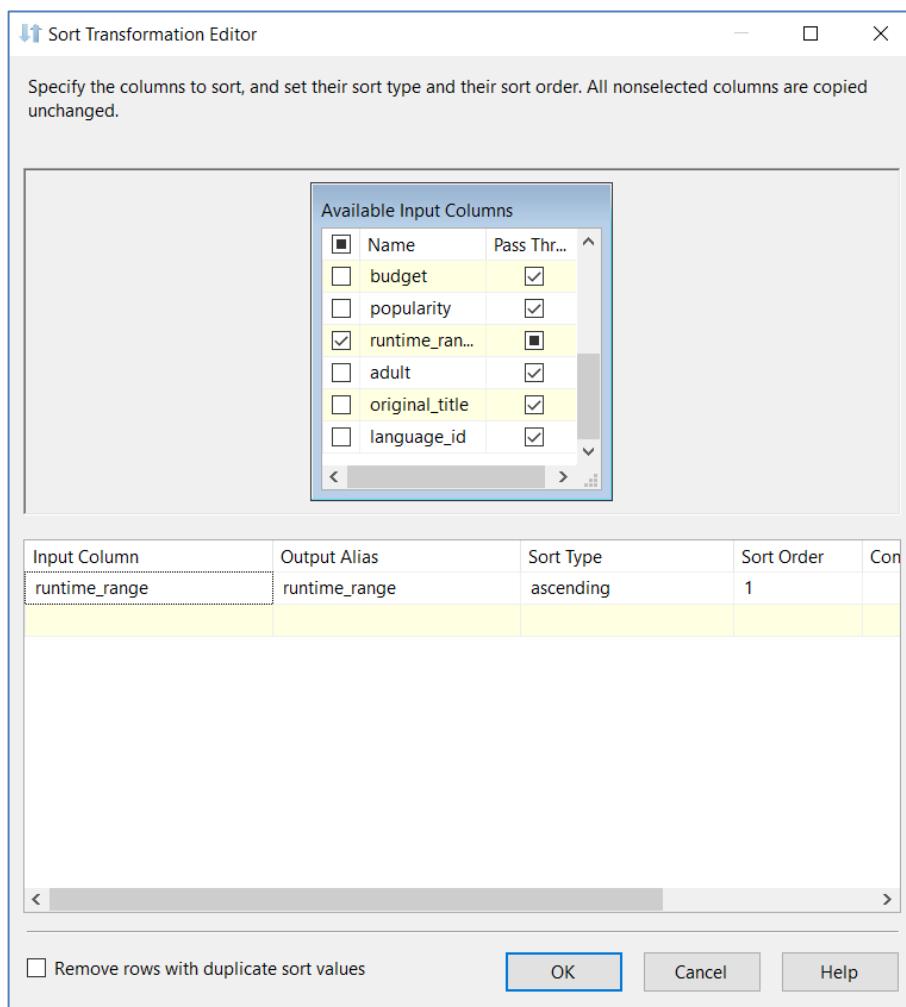
Chọn mục Columns để xem xét các cột được ánh xạ. Nhấn OK.



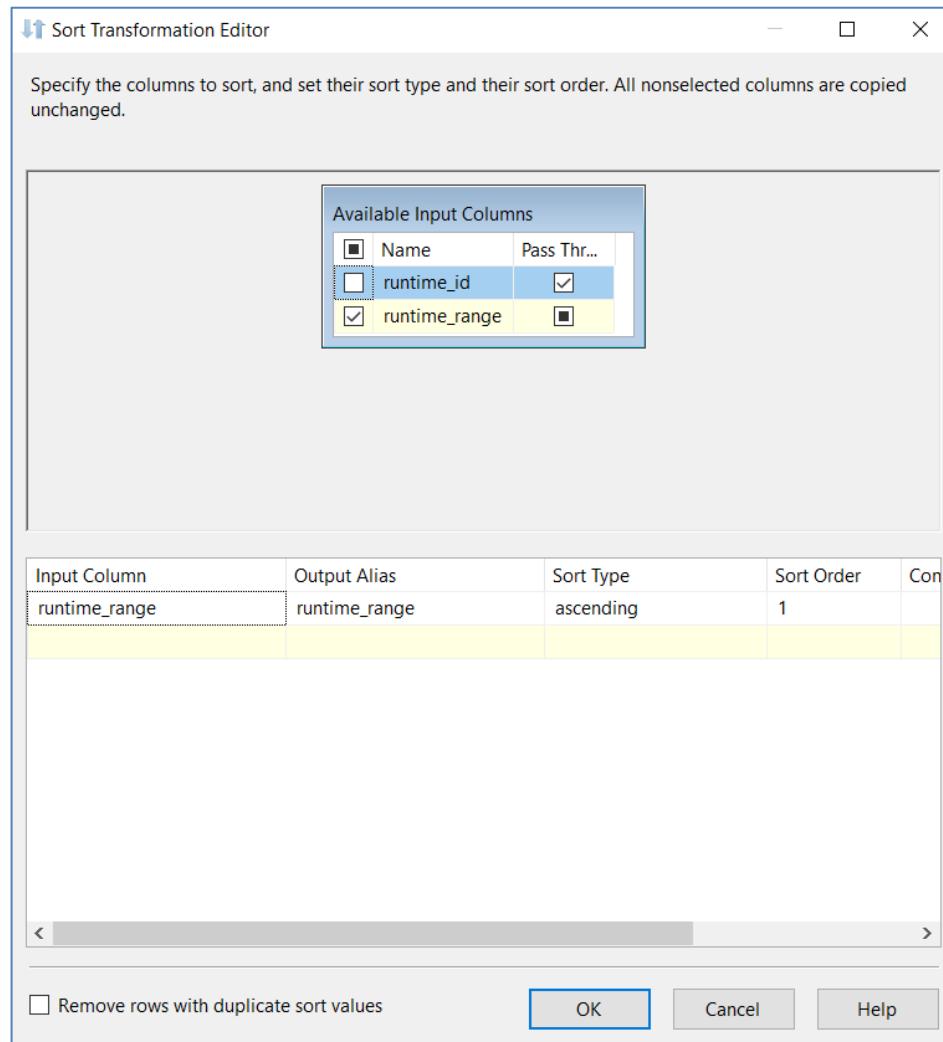
Bước 16: Tạo 2 Sort tương ứng với mỗi Source



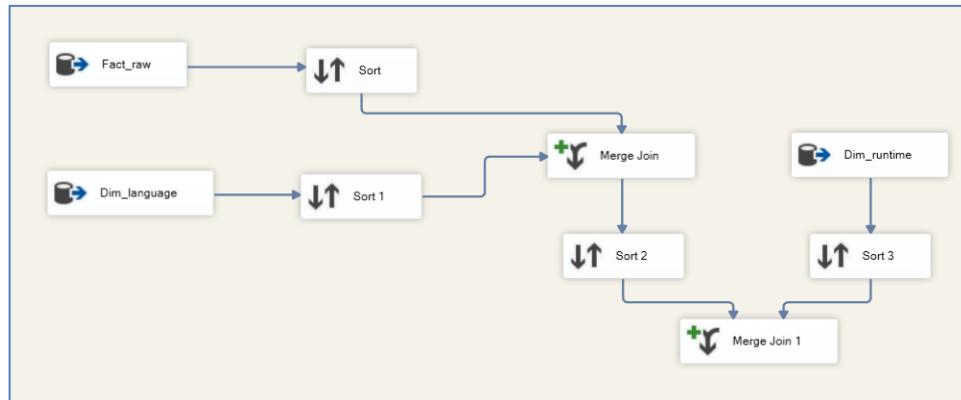
Bước 17: Tại Sort2, click chuột phải chọn Edit và chọn cột runtime_range để chuẩn bị cho quá trình merge.



Bước 18: Tương tự ta chọn cột runtime_range cho Sort3.

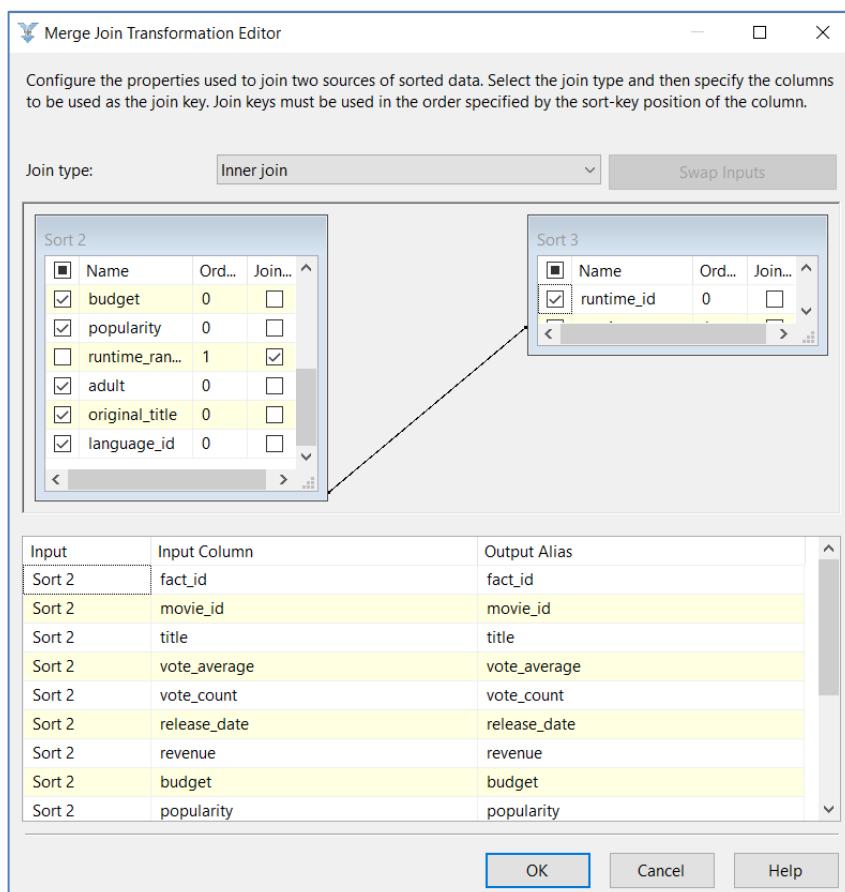


Bước 19: Tạo một Merge Join và nối với Sort2, tiếp theo chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong output của Merge Join bất kể có kết quả khi thực hiện phép kết trái với cột runtime_id của bảng Dim_runtime hay không. Nối Sort3 với Merge Join.

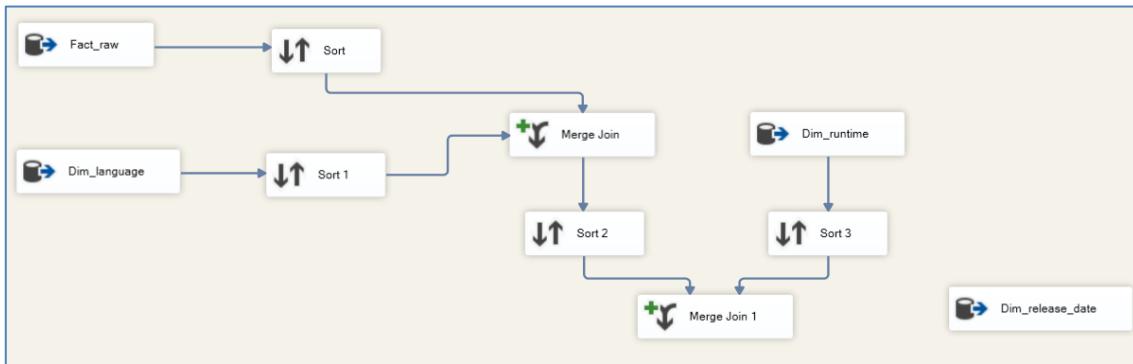


Bước 20: Chuột phải vào Merge Join và nhấn Edit, một hộp thoại merge editor xuất hiện: ở đây ta tick chọn tất cả các cột của Sort2 nhưng không lấy thuộc tính runtime_range.

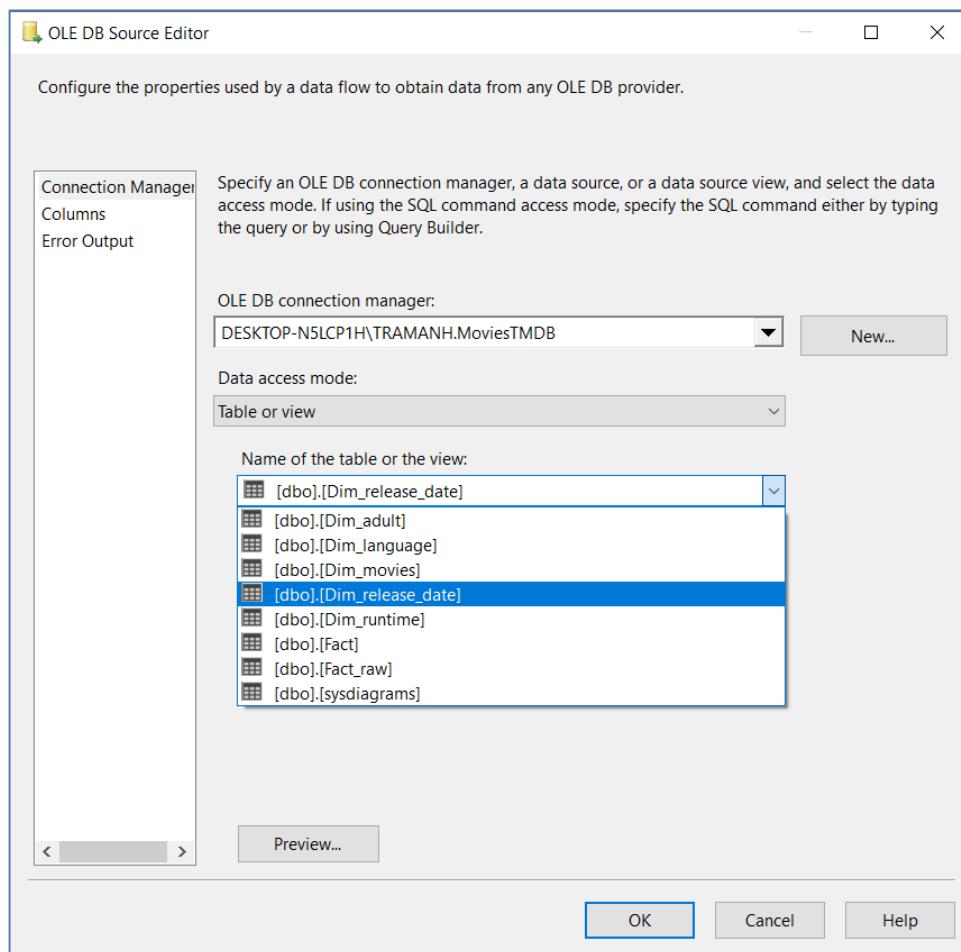
- Tiếp theo ta chọn runtime_ID ở Sort1 để merge
- Kết quả sau khi merge là bảng Fact1 không còn thuộc tính runtime_range và có thêm 1 thuộc tính mới là runtime_ID.



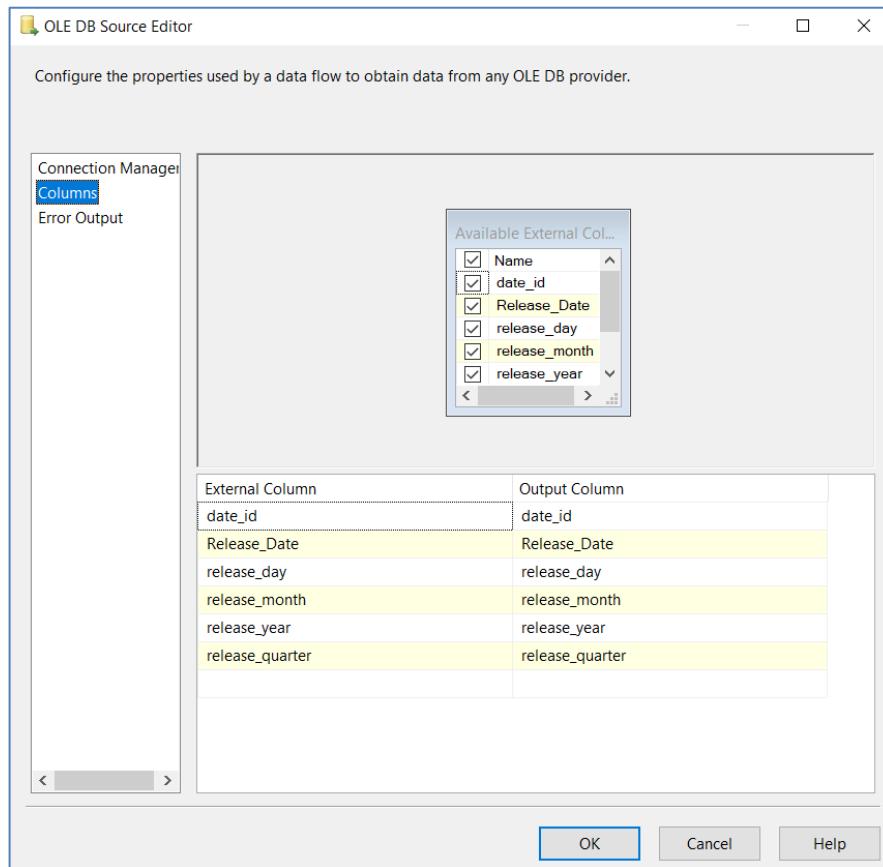
Bước 21: Click chuột phải vào Data Flow Task nói trên và chọn Edit, trong tab Data Flow ta tạo OLE DB Source và đổi thành Dim_release_date.



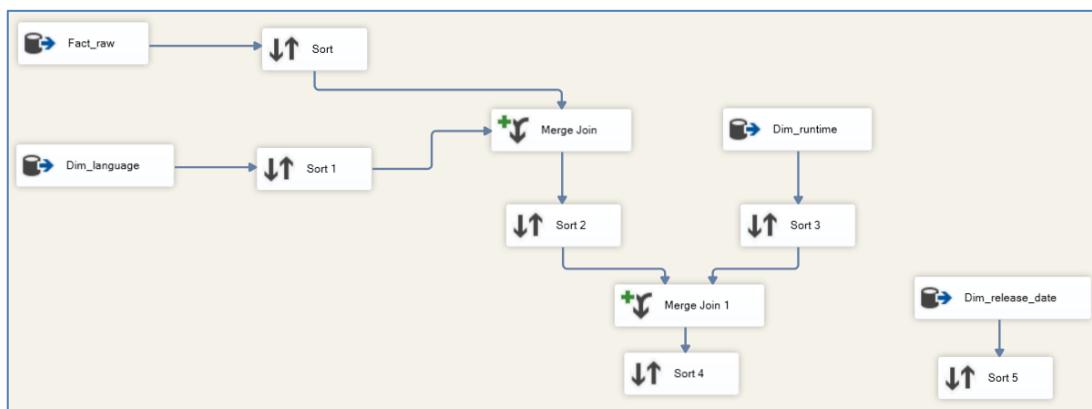
Bước 22: Thực hiện chọn ánh xạ các cột cho Dim_release_date.



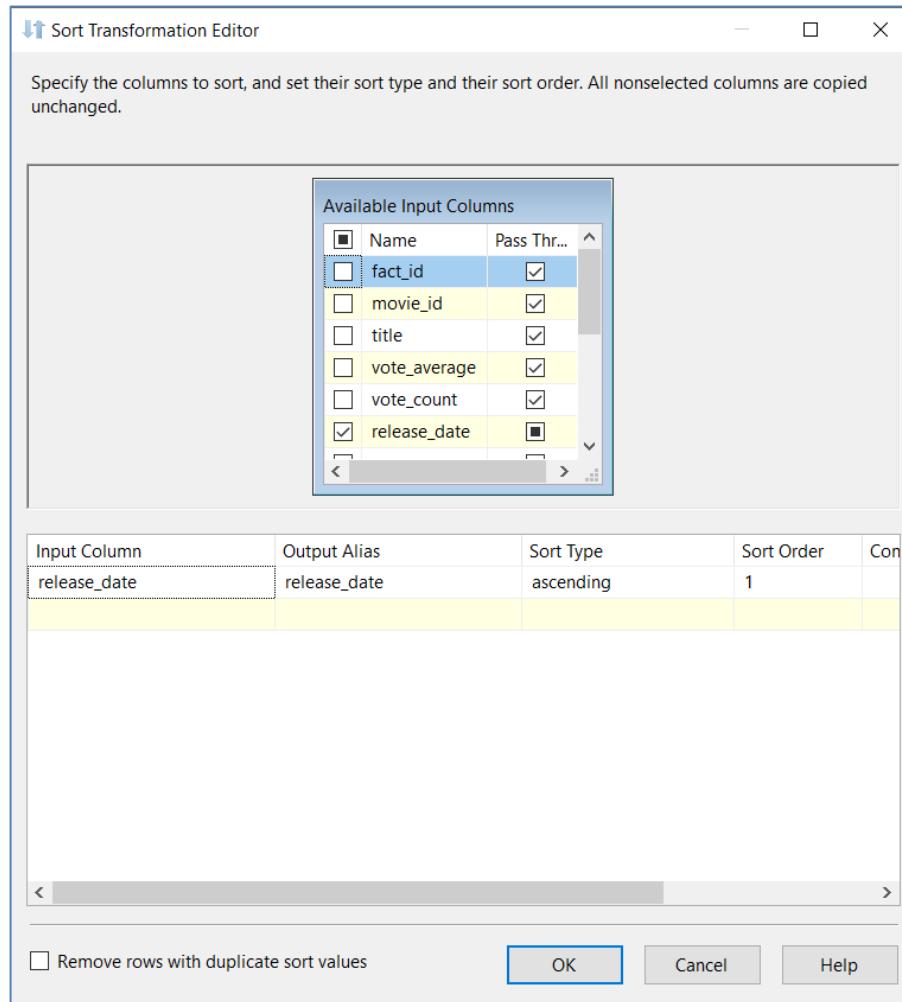
Chọn mục Columns để xem xét các cột được ánh xạ. Nhấn OK.



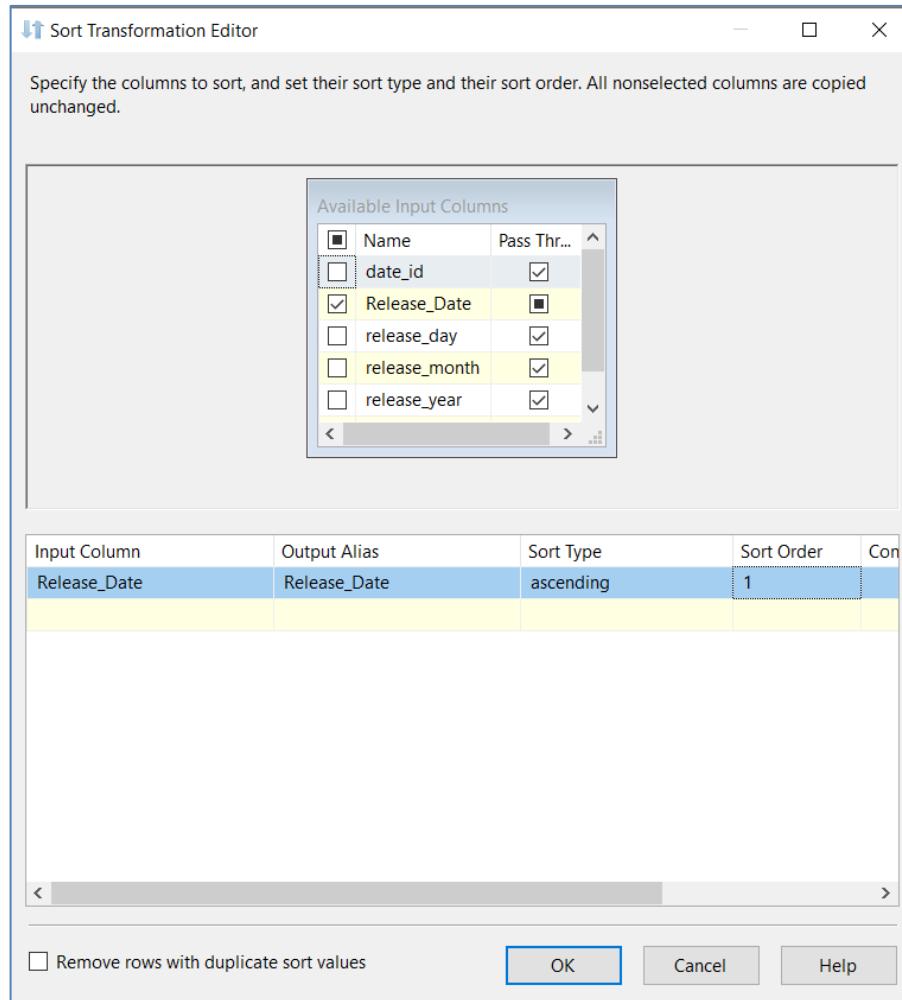
Bước 23: Tạo 2 Sort tương ứng với mỗi Merge Join 1 và Dim_release_date.



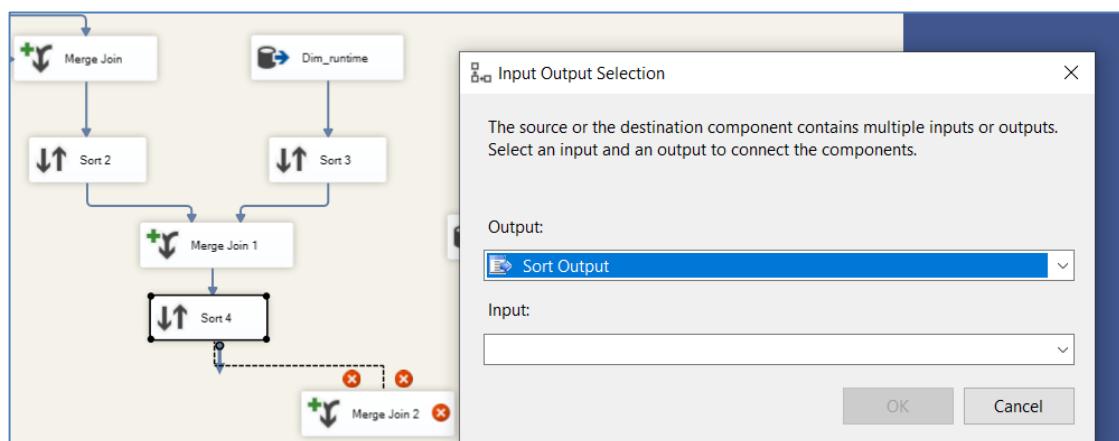
Bước 24: Tại Sort4, click chuột phải chọn Edit và chọn cột release_date để chuẩn bị cho quá trình merge.



Bước 25: Tương tự chọn cột Release_Date cho Sort 5



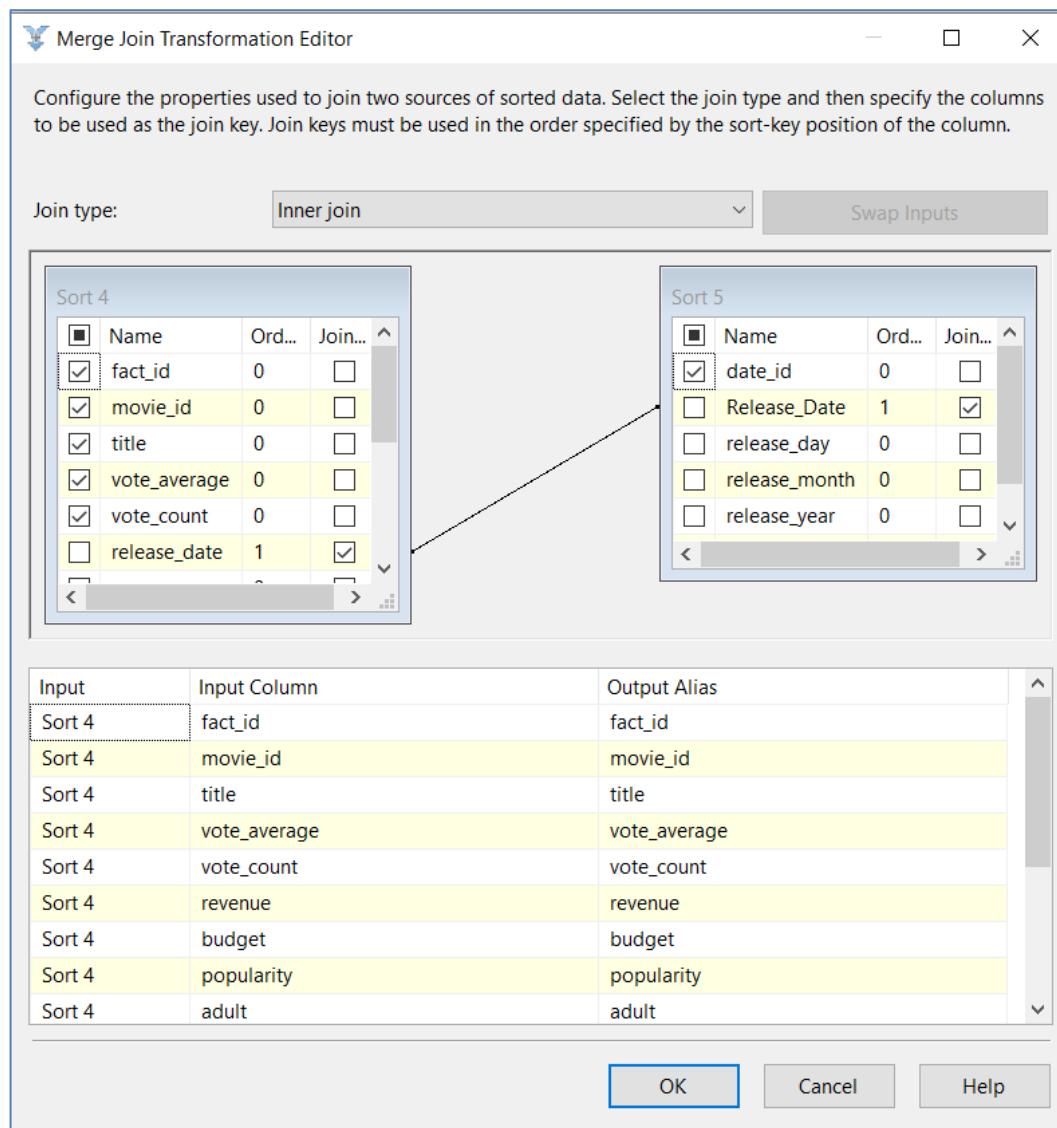
Bước 26: Tạo một Merge Join và nối với Sort 4, tiếp theo chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong output của Merge Join 1 bất kể có kết quả khi thực hiện phép kết trái với cột date_id của bảng Dim_release_date hay không.



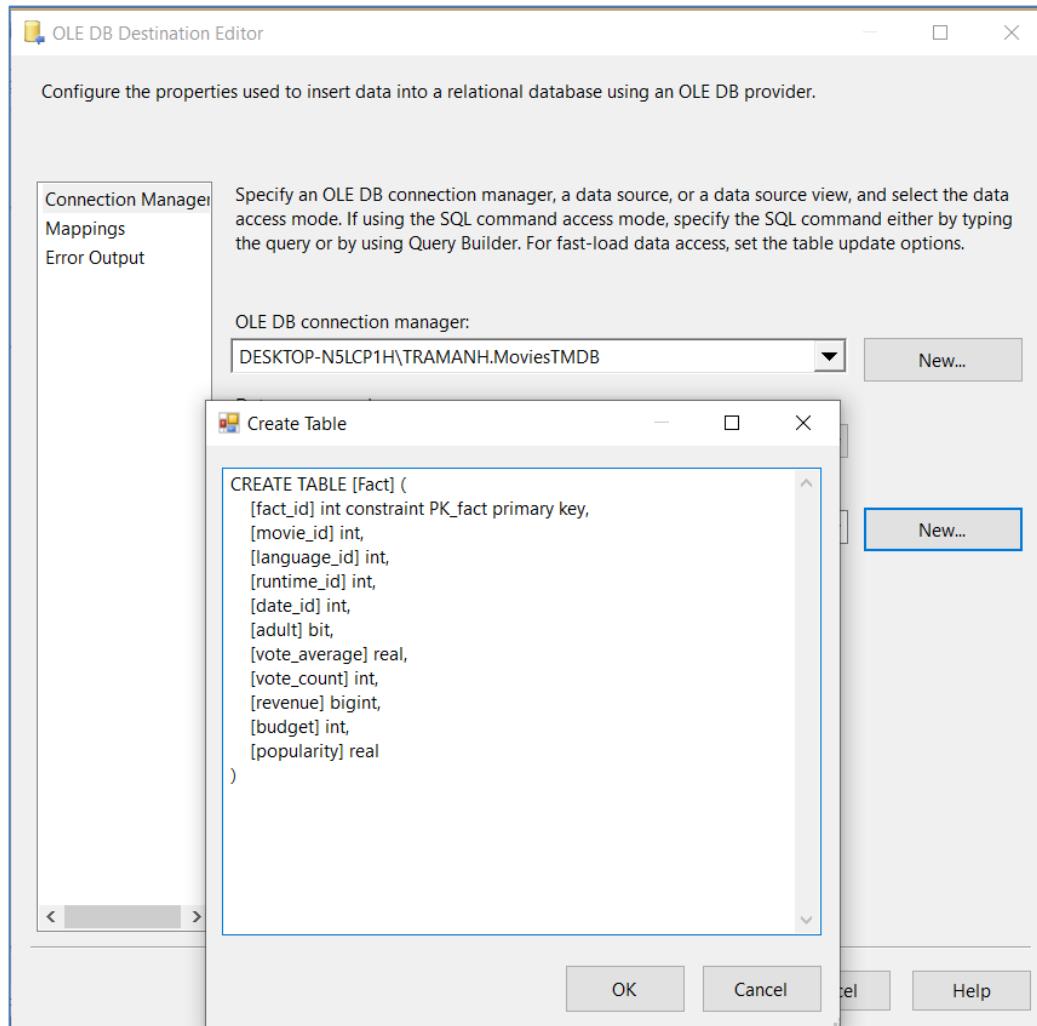
Bước 27: Nối Dim_release_date với Merge Join

Bước 28:

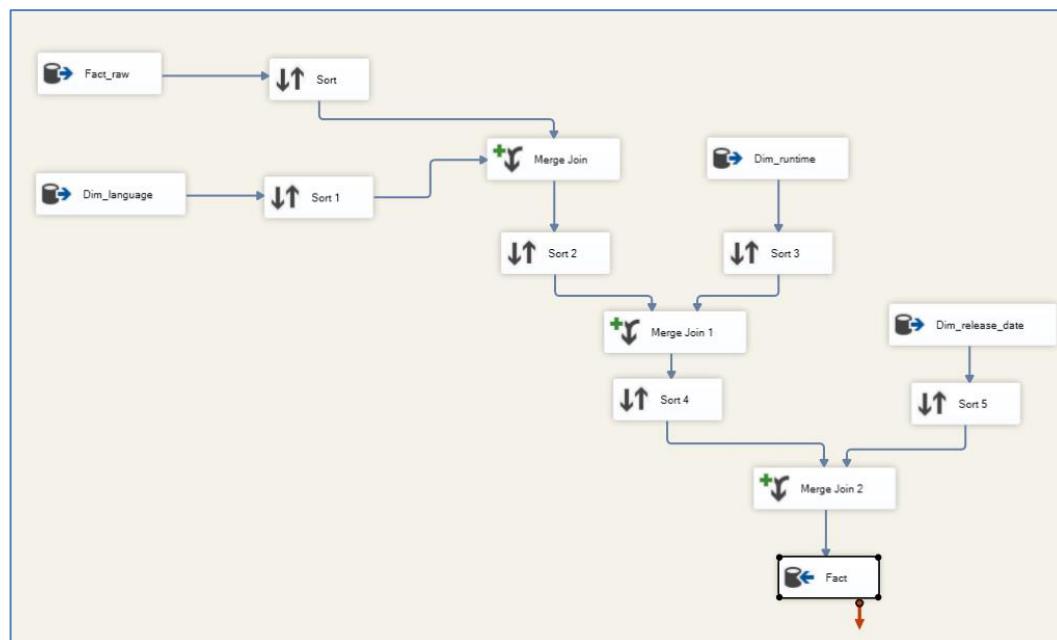
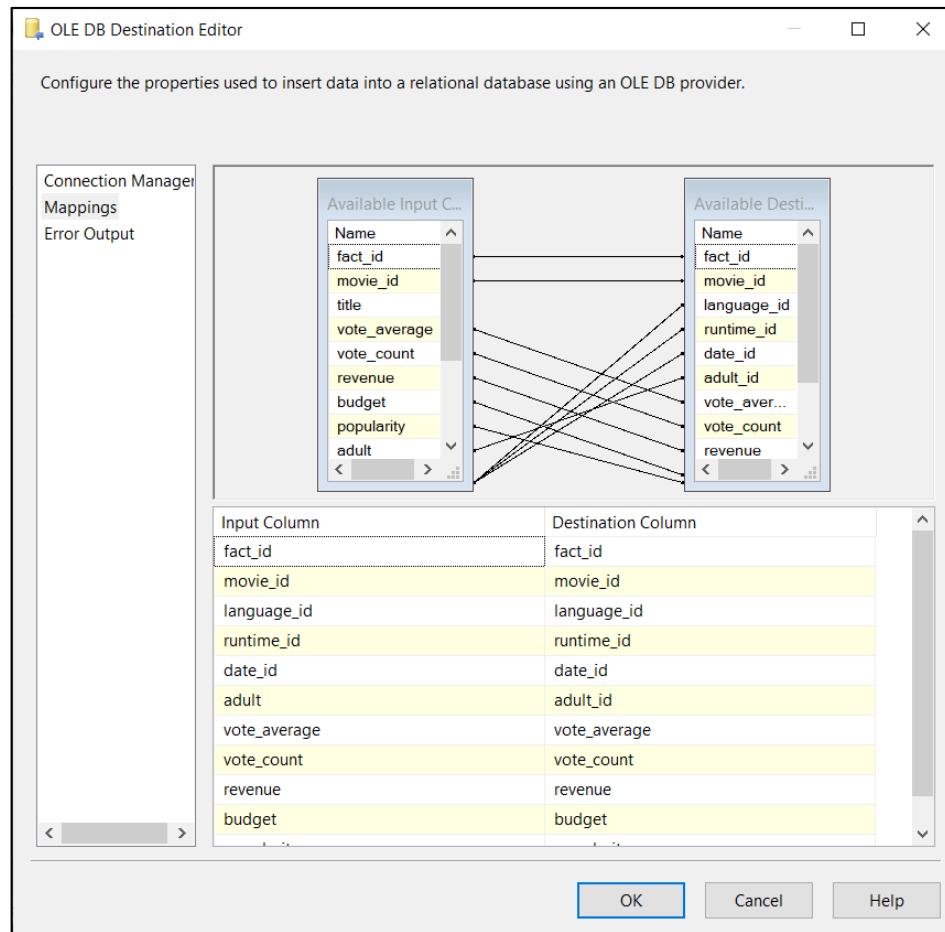
- Chuột phải vào Merge Join và nhấn Edit, một hộp thoại merge editor xuất hiện: ở đây ta tick chọn tất cả các cột của Sort 4 nhưng không lấy thuộc tính release_date.
- Tiếp theo ta chọn Airport_ID ở Sort 5 để merge
- Kết quả sau khi merge là bảng Fact không còn thuộc tính release_date và có thêm 1 thuộc tính mới là date_id



Bước 29: Tạo bảng Fact từ một OLE DB Destination để chứa tất cả những gì đã merge.

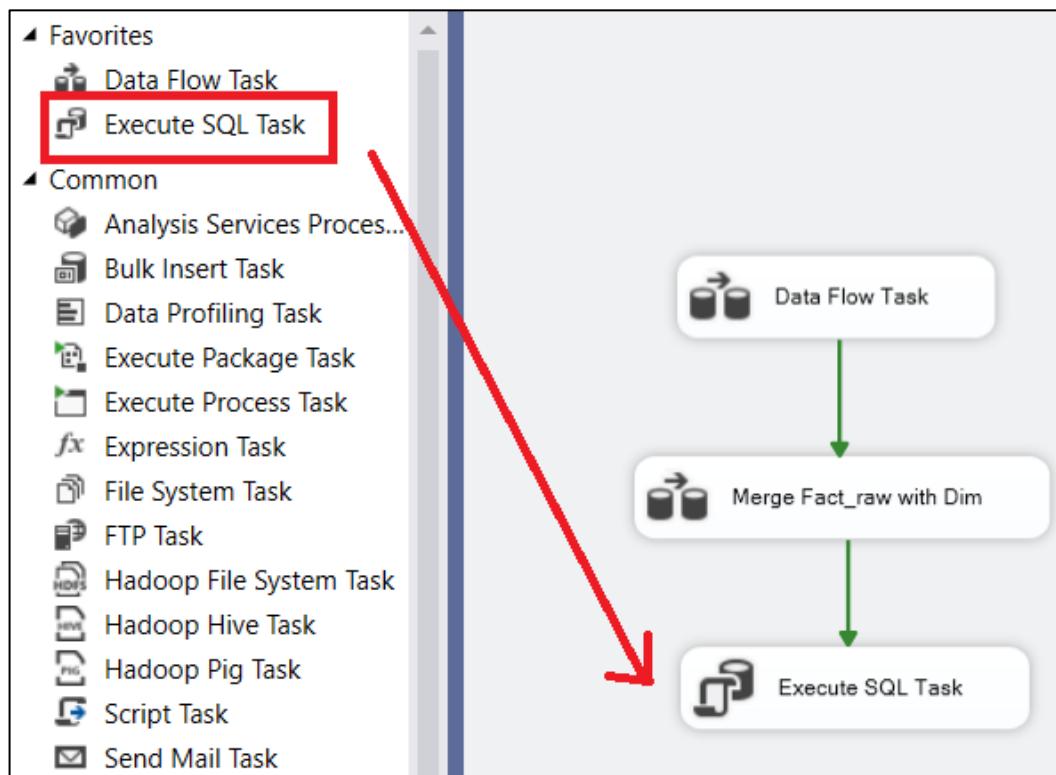


Chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu

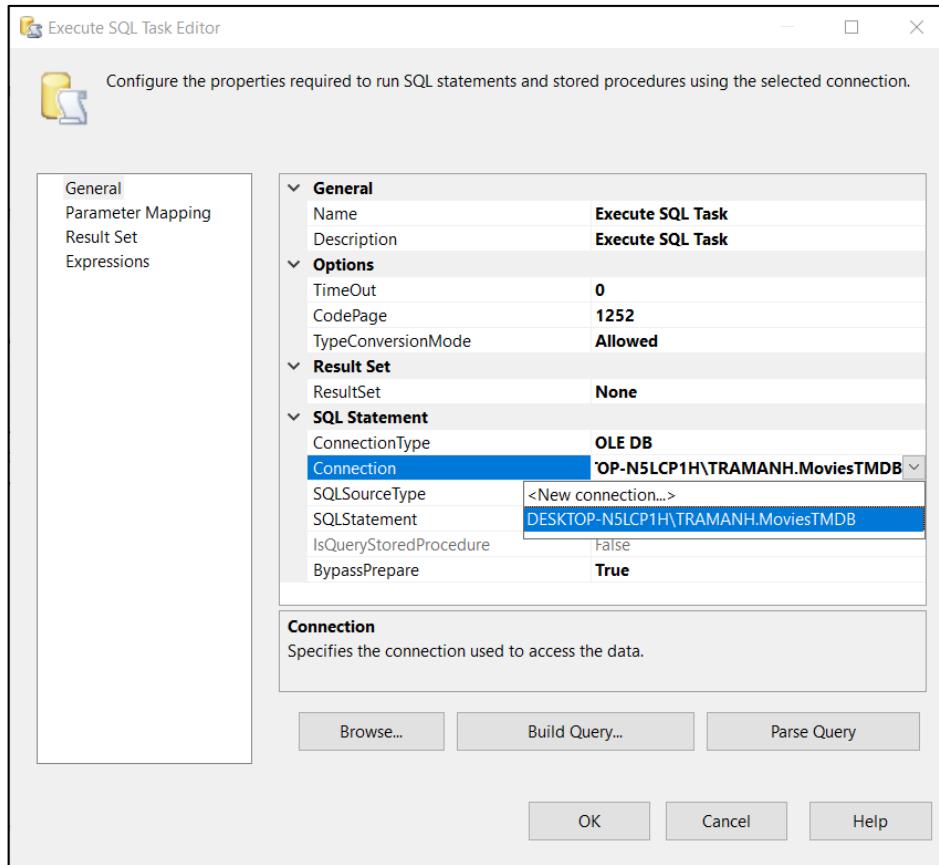


Tiếp theo đây, ta thực hiện quá trình tạo khóa ngoại giữa bảng Fact và Dimension.

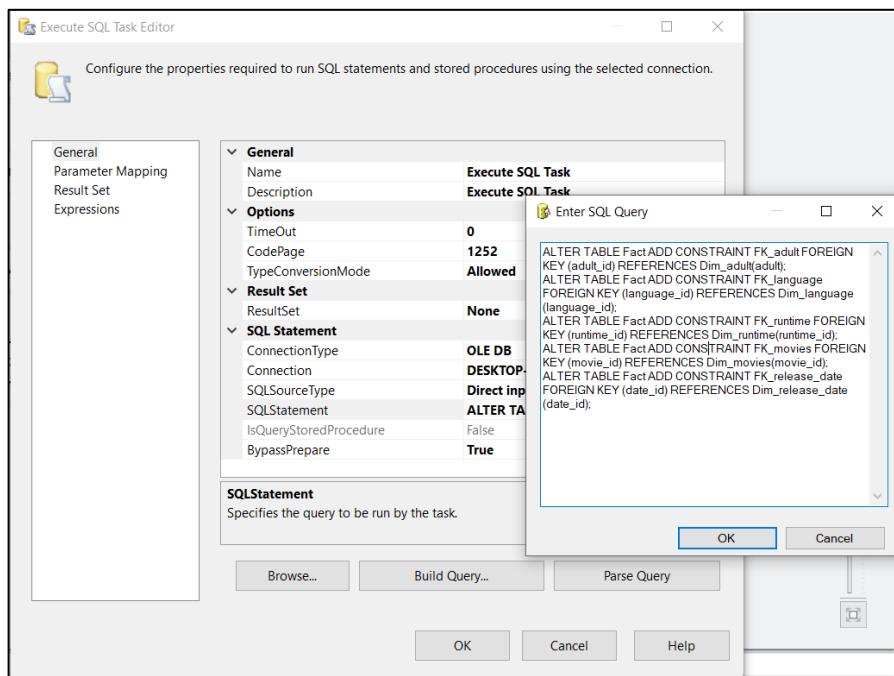
Bước 30: Tạo một Execute SQL Task để thực thi các câu lệnh SQL tạo các khóa ngoại từ các Dimension đến bảng Fact.



Bước 31: Nhấn chuột phải vào Execute SQL Task này và chọn Edit. Ở ô Connection, chọn connection đã thiết lập đến data warehouse trong SQL Server.

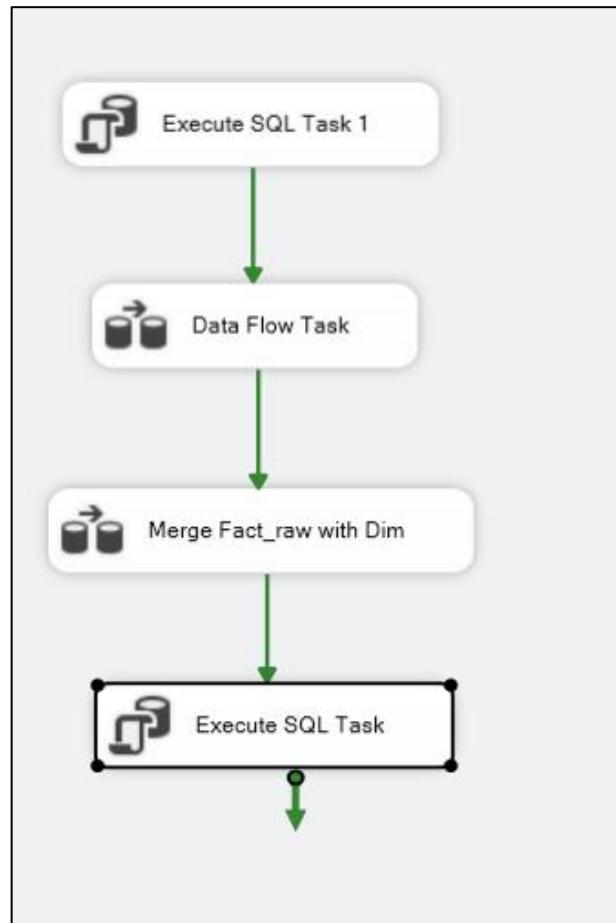


Bước 32: Ở ô SQLStatement, thêm các câu truy vấn SQL thực hiện tạo các khóa ngoại từ các Dimension đến bảng Fact. Nhấn OK để hoàn tất quá trình.

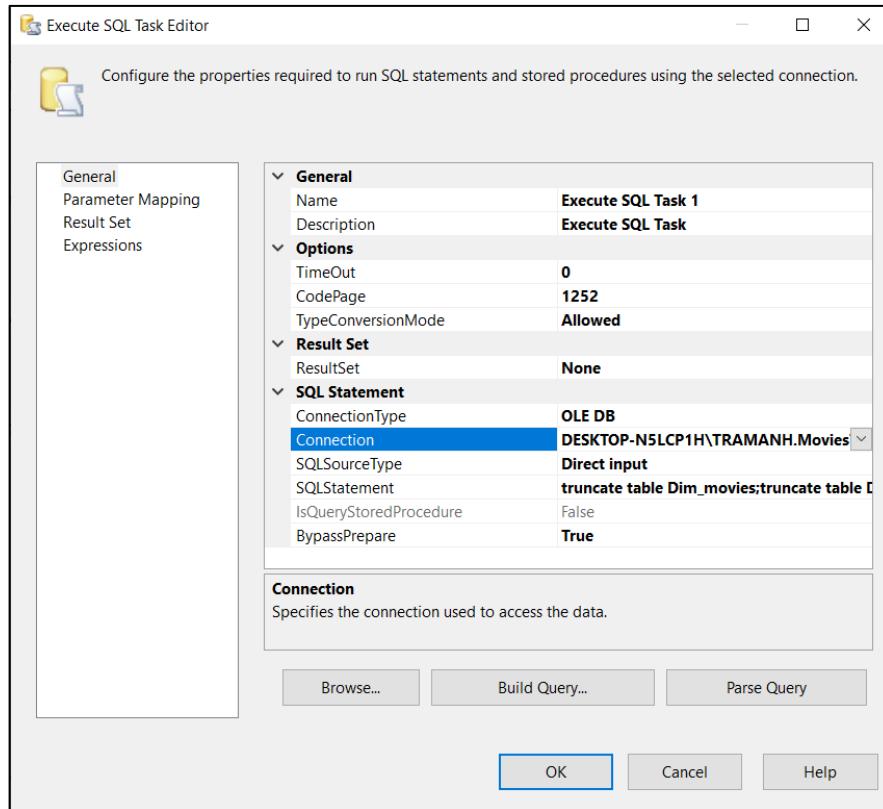


2.5.7. Chạy dự án SSIS

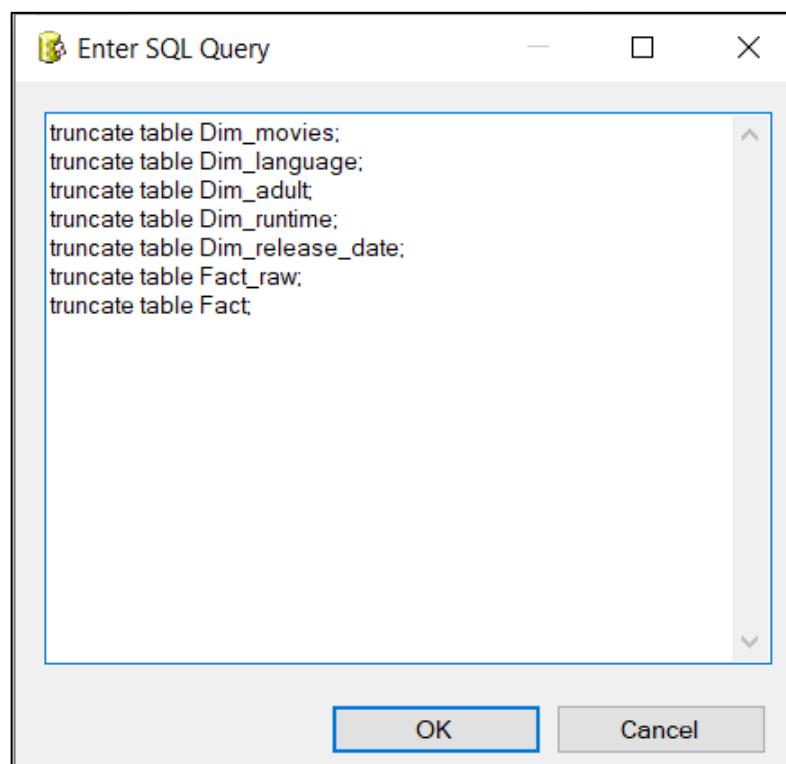
Bước 1: Thêm vào một Execute SQL Task nhằm thực hiện nhiệm vụ đảm bảo đồ dữ liệu mới hoàn toàn (không bị chồng chéo dữ liệu cũ) mỗi khi chạy project, trước quá trình chia bảng Fact và các Dimension.



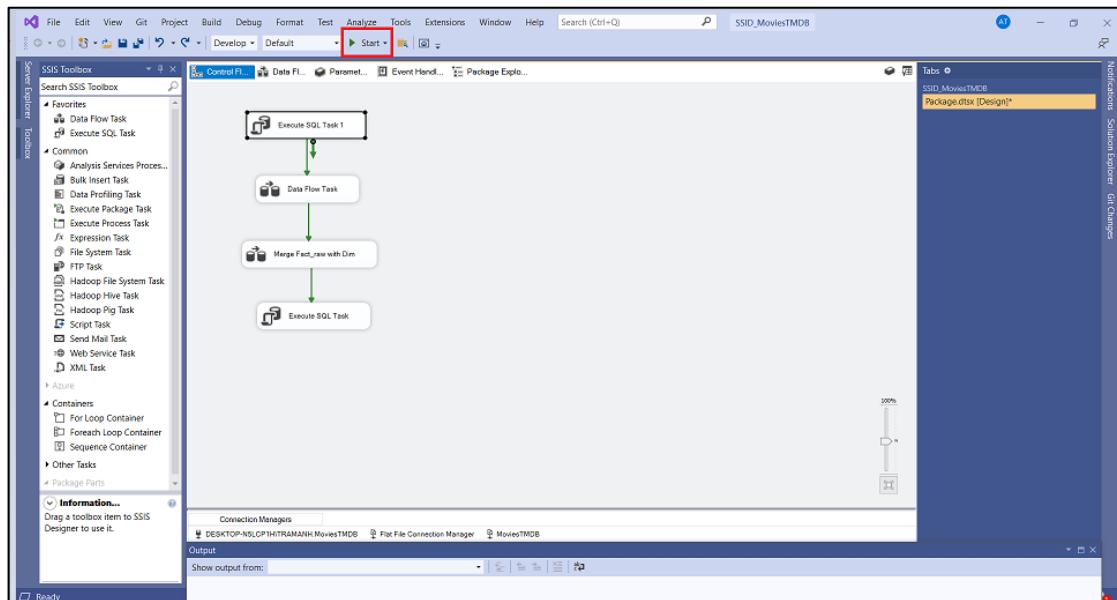
Bước 2: Nhấn chuột phải vào Execute SQL Task này và chọn Edit. Ở ô Connection, chọn connection đã thiết lập đến data warehouse trong SQL Server.



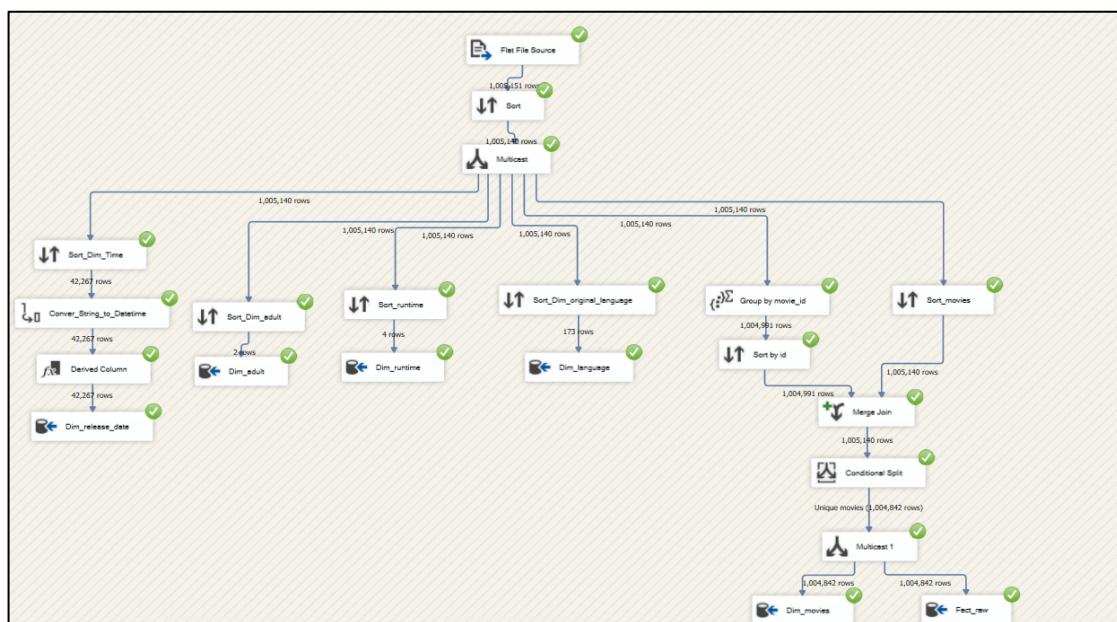
Bước 3: Ở ô SQLStatement, thêm các câu truy vấn SQL thực hiện xóa dữ liệu cũ trong các bảng mỗi khi chạy project. Nhấn OK để hoàn tất quá trình.

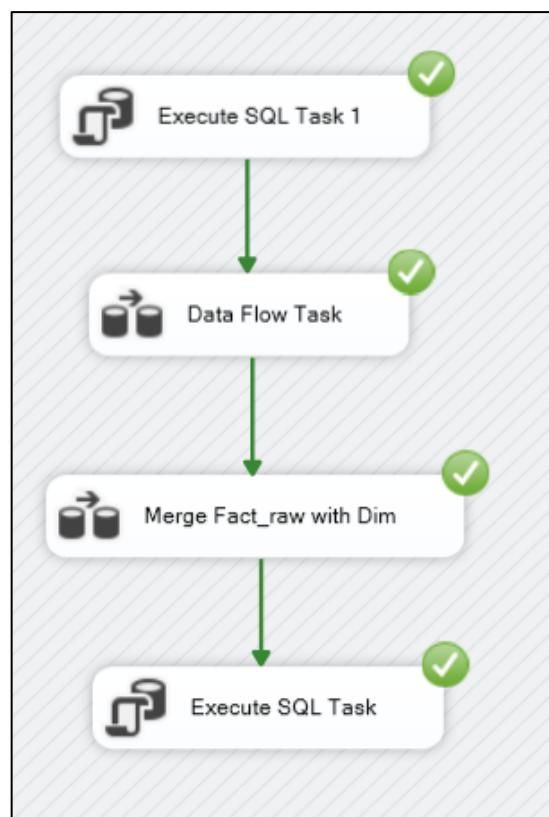
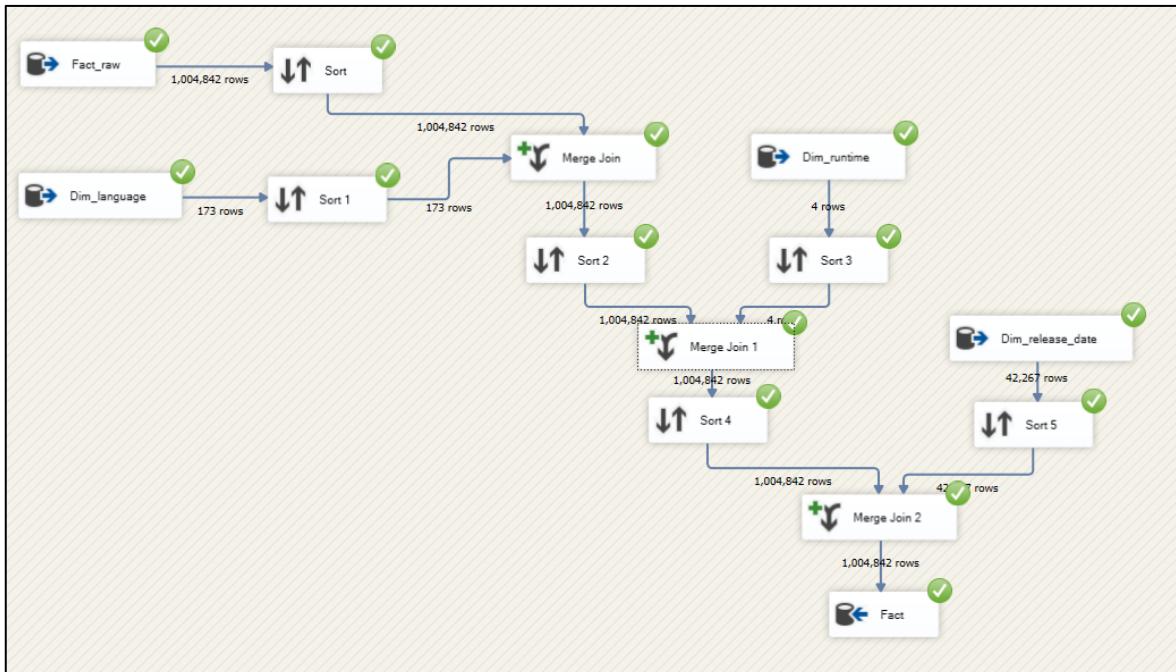


Bước 4: Nhấn nút Start trên thanh menu để tiến hành chạy project.

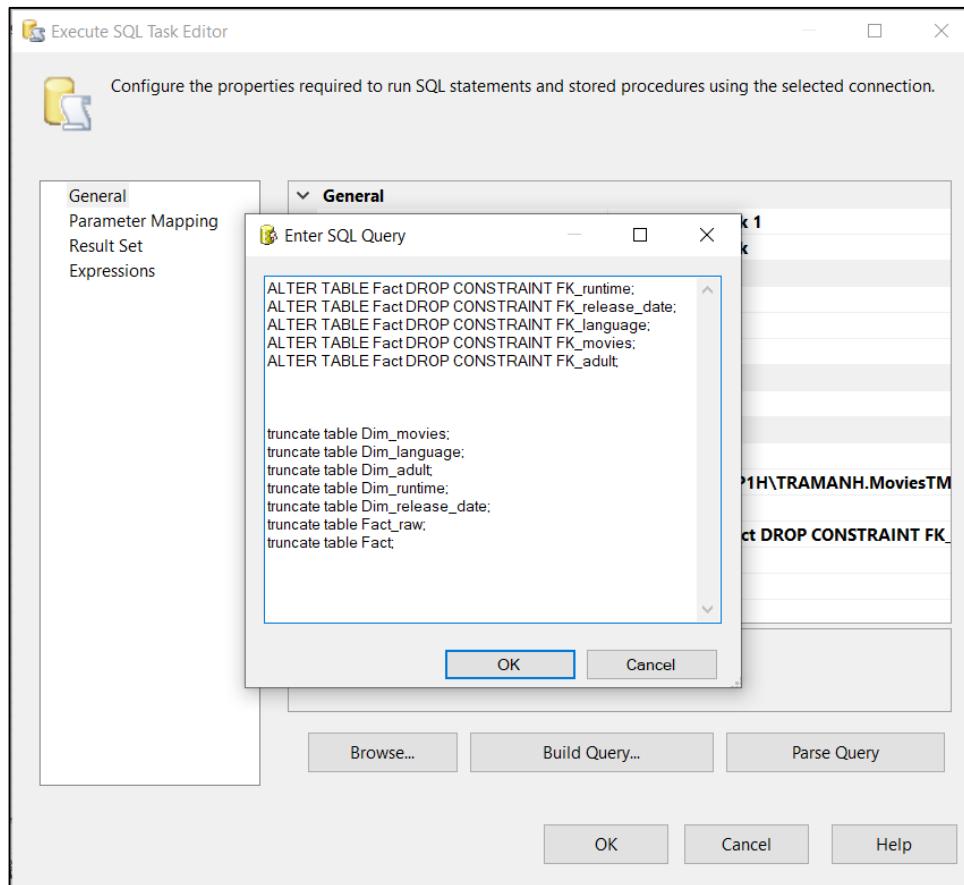


Kết quả chạy project:





Bước 5: Sau khi đã chạy project lần đầu thành công, khóa ngoại của bảng Fact tham chiếu đến các Dimension đã được tạo. Ta tiến hành thêm các lệnh SQL để xóa và thêm lại các khóa ngoại vào mỗi lần khởi chạy lại project tiếp theo.



2.5.8. Kiểm tra dữ liệu các bảng

Kiểm tra dữ liệu bảng Dim_adult

adult	
1	0
2	1

Kiểm tra dữ liệu bảng Dim_language

	language_id	original_language
1	1	aa
2	2	ab
3	3	af
4	4	ak
5	5	am
6	6	an
7	7	ar
8	8	as
9	9	av
10	10	ay
11	11	az
12	12	ba
13	13	be
14	14	bg
15	15	bi
16	16	bm

Kiểm tra dữ liệu bảng Dim_movies

	movie_id	title	original_title
1	2	Ariel	Ariel
2	3	Shadows in Paradise	Vanoja paratisissa
3	5	Four Rooms	Four Rooms
4	6	Judgment Night	Judgment Night
5	8	Life in Loops (A Megacities RMX)	Life in Loops (A Megacities RMX)
6	9	Sunday in August	Sonntag im August
7	11	Star Wars	Star Wars
8	12	Finding Nemo	Finding Nemo
9	13	Forrest Gump	Forrest Gump
10	14	American Beauty	American Beauty
11	15	Citizen Kane	Citizen Kane
12	16	Dancer in the Dark	Dancer in the Dark
13	17	The Dark	The Dark
14	18	The Fifth Element	The Fifth Element
15	19	Metropolis	Metropolis
16	20	My Life Without Me	My Life Without Me
17	21	The Endless Summer	The Endless Summer
18	22	Pirates of the Caribbean: The C...	Pirates of the Caribbean: The C...
19	24	Kill Bill: Vol. 1	Kill Bill: Vol. 1
20	25	Jarhead	Jarhead
21	26	Walk on Water	LaLehet Al HaMayim
22	27	9 Songs	9 Songs
23	28	Apocalypse Now	Apocalypse Now
24	33	Unforgiven	Unforgiven
25	35	The Simpsons Movie	The Simpsons Movie
26	38	Eternal Sunshine of the Spotles...	Eternal Sunshine of the Spotles...

Kiểm tra dữ liệu bảng Dim_movies

	date_id	Release_Date	release_day	release_month	release_year	release_quarter
1	1	NULL	NULL	NULL	NULL	NULL
2	2	1800-09-11 00:00:00.000	11	9	1800	3
3	3	1865-01-01 00:00:00.000	1	1	1865	1
4	4	1867-01-01 00:00:00.000	1	1	1867	1
5	5	1874-12-09 00:00:00.000	9	12	1874	4
6	6	1878-05-07 00:00:00.000	7	5	1878	2
7	7	1882-12-06 00:00:00.000	6	12	1882	4
8	8	1883-11-19 00:00:00.000	19	11	1883	4
9	9	1885-01-01 00:00:00.000	1	1	1885	1
10	10	1887-01-01 00:00:00.000	1	1	1887	1
11	11	1887-08-18 00:00:00.000	18	8	1887	3
12	12	1888-01-01 00:00:00.000	1	1	1888	1
13	13	1888-05-31 00:00:00.000	31	5	1888	2
14	14	1888-10-14 00:00:00.000	14	10	1888	4
15	15	1888-10-15 00:00:00.000	15	10	1888	4
16	16	1889-01-14 00:00:00.000	14	1	1889	1
17	17	1889-11-23 00:00:00.000	23	11	1889	4
18	18	1890-01-01 00:00:00.000	1	1	1890	1
19	19	1890-03-09 00:00:00.000	9	3	1890	1
20	20	1890-11-21 00:00:00.000	21	11	1890	4
21	21	1891-01-01 00:00:00.000	1	1	1891	1
22	22	1891-01-02 00:00:00.000	2	1	1891	1
23	23	1891-04-30 00:00:00.000	30	4	1891	2
24	24	1891-05-01 00:00:00.000	1	5	1891	2
25	25	1891-05-20 00:00:00.000	20	5	1891	2
26	26	1891-11-15 00:00:00.000	15	11	1891	4
27	27	1892-01-01 00:00:00.000	1	1	1892	1
28	28	1892-10-28 00:00:00.000	28	10	1892	4
29	29	1893-01-01 00:00:00.000	1	1	1893	1
30	30	1893-02-01 00:00:00.000	1	2	1893	1
31	31	1893-05-08 00:00:00.000	8	5	1893	2

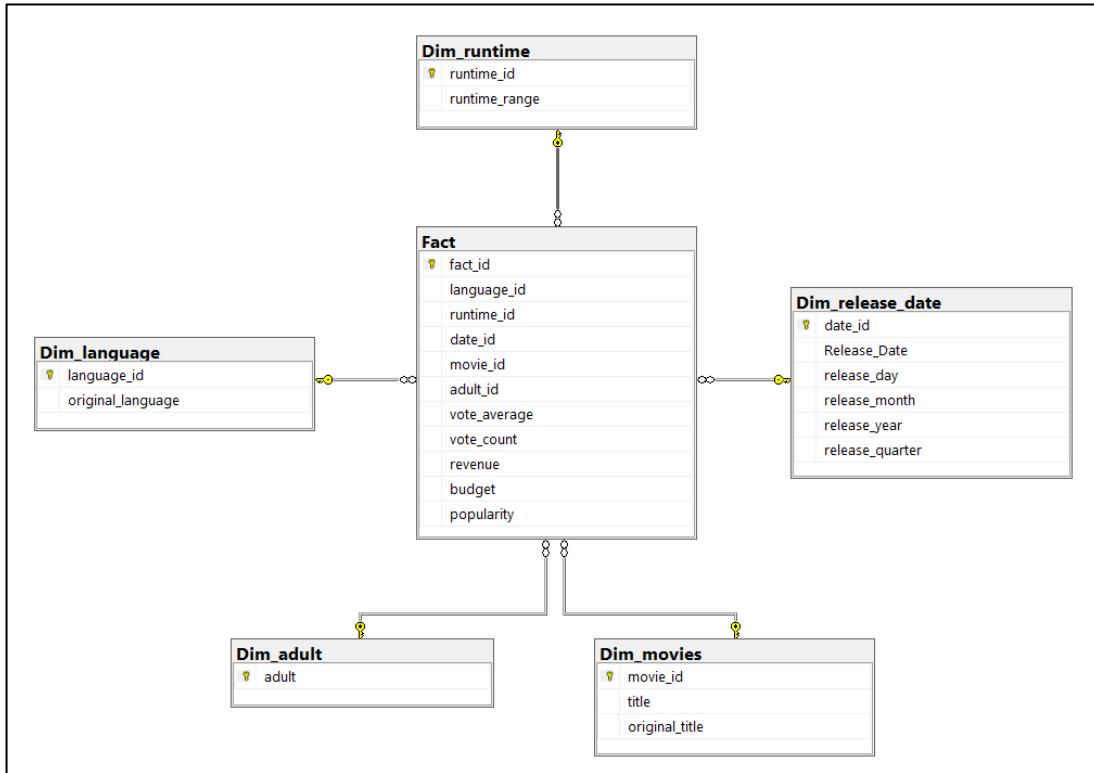
Kiểm tra dữ liệu bảng Dim_runtime

	runtime_id	runtime_range
1	1	1h - 2h
2	2	2h - 3h
3	3	over 3h
4	4	under 1h

Kiểm tra dữ liệu bảng Fact

	fact_id	language_id	runtime_id	date_id	movie_id	adult_id	vote_average	vote_count	revenue	budget	popularity
1	1	42	1	29031	2	0	7.1	262	0	0	8.155
2	2	42	1	28296	3	0	7.199	281	0	0	5.946
3	3	35	1	31635	5	0	5.784	2436	4257354	4000000	15.295
4	4	35	1	30850	6	0	6.533	302	12136938	21000000	13.564
5	5	35	1	35310	8	0	7.7	25	0	42000	1.587
6	6	31	4	34824	9	0	7.6	22	0	0	1.625
7	7	35	2	24875	11	0	8.204	19155	775398007	11000000	88.559
8	8	35	1	34363	12	0	7.824	18061	940335536	94000000	55.456
9	9	35	2	31101	13	0	8.477	25409	677387716	55000000	92.693
10	10	35	2	33010	14	0	8	11260	356296601	15000000	29.837
11	11	35	1	12100	15	0	8.015	5034	23218000	839727	28.218
12	12	35	2	33299	16	0	7.869	1618	40031879	12800000	18.067
13	13	35	1	35215	17	0	5.755	247	6593579	0	8.083
14	14	35	2	32145	18	0	7.53	9947	263920180	90000000	49.67
15	15	31	2	7237	19	0	8.153	2430	650422	1300000	22.817
16	16	35	1	34279	20	0	5.941	421	12300000	2500000	13.52
17	17	35	1	20884	21	0	7.245	98	0	0	8.638
18	18	35	2	34403	22	0	7.794	19233	655011224	140000...	80.509
19	19	35	1	34496	24	0	7.97	16188	180906076	30000000	36.315
20	20	35	2	35252	25	0	6.621	2618	97076152	72000000	26.349
21	21	54	1	34614	26	0	6.9	80	444	1400000	5.316
22	22	35	1	34776	27	0	5.653	462	1574623	1000000	30.475
23	23	35	2	25684	28	0	8.282	7485	150000000	31500000	37.715
24	24	35	2	30416	33	0	7.914	3960	159157447	14400000	31.894
25	25	35	1	35880	35	0	6.999	7430	527068851	75000000	59.717
26	26	35	1	34657	38	0	8.101	13709	72258126	20000000	54.124
27	27	37	2	33285	55	0	7.626	2335	20908467	2000000	26.344
28	28	35	2	35496	58	0	7.346	14868	1065659812	200000...	69.403
29	29	35	1	35210	59	0	7.168	2906	60740827	32000000	21.136
30	30	35	2	21539	62	0	8.076	10575	71923560	12000000	48.932
31	31	35	2	31655	63	0	7.6	7677	168841459	29000000	33.942

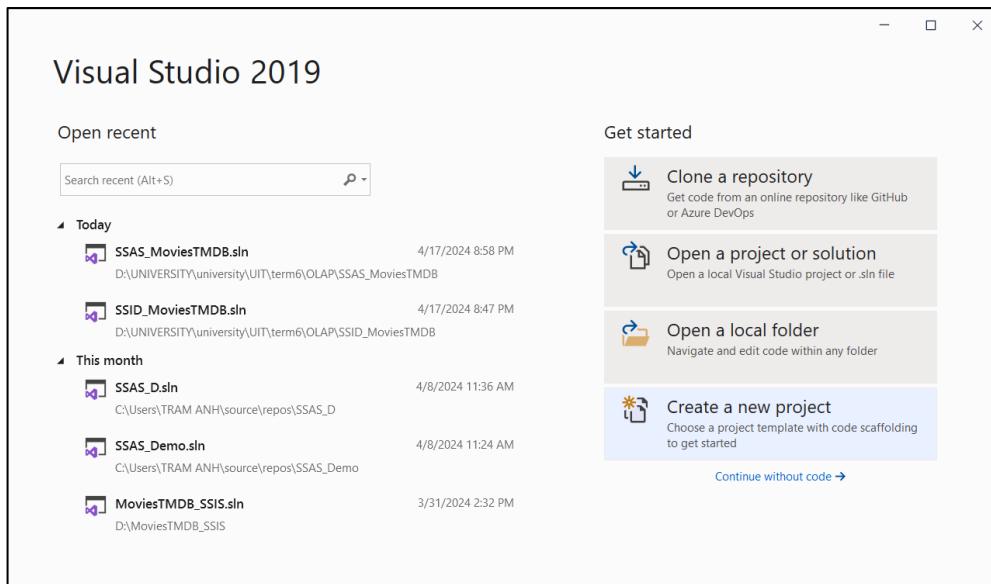
2.6. Lược đồ cơ sở dữ liệu sau khi hoàn thành



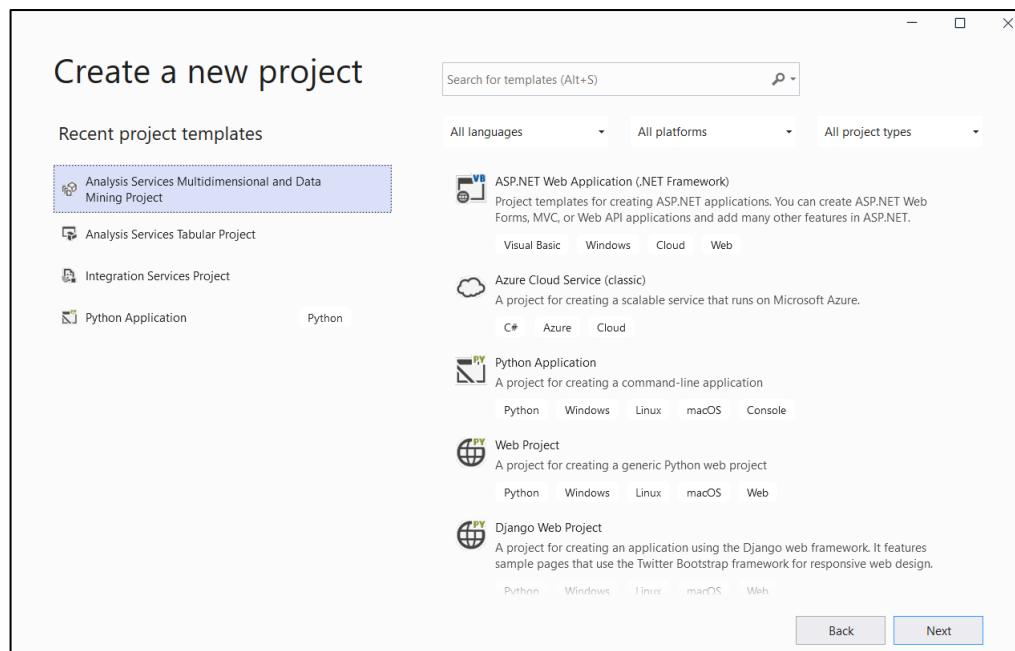
PHẦN 3: THỰC HIỆN PHÂN TÍCH TRÊN KHO DỮ LIỆU (SSAS)

3.1. Tạo project SSAS mới

Bước 1: Mở Visual Studio và chọn “Create a new project”.

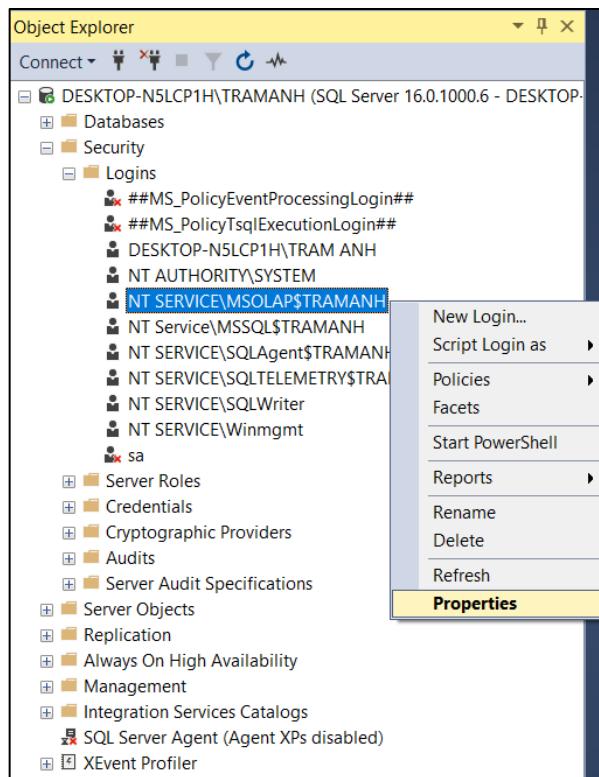


Bước 2: Chọn Analysis Services Multidimensional Project và chọn Next

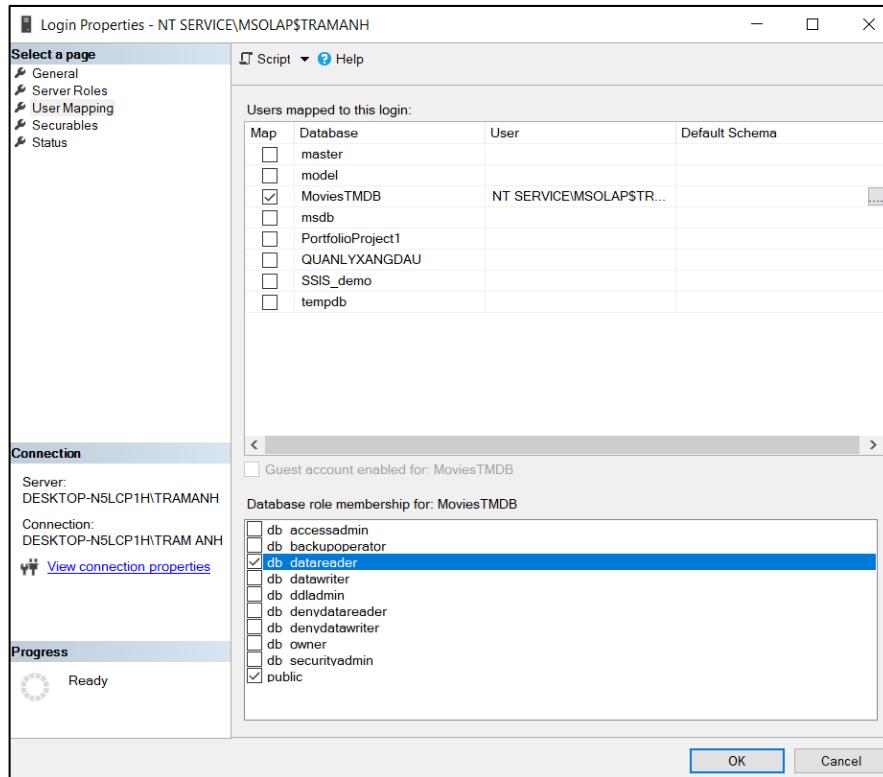


Bước 3: Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn Create.

Bước 4: Mở SQL Server Management. Kết nối Data Engine. Trong Object Explorer, ta chọn Security, rồi Logins và nhấp chuột phải vào User Service của OLAP. User của nhóm mặc định là NT SERVICE\MSOLAP\$TRAMANH. Click chuột phải và chọn Properties.

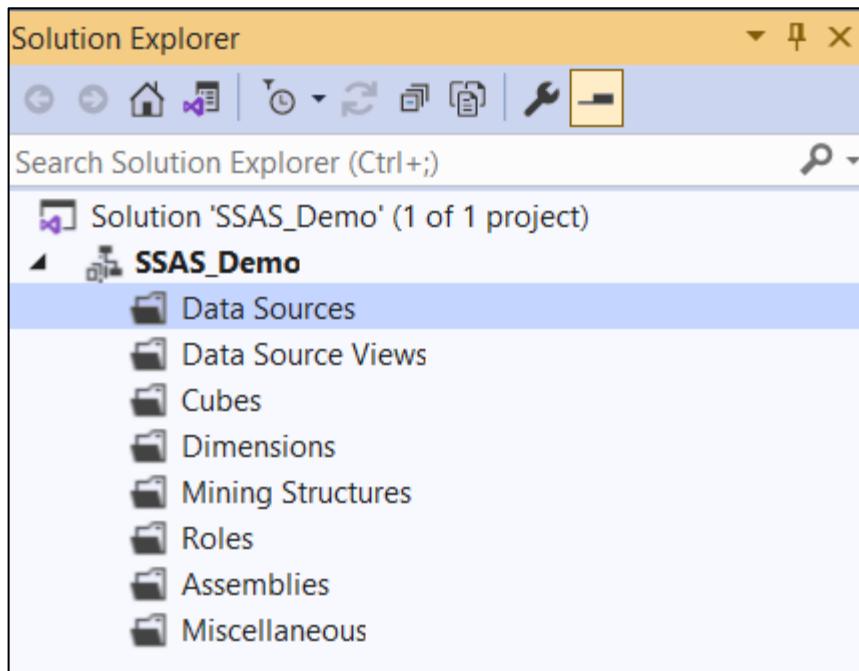


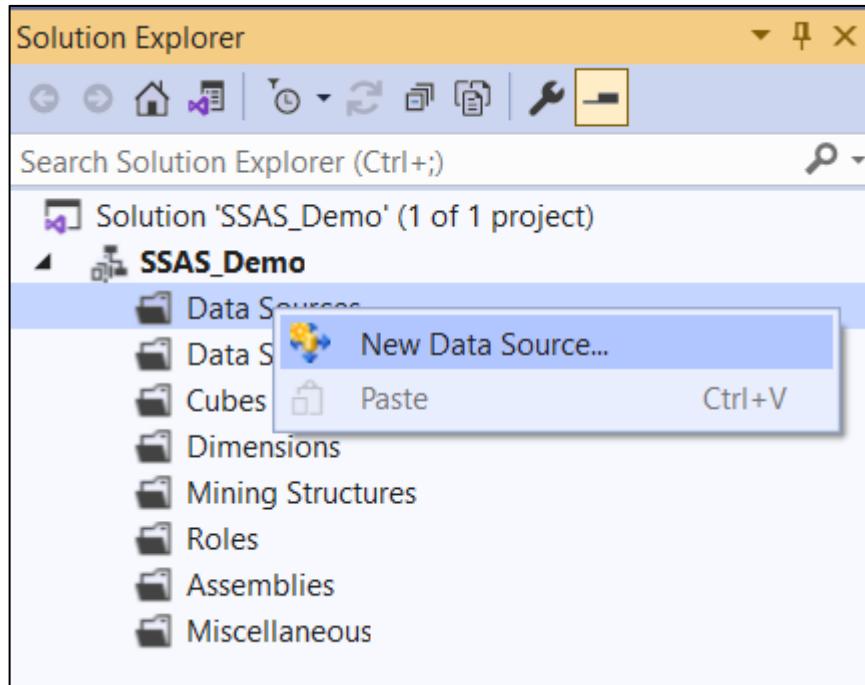
Bước 5: Màn hình ra cửa sổ Login Properties của tài khoản vừa chọn. Tại panel Select a page, chọn User Mapping. Tại cột Map, đánh dấu chọn cho cơ sở dữ liệu DWYoutubetrending. Bên dưới, chọn cấp quyền db datareader. Chọn Ok để xác nhận



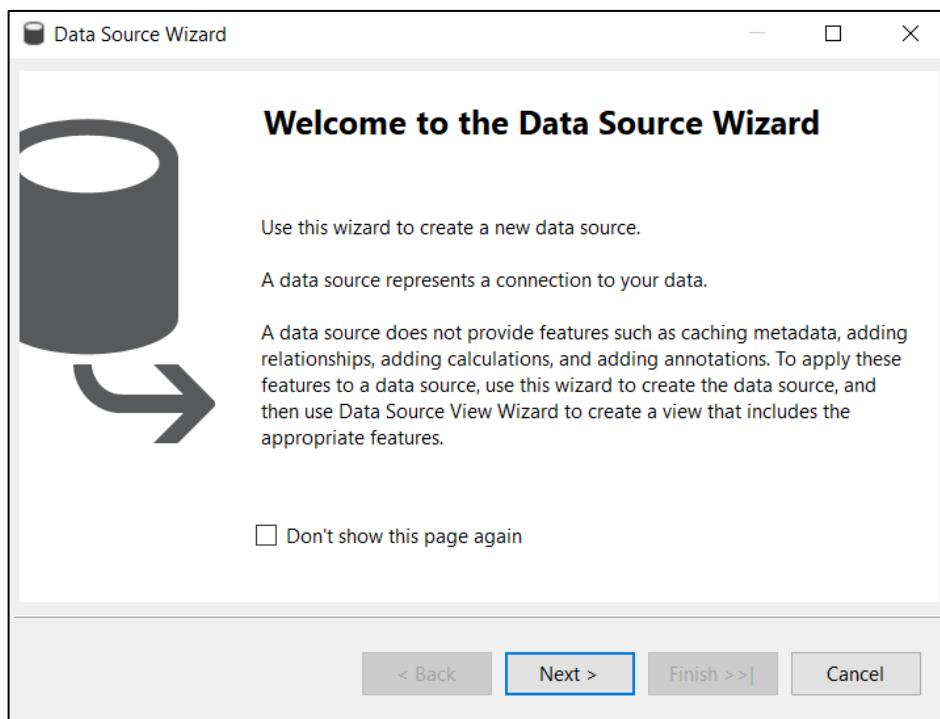
3.2. Xác định dữ liệu nguồn (Data Sources)

Bước 1: Tại Solution Explorer, ta click chuột phải vào thư mục Data Sources và chọn New Data Source.

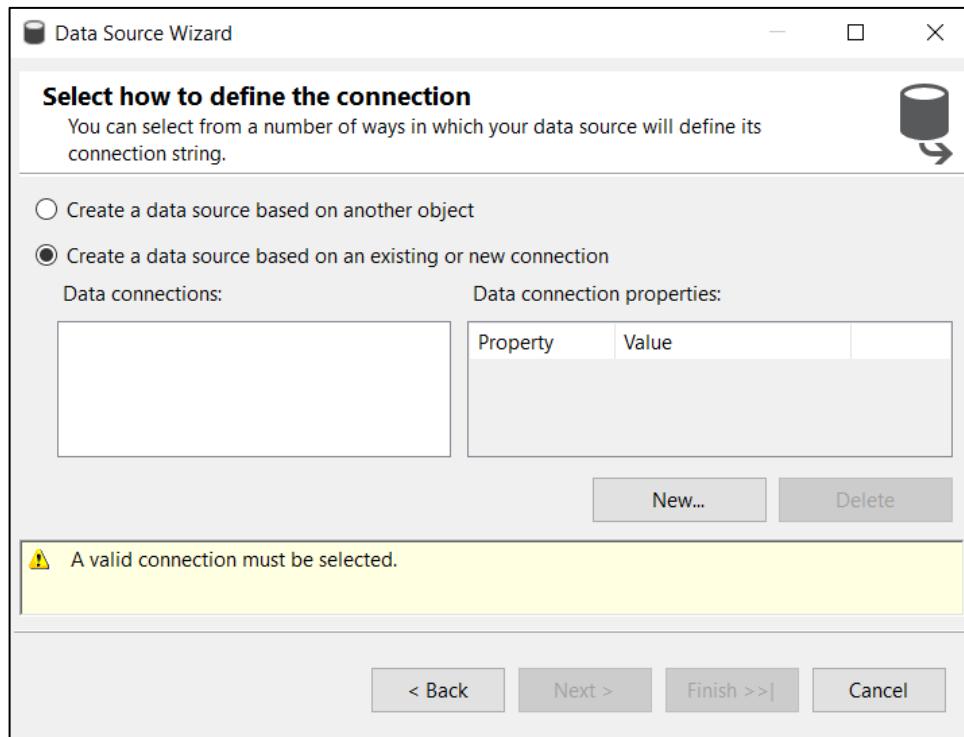




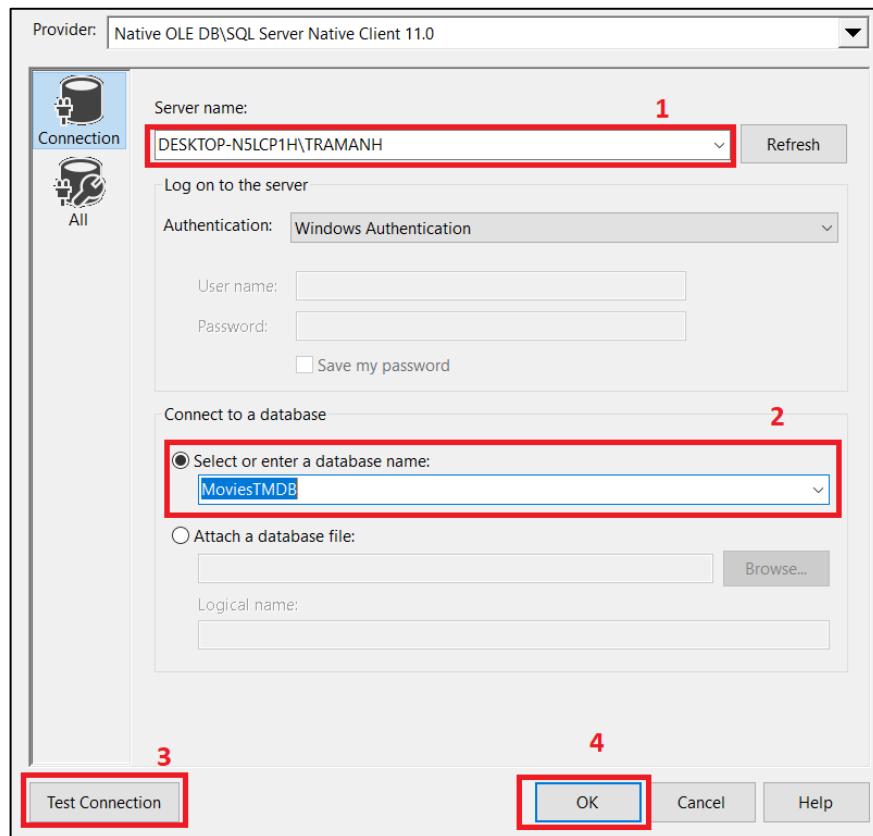
Bước 2: Hộp thoại Data Source Wizard xuất hiện, chọn Next để tiếp tục.



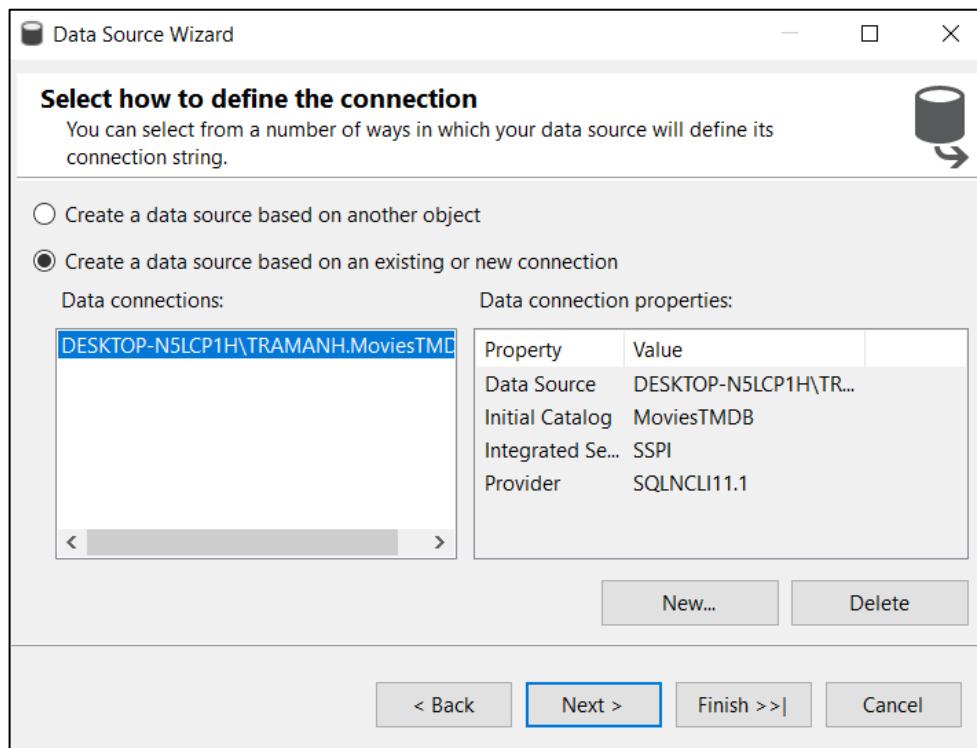
Bước 3: Chọn “Create a data source based on an existing or new connection” sau đó chọn New... để tạo kết nối với cơ sở dữ liệu đã được tạo từ quá trình SSIS.



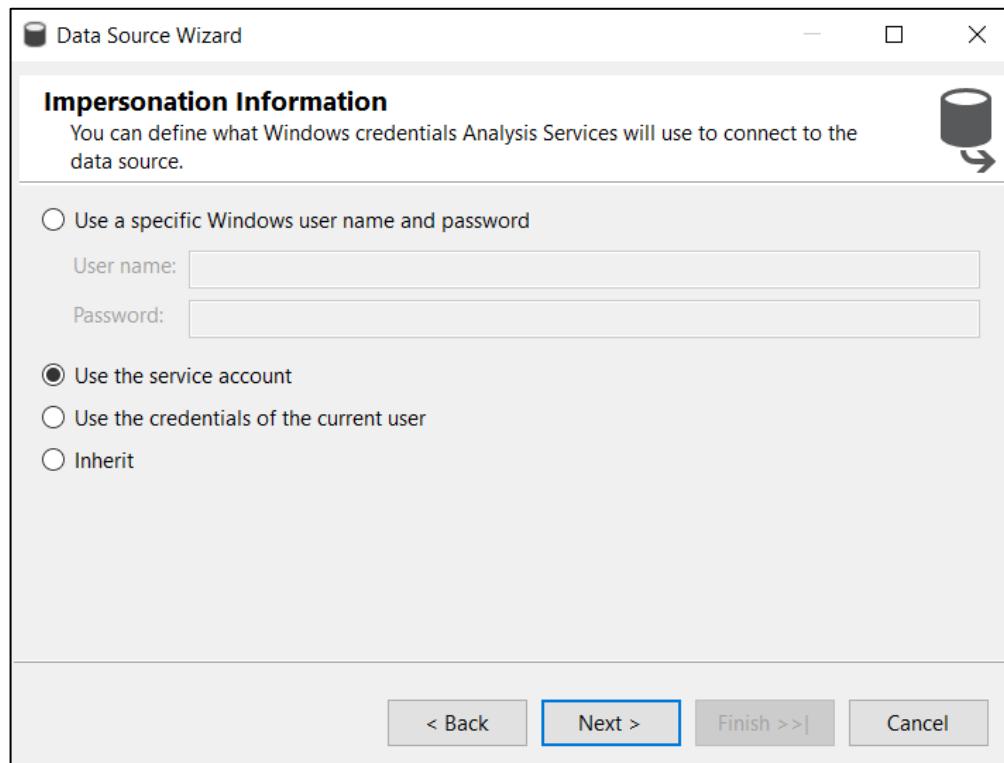
Bước 4: Hộp thoại Connection Manager xuất hiện, ta nhập Server name và chọn cơ sở dữ liệu mà ta đã tạo ra từ quá trình SSIS. Sau đó chọn OK.



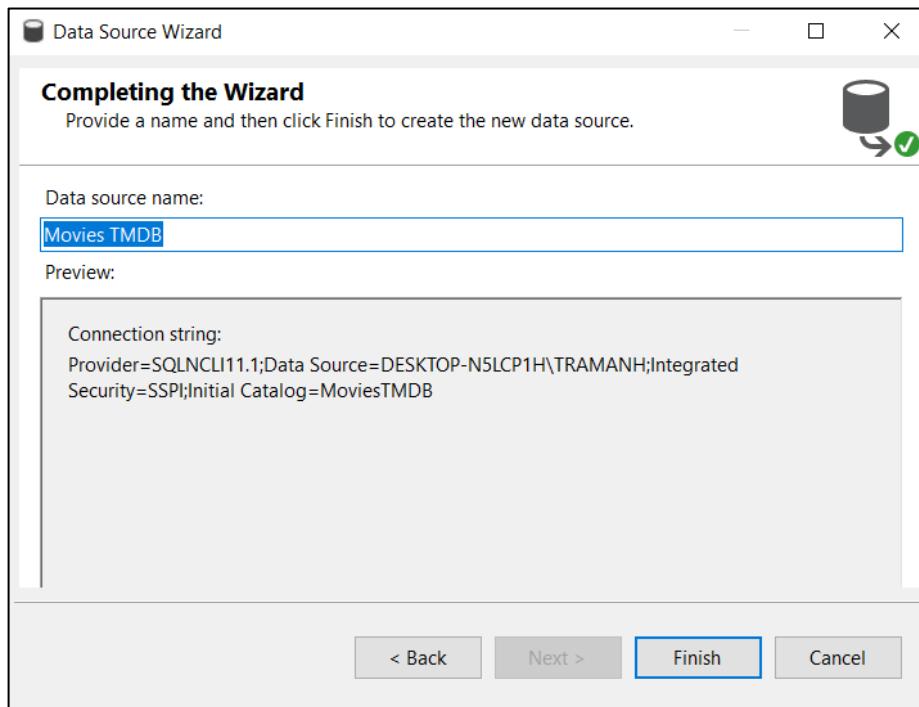
Chọn data source vừa tạo và chọn Next để tiếp tục.



Bước 5: Chọn “Use the service account”, sau đó chọn Next để tiếp tục.

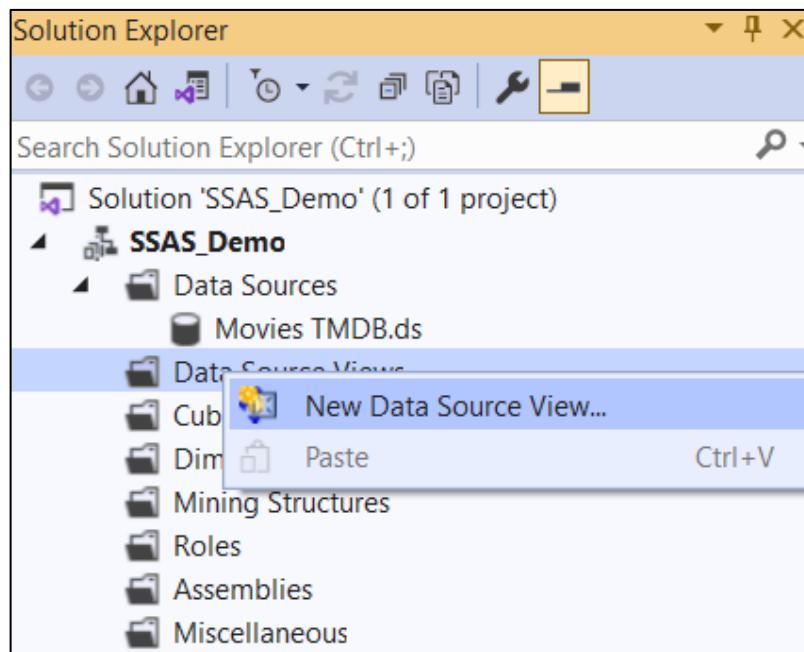


Bước 6: Cuối cùng ta chọn Finish để hoàn tất quy trình định nghĩa nguồn dữ liệu.

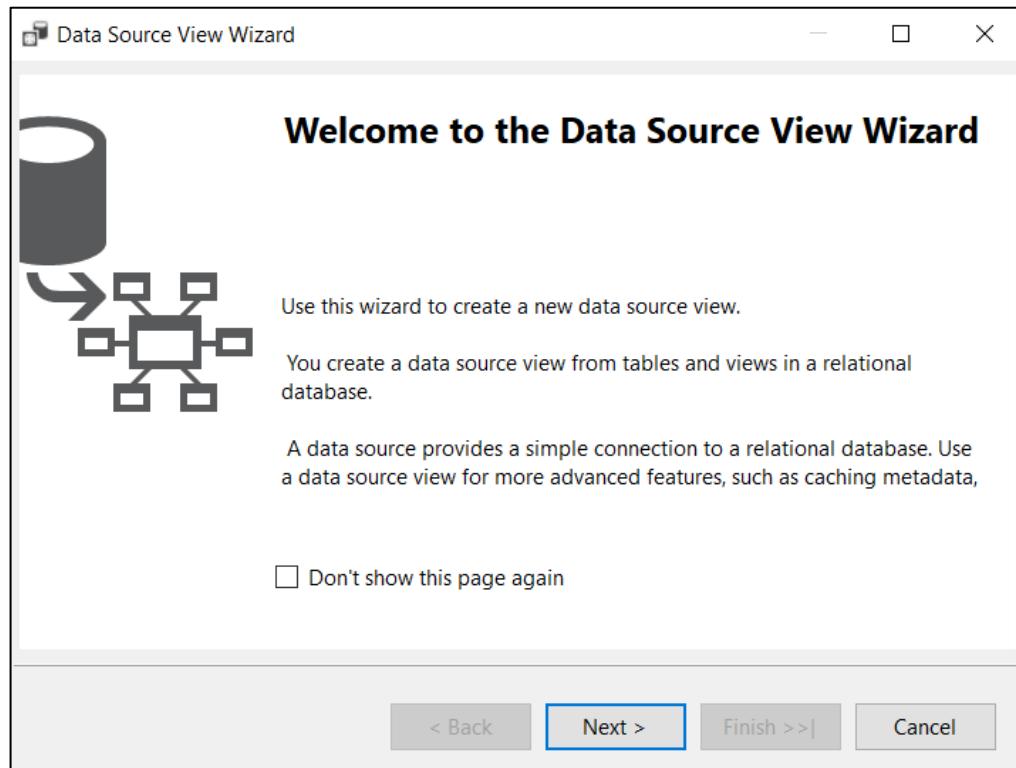


3.3. Xác định khung nhìn dữ liệu nguồn (Data source view)

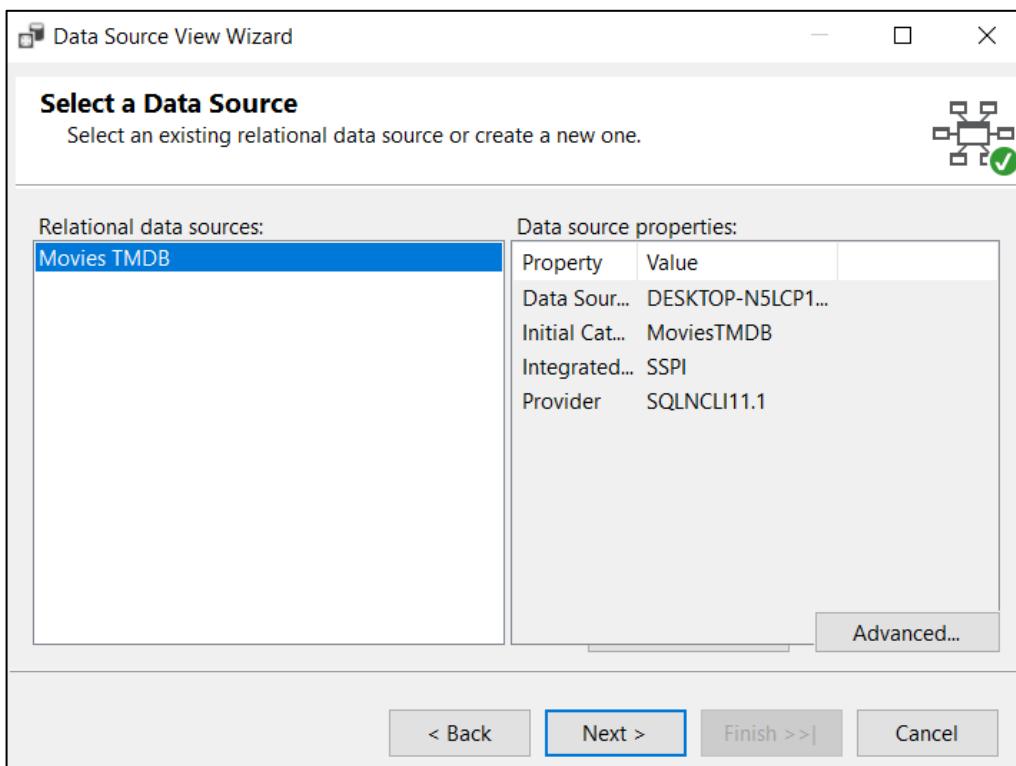
Bước 1: Tại Solution Explorer, ta click chuột phải vào thư mục Data Source Views và chọn New Data Source View.



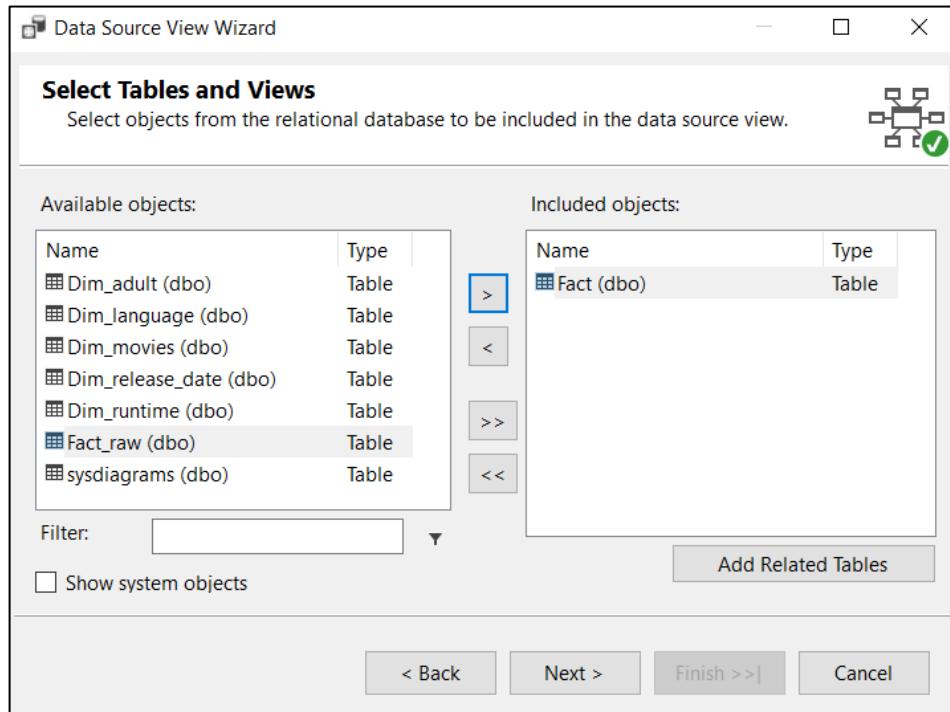
Bước 2: Hộp thoại Data Source View Wizard xuất hiện, chọn Next để tiếp tục



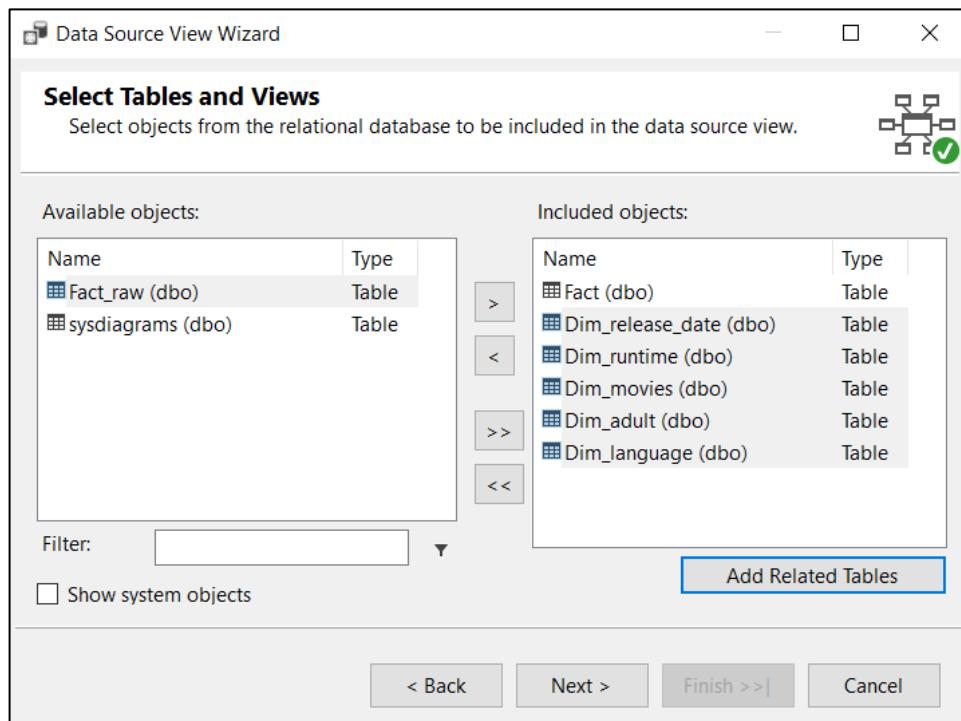
Bước 3: Chọn data source vừa tạo, sau đó chọn Next để tiếp tục.



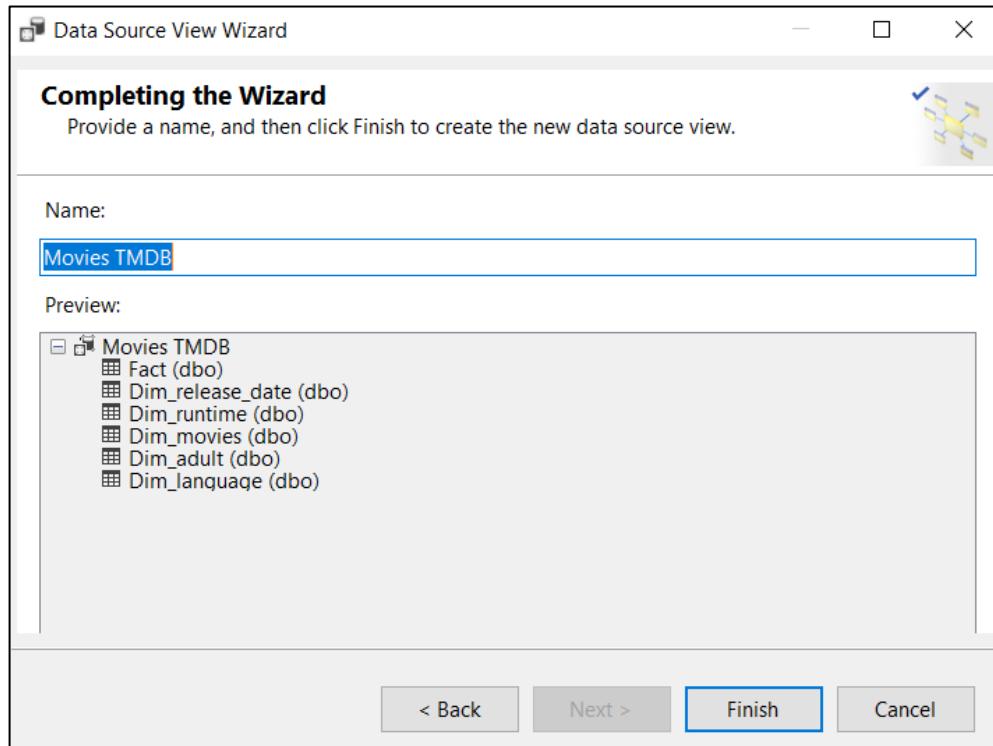
Bước 4: Chọn bảng Fact, sau đó chọn nút > để thêm bảng Fact vào data source view.



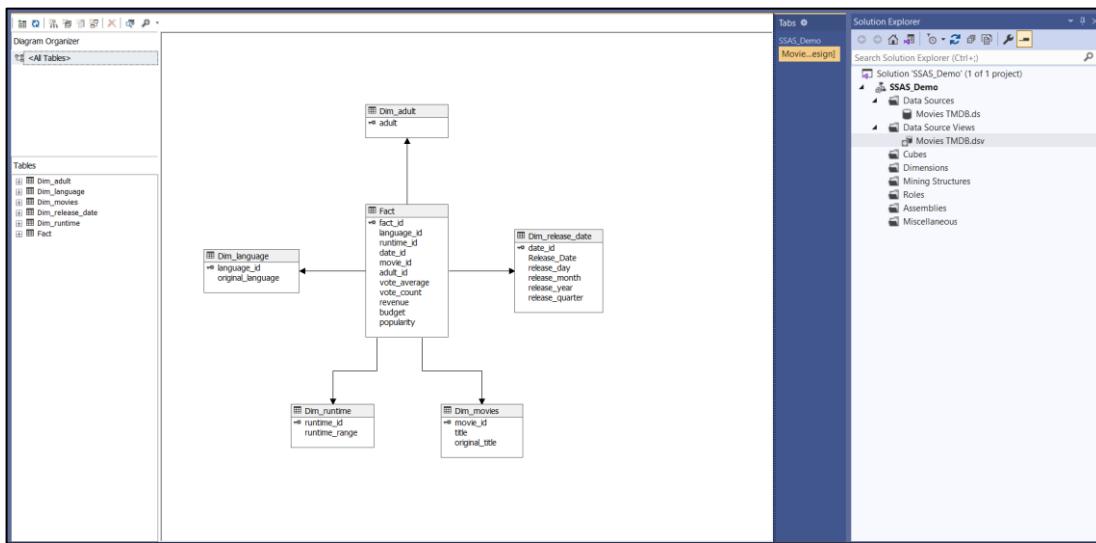
Bước 5: Tiếp theo, chọn nút Add Related Tables để thêm tất cả các bảng Dim vào data source view. Sau đó chọn Next để tiếp tục.



Bước 6: Chọn Finish để hoàn tất quá trình xác định khung nhìn dữ liệu nguồn (Data Source View).

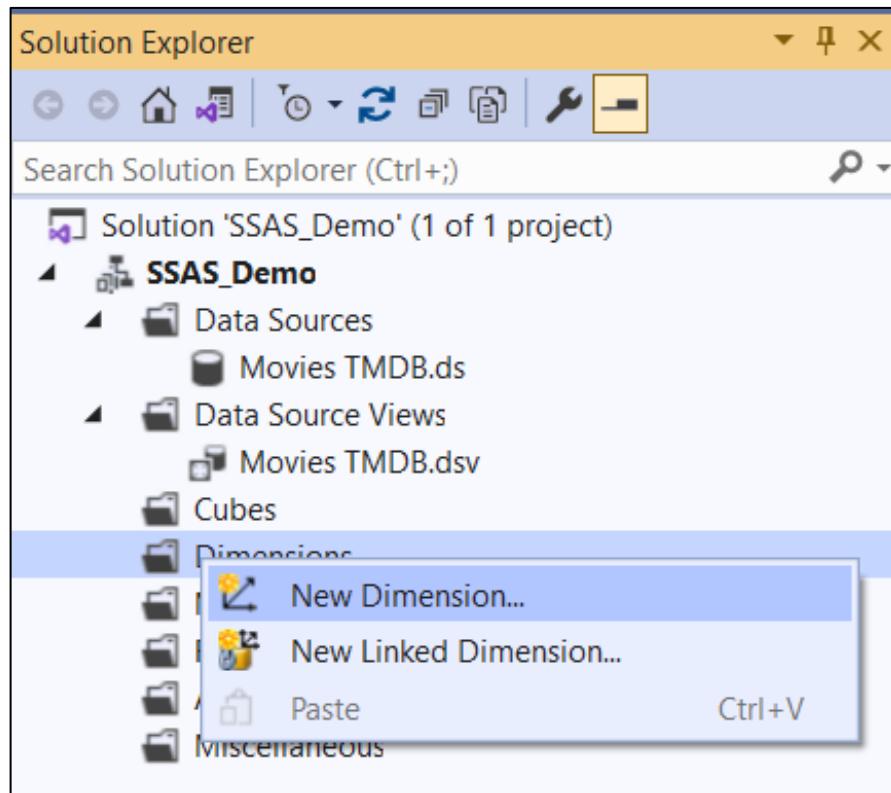


Sau khi kết thúc quá trình này, ta sẽ được data source view như hình sau

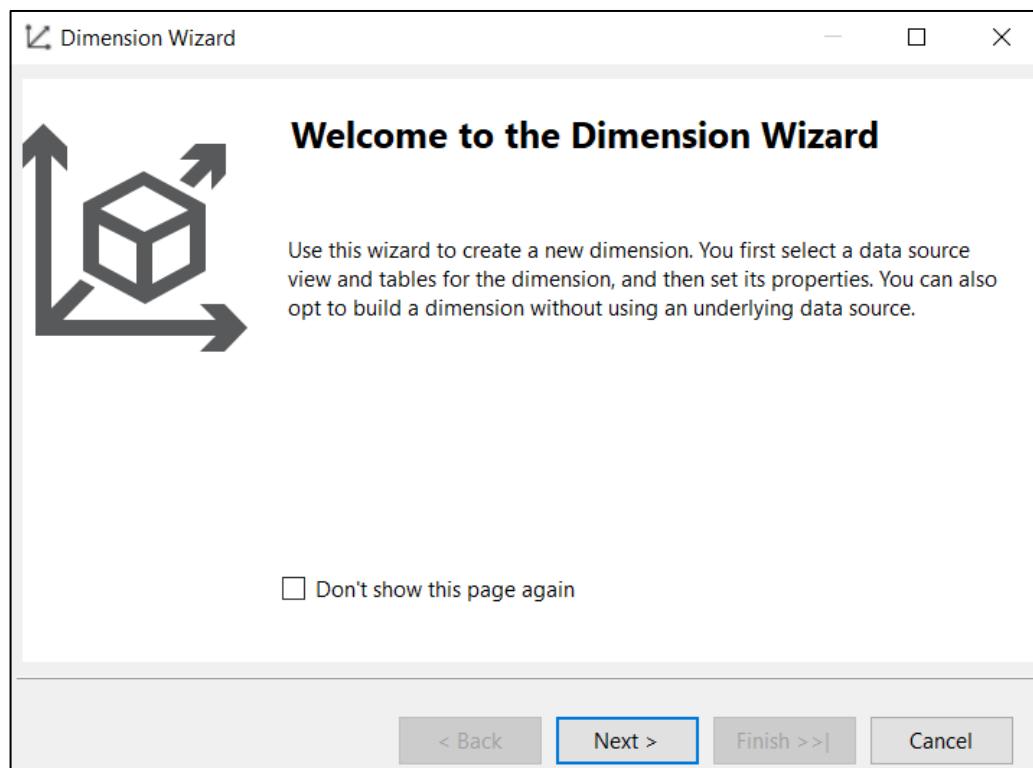


3.4. Tạo chiều thời gian

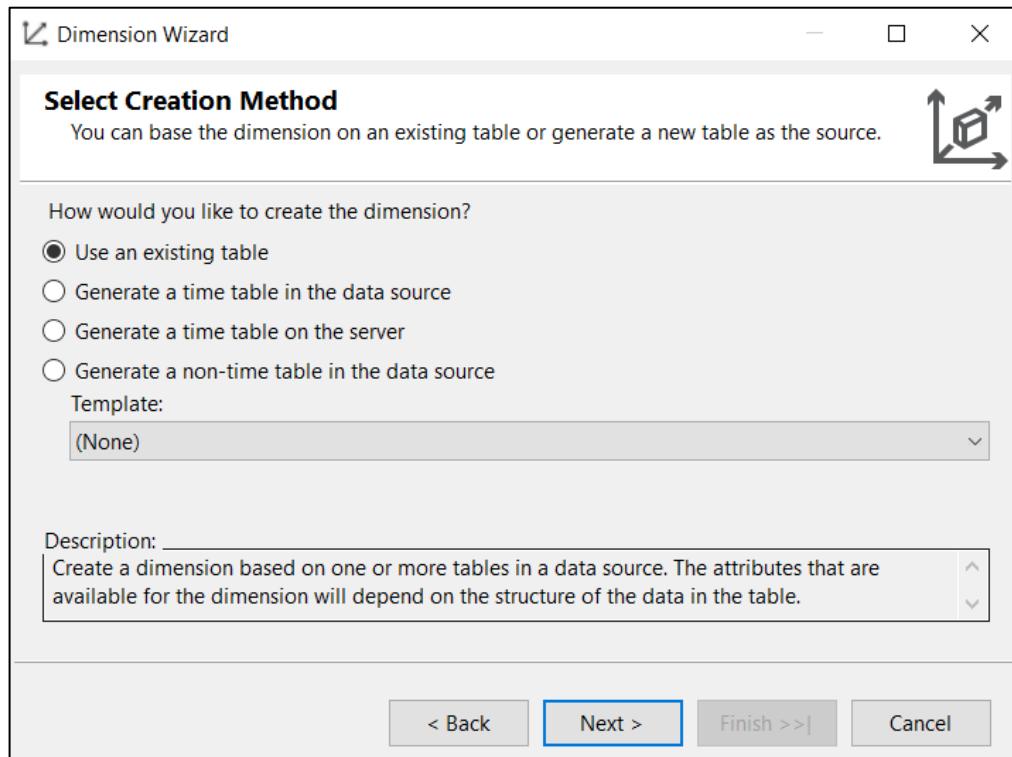
Bước 1: Nhấn chuột phải vào Dimensions. Chọn New Dimension.



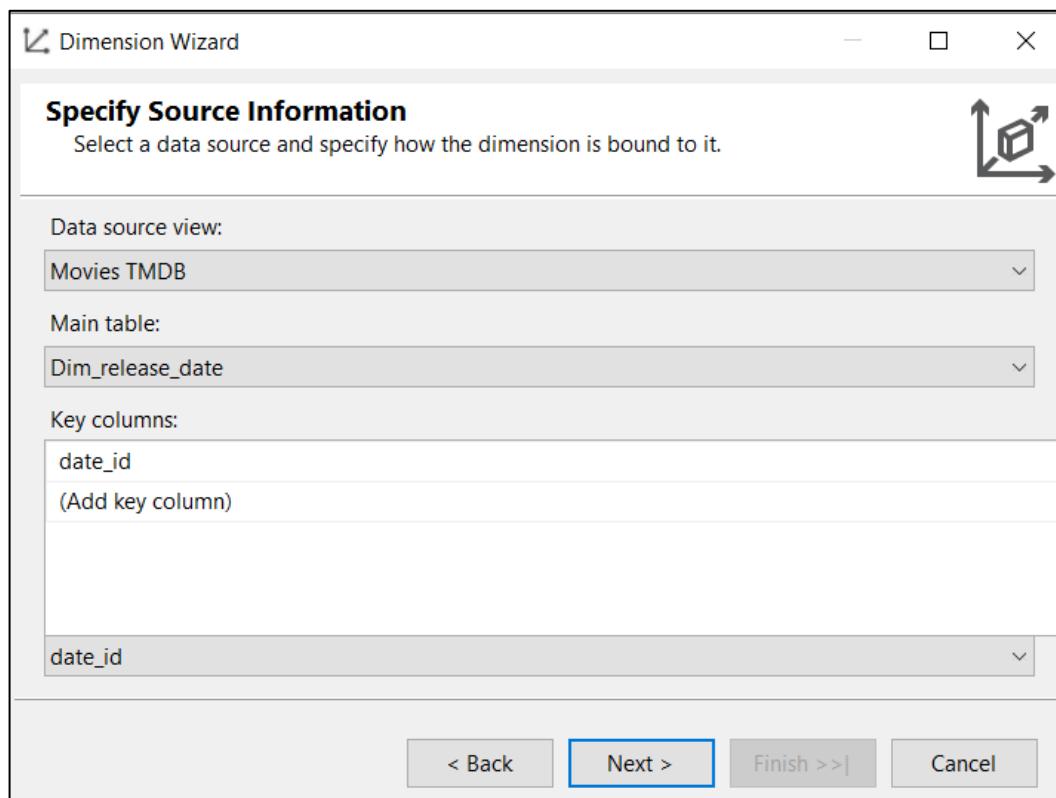
Bước 2: Cửa sổ Dimension Wizard xuất hiện. Nhấn Next để tiếp tục



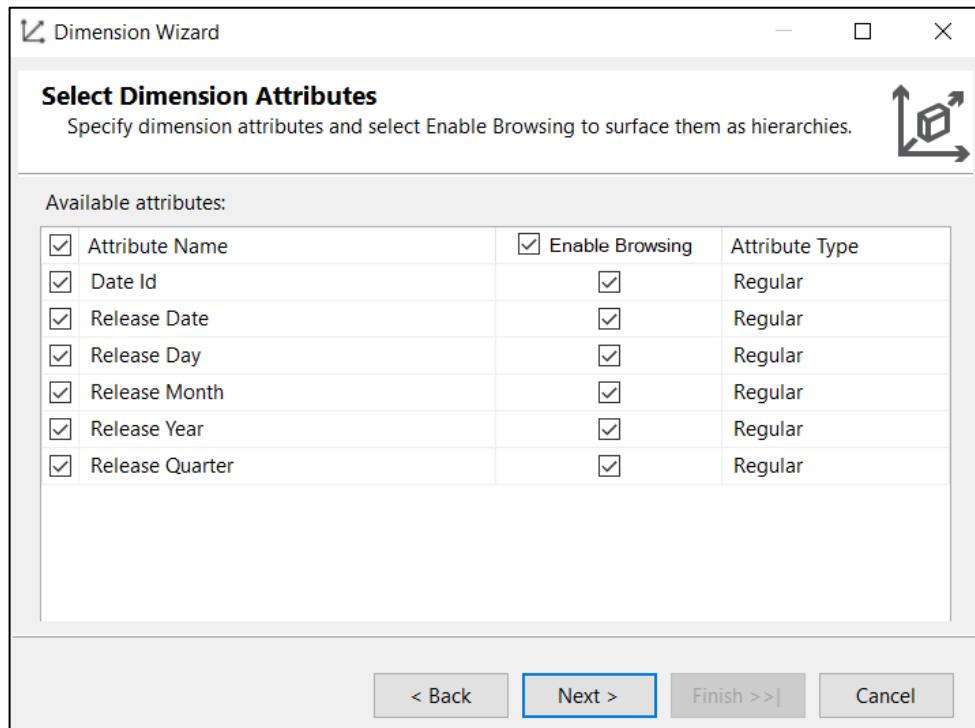
Bước 3: Giữ nguyên mặc định chọn Use an existing table. Chọn Next.



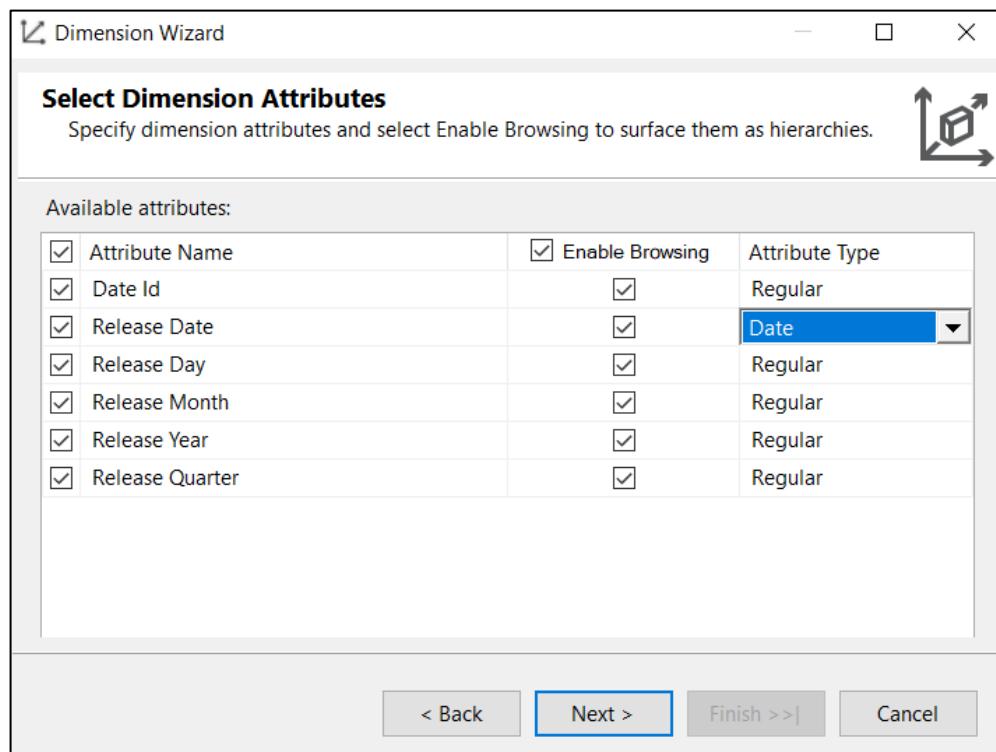
Bước 4: Tại Main table chọn Dim_release_date. Chọn Next để tiếp tục



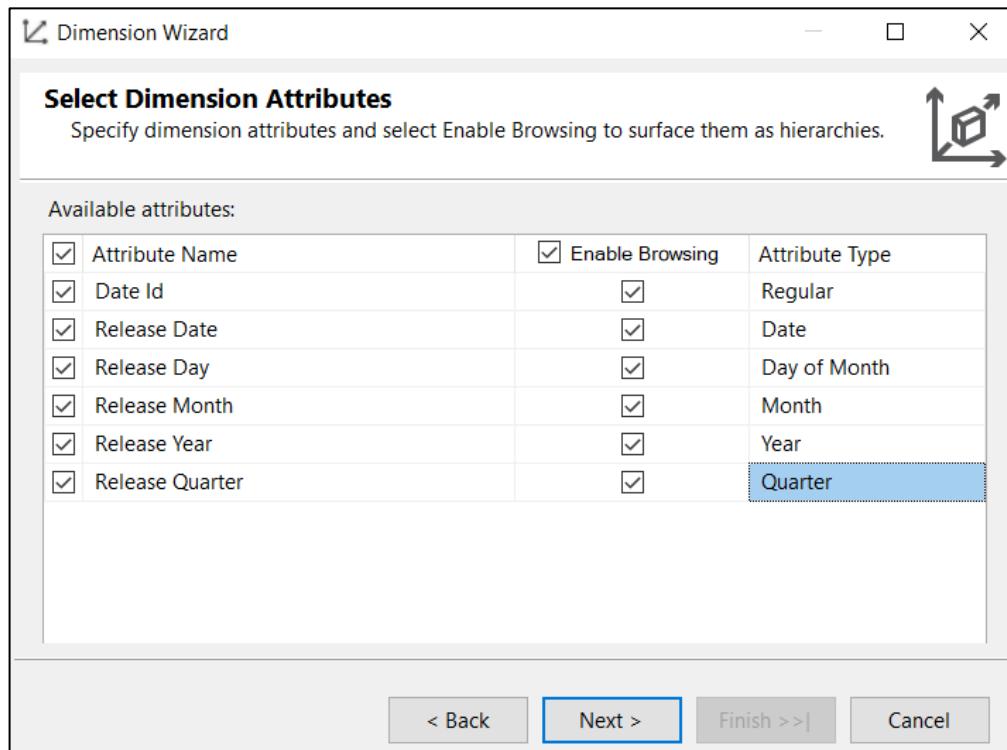
Bước 5: Chọn các thuộc tính cần thiết.



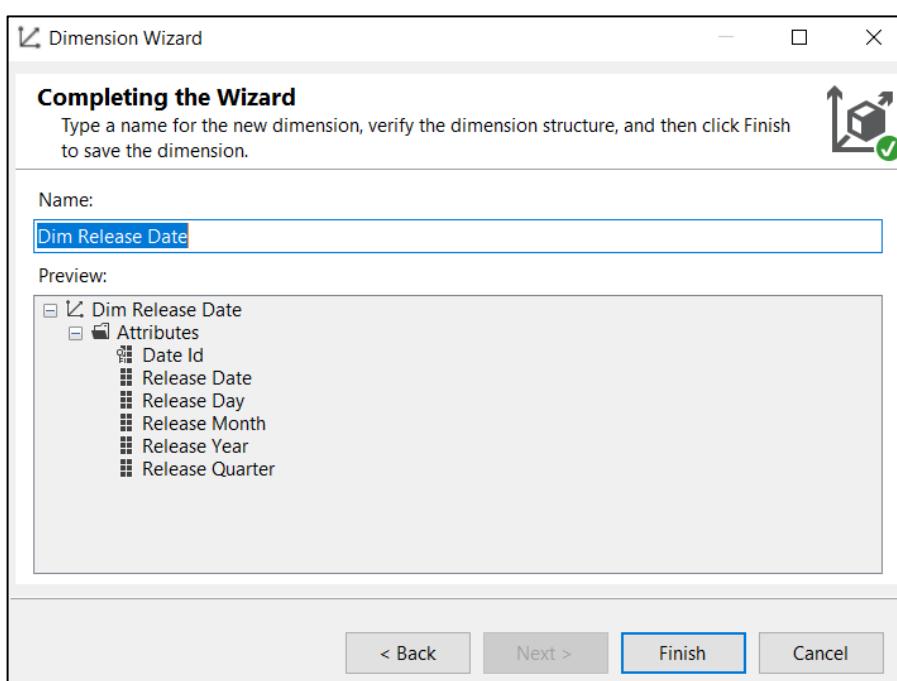
Bước 6: Tiến hành thay đổi Attribute Type. Đối với thuộc tính Release Date, chọn Date → Calender → Date.



Bước 7: Tương tự, Release Day có kiểu thuộc tính Day of Month, Release Month là Month, Release Quarter có thuộc tính là Quarter và Release Year là Year. Nhấn Next để tiếp tục.



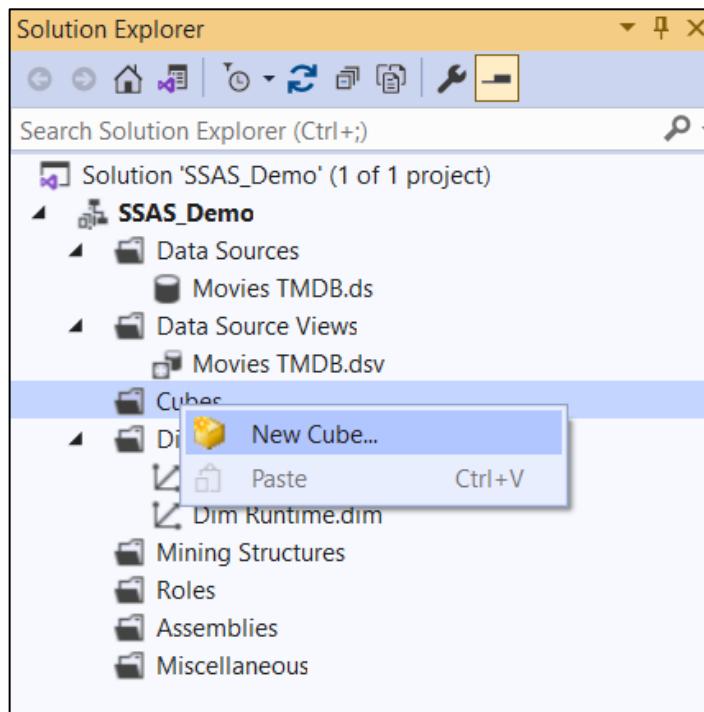
Đổi tên Cho Chiều nếu muốn. Nhấn Finish để kết thúc.



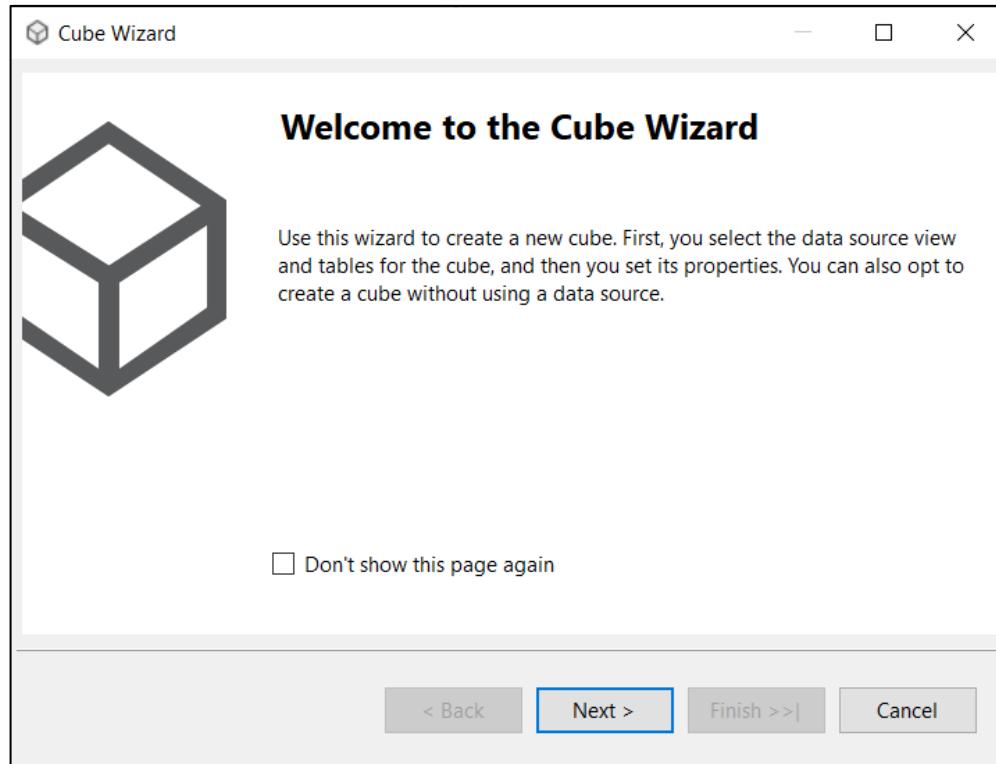


3.5. Tạo cube

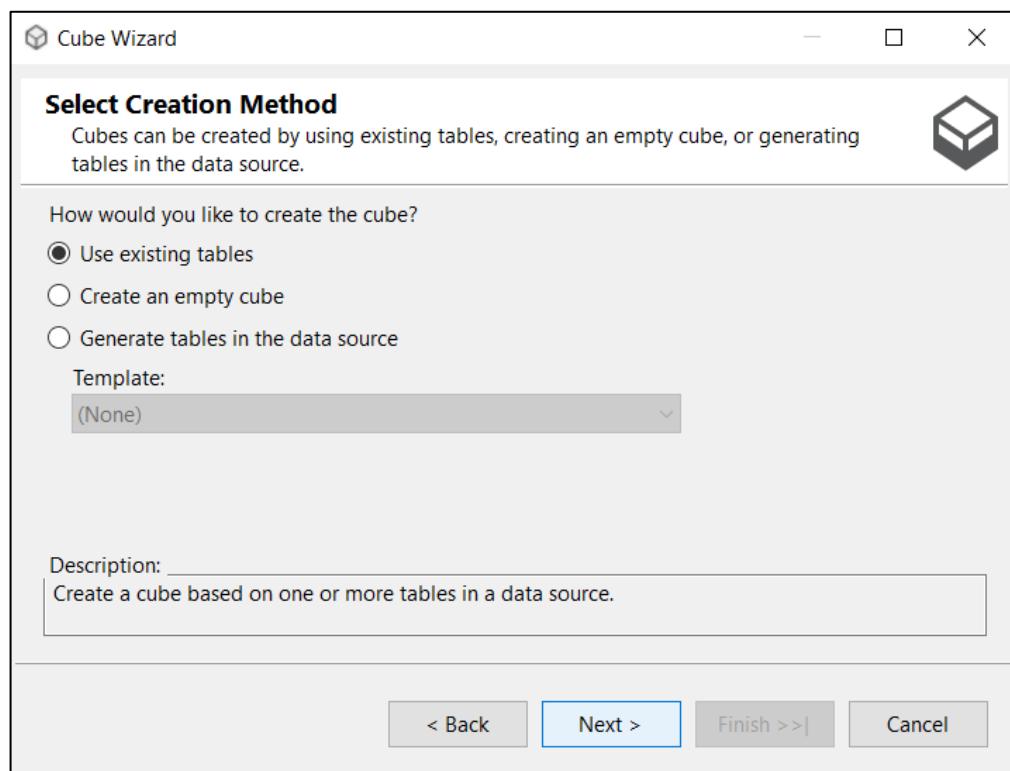
Bước 1: Nhấn chuột phải vào Cube. Chọn New Cube để tạo một cube.



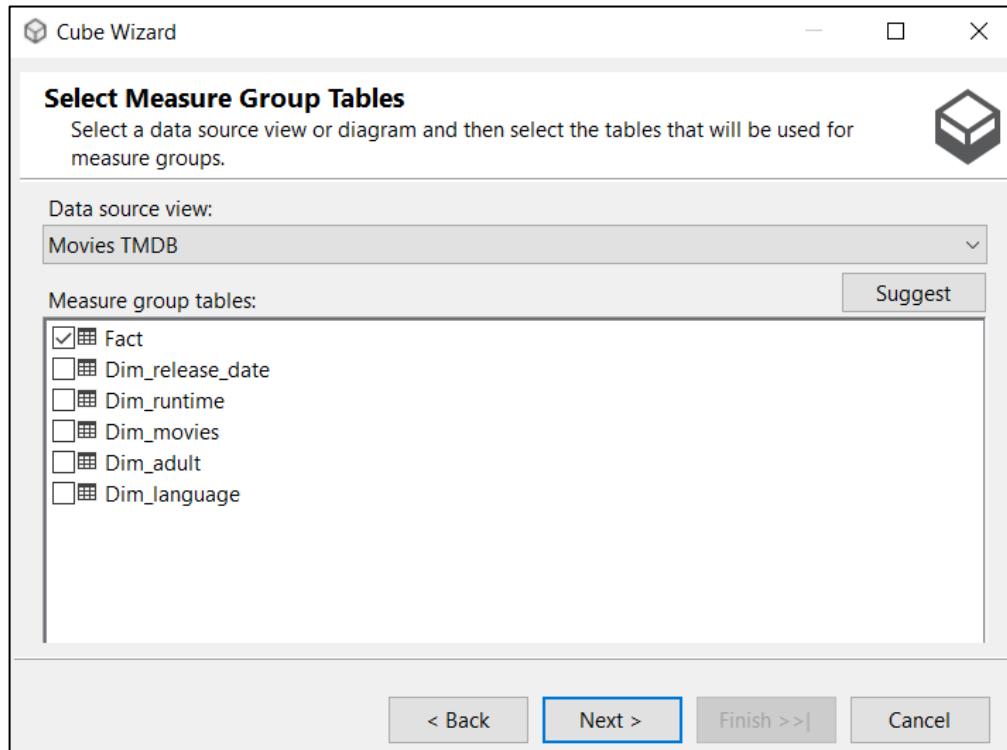
Bước 2: Cửa sổ Cube Wizard xuất hiện. Nhấn Next.



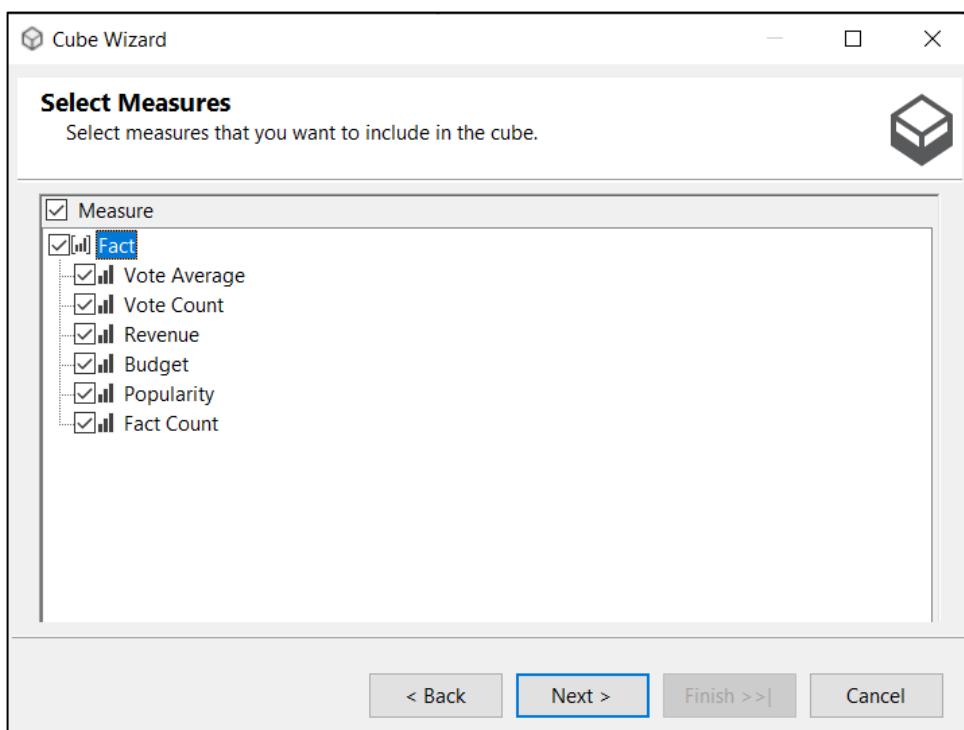
Bước 3: Tại Select Creation Method, giữ mặc định Use existing tables. Chọn Next.



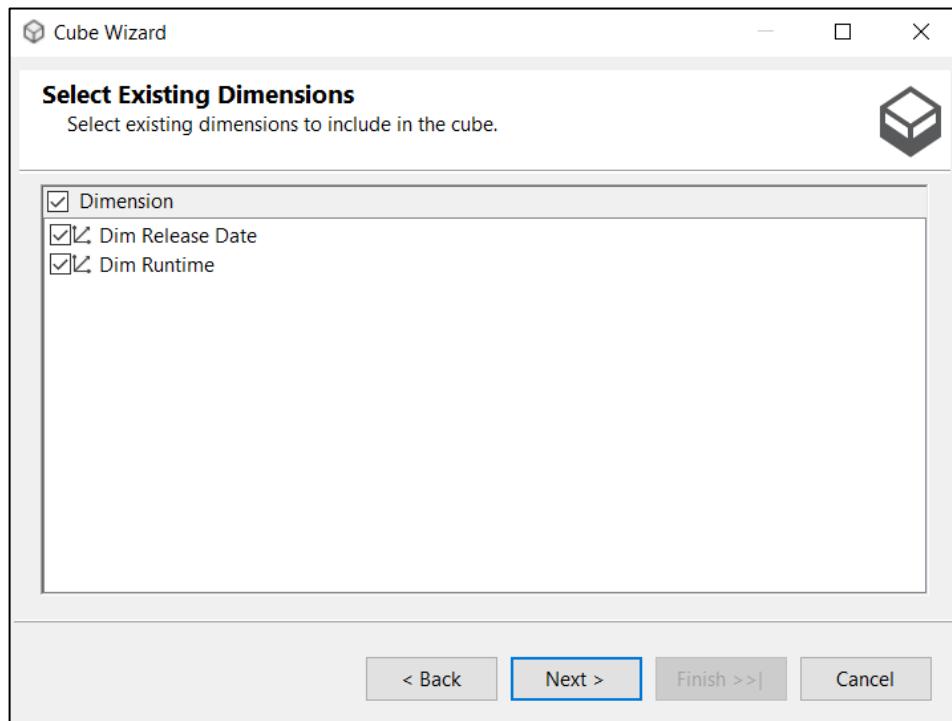
Bước 4: Tại Select Measure Group Tables, chọn Data Source View là MoviesTMDB, đánh dấu vào Fact. Nhấn Next.



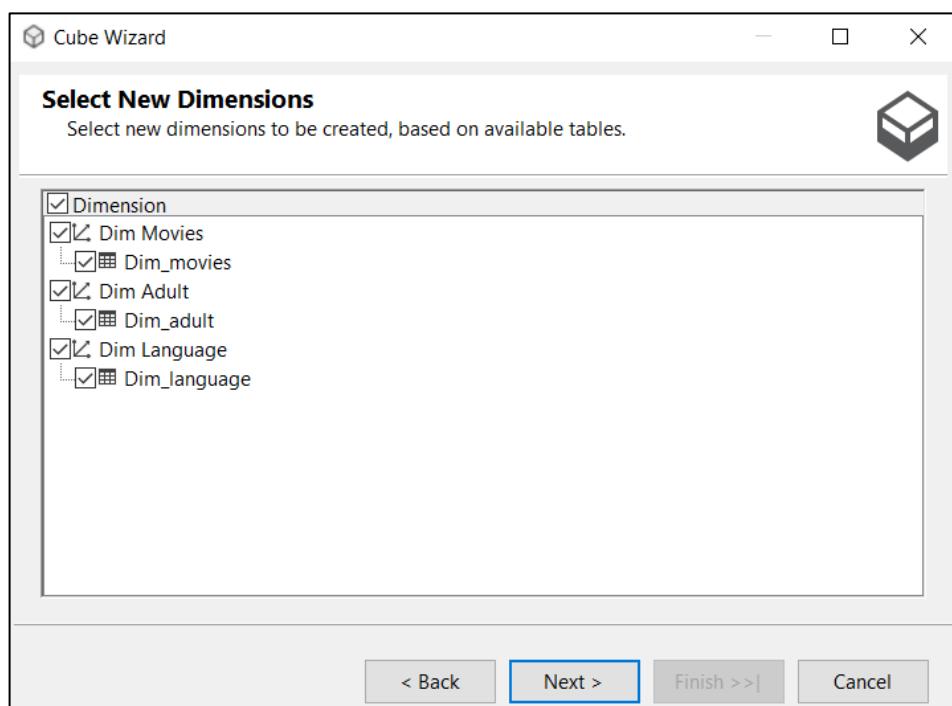
Bước 5: Trong phần Select Measures, giữ nguyên mặc định và nhấn Next.



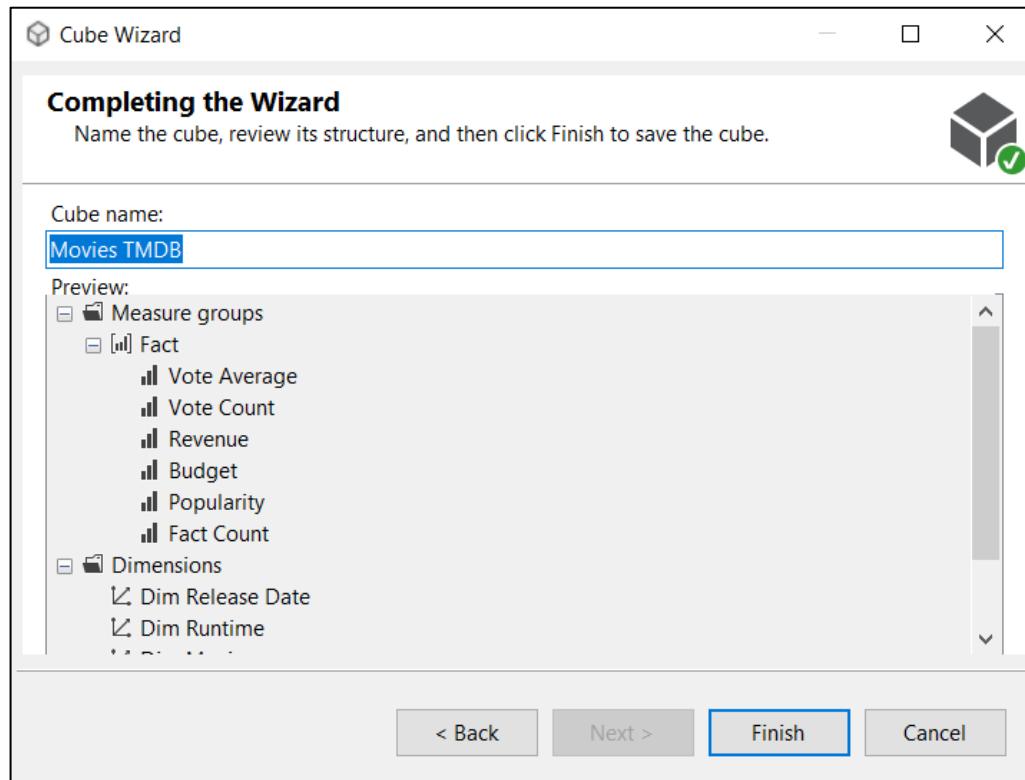
Bước 6: Chuyển sang phần Select Existing Dimension, kiểm tra đã chọn 2 Dimension thời gian hay chưa.



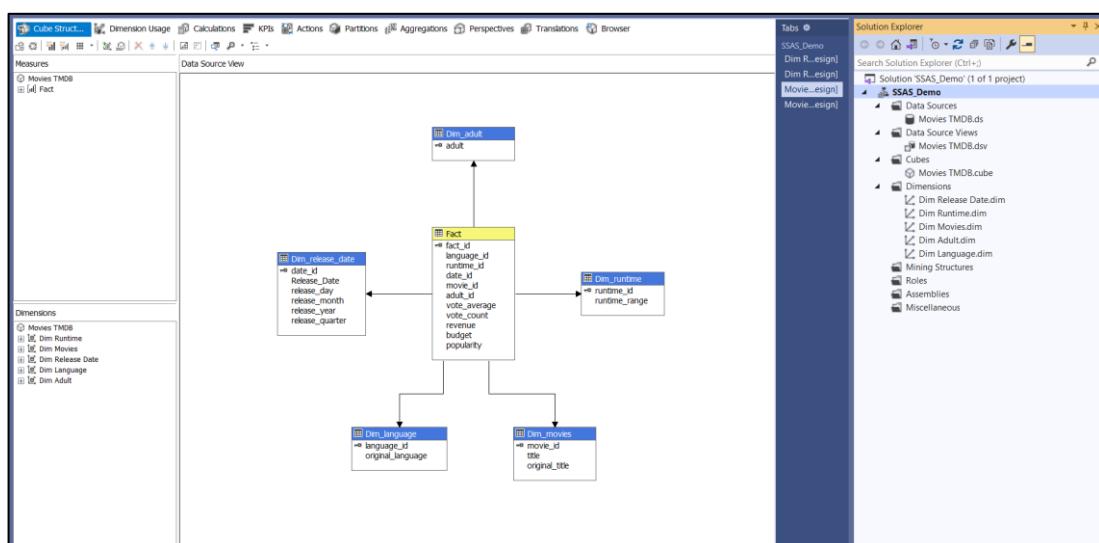
Bước 7: Tại Select New Dimension, đánh dấu chọn tất cả để tự động tạo chiều liên quan.

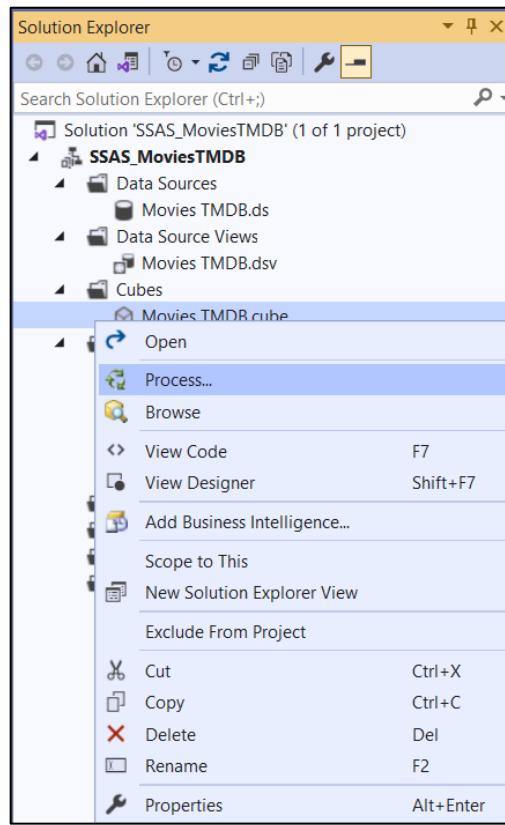


Bước 8: Đổi tên Cube nêu muộn và nhấn Finish.

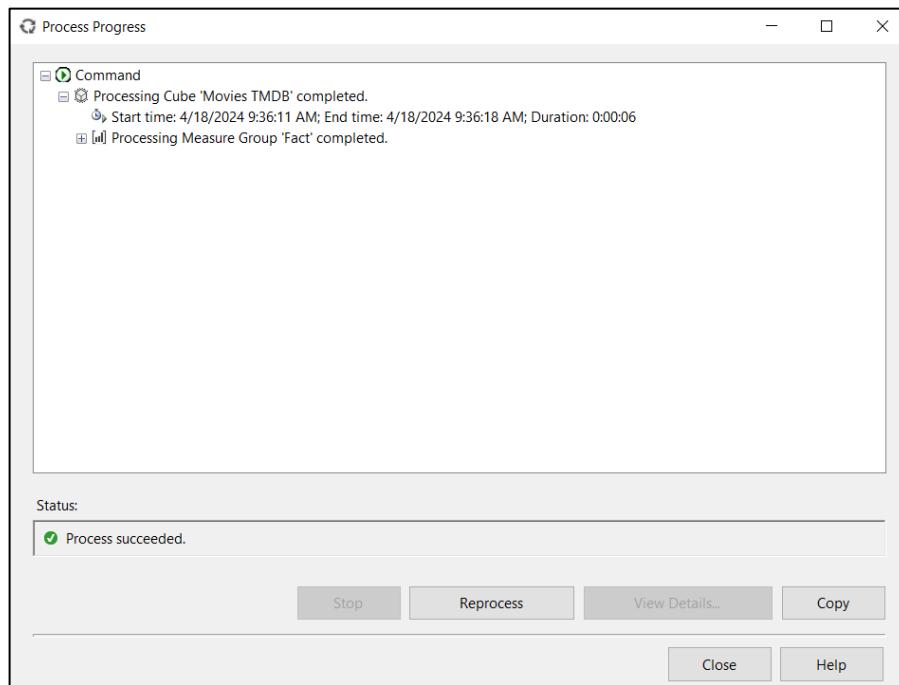


Ta được lược đồ như sau



Bước 9: Chuột phải vào MoviesTMDB.cube và chọn Process

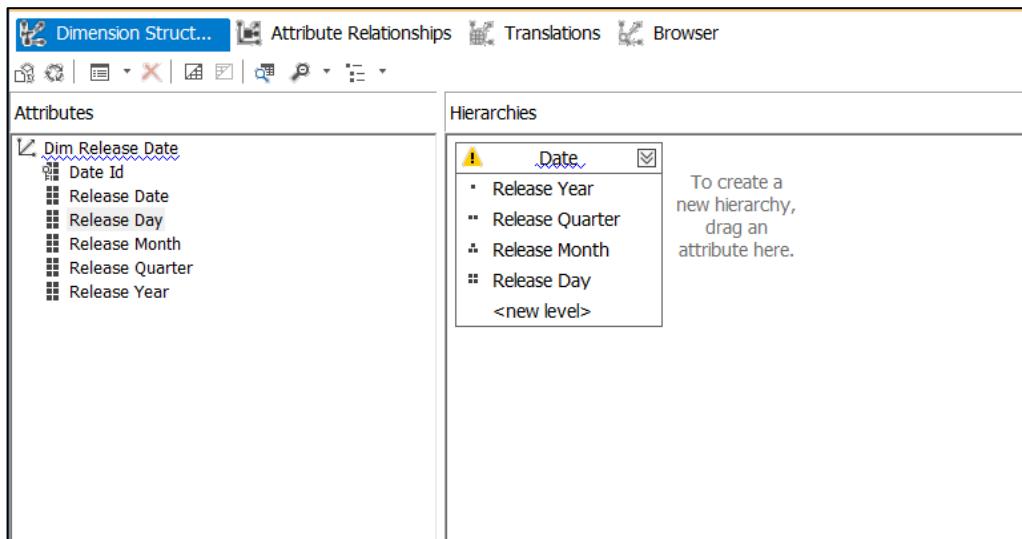
Nếu triển khai thành công, màn hình hiển thị thông báo tính trạng deploy như hình dưới



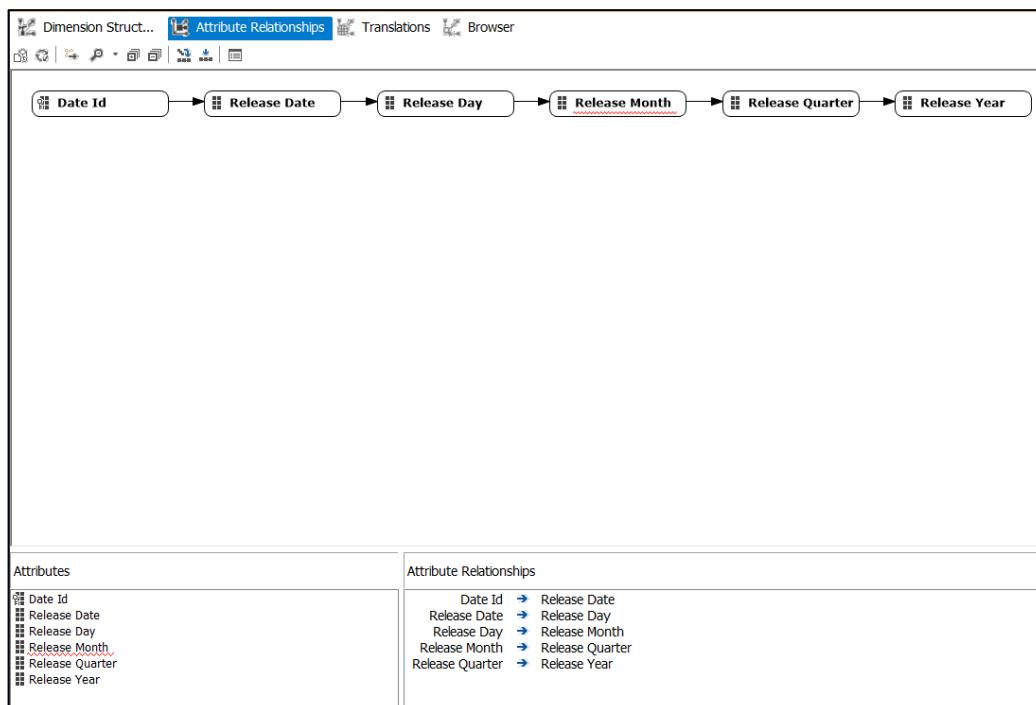
3.6. Điều chỉnh các thuộc tính của cube

3.6.1. Tạo hierarchy cho chiều thời gian

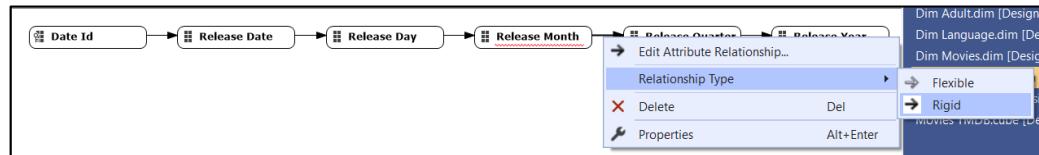
Bước 1: Đối với Dim Release Date , tạo hierarchies theo thứ tự Year > Quarter > Month > Day.



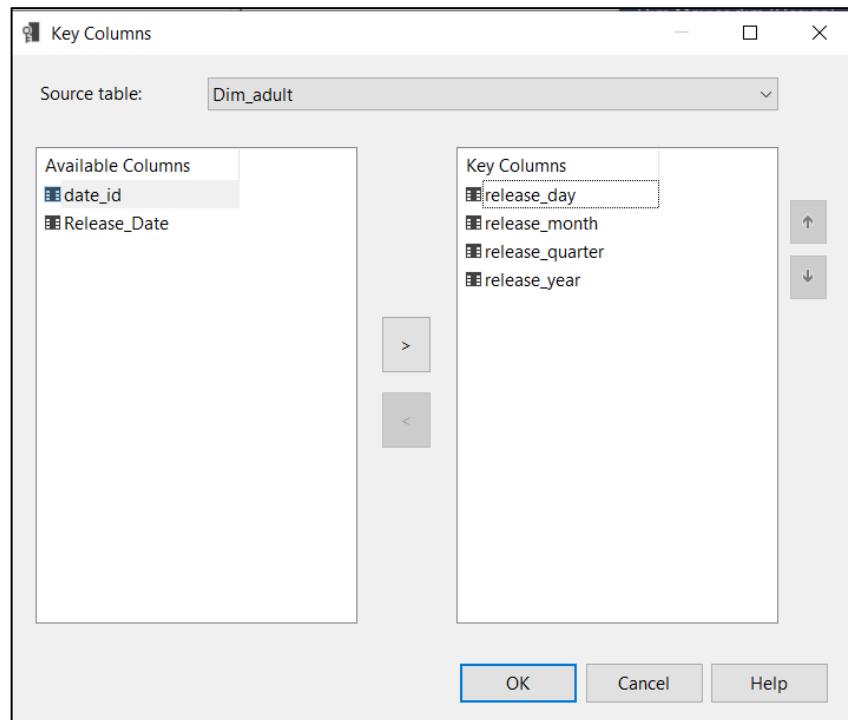
Bước 2: Tại panel Attribute Relationships, tạo mối quan hệ như sau



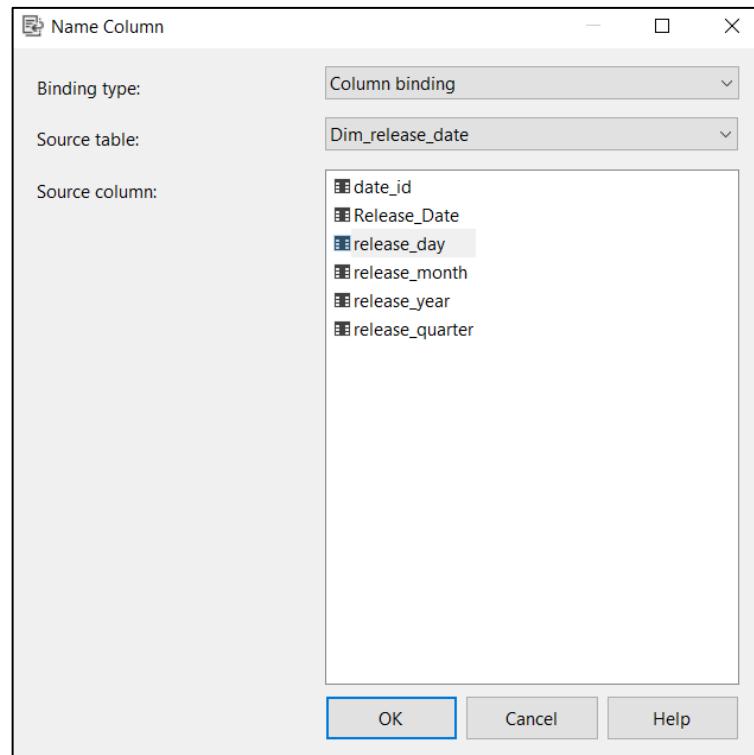
Bước 3: Nhấn chuột phải vào các đường dẫn. Trỏ chuột vào Relationship Type. Đổi thành Rigid.



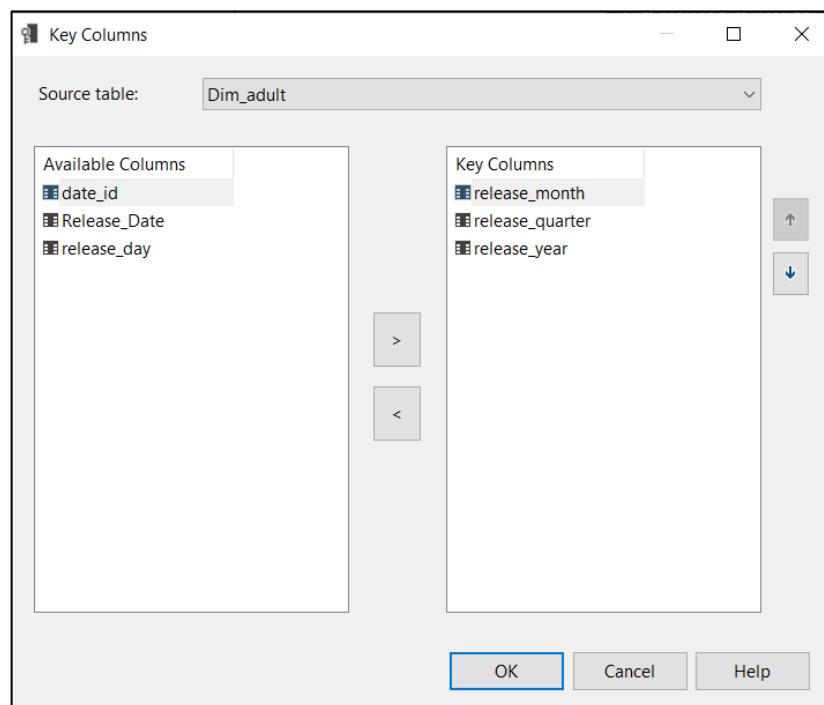
Bước 4: Thêm Key Columns là release_year, release_quarter, release_month cho release_day.



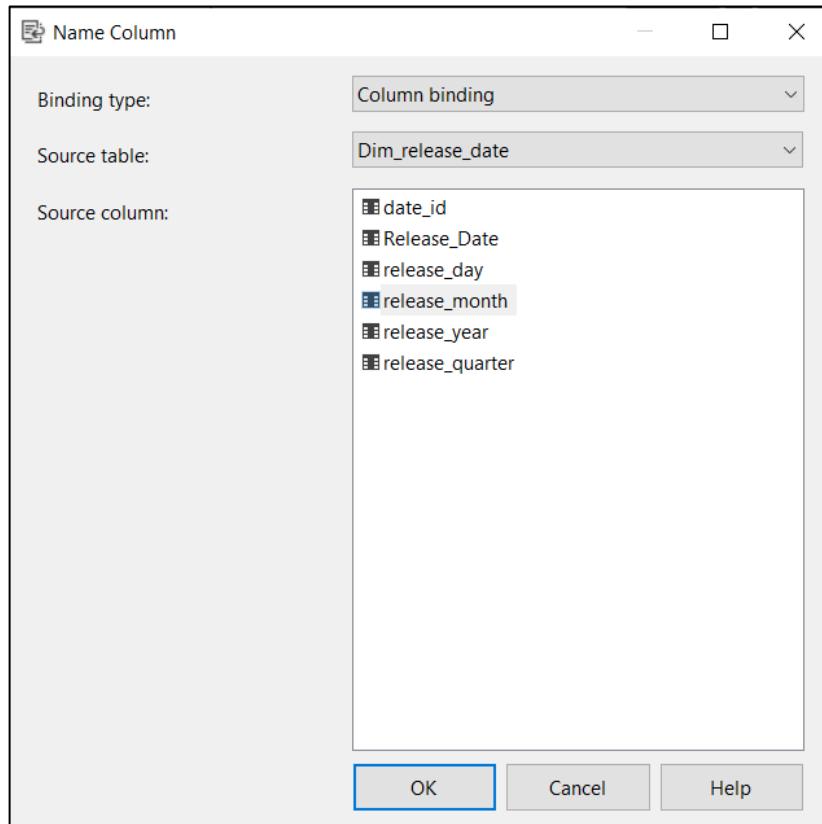
Bước 5: Tại cửa sổ Properties của thuộc tính release day, ta chọn Name Column và chọn tên thuộc tính là release_day.



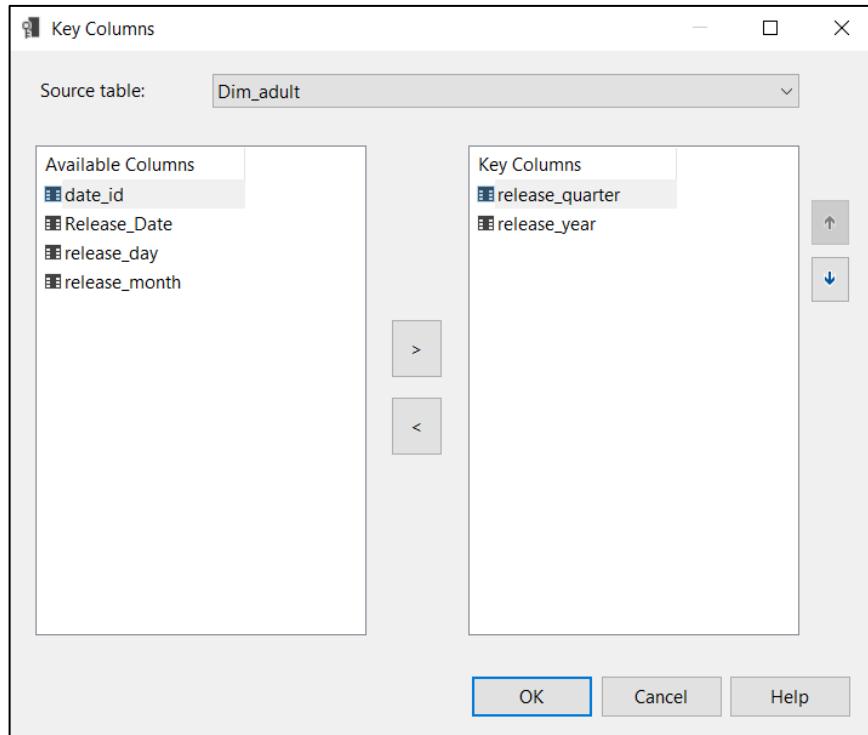
Bước 6: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính release_month. Vì thuộc tính release_month là thuộc tính cấp nhỏ hơn release_quarter, release_year nên sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.



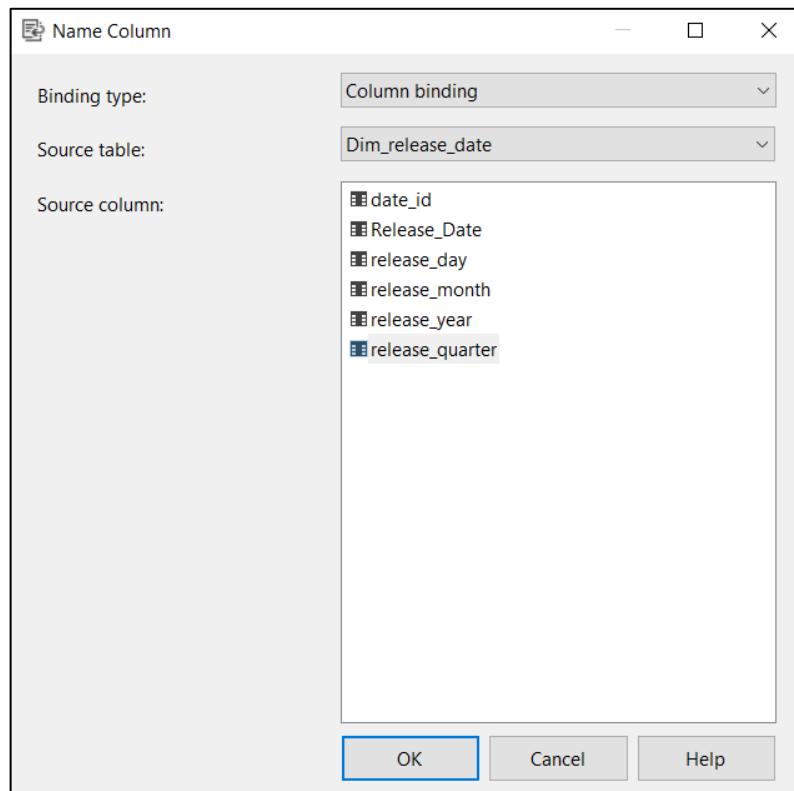
Bước 7: Tại cửa sổ Properties của thuộc tính release_month, ta chọn Name Column và chọn tên thuộc tính là release_month.



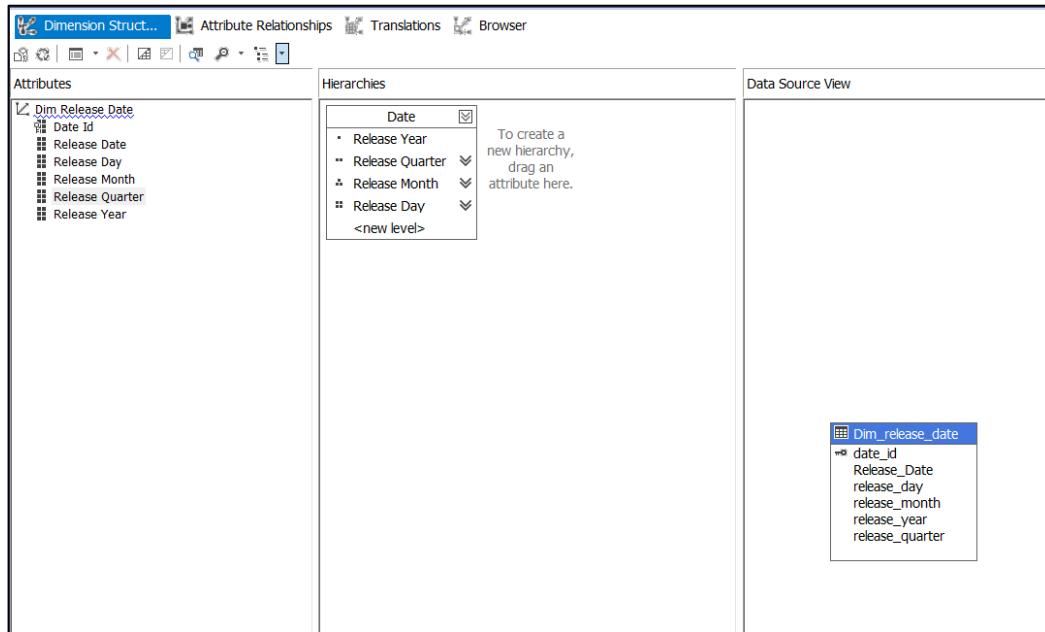
Bước 8: Chính khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính release_quarter. Vì thuộc tính release_quarter là thuộc tính cấp nhỏ hơn release_year nên sẽ lấy khóa cột gồm chính nó và thuộc tính cấp cao hơn.



Bước 9: Tại cửa sổ Properties của thuộc tính Start Time Quarter, ta chọn Name Column và chọn tên thuộc tính là release_quarter.

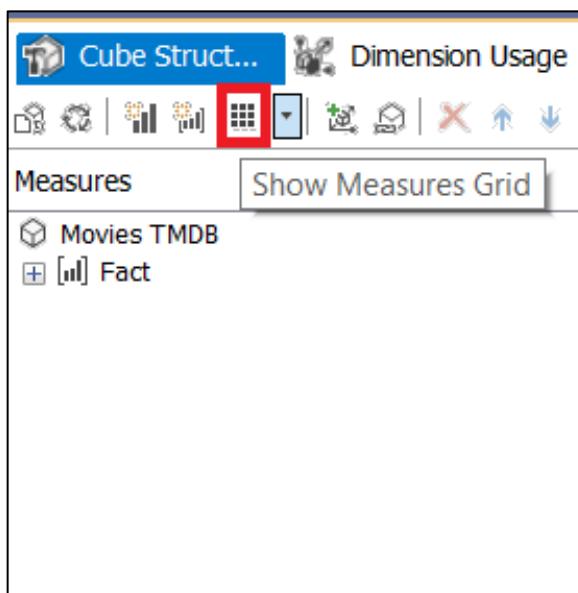


Ta thu được kết quả như hình dưới



3.6.2. Đổi tên một số thuộc tính

Bước 1: Tại khôi vừa tạo, chọn Show Measures Grid để hiện thị chi tiết các độ đo.



Chi tiết các độ đo sẽ hiển thị dưới dạng bảng, dễ dàng để tương tác.

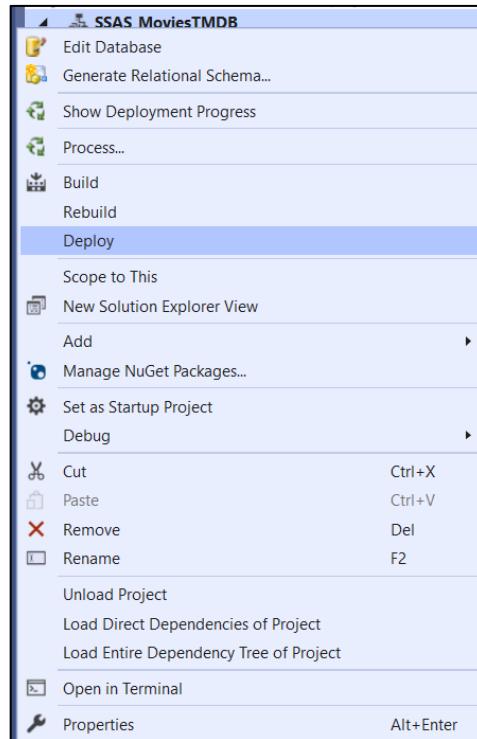
Measures				
	Name	Measure Group	Data Type	Aggregation
📊	Vote Average	Fact	Single	Sum
📊	Vote Count		Integer	Sum
📊	Revenue		BigInt	Sum
📊	Budget		Integer	Sum
📊	Popularity		Single	Sum
📊	Fact Count		Integer	Count

Bước 2: Ta đổi tên Fact Count (Số lượng bộ phim) đổi thành Number of Movies để rõ ràng và giữ nguyên Aggregation:

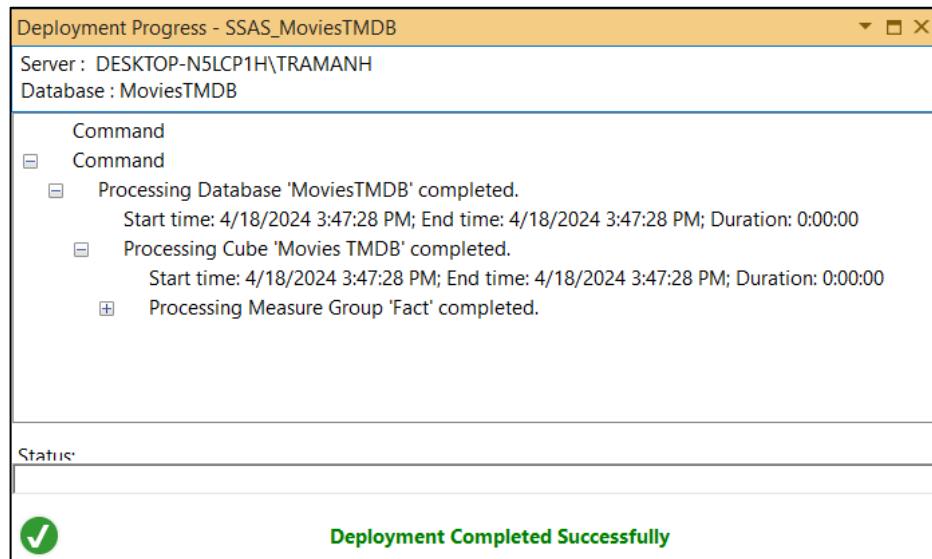
Measures				
	Name	Measure Group	Data Type	Aggregation
📊	Vote Average		Single	Sum
📊	Vote Count		Integer	Sum
📊	Revenue		BigInt	Sum
📊	Budget		Integer	Sum
📊	Popularity		Single	Sum
📊	Number of Movies	Fact	Integer	Count

3.7. Chạy dự án SSAS

Sau khi quá trình phân cấp cho các bảng chiều hoàn tất, ta thực hiện deploy project để đảm bảo không có lỗi xảy ra sau quá trình phân cấp. Nhấn chuột phải vào tên project (SSAS_MoviesTMDB) và nhấn Deploy như hình sau:



Khi deploy thành công, hệ thống sẽ hiển thị như hình sau và chúng ta bắt đầu thực hiện các câu truy vấn.



3.8. Thực hiện 15 câu truy vấn (MDX, Manual và Pivot Excel)

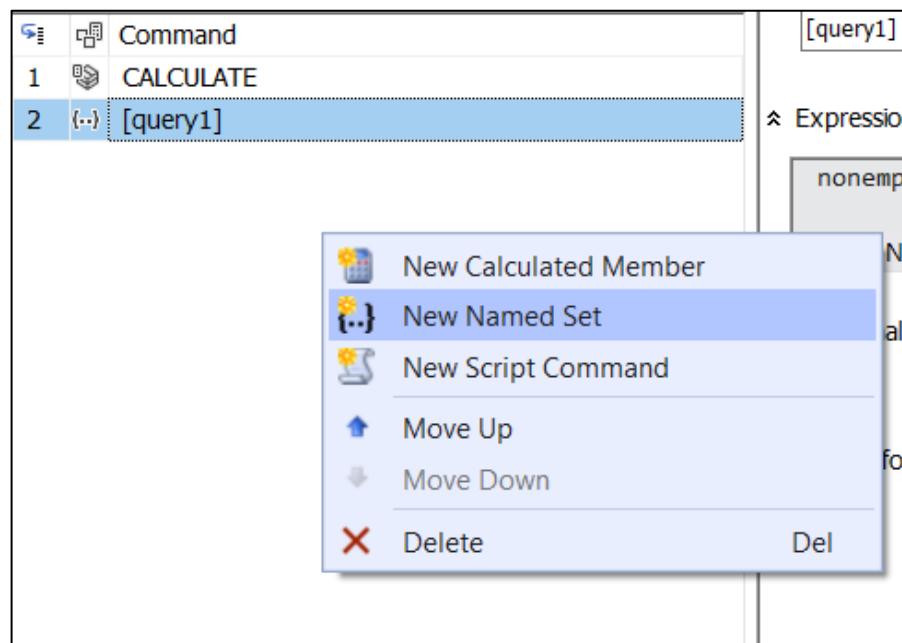
Câu 1: Tổng doanh thu và ngân sách của các phim theo các mốc thời lượng chiếu.

a. MDX:

```
select {[Measures].[Revenue], [Measures].[Budget]} on columns,
nonempty([Dim Runtime].[Runtime Range].children) on rows
from [Movies TMDB];
```

	Revenue	Budget
1h - 2h	412760184219	848111866
2h - 3h	292696425609	1554524119
over 3h	16039871686	2023501977
under 1h	22324079141	900872639

b. Manual



The screenshot shows the properties for the query '[query1]'. The 'Name' field is '[query1]' (marked with red box 1). The 'Expression' field contains 'nonempty([Dim Runtime].[Runtime Range].children)' (marked with red box 2). The 'Additional Properties' section shows 'Type: Dynamic' and 'Display folder:'.

Runtime Range	Revenue	Budget
1h - 2h	412760184219	848111866
2h - 3h	292696425609	1554524119
over 3h	16039871686	2023501977
under 1h	22324079141	900872639

c. Pivot Excel

Row Labels	Budget	Revenue
1h - 2h	848111866	4.1276E+11
2h - 3h	1554524119	2.92696E+11
over 3h	2023501977	16039871686
under 1h	900872639	22324079141

Câu 2: Liệt kê 5 bộ phim có doanh thu cao nhất, sắp xếp theo thứ tự giảm dần.

a. MDX

```
select {[Measures].[Revenue]} on columns,
order(
    topcount([Dim Movies].[Title].children, 5, [Measures].[Revenue]
    [Measures].[Revenue],
    DESC)on rows
from [Movies TMDB];
```

	Revenue
Bee Movie	3217594577
Adventures in Bora Bora	3000000000
Avatar	2923706026
Avengers: Endgame	2800000000
The Lion King	2426530962

b. Manual

Name: [query2]

Expression

```
order(
    topcount([Dim Movies].[Title].children, 5, [Measures].[Revenue]),
    [Measures].[Revenue])
✓ No issues found
```

Ln: 4 Ch: 7 Col: 10 TABS CRLF

Additional Properties

Type: Dynamic

Display folder:

Dimension	Hierarchy	Operator	Filter Expression	Para...
Dim Movies	Title	In	query2	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>
<hr/>				
Title	Revenue			
Adventures in Bora Bora	3000000000			
Avatar	2923706026			
Avengers: Endgame	2800000000			
Bee Movie	3217594577			
The Lion King	2426530962			

c. Pivot Excel

Row Labels	Revenue
Adventures in Bora Bora	3000000000
Avatar	2923706026
Avengers: Endgame	2800000000
Bee Movie	3217594577
The Lion King	2426530962

Câu 3: Thống kê 10 tên phim có ngân sách cao nhất được công chiếu vào năm 2023 sắp xếp giảm dần theo độ thịnh hành.

a. MDX

```

select {[Measures].[Budget],[Measures].[Popularity]} on columns,
order(
    topcount([Dim Movies].[Title].children, 10, [Measures].[Budget]),
    [Measures].[Popularity],
    DESC
) on rows
from [Movies TMDB]
where [Dim Release Date].[Release Year].&[2023];

```

	Budget	Popularity
Fast X	340000000	1175.267
Indiana Jones and the Dial of Destiny	294700000	579.252
The Flash	300000000	480.393
Mission: Impossible - Dead Reckoning Part One	291000000	401.083
The Little Mermaid	250000000	357.686
Guardians of the Galaxy Vol. 3	250000000	318.516
The Marvels	274800000	60.454
Lost in the Stars	417549000	23.732
Enea	888000000	4.319
Adventures in Bora Bora	800000000	0

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Release Date	# Release Year	Equal	{ 2023 }	<input checked="" type="checkbox"/>
Dim Movies	# Title	In	query3	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				

Title	Budget	Popularity
Adventures in Bora Bora	800000000	0
Enea	888000000	4.319
Fast X	340000000	1175.267
Guardians of the Galaxy Vol. 3	250000000	318.516
Indiana Jones and the Dial of Destiny	294700000	579.252
Lost in the Stars	417549000	23.732
Mission: Impossible - Dead Reckoning Part One	291000000	401.083
The Flash	300000000	480.393
The Little Mermaid	250000000	357.686
The Marvels	274800000	60.454

c. Pivot Excel

Row Labels	Budget	Popularity
2023		
Adventures in Bora Bora	800000000	0
Enea	888000000	4.318999767
Fast X	340000000	1175.266968
Guardians of the Galaxy Vol. 3	250000000	318.5159912
Indiana Jones and the Dial of Destiny	294700000	579.2520142
Lost in the Stars	417549000	23.73200035
Mission: Impossible - Dead Reckoning Part One	291000000	401.0830078
The Flash	300000000	480.3930054
The Little Mermaid	250000000	357.6860046
The Marvels	274800000	60.45399857

Câu 4: Số lượng phim được công bố vào các quý của năm 2023.

a. MDX

```
select [Measures].[Number of Movies] on columns,
[Dim Release Date].[Release Quarter].members on rows
from [Movies TMDB]
where [Dim Release Date].[Release Year].&[2023];
```

	Number of Movies
All	38107
1	9314
2	10506
3	9845
4	8442

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Param...										
Dim Release Date	Release Year	Equal	{ 2023 }	<input type="checkbox"/> <input type="checkbox"/>										
Dim Release Date	Release Quarter	In	query4	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>										
<Select dimension>														
<table border="1"> <thead> <tr> <th>Release Quarter</th> <th>Number of Movies</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>9314</td> </tr> <tr> <td>2</td> <td>10506</td> </tr> <tr> <td>3</td> <td>9845</td> </tr> <tr> <td>4</td> <td>8442</td> </tr> </tbody> </table>					Release Quarter	Number of Movies	1	9314	2	10506	3	9845	4	8442
Release Quarter	Number of Movies													
1	9314													
2	10506													
3	9845													
4	8442													

c. Pivot Excel

Row Labels		Number of Movies
2023		
1		9314
2		10506
3		9845
4		8442

Câu 5: Thống kê tên phim, số lượng bình chọn và điểm bình chọn trung bình của các bộ phim có ngân sách trên 1.5 tỷ đô la.

a. MDX

```
with member [Measures].[Mean Average Vote]
as (sum([Measures].[Vote Average]) / [Measures].[Number of movies])
select {[Measures].[Mean Average Vote], [Measures].[Vote Count]} on columns,
FILTER([Dim Movies].[Title].children, [Measures].[Revenue] > 1500000000) on rows
from [Movies TMDB];
```

	Mean Average Vote	Vote Count
Adventures in Bora Bora	0	0
Aladdin	4.71449995040894	20256
Avatar	6.44250011444092	29856
Avatar: The Way of Water	7.65399980545044	9830
Avengers: Endgame	8.26299953460693	23857
Avengers: Infinity War	8.25500011444092	27713
Beauty and the Beast	4.55081487584997	26819
Bee Movie	2.00533326466878	4500
Furious 7	7.23799991607666	9988
Joker	4.23392310509315	23507
Jurassic World	6.68200016021729	19414
Spider-Man: No Way Home	7.98999977111816	18299
Star Wars: The Force Awakens	7.29300022125244	18352
The Avengers	7.3630002339681	29805
The Lion King	7.67799949645996	26316
Titanic	5.7674446105957	23844

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Paramet...
Dim Movies	Title	In	query5	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				
Title	Mean Average Vote	Vote Count		
Adventures in Bora Bora	0	0		
Aladdin	4.71449995040894	20256		
Avatar	6.44250011444092	29856		
Avatar: The Way of Water	7.65399980545044	9830		
Avengers: Endgame	8.26299953460693	23857		
Avengers: Infinity War	8.25500011444092	27713		
Beauty and the Beast	4.55081487584997	26819		
Bee Movie	2.0053326466878	4500		
Furious 7	7.23799991607666	9988		
Joker	4.23392310509315	23507		
Jurassic World	6.68200016021729	19414		
Spider-Man: No Way Home	7.98999977111816	18299		
Star Wars: The Force Awakens	7.29300022125244	18352		
The Avengers	7.3630002339681	29805		
The Lion King	7.67799949645996	26316		
Titanic	5.7674446105957	23844		

c. Pivot Excel

Row Labels	Mean Average Vote	Vote Count
Adventures in Bora Bora	0	0
Aladdin	4.71449995	20256
Avatar	6.442500114	29856
Avatar: The Way of Water	7.653999805	9830
Avengers: Endgame	8.262999535	23857
Avengers: Infinity War	8.255000114	27713
Beauty and the Beast	4.550814876	26819
Bee Movie	2.00533265	4500
Furious 7	7.237999916	9988
Joker	4.233923105	23507
Jurassic World	6.68200016	19414
Spider-Man: No Way Home	7.989999771	18299
Star Wars: The Force Awakens	7.293000221	18352
The Avengers	7.363000234	29805
The Lion King	7.677999496	26316
Titanic	5.767444611	23844

Câu 6: Top 3 bộ phim có ngân sách cao nhất theo 3 ngôn ngữ Anh, Việt và Trung.

a. MDX

```

select {[Measures].[Budget]} on columns,
generate(
{
    [Dim Language].[Original Language].&[vi],
    [Dim Language].[Original Language].&[en],
    [Dim Language].[Original Language].&[cn]
},
    topcount(
        [Dim Language].[Original Language].currentmember * [Dim Movies].[Title].children,
        3,
        [Measures].[Budget]
    )
) on rows
from [Movies TMDB];

```

		Budget
vi	Em and Trinh	2156845
vi	Camellia Sisters 5 - Living Like Royalty	2000000
vi	Long Ruoi	2000000
en	Adventures in Bora Bora	8000000000
en	Playmobil The Movie 2: Maximus' Revenge	8000000000
en	Avatar 5	6000000000
cn	Chasing the Dragon	200005000
cn	Ip Man 4: The Finale	52000000
cn	Chinese Zodiac	48000000

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Paramet...
Dim Language	Original Language	Custom	generate({[Dim Language].[Original Language].&[vi],...})	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>
Original Language	Title	Budget		
cn	Chasing the Dragon	200005000		
cn	Chinese Zodiac	48000000		
cn	Ip Man 4: The Finale	52000000		
en	Adventures in Bora Bora	8000000000		
en	Avatar 5	6000000000		
en	Playmobil The Movie 2: Maximus' Revenge	8000000000		
vi	Camellia Sisters 5 - Living Like Royalty	2000000		
vi	Em and Trinh	2156845		
vi	Long Ruoi	2000000		

c. Pivot Excel

Row Labels	Budget
cn	
Chasing the Dragon	200005000
Chinese Zodiac	480000000
Ip Man 4: The Finale	52000000
en	
Adventures in Bora Bora	800000000
Avatar 5	600000000
Playmobil The Movie 2: Maximus' Revenge	800000000
vi	
Camellia Sisters 5 - Living Like Royalty	2000000
Em and Trinh	2156845
Long Ruoi	2000000

Câu 7: Thống kê 2 phim phổ biến nhất có tiêu đề bắt đầu bằng ký tự “H” ở từng năm từ năm 2020 đến 2024.

a. MDX

```
select {[Measures].[Popularity]} on columns,
generate(
    {[Dim Release Date].[Release Year].&[2020]:[Dim Release Date].[Release Year].&[2024]},
    topcount(
        filter(
            [Dim Release Date].[Release Year].currentmember * [Dim Movies].[Title].children,
            left([Dim Movies].[Title].currentmember.name, 1) = 'H'
        ),
        2,
        [Measures].[Popularity]
    )
) on rows
from [Movies TMDB];
```

		Popularity
2020	Host	31.421
2020	Her Uniform Is Soaking We...	30.316
2021	Halloween Kills	37.851
2021	House of Gucci	26.057
2022	H4Z4RD	198.952
2022	Hotel Transylvania: Transfo...	102.4
2023	Heart of Stone	337.527
2023	Heroic	170.27
2024	Haikyu!! THE MOVIE -Deci...	7.206
2024	How to Date Billy Walsh	5.501

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Release Date	Release Year	Range (Inclusive)	2020 : 2024	<input type="checkbox"/> <input checked="" type="checkbox"/>
Dim Movies	Title	Custom	generate({{[Dim Release Date].[Release Year].&...}}	<input type="checkbox"/> <input checked="" type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input checked="" type="checkbox"/>
Release Year	Title	Popularity		
2020	Her Uniform Is Soaking Wet, So She Had To Get Out Of The Rain My I	30.31599998		
2020	Host	31.42099953		
2021	Halloween Kills	37.851		
2021	House of Gucci	26.057		
2022	H4Z4RD	198.952		
2022	Hotel Transylvania: Transformania	102.4		
2023	Heart of Stone	337.527		
2023	Heroic	170.27		
2024	Haikyu!! THE MOVIE -Decisive Battle at the Garbage Dump-	7.206		
2024	How to Date Billy Walsh	5.501		

c. Pivot Excel

Row Labels	Popularity
2020	
Her Uniform Is Soaking Wet, So She Had To Get Out Of The Rain My I	30.31599998
Host	31.42099953
2021	
Halloween Kills	37.85100174
House of Gucci	26.05699921
2022	
H4Z4RD	198.9519958
Hotel Transylvania: Transformania	102.4000015
2023	
Heart of Stone	337.5270081
Heroic	170.2700043
2024	
Haikyu!! THE MOVIE -Decisive Battle at the Garbage Dump-	7.205999851
How to Date Billy Walsh	5.500999928
Grand Total	947.5009766

Câu 8: Số lượng phim, số lượng bình chọn và điểm bình chọn trung bình theo các mốc thời lượng có ngôn ngữ gốc là Anh, Trung, Việt.

a. MDX

```

with member [Measures].[Mean Average Vote]
as (sum([Measures].[Vote Average]) / [Measures].[Number of movies])
select {[Measures].[Number of movies], [Measures].[Vote Count], [Measures].[Mean Average Vote]} on columns,
nonempty(crossjoin([Dim Runtime].[Runtime Range].children,
{
    [Dim Language].[Original Language].&[vi],
    [Dim Language].[Original Language].&[en],
    [Dim Language].[Original Language].&[cn]
})
)on rows
from [Movies TMDB];

```

		Number of Movies	Vote Count	Mean Average Vote
1h - 2h	vi	402	1258	2.49031104851718
1h - 2h	en	183946	11764495	3.10154950365868
1h - 2h	cn	3517	72101	3.95909794169392
2h - 3h	vi	34	167	3.8712355669807
2h - 3h	en	37911	5783453	2.63556153721875
2h - 3h	cn	118	6839	4.54694017313294
over 3h	vi	10	3	1.31999998092651
over 3h	en	13506	261805	2.61103456334222
over 3h	cn	13	15	1.62823075514573
under 1h	vi	625	171	0.9072
under 1h	en	305199	381392	1.280151229198
under 1h	cn	1337	380	0.879053793474196

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Language	# Original Language	Custom	nonempty(crossjoin([Dim Runtime].[Runtime ...	<input type="checkbox"/> <input checked="" type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input checked="" type="checkbox"/>
Runtime Range	Original Language	Number of Movies	Vote Count	Mean Average Vote
1h - 2h	cn	3517	72101	3.95909794169392
1h - 2h	en	183946	11764495	3.10154950365868
1h - 2h	vi	402	1258	2.49031104851718
2h - 3h	cn	118	6839	4.54694017313294
2h - 3h	en	37911	5783453	2.6355617432935
2h - 3h	vi	34	167	3.8712355669807
over 3h	cn	13	15	1.62823075514573
over 3h	en	13506	261805	2.61103456334222
over 3h	vi	10	3	1.31999998092651
under 1h	cn	1337	380	0.879053793474...
under 1h	en	305199	381392	1.28015133159021
under 1h	vi	625	171	0.9072

c. Pivot Excel

Row Labels	Number of Movies	Vote Count	Mean Average Vote
1h - 2h			
cn	3517	72101	3.959097942
en	183946	11764495	3.101549504
vi	402	1258	2.490311049
2h - 3h			
cn	118	6839	4.546940173
en	37911	5783453	2.635561537
vi	34	167	3.871235567
over 3h			
cn	13	15	1.628230755
en	13506	261805	2.611034563
vi	10	3	1.319999981
under 1h			
cn	1337	380	0.879053793
en	305199	381392	1.280151229
vi	625	171	0.9072

Câu 9: Tổng doanh thu của các phim được công bố từ năm 2020 đến năm 2024, thực hiện drilldown từ năm sang quý.

a. MDX

```
select [Measures].[Revenue] on columns,
drilldownlevel(
    {[Dim Release Date].[Date].[Release Year].&[2020]:[Dim Release Date].[Date].[Release Year].&[2024]})
) on rows
from [Movies TMDB];
```

	Revenue
2020	7218698262
1	2629060808
2	545337966
3	1627419585
4	2416879903
2021	17049181126
1	3036950811
2	2731357426
3	6127234754
4	5153638135
2022	22745051929
1	4793170586
2	7051039912
3	4613616301
4	6287225130
2023	22402741064
1	4703671525
2	8437497742
3	9174147619
4	87424178
2024	100202878
1	100080235
2	122617
3	6
4	20

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Release Date	Release Year	Range (Inclusive)	2020 : 2024	<input type="checkbox"/> <input checked="" type="checkbox"/>
Dim Release Date	Release Quarter	Custom	dropdownlevel([Dim Release Date].[Date].[Rele...]	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input checked="" type="checkbox"/>
				<input type="checkbox"/> <input checked="" type="checkbox"/>
Release Year	Release Quarter	Revenue		
2020	1	2629060808		
2020	2	545337966		
2020	3	1627419585		
2020	4	2416879903		
2021	1	3036950811		
2021	2	2731357426		
2021	3	6127234754		
2021	4	5153638135		
2022	1	4793170586		
2022	2	7051039912		
2022	3	4613616301		
2022	4	6287225130		
2023	1	4703671525		
2023	2	8437497742		
2023	3	9174147619		
2023	4	87424178		
2024	1	100080235		
2024	2	122617		
2024	3	6		
2024	4	20		

c. Pivot Excel

Row Labels	Revenue
2020	
1	2629060808
2	545337966
3	1627419585
4	2416879903
2021	
1	3036950811
2	2731357426
3	6127234754
4	5153638135
2022	
1	4793170586
2	7051039912
3	4613616301
4	6287225130
2023	
1	4703671525
2	8437497742
3	9174147619
4	87424178
2024	
1	100080235
2	122617
3	6
4	20

Câu 10: Số lượt bình chọn và điểm thịnh hành của các bộ phim có ngôn ngữ gốc là tiếng Việt, tiếng Anh, có thời lượng dưới 1h công chiếu từ năm 2020 đến 2023.

a. MDX

```
select {[Measures].[Popularity], [Measures].[Vote Count]} on columns,
([Dim Release Date].[Release Year].&[2020]:[Dim Release Date].[Release Year].&[2023],
{
    [Dim Language].[Original Language].&[vi],
    [Dim Language].[Original Language].&[en]
})
on rows
from [Movies TMDB]
where [Dim Runtime].[Runtime Range].&[under 1h];
```

		Popularity	Vote Count
2020	vi	20.767	20
2020	en	8919.796	13213
2021	vi	17.768	9
2021	en	9314.811	11900
2022	vi	15.178	7
2022	en	10020	11218
2023	vi	35.424	2
2023	en	15926.51	3108

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Release Date	Release Year	Range (Inclusive)	2020 : 2023	<input type="checkbox"/> <input type="checkbox"/>
Dim Runtime	Runtime Range	Equal	{ under 1h }	<input type="checkbox"/> <input type="checkbox"/>
Dim Language	Original Language	Custom	[{[Dim Language].[Original Language].&[vi],[Dim Language].[Original Language].&[en]}]	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>

Release Year	Original Language	Popularity	Vote Count
2020	en	8919.796	13213
2020	vi	20.767	20
2021	en	9314.811	11900
2021	vi	17.768	9
2022	en	10020	11218
2022	vi	15.178	7
2023	en	15926.51	3108
2023	vi	35.424	2

c. Pivot Excel

Row Labels	Popularity	Vote Count
2020		
en	under 1h 8919.795898	13213
vi	under 1h 20.7670002	20
2021		
en	under 1h 9314.810547	11900
vi	under 1h 17.76799965	9
2022		
en	under 1h 10019.99707	11218
vi	under 1h 15.1779995	7
2023		
en	under 1h 15926.51465	3108
vi	under 1h 35.42399597	2

Câu 11: Thống kê số lượng và phần trăm các bộ phim có chi phí sản xuất lớn hơn doanh thu (gấp tình trạng lỗ) trong 5 năm gần nhất.

a. MDX

```

with
member [Measures].[Losses Movies Count] as
count(
    filter(
        [Dim Movies].[Movie Id].children,
        [Measures].[Revenue] < [Measures].[Budget]
    )
)
member [Measures].[Losses Movies Per] as
[Measures].[Losses Movies Count] / [Measures].[Number of Movies]
select {[Measures].[Measures].[Losses Movies Count], [Measures].[Number of Movies], [Measures].[Losses Movies Per]} on columns,
{[Dim Release Date].[Release Year].&[2020]:[Dim Release Date].[Release Year].&[2024]} on rows
from [Movies TMDB];

```

	Losses Movies Count	Number of Movies	Losses Movies Per
2020	2182	39062	5.58599150069121E-02
2021	2562	41787	6.13109340225429E-02
2022	3056	41849	0.073024445028555
2023	3208	38107	8.41840081874721E-02
2024	479	4252	0.112652869238006

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	Paramet...
Dim Release Date	Release Year	Range (Inclusive)	2020 : 2024	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				
Release Year	Losses Movies Count	Number of Movies	Losses Movies Per	
2020	2182	39062	0.05585991500...	
2021	2562	41787	0.06131093402...	
2022	3056	41849	0.07302444502...	
2023	3208	38107	0.08418400818...	
2024	479	4252	0.11265286923...	

c. Pivot Excel

Row Labels	Losses Movies Count	Number of Movies	Losses Movies Per
2020	2182	39062	0.055859915
2021	2562	41787	0.061310934
2022	3056	41849	0.073024445
2023	3208	38107	0.084184008
2024	479	4252	0.112652869

Câu 12: Thống kê số lượng phim và tổng số lượt bình chọn theo từng quý từ năm 2021 đến 2024 của hai ngôn ngữ là tiếng Anh và tiếng Việt.

a. MDX

```
select {[Measures].[Number of Movies], [Measures].[Vote Count]} on columns,
nonempty({[Dim Release Date].[Release Year].&[2021]: [Dim Release Date].[Release Year].&[2024]}) *
[Dim Release Date].[Release Quarter].children *
{
    [Dim Language].[Original Language].&[vi],
    [Dim Language].[Original Language].&[en]
} on rows
from [Movies TMDB];
```

			Number of Movies	Vote Count
2021	1	vi	21	20
2021	1	en	4851	89933
2021	2	vi	10	18
2021	2	en	4961	93461
2021	3	vi	6	0
2021	3	en	5019	145950
2021	4	vi	9	2
2021	4	en	6452	143604
2022	1	vi	12	63
2022	1	en	4314	89280
2022	2	vi	17	15
2022	2	en	5133	80154
2022	3	vi	13	6
2022	3	en	4979	92351
2022	4	vi	15	66
2022	4	en	6314	97717
2023	1	vi	12	4
2023	1	en	4886	45878
2023	2	vi	11	3
2023	2	en	5450	62910
2023	3	vi	10	3
2023	3	en	4955	29733
2023	4	vi	11	0
2023	4	en	4224	287
2024	1	vi	9	0
2024	1	en	1868	91
2024	2	en	198	0
2024	3	en	75	0
2024	4	en	86	0

b. Manual

Release Year	Release Quarter	Original Language	Number of Movies	Vote Count
2021	1	en	4851	89933
2021	1	vi	21	20
2021	2	en	4961	93461
2021	2	vi	10	18
2021	3	en	5019	145950
2021	3	vi	6	0
2021	4	en	6452	143604
2021	4	vi	9	2
2022	1	en	4314	89280
2022	1	vi	12	63
2022	2	en	5133	80154
2022	2	vi	17	15
2022	3	en	4979	92351
2022	3	vi	13	6
2022	4	en	6314	97717
2022	4	vi	15	66
2023	1	en	4886	45878
2023	1	vi	12	4
2023	2	en	5450	62910
2023	2	vi	11	3
2023	3	en	4955	29733
2023	3	vi	10	3
2023	4	en	4224	287
2023	4	vi	11	0
2024	1	en	1868	91
2024	1	vi	9	0
2024	2	en	198	0

2024	1	en	1868	91
2024	1	vi	9	0
2024	2	en	198	0
2024	3	en	75	0
2024	4	en	86	0

c. Pivot Excel

Row Labels	Number of Movies	Vote Count		
2021				
1				
en	4851	89933		
vi	21	20		
2				
en	4961	93461		
vi	10	18		
3				
en	5019	145950		
vi	6	0		
4				
en	6452	143604		
vi	9	2		
2022				
1				
en	4314	89280		
vi	12	63		
2				
en	5133	80154		
vi	17	15		
3				
en	4979	92351		
vi	13	6		
4				
en	6314	97717		
vi	15	66		
2023				
1				
en	4886	45878		
vi	12	4		
2				
en	5450	62910		
vi	11	3		
3				
en	4955	29733		
vi	10	3		
4				
en	4224	287		
vi	11	0		
2024				
1				
en	1868	91		
vi	9	0		
2				
en	198	0		
3				
en	75	0		
4				
en	86	0		

Câu 13: Thống kê số lượng phim và tổng doanh thu của theo từng năm từ 2021 đến 2024 và độ tuổi người xem cho phép(trên 18 tuổi hay không).

a. MDX: *False là mọi độ tuổi và True là trên 18 tuổi

```
select {[Measures].[Number of Movies], [Measures].[Revenue]} on columns,
nonempty(
{[Dim Release Date].[Release Year].&[2021]: 
[Dim Release Date].[Release Year].&[2024]} *
[Dim Adult].[Adult].members) on rows
from [Movies TMDB];
```

		Number of Movies	Revenue
2021	All	41787	17049181126
2021	False	37010	17048873594
2021	True	4777	307532
2022	All	41849	22745051929
2022	False	37307	22178344964
2022	True	4542	566706965
2023	All	38107	22402741064
2023	False	34802	22296207415
2023	True	3305	106533649
2024	All	4252	100202878
2024	False	3940	100197777
2024	True	312	5101

b. Manual

Dimension	Hierarchy	Operator	Filter Expression
Dim Release Date	Release Year	Range (Inclusive)	2021 : 2024
<Select dimension>			
<hr/>			
Release Year	Adult	Number of Movies	Revenue
2021	False	37010	17048873594
2021	True	4777	307532
2022	False	37307	22178344964
2022	True	4542	566706965
2023	False	34802	22296207415
2023	True	3305	106533649
2024	False	3940	100197777
2024	True	312	5101

c. Pivot Excel

Row Labels	Number of Movies	Revenue
2021	37010	17048873594
2021	4777	307532
2022	37307	22178344964
2022	4542	566706965
2023	34802	22296207415
2023	3305	106533649
2024	3940	100197777
2024	312	5101

Câu 14: top 5 các phim có lượt bình chọn thấp nhất được công chiếu trong khoảng từ năm 2021 đến 2024, có doanh thu và ngân sách lớn hơn 0.

a. MDX

```

select {[Measures].[Budget], [Measures].[Revenue], [Measures].[Vote Count]} on columns,
bottomcount(
    filter(
        nonempty([Dim movies].[Title].members
        * {[Dim Release Date].[Release year].&[2021]:[Dim Release Date].[Release year].&[2024]}),
        [Measures].[Budget] >= 0 and [Measures].[Revenue] > 0),
    5,
    [Measures].[Vote Count]
)
on rows
from [Movies TMDB];

```

		Budget	Revenue	Vote Count
Zombie Attack	2021	52	3300	0
Yo la matÃ©	2023	160	160	0
Wrong Bill	2022	30	350	0
Wrapped Up	2023	10	999999	0
World Cinema	2023	10	12	0

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	
Dim Release Date	Release Year	Range (Inclusive)	2021 : 2024	
Dim Movies	Title	Custom	bottomcount(filter(nonempty([Dim movies].[Title].me...))	
<Select dimension>				
Title	Release Year	Budget	Revenue	Vote Count
World Cinema	2023	10	12	0
Wrapped Up	2023	10	999999	0
Wrong Bill	2022	30	350	0
Yo la matÃ©	2023	160	160	0
Zombie Attack	2021	52	3300	0

c. Pivot Excel

Row Labels	Budget	Revenue	Vote Count
World Cinema			
2023	10	12	0
Wrapped Up			
2023	10	999999	0
Wrong Bill			
2022	30	350	0
Yo la matÃ©			
2023	160	160	0
Zombie Attack			
2021	52	3300	0

Câu 15: Thống kê tổng số lượt bình chọn, điểm thịnh hành, tổng ngân sách của các năm có số lượng phim công chiếu lớn 15000, sắp xếp theo thứ tự giảm dần theo tổng ngân sách.

a. MDX

```
select {[Measures].[Popularity], [Measures].[Vote Count],
         [Measures].[Budget], [Measures].[Number of Movies]} on columns,
order(
    [Dim Release Date].[Release Year].members,
    [Measures].[Budget],
    DESC
)
having [Measures].[Number of Movies] >= 15000
on rows
from [Movies TMDB];
```

	Popularity	Vote Count	Budget	Number of Movies
All	1358894	21445245	1032043305	1004842
1899	58129.93	10332	1429029617	117320
2022	66722.59	452277	1088941398	41849
2021	55228.64	572740	-408467899	41787
2019	52678.9	878869	1265496929	39924
2020	49619.04	491380	277533427	39062
2023	122321.3	175566	-1874454887	38107
2018	47009.01	950621	1263557790	35982
2017	44615.47	1072942	2116415969	33733
2016	39386	1057896	-598974539	30301
2015	37368.04	935245	1425765795	28791
2014	37054.11	1039020	1930782141	27399
2013	34243.43	965598	-2068006983	25400
2012	31916.92	802840	1720144768	23020
2011	30136.28	770049	-2010672541	21112
2010	27552	738231	-1877434546	19475
2009	27383.17	699557	2137433961	18786
2008	25457.5	638796	1745590727	17833
2007	24461.23	605553	795232563	16966
2006	23092	562297	454655208	16374

b. Manual

Dimension	Hierarchy	Operator	Filter Expression	
Dim Release Date	Release Year	Custom	[query 15]	
<Select dimension>				
Release Year	Popularity	Vote Count	Budget	Number of Movies
1899	58129.94	10332	1429029617	117320
2006	23092	562297	454655208	16374
2007	24461.23	605553	795232563	16966
2008	25457.5	638796	1745590727	17833
2009	27383.17	699557	2137433961	18786
2010	27552	738231	-1877434546	19475
2011	30136.28	770049	-2010672541	21112
2012	31916.92	802840	1720144768	23020
2013	34243.42	965598	-2068006983	25400
2014	37054.11	1039020	1930782141	27399
2015	37368.04	935245	1425765795	28791
2016	39386	1057896	-598974539	30301
2017	44615.48	1072942	2116415969	33733
2018	47009.02	950621	1263557790	35982
2019	52678.9	878869	1265496929	39924
2020	49619.04	491380	277533427	39062
2021	55228.64	572740	-408467899	41787
2022	66722.59	452277	1088941398	41849
2023	122321.3	175566	-1874454887	38107

c. Pivot Excel

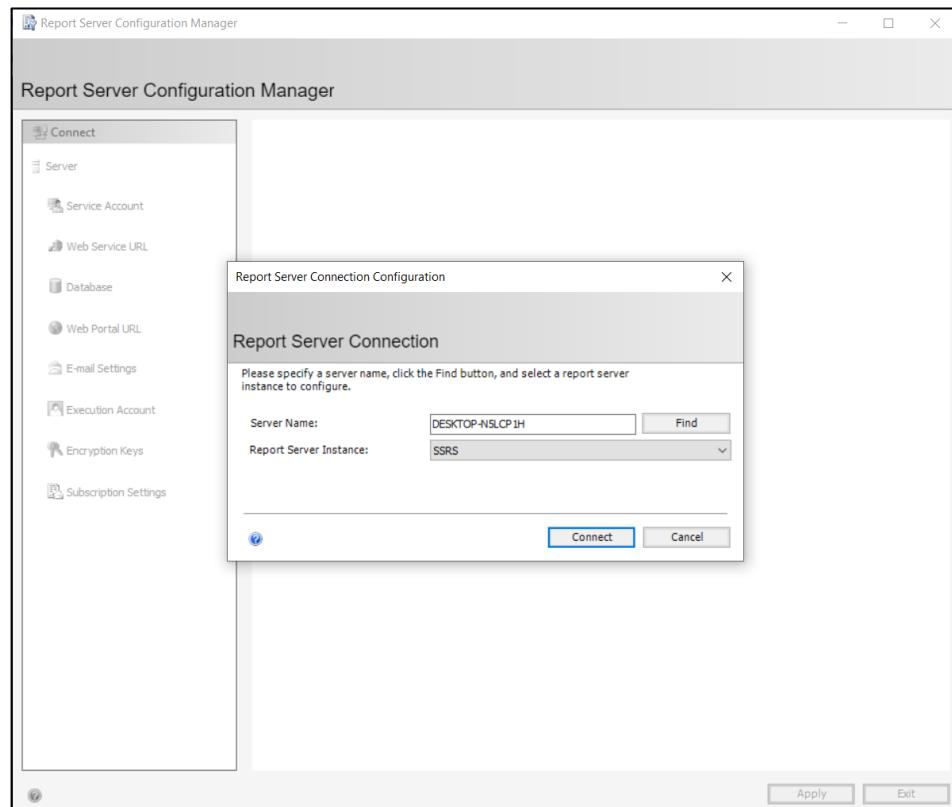
Row Labels	Popularity	Vote Count	Budget	Number of Movies
1899	58129.93359	10332	1429029617	117320
2006	23092	562297	454655208	16374
2007	24461.22852	605553	795232563	16966
2008	25457.50195	638796	1745590727	17833
2009	27383.16992	699557	2137433961	18786
2010	27551.99414	738231	-1877434546	19475
2011	30136.28125	770049	-2010672541	21112
2012	31916.92383	802840	1720144768	23020
2013	34243.42578	965598	-2068006983	25400
2014	37054.11719	1039020	1930782141	27399
2015	37368.05078	935245	1425765795	28791
2016	39386	1057896	-598974539	30301
2017	44615.47656	1072942	2116415969	33733
2018	47009.01172	950621	1263557790	35982
2019	52678.89453	878869	1265496929	39924
2020	49619.04297	491380	277533427	39062
2021	55228.64063	572740	-408467899	41787
2022	66722.59375	452277	1088941398	41849
2023	122321.3359	175566	-1874454887	38107

PHẦN 4: QUÁ TRÌNH TẠO BÁO CÁO DỮ LIỆU(SSRS)

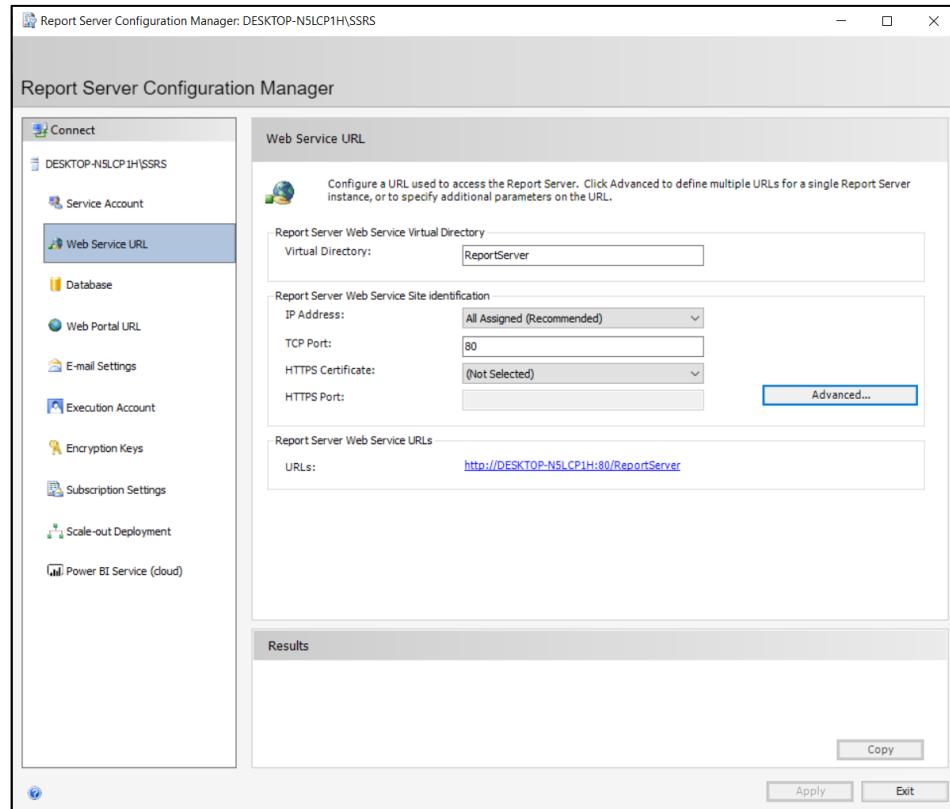
4.1. Tạo báo cáo với Reporting Service trên Visual Studio

4.1.1. Cấu hình Report Server Configuration Manager

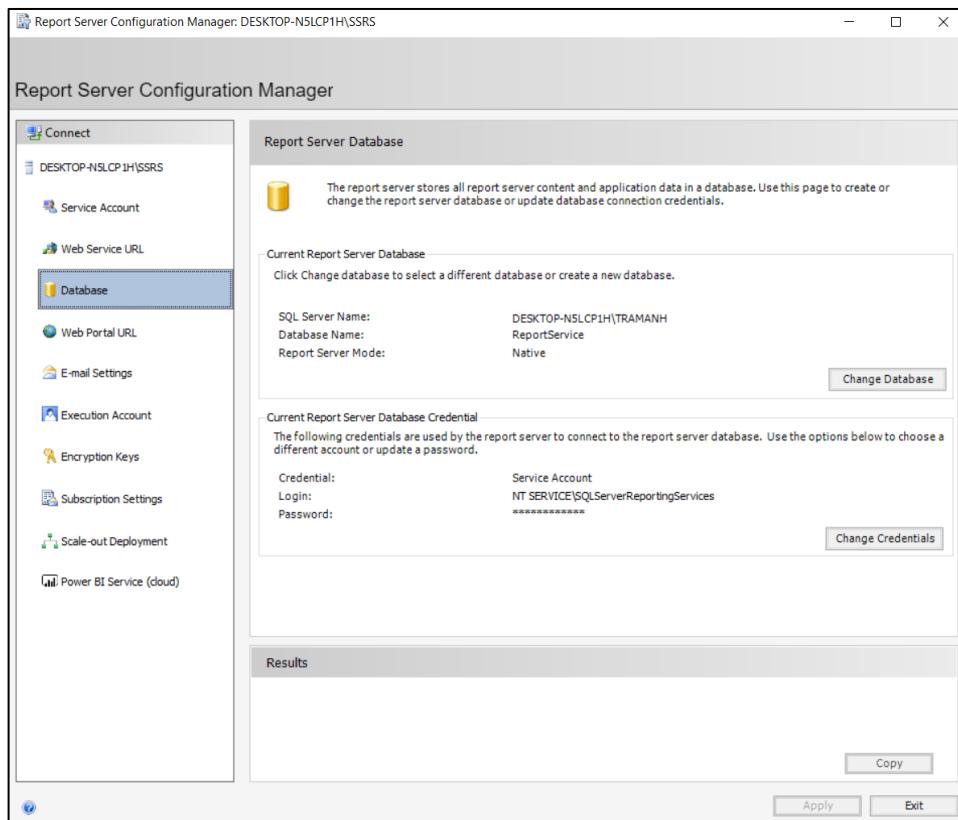
Tải về máy SQLReportingServices.exe. Chạy file.exe vừa tải về và cài đặt như bình thường. Sau khi cài đặt xong, mở Report Server Configuration Manager. Nhấn Connect để kết nối với Server. Mặc định là tên người dùng của hệ điều hành.



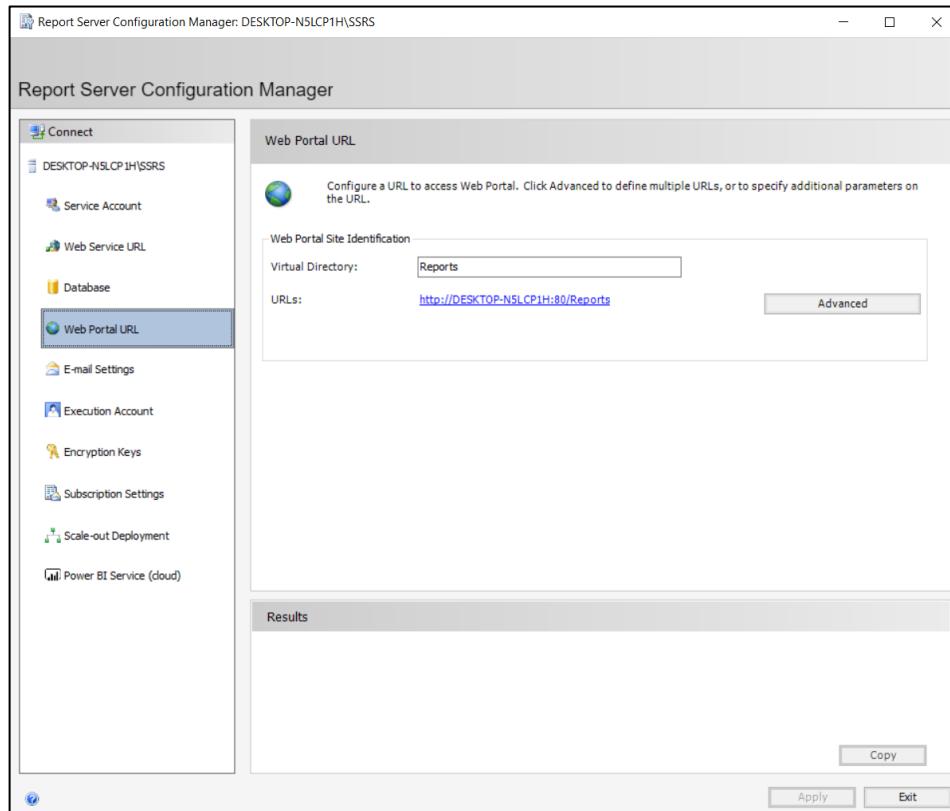
Trong phần Web Service URL, giữ nguyên mặc định bên dưới



Trong phần Database, nhấn chọn Change Database để thay đổi cơ sở dữ liệu.

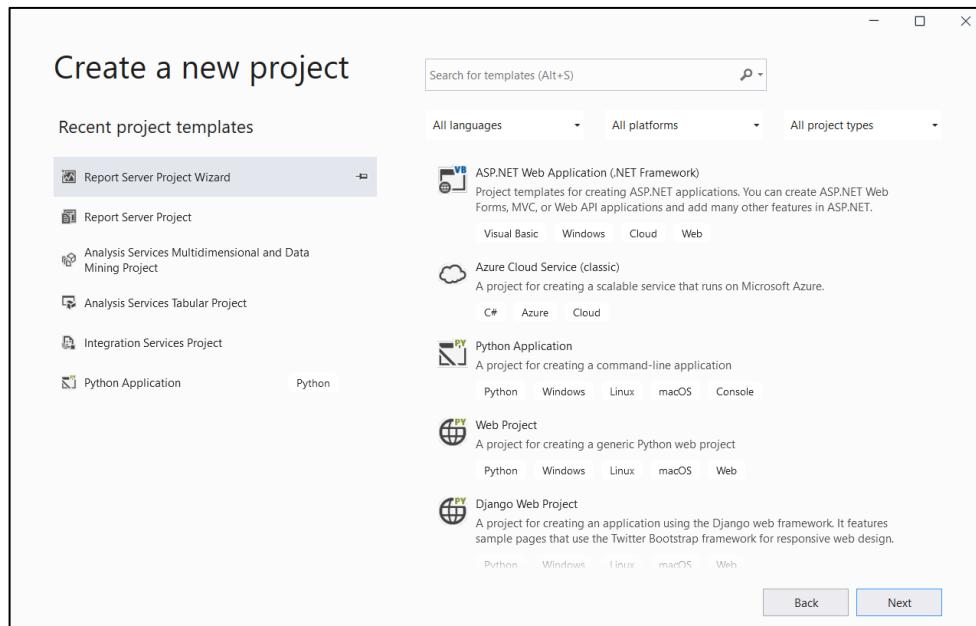


Trong phần Web Portal URL, kiểm tra các giá trị có trong hình dưới đây. Cuối cùng nhấn Apply.

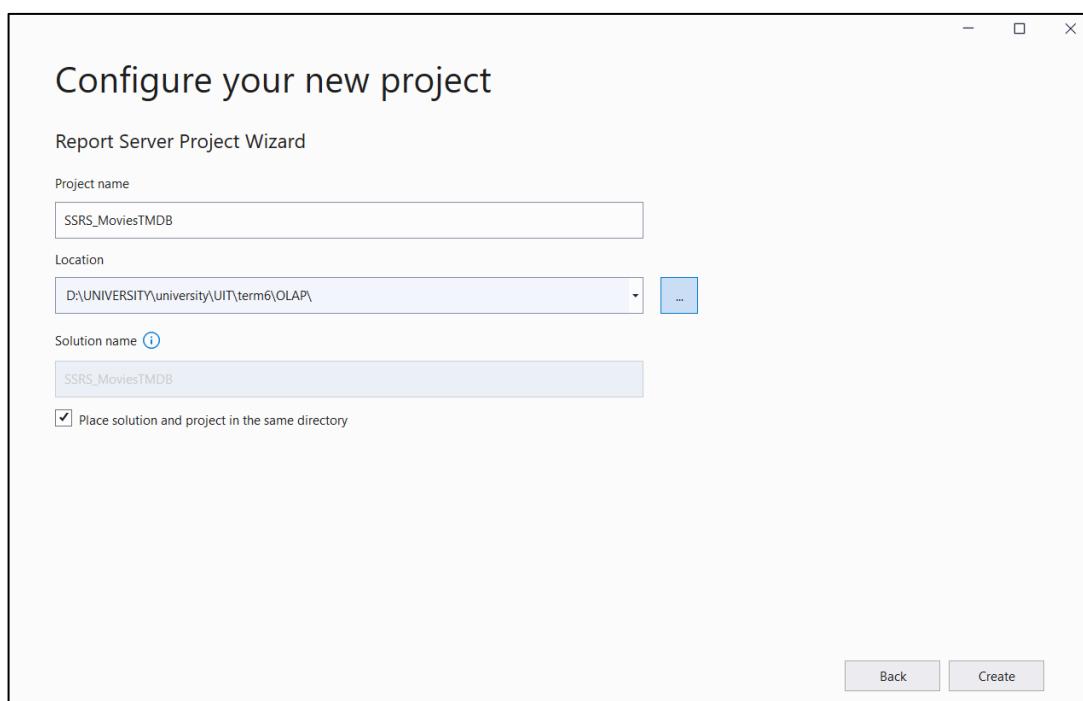


4.1.2. Khởi tạo và cấu hình project Visual Studio

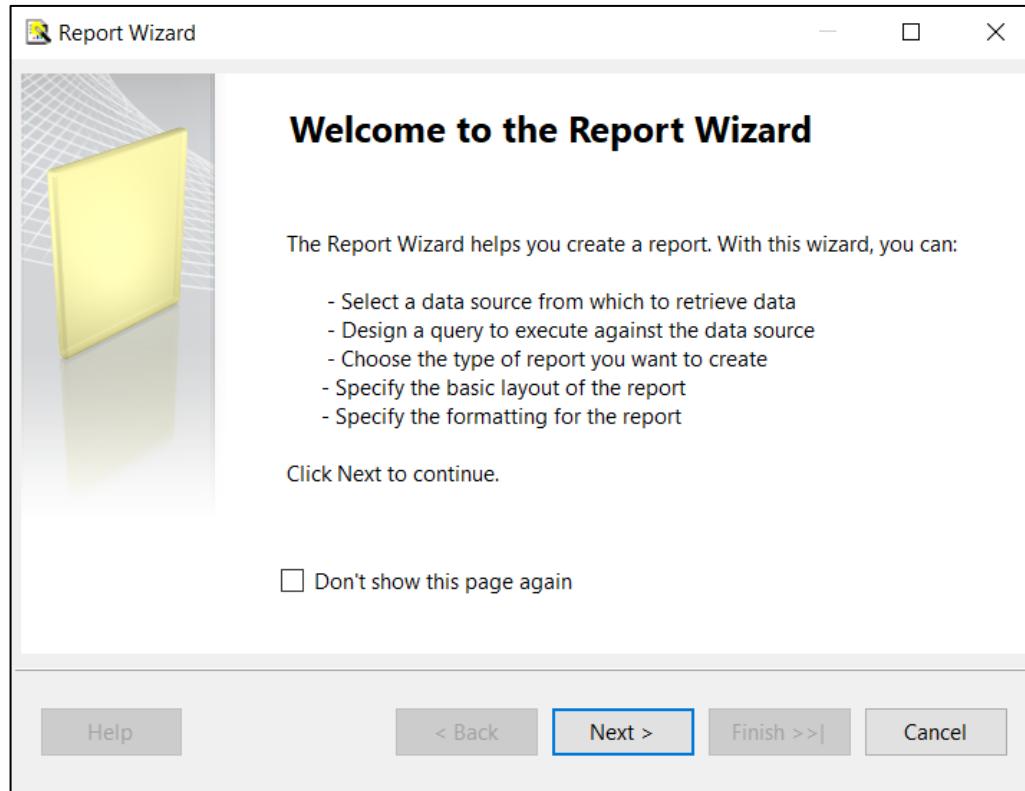
Mở Visual Studio, chọn File → New → Project để tạo mới một project. Tại cửa sổ New Project, chọn Report Server Project Wizzard, đặt tên và chọn thư mục thích hợp. Nhấn OK để xác nhận tạo Project SSRS.



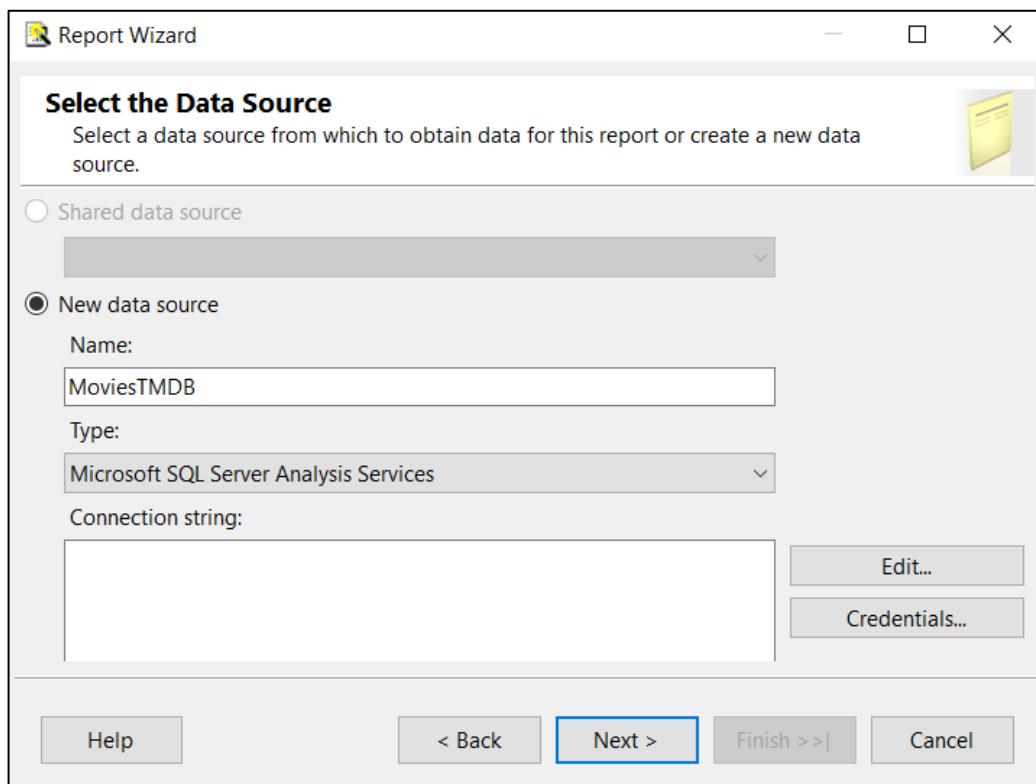
Màn hình xuất hiện khung làm việc như hình dưới và đặt tên cho project.



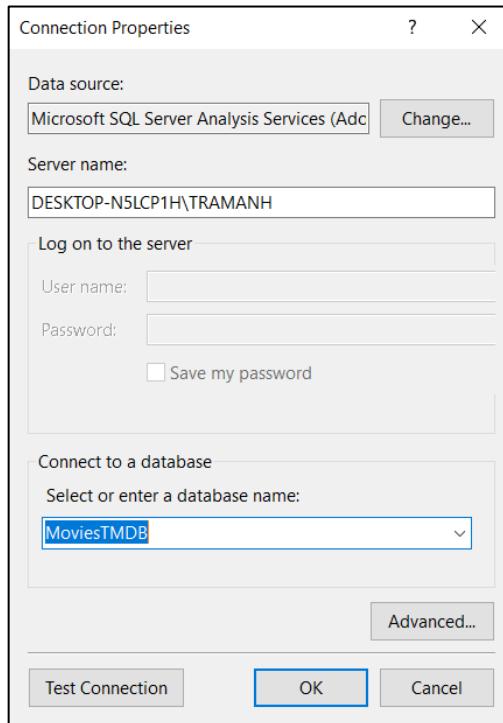
Màn hình xuất hiện Report Wizard và nhấn Next.



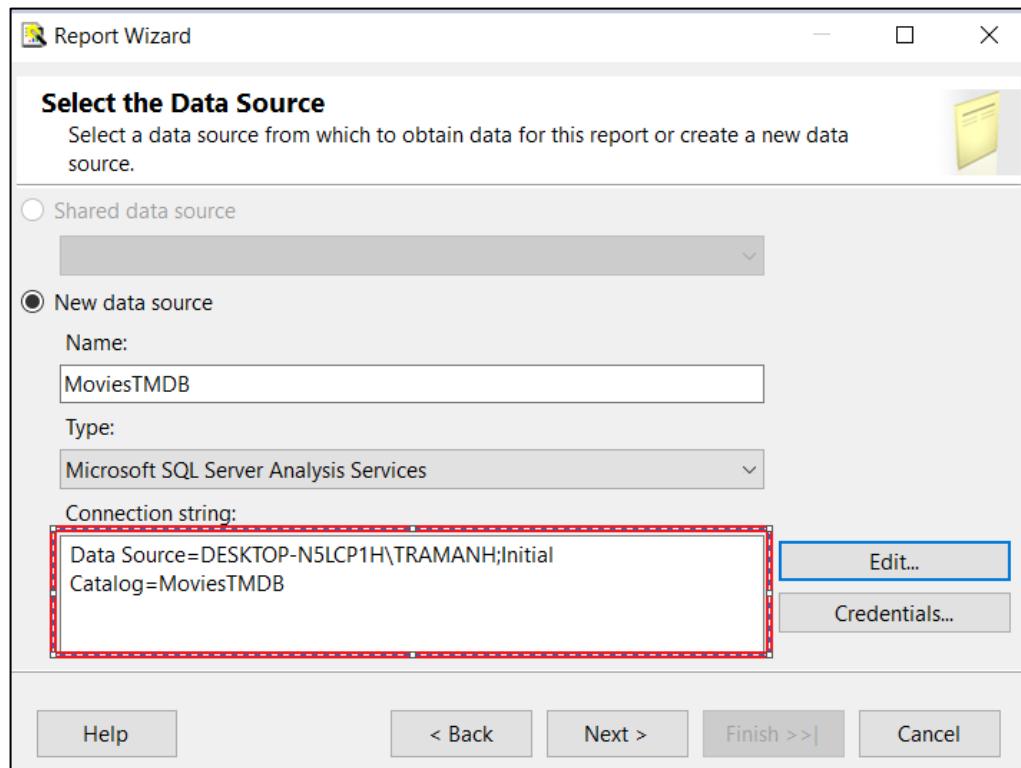
Ta đặt tên data soure và ở Type chọn Microsoft SQL Server Analysis Services, sau đó chọn Edit để tạo kết nối.



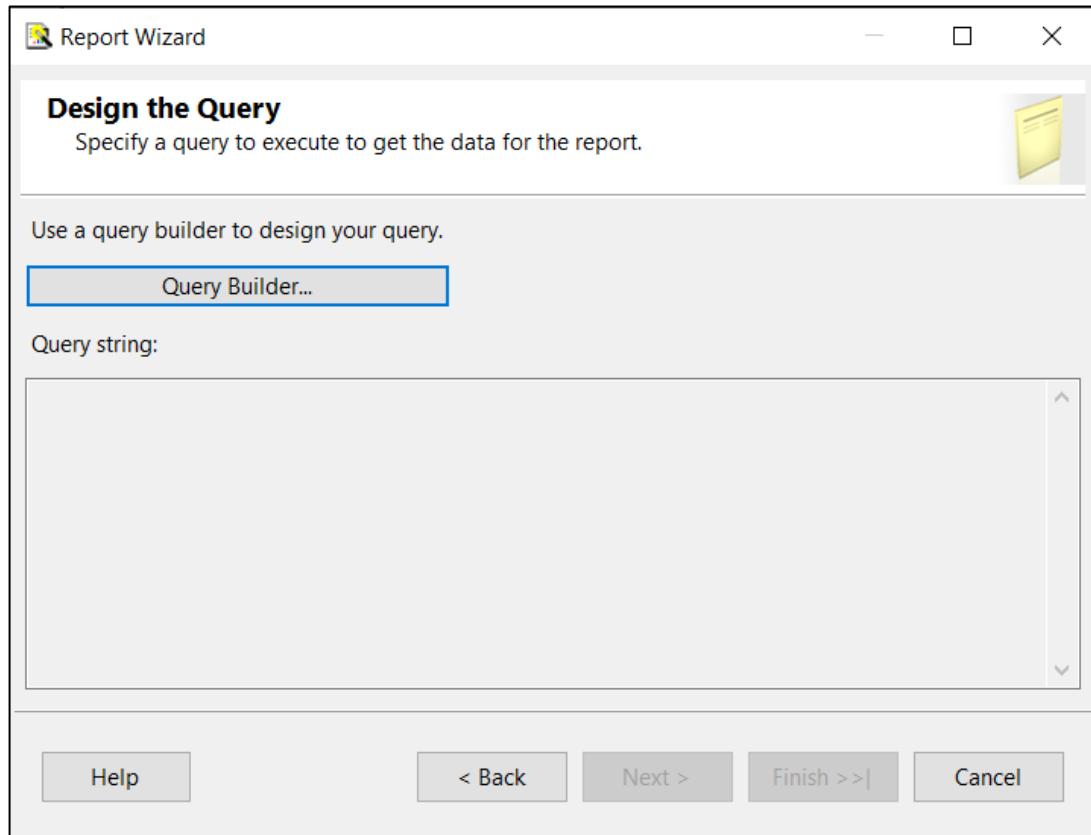
Màn hình xuất hiện Connection Properties, điền Server name, và ở mục Select or enter a database name, ta chọn database MoviesTMDB.

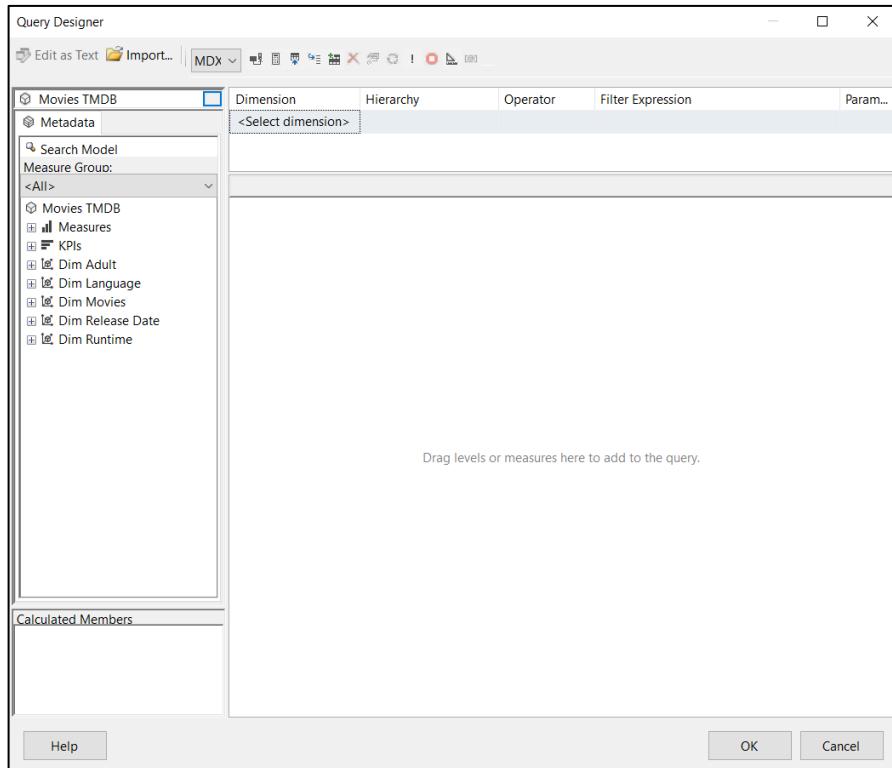


Khi đó, ta có kết quả ở Connection string.



Sau đó, ta thực hiện Design the Query bằng cách nhấp vào Query Builder, thực hiện kéo thả như Manual.





4.1.3. Báo cáo 1

Nội dung: thống kê số lượng phim và tổng doanh thu của theo từng năm từ 2014 đến 2024 và độ tuổi người xem cho phép (trên 18 tuổi hay không).

Tại Query Design, thực hiện kéo thả (Manual) và ấn OK.

Query Designer

Edit as Text Import... MDX

Movies TMDB

Dimension Hierarchy Operator Filter Expression Param...

Dim Release Date Release Year Range (Inclu...) 2014 : 2024

<Select dimension>

Release Year	Adult	Number of Movies	Revenue
2014	False	23954	3026772...
2014	True	3445	90
2015	False	24924	3117910...
2015	True	3867	0
2016	False	26098	3296090...
2016	True	4203	3
2017	False	29248	3399072...
2017	True	4485	24889
2018	False	31057	3284576...
2018	True	4925	318596
2019	False	34377	3215282...
2019	True	5547	7247128
2020	False	34326	7218676...
2020	True	4736	21650
2021	False	37010	1704887...
2021	True	4777	307532
2022	False	37307	2217834...
2022	True	4542	5667069...
2023	False	34802	2229620...
2023	True	3305	1065336...
2024	False	3940	1001977...
2024	True	312	5101

Calculated Members

Help OK Cancel

Report Wizard

Design the Query

Specify a query to execute to get the data for the report.

Use a query builder to design your query.

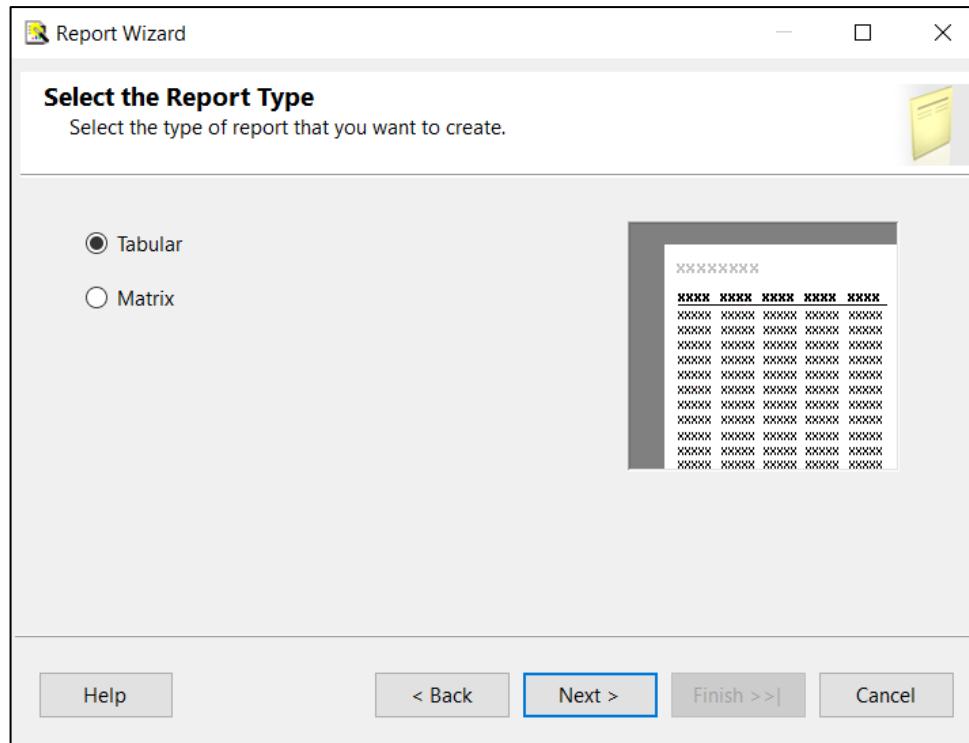
Query Builder...

Query string:

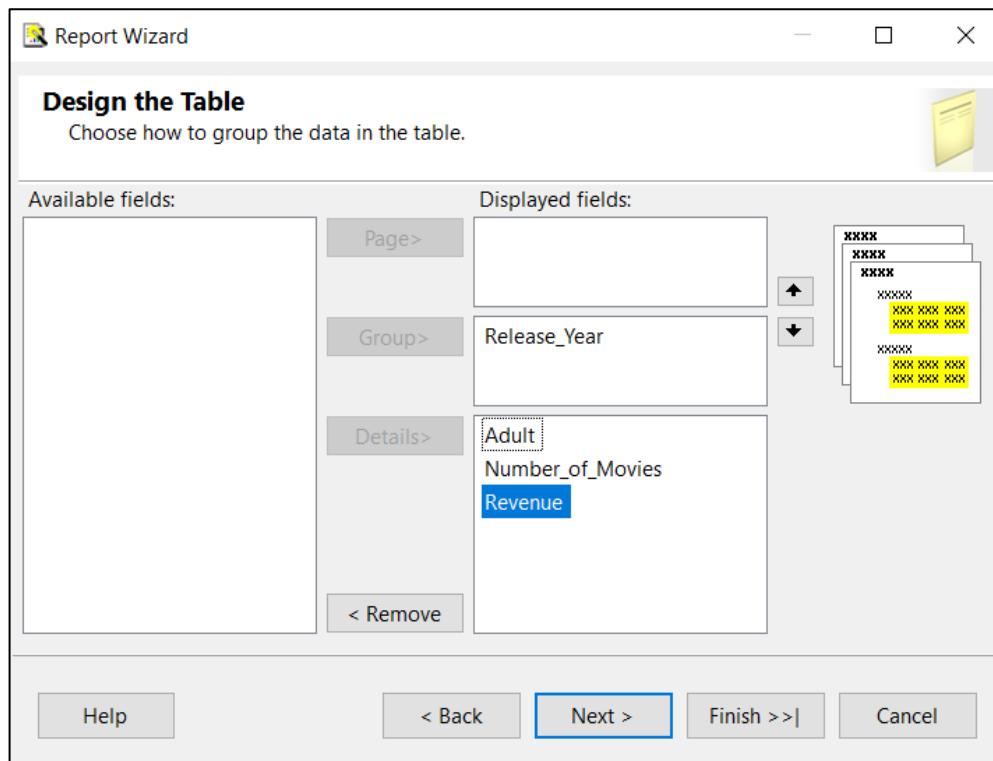
```
SELECT NON EMPTY { [Measures].[Number of Movies], [Measures].[Revenue] } ON COLUMNS, NON
EMPTY { ([Dim Release Date].[Release Year].[Release Year].ALLMEMBERS * [Dim Adult].[Adult],
[Adult].ALLMEMBERS) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_VALUE,
MEMBER_UNIQUE_NAME ON ROWS FROM ( SELECT ( [Dim Release Date].[Release Year].&[2014] :
[Dim Release Date].[Release Year].&[2024] ) ON COLUMNS FROM [Movies TMDB] ) CELL
PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING,
FONT_NAME, FONT_SIZE, FONT_FLAGS
```

Help < Back Next > Finish >> Cancel

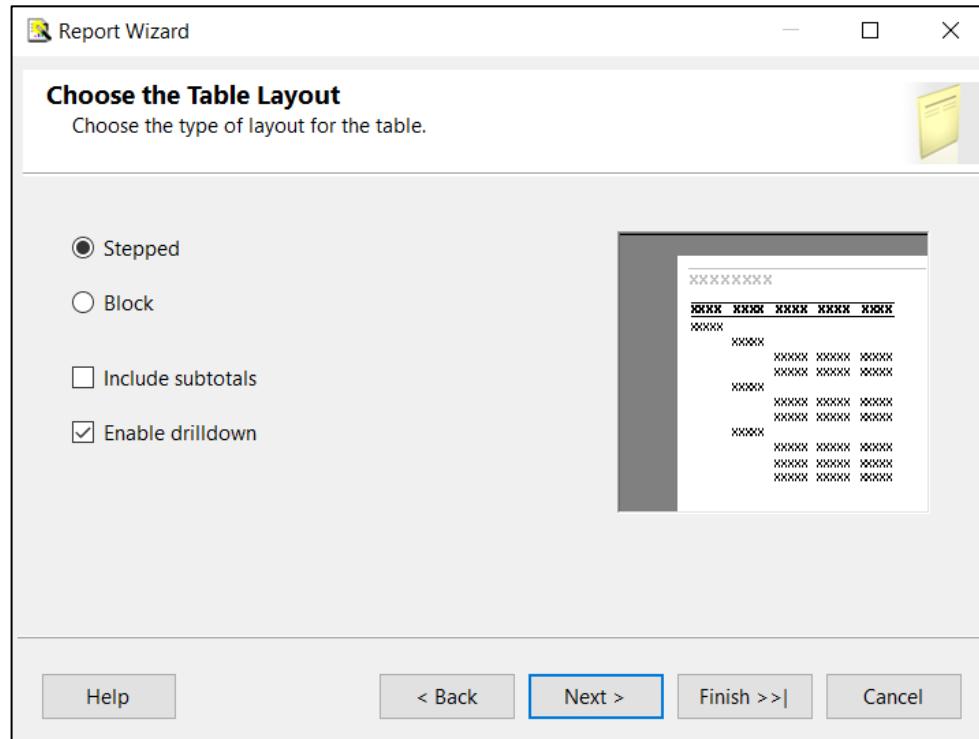
Tại Select the Report Type, ta chọn Tabular.



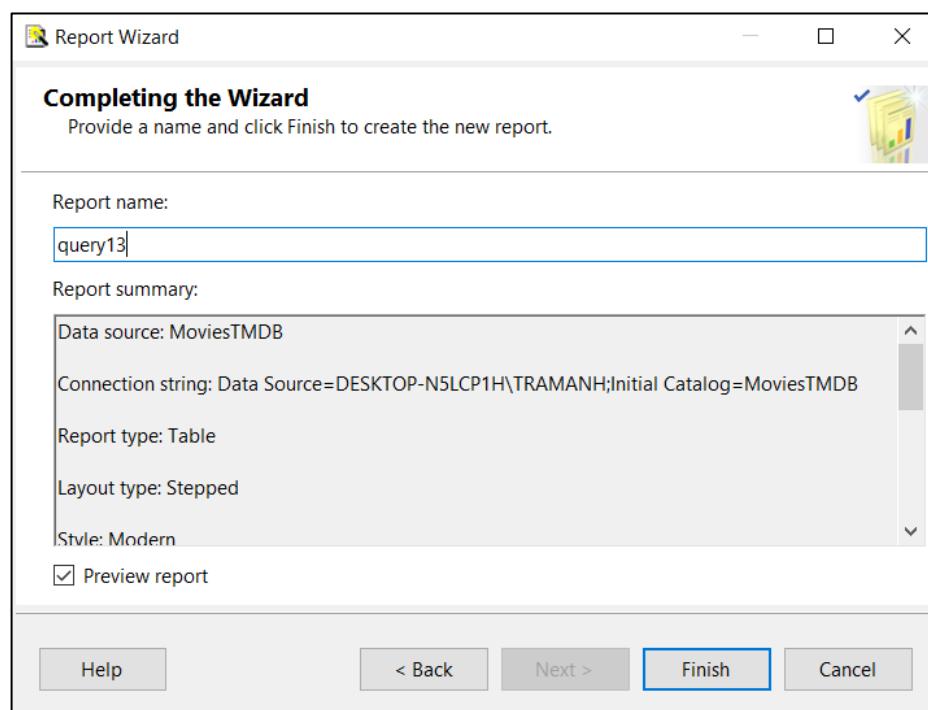
Tại phần Design the Table, ta thực hiện kéo thả các thuộc tính ở Available fields sang Group và Details (tùy nội dung query).

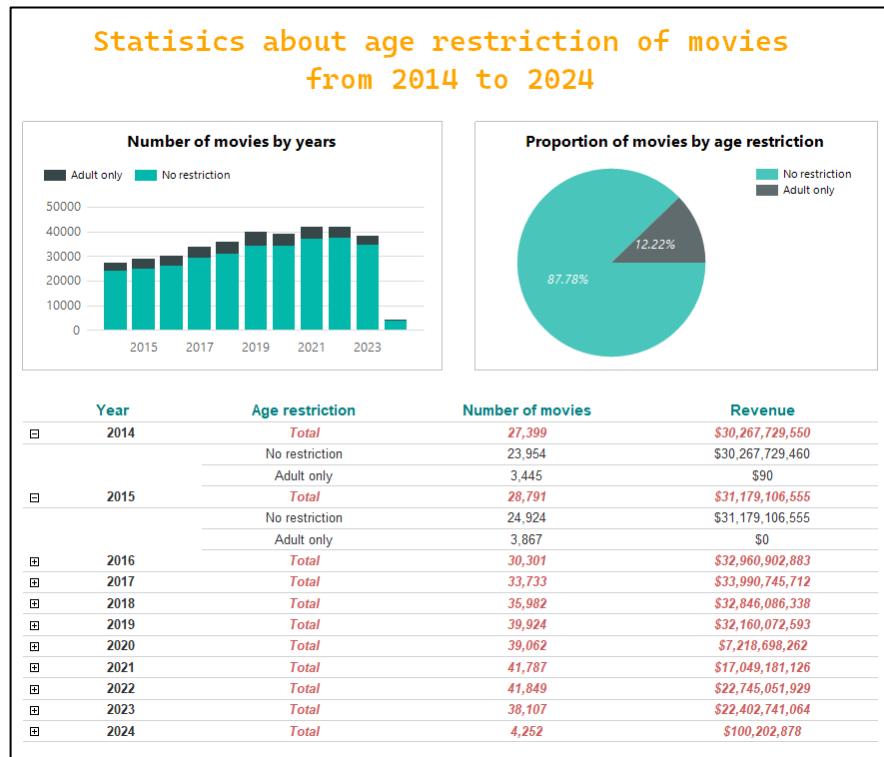
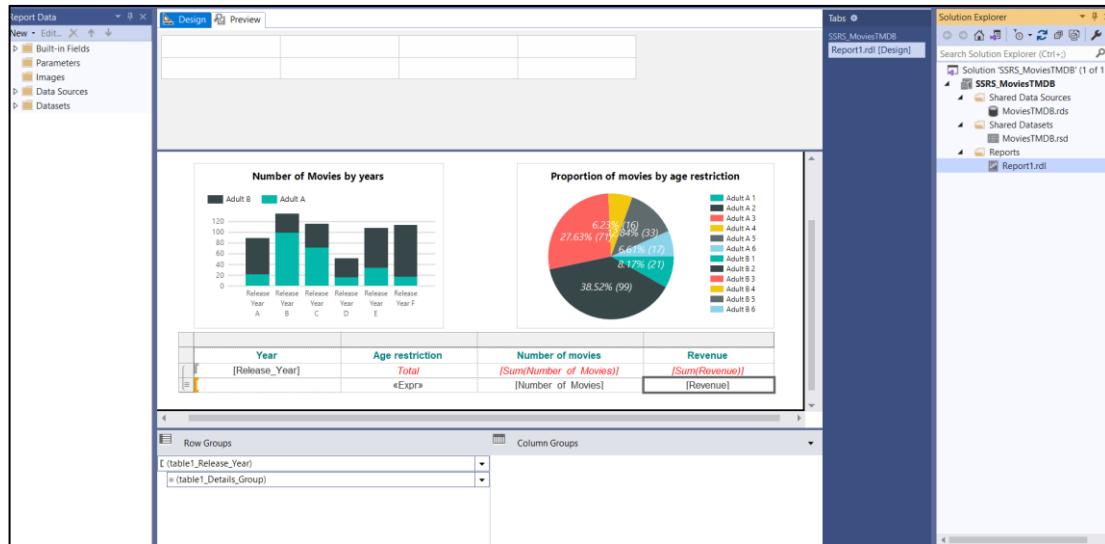


Tại phần Choose the Table Layout, ta chọn Stepped và Enable drilldown. Nhấn Next.



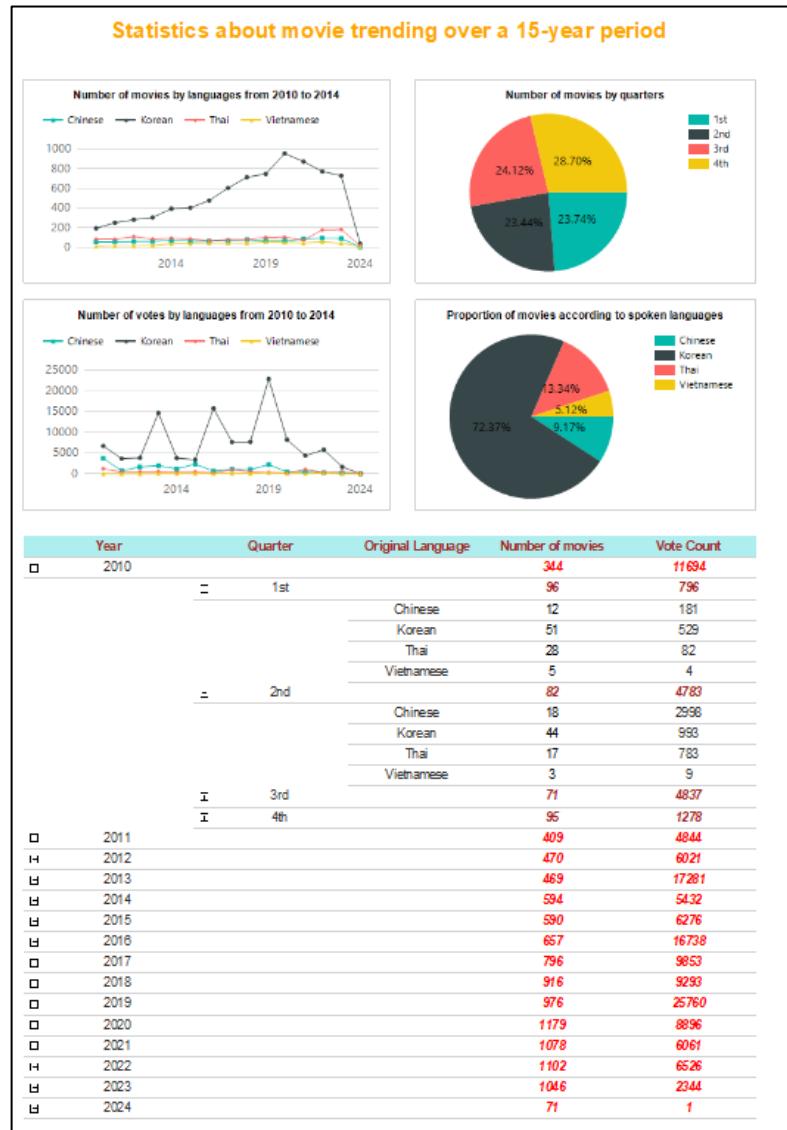
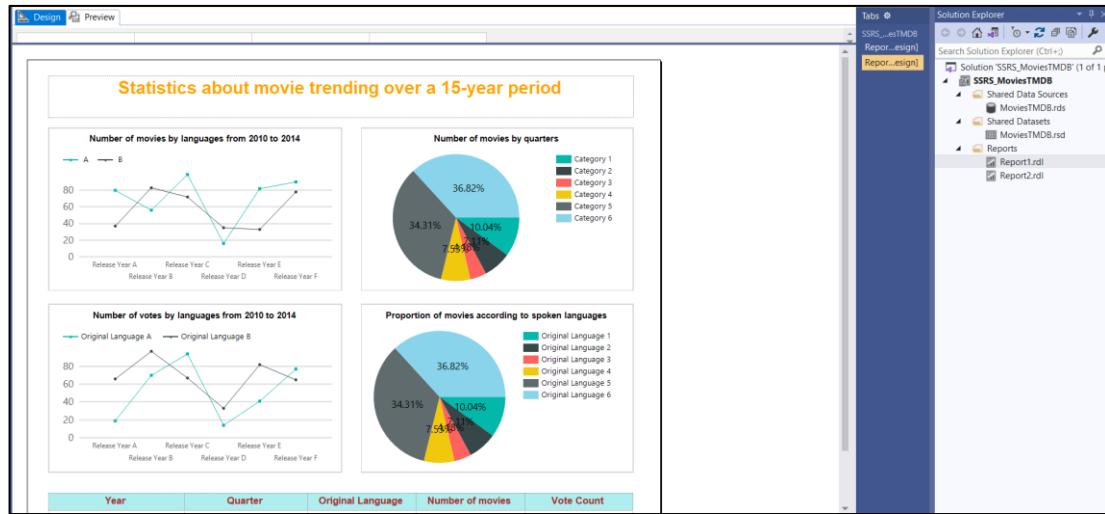
Tại phần Competing the Wizard, điền tên cho Report và chọn Preview report. Nhấn Finish.





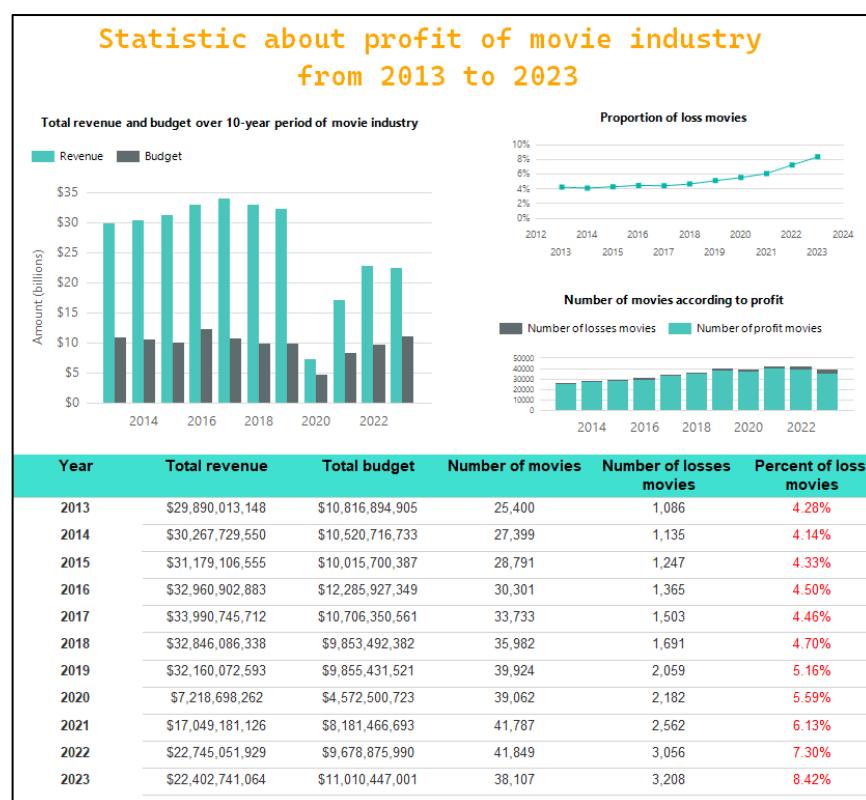
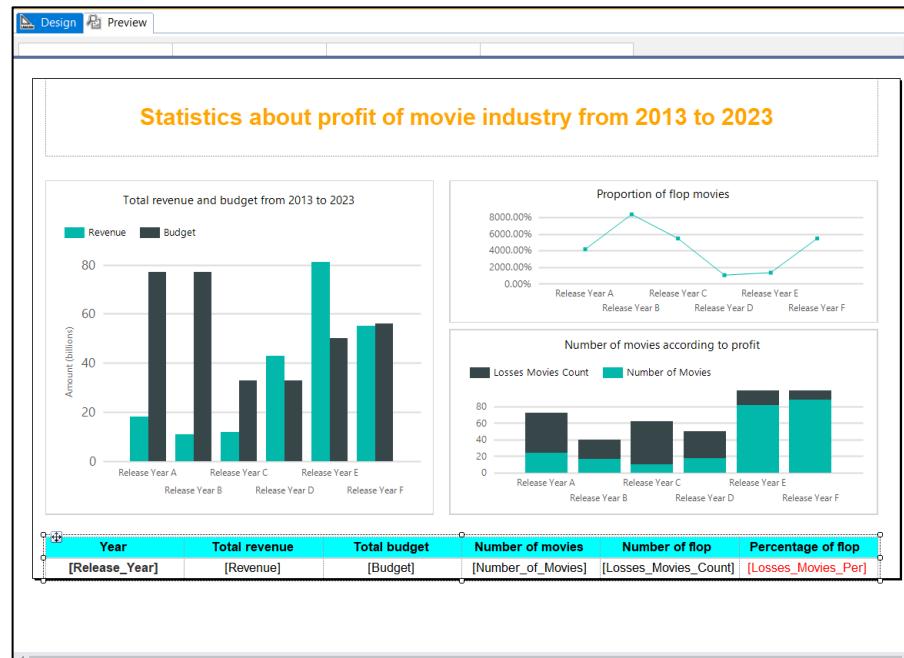
4.1.4. Báo cáo 2

Nội dung: thống kê số lượng phim và tổng số lượt bình chọn theo từng quý từ năm 2010 đến 2024 của hai ngôn ngữ là tiếng Trung Quốc, Hàn Quốc, Thái Lan và Việt Nam.



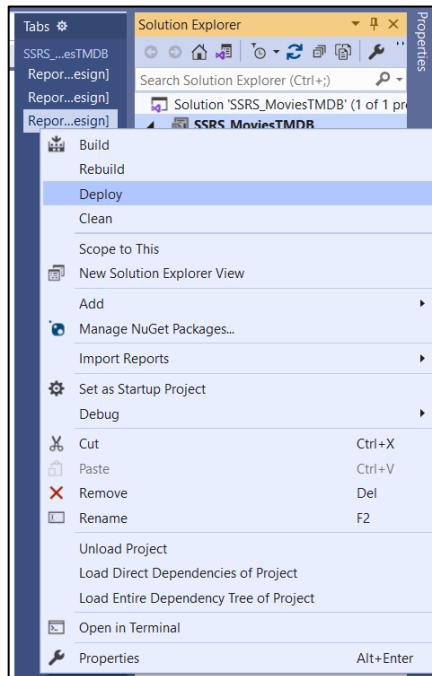
4.1.5. Báo cáo 3

Nội dung: Thông kê số lượng, phần trăm, doanh thu và ngân sách các bộ phim có chi phí sản xuất lớn hơn doanh thu (gặp tình trạng lỗ) từ năm 2013 đến 2023.(câu 11)

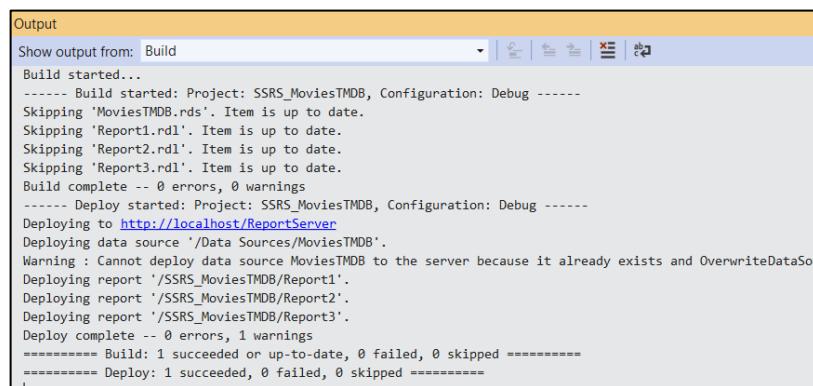


4.1.6. Deploy các báo cáo

Nhấn chuột phải chọn tên project, chọn Deploy để triển khai các báo cáo.



Nếu deploy thành công, pannel Output hiện ra thông báo như hình bên dưới.



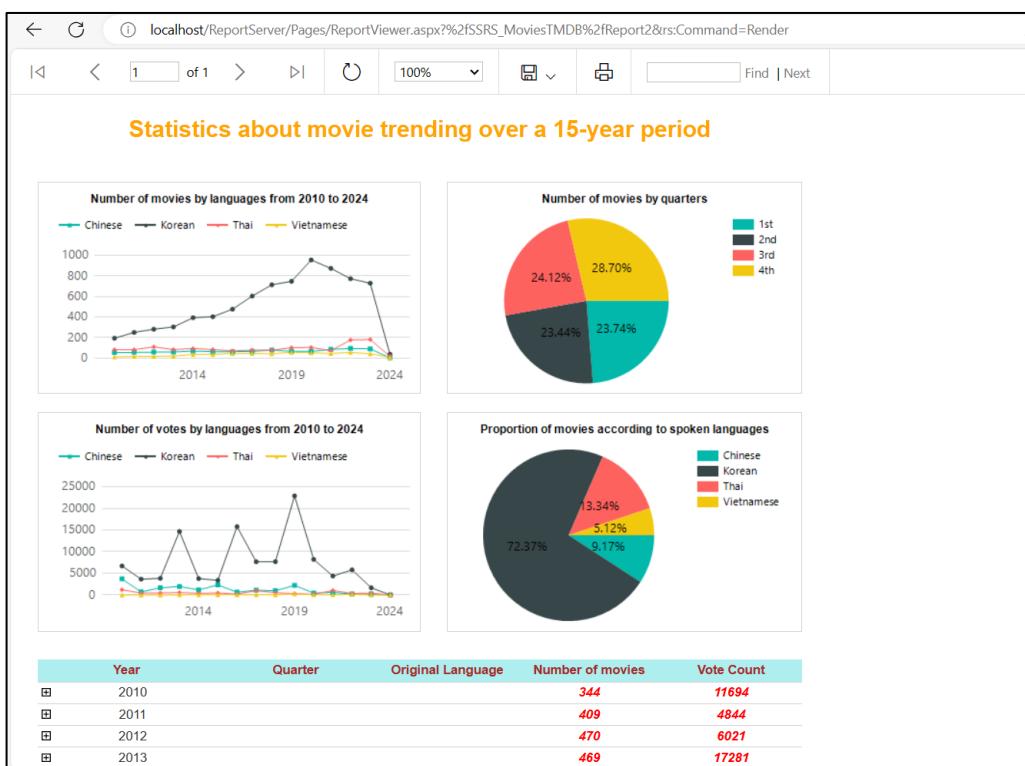
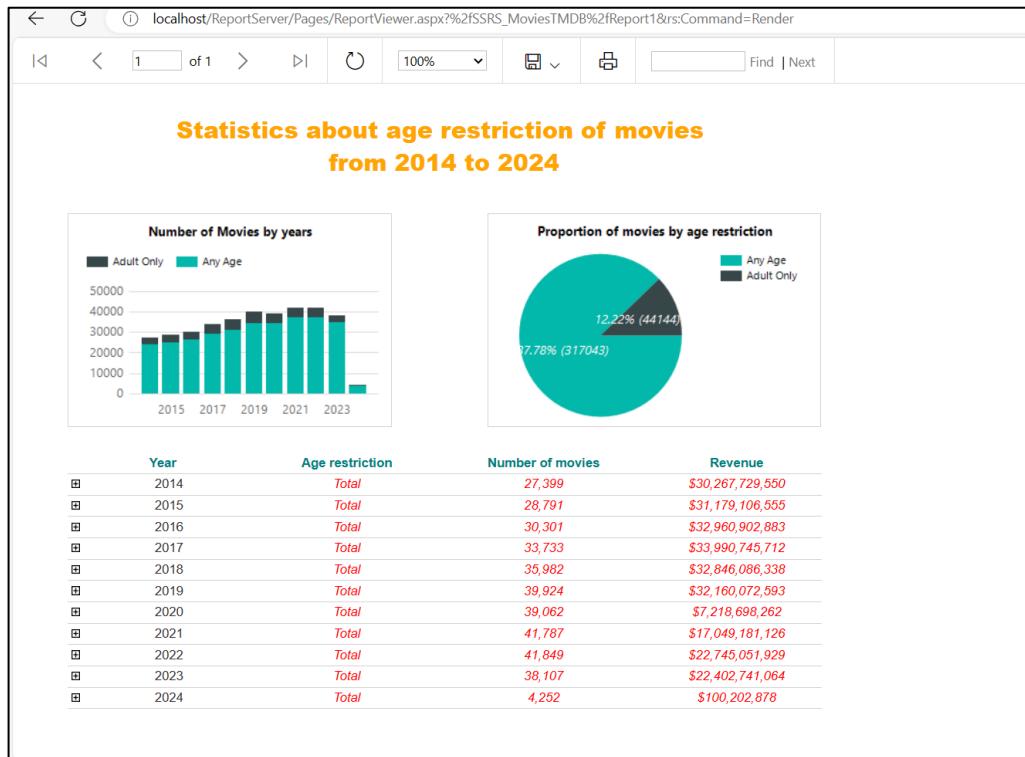
Truy cập vào đường link đã được cấu hình từ trước để truy cập vào các report. Đường link nhóm tạo được là: <http://localhost/ReportServer>

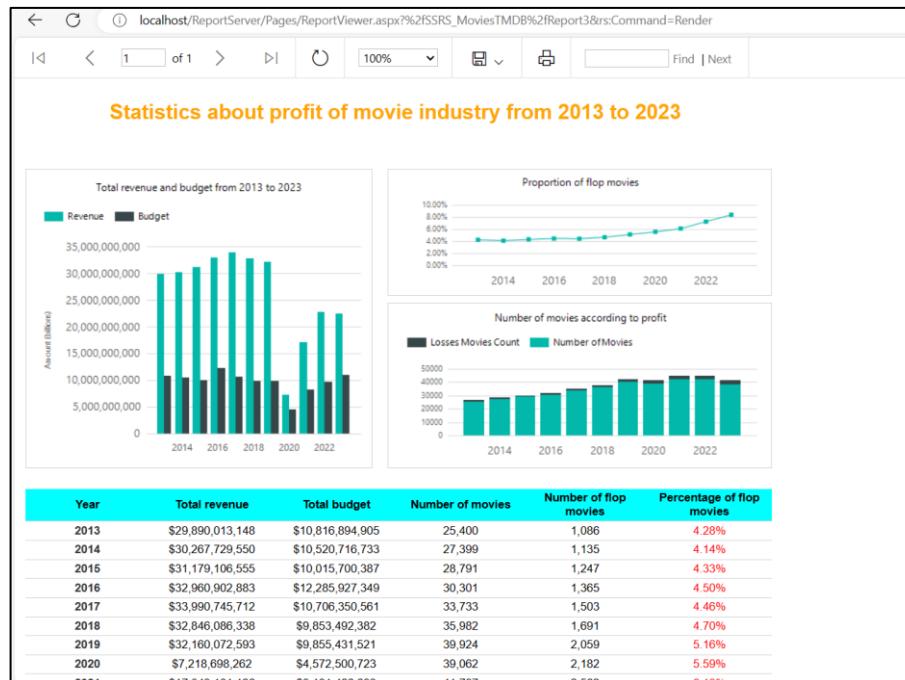
localhost/ReportServer - /

Tuesday, April 30, 2024 4:53 PM	<dir> Data Sources
Tuesday, April 30, 2024 4:55 PM	<dir> Datasets
Tuesday, April 30, 2024 11:18 AM	<dir> Query1
Tuesday, April 30, 2024 4:52 PM	<dir> SSRS_MoviesTMDB

Microsoft SQL Server Reporting Services Version 15.0.1102.1167

Review các báo cáo và lưu về máy





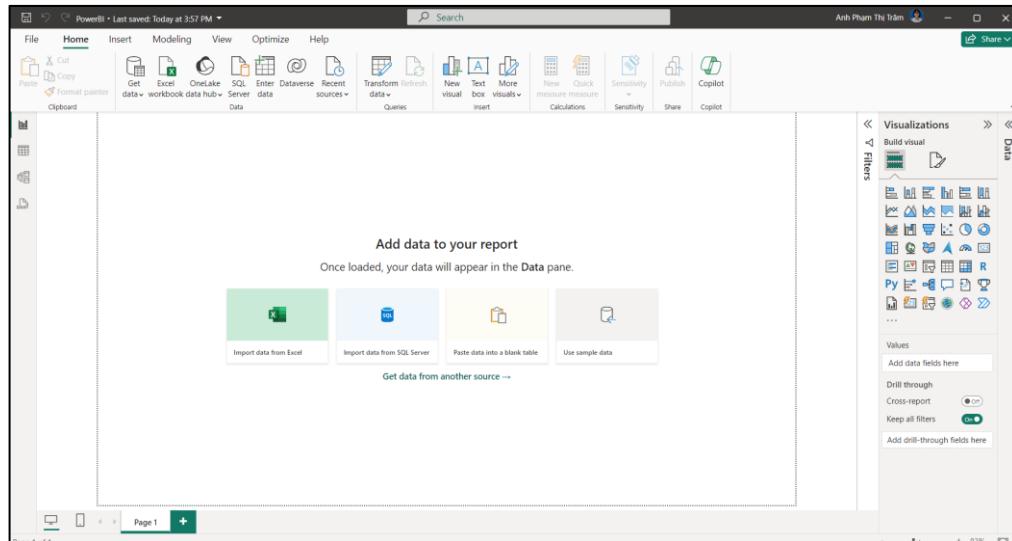
4.2. Tạo báo cáo với PowerBI

4.2.1. Cài đặt PowerBI

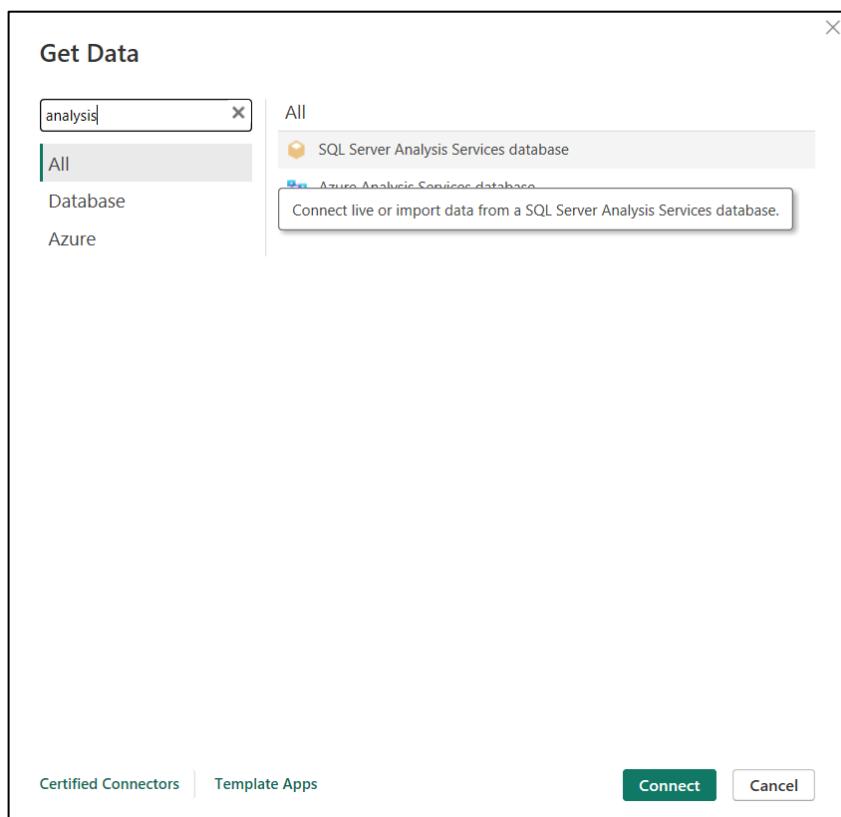
Tiến hành cài đặt Power BI về máy, truy cập [Download Microsoft Power BI Desktop from Official Microsoft Download Center](#). Cài đặt bình thường như các ứng dụng khác, sau khi cài đặt xong, đăng nhập bằng toàn khoản có đuôi @ms.uit.edu.vn và sử dụng.

4.2.2. Tạo project SSRS trên Power BI

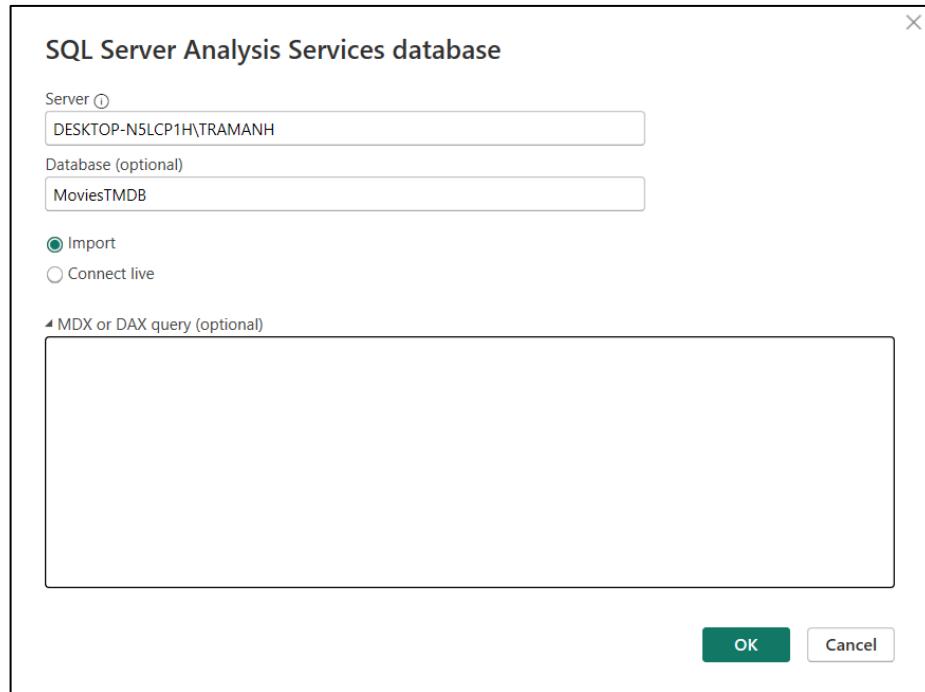
Mở Power BI, tạo mới một project và lưu vào thư mục phù hợp. Để tạo dataset, nhấn Get data from another source



Ta tìm từ khóa trong thanh tìm kiếm và chọn SQL Server Analysis Services database.



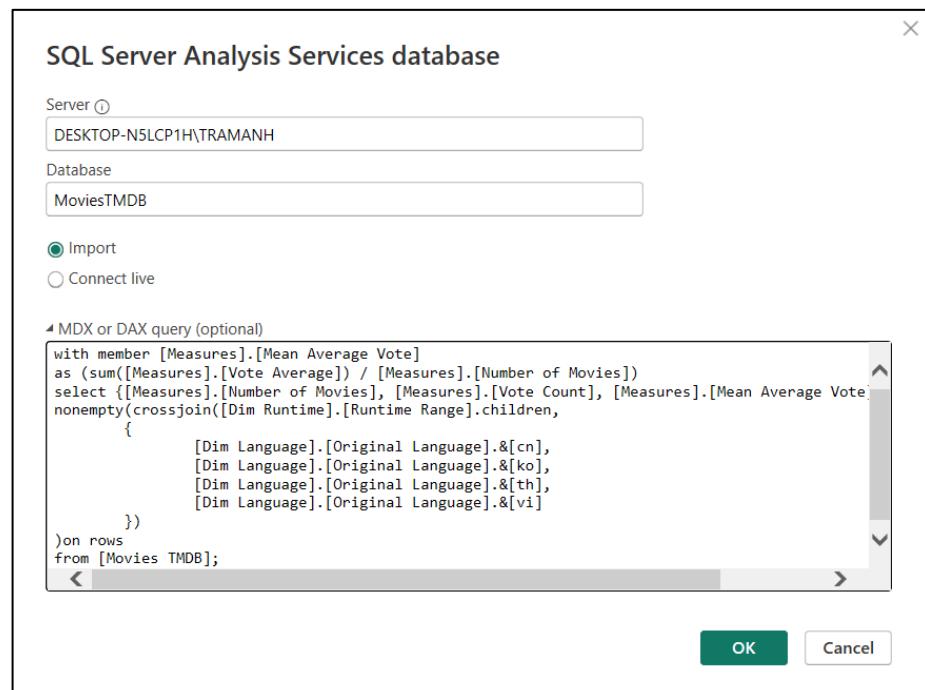
Nhập Server name và Database, sau đó chọn Import, bên dưới là MDX or DAX, nhập câu truy vấn MDX phù hợp cho report. Nhấn OK để nhận lấy dữ liệu.



4.2.3. Báo cáo 1

Nội dung: Thông kê tổng số phim, số lượt bình chọn, điểm bình chọn trung bình và điểm thịnh hành của các phim theo thời lượng phim và một số ngôn ngữ khác nhau (Trung Quốc, Hàn Quốc, Thái Lan và Việt Nam). (câu 8)

Nhập vào MDX câu truy vấn số 8 và nhấn OK để xác nhận.



Cửa sổ review dữ liệu hiện ra, nhấn Transform Data để biến đổi dữ liệu.

[Dim Runtime].[Runtime Range].[Runtime Range].[MEM]	[Dim Language].[Original Language].[Original Language]	[Measures].[Number of Movies]	[Measure]
1h - 2h	cn	3517	
1h - 2h	ko	5741	
1h - 2h	th	1367	
1h - 2h	vi	402	
2h - 3h	cn	118	
2h - 3h	ko	719	
2h - 3h	th	287	
2h - 3h	vi	34	
over 3h	cn	13	
over 3h	ko	131	
over 3h	th	20	
over 3h	vi	10	
under 1h	cn	1337	
under 1h	ko	5605	
under 1h	th	1164	
under 1h	vi	625	

Tiến hành đổi tên cột và đổi tên bộ dữ liệu này là Query8 tại mục Name.

runtime range	Original languages	Number of Movies	Vote Count	Mean Average Vote
1h - 2h	cn	3517	72101	3.959097942
1h - 2h	ko	5741	71623	3.388345141
1h - 2h	th	1367	12382	3.434124207
1h - 2h	vi	402	1258	2.490110745
2h - 3h	cn	118	6839	4.546940173
2h - 3h	ko	719	84734	4.930282741
2h - 3h	th	287	3419	2.641376236
2h - 3h	vi	34	167	3.871235367
over 3h	cn	13	15	1.628230755
over 3h	ko	131	632	3.46718112
over 3h	th	20	44	1.9875
over 3h	vi	10	3	1.319999981
under 1h	cn	1337	380	0.879053885
under 1h	ko	5605	1632	0.69631646
under 1h	th	1164	381	1.188280138
under 1h	vi	625	171	0.907200098

Nhấn chuột phải vào Mean Average Vote để đổi thành 2 chữ số thập phân. Sau khi biến đổi xong, chọn Close and Apply để lấy dữ liệu và chuyển vào màn hình làm việc.

Queries [1]

Query8

	ABC runtime range	ABC Original languages	ABC 123 Number of Movies
1	1h - 2h	cn	35
2	1h - 2h	ko	57
3	1h - 2h	th	13
4	1h - 2h	vi	4
5	2h - 3h	cn	1
6	2h - 3h	ko	7
7	2h - 3h	th	2
8	2h - 3h	vi	
9	over 3h	cn	

1.2 Mean Average Vote

PROPERTIES

STEPS

Query8
Renamed Columns
Selected Off

Transform

- Round
- Absolute Value
- Factorial
- Base-10 Logarithm
- Natural Logarithm
- Power
- Square Root
- Text Transforms

Remove
Remove Other Columns
Duplicate Column
Add Column From Examples...
Remove Duplicates
Remove Errors
Change Type
Replace Value
Replace Error
Group By...
Fill
Unpivot Columns
Unpivot Other Columns
Unpivot Only Selected Columns
Rename...
Move
Drill Down
Add as New Query

Statistic about vote and popularity of movies in four languages

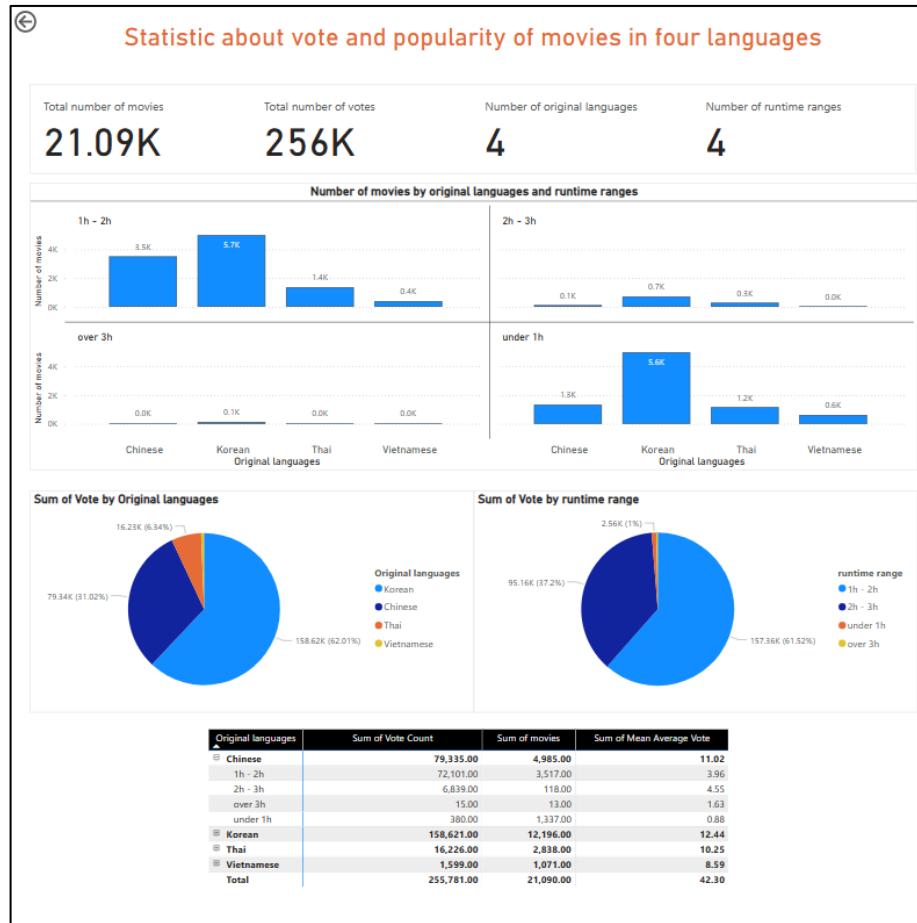
Total number of movies: 21.09K Total number of votes: 256K Number of original languages: 4 Number of runtime ranges: 4

Number of movies by original languages and runtime ranges

Sum of Vote by Original languages

Sum of Vote by runtime range

Original language	Type of Vote Count	Sum of votes	Sum of Mean Average Vote
Chinese	79,220,00	4,765,40	11,60
Korean	10,214,00	6,635,00	10,50
Thai	10,214,00	11,00	4,00
Vietnamese	10,214,00	1,337,00	1,00
Total	203,648,00	12,796,40	10,40
		107,846,00 (52%)	



4.2.4. Báo cáo 2

Nội dung: Tổng doanh thu và ngân sách phim từ năm 2014 đến năm 2024, theo các quý và các tháng. (câu 9)

Tương tự báo cáo 1, ta nhập câu truy vấn số 9 và tiến hành biến đổi dữ liệu, bao gồm:

- Đổi tên tất cả các cột
- Đặt tên kí hiệu:
 - Đối với cột quarter, ta đổi thành 1st Quarter, 2nd Quarter, 3rd Quarter, 4th Quarter
 - Đối với cột Month, ta đổi thành January, February, March, April, May, June, August, September, October, November and December
- Đổi kiểu dữ liệu của Budget và Revenue thành fixed decimal number

Screenshot of PowerBI desktop showing a query editor and a preview of data.

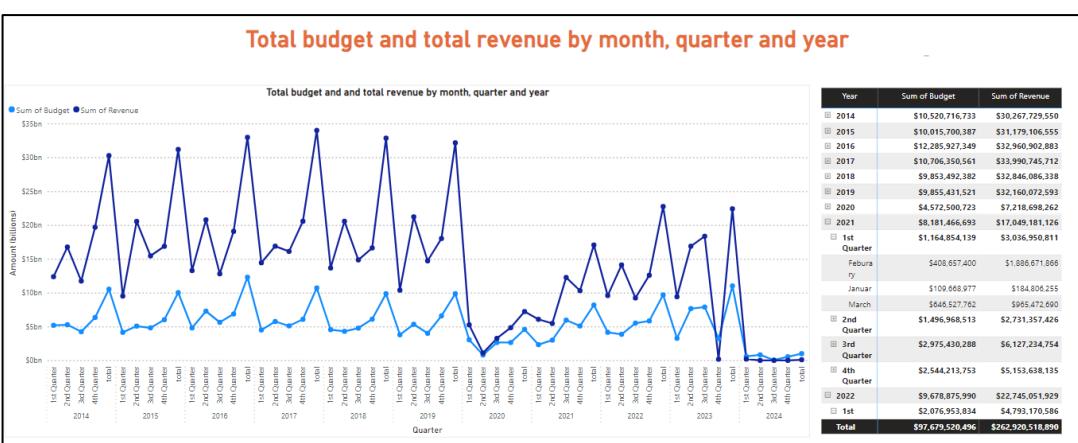
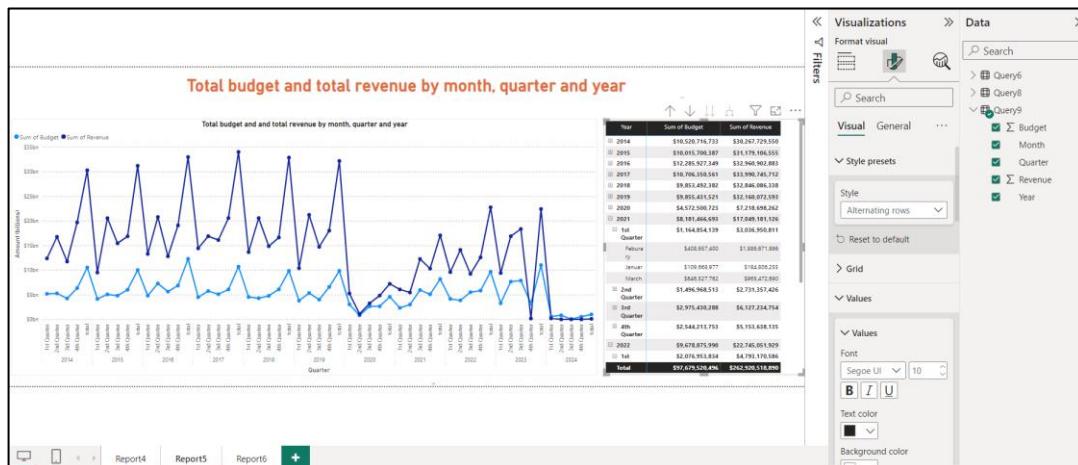
Query Editor:

- Queries [3]: Query8, Query9, Query6
- Current Query: Table.ReplaceValue("#Replaced Value17", "Feb", "February", Replacer.ReplaceText, {Month})
- Table Headers: Year, Quarter, Month, Budget, Revenue
- Data Preview (21 rows):

Year	Quarter	Month	Budget	Revenue
2014	total	null	10,520,716,733.00	30,267,729,500.00
2014	1st Quarter	null	2,593,431,871.00	6,184,197,518.00
2014	1st Quarter	January	710,640,981.00	1,730,499,231.00
2014	1st Quarter	February	778,692,386.00	1,717,738,607.00
2014	1st Quarter	March	1,104,098,504.00	2,735,959,680.00
2014	2nd Quarter	null	2,633,846,899.00	8,377,817,464.00
2014	2nd Quarter	April	593,936,574.00	1,683,114,227.00
2014	2nd Quarter	May	1,229,635,803.00	3,481,365,876.00
2014	2nd Quarter	June	812,274,522.00	3,213,337,361.00
2014	3rd Quarter	null	2,119,193,088.00	5,864,956,307.00
2014	3rd Quarter	July	682,350,316.00	2,716,418,892.00
2014	3rd Quarter	August	771,760,073.00	1,826,990,678.00
2014	3rd Quarter	September	665,082,699.00	1,319,546,737.00
2014	4th Quarter	null	3,172,244,875.00	9,840,758,261.00
2014	4th Quarter	October	983,575,328.00	2,724,230,338.00
2014	4th Quarter	November	810,278,129.00	2,872,286,981.00
2014	4th Quarter	December	1,378,391,418.00	4,244,240,442.00
2015	total	null	10,015,700,387.00	31,179,106,555.00
2015	1st Quarter	null	2,072,495,482.00	4,753,120,152.00
2015	1st Quarter	January	451,658,386.00	1,148,915,543.00
2015	1st Quarter	February	766,978,787.00	1,655,262,614.00

Query Settings:

- Properties: Name = Query9, All Properties
- Applied Steps: Replaced Value6, Replaced Value7, Replaced Value8, Replaced Value9, Replaced Value10, Replaced Value11, Replaced Value12, Replaced Value13, Replaced Value14, Replaced Value15, Changed Type, Renamed Columns1, Replaced Value16, Replaced Value17, Replaced Value18

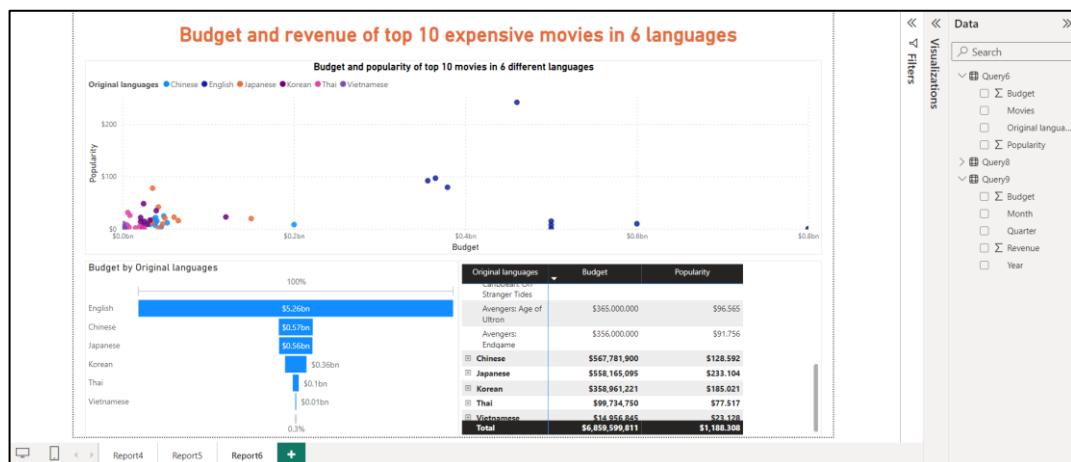


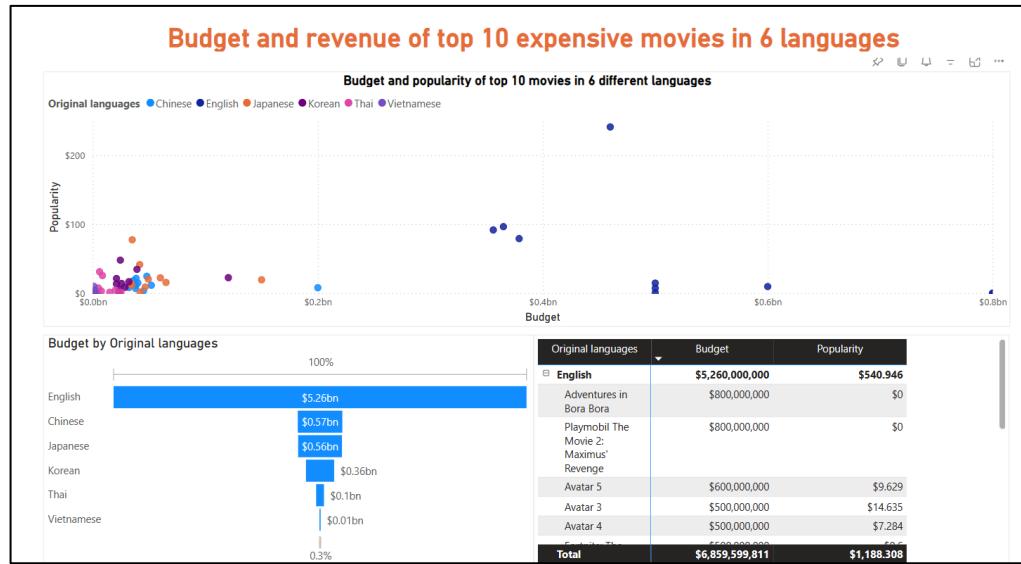
4.2.5. Báo cáo 3

Nội dung: Thông kê ngân sách và điểm thịnh hành của top 10 bộ phim có ngân sách lớn nhất theo 6 ngôn ngữ khác nhau. (câu 6)

Tương tự, ta nhập MDX query 6 và tiến hành biến đổi dữ liệu, bao gồm:

- Đổi tên các cột
- Đotted tên kí hiệu: vi → Vietnamese, en → English, cn → Chinese, th → Thai, ko → Korean và ja → Japanese
- Đổi kiểu dữ liệu của Budget và Popularity sang kiểu fixed decimal number



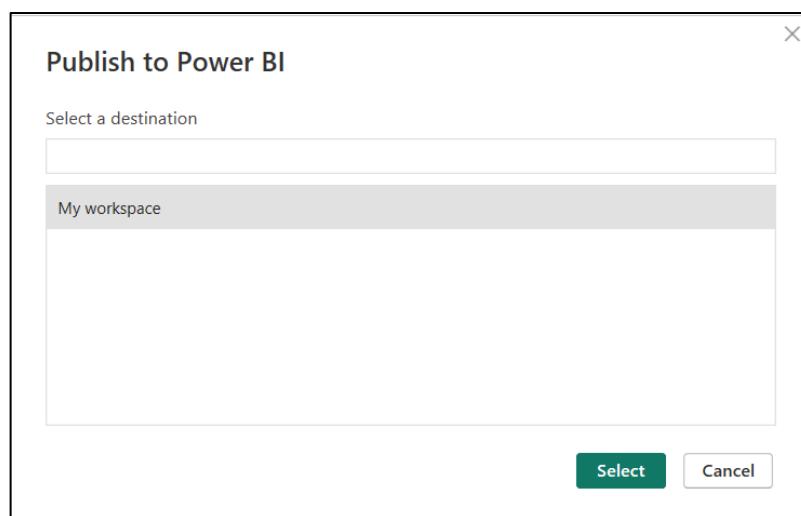
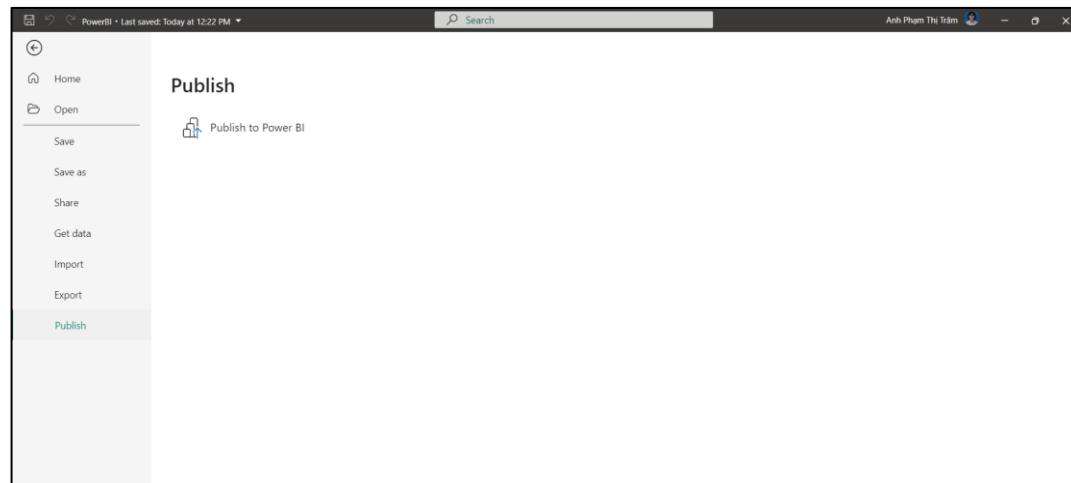


4.2.6. Deploy các báo cáo

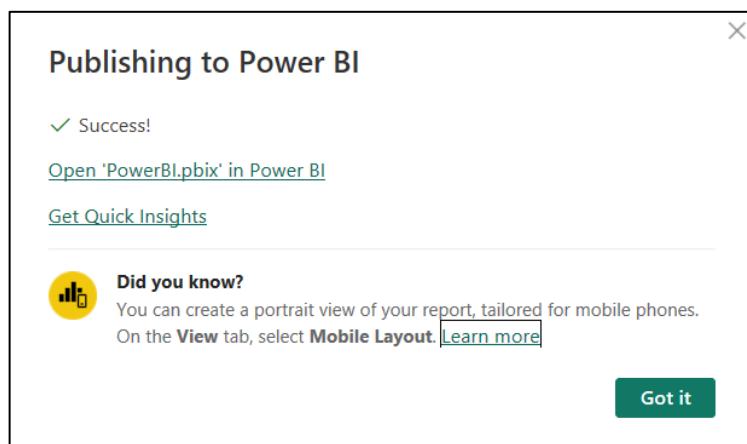
Chọn File → Publish → Publish to Power BI để đưa lên Server.



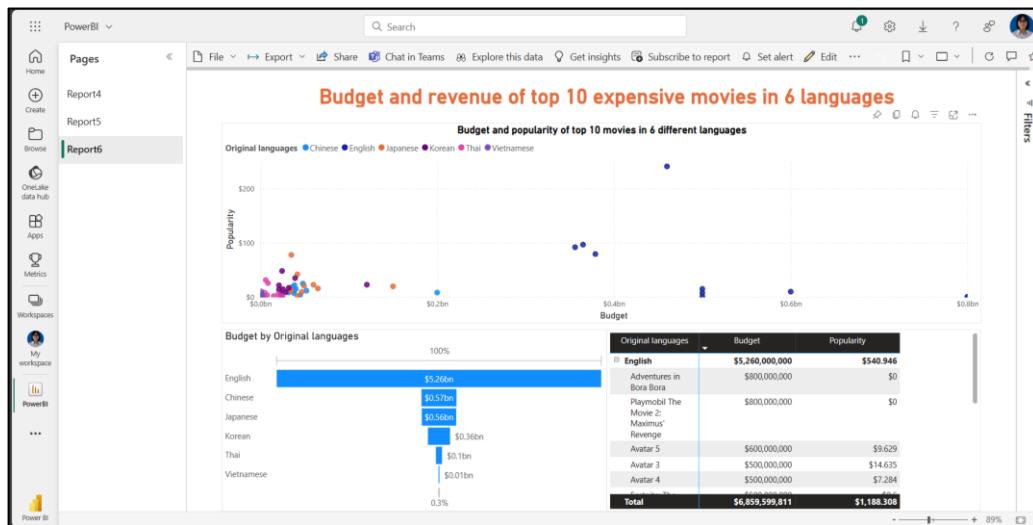
Tại cửa sổ publish to Power BI, ta chọn My workspace. Nhấn select để xác nhận nơi lưu trữ các report trên server.



Sau khi publish thành công, màn hình hiện ra thông báo bên dưới. Chọn Open ReportPowerBI.pbix in Power BI để mở các report mới được publish.



Trang chứa các report tự động mở ra, đăng nhập nếu cần thiết để xem lại các report.



PHẦN 5: QUÁ TRÌNH KHAI THÁC DỮ LIỆU(DATA MINING)